Decouple then Classify: A Dynamic Multi-view Labeling Strategy with Shared and Specific Information

Xinhang Wan [1] **Jiyuan Liu** [2] **Xinwang Liu** [1] **Yi Wen** [1] **Hao Yu** [1] **Siwei Wang** [3] **Shengju Yu** [1] **Tianjiao Wan** [1] **Jun Wang** [1] **En Zhu** [1]

Abstract

Sample labeling is the most primary and fundamental step of semi-supervised learning. In literature, most existing methods randomly label samples with a given ratio, but achieve unpromising and unstable results due to the randomness, especially in multi-view settings. To address this issue, we propose a Dynamic Multi-view Labeling Strategy with Shared and Specific Information. To be brief, by building two classifiers with existing labels to utilize decoupled shared and specific information, we select the samples of low classification confidence and label them in high priorities. The newly generated labels are also integrated to update the classifiers adaptively. The two processes are executed alternatively until a satisfying classification performance. To validate the effectiveness of the proposed method, we conduct extensive experiments on popular benchmarks, achieving promising performance. The code is publicly available at `https://github.com/wanxinhang/ICML2024_decouple_then_classify`.

1. Introduction

With the rapid development of information techniques, millions of unlabeled data are generated and collected. How to label them for the downstream tasks is vital in many fields (Wan et al., 2024b; Tu et al., 2023; Wan et al., 2024a; Liang et al., 2023; Liu et al., 2022). For instance, it is beneficial to recommend items personally for different types of users in the recommendation system. However, labeling them is challenging and expensive because the number of users is vast. In light of this, semi-supervised learning, which aims to partition data into several groups with limited labeled data, has been proposed recently and achieved great success (Cai et al., 2020; Holmes et al., 2021). Given an unlabeled dataset, most existing semi-supervised methods randomly select samples to label and train the model with them. However, for a given label ratio, the results frequently fluctuate a lot when the model is trained on different annotated instances. The reason is that samples contain different information and contribute variously to the model. Therefore, selecting samples that convey richer information to annotate is significant to reduce the label cost and attain better performance (Ren et al., 2021; Cao et al., 2020).

Active learning addresses the abovementioned problem by interactively selecting valuable samples from the unlabeled dataset to label and retraining the model (Wang et al., 2022a; He et al., 2023b;a). Nevertheless, most focus on data from a single source and overlook multi-view data. Multi-view data is widespread in the world with multimedia improves by leaps and bounds; for instance, a person can be described by his appearance, social networks, sound, etc. (Liu et al., 2021a; Wang et al., 2022c; Liang et al., 2022; Wan et al., 2023; Liu et al., 2024). How to label multi-view data and reduce the label cost is crucial to the analysis and utilization of multi-view data. Unfortunately, annotating data from multiple sources is challenging since the views usually contain two pieces of information, i.e., the shared and specific information (SSI), the efficient utilization of them and labeling the data is more complicated than labeling single-view data. Some samples might be easier to classify via shared information or the opposite. Existing semi-supervised multi-view (SSMV) methods frequently concentrate on the shared or specific information individually or get trouble decoupling and utilizing them (Shen et al., 2023; Xu et al., 2022). Despite some methods gaining two pieces of information adequately, the random selection of instances to annotate leads to unstable results (Wang et al., 2021a; Wu et al., 2023a).

To this end, we propose a Dynamic Multi-view Labeling Strategy with Shared and Specific Information (DMVLS) to handle these problems. Concretely, we attain SSI via auto-

[1]College of Computer, National University of Defense Technology, Changsha, China [2]Company Name, Location, Country [3]School of ZZZ, Institute of WWW, Location, Country. Correspondence to: Firstname1 Lastname1 <first1.last1@xxx.edu>, Firstname2 Lastname2 <first2.last2@www.uk>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

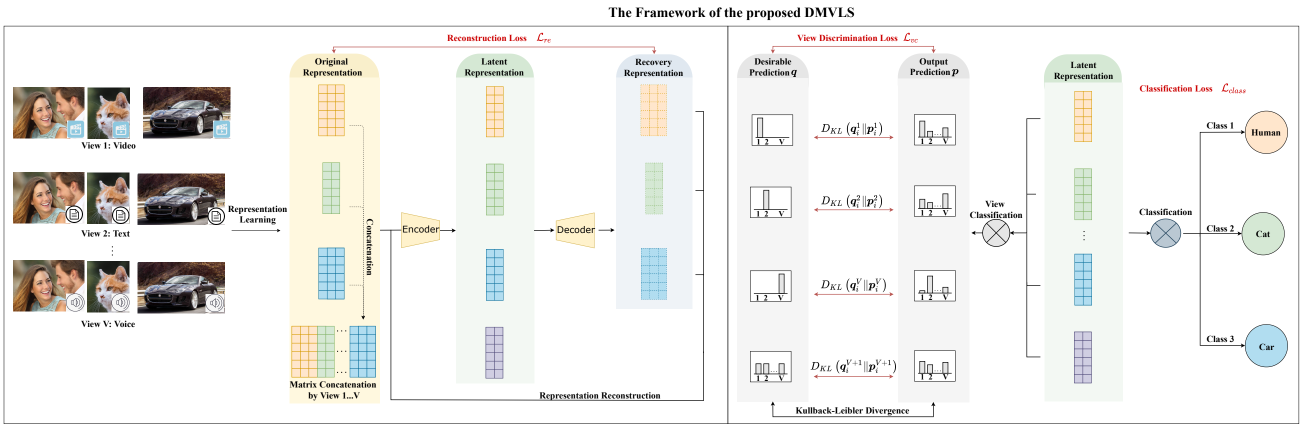

Figure 1. The basic framework of our proposed algorithm. First, we extract the shared and specific information (SSI) with the auto-encoder and decouple it with a view discriminator. Then, we utilize the two pieces of information to train two classifiers separately with existing labels. We interactively train the model and use it to select valuable samples to label until a promising classifier is achieved.

encoders to ensure the quality of the extracted knowledge. Meanwhile, a view discriminator is leveraged to conduct a view classification task to decouple them adequately. After attaining decoupled and high-quality information, we equip each piece of information with a classifier to attain the predicted label since some samples might be more accessible to predict via shared information or the opposite. In summary, our loss consists of three parts, i.e., the reconstruction loss, the view discrimination loss, and the classification loss. We first randomly select some samples as cold start to train the model. Then, for the rest of the unlabeled samples, if a sample is hard to predict via both SSI, it is regarded as an uncertain sample and needs to be labeled. Pseudo labels are generated with samples of high confidence and consistent results in the two classifiers. We interactively select samples to label and retrain the model with the newly generated labels to boost the classification performance. The framework of our proposed method is shown in Fig. 1. Our contributions are as follows:

1. We propose an efficient algorithm to tackle the sample labeling task in semi-supervised multi-view learning, termed the Dynamic Multi-view Labeling Strategy with Shared and Specific Information. The samples of low classification confidence are labeled as high priorities. Meanwhile, pseudo labels are generated with samples of high confidence and consistent results in the two classifiers.

2. We utilize a view discriminator to decouple the shared and specific information extracted via the encoder. To further ensure the generated information with high quality, the decoder is used to recover the original matrices and minimize the reconstruction loss.

3. To validate the effectiveness of the proposed method, we conduct extensive experiments on popular benchmarks, achieving promising performance.

2. Related Work

In this section, we introduce the work most related to our method, including semi-supervised multi-view classification and active learning.

2.1. Semi-supervised Multi-view Classification

Semi-supervised classification, which aims to partition samples into several categories with limited labeled samples and a considerable number of unlabeled samples, has achieved great attention in many fields such as anomaly detection, computer vision, medical diagnosis, etc (Xu et al., 2021; Zhmoginov et al., 2022). The early methods frequently pay attention to scenarios where data is collected from a single source. For instance, the authors in (Blum & Mitchell, 1998) train two predictors, and then each predictor generates pseudo labels as the training samples for the other to train. The process ends when all the unlabeled samples are annotated with pseudo labels. The emerging graph-based methods utilize label propagation with a graph structure, most of which are under the framework of GCN (Kipf & Welling, 2016; Wu et al., 2023b). GCNI overcomes the over-smoothing of GCN with a vanilla GCN model. Then, to enrich the supervision signals (Chen et al., 2020), CG^3 utilizes both data similarities and graph structure to improve the performance of GCN (Wan et al., 2021).

The burgeoning multi-view data benefits the development of semi-supervised multi-view classification (SSMVC) (Wang et al., 2021b; Huang et al., 2021b). MV-GCN (Yuan et al., 2021) integrates multiple graphs to fuse node representations by contrastive learning. MMatch (Wang et al., 2022d) achieves multi-view consistent classification among multiple views. SMDDRL combines deep metric learning and density clustering to generate pseudo labels for unlabeled samples (Jia et al., 2020). Despite current methods improving SSMVC from diverse perspectives, all of them randomly label samples with a given label ratio, resulting in unpromising and unstable results.

2.2. Active Learning

Active learning is proposed to reduce label costs and has been studied recently. The main issue of active learning is the design of the selector, which selects new data points to label from the unlabeled samples at each iteration (Zhou et al., 2021; Kothawade et al., 2021). It can be roughly divided into three types, including uncertainty (Nguyen et al., 2022), model influence (Liu et al., 2021b), and the sample distribution methods (Liu et al., 2021c). Among them, the most popular one is the uncertainty method, which can be further categorized into marginal sampling(Ducoffe & Precioso, 2018), best-versus-second best (BvSB) (Joshi et al., 2009), maximum confidence uncertainty (MCU) (Chen et al., 2018). MBAL selects instances based on a global margin or a combination of the margin of local classifiers (Roth & Small, 2006). MALF utilizes (AL-Sammarraie & Karaca, 2023) BvSB to select valuable samples and conducts dimension reduction as a pre-processing. MLAL designs a multi-label technique to take advantage of the correlations of speed measurements with MCU (Bellarmino et al., 2023). MVSS-AL extends active learning in the multi-view domain to handle hyperspectral image classification (Yu et al., 2022). M3L focuses on the multi-label task of multi-view multi-instance data (Yu et al., 2022) and selects the most informative bag-label pair for the query.

Despite some active learning methods being combined with multi-view applications, all of them fail to decouple and utilize shared and specific information simultaneously. To overcome this, we propose A Dynamic Multi-view Labeling Strategy with Shared and Specific Information (DMVLS) in the next section.

3. Methodology

The methodology of DMVLS is introduced in this section. We first give the notation summary of our method, then summarize the motivation and challenge. After that, the design of our method is provided.

3.1. Notation Summary

In our paper, the multi-view data with n samples is denoted as $\{\mathbf{X}^v\}_{v=1}^{V}$, where $\mathbf{X}^v \in \mathbf{R}^{d_v \times n}$, d_v represents the feature dimension of v-th view, and V is the view number. $\overline{\mathbf{X}} \in \mathbf{R}^{d \times n}$ is the is the concatenated feature and $d = \sum_{v=1}^{V} d_v$. The labeling budget of each iteration is set as B, which indicates that we select B samples each round to label. In the initial case, we randomly select some samples $\mathbf{X}_0^L$ to label and their label set is $\mathbf{Y}_0^L$. Then for the t-th round, we select the valuable samples to label based on the labeled sample set $\mathbf{X}_0^L \cup \mathbf{X}_1^L \cup \cdots \cup \mathbf{X}_{t-1}^L$ and the unlabeled sample set $\mathbf{X} - \mathbf{X}_0^L \cup \mathbf{X}_1^L \cup \cdots \cup \mathbf{X}_{t-1}^L$.

3.2. Motivation

Most SSMVC methods adopt GCN to attain the predicted labels, but they heavily rely on the input graph, and the space complexity is high, which is unsuitable for large-scale situations. Also, existing methods neglect SSI; some samples are more easily distinguished via shared or specific information, so decoupling the SSI and utilizing them subsequently is essential. How to choose the informative samples to annotate is another issue that needs to be addressed. The most popular method is random selection, but it results in unstable results, annotating valuable samples is beneficial to reduce the label cost and improve the classification accuracy. Therefore, it is necessary to propose a method to attain decoupled and high-quality SSI and utilize them to attain the classification results with valuable labeled data.

3.3. The extraction of SSI

Many multi-view methods attempt to attain high-quality SSI to conduct the downstream tasks. However, the traditional methods are trapped into a shallow model and fail to extract adequate knowledge (Luo et al., 2018b; Zhou et al., 2020), while deep models fail to provide the representation power with theoretical proof (Xia et al., 2021). Fortunately, the auto-encoder (AE) could encode the original feature into a low-dimensional latent space with the representative capacity of neural networks and decode it to fit the raw data. In our framework, we adopt it to extract SSI among views.

$V + 1$ AEs with four layers are utilized to extract the SSI from multi-view data. Considering that the obtained SSI is utilized to conduct the downstream classification task and reconstruct the raw features, we impose two layers with shared parameters among the AEs to facilitate the subsequent process. Let $\mathbf{z}_i^{0,v} = \mathbf{x}_i^v$ and $\mathbf{z}_i^{0,V+1} = \overline{\mathbf{x}}_i$, for the extraction of SSI, we attain it via the following formula:

$$\mathbf{z}_i^{1,v} = \sigma\left(\mathbf{W}_{ae}^{1,v}\mathbf{z}_i^{0,v} + \mathbf{b}_{ae}^{1,v}\right), \tag{1}$$

$$\mathbf{z}_i^{2,v} = \sigma\left(\mathbf{W}_{ae}^{2}\mathbf{z}_i^{1,v} + \mathbf{b}_{ae}^{2}\right). \tag{2}$$

where $1 \leq v \leq V+1$, $\mathbf{W}_{ae}^{1,v} \in \mathbf{R}^{d_{ae}^1 \times d_v}$, $\mathbf{W}_{ae}^{2} \in \mathbf{R}^{d_{ae}^2 \times d_{ae}^1}$, $\mathbf{b}_{ae}^{1,v} \in \mathbf{R}^{d_{ae}^1}$, and $\mathbf{b}_{ae}^{2} \in \mathbf{R}^{d_{ae}^2}$ are the weights and bias of corresponding layers.

Given that each view contains two pieces of information, i.e., SSI, we first combine them to attain the comprehensive knowledge of each view, then reconstruct the original features as follows:

$$\mathbf{z}_i^{3,v} = \sigma\left(\mathbf{W}_{ae}^{3}\left(\mathbf{z}_i^{2,v} + \mathbf{z}_i^{2,V+1}\right) + \mathbf{b}_{ae}^{3}\right), \tag{3}$$

$$\mathbf{z}_i^{4,v} = \sigma\left(\mathbf{W}_{ae}^{4,v}\mathbf{z}_i^{3,v} + \mathbf{b}_{ae}^{4,v}\right). \tag{4}$$

where $1 \leq v \leq V$.

To ensure the quality of SSI, we minimize the reconstruction loss as:

$$\mathcal{L}_{re} = \sum_{v=1}^{V}\sum_{i=1}^{n}\left\|\boldsymbol{x}_i^{v} - \mathbf{z}_i^{4,v}\right\|_2^2. \tag{5}$$

3.4. The decoupling of SSI

Most existing multi-view learning methods attempt to attain SSI, but they frequently neglect to decouple them, which inevitably affects the subsequent utilization of SSI (Luo et al., 2018a). Inspired by (Jia et al., 2020), we develop a view discriminator to separate them, and the input of the view discriminator is $\left\{\mathbf{Z}^{2,v}\right\}_{v=1}^{V+1}$.

Considering that the specific information is extracted from an individual view, it ought to be irrelevant to other views. Given a view classifier, it is expected to classify the specific information into the view to which it belongs. On the contrary, the shared information is obtained from all views, the view classifier tends to generate an indiscriminate result. We leverage a 3-layer network to design the view classifier. For the extracted SSI, we attain the result of the view discriminator as

$$\boldsymbol{r}_i^{l,v} = \sigma\left(\mathbf{W}_{vc}^{l}\mathbf{r}_i^{l-1,v} + \mathbf{b}_{vc}^{l}\right), \tag{6}$$

where $1 \leq v \leq V+1$, $1 \leq l \leq 3$, $\mathbf{r}_i^{0,v} = \mathbf{z}_i^{2,v}$.

We adopt Kullback-Leibler divergence D_{KL} to measure the loss of view classification. For the view-specific information, the loss is:

$$\mathcal{L}_{vc1} = \sum_{v=1}^{V}\sum_{i=1}^{n} D_{KL}\left(\boldsymbol{q}_i^{v} \| \boldsymbol{p}_i^{v}\right), \tag{7}$$

where $\boldsymbol{q}_i^{v}$ is a one-hot vector and the v-th element equal to 1 and other elements are zero, $\boldsymbol{p}_i^{v} = softmax(\boldsymbol{r}_i^{3,v})$.

The view discrimination loss of shared information is calculated via:

$$\mathcal{L}_{vc2} = \sum_{i=1}^{n} D_{KL}\left(\boldsymbol{q}_i^{V+1} \| \boldsymbol{p}_i^{V+1}\right), \tag{8}$$

where $\boldsymbol{q}_i^{V+1}$ is a V-dimensional vector and each element is $1/V$.

The total loss of view discriminator is:

$$\mathcal{L}_{vc} = \mathcal{L}_{vc1} + \mathcal{L}_{vc2}. \tag{9}$$

3.5. The utilization of SSI

After attaining the disassociated SSI, how to utilize it to conduct the subsequent semi-supervised classification task is significant. In a multi-view classification task, some samples might be easier to classify via specific knowledge or the opposite. Unlike current methods that concatenate SSI to train the classifier, we leverage two pieces of information to train the classifiers separately, which benefits the subsequent sample selection in the next section.

Two three-layer classifiers are adopted to train the SSI of labeled data individually, and each is equipped with the same dimension in each layer. In Section 3.3, we add SSI to attain comprehensive knowledge for each view. Therefore, for the classification of the specific information, we first sum the specific information in each view to attain a comprehensive understanding as:

$$\mathbf{Z}_{spe} = \sum_{v=1}^{V}\mathbf{Z}^{2,v}. \tag{10}$$

Then, we obtain the classification result concerning specific information via:

$$\mathbf{H}_{spe}^{l} = \sigma\left(\mathbf{W}_{spe}^{l}\mathbf{H}_{spe}^{l-1} + \mathbf{b}_{spe}^{l}\right), \tag{11}$$

where $\mathbf{H}_{spe}^{0} = \mathbf{Z}_{spe}$, $\mathbf{W}_{spe}^{1} \in \mathbf{R}^{d_{spe}^{1} \times d_{ae}^{2}}$.

For the shared information, we furnish the classifier with the same structure of specific information to obtain $\mathbf{H}_{share}^{3}$.

Then, the classification loss is calculated as follows:

$$\mathcal{L}_{class1} = \sum_{i=1}^{n_l} D_{KL}\left(\boldsymbol{y}_i \| \boldsymbol{p}_i^{spe}\right), \tag{12}$$

and

$$\mathcal{L}_{class2} = \sum_{i=1}^{n_l} D_{KL}\left(\boldsymbol{y}_i \| \boldsymbol{p}_i^{share}\right), \tag{13}$$

where n_l denotes the number of existing labeled data, $\boldsymbol{y}_i$ represents the label indicator matrix, $\boldsymbol{p}_i^{spe} = softmax(\mathbf{H}_{spe}^{3})$, $\boldsymbol{p}_i^{share} = softmax(\mathbf{H}_{share}^{3})$.

The final classification loss is the sum of them:

$$\mathcal{L}_{class} = \mathcal{L}_{class1} + \mathcal{L}_{class2}. \tag{14}$$

3.6. The Loss Function

In summary, the total loss of our framework comprises three parts: the reconstruction loss, view discrimination loss, and the classification loss. The quality of the extracted SSI is ensured via reconstruction loss, and the high-quality SSI is decoupled and utilized separately via view discrimination loss and classification loss. The loss function of our proposed method is as follows:

$$\mathcal{L} = \mathcal{L}_{re} + \lambda_1 \mathcal{L}_{vc} + \lambda_2 \mathcal{L}_{class}. \tag{15}$$

where λ_1 and λ_2 are the hyper-parameters to balance the losses.

3.7. Sample Labeling with Uncertainty

Sample labeling is the most primary and fundamental step of semi-supervised learning. However, current SSMVC methods often neglect the significance of sample labeling and randomly select samples to label, resulting in unpromising and unstable results (Zhang et al., 2020; Wang et al., 2022e). Inspired by active learning (AL), we interactively annotate highly uncertain samples from the unlabeled sample set until the classification performance is promising.

Given that two classifiers are leveraged to classify the samples, how to evaluate the uncertainty of each sample is different from the existing AL methods. In our scenario, we first train the classifiers via Eq. (15) and utilize it to predict the labels of the unlabeled data. Suppose a sample is challenging to predict via specific information and shared knowledge. In that case, it is regarded as a valuable sample for clarifying the decision boundary, and we annotate it and put it into the labeled set. The mathematical formulation is written as:

$$s_i = 1 - \max_{1 \leq k \leq K} \boldsymbol{p}_i^{spe} * \max_{1 \leq k \leq K} \boldsymbol{p}_i^{share}. \tag{16}$$

where s_i is the score of i-th sample, and K denotes the number of label categories.

After training classifiers with existing labels, we score the unlabeled samples via Eq.(16). Samples with higher scores are more uncertain, and we label the top B samples.

Conversely, a sample with the consistent predicted label in two classifiers of high confidence is regarded as a reliable sample. Therefore, we generate the pseudo label and add it to the labeled sample set with its pseudo label.

We alternately train the model and utilize it to choose samples to label to boost the classification performance. The whole process of our proposed method is summarized in Algorithm 1.

Additionally, compared with existing semi-supervised multi-view classification methods, our proposed method possesses the following merits:

- Sample selection. Unlike existing methods that randomly select samples to annotate, we interactively train the model with existing labels and utilize the trained model to label the samples of low confidence.
- Scalability. Compared with the graph-based methods, our method is under an auto-encoder framework; the space complexity is linear, respecting sample number, which is more suitable for large-scale data.
- High-quality representation. We extract the SSI in a latent space and reconstruct the raw features to ensure its quality with the auto-encoder. Meanwhile, a view discriminator is leveraged to decouple SSI for subsequent utilization.
- Reliability. Given that in multi-view classification, some samples are more accessible to classify via shared information or the opposite. We utilize decoupled SSI to train the classifiers separately, then leverage them to select the unreliable samples to label and reliable instances to generate the pseudo labels.

Algorithm 1 A Dynamic Multi-view Labeling Strategy with Shared and Specific Information.

Input: Multi-view dataset $\{\mathbf{X}^v\}_{v=1}^V$, initial labeled set $\mathbf{X}_0^l$, the maximize labeled number n_{max}, labeled number of each round B, hyper-parameters λ_1, λ_2, learning rate α, epoch number M.
Output: The predicted labels of the unlabeled sample set.
1: Initialize the network parameters Θ in the AEs, view discriminator, and the classifiers, $T = \frac{n_{max}}{B}$.
2: **for** $t = 1$ to T **do**
3:   **for** $m = 1$ to M **do**
4:     Obtain the outputs $\mathbf{Z}^{4,v}$, $\boldsymbol{P}^v$, $\boldsymbol{P}^{spe}$, $\boldsymbol{P}^{share}$.
5:     Compute the loss value via Eq. (5), (9), (14), and (15).
6:     Update Θ via back propagation.
7:   **end for**
8:   Obtain the output $\boldsymbol{p}_i^{spe}$, $\boldsymbol{p}_i^{share}$ of unlabeled samples.
9:   Calculate the scores of unlabeled samples via Eq. (16), then label the top B samples, add them to the labeled sample set $\mathbf{X}_t^l$.
10:   Generate the pseudo labels of high confidence and consistent results in the two classifiers, then add them to the labeled sample set $\mathbf{X}_t^l$.
11: **end for**
12: The final predicted label of the unlabeled sample i is $\arg\max_{1 \leq k \leq K} p_i^k$. ($\boldsymbol{p}_i = \max\left(\boldsymbol{p}_i^{spe}, \boldsymbol{p}_i^{share}\right)$).

4. Experiments

In this section, we first conduct comprehensive experiments on six widespread datasets to show the superiority of our proposed method over several state-of-the-art semi-supervised multi-view classification methods. After that, we record the classification performance changes with the selected sample number. Then, we conduct an ablation study to demonstrate the effectiveness of critical components of our framework. After that, we visualize the learned representation.

4.1. Experimental Setup

4.1.1. DATASETS

Six widespread datasets are used in our experiments, including Handwritten, BDGP, Cora, CiteSeer, STL10, and YTB10. The number of samples ranges from 2000 to 38654. The detailed information is summarized in Table 1. The source of these datasets is reported in the Appendix A.1 owing to the limited space.

Table 1. Datasets used in our experiments.

Dataset	Samples	Views	Categories
Handwritten	2000	2	10
BDGP	2500	3	5
Cora	2708	4	7
CiteSeer	3312	4	6
STL10	13000	4	9
YTB10	38654	4	10

4.1.2. COMPARED METHODS

To demonstrate the superiority of our proposed method, we conduct comprehensive experiments to compare our DMVLS with various baselines, including MVAR (Tao et al., 2017), Co-GCN (Li et al., 2020), ERL-MVSC (Huang et al., 2021a), DSRL (Wang et al., 2022b), IMvGCN(Wu et al., 2023a), the details of the compared algorithms are given in the Appendix A.2 because of the limited space.

4.1.3. SETTINGS

In our implementation, we fix the hyper-parameters $\lambda_1 = 1$, $\lambda_2 = 2^5$, and the learning rate α equals to 10^{-3}. Initially, we randomly select $B = 0.05n$ samples to label as cold start like existing semi-supervised classification methods. Additionally, we conduct the process ten times and report the average values to eliminate the randomness. Then, the same number of samples is annotated at each round. The training process ends when half of the samples are labeled. Additionally, $\tanh(\cdot)$ is adopted as the activation function for all datasets. Adam optimizer is applied to update the network parameters Θ.

For the compared methods, all baselines' parameter settings are tuned by their papers' suggestions. We randomly select the labeled samples ten times for each label ratio and report the average results and standard deviation (std). The accuracy (ACC) is adopted as the measured metric in our experiments.

All the experiments are conducted on a machine with Intel Core i9-10850K CPU @ 3.60GHz, 64GB RAM, and Nvidia RTX 3090 GPU.

4.2. Results

To show the superiority of our proposed method, we record the classification results under different label ratios on six datasets in Table 2. From the table, we have the following observations:

- Our proposed DMVLS is superior to existing methods on all datasets on different label ratios. For instance, it outperforms the second-best method over 1.60%, 15.12%, 13.98%, 20.36%, 18.15%, and 0.08% when half of the samples are labeled. The improvements in other label ratios are also promising, which shows the effectiveness of DMVLS.

- Compared with existing methods, our method is more suitable for large-scale data. Most Graph-based methods fail to handle extensive data since they utilize label propagation with the graph structure, which undergoes quadratic space complexity respecting the sample number. On the contrary, our proposed method adopts the AE structure, and the complexity is linear to the sample number.

- Our proposed method is more stable than existing methods. Compared with the five competitors, the std results of our method are smaller. The reason is that the performance is guaranteed since we select valuable samples to label, while the compared methods annotate samples randomly, and the results fluctuate a lot.

4.3. The Effect of Sample Selection

To analyze the effect of the label strategy of our method, we plot the classification performance varying with label ratios on six datasets in Fig. 2. Additionally, we document more detailed results in the Appendix A.3 due to the limited space. The figure shows that the classification results obviously increase with the growing label ratio, while the performance of the compared ones does not grow significantly. Due to the sample selection strategy, our method selects the uncertain samples to label and clarify the decision boundary. Then, we retrain the model to improve the classification performance. On the contrary, existing semi-supervised multi-view classification methods fail to choose valuable samples to label and are sensitive to the chosen labeled samples. For instance, MVAR attains worse performance with larger label ratios on Citeseer, and the performance of

Table 2. Empirical evaluation and comparison of our method with five compared methods on six benchmark datasets in terms of ACC. Note that '-' indicates the method fails to run smoothly due to the out-of-memory error, and the best results are marked in bold.

Ratio	Methods	Handwritten	BDGP	Cora	CiteSeer	STL10	YTB10
20%	MVAR	48.14±2.06	79.41±1.93	59.92±1.15	64.22±1.15	56.88±0.66	-
	Co-GCN	87.06±3.46	74.83±0.40	50.16±1.24	60.41±0.81	29.99±1.05	99.54±1.12
	ERL-MVSC	95.48±0.77	37.24±1.78	75.59±1.98	62.57±1.69	58.23±0.64	-
	DSRL	97.06±0.49	18.77±1.48	66.53±2.12	50.44±2.66	47.99±0.22	-
	IMvGCN	93.89±0.56	71.01±0.69	77.74±0.76	71.07±0.71	**60.12±0.32**	99.61±0.44
	Ours	**98.35±0.42**	**91.38±0.19**	**82.36±0.97**	**72.99±0.99**	54.46±0.87	**100.00±0.00**
30%	MVAR	35.21±1.22	80.74±0.73	63.45±0.94	63.35±0.94	60.59±0.40	-
	Co-GCN	89.81±2.58	71.13±3.41	46.96±4.16	62.43±0.30	33.07±2.14	99.77±0.10
	ERL-MVSC	97.00±0.50	40.79±2.21	81.70±0.65	66.81±1.18	61.69±0.39	-
	DSRL	98.21±0.15	20.02±0.36	71.75±0.02	52.72±2.55	49.87±0.34	-
	IMvGCN	94.39±0.44	69.90±1.62	77.99±0.59	70.97±1.18	60.59±0.19	99.92±0.05
	Ours	**99.48±0.32**	**95.67±0.25**	**88.58±0.58**	**78.68±0.87**	**63.00±0.44**	**100.00±0.00**
40%	MVAR	71.92±1.42	82.77±0.34	67.06±1.32	62.73±1.32	61.79±0.54	-
	Co-GCN	88.86±3.80	73.23±2.80	49.19±1.39	63.05±0.42	31.42±0.74	99.84±0.01
	ERL-MVSC	97.82±0.37	44.91±3.42	83.30±0.71	70.39±0.78	63.14±0.38	-
	DSRL	98.42±0.41	20.04±0.33	70.77±2.48	54.83±2.85	51.32±0.10	-
	IMvGCN	94.10±0.63	70.19±2.18	78.17±0.53	71.69±0.87	60.55±0.14	99.72±0.50
	Ours	**99.94±0.08**	**96.88±0.19**	**92.86±1.15**	**82.51±0.35**	**70.41±0.24**	**100.00±0.00**
50%	MVAR	82.16±1.24	84.67±1.54	70.52±1.31	61.45±1.31	62.88±0.79	-
	Co-GCN	89.33±2.38	74.76±1.27	48.40±1.10	63.63±1.10	31.48±0.75	99.86±0.02
	ERL-MVSC	97.98±0.44	45.70±1.59	84.42±1.05	72.26±1.47	64.01±0.43	-
	DSRL	98.40±0.29	20.08±0.75	74.37±0.58	55.90±2.47	52.54±0.35	-
	IMvGCN	93.84±0.78	69.81±1.21	78.67±0.32	71.97±0.76	60.59±0.23	99.92±0.07
	Ours	**99.97±0.05**	**97.47±0.24**	**96.22±0.63**	**86.97±0.78**	**75.63±0.37**	**100.00±0.00**

Co-GCN and IMvGCN almost stays unchanged on different label ratios on all datasets.

4.4. Ablation Study

In our framework, we first decouple the SSI via a view discriminator and then utilize it to conduct the subsequent classification task with two individual classifiers. Meanwhile, the sample selection strategy is beneficial for choosing valuable samples to label and clarifying the decision boundary. To investigate the effectiveness of the main components of our framework, we conduct an ablation study and compare our method with four methods when half of the samples are labeled, and the results on other label ratios are given in Appendix A.5. By removing the loss $\mathcal{L}_{vc}$, $\mathcal{L}_{class1}$, and $\mathcal{L}_{class1}$ separately, we study the validity of the view discriminator and the utilization of SSI. Also, we randomly select samples for each label ratio to annotate instead of labeling samples with a specific strategy to study the effectiveness of our label strategy. The results are reported in Table 3. From the table, it is seen that the performance diminishes when each element of our framework is removed, which shows the effectiveness of our framework. Furthermore, it is seen that some samples might be easier to classify via specific knowledge or the opposite in semi-supervised multi-view classification. For instance, the method merely utilizing shared information performs better on STL10 but works worse on CiteSeer than the one just leveraging specific information. Therefore, the attempt to decouple the SSI and exploit two pieces of information respectively is necessary in semi-supervised multi-view classification.

4.5. Visualization

In order to intuitively demonstrate the superiority of the representation learning of our proposed DMVLS, we visualize the learned representations via the t-SNE algorithm (van der Maaten & Hinton, 2008) when half of the samples are labeled on Handwritten, BDGP, Cora, and YTB10, and the results are plotted in Fig. 3. From the figure, we observe that our algorithm tends to generate separable representations on these datasets, and the learned representation is favourable to the subsequent classification process, ensuring the model assigns accurate class labels. We attribute the success of representation learning to the extraction and decoupling of SSI with AE and the view discriminator.

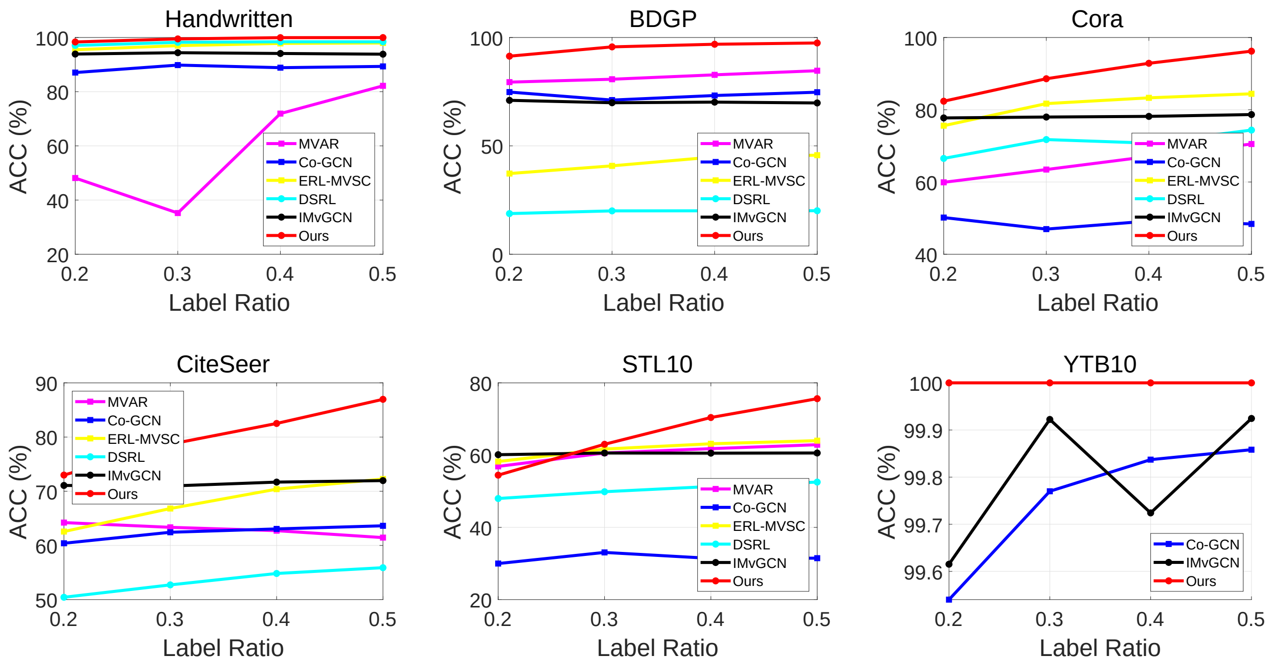

Figure 2. The classification performance of our method and the compared algorithms varies with different label ratios on six benchmark datasets.

Table 3. The ablation study of our method in terms of ACC when half of the samples are labeled. The best results are marked in bold.

Ratio	Methods	Handwritten	BDGP	Cora	CiteSeer	STL10	YTB10
	Remove $\mathcal{L}_{vc}$	99.97±0.05	97.00±0.36	94.83±0.50	84.77±1.64	74.77±0.78	100.00±0.00
	Remove $\mathcal{L}_{class1}$	99.87±0.05	95.87±0.34	95.65±0.55	73.62±13.01	73.54±2.04	100.00±0.00
50%	Remove $\mathcal{L}_{class2}$	99.87±0.12	96.96±0.34	95.62±0.34	86.09±0.21	64.65±7.33	100.00±0.00
	Random selection	96.42±1.02	87.60±0.38	82.01±0.55	69.57±1.61	62.79±0.69	99.99±0.01
	Ours	**99.97±0.05**	**97.47±0.24**	**96.22±0.63**	**86.97±0.78**	**75.63±0.37**	**100.00±0.00**

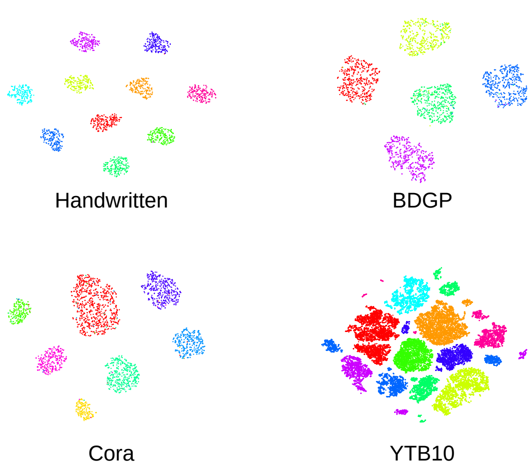

Figure 3. T-SNE visualization of our method on four datasets.

5. Conclusion

This paper proposes a Dynamic Multi-view Labeling Strategy with Shared and Specific Information method to label multi-view instances. Precisely, we extract and decouple the shared and specific information of multi-view data via the auto-encoder and view discriminator. Then, the two pieces of information are utilized to train the classifiers separately. We select the samples of low classification confidence and label them in high priorities. We interactively select samples to label and retrain the model with the newly generated labels to boost the classification performance. Comprehensive experiments demonstrate the superiority of our algorithm and the effectiveness of the critical components. In the future, we aim to study how to choose the initial training samples as a warm start.

6. Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

AL-Sammarraie, Z. D. and Karaca, A. C. Modified active learning framework for hyperspectral image classification with convolutions neural networks. In *2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–6. IEEE, 2023.

Bellarmino, N., Cantoro, R., Huch, M., Kilian, T., Martone, R., Schlichtmann, U., and Squillero, G. A multilabel active learning framework for microcontroller performance screening. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 42(10):3436–3449, 2023. doi: 10.1109/TCAD.2023.3245989.

Blum, A. and Mitchell, T. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pp. 92–100, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130570. doi: 10.1145/279943.279962. URL https://doi.org/10.1145/279943.279962.

Cai, J., Han, H., Cui, J., Chen, J., Liu, L., and Zhou, S. K. Semi-supervised natural face de-occlusion. *IEEE Transactions on Information Forensics and Security*, 16:1044–1057, 2020.

Cao, X., Yao, J., Xu, Z., and Meng, D. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4604–4616, 2020.

Chen, M., Wei, Z., Huang, Z., Ding, B., and Li, Y. Simple and deep graph convolutional networks. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1725–1735. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/chen20v.html.

Chen, N., Klushyn, A., Paraschos, A., Benbouzid, D., and Van der Smagt, P. Active learning based on data uncertainty and model sensitivity. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1547–1554. IEEE, 2018.

Ducoffe, M. and Precioso, F. Adversarial active learning for deep networks: a margin based approach. *arXiv preprint arXiv:1802.09841*, 2018.

He, R., Dai, Z., He, S., and Tang, K. Perturbation-based two-stage multi-domain active learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pp. 3933–3937, New York, NY, USA, 2023a. Association for Computing Machinery. ISBN 9798400701245. doi: 10.1145/3583780.3615222. URL https://doi.org/10.1145/3583780.3615222.

He, R., Liu, S., He, S., and Tang, K. Multi-domain active learning: Literature review and comparative study. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(3):791–804, 2023b. doi: 10.1109/TETCI.2022.3217753.

Holmes, C., Zhang, M., He, Y., and Wu, B. Nxmtransformer: Semi-structured sparsification for natural language understanding via admm. *Advances in neural information processing systems*, 34:1818–1830, 2021.

Huang, A., Wang, Z., Zheng, Y., Zhao, T., and Lin, C.-W. Embedding regularizer learning for multi-view semi-supervised classification. *IEEE Transactions on Image Processing*, 30:6997–7011, 2021a. doi: 10.1109/TIP.2021.3101917.

Huang, A., Wang, Z., Zheng, Y., Zhao, T., and Lin, C.-W. Embedding regularizer learning for multi-view semi-supervised classification. *IEEE Transactions on Image Processing*, 30:6997–7011, 2021b.

Jia, X., Jing, X.-Y., Zhu, X., Chen, S., Du, B., Cai, Z., He, Z., and Yue, D. Semi-supervised multi-view deep discriminant representation learning. *IEEE transactions on pattern analysis and machine intelligence*, 43(7):2496–2509, 2020.

Joshi, A. J., Porikli, F., and Papanikolopoulos, N. Multi-class active learning for image classification. In *2009 ieee conference on computer vision and pattern recognition*, pp. 2372–2379. IEEE, 2009.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Kothawade, S., Beck, N., Killamsetty, K., and Iyer, R. Similar: Submodular information measures based active learning in realistic scenarios. *Advances in Neural Information Processing Systems*, 34:18685–18697, 2021.

Li, S., Li, W.-T., and Wang, W. Co-gcn for multi-view semi-supervised learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):

4691–4698, Apr. 2020. doi: 10.1609/aaai.v34i04.5901. URL https://ojs.aaai.org/index.php/AAAI/article/view/5901.

Liang, K., Meng, L., Liu, M., Liu, Y., Tu, W., Wang, S., Zhou, S., Liu, X., and Sun, F. Reasoning over different types of knowledge graphs: Static, temporal and multi-modal. *arXiv preprint arXiv:2212.05767*, 2022.

Liang, K., Liu, Y., Zhou, S., Tu, W., Wen, Y., Yang, X., Dong, X., and Liu, X. Knowledge graph contrastive learning based on relation-symmetrical structure. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–12, 2023. doi: 10.1109/TKDE.2023.3282989.

Liu, S., Wang, S., Zhang, P., Xu, K., Liu, X., Zhang, C., and Gao, F. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7576–7584, 2022.

Liu, S., Zhang, J., Wen, Y., Yang, X., Wang, S., Zhang, Y., Zhu, E., Tang, C., Zhao, L., and Liu, X. Sample-level cross-view similarity learning for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 14017–14025, 2024.

Liu, X., Liu, L., Liao, Q., Wang, S., Zhang, Y., Tu, W., Tang, C., Liu, J., and Zhu, E. One pass late fusion multi-view clustering. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6850–6859. PMLR, 18–24 Jul 2021a. URL https://proceedings.mlr.press/v139/liu21l.html.

Liu, Z., Ding, H., Zhong, H., Li, W., Dai, J., and He, C. Influence selection for active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9274–9283, 2021b.

Liu, Z., Jiang, X., Luo, H., Fang, W., Liu, J., and Wu, D. Pool-based unsupervised active learning for regression using iterative representativeness-diversity maximization (irdm). *Pattern Recognition Letters*, 142:11–19, 2021c.

Luo, S., Zhang, C., Zhang, W., and Cao, X. Consistent and specific multi-view subspace clustering. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018a. ISBN 978-1-57735-800-8.

Luo, S., Zhang, C., Zhang, W., and Cao, X. Consistent and specific multi-view subspace clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018b. doi: 10.1609/aaai.v32i1.11617. URL https://ojs.aaai.org/index.php/AAAI/article/view/11617.

Nguyen, V.-L., Shaker, M. H., and Hüllermeier, E. How to measure uncertainty in uncertainty sampling for active learning. *Machine Learning*, 111(1):89–122, 2022.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., and Wang, X. A survey of deep active learning. *ACM computing surveys (CSUR)*, 54(9):1–40, 2021.

Roth, D. and Small, K. Margin-based active learning for structured output spaces. In Fürnkranz, J., Scheffer, T., and Spiliopoulou, M. (eds.), *Machine Learning: ECML 2006*, pp. 413–424, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-46056-5.

Shen, M., Huang, Y., Yin, J., Zou, H., Rajan, D., and See, S. Towards balanced active learning for multimodal classification. *arXiv preprint arXiv:2306.08306*, 2023.

Tao, H., Hou, C., Nie, F., Zhu, J., and Yi, D. Scalable multi-view semi-supervised classification via adaptive regression. *IEEE Transactions on Image Processing*, 26 (9):4283–4296, 2017. doi: 10.1109/TIP.2017.2717191.

Tu, W., Liao, Q., Zhou, S., Peng, X., Ma, C., Liu, Z., Liu, X., Cai, Z., and He, K. Rare: Robust masked graph autoencoder. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2023. doi: 10.1109/TKDE.2023.3335222.

van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL http://jmlr.org/papers/v9/vandermaaten08a.html.

Wan, S., Pan, S., Yang, J., and Gong, C. Contrastive and generative graph convolutional networks for graph-based semi-supervised learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(11): 10049–10057, May 2021. doi: 10.1609/aaai.v35i11.17206. URL https://ojs.aaai.org/index.php/AAAI/article/view/17206.

Wan, X., Liu, X., Liu, J., Wang, S., Wen, Y., Liang, W., Zhu, E., Liu, Z., and Zhou, L. Auto-weighted multi-view clustering for large-scale data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10078–10086, 2023.

Wan, X., Liu, J., Gan, X., Liu, X., Wang, S., Wen, Y., Wan, T., and Zhu, E. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2024a. doi: 10.1109/TNNLS.2024.3378194.

Wan, X., Xiao, B., Liu, X., Liu, J., Liang, W., and Zhu, E. Fast continual multi-view clustering with incomplete views. *IEEE Transactions on Image Processing*, 33:2995–3008, 2024b. doi: 10.1109/TIP.2024.3388974.

Wang, H., Huang, W., Wu, Z., Tong, H., Margenot, A. J., and He, J. Deep active learning by leveraging training dynamics. *Advances in Neural Information Processing Systems*, 35:25171–25184, 2022a.

Wang, S., Chen, Z., Du, S., and Lin, Z. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5042–5055, 2021a.

Wang, S., Wang, Z., Lim, K.-L., Xiao, G., and Guo, W. Seeded random walk for multi-view semi-supervised classification. *Knowledge-Based Systems*, 222:107016, 2021b.

Wang, S., Chen, Z., Du, S., and Lin, Z. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5042–5055, 2022b. doi: 10.1109/TPAMI.2021.3082632.

Wang, S., Liu, X., Liu, S., Jin, J., Tu, W., Zhu, X., and Zhu, E. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 5882–5895. Curran Associates, Inc., 2022c.

Wang, X., Fu, L., Zhang, Y., Wang, Y., and Li, Z. Mmatch: Semi-supervised discriminative representation learning for multi-view classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):6425–6436, 2022d. doi: 10.1109/TCSVT.2022.3159371.

Wang, X., Fu, L., Zhang, Y., Wang, Y., and Li, Z. Mmatch: semi-supervised discriminative representation learning for multi-view classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):6425–6436, 2022e.

Wu, Z., Lin, X., Lin, Z., Chen, Z., Bai, Y., and Wang, S. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE Transactions on Multimedia*, 2023a.

Wu, Z., Zhang, Z., and Fan, J. Graph convolutional kernel machine versus graph convolutional networks. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 19650–19672. Curran Associates, Inc., 2023b. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3ec6c6fc9065aa57785eb05dffe7c3db-Paper-Conference.pdf.

Xia, W., Wang, Q., Gao, Q., Zhang, X., and Gao, X. Self-supervised graph convolutional network for multi-view clustering. *IEEE Transactions on Multimedia*, 24:3182–3192, 2021.

Xu, M., Zhao, Q., and Jia, S. Multiview spatial–spectral active learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2021.3095292.

Xu, Y., Shang, L., Ye, J., Qian, Q., Li, Y.-F., Sun, B., Li, H., and Jin, R. Dash: Semi-supervised learning with dynamic thresholding. In *International Conference on Machine Learning*, pp. 11525–11536. PMLR, 2021.

Yu, G., Xing, Y., Wang, J., Domeniconi, C., and Zhang, X. Multiview multi-instance multilabel active learning. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9):4311–4321, 2022. doi: 10.1109/TNNLS.2021.3056436.

Yuan, J., Yu, H., Cao, M., Xu, M., Xie, J., and Wang, C. Semi-supervised and self-supervised classification with multi-view graph neural networks. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, pp. 2466–2476, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384469. doi: 10.1145/3459637.3482477. URL https://doi.org/10.1145/3459637.3482477.

Zhang, B., Qiang, Q., Wang, F., and Nie, F. Fast multi-view semi-supervised learning with learned graph. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):286–299, 2020.

Zhmoginov, A., Sandler, M., and Vladymyrov, M. Hypertransformer: Model generation for supervised and semi-supervised few-shot learning. In *International Conference on Machine Learning*, pp. 27075–27098. PMLR, 2022.

Zhou, T., Zhang, C., Peng, X., Bhaskar, H., and Yang, J. Dual shared-specific multiview subspace clustering. *IEEE Transactions on Cybernetics*, 50(8):3517–3530, 2020. doi: 10.1109/TCYB.2019.2918495.

Zhou, Y., Renduchintala, A., Li, X., Wang, S., Mehdad, Y., and Ghoshal, A. Towards understanding the behaviors of optimal deep active learning algorithms. In Banerjee, A. and Fukumizu, K. (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 1486–1494. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/zhou21b.html.

A. Appendix

A.1. Datasets

We conduct our experiments on the widely used datasets, including Handwritten[1], BDGP[2], Cora[3], CiteSeer[4], STL10[5], and YTB10 [6]. Specifically, Handwritten is an image dataset containing handwritten digits. BDGP dataset consists of images and text modalities. Cora and Citeseer are citation network datasets. STL10 and YTB10 are face image datasets.

A.2. The Compared Methods

The source code and the details of the compared methods are summarized as follows:

1. Scalable multi-view semi-supervised classification via adaptive regression (MVAR)[7]. It proposes a regression-based loss functions with $\ell_{2,1}$ matrix norm for each view and the combine the view loss with a linear weighted combination.

2. Co-GCN for Multi-View Semi-Supervised Learning (Co-GCN)[8]. It adaptively exploits the graph information from the multiple views with combined Laplacians and utilizes GCN to conduct classification.

3. Embedding Regularizer Learning for Multi-View Semi-Supervised Classification (ERL-MVSC)[9]. ERL-MVSC integrates diversity, sparsity, and consensus to manipulate multi-view data dexterously.

4. Learning Deep Sparse Regularizers With Applications to Multi-View Clustering and Semi-Supervised Classification (DSRL)[10]. It proposes a deep sparse regularizer learning model that learns data-driven sparse regularizers adaptively.

5. Interpretable Graph Convolutional Network for Multi-View Semi-Supervised Learning (IMvGCN)[11]. It equips GCN with a deep layer of interpretability.

A.3. The Effect of Sample Selection

To further investigate the effect of the sample selection strategy, we document the performance changes with each label ratio compared with the random selection method in Fig. 4. From the figure, it is observed that the performance of our proposed method consistently improves when the label ratio gets larger, and it outperforms the random selection method on all datasets. Therefore, the sample selection strategy is better than the random selection.

A.4. Convergence Analysis

To investigate the convergence of our proposed method, we conduct the convergence analysis in this section. The loss value varying with epochs is plotted in Fig. 5. From the figure, it is obtained that the loss gradually decreases and converges to a value, which verifies the convergence of our method experimentally.

A.5. Ablation Study

We report the results of our ablation study with different label ratios in Table 4. From the table, we can conclude that the performance drops when any of the components of our framework is removed, showing the effectiveness of the critical component of our method.

[1] https://archive.ics.uci.edu/ml/datasets/Multiple+Features
[2] https://www.fruitfly.org/
[3] http://linqs-data.soe.ucsc.edu/public/lbc/
[4] https://linqs.org/datasets/#citeseer-doc-classification
[5] https://cs.stanford.edu/~acoates/stl10
[6] http://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset
[7] https://github.com/taohong08/Scalable-Multi-View-Semi-Supervised-Classification-via-Adaptive-Regress
[8] https://github.com/cheunglei/AAAI-20-Co-GCN/blob/master/Co-GCN/train.py
[9] https://github.com/huangsuj/ERL-MVSC
[10] https://github.com/chenzl23/DSRL
[11] https://github.com/ZhihaoWu99/IMvGCN

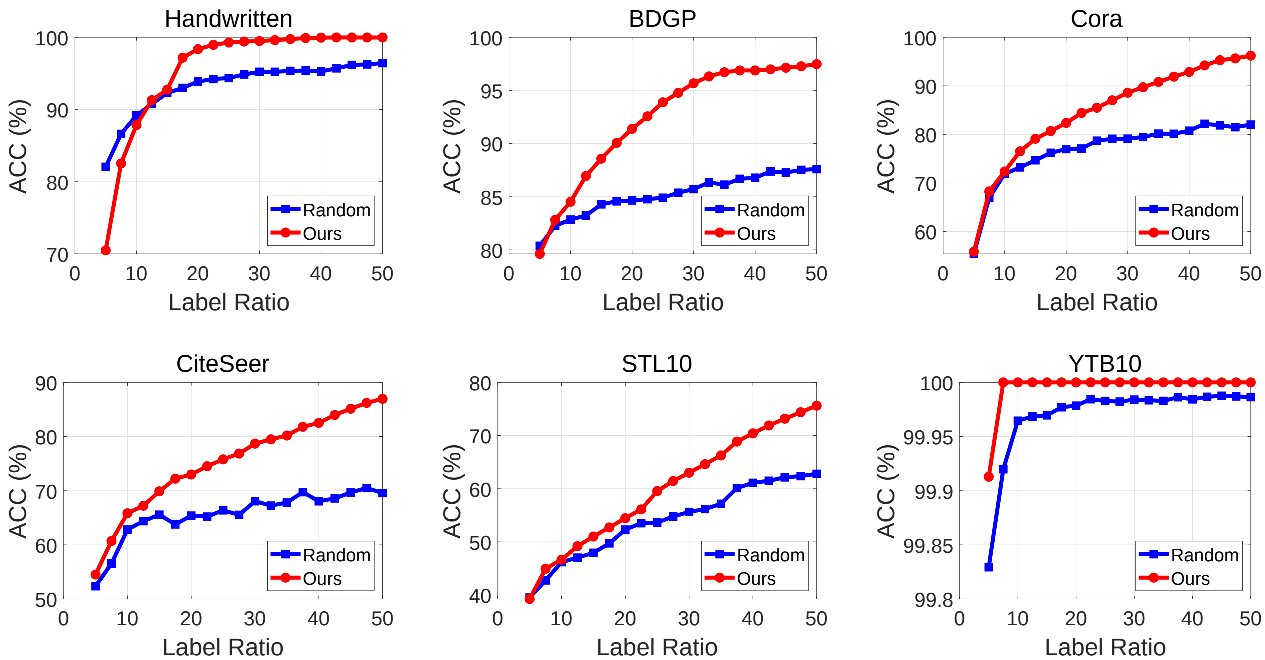

Figure 4. The classification performance varies with different label ratios on six benchmark datasets.

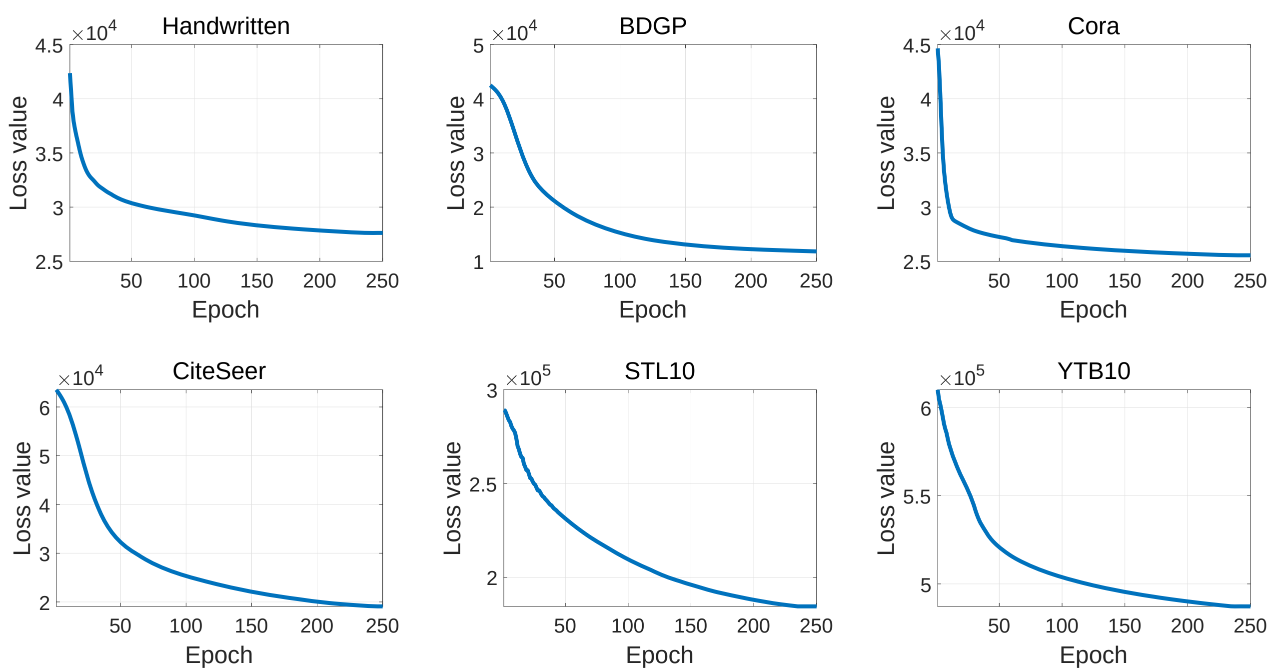

Figure 5. Convergence analysis of our method on six benchmark datasets.

Table 4. The ablation study of our method in terms of ACC. The best results are marked in bold.

Ratio	Methods	Handwritten	BDGP	Cora	CiteSeer	STL10	YTB10
20%	Remove $\mathcal{L}_{vc}$	97.83±0.11	91.08±0.32	**82.81±0.63**	70.2±0.63	54.32±2.58	100.00±0.00
	Remove $\mathcal{L}_{class1}$	95.52±4.04	91.08±0.48	80.81±2.71	58.56±11.73	51.90±5.07	100.00±0.00
	Remove $\mathcal{L}_{class2}$	92.25±9.15	90.85±1.65	81.39±1.04	71.24±0.81	44.73±6.60	100.00±0.00
	Random selection	93.87±1.18	84.65±0.53	76.98±1.44	65.39±1.13	52.31±3.56	99.98±0.01
	Ours	**98.35±0.42**	**91.38±0.19**	82.36±0.97	**72.99±0.99**	**54.46±0.87**	**100.00±0.00**
30%	Remove $\mathcal{L}_{vc}$	99.21±0.21	95.42±0.19	88.13±0.70	76.84±0.62	62.68±1.88	100.00±0.00
	Remove $\mathcal{L}_{class1}$	99.35±0.34	95.18±0.11	87.83±2.17	63.18±13.05	59.13±4.77	100.00±0.00
	Remove $\mathcal{L}_{class2}$	97.52±3.25	94.99±0.66	88.27±0.04	77.61±0.65	51.25±6.44	100.00±0.00
	Random selection	95.21±0.09	85.72±0.18	79.12±0.58	68.06±1.17	55.62±3.51	99.98±0.01
	Ours	**99.48±0.32**	**95.67±0.25**	**88.58±0.58**	**78.68±0.87**	**63.00±0.44**	**100.00±0.00**
40%	Remove $\mathcal{L}_{vc}$	99.92±0.07	96.44±0.74	92.56±0.16	81.23±1.12	70.24±1.26	100.00±0.00
	Remove $\mathcal{L}_{class1}$	99.78±0.10	95.96±0.23	92.27±1.27	68.41±13.35	67.56±2.76	100.00±0.00
	Remove $\mathcal{L}_{class2}$	99.72±0.22	96.75±0.09	92.54±0.71	82.31±0.40	57.63±7.54	100.00±0.00
	Random selection	95.26±0.41	86.79±0.18	80.76±1.11	68.05±0.72	61.10±1.00	99.98±0.01
	Ours	**99.94±0.08**	**96.88±0.19**	**92.86±1.15**	**82.51±0.35**	**70.41±0.24**	**100.00±0.00**
50%	Remove $\mathcal{L}_{vc}$	99.97±0.05	97.00±0.36	94.83±0.50	84.77±1.64	74.77±0.78	100.00±0.00
	Remove $\mathcal{L}_{class1}$	99.87±0.05	95.87±0.34	95.65±0.55	73.62±13.01	73.54±2.04	100.00±0.00
	Remove $\mathcal{L}_{class2}$	99.87±0.12	96.96±0.34	95.62±0.34	86.09±0.21	64.65±7.33	100.00±0.00
	Random selection	96.42±1.02	87.60±0.38	82.01±0.55	69.57±1.61	62.79±0.69	99.99±0.01
	Ours	**99.97±0.05**	**97.47±0.24**	**96.22±0.63**	**86.97±0.78**	**75.63±0.37**	**100.00±0.00**

A.6. Limitations and Future Work

We randomly label some samples and train the model as a cold start. Then, we interactively select samples to label under our selection strategy and retrain the model with existing labels. In the future, we plan to investigate how to select the initial training samples. Also, in our paper, we select 2.5% samples to label at each round. The number of samples to label is worth studying. Additionally, the cost between the label cost and classification loss could be balanced via a specific strategy. We could design a strategy to balance them in the future.