Neural Collapse Priors Driven Trust Semi-Supervised Multi-View Classification

Taotao Guo[1], Honglin Yuan[2], Xujian Zhao[2], Yuan Sun[3], Dongliang Wang[2], Zhenwen Ren[2],Xingfeng Li[2*]

[1]School of information and Control Engineering, Southwest University of Science and Technology, Mianyang, China
[2]School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China
[3]National Key Laboratory of Fundamental Algorithms and Models for Engineering Numerical Simulation, Sichuan University, Chengdu, China

Abstract

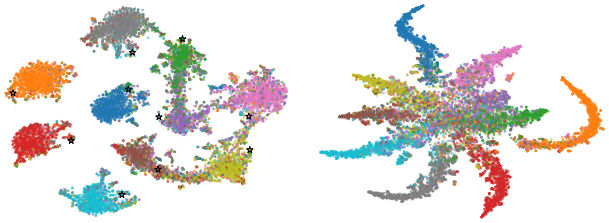
In semi-supervised multi-view classification (SMVC), scarce labels and noisy unlabeled data impair feature aggregation and compromise prediction reliability, while existing methods lack principled guidance and interpretability. To overcome these limitations, we propose a novel unified SMVC framework, Neural Collapse Priors Driven Trust Semi-Supervised Multi-View Classification (NCPD-TSMVC), building upon neural collapse–derived prototype priors and evidential opinion fusion. Concretely, we rigorously prove under neural collapse theory that normalized classifier weights from the labeled-data pre-training stage coincide with class centroids in feature space, conferring maximal inter-class separation and optimal within-class compactness. These prototype priors permeate the entire learning pipeline, calibrating the representation learning of unlabeled samples to obtain highly discriminative embeddings. Simultaneously, our evidential learning module quantifies epistemic uncertainty and fuses view-level opinions at the evidence level, yielding robust and transparent decision making. Extensive evaluations across diverse benchmarks demonstrate that NCPD-TSMVC surpasses state-of-the-art SMVC approaches in performance, robustness and interpretability.

Code — https://github.com/GTT-1/NCPD-TSMVC

Introduction

With the rapid development of information technology and continuous innovation in data acquisition methods, the availability of multi-source heterogeneous data has become increasingly diverse (Ren et al. 2021; Xu et al. 2022; Li et al. 2023c), which lays a solid foundation for research in Multi-View Learning (MVL). In recent years, MVL has demonstrated significant application potential and research achievements across various fields, including data mining (Xu et al. 2024a; Li et al. 2023b; Kang et al. 2024; Luo et al. 2024; Lou et al. 2025; Zhao et al. 2024; Wang et al. 2022a; Wen et al. 2025a; Wang et al. 2025; Qin et al. 2025; Li et al. 2024a), machine learning (Li et al. 2025; Wen et al. 2025b; Luo et al. 2025a; Yang, Chung, and Jang 2024; Wang et al. 2021a; Luo et al. 2025b; Hu et al. 2024; Wan et al. 2024a; Yuan et al. 2025; Wan et al. 2024b; Wen et al. 2023; Li et al. 2023c; Wang, Zhang, and Zhou 2025), and computer vision (Li et al. 2024b; Luo et al. 2021; Wang et al. 2020a; Luo, Xu, and Yang 2019; Wang et al. 2023; Li et al. 2024c; Xu et al. 2024b; Yang et al. 2022). Among the numerous MVL tasks, Multi-View Classification (MVC) stands out as one of the most representative research directions. The core of MVC methods (Andrew et al. 2013; Kan, Shan, and Chen 2016; Wang et al. 2020b) lies in fully exploiting the supervision from labels and the complementarity among views, learning a representation space that captures the intrinsic structure of samples, and constructing a discriminative classifier to achieve higher classification accuracy.

However, constructing high-quality multi-view datasets often relies heavily on extensive manual annotation, and the scarcity of expert annotators combined with complex labeling processes makes acquiring large-scale, high-quality labeled data extremely costly. In practical applications, only a small portion of the data is labeled, while the majority remains unlabeled. Therefore, how to effectively leverage vast amounts of unlabeled data under limited labeled data conditions to achieve desirable classification performance has become a critical challenge. To address this challenge, re-



(a) NC prior guidance (b) Without NC prior guidance

Figure 1: The t-SNE embeddings of MNIST dataset under Gaussian noise ($\sigma = 1.5$): (a) With Neural Collapse (NC) priors: Despite in noise case, digit classes form tight, well-separated clusters, preserving clear semantic structure under the guidance of prototype priors via NC theory; (b) Without NC priors: Embeddings scatter radially and overlap across classes, undermining discriminability under the same noise level. Comparatively, our NC priors could enhance the discrimination and robustness of classification.

*Corresponding authors.

searchers have proposed Semi-Supervised Multi-View Classification (SMVC) methods. Numerous SMVC approaches have emerged in this context. Existing methods for mining semantic consistency across multiple views can be broadly categorized into two types: (1) direct integration methods, which combine multiple original views through concatenation or weighted summation to generate a common label matrix representing the final classification results, thereby enforcing classification consistency across views (Ren et al. 2021; Xu et al. 2022; Li et al. 2023c); (2) sub-model collaboration methods, which construct independent sub-models for each view to extract high-level view-specific features, and impose cross-view consistency constraints to align feature representations and classification results among different sub-models. For example, (Wang et al. 2021b) integrate multiple views combined with attention mechanisms to generate improved pseudo-views that guide feature learning in each view; (Wang et al. 2022b) and (Wang et al. 2024) further consider discriminative structures across views, extracting view-specific features and aggregating classification results to produce pseudo-labels, which are then used to guide model learning in a feedback manner. However, existing SMVC methods often neglect the noise carried by unlabeled data during training, which causes model learning to deviate. As training progresses, such deviations accumulate erroneous information continuously, ultimately leading to performance degradation in classification.

To address the above issues, this paper proposes a novel framework: Neural Collapse Priors-Driven Trust Semi-Supervised Multi-View Classification. The overall framework is illustrated in Figure 2. The main objectives of this framework are as follows: 1) To extract shared classifier weights with semantic consistency; 2) To theoretically demonstrate the equivalence between the normalized weights of the shared classifier and prototype priors; 3) To calibrate the feature representations of unlabeled data using prototype priors; 4) To improve classification accuracy and enhance the interpretability of model decisions. Specifically, during a fully labeled pre-training stage, we calculate the class centers for each category based on ground-truth labels and leverage a shared classifier to extract semantically consistent weights. These weights serve as a stable and discriminative semantic prior for the subsequent training on unlabeled data. Based on the theory of Neural Collapse, we rigorously prove that the normalized classifier weights are equivalent to class prototypes in the feature space, ensuring that they possess stable and highly discriminative semantic characteristics (i.e., intra-class compactness and inter-class separability). To effectively calibrate the feature representations of unlabeled data and facilitate cross-view feature consistency learning, we introduce an attention layer and cross-view contrastive learning, which progressively align the unlabeled feature representations with the spatial distribution of the prototype priors. In summary, this work makes the following contributions:

- We propose a novel semi-supervised multi-view classification framework that leverages prototype priors derived from Neural Collapse to calibrate the feature representations of unlabeled data, significantly alleviating the adverse effects of noisy unlabeled samples on feature learning.
- We theoretically prove the equivalence between classifier weights and class prototypes, revealing their stable and discriminative properties. This enhances decision interpretability while improving classification performance.
- Through extensive experiments on six standard benchmarks and their noise-augmented variants, the robustness and superior performance are demonstrated.

Method

Prototype Priors via Neural Collapse Assumption

In semi-supervised classification tasks, the scarcity of supervision often leads to unstable model training and poor inter-class separability, which severely hinders effective learning. To alleviate this issue, we first compute the class centers $\{p_a^v\}$ for each view based on labeled data during the pre-training stage, and feed them into a shared classifier to obtain normalized classifier weights W_{NC}. These weights serve as stable and consistent prototype priors to guide the calibration of unlabeled data in subsequent stages.

Specifically, we further justify the following conclusion: after sufficient training, the normalized weights of the shared classifier in the pre-trained model can be regarded as class prototypes, that is, $p_j = \frac{w_{:,j}}{\|w_{:,j}\|_2}$. This allows us to directly use the weights of the shared classifier W_{NC} as prototype priors in the subsequent training stage.

The phenomenon of Neural Collapse is supported by four conclusions (Papyan, Han, and Donoho 2020; Chen et al. 2024; Zhu et al. 2021): Intra-class Variability Collapse (NC_1), Simplex ETF Alignment (NC_2), Self-duality (NC_3), and Simplify To Class Center (NC_4).

NC_1: As training progresses, the features of samples within the same class gradually concentrate, leading to a rapid decrease in intra-class variation. Eventually, the features approximately collapse to a single point. Formally, the within-class covariance matrix tends to zero, i.e., $\mathcal{NC}_1 : \Sigma_{NC}^{\dagger}\Sigma_D \to 0$, where Σ_D is the within-class covariance matrix and Σ_{NC} is the between-class covariance matrix.

NC_2: The class means are equidistant from each other and arranged in a regular simplex structure centered at the origin, thereby maximizing inter-class separability.

NC_3: Upon convergence, the weights of the linear classifier align with the class feature centers, forming a self-dual structure geometrically between the classifier and the Class center.

NC_4: Based on the structural convergence established above, the final classifier weights are equivalent to the Class center.

Prototype-Driven Bidirectional Optimization of Within-View Features

To fully achieve the driving effect of weight-based prototypes as prior information, we introduce an attention mechanism during the main training stage to calibrate the feature representations of unlabeled data. This attention mechanism

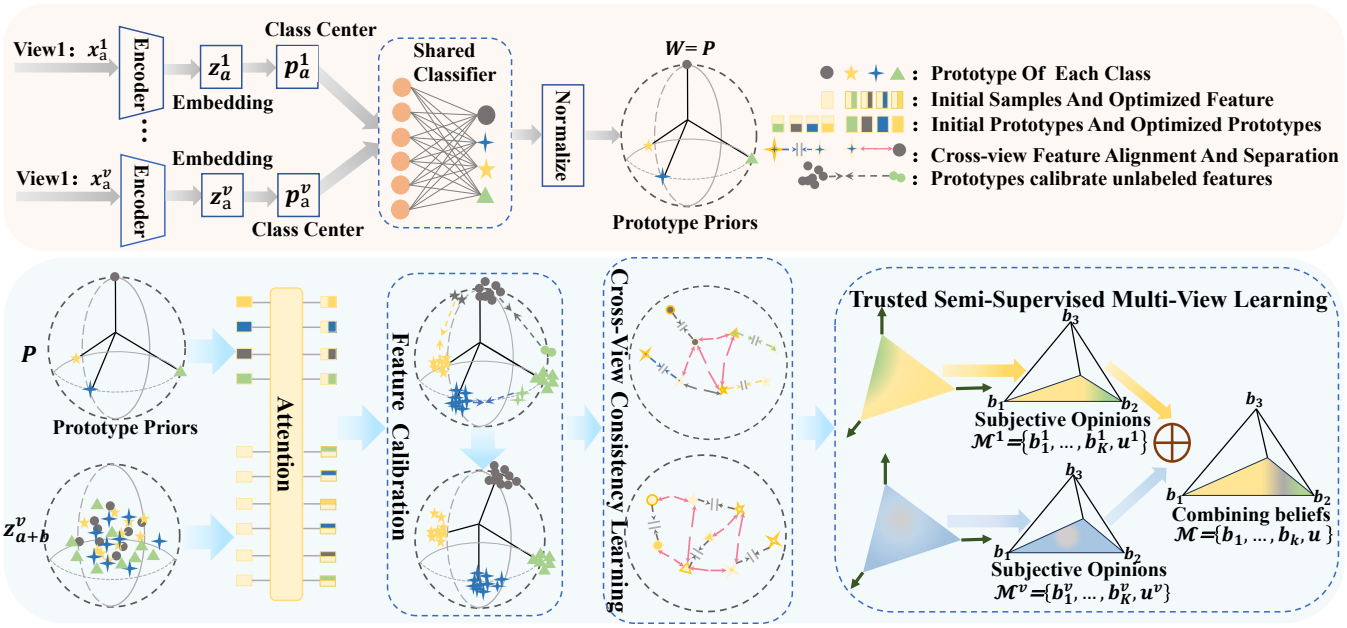


Figure 2: The NCPD-TSMVC method consists of two main stages. (a) In the pre-training stage (red part), only labeled samples $\{x_a^v, y_a^v\}$ are used, where each view extracts features through independent networks. Based on the labels, class centers $\{p_a^v\}$ are computed and a shared classifier obtains consistent weight priors. (b) In the main training stage (blue part), both labeled and unlabeled samples $\{x_{a+b}^v\}$ are utilized. The prior information from pre-training guides the feature learning of unlabeled samples in each view through attention and contrastive learning. Additionally, evidential learning is incorporated during the opinion aggregation to improve the decision-making capability and interpretability of multi-view fusion.

is centered on the weight prototypes, computing the similarity between unlabeled samples and the class-specific weight prototypes, and adaptively assigning aggregation weights. In this way, it achieves a prior-driven bidirectional optimization between sample features and prototypes, further enhancing the guidance of semantic priors on model training.

Specifically, taking a single view v as an example, we first apply two independent linear transformation layers, denoted as W_I^v and W_P^v, to the sample features X^v and the prototype priors P, respectively, thereby obtaining the Query matrix $Q^v = W_I^v X^v$ and the Key matrix $K^v = W_P^v P$. The attention weight matrix A^v is then computed using the scaled dot-product attention:

$$A^v = \text{Softmax}\left(\frac{Q^v (K^v)^\top}{\sqrt{d}}\right). \tag{1}$$

Sample Representation Update(Li et al. 2023a): we apply an independent linear transformation to the prototype priors C, resulting in V_P^v . We then aggregate it with the attention weights A^v and add the result to the original sample features X^v using a residual connection:

$$Z^v = X^v + A^v V_P^v. \tag{2}$$

Prototype Representation Update (Li et al. 2023a): we pass the sample features X^v through another linear transformation to obtain V_I^v. The attention weights are then transposed and multiplied with V^v, and the result is added to the original prototype priors:

$$U^v = C + (A^v)^\top V_I^v. \tag{3}$$

This attention mechanism not only calibrates the features of unlabeled samples but also enables the prototype priors to moderately adapt to the feature distribution of the unlabeled data, thereby achieving dynamic adjustment and more accurate representations.

Prototype-Driven Cross-View Contrastive Learning

Within each view, the attention mechanism driven by prototype priors has effectively captured the correlations between sample features and prototype priors. To further extend the prior information to cross-view feature consistency learning, we design a prototype-driven cross-view contrastive learning strategy. This strategy imposes consistency constraints on sample and prototype representations across views, enabling them to be uniformly mapped into a more discriminative and generalizable feature space, thereby significantly enhancing the quality of cross-view feature representations.

Sample-wise Contrastive Learning: To effectively capture semantic consistency across views, we design a sample-level contrastive learning loss. The goal is to encourage consistent representations of the same instance across different views, while pushing apart features from different instances. The loss is defined as:

$$L_i = \frac{1}{V(V-1)N} \sum_{i=1}^{N} \sum_{a=1}^{V} \sum_{\substack{b=1 \\ b \neq a}}^{V} L_i^{a,b}, \tag{4}$$

where $\mathcal{L}_i^{a,b}$ is the contrastive loss between the i-th instance in view a and view b, defined as:

$$L_i^{a,b} = -\log \frac{e^{s(z_i^a, z_i^b)/\tau_I}}{\sum_{j=1}^{N} \left[e^{s(z_i^a, z_j^a)/\tau_I} + e^{s(z_i^a, z_j^b)/\tau_I} \right]}, \quad (5)$$

here, $s(\cdot,\cdot)$ denotes cosine similarity, τ_I is the temperature parameter, z_i^a represents the feature of the i-th sample in view a, and z_i^b is its counterpart in view b.

Prototype-wise Contrastive Learning: The prototype-based attribution paradigm requires that prototypes be capable of capturing both the diversity and the complementarity of different views. After introducing unlabeled data in the formal training stage, it becomes necessary to balance cross-view consistency with diversity and complementarity. To this end, we impose a bounded optimization constraint on the similarity of cross-view prototypes, which not only prevents prototype representations from excessive collapse and preserves view-specific discriminative information, but also achieves moderate consistency among prototypes under shared semantics, thereby enabling richer and more robust cross-view prototype representations. The bounded prototype-level contrastive loss is defined as follows:

$$L_p = \frac{1}{V(V-1)} \sum_{a=1}^{V} \sum_{\substack{b=1 \\ b \neq a}}^{V} \left[\frac{2}{K} \sum_{i=1}^{K} \frac{\left| s\left(u_i^a, u_i^b\right) - \alpha \right|}{\tau_P} + \frac{1}{K^2} \sum_{i=1}^{K} \log \left(e^{\frac{\left| s\left(u_i^a, u_i^b\right) - \alpha \right|}{\tau_P}} + \sum_{\substack{j=1 \\ j \neq i}}^{K} e^{\frac{s\left(u_i^a, u_j^b\right)}{\tau_P}} \right) \right], \quad (6)$$

where $\alpha = 0.75$ is the similarity bound, and $\tau_P = 2.0$ is the temperature parameter.

Attention Regularization: From the computation of the attention weight matrix, each row represents the probability of the i-th sample belonging to the j-th cluster. To enable each sample to confidently indicate which cluster it belongs to, we introduce an attention regularization term to encourage the model to produce clearer and more definite cluster assignments. The regularization term is defined as:

$$L_a = \frac{1}{V} \sum_{v=1}^{V} \sum_{j=1}^{K} \left[A_{\cdot j}^v \log A_{\cdot j}^v - \beta \sum_{i=1}^{N} A_{ij}^v \log A_{ij}^v \right], \quad (7)$$

specifically, $A_{\cdot j}^v = \sum_{i=1}^{N} A_{ij}^v$, and β denotes a weight parameter that controls the trade-off between the sharpness and the uniformity of the attention distribution.

Trusted Semi-Supervised Multi-View Learning

Currently, traditional neural networks and most deep learning models typically use the softmax output as the prediction confidence. However, this often leads to overconfident predictions and lacks interpretability in multi-view opinion aggregation. To address this, this paper introduces an evidential learning framework based on Subjective Logic (SL) on top of prototype-driven feature optimization (Han et al. 2022): optimized features from each view, guided by prototype priors, are mapped through a non-negative activation function (e.g., ReLU) into evidence vectors $e^v = [e_1, \ldots, e_K]$, which are then mapped into the Dirichlet parameter space and aggregated based on the uncertainty of each view. This framework combines the structural guidance of prototype priors with the explicit uncertainty modeling of evidential learning, not only promoting effective feature optimization but also enhancing the interpretability and reliability of multi-view opinion fusion, thereby improving the classification accuracy of the model.

Mapping In The Dirichlet Space: SL defines the correspondence between the The evidence vector $e_i = [e_{ik}, \ldots, e_{ik}]$ corresponds to the parameters of the Dirichlet distribution $\alpha_i = [\alpha_{i1}, \ldots, \alpha_{ik}]$. In our framework, we map the evidence of each view to the parameters of the Dirichlet distribution as

$$\alpha^v = e^v + 1, \quad (8)$$

where $\alpha^v = [\alpha_1^v, \ldots, \alpha_K^v]$. Based on this distribution, the belief mass for each class and the overall uncertainty mass can be computed by

$$b_k^v = \frac{e_k^v}{S^v} = \frac{\alpha_k^v - 1}{S^v}, \quad u^v = \frac{K}{S^v}, \quad (9)$$

where $S^v = \sum_{k=1}^{K} e_k^v + K$. Furthermore, the uncertainty and the K belief masses satisfy the normalization condition $u^v + \sum_{k=1}^{K} b_k^v = 1$:

$$u^v + \sum_{k=1}^{K} b_k^v = 1, \quad u^v \geq 0, \quad b_k^v \geq 0. \quad (10)$$

Opinion Aggregation: According to the formula X, We can obtain the subjective opinion of each sample from each view as $\mathcal{M}_i^v = \left\{ \{b_{ik}^v\}_{k=1}^K, u_i^v \right\}_{v=1}^V$. Based on the Dirichlet-space opinion fusion rule, we aggregate the opinions from all views into a comprehensive joint opinion. This process not only integrates multi-view information but also provides interpretability for the reliability of multi-view classification decisions.

For two views with subjective opinions $\mathcal{M}_i^n = \left\{ \{b_{ik}^n\}_{k=1}^K, u_i^n \right\}$ and $\mathcal{M}_i^m = \left\{ \{b_{ik}^m\}_{k=1}^K, u_i^m \right\}$, the fusion is performed as

$$\mathcal{M}_i = \mathcal{M}_i^n \oplus \mathcal{M}_i^m, \quad n \neq m, \quad (11)$$

where the combined belief masses and uncertainty are calculated by

$$b_{ik} = \frac{b_{ik}^n b_{ik}^m + b_{ik}^n u_i^m + b_{ik}^m u_i^n}{1 - F}, u_i = \frac{u_i^n u_i^m}{1 - F}, n \neq m, \quad (12)$$

with $F = \sum_{i \neq j} b_{ik}^n b_{jk}^m$ representing the conflict degree between the two opinions, and $\frac{1}{1-F}$ being the normalization factor.

The analysis of opinion aggregation across multiple views is as follows: (a) when the predictions from two views are consistent and both exhibit high confidence, the fused opinion will also show high confidence, indicating a highly reliable classification decision for the sample; (b) when both

views have considerable uncertainty, the aggregated opinion will similarly reflect high uncertainty, signaling that the classification decision is less robust; (c) if one view has high confidence while the other is uncertain, the fusion mechanism will tend to rely more on the confident view, thereby maintaining the reliability and robustness of the final classification result.

This fusion rule can be naturally extended to the multi-view setting as

$$\mathcal{M}_i = \mathcal{M}_i^1 \oplus \mathcal{M}_i^2 \oplus \cdots \oplus \mathcal{M}_i^v. \tag{13}$$

Finally, after multi-view fusion, we obtain the comprehensive subjective opinion of each sample as $\mathcal{M}_i = \left\{\{b_{ik}\}_{k=1}^K, u_i\right\}$.

Evidence loss : In evidential learning networks, we follow previous methods, and the overall optimization objective consists of two core loss terms: the classification loss L_{acc} and the regularization loss L_{KL}, whose overall form is:

$$L_e = \sum_{i=1}^{N} \left[\big(L_{\text{acc}}(\alpha_i) + L_{\text{KL}}(\tilde{\alpha}_i)\big) + \sum_{v=1}^{V} \big(L_{\text{acc}}(\alpha_i^v) + L_{\text{KL}}(\tilde{\alpha}_i^v)\big) \right]. \tag{14}$$

First, the classification loss L_{acc} models the evidence vector output by the network using a Dirichlet distribution and treats it as a prior over the likelihood. Specifically, for a sample with a one-hot label y_i, the non-negative evidence vector e_i is converted into the parameter $\alpha_i = e_i + 1$. The classification loss is then computed in the form of a negative log-likelihood, which significantly enhances the evidence corresponding to the true class and thus improves the model's ability to distinguish the target class:

$$\begin{aligned} L_{\text{acc}}(\alpha_i) &= \int \left[-\sum_{k=1}^{K} y_{ik} \log p_{ik} \right] \frac{\prod_{k=1}^{K} p_{ik}^{\alpha_{ik}-1}}{B(\alpha_i)} \, dp_i \\ &= \sum_{k=1}^{K} y_{ik} \big(\psi(S_i) - \psi(\alpha_{ik})\big), \end{aligned} \tag{15}$$

where $S_i = \sum_{k=1}^{K} e_{ik} + K$, $\alpha_i = e_i + 1$, and $\psi(\cdot)$ denotes the digamma function.

Next, the regularization loss $L_{\text{KL}}(\tilde{\alpha}_i)$ is used to suppress the accumulation of evidence for incorrect classes. It minimizes the KL divergence between the predicted distribution and the uniform Dirichlet distribution, thereby shrinking the evidence for non-ground-truth classes toward zero and reducing the model's attention to irrelevant classes:

$$\begin{aligned} L_{\text{KL}}(\tilde{\alpha}_i) &= \lambda_t \, \text{KL}\big[\text{Dir}(p_i \mid \tilde{\alpha}_i) \,\|\, \text{Dir}(p_i \mid \mathbf{1})\big] \\ &= \lambda_t \Bigg(\log \frac{\Gamma\left(\sum_{k=1}^{K} \tilde{\alpha}_{ik}\right) \Gamma(K)}{\prod_{k=1}^{K} \Gamma(\tilde{\alpha}_{ik})} \\ &\quad + \sum_{k=1}^{K} (\tilde{\alpha}_{ik} - 1) \Big[\psi(\tilde{\alpha}_{ik}) - \psi\left(\sum_{m=1}^{K} \tilde{\alpha}_{im}\right)\Big] \Bigg), \end{aligned} \tag{16}$$

here, $\Gamma(\cdot)$ denotes the gamma function, which serves to regulate the shrinkage speed of evidence during the training process.

Overall Loss

By combining all the aforementioned losses, the overall loss of our model is formulated as:

$$L = L_e + \lambda_1 L_i + \lambda_2 L_p + \lambda_3 L_a, \tag{17}$$

where λ_1, λ_2, and λ_3 are the balance coefficients for the instance-level contrastive loss, the prototype-level contrastive loss, and the attention regularization loss, respectively. It is worth noting that this overall loss is applied during the formal training stage. During the pre-training stage, we optimize the evidence loss using only the labeled data and the class centers generated from them.

Experiments

Datasets

We conduct a comprehensive evaluation of the proposed method's classification performance on six widely used datasets: LandUse21 (Yang and Newsam 2010), Cifar10 (Krizhevsky, Sutskever, and Hinton 2012), Animal (Lampert, Nickisch, and Harmeling 2009), MNIST (Deng 2012), ACM (Wang et al. 2019), and Cora (Fang et al. 2023).

Compared Methods

In this study, we compare six representative semi-supervised multi-view classification methods, including MLAN (Nie et al. 2017), LMSSC (Bo et al. 2019), JCD (Zhuge et al. 2020), LACK (Yu et al. 2022), MMatch (Wang et al. 2022b), and TSMVC (Wang et al. 2024).

Implementation Details

In this work, a unified architecture is employed across all datasets. The encoder is implemented as a fully connected network, while the classifier consists of a linear layer followed by a Softplus activation function. The Adam optimizer is used for training, with a weight decay set to 0.00005. To ensure stable training and good generalization, the learning rate is adjusted based on the characteristics of each dataset. Specifically, a learning rate of 0.001 is used for the Animal and Landuse21 datasets, 0.0003 for CIFAR10 and MNIST, and 0.00003 for ACM and Cora. The hyperparameters λ_1, λ_2, and λ_3 are selected from the set $\{0.1, 0.2, 1, 10\}$. Other parameters are fixed as follows: $\tau_I = 0.5$, $\tau_P = 2$, $\alpha = 0.75$, $\beta = 0.02$, $\lambda_t = 1$. The model is implemented in Visual Studio using Python and trained in an environment equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB) and 32 GB of RAM. For all the datasets used, we run the experiments under 10 different random seeds and report the mean of the results. The evaluation metric is classification accuracy (ACC), where higher values indicate better performance.

Performance Evaluation

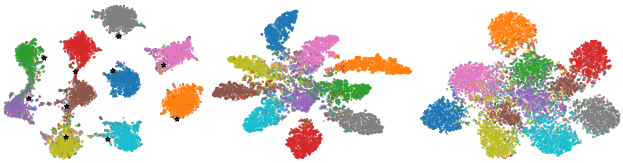
We compare the proposed NCPD-TSMVC method with six semi-supervised multi-view classification methods, and present the experimental results of all methods under different datasets and label rate conditions in Table 1. To further verify the robustness of NCPD-TSMVC in noisy environments, we also provide the t-SNE visualization comparison

Dataset	Rate	MLAN	LMSSC	JCD	LACK	MMatch	TSMVC	Ours
MNIST	0.1	95.74	95.81	96.77	95.94	99.26	99.24	**99.36**
	0.2	96.03	96.74	97.26	96.77	99.28	99.25	**99.38**
	0.3	96.12	97.02	97.95	97.08	99.33	99.29	**99.38**
	0.4	97.26	97.14	98.53	98.19	99.30	99.30	**99.39**
Cifar10	0.1	94.81	96.73	97.16	96.91	97.27	99.35	**98.38**
	0.2	95.46	97.24	98.08	97.53	98.30	**98.75**	99.43
	0.3	96.94	97.86	98.34	98.18	98.57	**98.80**	99.74
	0.4	97.26	98.33	98.47	98.31	98.58	98.82	**98.84**
Animal	0.1	24.07	25.97	27.49	26.88	31.46	27.88	**35.51**
	0.2	27.46	27.81	28.37	28.67	38.94	31.05	**45.44**
	0.3	29.89	29.95	29.44	29.74	43.31	33.42	**50.83**
	0.4	33.48	32.42	32.05	32.41	46.51	34.94	**53.14**
Landuse21	0.1	34.65	36.18	37.79	36.56	**44.65**	36.86	41.33
	0.2	41.19	41.52	40.94	41.29	50.88	43.56	**52.74**
	0.3	43.51	44.23	44.09	44.43	54.20	46.21	**58.69**
	0.4	44.17	46.37	45.59	46.79	55.60	47.80	**63.69**
ACM	0.1	67.49	68.34	73.02	71.28	79.65	**81.58**	80.12
	0.2	71.15	74.22	77.48	77.32	83.88	85.13	**85.29**
	0.3	74.34	76.53	79.94	80.14	85.20	87.04	**87.50**
	0.4	77.08	79.04	82.56	83.69	88.60	88.92	**89.54**
Cora	0.1	37.45	38.62	39.07	38.26	**51.65**	50.42	41.33
	0.2	43.61	45.89	46.15	44.52	55.93	57.25	**57.56**
	0.3	45.53	47.93	48.39	48.72	59.02	62.03	**66.59**
	0.4	49.71	51.84	43.26	53.02	63.50	66.92	**70.28**

Table 1: Classification accuracy (%) comparison under different labeled rates. The best results are shown in **bold**, and the second-best results are underlined.

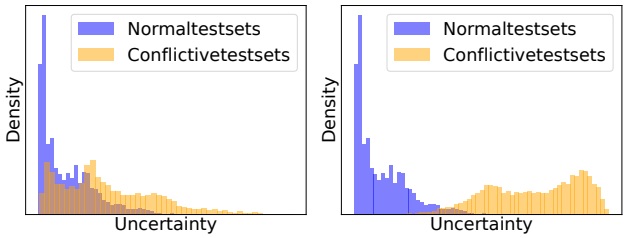
of NCPD-TSMVC with TSMVC and MMatch in Figure 3. From the experimental results, it can be observed that:

- NCPD-TSMVC achieves the best classification performance across most settings compared to all baseline methods. However, under low labeled sample rates, its performance on certain complex datasets is slightly inferior to some competing methods. This is mainly because NCPD-TSMVC relies on labeled data to construct the prototype weights. When the proportion of labeled data is too low, the learned prototypes may be insufficiently representative, limiting their ability to effectively guide the learning of unlabeled samples.
- From the t-SNE visualizations, we can observe that even in noisy environments, NCPD-TSMVC is still able to learn highly discriminative feature spaces. Compared with TSMVC and MMatch, our method exhibits clearer inter-class separation and intra-class compactness on the test samples, further demonstrating the effectiveness and robustness of the proposed method in feature learning.

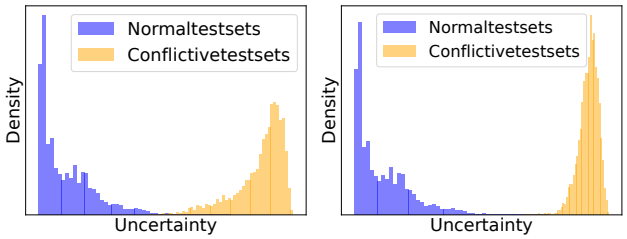


(a) NCPD-TSMVC (b) TSMVC (c) MMatch

Figure 3: t-SNE visualization comparison on the MNIST dataset with Gaussian noise ($\sigma = 1$) during the test phase.



(a) $\sigma = 0.1$ (b) $\sigma = 0.5$



(c) $\sigma = 1$ (d) $\sigma = 10$

Figure 4: Density of uncertainty.

	Components				Landuse21
ID	$\mathcal{L}_e$	$\mathcal{L}_i$	$\mathcal{L}_p$	$\mathcal{L}_a$	ACC
(A)	√		√	√	61.43
(B)	√	√		√	63.33
(C)	√	√	√		64.52
(D)	√	√	√	√	**66.19**

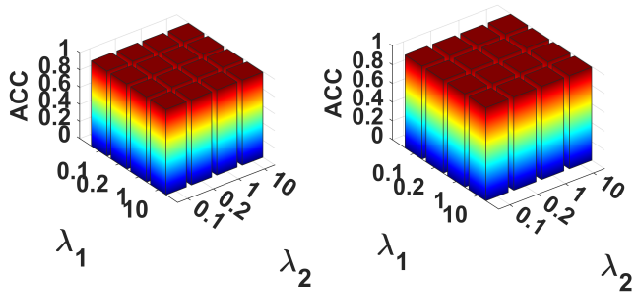
Table 2: Ablation study on the Landuse21 dataset.

Uncertainty Estimation

To provide an intuitive assessment of the uncertainty estimation capability of the proposed method, We visualize the uncertainty distributions of both in-distribution and out-of-distribution (OOD) samples on the Cifar10 dataset. Specifically, 50% of the training set is used as labeled data to train the model, while the test set is split evenly into two subsets. One subset is treated as in-distribution data, and the other is perturbed by adding scaled Gaussian noise $0.2 \times \mathcal{N}(0, \sigma)$ with standard deviations $\sigma = 0.1,\ 0.5,\ 1$ and 10 to simulate OOD samples. The experimental results are shown in Figure 4. When the noise level is low ($\sigma = 0.1$), the uncertainty distribution of the perturbed samples remains close to that of the clean samples. However, as the noise level increases, the uncertainty associated with the perturbed samples also rises. These results indicate that the proposed NCPD-TSMVC framework effectively distinguishes in-distribution data from OOD instances.

Ablation Study

This study conducts an ablation experiment on the Landuse21 dataset to verify the effectiveness of three loss functions: the sample-based contrastive loss ($\mathcal{L}_i$), the prototype-based contrastive loss ($\mathcal{L}_p$), and the attention regularization loss ($\mathcal{L}_a$), and Table 2 presents the experimental results under different combinations of these losses, and it is worth noting that the labeled evidence learning loss ($\mathcal{L}_e$) is used as the base loss and is not ablated, and the experimental results show that removing any one of the loss functions leads to a performance degradation while the model achieves the best performance when all three losses are employed simultaneously, and furthermore by comparing (A) with (B) and (C) a relatively large performance drop can be observed which indicates that although introducing the prototype-based contrastive loss ($\mathcal{L}_p$) effectively constrains prototype at a global level, from the perspective of feature distribution the sample-based contrastive loss ($\mathcal{L}_i$) directly pulls together samples of the same class and pushes apart those of different classes in the local feature space thus contributing more significantly to improving intra-class compactness and inter-class separability, and these findings demonstrate that all three loss functions are indispensable.



(a) $\lambda_3 = 0.1$ (b) $\lambda_3 = 1$

Figure 5: We conducted a parameter sensitivity analysis experiment on the MNIST dataset with a labeled rate of 0.1.

Parameter Sensitivity Analysis

Owing to space constraints, we only include two representative sensitivity plots for our model on the MNIST dataset with a label rate of 0.1 in Figure 5. These plots characterize how key hyperparameters affect clustering performance and reveal smooth, small-magnitude variations, providing initial evidence of the model's robustness and stability.

Conclusion

This study proposes a trusted semi-supervised multi-view classification framework (NCPD-TSMVC) driven by prototype priors derived from the Neural Collapse phenomenon. By leveraging the stability and high discriminability of these priors, the framework effectively calibrates the feature representations of unlabeled data, significantly mitigating the impact of noise and redundant information introduced by unlabeled data during training. We also provide a theoretical justification for the proposed prototype priors. Furthermore, in the multi-view information fusion stage, we design a trusted fusion strategy based on evidential learning, which not only improves decision accuracy but also enhances the interpretability of the model. Finally, we validate the effectiveness and superiority of NCPD-TSMVC on six benchmark datasets as well as datasets with added Gaussian noise.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 62576298), the Sichuan Science and Technology Program (Grant no. 2025ZNSFSC0474, 2024ZDZX0004), and the Mianyang Science and Technology Program (Grant no. 2025ZYDF096).

References

Andrew, G.; Arora, R.; Bilmes, J.; and Livescu, K. 2013. Deep canonical correlation analysis. In *International conference on machine learning*, 1247–1255. PMLR.

Bo, X.; Kang, Z.; Zhao, Z.; Su, Y.; and Chen, W. 2019. Latent multi-view semi-supervised classification. In *Asian Conference on Machine Learning*, 348–362. PMLR.

Chen, Z.; Zhang, M.; Cui, S.; Li, H.; Niu, G.; Gong, M.; Zhang, C.; and Zhang, K. 2024. Neural collapse inspired feature alignment for out-of-distribution generalization. *Advances in Neural Information Processing Systems*, 37: 93671–93689.

Deng, L. 2012. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6): 141–142.

Fang, S.-G.; Huang, D.; Cai, X.-S.; Wang, C.-D.; He, C.; and Tang, Y. 2023. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8): 11436–11447.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 2551–2566.

Hu, S.; Zhang, C.; Zou, G.; Lou, Z.; and Ye, Y. 2024. Deep multiview clustering by pseudo-label guided contrastive learning and dual correlation learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Kan, M.; Shan, S.; and Chen, X. 2016. Multi-view deep network for cross-view classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4847–4855.

Kang, Z.; Xie, X.; Li, B.; and Pan, E. 2024. CDC: A simple framework for complex data clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE conference on computer vision and pattern recognition*, 951–958. IEEE.

Li, H.; Hu, P.; Zhang, Q.; Peng, X.; Yang, M.; et al. 2025. Test-time Adaptation for Cross-modal Retrieval with Query Shift. In *The Thirteenth International Conference on Learning Representations*.

Li, H.; Li, Y.; Yang, M.; Hu, P.; Peng, D.; and Peng, X. 2023a. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*.

Li, X.; Pan, Y.; Sun, Y.; Sun, Q.; Sun, Y.; Tsang, I. W.; and Ren, Z. 2024a. Incomplete multi-view clustering with paired and balanced dynamic anchor learning. *IEEE Transactions on Multimedia*.

Li, X.; Pan, Y. P.; Sun, Y.; Sun, Q. S.; Tsang, I. W.; and Ren, Z. 2024b. Fast unpaired multi-view clustering. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*.

Li, X.; Ren, Z.; Sun, Q.; and Xu, Z. 2023b. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognition*, 134: 109083.

Li, X.; Sun, Y.; Sun, Q.; Ren, Z.; and Sun, Y. 2023c. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100: 101941.

Li, Y.; Hu, P.; Peng, D.; Lv, J.; Fan, J.; and Peng, X. 2024c. Image Clustering with External Guidance. In *International Conference on Machine Learning*, 27890–27902. PMLR.

Lou, Z.; Xue, H.; Wang, Y.; Zhang, C.; Yang, X.; and Hu, S. 2025. Parameter-Free Deep Multi-Modal Clustering With Reliable Contrastive Learning. *IEEE Transactions on Image Processing*.

Luo, X.; Chen, P.; Liu, C.; Jin, X.; Wen, J.; Liu, Y.; and Wang, J. 2025a. Enhancing Multimodal Protein Function Prediction Through Dual-Branch Dynamic Selection with Reconstructive Pre-Training. *arXiv preprint arXiv:2511.04040*.

Luo, X.; Pu, Z.; Xu, Y.; Wong, W. K.; Su, J.; Dou, X.; Ye, B.; Hu, J.; and Mou, L. 2021. MVDRNet: Multi-view diabetic retinopathy detection by combining DCNNs and attention mechanisms. *Pattern Recognition*, 120: 108104.

Luo, X.; Xu, Q.; Wang, Z.; Huang, C.; Liu, C.; Jin, X.; and Zhang, J. 2024. A lesion-fusion neural network for multi-view diabetic retinopathy grading. *IEEE Journal of Biomedical and Health Informatics*.

Luo, X.; Xu, Q.; Wu, H.; Liu, C.; Lai, Z.; and Shen, L. 2025b. Like an Ophthalmologist: Dynamic Selection Driven Multi-View Learning for Diabetic Retinopathy Grading. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 19224–19232.

Luo, X.; Xu, Y.; and Yang, J. 2019. Multi-resolution dictionary learning for face recognition. *Pattern Recognition*, 93: 283–292.

Nie, F.; Cai, G.; Li, J.; and Li, X. 2017. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 27(3): 1501–1511.

Papyan, V.; Han, X.; and Donoho, D. L. 2020. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40): 24652–24663.

Qin, Y.; Zhang, X.; Yu, S.; and Feng, G. 2025. A survey on representation learning for multi-view data. *Neural Networks*, 181: 106842.

Ren, Z.; Li, X.; Mukherjee, M.; Huang, Y.; Sun, Q.; and Huang, Z. 2021. Robust multi-view graph clustering in latent energy-preserving embedding space. *Information Sciences*, 569: 582–595.

Wan, X.; Liu, J.; Gan, X.; Liu, X.; Wang, S.; Wen, Y.; Wan, T.; and Zhu, E. 2024a. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3): 5774–5786.

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

Wang, C.; Xu, R.; Lu, K.; Xu, S.; Meng, W.; Zhang, Y.; Fan, B.; and Zhang, X. 2023. Attention weighted local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 10632–10649.

Wang, H.; Peng, J.; Chen, D.; Jiang, G.; Zhao, T.; and Fu, X. 2020a. Attribute-guided feature learning network for vehicle reidentification. *IEEE MultiMedia*, 27(4): 112–121.

Wang, J.; Zhang, L.; Wang, Q.; Chen, L.; Shi, J.; Chen, X.; Li, Z.; and Shen, D. 2020b. 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, Q.; Tao, Z.; Gao, Q.; and Jiao, L. 2022a. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5): 7244–7250.

Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2021a. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31: 556–568.

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

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference*, 2022–2032.

Wang, X.; Wang, Y.; Wang, Y.; Huang, A.; and Liu, J. 2024. Trusted semi-supervised multi-view classification with contrastive learning. *IEEE Transactions on Multimedia*, 26: 8268–8278.

Wang, X.; Zhang, Y.; Zhang, J.; and Zhou, Y. 2025. Incomplete Multiview Clustering using Discriminative Feature Recovery and Tensorized Matrix Factorization. *IEEE Transactions on Circuits and Systems for Video Technology*.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025. Bidirectional Probabilistic Multi-graph Learning and Decomposition for Multi-view Clustering. *IEEE Transactions on Image Processing*.

Wang, X.-l.; Zhu, Z.-f.; Song, Y.; and Fu, H.-j. 2021b. GR-Net: Graph-based remodeling network for multi-view semi-supervised classification. *Pattern Recognition Letters*, 151: 95–102.

Wen, J.; Long, J.; Lu, X.; Liu, C.; Fang, X.; and Xu, Y. 2025a. Partial Multiview Incomplete Multilabel Learning Via Uncertainty-Driven Reliable Dynamic Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wen, W.; Gong, T.; Dong, Y.; Yu, S.; and Zhang, W. 2025b. Towards the Generalization of Multi-view Learning: An Information-theoretical Analysis. *arXiv preprint arXiv:2501.16768*.

Wen, Y.; Wang, S.; Liao, Q.; Liang, W.; Liang, K.; Wan, X.; and Liu, X. 2023. Unpaired multi-view graph clustering with cross-view structure matching. *IEEE Transactions on Neural Networks and Learning Systems*.

Xu, C.; Si, J.; Guan, Z.; Zhao, W.; Wu, Y.; and Gao, X. 2024a. Reliable conflictive multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 16129–16137.

Xu, C.; Zhao, W.; Zhao, J.; Guan, Z.; Song, X.; and Li, J. 2022. Uncertainty-aware multiview deep learning for internet of things applications. *IEEE Transactions on Industrial Informatics*, 19(2): 1456–1466.

Xu, Q.; Luo, X.; Huang, C.; Liu, C.; Wen, J.; Wang, J.; and Xu, Y. 2024b. HACDR-Net: Heterogeneous-aware convolutional network for diabetic retinopathy multi-lesion segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 6342–6350.

Yang, J.; Chung, H.; and Jang, I. 2024. Uncertainty-Weighted Mutual Distillation for Multi-View Fusion. *arXiv e-prints*, arXiv–2411.

Yang, M.; Huang, Z.; Hu, P.; Li, T.; Lv, J.; and Peng, X. 2022. Learning with twin noisy labels for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14308–14317.

Yang, Y.; and Newsam, S. 2010. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 270–279.

Yu, Y.; Zhou, G.; Huang, H.; Xie, S.; and Zhao, Q. 2022. A semi-supervised label-driven auto-weighted strategy for multi-view data classification. *Knowledge-Based Systems*, 255: 109694.

Yuan, H.; Sun, Y.; Zhou, F.; Wen, J.; Yuan, S.; You, X.; and Ren, Z. 2025. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*.

Zhao, L.; Wang, X.; Liu, Z.; Wang, Z.; and Chen, Z. 2024. Learnable graph guided deep multi-view representation learning via information bottleneck. *IEEE Transactions on Circuits and Systems for Video Technology*.

Zhu, Z.; Ding, T.; Zhou, J.; Li, X.; You, C.; Sulam, J.; and Qu, Q. 2021. A geometric analysis of neural collapse with unconstrained features. *Advances in Neural Information Processing Systems*, 34: 29820–29834.

Zhuge, W.; Hou, C.; Peng, S.; and Yi, D. 2020. Joint consensus and diversity for multi-view semi-supervised classification. *Machine Learning*, 109(3): 445–465.