

CXR Segmentation by AdaIN-based Domain Adaptation and Knowledge Distillation: Supplementary Materials

Yujin Oh^[0000-0003-4319-8435] and Jong Chul Ye^[0000-0001-9763-9609]

Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea jong.ye@kaist.ac.kr

S1 Applications for DL-based Automatic CXR Analysis

Current deep learning (DL)-based CXR analysis tasks not just diagnose disease but provide explainable results such like saliency map or quantified severity level to assist clinicians [2, 3, 8, 9]. Accordingly, we further investigated applications of the proposed method using COVID-19 pneumonia CXR dataset [6]. For generating saliency map, we referenced a public source code¹. In addition, for severity quantification, we followed array-based methods [9, 14].

Fig. S1 shows that *Proposed* and *Proposed+ ℓ_{self}* models provide the most suitable segmentation mask for saliency map or severity array generation. The results demonstrate that our methods can be utilized for various automatic CXR analysis tasks.

S2 Network Architecture

We provide details of the proposed framework, as shown in Table S1, S2, S3 and S4. Note that each input and output dimension for the domain adaption task is single channel (C_{da}), whereas the segmentation task requires two channels (C_{seg}) to separate foreground and background channels.

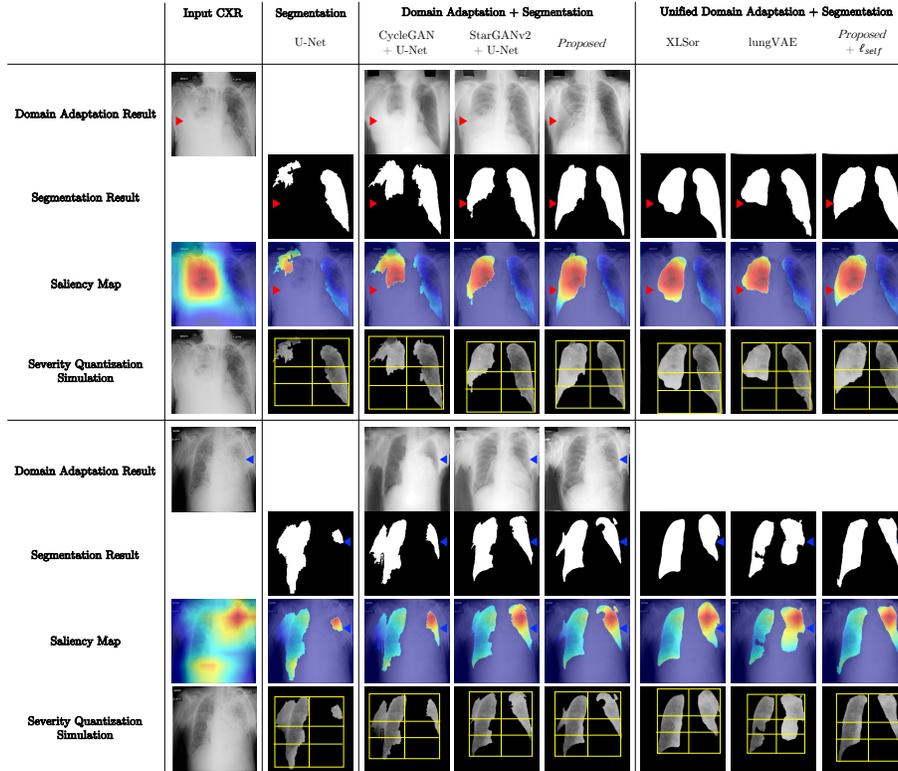
S3 Adaptive Instance Normalization

AdaIN has been proposed as an extension of the instance normalization [4]. AdaIN layer receives a content input x and a AdaIN code a , and simply aligns the channel-wise mean and variance of x to match those of desirable style by:

$$AdaIN = f(a) \left(\frac{x - \mu(x)}{\sigma(x)} \right) + g(a) \quad (1)$$

where f and g compute affine parameters from AdaIN code a , μ and σ represent mean and variance, respectively. In this way, AdaIN simply scales the normalized content input with $\sigma(a)$, and shifts with $\mu(a)$.

¹ <https://github.com/priyavrat-misra/xrays-and-gradcam.git>



Supplementary Figure S1. Model performance comparison on various DL-based automatic CXR analysis. Red and blue triangles indicate highly consolidated lung regions in CXRs. Yellow boxes indicate array-based six subdivisions of lung for severity quantification.

Supplementary Table S1. Generator network architecture.

Module	Layer	Norm	Resample	Input dimension ($C \times H \times W$)
In	Conv 1×1	-	-	$C_{da} \times 256 \times 256$
Encoder	ResBlock $\times 4$	AdaIN	Down	$64 \times 256 \times 256$
	ResBlock $\times 2$	AdaIN	-	$512 \times 16 \times 16$
Decoder	ResBlock $\times 2$	AdaIN	-	$512 \times 16 \times 16$
	ResBlock $\times 4$	AdaIN	Up	$512 \times 16 \times 16$
Unshared	Norm	IN	-	
	Leaky ReLU	-	-	$64 \times 256 \times 256$
	Conv 1×1	-	-	
Output	-	-	-	$C_{da} \times 256 \times 256$
	-	-	-	$C_{seg} \times 256 \times 256$

Supplementary Table S2. AdaIN code generator architecture.

Module	Layer	Input dimension (C)
In	Latent z	4
Shared	Linear \times 3	4
	Linear \times 1	512
Unshared	Linear \times 3	512
	Linear \times 1	512
Output	-	$16 \times K$

Supplementary Table S3. Style encoder architecture.

Module	Layer	Input channel (C)	Input size (H \times W)
Unshared	Conv 1×1	C_{da}, C_{seg}	256×256
Shared	ResBlock \times 6	64	256×256
	Leaky ReLU	512	4×4
	Conv 4×4	512	4×4
	Leaky ReLU	512	1×1
Unshared	Linear	512	1×1
Output	-	$16 \times K$	-

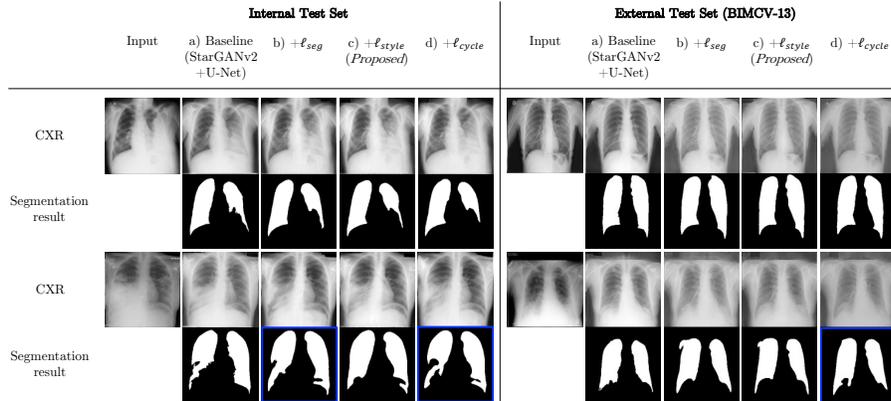
Supplementary Table S4. Discriminator architecture.

Module	Layer	Input channel (C_{da})	Input size (H \times W)
Shared	Conv 1×1	1	256×256
	ResBlock \times 6	64	256×256
	Leaky ReLU	512	4×4
	Conv 4×4	512	4×4
	Leaky ReLU	512	1×1
Unshared	Conv 1×1	512	1×1
Output	-	$1 \times K$	-

S4 Ablation Study

For ablation study, we analyzed contribution of different losses for the supervised segmentation task. We cumulatively added each loss to the baseline model, and compared segmentation performance on abnormal CXR.

Fig. S2 compares performance of each configuration. In configuration (b) with additional segmentation loss, domain adaptation performance was superior to the baseline. However, we observed that the segmentation results have several concave regions (blue boxes), which failed to resemble the general shape of lung structure. In configuration (c) with additional style loss, we observed that the segmentation results resemble normal lung better than (b), thanks to the style loss, which can extract common features of normal and abnormal CXR via the shared layer of the style encoder. In configuration (d) with additional cycle-



Supplementary Figure S2. Ablation study with different losses for the segmentation task.

consistency loss, we observed rather degraded lung segmentation performance as depicted as blue boxes, which have more concave regions compared to that of (c). The cycle-consistency loss, which tries to revert the generated lung mask back to the original image, may disturb the network to segment extremely consolidated lung regions.

Based on the ablation study results, we set the configuration (c) as our supervised segmentation loss.

S5 Domain Adaptation Loss

Domain adaptation loss basically follows StarGANv2 [1], given by

$$\ell_{da}(G, F_e, F_d, S, D) = \quad (2)$$

$$\ell_{adv}(G, D, F_d, S) \quad (3)$$

$$+ \lambda_{cycle} \ell_{cycle}(G, S) \quad (4)$$

$$+ \lambda_{style} \ell_{style}(G, S) \quad (5)$$

$$- \lambda_{div} \ell_{div}(G, F_d, S), \quad (6)$$

where $\lambda_{cycle}, \lambda_{style}$ and λ_{div} are hyper-parameters and ℓ_{adv} is the adversarial loss defined by

$$\ell_{adv}(G, D, F_d, S) = \mathbb{E}_{s \sim P_S} [\log D_S(s)] + \mathbb{E}_{s \sim P_S} [\log(1 - D_T(G(s, a_{da}^T)))] , \quad (7)$$

where \mathcal{S} and \mathcal{T} are source and target domains, which are chosen randomly from \mathcal{X} and \mathcal{Y} so that all domain combinations can be considered. Furthermore, the learnable part of the AdaIN code a_{da}^T is generated either from the encoder AdaIN coder generator F_d or the style encoder S given a reference target $t \in \mathcal{T}$.

The cycle-consistency loss ℓ_{cycle} is defined as follows:

$$\ell_{cycle}(G, S) = \mathbb{E}_{s \sim P_S} [\|x - G(G(s, a_{da}^T), a_{da}^S)\|_1]. \quad (8)$$

Similar to the cycle-consistency loss ℓ_{cycle} for the images, we introduce the style loss ℓ_{style} to enforce the cycle-consistency in the AdaIN code domain. More specifically, once a fake image is generated using a domain-specific AdaIN code, the style encoder with the fake image as an input should reproduce the original AdaIN code. This can be achieved by minimizing the following style loss:

$$\ell_{style}(G, S) = \mathbb{E}_{s \sim P_S} [\|a_{da}^T - S(G(s, a_{da}^T))\|_1]. \quad (9)$$

Finally, to make the generated fake images diverse, the difference between two fake images that are generated by two different AdaIN codes should be maximized. This can be achieved by maximizing the following loss:

$$\ell_{div}(G, F_d, S) = \mathbb{E}_{s \sim P_S} [\|G(s, a_{da}^T) - G(s, a'_{da}^T)\|_1], \quad (10)$$

where an additional a'_{da}^T is generated either from the encoder AdaIN coder generator F_d or the style encoder S given an additional reference image.

S6 Implementation Details

The proposed method was implemented with PyTorch library [10]. We applied Adam optimizer [7] to train the models and set the batch size 1. The model was trained using a NVIDIA GeForce GTX 1080 Ti GPU. Hyper parameters were chosen to be $\lambda_{cycle} = 2$, $\lambda_{style} = 1$, $\lambda_{div} = 1$, $\lambda_{seg} = 5$, $\lambda_{inter} = 10$ and $\lambda_{intra} = 1$. Learning rate was optimized to 0.0001. Once training iteration reaches certain fixed iteration points throughout the total iterations, the learning rate was reduced by factor of 10.

The network was trained for 20K iterations to simultaneously train domain adaptation and supervised segmentation tasks. We adopted early stopping strategy based on validation performance of abnormal chest X-ray radiograph (CXR) segmentation results. In terms of the training sequence, the self-supervised training started after training the domain adaptation and supervised segmentation tasks until they guaranteed certain performances. For self-supervised learning, the network was continued to be trained in self-supervised manner for additional 5K iterations.

At the inference phase, as for post-processing steps, two largest contours were automatically selected based on contour areas, and any holes within each contour were filled. The post-processing technique was identically applied to all the comparative model outputs for fair comparison.

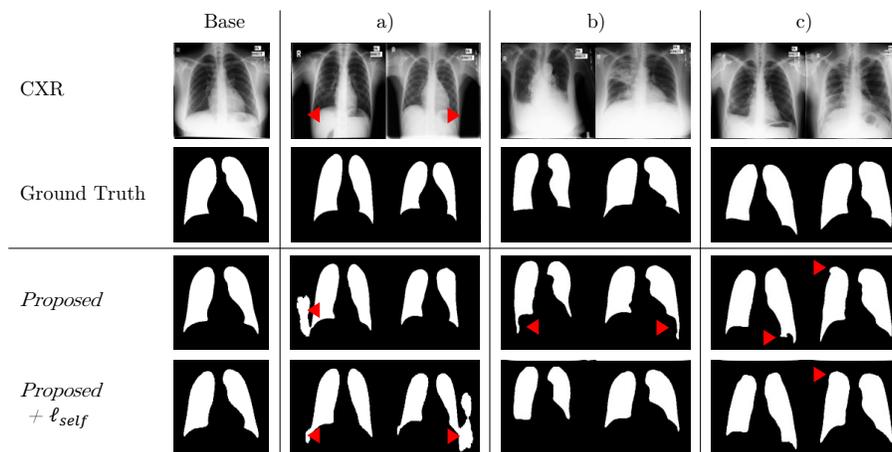
S7 Comparative Model Implementations

For comparative study, baseline models for domain adaptation and supervised segmentation tasks, i.e., CycleGAN [15], MUNIT [5], StarGANv2 [1] and U-Net [11], were trained with identical conditions to that of the proposed model.

For comparing performance of the unified domain adaptation and segmentation network, we inferred pre-trained networks optimized for abnormal CXR segmentation, i.e., XLSor [13] and lungVAE [12], by utilizing their official source codes.²³

S8 Error Analysis

We analyzed typical error cases, which failed to be segmented, and the error cases were grouped into three categories: (a) Over-segmentation on background pixels, (b) distorted lung shape, and (c) distorted lung boundary, as shown in Fig. S3.



Supplementary Figure S3. Representative error cases. (a) Over-segmentation on background pixels, (b) distorted lung shape and (c) distorted lung boundary. Base indicates standard segmentation result.

S9 Computational Costs

The proposed unified framework costs less training computation resources, compared to training individual domain adaptation and segmentation networks. Table S5 shows total network parameters utilized for either training or inference of comparative networks.

Once the model is trained, only the generator with pre-built AdaIN codes are used at the inference phase, thus the model costs only the single generator.

² <https://github.com/raghavian/lungVAE>

³ https://github.com/rsummers11/CADLab/tree/master/Lung_Segmentation_XLS
or

As shown in Table S5, *Proposed* and *Proposed+ ℓ_{self}* models need the least number of network parameters, with most promising segmentation performance. Specifically, compared to *Proposed* model, *Proposed+ ℓ_{self}* model only needs a single inference without preceding domain adaptation task, with comparable segmentation performance to that of *Proposed* model.

Supplementary Table S5. Number of trainable and inference parameters.

Method	Training			Inference(S)			Inference Time
	Generator(S)	Others	Total	Generator(S)	Others	Total	
SS							
U-Net [11]	29M	-	29M	29M	-	29M	× 1
XLSor [13]	71M	-	71M	71M	-	71M	× 1
DA							
CycleGAN [15]	-	29M	29M	-	-	-	
MUNIT [5]	-	47M	47M	-	-	-	
StarGANv2 [1]	-	78M	78M	-	-	-	
DA+SS							
CycleGAN + U-Net	29M	29M	58M	29M	11M	40M	× 2
StarGANv2 + U-Net	29M	78M	127M	29M	34M	63M	× 2
<i>Proposed</i>	34M	45M	79M	34M	-	34M	× 2
UDS/Self							
MUNIT + XLSor	71M	47M	118M	71M	-	71M	× 1
lungVAE [12]	34M	-	34M	34M	-	34M	× 1
<i>Proposed+ℓ_{self}</i>	34M	46M	80M	34M	-	34M	× 1

Note: SS, supervised segmentation; DA, domain adaptation; UDS, unified DA+SS; Self, self-supervised segmentation; (S), segmentation task; Others, other module parameters.

References

1. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: Stargan v2: Diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8188–8197 (2020)
2. Frid-Adar, M., Amer, R., Gozes, O., Nassar, J., Greenspan, H.: Covid-19 in cxr: From detection and severity scoring to patient disease monitoring. *IEEE journal of biomedical and health informatics* **25**(6), 1892–1903 (2021)
3. Hou, J., Gao, T.: Explainable dcnn based chest x-ray image analysis and classification for covid-19 pneumonia detection. *Scientific Reports* **11**(1), 1–15 (2021)
4. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510 (2017)
5. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European conference on computer vision (ECCV). pp. 172–189 (2018)
6. de la Iglesia Vayá, M., Saborit, J.M., Montell, J.A., Pertusa, A., Bustos, A., Cazorla, M., Galant, J., Barber, X., Orozco-Beltrán, D., García-García, F., Caparrós, M., González, G., Salinas, J.M.: Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients (2020)

7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **284**(2), 574–582 (2017)
9. Park, S., Kim, G., Oh, Y., Seo, J.B., Lee, S.M., Kim, J.H., Moon, S., Lim, J.K., Ye, J.C.: Vision transformer using low-level chest x-ray feature corpus for covid-19 diagnosis and severity quantification. arXiv preprint arXiv:2104.07235 (2021)
10. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in PyTorch. In: *NIPS Autodiff Workshop* (2017)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
12. Selvan, R., Dam, E.B., Detlefsen, N.S., Rischel, S., Sheng, K., Nielsen, M., Pai, A.: Lung segmentation from chest x-rays using variational data imputation (2020)
13. Tang, Y.B., Tang, Y.X., Xiao, J., Summers, R.M.: Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation. In: *International Conference on Medical Imaging with Deep Learning*. pp. 457–467. PMLR (2019)
14. Toussie, D., Voutsinas, N., Finkelstein, M., Cedillo, M.A., Manna, S., Maron, S.Z., Jacobi, A., Chung, M., Bernheim, A., Eber, C., et al.: Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with covid-19. *Radiology* **297**(1), E197–E206 (2020)
15. 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 (ICCV)* (Oct 2017)