

Abstract

Computer vision systems have been developed which perform well at recognizing and retrieving natural images as well as document images. However, these systems might not work well under certain scenarios, for instance, when the distribution of the training and the test data do not match. For example, an OCR system may not work on those target fonts which are very different from the fonts used while training. Moreover, such systems might not be able to tackle previously unseen categories. These scenarios limit the real world applications of computer vision systems to some extent. In this thesis, we tackle these problems by designing classifiers that work even in novel scenarios. We design algorithms for retrieving document images, as well as recognizing objects and digits in novel settings.

For the document image retrieval task, we consider two different scenarios. In the first scenario, we tackle the issue of novel query words in a classifier based retrieval system. We present a one-shot learning strategy for the learning of discriminative classifiers given a novel query word. This strategy utilizes the classifiers learned at the training time in order to obtain the classifier corresponding to the underlying query class. This extends the classifier based retrieval paradigm to an unlimited number of classes (words) present in a language. We validate our method on multiple datasets, and compare it with popular alternatives like OCR and wordspotting. In the second scenario, we tackle the problem of mismatch between the source and the target style(font). We tackle this problem by style(font)-content(word label) factorization strategy. Based on the style-content factorization, we present a semi-supervised style transfer strategy to transfer word images in the source font to the target font. We also present a nonlinear style content factorization for obtaining style independent representation of word images. We validate both these strategies on scanned document collections as well as multi-font synthetic datasets. We show mean average precision gains of upto 0.30 over the baseline using our nonlinear factorization strategy.

For the recognition task, we tackle the challenging problem of dataset shift between the source and the target data. Dataset shift is the scenario where the joint input-output distribution for the training and the test data are different. In such a scenario, the classifiers trained on the source data might perform poorly on the target data. We tackle two different tasks in this scenario, i.e. digits recognition and object recognition. The two domains we consider for the digits recognition task are handwritten and printed digits. To tackle the digits recognition task, we present a subspace alignment based strategy. In this approach, labeled source domain data and unlabeled target domain data is used to learn transformations for the two domain which reduces the mismatch between the two domains. A source domain classifier learned after applying the transformations would work well even on the target domain data. We consider

the simple nearest neighbor based classifier for validating this claim. For the object recognition task, we present a sparse representation based strategy. A dictionary learned from the source data might not be suitable for sparsely representing target domain data and vice-versa. Hence, we present a partially shared dictionary learning strategy which results in dictionaries which are suitable for representing the source as well as the target domains. We show our results on popular benchmark datasets and show improvement over the state of art approaches.