applied sciences

MDPI

Article

Synthesize and Segment: Towards Improved Catheter Segmentation via Adversarial Augmentation

Ihsan Ullah [1,†], Philip Chikontwe [1,†], Hongsoo Choi [1,2], Chang-Hwan Yoon [3] and Sang Hyun Park [1,*]

[1] Department of Robotics Engineering, DGIST, Daegu 42988, Korea; Ihsankhan@dgist.ac.kr (I.U.); philipchicco@dgist.ac.kr (P.C.); mems@dgist.ac.kr (H.C.)
[2] DGIST-ETH Microrobotics Research Center, DGIST, Daegu 42988, Korea
[3] Division of Cardiology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam-si 13620, Korea; changhwanyoon@gmail.com
* Correspondence: shpark13135@dgist.ac.kr
† These authors contributed equally to this work.

Abstract: Automatic catheter and guidewire segmentation plays an important role in robot-assisted interventions that are guided by fluoroscopy. Existing learning based methods addressing the task of segmentation or tracking are often limited by the scarcity of annotated samples and difficulty in data collection. In the case of deep learning based methods, the demand for large amounts of labeled data further impedes successful application. We propose a synthesize and segment approach with plug in possibilities for segmentation to address this. We show that an adversarially learned image-to-image translation network can synthesize catheters in X-ray fluoroscopy enabling data augmentation in order to alleviate a low data regime. To make realistic synthesized images, we train the translation network via a perceptual loss coupled with similarity constraints. Then existing segmentation networks are used to learn accurate localization of catheters in a semi-supervised setting with the generated images. The empirical results on collected medical datasets show the value of our approach with significant improvements over existing translation baseline methods.

Keywords: adversarial learning; catheter robot; convolutional neural networks; image translation

Citation: Ullah, I.; Chikontwe, P.; Choi, H.; Yoon, C.-H.; Park, S.H. Synthesize and Segment: Towards Improved Catheter Segmentation via Adversarial Augmentation. *Appl. Sci.* **2021**, *11*, 1638. https://doi.org/10.3390/app11041638

Academic Editor: Francesco Bianconi
Received: 31 December 2020
Accepted: 7 February 2021
Published: 11 February 2021

1. Introduction

In minimally invasive surgery (MIS), catheters and guidewires are often used for precise and targeted interventions offering several advantages over conventional procedures, such as faster recovery and less pain for patients (Hereon, 'catheter' and 'guidewire' will be used interchangeably). In general, MIS requires real-time imaging to visualize anatomy and accurately manipulate the tools that are involved in delivering treatment. For example, in cardiac catheterization, a catheter is inserted into the body with real-time continuous monitoring and guidance via X-ray imaging. However, because catheters are thin flexible tubes of varying stiffness, they can be easily mistaken for anatomy during placement and lead to severe complications for patients. Moreover, difficulties in segmenting the catheter may occur as dye is injected into the vein for the improved visibility of blocked or narrowed veins. Expert radiologists are often required to carefully monitor catheters in X-ray for the accurate maneuvering of surgical tools with minimal errors. Thus, an automated system is vital in augmenting the surgical ability of experts during interventions.

An automatic system is desired to detect, track, and segment the catheter in X-ray images to mitigate risks for both experts and patients during interventions. However, precise and reliable detection of catheters in these systems is a challenging task. Although most catheters provide radio-opaque markers to ease detection, they may be less visible due to projection angles [1]. The catheter can also be easily confused for wire-like structures and anatomies, such as surgical sutures, stipples, vessels, and ribs, etc., often showing similarities in structural appearance. Thus, existing works focus on improving visual analysis to reduce ambiguity via improved segmentation and detection techniques [2–4].

For example, Viswanathan et al. [5] proposed an approach to localize surgical tools in X-ray; however, they solely rely on primitive low level features, such as intensity, and often fail to accurately localize the catheter and other tools due to shape complexity and appearance ambiguity. To mitigate this, information regarding the shape of the catheter is used by hessian based line filtering [6] to improve performance. Following this line of work, Vandini et al. [7] developed a more sophisticated approach that is based on segment-like features (SEG-lets) to resolve large deformations between successive frames in video. However, this method requires domain knowledge and does not guarantee precise results under noisy conditions.

Latest advances in deep convolutional neural networks (CNN) have enabled significant improvements in the performance of catheter detection systems [8,9]. It is worth noting that the success of these deep learning based methods primarily depends on the availability of large and high-quality training data. Nevertheless, several areas, including medical imaging, suffer from a lack of sufficient data and require precise expert annotation to guarantee quality. Annotation is often non-trivial due to the heterogeneity and complexity of data per modality. To cope with these challenges, recent methods for segmentation follow the following trends; (i) employ extensive augmentation techniques when data is limited [10], (ii) the use of synthetic samples [11], or (iii) creating annotations that are based on simple low-level morphological operators [12]. However, training deep learning models with assumptions such as (iii) may reduce the ability of the models to generalize. Our work falls under categories (i) and (ii), and we aim to design a more general solution via Adversarial based augmentation.

In the literature, generative adversarial networks (GANs) [13] are the existing standard for synthetic image generation, with applications in both X-ray translation and pseudo data augmentation. For example, Tmenova et al. [14] proposed transfering the style of real X-ray acquired during interventions into artificial phantom arteries via CycleGAN [15]. However, artifacts are noticeable in the generated images and the method fails to represent fine details that are related to the structure of the arteries. Additionally, precise parameter tuning is required for effective generation. More recently, Lee et al. [16] designed a framework for data augmentation that adjusts a set of X-ray images with arbitrary intensity distribution to match the specific intensity distribution of chest X-ray images via GANs. Although impressive, GANs are known to be difficult to train and often require careful optimization of the adversarial objectives for stability.

In this work, we investigate and evaluate a deep learning based method for catheter synthesis and segmentation. In particular, we focus on the translation of in-painted catheter X-ray images to realistic X-ray images via the CycleGAN architecture as a form of augmentation for improved segmentation in limited settings. Aside from the standard cycle loss, we incorporate: (i) a perceptual loss to generate high-quality synthetic X-ray images and (ii) include a similarity loss to avoid large deformations from the original distribution of the real X-ray image as a constraint. A thorough evaluation of the generated images using state-of-the-art segmentation models demonstrates that the proposed method can indeed improve catheter segmentation performance. The main contributions of this work are highlighted below:

- Synthetic X-ray from labels: we propose to generate synthetic X-ray from in-painted catheter masks via adversarial learning with CycleGAN as data augmentation for segmentation.
- Improved generation with perceptual losses: to achieve more realistic generation from in-painted catheter masks, we incorporate a perceptual loss alongside the standard cycle loss.
- Enforcing semantic similarity: we further propose a similarity loss to alleviate large deviations in the semantic quality of the generated images from the original.
- Empirical results and several ablations show the effectiveness of the proposed training scheme with segmentation performance improving as synthetic augmentation is increased.

The presented work is an extended version of the study presented at a conference [17]. Herein, (i) the proposed CycleGAN X-ray translation method is further improved by using catheter masks in-painted in X-ray images with no real catheter present to mitigate the problem of generated images showing a progressively vanished catheter as training proceeds, (ii) catheter segmentation in fluoroscopic X-ray images is further explored based on the generated images, (iii) thorough quantitative and qualitative results are reported to validate the proposed methods, and (iv) several ablation studies are presented in order to assess the performance of catheter segmentation with the generated data.

The rest of the paper is organized, as follows. First, we revisit prior works regarding catheter segmentation, detection, and translation in Section 2. We present our proposed method in Section 3, and show the evaluation results regarding catheter synthesis and segmentation in Section 4. We conclude in Section 6.

2. Related Work

In this section, the current literature on segmentation and detection of catheters is discussed. In addition, we highlight works that are related to semi-/unsupervised synthesis for medical imaging.

2.1. Learning Based Methods for Segmentation and Detection

Several works have been proposed to address the localization of catheters in medical images. For example, Mercan et al. [18] introduced a CNN for catheter segmentation in chest X-ray images with curve fitting being used to connect line segments. Nguyen at al. [19] proposed to learn end-to-end temporal continuity between frames coupled with flow guided warping techniques for improved segmentation in endovascular interventions. Mountney et al. [20] further suggested a method for extracting the needle in X-ray images. However, this method might not be able to accurately track the tools in X-ray due to the flexible nature of the catheter. Wang et al. [21] detect catheter tips using region proposal networks (RPN); however, exact pixel wise locations are required for tracking to be effective. Recently, Ullah et al. [22] track catheter robot tips on successive frames via detection and segmentation in natural camera images. A more recent work by Lee et al. [23] segments the catheter tip position in chest X-ray via a fully convolutional network (FCN). Shaohan et al. [24] and Ambrosini et al. [25] employed similar approaches for catheter segmentation in ultrasound and X-ray images using a UNet [26] architecture. However, existing methods require large amounts of annotated data for supervised training in order to achieve significant improvements in accuracy. Moreover, collection is non-trivial due to privacy restrictions for patient data.

Recent works [27,28] have employed conventional data augmentation techniques as a strategy to increase existing data samples to avoid over-fitting and increase performance on limited datasets. In particular, some studies have investigated the effects of geometric transformations, such as translation, reflection, cropping, and the alteration of color schemes, to improve efficiency. Kooi et al. [29] suggested using scaling and translations to augment data for mammography lesions detection. Similar data augmentation techniques were employed with elastic deformations for surgical robot segmentation in [22]. However, the direct application of standard data augmentation techniques is often problem specific and it requires careful selection.

2.2. Image Translation in Medical Imaging

Alternatively, GANs [13] have been used to generate realistic images that boost the performance of vision tasks, such as segmentation. Moreover, several studies show that GANs are able to synthesize high-quality images by learning the data distribution in order to augment the training data. Recently, Zaman et al. [30] used the pix2pix [31] framework for ultrasound bone image generation and improved segmentation performance. Wolterink et al. [32] used an unpaired CycleGAN-based approach for brain computed tomography (CT) image synthesis from magnetic resonance (MR) images with notable improvements for

the task of cross modality-synthesis. Although image synthesis has been applied to various problems in the medical imaging domain [33–35], there are a few works that address catheter synthesis and segmentation in X-rays. A notable work is that of Gherardini et al. [36]; the authors showed the feasibility of using synthetic data to segment catheters via a transfer learning approach. Yi et al. [37] proposed simulated catheter data generation and then used recurrent neural networks to process multi-scale inputs for catheter segmentation. Frid-Adar et al. [38] used similar approaches for synthesizing endotracheal (ET) tubes in chest X-ray images with CNNs used for the classification and segmentation of ET tubes. However, these approaches may fail to detect the catheter or tube due to the large domain shift between the simulated training images and actual X-ray test images.

3. Methods

Figure 1 depicts the architecture of the proposed approach, which consists of two main stages. In the first stage, we generate realistic catheters in fluoroscopy from in-painted catheters in X-ray used as input via CycleGAN. Second, a segmentation network is trained on the generated images to segment the catheter in real X-ray images.

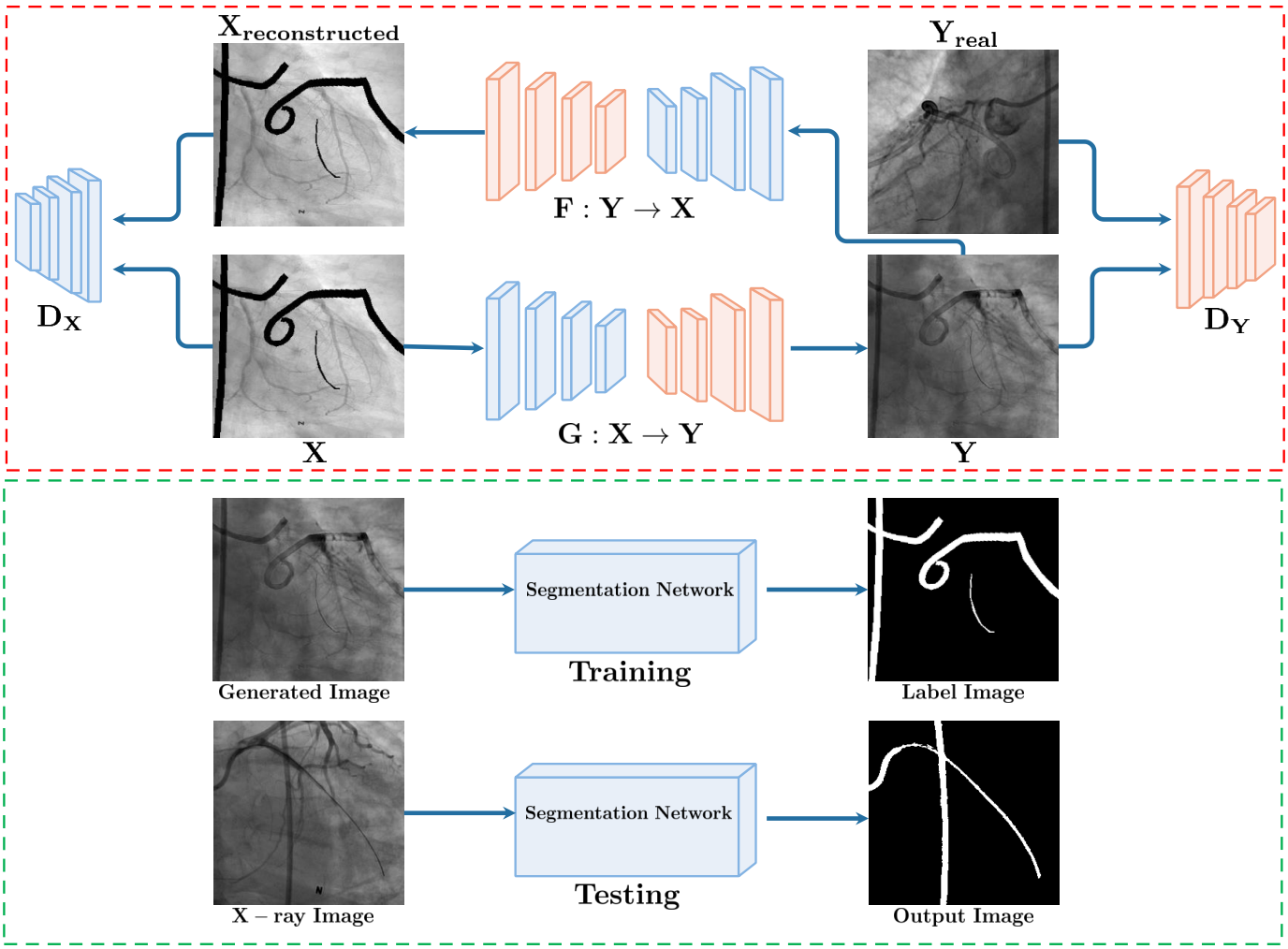


Figure 1. Overall framework of the proposed synthesis and segmentation of catheters in X-ray. The catheter sythesis is represented with red dotted line, while the catheter segmentation is represented with a sgreen dotted line.

3.1. Synthesize: GAN Based X-ray Translation

In this study, we achieve the task of synthesizing realistic X-ray from catheter masks using GANs. This can be considered as an image-to-image translation problem from real

to synthetic using a generator G and a discriminator network D in an unsupervised setting. Herein, G is trained to map random vectors $z \in \mathcal{R}^z$ to a synthetic vector $a = G(z)$, with D distinguishing real from synthetic samples. To achieve this, we employ a CycleGAN [15] architecture and augment the existing loss functions for realistic generation. Figure 1 (top) shows the framework.

Following the formulation presented in [17], we consider X as the domain of X-ray images composited with catheter masks $x_i \in X^n$ and Y as the domain of the original X-ray images $y_i \in Y^n$ with no existing catheter annotations i.e., in a single video, we select frames that do not show an inserted catheters. In a similar fashion, the composite is obtained via $x_i = \hat{c}_i \oplus y_i$, where $\hat{c}_i$ is the binary mask highlighting the pixel location of the catheter. Following, we use x_i as input in G.

In the adversarial framework, mappings between $G : X \rightarrow Y$ and $F : Y \rightarrow X$ are learned for arbitrary unpaired samples i.e., both in forward and backward cycles. A residual network (ResNet) [39] with skip connections is used as the network G and a network [31] consisting of five convolutional layers followed by batch normalization and leaky ReLU is used as D, respectively. The objective function of the forward cycle $G : X \rightarrow Y$ and D_Y is formally expressed as:

$$min_G max_{D_Y} \mathcal{L}_{adv}(G, D_Y, X, Y) = \mathbb{E}[logD_Y(y)] + \mathbb{E}[log(1 - D_Y(G(x))]. \quad (1)$$

On the other hand, for the backward cycle $F : Y \rightarrow X$, the objective is $min_F max_{D_X}$ $\mathcal{L}_{adv}(F, D_X, Y, X)$. In order to achieve cycle-consistency between the generated samples in both cycles, a consistency loss is employed following:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}[||F(G(x)) - x||_1] + \mathbb{E}[||G(F(y) - y)||_1], \quad (2)$$

collectively, the loss is defined as

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{adv}(G, D_Y, X, Y) + \mathcal{L}_{adv}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F). \quad (3)$$

We further propose to include additional terms based on the structural similarity (SSIM) [40] and perceptual losses [41] in order to enforce consistency in semantic quality in both the later and earlier levels of the networks. Formally, SSIM for a pixel p is defined as

$$SSIM(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \mathcal{L}_{ssim}(P) = \frac{1}{N} \sum_{p \in P} 1 - SSIM(p), \quad (4)$$

where μ and σ represent the mean and standard deviation of the inputs that are computed using a Gaussian filter. Moreover, SSIM has been shown to enforce visual consistency in generated images.

On the other hand, a perceptual loss enables the use of earlier level model features to improve learning. The loss is defined as the Euclidean distance between the feature maps of the original image X and the reconstructed image $\hat{X}$. This loss is formulated as:

$$\mathcal{L}_{prep}\theta^{i,j}(X, \hat{X}) = \frac{1}{H_{ij}W_{ij}} \sum_{x,y=1}^{H_{ij}W_{ij}} \theta, \text{with } \theta = (\theta_{ij}(X)_{xy} - (\hat{X})_{xy})^2, \quad (5)$$

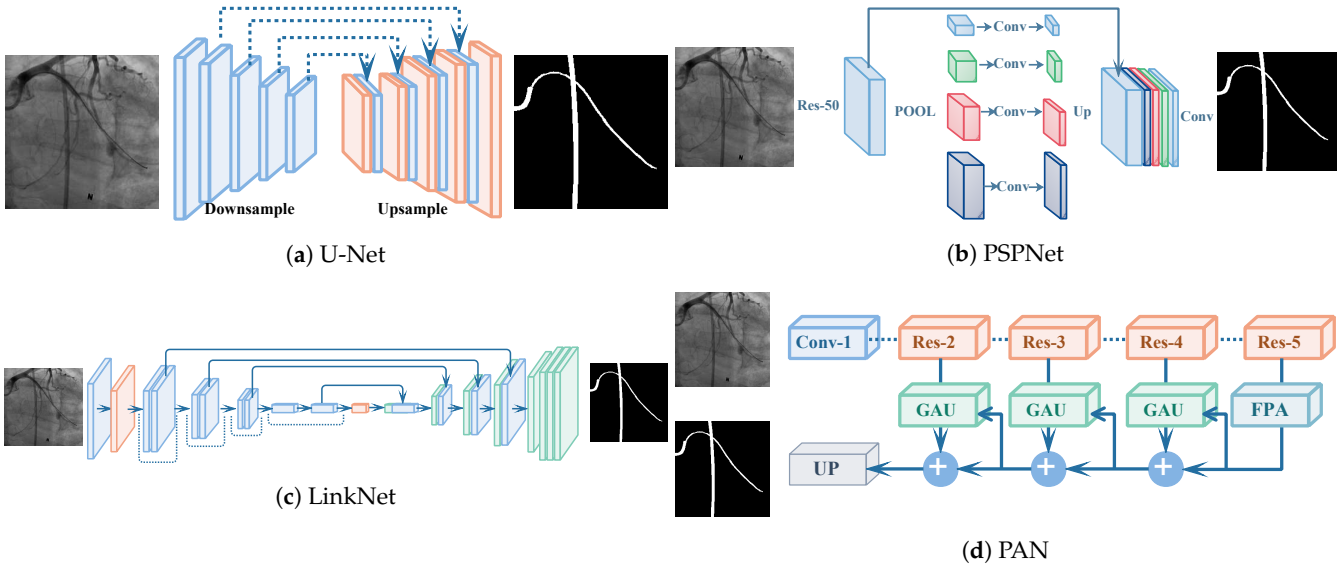
where H_{ij} and W_{ij} represent the size of the feature map θ for a particular layer in the pre-trained ResNet network.

Herein, the final loss $\mathcal{L}$ includes the GAN loss as well as the additional objectives that are computed between the forward and backward cycles. Formally,

$$\mathcal{L} = \mathcal{L}(G, F, D_X, D_Y) + \mathcal{L}_{ssim}(\mathcal{L}_{cyc}G, x) + \mathcal{L}_{ssim}(\mathcal{L}_{cyc}F, y) + \mathcal{L}_{perp}(\mathcal{L}_{cyc}G, x) + \mathcal{L}_{perp}(\mathcal{L}_{cyc}F, y) \quad (6)$$

3.2. Segment: From Synthesis to Segmentation

Figure 2 presents the segmentation networks that are utilized for the catheter segmentation. Given a synthetic X-ray image as input, the segmentation network is used to assign pixels to one of two classes i.e., foreground (catheter) and background (other). In this study, we consider four key architectures that are representative for segmentation tasks, (i) U-net [26], (ii) Linknet [42], (iii) PSPNet [43], and (iv) Pyramid Attention Network (PAN) [44], respectively.



(**a**) U-Net (**b**) PSPNet

(**c**) LinkNet (**d**) PAN

Figure 2. Overview of the segmentation networks employed to train and evaluate the augmentation techniques introduced.

U-net [26] (Figure 2a) is a popular architecture for segmentation tasks, especially in the medical imaging domain. An encoder-decoder network comprised of contracting and expanding paths enables the learning of semantic features across feature levels. The contracting path progressively extracts image representations and increases their dimension layer-by-layer, with decoding path leveraging previous layer information for high level learning. On the other hand, Linknet [42] (Figure 2c) is a U-shape variant, different from U-net in two aspects. First, it substitutes the standard convolutions of U-net with the residual modules. Second, it uses a summation of high and low level features instead of concatenation in the decoder. In this paper, we use a ResNet50 [39] as the encoder for Linknet.

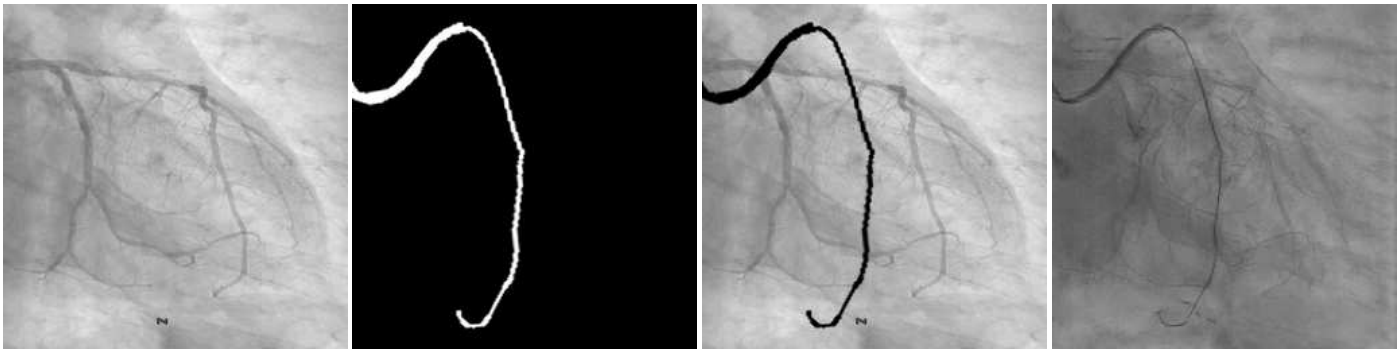
PSPNet [43] (Figure 2b) has a ResNet-50 pre-trained backbone with dilated convolutions along with a pyramid pooling module. Conventional convolutions are substituted by dilated convolutions in the last layers of the pre-trained network, which results in an increase of the receptive field. Pyramid pooling enables the model to capture more global context in a given image with feature maps pooled at different levels and scales. Recently, Li et al. [44] proposed Pyramid Attention Network (PAN) (Figure 2d), it uses spatial pyramid and attention mechanisms to capture dense features for segmentation. Feature Pyramid Attention (FPA) and Global Attention Up-sampling (GAU) modules were proposed to improve high level feature representation via spatial pyramid pooling for global context, while GAU utilizes low-level features later attached to each decoding layer.

4. Experiments

4.1. Datasets

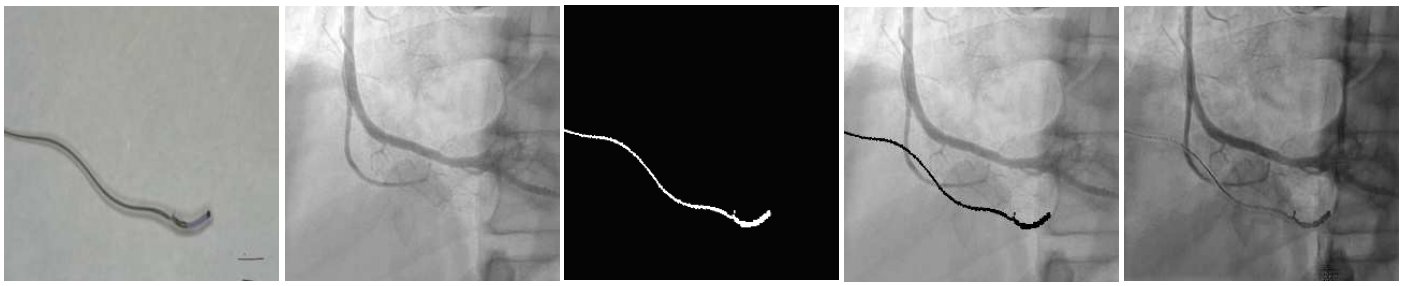
There are currently no publicly available catheter datasets. Thus, we made two datasets for evaluations. First, X-rays dataset consisted of two-dimensional (2D) angiograms with several cranial and caudal views acquired from different patients with heart disease at Seoul National University Hospital, Korea. Figures 3 shows the dataset. For the image generation, X-rays of 100 patients not showing any inserted catheter (Figure 3a)

were employed. Subsequently, we randomly composited masks (Figure 3b) with catheters to create the composited dataset (Figure 3c). Second, we considered the applicability of our method in a more challenging setting i.e., catheters in natural images. Figure 4 shows the samples obtained in our in-house micro-robotic research centre, which consists of 11 videos sequences with a total of 11884 catheter images (Figure 4a) where catheters exhibit varied movements in different directions in response to the magnet field alteration in coils. We composited the camera catheter mask (Figure 4c) obtained using vesseless filter[45] on the X-ray images (Figure 4b) to construct the composited X-ray dataset (Figure 4d). For evaluation, 200 X-ray images with manual annotations were split into train/validation/testing based on the ratios, i.e., 140, 10, and 50 images, respectively.



(**a**) (**b**) (**c**) (**d**)

Figure 3. (**a**) X-ray image without catheter (**b**) random groundtruth (**c**) composited X-ray with groundtruth (**d**) generated X-ray image with catheter.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 4. (**a**) Catheter in natural image (**b**) X-ray image without catheter (**c**) camera catheter mask generated using [45] given **(a)** as input (**d**) composited X-ray with camera catheter mask (**e**) generated X-ray image using camera catheter.

4.2. Experimental Setup

For training and testing CycleGAN, we selected X-ray images with no true catheter, including images with little to no visibly of artifacts i.e., 1000 training and 500 validation, respectively. During training, for any given image; randomly selected catheter masks (obtained offline) were augmented by random flips for composite creation. Adam [46] solver with a batch size of 8 was used with $\lambda = 10$ in the loss objective. The proposed method was trained with a learning rate of $2e - 4$ for 300 epochs. The generators and discriminators were trained alternately, with the discriminator later discarded during inference. The trained generator was later applied to synthesize images of size 256×256 following training on the composite dataset. The evaluation of generative models is often challenging, especially when ground-truth labels are absent. Instead, we qualitatively analyzed the generated images based on visual judgments and domain knowledge.

Furthermore, we present the quantitative results of the segmentation models trained in a supervised setting with synthetically generated images when the catheter mask is present, to confirm that applicability of synthetically generated data for catheter segmentation in real X-ray angiograms. The segmentation models were trained for 250 epochs with an initial learning rate was set to $1e-3$ for 100 epochs and gradually reduced to $1e-5$ for the remainder. The mini-batch size was set as 128 with Adam optimizer [46] used to minimize the Dice loss. The proposed synthesis and segmentation framework was implemented using Keras with a Tensorflow backend on a workstation with NVIDIA Titan XP GPU. The dice metric was used to evaluate segmentation performance; it measures the number of similar pixels divided by the total number of pixels present in both the target and predicted masks. Formally:

$$Dice = \frac{2 \times (A \cap B)}{A + B}, \tag{7}$$

where A is the ground truth and B is the predicated mask. The dice coefficient ranges from 0 to 1, where 1 means complete overlap.

4.3. Quantitative Results

This section provides quantitative analysis of the segmentation models trained with synthetic images. In Table 1, the performance of several state-of-the-art-methods (SOTA) evaluated on 50 real X-ray test images are presented in terms of Dice scores. Among the evaluated methods, Linknet trained with 140 labelled images showed the most improved performance (0.82960). Additionally, Linknet achieved a higher dice score when trained with labeled images and an additional generated images. As for the images generated by our catheter synthesis method, U-Net reports the highest dice score and largely outperforms the rest i.e., PSPNet, PAN, and Linknet.

Table 1. Comparison of segmentation models with labeled and generated images. All of the segmentation models are evaluated on real catheter X-Ray test set.

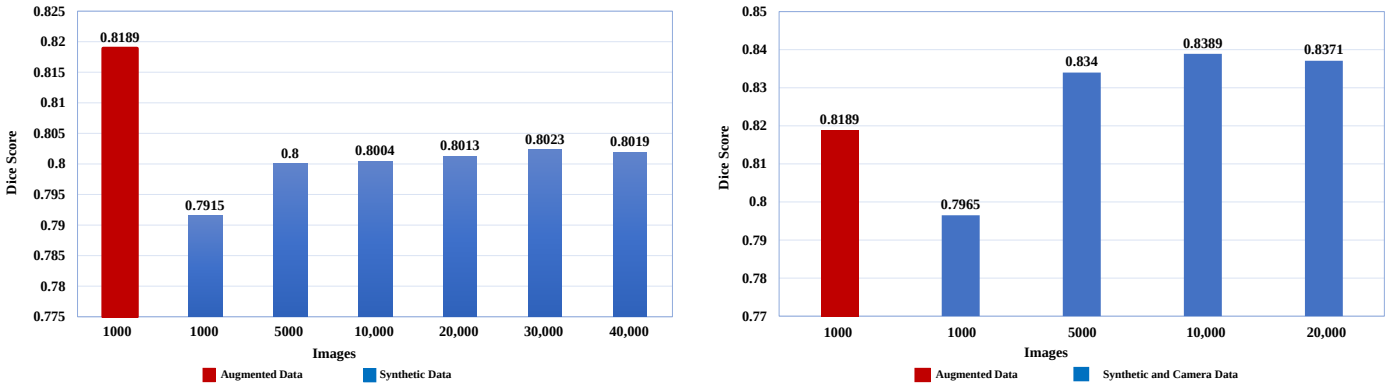
Model	Training Images	Dice Score
UNet [26]	140 Labeled Images (Baseline)	0.8156
	30,000 Synthetic catheter (Generated Images)	0.8023
	140 Labeled Images + 30,000 Synthetic catheter (Generated Images)	0.8595
PSPNet [43]	140 Labeled Images	0.7589
	30,000 Synthetic catheter (Generated Images)	0.7455
	140 Labeled Images + Synthetic catheter (Generated Images)	0.8133
PAN [44]	140 Labeled Images	0.8072
	30,000 Synthetic catheter (Generated Images)	0.7954
	140 Labeled Images + 30,000 Synthetic catheter (Generated Images)	0.8671
Linknet [42]	140 Labeled Images	0.8296
	30,000 Synthetic catheter (Generated Images)	0.8044
	140 Labeled Images + 30,000 Synthetic catheter (Generated Images)	0.8806

Based on Table 2, we noted that U-Net, PAN, and LinkNet report significant performance improvements when trained in mixed settings i.e., both camera and synthetic catheter samples with 140 annotated images, as compared to models trained with synthetic and annotated images only. Notably, camera samples enable the models to learn more complex and thin structures, and consequently result in an overall efficiency improvement when combined.

Table 2. Comparison of segmentation models using camera and synthetic catheters including labeled images. All the models are evaluated on real catheter X-Ray test set.

Model	Training Images	Dice Score
UNet [26]	5000 Camera catheter + 5000 Synthetic catheter	0.8389
	140 labeled images + 5000 Camera catheter + 5000 Synthetic catheter	0.8974
	140 labeled images + 10,000 Synthetic catheter	0.8544
PSPNet [43]	5000 Camera catheter + 5000 Synthetic catheter	0.7220
	140 labeled images + 5000 Camera catheter + 5000 Synthetic catheter	0.8112
	140 labeled images + 10,000 Synthetic catheter	0.8074
PAN [44]	5000 Camera catheter + 5000 Synthetic catheter	0.8210
	140 labeled images + 5000 Camera catheter + 5000 Synthetic catheter	0.8764
	140 labeled images + 10,000 Synthetic catheter	0.8598
Linknet [42]	5000 Camera catheter + 5000 Synthetic catheter)	0.8273
	140 labeled images + 5000 Camera catheter + 5000 Synthetic catheter	0.8894
	140 labeled images + 10,000 Synthetic catheter	0.8797

Figure 5a presents additional experiments with data augmentation included to confirm whether the synthetic images are useful. We selected subsets of the generated images sequentially to train the U-Net model to verify the effect of increasing the number of generated images on performance. The augmented data samples were created using transformations [47], such as scaling (from 0.5 to 1.5 ratio), horizontal flipping, blurring with gaussian filters, elastic deformations with different scaling factors and elasticity coefficients [48], as well as random rotations on each input image. We augmented 140 labeled images to 1000. The U-Net [26] model trained with augmented images achieved a higher dice score compared to the model using 1000 generated images. However, when we increased the number of the generated images to 5000, both models trained using traditional augmentation and synthetic images showed similar performance. Furthermore, even better results are obtained by training the segmentation model with 30,000 generated images. However, the performance saturated when trained with 40,000 images. We assume that saturation in performance occurs because of the error accumulation of inconsistently produced images.



(**a**) (**b**)

Figure 5. Test results of the U-Net model trained on (**a**) synthetic data only and a combination of (**b**) camera catheter and synthetic catheter images.

Additional experiments were performed to further analyze whether camera catheter images indeed enhance performance (see Figure 5b) by considering a setting with equal splits of samples i.e., a 50-50 split for camera and synthetic images generated from the labeled set. Herein, U-Net was trained on the aforementioned split. The best dice score (0.8389) was achieved when 10,000 Images (5000 Camera catheter + 5000 Synthetic catheter) were used to train model compared to the model that was only trained with 10,000 Synthetic catheter (see Figure 5a), the model trained with this split (50-50) achieved better accuracy,

since the camera catheter images enable the model to precisely localize and distinguish the thin part of the catheter in real x-ray images. However, the training images are further increased to 20,000 images (10,000 Camera catheter + 10,000 Synthetic catheter), performance decreased by 0.0018 points.

4.4. Qualitative Results

Synthesis: the qualitative results of our proposed synthesis method are illustrated in Figure 6. For both types of inputs provided to the synthesis model, the generated images show high visual quality overall content of the input image, such as blood vessels, even though no explicit supervision is provided. In comparison to the input images, the images generated by both U-Net and ResNet trained with $\mathcal{L}_{cyc}$ only appear darker, noisy, and introduced artifacts without any noticeable improvement. On the contrary, the proposed method with a perceptual loss showed improved visual quality without a significant loss of information. Moreover, translation results confirm the benefit of including additional losses to maintain visual quality and produce images that may be indistinguishable from real images.

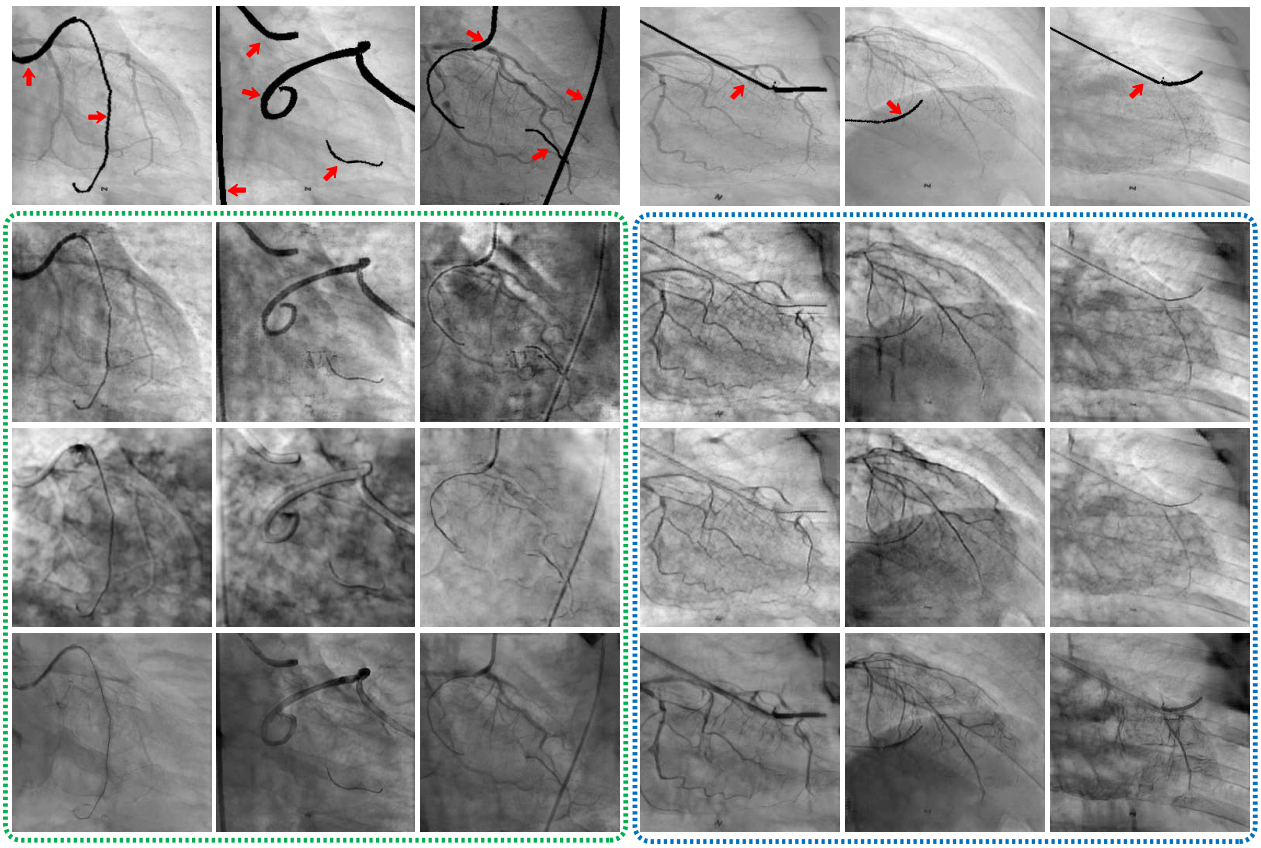


Figure 6. From top to bottom: inputs, U-Net $\mathcal{L}_{cyc}$, ResNet $\mathcal{L}_{cyc}$ and ResNet $\mathcal{L}_{cyc} + \mathcal{L}_{perp}$. (Red arrows) highlight the initial position of the composited catheter in X-ray angiograms and with the generated images using different methods presented in each subsequent rows. Synthetic catheters are represented with green dotted region, whereas the camera catheter is shown in the blue dotted region.

Segmentation: Figure 7 shows the segmentation results on the test set containing the real catheter. Overall, models that trained with synthetically generated images showed better visual segmentation compared to models trained on manually annotated data only. Among the evaluated methods, Linknet [42] showed consistent and improved performance over the rest. However, failure cases were noted in some samples where surgical sutures or stipples were segmented as a part of catheter. Moreover, although PSPNet [43] and PAN [44] showed improvement in removing the surgical sutures, they tend

to over-segment the catheter. We infer that this behavior is mainly due to the lack of samples with surgical sutures in the training data, thus the segmentation models treat the surgical sutures/stipples as a catheters.

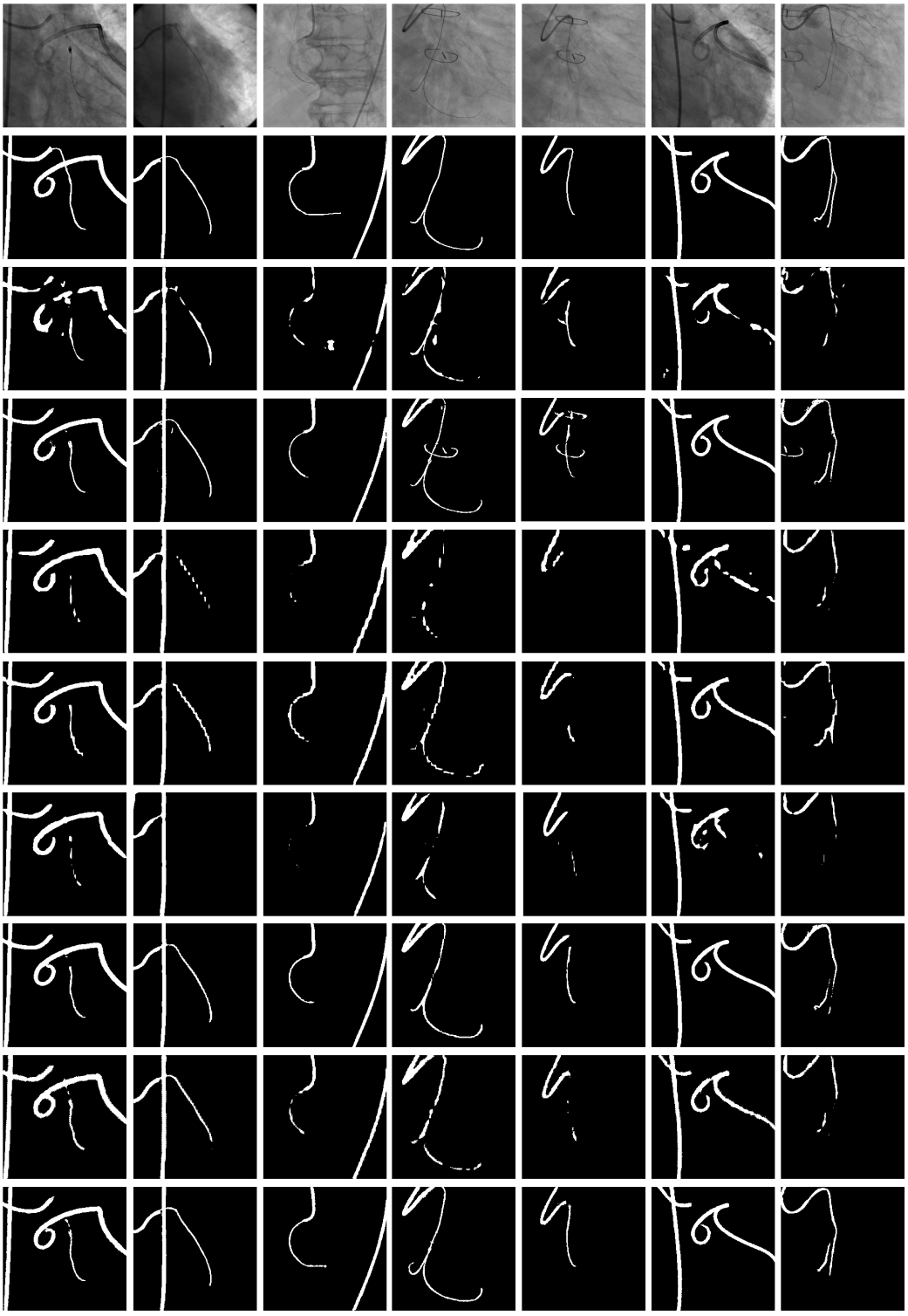


Figure 7. Test image results of a segmentation models trained with 30,000 generated images. From top to bottom: input (1st row), the label images (2nd row), U-Net [26] trained with 140 labeled images (3rd row), U-Net [26] trained with 30,000 generated images (4th row), PSPNet [43] trained with 140 labeled images (5th row), PSPNet [43] trained with 30,000 generated images (6th row),PAN [44] trained with 140 labeled images (7th row), PAN [44] trained with 30,000 generated images (8th row), Linknet [42] trained with 140 labeled images, (9th row), and Linknet [42] trained with 30,000 generated images (10th row).

5. Discussion

Based on our experiments, it is evident that the synthetic catheter images generated by our proposed method maintain visual quality and enable improved segmentation of real catheters compared to models trained on limited annotated data. We observed a significant performance gain of 7.61% when synthetically generated images were used alongside labeled data as compared to applying standard data augmentation techniques. Furthermore, using only synthetic catheter alongside camera catheter samples shows a performance increase of 2.33% and 2% in comparison to the U-Net model(baseline) and U-Net model trained with standard augmented data, respectively.

The baseline U-Net model [26] trained with 140 labeled images failed to accurately identify the catheter due to the limited variability of training samples, as shown in the Figure 8. An initial attempt was made to enable the model to generalize to unseen samples via standard data augmentation. However, this form of augmentation did not yield any potential improvements in segmenting the catheters on the test set. By augmenting the training data with synthetic images, we could improved the baseline results, since generation provides more diversity in the training set. Despite being successful in most cases, the models trained with generated images still failed to segment certain parts of the catheter especially for outliers, such as surgical sutures. To solve this, we further increased the initial training set with 30,000 generated images, leading to improved overall segmentation performance with less failure cases on unseen surgical sutures.

Despite the improved performance of the proposed method, we wish to highlight a few limitations that require careful attention. First, the performance of the proposed synthesis method is difficult to measure, due to the absence of labeled catheter data. Furthermore, the current synthesis method does not consider the segmentation step in the generation of synthetic images; however, we assume that the synthetic image generation can be improved by integrating catheter segmentation into the proposed synthesis method. Although we trained the segmentation models for both catheters and guidewires, a more accurate segmentation could be accomplished by using separate models per input type and later combine the outputs. In future, we plan to address these drawbacks and investigate the use of synthetic data as well as develop an architecture that incorporates both generation and segmentation in an unsupervised manner. Post-processing methods may be used to further boost segmentation and guarantee the robustness of the system when applied to low contrast X-ray fluoroscopic images. Moreover, we would like to compare our proposed approach to more sophisticated GAN based methods.

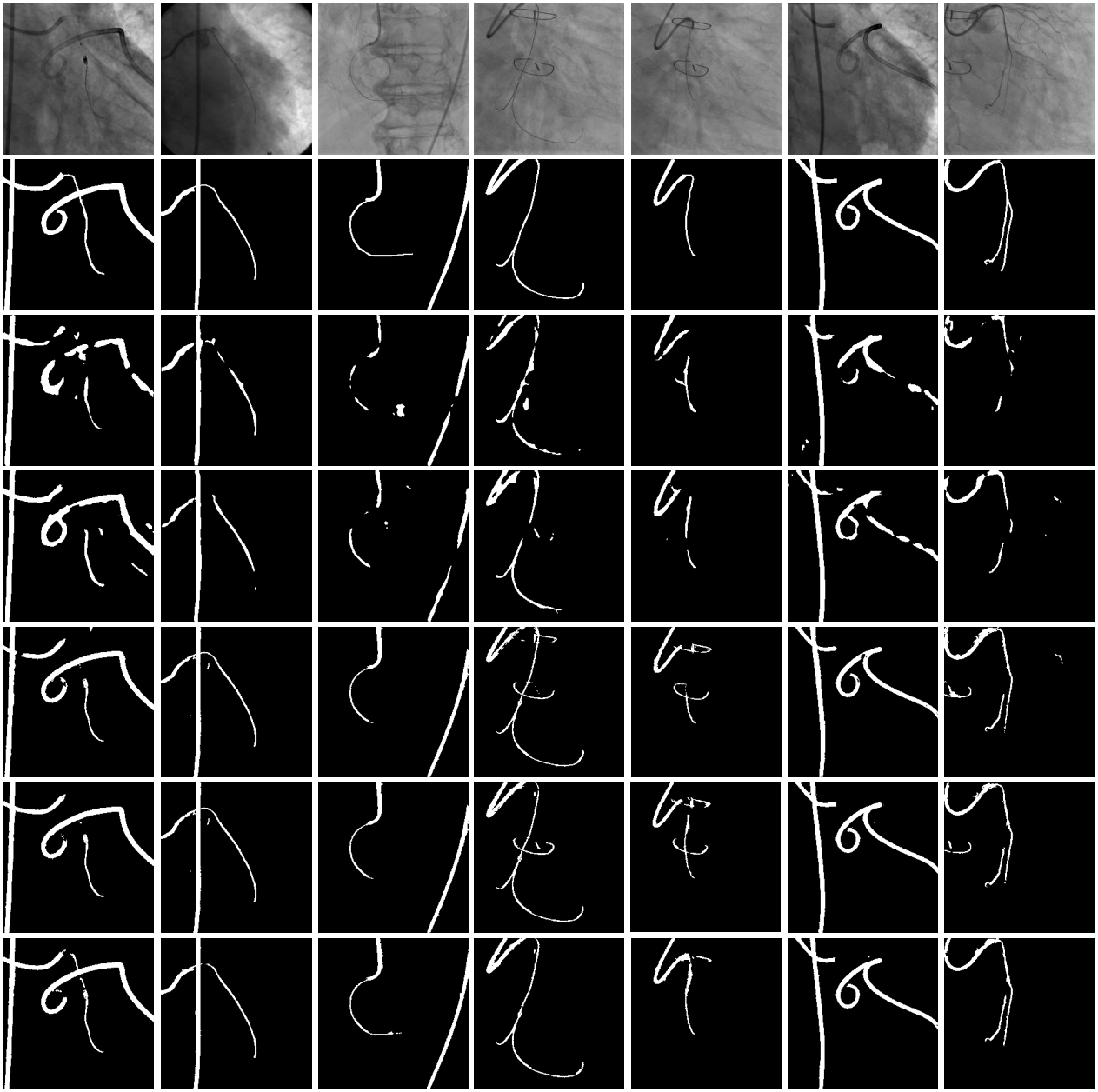


Figure 8. Test image results of a segmentation task using generated images. From top to bottom: input (1st row), the ground truth (2nd row), the baseline UNet [26] segmentation result (3rd row), augmented data segmentation result (4th row), the segmentation result of the model trained with the 5000 generated image (5th row), segmentation result of U-Net model trained with 30,000 genereated image (6th row), and the segmentation result of the model trained with 30,000 generated image plus 140 labeled image (7th row).

6. Conclusions

In this paper, we proposed a catheter synthesis and segmentation framework for X-ray fluoroscopic images under limited data settings. A CycleGAN based framework with a novel loss was introduced i.e., a perceptual loss coupled with similarity constraints to generate a realistic catheter in X-ray fluoroscopic images from composited catheters in X-ray. We evaluate four segmentation models to prove the effectiveness of the synthetically generated data, and found that all the models improved performance over models trained without synthetic data. Moreover, our approach is easily applicable in settings where data is scarce and labeling is expensive.

Author Contributions: I.U. and P.C. contributed equally in conducting the experiments and preparing the draft manuscript. H.C. and C.-H.Y. contributed to data acquisition and evaluation of the developed technique. S.H.P. designed the experimental plan, supervised the work and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Robot industry fusion core technology development project through the Korea Evaluation Institute of Industrial Technology(KEIT) funded by the Ministry of Trade, Industry and Energy of Korea(MOTIE) (NO. 10052980) and the DGIST R&D Program funded by the Ministry of Science and ICT (20-CoE-BT-02).

Institutional Review Board Statement: This study was approved by the Institutional Review Board (IRB) of Seoul National University Bundang Hospital (protocol code B-1902-522-104, date of approval 2019-02-18).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used in the current study are not publicly available because the permission of sharing patient data was not granted by the IRB but are available from the corresponding author on reasonable request.

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

References

1. Kern, M.M.; Gustafson, L.; Kapur, R.; Wasek, S. Angiographic projections made simple: an easy guide to understanding oblique views. *Cath Lab Digest* **2011**, *19*.
2. Zhou, Y.J.; Xie, X.L.; Bian, G.B.; Hou, Z.G.; Wu, Y.D.; Liu, S.Q.; Zhou, X.H.; Wang, J.X. Fully Automatic Dual-Guidewire Segmentation for Coronary Bifurcation Lesion. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–6.
3. Guo, S.; Tang, S.; Zhu, J.; Fan, J.; Ai, D.; Song, H.; Liang, P.; Yang, J. Improved U-Net for Guidewire Tip Segmentation in X-ray Fluoroscopy Images. In Proceedings of the 2019 3rd International Conference on Advances in Image Processing, Chengdu, China, 8–10 November 2019; pp. 55–59.
4. Kao, E.F.; Jaw, T.S.; Li, C.W.; Chou, M.C.; Liu, G.C. Automated detection of endotracheal tubes in paediatric chest radiographs. *Comput. Methods Programs Biomed.* **2015**, *118*, 1–10.
5. Viswanathan, R.R. Image-Based Medical Device Localization. US Patent 7,190,819, 13 March 2007.
6. Uherčík, M.; Kybic, J.; Zhao, Y.; Cachard, C.; Liebgott, H. Line filtering for surgical tool localization in 3D ultrasound images. *Comput. Biol. Med.* **2013**, *43*, 2036–2045.
7. Vandini, A.; Glocker, B.; Hamady, M.; Yang, G.Z. Robust guidewire tracking under large deformations combining segment-like features (SEGlets). *Med Image Anal.* **2017**, *38*, 150–164.
8. Wagner, M.G.; Laeseke, P.; Speidel, M.A. Deep learning based guidewire segmentation in x-ray images. In Proceedings of the Medical Imaging 2019: Physics of Medical Imaging. International Society for Optics and Photonics, San Diego, CA, USA, 16–21 February 2019; Volume 10948, p. 1094844.
9. Subramanian, V.; Wang, H.; Wu, J.T.; Wong, K.C.; Sharma, A.; Syeda-Mahmood, T. Automated Detection and Type Classification of Central Venous Catheters in Chest X-Rays. *arXiv* **2019**, arXiv:1907.01656.
10. Breininger, K.; Würfl, T.; Kurzendorfer, T.; Albarqouni, S.; Pfister, M.; Kowarschik, M.; Navab, N.; Maier, A. Multiple device segmentation for fluoroscopic imaging using multi-task learning. In *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*; Springer: Granada, Spain, 2018; pp. 19–27.
11. Gozes, O.; Greenspan, H. Bone Structures Extraction and Enhancement in Chest Radiographs via CNN Trained on Synthetic Data. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 858–861.
12. Vlontzos, A.; Mikolajczyk, K. Deep segmentation and registration in X-ray angiography video. *arXiv* **2018**, arXiv:1805.06406.
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
14. Tmenova, O.; Martin, R.; Duong, L. CycleGAN for style transfer in X-ray angiography. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 1785–1794.
15. 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.
16. Lee, D.H.; Li, Y.; Shin, B.S. Generalization of intensity distribution of medical images using GANs. *Hum. Centric Comput. Inf. Sci.* **2020**, *10*, 1–15.

17. Ullah, I.; Chikontwe, P.; Park, S.H. Catheter synthesis in X-Ray fluoroscopy with generative adversarial networks. In *International Workshop on PRedictive Intelligence In MEdicine*, Springer: Shenzhen, China, 2019; pp. 125–133.
18. Mercan, C.A.; Celebi, M.S. An approach for chest tube detection in chest radiographs. *IET Image Process.* **2013**, *8*, 122–129.
19. Nguyen, A.; Kundrat, D.; Dagnino, G.; Chi, W.; Abdelaziz, M.E.; Guo, Y.; Ma, Y.; Kwok, T.M.; Riga, C.; Yang, G.Z. End-to-End Real-time Catheter Segmentation with Optical Flow-Guided Warping during Endovascular Intervention. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 9967–9973.
20. Mountney, P.; Maier, A.; Ionasec, R.I.; Boese, J.; Comaniciu, D. Method and System for Obtaining a Sequence of X-ray Images Using a Reduced Dose of Ionizing Radiation. US Patent 9,259,200, 16 February 2016.
21. Wang, L.; Xie, X.L.; Bian, G.B.; Hou, Z.G.; Cheng, X.R.; Prasong, P. Guide-wire detection using region proposal network for X-ray image-guided navigation. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3169–3175.
22. Ullah, I.; Chikontwe, P.; Park, S.H. Real-time tracking of guidewire robot tips using deep convolutional neural networks on successive localized frames. *IEEE Access* **2019**, *7*, 159743–159753.
23. Lee, H.; Mansouri, M.; Tajmir, S.; Lev, M.H.; Do, S. A deep-learning system for fully-automated peripherally inserted central catheter (PICC) tip detection. *J. Digit. Imaging* **2018**, *31*, 393–402.
24. Chen, S.; Wang, S. Deep learning based non-rigid device tracking in ultrasound image. In Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, Shenzhen, China, 8–10 December 2018; pp. 354–358.
25. Ambrosini, P.; Ruijters, D.; Niessen, W.J.; Moelker, A.; van Walsum, T. Fully automatic and real-time catheter segmentation in X-ray fluoroscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Quebec City, QC, Canada, 2017; pp. 577–585.
26. 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: Munich, Germany, 2015; pp. 234–241.
27. Wu, Y.D.; Xie, X.L.; Bian, G.B.; Hou, Z.G.; Cheng, X.R.; Chen, S.; Liu, S.Q.; Wang, Q.L. Automatic guidewire tip segmentation in 2D X-ray fluoroscopy using convolution neural networks. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.
28. Breininger, K.; Albarqouni, S.; Kurzendorfer, T.; Pfister, M.; Kowarschik, M.; Maier, A. Intraoperative stent segmentation in X-ray fluoroscopy for endovascular aortic repair. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 1221–1231.
29. Kooi, T.; Litjens, G.; Van Ginneken, B.; Gubern-Mérida, A.; Sánchez, C.I.; Mann, R.; den Heeten, A.; Karssemeijer, N. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal.* **2017**, *35*, 303–312.
30. Zaman, A.; Park, S.H.; Bang, H.; Park, C.; Park, I.; Joung, S. Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 931–941.
31. 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.
32. Wolterink, J.M.; Dinkla, A.M.; Savenije, M.H.; Seevinck, P.R.; van den Berg, C.A.; Išgum, I. Deep MR to CT synthesis using unpaired data. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Québec City, QC, Canada, 2017; pp. 14–23.
33. Dar, S.U.; Yurt, M.; Karacan, L.; Erdem, A.; Erdem, E.; Çukur, T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans. Med Imaging* **2019**, *38*, 2375–2388.
34. Hiasa, Y.; Otake, Y.; Takao, M.; Matsuoka, T.; Takashima, K.; Carass, A.; Prince, J.L.; Sugano, N.; Sato, Y. Cross-modality image synthesis from unpaired data using CycleGAN. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Granada, Spain, 2018; pp. 31–41.
35. Chartsias, A.; Joyce, T.; Dharmakumar, R.; Tsaftaris, S.A. Adversarial image synthesis for unpaired multi-modal cardiac data. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Québec City, QC, Canada, 2017; pp. 3–13.
36. Gherardini, M.; Mazomenos, E.; Menciassi, A.; Stoyanov, D. Catheter segmentation in X-ray fluoroscopy using synthetic data and transfer learning with light U-nets. In *Computer Methods and Programs in Biomedicine*; Elsevier: Amsterdam, The Netherlands, 2020; p. 105420.
37. Yi, X.; Adams, S.; Babyn, P.; Elnajmi, A. Automatic Catheter and Tube Detection in Pediatric X-ray Images Using a Scale-Recurrent Network and Synthetic Data. *J. Digit. Imaging* **2019**, *33*, 181–190.
38. Frid-Adar, M.; Amer, R.; Greenspan, H. Endotracheal Tube Detection and Segmentation in Chest Radiographs using Synthetic Data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Shenzhen, China, 2019; pp. 784–792.
39. 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, 27–30 June 2016; pp. 770–778.
40. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57.
41. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*; Springer: Amsterdam, The Netherlands, 2016; pp. 694–711.
42. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.

43. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
44. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid attention network for semantic segmentation. *arXiv* **2018**, arXiv:1805.10180.
45. Frangi, A.F.; Niessen, W.J.; Vincken, K.L.; Viergever, M.A. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin, Germany, 1998; pp. 130–137.
46. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
47. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: fast and flexible image augmentations. *Information* **2020**, *11*, 125.
48. Simard, P.Y.; Steinkraus, D.; Platt, J.C.; others. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In Proceedings of the ICDAR, Edinburgh, UK, 3–6 August 2003; Volume 3.