Self-Supervised Transfer Learning for COVID-19 Detection from Chest X-Ray Images

Guang Li [†] **Ren Togo** [††] **Takahiro Ogawa** [†††] **Miki Haseyama**[†††]

[†] Graduate School of Information Science and Technology, Hokkaido University, Japan
[††] Education and Research Center for Mathematical and Data Science, Hokkaido University, Japan
[†††] Faculty of Information Science and Technology, Hokkaido University, Japan
E-mail: {guang, togo, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

Abstract

Under the global pandemic of Coronavirus Disease 2019 (COVID-19), computer-aided diagnosis for COVID-19 fast detection and patient triage is becoming critical. In this paper, we propose a new learning scheme called self-supervised transfer learning for COVID-19 detection from chest X-ray images. Considering that the representations learned by self-supervised learning on the target dataset are not enough, transfer learning from different datasets may make up for the shortcomings of self-supervised learning and help to obtain better representations. We show that the knowledge learned from natural images with transfer learning is beneficial for self-supervised learning on the chest X-ray images and boosts representation learning performance for COVID-19 detection. Our method can learn discriminative representations from chest X-ray images by combining transfer learning and self-supervised learning. Experimental results show that our method achieved an HM score of 0.985, an AUC of 0.999, and an Acc of 0.953 on the largest open COVID-19 chest X-ray dataset.

Introduction

The Coronavirus Disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) (Andersen et al. 2020) has emerged as one of the deadliest viruses of the century, resulting in about 242 million people infected with over 4.9 million death worldwide [1] as of 22 October, 2021. In the face of the unprecedented pandemic by COVID-19, medical facilities have faced many challenges, including a critical shortage of medical resources, and many healthcare providers have themselves been infected (Leung et al. 2020; Oh, Park, and Ye 2020). Due to the highly contagious nature of COVID-19, early screening for COVID-19 has become increasingly important to prevent the further spread of the disease and reduce the burden on medical facilities (Liu et al. 2020a).

Real-time Reverse Transcription Polymerase Chain Reaction (RT-PCR) is currently considered the gold standard in COVID-19 confirmation due to high sensitivity and specificity, but it takes several hours to get the result (Tahamtan and Ardebili 2020). As many patients with confirmed COVID-19 present radiological findings of pneumonia, radiologic examinations may be useful for fast diagnosis (Shi et al. 2020). Therefore, chest radiography has the potential for the fast screening of COVID-19 during the patient triage, determining the priority of patient's care to help saturated medical facilities in a pandemic situation (Park et al. 2021).

Recently, self-supervised learning methods have received widespread attention because they can learn good representations without manually designed labels (Jing and Tian 2020; Liu et al. 2020b). For example, the researches (Noroozi and Favaro 2016) and (Gidaris, Singh, and Komodakis 2018) play a jigsaw game on images and predict the rotation degrees of images for learning good representations, respectively. Furthermore, self-supervised learning methods have been shown to be effective on different medical datasets (Zhou et al. 2019; Azizi et al. 2021). Transfer learning (Pan and Yang 2009) is a technique where a model trained on one task is re-purposed on a second related task, which is also usually used in medical image analysis (Raghu et al. 2019). Since the representations learned by self-supervised learning on the target dataset are not enough (Yang et al. 2020), transfer learning from different datasets may make up for the shortcomings of self-supervised learning and help to obtain better representations.

In this paper, we propose a new learning scheme called self-supervised transfer learning for COVID-19 detection from chest X-ray images. Our method consists of three stages, first is supervised pre-training on labeled natural images, next is self-supervised pre-training on unlabeled chest X-ray images, and final is supervised fine-tuning on labeled chest X-ray images. We show that the knowledge learned from natural images with transfer learning is beneficial for self-supervised learning on the chest X-ray images and boosts representation learning performance for COVID-19 detection. Our method can learn discriminative representations from chest X-ray images by combining transfer learning and self-supervised learning. Experimental results show that our method can achieve an HM score of 0.985, an AUC of 0.999, and an Acc of 0.953 on the largest open COVID-19 chest X-ray dataset (Rahman et al. 2021).

Our contributions are summarized as follows.

- We propose a new learning scheme called self-supervised transfer learning for COVID-19 detection from chest X-ray images.
- We realize promising detection results on the largest open COVID-19 chest X-ray dataset, which may help to

[1] https://covid19.who.int

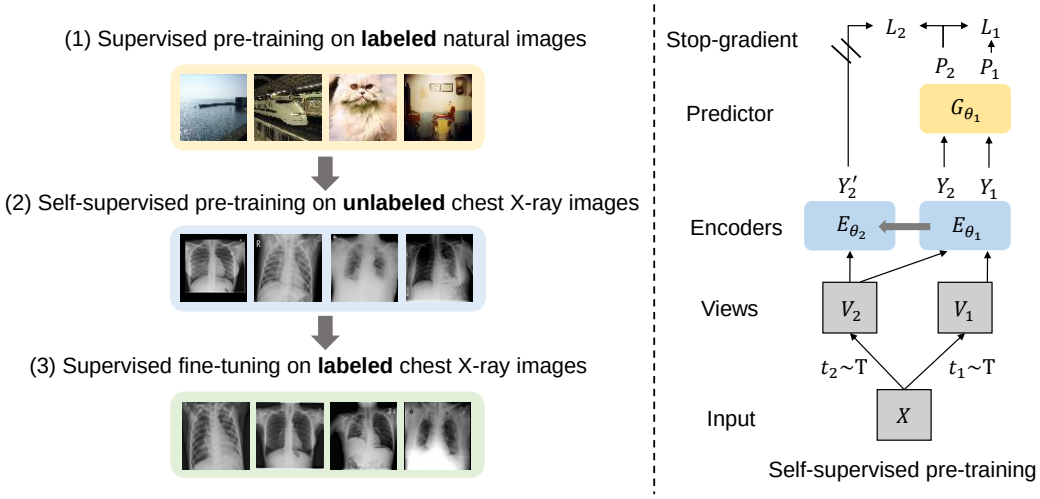


Figure 1: Overview of the proposed method. The left shows the concept illustration of our method and the right shows the self-supervised pre-training process of our method.

decrease transmission of COVID-19 and reduce the burden on healthcare providers and radiologists..

Self-Supervised Transfer Learning

Considering that the representations learned by self-supervised learning on the target dataset are not enough (Yang et al. 2020), transfer learning from different datasets may make up for the shortcomings of self-supervised learning and help to obtain better representations. We propose to learn discriminative representations from chest X-ray images by combining transfer learning and self-supervised learning. We show that the knowledge learned from natural images with transfer learning is beneficial for self-supervised learning on the chest X-ray images and boosts representation learning performance for COVID-19 detection. An overview of the proposed method is shown in Fig. 1. Our method consists of three stages, first is supervised pre-training on labeled natural images (e.g., ImageNet (Deng et al. 2009)), next is self-supervised pre-training on unlabeled chest X-ray images, and final is supervised fine-tuning on labeled chest X-ray images. Combining transfer learning from natural images and self-supervised learning on chest X-ray images, our method can learn discriminative representations for the final fine-tuning.

Based on our previous studies (Li et al. 2021a,b,c), we construct the self-supervised pre-training process as follows. Given an input chest X-ray image X without label information, two transformations t_1 and t_2 are randomly sampled from a distribution T to generate two views $V_1 = t_1(X)$ and $V_2 = t_2(X)$ (He et al. 2020; Chen et al. 2020c). Specifically, these transformations are combined with standard data augmentation methods such as cropping, resizing, flipping, and Gaussian blur (Chen et al. 2020a,b; Tian et al. 2020). Y_1, Y_2 and Y_2' are output representations processed by the encoders. P_1 and P_2 are output representations processed by the predictor. Predictor G_{θ_1} is designed to make the network structure asymmetric, which can prevent learning from collapsing (Grill et al. 2020). Finally, we define the losses L_1, L_2, and L to compare the normalized representations from two views of the same image as follows:

$$L_1 = ||\hat{P}_1 - \hat{P}_2||_2^2 = 2 - 2 \cdot \frac{\langle P_1, P_2 \rangle}{||P_1||_2 \cdot ||P_2||_2}, \quad (1)$$

$$L_2 = ||\hat{P}_2 - \hat{Y}_2'||_2^2 = 2 - 2 \cdot \frac{\langle P_2, Y_2' \rangle}{||P_2||_2 \cdot ||Y_2'||_2}, \quad (2)$$

$$L = L_1 + L_2, \quad (3)$$

where $\hat{P}_i = P_i / ||P_i||_2$ and $\hat{Y}_i' = Y_i' / ||Y_i'||_2$ denote the normalized representations of V_i $(i = 1, 2)$. Then we use the total loss L to update the parameters of the encoder E_{θ_1}

$$\theta_1 \leftarrow \mathrm{Opt}(\theta_1, \nabla_{\theta_1} L, \alpha), \quad (4)$$

where Opt and α denote the optimizer and the learning rate, respectively. The weights of E_{θ_2} are an exponential moving average (Tarvainen and Valpola 2017) of the weights of E_{θ_1} and are updated as follows:

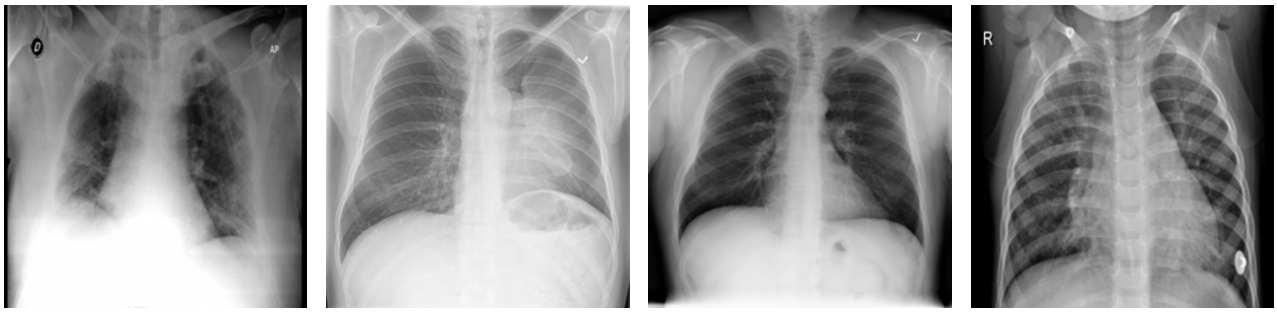
$$\theta_2 \leftarrow \tau\theta_2 + (1 - \tau)\theta_1, \quad (5)$$

where τ denotes the degree of moving average, and we update weights after every iteration. The gradient is not back-propagated through the encoder E_{θ_2} for stable training (Chen and He 2021). With the transfer learning from natural images and self-supervised learning on chest X-ray images, we can learn discriminative representations from chest X-ray images. After the self-supervised learning process, we use the encoder E_{θ_1} to fine-tune on labeled chest X-ray images for COVID-19 detection.

Experiments

Dataset and Settings

The dataset used in our study is the largest open COVID-19 chest X-ray dataset (Rahman et al. 2021). As shown in Table 1, the dataset has 4 classes (*i.e.*, COVID-19, Lung



(a) (b) (c) (d)

Figure 2: Examples of chest X-ray images: (a) COVID-19 , (b) Lung Opacity (c) Normal, and (d) Viral Pneumonia.

Table 1: Details of the COVID-19 chest X-ray dataset [9] used in our study. "C": COVID-19, "L": Lung Opacity,"N": Normal, and "V": Viral Pneumonia.

Class	Total	Training image	Test image
C	3,616	2,893	723
L	6,012	4,810	1,202
N	10,192	8,154	2,038
V	1,345	1,076	269

Opacity, Normal, and Viral Pneumonia) with a total number of 21,165 images. Figure 2 shows examples of chest X-ray images. These chest X-ray images are all gray scale and have a resolution of 224×224 pixels. We randomly select 80% chest X-ray images as the training set and the remaining 20% as the test set. We used sensitivity (Sen), specificity (Spe), their harmonic mean (HM), area under the ROC curve (AUC), and four-class classification accuracy (Acc) as evaluation metrics. For Sen, Spe, HM, and AUC, we took COVID-19 as positive and the others as negative.

The encoders in our method are ResNet50 (He et al. 2016), which is followed by a multilayer perceptron (MLP) whose architecture contains a linear layer with an output size of 512, a batch normalization layer, a ReLU activation function, and a linear layer with an output size of 128. The predictor is MLP whose architecture contains a linear layer with an output size of 4096, a batch normalization layer, a ReLU activation function, and a linear layer with an output size of 256. The optimizer used in our method was an SGD optimizer, whose learning rate α, momentum and weight decay were set to 0.03, 0.9, 0.0004, respectively. Hyperparameter τ was set to 0.996. The batch size and the generated view size were set to 256 and 112, respectively. We took self-supervised learning on the dataset for 40 epochs and fine-tuning on the dataset for 30 epochs, and tested with the average of the last 10 fine-tuning epochs.

We used several state-of-the-art self-supervised learning methods such as Cross (Li et al. 2021b), BYOL (Grill et al. 2020), SimSiam (Chen and He 2021), PIRL (Misra and Maaten 2020), and SimCLR (Chen et al. 2020a), transfer learning (using ImageNet (Deng et al. 2009) pre-trained weights), and training from scratch as comparative methods. Note that PIRL-Jigsaw and PIRL-Rotation are based on jigsaw and rotation pretext tasks, respectively. To verify the effectiveness of our method in a low-shot data regime, we randomly selected objects at 1%, 10%, 50%, and 100% of the total dataset size for the final fine-tuning process. All of our experiments were conducted by using the PyTorch framework with an NVIDIA Tesla P100 GPU that has 16G memory.

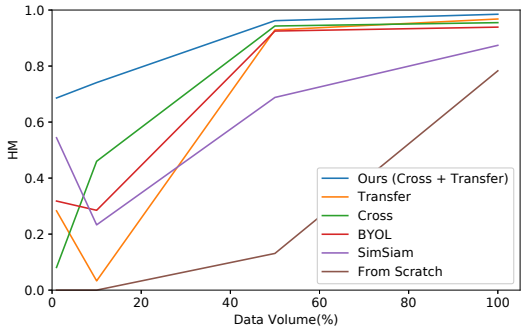
Experimental Results

Test results of COVID-19 detection using all of the training data are shown in Table 2. The results are the average and variance of the last 10 fine-tuning epochs. From Table 2, we can see that our method drastically outperformed other comparative methods and can significantly improve COVID-19 detection performance compared to using self-supervised learning or transfer learning alone. Specifically, when using all of the training data, transfer learning achieved an HM score of 0.968, an AUC of 0.997, and an Acc of 0.936, and the best self-supervised learning method Cross (Li et al. 2021a) achieved an HM score of 0.955, an AUC of 0.995, and an Acc of 0.908. However, when combing transfer learning and self-supervised learning, our method achieved an HM score of 0.985, an AUC of 0.999, and an Acc of 0.953 on the largest open COVID-19 chest X-ray dataset. Experimental results show that the knowledge learned from natural images with transfer learning is beneficial for self-supervised learning on the chest X-ray images and boosts representation learning performance for COVID-19 detection.

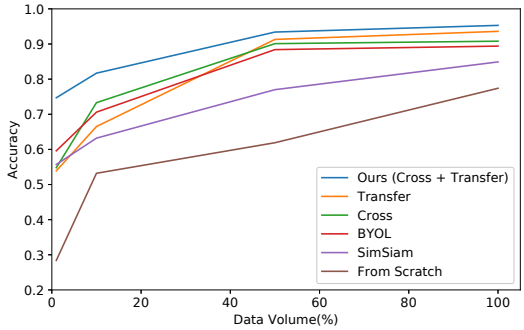
Test results of COVID-19 detection in a low-shot data regime are shown in Fig. 3. The results are the average of the last 10 fine-tuning epochs. As shown in Fig. 3, our method can significantly improve COVID-19 detection performance in different data volumes compared to other methods, and our method can achieve promising detection performance even using only 50% data of the training set. Fig. 4 shows the confusion matrix for the best model of our method with an HM score of 0.988, an AUC of 1.000, and an Acc of 0.956. We can see that our method achieved promising detection results on the largest open COVID-19 chest X-ray dataset.

Table 2: Test results of COVID-19 detection (average of the last 10 fine-tuning epochs).

Method	Sen	Spe	HM	AUC	Acc
Ours (Cross + Transfer)	**0.972±0.003**	**0.997±0.001**	**0.985±0.001**	**0.999±0.000**	**0.953±0.001**
Transfer	0.944±0.004	0.994±0.001	0.968±0.002	0.997±0.000	0.936±0.001
Cross	0.923±0.005	0.991±0.001	0.955±0.002	0.995±0.000	0.908±0.001
BYOL	0.895±0.005	0.987±0.001	0.939±0.003	0.991±0.000	0.894±0.001
SimSiam	0.794±0.013	0.972±0.002	0.874±0.007	0.972±0.000	0.849±0.001
SimCLR	0.778±0.006	0.965±0.002	0.862±0.003	0.996±0.000	0.876±0.001
PIRL-Jigsaw	0.685±0.014	0.973±0.003	0.804±0.009	0.954±0.000	0.821±0.001
PIRL-Rotation	0.760±0.009	0.962±0.002	0.849±0.005	0.960±0.001	0.817±0.001
From Scratch	0.665±0.013	0.954±0.003	0.783±0.008	0.935±0.001	0.774±0.002



(a)



(b)

Figure 3: Test results of COVID-19 detection in different data volumes: (a) HM and (b) Acc.

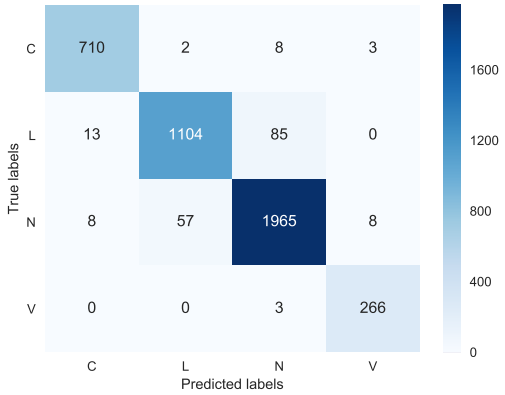


Figure 4: Confusion matrix for the best model of our method with an HM score of 0.988, an AUC of 1.000, and an Acc of 0.956. "C": COVID-19, "L": Lung Opacity, "N": Normal, and "V": Viral Pneumonia.

Discussion

Considering the large number of patients that are being screened due to the global pandemic of COVID-19, the use of artificial intelligence (AI) for computer-aided diagnosis has strong potential to assist in clinical workflow efficiency and reduce the burden on healthcare systems (Vaishya et al. 2020; Imran et al. 2020). Our work was born directly from the clinical needs, and the proposed method can not only use in COVID-19 detection but also in other diseases. Our findings show the promise of simply combining self-supervised learning and transfer learning for COVID-19 detection from chest X-ray images. Although the experimental results are promising, the proposed method should be evaluated on other COVID-19 datasets for any potential bias. Additionally, the proposed method required extensive computational resources, which may not necessarily be available in low-resource settings if further model fine-tuning is needed.

Conclusion

We have proposed a new learning scheme called self-supervised transfer learning for COVID-19 detection from chest X-ray images. We show that the knowledge learned from natural images with transfer learning is beneficial for the self-supervised learning on the chest X-ray images, and boosting the representation learning performance for COVID-19 detection. Our method can learn discriminative representations from chest X-ray images by combining transfer learning and self-supervised learning. Our method achieved promising results on the largest open COVID-19 chest X-ray dataset. Our method may help to decrease transmission of COVID-19 and reduce the burden on healthcare providers and radiologists.

Acknowledgments

This study was supported in part by AMED Grant Number JP21zf0127004. This study was conducted on the Data Science Computing System of Education and Research Center for Mathematical and Data Science, Hokkaido University.

References

Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; and Garry, R. F. 2020. The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4): 450–452.

Azizi, S.; Mustafa, B.; Ryan, F.; Beaver, Z.; Freyberg, J.; Deaton, J.; Loh, A.; Karthikesalingam, A.; Kornblith, S.; Chen, T.; et al. 2021. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 3478–3488.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020a. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Chen, T.; Kornblith, S.; Swersky, K.; Norouzi, M.; and Hinton, G. 2020b. Big Self-Supervised Models are Strong Semi-Supervised Learners. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*.

Chen, X.; Fan, H.; Girshick, R.; and He, K. 2020c. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*.

Chen, X.; and He, K. 2021. Exploring Simple Siamese Representation Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255.

Gidaris, S.; Singh, P.; and Komodakis, N. 2018. Unsupervised representation learning by predicting image rotations. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. 2020. Bootstrap your own latent-a new approach to self-supervised learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9729–9738.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Imran, A.; Posokhova, I.; Qureshi, H. N.; Masood, U.; Riaz, M. S.; Ali, K.; John, C. N.; Hussain, M. I.; and Nabeel, M. 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20: 100378.

Jing, L.; and Tian, Y. 2020. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Leung, M. S. T.; Lin, S. G.; Chow, J.; and Harky, A. 2020. COVID-19 and Oncology: Service transformation during pandemic. *Cancer Medicine*, 9(19): 7161–7171.

Li, G.; Togo, R.; Ogawa, T.; and Haseyama, M. 2021a. Cross-view Self-supervised Learning via Momentum Statistics in Batch Normalization. In *Proceedings of the IEEE International Conference on Consumer Electronics – Taiwan (ICCE-TW)*, 1–2.

Li, G.; Togo, R.; Ogawa, T.; and Haseyama, M. 2021b. Self-Supervised Learning for Gastritis Detection with Gastric X-Ray Images. *arXiv preprint arXiv:2104.02864*.

Li, G.; Togo, R.; Ogawa, T.; and Haseyama, M. 2021c. Triplet Self-Supervised Learning for Gastritis Detection with Scarce Annotations. In *Proceedings of the Global Conference on Consumer Electronics (GCCE)*, 787–788.

Liu, Q.; Luo, D.; Haase, J. E.; Guo, Q.; Wang, X. Q.; Liu, S.; Xia, L.; Liu, Z.; Yang, J.; and Yang, B. X. 2020a. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. *The Lancet Global Health*, 8(6): e790–e798.

Liu, X.; Zhang, F.; Hou, Z.; Wang, Z.; Mian, L.; Zhang, J.; and Tang, J. 2020b. Self-supervised learning: Generative or contrastive. *arXiv preprint arXiv:2006.08218*.

Misra, I.; and Maaten, L. v. d. 2020. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6707–6717.

Noroozi, M.; and Favaro, P. 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 69–84.

Oh, Y.; Park, S.; and Ye, J. C. 2020. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging*, 39(8): 2688–2700.

Pan, S. J.; and Yang, Q. 2009. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10): 1345–1359.

Park, S.; Kim, G.; Oh, Y.; Seo, J. B.; Lee, S. M.; Kim, J. H.; Moon, S.; Lim, J.-K.; and Ye, J. C. 2021. Vision Transformer for COVID-19 CXR Diagnosis using Chest X-ray Feature Corpus. *arXiv preprint arXiv:2103.07055*.

Raghu, M.; Zhang, C.; Kleinberg, J.; and Bengio, S. 2019. Transfusion: Understanding transfer learning for medical imaging. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*.

Rahman, T.; Khandakar, A.; Qiblawey, Y.; Tahir, A.; Kiranyaz, S.; Kashem, S. B. A.; Islam, M. T.; Al Maadeed, S.; Zughaier, S. M.; Khan, M. S.; et al. 2021. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, 132: 104319.

Shi, H.; Han, X.; Jiang, N.; Cao, Y.; Alwalid, O.; Gu, J.; Fan, Y.; and Zheng, C. 2020. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet Infectious Diseases*, 20(4): 425–434.

Tahamtan, A.; and Ardebili, A. 2020. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Review of Molecular Diagnostics*, 20(5): 453–454.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 1195–1204.

Tian, Y.; Sun, C.; Poole, B.; Krishnan, D.; Schmid, C.; and Isola, P. 2020. What makes for good views for contrastive learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 6827–6839.

Vaishya, R.; Javaid, M.; Khan, I. H.; and Haleem, A. 2020. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4): 337–339.

Yang, X.; He, X.; Liang, Y.; Yang, Y.; Zhang, S.; and Xie, P. 2020. Transfer Learning or Self-supervised Learning? A Tale of Two Pretraining Paradigms. *arXiv preprint arXiv:2007.04234*.

Zhou, Z.; Sodha, V.; Siddiquee, M. M. R.; Feng, R.; Tajbakhsh, N.; Gotway, M. B.; and Liang, J. 2019. Models genesis: Generic autodidactic models for 3d medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 384–393.