Computational Visual Media
https://doi.org/10.26599/CVM.2025.9450399 Vol. 11, No. 2, April 2025, 241–268

Review Article

Addressing missing modality challenges in MRI images: A comprehensive review

Reza Azad[1], **Mohammad Dehghanmanshadi**[2], **Nika Khosravi**[1], **Julien Cohen-Adad**[3,4,5], **and Dorit Merhof**[1,6,7] (✉)

Abstract Magnetic resonance imaging (MRI) is one of the most prevalent imaging modalities used for diagnosis, treatment planning, and outcome control in various medical conditions. MRI sequences provide physicians with the ability to view and monitor tissues at multiple contrasts within a single scan and serve as input for automated systems to perform downstream tasks. However, in clinical practice, there is usually no concise set of identically acquired sequences for a whole group of patients. As a consequence, medical professionals and automated systems both face difficulties due to the lack of complementary information from such missing sequences. This problem is well known in computer vision, particularly in medical image processing tasks such as tumor segmentation, tissue classification, and image generation. With the aim of helping researchers, this literature review examines a significant number of recent approaches that attempt to mitigate these problems. Basic techniques such as early synthesis methods, as well as later approaches that deploy deep learning, such as common latent space models, knowledge distillation networks, mutual information maximization, and generative adversarial networks (GANs) are examined in detail. We investigate the novelty, strengths, and weaknesses of the aforementioned strategies. Moreover, using a case study on the segmentation task, our survey offers quantitative benchmarks to further analyze the effectiveness of these methods for addressing the missing modalities challenge. Furthermore, a discussion offers possible future research directions.

Keywords missing modality; survey; deep learning; magnetic resonance imaging (MRI)

1 Introduction

Magnetic resonance imaging, widely known as MRI, is one of the most effective techniques used in biomedical imaging for obtaining high contrast images of soft human body tissues such as the brain [1–3], abdominal organs [4–6], legs [7, 8], spine [9], and other tissues [10, 11].

T1-weighted, contrast-enhanced T1-weighted (T1c-weighted), T2-weighted, fluid attenuation inversion recovery (FLAIR), magnetization prepared-rapid gradient echo (MP-RAGE), and proton density-weighted (PD-weighted) sequences are among the most frequently used MRI modalities and are of particular interest in our study because each reveals distinct characteristics of human tissue. From a clinical perspective [12], each kind of sequence provides vital and complementary information to the medical professional, and hence, more accurate diagnoses can be made when all modalities are present [13–17]. Similarly, in a machine learning context,

1 Faculty of Electrical Engineering and Information Technology, RWTH Aachen University, Aachen 52074, Germany. E-mail: R. Azad, reza.azad@rwth-aachen.de; N. Khosravi, khosravi@hia.rwth-aachen.de.
2 School of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran 14496-14535, Iran. E-mail: M. Dehghanmanshadi, m.dehghan9975@gmail.com.
3 NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal H3T 1J4, Canada. E-mail: J. Cohen-Adad, jcohen@polymtl.ca.
4 Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal H3T 1J4, Canada.
5 MILA, Quebec AI Institute, Montreal H2S 3H1, Canada.
6 Faculty of Informatics and Data Science, University of Regensburg, Regensburg 93053, Germany. E-mail: D. Merhof, dorit.merhof@ur.de (✉).
7 Fraunhofer Institute for Digital Medicine MEVIS, Bremen 28359, Germany.
Manuscript received: 2023-05-27; accepted: 2023-12-27

a missing modality can significantly reduce overall performance [18].

MRI, being one of the most effective methods for obtaining high-quality images of human tissue, is prone to artifacts for various reasons, which can result in one or more missing imaging sequences in practical scenarios. These artifacts are often caused by MRI hardware failure or interactions between the patients and imaging devices [19]. Some examples of these artifacts include the flow of cerebrospinal fluid in the brain and spinal canal, magnetic susceptibility artifacts, and different types of noise [20]. Additionally, the process of acquiring multiple sequences for a patient may be hindered by various other factors, such as scan time, different hospital protocols, limitations of imaging devices, or allergies to certain contrast materials [21, 22]. Consequently, the loss of complementary information provided by the missing modalities poses challenges for physicians and automated systems alike.

In this article, we define *missing modality* as the situation where one or more modalities are absent during inferencing, even though all modalities (e.g., T1w, T2w, FLAIR) are available during training. Following existing literature, we consider each MRI sequence (or contrast) as a unique modality and focus on approaches proposed to address the issue of missing modalities in MRI images. The importance of missing modalities in medical treatment is emphasized in Ref. [12]; annotating an object of interest (e.g., a brain tumor) in MRI images is prone to uncertainty and mistakes, leading to a decline in the radiologist's performance by up to 21% due to human errors during the annotation process in MRI images [12]. The situation may worsen if one or more modalities are missing, as each modality contains specific information that might not be entirely recoverable using the remaining modalities. Therefore, an automatic algorithm capable of compensating for the missing modality plays a significant role in clinical applications, which is the focus of our case study in this review paper.

Several methods for addressing the missing modalities problem in MR images have been presented over the years [21–26]. Early attempts suggested strategies for synthesizing or imputing absent input data. The early synthesis methods usually reconstruct the missing modality images by learning the most important features from an atlas image, and then utilize a traditional classification mechanism to perform voxel-by-voxel intensity prediction [27, 28]. Later methods used deep learning and took various approaches to solving the problem. These include translating the modalities to a shared latent subspace [29], knowledge distillation [18, 30–33], optimizing key feature information across all modalities [34–36], and employing conditional generative adversarial networks (cGAN) [37–40].

This review covers all of the aforementioned approaches, as well as various unique networks for each direction. In particular, we discuss recent works that attempt to compensate for missing modalities. Well-known techniques introduced prior to May 2023 are included, and are grouped into five categories, shown in Fig. 1. Section 2 includes a taxonomy of methods reviewed. A brief description of MRI sequence acquisition, MRI modalities, and probable MRI artifacts is also provided. Section 3 includes a wide range of innovative strategies, including early methods and deep learning-based strategies with a greater focus on deep learning approaches that represent the state of the art in semantic segmentation while addressing missing MRI modalities. The novelty, strengths, weaknesses, training datasets, network design, and major contributions of each approach are considered. Section 3 provides a comparative overview. Section 4 summarizes the most widely used semantic segmentation benchmarks and highlights their key aspects. Section 5 assesses the effectiveness of the proposed methods employing widely used evaluation metrics, which are provided at the start of this section. Section 6 reviews the obstacles and potential solutions of semantic segmentation networks with missing modalities. Section 7 concludes this review.

2 Taxonomy

In this section, we first present our search strategy, then outline our taxonomy (see Fig. 1) that categorizes different strategies to overcome the problem of semantic segmentation of MRI images with missing modalities. We will elaborate on these approaches in Section 3. In the rest of this section, we briefly explain the MRI technique, MRI modalities, and some of the artifacts commonly seen in MRI data.

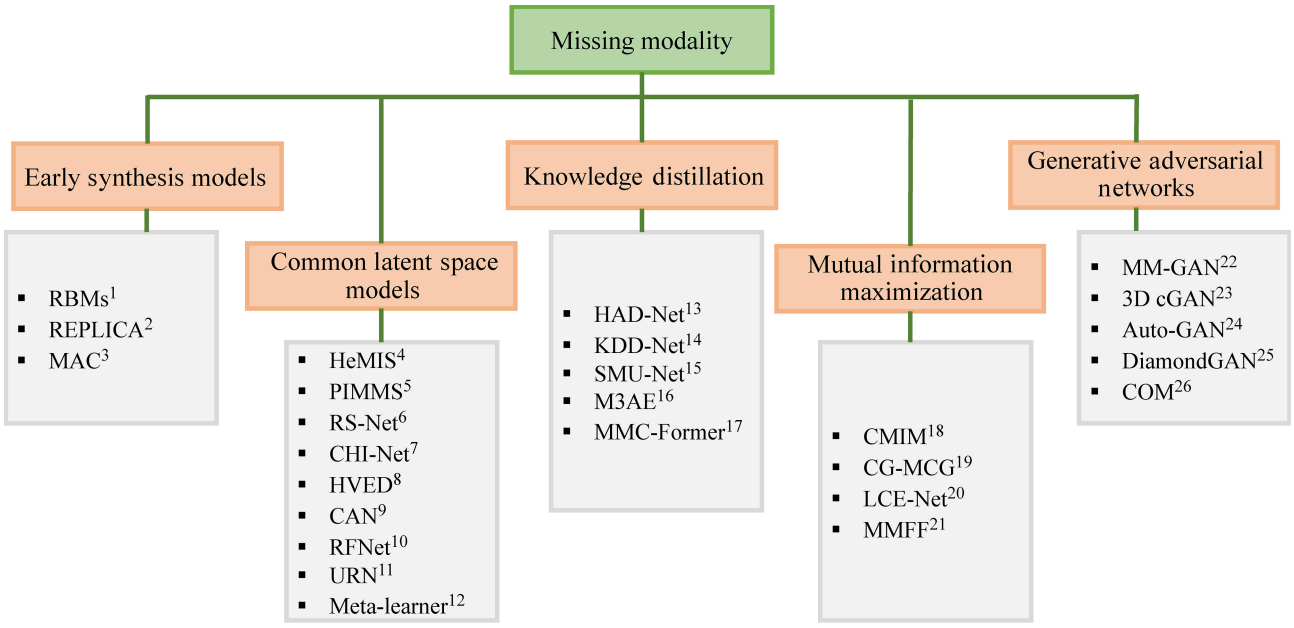

Fig. 1 The proposed taxonomy for the reviewed methods on MRI-based semantic segmentation with missing modalities. The superscripts in ascending order represent: 1. [27], 2. [28], 3. [41], 4. [29], 5. [42], 6. [43], 7. [44], 8. [45], 9. [46], 10. [47], 11. [48], 12. [49], 13. [31], 14. [32], 15. [18], 16. [50], 17. [33], 18. [35], 19. [51], 20. [34], 21. [52], 22. [37], 23. [38], 24. [53], 25. [54], 26. [55].

2.1 Search strategy

A comprehensive search strategy was employed to identify relevant scholarly publications on missing modality models. The search was conducted using DBLP, Google Scholar, and Arxiv Sanity Preserver, which allow for customized search queries and provide extensive lists of scholarly publications, including peer-reviewed journal papers, conference or workshop proceedings, non-peer-reviewed papers, and preprints. Our search query was formulated as: `(missing modality* deep| medical| MRI*) (compensation| MRI images*) (deep* | medical* | missing* | model*) (incomplete* | data* | MRI*)`. The search results were carefully filtered to eliminate false positives, focusing solely on papers directly related to missing modality models. Notably, the selection of papers for detailed examination was based on a thorough evaluation of their novelty, contribution, significance, and whether they represented a state-of-the-art (SOTA) approach. At least five of the highest-ranked papers meeting these criteria were chosen for in-depth analysis. It is important to acknowledge that other noteworthy papers in the field may exist that are not included in our review. However, our aim was to provide a comprehensive overview of the most important and impactful papers concerning missing modality models.

2.2 Magnetic resonance imaging

In this study, we focus on MRI, which is frequently used for acquiring high-contrast images of soft tissues with high spatial resolution.

MRI modalities encompass distinct imaging protocols that unveil different facets of human tissues. The amalgamation of various parameters, such as RF pulses, repetition times (TRs), and echo times (TEs), gives rise to diverse MRI series types, commonly referred to as modalities. These modalities play a pivotal role in medical imaging, each emphasizing specific tissue characteristics. For instance, alterations in RF pulses, TRs, and TEs lead to the creation of distinct MRI series types, offering varied insights into the human body [56]. MRI employs a technique known as nuclear magnetic resonance (NMR), which is based on the magnetization characteristics of the nuclei of hydrogen atoms ^{1}H, abundant in the human body, primarily in water but also in fat. A hydrogen atom consists of a single proton in its nucleus and a single electron orbiting the nucleus. Protons are the main focus of MRI, as the technique utilizes the inherent magnetic properties of protons. Protons spin around their arbitrarily aligned axis, and the resultant magnetic moment is responsible for the proton's magnetic characteristics. During the MRI

procedure, an external strong uniform magnetic field, over 30,000 times stronger than the Earth's magnetic field, is applied to the patient's body, causing the proton spins to align with the field's orientation and to precess around the field lines. The frequency at which protons precess, known as the Larmor frequency, is proportional to the intensity of the static magnetic field. Subsequently, a radio frequency (RF) current pulse is applied, which disturbs the proton alignment and induces coherent precession of the spins. The RF coils detect the induced magnetization, and the spatial information of the image is directly encoded in the Fourier domain through the manipulation of magnetic gradients [57, 58].

The inherent difference between pathological and healthy tissue is revealed in MRI scans by adopting specific settings and protocols. Other methods such as planar X-rays and computed tomography (CT) expose the body to ionizing radiation, making MRI a more attractive and safer option. MRI is widely acknowledged as the most effective technology for observing and segmenting human tissue, particularly the brain.

2.3 MRI modalities

There are several MRI sequences, each of which highlights unique features of human tissue. We briefly present the most common sequences.

2.3.1 T1-weighted

T1-weighted scans reveal fat within the human body: fatty tissues appear brighter in T1-weighted images than other anatomical tissues [59].

2.3.2 T2-weighted

T2-weighted images accentuate water as well as fat, resulting in both fat and water containing tissues appearing bright [60].

2.3.3 Contrast-enhanced T1 weighted (T1c-weighted)

In contrast-enhanced T1-weighted imaging, a gadolinium-based contrast medium is injected into the patient's bloodstream to shorten the T1-relaxation period and improve the detection of lesions with disruption of the blood–brain barrier.

2.3.4 Fluid attenuation inversion recovery (FLAIR)

The appearance of human body tissue in FLAIR scans is similar to that of T2-weighted scans, with the exception that cerebrospinal fluid appears dark rather than bright [61].

2.3.5 Magnetization prepared-rapid gradient echo (MP-RAGE)

MP-RAGE scans provide good contrast between white and gray tissues with a relatively short scan time. They are used in a large number of multicenter trials, such as the Alzheimer's disease neuroimaging initiative (ADNI) [62].

2.3.6 Proton density (PD-weighted)

Proton-density weighted sequences, or PD-weighted sequences, directly relate to the number of hydrogen nuclei present in the imaged body region. Since fat and fluids contain many protons, tissues with them appear bright in PD images [63].

Figure 2 compares different types of MRI modalities in two datasets, the longitudinal multiple sclerosis lesion dataset [64] and the BraTS dataset [65]. Furthermore, we include visualizations of multiple sclerosis tumors [66] in Fig. 3 to illustrate the distinctions between three different modalities used in the identification of multiple sclerosis biomarkers. From the visualization, it is evident that PD-weighted images play a crucial role in determining multiple sclerosis biomarkers. Despite the challenges in acquiring proton density data, PD-weighted images offer valuable insights into the distribution of hydrogen nuclei in the imaged tissue: they are useful for understanding certain pathological conditions, including multiple sclerosis, in clinical practice.

2.4 MRI artifacts

The MRI artifacts may be divided into three categories. The first category includes artifacts induced by movement, such as respiratory motion,

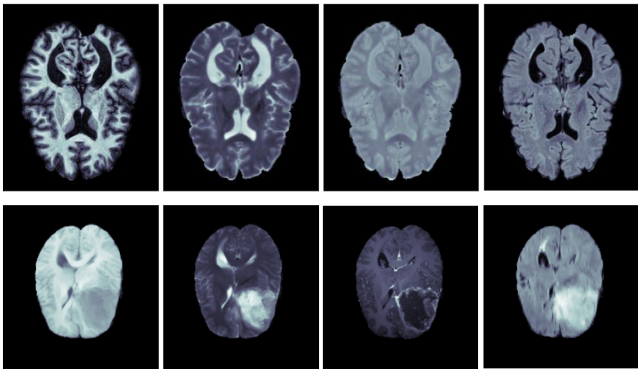

Fig. 2 Comparison of different MRI sequences. Left to right, above [64]: MP-RAGE, T2-w, PD-w, and FLAIR; below [65, 66]: T1, T2, T1c, and FLAIR. Each MRI modality reveals different characteristics of the soft tissue of the human brain: combining them provides clinical professionals with comprehensive information.

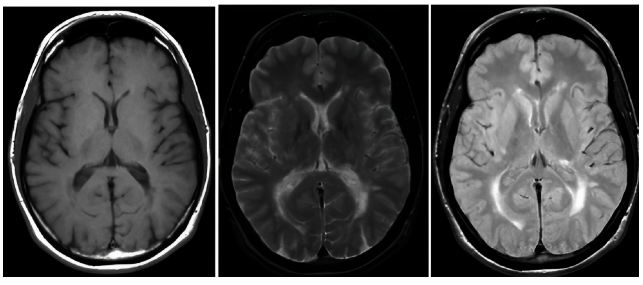

Fig. 3 Left to right: T1-weighted, T2-weighted, and proton density (PD)-weighted axial head scans of a patient with multiple sclerosis [66].

blood flow, and flow of cerebrospinal fluid in the brain and spinal canal. The second type of artifact can arise if the measuring technique or settings are not chosen precisely. Aliasing, chemical shift, phase cancellation, coherence, and magnetic susceptibility artifacts are examples of this kind. Finally, exogenous sources like magnetic field distortions, the hardware itself and noise may cause the third type of artifact [67]. It is worth noting that some of the aforementioned factors have a substantial impact on MRI scans and could result in one or more modalities being missing; this is known to be a common challenge in MRI. Please refer to Ref. [20] for more details of MRI artifacts.

Figure 4 depicts three of the most frequent MRI artifacts and how they impair MRI scans by reducing the visibility and detectability of a region of interest in the human brain in the obtained scans.

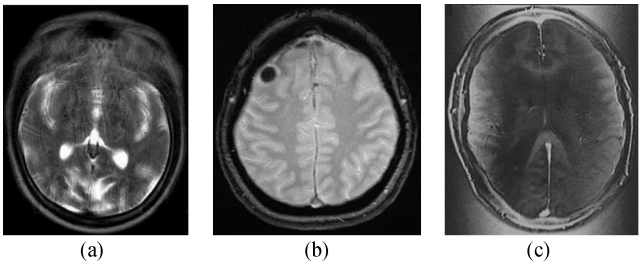

Fig. 4 The most frequent MRI artifacts: (a) caused by the patient's movement during an MRI scan of the brain [68], (b) magnetic susceptibility artifacts in an MRI scan [69], (c) RF overflow artifact due to MRI hardware [70].

2.5 Clinical importance

In addition to the aforementioned impact of missing MRI modalities on medical treatment, there are other factors that further emphasize the importance of addressing this issue from a clinical perspective. One such factor is the time-consuming and expensive process of gathering MRI images for patients. MRI scans are often associated with high costs and can be burdensome for patients, especially in resource-limited healthcare settings. Moreover, the acquisition of multiple MRI sequences for comprehensive analysis adds to the overall cost and duration of the diagnostic process.

Furthermore, the annotation burden in medical image analysis should not be overlooked. Manual annotation of MRI images, especially for complex structures like brain tumors, requires specialist expertise and considerable effort from radiologists. The process is labor-intensive, subject to human error, and may vary between experts, leading to inconsistent annotation.

The effect of missing modalities on the clinical decision-making process is significant. The absence of crucial information from one or more MRI sequences may result in incomplete assessments and potentially lead to misdiagnosis or suboptimal treatment plans. Radiologists heavily rely on comprehensive and detailed imaging data to accurately diagnose and monitor medical conditions. Studies show that human mistakes during the annotation process can deteriorate the radiologist's work by up to 21% in MRI images [12]. Therefore, any deficiency in the available data due to missing modalities can directly impact patient care and outcomes.

To address these challenges, the development of automatic algorithms that can effectively handle missing modalities in MRI images is of paramount importance. Such algorithms can provide a valuable and efficient solution for enhancing the diagnostic process by imputing the missing information and generating comprehensive images for analysis. By leveraging advanced techniques like deep learning and attention-based models, these algorithms have the potential to improve the accuracy and reliability of medical image analysis, ultimately benefiting patients, and healthcare providers.

3 Missing modality compensation approaches

As classified in Fig. 1, solutions utilized to overcome the missing modality problem can be divided into five main categories:

- Early synthesis methods.
- Common latent space models.
- Knowledge distillation networks.
- Mutual information maximization.
- Generative adversarial networks (GANs).

These methods usually perform operations in 3D space (our covered approaches) to address the missing modality problem in MRI images. In the following parts of Section 3, each of the five main categories will be discussed extensively, with numerous prominent networks introduced as examples and analyzed.

3.1 Early synthesis methods

Prior to the emergence of deep learning methods and their remarkable performance in medical image analysis, methods included image processing techniques, machine learning, or combinations thereof. This section discusses early attempts to overcome the problem of missing modalities, which included synthesizing them using traditional image processing techniques and machine learning. Generally, image synthesis refers to the process of generating an image which contains some specific content. In terms of missing modality, the synthesis approach tries to reconstruct the missing modality image based on the other available modalities [27].

Ref. [27] conducted experiments on two different classifiers: support vector machines (SVMs) and random forests (RFs), using two models, neural networks and restricted Boltzmann machines (RBMs), to synthesize the missing modality. The neural network used in Ref. [27] is a simple feed-forward network with just three layers that is able to predict a 3D patch. The second model used in Ref. [27] is an RBM. RBMs can be regarded as stochastic neural networks that learn critical characteristics of a probability distribution using relevant information from an unidentified probability distribution [71].

As shown by Ref. [27], inferring missing data at test time using a synthesis method, which is more adaptive than the classifier, may improve multimodal image segmentation by supplying data transformations for the classifier and also enlarging the training set. Through the use of synthesized data, random forests, a basic classifier presented in Ref. [27], exhibit improvements in segmentation outcomes.

Figure 5 depicts a model called REPLICA, which stands for regression ensembles with patch learning

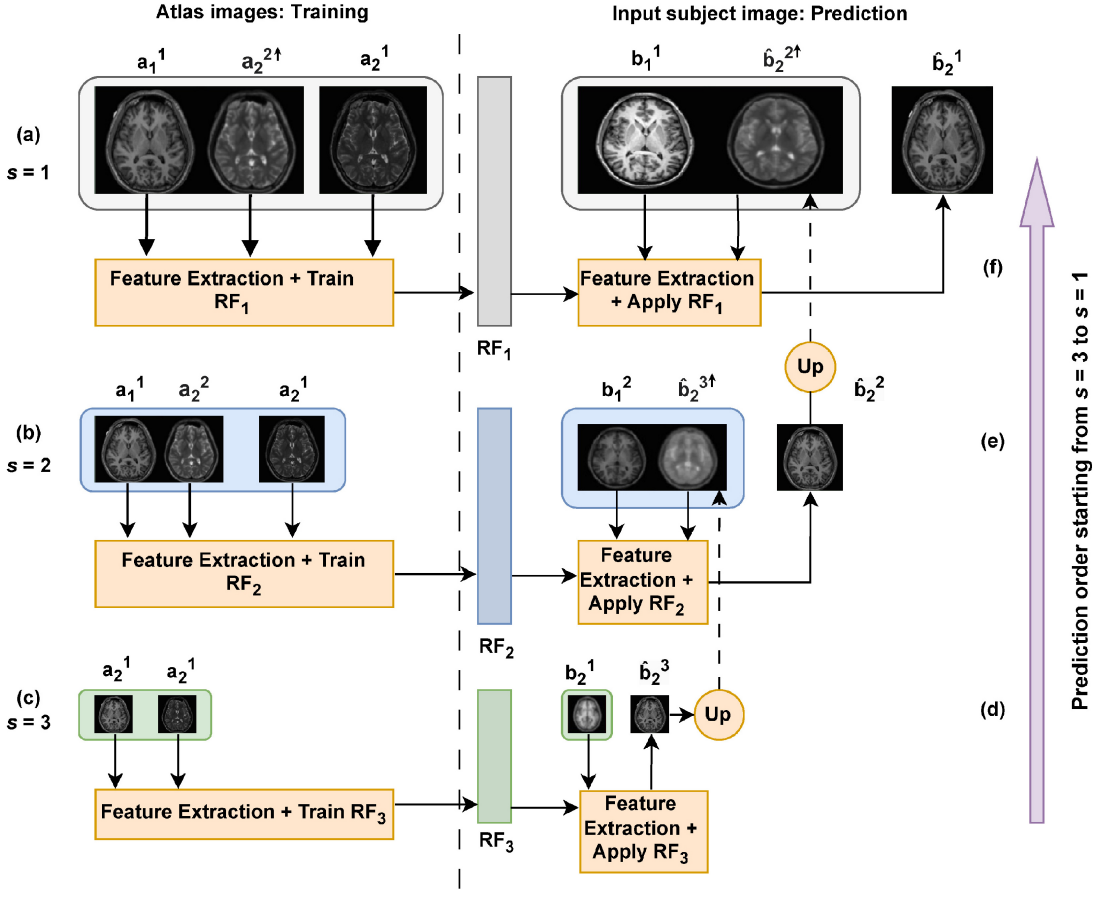

Fig. 5 The REPLICA algorithm [28] synthesizes T2w images from T1w images by first extracting a set of high-level features and then training a random forest at different scales. By matching a feature set at various scales, it synthesizes the T2w image.

for image contrast agreement [28]. REPLICA is a supervised random forest non-linear regression approach for synthesizing the missing modalities. It can synthesize T2 and FLAIR, which was previously thought to be a hurdle. REPLICA is structured to forecast tissue contrasts based on inputs with the same tissue contrast as the MRI image to be produced.

The REPLICA architecture predicts T2-weighted images from T1-w images. The training process is depicted in Fig. 5(left), and takes place at three different scales. Firstly, the most important features from the atlas image set are extracted at each scale, and the random forest is then trained to predict voxel-by-voxel intensity. After training, REPLICA aims to synthesize a missing modality, as follows: starting from the coarsest scale, the trained random forest at scale 3 is applied to the features extracted at the same scale, in order to synthesize the target MRI scan. With this synthesis at the lowest resolution, the features extracted from this scale are then up-sampled to the next level. The procedure proceeds to scale 1, which is the finest scale, with high resolution features, and then the trained random forest RF1 generates the final synthetic image.

Experiments conducted by Ref. [27] have demonstrated that in specific scenarios, replacing the absent modality with a synthesized or imputed sequence does not worsen outcomes, but it may also not lead to any improvement. For example, it was demonstrated that using a sequence generated by a three-layered feed forward neural network produced the same results as simply replacing it with zeros or performing the segmentation without it. Using more advanced vision transformer-based methods, Wang et al. [72] proposed a unified hybrid framework called SIFormer to synthesize MRI modalities in an end-to-end manner. SIFormer utilizes available low-resolution MRI contrasts to perform super-resolution of poor-quality MR images and impute missing sequences in a single forward process. The generator in SIFormer incorporates a dual branch attention block that combines the transformer's long-range dependency capabilities with the convolutional neural network's local information capture capabilities. Additionally, a learnable gating adaptation multi-layer perception is introduced to optimize information transmission efficiently.

3.2 Common latent space models

Adopting deep learning for biomedical image analysis was one of the significant steps for finding a viable strategy for dealing with missing modality issues in MR images. The objective of early deep learning methods for the missing modalities issue was to map modalities to a shared subspace and create a shared latent vector. Heteromodal image segmentation (HeMIS) [29] is a well-known example method that utilizes this concept. HeMIS consists of three main layers (see Fig. 6): a back-end layer, an abstraction layer, and a front-end layer. Each modality is directed into a specific set of convolutional layers in the network's back-end layer, which subsequently translate each modality into a common representation for all modalities. Arithmetic operations like mean and variance are computed in the abstraction layer. The mean and variance are then be combined and supplied to the front-end layer, which provides segmentation outputs.

Although establishing a common latent embedding for all available modalities is one of HeMIS' major goals, computing the mean and variance alone does not always suffice. Furthermore, HeMIS can only function properly in the absence of modalities if each modality input in the test set is labeled.

The authors of Ref. [42] were inspired by the HeMIS to create the permutation invariant multimodal segmentation (PIMMS) network, which in addition to missing modalities, tackles the issue of missing modality labels. PIMMS uses a classifier to build a distribution across modalities for the available inputs, then awards a score and labels each unlabeled item of input data. The inputs are then further adjusted by applying two different types of attention: soft and hard attention. The adjusted inputs are subsequently supplied to the second part of the network, which is an HeMIS model.

The regression-segmentation 3D CNN (RS-Net) model [43] is another method that builds a common representation of all modalities and synthesizes the missing modality. Three major blocks make up RS-Net. The first block is a 3D U-Net, very similar to the one presented in Ref. [73]. In this approach, the U-Net is fed with all of the present volumetric data, resulting in an intermediate latent representation of the data. The second block, the regression convolution block, synthesizes the missing modality using the latent

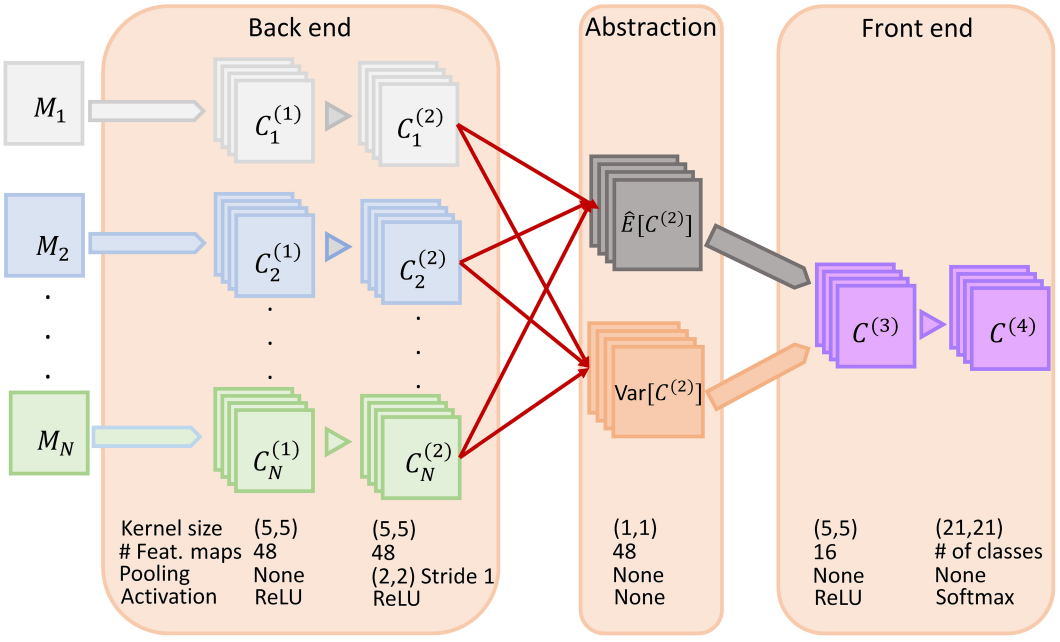

Fig. 6 The HeMIS architecture [29] applies a series of three connected blocks: a back end block to encode each modality into a latent space and to learn modality-specific features, an abstraction block to extract statistical features (first and second-order moments), and finally a front end block to generate the segmentation map based on the learned representation.

representation and also one of the existing volumetric data as input. The third block is the segmentation convolution block, which takes the produced latent representation as input and has multiple segmentation classes as outputs, each indicating a tumor subtype. The main shortcoming of RS-Net is that it results in errors while attempting to synthesize the T1c modality. Another method that utilizes a U-Net-based structure is introduced in Ref. [44]. This network uses four separate encoding paths to obtain initial feature maps for each MRI modality. Then, the final segmentation map of the missing modality is generated by combining the initial feature maps and fusing them with feature maps retrieved along the decoding path at multiple resolutions.

Subsequent attempts to create a shared latent space representation resulted in the development of the heteromodal variational encoder–decoder (HVED), which is a combination of the 3D U-Net and the multimodal variational auto-encoder (MVAE) [45]. The MVAE architecture contains four encoders, each of which separately computes variational parameters, or more specifically, mean and variance of each inference network, which are described as a Gaussian distribution after being merged [74], and forms a common subspace. Next, five decoders decode a randomly chosen latent variable from the common subspace. The first four decoders generate the desired modalities, while the fifth generates the segmentation map. In spite of the fact that HVED outflanks HeMIS and U-HeMIS (an HeMIS extension) it produces relatively inadequate results when more than one modality is lacking [46].

Calculating the first and second moments is not the only technique used to arrive at a shared latent representation. This aim may likewise be achieved using adversarial methods. In Ref. [46], a model referred to as the adversarial co-training network (ACN) is proposed. The ACN architecture utilizes a multimodal path with complete modalities and a unimodal path with the incomplete modality as inputs. Each path is trained individually and passes through a U-Net on its own. Then the joint learning process works by embedding an entropy adversarial learning module (EnA), a knowledge adversarial learning module (KnA), and a modality-mutual information knowledge transfer module (MMI) into the network's architecture. The segmentation maps created by each path are fed into the EnA module, which is located at the end of the networks. At each training epoch, the EnA acts as an adversarial discriminator, assisting the two networks to produce increasingly similar segmentation maps. The adversarial loss is calculated by the KnA module,

which, like EnA, assists the two networks to have more similar outputs. The MMI module's role is to compute the mean squared error (MSE) and prevent feature information loss for the path with the multimodal network.

The authors of Ref. [47] present RFNet, a feature fusion network. RFNet includes four encoders, each of which extracts features from a single modality. Then, in order to build a shared representation, a decoder that also shares the weights for the four modalities segments each modality individually. The retrieved features are then fused at different levels using a region-aware fusion model (RFM), and the produced fused representation is then segmented.

The method presented in Ref. [48] is a relatively simple feature fusion method. The unified representation network (URN) uses a U-Net to encode each modality independently, then uses a fusion module to combine the encoders' outputs. Following that, the newly formed unified representation is utilized to reconstruct and synthesize the missing modality. More recently, Knower et al. [49] introduced a novel approach for learning enhanced modality-agnostic representations by utilizing a meta-learning strategy (meta learner), even when only a fraction of patients have full modality data available. The proposed method enhances partial modality representations so that they resemble full modality representations through meta-training on partial modality data and meta-testing on a limited number of full modality samples. In addition, the method incorporates co-supervision through the introduction of an auxiliary adversarial learning branch. This branch involves a missing modality detector, which acts as a discriminator and learns to mimic the full modality setting.

In recent work by Liu et al. [75], a novel approach called multi-contrast multi-scale transformer (MMT) is proposed for missing data imputation in MRI images. The primary aim of MMT is to synthesize missing contrasts from existing ones, effectively addressing the challenges arising from varying contrast availability among patients. Unlike conventional CNN-based methods, MMT adopts a sequence-to-sequence learning framework, providing greater flexibility and interpretability. MMT consists of a multi-scale Transformer encoder and decoder, allowing it to handle any subset of input contrasts and generate the missing ones accurately. The proposed multi-contrast swin transformer blocks efficiently capture both intra- and inter-contrast dependencies, resulting in superior image synthesis performance. The interpretability of MMT is achieved through in-built attention maps in the decoder, providing insights into the significance of each input contrast in different regions. Extensive experiments conducted on large-scale multi-contrast MRI datasets demonstrate the superior performance of MMT compared to SOTA methods, both quantitatively and qualitatively. This validates the effectiveness and potential of the proposed approach in addressing missing data imputation challenges in MRI analysis.

3.3 Knowledge distillation networks

The authors of Ref. [76] presented *model compression* in 2006 as a novel strategy that allows simpler and faster models to learn from a larger, more sophisticated, and better performing one. The authors of Ref. [77] were later inspired by this approach; they recommended training a more intricate and larger model with also more accurate results, and then transfer the information gained from this model to a smaller model with the aim of enhancing its performance. The proposed approach, *knowledge distillation*, has subsequently been used in a variety of networks. This section outlines the methodology of two networks that employ knowledge distillation.

The hierarchical adversarial knowledge distillation network (HAD-Net) [31] uses the benefits of hierarchical adversarial training. HAD-Net is comprised of three sub-networks: a teacher network, a student network, and a hierarchical discriminator (HD). The teacher network is a 3D U-net [78], trained with full modalities. The same 3D U-Net is used to form the student network as done for the teacher network, however the T1c modality is not used during inference. At varying resolutions, the fully convolutional HD is in charge of distilling information from the teacher network into the student network. Thereafter, a mean squared error (MSE) adversarial loss is computed between the labels generated by the teacher network and the student network.

The authors of Ref. [32] introduced KDD-Net, a model that uses two teacher models, each of which is trained on all available data that includes samples with a complete set of modalities identified as X^{1c} and X^{2c}, as well as samples with missing modalities

identified as X^{1u} and X^{2u}, samples that only contain the first and second modality, respectively. The soft labels for the samples in X^{1c} and X^{2c} are then generated using these two unimodal teacher networks. The student model is a multimodal deep neural network that can learn a combined representation from the several modalities available. The student model is trained with the soft labels generated by the teacher networks and one hot label.

In other work, a style matching U-Net is proposed to overcome the problem of missing modalities. This approach builds its assumption on the idea that the feature representation in the latent space can be decomposed into style and content representations. It performs both style and content matching at different levels to distill the informative features from the full-modality path into a missing modality network. Similarly, Karimijafarbigloo et al. [33] proposed MMC-Former, a novel missing modality compensation network, which is designed to address in an end-to-end manner the challenge of missing information. Their approach leverages 3D efficient transformer blocks and employs a co-training strategy to effectively train a network that compensates for missing modalities. To ensure feature consistency across multiple scales, MMC-Former incorporates global contextual agreement modules at each encoder scale. Additionally, they proposed the use of auxiliary tokens in the bottleneck stage to capture the interaction between the full and missing-modality paths, thereby facilitating the transfer of modality-specific representations. Furthermore, their approach introduces feature consistency losses to minimize the domain gap in network prediction and enhance the prediction reliability for the missing modality path. The architecture of this approach is depicted in Fig. 7.

More recently, Liu et al. [50] proposed the M3AE framework. It consists of two stages. In the first stage, a multimodal masked auto-encoder (M3AE) is introduced. The M3AE uses self-supervised learning to train a robust multimodal representation by reconstructing random patches of the available modalities and applying modality dropout to simulate missing modalities. Additionally, a representative full-modal image is optimized using model inversion, which can be used as a substitute for the missing modalities during inferencing to improve segmentation performance. In the second stage, a memory-efficient self distillation approach is employed to transfer knowledge between different missing-modal situations. This process fine-tunes the model for supervised segmentation while incorporating information from heterogeneous missing modalities. The M3AE framework is designed to be versatile and applicable to any combination of modalities, making it cost-effective for both training and deployment. Experimental evaluations conducted on the BraTS 2018 and 2020 datasets demonstrate the superior performance of the proposed framework compared to other methods in scenarios with missing modalities.

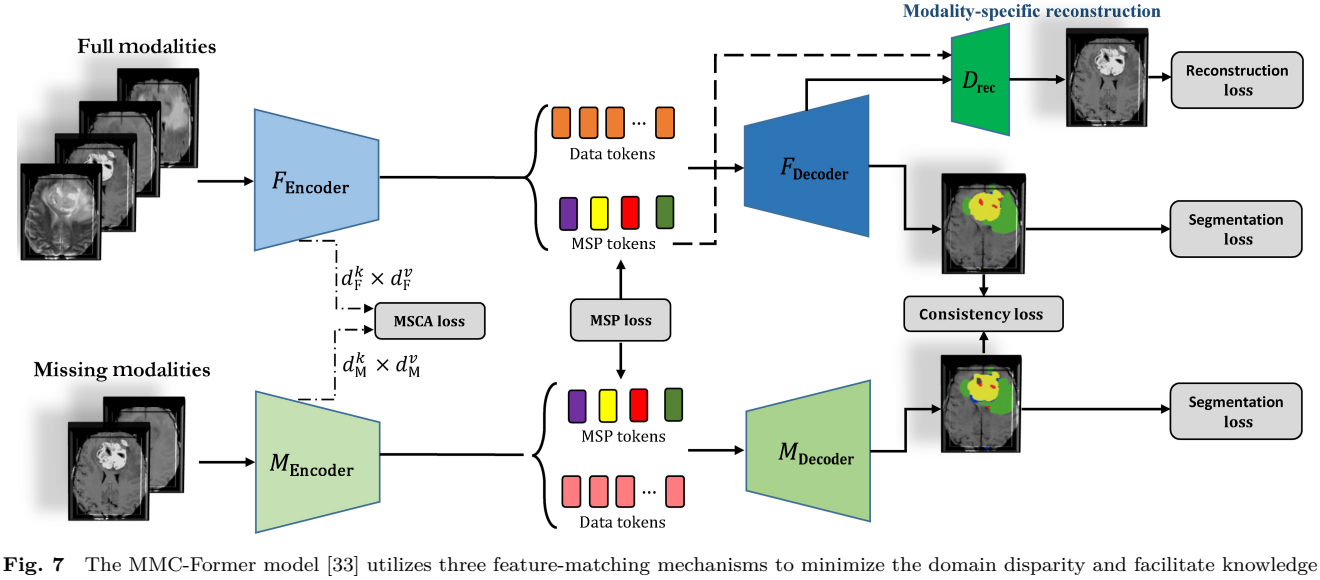

Fig. 7 The MMC-Former model [33] utilizes three feature-matching mechanisms to minimize the domain disparity and facilitate knowledge transfer from the full modality path to a network lacking certain modalities.

In a similar manner, Choi et al. [30] presented an innovative approach to address missing modality in MRI images through knowledge distillation. Their single-stage learning algorithm aims to improve brain tumor segmentation by utilizing information from the missing modalities. Unlike previous methods that employed a two-stage framework with pre-trained and student networks, where the latter were trained on limited image modalities, Choi et al. [30] proposed training both models concurrently using a single-stage knowledge distillation approach. The information transfer is achieved by reducing redundancy from a teacher network trained on full image modalities, to the student network through Barlow twins loss at the latent-space level. To perform pixel-level knowledge distillation, they also incorporate a deep supervision strategy, training the transformer-based backbone networks of both teacher and student paths using cross-entropy loss.

3.4 Mutual information maximization

In order to minimize information loss in the missing modality situation, the mutual information maximization strategy entails optimizing similarity metrics between available modalities during training.

Crossmodal information maximization for medical imaging (CMIM) [35] is one such network that utilizes the aforementioned strategy and maximizes mutual information between modalities instead of using a shared latent variable across all modalities. CMIM uses numerous modalities for training but just one modality at test time. The information about local and global features is then retrieved from the input images. For the semantic segmentation task, the local–local loss is computed, and for the classification task, both the local–local and local–global losses are calculated: see Fig. 8. The mutual information neural estimate approach (MINE) proposed in Ref. [79] is used by CMIM for mutual information estimation.

The CG-MCG network described in Ref. [51] takes advantage of the significant correlations between pairs of modalities. It comprises three sub-networks: the first is a conditional generator, or more precisely, a conditional U-Net with several encoders, capable of producing the missing modality in a more monitored manner by utilizing the index of the missing modality as the condition (indexes 0, 1, 2, 3 correspond to T2, T1c, FLAIR, and T1, respectively). The authors use a correlation constraint (CC) network

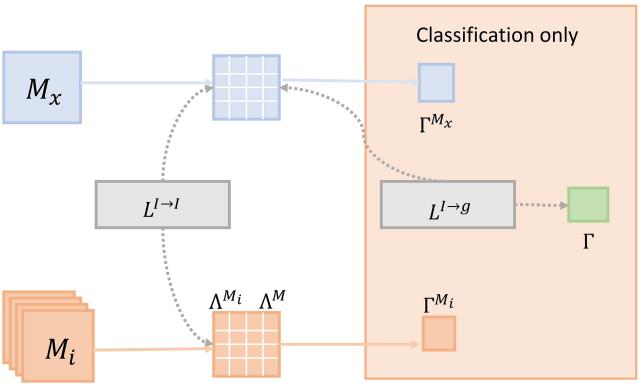

Fig. 8 CMIM architecture [35]. The model is trained on a set of modalities $M = \{M_x, M_1, ..., M_n\}$ but the only modality available at test time is M_x. Λ^{M_i} is the modality local feature for each modality, Γ^{M_i} is the modality global feature for each modality, Γ represents the multimodal global embedding shared over all available modalities, $L^{l\to l}$ is the cross-modal local-local loss and $L^{l\to g}$ is the cross-modal local-global loss.

as their second sub-network to compute the multi-source correlation, taking into consideration intensity distribution profiles and their correlations. The final segmentation is determined by the third sub-network, a segmentation network.

Similarly, Ref. [34] is a latent correlation representation learning method (LCE-Net) for addressing the missing modality problem. The latent correlation representations are created when each modality is encoded individually and provided to a model parameter estimation module (MPE Module), which is then fed to a linear correlation expression module (LCE Module). The correlation model is built by the MPE and LCE modules. The latent correlation representations then travel via the fusion block, resulting in a fused representation that spans all modalities. The fusion block employs a channel attention module and a spatial attention module. By decoding the fused representation, the network recreates the modalities and generates the segmentation map.

Recently, Zhou [52] proposed a framework that combines multimodal feature fusion (MMFF) and latent feature learning. The network serves two purposes: segmenting brain tumors when one or more modalities are missing and retrieving the missing modalities to compensate for incomplete data. The proposed network consists of three key components. First, a multimodal feature fusion module (MFFM) is introduced to effectively fuse complementary information from different modalities. This module

includes a cross-modality fusion module (CMFM) and a multi-scale fusion module (MSFM) to capture and integrate relevant information. Second, a spatial consistency-based latent feature learning module (SC-LFLM) is presented. This module exploits the correlations between multimodal latent features and extracts relevant features that contribute to segmentation accuracy. Finally, multi-task learning (MTL) paths are integrated into the network to supervise the segmentation task and facilitate the recovery of missing modalities.

In a different study, Dalmaz et al. [22] introduced ResViT, a novel generative adversarial approach for medical image synthesis. ResViT leverages the contextual sensitivity of vision transformers, the precision of convolution operators, and the realism of adversarial learning. The generator in ResViT incorporates a central bottleneck comprising novel aggregated residual transformer (ART) blocks, effectively combining residual convolutional and transformer modules. These ART blocks promote diverse representations through residual connections, while a channel compression module distills task-relevant information. To address computational challenges, the authors introduce a weight-sharing strategy among ART blocks, reducing the computational burden. Moreover, ResViT adopts a unified implementation, eliminating the need to construct separate synthesis models for different source-target modality configurations. Extensive evaluations on synthesizing missing sequences in multi-contrast MRI images demonstrate that ResViT surpasses competing CNN- and transformer-based methods in both qualitative observations and quantitative metrics, showcasing its superiority for medical image synthesis tasks.

3.5 Generative adversarial networks

Generative adversarial networks are a machine learning method initially presented in Ref. [80]. GANs comprise two networks: a generator G (that creates the missing modality in our case), and a discriminator D that determines whether a sample presented to it was created by the generator or is part of the original training data. Both the generator and the discriminator are trained simultaneously. This strategy increases the generator's performance and, as a consequence, it reconstructs a sample that contains and reflects the missing information.

The authors of Ref. [37] offered a generative adversarial network (GAN) modification that synthesizes the missing modality in a single forward pass, by just using one trained model, from the multiple inputs provided. In their approach they utilize the multimodal generative adversarial network (MM-GAN) that leverages implicit conditioning (IC) to enhance synthesis outcomes as follows: firstly, a U-Net as generator first imputes the missing modality, secondly, the $L1$ loss is determined for the scans produced by the generator, and thirdly the discriminator is a PatchGAN [39] with modality-selective $L2$ loss computation (a least squares GAN).

Using conditional generative adversarial networks (cGANs) to cope with the missing modality issue is another alternative. A cGAN is a traditional GAN extension that collects additional information in both generator and discriminator. The 3D cGAN proposed in Ref. [38] synthesizes the FLAIR sequence from only T1 MR images (see Fig. 9). However, due to their extensive training requirements, using GANs may not be the best approach.

Recently, Qian and Wang [55] proposed a framework called the contrastive masked-attention (COM) model. It employs cross-modal contrastive learning with GAN-based augmentation to reduce the modality gap caused by missing modalities. It also utilizes a masked-attention model to capture interactions between modalities effectively. The

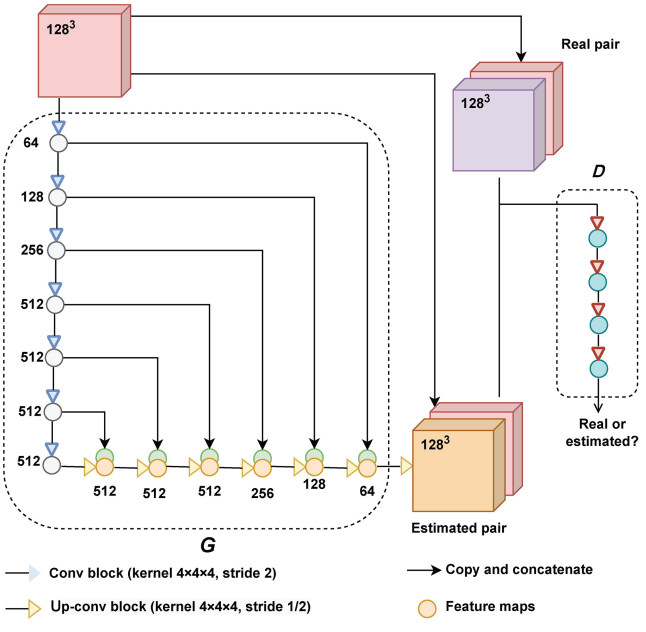

Fig. 9 3D conditional generative adversarial network [38].

augmentation strategy in COM adapts cross-modal contrastive learning to suit incomplete cases through a two-player game approach, enhancing the effectiveness of multimodal representations. Interactions between modalities are modeled using self-attention blocks, and attention masks are used to limit these interactions to the observed modalities, preventing the introduction of additional noise. A notable feature of COM is its flexibility to handle different missing patterns, as it utilizes a unified architecture that can accommodate various combinations of modalities.

In other work, Pan et al. [81] proposed a novel supervised deep learning approach for MR image synthesis. The method employs a transformer-based encoder and is specifically designed to translate 2D MR images from T1-weighted to T2-weighted modalities within a conditional GAN (cGAN) framework. Unlike traditional encoders, the authors utilize a transformer-based encoder and incorporate adjacent slices to enrich the model's spatial prior knowledge. By doing so, the proposed method aims to effectively synthesize missing modalities, enhancing the overall performance and accuracy of MR image synthesis. In the conducted experiments, the results demonstrate that the integration of a transformer-based encoder significantly facilitates the feature learning process and leads to notable improvements in performance.

Most of the analyzed method in this direction encounter challenges in handling multiple non-aligned imaging modalities and complex imaging sequences. To address this, DiamondGAN [54], a scalable and multimodal approach, was introduced. The model is capable of flexible non-aligned cross-modal synthesis and data infill, even when given multiple modalities or arbitrary subsets of them. It learns structured information in an end-to-end manner. Experiments on organ segmentation challenges synthesized two MRI sequences, namely double inversion recovery (DIR) and contrast-enhanced T1 (T1c), from three common sequences. The results were promising; trained radiologists are unable to distinguish synthetic DIR images from real ones, indicating the effectiveness of the proposed method.

3.6 Comparative overview

We now compare and contrast missing modality compensation approaches considered in Sections 3.1–3.5. Table 1 highlights their key techniques and describes the benefits and drawbacks of each approach to provide the reader a clearer picture. As detailed in Table 1, the synthesis approach performs compensation based on image reconstruction from an atlas sample. In practice, these approaches usually fail to reconstruct the missing information and result in no performance gain. On the other hand, common latent space representation methods perform information retrieval by modeling joint information from all modalities. However, these methods do not perform well when more than one modality is missing.

In the third category, knowledge distillation methods use the strength of the co-training approach and distill the informative features from a full-modality path into a missing modality network. Although knowledge distillation can encourage feature learning by the student network, important domain knowledge from the full-modal network is usually not gained by the student model. Hence, mutual information matching seems to be a necessary factor to include in the matching pipeline.

To this end, the fourth direction uses mutual information maximization algorithms. This approach calculates similarity metrics across available modalities and optimizes the mutual information. Regardless of the strong matching gain that can be derived from these methods, when insufficient modalities are available, this strategy will not necessarily assure that the lost data is recovered if there are not enough features to reconstruct the missing data.

In the last strategy, GAN methods are utilized along with segmentation networks to compensate for the missing information. It should be noted that the extensive training costs and noisy synthesis results are natural weaknesses of GAN methods, and can result in unstable reconstruction.

All in all, the choice of strategy for designing the network should accord with the clinical application and case of study. For a more robust network, the strengths of different directions should be unified into a single network. Such an approach is proposed in Ref. [18], where the author uses the knowledge distillation approach along with information maximization and adversarial losses. Figure 10 provides a timeline since 2016 for popular deep learning approaches for semantic segmentation

Table 1 Comparative overview of networks that compensate for missing modalities

Strategy	Approach	Novelty	Weakness
Early synthesis model	Why does synthesized data improve multi-sequence classification [27] REPLICA [28] MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration [41]	Using a more adaptable synthesis method such as an NN or RBM may result in performance enhancement Uses atlas registration methods	Results in no improvement when utilizing a classifier in a poorly adjustable model framework [27] The majority of these models do not alter results for downstream tasks such as segmentation [48] When using uniform atlases derived from healthy persons for glioma patients, distortion occurs [38]
Common latent space models	HeMIS [29] PIMSS [42] RS-Net [43] Brain Tumor Segmentation on MRI with Missing Modalities [44] HVED [45] Anatomy-Regularized Representation Learning for Crossmodality Medical Image Segmentation [82] ACN [46] RFNet [47] URN [48]	Maps the available modalities into a common latent subspace and aims to recover the missing information using the newly built latent representation	Unable to adequately recover the lost information by using methods such as computing first and second moments When more than one modality is lacking, many of these networks are inadequate [46] They usually fail to be resilient to missing modalities while also delivering accurate segmentation [37]
Knowledge distillation networks	HAD-Net [31] KDD-Net [32] Knowledge distillation from multimodal to monomodal segmentation networks [83] SMU-Net [18] MMC-Former [33] M3AE [50]	Transfers discriminative information from one or more teacher networks to a student network for recovering the missing data	Important domain knowledge from the full-modal network is usually not gained by the student model [46] Significant training costs when working with complex and large teacher networks [84] Capacity mismatch between teacher and student may result in no improvement in student network's outcomes [85]
Mutual information maximization	CMIM [35] Conditional generator and multi-source correlation guided brain tumor segmentation with missing MRI modalities [51] Latent correlation representation learning for brain tumor segmentation with missing MRI modalities [34] Brain graph synthesis by dual adversarial domain alignment and target graph prediction from a source graph [86]	Calculates similarity metrics across available modalities and optimizes their mutual information	When too few modalities are available, this strategy does not necessarily ensure that the lost data is recovered, as there may not be enough features to reconstruct the missing data [47] Some earlier models tended to confine the network's structure [35]
Generative adversarial networks (GANs)	MM-GAN [37] 3D Conditional generative adversarial network [38] Multimodal AsynDGAN: Learn from distributed medical image data without sharing private information [87] Auto-GAN [53] DiamondGAN [54] CoCa-GAN [88] COM [55]	Employs GAN and its modifications in the missing modality model framework	May generate undesirable imputation noise, when imputing or synthesizing the missing modality [32] GANs may exhibit a non-convergent nature [89] Extensive training costs [90] The generator may become unstable [89]

of MRI images with missing modalities. The timeline reveals increasing attention to the missing modality challenge in recent years due to the applicability of such approaches to clinical applications.

4 Datasets

In this section, we summarize the most common MRI datasets used for the task of semantic segmentation. These datasets include pixel-wise annotation to evaluate model performance. Some articles use data augmentation to increase the amount of annotated data, especially when dealing with small amounts of annotated data. Data augmentation increases the amount of training data by directly applying various transformations to images, feature space, or both. Some typical examples of these transformations include rotation, translation, scale, color jittering,

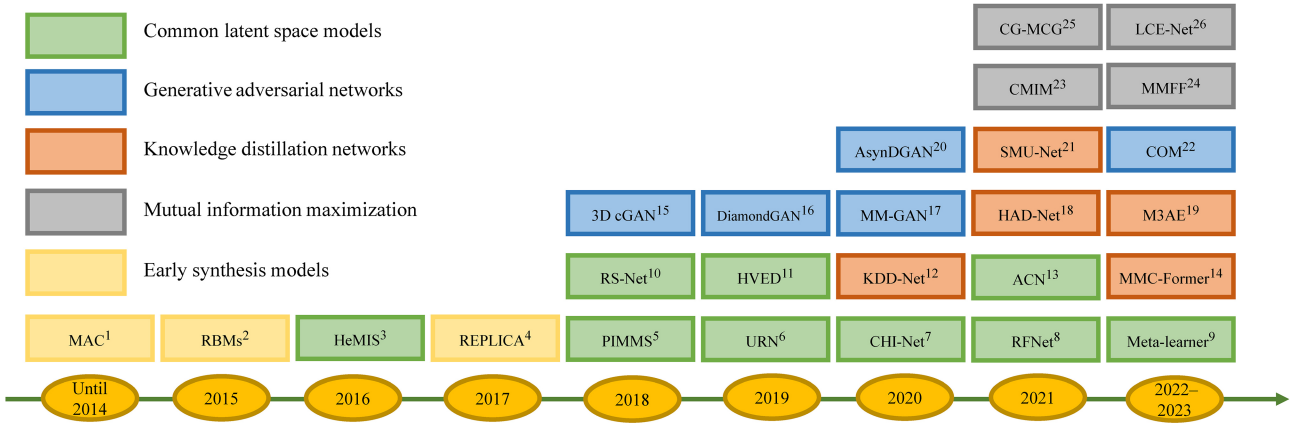

Fig. 10 The timeline of the deep learning methods proposed to compensate missing modalities in MRI-based semantic segmentation, from 2014 to 2023. The superscripts in ascending order represent: 1. [27], 2. [28], 3. [29], 4. [28], 5. [42], 6. [48], 7. [44], 8. [45], 9. [49], 10. [43], 11. [45], 12. [32], 13. [46], 14. [33], 15. [38], 16. [45], 17. [37], 18. [35], 19. [51], 20. [34], 21. [35], 22. [37], 23. [38], 24. [52], 25. [51], 26. [55].

cutting, and warping. In medical imaging, we usually have few images, and data augmentation helps to better train models. Other benefits of data enhancement include that it prevents overfitting, and leads to better generalization as well as faster convergence.

4.1 BraTS

The multimodal brain tumor image segmentation benchmark (BraTS) [65, 91] is a freely accessible dataset that comprises manually segmented images provided by clinical specialists from various institutes. It was initially released in 2012 and has since then been extensively used to improve automatic segmentation methods. The BraTS dataset concentrates on gliomas, a heterogeneous group of representing one of the most frequent kinds of primary brain tumor. T1-weighted, contrast-enhanced T1 weighted (also known as T1c-weighted), T2-weighted, and fluid attenuation inversion recovery (FLAIR) images provide the four kinds of MRI sequences included in the BraTS dataset, each of which reveals distinct characteristics of brain tumors. For evaluation, different versions of the BraTS dataset have been used by the literature. Each version of this dataset includes MRIs of several patients with four modalities (T1, T1c, T2, FLAIR). Each image's ground truth segmentation in the BraTS dataset includes labelling for four tissue classes: necrosis, edema, non-enhancing tumor, and enhancing tumor. Despite the fact that four distinct tumor labels are provided, they could well be categorized into three subregions for evaluation: the whole tumor (WT), the core tumor (CT), and the enhancing tumor (ET). Please refer to Table 2 later for more details of each version of the BraTS dataset. An example from the BraTS dataset is depicted in Fig. 11.

4.2 MSGC

The multiple sclerosis grand challenge (MSGC) [92] dataset consists of MR scans of 43 subjects from Boston Children's Hospital (CHB) and the University of North Carolina (UNC) with T1, T2, FLAIR, diffusion tensor imaging (DTI), and mean diffusivity (MD) images. There are 20 training images with manual ground truth lesion segmentation and a

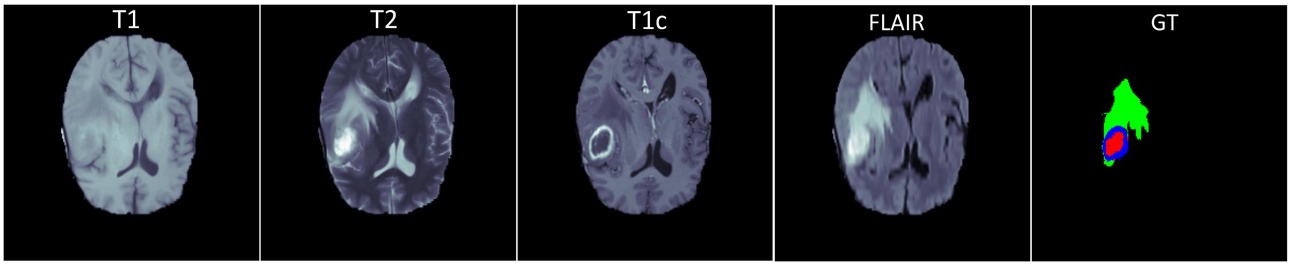

Fig. 11 Example from BraTS dataset [91] showing various modalities and the ground-truth annotation mask in which blue indicates GD-enhancing tumor, green represents peritumoral edema, and red is the tumor core.

further 23 testing images for which lesions masks are unavailable. The segmentation network output is evaluated using an automated system.

4.3 RRMS

The relapsing-remitting multiple sclerosis (RRMS) [29] dataset contains MRI sagittal plane scans of 300 RRMS patients, in T1, T2, and T1c modalities.

4.4 ADNI

The Alzheimer's disease neuroimaging initiative (ADNI) [42] database comprises MRI scans of 973 Alzheimer's patients, including T1 and FLAIR sequences, captured using scanners from only three manufacturers: GE, Philips, and Siemens. It is also relevant to note that ADNI suggests unified and specific criteria for the scans it contains, resulting in a lack of diversity across its contained scans.

4.5 SABRE

T1, T2, and FLAIR modalities are included in the Southall and Brent Revisited (SABRE) [93] dataset, which represents two longitudinal cohorts, one with low variation across images obtained from 586 participants and the other with high variation across images received from 1263 patients.

4.6 WMH

The MICCAI 2017 white Matter hyperintensity (WMH) [94] dataset includes 60 sets of brain MRIs, encompassing T1, and FLAIR modalities, with manual WMH annotations from three different institutes.

4.7 ISLES2015

The Ischemic Stroke Lesion Segmentation Challenge 2015 (ISLES2015) [95] dataset provides multi-spectral MRIs of stroke lesions in two different settings: sub-acute stroke lesion segmentation (SISS) and stroke perfusion estimation (SPES). The MRI sequences are stripped of their skulls, strictly co-registered with the FLAIR (SISS) and T1 (SPES) sequences, and re-sampled to a precise isotropic spacing for each setting.

4.8 CHAOS2019

The combined healthy abdominal organ segmentation (CHAOS) benchmark [96], comprises two databases, one containing CT scans and the other MRI images. The latter is of most interest in this paper. The MRI database comprises 120 Digital Imaging and Communications in Medicine (DICOM) datasets, including 40 T1-DUAL in phase datasets, 40 T2-SPIR datasets, and 40 T1-DUAL out phase datasets, all of which were obtained utilizing different RF Pulse and gradient combinations. All MRI scans were acquired with a 1.5T Philips machine.

4.9 MS Lesion

The MS Lesion dataset [97] includes T1, T2, FLAIR, double inversion recovery (DIR), and T1c modalities for 65 pathological brain MRIs of multiple sclerosis (MS) lesion patients.

4.10 IXI dataset

The IXI dataset [98] comprises almost 600 MR scans from healthy subjects, including T1, T2, and PD sequences as well as diffusion-weighted (DW) sequences, taken on two different vendors' systems: Philips 1.5 and 3T systems as well as a GE 1.5T system.

4.11 Duke liver dataset

The duke liver dataset [99] is a specialized collection of abdominal MRI scans from 105 patients, featuring 2146 image series and 113,280 unique images. The dataset, captured by Siemens and GE machines at 1.5T and 3T field strengths, includes a variety of series types, with liver segmentation performed on specific axial MR image series such as axial portal venous phase fat-suppressed T1-weighted and axial precontrast fat-suppressed T1-weighted. With a patients age range of 30–80 years, this dataset serves as a valuable resource for focused studies and advances in liver segmentation and abdominal radiology.

Table 2 summarizes the aforementioned datasets along with their numbers of samples, modalities and research utilizing these benchmarks.

5 Performance review

In this section, we undertake a case study focusing on the segmentation task to evaluate the effectiveness of various approaches in addressing the challenges posed by missing modalities in MRI images. To this end, we begin by presenting a summary of popular metrics commonly used to evaluate medical image segmentation networks. Subsequently, we discuss the quantitative performance of recent methodologies applied to the segmentation of medical images,

Table 2 Datasets commonly utilized in the literature to evaluate the performance of missing modality compensation networks

Dataset	Number of samples	Modalities	Articles
Multiple sclerosis grand challenge (MSGC) [92]	43 subjects: 20 training, 23 testing	T1, T2, FLAIR	[29]
Relapsing remitting multiple sclerosis (RRMS) [29]	300 subjects	T1, T2, T1C	[29]
BRATS2013 [65]	30 subjects: 20 high grade and 10 high grade	T1, T1C, T2, FLAIR	[27]
BraTS2015 [65]	274 subjects: 220 high grade and 54 low grade tumors	T1, T1C, T2, FLAIR	[22, 29, 35, 38, 43, 47, 100, 101]
BraTS2017 [65]	285 subjects: 210 high grade and 75 low grade tumors	T1, T1C, T2, FLAIR	[22, 43, 44, 102]
BraTS2018 [91]	285 subjects: 210 high grade and 75 low grade tumors	T1, T1C, T2, FLAIR	[23, 25, 31, 34, 37, 45–48, 51, 53, 87, 103]
BraTS2019 [91]	335 subjects: 259 high grade and 76 low grade tumors	T1, T1C, T2, FLAIR	[21, 22, 31, 34]
BraTS2020 [91]	369 subjects	T1, T1C, T2, FLAIR	[47]
Alzheimer's disease neuroimaging initiative (ADNI) [104]	973 subjects	T1, FLAIR	[32, 42]
SABRE [93]	586 subjects with same scanner, 1263 subjects with multiple scanners and multiple settings	T1, T2, FLAIR	[42]
MICCAI-WMH dataset [94]	60 subjects	T1, FLAIR	[42, 54, 105]
Ischemic stroke lesion segmentation challenge 2015 (ISLES2015) [95]	(1) SISS: 28 training and 36 testing (2) SPESS: 30 training and 20 testing	(1) FLAIR, T2 TSE, T1 TFE/ TSE, DWI; (2) T1C, T2, DWI, CBF, CBV, TTP, Tmax	[37]
Combined healthy organ segmentation (CHAOS2019) [96]	—	T1, T2	[106]
Brain tumor segmentation in medical segmentation decathlon [107]	750 subjects: 484 training, 266 testing	T1, T1C, T2, FLAIR	[106]
MS Lesions [97]	65 subjects	Flair, T1, T2, double inversion recovery (DIR), T1C	[54]
IXI dataset [98]	600 subjects	T1 , T2, PD	[98]
Duke liver dataset [98]	105 subjects	Axial portal venous phase fat-suppressed T1-weighted, axial precontrast fat-suppressed T1-weighted, axial in-phase, axial opposed-phase	[99, 108]

particularly when faced with the issue of missing modalities.

5.1 Performance metrics

Most articles in recent years have focused only on the issue of quantitative accuracy of the model and compare and report the performance of their model in terms of quantitative accuracy. They fail to include other important aspects such as speed (inference time) and the amount of memory required (which we will discuss in Section 6). In this section, we briefly introduce the most common evaluation metric used for evaluating the segmentation results of the missing modality compensating networks.

5.1.1 Dice score

In semantic segmentation, the Dice loss, based on the Dice coefficient of similarity, is well-known. In the segmentation of medical images, the region of interest (ROI) is typically a small part of the image. Therefore, models are prone to becoming trapped in a local minimum during training. Accordingly, the model will be biased to the background, so the object of interest will not be detected appropriately and so will often miss items of interest. The Dice loss was proposed to alleviate this problem [109]. The Dice loss is formulated for a 3D MRI image as

$$\boldsymbol{L}_{\text{Dice}} = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon} \tag{1}$$

where N is the number of voxels, p_i is the predicted binary segmentation volume, and g_i is the ground truth binary volume. The term ϵ is introduced to prevent division-by-zero causing issues.

5.2 Quantitative performance analysis

In this section, we report how well the reviewed approaches perform on the most commonly used benchmark (BraTS2018 dataset). To ensure a fair comparison, we only include methods on a particular dataset that use the same settings. To give a clear picture of the overall performance gained by the selected methods, we also provide experimental results with the extreme missing modalities scenario in the next subsection.

Table 3 shows two baselines along with four well-known approaches for missing modality compensation that were compared using BraTS2018 with the same setting. During inferencing, HeMIS and HVED approaches are significantly affected when there is a decrease in the number of available modalities, but this is not the case for SMU-Net, ACN, M3AE, and MMC-Former, which are able to compensate for missing modalities. Knowledge distillation, used in both the SMU-Net and MMC-Former methods, improved the network's robustness to missing modalities and its performance over classic latent space approaches. Furthermore, SMU-Net introduces a novel framework that involves the decomposition of the representational space into style and content vectors. This decomposition enables the minimization of the distribution disparity between the full and missing modality paths, thereby facilitating knowledge distillation. Similarly, ACN adopts a similar perspective and conducts knowledge distillation by considering both style and content modules without explicitly separating them. However, both approaches lack a mechanism to effectively preserve modality-specific features and distill such representations to the "missing modality" network. In contrast, MMC-Former incorporates MSP tokens along with a reconstruction head to ensure the preservation of such modality-specific information and achieves improved performance. Moreover, two crucial factors are believed to play pivotal roles in this context. Firstly, the co-training strategies employed in SMU-Net and ACN solely focus on modeling the one-to-one correlation between complete modalities and individual missing-modality scenarios. Conversely, M3AE and MMC-Former approaches implicitly capture the various correlations in all heterogeneous missing-modality situations. Secondly, the random modality dropout and patch masking techniques utilized in M3AE and the reconstruction head included in MMC-Former are likely to serve as effective data reconstruction methods that aid in model training. These techniques are not available in paired co-training (ACN and SMU-Net), where the absent modalities remain fixed.

It is also worthwhile to mention a limitation of ACN, SMU-Net, and MMC-Former, which is the

Table 3 Performance comparison on the BraTS2018 dataset using the Dice metric. HeM, HVE, MMC, and M3AE indicate the HeMIS [29], HVED [45], MMC-Former [33], and M3AE [50] models, respectively. In addition, SMU-Net [18] and ACN [46] performance are reported

Modalities				Complete						Core						Enhancing					
Flair	T1	T1c	T2	HeM	HVE	ACN	SMU	MAE	MMC	HeM	HVE	ACN	**SMU**	MAE	MMC	HeM	HVE	ACN	SMU	MAE	MMC
○	○	○	●	79.2	80.9	85.4	**85.7**	84.8	84.1	50.5	54.1	66.8	67.2	69.4	**69.7**	23.3	30.8	41.7	43.1	47.6	**50.7**
○	○	●	○	58.5	62.4	79.8	80.3	75.8	**80.4**	58.5	66.7	83.3	84.1	82.9	**86.6**	60.8	65.5	78.0	78.3	73.7	**79.0**
○	●	○	○	54.3	52.4	78.7	**78.6**	74.4	78.6	37.9	37.2	**70.9**	69.5	66.1	69.3	12.4	13.7	41.8	42.8	37.1	**42.9**
●	○	○	○	79.9	82.1	87.3	**87.5**	88.7	86.2	49.8	50.4	66.4	**71.8**	66.4	70.0	24.9	24.8	42.2	46.1	35.6	**49.6**
○	○	●	●	81.0	82.7	84.9	86.1	86.3	**86.3**	69.1	73.7	83.2	85.0	84.2	**86.7**	68.6	70.2	74.9	75.7	75.3	**79.1**
○	●	●	○	63.8	66.8	79.6	**80.3**	77.2	80.2	64.0	69.7	83.9	84.4	83.4	**86.6**	65.3	67.0	75.3	75.1	74.7	**78.7**
●	●	○	○	83.9	84.3	86.0	87.3	**89.0**	87.7	56.7	55.3	70.4	71.2	70.8	**72.0**	29.0	24.2	42.5	44.0	41.2	**48.1**
○	●	○	●	80.8	82.2	84.4	85.6	**86.7**	85.9	53.4	57.2	72.8	73.5	71.8	**74.0**	28.3	30.7	46.5	47.7	**48.7**	46.9
●	○	○	●	86.0	87.5	86.9	87.9	**89.9**	88.0	58.7	59.7	70.7	71.2	70.9	**72.4**	28.0	34.6	44.3	46.0	45.4	**48.3**
●	○	●	○	83.3	85.5	87.8	88.4	**89.7**	88.7	67.6	72.9	82.9	84.1	84.4	**86.7**	68.0	70.3	77.5	77.3	75.0	**79.5**
●	●	●	○	85.1	86.2	88.4	88.2	**88.9**	88.3	70.7	74.2	83.3	84.2	84.1	**86.6**	69.9	71.1	75.1	76.2	74.0	**78.1**
●	●	○	●	87.0	88.0	87.4	88.3	**90.2**	88.5	61.0	61.5	67.7	67.9	**72.7**	67.8	33.4	34.1	42.8	43.1	44.8	**50.7**
●	○	●	●	87.0	88.6	87.2	88.2	85.7	**88.9**	72.2	75.6	82.9	82.5	84.6	**86.6**	69.7	71.2	73.8	75.4	73.8	**79.1**
○	●	●	●	82.1	83.3	86.6	86.5	**90.1**	86.8	70.7	75.3	83.2	84.4	84.4	**86.7**	69.7	71.1	75.9	76.2	75.4	**79.2**
●	●	●	●	87.6	88.8	89.1	88.9	85.1	**89.0**	73.4	76.4	84.8	87.3	84.5	**87.4**	70.8	71.7	78.2	79.3	75.5	**80.1**
Mean				78.6	80.1	85.3	**85.9**	85.8	**85.9**	59.7	64.0	76.8	77.9	77.4	**79.2**	48.1	50.1	60.70	61.8	59.9	**64.7**

pre-defined missing modality scenario, unlike for the M3AE, HeMIS, and HVED methods. Like the ACN and SMU-Net methods, MMC-Former trains the network on each possible combination of modalities to compensate for the given missing modality scenario, while HeMIS and HVED try to handle this issue inside the network without training separate networks for each possible combination.

5.3 Extreme missing modalities

We next analyze the performance of several algorithms in cases with extreme missing modalities. More precisely we assume that during training, all modalities are presented, while inferencing only uses a single modality of data. This extreme scenario provides a strong benchmark to evaluate the effectiveness of different approaches to tackle the problem of missing information. We followed the approach outlined in Azad et al. [18] to design the extreme missing modality experiment. Table 4 provides experimental results on the BraTS2018 dataset. Here, for each modality, we report the average Dice score as the average of the whole, enhanced and core tumor Dice scores. The quantitative results clearly demonstrate that the MMC-Former model outperforms the U-Net baseline method by a significant margin. Additionally, both ACN and MMC-Former approaches achieve competitive results across all single modalities compared to other state-of-the-art (SOTA) methods. Specifically, for the T1c and T2 modalities, MMC-Former exhibits a substantial improvement over ACN, M3AE, and SMU-Net models. Similarly, for the Flair modality, MMC-Former outperforms all methods. In the case of the T1 modality, ACN demonstrates a significant improvement over the baseline and the MMC-Former method. These results validate the effectiveness of these approaches for learning task-specific patterns and recovering missing information.

To further validate the efficacy of incorporating a knowledge distillation strategy for performance enhancement and retrieval of missing information, we conducted a comparative evaluation between missing modality-specific networks with the well-established nnU-Net approach, as described by Isensee et al. [110]. The results, also presented in Table 4, clearly indicate that the knowledge distillation strategy yields significant performance improvements compared to conventional single-modality-based methods.

Table 4 Quantitative results for various models on the BraTS 2018 benchmark in the extreme missing modality scenario. We use the average of the whole, enhanced and core tumor segmentation scores to report the Dice score for each modality

Model	Dice score				
	T1	T1c	T2	FLAIR	AVG
U-HeMIS [29]	16.7	59.2	36.0	51.5	48.8
HVED [45]	34.4	64.8	55.2	52.4	51.7
ACN [46]	**63.8**	80.3	64.6	65.3	68.5
SMU-Net [18]	63.3	80.9	65.3	68.4	69.4
nnU-Net [110]	62.4	77.6	61.3	61.8	65.7
M3AE [50]	59.2	77.4	67.2	63.5	66.8
MMC-Former [33]	63.6	**82.0**	**68.2**	**68.6**	**70.6**

6 Challenges and opportunities

In recent years, promising deep-learning-based methods have been introduced to equip medical imaging with missing modalities. Here some perspectives on future research will be introduced that can further improve the methods of medical image segmentation with missing modalities.

6.1 More challenging datasets

Section 4 introduces the most popular MRI datasets for semantic segmentation. The BraTS dataset, for instance, contains a large number of 3D MRI sequences but appears to lack image format variations. This lack of variety may lead us to the conclusion that more challenging datasets, more representative of real-world situations, are needed to enhance the training process and lead models to provide better results.

6.2 Memory efficient models

The majority of the approaches discussed in this survey are primarily concerned with offsetting the negative consequences of operating with an incomplete set of MRI modalities and, as a result, increasing segmentation accuracy. However, these approaches often need a large amount of memory, not only during the training phase but also throughout inferencing. Knowledge distillation networks, as discussed in Section 3, are one of the key techniques for the issue at hand, and they may be used to transfer knowledge from a larger, more complicated model to a smaller, less memory-intensive one. A simplified network can be achieved using knowledge distillation [77] or network compression techniques [111], which can then be deployed in other devices such as smartphones.

6.3 Balance between accuracy and efficiency

In machine learning and deep learning models, there is a well-known trade-off between accuracy and efficiency, and semantic segmentation networks are no exception [112]. It is frequently the case that models that are capable of producing more accurate outcomes are less efficient. This is also truly the case in the inverse scenario, implying that more efficient models are less accurate [113]. Future work should consider this fact in their design.

6.4 Model complexity

As noted previously, few articles report information such as computational complexity, run time, and memory footprint, which are important for clinics that may have limited computing resources [114, 115]. Besides the lack of memory in some devices, inferencing time plays a critical role in some real-time applications. Thus, model complexity needs to be taken into account in such cases. The numbers of model parameters and floating-point operations (FLOPs), runtime, and frames per second (FPS) are all commonly used metrics to assess model complexity. The first two metrics are notably independent of the implementation, with larger values indicating lower efficiency. Because of their reliance on hardware and implementation environment, runtime and FPS are less useful metrics for comparing implementation speeds. For a real-world application, however, these measurements need to be considered.

6.5 Interpretable models

From a clinical perspective, it is highly desirable to understand how the deep learning method learns certain patterns to detect diseases in medical images. This fact can help the radiologist to understand the deep model and possibly model the pathology assumptions in deep network design. There have been several approaches proposed in the literature to visualize and depict feature maps learned by deep models [116, 117]. However, these feature maps are usually not interpretable by radiologists. Thus, potential opportunities exist for designing such methods to characterize the underlying assumptions deep models use, and incorporate radiologist feedback in the network design.

6.6 Software ecosystems

To prototype and compare architectures for dealing with missing modalities, a deep learning software ecosystem is critical. A few already incorporate some of the architectures discussed in this review. For example, ivadomed [118] includes the HeMIS framework and other Pytorch-based architectures. Other relevant libraries include MONAI [119], MIC-DKFZ [120], and DLTK [121].

6.7 Diffusion models

Diffusion models [122] have emerged as a valuable and promising approach for generating missing modality images in the medical domain. Dealing with incomplete medical data can be challenging, but diffusion models offer an effective solution by imputing missing modalities based on available information. These models are trained on a dataset of complete multimodal images, enabling them to learn the underlying joint distribution of the modalities. When presented with samples containing missing modalities, the diffusion model can utilize observed data to generate synthetic completions for the missing parts. This capability is especially valuable in medical imaging scenarios where acquiring certain modalities might be difficult or pose potential risks to patients. The ability of diffusion models to generate missing modality images can significantly aid clinicians in making more informed decisions, enhance research by augmenting datasets, and ultimately lead to improved patient care. For instance, CoLa-Diff [123] introduces a novel conditional multimodal MRI synthesis model that utilizes latent diffusion models and employs brain region structures to guide synthesis. Additionally, Meng et al. [124] proposed UMM-CSGM, a diffusion-based conditional SDE model, which synthesizes missing modalities based on all remaining modalities as conditions.

Moreover, diffusion models demonstrate remarkable proficiency in learning complex data distributions. In the context of multimodal MRI synthesis, where the relationships between different modalities are intricate, diffusion models excel in capturing the underlying joint distribution of the modalities. Their ability to generate high-quality, realistic images from noise is especially crucial in multimodal MRI synthesis, where the goal is to produce accurate and visually consistent images that are also anatomically plausible. By leveraging priors such as brain region masks, diffusion models can better maintain the anatomical structure in MRI

images, which is vital in medical imaging applications for accurate diagnosis and interpretation by clinicians. Experimental results show that diffusion models outperform other state-of-the-art MRI synthesis methods, underscoring their effectiveness in addressing the challenges of multimodal MRI synthesis and generating high-quality images.

As diffusion models continue to show promise, there is an exciting opportunity for academia to further explore and improve their efficiency and feasibility in the future. This research direction holds great potential for advancing medical imaging technologies and benefiting both clinicians and patients.

6.8 Neural implicit representations for missing modality synthesis

Given the recent promising results showcased by neural implicit representation methods [125], with a notable emphasis on the neural radiance fields (NeRF) approach, an intriguing avenue emerges for exploring their potential application in synthesizing missing modalities within the domain of medical imaging. This exploration could entail a comprehensive investigation into the foundational principles of neural implicit representations, aiming to discern how these techniques can be effectively employed to generate realistic and high-quality missing modality images. An examination of NeRF's capacity to model intricate 3D scenes [126] further adds depth to the potential utility of this approach, particularly in capturing nuanced anatomical details and diverse modalities. Moreover, delving into the challenges and opportunities tied to the seamless integration of neural implicit representations into the established framework of medical image analysis would offer valuable insights into the feasibility and efficiency of such an undertaking. As the evolution of NeRF and related methods continues, their exploration in the context of addressing challenges associated with missing modality in medical imaging stands as a promising trajectory for advancing the field.

7 Conclusions

This survey has presented a detailed discussion regarding missing modality compensation networks. Our taxonomy divided the literature into five categories: early synthesis models, common latent space, knowledge distillation networks, mutual information maximization, and GANs. For each strategy, a summary of the literature, network architecture, algorithms, and motivation along with its pros and cons are provided. Furthermore, a detailed discussion of these methods is provided to highlight the most important contribution of each strategy and point out the limitations they may face in their design. Moreover, we summarized the most common benchmarks, evaluation metrics and quantitative performance to provide a clear view of the application for a reader. Finally, our last section provided information regarding the challenges and potential research direction for future work.

Availability of data and materialss

All data used in this study are publicly available.

Funding

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under project numbers 191948804 and 455548460.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Author contributions

RA designed the study concept and the evaluation, and contributed to writing. MD and NK contributed to the writing. JCA and DM revised the study concept. All authors reviewed the manuscript.

References

[1] Ouyang, J.; Adeli, E.; Pohl, K. M.; Zhao, Q.; Zaharchuk, G. Representation disentanglement for multi-modal brain MRI analysis. In: *Information Processing in Medical Imaging. Lecture Notes in Computer Science, Vol. 12729.* Feragen, A.; Sommer, S.; Schnabel, J.; Nielsen, M. Eds. Springer Cham, 321–333, 2021.

[2] Dinsdale, N. K.; Bluemke, E.; Smith, S. M.; Arya, Z.; Vidaurre, D.; Jenkinson, M.; Namburete, A. Learning patterns of the ageing brain in MRI using deep convolutional networks. *NeuroImage* Vol. 224, 117401, 2021.

[3] Feng, C. M.; Wang, K.; Lu, S.; Xu, Y.; Li, X. Brain MRI super-resolution using coupled-projection residual network. *Neurocomputing* Vol. 456, 190–199, 2021.

[4] Conze, P. H.; Kavur, A. E.; Gall, E. C.; Gezer, N. S.; Le Meur, Y.; Selver, M. A.; Rousseau, F. Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. *arXiv preprint* arXiv:2001.09521, 2020.

[5] Biondetti, P.; Vangel, M. G.; Lahoud, R. M.; Furtado, F. S.; Rosen, B. R.; Groshar, D.; Canamaque, L G.; Umutlu, L.; Zhang, E. W.; Mahmood, U.; et al. PET/MRI assessment of lung nodules in primary abdominal malignancies: sensitivity and outcome analysis. *European Journal of Nuclear Medicine and Molecular Imaging* Vol. 48, No. 6, 1976–1986, 2021.

[6] Azad, R.; Asadi-Aghbolaghi, M.; Fathy, M.; Escalera, S. Bi-directional ConvLSTM U-Net with densley connected convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[7] Reyngoudt, H.; Marty, B.; Boisserie, J. M.; Le Louër, J.; Koumako, C.; Baudin, P. Y.; Wong, B.; Stojkovic, T.; Béhin, A.; Gidaro, T.; et al. Global versus individual muscle segmentation to assess quantitative MRI-based fat fraction changes in neuromuscular diseases. *European Radiology* Vol. 31, No. 6, 4264–4276, 2021.

[8] Lee, Y. H. Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks. *Journal of Digital Imaging* Vol. 31, No. 5, 604–610, 2018.

[9] Azad, R.; Rouhier, L.; Cohen-Adad, J. Stacked hourglass network with a multi-level attention mechanism: Where to look for intervertebral disc labeling. In: *Machine Learning in Medical Imaging. Lecture Notes in Computer Science, Vol. 12966.* Lian, C.; Cao, X.; Rekik, I.; Xu, X.; Yan, P. Eds. Springer Cham, 406–415, 2021.

[10] Bozorgpour, A.; Azad, R.; Showkatian, E.; Sulaiman, A. Multi-scale regional attention Deeplab3+: Multiple myeloma plasma cells segmentation in microscopic images. *arXiv preprint* arXiv:2105.06238, 2021.

[11] Feyjie, A. R.; Azad, R.; Pedersoli, M.; Kauffman, C.; Ben Ayed, I.; Dolz, J. Semi-supervised few-shot learning for medical image segmentation. *arXiv preprint* arXiv:2003.08462, 2020.

[12] Brady, A. P. Error and discrepancy in radiology: Inevitable or avoidable? *Insights into Imaging* Vol. 8, No. 1, 171–182, 2017.

[13] Yao, W.; Liu, C.; Wang, N.; Zhou, H.; Chen, H.; Qiao, W. An MRI-guided targeting dual-responsive drug delivery system for liver cancer therapy. *Journal of Colloid and Interface Science* Vol. 603, 783–798, 2021.

[14] Delli Pizzi, A.; Chiarelli, A. M.; Chiacchiaretta, P.; D'Annibale, M.; Croce, P.; Rosa, C.; Mastrodicasa, D.; Trebeschi, S.; Lambregts, D. M. J.; Caposiena, D.; et al. MRI-based clinical-radiomics model predicts tumor response before treatment in locally advanced rectal cancer. *Scientific Reports* Vol. 11, No. 1, Article No. 5379, 2021.

[15] Yao, W.; Liu, C.; Wang, N.; Zhou, H.; Chen, H.; Qiao, W. Anisamide-modified dual-responsive drug delivery system with MRI capacity for cancer targeting therapy. *Journal of Molecular Liquids* Vol. 340, Article No. 116889, 2021.

[16] Bleker, J.; Yakar, D.; van Noort, B.; Rouw, D.; de Jong, I. D.; Dierckx, R.; Kwee, T.; Huisman, H. Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer. *Insights into Imaging* Vol. 12, Article No. 150, 2021.

[17] Park, Y. M.; Lim, J. Y.; Koh, Y. W.; Kim, S. H.; Choi, E. C. Prediction of treatment outcome using MRI radiomics and machine learning in oropharyngeal cancer patients after surgical treatment. *Oral Oncology* Vol. 122, Article No. 105559, 2021.

[18] Azad, R.; Khosravi, N.; Merhof, D. SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities. *arXiv preprint* arXiv:2204.02961, 2022.

[19] Graves, M. J.; Mitchell, D. G. Body MRI artifacts in clinical practice: A physicist's and radiologist's perspective. *Journal of Magnetic Resonance Imaging* Vol. 38, No. 2, 269–287, 2013.

[20] Bekiesińska-Figatowska, M. Artifacts in magnetic resonance imaging. *Polish Journal of Radiology* Vol. 80, 93–106, 2015.

[21] Hamghalam, M.; Frangi, A. F.; Lei, B.; Simpson, A. L. Modality completion via gaussian process prior variational autoencoders for multi-modal glioma segmentation. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 12907.* De Bruijne, Marleen.; Cattin, P. C.; Cotin, S.; Padoy, N.; Speidel, S.; Zheng, Y.; Essert, C. Eds. Springer Cham, 442–452, 2021.

[22] Dalmaz, O.; Yurt, M.; Çukur, T. ResViT: Residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging* Vol. 41, No. 10, 2598–2614, 2022.

[23] Zhang, Y.; Yang, J.; Tian, J.; Shi, Z.; Zhong, C.; Zhang, Y.; He, Z. Modality-aware mutual learning for multi-modal medical image segmentation. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol.*

12901. De Bruijne, Marleen.; Cattin, P. C.; Cotin, S.; Padoy, N.; Speidel, S.; Zheng, Y.; Essert, C. Eds. Springer Cham, 589–599, 2021.

[24] Zhou, T.; Canu, S.; Vera, P.; Ruan, S. Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing MR modalities. *Neurocomputing* Vol. 466, 102–112, 2021.

[25] Zhu, Y.; Wang, S.; Lin, R.; Hu, Y.; Chen, Q. Brain tumor segmentation for missing modalities by supplementing missing features. In: Proceedings of the IEEE 6th International Conference on Cloud Computing and Big Data Analytics, 652–656, 2021.

[26] Yu, Z.; Zhai, Y.; Han, X.; Peng, T.; Zhang, X.-Y. MouseGAN: GAN-based multiple MRI modalities synthesis and segmentation for mouse brain structures. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 12901.* De Bruijne, Marleen.; Cattin, P. C.; Cotin, S.; Padoy, N.; Speidel, S.; Zheng, Y.; Essert, C. Eds. Springer Cham, 442–450, 2021.

[27] Tulder, G. V.; de Bruijne, M. Why does synthesized data improve multi-sequence classification? In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 9349.* Navab, N.; Hornegger, J.; Wells, W. M.; Frangi, A. Eds. Springer Cham, 531–538, 2021.

[28] Jog, A.; Carass, A.; Roy, S.; Pham, D. L.; Prince, J. L. Random forest regression for magnetic resonance image synthesis. *Medical Image Analysis* Vol. 35, 475–488, 2017.

[29] Havaei, M.; Guizard, N.; Chapados, N.; Bengio, Y. Hemis: Hetero-modal image segmentation. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 9901.* Ourselin, S.; Joskowicz, L.; Sabunce, M. R.; Unal, G.; Wells, W. Eds. Springer Cham, 469–477, 2016.

[30] Choi, Y.; Al-masni, M. A.; Jung, K. J.; Yoo, R. E.; Lee, S. Y.; Kim, D. H. A single stage knowledge distillation network for brain tumor segmentation on limited MR image modalities. *Computer Methods and Programs in Biomedicine* Vol. 240, Article No. 107644, 2023.

[31] Vadacchino, S.; Mehta, R.; Sepahvand, N.; Nichyporuk, B.; Clark, J. J.; Arbel, T. HAD-net: A hierarchical adversarial knowledge distillation network for improved enhanced tumour segmentation without post-contrast images. *arXiv preprint* arXiv:2103.16617, 2021.

[32] Wang, Q.; Zhan, L.; Thompson, P.; Zhou, J. Multimodal learning with incomplete modalities by knowledge distillation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1828–1838, 2020.

[33] Karimijafarbigloo, S.; Azad, R.; Kazerouni, A.; Ebadollahi, S.; Merhof, D. MMCFormer: Missing modality compensation transformer for brain tumor segmentation. In: Proceedings of the Machine Learning Research, 1144–1162, 2023.

[34] Zhou, T.; Canu, S.; Vera, P.; Ruan, S. Latent correlation representation learning for brain tumor segmentation with missing MRI modalities. *IEEE Transactions on Image Processing* Vol. 30, 4263–4274, 2021.

[35] Sylvain, T.; Dutil, F.; Berthier, T.; Di Jorio, L.; Luck, M.; Hjelm, D.; Bengio, Y. Cross-modal information maximization for medical imaging: CMIM. *arXiv preprint* arXiv:2010.10593, 2020.

[36] Pan, Y.; Chen, Y.; Shen, D.; Xia, Y. Collaborative image synthesis and disease diagnosis for classification of neurodegenerative disorders with incomplete multi-modal neuroimages. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 12905.* Navab, N.; Hornegger, J.; Wells, W. M.; Frangi, A. Eds. Springer Cham, 480–489, 2021.

[37] Sharma, A.; Hamarneh, G. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Transactions on Medical Imaging* Vol. 39, No. 4, 1170–1183, 2020.

[38] Yu, B.; Zhou, L.; Wang, L.; Fripp, J.; Bourgeat, P. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, 626–630, 2018.

[39] Isola, P.; Zhu, J. Y.; Zhou, T.; Efros, A. A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5967–5976, 2017.

[40] Zhan, B.; Li, D.; Wang, Y.; Ma, Z.; Wu, X.; Zhou, J.; Zhou, L. LR-cGAN: Latent representation based conditional generative adversarial network for multi-modality MRI synthesis. *Biomedical Signal Processing and Control* Vol. 66, Article No. 102457, 2021.

[41] Hofmann, M.; Steinke, F.; Scheel, V.; Charpiat, G.; Farquhar, J.; Aschoff, P.; Brady, M.; Schölkopf, B.; Pichler, B. J. MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine* Vol. 49, No. 11, 1875–1883, 2008.

[42] Varsavsky, T.; Eaton-Rosen, Z.; Sudre, C. H.; Nachev, P.; Cardoso, J. M. PIMMS: Permutation invariant multi-modal segmentation. In: *Deep Learning in Medical Imaging Analysis and Multimodal Learning for Clinical Decision Support. Lecture Notes in Computer*

Science, Vol. 11045. Stoyanov, D.; Taylor, Z.; Carneiro, G.; Syeda-Mahmood, T.; Martel, A.; Maier-Hein, L.; Tavares, J. M. R. S.; Bradley, A.; Papa, J. P.; Belagiannis, V.; et al. Eds. Springer Cham, 201–209, 2018.

[43] Mehta, R.; Arbel, T. RS-Net: Regression-segmentation 3D CNN for synthesis of full resolution missing brain MRI in the presence of tumours. *arXiv preprint* arXiv:1807.10972, 2018.

[44] Shen, Y.; Gao, M. Brain tumor segmentation on MRI with missing modalities. In: *Information Processing in Medical Imaging. Lecture Notes in Computer Science, Vol. 11492*. Chung, A. C. S.; Gee, J. C.; Yushkevich, P. A.; Bao, S. Eds. Springer Cham, 417–428, 2019.

[45] Dorent, R.; Joutard, S.; Modat, M.; Ourselin, S.; Vercauteren, T. Hetero-modal variational encoder-decoder for joint modality completion and segmentation. In: *Machine Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 11765*. Navab, N.; Hornegger, J.; Wells, W. M.; Frangi, A. Eds. Springer Cham, 74–82, 2019.

[46] Wang, Y.; Zhang, Y.; Liu, Y.; Lin, Z.; Tian, J.; Zhong, C.; Shi, Z.; Fan, J.; He, Z. ACN: Adversarial co-training network for brain tumor segmentation with missing modalities. *arXiv preprint* arXiv:2106.14591, 2021.

[47] Ding, Y.; Yu, X.; Yang, Y. RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 3955–3964, 2021.

[48] Lau, K.; Adler, J.; Sjölund, J. A unified representation network for segmentation with missing modalities. *arXiv preprint* arXiv:1908.06683, 2019.

[49] Konwer, A.; Hu, X.; Bae, J.; Xu, X.; Chen, C.; Prasanna, P. Enhancing modality-agnostic representations via meta-learning for brain tumor segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 21415–21425, 2023.

[50] Liu, H.; Wei, D.; Lu, D.; Sun, J.; Wang, L.; Zheng, Y. M3AE: Multimodal representation learning for brain tumor segmentation with missing modalities. *arXiv preprint* arXiv:2303.05302, 2023.

[51] Zhou, T.; Canu, S.; Vera, P.; Ruan, S. Conditional generator and multi-sourcecorrelation guided brain tumor segmentation with missing MR modalities. *arXiv preprint* arXiv:2105.13013, 2021.

[52] Zhou, T. Feature fusion and latent feature learning guided brain tumor segmentation and missing modality recovery network. *Pattern Recognition* Vol. 141, Article No. 109665, 2023.

[53] Cao, B.; Zhang, H.; Wang, N.; Gao, X.; Shen, D. Auto-GAN: Self-supervised collaborative learning for medical image synthesis. *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 34, No. 7, 10486–10493, 2020.

[54] Li, H.; Paetzold, J. C.; Sekuboyina, A.; Kofler, F.; Zhang, J.; Kirschke, J. S.; Wiestler, B.; Menze, B. DiamondGAN: Unified multi-modal generative adversarial networks for MRI sequences synthesis. *arXiv preprint* arXiv:1904.12894, 2019.

[55] Qian, S.; Wang, C. COM: Contrastive Masked-attention model for incomplete multimodal learning. *Neural Networks* Vol. 162, 443–455, 2023.

[56] Islam, S. K.; Nasim, M. D.; Hossain, I.; Ullah, M. A.; Gupta, K. D.; Bhuiyan, M. M. H. Introduction of medical imaging modalities. *arXiv preprint* arXiv:2306.01022, 2023.

[57] Webb, G. A. *Introduction to Biomedical Imaging*. John Wiley & Sons, 2017.

[58] Weishaupt, D.; Köchli, V. D.; Marincek, B. *How does MRI Work? An Introduction to the Physics and Function of Magnetic Resonance Imaging*, 2nd edn. Springer, 2006.

[59] Baba, Y.; Jones, J. A. T1 weighted image. 2009. Available at `https://radiopaedia.org/articles/5852`

[60] Haouimi, A.; Jones, J. T2 weighted image. 2009. Available at `https://radiopaedia.org/articles/6345`

[61] Baba, Y.; Niknejad, M. Fluid attenuated inversion recovery. 2013. Available at `https://radiopaedia.org/articles/21760`

[62] Tanner, M.; Gambarota, G.; Kober, T.; Krueger, G.; Erritzoe, D.; Marques, J. P.; Newbould, R. Fluid and white matter suppression with the MP2RAGE sequence. *Journal of Magnetic Resonance Imaging* Vol. 35, No. 5, 1063–1070, 2012.

[63] Yu, H.; Buch, K.; Li, B.; O'Brien, M.; Soto, J.; Jara, H.; Anderson, S. W. Utility of texture analysis for quantifying hepatic fibrosis on proton density MRI. *Journal of Magnetic Resonance Imaging* Vol. 42, No. 5, 1259–1265, 2015.

[64] Carass, A.; Roy, S.; Jog, A.; Cuzzocreo, J. L.; Magrath, E.; Gherman, A.; Button, J.; Nguyen, J.; Bazin, P. L.; Calabresi, P. A.; et al. Longitudinal multiple sclerosis lesion segmentation data resource. *Data in Brief* Vol. 12, 346–350, 2017.

[65] Menze, B. H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE*

Transactions on Medical Imaging Vol. 34, No. 10, 1993–2024, 2015.

[66] Ashton, E. A.; Kim, S. H. Evaluation of reproducibility for perfusion assessment of tumors in MRI. In: Proceedings of the 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 824–827, 2004.

[67] Brown, M. A.; Semelka, R. C. *MRI: Basic Principles and Applications*. John Wiley & Sons, 2015.

[68] Rauf, N.; Alam, D. Y.; Jamaluddin, M.; Samad, B. A. Improve image quality of transversal relaxation time PROPELLER and FLAIR on magnetic resonance imaging. *Journal of Physics*: *Conference Series* Vol. 979, Article No. 012079, 2018.

[69] Mudgal, P. Case courtesy of Dr. Prashant Mudgal. 2012. Available at `https://radiopaedia.org/cases/26952/studies/27131`

[70] Ballinger, J. R. Polycystic ovaries. 2013. Available at `https://radiopaedia.org/cases/polycystic-ovaries`

[71] Fischer, A.; Igel, C. An introduction to restricted boltzmann machines. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes in Computer Science, Vol. 7741.* Alvarez, L.; Mejail, M.; Gomez, L.; Jacobo, J. Eds. Springer Cham, 14–36, 2012.

[72] Wang, Y.; Hu, H.; Yu, S.; Yang, Y.; Guo, Y.; Song, X.; Chen, F.; Liu, Q. A unified hybrid transformer for joint MRI sequences super-resolution and missing data imputation. *Physics in Medicine & Biology* Vol. 68, No. 13, Article No. 135006, 2023.

[73] Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *arXiv preprint* arXiv:1606.06650, 2016.

[74] Wu, M.; Goodman, N. Multimodal generative models for scalable weakly-supervised learning. *arXiv preprint* arXiv:1802.05335, 2018.

[75] Liu, J.; Pasumarthi, S.; Duffy, B.; Gong, E.; Datta, K.; Zaharchuk, G. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. *IEEE Transactions on Medical Imaging* Vol. 42, No. 9, 2577–2591, 2023.

[76] Bucilua, C.; Caruana, R.; Niculescu-Mizil, A. Model compression. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 535–541, 2006.

[77] Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv preprint* arXiv:1503.02531, 2015.

[78] Isensee, F.; Kickingereder, P.; Martin Bendszus, W.; Maier-Hein, K. H. No new-net. *arXiv preprint* arXiv:1809.10483, 2018.

[79] Belghazi, M. I.; Baratin, A.; Rajeswar, S.; Ozair, S.; Bengio, Y.; Courville, A.; Hjelm, R. D. MINE: Mutual information neural estimation. *arXiv preprint* arXiv:1801.04062, 2018.

[80] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *arXiv preprint* arXiv:1406.2661, 2014.

[81] Pan, K.; Cheng, P.; Huang, Z.; Lin, L.; Tang, X. Transformer-based T2-weighted MRI synthesis from T1-weighted images. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, 5062–5065, 2022.

[82] Chen, X.; Lian, C.; Wang, L.; Deng, H.; Kuang, T.; Fung, S.; Gateno, J.; Yap, P. T.; Xia, J. J.; Shen, D. Anatomy-regularized representation learning for cross-modality medical image segmentation. *IEEE Transactions on Medical Imaging* Vol. 40, No. 1, 274–285, 2021.

[83] Hu, M.; Maillard, M.; Zhang, Y.; Ciceri, T.; La Barbera, G.; Bloch, I.; Gori, P. Knowledge distillation from multi-modal to mono-modal segmentation networks. *arXiv preprint* arXiv:2106.09564, 2020.

[84] Mirzadeh, S. I.; Farajtabar, M.; Li A.; Levine, N.; Matsukawa, A.; Ghasemzadeh, H. Improved knowledge distillation *via* teacher assistant. *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 34, No. 4, 5191–5198, 2020.

[85] Cho, J. H.; Hariharan, B. On the efficacy of knowledge distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 4793–4801, 2019.

[86] Cho, J. H.; Hariharan, B. On the efficacy of knowledge distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 4793–4801, 2019.

[87] Chang, Q.; Yan, Z.; Baskaran, L.; Qu, H.; Zhang, Y.; Zhang, T.; Zhang, S.; Metaxas, D. N. Multi-modal AsynDGAN: Learn from distributed medical image data without sharing private information. *arXiv preprint* arXiv:2012.08604, 2020.

[88] Huang, P.; Li, D.; Jiao, Z.; Wei, D.; Li, G.; Wang, Q.; Zhang, H.; Shen, D. CoCa-GAN: Common-feature-learning-based context-aware generative adversarial network for glioma grading. In: *Medical Imaging Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, Vol. 11766.* Shen, D.; Liu, T.; Peters, T. M.; Staib, L. H.; Essert, C.; Zhou, S.; Yap, P.-T.; Khan, A. Eds. Springer Cham, 155–163, 2019.

[89] Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. *arXiv preprint* arXiv:1606.03498, 2016.

[90] Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. *arXiv preprint* arXiv:2006.06676, 2020.

[91] Myronenko, A. 3D MRI brain tumor segmentation using autoencoder regularization. *arXiv preprint* arXiv:1810.11654, 2024.

[92] Styner, M.; Lee, J.; Chin, B.; Chin, M.; Commowick, O.; Tran, H.; Markovic-Plese, S.; Jewells, V.; Warfield, S. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *The MIDAS Journal* 2008, `http://hdl.handle.net/10380/1509`

[93] Tillin, T.; Forouhi, N. G.; McKeigue, P. M.; Chaturvedi, N. Southall and Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *International Journal of Epidemiology* Vol. 41, No. 1, 33–42, 2012.

[94] Kuijf, H. J. WMH segmentation challenge. 2017. Available at `https://wmh.isi.uu.nl/`

[95] Maier, O.; Menze, B. H.; von der Gablentz, J.; Häni, L.; Heinrich, M. P.; Liebrand, M.; Winzeck, S.; Basit, A.; Bentley, P.; Chen, L.; et al. A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis* Vol. 35, 250–269, 2017.

[96] Kavur, A. E.; Gezer, N. S.; Barıç, M.; Aslan, S.; Conze, P. H.; Groza, V.; Pham, D. D.; Chatterjee, S.; Ernst, P.; Özkan, S.; et al. CHAOS Challenge-combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis* Vol. 69, Article No. 101950, 2021.

[97] Commowick, O.; Kain, M.; Casey, R.; Ameli, R.; Ferré, J. C.; Kerbrat, A.; Tourdias, T.; Cervenansky, F.; Camarasu-Pop, S.; Glatard, T.; et al. Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset. *NeuroImage* Vol. 244, Article No. 118589, 2021.

[98] Brudfors, M.; Ashburner, J.; Nachev, P.; Balbastre, Y. Empirical bayesian mixture models for medical image translation. *arXiv preprint* arXiv:1908.05926, 2019.

[99] MacDonald, J. A.; Zhu Z.; Konkel, B.; Mazurowski, M. A.; Wiggins, W. F.; Bashir, M. R. Duke liver dataset: A publicly available liver MRI dataset with liver segmentation masks and series labels. *Radiology: Artificial Intelligence* Vol. 5, No. 5, Article No. e220275, 2023.

[100] Giacomello, E.; Loiacono, D.; Mainardi, L. Brain MRI tumor segmentation with adversarial networks. *arXiv preprint* arXiv:1910.02717, 2019.

[101] Chen, C.; Dou, Q.; Jin, Y.; Chen, H.; Qin, J.; Heng, P.-A. Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. *arXiv preprint* arXiv:2002.09708, 2019.

[102] Islam, M.; Wijethilake, N.; Ren, H. Glioblastoma multiforme prognosis: MRI missing modality generation, segmentation and radiogenomic survival prediction. *Computerized Medical Imaging and Graphics* Vol. 91, Article No. 101906, 2021.

[103] Zhou, T.; Fu, H.; Chen, G.; Shen, J.; Shao, L. Hi-net: Hybrid-fusion network for multi-modal MR image synthesis. *IEEE Transactions on Medical Imaging* Vol. 39, No. 9, 2772–2781, 2020.

[104] Jack Jr, C. R.; Bernstein, M. A.; Fox, N. C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P. R.; Whitwell, J. L.; Ward, C.; et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* Vol. 27, No. 4, 685–691, 2008.

[105] Orbes-Arteaga, M.; Cardoso, M. J.; Sørensen, L.; Modat, M.; Ourselin, S.; Nielsen, M.; Pai, A. Simultaneous synthesis of FLAIR and segmentation of white matter hypointensities from T1 MRIs. *arXiv preprint* arXiv:1808.06519, 2018.

[106] Yuan, W.; Wei, J.; Wang, J.; Ma, Q.; Tasdizen, T. Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. *Medical Image Analysis* Vol. 64, Article No. 101731, 2020.

[107] Simpson, A. L.; Antonelli, M.; Bakas, S.; Bilello, M.; Farahani, K.; van Ginneken, B.; Kopp-Schneider, A.; Landman, B. A.; Litjens, G.; Menze, B.; et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint* arXiv:1902.09063, 2019.

[108] Zhu, Z.; Mittendorf, A.; Shropshire, E.; Allen, B.; Miller, C.; Bashir, M. R.; Mazurowski, M. A. 3D pyramid pooling network for abdominal MRI series classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 44, No. 4, 1688–1698, 2022.

[109] Milletari, F.; Navab, N.; Ahmadi, S. A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th International Conference on 3D Vision, 565–571, 2016.

[110] Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P. F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; et al. nnU-Net: Self-adapting framework for U-Net-Based medical image segmentation. *arXiv preprint* arXiv:1809.10486, 2019.

[111] Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.; Adam, H.; Kalenichenko, D. Quantization

and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2704–2713, 2018.

[112] Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Speed/accuracy trade-offs for modern convolutional object detectors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3296–3297, 2017.

[113] Tan, M.; Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv preprint* arXiv:1905.11946, 2019.

[114] Matsuo, K.; Tanaka, Y.; Sarmenta, L. F. G.; Nakai, T.; Bagarinao, E. Enabling on-demand real-time functional MRI analysis using grid technology. *Methods of Information in Medicine* Vol. 44, No. 5, 665–673, 2005.

[115] Azad, R.; Aghdam, E. K.; Rauland, A.; Jia, Y.; Avval, A. H.; Bozorgpour, A.; Karimijafarbigloo, S.; Cohen, J. P.; Adeli, E.; Merhof, D. Medical image segmentation review: The success of U-Net. *arXiv preprint* arXiv:2211.14830, 2022.

[116] Wang, H.; Wang, Z.; Du, M.; Yang, F.; Zhang, Z.; Ding, S.; Mardziel, P.; Hu, X. Score-CAM: Score-weighted visual explanations for convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 24–25, 2020.

[117] Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, 618–626, 2017.

[118] Gros, C.; Lemay, A.; Vincent, O.; Rouhier, L.; Bucquet, A.; Cohen, J. P.; Cohen-Adad, J. Ivadomed: A medical imaging deep learning toolbox. *arXiv preprint* arXiv:2010.09984, 2020.

[119] Cardoso, M. J.; Li, W.; Brown, R.; Ma, N.; Kerfoot, E.; Wang, Y.; Murrey, B.; Myronenko, A.; Zhao, C.; Yang, D.; et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv preprint* arXiv:2211.02701, 2022.

[120] Jaeger, P. F.; Kohl, S. A. A.; Bickelhaupt, S.; Isensee, F.; Kuder, T. A.; Schlemmer, H. P.; Maier-Hein, K. H. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. *arXiv preprint* arXiv:1811.08661, 2018.

[121] Pawlowski, N.; Ktena, S. I.; Lee, M. C. H.; Kainz, B.; Rueckert, D.; Glocker, B.; Rajchl, M. DLTK: State of the art reference implementations for deep learning on medical images. *arXiv preprint* arXiv:1711.06853, 2017.

[122] Kazerouni, A.; Aghdam, E. K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *arXiv preprint* arXiv:2211.07804, 2022.

[123] Jiang, L.; Mao, Y.; Wang, X.; Chen, X.; Li, C. CoLa-Diff: Conditional latent diffusion model for multi-modal MRI synthesis. *arXiv preprint* arXiv:2303.14081, 2023.

[124] Meng, X.; Gu, Y.; Pan, Y.; Wang, N.; Xue, P.; Lu, M.; He, X.; Zhan, Y.; Shen, D. A novel unified conditional score-based generative framework for multi-modal medical image completion. *arXiv preprint* arXiv:2207.03430, 2022.

[125] Molaei, A.; Aminimehr, A.; Tavakoli, A.; Kazerouni, A.; Azad, B.; Azad, R.; Merhof, D. Implicit neural representation in medical imaging: A comparative survey. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2373–2383, 2023.

[126] Iddrisu, K.; Malec, S.; Crimi, A. 3D reconstructions of brain from MRI scans using neural radiance fields. In: *Artificial Intelligence and Soft Computing. Lecture Notes in Computer Science, Vol. 14126*. Rutkowski, L.; Scherer, R.; Korytkowski, M.; Pedrycz, W.; Tadeusiewicz, R.; Zurada, J. M. Eds. Springer Cham, 207–218, 2023.

Reza Azad is a Ph.D. student in the Electrical Engineering Department at RWTH Aachen University, Germany. His research interests include deep learning, medical image processing, and computer vision.

Mohammad Dehghanmanshadi is a research assistant specializing in AI in medicine at Iran University of Medical Sciences. His focus lies in deep learning and medical image analysis.

Nika Khosravi is currently a research associate and Ph.D. student at the Chair of Medical Information Technology at the Helmholtz Institute for Biomedical Engineering, RWTH Aachen University. Her research focuses on the deployment of AI and data-driven methods for the intelligent control and monitoring of artificial ventilation systems.

Julien Cohen-Adad is an associate professor at Polytechnique Montreal, a member of Mila (Univ. Montreal). He holds the Canada Research Chair in Quantitative Magnetic Resonance Imaging.

Dorit Merhof is a full professor and head of the Institute of Image Analysis and Computer Vision, Faculty of Informatics and Data Science, University of Regensburg, Germany. Her research interests include deep learning, biomedical image processing, and computer vision.

To submit a manuscript, please go to `https://jcvm.org`.