Deep Incomplete Multi-View Learning Network with Insufficient Label Information

Zhangqi Jiang[1], Tingjin Luo[1]*, Xinyan Liang[2]

[1]College of Science, National University of Defense Technology, Changsha 410073, Hunan, China
[2]Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, Shanxi, China
jiangzq@nudt.edu.cn, tingjinluo@hotmail.com, liangxinyan48@163.com

Abstract

Due to the efficiency of integrating semantic consensus and complementary information across different views, multi-view classification methods have attracted much attention in recent years. However, multi-view data often suffers from both the miss of view features and insufficient label information, which significantly decrease the performance of traditional multi-view classification methods in practice. Learning for such simultaneous lack of feature and label is crucial but rarely studied. To tackle these problems, we propose a novel Deep Incomplete Multi-view Learning Network (DIMvLN) by incorporating graph networks and semi-supervised learning in this paper. Specifically, DIMvLN firstly designs the deep graph networks to effectively recover missing data with assigning pseudo-labels of large amounts of unlabeled instances and refine the incomplete feature information. Meanwhile, to enhance the label information, a novel pseudo-label generation strategy with the similarity constraints of unlabeled instances is proposed to exploit additional supervisory information and guide the completion module to preserve more semantic information of absent multi-view data. Besides, we design view-specific representation extractors with the autoencoder structure and contrastive loss to learn high-level semantic representations for each view, promote cross-view consistencies and augment the separability between different categories. Finally, extensive experimental results demonstrate the effectiveness of our DIMvLN, attaining noteworthy performance improvements compared to state-of-the-art competitors on several public benchmark datasets. Code will be available at GitHub.

Introduction

For the potential to improve classification performance by exploiting the relevant information from multiple views, multi-view classification (MvC) (Zhao et al. 2017) has been widely applied in real-world applications, such as autonomous driving (Chen et al. 2017) and computational medicine (Gray et al. 2013). Generally, the target of MvC methods (Andrew et al. 2013; Kan, Shan, and Chen 2016; Wang et al. 2020) is to learn the view shared latent subspace and a discriminative classifier by exploring the consistency information from multi-view data. The traditional MvC methods can successfully learn the consistency information of multi-view data and typically depend on an assumption that all of the training instances are with complete features in all views and sufficient labels.

Unfortunately, in many real-world applications, multi-view data often simultaneously exhibit the view missing and label scarcity issues. For example, in the disease diagnostic system, only a small subset of patients receive a definite diagnosis of the specific disease (or label), while the majority of patients fall into the unknown state in the hospital database. Besides, not all of patients undergo the exact same and all physical examinations, where items not performed by the patient are regarded as missing views. To analyze this kind of data, there are at least two challenges: (1) lack of labeled data and (2) missing the view features. Due to the double missing of labels and features, it will dramatically decrease the classification performance and limit their wide application of the existing MvC methods. How to design a model for such incomplete multi-view semi-supervised classification (IMvSSC) problem is crucial but rarely studied. In this paper, we will propose a novel deep method to tackle these problems.

In literature, several methods have been proposed to address the two mentioned issues individually. To solve the multi-view semi-supervised classification (MvSSC) problem, there are many MvSSC approaches proposed and can be roughly divided into three major categories, i.e., co-training based methods (Cheng et al. 2016; Xia et al. 2020), alignment-based methods (Jing et al. 2017; Wu et al. 2019), and graph-based methods (Cai et al. 2013; Gong et al. 2016). These MvSSC methods concentrate on making full use of the extrinsic information contained in unlabeled data to enhance the classification performance, nonetheless, they neglect the view missing issue. For the incomplete issue, (Zhang et al. 2019; Liu et al. 2021; Lin et al. 2022; Wang et al. 2022b) have proposed many shallow and deep incomplete multi-view learning (IMvL) methods. These methods can effectively learn informative representations from data with missing views. However, these methods are unsupervised or fully supervised, and not suitable to the insufficient label scenarios. As far as we know, only a few works with the shallow model can tackle the double missing issues of the IMvSSC problem. Yang et al. (2018) proposed a semi-supervised multi-modal learning with incomplete modali-

*Corresponding author.

ties (SLIM) by learning view-specific projection and a common label matrix simultaneously based on the reconstruction of the incomplete similarity. Besides, Zhuge et al. (2023) proposed the absent multi-view semi-supervised classification (AMSC) method to jointly learn the view-specific and shared label indicator matrices, and classify the unlabeled data through a label propagation mechanism by combining the extra-view and intra-view similarity loss. However, both SLIM and AMSC employ traditional shallow methods to learn the information of features and labels, which are difficult to capture high-level semantic information hidden in all views. Moreover, these methods are limited by their training manner, which only enables them to predict the labels during the learning process and are unable to handle new incoming data without re-training.

To these problems, we propose a novel deep IMvSSC model based on the deep graph neural network (GNN) named as DIMvLN in this paper. The framework of DIMvLN consists of three major components, i.e., GNN-based completion module, multi-view representation learning module and semi-supervised learning module. Specifically, to leverage the hidden information from incomplete data to mitigate the negative influence of missing views, DIMvLN employs the view-specific GNN to recover the missing data based on the existing similarity relations, which also guarantees that the following modules can learn essential information. Based on the autoencoder structures, multi-view representation module is devised to learn the high-level semantic and complementary information from the recovered multi-view data. Moreover, to preserve the cross-view consensus and enhance the separability between different categories, DIMvLN employs the contrastive loss on both instance-level and category-level. In semi-supervised learning module, we establish a pseudo-label generation strategy to explore the additional supervisory information contained in ample unlabeled data by assigning pseudo-labels for the confident classified unlabeled instances. In addition, we develop the end-to-end training strategy to make our DIMvLN predict the labels of new coming data and handle the large-scale problem. Our main contributions are listed as follows:

- We propose the DIMvLN method to solve this crucial, but rarely studied problem that has arisen from many real applications. To our knowledge, this may be the first attempt concerning deep classification model in this simultaneous miss of feature and label scenario.
- DIMvLN is a unified framework of IMvSSC and designed to recover the absent information, preserve the high-level semantic representation and enlarge the extrinsic supervisory information of unlabeled instances, simultaneously.
- Extensive experimental results indicate that our DIMvLN outperforms other compared approaches in almost all cases and demonstrate its superiority and effectiveness.

Related Work

Incomplete Multi-view Learning

In literature, there are various methods proposed to solve IMvL, which can be split into the following groups. Matrix factorization-based methods (Li, Jiang, and Zhou 2014; Shao, He, and Yu 2015; Zhao, Liu, and Fu 2016; Hu and Chen 2018) are one group of them, which aim to learn a common latent representation that satisfies the low-rankness constraint. Another group is kernel learning and spectral clustering based methods. Liu et al. (2021) proposed efficient and effective incomplete multi-view clustering (EE-IMVC) to impute missing values and learn a consensus clustering matrix. Besides, Wen, Xu, and Liu (2020) combined spectral clustering and graph learning to jointly learn the low-dimensional common representations and view-specific similarity graphs. However, these IMvL methods are based on the shallow model and difficult to mine complex structured information of multi-view data. Recently, since deep neural networks are able to capture complex semantic information effectively, some deep learning based methods (Zhang et al. 2019; Lin et al. 2022; Wang et al. 2022b) have been proposed and achieved promising performance.

Multi-view Semi-supervised Classification

Semi-supervised learning (SSL) is a promising learning paradigm, which aims to tackle the scarcity of labeled data by leveraging a large amount of unlabeled data. To solve the MvSSC problem, several MvSSC methods have been proposed in literature. Cheng et al. (2016) proposed a diversity preserving co-training MvSSC algorithm by training two individual classifiers with initial labeled data to exploit the complementary information. To seek the optimal correlation between different views, Chen et al. (2012) and Jiang and Li (2017) proposed the alignment-based canonical correlation analysis and cross-modal hashing methods with correlation constraints. Then, to preserve the similarity information of instances, Cai et al. (2013) proposed a graph-based MvSSC method named as AMMSS, which learns a common label matrix and weights for each modality simultaneously. Meanwhile, to enhance the label information of unlabeled instances, Nie et al. (2018) proposed the multi-view learning with adaptive neighbors (MLAN) by learning a shared similarity graph with local structure learning and allocating weights for each view automatically. Recently, Jia et al. (2021) proposed a semi-supervised multi-view deep discriminant representation learning framework, which employs the orthogonality and similarity constraints to reduce the redundancy of learned representations.

Methodology

Notations and Problem Formulation. Suppose an incomplete multi-view dataset with N instances and V views, i.e., $\mathcal{X} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$, where $\boldsymbol{X}^{(v)} = \{\boldsymbol{x}_i^{(v)}\}_{i=1}^{N} \in \mathbb{R}^{d_v \times N}$ is d_v dimensional feature matrix of the v-th view. Let $\boldsymbol{M} \in \mathbb{R}^{N \times V}$ be the indicator matrix, where $\boldsymbol{M}_{i,v} = 1$ means the i-th instance has its feature of v-th view; otherwise $\boldsymbol{M}_{i,v} = 0$, that its feature is missing and set as 'NaN'. To be concise, let $\mathcal{C}^{(v)}$ and $\mathcal{U}$ be the index of the complete instances in the v-th view and the unlabeled instances, respectively. The ground truth of the labeled instances is $\mathcal{Y} = \{y_i \in \{1, \ldots, C\} | 1 \leq i \leq N, i \notin \mathcal{U}\}$, where C is the number of classes.

Definition 1: (Incomplete Multi-view Semi-supervise

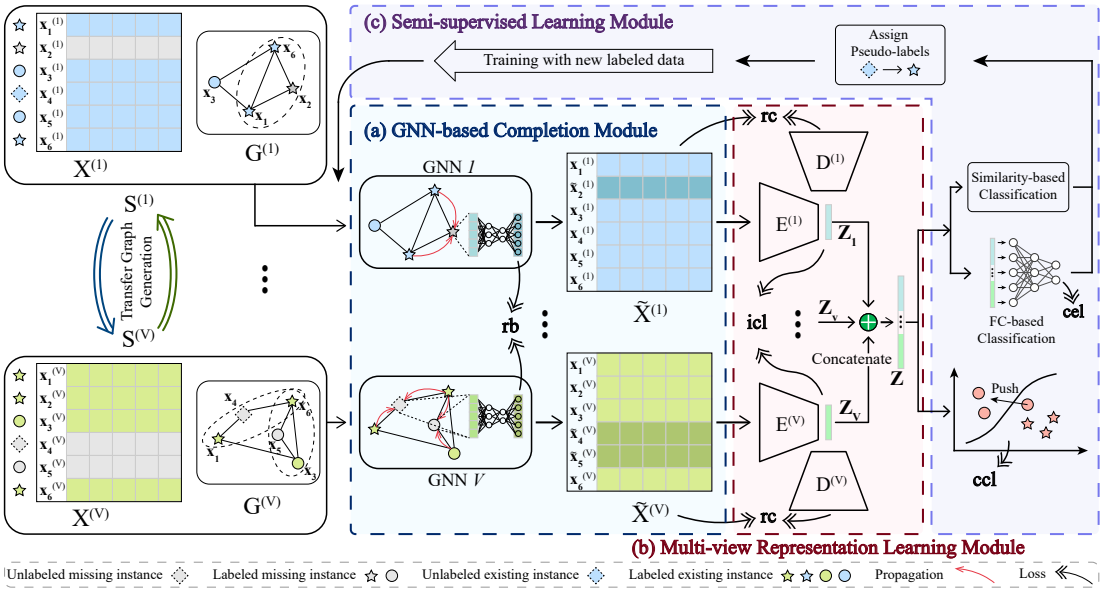


Figure 1: The main framework of our proposed DIMvLN, which is composed of three modules: (a) GNN-based completion module; (b) Multi-view representation learning module and (c) Semi-supervised learning module.

Learning). Given a training incomplete multi-view dataset $\mathcal{X}_{\text{train}} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ with its partial labels $\mathcal{Y}_{\text{train}} = \{y_1, \cdots, y_{n_l}\}$, where $n_l \ll N$ and its corresponding indicator matrix $\boldsymbol{M}_{\text{train}}$. The goal of our DIMvLN is to learn a discriminative multi-view classification model that can predict the label of the new instances $\mathcal{X}_{\text{test}}$.

Obviously, both the absence of view features and the scarcity of labeled data seriously decrease the multi-view classification performance and limit the practical application. The features missing of the instances will weaken the cross-view semantic consensus and degrade the generalization capability of the models. Besides, the scarcity of labels further degrades the performance of the models. To remedy this, we propose a novel deep learning framework named DIMvLN to address the simultaneous miss of feature and label. The main framework of our DIMvLN is illustrated in Fig. 1. Specifically, DIMvLN consists of three main modules:(a) GNN-based completion module for refining the incomplete feature information; (b) Multi-view representation learning module for capturing the high-level semantic consensus and complementary information in all views; (c) Semi-supervised learning module for extracting extrinsic supervisory information from unlabeled data. The concrete details are provided in the following.

GNN-based Completion Module

Data augmentation is a powerful technique to improve the performance and generalization capabilities of deep learning models (Shorten and Khoshgoftaar 2019). And data recovery can be considered as a variant of data augmentation tailored for handling incomplete multi-view data. Recently, GNN-based methods have gained popularity in data recovery due to their ability to mine the geometric information of data (Wang et al. 2022b; Liang et al. 2023). Sato (2023) further provides theoretical evidence supporting the effectiveness of GNNs in recovering hidden features. Therefore, we utilize the GNNs to recover the missing data based on the similarity relations between existing data for augmenting the incomplete data against the influence of view missing.

Firstly, DIMvLN constructs the view-specific graph $\boldsymbol{S}^{(v)} \in \mathbb{R}^{N\times N}$ to the existing data through k-nearest neighbors (k-NN) algorithm, where $\boldsymbol{S}_{i,j}^{(v)} = 1$ means $\boldsymbol{M}_{i,v}\boldsymbol{M}_{j,v} = 1$ and $\boldsymbol{x}_j^{(v)}$ is the neighbor of $\boldsymbol{x}_i^{(v)}$; otherwise, $\boldsymbol{S}_{i,j}^{(v)} = 0$. Inspired by (Wang et al. 2022b), we then transfer the known graph relations to find the available instances related to the missing ones in each view, and the transferred k-NN graph can be obtained by:

$$\boldsymbol{G}^{(v)} = \sum_{k=1,k\neq v}^{V} \boldsymbol{S}^{(k)}\text{diag}(\boldsymbol{M}_{:,k}), \tag{1}$$

where operator $\text{diag}(\cdot)$ creates a diagonal matrix, and $\boldsymbol{M}_{:,k}$ denotes the k-th column of matrix $\boldsymbol{M}$.

Next, we feed $\boldsymbol{G}^{(v)}$ as the adjacency matrix and the existing instance features $\boldsymbol{x}_j^{(v)}$ $s.t.$ $j \in \mathcal{C}^{(v)}$ as the attributes of nodes into the view-specific GNN to recover the missing data. After message propagation over $\boldsymbol{G}^{(v)}$ in the first layer of GNN, the initially reconstructed data can be obtained by:

$$\hat{\boldsymbol{x}}_i^{(v)} = \sigma\Big(\boldsymbol{b}_v + \sum_{\boldsymbol{G}_{i,j}^{(v)}\geq 1} \boldsymbol{G}_{i,j}^{(v)}\boldsymbol{\omega}_v\boldsymbol{x}_j^{(v)}\Big), \tag{2}$$

where $\boldsymbol{b}_v$ and $\boldsymbol{\omega}_v$ are the bias and transformation matrix of the v-th view, respectively; σ represents an activation function, which is set as rectified linear unit (ReLU) activation function in our experiments. In the remaining structure of GNN, instead of using the transferred relations, we adopt

two stacked fully-connected layers with ReLU activation function to further adapt the referred data. Finally, the reconstructed missing data is combined with the available data to obtain the recovered matrices $\{\tilde{\boldsymbol{X}}^{(v)}\}_{v=1}^{V}$.

To improve the reconstruction performance and maintain the stability of the subsequent modules, we pre-train the view-specific GNNs by minimizing the rebuilding loss L_{rb}. For a mini-batch, we denote the index of instances in a batch as $\mathcal{B}$ and the rebuilding loss can be written as:

$$L_{\text{rb}} = \sum_{v=1}^{V} \sum_{i \in \mathcal{B} \backslash \mathcal{C}^{(v)}} \sum_{\boldsymbol{G}_{i,j}^{(v)} \geq 1} \|\tilde{\boldsymbol{X}}_{:,i}^{(v)} - \boldsymbol{x}_j^{(v)}\|_2^2. \tag{3}$$

Multi-view Representation Learning Module

In our framework, we devise view-specific extractors in the autoencoder structures with contrastive learning, aiming to learn high-level discriminative representations from $\{\tilde{\boldsymbol{X}}^{(v)}\}_{v=1}^{V}$ rather than focusing on shallow-level features as commonly done in traditional approaches. For convenience, let $E^{(v)}(\cdot)$ and $D^{(v)}(\cdot)$ be the encoder and decoder for the v-th view, respectively. The representation matrix $\boldsymbol{Z}^{(v)} = E^{(v)}(\tilde{\boldsymbol{X}}) \in \mathbb{R}^{m \times N}$ can be learned by minimizing:

$$L_{\text{rc}} = \sum_{i=1}^{V} \sum_{k \in \mathcal{B}} \|\tilde{\boldsymbol{X}}_{:,k}^{(v)} - D^{(v)}(\boldsymbol{Z}_{:,k}^{(v)})\|_2^2, \tag{4}$$

where m denotes the dimension of representations. Then, we concatenate the view-specific representation matrices to obtain a common representation matrix $\boldsymbol{Z} \in \mathbb{R}^{mV \times N}$.

Furthermore, to enhance the discriminative of the learned representations, we employ the instance-level and category-level contrastive loss introduced in (Lin et al. 2022) to encourage the cross-view consistencies and bolster the separability between distinct categories. Specifically, the instance-level contrastive loss L_{icl} utilizes both labeled and unlabeled instances to maximize the mutual information between the representations of different views. To calculate the mutual information, $\boldsymbol{z}_i^{(v)}$ is treated as a distribution probability vector over m classes (Ji, Henriques, and Vedaldi 2019), by employing a Softmax activation function at the last layer of the encoder, and the loss L_{icl} can be defined as below:

$$\boldsymbol{P}^{(v,v^*)} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \boldsymbol{z}_i^{(v)} \left(\boldsymbol{z}_i^{(v^*)}\right)^T, \tag{5}$$

$$\ell_{v,v^*} = -\sum_{t=1}^{m} \sum_{t'=1}^{m} \boldsymbol{P}_{t,t'}^{(v,v^*)} \ln \left(\frac{\boldsymbol{P}_{t,t'}^{(v,v^*)}}{\left(\boldsymbol{P}_{t}^{(v)}\right)^{\alpha+1} \left(\boldsymbol{P}_{t'}^{(v^*)}\right)^{\alpha+1}} \right), \tag{6}$$

$$\mathcal{L}_{\text{icl}} = \frac{1}{V} \sum_{1 \leq v < v^* \leq V} \ell_{v,v^*}, \tag{7}$$

where $\boldsymbol{P}^{(v,v^*)} \in \mathbb{R}^{m \times m}$ represents the joint probability distribution over the v-th and v^*-th views; $\boldsymbol{P}^{(v)}$ and $\boldsymbol{P}^{(v^*)}$ are the marginal probability distribution of the v-th and v^*-th view, respectively; α is a balance parameter. In our experiments, we simply fix the α to 9.

For the category-level contrastive loss L_{ccl}, the available label information is used to make the decision boundaries between categories clearer by pushing the misclassified instances closer to the real within-class instances in the common latent representation space. We adopt a classification method based on cosine similarity metric to predict the labels of labeled instances and penalize the misclassification according to the ground truth $\mathcal{Y}$ and prediction $\mathcal{Y}_p$. The predicted labels can be obtained by:

$$y_{p,i} = \underset{y_{p,i} \in \{1,2,\ldots,C\}}{\operatorname{argmax}} \boldsymbol{z}_i^T \boldsymbol{Z}_l \boldsymbol{H} \operatorname{diag}(\mathbf{1}\boldsymbol{H})^{-1}, \tag{8}$$

where $\boldsymbol{z}_i$ and $y_{p,i}$ are the common representation and the predicted label of the i-th instance, respectively; $\boldsymbol{Z}_l$ denotes the matrix $\{\boldsymbol{Z}_{:,i} | i \in \mathcal{B} \backslash \mathcal{U}\}$; $\boldsymbol{H} \in \mathbb{R}^{|\mathcal{B} \backslash \mathcal{U}| \times C}$ denotes the real label indicator matrix of $\mathcal{Y}$ in mini-batch (i.e., $\boldsymbol{H}_{i,y_{gt,i}} = 1$, the other elements are zero); $\mathbf{1}$ is an all-one row vector. Similarly to $\boldsymbol{H}$, let $\boldsymbol{F}$ be the predicted label indicator matrix of labeled instances and then the L_{ccl} can be written as:

$$L_{\text{ccl}} = \sum_{i \in \mathcal{B} \backslash \mathcal{U}} \boldsymbol{A}_i (\boldsymbol{F}_{i,:} - \boldsymbol{H}_{i,:})^T + \mathbb{I}(y_{p,i} \neq y_{gt,i}), \tag{9}$$

where $\boldsymbol{A}_i = \boldsymbol{z}_i^T \boldsymbol{Z}_l \boldsymbol{H} \operatorname{diag}(\mathbf{1}\boldsymbol{H})^{-1}$ and $\mathbb{I}(\cdot)$ is the indicator function which guarantees the summand is positive.

Semi-supervised Learning Module

Due to the capability to supply new labeled data for training, pseudo-label style methods have gained significant prominence in deep semi-supervised classification tasks in recent years (Jia et al. 2021). In our work, we establish the pseudo-labeling mechanism with a deep classification network to annotate pseudo-labels for unlabeled data. The common representations are first fed into the deep classification network to predict soft labels of instances. However, during the early training stage, deep learning based classifiers usually exhibit low classification confidence. Hence, we employ the same similarity-based classification method, used in multi-view representation module, to re-check the predictions. Specifically, we assign pseudo-labels to the unlabeled instances where the two predictions agree with each other. Finally, the newly labeled instances are added to the labeled training data. In addition, to improve accuracy, the cross-entropy loss L_{cel} is applied to train the classifier.

Training Strategy

The training strategy employed in DIMvLN comprises two phases: pre-training and alternate optimization. In the pre-training phase, we only use the rebuilding loss $\mathcal{L}_{\text{rb}}$ to warm-up the view-specific GNNs. During the alternate optimization phase, the three modules of the proposed framework mutually complement each other and simultaneously enhance the classification performance. The overall loss of the alternate optimization is defined as:

$$\mathcal{L} = L_{\text{rc}} + \lambda_1 (L_{\text{icl}} + L_{\text{ccl}}) + \lambda_2 L_{\text{cel}}, \tag{10}$$

where the parameters λ_1 and λ_2 trade-off the importance of the contrastive loss ($L_{\text{icl}} + L_{\text{ccl}}$) and the classification loss L_{cel}. In our experiments, these two parameters are fixed to 1. The training strategy is summarized in Algorithm 1.

Algorithm 1: Training Strategy of DIMvLN

Input: Training data $\mathcal{X}_{\text{train}}$ with partial training label $\mathcal{Y}_{\text{train}}$; parameters of the whole framework Θ and batch size $|\mathcal{B}|$.

Initialization: Init $\{\boldsymbol{S}_{\text{train}}^{(v)}\}_{v=1}^{V}$ by k-NN graphs and the transferred graphs $\{\boldsymbol{G}_{\text{train}}^{(v)}\}_{v=1}^{V}$ by Eq. (1).

$\mathcal{Y}_{temp} \leftarrow \mathcal{Y}_{\text{train}}, \mathcal{U}_{temp} \leftarrow \mathcal{U}_{\text{train}}$;
for $epoch \leftarrow 1$ to $MaxIter$ **do**
  Obtain the recovered matrices $\{\tilde{\boldsymbol{X}}_{\text{train}}^{(v)}\}_{v=1}^{V}$ by Eq. (2);
  Learn its features $\{\boldsymbol{Z}_{\text{train}}^{(v)}\}_{v=1}^{V}$ by $\{E(\cdot)^{(v)}\}_{v=1}^{V}$;
  Obtain the common representation matrix $\boldsymbol{Z}_{\text{train}}$ by concatenating $\{\boldsymbol{Z}_{\text{train}}^{(v)}\}_{v=1}^{V}$;
  Compute the soft labels Y_s of $\boldsymbol{Z}_{\text{train}}$ by Eq. (8);
  Predict the labels Y_n of $\boldsymbol{Z}_{\text{train}}$ by deep neural network;
  Compute overall loss $\mathcal{L}$ by Eq. (10);
  for all $i \in \mathcal{U}_{temp}$ **do**
    if $Y_{s,i} = Y_{n,i}$ **then**
      $\mathcal{Y}_{temp} \leftarrow \mathcal{Y}_{temp} \cup \{Y_{s,i}\}, \mathcal{U}_{temp} \leftarrow \mathcal{U}_{temp} \setminus \{i\}$;
    end if
  end for
  Update $\Theta \leftarrow$ Optimizer($\mathcal{L}$, Θ);
  Stop training when the model converges.
end for

Experiments

Experimental Setup

Datasets. We conduct experiments on six public multi-view datasets as follows. **Caltech101-20** : It contains 2,386 images of 20 objects. Following (Lin et al. 2022), we select HOG and GIST features as two views. **CUB** : It includes 11,788 samples belonging to 200 bird species. Following (Zhang et al. 2019), we select the top 10 bird species with two views. **Wikipedia** : It contains image and text features from 2,866 documents on 29 topics. Following (Wang, Yang, and Li 2016), the top 10 most popular topics are selected for our experiment. **ALOI** : It collects 110,250 images for 1,000 small objects. Following (Huang, Wang, and Lai 2023), we use a subset that contains 10,800 images of 100 objects with four views. **Out-Scene** : It contains 4,485 images of 15 scene categories. Following (Huang, Wang, and Lai 2023), we select 8 outdoor categories with total 2,688 images with four views. **Animal** [1] : It contains 50 animals of 30,475 images and we use the subset of 11,673 images from the first 20 animals with four views.

Comparison Methods. To validate the effectiveness of DIMvLN, we compare it with six state-of-the-art approaches, which can be categorized into two groups: traditional methods and deep methods. Traditional methods include **SLIM** (Yang et al. 2018), **AMSC** (Zhuge et al. 2023), **AMMSS** (Cai et al. 2013), while deep methods include **CPM-Nets** (Zhang et al. 2019), **DCP** (Lin et al. 2022) and **TMC** (Han et al. 2023). Similar to (Zhuge et al. 2023), we introduce a grid search strategy to determine the optimal parameter within the recommended range for all compared methods and adopt the recommended network structures as their baselines. Note that CPM-Nets, DCP and TMC are supervised methods. Thus we train these methods using only labeled instances. Besides, for AMMSS and TMC, which can only handle complete multi-view data, we use its mean value to fill the missing data in each view.

Data Preparation. Each data can be split into training, validation, and test sets in the ratio of 7:1:2. Besides, to simulate the partial view setting, we randomly remove some view of samples from each set. Concretely, according to the pre-set partial example ratio (PER), PER% instances are randomly selected as incomplete instances, which randomly missing $1 \sim V$-1 views. To mimic the SSL situation, according to the pre-set labeled example ratio (LER), we randomly select LER% instances in training set as labeled instances.

Implementation Details. Adam optimizer with the initial learning rate of 0.0001 is used for optimization of all datasets. The k-NN graphs are constructed using k-NN algorithm with Euclidean distance metric, where the neighbor number k is fixed to 10 for all datasets. In addition, Accuracy (ACC), F1-score (F1), Precision (PREC), and AUC are adopted as the evaluation metrics. Due to the limit of space, we only present ACC results. The results of other three metrics are shown in the supplementary file. All the experimental results are obtained by independently running the methods ten times, and the final average results with standard deviations are reported. Our model is implemented by PyTorch on one NVIDIA Geforce A100 with GPU of 40GB memory.

Experimental Results

Performance Evaluation. To verify the effectiveness of our DIMvLN comprehensively, we compare it with six state-of-the-art competitive methods from two aspects: i) view missing and ii) label insufficient. For view missing, we fix LER to 10%, while PER is selected in {0%,10%,30%,50%,70%,90%}. The results of ACC are listed in Table 1 and we have the following observations: 1) When PER=0%, the proposed method achieves the highest performance on all datasets, validating that DIMvLN is also stable and effective for SSMvC task. 2) With PER increasing from 10% to 90%, DIMvLN outperforms the other six competitors on all datasets. 3) Our method is robust to incomplete multi-view data since DIMvLN consistently exhibits relatively promising performance with highly PER. For example, DIMvLN and the most competitive method DCP achieve ACC of 39.82% and 34.88% when PER=0% on Animal. As PER=90%, the performance of DIMvLN is 28.11% and remarkably superior to 17.46% of DCP.

For label insufficient, we fix PER to 50%, while varying LER from 5% to 35% with a gap of 5%. The results of all compared methods are shown in Fig. 2 and we could observe that: 1) Our method achieves the highest performance among all compared methods in almost all cases. 2) With decreasing LER, the performance degradation of the compared methods is much larger than that of ours. For example, on ALOI and Out-Scene, as LER is reduced from 30% to 5%, the performance decline of DIMvLN is less than 1%, while the decreases of other compared methods are more than 5%.

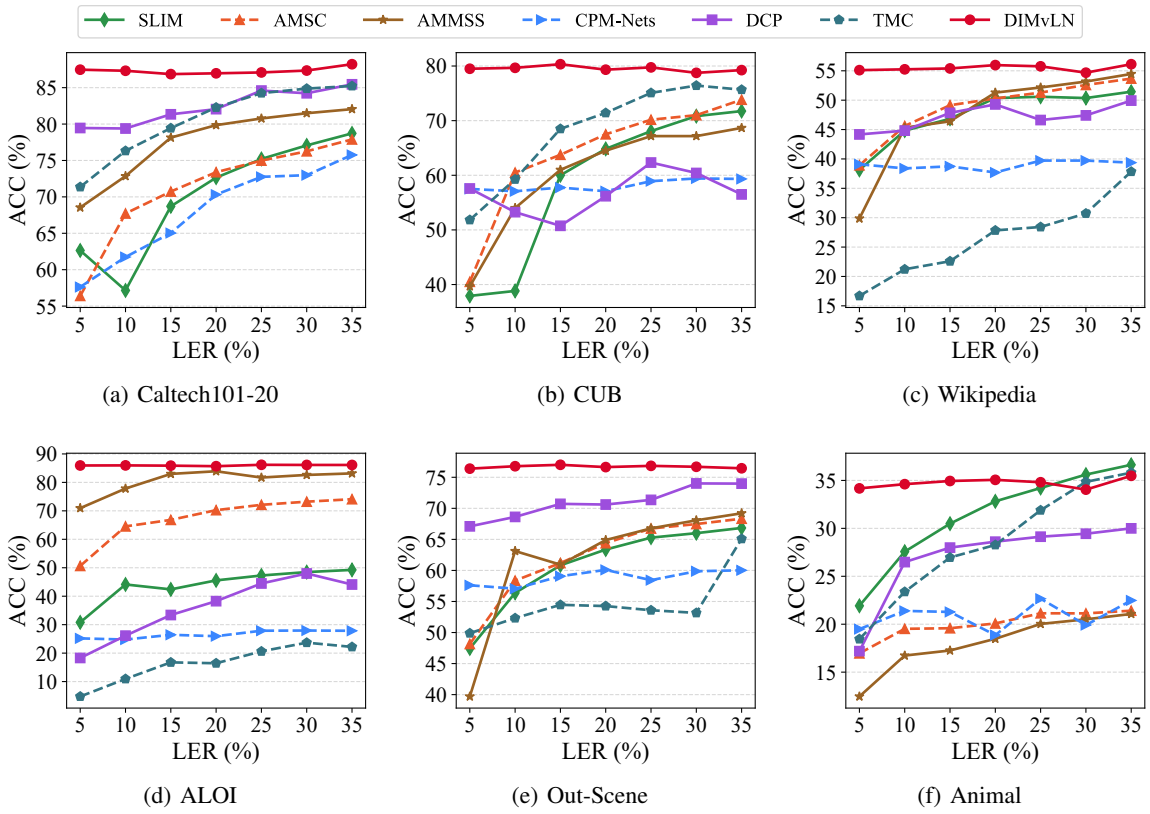
[1] https://cvml.ista.ac.at/AwA/

Datasets	Methods	PER (%)					
		0%	10%	30%	50%	70%	90%
Caltech101-20	SLIM	66.82 (1.73)	65.13 (2.91)	60.94 (1.34)	57.15 (2.31)	54.69 (2.00)	50.36 (1.22)
	AMSC	70.79 (1.70)	71.34 (2.35)	69.56 (1.73)	67.72 (1.76)	67.78 (3.28)	63.60 (1.78)
	AMMSS	79.23 (1.49)	75.92 (2.13)	75.52 (1.75)	72.85 (2.37)	70.08 (3.45)	65.46 (5.16)
	CPM-Nets	82.94 (3.53)	81.76 (2.03)	78.63 (3.72)	61.74 (4.06)	66.42 (5.28)	60.83 (6.79)
	DCP	81.99 (2.74)	83.37 (1.98)	82.89 (1.52)	79.39 (3.44)	79.00 (1.92)	73.85 (3.26)
	TMC	80.63 (1.43)	79.41 (1.43)	76.63 (2.09)	76.28 (2.80)	73.51 (2.46)	73.12 (2.81)
	DIMvLN	**91.28 (1.11)**	**90.40 (1.38)**	**88.49 (1.03)**	**87.32 (1.96)**	**83.89 (2.17)**	**85.23 (1.77)**
CUB	SLIM	48.58 (7.89)	43.00 (7.25)	41.42 (6.72)	38.83 (6.02)	37.92 (3.75)	36.92 (4.27)
	AMSC	58.42 (6.52)	63.17 (6.63)	58.50 (6.36)	60.50 (6.72)	53.42 (4.93)	49.92 (5.41)
	AMMSS	65.33 (9.11)	62.50 (8.14)	52.83 (5.76)	54.00 (5.98)	49.67 (8.77)	43.50 (5.06)
	CPM-Nets	85.25 (3.03)	81.75 (3.02)	70.17 (2.44)	57.08 (6.16)	56.75 (5.22)	56.58 (3.49)
	DCP	79.50 (4.79)	77.92 (6.46)	74.17 (5.65)	53.25 (8.17)	57.33 (4.76)	56.50 (6.32)
	TMC	73.25 (3.85)	71.50 (5.94)	65.25 (6.68)	59.25 (5.39)	57.33 (4.77)	46.92 (8.86)
	DIMvLN	**89.17 (2.84)**	**87.25 (2.94)**	**83.08 (2.84)**	**79.67 (2.67)**	**77.25 (3.78)**	**72.58 (4.49)**
Wikipedia	SLIM	50.86 (5.22)	48.99 (2.85)	47.05 (4.14)	44.82 (5.09)	39.57 (4.18)	35.18 (4.59)
	AMSC	57.48 (5.36)	55.83 (2.74)	51.15 (4.39)	45.61 (4.10)	40.86 (5.41)	35.11 (5.59)
	AMMSS	60.43 (5.41)	56.47 (1.80)	39.21 (13.13)	45.32 (7.11)	39.86 (7.91)	35.25 (4.15)
	CPM-Nets	63.85 (3.89)	59.79 (4.75)	48.39 (4.21)	38.39 (4.00)	38.74 (3.13)	38.74 (3.33)
	DCP	55.90 (6.00)	55.11 (3.27)	48.49 (3.63)	44.82 (3.90)	42.01 (5.52)	36.47 (4.41)
	TMC	21.80 (6.70)	22.23 (5.20)	21.29 (4.77)	21.22 (4.39)	18.35 (3.29)	16.69 (3.54)
	DIMvLN	**69.78 (2.86)**	**67.27 (2.41)**	**60.36 (4.29)**	**55.25 (2.63)**	**48.92 (3.95)**	**44.60 (5.47)**
ALOI	SLIM	54.04 (3.92)	51.33 (2.02)	47.29 (3.37)	44.13 (3.22)	40.07 (3.26)	34.62 (1.03)
	AMSC	75.87 (1.23)	72.45 (1.84)	69.19 (1.65)	64.54 (1.91)	60.00 (1.96)	54.49 (1.17)
	AMMSS	85.36 (1.11)	84.04 (0.79)	81.27 (1.05)	77.83 (0.65)	74.41 (0.90)	69.52 (1.03)
	CPM-Nets	54.00 (2.36)	47.89 (3.69)	36.65 (3.38)	24.77 (1.48)	15.36 (0.89)	12.17 (1.04)
	DCP	38.00 (7.25)	37.15 (4.08)	34.00 (3.05)	26.14 (3.87)	23.20 (3.45)	19.00 (1.81)
	TMC	16.10 (2.68)	15.78 (1.61)	12.90 (0.86)	10.93 (1.65)	8.76 (1.54)	6.43 (1.68)
	DIMvLN	**97.44 (0.41)**	**93.91 (0.54)**	**89.70 (1.23)**	**85.97 (1.29)**	**81.96 (1.04)**	**76.96 (1.38)**
Out-Scene	SLIM	63.27 (3.10)	61.90 (2.81)	59.83 (2.11)	56.38 (2.04)	51.15 (2.31)	45.50 (3.00)
	AMSC	65.67 (2.68)	62.34 (5.61)	59.65 (5.49)	58.36 (5.15)	59.26 (2.29)	53.66 (2.98)
	AMMSS	67.90 (3.45)	66.08 (5.10)	64.09 (4.48)	63.12 (3.80)	59.70 (2.25)	56.80 (2.14)
	CPM-Nets	63.91 (7.11)	62.81 (3.87)	61.57 (5.06)	57.06 (3.95)	43.20 (3.45)	33.50 (3.25)
	DCP	77.04 (2.21)	75.35 (2.30)	73.62 (0.97)	68.62 (2.84)	59.67 (1.78)	52.79 (2.51)
	TMC	61.86 (4.47)	57.73 (6.73)	54.70 (5.12)	52.32 (4.29)	47.29 (2.65)	41.99 (1.61)
	DIMvLN	**84.31 (0.72)**	**82.53 (1.37)**	**80.61 (1.47)**	**76.78 (1.59)**	**74.14 (1.57)**	**71.75 (1.79)**
Animal	SLIM	28.24 (1.18)	26.92 (1.97)	25.63 (1.53)	27.57 (0.94)	21.19 (1.38)	18.64 (0.89)
	AMSC	20.44 (2.57)	21.23 (2.42)	20.43 (1.17)	19.52 (1.39)	19.34 (1.29)	17.47 (1.39)
	AMMSS	20.90 (1.74)	18.50 (1.34)	17.65 (0.99)	16.73 (1.01)	16.00 (1.34)	14.66 (0.66)
	CPM-Nets	23.87 (1.25)	24.31 (1.11)	19.72 (3.78)	21.39 (3.06)	17.06 (2.82)	10.09 (1.21)
	DCP	34.88 (0.67)	33.61 (1.02)	30.73 (1.10)	26.48 (1.54)	21.50 (2.55)	17.46 (1.31)
	TMC	27.88 (1.46)	27.77 (0.96)	24.10 (1.53)	23.38 (0.72)	22.09 (1.35)	20.93 (1.50)
	DIMvLN	**39.82 (5.28)**	**40.36 (2.11)**	**38.12 (2.70)**	**34.61 (3.19)**	**32.67 (2.97)**	**28.11 (5.30)**

Table 1: ACC (%) comparisons on six datasets while PER is selected in {0%,10%,30%,50%,70%,90%} and LER is fixed to 10%. Standard deviation (%) is in parentheses. The best/second-best results are marked in bold/underline, respectively.

Ablation Study and Parameter Sensitivity. The ablation experiments on ALOI and Animal are conducted to deeply investigate the effect of the two crucial modules of DIMvLN, i.e., GNN-based completion module and semi-supervised learning module. Note that we retain the multi-view representation module since representation extraction is an integral part of multi-view learning. When the completion module is removed, we use the average strategy to fill the missing data. And the semi-supervised learning module is removed by excluding the pseudo-labeling mechanism. The ablation results are listed in Table 2. We can observe that: 1) When both modules are used, the model can obtain the highest performance, verifying the effectiveness of DIMvLN. 2) The semi-supervised module plays a crucial role in performance improvement, which indicates that it can guide the completion module of DIMvLN to exploit the extrinsic semantic information effectively.

Besides, we have two parameters in our DIMvLN, i.e., λ_1 and λ_2. We conduct experiments on Caltech101-20 and Out-Scene, where PER and LER are fixed to 50% and 10%,



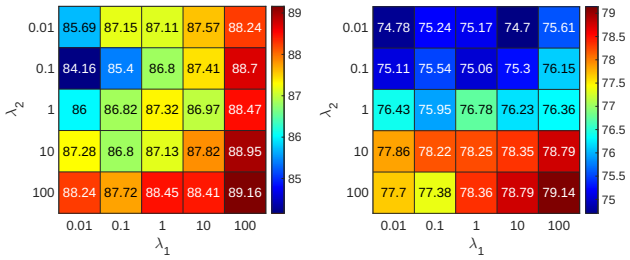
(a) Caltech101-20 (b) CUB (c) Wikipedia

(d) ALOI (e) Out-Scene (f) Animal

Figure 2: ACC (%) comparisons on six datasets with LER varying from 5% to 35% with a gap of 5% while PER=50%.

GNN Completion	Semi-supervised	ALOI ACC	ALOI F1	Animal ACC	Animal F1
✗	✗	53.08	52.98	19.22	16.04
✓	✗	55.44	54.81	20.37	17.15
✗	✓	82.04	82.82	31.50	27.23
✓	✓	**83.74**	**83.89**	**33.86**	**28.23**

Table 2: Ablation study on the ALOI and Animal datasets with PER=60% and LER=10%. '✓' and '✗' represent the used and not used module, respectively.



(a) Caltech101-20 (b) Out-Scene

Figure 3: Parameter analysis of the trade-off parameters λ_1 and λ_2 on Caltech101-20 and Out-Scene.

to analyze the sensitivity of these two parameters. To show the parameter influence more previously, we select two parameters both from the range of $\{0.01, 0.1, 1, 10, 100\}$. The results are reported in Fig. 3, and we can observe that DIMvLN performs well if these two parameters are selected from the predefined ranges. It means that our proposed DIMvLN is not so sensitive to both parameters.

Conclusions

In this paper, to solve the dual incomplete problem, we propose a novel deep absent multi-view semi-supervised method named DIMvLN. DIMvLN simultaneously incorporates the GNN-based completion and semi-supervised learning to recover the missing feature information and exploit the extrinsic label information. Therefore, our method is able to eliminate the negative influence caused by the incomplete data and insufficient labels and enhance its performance. Finally, extensive experimental results on six popular datasets show the effectiveness and superiority of DIMvLN. In the future, the meta-learning approach can be introduced to adaptively determine the optimal hyperparameters in DIMvLN. Besides, we will further extend our DIMvLN to solve the problems of the incomplete multi-view semi-supervised classification scenario, such as class-imbalance, noisy labels, and novel class discovery.

Acknowledgments

This work was supported by the National Science Foundation of China Grant [62036013, 62136005, 62376281, 62306171], and the NSF for Huxiang Young Talents Program of Hunan Province under Grant [2021RC3070].

References

Andrew, G.; Arora, R.; Bilmes, J.; and Livescu, K. 2013. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, 1247–1255.

Cai, X.; Nie, F.; Cai, W.; and Huang, H. 2013. Heterogeneous image features integration via multi-modal semi-supervised learning model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1737–1744.

Chen, X.; Chen, S.; Xue, H.; and Zhou, X. 2012. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data. *Pattern Recognition*, 45(5): 2005–2018.

Chen, X.; Ma, H.; Wan, J.; Li, B.; and Xia, T. 2017. Multi-view 3d object detection network for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1907–1915.

Cheng, Y.; Zhao, X.; Cai, R.; Li, Z.; Huang, K.; and Rui, Y. 2016. Semi-supervised multimodal deep learning for RGB-D object recognition. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 3345–3351.

Gong, C.; Tao, D.; Maybank, S. J.; Liu, W.; Kang, G.; and Yang, J. 2016. Multi-modal curriculum learning for semi-supervised image classification. *IEEE Transactions on Image Processing*, 25(7): 3249–3260.

Gray, K. R.; Aljabar, P.; Heckemann, R. A.; Hammers, A.; Rueckert, D.; Initiative, A. D. N.; et al. 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, 65: 167–175.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2023. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2551–2566.

Hu, M.; and Chen, S. 2018. Doubly aligned incomplete multi-view clustering. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2262–2268.

Huang, D.; Wang, C.-D.; and Lai, J.-H. 2023. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11388–11402.

Ji, X.; Henriques, J. F.; and Vedaldi, A. 2019. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9865–9874.

Jia, X.; Jing, X.-Y.; Zhu, X.; Chen, S.; Du, B.; Cai, Z.; He, Z.; and Yue, D. 2021. Semi-supervised multi-view deep discriminant representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7): 2496–2509.

Jiang, Q.-Y.; and Li, W.-J. 2017. Deep cross-modal hashing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3270–3278.

Jing, X.-Y.; Wu, F.; Dong, X.; Shan, S.; and Chen, S. 2017. Semi-supervised multi-view correlation feature learning with application to webpage classification. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 1374–1381.

Kan, M.; Shan, S.; and Chen, X. 2016. Multi-view deep network for cross-view classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4847–4855.

Li, S.-Y.; Jiang, Y.; and Zhou, Z.-H. 2014. Partial multi-view clustering. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 1968–1974.

Liang, W.; Li, Y.; Xie, K.; Zhang, D.; Li, K.-C.; Souri, A.; and Li, K. 2023. Spatial-temporal aware inductive graph neural network for C-ITS data recovery. *IEEE Transactions on Intelligent Transportation Systems*, 24(8): 8431–8442.

Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2022. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461.

Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; and Zhu, E. 2021. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2634–2646.

Nie, F.; Cai, G.; Li, J.; and Li, X. 2018. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 27(3): 1501–1511.

Sato, R. 2023. Graph neural networks can recover the hidden features solely from the graph structure. In *Proceedings of the 40th International Conference on Machine Learning*, 30062–30079.

Shao, W.; He, L.; and Yu, P. S. 2015. Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 318–334.

Shorten, C.; and Khoshgoftaar, T. M. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1): 1–48.

Wang, H.; Yang, Y.; and Li, T. 2016. Multi-view clustering via concept factorization with local manifold regularization. In *Proceedings of the IEEE 16th International Conference on Data Mining*, 1245–1250.

Wang, J.; Zhang, L.; Wang, Q.; Chen, L.; Shi, J.; Chen, X.; Li, Z.; and Shen, D. 2020. Multi-class ASD classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation. *IEEE Transactions on Medical Imaging*, 39(10): 3137–3147.

Wang, S.; Liu, X.; Liu, L.; Tu, W.; Zhu, X.; Liu, J.; Zhou, S.; and Zhu, E. 2022a. Highly-efficient incomplete largescale

multiview clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9766–9775.

Wang, Y.; Chang, D.; Fu, Z.; Wen, J.; and Zhao, Y. 2022b. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1): 367–378.

Wen, J.; Xu, Y.; and Liu, H. 2020. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics*, 50(4): 1418–1429.

Wu, F.; Jing, X.-Y.; Zhou, J.; Ji, Y.; Lan, C.; Huang, Q.; and Wang, R. 2019. Semi-supervised multi-view individual and sharable feature learning for webpage classification. In *Proceedings of the World Wide Web Conference*, 3349–3355.

Xia, Y.; Liu, F.; Yang, D.; Cai, J.; Yu, L.; Zhu, Z.; Xu, D.; Yuille, A.; and Roth, H. 2020. 3D semi-supervised learning with uncertainty-aware multi-view co-training. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 3635–3644.

Yang, Y.; Zhan, D.-C.; Sheng, X.-R.; and Jiang, Y. 2018. Semi-supervised multi-modal learning with incomplete modalities. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2998–3004.

Zhang, C.; Han, Z.; Cui, Y.; Fu, H.; Zhou, J. T.; and Hu, Q. 2019. CPM-Nets: Cross partial multi-view networks. In *Advances in Neural Information Processing Systems*, 559–569.

Zhao, H.; Liu, H.; and Fu, Y. 2016. Incomplete multimodal visual data grouping. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2392–2398.

Zhao, J.; Xie, X.; Xu, X.; and Sun, S. 2017. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38: 43–54.

Zhuge, W.; Luo, T.; Fan, R.; Tao, H.; Hou, C.; and Yi, D. 2023. Absent multiview semisupervised classification. *IEEE Transactions on Cybernetics*, 1–14.