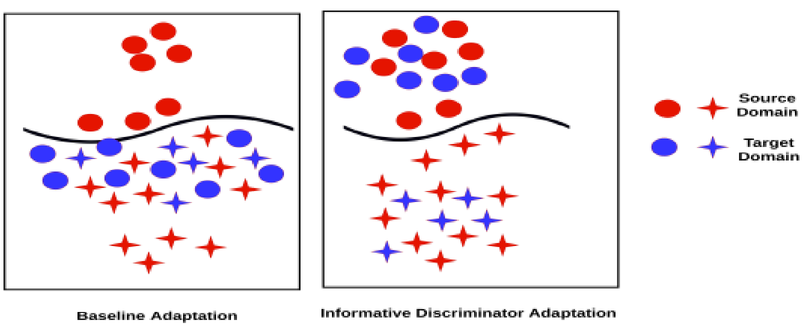
Title:	Understanding Transfer Learning between Domains and Tasks
Author(s):	Kurmi, Vinod
Supervisor(s):	Venkatesh, K S Namboodiri, Vinay P
Keyword(s):	Domain Adaptation Deep Learning Computer Vision Transfer Learning
Subject(s):	Deep Learning Computer Vision

Abstract:

Visual recognition has seen vast improvements based mainly on the success of deep learning-based models. These models are trained on very large annotated datasets such as Imagenet. The deployment of these generically trained models requires them to adapt to work in specific settings. However, the requirement of a large annotated dataset becomes a bottleneck for training networks in deep learning frameworks. Towards tackling these problems, transfer learning or domain adaptation based methods have been introduced. We consider the problem of domain adaptation for multi-class classification, where we are provided a labeled set of examples in a source dataset and a target dataset with no supervision. In this thesis, we consider various problems faced in domain adaptation, such as mode collapse and the fact that most adaptation is based on point estimates. We address these limitations by proposing a variety of techniques such as improved discriminator, considering distribution based adaptation, obtaining uncertainty, and through these obtaining localized adaptation and also addressing source data free adaptation. The first problem we addressed is mode collapse in domain adaptation using a multi-class discriminator. Adversarial learning-based techniques have shown their utility towards solving this problem using a discriminator that ensures source and target distributions are close. Our observation relies on the analysis that shows that if the discriminator has access to all the information available, including the class structure present in the source dataset, then it can guide the transformation of features of the target set of classes to a more structure adapted space. Using this formulation, we obtain state-of-the-art results for the standard evaluation on benchmark datasets. We further provide detailed analysis, which shows that using all the labeled information results in an improved domain adaptation. Another variant of this problem is proposed by an adversarial dropout discriminator. Here we suggest that rather than using a point estimate, it would be useful if a distribution based discriminator could be used to bridge this gap. This could be achieved using multiple classifiers or using traditional ensemble methods. In contrast, we suggest that a Monte Carlo (MC) dropout based ensemble discriminator could suffice to obtain the distribution based discriminator. Specifically, we propose a curriculum-based dropout discriminator that gradually increases the variance of the sample-based distribution, and the corresponding reverse gradients are used to align the source and target feature representations. All the regions of an image are not transferrable, and thus in the next method, we proposed a discriminator certainty based attention for localized domain adaptation. In an image, there would be regions which can be adapt better; for instance, the foreground object may be similar in nature. To obtain such regions, we propose

methods which consider the probabilistic certainty estimate of various regions and specifically focus on the same during classification for adaptation. We observe that just by incorporating the probabilistic certainty of the discriminator while training the classifier, we are able to obtain state of the art results on various datasets as compared against all the recent methods. In the next problem, we propose a source data free domain adaptation, i.e. the source data is not available for the adaptation, rather we have only provided the weights of the model that are optimized for source data prediction. In this case, first, we obtained the class impression of source data using the GAN setting. These impressions can work for the dummy source data, and these can be further used for the adaptation of target samples. All the adaptation frameworks have been proposed in the transductive setting, i.e., all the data is available in hands before the experiment. For exploring the inductive setting, first, we worked with the incremental learning scenario. Obtaining deep learning models that can be used for incremental or continual learning is a major challenge. There have been several approaches towards solving this problem, including those based on knowledge distillation. However, these methods have not widely considered approaches that are interpretable. We tackle these by proposing a method that uses attention based knowledge distillation and uncertainty based distillation to obtain an interpretable incremental learning approach. Our evaluation suggests that by adopting these techniques, we not only gain advantage in terms of interpretability of the incremental learning models so obtained, but we can also improve over the former proposed methods in terms of accuracy based on standard benchmarks. To conclude, we present different transfer learning methods between the domains and modalities. In addition, we also propose a framework for adaptation where we can not have the source data. These contributions surely help in further research to solve the transfer leaning based problems in general.



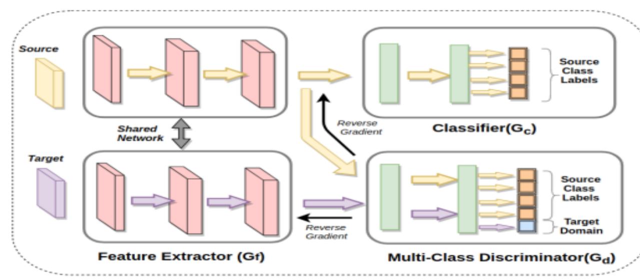


Figure 3.2: The proposed architecture consists of a feature extractor (G_f), a classifier (G_c) and a multiclass discriminator (G_d). The discriminator trained adversarial with the feature extractor by reverse the gradient.