Position Classifier: Rethinking Position Encoding on Chest X-ray Diseases Identification

Yu-Wen Fang[1] JEFFFANG07@GMAIL.COM
Fang-Yi Su[1] FANGYI@IIR.CSIE.NCKU.EDU.TW
Jung-Hsien Chiang[1] JCHIANG@MAIL.NCKU.EDU.TW
[1] *National Cheng Kung University, Tainan, Taiwan, R.O.C.*

Editors: Under Review for MIDL 2022

Abstract

The patch-based method of chest X-ray interpretation often suffers from the loss of information regarding global anatomy structures. We propose using a simple position classifier to train the model to encode the correct position information. Our model can accurately encode the position information with a 99.87% AUC score on patch positioning. With the help of position information, our model can filter out anatomy structures that are commonly misinterpreted as lesions. We believe the proposed method is both effective and easy to implement in common deep learning-based diseases identification framework with only slight modifications.

Keywords: Position Classifier, Position Embedding, Patch-based, Chest X-ray.

1. Introduction

Artificial intelligence (AI) based chest X-ray interpretation systems often suffer from low accuracy when identifying small lesions, such as lung nodules, and often insufficient training data. (Oh et al., 2020) proposed using the patching technique to increase the model robustness. However, the global anatomy positions of the given chest X-ray image are lost during patching. A common solution is to design position embeddings to encode the position of each patch. (Dosovitskiy et al., 2020) uses learnable parameters to encode the position information. We argue that with a simple position classification loss, the model can accurately encode the position information into the disease representations, further improving the accuracy of the identification of the disease. We demonstrate that by using a simple one-layer MLP as the position classifier, the model can not only accurately encode the position information, but also achieves better results in the identification of the disease.

2. Proposed Method

Assume the input chest X-ray images are split into N overlapping patches with disease labels for each patch, we assign the position label to each patch. A training sample can be viewed as $\{(x_i, y_i^d, y_i^l)\}_{i=1}^N$ where x_i is the patch data, y_i^d and y_i^l represent the patchwise disease label and the position label for the given patch. For feature extracting, a standard ResNet architecture without the final fully connected layer is used. The disease and the position classification share the same encoder, while the feature maps produced by the encoder

are passed through two different classifiers, i.e., the disease classifier, and the position classifier. Finally, the model produces the probabilities of the position $\hat{y}_i^l$ and the diseases $\hat{y}_i^d$ of the given patch. The model is trained end-to-end with the unified loss function $\mathcal{L} = \mathcal{L}_{disease} + \lambda \mathcal{L}_{position}$, where $\mathcal{L}_{disease} = -\sum_i y_i^d \log \hat{y}_i^d$ and $\mathcal{L}_{position} = -\sum_i y_i^l \log \hat{y}_i^l$ are the cross entropy between the model prediction and the label, and λ is the hyperparameter to balance the two losses.

3. Experiments

To demonstrate the improvement in diseases identification, a lung nodule identification model is trained on the VinDr-CXR dataset(Nguyen et al., 2020). Each chest X-ray image is split into 16 overlapping patches, resulting in 6176 training data, and 2000 testing data. We evaluate the model's ability to identify the lung nodules in each patch with Area Under ROC curve (AUC) score on the test set. To show the model's ability to encode the position information, the AUC score for position classification is also calculated. For the hyperparameters, ResNet-50 is chosen as the encoder, and a one-layer MLP is chosen as the position classifier. Finally, λ is set to 0.001 to balance the learning for the lung nodule and the patch position representations.

The model achieved a 0.9987 AUC score on the position classification, which shows that the position classifier can accurately identify the position of the given patch without the manually designed position embeddings. This proves that the encoder can encode the position information into the same embeddings used for diseases identifications, which we argue, will improve the performance of the identification of the disease.

Table 1: AUC score on the VinDr-CXR dataset lung nodule identification.

Method	**Result**	**Improvement**
Baseline	0.8219	0
Learnable Position Embedding	0.8291	0.0072
Coordinates Channels	0.7991	-0.0228
Position Classifier	**0.8342**	**0.0123**

We compare our method with the following methods: baseline (without position loss, i.e. $\lambda = 0$), learnable positional embeddings as in (Dosovitskiy et al., 2020), channels with coordinates information concatenate to the feature maps inspired by (Liu et al., 2018). As shown in Table 1, our model outperforms other methods, further proving that with the position supervision, the model can leverage the position information to better identify lung nodules without the help of additional position embeddings.

As shown in Figure 1, with the position classifier, the model learned to filter out anatomy structures that are commonly misinterpreted as lung nodules, such as the aortic arch or the lung hilum. Note that bounding box labels aren't used during the training phase.

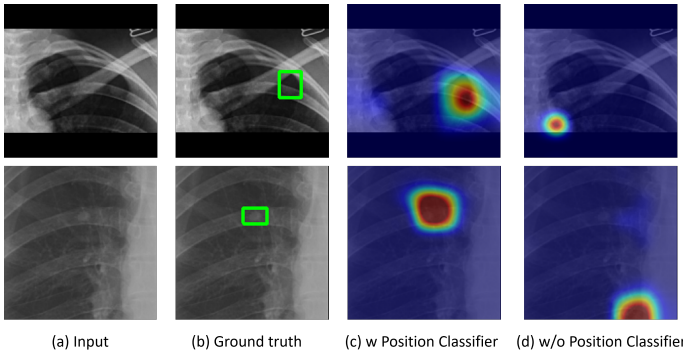

Figure 1: Visualization of Model Predictions using Grad-CAM.

4. Conclusion

In this work, we propose a simple position classifier to encode the position information into the disease representations. We also demonstrate the effectiveness of the position classifier and show that it can help the model to avoid misinterpreting anatomy structures as lesions. We believe that by improving the design of the position classifier, the performance of disease identification can be further improved.

Acknowledgments

We express our sincere gratitude to the Ministry of Science and Technology (MOST) for their support (grant number MOST 110-2634-F-006 -021). In addition, we acknowledge the dedication of Dr. Yau-Lin Tseng, Dr. Wei-Li Huang, and Dr. Chen-Yu Wu of the Chest Surgery Division of the Surgery Department, National Cheng Kung University Hospital, in this research.

References

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31, 2018.

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. *arXiv preprint arXiv:2012.15029*, 2020.

Yujin Oh, Sangjoon Park, and Jong Chul Ye. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging*, 39(8):2688–2700, 2020. doi: 10.1109/TMI.2020.2993291.