Statistical and Geometrical Alignment using Metric Learning in Domain Adaptation

Rakesh Kumar Sanodiya*, Alwyn Mathew+, Jimson Mathew +,Matloob Khushi°
*National Taipei University of Technology, Taipei 10608, Taiwan, Email: rakesh.pcs16@gmail.com
+Computer Science and Engineering, Indian Institute of Technology Patna, Patna, Bihar, India
°School of Computer Science, The University of Sydney, Sydney, Australia

***Abstract*—Domain adapted machine learning is driven by the possibilities of learning from source data distribution to understand different target data distributions. An assumption is made that one application (source) domain always has enough labeled information, but the other related application (target) may contain information that is partially labeled or completely unlabeled. Therefore, it is necessary to train the target domain classifier using enough labeled information of the source domain. However, contrary to primitive assumptions, the source domain and target domain data need not have the same distribution. Therefore, we can't directly use data of source domain to train classifier for data of target domain. Existing approaches can be deprived of one or more objectives: perform geometric diffusion on the manifold, align the cross-domain distributions, preserve the discriminative information using metric learning. Here, we have proposed a novel framework that aims to meet all such objectives. In this framework, we proposed two methods, statistical and geometrical alignment using metric learning with pseudo labels (SGA-MDAP) and without pseudo labels (SGA-MDA) in visual domain adaptation. It has been demonstrated through various experiments that our framework outperforms various state-of-the-art methods over four different real-world cross-domain visual identification datasets such as PIE face, ORL face, Yale face, and Office Caltech.**

***Index Terms*—Domain Adaptation, Transfer Learning, Classification, Metric learning, Manifold**

I. Introduction

Traditional statistical learning is based on the basic assumption that training and test data must be managed by a single underlying distribution (1). However, in real-world applications, e.g, face-to-face identification, speech recognition, statistical machine learning algorithms may fail because such applications follow different distributions and are less correlated with each other. For example, suppose we deploy an Android application to identify objects in an image taken with a mobile phone camera. Images such as C05 shown in Figure 1 were captured from old mobile phone cameras in different scenarios (visual and expressive) and annotated. If the Android app with the traditional classifier was trained with C05 images, then does this app work well with C27 images shown in Figure 1 that were captured with new mobile phone camera ?. Our intuition says no because both type of images C05 and C27 are differ in terms of their distributions. One solution for this problem can be to train the classifier with C27 images but it requires labeled images. However, in real-world scenarios it is a time consuming and manual task to annotate newly captured images.Another possible solution is to use old captured images in a way that it can be used to predict the label of its newly related captured images. To overcome these limitations, domain adaptation (DA) or transfer learning (TL) (2) plays an important role. These face or object recognition applications include both types of data sets, one taken from the source domain and the other taken from the target domain.

Domain adaptation is a robust approach which leverages source domain information for predicting the target domain information. It can be distinguished into two categories according to availability of the target domain data: Unsupervised domain adaptation (3; 4) and Semi-supervised domain adaptation (5). The key to domain adaptation is to exploit labeled source domain data and unlabeled target data effectively. Domain adaptation can be classified into three main sub-parts: (a) instance re-weighting (6)(b) feature matching (7) (c) metric-based transfer learning (8; 9).

The major contributions of this paper are:

- Our proposed framework is a new approach that overcomes the limitations of primitive and DA algorithms.
- In this framework, we propose two methods, statistical and geometrical alignment using metric learning with pseudo labels (SGA-MDAP) and without pseudo labels (SGA-MDA) in visual domain adaptation.
- Our proposed SGA-MDA approach achieved 83.62%, 86.16%, 73.62%, and 96.83% while SGA-MDAP attains 69.09%, 83.76%, 62.58%, and 84.86% mean accuracies over the PIE Face, Office-Caltech, ORL, and Yale datasets, respectively. The mean accuracies achieved by our methods are much higher than all other compared approaches.

II. Related Works

Here, we've discussed about the literature survey and the comparative analysis of the proposed methods with the current methods.

A. Existing works on Primitive and DA Approaches

In this section, we've given a brief explanation about the state-of-the-art approaches that are related to our proposed framework.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Si et al.(10) proposed Transfer Subspace Learning (TSL), where Bregman projection was used for divergence minimization. In (11), Ding et al. used effective knowledge transfer to facilitate learning process in the target domian. Transfer component analysis (TCA) (2) uses maximum mean miscrepancy (MMD) in the Hilbert space of the kernel. In (12), DA problem is solved by evaluating the relationships between given domains with transferred weights. Geodesic Flow Kernel (GFK) (13) performs domain adaptation using a kernel-based approach. Robust Visual Domain Adaptation with Low-Rank Reconstruction (RVDLR) (14) is an approach where source domain data is transformed in such a way that each of these points can be reconstructed by the target domain data points. Marginalized Stacked Denoising Autoencoder (mSDA) (15) takes corrupted input data, tries to eliminate noise and recovers original features after reconstruction. In Joint Distribution Adaptation (JDA) (7), PCA and MMD are used to jointly adapt marginal and conditional distributions. Subspace Alignment (SA) (16), on the other hand transforms source subspace into target subspace. When Cross-Domain Metric Learning Based on Information Theory (CDML) (17) is applied, the distributions are minimized using a Mahalanobis distance metric. Transfer Joint Matching (TJM) (18) uses feature matching and instance reweighting to reduce the domain difference. In Low-Rank Transfer Subspace Learning (LTSL) (19),the target data is represented in relation with the source data. Meherkanoon et al.(20) introduced a model that uses generalized eigen problems to reveal the general representation of both domains' data. In the case of Low-Rank and Sparse Representation (LRSR) (21), the source and target data are transformed into a common subspace such that the entire target domain can be reconstructed from the source domain. Robust transfer metric learning (RTML) framework (22) finds a transfer low rank metric and minimizes both marginal and conditional distributions. Zhang et al. (3) introduced joint Geometric and Statistical Alignment (JGSA), where geometric and statistical distributions are implemented simultaneously. Sun et al.(23) presented correlation alignment algorithm called (CORAL) that adopts a support vector machine (SVM) on the similarity matrix. Luo et al.(24) presented three domain adaptation (DA) methods, namely CDDA (Close yet Discriminative and Geometry Aware DA), GA-DA(Geometry Aware Discriminative and Geometry Aware Domain adaptation), and DGA-DA (Discriminative and Geometry Aware Domain Adaptation).

B. Shortcomings of Existing DA Methods

In cases of large difference between the two domains, the current DA methods deprived to the below stated targets:

1) Use of metric learning for preserving discrepancy between source and target domains.
2) Statistical alignment of the two domains by minimizing distributions such as marginal and conditional (2; 25; 26).
3) Geometric alignment of the given domains by utilizing the same geometric properties in the two domains.

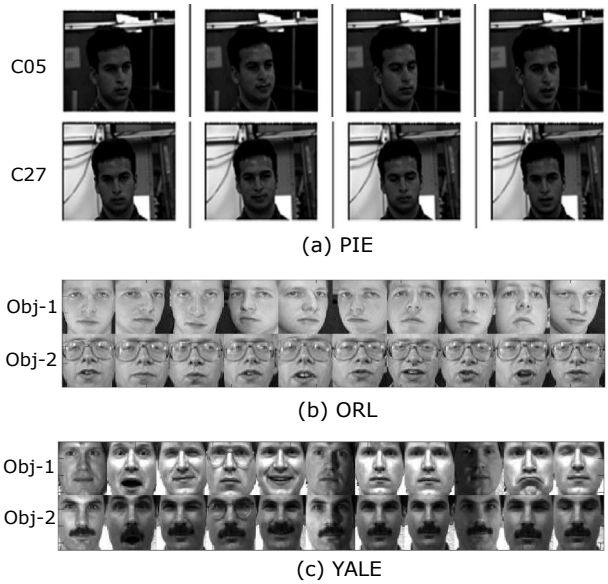


Fig. 1. (a) sample images from two different domains C05 and C27 of PIE face dataset (b) two persons (Obj-1 and Obj-2) samples with different expressions of ORL dataset (c) two persons (Obj-1 and Obj-2) samples with different expressions of Yale dataset

III. A Unified Framework for Statistical and Geometrical Alignment in Visual Domain Adaptation

To learn the appropriate metric for the target domain, many objectives as discussed in subsection II-B are required to be satisfied. In this paper, we provide a unified framework that can unify the limitations of existing transfer learning and domain adaptation approaches for statistical and geometrical alignment using metric learning in Visual Domain Adaptation. Here, we propose two methods, SGA-MDAP and SGA-MDA based on the availability of the target domain class-wise mean values. If the class-wise mean values of the target domain are not given, then we use the SGA-MDAP method otherwise use SGA-MDA method. In SGA-MDAP method, we first calculate class-wise mean values by generating the pseudo label of the target-domain in each iteration.

In this section, we first define the domain adaptation problem, then formulate the metric learning problem in DA and perform its optimization.

A. Notations and Problem Definition

We assume that X_s and X_t denote source domain and target domain data-sets, respectively.

As source and target domain images have different attributes such as view angles, pose, etc. there is a lot of discrepancy in their marginal and conditional distribution, i.e.$P_s(X_s) \neq P_t(X_t)$ and $P_s(Y_s|X_s) \neq P_t(Y_t|X_t)$. Hence, the target domain classifier can't be trained on source domain images.

B. Metric Learning in Domain Adaptation

Metric learning refers to learning a distance function that captures the geometry of data in addition to Euclidean dis-

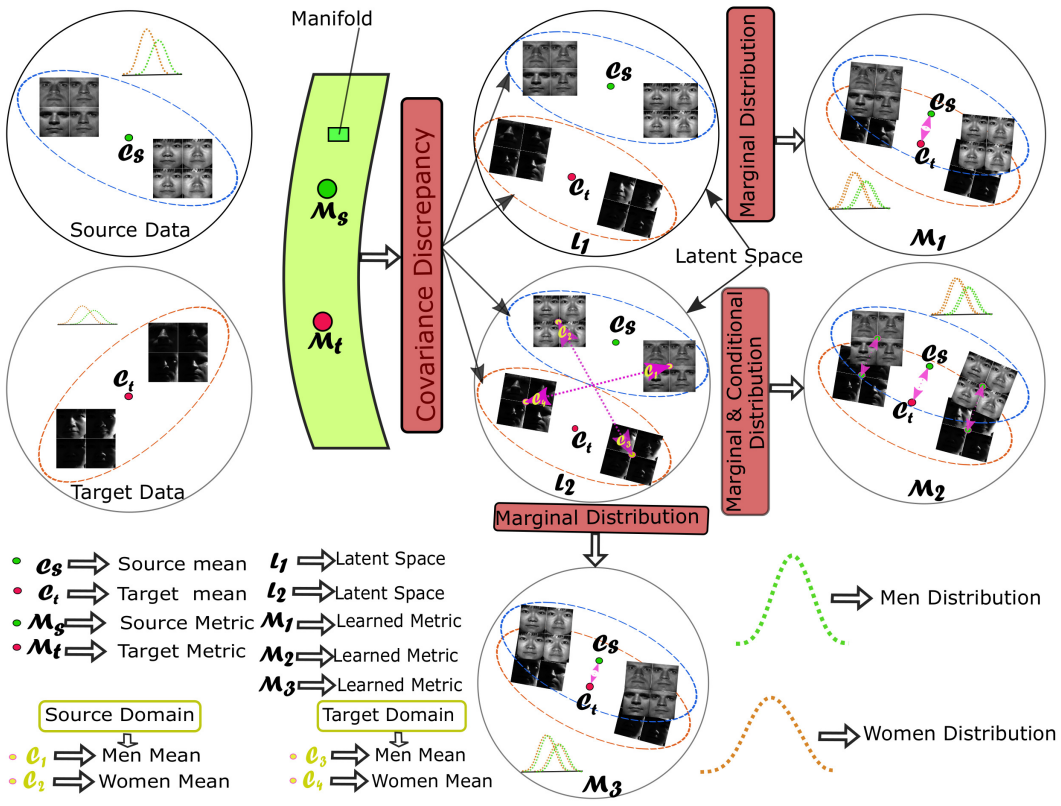


Fig. 2. A conceptual diagram of our proposed framework for statistical alignment is represented using some face data-set. The data-set consist of two classes men and women in which each class contains well-lit and dark-colored face images. We consider well-lit images as source domain and dark-colored images as target domain. The source and target domains are mapped to latent space using the M_s and M_t matrices. The possible matrices L_1 and L_2 are defined in latent space that are learned to maximize the discriminative of the data. After imposing a marginal distribution in latent space L_1, the learned metric is M_1, where the same class examples are close in both the domains. But when the marginal distribution is applied to L_2 space, where there is lot of discrepancy, the learned metric is M_3 where the samples of different classes collapse. So, we unify both marginal and conditional distributions into a statistical loss to minimize it. In the resultant, learned metric is M_2 where same class samples are close even after having lot of discrepancy in both the domains.

TABLE I
NOTATIONS WITH CORRESPONDING INTERPRETATIONS.

Interpretation	Notation
# samples for source and target domain	n_s, n_t
source/target domain Co-variance metrics	M_s, M_t
# classes	C
# nearest neighbor in KNN graph	K
source unlabelled samples' set and respective labelled set	X_s, Y_s
target unlabelled samples' set and respective labelled set	X_t, Y_t
# samples in both the domains and dimension of each sample and total	N, d
conditional and marginal distribution's alignment term	D_{st}
Pseduo label of taget domain	Y_p
proposed framework objective function	$J(M)$
source domain marginal distribution	$P_s(X_s)$
target domain marginal distribution	$P_t(X_t)$
source domain conditional distribution	$P_s(Y_s\|X_t)$
target domain conditional distribution	$P_t(Y_s\|X_t)$
eigenvector	Q
optimal learned metric	M
original input data in both the domains	X
Co-variance matrix of X.	$Cov(X)$
identity matrix	I
trade-off parameters	δ, t, γ
# iteration	T
accuracy	acc

tance. Mahalanobis distance metric, one of the most widely used metric learning algorithms measures the similarity between data points using covariance matrix and distance relations. Thus, the Mahalanobis distance d_m between a pair of data points x_i and x_j can be formulated as follows:

$$d_m(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \tag{1}$$

where M is a positive semi-definite matrix and it can be decomposed into it's eigenvectors as $M = Q^T Q$.

In our proposed framework, we formulate the problem of metric learning in domain adaptation using geometric mean metric learning (GMML) (27). In GMML, symmetric positive semi-definite matrix attempts to represent source domains with M and the target domain with M^{-1}. The problem of metric learning in DA is as follows:

$$\begin{aligned} \min_{M \prec 0} & \sum_{(x_i, x_j) \in X_s} \text{tr}(M(x_i - x_j)(x_i - x_j)^T) \\ + & \sum_{(x_i, x_j) \in X_t} \text{tr}(M^{-1}(x_i - x_j)(x_i - x_j)^T) \end{aligned} \tag{2}$$

In our framework,M_s and M_t are used that can be expressed as follows:

$$M_s = \sum_{(x_i, x_j) \in X_s} (x_i - x_j)(x_i - x_j)^T \tag{3}$$

$$M_t = \sum_{(x_i,x_j)\in X_t} (x_i - x_j)(x_i - x_j)^T \tag{4}$$

Now, the problem of metric learning in domain adaptation can be reduced to the following objective function:

$$\min_{M\prec 0} J(M) = \text{tr}(MM_s) + \text{tr}(M^{-1}M_t) \tag{5}$$

After optimizing the objective function J(M) in Equation 5 with respect to the metric M, the possible latent metrics, $L1$ and $L2$, are shown in Figure 2. So, we move towards statistical and geometrical learning metric.

1) Marginal and Conditional Distributions Alignment: The method proposed here takes into account both the distributions for identical statistical alignments of both the domains. The details of these distributions are given in Figure 2. MMD, used to reduce the marginal distributions between the two domains in a d-dimensional subspace is calculated as follows (2; 28; 29).

$$\min_{Q} \left\| \frac{1}{n_s} \sum_{x_i\in X_t} Q^T x_i - \frac{1}{n_t} \sum_{x_j\in X_t} Q^T x_j \right\|_F^2 = \text{tr}(Q^T X B_m X^T Q) \tag{6}$$

In cases where data distribution is in classes, minimizing conditional distribution is a must, which is calculated using equation given in Long et al.(18) as follows:

$$\min_{Q} \sum_{i=1}^{C} \left\| \frac{1}{n_s^i} \sum_{x_i\in X_s^i} Q^T x_i - \frac{1}{n_t^i} \sum_{x_j\in X_t^i} Q^T x_j \right\|_F^2 = \text{tr}(Q^T X B_c X^T Q) \tag{7}$$

Together the previous two Equations 6 and 7 can be summarised in the following way.

$$D_{st} = \text{tr}(MX(\delta B_m + (1-\delta)B_c)X^T) \tag{8}$$

Here B_m and B_c can be estimated as discussed in (7).

2) Laplacian Regularization: This term uses the similarity in the geometrical properties of the points which are near or far to each other to model the geometry of the manifold in both the domains. The similarity between any two samples, x_i and x_j, can be expressed as follows:

$$w_{ij} = \left\{ \begin{array}{cc} \exp(-\frac{\|x_i - x_j\|^2}{\sigma_i^2}) & \text{if } x_i \in \mathcal{N}_k(x_j) \text{ or } x_j \in \mathcal{N}_k(x_i) \\ 0 & \text{otherwise} \end{array} \right\} \tag{9}$$

where $\sigma_{ij}^2 = \sigma_i \sigma_j$ and σ_i equals the distance between element x_i and its k^{th} nearest neighbor and $\mathcal{N}(x_i)$ shows the set of k nearest neighbors to sample x_i.

Now based on the manifold assumption as discussed in (6), we can define a Laplacian regularization term as follows:

$$\begin{aligned} L =& \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \|x_i - x_j\|_M^2 \qquad (10)\\ =& \text{tr}(M \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(\|x_i\|^2 + \|x_j\|^2 - 2(x_j)^T x_i)) \\ =& 2\text{tr}(M(\sum_{i=1}^{N} h_i \|x_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_j)^T x_i)) \\ =& 2\text{tr}(MX(D-W)X^T) \\ =& 2\text{tr}(MXL_uX^T) \qquad (11) \end{aligned}$$

where W is a similarity and dissimilarity matrix between all data samples available in both the domains' dataset X, $h_i = \sum_{i=1}^{N} w_{ij}$, D is a diagonal matrix, $L_u = D - W$ is the Laplacian matrix of W, and w_{ij} is the similarity between $(i,j)^{th}$ sample of it.

3) Formulation of Objective Function: We have included two terms as discussed in subsection III-B1,III-B2 to the proposed objective function J III-B. These are D_{st} terms that reduces conditional and marginal distribution and a Laplacian regularization L. Thus, the objective function of proposed framework can be written as follows:

$$\min_{M\prec 0} J(M) = \text{tr}(MM_s) + \text{tr}(M^{-1}M_t) + D_{st} + L) \tag{12}$$

Further simplification and substitution gives us the below equation:

$$\begin{aligned} \min_{M\prec 0} J(M) =& \text{tr}(MM_s) + \text{tr}(M^{-1}M_t) + \text{tr}(MX(\delta B_m \\ &+ (1-\delta)B_c + \gamma L_u)X^T) \end{aligned} \tag{13}$$

C. Optimization

In this subsection, we have optimized the objective function $J(M)$ and estimated the value of M. The partial derivative of $J(M)$ w.r.t. M in Equation 13 can be written as:

$$\begin{aligned} \frac{\partial J(M)}{\partial M} =& \text{tr}(M_s) - \text{tr}(M^{-1}M_tM^{-1}) + \text{tr}(X(\delta B_m \\ &+ (1-\delta)B_c + \gamma L_u)X^T) \end{aligned} \tag{14}$$

By keeping the $\frac{\partial J(M)}{\partial M} = 0$, we find out M as follows:

$$M_t = (M_s + X(\delta B_m + (1-\delta)B_c + \gamma L_u)X^T))M^2 \tag{15}$$

$$M = (M_s + X(\delta B_m + (1-\delta)B_c + \gamma L_u)X^T))^{-1}\theta_{\frac{1}{2}}M_t \tag{16}$$

Where $\theta_{\frac{1}{2}}$ denotes the geometric mean of SPD matrices.

In order to find out meaningful intermediate subspaces between both the SPD matrices $(M_s + X(\delta B_m + (1-\delta)B_c + \gamma L_u)X^T))^{-1}$ and M_t, we use geodesic distance proposed by Gong et al. (13). This distance is represented as follows:

$$\varnothing(t) = G^{\frac{-1}{2}}(G^{\frac{-1}{2}} M_t G^{\frac{-1}{2}})^t G^{\frac{-1}{2}} \quad (17)$$

where $G = (M_s + X(\delta B_m + (1-\delta)B_c + \gamma L_u)X^T))$

From Equation 17, it can be seen that if the value of t=0 means $\varnothing(0) = G^{-1}$ and if t=1 then $\varnothing(1) = M_t$.

IV. Experimental Evaluations

This section describes the results and observations of the proposed methods for the four open real-world datasets, including PIE face, ORL face, Yale Face and Caltech. To show the strengths of SGA-MDAP and SGA-MDA, we compare their accuracies with other state-of-the-art methods' accuracies.

A. Benchmark data sets

We used the datasets of open-world problems such as PIE face, Office-Caltech with the current state-of-the-art Deep-net (VGG-Net), ORL, and Yale for experimentation and validation. Sample images of included datasets such as PIE, ORL, and Yale are shown in Figure 1.

PIE Face dataset: PIE Face Recognition Database consists of 68 subjects images, which were taken from light locations, having different view angles and expressions. All such images are cropped to 32×32 pixels. We adopt five subsets of poses: C05, C07, C09, C27, and C29. We can select one pose subset as the source and any rest one as a target. As a result, 20 source / target combinations of DA task, namely PIE 1 (5→7), PIE 2 (5→9), PIE 3 (5→27), etc. are obtained.

Office+Caltech dataset: The Office+Caltech dataset (13) contains images in ten different categories that are classified into four domains, i.e., Amazon, DSLR, Webcam, and Caltech. We have used the current state-of-the-art deep-net features (VGG-Net) for the experiment. In this experiment, we represent the dataset Caltech-256, Amazon, DSLR, and Webcam as C, A, D, and W and obtain 12 source / target combinations of DA tasks, namely A→C, A→D, A→W, etc.

ORL face dataset: The ORL face dataset (30) includes face images of 40 individual persons captured in different scenarios. Here also, all the images are normalized and cropped to 32×32 pixels for the experiment. While experimenting with this dataset, we consider some images as source domain and the remaining images as target domains and generate the following tasks : 1→9, 2→8, 3→7, and 4→6, where 1→9 means 1 image of each person will be taken as source and rest 9 images of corresponding person will be taken as target.

Yale face dataset: The Yale face dataset (31) includes a total of 165 images of 15 individuals persons where each person has 11 images. These images are captured with different facial expression and in different lighting conditions. All these images are normalised and cropped to 32×32 pixels. Similar to ORL dataset, we generate the tasks as follows: 2→9, 3→8, and 4→7, where 2→9 means 2 images of each person will be taken as source and rest 9 images of corresponding person will be taken as target.

B. Comparison with State-of-the-art Approaches

Our proposed approaches have been verified and compared with several state-of-the-art methods: NN, SVM , PCA, TSL (10), TCA (2), GFK and GFK-PLS (13), mSDA (15), RVDLR (14), JDA (7), ILS (32), SA (16), TJM (18), CDML (17), LTSL (19), LRSR (21), CORAL(23), RTML (22), JGSA (3), and CDDA,GA-DA and DGA-DA (24)

C. Parameter Sensitivity Tests

A parameter sensitivity analysis was performed to find the appropriate values of all the parameters for the proposed methods. However, we have shown the performance of SGA-MDA with respect to the parameters in Fig. 3 and 4 for PIE face and Caltech dataset respectively. The parameters values of SGA-MDAP are similar to SGA-MDA method.

1) Experimental analysis on t, γ and δ: For finding optimal trade-off parameters, we vary parameters t, γ and δ, values from 0.1 to 0.9, from 10^{-3} to 10^3, and from 0 to 1, respectively for PIE face and Caltech dataset. The corresponding graphs have been shown in Fig. 3(a,c, and d). From Fig.3, the following conclusions can be made:

The graph provides evidence that the t parameter (Fig.3(a)) gives the best results for parameter values in the range $[0.1, 0.6] \in t$.

From 3(c) γ parameter variation graph, the proposed method is outperforming at 10^{-2} for all tasks.

The graph (Fig. 3(d)) shows that δ parameter gives the best results when it's value is higher than 0.

2) Parameter K of KNN Graph : We analyze the performance of our method with the variation of the K-parameter (no. of nearest neighbors) values from 1 to 10 and report the performance in the Figure 3 (b). From Fig. 3(b), we can determine that our approach performs better on K = 1.

3) Parameter D for dimension of data: In addition to these parameters included in SGA-MDA, we use dimensional reduction method called Principle component analysis (PCA). Therefore, we conduct parameter sensitivity analysis test for parameter D on all the datasets and report performance of our method in Fig. 4. Fig. 4 shows that an increase in the value of D for PIE face datasets increases the performance but our proposed method for Caltech datasets performs well on parameter $D = 40$.

Based on the experiments performed, we summarize the suggested parameters values in Table V for all the data-sets.

D. Comparing the Results with State-of-the-art Approaches

For the PIE face dataset, proposed method SGA-MDA attains 83.62% mean accuracy while SGA-MDAP attains 69.09%, each more efficient than current techniques.

In Office-Caltech dataset's case, our SGA-MDA gives 86.16 % mean accuracy followed by our SGA-MDAP with 83.76%,each greater than all previous methods' accuracies.

In case of Yale data set, proposed SGA-MDA and SGA-MDAP outperform all other techniques with respective mean accuracies of 73.62% and 62.58%.

TABLE II
COMPARISON OF ACCURACIES OF THE PROPOSED METHODS SGA-MDAP AND SGA-MDA WITH OTHER APPROACHES FOR DIFFERENT TASKS OF PIE FACE DATASET.

Task	NN	PCA	RTML	GFK	CDML	RDALR	LTSL	mSDA	JDA	TSL	LRSR	TCA	GA-DA	CDDA	DGA-DA	TJM	JGSA	**SGA-MDAP**	**SGA-MDA**
PIE 1 5→7	26.09	24.80	60.12	26.15	53.22	40.76	22.96	28.35	58.81	44.08	65.87	40.76	57.40	60.22	65.32	29.52	68.07	65.00	**89.25**
PIE 2 5→9	26.59	25.18	55.21	27.27	53.12	41.79	20.65	26.91	54.23	47.49	64.09	41.79	60.54	58.70	62.81	33.76	67.52	71.38	**89.39**
PIE 3 5→27	30.67	29.26	85.19	31.15	80.12	59.63	31.81	30.39	84.50	62.78	82.03	59.63	84.05	83.48	83.54	59.20	82.87	92.82	**94.29**
PIE 4 5→29	16.67	16.30	52.98	17.59	48.23	29.35	12.07	21.76	49.75	36.15	54.90	29.35	52.21	54.17	56.07	26.96	46.50	48.77	**84.92**
PIE 5 7→5	24.49	24.22	58.13	25.24	52.39	41.81	18.25	28.27	57.62	46.28	45.04	41.81	57.89	62.33	63.69	39.40	25.21	67.58	**79.71**
PIE 6 7→9	46.63	45.53	63.92	47.37	54.23	51.47	16.05	44.19	62.93	57.60	53.49	51.47	61.58	64.64	61.27	37.74	54.77	74.81	**87.50**
PIE 7 7→27	54.07	53.35	76.16	54.25	68.36	64.73	45.15	55.39	75.82	71.43	71.43	64.73	82.34	79.90	82.37	49.80	58.96	85.43	**89.21**
PIE 8 7→29	26.53	25.43	40.38	27.08	37.34	33.70	17.52	28.08	39.89	35.66	47.97	33.70	41.42	44.00	46.63	17.09	35.41	56.80	**79.35**
PIE 9 9→5	21.37	20.95	53.12	21.82	43.54	34.69	22.36	24.83	50.96	36.94	52.49	34.69	54.14	58.46	56.72	37.39	22.81	57.02	**75.57**
PIE 10 9→7	41.01	40.45	58.67	43.16	54.87	47.70	20.26	42.59	57.95	47.02	55.56	47.70	60.77	59.73	61.26	35.29	44.19	60.58	**79.68**
PIE 11 9→27	46.53	46.14	69.81	46.41	62.76	56.23	57.34	50.25	68.45	59.45	77.50	56.23	77.23	77.20	77.83	44.03	56.86	81.22	**88.22**
PIE 12 9→29	26.23	25.31	42.13	26.78	38.21	33.15	24.57	27.83	39.95	36.34	54.11	33.15	43.50	47.24	44.24	17.03	41.36	50.91	**77.08**
PIE 13 27→5	32.95	31.96	81.12	34.24	75.12	55.64	51.20	32.89	80.58	63.66	81.54	55.64	79.83	83.10	81.84	59.51	72.14	84.51	**88.35**
PIE 14 27→7	62.68	60.96	8.92	62.92	80.53	67.83	70.10	63.01	82.63	72.68	85.39	67.83	84.71	82.26	85.27	60.58	88.27	84.96	**89.93**
PIE 15 27→9	73.22	72.18	89.51	73.35	83.72	75.86	72.00	74.70	87.25	83.52	82.23	75.86	89.17	86.64	90.95	64.88	86.09	92.03	**93.25**
PIE 16 27→29	37.19	35.11	56.26	37.38	52.78	40.26	48.28	34.81	54.66	44.79	72.61	40.26	53.62	58.33	53.80	25.06	74.32	69.91	**84.68**
PIE 17 29→5	18.49	18.85	29.11	20.35	27.34	26.98	13.06	25.85	46.46	33.28	52.19	26.98	52.73	48.02	57.44	32.86	17.52	45.61	**68.84**
PIE 18 29→7	24.19	23.39	33.28	24.62	30.82	29.90	21.61	26.33	42.05	34.13	49.41	29.90	47.64	45.61	53.84	22.89	41.06	55.98	**74.76**
PIE 19 29→9	28.31	27.21	39.85	28.49	36.34	29.90	17.03	28.63	53.31	36.58	58.45	29.90	51.66	52.02	55.27	22.24	49.20	65.87	**78.61**
PIE 20 29→27	31.24	30.34	47.13	31.33	40.61	33.64	29.59	32.98	57.01	38.75	64.31	33.64	58.82	55.99	61.82	30.72	34.75	70.62	**79.75**
Average	34.76	33.85	58.80	35.35	53.69	44.75	31.59	36.41	60.24	49.43	63.53	44.75	62.56	63.10	65.09	37.29	53.39	69.09	**83.62**

TABLE III
COMPARISON OF ACCURACIES OF THE PROPOSED METHODS SGA-MDAP AND SGA-MDA WITH OTHER APPROACHES FOR DIFFERENT TASKS OF CALTECH DATASET.

Tasks	**NN**	**PCA**	**SVM**	**TJM**	**GFK**	**GFK-PLS**	**JDA**	**TCA**	**SA**	**CORAL**	**ILS**	**SGA-MDAP**	**SGA-MDA**
A→C	70.1	76.49	74.2	82.45	77.73	77.7	82.01	80.14	77.1	79.0	78.9	80.40	**82.45**
A→D	52.3	59.87	51.7	72.61	59.23	63.5	70.06	65.60	64.9	67.1	72.5	71.97	**82.80**
A→W	69.9	69.15	63.1	82.71	73.89	74.1	83.72	76.94	76.0	74.8	82.4	81.35	**83.05**
C→A	81.9	86.43	86.7	85.80	86.01	86.2	88.10	86.63	83.9	89.4	87.6	88.83	**88.83**
C→D	55.6	61.14	61.5	75.79	62.42	66.5	72.61	69.42	66.2	67.6	73.0	**78.98**	**78.98**
C→W	65.9	74.23	74.8	77.96	74.91	76.5	80.67	74.91	76.0	77.6	84.4	82.03	**84.74**
D→A	57.0	67.43	58.7	80.79	68.58	69.9	77.13	75.15	69.0	75.6	79.2	80.48	**83.71**
D→C	48.0	58.50	55.5	74.44	59.57	64.0	70.52	69.18	62.3	64.7	66.5	75.33	**79.51**
D→W	86.7	95.59	91.8	96.94	95.93	92.4	97.62	96.61	90.5	94.6	94.2	**99.32**	**99.32**
W→A	62.4	75.15	69.8	82.25	79.01	77.9	84.2	80.27	76.6	81.2	85.9	86.11	**88.30**
W→C	57.5	69.01	64.7	78.45	70.16	71.3	74.79	75.24	70.7	75.2	77.0	80.32	**82.27**
W→D	83.9	94.90	89.4	94.90	94.90	92.6	96.81	93.63	90.4	92.6	87.4	**100**	**100**
Average	65.93	73.9	70.15	82.09	75.19	76.05	81.52	78.64	75.3	78.2	80.7	83.76	**86.16**

TABLE IV
COMPARISON OF ACCURACIES OF THE PROPOSED METHODS SGA-MDAP AND SGA-MDA WITH OTHER APPROACHES FOR DIFFERENT TASKS OF ORL AND YALE DATASETS

Task	NN	PCA	GFK	JDA	TCA	TJM	JGSA	**SGA-MDAP**	**SGA-MDA**
Yale 1 2→9	51.11	53.33	54.07	53.33	40.74	42.22	25.15	60.74	**75.76**
Yale 2 3→8	57.50	60.00	60.00	45.83	46.67	47.50	49.15	57.50	**70.83**
Yale 3 4→7	61.90	64.76	64.76	41.88	53.33	52.38	58.10	69.52	**74.29**
Average	56.83	59.36	59.61	47.01	46.91	47.36	44.13	62.58	**73.62**
ORL 1 1→9	58.89	59.17	59.54	41.94	41.64	42.50	64.72	74.44	**97.50**
ORL 2 2→8	72.19	71.56	71.54	50.31	50.63	50.94	73.12	82.81	**96.25**
ORL 3 3→7	76.43	75.36	75.71	51.43	51.79	51.78	82.14	86.43	**96.07**
ORL 4 4→6	85.83	84.17	85.00	66.25	67.08	67.08	87.50	94.17	**97.50**
Average	73.33	72.56	72.96	52.48	52.78	53.07	76.87	84.86	**96.83**

For the ORL dataset, our proposed SGA-MDA achieves 96.83% accuracy and SGA-MDAP achieves 84.86% accuracy, each outperforming current methods.

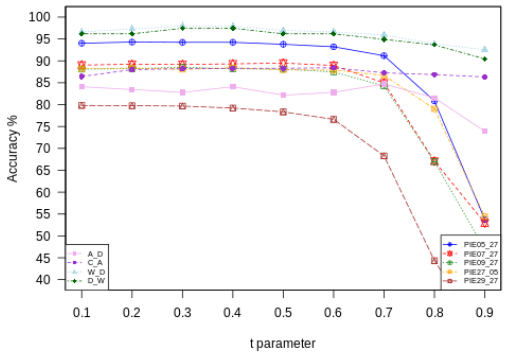
TABLE V
SUGGESTED VALUES OR RANGES FOR PARAMETERS.

Dataset	K	D	t	δ	γ
PIE	1	$[100, 200]$	$[0.1, 0.5]$	$[0.1, 1]$	10^{-2}
Caltech	1	40	$[0.1, 0.5]$	$[0.1, 1]$	10^{-2}
ORL	1	$[20, 100]$	$[0.2, 0.5]$	$[0, 1]$	$[10^{-2}, 10^{0}]$
Yale	1	$[20, 100]$	$[0.2, 0.5]$	$[0, 1]$	$[10^{-2}, 10^{0}]$

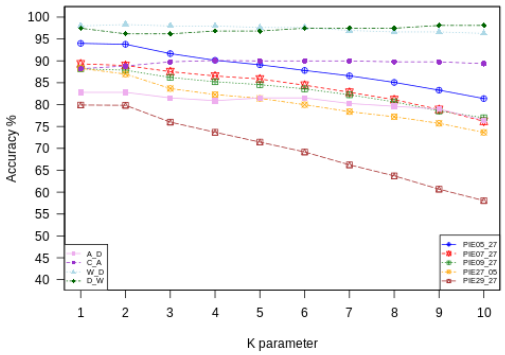
V. CONCLUSION

In this paper, we proposed a novel framework for the visual domain classification problem in machine learning, based on overcoming the limitations of existing methods. Our proposed methods SGA-MDAP and SGA-MDA, were tested on four well-known data sets. These data sets include Office-Caltech with Deep-Net Features (VGG-Net), PIE Face Recognition , ORL Face, and Yale Face. The experiments conducted proved that SGA-MDAP and SGA-MDA achieved higher accuracy compared to various other state-of-the-art learning methods.

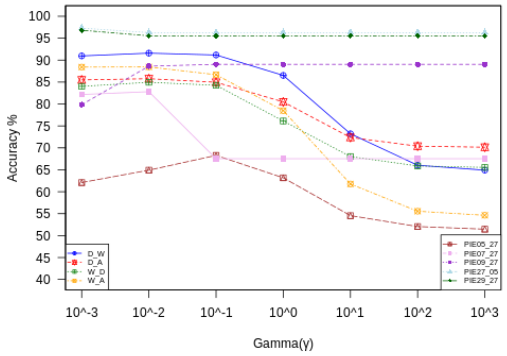
In future, we will extend our model to the multi-metric



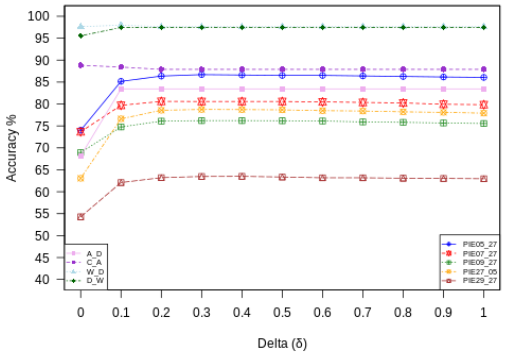
(a) t parameter variation plot



(b) K parameter variation plot

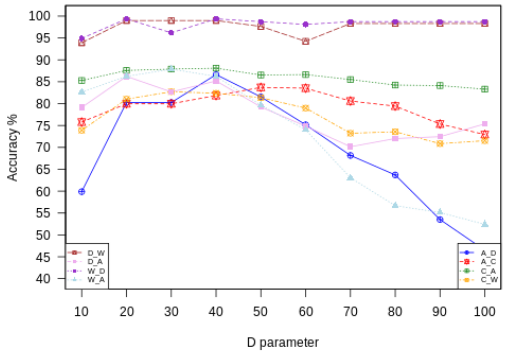


(c) γ parameter variation plot

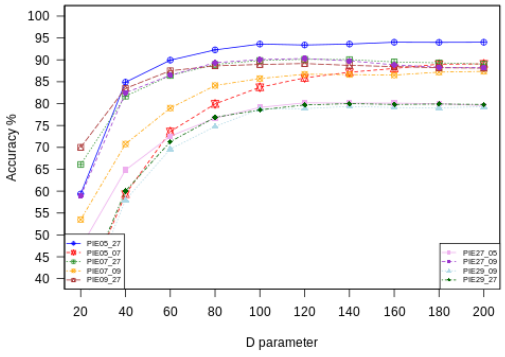


(d) δ parameter variation plot

Fig. 3. Impact of t, K,γ, and δ parameters on the PIE face and Office-Caltech data sets.



(a) D parameter variation plot for Caltech dataset



(b) D parameter variation plot for PIE face data set

Fig. 4. Impact of D parameter on the PIE face and Office-Caltech data sets.

learning environment, where the generalized performance of all the metric can be enhanced at the same time.

REFERENCES

[1] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–7.

[2] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[3] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1859–1867.

[4] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *2011 international conference on computer vision*. IEEE, 2011, pp. 999–1006.

[5] R. K. Sanodiya, J. Mathew, S. Saha, and M. D. Thalakottur, "A new transfer learning algorithm in semi-supervised setting," *IEEE Access*, 2019.

[6] R. K. Sanodiya and J. Mathew, "A framework for semi-supervised metric transfer learning on manifolds," *Knowledge-Based Systems*, 2019.

[7] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2200–2207.

[8] Y. Xu, S. J. Pan, H. Xiong, Q. Wu, R. Luo, H. Min, and H. Song, "A unified framework for metric transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 6, pp. 1158–1171, 2017.

[9] S. Mahadevan, B. Mishra, and S. Ghosh, "A unified framework for domain adaptation using metric learning on manifolds," *arXiv preprint arXiv:1804.10834*, 2018.

[10] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 7, p. 929, 2010.

[11] Z. Ding, M. Shao, and Y. Fu, "Incomplete multisource transfer learning," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 2, pp. 310–323, 2018.

[12] L. Duan, D. Xu, and I. W.-H. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 504–518, 2012.

[13] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2066–2073.

[14] I.-H. Jhuo, D. Liu, D. Lee, and S.-F. Chang, "Robust visual domain adaptation with low-rank reconstruction," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2168–2175.

[15] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," *arXiv preprint arXiv:1206.4683*, 2012.

[16] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2960–2967.

[17] H. Wang, W. Wang, C. Zhang, and F. Xu, "Cross-domain metric learning based on information theory," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[18] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1410–1417.

[19] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *International Journal of Computer Vision*, vol. 109, no. 1-2, pp. 74–93, 2014.

[20] S. Mehrkanoon and J. A. Suykens, "Regularized semi-paired kernel cca for domain adaptation," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 7, pp. 3199–3213, 2018.

[21] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 850–863, 2016.

[22] Z. Ding and Y. Fu, "Robust transfer metric learning for image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 660–670, 2017.

[23] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[24] L. Luo, L. Chen, S. Hu, Y. Lu, and X. Wang, "Discriminative and geometry aware unsupervised domain adaptation," *arXiv preprint arXiv:1712.10042*, 2017.

[25] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The annals of statistics*, pp. 1171–1220, 2008.

[26] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. Mar, pp. 723–773, 2012.

[27] P. Zadeh, R. Hosseini, and S. Sra, "Geometric mean metric learning," in *International conference on machine learning*, 2016, pp. 2464–2471.

[28] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in neural information processing systems*, 2007, pp. 513–520.

[29] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction." in *AAAI*, vol. 8, 2008, pp. 677–682.

[30] S. Wang, J. Lu, X. Gu, H. Du, and J. Yang, "Semi-supervised linear discriminant analysis for dimension reduction and classification," *Pattern Recognition*, vol. 57, pp. 179–189, 2016.

[31] A. Georghiades, P. Belhumeur, and D. Kriegman, "Yale face database," *Center for computational Vision and Control at Yale University, http://cvc. yale. edu/projects/yalefaces/yalefa*, vol. 2, p. 6, 1997.

[32] S. Herath, M. Harandi, and F. Porikli, "Learning an invariant hilbert space for domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3845–3854.