Chest Radiograph Disentanglement for COVID-19 Outcome Prediction

Lei Zhou[1(✉)], Joseph Bae[2], Huidong Liu[1], Gagandeep Singh[3], Jeremy Green[3], Dimitris Samaras[1], and Prateek Prasanna[2]

[1] Department of Computer Science, Stony Brook University, Stony Brook, NY, USA
lezzhou@cs.stonybrook.edu
[2] Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA
[3] Department of Radiology, Newark Beth Israel Medical Center, Newark, NJ, USA

Abstract. Chest radiographs (CXRs) are often the primary front-line diagnostic imaging modality. Pulmonary diseases manifest as characteristic changes in lung tissue texture rather than anatomical structure. Hence, we expect that studying changes in only lung tissue texture without the influence of possible structure variations would be advantageous for downstream prognostic and predictive modeling tasks. In this paper, we propose a generative framework, Lung Swapping Autoencoder (LSAE), that learns a factorized representation of a CXR to *disentangle* the tissue texture representation from the anatomic structure representation. Upon learning the disentanglement, we leverage LSAE in two applications. 1) After adapting the texture encoder in LSAE to a thoracic disease classification task on the large-scale ChestX-ray14 database (N = 112,120), we achieve a competitive result (mAUC: 79.0%) with unsupervised pre-training. Moreover, when compared with Inception v3 on our multi-institutional COVID-19 dataset, COVOC (N = 340), for a COVID-19 outcome prediction task (estimating need for ventilation), the texture encoder achieves 13% less error with a 77% smaller model size, further demonstrating the efficacy of texture representation for lung diseases. 2) We leverage the LSAE for data augmentation by generating hybrid lung images with textures and labels from the COVOC training data and lung structures from ChestX-ray14. This further improves ventilation outcome prediction on COVOC.

The code is available here: https://github.com/cvlab-stonybrook/LSAE.

Research reported in this publication was enabled by the Renaissance School of Medicine at Stony Brook University's "COVID-19 Data Commons and Analytic Environment", a data quality initiative instituted by the Office of the Dean, and supported by the Department of Biomedical Informatics. Research was supported by SBU OVPR and IEDM seed grant 2019 (P.P, D.S), NIGMS T32GM008444 (J.B). D.S was partially supported by the Partner University Fund, the SUNY2020 Infrastructure Transportation Security Center, and a gift from Adobe.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87234-2_33) contains supplementary material, which is available to authorized users.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12907, pp. 345–355, 2021.
https://doi.org/10.1007/978-3-030-87234-2_33

Keywords: Chest radiographs · Disentanglement · Lung swapping autoencoder · Unsupervised learning

1 Introduction

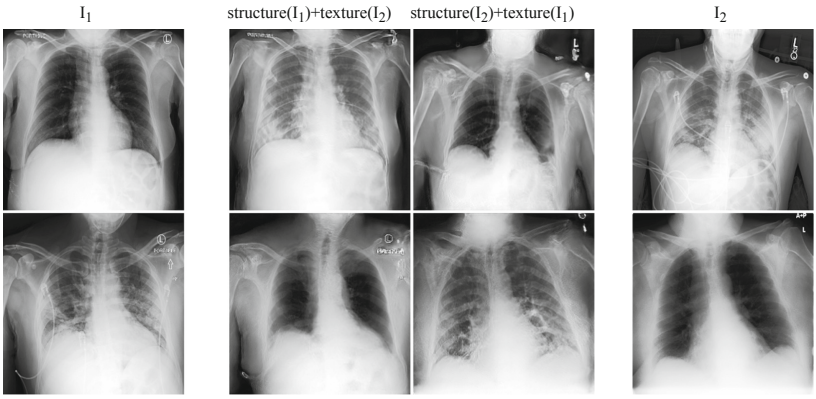

Fig. 1. Lung Swapping Result. Two examples of lung swapping between images in column I_1 and images in column I_2. The Lung Swapping Autoencoder (LSAE) is able to successfully transfer target lung textures without affecting the lung shape. The swapping results are shown in the second and the third columns.

Chest radiographs (CXRs) are the primary diagnostic tool for COVID-19 pneumonia because they are widely available, and have lower risk of cross infection compared to Computed Tomography (CT) scans. Despite these advantages, CXRs are less sensitive to subtle disease changes compared to CT scans [12].

In COVID-19 CXRs, we observe that lung tissue texture may change drastically during hospitalization due to varying infiltrate levels. However, chest anatomy remains mostly unchanged. Therefore, we hypothesize that *disease information is more related to lung tissue texture rather than the anatomical structure of the lung*. To be concise, we use the terms *texture* and *structure* in the rest of the article. Our hypothesis is also supported by recent findings that COVID-19 on CXR is observed as opacities within lung regions, and their extent and location is associated with disease severity and progression [22,25].

Previous medical image analysis in COVID-19 has mostly focused on diagnosis [7,13]. CT-based models are better at predicting COVID-19 outcomes, compared to CXR-based models [1,11,13]. This is largely due to the lack of COVID-19 CXR datasets with relevant endpoints and the limited information that CXRs contain relative to CT scans. Several data augmentation techniques have been proposed for CT scans in the COVID-19 setting [11], but CXR approaches have continued to rely on publicly sourced datasets [2,13]. These datasets tend to be homogeneous, and often lack disease outcome labels (hospitalization, mechanical ventilation requirement, etc.) [2,13]. Generative Adversarial Networks (GANs) [3] and

Autoencoders [10] have been widely used for data augmentation, including in medical image applications [6,19,20]. However, standard GAN-based methods [8,26] are not suitable for CXR generation due to the lack of explicit structure supervision, which can lead to generating distorted shapes.

In this paper, we propose the Lung Swapping AutoEncoder (LSAE), which learns a factorized representation of a CXR to *disentangle* the texture factor from the structure factor. LSAE shares the same core idea as the recently-proposed Swapping AutoEncoder (SAE) [15], namely that a successful disentanglement model should be able to generate a realistic hybrid image that merges the structure of one image and the texture of another. To achieve this, images are encoded as a combination of two latent codes representing structure and texture respectively. The SAE is trained to generate realistic images from the swapped codes of arbitrary image pairs. Moreover, the SAE is also forced to synthesize the target texture supervised by patches sampled from the target image. However, this vanilla SAE does not work well for CXR disentanglement. First, by sampling texture patches from the whole target image, irrelevant out-of-lung textures diminish the effect of the in-lung texture transfer. More importantly, because texture supervision is derived from image patches, irrelevant structure clues may leak into the hybrid image, resulting in undesired lung shape distortion and interference with successful disentanglement.

We address these problems by: 1) Sampling patches from the lung area instead of the whole image for texture supervision (as the infiltrates of interest are located within lung zones), and 2) Adding a patch contrastive loss to explicitly force out-of-lung local patches in the hybrid image to mimic the corresponding patch in the structure image (to prevent structure information in texture patches from leaking into the hybrid image). LSAE, trained on a large public CXR dataset, ChestX-ray14, can generate realistic and plausible hybrid CXRs with one patient's lung structure and another patient's disease texture (see Fig. 1). We further provide quantitative results for disentanglement in the experiment section.

The trained disentanglement model is used in two applications: 1) If textures represent disease infiltrates, the texture encoder in LSAE, Enc^t, should be discriminative in downstream CXR semantic tasks. To prove this, we finetune Enc^t in LSAE on both the large ChestX-ray14 database (N = 112,120) and our multi-institutional COVID-19 outcome prediction dataset, COVOC (N = 340). For thoracic disease classification on ChestX-ray14, Enc^t achieves competitive results without supervised pre-training. On COVOC, compared with a strong baseline Inception v3 [21], Enc^t reduces error by 13% with a much smaller model size. Results show that Enc^t learns effective and transferable representations of lung diseases. 2) To exploit the generative potential of LSAE, we generate hybrid images with textures and labels from COVOC training data and lung structures from ChestX-ray14. Augmenting with these hybrid images further improves ventilation prediction on COVOC.

In summary: 1) LSAE is the first approach to disentangle chest CXRs into structure and texture representations. LSAE succeeds by explicit in-lung texture supervision and out-of-lung distortion suppression, 2) We achieve superior performance for COVID-19 outcome prediction with an efficient model, and 3) We

propose a hybrid image augmentation technique, to further improve ventilation prediction.

2 Methodology

2.1 Swapping AutoEncoder (SAE)

The recent Swapping AutoEncoder (SAE) [15] consists of an encoder Enc, and a decoder Dec, where Enc is composed of a structure branch Enc^s and a texture branch Enc^t to encode the input into structure and texture codes z^s and z^t.

Latent Code Swapping. SAE aims to generate realistic images from swapped latent codes of arbitrary image pairs. Two sampled images I_1 and I_2 are first encoded as (z_1^s, z_1^t) and (z_2^s, z_2^t). Then, the latent codes are swapped to get a hybrid code (z_1^s, z_2^t). Finally, the hybrid code is decoded to yield a hybrid image $I_{\texttt{hybrid}}$ which is expected to maintain the structure of I_1 but present the texture of I_2. We use G to denote the composite generation process as $G(I_1, I_2) \stackrel{\text{def}}{=} Dec(Enc^s(I_1), Enc^t(I_2))$. The objective has reconstruction and GAN loss, i.e., $\mathcal{L}_{\texttt{recon}} = \mathbb{E}_{I_1 \sim \mathbf{X}} \|G(I_1, I_1) - I_1\|_1$, and $\mathcal{L}_{\texttt{G}} = \mathbb{E}_{I_1, I_2 \sim \mathbf{X}} - \log\big(D\big(G(I_1, I_1)\big)\big) - \log\big(D\big(G(I_1, I_2)\big)\big)$, where $\mathbf{X}$ denotes the image training set, and D is a discriminator [3]. Note that the complete GAN loss also includes a discriminator part. To be concise, we only show the generator part in the paper.

Texture Supervision. To ensure $I_{\texttt{hybrid}}$ texture matches that of I_2, SAE samples patches from I_2 to supervise the $I_{\texttt{hybrid}}$ texture. Patch discriminator $D_{\texttt{patch}}$ is trained to distinguish patches in $I_{\texttt{hybrid}}$ from patches in I_2. SAE is trained adversarially to generate a $I_{\texttt{hybrid}}$ whose patches can confuse $D_{\texttt{patch}}$ by mimicking the texture of I_2. The texture loss is formulated as $\mathcal{L}_{\texttt{tex}} = \mathbb{E}_{\substack{\tau_1, \tau_2 \sim \mathcal{T} \\ I_1, I_2 \sim \mathbf{X}}}\big[- \log\big(D_{\texttt{patch}}\big(\tau_1\big(I_2\big), \tau_2\big(G(I_1, I_2)\big)\big)\big)\big]$ where $\mathcal{T}$ is the distribution of multi-scale random cropping, and τ_1 and τ_2 are two random operations sampled from $\mathcal{T}$.

2.2 Lung Swapping AutoEncoder (LSAE)

Using SAE to generate a hybrid CXR from an image pair does not work as desired. First, the disease level in $I_{\texttt{hybrid}}$ is usually diminished when compared with I_2. We hypothesize that the out-of-lung irrelevant texture patterns may hinder target texture synthesis in $I_{\texttt{hybrid}}$. Results in Table 1 (left) support our analysis. Second, since there is no structure supervision in SAE, $I_{\texttt{hybrid}}$ shows undesired lung shape distortion towards I_2. This leads us to design two new features, in-lung texture supervision and out-of-lung structural distortion suppression. We name the new model as *Lung Swapping AutoEncoder* (LSAE) (Fig. 2).

In-lung Texture Supervision. As infiltrates on CXRs are observed inside lung regions, we should only sample texture patches from within the lung zone. Guided by a lung segmentation mask, we rewrite the texture supervision loss as,

$$\mathcal{L}_{\texttt{inTex}} = \mathop{\mathbb{E}}_{\substack{\tau_1, \tau_2 \sim \mathcal{T}^{\texttt{in}}_{\texttt{LungMask}} \\ I_1, I_2 \sim \mathbf{X}}} \left[- \log\left(D_{\texttt{patch}}\Big(\tau_1\big(I_2\big), \tau_2\big(G(I_1, I_2)\big)\Big)\right)\right] \tag{1}$$

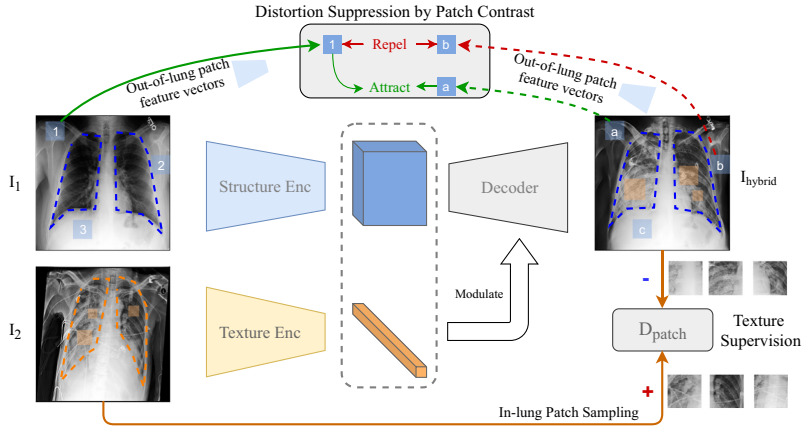

Fig. 2. Lung Swapping AutoEncoder (LSAE) consists of a structure encoder, a texture encoder, and a decoder. After inputing images (I_1, I_2), the LSAE generates a hyrbid image with factorized representations from I_1 and I_2 respectively. To ensure the texture in $I_{\texttt{hybrid}}$ matches I_2, we adversarially train a patch discriminator $D_{\texttt{patch}}$ to supervise the texture synthesis within lungs. To ensure $I_{\texttt{hybrid}}$ maintains the structure of I_1, we apply a patch contrastive loss outside the lungs to minimize structural distortion.

where $\mathcal{T}$ in $\mathcal{L}_{\texttt{tex}}$ is replaced by $\mathcal{T}^{\texttt{in}}_{\texttt{LungMask}}$, indicating that we only sample patches from in-lung regions to train the generator G.

Out-of-Lung Structural Distortion Suppression. To suppress structural distortion, we introduce a patch contrastive loss [14] in LSAE. The patch contrastive loss, based on Noise Contrastive Estimation [4], preserves image content in image-to-image translation tasks. We first encode $I_{\texttt{hybrid}}$ with structure encoder Enc^s. Then, we randomly sample a feature vector q_i from the h_{th} layer's output $Enc^s_h(I_{\texttt{hybrid}})$, and a bag of feature vectors $\{p^+, p^-_1, p^-_2, \cdots, p^-_{N-1}\}$ from $Enc^s_h(I_1)$, where p^+ is from the corresponding position with q_i, and $\{p^-\}$ from other positions. The loss objective is to force q_i from $I_{\texttt{hybrid}}$ to attract to the corresponding feature vector p^+ but repel the others $\{p^-\}_{N-1}$ from I_1. Therefore, we write it as a cross-entropy loss by regarding the process as an N-way classification problem where q_i is the query and p^+ is the target:

$$l_{\texttt{NCE}}(q_i, \{p\}_N) = -\log\left(\frac{\exp(q_i \cdot p^+/\alpha)}{\exp(q_i \cdot p^+/\alpha) + \sum_{j=1}^{N-1}\exp(q_i \cdot p^-_j/\alpha)}\right) \tag{2}$$

where α is the temperature, and $\{p\}_N = \{p^+\} \cup \{p^-\}_{N-1}$. To prevent $l_{\texttt{NCE}}$ from affecting the in-lung texture transfer, we only apply it outside the lung region. Thus, the structural distortion suppression loss is expressed as

$$\mathcal{L}_{\texttt{sup}} = \mathop{\mathbb{E}}_{\substack{\{\tau\}_N \sim \mathcal{T}^{\texttt{out}}_{\texttt{LungMask}} \\ I_1, I_2 \in \mathbf{X}}} \sum_h^H \sum_i^N l_{\texttt{NCE}}\Big(\tau_i\big(Enc^s_h(I_{\texttt{hybrid}})\big), \{\tau\big(Enc^s_h(I_1)\big)\}_N\Big) \tag{3}$$

where $\mathcal{T}^{\texttt{out}}_{\texttt{LungMask}}$ is the distribution of position sampling outside the lung region guided by the mask, and H is the number of layers we apply the loss to.

Overall Objective. Combining the above, we term the final loss as $\mathcal{L} = \mathcal{L}_{\texttt{recon}} + \lambda_1 \mathcal{L}_{\texttt{G}} + \lambda_2 \mathcal{L}_{\texttt{inTex}} + \lambda_3 \mathcal{L}_{\texttt{sup}}$, where λ_1, λ_2 and λ_3 are tunnable hyper-parameters.

2.3 Hybrid Image Augmentation

If the texture in I_2 can be transferred to $I_{\texttt{hybrid}}$, we assume the label (e.g., ventilation) of I_2 is also attached to $I_{\texttt{hybrid}}$. Based on this hypothesis, we design a new data augmentation method: Given an image $I^{\texttt{dst}}$ in the target domain, e.g., COVOC, we sample K images $\{I^{\texttt{src}}\}$ from a source domain, e.g., ChestX-ray14. Then, we use LSAE to take $I^{\texttt{dst}}$ as the texture template and images from $\{I^{\texttt{src}}\}$ as structure templates to generate K hybrid images $\{I_{\texttt{hybrid}}\}$. We label $\{I_{\texttt{hybrid}}\}$ with the label of $I^{\texttt{dst}}$. Following this protocol, we can enlarge the training set in target domain K times. Experimental results show the augmentation method can improve performance further on the COVID-19 ventilation prediction task.

3 Experimental Design and Results

3.1 Dataset Description

ChestX-ray14 [24] is a large-scale CXR database consisting of 112,120 frontal-view CXRs from 32,717 patients. We report all the results based on the official split which consists of training (∼70%), validation (∼10%), and testing (∼20%) sets. Images from the same patient will only appear in one of the sets.

COVID-19 Outcome (COVOC) is a COVID-19 CXR dataset curated from two institutions. It consists of 340 CXRs from 327 COVID-19 patients acquired upon disease presentation [9]. Each CXR in COVOC is labeled based upon whether the patient required mechanical ventilation (henceforth ventilation) or not. We separate COVOC randomly into 3 splits. Each split has 250 samples for training, 30 for validation, and 60 for testing.

3.2 Implementation

For LSAE, both decoder and discriminator architectures follow StyleGAN2. Encoders are built with residual blocks [5]. The texture encoder outputs a flattened vector while the structure encoder's output preserves spatial dimension. The input image size is of 256×256. Patch sizes sampled for texture supervision vary from 16×16 to 64×64. We set $\lambda_1 = 0.5$, $\lambda_2 = 1$, and $\lambda_3 = 1$. The temperature α is set to 0.07. LSAE is optimized by Adam with learning rate 1e-3. Our code is based on PyTorch 1.7. We preprocess all CXRs with histogram equalization.

3.3 Hybrid CXR Generation

We start with a larger dataset, i.e., ChestX-ray14, to pre-train LSAE by learning to generate hybrid CXRs from swapped latent code.

Experimental Settings. We train LSAE on the training set of ChestX-ray14 with a batch of 16 images for 150K iterations. To evaluate performance, we define a lung swapping task by creating two sets of 9,000 images, $\{I^+\}$ and $\{I^-\}$, from the test set of ChestX-ray14. $\{I^+\}$ are sampled from the images diagnosed with at least one of the diseases: infiltration, pneumonia, and fibrosis, which are tightly related with COVID-19 [18]. In contrast, $\{I^-\}$ are sampled from healthy lungs. We generate hybrid image sets by mixing texture and structure from $\{I^+\}$ and $\{I^-\}$ in both directions. We measure the distance of hybrid image set between both the target texture image set and the structure image set.

Evaluation Protocol. First, we propose a new metric, Masked SIFID, to measure the disease level distance between $I_{\texttt{hybrid}}$ and I_2. Masked SIFID is based on the SIFID [17] which calculates the FID distance between two images in the Inception v3 feature space. To customize it to disease level distance, we only consider features within the lung region. Additionally, we use the ChestX-ray14 pre-trained Inception v3 (mAUC: 79.56%) to infer features. Second, to quantify structural distortion, we use lung segmentation metrics as surrogates. Given a lung segmentation model Seg, we can compute segmentation metrics by treating $Seg(G(I_1, I_2))$ as the query and $Seg(I_1)$ as the ground truth. Third, we solicit feedback from 5 radiologists through a 4-question survey. Please refer to the supplementary material for details. In the first two questions regarding image quality, radiologists misconstrue 56% of the generated images as real, and 74% of the hybrid image patches as real. The third question is to verify the correlation between Masked SIFID and disease level distance. When picking which of two query images is closer to the reference, 78.67% of the radiologists' answers match with Masked SIFID. The fourth question is to ascertain whether our method can transfer disease correctly. When the radiologists were shown the original and the hybrid image with the same texture, 60% of the images passed the test.

Results. To get reference values, we first report initial results by directly comparing $\{I^-\}$ with $\{I^+\}$. In Table 1 (left), LSAE achieves lower Masked SIFID when compared with SAE, which demonstrates that our design of in-lung texture supervision works as expected. LSAE also outperforms SAE by a large margin in the segmentation metrics, and achieves over 90% in all segmentation metrics, which proves out-of-lung patch contrastive loss can effectively suppress structural distortion. Please refer to the supplementary material for details.

3.4 Semantic Prediction in CXRs by Texture Encoder

Based on hypothesis that pulmonary diseases are tightly related to CXR texture, the texture encoder in a well-trained LSAE should be discriminative on CXR semantic tasks. To verify this hypothesis, we evaluate texture encoder, Enc^t, on both lung disease classification and COVID-19 outcome prediction tasks.

Table 1. Left. On the hybrid image generation task, LSAE surpasses SAE in both texture synthesis and structure maintenance. We report the average of two directions' texture transfer. **Right.** For the 14 pulmonary diseases classification task on ChestX-ray14, the texture encoder in LSAE achieves competitive results with a smaller model size. *Note that the data split in [16] is not released. We reimplemented CheXNet on the official split. The Inception v3 model is also self-implemented.

Method	Masked SIFID ↓	mIoU ↑	Pixel Acc ↑	Dice ↑
Init	0.0335	0.60	0.82	0.72
SAE	0.0257	0.76	0.91	0.85
LSAE	**0.0245**	**0.91**	**0.97**	**0.95**

Pre-train	Method	Params	mAUC ↑
Sup	CXR14-R50[24]	23M	0.745
	ChestNet [23]	60M	0.781
	CheXNet* [16]	7M	0.789
	Inception v3	22M	**0.796**
Unsup	Enc^t	**5M**	0.790

Table 2. COVOC Outcome Prediction. We evaluate models with Balanced Error Rate (BER) and average AUC (mAUC). The texture encoder surpasses Inception v3 by a large margin. It also outperforms the texture encoder in a baseline model SAE, which demonstrates that better disentanglement does lead to better discrimination. We report mean and std of 5 random runs.

BER(%)↓	Inception v3	Enc^t in SAE	Enc^t in LSAE
split 1	20.25 ± 1.46	20.25 ± 1.63	19.00 ± 1.84
split 2	19.25 ± 3.67	20.50 ± 1.12	17.75 ± 1.66
split 3	17.75 ± 3.48	14.00 ± 1.85	12.75 ± 2.15
Avg	19.08	18.25	16.50

mAUC(%)↑	Inception v3	Enc^t in SAE	Enc^t in LSAE
split 1	85.45 ± 1.89	89.03 ± 2.16	89.17 ± 0.68
split 2	86.02 ± 1.27	85.63 ± 1.77	87.07 ± 1.91
split 3	89.12 ± 1.38	92.60 ± 1.25	95.00 ± 0.29
Avg	86.86	89.09	90.41

Disease Classification on ChestX-ray14. We finetune Enc^t on the training set of ChestX-ray14 with 14 disease labels, and report mean AUC in Table 1 (right). We achieve competitive results, despite having a smaller model size.

Outcome Prediction on COVOC. Considering the possible domain discrepancy between ChestX-ray14 and COVOC, we first adapt LSAE to the new domain by training to generate hybrid images on COVOC for 10K iterations. Then, we evaluate the texture encoder Enc^t on the outcome prediction task by further finetuning. As COVOC is imbalanced, we report the ventilation prediction Balanced Error Rate (BER) together with mAUC in Table 2. Compared with Inception v3, Enc^t reduces BER by 13.5%, and improves mAUC by 4.1%. When comparing with the texture encoder in baseline model SAE, LSAE also

performs better. It demonstrates that better disentanglement leads to better discrimination. We also report the prediction of mortality in Table 2 in SM.

3.5 Data Augmentation with Hybrid Images

As described in Sect. 2.3, we generate hybrid images to augment the training data in COVOC. To control the training budget, we set $K = 2$, i.e., the training data of COVOC is augmented 2 times. To avoid introducing irrelevant diseases, we only sample structure images from the healthy lungs in ChestX-ray14. With the same experimental setup with Table 2, the augmentation method can reduce error rate further from 16.50% to 15.67%, and improve mAUC from 90.41% to 92.04% on the ventilation prediction task. Moreover, we implement Mixup independently for training on COVOC which achieves 16.41% BER/90.82% mAUC. Our method still shows superior performance.

4 Conclusion

We propose LSAE to disentangle texture from structure in CXR images, enabling analysis of disease-associated textural changes in COVID-19 and other pulmonary diseases. We also create a data augmentation technique which synthesizes images with structural and textural information from two CXRs. This technique can be used to augment data for machine learning applications. Our resulting predictive model for mechanical ventilation in COVID-19 patients outperforms conventional methods and may have clinical significance in enabling improved decision making. We will apply our texture disentanglement and data augmentation methods to study other pulmonary diseases in the future.

Acknowledgement. We thank Amit Gupta, Nicole Sakla, and Rishabh Gattu for their expert evaluation of the generated radiographs.

References

1. Bae, J., et al.: Predicting mechanical ventilation requirement and mortality in COVID-19 using radiomics and deep learning on chest radiographs: a multi-institutional study. arXiv preprint arXiv:2007.08028 (2020)
2. Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M.: COVID-19 image data collection: Prospective predictions are the future. arXiv:2006.11988 (2020)
3. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
4. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 297–304. JMLR Workshop and Conference Proceedings (2010)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

6. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433 (2016)
7. Hu, Q., Drukker, K., Giger, M.L.: Role of standard and soft tissue chest radiography images in COVID-19 diagnosis using deep learning. In: Medical Imaging 2021: Computer-Aided Diagnosis, vol. 11597, p. 1159704. International Society for Optics and Photonics, February 2021
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017
9. Konwer, A., et al.: Predicting COVID-19 lung infiltrate progression on chest radiographs using spatio-temporal LSTM based encoder-decoder network. In: Medical Imaging with Deep Learning (2021)
10. Li, Y., Liu, S., Yang, J., Yang, M.H.: Generative face completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3911–3919 (2017)
11. Li, Z., et al.: A novel multiple instance learning framework for COVID-19 severity assessment via data augmentation and self-supervised learning. arXiv:2102.03837 [cs, eess], February 2021
12. Litmanovich, D.E., Chung, M., Kirkbride, R.R., Kicska, G., Kanne, J.P.: Review of chest radiograph findings of COVID-19 pneumonia and suggested reporting language. J. Thorac. Imaging **35**(6), 354–360 (2020)
13. López-Cabrera, J.D., Orozco-Morales, R., Portal-Diaz, J.A., Lovelle-Enríquez, O., Pérez-Díaz, M.: Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging. Heal. Technol. **11**(2), 411–424 (2021). https://doi.org/10.1007/s12553-021-00520-2
14. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 319–345. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_19
15. Park, T., Zhu, J.Y., Wang, O., Lu, J., Shechtman, E., Efros, A.A., Zhang, R.: Swapping autoencoder for deep image manipulation. arXiv preprint arXiv:2007.00653 (2020)
16. Rajpurkar, P., et al.: Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
17. Rott Shaham, T., Dekel, T., Michaeli, T.: Singan: learning a generative model from a single natural image. In: IEEE International Conference on Computer Vision (ICCV) (2019)
18. Salehi, S., Abedi, A., Balakrishnan, S., Gholamrezanezhad, A.: Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. Am. J. Roentgenol. **215**(1), 87–93 (2020)
19. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci. Rep. **9**(1), 1–9 (2019)
20. Shin, H., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. CoRR abs/1807.10225 (2018)
21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

22. Toussie, D., et al.: Clinical and chest radiography features determine patient outcomes in young and middle age adults with COVID-19. Radiology 201754 (2020)
23. Wang, H., Xia, Y.: Chestnet: a deep neural network for classification of thoracic diseases on chest radiography. arXiv preprint arXiv:1807.03058 (2018)
24. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106 (2017)
25. Wong, H.Y.F., et al.: Frequency and distribution of chest radiographic findings in COVID-19 positive patients. Radiology 201160 (2020)
26. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)