Journal of Digital Imaging (2023) 36:1376–1389
https://doi.org/10.1007/s10278-023-00811-2

CCS-GAN: COVID-19 CT Scan Generation and Classification with Very Few Positive Training Images

Sumeet Menon[1] · Jayalakshmi Mangalagiri[1] · Josh Galita[1] · Michael Morris[1,2,3,4,5] · Babak Saboury[1,2,5] · Yaacov Yesha[1] · Yelena Yesha[1,2] · Phuong Nguyen[1] · Aryya Gangopadhyay[1] · David Chapman[1]

Received: 24 May 2022 / Revised: 2 March 2023 / Accepted: 3 March 2023 / Published online: 17 April 2023

Abstract

We present a novel algorithm that is able to generate deep synthetic COVID-19 pneumonia CT scan slices using a very small sample of positive training images in tandem with a larger number of normal images. This generative algorithm produces images of sufficient accuracy to enable a DNN classifier to achieve high classification accuracy using as few as 10 positive training slices (from 10 positive cases), which to the best of our knowledge is one order of magnitude fewer than the next closest published work at the time of writing. Deep learning with extremely small positive training volumes is a very difficult problem and has been an important topic during the COVID-19 pandemic, because for quite some time it was difficult to obtain large volumes of COVID-19-positive images for training. Algorithms that can learn to screen for diseases using few examples are an important area of research. Furthermore, algorithms to produce deep synthetic images with smaller data volumes have the added benefit of reducing the barriers of data sharing between healthcare institutions. We present the cycle-consistent segmentation-generative adversarial network (CCS-GAN). CCS-GAN combines style transfer with pulmonary segmentation and relevant transfer learning from negative images in order to create a larger volume of synthetic positive images for the purposes of improving diagnostic classification performance. The performance of a VGG-19 classifier plus CCS-GAN was trained using a small sample of positive image slices ranging from at most 50 down to as few as 10 COVID-19-positive CT scan images. CCS-GAN achieves high accuracy with few positive images and thereby greatly reduces the barrier of acquiring large training volumes in order to train a diagnostic classifier for COVID-19.

Keywords COVID-19 · CT · CCS-GAN · Pulmonary segmentation · Synthetic data

Introduction

Although deep learning algorithms have achieved high performance in cross-validated diagnostic tasks including screening for COVID-19 from X-ray and CT modalities, these results have been obtained overwhelmingly using extensive volumes of training data, including large volumes of COVID-19-positive images obtained from many cases. The use of large volumes of positive cases for training, however, is problematic, especially for a novel disease, as there may be a substantial lag between when the disease becomes a major public health concern and when large training datasets become publicly available, especially given HIPAA and IRB considerations [1, 2]. This has led to a great deal of research in order to determine how a deep learning algorithm can best screen for a disease when few positive training samples are available [3–6]. Successful deep learning–based COVID-19 diagnostic classification from CT scans has been demonstrated using hundreds of positive cases for training [3–6], but it is desirable to demonstrate that accurate classification is possible using even fewer training cases, especially as a preventative measure for a potential future pandemic for which adequate training

✉ Sumeet Menon
sumeet1@umbc.edu

[1] University of Maryland, 1000 Hilltop Circle, 21250 Baltimore, MD, USA

[2] Institute for Data Science and Computing, University of Miami, 33124 Coral Gables, FL, USA

[3] University of Miami Miller School of Medicine, Miami, FL, USA

[4] Networking Health, Oak Manor Drive, Suite 201, 21061 Glen Burnie, MD, USA

[5] National Institutes of Health Clinical Center, 9000 Rockville Pike, Building 10, Room 1C455, Bethesda, MD, USA

examples may once again be difficult to obtain. As such, throughout this paper, an assumption is made that negative (normal) images are prevalent and accessible, whereas positive (COVID-19) images are rare and/or difficult to obtain. Under such an assumption, reasonable metrics for the evaluation of an AI-based screening algorithm including accuracy and AUC are obtainable using a given number of positive images, even if larger volumes of negative images may be used freely as necessary for training of a deep neural network in order to achieve adequate classification performance. To the best of our knowledge, we present the first deep learning–based COVID-19 classification algorithm capable of achieving a cross-validated test accuracy of 99.00% (VGG19) and 98.17% (AlexNet) while using only 10 COVID-19-positive CT scan cases for training. Furthermore, the algorithm presented obtains these results using only a single CT scan slice per positive training case. These results are obtained by using a cycle-consistent segmentation-generative adversarial network (CCS-GAN) that is designed to generate high-quality COVID-19-infected pulmonary images using as few positive training examples as possible. CCS-GAN incorporates style transfer based on CycleGAN, automated intensity-based pulmonary segmentation, and transfer learning from relevant CT scan images, all of which greatly reduce the need for positive samples and allow the method to learn from a highly skewed training dataset with many more negative samples. An ablation study included in this work demonstrates that each of these underlying techniques alone is insufficient to achieve the presented results, and rather, all these techniques must be combined as part of the CCS-GAN methodology in order for high classification accuracy to be possible using very few positive cases for training.

Literature Review

Generative Methods for COVID-19 Pneumonia Screening

There have been a number of related works that have investigated the use of GANs to improve the performance of COVID-19 screening from CT scans with reduced training volumes. However, to the best of our knowledge, all recent studies have made use of hundreds of positive cases for training [3–8]. Hundreds of cases, although small by deep learning standards, are still a substantial training volume to obtain during a pandemic, and the intent of CCS-GAN is to determine how advanced methods may be able to greatly reduce the number of positive images needed for potential future events. Loey et al. introduced the use of conditional GANs (cGAN) for the generation of deep synthetic COVID-19 CT scans [3]. Their cGAN methodology generates both normal and COVID-19 images by conditioning using the category label (COVID/non-COVID). Goel et al. developed a similar approach making use of the InceptionV3 network with whale optimization for hyperparameter tuning [4]. Li et al. [7] extended these approaches by combining GANs with ensemble learning and attention mechanisms. Mangalagiri et al. also proposed an algorithm for generating 3D diagnostic quality COVID-19 CT scans with a conditional GAN architecture [8]. This method mainly focused on generating the entire CT volume through subdivision into blocks and focusing on blockwise synthesis rather than slice-wise synthesis. All of these methods have demonstrated improved performance of a binary classifier with a limited number of positive cases for training. However, all of these methods still require hundreds of positive cases or more to achieve their reported classification accuracies over a withheld testing set. We are unaware of any works prior to CCS-GAN that have been able to demonstrate comparable diagnostic classification accuracy using on the order of 10 positive cases for training. Several recent works that have looked at the use of style transfer as a foundation for deep fake CT image synthesis. Similar to CCS-GAN, most of these approaches have made use of the cycle-consistent loss approach from CycleGAN as a backbone approach. However, none of these works has combined this approach with automated pulmonary segmentation, and as such, these approaches may be susceptible to artifacts in regions unrelated to COVID-19 disease presentation []. Sandfort et al. show that CycleGAN can be used to generate synthetic CT images by learning the transformation of contrast to non-contrast CT [9]. Ghassemi et al. [10] also show the use of CycleGANs for improved COVID-19 classification with transfer learning, but they have used a total of 3163 images, which is far more than the proposed approach. Jin et al. incorporates prior training on a similar dataset in order to regularize the model and improve performance, especially with limited training volumes [11]. Liu et al. [6] generates full CT scan volumes using feature in-painting to insert COVID-19 opacities via alpha-opacity blending. Transferring disease presentation as a style is an active area of research, but more work is necessary to demonstrate that style transfer is feasible for the CT scan modality using a small number of positive images, as well as to prevent style transfer from introducing artifacts in background regions that are not affected by the target disease.

Very few works have looked at the possibility of using pulmonary segmentation in order to greatly simplify the problem of generating high-quality CT scan slices of COVID-19 pneumonia, and instead most have focused on attempting to generate the full CT scan slices including irrelevant anatomy. We demonstrate that pulmonary segmentation, however, can have a major positive impact on the performance of GAN techniques, especially when training

with few examples, because when training data volumes are small, the GAN could be distracted by learning to generate bones, organs, and other non-lung anatomy which are irrelevant to COVID-19 at the expense of decreased ability to learn relevant pulmonary features.

Although Jiang et al. [5] was the first to make use of pulmonary segmentation to generate synthetic COVID-19 CT scan slices using a dual generator/discriminator with dynamic element-wise sum, their segmentation approach requires annotation of the infected region of the lung which may be tedious for a radiologist to annotate for training data purposes [5]. We are unaware of a comparable GAN approach that takes advantage of automated pulmonary segmentation without the need for pixelwise annotated training data. Although there have been very few works to take advantage of pulmonary segmentation for the purpose of deep fake COVID-19 synthesis, there are several recent works that have explored segmentation toward the identification of lesions. However, we do not believe these works to be directly relevant to the proposed research aims because lesion segmentation (CADe) is a very different task from diagnostic classification (CADx). Deep convolutional GAN (DCGAN) [12] and conditional-GAN [13] were used to augment medical CT images of liver lesions and mammograms, yielding improved CNN-based classification accuracy of malignancy [14, 15]. Z. Xu et al. proposed GASNet in [16] which is a 3D segmentation framework containing a segmentation network with an embedded GAN to segment the pixel boundaries of COVID-19 opacities in CT scans. Although substantial progress has been made for the purposes of lesion segmentation, more work is necessary to determine the extent to which segmentation can improve the performance of GAN-based deep fake image synthesis at low training volumes.

Relation to Few-Shot Image Classification

It is important to discuss how the task of COVID-19 image synthesis with few positive cases is related to a branch of machine learning known as *few-shot image classification* [17–20]. *Few-shot* image classification is the task of constructing an *N-shot K-way* classifier using very few examples (small N) of the *target dataset*. Usually, other *relevant* datasets are available for *meta-learning* of a low-dimensional *feature embedding* [17–20]. One of the highly influential *metric-based* few-shot classifiers is the Siamese neural network [18] for which two identical network branches are pre-conditioned to learn a *low-dimensional embedding* suitable for a pairwise *cosine similarity metric* to distinguish between the image classes [18]. *Matching networks* improve on *metric-based* image classification, by combining a *low-dimensional embedding* network with a differentiable *k-nearest neighbor* distance metric to enable end-to-end gradient optimization for the entire problem [19]. *Prototypical networks* further expand upon *matching networks* by restricting the distance metric to compare against a single *M-dimensional* cluster *centroid* or *prototype* per image category within the *low-dimensional* embedding space [20]. Although CCS-GAN is designed to reduce the number of positive samples necessary for COVID-19 image classification, the CCS-GAN technique greatly differs in its design from these *metric-based* algorithms, because rather than learning a *low-dimensional embedding* for a distance-metric classifier, CCS-GAN instead learns a manifold in order to generate a large number of *deep synthetic* images to enable a standard DNN classifier to achieve adequate performance.

Our task also somewhat differs from *few-shot image classification*, because of the prevalence of substantial *class imbalance* which does not adhere to the standard *N-shot K-way* problem definition. More specifically, we assume a large number of *negative* (normal) cases but very few *positive* (covid) cases. The distinction is that both the *normal* (large-N) and *covid* (small-N) samples are part of the *target* categories (for binary classification), as opposed to the *N-shot K-way* problem definition which assumes small-N for both *target* categories simultaneously. The prevalence of substantial *class imbalance* is one of the motivating factors for the use of a GAN to generate additional data to produce a *balanced dataset* of *normal* and *covid* image slices to train an accurate image classifier model which in our case is a standard DNN.

Significance

In this study, we demonstrate that it is possible to generate *deep synthetic* CT scan slices of COVID-19 pneumonia by using only 10 positive examples. In contrast, previous studies have required on the order of 100 positive examples or more to produce adequate results for an external DNN classifier. The lack of available COVID-19 images was a major hurdle at the onset of the pandemic due to the inherent difficulties in acquiring and sharing medical image data. *Deep synthetic* image generation is considered to be one of the most promising techniques to overcome the data-sharing bottleneck not just for COVID-19 pneumonia, but many other diseases. The smaller the minimum adequate sample size, the more likely it is for a single healthcare institution to have sufficient cases available to generate deep synthetic datasets that are easier to share with other research groups. As such CCS-GAN demonstrates that *deep synthetic* images for COVID-19 can be generated with an order of magnitude fewer examples than previous studies.

Contributions

CCS-GAN is the first architecture for COVID-19 pneumonia image synthesis that can produce images of high quality for binary classification with only 10 positive training samples. CCS-GAN is data-efficient because it integrates style transfer with pulmonary segmentation thereby allowing the model to focus exclusively on generating the COVID-19 pneumonia features, rather than attempting to generate the entire CT scan slice including irrelevant anatomy. The architecture of CCS-GAN simplifies the learning task of generating CT-slices for COVID-19 pneumonia by reducing the need to generate irrelevant anatomy including anatomy outside the lung such as bone and organs, as well as anatomy within the lung such as blood vessels. This simplification of the learning task thereby allows the model to focus exclusively on generating relevant COVID-19 pneumonia features such as ground glass opacities. We summarize the contributions of CCS-GAN as follows:

1. We present the design of CCS-GAN which can generate deep synthetic COVID-19 pneumonia using an order of magnitude fewer positive CT slices relative to prior works. The synthetic images are demonstrated to be of adequate quality for a DNN classifier to achieve high accuracy on a COVID-19 pneumonia binary image classification task.
2. The design of CCS-GAN is data-efficient because it simplifies the learning task for deep synthetic COVID-19 image generation by eliminating irrelevant anatomical features both within the lung as well as outside of the lung. This irrelevant anatomy would otherwise greatly increase the amount of training data required to generate high-quality synthetic images.
3. We find in our ablation experiment that the key features of CCS-GAN include *pulmonary segmentation* which eliminates irrelevant anatomical features outside of the lung, as well as *cycle consistent loss* which eliminates irrelevant anatomical features within the lung from the learning process. If either of these key features are disabled, the quality of the generated images substantially degrades and the classifier degrades to random guessing.
4. Our quantitative and qualitative comparison shows that CCS-GAN is able to achieve high-quality synthetic COVID-19 images with far fewer training images relative to COVID-CT-GAN [7] which is a state of the art method, as well as baseline methods ResNet-50 GAN, U-Net GAN, and CycleGAN.

Limitations

We summarize the key limitations of this study as follows:

1. Although the CCS-GAN-generated deep synthetic images are of adequate quality for training a DNN classifier, these images are not intended to be of adequate quality to be used for clinical purposes by human radiologists. This is because the generated CT slices are easily distinguishable from real COVID-19 slices by a human, even though they are of adequate quality to augment a training sample for binary classification.
2. CCS-GAN operates on individual CT scan slices rather than whole images, which has both advantages and limitations. One notable limitation is that there is no guarantee that running CCS-GAN on consecutive slices will produce a coherent 3D volume, as it is likely that consecutive slices could generate pneumonia features in different locations in the lung leading to discontinuities between consecutive slices.
3. We have not evaluated the performance of CCS-GAN for binary classification of diseases other than COVID-19 pneumonia with CT-slices. Although we anticipate that CCS-GAN may be applicable to other similar pulmonary diseases using the CT modality, the use of intensity-based segmentation and style transfer may limit the applicability of CCS-GAN to other modalities and/or diseases with substantial differences in appearance.

Methods

CCS-GAN

Figure 1 describes the CCS-GAN approach. The input dataset is defined as a tuple (X_N, X_C) where X_N is the set of normal images and X_C is the set of images exhibiting COVID-19 pneumonia infection. Both the normal and COVID-19-infected images are segmented using binary K-means/OTSU thresholding to extract the pulmonary regions by creation of a binary segmentation mask. This segmentation mask can be extracted due to the large intensity difference in radiodensities between lung (less dense) and tissue (denser), even when the lung region is potentially affected by ground glass opacities due to pneumonia. After this, the generator models are pre-trained using a cycle-consistent pre-training procedure in which the normal images are employed for both the X and Y categories. Subsequently, the CycleGAN component is trained using unmatched pairs of normal images X_N and COVID-19 images X_C. Finally, the dataset for use with the

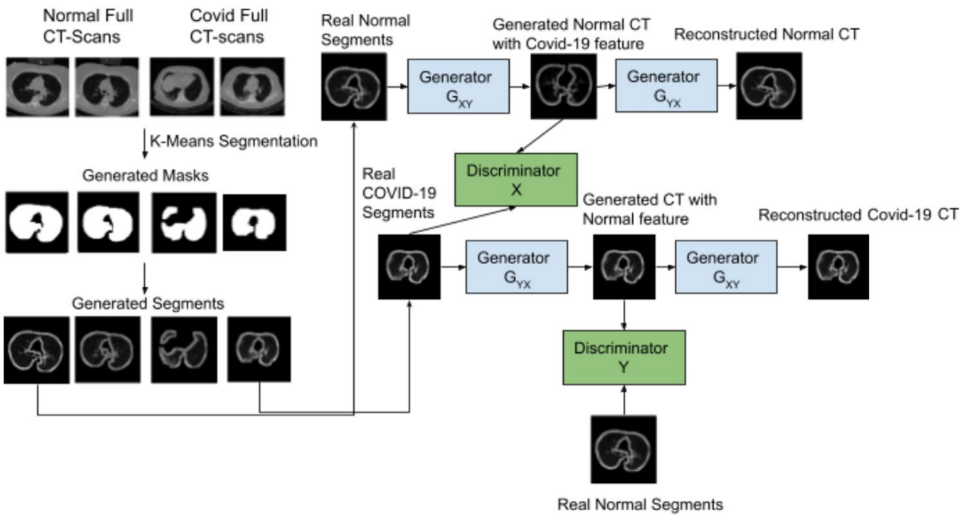

Fig. 1 CCS-GAN approach

classifier is augmented through the generation of deep-fake COVID-19-infected images through style transfer from additional normal images. This multi-faceted approach minimizes the number of positive real cases necessary for training either the GAN or the classifier and allows the methodology to achieve high classification accuracy in the presence of extremely high class imbalance between positive and negative examples. The training set only exhibits between 10 and 50 positive images, with up to 2000 normal images Figs. 2 and 3.

Intensity-Based Pulmonary Segmentation

Intensity-based pulmonary segmentation is an effective training-free way to extract the lung region. Intensity-based segmentation is effective due to the large differences in radiodensity between the lung region, which is mostly air, and the denser surrounding tissue. In order to extract the lung region for purposes of pulmonary segmentation, binary K-means thresholding was performed. K-means attempts to minimize the intra-cluster variance as follows:

$$J = \sum_{j=1}^{k} \sum_{1=1}^{n} \parallel x_i^j - c_j \parallel^2 \quad (1)$$

In this equation, J is the objective function, k is the number of clusters, n is the number of cases, x_i^j is the ith case, c_j is the centroid for cluster j, and $||x_i^{(j)} - c_j||^2$ is the distance function. The special case of k-means clustering with only two cluster centers is mathematically equivalent to the Otsu thresholding technique for binary image segmentation, which minimizes the intra-class variance between pixel intensity histograms over the image as follows:

$$\sigma_w^2(t) = q_1(t)\,\sigma_1^2(t) + q_2(t)\,\sigma_w^2(t) \quad (2)$$

where μ is the mean of the pixel, σ is the standard deviation, q is the sum of the probabilities, and t is the threshold ranging from the minimum value of the pixel to the highest value of pixels. Subsequent to thresholding, a fixed series of erosion and dilation steps are performed to suppress noise particularly over blood vessels in the lung. Air outside the patient is also removed, and the lung mask is identified due to its central positioning within the image frame. All pixels outside of the lung region are zeroed out to prevent non-lung anatomy tissue from contaminating the GAN and classifier training.

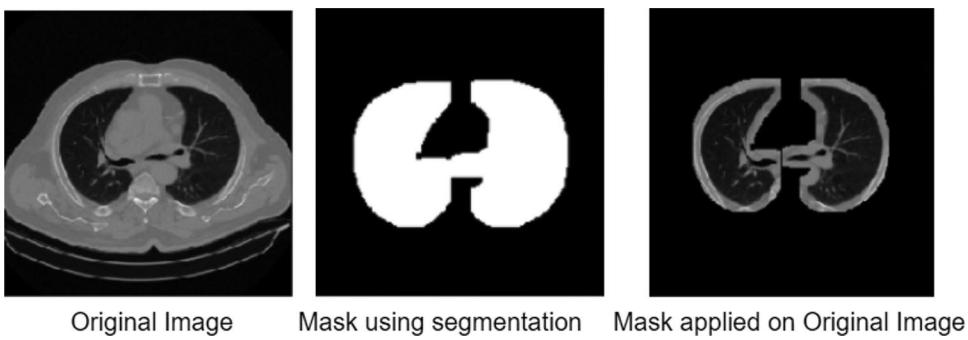

Fig. 2 Pulmonary segmentation

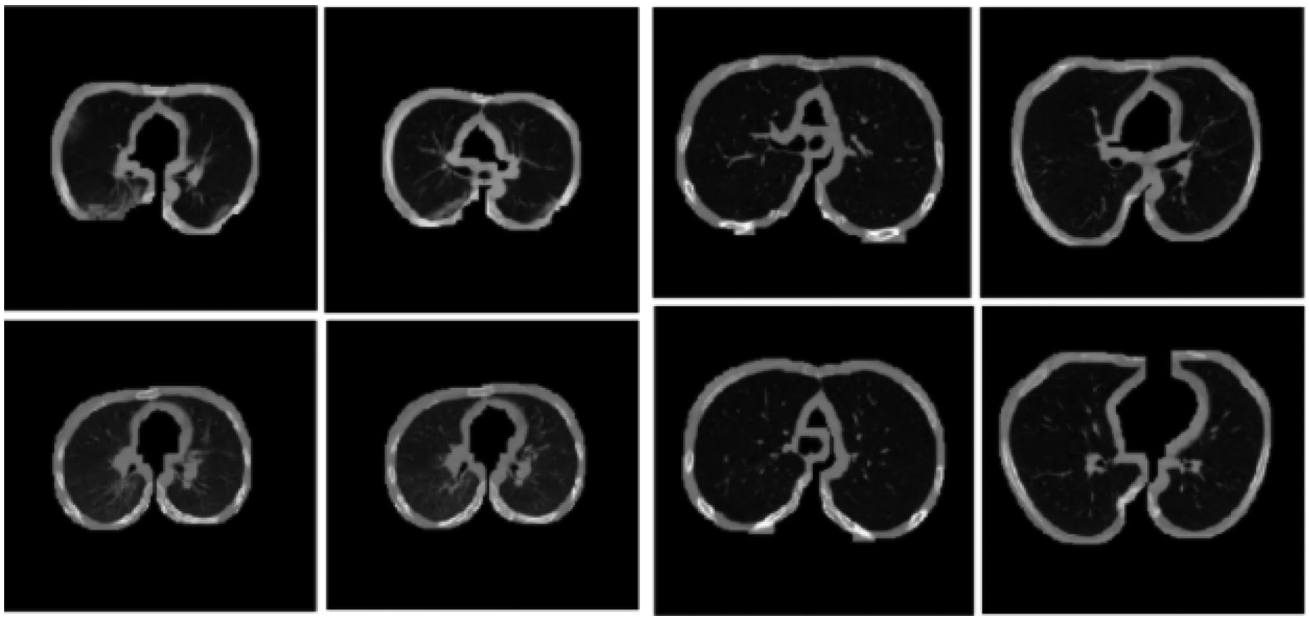

Fig. 3 Segmented full CT-slices with Covid-19 Pneumonia (left) versus Normal (right)

Adversarial and Cycle Consistent Loss

CCS-GAN makes use of combined adversarial and cycle-consistent loss functions as first introduced by Cycle-GAN [21]. The purpose of the cycle-consistent loss is that one should be able to apply style transfer from normal to COVID-19 and vice versa. This is accomplished by having two generators G and F, and two discriminators D_X and D_Y. Each generator and discriminator makes use of the min/max loss as proposed by Goodfellow as follows [21, 22]:

$$L_{GAN}(G, D_Y, X, Y) = E_{Y \to Pdata(Y)} [\log D_Y(Y)] + E_{X \to P\, data(X)} [1 - \log D_Y G(X)] \quad (3)$$

$$L_{GAN}(F, D_X, Y, X) = E_{X \to P\, data(X)} [\log D_X(X)] + E_{Y \to P\, data(Y)} [1 - \log D_X G(Y)] \quad (4)$$

The cycle-consistent loss ensures that if both generators are applied in a row, then the resulting image should be indistinguishable from the original image; i.e., for any images:$x \in X$, and $y \in Y$, that $F(G(x)) \approx x$, and that $F(G(y)) \approx y$ as follows [21]:

$$L_{cyc}(G, F) = E_{X \to P\, data(X)} |F(G(X)) - X| + E_{Y \to P\, data(Y)} |G(F(Y)) - Y| \quad (5)$$

The overall loss function accounts for all the losses is the sum of the constituent adversarial and cycle consistent loss functions as follows:

$$L_{GAN}(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + L_{cyc}(G, F) \quad (6)$$

Generator Architecture

Figure 4 describes the generator architecture of CCS-GAN, which is inspired by U-Net [23]. This architecture is arranged as downsampling blocks (yellow), upsampling blocks (red), and finalizing blocks (green). The first layer is a conv2D layer which accepts as input a CT scan slice of size (256,256,1) and outputs features of shape (128,128,1). The next layer is a batch normalization layer, followed by a Leaky ReLU activation. These three layers comprise each of the downsampling blocks as seen in Fig. 4 Sect. 1 (yellow). Furthermore, the Leaky ReLU activation layer has 2 output branches: a downsampling connection and a skip connection. The skip connection connects the output of a downsampling block of Sect. 1 (yellow) directly to the input of an upsampling block of Sect. 2 (red). Each upsampling block consists of a series of conv2d, transpose, and concatenation layers. Section 3 (green) exhibits finalizing blocks of alternating conv2D and transpose layers in order to generate a synthetic output image of size (256,256,1).

Figure 5 compares real COVID-19 images versus deep synthetic images as generated by CCS-GAN as trained using a varying number of positive cases. Left shows real positive cases. Center left shows generated images as trained with 10

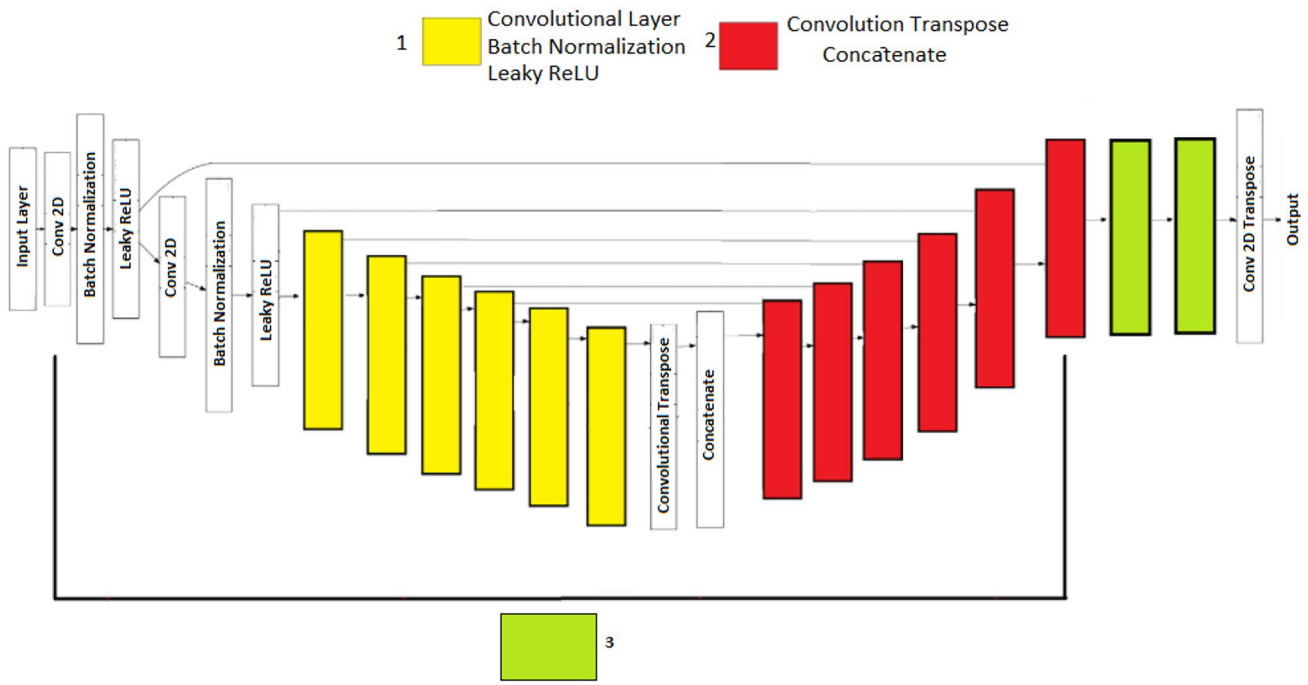

Fig. 4 Generator architecture

positive cases. Center right shows generated images as trained with 50 positive cases. Right shows generated images as trained with 500 positive cases. As expected, a general improvement in image quality is observed as the number of COVID-19-positive examples is increased, showing detail of blood vessels. Nevertheless, many pulmonary features are observable using very few positive cases, including those images when CCS-GAN is trained using 50 cases or even 10 cases.

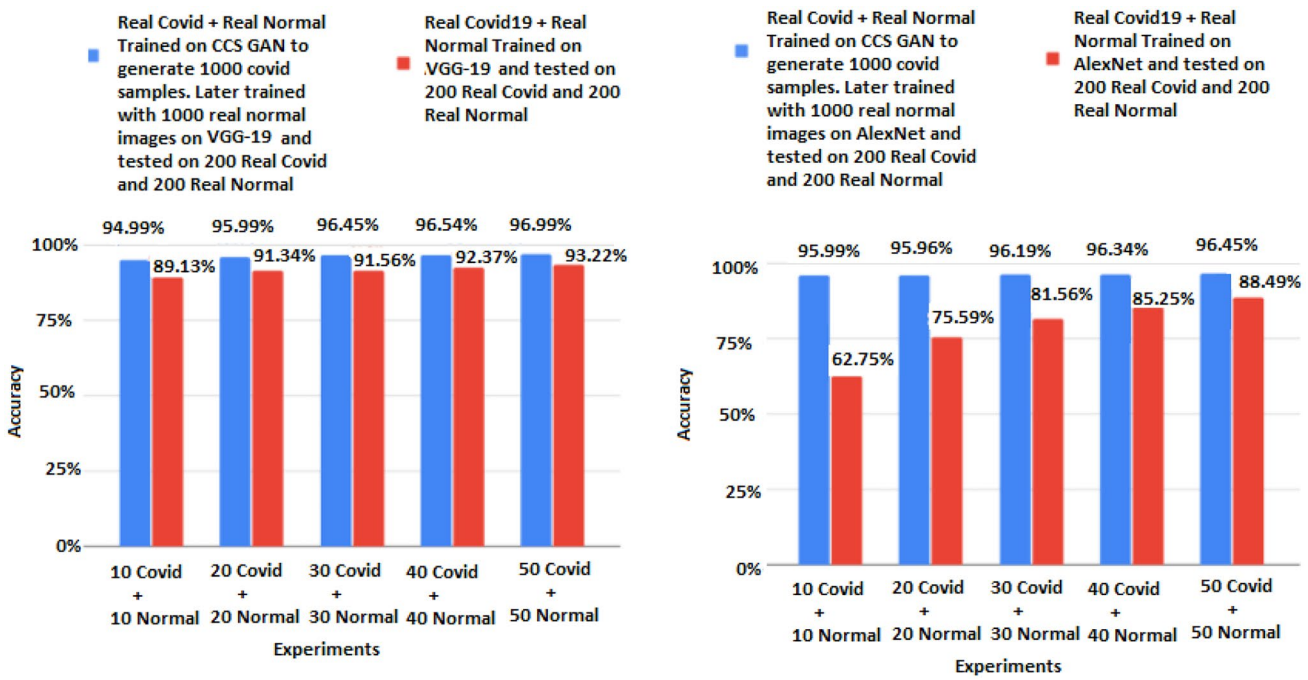

Fig. 5 AlexNet and VGG-19 Stress Test

Experimental Design

CCS-GAN was evaluated both quantitatively and qualitatively, with the overarching goal to determine the extent to which CCS-GAN can improve the ability of a classifier to identify COVID-19 infection using as few positive training images as possible. Furthermore, a stress test is included in which the number of positive images is reduced to as low as 10. Two CNN models were used for evaluation of diagnostic classification: AlexNet and VGG-19. These classifiers were trained for 50 epochs from scratch for every experiment. The Adam optimizer was used with a learning rate of $10-5$. Additionally, another experiment was performed using a form of transfer learning from the normal images only. For this transfer learning experiment, CCS-GAN was initially pre-trained using 500 random unmatched samples of normal images in which both cycle-consistent classes X and Y consisted of normal CT slices. The pre-trained model was subsequently fine-tuned with 1000 normal images and 10 positive COVID-19 images. Finally, an ablation study was performed to compare the results produced by the proposed CCS-GAN method versus a baseline GAN. Throughout this ablation study, individual features of the CCS-GAN are disabled including the cycle-consistent training, the pulmonary segmentation, the cycle-consistent transfer learning with unmatched normal pairs, and the U-Net inspired generator architecture (as replaced by ResNet-50). The key finding of this ablation study is that the entirety of the CCS-GAN approach is necessary to obtain the high-quality results, and that if any of these techniques is disabled, the image quality dramatically suffers and becomes unsuitable for the intended use cases. As such, all of the underlying techniques of CCS-GAN are necessary to achieve the reported performance.

Dataset

For training and evaluation of the GAN, a dataset was provided by the Networking Health, a nonprofit organization, consisting of deidentified imagery from 944 CT scans from patients with COVID-19 as collected across a diverse set of hospital institutions (Fig. 6). However, only between 10 and 50 positive slices were employed for training the GAN in almost all experiments, the only exception being a control experiment in which 500 positive slices were used to observe how this larger number of cases affects qualitative image quality. From these CT scans, only the axial slices in the scan which exhibit substantial lung volume were considered, as determined by the pixel area of the automated lung segmentation mask. The normal CT scans were extracted from the Kaggle lung cancer dataset. This dataset consists of 1020 normal CT scans. All the slices of both the classes were resized to 256×256. The training dataset for the CCS-GAN consists of COVID-19 and normal CT slices ranging from 10 to 50 COVID-19 images for each experiment and the testing dataset consisted of 200 images of each class.

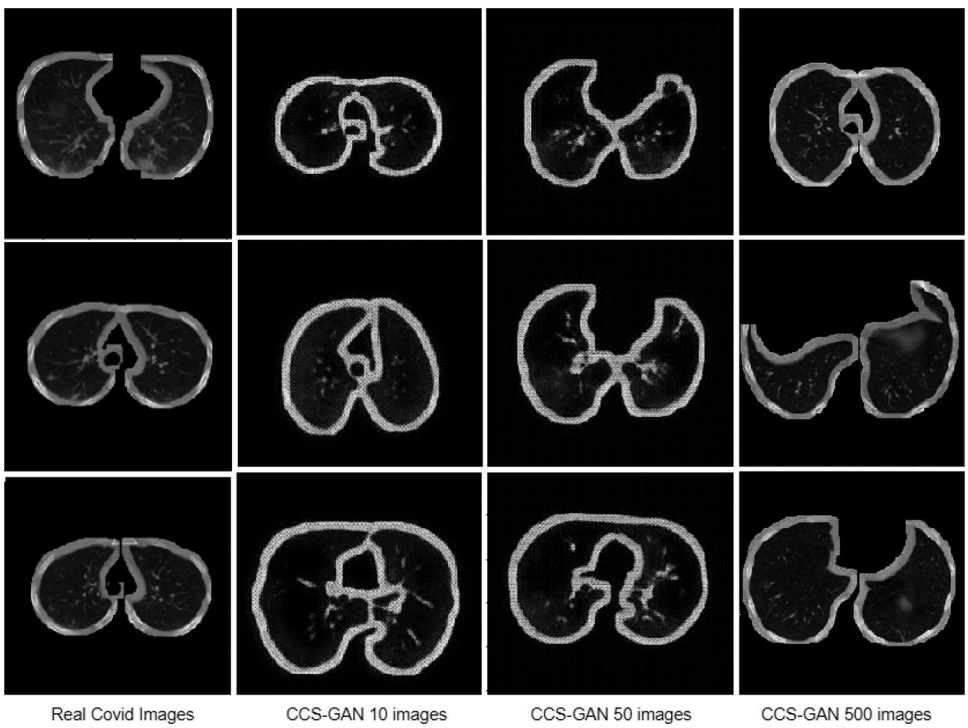

Fig. 6 Results of generated slices using CCS-GAN versus real images of Covid-19 Pneumonia

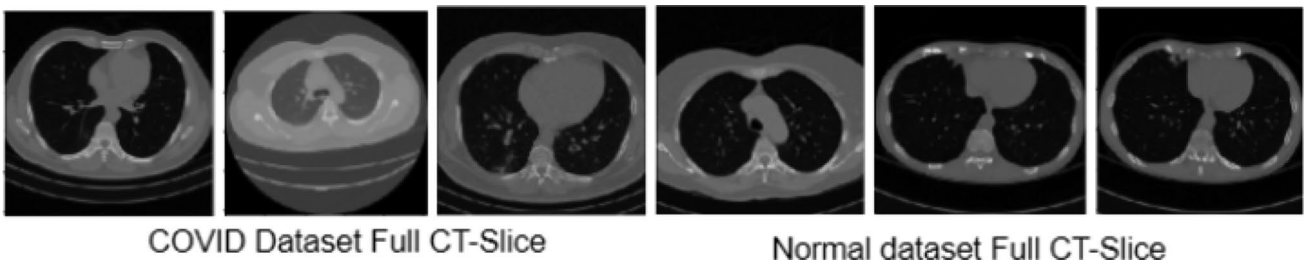

Fig. 7 Examples of real Covid-19 versus Normal CT-Slices.

Statistical Inference and Hypothesis testing

CCS-GAN operates on individual 2D CT scan slices, not whole 3D CT scan volumes. In order to ensure sample independence during our evaluation, we used only one axial slice per CT scan image for training and evaluations. The use of only one axial slice per image during training and evaluation eliminates the potential for auto-correlated results due to the similarity between nearby slices within the same CT scan image and thereby ensures sample independence Fig. 7.

All experimental results include reported accuracies as well as confidence intervals using the Clopper-Pearson method with $\alpha = 0.05$. Confidence intervals are designated by ($S_{\geq}$, $S_{\leq}$), where

$$\begin{aligned} S \geq & \text{ is minimum } p \text{ s.t. } P[Bin(n;p) < x] > \frac{\alpha}{2} \\ S \leq & \text{ is minimum } p \text{ s.t. } P[Bin(n;p) < x] > \frac{\alpha}{2} \end{aligned} \tag{7}$$

where x is the number of correctly identified samples, n is the total number of samples, and p is the unobserved population success probability (unobserved accuracy). When comparing accuracies, p values are also reported using the two-tailed exact binomial test.

Quantitative Analysis

A quantitative analysis was performed to determine the extent to which CCS-GAN can improve the performance of a binary classifier with very few COVID-19 training images. To test the effectiveness, the classifier with CCS-GAN augmented images was compared with a baseline. The backbone classifiers used for this task were AlexNet and VGG-19. The baseline classifier was trained on a balanced dataset of normal and COVID-19 CT slices ranging from 50 to 10 images per class. This baseline was compared against a classifier with additional normal images as well as augmented with synthetic COVID-19 images as generated from CCS-GAN. CCS-GAN was trained using the exact same COVID-19 images as were available to the classifier. The reported performance metrics are accuracy and confidence intervals. Figure 4 shows a stress test which compares the accuracy of AlexNet and VGG-19 baselines versus the addition of synthetic COVID-19 images from CCS-GAN without transfer learning. In this stress test, AlexNet and VGG-19 were trained with a small sample of COVID-19 images ranging from 10 images up to 50 images. In all cases, the addition of CCSGAN augmented images improves performance relative to the baseline. In both cases, the most dramatic improvements were obtained when only 10 COVID-19 images were used. Furthermore, AlexNet + CCS-GAN presents a more dramatic improvement in the classification accuracy relative to VGG-19 + CCS-GAN. Baseline AlexNet achieves an accuracy of 62.75% (0.5780, 0.6750) with 10 COVID-19 images, but this improves to 95.99% (0.935,0.9769) with the addition of CCS-GAN. The accuracy of baseline VGG19 is 89.13% (0.8551, 0.9189) which improves to 94.99% (0.9238,0.9691) with the use of synthetic images as generated by CCS-GAN which is statistically significant. Slightly higher accuracies, with somewhat less substantial improvements, are obtained when the models have access to a larger sample size of COVID-19 images. With 50 COVID-19 images, baseline AlexNet achieves accuracy of 88.49% (0.8496, 0.9145), whereas the inclusion of CCS-GAN improves accuracy to 96.45% (0.9419, 0.9807) which is statistically significant. Moreover, baseline VGG-19 achieves accuracy of 93.22% (0.9033, 0.9550) whereas enhancement with CCS-GAN improves accuracy to 96.99% (0.9481, 0.9844) which is statistically significant. As such, we observe that CCS-GAN greatly improves classification accuracy, and this improvement is particularly pronounced when using only 10 COVID-19 training images relative to 50 COVID-19 images. Table 1 shows a final experiment that was performed in which the CCS-GAN was pre-trained using 1000 normal images (split into two groups of 500 images each), prior to fine-tuning on 10 COVID-19 CT-scan images and 1000 normal CT-scan images to generate 990 synthetic COVID-19 images for the purposes of training the classifier. The classifier was then trained using 1000

Table 1 Comparative performance of classification using CCS GAN with 10 real COVID-19 images

Model	Test data CT slices	10 real COVID-19 CT slices + 10 real normal CT slices	1000 generated COVID-19 CT slices by CCS-GAN using 10 real COVID-19 CT slices + 1000 real normal CT slices	Transfer learning 1000 generated COVID-19 CT slices by CCS-GAN using 10 real COVID-19 CT slices + 1000 real normal CT slices
VGG-19	**200 covid + 200 normal**	**89.13% (0.8551, 0.9189)**	**94.99% (0.9238,0.9691)**	**98.17% (0.9609, 0.9913)**
AlexNet	**200 covid + 200 Normal**	**62.75% (0.5780, 0.6750)**	**95.99% (0.935,0.9769)**	**99.00% (0.9746, 0.9972)**

The bold format was to signify the best performing model

normal images and 1000 COVID19 images (990 generated). In this configuration, the classifier outperforms all of the prior experiments and yields an accuracy of 98.17% (0.9609, 0.9913) using AlexNet, as well as an accuracy of 99.00% (0.9746, 0.9972) using VGG19 which is not statistically significant.

Comparative Comparison and Ablation Study

We have included a comparison study including the COVID-CT-GAN architecture of Li et al. [7] as well as several strong baseline methods including CycleGAN [21], a ResNet-50-GAN, and a U-Net-based GAN. We present generated images from each of these methods using 10 positive and 50 positive training samples along with a larger sample size of negative images. We qualitatively and quantitatively compare CCS-GAN against the following GAN implementations.

COVID-CT-GAN [8]

COVID-CT-GAN is a state-of-the-art generator architecture proposed by Li et al. [7] for the purposes of COVID-19 binary image classification from CT scans. COVID-CT-GAN is based on the auxiliary classifier GAN (ACGAN) [24] including the concatenated version of feature pyramid networks (FPNs) [25] to improve inference across scales, as well as the convolutional block attention module (CBAM) of Woo et al. [26].

Cycle-GAN [21]

Cycle-GAN is a generator architecture for image-to-image translation that makes use of a cycle-consistent loss as well as adversarial loss and perceptual loss. CCS-GAN makes use of the cycle-consistent loss term first presented in CycleGAN, and as such the comparison against CycleGAN is included as a part of the ablation experiment.

ResNet 50 GAN

The ResNet 50 GAN is a baseline method that uses the general purpose ResNet 50 architecture as the backbone generator for the GAN model. The ResNet 50 GAN is therefore included as a comparison against the basic.

U-Net GAN

U-Net GAN is also a component of CCS-GAN as the CCS-GAN generator architecture is based on a U-Net generator model. The U-Net GAN however does not exhibit the cycle-consistent loss, nor does it exhibit the pulmonary segmentation methods that are part of the full CCS-GAN technique, and as such U-Net GAN can be presented as part of the Ablation experiment.

The selection of methods also has the benefit of constituting an Ablation study to determine if the entire CCS-GAN methodology is necessary to obtain the reported results. In particular, we wish to determine if both the pulmonary segmentation and cycle-consistent adversarial loss techniques are necessary, or if a simplified version of CCS-GAN using only one of these methods would be adequate for the intended purposes. The first set of ablation experiments were to generate synthetic COVID-19 images using only pulmonary segmentation, only cycle-consistent adversarial loss, with either the U-Net like proposed generator or using a ResNet 50 generator architecture. The qualitative outcome of these experiments is presented and all of which yield very poor image quality relative to the full CCS-GAN which combines all of these techniques into a single methodology. Figure 8 shows the image quality obtained if cycle-consistent adversarial loss is disabled. As such, the pulmonary segmentation is used in both cases, but the generator is only trained to generate COVID-19 images using the 10 COVID-19 input image slices. We also made use of two generator models, the U-Net-like architecture as used in CCS-GAN, as well as a ResNet50 architecture which is a commonly employed alternative in related literature. As we see in Fig. 8, when

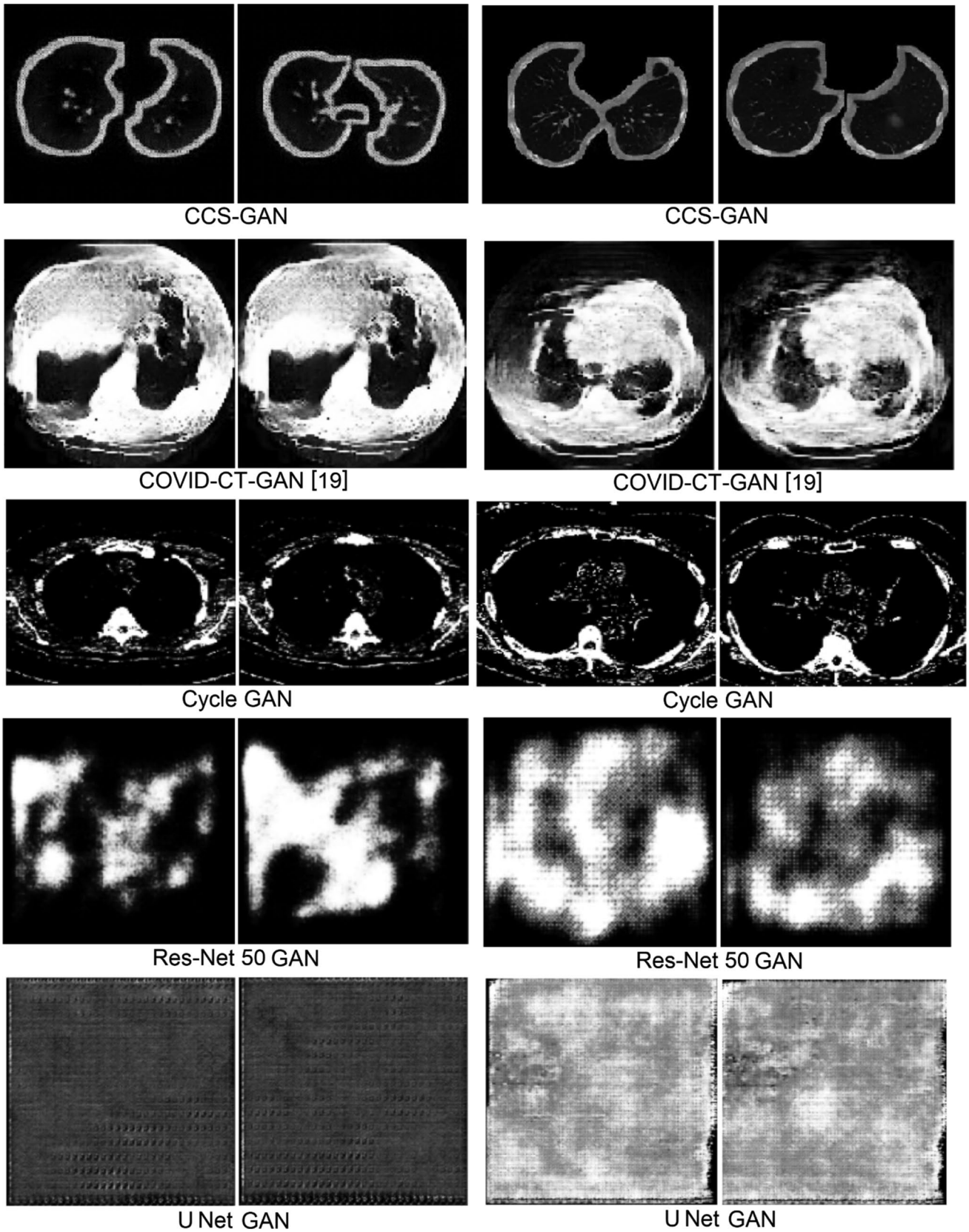

Fig. 8 Comparison of generated synthetic CT slices with models trained with (left) 10 positive CT slices and (right) 50 positive CT slices

Table 2 Comparative accuracy of CCS-GAN versus related advanced and baseline methods

Model	Accuracy using 10 real COVID-19 CT slices + 1000 real normal CT slices	Accuracy using 50 real COVID-19 CT-slices + 1000 real normal CT slices
CCS-GAN with AlexNet	94.99% (0.92383, 0.96919)	**96.99%** (0.94818,0.98440)
CCS-GAN with VGG19	**95.99%** (0.93585, 0.97697)	96.45% (0.94197,0.98074)
COVID-CT-GAN [7] AlexNet	50.25% (0.4523,0.5525)	50.25% (0.4523,0.5525)
COVID-CT-GAN [7] VGG19	50.50% (0.4548, 0.5550)	50.50% (0.4548, 0.5550)
Cycle GAN with AlexNet	50.25% (0.45239, 0.55257)	50.50% (0.45487, 0.55505)
Cycle GAN with VGG-19	50.50% (0.45487, 0.55505)	51.25% (0.46233, 0.56248)
ResNet-50 GAN with AlexNet	50.50% (0.45487, 0.55505)	51.25% (0.46233, 0.56248)
ResNet-50 GAN with VGG-19	50.25% (0.45239, 0.55257)	51.25% (0.46233, 0.56248)
U-Net GAN with AlexNet	50.00% (0.4286, 0.5713)	50.00% (0.4286, 0.5713)
U-Net GAN with VGG-19	50.00% (0.4286, 0.5713)	50.25% (0.45239,0.55257)

The bold format was to signify the best performing model

a regular GAN is trained using pulmonary segmentation, in both cases the resulting images are unreasonable and do not show any meaningful anatomic structure. The U-Net generator is able to produce the rough silhouette of a lung region, but the structure within this silhouette appears more similar to woven fabric than to pulmonary anatomy.

The ResNet50 generator produces red noise and is unable to even construct the pulmonary silhouette at such low volumes. By comparison, the CCS-GAN is able to produce qualitatively reasonable results showing pulmonary features. Figure 8 shows the qualitative results of training CycleGAN (without pulmonary segmentation) versus CCS-GAN with two training configurations. Configuration 1 is with a dataset of 10 slices of the COVID-19 class and 10 slices of the normal class, and Configuration 2 is with a dataset of 50 slices of the COVID-19 class and 50 slices of the normal class. All images are displayed with a standard grayscale window of [− 1000 to 1000] HU. We observe that without pulmonary segmentation, baseline CycleGAN does not generate qualitatively reasonable images using such a small COVID-19 training sample, whereas the images generated by CCS-GAN are qualitatively reasonable and constitute a substantial improvement, although they remain less than clinically accurate. It can be seen that basic CycleGAN at such low training volumes attempts to replicate bones and non-lung features. Furthermore, basic CycleGAN appears to have unreasonably high contrast exaggerating the radiodensity of bones, while unreasonably suppressing the radiodensity of non-bone anatomy. Finally, basic CycleGAN is only able to synthesize few if any pulmonary features. Conversely CCS-GAN is able to synthesize relatively superior pulmonary features. With 50 COVID-19 images, it appears to generate ground glass opacities. As such, it is clear that CCS-GAN, as it includes pulmonary segmentation, is vastly superior to a basic CycleGAN in this qualitative comparison, as the basic CycleGAN does not yield reasonable image quality at these low training volumes. Quantitative analysis using these generated images was performed using AlexNet and VGG19; however, the results are not shown as baseline models achieve very poor performance, as anticipated because the synthetic images using only part of the CCS-GAN methodology are of unacceptable quality. As such, we conclude from these qualitative ablation experiments that the entire CCS-GAN approach is necessary to achieve the reported qualitative results and quantitative performance.

For the quantitative analysis, we trained AlexNet and VGG-19 on 1000 generated COVID-19-positive and 1000 negative samples, then tested the models on 200 positive and 200 negative COVID-19 samples. When AlexNet and VGG-19 were trained on positive samples generated by CCS-GAN (using 10 samples), they had an accuracy of 94.99% (confidence interval of 92.38–96.91%) and 95.99% (confidence interval of 93.55–97.69%) respectively which is not statistically significant. We also trained AlexNet and VGG-19 on images generated by CCS-GAN without segmentations (using 10 and 50 samples, respectively). The models had an accuracy of 50.25% (confidence interval of 45.2–55.2%) and 50.50% (confidence interval of 45.4–55.5%) for AlexNet, and 50.50% (confidence interval of 45.48–55.55%) and 51.25% (confidence interval of 46.23–56.24%) for VGG-19 which is not statistically significant. Finally, we performed a quantitative analysis on images generated by ResNet-50 GAN. AlexNet trained on positive samples generated by ResNet50 (using 10 slices) had an accuracy of 50.50% (confidence interval of 45.4–55.5%) and VGG-19 had an accuracy of 50.25% (confidence interval of 45.2–55.2%) which is not statistically significant. When trained on positive samples generated by ResNet50 (using 50 slices), AlexNet had an accuracy of 51.25% (confidence interval of 46.23–56.24%) and VGG-19 had an accuracy of 51.25% (confidence interval

of 46.23–56.24%) which is not statistically significant. While comparing the performance of CCS-GAN to other related methodologies, it was observed that CCS-GAN outperformed the other methodologies. The results were better and the statistical analysis showed a p value of less than 0.0001 Table 2.

Conclusion and Future Work

Diagnostic classification using few positive training example images is an important problem, because in the early stages of a pandemic, there may be a substantial lag between when the disease is of global health concern and when large datasets are publicly available. A similar problem may also be encountered if screening for rare diseases in which few example images have been collected. A novel methodology is presented that combines intensity based pixel wise segmentation with cycle-consistent segmentation-generative adversarial networks to generate synthetic COVID-19 CT scans with one order of magnitude fewer positive training examples than have been previously demonstrated. As such, CCS-GAN allows CNN classifiers to achieve high performance in COVID-19 diagnostic classification from the CT modality using only 10 CT scan slices from 10 positive cases. As future work, we wish to extend this approach to be able to classify with only a single example of the target disease. Furthermore, in future wor,k we wish to determine if the reported results are only valid for detection of COVID-19 or if the CCS-GAN approach is capable of screening the presence of other pulmonary diseases with few examples. We remain optimistic that in the near future, the need for very large training datasets of a target disease will no longer be a hindrance to the timely development of accurate AI-based screening algorithms [27–31].

Funding This research was supported by NSF award titled RAPID: Deep Learning Models for Early Screening of COVID-19 using CT Images, award # 2027628. This work was also supported in part by the NSF IUCRC program as part of the Center for Advanced Real Time Analytics (CARTA) Research Experience for Undergraduates (REU).

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Shuja, J., Alanazi, E., Alasmary, W., & Alashaikh, A. (2021). COVID19 open source data sets: a comprehensive survey. Applied Intelligence, 51(3), 1296-1325.
2. Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). COVID-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988.
3. Mohamed Loey, Gunasekaran Manogaran, and Nour Eldeen M Khalifa. "A deep transfer learning model with classical data augmentation and cgan to detect COVID-19 from chest CT radiography digital images". In: Neural Computing and Applications (2020), pp. 1–13.
4. Tripti Goel, R Murugan, Seyedali Mirjalili, et al. "Automatic Screening of COVID-19 Using an Optimized Generative Adversarial Network". In:Cognitive Computation (2021), pp. 1–16.
5. Yifan Jiang, Han Chen, MH Loew, et al. "COVID-19 CT Image Synthesis with a Conditional Generative Adversarial Network". In: IEEE Journal of Biomedical and Health Informatics (2020).
6. Siqi Liu, Bogdan Georgescu, Zhoubing Xu, et al. "3d tomographic pattern synthesis for enhancing the quantification of COVID-19". In: arXiv preprint arXiv:2005.01903 (2020)
7. Li Z, Zhang J, Li B, Gu X, Luo X. COVID-19 diagnosis on CT scan images using a generative adversarial network and concatenated feature pyramid network with an attention mechanism. Med Phys. 2021 Aug;48(8):4334–4349. https://doi.org/10.1002/mp.15044. Epub 2021 Jul 9. PMID: 34117783; PMCID: PMC8420535.
8. J. Mangalagiri et al., "Toward Generating Synthetic CT Volumes using a 3D-Conditional Generative Adversarial Network," 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 2020, pp. 858-862, https://doi.org/10.1109/CSCI51800.2020.00160.
9. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.
10. Ghassemi, N.; Shoeibi, A.; Khodatars, M.; Heras, J.; Rahimi, A.; Zare, A.; Pachori, R.B.; Gorriz, J.M. Automatic Diagnosis of COVID-19 from CT Images using CycleGAN and Transfer Learning. arXiv 2021, arXiv:2104.11949.
11. Jin, Q., Cui, H., Sun, C., Meng, Z., Wei, L., & Su, R. (2021). Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images. Expert Systems with Applications, 176, 114848
12. Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in CT segmentation tasks. Scientific reports, 9(1):1–9, 2019
13. Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.
14. Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing, 321:321–331, 2018.
15. A. Degerli, et al., "COVID-19 Infection Map Generation and Detection from Chest X-ray Images," arXiv preprint arXiv:2009.12698, 2020.
16. Z. Xu, et al., "GASNET: Weakly-Supervised Framework for COVID-19 Lesion Segmentation," arXiv preprint arXiv:2010.09456, 2020
17. Parnami, A., & Lee, M. (2022). Learning from few examples: A summary of approaches to few-shot learning. *arXiv preprint* arXiv:2203.04291.
18. Koch, G., Zemel, R., & Salakhutdinov, R. (2015, July). Siamese neural networks for one-shot image recognition. In ICML deep learning workshop (Vol. 2, No. 1).
19. Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. *Advances in neural information processing systems*, *29*.
20. Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. Advances in neural information processing systems, 30.
21. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223–2232).
22. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

23. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015.
24. Odena, A., Olah, C., & Shlens, J. (2017, July). Conditional image synthesis with auxiliary classifier gans. In International conference on machine learning (pp. 2642–2651). PMLR.
25. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the Euro0070ean conference on computer vision (ECCV) (pp. 3–19).
26. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3–19).
27. Q. Hu, L. F. de F. Souza, G. B. Holanda, S. S. A. Alves, F. H. dos S. Silva, T. Han, and F. P. P. Rebouc¸as, "An effective approach for CT lung segmentation using mask region-based convolutional neural networks," Artif. Intell. Med., vol. 103, Mar. 2020, Art. no. 101792
28. S. Menon, J. Galita, D. Chapman, A. Gangopadhyay, J. Mangalagiri, P. Nguyen, and M. Morris, "Generating Realistic COVID-19 X-rays with a Mean Teacher
29. A. Likas, N. Vlassis and J. Verbeek, " The global K -means clustering algorithm ", Pattern Recognit., vol. 36, no. 2, pp. 451-461, 2003
30. Oh, Y., Ye, J.C., 2021. Unifying domain adaptation and self-supervised learning for cxr segmentation via adain-based knowledge distillation. arXiv:2104.05892.
31. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117–2125).