

Distance Metric Learning for Visual Recognition

How to design a favorable distance function plays an important role in many computer vision and pattern recognition tasks. Over the past decade, a variety of distance metric learning algorithms have been proposed in the literature and most of them have obtained the reasonable success in many visual recognition applications, such as face recognition, image classification, person re-identification and visual search. However, there are still several issues to be addressed to distance metric learning. In this thesis, we propose several deep metric learning methods and multi-view metric learning methods, and apply them to various visual recognition applications to demonstrate their effectiveness.

First, most metric learning methods usually seek a single linear transformation to map data points into a new feature space such that they may not be powerful enough to exploit the nonlinear relationship of data points. Even if the kernel trick is employed to map data points into a high-dimensional feature space under which a discriminative distance metric is learned, the kernel-based metric learning methods still suffer from the scalability problem as they cannot obtain the explicit nonlinear mapping functions. To explicitly address both the nonlinearity and scalability problems, we propose a discriminative deep metric learning (DDML) method by using deep neural network architecture. The DDML trains a neural network which learns a set of hierarchical nonlinear transformations to project sample pairs into the same feature subspace, under which the distance of each positive sample pair is less than a smaller threshold and that of each negative pair is higher than a larger threshold respectively, such that more discriminative information can be used in the neural network. Extensive experiments for face verification and kinship verification demonstrate the efficacy of the proposed method.

Second, most existing metric learning methods usually assume that the training and test samples are from the same distribution. This assumption doesn't hold in many real-world visual recognition applications. To address this problem, we propose a deep transfer metric learning (DTML) method to learn a set of hierarchical nonlinear transformations for cross-domain visual recognition by transferring discriminative knowledge from the labeled source domain to the unlabeled target domain. Specifically, the proposed DTML learns a deep metric network by maximizing the inter-class variations and minimizing the intra-class variations, and minimizing the distribution divergence between the source domain and the target domain at the top layer of the network. To better exploit the discriminative information from the source domain, we further develop a deeply supervised transfer metric learning (DSTML) method by including an additional objective on DTML, where the

output of both the hidden layers and the top layer are optimized jointly. To preserve the local manifold of input data points in the metric space, we present two new methods, DTML with autoencoder regularization (DTML-AE) and DSTML with autoencoder regularization (DSTML-AE). Experimental results on face verification, person re-identification, and handwritten digit recognition validate the effectiveness of the proposed methods.

Third, it is desirable to learn distance metrics from multiple feature representations such that more discriminative information can be exploited. To explore multiple feature representations in metric learning, we first propose a large-margin multi-metric learning (LM^3L) method to collaboratively learn multiple distance metrics from multiple feature representations of data, where one distance metric is learned for each feature and the correlations of different feature representations of each sample are maximized, and under the learned metric spaces the distance of each positive pair is less than a smaller threshold and that of each negative face pair is more than a larger threshold, respectively. In addition, we also propose two local distance metric learning approaches, i.e., local metric learning (LML) and local large-margin multi-metric learning (L^2M^3L), to better exploit the local manifold structures of data points. Experimental results on face verification and kinship verification tasks show the efficacy of the proposed methods.