

Titre: Détection d'anomalies dans les photographies du segment
Title: antérieur de l'oeil par apprentissage profond non supervisé

Auteur: Philippe Baumstimler
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Baumstimler, P. (2024). Détection d'anomalies dans les photographies du
Citation: segment antérieur de l'oeil par apprentissage profond non supervisé [Mémoire de
maîtrise, Polytechnique Montréal]. PolyPublie.
<https://publications.polymtl.ca/61604/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/61604/>
PolyPublie URL:

**Directeurs de
recherche:** Lama Séoud
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Détection d'anomalies dans les photographies du segment antérieur de l'œil par
apprentissage profond non supervisé**

PHILIPPE BAUMSTIMLER

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Décembre 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Détection d'anomalies dans les photographies du segment antérieur de l'œil par
apprentissage profond non supervisé**

présenté par **Philippe BAUMSTIMLER**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Guillaume-Alexandre BILODEAU, président

Lama SÉOUD, membre et directrice de recherche

Clément PLAYOUT, membre

DÉDICACE

*À ma famille, pour sa confiance sans limites,
À mes amis, pour leur soutien sans égal,
À vous tous, pour avoir fait de moi, celui que je suis aujourd'hui . . .*

REMERCIEMENTS

Je tiens tout d’abord à remercier ma directrice de recherche, Lama Séoud, pour ces deux années de maîtrise passées au sein de son équipe. Que ce soit dans les moments les plus joyeux ou les plus difficiles, elle a toujours su m’apporter un soutien indéfectible, tant sur le plan technique que moral.

Je souhaite également exprimer ma gratitude envers toute l’équipe de LightX Innovations, en particulier Jean-Mathieu Gagnon, Mathieu Champagne et Sébastien Gagné, qui m’ont fait confiance pour la réalisation de ce projet. Leur optimisme et leur ferveur au quotidien ont su inspirer et motiver mon travail.

Un grand merci à tous les membres du VisionIC, sans qui cette aventure n’aurait pas été la même. Votre joie et votre bonne humeur ont été un véritable réconfort durant la réalisation de mes travaux. Je tiens à remercier chaleureusement Doha Zrouki, Hugo Rodet, Gaspar Faure, Etienne Lesarbeault, Cyprien Arnold, Victor Nogues, Clément Galaup et Yu-Chi Kung. Je suis heureux de vous avoir rencontrés et fier de vous compter parmi mes plus proches amis.

Je remercie également les membres des laboratoires voisins MAGNU et Liv4D, ainsi que toutes les personnes que j’ai pu rencontrer durant ma maîtrise à Polytechnique Montréal. Loin de mes proches, vous avez constitué une seconde grande famille, et je suis convaincu qu’elle continuera de grandir et de s’épanouir.

Enfin, je tiens à exprimer toute ma reconnaissance envers mes parents et mon frère, ainsi qu’à tous mes proches et amis outre-Atlantique. Leur soutien constant, malgré la distance, a été une source précieuse de motivation. C’est en les gardant dans mon cœur que je poursuis mon chemin, avec l’espoir de toujours devenir une meilleure version de moi-même.

RÉSUMÉ

Le diagnostic des pathologies oculaires menant à la perte de vision constitue un enjeu économique et social majeur au Canada. En raison des contraintes liées au manque de personnel spécialisé, notamment dans les régions éloignées, et du manque de ressources financières, la téléophtalmologie offre une solution prometteuse en combinant imagerie médicale et analyse à distance. Largement fondée sur l'imagerie du fond d'œil, qui bénéficie d'un riche corpus d'études scientifiques, elle nécessite cependant des équipements spécialisés et coûteux, freinant ainsi son déploiement à grande échelle.

Face à ces limitations, l'imagerie du segment antérieur de l'œil se présente comme une alternative intéressante. Plus accessible et économique, elle permet l'identification de pathologies grâce à des photographies prises à l'aide d'une lampe à fente. Ce mémoire se concentre sur le développement d'une méthode de détection d'anomalies non supervisée pour les photographies de la conjonctive. En s'appuyant sur des approches par reconstruction, l'objectif est d'identifier toute structure anatomique sortant de la définition de la normalité apprise. À cet effet, nous avons d'abord constitué une base de données adaptée, puis développé SiamAAE, un modèle d'apprentissage profond innovant qui combine reconstruction et auto-distillation surmontant les biais liés à l'apprentissage de l'identité observés dans les modèles de la littérature, tout en conservant les performances en détection d'anomalies de ces dernières.

Cette étude marque une première étape dans l'exploration des méthodes non supervisées de détection d'anomalies appliquée aux photographies du segment antérieur de l'œil. Elle vise à promouvoir la téléophtalmologie et à faciliter la détection précoce des pathologies oculaires, contribuant ainsi à une meilleure prise en charge médicale.

ABSTRACT

Vision loss is a major economic and social challenge in Canada, often caused by undiagnosed or untreated ocular pathologies. Limited access to specialized personnel, particularly in remote areas, and financial constraints have made teleophthalmology a promising solution. This approach combines medical imaging with remote analysis, relying primarily on retinal imaging, a well-researched field in the scientific community. However, the high cost and specialized nature of the required equipment remain significant barriers to widespread adoption.

Given these limitations, anterior eye segment photography offers an attractive alternative. It is more accessible and cost-effective, enabling the detection of pathologies only using slit lamps. This project focuses on developing an unsupervised anomaly detection method in conjunctival images. By leveraging reconstruction-based approaches, the aim is to identify anatomical structures that deviate from the learned definition of normality. To support this work, a tailored database was created, and SiamAAE, an innovative deep learning network, was developed. SiamAAE combines reconstruction and self-distillation approaches, addressing identity-related biases found in existing models while maintaining high performance in anomaly detection.

This study serves as an initial step in developing unsupervised methods for detecting anomalies in anterior eye segment photographs. It aims to advance teleophthalmology and support the early detection of ocular pathologies, ultimately improving medical care in Canada.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
LISTE DES ANNEXES	xvi
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 L'œil	4
2.1.1 Anatomie	4
2.1.2 Modalités d'imagerie et bases de données	7
2.1.3 Pathologies du segment antérieur de l'œil	11
2.2 Détection d'anomalies	13
2.2.1 Approches classiques basées sur l'extraction d'attributs	14
2.2.2 Apprentissage profond supervisé bout-en-bout	17
2.2.3 Apprentissage profond non supervisé ou faiblement supervisé de la normalité	21
2.3 Interprétabilité	28
2.4 Métriques d'évaluation	31
2.4.1 Précision	31
2.4.2 Rappel ou sensibilité	32
2.4.3 Spécificité	32
2.4.4 Aire sous la courbe ROC	32

2.4.5	Aire sous la courbe PR	33
CHAPITRE 3	HYPOTHÈSES ET OBJECTIFS DE RECHERCHE	34
CHAPITRE 4	MÉTHODOLOGIE	36
4.1	Création de la base de données	36
4.1.1	Données existantes	36
4.1.2	Collecte de données	38
4.1.3	Annotations	39
4.1.4	Prétraitement	40
4.2	Notre approche : SiamAAE	41
4.2.1	Module de guidage SSL	42
4.2.2	Module de reconstruction	45
4.2.3	Module antagoniste	46
4.2.4	Inférence	46
CHAPITRE 5	PROTOCOLE D'ENTRAÎNEMENT ET D'ÉVALUATION	47
5.1	Sélection des modèles de la littérature	47
5.2	Configuration unifiée d'entraînement	52
5.3	Procédés d'évaluation	55
5.3.1	Répartition des données	56
5.3.2	Tâche de détection d'anomalies	56
5.3.3	Tâche de reconstruction	57
CHAPITRE 6	RÉSULTATS & DISCUSSION	58
6.1	À propos de la base de données	58
6.2	Analyse et comparaison des capacités de reconstruction et de détection d'anomalies	60
6.2.1	Influence des implémentations	61
6.2.2	Robustesse face à l'apprentissage de l'identité	68
6.3	Analyses complémentaires	74
6.3.1	Influence du module de guidage SSL	74
6.3.2	Analyse qualitative des cartes d'anomalies	78
CHAPITRE 7	CONCLUSION	83
7.1	Synthèse des travaux	83
7.2	Limitations de la solution proposée	84
7.3	Améliorations futures	84

RÉFÉRENCES 86

ANNEXES 98

LISTE DES TABLEAUX

Tableau 6.1	Récapitulatif des photographies exploitables à la détection d'anomalies et leur provenance	58
Tableau 6.2	Résultats de la littérature et de notre approche sur la tâche de détection d'anomalies et de reconstruction sur les photographies du segment antérieur de l'œil	61
Tableau 6.3	Résultats de la littérature et de notre approche sur la tâche de détection d'anomalies et de reconstruction sur les photographies du fond d'œil .	62
Tableau 6.4	Résultats de l'étude du module de guidage SSL sur la tâche de détection d'anomalies et de reconstruction sur les photographies du segment antérieur de l'œil	75

LISTE DES FIGURES

Figure 2.1	Schéma de la segmentation du globe oculaire [1]	5
Figure 2.2	Schéma détaillé du globe oculaire [1]	6
Figure 2.3	Exemple de photographies sur lampe à fente du segment antérieur de l'œil	9
Figure 2.4	Exemple d'une photographie de la rétine centrée sur la fovéa	10
Figure 2.5	Exemple de scan OCT de la rétine centré sur la fovéa	11
Figure 4.1	Système de capture d'images du segment antérieur - LightX Innovations	37
Figure 4.2	Photographies de la conjonctive selon le positionnement de l'œil	37
Figure 4.3	Exemple d'annotations depuis la plateforme d'annotation personnalisée	41
Figure 4.4	Exemple des étapes du prétraitement sur une photographie issue de IOL	42
Figure 4.6	Schéma de l'architecture complète de SiamAAE durant la phase d'entraînement	43
Figure 4.7	Schéma du guidage SSL et visualisation de la fonction de perte L_{SSL} en termes d'espace de représentation	44
Figure 4.8	Exemples de pseudo-anomalies SiamAAE	45
Figure 5.1	Schéma de l'architecture unifiée de l'auto-encodeur	53
Figure 5.2	Schéma de l'architecture du bloc ResNet avec pré-activation	54
Figure 5.3	Schéma de l'architecture du discriminateur	55
Figure 5.4	Exemple de planification cosinus du taux d'apprentissage - de 1.0E-3 à 1.0E-8	56
Figure 6.1	Étude démographique des participants de la collecte de données POLY	59
Figure 6.3	Histogramme de répartition de l'âge des participants à la collecte de données POLY	60
Figure 6.4	Exemple de reconstructions de photographies "normales" du segment antérieur de l'œil	66
Figure 6.5	Exemple de reconstructions de photographies "normales" du fond d'œil	67
Figure 6.6	Histogrammes des erreurs de reconstruction pour l'évaluation de la tendance à l'apprentissage de l'identité	72
Figure 6.8	Exemples de reconstructions pour l'évaluation de la tendance à l'apprentissage de l'identité	73
Figure 6.9	Visualisations UMAP de l'espace de représentation pour différents coefficients λ_{SSL}	77

Figure 6.10	Histogrammes de densité de probabilité du score d'anomalies SiamAAE sur les photographies du segment antérieur de l'œil	81
Figure 6.12	Exemples de reconstructions et de cartes d'erreur pour notre approche SiamAAE, sur des photographies anormales du segment antérieur de l'œil	82

LISTE DES SIGLES ET ABRÉVIATIONS

EOM	École d'optométrie de Montréal
IOL	Institut de l'œil des Laurentides
OCT	Tomographie à cohérence optique - Optical Coherence Tomography
HOG	Histogramme des gradients orientés - Histogram of Oriented Gradients
LBP	Motif binaire local - Local Binary Patterns
SIFT	Transformation de caractéristiques visuelles invariante à l'échelle - Scale-Invariant Features Transform
BRIEF	Caractéristiques élémentaires binaires robustes et indépendantes - Binary Robust Independent Elementary Features
KNN	K plus proches voisins - K Nearest Neighbours
GMM	Modèle de mélange gaussien - Gaussian Mixture Model
SVM	Machine à vecteurs de support - Support Vector Machine
MLP	Perceptron multicouches - Multilayer Perceptron
PCA	Analyse en composantes principales - Principal Component Analysis
t-SNE	t-distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection
AE	Auto-encodeur - AutoEncoder
VAE	Auto-encodeur variationnel - Variational AutoEncoder
GAN	Réseau antagoniste génératif - Generative Adversarial Network
SSL	Apprentissage auto-supervisé - Self-Supervised Learning
DML	Apprentissage profond de métrique - Deep Metric Learning
TP	Vrai positif - True Positive

FP	Faux positif - False Positive
TN	Vrai négatif - True Negative
FN	Faux négatif - False Negative
MAE	Erreur absolue moyenne - Mean Absolute Error
MSE	Erreur quadratique moyenne - Mean Squared Error
DSSIM	Dissimilarité structurelle - Structural Dissimilarity
SSIM	Similarité structurelle - Structural SIMilarity
MS-SSIM	Similarité structurelle multi-échelle - MultiScale Structural SIMilarity
MRPE	Erreur perceptuelle relative moyenne - Mean Relative Perceptual Error
ROC	Fonction d'efficacité du récepteur - Receiver Operating Characteristic
PR	Précision-rappel - Precision-Recall
AUC	Aire sous la courbe - Area Under the Curve
NCS	Non Cliniquement Significatif

LISTE DES ALGORITHMES

1	Algorithme de détection de contours	109
2	Algorithme de détection de cercles	110
3	Algorithme de sélection du meilleur candidat	111
4	Algorithme de rognage circulaire	112

LISTE DES ANNEXES

Annexe A	Approbation éthique - Projet CER-2223-75-D	98
Annexe B	Formulaire d'information et de consentement	100
Annexe C	Prétraitement : description & algorithmes	109

CHAPITRE 1 INTRODUCTION

La perte de vision totale, causée par des pathologies oculaires comme la rétinopathie diabétique, la cataracte, la dégénérescence maculaire ou encore le glaucome, représente un enjeu économique et social important au Canada. Le Conseil canadien des aveugles estime à plus de 8 millions le nombre de personnes vivant au Canada avec une maladie pouvant mener à la perte de vision et évalue à 56 milliards de dollars le coût associé d'ici 2050 [2]. Un diagnostic tardif de ces pathologies est d'autant plus critique qu'il engendre des complications lors de la chirurgie ainsi qu'une hausse importante des coûts de soin. C'est pourquoi l'Association canadienne des optométristes recommande un examen annuel pour les personnes âgées de moins de 20 ou de plus de 65 ans, et un examen tous les deux ans pour celles qui ont entre 20 et 65 ans. Toutefois, depuis le début de la pandémie de la COVID-19, de nombreuses préoccupations concernant la santé oculaire au Canada ont émergé. En effet, le nombre de personnes âgées de plus de 65 ans tend à croître plus vite que l'augmentation annuelle du nombre d'ophtalmologistes. S'ajoute à cela le stress supplémentaire imposé aux professionnels de la santé par la pandémie, l'expansion des déserts médicaux et la baisse des consultations régulières pour des examens de routine en raison des insécurités financières [3].

C'est dans un tel contexte que le concept de téléophtalmologie s'est grandement développé. En connectant à distance plusieurs centres d'optométrie et ophtalmologistes par l'infonuagique, il est possible d'étendre leur champ d'action et ainsi réduire les déserts médicaux. Suite au programme VISION 2020 initié par l'Organisation mondiale de la santé, il a été prouvé que l'utilisation des appareils portatifs, comme les téléphones intelligents, rend possible l'acquisition de données médicales utiles au diagnostic, et cela pour un coût et une main-d'œuvre significativement réduits [4]. Lors d'une étude menée au Royaume-Uni [5], 24 257 examens de glaucome ont été réalisés en deux étapes. Un premier diagnostic a été réalisé directement en clinique par des optométristes, suivi directement d'un second diagnostic réalisé cette fois-ci par des ophtalmologistes spécialistes du glaucome en utilisant la téléophtalmologie. L'analyse effectuée par les spécialistes à distance a permis une réduction de 2,4% du nombre de visites nécessaires, mais surtout 6,5% des patients considérés à risque faible de glaucome par les optométristes se sont révélés être des cas instables à surveiller, et 15% des patients considérés sains ont été diagnostiqués avec un glaucome. La téléophtalmologie permet donc de combler cet écart d'expertise, qui est crucial au suivi de la santé oculaire de la population mondiale. Néanmoins, une évaluation similaire réalisée dans des régions reculées du Kenya [6] montre en réalité un faible taux de confiance au diagnostic par téléophtalmologie, avec 41.3% de

sensibilité, en raison de la mauvaise qualité des images à l'acquisition. Sur 309 images acquises, 74, soit 24% d'entre elles étaient inutilisables. Dans la grande majorité des cas, les images utilisées pour le diagnostic de pathologies oculaires sont des images du fond d'œil, du fait de la présence de nombreux biomarqueurs au niveau de la rétine. Acquérir des images du fond d'œil nécessite un matériel spécialisé adéquat, dispendieux et du personnel qualifié pour la prise de données. Cette limitation nuit radicalement à la démocratisation de la téléophtalmologie, qui repose entièrement sur la qualité des images prises en amont.

Une alternative à l'imagerie du fond d'œil est la photographie du segment antérieur de l'œil. Le segment antérieur de l'œil est généralement observé durant les examens de routine à l'aide d'un appareil à lampe à fente, permettant le diagnostic de nombreuses pathologies, dont celles de la conjonctive. Facile à acquérir, peu coûteuse et au potentiel diagnostique démontré [7], l'imagerie du segment antérieur de l'œil représente une véritable opportunité au déploiement de la téléophtalmologie au Canada.

Ce projet de recherche est réalisé en partenariat avec LightX Innovations, une startup mont-réalaise, qui dispose d'une plateforme d'acquisition de photographies du segment antérieur de l'œil, au moyen d'un système de capture stéréoscopique monté sur un appareil à lampe à fente, permettant de numériser l'examen traditionnel de l'œil. Combinées avec des algorithmes de traitement d'images et d'apprentissage machine, les images issues de leur plateforme d'acquisition pourront être analysées automatiquement et permettre une première étape de préfiltrage des patients à risque pour accélérer et faciliter le processus de diagnostic. Le projet de recherche se divise en deux modules : un premier visant à estimer automatiquement et de façon fiable l'angle entre la cornée et l'iris dans les images issues du test dit de Van Herick, jouant un rôle important dans l'estimation du glaucome à angle fermé, et un second module, se concentrant sur les tests dits de conjonctive, et sur l'apprentissage des structures anatomiques de l'œil à des fins de détection de tout biomarqueur sortant de l'espace de distribution d'un œil sain. Ce mémoire vous présente le travail effectué sur ce second module.

L'ensemble des travaux de la littérature portant sur les outils d'analyse des photographies du segment antérieur de l'œil se fondent sur des approches d'apprentissage profond supervisé, en raison de leur efficacité et de leur rapidité de diagnostic. Cependant, ces approches nécessitent de grandes quantités d'annotations et sont limitées dans leur diagnostic aux données d'entraînement. De plus, elles sont considérées comme des "boîtes noires", dont il est difficile d'estimer les liens entre la prédiction et les causes qui mènent à celle-ci. Dans un contexte médical comme le notre, la notion d'interprétabilité est d'autant plus cruciale que les prédictions des modèles ont un enjeu sur la santé même d'êtres humains. C'est pourquoi nous

nous intéresserons particulièrement à la détection d'anomalies par apprentissage profond non supervisé, dans l'objectif de concevoir un modèle capable de mettre en évidence toute structure anatomique du segment antérieur de l'œil qui pourrait présenter un signe de pathologie, connue ou non, et de potentiellement permettre la découverte de nouveaux biomarqueurs cliniquement significatifs, sans pour autant requérir à des annotations.

La suite du mémoire sera structurée de la manière suivante. Tout d'abord, au travers du **chapitre 2**, nous présenterons les travaux existants de la littérature scientifique en lien avec ce projet. À partir de cette revue de la littérature, les hypothèses et objectifs de recherche seront définis dans le **chapitre 3**. Les choix de conception pour notre modèle de détection d'anomalies appliqué aux photographies du segment antérieur de l'œil seront alors présentés au **chapitre 4**. Ensuite, dans le **chapitre 5**, nous établirons un protocole d'évaluation prenant en compte une sélection des modèles de la littérature pour une comparaison à notre approche. Les expérimentations et les résultats obtenus seront alors exposés et discutés dans le **chapitre 6**. Enfin, le mémoire se conclura par le **chapitre 7**, qui résumera les travaux effectués, leurs limitations et les pistes d'améliorations futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre fait état des connaissances et des travaux existants, en lien avec le projet de recherche présenté dans ce mémoire. Dans un premier temps vous sera présentée une description clinique de l'œil, comprenant son anatomie, ses pathologies communes et les modalités d'imageries existantes pour l'examen. Une description des bases de données publiques existantes et leur utilité au projet seront également suggérées. Ensuite, nous nous attarderons à la revue technique des méthodes de détection d'anomalies par apprentissage machine. Y seront décrites les approches les plus communes de la détection d'anomalies ainsi que leurs avantages et inconvénients respectifs. Le sujet étant peu exploré pour l'application spécifique aux photographies du segment antérieur de l'œil, nous ciblerons autant que possible la revue aux applications du domaine médical et à l'analyse automatique du segment postérieur de l'œil, largement explorés par la communauté scientifique. Enfin, une brève description des métriques d'évaluation couramment utilisées en détection d'anomalies vous sera proposée à des fins pédagogiques.

2.1 L'œil

L'œil est un organe fascinant. Fenêtre sur le monde extérieur, il est également une fenêtre sur les mécanismes internes du corps humain. Avec une certaine expertise et les bons outils, il est possible de dépister de nombreux troubles oculaires menant à la cécité [8], mais également certains troubles systémiques [7]. Dans cette section est détaillée l'anatomie de l'œil, ainsi que les modalités d'imagerie courantes, les bases de données publiques existantes et les troubles oculaires couramment dépistés.

2.1.1 Anatomie

L'œil est un organe permettant la captation de la lumière et sa transformation en information interprétable par le cerveau. Présents par paire et de longueur moyenne de 24mm, les yeux sont situés dans des cavités de part et d'autre du crâne humain, appelées orbites. Contrairement aux représentations généralement faites de l'œil, celui-ci ne prend pas la forme d'une sphère parfaite. Comme indiqué sur la vue de coupe figure 2.1, l'œil est divisé en deux segments, appelés segment postérieur, et segment antérieur, eux-mêmes composés de plusieurs systèmes organiques aux rôles divers et variés.

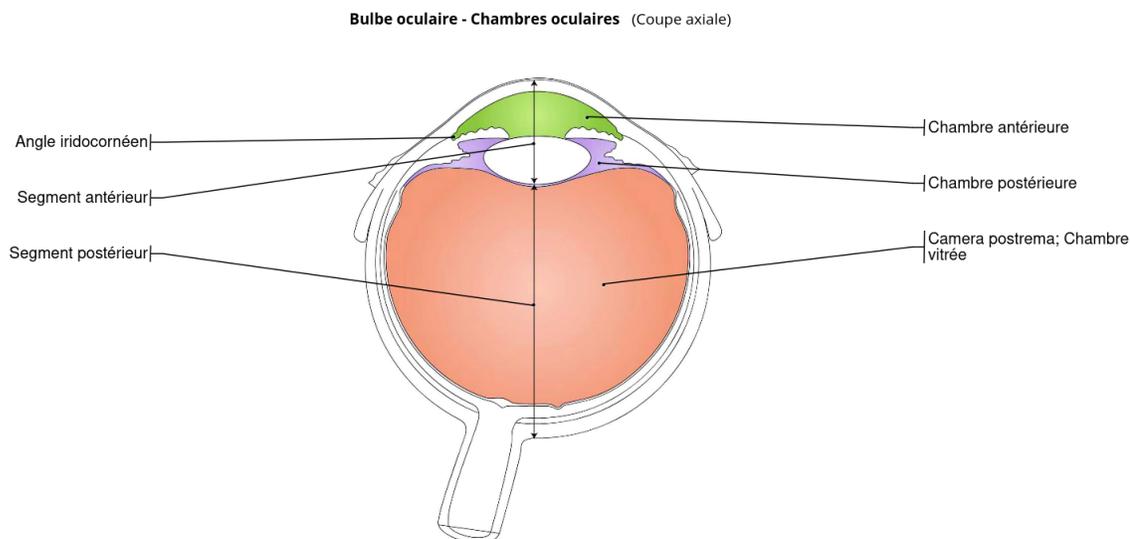


FIGURE 2.1 Schéma de la segmentation du globe oculaire [1]

Segment antérieur

Situé en avant de l'œil, le segment antérieur est composé des paupières, la cornée, l'iris, la sclère, le corps ciliaire, et le cristallin (Figure 2.2).

Les paupières : situées par-dessus le segment antérieur, elles ont pour rôle principal la protection de l'intégrité du globe oculaire, la limitation de la quantité de lumière pénétrant l'œil et au nettoyage des yeux au moyen du film lacrymal. Les paupières sont des structures musculo-membraneuses à grande vascularisation. Elles sont également bordées de deux rangées de cils.

La cornée : partie antérieure transparente du globe oculaire, la cornée est le premier élément avec lequel la lumière entre en contact. Elle n'admet aucune vascularisation et correspond à deux tiers du pouvoir réfractif, le cristallin en comptant pour le dernier tiers. Constituée de cinq couches successives (épithélium cornéen de surface, membrane de Bowman, stroma, membrane de Descemet et endothélium), la cornée a pour fonction principale la convergence des rayons lumineux vers la pupille et le cristallin. Elle joue également un rôle de barrière protectrice entre le monde extérieur et le segment antérieur.

L'iris : l'iris est une membrane pigmentée circulaire, contractile et percée en son centre par un orifice circulaire, la pupille, qui régule la quantité de lumière entrante. L'iris a pour particularité d'avoir des motifs uniques pour chaque individu et chaque œil, et facilement visible, ce qui en fait un bon candidat pour de l'identification biométrique. L'angle entre l'iris et la sclère est appelé angle iridocornéen, où a lieu l'excrétion de l'humeur aqueuse.

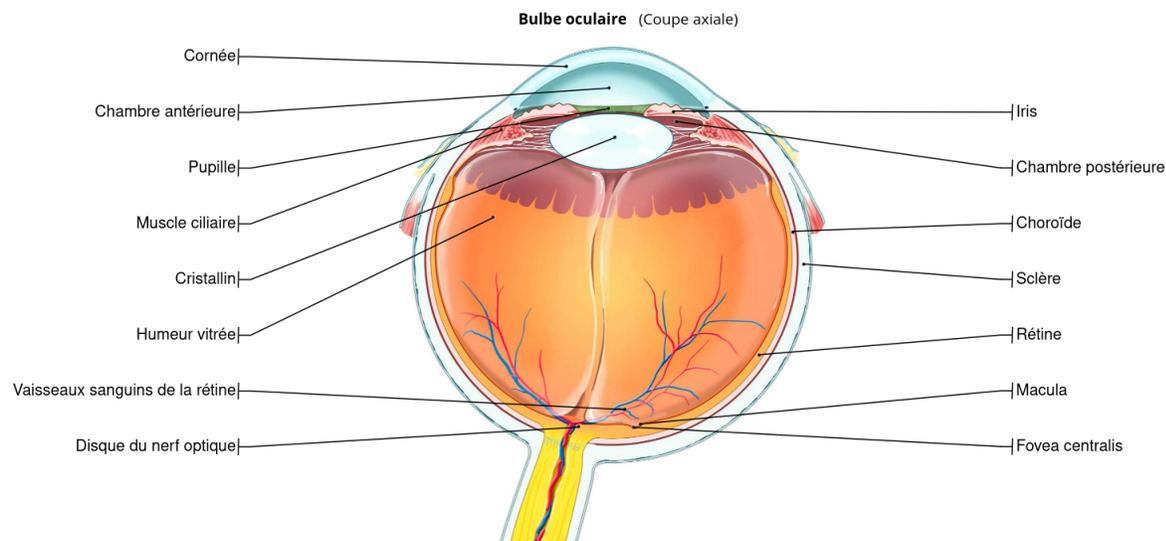


FIGURE 2.2 Schéma détaillé du globe oculaire [1]

La sclère : elle correspond à la surface externe blanche et opaque de l'œil, occupant la majorité de sa surface, antérieure et postérieure. Elle a pour fonction de protéger l'intégrité et maintenir la forme du globe oculaire. Elle est caractérisée par sa forte vascularisation, et elle est traversée à l'arrière par le nerf optique. La sclère antérieure est recouverte d'une fine membrane transparente appelée la conjonctive.

Le corps ciliaire : composé des muscles ciliaires et des procès ciliaires, le corps ciliaire se situe à l'arrière de l'iris. Les muscles ciliaires ont pour rôles de moduler la forme du cristallin pour modifier son pouvoir convergent et permettre l'accommodation de la vision. Les procès ciliaires sont quant à eux responsables de la sécrétion d'humeur aqueuse.

Le cristallin : situé derrière la pupille, le cristallin est une lentille biconvexe, dont le rôle est la convergence des rayons lumineux entrant par la pupille vers la rétine. Transparent, sans vascularisation ni innervation, le cristallin peut adapter sa forme par compression des muscles ciliaires afin d'assurer l'accommodation, c'est-à-dire assurer la netteté de la vision dépendamment de la distance des objets.

Segment postérieur

Situé en arrière de l'œil, le segment postérieur correspond à la section non visible de l'œil. Il est composé de la rétine, de l'humeur vitrée et de la choroïde (Figure 2.2).

La rétine : recouvrant la majeure partie de la surface interne de l'œil, la rétine est la région anatomique sensible à la lumière chargée de la vision. Située entre l'humeur vitrée

et la choroïde, la rétine est constituée d'une couche neurosensorielle, elle-même composée de photorécepteurs. Appelés cônes et bâtonnets, ils sont responsables de la transformation du signal lumineux projeté sur la rétine en un signal électrochimique interprétable par le système cérébral. On distingue trois régions d'intérêts :

- la macula, région centrale de la rétine proche de l'axe optique ;
- la fovéa, région centrale de dépression de la macula où la concentration de cônes est maximale ;
- la papille optique, région d'apparition du nerf optique dépourvue de tout photorécepteur.

L'humeur vitrée : l'humeur vitrée ou corps vitré est une structure gélatineuse transparente qui occupe la majorité du volume de l'œil, située en arrière du cristallin et en avant de la rétine. Formée dès la naissance, l'humeur vitrée est, contrairement à l'humeur aqueuse, stagnante et ne se renouvelle pas.

La choroïde : siège d'une grande partie de la vascularisation de l'œil et située entre la sclère et la rétine, la choroïde est l'interface d'échange dont le rôle premier est l'acheminement des nutriments nécessaires aux différents systèmes du globe oculaire. Elle sert également d'écran de protection des rayons lumineux externes, tout en piégeant les rayons lumineux à l'intérieur, telle une chambre noire.

2.1.2 Modalités d'imagerie et bases de données

Il existe plusieurs modalités d'imagerie, chacune adaptée à l'observation d'une ou plusieurs régions anatomiques de l'œil, permettant de faciliter l'analyse, mais aussi de promouvoir la téléophtalmologie et démocratiser l'accès aux soins oculaires. Sont décrites ici les modalités d'imagerie les plus explorées dans la littérature, comprenant les photographies du segment antérieur, les photographies du fond d'œil et les scans tomographiques par cohérence optique (OCT ou "Optical Coherence Tomography") de la rétine. Une description des bases de données disponibles à la recherche et de leur utilité à notre projet y est également proposée. Ultiment, le projet de recherche se limitera à l'analyse des photographies du segment antérieur de l'œil uniquement.

Photographies du segment antérieur

La lampe à fente, également appelée biomicroscope, est un outil incontournable dans le domaine de l'ophtalmologie. Munie de plusieurs lentilles et d'un faisceau réglable et amovible, elle permet un examen détaillé et structuré des différentes parties de l'œil, allant des régions

les plus externes jusqu'au cristallin. Grâce à cet instrument, il est possible de mener une série de tests ophtalmologiques standards, qui consistent à analyser méthodiquement les régions anatomiques de l'œil pour en évaluer l'état et détecter d'éventuels biomarqueurs pathologiques. L'imagerie du segment antérieur consiste alors simplement en l'acquisition de photographies détaillées des différentes structures anatomiques situées entre la surface externe de l'œil et le cristallin, à partir de la lentille du biomicroscope. Pendant un examen de routine, on dénombre en général 8 tests ophtalmologiques, partant des régions les plus antérieures jusqu'au cristallin (Figure 2.3) :

- Observation des paupières et cils, par lumière diffuse et vue de face pour une vue globale,
- Observation de la sclère et de la conjonctive, avec également une lumière diffuse et vue de face,
- Observation de la cornée, par balayage d'un faisceau étroit pour mettre en évidence sa topologie,
- Observation de la cornée avec fluorescéine et lumière bleue, pour une mise en évidence des dégâts sur la cornée,
- Observation de la chambre antérieure, par projection d'un faisceau circulaire dans la pupille et mettre en évidence la présence de corps étranger,
- Observation de l'iris par transillumination, c'est-à-dire par la projection d'un faisceau circulaire fort dans la pupille et sa réflexion sur l'arrière de l'iris,
- Observation de la marge pupillaire, avec un faisceau large fort sur l'entièreté de l'iris,
- Observation du cristallin, par balayage d'un faisceau étroit pour mettre en évidence sa topologie.

La plupart des bases de données publiques disponibles sont spécialisées dans l'identification d'individus ou dans la segmentation des régions anatomiques de l'œil [9–11]. Les bases de données construites pour l'analyse automatique de pathologies sont généralement privées, en raison du caractère identificatoire (biométrique) des images. Pour en citer l'une d'entre elles, la base de données privée EyePACS [12], regroupe des photographies du segment antérieur de l'œil pour des applications de vision par ordinateur, avec une grande quantité d'annotations couvrant le dossier médical de chaque individu. Cette base de données présente un fort potentiel pour notre projet, mais son utilisation ne sera pas explorée dans ce mémoire, car son acquisition n'a malheureusement pas pu se faire à temps. Nous nous concentrerons donc sur la base de données privée de notre partenaire industriel LightX Innovations, pour laquelle nous avons obtenu un certificat d'approbation éthique auprès du comité d'éthique de la recherche de Polytechnique Montréal (Annexe A). L'acquisition de cette base de données s'est faite à la clinique ophtalmologique de *l'Institut de l'œil des Laurentides (IOL)* et à *l'École d'Optométrie*

de Montréal (EOM), et comprend au total 10 779 photographies des 8 tests ophtalmiques décrits précédemment, et provenant de 255 patients.

Afin de réduire l'étendue du projet, nous nous limiterons à l'étude des photographies issues de l'observation de la conjonctive, de la sclère et des cils et paupières. Ces photographies ont l'avantage d'offrir une vue globale de l'ensemble du segment antérieur et de la majorité des systèmes qui le composent, incluant l'iris, la cornée et la chambre antérieure.

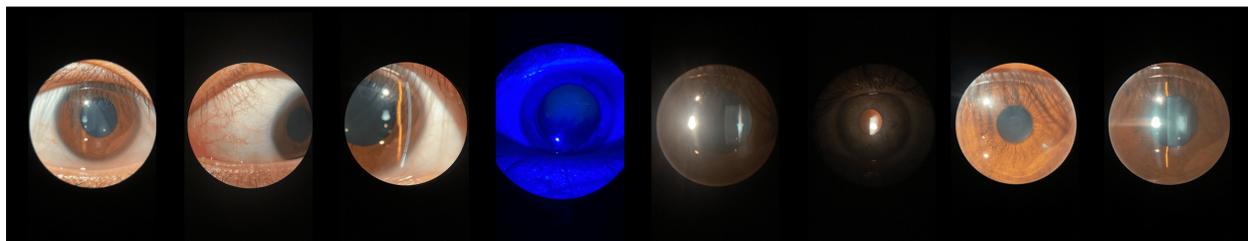


FIGURE 2.3 Exemple de photographies sur lampe à fente du segment antérieur de l'œil

Photographies issues de la base de données du partenaire industriel LightX Innovations - De gauche à droite, observation des paupières et cils ; de la sclère et de la conjonctive ; de la cornée ; de la cornée avec fluorescéine sous lumière bleue ; de la chambre antérieure ; de l'iris par transillumination ; de la marge pupillaire ; du cristallin.

Photographies de la rétine

L'examen du fond d'œil, ou ophtalmoscopie, consiste en l'étude et l'observation des systèmes en arrière du cristallin, en particulier la rétine. L'examen du fond d'œil est réalisé directement au moyen d'un appareil à lampe à fente ou d'un ophtalmoscope de poche, mais nécessite la présence d'un spécialiste et l'instillation de gouttes mydriatiques pour dilater la pupille [8]. La photographie de la rétine ou rétinographie est quant à elle réalisée au moyen d'un rétinographe. Elle nécessite également la présence d'un spécialiste, mais l'application de gouttes mydriatique n'est dans ce cas pas obligatoire. La prise de photos par rétinographe est automatique, et est obtenue par flashes lumineux successifs. Un exemple de photographie rétinienne est donné en figure 2.4.

L'analyse automatique de la rétine est un sujet largement exploré par la communauté scientifique. Il existe par conséquent de nombreuses bases de données publiques disponibles à la recherche. On en retrouve certaines spécialisées dans la gradation de la rétinopathie diabétique [14–18], le diagnostic d'œdème maculaire [15, 16], la segmentation de lésions [19] ou même le diagnostic d'une variété de pathologies de la rétine, au sens large [13]. L'utilisation de ces bases de données, en particulier RFMiD pour une application à la détection d'anomalies,



FIGURE 2.4 Exemple d'une photographie de la rétine centrée sur la fovéa

Image tirée de la base de donnée RFMiD [13]

s'avère intéressante pour l'exploration de la généralisabilité de nos algorithmes.

Scans OCT de la rétine

Système d'imagerie non invasive, les OCT sont utilisés pour l'examen ophtalmique en produisant des images de l'œil en vue de coupe, et ce, en quelques secondes. Elles sont obtenues par rayonnements infrarouges et reposent sur la notion d'interférométrie laser, c'est-à-dire la comparaison des longueurs d'ondes émises par rapport aux longueurs d'onde réfléchies. Les images OCT peuvent être appliquées aux deux segments, antérieur ou postérieur, dépendamment des longueurs d'onde d'émission. Elles sont les plus couramment utilisées pour l'examen de la rétine (Figure 2.5). Les images OCT détiennent une information de volume, que la photographie rétinienne omet. Elles peuvent permettre la reconstruction 3D des tissus formant le segment antérieur.

Tout comme la rétinographie, l'analyse automatique des scans OCT de la rétine est un champ très exploré par la communauté scientifique. Il existe donc plusieurs bases de données publiques disponibles à la recherche, pour le diagnostic de pathologies oculaires, comme la dégénérescence maculaire ou la rétinopathie diabétique [18, 20]. Cependant, les scans OCT sont une modalité d'imagerie très différente de la rétinographie ou de la photographie du segment antérieur. Ces deux dernières sont prises en RGB et présentent des structures anatomiques similaires (vascularisation). Les bases de données OCT ne seront donc pas explorées dans nos travaux.

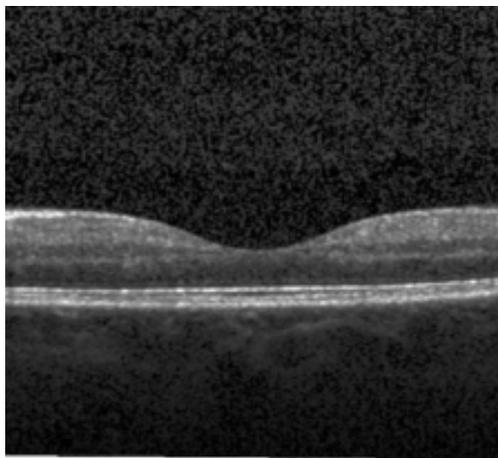


FIGURE 2.5 Exemple de scan OCT de la rétine centré sur la fovéa

Image tirée de la base de donnée MedMNIST [18]

2.1.3 Pathologies du segment antérieur de l'œil

Nos travaux se penchent sur un champ spécifique de l'analyse automatique, qui est la détection d'anomalies. L'objet de l'analyse ne porte donc pas sur une pathologie en particulier, mais sur l'ensemble des pathologies possibles du segment antérieur. C'est pourquoi nous proposons une brève description des pathologies les plus communes, observables à l'examen du segment antérieur.

Cataracte

La cataracte [21] est l'une des principales causes de la perte de vision au Canada [3], et nécessite donc un suivi important. Elle se traduit chez l'individu par une opacité du cristallin congénitale ou dégénérative, dont l'apparition s'étend souvent sur plusieurs années, due au vieillissement. La cataracte est indolore et engendre la perte progressive de la vision, jusqu'à la possible cécité.

La cataracte est diagnostiquée lors d'une ophtalmoscopie réalisée à la lampe à fente. Il en existe de nombreux types, dont le diagnostic s'établit essentiellement selon le degré d'opacité du cristallin, affectant l'acuité visuelle.

Glaucome

Alors que la cataracte est la principale cause de cécité réversible, le glaucome quant est la première cause de cécité irréversible au Canada [3,22]. Sans un dépistage adéquat et une prise en charge précoce, le glaucome affecte de manière permanente la vision. Il s'agit d'un groupe de pathologies oculaires graves touchant le nerf optique, ayant notamment comme facteur de risque l'hypertension oculaire. On catégorise généralement le glaucome en 2 grandes catégories, soit le glaucome à angle ouvert et le glaucome à angle fermé, elles-mêmes étant classées en deux sous catégories, primitives (causes inconnues) ou secondaires (causes connues).

En général, le diagnostic du glaucome nécessite une gonioscopie, examen de vérification de l'angle iridocornéen par instillation de gouttes anesthésiantes et l'utilisation d'une lentille déposée sur l'œil. La mesure de la pression intraoculaire est également un indicateur important dans les cas extrêmes, tout comme l'examen de la présence de dommage sur le nerf optique par ophtalmoscopie. Néanmoins, l'évaluation grossière du risque d'un glaucome à angle fermé peut se faire au moyen d'un appareil à lampe à fente, par test dit de "Van Herick". Il consiste en l'estimation de l'angle iridocornéen par la comparaison de la fente reflétée sur la cornée, et son image sur l'iris. Comme expliqué en introduction, l'étude spécifique de l'exploitation des photographies du segment antérieur pour l'évaluation du glaucome à angle fermé fait l'objet de travaux de recherche à part entière, par un autre membre de notre équipe de recherche.

Troubles cornéens

Les troubles cornéens [23] correspondent à l'ensemble des pathologies et lésions situées au niveau de la cornée. En général, ces troubles peuvent autant engendrer des douleurs, que des larmoiements ou une baisse d'acuité visuelle.

L'apparition de ces pathologies peut être infectieuse ou inflammatoire (ulcères, kératite), dégénérative (kératocône), héréditaire (syndrome de Cogan) ou bien encore traumatique. Elles sont diagnostiquées par examen à la lampe à fente, et nécessitent dans la plupart des cas l'administration de fluorescéine, un agent colorant permettant la mise en évidence de lésions et dommages sur la cornée.

Troubles conjonctivaux et scléaux

Comme leurs noms l'indiquent, les troubles conjonctivaux et scléaux [24] regroupent l'ensemble des pathologies touchant la conjonctive ou la sclère. Les plus fréquents sont infectieux ou inflammatoires.

Conjonctivite : La conjonctivite est une inflammation de la conjonctive, pouvant être causée par une infection (virale ou bactérienne), des allergies ou à l'exposition à des produits irritants. Elle peut entraîner une dilatation vasculaire conjonctivale, l'augmentation des larmoiements et de sécrétions épaisses pouvant troubler la vision, et se caractérise souvent par un œil rouge. Son diagnostic repose en général sur l'examen de l'œil à la lampe à fente avec fluorescéine et la description des symptômes de l'individu atteint.

Épisclérite : l'épisclérite est une inflammation de l'épisclère, membrane vasculaire située entre la conjonctive et la sclère. Il s'agit d'une inflammation récurrente et habituellement idiopathique, c'est-à-dire qui ne provient ni n'est la conséquence d'une autre. Contrairement à la conjonctivite, l'épisclérite se traduit par l'apparition de rougeurs localisées au niveau de la sclère et une diminution des larmoiements. Bien qu'elle ne menace pas la vision, elle peut néanmoins être le résultat d'une maladie auto-immune, soit la perte de tolérance immunologique du corps humain face à son propre organisme. Le diagnostic de pathologies systémiques graves à la suite d'une épisclérite représente environ 15% des patients [25]. Un diagnostic précoce en est d'autant plus important. Son diagnostic est purement clinique, et par examen à la lampe à fente. Elle peut être différenciée de la sclérite par instillation de phényléphrine. La phényléphrine colore les vaisseaux conjonctivaux et épiscléaux sans toucher à la sclère.

Sclérite : finalement, la sclérite est une inflammation grave de la sclère, principalement due à des pathologies systémiques, en particulier des maladies rhumatologiques. Elle engendre des douleurs intenses dans des régions locales ou sur l'entièreté de l'œil, et la formation de lésions hyperémiques (accumulation de sang) sous la conjonctive bulbaire. Dans les cas les plus graves, la sclérite peut être nécrosante, et induire la perte de l'œil par perforation du globe. Le diagnostic de la sclérite se fait par évaluation clinique et par examen à la lampe à fente, les lésions hyperémiques apparaissant comme des régions violacées.

D'autres pathologies bénignes de la conjonctive peuvent être recensées au moyen d'un examen à la lampe à fente, comme la pinguécula ou le ptérygion. Il s'agit de proliférations de la conjonctive en général à faible risque. Le ptérygion peut s'étendre et déformer la cornée, pouvant induire une diminution de l'acuité visuelle. De même, des hémorragies sous-conjonctivales peuvent être observées, généralement causées par des traumatismes locaux mineurs, mais ne présentent aucun risque pathologique, sauf en cas de lésions oculaires.

2.2 Détection d'anomalies

En analyse de données, la détection d'anomalies est la tâche qui consiste à identifier tout évènement, objet ou observation sortant de la distribution moyenne des données en entrée.

En général rares et proches de la distribution d'entrée, la détection automatique des données anormales peut s'avérer être une tâche complexe présente dans de nombreux domaines, comme la vidéosurveillance [26, 27], la détection de défaut en industrie [28, 29] ou encore, dans notre cas, le diagnostic médical [30, 31]. Dans cette section seront explicitées en détail les différentes approches de détection automatique d'anomalies de la littérature, en particulier appliquées au domaine de la vision par ordinateur et au domaine médical. Elles sont regroupées en trois catégories, chacune d'entre elles faisant l'objet d'une sous-section ci-dessous. Tout d'abord, on distingue les approches en deux temps, se fondant sur un premier travail d'extraction des attributs de l'image, suivie d'une étape de discrimination des dits attributs établis en amont. Ensuite s'en vient un deuxième groupe d'approches, qui rassemblent les deux précédentes étapes en une seule étape de détection d'anomalies bout-en-bout, au travers de l'entraînement de réseaux convolutifs entièrement supervisé. Enfin, plus récemment, la grandissante effervescence autour des méthodes d'apprentissage profond non supervisé et faiblement supervisé a vu naître un troisième groupe d'approches, visant à l'apprentissage de la normalité, souvent au moyen de tâches connexes.

2.2.1 Approches classiques basées sur l'extraction d'attributs

Les approches par modélisation de vecteurs de caractéristiques décomposent la détection d'anomalies en deux temps. Elles se fondent sur l'hypothèse qu'il existe une sous-représentation pertinente de l'entrée, permettant de distinguer les données normales aux données anormales. Il s'agit donc, dans un premier temps, d'établir une méthode d'extraction d'attributs caractéristiques à l'image d'entrée puis, dans un second temps, d'établir une méthode de discrimination de ces derniers.

Extraction d'attributs

En temps normal, une image est représentée sous la forme de tableaux d'entiers à trois dimensions. Analyser une image telle qu'elle pose problème en raison du fléau de la dimension : pour un trop grand nombre d'attributs, l'espace de représentation croît rapidement, au point de rendre éparse chaque donnée, et rendant inefficace, voire nulle, toute méthode d'analyse ou d'apprentissage. L'objectif est donc d'établir une fonction de passage $f : \mathbb{R}^D \mapsto \mathbb{R}^A$ d'un espace de départ de haute dimension \mathbb{R}^D à un espace d'arrivée de plus faible dimension \mathbb{R}^A avec $D \gg A$.

Traditionnellement, la description d'une image se fait par l'extraction d'attributs explicites comme des caractéristiques structurelles, géométriques ou colorimétriques. Un descripteur simpliste est la représentation par histogramme de couleurs pour la reconnaissance d'ob-

jets [32], particulièrement adapté aux espaces contrôlés, industriels par exemple, où les principales anomalies sont des forts contrastes de couleurs. Pour une description locale plus fine, on retrouve d'autres descripteurs, comme les histogrammes de gradient orienté (HOG ou "Histogram of Oriented Gradients") [33] et les motifs binaires locaux (LBP ou "Local Binary Patterns") [34] qui s'intéressent aux attributs structurels et de textures de l'image par l'estimation des gradients locaux. Combinées, ces deux méthodes se sont notamment avérées très puissantes pour la classification et la détection d'objet dans des images naturelles [35]. Enfin, une autre alternative sont les descripteurs de points clés, comme SIFT [36] et BRIEF [37]. Ils permettent de décrire des régions très précises d'une image, mais nécessitent en général une première étape supplémentaire de détection de points clés. Ces approches présentent l'avantage d'être conceptuellement compréhensibles pour un être humain. Cependant, elles nécessitent pour la majorité un environnement très contrôlé où les régions d'intérêts sont connues à l'avance et peuvent être computationnellement lourdes. De plus, une application au domaine médical nécessite une bonne capacité de généralisation aux données inconnues et une forte sensibilité aux légères variations, rendant l'utilisation des descripteurs traditionnels difficile, voire impossible, dans la majorité des cas.

Plutôt que d'établir une représentation arbitraire, il paraît pertinent de vouloir extraire directement une représentation spécifique à la base de données et la tâche considérée. En se servant des réseaux de neurones profonds préentraînés, il est possible d'obtenir une représentation abstraite d'une image. L'information disponible dans les couches intermédiaires de ces réseaux se veut spécifique et pertinente à la généralisation à d'autres tâches que celle d'entraînement. Les travaux menés par Zhang et. al. [38] témoignent de la capacité émergente des caractéristiques profondes comme métrique perceptuelle équivalente à la perception humaine, permettant d'accomplir des tâches de super-résolution [39] mais aussi de détection d'anomalies [40]. Le point central réside dans le choix de la base de données de préentraînement, en général ImageNet [41]. Celle-ci contient un large nombre d'images naturelles, permettant à ces réseaux d'apprendre une sémantique globale d'une image. De plus, il a été montré que cette capacité est indépendante de l'architecture choisie, et se partage entre de nombreux réseaux comme SqueezeNet [42], AlexNet [43], VGG [44] ou ResNet [45], dont l'utilisation et la disponibilité sont largement facilitées de nos jours. Le préentraînement peut suffire de lui-même dans les cas simples, mais un raffinement sur la nouvelle base de données permet en général d'obtenir une représentation encore plus fine et adaptée à la tâche considérée.

Discrimination des attributs

L'une des premières approches considérées en détection d'anomalies est l'approche par estimation de densité de probabilité. Par définition, il est considéré que l'ensemble des données anormales se situeront dans les régions à faible densité, en raison de leur rareté, à l'inverse des données normales. L'objectif de ces méthodes est donc d'estimer la densité de probabilité des données en entrée et de discriminer toutes aberrations n'appartenant pas aux régions de densité élevées. Une première méthode d'estimation de densité pour la détection d'anomalies est celle des k plus proches voisins (KNN ou "K-Nearest Neighbors") [46]. En considérant l'ensemble des vecteurs caractéristiques du jeu de données et leur classe associée, un nouvel échantillon est classifié en accord avec la classe majoritaire de ses k plus proches voisins. Les anomalies étant par hypothèse dans les régions à plus faible densité, il est possible d'estimer un seuil de similarité moyen des k plus proches voisins permettant de maximiser la discrimination des valeurs aberrantes. Une deuxième approche simple est celle de l'algorithme des K-moyennes (ou "K-Means") [47]. Il s'agit d'un algorithme de regroupement, qui, pour un nombre de groupes K arbitraire, cherche à estimer les centroïdes respectifs aux K groupes ayant le plus de similarité, parmi la base d'attributs. Dès lors, les centroïdes définis, il est possible de procéder à la discrimination des anomalies par un seuillage autour de ces derniers. Enfin, une dernière famille d'approches très populaire dans le domaine sont les modèles de mélanges gaussiens (GMM ou "Gaussian Mixture Models") [48]. Il s'agit de modèles statistiques qui visent à estimer la distribution des représentations, comme la somme de plusieurs gaussiennes, à moyenne, variance et amplitude variable. Comme pour l'algorithme K-moyennes, l'espace de représentation est ainsi regroupé en clusters centrés autour de gaussiennes, qui modélisent un espace de probabilité. Si la probabilité d'un nouvel échantillon est faible par rapport à un seuil à déterminer, alors ce dernier sera considéré comme une anomalie. Le nombre de gaussiennes est un hyperparamètre à définir avant entraînement.

Un second paradigme très apprécié dans les approches populaires de la détection d'anomalies est la classification à classe unique (OCC ou "One-class classification"). Ici, l'hypothèse réalisée est que les données normales appartiennent à une seule et unique classe compacte dans l'espace de représentation, auxquelles les données anormales ne se conforment pas. La majorité d'entre elles sont fondées sur le concept de machine à vecteurs de support (SVM ou "Support Vector Machine") [49]. Classiquement, l'algorithme SVM vise à établir un hyperplan dans l'espace de représentation qui maximise la distance à la frontière entre les données de la classe unique et le reste selon une certaine marge définie en amont. Dès lors, un nouvel échantillon sera classifié selon sa position par rapport à l'hyperplan trouvé. Une méthode simple qui toutefois évite de calculer une à une les distances entre l'ensemble des échantillons. L'uti-

lisation de noyaux permet de généraliser l'utilisation des SVM aux cas non linéaires, avec la définition d'hyperplans plus adaptés aux problèmes les plus complexes. Une variante existe dans le cas non supervisé, c'est-à-dire en l'absence de données anormales à l'entraînement. Il s'agit des Support Vector Data Description (SVDD). À la différence des SVM, on cherche à minimiser la distance à la frontière de l'hypersphère qui englobe l'espace de représentation des données normales. Dans les deux cas, il est nécessaire d'établir quel hyperplan permettra de mieux séparer les représentations, que ce soit avec le choix du noyau ou des critères de calcul de distance à la marge. Rien ne garantit cependant que la représentation permettra une telle séparation de l'espace latent.

De nombreux travaux se sont attardés à l'application de ces méthodes sur des images ophtalmiques. On retrouve des applications pour le diagnostic de la cataracte et du gérontoxon [50], la classification post-chirurgie du ptérygion [51] ou en encore la classification du film lacrymal pour le diagnostic des yeux secs [52]. Le point commun dans ces travaux est l'importance et la minutie accordées à la définition de la représentation, basée sur des facteurs médicaux connus et fiables, puis à l'utilisation des outils d'apprentissage machine, en particulier les SVM pour le travail de discrimination. Néanmoins, les cas cités font le diagnostic d'un sous-ensemble précis des pathologies de l'œil, et dont les biomarqueurs sont bien connus et explicites. En réalité, malgré leur rareté, les biomarqueurs synonymes de pathologies sont, dans la majorité des cas, suffisamment subtils pour appartenir à l'espace de probabilité des données normales, sans compter l'inconnu.

2.2.2 Apprentissage profond supervisé bout-en-bout

La détection d'anomalies en deux temps présentée dans la section précédente peut être combinée en une seule et unique étape. On parle alors de modèles de détection d'anomalies bout-en-bout. Il s'agit ici d'apprendre directement à extraire les attributs pertinents de l'image en même temps que l'apprentissage d'un score d'anomalies. Soit une base de données de départ D , formées de paires image et annotation associée $(x_n, y_n)_{n \in \mathbb{N}} \in \mathcal{X} \times \mathcal{Y}$ tel que $\mathbb{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$. Il est question de trouver l'ensemble des paramètres optimaux θ^* de $f_\theta : X \mapsto \mathcal{Y}$, permettant de maximiser une fonction objectif connue $l : \mathbb{Y} \times \mathbb{Y} \mapsto \mathbb{R}$, tel que

$$\theta^* = \arg \max_{\theta} \sum_{(x,y) \in \mathbb{D}} l(f_\theta(x), y) \quad (2.1)$$

De nos jours, ces approches se font dans la majorité des cas par apprentissage profond, où f_θ représente un réseau de neurones et f_{θ^*} sa version entraînée. L'atteinte des objectifs réside

donc fondamentalement dans le choix de l'architecture du réseau et de l'objectif d'entraînement l , permettant de guider l'apprentissage vers une solution optimale.

Classification

Les méthodes par classification partent de l'hypothèse que chaque échantillon de la base de données appartient à une classe $k \in \mathbb{N}$ bien défini. Usuellement en détection d'anomalies, on se retrouve dans des situations de classification binaire avec un groupe de données "normal" et un autre "anormal". Elles peuvent aussi se dériver sous des formes plus subtiles, avec des labels précis, associés à une voire plusieurs pathologies. De façon générale, on souhaite trouver la fonction $f_\theta : \mathcal{X} \mapsto \mathbb{R}^K$, estimant la probabilité d'un échantillon d'appartenir à la classe $k \in \mathbb{N}^*$. Dans le cadre d'une application médicale, c'est un moyen efficace de réaliser une étape de préfiltrage des patients et de diagnostic des cas les plus à risque.

Il s'agit d'un sujet largement exploré dans la littérature, également pour les images du segment antérieur de l'œil. Xu et. al. [53] ont proposé une revue des algorithmes d'intelligence artificielle existants, comprenant un large éventail de méthodes d'apprentissage machine et d'apprentissage profond, appliquées au diagnostic des pathologies du segment antérieur. Seules quelques-unes, cependant, traitent des images par lampe à fentes. On retrouve tout d'abord les travaux de Fang et. al. [54], traitant du diagnostic du ptérygion. L'approche proposée est une approche bout-à-bout entièrement supervisée, fondée sur une classification binaire de base par extraction de vecteurs caractéristiques VGG [44] suivie d'une tête de prédiction perceptron multicouche (MLP ou "Multilayer Perceptron") [55]. La restriction à une seule pathologie permet d'obtenir des modèles spécialistes, ayant de hautes capacités de prédiction. La puissance des modèles spécialistes se confirme avec l'étude menée par Kuo et. al. [56]. Au travers d'un cadre expérimental identique ont été raffinées plusieurs architectures pré-entraînées sur ImageNet, sur le diagnostic de la kératite bactériologique à partir d'images de lampe à fentes, contenant un ensemble de données normales, de kératites bactériologiques et de kératites autres. Malgré des performances de même ordre de grandeur, il en ressort que le choix de l'architecture peut impacter significativement le processus de diagnostic des modèles, avec par exemple une plus grande sensibilité pour les modèles ResNet [45] et DenseNet [57], et une plus grande spécificité pour les modèles EfficiencyNet [58].

Malgré l'efficacité globale des modèles spécialistes, ils restent néanmoins limités dans leur prédiction à la pathologie en question, ne prenant pas en compte la potentielle existence d'autres pathologies qui seraient considérées par ces modèles comme de faux négatifs. Pour répondre à cette éventualité, une forme de généralisation est proposée par Li et. al. [59], en ajoutant un label "autres" à la prédiction de la kératite, en plus des labels "présence de

kératite" et "normal". De fait, en plus d'avoir une spécialité, le modèle est capable d'avoir une sensibilité sur d'autres biomarqueurs pathologiques récurrents. Cette approche peut encore être raffinée, aux moyens d'une classification par niveaux comme proposée par Gu et. al. [60]. Trois niveaux d'abstraction des annotations ont été définis, allant du plus général ("normal" contre "anormal") vers le plus spécifique (31 labels de pathologies). Par une approche hiérarchique, leur méthode permet l'obtention d'un modèle pluridisciplinaire qui couvre un large éventail de pathologies de la cornée. Un œil est une structure complexe, formée de plusieurs régions anatomiques distinctes, desquelles peut découler l'apparition d'un ou plusieurs marqueurs pathologiques qui peuvent leur être spécifiques. Pour prendre en compte cette notion de région anatomique, le pipeline Visionome [61] reprend la notion de prédiction par niveau et y intègre une étape de segmentation de région d'intérêts. De ce fait, la prédiction ainsi réalisée se fait non seulement au niveau de l'image (prédiction "normal" ou "anormal"), mais aussi localement au niveau de chaque région et sous-région anatomique.

Jusqu'à présent, les approches présentées sur le segment antérieur de l'œil se concentraient sur des pathologies qui lui sont spécifiques et dont les signes biologiques sont bien connus. En réalité, le segment antérieur peut être révélateur de troubles systémiques, observables également dans les images du fond de l'œil, mais dont les facteurs sur la paroi externe de l'œil sont peu connus. Cette piste a été explorée au travers des travaux de Babenko et. al. [7], par apprentissage profond supervisé d'un modèle Inception-V3 [62] sur la base de données EyePACS [12]. Les auteurs ont démontré la capacité du réseau à détecter efficacement des cas de rétinopathie diabétique, de mauvais contrôle de la glycémie, mais aussi du cholestérol, et cela au travers de populations ethniques et d'âges variés. L'approche par apprentissage profond pour la classification des taux de cholestérol et de glycémie a largement surpassé les méthodes classiques de régression linéaire sur les données démographiques disponibles. L'intuition derrière la capacité prédictive de leur modèle est la capacité des réseaux convolutifs à se servir de l'ensemble des informations utiles de l'image, en particulier en ciblant les régions anatomiques pertinentes au diagnostic pour la tâche considérée, que ce soit la pupille pour le diagnostic de la cataracte, ou la sclère, l'iris et les paupières pour d'autres.

La tâche de classification en apprentissage profond présente l'avantage de se servir efficacement d'annotations d'experts pour guider l'entraînement. Il en résulte des modèles spécialistes, capables d'atteindre de très hautes performances sur les pathologies individuelles. Le cas de l'inconnu pose cependant problème. La prédiction est limitée aux annotations disponibles à l'entraînement, et ne permet pas une généralisation à l'ensemble des anomalies possibles. De plus, l'obtention d'annotations médicales n'est pas une tâche aisée, qui requiert le travail de spécialistes pendant une période prolongée et des données difficilement partageables considérant leur caractère identificatoire.

Segmentation

La tâche de classification vue précédemment n’offre en soi aucune information d’ordre spatial des anomalies. Or, dans l’optique d’une application au domaine médical, il est intéressant de connaître la région de l’image associée à la prédiction, que ce soit pour guider le professionnel dans son diagnostic, mais surtout pour valider que le modèle se réfère effectivement aux régions anatomiques liées à la pathologie prédite. Plutôt que de prédire un unique score pour toute l’image, on prédit un score par pixel, avec $f_\theta : \mathcal{X} \mapsto \mathbb{R}^{K \times H \times W}$ où H et W correspondent aux dimensions spatiales de l’image et K au nombre de classes à prédire. Cela devient donc une tâche de segmentation.

L’architecture la plus répandue pour la segmentation sémantique est l’architecture UNet [63], reprenant le principe des auto-encodeurs. Une image donnée en entrée d’un encodeur est successivement sous-échantillonnée, jusqu’à obtenir une représentation $z \in \mathbb{R}^P$ où P est très petit devant les dimensions de l’image d’entrée. Cette représentation est ensuite suréchantillonnée au travers d’un décodeur symétrique, jusqu’à revenir aux dimensions souhaitées $\mathbb{R}^{K \times H \times W}$. La particularité de cette architecture est l’utilisation de connexions résiduelles à chaque niveau de l’encodeur vers le décodeur, facilitant la descente de gradient à l’optimisation et permettant le transfert d’information sur plusieurs niveaux d’abstraction. L’utilisation de cette architecture s’est prouvée efficace sur des photographies du segment antérieur, en particulier pour la segmentation du ptérygion [64]. Une variante de UNet pour les images médicales, UNet++ [65], visant à améliorer le transfert d’informations dans les connexions résiduelles par ajout de blocs convolutifs intermédiaires, a permis une amélioration de la segmentation du ptérygion [66]. On retrouve plus communément des travaux pour la segmentation de régions anatomiques, comme l’iris, la sclère ou la pupille, pour du suivi du regard [67] ou de la reconnaissance biométrique [68–70]. De nombreux travaux ont cependant été réalisés sur les images du fond d’œil, que ce soit pour la segmentation des vaisseaux sanguins et du disque optique [71, 72] mais aussi pour le diagnostic de la rétinopathie diabétique [73].

La tâche de segmentation a l’avantage de fournir non seulement l’information de la présence d’une pathologie dans une image, mais elle fournit aussi une information spatiale importante pour le traçage de cette dernière par des spécialistes. Le problème reste cependant l’obtention d’annotations sémantiques pour l’apprentissage et l’évaluation. Là où l’annotation niveau image peut déjà s’avérer longue et fastidieuse dans le domaine médical, l’annotation niveau pixel est d’autant plus difficile qu’elle nécessite l’attention prolongée de spécialistes pour s’assurer de leur validité. C’est pourquoi la plupart des bases de données médicales publiques pour la détection d’anomalies ne contiennent pas d’annotations au niveau du pixel.

2.2.3 Apprentissage profond non supervisé ou faiblement supervisé de la normalité

L'entraînement de réseaux de neurones supervisés nécessite une quantité non négligeable d'annotations. L'obtention de ces dernières est une tâche pénible, et qui s'avère d'autant plus difficile que les données anormales sont par définition rares. C'est là qu'intervient l'apprentissage non supervisé. Au lieu de se limiter à un ensemble d'annotations, l'objectif est d'apprendre la notion de normalité, au moyen d'un objectif connexe, qui n'a aucun lien direct avec la détection d'anomalies. Tout échantillon sortant de la normalité apprise par le réseau se verra classifié comme anormal. De fait, on se débarrasse de la contrainte des annotations, et on se limite à l'entraînement uniquement aux données normales. À notre connaissance, il n'existe aucun travaux sur l'apprentissage connexe de la normalité appliqué aux photographies du segment antérieur de l'œil. Toutefois, c'est un champ de recherche largement exploré en imagerie médicale, ce sur quoi cette sous-section se concentrera. Ces approches peuvent être divisées en trois sous-catégories, chacune desquelles étant développées dans cette sous-section. La première catégorie regroupe l'ensemble des méthodes d'apprentissage de la normalité par une tâche connexe de reconstruction, allant des auto-encodeurs classiques et des modèles variationnels jusqu'aux réseaux antagonistes génératifs. La deuxième catégorie, quant à elle, rassemble les approches faiblement supervisées par génération de pseudo-anomalies. Enfin, la dernière catégorie traite des méthodes de détection d'anomalies par auto-supervision, au travers des modèles d'apprentissage profond de métriques et d'auto-distillation.

Reconstruction

Actuellement, les méthodes par reconstruction font partie des méthodes les plus populaires en détection d'anomalies. Le principe est d'apprendre une représentation des données normales à travers une tâche connexe de reconstruction. Les images issues de l'espace des données normales \mathcal{X}_a seront a fortiori bien reconstruites par le modèle, car ce dernier a été entraîné pour cela. Au contraire, les données anormales \mathcal{X}_b , verront leur reconstruction erronée dans les régions d'anomalies, car non présentes durant l'entraînement. De nombreuses variantes ont été expérimentées et sont détaillées dans la suite de cette sous-section.

Auto-Encodeurs : Les réseaux auto-encodeurs (AE ou "Auto-encoders") sont des modèles en U composés d'un encodeur, se chargeant de projeter l'image dans un espace de représentation de petite dimension, et d'un décodeur, en général symétrique à l'encodeur, qui effectue le processus inverse de reconstruction. On considère notre ensemble d'échantillons de départ \mathcal{X} , composé d'un sous-ensemble d'échantillons normaux $\mathcal{X}_n \subset \mathcal{X}$ et un autre d'échantillons anormaux $\mathcal{X}_a \subset \mathcal{X}$. Soit une fonction d'encodage $f_\phi : \mathcal{X} \mapsto \mathbb{R}^D$, avec $D \ll \dim(\mathcal{X})$, et

une fonction de décodage $g_\theta : \mathbb{R}^D \mapsto \mathcal{X}$. Pour un score de reconstruction ou de similarité inter-image connu $l : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, la tâche de reconstruction par auto-encodeur se traduit par l'équation suivante :

$$\arg \max_{\theta, \phi} \sum_{x \in \mathcal{X}_n} l(g_\theta(f_\phi(x)), x) \quad (2.2)$$

Contrairement aux modèles UNet vus précédemment, les AE n'admettent pas de connexions résiduelles. La raison est simple, on souhaite éviter que le réseau apprenne l'identité, et reproduise toute entrée à l'identique, sans se soucier de sa distribution d'origine. Entraînés à minimiser l'erreur absolue moyenne (MAE ou "Mean Absolute Error") ou l'erreur quadratique moyenne (MSE ou "Mean Squared Error"), les auto-encodeurs convolutifs classiques se sont avérés théoriquement robustes à la détection d'anomalies sur des bases de données simples comme MNIST [74]. Cependant, les reconstructions souvent floues qui résultent de ces modèles ne permettent pas une validation adéquate sur des bases de données médicales. En effet, pour des images contenant beaucoup de structures fines ou de textures subtiles, l'utilisation d'une métrique structurelle comme SSIM (ou "Structural SIMilarity") [75], à la place du score MAE ou même d'architecture plus complexes, peut s'avérer largement supérieur, comme démontré par [29] sur des données de nanofibres. Plus encore, l'utilisation de scores perceptuels basés sur des réseaux VGG pré-entraînés sur ImageNET à la place des scores de reconstruction habituels améliore tout autant les capacités de reconstruction que les performances de détection d'anomalies. Les modèles PIAD [76] et DPAE [30] témoignent de leur efficacité, que ce soit sur des images naturelles, des images histologiques ou des scans aux rayons X. L'ajout d'information connue a priori peut également s'avérer utile pour la tâche de reconstruction. Appliqués à l'imagerie IRM cérébrale [77] et aux OCT [78], les "denoising auto-encoders" (DAE) permettent d'augmenter la robustesse du modèle aux légères variations, et à produire des sorties plus fidèles et proches de l'espace de probabilité des images normales, en transformant l'objectif de reconstruction en un objectif de débruitage, avec l'ajout d'un bruit normal à l'entrée du réseau. Le format du vecteur de représentation $z \in \mathbb{R}^D$ influe aussi sur les capacités de reconstruction. En général, ce dernier est représenté sous la forme d'un vecteur à une dimension $z \in \mathbb{R}^D$. Toute notion de spatialité des images est alors perdue. L'utilisation d'AE convolutifs, permet d'obtenir des vecteurs de représentation sous forme matricielle $z \in \mathbb{R}^{C' \times H' \times W'}$, conservant ainsi l'information spatiale à travers les différentes couches du réseau. Une étude d'ablation menée par Baur et. al. [31] sur des images IRM cérébrales témoigne du gain en qualité de reconstruction pour des représentations spatiales plutôt qu'à une dimension unique.

Auto-Encodeurs Variationnels : Les modèles auto-encodeurs variationnels (VAE ou "Variational auto-encoder") visent à transformer la représentation déterministe des AE en une représentation stochastique, où les attributs du vecteur de représentation sont dorénavant définis par une distribution. Le décodeur est décrit comme un générateur probabiliste, dont l'objectif est d'estimer, pour un vecteur d'entrée x , la probabilité $p_\theta(x|z)$. De son côté, l'encodeur est chargé d'établir la probabilité $p_\theta(z|x) = \frac{p_\theta(x|z)p_\theta(z)}{p_\theta(x)}$. Cette dernière n'est malheureusement pas soluble, en raison de la nécessité de connaître la distribution d'entrée $p_\theta(x)$. On cherche donc à estimer cette dernière, en considérant une fonction paramétrique tierce q_ϕ , telle que $q_\phi(z|x) \sim p_\theta(z|x)$, symbolisé par l'encodeur. On cherche alors à minimiser la divergence de Kullback–Leibler $D_{KL}(q_\phi(z|x)||p_\theta(z|x))$. En développant la formule de la divergence KL et en la simplifiant, le nouvel objectif est défini par :

$$\arg \max_{\theta, \phi} \mathbb{E}_{z \sim q_\phi(\cdot|x)} [\ln p_\theta(x|z)] - D_{KL}(q_\phi(\cdot|x)||p_\theta(\cdot)) \quad (2.3)$$

Le premier terme, correspondant au terme de maximum de vraisemblance, qui peut être approximé par l'erreur de reconstruction $\|x - D_\theta(z)\|_p$. Le second correspond à la divergence de Kullback–Leibler, qui suppose que $p_\theta(z)$ est une loi normale centrée réduite $\mathcal{N}(0, I)$. Le changement du paradigme stochastique vers un paradigme probabiliste est en faveur de la détection d'anomalies. An et. al [79] appuie dans son étude que les VAE ont un pouvoir expressif plus grand que les AE, car ils prennent en compte la variabilité de l'espace latent. Par définition, les anomalies admettent des attributs moyens similaires, mais une plus grande variance, qui est dorénavant prise en compte lors de la phase d'échantillonnage $z \sim q_\phi(\cdot|x)$. Appliquées aux volumes IRM cérébrales [31], les approches VAE sont comparativement plus performantes que leurs homologues AE classiques. La régularisation de l'espace latent semble en être la principale raison. Néanmoins, le compromis capacité de reconstruction contre capacité de détection d'anomalies subsiste, et les générations VAE forment en général une représentation moyenne de la base de données avec peu de détails [31].

Réseaux antagonistes génératifs : Les réseaux antagonistes génératifs (GAN ou "Generative adversarial networks") sont des réseaux connus pour leur haute capacité de génération et leur fidélité à l'espace de probabilité d'origine. Ils se démarquent par l'entraînement d'un réseau tierce, appelé discriminateur D , en plus du générateur G . D'un côté, à partir d'un vecteur aléatoire gaussien $z \sim \mathcal{N}(0, I)$, le générateur tente de produire des images suffisamment réalistes pour tromper le discriminateur. De l'autre, le discriminateur cherche à distinguer les images générées des images réelles de la base d'entraînement. L'entraînement en parallèle de ces deux réseaux se traduit par un jeu min-max entre le générateur et le discriminateur,

exprimé par

$$\min_G \max_D \mathbb{E}_{x \sim p_{réel}(x)} [\log(D(x))] + \mathbb{E}_{z \sim p_{générée}(z)} [1 - \log(D(G(z)))] \quad (2.4)$$

où $D(x)$ représente la probabilité que l'échantillon x appartienne à la distribution des données réelles $p_{réel}$ et $D(G(z))$ la probabilité que la génération $G(z)$ appartienne à la distribution des données générées $p_{générée}$. Théoriquement, l'approche GAN permet donc de rapprocher la distribution des images générées de la distribution des images réelles. Ils sont par définition parfaitement adaptés à une application à la détection d'anomalies, en remplaçant la distribution des données réelles $p_{réel}$ par la distribution des données normales p_{normal} . Cette idée a été explorée pour de la détection d'anomalies d'images OCT avec AnoGAN [80]. Entraîné sur des patchs normaux et selon la structure classique des GAN le modèle a appris à générer des images appartenant à l'espace de probabilité des images normales. Est ensuite introduit une étape de recherche du vecteur latent z_x , qui, pour une image du jeu de validation x , engendre la génération $G(z_x)$ la plus proche de x . À partir de là, un score de reconstruction entre la génération et l'image d'étude permet d'établir le score d'anomalies et d'identifier les régions anormales. La recherche du vecteur latent idéal z_x se fait incrémentalement à l'inférence pour chaque image de la validation. C'est pourquoi les mêmes auteurs ont proposé une amélioration appelée f-AnoGAN [81], qui introduit un encodeur entraîné séparément à trouver le vecteur aléatoire engendrant la génération la plus proche de la donnée d'entrée. Plus rapide, cette approche est également nettement plus performante sur la détection d'anomalies dans les images OCT. Du fait de leur entraînement compétitif, les GAN nécessitent une quantité de données bien supérieure que d'autres architectures et font souvent l'objet de problèmes de stabilité. De nombreuses astuces existent cependant pour éviter ce dernier problème, comme notamment le passage aux Wasserstein GAN [82, 83], présents dans l'implémentation de f-AnoGAN. L'architecture GAN peut également être combinée à une architecture AE; et se comporter comme une forme de régularisation de l'entraînement pour améliorer la fidélité des reconstructions. AnoVAEGAN, présenté dans [84] pour une application sur les volumes IRM cérébrales, est une architecture VAE à laquelle les auteurs ont greffé un discriminateur. Ce simple ajout à l'entraînement ne modifie pas le nombre de paramètres final du modèle, puisqu'en général le discriminateur n'est pas considéré à l'inférence, et permet une amélioration des métriques de reconstruction. En effet, la comparaison des modèles combinés GAN-AE, aux modèles VAE et AE, faite par les mêmes auteurs [31] confirme une amélioration de la reconstruction au niveau des contours des volumes IRM cérébrales, mais aucune garantie n'est donnée sur la cohérence anatomique fine de la reconstruction. Une dernière architecture combinée GAN-AE que l'on peut citer est GANomaly [27]. Ici, l'entraînement du

discriminateur est modifié pour prendre en compte l'information perceptuelle de ses couches intermédiaires, comme expliqué en [85], afin de faciliter la convergence du générateur. En outre, pour résoudre le manque d'expressivité du score d'anomalies par score de reconstruction, un encodeur supplémentaire est entraîné à reproduire l'espace latent du générateur pour la même entrée. L'hypothèse avancée est que le vecteur caractéristique contient l'ensemble de l'information pertinente à la discrimination, et que le score d'anomalies à ce niveau de profondeur aura plus de sens.

En résumé, il existe de nombreuses variantes d'apprentissage de la reconstruction, toutes cherchant à faire correspondre la sortie à l'espace de distribution de l'entrée. Cependant, il semble se dégager un compromis. Le gain en reconstruction pure, implique en général une perte de capacité de détection d'anomalies. Pour une meilleure capacité de reconstruction, les modèles ont alors une tendance à tendre vers l'identité, et à reconstruire toute structure en entrée, anomalies comprises. Cette tendance est explicitée dans [86], où les modèles AE entraînés sur des radiographies thoraciques sont capables de reconstruire des images issues d'autres bases de données médicales ou autre. Il s'agit d'une forme de surapprentissage de la reconstruction, qui se généralise à toutes les entrées, peu importe leur distribution d'origine. Au contraire, une baisse des capacités de reconstruction peut engendrer de meilleures capacités de détection d'anomalies, souvent en raison d'une représentation plus abstraite de l'entrée. Néanmoins, la capacité de reconstruction joue un rôle important dans l'interprétation du comportement du réseau, et une très mauvaise reconstruction ne donne aucune information anatomique des raisons qui poussent à la prédiction normale ou anormale.

Création de pseudo-anomalies

Les cas anormaux étant par définition rares, il est très difficile d'obtenir une base de données homogène représentant toutes les anomalies connues. Dans le cas d'un apprentissage supervisé, un déséquilibre trop grand des données engendrera une chute des performances, en général causée par un surapprentissage sur la classe dominante. Une façon de se débarrasser de cette contrainte et de diversifier au maximum le jeu d'entraînement, est de considérer une approche faiblement supervisée, en établissant une fonction de création de pseudo-anomalies $\phi : \mathcal{X} \mapsto \tilde{\mathcal{X}}$, visant à augmenter le jeu de données de départ \mathcal{D} sans ou avec peu d'anomalies en un jeu de données augmenté $\tilde{\mathcal{D}}$ avec un nombre suffisant de données anormales.

La création de pseudo-anomalies peut se faire de manière entièrement non supervisée, en considérant un jeu d'augmentations à appliquer à nos échantillons. Habituellement, les augmentations sont utilisées pour multiplier artificiellement le nombre de données d'entrées, en appliquant des transformations de couleurs, de position ou de structure, et rendre plus

robuste les réseaux de neurones [87] à ces dernières. En détection d'anomalies, les régions anormales ne concernent qu'un sous-ensemble réduit de l'image d'entrée, et occupent rarement toute l'image. C'est pourquoi l'implémentation d'augmentations locales comme Cut-Paste [28], CutMix [88] ou CutOut [89] qui consistent respectivement à coller des patches d'une même image à une autre position, coller des patches d'images similaires dans une autre, et remplacer des patches par des valeurs constantes ; s'avère efficace. En particulier, les auteurs de CutPaste [28] traitent particulièrement du cas de la détection d'anomalies, et mettent en avant l'importance d'utiliser des pseudo-anomalies avec une notion de structure et de texture similaires à la distribution d'entrée. Leur méthode de pseudo-anomalies est appliquée à un réseau convolutif profond pour de la prédiction d'anomalie binaire, et améliore grandement les performances de détection d'anomalies. Le risque reste cependant que le modèle apprenne à reconnaître ce jeu de transformations et ne parvienne pas à identifier des anomalies plus subtiles. Les auteurs de AugPaste [90] proposent une approche plus spécifique à la base de données. Dans le cadre du diagnostic de la rétinopathie diabétique dans les images de fond d'œil, les auteurs proposent de sélectionner un sous-ensemble d'images anormales, d'en extraire les signes pathologiques, et de créer à partir d'images normales de nouvelles images anormales en collant et transformant aléatoirement le jeu d'anomalies extraites au préalable. De cette façon, on s'assure que le modèle base sa prédiction sur des transformations qui ont un sens clinique, mais cela requiert une certaine quantité d'images anormales de départ et une tâche d'extraction par annotation fastidieuse pas toujours réalisable. Enfin, les pseudo-anomalies peuvent également être considérées dans une tâche de reconstruction. Selon le même principe que les DAE, on souhaite rendre robuste la reconstruction à l'apparition d'anomalies et éviter de l'apprentissage de l'identité. Pour cela, Astrid et. al. [26] ont proposé d'entraîner un AE à la fois à reconstruire des images normales, mais aussi à corriger les pseudo-anomalies créées à partir de la même image. De cette manière, l'encodeur apprend une représentation commune plus robuste, et tente non seulement de reconstruire, mais de transformer l'entrée de sorte qu'elle corresponde à la distribution d'entrée.

La création de pseudo-anomalies est une tâche qui, de prime abord, paraît simple, mais dont les issues sont plus complexes. Le choix des augmentations influe grandement les résultats, et peut tout aussi bien les améliorer que de les diminuer.

Auto-supervision

Comme son nom l'indique, l'apprentissage auto-supervisé (SSL ou "self-supervised learning") est une branche du non supervisé, qui consiste à créer son propre signal de supervision à partir des données elles-mêmes. Les méthodes SSL visent à construire un espace de représentation

générique des données et transférable sur de nombreuses tâches en aval, dont les principales caractéristiques et enjeux sont explicités dans un même "livre de cuisine" [91]. Elles s'avèrent particulièrement utiles pour une application au domaine médical, où l'obtention d'annotations y est particulièrement difficile et où de nombreux inconnus demeurent. On s'intéressera ici aux applications SSL au domaine médical et plus précisément à ses variantes pour la détection d'anomalies, séparables en deux familles.

Apprentissage profond de métriques : L'apprentissage profond de métriques (DML ou "Deep Metric Learning") auto-supervisé est une famille de méthodes SSL dont l'objectif est d'encourager la proximité dans l'espace de représentation des données sémantiquement proches, et d'éloigner les autres. On y retrouve des méthodes comme SimCLR [92] ou MoCo [93], les deux fondées sur le principe d'une fonction objectif contrastive. On considère à l'entraînement un ensemble de paires positives et de paires négatives. L'approche contrastive consiste à maximiser la similarité des paires positives, et d'éloigner la représentation de celles des paires négative. Le défi repose dans la définition des paires positives et négatives, là où aucune annotation n'est disponible en SSL. Les paires positives sont en général deux transformations aléatoires ou deux "vues" issues d'une même image d'origine, et les paires négatives une combinaison avec les autres paires. Les représentations issues de l'entraînement d'un encodeur sont alors transférées sur d'autres tâches de prédiction plus concrètes. Cho et al. [94] ont ainsi entraîné un modèle MoCo v2 sur des radiographies thoraciques et validé leur représentation générique sur des tâches de classification de pathologies et une tâche de traduction image-à-image de suppression osseuse. Les résultats sont formels, l'utilisation d'encodeurs pré-entraînés par tâche SSL pour du transfert de tâche en aval est bien supérieure à un entraînement de rien ou à un pré-entraînement sur ImageNet comme d'ordinaire. Plus encore, les méthodes contrastives peuvent être repensées pour une application directe à la détection d'anomalies. Les auteurs de SPD [95] proposent de raffiner la représentation apprise par méthode contrastive pour la détection d'anomalies, en considérant pour les paires positives une transformation faible qui ne change quasiment pas la sémantique de l'image, et pour les paires négatives une transformation locale forte qui pourrait correspondre à une anomalie. Les auteurs montrent ainsi que les représentations apprises admettent une plus grande sensibilité aux perturbations locales, et leur transfert à des modèles de détection d'anomalies améliore leurs performances.

Auto-distillation : L'auto-distillation est une seconde famille de méthodes SSL qui se débarrasse de la contrainte des paires négatives. En effet, un désavantage des DLM est la nécessité d'avoir une grande quantité de paires négatives face aux paires positives, ce qui engendre des coûts computationnels et de mémoire non négligeables. Ici, on souhaite simplement rapprocher la représentation de deux vues d'une même image en entrée de deux encodeurs.

On retrouve des méthodes comme SimSiam [96] et BYOL [97]. Le défi de ces approches est d'éviter que la sortie soit constante pour toutes données en entrée. Plusieurs approches sont considérées, mais reposent fondamentalement sur l'asymétrie des deux encodeurs, le chemin emprunté par la descente de gradient à l'optimisation et la méthode de prédiction en sortie des encodeurs. Tout comme les approches DLM, les approches par auto-distillation sont en général utilisées comme pré-entraînement pour un transfert des tâches en aval. Toutefois, la simplicité architecturale de ces approches a rendu possible leur fusion à d'autres architectures. SiAD [98] est un exemple de fusion d'un AE pour une tâche de reconstruction avec une architecture SimSiam [96]. L'idée principale est d'apporter un niveau d'abstraction supplémentaire à l'espace de représentation, en rapprochant les paires positives l'une de l'autre et en leur imposant de contenir de l'information structurelle caractéristique. Li et. al. proposent ainsi SSL-AnoVAE [99], un modèle fusionnant une approche VAE, GAN et BYOL pour la prédiction de la rétinopathie diabétique dans les images du fond de l'œil. L'approche SSL apprise en parallèle de la tâche de reconstruction, vise à apprendre une information à priori de la base de données, spécifique à l'apprentissage de la normalité et complémentaire à la reconstruction.

Les méthodes SSL ont largement gagné en essor ces dernières années, grâce à leur capacité de généralisation remarquable. Couplées à des méthodes SSL, les méthodes de détection d'anomalies gagnent en sémantique de représentation et mènent à de meilleures capacités de généralisation. De plus, ces dernières apportent un certain contrôle sur l'entraînement.

2.3 Interprétabilité

L'évaluation des réseaux de neurones est fondée sur un jeu de données test, indépendant du jeu d'entraînement, permettant d'évaluer le comportement du modèle face à de nouvelles données jusque là jamais rencontrées, et seule la prédiction finale du réseau est en général évaluée. Or, le processus d'analyse de l'entrée qui a mené à ladite prédiction est tout aussi important, voire plus, surtout pour une application médicale ayant des enjeux sur la santé d'êtres humains : on souhaite non seulement savoir si un patient présente un risque de pathologie grave, mais surtout si ce risque se base sur des faits médicaux valables. Malgré des performances époustouflantes, les réseaux de neurones sont souvent qualifiés de "boîtes noires", en raison de leur complexité interne, et du manque de contrôle sur leur processus interne. Dans cette section, on s'intéresse aux méthodes d'analyse d'interprétabilité post-hoc [100], c'est-à-dire après entraînement du modèle.

Une première forme d'analyse du comportement des réseaux de neurones est la visualisation de l'espace de représentation. Le vecteur de représentation $z \in \mathbb{R}^D$ est un vecteur carac-

térisant les attributs pertinents de l'entrée, dans une dimension D très petite devant les dimensions de l'entrée. À l'aide d'outils de réduction de la dimensionnalité, le vecteur de représentation z peut-être transposé de son espace en D dimensions, à un espace interprétable par un être humain, en deux ou trois dimensions sous la forme d'un nuage de points. L'important est, pour la phase de réduction de dimension, de conserver au maximum les ressemblances et les dissimilarités, pour une représentation fidèle à la réalité. Il existe de nombreux outils de réduction de dimension satisfaisants ces critères. Tout d'abord l'analyse en composantes principales (PCA ou "Principal Component Analysis") [101], est une technique d'analyse statistique, qui vise à transformer un ensemble de variables corrélées, ici les attributs de notre vecteur de représentation, en ensemble de variables décorréelées, appelées axes ou composantes principaux. Les premières composantes de la transformation correspondent aux attributs dominants de l'espace de représentation, et peuvent directement être utilisées pour l'affichage en deux ou trois dimensions. Un autre outil communément utilisé est l'algorithme t-SNE (ou "t-distributed stochastic neighbor embedding") [102], qui se fonde sur une approche probabiliste, visant à transformer les distances inter-échantillons en probabilités dans l'espace de haute dimension, puis à minimiser la divergence de Kullback-Leibler dans la représentation en deux ou trois dimensions. De cette façon, l'algorithme vise à conserver les densités de probabilités entre les deux représentations. Enfin, une dernière alternative couramment utilisée similaire à t-SNE est l'algorithme UMAP (ou "Uniform Manifold Approximation and Projection for Dimension Reduction") [103]. Il diffère dans l'analyse haute dimension, en réalisant un graphe, dont les poids et les liaisons correspondent à la vraisemblance que deux points soient proches, déterminés en regardant les n plus proches voisins. De par sa théorie, UMAP présente l'avantage de mieux conserver les structures globales lors de la projection que t-SNE, et un temps de calcul significativement plus faible. Une application de l'algorithme de réduction de dimensionnalité t-SNE aux photographies du segment antérieur dans les approches par classification [59, 60] permettent de visuellement confirmer une séparation dans l'espace de représentation entre les échantillons présentant des signes de pathologies, souvent dans des regroupements distincts et avec peu de superposition. Cette analyse visuelle permet de valider la pertinence de la représentation pour la tâche de prédiction considérée, mais admet tout de même certaines limitations. Les sous-représentations t-SNE ou UMAP ne sont aucunement une garantie de la représentation de l'espace en haute dimension, elles doivent donc être analysées avec parcimonie et un certain recul.

La visualisation de l'espace de représentation est un bon indicateur de la séparation des attributs communs pertinents à l'entraînement, mais rien ne garantit que la discrimination se fonde sur des facteurs discriminatifs réels plutôt que de biais propres à la base de données d'entraînement. Pour obtenir une visualisation des régions supposément d'intérêts de

l'image d'entrée selon la prédiction d'un modèle, la méthode la plus populaire est l'algorithme Grad-CAM (ou "Gradient-weighted Class Activation Mapping") [104]. Soit un réseau convolutif entraîné sur une tâche de classification. L'objectif est d'établir, pour une image d'entrée donnée, une carte de chaleur indiquant la région d'intérêt dans l'image la plus discriminative pour la prédiction souhaitée en sortie. L'algorithme Grad-CAM propose une méthode pour y arriver, sans modifier la structure du réseau et sans entraînement supplémentaire. La méthode se fonde sur l'hypothèse que les couches les plus profondes d'un réseau convolutif contiennent de l'information caractéristique abstraite, liée directement à la tâche de prédiction. Observer l'activation des neurones de ces dernières couches, reviendrait à déterminer les régions discriminatives de l'image d'entrée pour la prédiction souhaitée. Pour déterminer l'importance relative à la classe de prédiction souhaitée des cartes caractéristiques profondes, les auteurs proposent de procéder à une rétropropagation de la sortie souhaitée, permettant d'associer un poids à chacun des neurones. Dès lors, en combinant l'ensemble de ces cartes avec leurs poids associés, on obtient une carte de chaleur indiquant la région la plus sensible pour la prédiction désirée. La majorité des travaux de classification présentés précédemment font état d'une analyse par Grad-CAM des régions sensibles à la prédiction des pathologies de la conjonctive [7, 7, 54, 56, 59, 60]. Les cartes de chaleurs permettent de donner un sens clinique aux prédictions, et s'assurer que le biomarqueur pathologique se retrouve effectivement dans la région d'intérêt. Surtout, elles permettent de mettre en avant de potentiels biais de la base de données qui peuvent induire en erreur le processus de diagnostic du réseau.

Récemment, une alternative élégante à l'approche Grad-CAM a été développée par Lang et al. [105], en se servant de la puissance générative des modèles StyleGAN [106]. L'idée est d'entraîner un modèle StyleGAN à construire un espace latent désentrelacé conditionné sur les prédictions du classificateur pré-entraîné à analyser. En raison de leur espace de style, les StyleGAN ont la propriété intéressante de produire un espace de représentation désentrelacé, c'est-à-dire dont chaque élément de représentation correspond à un unique attribut dans l'espace de représentation de l'image originale. En conditionnant l'espace de style à la prédiction du classificateur, le modèle StyleGAN apprend à reconnaître les attributs colorimétrique, structurel et géométrique de l'image qui induisent des changements dans la prédiction. Il est alors possible de faire varier un à un les attributs dominants, de générer les images correspondantes, et de révéler des attributs sémantiques complexes directement dans l'espace de l'image. Comme expliqué ci-dessus, les travaux de Babenko [?] ont permis de révéler la capacité des réseaux de neurones à prédire des facteurs systémiques à partir d'images du segment antérieur de l'œil. Les auteurs avaient initialement réalisé une analyse Grad-CAM, qui aboutit à la conclusion que la prédiction se fondait sur des régions anatomiques pertinentes, comme la sclère, la pupille, l'iris ou les paupières. Cependant, ces régions sont relativement

larges, ne fournissant pas d'information locale sur les biomarqueurs mêmes. Suite à cela, l'approche StyleGAN présentée a été appliquée au même classificateur et les attributs mis en avant ont été analysés par un ensemble de spécialistes [107]. Cette analyse a permis de confirmer la validité des attributs discriminatifs à la prédiction pour des cas simples, comme la prédiction de cataracte. Mais elle a aussi révélé certains biais de la base de données, notamment pour le diagnostic du taux d'hémoglobine, un des attributs dominants pour de hauts taux d'hémoglobine étant l'apparition d'eye-liner.

2.4 Métriques d'évaluation

Cette section présente une courte introduction aux métriques d'évaluation couramment utilisées en détection d'anomalies. Soit un problème de classification binaire, où les cas négatifs correspondent aux données dites normales et les cas positifs aux données dites anormales, et un classificateur entraîné à la détection d'anomalies. On notera pour la suite :

- **Vrais positifs (TP ou "True Positives")** : L'ensemble des données positives prédites positives,
- **Faux positifs (FP ou "False Positives")** : L'ensemble des données positives prédites négatives,
- **Vrais négatifs (TN ou "True Negatives")** : L'ensemble des données négatives prédites négatives,
- **Faux négatifs (FN ou "False Negatives")** : L'ensemble des données négatives prédites positives.

2.4.1 Précision

La précision évalue le ratio de données correctement prédites positives par rapport à toutes les prédictions positives du modèle.

$$Precision = \frac{TP}{(TP + FP)} \quad (2.5)$$

Ici, l'accent est mis sur l'évaluation de la capacité du modèle à détecter les cas effectivement pathologiques. Cette métrique est à privilégier dans les situations où l'on souhaite minimiser le taux de faux positifs, sans se soucier de l'éventuelle prédiction de faux négatifs.

2.4.2 Rappel ou sensibilité

Le rappel ou la sensibilité évalue le taux de données positives correctement prédites par le modèle.

$$\text{Rappel/Sensibilite} = \frac{TP}{(TP + FN)} \quad (2.6)$$

Ici, l'accent est mis sur l'évaluation de la capacité du modèle à détecter correctement tous les cas pathologiques. Cette métrique est à privilégier notamment pour les applications médicales, où l'enjeu principal est de réduire le taux des faux négatifs.

2.4.3 Spécificité

La spécificité évalue le taux de données négatives correctement prédites par le modèle

$$\text{Specificite} = \frac{TN}{(TN + FP)} \quad (2.7)$$

Ici, l'accent est mis sur l'évaluation de la capacité du modèle à diagnostiquer correctement tous les cas sains. Cette métrique est à privilégier dans les situations où la prédiction des cas sains est plus importante que celles des cas pathologiques, sans se soucier des éventuels faux négatifs.

2.4.4 Aire sous la courbe ROC

L'ensemble des métriques décrites précédemment se fonde sur une classification binaire à l'évaluation. Or, la majorité des applications de détection d'anomalies propose en réalité un score, duquel un seuillage permet d'obtenir une classification binaire. Dépendamment de la valeur du seuil choisi, il est possible de plus ou moins favoriser une métrique plutôt qu'une autre, selon le cas d'utilisation souhaité.

Pour se débarrasser de la contrainte du choix d'un seuil adéquat, on établit la fonction d'efficacité du récepteur (ROC ou "Receiver Operating Characteristic"). Pour les différentes valeurs de seuil possible, on trace en abscisse le taux de faux positifs (égal à $1 - \text{Specificite}$), et en ordonnée le taux des vrais positifs (égal à la *Sensibilite*). Enfin, il suffit de calculer l'aire sous cette courbe, notée AUC (ou "Area under the curve") pour obtenir une métrique globale sans la contrainte de seuillage.

La ROC AUC est une métrique particulièrement appréciée pour des applications au domaine

médical, car elle donne autant de poids aux classes positives qu'aux classes négatives, et permet d'établir un seuil en fonction des particularités de l'application. Cependant, il est souvent déconseillé de l'utiliser sur un ensemble de données déséquilibré, c'est-à-dire dans notre cas avec une classe dominante en général les données normales, car le taux de faux positifs sera tiré vers le bas en raison d'un grand nombre de vrais négatifs comme le suggère l'équation 2.7.

2.4.5 Aire sous la courbe PR

Similaire à la courbe ROC, la courbe précision-rappel (PR ou "Precision-Recall") trace aux différentes valeurs de seuils possibles la valeur du rappel en abscisse, et la valeur de la précision en ordonnée. L'aire sous la courbe de la courbe PR est souvent confondue avec la précision moyenne (AP ou "Average Precision"), et est un indicateur global des performances du modèle sans seuillage. Contrairement à la ROC AUC, la PR AUC se focalise sur la classe positive et se soucie moins des classes négatives. Elle est donc plus adaptée pour les bases de données avec un débalancement de classe. Elle est néanmoins moins appréciée dans un contexte médical de détection d'anomalies, où les statistiques de prédiction des classes négatives normales est tout aussi voire plus importante que celle des classes positives anormales.

CHAPITRE 3 HYPOTHÈSES ET OBJECTIFS DE RECHERCHE

De la revue de la littérature, nous constatons les faits suivants :

1. Les photographies du segment antérieur admettent un pouvoir diagnostique certain, autant pour des troubles oculaires graves que certains troubles systémiques.
2. La majorité des bases de données du segment antérieur sont privées, et sont en général spécialisées sur une ou quelques pathologies.
3. La conception manuelle d'attributs caractéristiques discriminants pour le segment antérieur paraît difficilement réalisable, aux vues de la diversité des pathologies et de l'importance clinique de la robustesse à l'inconnu.
4. L'utilisation d'approches entièrement supervisées s'avère dans la majorité des cas cliniquement insuffisante, et est limitée par la disponibilité des annotations.
5. Les approches non supervisées semblent nécessiter une conception spécifique aux données étudiées, aucune d'entre elles ne semblant faire véritablement consensus dans la littérature.

De fait, il n'existe à notre connaissance aucuns travaux sur la détection d'anomalies non supervisée appliquée aux photographies du segment antérieur de l'œil, l'entièreté de la communauté scientifique s'étant concentrée sur des approches entièrement supervisées. Explorer toutes les méthodes existantes est temporellement inconcevable. Afin de réduire notre champ de recherche, et uniquement conserver un sous-ensemble pertinent de modèles de base de détection d'anomalies, nous établissons ci-dessous un corpus d'hypothèses, qui nous guidera dans le choix des méthodes à explorer.

1. L'apprentissage profond garantit une plus grande expressivité des caractéristiques discriminantes spécifiques à la modalité d'entrée, en particulier en vue de la diversité des pathologies existantes pour le segment antérieur.
2. Les approches non supervisées garantissent la robustesse de la détection d'anomalies face à l'inconnu, là où les approches supervisées voient leur pouvoir prédictif limité aux annotations disponibles.
3. Le score de reconstruction fournit une interprétation clinique valable de l'anormalité, aussi bien au niveau de l'image qu'au niveau des pixels.

De plus, dans le cadre de notre partenariat avec LightX Innovations, il nous est demandé de respecter certaines contraintes.

1. Le modèle livré doit pouvoir fournir un score d'anomalie aussi bien au niveau de l'image, qu'au niveau des pixels.
2. La taille du modèle doit être limitée au minimum, dans le cadre d'un éventuel déploiement sur des appareils portatifs.
3. L'inférence doit pouvoir se faire en un temps raisonnable, dans le cadre d'un éventuel diagnostic sur le vif.

Basés sur ces hypothèses et ces contraintes, nous nous limiterons dans cette étude à l'apprentissage de la normalité par reconstruction (section 2.2.3). Spécifiquement, on s'intéresse aux méthodes construites autour d'une architecture auto-encodeur, pour leur rapidité d'exécution, leur modularité et leur capacité à fournir une erreur autant au niveau de l'image qu'au niveau pixels.

En partenariat avec l'entreprise LightX Innovations, nous avons défini les trois objectifs de recherche suivants :

O1 : Établir une base de données adéquate à l'entraînement de méthodes de détection d'anomalies non supervisées.

O2 : Réaliser une étude comparative d'un ensemble pertinent de méthodes de la littérature sur la tâche de détection d'anomalies et de reconstruction sur les photographies du segment antérieur de l'œil.

O3 : Concevoir et évaluer une nouvelle approche spécifique à notre cas d'étude en se basant sur les résultats de O2.

La méthodologie adoptée pour les objectifs **O1** et **O3** sera présentée dans le **chapitre 4**, comprenant le procédé d'acquisition des données et de leur prétraitement (**O1**) et la conception d'un nouveau modèle spécifique à notre cas d'étude (**O3**). Ensuite, le protocole d'évaluation ainsi que la sélection des modèles de la littérature répondant aux hypothèses de recherche mentionnées ci-dessus dans le cadre des objectifs **O2** et **O3** seront présentés au **chapitre 5**. Enfin, les résultats de l'évaluation seront analysés au travers du **chapitre 6**, dont la discussion mettra en lumière les avantages, les inconvénients et les limites des différentes approches considérées.

CHAPITRE 4 MÉTHODOLOGIE

Dans ce chapitre, nous détaillons l'ensemble des choix méthodologiques faits en vue de l'accomplissement des objectifs de recherche établis au chapitre précédent. Y sont décrits la création et le prétraitement de la base de données du partenaire pour une application à la tâche de détection d'anomalies aux photographies du segment antérieur de l'œil (**O1**), ainsi que la description d'une nouvelle approche spécifique à la tâche et aux données, résultant de l'analyse des méthodes existantes de la littérature (**O3**).

4.1 Création de la base de données

Cette section décrit la création d'une base de données pour la détection d'anomalies dans des photographies du segment antérieur par approche non supervisée. Ce travail a été fait collaborativement avec LightX Innovations, au moyen de leur plateforme d'acquisition. Le projet de recherche fait l'objet d'une approbation éthique délivrée par le comité d'éthique de la recherche de Polytechnique Montréal CER-22223-75-D (Annexe A), pour l'utilisation de données existantes et la collecte de nouvelles données.

4.1.1 Données existantes

Notre partenaire, LightX Innovations, dispose d'une plateforme d'acquisition de photographies du segment antérieur. Cette plateforme utilise deux appareils portatifs, dans notre cas des iPhone, montés simultanément sur la lentille d'un biomicroscope à lampe à fente à l'aide d'un socle imprimé en 3D (Figure 4.1). Cette configuration permet de capturer une vue stéréoscopique de chaque œil, ouvrant ainsi la possibilité d'une reconstruction 3D. Cependant, dans le cadre de ce projet, nous nous limiterons à l'analyse 2D, la stéréoscopie étant uniquement utilisée à des fins d'augmentation du nombre d'échantillons de la base de données.

Pour rappel de la section 2.1.2, la base de données initiale, mise à disposition par le partenaire, est constituée de 10 779 images RGB d'une résolution avoisinant 2000x4000 pixels, capturées sur les deux yeux de 255 patients, et réparties selon 8 tests ophtalmologiques (Figure 2.3). Cette base est scindée en deux sous-ensembles, issues de deux cliniques distinctes : *l'École d'optométrie de Montréal* (EOM) comprenant 778 photographies et *l'Institut de l'œil des Laurentides* (IOL) comprenant 10 001 photographies. Pour ce projet, nous nous limiterons aux photographies issues de l'observation de la conjonctive, de la sclère, des cils et paupières.



FIGURE 4.1 Système de capture d'images du segment antérieur - LightX Innovations

Ce choix est fait afin de réduire l'étendue du projet et d'assurer une certaine homogénéité des échantillons, tout en assurant une couverture suffisamment large du segment antérieur pour le diagnostic d'une majorité des pathologies du segment antérieur. La particularité de cette modalité réside dans le positionnement de l'œil. Pour chaque patient et chaque œil, il leur a été demandé de regarder dans cinq directions différentes — tout droit, en haut, en bas, à gauche et à droite - comme illustré en figure 4.2. Dans la suite, l'entraînement de nos modèles se basera exclusivement sur ce type d'image.

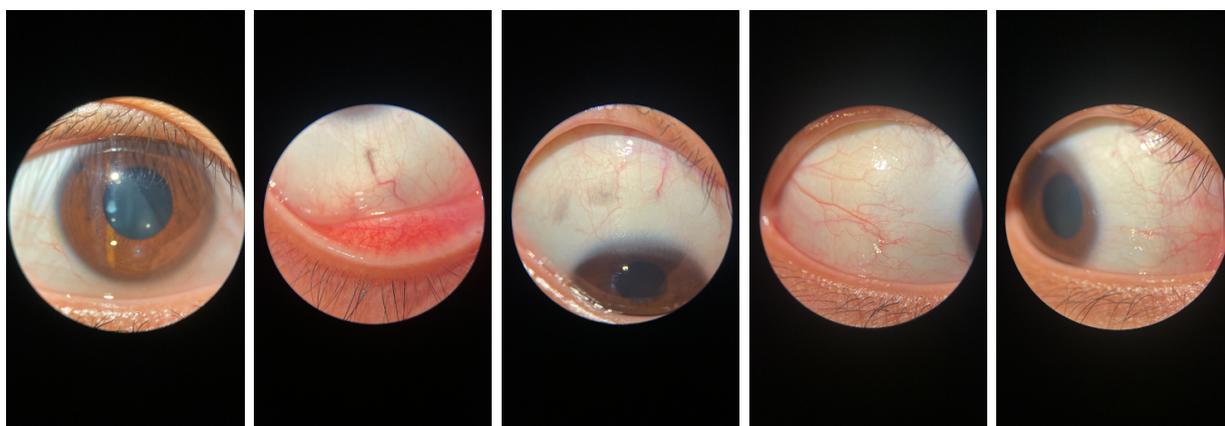


FIGURE 4.2 Photographies de la conjonctive selon le positionnement de l'œil

Les modèles de détection d'anomalies par reconstruction reposent sur l'apprentissage de la notion de normalité à partir des données d'entraînement et sont conçus pour identifier comme anomalies tout échantillon qui dévierait de la distribution apprise. Par conséquent, l'entraîne-

ment nécessite exclusivement des données dites "normales", en quantité suffisante pour assurer la généralisabilité du modèle. Au contraire, les données dites "anormales" n'interviennent qu'à l'étape de validation du modèle. Cependant, la base de données fournie par notre partenaire ne comporte aucune annotation spécifique pour distinguer les données normales des données anormales. En réalité, toutes les photographies disponibles proviennent de cliniques, ce qui implique un fort risque de présence de pathologies oculaires. Ces pathologies peuvent être visibles sur une ou plusieurs vues, selon leur nature et la région anatomique concernée. De plus, l'absence d'annotations cliniques précises empêche une validation rigoureuse des algorithmes développés. Pour pallier ce problème, nous avons adopté une approche en deux étapes :

1. Réaliser une collecte de données auprès de la population de Polytechnique Montréal, pour l'obtention de photographies du segment antérieur de l'œil d'individus sains comme base de données d'échantillons "normaux".
2. Annoter les données des bases de données IOL et EOM au niveau image ("normal" ou "anormal") et au niveau pixel (masques de segmentation sémantique) pour assurer une analyse et une évaluation précise des modèles proposés.

4.1.2 Collecte de données

La collecte de données a été réalisée dans les locaux de Polytechnique Montréal, 2500 chemin de Polytechnique, Montréal, sur 3 jours non consécutifs, les 31 janvier 2024, 2 février 2024 et 22 février 2024.

La campagne a été réalisée dans le but simple et précis d'acquérir des photographies du segment antérieur sur des individus sains, afin de constituer notre base de normalité. Est éligible à la collecte tout individu ne présentant aucun trouble oculaire connu, à l'exception de simples troubles de la vision. Plus précisément, cela inclut les sujets exempts d'atteintes ou d'antécédents de pathologies, telles que le glaucome, la cataracte, la rétinopathie diabétique ou la dégénérescence maculaire. C'est pourquoi le recrutement a majoritairement ciblé la communauté étudiante, majeure, moins sujette à ce type de troubles. Le nombre de participants attendu était de 100, un nombre que nous avons jugé pertinent pour l'entraînement de nos modèles tout en ayant une variabilité démographique suffisante pour l'élimination de biais.

La collecte de données prend la forme d'un examen ophtalmique de routine, d'une durée de 10 minutes environ et réalisé au moyen d'un appareil à lampe à fente combiné au système d'acquisition, fourni par notre partenaire industriel, et entièrement manipulé par un ophtalmologiste. Pour des questions de praticité et d'éthique, aucune goutte mydriatique pour la dilatation de la pupille ou de fluorescéine n'a été administrée. Pour chaque participant, 7

tests ophtalmologiques ont été réalisés pour chaque œil, comprenant :

- Test de la conjonctive (observation de la sclère, de la conjonctive, des paupières et des cils),
- Test de la cornée,
- Test de la chambre antérieure,
- Test de Van Herick,
- Test du cristallin,
- Test de la marge pupillaire,
- Test de transillumination de l'iris.

Pour rappel, seules les images du test de la conjonctive sont étudiées dans nos travaux, les autres étant mises de côté pour une éventuelle suite au projet. Cela correspond à un minimum de 20 photographies par participant (2 yeux et 5 positions de l'œil en vue stéréoscopique), soit un total minimum de 2000 images. En parallèle, pour chaque participant, nous récupérons un ensemble d'information démographique, dont l'âge, le sexe, l'ethnicité et la correction optique si connue, à des fins purement statistiques et d'étude de biais. Les données ont été récoltées après consentement signé des participants, au moyen du formulaire d'information et de consentement disponibles en annexe B.

Dans la suite du mémoire, nous appellerons cette nouvelle base de données "POLY", en référence à sa provenance, tout comme EOM et IOL. Pour faire référence à la totalité des données, c'est-à-dire la combinaison des trois bases données POLY, EOM et IOL, nous utiliserons la formulation "base de données LightX".

4.1.3 Annotations

La plateforme d'acquisition du partenaire industriel construit automatiquement la base de données avec un ensemble d'annotations pour chaque image, dont voici la liste :

- *mediaID* : identifiant unique propre à l'image.
- *examID* : identifiant unique d'examen.
- *patientID* : identifiant unique de patient.
- *eye* : œil droit ou œil gauche.
- *testType* : test ophtalmologique réalisé (conjonctive, Van Herick, ...).
- *isPrimary* : vue stéréoscopique primaire.
- *isSecondary* : vue stéréoscopique secondaire.

Comme constaté, la liste initiale ne présente aucune annotation clinique. Le minimum nécessaire à l'élaboration d'un modèle de détection d'anomalies serait d'avoir, pour chaque échantillon, une information binaire de normalité, valant 0 pour les photographies exemptes

de tout biomarqueur pathologique, et valant 1 lorsqu’au moins un biomarqueur est présent. Plus encore, une information au niveau pixel de la localisation de l’anomalie serait d’autant plus bénéfique dans le cas d’une étude de l’interprétabilité de nos algorithmes. En partenariat avec notre partenaire industriel, nous avons donc réalisé une étape supplémentaire d’annotations des données, auprès d’un ophtalmologiste.

L’annotation a été réalisée sur 3 jours non consécutifs, les 12 mars, 22 mars et 19 avril 2024, dans les locaux de Polytechnique Montréal. Cette annotation concerne uniquement les photographies issues des bases IOL et EOM, les images de POLY étant par définition considérées comme "normales". Au moyen de la plateforme d’annotation personnalisable et libre de droits *Label Studio* [108], il nous a été possible de récolter des annotations au niveau pixel par masque de segmentation et des annotations niveau image par simple case à cocher. Un exemple de visualisation de la plateforme est donné en figure 4.3. Un total de 36 labels a été défini par l’ophtalmologiste, comprenant des biomarqueurs pathologiques (cataracte, ptérygion, hémorragies ...), des artéfacts (mascara, doigt ...) et des biomarqueurs non cliniquement significatifs, caractérisé par le sigle NCS. Ces derniers représentent une classe de biomarqueurs suspects, qui ne présentent aucun lien de cause-conséquence à des troubles oculaires ou systémiques connus à ce jour.

En parallèle, nous avons réalisé une annotation semi-automatique de la position de l’iris (au centre, à gauche, à droite, en haut, en bas), au moyen d’un modèle ResNet18 [45] entraîné de manière supervisée sur un sous-échantillon des bases de données POLY, EOM et IOL annotés à la main. Elle sera utile uniquement à des fins d’analyse de l’espace de représentation.

4.1.4 Prétraitement

Le prétraitement est une étape importante d’harmonisation et de sélection des données pour l’entraînement des modèles d’apprentissage profond. Pendant l’analyse de nos premiers modèles, nous avons remarqué une forte variation dans le format d’acquisition des données issues des trois bases de données IOL, EOM et POLY. Le masque circulaire observable dans les photographies est un masquage naturel, causé par le support 3D et la lentille du biomicroscope sur lequel il est monté. Il s’avère que la position du masque varie entre les bases de données, en raison des différents biomicroscopes et supports 3D utilisés, mais aussi au sein même des bases de données, dépendamment de la vue stéréoscopique observée et des possibles mouvements du support lors de l’acquisition. De plus, il ne s’agit pas d’un noir absolu, dans de nombreuses situations, on retrouve des fuites de lumière, rendant impossible un traitement par simple seuillage. Nous proposons ici un pipeline de prétraitement personnalisé, dont l’objectif est l’harmonisation du masque entre toutes les données EOM, IOL et POLY.

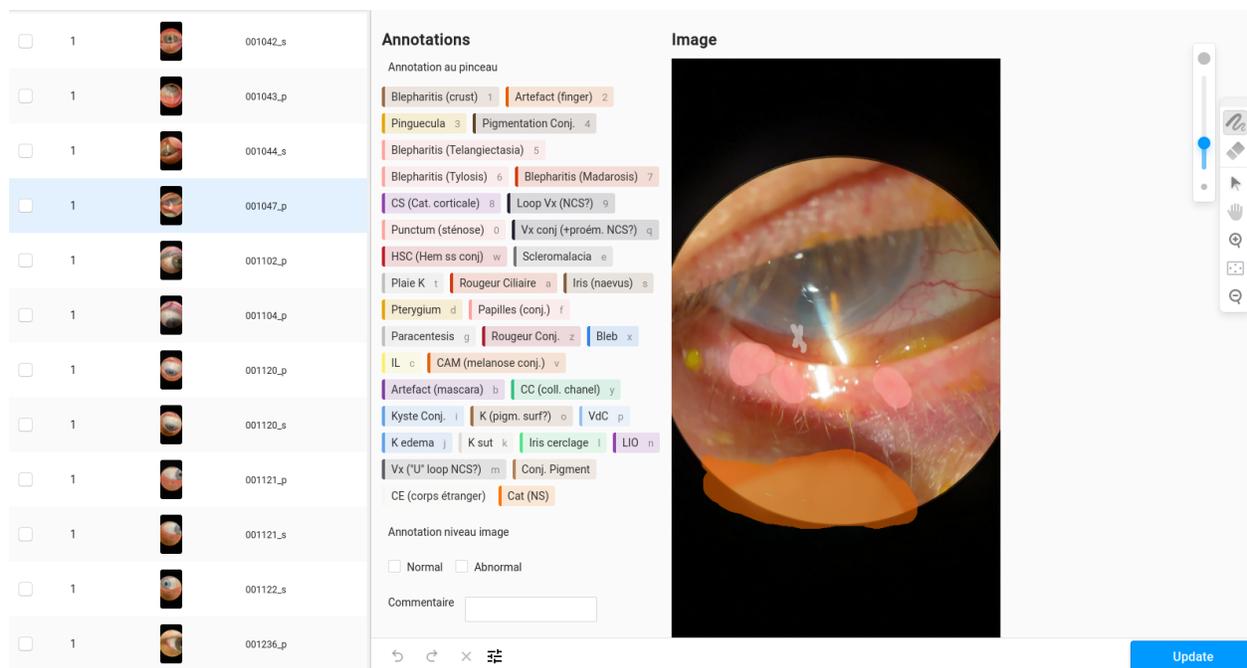


FIGURE 4.3 Exemple d’annotations depuis la plateforme d’annotation personnalisée

Plateforme *Label Studio* [108] - **Section gauche** : Sélection des échantillons de la base de données
- **Section droite** : Choix des annotations personnalisables niveau pixel et niveau image, et outils de peinture à même l’échantillon sélectionné

Ce pipeline sera entièrement réalisé en python, en majeure partie grâce à la librairie *OpenCV* [109], spécialisée dans le traitement de l’image. L’implémentation du pipeline est détaillée dans l’annexe C et consiste en la succession d’algorithmes de détection de contours (C-1), de détection de cercles (C-2), de sélection du meilleur candidat (C-3) et de rognage (C-4). Le prétraitement est appliqué sur toutes les photographies du segment antérieur de l’œil des bases de données IOL, EOM et POLY, et sauvegardé dans un espace de stockage sécurisé, prêt à l’utilisation pour l’entraînement de nos modèles. Une visualisation des étapes du prétraitement sur un exemple de photographies du segment antérieur de l’œil est donnée en figure 4.4.

4.2 Notre approche : SiamAAE

De l’analyse des modèles de la littérature sur la tâche de détection d’anomalies dans les photographies du segment antérieur de l’œil, dont les résultats seront présentés au **chapitre 6**, nous avons constaté une tendance des modèles à privilégier la tâche de reconstruction et tendre vers une fonction identité, peu importe la distribution de l’entrée. Nous proposons

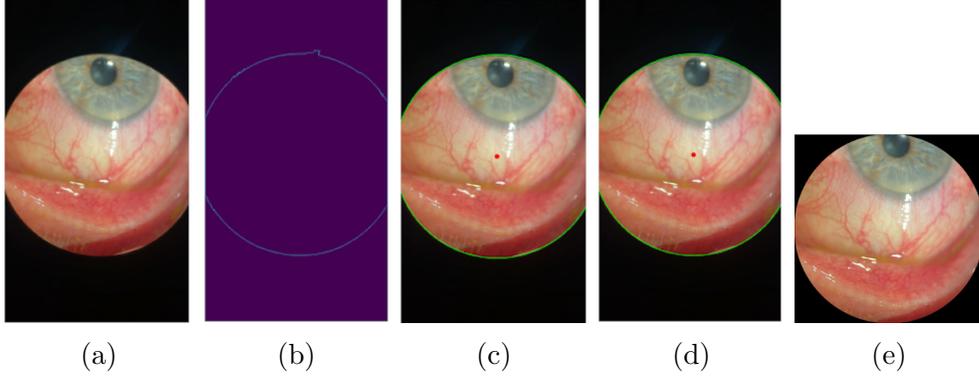


FIGURE 4.4 Exemple des étapes du prétraitement sur une photographie issue de IOL

(a) Image d'origine - (b) Tracé grossier du contour du masque circulaire après seuillage - (c) Ensemble des cercles plausibles détectés par transformée de Hough circulaire, ici un unique candidat. En vert, le contour du cercle ; en rouge son centre - (d) Sélection du meilleur candidat - (e) Rognage circulaire final

une nouvelle méthode spécialement conçue pour répondre à ce problème, que l'on nomme SiamAAE (ou "Siamese Adversarial AutoEncoder"). Elle est inspirée de SSLAnoVAE [99], et vise à grandement simplifier l'approche, tout en gardant ses avantages. L'architecture complète à l'entraînement est illustrée en figure 4.6, et se base sur la même architecture unifiée présentée à la section précédente. Elle est composée de trois modules principaux : le module de guidage SSL basé sur une approche par auto-supervision SimSiam [96] combiné à la création de pseudo-anomalies [28, 88, 89] ; le module de reconstruction basé sur les travaux [26, 77] et les erreurs de reconstruction de DPAE [30] et AE-SSIM [29] ; le module antagoniste basé sur les travaux de GANomaly [27]. La fonction objectif de notre modèle SiamAAE est définie par

$$L_{SiamAAE} = \lambda_{SSL}L_{SSL} + \lambda_{rec_total}L_{rec_total} + \lambda_{adv}L_{adv} \quad (4.1)$$

où l'intuition et l'implémentation derrière chacun de ces termes sont données dans la suite de cette section.

4.2.1 Module de guidage SSL

SimSiam [96] est une méthode d'apprentissage auto-supervisée de la famille des approches à auto-distillation, c'est-à-dire qui ne considère que des paires positives à l'entraînement. Pour deux vues différentes d'une même image de référence, l'architecture SimSiam vise à les

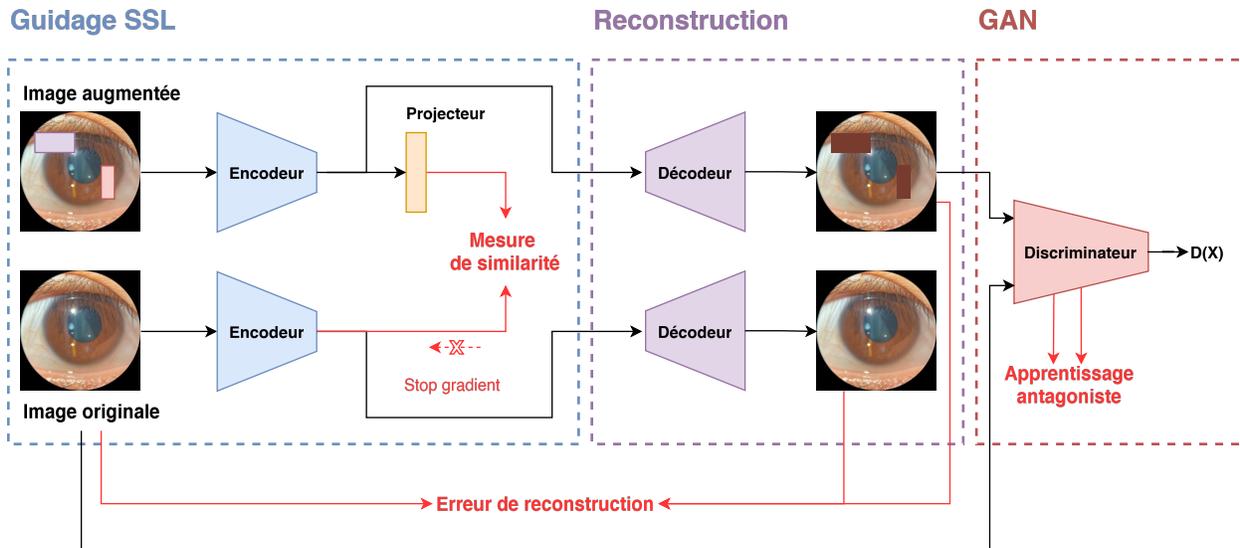


FIGURE 4.6 Schéma de l'architecture complète de SiamAAE durant la phase d'entraînement

Flèches noires : circulation de l'information dans le réseau - Flèches rouges : calcul des fonctions objectifs pour optimisation à l'entraînement

rapprocher dans l'espace de représentation, afin de conserver l'information sémantique les liant. Au travers du module de guidage SSL, nous souhaitons retranscrire ce principe à la tâche de détection d'anomalies.

Par hypothèse, apprendre à reconstruire, c'est apprendre les attributs structuraux et sémantiques essentiels de la base de données d'entraînement. En particulier, pour la détection d'anomalies, il s'agit d'apprendre les attributs spécifiques à la normalité, en restreignant les données d'entraînement aux données normales, de sorte que toute anomalie en entrée ne soit pas présente à la reconstruction, ce qui est en réalité très peu souvent réalisé, en raison de la nature de la tâche de reconstruction qui pousse les modèles à apprendre une forme de fonction identité. Pour pallier ce problème, on souhaite réaliser un modèle capable d'encoder toute image d'entrée anormale, vers une représentation de la normalité la plus proche. C'est-à-dire, pour une image donnée x , et une image pseudo-anormale construite à partir de cette dernière $x_a = \phi(x)$, on souhaite que leur représentation dans l'espace latent soit égale, soit que $E(x) = E(x_a)$ où E représente l'encodeur dans la structure de l'auto-encodeur. Pour ce faire, nous nous appuyons sur l'approche SimSiam [96], conçu précisément pour rapprocher dans l'espace de représentation deux vues d'une même image. Ici, les deux vues considérées sont donc l'image de référence x et l'image pseudo-anormales $x_a = \phi(x)$. Une visualisation de l'architecture et du concept de rapprochement dans l'espace de représentation sont illustrés

en figure 4.7.

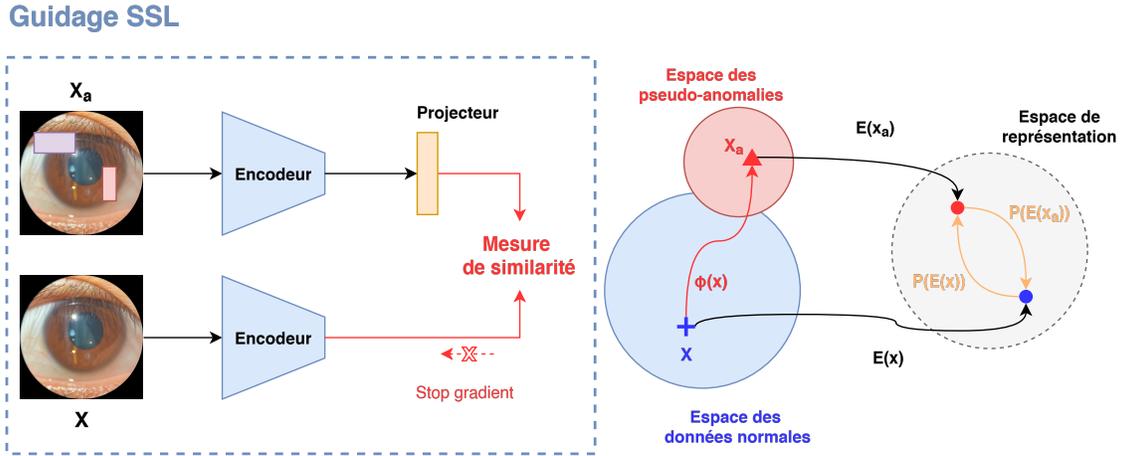


FIGURE 4.7 Schéma du guidage SSL et visualisation de la fonction de perte L_{SSL} en termes d'espace de représentation

La fonction objectif de ce module permettant le rapprochement des vecteurs de représentation z et z_a peut alors être définie par

$$L_{SSL} = \frac{1}{2}(d_{cos}(P(E(x)), E(\phi(x))) + d_{cos}(P(E(\phi(x))), E(x))) \quad (4.2)$$

où $d_{cos}(x, y) = 1 - \frac{x \cdot y}{\|x\| \cdot \|y\|}$ correspond à la distance cosinus et P correspond à une tête de projection, nécessaire à l'asymétrie du modèle pour empêcher un effondrement du modèle à l'entraînement [96].

Concernant la fonction de création de pseudo-anomalies ϕ , nous prenons inspiration des méthodes de la littérature [26, 28, 88, 89]. Pour chaque échantillon, nous sélectionnons aléatoirement entre 1 et N sous-régions rectangulaires de dimensions également aléatoires (Figure 4.8). Pour chacune d'entre elles, un ensemble de transformations est appliqué, choisi de manière aléatoire parmi les suivantes :

- Modification de la teinte, saturation et du contraste.
- Retourneement horizontal ou vertical.
- Mise à zéro de tous les pixels.
- Floutage gaussien.
- Mise à zéro d'un ou plusieurs canaux séparément (équivalent à un Dropout).

Durant la phase d'entraînement, le nombre maximal de sous-régions transformables est augmenté de 1 toutes les deux epochs jusqu'à atteindre la valeur maximale $N = 15$ définie empiriquement. Cela permet de faciliter la convergence du modèle en commençant par un

"échauffement" sur des cas simples.

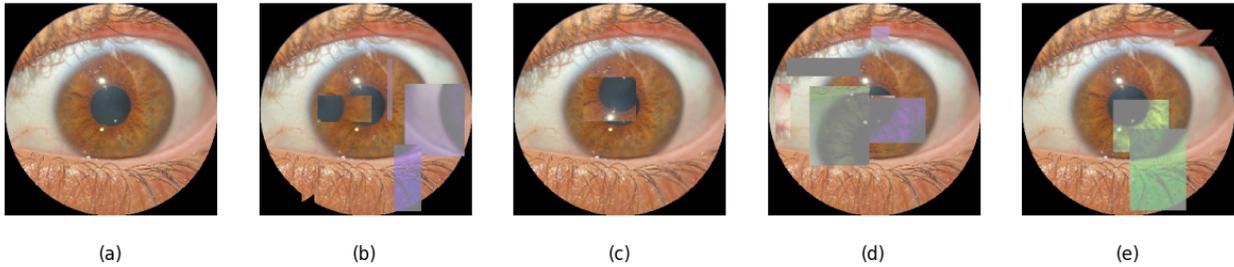


FIGURE 4.8 Exemples de pseudo-anomalies SiamAAE

(a) : Photographie de référence - (b) à (e) : Pseudo-anomalies aléatoires appliquées sur la photographie de référence selon le protocole de SiamAAE

4.2.2 Module de reconstruction

Couplé au module de guidage SSL, le module de reconstruction permet l'apprentissage d'attributs sémantiquement pertinents, tout en offrant la possibilité d'établir un score d'anomalies au niveau de l'image et au niveau des pixels. L'analyse des modèles de la littérature nous a permis de mettre en évidence les fortes capacités de reconstruction des modèles DPAAE [30] et AE SSIM [29] sur les photographies du segment antérieur de l'œil. Le premier modèle se sert de la puissance perceptuelle des réseaux de neurones pré-entraînés pour établir l'erreur de reconstruction, en réalisant la différence au niveau des couches moyennes profondes du dit réseau. Le second utilise la métrique structurelle SSIM [75] à la place de l'erreur moyenne absolue ou quadratique généralement utilisées, en raison de ses meilleures capacités à définir la différence perceptuelle entre deux images, en particulier dans les images contenant beaucoup de structures fines ou de textures. La fonction de perte de reconstruction L_{rec} utilisée dans notre approche peut alors être définie comme une combinaison linéaire de ces deux erreurs, soit

$$L_{rec} = \alpha L_{SSIM} + \beta L_{DPAAE} \quad (4.3)$$

où α et β sont des coefficients, fixés empiriquement à $\alpha = \beta = 1$. Les fonctions objectifs L_{SSIM} ainsi que L_{DPAAE} seront présentées plus en détail dans la section 5.1. Comme expliqué à la section précédente, le module de guidage SSL vise à rapprocher au maximum la représentation de l'image de référence de celle de l'image pseudo-anormale. Toutefois, l'égalité n'est pas garantie et une marge d'erreur existe. Pour prendre en compte cette erreur, tout comme

dans les travaux de [26], la reconstruction est appliquée aux deux représentations z et z_a , et la fonction de perte finale inclut l’erreur des deux reconstructions par rapport à l’image de référence, soit

$$L_{rec_total} = L_{rec}(x, D(E(x))) + L_{rec}(x, D(E(x_a))) \quad (4.4)$$

4.2.3 Module antagoniste

La reconstruction, ou plutôt la correction de l’image avec pseudo-anomalie, est une tâche qui reste tout de même très sensible. La fonction de création de pseudo-anomalies est plus ou moins destructive, dépendamment du nombre de sous-régions sélectionnées, de leur taille et des transformations appliquées. La perte d’information qui en résulte rend plus ou moins difficile, voir quasiment impossible leur correction totale. Néanmoins, l’image générée doit tout de même faire du sens et appartenir à l’espace de distribution des données réelles. C’est pourquoi, par-dessus les deux modules présentés ci-dessus, est greffé un discriminateur, dont l’objectif est de ramener la représentation de la reconstruction pseudo-anormale vers l’espace de distribution des données normales. Nous reprenons pour cela la fonction objectif antagoniste présentée dans GANomaly [27], elle-même s’appuyant sur les travaux de Salimans et al. [85]. Pour une couche intermédiaire du discriminateur f_D , la fonction de perte antagoniste pour le modèle SiamAAE s’écrit alors :

$$L_{adv} = \mathbb{E}_{x \sim p(x)} \|f_D(x) - f_D(\hat{x})\|_1 \quad (4.5)$$

4.2.4 Inférence

Après entraînement, le projecteur ainsi que le discriminateur sont mis de côté, et seule l’architecture auto-encodeur est conservée. La fonction de score d’anomalies considérée sera empruntée des travaux de DPAAE, soit la fonction d’erreur perceptuelle relative.

$$\mathcal{A}_{SiamAAE}(x, \hat{x}) = \mathcal{A}_{DPAAE}(x, \hat{x}) \quad (4.6)$$

CHAPITRE 5 PROTOCOLE D'ENTRAÎNEMENT ET D'ÉVALUATION

Dans ce chapitre est présenté le protocole d'entraînement, d'évaluation et de comparaison de notre modèle à un ensemble de modèles de la littérature, dans le cadre de l'objectif **O2**. Étant donné l'absence de travaux comparables à notre méthode sur les données du segment antérieur de l'œil, nous proposons une première étape de sélection d'un ensemble pertinent de modèles de la littérature basés sur la tâche de reconstruction. Dès lors, nous pourrons établir le protocole d'entraînement, comprenant les choix architecturaux et d'optimisation retenus, ainsi que le protocole d'évaluation, mettant en avant la configuration de validation et les métriques choisies.

5.1 Sélection des modèles de la littérature

Cette section vise à définir un sous-ensemble pertinent de modèles issus de la littérature pour la détection d'anomalies non supervisée dans les photographies du segment antérieur de l'œil. Dans cette section est présenté le processus de sélection des modèle, en conformité avec les hypothèses de recherche formulées au **chapitre 3**, tout en détaillant leurs caractéristiques d'implémentation individuelle.

Conformément aux hypothèses de recherche, nous limitons le domaine d'étude de ce projet aux méthodes d'apprentissage profond dites par reconstruction, reposant sur des architectures auto-encodeurs. Ces modèles présentent une grande modularité permettant l'ajout de différents modules à l'entraînement, sans affecter leur taille en phase d'inférence. Ils offrent également une inférence rapide des données en entrée, tout en fournissant un diagnostic à la fois au niveau de l'image et au niveau des pixels. Dans ce qui suit, nous détaillons pour chaque modèle sélectionné dans la littérature, sa fonction d'optimisation, le score d'anomalies utilisé pour la validation, ainsi que les détails d'implémentation, lorsque nécessaires.

AE [74] : Auto-encodeur classique, il s'agit de l'architecture la plus classique des approches par reconstruction, minimisant la distance pixel à pixel entre l'entrée x et sa reconstruction \hat{x} . La plus commune restant la norme L_1 , ou MAE, la fonction objectif de ces modèles peut alors être décrite par l'équation suivante.

$$L_{AE} = \mathbb{E}_{x \sim p(x)} \|x - \hat{x}\|_1 \quad (5.1)$$

La fonction de score d'anomalies final utilisée dans cette approche est équivalente à la fonction

objective :

$$\mathcal{A}_{AE}(x, \hat{x}) = \|x - \hat{x}\|_1 \quad (5.2)$$

AE SSIM [29] : Similaire à un AE classique, AE SSIM privilégie non pas la minimisation d'une distance pixel à pixel, mais la minimisation d'une distance structurelle DSSIM, basée sur une mesure de similarité de voisinage, appelée SSIM [75]. Elle favorise la reconstruction des données présentant des structures ou des textures fines, comme la vascularisation, les cils ou les structures de l'iris. Pour deux patches (x, y) de même position entre deux images, la distance DSSIM est définie par

$$DSSIM(x, y) = \frac{1 - SSIM(x, y)}{2} \quad (5.3)$$

avec

$$SSIM(x, y) = \frac{(2\mu_x, \mu_y + c_1)(2\sigma_x\sigma_y + c_2)(cov(x, y) + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x\sigma_y + c_3)} \quad (5.4)$$

où μ_x et μ_y représentent la moyenne de x et y , σ_x^2 et σ_y^2 la variance, $cov(x, y)$ la covariance, et c_1 , c_2 et c_3 des hyperparamètres de stabilisation. La fonction objectif finale entre l'entrée x et sa reconstruction \hat{x} , est obtenue en calculant la valeur de la distance DSSIM en chaque pixel pour une fenêtre de taille $K \times K$. Les valeurs des hyperparamètres K , c_1 , c_2 et c_3 seront directement repris de l'article introduisant AE SSIM [29]. Le score d'anomalies utilisé sera la version multi-échelle de la SSIM, aussi appelée MS-SSIM [110]. Elle consiste simplement à calculer la valeur SSIM sur les deux images et leurs versions sous-échantillonnées, pour capturer un ensemble d'informations haut niveau et bas niveau.

$$\mathcal{A}_{AE-SSIM}(x, \hat{x}) = 1 - MS-SSIM(x, \hat{x}) \quad (5.5)$$

DPAE [30] : Le modèle DPAE proposé par Shvetsova et. al. exploite la puissance perceptuelle des modèles convolutifs pré-entraînés sur de grandes bases de données, comme ImageNet [41], pour minimiser une représentation profonde intermédiaire plutôt que l'erreur entre l'entrée et la sortie. Les auteurs ont démontré la généralisabilité de leur implémentation sur plusieurs bases de données médicales, et donc potentiellement transférable sur notre base de données. Leur fonction objectif, appelée fonction de perte perceptuelle relative, est définie par :

$$L_{DPAE} = \mathbb{E}_{x \sim p(x)} \frac{\|\tilde{f}(x) - \tilde{f}(\hat{x})\|_1}{\|\tilde{f}(x)\|_1} \quad (5.6)$$

avec

$$\tilde{f}(x) = \frac{f(x) - \mu}{\sigma} \quad (5.7)$$

où f représente une couche intermédiaire d'un modèle VGG19 [44] pré-entraîné, et μ et σ^2 sa moyenne et sa variance, précalculées sur un sous-ensemble d'ImageNet [41]. Nous reprendrons directement leur implémentation, et nous considérerons donc la couche $relu_{42}$, soit la vingt-deuxième couche du réseau, et les valeurs de σ et μ mis à disposition dans [76]. Le score d'anomalies est équivalent à la fonction objectif :

$$\mathcal{A}_{DPAE}(x, \hat{x}) = \frac{\|\tilde{f}(x) - \tilde{f}(\hat{x})\|_1}{\|\tilde{f}(x)\|_1} \quad (5.8)$$

DAE [77] : Une approche par débruitage plutôt que par reconstruction pure, se veut plus robuste face à la fâcheuse tendance des modèles auto-encodeur conventionnel à tendre vers une fonction identité. Nous incorporons donc l'architecture DAE dans notre ensemble de modèles de base pour une exploration de l'impact des tâches de débruitage comparativement à la tâche de reconstruction. Encore une fois, seule la fonction objectif diffère, par ajout d'un bruit gaussien $\epsilon \sim \mathcal{N}(0, I)$ à l'entrée, tel que :

$$L_{DAE} = \mathbb{E}_{x \sim p(x)} \|(x + \epsilon) - \hat{x}\|_1 \quad (5.9)$$

De même, la fonction de score d'anomalies sera celle utilisée pour l'optimisation.

$$\mathcal{A}_{DAE}(x, \hat{x}) = \|(x + \epsilon) - \hat{x}\|_1 \quad (5.10)$$

GANomaly [27] : Le pouvoir génératif des GAN n'est plus à prouver, et leur utilité dans l'application aux photographies du segment antérieur est à explorer. Nous faisons le choix d'introduire à notre jeu de modèles de base le modèle GANomaly. En ajoutant un entraînement antagoniste par discriminateur à un auto-encodeur usuel, la sortie du modèle est contrainte de sorte à tendre vers le même espace de distribution que l'entrée et ainsi favoriser une reconstruction plus fidèle, et donc théoriquement une meilleure détection d'anomalies. Plus encore, les auteurs de GANomaly proposent l'ajout d'un encodeur secondaire E_S , indépendant du reste de l'architecture, qui vise à estimer le vecteur latent $z = E(x)$ de l'encodeur principal E , à partir de la reconstruction. La fonction objectif finale se traduit par :

$$L_{GANomaly} = L_{AE} + L_{adv} + L_{enc} \quad (5.11)$$

avec

$$L_{adv} = \mathbb{E}_{x \sim p(x)} \|f_D(x) - f_D(\hat{x})\|_2 \quad (5.12)$$

$$L_{enc} = \mathbb{E}_{x \sim p(x)} \|E(x) - E_S(\hat{x})\|_2 \quad (5.13)$$

La subtilité ici réside dans le choix de la fonction de perte antagoniste, différente de l'optimisation classique définie par l'équation 2.4. À la place, les auteurs proposent de suivre les conseils d'optimisation de Salimans et. al. [85], et de minimiser la correspondance des caractéristiques intermédiaires du discriminateur pour faciliter la convergence du modèle, exprimé par L_{adv} . Enfin, plutôt que d'estimer l'erreur pixel à pixel entre l'entrée et sa reconstruction, il est proposé de comparer l'erreur des vecteurs de représentations, plus expressifs des attributs discriminatifs des deux images. La fonction de score d'anomalies s'exprime alors comme suit :

$$\mathcal{A}_{GANomaly}(x, \hat{x}) = \|E(x) - E_S(\hat{x})\|_1 \quad (5.14)$$

VAE [79] : Des additions importantes à nos modèles de base sont les modèles variationnels, notamment leur version la plus basique avec les VAE. La difficulté dans l'implémentation de ces modèles réside dans la notion d'échantillonnage du vecteur latent $z \sim q_\phi(\cdot|x)$, opération non différentiable. Pour la rendre différentiable, on procède à une reparamétrisation du vecteur latent, et on scinde le problème en deux étapes :

1. La paramétrisation : on définit la distribution directement par ses paramètres, soit μ et σ puisqu'il s'agit d'une distribution normale. Ces deux paramètres sont donnés directement en sortie de l'encodeur.
2. L'échantillonnage différentiable : on récupère un échantillon de notre distribution à partir d'un échantillon secondaire $\epsilon \in \mathcal{N}(0, I)$ que l'on recentre selon les paramètres obtenus précédemment, de sorte que

$$z = \mu + \sigma \cdot \epsilon$$

De cette manière, le processus d'échantillonnage devient entièrement différentiable, et le réseau peut être entraîné selon les méthodes classiques d'optimisation par descente de gradient. En reprenant l'équation 2.3 des VAE, la fonction objectif du modèle peut alors s'écrire :

$$L_{VAE} = L_{AE} + L_{KL} \quad (5.15)$$

avec

$$L_{KL} = D_{KL}(z||\mathcal{N}(0, I)) = 0.5 * (\mu_z^2 + \sigma_z^2 - \log(\sigma_z)^2 - 1) \quad (5.16)$$

Les VAE étant stochastiques, la reconstruction en sortie ne sera pas la même pour deux entrées consécutives identiques. De fait, la sortie n'est pas une reconstruction à proprement parler, mais une carte de probabilité de reconstruction. Pour pallier ce problème, on considèrera l'erreur de probabilité de reconstruction entre x et sa reconstruction \hat{x} , sur une moyenne de N échantillonnages. Nous fixons arbitrairement N à 50 échantillonnages pour l'ensemble des méthodes variationnelles qui suivront.

$$\mathcal{A}_{VAE}(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N \|x - \hat{x}_i\|_1 \quad (5.17)$$

AnoVAEGAN [84] : Le modèle AnoVAEGAN proposé par Baur et. al. est une combinaison des modèles variationnels VAE et des modèles antagonistes GAN. L'idée derrière cette architecture est de combler le manque de capacité générative des VAE, en ajoutant un module antagoniste pour une meilleure correspondance des espaces de probabilités d'entrée et de sortie. La fonction objective est simplement une combinaison linéaire des termes des deux approches, soit

$$L_{AnoVAEGAN} = \lambda_1 L_{AE} + \lambda_2 L_{KL} + \lambda_3 L_{adv} \quad (5.18)$$

avec $\lambda_1 = 10$, $\lambda_2 = 1$ et $\lambda_3 = 0.1$, des poids définis empiriquement durant la validation sur nos données. Nous proposons pour cette implémentation une légère variation, comme proposée dans [31], en remplaçant la fonction objectif antagoniste classique par celle de Wasserstein [111], avec pénalité de gradient [83]. Cette variante offre une meilleure stabilité d'entraînement et évite les problèmes récurrents de réduction du mode (ou "mode collapse"). Elle consiste à remplacer le discriminateur binaire en un critique, dont le score réel de sortie indique, s'il est positif, que l'entrée provient de l'espace de distribution des données réelles, et, inversement s'il est négatif, qu'elle est une donnée générée. La divergence de Jensen-Shannon, classiquement utilisée dans la définition des problèmes d'optimisation des GAN est remplacée par la distance de Wasserstein (ou "Earth Mover's distance"). Cette dernière implique cependant que le critique soit k -Lipschitz, d'où l'introduction de la pénalité de gradient par Gulrajani et. al. [83] dans l'entraînement du critique pour imposer cette propriété. Nous utiliserons directement

les implémentations disponibles sur [112], ainsi que les mêmes coefficients utilisés dans [83]. Le score d’anomalies final suit le même processus que celui des VAE :

$$\mathcal{A}_{AnoVAEGAN}(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N \|x - \hat{x}_i\|_1 \quad (5.19)$$

SSLAnoVAE [99] : La dernière méthode de la littérature que nous considérerons est SSLAnoVAE, un modèle combinant apprentissage auto-supervisé et reconstruction, et dont les auteurs démontrent son efficacité pour le diagnostic de la rétinopathie diabétique sur des images du fond d’œil. Il est fondé sur le modèle P-Net [113], une architecture en deux branches d’encodeur, qui permet l’ajout d’une information à priori à l’espace de représentation pour faciliter la tâche de reconstruction. Pour les images du fond d’œil, l’information à priori considérée est le réseau vasculaire, information fine souvent perdue au travers des couches de l’encodeur. Cependant, elle suppose l’existence d’une donnée à priori à l’entraînement, sans garantie que celle-ci soit véritablement pertinente à la détection d’anomalies. Les auteurs de SSLAnoVAE proposent ainsi d’apprendre cette information à priori par approche SSL au moyen d’un module BYOL [97], en parallèle du travail de reconstruction par double VAE, se traduisant par une fonction objectif de la forme.

$$L_{SSLAnoVAE} = \lambda_1 L_{AE} + \lambda_2 L_{KL1} + \lambda_2 L_{KL2} + \lambda_3 L_{adv} + \lambda_4 L_{SSL} \quad (5.20)$$

où L_{adv} est la fonction objective d’un GAN classique, et L_{SSL} la fonction objectif explicitée par les auteurs de BYOL [97]. Malgré les bonnes performances apparentes de SSLAnoVAE [99], aucune implémentation n’a été rendue disponible et de nombreux détails techniques sont manquants. On se propose de réimplémenter l’architecture au mieux avec les informations à notre disposition. Puisqu’il s’agit d’une architecture avec échantillonnage, nous utiliserons la même fonction de score d’anomalies que précédemment.

$$\mathcal{A}_{SSLAnoVAE}(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N \|x - \hat{x}_i\|_1 \quad (5.21)$$

5.2 Configuration unifiée d’entraînement

Bien que tous basés sur le principe de reconstruction par auto-encodeur, les modèles de la littérature sélectionnés ainsi que notre approche SiamAAE présentée en section 4.2, présentent des conceptions d’architecture de réseau et d’optimisation diverses et variées. Pour assurer une comparaison équitable entre ces modèles de base, nous proposons d’unifier leur entraîne-

ment, comprenant le choix des transformations des données en entrée des modèles, les choix architecturaux et les choix de configurations d’optimisation.

Transformations initiales

Les données d’entraînement sont issues du prétraitement décrit en section 4.1.4. Elles sont constituées d’images carrées de dimensions $3 \times 2000 \times 2000$, et de valeurs comprises entre 0 et 255. En raison des contraintes de ressources physiques et temporelles, un redimensionnement par interpolation bilinéaire est appliqué à l’image pour obtenir une taille finale de $3 \times 256 \times 256$. Bien que cette réduction de dimension entraîne une perte d’information, elle est jugée négligeable pour une taille de 256×256 . De plus, les données sont ramenées entre -1 et 1 pour des questions d’échelle et de facilité de convergence.

Architecture unifiée

Tous les modèles étudiés seront ré-entraînés selon l’architecture unifiée décrite dans la figure 5.1, inspirée des travaux de Baur et al. [31] sur la détection d’anomalies dans les volumes IRM cérébraux. Cette architecture a été élaborée à partir des conceptions standards de la littérature et d’une série d’essais empiriques appliqués à la tâche de reconstruction de photographies du segment antérieur de l’œil.

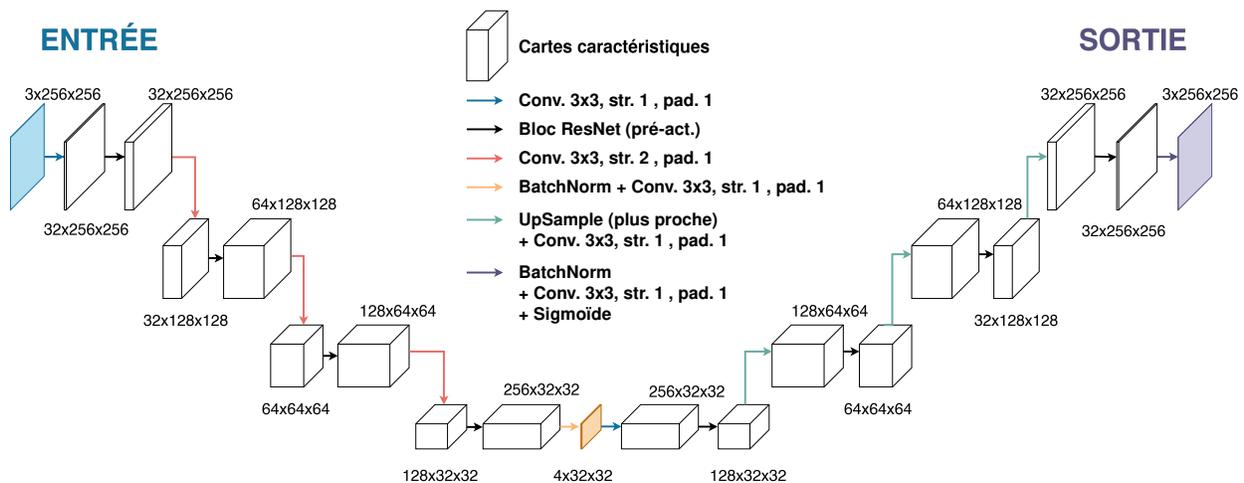


FIGURE 5.1 Schéma de l’architecture unifiée de l’auto-encodeur

Il s’agit d’une architecture entièrement convolutive, ce qui lui confère l’avantage d’être indépendante des dimensions d’entrée et de préserver l’information spatiale à travers les différents niveaux de profondeur. Ce choix repose sur les conclusions des analyses de [31, 84] et sur nos

propres expérimentations, qui montrent une meilleure capacité de reconstruction pour les architectures entièrement convolutives par rapport à celles utilisant une transformation linéaire du vecteur de représentation au niveau du goulot d'étranglement. Pour améliorer l'expressivité des modèles lors de l'apprentissage et prévenir les problèmes de disparition du gradient au travers des différentes couches de l'auto-encodeur, nous optons pour l'implémentation de blocs ResNet à chaque niveau du réseau [45]. L'ajout d'une connexion résiduelle facilite le transfert du gradient à travers les couches lors de la rétropropagation. Plus précisément, notre implémentation du bloc ResNet, décrite en figure 5.2, est basée sur sa variante avec pré-activation [114], offrant de meilleures performances de généralisation et de convergence.

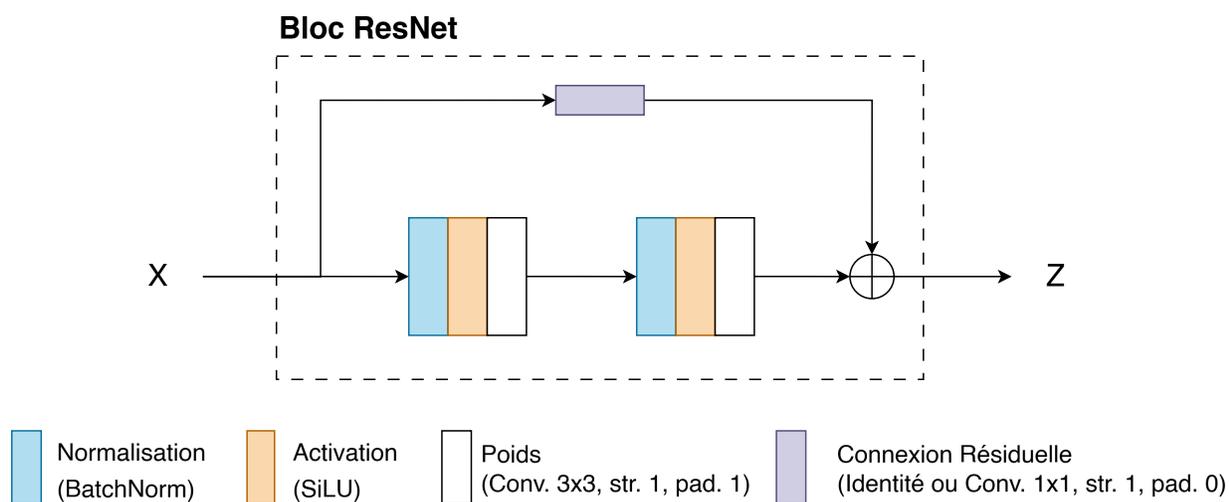


FIGURE 5.2 Schéma de l'architecture du bloc ResNet avec pré-activation

Nous avons également choisi d'utiliser la fonction d'activation SiLU (ou "Sigmoid-Weighted Linear Unit") [115], au lieu des fonctions plus classiques, telles que ReLU (ou "Rectified Linear Unit") ou encore LeakyReLU (ou "Leaky Rectified Linear Unit") [116]. Contrairement à ces dernières, qui présentent une discontinuité en $x = 0$ pouvant engendrer des sauts dans l'optimisation par descente de gradient, la fonction SiLU est lisse et différentiable en tout point. Cela favorise une meilleure convergence durant l'entraînement et améliore la conservation de l'information autour de zéro, ce qui est bénéfique pour des tâches d'extraction de caractéristiques fines, comme la détection d'anomalies.

Dans les modèles variationnels, le goulot d'étranglement (indiqué en jaune dans la figure 5.1) est scindé en deux branches suivies d'une couche de convolution supplémentaire, sans activation, pour prédire les paramètres μ et σ selon l'astuce de reparamétrisation présentée dans la section précédente.

Enfin, pour les modèles entraînés de manière antagoniste, une architecture commune du discriminateur est également proposée. Nous avons opté, là encore, pour une architecture entièrement convolutive, illustrée en figure 5.3. Il s'agit d'un discriminateur classique, qui produit en sortie une probabilité $D(X) \in [0, 1]$ indiquant si l'image d'entrée X est réelle ($D(X) = 1$) ou générée ($D(X) = 0$).

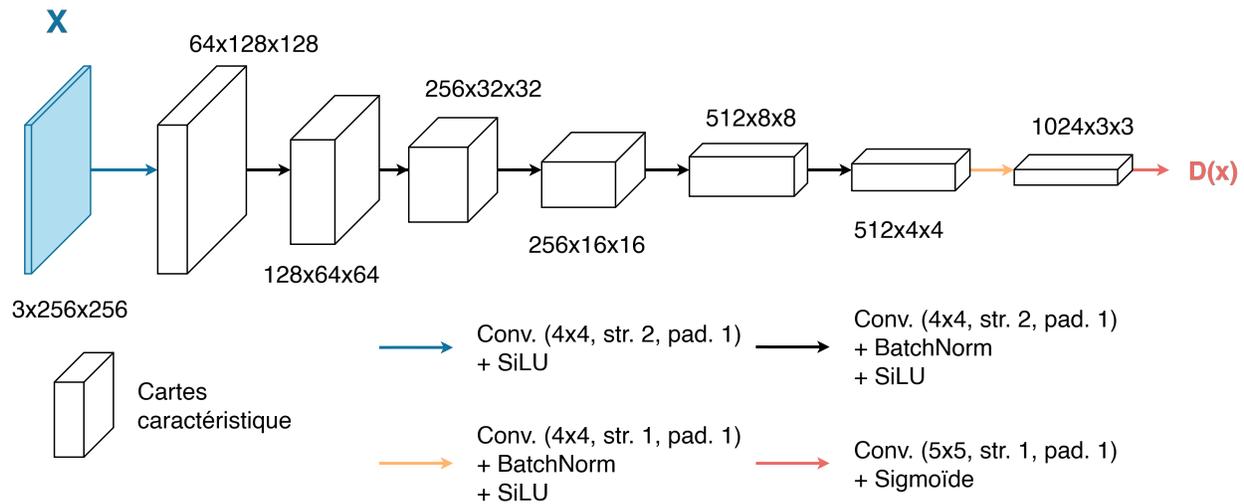


FIGURE 5.3 Schéma de l'architecture du discriminateur

Configuration d'optimisation

Pour tous les modèles considérés, nous utilisons l'optimiseur stochastique Adam [117], couplé à une planification cosinus du taux d'apprentissage (illustrée en figure 5.4) pour une meilleure convergence. Le choix des hyperparamètres finaux est réalisé par optimisation des paramètres à la validation qui maximisent la tâche de détection d'anomalies.

5.3 Procédés d'évaluation

Les méthodes d'évaluation des modèles de la littérature sélectionnés (section 5.1) et notre approche SiamAAE (section 4.2) sont décrites dans la section qui suit. Elle comprend la description des choix de partitionnement des données pour l'entraînement, la validation et le test, mais aussi le choix des métriques d'évaluations pour la tâche de détection d'anomalies, et celles pour la tâche de reconstruction.

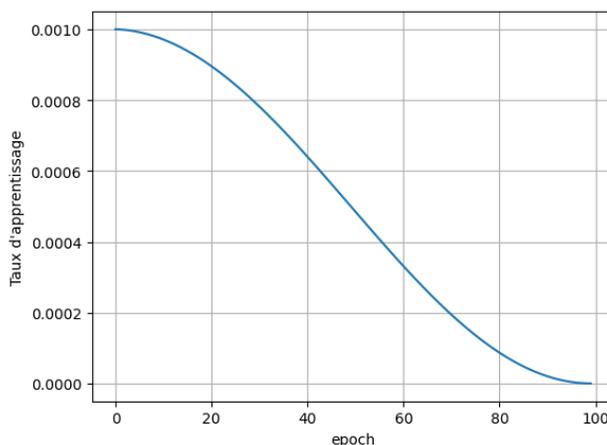


FIGURE 5.4 Exemple de planification cosinus du taux d'apprentissage - de $1.0E-3$ à $1.0E-8$

5.3.1 Répartition des données

L'évaluation de modèles d'apprentissage profond nécessite la répartition des données en trois jeux de données distincts, soient les ensembles d'entraînement, de validation et de test. Afin de garantir l'absence de chevauchement entre ces groupes et pour éviter toute fuite d'information, nous commençons par établir l'ensemble de test, qui sera composé de 15% des individus de la base de données LightX, répartis équitablement entre les sous bases POLY, EOM et IOL. Cet ensemble a été fixé dès le début du projet et n'a été utilisé que lors de la phase finale d'évaluation. Puisque nous nous intéressons uniquement à des approches non supervisées, nous séparons le reste des données entre les individus normaux et anormaux. Un individu est considéré anormal si au moins une photographie est annotée comme anormale. Parmi les individus normaux restants, 80% d'entre eux sont gardés pour l'entraînement, et le reste pour la validation. Tous les individus anormaux sont quant à eux gardés pour la validation. De cette manière, l'ensemble d'entraînement ne contient que des données normales, et aucun individu ne se retrouve dans deux ensembles différents. Les modèles seront ensuite entraînés sur 100 époques et avec 5 graines aléatoires différentes, entraînant une répartition aléatoire des individus entre les ensembles d'entraînement et de validation.

5.3.2 Tâche de détection d'anomalies

Chaque modèle étudié est muni d'une fonction de score d'anomalies, qui, pour une image d'entrée fournit un score réel, dont les valeurs les plus élevées indiquent une plus grande probabilité d'anomalies.

Afin d'éviter la dépendance à un seuil spécifique, nous évaluerons la tâche de détection

d'anomalies à l'aide des métriques ROC AUC et PR AUC présentées en section 2.4. Celles-ci seront calculées à partir du score d'anomalie de chaque modèle, en se basant sur les annotations binaires « normal » ou « anormal » disponibles. Pour les 5 graines aléatoires, nous établirons la moyenne et l'écart type des métriques retenues.

5.3.3 Tâche de reconstruction

La tâche de reconstruction, intrinsèquement liée à la détection d'anomalies, sera également évaluée à travers diverses expérimentations en utilisant des métriques de similarité entre deux images. Les métriques retenues sont les suivantes :

- L'erreur moyenne absolue MAE (équation 5.2), pour une évaluation de l'erreur pixel à pixel.
- Les mesures de similarités SSIM et MS-SSIM (équation 5.4), pour une évaluation de l'erreur structurelle.
- L'erreur moyenne perceptuelle relative tirée des travaux de DPAE [30] (équation 5.7), que l'on nommera dorénavant MRPE (ou "Mean Relative Perceptual Error"), pour une évaluation de l'erreur perceptuelle et sémantique à partir d'un réseau VGG pré-entraîné sur ImageNet.

Les capacités de reconstruction seront évaluées sur le score de reconstruction des images "normales" uniquement, car c'est sur cette tâche que les modèles ont été entraînés. De même, pour les 5 graines aléatoires, nous établirons la moyenne et l'écart type des métriques retenues.

Ces métriques seront également appliquées pixel à pixel, pour la visualisation des cartes d'anomalies et une analyse qualitative de l'erreur de reconstruction se fera par comparaison avec les annotations cliniques obtenues (section 4.1.3).

CHAPITRE 6 RÉSULTATS & DISCUSSION

Ce chapitre présente les expérimentations ainsi que les résultats pour chaque modèle de la littérature évalué, ainsi que pour notre approche SiamAAE, conformément au protocole d’entraînement et d’évaluation défini dans le chapitre précédent. Les résultats obtenus sont ensuite analysés et discutés afin de mettre en évidence les avantages et les limites de chaque approche étudiée. L’analyse inclut l’évaluation des résultats de la collecte de données, de la détection d’anomalies, ainsi que des capacités de reconstruction.

6.1 À propos de la base de données

La collecte de données a été réalisée auprès de 99 participants, pour un total de 10 825 photographies prises. Pour notre étude, ce nombre a été réduit à 2174 photographies du test de la conjonctive, qui constituent notre base de normalité pour l’entièreté du projet. Concernant les annotations, un total de 473 images ont été annotées au niveau pixel et au niveau image par l’ophtalmologiste. Cela correspond à l’entièreté des photographies du test de la conjonctive de la base de données EOM, soit 166 échantillons, pour 307 provenant de la base de données IOL. De cette étape supplémentaire, nous avons pu identifier que 209 d’entre elles pouvaient être considérées comme normales, car aucun signe pathologique n’était visible. La fusion de toutes ces données forme notre base de données définitive, conçue spécifiquement pour une tâche de détection d’anomalies des photographies du segment antérieur de l’œil. Ses caractéristiques sont récapitulées au travers de la table 6.1.

Provenance	Classe	Nombre de données
EOM	Normal	68
	Anormal	98
IOL	Normal	141
	Anormal	166
POLY	Normal	2174
	Anormal	0

TABLEAU 6.1 Récapitulatif des photographies exploitables à la détection d’anomalies et leur provenance

L’annotation des données est une tâche longue et fastidieuse, et ce, même pour un spécialiste

habitué à l'observation de données médicales. C'est pourquoi notre base de données admet un tel déséquilibre : elle est composée de 90% d'images normales pour seulement 10% d'images anormales. Cette répartition justifie le choix d'utiliser des approches d'apprentissage de la normalité non supervisées, qui nécessitent uniquement des données normales à l'entraînement.

Cependant, il est important de noter que 91% des images normales sont issues de la collecte de données de Polytechnique Montréal. La base de données POLY qui en découle présente une bonne diversité ethnique avec une représentation supérieure à 10% d'au moins 4 groupes ethniques distincts, ainsi qu'une répartition presque égale entre hommes (53,5%) et femmes (46,5%), comme le montrent les diagrammes en figure 6.1. La population cible de la collecte étant la communauté étudiante, l'âge moyen des participants tourne autour des 24 ans, avec très peu de participants recensés au delà des 30 ans (Figure 6.3). En revanche, les photographies des bases de données IOL et EOM ont été prises lors de rendez-vous cliniques, et, bien qu'aucune statistique précise ne soit disponible, il nous a été indiqué que ces patients sont généralement plus âgés, avec une moyenne d'âge estimée autour de 50 ans, voire davantage. D'autre part, la collecte de données POLY a été réalisée sans l'instillation de gouttes de fluorescéine, tandis que les bases de données IOL et EOM contiennent des photographies avec fluorescéine, qui se manifeste sous la forme d'un film jaunâtre au niveau de la sclère et des paupières.

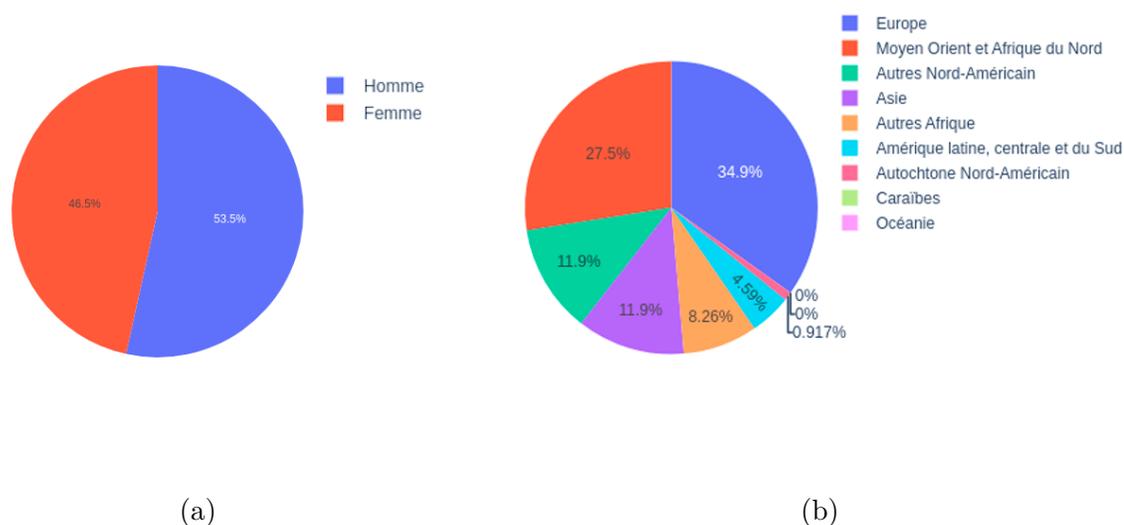


FIGURE 6.1 Étude démographique des participants de la collecte de données POLY

(a) Diagramme circulaire de la répartition des genres - (b) Diagramme circulaire des origines ethniques

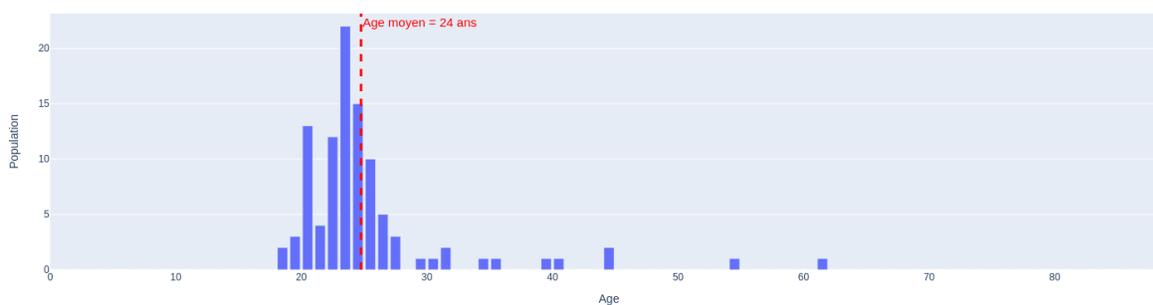


FIGURE 6.3 Histogramme de répartition de l'âge des participants à la collecte de données POLY

Ces différences entre les bases de données POLY, IOL et EOM sont des biais de notre base de données, et peuvent impacter directement l'entraînement de nos modèles. Les attributs de la normalité appris seront les attributs principaux des données de POLY, et la présence de fluorescéine ou l'âge de l'individu pourront être utilisés comme facteurs discriminants durant de la détection d'anomalies, bien qu'ils ne soient pas nécessairement la cause de l'anormalité de l'image. Cet effet pourrait être estompé en augmentant le nombre de données issues des bases de données IOL et EOM à l'entraînement, ce qui nécessiterait une phase d'acquisition et d'annotation supplémentaire.

6.2 Analyse et comparaison des capacités de reconstruction et de détection d'anomalies

Dans cette section, sont présentées et discutées les expérimentations vouées à l'étude de l'impact des choix d'implémentation des modèles de la littérature, ainsi que de notre approche SiamAAE, sur leur performance en détection d'anomalies et en reconstruction d'images, selon la méthodologie d'évaluation établie dans la section 5.3.

Pour s'assurer de la généralisation de nos implémentations, l'expérimentation est également reproduite sur la base de données publique de photographies du fond d'œil RFMiD [13], selon le même protocole. La synthèse des résultats sur la base de données LightX est accessible dans le tableau 6.2 et une même synthèse pour la base de données RFMiD est disponible au tableau 6.3. Une visualisation des reconstructions après entraînement pour chaque modèle sur des photographies du segment antérieur de l'œil est illustrée à la figure 6.4, ainsi que sur des photographies du fond d'œil à la figure 6.5.

LightX	Détection d'anomalies		Reconstruction (images normales uniquement)			
	ROC AUC \uparrow	PR AUC \uparrow	MAE \downarrow	SSIM \uparrow	MS-SSIM \uparrow	MRPE \downarrow
AE MAE	0,547 \pm 0,042	0,089 \pm 0,007	1.71e-02 \pm 2.81e-03	0.675 \pm 0.027	0.907 \pm 0.016	4.72e-01 \pm 2.27e-02
AE SSIM	0,563 \pm 0,014	0,097 \pm 0,005	1.85e-02 \pm 7.02e-04	0.751 \pm 0.009	0.923 \pm 0.003	4.20e-01 \pm 1.14e-02
DPAE	<u>0,713 \pm 0,008</u>	0,237 \pm 0,008	2.87e-02 \pm 9.10e-04	0.476 \pm 0.009	0.844 \pm 0.007	2.46e-01 \pm 5.06e-03
DAE	0,663 \pm 0,026	<u>0,143 \pm 0,019</u>	2.98e-02 \pm 8.75e-04	0.526 \pm 0.005	0.692 \pm 0.006	5.44e-01 \pm 4.38e-03
VAE	0,511 \pm 0,022	0,084 \pm 0,001	<u>1.75e-02 \pm 2.99e-03</u>	0.646 \pm 0.03	0.883 \pm 0.031	4.82e-01 \pm 1.95e-02
AnoVAEGAN	0,511 \pm 0,023	0,086 \pm 0,03	1.78e-02 \pm 2.70e-03	0.637 \pm 0.04	0.889 \pm 0.019	4.76e-01 \pm 2.33e-02
GANomaly	<u>0,695 \pm 0,22</u>	0,131 \pm 0,01	1.57e-02 \pm 7.25e-04	<u>0.679 \pm 0.014</u>	<u>0.916 \pm 0.006</u>	<u>4.58e-01 \pm 2.22e-02</u>
SSLAnoVAE	0,41 \pm 0,174	0,072 \pm 0,028	1.04e-01 \pm 4.80e-02	0.166 \pm 0.253	0.533 \pm 0.247	7.02e-01 \pm 1.51e-01
SiamAAE	0,726 \pm 0,008	<u>0,229 \pm 0,01</u>	2.07e-02 \pm 1.2e-03	<u>0.704 \pm 0.007</u>	0.901 \pm 0.004	<u>2.91e-01 \pm 1.41e-02</u>

TABLEAU 6.2 Résultats de la littérature et de notre approche sur la tâche de détection d'anomalies et de reconstruction sur les photographies du segment antérieur de l'œil

Moyennes et écarts types obtenus sur 5 graines aléatoires différentes, sur l'ensemble de test de la base de données de LightX Innovations - Les modèles ont été ré-entraînés selon le protocole établi au chapitre 5 - **En gras** les meilleurs résultats - En souligné les deuxième et troisième meilleurs résultats

6.2.1 Influence des implémentations

Voici un résumé des résultats importants à retenir des deux tables 6.2 et 6.3. Pour la base de données LightX, les modèles de la littérature les plus performants en termes de détection d'anomalies et de capacité de reconstruction sont respectivement DPAE [30] (ROC AUC : 0.713 - PR AUC : 0.237) et GANomaly [27] (MAE : 1.57e-2 - SSIM : 0.679 - MS-SSIM : 0.916 - MRPE : 4.58e-1). De même pour la base de données RFMiD, il s'agit respectivement de DAE [77] (ROC AUC : 0.589 - PR AUC : 0.862) et de AE SSIM [29] (MAE : 2.28e-2 - SSIM : 0.651 - MS-SSIM : 0.888 - MRPE : 5.12e-1). Notre approche SiamAAE, quant elle, est la deuxième plus performante pour la détection d'anomalies (ROC AUC : 0.563 - PR AUC : 0.852), derrière DAE pour la base de données RFMiD, et première selon la ROC AUC sur la base de données LightX (ROC AUC : 0.726 - PR AUC : 0.229), tout en ayant les deuxièmes meilleures performances sur les deux bases de données selon les métriques de reconstruction structurelle SSIM et perceptuelle MRPE. Elle présente également une des meilleures, si ce n'est la meilleure stabilité entre chaque entraînement (écarts types de l'ordre de 1%).

Les résultats étant dans l'ensemble compacts, une discussion plus fine des mécanismes sous-jacents de chacun des modèles évalués est proposée à la suite. À partir des modèles de la littérature évalués, nous distinguons trois facteurs majeurs possible d'influence des performances de détection d'anomalies et de reconstruction, qui sont :

- la fonction objectif de reconstruction,

RFMiD	Détection d'anomalies		Reconstruction (images normales uniquement)			
	ROC AUC \uparrow	PR AUC \uparrow	MAE \downarrow	SSIM \uparrow	MS-SSIM \uparrow	MRPE \downarrow
AE MAE	0,555 \pm 0,031	0,848 \pm 0,009	<u>2.12e-02 \pm 1.59e-03</u>	0.523 \pm 0.028	<u>0.846 \pm 0.017</u>	5.70e-01 \pm 3.21e-02
AE SSIM	0,384 \pm 0,015	0,728 \pm 0,012	2.28e-02 \pm 2.38e-03	0.651 \pm 0.024	0.888 \pm 0.012	<u>5.12e-01 \pm 3.10e-02</u>
DPAE	0,459 \pm 0,093	0,806 \pm 0,05	1.23e-01 \pm 6.12e-02	0.013 \pm 0.26	0.581 \pm 0.295	2.79e-01 \pm 4.46e-02
DAE	0,589 \pm 0,024	0,862 \pm 0,007	2.85e-02 \pm 1.10e-03	0.439 \pm 0.014	0.673 \pm 0.013	6.05e-01 \pm 1.04e-02
VAE	0,546 \pm 0,076	0,843 \pm 0,031	2.22e-02 \pm 3.07e-03	0.511 \pm 0.019	0.784 \pm 0.043	5.62e-01 \pm 6.32e-03
AnoVAEGAN	0,505 \pm 0,073	0,825 \pm 0,036	1.97e-02 \pm 1.86e-03	<u>0.534 \pm 0.014</u>	0.814 \pm 0.009	5.58e-01 \pm 9.48e-03
GANomaly	<u>0,56 \pm 0,061</u>	0,832 \pm 0,03	<u>2.11e-02 \pm 1.49e-03</u>	0.522 \pm 0.019	<u>0.837 \pm 0.011</u>	5.88e-01 \pm 7.95e-03
SSLAnoVAE	0,405 \pm 0,104	0,764 \pm 0,031	1.00e-01 \pm 5.70e-02	-0.162 \pm 0.402	0.47 \pm 0.225	6.12e-01 \pm 5.64e-02
SiamAAE	<u>0,563 \pm 0,019</u>	<u>0,852 \pm 0,006</u>	4.12e-02 \pm 2.19e-03	<u>0.586 \pm 0.014</u>	0.836 \pm 0.028	<u>2.81e-01 \pm 1.00e-02</u>

TABLEAU 6.3 Résultats de la littérature et de notre approche sur la tâche de détection d'anomalies et de reconstruction sur les photographies du fond d'œil

Moyennes et écarts types obtenus sur 5 graines aléatoires différentes, sur l'ensemble de test de la base de données RFMiD [13] - Les modèles ont été ré-entraînés selon le protocole établi au chapitre 5 - **En gras** les meilleurs résultats - En souligné les deuxième et troisième meilleurs résultats

- la régularisation de l'espace latent par approche variationnelle,
- la régularisation de la génération par approche antagoniste.

De cette analyse, nous pourrions ensuite justifier et analyser la conception de notre approche.

Objectif de reconstruction

L'objectif de reconstruction influence directement la capacité des modèles à reconstruire les images d'entrée. L'utilisation de la fonction de perte basée sur la similarité structurelle (SSIM), par rapport à l'erreur absolue moyenne (MAE), améliore notablement la qualité de reconstruction, tant sur les données LightX que sur celles de RFMiD (augmentation de 10% des métriques SSIM et de 5% des métriques MS-SSIM sur les deux bases de données). Contrairement aux reconstructions floues produites par AE MAE [74], les reconstructions de AE SSIM [29] sont beaucoup plus nettes, en particulier pour les détails fins, comme le réseau vasculaire (figures 6.4 et 6.5). Les photographies du segment antérieur et du fond d'œil présentent en effet de nombreuses structures fines, ce qui explique l'avantage d'une fonction de perte basée sur la similarité structurelle SSIM plutôt que sur une simple erreur pixel à pixel. Cependant, ce gain en précision de reconstruction se traduit par une amélioration peu significative des capacités de détection d'anomalies sur LightX (table 6.2) et par une diminution des performances sur RFMiD (table 6.3).

L'utilisation de l'erreur perceptuelle profonde, avec DPAE [30], montre également des caracté-

ristiques intéressantes. Bien que le modèle ne présente pas de performances de reconstruction élevées sur les métriques MAE, SSIM et MS-SSIM, il se distingue par de meilleurs scores en MRPE (quasiment divisé par 2), la métrique sur laquelle il a été entraîné. Visuellement, cela se traduit par des reconstructions fidèles à l'image d'entrée, qualitativement supérieures à celles produites par AE MAE, avec une meilleure conservation des détails et des structures anatomiques fines (figures 6.4 et 6.5). Les mauvais résultats affichés sont dus à la sensibilité des métriques MAE, SSIM et MS-SIM aux écarts au niveau pixel des couleurs, de l'illumination ou du contraste. En effet, la différence au niveau des cartes caractéristiques profondes permet de minimiser l'écart sémantique entre deux images, mais induit nécessairement une perte d'information au niveau des pixels. Toutefois, cette notion de minimisation de l'écart sémantique semble être la raison derrière l'amélioration des performances en ce qui concerne la détection d'anomalies pour le modèle DPAAE en comparaison aux autres approches.

Enfin, la transformation de la tâche de reconstruction en une tâche de débruitage, par ajout de bruit en entrée, réduit sensiblement les capacités de reconstruction comparées aux autres approches. Cette diminution, attendue en raison de la perte d'information introduite dans les données d'entrée, résulte en des reconstructions floues, dépourvues de structures fines (figures 6.4 et 6.5). Néanmoins, cette réduction de précision dans la reconstruction est compensée par une meilleure performance en détection d'anomalies. En débruitant, le modèle DAE est plus apte à "normaliser" les anomalies en les rapprochant de l'espace de distribution des données saines, d'où ses meilleures performances en détection d'anomalies sur les bases de données RFMiD et LightX.

Régularisation de l'espace de représentation

Les modèles variationnels, tels que VAE [79], AnoVAEGAN [84] et SSLAnoVAE [99], contraignent l'espace de représentation en réduisant la quantité d'information encodée pour ne conserver que l'essentiel. En comparant les modèles AE MAE et VAE, on observe une diminution des performances de reconstruction par ajout du terme variationnel, particulièrement pour les métriques SSIM et MS-SSIM, reflétant une perte d'information structurelle tant fine que globale. Cela se reflète particulièrement bien dans les reconstructions des images du fond d'œil à la figure 6.5), où les reconstructions pour ses modèles apparaissent très floues, et où la vascularisation est difficile à discerner, comparativement aux reconstructions des modèles non variationnels.

Contrairement aux observations souvent rapportées dans la littérature, l'ajout d'une contrainte variationnelle n'a pas amélioré les performances de détection d'anomalies dans notre cas. Une explication possible de ce constat réside dans nos choix d'implémentation. Nous avons adopté

une architecture entièrement convolutive, commune à tous les modèles, avec un espace de représentation préservant la structure spatiale de l'image (figure 5.1). Ce choix répondait d'une part à un besoin d'uniformiser l'architecture afin de permettre une comparaison équitable des modèles, et d'autre part, à un objectif d'optimiser les capacités de reconstruction pour des images riches en détails structurels et anatomiques fins. Il est possible que cette préservation de la spatialité complique la régularisation de l'espace de représentation en raison du grand nombre de paramètres. Une piste d'amélioration pourrait consister à explorer les β -VAE, qui introduisent un coefficient $\beta > 1$ pour renforcer la régularisation de l'espace latent en augmentant la force de la contrainte de divergence de Kullback-Leibler (équation 2.3).

Réseaux génératifs antagonistes

Les modèles antagonistes présents dans notre ensemble de modèles de la littérature sont GANomaly [27], AnoVAEGAN [84] et SSLAnoVAE [99]. La fonction de perte antagoniste a pour objectif de rapprocher l'espace de distribution des données générées à l'espace de distribution de la normalité. Une amélioration des performances de reconstruction est donc attendue. On retrouve cette logique au travers des résultats de reconstruction de GANomaly comparé à AE MAE, de même pour AnoVAEGAN sur la base de données RFMiD, comparé à VAE. Visuellement parlant, il est difficile d'identifier les différences de reconstruction entre ces modèles donnés dans les figures 6.4 et 6.5. Et pour cause, la comparaison des métriques de reconstruction dans les tables 6.2 et 6.3, montre que l'amélioration, si elle existe, n'est que très légère, les ordres restant sensiblement les mêmes.

Un fait marquant cependant est le manque de stabilité de ces approches. En effet, on remarque des écarts types très élevés pour GANomaly et SSLAnoVAE, de l'ordre de 20% pour la ROC AUC lors de l'entraînement sur la base de données de LightX sur les 5 graines aléatoires. Il s'agit d'un problème récurrent des GAN, dont il est difficile de s'extraire. AnoVAEGAN est quant à lui plutôt stable, ce qui s'explique par l'utilisation des Wasserstein GAN [111], créés dans cet objectif.

Notre approche

La comparaison des performances des modèles de la littérature met en avant le compromis classique entre reconstruction et détection d'anomalies [31,86]. En général, augmenter les performances de reconstruction résulte en la dégradation des performances de détection d'anomalies, que ce soit au travers du choix de la fonction objectif de reconstruction (AE MAE, AE SSIM, DPAAE et DAAE), de la régularisation de l'espace de représentation (VAE, AnoVAEGAN et SSLAnoVAE) ou de l'introduction d'un entraînement antagoniste (GANomaly,

AnoVAEGAN et SSLAnoVAE). Nous avons conçu notre approche dans l'objectif d'optimiser ce compromis. Dans l'étude menée jusqu'à maintenant, il semblerait qu'une reconstruction optimisée d'une photographie du segment antérieur de l'œil pourrait être obtenue avec une fonction de perte SSIM (AE SSIM) ou un apprentissage antagoniste (GANomaly), tandis qu'une détection d'anomalies plus généralisable pourrait être obtenue avec une fonction de perte perceptuelle profonde (DPAE) ou une tâche de débruitage (DAE). Notre approche SiamAAE prend en compte ces deux aspects, en minimisant simultanément l'erreur structurelle et perceptuelle, et en introduisant une tâche de correction de pseudo-anomalie, à l'instar du débruitage, pour atteindre un compromis intéressant entre capacité de reconstruction et de détection d'anomalies. Sa stabilité interne et entre les bases de données en fait également un très bon choix pour une application à la détection d'anomalies.

L'impact du module de guidage SSL aurait pu être comparé au modèle SSLAnoVAE, qui comprend également un module contrastif secondaire dans son entraînement. Cependant, son manque crucial de stabilité (écart type allant de 10% à 20% sur la plupart des métriques) empêche toute comparaison. L'analyse du module de guidage SSL fera l'objet d'une section spécifique dans la suite de ce chapitre.



FIGURE 6.4 Exemple de reconstructions de photographies "normales" du segment antérieur de l'œil

Données issues de l'ensemble test de la base de données LightX - **Première ligne** : Images d'entrées des modèles - **Lignes suivantes** : Reconstructions correspondantes à chaque entrée pour le modèle considéré, indiqué à gauche



FIGURE 6.5 Exemple de reconstructions de photographies "normales" du fond d'œil

Données issues de l'ensemble test de la base de données RFMiD [13] - **Première ligne** : Images d'entrées des modèles - **Lignes suivantes** : Reconstructions correspondantes à chaque entrée pour le modèle considéré, indiqué à gauche

6.2.2 Robustesse face à l'apprentissage de l'identité

Les approches non supervisées d'apprentissage de la normalité par reconstruction, visent à détourner la tâche de reconstruction en une tâche de détection d'anomalies, en se basant sur l'hypothèse forte que, toute anomalie en entrée ne sera pas reconstruite, car en dehors de l'espace de distribution de la normalité apprise. Toutefois, les travaux de Bercea et. al. [86] montrent que les modèles auto-encodeurs sont en proie à une forme de surapprentissage sur la tâche reconstruction, qui résulte en des modèles capables de reconstruire toute entrée, sans distinction de leur distribution de provenance. Cette tendance à l'identité pose un questionnement sur la validité de ces modèles dans leur diagnostic final. En quoi un modèle entraîné sur des photographies du segment antérieur de l'œil est-il cliniquement recevable, s'il est tout aussi capable de reconstruire des radiographies thoraciques, alors qu'il n'en a jamais vu ?

Description de l'expérimentation

Basé sur ce questionnement et les travaux de Bercea et. al. [86], nous proposons une évaluation de la tendance à l'apprentissage de l'identité. Nous considérons un ensemble de bases de données médicales publiques, différentes des photographies du segment antérieur de l'œil. Nous utiliserons l'agglomération de base de données médicale MedMNIST [18], pour sa simplicité d'utilisation, et en particulier, nous nous intéresserons aux sous-bases de données :

- PneumoniaMNIST, radiographies thoraciques pour le diagnostic de la pneumonie.
- RetinaMNIST, photographies du fond d'œil pour le diagnostic de la rétinopathie diabétique.
- BreastMNIST, échographies de la poitrine pour le diagnostic de tissus pathologiques.

Pour chacune d'entre elles sont calculées l'erreur MAE, la distance DSSIM et l'erreur MPRE, entre les images d'entrée classifiées comme "normales" et leur reconstruction, et cela pour tous les modèles entraînés sur les photographies du segment antérieur de l'œil. Ce choix est fait afin de valider notre analyse sous tous les attributs de reconstruction à notre disposition, à savoir au niveau pixel (MAE), au niveau structurel (DSSIM) et au niveau sémantique ou perceptuel (MRPE). L'utilisation de DSSIM (equation 5.3) plutôt que SSIM est fait principalement pour des raisons de lisibilité des résultats. Certains modèles présentent des problèmes de stabilité (SSLAnoVAE, GANomaly), qui peuvent impacter l'analyse des résultats. C'est pourquoi cette analyse se fera sur une seule graine aléatoire, pour laquelle tous les modèles ont convergé à l'entraînement.

Afin de prendre en compte la notion plus fine d'anomalies, nous proposons de réaliser cette même expérience sur la base de données de photographies du segment antérieur de l'œil LightX, auquel on applique des transformations locales à l'instar de ce qui a été fait pour

la création de pseudo-anomalies dans la conception de SiamAAE (section 4.2). Pour ne pas biaiser l'évaluation, l'augmentation appliquée sera une transformation élastique, modifiant la morphologie de la région concernée, augmentation qui n'a jamais été vue à l'entraînement de SiamAAE. L'erreur sera alors calculée entre la reconstruction, et la photographie de référence, avant création de pseudo-anomalies.

Discussion des résultats

Les résultats de l'expérimentation sont rapportés dans la figure 6.6 sous la forme d'histogrammes. Pour chaque histogramme, une couleur est associée à un modèle, dont la tâche de reconstruction est ensuite évaluée sur les bases de données LightX, LightX augmentée, RetinaMNIST, PneumoniaMNIST et BreastMNIST, pour les trois métriques MAE (figure 6.6a), DSSIM (figure 6.6b) et MRPE (figure 6.6c). LightX augmentée désigne ici la base de données LightX pour laquelle on applique une transformation élastique locale aléatoire sur chaque photographie avant l'entrée du modèle. Une visualisation des reconstructions de cette expérimentation est également disponible à la figure 6.8.

Analysons les résultats pour chacun des cas possibles.

- **LightX** : Tout d'abord, les résultats affichés pour la base de données LightX correspondent à ceux déjà observés dans le tableau 6.2. Notre approche SiamAAE, présente de bonnes capacités de reconstruction, en particulier sur les métriques structurelle et perceptuelle, en compétition avec les modèles GANomaly, DPAE et AE SSIM. Cette affirmation est qualitativement vérifiée au travers des visualisations de reconstruction présentées dans la section précédente, à la figure 6.4.
- **LightX Augmentée** : Ensuite vient la base de données LightX augmentée. Puisqu'il s'agit d'images pseudo-anormales, fabriquées à partir des données d'entraînement, on s'attend à ce que l'erreur entre la reconstruction et l'image de référence sans anomalies soit la plus proche possible de l'erreur de référence sur la base de données LightX. Dans l'ensemble, tous les modèles admettent une augmentation de leur erreur de reconstruction, mais seuls les modèles SiamAAE et DAE présentent un accroissement faible comparé à leur valeur de référence. C'est-à-dire que ces deux modèles ont une capacité de correction des anomalies plus puissante que les autres. En effet, pour les reconstructions affichées à la figure 6.8, tous les modèles semblent copier simplement l'entrée sans en modifier le contenu pour "corriger" la pseudo-anomalie. Au contraire, les reconstructions issues de DAE et de SiamAAE montrent une tentative d'effacer et de corriger les pseudo-anomalies en accord avec la notion d'anatomie d'un œil sain, apprise à l'entraînement. Là où notre méthode prend le dessus, c'est dans son erreur

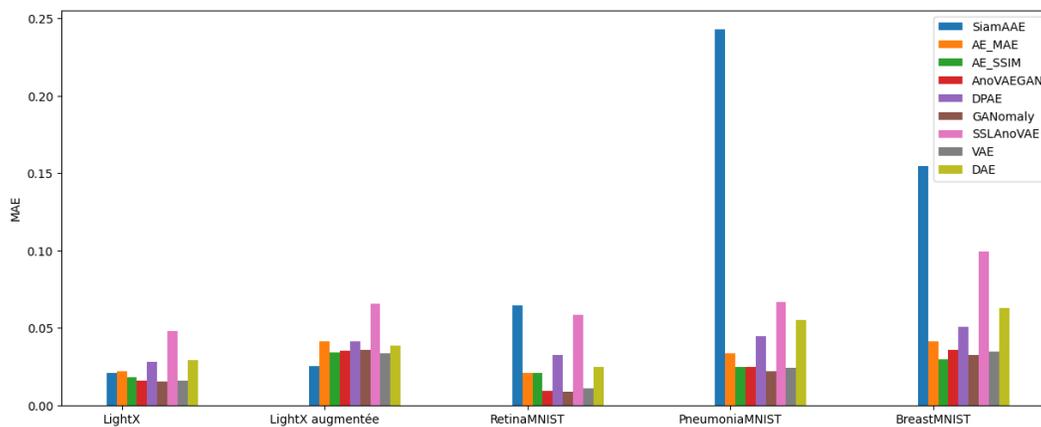
minimale dans la reconstruction, face à tous les autres modèles.

- **RetinaMNIST** : La base de données RetinaMNIST est composée de photographies du fond d’œil. Bien que sémantiquement proches des photographies du fond d’œil, les modèles ont été entraînés sur des photographies du segment antérieur de l’œil. C’est pourquoi on s’attend ici à une augmentation de l’erreur de reconstruction sur toutes les métriques, car les entrées sont techniquement en dehors de la distribution de la normalité apprise, mais cette augmentation devrait être minimale, car elles restent sémantiquement proches. C’est le comportement exact observé pour notre modèle SiamAAE. Cependant, l’ensemble des modèles de la littérature admettent une forte diminution de leur erreur moyenne. Cette diminution peut être expliquée en raison du changement brutal de la distribution d’entrée, mais ne peut être justifiée dans le cadre de la détection d’anomalies. Les modèles SSLAnoVAE, DAE et DPAE semblent sur certaines métriques adopter un comportement similaire. Cependant, l’observation qualitative des reconstructions à la figure 6.8 va à l’encontre de notre intuition. Encore une fois, les modèles de la littérature semblent copier l’entrée sans se soucier de sa distribution de provenance. La reconstruction de notre modèle SiamAAE est quant à elle intéressante, car les rétines reconstruites semblent avoir été blanchies, comme si le fond d’œil a été interprété comme de la sclère.
- **PneumoniaMNIST & BreastMNIST** : Les deux bases de données PneumoniaMNIST et BreastMNIST agissent comme un exemple d’une donnée d’entrée extrêmement anormal, dont la structure, la colorimétrie et la sémantique sont à l’opposé de ce que l’on retrouve dans les photographies du segment antérieur de l’œil. Cette fois-ci, on s’attend à une très grande erreur de reconstruction, et ce au travers des trois métriques. L’erreur de reconstruction de notre approche SiamAAE explose, comparativement aux autres modèles de la littérature, pour qui l’erreur de reconstruction augmente peu voir même stagne. Seules les approches DAE et SSLAnoVAE présentent une forte augmentation de leur erreur comparativement aux valeurs précédentes, mais seulement pour une ou deux des métriques. Visuellement parlant, les reconstructions affichées à la figure 6.8 des modèles de la littérature ressemblent trait pour trait aux images d’entrées. Les sorties de notre approche SiamAAE quant à elles, présentent un masque circulaire tout comme les données d’entraînement, et admettent une couleur rougeâtre, également très présente dans les images d’entraînement. La structure semble néanmoins avoir été conservée, on retrouve l’information du squelette et des tissus dans l’image de sortie, bien qu’elle soit décousue. Enfin, il est intéressant de noter qu’un iris ainsi qu’une pupille semblent apparaître dans la reconstruction de la deuxième image issue de BreastMNIST.

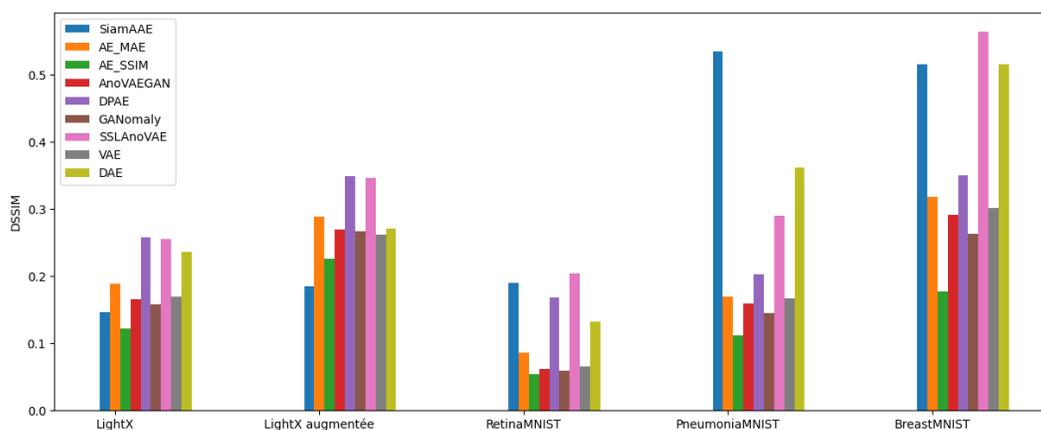
Conclusion de l'expérimentation

En somme, l'apprentissage de la normalité par reconstruction bascule très facilement dans l'apprentissage d'une forme de fonction identité. Tous les modèles de la littérature y font plus ou moins face. Seules les approches SSLAnoVAE et DAE semblent s'en éloigner un peu. Pour le premier, il s'agit sûrement de l'effet de l'apprentissage contrastif secondaire réalisé à l'entraînement. Pour le second, il s'agit des capacités émergentes de la tâche de débruitage, plutôt que d'une simple reconstruction.

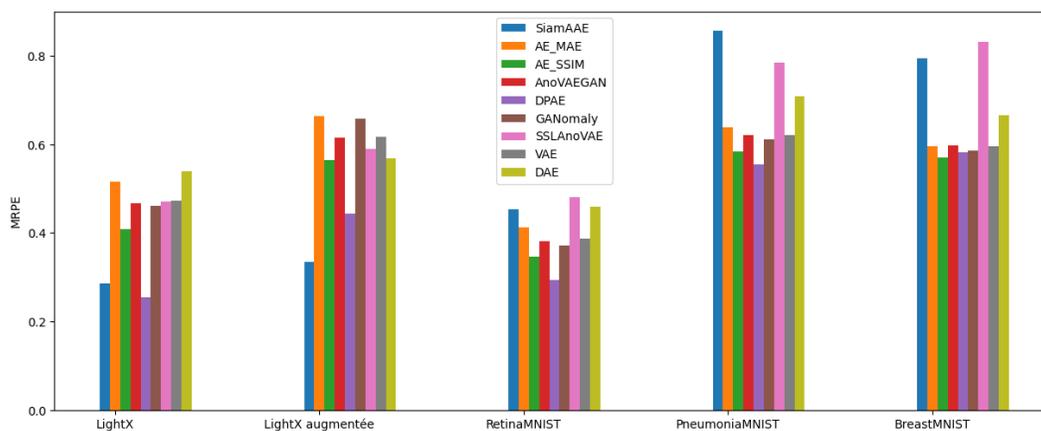
Nos choix d'implémentation pour notre modèle SiamAAE semblent avoir réussi l'objectif de s'éloigner de l'apprentissage d'une fonction identité à l'entraînement, tout en conservant de hautes capacités de reconstruction sur les données d'entraînement. Cela est clairement dû à la combinaison du module de guidage SSL et du module de reconstruction avec pseudo-anomalies, à l'image des performances des modèles SSLAnoVAE et DAE.



(a) MAE



(b) DSSIM



(c) MRPE

FIGURE 6.6 Histogrammes des erreurs de reconstruction pour l'évaluation de la tendance à l'apprentissage de l'identité

Erreurs de reconstruction moyennes (MAE, DSSIM, MRPE) calculées pour les modèles entraînés sur la base de données des photographies du segment antérieur de l'œil, et évalués sur la base de données LightX, LightX augmentée (avec pseudo-anomalies), RetinaMNIST, PneumoniaMNIST et BreastMNIST [18].

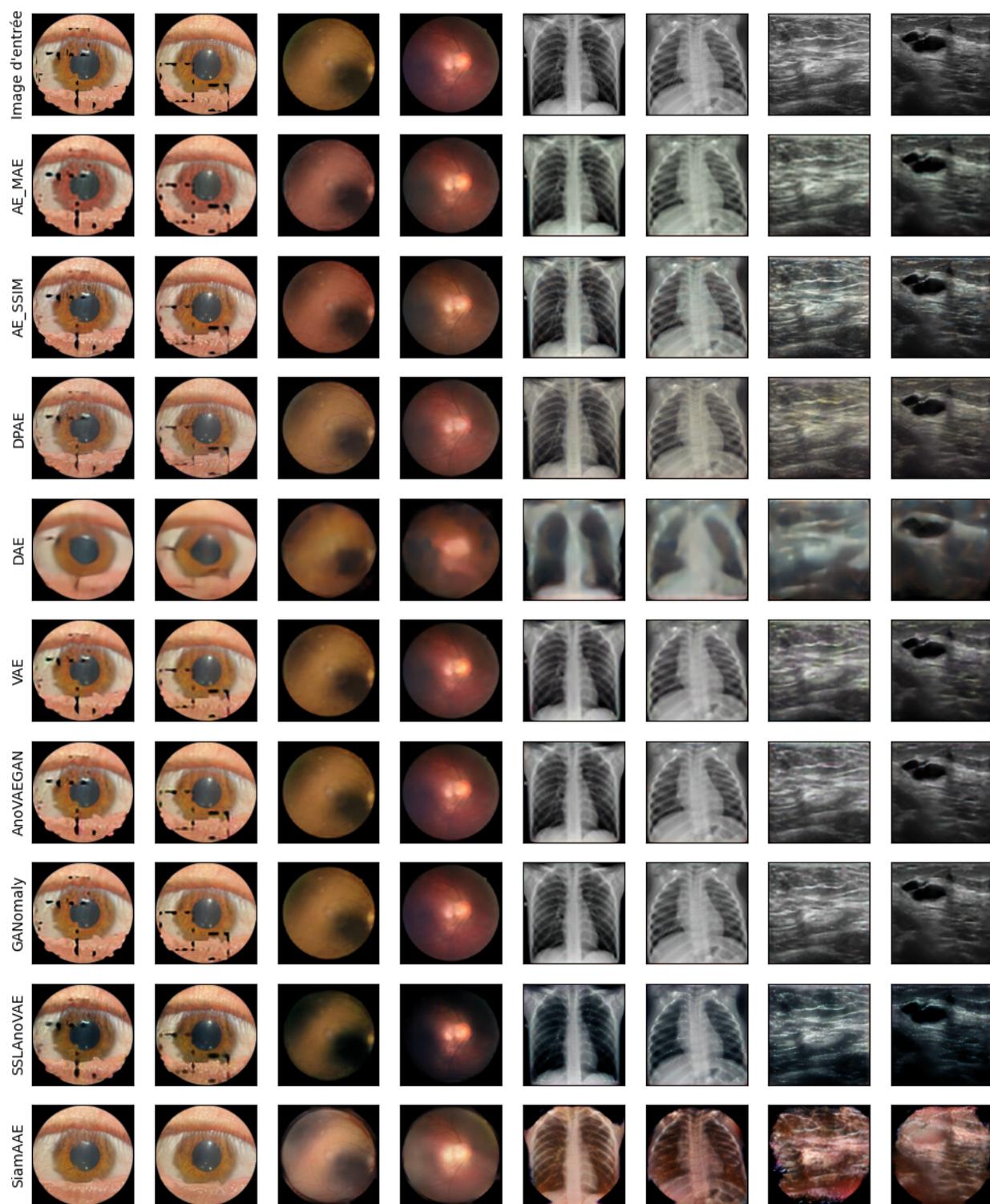


FIGURE 6.8 Exemples de reconstructions pour l'évaluation de la tendance à l'apprentissage de l'identité

Première ligne : Images d'entrées des modèles. Les photographies sont deux à deux et de gauche à droite issues des bases de données LightX (avec pseudo-anomalie élastique), RetinaMNIST, PneumoniaMNIST et BreastMNIST - **Lignes suivantes** : Reconstructions correspondantes à chaque entrée pour le modèle considéré, indiqué à gauche

6.3 Analyses complémentaires

Les expérimentations présentées jusqu’à maintenant ont permis de mettre en avant les avantages de notre approche SiamAAE comparée aux modèles d’apprentissage de la normalité par reconstruction de la littérature, que ce soit en matière de détection d’anomalies ou de reconstruction. Dans cette section, nous présentons deux analyses complémentaires visant à discuter plus en détail de notre approche SiamAAE. En particulier, une première discussion tournera autour du module de guidage SSL, élément central de notre modèle, et de son influence sur l’apprentissage de la normalité. Enfin, la dernière analyse portera sur l’étude des performances qualitatives de détection d’anomalies, au travers d’exemples pathologiques anormaux, et de la visualisation des cartes d’anomalies en comparaison aux masques de segmentation réels annotés par un spécialiste (section 4.1.3).

6.3.1 Influence du module de guidage SSL

Le module de guidage SSL a été conçu pour contraindre l’encodeur à générer un vecteur de représentation qui appartient à l’espace de la normalité apprise lors de l’entraînement, quel que soit le type d’entrée, normale ou anormale. Pour mieux comprendre les mécanismes derrière ce module, nous proposons de modifier la valeur du coefficient λ_{SSL} du terme L_{SSL} dans la fonction objectif définie à l’équation 4.1, et de réentraîner le modèle avec différentes valeurs de λ_{SSL} pour une graine aléatoire fixe. Les valeurs testées sont

$$\lambda_{SSL} \in [0.0, 0.1, 1.0, 2.0, 5.0, 10.0]$$

où $\lambda_{SSL} = 0$ signifie que l’on retire le module SSL de l’entraînement, tout en conservant les autres modules. Puisqu’il s’agit d’une forme de contrainte sur l’espace de représentation, l’analyse portera à la fois sur les performances de détection d’anomalies sur les photographies du segment antérieur de l’œil, disponible à la table 6.4 et sur la visualisation de l’espace latent, réalisée au moyen de la méthode de réduction de dimension UMAP [103], illustrée à la figure 6.9.

Tout d’abord, la lecture des données de la table 6.4 nous permet d’établir les faits suivants.

- L’ajout du module de guidage SSL à l’entraînement pour un coefficient $\lambda_{SSL} > 0$ semble être bénéfique à la tâche de détection d’anomalies. En effet, pour le passage de $\lambda_{SSL} = 0$ à $\lambda_{SSL} = 0.1$, on observe une augmentation de la ROC AUC de 0.5% (de 0.718 à 0.723) et également de la PR AUC de 1.2% (de 0.225 à 0.237). De plus, les meilleures performances sont atteintes pour $\lambda_{SSL} = 5$ et $\lambda_{SSL} = 10.0$, avec respectivement une ROC AUC de 0.726 et 0.728, et une PR AUC de 0.241 et 0.237.

LIGHTX	Détection d'anomalies		Reconstruction (images normales uniquement)			
	ROC AUC \uparrow	PR AUC \uparrow	MAE \downarrow	SSIM \uparrow	MS-SSIM \uparrow	MPRE \downarrow
0	0,718	0,225	<u>1,84E-02</u>	0,72	0,912	2,66E-01
0,1	0,723	<u>0,237</u>	1,82E-02	<u>0,719</u>	<u>0,911</u>	<u>2,67E-01</u>
1	0,718	0,234	1,86E-02	0,717	0,908	2,73E-01
2	0,714	0,226	1,89E-02	0,716	0,909	2,69E-01
5	<u>0,726</u>	0,241	1,88E-02	0,717	0,909	<u>2,67E-01</u>
10	0,728	<u>0,237</u>	2,07E-02	0,707	0,9	2,86E-01

TABLEAU 6.4 Résultats de l'étude du module de guidage SSL sur la tâche de détection d'anomalies et de reconstruction sur les photographies du segment antérieur de l'œil

Évaluation à graine aléatoire fixe, sur l'ensemble de test de la base de données LightX - **En gras** les meilleurs résultats - En souligné le deuxième meilleur résultat

- Au contraire, l'ajout du module de guidage SSL à l'entraînement pour un coefficient $\lambda_{SSL} > 0$ semble détériorer les capacités de reconstruction du modèle. Les meilleures performances sont obtenues en l'absence du module SSL, c'est-à-dire pour un coefficient $\lambda_{SSL} = 0$. Entre $\lambda_{SSL} = 0$ à $\lambda_{SSL} = 10$, on observe une augmentation de 12% de la MAE et de 7% de la MRPE, ainsi qu'une diminution de 1.3% de la SSIM et de 1.2% de la MS-SSIM.

Cette double tendance est une mise en évidence du compromis entre performance de détection d'anomalies et de capacité de reconstruction. Plus le coefficient λ_{SSL} augmente, plus la contrainte sur l'espace de représentation est forte, et plus le modèle concentre son apprentissage sur l'établissement d'attributs communs, entre les images normales et les pseudo-anomalies. Il paraît donc normal que les capacités de reconstruction soient détériorées, car moins priorisées. Les visualisations UMAP en deux dimensions de l'espace de représentation pour les différents coefficients corroborent cette analyse (Figure 6.9). Sans module de guidage SSL, l'espace de représentation se compose de 5 groupements compacts, un pour chaque position de l'œil présent dans la base de données. Il s'agit d'un critère visuellement très discriminant, il fait donc sens que les données soient réparties de telle sorte. Avec l'augmentation du coefficient, on observe un étalement et un rapprochement des groupements composé de chaque position de l'œil. L'information de haut niveau reste présente, mais le modèle semble faire meilleur usage de son espace de représentation. En apprenant à rapprocher deux vues, normale et pseudo-anormale, dans l'espace de représentation, le modèle semble donc favoriser l'apprentissage d'attributs qui rapprochent deux données issues de deux groupements aupa-

ravant très distincts, potentiellement des attributs anatomiques qui favorisent la détection d'anomalies, comme laissent à penser les résultats quantitatifs.

L'ensemble des résultats présentés dans les sections précédentes et celles qui suivent ont été réalisés pour un coefficient $\lambda_{SSL} = 10$, car nous considérons que la perte en qualité de reconstruction affichée n'est pas significative comparée au gain en matière de détection d'anomalies.

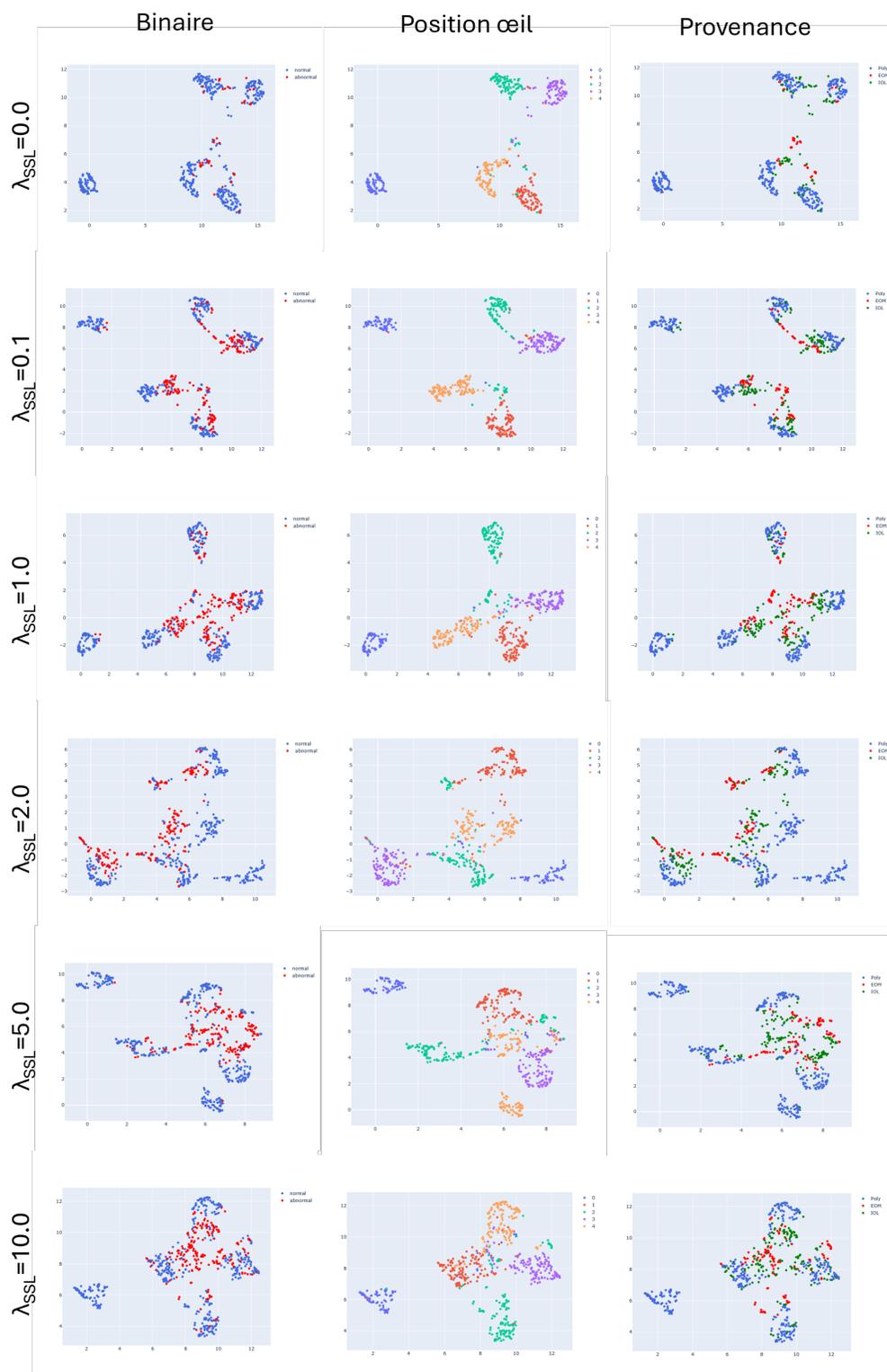


FIGURE 6.9 Visualisations UMAP de l'espace de représentation pour différents coefficients λ_{SSL}

Première colonne : Coloration binaire des échantillons, normaux (bleu) et anormaux (rouge) -
Deuxième colonne : Coloration des échantillons selon la position de l'œil, milieu (bleu), haut (rouge), bas (vert), droite (violet) et gauche (orange) -
Troisième colonne : Coloration des échantillons selon leur provenance, POLY (bleu), EOM (rouge) et IOL (vert)

6.3.2 Analyse qualitative des cartes d'anomalies

Grâce à l'expérimentation de robustesse face à l'apprentissage de l'identité faite en section 6.2.2, nous avons pu mettre en évidence la lacune des approches par reconstruction à apprendre une forme de fonction identité, de laquelle notre approche SiamAAE s'éloigne. Néanmoins, les cas étudiés (pseudo-anomalies, RetinaMNIST, PneumoniaMNIST et BreastMNIST) sont des cas d'anomalies extrêmes, qui ne représentent pas la réelle distribution des pathologies du segment antérieur de l'œil. À partir du score d'anomalies de notre approche (équation 4.6), on calcule les histogrammes de densité de distribution de chacune des classes "normales" et "anormales", illustrés à la figure 6.10.

Idéalement, pour une détection d'anomalies optimale, les deux distributions doivent être séparables, de sorte qu'un simple seuillage sur la valeur du score fournisse la prédiction attendue. En réalité, ce n'est jamais totalement séparable. Dans notre cas, que ce soit sur l'ensemble de validation ou celui de test, on observe un décalage entre les centres de masse des deux distributions, mais avec tout de même une superposition des deux distributions. Les raisons possibles derrière ce chevauchement sont les suivantes :

1. La première raison pourrait être que notre approche n'a en réalité pas appris la notion de normalité des photographies du segment antérieur de l'œil. On se retrouve dans une situation où les anomalies sont entièrement reconstruites, et où le score d'anomalies reflète simplement la perte d'information entre l'encodeur et le décodeur, sans notion anatomique pertinente.
2. Une autre raison possible concerne les anomalies elles-mêmes. Il se peut que ces dernières soient tellement subtiles, qu'en réalité l'erreur liée à leur reconstruction est très faible devant l'erreur de reconstruction moyenne due à la perte d'information mentionnée juste avant. Au contraire, si le score d'anomalies est élevé, mais concerne une région très petite de l'image, alors la moyenne du score pourrait ne pas prendre en compte cette erreur dans la prédiction.
3. Enfin, une dernière raison pourrait être l'apprentissage malencontreux de biais de la base de données d'entraînement. Si cette dernière présente des biais très importants, ils peuvent influencer la notion même de normalité apprise. Ainsi, les anomalies détectées pourraient être en rapport avec ces biais, et non des facteurs cliniquement significatifs.

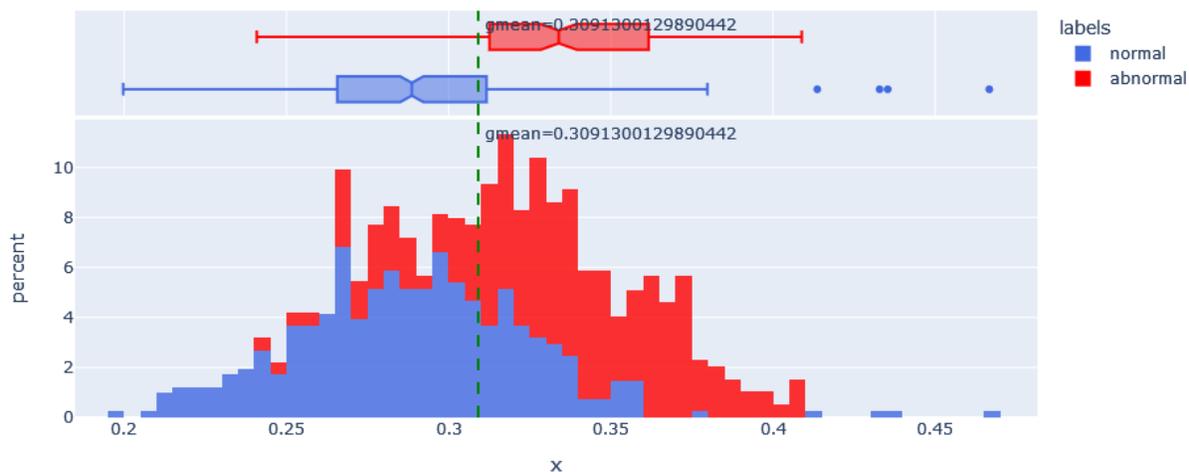
Dans tous les cas, il nous faut analyser les carte d'erreur, ou cartes résiduelles, entre les reconstructions et les photographies du segment antérieur de l'œil contenant des anomalies, pour en connaître les tenants et aboutissants. À la figure 6.12, est présenté un ensemble de photographies du segment antérieur de l'œil, leur reconstruction et leurs cartes résiduelles, calculées à partir des métriques usuelles MAE, SSIM, MS-SSIM et MRPE, ainsi que les

masques de segmentation réalisés par un spécialiste lors de la phase d’annotation (section 4.1.3). En réalité, au travers des reconstructions et cartes résiduelles des échantillons anormaux de la figure 6.12, on observe les trois hypothèses mentionnées.

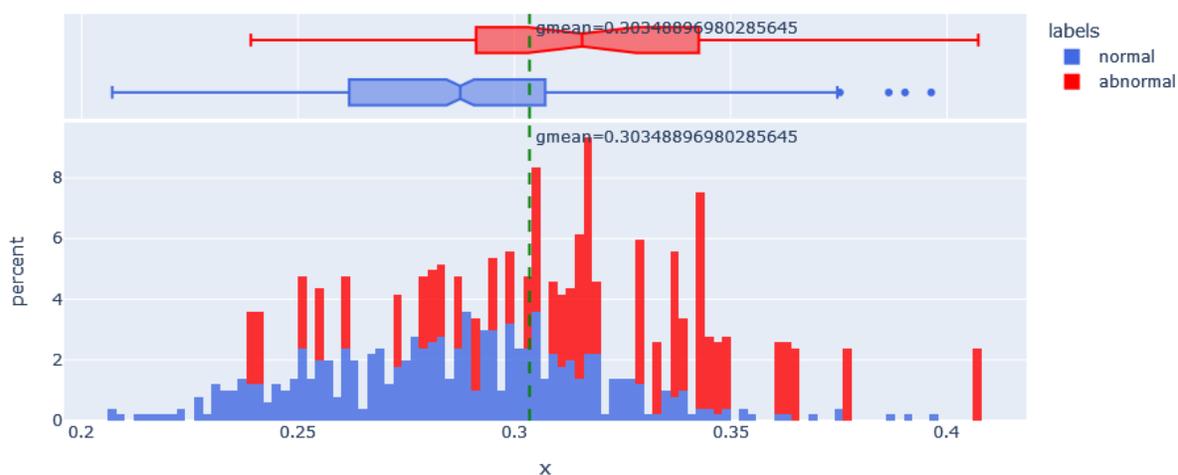
- Notre approche SiamAAE est plus robuste à l’apprentissage de l’identité. Dans certains cas, le travail de reconstruction semble correspondre aux attentes cliniques, comme c’est le cas pour la photographie 1. Les deux pathologies de la sclère semblent avoir été effacées à la reconstruction, et ressortent dans les cartes résiduelles, dans les mêmes régions que celles du masque de segmentation. Le contraire est malheureusement tout aussi possible. Les photographies 3 et 5 présentent au niveau de la sclère des hémorragies, de couleur caractéristique rouge vif. On observe effectivement un blanchissement léger de ces dernières à la reconstruction, mais elles restent tout de même très visibles, en particulier pour la photographie 4. Néanmoins, dans d’autres cas, comme la photographie 6, l’hémorragie présente a été affinée, tels des vaisseaux sanguins, sûrement pour faire sens avec la notion de vascularisation apprise à l’entraînement. Enfin, la cataracte présente dans la photographie 2, caractérisée par un voile blanchâtre dans la pupille, a induit en erreur notre modèle, qui semble avoir confondu la pupille avec l’iris elle-même. Cette transformation ne fait pas de sens cliniquement, car il aurait fallu que la pupille soit assombrie par le modèle, mais engendre par chance une erreur dans la bonne région anatomique.
- Dans beaucoup de cas, les erreurs de reconstruction locales semblent noyées dans l’erreur de reconstruction moyenne, causée par de nombreuses structures fines dans les photographies. Par exemple, les photographies 5 et 6 englobent une bonne partie du segment antérieur de l’œil, y compris les paupières. La peau ainsi que les cils engendrent de fortes erreurs localisées, qui accumulées, noient complètement l’information des cartes résiduelles et la potentielle lecture de biomarqueurs pathologiques. De plus, il apparaît que beaucoup d’erreurs retranscrites dans les cartes résiduelles sont liées aux cils, à l’iris et aux illuminations. Les deux premiers en raison de leurs structures fines irrégulières, et la dernière en raison du fort écart de valeur entre les reflets et les régions anatomiques de l’œil. La réduction de la dimension des images en entrée est un autre facteur qui joue sur la résolution des biomarqueurs pathologiques. Bien qu’ils soient toujours visibles, la perte d’information qui en découle rend nécessairement plus difficile de les distinguer pour nos modèles.
- Un des biais de notre base de données, présenté à la première section de ce chapitre (section 6.1) est l’utilisation de goutte de fluorescéine lors de l’acquisition des données IOL et EOM, et leur absence dans la base de données POLY. De fait, la surreprésentation des données de POLY à l’entraînement résulte en un modèle qui efface toute

apparition de la fluorescéine dans les photographies du segment antérieur. La fluorescéine, en tant que telle, ne représente pas une anomalie. Elle peut indiquer des régions anormales dans certaines conditions. Par exemple, dans la photographie 5, la fluorescéine, caractéristique par sa couleur jaunâtre, présente dans la sclère ne présente aucune anomalie, mais engendre une forte erreur de reconstruction dans cette région. De même pour les exemples 7 et 8. En particulier, la photographie 8 admet une forte accumulation de fluorescéine à l'endroit de l'anomalie segmentée par le spécialiste. La question alors se pose, l'erreur de reconstruction affichée est-elle liée à la fluorescéine, ou bien au biomarqueur segmenté. Dans cet exemple, il s'agit d'une combinaison des deux erreurs au vu de la reconstruction, mais rien ne garantit que cela soit toujours le cas.

En résumé, l'évaluation de la tâche d'anomalies dans les photographies du segment antérieur de l'œil s'avère être une tâche beaucoup plus subtile qu'attendu. Même si les métriques peuvent s'avérer positives, rien ne garantit que le diagnostic final se fonde sur des facteurs cliniques pertinents, et il est difficile de s'en assurer pour un entraînement non supervisé, qui ne contient aucune autre information que la notion de normalité, difficile à estimer même pour un spécialiste, dans un organe aussi complexe que l'œil. De plus, les cartes résiduelles en général utilisées offrent toute une information différente, et se focalisent sur des types d'erreurs différentes, dépendamment de leur implémentation. Il devient donc difficile de justifier le choix de l'une d'entre elles plutôt qu'une autre, et à fortiori, un questionnement sur le score d'anomalies se pose, puisque lui-même est construit à partir d'une moyenne de ces cartes résiduelles.



(a) Ensemble de validation



(b) Ensemble de test

FIGURE 6.10 Histogrammes de densité de probabilité du score d'anomalies SiamAAE sur les photographies du segment antérieur de l'œil

(a) Histogramme calculé sur l'ensemble de validation - (b) Histogramme calculé sur l'ensemble de test - En bleu : distribution du score d'anomalies des échantillons normaux - En rouge : distribution du score d'anomalies échantillons anormaux - En vert : valeur de seuil idéal, calculé selon la moyenne quadratique $gmean = \sqrt{Sensibilité \times (1 - Spécificité)}$

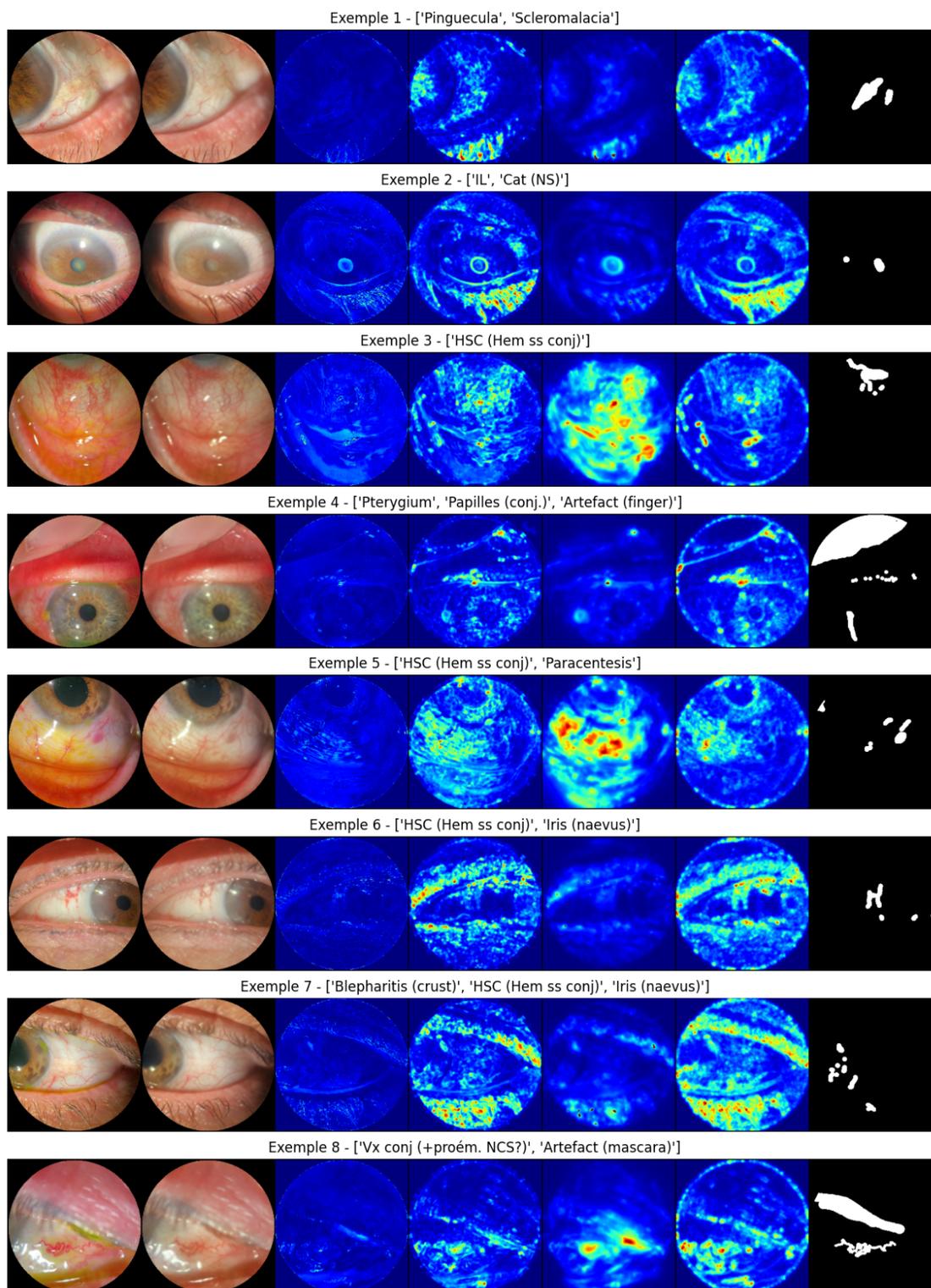


FIGURE 6.12 Exemples de reconstructions et de cartes d'erreur pour notre approche SiamAAE, sur des photographies anormales du segment antérieur de l'œil

De gauche à droite : Images d'entrées - Reconstructions SiamAAE - Cartes d'erreur MAE - Cartes d'erreur SSIM - Cartes d'erreur MS-SSIM - Cartes d'erreur MRPE - Cartes de segmentation des anomalies annotées par un spécialiste (section 4.1.3) dont les annotations sont données en titre de chaque exemple

CHAPITRE 7 CONCLUSION

Ce dernier chapitre conclut notre étude en récapitulant les contributions et résultats obtenus, tout en ouvrant la réflexion sur les perspectives d'amélioration. La première section propose une synthèse des travaux réalisés, mettant en lumière les avancées apportées par notre modèle SiamAAE comparé aux modèles étudiés de la littérature. Ensuite, les limitations identifiées dans notre solution sont discutées, qu'elles soient liées à l'approche méthodologique, aux données utilisées, ou aux performances obtenues. Enfin, des pistes d'améliorations futures sont proposées, visant à répondre aux dites limitations.

7.1 Synthèse des travaux

Au cours de ce projet, nous avons :

1. Annoté et construit une base de données dédiée à la détection d'anomalies des photographies du segment antérieur de l'œil. Cette base de données résulte d'acquisitions cliniques réalisées par notre partenaire LightX Innovations (IOL et EOM) et d'une collecte de données "normales" (POLY), issues d'une population démographiquement diverse sans antécédents pathologiques graves, pour l'entraînement de modèles d'apprentissage profond non supervisé.
2. Analysé et comparé des méthodes d'apprentissage de la normalité par reconstruction issues de la littérature, appliquées aux photographies du segment antérieur de l'œil.
3. Proposé, développé et étudié un nouveau modèle, SiamAAE, combinant les approches par reconstruction et auto-distillation, aboutissant à une meilleure robustesse contre l'apprentissage de l'identité tout en conservant, voire améliorant, les performances de détection d'anomalies et de reconstruction par rapport aux modèles existants.

Ces travaux ont permis d'explorer l'application des méthodes de détection d'anomalies par reconstruction sur des photographies du segment antérieur, un domaine jusqu'alors peu étudié par la communauté scientifique (présenté en section 6.2). Nos expérimentations ont ainsi mis en évidence la tendance indésirable des modèles existants à apprendre l'identité, traduisant un surapprentissage de la reconstruction pour toute donnée d'entrée (expérimentation présentée en section 6.2.2). Notre approche SiamAAE a démontré sa capacité à surmonter cette lacune, offrant une meilleure robustesse face à l'apprentissage de l'identité tout en conservant une architecture classique d'auto-encodeur pour l'inférence.

Le modèle développé répond ainsi aux attentes techniques de notre partenaire, telles qu'énon-

cées au chapitre 3, en fournissant une détection d’anomalies à l’échelle globale (image) et locale (pixel), avec une taille modérée (environ 2,3 millions de paramètres) et une vitesse d’inférence rapide (moins d’une seconde par image).

7.2 Limitations de la solution proposée

Il est important de reconnaître qu’il s’agit d’un projet ambitieux, aux limitations certaines.

1. **Déséquilibre et biais de la base de données** : La première limitation réside dans la base de données elle-même. Quelle que soit la modalité d’entraînement (supervisée ou non supervisée), les annotations restent essentielles pour l’évaluation. Bien que soigneusement construite, notre base de données souffre d’un déséquilibre marqué par un manque de données anormales. Par ailleurs, les sous-bases IOL et EOM incluent un protocole d’acquisition avec instillation de gouttes de fluorescéine, absente lors de la collecte de donnée de la base de normalité. Les modèles, entraînés sur un ensemble sur-représenté par ces images, ont tendance à considérer toute apparition de la fluorescéine comme une anomalie, ce qui n’est cliniquement pas le cas.
2. **Complexité et variabilité des données d’entraînement** : Les photographies du segment antérieur de l’œil présentent une grande diversité anatomique et des attributs uniques (notamment biométriques, tels que l’iris). Bien que notre modèle SiamAAE vise à apprendre la normalité sans tomber dans l’apprentissage de l’identité, l’analyse qualitative montre que la détection d’anomalies ne se généralise pas de manière homogène à tous les biomarqueurs. La haute variabilité des données d’entrée pourrait expliquer les difficultés d’apprentissage de la normalité.
3. **Disparité des cartes résiduelles** : L’exploitation brute des cartes résiduelles pour l’estimation spatiale des anomalies ne semble pas adaptée à notre contexte. Les structures fines (cils, iris, vaisseaux sanguins) et les reflets cornéens génèrent des erreurs de reconstruction importantes, qui ne reflètent pas nécessairement des anomalies cliniques. L’information pertinente est alors souvent noyée dans l’erreur moyenne de reconstruction. De plus, les informations contenues dans ces cartes varient selon la métrique utilisée, rendant difficile la sélection d’une métrique optimale pour extraire les anomalies pertinentes.

7.3 Améliorations futures

Ce projet est une première exploration des méthodes de détection d’anomalies non supervisées appliquées aux photographies du segment antérieur de l’œil. De nombreuses pistes

d'améliorations peuvent être envisagées, parmi lesquelles :

- Minimiser l'impact des biais dans les données d'entraînement, en collectant davantage de données, notamment pour des cas normaux incluant l'instillation de fluorescéine.
- Assurer la généralisabilité des modèles sur des données acquises sous d'autres formats, telles que la base de données EyePACS [12].
- Réduire la variabilité inter et intra-sujet, tant au niveau colorimétrique que géométrique, afin de standardiser les images d'entrée et faciliter l'apprentissage de la normalité.
- Exploiter la résolution totale des images en adoptant une approche par patch, mettant ainsi l'accent sur des caractéristiques anatomiques spécifiques du segment antérieur de l'œil, plutôt que d'apprendre la normalité sur une sémantique globale de l'œil.
- Explorer de nouvelles approches pour l'estimation des erreurs de reconstruction, combinant les informations pixels à pixels, structurelles et perceptuelles dans une seule carte sans en altérer la signification.

Il s'agit ici des pistes qui nous semblent les plus pertinentes à explorer suite à nos travaux, mais seule une analyse rigoureuse pourra déterminer leur efficacité dans le cadre d'une future recherche.

RÉFÉRENCES

- [1] “Orbits and eyes Illustrations : normal anatomy| e-Anatomy.” [En ligne]. Disponible : <https://www.imaios.com/en/e-anatomy/head-and-neck/eye>
- [2] K. D. Gordon, “The Cost of Vision Loss and Blindness in Canada,” The Canadian Council of the Blind, Summary Report, mai 2021. [En ligne]. Disponible : <https://www.fightingblindness.ca/wp-content/uploads/2021/05/KG-EN-ACC-Cost-of-Vision-loss-and-Blindness-in-Canada-Final.pdf>
- [3] —, “The Impact of the Covid-19 Pandemic on Vision Health in Canada 2021,” The Canadian Council of the Blind and Fighting Blindness Canada, Rapport technique, oct. 2022. [En ligne]. Disponible : <https://ccbnational.net/shaggy/wp-content/uploads/2023/07/15-Report-Card-on-Vision-Health-in-Canada-2021-Part-1-English-Oct-14-2022.pdf>
- [4] M. Mohammadpour, Z. Heidari, M. Mirghorbani et H. Hashemi, “Smartphones, tele-ophthalmology, and VISION 2020,” *International Journal of Ophthalmology*, déc. 2017. [En ligne]. Disponible : http://www.ijo.cn/gjyken/ch/reader/view_abstract.aspx?file_no=20171219&flag=1
- [5] H. R. Wright et J. P. Diamond, “Service innovation in glaucoma management : using a web-based electronic patient record to facilitate virtual specialist supervision of a shared care glaucoma programme,” *British Journal of Ophthalmology*, vol. 99, n^o. 3, p. 313–317, mars 2015. [En ligne]. Disponible : <https://bjo.bmj.com/lookup/doi/10.1136/bjophthalmol-2014-305588>
- [6] D. Kiage, I. Kherani, S. Gichuhi, K. Damji et M. Nyenze, “The muranga teleophthalmology study : Comparison of virtual (teleglaucoma) with in-person clinical assessment to diagnose glaucoma,” *Middle East African Journal of Ophthalmology*, vol. 20, n^o. 2, p. 150, 2013. [En ligne]. Disponible : <https://journals.lww.com/10.4103/0974-9233.110604>
- [7] B. Babenko, A. Mitani, I. Traynis, N. Kitade, P. Singh, A. Y. Maa, J. Cuadros, G. S. Corrado, L. Peng, D. R. Webster, A. Varadarajan, N. Hammel et Y. Liu, “Detection of signs of disease in external photographs of the eyes via deep learning,” *Nature Biomedical Engineering*, vol. 6, n^o. 12, p. 1370–1383, mars 2022. [En ligne]. Disponible : <https://www.nature.com/articles/s41551-022-00867-5>
- [8] L. M. Khazaeni, *Examen du patient en ophtalmologie*, Le Manuel Merck, 2019.
- [9] M. Dobes et L. Machala, “Iris database,” <http://phoenix.inf.upol.cz/iris/>, UPOL, 2004.

- [10] C. A. of Sciences, “Casia iris image database,” <http://biometrics.idealtest.org/>.
- [11] H. Proença et L. Alexandre, “UBIRIS : A noisy iris image database,” dans *13th International Conference on Image Analysis and Processing - ICIAP 2005*, vol. LNCS 3617. Cagliari, Italy : Springer, September 2005, p. 970–977.
- [12] J. Cuadros et G. Bresnick, “Eyepacs : An adaptable telemedicine system for diabetic retinopathy screening,” *Journal of Diabetes Science and Technology*, vol. 3, n^o. 3, p. 509–516, mai 2009. [En ligne]. Disponible : <http://dx.doi.org/10.1177/193229680900300315>
- [13] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, L. Giancardo, G. Quellec et F. Mériaudeau, “Retinal fundus multi-disease image dataset (rfmid),” 2020. [En ligne]. Disponible : <https://dx.doi.org/10.21227/s3g7-st65>
- [14] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu et H. Kang, “Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening,” *Information Sciences*, vol. 501, p. 511 – 522, 2019. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0020025519305377>
- [15] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton et J.-C. Klein, “Feedback on a publicly distributed image database : The messidor database,” *Image Analysis & Stereology*, vol. 33, n^o. 3, p. 231, août 2014. [En ligne]. Disponible : <http://dx.doi.org/10.5566/ias.1155>
- [16] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe et F. Meriaudeau, “Indian diabetic retinopathy image dataset (idrid),” 2018. [En ligne]. Disponible : <https://dx.doi.org/10.21227/H25W98>
- [17] Y. Zhou, B. Wang, L. Huang, S. Cui et L. Shao, “A benchmark for studying diabetic retinopathy : Segmentation, grading, and transferability,” *IEEE Transactions on Medical Imaging*, vol. 40, n^o. 3, p. 818–828, 2021.
- [18] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister et B. Ni, “Medmnist v2- a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific Data*, vol. 10, n^o. 1, p. 41, 2023.
- [19] Q. Wei, X. Li, W. Yu, X. Zhang, Y. Zhang, B. Hu, B. Mo, D. Gong, N. Chen, D. Ding et Y. Chen, “Learn to segment retinal lesions and beyond,” dans *International Conference on Pattern Recognition (ICPR)*, 2020.
- [20] P. Gholami et V. Lakshminarayanan, “Optical coherence tomography image retinal database,” 2019. [En ligne]. Disponible : <https://www.openicpsr.org/openicpsr/project/108503/version/V1/view>

- [21] L. M. Khazaeni, “Cataracte,” 2024/01 2024. [En ligne]. Disponible : <https://www.merckmanuals.com/fr-ca/professional/troubles-oculaires/ataracte/ataracte>
- [22] D. J. Rhee, “Revue générale des glaucomes,” 2023/04 2023. [En ligne]. Disponible : <https://www.merckmanuals.com/fr-ca/professional/troubles-oculaires/glaucome/revue-gnrale-des-glaucomes>
- [23] V. Y. Bunya, “Introduction aux troubles cornéens,” 2024/07 2024. [En ligne]. Disponible : <https://www.merckmanuals.com/fr-ca/professional/troubles-oculaires/troubles-cornens/introduction-aux-troubles-cornens>
- [24] Z. A. Syed, “Revue générale des troubles conjonctivaux et sclé-
raux,” 2023/04 2023. [En ligne]. Disponible : <https://www.merckmanuals.com/fr-ca/professional/troubles-oculaires/troubles-conjonctivaux-et-sclraux/revue-gnrale-des-troubles-conjonctivaux-et-sclraux>
- [25] —, “Épisclérite,” 2023/04 2023. [En ligne]. Disponible : <https://www.merckmanuals.com/fr-ca/professional/troubles-oculaires/troubles-conjonctivaux-et-sclraux/lpisclrite>
- [26] M. Astrid, M. Zaheer, J.-Y. Lee et S.-I. Lee, “Learning not to reconstruct anomalies,” dans *British Machine Vision Conference*, 2021. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:239024474>
- [27] S. Akcay, A. Atapour-Abarghouei et T. P. Breckon, “GANomaly : Semi-Supervised Anomaly Detection via Adversarial Training,” nov. 2018, arXiv :1805.06725 [cs]. [En ligne]. Disponible : <http://arxiv.org/abs/1805.06725>
- [28] C.-L. Li, K. Sohn, J. Yoon et T. Pfister, “Cutpaste : Self-supervised learning for anomaly detection and localization,” dans *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, juin 2021. [En ligne]. Disponible : <http://dx.doi.org/10.1109/CVPR46437.2021.00954>
- [29] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger et C. Steger, “Improving unsupervised defect segmentation by applying structural similarity to autoencoders,” dans *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2019. [En ligne]. Disponible : <http://dx.doi.org/10.5220/0007364503720380>
- [30] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz et D. V. Dylov, “Anomaly detection in medical imaging with deep perceptual autoencoders,” *IEEE Access*, vol. 9, p. 118571–118583, 2021. [En ligne]. Disponible : <http://dx.doi.org/10.1109/ACCESS.2021.3107163>

- [31] C. Baur, S. Denner, B. Wiestler, N. Navab et S. Albarqouni, “Autoencoders for unsupervised anomaly segmentation in brain mr images : A comparative study,” *Medical Image Analysis*, vol. 69, p. 101952, 2021. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1361841520303169>
- [32] M. J. Swain et D. H. Ballard, “Swain1.pdf,” *Readings in Multimedia Computing Networking*, p. 265–277, 2002. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/B9781558606517501091?via%3Dihub>
- [33] N. Dalal et B. Triggs, “Histograms of oriented gradients for human detection,” dans *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1, juin 2005, p. 886–893 vol. 1, iSSN : 1063-6919. [En ligne]. Disponible : <https://ieeexplore.ieee.org/document/1467360>
- [34] T. Ojala, M. Pietikainen et D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, p. 582–585, 1994, conference Name : 12th International Conference on Pattern Recognition ISBN : 9780818662652 Place : Jerusalem, Israel Publisher : IEEE Comput. Soc. Press. [En ligne]. Disponible : <http://ieeexplore.ieee.org/document/576366/>
- [35] X. Wang, T. X. Han et S. Yan, “An hog-lbp human detector with partial occlusion handling,” dans *2009 IEEE 12th International Conference on Computer Vision*, 2009, p. 32–39.
- [36] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, n^o. 2, p. 91–110, nov. 2004. [En ligne]. Disponible : <http://link.springer.com/10.1023/B:VISI.0000029664.99615.94>
- [37] D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, M. Calonder, V. Lepetit, C. Strecha et P. Fua, “BRIEF : Binary Robust Independent Elementary Features,” dans *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos et N. Paragios, édit. Berlin, Heidelberg : Springer Berlin Heidelberg, 2010, vol. 6314, p. 778–792, series Title : Lecture Notes in Computer Science. [En ligne]. Disponible : http://link.springer.com/10.1007/978-3-642-15561-1_56
- [38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman et O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” dans *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA : IEEE Computer Society, jun 2018, p. 586–595. [En ligne]. Disponible : <https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00068>

- [39] J. Johnson, A. Alahi et L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” dans *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe et M. Welling, édit. Cham : Springer International Publishing, 2016, p. 694–711.
- [40] N. Tuluptceva, B. Bakker, I. Fedulova et A. Konushin, “Perceptual image anomaly detection,” dans *Pattern Recognition*, S. Palaiiahnakote, G. Sanniti di Baja, L. Wang et W. Q. Yan, édit. Cham : Springer International Publishing, 2020, p. 164–178.
- [41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li et L. Fei-Fei, “Imagenet : A large-scale hierarchical image database,” dans *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, p. 248–255.
- [42] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally et K. Keutzer, “Squeezenet : Alexnet-level accuracy with 50x fewer parameters and <1mb model size,” *ArXiv*, vol. abs/1602.07360, 2016. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:14136028>
- [43] A. Krizhevsky, I. Sutskever et G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” dans *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou et K. Weinberger, édit., vol. 25. Curran Associates, Inc., 2012. [En ligne]. Disponible : https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [44] K. Simonyan et A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” dans *International Conference on Learning Representations*, 2015.
- [45] K. He, X. Zhang, S. Ren et J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770–778, 2015. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:206594692>
- [46] T. Cover et P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, n^o. 1, p. 21–27, 1967.
- [47] X. Jin et J. Han, *K-Means Clustering*. Boston, MA : Springer US, 2010, p. 563–564. [En ligne]. Disponible : https://doi.org/10.1007/978-0-387-30164-8_425
- [48] D. Reynolds, *Gaussian Mixture Models*. Boston, MA : Springer US, 2009, p. 659–663. [En ligne]. Disponible : https://doi.org/10.1007/978-0-387-73003-5_196
- [49] C. Cortes et V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, n^o. 3, p. 273–297, 1995.
- [50] M. K. S.V. et G. R., “Computer-aided diagnosis of anterior segment eye abnormalities using visible wavelength image analysis based machine learning,” *Journal of Medical Systems*, vol. 42, n^o. 7, juin 2018. [En ligne]. Disponible : <http://dx.doi.org/10.1007/s10916-018-0980-z>

- [51] F. N. Jais, M. Z. Che Azemin, M. R. Hilmi, M. I. Mohd Tamrin et K. M. Kamal, "Postsurgery classification of best-corrected visual acuity changes based on pterygium characteristics using the machine learning technique," *The Scientific World Journal*, vol. 2021, p. 1–7, nov. 2021. [En ligne]. Disponible : <http://dx.doi.org/10.1155/2021/6211006>
- [52] L. B. da Cruz, J. C. Souza, A. C. de Paiva, J. D. S. de Almeida, G. B. Junior, K. R. T. Aires, A. C. Silva et M. Gattass, "Tear film classification in interferometry eye images using phylogenetic diversity indexes and ripley's k function," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, n^o. 12, p. 3491–3498, 2020.
- [53] Z. Xu, J. Xu, C. Shi, W. Xu, X. Jin, W. Han, K. Jin, A. Grzybowski et K. Yao, "Artificial Intelligence for Anterior Segment Diseases : A Review of Potential Developments and Clinical Applications," *Ophthalmology and Therapy*, vol. 12, n^o. 3, p. 1439–1455, juin 2023. [En ligne]. Disponible : <https://link.springer.com/10.1007/s40123-023-00690-4>
- [54] X. Fang, M. Deshmukh, M. L. Chee, Z.-D. Soh, Z. L. Teo, S. Thakur, J. H. L. Goh, Y.-C. Liu, R. Husain, J. S. Mehta, T. Y. Wong, C.-Y. Cheng, T. H. Rim et Y.-C. Tham, "Deep learning algorithms for automatic detection of pterygium using anterior segment photographs from slit-lamp and hand-held cameras," *British Journal of Ophthalmology*, vol. 106, n^o. 12, p. 1642–1647, déc. 2022, publisher : BMJ Publishing Group Ltd Section : Clinical science. [En ligne]. Disponible : <https://bjo.bmj.com/content/106/12/1642>
- [55] S. Haykin, *Neural networks : a comprehensive foundation*. Prentice Hall PTR, 1994.
- [56] M.-T. Kuo, B. W.-Y. Hsu, Y.-S. Lin, P.-C. Fang, H.-J. Yu, A. Chen, M.-S. Yu et V. S. Tseng, "Comparisons of deep learning algorithms for diagnosing bacterial keratitis via external eye photographs," *Scientific Reports*, vol. 11, n^o. 1, p. 24227, déc. 2021, publisher : Nature Publishing Group. [En ligne]. Disponible : <https://www.nature.com/articles/s41598-021-03572-6>
- [57] G. Huang, Z. Liu et K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2261–2269, 2016. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:9433631>
- [58] M. Tan et Q. Le, "EfficientNet : Rethinking model scaling for convolutional neural networks," dans *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri et R. Salakhutdinov, édit., vol. 97. PMLR, 09–15 Jun 2019, p. 6105–6114. [En ligne]. Disponible : <https://proceedings.mlr.press/v97/tan19a.html>

- [59] Z. Li, J. Jiang, K. Chen, Q. Chen, Q. Zheng, X. Liu, H. Weng, S. Wu et W. Chen, “Preventing corneal blindness caused by keratitis using artificial intelligence,” *Nature Communications*, vol. 12, n^o. 1, p. 3738, juin 2021, publisher : Nature Publishing Group. [En ligne]. Disponible : <https://www.nature.com/articles/s41467-021-24116-6>
- [60] H. Gu, Y. Guo, L. Gu, A. Wei, S. Xie, Z. Ye, J. Xu, X. Zhou, Y. Lu, X. Liu et J. Hong, “Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs,” *Scientific Reports*, vol. 10, n^o. 1, p. 17851, oct. 2020, publisher : Nature Publishing Group. [En ligne]. Disponible : <https://www.nature.com/articles/s41598-020-75027-3>
- [61] W. Li, Y. Yang, K. Zhang, E. Long, L. He, L. Zhang, Y. Zhu, C. Chen, Z. Liu, X. Wu, D. Yun, J. Lv, Y. Liu, X. Liu et H. Lin, “Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders,” *Nature Biomedical Engineering*, vol. 4, n^o. 8, p. 767–777, juin 2020. [En ligne]. Disponible : <http://dx.doi.org/10.1038/s41551-020-0577-y>
- [62] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens et Z. Wojna, “Rethinking the inception architecture for computer vision,” dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, p. 2818–2826.
- [63] O. Ronneberger, P. Fischer et T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” dans *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells et A. F. Frangi, édit. Cham : Springer International Publishing, 2015, p. 234–241.
- [64] K.-H. Hung, C. Lin, J. Roan, C.-F. Kuo, C.-H. Hsiao, H.-Y. Tan, H.-C. Chen, D. H.-K. Ma, L.-K. Yeh et O. K.-S. Lee, “Application of a Deep Learning System in Pterygium Grading and Further Prediction of Recurrence with Slit Lamp Photographs,” *Diagnostics*, vol. 12, n^o. 4, p. 888, avr. 2022, number : 4 Publisher : Multidisciplinary Digital Publishing Institute. [En ligne]. Disponible : <https://www.mdpi.com/2075-4418/12/4/888>
- [65] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh et J. Liang, “Unet++ : A nested u-net architecture for medical image segmentation,” 2018. [En ligne]. Disponible : <https://arxiv.org/abs/1807.10165>
- [66] C. Wan, Y. Shao, C. Wang, J. Jing et W. Yang, “A novel system for measuring pterygium’s progress using deep learning,” *Frontiers in Medicine*, vol. 9, févr. 2022. [En ligne]. Disponible : <http://dx.doi.org/10.3389/fmed.2022.819971>
- [67] A. K. Chaudhary, R. Kothari, M. Acharya, S. Dangi, N. Nair, R. Bailey, C. Kanan, G. Diaz et J. B. Pelz, “Ritnet : Real-time semantic segmentation of the eye for gaze tra-

- cking,” dans *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, p. 3698–3702.
- [68] J. Lozej, B. Meden, V. Struc et P. Peer, “End-to-end iris segmentation using u-net,” dans *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOB)*, 2018, p. 1–6.
- [69] M. Poovayar Priya et M. Ezhilarasan, *IRIS Segmentation Technique Using IRIS-UNet Method*. Springer Nature Switzerland, 2023, p. 235–249. [En ligne]. Disponible : http://dx.doi.org/10.1007/978-3-031-45382-3_20
- [70] P. Rot, Z. Emersic, V. Struc et P. Peer, “Deep multi-class eye segmentation for ocular biometrics,” dans *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOB)*, 2018, p. 1–8.
- [71] C. Chen, J. H. Chuah, R. Ali et Y. Wang, “Retinal vessel segmentation using deep learning : A review,” *IEEE Access*, vol. 9, p. 111 985–112 004, 2021.
- [72] A. E. Ilesanmi, T. Ilesanmi et G. A. Gbotoso, “A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks,” *Healthcare Analytics*, vol. 4, p. 100261, 2023. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S2772442523001284>
- [73] Q. Wu et A. Cheddad, “Segmentation-based deep learning fundus image analysis,” dans *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2019, p. 1–5.
- [74] C. Zhou et R. C. Paffenroth, “Anomaly Detection with Robust Deep Autoencoders,” dans *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '17. New York, NY, USA : Association for Computing Machinery, août 2017, p. 665–674. [En ligne]. Disponible : <https://dl.acm.org/doi/10.1145/3097983.3098052>
- [75] Z. Wang, A. Bovik, H. Sheikh et E. Simoncelli, “Image quality assessment : from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, n°. 4, p. 600–612, 2004.
- [76] N. Tuluptceva, B. Bakker, I. Fedulova et A. Konushin, *Perceptual Image Anomaly Detection*. Springer International Publishing, 2020, p. 164–178. [En ligne]. Disponible : http://dx.doi.org/10.1007/978-3-030-41404-7_12
- [77] A. Kascenas, N. Pugeault et A. Q. O’Neil, “Denoising Autoencoders for Unsupervised Anomaly Detection in Brain MRI,” févr. 2022. [En ligne]. Disponible : https://openreview.net/forum?id=Bm8-t_ggzPD

- [78] P. Seebock, S. M. Waldstein, S. Klimesch, H. Bogunovic, T. Schlegl, B. S. Gerendas, R. Donner, U. Schmidt-Erfurth et G. Langs, “Unsupervised Identification of Disease Marker Candidates in Retinal OCT Imaging Data,” *IEEE Transactions on Medical Imaging*, vol. 38, n^o. 4, p. 1037–1047, avr. 2019. [En ligne]. Disponible : <https://ieeexplore.ieee.org/document/8502086/>
- [79] J. An et S. Cho, “Variational autoencoder based anomaly detection using reconstruction probability,” 2015. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:36663713>
- [80] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth et G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” dans *Information Processing in Medical Imaging*, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap et D. Shen, édit. Cham : Springer International Publishing, 2017, p. 146–157.
- [81] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs et U. Schmidt-Erfurth, “f-AnoGAN : Fast unsupervised anomaly detection with generative adversarial networks,” *Medical Image Analysis*, vol. 54, p. 30–44, mai 2019. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1361841518302640>
- [82] M. Arjovsky, S. Chintala et L. Bottou, “Wasserstein generative adversarial networks,” dans *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup et Y. W. Teh, édit., vol. 70. PMLR, 06–11 Aug 2017, p. 214–223. [En ligne]. Disponible : <https://proceedings.mlr.press/v70/arjovsky17a.html>
- [83] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin et A. Courville, “Improved training of wasserstein GANs,” dans *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17. Red Hook, NY, USA : Curran Associates Inc., déc. 2017, p. 5769–5779.
- [84] C. Baur, B. Wiestler, S. Albarqouni et N. Navab, *Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images*. Springer International Publishing, 2019, p. 161–169. [En ligne]. Disponible : http://dx.doi.org/10.1007/978-3-030-11723-8_16
- [85] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen et X. Chen, “Improved Techniques for Training GANs.”
- [86] C. I. Bercea, D. Rueckert et J. A. Schnabel, “What do we learn ? Debunking the Myth of Unsupervised Outlier Detection,” févr. 2023, arXiv :2206.03698 [cs, eess]. [En ligne]. Disponible : <http://arxiv.org/abs/2206.03698>

- [87] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao et F. Shen, “Image data augmentation for deep learning : A survey,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/2204.08610>
- [88] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe et Y. J. Yoo, “Cutmix : Regularization strategy to train strong classifiers with localizable features,” *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, p. 6022–6031, 2019. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:152282661>
- [89] T. Devries et G. W. Taylor, “Improved regularization of convolutional neural networks with cutout,” *ArXiv*, vol. abs/1708.04552, 2017. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:23714201>
- [90] J. Qiu, W. Huang, Y. Huang, N. Yu et X. Tang, “AugPaste : A one-shot approach for diabetic retinopathy detection,” *Biomedical Signal Processing and Control*, vol. 96, p. 106489, oct. 2024. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1746809424005470>
- [91] R. Balestrieri, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian, A. Schwarzschild, A. G. Wilson, J. Geiping, Q. Garrido, P. Fernandez, A. Bar, H. Pirsiavash, Y. LeCun et M. Goldblum, “A cookbook of self-supervised learning,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/2304.12210>
- [92] T. Chen, S. Kornblith, M. Norouzi et G. Hinton, “A simple framework for contrastive learning of visual representations,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2002.05709>
- [93] K. He, H. Fan, Y. Wu, S. Xie et R. Girshick, “Momentum contrast for unsupervised visual representation learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/1911.05722>
- [94] K. Cho, K. D. Kim, Y. Nam, J. Jeong, J. Kim, C. Choi, S. Lee, J. S. Lee, S. Woo, G.-S. Hong, J. B. Seo et N. Kim, “Chess : Chest x-ray pre-trained model via self-supervised contrastive learning,” *Journal of Digital Imaging*, vol. 36, n^o. 3, p. 902–910, janv. 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1007/s10278-023-00782-4>
- [95] Y. Zou, J. Jeong, L. Pemula, D. Zhang et O. Dabeer, *SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation*. Springer Nature Switzerland, 2022, p. 392–408. [En ligne]. Disponible : http://dx.doi.org/10.1007/978-3-031-20056-4_23
- [96] X. Chen et K. He, “Exploring simple siamese representation learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2011.10566>

- [97] J.-B. Grill, F. Strub, F. Alché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos et M. Valko, “Bootstrap your own latent : A new approach to self-supervised learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2006.07733>
- [98] F. Baier, S. Mair et S. G. Fadel, *Self-supervised Siamese Autoencoders*. Springer Nature Switzerland, 2024, p. 117–128. [En ligne]. Disponible : http://dx.doi.org/10.1007/978-3-031-58547-0_10
- [99] Y. Li, Q. Lao, Q. Kang, Z. Jiang, S. Du, S. Zhang et K. Li, “Self-supervised anomaly detection, staging and segmentation for retinal images,” *Medical Image Analysis*, vol. 87, p. 102805, 2023. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S136184152300066X>
- [100] B. M. de Vries, G. J. Zwezerijnen, G. L. Burchell, F. H. van Velden, C. W. Menke-van der Houven van Oordt et R. Boellaard, “Explainable artificial intelligence (xai) in radiology and nuclear medicine : a literature review,” *Frontiers in medicine*, vol. 10, p. 1180773, 2023.
- [101] K. P. F.R.S., “Liii. on lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, n°. 11, p. 559–572, 1901.
- [102] L. van der Maaten et G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, n°. 86, p. 2579–2605, 2008. [En ligne]. Disponible : <http://jmlr.org/papers/v9/vandermaaten08a.html>
- [103] L. McInnes, J. Healy, N. Saul et L. Großberger, “Umap : Uniform manifold approximation and projection,” *Journal of Open Source Software*, vol. 3, n°. 29, p. 861, 2018. [En ligne]. Disponible : <https://doi.org/10.21105/joss.00861>
- [104] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh et D. Batra, “Grad-cam : Visual explanations from deep networks via gradient-based localization,” dans *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, p. 618–626.
- [105] O. Lang, Y. Gandelsman, M. Yarom, Y. Wald, G. Elidan, A. Hassidim, W. T. Freeman, P. Isola, A. Globerson, M. Irani et I. Mosseri, “Explaining in style : Training a gan to explain a classifier in stylespace,” *arXiv preprint arXiv :2104.13369*, 2021.
- [106] T. Karras, S. Laine et T. Aila, “A style-based generator architecture for generative adversarial networks,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 4396–4405, 2018. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:54482423>
- [107] O. Lang, D. Yaya-Stupp, I. Traynis, H. Cole-Lewis, C. R. Bennett, C. R. Lyles, C. Lau, M. Irani, C. Semturs, D. R. Webster, G. S. Corrado, A. Hassidim, Y. Matias,

- Y. Liu, N. Hammel et B. Babenko, “Using generative ai to investigate medical imagery models and datasets,” *eBioMedicine*, vol. 102, p. 105075, avr. 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.ebiom.2024.105075>
- [108] M. Tkachenko, M. Malyuk, A. Holmanyuk et N. Liubimov, “Label Studio : Data labeling software,” 2020-2024, open source software available from <https://github.com/HumanSignal/label-studio>. [En ligne]. Disponible : <https://github.com/HumanSignal/label-studio>
- [109] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [110] Z. Wang, E. Simoncelli et A. Bovik, “Multi-scale structural similarity for image quality assessment,” *Proceedings of the IEEE Asilomar Conference Signals, Systems and Computers*, 02 2004.
- [111] M. Arjovsky, S. Chintala et L. Bottou, “Wasserstein generative adversarial networks,” dans *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup et Y. W. Teh, édit., vol. 70. PMLR, 06–11 Aug 2017, p. 214–223. [En ligne]. Disponible : <https://proceedings.mlr.press/v70/arjovsky17a.html>
- [112] E. Linder-Norén, “Pytorch implementations of generative adversarial networks,” <https://github.com/eriklindernoren/PyTorch-GAN>, [Accessed 04-11-2024].
- [113] K. Zhou, Y. Xiao, J. Yang, J. Cheng, W. Liu, W. Luo, Z. Gu, J. Liu et S. Gao, “Encoding structure-texture relation with p-net for anomaly detection in retinal images,” dans *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox et J.-M. Frahm, édit. Cham : Springer International Publishing, 2020, p. 360–377.
- [114] K. He, X. Zhang, S. Ren et J. Sun, “Identity mappings in deep residual networks,” dans *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe et M. Welling, édit. Cham : Springer International Publishing, 2016, p. 630–645.
- [115] S. Elfving, E. Uchibe et K. Doya, “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” 2017. [En ligne]. Disponible : <https://arxiv.org/abs/1702.03118>
- [116] B. Xu, N. Wang, T. Chen et M. Li, “Empirical evaluation of rectified activations in convolutional network,” 2015. [En ligne]. Disponible : <https://arxiv.org/abs/1505.00853>
- [117] D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization,” 2017. [En ligne]. Disponible : <https://arxiv.org/abs/1412.6980>
- [118] R. O. Duda et P. E. Hart, “Use of the hough transformation to detect lines and curves in pictures,” *Communications of the ACM*, vol. 15, n°. 1, p. 11–15, 1972.

ANNEXE A APPROBATION ÉTHIQUE - PROJET CER-2223-75-D

CER-2223-75-D



Montréal, le 1er novembre 2023

Objet: Approbation éthique – « Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil » - Projet CER-2223-75-D

Mme Lama Seoud,

J'ai le plaisir de vous informer que le Comité d'éthique de la recherche, selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche susmentionné et conclu que ce dernier répond aux normes en vigueur au chapitre de l'éthique de la recherche énoncées dans la *Politique en matière d'éthique de la recherche avec des êtres humains* de Polytechnique Montréal.

Veillez noter que le présent certificat est valable pour une durée d'un an, soit du 1er novembre 2023 au 2 novembre 2024, pour le projet tel qu'approuvé au Comité d'éthique de la recherche avec des êtres humains.

Veillez noter que conformément aux exigences auxquelles l'institution et son personnel sont assujettis afin d'être admissibles aux fonds des organismes subventionnaires, il est de votre responsabilité de déposer au CÉR un rapport annuel ou un rapport final avant l'expiration de la présente approbation éthique afin de l'informer de l'avancement de vos travaux. Le formulaire à remplir est disponible à l'adresse suivante : (<http://www.polymtl.ca/recherche/formulaires-et-guides>).

De plus, il est de votre responsabilité d'informer le CER de toute modification importante qui pourrait être apportée au protocole expérimental avant sa mise en œuvre, de même que de tout élément ou événement imprévu pouvant avoir une incidence sur le bien-être ou l'intégrité des participant(e)s impliqué(e)s dans le projet de recherche. Nous vous invitons aussi à nous signaler tout problème susceptible d'avoir une incidence sur les membres de l'équipe de recherche.

Je vous souhaite bonne chance dans la poursuite de vos travaux.

Nous vous prions d'agréer, Madame, l'expression de nos sentiments les meilleurs,

Philippe Doyon-Poulin, président
Comité d'éthique de la recherche
Polytechnique Montréal

c.c. Direction de la formation et de la recherche; Service des Finances

p.j. Certificat # CER-2223-75-D

**Comité d'éthique de la recherche
avec des êtres humains**
Tél.: 514 340-4711 poste : 3755
Fax : 514 340-4992
Courriel : ethique@polymtl.ca

Adresse postale
C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7

Campus de l'Université de Montréal
2900, boul. Édouard-Montpetit
2500, chemin de Polytechnique
Montréal (Québec) Canada H3T1J4

CER-2223-75-D



CERTIFICAT D'APPROBATION ÉTHIQUE

Le Comité d'éthique de la recherche de Polytechnique Montréal, selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche suivant et conclu qu'il respecte les règles d'éthique énoncées dans sa Politique en matière d'éthique de la recherche avec des êtres humains.

Projet	
Titre du projet	Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil CER-2223-75-D
Chercheuse requérante	Lama Seoud , Professeure adjointe, Département de génie informatique et de génie logiciel
Étudiant(e)s couvert(e)s:	Doha Zrouki, Philippe Baumstimler (2e cycle, Poly)
Agissant à titre d'associé(e) de recherche:	Philippe Debanne

Financement	
Organisme	MITACS
No de UBR	3320899
Programme	accélération
No d'octroi:	IT34496
Titre original de l'octroi:	
Chercheur principal:	

MODALITÉS D'APPLICATION

Toute modification importante qui pourrait être apportée au protocole expérimental doit être transmise au Comité avant sa mise en œuvre.

L'équipe de recherche doit informer le Comité de tout élément ou évènement imprévu pouvant avoir une incidence sur le bien-être ou l'intégrité des participant(e)s impliqué(e)s dans le projet de recherche ainsi que tout problème susceptible d'avoir une incidence sur les membres de l'équipe de recherche.

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du Comité.


Philippe Doyon-Poulin, président
Comité d'éthique de la recherche
Polytechnique Montréal

Date de délivrance :
1er novembre 2023

Date de fin de validité :
1er décembre 2024

Date du prochain
suivi :
2 novembre 2024

Comité d'éthique de la recherche
avec des êtres humains
Tél.: 514 340-4711 poste : 3755
Fax : 514 340-4992
Courriel : ethique@polymtl.ca

Adresse postale
C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7

Campus de l'Université de Montréal
2900, boul. Édouard-Montpetit
2500, chemin de Polytechnique
Montréal (Québec) Canada H3T 1J4

ANNEXE B FORMULAIRE D'INFORMATION ET DE CONSENTEMENT

**Formulaire d'information et de consentement**

Titre de l'activité de recherche :

Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil

Équipe de recherche :**Responsable de l'activité de recherche**

Doha Zrouki

Étudiante à la maîtrise

Polytechnique Montréal – Département GIGL

Adresse courriel: doha.zrouki@polymtl.ca

Responsable de l'activité de recherche

Philippe Baumstimler

Étudiant à la maîtrise

Polytechnique Montréal – Département de GIGL

Adresse courriel: philippe.baumstimler@polymtl.ca

Sous la direction de

Lama Séoud

Professeure superviseure

Polytechnique Montréal – Département GIGL

Adresse courriel: lama.seoud@polymtl.ca

Ce texte est rédigé au masculin dans le seul but d'alléger le texte.

Financement de l'activité de recherche :

La présente activité de recherche est financée par l'organisme fédéral de recherche sans but lucratif MITACS en partenariat avec le partenaire industriel LightX Innovations, entreprise qui propose une application mobile pour capter des images du segment antérieur de l'œil et qui développe des algorithmes d'intelligence artificielle pour analyser ces images.

Conflits d'intérêts

Dépendamment des résultats, les avancées et améliorations issues de la recherche pourront être utilisées par le partenaire à des fins commerciales. Les données collectées serviront à entraîner des algorithmes de détection de pathologies, ces derniers (et non les données) pouvant faire l'objet d'une commercialisation.

Les intérêts de Polytechnique sont liés la formation de personnel hautement qualifié et à l'avancement des connaissances. Ni l'équipe de recherche, ni Polytechnique ne recevront de redevance advenant la commercialisation des algorithmes par le partenaire.

Préambule

Nous vous invitons à participer à une activité de recherche qui vise à évaluer le pouvoir diagnostique des images du globe oculaire externe combinées à l'apprentissage machine afin d'assister les professionnels médicaux en ophtalmologie.

Cependant, avant d'accepter de participer à cette activité et de signer le présent formulaire d'information et de consentement, veuillez prendre le temps de lire toute l'information présentée.

Si vous consentez à participer à cette activité de recherche, un deuxième formulaire d'information et de consentement vous sera proposé par le partenaire industriel. Celui-ci concerne la collecte par la compagnie des images désidentifiées à des fins de recherche (autre que le présent projet). La gestion des données collectées par le partenaire se fera selon les termes établis dans leur formulaire d'information et de consentement, l'École Polytechnique n'ayant aucune autorité sur ce dernier. Le second formulaire est optionnel et le non-consentement à ce dernier n'empêche pas la participation à l'activité de recherche après signature du présent formulaire de consentement et d'information.

Nous vous invitons à poser toutes les questions que vous jugerez utiles à la personne responsable de l'activité de recherche ou à tout autre membre de l'équipe de recherche et à leur demander de vous expliquer tout mot ou renseignement qui ne serait pas clair. Nous vous invitons également à prendre conseil auprès de toute autre personne de qui vous aimeriez obtenir un avis à propos de votre éventuelle participation.

Présentation générale du projet de recherche

Des recherches récentes démontrent qu'il est possible de détecter des signes de pathologies comme la rétinopathie diabétique à partir d'images externes du globe oculaire. Ces images ont l'avantage d'être acquises avec des caméras standards, sans besoin de dilater la pupille, contrairement à des images du fond d'œil pour capturer la rétine.

Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil

L'objectif de notre projet est d'évaluer le pouvoir diagnostique des images de la partie antérieure du globe oculaire pour le glaucome en particulier, et pour des pathologies au sens large. Pour cela nous exploitons la base de données d'images du globe oculaire dont dispose notre partenaire industriel de sorte à développer des algorithmes de détection de pathologies au moyen de méthodes d'apprentissage machine. Cependant, cette base de données ne présente que très peu de données issues de sujets n'ayant pas de pathologie oculaire, ce qui représente un frein pour notre recherche. C'est pourquoi nous souhaitons bâtir une base de données « normative », comportant des images acquises chez des personnes dites « saines » c'est-à-dire sans pathologie oculaire connue, desquelles nous pourrions constituer les ensembles d'entraînement, de validation et de test de nos algorithmes. Ces derniers seront utilisés pour entraîner et évaluer des réseaux de neurones profonds à classifier les images selon le diagnostic associé, et à reconnaître des images présentant des maladies.

Les principaux objectifs de l'étude que nous réalisons sont les suivants :

- Développer des outils pour analyser de manière automatique les images du globe oculaire externe et détecter la présence de PACS (*Primary Angle Closure Suspect*) qui est un signe précurseur du glaucome ;
- Développer des outils de détection d'anomalies des images du globe oculaire externe.

À plus long terme, nos travaux de recherche pourront augmenter l'accessibilité au diagnostic ophtalmologique, par exemple à partir d'images de lampes à fentes dans des cliniques d'optométrie, ou par un suivi du patient à distance.

Afin d'atteindre nos objectifs, nous effectuerons une collecte qui vise à recruter 100 participants, sur lesquels seront réalisés des tests oculaires par examen de lampe à fente. Les tests seront réalisés par un ophtalmologue et chaque test résultera en une paire de photos stéréoscopiques (une pour chaque œil). Seuls vos yeux seront pris en photo. En addition aux images acquises, nous vous demanderons de nous fournir les informations suivantes :

- Votre âge
- Votre sexe
- Votre origine ethnique
- Votre correction optique (si connue)

Ces informations ne sont collectées qu'à des fins d'analyse de sensibilité des modèles d'apprentissage et ne seront transmises à aucune personne en dehors de l'équipe de recherche.

Critères d'inclusion et d'exclusion

Dans le cadre de cette activité, nous cherchons à recruter des personnes majeures, aptes à donner leur consentement, francophones ou anglophones, voyantes, de tout genre et de toute origine. Vous ne pouvez **pas** participer à cette activité si vous êtes atteint, ou avez déjà été atteint, d'une des maladies suivantes :

- Le glaucome
- La rétinopathie diabétique
- La dégénérescence maculaire
- Des cataractes.

Nature et durée de votre participation à l'activité de recherche

Votre participation à l'activité de recherche prendra la forme d'une acquisition d'images ophtalmiques moyennant un dispositif, appelé lampe à fente, qui sera manipulé par un médecin ophtalmologue. Une paire d'images stéréoscopique de chaque œil sera acquise dans 7 configurations de lumière différentes, permettant chacune de visualiser une partie spécifique de l'œil (conjonctive, cornée, iris, cristallin, chambre antérieure et humeur vitrée). Au total, ce seront 28 images qui seront collectées par participant. La procédure d'acquisition au complet durera entre 5 et 10 minutes.

Des informations additionnelles vous concernant (âge, sexe, ethnicité, correction optique si connue) seront recueillies par l'équipe de recherche à des fins d'analyse de sensibilité des modèles.

Risques pouvant découler de votre participation à l'activité de recherche

La présente activité ne devrait pas entraîner des risques plus grands que ceux rencontrés dans votre vie de tous les jours. La manipulation de l'appareil d'acquisition sera entièrement réalisée par un expert du domaine médical (un médecin ophtalmologue d'expérience), et ne requiert pas l'administration de gouttes oculaires. Les tests réalisés impliquant uniquement de la lumière dans le spectre du visible, une gêne associée à l'illumination pourrait être ressentie.

Inconvénients pouvant découler de votre participation à l'activité de recherche

Les inconvénients liés à la participation à la recherche sont le temps nécessaire à l'acquisition (maximum 10 minutes), le déplacement au laboratoire où se tiendra la prise d'acquisition et la possible gêne causée par l'illumination de la lampe à fente. De plus, les personnes portant des verres de contacts devront également les retirer le temps de la collecte.

Avantages pouvant découler de votre participation à l'activité de recherche

Votre participation à l'étude vous permettra d'effectuer un test oculaire avec un ophtalmologue, qui, s'il observe une quelconque anomalie, pourra vous le communiquer.

Découverte fortuite:

En cas de découverte fortuite d'un signe de pathologie pendant l'acquisition, l'ophtalmologue responsable de l'acquisition pourra le communiquer au participant et le référer à son médecin traitant. Toutefois, la détection éventuelle de signes de pathologies après la collecte d'images

Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil

(découlant des algorithmes développés par l'équipe de recherche) ne sera pas communiquée aux participants en vertu de la confiance accordée à l'avis du professionnel de la santé lors de l'acquisition des données.

Compensation financière

Vous recevrez une compensation financière de 20\$ pour votre temps.

Indemnisation en cas de préjudice et droits des participant(e)s

Si vous deviez subir quelque préjudice que ce soit par suite de votre participation à cette activité de recherche, vous ne renoncez à aucun de vos droits ni ne libérez les chercheurs, l'organisme de financement ou Polytechnique Montréal de leurs responsabilités légales et professionnelles.

Participation volontaire et possibilité de retrait :

Votre participation à la présente activité de recherche est volontaire. Vous êtes donc libre de refuser d'y participer et pouvez à tout moment décider de vous en retirer sans avoir à motiver votre décision et sans risquer d'en subir de préjudice. Vous n'avez qu'à en informer la personne-ressource de l'équipe de recherche et ce, par simple avis verbal.

En cas de retrait, vous pouvez demander la destruction des données vous concernant. Cependant, il sera impossible de retirer vos données ou votre matériel des analyses menées une fois ces dernières publiées ou diffusées. La compensation financière obtenue pourra être gardée par le participant après son retrait.

Tout au long des activités de recherche, et si vous y consentez, vous recevrez en temps opportun l'information pertinente en lien avec votre participation.

L'équipe de recherche s'assurera du bon établissement du processus de consentement dans un cadre privé et non-coercif.

L'équipe de recherche et le comité d'éthique de la recherche se réservent le droit de vous retirer de l'étude si vous ne respectez pas les consignes, s'il existe des raisons administratives d'abandonner l'activité, si votre participation n'est plus sécuritaire ou pour toutes autres raisons concernant la faisabilité de l'étude. Si une telle situation survient, l'équipe de recherche vous en informera dès que possible.

Confidentialité et protection de vos données

L'équipe de recherche recueillera et consignera toutes vos données (images et informations démographiques) de manière sécuritaire de façon à en protéger le caractère confidentiel.

Voici comment nous protégerons vos données **lors de la collecte** :

- Vos informations seront désidentifiées et une clé d'identification reliant vos données à

Analyse automatique par apprentissage profond d'images du segment antérieur de l'œil

vos nom et prénom sera conservée uniquement par la chercheuse responsable de ce projet.

- La clé d'identification ne sera utilisée que pour retracer vos données et les supprimer suite à un retrait éventuel de votre consentement.
- Les données désidentifiées seront sauvegardées sur un disque dur avec clé d'encryption appartenant à l'équipe de recherche.

Voici comment nous protégerons vos données **lors des analyses et du transfert des données** entre les membres de l'équipe :

- Les informations recueillies auprès des participants resteront confidentielles et ne seront partagées à aucune personne en dehors de l'équipe de recherche.
- Les informations démographiques ne serviront qu'à des fins d'analyse de sensibilité des modèles et statistiques.
- Le partage des données se fera sur des plateformes sécurisées ou par disque dur externe avec clé d'encryption.
- Les données pourront être copiées sur le disque dur local des ordinateurs du laboratoire de recherche le temps de l'entraînement et de la validation des modèles, mais seront intégralement supprimées une fois ces étapes réalisées.

Voici comment nous protégerons vos données **lors des publications** :

- Les données seront désidentifiées, et aucune mention de l'identité ni des coordonnées des participants ne sera faite lors des éventuelles publications.
- Les images ne comprennent qu'un seul œil à la fois et non le reste du visage; il sera donc impossible de retracer l'identité du participant à partir de ces dernières.

Enfin, voici comment nous protégerons vos données **après le projet de recherche** :

- Vos données seront conservées sur un disque dur avec clé d'encryptions détenu par la chercheuse responsable du projet, Mme Lama Séoud, pour une période de 10 ans après la fin de ce projet de recherche, en vertu des normes de pratique professionnelles.

Vous avez le droit de consulter votre dossier de recherche pour vérifier l'exactitude des renseignements recueillis aussi longtemps que l'équipe de recherche ou Polytechnique Montréal détiendront ces informations. Cependant, afin de préserver l'intégrité scientifique du projet de recherche, certaines informations seront accessibles seulement à la fin du projet de recherche.

Diffusion des résultats de la recherche

Le travail effectué dans le cadre de ce projet de recherche à partir des données récoltées fera l'objet de possibles publications dans des mémoires et des revues scientifiques ou présentés lors de congrès scientifiques. Les participants seront tenus informés des différentes publications et présentations des résultats de l'étude par courriel uniquement s'ils consentent à être recontactés à cet effet.

Personnes-ressources

Si vous avez des questions sur les **aspects scientifiques** du projet de recherche ou pour vous **retirer de l'étude**, vous pouvez contacter la professeure superviseure Lama Seoud par courriel à lama.seoud@polymtl.ca.

Pour toute préoccupation sur vos droits ou sur les responsabilités de l'équipe de recherche concernant votre participation à ce projet, vous pouvez contacter le Comité d'éthique de la recherche de Polytechnique Montréal au (514) 340-4711, poste 4420 ou encore par courriel à ethique@polymtl.ca

Consentement à la participation au projet de recherche

1. J'ai pris connaissance de la documentation ci-jointe, décrivant la nature et le déroulement du projet de même que les risques et les inconvénients qui pourraient survenir.
2. Je comprends que j'ai droit à des réponses satisfaisantes aux questions que je poserais quant à mon implication dans ce projet tout au long de ma participation.
3. Je consens à participer librement à ce projet, après avoir obtenu et pris le temps d'y réfléchir à ma satisfaction et sans avoir subi de pression à cet effet.
4. Je comprends qu'en participant à ce projet de recherche, je ne renonce à aucun de mes droits ni ne dégage les chercheurs de leurs responsabilités.
5. Je comprends que je peux consulter le dossier que l'équipe de recherche constitue sur moi.
6. Je pourrai à tout moment, sur simple avis de ma part, revenir sur ma décision de participer et serai alors immédiatement libéré de mon engagement.
7. J'ai reçu une copie du présent document.
8. Je consens à fournir les informations suivantes me concernant à condition qu'elles soient anonymisées : âge, sexe, origine ethnique et correction visuelle (si connue).
9. Je consens à ce que des photos de mes yeux soient utilisées à des fins d'entraînement de modèles d'apprentissage machine comprenant l'annotation des données anonymisées.
10. Je consens à ce que des photos de mes yeux soient utilisées dans des publications scientifiques à condition qu'aucune mention de mon identité ne soit faite.
11. Je comprends que les résultats de l'activité de recherche pourraient mener à une éventuelle commercialisation par le partenaire industriel, mais que ma participation à la collecte ne me donne aucun droit aux bénéfices qui en découleraient.

(Veuillez cocher votre choix) Oui Non

Consentement à partager mes informations de contact

J'autorise le groupe de recherche à me recontacter par courriel pour me transmettre leurs résultats de recherche.

(Veuillez cocher votre
choix)

Oui _____ Non

Adresse courriel (caractères
d'imprimerie)

Prénom et nom du participant
(caractère d'imprimerie)

Signature du participant

Date :

Engagement de l'équipe de recherche

Je confirme que moi ou mon représentant avons expliqué à la personne précitée la nature de sa participation à la présente activité de recherche, demandé si elle avait des questions, et répondu à ses questions. Nous avons clairement indiqué qu'elle ou il demeurerait libre de participer et de mettre un terme à sa participation à tout moment, par simple avis verbal. Je m'engage, avec l'équipe de recherche, à respecter les modalités décrites dans le présent formulaire d'information et de consentement et déclare en avoir remis une copie signée à la personne.

Lama Séoud
(caractères d'imprimerie)

Signature de la responsable

Date : _____

Doha Zrouki / Philippe Baumstimler
(caractères d'imprimerie)

Signature de l'étudiant(e)

Date : _____

ANNEXE C PRÉTRAITEMENT : DESCRIPTION & ALGORITHMES

Dans cette annexe sont décrites en détail les étapes du pré-traitement ainsi que leur implémentation.

Détection de contours

La première étape du pré-traitement est celle de la détection des contours du masque circulaire dans nos images. Elle a pour but de simplifier le travail de la détection de cercle, en ne gardant que l'information utile. En effet, le traitement se fait sur des images RGB contenant beaucoup d'information structurelle, dont des structures circulaires comme l'iris ou la pupille. On fait donc l'hypothèse que la majorité de l'information du masque correspond aux régions d'intensité lumineuse supérieures à un certain seuil s , à déterminer empiriquement. En se servant de la fonction `cv2.findContours` de la librairie *OpenCV* [109], on ne garde alors que les pixels associés aux contours du masque circulaire. Pour des raisons de temps de calcul, l'image des contours est redimensionnée par un ratio $r < 1$. Le redimensionnement conserve l'information de la position du cercle dans l'image, mais réduit drastiquement le nombre de pixels à traiter pour la suite du pipeline.

L'implémentation est détaillée dans le pseudo-code 1.

Algorithme 1 Algorithme de détection de contours

Entrée: <code>img</code>	▷ Photographie du segment antérieur de l'œil
Sortie: <code>img_contours</code>	▷ Image de contours grossiers du masque circulaire

- 1: $r \leftarrow 0.25$
- 2: $s \leftarrow 30$
- 3: `img_gris` ← Convertir `img` en niveau de gris
- 4: `img_gris` ← Redimensionner `img_gris` par un ratio r
- 5: `contours` ← Appliquer un seuillage binaire à `img_gris` de seuil s
- 6: `contours` ← Trouver les contours externes dans `contours`
- 7: `img_contours` ← Tracer `contours` dans une image noire (0) de même dimension que `img_gris` et configurer les contours en blanc (255)
- 8: **retourne** `img_contours`

Détection de cercles

À partir de l'image de contours grossiers obtenue à l'étape précédente, on souhaite estimer l'équation de l'ensemble des cercles plausibles. Pour ce faire, nous allons nous servir de la

transformée de Hough circulaire [118], pré implémenté dans *OpenCV* [109]. Soit l'équation d'un cercle \mathcal{C} , de centre (c_x, c_y) et de rayon r , définie par

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid (x - c_x)^2 + (y - c_y)^2 = r^2\} \quad (\text{C.1})$$

La transformée de Hough circulaire considère un espace paramétrique en trois dimensions (c_x, c_y, r) , correspondant aux paramètres du cercle \mathcal{C} . À rayon fixé, l'espace paramétrique se réduit à un espace en deux dimensions correspondant aux coordonnées du centre du cercle d'un même rayon, et de dimensions égales à l'image d'entrée. Les valeurs en chaque pixel (c_x, c_y) dans cet espace paramétrique correspondent alors au nombre de pixels de l'image d'origine qui intersecte avec le cercle de centre (c_x, c_y) et de rayon r . On parle alors de matrice d'accumulation. Les maximums locaux représentent les paramètres (c_x, c_y, r) les plus probables pour l'image d'entrée. De fait, l'utilisation d'une telle méthode s'avère en général très lourde, en raison de la taille de l'accumulateur, en particulier pour nos images en haute résolution (2000x4000 pixels). Pour réduire au maximum le temps de calcul, nous avons opté pour un redimensionnement de l'image de contours, et un choix d'hyperparamètres astucieux, par exemple en ne regardant que les rayons compris dans un certain intervalle $[\text{rayon_min}, \text{rayon_max}]$ déterminé empiriquement. Les détails de l'implémentation sont décrits dans le pseudo-code 2.

Algorithme 2 Algorithme de détection de cercles

Entrée: `img_contours`

▷ Image des contours

Sortie: `cercles`

▷ Liste des cercles présents dans l'image

```

1: cercles ← None
2: rayon_min ← min(hauteur img_contours, largeur img_contours)//4
3: rayon_max ← max(hauteur img_contours, largeur img_contours)
4: dist_min ← 1000
5: dp ← 1.0
6: param1 ← 1
7: param2 ← 10
8: r ← 0.25
9: cercles ← cv2.HoughCircles(img_contours, cv.HOUGH_GRADIENT, dp, min_dist,
   param1, param2, rayon_min, rayon_max)
10: si cercles n'est pas None alors
11:   cercles ← Redimensionner selon le ratio 1/r
12:   retourne cercles
13: sinon
14:   retourne None
15: fin si

```

Sélection du meilleur candidat

L'algorithme 2 nous renvoie une liste de cercles plausibles pour le masque circulaire des photographies du segment antérieur de l'œil de notre base de données. Cependant, seul un seul correspond à celui qui nous intéresse. Empiriquement, nous avons observé que la région d'intérêt est relativement bien centrée dans nos images, et ce, dans la majorité des cas. Dès lors, la sélection se fait en choisissant le cercle candidat dont le centre est le plus proche de l'image d'origine. Pour éviter les cas où le cercle le plus proche se situe dans un des coins, on restreint la sélection dans la région centrale de l'image définie par un ratio de ses dimensions. Un cercle dont le centre se situe en dehors de cette région ne sera donc pas considéré, car très certainement erroné. Les détails de l'implémentation sont décrits dans le pseudo-code 3.

Algorithme 3 Algorithme de sélection du meilleur candidat

Entrée: `img` ▷ Photographie du segment antérieur de l'œil
Entrée: `cercles` ▷ Liste des cercles détectés à l'étape précédente
Sortie: `cercle_final` ▷ Cercle final sélectionné

```

1: h ← hauteur de img
2: w ← largeur de img
3: centre_x ← h/2
4: centre_y ← w/2
5: ratio_central ← 0.2
6: si cercles n'est pas None alors
7:   d ← Distances euclidiennes entre les centres des cercles et (centre_x, centre_y)
8:   i ← index de la valeur minimale dans d
9:   si d[i] < ratio_central × min(h, w) alors
10:    retourne cercles[id]
11:   fin si
12: fin si
13: retourne None

```

Rognage circulaire

Enfin, une fois les paramètres du cercle définis, il nous suffit de rogner l'image autour du cercle inscrit le plus grand, afin de perdre le minimum d'information utile. De cette manière, l'ensemble des données est mis au format carré, centré autour d'un nouveau masque circulaire, et bordé de noir absolu, soit une valeur pixel de 0. Dans les cas où la détection de cercle a failli, le masque circulaire choisi sera un disque centré autour du centre de la photographie, et de rayon égal à la plus petite des dimensions spatiales de cette dernière. L'implémentation est détaillée dans le pseudo-code 4.

Algorithme 4 Algorithme de rognage circulaire

Entrée: `img` ▷ Photographie du segment antérieur de l'œil
Entrée: `cercle` ▷ Équation du cercle (x,y,r)
Sortie: `prep_img` ▷ Image rognée autour des coordonnées du cercle

- 1: `w` ← largeur de `img`
- 2: **si** `cercle` n'est pas `None` **alors**
- 3: `x,y,r` ← `cercle`
- 4: `d` ← $\min(x + r, w) - \max(x - r, 0)$
- 5: `nr` ← $d/2$
- 6: `nx` ← $nr + \max(x - r, 0)$
- 7: `mask` ← Dessiner un disque blanc (255) sur une image noire (0) de mêmes dimensions que `img`, de centre (nx, y) et de rayon `nr`
- 8: `prep_img` ← `img` × `mask`
- 9: `prep_img` ← Rogner l'image autour des coordonnées du cercle (nx,y,nr)
- 10: **sinon**
- 11: `h` ← hauteur de `img`
- 12: `x` ← $w/2$
- 13: `y` ← $h/2$
- 14: `r` ← $\min(h, w)/2$
- 15: `mask` ← Dessiner un disque blanc (255) sur une image noire (0) de mêmes dimensions que `img`, de centre (x, y) et de rayon `r`
- 16: `prep_img` ← `img` × `mask`
- 17: `prep_img` ← Rogner l'image autour des coordonnées du cercle (x,y,r)
- 18: **fin si**
- 19: **retourne** `prep_img`
