Robust Medical Image Colorization with Spatial Mask-Guided Generative Adversarial Network

Zuyu Zhang, Yan Li and Byeong-Seok Shin *

Department of Electrical and Computer Engineering, Inha University, Incheon 22212, Republic of Korea
* Correspondence: bsshin@inha.ac.kr

Abstract: Color medical images provide better visualization and diagnostic information for doctors during clinical procedures than grayscale medical images. Although generative adversarial network-based image colorization approaches have shown promising results, in these methods, adversarial training is applied to the whole image without considering the appearance conflicts between the foreground objects and the background contents, resulting in generating various artifacts. To remedy this issue, we propose a fully automatic spatial mask-guided colorization with generative adversarial network (SMCGAN) framework for medical image colorization. It generates colorized images with fewer artifacts by introducing spatial masks, which encourage the network to focus on the colorization of the foreground regions instead of the whole image. Specifically, we propose a novel spatial mask-guided method by introducing an auxiliary foreground segmentation branch combined with the main colorization branch to obtain the spatial masks. The spatial masks are then used to generate masked colorized images where most background contents are filtered out. Moreover, two discriminators are utilized for the generated colorized images and masked generated colorized images, respectively, to assist the model in focusing on the colorization of foreground regions. We validate our proposed framework on two publicly available datasets, including the Visible Human Project (VHP) dataset and the prostate dataset from NCI-ISBI 2013 challenge. The experimental results demonstrate that SMCGAN outperforms the state-of-the-art GAN-based image colorization approaches with an average improvement of 8.48% in the PSNR metric. The proposed SMCGAN can also generate colorized medical images with fewer artifacts.

Keywords: image colorization; generative adversarial network; medical images

1. Introduction

Imaging technology in biomedical engineering has made the interior portions of the body observable by doctors for disease diagnoses without having to invade the bodies of patients [1,2]. Medical imaging has also been used to guide and assist surgical procedures [3]. For instance, in keyhole surgeries, can help doctors reach the interior parts without really opening too much of the body. Medical imaging utilizes fundamental physical phenomena, including acoustic wave dissemination and X-ray propagation, to obtain the health parameters of patients. With the emergence of advanced medical imaging devices, a large number of medical images have been generated and collected at an unprecedented speed and scale using imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound imaging (UI) [4]. Thus, it is critical to develop medical image processing algorithms to help doctors diagnose and analyze diseases rapidly. Among them, medical image colorization is an important topic attracting more and more attention [5]. Synthesized color images can enrich the details of organs and tissues much more compared with gray images [6,7]. They can help doctors identify problems more accurately and avoid misjudgment [8].

Image colorization is the process of assigning colors to each pixel based on the intensity in a grayscale image. In the last decade, many methods have been proposed to solve this
problem. These algorithms can be roughly divided into three categories: scribble-based methods [9,10], exemplar-based methods [11,12], and fully automatic methods [13–16]. The scribble-based methods utilize the color hints provided by users to assign different colors to the objects in an image. In contrast, to reduce human interactions during the process of colorization, exemplar-based methods infer the RGB color of each region in the input image by selecting a reference image. However, the performance of example-based methods depends highly on the quality of the selected reference image.

Different from scribble-based methods and exemplar-based methods, fully automatic approaches based on deep learning perform colorization in an end-to-end manner without any human intervention [15,16]. Recently, GAN-based colorization methods have been explored to perform fully automatic colorization. For example, Nazeri et al. proposed the deep convolutional generative adversarial network (DCGAN) [15] to perform end-to-end colorization by directly learning the mapping between the input grayscale images and the corresponding colorized ones. Vitoria et al. [16] proposed ChromaGAN, an adversarial learning colorization approach with semantic information incorporated into it to colorize images more realistically. Although the aforementioned GAN-based methods have achieved promising results, these GAN-based methods mainly focus on adversarial training operated on the whole image while neglecting the various generated artifacts [17,18]. Some artifacts are produced because of appearance conflicts between the foreground objects and the background contents [19]. In medical images, background regions may not contain any tissues or organs and can be regarded as noise during the colorization process. Several works have been proposed to employ foreground-aware adversarial training to suppress the generated artifacts [19,20]. However, the effectiveness of foreground-aware adversarial training has not been investigated in image colorization for medical images. In addition, unlike most previous works that were limited to natural image colorization, this work focuses on the colorization techniques of medical images.

To this end, we propose a novel end-to-end SMCGAN framework for medical image colorization by introducing a spatial mask-guided generative adversarial network, in which the model is forced to focus on the colorization of foreground regions and reduce visual artifacts. Specifically, we employ a generative adversarial network as the main image colorization network to learn the mapping between the grayscale values and chromatic values. A spatial mask derived from the auxiliary segmentation network is used to obtain a weighted synthesized color image where most contents of background regions are filtered out. The weighted synthesized color image assists the image colorization network to focus the colorization of the foreground regions by adversarial training.

The main contributions of this paper are as follows:

- We demonstrate that the foreground-aware module composed of a spatial mask embedded in a GAN-based colorization framework makes the model emphasize the foreground regions during the colorization process.
- A novel adversarial loss function is devised to assist the colorization model to focus on the colorization of foreground regions, reducing visual artifacts to improve the performance of colorization.

The remainder of this paper is organized as follows: we introduce the existing methods of image colorization in Section 2. In Section 3, we detail the architecture and the loss functions of our proposed SMCGAN framework. Performance comparisons and ablation studies of the proposed SMCGAN framework are performed in Section 4. We also provide a brief analysis of the limitations of our framework in Section 5. Finally, in Section 6, we present our conclusions.

2. Related Work

The existing methods of image colorization can be divided into three categories: scribble-based colorization, exemplar-based colorization, and fully automatic-based colorization.

Scribble-based methods attempt to annotate the grayscale image in a straightforward way with color scribbles. These color scribbles serve as landmarks for colorization, and
color from the scribble is propagated to the rest of the image. Levin et al. [9] regarded scribble-based colorization as an optimization problem with linear constraints based on the assumption that the adjacent pixels with the same intensity should have a similar color. To capture the long-range color pixels, both the use of locally linear embedding [21] and the utilization of an affinity-based edit scheme [22] are proposed by modeling the linear combination of adjacent pixels in a featured space. To maintain structural information, Sangkloy et al. [23] used a novel deep generative adversarial architecture with sketches and color strokes as user input. To simultaneously utilize global and local information, Xiao et al. [24] developed an interactive colorization model based on the U-Net [25] architecture, which is composed of a feature extraction module, a dilated module, a global input module, and a reconstruction module. However, the main weakness of the aforementioned scribble-based methods is that the results are highly related to the position and number of given color scribbles. In addition, most of these methods demand significant human efforts to provide hints for ensuring plausible colorization results.

In contrast, the exemplar-based colorization approaches exploit color information from a referenced source image to guide the colorization of the target grayscale image. They reduce human effort in choosing many color scribbles and mainly focus on matching local spatial features between the reference image and the input grayscale image by using statistical analysis [26]. Recently, deep learning techniques have been employed in the exemplar-based colorization methods to further reduce human intervention during image colorization. He et al. [27] proposed the first deep exemplar-based method to transfer the colors from a reference image to the grayscale one, where the network is composed of a similarity subnetwork for automatically recommending references and a colorization subnetwork for colorizing images. A faster version of deep exemplar colorization is proposed by Xu et al. [28] with a stylization-based architecture. To generate semantically related colorized images from reference images, both Gray2ColorNet [29] with an attention-gating mechanism-based color fusion network and reference-based sketch image colorization [30] with augmented-self reference are proposed. However, the results generated by these exemplar-based methods are highly dependent on the quality of the reference images. In other words, unnatural output images would be obtained when a given reference image exhibits a large variance from the input image.

Recently, various deep learning-based models have been proposed for full automatic colorization without any human efforts [31–33]. Most early works employ a simple, straightforward architecture with stacked convolutional layers to learn the mapping from grayscale to color embeddings [34]. For example, Cheng et al. [31] first introduced deep neural networks to implement image colorization by learning a mapping function between features extracted from patches in a grayscale image and color values of the source image. An et al. [35] developed a fully automatic learning-based colorization algorithm on the VGG-16 CNN model based on the classification with the loss of cross-entropy. To extract features at different levels, multi-path networks [32,36] have been proposed. For instance, to preserve global features during model training, Iizuka et al. [32] developed two-stream networks to extract local features and global features, respectively. A fusion layer was utilized to fuse local and global information together. However, these CNN-based image colorization methods may produce blurry results because the Euclidean distance is minimized by averaging all plausible outputs [37]. More recently, to produce vivid colorization results, some generative models have been proposed for image colorization [13,16,38]. Isola et al. [13] proposed a general image-to-image translation framework based on conditional GAN to produce plausible results by adversarial training. Vitoria et al. [16] proposed ChromaGAN to colorize by combining both the perceptual and semantic understanding of color and class distributions in an adversarial training manner. To minimize semantic confusion and color bleeding, Zhao et al. [39] proposed a fully automatic saliency map-guided colorization with a generative adversarial network (SCGAN) framework, which jointly predicts the colorization and saliency map. However, most of these GAN-based image colorization approaches emphasize adversarial training to mimic the distribution of
real color images while various generated artifacts are neglected. The proposed SMCGAN framework produces high-quality colorized images while reducing the generated artifacts resulting from various backgrounds.

3. Methodology

Image colorization is an image-to-image translation problem that maps a grayscale image to a color image. In this work, we follow the work of Nazeri et al. [15] and utilize the YUV color space for the colorization task. Thus, image colorization can also be seen as a pixel-wise regression problem that maps the grayscale value of each pixel to the chromatic value of the corresponding pixel. Different from the GAN-based image colorization methods focusing on adversarial training [31], our work aims to reduce the generated artifacts caused by the various backgrounds when learning the mapping between grayscale values and chromatic values. Motivated by the work of Wang et al. [40], we introduce spatial masks in our proposed SMCGAN framework to identify foreground organs and tissues, as illustrated in Figure 1. The work in [40] aims to prune redundant computation in flat regions for CNN-based super-resolution by using spatial masks to identify important regions in feature maps. While our proposed method aims to learn robust colorization for medical images by using spatial masks on the synthesized color images. In this way, during the process of color transfer, the image colorization model is forced to focus on the foreground regions containing organs or tissues rather than on the background regions.

![Figure 1. Overview of the proposed SMCGAN. It receives a grayscale image as input and predicts a colorized image and a corresponding spatial mask. It comprises a main colorization network and an auxiliary segmentation network.](image)

3.1. SMCGAN Architecture

An overview of the SMCGAN framework is shown in Figure 1. Our method is extended from GAN architecture [41] that simultaneously produces colorized images and spatial masks from grayscale images. It is composed of a main colorization network and an auxiliary segmentation network. As we regard image colorization as image-to-image translation, we employ an image transformation network first proposed by Johnson et al. [42] as the main colorization network and the generator $G$ to colorize the input grayscale images. It comprises five residual blocks with non-residual convolutional layers followed by batch normalization and ReLU nonlinearities with the exception of the output layer. We employ the plain U-Net structure [25] as the auxiliary segmentation network. U-Net employs skip connections, which connect the output feature map generated from each level of the encoder to the corresponding level of the decoder. These skip structures have been proven to be effective in preventing gradient vanishment and simultaneously fusing low-level features and global features [43]. Both low-level features and global features play an important
role in medical image segmentation. We also follow Dong et al. and Lee et al. [44, 45] and share the same ResNet18 architecture [46] for the two discriminators. ResNet18 is widely used as a discriminator to distinguish images due to its strong feature extraction ability, which is composed of residual building blocks. The first discriminator, \( D_1 \), judges the synthesized color image and ground truth color image. A weighted colorized image is obtained by performing an element-wise product between the synthesized color image and the generated spatial mask. Similarly, a weighted ground truth color image can also be obtained by performing an element-wise product between the ground truth color image and the generated spatial mask. Thereafter, we feed the paired weighted color images to the second discriminator, \( D_2 \), which judges whether the input is real weighted color images or not. For example, given a grayscale medical image \( x \), the main colorization network first translates the grayscale medical image into a colorized image \( G(x) \) that cannot be distinguished from the ground truth color image \( x \). The discriminator \( D_1 \) is then trained to distinguish between the synthesized color image and the real color image well. The spatial mask is obtained from the auxiliary foreground segmentation task and utilized to produce the weighted ground truth color image and the weighted synthesized color image. Finally, discriminator \( D_2 \) is utilized to distinguish between the weighted ground truth color image and the weighted synthesized color image. Regarding the loss functions, a segmentation loss based on cross-entropy, an adversarial loss, and a color loss based on an L1 term is defined. We will detail each loss term in the next section.

To reduce the impact of the background when performing colorization, we introduce a spatial mask embedding in the GAN-based image colorization structure to identify the foreground regions containing organs and tissues. The spatial mask is used to act on the synthesized color image to obtain a weighted synthesized color image where most background contents are masked. A spatial mask is defined to be a binary matrix where only the pixels within the bounding box area are nonzero [47]. The spatial mask \( M_s \) has the same spatial size as the synthesized color image \( \hat{x} \), which can be written as follows:

\[
M_s(i, j) = \begin{cases} 
1, & S(\hat{x}(i, j)) \geq 0.5 \\
0, & \text{Otherwise} 
\end{cases}
\]

where \( S \) denotes the foreground segmentation network, and \( i \) and \( j \) represent the vertical and horizontal indices of an image, respectively. The term \( S(\hat{x}(i, j)) \) means the output of the foreground segmentation network using the synthesized color image \( \hat{x} \) as the input.

However, during training, there is a challenge that the binary spatial mask cannot directly backpropagate, as it is non-differentiable. Here, we utilize the Gumbel-softmax reparameterization technique [48, 49] to relax the discrete binary masks to continuous variables. Specifically, the probability of the foreground regions being selected is \( P_1^s = S(\hat{x}(i, j)) \). In contrast, the probability of the background regions being selected is \( P_0^s = 1 - S(\hat{x}(i, j)) \). Then, the sampling process of the spatial mask \( M_s \) can be reparameterized as:

\[
M_s = \arg \max_k \left( \log \left( P_k^s \right) + g_k \right), \forall k = 0, 1
\]

where \( \{g_k\}_{k=0,1} \) are random variables that follow the Gumbel distribution. To make the spatial mask \( M_s \) continuous, we replace the discontinuous function argmax with a softmax. Then, the binary learnable spatial mask from the Gumbel-softmax relaxation can be expressed as follows:

\[
M_s = \frac{\exp \left( \frac{\log (P_1^s) + g_1}{\tau} \right)}{\sum_{k \in \{0,1\}} \exp \left( \frac{\log (P_k^s) + g_k}{\tau} \right)}
\]

where \( \tau \in (0, \infty) \) is a temperature parameter. The Gumbel-softmax distribution becomes a uniform distribution when \( \tau \to \infty \). Conversely, samples from the Gumbel-softmax distri-
bution become one-hot. In other words, $M_s$ becomes a binary mask. In the experiments, we empirically start $\tau$ with a high temperature and gradually decrease it to a lower value.

To emphasize the foreground areas where organs or tissues are located, we perform element-wise product between synthesized color images and the spatial masks generated from the foreground segmentation network. The weighted output synthesized images contain regions filtering out most background contents and are sent to the discriminator for adversarial training. In this way, the foreground segmentation network is capable of assisting the main colorization network to focus on the colorization of the foreground regions instead of the background regions.

### 3.2. Loss Functions

Here, $x^y$ denotes the luminance (Y) of the input image $x$ on the YUV color space, while $x^{uv}$ represents the chrominance (UV). We propose a color loss based on the L1 loss to guide the color of the synthesized image.

$$L_{col} = \mathbb{E}[\|G(x^y) - x^{uv}\|_1 + \|G(x^y) \odot M_s - x^{uv} \odot M_s\|_1],$$  \hspace{1cm} (4)

where the operator $\odot$ means the element-wise product and $\|\cdot\|_1$ denotes the L1 norm. The color loss makes the colorization model learn the mapping between grayscale values and chromatic values. To generate target-like colored images from the luminance of the input grayscale image, we also propose two additional adversarial losses derived from two discriminators, $D_1(\cdot)$ and $D_2(\cdot)$, respectively. The proposed adversarial loss items are defined as:

$$L_{adv} = L_{gan1} + L_{gan2},$$  \hspace{1cm} (5)

$$L_{gan1} = \mathbb{E}[\log D_1(x) + \mathbb{E}[\log(1 - D_1(fc(G(x^y), x^y)))]],$$  \hspace{1cm} (6)

$$L_{gan2} = \mathbb{E}[\log D_2(x \odot M_s) + \mathbb{E}[\log(1 - D_2(fc(G(x^y), x^y) \odot M_s))]],$$  \hspace{1cm} (7)

where $D$ denotes the discriminator of the networks and $fc(\cdot)$ is a concatenation function used to concatenate the luminance and chrominance to regenerate an image. $L_{gan1}$ means the adversarial loss between the ground truth color image and the synthesized color image. $L_{gan2}$ represents the adversarial loss between the weighted ground truth color image and the weighted synthesized color image.

In the task of foreground segmentation, we adopt U-Net [25] to perform foreground segmentation with the cross-entropy loss. The spatial masks are derived from the auxiliary segmentation network. Although the learned foreground maps are not accurate during the first several iterations, the synthesized colorized medical images can still reduce visual artifacts by combining both adversarial losses and cross-entropy loss after adequate iterations. The segmentation loss can be written as:

$$L_{seg} = \mathbb{E}[m[-\log(p)] + (1 - m)[-\log(1 - p)]],$$  \hspace{1cm} (8)

where $p = S(fc(G(x^y), x^y))$ is the output of the foreground segmentation, and $m$ denotes the ground truth map of foreground segmentation (i.e., 0 for background regions and 1 for foreground regions).

By combining all the aforementioned losses, the total loss function of the proposed SMCGAN framework can be described as follows:

$$L_{total} = L_{adv} + \lambda_{col}L_{col} + \lambda_{seg}L_{seg},$$  \hspace{1cm} (9)

where $\lambda_{col}$ and $\lambda_{seg}$ are the weight parameters of the color loss function and segmentation loss function, respectively.
4. Experiments and Analysis

4.1. Experimental Settings

For datasets, we adopt the publicly available Visible Human Project (VHP) [50] dataset to evaluate our methodology. This dataset provides cross-sectional cryosection, MRI, and CT images of two cadavers, including one male cadaver and one female cadaver. We also adopt prostate T2-weighted MRI images from the NCI-ISBI 2013 challenge [51] for evaluation. This dataset contains 30 samples (578 images) collected by Radboud University Nijmegen Medical Centre with a resolution of in-plane of 0.6–0.625 mm and through-plane of 3.6–4.0 mm. For the cross-sectional cryosection images of VHP, only the thorax images and the abdomen images are selected for colorization performance evaluation, as these images contain a number of organs and tissues. The foreground masks of the cryosection images are first obtained by feeding into the pre-trained salient object detection model [52] to reduce the workload of annotations and then refined by professional doctors. We train the model on the cross-sectional cryosection datasets with original backgrounds and test it on the cross-sectional cryosection datasets with different backgrounds and the remaining datasets. All the images are rescaled to 256 × 256 and normalized within a [-1, 1] range.

For quantitative metrics, we adopt pixel-level MAE (mean absolute error) to evaluate the prediction accuracy of the synthesized color images, and PSNR (peak signal to noise ratio) and SSIM (structural similarity index) metrics to evaluate the pixel fidelity of an image [15,39,53]. For perceptual evaluation, we adopt color naturalness and color bleeding removal [39]. The color naturalness denotes whether the color of the colorized images is reasonable. For instance, the color of the same tissues should be the same. Different from color naturalness, color bleeding artifacts exist around the region boundaries of colorized images. A robust colorization system is capable of reducing such artifacts. For comparison methods, we adopt six GAN-based fully automatic colorization methods, including GAN [41], DCGAN [15], ChromaGAN [16], WGAN [54], WGAN-GP [55], and CycleGAN [56], for comparisons.

4.2. Implementation Details

For datasets, the ground truth foreground segmentation maps are generated by a pre-trained BASNet [52]. The cross-sectional cryosection images with different backgrounds are obtained by replacing their backgrounds. For network architecture, we employ the image transformation network in [42] as the generator (main colorization network), the ResNet18 architecture [46] as the two discriminators, and the U-Net structure [25] as the auxiliary segmentation network. For optimization details, we train the generator, segmentation network, and discriminators collaboratively for 10 epochs. We use the Adam optimizer with a batch size of 1 and a learning rate of 1 × 10^{-4}. The trade-off parameters \(\lambda_{col}\) and \(\lambda_{seg}\) are empirically set to 1.0 and 1.0, respectively. We start the temperature \(\tau\) at 1.0 and decrease it by \(1.0 - t/t_{max}\), where \(t\) and \(t_{max}\) denote the current and maximum training epochs, respectively. For quantitative tasks, we repeat the experiments three times and report the average performance. We implement the whole framework in Pytorch framework, using a single NVIDIA 1080Ti GPU.

4.3. Experimental Results

Comparison with state-of-the-art methods. The comparison colorization results using cross-sectional cryosection images with different backgrounds between SMCGAN and the other GAN-based colorization methods are shown in Table 1. From these results, we can observe that our proposed SMCGAN outperforms the other colorization methods in terms of MAE, achieving an average MAE of 0.019. It means that the proposed method can accurately learn the mapping between grayscale values and chromatic values. In addition, our proposed SMCGAN also ranks first in the PSNR metric as well as the SSIM metric, achieving an average PSNR of 27.42 and an average SSIM of 0.983. It demonstrates that our proposed SMCGAN could accurately model the perceptual structure of grayscale input images and also generate colorized images of good quality. DCGAN performs better than
GAN in terms of MAE, with an improvement of 0.040. WGAN and WGAN-GP perform better than both GAN and DCGAN in terms of the MAE, PSNR, and SSIM metrics as they mitigate the issue of unstable training and synthesize images with better quality. CycleGAN obtains similar performance to WGAN and WGAN-GP as the cycle consistency is utilized to generate images with content information reserved. ChromaGAN performs better than the other baseline methods in terms of PSNR metric, as the semantic information is considered when performing colorization. This indicates that a good semantic understanding of color can help the model accurately learn the mapping function for colorization. The proposed SMCGAN can be considered robust because even compared with ChromaGAN, it still has an average improvement of 0.011 and 1.17 over the MAE and PSNR metrics, respectively. This demonstrates that our method is capable of being more accurate and obtaining more robust colorization results than the other comparative methods.

**Table 1.** Comparison results using different background images as input over cross-sectional cryosection images in terms of the MAE, PSNR, and SSIM metrics. The small p-values ($p < 0.001$) calculated between our method and the comparison approaches in terms of MAE indicate the improvements are significant. ↑ denotes higher the better and ↓ represents lower the better. The best results are marked in bold.

| Methods     | Black Background | White Background | Average | $p$-Value |
|-------------|------------------|-----------------|---------|-----------|
|             | MAE ↓ PSNR ↑ SSIM ↑ | MAE ↓ PSNR ↑ SSIM ↑ | MAE ↓ PSNR ↑ SSIM ↑ |          |
| GAN         | 0.080 13.58 0.886 | 0.074 14.06 0.896 | 0.077 13.82 0.891 | $2.01 \times 10^{-15}$ |
| DCGAN       | 0.040 13.55 0.726 | 0.031 11.21 0.835 | 0.035 12.38 0.771 | $1.92 \times 10^{-10}$ |
| ChromaGAN   | 0.029 26.5 0.964 | 0.026 24.56 0.999 | 0.028 25.53 0.976 | $3.74 \times 10^{-17}$ |
| WGAN        | **0.014** 26.46 **0.989** | 0.024 23.94 0.988 | 0.019 25.20 0.988 | $1.77 \times 10^{-20}$ |
| WGAN-GP     | 0.021 24.13 0.985 | 0.024 23.55 0.987 | 0.022 23.84 0.986 | $5.00 \times 10^{-21}$ |
| CycleGAN    | **0.014** 26.21 **0.989** | 0.025 23.66 0.986 | 0.019 24.94 0.988 | $7.53 \times 10^{-12}$ |
| Ours        | 0.019 27.42 0.983 | 0.015 27.97 0.995 | 0.017 27.70 0.989 | - |

The visualization colorization results of the proposed SMCGAN and the other methods using cross-sectional cryosection images with different backgrounds are illustrated in Figure 2 for qualitative evaluation. The results from GAN and DCGAN between the two backgrounds are inconsistent in chromatic value. This demonstrates that the background of color images has a significant impact on the colorization process for GAN-based colorization algorithms. As GAN-based colorization approaches aim to synthesize the whole colorized images that are indistinguishable from the real colorized images, the perturbation in the background would lead to generating inconsistent results. Moreover, the results from GAN and DCGAN between two backgrounds exhibit a large shift in chromatic value compared with the ground truth color images. This shows that neither GAN nor DCGAN is capable of accurately learning the mapping between grayscale values and chromatic values. This inaccurate mapping between grayscale values and chromatic values also exists in WGAN, WGAN-GP, and CycleGAN. In contrast, the synthesized color images from ChromaGAN are similar to the ground truth color images. This implies that the incorporation of semantic information results in improving the ability to learn the mapping function for colorization, even if some perturbations exist in the background of color images. Compared with ChromaGAN, our proposed SMCGAN produces more similar colorized images to the ground truth color images. Although semantic information is not utilized in our proposed SMCGAN, it is still able to accurately learn the mapping between grayscale values and chromatic values. This demonstrates the utility of introducing spatial masks, which force the colorization model to emphasize the foreground regions and improve the performance of image colorization.
To evaluate the generalization capability of the proposed SMCGAN in colorization qualitatively, we also produce visualization colorization results of different GAN-based image colorization methods using MRI and CT datasets, as illustrated in Figures 3 and 4. It is to be noted that ground truth color images are not available in these datasets. Consequently, we evaluated the performance through a perceptual analysis, including color naturalness and color bleeding removal. From the results in Figure 3a, we can observe that the color of the colorized images generated by GAN are almost the same. It is not reasonable because different types of tissues should be rendered in different colors. This indicates that GAN fails to learn the mapping between different grayscale values and different chromatic values. In contrast, different tissues of colorized images in DCGAN are rendered in different colors. However, some color-bleeding artifacts exist between the borders of different tissues. For instance, the colors of some tissues bleed into the other tissues, as seen in the first column image of DCGAN. This shows that DCGAN fails to capture the semantic information when performing image colorization. Similar artifacts also exist in WGAN, WGAN-GP, and CycleGAN. The colorized images generated...
by ChromaGAN contain fewer artifacts than the other baseline methods. As ChromaGAN considers semantic information during image colorization, it is able to learn the relationship between tissues and colors. Compared with ChromaGAN, the colorized images generated by our proposed SMCGAN are more reasonable and natural as the colors are diverse and the structural details of the tissues are clearer. The results in Figure 3b are similar to those in Figure 3a. Both ChromaGAN and our proposed SMCGAN achieve better performance than GAN and DCGAN in terms of color naturalness and color-bleeding removal. As many details are missing in the CT images, it is more practical to compare the MRI images to test the colorization performance. From the image colorization results of prostate T2-weighted MRI in Figure 4, we can observe that the colorized images generated by our approach contain a clearer structure and fewer artifacts compared with those synthesized by the other image colorization algorithms. From these results, we can conclude that our approach is capable of generating robust results and reducing color-bleeding artifacts.

Figure 3. Qualitative colorization results of MRI images and CT images of the VHP dataset using different GAN-based image colorization methods.

The effectiveness of using the foreground maps. Although the generalization ability of our proposed approach has been validated, the effectiveness of using the foreground map is still required to be investigated. Therefore, we perform a comparison between using the coarse foreground maps and refined maps, as shown in Table 2. Note that the coarse foreground maps directly obtained from the pre-trained salient object detection model [52] contain many noise labels, while the refined maps have accurate annotations on the cryosection images. From the results, we can conclude that the performance of the model using the coarse foreground maps decreased significantly as they may mislead the model in learning to distinguish the foreground regions and background regions.
Figure 4. Qualitative colorization results of the prostate T2-weighted MRI using different image colorization approaches.

Table 2. Performance comparison between using refined foreground maps and coarse foreground maps. ↑ denotes higher the better and ↓ represents lower the better.

| Methods  | MAE ↓ | PSNR ↑ | SSIM ↑ | MAE ↓ | PSNR ↑ | SSIM ↑ | MAE ↓ | PSNR ↑ | SSIM ↑ |
|----------|-------|--------|--------|-------|--------|--------|-------|--------|--------|
| Ours     | 0.025 | 21.31  | 0.972  | 0.042 | 19.34  | 0.970  | 0.033 | 20.33  | 0.971  |
| Ours     | 0.019 | 27.42  | 0.983  | 0.015 | 27.97  | 0.995  | 0.017 | 51.81  | 0.989  |

Ablation analysis of our approach. To further investigate the effectiveness of several loss functions, we analyzed the different loss functions of our proposed SMCGAN on cross-sectional cryosection dataset quantitatively. Basically, there are three settings to exclude some parts from the original structure: (1) Drop the color loss to investigate the effect of the color loss $L_{col}$. This setting will not affect the network architecture. (2) Drop the discriminators and generators for colorized images, with adversarial training to analyze the effect of the adversarial loss $L_{gan1}$ in SMCGAN. (3) Drop the discriminators and generators for weighted colorized images, with adversarial training to analyze the effect of the adversarial loss $L_{gan2}$ in SMCGAN.

The quantitative analysis of the ablation study is summarized in Table 3. First, if the color loss is removed, the model tends to fail to accurately learn the mapping between the grayscale values and the chromatic values. The performance drops significantly compared with full losses in terms of MAE, PSNR, and SSIM. Secondly, if the gan1 loss is dropped, the performance is still inferior to full losses. The gan1 loss plays a key role in modeling the intensity distribution of colorized image values. Finally, the SMCGAN without gan2 loss produces results that are worse than the SMCGAN without gan1 loss. This result also demonstrates the effectiveness of the spatial mask in the improvement of colorization performance. In conclusion, each component of the proposed SMCGAN is indispensable.
Table 3. Quantitative results of the ablation study on the cryosection dataset. ↑ denotes higher the better and ↓ represents lower the better.

| Methods          | Black Background | White Background | Average |
|------------------|------------------|------------------|---------|
|                  | MAE ↓ | PSNR ↑ | SSIM ↑ | MAE ↓ | PSNR ↑ | SSIM ↑ | MAE ↓ | PSNR ↑ | SSIM ↑ |
| w/o color        | 0.040 | 20.59  | 0.947  | 0.049 | 19.28  | 0.966  | 0.045 | 19.94  | 0.957  |
| w/o gan1         | 0.015 | 27.88  | 0.987  | 0.022 | 25.37  | 0.994  | 0.019 | 26.63  | 0.991  |
| w/o gan2         | 0.049 | 20.97  | 0.901  | 0.028 | 23.50  | 0.991  | 0.039 | 22.24  | 0.946  |
| Ours             | 0.019 | 27.42  | 0.983  | 0.015 | 27.97  | 0.995  | 0.017 | 27.70  | 0.989  |

As shown in Figure 5, the full SMCGAN produces the best perceptual results compared with the three settings of the ablation study. If the color loss is removed, the color of the generated images exhibits a large discrepancy in chromatic value compared with the ground truth color images. Moreover, the generated images are less colorful than the full SMCGAN. This demonstrates the importance of color loss in learning the mapping between grayscale values and chromatic values. If gan1 loss is reduced, the color of the synthesized images still has less shift in chromatic values than full SMCGAN. If gan2 loss is reduced, color-bleeding artifacts are generated in the samples. This result also validates the effectiveness of introducing spatial masks to make the model focus on the colorization of foreground regions and reduce generated artifacts.

Figure 5. A comparison of colorization results under different ablation study settings. (a) Colorization results using images with black background as input; (b) Colorization results using images with white background.

5. Discussion

Although the proposed SMCGAN has achieved robust colorization results over cryosection images, CT images, and MRI images, there are still several issues needed to be further discussed.

Why the BASNet is utilized to obtain segmentation maps. Background subtraction (BGS) aims to segment the foreground objects from their surroundings in a given image [57–59]. The key issue of BGS is how to improve the accuracy of the detection of the foreground [57]. The Gaussian mixture model (GMM) is one of the widely used models for background modeling [58,59], which attempts to model color intensity variations as a mixture of Gaussians at the pixel level. However, the GMM-based methods are not robust and suffer from performance degradation when they encounter complex scenes. Recently, CNN-based methods have demonstrated significant performances for foreground
detection [60–62]. However, most of these methods follow the trend of recent generic image segmentation, including multi-scale feature aggregation, concatenated features from different layers, and multi-scale inputs [63]. Different from these aforementioned methods, BASNet pays more attention to the finer structures rather than the large structures [52]. We thus employed a pre-trained BASNet to obtain segmentation maps of the VHP dataset as ground truth, considering small tissues such as blood vessels are required to be treated carefully during the image segmentation process.

**Why not utilize the foreground map as a mask?** To guide the image generation network focusing on the colorization of organs or tissues, we embed the spatial mask obtained from the foreground map into the whole network. As the foreground map is a non-differentiable binary spatial map, we can not directly employ it in the network, considering the backpropagation during the model training process. We thus propose to utilize the Gumbel-softmax reparameterization technique [48,49] to relax the discrete binary masks to continuous variables. In this way, the segmentation subnetwork is capable of assisting the image colorization network in paying more attention to the colorization of foreground organs or tissues instead of the background noises.

**How to avoid miss-coloring of the regions with indistinct border margins?** We have proposed a fully automatic approach utilizing end-to-end learning to directly learn the mapping between an input grayscale image and the corresponding color image without the requirement of user intervention. Like other popular image colorization approaches [64], we also adopt the widely-used L1 loss to regularize the difference between the synthetic color image and the real color image on the chrominance space. Nevertheless, the L1 loss treats the colorization as a regression or classification problem at the pixel level, which may fail in accurately coloring the regions with indistinct border margins. To alleviate this issue, some previous works [32,65] encoded the semantic information to guide image colorization at the image level. However, most of these methods are designed for specific scenes relying on pre-trained models utilized for feature extraction. Different from these methods, we propose to employ an auxiliary segmentation branch with spatial mask to help focus more on the organs or tissues while they are less influenced by the background. In this way, our proposed method is capable of accurately coloring the regions between the foreground objects and the backgrounds, even with indistinct border margins.

**Disadvantages and future works.** Although the proposed SMCGAN can generate relatively robust colorized images in most cases, there are still some failure cases, as shown in Figure 2. The color breeding artifacts still exist in some generated images. As there is no specific loss item for enhancing colors by considering the semantic information of medical images, it is difficult to identify the colors for organs or tissues at the pixel level when the background content changes significantly. Some previous works [32,66] have employed pre-trained VGG models to extract semantic information; these VGG models are trained on natural image datasets, which may not be capable of effectively capturing the semantic features for medical images due to the large discrepancy compared with natural images. In the future, we will develop new methods for medical image colorization while considering semantic information.

6. Conclusions

We presented a fully automatic SMCGAN colorization framework for medical images to reduce the generated artifacts. It simultaneously generates colorized images and their corresponding spatial masks from grayscale input images by introducing an auxiliary foreground segmentation network combined with the main colorization network. The generated spatial masks can be used to generate weighted synthesized color images where most background contents are filtered out, which assists the discriminator in emphasizing the colorization of foreground regions and reducing the generated artifacts. We validated our proposed framework on the publicly available VHP dataset and the prostate dataset from NCI-ISBI 2013 challenge, compared with six state-of-the-art GAN-based image colorization.
The experimental results demonstrated that SMCGAN can generate robust colorized medical images and reduce generated artifacts.

**Author Contributions:** Conceptualization, Z.Z.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z.; formal analysis, Z.Z.; investigation, Z.Z.; resources, Z.Z.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.L. and B.-S.S.; visualization, Z.Z.; supervision, Y.L. and B.-S.S.; project administration, Y.L. and B.-S.S.; funding acquisition, Y.L. and B.-S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (No. NRF-2022R1A2B5B0101553 and No. NRF-2022R1A4A1033549) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the data used in this study are publicly available at 18 November 2022 https://www.nlm.nih.gov/research/visible/visible_human.html.

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

**References**

1. Zaffino, P.; Marzullo, A.; Moccia, S.; Calimeri, F.; De Momi, E.; Bertucci, B.; Arcuri, P.P.; Spadea, M.F. An open-source covid-19 ct dataset with automatic lung tissue classification for radiomics. *Bioengineering* 2021, 8, 26. [CrossRef]

2. Lee, J.; Kim, J.N.; Gomez-Perez, L.; Gharaiheb, Y.; Motoairek, I.; Pereira, G.T.; Zimin, V.N.; Dallan, L.A.; Hoori, A.; Al-Kindi, S.; et al. Automated segmentation of microvessels in intravascular OCT images using deep learning. *Bioengineering* 2022, 9, 648. [CrossRef]

3. Tang, Y.; Cai, J.; Lu, L.; Harrison, A.P.; Yan, K.; Xiao, J.; Yang, L.; Summers, R.M. CT image enhancement using stacked generative adversarial networks and transfer learning for lesion segmentation improvement. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 46–54.

4. Luo, J.; Wu, M.; Gopukumar, D.; Zhao, Y. Big data application in biomedical research and health care: A literature review. *Biomed. Inform. Insight* 2016, 8, 1–10. [CrossRef]

5. Wei, W.; Zhou, B.; Polap, D.; Woźniak, M. A regional adaptive variational PDE model for computed tomography image reconstruction. *Pattern Recognit.* 2019, 92, 64–81. [CrossRef]

6. Kaur, M.; Singh, M. Contrast Enhancement and Pseudo Coloring Techniques for Infrared Thermal Images. In Proceedings of the 2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 22–24 October 2018; IEEE: Manhattan, NY, USA, 2018; pp. 1005–1009.

7. Dabass, J.; Vig, R. Biomedical image enhancement using different techniques-a comparative study. In *International Conference on Recent Developments in Science, Engineering and Technology*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 260–286.

8. Wang, H.; Liu, X. Overview of image colorization and its applications. In Proceedings of the IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; IEEE: Manhattan, NY, USA, 2021; pp. 1561–1565.

9. Kaur, M.; Singh, M. Contrast Enhancement and Pseudo Coloring Techniques for Infrared Thermal Images. In Proceedings of the 2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 22–24 October 2018; IEEE: Manhattan, NY, USA, 2018; pp. 1005–1009.

10. Iizuka, S.; Simo-Serra, E. Deepremaster: Temporal source-reference attention networks for comprehensive video enhancement. *ACM Trans. Graph. (TOG)* 2019, 38, 1–13. [CrossRef]

11. 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; pp. 1125–1134.

12. Lei, C.; Chen, Q. Fully automatic video colorization with self-regularization and diversity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3753–3761.

13. Nazeri, K.; Ng, E.; Ebrahimi, M. Image colorization using generative adversarial networks. In *International Conference on Articulated Motion and Deformable Objects*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 85–94.
16. Vitoria, P.; Raad, L.; Ballester, C. Chromagan: Adversarial picture colorization with semantic class distribution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2445–2454.

17. Zhang, X.; Karaman, S.; Chang, S.F. Detecting and simulating artifacts in gan fake images. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Delft, The Netherlands, 9–12 December 2019; pp. 1–6.

18. Marra, F.; Saltori, C.; Boato, G.; Verdoliva, L. Incremental learning for the detection and classification of gan-generated images. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Delft, The Netherlands, 9–12 December 2019; pp. 1–6.

19. Zhan, F.; Zhu, H.; Lu, S. Spatial fusion gan for image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3653–3662.

20. Xiong, W.; Yu, J.; Lin, Z.; Yang, J.; Lu, X.; Barnes, C.; Luo, J. Foreground-aware image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5840–5848.

21. Chen, X.; Zou, D.; Zhao, Q.; Tan, P. Manifold preserving edit propagation. ACM Trans. Graph. (TOG) 2021, 31, 1–7. [CrossRef]

22. Xu, K.; Li, Y.; Ju, T.; Hu, S.M.; Liu, T.Q. Efficient affinity-based edit propagation using kd tree. ACM Trans. Graph. (TOG) 2009, 28, 1–6.

23. Sangkloy, P.; Lu, J.; Fang, C.; Yu, F.; Hays, J. Scribbler: Controlling deep image synthesis with sketch and color. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5400–5409.

24. Xiao, Y.; Zhou, P.; Zheng, Y.; Leung, C.S. Interactive deep colorization using simultaneous global and local inputs. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Manhattan, NY, USA, 2019; pp. 1887–1891.

25. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

26. Li, B.; Zhao, F.; Su, Z.; Liang, X.; Lai, Y.K.; Rosin, P.L. Example-based image colorization using locality consistent sparse representation. IEEE Trans. Image Process. 2017, 26, 5188–5202. [CrossRef]

27. He, M.; Chen, D.; Liao, J.; Sander, P.V.; Yuan, L. Deep exemplar-based colorization. ACM Trans. Graph. (TOG) 2018, 37, 1–16. [CrossRef]

28. Xu, Z.; Wang, T.; Fang, F.; Sheng, Y.; Zhang, G. Stylization-based architecture for fast deep exemplar colorization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9363–9372.

29. Lu, P.; Yu, J.; Peng, X.; Zhao, Z.; Wang, X. Gray2colornet: Transfer more colors from reference image. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3210–3218.

30. Lee, J.; Kim, E.; Lee, Y.; Kim, D.; Chang, I.; Choo, J. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5801–5810.

31. Cheng, Z.; Yang, Q.; Sheng, B. Deep colorization. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 415–423.

32. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Trans. Graph. (TOG) 2016, 35, 1–11. [CrossRef]

33. Zhang, R.; Isola, P.; Efros, A.A. Colorful image colorization. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

34. Anwar, S.; Tahir, M.; Li, C.; Mian, A.; Khan, F.S.; Muzaffar, A.W. Image colorization: A survey and dataset. arXiv 2020, 2008.10774, arXiv:2008.10774.

35. An, J.; Kpeiyou, K.G.; Shi, Q. Grayscale images colorization with convolutional neural networks. Soft Comput. 2020, 24, 4751–4758. [CrossRef]

36. Zhao, J.; Han, J.; Shao, L.; Snoek, C.G. Pixelated semantic colorization. Int. J. Comput. Vis. 2020, 128, 818–834. [CrossRef]

37. 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.

38. Liang, Y.; Lee, D.; Li, Y.; Shin, B.S. Unpaired medical image colorization using generative adversarial network. Multimed. Tools Appl. 2021, 81, 26669–26683. [CrossRef]

39. Zhao, Y.; Po, L.M.; Cheung, K.W.; Yu, W.Y.; Rehman, Y.A.U. SCGAN: Saliency map-guided colorization with generative adversarial network. IEEE Trans. Circuits Syst. Video Technol. 2020, 31, 3062–3077. [CrossRef]

40. Wang, L.; Dong, X.; Wang, Y.; Ying, X.; Lin, Z.; An, W.; Guo, Y. Exploring Sparsity in Image Super-Resolution for Efficient Inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 4917–4926.

41. Goodfellow, I.; 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]

42. 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; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.
43. Du, G.; Cao, X.; Liang, J.; Chen, X.; Zhan, Y. Medical image segmentation based on u-net: A review. *J. Imaging Sci. Technol.* 2020, 64, 1–12. [CrossRef]

44. Dong, N.; Xu, M.; Liang, X.; Jiang, Y.; Dai, W.; Xing, E. Neural architecture search for adversarial medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 828–836.

45. Lee, H.H.; Tang, Y.; Tang, O.; Xu, Y.; Chen, Y.; Gao, D.; Han, S.; Gao, R.; Savona, M.R.; Abramson, R.G.; et al. Semi-supervised multi-organ segmentation through quality assurance supervision. In *Proceedings of the Medical Imaging 2020: Image Processing*; SPIE, Houston, TX, USA, 15–20 February 2020; Volume 11313, pp. 365–369.

46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

47. Liang, K.; Gao, Y.; Chang, H.; Chen, X. Visual relationship detection with deep structural ranking. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, 2–7 February 2018; pp. 7098–7105.

48. Maddison, C.J.; Mnih, A.; Teh, Y.W. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv* 2016, arXiv:1611.00712.

49. Li, F.; Li, G.; He, X.; Cheng, J. Dynamic Dual Gating Neural Networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, 10–17 October 2021; pp. 5330–5339.

50. Spitzer, V.M.; Whitlock, D.G. The Visible Human Dataset: The anatomical platform for human simulation. *Anat. Rec. Off. Publ. Am. Assoc. Anat.* 1998, 253, 49–57. [CrossRef]

51. Liu, Q.; Dou, Q.; Yu, L.; Heng, P.A. MS-Net: Multi-site network for improving prostate segmentation with heterogeneous MRI data. *IEEE Trans. Med. Imaging* 2020, 39, 2713–2724. [CrossRef]

52. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. Basnet: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15–20 June 2019; pp. 7479–7489.

53. Zeng, X.; Tong, S.; Lu, Y.; Xu, L.; Huang, Z. Adaptive Medical Image Deep Color Perception Algorithm. *IEEE Access* 2020, 8, 56559–56571. [CrossRef]

54. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In *Proceedings of the International Conference on Machine Learning*, PMLR, Sydney, NSW, Australia, 6–11 August 2017; pp. 214–223.

55. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. In *Proceedings of the Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.

56. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

57. Sultana, M.; Mahmood, A.; Javed, S.; Jung, S.K. Unsupervised deep context prediction for background estimation and foreground segmentation. *Mach. Vis. Appl.* 2019, 30, 375–395. [CrossRef]

58. Stauffer, C.; Grimson, W.E.L. Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ft. Collins, CO, USA, 23–25 June 1999; IEEE: Manhattan, NY, USA, 1999; Volume 2, pp. 246–252.

59. Lu, X. A multiscale spatio-temporal background model for motion detection. In *Proceedings of the IEEE International Conference on Image Processing* (ICIP), Paris, France, 27–30 October 2014; IEEE: Manhattan, NY, USA, 2014; pp. 3268–3271.

60. Wang, Y.; Luo, Z.; Jodoin, P.M. Interactive deep learning method for segmenting moving objects. *Pattern Recognit. Lett.* 2017, 96, 66–75. [CrossRef]

61. Zeng, D.; Zhu, M. Background subtraction using multiscale fully convolutional network. *IEEE Access* 2018, 6, 16010–16021. [CrossRef]

62. Lim, L.A.; Keles, H.Y. Foreground segmentation using convolutional neural networks for multiscale feature encoding. *Pattern Recognit. Lett.* 2018, 112, 256–262. [CrossRef]

63. Sakkos, D.; Ho, E.S.; Shum, H.P. Illumination-aware multi-task GANs for foreground segmentation. *IEEE Access* 2019, 7, 10976–10986. [CrossRef]

64. Zhang, B.; He, M.; Liao, J.; Sander, P.V.; Yuan, L.; Bermak, A.; Chen, D. Deep exemplar-based video colorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 16–20 June 2019; pp. 8052–8061.

65. Larsson, G.; Maire, M.; Shakhnarovich, G. Learning representations for automatic colorization. In *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 577–593.