
DAug: Diffusion-based Channel Augmentation for Radiology Image Retrieval and Classification

Supplementary Material

Anonymous Author(s)

Affiliation

Address

email

A Details on Image Generation and Image-to-Image Translation

Common image generation methods include Generative Adversarial Networks (GANs) [1], Variational Autoencoders (VAEs) [2] and Diffusion Models [3]. Diffusion models became the current main stream due to the ease of training and the superior image quality. During training, the diffusion model learns to remove noise from a noisy input and therefore can gradually turn a Gaussian noise into an image. Such denoising steps can be guided by an image classifier trained separately, whose gradients are used to determine the direction of the denoising process, encouraging the output to maximize the probability of a certain class based on the classifier. The result will be an image of the chosen class.

To train a diffusion model, we first conduct a forward diffusion process which gradually adds Gaussian noise to the original image. The forward process has T steps (usually, $T = 1000$), producing a sequence of noisy samples $\mathbf{x}_1, \dots, \mathbf{x}_T$. $\mathbf{x}_0 = \mathbf{I}$ is the original image and $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ becomes a Gaussian noise. Then, a U-Net model was trained to predict the noise added per step, in order to reverse the steps by removing the added noise from the noisy image. The reverse process gradually turns a Gaussian noise to an image. Diffusion models can be guided by an image classifier to generate an image of a particular class. The image classifier is trained to produce class probabilities given a noisy image $f_\phi(y|\mathbf{x}_t, t)$. The gradients of the classifier is used to alter the denoising at each step so that the output maximizes the probability of the target class.

A classifier-guided diffusion model can be used for image-to-image translation. Specifically, we obtain a half-noised image at time step $x = 500$, and then conducts denoising guided by a classifier. As \mathbf{x}_{500} maintains key distinguishable features of the original image, the output still maintains the identity of the original input, but changes it in a way that will be classified to the target class. Related to medical images, this is about converting a healthy CXR to a diseased one, and vice versa.

B Disease Super-classes

In our classifier-guided diffusion model, the classifier was trained on disease super-classes where each super-class consists of one or multiple related diseases. The super-classes were defined with radiologists to align with medical knowledge. For additional rationale behind this definition, CheXpert [4] provides a hierarchical structure of the 14 disease classes, which aligns with our super-class definition. Take super-class #4 as an example, they are grouped together because they are shown as increased density in the X-ray, although for different reasons. The goal is to let the classifier focus only on the appearance instead of attempting to distinguish the cause. For example, in [4], Atelectasis is another type of lung opacity abnormality. We categorized it separately because it looks different (an absence of density).

We show the classifier performance on the super-classes in Table 1. Upon empirical examinations, we found that the quality of the heatmaps are highly correlated with the classifier performance of the

Super Class	Disease classes	AP
1	No Finding	0.631
2	Enlarged Cardiomediatinum, Cardiomegaly	0.885
3	Lung Lesion	0.407
4	Consolidation, Edema, Pneumonia	0.857
5	Atelectasis	0.778
6	Pleural Effusion, Pleural Other	0.746
7	Support Device	0.633

Table 1: Multi-label classifier performance in Average Precision (AP). Please note that this is the classifier which takes in noisy image and is trained to guide the diffusion model. There are totally seven classes, where each one is a super-class consisting of disease classes with similar visual features. For example, super class 4 includes Edema and Pneumonia, which are sub-categories of Consolidation. Training the classifier with merged classes reduces class imbalance and improves performance, and therefore provides better guidance for the diffusion model.

selected class. This observation supports the decision to group sub-classes together to improve the classifier’s robustness. Please note that the performance in the table is expected to be low, as input to the classifier are noisy images instead of the original image (see Appendix A for details about classifier-guided diffusion).

C Pseudo-label Quality

The class labels were not human annotated but were generated with CheXpert, a text classification model which converts a radiology report into disease classes. According to human evaluation in [4], the label quality is claimed to have a 96.9% F1 score.

D Ethical Considerations and Limitations

Our use of the MIMIC-CXR dataset was approved through PhysioNet¹. All authors who accessed the data have obtained the permission.

We identify two limitations of our work. First, to be compatible with the pretrained models, we configure the input image to be 3 channels. The method may achieve even better result if heatmaps of all supported super-classes are used. Second, instead of using the diffusion model’s output heatmaps as input, it could be valuable to explore using these heatmaps as a supervision to avoid introducing extra delay in waiting for the diffusion model to inference. We will make these two limitation as our future work.

References

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [2] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [3] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” *Advances in neural information processing systems*, vol. 32, 2019.
- [4] A. Smit, S. Jain, P. Rajpurkar, A. Pareek, A. Y. Ng, and M. P. Lungren, “Chexbert: combining automatic labelers and expert annotations for accurate radiology report labeling using bert,” *arXiv preprint arXiv:2004.09167*, 2020.

¹<https://physionet.org/content/mimic-cxr/2.0.0/>