Lesion in-and-out painting for medical image augmentation

Anonymous Author(s)
Affiliation
Address
email

Abstract

Deep learning(DL) in the medical imaging field suffers from lack of usable data compared to natural image because of the private and sensitive nature of medical data. Also it is a highly imbalanced data because for almost any disease, medical imaging has more patients not having it rather than having it. To address these problems, synthetic data generation is considered to be a promising solution. In this study, we present Lesion In-aNd-Out Painting (LINOP) to generate synthetic medical images for data augmentation. Generative model based on Mask Aware Transformer (MAT) architecture was used to synthesize lesions onto normal image (inpainting) and synthesis outside of lesion area (outpainting). We train and validate a lesion inpainting pipeline on mammography dataset and a lesion outpainting pipeline on chest X-ray dataset. For mammography, proposed augmentation showed up to 30.3% improvements on mass localization in terms of mAP@50, and for CXR, up to 10.3% improvements on disease classification in terms of AUROC.

1 Introduction

In the middle of the remarkable success of deep learning, there has always been a big, well-defined dataset. For instance, ImageNet[1] plays an important role for developing and validating new deep learning algorithms in computer vision. It is also true for the medical imaging field. The release of multiple, large, publicly available Chest X-ray (CXR) datasets has encouraged research interest and boosted the number of publications [2]. This has increased the dialogue among radiologists and data scientists, which serves to guide and move the field forward [3].

However, in most cases, the medical imaging field suffers from lack of usable data compared to natural image because of the private and sensitive nature of medical data. Also it is a highly imbalanced data because for almost any disease, medical imaging has more patients not having it rather than having it. To address these problems in medical data, synthetic data generated from DL models such as generative adversarial networks (GAN) is considered to be a promising solution. But generating images showing certain types of disease is challenging and difficult to ensure that the imaging feature of certain disease has been correctly generated.

In this study, we present Lesion In-aNd-Out Painting (LINOP) to generate synthetic medical images for data augmentation. Using the inpainting method, it is possible to accurately generate a lesion of the desired size in the desired location. Also, using the outpainting method, the imaging feature of the disease can be clearly preserved. These aspects of inpainting and oupainting approach can improve the reliability and controllability of the generated image compared to the generating entire images. MAT[4] architecture based generative model was used to synthesize lesions onto normal image (inpainting) and synthesis outside of lesion area (outpainting).

Submitted to the Workshop on Deep Generative Models for Health at NeurIPS 2023. Do not distribute.

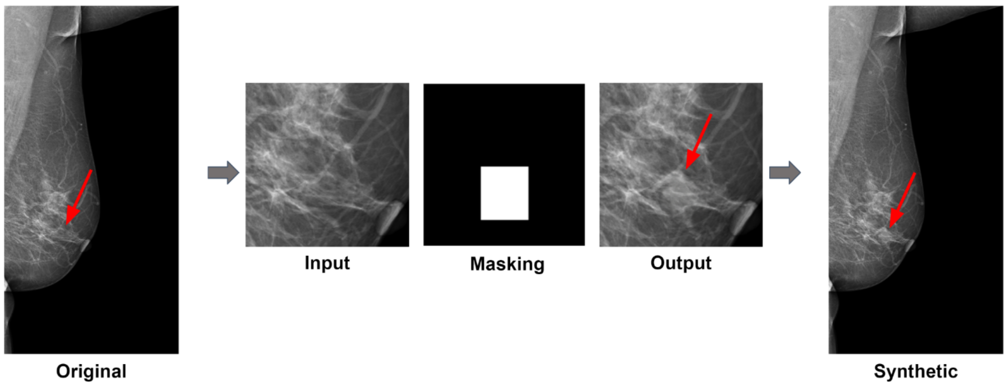

Figure 1: Example of a lesion inpainting on mammography image. The red arrow indicates the synthetic mass.

We train and validate a lesion inpainting pipeline on mammography dataset and a lesion outpainting pipeline on CXR dataset. For mammography, proposed inpainting augmentation showed up to 30.3% improvements on mass localization in terms of mAP@50, and for CXR, proposed outpainting augmentation showed improvement on 4 class disease classification both on accuracy and AUROC, up to 11.2% and 10.3% increases respectively. The example of synthetic data generation process is depicted in Figure 1

2 Methods

2.1 Dataset

VinDr-Mammo VinDr-Mammo[5] is an a large-scale full-field digital mammography dataset of 5,000 four-view exams, which are double read by experienced mammographers to provide cancer assessment and breast density following the Breast Imaging Report and Data System (BI-RADS). Breast abnormalities that require further examination are also marked by bounding box. In this study, we used "Mass" only images to train the lesion inpainting model, and "No finding" images to generate inpainting results for data augmentation.

VinDr-CXR VinDr-CXR[6] is an open large-scale dataset of chest X-rays with radiologist's annotations. The published dataset consists of 18,000 postero-anterior view chest x-ray scans that come with both the localization of critical findings and the classification of common thoracic diseases. These images were annotated by a group of 17 radiologists with at least 8 years of experience for the presence of 22 critical findings and each finding is localized with a bounding box. The dataset is divided into the training set of 15,000 scans and the test set of 3,000 scans. In this study, we used 10,478 of "No finding" only images from training set to train the outpainting model, and 4,522 not "No finding" images to generate outpainting results for data augmentation.

2.2 Mask Aware Transformer (MAT)

To perform realistic-looking inpainting and outpainting, we used a DL-model based on MAT[4] architecture. It consists of a convolutional head designed for tokenization, a transformer body that extracts information through multi-head contextual attention and window shifting, and a Conv-U-Net used for reconstruction. The shifted window enables cross-window connections to conduct non-local interactions and the multi-head contextual attention module employs the feature from the partial valid tokens. In the original work, non-saturating adversarial loss was adopted to enhance the quality and diversity of the texture synthesis. The non-saturating adversarial loss is formulated as,

$$
\begin{aligned}
L_G &= -\mathbb{E}_{\hat{x}}\left[\log\left(D\left(\hat{x}\right)\right)\right], \\
L_D &= -\mathbb{E}_{x}\left[\log\left(D\left(x\right)\right)\right] - \mathbb{E}_{\hat{x}}\left[log\left(1 - D\left(\hat{x}\right)\right)\right],
\end{aligned}
$$

where x and $\hat{x}$ are the real and generated outputs. In addition, the generator was optimized by employing a perceptual loss. The perceptual loss is expressed as

$$L_p = \sum_t s_i \left\| \phi_i\left(\hat{x}\right) - \phi_i\left(x\right) \right\| ,$$

where $\phi_i(\cdot)$ is the layer activation of pre-trained VGG-19[7] network, with scaling coefficients s_i.

3 Results

The MAT-based LINOP model was trained with VinDr-Mammo and VinDr-CXR dataset separately for 1,000 kimgs on 4 NVIDIA V100 GPUs and choosed the checkpoint that shows the best FID score. The best FID score for VinDr-Mammo and VinDr-CXR was 6.76 at 840 kimgs and 3.83 at 960 kimgs, respectively.

3.1 Synthetic data generation for mass detection

We generated synthetic mass onto a normal mammography image using the proposed inpainting pipeline. Normal mammography images were fed into the LINOP model trained with VinDr-Mammo with masking images. The size and location of the masking area determines the size and location of the mass to be generated. The examples with various mask sizes are shown in Figure 2.

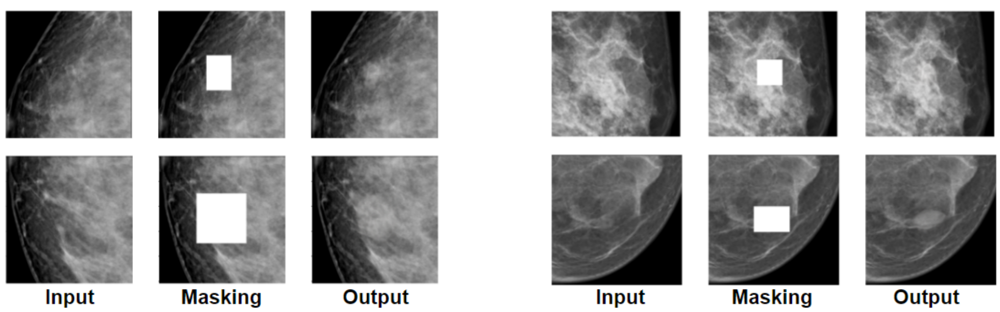

Figure 2: Examples of mass inpainting results with various mask sizes.

To evaluate the effectiveness of inpainting augmentation, mass localization of mammography was performed on images crrespoding to normal and mass cases. Cascade R-CNN was trined for object detection, and original VinDr-Mammo data was randomly split it to 80% training, 10% validation and 10% test set, and the 976 inpainting data was used as training data. Table 1 shows the experimental results of original only baseline and proposed inpainting augmentation for different portions of data used. Proposed augmentation showed improvement on mAP@50, up to 30.3% increases.

Table 1: Mass localization performance of original only and proposed inpainting augmentation.

Dataset	mAP@50
Baseline	0.228
+ 50% of inpainting augmentation	0.2814
+ 100% of inpainting augmentation	**0.297**

3.2 Synthetic data generation for disease classification

We also generated synthetic abnormal chest x-ray images using the proposed outpainting pipeline. Chest X-rays with abnormal findings were fed into the LINOP model trained with VinDr-CXR with masking images. For outpainting, the masking area is where the abnormal finding is, so the generated image also contains the information of abnormal findings from original images while changing the other areas which have no information of abnormal findings. The example of this outpainting process is shown in Figure 3.

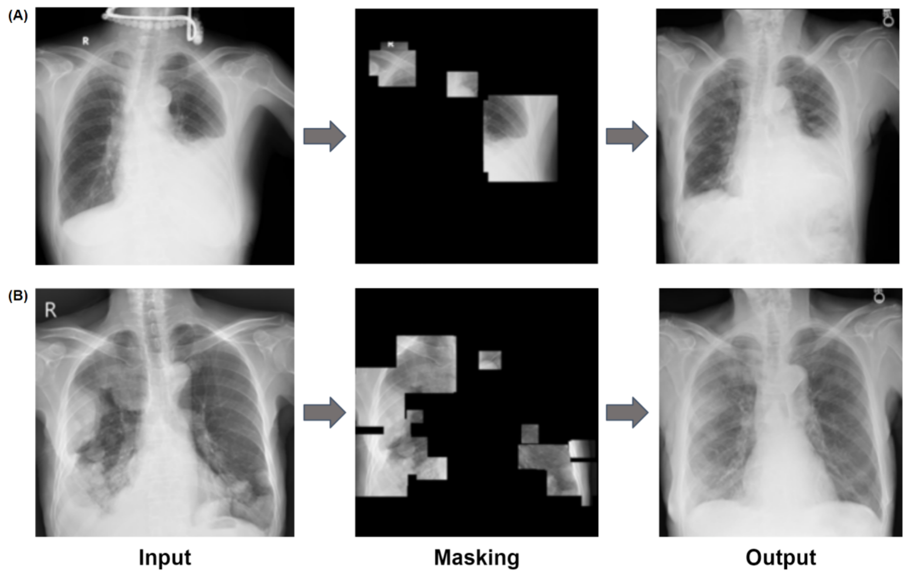

Figure 3: Examples of the proposed outpainting pipeline on chest X-ray image. (A) shows outpainting process of pleural effusion case and (B) shows outpainting process of pneumonia case.

To evaluate the effectiveness of outpainting augmentation, 4 class classification of CXR was performed on images corresponding to normal, pleural effusion, pneumothorax and pneumonia among the total data. Densenet121 was trained for classification and original VinDr-CXR data was randomly split it to 80% training, 10% validation and 10% test set, and the outpainting data was used as training data. Table 2 shows the experimental results of original only baseline and proposed outpainting augmentation for different portions of data used, from 100% to 12.5%. For all portions, proposed augmentation showed improvement both on accuracy and AUROC, up to 11.2% and 10.3% increases respectively.

Table 2: Classification performance of original only and proposed augmentation for different portions of dataset used.

Portion of dataset used:		**100%**	**50%**	**25%**	**12.5%**
Baseline	Accuracy	0.8113	0.7506	0.723	0.661
	AUROC	0.933	0.8845	0.8387	0.8023
Outpainting augmentation	Accuracy	**0.8571**	**0.8349**	**0.7912**	**0.7102**
	AUROC	**0.9638**	**0.9528**	**0.9255**	**0.8619**

4 Conclusion

In this study, We present the LINOP model, lesion inpainting and outpainting model for medical image data augmentation. We generated lesion inpainting results using mammography data with mass and outpainting results with CXR data with various abnormal findings. Using the proposed method, it is possible to accurately generate a lesion of the desired size in the desired location or keep the imaging feature of the disease. This approach can improve the reliability and controllability of the generated image compared to the generating entire images, and is expected to further improve the quality of synthetic data for data augmentation. Further studies on more modalities of medical images and lesions, as well as studies verifying the data augmentation effect of generated images, should be conducted.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[2] Erdi Çallı, Ecem Sogancioglu, Bram van Ginneken, Kicky G van Leeuwen, and Keelin Murphy. Deep learning for chest x-ray analysis: A survey. *Medical Image Analysis*, 72:102125, 2021.

[3] Luciano M Prevedello, Safwan S Halabi, George Shih, Carol C Wu, Marc D Kohli, Falgun H Chokshi, Bradley J Erickson, Jayashree Kalpathy-Cramer, Katherine P Andriole, and Adam E Flanders. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiology: Artificial Intelligence*, 1(1):e180031, 2019.

[4] Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. Mat: Mask-aware transformer for large hole image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10758–10768, 2022.

[5] Hieu Huy Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. Vindr-mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography. *Physionet https://doi.org/10.13026/br2v-7517*, 2022.

[6] Ha Q Nguyen, Khanh Lam, Linh T Le, Hieu H Pham, Dat Q Tran, Dung B Nguyen, Dung D Le, Chi M Pham, Hang TT Tong, Diep H Dinh, et al. Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. *Scientific Data*, 9(1):429, 2022.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.