Expert Systems With Applications 296 (2026) 129062

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

DCMTL network: A double-contrast multi-task learning network for semi-supervised multi-source data classification

Yongjie Huang, Man Chen, Jun Chen, Zhisong Pan*

Command Control Engineering College, Army Engineering University of PLA, Nanjing, 210007, China

ARTICLE INFO

Keywords:
Multi-task learning
Semi-supervised learning
Multi-source data
Image classification
Time series classification

ABSTRACT

Semi-supervised learning (SSL) has demonstrated remarkable success across various fields but is primarily limited to single data sources, which constrains its performance in complex multi-source scenarios with limited labeled data. To meet this challenge, we propose a novel double-contrast multi-task learning (DCMTL) network for semi-supervised multi-source data classification. Specifically, DCMTL considers each data source as an individual task and comprises two main modules: multi-source prototype contrastive learning (MPCL) and contrastive multi-task learning (CML). The MPCL module extends prototype contrastive learning (PCL) for multi-source settings, providing efficient feature representation. It leverages abundant multi-source unlabeled data for instance discrimination and semantic clustering, while also incorporates multi-source index information to assist model training. To effectively utilize limited labeled data, the CML defines an encoder-decoder multi-task learning (MTL) network to jointly learn knowledge across tasks. Meanwhile, it contrasts features across categories to enhance the feature compactness and obtain better classification boundaries. Moreover, the CML shares its encoder with the MPCL to improve the model's performance by synergistic utilization of both labeled and unlabeled data. Extensive experiments on both image datasets and time series datasets show that our method significantly outperforms state-of-the-art methods.

1. Introduction

In scenarios with limited labeled data, SSL has achieved great success by leveraging large amounts of unlabeled data to assist model training (Yang et al., 2022). However, with the rapid development of various sensing devices, the Internet, the Internet of Things, and other technologies, the characteristics of data generated in current society are increasingly complex (Misra et al., 2020). Specifically, the data originates from various independent sources, leading to the formation of multi-source datasets that exhibit differing feature distributions yet maintain underlying correlations. This difference can be attributed to variance in locations, time periods, and operational conditions that arise during data collection. Additionally, disparities in sensor technologies and data processing methodologies further exacerbate these differences (Maharana et al., 2022; Qiu et al., 2022). In this context, each data source typically also possesses a limited amount of labeled data and corresponds to an independent task. This limited labeled multi-source data scenario poses a significant challenge for SSL.

SSL is a learning paradigm that employs limited labeled data and abundant unlabeled data to construct models (Yang et al., 2022). A common strategy within this framework is to extend self-supervised learning by integrating labeled instances into the training process (Chen et al., 2020b; Huang et al., 2023). Self-supervised learning aims to harness large amounts of unlabeled data to develop a feature extractor with robust generalization capabilities (Ericsson et al., 2022). Contrastive learning, a discriminative method, seeks to achieve the objective of self-supervised learning by setting instance discriminative pretext tasks (Jaiswal et al., 2020). For instance, SimCLR (Chen et al., 2020a) and MoCo (He et al., 2020) enhance the feature extraction capability of the model by calculating the feature contrast between pairs of image instances that have been separately augmented from the same image. However, the aforementioned methods treat augmented data views from disparate instances as negative pairs, neglecting their semantic similarities. This oversight results in the unnecessary distancing of numerous negative sample pairs with similar semantics in the feature space. Fortunately, recent research has found that considering both instance and semantic similarity alleviate these problems (Gao et al., 2021; Pu et al., 2023). The PCL (Li et al., 2021), a variant of contrastive learning, addresses these limitations by enabling the model to learn instance discriminability and capture semantic features associated with hierarchical clustering. This framework links contrastive learning and clustering, providing new ideas for the study of self-supervised learning.

* Corresponding author.
E-mail addresses: yongjiehforever@163.com (Y. Huang), chenmanhjdx@163.com (M. Chen), chenjun_sc@aeu.edu.cn (J. Chen), panzs@nuaa.edu.cn (Z. Pan).

https://doi.org/10.1016/j.eswa.2025.129062
Received 2 December 2024; Received in revised form 26 June 2025; Accepted 16 July 2025
Available online 19 July 2025

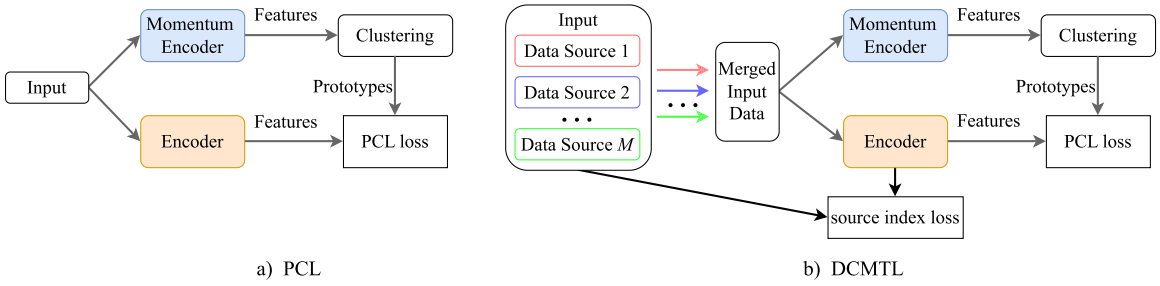

Fig. 1. Comparison between the typical prototype contrastive learning (PCL) and the double-contrast multi-task learning (DCMTL). In the DCMTL setting, there are multiple data sources with different features distributions and utilizing their source indexes as additional supervisory information when compared with PCL.

Nevertheless, as shown in Fig. 1, this method is primarily focused on a single data source and does not take into account the additional auxiliary information provided by multiple data sources. This gap highlights the need to integrate methods from multi-source data scenarios in order to fully utilize the potential of SSL in complex scenarios. Our goal is to address these limitations by developing a method that combines semi-supervised learning with multi-source data learning to improve model performance in semi-supervised multi-source data scenarios.

Furthermore, the limited labeled data is employed to train the model for achieving SSL. For multi-source semi-supervised data scenarios, MTL represents a promising solution, where each individual data source can be regarded as a distinct task. This framework facilitates the sharing of common knowledge across multiple data sources, thereby enhancing the model's generalization performance for each data source (Caruana, 1997). However, most existing MTL methods are designed for a single data source with multiple labels and thus cannot be directly applied to multi-source data scenarios. For example, Mti-net (Vandenhende et al., 2020) considers a single image for depth estimation and image segmentation tasks under multi-scale interaction. Moreover, there is an additional critical issue worth noting. In this scenario, MTL has to learn both limited labeled data and utilize abundant unlabeled data. Typical MTL networks usually employ a single encoder that learns features from these multi-source data to enable knowledge sharing across data sources (Lee & Son, 2022; Liu et al., 2019). However, a single encoder learns features from data in multiple source domains, which may lead to difficulties in distinguishing features with fuzzy classification boundaries when identifying unseen data, due to inconsistent distributions among the data sources (Zhou et al., 2021).

In this work, we adopt double contrastive learning to enhance the classification ability of the MTL model for semi-supervised multi-source data. A novel MTL network is designed called the DCMTL network: double-contrast multi-task learning network. Unlike existing methods, the motivation of DCMTL network comes from three aspects: 1) PCL is a contrastive learning method that integrates both instance discrimination and semantic hierarchical clustering (Li et al., 2021). Therefore, we introduce source indexes of multi-source data as supervised information into the PCL, and extend it as a SSL method for multi-source data classification. 2) Contrastive learning possesses the capability to effectively cluster instances of the same category within the feature space while simultaneously separating instances from different categories (Wang et al., 2023). Based on this, we contrast features across categories within each data source to enhance the compactness of the feature representations, thus alleviating the problem of fuzzy classification boundaries. 3) In MTL, the useful task-specific knowledge contained within common knowledge is key to improving model performance (Huang et al., 2024). Inspired by this, we consider whether an encoder that identifies the source of all samples can further improve the performance of the MTL models. Consequently, we share an encoder between the extended PCL and the MTL network.

Specifically, as shown in Fig. 2, the DCMTL network consists of two parts, the MPCL module and the CML module. Initially, in the MPCL module, features extracted by the momentum encoder are hierarchically clustered to obtain the prototypes, following the training steps of PCL. Then, the distances between these prototypes and features extracted by the encoder are contrasted to compute the PCL loss. On this basis, the network identifies the sources of all samples to compute the source index loss. In the CML module, a MTL network with an encoder-decoder structure is designed to learn features from labeled samples across multiple data sources. Subsequently, a contrastive learning technique is applied to enhance the compactness of the decoded features, and computes the feature compactness contrast loss. Finally, the decoded features are also used to complete the learning of each data source. In addition, the two modules share a single encoder. We conduct extensive experiments with the the DCMTL network on two image datasets and generalized it to two time series datasets. Experimental results show the significant superiority of our method over existing SSL methods.

We summarize the our contributions as follows:

- A dual-contrast multi-task learning network is proposed that combines multi-task learning and contrastive learning techniques to encourage the model to learn effective feature representations for semi-supervised classification in multi-source data scenarios.
- A multi-source contrastive learning approach is designed to integrate both instance discrimination and semantic clustering, while also incorporating source index information from diverse data sources.
- A contrastive learning method is introduced to encourage feature compactness for better classification boundaries.
- Experimental results show that our proposed DCMTL network achieves state-of-the-art results on several multi-source image classification datasets for semi-supervised learning, and generalizes to multi-source time series classification datasets.

2. Related works

2.1. Semi-supervised learning

In recent years, SSL has achieved remarkable success across various fields (Chen et al., 2024a; Deng et al., 2024). A common approach is to incorporate labeled samples into self-supervised learning strategies to facilitate the training of semi-supervised learning (Georgescu et al., 2021; Goyal et al., 2019). Contrastive learning is a contrastive self-supervised learning method that focuses on learning discriminative features in the feature space. For example, SimCLR (Chen et al., 2020a) considers different data augmentations of the same image as positive samples and different images as negative samples, while measuring their distance as a contrast loss to guide the model to learn by itself. MOCO (He et al., 2020) employs a dynamic queue to store negative samples, enabling the model to leverage rich negative sample information even when processing smaller batches of data. The authors (Huynh et al., 2022) eliminate the most similar false negative samples to avoid their adverse effects on contrast training. MDCL (He et al., 2023) not only

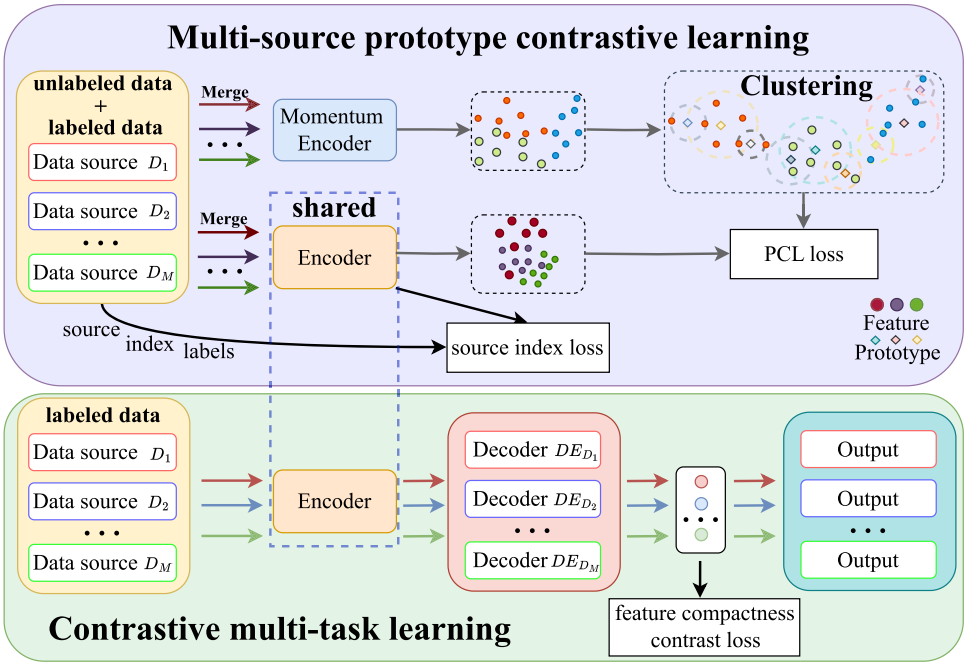

Fig. 2. Illustration of double-contrast multi-task learning (DCMTL).

focuses on the contrast between instances, but also introduces multi-domain learning approaches for intra- and inter-domain alignment. It also considers domain index information to assist in model training. Another approach to achieve semi-supervised learning is to generate pseudo-labels for unlabeled data and train the model by combining labeled and pseudo-labeled data (Chen et al., 2024b). FixMatch (Sohn et al., 2020) performs consistent regularisation on unlabeled data using both weak and strong data augmentations to generate high-quality pseudo-labels, and uses these pseudo-labels alongside ground truth labels for training. FlexMatch (Zhang et al., 2021) assigns trusted pseudo-labels to unlabeled data and dynamically adjusts thresholds of different categories to optimise the selection process for unlabeled data. MultiMatch (Qi et al., 2024) extends FixMatch to a multi-task learning framework by defining independent batch normalization for each data source. This approach reduces interference between data sources, leading to the generation of high-quality pseudo-labels. Recently, several SSL models in the adversarial setting have been developed. For instance, SeiMtl (Wang et al., 2022) jointly learns data from multiple data sources. Meanwhile, an adversarial learning scheme is designed to train the discriminator using multi-source index information as supervised signals for unlabeled data, and forcing the model to predict close to the true distribution. MTSS-AAE (Ullah et al., 2023) uses an adversarial learning technique to identify whether unlabeled data generates samples.

Most of these methods are designed for single data sources, while MDCL, SeiMtl and MultiMatch can be applied in multi-source data scenarios. MDCL considers a relatively singular data augmentation strategy and does not fully take into account semantic-level contrasts. SeiMtl is unstable and difficult to achieve converge during the training process. MultiMatch is sensitive to the threshold settings of pseudo-labels. On the other hand, pseudo-labels are generated by model predictions and therefore may contain false labels. Our proposed network relies on self-supervised learning and incorporates MTL to jointly learn knowledge from multi-source data. In addition, multi-source indexes are used as supervised information to further assist model training.

2.2. Contrastive learning

Contrastive learning has been demonstrated to be highly effective for representation learning in a self-supervised way (Chen et al., 2020a; He et al., 2020; Jaiswal et al., 2020). Previous research has typically concentrated on instance-level discrimination. However, recent works on semantic level contrasts provides new perspectives for contrastive learning research (Quan et al., 2023; Xiao et al., 2022). PCL(Li et al., 2021), which contrasts instance-level features and also contrasts instance features to a set of prototypes, has been shown to provide efficient feature representation. FEND (Wang et al., 2023) and DCCL (Pu et al., 2023) are related to prototype contrastive learning. FEND uses an offline cluster module to obtain prototypes and also contrasts instance features with these prototypes to improve the trajectory prediction accuracy. DCCL designs a dynamic prototype generation and update mechanism that ensures consistency in prototype learning, thus further facilitating model optimization.

Contrastive learning has also been extended to supervised setting to exploit label similarities (Khosla et al., 2020). The authors (Motiian et al., 2017) use contrast loss to encourage instances from the same class to be close to each other in the feature space. The authors (Kang et al., 2019) propose a contrastive adaptation network and a new metric method to leverage the class similarities. Our proposed network extends PCL to multi-source data scenarios, while introducing multi-source index information to assist model training. Meanwhile, the compactness of the features is enhanced using a supervised contrastive learning technique, resulting in obtaining better classification boundaries.

3. Methodology

3.1. Problem formulation

SSL aims to utilize both labeled and unlabeled data to overcome the limitations imposed by the use of only limited labeled information and to improve the performance of the model. The problem of SSL in multi-source data scenarios is described as follows. Suppose that we have M different data sources $\{D_1, D_2, \dots, D_M\}$. Meanwhile, each independent task is associated with a dataset D_i consisting of two subdatasets $D_{i,L}$ and $D_{i,U}$. The labeled subdataset $D_{i,L} = \{\mathbf{x}^h_{D_{i,L}}, y^h_{D_i}\}_{h=1}^{N_{D_{i,L}}}$ contains $N_{D_{i,L}}$ labeled samples, where $\mathbf{x}^h_{D_{i,L}}$ and $y^h_{D_i}$ are one sample and its corresponding label, respectively. There is a consistent label space $\mathbf{Y} = \{1, 2, \dots, c\}$, $\forall y^h_{D_i} \in \mathbf{Y}$, where c denotes the number of categories.

Similarly, the unlabeled subdataset $D_{i,U} = \{\mathbf{x}^h_{D_{i,U}}\}_{h=N_{D_{i,L}}+1}^{N_{D_{i,L}}+N_{D_{i,U}}}$ contains $N_{D_{i,U}}$ unlabeled samples, where $\mathbf{x}^h_{D_{i,U}}$ is a unlabeled sample. On this basis, we assume that one sample and its corresponding source index label are denoted as $\mathbf{x}_{D_i} \in \mathbf{x}_{D_{i,L}} \cup \mathbf{x}_{D_{i,U}}$ and $y^h_{source} \in \{1, 2, \ldots, M\}$, respectively. It is worth noting that the samples in each data source are independent, i.e. $P(\mathbf{x}_{D_i}, \mathbf{x}_{D_j}) = P(\mathbf{x}_{D_i}) \cdot P(\mathbf{x}_{D_j}), \forall D_i \neq D_j$. Our goal is to train a MTL model using both labeled and unlabeled samples from diverse data sources.

3.2. Overview

Our work addresses the problem of SSL in multi-source data scenarios where the input data is acquired from several different data sources and exists a large amount of unlabeled data. We design a MPCL module to learn feature representations from multiple data sources. The MPCL follows the training steps of the original PCL and utilizes a large amount of unlabeled data from different data sources to obtain an encoder with strong feature extraction capability in a self-supervised manner. Furthermore, the source indexes of multi-source data serve as supervised information to facilitate the encoder learning. After self-supervised learning, a CML module learns limited labeled information from multiple data sources. The CML includes a MTL network with an encoder-decoder structure. This MTL network shares its encoder with the MPCL to utilize the large amount of unlabeled data. Meanwhile, the task-specific decoder combines contrastive learning at the category level to promote the features compactness.

3.3. Multi-source prototype contrastive learning

The MPCL extends PCL to multi-source data scenarios, and shares a single encoder with the CML. PCL's high-dimensional feature comparison and clustering are computationally inefficient. At the same time, training multiple tasks together is obviously more time-consuming than training a single task. Therefore, to make our method can be more widely applied, we follow FEND (Wang et al., 2023) and use an offline MPCL module for self-supervised learning.

3.3.1. Prototypical contrastive learning

The PCL is the foundation of the MPCL module. For data source D_i, the training steps of the original PCL are as follows. Firstly, an encoder E with parameters θ and a momentum encoder E' with parameters θ' are defined to obtain the features $\hat{\mathbf{v}}^h_{D_i} = E(aug(\mathbf{x}^h_{D_i}))$ and the positive features $\tilde{\mathbf{v}}^h_{D_i} = E'(aug(\mathbf{x}^h_{D_i}))$ for sample $\mathbf{x}^h_{D_i}$, where $aug(\cdot)$ denotes data augmentation strategy. On this basis, we perform the computation of a fully connected layer to generate the fixed-dimensional feature embedding $\mathbf{v}^h_{D_i} = FC_{emb}(\hat{\mathbf{v}}^h_{D_i})$ and the positive feature embedding $\bar{\mathbf{v}}^h_{D_i} = FC_{emb}(\tilde{\mathbf{v}}^h_{D_i})$. Consistent with MOCO (He et al., 2020), the momentum encoder E' has the same parameter initialization as encoder E, and parameters θ' update do not participate in back-propagation, which is as follows:

$$\theta' = m\theta' + (1-m)\theta \tag{1}$$

here, $m = 0.999$ (MOCO default). Therefore, the instance-wise contrastive loss is defined as follows:

$$\mathcal{L}^{D_i}_{ins} = -\sum_{h=1}^{N_{D_{i,L}}+N_{D_{i,U}}} \log \frac{\exp\left(\mathbf{v}^h_{D_i} \cdot \bar{\mathbf{v}}^h_{D_i}/\tau_{ins}\right)}{\sum_{j=1}^{r} \exp\left(\mathbf{v}^h_{D_i} \cdot \mathbf{v}^j_{D_i}/\tau_{ins}\right)} \tag{2}$$

where, τ_{ins} denotes the contrastive temperature of the instance-wise contrastive loss, $\mathbf{v}^j_{D_i}$ denotes one of r negative embeddings. Then, we perform K times Kmeans on the momentum feature embeddings $\{\bar{\mathbf{v}}^h_{D_i}\}_{h=1}^{N_{D_{i,L}}+N_{D_{i,U}}}$ to obtain $\{S_k\}_{k=1}^{K}$ clusters. Finally, the prototype-wise contrastive loss $\mathcal{L}^{D_i}_{proto}$ is defined as follows:

$$\mathcal{L}^{D_i}_{proto} = -\frac{1}{K}\sum_{h=1}^{N_{D_{i,L}}+N_{D_{i,U}}}\sum_{k=1}^{K} \log \frac{\exp\left(\mathbf{v}^h_{D_i} \cdot \mathbf{c}_m/\phi_m\right)}{\sum_{s=1}^{S_k} \exp\left(\mathbf{v}^h_{D_i} \cdot \mathbf{c}_s/\phi_s\right)} \tag{3}$$

where, $\mathbf{c}_m$ denotes the prototype of the sample $\mathbf{x}^h_{D_i}$, and $\mathbf{c}_s$ denotes the prototype of an arbitrary cluster. The prototype $\mathbf{c}$ is calculated by taking an average of all the features in a cluster. ϕ denotes the density of a cluster, which is calculated as below:

$$\phi = \frac{\sum_{z=1}^{Z} \|\bar{\mathbf{v}}_z - \mathbf{c}\|_2}{Z\log(Z+\alpha)} \tag{4}$$

where Z denotes the number of instances in the cluster, α denotes a smoothing factor, and $\bar{\mathbf{v}}_z$ denote one momentum feature embedding (obtained using momentum encoder E') within the cluster corresponding to prototype $\mathbf{c}$. Therefore, the original PCL loss for a single data source D_i is defined as:

$$\mathcal{L}^{D_i}_{pcl} = \mathcal{L}^{D_i}_{ins} + \mathcal{L}^{D_i}_{proto} \tag{5}$$

3.3.2. Multi-source data supervised information

The majority of SSL research focuses on a single data source, while supervised information from multi-source indexes is often ignored in multi-source data scenarios (Huynh et al., 2022; Zhang et al., 2021). This supervised information from multi-source data is crucial for training a encoder that recognizes the data sources.

The original PCL is designed to be oriented towards a single data source. Extended to multi-source data scenarios, the dataset $D = \bigcup_{i=1}^{M} D_i$ consists of merged data from all data sources, and the neural network is trained following the steps described above. The PCL loss for multi-source data is defined as:

$$\mathcal{L}_{pcl} = \mathcal{L}^{D}_{pcl} \tag{6}$$

It should be noted that the $\mathcal{L}_{pcl}$ does not consider supervised information derived from multiple data sources. Instead, an encoder that recognizes the source of the task is important for MTL models (Huang et al., 2024). In order to obtain an efficient feature extraction encoder shared with the MTL model, we add a task discriminator after the encoder E to predict the source indexes:

$$\hat{\mathbf{y}}^h_{source} = SD(\hat{\mathbf{v}}^h_D) \tag{7}$$

where, $\hat{\mathbf{y}}^h_{source}$ denotes the output of the prediction source index and $SD(\cdot)$ denotes a source discriminator. In our method, the source discriminator is simply set up as a fully connected layer. In addition, the MPCL follows the standard optimization procedure for neural networks by minimizing the cross-entropy loss:

$$\mathcal{L}_{source} = -\sum_{h\in D} y^h_{source} \log\left(\hat{\mathbf{y}}^h_{source}\right) \tag{8}$$

where $\mathcal{L}_{source}$ denotes the loss of the predicted source index. Finally, the total loss of the MPCL is:

$$\mathcal{L}_{mpcl} = \mathcal{L}_{source} + \mathcal{L}_{pcl} \tag{9}$$

3.4. Contrastive multi-task learning

3.4.1. Multi-task learning

In our approach, a typical encoder-decoder structure of the MTL is used to learn knowledge from the limited labeled samples in multi-source data. This MTL network shares an encoder E with the MPCL and achieves common knowledge sharing among tasks. Therefore, the knowledge of data source D_i learned from encoder E can be formulated as:

$$\mathbf{z}^h_{D_i} = E\left(\mathbf{x}^h_{D_{i,L}}\right) \tag{10}$$

where, $\mathbf{z}_{D_i}^h$ denotes the features captured by the shared encoder. Subsequently, each data source is assigned a separate task-specific decoder responsible for extracting features associated with that data source, calculated as follows:

$$\hat{\mathbf{z}}_{D_i}^h = DE_{D_i}\left(\mathbf{z}_{D_i}^h\right) \tag{11}$$

where, $\hat{\mathbf{z}}_{D_i}^h$ denotes the task-specific features of data source D_i and $DE_{D_i}(\cdot)$ denotes a task-specific decoder.

3.4.2. Contrastive learning for feature compactness

Multiple data sources sharing a single network may lead to fuzzy classification boundaries when applied to unseen data (Zhou et al., 2021). In this case, the performance boosted by the MTL network is greatly reduced when the MTL network employ the shared encoder.

In order to solve this issue, we also encourage feature compactness via contrastive learning to enhance feature alignment. Specifically, data belonging to the same category should be positioned close to each other in the feature space, while data belonging to different categories should be positioned at a distance. In our MTL setup, the model is capable of acquiring labels, learning contrasting features in a supervised training mode. The feature compactness contrast loss is defined as follows:

$$\mathcal{L}_{con}^{D_i} = -\sum_{h=1}^{N_{D_{i,L}}} \frac{1}{N_p} \sum_{p=1}^{N_p} \log \frac{\exp\left(\hat{\mathbf{z}}_{D_i}^h \cdot \mathbf{z}_{D_i}^p / \tau_{con}\right)}{\sum_{b=1}^{B} \exp\left(\hat{\mathbf{z}}_{D_i}^h \cdot \mathbf{z}_{D_i}^b / \tau_{con}\right)} \tag{12}$$

where, τ_{con} denotes the contrastive temperature of the contrast loss, N_p denotes the number of positive samples to in a batch, B denotes the batch size, $\mathbf{z}_{D_i}^p$ denotes the positive sample, and $\mathbf{z}_{D_i}^b$ denotes the negative sample. The positive samples are instances from the same category as the sample $\mathbf{x}_{D_{i,L}}^h$, while other instances in the batch are considered negative samples.

3.4.3. Objective function

The classification loss $\mathcal{L}_{ce}^{D_i}$ for data source D_i is denoted as:

$$\mathcal{L}_{ce}^{D_i} = -\sum_{h=1}^{N_{D_{i,L}}} y_{D_i}^h \left(\log \hat{\mathbf{y}}_{D_i}^h\right) \tag{13}$$

where, $\hat{\mathbf{y}}_{D_i}^h = FC(\hat{\mathbf{z}}_{D_i}^h)$ denotes the predicted output of sample $\mathbf{x}_{D_{i,L}}^h$, $FC(\cdot)$ denotes a fully connected layer. Consequently, the objective function for CML has two main components:

$$\mathcal{L} = \sum_{i=1}^{M} \mathcal{L}_{\text{ce}}^{D_i} + \lambda \mathcal{L}_{con}^{D_i} \tag{14}$$

where, λ denotes the hyperparameter of balance the contrast loss and classification loss. The training procedure of DCMTL network are summarized in Algorithm 1. The MPCL module is trained separately from the CML module. The MPCL module mixes labeled and unlabeled samples from all data sources as input, but does not use label information from labeled samples. In this module, not only instance-level contrasts are considered, but also prototypes are introduced for semantic-level contrasts. In addition, the model training process utilizes multi-source index information to provide auxiliary support. The CML module shares an encoder with the MPCL module so that the model combines labeled and unlabeled samples for training. In this module a multi-task learning network with an encoder-decoder structure is jointly trained on multi-source data and features of different categories are contrasted to increase the feature compactness.

The DCMTL network needs to apply two contrasts during the training process, where the MPCL module involves instance-level and semantic-level contrasts and the CML module involves category-level contrasts. These two contrasts undoubtedly require significant computational cost. However, during the test phase, only the CML module is involved and no contrasts are required. The computational cost is consistent with that of a common multi-task learning network with an encoder-decoder structure.

Algorithm 1 The training procedure of DCMTL network.

Require: Training datasets $D_i = \left\{D_{i,L}, D_{i,U}\right\}_{i=1}^{M}$
1: Random initialize the parameters of the network
2: **repeat**
3: {Use all samples to train MPCL parameters}
4: **for** all $\mathbf{x}_D^h \in D, D = \bigcup_{i=1}^{M} D_i$ **do**
5: Obtain feature embeddings by encoder E and momentum encoder E'.
6: Compute the instance-wise contrastive loss $\mathcal{L}_{ins}^{D}$ by 2
7: Compute the prototype-wise contrastive loss $\mathcal{L}_{proto}^{D}$ by 3
8: Predict multi source indexes and calculate the source index loss $\mathcal{L}_{source}$ by 8
9: Update the MPCL parameters on the network by 9
10: **end for**
11: {Use labeled samples to train CML parameters}
12: **for** all $\left(\mathbf{x}_{D_{i,L}}^h, y_{D_i}^h\right) \in D_{i,L}, i = 1, 2, \ldots, M$ **do**
13: Extract features using shared encoder E by 10
14: Extract task-specific features by 11
15: Compute the contrast loss $\mathcal{L}_{con}^{D_i}$ by 12
16: Compute the classification loss $\mathcal{L}_{ce}^{D_i}$ by 13
17: Update the CML parameters on the network by 14
18: **end for**
19: **until** converged
Output: The final neural network

4. Experiments

4.1. Datasets

We evaluated our network on commonly used datasets(see Table 1). Specifically, we tested office-home[1] and office-31[2] for image classification and generalized to the time series datasets Autism Brain Imaging Data Exchange (ABIDE)[3] and EMG physical actions[4] for time series classification. In the experimental setup, we divided all the samples in these datasets into training and test dataset in the ratio of 8:2.

Office-Home dataset consists of image subsets from four different domains: Art, Clipart, Product and Real-World. Each subset has 65 categories and a total of 15,500 images. In our experiments, these domains of data are defined as four different data sources. Meanwhile, two sets of experiments, Office-Home 20 % and Office-Home 10 %, are set up based on 20 % and 10 % labeled samples in the training dataset.

The Office-31 dataset is the mainstream dataset in machine vision, which contains 4110 images of 31 categories of objects commonly found in office environments. These images are mainly from Amazon (online e-commerce images), Webcam (low resolution images taken by webcams), and DSLR (high resolution images taken by SLR cameras). Experiments are set up based on the 20 % labeled data in the training dataset.

EMG physical action dataset includes 10 normal physical actions that measure the human activity. The data was collected by four subjects (age 25 to 30) using the EMG wireless apparatus. The subjects are placed 8 skin-surface electrodes and collected 1000 time series length data. These subsets are further preprocessed into samples with 50 timesteps and 8 feature variables. These four subjects are defined as four different data sources. In addition, experiments are set up based on 20 % and 10 % labeled data in the training dataset.

[1] http://hemanthdv.org/OfficeHome-Dataset/
[2] https://github.com/jindongwang/transferlearning/tree/master/data
[3] http://preprocessed-connectomes-project.org/abide/
[4] http://archive.ics.uci.edu/ml/datasets/EMG + Physical + Action + Data + Set

Table 1
Characteristics of the experimental datasets.

Dataset	Samples	Number of data sources	Data type	Categories
office-home	15500	4	Image	65
office-31	4110	3	Image	31
Normal EMG	7916	4	time series	10
ABIDE	871	2	time series	2

ABIDE dataset collect a total of 871 quality functional magnetic resonance imaging data from 17 different imaging sites. Since the limited number of subjects in several sites, we select subjects from 2 different sites with the number of subjects >100, including NYU (172 subjects) and UM (120 subjects). These subjects are preprocessed into samples with 80 timesteps and 116 feature variables, and divided into 2 categories. In our experiment, the data from these two different sites are defined as samples from two different sources. At the same time, experiments are also set up based on the 10 % labeled data in the training dataset.

4.2. Implementation

The environment for all experiments is as follows: PyTorch 1.8 framework, Python version 3.8, and a GeForce RTX 2080 Ti GPU for training.

For the image dataset, we scale the size of the image to 224 * 224 and follow He et al. (2020), Wu et al. (2018) to perform data augmentation with random crop, random color jittering, random horizontal flip, and random grayscale conversion. A ResNet-50 is adopted as the encoder. For time series dataset, the hidden dimension of LSTM is set to 128 and follow Chen et al. (2023) to perform data augmentation with masking and mixing in the frequency domain. The encoder consists of 5 layers of LSTM. Both their decoders consist of two fully connected layers with output dimension [4096, 4096].

For Kmeans clustering, following Wang et al. (2023), we use $K = \{20, 50, 100\}$ as the cluster numbers for getting hierarchical clusters, and set r to be consistent with the batch sizes. In addition, following Li et al. (2021), the fully connected layer processes the features extracted by the encoder and outputs 128-D and normalized fixed-dimensional feature embeddings. SGD is used as experimental optimizer, with a weight decay of 0.0001, a momentum of 0.9. The initial learning rates for the image and time series classification experiments are 0.01 and 0.001, respectively, and both decayed to 0 in a cosine shape. Meanwhile, following Li et al. (2021), the hyperparameters τ_{ins} and τ_{con} are both set to 0.1, and α is set to 10. For image data and time series data, the λ is 0.3 and 0.8, respectively. The remaining hyperparameters are set as follows: the datasets have the batch sizes of 32 and the epochs of 600. Except for the ABIDE dataset, which is set to 4 for the batch sizes and 300 for the epochs due to its small sample size.

4.3. Results of model performance

In order to demonstrate the effectiveness of our method, we have selected state-of-the-art semi-supervised classification methods for comparison, including contrastive learning semi-supervised classification methods: SimCLR (Chen et al., 2020a), PCL (Li et al., 2021), and MDCL (He et al., 2023); adversarial learning semi-supervised classification methods: SeiMtl (Wang et al., 2022) and MTSS-AAE (Ullah et al., 2023); and pseudo-label semi-supervised classification methods: FixMatch (Sohn et al., 2020) and MultiMatch (Qi et al., 2024). It is worth noting that the experimental setup for SimCLR and PCL is consistent with the original paper, with a fully connected layer added to the network as a classifier to learn limited labeled information, thereby completing semi-supervised learning. To ensure the fairness and comparability of the experiments, these compared methods adopt the same settings and hyperparameters as those used in the DCMTL.

Table 2 reports the test accuracy on Office-Home 20 % dataset. From the experimental results, it can be seen that: 1) Among these semi-supervised methods, DCMTL has the best performances. 2) Compared to current semi-supervised methods, the DCMTL achieves optimal results across the majority of data sources. 3) Compared to FixMatch, the performance of DCMTL in the Product data source is slightly inferior, although the gap is not significant. The experimental results demonstrate the effectiveness of our method. Combined with Table 3 experimental results on Office-Home 10 % dataset, the DCMTL still shows remarkable generalization performance when applied to a considerably smaller amount of labeled multi-source data.

Table 4 reports the test accuracy on Office-31 dataset. As can be seen from the results, our algorithm can outperform the methods for contrastive methods (SimCLR and PCL) and adversarial methods (SeiMtl and MTSS-AAE). However, the performance is slightly lower compared to the current state-of-the-art pseudo-label method (MultiMatch). This may be due to the smaller number of total samples, samples in each data source, and samples in each category in Office-31, compared to Office-home. The MultiMatch combines high-quality pseudo-labeled and MTL technology, leading to significant experimental results. Therefore, the MTL technique is also demonstrated to be an effective method for processing small samples of data from multiple sources. However, it is worth noting that the DCMTL achieves the optimal result in Webcam data source. This again shows the effectiveness of DCMTL for semi-supervised multi-source image classification.

Table 2
Experiment results on Office-Home 20 % dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	Art	Clipart	Product	Real-World	Average
SimCLR	22.61 ± 0.81	42.51 ± 0.12	66.27 ± 0.41	41.89 ± 0.20	43.32 ± 0.38
PCL	20.35 ± 0.45	41.18 ± 0.12	64.73 ± 0.71	39.89 ± 0.38	41.54 ± 0.41
MDCL	22.99 ± 0.52	30.02 ± 0.69	47.28 ± 0.61	35.73 ± 0.48	34.01 ± 0.58
SeiMtl	27.52 ± 0.79	43.95 ± 0.72	66.97 ± 0.38	44.05 ± 1.44	45.62 ± 0.83
MTSS-AAE	22.09 ± 0.72	39.20 ± 0.25	60.11 ± 1.58	37.17 ± 1.33	39.64 ± 0.97
FixMatch	14.54 ± 0.56	38.15 ± 0.79	**67.64 ± 0.35**	38.57 ± 1.32	39.72 ± 0.76
MultiMatch	19.90 ± 1.19	40.05 ± 0.78	67.26 ± 0.38	40.05 ± 0.96	41.81 ± 0.83
DCMTL	**27.97 ± 0.34**	**44.27 ± 0.50**	67.33 ± 0.47	**45.41 ± 0.13**	**46.24 ± 0.36**

Table 3
Experiment results on Office-Home 10 % dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	Art	Clipart	Product	Real-World	Average
SimCLR	16.20 ± 0.65	30.75 ± 0.24	52.54 ± 0.27	29.17 ± 0.31	32.16 ± 0.37
PCL	14.24 ± 0.45	28.66 ± 0.18	50.46 ± 0.58	28.61 ± 0.73	30.49 ± 0.48
MDCL	15.68 ± 0.25	17.18 ± 0.07	33.01 ± 1.89	24.60 ± 0.62	22.62 ± 0.71
SeiMtl	17.64 ± 0.59	28.33 ± 0.42	47.64 ± 0.73	29.73 ± 1.15	30.83 ± 0.72
MTSS-AAE	15.53 ± 0.51	26.97 ± 0.38	42.97 ± 0.99	24.12 ± 0.43	27.39 ± 0.58
FixMatch	10.62 ± 0.22	25.11 ± 1.16	**55.56 ± 0.36**	23.04 ± 0.83	28.58 ± 0.64
MultiMatch	13.34 ± 0.22	26.56 ± 0.79	50.03 ± 0.44	25.00 ± 0.98	28.73 ± 0.61
DCMTL	**19.68 ± 0.45**	**32.08 ± 0.91**	53.01 ± 0.78	**33.81 ± 0.61**	**34.64 ± 0.69**

Table 4
Experiment results on Office-31 dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	Amazon	Webcam	DSLR	Average
SimCLR	62.92 ± 0.46	55.55 ± 1.23	70.94 ± 0.40	63.14 ± 0.70
PCL	62.79 ± 0.55	55.14 ± 0.71	65.94 ± 0.40	61.29 ± 0.55
MDCL	48.30 ± 0.27	59.66 ± 1.88	58.33 ± 0.41	55.43 ± 0.85
SeiMtl	60.28 ± 2.36	55.96 ± 1.42	64.04 ± 1.79	60.09 ± 1.86
MTSS-AAE	58.00 ± 0.73	60.49 ± 2.47	63.56 ± 1.24	60.68 ± 1.48
FixMatch	66.96 ± 0.59	51.02 ± 0.71	67.61 ± 0.82	61.86 ± 0.71
MultiMatch	**67.15 ± 0.53**	62.13 ± 0.71	**72.37 ± 0.40**	**67.22 ± 0.55**
DCMTL	65.50 ± 0.74	**64.60 ± 1.42**	68.80 ± 0.40	66.30 ± 0.80

Table 5
Experiment results on EMG 20 % dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	Subject 1	Subject 2	Subject 3	Subject 4	Average
SimCLR	32.99 ± 0.00	60.50 ± 0.43	39.64 ± 0.42	29.45 ± 0.28	40.64 ± 0.28
PCL	35.42 ± 0.14	63.16 ± 0.28	40.06 ± 0.28	32.23 ± 0.28	42.72 ± 0.25
SeiMtl	47.60 ± 0.00	65.25 ± 0.00	48.12 ± 0.00	44.19 ± 0.00	51.29 ± 0.00
MTSS-AAE	25.18 ± 0.00	55.50 ± 0.00	32.66 ± 0.00	25.50 ± 0.00	34.71 ± 0.00
FixMatch	46.00 ± 0.14	**69.00 ± 0.00**	49.12 ± 0.00	37.87 ± 0.00	50.49 ± 0.03
MultiMatch	46.59 ± 0.00	67.00 ± 0.00	44.88 ± 0.00	43.43 ± 0.00	50.47 ± 0.00
DCMTL	**84.63 ± 0.00**	66.75 ± 0.00	**71.82 ± 0.00**	**70.45 ± 0.00**	**73.41 ± 0.00**

Table 6
Experiment results on EMG 10 % dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	Subject 1	Subject 2	Subject 3	Subject 4	Average
SimCLR	24.18 ± 0.00	47.00 ± 0.00	25.18 ± 0.00	22.72 ± 0.00	29.77 ± 0.00
PCL	27.45 ± 0.00	51.75 ± 0.00	28.92 ± 0.00	25.25 ± 0.00	33.34 ± 0.00
SeiMtl	29.97 ± 0.00	**58.50 ± 0.00**	32.66 ± 0.00	30.55 ± 0.00	37.92 ± 0.00
MTSS-AAE	18.13 ± 0.00	46.00 ± 0.00	23.94 ± 0.00	19.44 ± 0.00	26.87 ± 0.00
FixMatch	30.81 ± 0.29	58.00 ± 0.00	32.16 ± 0.00	26.01 ± 0.00	36.74 ± 0.07
MultiMatch	26.27 ± 0.28	53.83 ± 0.28	31.42 ± 0.43	27.27 ± 0.43	34.69 ± 0.36
DCMTL	**67.00 ± 0.00**	57.25 ± 0.00	**50.87 ± 0.00**	**52.27 ± 0.00**	**56.84 ± 0.00**

To validate the generalization performance of our method, the experimental setup is generalized from image datasets to time series datasets. Table 5 reports the test accuracy of DCMTL network on EMG 20 % dataset. The experimental results show that: (1) Among these semi-supervised methods, the DCMTL also performs best in multi-source time series classification tasks. (2) DCMTL significantly outperforms current semi-supervised methods on the majority of data sources. When comparing the results of Table 6 on the EMG 10 % dataset, DCMTL demonstrates remarkable competitiveness even with fewer labeled data in multi-source time series dataset. Table 7 shows that the DCMTL exhibits similar phenomenon on ABIDE dataset, a medical dataset with smaller samples. The experimental results further validate the effectiveness of our method, while the performance in image data and time series data demonstrates its good generalization ability.

These experimental results described above convincingly validate the effectiveness and generalization of our proposed DCMTL network. This network offers significant advantages over current semi-supervised methods. It incorporates contrastive learning and multi-task learning to address the challenges of SSL in multi-source data scenarios. Based on instance discrimination and semantic clustering, multi-source index information is added to assist with model training. At the same time, features across categories are contrasted to enhance feature compactness for better classification boundaries when jointly learning knowledge across data sources. This is the primary reason why our method outperforms other semi-supervised learning methods in multi-source data scenarios.

Table 7
Experiment results on ABIDE dataset. We report the average accuracy and standard error over three times. The optimal performances are bolded.

Method	NYU	UM	Average
SimCLR	**71.56 ± 1.69**	69.33 ± 2.30	70.44 ± 2.00
PCL	61.76 ± 0.00	68.00 ± 0.00	64.88 ± 0.00
SeiMtl	70.58 ± 0.00	**72.00 ± 0.00**	**71.29 ± 0.00**
MTSS-AAE	58.82 ± 0.00	**72.00 ± 0.00**	65.41 ± 0.00
FixMatch	68.62 ± 1.69	70.66 ± 2.30	69.64 ± 2.00
MultiMatch	70.58 ± 0.00	**72.00 ± 0.00**	**71.29 ± 0.00**
DCMTL	70.58 ± 0.00	**72.00 ± 0.00**	**71.29 ± 0.00**

4.4. Ablation studies

In this section, we perform ablation experiments for all the experiments to verify the validity of key components in the proposed method. Experimental results are listed in Table 8, where "wo/ MPCL", "wo/ source" and "wo/ cont" indicate without using the MPCL module, the source index loss $\mathcal{L}_{source}$ and the contrast loss $\mathcal{L}_{con}$ in our method, respectively. As observed in this table, the removal of the MPCL module results in a significant decrease in average accuracy, confirming its critical role in enhancing model performance through feature representations acquired from contrastive learning and multi-source index information. Furthermore, the combined utilization of both the source index loss and the contrast loss achieves superior performance compared to using either loss in isolation. This demonstrates the effectiveness of introducing multi-source index information and enhancing feature compactness for improving model performance.

When comparing the results of Office-Home 20 % and Office-Home 10 %, we find that the effect of the source index loss on model performance is more significant as the number of labeled samples decreases. We infer that the reduction in supervised information makes the multi-source index information play a more important role in improving model performance as an additional supervised signal. The EMG dataset defines each subject as a data source, and the differences between subjects may be relatively small compared to other datasets, so training the model to correctly recognize each subject provides a more effective feature representation and has a greater impact on performance. Furthermore, we observe that our proposed method outperforms the average accuracy of most semi-supervised learning methods even without using the source index loss. This proves the effectiveness of PCL within the MPCL module.

4.5. Parameter analysis experiments

To further explore the effect of the balance parameter λ on DCMTL network, we conducted experiments at different ranges of value. Specifically, we conducted extensive experiments on the image dataset Office-Home 20 % and the time series dataset EMG 10 % to explore the effect of taking values of λ in the range of 0.1 to 1 at intervals of 0.1 on the DCMTL. The results are presented in Fig. 3. From this figure, we can observe the following results: (1) The parameter λ affects the classification performance of DCMTL, resulting in different values of λ producing widely varying classification results. (2) Parameter λ has a greater impact on time series data than on image data. (3) When the λ ranges between 0.2 and 0.4, the image classification accuracy is significantly superior compared to other classification experiments with λ values. (4) When the λ ranges between 0.7 and 0.9, the time series classification accuracy is significantly superior compared to other classification experiments with λ values. (5) Compared to image datasets, time series datasets require a higher value of parameter λ. We infer that time series

Table 8
Results of ablation experiments on all datasets. We report the average accuracy and standard error for all tasks over three times. The optimal performances are bolded.

Module	Office-Home 20%	Office-Home 10%	Office-31	EMG 20%	EMG 10%	ABIDE
wo/ MPCL	42.47 ± 0.96	28.43 ± 0.96	58.01 ± 0.91	49.88 ± 0.14	36.97 ± 0.10	66.35 ± 0.00
wo/ source	45.46 ± 0.35	30.27 ± 0.49	62.64 ± 1.31	50.53 ± 0.00	37.32 ± 0.07	68.35 ± 0.00
wo/ cont	44.95 ± 0.31	33.60 ± 0.43	61.79 ± 0.74	68.76 ± 0.00	52.45 ± 0.00	68.48 ± 1.15
DCMTL	**46.24 ± 0.36**	**34.64 ± 0.69**	**66.30 ± 0.80**	**73.41 ± 0.00**	**54.90 ± 0.00**	**71.29 ± 0.00**

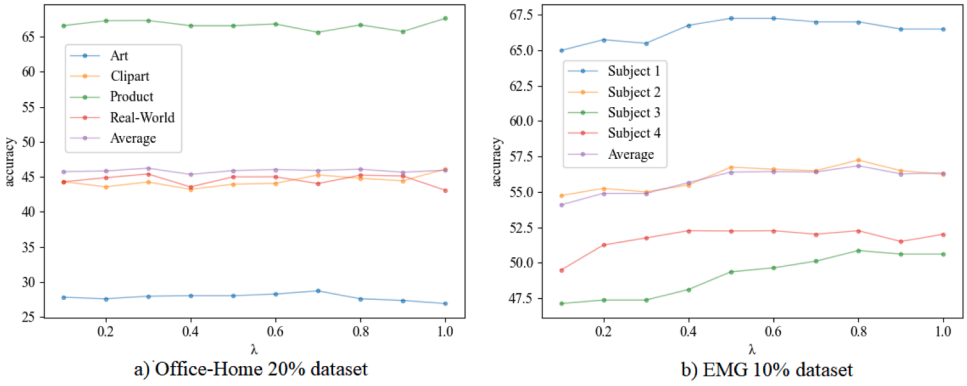

Fig. 3. Experimental results of the parameter λ on Office-Home 20 % and EMG 10 % datasets.

Table 9
Experimental results of the parameter K on Office-Home 20 %. The optimal performances are bolded.

K	Art	Clipart	Product	Real-World	Average
{2, 5, 10}	26.69	44.32	66.35	45.49	45.71
{4, 10, 20}	28.05	**44.92**	65.88	45.61	46.11
{20, 50, 100}	28.28	44.68	**67.76**	45.49	**46.55**
{10, 250, 500}	**28.50**	44.32	67.64	45.73	46.54
{200, 500, 1000}	28.28	43.84	66.94	**46.33**	46.34

Table 10
Experimental results of the parameter K on EMG 10 % dataset. The optimal performances are bolded.

K	Subject 1	Subject 2	Subject 3	Subject 4	Average
{2, 5, 10}	61.94	55.00	47.63	51.76	54.08
{4, 10, 20}	67.50	56.75	48.37	53.03	56.41
{20, 50, 100}	67.00	57.25	**50.87**	52.27	56.84
{10, 250, 500}	**67.75**	56.50	48.62	51.01	55.97
{200, 500, 1000}	67.00	**59.25**	48.62	**53.53**	**57.10**

data is more susceptible to noise and exhibits greater complexity in the dependencies between data points. This complexity can hinder effective feature extraction, and a higher λ is required to adequately contrast samples during model training.

To evaluate the impact of cluster quantity on the DCMTL network performance, we set up four sets of comparison experiments based on $K = \{20, 50, 100\}$ to determine the optimal cluster configuration. Tables 9 and 10 show the results of the experiments on the Office-Home 20 % dataset and EMG 10 % dataset, respectively. We observe that $K = \{20, 50, 100\}$ achieves favorable classification performance, while increasing the cluster quantity does not significantly improve performance. Therefore, $K = \{20, 50, 100\}$ is the appropriate parameter setting.

We test the effect of contrastive temperatures on the DCMTL network performance. As shown in Fig. 4, we find that the contrastive temperature setting of 0.1 is valid for all data sources. Higher temperatures do not have a significant effect. These hyperparameter settings are consistent with the SimCLR optimal hyperparameter settings. Meanwhile, the optimal contrastive temperature is different for each data source, which is consistent with previous research (Mu et al., 2024).

4.6. Multi-task learning effectiveness experiments

We introduce a multi-task learning approach that enables knowledge sharing across multiple data sources through parameter sharing in a shared encoder, thereby enhancing the model's generalization performance for each individual data source. To validate the effectiveness of multi-task learning, we compared the performance of single-task learning on individual data sources with that of multi-task learning across multiple data sources using the CML module. Fig. 5 a and b show the experimental results on Office-Home 20 % and EMG 10 % datasets (using only labeled data), respectively. We observe that multi-task joint learning with a shared encoder significantly outperforms single-task individual learning. This validates the effectiveness of multi-task learning in our method.

4.7. Feature visualization

We visualize the features extracted by our proposed method and the PCL method for two categories in the Office-home 20 % dataset and the EMG 10 % dataset by t-SNE (Van der Maaten & Hinton, 2008).

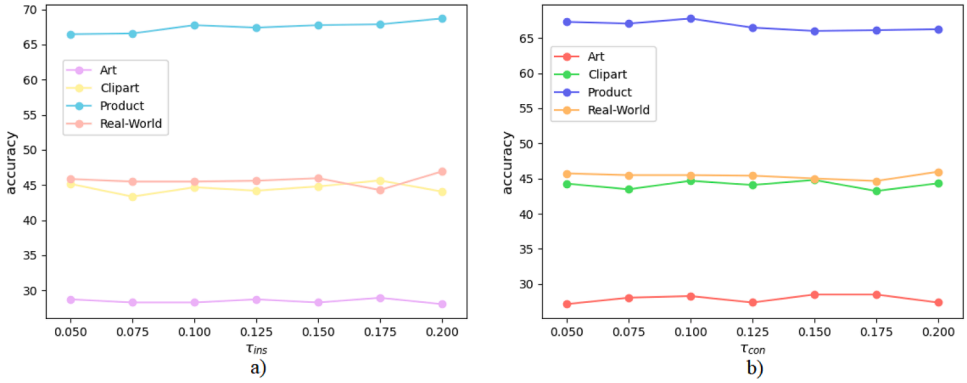

Fig. 4. Experimental results of the contrastive temperature on Office-Home 20 % datasets. a) and b) denote the experimental results for parameters τ_{ins} and τ_{con}, respectively.

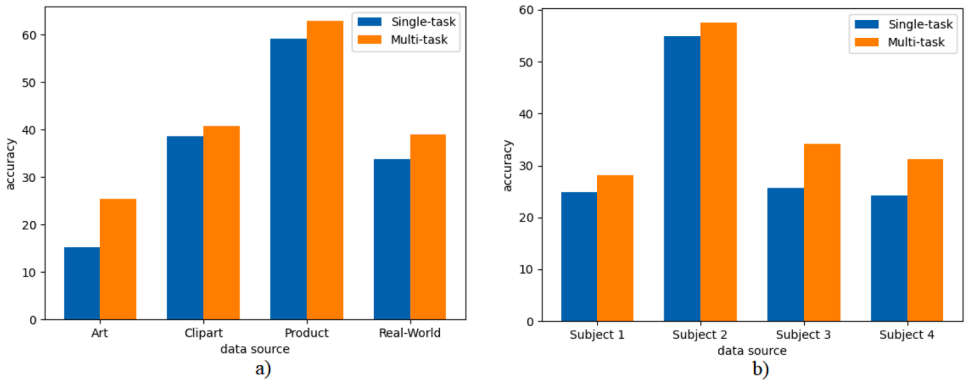

Fig. 5. Experimental results of multi-task learning effectiveness on Office-Home 20 % and EMG 10 % datasets.

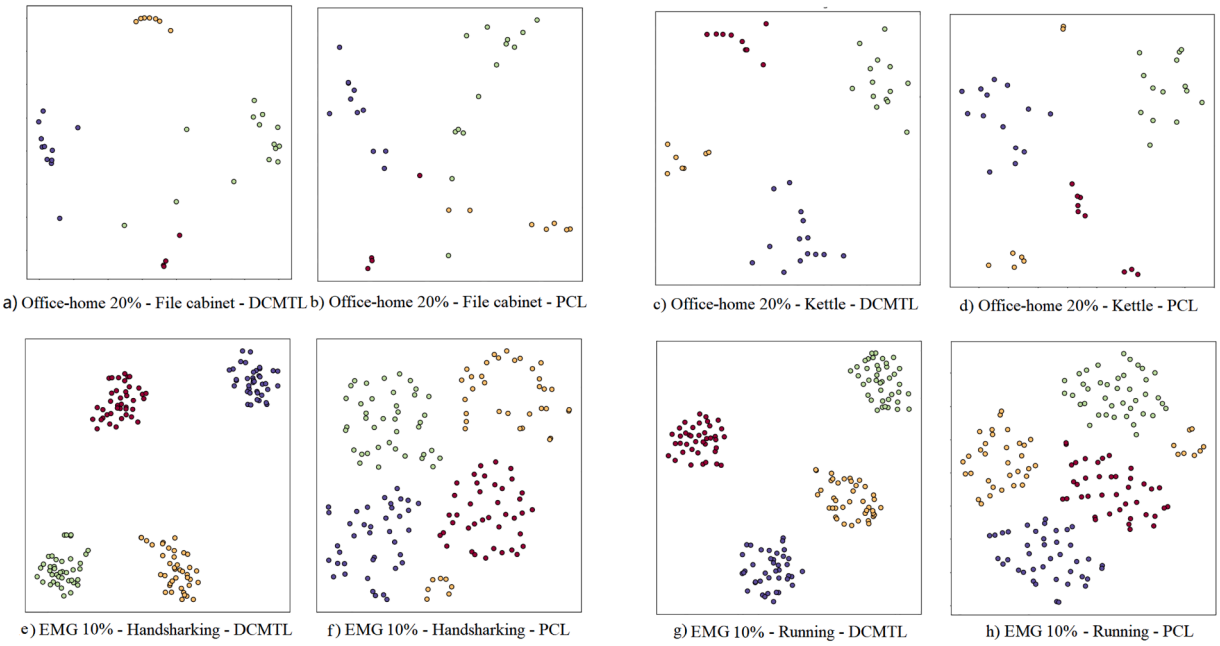

Fig. 6. Results of visualization experiments on Office-Home 20 % and EMG 10 % datasets.

Different colours indicate different data sources. The results are reported in Fig. 6. We observe that the same category appears in different locations for different data sources, which confirms that the classification boundary differs for each data source. Therefore, a network that extracts features from all data sources and is applied to unseen data will suffer from fuzzy classification boundaries. Furthermore, the features of our method are more discriminative. The method is extended based on PCL, which verifies that incorporating multi-source index information, enhancing feature compactness and joint multi-task learning are effective strategies to improve model performance.

5. Conclusion

In this paper, we provide a novel semi-supervised MTL network DCMTL. Unlike previous SSL research, this network is applied to multi-source data scenarios. The DCMTL not only utilizes multi-source unlabeled data for instance and semantic contrasts, but also introduces multi-source index information to enhance the model's feature representation capability. Meanwhile, it jointly learns knowledge from multi-source labeled data and performs category-level contrasts in feature space to enhance feature compactness. To evaluate the proposed DCMTL network, we compare it with other semi-supervised learning methods on image datasets and also generalize to time series datasets. The experimental results demonstrate the effectiveness of our DCMTL. In summary, our work has been richly researched in the following three aspects: (1) Based on instance discrimination and semantic clustering, multi-source index information is introduced to obtain efficient feature representations. (2) To overcome the indiscriminative features learned by the model from multi-source data, supervised contrastive learning is employed to enhance feature compactness for better classification boundaries. (3) A novel semi-supervised MTL method is designed to accommodate multi-source data scenarios. The proposed method also exists limitations. In the multi-source data scenarios considered in this paper, each data source has different feature distributions that belong to homogeneous data. However, in the real world, most multi-source data are heterogeneous. In the future, we will delve into the problems faced in semi-supervised multi-source data scenarios: 1) the presence of heterogeneous data between data sources; 2) the existence of high imbalance between data sources; and 3) the combination of multi-domain learning methods (Schoenauer-Sebag et al., 2019) to mitigate the data drift that exists between data sources.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Yongjie Huang: Conceptualization, Methodology, Software, Writing – original draft; **Man Chen:** Formal analysis, Visualization; **Jun Chen:** Data curation, Investigation; **Zhisong Pan:** Funding acquisition, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 62076251).

References

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*, 41–75.

Chen, M., He, Y., Wang, T., Hu, Y., Chen, J., & Pan, Z. (2024a). Scale-mixing enhancement and dual consistency guidance for end-to-end semi-supervised ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 15685–15701.

Chen, M., Wang, T., Xu, C., Chen, J., Chen, E., & Pan, Z. (2024b). Gradient prior guidance and image adaptation enhancement for semi-supervised SAR ship instance segmentation. *IEEE Sensors Journal*, *24*(21), 36216–36229.

Chen, M., Xu, Z., Zeng, A., & Xu, Q. (2023). FrAug: Frequency domain augmentation for time series forecasting. arXiv preprint arXiv:2302.09292.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020a). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597–1607). PMLR.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. E. (2020b). Big self-supervised models are strong semi-supervised learners. *Advances in Neural Information Processing Systems*, *33*, 22243–22255.

Deng, Q., Guo, Y., Yang, Z., Pan, H., & Chen, J. (2024). Boosting semi-supervised learning with contrastive complementary labeling. *Neural Networks*, *170*, 417–426.

Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, *39*(3), 42–62.

Gao, Y., Fei, N., Liu, G., Lu, Z., & Xiang, T. (2021). Contrastive prototype learning with augmented embeddings for few-shot learning. In *Uncertainty in artificial intelligence* (pp. 140–150). PMLR.

Georgescu, M.-I., Barbalau, A., Ionescu, R. T., Khan, F. S., Popescu, M., & Shah, M. (2021). Anomaly detection in video via self-supervised and multi-task learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12742–12752).

Goyal, P., Mahajan, D., Gupta, A., & Misra, I. (2019). Scaling and benchmarking self-supervised visual representation learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6391–6400).

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729–9738).

He, R., Liu, S., Wu, J., He, S., & Tang, K. (2023). Multi-domain learning from insufficient annotations. In *ECAI 2023* (pp. 1028–1035). IOS Press.

Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digital Medicine*, *6*(1), 74.

Huang, Y., Han, X., Chen, M., & Pan, Z. (2024). TGGS network: A multi-task learning network for gradient-guided knowledge sharing. *Knowledge-Based Systems*, *301*, 112254.

Huynh, T., Kornblith, S., Walter, M. R., Maire, M., & Khademi, M. (2022). Boosting contrastive self-supervised learning with false negative cancellation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2785–2795).

Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2020). A survey on contrastive self-supervised learning. *Technologies*, *9*(1), 2.

Kang, G., Jiang, L., Yang, Y., & Hauptmann, A. G. (2019). Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4893–4902).

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. *Advances in Neural Information Processing Systems*, *33*, 18661–18673.

Lee, S., & Son, Y. (2022). Multitask learning with single gradient step update for task balancing. *Neurocomputing*, *467*, 442–453.

Li, J., Zhou, P., Xiong, C., & Steven, C. H. H. (2021). Prototypical contrastive learning of unsupervised representations. In *ICLR*.

Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 4487–4496).

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99.

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2020). IoT, big data, and artificial intelligence in agriculture and food industry. *IEEE Internet of things Journal*, *9*(9), 6305–6324.

Motiian, S., Piccirilli, M., Adjeroh, D. A., & Doretto, G. (2017). Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 5715–5725).

Mu, E., Guttag, J., & Makar, M. (2024). Multitask contrastive learning. In *The international conference on learning representations*.

Pu, N., Zhong, Z., & Sebe, N. (2023). Dynamic conceptional contrastive learning for generalized category discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7579–7588).

Qi, L., Yang, H., Shi, Y., & Geng, X. (2024). MultiMatch: Multi-task learning for semi-supervised domain generalization. *ACM Transactions on Multimedia Computing, Communications and Applications*, *20*(6), 1–21.

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, *80*, 241–265.

Quan, S., Hirano, M., & Yamakawa, Y. (2023). Semantic information in contrastive learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5686–5696).

Schoenauer-Sebag, A., Heinrich, L., Schoenauer, M., Sebag, M., Wu, L. F., & Altschuler, S. J. (2019). Multi-domain adversarial learning. arXiv preprint arXiv:1903.09239.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., & Li, C.-L. (2020). FixMatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, *33*, 596–608.

Ullah, Z., Usman, M., & Gwak, J. (2023). MTSS-AAE: Multi-task semi-supervised adversarial autoencoding for COVID-19 detection based on chest x-ray images. *Expert Systems with Applications*, *216*, 119475.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Vandenhende, S., Georgoulis, S., & Van Gool, L. (2020). MTI-Net: Multi-scale task interaction networks for multi-task learning. In *Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part IV 16* (pp. 527–543). Springer.

Wang, Y., Tsai, Y.-H., Hung, W.-C., Ding, W., Liu, S., & Yang, M.-H. (2022). Semi-supervised multi-task learning for semantics and depth. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2505–2514).

Wang, Y., Zhang, P., Bai, L., & Xue, J. (2023). FEND: A future enhanced distribution-aware contrastive learning framework for long-tail trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1400–1409).

Wu, Z., Xiong, Y., Yu, S. X., & Lin, D. (2018). Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3733–3742).

Xiao, T., Liu, S., De Mello, S., Yu, Z., Kautz, J., & Yang, M.-H. (2022). Learning contrastive representation for semantic correspondence. *International Journal of Computer Vision*, *130*(5), 1293–1309.

Yang, X., Song, Z., King, I., & Xu, Z. (2022). A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(9), 8934–8954.

Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., & Shinozaki, T. (2021). FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, *34*, 18408–18419.

Zhou, F., Jiang, Z., Shui, C., Wang, B., & Chaib-draa, B. (2021). Domain generalization via optimal transport with metric similarity learning. *Neurocomputing*, *456*, 469–480.