Transfer Learning via Parameter Regularization for Medical Image Segmentation

Nimrod Sagie
Department of Medical Engineering
Tel-Aviv University, Israel
nimrodsagie@mail.tau.ac.il

Hayit Greenspan
Department of Medical Engineering
Tel-Aviv University, Israel
hayitg@tauex.tau.ac.il

Jacob Goldberger
Faculty of Electrical Engineering
Bar-Ilan University, Israel
jacob.goldberger@biu.ac.il

***Abstract*—Transfer learning is a popular strategy to overcome the difficulties posed by limited training data. It uses the parameters of the source task to initialize the parameters of the target task. In this study, we cast transfer learning as a regularization procedure. In addition to initialization, we incorporate the source task parameters into the cost function used to train the target task. We regularize the learned parameters by penalizing them if they deviate too much from their initial values. We demonstrate the power of the proposed transfer learning scheme on the task of COVID-19 opacity https://www.overleaf.com/projectsegmentation. Specifically, we show that it can improve the segmentation of coronavirus lesions in chest CT scans.**

***Index Terms*—transfer learning, segmentation, regularization, Covid-19**

I. Introduction

One of the main differences between the medical imaging domain and computer vision is the need to cope with small datasets and a limited amount of annotated samples [1] [2] [3] [4]. Collecting medical data is usually an expensive procedure that requires the collaboration of radiologists and researchers. This problem is even more acute in the case of supervised machine learning algorithms, in which the training requires labeled data and larger sets of training examples. Training an automatic medical imaging system is based on annotations, which can only be made by radiologists with vast expertise about the data and the task. Such annotations are time-consuming, even when done by experts. This is true for image classification tasks and is even more problematic in organ or lesion segmentation tasks. For some medical tasks, medical datasets are available online, and grand challenges have been publicized for others. Today, however, most datasets are still collected from a single hospital or several hospitals who have datasharing protocols. These datasets are limited in size and only applicable to specific medical problems.

Transfer learning is a popular strategy to overcome the difficulties posed by limited training data. The goal of transfer learning is to transfer knowledge from a source task to a target task by using the parameter set of the source task in the process of learning the target task. Transfer learning utilizes models that are pre-trained on large datasets, that can either be scenery datasets such as ImageNet or medical datasets from a similar domain. The pre-trained network is further trained on the specific target task of interest, which has fewer labeled examples than the source task. There is a plethora of work on using transfer learning in different medical imaging applications (e.g. [5] [6]). Despite the popularity of transfer learning in medical imaging, there has been little work analyzing its precise effects [7].

In this study, we examined whether using the source parameters to initialize the target parameters is the best way to transfer knowledge. Injecting information into a network via parameter initialization is problematic since this information can be lost during the optimization procedure. Catastrophic forgetting [8] is the tendency of a neural network to completely and abruptly forget previously learned information upon learning new information. In transfer learning, the target model is not expected to handle the source task and it tends to forget it. However, we still expect that the feature processing, which is done by the source network, not to be completely forgotten. Another variant of transfer learning is freezing the first layers of the source network and only training the last layers that perform the actual classification required by the target task. This strategy does not suffer from knowledge forgetting. However, if the target domain data are substantially different from the original data, the low-level processing performed by the first layers will not be suitable for the new data.

In this study, we propose a transfer learning method that addresses the information forgetting problems described above. Instead of using the parameters of the source network as an initialization, we use them as a regularization term. We penalize the learned parameters of the target task if they are too far from the source network. This way, we overcome the catastrophic forgetting problem and benefit from the knowledge acquired by the source task.

Closely related to our work, Li et al. [9], investigated several regularization schemes that explicitly promote the similarity of the fine-tuned model with the original pre-trained model. Different from that method, our proposed approach concentrates on the relevant subset of layers to fine-tune in a pretrained model on a new task. We concentrate here on U-net networks that are applied to the task of medical image segmentation. We apply regularized transfer learning only on the encoder while training the decoder from scratch.

We applied this transfer learning strategy to the task of COVID-19 opacity segmentation and show that it improves the segmentation of coronavirus lesions in chest CT scans.

II. Transfer learning regularization

The U-net architecture was introduced by Olaf Ronneberger et al. [10] and has become the state-of-the-art for medical image semantic segmentation. It is composed of two main pathways: a contraction path (the encoder) that captures context by processing low-level information, and the expanding path (the decoder), which enables precise localization. The U-net encoder performs low and mid-level processing of the pixel map leading to a latent image representation. In contrast, the U-net decoder, generates the network's decisions based on the computed representation and it is focused on a specific task accomplished by the network. The most common way of utilizing transfer learning with U-net is by initializing the encoder with pre-trained weights and then either freezing it, or allowing re-training, depending on the target data size and computational power limitations. The decoder, which is task-dependent, is trained from scratch.

Regularization is a method for preventing overfitting to the training data. Let θ be the parameter-set of a given neural network. L_2 regularization is a popular scheme, that modifies the (cross-entropy) loss function $\text{Loss}(\theta)$ which we minimize by adding a regularization term that penalizes large weights:

$$\text{Cost}(\theta) = \text{Loss}(\theta) + \lambda\|\theta\|^2, \tag{1}$$

where λ is the regularization coefficient. Adding the L_2 term usually results in much smaller weights across the entire model, and for this reason is known as weight decay. Adding L_2 can be viewed as imposing a zero-mean Gaussian prior to the parameter set. This transforms the optimization problem from one that involves performing maximum likelihood estimation (MLE) to one that involves maximum a posteriori (MAP) estimation; i.e. a shift from using frequentist inference to Bayesian inference.

We propose to exploit the full potential of the knowledge already acquired by the model on the source task, by enabling changes in weights, but under a certain constraint. We formalize this intuition by considering the parameters of the source network as the means of a Gaussian prior on the parameters of the target task. The proposed cost function is:

$$\text{Cost}(\theta) = \text{Loss}(\theta) + \lambda\|\theta_{\text{encoder}} - \bar{\theta}_{\text{encoder}}\|^2 \tag{2}$$

s.t. $\theta = (\theta_{encoder}, \theta_{decoder})$ and $\bar{\theta} = (\bar{\theta}_{encoder}, \bar{\theta}_{decoder})$ are the parameters of the source and target networks, respectively. In other words, the parameters of the source network form a Gaussian prior on the target parameters:

$$\theta_{encoder} \sim \mathcal{N}(\bar{\theta}_{encoder}, \frac{1}{\lambda}I). \tag{3}$$

Setting the hyper-parameter λ to ∞ results in freezing the regularized parameters. By setting the hyper-parameter λ to zero, we obtain standard transfer learning where the only way knowledge is transferred to the target task is via parameter initialization. Choosing a value for λ in the range of $(0, \infty)$ controls the amount of knowledge we want to transfer from the source to the target task. In practice, λ is a hyper-parameter that can be tuned using cross-validation.

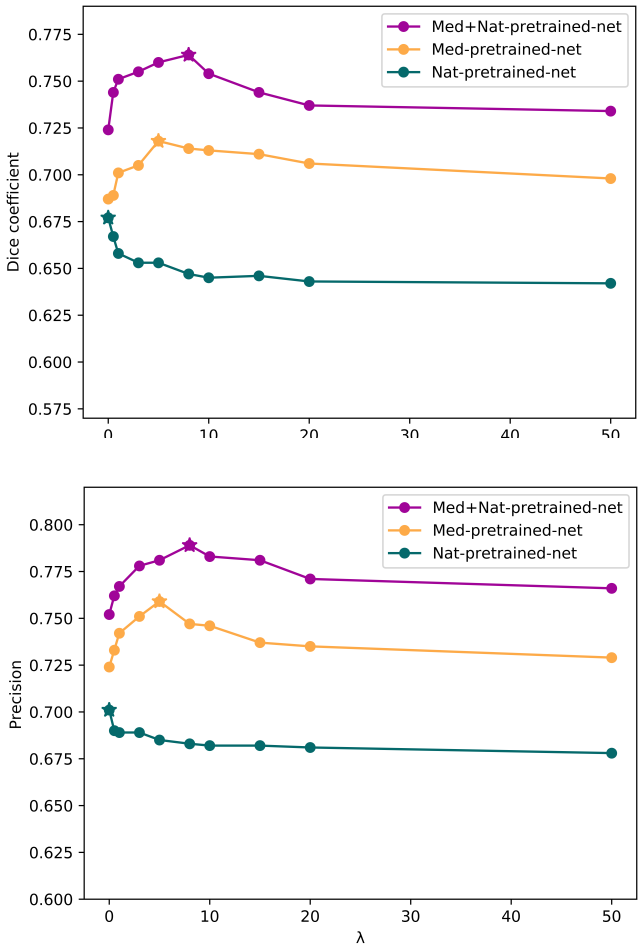

Fig. 1. Segmentation results for coronavirus lesions in chest CT scans as a function of the regularization coefficient λ, for three different pre-training scenarios.

One of the major components of the transfer learning scheme is the identity of the source task and its degree of similarity with the target task. This similarity can be measured in terms of the structure of the input image, whether it is an image of scenery, or a medical image (e.g. MRI, Ultrasound, X-ray). This similarity can also be measured by the task type of the source; e.g. the same organ, similar pathology. It is our hypothesis that the greater the similarity between the source and target tasks, the larger the optimal value of λ will be. We therefore explore network regularization in several scenarios: from natural image pre-training using ImageNet, to a network pre-trained with medical context. We next describe the network architecture and pretraining used.

A. Implementation details

We focused here on the task of COVID-19 opacity segmentation. We used a 2-D U-net [10] with a DenseNet121 [11] backbone. In our implementation, the decoder was composed of decoder blocks and a final segmentation head, which consisted of a convolutional layer and softmax activation. Each decoder block consisted of a transpose convolution

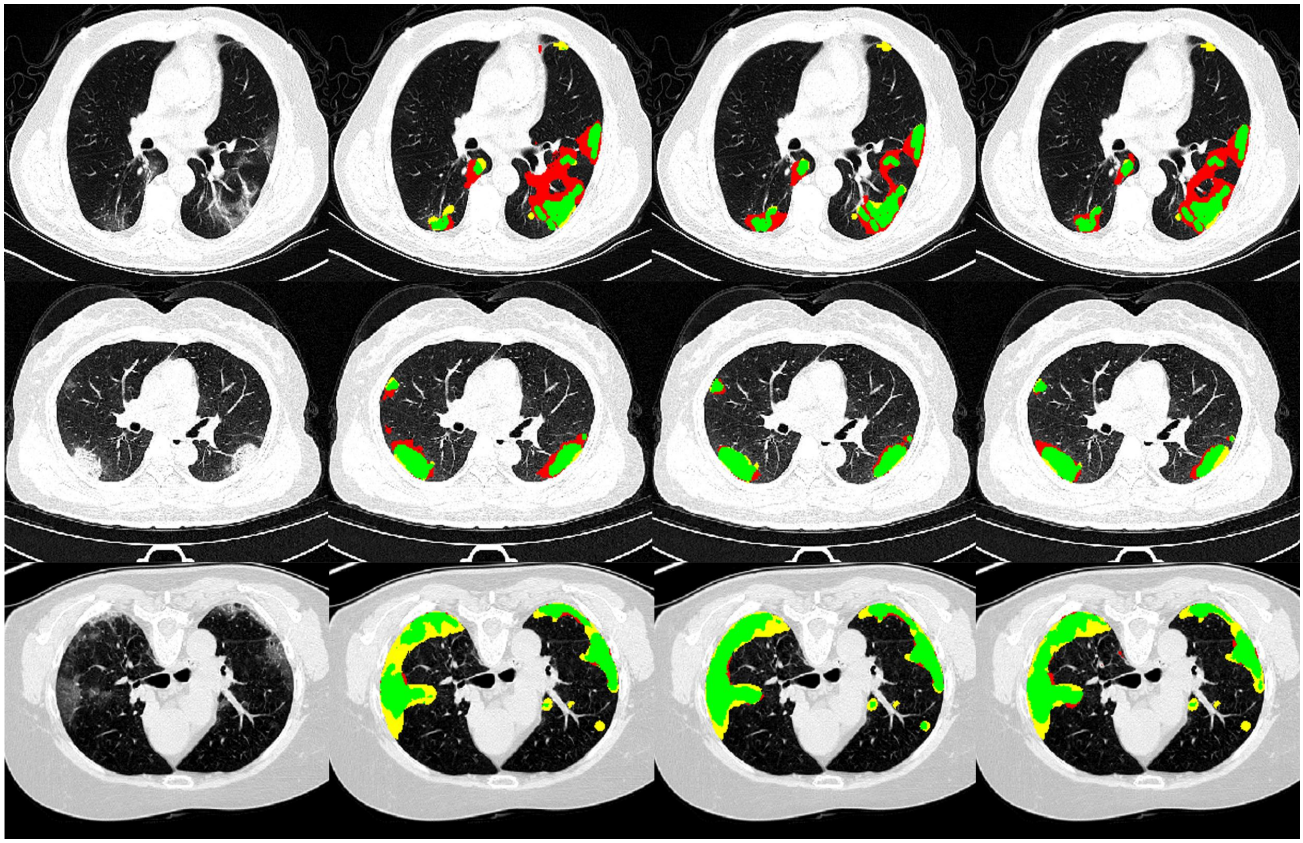

Image slice $\lambda = 0$ $\lambda = 8$ $\lambda = 50$

Fig. 2. Qualitative comparison of COVID-19 opacity segmentation with different transfer learning regularization. Three examples are shown. Green, red, and yellow represent TP, FP, and FN prediction respectively.

layer, followed by two blocks of convolutional layers, batch normalization, and ReLU activation. For the cost function, we used weighted cross-entropy, where the weights were calculated using the class ratio in the dataset.

In our experimentation we investigate regularization in varying pre-training scenarios. We implemented three source tasks and used them to pre-train the encoder on the target task (the decoder was trained from scratch). The three source tasks were as follows:

- **Natural-image pre-trained network**: U-net with encoder that was trained on ImageNet. Hereon we term this network "Nat-pretrained-net".
- **Medical-image pre-trained network**: U-net with encoder that was trained from scratch on several publicly available medical imaging segmentation tasks [12]. This is based on our earlier work [13], in which we introduced the "HydraNet", a U-net-based, fully convolutional neural network, with a shared encoder for global feature extraction followed by several task-specific decoders. Hereon, we term this network "Med-pretrained-net".
- **Combined natural and medical image pre-training network**: Encoder was initialized with ImageNet weights and then trained on the medical datasets as above. We term this network "Med+Nat-pretrained-net".

The overall system consisted of the trained model and a series of image processing techniques for both the pre, and the post-processing stages. For pre-processing, all the input slices were clipped and normalized to [0,1] using a window of [-1000, 0] HU and then resized to a fixed spatial input size of 384×384. The trained network was applied on each slice separately. To construct the 3-D segmentation, we first concatenated the slice-level probabilities generated by the model, and then applied a post-processing pipeline that includes morphological operations and false-positive reduction (removal of opacities outside the lungs).

III. Experiments and Results

We evaluated the system on the task of COVID-19 opacity segmentation using a small COVID-19 dataset [14] containing 29 non-contrast CT scans from three different distributions, from which 3,801 slices were extracted. Lungs and areas of infection were labeled by two radiologists and verified by an experienced radiologist. The given labels were of the lungs and infection. The train-validation-test split was: 21 cases for training, 3 cases for validation, and 5 cases for testing, chosen at random.

We performed experiments with several values of λ, starting with $\lambda = 0$; i.e., standard transfer learning via parameter

initialization, up to $\lambda = 50$; i.e., a high penalty for deviation from the learned weights, which can be considered as basically freezing the encoder. Given a 3-D chest CT scan, the system produced the correlated 3-D prediction mask for the lungs, as well as the COVID-19 related infections. Once the 3-D segmentation mask for the test set had been extracted, we compared it to the ground truth (GT) reference mask for the opacity class, in terms of Dice and Precision metrics. Consistent pre- and post-processing stages were applied to all the network solutions that were compared.

Figure 1 shows the average segmentation results for the opacity class as a function of the regularization coefficient λ. For the ImageNet experiments, both Dice and precision were optimal at $\lambda = 0$. For the Med- and Med+Nat-pretrained-net experiments, the segmentation performance improved up to a certain optimal regularization weight. The best segmentation results were attained with Med+Nat-pretrained-net, and $\lambda = 8$ was obtained as the optimal value. At $\lambda = 8$ the dice score was improved by 5.5% from 0.724 ($\lambda = 0$) to 0.764, with a p-value of 0.030. The precision was improved by 4.9% from 0.752 ($\lambda = 0$) to 0.789, with a p-value of 0.019.

These results demonstrate that using extra regularization penalization for transfer learning overpowers initialization on its own, since the source task and the target task's distributions are more similar. Thus, by using the regularization term, the segmentation results can be improved in cases where the transfer learning is from a source domain close to the target domain. In cases where the transfer learning comes from a source domain with a very different distribution than the target domain, as in the case from natural images to non-contrast chest CT images, it is better to allow deviation from the learned weights.

Qualitative results are shown in Figure 2. For each input slice, we see the CT slice and the segmentation results for several values of λ obtained by using the Med+Nat as source task. The given examples show the system prediction for slices from three different test cases with different disease manifestations, demonstrate its generalization capabilities. It can be also seen that at $\lambda = 8$ the red and the yellow regions are minimal compared to at $\lambda = 0$ and at $\lambda = 50$, indicates the improved results of the optimal regularization term.

There are several published results on the same COVID-19 dataset [14]. Wang et al. [15] suggested a Hybrid-encoder transfer learning approach. Laradji et al. [16] used a weakly supervised consistency-based strategy with point-level annotations. Muller et al. [17] implemented a 3-D U-Net and using a patch-based scheme. Paluru et al. [18] recently suggested an anamorphic depth embedding-based lightweight model. The reported Dice scores were 0.704 [15], 0.750 [16], 0.761 [17], 0.798 [18] and 0.698 [19]. Comparison here, however, is problematic due to different data-splits and different source tasks used for transfer learning. We note, however, that our transfer regularization approach is complementary to previous works and can be easily integrated into their training procedure.

IV. Conclusions

This study described a transfer learning scheme based on using the parameters of the source task as a regularization term while learning the target task. We concentrated on image segmentation problems that are handled by the U-net architecture where the encoder and the decoder need to be treated differently. We addressed the specific task of segmenting COVID-19 lesions in chest CT images and showed that adding an extra regularization term to the cost function leads to improved segmentation results in cases of sufficient similarity between the source and target tasks. The presented method is general and can be incorporated in any instance where transfer learning from the source task to the target task is implemented.

Acknowledgment

The research was partially supported by the Israeli Ministry of Science & Technology.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A.W.M van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, 2017.

[2] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[3] 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, May 2016.

[4] V. Cheplygina, M.n de Bruijne, and J. P.W. 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.

[5] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "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*, 2017.

[6] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[7] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Advances in neural information processing systems*, 2019.

[8] M. McCloskey and N. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," *The Psychology of Learning and Motivation*, vol. 24, pp. 109–164, 1989.

[9] Xuhong Li, Yves Grandvalet, and Franck Davoine, "Explicit inductive bias for transfer learning with convolutional networks," in *International Conference on Machine Learning*, 2018, pp. 2825–2834.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[12] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[13] N. Sagie, S. Almog, A. Talby, and H. Greenspan, "Covid-19 opacity segmentation in chest ct via hydranet: a joint learning multi-decoder network," in *Medical Imaging 2021: Computer-Aided Diagnosis*. SPIE, 2021, vol. 11597.

[14] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, et al., "COVID-19 CT lung and infection segmentation dataset," *Zenodo, Apr*, vol. 20, 2020.

[15] Yixin Wang, Yao Zhang, Yang Liu, Jiang Tian, Cheng Zhong, Zhongchao Shi, Yang Zhang, and Zhiqiang He, "Does non-COVID19 lung lesion help? investigating transferability in COVID-19 CT image segmentation," *Computer Methods and Programs in Biomedicine*, p. 106004, 2021.

[16] Issam Laradji, Pau Rodriguez, Oscar Manas, Keegan Lensink, Marco Law, Lironne Kurzman, William Parker, David Vazquez, and Derek Nowrouzezahrai, "A weakly supervised consistency-based learning method for covid-19 segmentation in CT images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2453–2462.

[17] Dominik Müller, Iñaki Soto Rey, and Frank Kramer, "Automated chest ct image segmentation of Covid-19 lung infection based on 3D U-net," *arXiv preprint arXiv:2007.04774*, 2020.

[18] Naveen Paluru, Aveen Dayal, Håvard Bjørke Jenssen, Tomas Sakinis, Linga Reddy Cenkeramaddi, Jaya Prakash, and Phaneendra K Yalavarthy, "Anam-net: Anamorphic depth embedding-based lightweight cnn for segmentation of anomalies in Covid-19 chest ct images," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[19] Keno K. Bressem, Stefan Markus Niehues, Bernd Hamm, Marcus R. Makowski, Janis Lucas Vahldiek, and Lisa C. Adams, "3d u-net for segmentation of COVID-19 associated pulmonary infiltrates using transfer learning: State-of-the-art results on affordable hardware," *CoRR*, vol. abs/2101.09976, 2021.