sensors

MDPI

*Article*

# Local Style Preservation in Improved GAN-Driven Synthetic Image Generation for Endoscopic Tool Segmentation

**Yun-Hsuan Su [1,*], Wenfan Jiang [1], Digesh Chitrakar [2], Kevin Huang [2], Haonan Peng [3] and Blake Hannaford [3]**

[1] Department of Computer Science, Mount Holyoke College, 50 College Street, South Hadley, MA 01075, USA; jiang24w@mtholyoke.edu
[2] Department of Engineering, Trinity College, 300 Summit St., Hartford, CT 06106, USA; digesh.chitrakar@trincoll.edu (D.C.); kevin.huang@trincoll.edu (K.H.)
[3] Department of Electrical and Computer Engineering, University of Washington, 185 Stevens Way, Paul Allen Center, Seattle, WA 98105, USA; penghn@uw.edu (H.P.); blake@uw.edu (B.H.)
* Correspondence: msu@mtholyoke.edu; Tel.: +1-413-538-3468

**Abstract:** Accurate semantic image segmentation from medical imaging can enable intelligent vision-based assistance in robot-assisted minimally invasive surgery. The human body and surgical procedures are highly dynamic. While machine-vision presents a promising approach, sufficiently large training image sets for robust performance are either costly or unavailable. This work examines three novel generative adversarial network (GAN) methods of providing usable synthetic tool images using only surgical background images and a few real tool images. The best of these three novel approaches generates realistic tool textures while preserving local background content by incorporating both a style preservation and a content loss component into the proposed multi-level loss function. The approach is quantitatively evaluated, and results suggest that the synthetically generated training tool images enhance UNet tool segmentation performance. More specifically, with a random set of 100 cadaver and live endoscopic images from the University of Washington Sinus Dataset, the UNet trained with synthetically generated images using the presented method resulted in 35.7% and 30.6% improvement over using purely real images in mean Dice coefficient and Intersection over Union scores, respectively. This study is promising towards the use of more widely available and routine screening endoscopy to preoperatively generate synthetic training tool images for intraoperative UNet tool segmentation.

**Keywords:** robot-assisted minimally invasive surgery; surgical tool segmentation; generative adversarial networks; UNet; medical imaging



**Citation:** Su, Y.-H.; Jiang, W.; Chitrakar, D.; Huang, K.; Peng, H.; Hannaford, B. Local Style Preservation in Improved GAN-Driven Synthetic Image Generation for Endoscopic Tool Segmentation. *Sensors* **2021**, *21*, 5163. https://doi.org/10.3390/s21155163

Academic Editors: Tamás Haidegger and Axel Krieger

Received: 1 July 2021
Accepted: 27 July 2021
Published: 30 July 2021



## 1. Introduction

Computer vision and machine learning have experienced rapid development and growth in the last decade. While applications in the medical imaging field are growing, challenges still exist. This manuscript focuses on the use of the UNet, the most widely adopted image segmentation tool for medical imaging. In the context of robot-assisted minimally invasive procedures, accurate surgical tool segmentation is a key component of numerous computer-assisted interventions [1], and may enable robust reconstruction [2] potentially from multiple simultaneous viewpoints [3,4]. Due to the challenges of dynamic deformation, specular reflections, and partial blurriness [5], accurate tool segmentation often requires large surgical image data sets to achieve desirable segmentation performance through data-driven approaches. Such data sets are difficult to acquire due to lack of expert annotation, under-representation of rare conditions, and poor standardization. Furthermore, large surgical image data sets are expensive and oftentimes impractical to acquire from clinical robot-assisted minimally invasive surgeries [6]. Concerns include potential interruption of operation workflow as well as sterilization and data privacy concerns.

### 1.1. UNet in Medical Image Segmentation

The primary purpose of medical image segmentation tasks is to separate different objects or anatomical structures of interest from the rest of the image. These structures often need to be isolated for proper diagnosis of conditions [7] or to remove occluding elements. The use of image segmentation tools, and UNet in particular, is most prominent in the medical imaging field for cardiovascular and brain systems. These anatomical structures are oftentimes imaged using 3D imaging methods, such as with computed tomography (CT) and magnetic resonance imaging (MRI), and thus the UNet has been adapted to a variety of medical imaging modalities.

Brain tumor imaging is most prominently achieved with MRI, and identifying the boundaries of cancerous and healthy tissue is necessary for proper resection of the diseased tissue. Several implementations of UNet have been developed and implemented to successfully segment brain tumor structures in MRI [8–11]. Similarly, UNet has been used in neural MRI to identify and segment brain lesions [12,13] and to analyze brain development [14,15]. Three-dimensional imaging of the cardiovascular system exhibits a broader range of imaging modalities, including CT and MRI. Lung and pulmonary structures were segmented using the UNet on CT scans [9,16–18] and cardiovascular structures with MRI [19–22]. Three-dimensional UNet segmentation has also been used for segmentation of liver tumors in CT scans [23,24] and MRI [25], prostate and breast cancer in MRI [26,27], multi-organ segmentation from CT [28–30], and osteosarcoma [31] and vertebrae [32] from CT. The application of UNet for 2D medical image segmentation covers a range of tasks, including skin lesion segmentation [33], segmentation in microscopy [16,34–36], and retinal imaging [37–40] to name a few. Endoscopic imaging is the modality of interest for 2D image segmentation, with several uses of UNet for these types of images [41,42].

### 1.2. Endoscopic Surgical Tool Segmentation

In robot-assisted minimally invasive surgery, the endoscope is typically the primary imaging modality, and in itself provides a restricted field of view [43]. Understanding the surgical tool location within the endoscope frame and with respect to the surgical anatomy can be of vital importance to provide intelligent, computer-aided assistance or intervention. Tool proximity and localization with respect to anatomy can, for example, inform operation procedure or even be used to isolate imaging of the anatomy for image registration and reconstruction. Several auxiliary sensing methods can be used for tool pose estimation, including robot kinematic information [44] and electromagnetic tracking [45]. However, endoscopic and image-based methods coincide with the operator's frame of reference and do not require augmentations to the surgical tools or instruments. With sophisticated developments in deep learning, the use of machine vision is an attractive avenue for tool tracking and segmentation. However, the lack of sufficient and established endoscopic image data and standard evaluation or ranking are challenges. Within the available datasets, imaging may be sourced from simulation environments [46], a phantom [47], ex vivo [48,49], or in vivo [50]. Furthermore, the image resolutions, surgical operation type, task conditions (e.g., lighting, smoke, occlusions, and blur), image set size, and tool labeling (e.g., tool tip or bounding box) vary between datasets. Several natural image features can be used for tool detection and segmentation. Image color can provide a natural feature for discrimination [51–53], and the image gradient, textures and shapes can also be used [54–56]. Early approaches used these features to assist in tool detection and segmentation using support vector machines [54,57] and decision forests [50]. More recently, neural network and UNet-based methods have emerged as promising directions [58–63].

#### 1.2.1. Surgical Image Augmentation

Lacking sufficient numbers of real data has been addressed in conventional vision applications with various synthetic image generation approaches [64–66]. Unfortunately, most use simple morphological augmentations [67] unsuitable for surgical images, which are rich with the complex and diverse features of real human tissue [68]. The problem re-

mains an open challenge, and several approaches exist in the literature. Surgical simulators, such as the 3D-slicer [69], the RobotiX mentor [70], the dV-Trainer [71], and the AMBF [72], enable users to readily capture large quantities of synthetic training images. However, because these images depict a purely artificial scene, they often lack the visual artifacts and imperfections required to train strong tool segmentation models [73,74]. It is possible to convert the working domain to a synthetic one by training the segmentation network on a large set of synthetic images and real test images converted to synthetic through pre-processing domain transfer [75]. The results are robust, however, the real-to-synthetic domain transfer loses textual cues and details. A similar concept was a adopted with a more readily available target image domain-cadaver images [76]. Labeling was expedited for the cadaver endoscopic imaging by using robot kinematic information [6,58,77,78]. Although both image domains contain some realistic visual details and texture, cadaver data acquisition is expensive.

#### 1.2.2. Rendering via Adversarial Learning

Generative adversarial networks (GANs) [79] have gained traction in the medical imaging field for data generation without explicit need for probability density functions or labeled samples [80]. Compared with traditional training image augmentation methods like scaling, rotation, flipping, and elastic deformation [81], GAN-driven approaches afford the capability to enforce domain specific conditions on the generated images to abide by the surgical workflow, appearance of a particular pathology, or various imaging protocols [82,83]. Conditional GAN-based medical image synthesis research has been explored in numerous medical imaging domains such as CT [84], MRI [85–87], Ultrasound [88,89], X-rays [90,91], and retinal fundus imaging [92]. This is useful to synthesize images in uncommon conditions, such as lung nodules along the lung border [93]. However, little work has been done in the endoscopic imaging modality, and even less for surgical tool segmentation. In [63], an image-to-image (I2I) model and robot simulator transferred the realistic style of ex vivo and in vivo RMIS images onto simulated tool images. However, since surgical tool pixels were processed independently of the background, visual effects from reflection, motion blur and tool-tissue interactions were not well modeled. The absence of realistic visual artifacts in the generated images can be addressed by collecting a large number of cadaver images in a similar mock sinus endoscopic surgery setup and conducting cadaver-to-real cross domain image synthesis. This ensures that similar visual effects exist in both the source (cadaver) and target (real) domain [76]. Although realistic images are generated, a generic solution for synthetic endoscopic image generation with pure real dataset remains an open challenge. In another approach, the image synthesis step was bypassed and surgical tool segmentation was implemented directly using GAN-based domain adaptation [94]. The surgical image was the real image domain and the segmented mask was the target domain.

### *1.3. Contributions*

This work investigated three novel GAN-driven surgical image augmentation approaches. The best method utilizes the proposed loss function that incorporates both local neural style transfer [95] and a modified CycleGAN [96]-like structure with custom component-level losses. The method was evaluated on a classic tool segmentation model, UNet, with varying levels of synthetic training data composition. To the best of the authors' knowledge, this work is the first to provide simultaneously:

1. A GAN-driven synthetic surgical endoscopic image generation framework without requiring cross-domain sample images;
2. The development of a custom multi-level loss function that:
    - On the component level, adopts realistic tool textural style, minimizes background content changes and preserves synthetic tool shape;
    - On the image level, incorporates visual artifacts to mimic realistic tool-tissue interaction regions;

3. A systematic guide to evaluate generated synthetic images and identify the ideal composition of real and synthetic training images;
4. Open access of all source code [97,98].

## 2. Methods

Three modified CycleGAN approaches for generating synthetic tool images for endoscopic image segmentation were investigated. All three strategies are improvements upon the baseline tool augmentation method described in Section 2.1. The set of baseline synthetic images $\mathcal{S}$ is generated by overlaying real endoscopic surgical backgrounds, the collection of which is denoted $\mathcal{R}_\mathcal{B}$, with a randomly placed artificial surgical tool. The task then is to enhance the baseline artificial tool pixels with realistic appearance. The first strategy performs transfer from the domain of baseline synthetic images $\mathbb{S}$ to the domain of real surgical images $\mathbb{R}$ through CycleGAN (note: $\mathbb{R}$ does not refer to the set of real numbers in this context). The second strategy executes partial GAN application on only the baseline surgical tool pixels. Finally, the third approach utilizes a modified CycleGAN loss design that balances (a) partial style preservation of the background and (b) realistic generation of the tool (visual artifacts and texture) and (c) tool tissue border smoothness of the generated synthetic image.

### *2.1. Baseline Synthetic Image Generation, S*

A baseline synthetic image $s_i \in \mathcal{S} \subset \mathbb{S}$ is constructed using two main steps on a preselected real endoscopic background image in $\mathcal{R}_\mathcal{B}$. These steps are

1. Surgical tool augmentation;
2. Circle border pre-processing.

#### 2.1.1. Surgical Tool Augmentation

As depicted in Figure 1a, the synthetic tool shape is defined by 5 key geometric points, which are mostly connected by straight lines. Exceptions exist between key points 2, 3 and 4, which are connected via a 2nd order polynomial. The tools are randomly scaled, shifted and rotated before being placed on each background image. The tool colors $I_T$ were rendered from the fusion of a metallic texture image $I_M$ and a normalized reflection background image $I_{NR}$.

$$\begin{aligned} I_T &= I_M * I_{NR} \\ I_{NR} &= \alpha \frac{I_R}{255} + (1 - \alpha) \end{aligned} \tag{1}$$

where $*$ indicates elemental-wise multiplication, $\alpha$ is an empirically chosen reflection coefficient, and $I_R$ represents the reflection background image. A larger $\alpha$ can result in a stronger reflection effect. Once the color is rendered, the tool is overlaid onto the surgical backgrounds in $\mathcal{R}_\mathcal{B}$ with added modifications such as glare and shading to enhance realism as shown in Figure 1b.

#### 2.1.2. Circular Border Pre-Processing

The endoscopic background images in $\mathcal{R}_\mathcal{B}$ used to generate $\mathcal{S}$ images are selected from the University of Washington Sinus dataset [49]. Since endoscopes exhibit a circular field of view, each original rectangular background image contains endoscopic information with a circular border contained within the entire image. Between varying endoscopes, the circular borders are inconsistent in size and sometimes off-center. Without first isolating only relevant image sections, unwanted overfit of the tool location based on circle location may occur. To that end, proceeding the surgical tool augmentation step, only the largest square within the circular image is stored as a synthetic baseline image in $\mathcal{S}$.

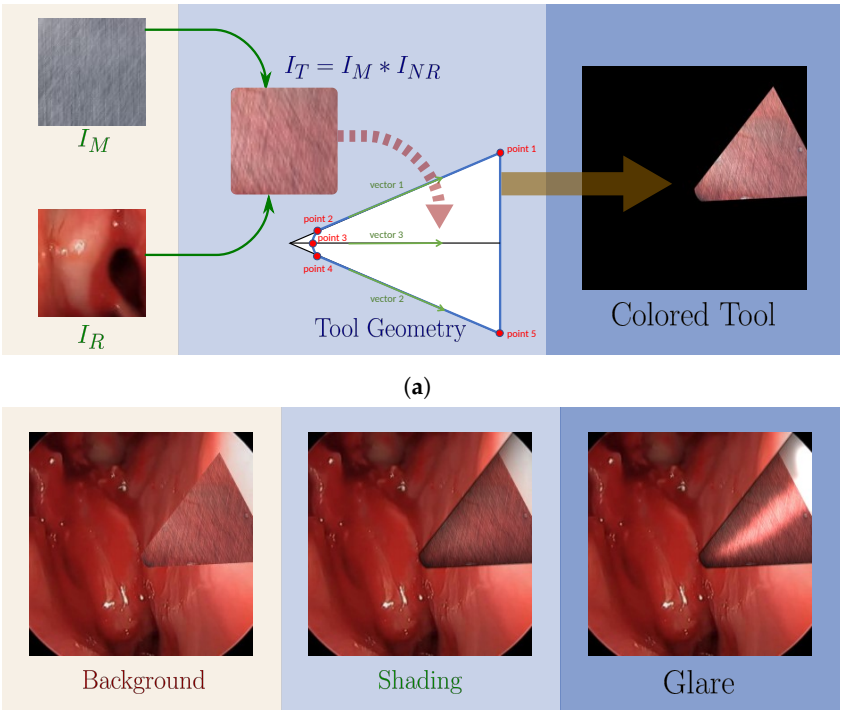



**Figure 1.** The baseline tool image generation procedure (**a**) tool color ($I_M * I_{NR}$) and tool geometry fused to create colored tool patch. (**b**) Post-processing steps by overlaying tool patch on background, shading, and glare modifications.

### *2.2. GAN-Driven Augmentations*

The baseline synthetic tool images as depicted in Figure 1 lack sufficient realism. The automated glare and shade additions are incapable of deceiving the human eye. The information used from the surgical background is applied in an inflexible and non-adaptive manner. To improve synthetic tool realism, three modified implementations of the CycleGAN network were developed, which execute transfer between the domain of real surgical images $\mathbb{R}$ and the domain of synthetic baseline tool images $\mathbb{S}$.

#### 2.2.1. Naive Global GAN Application (***Strategy I***)

The CycleGAN network consists of four networks: two generators and two discriminators. Given synthetic images $\{s \in \mathcal{S} \subset \mathbb{S}\}$ and real surgical scenes $\{r \in \mathcal{R} \subset \mathbb{R}\}$ with training domains $\mathbb{S}$ and $\mathbb{R}$, the generators $\mathcal{G} : \mathbb{R} \rightarrow \mathbb{S}$ and $\mathcal{F} : \mathbb{S} \rightarrow \mathbb{R}$ seek a bijective mapping. Furthermore, they ideally are inverses. The discriminators $D_S$ and $D_R$ serve the role of two inspectors, where $D_S$ and $D_R$ evaluate the likelihood that an image belongs to $\mathbb{S}$ and $\mathbb{R}$, respectively.

The four networks are trained sequentially with pairs of images $(s_i, r_i)$, where $s_i \in \mathcal{S}$ and $r_i \in \mathcal{R}$. During training, four loss functions are used for optimization, one for each network. First define the following expressions

$$f_{\text{cyc}}(I_x, I_y) = \frac{\sum |p_{xi} - p_{yi}|}{N_x} \quad (2)$$

$$f_{\text{gen}}(x) = (x-1)^2 \quad (3)$$

$$f_{\text{dis}}(x, y) = (x-1)^2 + \frac{y^2}{2} \quad (4)$$

where $I_x$, $I_y$ are images, $N_x$ is the number of pixels in $I_x$, $p_{xi}$ and $p_{yi}$ are the $i$th pixels in $I_x$, $I_y$, respectively, and $x, y$ are real numbers. For each training image pair $(s_i, r_i)$, loss functions for each network are computed. First define

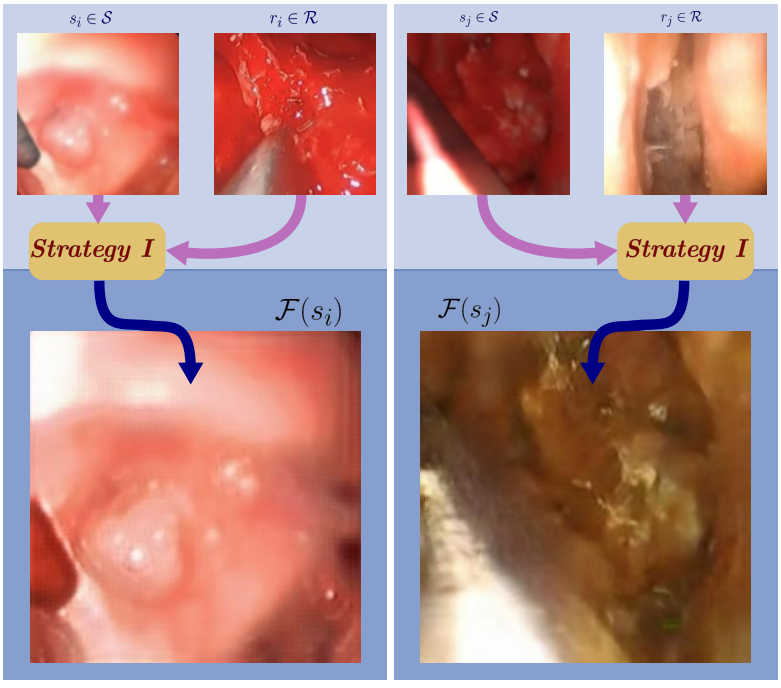
$$L_{\text{cyc}} = \lambda_1 f_{\text{cyc}}\big(r_i, (\mathcal{F} \circ \mathcal{G})(r_i)\big) + \lambda_2 f_{\text{cyc}}\big(s_i, (\mathcal{G} \circ \mathcal{F})(s_i)\big) \tag{5}$$

where $\circ$ is the composition operator, $\lambda_1, \lambda_2$ are heuristically tuned weights. Then the four loss functions for each network are computed as

$$\begin{aligned} L_{\mathcal{G}} &= L_{\text{cyc}} + f_{\text{gen}}\big((D_s \circ \mathcal{G})(r_i)\big) && (6)\\ L_{\mathcal{F}} &= L_{\text{cyc}} + f_{\text{gen}}\big((D_r \circ \mathcal{F})(s_i)\big) && (7)\\ L_{D_s} &= f_{\text{dis}}\Big(D_s(s_i), (D_s \circ \mathcal{G})(r_i)\Big) && (8)\\ L_{D_r} &= f_{\text{dis}}\Big(D_r(r_i), (D_r \circ \mathcal{F})(s_i)\Big) && (9) \end{aligned}$$

where $L_{\mathcal{G}}, L_{\mathcal{F}}$ are cycle consistency losses associated with $\mathcal{G}, \mathcal{F}$, respectively, and $L_{D_s}, L_{D_r}$ are adversarial losses for $D_s, D_r$, respectively. Adversarial losses characterize the deviation between the distribution of the generated data and that of the original data. On the other hand, the cycle consistency loss ensures that the network has the flexibility to map a set of data to all possible permutations in the target domain [96].

When training the CycleGAN network, images from each domain were taken in pairs. Figure 2 depicts two example input image pairs ($s_i \in \mathcal{S} \subset \mathbb{S}$, $r_i \in \mathcal{R} \subset \mathbb{R}$), and the resultant synthetic image generated by $\mathcal{F}(s_i)$ after the model is fully trained. Because the approach lacks semantic knowledge of the image, tool pixels and background pixel attributes were often interchanged. In Figure 2a, the final synthetic tool image tool pixels in $\mathcal{F}(s_i)$ inherit tissue color tones, and in Figure 2b the final synthetic tool image background pixels adopt tool color tones. This strategy's drawbacks made it unacceptable.




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

**Figure 2.** CycleGAN generated synthetic images from ***Strategy I***: (**a**) The generated tool image adopts texture from the background; (**b**) the background adopts texture from the tool.

#### 2.2.2. Tool Localized GAN (***Strategy II***)

To address the mismatching of color compositions for different semantic elements within the images, the CycleGAN approach was modified to first execute only on isolated synthetic and real tool pixels. For each $s_i \in \mathcal{S}$, there is an associated binary tool mask ${}_s m_i$. The isolated tool image from $s_i \in \mathcal{S}$, call it $s_{t_i}$, is then calculated as

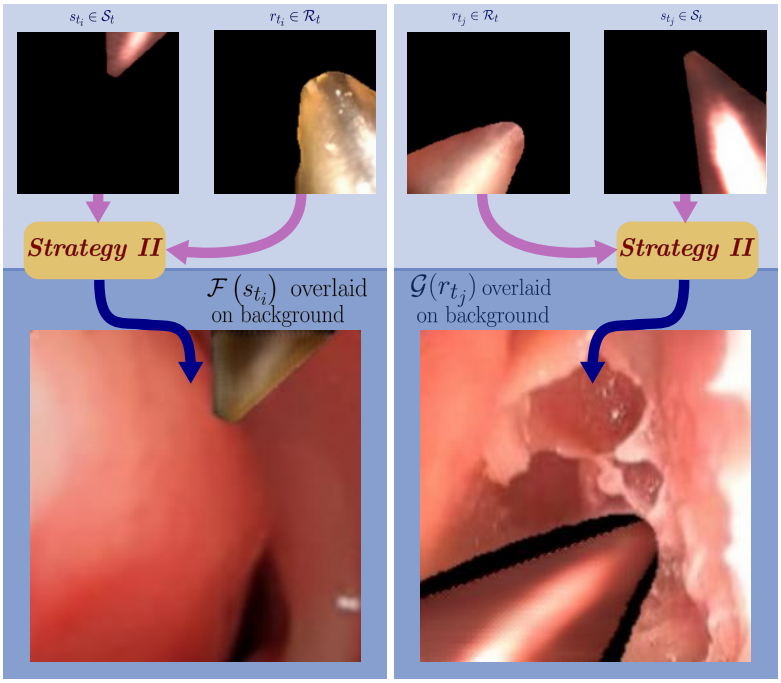
$$s_{t_i} = s_i * {}_s m_i \tag{10}$$

where $*$ denotes pixel-wise multiplication. Isolated tool image for $r_i \in \mathcal{R}$ are similarly defined as

$$r_{t_i} = r_i * {}_r m_i \tag{11}$$

Let $\mathcal{S}_t = \{s_{t_i} | s_i \in \mathcal{S}\}$ and $\mathcal{R}_t = \{r_{t_i} | r_i \in \mathcal{R}\}$. The CycleGAN algorithm as described in the previous subsection was then implemented again on the entire images with separately enhanced tool pixels overlaid on background images from $\mathcal{R}_\mathcal{B}$.

As depicted in Figure 3, the separately enhanced tool pixels are non-ideal. In Figure 3a, lack of context from the background during isolated tool image training result in tool pixels that do not reflect the surgical scene colors. Furthermore, since morphology of synthetic and real tools vary, when isolated with a black background tool borders are not well incorporated. This is observed in Figure 3b. Because of these faults, this strategy was also deemed unacceptable.




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

**Figure 3.** CycleGAN generated synthetic images from ***Strategy II***: (**a**) The tool image was generated ignorant of the background; (**b**) the tool image borders are not preserved.

#### 2.2.3. GAN with Partial Style Preservation (***Strategy III***)

The results of the previous two methods demonstrated an instability in the texture, color or shape of the artificial surgical tool. To address these issues, methods for partial style preservation of tool pixels and content preservation of tissue pixels were incorporated while continuing to train the entire image through CycleGAN. This approach aims to minimize textural disparity between generated and real tool pixels while preserving background surgical scene content, processing each of the two semantic portions separately.

Style differences between the generated and real tool pixels are minimized within each activation layer.

For content preservation, first define

$$f_{\text{con}}(I_x, I_y, I_z) = \frac{\sum(1 - p_{zi}) * |p_{xi} - p_{yi}|}{N_x} \tag{12}$$

where $I_x, I_y$ are images and $I_z$ is a binary labeled mask for $I_x$, $p_{xi}, p_{yi}, p_{zi}$ are the $i$th pixels in $I_x, I_y, I_z$, and $*$ denotes element-wise multiplication.

Then for each of the two generators a content loss is assigned as

$$L_{\mathcal{G}_{\text{con}}} = f_{\text{con}}\left(r_i, \mathcal{G}(r_i), {}_r m_i\right) \tag{13}$$

$$L_{\mathcal{F}_{\text{con}}} = f_{\text{con}}\left(s_i, \mathcal{F}(s_i), {}_s m_i\right) \tag{14}$$

To formulate style preservation, let

$$f_{\text{vgg}}(I_x, I_y) = G\left(V(I_x) * V(I_y)\right) \tag{15}$$

$$f_{\text{sty}}({}_x G, {}_y G) = \sum_{l<=L} \frac{\omega_l({}_x G_l - {}_y G_l) * ({}_x G_l - {}_y G_l)}{4N_{xGl}^2} \tag{16}$$

where $*$ again indicates element-wise multiplication, ${}_x G, {}_y G$ denote the Gramian of images $I_x, I_y$, respectively, $l$ iterates through layers, $N_{xGl}$ is the number of elements in $xGl$, $L$ is the number of style layers, the weighting factor for each layer $\omega_l = \frac{1}{|L|}$ in this work, and $V$ returns pretrained VGG19 neural network per-layer output.

Recall that CycleGAN is trained sequentially through pairs of images, $(s_i, r_i)$. Each pair is associated with a pair of binary tool masks $({}_s m_i, {}_r m_i)$. Then, for each of the two generators, a style loss is assigned as

$$L_{\mathcal{G}_{\text{sty}}} = f_{\text{sty}}\Big(f_{\text{vgg}}\left(\mathcal{G}(r_i), {}_r m_i\right), f_{\text{vgg}}\left(s_i, {}_s m_i\right)\Big) \tag{17}$$

$$L_{\mathcal{F}_{\text{sty}}} = f_{\text{sty}}\Big(f_{\text{vgg}}\left(\mathcal{F}(s_i), {}_s m_i\right), f_{\text{vgg}}\left(r_i, {}_r m_i\right)\Big) \tag{18}$$

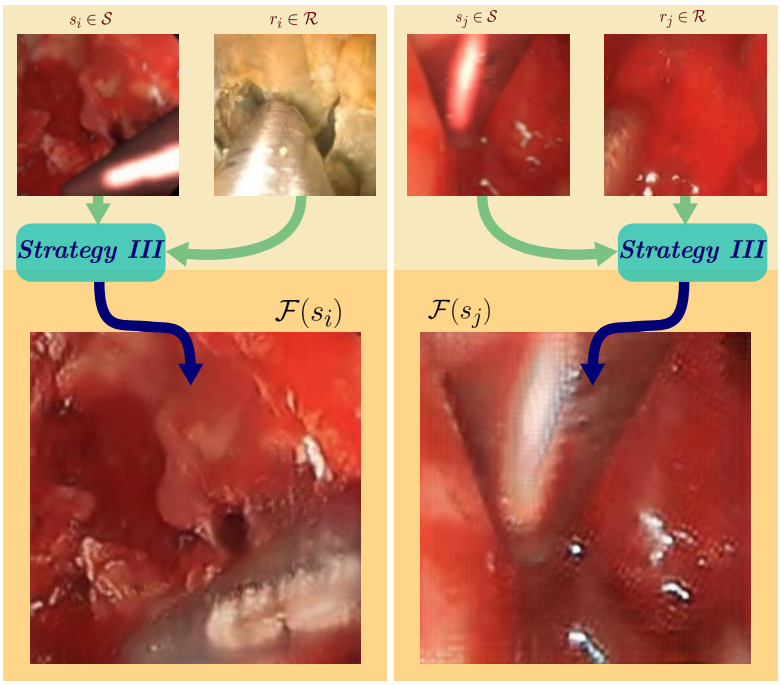
With these parameters defined, the cycle consistency loss functions for training generators $\mathcal{G} : \mathbb{R} \to \mathbb{S}$ and $\mathcal{F} : \mathbb{S} \to \mathbb{R}$ are modified by augmenting the original expressions in (6) and (7) to style preserving cycle consistency loss functions

$$L_{\mathcal{G}_3} = L_{\mathcal{G}} + L_{\mathcal{G}_{\text{sty}}} + L_{\mathcal{G}_{\text{con}}} \tag{19}$$

$$L_{\mathcal{F}_3} = L_{\mathcal{F}} + L_{\mathcal{F}_{\text{sty}}} + L_{\mathcal{F}_{\text{con}}} \tag{20}$$

The style representation of an image is described as the correlation between various filter responses as determined by the image Gramian.

***Strategy III*** is depicted in Figure 4, the structure of which is illustrated in Figure 5. The modified CycleGAN model contains four loss functions: the image level (1) cycle-consistency loss $L_{\text{cyc}}$ and (2) adversarial loss $L_{D_s}, L_{D_r}$ preserve the semantic meaning of the whole image; the component-level (3) style loss of tool $L_{\mathcal{G}_{\text{sty}}}, L_{\mathcal{F}_{\text{sty}}}$ and (4) content loss of tissue $L_{\mathcal{G}_{\text{con}}}, L_{\mathcal{F}_{\text{con}}}$ trace back in the hidden layer activations to perform deep restricted style transfer locally in the surgical tool region of the images while ensuring minimal modifications to the background. The separation of foreground and background were provided in the data set as prior knowledge.




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

**Figure 4.** CycleGAN generated synthetic images from ***Strategy III***: (**a**) Tool image style is adopted from the real image, background content is preserved resulting in realistic synthetic image; (**b**) tool border is retained even if $s_j$ and $r_j$ tool shapes vary.

### *2.3. Synthetic Image Verification: Tool Segmentation*

Only images generated using ***Strategy III*** were evaluated quantitatively—the two other GAN-driven strategies were investigated but eliminated as viable methods via preliminary tests. To evaluate the utility of the synthetically generated endoscopic images as augmented training data, various combinations of real and synthetic images were used to train a U-Net, a classic surgical tool segmentation network [67]. The network was then tested on real images. These training and testing procedures were designed to answer the following two questions:

**(i)** Does incorporating synthetic images with real training images improve segmentation performance?

**(ii)** Can a large synthetic training set alone be used to train a successful segmentation network to segment test images from $\mathcal{R}$?

Table 1 shows the training set composition for the nine UNet experiments conducted. In each experiment, the UNet was trained on a different proportion of real images from $\mathcal{R}$ and GAN-generated synthetic surgical images. The resultant networks were comparatively evaluated on a fixed separate set of real surgical images.

**Table 1.** Experimental training set compositions.

| Exp | Real Images | Synthetic Images | Total Training Set |
|---|---|---|---|
| 1 | **300** | 0 | 300 |
| 2 | **300** | 100 | 400 |
| 3 | **300** | 200 | 500 |
| 4 | **300** | 300 | 600 |
| 5 | **300** | 400 | 700 |
| 6 | **300** | 500 | 800 |
| 7 | **300** | 600 | 900 |
| 8 | **300** | **5665** | 5965 |
| 9 | 0 | **5965** | 5965 |

Note: The fixed number of real training images in Exp. 1–8 (**blue**) were designed to aid in addressing research question (i), whereas Exp. 8–9 utilize training sets heavily composed of synthetic data (**gold**) to aid in addressing (ii).

## 3. Experiments

### *3.1. Raw and Baseline Data*

The image data used for this study were drawn from the publicly available, de-identified University of Washington Sinus Cadaver/Live Dataset [49,99]. This data set contains a total of 4345 cadaver and 4658 live sinus endoscopic images, denoted $\mathcal{R}$. Each endoscopic image in $\mathcal{R}$ is accompanied by a manually labeled annotation, i.e., a pixel-wise labeled mask of the surgical tool.

Section 2.1 describes the baseline synthetic image generation process combining synthetic tool and background images. For this process, a total of 354 images containing only surgical environment pixels and no surgical tools pixels were selected as background baselines. Within this baseline set of images, six categories were defined based on color composition. In total, 30,000 baseline synthetic images (i.e., synthetic tool combined with background image) were generated with uniform color theme distribution, and are denoted as $\mathcal{S}$.

### *3.2. System Workflow*

The baseline synthetic images in $\mathcal{S}$, as described in Section 2.1, exhibit realistic surgical tool pixel placement in the spatial and morphological sense. Tool pixel colorization for baseline images in $\mathcal{S}$ is uninformed of the background. To better mimic realistic endoscopic data, tool pixels must be modified to reflect the surgical environment as depicted by the surrounding background. This research presents a reliable method for enhancing synthetic baseline images in $\mathcal{S}$ into realistic ones with life-like tool pixel colorization.

In general, the approach is GAN-driven with domain transfer between $\mathcal{R}$ and $\mathcal{S}$, and several modifications of the CycleGAN code base [96] are experimentally evaluated. Empirically, ***Strategy III*** described in Section 2.2 was found to return the best synthetic image enhancement results. Thus, the following experiments were conducted to quantify the utility of GAN-generated synthetic images using ***Strategy III***. Note that since the CycleGAN approach is bidirectional, a byproduct set of images denoted $\mathcal{G}_{\mathcal{S}}$ are generated as well, as depicted in Figure 5.

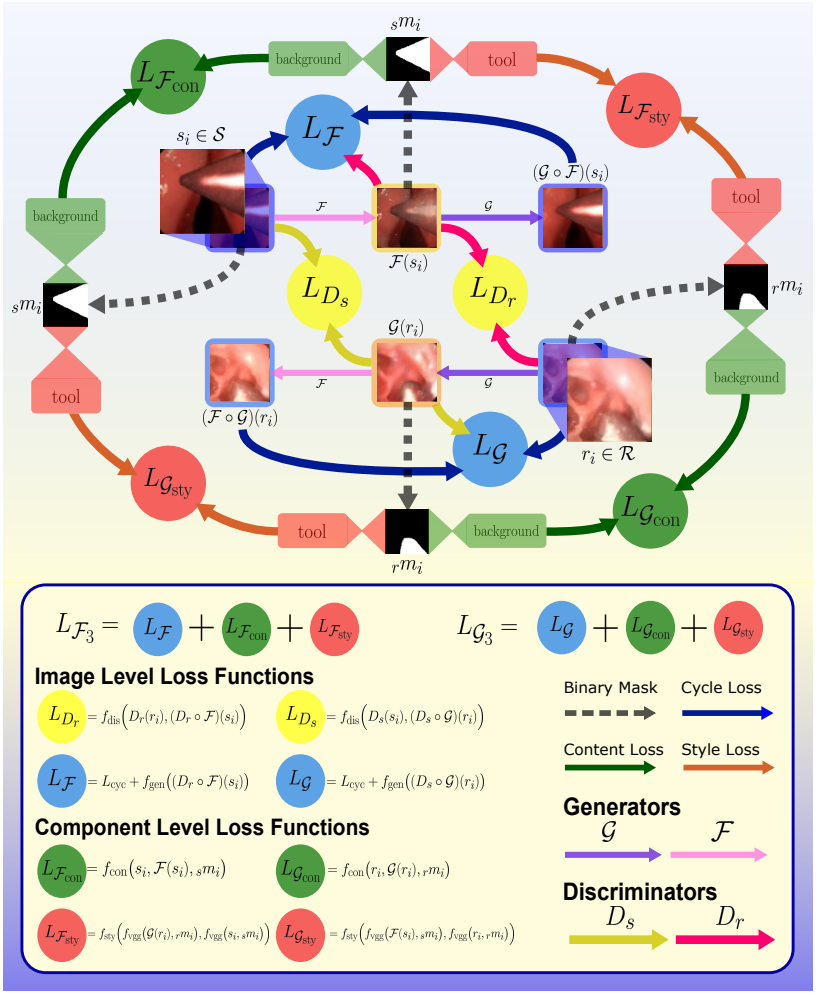



**Figure 5.** Flowchart diagram depicting ***Strategy III*** loss structure. $L_{D_r}$, $L_{D_s}$, $L_{\mathcal{F}_3}$, and $L_{\mathcal{G}_3}$ represent the loss functions used for $D_r$, $D_s$, $\mathcal{F}$, and $\mathcal{G}$, respectively.

The GAN-driven synthetic image generation procedure ideally extends the scope of otherwise limited endoscopic surgical image data sets with additional training data. If the data are realistic enough, training using the synthetically generated images should result in improved segmentation of real test images. The goal of the experimental tasks was to analytically verify that incorporating the synthetic images into training does indeed increase segmentation performance during testing. This experimental system workflow of this study can be found in Figure 6.

### *3.3. Implementation Details*

#### 3.3.1. CycleGAN Image Generation

To train the GAN-driven synthetic image generation model using strategies described in Section 2.2, a total of 5965 training images from each of the two domains, real images $\mathcal{R}$ and baseline synthetic $\mathcal{S}$, were randomly selected. Note that training images from domain $\mathcal{R}$ were chosen from a random mix of cadaver and live endoscopic images. Meanwhile, the hyperparameters including adaptive base learning rate of 0.0002, max step of 200, pool size of 200, and $\lambda$ of 10 were heuristically tuned for best performance. The GAN models were then tested on 88 images per domain.

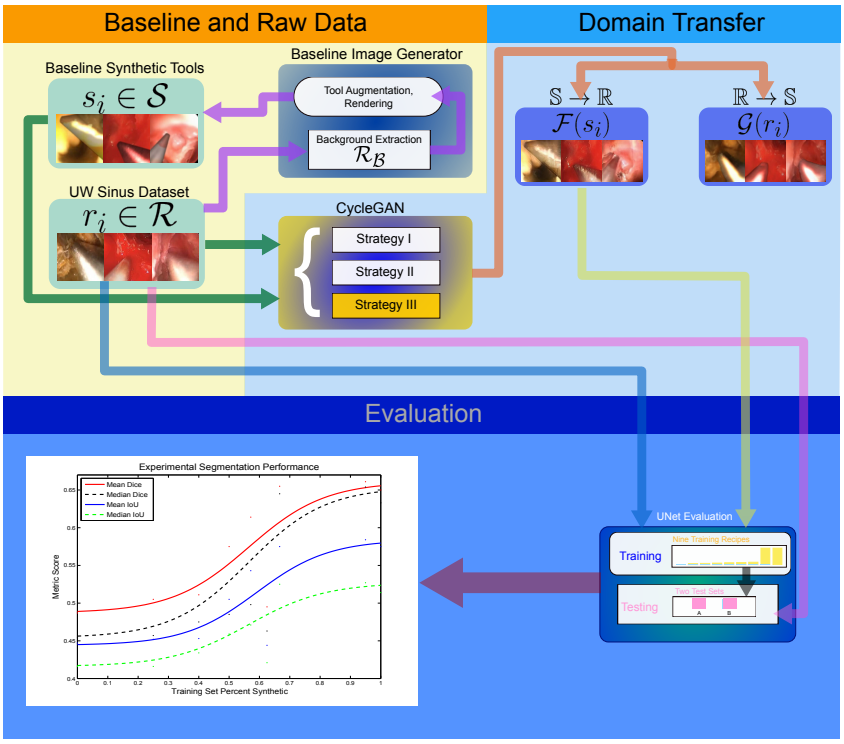



**Figure 6.** The GAN-driven synthetic image generation framework for endoscopic surgical tool segmentation.

#### 3.3.2. UNet Image Evaluation

The nine experiments as described in Table 1 were evaluated on the same UNet structure. Specifically, the hyperparameters were set to 50 epochs, 100 steps per epoch, 2 batches, Adam's optimizer learning rate of 0.0001 for optimal results and the binary crossentropy loss was set as the cost function.

The 300 real training images selected from $\mathcal{R}$ were consistent across Exp. 1 through 8. Two test sets were set aside to evaluate the UNet performances using the designated training set mixtures. In particular:

- ***Test Set A*** contains 100 random mix of cadaver and live endoscopic images in $\mathcal{R}$;
- ***Test Set B*** contains 100 selected real image frames from $\mathcal{R}$ that neighbor the 300 real training images in the University of Washington Sinus Dataset video sequence.

## 4. Results

Two widely accepted and commonly used image segmentation metrics were employed to evaluate performance of the UNet tool segmentation and thus usability of the proposed GAN-driven synthetic image generation framework in surgical tool segmentation tasks:

1. Dice coefficient [100];
2. Intersection over Union (IoU) score [101].

The average Dice and IoU scores segmenting the test set with varying training set synthetic composition is shown in Figure 7.

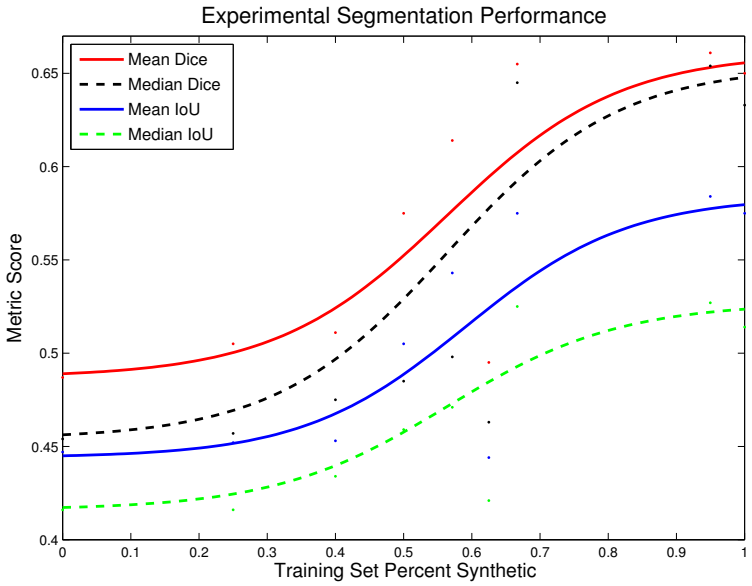



**Figure 7.** Average Dice and IoU scores for tests on ***Test Set A*** as a greater percentage of the training set is composed of synthetically generated endoscopic images using the presented method. Traces are logistic regression fits to the data.

Figures 8 and 9 depict the evaluation histograms for Dice Coefficient and for IoU, while Figure 10 shows sample predictions across Exp. 1–9. The mean and median scores are summarized in Table 2.

**Table 2.** Experimental segmentation performance results.

| Metric | Training Loss | Dice Coeff. | | | |
|---|---|---|---|---|---|
| **Exp** | **Mean** | **Mean** | **Median** | **Mean** | **Median** |
| 1 | 0.130 | 0.487 | 0.454 | 0.447 | 0.416 |
| | | 0.824 | 0.920 | 0.776 | 0.856 |
| 2 | 0.202 | 0.505 | 0.457 | 0.452 | 0.416 |
| | | 0.821 | 0.913 | 0.772 | 0.844 |
| 3 | 0.283 | 0.511 | 0.475 | 0.453 | 0.434 |
| | | 0.825 | 0.904 | 0.771 | 0.830 |
| 4 | 0.336 | 0.575 | 0.485 | 0.505 | 0.459 |
| | | 0.701 | 0.719 | 0.637 | 0.596 |
| 5 | 0.289 | 0.614 | 0.498 | 0.543 | 0.471 |
| | | 0.746 | 0.791 | 0.681 | 0.676 |
| 6 | 0.332 | 0.495 | 0.463 | 0.444 | 0.421 |
| | | 0.576 | 0.486 | 0.525 | 0.452 |
| 7 | 0.325 | 0.655 | 0.645 | 0.575 | 0.525 |
| | | 0.584 | 0.499 | 0.524 | 0.473 |
| 8 | 0.308 | 0.661 | 0.654 | 0.584 | 0.527 |
| | | 0.566 | 0.483 | 0.513 | 0.449 |
| 9 | 0.321 | 0.650 | 0.633 | 0.575 | 0.514 |
| | | 0.524 | 0.455 | 0.487 | 0.418 |

The scores displayed with white background denote results from ***Test Set A***, and those with gray tinted background indicate results from ***Test Set B***. The best performances within each test set and metric are specified with blue text.

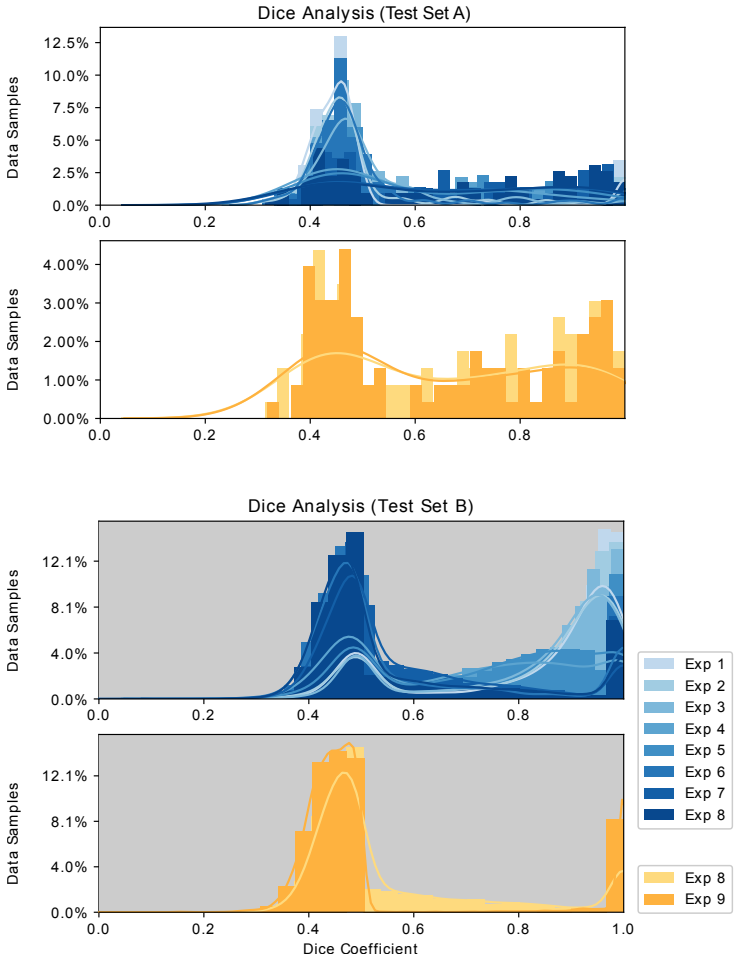



**Figure 8.** Histograms and distribution curves of the Dice coefficients of Exp. 1–8 and Exp. 8,9 evaluated on ***Test Set A*** and ***Test Set B***.

These results indicate that the composition in Exp. 1 resulted in the least training loss and the best performance on ***Test Set B***, while Exp. 8 achieved the best scores on ***Test Set A***.

## 5. Discussion

The experimental results summarized in Table 2, Figures 8 and 10 lead to several observations about the training image compositions and segmentation performances on the two test sets. These are described below.

### *5.1. Dilution of Real Training Images*

***Test Set A*** is an unbiased analysis of the UNet model segmentation performance with random endoscopic data samples, while ***Test Set B*** provides an indication of model performances for segmenting images similar to the real training images. From Exp. 1 through Exp. 9, the training set was augmented with increasing proportion of randomly selected synthetic images. As such, greater variance is introduced into the training set, and hence the trained model is more generalizable. The segmentation performance enhances with increased synthetic data augmentation using ***Test Set A***. On the other hand, the 300 real training image samples are increasingly diluted within the training set with increasing Exp. number. Thus, the trained models perform progressively worse on test images similar to the original real training images, i.e., ***Test Set B***.

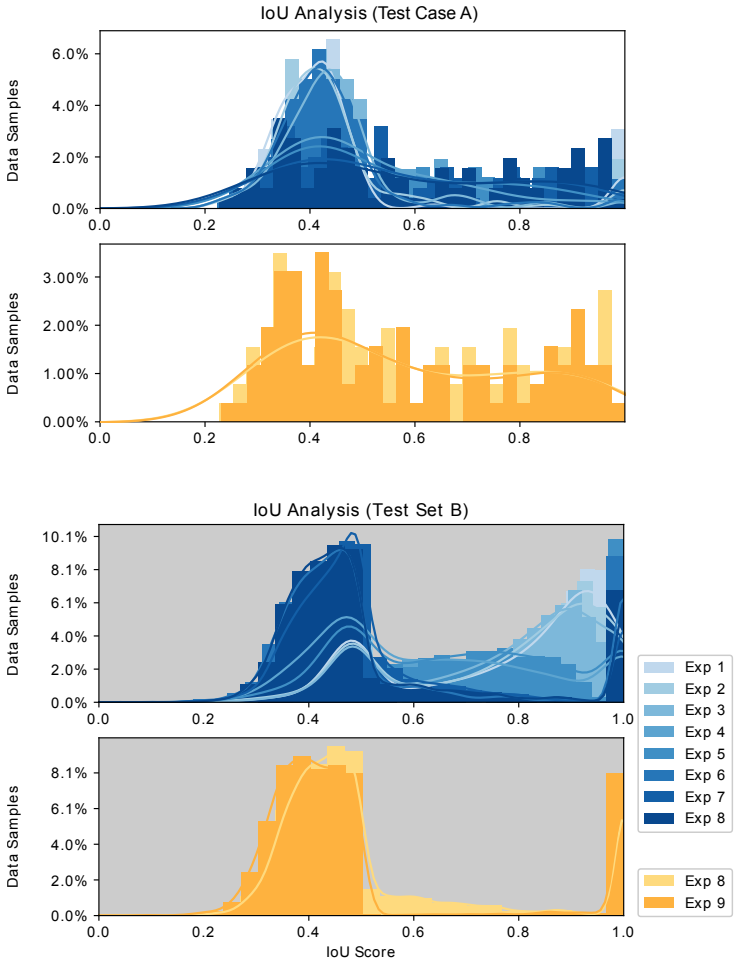



**Figure 9.** Histograms and distribution curves of the IoU scores of Exp. 1–8 and Exp. 8,9 evaluated on ***Test Set A*** and ***Test Set B***.

### *5.2. Overfitting with Small Training Set*

In Table 2, overfitting is observed in the first few experiments, as indicated by lower training loss, and good segmentation performance on ***Test Set B*** but poor segmentation with ***Test Set A***. The overfitting is also observed in the blue histograms in Figure 8, as darker histograms perform better with ***Test Set A*** but worse with ***Test Set B***. This overfitting issue unfortunately is a common problem in data-driven semantic medical image segmentation tasks when the training size is too small.

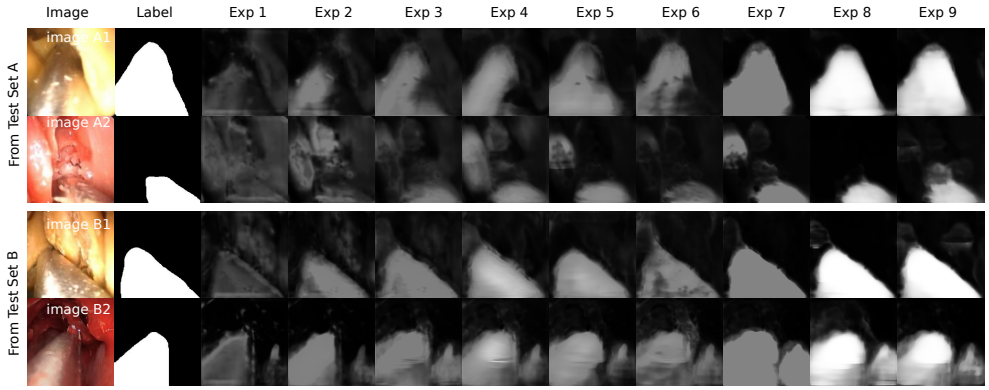



**Figure 10.** Sample predictions from ***Test Set A*** and ***Test Set B*** in Exp. 1–9. The prediction confidence is greater in later experiments. Images A1 and A2 show poor results in early experiments. Image B2 demonstrates a phantom 2nd tool falsely detected from Exp. 7 onward.

Exp. 1–8 also indicate several performance trends with increased synthetic composition of training data:

- From Exp. 5 onward, a consistent negative correlation between training loss and ***Test Set A*** scores is observed;
- From Exp. 6 onward, training loss is monotonically decreasing, ***Test Set A*** metrics monotonically increasing;
- From Exp. 7 onward, ***Test Set A*** performance surpassed that of ***Test Set B***.

These observations indicate that, in this experiment using 300 real training images, the addition of 400 (Exp. 5) or more synthetic images to the training set effectively neutralizes detrimental variance introduced by the GAN-driven synthetic images. Furthermore, the addition of 500 (Exp. 6) or more synthetic images steadily enhances general endoscopic tool segmentation performance. Meanwhile, interpreting ***Test Set B*** scores as an indicator of the training performance, Exp. 7 marks the point when the validation surpasses training performance, and therefore when overfitting is resolved.

### *5.3. Training with Synthetic Images*

In Exp. 1–8, the training set contained a fixed number of real images from $\mathcal{R}$, with increasing addition of synthetic images. The results show with increased proportion of synthetic training data, an overall enhancement of segmentation performance is observed when testing with arbitrarily selected real images, i.e., ***Test Set A***. Furthermore, with a purely synthetic training set, results are promising in Exp. 9 with performance of up to 0.650 Dice Coefficient, which is comparable to that of Exp. 8, the best performing training composition for ***Test Set A***.

### *5.4. Purely Synthetic Training*

The results from this work indicate that GAN-driven synthetic images provided enhanced surgical tool segmentation performance. However, the presence of real images in Exp. 8 resulted in superior performance to models trained with purely synthetic data, Exp. 9. This was true for both ***Test Set A*** and ***Test Set B*** measured performance, as observed in the gold histograms in Figure 8. This suggests that a large quantity of purely synthetic images generated per ***Strategy III*** does not completely replace the value of even a small number of real training images.

### *5.5. Implications*

To evaluate the practical feasibility of enhancing tool segmentation performance in robot-assisted surgical procedures using partial-synthetic training sets, nine UNet models were trained with different proportions of the synthetic and real images designed to address two queries of interest as described in Section 2.3.

Based on the numerical results in Section 4 and the aforementioned analysis in Section 5, insight is gained with regard to answering these two questions:

**(i)** The addition of the generated synthetic images to a small set of real images can indeed enhance segmentation performance. To maximize this improvement, two conditions should be followed:
  - (a) The test images share a broader variance than the set of available real training images;
  - (b) The number of synthetic images is sufficiently large.

**(ii)** A large set of purely synthetic training images as generated in this work does not eliminate the benefit of real surgical training images. With large purely synthetic training set, performance is satisfactory.

## 6. Conclusions

In summary, this research showcased a promising GAN-driven approach for generating reliable synthetic training data for surgical tool segmentation. As depicted in Table 1,

Exp. 1–8 were designed to train the UNet segmentation network with the same, high-cost real endoscopic images and with varying number of low-cost synthetic endoscopic images generated via the proposed method. Exp. 9 used a training set with purely synthetic images. As shown in Table 2, the best mean Dice and IoU scores for random test images, i.e., ***Test Set A***, were achieved with a training set consisting of 95% synthetic images. Figure 7 shows that addition of the generated synthetic data tends to increase performance. To summarize, for ***Test Set A*** mean Dice and IoU scores of:

- 0.487 and 0.447 for 0% synthetic training set;
- 0.661 and 0.584 for 95% synthetic training set

were observed, respectively. This corresponds to a 35.7% and 30.6% increase. The results are promising and suggest that the proposed method can enhance segmentation results with limited availability of real training data. The availability of even a small amount of real training data is still beneficial.

In this exploratory, baseline experiment, the addition of synthetic training images generated by the novel framework were evaluated by incorporating increasing proportion of synthetic data to the training set, which originated from a publicly available set of endoscopic surgical images. A widely used and accepted network, the UNet, was used identically across experimental conditions to execute the segmentation, and results suggest that the generated synthetic data are indeed useful and benefit tool segmentation performance. Comparable methods in the domain of endoscopic tool segmentation synthetic image generation are not readily evaluated on the same baseline dataset. In other works, methods may have been quantitatively evaluated only on a private dataset, or the source code was not made available for public use to replicate and compare.

The designed experiments do provide a guideline for systematically quantifying the usability and requirements of synthetic training data for other applications. Enhancing surgical tool segmentation can enable broader research efforts in multicamera surgical reconstruction [102,103] within the context of vision-based force estimation [104–106] and other robot-assisted medical procedures.

Two possible directions to advance this study are of interest:

(a) Compare the proposed method with other synthetic endoscopic surgical image generation approaches and combinations thereof;
(b) Automate the GAN-driven synthetic image generation process to adaptively optimize the number and distribution of training images based on training and validation sets.

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

**Funding:** This material is based upon work supported by the National Science Foundation under Grant No. IIS-2101107. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**Acknowledgments:** The authors thank Shan Lin from the University of Washington BioRobotics Laboratory and Fangbo Qin from the Institute of Automation, Chinese Academy of Sciences for their technical assistance.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References


1. Delp, S.L.; Loan, J.P.; Robinson, C.B.; Wong, A.Y.; Stulberg, S.D. Computer-Assisted Surgical System. U.S. Patent 5,682,886, 4 Novmber 1997.
2. Su, Y.H.; Lindgren, K.; Huang, K.; Hannaford, B. A Comparison of Surgical Cavity 3D Reconstruction Methods. In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 329–336.
3. Su, Y.H.; Huang, K.; Hannaford, B. Multicamera 3d reconstruction of dynamic surgical cavities: Camera grouping and pair sequencing. In Proceedings of the 2019 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, 3–5 April 2019; pp. 1–7.
4. Su, Y.H.; Huang, K.; Hannaford, B. Multicamera 3D Viewpoint Adjustment for Robotic Surgery via Deep Reinforcement Learning. *J. Med. Robot. Res.* **2021**, 2140003. [CrossRef]
5. Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J. Digit. Imaging* **2019**, *32*, 582–596.
6. Colleoni, E.; Edwards, P.; Stoyanov, D. Synthetic and Real Inputs for Tool Segmentation in Robotic Surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2020; pp. 700–710.
7. Ciecholewski, M.; Kassjański, M. Computational Methods for Liver Vessel Segmentation in Medical Imaging: A Review. *Sensors* **2021**, *21*, 2027.
8. Xue, Y.; Xu, T.; Zhang, H.; Long, L.R.; Huang, X. Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. *Neuroinformatics* **2018**, *16*, 383–392.
9. Zhang, Z.; Wu, C.; Coleman, S.; Kerr, D. DENSE-INception U-net for medical image segmentation. *Comput. Methods Programs Biomed.* **2020**, *192*, 105395.
10. Li, H.; Li, A.; Wang, M. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Comput. Biol. Med.* **2019**, *108*, 150–160.
11. Chen, S.; Ding, C.; Liu, M. Dual-force convolutional neural networks for accurate brain tumor segmentation. *Pattern Recognit.* **2019**, *88*, 90–100.
12. Dev, K.B.; Jogi, P.S.; Niyas, S.; Vinayagamani, S.; Kesavadas, C.; Rajan, J. Automatic detection and localization of Focal Cortical Dysplasia lesions in MRI using fully convolutional neural network. *Biomed. Signal Process. Control* **2019**, *52*, 218–225.
13. Karthik, R.; Gupta, U.; Jha, A.; Rajalakshmi, R.; Menaka, R. A deep supervised approach for ischemic lesion segmentation from multimodal MRI using Fully Convolutional Network. *Appl. Soft Comput.* **2019**, *84*, 105685.
14. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging. *IEEE Trans. Med. Imaging* **2017**, *36*, 2319–2330.
15. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573.
16. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; pp. 3–11.
17. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292.
18. Tong, G.; Li, Y.; Chen, H.; Zhang, Q.; Jiang, H. Improved U-NET network for pulmonary nodules segmentation. *Optik* **2018**, *174*, 460–469.
19. Vigneault, D.M.; Xie, W.; Ho, C.Y.; Bluemke, D.A.; Noble, J.A. Ω-net (omega-net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med. Image Anal.* **2018**, *48*, 95–106.
20. Zhang, J.; Du, J.; Liu, H.; Hou, X.; Zhao, Y.; Ding, M. LU-NET: An Improved U-Net for ventricular segmentation. *IEEE Access* **2019**, *7*, 92539–92546.
21. Liu, T.; Tian, Y.; Zhao, S.; Huang, X.; Wang, Q. Automatic whole heart segmentation using a two-stage u-net framework and an adaptive threshold window. *IEEE Access* **2019**, *7*, 83628–83636.
22. Curiale, A.H.; Colavecchia, F.D.; Mato, G. Automatic quantification of the LV function and mass: A deep learning approach for cardiovascular MRI. *Comput. Methods Programs Biomed.* **2019**, *169*, 37–50.
23. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.W.; Heng, P.A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674.
24. Huang, Q.; Sun, J.; Ding, H.; Wang, X.; Wang, G. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *Comput. Biol. Med.* **2018**, *101*, 153–162.
25. Wang, Y.; Song, Y.; Wang, F.; Sun, J.; Gao, X.; Han, Z.; Shi, L.; Shao, G.; Fan, M.; Yang, G. A two-step automated quality assessment for liver MR images based on convolutional neural network. *Eur. J. Radiol.* **2020**, *124*, 108822.
26. Baldeon-Calisto, M.; Lai-Yuen, S.K. AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation. *Neurocomputing* **2020**, *392*, 325–340.
27. Lee, P.Q.; Guida, A.; Patterson, S.; Trappenberg, T.; Bowen, C.; Beyea, S.D.; Merrimen, J.; Wang, C.; Clarke, S.E. Model-free prostate cancer segmentation from dynamic contrast-enhanced MRI with recurrent convolutional networks: A feasibility study. *Comput. Med. Imaging Graph.* **2019**, *75*, 14–23.

28. Weng, Y.; Zhou, T.; Li, Y.; Qiu, X. Nas-unet: Neural architecture search for medical image segmentation. *IEEE Access* **2019**, *7*, 44247–44257.
29. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207.
30. Heinrich, M.P.; Oktay, O.; Bouteldja, N. OBELISK-Net: Fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions. *Med. Image Anal.* **2019**, *54*, 1–9.
31. Zhang, R.; Huang, L.; Xia, W.; Zhang, B.; Qiu, B.; Gao, X. Multiple supervised residual network for osteosarcoma segmentation in CT images. *Comput. Med. Imaging Graph.* **2018**, *63*, 1–8.
32. Bae, H.J.; Hyun, H.; Byeon, Y.; Shin, K.; Cho, Y.; Song, Y.J.; Yi, S.; Kuh, S.U.; Yeom, J.S.; Kim, N. Fully automated 3D segmentation and separation of multiple cervical vertebrae in CT images using a 2D convolutional neural network. *Comput. Methods Programs Biomed.* **2020**, *184*, 105119.
33. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87.
34. Kablan, E.B.; Dogan, H.; Ercin, M.E.; Ersoz, S.; Ekinci, M. An ensemble of fine-tuned fully convolutional neural networks for pleural effusion cell nuclei segmentation. *Comput. Electr. Eng.* **2020**, *81*, 106533.
35. Rad, R.M.; Saeedi, P.; Au, J.; Havelock, J. Trophectoderm segmentation in human embryo images via inceptioned U-Net. *Med. Image Anal.* **2020**, *62*, 101612.
36. Colonna, A.; Scarpa, F.; Ruggeri, A. Segmentation of corneal nerves using a u-net-based convolutional neural network. In *Computational Pathology and Ophthalmic Medical Image Analysis*; Springer: Cham, Switzerland, 2018; pp. 185–192.
37. Wang, B.; Qiu, S.; He, H. Dual Encoding U-Net for Retinal Vessel Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*; Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 84–92.
38. Civit-Masot, J.; Luna-Perejon, F.; Vicente-Diaz, S.; Corral, J.M.R.; Civit, A. TPU cloud-based generalized U-Net for eye fundus image segmentation. *IEEE Access* **2019**, *7*, 142379–142387.
39. Zhang, S.; Zheng, R.; Luo, Y.; Wang, X.; Mao, J.; Roberts, C.J.; Sun, M. Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network. *IEEE Access* **2019**, *7*, 57561–57573.
40. Jin, Q.; Meng, Z.; Pham, T.D.; Chen, Q.; Wei, L.; Su, R. DUNet: A deformable network for retinal vessel segmentation. *Knowl.-Based Syst.* **2019**, *178*, 149–162.
41. Laves, M.H.; Bicker, J.; Kahrs, L.A.; Ortmaier, T. A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 483–492.
42. Ji, B.; Ren, J.; Zheng, X.; Tan, C.; Ji, R.; Zhao, Y.; Liu, K. A multi-scale recurrent fully convolution neural network for laryngeal leukoplakia segmentation. *Biomed. Signal Process. Control* **2020**, *59*, 101913. [CrossRef]
43. Baumhauer, M.; Feuerstein, M.; Meinzer, H.P.; Rassweiler, J. Navigation in endoscopic soft tissue surgery: Perspectives and limitations. *J. Endourol.* **2008**, *22*, 751–766.
44. Reiter, A.; Allen, P.K.; Zhao, T. Feature classification for tracking articulated surgical tools. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 592–600.
45. Lahanas, V.; Loukas, C.; Georgiou, E. A simple sensor calibration technique for estimating the 3D pose of endoscopic instruments. *Surg. Endosc.* **2016**, *30*, 1198–1204.
46. Allan, M.; Ourselin, S.; Thompson, S.; Hawkes, D.J.; Kelly, J.; Stoyanov, D. Toward detection and localization of instruments in minimally invasive surgery. *IEEE Trans. Biomed. Eng.* **2012**, *60*, 1050–1058.
47. Zhou, J.; Payandeh, S. Visual tracking of laparoscopic instruments. *J. Autom. Control. Eng. Vol* **2014**, *2*, 234–241.
48. Allan, M.; Chang, P.L.; Ourselin, S.; Hawkes, D.J.; Sridhar, A.; Kelly, J.; Stoyanov, D. Image based surgical instrument pose estimation with multi-class labelling and optical flow. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 331–338.
49. Lin, S.; Qin, F.; Bly, R.A.; Moe, K.S.; Hannaford, B. UW Sinus Surgery Cadaver/Live Dataset (UW-Sinus-Surgery-C/L). 2020. Available online: https://digital.lib.washington.edu/researchworks/handle/1773/45396 (accessed on 28 October 2020).
50. Rieke, N.; Tan, D.J.; Alsheakhali, M.; Tombari, F.; di San Filippo, C.A.; Belagiannis, V.; Eslami, A.; Navab, N. Surgical tool tracking and pose estimation in retinal microsurgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 266–273.
51. Reiter, A.; Allen, P.K. An online learning approach to in-vivo tracking using synergistic features. In Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 3441–3446.
52. McKenna, S.; Charif, H.N.; Frank, T. Towards video understanding of laparoscopic surgery: Instrument tracking. In Proceedings of the Image and Vision Computing, Auckland, New Zealand, 23–24 November 2005.
53. Alsheakhali, M.; Yigitsoy, M.; Eslami, A.; Navab, N. Surgical tool detection and tracking in retinal microsurgery. In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*; International Society for Optics and Photonics: Orlando, FL, USA, 2015; Volume 9415, p. 941511.
54. Bouget, D.; Benenson, R.; Omran, M.; Riffaud, L.; Schiele, B.; Jannin, P. Detecting surgical tools by modelling local appearance and global shape. *IEEE Trans. Med. Imaging* **2015**, *34*, 2603–2617.

55. Sznitman, R.; Becker, C.; Fua, P. Fast part-based classification for instrument detection in minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2014; pp. 692–699.
56. Wolf, R.; Duchateau, J.; Cinquin, P.; Voros, S. 3D tracking of laparoscopic instruments using statistical and geometric modeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 203–210.
57. Kumar, S.; Narayanan, M.S.; Singhal, P.; Corso, J.J.; Krovi, V. Product of tracking experts for visual tracking of surgical tools. In Proceedings of the 2013 IEEE International Conference on Automation Science and Engineering (CASE), Madison, WI, USA, 17–20 August 2013; pp. 480–485.
58. Qin, F.; Li, Y.; Su, Y.H.; Xu, D.; Hannaford, B. Surgical instrument segmentation for endoscopic vision with data fusion of cnn prediction and kinematic pose. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 9821–9827.
59. Gupta, S.; Ali, S.; Goldsmith, L.; Turney, B.; Rittscher, J. Mi-unet: Improved segmentation in ureteroscopy. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 212–216.
60. Jha, D.; Ali, S.; Emanuelsen, K.; Hicks, S.A.; Thambawita, V.; Garcia-Ceja, E.; Riegler, M.A.; de Lange, T.; Schmidt, P.T.; Johansen, H.D.; et al. Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy. In *International Conference on Multimedia Modeling*; Springer: Cham, Switzerland, 2021, pp. 218–229.
61. Roß, T.; Reinke, A.; Full, P.M.; Wagner, M.; Kenngott, H.; Apitz, M.; Hempe, H.; Mindroc-Filimon, D.; Scholz, P.; Tran, T.N.; et al. Comparative validation of multi-instance instrument segmentation in endoscopy: Results of the ROBUST-MIS 2019 challenge. *Med. Image Anal.* **2021**, *70*, 101920.
62. Islam, M.; Atputharuban, D.A.; Ramesh, R.; Ren, H. Real-time instrument segmentation in robotic surgery using auxiliary supervised deep adversarial learning. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2188–2195.
63. Colleoni, E.; Stoyanov, D. Robotic instrument segmentation with image-to-image translation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 935–942.
64. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48.
65. Bloice, M.D.; Stocker, C.; Holzinger, A. Augmentor: An image augmentation library for machine learning. *arXiv* **2017**, arXiv:1708.04680.
66. Lindgren, K.; Kalavakonda, N.; Caballero, D.E.; Huang, K.; Hannaford, B. Learned hand gesture classification through synthetically generated training samples. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 3937–3942.
67. 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: Cham, Switzerland, 2015; pp. 234–241.
68. Eaton-Rosen, Z.; Bragman, F.; Ourselin, S.; Cardoso, M.J. Improving data augmentation for medical image segmentation. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL), Amsterdam, The Netherlands, 4–6 July 2018.
69. Kikinis, R.; Pieper, S.D.; Vosburgh, K.G. 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative Imaging and Image-Guided Therapy*; Springer: New York, USA, 2014; pp. 277–289.
70. Whittaker, G.; Aydin, A.; Raison, N.; Kum, F.; Challacombe, B.; Khan, M.S.; Dasgupta, P.; Ahmed, K. Validation of the RobotiX mentor robotic surgery simulator. *J. Endourol.* **2016**, *30*, 338–346.
71. Perrenot, C.; Perez, M.; Tran, N.; Jehl, J.P.; Felblinger, J.; Bresler, L.; Hubert, J. The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. *Surg. Endosc.* **2012**, *26*, 2587–2593.
72. Munawar, A.; Srishankar, N.; Fischer, G.S. An Open-Source Framework for Rapid Development of Interactive Soft-Body Simulations for Real-Time Training. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 6544–6550.
73. Choueib, S.; Pinter, C.; Lasso, A.; Fillion-Robin, J.C.; Vimort, J.B.; Martin, K.; Fichtinger, G. Evaluation of 3D slicer as a medical virtual reality visualization platform. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. International Society for Optics and Photonics: San Diego, CA, USA, 2019; Volume 10951, p. 1095113.
74. Hertz, A.M.; George, E.I.; Vaccaro, C.M.; Brand, T.C. Head-to-head comparison of three virtual-reality robotic surgery simulators. *JSLS J. Soc. Laparoendosc. Surg.* **2018**, *22*, e2017.00081. [CrossRef]
75. Mahmood, F.; Chen, R.; Durr, N.J. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Trans. Med. Imaging* **2018**, *37*, 2572–2581.
76. Lin, S.; Qin, F.; Li, Y.; Bly, R.A.; Moe, K.S.; Hannaford, B. LC-GAN: Image-to-image Translation Based on Generative Adversarial Network for Endoscopic Images. *arXiv* **2020**, arXiv:2003.04949.
77. Su, Y.H.; Huang, K.; Hannaford, B. Real-time vision-based surgical tool segmentation with robot kinematics prior. In Proceedings of the 2018 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, 1–3 March 2018; pp. 1–6.
78. Su, Y.H.; Huang, I.; Huang, K.; Hannaford, B. Comparison of 3d surgical tool segmentation procedures with robot kinematics prior. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 4411–4418.
79. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. *arXiv* **2017**, arXiv:1703.00848.
80. Yi, X.; Walia, E.; Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **2019**, *58*, 101552.

81. Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. Icdar. 2003; Volume 3. Available online: https://www.researchgate.net/profile/John-Platt-2/publication/2880624_Best_Practices_for_Convolutional_Neural_Networks/links/00b49524c79b1afb07000000/Best-Practices-for-Convolutional-Neural-Networks.pdf (accessed on 12 October 2020).
82. Zhang, X.; Smith, N.; Webb, A. Medical imaging. In *Biomedical Information Technology*; Elsevier: London, UK, 2008; pp. 3–27.
83. Ha, E.; Shin, J.; Paik, J. Gated Dehazing Network via Least Square Adversarial Learning. *Sensors* **2020**, *20*, 6311.
84. Emami, H.; Dong, M.; Nejad-Davarani, S.P.; Glide-Hurst, C.K. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med. Phys.* **2018**, *45*, 3627–3636.
85. Mok, T.C.; Chung, A.C. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2018; pp. 70–80.
86. Shin, H.C.; Tenenholtz, N.A.; Rogers, J.K.; Schwarz, C.G.; Senjem, M.L.; Gunter, J.L.; Andriole, K.P.; Michalski, M. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Cham, Switzerland, 2018; pp. 1–11.
87. Gu, X.; Knutsson, H.; Nilsson, M.; Eklund, A. Generating diffusion MRI scalar maps from T1 weighted images using generative adversarial networks. In *Scandinavian Conference on Image Analysis*; Springer: Cham, Switzerland, 2019; pp. 489–498.
88. Hu, Y.; Gibson, E.; Lee, L.L.; Xie, W.; Barratt, D.C.; Vercauteren, T.; Noble, J.A. Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks. In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*; Springer: Cham, Switzerland, 2017; pp. 105–115.
89. Tom, F.; Sheet, D. Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1174–1177.
90. Galbusera, F.; Niemeyer, F.; Seyfried, M.; Bassani, T.; Casaroli, G.; Kienle, A.; Wilke, H.J. Exploring the potential of generative adversarial networks for synthesizing radiological images of the spine to be used in in silico trials. *Front. Bioeng. Biotechnol.* **2018**, *6*, 53.
91. Mahapatra, D.; Bozorgtabar, B.; Thiran, J.P.; Reyes, M. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2018; pp. 580–588.
92. Burlina, P.M.; Joshi, N.; Pacheco, K.D.; Liu, T.A.; Bressler, N.M. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA Ophthalmol.* **2019**, *137*, 258–264.
93. Jin, D.; Xu, Z.; Tang, Y.; Harrison, A.P.; Mollura, D.J. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2018; pp. 732–740.
94. Zhang, Z.; Rosa, B.; Nageotte, F. Surgical Tool Segmentation using Generative Adversarial Networks with Unpaired Training Data. *IEEE Robot. Autom. Lett.* **2021**, *6*, 6266–6273.
95. Wang, H.; Xiong, H.; Cai, Y. Image Localized Style Transfer to Design Clothes Based on CNN and Interactive Segmentation. *Comput. Intell. Neurosci.* **2020**, *2020*. [CrossRef]
96. 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.
97. Su, Y.H.; Chitrakar, D.; Jiang, W.; Huang, K. The Modified UNet Source Code for This Research. 2021. Available online: https://github.com/MHC-CycleGAN-Research/Our-UNet-Code (accessed on 10 February 2021).
98. Jiang, W.; Su, Y.H. The Modified CycleGAN Source Code for This Research. 2021. Available online: https://github.com/MHC-CycleGAN-Research/Our-CycleGAN-Code (accessed on 28 November 2020).
99. Qin, F.; Lin, S.; Li, Y.; Bly, R.A.; Moe, K.S.; Hannaford, B. Towards better surgical instrument segmentation in endoscopic vision: multi-angle feature aggregation and contour supervision. *IEEE Robot. Autom. Lett.* **2020**, *5*, 6639–6646.
100. Shamir, R.R.; Duchin, Y.; Kim, J.; Sapiro, G.; Harel, N. Continuous dice coefficient: A method for evaluating probabilistic segmentations. *arXiv* **2019**, arXiv:1906.11031.
101. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
102. Su, Y.H.; Huang, K.; Hannaford, B. Multicamera 3d reconstruction of dynamic surgical cavities: Non-rigid registration and point classification. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 7911–7918.
103. Su, Y.H.; Huang, K.; Hannaford, B. Multicamera 3d reconstruction of dynamic surgical cavities: Autonomous optimal camera viewpoint adjustment. In Proceedings of the 2020 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, 18–20 November 2020; pp. 103–110
104. Huang, K.; Chitrakar, D.; Mitra, R.; Subedi, D.; Su, Y.H. Characterizing limits of vision-based force feedback in simulated surgical tool-tissue interaction. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4903–4908.

105. Nazari, A.A.; Janabi-Sharifi, F.; Zareinia, K. Image-Based Force Estimation in Medical Applications: A Review. *IEEE Sens. J.* **2021**, *21*, 7. [CrossRef]
106. Su, Y.H.; Sosnovskaya, Y.; Hannaford, B.; Huang, K. Securing Robot-assisted Minimally Invasive Surgery through Perception Complementarities. In Proceedings of the 2020 Fourth IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 9–11 November 2020; pp. 41–47.