# 1 Summary

The paper addresses domain adaptation in deep networks by incorporating an empirical multi-kernel MMD based loss (between source and target domains) along with the usual Cross-Entropy loss. The MMD loss is calculated only from the deeper layers since they are responsible for domain-specific learning while Shallower layers' weights are either frozen or fine-tuned to target domain.

# 2 Strengths

- Using the kernel trick, exact computation of MK-MMD takes $O(n^2)$ time-complexity which is rather undesirable. The paper uses unbiased estimate of MK-MMD which can be computed in $O(n)$.
- The paper uses multiple kernel for domain adaptation of deep feature representations. They experimentally show that this substantially increases performance of the model.

# 3 Weaknesses

- The paper uses MMD loss in the last few layers of AlexNet network. It does not mention the effect of using MMD loss in all of the layers network. Will that cause non-significant improvement in domain generalization or will it hamper the network's ability to learn on source domain ?
- Use of unbiased estimate of MK-MMD in $O(n)$ does lead to faster computation but it also has a higher variance than the $O(n^2)$ exact computation. The paper does not address this issue nor does it say that it wont lead to decrease in performance.

# 4 Critique of Experiments

The author takes AlexNet architecture that is pre-trained on ImageNet dataset. The weights of first 3 layers are frozen and the layers 4-5 are fine-tuned on the target domain. To study the efficacy of multi-layer adaptation and multi-kernel MMD, he considers variants of DAN.

- Multi-kernel adaptation only in the last or second-last layer of network. (DAN$_7$ and DAN$_8$)
- Single-kernel MMD adaptation in the last 3 layers of network. (DAN$_{SK}$)
- Multi-kernel adaptation in the last 3 layers of network. (DAN)

It turns out that all DAN variants beat the performance of domain adaptation methods that were standard at the time (TCA, CNN, LapCNN). Among DAN variants, as expected, DAN obtains the best performance. This shows that multi-kernel MMD with multi-layer adaptation significantly effects the performance.

The t-SNE plots for DAN and DDC model are also analyzed. Its observed that DAN features on target domain are much more easily discriminated than DDC features on target domain. Also, DAN features are between source and target domain are much more better aligned categorically than DDC features.

# 5 Follow Ups/Extensions

- Instead of using MMD, we can use a different metric for quantifying domain discrepancy.
- To study task specific learning capability of layers, we can use MMD for shallower layers also and analyze the changes in network performance.