

# A review on multi-view learning

Zhiwen YU (✉)<sup>1,2</sup>, Ziyang DONG<sup>1,2</sup>, Chenchen YU<sup>1,2</sup>, Kaixiang YANG<sup>1</sup>,  
Ziwei FAN<sup>1,2</sup>, C. L. Philip CHEN<sup>1</sup>

1 School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

2 Pengcheng Laboratory, Shenzhen 518066, China

© The Author(s) 2024. This article is published with open access at [link.springer.com](http://link.springer.com) and [journal.hep.com.cn](http://journal.hep.com.cn)

**Abstract** Multi-view learning is an emerging field that aims to enhance learning performance by leveraging multiple views or sources of data across various domains. By integrating information from diverse perspectives, multi-view learning methods effectively enhance accuracy, robustness, and generalization capabilities. The existing research on multi-view learning can be broadly categorized into four groups in the survey based on the tasks it encompasses, namely multi-view classification approaches, multi-view semi-supervised classification approaches, multi-view clustering approaches, and multi-view semi-supervised clustering approaches. Despite its potential advantages, multi-view learning poses several challenges, including view inconsistency, view complementarity, optimal view fusion, the curse of dimensionality, scalability, limited labels, and generalization across domains. Nevertheless, these challenges have not discouraged researchers from exploring the potential of multi-view learning. It continues to be an active and promising research area, capable of effectively addressing complex real-world problems.

**Keywords** multi-view learning, multi-view clustering, ensemble learning, semi-supervised learning

## 1 Introduction

In the era of big data, the methods and perspectives for data acquisition have become increasingly diverse. For instance, an individual can describe something from various visual angles; an image can be represented by different features such as colors, text; a social media user can be characterized by their behavioral data, social network graph, and textual content. These data, describing entities from different perspectives, are referred to as multi-view data. Different views represent various aspects and features of the data, yet they are intertwined and fundamentally consistent. Simply concatenating multiple view features and applying traditional machine learning methods not only leads to the curse of dimensionality, making it difficult to uncover latent information, but also overlooks the interrelations among

multiple views. Therefore, multi-view learning is gradually becoming a well-established domain within machine learning that tackles problems involving the availability of multiple views or sources of data.

The utilization of these multiple views provides an opportunity to leverage complementary information, thereby resulting in enhanced learning performance compared to single-view methods. For instance, Wei et al. [1] conducted a study demonstrating that multi-view learning methods outperform single-view approaches when applied to both uni-view and multi-view surface electromyography data streams. Similarly, Tian et al. [2] delved into the realm of multi-view deep feature learning and achieved superior performance in the context of Encephalogram (EEG)-Based Epileptic Seizure Detection when compared to single-view deep feature learning techniques. In another study, Kong et al. [3] employed a multi-view learning method using deep models for action recognition, attaining improved results compared to a single-view learning algorithm. Furthermore, by taking into account the deep Gaussian process, Sun et al. [4] presented a new multi-view representation learning approach, exhibiting promising performance on real-world multi-view datasets. Additionally, in image classification tasks, Zhang et al. [5] investigated multi-view visual classification methods and attained noteworthy performance gains. Overall, integrating different views during the learning process helps capture diverse aspects of the data, leading to heightened accuracy, robustness, and generalization capabilities.

In the exploration of multi-view learning, researchers have been striving to find an effective method to fully utilize the information among views. With the deepening research on multi-view learning, it is gradually recognized that the key to multi-view learning lies in balancing the consistency and complementarity between views [6], namely the two principles of multi-view learning: 1) *Consensus principle*: The consensus principle refers to the consistency among different views. This means that multiple views should agree to some extent in order to effectively capture common characteristics and patterns in the data. The basic idea of the consensus principle is that there exists some degree of correlation among different views, which can be leveraged to improve learning performance. For example, in a multi-view dataset containing

both images and text descriptions, there may be some correlation between the objects in the images and the content described in the text. 2) *Complementarity principle*: The complementarity principle refers to the complementary or supplementary nature of different views. This means that multiple views should provide different yet related information to collectively enhance learning performance. The basic idea of the complementarity principle is that each view may contain information about specific aspects or local details of the data, and combining this information can provide a more comprehensive and rich data representation. For example, in image recognition tasks, in addition to the pixel information of the images themselves, text descriptions of the images or semantic labels associated with the images can be utilized to improve recognition accuracy. Both two principles jointly facilitate the effective integration and utilization of information across multiple views. While the consensus principle emphasizes the consistency and commonality among different views, the complementary principle emphasizes the complementarity and diversity between them. In order to effectively leverage both two principles, a series of methods were proposed by simultaneously considering these two principles. Luo et al. [7] decomposed the representations of each view into an all views' shared consensus representation and a set of view-specific distinctive representations. Wang et al. [8] leveraged complementary information between different views by introducing a position-aware exclusive term, while ensuring their consistency based on a common indicator matrix. Liang et al. [9] decomposed the graphs of each view into consistent and inconsistent parts to fuse the unified graph by introducing an inconsistency measurement.

Based on the time of multi-view fusion during the learning process, it can be divided into two primary types: early fusion and late fusion [6,10–18], as illustrated in Figs. 1 and 2. In

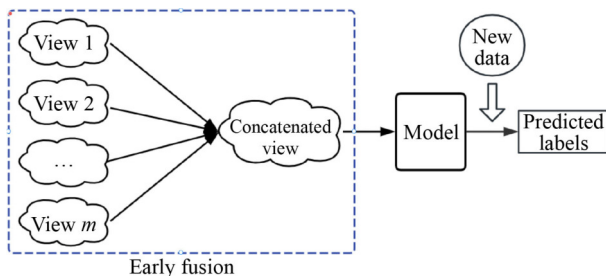


Fig. 1 The early fusion scheme for multi-view learning

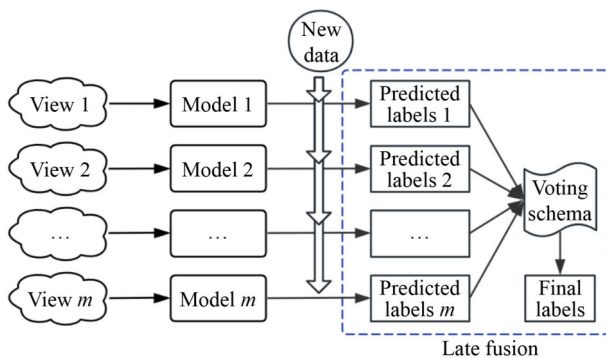


Fig. 2 The late fusion scheme for multi-view learning

early fusion methods, the representations from different views are concatenated or aggregated, and on the fused representation, a single model is trained using a joint objective function that integrates information from multiple views. This approach assumes equal contribution of all views to the label predictions. In contrast, late fusion combines the predictions from different views after training separate models for each view. The final predictions are derived by combining the outputs of each classifier that has been trained on each view. Fusion can be performed through averaging probabilities, voting schemes, or more advanced fusion techniques such as stacking or boosting, similar to ensemble learning approaches [19–25]. Both early fusion and late fusion methods consider the consensus and complementarity between views. The learned models are encouraged to agree on shared aspects while maintaining diversity in their respective views.

In recent years, several reviews on multi-view learning have been published, discussing the theories, methodologies, classifications, and applications of multi-view learning methods. With a focus on classification and clustering tasks, Table 1 provides an overview of different surveys on multi-view learning. Many surveys on methods primarily concentrate on multi-view classification tasks. For instance, Xu et al. [6] reviewed multi-view learning approaches and classified these approaches into three categories: co-training, multiple kernel learning, and subspace learning. Li et al. [10] provided an extensive survey on multi-view representation learning, encompassing topics such as multi-view representation alignment and multi-view representation fusion. They reviewed representative methods and theories, explored various approaches including correlation-based alignment and generative/neural network-based fusion, and highlighted important applications in the field. In the domain of hyperspectral image classification, Li et al. [11] provided an extensive review and summarized the methods into three steps: multi-view construction, interactivity enhancement, and multi-view fusion. Zhao et al. [12] presented a comprehensive survey on recent developments in multi-view learning, dividing them into three categories: co-training methods, co-regularization methods, and margin-consistency methods. Yan et al. [13] conducted an extensive survey on the topic of deep multi-view learning, covering representative methods in this context, including auto-encoders, conventional neural networks, brief networks, canonical correlation analysis, matrix factorization, and information bottleneck.

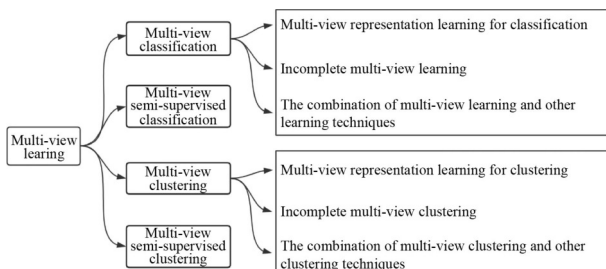
Some reviews on multi-view learning methods specifically pay attention to multi-view clustering. For instance, Yang et al. [14] presented a survey and introduction to multi-view clustering, presenting a taxonomy of techniques based on mechanisms and principles. They categorized them into five types: co-training style, multi-kernel learning, graph clustering, subspace clustering, and multi-task clustering. Fu et al. [15] reviewed multi-view clustering approaches, summarizing them as three classes: graph-based models, space learning-based models, and binary code learning-based models. Wen et al. [16] studied incomplete multi-view clustering (IMC) and categorized it into five groups: matrix factorization-based IMC, kernel learning-based IMC, graph

**Table 1** The overview of different surveys for multi-view learning

References	Task	Categories
Xu et al. [6]	Multi-view classification	co-training, multiple kernel learning, subspace learning
Li et al. [10]	Multi-view classification	multi-view representation alignment, multi-view representation fusion(multi-modal topic learning, multi-view sparse coding, multi-view latent space Markov networks, multi-modal autoencoders, multi-view convolutional neural networks, and multi-modal recurrent neural networks)
Li et al. [11]	Multi-view classification	multi-view construction methods, interactivity enhancement methods, multi-view fusion methods
Zhao et al. [12]	Multi-view classification	co-training, co-regularization, margin-consistency
Yan et al. [13]	Multi-view classification	multi-view auto-encoder, multi-view conventional neural networks, multi-view deep brief networks, deep multi-view canonical correlation analysis, deep multi-view matrix factorization, deep multi-view information bottleneck
Yang et al. [14]	Multi-view clustering	co-training, multi-kernel learning, graph clustering, subspace clustering, multi-task clustering
Fu et al. [15]	Multi-view clustering	graph based model, space learning based model, binary code learning based model
Wen et al. [16]	Multi-view clustering	matrix factorization-based incomplete multi-view clustering (IMC), kernel learning-based IMC, graph learning-based IMC, deep learning-based IMC
Chao et al. [17]	Multi-view clustering	generative methods, discriminative methods (five groups: common eigenvector matrix, common coefficient matrix, common indicator matrix, direct combination, combination after projection)
Fang et al. [18]	Multi-view clustering	heuristic-based approaches, neural network-based approaches
Ours	Multi-view classification Multi-view clustering	multi-view classification methods, multi-view semi-supervised classification methods, multi-view clustering methods, multi-view semi-supervised clustering methods

learning-based IMC, and deep learning-based IMC. Chao et al. [17] formulated a new taxonomy, categorizing them into generative and discriminative classes. Fang et al. [18] conducted a comprehensive review of recent multi-view clustering techniques, categorizing them into heuristic-based and neural network-based approaches.

In summary, existing multi-view learning methods can be classified from different perspectives. Most reviews focus on specific tasks, such as classification or clustering, or on specific methodologies, such as representation learning or deep learning, and they primarily classify and explain existing methods based on algorithm principles or styles. However, since few surveys systematically elucidate existing methods based on learning paradigms, this survey emphasizes the perspective of learning paradigms, categorizing existing multi-view learning research into four groups: multi-view classification methods, multi-view semi-supervised classification methods, multi-view clustering methods and multi-view semi-supervised clustering methods. In this survey, the basic concepts, technologies, methods and categorizations of multi-view learning are comprehensively discussed. Moreover, the relationships and differences between groups are also analyzed. In addition, this survey also elaborates on the current applications and challenges of multi-view learning. We hope that this survey, through its clear categorization, will help researchers and practitioners better understand the different methods of multi-view learning and their applicable scenarios, thereby aiding in the selection of appropriate techniques and methods for specific problems. [Figure 3](#)

**Fig. 3** The outline of the survey on multi-view learning

presents an overview of the survey structure, highlighting the main sections and their organization.

The subsequent sections of the survey are structured as follows. Section 2 provides a comprehensive review of multi-view classification methods, covering the various approaches and algorithms employed in this area. Section 3 expounds the concept of multi-view semi-supervised learning and discusses the approaches used in this context. Section 4 focuses on multi-view clustering approaches, providing an in-depth review of the different methods employed for clustering data from multiple views. Section 5 delves into multi-view semi-supervised clustering methods, exploring the techniques and algorithms utilized to perform clustering in scenarios where labeled and unlabeled data from multiple views are available. The applications of multi-view learning are introduced in Section 6. We address the challenges associated with multi-view learning, discussing the limitations and open research problems in the field in Section 7. Finally, in Section 8, we conclude the survey by summarizing the key findings and highlighting potential future directions for research and application in multi-view learning. By following this organized structure, the survey seeks to present a thorough knowledge of multi-view learning, its various methodologies, and the challenges that lie ahead.

## 2 Multi-view classification

Multi-view classification, where the input data consists of multiple views or sources, aims to classify or predict the labels of the samples based on these multiple views. Each view offers a distinctive perspective or representation of the data. By considering multiple views, the classifier can potentially capture more comprehensive and discriminative information.

Seven crucial processes make up the multi-view categorization process, as shown in [Fig. 4](#). Each step plays a crucial role in achieving effective classification results:

In the data acquisition and representation step, multi-view data is gathered from a variety of sources or sensors, and each view represents a distinct aspect or perspective of the data.

In the feature extraction step, relevant features are extracted from each view to capture discriminative information.

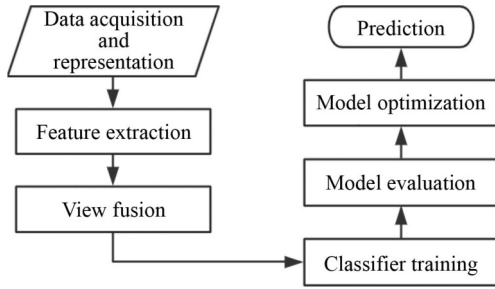


Fig. 4 The process of multi-view classification

Techniques such as dimensionality reduction, filtering, or view-specific transformations may be employed [26–30], aiming to retain important discriminative characteristics.

In the view fusion step, the features derived from different views are combined into a single representation that captures complementing information. View fusion techniques, such as early fusion and late fusion, exploit correlations and dependencies between views. Common fusion techniques include concatenation, weighted averaging, and learning-based approaches.

In the classifier training step, the fused representation is utilized to train a classifier. The classifier learns the mapping between the fused features and the corresponding class labels. The specific problem and data characteristics determine the classifier to use. Popular classifiers include support vector machines, random forests, neural networks, and ensemble methods [31–34]. By integrating the initial steps, it is possible to develop an end-to-end deep multi-view classification approach within deep learning frameworks.

In the model evaluation step, after training the classifier, its performance is assessed using suitable metrics, including accuracy, precision, recall, F1 score, and other relevant evaluation measures. Additionally, considerations are given to consistency, complementarity, and fusion quality. Consistency assesses the agreement or alignment between different views, complementarity measures the extent to which views provide unique and non-redundant information, and fusion quality evaluates the effectiveness of fusion methods in combining multiple views. Cross-validation or a separate validation set is typically used to assess performance and generalization ability.

In the model optimization and refinement step, hyperparameter tuning, regularization techniques, or optimization methods are used to augment the model's discriminative capability. Overall performance can be improved by fine-tuning the classifier or refining the fusion process.

In the prediction step, the trained model can be utilized to make predictions and assign class labels to new, unseen samples with multiple views after being trained and optimized.

In summary, multi-view classification is a challenging task, but when approached systematically, it can enhance overall classification performance by utilizing the consensus between views.

Multi-view classification faces numerous challenges that need to be addressed: (1) How to capture a more comprehensive representation of the data, leading to better

performance in learning tasks. (2) How to handle missing or unreliable data across the views. (3) How to leverage the advantages of multi-view learning and other learning technique simultaneously. (4) How to apply multi-view learning algorithm to solve various classification problems in different multi-view scenarios. As a way to solve these challenges, the researchers take into account the multi-view representation learning, the incomplete multi-view learning, the combination of multi-view learning and other learning techniques, and the applications of multi-view learning, and they will be introduced in the following subsections.

## 2.1 Multi-view representation learning for classification

In recent years, multi-view representation learning has gained significant attention as a powerful technique for understanding complex data with multiple perspectives. In many real-world scenarios, data is represented in various ways since it is inherently multi-faceted. For instance, in computer vision, an image have visual, textual, and spatial views, while text data can be represented in terms of words, characters, or syntax in natural language processing. By considering multiple views, a model can capture diverse aspects of data that may be missed by single-view approaches, leading to enhanced performance and generalization. Multi-view representation learning aims to extract informative and complementary representations from different views and combine these representations to attain a model that exhibits both higher accuracy and increased robustness.

The core concept of multi-view representation learning for classification is to utilize label information from multi-view data and complementary information between different views to jointly learn a representation space that reflects the intrinsic structure of the data. This approach attempts to address the problem where a single view may not fully capture the complexity of the data by integrating information from multiple perspectives to enhance classification performance. A mainstream schema of multi-view representation learning for classification is shown in Fig. 5. Firstly, multi-view representation learning requires learning representations for different views and integrate them, adopting traditional representation learning methods or neural networks, while preserving the uniqueness and complementarity of each view. Secondly, in classification tasks, the representation learning process is guided by utilizing label information to ensure that the learned representations can differentiate between different classes and possess good generalization ability. This is

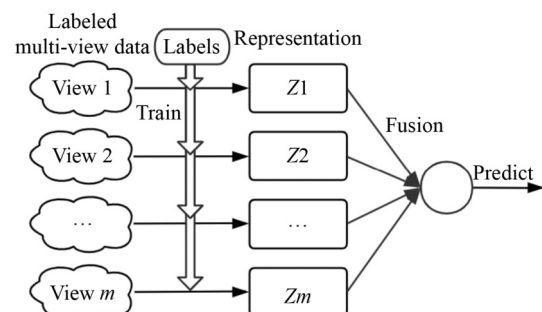


Fig. 5 The scheme of multi-view representation learning for classification

typically achieved by incorporating label-related terms into the optimization objective function, such as minimizing reconstruction error or utilizing discriminative constraints. Additionally, regularization terms may be included to prevent overfitting and ensure that the representation space possesses the desired properties, such as sparsity or low-rankness.

Traditional machine learning methods typically employ linear or nonlinear shallow models to integrate information from multiple views, such as kernel methods, graph models, or subspace learning. These approaches are often combined with dimensionality reduction techniques like principal component analysis, linear discriminant analysis, and non-negative matrix factorization to learn data representations. Such methods are known for their high computational efficiency, simple model structures, and ease of interpretation. For instance, Based on maximum entropy discrimination, Chao et al. [35] proposed a novel multi-view classification method which learns a common subspace representation from multiple views and augments it with the original features, combining the principles of consensus and complementarity. This method leverages the principle of maximum entropy discrimination, thereby fully utilizing the advantages of discriminative estimation. However, it is only applicable to two views and cannot be extended to more-views cases. Guan et al. [36] investigated the field of multi-view concept learning and devised a novel method for nonnegative latent representation learning, aiming to extract conceptual factors from multi-view data. This method learns a shared latent space representation through graph embedding regularization and sparsity constraints. It encourages samples of the same class to be close to each other in the latent space, while samples of different classes are encouraged to be distant, thereby capturing the semantic relationships between samples. Wang et al. [37] investigated multi-view analysis dictionary learning, a framework that incorporates semantic information into the fundamental dictionary learning model for image classification purposes. Liu et al. [38] studied multi-view dictionary learning based on the consensus between views. Both two methods employ dictionary learning to obtain sparse representations for samples in each view and then explore consensus between different views through regularization terms, aiming to make the learned dictionaries from different views as close as possible. However, such approach is sensitive to parameter selection and the scale of the data.

Moreover, some methods utilize deep neural networks to learn representations. Different from traditional machine learning methods, deep learning methods are more effective in handling large-scale, high-dimensional data. They can learn complex patterns and structures from vast amounts of data, improving feature extraction and representation learning. However, deep methods often require substantial computational resources and have poorer interpretability compared to traditional methods. For instance, Jia et al. [39] provided a multi-view representation learning method that draws inspiration from human collective intelligence and group decision-making processes. By facilitating multi-round view communication, this approach enables each view to utilize additional information from other views and achieve

mutual assistance. But the increase in dataset size and the number of views may exacerbate computational efficiency and resource consumption in multi-round communication. Zheng et al. [40] introduced the collaborative unsupervised multi-view representation learning method, which uses auto-encoders to learn view-specific compact representations and then integrates them into a unified representation. This method captures the high-order view correlations by utilizing a low-rank tensor constraint and achieves competitive multi-view representation compared to other methods, showing its effectiveness in the field. Ma et al. [41] presented a novel multi-attributed-view graph convolutional auto-encoder model that effectively handles multiple attributed views in complex data by learning node-pairwise proximity and embeddings. Although this method introduces a multi-attribute view proximity measurement, it may not fully explore the complex relationships between different attribute views. Sun et al. [4] presented a deep Gaussian process-based multi-view representation learning approach in their research, which adopts a deep network composed of a series of Gaussian processes to learn the latent representations of each view. Compared to standard DGPs, multi-view DGPs face challenges in simultaneously learning representations for different views. Huang et al. [42] proposed the multi-view Laplacian network, an extension of the multi-view spectral representation learning approach, which aims to enhance the representative capacity of deep learning models.

Researchers also explore representation learning techniques tailored to specific types of datasets and domains. Yang et al. [43] built a multi-view deep auto-encoder with tensor factorization for image representation learning. Zhang et al. [44] applied a hierarchical attentive multi-view learning model to extract multi-scale features from the keyframe perspective, with the specific aim of coronary artery stenosis quantification. Lyu et al. [45] investigated the application of a multi-view group representation learning framework in the area of location-aware group recommendation. Based on Kolmogorov-Smirnov, Tan et al. [46] designed a representation learning approach for default prediction on imbalanced and complex loan multi-view datasets. These examples highlight the diverse range of research in multi-view representation learning, addressing the challenges and nuances associated with different datasets and application domains.

Overall, multi-view representation learning has shown great promise in extracting rich and complementary information from diverse data sources. With ongoing advancements and novel methodologies, it is expected to play a crucial role in various domains, enabling deeper insights and more accurate predictions.

## 2.2 Incomplete multi-view learning

In many real-world applications, data may not be entirely available. For example, certain features might be missing for some samples, or the data may originate from different sensors, each of which can only observe a subset of the sample's attributes. Incomplete multi-view learning frameworks acknowledge this incompleteness, aiming to learn useful information from it and impute missing samples.

Incomplete multi-view learning refers to the scenario where not all views of the data are available or complete for every instance. In this setting, the multi-view data may have missing or partially observed views, resulting in incomplete information for learning and modeling, thus complete multi-view learning methods are no longer applicable. Incomplete multi-view learning aims to develop methods and techniques that can effectively handle and utilize the available information from the incomplete views to make accurate predictions or perform other learning tasks. As shown in Fig. 6, the goal is to integrate multiple incomplete views, utilizing partial complete information from each view and merging them while mitigating the negative impact of missing samples during the learning process to construct a more comprehensive global representation. For example, an indicator matrix is introduced to only consider the reconstruction loss of observed instances, or the model is trained with only observed instances. Meanwhile, it is necessary to handle missing data, which involves imputing missing values. This may include ignoring missing views, directly setting missing values to zero, using information from other views to fill in missing values, or inferring missing sample values using the learned unified representation. For classification tasks, it is worth noting that even if an instance is missing on some views, its label information is still considered available by most methods. Incomplete multi-view learning utilizes complete samples and label information to learn representations for each view, enabling the model to learn representations that differentiate between labels. Then, it integrates representations from other complete views of missing samples to generate or infer values for the missing views.

Researchers have proposed various multi-view learning algorithms to tackle the challenges associated with incomplete views. For examples, Qin et al. [47] introduced the noise-aware incomplete multi-view learning networks framework. This approach reduces the influences of missing views, develops a consistent and useful representation while mitigating the effects of inherent noise. The framework considers the issue of view quality arising from inherent noise, exhibiting robustness on noisy data. However, it relies on the selection of parameters related to the noise distribution. Lin et al. [48] developed a unified framework that addresses the challenges of consistency learning and missing view recovery in incomplete multi-view representation learning. This method

utilizes the minimization of conditional entropy to recover missing views while simultaneously maximizing the mutual information between different views to learn consistent representations. However, the method only experiments with two views, and further research is needed to extend it to multiple views. Xu et al. [49] explored an incomplete multi-view learning method, assuming that all views share a common subspace. They leverage the assumption to learn a joint representation and handle incomplete views through iteratively updating representations by using a successive over-relaxation method. Zhu et al. [50] designed a latent heterogeneous graph network that incorporates a neighborhood constraint and a view-existence constraint for incomplete multi-view learning. This approach leverages the graph structure to exploit relationships between views and handle missing information.

In addition to handling incomplete views, some researchers also address the issue of missing labels. To address the dual incompleteness issue of multi-view features and labels, Wen et al. [51] developed an incomplete multi-view multi-label learning network. This method integrates missing information and explores available data and label information through view-specific feature extraction, weighted fusion, and classification modules. Although this method combines available information from different views, simply ignoring missing values and labels and only utilizing known information may perform poorly in cases where the missing proportion is significant. Furthermore, Liu et al. [52] presented an instance-level contrastive network for this case, building an end-to-end multi-view feature extraction framework using stacked auto-encoders. During the handling of missing cases, this method employs an instance-level contrastive learning approach to recover missing values. However, for dealing with missing labels, it only introduces indicator matrices to mitigate loss caused by missing labels. Li et al. [53] presented a concise yet effective multi-view learning model to address problems caused by three incomplete cases: incomplete views, missing labels, and non-aligned views. Their model takes into account these challenges and aims to learn robust representations and accurate predictions. These examples demonstrate the diverse range of approaches taken by researchers in incomplete multi-view learning. These algorithms enhance the robustness and effectiveness of multi-view learning approaches in real-world scenarios by addressing these challenges.

In summary, incomplete multi-view learning involves leveraging multiple views for learning, where each view provides partial information, and specialized algorithms are used to handle missing or unreliable data across the views, resulting in improved performance in classification and clustering tasks.

### 2.3 The combination of multi-view learning and other learning techniques

Combining multi-view learning with other learning techniques can be a fruitful approach to enhance model performance and expand its capabilities. The specific combination of techniques depends on the problem domain, data characteristics, and

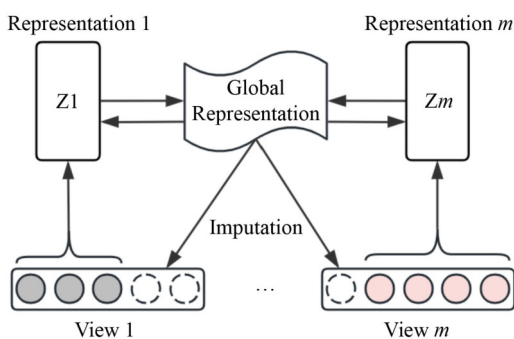


Fig. 6 The scheme of incomplete multi-view learning

learning task goals. Several methods are summarized as follows:

**Multi-task multi-view learning:** Multi-task multi-view learning is an extension of traditional multi-view learning that incorporates the concept of learning multiple tasks simultaneously. In this context, each view represents a different representation or feature set of the same underlying data, while each task corresponds to a different prediction problem or learning objective that needs to be solved. The key idea behind multi-task multi-view learning is to leverage the complementary information present in multiple views and the relationships between tasks to improve the overall learning performance. For multi-variate time series forecasting, Deng et al. [54] proposed a multi-view multi-task learning approach. Their framework assigns specific affine transformations and normalization to each task from multiple views in time and space, aiding the model in adaptively extracting multi-view and multi-task information during prediction. He et al. [55] proposed a graph-based multi-task multi-view learning framework, addressing both feature heterogeneity and task heterogeneity. By constructing a graph model, this framework effectively integrates the correlations between tasks and the consistency between views, requiring multiple related tasks to simultaneously share common views and have specific task views. Zhang et al. [56] proposed an inductive learning framework, ensuring the model learns similar functions across different views for each task by introducing regularization. Additionally, they considered scenarios involving missing views and uneven relationships between tasks.

**Multi-view multi-label learning:** To overcome the issue of missing labels, researchers have concentrated on multi-view multi-label learning, where each instance in the dataset is linked with multiple labels. This setting is common in real-world applications where data is often heterogeneous and complex. For example, Zhao et al. [57] explored the multi-view multi-label classification tasks and explicitly extracted view-specific label information and low-rank label structure from misaligned views within a unified model framework. Zhang et al. [58] proposed a multi-view multi-label learning method with the basis of sparse feature selection, capturing discriminative features according to label correlations and view relations. Yuan et al. [59] addressed the multi-view partial multi-label learning problem by embedding the learned unified similarity graph into the label disambiguation process. Based on label correlation, Liu et al. [60] designed a multi-view multi-label learning method and applied manifold regularization terms to preserve the intrinsic structure of data samples in low-dimensional space. In summary, multi-view multi-label learning is a specialized variant of multi-view learning that proves valuable in scenarios where instances have multiple associated labels. Combining multi-view learning with multi-label techniques can lead to improved performance and better generalization, particularly when dealing with missing labels.

Additionally, There are also many other methods that combine multi-view learning with other learning methods or technical frameworks. For example, Li et al. [61] integrates multiple context structures into a unified framework, explored

the concept of multi-view multi-instance learning. Xu et al. [62] presented the multi-view intact space learning approach, which combines multiple views of data with the aim of learning a latent intact representation, demonstrating its effectiveness through theoretical analysis and experiments. Hu et al. [63] developed a multi-view metric learning framework which learns the combination of different distance metrics from multi-view representations. Wu et al. [64] conducted research on online multi-view learning using knowledge registration units as a basis for their investigation. Fan et al. [65] studied the problem of incomplete multi-view learning under label shift, which estimates the importance weight by learning the bidirectional complete representation of multi-view data, achieving label distribution alignment between the source domain and the target domain. Fu et al. [66] developed a transductive multi-view zero-shot learning system to address the concerns of projection domain shift and prototype sparsity. Shi et al. [67] presented a method by extending the broad learning system to multi-view learning. Yan et al. [68] studied a multi-view-oriented multiple kernel learning technique, capturing complex hierarchical information. Huang et al. [69] integrated federated learning into multi-view learning, aiming to address the training of machine learning models on multi-view data across multiple devices in distributed networks. Researchers continue to explore and develop new methods by combining multi-view learning with various learning techniques to address diverse challenges.

### 3 Multi-view semi-supervised classification

In real life, the acquisition of annotated data is often time-consuming and expensive. How to enhance the performance of multi-view methods with limited label information is an issue worth studying. Multi-view semi-supervised learning refers to a learning paradigm where the task is to leverage both labeled and unlabeled data from multiple views to enhance the predictive model's performance. In this setting, although each sample has multiple views, its labels are publicly unique. Different from supervised methods, how to make full use of limited labels and combine unlabeled data and multi-view information is the key to it.

A limited amount of labeled data can be used for training, while a large amount of unlabeled data is used to enhance the learning process by capturing shared information or discovering the underlying structure in different views. Since labels are shared across views, a key idea is to learn a consistent representation of views. By aligning data from different views into a common representation space, label propagation or pseudo-label assignment is achieved while capturing complementary information. Furthermore, leveraging additional unlabeled data can improve model generalization and robustness, especially when labeled data is scarce or expensive to acquire.

Recently, there has been increasing attention from researchers on designing new multi-view semi-supervised learning methods. For instance, Nie et al. [70] presented an auto-weighted multi-view learning algorithm that combines semi-supervised classification and local structure learning.

Based on their ability to discriminate, these views are given weights by this algorithm, which incorporates them into the learning process. Additionally, they proposed two other methods [71,72], the core idea of which is constructing a shared label indicator matrix and learning a common representation from multiple views. By utilizing graph regularization and the learned common representation, labels from fixed labeled data are propagated to unlabeled data. Xu et al. [73] studied a multi-view weakly labeled learning approach, it effectively utilizes weakly labeled multi-view data. Their method generates pseudo-label vectors and incorporates different strategies and iterations to leverage the weak labels. Wang et al. [74] developed a semi-supervised discriminative representation learning approach which simultaneously learns class probabilities of training data and view-specific representations for multi-view classification. This method enhances classification performance by integrating discriminative information from multiple views. Chao et al. [75] presented a multi-view semi-supervised learning approach on the basis of maximum entropy discrimination, which enhances classification performance by leveraging the geometric information of unlabeled data through expected Laplacian regularization. Zhang et al. [76] proposed a fast multi-view semi-supervised learning approach based on an anchor-based strategy. This approach accelerates the learning process and enhances performance by effectively leveraging labeled and unlabeled data in a multi-view context. Using an embedding regularizer learning approach, Huang et al. [77] presented a multi-view semi-supervised classification method, which manipulates data with limited labels by learning an embedding regularizer to guide the learning process. Wang et al. [78] presented a deep sparse regularizer learning model. To handle the challenges of semi-supervised learning in a multi-view scenario, their model adaptively learns data-driven sparse regularizers. Qian et al. [79] presented a framework for semi-supervised dimension reduction in multi-label and multi-view learning settings. Their approach explicitly models the information combining mechanism to effectively reduce dimensionality and enhance classification performance.

Some researchers have explored multi-view graph-based semi-supervised learning approaches. Zheng et al. [80] explored a multi-view graph-based semi-supervised extreme learning machine. Their approach leverages graph-based regularization to boost learning of information from multiple views. Guo et al. [81] designed a unified robust graph-based semi-supervised multi-view learning scheme. Their approach reduces the impact of noise and makes use of the understanding of multi-view data to increase the model's robustness and effectiveness. Li et al. [82] proposed a multi-view semi-supervised learning scheme with the basis of a relaxation regularization term. Their method learns a well-structured unified graph that captures the relationships between instances and incorporates it into the semi-supervised learning process. These examples highlight the ongoing efforts of researchers to develop effective multi-view semi-supervised learning approaches, leveraging the complementary between views and incorporating unlabeled data to enhance

classification performance and handle real-world scenarios where labeled data is limited.

Some research works explore the application of multi-view semi-supervised learning algorithms in different research areas. For example, Jia et al. [83] proposed a semi-supervised multi-view deep discriminant representation learning approach that effectively leverages the consensus and complementary properties of multi-view data. Their method learns shared and specific representations, addresses redundancy, and incorporates unlabeled data to improve performance in webpage, image, and document classification tasks. Cui et al. [84] presented a multi-view semi-supervised learning approach with the basis of a multi-objective scheme for large-vocabulary continuous speech recognition. Their approach leverages multiple views of speech data and incorporates unlabeled data to enhance speech recognition performance. Thammasorn et al. [85] designed the multi-view triplet network based on limited data for medical imaging data classification. Their method utilizes multiple views of medical imaging data and incorporates unlabeled samples for image classification tasks.

In summary, the key advantage of multi-view semi-supervised learning is its ability to leverage limited labeled samples to capture more information and enhance the performance of predictive models. By incorporating both labeled and unlabeled data, the models can learn more robust and discriminative representations that generalize better to unseen instances. However, multi-view semi-supervised learning methods face challenges in two key aspects: First, when addressing high-dimensional data, the curse of dimensionality can become more pronounced. As the number of views and dimensions increases, effectively leveraging the unlabeled data to learn meaningful representations becomes more challenging. Second, designing effective algorithms and models for semi-supervised multi-view learning requires careful consideration of feature fusion, data alignment, and model regularization techniques. These factors add complexity to the learning process, making it more intricate and computationally demanding.

Addressing these challenges requires innovative approaches that handle high-dimensional data and effectively incorporate multiple views and unlabeled data while ensuring model efficiency and generalization performance. Researchers continue to explore and develop new techniques to overcome these challenges and advance the field of multi-view semi-supervised learning.

## 4 Multi-view clustering

Multi-view clustering is a subfield of machine learning that focuses on clustering data using multiple views or representations. Different from multi-view classification, the idea behind multi-view clustering is to cluster samples from different views to uncover latent structures and similarities in the data without specifying sample categories or labels in advance. The goal of multi-view clustering is to group samples into different clusters, where samples within the same cluster exhibit high similarity, while samples from different clusters show distinct differences. By integrating information

from different views, multi-view clustering can capture the diversity and complexity of the data, thus better revealing the data's intrinsic structure and features.

Unlike traditional clustering algorithms that operate on a single view of data, multi-view clustering leverages the complementary data from multiple views to improve clustering accuracy and robustness. For examples, Xie et al. [86] presented a deep multi-view clustering method and demonstrated its superiority over single-view baseline methods for image clustering. By utilizing multiple views of image data, their approach improved the clustering performance and captured more fine-grained patterns. Liang et al. [87] showed that multi-view graph learning has better robustness and effectiveness compared to single-view for clustering tasks. Their method effectively integrates views' information into a unified graph representation, enhancing the clustering performance by capturing diverse relationships between data points. Huang et al. [88] explored the multi-view spectral clustering approach, which outperformed single-view clustering methods for face recognition tasks. Considering multiple views of face images, their method improved the discriminative power and robustness of the clustering process. Tang et al. [89] demonstrated that multi-view unsupervised feature selection approaches outperform single-view approaches. By incorporating multiple views, their approach identified relevant and complementary features between views, resulting in improved clustering performance and better representation of the data. In general, different views capture distinct perspectives of the data, thus combining them can provide a more comprehensive understanding of complex patterns. Multi-view clustering techniques exploit the complementary nature of multiple views, enabling more accurate and robust clustering results. These methods contribute to uncovering underlying structures and discovering valuable insights in various applications.

The procedure of multi-view clustering typically involves the following steps [90–97]: data representation, view integration, similarity or dissimilarity computation, multi-view clustering algorithm, cluster validation and evaluation, refinement and iteration, as shown in Fig. 7. Specifically, the data from each view is represented in a suitable format at first. This may involve preprocessing steps such as feature extraction, dimensionality reduction, or normalization for each view separately. The goal is to depict the raw data in a way that captures the essential characteristics of each view.

The next step is to integrate the multiple views into a unified representation. This can be achieved by concatenating the representations from each view into a single feature vector or

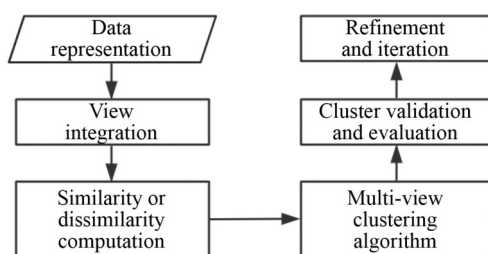


Fig. 7 The process of multi-view clustering

by using more advanced techniques such as feature fusion, subspace learning, or kernel methods. The goal is to combine the complementary information from different views to enhance the performance.

Once the views are integrated, the similarity or dissimilarity matrix between the data points is computed, which is comparable to clustering ensemble [25,98–101]. This matrix quantifies the pairwise relationships between the data points via their integrated representations. Various similarity measures can be used, such as Euclidean distance, cosine similarity, or correlation coefficients, with the basis of the nature of the data and the unique needs of the clustering technique.

Once the similarity or dissimilarity matrix has been computed, a multi-view clustering approach is used to partition the data into clusters. There are several algorithms specifically designed for multi-view clustering, such as co-regularized spectral clustering, consensus clustering, or multiple kernel k-means. These algorithms leverage the information from multiple views to discover consistent and robust clusters.

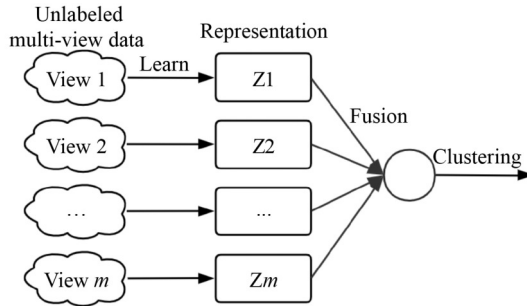
After clustering, it is important to evaluate the clusters obtained. There are various effective metrics, such as clustering stability, cluster compactness, or external validation measures if ground truth labels are available. The evaluation helps assess the model effectively and provides insights into the clustering performance.

Depending on the results of the evaluation, it may be necessary to refine the clustering procedure or revisit the earlier steps. This could involve adjusting the data representation, integrating additional views, fine-tuning the similarity/dissimilarity computation, or exploring alternative clustering algorithms. Iteration allows for improving the clustering results iteratively.

By following these steps, multi-view clustering seeks to leverage the complementary information from multiple views to realize more accurate and comprehensive clustering results compared to single-view clustering techniques.

#### 4.1 Multi-view representation learning for clustering

In the task of clustering, multi-view representation learning is essential for obtaining informative and complementary representations from multiple views. Unlike multi-view representation learning for classification, multi-view representation learning for clustering is unsupervised learning. The core idea of multi-view representation learning for clustering is to leverage information from multiple views to better discover the latent structure and similarities within the data, without relying on predefined labels. The scheme of multi-view representation learning for clustering is shown in Fig. 8. Its key lies in how to integrate data from different views to obtain a representation space that reveals the underlying structure of the data, thus enhancing clustering performance. Specifically, it requires designing suitable optimization objectives to maximize the clustering performance of the data. This may involve considering weights for different views, regularization terms, and enhancing the model's robustness to noise.



**Fig. 8** The scheme of multi-view representation learning for clustering

Researchers have proposed various multi-view clustering algorithms based on representation learning. For instance, Chen et al. [102] proposed a multi-view representation learning approach for clustering data streams. Their approach incorporates collaborative representation, individual global affinity matrix construction, and fused sparse affinity matrix calculation. By effectively capturing the evolving patterns and structures of data streams, their method shows an excellent performance in dynamic environments. Zhao et al. [103] presented an efficient multi-view dictionary learning method for multi-view clustering. By utilizing a partially shared model with a flexible ratio of shared sparse coefficients and a differentiable scale-invariant function as the sparsity regularizer, their approach achieves improved clustering performance. Zheng et al. [104] presented a method that addresses the challenge of integrating underlying information from multiple views. Their approach leverages graph-guided unsupervised multi-view representation learning, which effectively captures the relationships and dependencies between views. This leads to improved clustering and classification performance compared to existing methods.

Additionally, another multi-view representation learning method for clustering is called multi-view subspace clustering. By adopting self-representation of the raw data, the model learns the relationship between samples, thereby converting the dimension of each sample from original features to sample size, which is an effective way to achieve clustering of high-dimensional datasets. After learning the self-representation matrix of each view, the representations of each view are fused to construct a common similarity matrix, and then spectral clustering is performed to obtain the clustering results. Recently, researchers have proposed various multi-view subspace clustering methods. For instance, Zheng et al. [105] explored large-scale multi-view clustering based on subspace representation learning. Their method leverages the underlying subspaces of multiple views to effectively capture the data's intrinsic structure. Their approach improves clustering accuracy and scalability in large-scale scenarios by considering the subspaces. Zhang et al. [106] proposed a low-rank tensor constrained multi-view subspace clustering. By introducing low-rank tensor constraints, they stack the subspace representations of all views to construct a 3D tensor, learning the subspace representations of each view and exploring high-order correlations between views, thus improving consistency. Although this method enhances the similarity between views by constructing tensors, it does not

consider the differences between views. Cao et al. [107] introduced the Hilbert-Schmidt Independence Criterion (HSIC) as a measure to evaluate the diversity of subspace representation matrices from different views, thereby enhancing diversity between views. However, this also minimizes the consistency between views as much as possible, making it difficult to extract consensus information. Zhang et al. [108] assumed that all views can be mapped from a latent low-dimensional common representation, thereby enhancing the consistency among views. They then employ low-rank representation on this latent space to ensure that samples within the same cluster are as close as possible.

Overall, by leveraging multi-view representation learning, clustering algorithms can benefit from a more comprehensive and diverse understanding of the data. These algorithms capture different aspects and perspectives, resulting in improved clustering accuracy and robustness. Multi-view representation learning techniques play a crucial role in enhancing the effectiveness of clustering algorithms across various domains and applications.

#### 4.2 Incomplete multi-view clustering

Incomplete multi-view clustering is challenging because there exist missing or incomplete views which may result in information loss and affect the clustering results. Handling this incompleteness requires methods that can effectively utilize the available information from the incomplete views while mitigating the impact of missing data. Specifically, during the model training or representation learning phase, similar to incomplete multi-view learning, most incomplete multi-view clustering methods only utilize observed samples to separately learn different views and construct a common representation. However, during the missing data completion phase, incomplete multi-view learning generate missing values based on the global representation learned from other views of the missing samples, while incomplete multi-view clustering relies more on learning the latent relationships or similarity information between samples. It utilizes these relationships to find several samples most similar to the unobserved instances and recover their original values or representation based on this similarity.

Researchers have proposed various methods to handle missing or partially available views in multi-view clustering tasks. For instance, Chao et al. [109] presented a multi-view co-clustering approach by introducing an indicator matrix to ignore missing values, thus reducing the impact of incomplete views. Their method can handle any pattern of incomplete data and performs excellently in the treatment study of opioid dependence which would frequently deal with the large number of missing data items. Furthermore, to address the case of missing arbitrary values in views, they decomposed the problem into a two-stage algorithm, including multiple imputation and ensemble clustering [110]. The algorithm first completes the missing values and then treats it as complete multi-view data to weighted ensemble clustering. Fang et al. [111] studied incomplete multi-view clustering by considering view variation and view heredity. Their approach takes into account the variations between views and the relationships

between data samples. Yang et al. [112] introduced a new incomplete multi-view clustering method that specifically deals with the challenges posed by partially view-unaligned and partially sample-missing data. They utilize a robust contrastive learning paradigm that handles false negatives caused by random sampling, improving the robustness. Liu et al. [113] introduced a fast incomplete multi-view clustering method that utilizes view-independent anchors learned from the diversity of distribution among each incomplete view. By constructing a unified anchor graph, their approach efficiently handles large-scale tasks with improved complexity and effectiveness compared to other multi-view clustering methods. Zhang et al. [114] introduced an isomorphic linear correlation analysis method and an identical distribution pursuit completion model. Their approach is developed for feature-level completion of missing views in multi-view data. For incomplete multi-view clustering, Yin et al. [115] designed a learning latent embedding by weighted projection matrix alignment. Their method incorporates a view completion model and enforces alignment of different projection matrices to cluster centers, addressing the limitations of existing approaches. Shang et al. [116] introduced a novel generalized incomplete multi-view clustering method that combines latent representation learning, spectral embedding, and optimal graph clustering. Their approach leverages multiple techniques to handle incomplete views and improve clustering performance. Chao et al. [117] proposed a multi-view clustering method, which utilizes high-confidence guidance and graph convolutional networks to handle incomplete data and simultaneously leverages complementary and consistent information for end-to-end clustering optimization. This approach effectively integrates the handling of missing data across multiple views, representation learning, and cluster assignment, thereby enhancing clustering performance.

In general, the technique chosen is determined by the specific characteristics of the incomplete multi-view data and the objectives of the clustering task. It is important to carefully handle the missing or incomplete views to ensure that the clustering results are meaningful and robust. Researchers continue to develop innovative methods to address the challenges of incomplete multi-view clustering and improve the performance of clustering algorithms in the actual world.

#### 4.3 The combination of multi-view clustering and other learning techniques

By integrating multi-view clustering with traditional clustering algorithms, we can leverage the complementary aspects of each approach and potentially achieve more accurate and robust clustering results. Here are a few examples of how multi-view clustering can be combined with other learning techniques:

**Multi-view graph clustering:** The core idea of multi-view graph clustering is to utilize graph structure information from multiple data views for clustering. It represents different data views as graphs, where each node in the graph represents a data sample, and edges represent relationships or similarities between samples. By integrating information from multiple

graphs, multi-view graph clustering aims to discover the intrinsic structure and patterns of the data, thereby assigning samples to appropriate clustering clusters. For example, Wang et al. [118] studied multi-view clustering based on multi-order structured graph learning. By incorporating structured graph learning, their approach captures the complex relationships between data points across different views, resulting in improved accuracy. Wang et al. [119] proposed parameter-free weighted multi-view projected clustering approach. Their approach effectively addresses the challenges of high-dimensional heterogeneous data by combining structured graph learning and dimensionality reduction, leading to improved performance. Xia et al. [120] introduced a multi-view clustering algorithm with the basis of tensorized bipartite graph learning. By considering both inter-view and intra-view similarity, their approach addresses the drawbacks of existing methods and improves the clustering results. Jiang et al. [121] designed a tensorial multi-view clustering method that incorporates a low-rank tensor constraint and high-order graph learning techniques. Their method allows for the efficient synthesis of crucial information from multiple views, leading to enhanced clustering accuracy. For multi-view graph learning, Huang et al. [122] introduced a unified framework that combines multi-view consistency and diversity. Their approach leverages both consistency and diversity measures by effectively integrating information from multiple views.

**Multi-task multi-view clustering:** Multi-task multi-view clustering is a method that integrates multiple data views and multiple clustering tasks. It not only utilizes information from different views but also considers the objectives of multiple clustering tasks. For instance, Zhang et al. [123] proposed multi-task multi-view clustering, which combines the advantages of both multi-task clustering and multi-view clustering. Their approach simultaneously considers the shared and individual structures across multiple views, leading to improved clustering performance. Furthermore, Zhang et al. [124] independently cluster samples and features within each view of each task, aiming to learn shared information between tasks. They simultaneously minimize the clustering results between different views to enhance consistency, thereby addressing the multi-task multi-view clustering problem.

**Multi-view spectral clustering:** Multi-view spectral clustering performs spectral clustering on the similarity matrices constructed from subspace learning or graph learning. This method often incorporates an indicator matrix-based graph regularization term into the objective function, integrating spectral clustering with multi-view representation learning into a unified framework. For instance, Jiang et al. [125] developed a graph-based auto-weighted multi-view consensus spectral clustering. By continuously updating similarity matrices and optimizing different weights, their approach addresses the limitations of existing multi-view learning approaches and improves clustering accuracy. Mei et al. [126] proposed a new multi-order similarity learning model, that effectively captures the local structure, adjacent structure, specificity and consensus of views.

The decision to combine multi-view clustering with other techniques is driven by the specific characteristics of the data,

the availability and quality of the views, and the objectives of the clustering task. More attention has to be attached to the strengths and limitations of both approaches and design an integration strategy that maximizes the benefits and overcomes the challenges of the data at hand. Researchers continue to explore and develop innovative methods that combine multi-view clustering with other techniques to address various challenges in clustering tasks.

## 5 Multi-view semi-supervised clustering

Multi-view semi-supervised clustering is a variant of clustering. Unlike multi-view semi-supervised classification methods, multi-view semi-supervised clustering methods aim to partition similar samples into the same cluster based on some prior knowledge. This prior knowledge can be labels or pairwise constraints between samples, such as must-link and cannot-link constraints, indicating whether samples belong to the same cluster. By using the prior knowledge as a hard constraint to force the learned representations of two samples to be same (different), if they have must-link (cannot-link). It combines the benefits of utilizing multiple views and incorporating partial supervision or prior knowledge into the clustering process. The core idea is to utilize information from different data views and partially labeled data or prior knowledge to discover the latent structure and similarity within the data, and to assign data samples to different clustering clusters.

In multi-view semi-supervised clustering, the data is represented by multiple views, each capturing different aspects or perspectives of the data. The clustering algorithm utilizes both unsupervised information from the views and labeled information for a subset of the data points. Here are some examples of research works in this area. Qin et al. [127] proposed a multi-view semi-supervised clustering algorithm based on structured subspace learning. In their approach, specific affinity matrices of views are normalized to a shared affinity matrix, and they use the indicator matrix constructed from labeled data to guide the learning of corresponding representations in the common subspace, promoting the formation of block diagonal structure. This idea enables the efficient integration of information from multiple views, leveraging the available labeled data to enhance the performance. Zhu et al. [128] explored a semi-supervised multi-view spectral clustering method. Their approach leverages pre-set labels and tensor minimization to uncover hidden mutual information in multiple views. It achieves improved clustering performance compared to existing methods while maintaining a relatively fast computational complexity, indicating its potential for various applications. Zhang et al. [129] proposed a tensorized semi-supervised multi-view subspace clustering, where must-link constraints are introduced as hard constraints to enforce consistency of representations for data in the same cluster, while the remaining unconstrained data are learned separately as a subspace. Tang et al. [130] studied a novel regularization approach that integrates weakly supervised sample pair constraints into multi-view subspace clustering, such as must-link and cannot-link. This approach improves performance

and gives semi-supervised subspace clustering an adaptable framework. Its efficacy has been extensively tested on datasets from the actual world.

In summary, by integrating multiple views and incorporating labeled information, multi-view semi-supervised clustering can leverage the complementary information and limited supervision to achieve improved clustering results. By effectively combining unsupervised and supervised learning, these approaches enhance clustering performance and capture the underlying structure of the data more accurately.

## 6 Applications of multi-view learning

Multi-view learning finds applications in various domains, including multimedia analysis, human activity recognition, medical diagnosis, traffic monitoring, fraud detection, and more. It proves to be valuable in scenarios where leveraging different perspectives or data sources can enhance learning and decision-making capabilities.

**1) Multimedia analysis:** In the past few years, multi-view learning has been widely utilized in the field of multimedia analysis, including image annotation, audio recognition, video classification, and more. By considering multiple views, such as visual, textual, and acoustic information, multi-view learning improves the accuracy and richness of multimedia analysis models. Here are some examples: For facial expression identification, Zhang et al. [131] developed a deep neural network-based multi-view method. Their method extracted SIFT features from facial images and employed a well-designed DNN model, resulting in improved performance compared to existing methods on two non-frontal facial expression databases. For 3D form analysis, Wei et al. [132] used a multi-view based graph convolutional network, leveraging multiple views of shape data to enhance the understanding and analysis of complex 3D objects. For image manipulation detection, Dong et al. [133] developed multi-view multi-scale supervised networks. Their approach utilizes multiple views of images at different scales to improve the detection accuracy of manipulated images. In multimedia analysis tasks, integrating information from different aspects is essential to enhance performance and gain a comprehensive understanding of the data. Allowing simultaneous consideration of these views during the learning process, multi-view learning provides a framework to leverage the complementary nature of multiple views. This makes it possible for the models to obtain more detailed and reliable representations of the multimedia data, leading to improved analysis and recognition capabilities. Overall, multi-view learning plays a critical role in advancing multimedia analysis by effectively combining multiple views and exploiting the synergies between them to improve performance and enable more comprehensive data understanding.

**2) Human activity recognition:** Multi-view learning finds applications in human activity recognition systems, particularly in activity monitoring for healthcare or sports applications. By considering multiple sensor modalities such as accelerometers, gyroscopes, and video cameras, multi-view learning enhances the accuracy of activity recognition and enables a more detailed and comprehensive understanding of

human behaviors. For example, for the purpose of recognizing human activity, Tran et al. [134] developed the multi-view discriminant analysis approach, leveraging multiple views of sensor data to improve action recognition performance. Wang et al. [135] explored multi-view multi-instance learning for 3D action recognition, utilizing multiple views of action sequences to capture diverse spatio-temporal patterns. For video-based person re-identification, Chen et al. [136] presented a multi-view metric learning method, considering multiple camera views to enhance person matching across different video sequences.

**3) Medical diagnosis:** In the domain of medical diagnosis, multi-view learning is employed to integrate heterogeneous medical data sources, including patient demographics, medical imaging, genetic information, and electronic health records. By considering multiple views, multi-view learning enhances disease diagnosis, prognosis, and treatment prediction. Here are some examples. To enhance the accuracy of seizure detection, Yuan et al. [137] developed a multi-view deep learning method to address the electroencephalogram (EEG) seizure detection problem. Yang et al. [138] proposed a multi-view multi-scale convolutional neural network for ECG classification, leveraging multiple views of ECG signals at different scales to capture global and local cardiac patterns. Puyol-Antón et al. [139] applied a multi-view learning approach to detect patients with dilated cardiomyopathy, integrating cardiac imaging data from different sources to enhance the accuracy of disease detection. Zhang et al. [140] designed a multi-view learning approach with  $l_1$ -norm co-regularization to predict the drug-induced QT prolongation effect, using diverse views of drug data to enhance the prediction accuracy. To sum up, in the context of medical data, each view can provide valuable insights into different aspects of the same patient or disease. Multi-view learning techniques aim to effectively integrate these diverse perspectives, enabling a comprehensive understanding and enhancing the diagnostic and predictive capabilities of medical systems.

**4) Traffic monitoring:** Multi-view learning is important in traffic monitoring and the development of intelligent transportation systems. By integrating data from multiple sources such as traffic sensors, vehicle probes, weather information, and incident reports, multi-view learning models enable real-time traffic monitoring, adaptive signal control, dynamic route guidance, and efficient resource allocation. Here are some examples of its applications in recent years. Jin et al. [141] conducted research on multi-view vehicle re-identification based on a multi-center metric learning framework. Their approach leverages multiple views to improve the accuracy of vehicle identification in traffic surveillance systems. Zhu et al. [142] studied contrastive multi-view learning and its applications in vehicle recognition using carrier-free ultrawideband radar. Their method utilizes multiple views to enhance vehicle recognition capabilities. For adaptive traffic signal control, Ge et al. [143] combined multi-view encoders with multi-agent transfer reinforcement learning and applied them. By leveraging multiple views and reinforcement learning techniques, they optimize traffic signal

control strategies to improve traffic flow efficiency. Yang et al. [144] investigated multi-view learning based on quadruplet loss for baggage re-identification. Their approach utilizes multiple views of baggage images to enhance re-identification accuracy in security settings. The applications of multi-view learning in the traffic domain lead to improved traffic efficiency, reduced congestion, and enhanced safety. By integrating and leveraging information from various sources, these models contribute to more effective and intelligent transportation systems, enabling better management and optimization of traffic conditions.

In addition to the above applications, as related technologies mature, multi-view learning will have broad application prospects in the future and has potential application value in many fields.

- **Industrial manufacturing:** In the industrial manufacturing process, traditional anomaly detection methods usually only consider a single data source, while multi-view learning can combine different views such as sensor data, equipment operation data and historical fault data to effectively improve production efficiency and quickly detect anomalies.
- **Finance:** By combining market data, transaction data, customer behavior data, macroeconomic data, social media data, etc., multi-view learning can realize many applications in the financial field, such as risk management, fraud detection, investment decision-making and market analysis and forecasting.
- **Cyber security:** In the field of cyber security, multi-view learning can utilize multiple data sources such as network traffic data, system logs, user behavior data, and threat intelligence to improve the accuracy and efficiency of intrusion detection, network threat identification, and user behavior analysis, comprehensively enhancing cyber security protection capabilities.

As a whole, these applications demonstrate the versatility and effectiveness of multi-view learning in utilizing many information sources to enhance the performance. The ability to combine multiple views allows for a more comprehensive understanding of complex data and leads to better decision-making in diverse domains.

## 7 Challenges

Multi-view learning presents several challenges that necessitate attention from researchers and practitioners. These challenges encompass view inconsistency, view complementarity, optimal view fusion, the curse of dimensionality, scalability, limited labels, generalization across domains, and others.

**1) View inconsistency:** Within multi-view learning, each view typically describes different aspects or perspectives of the data. For instance, image data can be represented by multiple views constructed from extracting different features (such as Gabor, LBP, Intensity). Similarly, a webpage can be described using different types of information such as images, text, symbols. Although these different types of data can be

transformed into a unified vector representation through data preprocessing, some intrinsic information of the data may be lost in this process, leading to inconsistencies among the views. These inconsistencies may arise due to differences in data preprocessing procedures, choices of embedding methods, and variations in feature scales. Additionally, inconsistencies can also stem from data missing, noisy data or labels, and partially view-unaligned. Recently, there have been several methods [112,145] to address inconsistencies such as partially view-unaligned, data unmapped, and noise correspondence, however, relevant research remains scarce. Furthermore, the relevance of different views or views obtained through different feature extraction methods to the learning task may vary, further complicating the issue of view inconsistency. Therefore, effectively addressing these inconsistencies is a critical challenge in multi-view learning.

**2) View complementarity:** While each view offers unique information, there is also the potential for complementarity among views. The challenge lies in identifying and effectively utilizing the complementary information derived from different views to enhance overall learning performance. This entails in-depth analysis of the information provided by each view and their interrelations. Existing multi-view learning methods often encounter limitations in effectively leveraging view complementarity. One major limitation is the oversimplified assumption that views are independent and complementary. In reality, views may exhibit complex interdependencies, and their complementarity may not be fully exploited. Additionally, many methods focus solely on integrating information from different views without adequately considering their inherent complementarity. This may result in suboptimal performance, as the complementary information across views is not fully harnessed.

**3) Optimal view fusion:** The integration of multiple views into a single representation or model is a non-trivial task. As we discussed in Section 1, existing methods can be categorized into two types based on the timing of view fusion: early fusion and late fusion. Early fusion primarily integrates features or representations learned from different views, while late fusion shares similarities with ensemble learning. It remains uncertain which fusion method is more effective, prompting the need for comprehensive theoretical research. Additionally, assessing the significance of each view during the fusion process warrants further investigation. Existing methods primarily employ adaptive learning of view weights through the introduction of weighting parameters, yet frequently lack a certain level of rationality or interpretability. Determining how to combine the views, assessing their importance, or selecting relevant features from each view presents a challenge. The fusion process should appropriately capture shared information while reducing noise and redundancy.

**4) The curse of dimensionality:** The dimensionality of the data can significantly increase when multiple views are involved. Addressing the curse of dimensionality and avoiding overfitting necessitate careful feature selection, dimensionality reduction, or regularization techniques when dealing with high-dimensional data. Different views may contain similar or

correlated features, or may provide different feature subsets for the same learning task. How to maintain consistency in feature selection across different views, address the correlation and redundancy issues between features, and effectively perform feature selection, dimensionality reduction, and fusion on high-dimensional data to retain the most informative features, poses a challenge.

**5) Scalability:** In practical scenarios, the widespread application of large-scale data has become a reality, encompassing not only traditional data domains but also various fields such as the internet, social media, and biomedicine. However, existing multi-view learning methods primarily focus on small-scale datasets. As the number of views or data instances increases, the computational complexity of multi-view learning methods becomes a challenge. Therefore, ongoing research is focused on developing scalable algorithms and techniques capable of effectively handling large-scale datasets. Recently, Huang et al. [146] proposed a novel ensemble-based fast multi-view clustering method with nearly linear time and space complexity, demonstrating excellent performance on massive datasets. Further research is warranted to enable multi-view learning methods to effectively adapt to large-scale datasets.

**6) Limited labels:** Obtaining labeled data for every view in a multi-view setting is often challenging and costly. In supervised or semi-supervised learning scenarios, the shortage of sufficient labeled data presents challenges. Overcoming this limitation involves developing strategies to effectively utilize the limited labeled information or exploring weakly supervised learning methods. Tan et al. [147] attempted to learn a shared subspace from incomplete views and weak labels by exploiting the relationships between views and the relevance of local labels to approximate the original space. However, in the field of multi-view learning, weakly supervised learning is still at an early stage, making it worthy of further exploration and in-depth investigation.

**7) Generalization across domains:** In practical applications, multi-view learning frequently deals with heterogeneous data sources or different domains, such as images, text. The issue of domain shifts or discrepancies limits the ability to generalize the learned models or representations to diverse domains. Further, many methods overlook the interrelationships and interactions between different domains, resulting in models unable to capture the complex relationships in the real world. Meanwhile, the acquisition of label information is also an important factor limiting the model's ability to generalize across domains. Therefore, developing more robust and versatile multi-view learning methods is of great significance.

In summary, effectively addressing these challenges necessitates the development of novel algorithms, models, and evaluation methodologies. Such advancements should be capable of accommodating the complexity and diversity of multi-view data, thereby enhancing learning performance and scalability.

## 8 Conclusion

Multi-view learning is a rapidly expanding field that leverages

multiple views or data sources to enhance learning performance across diverse domains. By integrating information from different views, multi-view learning methods can enhance accuracy, robustness, and generalization capabilities. Research in this area can be categorized into four types: multi-view classification methods, multi-view semi-supervised classification methods, multi-view clustering methods, and multi-view semi-supervised clustering methods.

However, multi-view learning also presents several challenges. These challenges encompass handling view inconsistency, leveraging view complementarity, achieving optimal view fusion, addressing the curse of dimensionality, ensuring scalability, dealing with limited labeled data, and enabling generalization across different domains. Overcoming these challenges remains an active area of research, and researchers continue to explore innovative solutions.

Despite the inherent challenges, multi-view learning holds great potential for solving complex real-world problems. The versatility of this approach allows for the integration of diverse perspectives and the handling of high-dimensional data. Moving forward, advancements in view selection, alignment, consistency, and other aspects will further propel the field of multi-view learning, resulting in improved performance and increased applicability across various domains.

**Acknowledgements** This work was supported in part by the Major Key Project of PCL, China (PCL2023AS7-1 and PCL2023A09), in part by the National Key R&D Program of China (2023YFA1011601), in part by the National Natural Science Foundation of China (Grant Nos. 62106224 and U21A20478), and in part by the Guangzhou Science and Technology Plan Project (2024A04J3749).

**Competing interests** The authors declare that they have no competing interests or financial conflicts to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Wei W, Dai Q, Wong Y, Hu Y, Kankanhalli M, Geng W. Surface-electromyography-based gesture recognition by multi-view deep learning. *IEEE Transactions on Biomedical Engineering*, 2019, 66(10): 2964–2973
- Tian X, Deng Z, Ying W, Choi K S, Wu D, Qin B, Wang J, Shen H, Wang S. Deep multi-view feature learning for EEG-based epileptic seizure detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2019, 27(10): 1962–1972
- Kong Y, Ding Z, Li J, Fu Y. Deeply learned view-invariant features for cross-view action recognition. *IEEE Transactions on Image Processing*, 2017, 26(6): 3028–3037
- Sun S, Dong W, Liu Q. Multi-view representation learning with deep Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(12): 4453–4468
- Zhang C, Cheng J, Tian Q. Multi-view image classification with visual, semantic and view consistency. *IEEE Transactions on Image Processing*, 2020, 29: 617–627
- Xu C, Tao D, Xu C. A survey on multi-view learning. 2013, arXiv preprint arXiv: 1304.5634
- Luo S, Zhang C, Zhang W, Cao X. Consistent and specific multi-view subspace clustering. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018
- Wang X, Guo X, Lei Z, Zhang C, Li S Z. Exclusivity-consistency regularized multi-view subspace clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, 1–9
- Liang Y, Huang D, Wang C D. Consistency meets inconsistency: A unified graph learning framework for multi-view clustering. In: *Proceedings of the IEEE International Conference on Data Mining*. 2019, 1204–1209
- Li Y, Yang M, Zhang Z. A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(10): 1863–1883
- Li X, Liu B, Zhang K, Chen H, Cao W, Liu W, Tao D. Multi-view learning for hyperspectral image classification: an overview. *Neurocomputing*, 2022, 500: 499–517
- Zhao J, Xie X, Xu X, Sun S. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 2017, 38: 43–54
- Yan X, Hu S, Mao Y, Ye Y, Yu H. Deep multi-view learning methods: A review. *Neurocomputing*, 2021, 448: 106–129
- Yang Y, Wang H. Multi-view clustering: a survey. *Big Data Mining and Analytics*, 2018, 1(2): 83–107
- Fu L, Lin P, Vasilakos A V, Wang S. An overview of recent multi-view clustering. *Neurocomputing*, 2020, 402: 148–161
- Wen J, Zhang Z, Fei L, Zhang B, Xu Y, Zhang Z, Li J. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(2): 1136–1149
- Chao G, Sun S, Bi J. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2021, 2(2): 146–168
- Fang U, Li M, Li J, Gao L, Jia T, Zhang Y. A comprehensive survey on multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(12): 12350–12368
- Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. *Frontiers of Computer Science*, 2020, 14(2): 241–258
- Xu Y, Yu Z, Cao W, Chen C L P. A novel classifier ensemble method based on subspace enhancement for high-dimensional data classification. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 16–30
- Jiang J, Liu F, Ng W W Y, Tang Q, Wang W, Pham Q V. Dynamic incremental ensemble fuzzy classifier for data streams in green internet of things. *IEEE Transactions on Green Communications and Networking*, 2022, 6(3): 1316–1329
- Xu Y, Yu Z, Cao W, Chen C L P, You J. Adaptive classifier ensemble method based on spatial perception for high-dimensional data classification. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(7): 2847–2862
- Yu Z, Luo P, Liu J, Wong H S, You J, Han G, Zhang J. Semi-supervised ensemble clustering based on selected constraint projection. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(12): 2394–2407
- Jiang J, Liu F, Liu Y, Tang Q, Wang B, Zhong G, Wang W. A dynamic ensemble algorithm for anomaly detection in IoT imbalanced data streams. *Computer Communications*, 2022, 194: 250–257

25. Yang K, Yu Z, Wen X, Cao W, Chen C L P, Wong H S, You J. Hybrid classifier ensemble for imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(4): 1387–1400
26. Jiang B, Xiang J, Wu X, Wang Y, Chen H, Cao W, Sheng W. Robust multi-view learning via adaptive regression. *Information Sciences*, 2022, 610: 916–937
27. Zhao L, Yang T, Zhang J, Chen Z, Yang Y, Wang Z J. Co-learning non-negative correlated and uncorrelated features for multi-view data. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(4): 1486–1496
28. Chen W, Yang K, Yu Z, Shi Y, Chen C L P. A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, 2024, 57(6): 1–51
29. Li G, Yu Z, Yang K, Lin M, Chen C L P. Exploring feature selection with limited labels: a comprehensive survey of semi-supervised and unsupervised approaches. *IEEE Transactions on Knowledge and Data Engineering*, 2024, doi: 10.1109/TKDE.2024.3397878
30. Li W, Wang R, Luo X. A generalized nesterov-accelerated second-order latent factor model for high-dimensional and incomplete data. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, doi: 10.1109/TNNLS.2023.3321915
31. Luo D, Xu H, Carin L. Differentiable hierarchical optimal transport for robust multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 7293–7307
32. Xie X, Sun S. Multi-view support vector machines with the consensus and complementarity information. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 32(12): 2401–2413
33. Hu P, Peng D, Sang Y, Xiang Y. Multi-view linear discriminant analysis network. *IEEE Transactions on Image Processing*, 2019, 28(11): 5352–5365
34. Jia K, Lin J, Tan M, Tao D. Deep multi-view learning using neuron-wise correlation-maximizing regularizers. *IEEE Transactions on Image Processing*, 2019, 28(10): 5121–5134
35. Chao G, Sun S. Consensus and complementarity based maximum entropy discrimination for multi-view classification. *Information Sciences*, 2016, 367–368: 296–310
36. Guan Z, Zhang L, Peng J, Fan J. Multi-view concept learning for data representation. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(11): 3016–3028
37. Wang Q, Guo Y, Wang J, Luo X, Kong X. Multi-view analysis dictionary learning for image classification. *IEEE Access*, 2018, 6: 20174–20183
38. Liu B, Chen X, Xiao Y, Li W, Liu L, Liu C. An efficient dictionary-based multi-view learning method. *Information Sciences*, 2021, 576: 157–172
39. Jia X, Jing X Y, Sun Q, Chen S, Du B, Zhang D. Human collective intelligence inspired multi-view representation learning—Enabling view communication by simulating human communication mechanism. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 46(6): 7412–7429
40. Zheng Q, Zhu J, Li Z. Collaborative unsupervised multi-view representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4202–4210
41. Ma X, Xue S, Wu J, Yang J, Paris C, Nepal S, Sheng Q Z. Deep multi-attributed-view graph representation learning. *IEEE Transactions on Network Science and Engineering*, 2022, 9(5): 3762–3774
42. Huang Z, Zhou J T, Zhu H, Zhang C, Lv J, Peng X. Deep spectral representation learning from multi-view data. *IEEE Transactions on Image Processing*, 2021, 30: 5352–5362
43. Yang S, Li L, Wang S, Zhang W, Huang Q, Tian Q. SkeletonNet: A hybrid network with a skeleton-embedding process for multi-view image representation learning. *IEEE Transactions on Multimedia*, 2019, 21(11): 2916–2929
44. Zhang D, Yang G, Zhao S, Zhang Y, Ghista D, Zhang H, Li S. Direct quantification of coronary artery stenosis through hierarchical attentive multi-view learning. *IEEE Transactions on Medical Imaging*, 2020, 39(12): 4322–4334
45. Lyu Z, Yang M, Li H. Multi-view group representation learning for location-aware group recommendation. *Information Sciences*, 2021, 580: 495–509
46. Tan Y, Zhao G. Multi-view representation learning with Kolmogorov-Smirnov to predict default based on imbalanced and complex dataset. *Information Sciences*, 2022, 596: 380–394
47. Qin Y, Qin C, Zhang X, Qi D, Feng G. NIM-Nets: noise-aware incomplete multi-view learning networks. *IEEE Transactions on Image Processing*, 2023, 32: 175–189
48. Lin Y, Gou Y, Liu X, Bai J, Lv J, Peng X. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 4447–4461
49. Xu C, Tao D, Xu C. Multi-view learning with incomplete views. *IEEE Transactions on Image Processing*, 2015, 24(12): 5812–5825
50. Zhu P, Yao X, Wang Y, Cao M, Hui B, Zhao S, Hu Q. Latent heterogeneous graph network for incomplete multi-view learning. *IEEE Transactions on Multimedia*, 2023, 25: 3033–3045
51. Wen J, Liu C, Deng S, Liu Y, Fei L, Yan K, Xu Y. Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(8): 11396–11408
52. Liu C, Wen J, Luo X, Huang C, Wu Z, Xu Y. DICNet: deep instance-level contrastive network for double incomplete multi-view multi-label classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023, 8807–8815
53. Li X, Chen S. A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(10): 5918–5932
54. Deng J, Chen X, Jiang R, Song X, Tsang I W. A multi-view multi-task learning framework for multi-variate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8): 7665–7680
55. He J, Lawrence R. A graph-based framework for multi-task multi-view learning. In: *Proceedings of the 28th International Conference on Machine Learning*. 2011, 25–32
56. Zhang J, Huan J. Inductive multi-task learning with multiple view data. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2012, 543–551
57. Zhao D, Gao Q, Lu Y, Sun D. Non-aligned multi-view multi-label classification via learning view-specific labels. *IEEE Transactions on Multimedia*, 2023, 25: 7235–7247
58. Zhang Y, Wu J, Cai Z, Yu P S. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Transactions on Multimedia*, 2020, 22(11): 2844–2857
59. Yuan J, Liu W, Gu Z, Feng S. A unified framework for graph-based multi-view partial multi-label learning. *IEEE Access*, 2023, 11: 49205–49215
60. Liu B, Li W, Xiao Y, Chen X, Liu L, Liu C, Wang K, Sun P. Multi-view multi-label learning with high-order label correlation. *Information Sciences*, 2023, 624: 165–184
61. Li B, Yuan C, Xiong W, Hu W, Peng H, Ding X, Maybank S. Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2554–2560
62. Xu C, Tao D, Xu C. Multi-view intact space learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(12): 2531–2544

63. Hu J, Lu J, Tan Y P. Sharable and individual multi-view metric learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(9): 2281–2288
64. Wu S, Wu A, Zheng W S. Online multi-view learning with knowledge registration units. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, doi: 10.1109/TNNLS.2023.3256390
65. Fan R, Ouyang X, Luo T, Hu D, Hou C. Incomplete multi-view learning under label shift. *IEEE Transactions on Image Processing*, 2023, 32: 3702–3716
66. Fu Y, Hospedales T M, Xiang T, Gong S. Transductive multi-view zero-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(11): 2332–2345
67. Shi Z, Chen X, Zhao C, He H, Stuphorn V, Wu D. Multi-view broad learning system for primate oculomotor decision decoding. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020, 28(9): 1908–1920
68. Yan W, Li Y, Yang M. Towards deeper match for multi-view oriented multiple kernel learning. *Pattern Recognition*, 2023, 134: 109119
69. Huang S, Shi W, Xu Z, Tsang I W, Lv J. Efficient federated multi-view learning. *Pattern Recognition*, 2022, 131: 108817
70. Nie F, Cai G, Li J, Li X. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 2018, 27(3): 1501–1511
71. Nie F, Li J, Li X. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence*. 2016, 1881–1887
72. Nie F, Tian L, Wang R, Li X. Multiview semi-supervised learning model for image classification. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 32(12): 2389–2400
73. Xu X, Li W, Xu D, Tsang I W. Co-labeling for multi-view weakly labeled learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(6): 1113–1125
74. Wang X, Fu L, Zhang Y, Wang Y, Li Z. MMatch: Semi-supervised discriminative representation learning for multi-view classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(9): 6425–6436
75. Chao G, Sun S. Semi-supervised multi-view maximum entropy discrimination with expectation Laplacian regularization. *Information Fusion*, 2019, 45: 296–306
76. Zhang B, Qiang Q, Wang F, Nie F. Fast multi-view semi-supervised learning with learned graph. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(1): 286–299
77. Huang A, Wang Z, Zheng Y, Zhao T, Lin C W. Embedding regularizer learning for multi-view semi-supervised classification. *IEEE Transactions on Image Processing*, 2021, 30: 6997–7011
78. Wang S, Chen Z, Du S, Lin Z. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 5042–5055
79. Qian B, Wang X, Ye J, Davidson I. A reconstruction error based framework for multi-label and multi-view learning. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(3): 594–607
80. Zheng F, Liu Z, Chen Y, An J, Zhang Y. A novel adaptive multi-view non-negative graph semi-supervised ELM. *IEEE Access*, 2020, 8: 116350–116362
81. Guo W, Wang Z, Du W. Robust semi-supervised multi-view graph learning with sharable and individual structure. *Pattern Recognition*, 2023, 140: 109565
82. Li Z, Qiang Q, Zhang B, Wang F, Nie F. Flexible multi-view semi-supervised learning with unified graph. *Neural Networks*, 2021, 142: 92–104
83. Jia X, Jing X Y, Zhu X, Chen S, Du B, Cai Z, He Z, Yue D. Semi-supervised multi-view deep discriminant representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(7): 2496–2509
84. Cui X, Huang J, Chien J T. Multi-view and multi-objective semi-supervised learning for hmm-based automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(7): 1923–1935
85. Thammasorn P, Chaovaitwongse W A, Hippe D S, Wootton L S, Ford E C, Spraker M B, Combs S E, Peeken J C, Nyflot M J. Nearest neighbor-based strategy to optimize multi-view triplet network for classification of small-sample medical imaging data. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(2): 586–600
86. Xie Y, Lin B, Qu Y, Li C, Zhang W, Ma L, Wen Y, Tao D. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(11): 3594–3606
87. Liang Y, Huang D, Wang C D, Yu P S. Multi-view graph learning by joint modeling of consistency and inconsistency. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(2): 2848–2862
88. Huang L, Lu J, Tan Y P. Co-learned multi-view spectral clustering for face recognition based on image sets. *IEEE Signal Processing Letters*, 2014, 21(7): 875–879
89. Tang C, Zheng X, Liu X, Zhang W, Zhang J, Xiong J, Wang L. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(10): 4705–4716
90. Nie F, Shi S, Li J, Li X. Implicit weight learning for multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(8): 4223–4236
91. Zhao L, Zhao T, Sun T, Liu Z, Chen Z. Multi-view robust feature learning for data clustering. *IEEE Signal Processing Letters*, 2020, 27: 1750–1754
92. Liu B Y, Huang L, Wang C D, Lai J H, Yu P S. Multi-view consensus proximity learning for clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(7): 3405–3417
93. Hou C, Nie F, Tao H, Yi D. Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(9): 1998–2011
94. Hu S, Lou Z, Ye Y. View-wise versus cluster-wise weight: Which is better for multi-view clustering? *IEEE Transactions on Image Processing*, 2022, 31: 58–71
95. Deng Z, Liu R, Xu P, Choi K S, Zhang W, Tian X, Zhang T, Liang L, Qin B, Wang S. Multi-view clustering with the cooperation of visible and hidden views. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(2): 803–815
96. Yu X, Liu H, Lin Y, Liu N, Sun S. Sample-level weights learning for multi-view clustering on spectral rotation. *Information Sciences*, 2023, 619: 38–51
97. Liang C, Wang L, Liu L, Zhang H, Guo F. Multi-view unsupervised feature selection with tensor robust principal component analysis and consensus graph learning. *Pattern Recognition*, 2023, 141: 109632
98. Dai D, Yu Z, Huang W, Hu Y, Chen C L P. Multi-objective cluster ensemble based on filter refinement scheme. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8): 8257–8269
99. Yu Z, Kuang Z, Liu J, Chen H, Zhang J, You J, Wong H S, Han G. Adaptive ensembling of semi-supervised clustering solutions. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(8): 1577–1590
100. Shi Y, Yu Z, Chen C L P, Zeng H. Consensus clustering with co-association matrix optimization. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(3): 4192–4205
101. Yu Z, Wang D, Meng X B, Chen C L P. Clustering ensemble based on hybrid multiview clustering. *IEEE Transactions on Cybernetics*, 2022,

- 52(7): 6518–6530
102. Chen J, Yang S, Wang Z. Multi-view representation learning for data stream clustering. *Information Sciences*, 2022, 613: 731–746
  103. Zhao H, Li Z, Chen W, Zheng Z, Xie S. Accelerated partially shared dictionary learning with differentiable scale-invariant sparsity for multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(11): 8825–8839
  104. Zheng Q, Zhu J, Li Z, Tang H. Graph-guided unsupervised multiview representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(1): 146–159
  105. Zheng Q. Large-scale multi-view clustering via fast essential subspace representation learning. *IEEE Signal Processing Letters*, 2022, 29: 1893–1897
  106. Zhang C, Fu H, Liu S, Liu G, Cao X. Low-rank tensor constrained multiview subspace clustering. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, 1582–1590
  107. Cao X, Zhang C, Fu H, Liu S, Zhang H. Diversity-induced multi-view subspace clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, 586–594
  108. Zhang C, Hu Q, Fu H, Zhu P, Cao X. Latent multi-view subspace clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, 4333–4341
  109. Chao G, Sun J, Lu J, Wang A L, Langleben D D, Li C S, Bi J. Multi-view cluster analysis with incomplete data to understand treatment effects. *Information Sciences*, 2019, 494: 278–293
  110. Chao G, Wang S, Yang S, Li C, Chu D. Incomplete multi-view clustering with multiple imputation and ensemble clustering. *Applied Intelligence*, 2022, 52(13): 14811–14821
  111. Fang X, Hu Y, Zhou P, Wu D O. V<sup>3</sup>H: View variation and view heredity for incomplete multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2020, 1(3): 233–247
  112. Yang M, Li Y, Hu P, Bai J, Lv J, Peng X. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 1055–1069
  113. Liu S, Liu X, Wang S, Niu X, Zhu E. Fast incomplete multi-view clustering with view-independent anchors. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(6): 7740–7751
  114. Zhang L, Zhao Y, Zhu Z F, Shen D G, Ji S W. Multi-view missing data completion. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(7): 1296–1309
  115. Yin M, Liu X, Wang L, He G. Learning latent embedding via weighted projection matrix alignment for incomplete multi-view clustering. *Information Sciences*, 2023, 634: 244–258
  116. Shang M, Liang C, Luo J, Zhang H. Incomplete multi-view clustering by simultaneously learning robust representations and optimal graph structures. *Information Sciences*, 2023, 640: 119038
  117. Chao G, Jiang Y, Chu D. Incomplete contrastive multi-view clustering with high-confidence guiding. In: *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. 2024, 11221–11229
  118. Wang R, Wang P, Wu D, Sun Z, Nie F, Li X. Multi-view and multi-order structured graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, doi: 10.1109/TNNLS.2023.3256390
  119. Wang R, Nie F, Wang Z, Hu H, Li X. Parameter-free weighted multi-view projected clustering with structured graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 32(10): 2014–2025
  120. Xia W, Gao Q, Wang Q, Gao X, Ding C, Tao D. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 5187–5202
  121. Jiang G, Peng J, Wang H, Mi Z, Fu X. Tensorial multi-view clustering via low-rank constrained high-order graph learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(8): 5307–5318
  122. Huang S, Tsang I W, Xu Z, Lv J. Measuring diversity in graph learning: A unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(12): 5869–5883
  123. Zhang X, Zhang X, Liu H, Liu X. Multi-task multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(12): 3324–3338
  124. Zhang X, Zhang X, Liu H. Multi-task multi-view clustering for non-negative data. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. 2015, 4055–4061
  125. Jiang Z, Liu X. Adaptive KNN and graph-based auto-weighted multi-view consensus spectral learning. *Information Sciences*, 2022, 609: 1132–1146
  126. Mei Y, Ren Z, Wu B, Yang T, Shao Y. Multi-order similarity learning for multi-view spectral clustering. *Pattern Recognition*, 2023, 137: 109264
  127. Qin Y, Wu H, Zhang X, Feng G. Semi-supervised structured subspace learning for multi-view clustering. *IEEE Transactions on Image Processing*, 2022, 31: 1–14
  128. Zhu Z, Gao Q. Semi-supervised clustering via cannot link relationship for multiview data. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(12): 8744–8755
  129. Zhang C, Fu H, Wang J, Li W, Cao X, Hu Q. Tensorized multi-view subspace representation learning. *International Journal of Computer Vision*, 2020, 128(8): 2344–2361
  130. Tang Y, Xie Y, Zhang C, Zhang W. Constrained tensor representation learning for multi-view semi-supervised subspace clustering. *IEEE Transactions on Multimedia*, 2022, 24: 3920–3933
  131. Zhang T, Zheng W, Cui Z, Zong Y, Yan J, Yan K. A deep neural network-driven feature learning method for multi-view facial expression recognition. *IEEE Transactions on Multimedia*, 2016, 18(12): 2528–2536
  132. Wei X, Yu R, Sun J. Learning view-based graph convolutional network for multi-view 3D shape analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 7525–7541
  133. Dong C, Chen X, Hu R, Cao J, Li X. MVSS-Net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(3): 3539–3553
  134. Tran H N, Nguyen H Q, Doan H G, Tran T H, Le T L, Vu H. Pairwise-covariance multi-view discriminant analysis for robust cross-view human action recognition. *IEEE Access*, 2021, 9: 76097–76111
  135. Wang Y, Xiao Y, Lu J, Tan B, Cao Z, Zhang Z, Zhou J T. Discriminative multi-view dynamic image fusion for cross-view 3-D action recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(10): 5332–5345
  136. Chen J, Wang Y, Tang Y Y. Person re-identification by exploiting spatio-temporal cues and multi-view metric learning. *IEEE Signal Processing Letters*, 2016, 23(7): 998–1002
  137. Yuan Y, Xun G, Jia K, Zhang A. A multi-view deep learning framework for EEG seizure detection. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(1): 83–94
  138. Yang S, Lian C, Zeng Z, Xu B, Zang J, Zhang Z. A multi-view multi-scale neural network for multi-label ECG classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023, 7(3): 648–660
  139. Puyol-Antón E, Ruijsink B, Gerber B, Amzulescu M S, Langet H, De Craene M, Schnabel J A, Piro P, King A P. Regional multi-view learning for cardiac motion analysis: Application to identification of dilated cardiomyopathy patients. *IEEE Transactions on Biomedical*

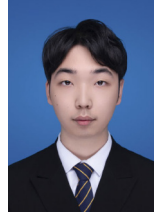
- Engineering, 2019, 66(4): 956–966
140. Zhang J, Huan J. Predicting drug-induced QT prolongation effects using multi-view learning. *IEEE Transactions on NanoBioscience*, 2013, 12(3): 206–213
  141. Jin Y, Li C, Li Y, Peng P, Giannopoulos G A. Model latent views with multi-center metric learning for vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(3): 1919–1931
  142. Zhu Y, Zhang S, Chen S. Vehicle recognition based on carrier-free UWB radars using contrastive multi-view learning. *IEEE Microwave and Wireless Technology Letters*, 2023, 33(3): 343–346
  143. Ge H, Gao D, Sun L, Hou Y, Yu C, Wang Y, Tan G. Multi-agent transfer reinforcement learning with multi-view encoder for adaptive traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(8): 12572–12587
  144. Yang H, Chu X, Zhang L, Sun Y, Li D, Maybank S J. QuadNet: Quadruplet loss for multi-view learning in baggage re-identification. *Pattern Recognition*, 2022, 126: 108546
  145. Zhang X, Zong L, Liu X, Yu H. Constrained NMF-based multi-view clustering on unmapped data. In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 2015
  146. Huang D, Wang C D, Lai J H. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(11): 11388–11402
  147. Tan Q, Yu G, Domeniconi C, Wang J, Zhang Z. Incomplete multi-view weak-label learning. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 2018, 2703–2709



Zhiwen Yu is a professor in the School of Computer Science and Engineering, South China University of Technology, China. He received the PhD degree from City University of Hong Kong, China in 2008. Dr. Yu has authored or coauthored more than 200 refereed journal articles and international conference papers, including more than 70 articles in the journals of *IEEE Transactions*. He is an Associate Editor of the *IEEE Transactions on systems, man, and cybernetics: systems*. He is a senior member of IEEE and ACM, a Member of the Council of China Computer Federation (CCF).



Ziyang Dong received the BS degree from the School of Computer Science and Engineering, South China University of Technology, China in 2011, and received the MASc degree from the School of Computer Science and Engineering, South China University of Technology, China in 2016. He is currently pursuing the PhD degree in the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include machine learning and data mining.



Chenchen Yu received the BS degree in 2023 from the Hohai University, China and he is currently working toward the MS degree with the South China University of Technology, China. His research interests include data mining and machine learning.



Kaixiang Yang received the BS degree and MS degree from the University of Electronic Science and Technology of China and Harbin Institute of Technology, China in 2012 and 2015, respectively, and the PhD degree from the School of Computer Science and Engineering, South China University of Technology, China in 2020. He has been a Research Engineer with the 7th Research Institute, China Electronics Technology Group Corporation, China from 2015 to 2017, and has been a postdoctoral researcher with Zhejiang University, China from 2021 to 2023. He is currently an associate professor with the School of Computer Science and Engineering, South China University of Technology, China. His research interests include pattern recognition, machine learning, and industrial data intelligence.



Ziwei Fan received the BS degree of Natural Sciences from the University of Cambridge, United Kingdom in 2011. She is currently pursuing the PhD degree in the School of Computer Science and Engineering, South China University of Technology, China. Her research interests include machine learning and AI based imaging diagnosis.



C. L. Philip Chen is the Chair Professor and Dean of the College of Computer Science and Engineering, South China University of Technology, China. He is a Fellow of IEEE, AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea (AE), and a member of European Academy of Sciences and Arts (EASA). He received IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learnings, and he is a highly cited researcher by Clarivate Analytics from 2018–2022. His current research interests include cybernetics, systems, and computational intelligence. He was the Editor-in-Chief of the *IEEE Transactions on Cybernetics*, the Editor-in-Chief of the *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, President of IEEE Systems, Man, and Cybernetics Society. Dr. Chen was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University, USA in 1988, after he graduated from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA in 1985.