Applications of cyclic invariant style transfers in medical imaging

ECE4179 - Neural Networks and Deep Learning

Thomas Clancy, Benjamin Milanko
Monash University

October 21, 2019

We explore the use of CycleGAN, a image-to-image Generative Adversarial Network (GAN), in the style transfer between chest x-rays of healthy patients and those with confirmed cases of bacterial or viral pneumonia. Furthermore, we discuss the applications of these style transferred images as a means of assisting clinicians in their diagnosis of these pathologies by using the network to identify regions of interest in presenting patients.

Contents

1 Introduction

1.1 CycleGAN

CycleGAN [1] is a Generative Adversarial Network (GAN) that performs image-to-image translation. More formally, this can be expressed as the learning of a mapping:

$$G : X \to Y$$

where X is a style of image (e.g. photos of landscapes) and Y is another style of image (e.g. Van Gogh landscape paintings) to which we would like to perform style transfer from the domain X. This type of style transfer is demonstrated in Figure 1.

(a) Original

(b) Generated style transfer

Figure 1: An example of style transfer of a landscape photo using a CycleGAN network trained on Van Gogh paintings

Unlike other image-to-image translation networks that require paired training examples that express a change in style with some invariance to other properties of the image (e.g. a photo and a painting of the same scene), CycleGAN learns on two classes of images that are unpaired. A network of this kind is of particular interest for problems where the collection of paired training samples is unfeasible, or limits the quantity of data that can be collected.

The lack of paired training data presents a significant problem if the standard adversarial loss is used for the training of CycleGAN; it does not enforce that translated images maintain their original properties that do not relate to the style transfer. By way of example, if a style transfer of photos of people to paintings of people was being learned, adversarial loss would not place a penalty on the painted person being unrecognisable as long as the output image resembled the trained painting style. The network will learn to replicate samples from the painting domain, independent of the input, in a process known as “Mode Collapse”. To address this problem, the loss function that CycleGAN uses includes an additional **cycle consistency loss**:

$$\mathcal{L}_{\text{cyc}}(G, F, X, Y) = \frac{1}{m}\sum_{i=1}^{m}[F(G(x_i) - x_i] + [G(F(y_i) - y_i]$$

This requires the learning of an inverse mapping

$$F : Y \to X$$

so that the loss term can penalise the network for the difference between an original image in one class, and the same image that has been transformed to the other class, and back to its original class.

1.2 Style Transfer in Medical Imaging

Many applications of deep learning in the field of medical imaging relate directly to building networks to classify or grade certain pathologies based on their presentation in imaging [2]. Given, however, our own prior interest in style transfer we chose to explore how an image-to-image translation network could learn the features of a particular pathology as a form of image style. Furthermore, we propose that a network of this kind could overcome a perceived mistrust of “black box” classifiers in the medical profession, by instead acting as a diagnostic tool that identifies regions of interest in a given image by observing changes to the image after performing style transfer to the healthy class and the pathological class.

A network that doesn’t require paired training examples was essential for exploring this question given that we were unable to source a suitable medical dataset included paired examples of the same patient in a healthy and pathological state. The use of CycleGAN was a natural choice since it allowed us instead to use a dataset that included general classes of healthy images and pathological images, with no relationship between these images.

1.3 Dataset

For this project, we utilised a dataset of labelled chest x-ray images from a paediatric hospital including healthy patients, and those presenting with pneumonia[3]. These authors established the use of a deep learning tool for classification, rather than our intended purpose of style transfer. The dataset includes 5,863 JPEG images of chest x-rays of paediatric patients from Guangzhou Women and Children’s Medical Center, Guangzhou. An example of our use of CycleGAN to generate style transfers on this dataset is displayed in Figure 2.

(a) Original (b) Generated style transfer (c) Reconstruction

Figure 2: An example of the network taking an image of a patient presenting with pneumonia (a), generating a healthy style transfer (b), and attempting to re-construct the original image (c).

2 Method

2.1 Pre-processing

The dataset images were provided de-identified, and cropped around the chest area. All images in the set were high resolution (greater than 1000x1000) and had varying aspect ratios. The images were first scaled to have minimum dimensions 283, then cropped to 256x256. The images are also randomly flipped horizontally to augment the data.

2.2 Network Architecture

Assuming that the network direction is A to B to A. The original image is transformed via generator A to B, this fake image in B domain is evaluated against a real sample from the B domain. The discriminators loss is propagated through the network and is also used for generator A. The fake image is then converted back to domain A via generator B to A where the same discriminator evaluation occurs.

The reconstructed image is then compared against the original image with the loss function described above. This loss is applied with a factor λ in conjunction with the adversarial loss to train the network. λ defines the importance of maintaining cycle consistent results. The propogation of loss in the network is summarised in Figure 3. A 9 block ResNet [4] architecture is used for the generators and PatchGAN is used for the discriminators.

2.3 Training and Evaluation

Training occurred over 200 epochs, with 4000 iterations per epoch. Each epoch took approximately 4200 seconds to run, making the total training time close to 240 hours. Training took place on a GPU-accelerated PC with 5.2 Cuda Compute Capability.

Due to the qualitative nature of the image output the there is no definitive accuracy metrics that can be applied. Furthermore, CycleGANs loss metrics over epochs were unhelpful in providing a representation of the performance of the network. To evaluate the network, the images were upscaled to 1024x1024 using bicubic interpolation, then the difference was taken of each to highlight regions of greater and lesser change. The results were analysed by us with additional comments from medical professionals.

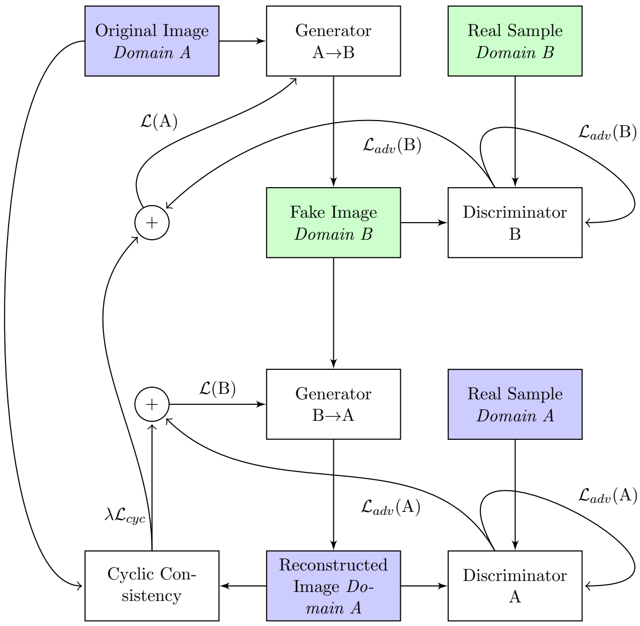
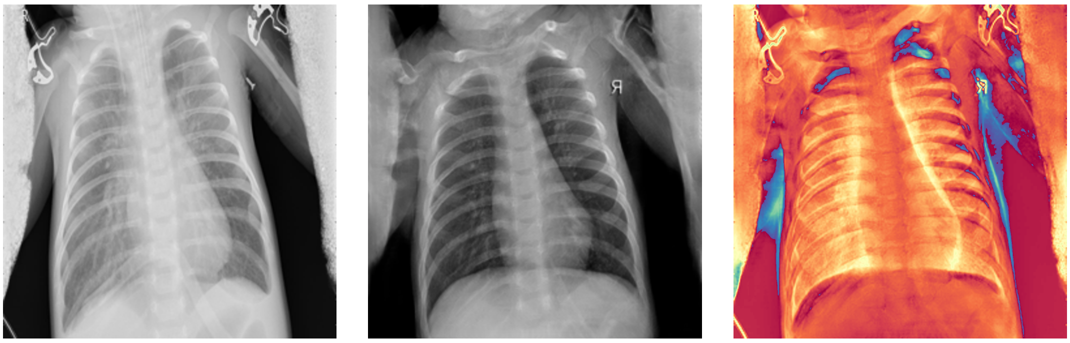


Figure 3: A visual representation of the propagation of loss through the network.

3 Results

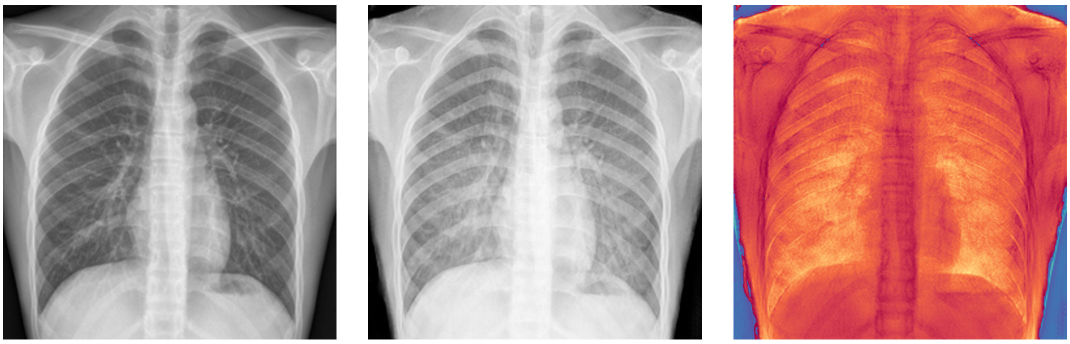
Figure 4 demonstrates an example of style transfer from imaging of a patient diagnosed with pneumonia, to a generated healthy counterpart. The original image shows fuzzy heart border bilaterally, fuzzy diaphragm borders, blunted left costophrenic angle (suggesting fluid in that region) and consolidation suggesting bilateral pneumonia. These clinical signs are absent from the second, generated image. However, the network has performed additional interesting changes; The ECG clips seen in the first image have disappeared, along with the intravenous tube. The generator has also mistaken a screw behind the babies left armpit as the R (Right) symbol in the x-ray. Finally, the network has also distorted the bone structure, fusing the uppermost rib bones and distorting the shoulders.



(a) Original (b) Generated style transfer (c) Delta

Figure 4: The above set of images demonstrates the style transfer of x-ray imaging that indicates pneumonia, to an image of a (fake) healthy counterpart. (c) Shows the difference between these images.

Figure 5 demonstrates an example of style transfer from imaging of a healthy patient, to a generated counterpart with pneumonia. Of note, image (b) shows bilateral haziness suggesting pneumonia. Interestingly this introduced haziness is not uniform, this can be seen as yellow highlights outlining the diaphragm and heart. The introduced haziness in the lungs is particularly noticeable in the delta image.



(a) Original (b) Generated style transfer (c) Delta

Figure 5: The above set of images demonstrates the style transfer of x-ray imaging of a healthy patient, to an image of a (fake) counterpart that has been diagnosed with pneumonia. (c) Shows the difference between these images.

4 Discussion

Through examination of the deltas from each image set, we can determine that the basic lightening/darkening pattern associated with the disease is non-linear and follows some complex, network defined function. The increase or removal of fuzziness specifically around the heart and diaphragm is consistent with the clinical signs of pneumonia, which demonstrates our networks ability to recreate certain medically accurate phenomena.

There are certain decisions made by the network which are influenced by the training data. In sick to healthy transfers, the network removed ECGs, IV drips and intubation tubing, this can be explained by a representation of these medical devices among the sick samples and a lack among the healthy samples. This feature does not affect the performance of the network negatively, however, there are additional changes which reduce the quality of the output. From the sick to healthy transform the network can be observed shrinking the chest cavity, oftentimes fusing the upper rib bones to the neck. Additionally, the network will also fuse or distort the shoulder bones together. We believe this is due to imbalances in the data set which are discussed further below.

Ultimately the network was able to recreate aspects of the disease from the healthy to sick transfer and remove them from the reverse transfer. These features are however, low resolution and difficult to apply to a clinical setting. The unexpected learned behaviour is ultimately focused just outside the area of interest and should not cause the network to be deemed a failure.

4.1 Dataset Asymmetry

Due to the lack of metadata in the dataset, a numerical conclusion is difficult to reach. However, it appears that there is an over representation of sick younger children, to healthy younger children and also an over representation of healthy to sick older children. These biases in the dataset causes the network to skew results. It associates the sick domain to features found in younger children, while associating the healthy domain with older characteristics. This is apparent when viewing bone structures around the neck and shoulders. The shoulder of a young child is made up of many separate bones which fuse together as they grow older. Oftentimes, the network will attempt to transform this area to more resemble a young child, in the case of a healthy to sick transform or apply older features in the case of a sick to healthy generator. The overall shape of the chest is also distorted, the sick samples having an overall elongated shape.

4.2 Limited Resolution

Training large convolution based networks is computationally costly, due to time and hardware limitations, 256×256 was the chosen resolution. At this size, the image is of reduced use to medical personnel. The first modification was to change the load and crop size of the network. As the network is fully convolutional it should be scale invariant, however, using a higher quality image as inputs to the network caused the outputs to be distorted by high frequency noise as seen in Figure 6. The solution was to upscale the images with a bicubic interpolation technique, while not introducing new information, it allowed existing features to become more visible.

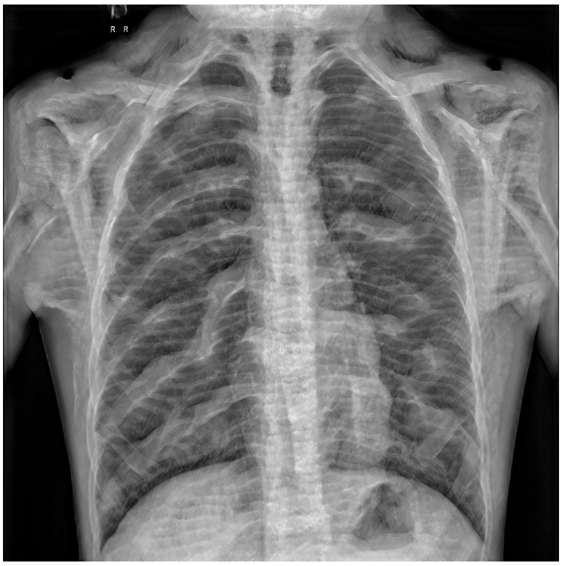


Figure 6: Distortion of the image is visible when a higher resolution image is used than the resolution on which the network was trained

4.3 Cheating Network

Due to the adversarial nature of the network, researchers [5] have found the network to exhibit steganographic behaviour. The network was "hiding" imperceptible information in high frequency and low contrast bands when generating the fake image. This allowed the network to recreate the original image near perfectly, even when the domain of the fake image lost data from the original. Comparing the adaptive equalized histograms of the paired dataset revealed the trickery. As we are using an unpaired dataset we are only able to speculate on any self-derived encoding schema, however, the high frequency noise seen in the enlarged images as well as certain aspects of the adaptive equalization histogram suggest there may be some level of encoding happening.

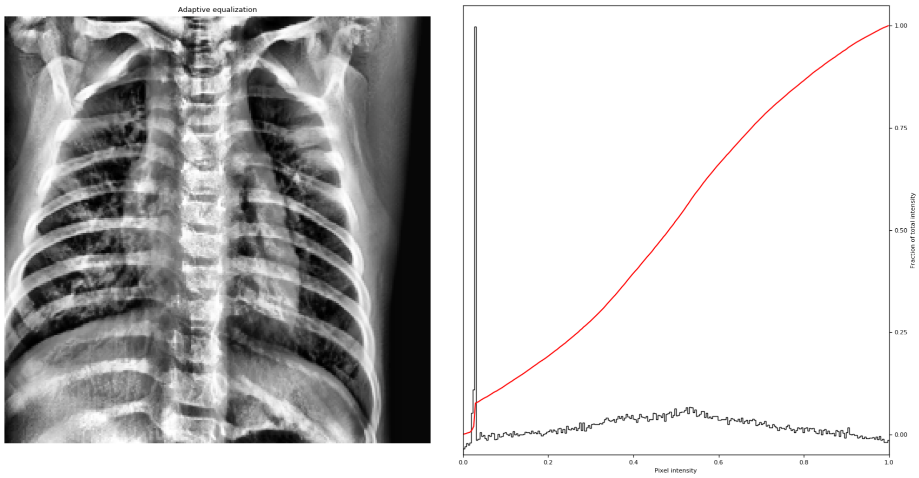


Figure 7: The above analysis indicates the possibility of CycleGAN engaging in steganographic behaviour, based on similar observations by researchers [5]

4.4 Future Study

While the scope of the research carried out in this paper is limited, there are many exciting avenues of investigation to continue pursuing. These could lead to major improvements in the fields of AI-Assisted Medical Imaging and Machine Learning. Using a variable sample size, with datasets as low as 50 images will test the networks ability to accurately generate medical images. The network has already demonstrated accurate style transfers using small sample sets of renowned artists, which gives promise to the possibility of performing well on diseases. The ability to generate additional samples would be invaluable to medical students and researchers investigating rare diseases.

For a medical network which aids doctors, high quality images are essential. Doctors train with high quality x-rays and are restricted in the diagnoses that can be inferred from low quality images. To achieve this, a “Growing GANs” approach is proposed. Growing GANs [6] are most well known for generating photorealistic human faces; they use many GANs which start small (4×4 pixels), each network doubling in size until a high resolution image is produced.

This technique removes the difficulty of placing semantic tasks on a single network, instead, spreading the load among many. For the use case of generating sick or healthy chest x rays, the lower levels of the network could set out the structure of the image, such as, the overall chest shape. At the next level, the more specific details such as bones could be filled in. However, the network does not need to know where on the image to place the bones as the previous network has already created a template. Each successive layer fills in more and more details until a high quality image is achieved.

This negates a major pitfall in using CycleGAN where it struggles to make structural changes to the image, a canonical example being style transfer of dog images to a cat images. To improve this growing GAN approach, cyclic consistency can be introduced to avoid mode collapse.

A clear next step for the network is to apply datasets of different diseased samples. Changing the disease represented in the sick samples is trivial for the network and would simply be limited by training time. Additionally, the dataset could be modified to be specific to a subset of a disease, increasing network fidelity. The network could also be applied to certain treatments of the disease, which could visualise treatment outcomes ahead of time.

As improvements continue, we see possibilities for completely novel fields in medical imaging. “Predictive medical imaging” would allow doctors to use GAN-Like structures to identify and eliminate health problems before they arise. Doctors could view a patient’s healing process from day one, using sequential models to introduce external factors, with instantaneous results predicting health outcomes. Additionally this structure could be applied as a prognostic tool to predict how diseases will develop through medical imaging.

References

[1] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017.

[2] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

[3] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[5] C. Chu, A. Zhmoginov, and M. Sandler, "Cyclegan, a master of steganography," *CoRR*, vol. abs/1712.02950, 2017.

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *CoRR*, vol. abs/1710.10196, 2017.