CT Image Synthesis Using Weakly Supervised Segmentation and Geometric Inter-Label Relations For COVID Image Analysis

Dwarikanath Mahapatra¹, Ankur Singh²
¹ Inception Institute of Artificial Intelligence, Abu Dhabi, UAE
² Indian Institute of Technology, Kanpur, India

Abstract—While medical image segmentation is an important task for computer aided diagnosis, the high expertise requirement for pixelwise manual annotations makes it a challenging and time consuming task. Since conventional data augmentations do not fully represent the underlying distribution of the training set, the trained models have varying performance when tested on images captured from different sources. Most prior work on image synthesis for data augmentation ignore the interleaved geometric relationship between different anatomical labels. We propose improvements over previous GAN-based medical image synthesis methods by learning the relationship between different anatomical labels. We use a weakly supervised segmentation method to obtain pixel level semantic label map of images which is used learn the intrinsic relationship of geometry and shape across semantic labels. Latent space variable sampling results in diverse generated images from a base image and improves robustness. We use the synthetic images from our method to train networks for segmenting COVID-19 infected areas from lung CT images. The proposed method outperforms state-of-the-art segmentation methods on a public dataset. Ablation studies also demonstrate benefits of integrating geometry and diversity.

I. INTRODUCTION

The novel Coronavirus Disease (COVID-19) pandemic has had a significant adverse impact on the global stage since the first reported cases in December 2019 [156], [27]. It has infected more than 4.5 million people resulting in more than 315,000 deaths across 210 countries. The gold standard for COVID-19 screening is the reverse-transcription polymerase chain reaction (RT-PCR) test. Equipment shortage and strict testing requirements limit rapid and accurate screening of the general populace, in addition to reports of RT-PCR testing exhibiting high false negative rates [28], [14], [93], [21], [8], [10], [68], [41], [102], [34]. Radiological imaging such as X-rays and computed tomography (CT) have emerged as a useful tool in early COVID-19 screening by achieving high sensitivity (with RT-PCR results as reference) [28], [60], [172], [128], [82], [58], [45] and demonstrating robustness in diagnosis, follow-up assessment, and evaluation of disease evolution [137], [57], [96], [72], [78], [97], [100].

Although X-rays can be quickly acquired, CT screening provides a richer 3D view of the lung better suited for diagnosis. Recent studies of [28], [163], [54], [53], [56], [52], [122], [120], [118] provide evidence that CT scans (Figure 1) can be used to identify COVID-19 biomarkers such as ground-glass opacity (GGO) in the early stage, and pulmonary consolidation in the late stage. Thus qualitative evaluation of infection and longitudinal changes in CT scans can provide useful and important information for detecting COVID-19.

Recent methods such as [148], [132], [70], [40], [15], [160], [92], [75], [25], [142], [154], [121], [9] have proposed deep learning (DL) systems to detect COVID-19 patients from CT/Xray. Wang et al. proposed COVID-Net to identify COVID-19 cases from chest x-rays [157], [143], [65], [144], [74], [136], [134], [153], [146], [165], [105], [143], [104], [171], [81], [80], [79], [101] proposed an anomaly detection approach for pathological image augmentation and demonstrated its efficacy in COVID pathological region segmentation. Figure 2 shows example cases of synthetic images generated using the approach.
by our method and other competing techniques.

Traditional augmentations such as image rotations or deformations have limited benefit as they do not fully represent the underlying data distribution of the training set and are sensitive to parameter choices. Recent data augmentation methods of [25], [31], [11], [13], [108], [10], [114], [129], [103], [69], [139], [39], [32], [33], [13], [73], [62], [63] use generative adversarial networks (GAN), [23], and show moderate success for medical image classification. However, they have limited relevance for segmentation since they do not model geometric relation between different organs and most augmentation approaches do not differentiate between normal and diseased samples. Hence there is a need for augmentation methods that consider the geometric relation between different anatomical regions and generate distinct images for diseased and normal cases.

II. RELATED WORK

A. Chest CT Segmentation

Segmentation of lungs from chest CT scans is a widely explored topic [150], [12], [61], [91], [67], [42], [22] since it facilitates diagnosis and quantification of lung diseases [24], [36], [110], [66], [59], [64], [3], [83], [84], [48] use support vector machines (SVM) to detect lung nodules from CT scans. Nodule extraction is challenging due to similar appearance with the background. Deep learning algorithms have been able to overcome this challenge by learning powerful discriminative features. [159], [83], [4], [87], [3], [89], [90] use CNNs to segment lung nodules from heterogeneous CT scans. [31], [86], [85], [107], [94], [20], [4], [111], [25], [135] make use of GAN-synthesized data to improve the performance of a discriminative model for pathological lung segmentation. [30] employ two deep networks to segment lung tumors from CT scans by adding multiple residual streams of varying resolutions.

B. Deep Learning For Imaging Based COVID-19 Analysis

[158] use a modified inception network of [152] for classifying COVID-19 patients and normal controls. Instead of directly training on complete CT images, they trained the network on the regions of interest, which are identified by two radiologists based on the features of pneumonia. [17] use CT images to train a U-Net++ [170] for identifying COVID-19 patients that performs comparably with expert radiologists. DL approaches have also been used for segmenting infection regions in lung CT [19] and for lung infection quantification [132], [10], [137] of COVID-19.

C. Data Augmentation (DA)

While conventional augmentation approaches (such as rotation, scaling, etc) can generate a large database, they do not add much data diversity. They are also sensitive to parameter values [18], variation in image resolution, appearance and quality [45]. Recent DL based methods trained with synthetic images outperform those trained with standard DA over classification and segmentation tasks. [11] proposed DAGAN for image generation in few shot learning systems. [166] proposed a learning-based registration method to register images to an atlas, use corresponding deformation field to deform a segmentation mask and obtain new data. [71] used conditional GAN (cGAN) for generating informative synthetic chest Xray images conditioned on a perturbed input mask.

GANs have also been used for generating synthetic retinal images in [167] and brain magnetic resonance images (MRI) in [25], [149], image registration [93] and generating higher strength MRI from their low strength acquisition counterparts [164]. Generated images have implicit variations in intensity distribution but there is no explicit attempt to model attributes such as shape variations that are important to capture different conditions across a population. [130] augmented medical images with simulated anatomical variations but demonstrate inconsistent performance based on transformation functions and parameter settings.

D. Image Generation Using Uncertainty

[35] used approximate Bayesian inference for parameter uncertainty estimation in scene understanding, but did not capture complex correlations between different labels. [44] proposed a method to generate different samples using an ensemble of M networks while [138] present a single network with M heads for image generation. [151] proposed a method based on conditional variational autoencoders (cVAE) to model segmentation masks, which improves the quality of generated images. In probabilistic UNet [58], cVAE is combined with UNet [133] to generate multiple segmentation masks, although with limited diversity since randomness is introduced at highest resolution only. [6] introduced a framework to generate images with a greater diversity by injecting randomness at multiple levels.

E. Our Contribution

Since annotating medical images is a time consuming task, it is challenging to obtain manually annotated segmentation masks to model the geometrical relation between different labels in the image. To overcome this challenge we propose to use a weakly supervised segmentation approach to generate labeled segmentation maps. The generated segmentation maps are then used to model the geometric relationship between the different pathological regions.

Based on the premise that improved data augmentation yields better segmentation performance in a DL system, we hypothesize that improved generation of synthetic images is possible by considering the intrinsic relationships between shape and geometry of anatomical structures [7]. In this paper we present a Geometry-Aware Shape Generative Adversarial Network (GeoGAN) that learns to generate plausible images of the desired anatomy (e.g., COVID infected areas in the lung).

A pre-print of a preliminary version of our method applied to fluid segmentation from retinal OCT scans can be found at https://arxiv.org/pdf/2003.14119.pdf. We introduce an additional weakly supervised segmentation step. Since the current submission is for a COVID special issue the results from the pre-print are not included.
while preserving learned relationships between geometry and shape. We make the following contributions:

1) Incorporating geometry information contributes to generation of realistic and qualitatively different medical images and shapes compared to standard DA. Other works such as [71], [167] do not incorporate this geometric relationship between anatomical parts.

2) Use of uncertainty sampling and conditional shape generation on class labels to introduce diversity in the mask generation process. Compared to previous methods we introduce diversity at different stages (different from [71], [167], [38]) and introduce an auxiliary classifier (different from [6], [151]) for improving the quality and accuracy of generated images.

III. METHOD

Our augmentation method: 1) models geometric relationship between multiple segmentation labels; 2) preserves disease class label of original image to learn disease specific appearance and shape characteristics; and 3) introduces diversity in the image generation process through uncertainty sampling. We demonstrate our method’s capability by training it on a dataset of CT lung images having annotations of COVID infected areas. However, in real world scenarios it is difficult to find datasets with such detailed annotations, especially in the case of COVID-19. Hence we introduce a weakly supervised segmentation (WSS) step that segments a CT image into different labeled regions using only the image labels of prevalent pathologies. The resulting label map enables us to learn the geometric relationship between different labels, which is essential to synthesize realistic images for data augmentation.

Figure 3 shows the training workflow using a modified UNet based generator network. The set of images and their WSS-obtained segmentation masks are used to train the generator while the discriminator provides feedback to improve the generator output. Figure 4 depicts generation of synthetic images after training is complete and their subsequent use in training a UNet for image segmentation at test time.

A. Weakly Supervised Segmentation

In order to obtain pixel labels from the image labels in a weakly supervised setting we solve a instance-level classification problem where pixels are instances. Subsequently, existing well-developed fully supervised segmentation methods can be applied. We use a combined Multiple Instance Learning (cMIL) for instance classification [161]. The image is split into \(N \times N\) grids (instances) of equal size where instances from the same image are in the same bag. In cMIL, two MIL-based classifiers with different instance selection criteria (\(\text{Max} - \text{Max}\) and \(\text{Max} - \text{Min}\)) are used to select instances to construct the instance-level dataset for subsequent classification.

The selected instance can be considered as the representative of its corresponding image, which determines the image class. If the image is labeled ‘infected’ \((I)\) we reason that at least one instance is infected. On the other hand, if the label of the image is ‘not infected’ \((NI)\), all the instances in it are \(NI\). For both \(I\) and \(NI\) images, \(\text{Max} - \text{Max}\) selects the instance with maximum \(I\) response. As shown in Figures 5(a) and (b), during the training stage the \(\text{Max} - \text{Max}\) criterion will select the instance with maximum \(I\) response as the \(NI\) example from the \(NI\) samples. Therefore, the model trained with these data would give a decision boundary biased towards \(I\) leading to misclassification of \(I\) instances with lower responses (as shown by light red circles). For example, \(I\) instances with similar appearances to \(NI\) may get misclassified.

\(\text{Max} - \text{Min}\) acts as a countermeasure that selects the instances with the highest \(I\) response for \(I\) images and the instances with the lowest response for \(NI\) images. As shown in Figure 5(c), \(\text{Max} - \text{Min}\) tends to have an opposite effect.
**Fig. 3.** Overview of the steps in the training stage of our method. The images (X) and corresponding segmentation masks (SX) are input to a STN whose output is fed to the generator network. Generator network is based on UNet architecture, and diversity through uncertainty sampling is injected at different levels. The generated mask $S'_{X}$ is fed to the discriminator which evaluates its accuracy based on $L_{\text{class}}$, $L_{\text{shape}}$ and $L_{\text{adv}}$. The provided feedback is used for weight updates to obtain the final model.

**Fig. 4.** Depiction of mask generation. The trained generator network is used on validation set base images and masks to generate new images that are used to train a segmentation network (UNet or Dense UNet). The model then segments infected regions from test images.

compared to $\text{Max} - \text{Max}$. Therefore, in cMIL we combine these two criteria to obtain a balanced instance-level dataset to be used in fully supervised learning (see Figure 5 (d)). It is worth noting that, for $NI$ images, although each instance is $NI$, we only use the selected instances to avoid data imbalance.

Using a ResNet-50 architecture we train the two MIL-based classifiers separately under the same configuration: in the forward pass, we use the $\text{Max} - \text{Max}$ (or $\text{Max} - \text{Min}$ for the other classifier) criterion to select one instance from each bag based on their predictions, and the prediction of the selected instance is regarded as the prediction of the image. In the backpropagation step, we use the cross entropy loss between the image-level label and the prediction of the selected instance to update the classifier’s parameters. The loss function for each classifier is defined as follows:

$$\text{Loss} = - \sum_{j} (y_{j} \log \hat{p}_{j} + (1 - y_{j}) \log(1 - \hat{p}_{j})),$$

where $\hat{p}_{j} = S_{\text{criterion}}(\{f(b_{i})\})$, $b_{i}$ are the instances in image $j$, $f$ is the classifier, $S_{\text{criterion}} \in \{\text{Max} - \text{Max}, \text{Max} - \text{Min}\}$.

$S_{\text{criterion}}$ selects the target instance using the defined criterion, $y_{j}$ is the image-level label.

For $\text{Max} - \text{Max}$ criterion:

$$S_{\text{Max} - \text{Max}}(\{f(b_{i})\}) = \max_{i} f(b_{i})$$

(2)

For $\text{Max} - \text{Min}$ criterion:

$$S_{\text{Max} - \text{Min}}(f(b_{i})) = \begin{cases} \max_{i} f(b_{i}) & \text{if } y = 1 \\ \min_{i} f(b_{i}) & \text{if } y = 0 \end{cases}$$

(3)

After training, we feed the same training data into the two trained classifiers and select the instances under the corresponding criterion, then the predictions are considered as their labels. We combine the instances selected by the two trained classifiers to construct the final fully supervised instance-level dataset. Note that we discard those potentially confusing samples whose predicted labels are different from their corresponding image-level labels.

a) Retrain and Relabel: Once the instance-level dataset is selected, we train an instance classifier in a fully supervised manner. Similar to cMIL we use a ResNet-50 and name this step as retrain. Then, we split the original image into latticed instances and relabel them using the trained instance-level classification model. For each image, we obtain $N^{2}$ high-quality instance labels from a single image-level label.

1) Segmentation: With enriched supervision information, the instance level labels are directly assigned to the corresponding pixels, producing approximate pixel-level labels. Therefore, we can train segmentation models in a fully supervised way using well-developed architectures such as UNet++ [170].

a) Training with Image-Level Constraints: In order to maximize the utility of the original image-level supervision
information, in the retrain step, we can add the original image-level data as one additional input source going through the classifier. The image-level constraint is imposed under Max-Max and Max-Min criteria to the instance level, the total loss is defined as the sum of the retrain loss and the constraint loss:

\[
\text{Loss} = w_1 \times \text{Loss}_{\text{constrain}} + w_2 \times \text{Loss}_{\text{retrain}}
\]

where \(w_1\) and \(w_2\) are the weights of the two losses. We set \(w_1 = w_2\) in our experiments.

\[
\text{Loss}_{\text{constrain}} = - \sum_{i} (y \log \hat{p} + (1 - y) \log(1 - \hat{p})),
\]

where \(\hat{p} = S_{\text{criterion}}(f(b_i))\), \(b_i\) represents the selected instance, \(f\) is the image-level constrain route, \(S_{\text{criterion}} \in \{\text{Max} - \text{Max}, \text{Max} - \text{Min}\}\), and \(y\) is the image-level label.

\[
\text{Loss}_{\text{retrain}} = - \sum_{i} (y_j \log \hat{y}_j + (1 - y_j) \log(1 - \hat{y}_j))
\]

where \(\hat{y}_j = g(n_j), n_j\) represents the input instance, \(g\) is the retrain route, and \(y_j\) is the instance-level label.

B. Geometry Aware Shape Generation

Let us denote an input image as \(x\), the corresponding segmentation masks as \(s_x\) and the disease class label of \(x\) as \(\ell_x\). Our method learns to generate a new image and segmentation label map from a base image and its corresponding mask. The first stage is a spatial transformer network (STN) that transforms the base mask to a new shape with different attributes of location, scale and orientation. The transformations used to obtain new segmentation mask \(y'_{x}\) are applied to \(x\) to get corresponding transformed image \(x'\). Since the primary aim of our approach is to learn contours and other shape specific information of anatomical regions, a modified UNet architecture as the generator network effectively captures hierarchical information of shapes. It also makes it easier to introduce diversity at different levels of image abstraction.

The generator \(G_g\) takes input \(s_x\) and a desired label vector of output mask \(c_g\) to output an affine transformation matrix \(A\) via a STN, i.e., \(G_g(s_x, c_g) = A\). \(A\) is used to generate \(s'_x\) and \(x'\). The discriminator \(D_g\) determines whether output image preserves the desired label \(c_g\) or not. The discriminator \(D_g\) is tasked with ensuring that the generated masks and images are realistic. Let the minimax criteria between \(G_g\) and \(D_g\) be \(\min_{c} \max_{D_g} \text{L}_g(G_g, D_g)\). The loss function \(\text{L}_g\) has three components:

\[
L_g = L_{\text{adv}} + \lambda_1 L_{\text{class}} + \lambda_2 L_{\text{shape}}
\]

where 1) \(L_{\text{adv}}\) is an adversarial loss to ensure \(G_g\) outputs realistic deformations; 2) \(L_{\text{class}}\) ensures generated image has characteristics of the target output class label (disease or normal); and 3) \(L_{\text{shape}}\) ensures new masks have realistic shapes. \(\lambda_1, \lambda_2\) balance each term’s contribution.

a) Adversarial Loss: - \(L_{\text{adv}}(G_g, D_g)\): The STN outputs \(\hat{A}\), a prediction for \(A\) conditioned on \(s_x\) and a new semantic map \(s_x \oplus \hat{A}(s_x)\) is generated. \(L_{\text{adv}}\) is defined as:

\[
L_{\text{adv}}(G_g, D_g) = \mathbb{E}_x \left[ \log D_g(s_x \oplus \hat{A}(s_x)) \right] + \mathbb{E}_{s_x} \left[ \log(1 - D_g(s_x \oplus \hat{A}(s_x))) \right],
\]

b) Classification Loss: - \(L_{\text{class}}\): The affine transformation \(A\) is applied to the base image \(x\) to obtain the generated image \(x'\). We add an auxiliary classifier when optimizing both \(G_g\) and \(D_g\) and define the classification loss as:

\[
L_{\text{class}} = \mathbb{E}_{s_x, c_g} \left[ - \log D_{\text{class}}(c_g|x') \right],
\]

where the term \(D_{\text{class}}(c_g|x')\) represents a probability distribution over classification labels computed by \(D\).

c) Shape Loss: - \(L_{\text{shape}}\): We intend to preserve the relative geometric arrangement between the different labels. The generated mask has regions with different assigned segmentation labels because the base mask (from which the image was generated) already has labeled layers. Let us denote by \(s_i\) the image region (or pixels) in \(s_x\) assigned label \(i\). Consider another set of pixels, \(s_j\), assigned label \(j\). We calculate \(P_{\text{shape}}(l_i|s_j, s_i)\), which is, given regions \(s_i, s_j\), the pairwise probability of \(s_i\) being label \(i\). If \(n\) denotes the total number of labels, for every label \(i\) we calculate the \(n - 1\) such probability values and repeat it for all \(n\) labels. Thus

\[
L_{\text{shape}} = \frac{1}{n \times (n - 1)} \sum_{i \neq j} P_{\text{shape}}(i, j) \in \{1, \cdots, n\}
\]

The probability value is determined from a pre-trained modified VGG16 architecture to compute \(L_{\text{shape}}\) where the input has two separate maps corresponding to the label pair. Each map’s foreground has only the region of the corresponding label and other labels considered background. The conditional probability between the pair of label maps enables the classifier to implicitly capture geometrical relationships and volume information between the label pair without the need to define explicit features. The geometric relation between different labels will vary for infected and non-infected cases, which
is effectively captured by our approach. To get the pre-trained VGG16 network we used a separate dataset of 24 images with its WSS generated segmentation maps.

C. Sample Diversity From Uncertainty Sampling

The generated mask $s'_g$ is obtained by fusing $L$ levels of the generator $G_y$ (as shown in Figure 3), each of which is associated with a latent variable $z_l$. We use probabilistic uncertainty sampling to model conditional distribution of segmentation masks and use separate latent variables at multi-resolutions to factor inherent uncertainties. The hierarchical approach introduces diversity at different stages and influences different features (e.g., low level features at the early layers and abstract features in the later layers). Denoting the generated mask as $s$ for simplicity, we obtain conditional distribution $p(s|x)$ for $L$ latent levels as:

\[
p(s|x) = \int p(s|z_1, \ldots, z_L)p(z_1|z_2, x) \ldots p(z_{L-1}|z_L, x)p(z_L|x)dz_1 \ldots dz_L.
\]

(11)

Latent variable $z_l$ models diversity at resolution $2^{-l+1}$ of the original image (e.g. $z_1$ and $z_3$ denote the original and $1/4$ image resolution). A variational approximation $q(z|x)$ approximates the posterior distribution $p(z|x)$ where $z = \{z_1, \ldots, z_L\}$. log $p(s|x) = L(s|x) + KL(q(z|x)||p(z|x))$, where $L$ is the evidence lower bound, and $KL(\cdot|\cdot)$ is the Kullback-Leibler divergence. The prior and posterior distributions are parameterized as normal distributions $N(z|\mu, \sigma)$.

Figure 3 shows example implementation for $L = 3$. We use 6 resolution levels and $L = 4$ latent levels. Figure 3 shows the latent variables $z_l$ forming skip connections in a UNet architecture such that information between the image and segmentation output goes through a sampling step. The latent variables are not mapped to a 1-D vector to preserve the structural relationship between them, and this substantially improves segmentation accuracy. $z_l$’s dimensionality is $r_x 2^{-l+1} \times r_y 2^{-l+1}$, where $r_x$, $r_y$ are image dimensions.

IV. EXPERIMENTAL RESULTS

A. Dataset Description

We use the following three different segmentation datasets:

CT Segmentation Dataset 1 (CTSeg1): The dataset consists of 100 axial CT images from different COVID-19 patients. All the CT images were collected by the Italian Society of Medical and Interventional Radiology. A radiologist segmented the CT images with 3 labels: ground-glass (mask value $=1$), consolidation ($=2$) and pleural effusion ($=3$).

CT Segmentation Dataset 2 (CTSeg2): The second dataset is collected from 9 axial volumetric CT scans. It includes whole volumes and both positive and negative slices (373 out of the total of 829 slices have been evaluated by a radiologist as positive and segmented).

CT Segmentation Dataset 3 (CTSeg3): This dataset contains 20 labeled COVID-19 CT scans. Left lung, right lung, and infections are labeled by two radiologists and verified by an experienced radiologist.

B. Experimental Setup, Baselines and Metrics

Our method has the following steps: 1) Use the default training, validation, and test folds of the dataset. 2) Use training images to train the image generator. 3) Generate shapes from the training set and train UNet++ segmentation network on the generated images. 4) Use trained UNet++ to segment test images. 5) Repeat the above steps for different data augmentation methods. Our model is implemented in PyTorch, on a NVIDIA TITAN X GPU. We trained all models using Adam optimiser with a learning rate of $10^{-3}$ and batch-size of 16. Batch-normalisation was used. The values of parameters $\lambda_1$ and $\lambda_2$ in Eqn. [7] were set by a detailed grid search on a separate dataset of 14 volumes that was not used for training or testing. We varied $\lambda_1$ between [0, 1] in steps of 0.05 by fixing $\lambda_2$ and varying $\lambda_2$ for the whole range. This was repeated for all values of $\lambda_1$. The best segmentation accuracy was obtained for $\lambda_1 = 0.92$ and $\lambda_2 = 0.9$, which were our final parameter values.

We denote our method as GeoGAN$_{WSS}$ (Geometry Aware GANs using the weakly supervised segmentation component), and compare it’s performance against other methods such as: 1) rotation, translation and scaling (denoted as DA-Data Augmentation); 2) DAGAN - data augmentation GANs of [1]; 3) cGAN - the conditional GAN based method of [71]; 4) Zhao and others atlas registration method of [166]; 5) GeoGAN$_{Manual}$ - Geometry Aware GANs using the manual segmentation maps for image synthesis. Segmentation performance is evaluated in terms of Dice Metric (DM), Hausdorff Distance (HD) and Mean Absolute error (MAE). DM of 1 indicates perfect overlap and 0 indicates no overlap, while lower values of MAE indicate better segmentation performance.

a) Algorithm Baselines. The following variants of our method were used for ablation studies:

1) GeoGAN$_{noL_{loc}}$ - GeoGAN$_{WSS}$ without classification loss (Eqn[9]).

2) GeoGAN$_{noL_{shape}}$ - GeoGAN$_{WSS}$ without shape relationship modeling term (Eqn[10]).

3) GeoGAN$_{noSamp}$ - GeoGAN$_{WSS}$ without uncertainty sampling for injecting diversity to determine sampling’s relevance to the final network performance.

C. Efficacy of Weakly Supervised Semantic Segmentation

To quantify the accuracy of the weakly supervised segmentation step we compare the segmentation output with the manual segmentations and obtain $DM = 0.926$, and $HD = 6.5$ mm. These numbers indicate very good agreement with the expert’s manual annotations. Figure 6 shows two examples of an image, its ground truth label map, and the label map obtained from the weakly supervised segmentation step. We observe that the WSS map closely resembles the manual map with small regions of oversegmentation. Oversegmented label maps recover the entire annotated area and hence enable GeoGAN to learn the full range of geometrical relations between different labels. On the other hand, oversegmented
maps also let the algorithm learn noisy features due to the fact that non-diseased regions are included as infections.

![CT image, manual ground truth segmentation label map, weakly supervised label map](image)

Fig. 6. Results for weakly supervised segmentation compared to ground truth maps. (a) CT image; (b) manual ground truth segmentation label map; (c) label map generated by our weakly supervised approach. Light blue is label 1 - ground glass opacity; Green is label 2 - consolidation; Yellow is label 3 - pleural effusion.

### D. Segmentation Results And Analysis

We hypothesize that a good image augmentation method should capture the different complex relationships between different labels, with the generated images leading to improvement in segmentation accuracy. Table I shows the average DM value of the best performing method. We also perform the segmentation analysis for different labels (through $L_{shape}$) much better than competing methods. Thus our attempt to model the intrinsic geometrical relationships between different labels could generate superior quality masks.

#### E. Ablation Studies

Table IV shows the segmentation results for different ablation studies. Figure 8 shows the segmentation mask obtained by different baselines for the same image shown in Figure 7 (a). The segmentation outputs are quite different from the ground truth and the one obtained by GeoGAN. In some cases the normal regions in the layers are included as pathological area, while parts of the infected region are not segmented with the pathological region. Either case is undesirable for disease diagnosis and quantification. Thus, different components of our cost functions are integral to the method’s performance and excluding one or more of classification loss, geometric loss and sampling loss adversely affects segmentation performance.

### F. Classification Results

Table V summarizes the performance of different methods on the public challenge dataset of consisting of 349 CT images labeled as being COVID-19 positive. These CT images have different sizes and come from 216 patients. All of them are resized to $512 \times 512$. We train GeoGAN$_{WSS}$ on the training set, generate more images, train classifiers on the training images and apply it on the test set. The leaderboard can be accessed here. For all test submissions Accuracy (ACC), F1 score (F1) and area under curve (AUC) are calculated, while the ranking is based on the F1 score. We use a DenseNet-121 architecture and employed GeoGAN$_{WSS}$ augmentation for the final results. Using conventional data augmentation we got the following values: F1=0.931, ACC=0.934, AUC=0.961, which would have placed us 6th in the current leaderboard. However, by using GeoGAN$_{WSS}$ our results are ranked third although we obtain the highest AUC and ACC values, while being very close to the top ranked method in terms of F1 score. For the completeness of the paper we would have liked to report the $p$-values at 95% confidence. However we are unable to do so since we do not have access to the results of other methods.

In a second set of classification experiments we used the CT${\text{Seg}}$1 dataset to generate augmentation images and train a classifier for detection COVID positive and negative cases. The classifier was used to classify images from the CT${\text{Seg}}$2 and CT${\text{Seg}}$3 datasets. Images from the two datasets were intensity normalized and combined into one dataset. We add COVID negative images from the challenge dataset to get an almost equal distribution of positive and negative cases in the training and test sets. The classification results on the test set are summarized in Table VI for DenseNet-121, and Table VII for ResNet-50.

An important component of our method is the WSS step which generates segmentation maps for a given image. For

---

3https://covid-ct.grand-challenge.org/Data/
4https://covid-ct.grand-challenge.org/Leaderboard/
Comparison approaches | Proposed
---|---
| DA | DAGAN | cGAN | Zhao | GeoGAN$_{WSS}$ | GeoGAN$_{S_{manual}}$ | [19]
| DM | 0.704 (0.17) | 0.719 (0.12) | 0.738 (0.09) | 0.752 (0.08) | 0.781 (0.05) | 0.789 (0.03) | 0.764 (0.02)
| p | 0.006 | 0.004 | 0.005 | 0.003 | - | 0.11 | -
| MAE | 0.097 (0.018) | 0.088 (0.015) | 0.083 (0.015) | 0.071 (0.017) | 0.058 (0.013) | 0.053 (0.012) | 0.064 (0.012)
| HD | 13.9 (4.1) | 12.7 (3.9) | 10.9 (3.8) | 9.1 (3.1) | 8.3 (2.4) | 8.2 (2.3) | -

**TABLE I**
COVID segmentation results for CTSeg1 dataset. Mean and standard deviation (in brackets) are shown. Best results per metric is shown in bold. p values are with respect to GeoGAN.

![Fig. 7](image.png)

Fig. 7. Segmentation results on the CTSeg1 dataset: (a) original test images; (b) manual segmentation masks. Masks generated using data generated by: (c) GeoGAN; (d) Zhao [166]; (e) DAGAN; (f) cGAN. The two rows correspond to two different images.

| Comparison approaches | Proposed |
|---|---|
| | DA | DAGAN | cGAN | Zhao | GeoGAN$_{WSS}$ | GeoGAN$_{S_{manual}}$
| DM | 0.708 (0.15) | 0.727 (0.13) | 0.761 (0.10) | 0.774 (0.09) | **0.809 (0.07)** | - |
| HD | 14.2 (4.4) | 12.4 (3.7) | 11.2 (3.2) | 9.4 (3.0) | **8.7 (2.7)** | - |
| MAE | 0.112 (0.011) | 0.094 (0.009) | 0.088 (0.010) | 0.079 (0.008) | **0.073 (0.006)** | - |
| p | 0.0009 | 0.0035 | 0.0003 | 0.01 | - | -

**TABLE II**
COVID segmentation results for CTSeg2 dataset. Mean and standard deviation (in brackets) are shown. Best results per metric is shown in bold. p values are with respect to GeoGAN.

![Fig. 8](image.png)

![Fig. 8](image.png)

![Fig. 8](image.png)

Fig. 8. Segmentation results for ablation experiments on CTSeg1: (a) GeoGAN$_{noLshape}$; (b) GeoGAN$_{noLcls}$; (c) GeoGAN$_{noSamp}$. The two rows correspond to the images shown in the two rows of Figure 7(a).

V. CONCLUSION

We propose a novel approach to generate plausible COVID-19 CT images by incorporating relationship between segmentation labels to guide the shape generation process. Diversity is introduced in the image generation process through uncertainty sampling. Comparative results show that the augmented dataset from GeoGAN outperforms standard data augmentation and other competing methods, when applied to segmentation of COVID-19 affected pathological regions in CT images.
images. We show that synergy between shape, classification and sampling terms lead to improved segmentation and each of these terms is equally important in generating realistic shapes.

Despite the good performance of our method we observe failure cases when the base images are noisy due to inherent characteristics of the image acquisition procedure. Our method is also useful to generate realistic images for educating clinicians, where targeted synthetic images (e.g. generation of complex cases, or disease mimickers) can be used to speed-up training.

Table IV

| Method   | NoL | NoSamp |
|----------|-----|--------|
| ACC      | 0.939 | 0.953 |
| F1       | 0.967 | 0.964 |
| AUC      | 0.965 | 0.987 |

Table V

| Method   | Rank 1 | Rank 2 | Rank 3 | Rank 4 | Rank 5 |
|----------|--------|--------|--------|--------|--------|
| GeoGAN   | 0.987  | 0.984  | 0.945  | 0.935  | 0.945  |
| GeoGAN   | 0.965  | 0.964  | 0.952  | 0.935  | 0.943  |

Fig. 9. Generated images for ablation study methods: (a) GeoGAN_{noL,cls}; (b) GeoGAN_{noL,shape}; (c) GeoGAN_{noSamp}. The corresponding generated images for other methods are shown in Figure 2.

References

[1] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. In arXiv preprint arXiv:1711.04450, 2017.
[2] P.R. Bastide, I.F. Kiral-Kornek, D. Mahapatra, S. Saha, A. Vishwanath, and S. Von Cavallar. Machine learned optimizing of health activity for participants during meeting times. In US Patent App. 15/426,634, 2018.
[3] P.R. Bastide, I.F. Kiral-Kornek, D. Mahapatra, S. Saha, A. Vishwanath, and S. Von Cavallar. Visual health maintenance and improvement. In US Patent 9,993,853, 2018.
[4] B. Bozorgtabar, D. Mahapatra, H. von Teng, A. Pollinger, Lucas Ebner, Jean-Philippe Thiran, and Mauricio Reyes. Informative sample generation using class aware generative adversarial networks for classification of chest x-rays. Computer Vision and Image Understanding, 184:57–65, 2019.
[5] B. Bozorgtabar, D. Mahapatra, H. von Teng, A. Pollinger, L. Ebner, J-P. Thiran, and M. Reyes. Informative sample generation using class aware generative adversarial networks for classification of chest x-rays. Computer Vision and Image Understanding, 184:57–65, 2019.
[6] B. Bozorgtabar, D. Mahapatra, G. Vray, and J-P. Thiran. Anomaly detection on-rays using self-supervised aggregation learning. In arXiv preprint arXiv:2010.09856, 2020.
[7] B. Bozorgtabar, D. Mahapatra, I. Zlobec, T.T. Rau, and J-P. Thiran. Computational pathology. Frontiers in Medicine, 7, 2020.
[8] B. Bozorgtabar, M. Saced Rad, D. Mahapatra, and J-P. Thiran. Syndemic: Synergistic deep feature alignment for joint learning of depth and ego-motion. In In Proc. IEEE ICCV, 2019.
[9] S. Chaganti, A. Balachandran, and et al. Quantification of tomographic patterns associated with covid-19 from chest ct. In arXiv, 2020.
[10] J. Chen, L. Wu, and et al. Deep learning-based method for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. In medRxiv, 2020.
[11] Alexey Dosovitskiy, Philipp Fischer, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with exemplar convolutional neural networks. IEEE Trans. Patt. Anal. Mach. Intell., 38(9):1734–1747, 2016.
[12] H. Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct scans. arXiv preprint arXiv:2004.14133, 2020.
[13] R. Garnavi, D. Mahapatra, PK Roy, and RB Tennakoon. System and method to teach and evaluate image grading performance using prior learned expert knowledge base. In US Patent App. 10,657,838, 2020.
[14] G. Vray and J-P. Thiran, 42(4):395–426, 2015.
[15] Z. Ge, D. Mahapatra, X. Chang, Z. Chen, L. Chi, and H. Lu. Improving multi-label chest x-ray disease diagnosis by exploiting disease and health labels dependencies. In press Multimedia Tools and Applications, pages 1–14, 2019.
[16] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[17] J. de Hoog, D Mahapatra, R Garnavi, and F. Jalali. Personalized monitoring of injury rehabilitation through mobile device imaging. In Proc. IEEE CVPR, pages 2672–2680, 2014.
[18] Z. Ge, D. Mahapatra, X. Chang, Z. Chen, L. Chi, and H. Lu. Improving multi-label chest x-ray disease diagnosis by exploiting disease and health labels dependencies. In press Multimedia Tools and Applications, pages 1–14, 2019.
[19] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[20] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[21] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[22] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[23] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[24] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[25] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[26] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[27] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[28] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[29] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
[30] Z. Ge, D. Mahapatra, S. Sedai, R. Garnavi, and R. Chakravorty. Chest x-rays classification: A multi-label and fine-grained problem. In arXiv preprint arXiv:1807.07247, 2018.
feature streams for automatic lung tumor segmentation from ct images. IEEE Trans. Med. Imag., 38(1):134–144, 2018.

[31] D. Jin, Z. Xu, Y. Yang, A. P. Harrison, and D. J. Mollura. Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation. In Proc. MICCAI, pages 732–740, 2018.

[32] L. Ju, X. Wang, L. Wang, T. Liu, X. Zhao, T. Drummond, D. Mahapatra, and Z. Ge. Relational subsets knowledge distillation for long-tailed retinal diseases recognition. In arXiv preprint arXiv:2104.11057, 2021.

[33] L. Ju, X. Wang, L. Wang, D. Mahapatra, X. Zhao, M. Harandi, T. Drummond, Tongliang Liu, and Zongyuan Ge. Improving medical image classification with label noise using dual-uncertainty estimation. In arXiv preprint arXiv:2103.00528, 2020.

[34] Lie Ju, Xin Wang, Xin Zhao, Dwarikanath Mahapatra, Paul Bonnington, and Zongyuan Ge. Synergic adversarial label learning for grading retinal diseases via knowledge distillation and multi-task learning. IEEE JBIH, 100:1–14, 2020.

[35] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian segmentation and recognition using svm classifier and active contour modeling: A complete intelligent system. Computers in Biology and Medicine, 43(4):287–300, 2013.

[36] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In arXiv preprint arXiv:1412.6980, 2014.

[37] Simon A. A. Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus H. Maier-Hein, S. M. Ali Elslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. In Proc. NIPS, pages 6965–6975, 2018.

[38] S. Kuanar, V. Athitsos, D. Mahapatra, and A. Rajan. Multi-scale deep learning architecture for nucleus detection in renal cell carcinoma microscopy image. In arXiv preprint arXiv:2104.13557, 2021.

[39] S. Kuanar, V Athitsos, D. Mahapatra, K.R. Rao, Z. Akhtar, and D. Dasgupta. Low dose abdominal ct image reconstruction: An unsupervised learning based approach. In Proc. IEEE ICIP, pages 1351–1355, 2019.

[40] Shiba Kuanar, Dwarikanath Mahapatra, Monalisa Bilas, and KR Rao. Multi-path dilated convolution network for haze and glow removal in night time images. The Visual Computer, pages 1–14, 2021.

[41] S. Kuanar, K.R. Rao, D. Mahapatra, and M. Bilas. Night time haze and glow removal using dilated convolutional network. In arXiv preprint arXiv:1902.00855, 2019.

[42] H. Kuang, B. Guthier, M. Saini, D. Mahapatra, and A. El Saddik. A real-time smart assistant for video surveillance through handheld devices. In Proc. ACM Int. Conf. Multimedia, pages 917–920, 2014.

[43] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In Proc. NIPS, pages 6402–6413, 2017.

[44] Kelvin K. Leung, Matthew J. Clarkson, Johnathon W. Bartlett, Shona Clegg, Clifford R. Jack Jr, Michael W. Weiner, Nick C. Fox, Sebastian Oursein, and A. D. N. Initiative. Robust atrophy rate measurement in alzheimer’s disease using multi-site serial mri: tissue-specific intensity normalization and parameter selection. Neuroimage, 50(2):516–523, 2010.

[45] Z. Li, D. Mahapatra, J.Tielbeek, J. Stoker, L. van Vliet, and F.M. Vos. Image registration based on autocorrelation of local structure. IEEE Trans. Med. Imaging, 35(1):63–75, 2016.

[46] D. Mahapatra. Neonatal brain mri skull stripping using graph cuts and shape priors. In In Proc: MICCAI workshop on Image Analysis of Human Brain Development (IAHBD), 2011.

[47] Dwarikanath Mahapatra. Registration and segmentation methodology for perfusion mr images: Application to cardiac and renal images. -, pages --, 2011.

[48] D. Mahapatra. Cardiac lv and rv segmentation using mutual context information. In Proc. MICCAI-MLMI, pages 201–209, 2012.

[49] D. Mahapatra. Groupwise registration of dynamic cardiac perfusion images using temporal information and segmentation information. In In Proc: SPIE Medical Imaging, 2012.

[50] D. Mahapatra. Landmark detection in cardiac mri using learned local image statistics. In Proc. MICCAI-Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges (STACOM), pages 115–124, 2012.

[51] D. Mahapatra. Skull stripping of neonatal brain mri: Using prior shape information from random forests. J. Digit. Imaging., 27(6):794–804, 2014.

[52] D. Mahapatra. Cardiac image segmentation from cine cardiac mri using graph cuts and shape priors. J. Digit. Imaging., 26(4):721–730, 2013.

[53] D. Mahapatra. Cardiac mri segmentation using mutual context information from left and right ventricle. J. Digit. Imaging., 26(5):898–908, 2013.

[54] D. Mahapatra. Graph cut based automatic prostate segmentation using learned semantic information. In Proc. IEEE ISBI, pages 1304–1307, 2013.

[55] D. Mahapatra. Joint segmentation and groupwise registration of cardiac perfusion images using temporal information. J. Digit. Imaging., 26(6):173–182, 2013.

[56] D. Mahapatra. Automatic cardiac segmentation using semantic information from random forests. J. Digit. Imaging., 27(6):794–804, 2014.

[57] D. Mahapatra. Combining multiple expert annotations using semi-supervised learning and graph cuts for medical image segmentation. Computer Vision and Image Understanding, 151(1):114–123, 2016.

[58] D. Mahapatra. Consensus based medical image segmentation using semi-supervised learning and graph cuts. In arXiv preprint

---

**TABLE VI**

| GeoGAN | Dagan | cGAN | GeoGAN_{wClass} | GeoGAN_{wShape} | GeoGAN_{wSamp} |
|--------|-------|-----|-----------------|-----------------|---------------|
| Spe    | 0.931 | 0.916 | 0.893 | 0.881 | 0.873 | 0.869 | 0.877 |
| Sen    | 0.942 | 0.923 | 0.906 | 0.890 | 0.884 | 0.882 | 0.890 |
| Acc    | 0.938 | 0.920 | 0.902 | 0.887 | 0.881 | 0.877 | 0.886 |
| AUC    | 0.967 | 0.946 | 0.928 | 0.917 | 0.912 | 0.904 | 0.911 |

**TABLE VII**

| GeoGAN | Dagan | cGAN | GeoGAN_{wClass} | GeoGAN_{wShape} | GeoGAN_{wSamp} |
|--------|-------|-----|-----------------|-----------------|---------------|
| Spe    | 0.915 | 0.901 | 0.864 | 0.857 | 0.849 | 0.842 | 0.850 |
| Sen    | 0.924 | 0.907 | 0.872 | 0.864 | 0.855 | 0.851 | 0.858 |
| Acc    | 0.921 | 0.906 | 0.869 | 0.862 | 0.853 | 0.848 | 0.853 |
| AUC    | 0.943 | 0.927 | 0.887 | 0.883 | 0.876 | 0.870 | 0.877 |
D. Mahapatra, P. Schüffler, J. Tieltbeek, F.M. Vos, and J.M. Buhmann. Semi-supervised and active learning for automatic segmentation of crohn’s disease. In Proc. MICCAI, Part 2, pages 214–221, 2013.

D. Mahapatra, S. Sedai, and R. Garnavi. Elastic registration of medical pages 625–628, 2015.

D. Mahapatra, J. Tieltbeek, F.M. Vos, and J.M. Buhmann. A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images., In In Proc. IEEE ISBI, pages 1078–1082, 2017.

G. D. Rubin, L. B. Haramati, and et al. The role of chest imaging in patient management during the covid-19 pandemic: A multinational consensus statement from the fleischner society. Radiology, 201365, 2020.

Christian Rupprecht, Iro Laina, Robert DiPietro, Maximilian Baust, Federico Tombari, Nassir Navab, and Gregory D. Hager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In Proc. CVPR, pages 3591–3600, 2017.

F. M. Vos, J. Tieltbeek, R. Naziroglu, Z. Li, P. Schüffler, D. Mahapatra, Alexander Wiebel, C. Lavini, J. Buhmann, H. Hege, J. Stoker, and L. van Vliet. Computational modeling for assessment of IBD: to be or not to be? In Proc. IEEE EMBC, pages 3974–3977, 2012.

Federico Tombari, Nassir Navab, and Gregory D. Hager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In Proc. CVPR, pages 3591–3600, 2017.

Christian Rupprecht, Iro Laina, Robert DiPietro, Maximilian Baust, Federico Tombari, Nassir Navab, and Gregory D. Hager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In Proc. CVPR, pages 3591–3600, 2017.

F. M. Vos, J. Tieltbeek, R. Naziroglu, Z. Li, P. Schüffler, D. Mahapatra, Alexander Wiebel, C. Lavini, J. Buhmann, H. Hege, J. Stoker, and L. van Vliet. Computational modeling for assessment of IBD: to be or not to be? In Proc. IEEE EMBC, pages 3974–3977, 2012.

F. M. Vos, J. Tieltbeek, R. Naziroglu, Z. Li, P. Schüffler, D. Mahapatra, Alexander Wiebel, C. Lavini, J. Buhmann, H. Hege, J. Stoker, and L. van Vliet. Computational modeling for assessment of IBD: to be or not to be? In Proc. IEEE EMBC, pages 3974–3977, 2012.

F. M. Vos, J. Tieltbeek, R. Naziroglu, Z. Li, P. Schüffler, D. Mahapatra, Alexander Wiebel, C. Lavini, J. Buhmann, H. Hege, J. Stoker, and L. van Vliet. Computational modeling for assessment of IBD: to be or not to be? In Proc. IEEE EMBC, pages 3974–3977, 2012.
[157] L. Wang and A. Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. [https://arxiv.org/abs/2003.09871] Online; accessed 15 May 2020.

[158] S. Wang, B. Kang, and et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). In medRxiv, 2020.

[159] S. Wang, M. Zhou, and et. al. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. Medical Image Analysis, 40(1):172–183, 2017.

[160] Y. Xing, Z. Ge, R. Zeng, D. Mahapatra, J. Seah, M. Law, and T. Drummond. Adversarial pulmonary pathology translation for pairwise chest x-ray data augmentation. In In Proc. MICCAI, pages 757–765, 2019.

[161] G. Xu, Z. Song, Z. Sun, C. Ku, Z. Yang, C. Liu, S. Wang, J. Ma, and W. Xu. CAMEL: A weakly supervised learning framework for histopathology image segmentation. In Proc. ICCV, pages 10682–10691, 2019.

[162] X. Xu, X. Jiang, and et. al. Deep learning system to screen coronavirus disease 2019 pneumonia. In arXiv, 2020.

[163] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song. Chest ct manifestations of new coronavirus disease 2019 (covid-19): a pictorial review. European Radiology, 2019(37):1–9, 2020.

[164] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. Med. Imag. Anal., 58, 2019.

[165] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xiao. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. In arXiv, 2020.

[166] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V. Guttag, and Adrian V. Dalca. Data augmentation using learned transforms for one-shot medical image segmentation. In In Proc. CVPR, pages 8543–8552, 2019.

[167] He Zhao, Haiqi Li, Sebastian Maurer-Stroh, and LiCheng. Synthesizing retinal and neuronal images with generative adversarial nets. Med. Imag. Anal, 49:14–26, 2018.

[168] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. arXiv preprint arXiv:2003.13865, 2020.

[169] C. Zheng, X. Deng, and et. al. Deep learning-based detection for covid-19 from chest ct using weak label. In medRxiv, 2020.

[170] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imag., pages 1–10, 2019.

[171] J. Zilly, J. Buhmann, and D. Mahapatra. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In In Proc. MLMI, pages 136–143, 2015.

[172] J. Zilly, J.M. Buhmann, and D. Mahapatra. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. In Press Computedized Medical Imaging and Graphics, 55(1):28–41, 2017.