Electronic Research Archive

http://www.aimspress.com/journal/era

ERA, 31(9): 5425–5441.
DOI: 10.3934/era.2023275
Received: 11 April 2023
Revised: 07 July 2023
Accepted: 23 July 2023
Published: 31 July 2023

Research article

Cross-view learning with scatters and manifold exploitation in geodesic space

Qing Tian[1,2,3,*], **Heng Zhang**[1,2,†], **Shiyu Xia**[4,†], **Heng Xu**[1,2,†], **and Chuang Ma**[1,2]

[1] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China.

[2] Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China.

[3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China.

[4] School of Computer Science and Engineering, Southeast University, Nanjing 210096, China.

[†] These authors contributed equally to this work.

[*] **Correspondence:** Email: tianqing@nuist.edu.cn.

Abstract: Cross-view data correlation analysis is a typical learning paradigm in machine learning and pattern recognition. To associate data from different views, many approaches to correlation learning have been proposed, among which canonical correlation analysis (CCA) is a representative. When data is associated with label information, CCA can be extended to a supervised version by embedding the supervision information. Although most variants of CCA have achieved good performance, nearly all of their objective functions are nonconvex, implying that their optimal solutions are difficult to obtain. More seriously, the discriminative scatters and manifold structures are not exploited simultaneously. To overcome these shortcomings, in this paper we construct a Discriminative Correlation Learning with Manifold Preservation, DCLMP for short, in which, in addition to the within-view supervision information, discriminative knowledge as well as spatial structural information are exploited to benefit subsequent decision making. To pursue a closed-form solution, we remodel the objective of DCLMP from the Euclidean space to a geodesic space and obtain a convex formulation of DCLMP (C-DCLMP). Finally, we have comprehensively evaluated the proposed methods and demonstrated their superiority on both toy and real datasets.

Keywords: canonical correlation analysis; cross-view data correlation analysis; convex discriminative correlation learning; cross-view representation

1. Introduction

Correlation analysis deals with data with cross-view feature representations. To handle such tasks, many correlation learning approaches have been proposed, among which canonical correlation analysis (CCA) [1–5] is a representative method and has been widely employed [6–11] . To be specific, given training data with two or more feature-view representations, the traditional CCA method comes to seek a projection vector for each of the views while maximizing the cross-view correlations. After the data are mapped along the projection directions, subsequent cross-view decisions can be made [4]. Although CCA yields good results, a performance room is left since the data labels are not incorporated in learning.

When class labels information is also provided or available, CCA can be remodeled to its discriminant form by making use of the labels. To this end, Sun et al. [12] proposed a discriminative variant of CCA (*i.e.*, DCCA) by enlarging distances between dissimilar samples while reducing those of similar samples. Subsequently, Peng et al. [13] built a locally-discriminative version of CCA (*i.e.*, LDCCA) based on the assumption that the data distributions follow low-dimensional manifold embedding. Besides, Su et al. [14] established a multi-patch embedding CCA (MPECCA) by developing multiple metrics rather than a single one to model within-class scatters. Afterwards, Sun et al. [15] built a generalized framework for CCA (GCCA). Ji et al. [16] remodeled the scatter matrices by deconstructing them into several fractional-order components and achieved performance improvements.

In addition to directly constructing a label-exploited version of CCA, the supervised labels can be utilized by embedding them as regularization terms. Along this direction, Zhou et al. [17] presented CECCA by embedding LDA-guided [18] feature combinations into the objective function of CCA. Furthermore, Zhao et al. [19] constructed HSL-CCA by reducing inter-class scatters within their local neighborhoods. Later, Haghighat et al. [20] proposed the DCA model by deconstructing the inter-class scatter matrix guided by class labels. Previous variants of CCA were designed to cater for two-view data and cannot be used directly to handle multi-view scenarios. To overcome this shortcoming, many CCA methods have been proposed, such as GCA [21] , MULDA [22] and FMDA [23] .

Although the aforementioned methods have achieved successful performances of varying extent, unfortunately, the objective functions of nearly all of them are not convex [14, 24, 25]. Although CDCA [26] yields closed-form solutions and better results than the previous methods.

To overcome these shortcomings, we firstly design a discriminative correlation learning with manifold preservation, coined as DCLMP, in which, not only the cross-view discriminative information but also the spatial structural information of training data is taken into account to enhance subsequent decision making. To pursue closed-form solutions, we remodel the objective of DCLMP from the Euclidean space to a geodesic space. In this way, we obtain a convex formulation of DCLMP (C-DCLMP). Finally, we comprehensively evaluated the proposed methods and demonstrated their superiority on both toy and real data sets. To summarize, our contributions are three-fold as follows:

1. A DCLMP is constructed by modelling both cross-view discriminative information and spatial structural information of training data.
2. The objective function of DCLMP is remodelled to obtain its convex formulation (C-DCLMP).
3. The proposed methods are evaluated with extensive experimental comparisons.

This paper is organized as follows. Section 2 reviews related theories of CCA. Section 3 presents

models and their solving algorithms. Then, experiments and comparisons are reported to evaluate the methods in Section 4. Section 5 concludes and provides future directions.

2. Related work

2.1. Multi-view learning

In this section, we briefly review the works on multi-view learning, which aims to study how to establish constraints or dependencies between views by modeling and discovering the interrelations between views. There exist studies about multi-view learning. Tang et al. [27] proposed a multi-view feature selection method named CvLP-DCL, which divided the label space into a consensus part and a domain-specific part and explored the latent information between different views in the label space. Additionally, CvLP-DCL explored how to combine cross-domain similarity graph learning with matrix-induced regularization to boost the performance of the model. Tang et al. [28] also proposed UoMvSc for multi-view learning, which mined the value of view-specific graphs and embedding matrices by combining spectral clustering with k-means clustering. In addition, Wang et al. [29] proposed an effective framework for multi-view learning named E^2OMVC, which constructed the latent feature representation based on anchor graphs and the clustering indicator matrix about multi-view data to obtain better clustering results.

2.2. Canonical correlation analysis

We briefly review related theories of CCA [1, 2]. Given two-view feature representations of training data, CCA seeks two projection matrices respectively for the two views, while preserving the cross-view correlations. To be specific, let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N] \in \mathbb{R}^{p\times N}$ and $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_N] \in \mathbb{R}^{q\times N}$ be two view representations of N training samples, with $\mathbf{x}_i$ and $\mathbf{y}_i$ denoting normalized representations of the ith sample. Besides, let $\mathbf{W}_x \in \mathbb{R}^{p\times r}$ and $\mathbf{W}_y \in \mathbb{R}^{q\times r}$ denote the projection matrices mapping the training data from individual view spaces into a r-dimensional common space. Then, the correlation between $\mathbf{W}_x^T\mathbf{x}_i$ and $\mathbf{W}_y^T\mathbf{y}_i$ should be maximized. Consequently, the formal objective of CCA can be formulated as

$$\max_{\{\mathbf{W}_x,\mathbf{W}_y\}} \frac{\mathbf{W}_x^T\mathbf{C}_{xy}\mathbf{W}_y}{\sqrt{\mathbf{W}_x^T\mathbf{C}_{xx}\mathbf{W}_x\mathbf{W}_y^T\mathbf{C}_{yy}\mathbf{W}_y}}, \tag{2.1}$$

where $\mathbf{C}_{xx} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$, $\mathbf{C}_{yy} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^T$, and $\mathbf{C}_{xy} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{y}_i - \overline{\mathbf{y}})^T$, where $\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$ and $\overline{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i$ respectively denote the sample means of the two views. The numerator describes the sample correlation in the projected space, while the denominator limits the scatter for each view. Typically, Eq (2.1) is converted to a generalized eigenvalue problem as

$$\begin{pmatrix} & \mathbf{X}\mathbf{Y}^T \\ \mathbf{Y}\mathbf{X}^T & \end{pmatrix}\begin{pmatrix}\mathbf{W}_x \\ \mathbf{W}_y\end{pmatrix} = \lambda\begin{pmatrix}\mathbf{X}\mathbf{X}^T & \\ & \mathbf{Y}\mathbf{Y}^T\end{pmatrix}\begin{pmatrix}\mathbf{W}_x \\ \mathbf{W}_y\end{pmatrix}. \tag{2.2}$$

Then, $\begin{pmatrix}\mathbf{W}_x \\ \mathbf{W}_y\end{pmatrix}$ can be achieved by computing the largest r eigenvectors of

$$\begin{pmatrix}\mathbf{X}\mathbf{X}^T & \\ & \mathbf{Y}\mathbf{Y}^T\end{pmatrix}^{-1}\begin{pmatrix} & \mathbf{X}\mathbf{Y}^T \\ \mathbf{Y}\mathbf{X}^T & \end{pmatrix}.$$

After $\mathbf{W}_x$ and $\mathbf{W}_y$ are obtained, $\mathbf{x}_i$ and $\mathbf{y}_i$ can be concatenated as $\mathbf{W}_x^T\mathbf{x}_i + \mathbf{W}_y^T\mathbf{y}_i = \begin{pmatrix}\mathbf{W}_x \\ \mathbf{W}_y\end{pmatrix}^T \begin{pmatrix}\mathbf{x}_i \\ \mathbf{y}_i\end{pmatrix}$. With the concatenated feature representations are achieved, subsequent classification or regression decisions can be made.

2.3. Variants of CCA

The most classic work of discriminative CCA is DCCA [12] , which is shown as follows:

$$\begin{aligned} &\max_{w_x, w_y} \left(\boldsymbol{w}_x^T \boldsymbol{C}_w \boldsymbol{w}_y - \eta \cdot \boldsymbol{w}_x^T \boldsymbol{C}_b \boldsymbol{w}_y\right) \\ &\text{s.t. } \boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}_x = 1, \boldsymbol{w}_y^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}_y = 1 \end{aligned} \tag{2.3}$$

It is easy to find that DCCA is discriminative because DCCA needs instance labels to calculate the relationship between each class. Similar to DCCA, Peng et al. [13] proposed LDCCA which is shown as follows:

$$\begin{aligned} &\max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\boldsymbol{w}_x^T \boldsymbol{C}_{xy} \boldsymbol{w}_y}{\sqrt{\left(\boldsymbol{w}_x^T \tilde{\boldsymbol{C}}_{xx} \boldsymbol{w}_x\right)\left(\boldsymbol{w}_y^T \boldsymbol{C}_{yy} \boldsymbol{w}_y\right)}} \\ &\text{s.t. } \boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}_x = 1, \boldsymbol{w}_y^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}_y = 1 \end{aligned} \tag{2.4}$$

where $\tilde{\boldsymbol{C}}_{xy} = \boldsymbol{C}_w - \eta \boldsymbol{C}_b \cdot \boldsymbol{C}_w$. Compared with DCCA, LDCCA consider the local correlations of the within-class sets and the between-class sets. However, these methods do not consider the problem of multimodal recognition or feature level fusion. Haghighat et al. [20] proposed DCA which incorporates the class structure, i.e., memberships of the samples in classes, into the correlation analysis. Additionally, Su et al. [14] proposed MPECCA for multi-view feature learning, which is shown as follows:

$$\begin{aligned} &\max_{u,v,w_j^{(x)},w_r^{(y)}} u^T \left(\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{r=1}^{M}\left(w_j^{(x)} w_r^{(y)}\right) X S_{ij}^{(x)} L_i S_{ir}^{(y)T} Y^T\right) v \\ &\quad \text{s.t. } u^T S_{wx} u = 1, v^T S_{wy} v = 1 \\ &\quad \sum_{j=1}^{M} w_j^{(x)} = 1, w_j^{(x)} \geqslant 0 \\ &\quad \sum_{r=1}^{M} w_r^{(y)} = 1, w_r^{(y)} \geqslant 0 \end{aligned} \tag{2.5}$$

where u and v means correlation projection matrices. Considering combining LDA and CCA, CECCA was proposed [17] . The optimization objective of CECCA was shown as follows:

$$\begin{aligned} &\max_{w_x, w_y} \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{X}(\boldsymbol{I} + 2\boldsymbol{A})\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y + \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{A}\boldsymbol{T}^{\mathrm{T}} \boldsymbol{w}_x + \boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_y \\ &\text{s. t. } \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x + \boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y = 2 \end{aligned} \tag{2.6}$$

where $\boldsymbol{A} = 2\boldsymbol{U} - \boldsymbol{I}$, $\boldsymbol{I}$ means Identity matrix. On the basis of CCA, CECCA combined with discriminant analysis to realize the joint optimization of correlation and discriminant of combined features, which makes the extracted features more suitable for classification. However, these methods cannot achieve

the closed form solution. CDCA [26] combined GMML and discrimative CCA and then achieve the closed form solution in Riemannian manifold space, the optimization objective was shown as follows:

$$\begin{aligned}\min_{\mathbf{A}>\mathbf{0}} \gamma \operatorname{tr}(\boldsymbol{AC}) + (1-\gamma)\left(\operatorname{tr}(\boldsymbol{AS}_Z) + \operatorname{tr}\left(\boldsymbol{A}^{-1}\boldsymbol{D}_Z\right)\right) = \\ \operatorname{tr}\left(\boldsymbol{A}\left(\gamma\boldsymbol{C} + (1-\gamma)\boldsymbol{S}_Z\right)\right) + \operatorname{tr}\left(\boldsymbol{A}^{-1}(1-\gamma)\boldsymbol{D}_Z\right)\end{aligned} \tag{2.7}$$

From Eq (2.7) and CDCA [26] we can find with the help of discrimative part and closed form solution, the multi-view learning will easily get the the global optimality of solutions and achieve a good result.

3. Proposed method

CCA suffer from three main problems: (1) the similarity and dissimilarity across views are not modeled; (2) although the data labels can be exploited by imposing supervised constraints, their objective functions are nonconvex; (3) the cross-view correlations are modeled in Euclidean space through RKHS kernel transformation [30, 31] whose discriminating ability is obviously limited.

We present a novel cross-view learning model, called DCLMP, in which not only the with-class and between-class scatters are characterized, but also the similarity and dissimilarity of the training data across views are modelled for utilization. Although many preferable characteristics are incorporated in DCLMP, it still suffers from non-convexity for its objective function. To facilitate pursuing global optimal solutions, we further remodel DCLMP to the Riemannian manifold space to make the objective function convex. The proposed method is named as C-DCLMP.

Assume we are given N training instances sampled from K classes with two views of feature representations, *i.e.*, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K] \in \mathbb{R}^{p\times N}$ with $\mathbf{X}_k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \cdots, \mathbf{x}_{N_k}^k]$ being N_k x-view instances from the k-th class and $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_K] \in \mathbb{R}^{q\times N}$ with $\mathbf{Y}_k = [\mathbf{y}_1^k, \mathbf{y}_2^k, \cdots, \mathbf{y}_{N_k}^k]$ being N_k y-view instances from the k-th class, where $\mathbf{y}_1^k$ and $\mathbf{x}_1^k$ stand for two view representations from the same instance. In order to concatenate them for subsequent classification, we denote $U \in \mathbb{R}^{p\times r}$ and $V \in \mathbb{R}^{q\times r}$ as projection matrices for the two views to transform their representations to a r-dimensional common space.

3.1. Discriminative correlation learning with manifold preservation (DCLMP)

To perform cross-view learning while exploring supervision knowledge in terms of similar and dissimilar relationships among instances in each view and across the views, as well as sample distribution manifolds, we construct DCLMP. To this end, we should construct the model by taking into account the following aspects: 1) distances between similar instances from the same class should be reduced while those among dissimilar from different classes should be enlarged, in levels of intra-view and inter-view; 2) manifold structures embedded in similar and dissimilar instances should be preserved. These modelling considerations are intuitively demonstrated in Figure 1.

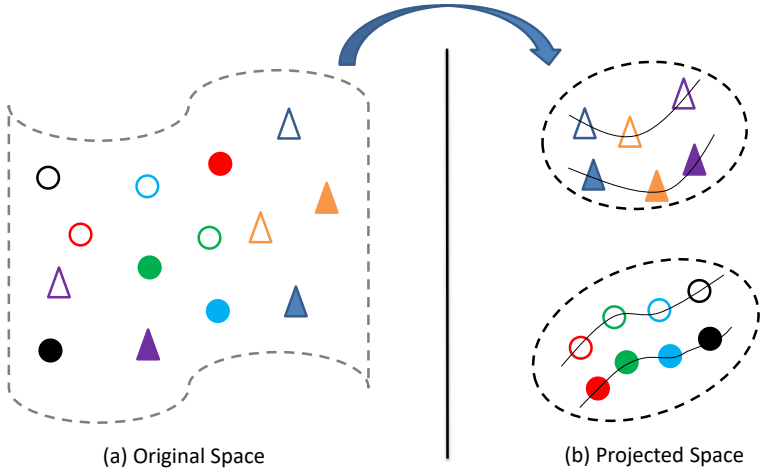

Figure 1. Modelling strategy of DCLMP. Here, circular and triangular shapes represent samples from two different classes, while filling in different colors represents different view representations. The samples are distributed dispersedly in original feature representation space (a); however, in DCLMP projection space (b), similar samples are pushed nearer while dissimilar from different classes are pulled apart from each other, while their manifold relations are preserved.

Along this line, we construct the objective function of DCLMP as follows:

$$\begin{aligned}
\min_{\{\mathbf{U},\mathbf{V}\}} & \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}\|\mathbf{U}^T\mathbf{x}_i-\mathbf{V}^T\mathbf{y}_j\|_2^2\cdot\mathbf{L}_{ij} \\
&+\frac{\lambda_1}{K}\sum_{k=1}^{K}\frac{1}{N_k}\sum_{i=1}^{N_k}\frac{1}{k_n}\sum_{j=1}^{k_n}\left\{\|\mathbf{U}^T\mathbf{x}_i^k-\mathbf{U}^T\mathbf{x}_j^k\|_F^2\cdot\mathbf{S}_{ij}^{w_x}\right. \\
&\left.+\|\mathbf{V}^T\mathbf{y}_i^k-\mathbf{V}^T\mathbf{y}_j^k\|_F^2\cdot\mathbf{S}_{ij}^{w_y}\right\} \\
&-\frac{\lambda_2}{K}\sum_{k=1}^{K}\sum_{h\neq k}\frac{1}{N_k}\sum_{i=1}^{N_k}\frac{1}{k_n}\sum_{j=1}^{k_n}\left\{\|\mathbf{U}^T\mathbf{x}_i^k-\mathbf{U}^T\mathbf{x}_j^h\|_F^2\cdot\mathbf{S}_{ij}^{b_x}\right. \\
&\left.+\|\mathbf{V}^T\mathbf{y}_i^k-\mathbf{V}^T\mathbf{y}_j^h\|_F^2\cdot\mathbf{S}_{ij}^{b_y}\right\}
\end{aligned} \tag{3.1}$$

where $\mathbf{U}$ and $\mathbf{V}$ denote the projection matrices in the r-dimensional common space of two views and k_n denotes the k-nearest neighbors of an instance. $\mathbf{L}$ is the discriminative weighting matrix. $\mathbf{S}^{w_x}$ and $\mathbf{S}^{w_y}$ stand for the within-class manifold weighting matrices of two different views of feature representations, and $\mathbf{S}^{b_x}$ and $\mathbf{S}^{b_y}$ stand for the between-class manifold weighting matrices of two different views of feature representations. Their elements are defined as follows

$$\mathbf{L}_{ij}=\begin{cases}\frac{1}{N_k} & \mathbf{x}_i \text{ and } \mathbf{y}_j \text{ are from the same class}\\ 0 & \mathbf{x}_i \text{ and } \mathbf{y}_j \text{ are from different classes}\end{cases} \tag{3.2}$$

$$\mathbf{S}_{ij}^{w_x} = \begin{cases} exp\left(-\frac{\|\mathbf{x}_i^k - \mathbf{x}_j^k\|^2}{\sigma_x^2}\right) & \mathbf{x}_j^k \in \mathrm{KNN}_{k_n}(\mathbf{x}_i^k) \\ 0 & \mathbf{x}_j^k \notin \mathrm{KNN}_{k_n}(\mathbf{x}_i^k) \end{cases} \tag{3.3}$$

$$\mathbf{S}_{ij}^{w_y} = \begin{cases} exp\left(-\frac{\|\mathbf{y}_i^k - \mathbf{y}_j^k\|^2}{\sigma_y^2}\right) & \mathbf{y}_j^k \in \mathrm{KNN}_{k_n}(\mathbf{y}_i^k) \\ 0 & \mathbf{y}_j^k \notin \mathrm{KNN}_{k_n}(\mathbf{y}_i^k) \end{cases} \tag{3.4}$$

$$\mathbf{S}_{ij}^{b_x} = \begin{cases} exp\left(-\frac{\|\mathbf{x}_i^k - \mathbf{x}_j^h\|^2}{\sigma_x^2}\right) & \mathbf{x}_j^h \in \mathrm{KNN}_{k_n}(\mathbf{x}_i^k) \\ 0 & \mathbf{x}_j^h \notin \mathrm{KNN}_{k_n}(\mathbf{x}_i^k) \end{cases} \tag{3.5}$$

$$\mathbf{S}_{ij}^{b_y} = \begin{cases} exp\left(-\frac{\|\mathbf{y}_i^k - \mathbf{y}_j^h\|^2}{\sigma_y^2}\right) & \mathbf{y}_j^h \in \mathrm{KNN}_{k_n}(\mathbf{y}_i^k) \\ 0 & \mathbf{y}_j^h \notin \mathrm{KNN}_{k_n}(\mathbf{y}_i^k) \end{cases} \tag{3.6}$$

where KNN_{k_n} denotes the k_n-nearest neighbors of an instance. σ_x and σ_y stand for width coefficients to normalize the weights.

In Eq (3.1), the first part characterizes the cross-view similarity and dissimilarity discriminations, the second part preserves the manifold relationships within each class scatters, while the third part magnifies the distribution margins for a dissimilar pair of instances. In this way, both the discriminative information and the manifold distributions can be modelled in a joint objective function.

For convenience of solving Eq (3.1), we transform it as the following concise form

$$\min_{\mathbf{A}\succ 0} \; tr(\mathbf{A}(\mathbf{C} + \lambda_1 \mathbf{S}_z - \lambda_2 \mathbf{D}_z)) \tag{3.7}$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}^T \tag{3.8}$$

$$\begin{aligned} \mathbf{C} = &\begin{bmatrix} \mathbf{1}_{p\times p} \\ \mathbf{0}_{q\times p} \end{bmatrix} \mathbf{X}\mathbf{M}^L\mathbf{X}^T[\mathbf{0}_{p\times q}, \mathbf{1}_{q\times q}] + \begin{bmatrix} \mathbf{0}_{p\times q} \\ \mathbf{1}_{q\times q} \end{bmatrix} \mathbf{Y}\mathbf{M}^L\mathbf{Y}^T[\mathbf{0}_{q\times p}, \mathbf{1}_{q\times q}] - \\ &\begin{bmatrix} \mathbf{1}_{p\times p} \\ \mathbf{0}_{q\times p} \end{bmatrix} \mathbf{X}\mathbf{L}\mathbf{Y}^T[\mathbf{0}_{q\times p}, \mathbf{1}_{q\times q}] - \begin{bmatrix} \mathbf{0}_{p\times q} \\ \mathbf{1}_{q\times q} \end{bmatrix} \mathbf{Y}\mathbf{L}\mathbf{X}^T[\mathbf{1}_{p\times p}, \mathbf{0}_{p\times q}] \end{aligned} \tag{3.9}$$

$$\begin{aligned} \mathbf{S}_z = \sum_{k=1}^{K} \Bigg\{ &\begin{bmatrix} \mathbf{1}_{p\times p} \\ \mathbf{0}_{q\times p} \end{bmatrix} \mathbf{X}^k \left(\mathbf{M}^{w_x} + \mathbf{M}^{w_x T} - \mathbf{S}^{w_x} - \mathbf{S}^{w_x T}\right) \mathbf{X}^{kT}[\mathbf{1}_{p\times p}, \mathbf{0}_{p\times q}] \\ &+ \begin{bmatrix} \mathbf{0}_{p\times q} \\ \mathbf{1}_{q\times q} \end{bmatrix} \mathbf{Y}^k \left(\mathbf{M}^{w_y} + \mathbf{M}^{w_y T} - \mathbf{S}^{w_y} - \mathbf{S}^{w_y T}\right) \mathbf{Y}^{kT}[\mathbf{0}_{q\times p}, \mathbf{1}_{q\times q}] \Bigg\} \end{aligned} \tag{3.10}$$

$$\begin{aligned} \mathbf{D}_z = \Bigg\{ &\begin{bmatrix} \mathbf{1}_{p\times p} \\ \mathbf{0}_{q\times p} \end{bmatrix} \mathbf{X} \left(\mathbf{M}^{b_x} + \mathbf{M}^{b_x T} - \mathbf{S}^{b_x} - \mathbf{S}^{b_x T}\right) \mathbf{X}^T[\mathbf{1}_{p\times p}, \mathbf{0}_{p\times q}] \\ &+ \begin{bmatrix} \mathbf{0}_{p\times q} \\ \mathbf{1}_{q\times q} \end{bmatrix} \mathbf{Y} \left(\mathbf{M}^{b_y} + \mathbf{M}^{b_y T} - \mathbf{S}^{b_y} - \mathbf{S}^{b_y T}\right) \mathbf{Y}^T[\mathbf{0}_{q\times p}, \mathbf{1}_{q\times q}] \Bigg\} \end{aligned} \tag{3.11}$$

where $\mathbf{M}_{ii}^{w_x} = \sum_{j=1}^{N} \mathbf{S}_{ij}^{w_x}$, $\mathbf{M}_{ii}^{w_y} = \sum_{j=1}^{N} \mathbf{S}_{ij}^{w_y}$, $\mathbf{M}_{ii}^{b_x} = \sum_{j=1}^{N} \mathbf{S}_{ij}^{b_x}$, $\mathbf{M}_{ii}^{b_y} = \sum_{j=1}^{N} \mathbf{S}_{ij}^{b_y}$, and $\mathbf{M}_{ii}^{L} = \sum_{j=1}^{N} \mathbf{L}_{ij}$.

We let $\mathcal{J}$ record the objective function value of Eq (3.7) and introduce $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ to replace $\mathbf{A}$ and Λ to rewrite Eq (3.7) as

$$\mathcal{J}_{\{\mathbf{Q},\Lambda\}} = tr(\mathbf{Q}^T(\mathbf{C} + \lambda_1 \mathbf{S}_z - \lambda_2 \mathbf{D}_z)\mathbf{Q}) - tr(\Lambda(\mathbf{Q}^T\mathbf{Q} - \mathbf{I})). \tag{3.12}$$

Calculating the partial derivative of $\mathcal{J}_{\{\mathbf{Q},\Lambda\}}$ with regard to $\mathbf{Q}$ and making it to zero yields

$$(\mathbf{C} + \lambda_1\mathbf{S}_z - \lambda_2\mathbf{D}_z)\mathbf{Q} = \mathbf{Q}\Lambda, \tag{3.13}$$

The projection matrix $\mathbf{Q}$ can be obtained by calculating a required number of smallest eigenvectors of $\mathbf{C} + \lambda_1\mathbf{S}_z - \lambda_2\mathbf{D}_z$. Finally, we can recover $\mathbf{A} = \mathbf{Q}\mathbf{Q}^T$. Then, $\mathbf{U}$ and $\mathbf{V}$ can be obtained through Eq (3.8).

3.2. C-DCLMP

We find that such a objective function may be not convex [32, 33] . The separability of nonlinear data patterns in the geodesic space can be significantly improved and thus benefits their subsequent recognitions. Referring to $\min_{\mathbf{A}>0}\ tr(\mathbf{A}^{-1}\bullet) \Leftrightarrow \max_{\mathbf{A}>0}\ tr(\mathbf{A}\bullet)$ [34] , we reformulate DCLMP in (3.7) equivalently as

$$\min_{\mathbf{A}>0}\ tr(\mathbf{AC} + \lambda_1\mathbf{AS}_z + \lambda_2\mathbf{A}^{-1}\mathbf{D}_z) \Leftrightarrow \min_{\mathbf{A}>0}\ tr(\mathbf{AC}) + \lambda_1 tr(\mathbf{AS}_z) + \lambda_2 tr(\mathbf{A}^{-1}\mathbf{D}_z), \tag{3.14}$$

Minimizing the third term $\lambda_1 tr(\mathbf{A}^{-1}\mathbf{D}_z)$ is equivalent to minimizing $-\lambda_1 tr(\mathbf{AD}_z)$ of Eq (3.7). Although the last term is nonlinear, it is defined in the convex cone space [35] and thus is still convex. As a result, Eq (3.14) is entirely *convex* regarding A. It enjoys closed-form solution [36–38] . To distinguish Eq (3.14) from DCLMP, we call it C-DCLMP.

For convenience of deriving the closed-form solution, we reformulate Eq (3.14) as

$$\min_{\mathbf{A}>0}\ \gamma tr(\mathbf{A}^{-1}\mathbf{D}_z) + (1-\gamma)\left(tr(\mathbf{AS}_z) + \alpha tr(\mathbf{AC})\right), \tag{3.15}$$

where we set $\gamma \in (0, 1)$ [34] . Let $J(\mathbf{A}) := \gamma tr(\mathbf{A}^{-1}\mathbf{D}_z) + (1-\gamma)\left(tr(\mathbf{AS}_z) + \alpha tr(\mathbf{AC})\right)$.

$$(1-\gamma)\mathbf{A}(\mathbf{S}_z + \alpha\mathbf{C})\mathbf{A} = \gamma\mathbf{D}_z, \tag{3.16}$$

whose solution is the midpoint of the geodesic jointing $((1-\gamma)(\mathbf{S}_z + \alpha\mathbf{C}))^{-1}$ and $\gamma\mathbf{D}_z$, that is

$$\mathbf{A} = \left((1-\gamma)(\mathbf{S}_z + \alpha\mathbf{C})\right)^{-1} \sharp_{1/2} (\gamma\mathbf{D}_z), \tag{3.17}$$

$(\cdot)\sharp_{1/2}(\cdot)$ denotes the midpoint. We extend the geodesic mean solution (3.17) to the geodesic space by replacing $(\cdot)\sharp_{1/2}(\cdot)$ with $(\cdot)\sharp_t(\cdot)$, $0 \leqslant t \leqslant 1$.

We add a regularizer with prior knowledge to (3.15). Here, we incorporate symmetrized LogDet divergence and consequently (3.15) becomes

$$\begin{aligned}\min_{\mathbf{A}>0}\ & \gamma tr(\mathbf{A}^{-1}\mathbf{D}_z) + (1-\gamma)(tr(\mathbf{AS}_z) + \alpha tr(\mathbf{AC})) \\ & + \lambda D_{sld}(\mathbf{A}, \mathbf{A}_0),\end{aligned} \tag{3.18}$$

$$D_{sld}(\mathbf{A}, \mathbf{A}_0) = tr(\mathbf{A}\mathbf{A}_0^{-1}) + tr(\mathbf{A}^{-1}\mathbf{A}_0) - 2(p+q), \tag{3.19}$$

where $(p+q)$ is the dimension of the data. Fortunately, complying with the definition of geometric mean [36], Eq (3.18) is still convex. We let $G(\mathbf{A}) := \gamma tr(\mathbf{A}^{-1}\mathbf{D}_z) + (1-\gamma)(tr(\mathbf{AS}_z) + \alpha tr(\mathbf{AC})) + \lambda D_{sld}(\mathbf{A}, \mathbf{A}_0)$. Then we set the gradient of $G(\mathbf{A})$ regarding to $\mathbf{A}$ to zero and obtain the equation as

$$(1-\gamma)\mathbf{A}(\mathbf{S}_z+\alpha\mathbf{C})\mathbf{A}+\lambda\mathbf{A}\mathbf{A}_0^{-1}\mathbf{A}=\gamma\mathbf{D}_z+\lambda\mathbf{A}_0, \tag{3.20}$$

we calculate the closed-form solution as

$$\mathbf{A}=((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1})^{-1}\sharp_t(\gamma\mathbf{D}_z+\lambda\mathbf{A}_0). \tag{3.21}$$

More precisely, according to the definition of $(\cdot)\sharp_t(\cdot)$, namely the geodesic mean jointing two matrices, we can directly expand the final solution of our C-DCLMP in Eq (3.18) as

$$\begin{aligned}\mathbf{A}&=((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1})^{-1}\sharp_t(\gamma\mathbf{D}_z+\lambda\mathbf{A}_0)\\&=\left((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1}\right)^{1/2}\\&\quad\left(\left((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1}\right)^{-1/2}(\gamma\mathbf{D}_z+\lambda\mathbf{A}_0)\right.\\&\quad\left.\left((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1}\right)^{-1/2}\right)^t\\&\quad\left((1-\gamma)(\mathbf{S}_z+\alpha\mathbf{C})+\lambda\mathbf{A}_0^{-1}\right)^{1/2}.\end{aligned} \tag{3.22}$$

where we set $\mathbf{A}_0$ to be a $(p+q)$-order identity matrix $\mathbf{I}_{p+q}$. When obtaining $\mathbf{A}$, $\mathbf{U}$ and $\mathbf{V}$ are recovered.

Its concatenated representation can be generated by $\mathbf{U}^T\mathbf{x}+\mathbf{V}^T\mathbf{y}=\begin{bmatrix}\mathbf{U}\\\mathbf{V}\end{bmatrix}^T\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}$ and the classification decision using a classifier (e.g., KNN) can be made on this fused representation.

4. Experiment

To comprehensively evaluate the proposed methods, we first performed comparative experiments on several benchmark and real face datasets. Besides, we also performed sensibility analysis on the model parameters.

4.1. Setup

For evaluation and comparisons, CCA [1], DCCA [12], MPECCA [14], CECCA [17], DCA [20] and CDCA [26] were implemented. All hyper-parameters were cross-validated in the range of [0, 0.1, ..., 1] for t and γ, and [1e-7, 1e-6, ..., 1e3] for α and λ. For concatenated cross-view representations, a 5-nearest-neighbors classifier was employed for classification. Additionally, recognition accuracy (%, higher is better) and mean absolute errors (MAE, lower is better) were adopted as performance measures.

4.2. Comparisons on non-face datasets

We first performed experiments on several widely used non-face multi-view datasets, i.e., MFD [39] and USPS [40], AWA [41] and ADNI [42]. We report the results in Table 1.

The proposed DCLMP method yielded the second-lowest estimation errors in most cases, slightly higher than the proposed C-DCLMP. The improvement achieved by C-DCLMP method is significant, especially on AWA and USPS datasets.

Table 1. Recognition accuracy (%) comparison on the non-face datasets.

Dataset	View Represenations		*CCA*	*DCA*	*MPECCA*	*DCCA*	*CECCA*	*CDCA*	*DCLMP (ours)*	*C-DCLMP (ours)*
MFD	fac	fou	80.22 ± 0.9	80.00 ± 0.2	90.64 ± 1.3	95.15 ± 0.9	96.46 ± 2.4	**98.11 ± 0.3**	94.49 ± 1.7	98.03 ± 0.3
	fac	kar	92.12 ± 0.5	90.10 ± 0.8	95.39 ± 0.6	95.33 ± 0.7	96.52 ± 1.2	97.06 ± 0.4	96.86 ± 0.5	**97.93 ± 0.6**
	fac	mor	78.22 ± 0.8	63.22 ± 4.3	72.32 ± 2.4	95.22 ± 0.9	94.23 ± 1.0	**98.13 ± 0.3**	90.97 ± 2.5	97.63 ± 0.3
	fac	pix	83.02 ± 1.2	90.20 ± 0.5	94.65 ± 0.5	65.60 ± 1.1	93.67 ± 2.9	**97.52 ± 0.4**	97.45 ± 0.5	97.21 ± 0.4
	fac	zer	84.00 ± 0.6	71.50 ± 2.2	93.79 ± 0.7	96.00 ± 0.6	97.04 ± 0.6	97.03 ± 0.4	95.98 ± 0.3	**97.75 ± 0.4**
	fou	kar	90.11 ± 1.0	75.42 ± 5.6	93.98 ± 0.4	89.12 ± 4.3	96.90 ± 0.5	97.19 ± 0.6	97.45 ± 0.4	**97.45 ± 0.3**
	fou	mor	70.22 ± 0.4	55.82 ± 4.6	60.62 ± 1.6	82.30 ± 0.9	78.25 ± 0.6	83.81 ± 0.7	82.09 ± 1.0	**84.80 ± 0.6**
	fou	pix	68.44 ± 0.4	76.10 ± 4.7	78.24 ± 1.1	90.41 ± 3.2	76.28 ± 1.3	96.11 ± 0.5	97.62 ± 0.4	**97.74 ± 0.3**
	fou	zer	74.10 ± 0.9	62.80 ± 4.1	79.38 ± 1.2	79.53 ± 4.5	83.16 ± 1.4	85.98 ± 0.9	85.33 ± 1.1	**86.56 ± 1.0**
	kar	mor	64.09 ± 0.6	82.00 ± 1.6	72.92 ± 2.7	91.95 ± 2.8	91.89 ± 0.6	**97.28 ± 0.5**	96.83 ± 0.5	97.14 ± 0.4
	kar	pix	88.37 ± 0.9	88.85 ± 0.8	95.07 ± 0.6	92.59 ± 2.0	95.98 ± 0.3	94.68 ± 0.5	**97.54 ± 0.4**	97.31 ± 0.5
	kar	zer	90.77 ± 1.0	75.97 ± 2.8	94.17 ± 0.6	88.47 ± 2.9	93.57 ± 0.9	96.69 ± 0.4	96.98 ± 0.4	**97.42 ± 0.4**
	mor	pix	68.66 ± 1.5	82.01 ± 2.1	67.21 ± 2.3	93.04 ± 0.7	90.08 ± 1.0	96.89 ± 0.4	**97.20 ± 0.5**	97.19 ± 0.4
	mor	zer	73.22 ± 0.6	50.35 ± 1.8	60.95 ± 1.4	84.55 ± 0.9	80.59 ± 0.9	84.19 ± 0.8	81.75 ± 1.1	**84.29 ± 0.7**
	pix	zer	82.46 ± 0.6	71.16 ± 2.8	82.81 ± 1.2	91.67 ± 2.1	91.81 ± 1.2	96.30 ± 0.5	**97.35 ± 0.5**	97.30 ± 0.5
AWA	cq	lss	73.11 ± 2.1	62.08 ± 0.3	76.19 ± 1.0	70.51 ± 1.3	77.53 ± 1.7	87.80 ± 2.8	89.03 ± 1.4	**89.80 ± 1.2**
	cq	phog	65.21 ± 1.4	73.10 ± 1.2	72.42 ± 1.6	70.15 ± 0.9	74.51 ± 2.1	85.58 ± 2.7	86.71 ± 2.3	**86.81 ± 1.2**
	cq	rgsift	60.22 ± 1.3	61.40 ± 1.7	78.04 ± 1.3	82.87 ± 2.4	82.83 ± 1.4	90.99 ± 3.0	93.44 ± 0.6	**94.34 ± 0.8**
	cq	sift	74.33 ± 1.3	61.28 ± 1.9	77.85 ± 1.4	83.19 ± 2.1	80.05 ± 1.7	81.59 ± 5.2	87.17 ± 0.8	**90.68 ± 0.8**
	cq	surf	75.86 ± 1.7	69.30 ± 2.1	79.07 ± 0.8	73.55 ± 2.3	81.59 ± 1.5	93.58 ± 1.1	94.36 ± 1.0	**95.35 ± 0.5**
	lss	phog	69.96 ± 1.7	59.72 ± 0.2	68.12 ± 1.2	64.86 ± 2.6	71.36 ± 1.4	80.48 ± 2.0	**81.76 ± 1.1**	81.62 ± 1.1
	lss	rgsift	78.65 ± 0.9	63.21 ± 1.3	73.64 ± 1.0	78.28 ± 2.8	77.28 ± 1.4	87.38 ± 4.3	**90.13 ± 0.7**	89.95 ± 1.0
	lss	sift	73.49 ± 1.0	65.72 ± 2.1	73.12 ± 1.4	66.21 ± 1.6	76.69 ± 1.7	81.56 ± 2.4	84.05 ± 0.9	**84.07 ± 1.9**
	lss	surf	76.30 ± 1.4	65.33 ± 1.8	74.84 ± 1.6	79.06 ± 2.8	78.52 ± 1.3	89.81 ± 2.5	89.75 ± 0.8	**91.12 ± 0.7**
	phog	rgsift	68.18 ± 1.1	48.38 ± 1.0	69.49 ± 2.3	77.37 ± 1.5	74.41 ± 1.5	82.76 ± 1.1	83.57 ± 1.6	**83.68 ± 1.2**
	phog	sift	68.26 ± 1.1	70.24 ± 1.1	68.97 ± 1.3	63.16 ± 1.3	72.14 ± 1.5	80.50 ± 1.2	83.57 ± 1.1	**83.75 ± 1.5**
	phog	surf	64.57 ± 1.4	56.94 ± 0.5	71.55 ± 1.4	75.68 ± 1.9	74.43 ± 2.1	84.97 ± 2.6	**88.02 ± 1.8**	87.34 ± 0.8
	rgsift	sift	71.35 ± 1.3	58.56 ± 2.3	72.85 ± 1.1	75.28 ± 2.5	76.69 ± 1.7	90.76 ± 2.2	93.44 ± 0.4	**93.79 ± 1.0**
	rgsift	surf	75.55 ± 1.3	67.22 ± 1.6	76.94 ± 2.2	84.10 ± 2.4	80.46 ± 1.7	93.25 ± 1.2	92.82 ± 0.8	**93.66 ± 0.8**
	sift	surf	75.33 ± 1.3	63.36 ± 1.6	74.27 ± 1.2	82.14 ± 2.7	75.51 ± 1.1	90.07 ± 3.4	90.67 ± 1.0	**91.69 ± 1.1**
ADNI	AV	FDG	65.47 ± 1.8	73.28 ± 2.1	75.28 ± 2.6	76.25 ± 2.1	76.26 ± 2.5	79.59 ± 1.9	68.64 ± 3.3	**80.86 ± 2.1**
	AV	VBM	71.02 ± 2.4	71.02 ± 2.8	73.24 ± 3.1	63.47 ± 2.1	60.67 ± 2.7	**81.59 ± 2.5**	78.38 ± 2.5	80.70 ± 2.8
	FDG	VBM	61.37 ± 1.2	65.28 ± 1.6	70.37 ± 2.6	64.05 ± 1.6	70.95 ± 1.8	80.12 ± 2.0	74.97 ± 2.9	**80.21 ± 1.7**
USPS	left	right	62.14 ± 0.6	80.11 ± 1.2	66.67 ± 0.9	63.96 ± 2.0	82.89 ± 1.9	89.76 ± 0.3	**96.19 ± 0.7**	96.03 ± 0.6

4.3. Comparisons on face datasets

We also conducted age estimation experiments on AgeDB [43] , CACD [44] and IMDB-WIKI [45] . These three databases are illustrated in Figure 2.

(a) AgeDB (b) CACD (c) IMDB-WIKI

Figure 2. Face examples from (a) AgeDB, (b) CACD datasets and (c) IMDB-WIKI dataset.

We extracted BIF [46] and HoG [47] feature vectors and reduced dimensions to 200 by PCA as two view representations. We randomly chose 50, 100, 150 samples for training. Also, we use VGG19 [48] and Resnet50 [45] to extract deep feature vectors from AgeDB, CACD and IMDB-WIKI databases. We report results in Tables 3, 5 and 6.

The estimation errors (MAEs) of all the methods reduced monotonically. The age MAEs of DCLMP

are the second lowest, demonstrating the solidness of our modelling cross-view discriminative knowledge and data manifold structures. We can also observe that C-DCLMP yields the lowest estimation errors, demonstrating its effectiveness and superiority.

Table 2. Age estimation results (MAE±STD) on AgeDB.

training samples	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
50	17.70 ± 0.5	17.78 ± 0.5	16.10 ± 0.4	15.93 ± 0.4	15.62 ± 0.5	15.48 ± 0.2	15.59 ± 0.1	**15.16 ± 0.4**
100	16.81 ± 0.5	17.23 ± 0.6	14.74 ± 0.5	14.79 ± 0.5	14.67 ± 0.4	14.57 ± 0.2	14.60 ± 0.2	**14.13 ± 0.2**
150	15.43 ± 0.5	16.25 ±0 .6	13.83 ± 0.5	13.49 ± 0.4	13.43 ± 0.4	13.21 ± 0.2	13.48 ± 0.2	**13.19 ± 0.3**

Table 3. Age estimation results (MAE±STD) on AgeDB (deep features).

training samples	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
50	16.17 ± 0.5	16.27 ± 0.46	15.42 ± 0.5	15.09 ± 0.5	14.78 ± 0.5	14.67 ± 0.3	14.75 ± 0.2	**14.52 ± 0.2**
100	15.86 ± 0.5	15.79 ± 0.6	14.89 ± 0.8	14.23 ± 0.4	14.09 ± 0.4	13.78 ± 0.3	14.07 ± 0.3	**13.68 ± 0.2**
150	15.09 ± 0.5	14.81 ± 0.3	13.97 ± 0.5	13.41 ± 0.6	13.34 ± 0.5	**13.16 ± 0.3**	13.46 ± 0.3	**13.15 ± 0.3**

Table 4. Age estimation results (MAE±STD) on CACD.

training samples	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
50	16.28 ± 0.5	16.78 ± 0.4	15.79 ± 0.4	14.98 ± 0.4	14.38 ± 0.4	14.28 ± 0.4	14.10 ± 0.3	**13.95 ± 0.3**
100	15.45 ± 0.4	16.52 ± 0.5	15.04 ± 0.5	14.44 ± 0.4	13.99 ± 0.4	13.98 ± 0.3	13.85 ± 0.2	**13.74 ± 0.2**
150	15.20 ± 0.5	15.41 ± 0.5	14.79 ± 0.4	14.02 ± 0.5	13.73 ± 0.5	13.79 ± 0.2	13.67 ± 0.1	**13.63 ± 0.3**

Table 5. Age estimation results (MAE±STD) on CACD (deep features).

training samples	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
50	16.07 ± 0.6	16.27 ± 0.4	15.35 ± 0.4	14.21 ± 0.6	13.39 ± 0.4	13.49 ± 0.3	13.52 ± 0.3	**13.27 ± 0.2**
100	15.69 ± 0.5	15.75 ± 0.3	14.65 ± 0.5	14.17 ± 0.5	13.28 ± 0.3	13.26 ± 0.3	13.24 ± 0.2	**12.97 ± 0.4**
150	15.22 ± 0.4	15.32 ± 0.4	14.45 ± 0.3	14.01 ± 0.6	13.01 ± 0.3	12.94 ± 0.3	12.91 ± 0.3	**12.76 ± 0.4**

Table 6. Age estimation results (MAE±STD) on IMDB-WIKI.

training samples	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
50	14.29 ± 0.5	14.39 ± 0.5	13.49 ± 0.3	13.04 ± 0.5	12.26 ± 0.4	12.37 ± 0.3	11.84 ± 0.3	**11.65 ± 0.3**
100	13.97 ± 0.5	13.87 ± 0.4	12.79 ± 0.5	12.35 ± 0.3	11.96 ± 0.3	11.86 ± 0.3	11.53 ± 0.2	**11.13 ± 0.3**
150	13.43 ± 0.5	13.56 ± 0.5	12.48 ± 0.3	12.26 ± 0.4	11.66 ± 0.3	11.65 ± 0.3	11.45 ± 0.2	**10.98 ± 0.3**

4.4. Parameters analysis

For the proposed methods, we performed parameter analysis t, γ and λ involved in (3.21), respectively. Specifically, we conducted age estimation experiments on both AgeDB and CACD . The results are plotted in Figures 3–5.

Geometric weighting parameter t of C-DCLMP: We find some interesting observations from Figure 3. That is, with t increasing from 0 to 1, the estimation error descended first and then rose again. It shows that the similar manifolds within class and the inter-class data distributions are helpful in regularizing the model solution space.

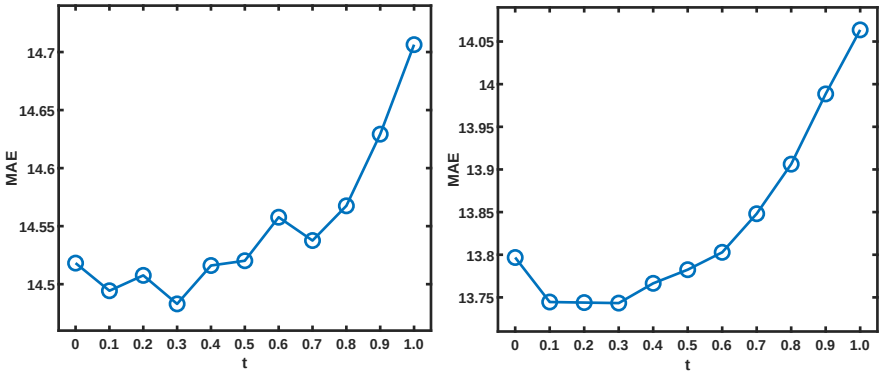

Figure 3. Age estimation MAE on AgeDB (left) and CACD (right) with varying t.

Metric balance parameter γ of C-DCLMP: We can observe from Figure 4 that, the age estimation error (MAE) achieved the lowest values when $0.1 < \gamma < 0.9$. This observation illustrates that preserving the data cross-view discriminative knowledge and the manifold distributions is useful and helps improve the estimation precision.

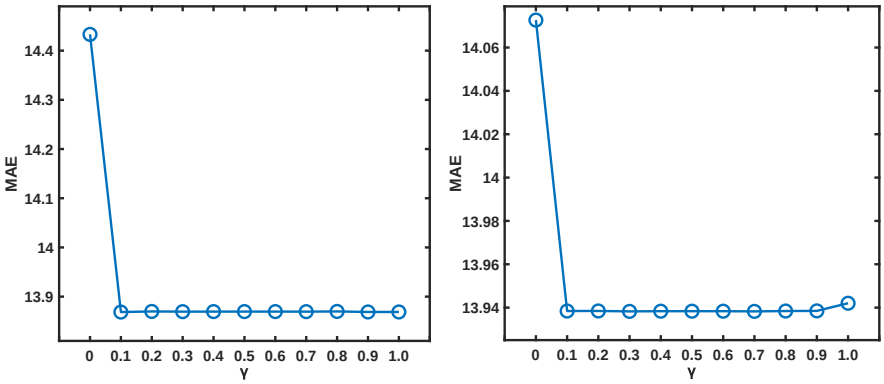

Figure 4. Age estimation MAE on AgeDB (left) and CACD (right) with varying γ.

Metric prior parameter λ of C-DCLMP: Figure 5 shows that, with increased λ value, age estimation error descended to its lowest around λ = 1e-1 and then increased steeply. It demonstrates that incorporating moderate metric prior knowledge can regularize the model solution positively, but excess prior knowledge may dominate the entire data rule and mislead the training of the model.

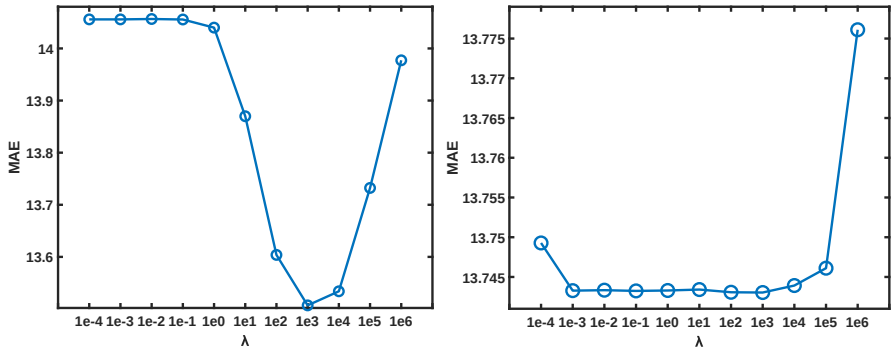

Figure 5. Age estimation MAE on AgeDB (left) and CACD (right) with varying λ.

4.5. Time complexity analysis

For the proposed methods and the comparison methods mentioned above, we performed time complexity analysis. Specifically, we conducted age estimation experiments on both AgeDB and CACD by choosing 100 samples from each class for training while taking the rest for testing, respectively. We reported the averaged results in Table 7.

Table 7. Running time results (MAE±STD) on AgeDB and CACD.

Dataset	CCA	DCA	MPECCA	DCCA	CECCA	CDCA	DCLMP (ours)	C-DCLMP (ours)
AgeDB	0.10 ± 0.12	0.06 ± 0.05	0.41 ± 0.03	0.18 ± 0.03	0.52 ± 0.02	0.11 ± 0.10	57.74 ± 0.71	54.94 ± 0.32
CACD	0.09 ± 0.13	0.06 ± 0.10	0.38 ± 0.09	0.15 ± 0.04	0.47 ± 0.03	0.07 ± 0.01	30.86 ± 0.61	31.04 ± 0.68

4.6. Ablation experiments

For the proposed methods, we performed ablation experiments. Specifically, we conducted age estimation experiments on both AgeDB and CACD. We repeated the experiment 10 times with random data partitions and reported the averaged results in Table 8. In Table 8, each referred part corresponds to Eq (3.7).

Table 8. Ablation experiment results (MAE±STD) on AgeDB and CACD.

Dataset	First part	Second part	Third part	C-DCLMP (ours)(ours)
AgeDB		✓	✓	14.49 ± 0.16
	✓		✓	14.51 ± 0.10
	✓	✓		14.47 ± 0.22
	✓	✓	✓	14.18 ± 0.32
CACD		✓	✓	14.07 ± 0.25
	✓		✓	14.08 ± 0.32
	✓	✓		14.04 ± 0.21
	✓	✓	✓	13.73 ± 0.24

5. Conclusion

In this paper, we proposed a DCLMP, in which both the cross-view discriminative information and the spatial structural information of training data is taken into consideration to enhance subsequent decision making. To pursue closed-form solutions, we remodeled the objective of DCLMP to nonlinear geodesic space and consequently achieved its convex formulation (C-DCLMP). Finally, we evaluated the proposed methods and demonstrated their superiority on various benchmark and real face datasets. In the future, we will consider exploring the latent information of the unlabeled data from the feature and label level, and study how to combine related advanced multi-view learning methods to reduce the computational consumption of the model and further improve the generalization ability of the model in various scenarios.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62176128, the Open Projects Program of State Key Laboratory for Novel Software Technology of Nanjing University under Grant KFKT2022B06, the Fundamental Research Funds for the Central Universities No. NJ2022028, the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund, as well as the Qing Lan Project of the Jiangsu Province.

References

1. P. L. Lai, C. Fyfe, Kernel and nonlinear canonical correlation analysis, International Journal of Neural Systems, *Int. J. Neural Syst.*, **10** (2000), 365–377. https://doi.org/10.1142/S012906570000034X
2. D. R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, *Neural Comput.*, **16** (2004). https://doi.org/10.1162/0899766042321814
3. Q. Tian, C. Ma, M. Cao, S. Chen, H. Yin, A Convex Discriminant Semantic Correlation Analysis for Cross-View Recognition, *IEEE Trans. Cybernetics*, **52** (2020), 1–13. https://doi.org/10.1109/TCYB.2020.2988721
4. Q. Tian, S. Xia, M. Cao, K. Chen, Reliable sensing data fusion through robust multiview prototype learning, *IEEE Trans. Ind. Inform.*, **18** (2022), 2665–2673. https://doi.org/10.1109/TII.2021.3064358
5. P. Zhuang, J. Wu, F. Porikli, C. Li, Underwater image enhancement with hyper-laplacian reflectance priors, *IEEE Trans. Image Process.*, **31** (2022), 5442–5455. https://doi.org/10.1109/TIP.2022.3196546
6. V. Sindhwani, D. S. Rosenberg, An RKHS for multi-view learning and manifold co-regularization, *IEEE Trans. Cybernetics*, **99** (2020), 1–33. https://doi.org/10.1145/1390156.1390279
7. M. H. Quang, L. Bazzani, V. Murino, A unifying framework for vector-valued manifold regularization and multi-view learning, in *Proceedings of the 30th International Conference on Machine Learning*, (2013), 100–108.
8. J. Zhao, X. Xie, X. Xu, S. Sun, Multi-view learning overview: Recent progress and new challenges, *Inform. Fusion*, **38** (2017), 43–54. https://doi.org/10.1016/j.inffus.2017.02.007
9. D. Zhang, T. He, F. Zhang, Real-time human mobility modeling with multi-view learning, *ACM Trans. Intell. Syst. Technol.*, **9** (2017), 1–25. https://doi.org/10.1145/3092692
10. D. Zhai, H. Chang, S. Shan, X. Chen, W. Gao, Multiview metric learning with global consistency and local smoothness, *ACM Trans. Intell. Syst. Technol.*, **3** (2012), 1–22. https://doi.org/10.1145/2168752.2168767
11. P. Zhuang, X. Ding, Underwater image enhancement using an edge-preserving filtering retinex algorithm, *Multimed. Tools Appl.*, **79** (2020), 17257–17277. https://doi.org/10.1007/s11042-019-08404-4

12. T. Sun, S. Chen, J. Yang, P. Shi, A novel method of combined feature extraction for recognition, in *2008 Eighth IEEE International Conference on Data Mining*, (2008), 1043–1048. https://doi.org/10.1109/ICDM.2008.28

13. Y. Peng, D. Zhang, J. Zhang, A new canonical correlation analysis algorithm with local discrimination, *Neural Process. Lett.*, **31** (2010), 1–15. https://doi.org/10.1007/s11063-009-9123-3

14. S. Su, H. Ge, Y. H. Yuan, Multi-patch embedding canonical correlation analysis for multi-view feature learning, *J. Vis. Commun. Image R.*, **41** (2016), 47–57. https://doi.org/10.1016/j.jvcir.2016.09.004

15. Q. S. Sun, Z. D. Liu, P. A. Heng, D. S. Xia, Rapid and brief communication: A theorem on the generalized canonical projective vectors, *Pattern Recogn.*, **38** (2005), 449–452. https://doi.org/10.1016/j.patcog.2004.08.009

16. H. K. Ji, Q. S. Sun, Y. H. Yuan, Z. X. Ji, Fractional-order embedding supervised canonical correlations analysis with applications to feature extraction and recognition, *Neural Process. Lett.*, **45** (2017), 279–297. https://doi.org/10.1007/s11063-016-9524-z

17. X. D. Zhou, X. H. Chen, S. C. Chen, Combined-feature-discriminability enhanced canonical correlation analysis, *Pattern Recogn. Artif. Intell.*, **25** (2012), 285–291.

18. P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.*, **19** (1997), 711–720. https://doi.org/10.1109/34.598228

19. F. Zhao, L. Qiao, F. Shi, P. Yap, D. Shen, Feature fusion via hierarchical supervised local CCA for diagnosis of autism spectrum disorder, *Brain Imaging Behav.*, **11** (2017), 1050–1060. https://doi.org/10.1007/s11682-016-9587-5

20. M. Haghighat, M. Abdel-Mottaleb, W. Alhalabi, Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition, *IEEE Trans. Inform. Foren. Sec.*, **11** (2016), 1984–1996. https://doi.org/10.1109/TIFS.2016.2569061

21. A. Sharma, A. Kumar, H. Daume, D. W. Jacobs, Generalized multiview analysis: A discriminative latent space, in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (2012), 2160–2167. https://doi.org/10.1109/CVPR.2012.6247923

22. S. Sun, X. Xie, M. Yang, Multiview uncorrelated discriminant analysis, *IEEE Trans. Cybernetics*, **46** (2016), 3272–3284. https://doi.org/10.1109/TCYB.2015.2502248

23. P. Hu, D. Peng, J. Guo, L. Zhen, Local feature based multi-view discriminant analysis, *Knowl.-Based Syst.*, **149** (2018), 34–46. https://doi.org/10.1016/j.knosys.2018.02.008

24. X. Fu, K. Huang, M. Hong, N. D. Sidiropoulos, A. M. C. So, Scalable and flexible multiview MAX-VAR canonical correlation analysis, *IEEE Trans. Signal Process.*, **65** (2017), 4150–4165. https://doi.org/10.1109/TSP.2017.2698365

25. D. Y. Gao, Canonical duality theory and solutions to constrained nonconvex quadratic programming, *J. Global Optim.*, **29** (2004), 377–399. https://doi.org/10.1023/B:JOGO.0000048034.94449.e3

26. J. Fan, S. Chen, Convex discriminant canonical correlation analysis, *Pattern Recogn. Artif. Intell.*, **30** (2017), 740–746. https://doi.org/10.16451/j.cnki.issn1003-6059.201708008

27. C. Tang, X. Zheng, X. Liu, W. Zhang, J. Zhang, J. Xiong, et al., Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection, *IEEE Trans. Knowl. Data Eng.*, **34** (2022), 4705–4716. https://doi.org/10.1109/TKDE.2020.3048678

28. C. Tang, Z. Li, J. Wang, X. Liu, W. Zhang, E. Zhu, Unified one-step multi-view spectral clustering, *IEEE Trans. Knowl. Data Eng.*, **35** (2023), 6449–6460. https://doi.org/10.1109/TKDE.2022.3172687

29. J. Wang, C. Tang, Z. Wan, W. Zhang, K. Sun, A. Y. Zomaya, Efficient and Effective One-Step Multiview Clustering, *IEEE Trans. Neur. Net. Learn. Syst.*, (2023), 1–12. https://doi.org/10.1109/TNNLS.2023.3253246

30. P. L. Lai, C. FyFe, KERNEL AND NONLINEAR CANONICAL CORRELATION ANALYSIS, International Journal of Neural Systems, **10** (2000), 365–377.

31. K Fukumizu, FR Bach, A Gretton, Statistical consistency of kernel canonical correlation analysis, *J. Mach. Learn. Res.*, **8** (2007), 361–383.

32. T. Liu, T. K. Pong, Further properties of the forward Cbackward envelope with applications to difference-of-convex programming, *Comput. Optim. Appl.*, **67** (2017), 480–520. https://doi.org/10.1007/s10589-017-9900-2

33. T. P. Dinh, H. M. Le, H. A. Le Thi, F. Lauer, A difference of convex functions algorithm for switched linear regression, *IEEE Trans. Automat. Contr.*, **59** (2014), 2277–2282. https://doi.org/10.1109/TAC.2014.2301575

34. P. Zadeh, R. Hosseini, S. Sra, Geometric mean metric learning, in *Proceedings of The 33rd International Conference on Machine Learning*, (2016), 2464–2471.

35. B. Stephen, V. Lieven, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

36. V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, *SIAM J. Matrix Anal. Appl.*, **29** (2007), 328–347. https://doi.org/10.1137/050637996

37. A. Papadopoulos, *Metric Spaces, Convexity and Nonpositive Curvature*, European Mathematical Society, Zurich, 2005.

38. T. Rapcsák, Geodesic convexity in nonlinear optimization, *J. Optim. Theory Appl.*, **69** (1991), 169–183. https://doi.org/10.1007/BF00940467

39. C. L. Liu, K. Nakashima, H. Sako, H. Fujisawa, Handwritten digit recognition: investigation of normalization and feature extraction techniques, *Pattern Recogn.*, **37** (2004), 265–279. https://doi.org/10.1016/S0031-3203(03)00224-3

40. Pawlicki, D. S. Lee, Hull, Srihari, Neural network models and their application to handwritten digit recognition, in *IEEE 1988 International Conference on Neural Networks*, **2** (1988), 63–70. https://doi.org/10.1109/ICNN.1988.23913

41. C. H. Lampert, H. Nickisch, S. Harmeling, Learning to detect unseen object classes by between-class attribute transfer, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009),951–958. https://doi.org/10.1109/CVPR.2009.5206594

42. C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, *J. Magn. Reson. Imaging*, **27** (2008), 685–691. https://doi.org/10.1002/jmri.21049

43. S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, S. Zafeiriou, Agedb: the first manually collected, in-the-wild age database, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, (2017), 51–59.

44. B. C. Chen, C. S. Chen, W. H. Hsu, Cross-age reference coding for age-invariant face recognition and retrieval, in *Computer Vision – ECCV 2014.*, Springer, (2014), 768–783. https://doi.org/10.1007/978-3-319-10599-4_49

45. R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, *Int. J. Comput. Vis.*, **126** (2018),144–157. https://doi.org/10.1007/s11263-016-0940-3

46. G. Guo, G. Mu, Y. Fu, T. S. Huang, Human age estimation using bio-inspired features, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009), 112–119. https://doi.org/10.1109/CVPR.2009.5206681

47. Q. Zhu, M. C. Yeh, K. T. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, (2006), 1491–1498. https://doi.org/10.1109/CVPR.2006.119

48. K. Simonyan, A. Zisserma, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556.

AIMS Press