

## Article

# Multi-Task Learning for Medical Image Inpainting Based on Organ Boundary Awareness

Minh-Trieu Tran , Soo-Hyung Kim, Hyung-Jeong Yang  and Guee-Sang Lee \*

Department of Artificial Intelligence Convergence, Chonnam National University, 77 Yongbong-ro, Gwangju 500-757, Korea; tmtvaa@gmail.com (M.-T.T.); shkim@jnu.ac.kr (S.-H.K.); hjyang@jnu.ac.kr (H.-J.Y.)  
\* Correspondence: gslee@jnu.ac.kr

**Abstract:** Distorted medical images can significantly hamper medical diagnosis, notably in the analysis of Computer Tomography (CT) images and organ segmentation specifics. Therefore, improving diagnostic imagery accuracy and reconstructing damaged portions are important for medical diagnosis. Recently, these issues have been studied extensively in the field of medical image inpainting. Inpainting techniques are emerging in medical image analysis since local deformations in medical modalities are common because of various factors such as metallic implants, foreign objects or specular reflections during the image captures. The completion of such missing or distorted regions is important for the enhancement of post-processing tasks such as segmentation or classification. In this paper, a novel framework for medical image inpainting is presented by using a multi-task learning model for CT images targeting the learning of the shape and structure of the organs of interest. This novelty has been accomplished through simultaneous training for the prediction of edges and organ boundaries with the image inpainting, while state-of-the-art methods still focus only on the inpainting area without considering the global structure of the target organ. Therefore, our model reproduces medical images with sharp contours and exact organ locations. Consequently, our technique generates more realistic and believable images compared to other approaches. Additionally, in quantitative evaluation, the proposed method achieved the best results in the literature so far, which include a PSNR value of 43.44 dB and SSIM of 0.9818 for the square-shaped regions; a PSNR value of 38.06 dB and SSIM of 0.9746 for the arbitrary-shaped regions. The proposed model generates the sharp and clear images for inpainting by learning the detailed structure of organs. Our method was able to show how promising the method is when applying it in medical image analysis, where the completion of missing or distorted regions is still a challenging task.



**Citation:** Tran, M.-T.; Kim, S.-H.; Yang, H.-J.; Lee, G.-S. Multi-Task Learning for Medical Image Inpainting Based on Organ Boundary Awareness. *Appl. Sci.* **2021**, *11*, 4247. <https://doi.org/10.3390/app11094247>

Academic Editors: Hee-Deok Yang and Manuel Armada

Received: 8 March 2021

Accepted: 30 April 2021

Published: 7 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** multi-task learning; medical image inpainting; medical image analysis; deep learning; arbitrary-shaped inpainting; medical prognosis

## 1. Introduction

Computed Tomography (CT) has been one of the essential medical imaging systems and utilized for expert diagnoses. However, the CT images are often distorted by reflection from metallic implants or foreign objects such as pacemakers, catheters, and drainage tubes. Moreover, medical images are sometimes degraded due to the sudden movements of the patient during the scanning phase. Many approaches have been proposed for the restoration of deformed images, which include research results on noise reduction, image translation, or inpainting. Among these methods, inpainting has emerged as a reasonably effective and popular method today. Several studies on medical image inpainting have been proposed, including the technique of handling damaged square-shaped regions [1,2]. However, in the real situation, the defects are mostly not of the squares, but of arbitrary-shaped regions, which launched the study of medical image inpainting with any damaged forms [3], resulting in the restoration of practical failures with any deformation. These techniques still suffer from incomplete restorations, such as blurred boundaries and loss of

the organ structures inside the deformed part. To overcome such problems of restoration failures, structural information has been exploited, where the information of edges is the main tool for implementing the learning process [4–8]. EdgeConnect [4] is a two-stage adversarial model that comprises an edge generator followed by an image completion network. The edge generator hallucinates the edges in the missing area (either square-shaped or arbitrary-shaped). The image completion network fills in the distorted regions using hallucinated edges as a priori for the inpainting. Edge structures and color-aware maps are fused in a two-stage generative adversarial network (GAN) [5]. In the first stage, edges with the missing regions are used to train an edge structure generator. Meanwhile, distorted input images with the missing part are transformed into a global color feature map by the content-aware fill algorithm. In the second stage, the edge map and the colormap are fused to generate the refined image. The authors in [6] proposed a foreground-aware image inpainting system that explicitly disentangles structure inference and content completion. The foreground contours are predicted first and then the inpainting is performed using the predicted contours as a guidance. The Edge-Guided GAN [7] method is an edge-guided generative adversarial network to restore brain MRI images which have distortion or missing parts. A multi-task learning framework with auxiliary tasks of edge finding and gradient map prediction is used to incorporate the knowledge of the image structure to assist inpainting [8]. Even though the edges in the image provide a part of structural information, there are several drawbacks in such methods. First, the edges are obtained from any objects in the image and they do not represent the spatial structures of the organs of interest. Second, the edges are more complex than the specific organ descriptions and this could make it harder for the model to understand the structure of organs. Therefore, we believe that the edges alone cannot provide sufficient knowledge of organ structures in the body, resulting in the still poor restoration quality.

The proposed method is trained and predicted in an end-to-end framework. Our contributions can be summarized as follows:

- We propose a framework based on edge and organ boundary awareness to reconstruct deformed regions in CT images.
- We newly introduced the use of organ boundaries in addition to edges to establish enough structural knowledge for the inpainting of damaged regions, including the part of the organs. Specifically, multi-task learning is employed to train the network simultaneously for the prediction of edges and organ boundaries. The use of organ boundaries for the learning of structural information in medical image inpainting has never been tried before, and it is adopted for the first time in the literature.
- Our method generates more realistic and believable images compared to other approaches. In both quantitative and qualitative evaluation, the proposed method outperforms the state-of-the-art methods.

The rest of the paper is organized as follows. In Section 2, we introduce the related literature in the field of general inpainting and medical inpainting. The details of our architecture are presented in Section 3. The experimental results are given in Section 4. Finally, conclusions are shown in Section 5.

## 2. Related Works

### 2.1. Inpainting in General Field

For inpainting images, we can put them into two main groups: traditional and learning-based approaches. The traditional methods employ diffusion-based or patch-based methods with low-level features, while the learning-based approaches try to understand the semantics of the image to fulfill the inpainting task. The success of deep learning has made the second approach effective and very popular in recent years. We introduce the details of studies based on both approaches in the following sections.

### 2.1.1. Traditional Approach

With conventional methods, the algorithms try to find components from the background area, then compute the similarity levels and fill in the hole. It uses the information available in the image containing the deformed part to generate the missing area [9,10], which provides a simple algorithm, responding relatively well to the inpainting of small areas in case the scene is not too complicated. Conventional methods also do not require a high amount of training data. However, in some cases, such as large or arbitrary-shaped holes possibly with the background of complex structures, these methods can fail to produce a good recovery.

### 2.1.2. Learning-Based Approach

In learning-based approaches, different types of features can be learned from a large spectrum of sample images, leading to better predictions compared to conventional methods. The deep learning approaches have been studied extensively, and the results showed a significant improvement in the performance. The context encoder (CE) network [11] uses adversarial training [12] with a novel autoencoder. Most of the early deep learning methods use standard convolutional networks over the corrupted image, using convolutional filter responses on the pixels in the masked holes, which often lead to artifacts such as color discrepancy and blurriness. Partial convolution [13] was proposed, where the convolution is masked and renormalized to be conditioned on only valid pixels. In this model, an updated mask was automatically generated for the next layer as part of the forward pass. Later, partial convolution has been generalized to a gated convolution [14] by providing a learnable dynamic feature selection mechanism for each channel and each spatial location for free-form image inpainting. In early studies using deep learning networks, the missing parts were predicted by propagating the surrounding convolutional features into the missing region to produce semantically plausible images, but they often resulted in blurry images. Spatial attention has been applied to consider the contextual relationship between the background and the hole region. The Shift-Net model [15] introduced a special shift layer to the U-Net architecture to shift the encoder feature of the known region for an estimation of the missing parts, resulting in sharper images with detailed textures. A learnable bidirectional attention map module (LBAM) [16] learned feature re-normalization on both the encoder and decoder of the U-net [17] architecture. A recurrent feature network (RFN) [18] was proposed to exploit the correlation between adjacent pixels and strengthen the constraints for estimating deeper pixels. However, these studies have not fully utilized the structural knowledge of the image. There are several approaches to exploit the inherent structure of information in the input images by using the edge or object boundaries for inpainting [4–6,8]. EdgeConnect [4] is a two-stage adversarial model that comprises an edge generator followed by an image completion network. The edge generator hallucinates the edges in the missing area (either square-shaped or arbitrary-shaped). The image completion network fills in the distorted regions using hallucinated edges as a priori for the inpainting. Edge structures and color-aware maps are fused in a two-stage generative adversarial network (GAN) [5]. In the first stage, edges with the missing regions are used to train an edge structure generator. Meanwhile, distorted input images with the missing part are transformed into a global color feature map by the content-aware fill algorithm. In the second stage, the edge map and the colormap are fused to generate the refined image. The authors in [6] proposed a foreground-aware image inpainting system that explicitly disentangles structure inference and content completion. The foreground contours are predicted first and then the inpainting is performed using the predicted contours as a guidance. A multi-task learning framework with auxiliary tasks of edge finding and gradient map prediction is used to incorporate the knowledge of the image structure to assist inpainting [8].

## 2.2. Inpainting in Medical Field

### 2.2.1. Traditional Approach

The use of inpainting techniques is emerging in medical image analysis since local deformations in medical modalities are common because of various factors such as metallic implants or specular reflections during the image captures. The completion of such missing or distorted regions is important to enhance post-processing tasks such as segmentation or classification. Traditional approaches for medical image inpainting focus on interpolation, non-local means, diffusion techniques, and texture synthesis [19–23]. However, the conventional methods are confined to a single image and they do not learn from images with similar features.

### 2.2.2. Learning-Based Approach

These days, medical image inpainting has been studied extensively with deep learning models [1–3,24,25]. GAN is used to incorporate two patch-based discriminator networks with style and perceptual losses for the inpainting of missing information in positron emission tomography–magnetic resonance imaging (PET-MRI) [1]. A generative framework is proposed to handle the inpainting of arbitrary-shaped regions without a prior localization of the regions of interest [3]. Several improvements are made to deep learning models and are reported with better performance than conventional methods. However, these methods do not use the inherent structure of information in the medical images, resulting in blurry images and often lacking detail. The authors in [7] proposed a method using structural information which is represented by the edges of the image. The network decouples image repair into two separate stages: edge connection and contrast completion. The first stage is to predict the edges inside the missing region. The result edge map is used for inpainting. Even though the use of edges succeeded in improving the performance, it does not provide deeper knowledge of organ structures in the body, resulting in still poor quality of restoration. Recently, a deep neural network for medical inpainting has been proposed in [26]. This framework generates 3D images from sparsely sampled 2D images. They employed an inpainting deep neural network based on a U-net-like structure and DenseNet sub-blocks. However, because of ignoring boundary information in training, their method meets the problem of boundary artifacts. Additionally, since [26] was trained and tested on a dataset that is not publicly available, it is hard to compare performance with this study.

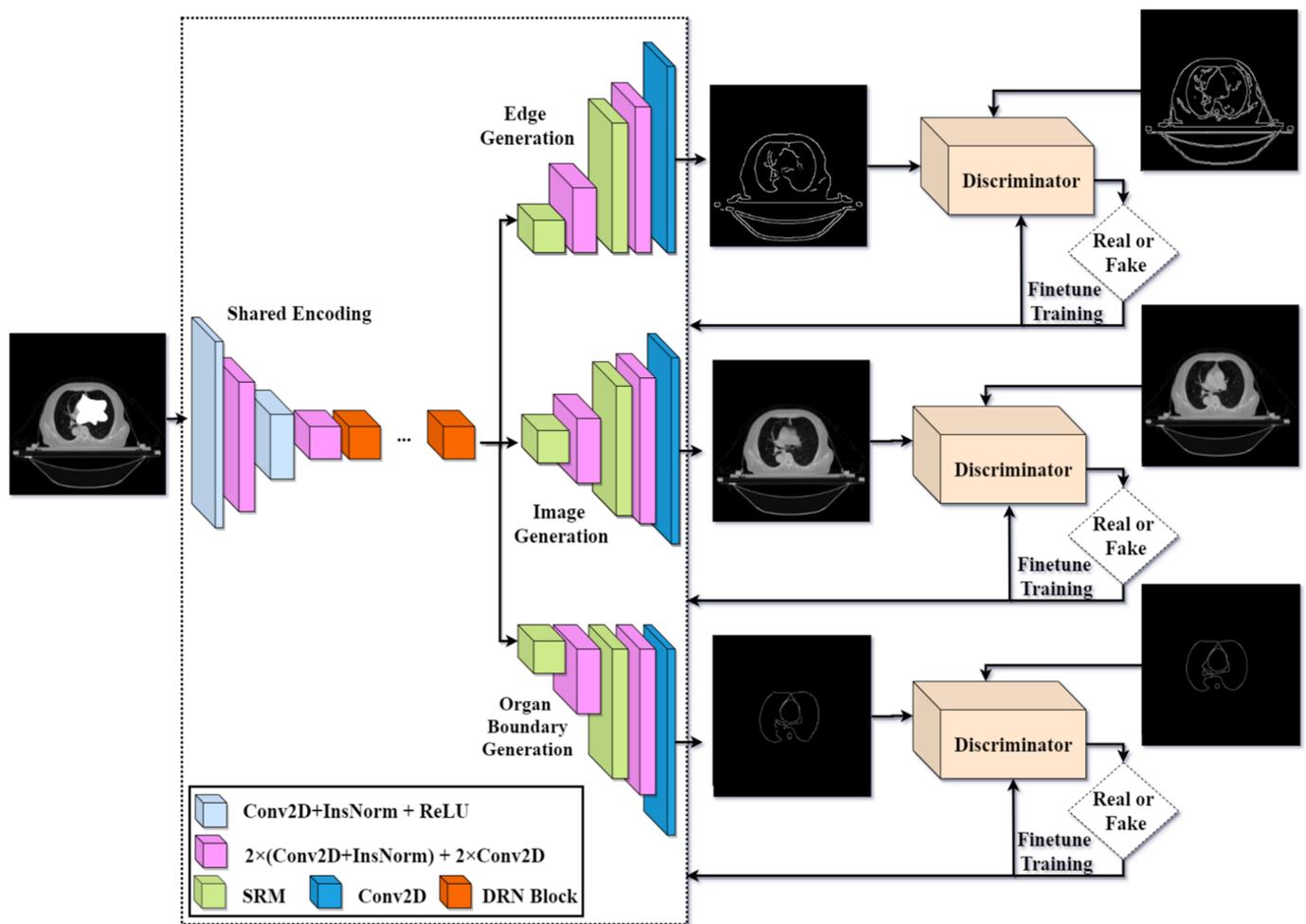
In this paper, we propose a multi-task learning model based on auxiliary tasks of edge reconstruction, and organ boundary prediction with the main task of CT image inpainting. The proposed method is more consistent and effective for image inpainting through simultaneous training of the prediction of edges and organ boundaries.

## 3. Proposed Method

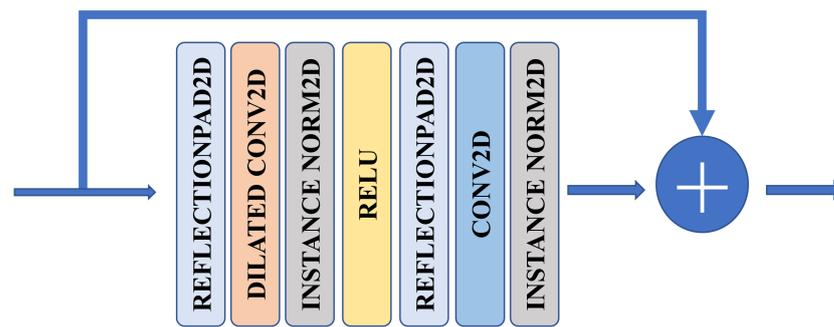
### 3.1. Network Architecture

There have been several methods for boundary detection, such as sketch generation using GAN [27,28]. A contour generation algorithm is used to output contour drawings of arbitrary input images [27]. An application for face photo-sketch synthesis based on the composition-aided GAN is introduced by [28]. The proposed network consists of three GANs for edge reconstruction, organ boundary prediction and image inpainting. Our multi-task learning model is built on an adversarial framework, where the three discriminators feedback the discrimination results to the generator as well as the discriminator. Figure 1 shows the detailed architecture of our model. After encoding the input image, three decoder networks predict the edge map, the organ boundary, and the completed image simultaneously. These results are fed into the discriminator networks, whose feedbacks are directed to the generators. The generator network is a modified autoencoder with one shared encoding and three decoding parts. The Dilated Residual Network (DRN) block of an upgraded ResNet block [4] is constructed by replacing the first convolutional layer with the dilated convolutional layer. Dilated convolutions are used with a dilation

factor of two instead of original convolutions in the residual layers to effectively expand the receptive field without losing resolution in subsequent layers [29–31]. Figure 2 shows the detailed architecture of the DRN block. In the training process, the decoding part is usually difficult to generate feature maps with enough detailed information. Therefore, we employ a super-resolution module (SRM) inside the decoding parts for helping the network learning feature efficiently and produce feature maps with more details. SRM is a modification of the fast super-resolution convolutional neural network (FSRCNN) [32], which makes our model faster with better-reconstructed image quality. Our model is based on the pix2pix GAN [33]. The proposed network takes images from one domain as input and outputs the corresponding image in the other domain, rather than a fixed-size vector. Unlike the initially proposed architecture which classifies a whole image as real or fake, the pix2pix GAN-based model tries to classify patches of an image as real or fake. Therefore, the output is a matrix of values instead of a single value.



**Figure 1.** The overall architecture of our multi-task framework. The network is built on an adversarial framework. It leverages the edge and organ boundary knowledge with multi-task learning (simultaneous image, edge and organ boundary generation).



**Figure 2.** The architecture of Dilated Residual Network Block (DRN Block).

### 3.2. Discriminator

The discriminator is the network to distinguish whether data are from a dataset or generated from generators. Thanks to the discriminator, the model learns the association between input and output. Therefore, the generated images are better and more plausible in detail. We use three discriminators separately during training for better learning features. Each discriminator consists of several convolution layers with a sigmoid activation function.

### 3.3. Loss Function

The loss function returns a non-negative real number representing the difference between two quantities: the predicted label and the correct label. The loss function is like a form to force the model to pay the penalty every time it predicts its mistake, and the number of penalties is proportional to the severity of the error. In all supervised learning problems, our goal always includes minimizing the total penalty payable. Ideally, the loss function should return the minimum value of zero. During the training process, we used many different types of loss for various purposes.

In our network, the input uses the distorted image  $\widetilde{I}_{gt} = I_{gt} \odot (1 - M)$ , where  $I_{gt}$  is the ground truth image and  $M$  is the mask image with 1 value for missing region and 0 for background. The symbol  $\odot$  denotes the Hadamard product. Similarly, we have  $I_{edge\_in} = I_{edge\_gt} \odot (1 - M)$ , where  $I_{edge\_gt}$  is the edge map extracted from ground truth images by the Canny edge detector. Our network generates three images: completed image  $I_{image\_pred}$ , organ boundary map  $I_{organs\_pred}$  and edge map  $I_{edge\_pred}$  with missing regions filled in. Those images have the same resolution as the input image. Let  $G$ ,  $D_1$ ,  $D_2$ , and  $D_3$  be the generator and the discriminator of the image generator, edge generator and organ boundary generator, respectively.

$$I_{image\_pred}, I_{edge\_pred}, I_{organs\_pred} = G(\widetilde{I}_{gt}, I_{edge\_in}) \quad (1)$$

First, we analyze the network with a decoding part which generates completed image  $I_{image\_pred}$ . We employed two losses proposed in [34,35], commonly known as perceptual loss  $Loss_{image\_perceptual}$  and style loss  $Loss_{image\_style}$ . Perceptual loss is defined as follows:

$$Loss_{image\_perceptual} = E \left[ \sum_i \frac{1}{N_i} \|\delta_i(I_{image\_gt}) - \delta_i(I_{image\_pred})\|_1 \right] \quad (2)$$

where  $\delta_i$  is the activation map in the  $i$ th layer of a pre-trained network. These activation maps are also employed to calculate style loss, which measures the differences between covariances of activation maps. Given feature maps of size  $N_i = C_j \times H_j \times W_j$ , style loss is calculated by:

$$Loss_{image\_style} = E \left[ \sum_i \|G_j^\delta(I_{image\_pred}) - G_j^\delta(I_{image\_gt})\|_1 \right] \quad (3)$$

where  $G_j^\delta$  is a  $C_j \times C_j$  Gram matrix generated from activation maps  $\delta_i$ . We also used the reconstruction loss  $Loss_{image\_reconstruction}$  in our model. We chose  $l_1$  loss for this reconstruction loss. We also used a discriminator for our image completion part. Typically, generators' gradients often disappear quickly in generative adversarial networks [12]. To fix this problem, we employed Hinge loss [36], which is useful for classifiers. These loss functions are defined as:

$$Loss_{image\_gen} = -E_{edge\_in} \left[ D_1 \left( I_{image\_pred}, I_{edge\_in} \right) \right] \quad (4)$$

$$Loss_{D_1} = E_{image\_gt, edge\_in} \left[ \max(0, 1 - D_1 \left( I_{image\_gt}, I_{edge\_in} \right)) \right] \\ + E_{edge\_in} \left[ \max(0, 1 + D_1 \left( I_{image\_pred}, I_{edge\_in} \right)) \right] \quad (5)$$

In the components that generate the completed map and completed organ boundary, the structures are similar. The completed edge map and completed organ boundary map are denoted as  $I_{edge\_pred}$ ,  $I_{organs\_pred}$ , respectively. We still use the perceptual losses  $Loss_{edge\_perceptual}$ ,  $Loss_{organs\_perceptual}$ , style losses  $Loss_{edge\_style}$ ,  $Loss_{organs\_style}$  and reconstruction losses  $Loss_{edge\_reconstruction}$ ,  $Loss_{organs\_reconstruction}$  in our model for training the whole model. Finally, our overall loss function is calculated by:

$$Loss_{total} = \varepsilon_1 Loss_{image\_perceptual} + \varepsilon_2 Loss_{image\_style} \\ + \varepsilon_3 Loss_{image\_reconstruction} \\ + \varepsilon_4 Loss_{image\_gen} + \varepsilon_5 Loss_{edge\_perceptual} + \varepsilon_6 Loss_{edge\_style} \\ + \varepsilon_7 Loss_{edge\_reconstruction} + \varepsilon_8 Loss_{edge\_gen} \\ + \varepsilon_9 Loss_{organs\_perceptual} + \varepsilon_{10} Loss_{organs\_style} \\ + \varepsilon_{11} Loss_{organs\_reconstruction} + \varepsilon_{12} Loss_{organs\_gen} \quad (6)$$

where  $\varepsilon$  is the weight of each loss component. From the experiment, we choose  $\varepsilon_1 = \varepsilon_5 = \varepsilon_9 = 0.1$ ,  $\varepsilon_2 = \varepsilon_6 = \varepsilon_{10} = 250$ ,  $\varepsilon_3 = \varepsilon_7 = \varepsilon_{11} = 1$ ,  $\varepsilon_4 = \varepsilon_8 = \varepsilon_{12} = 0.1$  for the training of our model.

## 4. Experiments and Results

### 4.1. Experimental Environment and Datasets

We have used masks of arbitrary-shaped regions and square-shaped regions in this study. One hundred random mask images are created for each mask type for training and 50 mask images are generated for the testing. To make the comparison fair, we use these same masks for training and testing. We conducted a review of our methodology on a publicly available medical dataset of StructSeg2019 [37]. In the StructSeg2019 dataset, there are 50 3D images of CT scans from 50 patients. Fifty of the voxel representations in 3D images are converted into 4775 2D images, among which 1000 2D images are used for the testing, and the rest are used for the training. The input sizes for training and testing are uniformly set as 256 in width and 256 in height. The Canny edge detector [38] is used to generate edge map ground-truth from the input image and organ boundaries ground-truth from the organ segmentations given in the dataset. We employed the Adam algorithm with a batch size of 4 to optimize the network. The proposed method was trained with 30 epochs and the initial learning rate was set at 0.0002. During the training of the model, we use two types of augmentation which are rotation and horizontal reflection. With rotation, we rotate the image in angles 90, 180, and 270 degrees. Our method was implemented in Python language and Pytorch framework. Table 1 shows the details of our experimental environment and the configuration of the training model.

**Table 1.** The details of our experimental environment and the configuration of training model.

Batch Size	Num. of Epoch	Learning Rate	Programming Language	Framework
4	30	0.0002	Python	Pytorch

#### 4.2. Evaluation Criterion

Following research [3] in the medical inpainting field, we use common evaluation metrics such as structural similarity index for measuring image quality (SSIM) [39], peak signal-to-noise ratio (PSNR), mean squared error (MSE), and universal image quality index (UQI) [40] to quantify the performance of the models. Our research conducted experiments on both settings of square-shaped holes and arbitrary-shaped holes. The metric PSNR and MSE are defined as:

$$PSNR = 10 \log_{10} \left( \frac{k_{max}^2}{MSE} \right) \quad (7)$$

$$MSE = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n (k_{ij} - (k_0)_{ij})^2 \quad (8)$$

where MSE represents mean squared error, and the maximum value is denoted by  $k_{max}$ , particularly for 8-bit images  $k_{max} = 255$ . The better the image quality, the higher the PSNR. Structural similarity (SSIM) is seen to be a stronger parameter for evaluating picture consistency which is within the range [0,1], with a score close to 1 indicating better conservation of the structure. This metric is based on the visual perception characteristic of humans. The SSIM is calculated between two commonly sized windows  $A \times B$ .

$$SSIM = \frac{(2\mu_{\omega_1}\mu_{\omega_2} + c_1)(2\sigma_{\omega_1\omega_2} + c_2)}{(\mu_{\omega_1}^2 + \mu_{\omega_2}^2 + c_1)(\sigma_{\omega_1}^2 + \sigma_{\omega_2}^2 + c_2)} \quad (9)$$

where  $\mu_{\omega_i}, \sigma_{\omega_i}^2$  are the average and the variance of window  $\omega_i$ , respectively. The covariance is denoted by  $\sigma_{\omega_1\omega_2}$  and  $c_1, c_2$  are numerical stabilizing parameters. We also used the metric UQI, which is the predecessor of SSIM, to evaluate our proposed methods with other methods. Let  $I_{gt} = \{I_{gt_i} | i = 1, 2, \dots, Z\}$  and  $I_{pred} = \{I_{pred_i} | i = 1, 2, \dots, Z\}$  be the ground truth and the predicted image, respectively. The metric UQI is defined as:

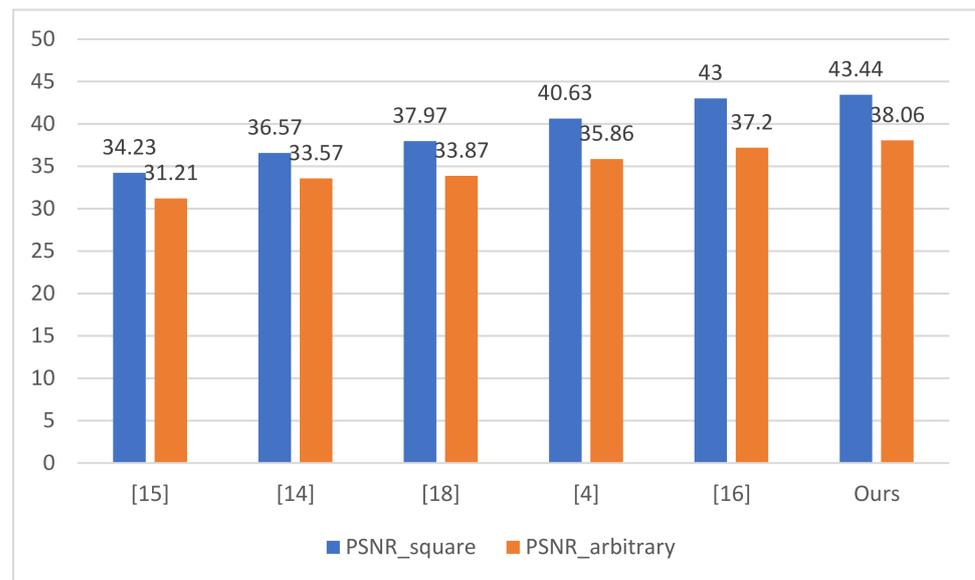
$$UQI = \frac{\sigma_{I_{gt}I_{pred}}}{\sigma_{I_{gt}}\sigma_{I_{pred}}} \cdot \frac{2 \overline{I_{gt}} \overline{I_{pred}}}{(\overline{I_{gt}})^2 + (\overline{I_{pred}})^2} \cdot \frac{2 \sigma_{I_{gt}}\sigma_{I_{pred}}}{\sigma_{I_{gt}}^2 + \sigma_{I_{pred}}^2}, \quad (10)$$

where the dynamic range of UQI is [-1,1]. One can achieve the highest value 1 if and only if  $I_{pred_i} = I_{gt_i}$  for all  $i = 1, 2, \dots, Z$ . The lowest value of -1 occurs when  $I_{pred_i} = 2 \overline{I_{gt}} - I_{gt_i}$  for all  $i = 1, 2, \dots, Z$ . The element  $\frac{\sigma_{I_{gt}I_{pred}}}{\sigma_{I_{gt}}\sigma_{I_{pred}}}$  is defined as the correlation coefficient between ground truth image  $I_{gt}$  and predicted image  $I_{pred}$ , and the value of this element is in the range [-1,1]. The element  $\frac{2 \overline{I_{gt}} \overline{I_{pred}}}{(\overline{I_{gt}})^2 + (\overline{I_{pred}})^2}$ , with a value range of [0,1], computes how close the mean luminance is between ground truth image  $I_{gt}$  and predicted image  $I_{pred}$ . It equals 1 if and only if  $\overline{I_{gt}} = \overline{I_{pred}}$ . The last element  $\frac{2 \sigma_{I_{gt}}\sigma_{I_{pred}}}{\sigma_{I_{gt}}^2 + \sigma_{I_{pred}}^2}$  presents how similar the contrasts of the images are. The value of this element is in the range [-1,1], where the best value 1 is obtained if and only if  $\sigma_{I_{gt}} = \sigma_{I_{pred}}$ .

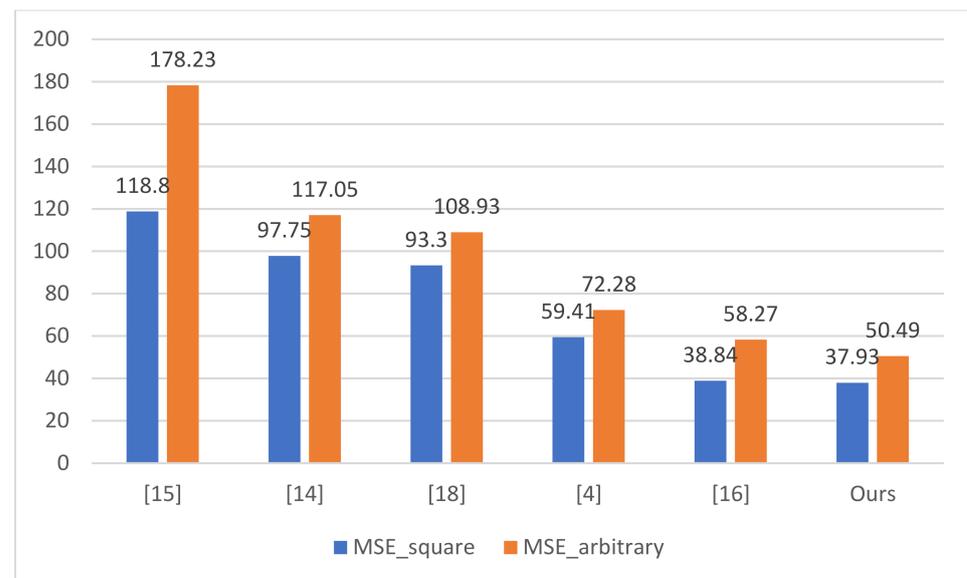
#### 4.3. Results

Our results in PSNR and MSE metrics are presented by graphs which are shown in Figures 3 and 4. Figure 3 shows the PSNR score of our method for a square-shaped and arbitrary-shaped masked image compared with the others. The higher value is the better. Figure 4 presents the MSE score of our method for a square-shaped and arbitrary-shaped masked image compared with the others. The lower value is the better. For square-shaped inpainting, the results are presented in Figure 5 and Table 2, respectively. Partial convolution [13] resulted in the worst inpainting results from both a quantitative and qualitative perspective. The proposed method achieved the highest performance compared to the others. In Figure 6 and Table 3, the qualitative and quantitative results for

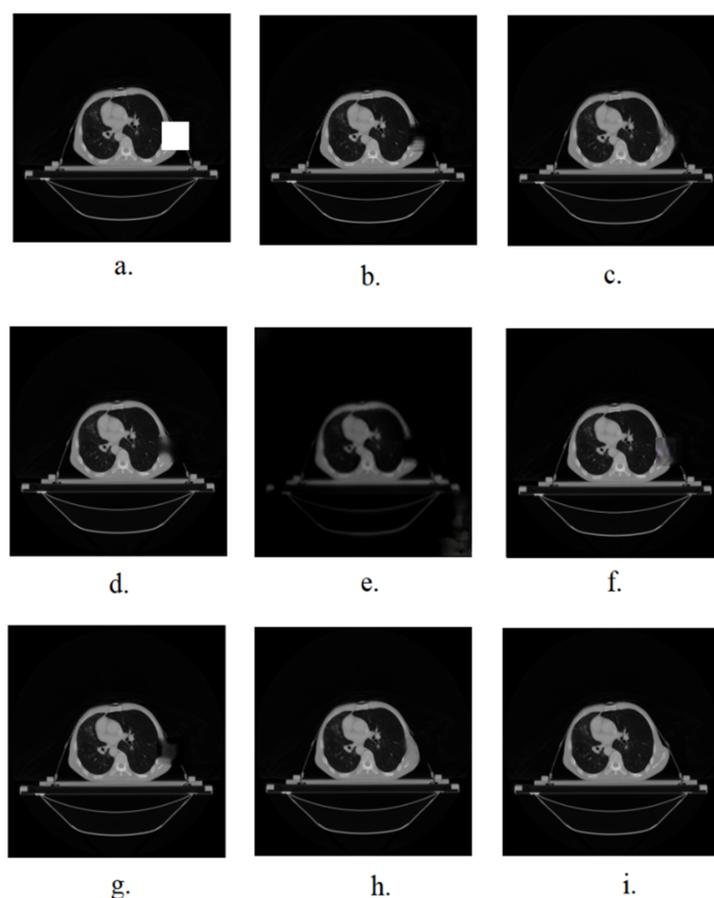
arbitrary-shaped inpainting are given, respectively. Our approach still outperforms other methods. The PSNR results achieved 43.44 and 38.06 dB in the property of square-shaped and arbitrary-shaped masks, respectively. These results demonstrate the effectiveness of the proposed method for both square and arbitrary masks. Table 4 shows the comparison of results from different types of loss function using a discriminator. Table 5 shows the quantitative results of the proposed method between using SRM and without using it. Table 6 presents the quantitative comparison of PSNR/SSIM/MSE/UQI between multi-task and mono-task framework in property of arbitrary-shaped regions. Table 7 introduces the quantitative comparison of PSNR/SSIM/MSE/UQI between multi-task and mono-task framework in property of square-shaped regions.



**Figure 3.** This graph shows PSNR score of our method for a square-shaped and arbitrary-shaped masked image compared with the others. The higher value is the better.



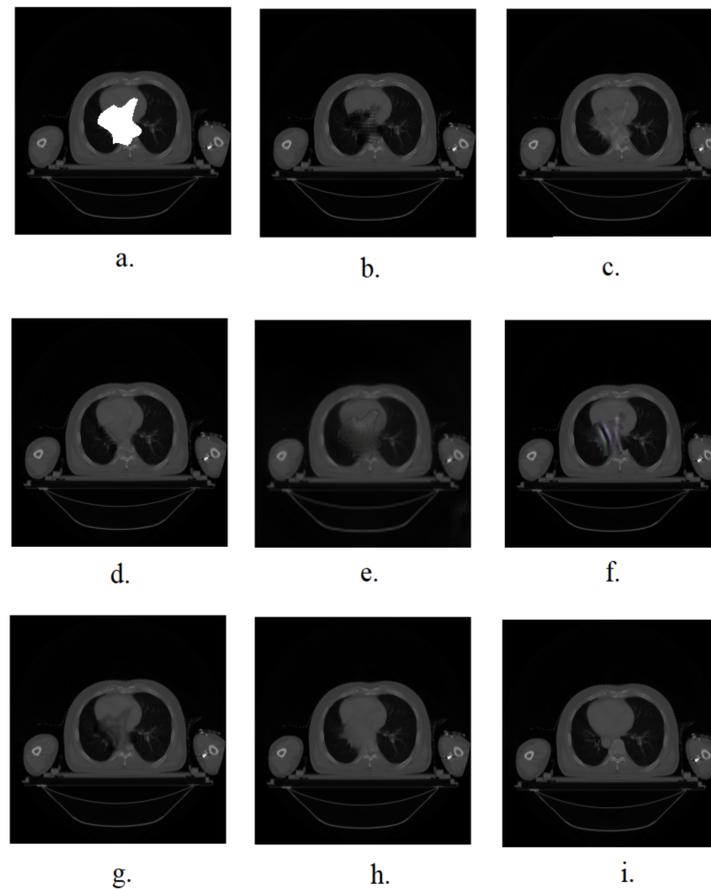
**Figure 4.** This graph shows MSE score of our method for a square-shaped and arbitrary-shaped masked image compared with the others. The lower value is the better.



**Figure 5.** The inpainting results of our method for a square-shaped masked image compared with the others. (a) The masked image. (b) Result obtained from using public source code of method [15]. (c) Result obtained from using public source code of method [16]. (d) Result obtained from using public source code of method [4]. (e) Result obtained from using public source code of method [13]. (f) Result obtained from using public source code of method [14]. (g) Result obtained from using public source code of method [18]. (h) Result of our method. (i) Ground truth image.

**Table 2.** The quantitative comparison of PSNR/SSIM/MSE/UQI between proposed method and other methods in property of squared-shaped regions. The results from other methods are derived from using their public code.

	[4]	[13]	[14]	[15]	[16]	[18]	Ours
PSNR	40.63	22.34	36.57	34.23	43.00	37.97	43.44
SSIM	0.9785	0.2335	0.9002	0.7032	0.9811	0.9764	0.9818
MSE	59.41	1223.08	97.75	118.80	38.84	93.30	37.93
UQI	0.9938	0.2666	0.8767	0.9393	0.9951	0.9899	0.9960



**Figure 6.** The inpainting results of our method for an arbitrary-shaped masked image compared with the others. (a) The masked image. (b) Result obtained from using public source code of method [15]. (c) Result obtained from using public source code of method [16]. (d) Result obtained from using public source code of method [4]. (e) Result obtained from using public source code of method [13]. (f) Result obtained from using public source code of method [14]. (g) Result obtained from using public source code of method [18]. (h) Result of our method. (i) Ground truth image.

**Table 3.** The quantitative comparison of PSNR/SSIM/MSE/UQI between proposed method and other methods in property of arbitrary-shaped regions. The results from other methods are derived from using their public code.

	[4]	[13]	[14]	[15]	[16]	[18]	Ours
PSNR	37.20	29.95	33.87	31.21	35.86	33.57	38.06
SSIM	0.9731	0.3916	0.8941	0.7222	0.9729	0.9716	0.9746
MSE	58.27	241.09	108.93	178.23	72.28	117.05	50.49
UQI	0.9964	0.7228	0.8790	0.9601	0.9966	0.9939	0.9972

**Table 4.** The effect of loss function using a discriminator.

	Arbitrary-Shaped Regions			Square-Shaped Regions		
	BCE	MSE	Hinge (Ours)	BCE	MSE	Hinge (Ours)
PSNR	37.00	37.58	38.06	41.78	42.51	43.44
SSIM	0.9723	0.9742	0.9746	0.9802	0.9812	0.9818
MSE	57.75	53.64	50.49	50.36	46.44	37.93
UQI	0.9965	0.9970	0.9972	0.9948	0.9953	0.9960

**Table 5.** The effect of SRM in our network.

	Square-Shaped Regions		Arbitrary-Shaped Regions	
	w/o SRM	Ours	w/o SRM	Ours
PSNR	41.98	43.44	36.99	38.06
SSIM	0.9804	0.9818	0.9726	0.9746
MSE	44.16	37.93	59.23	50.49
UQI	0.9954	0.9960	0.9967	0.9972

**Table 6.** The quantitative comparison of PSNR/SSIM/MSE/UQI between multi-task and mono-task framework in property of arbitrary-shaped regions.

	Mono-Task	Multi-Task with Edge Information	Multi-Task with Organs' Boundary Information	Multi-Task with Edge and Boundary Information (Ours)
PSNR	36.62	37.20	36.94	38.06
SSIM	0.9724	0.9745	0.9734	0.9746
MSE	69.78	69.25	63.73	50.49
UQI	0.9963	0.9962	0.9967	0.9972

**Table 7.** The quantitative comparison of PSNR/SSIM/MSE/UQI between multi-task and mono-task framework in property of square-shaped regions.

	Mono-Task	Multi-Task with Edge Information	Multi-Task with Organs' Boundary Information	Multi-Task with Edge and Boundary Information (Ours)
PSNR	40.85	42.32	42.77	43.44
SSIM	0.9797	0.9812	0.9817	0.9818
MSE	61.36	61.64	36.35	37.93
UQI	0.9946	0.9947	0.9963	0.9960

Tables 2 and 3 show the experimental results compared to other methods based on both arbitrary-shaped and square-shaped masks. Compared to recent inpainting studies, our method produced promising results. Particularly, we compared it to methods [13,15] proposed in 2018, methods [4,14,16], introduced in 2019, and method [18] presented in 2020.

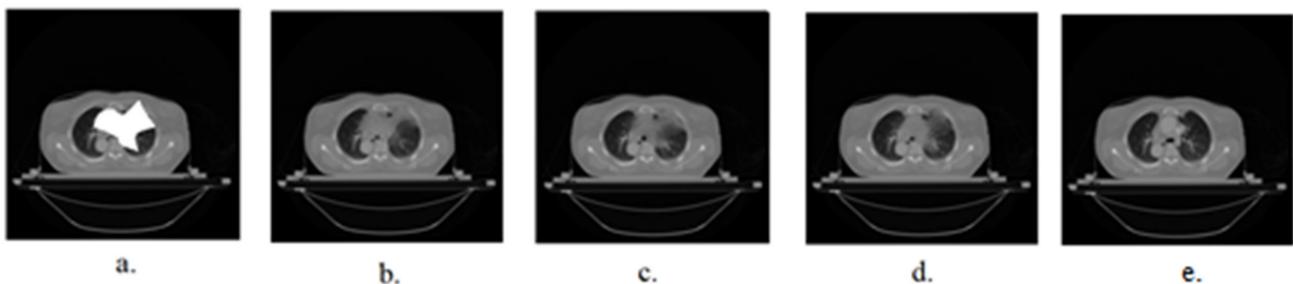
From Tables 2 and 7, for square mask shapes, with using the only mono-task framework, our results are only 40.85 dB PSNR and lower than the method results from LBAM [16], with a PSNR value reaching 43 dB. Although we integrated the edge knowledge learning task or organ boundary awareness task, the results show very little performance increase with the PSNR metric increased to 42.32 and 42.77 dB, respectively. However, it does prove the positive value of adding the auxiliary task into the network. We continued to survey the inpainting results using a multi-task learning framework based on edge awareness and organ boundary knowledge in the medical image. We achieved promising results when the PSNR metric value increased to 43.44 dB, which surpassed all remaining methods.

Tables 3 and 6 show that our results still obtained the best value for the mask shape in an arbitrary form. The proposed method helped us achieve a PSNR value exceeding 38.06 dB compared with the highest value of the remaining methods of 37.20 dB belonging to the EdgeConnect method [4]. Our approach has outstanding practicality and outstanding performance from the above quantitative comparisons compared to inpainting methods in recent years.

We also present qualitative comparisons in Figures 5 and 6. We apply a mask shape of a square in Figure 5 and then we reproduce images with very high authenticity. The structure of the right lung is preserved relatively intact while the remaining methods reproduce low plausible images. The other methods generate images that lost so much detailed information. The left lung in Figure 6 has degraded a lot when we use an arbitrary-

shaped mask. However, thanks to the multi-task framework based on edge awareness and organ boundary knowledge, our method can reproduce the image with a plausible result. Especially, the structures of the right lung and left lung are pretty well reconstructed. The left lung boundary area was reproduced quite sharply without any blurring or distortion, while the rest of the methods were not capable.

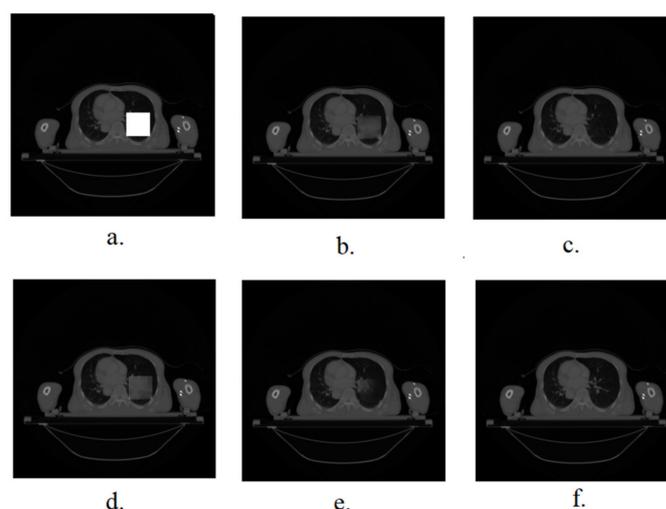
Table 4 shows the comparison of results from different types of loss function using a discriminator. By using binary cross-entropy loss, our model can only generate results with 37.00 and 41.78 dB in PSNR with the property of arbitrary-shaped mask and square-shaped mask, respectively. These results are lower than the results of methods [4,16]. When we change the binary cross-entropy loss with mean square error loss, our inpainting task results are a little bit better, but these results in the property of square-shaped mask are still lower than the method [16]. Finally, we replace the mean square error loss with Hinge loss. Our inpainting results outperform others, with 43.44 and 38.06 dB in the property of square-shaped and arbitrary-shaped masks, respectively. Figure 7 shows the qualitative comparison of inpainting results between using binary-cross entropy loss, L2 (mean square error) loss and Hinge loss in discriminators. For the best-generated results, we chose Hinge loss as our discriminators during training the model. Figure 8 introduces the qualitative comparison inpainting results between using the mono-task framework, multi-task with organ boundary knowledge framework, multi-task with edge knowledge framework, and multi-task with boundary combined with edge knowledge framework.



**Figure 7.** Qualitative comparison inpainting results between using binary cross entropy loss, L2 loss and Hinge loss in discriminators. (a) Input masked image. (b) Result of using binary cross entropy loss. (c) Result of using L2 loss. (d) Result of using Hinge loss. (e) Ground truth image.

We also examine the effect of SRM on our model. Table 5 shows the quantitative results of the proposed method between using SRM and without using it. The results show that generated images by the model with SRM are pretty much better in terms of either square-shaped or arbitrary-shaped masks. This proved the positive effect of SRM in making the model generate better high-resolution features with more useful details. From the above comparison and analysis, we find that the proposed model has superior performance in research on inpainting on medical CT images compared with other studies in recent years.

Table 8 shows the detailed structure of the encoding part. The components of the discriminator network are introduced in Table 9. The detailed information of the decoding parts is given in Tables 10–12.



**Figure 8.** Qualitative comparison inpainting results between using mono-task framework, multi-task with organ boundary knowledge framework, multi-task with edge knowledge framework and multi-task with boundary combined with edge knowledge framework. (a) Input masked image. (b) Result of using mono-task framework. (c) Result of using multi-task with organ boundary knowledge framework. (d) Result of using multi-task with edge knowledge framework. (e) Result of using multi-task with boundary combined with edge knowledge framework. (f) Ground truth image.

**Table 8.** Architecture of encoding part in our network.

	Layer	Kernel Size	Stride
	ReflectionPad2d		
	Conv2d+IN+ReLU	[7,7]	[1,1]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d+IN+ReLU	[4,4]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
8 × DRN blocks			

**Table 9.** Architecture of discriminator network.

Layer	Kernel Size	Stride
Conv2d + LeakyReLU	[4,4]	[2,2]
Conv2d + LeakyReLU	[4,4]	[2,2]
Conv2d + LeakyReLU	[4,4]	[2,2]
Conv2d + LeakyReLU	[4,4]	[1,1]
Conv2d	[4,4]	[1,1]
Sigmoid		

**Table 10.** Architecture of decoding part 1 in our network.

	Layer	Kernel Size	Stride
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
	ReflectionPad2d		
	Conv2d	[7,7]	[1,1]

**Table 11.** Architecture of decoding part 2 in our network.

	Layer	Kernel Size	Stride
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
	ReflectionPad2d		
	Conv2d	[7,7]	[1,1]

**Table 12.** Architecture of decoding part 3 in our network.

	Layer	Kernel Size	Stride
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
SRM	Conv2d + PReLU	[6,6]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d	[3,3]	[1,1]
	Conv2d + PReLU	[3,3]	[1,1]
	Conv2d + PReLU	[1,1]	[1,1]
	ConvTranspose2d + IN +ReLU	[9,9]	[2,2]
Deep Block	Conv2d + IN	[8,8]	[1,1]
	Conv2d + IN	[8,8]	[1,1]
	Conv2d	[1,1]	[1,1]
	Conv2d	[1,1]	[1,1]
	ReflectionPad2d		
	Conv2d	[7,7]	[1,1]

#### 4.4. Ablation Study

Binary Cross-Entropy (BCE), Mean Square Error (MSE), and Hinge loss are popular loss functions for classification. Cross-entropy calculated a score that summarizes the average difference between the actual and predicted probability distributions for predicting class 1. The score is minimized, and the perfect cross-entropy value is 0. The range of BCE output is from 0 to 1. MSE computes the sum of squared distances between the ground truth value and the predicted value. The Hinge loss function emphasizes examples to have the correct sign, adding more error when there is a difference in the sign between the ground truth and the predicted value. Figure 5 shows qualitative comparison inpainting results between using binary cross entropy loss, L2 loss and Hinge loss in discriminators. Table 4 compares the results from different types of loss functions used in discriminators with the property of square-shaped and arbitrary-shaped regions. Our results outperformed others when we chose Hinge loss for our discriminators during training the model. We also validate the effectiveness of SRM in our model. Table 5 shows the quantitative results with SRM and without using it, with the property of square-shaped and arbitrary-shaped regions. The results show that generated images by the model with SRM are of better performance, which proved the positive effect of SRM in making the model generate features with useful details. Tables 6 and 7 show the quantitative comparison of PSNR/SSIM/MSE/UQI between the multi-task and mono-task framework in the property of arbitrary-shaped and square-shaped regions. Although we integrated the edge learning task or organ boundary task, the results are slightly increased in the PSNR metric. Our method outperformed the rest of the methods when we used a multi-task learning framework based on both edge and organ boundary learning. It does prove the positive value of adding the auxiliary task into the network. Figure 6 presents qualitative comparison inpainting results between using the mono-task framework, multi-task with organ boundary knowledge framework, multi-task with edge knowledge framework and multi-task with boundary combined with edge knowledge framework.

## 5. Conclusions

This paper presented an efficient multi-task learning network for medical image inpainting based on organ boundary awareness. We utilized the auxiliary tasks of edge and organ boundary prediction to make the model generate the sharp and clear images for inpainting by learning the detailed structure of organs. Our model proved itself efficient in the reconstruction of the degraded or distorted organs and generates plausible boundaries for the inpainting. Based on detailed experimental evaluation, we demonstrated that the proposed method outperforms the state-of-the-art methods on medical image inpainting. It achieved the best results in literature so far, with the highest PSNR and lowest MSE value for both of the arbitrary-shaped and the square-shaped regions. The proposed model generates the sharp and clear images for inpainting by learning the detailed structure of organs. Therefore, our method was able to show how promising the method is when applying it in medical image analysis, where the completion of missing or distorted regions is still a challenging task. The research is a good foundation for future medical imaging analysis and helps the diagnostic and prognostic capabilities of medical experts. We hope to extend the proposed method of handling other systems of medical images such as X-rays, magnetic resonance, or ultrasound images. Additionally, in medical analysis, there are so many datasets that are quite small in size. Therefore, it is essential to optimize the model when applied to small datasets to achieve good results. Some research directions use relatively small datasets but still achieve good performance, such as [41–43]. In the future, we will optimize the proposed model to apply it to small datasets.

**Author Contributions:** Conceptualization, M.-T.T.; Methodology, M.-T.T.; Writing—review and editing, M.-T.T. and G.-S.L.; Supervision, G.-S.L., S.-H.K. and H.-J.Y.; Project administration, G.-S.L., S.-H.K. and H.-J.Y.; Funding acquisition, G.-S.L., S.-H.K. and H.-J.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1A4A1019191) and also by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (MSIT) (NRF-2019M3E5D1A02067961).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Karim, A.; Youssef, M.; Sergios, G.; Bin, Y. Adversarial Inpainting of Medical Image Modalities. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3267–3271.
2. Ecem, S.; Shi, H.; Davide, B.; Bram, G. Chest X-Ray Inpainting with Deep Generative Models. *arXiv* **2018**, arXiv:1809.01471.
3. Karim, A.; Vijeth, K.; Sherif, A.; Tobias, H.; Sergios, G.; Bin, Y. IPA-Medgan: Inpainting of Arbitrary Regions in Medical Imaging. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 3005–3009.
4. Kamyar, N.; Eric, N.; Tony, J.; Faisal, Q.; Mehran, E. Edgeconnect: Structure Guided Image Inpainting Using Edge Prediction. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019.
5. Shao, H.; Wang, Y.; Fu, Y.; Yin, Z. Generative Image Inpainting Via Edge Structure and Color Aware Fusion. *Signal Process. Image Commun.* **2020**, *87*, 115929. [[CrossRef](#)]
6. Xiong, W.; Yu, J.; Lin, Z.; Yang, J.; Lu, X.; Barnes, C.; Luo, J. Foreground-Aware Image Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5840–5848.
7. Chai, Y.; Xu, B.; Zhang, K.; Lepore, N.; Wood, J.C. MRI Restoration Using Edge-Guided Adversarial Learning. *IEEE Access* **2020**, *8*, 83858–83870. [[CrossRef](#)] [[PubMed](#)]
8. Yang, J.; Qi, Z.; Shi, Y. Learning to Incorporate Structure Knowledge for Image Inpainting. In Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 12605–12612.
9. Huang, J.B.; Kang, S.B.; Ahuja, N.; Kopf, J. Image Completion Using Planar Structure Guidance. *ACM Trans. Graph. (TOG)* **2014**, *33*, 1–10. [[CrossRef](#)]

10. Le Meur, O.; Gautier, J.; Guillemot, C. Exemplar-Based Inpainting Based on Local Geometry. In Proceedings of the IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 3401–3404.
11. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
12. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
13. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image Inpainting for Irregular Holes Using Partial Convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
14. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-Form Image Inpainting with Gated Convolution. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4471–4480.
15. Yan, Z.; Li, X.; Li, M.; Zuo, W.; Shan, S. Shift-Net: Image Inpainting via Deep Feature Rearrangement. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 1–17.
16. Xie, C.; Liu, S.; Li, C.; Cheng, M.M.; Zuo, W.; Liu, X.; Wen, S.; Ding, E. Image Inpainting with Learnable Bidirectional Attention Maps. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8858–8867.
17. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
18. Li, J.; Wang, N.; Zhang, L.; Du, B.; Tao, D. Recurrent Feature Reasoning for Image Inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7760–7768.
19. Alsalamah, M.; Amin, S. Medical Image Inpainting with RBF Interpolation Technique. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 91–99. [[CrossRef](#)]
20. Feng, Z.; Chi, S.; Yin, J.; Zhao, D.; Liu, X. A Variational Approach to Medical Image Inpainting Based on Mumford-Shah Model. In Proceedings of the International Conference on Service Systems and Service Management, Chengdu, China, 9–11 June 2007; pp. 1–5.
21. Guizard, N.; Nakamura, K.; Coupé, P.; Fonov, V.S.; Arnold, D.L.; Collins, D.L. Non-Local Means Inpainting of MS Lesions in Longitudinal Image Processing. *Front. Neurosci.* **2015**, *9*, 456. [[CrossRef](#)] [[PubMed](#)]
22. Vlašánek, P. Fuzzy Image Inpainting Aimed to Medical Images. In Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Pilsen/Prague, Czech Republic, 28 May–1 June 2018.
23. Arnold, M.; Ghosh, A.; Ameling, S.; Lacey, G. Automatic Segmentation and Inpainting Of Specular Highlights for Endoscopic Imaging. *J. Image Video Process.* **2010**. [[CrossRef](#)]
24. Tran, M.-T.; Kim, S.H.; Yang, H.-J.; Lee, G.-S. Medical Image Inpainting with Deep Neural Network. In Proceedings of the Smart Media Spring Conference, Gwangju, Korea, 22–23 May 2020.
25. Tran, M.T.; Kim, S.H.; Yang, H.J.; Lee, G.S. Deep Learning-Based Inpainting for Chest X-Ray Image. In Proceedings of the International Conference on Smart Media and Applications (SMA), Jeju, Korea, 17–19 September 2020.
26. Kang, S.K.; Shin, S.A.; Seo, S.; Byun, M.S.; Lee, D.Y.; Kim, Y.K.; Lee, D.S.; Lee, J.S. Deep learning-based 3d inpainting of brain mr images. *Sci. Rep.* **2021**, *11*, 1673. [[CrossRef](#)] [[PubMed](#)]
27. Li, M.; Lin, Z.; Mech, R.; Yumer, E.; Ramanan, D. Photo-Sketching: Inferring Contour Drawings from Images. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1403–1412.
28. Yu, J.; Xu, X.; Gao, F.; Shi, S.; Wang, M.; Tao, D.; Huang, Q. Toward Realistic Face Photo-Sketch Synthesis via Composition-Aided Gans. *IEEE Trans. Cybern.* **2020**. [[CrossRef](#)] [[PubMed](#)]
29. Chen, C.; Liu, X.; Ding, M.; Zheng, J.; Li, J. 3D Dilated Multi-Fiber Network for Real-Time Brain Tumor Segmentation in MRI. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 184–192.
30. Ngo, D.K.; Tran, M.T.; Kim, S.H.; Yang, H.J.; Lee, G.S. Multi-Task Learning for Small Brain Tumor Segmentation from MRI. *Appl. Sci.* **2020**, *10*, 7790. [[CrossRef](#)]
31. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2015**, arXiv:1511.07122.
32. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 391–407.
33. 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, Honolulu, HI, USA, 21–26 July 2017.
34. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
35. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses For real-Time Style Transfer and Super-Resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
36. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 7354–7363.
37. Li, H.; Chen, M. Automatic Structure Segmentation for Radio Therapy Planning Challenge 2020. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020.

38. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, 679–698. [[CrossRef](#)]
39. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
40. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
41. Comelli, A.; Dahiya, N.; Stefano, A.; Benfante, V.; Gentile, G.; Agnese, V.; Raffa, G.M.; Pilato, M.; Yezzi, A.; Petrucci, G.; et al. Deep learning approach for the segmentation of aneurysmal ascending aorta. *Biomed. Eng. Lett.* **2021**, *11*, 15–24. [[CrossRef](#)]
42. Comelli, A.; Dahiya, N.; Stefano, A.; Vernuccio, F.; Portoghese, M.; Cutaia, G.; Bruno, A.; Salvaggio, G.; Yezzi, A. Deep learning-based methods for prostate segmentation in magnetic resonance imaging. *Appl. Sci.* **2021**, *11*, 782. [[CrossRef](#)]
43. Comelli, A.; Coronello, C.; Dahiya, N.; Benfante, V.; Palmucci, S.; Basile, A.; Vancheri, C.; Russo, G.; Yezzi, A.; Stefano, A. Lung segmentation on high-resolution computerized tomography images using deep learning: A preliminary step for radiomics studies. *J. Imaging* **2020**, *6*, 125. [[CrossRef](#)]