3 Literature Review

Given the significant growth in the incorporation of ML into imputation, inpainting and completion algorithms, this chapter aims to provide a comprehensive overview of the current landscape of image imputation technology, highlighting recent research endeavours with a special emphasis on applications in medical imaging.

The research process was conducted across several academic databases, including *IEEE Xplore*, *PubMed*, *ScienceDirect*, *Scopus* and *Web of Science*. Articles were selected using a combination of keywords such as "Completion", "Imputation", "Inpainting", "Machine Learning", and "Medical Imaging", among others.

In architectural terms, GANs and CNNs emerged as the most frequently used, although some less prevalent works based on Variational Autoencoders (VAE) and other architectures were also noted [68]. The preference for GANs and CNNs was largely due to the larger number of recent findings associated with these architectures.

Knowing in advance that recent overviews have introduced several classifications of image imputation techniques — typically linked to the architectural families utilized [6, 9], the practical application domains of the models [11] or the stratification of the reconstruction process [7] — this section begins by defining and categorizing imputation algorithms based on the type of computational structure responsible for the reconstructive process. Subsequently, when discussing studies on medical images, the focus shifts towards the practical application of these techniques.

3.1 General Approaches

More recently, alongside advancements in computational learning techniques such as CNNs and GANs, which are capable of capturing high-level patterns and semantics, new learning-based image reconstruction techniques have emerged to address the limitations of traditional approaches [6–11]. In this context, these deep learning approaches can be further divided into subgroups: CNN-based and GAN-based methods [9].

3.1.1 CNN-based Approaches

In the case of CNN-based techniques, the most extensively studied architectures, namely Fully Convolutional Networks (FCNs) and U-Nets, stem from the Encoder-Decoder model. When applied to an inpainting problem, this architecture family is characterized by efficient down-sampling of damaged images, followed by an up-sampling to a new image completely reconstructed. [9, 10]. Consequently, the fundamental principle of the CNN-based techniques is to construct a feature map that efficiently integrates all image details, encompassing structural and textural information [9, 10]. Historically, between the years 2012 and 2014, pioneering studies by Xie et al. [69], Eigen et al. [71] and Xu et al. [70] provided initial evidence of CNNs' efficacy in image reconstruction within denoising algorithms. The adaptation of architectures across various research domains, such as FCNs — first introduced by Long et al. [72] in 2015 for image segmentation and later modified for image inpainting [73] — has led to a significant surge in research on CNN-based algorithms.

In this manner, ahead of its time in the image inpainting domain, Cai et al. [73] introduced a novel network in 2015 called Blind Image Convolutional Neural Network (BICNN). This network utilized 0.3 million ground truth grey-scale images from *ImageNet*, enabling it to autonomously learn the mapping relationship between corrupted and ground truth sub-images, even without prior knowledge of the damaged pixels. Notably, it achieved an average PSNR and SSIM of 36.48 dB and 0.9809, respectively [73].

In 2017, Chaudhury and Roy [74] employed a FCNs for denoising and reconstructing three-colour channel images, after the same network demonstrated high performance in segmentation tasks [74]. The research findings indicated that the model exhibited marginally superior performance compared to other denoising models, as assessed by the PSNR metric [74]. For white noise levels of $\sigma = 25$ and $\sigma = 50$ in images sourced from the *ImageNet* and *MSCOCO* datasets, the average PSNR values were 30.42 dB and 27.12 dB, respectively [74]. However, cases involving missing pixel inpainting and text removal demonstrated slightly inferior results (PSNR=30.95 dB) compared to existing techniques [74].

In the same year, Yang et al. [75] introduced a multi-scale neural patch synthesis approach using the correlations between mid-layer features, together with a pre-trained network (*VGG* network) application to enhance the reconstructed image texture [75]. The study demonstrated the best performance among the compared techniques, both qualitatively and quantitatively (MAE=10.01%; MSE=2.21%; PSNR=18.00 dB), highlighting an efficient high-frequency details capture [75]. However, the method exhibited some weaknesses, mainly for larger hole sizes, derived from the weakening correlations between pixels, caused by the distance increase between known and unknown data [75].

To overcome the limitation of rectangular masks typically employed in previous approaches, Liu et al. [76] introduced Partial Convolution (PConv) layers for image inpainting. These innovative layers enabled the implementation of irregular masks, with a normalization mechanism applied solely to the pixels within non-hole regions [76]. In their study, a qualitative evaluation of the style and perceptual losses' impact on the overall loss was initially conducted, followed by a quantitative comparison with other literature approaches [76]. Then, the authors examined the performance of each model across different hole-to-image area ratios. For hole areas ranging from 1% to 10%, PConv exhibited MAE, PSNR, SSIM and IS values of 0.49%, 33.75 dB, 0.946 and 0.051, respectively, for masks positioned away from the edges [76]. Conversely, for scenarios where the holes were near or touching the image edges, the results were 0.47%, 34.34 dB, 0.945 and 0.032, respectively [76]. Further analysis revealed that in "no border" situations with hole-to-image area ratios between 20% to 30% or 40% to 50%, the PSNR and SSIM values were 24.54 dB, 20.34 dB and 0.775, 0.583, respectively. This suggests that as the area covered by the mask increases, the model's reconstruction capability diminishes [76]. Nonetheless, PConv consistently outperformed other techniques in most cases, yielding superior metric results. It is crucial to note that the comparison might not be entirely exhaustive since only *PatchMatch*, a non-learning-based technique, could directly be applied to irregular masks [76].

More recently, in 2021, Deng et al. [77] introduced a model based on FCNs that incorporated attention modules, referred to as transformer blocks, in place of some convolution layers. These transformer blocks utilized a multi-scale multi-head attention mechanism [77]. In their study, three different datasets (*CelebA-HQ*, *Paris StreetView* and *Places2*) were tested across several missing ratios [77]. According to the evaluation metrics, including MAE, SSIM and PSNR, this model outperformed all image inpainting techniques compared, such as the PConv approach [77]. To highlight some notable results, when evaluated on the *CelebA-HQ* dataset, the model achieved PSNR values of 32.84, 29.75, 27.35 and 25.43 dB for mask ratios of $10-20\%$, $20-30\%$, $30-40\%$ and $40-50\%$, respectively [77]. Additionally, for the same mask ratios, SSIM values were recorded as 0.981, 0.964, 0.940 and 0.909 [77].

Subsequently, Table 2 was created to consolidate the findings from the aforementioned papers. It is important to note that direct comparisons between the results of the previous models may not be feasible due to variations in the training conditions, like the datasets and the missing area dimensions.

3.1.2 GAN-based Approaches

As mentioned earlier, the first GAN framework was introduced by Goodfellow et al. [48], in 2014, marking the beginning of a new era in algorithms such as those used for imputation problems [48].

However, the effective application of this new adversarial model to image reconstruction mechanisms only emerged two years later, in 2016, introduced by Pathak et al. [78]. In this study, the authors utilized an encoder-decoder mechanism with internally fully connected layers to enable a square mask

Table 2: Main topics of research projects related to CNN-based imputation techniques.

AUTHORS	**OBJECTIVE**	**DATASET**	**EVALUATION METRICS**	**RESULTS**
Cai et al. [73]	Remove corrupting patterns	*ImageNet*	PSNR SSIM	36.48 dB 0.9809
Chaudhury and Roy [74]	Remove corrupting patterns	*ImageNet* *MSCOCO*	PSNR	Denoising: ($\sigma = 25$) 30.42 dB; ($\sigma = 50$) 27.12 dB Completion: 30.95 dB
Yang et al. [75]	Filling large holes in natural images	*Paris StreetView* *ImageNet*	——	——
Liu et al. [76]	Irregular masks inpainting	*ImageNet* *Places2* *CelebA-HQ*	MAE PSNR SSIM IS	No border: (Mask ratio: PSNR/ SSIM) [1%, 10%]: 33.75 dB/ 0.946 [20%, 30%]: 24.54 dB/ 0.775 [40%, 50%]: 20.34 dB/ 0.583 Border: [1%, 10%]: 34.34 dB/ 0.945 [20%, 30%]: 25.25 dB/ 0.779 [40%, 50%]: 21.38 dB/ 0.595
Deng et al. [77]	Fill missing image regions with plausible contents	*Paris StreetView* *CelebA-HQ* *Place2*	MAE PSNR SSIM	*CelebA-HQ* (Mask ratio: PSNR/ SSIM) [10%, 20%]: 32.84 dB/ 0.981 [20%, 30%]: 29.75 dB/ 0.964 [30%, 40%]: 27.35 dB/ 0.940 [40%, 50%]: 25.43 dB/ 0.909 *Paris StreetView* [10%, 20%]: 31.22 dB/ 0.955 [20%, 30%]: 28.62 dB/ 0.921 [30%, 40%]: 26.62 dB/ 0.872 [40%, 50%]: 24.91 dB/ 0.812 *Places2* [10%, 20%]: 27.85 dB/ 0.950 [20%, 30%]: 25.17 dB/ 0.908 [30%, 40%]: 23.20 dB/ 0.856 [40%, 50%]: 21.52 dB/ 0.789

reconstruction in the input image central area, termed Context Encoder (CE) [78]. The model's weights and bias adjustments were based on the backpropagation of the weighted sum between the adversarial and the reconstruction loss [78]. In terms of results, as the pioneering technique, the proposed model evidenced MAE=9.37%, MSE=1.96% and PSNR=18.58 dB, indicating a revolutionary reconstruction capability compared to the *Best Nearest Neighbour Patch* technique, a sequential-based approach, which yielded MAE=19.92%, MSE=6.92% and PSNR=12.79 dB [78].

Similarly, Yeh et al. [79] introduced an approach in the same year called Structural Inpainting (SI), which differs from CE because it does not require information about the input image holes during the training phase [79]. The network underwent quantitative testing across multiple datasets and various missing mask types, resulting in a broader scope of data evaluation [79]. As an illustration, the proposed model achieved PSNR values of 22.8 dB, 33.0 dB and 18.9 dB for the *CelebA*, *Street View House Numbers* and *Stanford Cars* datasets, respectively, to an input with 80% of missing pixels [79]. Upon a complete analysis, the authors found that, in most cases, CE outperformed SI, except for scenarios involving random masks. However, it was also concluded that SI generated more realistic outputs, concerning image edges at a qualitative level [79].

A year later, Iizuka et al. [80] introduced a model named Global and Local Consistent Image Completion (GLCIC), featuring two networks capable of discriminating between an original and an artificially inpainted image [80]. In this setup, the discriminators operated independently on both the non-information region (local discriminator) and the entire output image of the model (global discriminator) [80]. This particular enhancement to the technique proposed by Pathak et al. [78] resulted in a significant improvement in the global and local consistency of the generated output, as well as in the handling of several rectangular masks randomly scattered throughout the original image [80]. However, one limitation of this study is the absence of a quantitative evaluation based on comparative metrics between techniques [80]. The user study conducted is perceived to be biased and may not offer an accurate interpretation of the approach's feasibility.

To mitigate the emergence of distorted structures observed in methods published until 2018, Yu et al. [81] introduced a novel deep generative approach named Contextual Attention (CA). This approach incorporates structures by leveraging surrounding image features as references for reconstructing multiple holes at arbitrary locations and sizes [81]. The inpainting process involves two stages: the Coarse Network and the Refinement Network [81]. A notable feature of this model is the Contextual Attention layer introduction and the modification of the adversarial loss calculation by incorporating global and local Wasserstein GANs. Through these adjustments, the network achieved MAE, MSE and PSNR values of 8.6%, 2.1% and 18.91 dB, respectively, when tested with the *Places2* dataset [81].

In 2018, Liao et al. [82] introduced a novel approach to CE [78], termed Edge-aware Context Encoder (E-CE), which use information from edge structures as a prior [82]. The E-CE method generates an edge map initially, which preserves main structures and effectively guides the filling of regions with their respective textures [82]. Compared to CE, the integration of edge information led to notable enhancements in metrics such as MAE, MSE, PSNR and SSIM (22.50%, 12.19%, 16.07 dB and 0.549, respectively), enabling precise allocation of missing textures [82].

Building upon a similar ideology, in 2019, Nazeri et al. [83] developed a two-stage model, called Edge-Connected (EC), that divides the inpainting problem into structure prediction and image completion [83]. The study revealed that this model surpassed CA, GLCIC and PConv models across all evaluation metrics in test scenarios with different hole-to-image ratios [83]. Moreover, the research identified a performance improvement in the model with the inclusion of edge information. Specifically, for the *Places2* dataset, the results improved from MAE=6.69, SSIM=0.682, PSNR=19.59 dB and FID=32.18 when edge structure information was absent, to MAE=5.14, SSIM=0.731, PSNR=21.16 dB and FID=14.98 when edge information was incorporated [83].

Zeng et al. [84] introduced an adaptation based on [78] termed the Pyramid-Context Encoder Network (PEN-Net) [84]. The central concept of this model revolved around progressive affinity learning between image regions, which was subsequently transferred from high-level semantic feature maps to the preceding low-level feature map [84]. By employing this approach alongside other structural enhancements, the PEN-Net demonstrated superior performance compared to contemporary techniques at the time of the article's publication (such as PatchMatch, GLCIC, PConv and CA), with metrics including MAE=9.94, Multi-Scale SSIM=78.09%, IS=50.51 and FID=15.19 [84].

In 2019, Sagong et al. [85] and Shin et al. [86] jointly explored an approach to solve the high computational demand problem associated with coarse-to-fine frameworks, analogous to [81, 85, 86]. Both studies advocate for the implementation of a Parallel Extended-Decoder Path for Semantic Inpainting (PEPSI), which reduces the number of convolutions through a shared encoding component, thereby reducing computational time [85, 86]. Moreover, the initial model and its variant, Diet-PEPSI, exhibited superior quantitative results compared to the network introduced by Yu et al. [81]. During the evaluation, the techniques yielded PSNR values of 25.56 dB and 25.5 dB, respectively, along with SSIM values of 0.901 and 0.898 for square masks [85, 86]. Similarly, for free-shape holes, the results were 28.6 dB and 28.5 dB for PSNR, and 0.929 and 0.928 for SSIM [85, 86]. It is noteworthy that the adaptation proposed by Yu et al. [81] did not exhibit a significant decline in reconstruction capacity, thus justifying the enhancement over the method suggested by Shin et al. [85, 86].

In the resume, Table 3 provides a concise overview of the previously mentioned publications.

Table 3: Main topics of research projects related to GAN-based imputation techniques.

AUTHORS	OBJECTIVE	DATASET	EVALUATION METRICS	RESULTS
Pathak et al. [78]	Generate the contents of an arbitrary image region	*Paris StreetView* *ImageNet*	MAE MSE PSNR	9.37% 1.96% 18.58 dB
Yeh et al. [79]	Generate the contents of an arbitrary image region	*CelebA* *Paris Street View* *House Numbers* *Stanford Cars*	PSNR	Dataset 1: 22.8 dB Dataset 2: 33.0 dB Dataset 3: 18.9 dB
Iizuka et al. [80]	Image completion with local and global consistent	*Places2*	——	——
Yu et al. [81]	Inpainting large missing regions in an image	*CelebA* *CelebA-HQ* *DTD* *ImageNet* *Places2*	MAE MSE PSNR TV Loss	8.6% 2.1% 18.91 dB 25.3%
Liao et al. [82]	Inpainting based on edge structures	*Amazon Handbag* *ImageNet*	MAE MSE PSNR SSIM	22.50% 12.19% 16.07 dB 0.549
Nazeri et al. [83]	Inpainting based on edge structures	*CelebA* *CelebHQ* *Places2* *Paris StreetView*	MAE SSIM PSNR FID	Without edge information (MAE/ SSIM/ PSNR/ FID) 6.69/ 0.682/ 19.59 dB/ 32.18 With edge information 5.14/ 0.731/ 21.16 dB/ 14.98
Zeng et al. [84]	Fill missing regions in a damaged image	*Facade* *DTD* *CelebA-HQ* *Places2*	MAE MS-SSIM IS FID	9.94 78.09% 50.51 15.19
Sagong et al. [85]	Image inpainting with coarse-to-fine network	*CelebA-HQ* *ImageNet* *Place2*	PSNR SSIM	Square masks (PSNR/ SSIM) 25.6 dB/ 0.901 Free-shape masks 28.6 dB/ 0.929
Shin et al. [86]	Image inpainting with coarse-to-fine network	*CelebA-HQ* *ImageNet* *Place2*	PSNR SSIM	Square masks (PSNR/ SSIM) 25.5 dB/ 0.898 Free-shape masks 28.5 dB/ 0.928

3.2 Medical Images Specialized Approaches

Following the outcomes achieved with non-specific images, several inpainting techniques swiftly transitioned into the medical imaging field [11]. Their objective was to aid other image post-processing mechanisms or, ultimately, to serve as an additional tool for extracting medical insights. Image imputation techniques found diversified practical applications within the medical context, including medical image completion, non-anatomic artefact removal and segmentation via inpainting. This section will explore each mentioned application individually, citing several promising studies on medical image imputation (Table 4, 5 and 6).

3.2.1 Medical Image Completion

Pioneering the inpainting techniques appliance on medical datasets, Tran et al. [87] compared three GAN-based frameworks: CE [78], SI [79], CA [81]; using 1.2 million patches of size 128 × 128 extracted from 60 thousand healthy frontal chest X-ray images (*ChestXray14* dataset) [87]. The article assessed the networks' capability to reconstruct a small hole of size 64 × 64 in the centre of images from patients with both healthy and abnormal conditions [87]. Quantitatively, the PSNR values indicated a slight superiority of SI (33.85 ± 4.67 dB), followed by CA (31.80 ± 5.19 dB) and finally CE (26.31 ± 4.48 dB) for 880 healthy inpainted patches [87]. The same trend was observed for images containing anomalies, even though the results were lower (SI with PSNR=30.18 ± 3.28 dB, CA with PSNR=26.79 ± 3.47 dB and CE with PSNR=22.22 ± 4.26 dB), as expected [87]. Nevertheless, a global consistency was visually observed in the CA approach, as evidenced by a visual-perception study, with a value of 37.03%, assuming 100% accuracy indicates the observer can perfectly identify the unaltered chest X-ray in a pair [87].

In 2019, recognizing the extensive utility of image inpainting in the medical field, Armanious et al. [88] introduced a new technique inherent to medical modalities reconstruction, such as MRI and CT, called ip-MedGAN [88]. The proposed framework, based on a Conditional GANs, incorporated two patch-based discriminators with additional style and perceptual losses, allowing a detailed and contextually consistent filling of a 64 × 64 pixel square region within the 256 × 256 image [88]. These non-adversarial losses were calculated from the *VGG-19* feature maps extraction [88]. This study used 50 volunteers for CT image acquisition and 44 T2-weighted datasets of the head region [88]. After testing the generated models, it was concluded that both CE and GLCIC presented imputed regions with blurriness and lacked sharpness [88]. Alternatively, MedGAN (image-to-image GAN) showed decent results, although with some failures in image details and the creation of tiny unrealistic artefacts [88]. The introduced technique, in turn, achieved the best results for both datasets under study, revealing a PSNR of 31.45 dB, 18.32 dB, an SSIM of 0.8346, 0.3818 and Universal Quality Image Index (UQI) of 0.974, 0.926 for CT and MRI scans, respectively [88]. Despite that, the method revealed two main limitations: the manual segmentation of non-informative areas requirement during training and the restriction of fixed shape holes [88].

A year later, the same authors proposed adapting their previous model to make the framework, named ipA-MedGAN, capable of reconstructing arbitrary missing regions, without prior knowledge of their location [89]. Similar to the earlier study [88], this new approach was trained with 3028 slices from 33 MRI acquisition volunteers and tested with 1072 slices from 11 other volunteers [89]. In the initial phase, the present framework was compared with functional techniques only for reconstructing square regions and subsequently for arbitrary missing with a 1.36% to 5.46% total image size range [89]. In all cases, ipA-MedGAN exhibited better performance for free-shape holes (PSNR=35.12 dB; SSIM=0.9818; UQI=0.989), eliminating artefacts at the border without requiring prior information about the masked regions [89]. The authors of this article encouraged future studies to investigate the effect of these inpainting techniques on the performance of real-world MRI processing applications and to introduce spatial information relevant to image reconstruction [89].

Conversely, Wang et al. [90] investigated the applicability of a model based on medical image edge and structure information, especially for CT and MRI images [90]. This proposal, named Edge and Structure Information for Medical Image Inpainting (ESMII) and emerged in 2021, takes advantage of prior information from the edges obtained by the *Canny* algorithm and the structure obtained by removing textures [90]. Simultaneously, to avoid training degradation issues, image features at three scales are extracted through a Multi-Scale Residual Block [90]. The model was trained on three distinct datasets: COVID CT (275 CT images from a COVID-19 study); Abdominal CT (images from 40 patients in the healthy abdominal organ segmentation dataset); Abdominal MRI (120 DICOM datasets from 2 different MRI sequences, also from the healthy abdominal organ segmentation dataset) [90]. During the studies for square and arbitrary missing patterns reconstruction, the ESMII model's results were generally superior for all datasets to all calculated metrics: MSE, SSIM, PSNR, FID and IS; when compared with imputation techniques, such as GLCIC, CA, PConv and EC [90]. Regarding the COVID CT dataset, it is worth noting that the previously measured metrics values were SSIM=0.889 and PSNR=34.442 for centred masks and SSIM=0.939 and PSNR=35.202 for irregular masks. In addition, the article conducted a procedure to understand the effect of edge and structure information on the reconstructed images, concluding that they significantly contributed to the imputation process, avoiding the output image diffusion and blurring [90]. Overall, the model proved to be a potential approach for a preprocessing stage of these medical modalities types, serving as an automatic artefact-free tool and enhancing the performance of segmentation and classification post-processing models [90].

A similar approach was also developed by Tran et al. [91]. However, the input to the framework was exclusively the masked image because the information related to edges and organ boundaries was obtained only during the decoding phase of the respective network. This study used the *StructSeg2019* dataset, with 50 volumes of CT scans from 50 patients, where it was exploited to 4775 2D usable images. Regarding the presented results, compared to the ESMII method, this model showed a PSNR of 43.44 dB for square-shaped and 38.06 dB for arbitrary-shaped masks, while the ESMII framework presented a PSNR value of 33.88 dB for abdominal CT images with square holes in the centre. As advised, it would be interesting to parallel the last two referenced studies for future research endeavours and expand and refine the Tran et al. [91] study for other medical image modalities.

Very recently, Yuan and Dai [92], following preceding works, developed a missing areas filling method in medical images, ensuring structural, textural and transitional information in boundary regions, named MII-DFNet [92]. This framework was based on a U-net, with an associated fusion block in its decoder, inspired by the article [93]. This research used open-source datasets from The Cancer Imaging Archive (TCIA), selecting CT and MRI images of brain, lung and abdomen diseases [92]. One weakness of the framework, which should be addressed in the future, is the lack of quantitative metrics between compared techniques [92]. Nevertheless, MII-DFNet solves the problem of the boundaries discontinuous at the input image holes, clearly visualized with the EC network [92].

3.2.2 Non-Anatomical Artefact Removal

Based on the results of medical image completion, many authors began researching the contribution of these approaches to limiting non-anatomic artefact interference in imaging exams. As expected, the first records of these applications were based on traditional methodologies, such as those used by Duan et al. [94] and Peng et al. [95].

However, during the 2019 year, an early pioneer learning-based approach was introduced by Li et al. [96]. The main idea was to remove cross symbols manually placed by doctors for thyroid nodule location in ultrasound images. Recognizing how these artefacts significantly affected computational diagnostic mechanisms, Li et al. [96] developed a Pyramidal GAN (Py-GAN) to reconstruct texture and structure in artifact-affected regions. The study used 1550 ultrasound images of thyroid nodules provided by the Tianjin Medical University Cancer Institute and Hospital, with 1130 images for model training and 420 for testing [96]. Similar to other studies, Li et al. [96] expanded the training dataset by 10 times factor. Notably, the proposed framework was preceded by autonomous cross symbols detection and removal mechanisms [96]. In the testing phase, in addition to the reconstruction quality comparison between different techniques, the authors investigated the pyramidal structure influence after manipulating the training loss function [96]. Py-GAN demonstrated better qualitative and

Table 4: Main topics of research projects related to Medical Image Completion techniques.

AUTHORS	**OBJECTIVE**	**DATASET**	**EVALUATION METRICS**	**RESULTS**
Tran et al. [87]	Inpainting in medical imaging domain	1.2M 128 × 128 patches from 60K healthy X-rays (ChestXray14)	PSNR	Normal Patches: SI: 33.85 ± 4.67 CE: 26.31 ± 4.48 CA: 31.80 ± 5.19 Abnormal Patches: SI: 30.18 ± 3.28 CE: 22.22 ± 4.26 CA: 26.79 ± 3.47
Armanious et al. [88]	Inpainting of medical images via GANs	50 CT volumes and 44 T2-weighted MRI volumes	SSIM PSNR MSE UQI	CT (SSIM/ PSNR/ MSE/ UQI) 0.8346/ 31.45 dB/ 284.4/ 0.974 MRI 0.3818/ 18.32 dB/ 1121.2/ 0.926
Armanious et al. [89]	Completion of the missing or distorted regions	44 T2-weighted MRI volumes	SSIM PSNR MSE UQI	Square mask (SSIM/ PSNR/ MSE/ UQI) 0.9606/ 30.62 dB/ 65.81/ 0.962 Free-shape mask 0.9818/ 35.12 dB/ 24.90/ 0.989
Wang et al. [90]	Repair the structure rationally based on edge and structure information for medical image	275 COVID CT, 40 abdominal CT and 120 Abdominal MRI DICOM	MSE SSIM PSNR FID IS	COVID CT (MSE/ SSIM/ PSNR/ FID/ IS) **[Centered hole]:** 17.828/ 0.889/ 34.442 dB/ 97.822/ 2.355 **[Irregular hole]:** 15.152/ 0.939/ 35.202 dB/ 56.753/ 2.483 Abdominal CT **[Centered hole]:** 23.595/ 0.882/ 33.881 dB/ 47.959/ 1.940 Abdominal MRI **[Irregular hole]:** 12.850/ 0.965/ 36.398 dB/ 41.529/ 1.929
Tran et al. [91]	Repair the structure rationally based on edge and structure information for medical image	50 CT scans (converted into 4775 2D images) from *StructSeg2019*	PSNR SSIM MSE UQI	Square masks (PSNR/ SSIM/ MSE/ UQI) 43.44 dB/ 0.9818/ 37.93/ 0.9960 Free-shape masks 38.06 dB/ 0.9746/ 50.49/ 0.9972
Yuan and Dai [92]	Restore original information of medical images covered by masks	CT and MRI images from TCIA	——	——

quantitative results, achieving PSNR values of 25.54 dB and 25.73 dB for inpainting central and marginal regions, respectively. In comparison, techniques like GLCIC (PSNR equal to 20.51 dB and 20.70 dB), CA (PSNR equal to 23.82 dB and 24.21 dB) and PConv (PSNR equal to 22.16 dB and 23.32 dB) showed a lower capacity for realistic and natural image recovery [96].

In 2020, following traditional techniques for removing metal artefacts in CT images, commonly referred to as Metal Artifact Reduction (MAR), Peng et al. [97] developed a framework to reconstruct irregular metal traces directly on sinograms, followed by a FBP for artefact-free CT image visualization [97]. This framework utilized a Partial Sub-Network (PN) to complete the cropped sinogram, followed by concatenation with the result obtained using an Auxiliary Inpainting Subnetwork (AIN), aiming to gather more information in the image domain [97]. The model training was executed using CT images with dimensions of 512 × 512, with synthetically generated metal artefacts for four different types of datasets, totalling 9218 samples [97]. In the CT sinogram domain, Peng et al. [97] considered 1024 as the number of detector channels and 720 as the projection perspectives, resulting in processed sinograms with 720 × 1024 dimensions. Despite achieving well-performed results compared to existing techniques, this framework exhibited limitations, such as not effectively reducing metal artefacts for CT images with complex bone and tissue structures [97].

In the same year, an approach, published by Pimkin et al. [98], focused on the adverse effects that embolic agents could have on radiotherapy dosage calculations, during the treatment of pathologies, such as Cerebral Arteriovenous Malformations (AVMs) [98]. Previously, professionals had to manually segment errors in each slice of CT volumes, a non-efficient and time-consuming process [98]. As a result, Pimkin et al. [98] proposed a model capable of automatically removing artificially generated artefacts, using a sinogram space completion process, where the input algorithm resulted from the difference between the original image sinogram and the artefacts masked image sinogram [98]. After the inpainting loop, followed by projection into the CT image domain, the framework used a training cycle to eliminate residues resulting from the artefact removal process. The outcomes showed successful artefact removal, with MAE, MSE and SSIM values of 9.6 HU, 831 HU^2 and 0.99, correspondingly, after comparing the original artefact-free image and the reconstructed image post-artefact addition [98]. Furthermore, it was observed that this learning-based method exhibited a substantial MSE drop compared to the traditional Li-MAR technique [98].

In 2021, to counter potential errors from the detector ring, which introduce distortions in PET images and consequently affect the results of post-processing algorithms, Amirrashedi et al. [99] investigated the modified U-Net feasibility to generate artefact-free PET scans in the presence of several dead regions between adjacent detector blocks on a high-resolution preclinical PET scanner [99]. Differently from the previous articles, this research was conducted on PET images of 30 mice, resulting in

2000 samples for training, 270 for validation and 280 for testing [99]. In terms of results, this article did not compare its performance with other existing inpainting techniques but, instead, promoted a comparison between the effectiveness of applying the U-Net on the image domain or the sinogram domain [99]. For this, the authors used the metrics PSNR, SSIM and Root Mean Squared Erro (NRMSE), finding that sinogram completion (PSNR=45.87 ± 5 dB; SSIM=0.98 ± 0.01) was more effective than image domain completion (PSNR=44.83 ± 5.1; SSIM=0.97 ± 0.01). Despite the projection domain learning advantage, Amirrashedi et al. [99] indicated that it is computationally more demanding when compared to image domain learning [99].

From a different perspective, Xia et al. [100] proposed the I2-GAN to restore lost or altered information due to inter-slice motion artefacts by learning the key features of the *Cardiac Short Axis* (SAX). This approach was initially designed for a 3D regression network to predict the missing slice position and extract relevant features in the missing slice neighbourhood. Following, the GAN's generator model took the pre-computed intrinsic slice features as conditioning input to synthesize a cardiac MRI slice at the corresponding position. Finally, a multi-scale discriminator is used to distinguish the generated samples from the real ones, simultaneously matching the features of the inferred slices with those of the original samples. For this purpose, the framework used cardiac MRI images from the *UK Biobank* as training and validation data, with the subsequent specifications: in-plane spatial resolution of 1.8 × 1.8 mm, slice thickness of 8 mm, slice gap of 2 mm and image size of 198×208. In the testing phase, images from 100 subjects of the *Automatic Cardiac Diagnosis Challenge* (ACDC) dataset were used. However, before applying I2-GAN, a preprocessing step was performed, resizing the slices to 256×256 and removing random slices to simulate the presented problem. In terms of results, this article demonstrated several steps showing the effectiveness of this approach compared to existing techniques. In the Intra-Phase evaluation, the model's completion ability was studied using SSIM and PSNR metrics and the statistical effects of the restored slice incorrect estimation. It was found that I2-GAN generates visually comparable results to the reference slices and yields more plausible results in terms of preserving fine structural and textural details. However, after the reconstructive process, it verified a smooth inaccurate slice position estimation. Subsequently, in the Inter-Phase evaluation, the model was evaluated to different cardiac phases and didn't demonstrate any significant differences in the synthesizing slice quality. Finally, Xia et al. [100] evaluated the test results on the testing dataset and promoted an ablation study, applying a post-processing method for calculating *LV/RV* mass and volume after the image imputation process.

Two years later, Xie et al. [101] introduced a model based on a Patch GAN, whose generator had a U-net structure with a dual-path encoder, to extract metallic artefacts from input images [101]. This framework used original MRI and flipped registered images as inputs from the study dataset

[101]. The data corresponded to head MRI scans of 100 patients (61 males and 39 females), realized from 2017 to 2021 by the Hospital of Changzhou [101]. Subsequently, the article's authors divided the dataset into training, validation and test sets with 78, 12 and 12 patients, respectively and introduced an augmentation mechanism (such as horizontal or vertical flips) to avoid overfitting, increasing the available slices for model learning from 3000 to 9000 samples [101]. Adopting an evaluation approach based on several metrics: MAE, PSNR, SSIM, FID and SAMScore; it was evidenced that the proposed technique outperformed other methods, such as PConv and Gated Convolution (GConv), in quantitative results. This became particularly evident in masked MAE and masked PSNR values, where the proposed methodology achieved 0.1834 and 17.60 dB, respectively, compared to 0.1938 and 17.39 dB for PConv, and 0.1904 and 17.40 dB for GConv [101]. Additionally, it was shown through the developed statistical tests that the observed differences between metrics values were almost entirely significant, reinforcing the completion quality of the model proposed by Xie et al. [101].

In the same year, the authors of the previously described work proposed a divergent approach, which aimed to reconstruct areas omitted due to FOV artefacts in CT exams to determine if this procedure could help mitigate dosimetric calculation errors [102]. The research involved data from patients with oesophagal cancer, with 85 CT scans allocated for the training and 15 for the testing phase. The imputation models compared: U-Net, pix2pix, PConv and GConv; were evaluated regarding their reconstructive capacity and dosimetric performance [102]. The first round of evaluations utilized metrics such as MAE, PSNR, SSIM and DICE. Among the top results, GConv stood out, achieving MAE of 3.71 ± 1.40, PSNR of 33.77 ± 2.47 dB, SSIM of 98.96 ± 0.60 % and DICE of 94.63 ± 3.77% for the global image [102]. For the reconstructed region, GConv achieved MAE of 158.45 ± 31.65 HU and PSNR of 18.01 ± 2.59 dB. From a dosimetric perspective, it was found that the dose distribution in the images reconstructed by GConv was identical to that calculated for the ground truth images, meaning an improvement in the results for images with FOV limitation artefacts [102].

Also in 2023, Xu et al. [103] investigated reconstructing missing areas in chest CT scans caused by pose limitations. Their approach differed from projection domain applications by employing a dual mechanism: first estimating the delineation of body structures to be reconstructed and then recreating the truncated regions. Two datasets of non-artefact CT images of lung cancer were used for this study: the National Lung Screening Trial (NLST) and the Vanderbilt Lung Screening Program (VLSP). The NLST dataset included 3586 scans from 1280 patients and the VLSP dataset consisted of 1490 scans from 887 patients [103]. As a result, the limited FOV artefacts were artificially applied to these images using specific masks, allowing their use as input of the imputation models compared: pix2pix, PConv and Recurrent Feature Reasoning (RFR). At the evaluation, metrics such as the square root of the MSE (RMSE) were used, having recorded 7.48 ± 4.48, 6.52 ± 4.05 and 6.12 ± 4.09 HU, respectively to the order presented previously, and also the DICE with values of 0.96 ± 0.04, 0.96 ± 0.04 and 0.97 ± 0.03 %. Additionally, Xu et al. [103] utilized a visual assessment conducted independently by two experts and applied a deep learning Body Composition pipeline. Notably, the second evaluation method revealed improvements in intra-subject consistency by reducing errors associated with truncated images [103].

Similarly, a year later, Kim et al. [62] conducted a study using $10,005$ slices from 322 lung cancer patients for training their imputation models and 166 slices from 15 patients for testing, sourced from TCIA. The study explored three approaches: the standalone GLCIC framework, the Patient Body Outline (PBO) method and a hybrid algorithm combining both methods [62]. To assess the performance of the imputation model, Kim et al. [62] employed metrics such as the Root of MSE (RMSE), PSNR and SSIM. In the final phase of training, the model achieved a RMSE of 7.4, a PSNR of 30.9 dB, and an SSIM of 0.97 for the global image [62]. Additionally, the study included a dosimetric evaluation, revealing that the hybrid approach improved the precision and effectiveness of radiotherapy, as the planned dose more closely matched the applied dose [62].

3.2.3 Inpainting-based Segmentation

One of the initial approaches, which demonstrates the complementation of segmentation models by inpainting mechanisms, was published by Manjón et al. [104]. This research developed a model capable of automatically removing brain lesions through image inpainting, consequently improving the *SPM12* pipeline performance, used for several brain regions segmentation [104]. The model training step used two datasets representing brain MRI images with and without Multiple Sclerosis lesions, enabling the implemented approach's effects comparison[104].

In 2021, to segment brain tumours, Nguyen et al. [105] proposed an inpainting-based framework, fully autonomous and unsupervised segmentation, from T1-weighted MRI samples [105]. The proposed system considered an initial imputation training block, based on Deep CNN, to reconstruct missing healthy brain regions, followed by a localization system to image hole identification [105]. After the previous step, a super-pixel segmentation and a heat map creation were implemented, resulting in highly successful structural delineation compared to the ground truth data [105]. During the imputation processor, several missing area sizes were tested. The research also found that the best segmentation was achieved for a 32×32 pixel window reconstruction, with 35.63 ± 5.413 and 0.97 ± 0.002, as PSNR and SSIM values, respectively.

Table 5: Main topics of research projects related to Non-Anatomical Artefact Removal imputation-based techniques.

AUTHORS	OBJECTIVE	DATASET	EVALUATION METRICS	RESULTS
Li et al. [96]	Recovering the thyroid ultrasound image with special cross symbols	1550 thyroid nodules ultrasound from Tianjin Medical University Cancer Institute and Hospital	PSNR	Central Masks: 25.54 dB Border Masks: 25.73 dB
Pimkin et al. [98]	Synthetic data generation pipeline to avoid problems with overfitting to metal shapes set and an artefacts formation technique	163 CT scans of the patients treated with radiosurgery for trigeminal neuralgia	MAE MSE SSIM	9.6 (HU) 831 (HU^2) 0.99
Peng et al. [97]	Recover corrupt regions generated by metal artefacts in CT images	9218 CT images with synthesized metal artefacts	PSNR SSIM	Panoramic Dental CT ([N metals] PSNR/ SSIM) [1] 30.1739 dB/ 0.9012 [2] 30.7217 dB/ 0.9003 [3] 30.0722 dB/ 0.8746 Local Dental CT [1] 36.4779 dB/ 0.9909 [2] 36.4779 dB/ 0.9872 [3] 36.4779 dB/ 0.9869 Hip CT [1] 20.6914 dB/ 0.8919 [2] 17.4508 dB/ 0.9497 Vertebra CT 30.0635 dB/ 0.6852
Amirrashedi et al. [99]	Impaired sinograms and removing the streaking artefacts	30 mice PET scans (85 slices/scan)	NRMSE PSNR SSIM	Sinogram completion (NRMSE/ PSNR/ SSIM) 0.051 ± 0.01/ 45.87 ± 5 dB/ 0.98 ± 0.01 Image domain completion 0.054 ± 0.009/ 44.83 ± 5.1 dB/ 0.97 ± 0.01

* Continued on next page

Table 5 – continued from previous page

AUTHORS	OBJECTIVE	DATASET	EVALUATION METRICS	RESULTS
Xia et al. [100]	Recover of missing or unusable slices owing to the presence of image artefacts	Cardiac MRI from the UK Biobank and ACDC	PSNR SSIM	26.88 ± 1.63 dB 0.872 ± 0.027
Xie et al. [101]	Generate normal MRI images from MRI scans with metallic artefacts	T1-weighted MRI images from 100 patients were collected (9000 slices after data augmentation)	MAE PSNR SSIM FID SAMscore	$(18.34 \pm 2.75) \times 10^{-2}$ 17.62 ± 0.96 dB $(98.63 \pm 0.46)\%$ 6.92 ± 0.39 $(99.69 \pm 0.10)\%$
Xie et al. [102]	Reconstruct omitted areas due to FOV artefacts via GConv	100 CT volumes from patients with esophageal cancer	MAE PSNR SSIM DICE	3.71 ± 1.40 HU 33.77 ± 2.47 dB 98.96 ± 0.60 % $94.63 \pm 3.77\%$
Xu et al. [103]	Reconstruct missing areas in chest CT scans caused by pose limitations	National Lung Screening Trial (NLST): 3586 scans from 1280 patients Vanderbilt Lung Screening Program (VLSP): 1490 scans from 887 patients	RMSE DICE	[pix2pix] 7.48 ± 4.48 HU / 0.96 ± 0.04 % [PConv] 6.52 ± 4.05 HU / 0.96 ± 0.04 % [RFR] 6.12 ± 4.09 HU / 0.97 ± 0.03 %
Kim et al. [62]	Reconstruct omitted areas due to FOV artefacts adding a hybrid approach with Patient Body Outline (PBO)	10005 slices from 322 lung cancer patients for training and 166 slices from 15 patients for testing	RMSE PSNR SSIM	7.358 30.916 dB 0.975

An even more recent example is the research by Lu et al. [106], published in 2023. This study was motivated by introducing segmentation mechanisms with a trade-off between performance and label dependency. Therefore, a framework was developed exclusively dependent on bounding box labels for precise segmentation of COVID-19 lesions in CT images [106]. The authors developed a model, designated as Faster RCNN, responsible for creating boxes around potential lung lesions, followed by the box-enclosed information removal [106]. Subsequently, the Initial Patch Generation Network (IPGN) was implemented to establish the primary fill of the previously generated holes with random noise [106]. The newly filled images then underwent the action of the Progressive CT Inpainting Network (PCIN) from a dataset of healthy patients, along with the action of the Structure Extraction Module (SEM) [106]. The model concluded that the difference between the ground truth image (with the lesion) and the inpainted image (without the lesion) resulted in the affected tissue demarcation. Lu et al. [106] developed their approach based on two public datasets: *Zenodo*, with 20 COVID-19 CT scans with lung lesions and segmentation labels performed by two qualified radiologists, and *MosMedData* containing CT images from 1110 patients, where only 50 healthy and 50 COVID-19-infected individuals were used. As a result, the authors tested multiple proposed framework subsystems [106]. In terms of the inpainting process, the PCIN was quantitatively compared using MAE, PSNR and SSIM with other well-known techniques (EC, CE, PConv and GConv), as well as different initialization parameters for regions without information, showing the best result (MAE=14.93, PSNR=21.47 dB and SSIM=0.4182) when using the PCIN with holes previously filled by the IPGN [106]. Additionally, the missing reconstruction performance was evaluated based on the missing data ratio [106]. Contrary to expectations, MAE, PSNR and SSIM values showed better results with an increase of the regions requiring reconstruction size [106]. One of the unambiguous model's weaknesses was its high dependence on the initial lesions boxes detection network because an error in this mechanism immediately avoids the subsequent segmentation procedure [106].

3.3 Main Findings

Current developments in image imputation have shown a near-exponential increase in the use of DL algorithms for this problem. Among these algorithms, CNNs and GANs stand out as the most common architectures, frequently applied in various contexts. Despite their structural differences, both architectures generally follow a similar approach: using downsampling and upsampling mechanisms to reconstruct missing or distorted portions of an image.

From an overall perspective, the review revealed that CNN-based approaches were the first type of architecture applied to image imputation tasks. Since then, there has been a growing number of proposals based on these methods, with various optimizations introduced at diverse levels, ranging from modifying the network layers' architecture to reconstructing missing or corrupted regions with irregular shapes. Due to the inherent simplicity of CNN-based networks, which often leads to outputs with limited detail, several studies have proposed mechanisms to counteract this issue. Notably, one common enhancement involves replacing standard convolutional layers with attention modules with multi-scale and multi-head attention mechanisms to improve detail preservation and image quality.

Still on overall applications, GANs, despite their later emergence compared to CNN-based techniques, have become the most extensively studied approaches in image inpainting. They have demonstrated the ability to overcome the primary limitations of CNN-based methods — generating images with low-level detail — by ensuring more realistic and visually convincing results. In evolutionary terms, GAN-based algorithms started with basic architectures consisting of two feed-forward networks, a generator and a discriminator, trained through an adversarial competition mechanism. With

Table 6: Main topics of research projects related to Medical Image Segmentation, based on imputation techniques.

AUTHORS	OBJECTIVE	DATASET	EVALUATION METRICS	RESULTS
Manjón et al. [104]	Blind inpainting lesions in brain images automatically allowing current pipelines to robustly operate under pathological conditions	298 healthy MRI from *IXI* dataset; 43 MRI with manual segmentation of Multiple Sclerosis (MS) lesions; 200 MRI with automatic segmentation of MS lesions	PSNR Correlation Coefficient —— DICE	Blind inpainting (PSNR/ Correlation Coef.) [**MS cases**] 49.57 dB/ 0.9998 [**Normal cases**] 44.97 dB/ 0.9999 Segmentation (DICE) [**Without inpainting**] 0.9718 ± 0.0172 [**With inpainting**] 0.9834 ± 0.0128
Nguyen et al. [105]	Fully automatic, unsupervised inpainting-based brain tumour segmentation	125 T1-weighted MRI scans of normal brain tissue from *Neurofeedback* and 22 T1-weighted MRI scans provided by the Centre for Clinical Brain Sciences from the University of Edinburgh	PSNR SSIM DICE	([Mask size] PSNR/ SSIM/ DICE) [**8px**] (38.61 ± 3.159) dB/ 0.99 ± 0.006/ 0.49 ± 0.396 [**16px**] (37.94 ± 4.664) dB/ 0.98 ± 0.022/ 0.66 ± 0.346 [**32px**] (35.63 ± 5.413) dB/ 0.97 ± 0.002/ 0.77 ± 0.176 [**64px**] (31.66 ± 6.491) dB/ 0.96 ± 0.003/ 0.55 ± 0.380
Lu et al. [106]	Weakly supervised inpainting-based learning for accurate segmentation	20 COVID-19 CT scans with lung and lesion segmentation from *Zenodo* and 50 normal cases and 50 COVID-19 cases from *MosMedData*	MAE PSNR SSIM	14.93 21.47 dB 0.4182

the optimization of these applications, various improvements have been observed compared to the baseline model implemented by Pathak et al. [78]. These include the development of more complex dual discriminators, the introduction of generators based on coarse-to-fine refinement mechanisms and priors' addition, such as edges, as guiding elements to the reconstruction process. However, GANs generally require more computational resources than CNNs and present greater challenges in stabilizing the training phase due to the adversarial nature of the interaction between the generator and discriminator networks.

In a medical context, the primary applications of these techniques have been in CT and MRI imaging, with some studies also exploring their use in ultrasound, conventional X-rays and PET samples. These algorithms have been particularly developed to remove non-anatomic artefacts, such as cross symbols, streaking artefacts caused by metal, motion artefacts and imaging issues related to FOV and mispositioning artefacts. Image imputation techniques have generally demonstrated good efficacy in removing such errors, resulting in an overall increase in image quality. Beyond artefact removal, studies exploring image completion tendentiously have aimed to understand the broader applications of inpainting in medical imaging. More recently, there has been a trend toward incorporating these techniques into more complex algorithms, such as segmentation models, where image imputation serves as an intermediate step in the processing pipeline.

Despite these advancements, several limitations remain, suggesting directions for future research. A key challenge is the development of blind inpainting mechanisms that can identify and reconstruct only the regions needing restoration, as current approaches generally rely on human intervention for accurate localization. Additionally, there is a need for techniques capable of removing multiple artefacts simultaneously within a single sample, as current methods are often designed to address specific artefact types. Contrary to current approaches, which operate uniformly regardless of the context or specific area needing reconstruction, a promising future direction in medical imaging could involve developing hybrid mechanisms that adapt to the particular characteristics of the tissue being reconstructed, potentially enhancing both the effectiveness and efficiency of the reconstruction process of medical images. Finally, with the current emphasis on 2D applications, extending these models to 3D and even 4D medical data would represent a significant advancement, augmenting their usefulness in medical imaging scenarios.

In summary, the current review highlights that image imputation is a growing field that needs special attention. The results have been promising, especially within medical applications, where its use has been extensive, diverse and impactful.