

Does Non-COVID19 Lung Lesion Help? Investigating Transferability in COVID-19 CT Image Segmentation

Yixin Wang, Yao Zhang, Yang Liu, Jiang Tian, Cheng Zhong, Zhongchao Shi, Yang Zhang, Zhiqiang He

Abstract—Coronavirus disease 2019 (COVID-19) is a highly contagious virus spreading all around the world. Deep learning has been adopted as an effective technique to aid COVID-19 detection and segmentation from computed tomography (CT) images. The major challenge lies in the inadequate public COVID-19 datasets. Recently, transfer learning has become a widely used technique that leverages the knowledge gained while solving one problem and applying it to a different but related problem. However, it remains unclear whether various non-COVID19 lung lesions could contribute to segmenting COVID-19 infection areas and how to better conduct this transfer procedure. This paper provides a way to understand the transferability of non-COVID19 lung lesions. Based on a publicly available COVID-19 CT dataset and three public non-COVID19 datasets, we evaluate four transfer learning methods using 3D U-Net as a standard encoder-decoder method. The results reveal the benefits of transferring knowledge from non-COVID19 lung lesions, and learning from multiple lung lesion datasets can extract more general features, leading to accurate and robust pre-trained models. We further show the capability of the encoder to learn feature representations of lung lesions, which improves segmentation accuracy and facilitates training convergence. In addition, our proposed Multi-encoder learning method incorporates transferred lung lesion features from non-COVID19 datasets effectively and achieves significant improvement. These findings promote new insights into transfer learning for COVID-19 CT image segmentation, which can also be further generalized to other medical tasks.

Index Terms—COVID-19, CT image, Segmentation, Transfer learning.

I. INTRODUCTION

IN December 2019, the coronavirus disease 2019 (COVID-19) broke out and has become a global challenge since then. This new virus was spreading rapidly, affecting countries worldwide. Up to 13 June 2020, 7,533,177 identified cases of COVID-19 have been reported in over 216 countries and territories, resulting in 422,770 deaths. This severe disease has been declared as a Public Health Emergency of International Concern by the World Health Organization (WHO).

A gold standard method to screen the COVID-19 patients is the real-time reverse transcription-polymerase chain

reaction (rRT-PCR) [1], which requires inconvenient repeated tests. Researchers found that chest computed tomography (CT) images show strong ability to capture ground glass and bilateral patchy shadows which are typical CT features in affected patients [2]. Thus, chest CT can help identify distinguishing patterns and features in patients. Given that traditional CT imaging analysis methods are time-consuming and laborious, it is of great importance to develop artificial intelligence (AI) systems to aid COVID-19 diagnosis [3].

Segmentation in CT slices is a key component of the diagnostic work-up for patients with COVID-19 infection in clinical practice [4]. Recently, deep learning with CNNs has showed significant performance improvements on the automatic detection and automatic extraction of essential features from CT images, related to the diagnosis of the Covid-19. Though deep learning has made great progress in medical image segmentation, it remains a challenging task in the field of COVID-19 lung infection, as existing public datasets on COVID-19 are relatively small and weakly labelled. Thus, training deep networks from scratch with inadequate data and task-specific nature such as COVID-19 infection may lead to over-fitting and poor generalization.

Transfer learning is an effective method to solve this problem, which helps to leverage knowledge and latent features from other datasets and avoids over-parameterization. In transfer learning, successful deep learning models such as ResNet, DenseNet and GoogLeNet have been trained on large datasets such as ImageNet. These pre-trained models have proven impressive performance on natural image downstream tasks. However, there exists no large-scale annotated medical image datasets as data acquisition is difficult, and high-quality annotations are expensive. Recent research [5] shows that transfer learning from natural image datasets to medical tasks produces very limited performance gain. In particular, though with large medical datasets for pre-training, transfer learning in the task of COVID-19 infection segmentation is still much more difficult: 1) The shape, texture and position of COVID-19 infections are in high variation. 2) Existing large medical CT datasets differ in domains with COVID-19 datasets. Thus, similar in domain, a pre-trained model from lung lesions may share more knowledge with COVID-19 infection and learn some general-purpose visual representations for lung lesions.

Evidence shows that larger datasets are necessarily better for pre-training and the diversity of datasets is extremely important [6]. In medical domain, pre-training from medical datasets, especially chest CT datasets tends to be more

Corresponding authors: Zhiqiang He and Yang Zhang.

Yixin Wang, Yao Zhang and Yang Liu are with Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China.

Jiang Tian, Cheng Zhong and Zhongchao Shi are with AI Lab, Lenovo Research, Beijing, China.

Zhiqiang He and Yang Zhang are with Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China, and also with Lenovo Corporate Research & Development, Lenovo Ltd., Beijing, China.

homogeneous compared to non-medical and other medical areas' data. Thus, non-COVID19 lung lesion CT imaging manifestations may serve as potential profit for COVID-19 segmentation. Existing works have proven that multi-task training through simply fusing different lesions with COVID-19 affects model's representation ability for COVID-19 infection segmentation [7]. Therefore, with limited COVID-19 datasets, 1) whether these non-COVID19 lesions help and to what extent they can contribute to COVID-19 infection segmentation? 2) How to train a better pre-trained model using these non-COVID19 datasets for transfer learning? 3) In what manners can the pre-trained models fitted on non-COVID19 lesions be effectively transferred to COVID-19?

In this paper, we aim to answer the above questions which are significant for COVID-19 segmentation. To our best knowledge, this is the first study to explore the transferability of non-COVID19 datasets for COVID-19 CT images segmentation.

Our contributions are as follows.

- We experimentally assess the extent of contributions from non-COVID19 to COVID-19 infection segmentation. We found that despite the disparity between non-COVID19 lung lesion images and COVID-19 infection images, pre-training on the large scale well-annotated lung lesions may still be transferred to benefit COVID-19 recognition and segmentation.
- We conduct extensive experiments using various non-COVID19 lung lesion datasets. Although pre-training on a single non-COVID19 dataset is unstable among different transfer strategies, learning from different non-COVID19 lesions demonstrates promising performance since such a multi-task learning process can share knowledge from different related tasks and discover common and general representations of lung lesions.
- We design four different transfer learning strategies with various performance gains and training time costs: Continual Learning, Body Fine-tuning, Pre-trained Lesion Representations and Multi-encoder Learning. The Multi-encoder strategy effectively combines both non-COVID19 and COVID-19 features and shows the best performance. We further conclude that it is possible to freeze the encoder and train only the decoder on COVID-19 tasks during fine-tuning, which provides significant performance gains and fast convergence speed.

II. RELATED WORKS

In this section, we review the recent related research on COVID-19 CT images, COVID-19 segmentation and transfer learning for COVID-19.

A. Research on COVID-19 CT Images.

CT is widely used in the screening and diagnosis of COVID-19 since ground-glass opacities and bilateral patchy shadows are the most relative imaging features in pneumonia associated with infections. Recent research on CT-based diagnosis for COVID-19 has indicated great performance. Compared with traditional CT images processing, artificial intelligence (AI)

serves as a core technology, enabling a more accurate and efficient solution. The machine learning-based CT radiomics models such as random forest, logistic regression showed feasibility and accuracy for predicting hospital stay in patients affected in the work of [8]. Machine learning method was also adopted by Tang *et al.* [9] to realize automatic severity assessment (non-severe or severe) of COVID-19 based on chest CT images, and to explore the severity-related features from the resulting assessment model. Gozes *et al.* [10] presented a system that utilized robust 2D and 3D deep learning models, modifying and adapting existing AI models and combining them with clinical understanding. Huang *et al.* [11] monitored the disease progression and understood the temporal evolution of COVID-19 quantitatively using serial CT scan by an automated deep learning method.

B. Research on COVID-19 Segmentation.

Segmentation is an essential step in AI-based image processing and analysis [4]. In particular, segmenting the regions of interests (ROIs) of COVID-19 infections is crucial and helpful for doctors to make further assessment and quantification. However, manual contouring of these infections is time-consuming and tedious. Although plenty of methods have been explored on COVID-19 diagnosis and classification, there are very few works on the segmentation of COVID-19 infection due to its great annotation challenges.

Shan *et al.* [12] developed a DL-based system for automatic segmentation and quantification of infection regions and adopted a human-in-the-loop (HITL) strategy to accelerate the manual delineation of CT training. Zheng *et al.* [13] designed a weakly-supervised deep learning algorithm to investigate the potential of a deep learning-based model for automatic COVID-19 detection on chest CT volumes using the weak patient-level label. Based on semi-supervised learning, Fan *et al.* [14] presented a COVID-19 Lung Infection Segmentation Deep Network (Inf-Net) for CT slices based on randomly selected propagation. Yan *et al.* [15] introduced a feature variation block which adaptively adjusted the global properties of the features for segmenting COVID-19 infection.

C. Research on COVID-19 Transfer Learning.

In transfer learning, deep models are first trained on large datasets such as ImageNet, then these pre-trained models are fine-tuned on different downstream tasks. Several studies [6], [16], [17] have investigated transfer learning methodologies on deep neural networks applied to medical image analysis tasks. Plenty of works used networks pre-trained on natural images to extract features and followed by another classifier [18], [19]. Carneiro *et al.* [19] replaced the fully connected layers of a pre-trained model with a new logistic layer and only trained the appended layer, yielding promising results for classification of unregistered multi-view mammograms. Other studies performed layer fine-tuning on the pre-trained networks for adapting the learned features to the target domain. In [20], CNNs were pre-trained as a feature generator for chest pathology identification. Gao *et al.* [21] fine-tuned all the layers of a pre-trained CNN to classify interstitial lung

diseases. Ghafoorian *et al.* [22] trained a CNN on legacy MR images of brain and evaluated the performance of the domain-adapted network on the same task with images from a different domain.

Due to the limited labeled COVID-19 data, several transfer learning methods have been applied to address this problem. Chouhan *et al.* [23] proposed an ensemble approach of transfer learning using pre-trained models trained on ImageNet. Researchers used five different pre-trained models and analyzed their performance in chest X-ray images. In the work of [24], several deep CNNs were employed for automatic COVID-19 infection detection from X-ray images through tuning parameters. Majeed *et al.* [25] presented a critical analysis for 12 off-the-shelf CNN models and proposed a simple CNN architecture with a small number of parameters that performed well on distinguishing COVID-19 from normal X-rays. By combining three different models which were fine-tuned on 3 datasets, Misra *et al.* [26] designed a multi-channel ensemble TL method based on ResNet-18 in such a way that the model could extract more relevant features for each class and identify COVID-19 features more accurately from the X-ray images.

Even though the above studies on transfer learning present inspiring achievement for COVID-19 research, there are several limitations: 1) They only focus on the ensemble of existing CNNs, but ignore the contribution of various datasets for pre-training. 2) Their studies are limited to X-ray dataset and only dedicate to COVID-19 detection and classification. 3) They lack an in-depth study on the transferability of different transfer manners related to COVID-19 infections. Our work contributes to a much more difficult task of semantic segmentation in COVID-19 CT images. We explore the transferability from the perspective of transferring knowledge from various non-COVID19 lung lesions. Moreover, we investigate better transfer methods to assist COVID-19 segmentation.

III. METHODOLOGY

In this section, we briefly describe our backbone network, then introduce the multi-task learning method to get a multi-lesion pre-trained model. We further give detailed illustration on four transfer strategies employed in our work.

A. Encoder-Decoder Network

The U-Net is a commonly used network for medical semantic segmentation. As an advanced architecture, it has a U-shape like structure with an encoding and a decoding signal path. The encoder serves as a contraction to capture semantically image contextual features. The decoder is a symmetric expanding path recovering spatial information. The two paths are connected using skip connections on each same level, which recombine with essential high-resolution features from the encoding path. In this work, we make some minor changes following nnU-Net [27]. Original batch normalization and ReLU are replaced by instance normalization and leaky ReLU. What's more, deep supervision loss [28] is aggregated to obtain multi-level deep supervision and facilitate the training process.

B. Pre-trained Multi-lesion Learning

Multi-task learning is an effective method to share knowledge among different related tasks. As for segmentation tasks, the performance of each task highly depends on the similarity among these tasks. Due to the large domain distance among those existing non-COVID19 lung lesion datasets, fusing these lesions to train a multi-task segmentation model tends to underperform on each single task. However, this multi-task training can exploit the shared knowledge which is essential for learning some general-purpose visual representations about lung lesions. Therefore, in our work, besides separately training segmentation models on each non-COVID19 dataset as pre-trained models for transfer learning, we provide a multi-lesion model learning from multiple lung lesions. Compared with learning from separate tasks, this multi-task strategy leads to a more robust pre-trained model across all lesion tasks and empowers downstream COVID-19 task.

C. Continual Learning

Continual Learning aims to learn from an endless stream of tasks [29]. This paradigm is capable of learning consecutive tasks without forgetting how to perform previously trained tasks. This is challenging that the training process tends to lose knowledge from the previous tasks due to the information relevant to current tasks. To avoid this, we adopt a training schedule to pre-train the model. The value of the initial learning rate is set as 0.01 and decays throughout the training process following the poly learning rate policy $(1 - \text{epoch} / \text{epoch}_{\max})^{0.9}$. In this way, the learning rate is decreasing continuously so that the weight parameters after training non-COVID19 lesion tasks tend to follow its training process while slightly being updated by the current COVID-19 infection task.

D. Fine-tuning

Fine-tuning is the most standard strategy to transfer knowledge from domains where data are abundant. In general, it is conducted by copying the weights from a pre-trained network and tuning them on the downstream task. Recent work [30] shows that fine-tuning can enjoy better performance on small datasets. It is of additional interest to assess the contribution of the encoder and decoder relative to learning COVID-19 knowledge. While the progressive reinitializations demonstrate the incremental effect of each layers, it is unnecessary to prob the extent of localized reinitialization cause our encoder-decoder network is not very deep. Therefore, we adopt two strategies for tuning a non-COVID19 lesion model on the COVID-19 downstream task.

1) *Body Fine-tuning*: Due to the large domain difference between pre-trained tasks and downstream tasks, the most secured fine-tuning method is Body Fine-tuning, which means all parameters of the pre-trained models are used as the initial values to complete the training process of the model. When we train COVID-19 infection networks, all of the parameters

are assigned initially from the pre-trained models on non-COVID19 lesions. In this fine-tuning strategy, the update of parameters largely depends on COVID-19 infection training process itself. Thus, it is a conservative fine-tuning approach that the training task of COVID-19 infection is not affected by the upstream pre-trained non-COVID19 models too much.

2) *Pre-trained Lesion Representations*: Our segmentation model is an encoder-decoder architecture and the encoder serves as a series of convolution operations to encode image features into context representations. These representations are trained on large relative datasets from upstream tasks, and fed as features to downstream ones. In natural language processing tasks, features extracted from internal representations of sequence language models are encoded as pre-trained text representations [29], [30]. In this fine-tuning strategy, we aim to learn general lung lesion features, which we call Pre-trained Lesion Representations. We first train models on non-COVID19 lung lesion datasets. The encoders of these models are capable of encoding lesion features. In other words, the encoders' parameters can exhibit some transferability. Thus, in the following fine-tuning process on COVID-19 dataset, we preserve and freeze them while only fine-tuning and re-training the decoding parts.

E. Multi-encoder Learning

When performing the above fine-tuning strategies, there exist two problems: 1) Body Fine-tuning easily falls into over-fitting cause the downstream COVID-19 infection dataset is much smaller. 2) Utilizing the encoder of pre-trained models to capture feature representation is unstable, because the label spaces and losses for the upstream and downstream tasks differ (Though they are all lung lesions but differ in appearance and shape). Therefore, we present a new method for transfer learning. Our method has three key properties:

- It leverages transferred knowledge from other datasets and combines knowledge from COVID-19 in a better manner.
- It gains performance improvement in terms of training from scratch and other transfer learning methods.
- It does not cost more training time.

To achieve these properties, we propose a Multi-encoder architecture. As shown in Fig.1, this standard U-Net network has two encoders: self-encoder and adapted-encoder. These two encoders have the same architecture and share the same decoder. Self-encoder is a task-specific feature extractor, focusing on segmenting COVID-19 infection through reinitializing all the parameters. These parameters are COVID19-specific which enable update to the network. It is better to assign initial values from pre-trained models to these parameters for convergence, which is adopted in our work. The adapted-encoder is totally frozen while totally using the pre-trained encoder to capture general lung lesion features. Features from these two paralleled encoders are concatenated as fusion of COVID-19 feature representations and then fed into the decoder part.

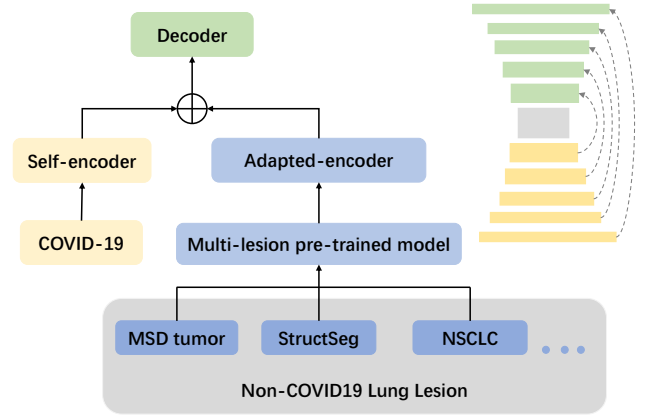


Fig. 1. An overview of Multi-encoder transfer learning strategy. The U-Net network consists of a parallel encoder and a shared decoder. Features from Self-encoder and Adapted-encoder are concatenated.

IV. EXPERIMENTS

In this section, we will describe in detail the datasets, experimental setup and results of our investigation.

A. Dataset Introduction

1) *COVID-19 Dataset*: This dataset is released by Coronacases Initiative and Radiopaedia¹. It is a publicly available COVID-19 volume dataset which contains 20 COVID-19 CT scans. In the work of [7], [31], left lung, right lung, and infections are well-labelled by two radiologists and verified by an experienced radiologist. Thus, with over 1800 annotated slices, this dataset serves as a downstream COVID-19 infection segmentation dataset for transfer learning in our work.

2) *Non-COVID19 Lung Lesion Datasets*: In order to better explore the transferability from various non-COVID19 lung lesions to COVID-19 infections, the following relationships need to be satisfied among these datasets: 1) The size of different lesion datasets are similar. 2) The shape, size and location of different lesion areas are relatively distinguishing. Therefore, in this paper, we introduce three public datasets.

MSD Lung Tumor This dataset was used in a crowd-sourced challenge of generalized semantic segmentation algorithms called the Medical Segmentation Decathlon (MSD) held in MICCAI 2018². This dataset includes patients with non-small cell lung cancer from Stanford University (Palo Alto, CA, USA) publicly available through TCIA and previously utilized to create a radiogenomic signature [32], [33], [34]. The tumor regions are denoted by an expert thoracic radiologist on a representative CT cross section using OsiriX [35]. 63 3D CT scans with corresponding tumor segmentation masks are utilized in our paper.

StructSeg Lung Cancer StructSeg organ dataset is a collection of 3D organ CT scans along with their segmentation ground-truth from 2019 MICCAI challenge³. This dataset contains two types of cancers, nasopharyngeal cancer and lung

¹<https://github.com/ieee8023/covid-chestxray-dataset>

²<http://medicaldecathlon.com/>

³<https://structseg2019.grand-challenge.org>

cancer. We adopt the gross target volume segmentation of lung cancer from 50 patients. Each CT scan is annotated by an experienced oncologist and verified by another one.

NSCLC Pleural Effusion This dataset is developed from a Non-Small Cell Lung Cancer (NSCLC) cohort of 211 subjects⁴. 78 cases with pleural effusion are selected with their segmentation masks.

B. Experimental Settings

All the experiments are implemented in Pytorch and trained on NVIDIA Tesla V100 32GB GPU. For fair comparison, we follow the settings on COVID-19 dataset benchmarks in [7].

For the COVID-19 dataset, we use 5-fold cross validation based on a pre-defined dataset split file. Each fold contains 4 scans (20%) for training and 16 (80%) for testing.

For non-COVID19 lung lesion datasets, we all randomly select 80% of the data for training and the rest of 20% for validation. Pre-trained models based on these non-COVID19 lesions are all trained from scratch with random initial parameters using 3D U-Net network. Due to the limited number of different lesion cases, we do data pre-processing following nnU-Net [27]. The input patch size is set as $50 \times 160 \times 192$ and batch size as 2. Stochastic gradient descent optimizer with an initial learning rate of 0.01 and a nesterov momentum of 0.99 are used for non-COVID19 pre-training. We adopt a summation of Dice loss and Cross entropy loss as loss function.

C. Evaluation Metrics

Diagnostic evaluation is often used in clinical practice for disease diagnosis, patient follow-up or efficacy monitoring. Whether the results of a certain diagnostic evaluation are true, reliable and practical, will largely determine a reasonable medical decision. In this work, we introduce five evaluation metrics for the exploration of transferability.

Dice similarity coefficient (DSC) measures volumetric overlap between segmentation results and annotations. It is computed by:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \quad (1)$$

where A is the sets of foreground voxels in the annotation and B is the corresponding sets of foreground voxels in the segmentation result, respectively.

Normalized surface distance (NSD) [36] serves as a distance-based measurement to assess performance. It is computed by:

$$NSD(A, B) = \frac{|\partial A \cap R_{\partial B}^{(\tau)}| + |\partial B \cap R_{\partial A}^{(\tau)}|}{|\partial A| + |\partial B|}, \quad (2)$$

where $R_{\partial A}^{(\tau)}$ and $R_{\partial B}^{(\tau)}$ represent the borders of ground-truth and segmentation masks which use a threshold τ to tolerate the inter-rater variability of the annotators. We set $\tau = 3mm$ for COVID-19 infection. In contrast to the DSC, which measures

the overlap of volumes, the NSD measures the overlap of two surfaces.

We also consider three other evaluation metrics. Accuracy denotes the correct rate for both positive and negative predictions. Sensitivity shows the percentage of positive instances correctly identified positive. F1-score is the weighted harmonic average of Precision and Sensitivity which is an effective and comprehensive evaluation.

D. Analysis of Different Non-COVID19 Pre-trained Models

All of the four pre-trained models (MSD, StructSeg, NSCLC, Multi-lesion) are utilized to investigate the ability of transfer learning to COVID-19 datasets. With extensive experiments in Table I, we observe that different pre-trained models show different transferability. The best scores of models corresponding to each transfer learning strategies are highlighted along with the detailed comparative analysis of 5 validation folds.

1) *Single-lesion Pre-trained Model*: Among pre-trained models on each single lesion (MSD, StructSeg, NSCLC), compared with training COVID-19 datasets from scratch, MSD tumor pre-trained model improves the segmentation by 3.0% DSC and 3.2% NSD at most. Meanwhile, StructSeg and NSCLC lung lesion pre-trained models show instability under different transfer strategies. As shown in Table I, with Continual Learning and Body Fine-tuning strategies, StructSeg and NSCLC pre-trained models can achieve promising improvement in most folds, but instead, they obtain lower performance with the strategy of Pre-trained Lesion Representations. The rationale is that there exists a large domain distance between these two lesions and COVID-19 infection. The strategy of Pre-trained Lesion Representations totally uses encoded non-COVID19 lesion features. Thus, the transferability of a single-lesion dataset model largely depends on their domain difference.

2) *Multi-lesion Pre-trained Model*: As shown in Table I, we further notice that among all transfer learning strategies, multi-lesion pre-trained model not only achieves a high percentage for DSC and NSD values, but also performs the most stably among all pre-trained models, with the average DSC value of 0.696, 0.696, 0.693, 0.703, respectively. This shows the robustness of a multi-lesion pre-trained model used for transfer learning to COVID-19 infection. Table II further verifies this demonstration using Sensitivity, F1-score and Accuracy concerned with different transfer strategies. It is observed that transferring from multi-lesion pre-trained model outperforms training from scratch on all these evaluation metrics among all strategies. Significantly, when it comes to Pre-trained Lesion Representations, where the encoder of pre-trained models is totally frozen and serves as a non-COVID19 lesion features extractor, the multi-lesion pre-trained model can still perform well. This confirms the effectiveness of multi-task training for multiple lung lesions, which generates more robust and general-purpose representations to help COVID-19 infection tasks. Fig.2(a)-(e) show some examples of segmentation results of the above pre-trained models. It is clear that compared with training from scratch, pre-training from single non-COVID19 models can obtain more accurate massive structures

⁴<https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics>

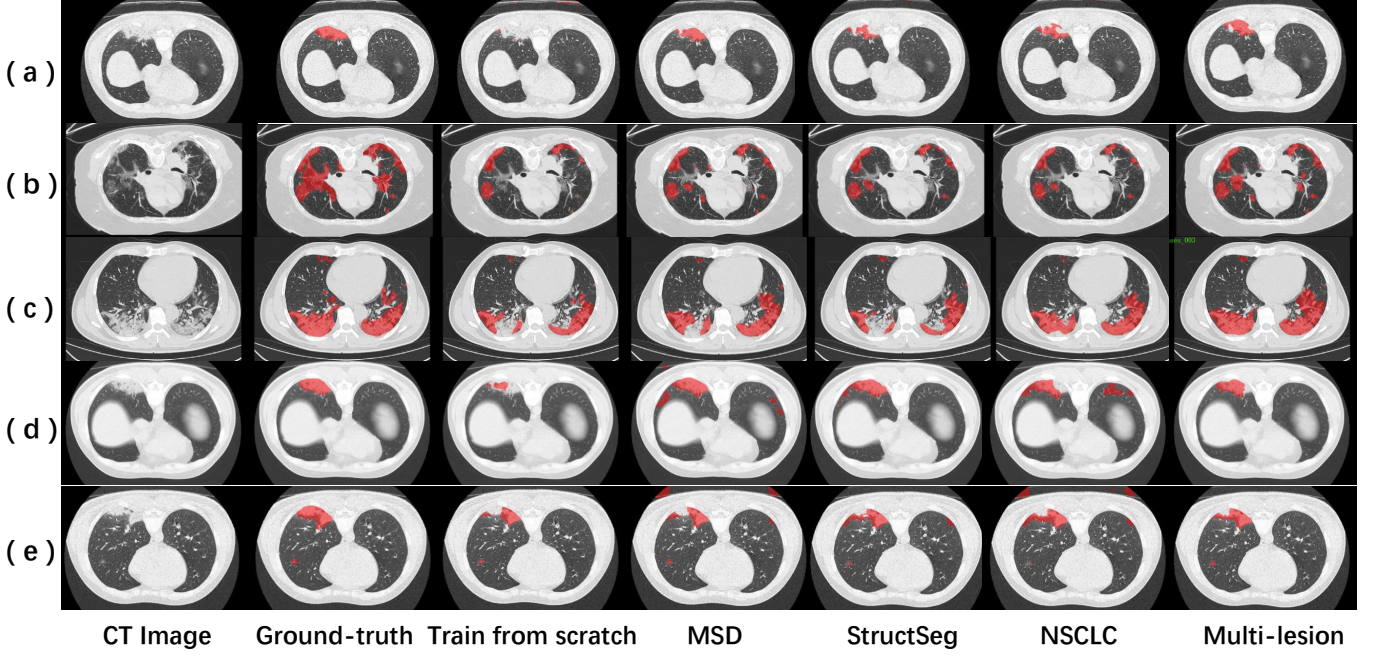


Fig. 2. Visual comparison of COVID-19 infection segmentation results. Training from scratch means using no pre-trained models. MSD, StructSeg, NSCLC and Multi-lesion mean using pre-trained models from corresponding non-COVID19 lesion datasets.

TABLE I
RESULTS OF 5-FOLD CROSS VALIDATION OF DIFFERENT TRANSFER LEARNING STRATEGIES UNDER DIFFERENT NON-COVID19 LUNG LESION PRE-TRAINED MODELS. THE BEST RESULTS ARE SHOWN IN **RED** FONT.

Transfer Strategy	Pre-trained model	fold 0		fold 1		fold 2		fold 3		fold 4		AVG	
		DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD
Train from scratch	None	0.681±0.205	0.709±0.213	0.713±0.205	0.718±0.230	0.662±0.217	0.717±0.242	0.681±0.231	0.708±0.271	0.627±0.269	0.649±0.282	0.673±0.223	0.700±0.224
Continual Learning	MSD	0.696±0.211	0.725±0.219	0.731±0.206	0.734±0.238	0.724±0.166	0.782±0.180	0.720±0.205	0.745±0.224	0.646±0.235	0.672±0.242	0.703±0.203	0.732±0.219
	StructSeg	0.672±0.234	0.704±0.231	0.712±0.206	0.716±0.234	0.704±0.163	0.762±0.183	0.717±0.192	0.745±0.208	0.600±0.264	0.616±0.276	0.681±0.214	0.709±0.228
	NSCLC	0.673±0.215	0.702±0.219	0.712±0.209	0.720±0.244	0.722±0.145	0.775±0.169	0.720±0.209	0.745±0.235	0.581±0.283	0.605±0.292	0.681±0.218	0.709±0.237
	Multi-lesion	0.709±0.209	0.726±0.220	0.740±0.209	0.735±0.232	0.692±0.191	0.745±0.198	0.721±0.208	0.749±0.219	0.618±0.251	0.638±0.260	0.696±0.213	0.718±0.225
Body Fine-tuning	MSD	0.679±0.214	0.706±0.224	0.706±0.210	0.708±0.250	0.724±0.157	0.785±0.167	0.708±0.178	0.724±0.228	0.623±0.247	0.642±0.253	0.688±0.201	0.713±0.225
	StructSeg	0.689±0.203	0.710±0.222	0.719±0.207	0.721±0.238	0.713±0.161	0.770±0.173	0.724±0.209	0.743±0.231	0.603±0.265	0.617±0.271	0.689±0.211	0.712±0.229
	NSCLC	0.696±0.197	0.714±0.213	0.716±0.206	0.720±0.239	0.673±0.208	0.734±0.219	0.690±0.229	0.707±0.266	0.579±0.288	0.590±0.298	0.671±0.228	0.693±0.248
	Multi-lesion	0.702±0.209	0.715±0.221	0.736±0.209	0.730±0.235	0.703±0.182	0.755±0.188	0.725±0.211	0.754±0.225	0.612±0.254	0.628±0.263	0.696±0.213	0.717±0.227
Pre-trained Lesion Representations	MSD	0.709±0.210	0.735±0.225	0.741±0.207	0.751±0.233	0.713±0.159	0.763±0.173	0.693±0.217	0.689±0.262	0.618±0.249	0.637±0.258	0.695±0.209	0.715±0.232
	StructSeg	0.657±0.231	0.670±0.243	0.696±0.215	0.687±0.249	0.657±0.233	0.701±0.246	0.682±0.231	0.701±0.237	0.536±0.271	0.535±0.276	0.645±0.238	0.659±0.252
	NSCLC	0.663±0.214	0.668±0.224	0.695±0.209	0.683±0.241	0.626±0.240	0.664±0.251	0.658±0.229	0.670±0.254	0.544±0.285	0.543±0.294	0.637±0.237	0.646±0.253
	Multi-lesion	0.704±0.210	0.728±0.220	0.731±0.208	0.737±0.235	0.705±0.193	0.748±0.213	0.725±0.209	0.751±0.221	0.602±0.265	0.618±0.271	0.693±0.218	0.716±0.232
Multi-encoder	Multi-lesion	0.712±0.210	0.735±0.223	0.732±0.208	0.740±0.235	0.721±0.148	0.785±0.159	0.742±0.112	0.772±0.145	0.608±0.278	0.629±0.286	0.703±0.201	0.732±0.218

of COVID-19 but shows dissatisfying instability. However, pre-training from multi-lesion model shows high precision and smooth boundary like manually annotated.

E. Analysis of Different Transfer Learning Strategies

Based on the above conclusion that multi-lesion pre-trained model brings consistently higher accuracy and robustness, we further analyze the performance of different transfer strategies adopted in this paper.

Table III shows the results of COVID-19 infection segmentation using the same multi-lesion pre-trained model under different transfer strategies. These results suggest that all the transfer strategies improve the performance of training from

scratch on average DSC, NSD and Sensitivity. In particular, in fold 3, they improve the segmentation by more than 6.0% DSC and 6.4% NSD on maximum. The strategies of Continual Learning and Body Fine-tuning get similar promising results, which both improve the segmentation by 2.3% DSC, 1.8% NSD and 3.4% Sensitivity on average.

An interesting observation is that, compared with Continual Learning and Body Fine-tuning, where all the parameters are updated on COVID-19 dataset, the strategy of Pre-trained Lesion Representations still achieves a competitive performance with an entirely frozen encoder and inherited weights of pre-trained non-COVID19 models. It is observed to improve the DSC from 0.673 to 0.693, NSD from 0.700 to 0.716 and

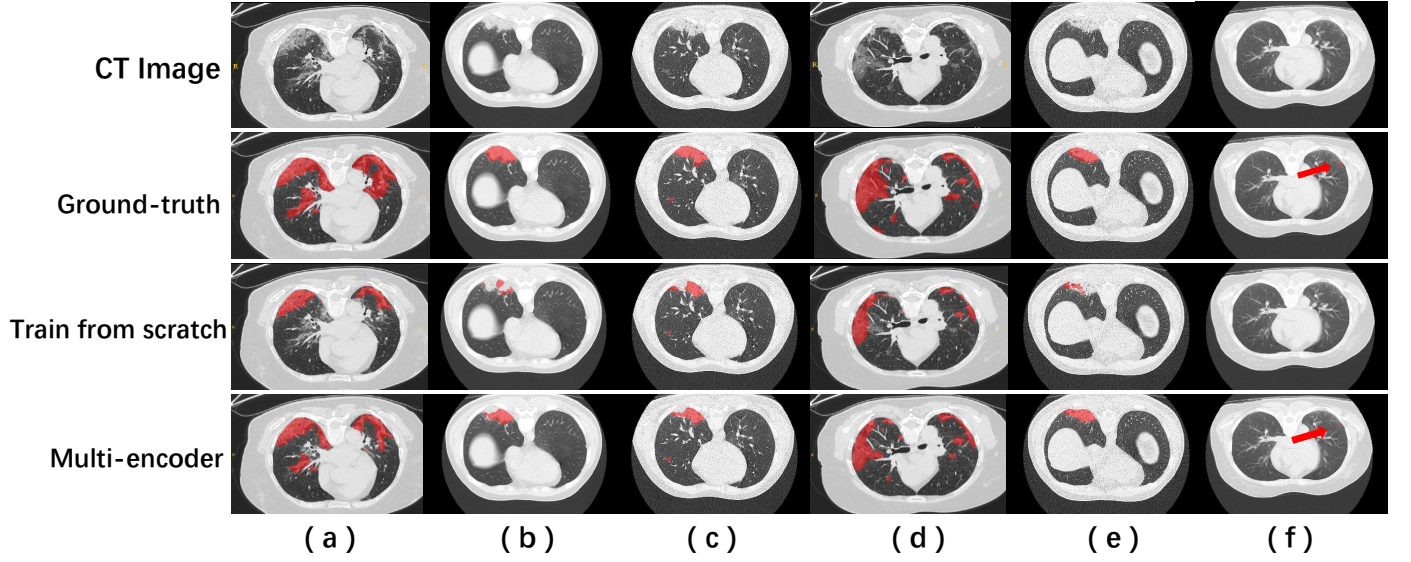


Fig. 3. Visual comparison of COVID-19 infection segmentation results between training from scratch and proposed Multi-encoder transfer learning strategy based on multi-lesion pre-trained model.

TABLE II
RESULTS OF AVERAGE SENSITIVITY, F1-SCORE AND ACCURACY OF DIFFERENT TRANSFER LEARNING STRATEGIES UNDER DIFFERENT NON-COVID19 LUNG LESION PRE-TRAINED MODELS. THE BEST RESULTS ARE SHOWN IN **RED** FONT.

	Sensitivity	F1-score	Accuracy
Train from scratch	0.6204±0.2369	0.6728±0.2227	0.9939±0.0086
Continual Learning			
MSD	0.6592±0.2148	0.7033±0.2029	0.9942±0.0083
StructSeg	0.6317±0.2224	0.6809±0.2139	0.9941±0.0081
NSCLC	0.6371±0.2249	0.6814±0.2180	0.9939±0.0086
Multi-lesion	0.6545±0.2244	0.6962±0.2134	0.9941±0.0083
Body Fine-tuning			
MSD	0.6482±0.2186	0.6877±0.2170	0.9934±0.0100
StructSeg	0.6392±0.2312	0.6894±0.2112	0.9940±0.0083
NSCLC	0.6270±0.2041	0.6709±0.2277	0.9941±0.0086
Multi-lesion	0.6528±0.2252	0.6956±0.1328	0.9941±0.0083
Pre-trained Lesion Representations			
MSD	0.6809±0.2055	0.6949±0.2093	0.9940±0.0083
StructSeg	0.6218±0.2362	0.6455±0.2382	0.9926±0.0118
NSCLC	0.6200±0.2376	0.6371±0.2368	0.9929±0.0096
Multi-lesion	0.6622±0.2117	0.6934±0.2185	0.9940±0.0082
Multi-encoder			
Multi-lesion	0.6635±0.2076	0.7035±0.1995	0.9942±0.0079

Sensitivity from 0.643 to 0.662. In particular, in terms of training cost, Table III shows the training time (per epoch) for each transfer strategy. It can be clearly seen that the strategy of Pre-trained Lesion Representations spends just 148s per epoch on average, much less than training from scratch and other transfer strategies. Due to the frozen encoder, the strategy of Pre-trained Lesion Representations cuts down nearly a half of parameters that need to be updated. Thus, it is promising

to adopt this strategy for COVID-19 transfer learning to save training costs and gain fast convergence.

It is also observed in Table III that our proposed Multi-encoder transfer learning strategy exhibits significantly better segmentation performance than other strategies using multi-lesion pre-trained model. It improves the average DSC from 0.673 to 0.703 and NSD from 0.700 to 0.732, which also performs best among all the pre-trained models in Table I. In terms of F1-score and Accuracy in Table II, proposed Multi-encoder learning outperforms other strategies and enhances the values to 0.7035 and 0.9942, respectively. The transfer ability of Multi-encoder is also confirmed by Fig.3. Compared with training from scratch, Multi-encoder learning yields segmentation results with more accurate boundaries in Fig.3(b)(c)(e) and identifies some minor COVID-19 infection areas in Fig.3(a)(d)(f). The success of proposed Multi-encoder learning strategy is owed to the designed parallel encoders, where the COVID-19 and non-COVID19 lesions are both employed for encoding feature representations, leading to better generalization and lower over-fitting risks.

V. DISCUSSION AND CONCLUSION

In this paper, we investigate the transferability in COVID-19 CT segmentation. We present a set of experiments to better understand how different non-COVID19 lung lesions influence the performance of COVID-19 infection segmentation and their different transfer ability under different transfer learning strategies. Our results reveal clear benefits of pre-training on non-COVID19 datasets.

A valuable demonstration is that a multi-lesion pre-trained model can make the best of multiple lesion representations, and advance the generalization and robustness of pre-trained models. There is value in recognizing that the CT appearance of these different lung lesions share some similarity. Thus, with more kinds of lung lesion datasets incorporated to pre-train a

TABLE III

RESULTS OF 5-FOLD CROSS VALIDATION OF DIFFERENT TRANSFER LEARNING STRATEGIES BASED ON MULTI-LESION PRE-TRAINED MODEL. **TIME** REPRESENTS THE TRAINING TIME FOR AN EPOCH. THE BEST RESULTS ARE SHOWN IN **RED** FONT.

		Train from scratch				Continual Learning				Body Fine-tuning				Lesion Representations				Multi-encoder			
		DSC↑	NSD↑	Sen.↑	Time↓	DSC↑	NSD↑	Sen.↑	Time↓	DSC↑	NSD↑	Sen.↑	Time↓	DSC↑	NSD↑	Sen.↑	Time↓	DSC↑	NSD↑	Sen.↑	Time↓
fold 0	avg	0.681	0.709	0.639	224s	0.709	0.726	0.683	194s	0.702	0.715	0.668	196s	0.704	0.728	0.682	145s	0.712	0.735	0.693	206s
	std	0.205	0.213	0.162	–	0.209	0.220	0.161	–	0.209	0.221	0.160	–	0.210	0.220	0.156	–	0.210	0.223	0.151	–
fold 1	avg	0.713	0.718	0.707	194s	0.740	0.735	0.752	193s	0.736	0.730	0.754	196s	0.731	0.737	0.762	145s	0.732	0.740	0.769	201s
	std	0.205	0.230	0.189	–	0.209	0.232	0.204	–	0.209	0.235	0.172	–	0.208	0.235	0.126	–	0.208	0.235	0.129	–
fold 2	avg	0.662	0.717	0.621	224s	0.692	0.745	0.667	193s	0.703	0.755	0.688	194s	0.705	0.748	0.676	146s	0.721	0.785	0.685	208s
	std	0.217	0.242	0.295	–	0.191	0.198	0.272	–	0.182	0.188	0.262	–	0.193	0.213	0.260	–	0.148	0.159	0.232	–
fold 3	avg	0.681	0.708	0.604	194s	0.721	0.749	0.657	193s	0.725	0.754	0.648	196s	0.725	0.751	0.696	148s	0.742	0.772	0.659	208s
	std	0.231	0.271	0.241	–	0.208	0.219	0.164	–	0.211	0.225	0.203	–	0.209	0.221	0.143	–	0.112	0.145	0.152	–
fold 4	avg	0.627	0.649	0.531	194s	0.618	0.638	0.514	223s	0.612	0.628	0.506	194s	0.602	0.618	0.495	157s	0.608	0.629	0.512	208s
	std	0.269	0.282	0.267	–	0.251	0.260	0.254	–	0.254	0.263	0.258	196s	0.265	0.271	0.255	–	0.278	0.286	0.270	–
AVG	avg	0.673	0.700	0.620	206s	0.696	0.718	0.655	199s	0.696	0.717	0.653	195s	0.693	0.716	0.662	148s	0.703	0.732	0.663	206s
	std	0.223	0.224	0.237	16s	0.213	0.225	0.224	13s	0.213	0.227	0.225	1s	0.218	0.232	0.212	5s	0.201	0.218	0.208	3s

model, we could achieve better performance. This exploration is an important contribution, enabling more research on transfer learning to COVID-19 infection from the perspective of utilizing non-COVID19 lung lesions.

Moreover, this paper examines a series of different transfer learning strategies, including Continual Learning, Body Fine-tuning, Pre-trained Lesion Representations and the proposed Multi-encoder Learning. We observe segmentation improvement in all performance metrics. It is also noted that the strategy of Pre-trained Lesion Representations with a frozen encoder enhances performance as well. This gains more insight into the significant transferability exhibited by the encoding parameters. Though the encoding layers do not contain any explicit knowledge of the COVID-19 infection, their parameters still enable the optimizer to reach a higher performance while fine-tuning. The rationale is that the encoded multi-lesion representations contain more high-level and abundant encoding information of the medically relevant lung lesion observed in CT images. Through combining the multi-lesion representations and COVID-19 infection features, the proposed Multi-encoder achieves significant improvement. These observation and exploration are important not only in COVID-19 transfer learning but also in the general medical domain, because feature reuse from pre-training out-of-domain datasets shows significant improvement for task performance and training convergence.

Future research directions include utilizing more various non-COVID19 lung lesion datasets and investigating better transfer learning methods, so that non-COVID19 lung lesions can be effectively used to improve the quality of COVID-19 infection segmentation in the absence of sufficient high-quality COVID-19 datasets.

REFERENCES

- [1] J. F.-W. Chan, C. C.-Y. Yip, K. K.-W. To, T. H.-C. Tang, S. C.-Y. Wong, K.-H. Leung, A. Y.-F. Fung, A. C.-K. Ng, Z. Zou, H.-W. Tsoi, G. K.-Y. Choi, A. R. Tam, V. C.-C. Cheng, K.-H. Chan, O. T.-Y. Tsang, and K.-Y. Yuen, "Improved molecular diagnosis of covid-19 by the novel, highly sensitive and specific covid-19-rdrp/hel real-time reverse transcription-pcr assay validated in vitro and with clinical specimens," *Journal of Clinical Microbiology*, vol. 58, no. 5, 2020. [Online]. Available: <https://jcm.asm.org/content/58/5/e00310-20>
- [2] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad, A. Jacobi, K. Li, S. Li, and H. Shan, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020, pMID: 32017661. [Online]. Available: <https://doi.org/10.1148/radiol.202002030>
- [3] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, W. Sun, L. Lan, W. Zhang, X. Mu, C. Shi, Z. Wang, J. Lee, Z. Jin, M. Lin, H. Jin, L. Zhang, J. Guo, B. Zhao, Z. Ren, S. Wang, Z. You, J. Dong, X. Wang, J. Wang, and W. Xu, "Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks," *medRxiv*, 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/03/23/2020.03.19.20039354>
- [4] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, pp. 1–1, 2020.
- [5] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," 2019.
- [6] V. Cheplygina, M. [de Bruijne], and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280 – 296, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1361841518307588>
- [7] M. Jun, W. Yixin, A. Xingle, G. Cheng, Y. Ziqi, C. Jianan, Z. Qiongjie, D. Guoqiang, H. Jian, H. Zhiqiang, N. Ziwei, and Y. Xiaoping, "Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation," *arXiv preprint arXiv:2004.12537*, 2020.
- [8] X. Qi, Z. Jiang, Q. YU, C. Shao, H. Zhang, H. Yue, B. Ma, Y. Wang, C. Liu, X. Meng, S. Huang, J. Wang, D. Xu, J. Lei, G. Xie, H. Huang, J. Yang, J. Ji, H. Pan, S. Zou, and S. Ju, "Machine learning-based ct radiomics model for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: A multicenter study," *medRxiv*, 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/03/03/2020.02.29.20029603>
- [9] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, J. Liu, and D. Shen, "Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest ct images," 2020.
- [10] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," 2020.
- [11] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: Deep-learning approach," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200075, 2020. [Online]. Available: <https://doi.org/10.1148/rct.202000075>
- [12] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," 2020.

- [13] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/03/26/2020.03.12.20027185>
- [14] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," 2020.
- [15] Q. Yan, B. Wang, D. Gong, C. Luo, W. Zhao, J. Shen, Q. Shi, S. Jin, L. Zhang, and Z. You, "COVID-19 Chest CT Image Segmentation – A Deep Convolutional Neural Network Solution," *arXiv e-prints*, p. arXiv:2004.10987, Apr. 2020.
- [16] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [17] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," 2016.
- [18] V. Cheplygina, I. P. Pena, J. H. Pedersen, D. A. Lynch, L. Sorensen, and M. de Bruijne, "Transfer learning for multicenter classification of chronic obstructive pulmonary disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, p. 14861496, Sep 2018. [Online]. Available: <http://dx.doi.org/10.1109/JBHI.2017.2769800>
- [19] G. Carneiro, J. C. Nascimento, and A. P. Bradley, "Unregistered multi-view mammogram analysis with pre-trained deep learning models," in *MICCAI*, 2015.
- [20] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology identification using deep feature selection with non-medical training," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 259–263, 2018. [Online]. Available: <https://doi.org/10.1080/21681163.2016.1138324>
- [21] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, and D. J. Mollura, "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018, pMID: 29623248. [Online]. Available: <https://doi.org/10.1080/21681163.2015.1124249>
- [22] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken, and et al., "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," *Lecture Notes in Computer Science*, p. 516524, 2017.
- [23] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaevius, and V. H. C. de Albuquerque, "A novel transfer learning based approach for pneumonia detection in chest x-ray images," *Applied Sciences*, vol. 10, no. 2, p. 559, Jan 2020. [Online]. Available: <http://dx.doi.org/10.3390/app10020559>
- [24] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, Apr 2020. [Online]. Available: <http://dx.doi.org/10.1007/s13246-020-00865-4>
- [25] T. Majeed, R. Rashid, D. Ali, and A. Asaad, "Covid-19 detection using cnn transfer learning from x-ray images," *medRxiv*, 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/05/19/2020.05.12.20098954>
- [26] S. Misra, S. Jeon, S. Lee, R. Managuli, and C. Kim, "Multi-Channel Transfer Learning of Chest X-ray Images for Screening of COVID-19," *arXiv e-prints*, p. arXiv:2005.05576, May 2020.
- [27] F. Isensee, J. Petersen, S. A. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnu-net: Breaking the spell on successful medical image segmentation," *CoRR*, vol. abs/1904.08128, 2019. [Online]. Available: <http://arxiv.org/abs/1904.08128>
- [28] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3d deeply supervised network for automatic liver segmentation from ct volumes," 2016.
- [29] N. Houlsby, A. Giurghi, S. Jastrzebski, B. Morrone, Q. de Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," 2019.
- [30] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?" 2019.
- [31] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Mingqing, L. Xin, D. Xueyuan, C. Shucheng, W. Hao, M. Sen, Y. Xiaoyu, N. Ziwei, L. Chen, T. Lu, Z. Yuntao, Z. Qiongjie, D. Guoqiang, and H. Jian, "COVID-19 CT Lung and Infection Segmentation Dataset," Apr. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3757476>
- [32] S. Napel and S. K. Plevritis, "Nslc radiogenomics: Initial stanford study of 26 cases," 2014. [Online]. Available: <https://wiki.cancerimagingarchive.net/x/Gglp>
- [33] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. Benson, W. Zhang, A. Leung, M. Kadoch, C. D Hoang, J. Shrager, A. Quon, D. Rubin, S. Plevritis, and S. Napel, "A radio-genomic dataset of non-small cell lung cancer," *Scientific data*, vol. 5, Oct. 2018.
- [34] O. Gevaert, J. Xu, C. D. Hoang, A. N. Leung, Y. Xu, A. Quon, D. L. Rubin, S. Napel, and S. K. Plevritis, "Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results," *Radiology*, vol. 264, no. 2, p. 387396, August 2012. [Online]. Available: <https://europepmc.org/articles/PMC3401348>
- [35] A. Rosset, L. Spadola, and O. Ratib, "Osirix: an open-source software for navigating in multidimensional dicom images," *Journal of digital imaging*, vol. 17, no. 3, p. 205216, September 2004. [Online]. Available: <https://europepmc.org/articles/PMC3046608>
- [36] S. Nikolov, S. Blackwell, R. Mendes, J. D. Fauw, C. Meyer, C. Hughes, H. Askham, B. Romera-Paredes, A. Karthikesalingam, C. Chu, D. Carnell, C. Boon, D. D'Souza, S. A. Moinuddin, K. Sullivan, D. R. Consortium, H. Montgomery, G. Rees, R. Sharma, M. Suleyman, T. Back, J. R. Ledsam, and O. Ronneberger, "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy," 2018.