mathematics

MDPI

Article

A Multi-Stage GAN for Multi-Organ Chest X-ray Image Generation and Segmentation

Giorgio Ciano [1,2,*], Paolo Andreini [2], Tommaso Mazzierli [3], Monica Bianchini [2] and Franco Scarselli [2]

[1] Department of Information Engineering, University of Florence, 50121 Florence, Italy
[2] Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy; paolo.andreini@unisi.it (P.A.); monica.bianchini@unisi.it (M.B.); franco@diism.unisi.it (F.S.)
[3] Department of Nephrology, AOU Careggi, University of Florence, 50121Florence, Italy; tommaso.mazzierli@unifi.it
* Correspondence: giorgio.ciano@unifi.it

Abstract: Multi-organ segmentation of X-ray images is of fundamental importance for computer aided diagnosis systems. However, the most advanced semantic segmentation methods rely on deep learning and require a huge amount of labeled images, which are rarely available due to both the high cost of human resources and the time required for labeling. In this paper, we present a novel multi-stage generation algorithm based on Generative Adversarial Networks (GANs) that can produce synthetic images along with their semantic labels and can be used for data augmentation. The main feature of the method is that, unlike other approaches, generation occurs in several stages, which simplifies the procedure and allows it to be used on very small datasets. The method was evaluated on the segmentation of chest radiographic images, showing promising results. The multi-stage approach achieves state-of-the-art and, when very few images are used to train the GANs, outperforms the corresponding single-stage approach.

Keywords: deep learning; convolutional neural networks; semantic segmentation; generative adversarial networks; chest X-ray; image augmentation

Citation: Ciano, G.; Andreini, P.; Mazzierli, T.; Bianchini, M.; Scarselli, F. A Multi-Stage GAN for Multi-Organ Chest X-ray Image Generation and Segmentation. *Mathematics* **2021**, *9*, 2896. https://doi.org/10.3390/math9222896

Academic Editor: Ezequiel López-Rubio

Received: 9 October 2021
Accepted: 11 November 2021
Published: 14 November 2021

1. Introduction

Chest X-ray (CXR) is one of the most used techniques worldwide for the diagnosis of various diseases, such as pneumonia, tuberculosis, infiltration, heart failure and lung cancer. Chest X-rays have enormous advantages: they are cheap, X-ray equipment is also available in the poorest areas of the world and, moreover, the interpretation/reporting of X-rays is less operator-dependent than the results of other more advanced techniques, such as computed tomography (CT) and magnetic resonance (RMI). Furthermore, undergoing this examination is very fast and minimally invasive [1]. Recently, CXR images have gained even greater importance due to COVID-19, which mainly causes lung infection and, after healing, often leaves widespread signs of pulmonary fibrosis: the respiratory tissue affected by the infection loses its characteristics and its normal structure. Consequently, CXR images are often used for the diagnosis of COVID-19 and for treatment of the after-effects of SARS-CoV-2 [2–4].

Therefore, with the rapid growth in the number of CXRs performed per patient, there is an ever-increasing need for computer-aided diagnosis (CAD) systems to assist radiologists, since manual classification and annotation is time-consuming and subject to errors. Recently, deep learning (DL) has radically changed the perspective in medical image processing, and deep neural networks (DNNs) have been applied to a variety of tasks, including organ segmentation, object and lesion classification [5], image generation and registration [6]. These DL methods constitute an important step towards the construction of CADs for medical images and, in particular, for CXRs.

Semantic segmentation of anatomical structures is the process of classifying each pixel of an image according to the structure to which it belongs. In CAD, segmentation plays a fundamental role. Indeed, segmentation of CXR images is usually necessary to obtain regions of interest and allows the extraction of size measurements of organs (e.g., cardiothoracic ratio quantification) and irregular shapes, which can provide meaningful information on important diseases, such as cardiomegaly, emphysema and lung nodules [7]. Segmentation may also help to improve the performance of automatic classification: in [8], it is shown that, by exploiting segmentation, DL models focus their attention primarily on the lung, not taking into account unnecessary background information and noise.

Modern state-of-the-art segmentation algorithms are largely based on DNNs [9–11]. However, to achieve good results, DNNs need a fairly large amount of labeled data. Therefore, the main problem with segmentation by DNNs is the scarce availability of appropriate datasets to help solve a given task. This problem is even more evident in the medical field, where data availability is affected by privacy concerns and where a great deal of time and human resources are required to manually label each pixel of each image.

A common solution to cope with this problem is the generation of synthetic images, along with their semantic label maps. This task can be carried out by Generative Adversarial Networks (GANs) [12], which can learn, using few training examples, the data distribution in a given domain. In this paper, we present a new model, based on GANs, to generate multi-organ segmentation of CXR images. Unlike other approaches, the main feature of the proposed method is that generation occurs in three stages. In the first stage, the position of each anatomical part is generated and represented by a"dot" within the image; in the second stage, semantic labels are obtained from the dots; finally, the chest X-ray image is generated. Each step is implemented by a GAN. More precisely, we adopt Progressively Growing GANs (PGGANs) [13], a recent extension of GANs that allows the generation of high resolution images, and Pix2PixHD [14] for the translation steps. The intuitive idea underlying the approach is that generation benefits by the multi-stage procedure, since the GAN used in each single step faces a subproblem, and can be trained using fewer data. Actually, the generalization capability of neural networks, and more generally of deep learning approaches, has a solid mathematical foundation (see, e.g., the seminal work [15] and the more recent papers [16,17]). The most general rule states that the simpler the model the better its generalization capability. In our approach, the simplification lies in that, in the three-stage method, the tasks to be solved in each of the three steps are simpler and require less effort.

In order to evaluate the performance of the proposed method, synthetic images were used to train a segmentation network (here, we use the Segmentation Multiscale Attention Network (SMANet) [18], a deep convolutional neural network based on the Pyramid Scene Parsing Network [11]), subsequently applied to a popular benchmark for multi-organ chest segmentation, the Segmentation in Chest Radiographs (SCR) dataset [6]. The results obtained are very promising and exceed (to the best of our knowledge) those obtained by other previous methods. Moreover, the quality of the produced segmentation was confirmed by physicians. Finally, to demonstrate the capabilities of our approach, especially having little data available, we compared it to two other methods, using only 10% of the images in the dataset. In particular, the multi-stage approach was compared with a single-stage method—in which chest X-ray images and semantic label maps are generated simultaneously—and with a two-stage method—where semantic label maps are generated and then translated into X-ray images. The experimental results show that the proposed three-stage method outperforms the two-stage method, while the two-stage overcomes the single-stage approach, confirming that splitting the generation procedure can be advantageous, particularly when few training images are available.

The paper is organized as follows. In Section 2, the related literature is reviewed. Section 3 presents a description of the proposed image generation method. Section 4 shows and discusses the experimental results. Finally, in Section 5, we draw some conclusions and describe future research.

2. Related Works

In the following, recent works related to the topics addressed in this paper are briefly reviewed, namely regarding synthetic image generation, image-to-image translation, and the segmentation of medical images.

2.1. Synthetic Image Generation

Methods for generating images are by no means new and can be classified into two main categories: model-based and learning-based approaches. A model-based method consists of formulating a model of the observed data to render the image by a dedicated engine. This approach has been widely adopted to generate images in many different domains [19–21]. Nonetheless, the design of specialized engines for data generation requires a deep knowledge of the specific domain. For this reason, in recent years, the learning-based approach has attracted increasing research interest. In this context, machine learning techniques are used to capture the intrinsic variability of a set of training images, so that the specific domain model is acquired implicitly from the data. Once the probability distribution that underlies the set of real images has been learned, the system can be used to generate new images that are likely to mimic the original ones. One of the most successful machine learning models for data generation is the Generative Adversarial Network (GAN) [12]. A GAN is composed by two networks: a generator G and a discriminator D. The former learns to generate data starting from a latent random variable $\mathbf{z} \in \mathbb{R}^Z$, while the latter aims at distinguishing real data from generated ones. Training GANs is difficult, because it consists of a min-max game between two neural networks and convergence is not guaranteed. This problem is compounded in the generation of high resolution images, because the high resolution makes it easier to distinguish generated images from training images [22]. One of the most successful approaches to face this problem is represented by Progressively Growing GANs (PGGANs) [13]. This model, in fact, is based on a multi-stage approach that aims to simplify and stabilize the training and allows it to generate high resolution images. More specifically, in a PGGAN, the training starts at low resolution, while new blocks are progressively introduced into the system to increase the resolution of the generation. The generator and discriminator grow symmetrically until the desired resolution is reached. Based upon PGGANs, many different approaches have been proposed. For instance, StyleGANs [23] maintain the same discriminator as PGGANs, but introduce a new generator which is able to control the style of the generated images at different levels of detail. In StyleGAN2s [24], an improved training scheme is introduced, which achieves the same goal—training starts by focusing on low resolution images and then progressively shifts the focus to higher and higher resolutions—without changing the network topology during training. In this way, the updated model shows improved results at the expense of longer training times and more computing resources.

In this paper, we use PGGANs in three different ways. For the single-stage method, a PGGAN simultaneously generates semantic label maps and CXR images. For the two-stage method, only semantic label maps are generated, while for the three-stage method we use a PGGAN to generate "dots" that correspond to different anatomical parts.

2.2. Image-to-Image Translation

Recently, besides image generation, adversarial learning has also been employed for image-to-image translation, the goal of which is to translate an input image from one domain to another. Many computer vision tasks, such as image super-resolution [25], image inpainting [26], and style transfer [27], can be cast into the image-to-image translation framework. Both unsupervised [28–31] and supervised approaches [13,32,33] can be used but, for the proposed application to CXR image generation, the unsupervised category is not relevant. Supervised training uses a set of pairs of corresponding images $\{(s_i, t_i)\}$, where s_i is an image of the source domain and t_i is the corresponding image in the target domain. In the original GAN framework, there is no explicit way of controlling what to generate, since the output depends only on the latent vector $\mathbf{z}$. For this reason, in conditional GANs

(cGANs) [34], an additional input $\mathbf{c}$ is introduced to guide the generation. In a cGAN, the generator can be defined accordingly as $G(\mathbf{c}, \mathbf{z})$. Pix2Pix [32] is a general approach for image-to-image translation and consists of a conditional GAN that operates in a supervised way. Pix2Pix uses a loss function that allows it to generate plausible images in relation to the destination domain, which are also credible translations of the input image. With respect to supervised image-to-image translation techniques, in addition to the aforementioned Pix2Pix, the most used models are CRN [33], Pix2PixHD [14], BycicleGAN [35], SIMS [36], and SPADE [37]. In particular, Pix2PixHD [14] improves upon Pix2Pix by employing a coarse-to-fine generator and discriminator, along with a feature-matching loss function, allowing it to translate images with higher resolution and quality.

For the image-to-image translation phase, we use the Pix2PixHD network. The single-stage method does not require a translation step, while for the two-stage method we use Pix2PixHD to obtain a CXR image from the label map. Finally, in the three-stage method, Pix2PixHD is used in two steps: for the translation from "dots" to semantic label maps and, after that, for the translation of label maps into CXR images.

2.3. Medical Image Generation

In recent years, GANs have attracted the attention of medical researchers, their applications ranging from object detection [38–40] to registration [41–43], classification [44–46] and segmentation [47,48] of images. For instance, in [49], different GANs have been used for the synthesis of each class of liver lesion (cysts, metastases and hemangiomas). However, in the medical domain, the use of complex machine learning models is often limited by the difficulty of collecting large sets of data. In this context, GANs can be employed to generate synthetic data, realizing a form of data augmentation. In fact, GAN generated data can be used to enlarge the available datasets and improve the performance in different tasks. As an example, GAN generated images have been successfully used to improve the performance in classification problems, by combining real and synthetic images during the training of a classifier. In [50], Wasserstein GANs (WGANs) and InfoGANs have been combined to classify histopathological images, whereas in [44] WGAN and CatGAN generated images were used to improve the classification of dermoscopic images. Only in a few cases have GANs been used to generate chest radiographic images, as in [45], where images for cardiac abnormality classification were obtained with a semi-supervised architecture, or in [51], where GANs were used to generate low resolution (64×64) CXRs to diagnose pneumonia. More related to this work, in [19], high-resolution synthetic images of the retina and the corresponding semantic label maps have been generated. Moreover, synthesizing images has been proven to be an effective method for data augmentation, that can be used to improve performance in retinal vessel segmentation.

In this paper, chest X-ray images were generated with the corresponding semantic label maps (which correspond to different organs). We then used such images to train a segmentation network, with very promising results.

2.4. Organ Segmentation

X-rays are one of the most used techniques in medical diagnostics. The reasons are medical and economic, since they are cheap, noninvasive and fast examinations. Many diseases, such as pneumonia, tuberculosis, lung cancer, and heart failure are commonly diagnosed from CXR images. However, due to overlapping organs, low resolution and subtle anatomical shape and size variations, interpreting CXRs accurately remains challenging and requires highly qualified and trained personnel. Therefore, it is of a great clinical and scientific interest to develop computer-based systems that support the analysis of CXRs. In [52], a lung boundary detection system was proposed, building an anatomical atlas to be used in combination with graph cut-based image region refinement [53–55]. A method for lung field segmentation, based on joint shape and appearance sparse learning, was proposed in [56], while a technique for landmark detection was presented in [57]. Haar-like features and a random forest classifier were combined for the appearance of

landmarks. Furthermore, a Gaussian distribution augmented by shape-based random forest classifiers was adopted for learning spatial relationships between landmarks. *InvertedNet*, an architecture able to segment the heart, clavicles and lungs, was introduced in [58]. This network employs a loss function based on the Dice Coefficient, Exponential Linear Units (ELUs) activation functions, and a model architecture that aims at containing the number of parameters. Moreover, the UNet [59] architecture has been widely used for lung segmentation, as in [60–62]. In the Structure Correcting Adversarial Network (SCAN) [63] a segmentation network and a critic network were jointly trained with an adversarial mechanism for organ segmentation in chest X-rays.

3. Chest X-ray Generation

The main goal of this study is to prove that by dividing the generation problem into multiple simpler stages, the quality of the generated images improves, so that they can be more effectively employed as a form of data augmentation. More specifically, we compare three different generation approaches. The first method, described in Section 3.1, consists of generating chest X-ray images and the corresponding label maps in a single stage. In the second approach, presented in Section 3.2, the generation procedure is divided into two stages, where the label maps are initially generated and then translated into images. The third method, reported in Section 3.3, consists of a three-stage approach, that starts by generating the position of the objects in the image, then the label maps and, finally, the X-ray images. The images generated employing each of the three approaches are comparatively evaluated by training a segmentation network.

To increase the descriptive power of real images, especially with regards to the position of the various organs, standard data augmentation has preventively been applied. Therefore, the original X-ray images, along with their corresponding masks, were augmented by applying random rotations in the interval $[-2, 2]$ degrees, random horizontal, vertical and combined translations from -3% to $+3\%$ of the number of pixels, and adding a Gaussian noise—only to the original images—with a zero mean and variance between 0.01 and 0.03×255. For the generation of images, we essentially used two networks well known in the literature, namely PGGANs [13] and Pix2PixHD [14], and their details are given in the following sections. In particular, in Sections 3.1–3.3, we extensively describe the three different generation procedures, respectively the single-stage, two-stage and three-stage methods. The next Section 3.4 presents the semantic segmentation network that was employed. Finally, some details on the training method are collected in Section 3.5.

3.1. Single-Stage Method

This baseline approach consists of stacking X-ray images and labels into two different channels, which are simultaneously fed into the PGGAN. Therefore, the PGGAN is trained to generate pairs composed by an X-ray image and its corresponding label (see Figure 1).

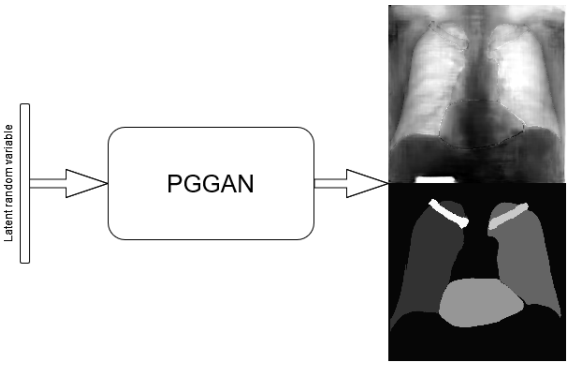


Figure 1. The one-stage image generation scheme. The input of the network is a latent vector, while the PGGAN simultaneously produces the label map and the X-ray image.

3.2. Two-Stage Method

In this approach, the generation procedure is divided into two steps. The first one consists of generating the labels through a PGGAN, while, in the second, the translation from the label to the corresponding chest X-ray image is carried out using Pix2PixHD (see Figure 2).

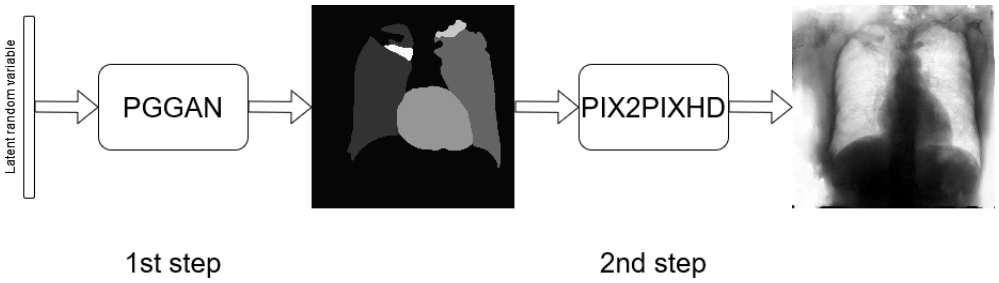


Figure 2. The two-stage image generation scheme. In the first step, the PGGAN takes in input as a latent vector and produces the label map. The generated label map is then used as input to a Pix2PixHD module, which is trained to output the X-ray image.

3.3. Three-Stage Method

It consists of further subdividing the generation procedure, with a first phase consisting of generating the position and type of the objects that will be generated later, regardless of their shape or appearance. This is obtained by generating label maps that contain "dots" in correspondence with different anatomical parts (lungs, heart, clavicles). The dots can be considered as "seeds", from which, through the subsequent steps, the complete label maps are realized (second phase). Finally, in the last step, chest X-ray images are generated from the label maps. The exact procedure is described in the following. Initially, label maps containing "dots", with a specific value for each anatomical part, are created. The position of the "dot" center is given by the centroid of each labeled anatomical part. The label maps generated in this phase have a low resolution (64×64), as a high level of detail is not necessary, because the exact object shapes are not defined—but only their centroid positions. It should be observed that this also allows a significant reduction in the computational burden of this stage and speeds up the computation. The generated label maps must be subsequently resized to the original image resolution—required in the following stages of generation (a nearest neighbour interpolation was used to maintain the original label codes)—and translated into labels, which will be finally translated into images, using Pix2PixHD (see Figure 3).

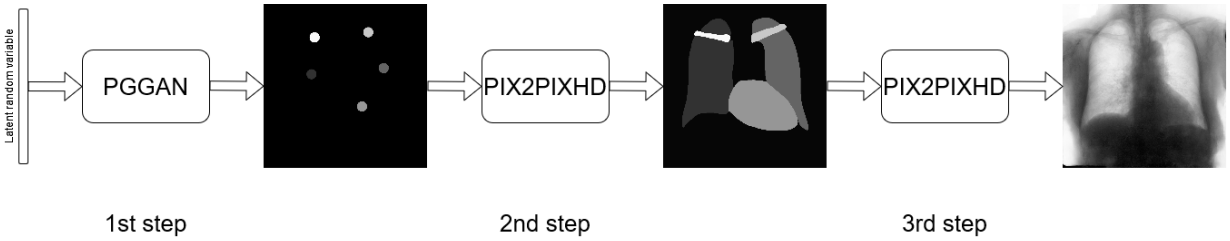


Figure 3. The three-stage image generation scheme. In the first step, dots are generated from a latent vector. Then, Pix2PixHD translates dots into a label map, and finally the label map is translated into an X-ray image.

3.4. Segmentation Multiscale Attention Network

After generating the label maps of the corresponding chest X-ray images, we use a semantic segmentation network to prove the effectiveness of the synthetic images during training, and to compare the three-stage approach with the one- and two-stage methods, proving its superior performance. In this paper, the Segmentation Multiscale Attention Network (SMANet) [18] was employed. The SMANet is composed of three main compo-

nents, a ResNet encoder, a multi-scale attention module, and a convolutional decoder (see Figure 4).

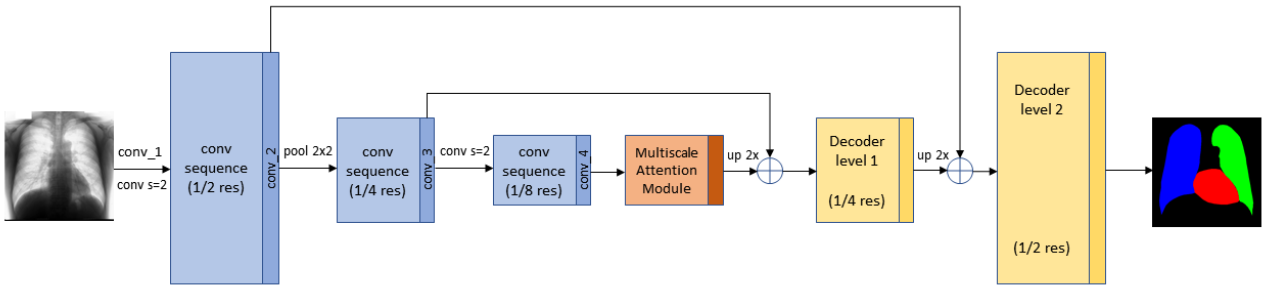


Figure 4. Scheme of the SMANet segmentation network.

This architecture, initially proposed for scene text segmentation, is based on the Pyramid Scene Parsing Network (PSPNet) [11], a deep fully convolutional neural network with a ResNet [64] encoder. Dilated convolutions (i.e. atrous convolutions [65]) are used in the ResNet backbone, to widen the receptive field of the neural network in order to avoid an excessive reduction of the spatial resolution due to down-sampling. The most characteristic part of the PSPNet architecture is the pyramid pooling module (PSP), which is employed to capture features of different scale in the image. In the SMANet, the PSP module is replaced with a multi-scale attention mechanism to better focus on the relevant objects present in the image. Finally, a two-level convolutional decoder is added to the architecture to improve the recognition of small objects.

3.5. Training Details

The PGGAN architecture, proposed in [13], was employed for image generation; the number of parameters were modified to speed up learning and reduce overfitting. More specifically, the maximum number of feature maps for each layer was reduced to 64. Furthermore, since the PGGAN was used to generate seeds and labels, obtaining only the semantic label maps in both cases, the output image has only one channel instead of three. The generation procedure (PGGAN and Pix2PixHD) was stopped by visually examining the generated samples during the training phase. The images, generated in the various steps for all the methods, have a resolution of 1024×1024, except in the case of the "dot" label maps, which, as mentioned before, are generated at a 64×64 resolution.

The SMANet is implemented in TensorFlow. Random crops of 377×377 pixels were employed during training, whereas a sliding window of the same size was used for testing. The Adam optimizer [66], based on an initial learning rate of 10^{-4} and a mini batch of 17 examples, was used to train the SMANet. All the experiments were carried out in a Linux environment on a single NVIDIA Tesla V100 SXM2 with 32 GB RAM. The SMANet's goal is to produce the semantic segmentation of the lungs and heart. The network is trained by a supervised approach, both in the case of real and synthetic images. In particular, for the images generated by the three different methods, we are able to use this approach thanks to the generation of both the images and the label maps.

4. Experiments and Results

In this section, after describing the dataset on which our new proposed method was tested, we evaluate the results obtained, both qualitatively—based on the judgment of three physicians—and quantitatively, comparing them with related approaches present in the literature.

4.1. Dataset

Chest X-ray images are available thanks to the Japanese Society of Radiological Technology (JSRT) [67]. The dataset they provide consists of 247 chest X-ray images. The res-

olution of the images is 2048 × 2048 pixels, with a spatial resolution of 0.175 mm/pixel and 12 bit gray levels. Furthermore, segmentation supervisions for the JSRT database are available in the Segmentation in the Chest Radiographs (SCR) dataset [6]. More precisely, this dataset provides chest X-ray supervisions which correspond with the pixel-level positions of the different anatomical parts. Such supervisions were produced by two observers who segmented five objects in each image: the two lungs, the heart and the two clavicles. The first observer was a medical student and his segmentation was used as the gold standard, while the second observer was a computer science student, specialized in medical imaging, and his segmentation was considered that of a human expert.

The SCR dataset comes with an official splitting, which is employed in this paper and consists of 124 images for learning and 123 for testing. We use two different experimental configurations. In the former, called FULL_DATASET, all the training images are exploited. More precisely, the PGGAN generation network is trained on the basis of 744 images, available in the SCR training set and obtained with the augmentation procedure described above. The SMANet is trained on 7500 synthetic images, generated by the PGGAN, and fine-tuned on the 744 images extracted from the SCR training set, while 2500 synthetic images are used for validation. For the second configuration, called TINY_DATASET, only 10% of the SCR training set is used and the PGGAN is trained on only 66 images (obtained both from SCR and with augmentation); furthermore, the SMANet is trained exactly as above, except for the fine-tuning, which is carried out on 66 images.

4.2. Quantitative Results

Generated images were employed to train a deep semantic segmentation network. The rationale behind the approach is that the performance of the network trained on the generated data reflects the data quality and variety. A good performance of the segmentation network indicates that the generated data successfully capture the true distribution of the real samples. To assess the segmentation results, some standard evaluation metrics were used. The Jaccard Index, J, also called Intersection Over Union (IOU), measures the similarity between two finite sample sets—the predicted segmentation and the target mask in this case—and is defined as the size of their intersection divided by the size of their union. For binary classification, the Jaccard index can be framed in the following formula:

$$J = \frac{TP}{TP + FP + FN}$$

where TP, FP and FN denote the number of true positives, false positives and false negatives, respectively. Furthermore, the Dice Score, DSC, is defined as:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$

DSC is a *quotient of similarity between sets* and ranges between 0 and 1.

The experiments can be divided into two phases: first, we evaluated the generation procedure described in Section 3.3 using the FULL_DATASET, then, we compared this approach with the other two methods described in Sections 3.1 and 3.2 using the TINY_DATASET. The purpose of this latter experiment was to evaluate whether multi-stage generation methods are actually more effective in producing data suitable for semantic segmentation with a limited amount of data. In particular, in the experimental setup based on the FULL_DATASET, for the three-stage method, the generation network was trained on all the SCR training images, to which the augmentation procedure described in Section 3 was applied. Then, 10,000 synthetic images were generated and used to train the semantic segmentation network. Moreover, we evaluated a fine-tuning of the network on the SCR real images after the pre-training on the generated images. The results, shown in Table 1, are compared with those obtained using only real images to train the semantic segmentation network, which can be considered as a baseline.

Table 1. Evaluation of the proposed methods based on the FULL_DATASET, using 2500 generated images for the validation set. Real corresponds to the results obtained using the official training set; *Synth 3* corresponds to the results obtained using only the generated images, while in the *Finetune* column, real data are employed for fine-tuning.

		Real	Three-Stage	
			Synth 3	Finetune
J	Left Lung	96.10	95.30	**96.22**
	Heart	90.78	87.25	**91.11**
	Right Lung	**96.85**	96.15	96.79
	Average	94.58	92.90	**94.71**
DSC	Left Lung	98.01	97.6	**98.07**
	Heart	95.17	93.19	**95.35**
	Right Lung	**98.40**	98.04	98.37
	Average	97.19	96.28	**97.26**

Next, the TINY_DATASET was used in order to evaluate the performance of the methods with a very small dataset. More precisely, the following experimental setups, the results of which are shown in Table 2, are considered:

- REAL—only real images are used for training the semantic segmentation network;
- SINGLE-STAGE—the segmentation network uses the images generated by the single-stage method (Synth 1 in the tables) for training while real images are employed for fine-tuning (Finetune in the tables);
- TWO-STAGES—the images generated with the two-stage method are used to pre-train the segmentation network (Synth 2) while real images are used for fine-tuning;
- THREE-STAGE—the images generated with the three-stage method are used for training the segmentation network (Synth 3), while real images are employed for fine-tuning.

In this case, the PGGAN was trained on 66 images, based on 11 images randomly chosen from the entire training set to which the augmentation described above was applied.

Table 2. Evaluation of the proposed methods based on the TINY_DATASET, using 2500 generated images for the validation set. **Real** corresponds to the results obtained using the official training set; *Synth 1*, *Synth 2*, *Synth 3*, correspond to the results obtained using only the generated images, while in the *Finetune* columns, real data are employed for fine-tuning.

		Real	Single-Stage		Two-Stage		Three-Stage	
			Synth 1	Finetune	Synth 2	Finetune	Synth 3	Finetune
J	Left Lung	93.70	55.59	74.11	94.91	94.4	94.96	**95.29**
	Heart	85.50	0.07	37.47	86.98	85.21	87.27	**87.47**
	Right Lung	93.70	52.78	79.99	95.90	95.44	95.90	**95.92**
	Average	90.97	36.15	63.86	92.60	91.68	92.71	**92.89**
DSC	Left Lung	96.75	71.46	85.13	97.39	97.12	97.42	**97.59**
	Heart	92.18	0.13	54.51	93.04	92.02	93.20	**93.32**
	Right Lung	96.74	69.09	88.89	97.91	97.66	97.90	**97.92**
	Average	95.22	46.89	76.18	96.11	95.60	96.17	**96.28**

In general, we can see that the best results are obtained with the three-stage method followed by fine-tuning. From Table 1, we observe a small improvement in results using a fine-tune on a network previously trained with images generated using the three-stage method. Therefore, the three-stage method provides good synthetic data, but the advantage given by generated images is low when the training set is large. Conversely, when few training images are available, in the TINY_DATASET setup, multi-stage methods outperform the baseline (column REAL of Table 2) and this happens even without fine-tuning. Thus, in this case, the advantage provided by synthetic images is evident. Moreover,

the three-stage method outperforms the two-stage approach, even with fine-tuning, which confirms our claim that splitting the generation procedure may provide a performance increase when few training images are available.

Finally, it is worth noting that fine-tuning improves the performance of the three-stage method, both in the FULL_DATASET and in the TINY_DATASET framework, which does not hold for the two-stage method. This behaviour may be explained by some complementary information that is captured from real images only with the three-stage method. Actually, we may argue that, in different phases of a multi-stage approach, different types of information can be captured: such a diversification seems to provide an advantage to the three-stage method, which develops some capability to model the data domain with more orthogonal information.

4.3. Comparison with Other Approaches

Table 3 shows our best results and the segmentation performance published by all recent methods, of which we are aware, on the SCR dataset. According to the results in the table, the three-stage method obtained the best performance score both for the lungs and the heart.

However, it is worth mentioning that Table 3 gives only a rough idea of the state-of-the-art, since a direct comparison between the proposed method and other approaches is not feasible, our primary focus being on image generation, in contrast with the comparative approaches that are mainly devoted to segmentation, and for which no results are reported on small image datasets. Moreover, the previous methods used different partitions of the SCR dataset to obtain the training and the test set, such as two-fold, three-fold, five-fold cross-validation or ad hoc splittings, which are often not publicly available, while, in our experiments, we preferred to use the original partition, provided with the SCR dataset (note that, compared to most of the other solutions used in comparative methods, the original subdivision has the disadvantage of producing a smaller training set, which is not in conflict, however, with the purpose of the present work). Finally, a variety of different image sizes have also been used, ranging from 256×256, to 400×400, and to 512×512—the resolution used in this work.

Table 3. Comparison of segmentation results among different methods on the SCR dataset (CV stands for cross-validation).

Method	Image Size	Augmentation	Evaluation Scheme	Lungs		Heart	
				DSC	J	DSC	J
Human expert [6]	2048×2048	No	-	-	94.6	-	87.8
U-Net [60]	256×256	No	5-fold CV	-	95.9	-	89.9
InvertedNet [58]	256×256	No	3-fold CV	97.4	95	93.7	88.2
SegNet [62]	256×256	No	5-fold CV	97.9	95.5	94.4	89.6
FCN [62]	256×256	No	5-fold CV	97.4	95	94.2	89.2
SCAN [58]	400×400	No	training/test split (209/38)	97.3	94.7	92.7	86.6
Our three-stage method	512×512	Yes	official split	**98.2**	**96.5**	**95.36**	**91.1**

4.4. Qualitative Results

In this section, some examples of images and corresponding segmentations, generated with the approaches described in Section 3, are qualitatively examined. We also report some comments from three physicians on the generated segmentations, to provide a medical assessment of the quality of our method.

Figures 5 and 6 display some examples—randomly chosen from all the generated images—of the label maps and the corresponding chest X-ray images generated with the three methods described in Section 3, using the FULL_DATASET and the TINY_DATASET, respectively. We can observe that, with the single and two-stage methods, the images tend to be more similar to those belonging to the training set. For example, in most of

the generated images there are white rectangles, which resemble those present in the training images, used to cover the names of both the patient and the hospital. Instead, the three-stage method does not produce such artifacts, suggesting that it is less prone to overfitting.

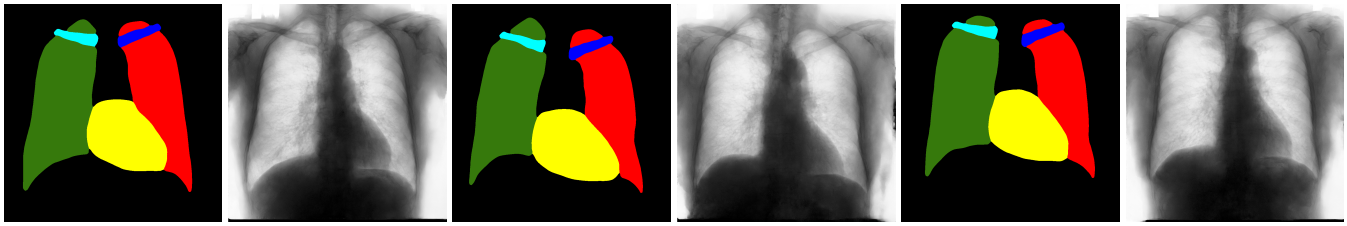
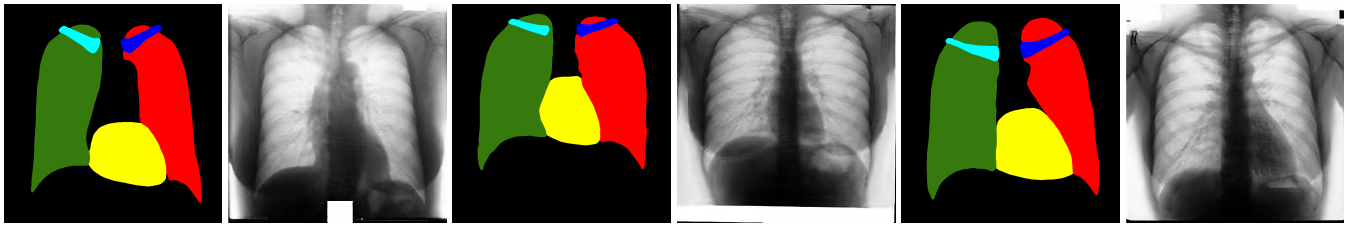
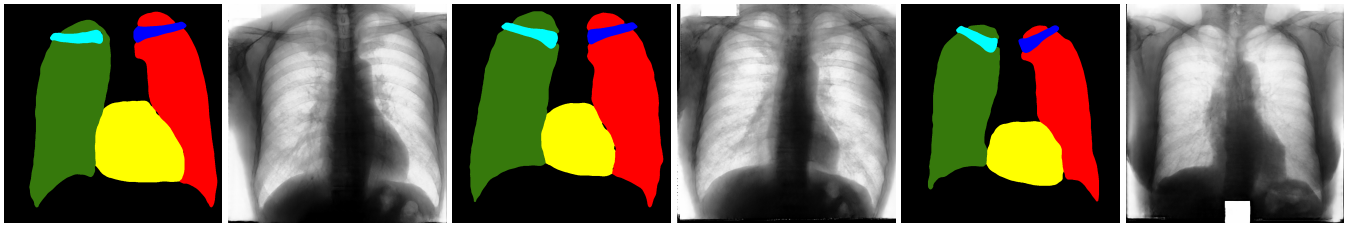


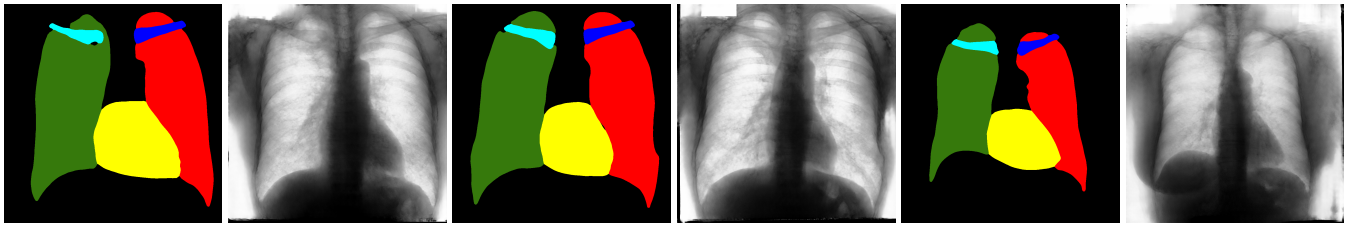
Figure 5. Examples three-stage generated images based on the FULL_DATASET.



(**a**)



(**b**)



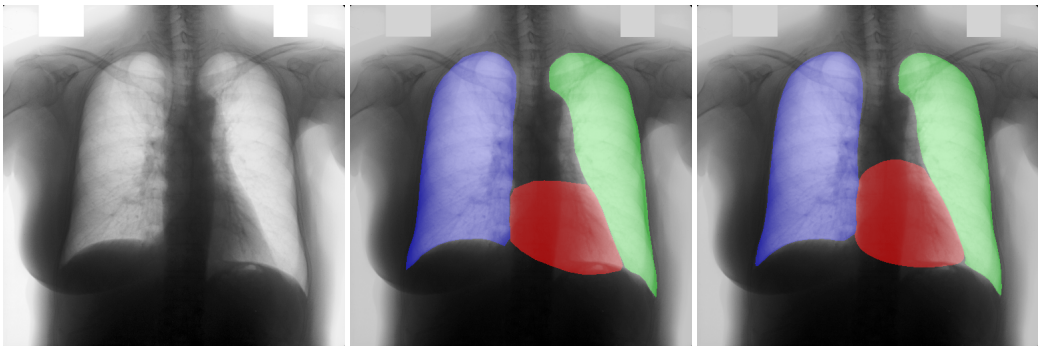
(**c**)

Figure 6. Examples of generated images based on the TINY_DATASET. (**a**) Single-stage 10% of generated images, (**b**) Two-stage 10% of generated images, (**c**) Three-stage 10% of generated images.

Moreover, in order to clarify the limits of the three-stage method, we assessed the quality of the segmentation results based on three human experts, who were asked to check 20 chest X-ray images, along with the corresponding supervision and the segmentation obtained by the SMANet network. Such images were chosen among those that can be considered difficult, at least based on the high error obtained by the segmentation algorithm. Figures 7 and 8 show different examples of the images evaluated by the experts. The first column represents the chest X-ray image, while the second and the third columns, the order of which was randomly exchanged during the presentation to the experts, represent the target segmentation and our prediction, respectively. The three physicians were asked to choose the best segmentation and to comment about their choice. Apart from a general agreement of all the doctors on the good quality of both the target segmentation and the segmentation provided by the three-stage method, surprisingly, they often chose the second

one. For the examples in Figure 7, for instance, all the experts shared the same opinion, preferring the segmentation obtained by the SMANet over the ground-truth segmentation. To report the results of the qualitative analysis, we numbered the target and predicted segmentation with numbers 1 and 2, respectively, while doctors were assigned unordered pairs to obtain an unbiased result. Then, with respect to Figure 7a, the comments reported by the experts were: (1) In segmentation 1, a fairly large part of the upper left ventricle is missing; (2) I choose the segmentation number 2 because the heart profile does not protrude to the left of the spine profile; (3) The best is number 2, the other leaves out a piece of the left free edge of the heart, in the cranial area. Furthermore, for Figure 7b, we obtained: (1) The second image is the best for the cardiac profile. For lung profiles, the second image is always better. The only flaw is that it leaks a bit on the right and left costophrenic sinuses. (2) Image 2 is the best, because the lower cardiac margin is lying down and does not protrude from the diaphragmatic dome. Image number 1 has a too flattened profile of the superior cardiac margin. (3) Number 2, for the cardiac profile is more faithful to the real contours.

(**a**)

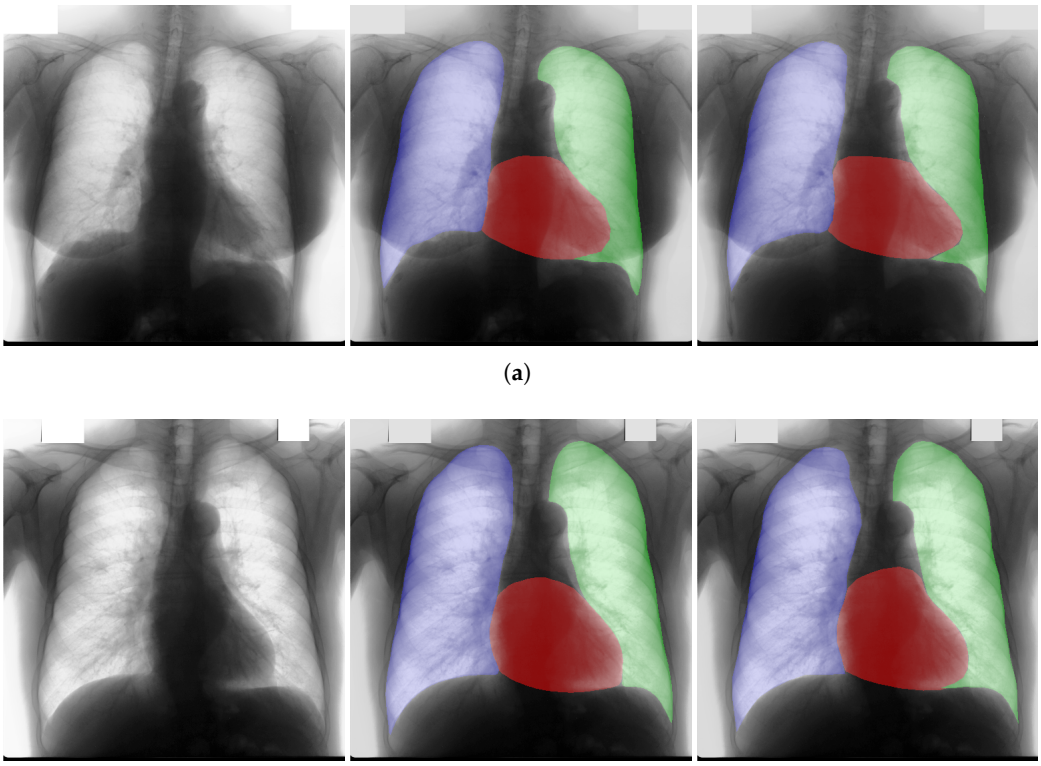


(**b**)

Figure 7. Examples of segmented images for which doctors shared the same opinion. The first column represents the chest X-ray image, while the second and third columns are the target and our predicted segmentation, respectively. (**a**) NODULES001, (**b**) NODULES066.

Furthermore, they reported conflicting opinions or decided not to give a preference with respect to the examples in Figure 8. When they agreed, they generally found different reasons for choosing one segmentation over the other. With respect to Figure 8a the comments reported by the experts were: (1) I prefer not to indicate any options because the heart image is completely subverted; (2) Segmentation number 2 is better, even if it is complicated to read because there is a "bottle-shaped" heart. The only thing that can be improved in image 2 is that a small portion of the right side of the heart is lost; (3) Number 1 respects more what could be the real contours of the heart image. Furthermore, for Figure 8b we obtained: (1) I prefer number 2 because the tip of the heart is well placed on the diaphragm and does not let us see that small wedge-shaped image that incorrectly insinuates itself between heart and diaphragm in image 1 and which has no correspondence

in the RX; (2) Both are good segmentations. Both have small problems, for example, in segmentation 1 a small portion of the tip (bottom right of the image) of the heart is missing, in segmentation 2 a part of the outflow cone (the "upper" part of the heart) is missing. It is difficult to choose, probably better number 1 because of the heart; (3) Number 2 because number 1 canal probably exceeds the real dimensions of the cardiac image, including part of the other mediastinal structures.



(**a**)

(**b**)

Figure 8. Examples of segmented images for which doctors gave conflicting opinions. The first column represents the chest X-ray image, while the second and third columns are the target and our predicted segmentations, respectively. (**a**) NODULES014, (**b**) NODULES015.

These different evaluations, albeit limited by the small number of examined images, confirm the difficulty of segmenting CXRs, a difficulty that is likely to be more evident in the case of the images selected for our quality analysis, which were chosen based on the large error produced by the segmentation algorithm.

5. Conclusions

In this paper, we have proposed a multi-stage method based on GANs to generate multi-organ segmentation of chest X-ray images. Unlike existing image generation algorithms, in the proposed approach, generation occurs in three stages, starting with "dots", which represent anatomical parts, and initially involves low-resolution images. After the first step, the resolution is increased to translate "dots" into label maps. We performed this step with Pix2PixHD, thus making the information grow and obtaining the labels for each anatomical part taken into consideration. Finally, Pix2PixHD is also used for translating the label maps into the corresponding chest X-ray images. The usefulness of our method was demonstrated especially when there were few images in the training set, an affordable problem thanks to the multi-stage nature of the approach.

It is worth observing that our method can be employed for any type of image, not exclusively medical ones, while synthetic and real images can concur in solving the segmentation problem (being used for pre-training and for fine-tuning the segmentation network, respectively), with a significant increase in performance. As a matter of future research,

the proposed approach will be extended to other, more complex domains, such as that of natural images.

Author Contributions: Conceptualization, G.C. and P.A.; methodology, G.C. and P.A.; software, G.C. and P.A.; validation, G.C., P.A., T.M., M.B. and F.S.; formal analysis, G.C. and P.A.; investigation, G.C.; resources, P.A., M.B. and F.S.; data curation, G.C.; writing—original draft preparation, G.C.; writing—review and editing, G.C., P.A., T.M., M.B. and F.S.; visualization, G.C., P.A., T.M., M.B. and F.S.; supervision, M.B. and F.S.; project administration, M.B. and F.S.; funding acquisition, G.C., M.B. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Università degli Studi di Firenze.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: http://db.jsrt.or.jp/eng.php.

Acknowledgments: In addition to Tommaso Mazzierli, who is one of the authors of this work, we would like to thank Gabriella Gaudino and Valentina Vellucci for their contribution in the analysis of the segmentations.

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

References

1. Mettler, F.A., Jr,; Huda, W.; Yoshizumi, T.T.; Mahesh, M. Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology* **2008**, *248*, 254–263.
2. Hussain, E.; Hasan, M.; Rahman, M.A.; Lee, I.; Tamanna, T.; Parvez, M.Z. CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos Solitons Fractals* **2021**, *142*, 110495.
3. Ismael, A.M.; Şengür, A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst. Appl.* **2021**, *164*, 114054.
4. Nayak, S.R.; Nayak, D.R.; Sinha, U.; Arora, V.; Pachori, R.B. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomed. Signal Process. Control* **2021**, *64*, 102365.
5. Bonechi, S.; Bianchini, M.; Bongini, P.; Ciano.; G., Giacomini, G.; Rosai, R.T.; Rossi, A.R.; Andreini, P. Fusion of Visual and Anamnestic Data for the Classification of Skin Lesions with Deep Learning. In *Lecture Notes in Computer Science*; Cristani, M., Prati, A., Lanz, O., Messelodi, S., Sebe, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11808, pp. 211–219.
6. Van Ginneken, B.; Stegmann, M.B.; Loog, M. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. *Med. Image Anal.* **2006**, *10*, 19–40.
7. Qin, C.; Yao, D.; Shi, Y.; Song, Z. Computer-aided detection in chest radiography based on artificial intelligence: A survey. *Biomed. Eng. Online* **2018**, *17*, 1–23.
8. Teixeira, L.O.; Pereira, R.M.; Bertolini, D.; Oliveira, L.S.; Nanni, L.; Cavalcanti, G.D.; Costa, Y.M. Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images. *arXiv* **2020**, arXiv:2009.09780.
9. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
10. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848.
11. 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.
12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
13. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
14. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8798–8807.
15. Vapnik, V.N. *Statistical Learning Theory*; Wiley-Interscience: Hoboken, NJ, USA, 1998.
16. Neyshabur, B.; Bhojanapalli, S.; Mcallester, D.; Srebro, N. Exploring Generalization in Deep Learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5947–5956.
17. Kawaguchi, K.; Kaelbling, L.P.; Bengio, Y. Generalization in Deep Learning. *arXiv* **2017**, arXiv:1710.05468.
18. Bonechi, S.; Bianchini, M.; Scarselli, F.; Andreini, P. Weak supervision for generating pixel–level annotations in scene text segmentation. *Pattern Recognit. Lett.* **2020**, *138*, 1–7.

19. Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F.; Sodi, A. A two stage gan for high resolution retinal image generation and segmentation. *arXiv* **2019**, arXiv:1907.12296.
20. Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F. Image generation by GAN and style transfer for agar plate image segmentation. *Comput. Methods Programs Biomed.* **2020**, *184*, 105268, doi:10.1016/j.cmpb.2019.105268.
21. Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F. A Deep Learning Approach to Bacterial Colony Segmentation. In *Artificial Neural Networks and Machine Learning—ICANN 2018*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 522–533.
22. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier gans. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 2642–2651.
23. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
24. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
25. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; others. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 4681–4690.
26. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
27. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576.
28. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. *arXiv* **2017**, arXiv:1703.00848.
29. Liu, M.Y.; Tuzel, O. Coupled generative adversarial networks. *arXiv* **2016**, arXiv:1606.07536.
30. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
31. 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.
32. 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.
33. Chen, Q.; Koltun, V. Photographic image synthesis with cascaded refinement networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1511–1520.
34. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
35. Zhu, J.Y.; Zhang, R.; Pathak, D.; Darrell, T.; Efros, A.A.; Wang, O.; Shechtman, E. Toward multimodal image-to-image translation. *arXiv* **2017**, arXiv:1711.11586.
36. Qi, X.; Chen, Q.; Jia, J.; Koltun, V. Semi-parametric image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8808–8816.
37. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 2337–2346.
38. Sun, L.; Wang, J.; Ding, X.; Huang, Y.; Paisley, J. An adversarial learning approach to medical image synthesis for lesion removal. *arXiv* **2018**, arXiv:1810.10850.
39. Chen, X.; Konukoglu, E. Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. *arXiv* **2018**, arXiv:1806.04972.
40. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *arXiv* **2017**, arXiv:1703.05921.
41. Zhang, X.; Jian, W.; Chen, Y.; Yang, S. Deform-GAN:An Unsupervised Learning Model for Deformable Registration. *arXiv* **2020**, arXiv:2002.11430.
42. Fan, J.; Cao, X.; Xue, Z.; Yap, P.; Shen, D. Adversarial Similarity Network for Evaluating Image Alignment in Deep Learning Based Registration. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018—21st International Conference, Granada, Spain, 16–20 September 2018; Lecture Notes in Computer Science; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Proceedings, Part I, Volume 11070, pp. 739–746, doi:10.1007/978-3-030-00928-1_83.
43. Tanner, C.; Ozdemir, F.; Profanter, R.; Vishnevsky, V.; Konukoglu, E.; Goksel, O. Generative Adversarial Networks for MR-CT Deformable Image Registration. *arXiv* **2018**, arXiv:1807.07349.
44. Yi, X.; Walia, E.; Babyn, P. Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by wasserstein distance for dermoscopy image classification. *arXiv* **2018**, arXiv:1804.03700.
45. Madani, A.; Moradi, M.; Karargyris, A.; Syeda-Mahmood, T. Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1038–1042.
46. Lecouat, B.; Chang, K.; Foo, C.S.; Unnikrishnan, B.; Brown, J.M.; Zenati, H.; Beers, A.; Chandrasekhar, V.; Kalpathy-Cramer, J.; Krishnaswamy, P. Semi-Supervised Deep Learning for Abnormality Classification in Retinal Images. *arXiv* **2018**, arXiv:1812.07832.

47. Li, Y.; Shen, L. cC-GAN: A robust transfer-learning framework for HEp-2 specimen image segmentation. *IEEE Access* **2018**, *6*, 14048–14058.
48. 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.
49. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331.
50. Hu, B.; Tang, Y.; Eric, I.; Chang, C.; Fan, Y.; Lai, M.; Xu, Y. Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1316–1328.
51. Srivastav, D.; Bajpai, A.; Srivastava, P. Improved Classification for Pneumonia Detection using Transfer Learning with GAN based Synthetic Image Augmentation. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 433–437.
52. Candemir, S.; Jaeger, S.; Palaniappan, K.; Musco, J.P.; Singh, R.K.; Xue, Z.; Karargyris, A.; Antani, S.; Thoma, G.; McDonald, C.J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans. Med Imaging* **2013**, *33*, 577–590.
53. Boykov, Y.; Funka-Lea, G. Graph cuts and efficient ND image segmentation. *Int. J. Comput. Vis.* **2006**, *70*, 109–131.
54. Candemir, S.; Akgül, Y.S. Statistical significance based graph cut regularization for medical image segmentation. *Turk. J. Electr. Eng. Comput. Sci.* **2011**, *19*, 957–972.
55. Boykov, Y.; Jolly, M. Interactive graph cuts for optimal boundary and region segmentation of objects in nd images. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; pp. 105–112.
56. Shao, Y.; Gao, Y.; Guo, Y.; Shi, Y.; Yang, X.; Shen, D. Hierarchical lung field segmentation with joint shape and appearance sparse learning. *IEEE Trans. Med Imaging* **2014**, *33*, 1761–1780.
57. Ibragimov, B.; Likar, B.; Pernuš, F.; Vrtovec, T. Accurate landmark-based segmentation by incorporating landmark misdetections. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1072–1075.
58. Novikov, A.A.; Lenis, D.; Major, D.; Hladůvka, J.; Wimmer, M.; Bühler, K. Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs. *IEEE Trans. Med Imaging* **2018**, *37*, 1865–1876, doi:10.1109/TMI.2018.2806086.
59. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
60. Wang, C. Segmentation of multiple structures in chest radiographs using multi-task fully convolutional networks. In Proceedings of the Scandinavian Conference on Image Analysis, Tromsø, Norway, 12–14 June 2017; pp. 282–289.
61. Oliveira, H.; dos Santos, J. Deep transfer learning for segmentation of anatomical structures in chest radiographs. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Paraná, Brazil, 29 October–1 November 2018; pp. 204–211.
62. Islam, J.; Zhang, Y. Towards robust lung segmentation in chest radiographs with deep learning. *arXiv* **2018**, arXiv:1811.12638.
63. Dai, W.; Dong, N.; Wang, Z.; Liang, X.; Zhang, H.; Xing, E.P. Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 263–273.
64. 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.
65. Papandreou, G.; Kokkinos, I.; Savalle, P.A. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv* **2014**, arXiv:1412.0296.
66. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
67. Shiraishi, J.; Katsuragawa, S.; Ikezoe, J.; Matsumoto, T.; Kobayashi, T.; Komatsu, K.i.; Matsui, M.; Fujita, H.; Kodera, Y.; Doi, K. Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *Am. J. Roentgenol.* **2000**, *174*, 71–74.