
Chest X-Ray Image Enhancement Using The LED Framework

Ted Arlide¹ Dawid Grzywocz¹ Pascal U¹

Abstract

Using the LED diffusion model as a framework, we evaluate its effectiveness in enhancing low quality chest x-rays (CXR). The dataset used is a subset of the PadChest dataset containing 1000 super-negative exposure (low-quality) CXR images, and 1000 near-zero exposure (high-quality). Following the LED framework, we first implement a CycleGAN to learn a degradation mapping between unpaired high-quality and low-quality images, creating a paired dataset of high-quality images and their generated mapping. Then a conditional diffusion model is trained to learn the inverse. Evaluating on a pre-trained ResNet-18 model, our enhancing diffuser demonstrates improvement in diagnostic accuracy on low-quality CXRs from 56.6% to 66.3%, narrowing most of the gap to high-quality images (73.9%).

1. Introduction

Chest X-ray images frequently suffer from noise, low contrast, blurring, and various other degradation issues. These issues may arise due to varying factors such as poor hardware maintenance and low-dose imaging. Enhancement to these images can offer medical practitioners higher accuracy in diagnosing patients but must be paired with responsible techniques used in regenerating images. The goal of the project is to develop a diffusion-based restoration pipeline for chest X-rays such that low quality images can be restored to sufficient quality for evaluation.

Inspired by the Learning Enhancement from Degradation (LED) framework in (Cheng et al., 2023), which originally was used on retinal fundus images. We proposed utilising this model on chest X-ray images can yield similar promising results. The proposed LED framework first learns a degradation mapping from unpaired high quality images onto low quality ones through a CycleGAN-based (Zhu et al., 2020) degradation network. After this, the framework

embeds the high quality image into a Gaussian distribution with a paired low-quality image, and aims to provide enhancement while preserving relevant characteristics and minimising hallucinations. This coursework will explore the feasibility, existing and relevant literature, architectural details, evaluation metrics, and existing constraints of the proposed model.

1.1. Literature Review

The problem of low quality medical imaging is a topic of high concern among medical researchers, recent success with deep learning models for image processing suggests that these methods may obtain similar successes within the medical domain.

Traditionally, medical image practitioners have relied on the use of global enhancement methods including histogram equalisation (HE) (Abdullah-Al-Wadud et al., 2007) and contrast limited adaptive histogram equalisation (CLAHE) (Zuiderveld et al., 1994). Which both operationally increase the contrast of an image, however, CLAHE further adds an explicit clip limit to every section of the image's histogram so local contrast increases but restricts noise growth. (Nefoussi et al., 2020) compared HE and CLAHE with unsharpened images on a CNN pneumonia classifier and found that both methods gave a boost in recall but in so doing lost precision and had no significant benefit in accuracy on the RSNA Pneumonia dataset (Anouk Stein et al., 2018).

There exist other basic image pre-processing techniques that remain in use for CXR processing, the results in (Hadef et al., 2024) conclude that geometric and colour transformations can offer minor but significant increases in precision, recall and F-score, but did not explore any SOTA deep learning approaches.

(Anand et al., 2023) proposed a deep contrast diffusion network which utilised a multi-level CLAHE process to determine the optimal amount of contrast to diffuse, and then is followed by a CNN that diffuses that contrast back into each region of the original chest X-ray. Though the results are promising, it is still unclear whether contrast alone is enough for medical diagnoses, and room for enhancing additional features for extraction must be considered, which may not be captured by contrast.

¹Computer Science and Electronic Engineering, University of Surrey, Guildford, UK. Correspondence to: Lilian H. Tang <h.tang@surrey.ac.uk>.

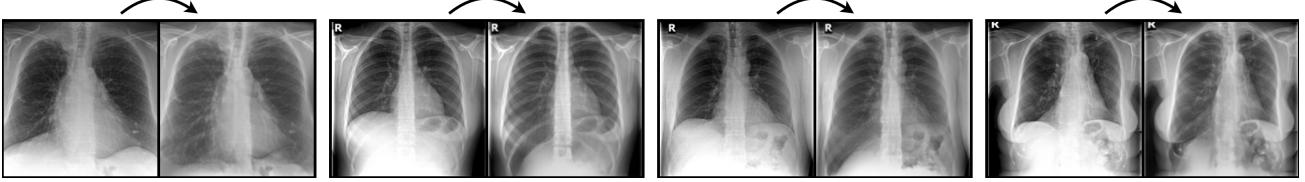


Figure 1. Example of chest-X-ray enhancement using the LED framework.

(Ma et al., 2021) proposed a novel bi-directional GAN called the structure and illumination constrained GAN (StillGAN) for medical image enhancement, and argued that most existing bi-directional GANs are ineffective in capturing local detail. Though this may be true, it is easier to attach lightweight add-ons through the diffusion process (as LED does) to address finer details than to do this through the initial enhancement learning process. Further, (Cheng et al., 2023) show that the degraded fundus images using CycleGAN seem to occupy the same feature space in a t-SNE (Van der Maaten & Hinton, 2008) mapping, suggesting that it may be possible to extract these features in reversing the degradation process.

2. Model Architecture

The LED framework consists of a two step pipeline consisting of an initial bi-directional GAN (in our case using CycleGAN) to learn the data-driven degradation mapping that converts high quality images into realistic low quality counterparts, and the inverse low-to-high enhancement mapping. Followed by a subsequent conditional denoising diffusion model trained on the paired examples. This section outlines the details of these models.

2.1. CycleGAN

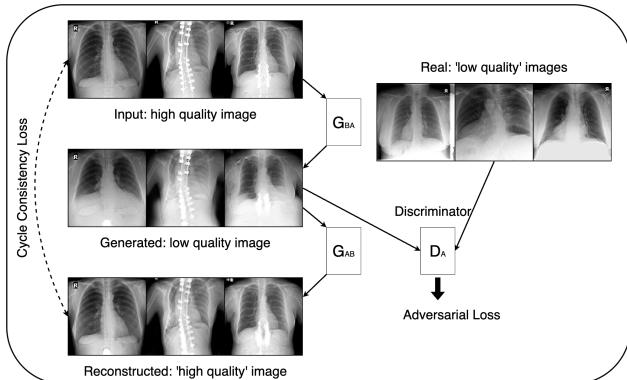


Figure 2. CycleGAN Architecture

As detailed in (Zhu et al., 2020), CycleGAN was initially

proposed as a model for unpaired image-to-image translation but has proved to be versatile in its application for use in medical imagery.

Shown in Figure 2 is an overview of the architecture of the forward process in CycleGAN. G_{BA} is the generator mapping function which maps the high quality images B to low quality images A , and the generator G_{AB} is the corresponding inverse mapping function. D_A represents the discriminator, which tries to distinguish generated images from real ones through the objective function;

$$\begin{aligned} \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) = & \mathbb{E}_{a \sim p_{data}(a)} [\log D_A(a)] \\ & + \mathbb{E}_{b \sim p_{data}(b)} [\log(1 - D_A(G_{BA}(b)))] \end{aligned}$$

D_A serves as an adversary which tries to maximise \mathcal{L}_{GAN} and G_{BA} conversely tries minimising it.

The cycle consistency loss (which has an objective function shown below) simply describes the absolute error between the input and reconstructed images and is the main distinction between CycleGAN and other GAN architectures.

$$\begin{aligned} \mathcal{L}_{cyc}(G_{BA}, G_{AB}) = & \mathbb{E}_{b \sim p_{data}(b)} [\|G_{AB}(G_{BA}(b)) - b\|_1] \\ & + \mathbb{E}_{a \sim p_{data}(a)} [\|G_{BA}(G_{AB}(a)) - a\|_1] \end{aligned}$$

The objective of which is to ensure the image translation cycle should be able to bring the generated image back to the original, which is called forward cycle consistency.

The full objective function is therefore;

$$\begin{aligned} \mathcal{L}(G_{BA}, G_{AB}, D_B, D_A) = & \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) \\ & + \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) \\ & + \lambda \mathcal{L}_{cyc}(G_{BA}, G_{AB}) \end{aligned}$$

where λ controls the relative importance of either objectives. The optimal solution can be given by solving for;

$$G_{BA}^*, G_{AB}^* = \arg \min_{G_{BA}, G_{AB}} \max_{D_B, D_A} \mathcal{L}(G_{BA}, G_{AB}, D_B, D_A)$$

2.2. Denoising Diffusion Probabilistic Model (DDPM)

(Ho et al., 2020) describe the DDPM diffusion model as a parameterized markov chain trained using variational inference to produce samples matching data after a finite amount of time.

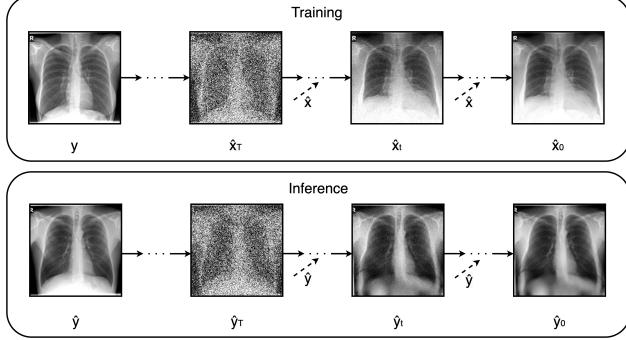


Figure 3. The diffusion process

Simply put, the diffusion model is a generative process that generates samples from gaussian noises through the process of denoising. This is described as the reverse process, the inverse being the forward process which is the gradual addition of noise to images which has the formulation;

$$q(x_t | x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I)$$

Where $q(x_0)$ is the distribution of x_0 which is used to generate the sequence of samples $\{x_t\}_{t=1}^T$, $[\beta_1, \dots, \beta_T]$ denotes a variance schedule for the addition of gaussian noise, and T being a hyper-parameter denoting total number of time steps. The forward process samples x_t at an arbitrary timestep t as given by

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t} x_0, (1 - \alpha_t) I)$$

Where $\alpha_t = \prod_{i=1}^t (1 - \beta_i)$. The reverse process $p(x_{t-1} | x_t)$ is approximated as

$$p_\theta(x_{t-1} | x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \tilde{\beta}_t I)$$

where $\tilde{\beta}_t = \frac{1 - \alpha_{t-1}}{1 - \alpha_t} \beta_t$, and θ represents the learnable parameters. Here the reverse process may be reformulated as

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{1 - \beta_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right)$$

where ϵ_θ is the noise estimation network. (Cheng et al., 2023) derive the objective function of DDPM by reformulating and reparameterizing the forward process to give

$$\mathcal{L}_\theta(x_0, t) = \mathbb{E}_{x_0, \epsilon, t} \left[\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} \epsilon, t)\|^2 \right]$$

2.3. The Overall Framework

The LED framework's proposed diffusion model aims to identify a mapping of low to high quality images using the reverse process. Therefore the model must be conditioned on low quality images x paired with corresponding high quality images y . The degradation model $d(y) = G_{BA}$ generates \hat{x} which trains the diffusion model to learn the reverse process minimizing

$$\mathcal{L}_\theta(\hat{x}, y, t) = \mathbb{E}_{\hat{x}, y, \epsilon, t} \left[\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t} y + \sqrt{1 - \alpha_t} \epsilon, t, \hat{x})\|^2 \right]$$

The inference phase conversely requires an enhancement model $f(x) = G_{AB}$ to generate an initial 'high quality' image \hat{y} from a low-quality image x , which the diffusion model further refines into a final high-quality result \hat{y}_0 .

3. Implementation

Different to the original implementation of LED, we opted to pre-train the CycleGAN degradation model for 80 epochs and set λ to 10 for the full objective function in Section 2.1. We chose our learning rate to be set to 0.0001, Adam (Kingma & Ba, 2014) as our optimizer, and PatchGAN 70 x 70 (Isola et al., 2017) as the discriminator for D_A and D_B .

Our diffusion model is set to train for 150 epochs with early stopping implemented, with learning rate set to 0.00001. Timescale T was set to 1000, with a batch size of 32.

3.1. Datasets

There are currently no publicly available datasets of X-ray images classified into high and low quality. As a result, the dataset we used for this project is a subset of the PadChest dataset (Bustos et al., 2020), which we pruned down to 1000 high quality images and another 1000 low quality images based on the relative exposure. The greyscale CXR images were originally sized at 224x224, but transformed to 256x256 for processing through the diffusion training.

Separate to the high and low quality datasets, we used another subset of data which comprised of two sets of 5000 CXR images labelled as 'normal' and 'COPD', totalling at a 10,000 image dataset used for training the classifier.

3.2. Code Execution

All training, inferencing and evaluation was done on Google Colab (Google LLC, 2025).

4. Results

There are minimal ways of quantitatively evaluating the quality of CXR images. As such, we resort to using classification accuracy to gauge how well or difficult these images are classified correctly. We evaluated performance using the ResNet-18 architecture to correctly classify COPD and normal images as identified by the authors of (Bustos et al., 2020). We present our accuracies over 4 different test sets in Table 1.

Test Set	Test Accuracy
Near Zero (Target)	73.9%
Super Negative (Baseline)	56.6%
Enhanced Super Negative	66.3%

Table 1. Test accuracy across different datasets

The "Super Negative" test set shows the lowest accuracy, suggesting these images are more difficult to classify correctly. The "Enhanced Super Negative" set shows improved accuracy of 9.7% to the "Super Negative" set, indicating that the enhancement applied to these images has a correlating effect on enhancing identifiable medical features.

While structural consistency has not been analysed, the improvement in classification accuracy using ResNet-18 implies that the key features necessary for accurate classification are preserved in the enhanced images.

4.1. Critical Evaluation

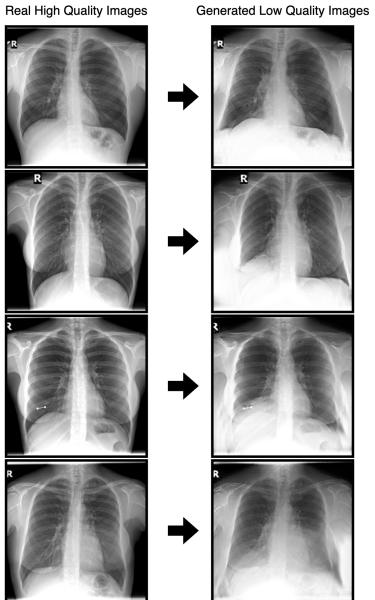


Figure 4. body shape hallucinations in the CycleGAN degradation process

The integrity and structure of medical images are impor-

tant not just for performance evaluations, but it is unethical for medical practitioners to diagnose based on models which change these important features. Unfortunately, due to reasons beyond our control, the dataset we conducted this experiment on had an imbalance of body types between the high- and low-quality images. As such, our degradation model was trained to learn a mapping that, in a select few cases, morphs the skeletal structure of the body (see Figure 4). However, this is only partially observable (and potentially negligible) in the reverse (enhancement) process. Considering the amount of hallucination in the reverse process is shown to be minimal, and that our results show a positive result, we maintain that our implementation has been successful. However, for ethical and performance reasons, more work must be done on better procurement of CXR dataset samples so that there is negligible hallucination.

5. Conclusion

This project adapts the Learning Enhancement from Degradation (LED) framework to the task of CXR image restoration by combining a CycleGAN with a conditional denoising diffusion model. This report explains our motivations and findings from reviewing SOTA methods published by external research (Section 1) and details the architecture of the proposed model (Section 2). We attempted training the model on a rudimentary dataset and yielded promising results, suggesting that the proposed framework has potential for real-world medical use, if properly conducted within safe and responsible implementation parameters.

We show that the results of our evaluation (Section 4) quantitatively conclude that the enhanced images yield significant improvements in being labeled by the ResNet-18 classifier.

5.1. Future Work

Further research must be done on different variations of the LED framework for a more comprehensive overview of the method. This may include a different model to CycleGAN (e.g.; using StillGAN (Ma et al., 2021) may potentially improve the model) to learn the data-driven degradation mapping, and experimenting with different diffusion architectures. The hyperparameters chosen for this project were arbitrarily chosen to match time constraints and computing power, and it is likely that a better enhancer can be produced given more optimized parameters.

Future work will most likely require a more comprehensive and carefully audited dataset to reduce imbalance of features in the degradation mapping. To be commercially viable, this will likely require more commitment on the researcher's end to procure such a dataset, as there are likely ethical concerns as mentioned in Section 4.1.

6. Acknowledgements

Ted worked on the code generating the dataset, equal contribution of the training code for both the CycleGAN and Diffusion models. Also wrote the Abstract for this report.

Dawid worked on the dataset generation code. He also had equal contribution toward writing training code for both the CycleGAN and Diffusion models, wrote the bulk of the evaluation code, and parts of the results section to this report.

Pascal wrote and researched the bulk of this report, including the Introduction, Model Architecture, Implementation and Results. Assisted in coding both the inferencing models for CycleGAN and Diffusion, and was also part of formulating and writing the evaluation code.

References

- Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A. A., and Chae, O. A dynamic histogram equalization for image contrast enhancement. *IEEE transactions on consumer electronics*, 53(2):593–600, 2007.
- Anand, S., Roshan, R., et al. Chest x ray image enhancement using deep contrast diffusion learning. *Optik*, 279:170751, 2023.
- Anouk Stein, M., Wu, C., Carr, C., Shih, G., Dulkowski, J., kalpathy, Chen, L., Prevedello, L., Marc Kohli, M., McDonald, M., Peter, Culliton, P., MD, S. H., and Xia, T. Rsna pneumonia detection challenge. <https://kaggle.com/competitions/rsna-pneumonia-detection-challenge>, 2018. Kaggle.
- Bustos, A., Pertusa, A., Salinas, J.-M., and De La Iglesia-Vaya, M. Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66:101797, 2020.
- Cheng, P., Lin, L., Huang, Y., He, H., Luo, W., and Tang, X. Learning enhancement from degradation: A diffusion model for fundus image enhancement, 2023. URL <https://arxiv.org/abs/2303.04603>.
- Google LLC. Google colaboratory, 2025. URL <https://colab.research.google.com/>.
- Hadef, M., Boulahia, S. Y., and Amamra, A. A comparative analysis of eleven augmentation techniques for enhanced retinal pathology recognition. In *International Conference on AI in Healthcare*, pp. 110–119. Springer, 2024.
- Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Ma, Y., Liu, J., Liu, Y., Fu, H., Hu, Y., Cheng, J., Qi, H., Wu, Y., Zhang, J., and Zhao, Y. Structure and illumination constrained gan for medical image enhancement. *IEEE Transactions on Medical Imaging*, 40(12):3955–3967, 2021.
- Nefoussi, S., Amamra, A., and Amarouche, I. A. A comparative study of chest x-ray image enhancement techniques for pneumonia recognition. In *International Conference on Computing Systems and Applications*, pp. 276–288. Springer, 2020.
- Van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020. URL <https://arxiv.org/abs/1703.10593>.
- Zuiderveld, K. J. et al. Contrast limited adaptive histogram equalization. *Graphics gems*, 4(1):474–485, 1994.