Paper: How transferable are features in deep neural networks?

Summary

In this paper, the authors present a method for quantifying the transferability of features from each layer of a neural network, which reveals their generality or specificity. They show that the transferability is negatively affected by optimization difficulties related to splitting networks in the middle of fragilely co-adapted layers and the specialization of higher layer features to the original task at the expense of performance on the target task. By training their network on the ImageNet dataset, they showed that either of these two issues may dominate depending on which layer you transfer i.e. bottom, top, or middle. They also show that the transferability of features decreases as the base and target dataset differ from each other. However, they conclude that transferring the features is always better than using random features.

The authors note that the first layer feature of a modern deep neural network resembles closely to common features such as Gabor filters or color blobs regardless of the exact cost function and dataset. The authors call such layer general. They call the last-layer features specific because these features depend greatly on the chosen dataset and task. What they are trying to study in this paper is the transition from general to specific features in the network and how it can be used in some other dataset. This technique of transferring features from the base network on a base dataset and task to a second target network to be trained on a target dataset and task is called transfer learning.

The authors created pairs of classification tasks A and B by constructing pairs of non-overlapping subsets of the ImageNet dataset. They constructed various CNN which different settings. baseA and baseB are the normal 8 layered CNN networks on A and B respectively. A selffer network B3B in which the first 3 layers are copied from baseB and frozen. The five higher layers are initialized randomly and trained on dataset B itself. A transfer network A3B in which the first 3 layers are copied from baseA and frozen. The five higher layers are initialized randomly and trained toward dataset B. There were also two other additional types of the network called B3B+ and A3B+ which are just like B3B and A3B respectively. However, the difference is that all layers in the CNN learn. The authors trained various versions of this network by varying the number of transfer layers chosen in the target network from the base network. Generally, for example, such networks are called BnB and AnB.

From their experiment, they found out that BnB's performance at layer one is the same as the baseB points. However, layers 3, 4, 5, and 6 exhibits worse performance. The performance drop is evidence that the original network contained fragile co-adapted features on successive layers. They also found out that layers 1 and 2 transfer almost perfectly from A to B in the AnB network. However, the performance drops as we include layer 3-7 which can also be tied back to the loss from co-adaptation and the drop from features that are less and less general. The result from the AnB+ network shows that transferring features and then fine-tuning them results in networks that generalize better than those trained directly on the target dataset, even if the target dataset is large.

They also tested their models on the dissimilar dataset: they split the ImageNet dataset into A and B, A being dataset with man-made classes and B being dataset with natural classes. They found out that the transferability gap when using frozen features grows more quickly as n

increases, n being the number of transferred layers, for dissimilar tasks than similar tasks. They also found out that transferring even from a distant task is better than using random filters.

Strengths

- The paper is really easy to follow, even for someone who may not have an extensive background in deep learning.
- I liked how they broke down their questions and arguments into bullet points. It makes it easier to understand what they are focusing on.
- The added supplementary materials are also very helpful for someone who wants to study this work further.

Weaknesses

- The paper could have made a much stronger argument if they have experimented on two separate datasets as well regarding transfer leaning on the dissimilar dataset.

Confusions

- I did not understand their reasoning behind choosing the same dataset and splitting it into 2 as man-made and natural classes and considering them dissimilar datasets. Why wouldn't the chose a completely different dataset?

Discussions

- How will the transfer learning perform in a much more dissimilar dataset than the one mentioned in the paper?
- The paper focuses specifically on object detection application using CNN. Can transfer learning be used in a different deep neural network for some different applications?
- How has transfer learning been used in present-day applications? Any popular ones?