

# What is Transfer Learning?

Anton Zhitomirsky

January 3, 2024

A review of the following [3, 2, 5, 16, 10, 6, 4, 8, 7, 1, 12, 9, 14, 11, 13]

## 1 What is Transfer Learning?

We can use models that have been trained for one task, and use it as a starting point for a model on a second task. This can be useful when the second task is similar to the first task, or when there is a limited amount of data available [15]. The intuitive reason why transfer learning works is because the early layers of the network a deep learning model attempts to learn very low-level features. All these features occur regardless of the exact cost function or image dataset [15], which makes it ideal for transfer learning because the set of low-level features learnt will be similar.

### 1.1 Advantages

- By using the learned features from the first task as a starting point, the model can learn more quickly and effectively on the second task [15] (it already has a good understanding of the features and patterns in the data).
- can also help to prevent overfitting, as the model will have already learned general features that are likely to be useful in the second task [15].

### 1.2 Disadvantages

- Domain mismatch if the two models are trained on domains that are vastly different or with a different data distribution [15].
- Overfitting ‘Transfer learning can lead to overfitting if the model is fine-tuned too much on the second task, as it may learn task-specific features that do not generalize well to new data.’ [15].

### 1.3 General Approach

1. Obtain a pre-trained ‘base’ model. This model has been trained on extensive data and has identified general features and patterns relevant to numerous related jobs.
2. Identify the transfer-layers. These capture generic information relevant to the new task as well as the previous one. These are ‘frozen’ in the final deliverable (see Figure 1) as we want to preserve the low-level learnt features. A layer is frozen or ‘fixed’ when it is no longer available for training, hence the weights of these layers will not be available for updates.

3. Fine-tune and retrain the remaining layers. The goal is to preserve the knowledge from the pre-training while enabling the model to modify its parameters to better suit the demands of the current assignment. We can either
  - freeze a few layers of the pre-trained model and train other layers on our new dataset for the new task
  - or make a new model, but also take out some features from the layers in the pre-trained model and use them in a newly created model

The number of frozen layers depend on how much you want to inherit from the pre-trained model. If the networks are significantly different (i.e. use a human face detector to detect cars) selecting many layers for freezing will not only give low level features but also give high-level features like nose, eyes, etc which are useless for new dataset (car detection). Therefore, in this case you only copy the low level features [15].

```
# freezing layers in python
for layer in base_model.layers:
    layer.trainable = False
```

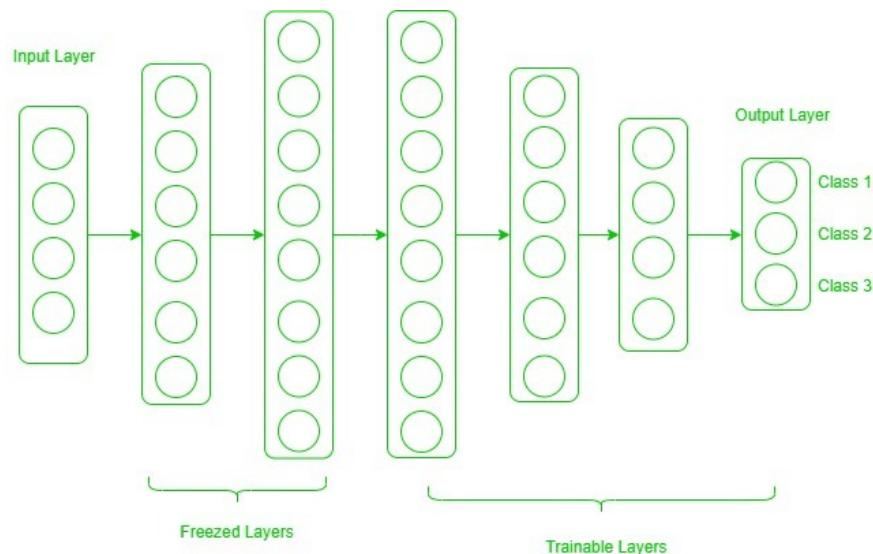


Figure 1: Figure taken from [15] illustrating frozen and trainable layers from pre-trained and fine-tuned network.

## 1.4 Recommendations

- **The target dataset is small and similar to the base network dataset:** There may be an issue of overfitting because fine-tune the pre-trained network with the target dataset may not generalise to the global population. Also, there may be some changes in the number of classes in the target task; remove the fully connected layers from the end and add a new fully connected layer satisfying the number of new classes. Now, we freeze the rest of the model and only train newly added layers [15].

- **The target dataset is small and different from the base network dataset:** Since the target dataset is different, using high-level features of the pre-trained model will not be useful. In such a case, remove most of the layers from the end in a pre-trained model, and add new layers a satisfying number of classes in a new dataset. This way we can use low-level features from the pre-trained model and train the rest of the layers to fit a new dataset. Sometimes, it is beneficial to train the entire network after adding a new layer at the end [15].
- **Set the learning rate to be low:** It is essential to utilize a lower learning rate when compiling the new model because you are training a much larger model and want to readjust the pretrained weights. If not, your model may rapidly become overfit [15].

## References

- [1] Nikolas Adaloglou. *Transfer learning in medical imaging: classification and segmentation*. 2020. URL: <https://theaisummer.com/medical-imaging-transfer-learning/>.
- [2] Christopher M. Bishop and Hugh Bishop. *Deep Learning, Foundations and Concepts*. Springer, 2023.
- [3] Jason Brownlee. “A Gentle Introduction to Transfer Learning for Deep Learning”. In: (2019). URL: <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>.
- [4] Sasank Chilamkurthy. *Transfer Learning for Computer Vision Tutorial*. URL: [https://pytorch.org/tutorials/beginner/transfer\\_learning\\_tutorial.html](https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html).
- [5] Abolfazl Farahani et al. *A Concise Review of Transfer Learning*. Tech. rep. 2021. URL: <https://arxiv.org/abs/2104.02144v1>.
- [6] Michal Heker and Hayit Greenspan. *Joint Liver Lesion Segmentation and Classification via Transfer Learning*. Tech. rep. 2020. URL: <https://arxiv.org/pdf/2004.12352.pdf>.
- [7] Hee E. Kim et al. *Transfer learning for medical image classification: a literature review*. Tech. rep. 2022. URL: <https://bmcmimedimaging.biomedcentral.com/articles/10.1186/s12880-022-00793-7>.
- [8] Pedro Marcelino. *Transfer learning from pre-trained models*. 2018. URL: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>.
- [9] Basil Mustafa et al. *Supervised Transfer Learning at Scale for Medical Imaging*. 2021. URL: <https://arxiv.org/abs/2101.05913>.
- [10] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. *What is being transferred in Transfer Learning?* Tech. rep. 2020. URL: <https://arxiv.org/abs/2008.11687>.
- [11] Sinno Jialin Pan and Qiang Yang. “A Survey on Transfer Learning”. In: (2009). URL: <https://ieeexplore.ieee.org/abstract/document/5288526>.
- [12] Maithra Raghu et al. *Transfusion: Understanding Transfer Learning for Medical Imaging*. 2019. URL: <https://arxiv.org/abs/1902.07208>.
- [13] Richard Szeliski. *Computer Vision: Algorithms and Applications, 2nd ed*. Springer, 2022. URL: <http://szeliski.org/Book/>.
- [14] Abdel Aziz Taha and Allan Hanbury. “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool”. In: (2015). URL: <https://ncbi.nlm.nih.gov/pmc/articles/PMC4533825/>.

- [15] “What is Transfer Learning?” In: (). URL: <https://www.geeksforgeeks.org/ml-introduction-to-transfer-learning/>.
- [16] Fuzhen Zhuang et al. *A Comprehensive Survey on Transfer Learning*. Tech. rep. 2019. URL: <https://arxiv.org/abs/1911.02685>.