Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/45/20241170

SRGAN and CNN integration for enhanced chest CT diagnostics

Mengliang Tan[1,4], Haoyang Wang[2] and Wentao Zhang[3]

[1]The Department of Anthropology, University of California, San Diego, San Diego, 92093, USA
[2]The Department of Mathematics, University of California, San Diego, San Diego, 92093, USA
[3]The Department of Statistics and Applied Probability, University of California, Santa Barbara, Santa Barbara, 93106, USA

[4]metan@ucsd.edu

Abstract. Within the domain of Convolutional Neural Networks, there exists no prior application of ESRGAN for the pre-processing of CT images aimed at CNN diagonalization. The quality of CT scan images holds a direct correlation with the precision of diagnoses. Presently employed methods to enhance image quality are antiquated and have yielded less favorable outcomes. Our objective is to amalgamate ESRGAN, renowned for its prowess in enhancing image resolution, with CNN, in pursuit of a more effective and precise diagnostic procedure. In this project, we first gather datasets from the internet. And then in order to test if ESRGAN can actually improve the functionality of CNN, we preprocess images to make images in the dataset blurry. This provides a dataset served for comparison. Later, we will use ESRGAN model from the internet to improve the quality of blurred images and test CNN on both of them for comparison. As a result, using Enhanced ESRGAN can help improve the accuracy of detecting whether there is a disease in the CT images.

Keywords: Computerized Tomography, Generative Adversarial Networks, Convolutional Neural Networks.

1. Introduction

In the realm of medical imaging, especially in chest Computerized Tomography (CT) scan interpretations, traditional methods have relied on conventional image resolution techniques such as Spatial Resolution Enhancement, Contrast Enhancement, Noise Reduction. These standard resolution mechanisms, though effective in many circumstances, may struggle to produce clear and accurate visual representations of finer anatomical details [1]. The underlying challenge has been the limited capacity of these methods to transform low-resolution images into high-definition visuals without losing vital details or introducing artificial ones. The adverse effect of this limitation is the possibility of misdiagnosis or oversight of critical abnormalities due to inadequate image clarity [2].

Over the past decade, advancements in image processing techniques have witnessed the adoption of Convolutional Neural Networks (CNN) for a variety of medical imaging tasks. A notable study implemented by Rutwik Gulakala employed the strengths of CNN which can deal with the randomness of luminosity that arises due to different sources of X-ray images[3]. However, the approach still depended on the inherent resolution of the base image. Since there exist some limitations of generative adversarial networks (GANs). For example, in GANs, the generator may produce a limited range of output and ignore some modes present in data distribution. Also, GANs are sensitive to hyperparameters and finding the right set of hyperparameters can be time consuming. The breakthrough came with the introduction of Super-Resolution Generative Adversarial Networks (SRGAN) in medical imaging [4]. By employing SRGAN, researchers could successfully upscale low-resolution images to much higher resolution, potentially aiding in better disease detection and diagnosis [5].

However, a significant gap remains. While SRGAN techniques show promise in increasing the image quality, their direct application to chest CT scans, specifically for improving diagnosis accuracy through a combination with CNN models, has not been thoroughly explored. Current studies either focus solely on SRGAN's image enhancement capabilities or purely on CNN based diagnosis without considering image resolution upscaling [6].

Our study aims to bridge this gap by integrating the SRGAN mechanism to upscale low-resolution chest CT scans, followed by applying a trained CNN model for improved diagnostic accuracy. The proposed technique integrates SRGAN with CNN to process chest CT scans. Initially, low-resolution images undergo a super-resolution process using SRGAN, enhancing finer anatomical details without introducing artificial artifacts. Post upscaling, the high-resolution images are fed into a CNN model trained specifically for chest CT scan interpretations. The CNN model, equipped with multiple convolutional layers, filters, and pooling operations, extracts essential features from the upscaled images. We conclude that by enhancing the image quality, the CNN model will have access to more refined details, subsequently leading to better diagnostic outcomes. Such an approach could improve the way chest CT scans are interpreted, moving towards a future with reduced misdiagnoses and ensuring better patient outcomes.

2. Method

2.1. Data preparation

The data we chose was sourced from Kaggle [7]. There are 1000 Chest CT scan images in the data and each image represents 4 different conditions of a human's chest. They are chest with adenocarcinoma, chest with large cell carcinoma, chest with squamous cell carcinoma, and healthy chest. The resolution of each image is 96×96. The images are all in gray and black color, and Figure 1 is a sample of each image.

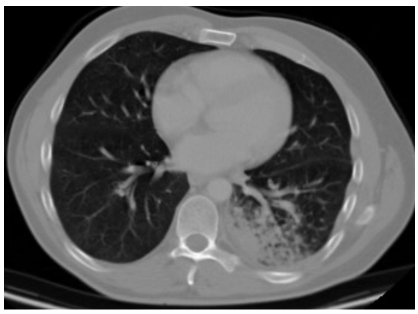

Figure 1. Example CT image.

All the images are divided into three parts, 70 percent of the images are in the training set, 20 percent of the images are in the test set and 10 percent of the images are in the validation set. Since the research aims to find out whether using SRGAN can help make better judgments of cancer in the chest when there is a blur, we then add Gaussian blur to each graph in the test set.

Then all we need to implement is to use the SRGAN to improve the image that we add blur and use CNN to check if SRGAN help improve the accuracy.

DOI: 10.54254/2755-2721/45/20241170

2.2. SRGAN

GANs are a class of artificial intelligence algorithms used in unsupervised machine learning. It was proposed by Goodfellow et al to address difficulties met by Generative models such as certain probabilistic computations and the use of piecewise linear units [8]. GANs operate based on a unique framework that involves two neural networks: the Generator and the Discriminator. The Generator's role is to generate data, while the Discriminator's task is to evaluate whether the data is genuine or fabricated by the Generator. SRGAN is a type of GAN used for super-resolution tasks. Its purpose is to improve the resolution of input pictures through upscaling and enhancing details in pictures. The thing that makes SRGAN special among all types of GANs is the usage of perceptual loss. Perceptual loss is a measure used in image processing that evaluates the difference between images based on their perceptual and semantic differences, rather than just pixel-by-pixel discrepancies.

We use a pre-trained model from [9]. It is an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) model. Its difference from SRGAN is that its architecture is distinctive, comprising several RDDB blocks and integrating residual scaling and smaller initialization techniques to enhance training ability [10].

The implemented code first loads the ESRGAN model and sets up an output directory for the enhanced images. The code then iterates over specific subfolders, each containing images to be enhanced. For each image in these subfolders, the code loads the image, preprocesses it by resizing and normalizing, and then feeds it to the SRGAN model for enhancement. The enhanced image is later saved in the new folder

2.3. CNN

CNNs are widely used in the various computer vision tasks due to their excellent performance [11, 12]. In terms of the CNN part, we would like to create a model to diagnose each CT image in which kind of disease. After that, we can compare the result of accuracy we get from the CT images with Gaussian blur and CT images improved by Enhanced Super- ESRGAN. The model we created is defined as a sequential model. It makes the layers stacked sequentially. When everything is set up, we create 4 blocks for the model. The first one consists of two Convolutional layers with 64 filters each, both using a 3x3 kernel, 'same' padding, and ReLU activation. After these two convolutional layers, a MaxPooling2D layer is applied with a 2x2 pooling size, reducing the spatial dimensions by half. Batch normalization is also applied after the max-pooling layer. And for the second, third, fourth block, they are all similar to the first block, the only difference is that the second, third, fourth block consists of two Convolutional layers with 128, 256, 512 filters each and all followed by a MaxPooling2D layer and batch normalization. After the convolutional layers, the Flatten layer is used to convert the 2D feature maps into a 1D vector. Then, there is a fully connected layer with 64 units and ReLU activation. A Dropout layer is added to prevent overfitting, with the dropout ratio specified by the Dropout_ratio parameter. Also, we use the "trainable" argument for each Conv2D layer which allows us to control whether each block of layer is trainable or not. The final output layer has 4 units, which means that the model is designed for a classification task with four classes, and it uses softmax activation to produce class probabilities. In the end we use the default settings for Adam optimizer, we use CategoricalCrossentropy as loss function and we specify what metrics should be used to evaluate the model's performance during training and testing. With all we implemented, we can output the loss and accuracy in each training.

DOI: 10.54254/2755-2721/45/20241170

3. Result and discussion

3.1. ESRGAN

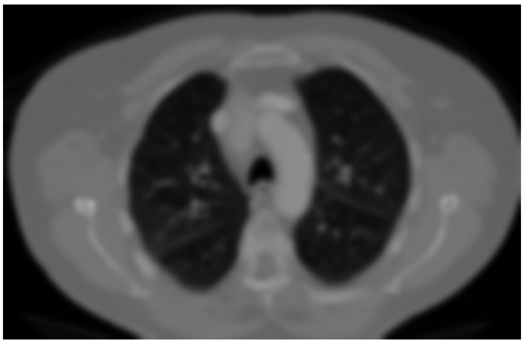

Figure 2. Example Gaussian blur image.

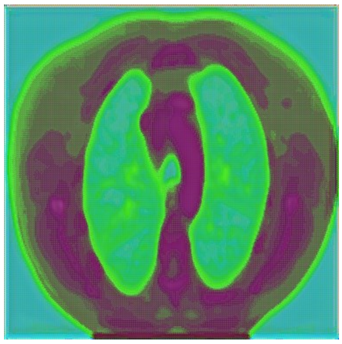

Figure 3. Example Gaussian blur image after ESRGAN.

Figure 2 and Figure 3 are examples of images before being processed by ESRGAN and after being processed by Enhanced Super-ESRGAN. The left one is blurred. It has a resolution of 409 x 264, which is considered to be very low. The right one has a resolution of 896 x 896 pixels, and it is also resized to the input shape that the model expects, otherwise, there would be a problem with the enhancement process. As the image resolution improves, the image's data content also escalates, leading to a proportional enlargement in file size. Additionally, a notable alteration has occurred in the image's color palette, transitioning from the original black and white to a new combination of green and purple. The precise cause of this color shift remains uncertain.There could be innate flaws with the model used or problems with functions from imported libraries. If simply observing the pictures with human eyes, there is indeed a decrease in the blurriness of images, which is some improvement in the clarity. However, it is not optimistic. The picture still does not reach a very high level of detail and sharpness.

3.2. The performance of CNN

After we get 50 epochs of each data set, we get the accuracy of two datasets. To compare these two results, we decided to analyze them in 2 ways: maximum number, mean number. For the data that with Gaussian blur, the maximum accuracy is 0.5041, the mean of accuracy is 0.2745. However, for the accuracy we got for data after implementing ESRGAN, the maximum accuracy is 0.6111, mean accuracy is 0.29191.

Here is the plot of accuracy of data with Gaussian blur and data after doing ESRGAN.

DOI: 10.54254/2755-2721/45/20241170

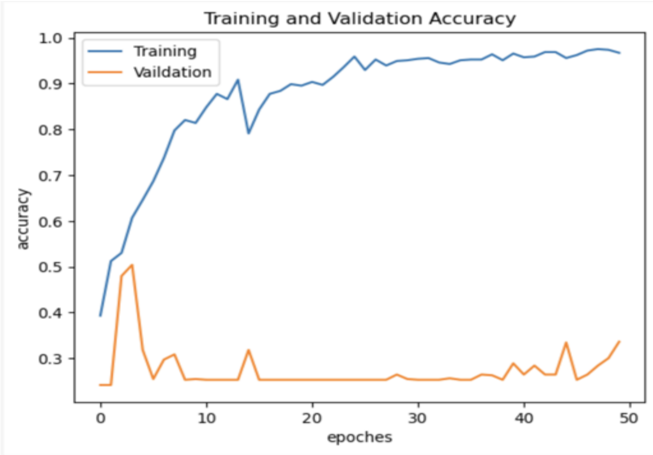

Figure 4. Training and Validation Accuracy (without ESRGAN).

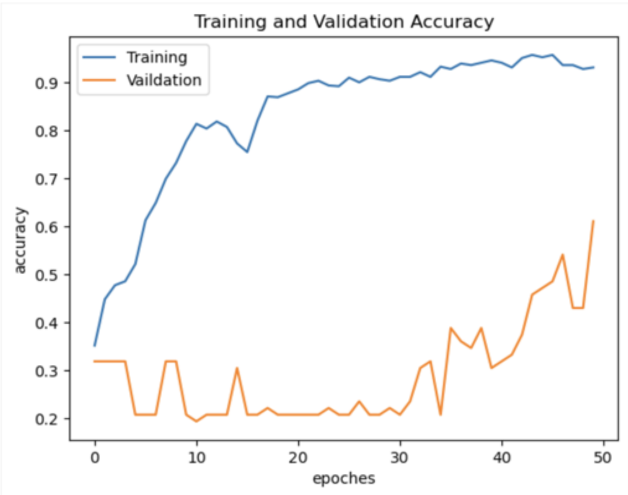

Figure 5. Training and Validation Accuracy (with ESRGAN).

The Figure 4 is an accuracy plot for data without doing ESRGAN, and the Figure 5 is accuracy for data after doing ESRGAN. As we can see in two graphs, most of the accuracy is around 0.25, but the highest accuracy for a graph after doing ESRGAN is much higher than the graph without ESRGAN.

Figure 6. Training and Validation Losses (without ESRGAN).

Figure 7. Training and Validation Losses (with ESRGAN).

Also, we see the loss graph shown in Figure 6 and Figure 7, in which the first one is a graph of data with Gaussian blur and the second one is for data after doing ESRGAN. It is apparent that in the first graph, the validation loss goes high as the training loss decreases. However, for the second graph, the validation loss is still higher than the training loss but as the training loss decreases, the validation loss has a downtrend.

In addition, we calculate the confusion matrix for both model's predictions and calculate the accuracy scores of our model's predictions. We find that by doing ESRGAN in the data, the accuracy score improved around 0.5, which indicates that ESRGAN can help a little bit for CT images with Gaussian blur.

The rationale underlying our conjecture regarding the potential efficacy of ESRGAN in enhancing predictive capabilities is as follows: In the presence of Gaussian blur, critical diagnostic information within the graphical data may be obscured, thus compromising the accuracy of disease detection. However, the application of ESRGAN to the graphical data holds promise for the restoration of obscured information. Consequently, this approach stands to contribute to the refinement and augmentation of predictive accuracy within CNN models.

4. Conclusion

In this research project, we enhanced the quality of blurred images through ESRGAN to let CNN access pictures with more details and clarity so its accuracy of diagnosis may be improved. We preprocessed CT images of the chest by making it blurrier and then used Enhanced Super-Resolution ESRGAN to make them more detailed. And then we use CNN to diagnose processed images. We found out that by using ESRGAN, the accuracy of the prediction improved, and it showed that we can use ESRGAN to deal with CT images. In the future, we plan to refine the upscaling process of ESRGAN to avoid color change of original image for more accuracy and apply the combination of ESRGAN and CNN to diagnose other kinds of CT Images.

Authors Contribution

All the authors contributed equally, and their names were listed in alphabetical order.

References

[1] Alwakid G Gouda W Humayun M & Sama N (2022) Melanoma Detection Using Deep Learning-Based Classifications Healthcare 10(12)
[2] El-shafai W Ali A M El-Rabaie E M Soliman N F Algarni A D & Abd El-Samie F E (2022) Automated COVID-19 Detection Based on Single-Image Super-Resolution and CNN Models Computers Materials & Continua 70(1)
[3] Gulakala R Markert B & Stoffel M (2022) Generative adversarial network based data augmentation for CNN based detection of Covid-19 Scientific Reports 12
[4] Yu H Wang S Fan Y Wang G Li J Liu C Li Z & Sun (2022) Large-factor Micro-CT super-resolution of bone microstructure. Frontiers in Physics 10

DOI: 10.54254/2755-2721/45/20241170

[5] Polat H Özerdem M S Ekici F & Akpola V (2021) Automatic detection and localization of COVID-19 pneumonia using axial computed tomography images and deep convolutional neural networks

[6] Ahmed S T & Kadhem S M (2021) Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging A Survey

[7] Kaggle 2020 https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[8] Goodfellow I J Pouget-Abadie J Mirza M Xu B Warde-Farley D Ozair S Courville A &amp Bengio Y (2014 June 10) Generative Adversarial Networks. arXiv.org. https://arxiv.org/abs/1406.2661

[9] Tensorflow hub. (n.d.). https://tfhub.dev/captain-pool/esrgan-tf2/1

[10] Wang X Yu K Wu S Gu J Liu Y Dong C Loy C C Qiao Y &amp Tang X (2018, September 17) Esrgan: Enhanced super-resolution generative adversarial networks. arXiv.org. https://arxiv.org/abs/1809.00219

[11] Qiu Y Wang J Jin Z et al. 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72: 103323.

[12] Yuan F Zhang Z and Fang Z 2023 An effective CNN and Transformer complementary network for medical image segmentation Pattern Recognition 136 109228