

Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks

Link

[pdf](#)

Summary

- Training large CNN models require lots of training data which are not always available for the task at hand. To address this issue we can train a model on large scale image dataset and transfer the learned image representation to a task with limited amount of data. Here ImageNet is used for training a deep CNN model whose features are then transferred to Pascal VOC dataset.
- When transferring image features we need to be aware of following issues that might arise
 1. *Label bias*: Labels(class) in source task may be very different from labels in the target task.
 2. *Dataset capture bias*: The distributions of object orientations, sizes, mutual exclusion patterns etc. may be very different between the two tasks.
 3. *Negative data bias*: The target task may contain many other objects in the background that are not present in the source task.
- To address label bias we design an architecture that remaps the class labels between source and target task. To achieve this we remove the output layer of the pre-trained network and add an adaptation layer formed by two fully connected layers. The transferred layers are kept fixed (frozen) when training on the target task and only adaptation layer is trained.
- To address dataset bias we employ a sliding window strategy and extract around 500 square patches from each image for Pascal VOC dataset by sampling eight different scales on a regularly spaced grid with at least 50% overlap between neighboring patches.
- During test time we apply the network on each of the approximately 500 overlapping multi-scale patches and compute the overall contributions for each class.
- There was a 8% performance drop when training the identical network on Pascal VOC dataset from scratch(no pretrain) which indicates benefit of feature transfer.
- Overlap between classes in the source and target domains may have a positive effect on the transfer.