Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.059115

ARTICLE

Dual-Task Contrastive Meta-Learning for Few-Shot Cross-Domain Emotion Recognition

Yujiao Tang[1], Yadong Wu[1,*], Yuanmei He[2], Jilin Liu[1] and Weihan Zhang[1]

[1]School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin, 644002, China

[2]School of Mechanical and Power Engineering, Chongqing University of Science and Technology, Chongqing, 401331, China

*Corresponding Author: Yadong Wu. Email: selobdvat@gmail.com

Received: 28 September 2024 Accepted: 18 November 2024 Published: 17 February 2025

ABSTRACT

Emotion recognition plays a crucial role in various fields and is a key task in natural language processing (NLP). The objective is to identify and interpret emotional expressions in text. However, traditional emotion recognition approaches often struggle in few-shot cross-domain scenarios due to their limited capacity to generalize semantic features across different domains. Additionally, these methods face challenges in accurately capturing complex emotional states, particularly those that are subtle or implicit. To overcome these limitations, we introduce a novel approach called Dual-Task Contrastive Meta-Learning (DTCML). This method combines meta-learning and contrastive learning to improve emotion recognition. Meta-learning enhances the model's ability to generalize to new emotional tasks, while instance contrastive learning further refines the model by distinguishing unique features within each category, enabling it to better differentiate complex emotional expressions. Prototype contrastive learning, in turn, helps the model address the semantic complexity of emotions across different domains, enabling the model to learn fine-grained emotions expression. By leveraging dual tasks, DTCML learns from two domains simultaneously, the model is encouraged to learn more diverse and generalizable emotions features, thereby improving its cross-domain adaptability and robustness, and enhancing its generalization ability. We evaluated the performance of DTCML across four cross-domain settings, and the results show that our method outperforms the best baseline by 5.88%, 12.04%, 8.49%, and 8.40% in terms of accuracy.

KEYWORDS

Contrastive learning; emotion recognition; cross-domain learning; dual-task; meta-learning

1 Introduction

With the increasing use of the internet, emotion recognition is expanding into a variety of new applications, especially on social media platforms where people frequently share their opinions and emotions. By analyzing user-generated content such as comments, posts, and expressions, we can gain valuable insights into their emotional states and interests, which can be used to enhance personalized recommendations and services. In mental health, analyzing social media posts can aid in the early detection of potential depressive symptoms, enabling timely intervention. Additionally, emotion recognition helps businesses understand consumer emotional needs, guiding adjustments in product and service strategies. However, practical applications often face challenges such as variations in data distribution and limited training data, caused by differences in data sources and volumes. Consequently, improving model robustness in the face of data scarcity, as well as effectively transferring emotional information across domains [1–4], remains a significant challenge in emotion recognition research.

In emotion recognition, addressing data limitations has led to the widespread use of few-shot learning methods. Meta-learning, which is essential in few-shot scenarios, has significantly advanced the field of emotion recognition with limited data [5–7]. The main advantage of meta-learning is its ability to enable rapid learning, allowing models to quickly acquire task-specific knowledge from a small number of samples and generalize to new tasks. Early research in data scarcity addressed the issue through data augmentation techniques [8,9], but the scope and effectiveness of these methods are often limited. In contrast, meta-learning offers a more adaptive approach, allowing models to personalize learning strategies and quickly adapt to new tasks, thereby increasing model flexibility.

In cross-domain emotion recognition, the limited availability of labeled data across domains, along with differences in style, vocabulary, and context, often results in decreased model performance when transferring from one domain to another. The challenge lies in effective feature transfer between the source and target domains [10–13], which significantly affects emotion recognition accuracy. Most research has focused on domain adaptation, a form of transfer learning [14,15] aimed at enhancing target domain performance through model transfer and cross-domain emotional feature capture. Bozorgtabar et al. [16] proposed an adversarial domain adaptation approach to achieve cross-domain emotion recognition in facial expression analysis by aligning features across domains. Similarly, Han et al. [17] introduced a model that combines meta-learning with adversarial domain adaptation (MLADA), using a meta-knowledge generator and adversarial domain discriminator to produce features that bridge source and target domains, enabling effective domain adaptation. However, adversarial learning in current approaches often faces training instability, particularly in emotion recognition, where the inherent complexity and ambiguity of emotions can cause model collapse during adversarial training. While adversarial learning is effective in addressing global distribution differences, it frequently lacks the ability to capture fine-grained emotional details. By contrast, contrastive learning facilitates cross-domain knowledge transfer by distinguishing unique sample features, helping the model better capture and understand subtle emotional nuances. Additionally, a dual-task approach enables the model to learn from data across both domains, allowing it to capture richer and more diverse features, which can improve performance in the source domain as well.

Based on this, we developed a dual-task contrastive learning framework to enhance the cross-domain generalization and emotion recognition abilities of deep learning models. In our dual-task design, we construct two unique meta-learning tasks from different domains, creating a dual-task structure that broadens the model's adaptability. Our framework, which combines dual-task and meta-learning components, allows the model to process data from both domains concurrently. Each iteration includes learning from both domain datasets and performing emotion recognition on these, dividing the data into support and query sets with data augmentation applied to prevent overfitting. Specifically, we use random token replacement and integrate both the augmented and original data into the encoder, creating positive and negative cases for each text. The inclusion of contrastive learning further strengthens the model's semantic comprehension and resilience. Prototype contrastive learning aids in accurately identifying distinct features for each emotion category, refining the boundaries between categories. In addition, instance-level contrastive learning brings representations of individual emotional instances closer to their augmented versions while distancing them from other instances. This approach enhances the model's sensitivity to subtle emotional differences. To make final predictions, we utilize a matching network with cosine similarity and convert labels into one-hot encoding, deriving classification results through the cross-entropy loss function. The methodology and technical innovations of our research are detailed in this paper, including each of the above techniques and steps.

- We combine meta-learning with a dual framework to enhance effective feature transfer across different domains. At the same time, meta-learning continuously adjusts the relationships between tasks, helping the model quickly adapt to new emotional tasks.
- We introduce the concept of prototype contrastive learning, which enables better differentiation of similar emotions that may exist within the same category by learning the prototypes for each emotional category.
- We introduce the concept of instance contrastive learning, which allows the model to distinguish the emotional states of each specific instance during training, enabling the learning of more latent emotional features and effectively handling more complex and diverse emotional expressions.
- Our method outperforms state-of-the-art models on four few-shot datasets, including a multi-category emotion dataset that we processed and a food-related dataset.

The rest of the paper is structured as follows. Section 2 reviews the relevant conceptual knowledge of this study, Section 3 describes DTCML and introduces the construction of the emotion recognition model, Section 4 provides a detailed description of the validation process of the emotion recognition model, including stability analysis, ablation experiments, and final visualizations and visual analysis. Section 5 discusses the main conclusions drawn and future work.

2 Related Work

2.1 Few-Shot Learning

Few-shot learning [18,19] enables models to quickly generalize to new tasks by leveraging prior knowledge and experience, even when only a limited number of samples are available. The core idea is to learn and generalize from just a few examples, allowing for accurate predictions or classifications when encountering new instances. This approach emulates human learning when facing new tasks, making it well-suited for data-scarce situations. To address this challenge, researchers have developed several methods, including the use of meta-learning. Meta-learning [20–22] simulates few-shot learning scenarios during training, enabling models to quickly acquire task-specific knowledge from limited samples. Zhang et al. [23] proposed a two-stage framework consisting of a meta-encoder and a base learner, which initializes label word embeddings using an external knowledge graph and continuously refines these embeddings, thereby enhancing the model's generalization and semantic representation for few-shot text classification. Another approach is the use of prior knowledge [24,25], such as category structures and similarities, which can provide additional constraints and assumptions to improve generalization in few-shot learning. For example, Qin et al. [26] introduced prior knowledge and attention mechanisms to meta-learning, proposing three stepwise methods Attention-based Meta-Learning (AML), Representation and Attention-based Meta-Learning (RAML), and Unsupervised Representation and Attention-based Meta-Learning (URAML) to integrate attention mechanisms and prior knowledge into meta-learning. Qin also identified overfitting issues in existing meta-learning methods and developed a new Cross-Entropy across Task (CET) metric to measure the impact of the Task-Over-Fitting (TOF) problem on these methods. Combining meta-learning with contrastive learning has become common in emotion recognition. Traditional contrastive learning often relies heavily on large sample sizes for model optimization, making it sample-dependent. This dependency poses limitations in few-shot and cross-domain tasks, and traditional contrastive learning can also lack precision in fine-grained emotion recognition. Integrating meta-learning with prototype contrastive learning and instance contrastive learning offers a solution to these challenges.

2.2 Cross-Domain Emotion Recognition

Cross-domain emotion recognition refers to the task of recognizing emotions across different domains. Previously, text emotion recognition primarily focused on modeling and training within specific domains. However, current emotion recognition efforts are beginning to address the issue of data distribution differences between different domains. As a result, various methods have been proposed to learn domain-invariant features, including adversarial learning and contrastive learning. Moreover, cross-domain emotional samples are scarce, particularly labeled data, making it challenging for emotion recognition models to transfer to other domains. Therefore, semi-supervised learning [27], unsupervised learning, and few-shot learning have gradually become hot topics, with multi-task learning [28,29] and reinforcement learning also being widely utilized. Nguyen et al. [30] proposed a deep cross-domain transfer learning framework that experimentally addresses emotion recognition through joint learning. Specifically, he first evaluated the performance of pre-trained models on the same data sources and different data sources. Then, he presented results for cross-domain transfer between visual and auditory domains. Finally, he validated the effectiveness of joint learning across multiple datasets. Fan et al. [31] introduced a cross-domain discriminative subspace classification algorithm specifically for text emotion recognition. After extracting deep features using a BiLSTM (Bidirectional Long Short-Term Memory) model, he incorporated conditional distribution adaptation and the distance between conditional distributions to enhance intra-class compactness and inter-class separability. Wang et al. [32] designed a decoupled loss function to learn domain information within emotion-specific features, and then further enhanced the quality of these features through prototype learning. From a causal perspective. Wang et al. [33] considered the causal relationships among text, emotion labels, and domains, and employed a backdoor adjustment method to eliminate domain bias, extracting the pure causal effects between text and emotion to address the issue of domain generalization. However, none of these methods effectively address the contextual semantic issues inherent in cross-domain challenges. To this end, we introduced a dual-task strategy in our model, which helps the model learn discriminative features across domains by processing data from both domains. This approach encourages the model to acquire richer semantic knowledge in different domains, thereby enhancing its generalization capability in the target domain.

2.3 Contrastive Learning

Contrastive learning [34–36] is a machine learning approach aimed at enhancing model performance by discerning the similarities and differences between samples. In this method, models are trained to maximize the similarity between positive sample pairs while minimizing the similarity between positive and negative pairs. Contrastive learning has demonstrated state-of-the-art performance in the unsupervised training of deep image models and is increasingly applied to text classification tasks, where it addresses data imbalance, strengthens feature learning, and effectively improves model classification performance and adaptability to real-world applications. When combined with self-supervised learning, contrastive learning can generate pseudo-labels or apply data augmentation to derive effective emotional representations from unlabeled data, reducing the need for large-scale labeled datasets. Khosla et al. [37] enhanced supervised classification tasks through supervised contrastive learning, noting that traditional methods using cross-entropy loss often fall short with large datasets and complex tasks. Azuma et al. [38] tackled domain shift by integrating contrastive learning with domain adaptation techniques, combining contrastive models with domain discriminators to achieve domain-invariant feature representations using adversarial loss.

However, these methods may struggle to capture nuanced or implicit emotional expressions, such as distinguishing "sadness" from "depression." To address this, we employ instance and prototype contrastive learning strategies. Prototype contrastive learning optimizes the model by clustering similar samples closer to their respective class prototypes, which is essential for fine-grained emotion recognition. Instance contrastive learning further refines the model by distinguishing unique features within each category, enabling it to better differentiate complex emotional expressions.

3 Methods

In this section, we will introduce the proposed DTCML (Dual-Task Contrastive Meta-Learning) framework. As shown in Fig. 1, we aim to combine contrastive learning to achieve efficient and stable accuracy in cross-domain emotion recognition with our model. To this end, Section 3.1 introduces the dual-task, Section 3.2 describes the data augmentation methods, and Section 3.3 details how to fine-tune BERT (Bidirectional Encoder Representations from Transformers) to overcome the impact of unknown domain samples on model performance. In Sections 3.4 and 3.5, we provide detailed explanations on how to apply contrastive learning to cross-domain emotion recognition. Finally, in Section 3.6, we describe how to predict emotion labels.

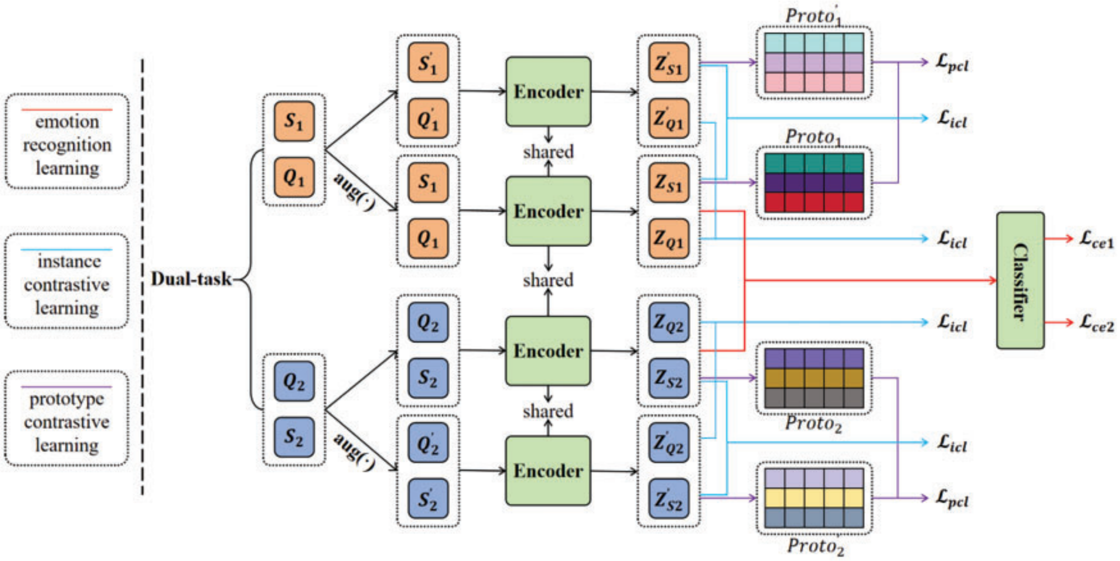


Figure 1: DTCML model. *icl* for instance contrastive learning loss, *ce* for classification loss and *pcl* for prototype contrastive learning loss

3.1 Problem Definition

In the few-shot cross-domain emotion recognition task, the entire dataset is divided into three distinct subsets: the training Set_{train}, the validation Set_{val}, and the test Set_{test}. The training Set_{train} provides essential learning resources for the Dual-Task Contrastive Meta-Learning (DTCML) model. The validation Set_{val} serves to assess the model's learning progress during training, while the test Set_{test} is utilized to evaluate the final performance of the model. Within the DTCML framework, each meta-learning task requires defining a support set S and a query set Q. For the $n - way\ k - shot$ emotion recognition, n categories are selected from the entire dataset, with $k \times m$ sample randomly drawn from each category. The support set consists of $n \times k$ labeled samples that provide essential feature learning for the DTCML model, whereas the query set contains $n \times m$ unlabeled samples. The main objective of the DTCML model is to utilize the limited labeled information in the support set S to enhance the classification capability on the samples in the query set Q. This approach significantly boosts the model's generalization capability, enabling effective feature recognition across categories even when labeled samples are scarce.

3.2 Dual-Task

Dual-task [39,40] strategy typically constructs a symmetrical model learning framework that inputs two different learning tasks simultaneously into the model to aid in acquiring domain knowledge. In our approach, we incorporate dual tasks into our framework to enhance the model's generalization capability and domain adaptability. Specifically, for the support set and query set of a single task, we generate augmented samples through data augmentation techniques. These samples help the model capture a greater diversity of emotional expressions. Subsequently, we input all these data into the model to obtain its semantic representations. Next, we compute the prototype representations, which capture the overall characteristics of each emotional category, and estimate the prototype contrastive learning loss pcl, as well as the instance contrastive learning loss icl. Crucially, by calculating the similarity matrix between the support set and the query set, we learn cross-domain emotion recognition, aiding the model in performing this task more effectively. Due to the symmetry of the dual-task strategy, the learning steps for the other task are identical to those described above, further enhancing the capacity for feature sharing and transfer. The dual-task strategy encourages the model to learn richer and more general features by simultaneously addressing two related tasks. This learning mechanism enables the model to better adapt to different tasks and domains, thereby improving its generalization ability. Additionally, by establishing relationships between samples through contrastive learning, the model can more effectively differentiate between different categories.

In addition, by incorporating contrastive learning, we introduce prototype contrastive loss and instance contrastive loss. The prototype contrastive loss is used to align the overall emotional features across domains, while the instance contrastive loss helps the model refine the distinctions between different emotional categories within a domain. By jointly optimizing these two types of losses, the model can capture domain-independent general emotional features while also learning to address the subtle differences between domains, leading to superior performance in cross-domain emotion recognition.

3.3 Datas Augmentation

Data augmentation has been shown to be beneficial for contrastive learning [41]. To enhance data diversity and improve the model's generalization ability, we used a robust data augmentation technique in our experiments. Specifically, before feeding data into the encoder, we randomly replaced a small portion of tokens in sentences to generate new sentences that maintain similar meanings with slight variations. This data augmentation strategy aims to introduce more sentence variations, enriching the training data and helping the model better capture semantic information and sentence representations. Furthermore, by reducing dependency on specific sentence structures, this approach enhances the model's understanding and clustering of input data, thus boosting its performance. In our method, we treat the augmented data as positive samples and the original unaugmented data as negative samples, allowing the model to compute similarity between the two domains through contrastive learning. By clustering similar samples closer and aligning samples with their respective category prototypes, the model effectively differentiates features across samples.

3.4 Fine-Tuning BERT

BERT [42] is a pre-trained encode language model based on the Transformer [43]. Our fine-tuning task utilizes the 10th, 11th, and 12th layers of the Encoder. This fine-tuning approach helps reduce the number of parameters and lowers the computational burden. Selecting specific layers for fine-tuning strikes a good balance between model complexity and performance, resulting in features that often exhibit stronger generalization capabilities. This means the model not only performs well on the training set but also maintains high performance on unseen data, thereby enhancing the model's adaptability. Here, BERT is used as a text encoder to map textual information into latent space. After inputting the augmented and original data into the encoder, we obtain the corresponding token embeddings. These embeddings are represented as points in a low-dimensional continuous space within the encoder network, where each point represents the semantic representation of a sentence. By encoding both augmented and original data, we obtain rich and expressive sentence embeddings that capture the semantic information and features of the sentences. These embeddings will serve as inputs for subsequent tasks, specifically for contrastive learning. In a given set, two N-way-K-shot tasks are provided $D_1 = \left\{\left(M_1^s, N_1^s\right), \left(M_1^q, N_1^q\right)\right\}$ and $D_2 = \left\{\left(M_2^s, N_2^s\right), \left(M_2^q, N_2^q\right)\right\}$, M_1^s is given by:

$$Eb_1^s = BERT\left(M_1^s\right), Eb_1^s \in R^{NK\times s\times h} \tag{1}$$

where s represents the length of BERT sentence embeddings, and h represents the dimensionality of BERT embeddings' hidden layers. Similarly, the embedding representations of M_1^q, M_2^s and M_2^q can both be derived from the aforementioned formula. They are respectively: $Eb_1^q \in R^{NK\times s\times h}$, $Eb_2^s \in R^{NK\times s\times h}$ and $Eb_2^q \in R^{NK\times s\times h}$.

3.5 Prototype Contrastive Learning

We employ prototype contrastive learning, aiming to bring similar samples closer and push dissimilar samples farther apart. Therefore, after obtaining the embedding vector representations, we use datas augmentation separately for Z_{S1} and Z'_{S1}, Z_{Q1} and Z'_{Q1}, Z_{S2} and Z'_{S2}, Z_{Q2} and Z'_{Q2}, as a positive pair, and similarly, Z_{Si} and Z'_{Si} as well. For each sample B = $\{Z_{Si}, Z_{Qi}\}_{i=1,2}^{M}$, we compute their similarity by comparing the distances between sample pairs. Specifically, we measure the similarity between them by comparing the distances between sample pairs and then use these similarity measures to construct the loss function. To compute the contrastive loss, we construct an adjacency matrix A' for B', which is a binary matrix of size 2M × 2M. For each sample pair, if they belong to the same category (positive sample), we want their distance to be closer to 0; if they belong to different categories (negative sample), we want their distance to be farther from 0. Therefore, we can express the contrastive loss as:

$$\mathrm{L}_{\mathrm{icl}} = \frac{1}{2M}\sum_{i=1}^{2M}\left[-\frac{1}{|C_i|}\sum_{j\in C_i}\log\frac{\exp\left(sim\left(Z_{Si}, Z_{Qi}\right)/\tau\right)}{\sum_{k\neq i}^{2M}\exp\left(sim\left(Z_{Si}, Z_{Qi}\right)/\tau\right)}\right] \tag{2}$$

$C_i \equiv \left\{A'_{i,j} = 1 | j \in \{1, \ldots, 2M\}\right\}$ represents the cardinality of the set of instances, where Z_{Si}, Z_{Qi} are positively correlated, Z_{Si}, Z_{Qi} are vector embeddings. τ is temperature, $sim(\cdot, \cdot)$ is a similarity function on a pair of normalized feature vectors.

Meanwhile, we partition the augmented and original datas into two groups each, specifically, Z_{S1} and Z_{Q1} as one group, totaling four groups, and compute the prototype as $P_i\ (i = 1, 2, 3, 4)$. For each sample x, we calculate its distance to each class prototype P_i. In our model, we employ cosine similarity as the distance metric to measure the similarity between samples and prototypes. The calculation of cosine similarity is as follows:

$$sim\left(x, P_i\right) = \frac{x \cdot P_i}{\|x\| \cdot \|P_i\|} \tag{3}$$

We utilize the computed similarity between samples and class prototypes to calculate the prototype contrastive learning loss. We chose prototype contrastive learning because it effectively clusters and classifies emotional features, thereby enhancing emotion recognition performance. By constructing prototypes, the model can learn the central characteristics of categories, which improves performance in cross-domain emotion recognition.

3.6 Instance Contrastive Learning

Instance Contrastive Learning (ICL) is a method designed to enhance data representation by examining the relationships between samples. Its main aim is to assist the model in acquiring meaningful representations by leveraging the similarities observed in different perspectives of a single sample. By increasing the similarity between various views of the same sample while decreasing the similarity across different samples, the model can effectively identify significant features, thereby improving its capacity to distinguish between different categories or semantic concepts in the representation space. Unlike prototype contrastive learning, which focuses on optimizing model performance using representative prototypes, instance contrastive loss functions specifically optimize the model by assessing the spatial relationships between individual sample instances. Typically, this loss function works by minimizing the distance between pairs of samples from the same class and maximizing the distance between pairs from different classes. After calculating the similarities of samples in set $B = \{Z_{Si}, Z_{Qi}\}_{i=1,2}^{M}$, based on the similarity matrix S, the support set sample with the highest similarity to each query set sample is selected as its contrast sample. The generated instance contrast samples are used to construct the contrastive loss function:

$$CLoss\left(x_q, x_p, x_n\right) = \frac{1}{N}\sum_{i=1}^{N}\left(1 - y_i\right)\frac{1}{2}D^2\left(x_q^i, x_p^i\right) + y_i\frac{1}{2}\max\left(0, m - D^2\left(x_q^i, x_p^i\right)\right) \tag{4}$$

Here, x_q denotes a query set sample, x_p denotes a contrast sample, x_n denotes a negative sample, N represents the number of samples, y_i is the binary indicator function, which equals 1 when x_q^i and x_p^i are of the same class and 0 otherwise, and D represents the euclidean distance or other distance metric.

In combining prototype contrastive learning and instance contrastive learning, prototype contrastive learning captures the central features of categories, providing a global perspective, while instance contrastive learning focuses on the fine-grained differences between samples. This combination allows the model to learn simultaneously at different levels, effectively enhancing the accuracy of emotion recognition. The choice of these two contrastive learning strategies lies in their complementary nature, collectively improving performance in cross-domain emotion recognition. Through prototype contrastive learning, the model can learn representative features of categories, while instance contrastive learning deepens the understanding of subtle differences among similar samples, enabling more accurate emotion classification.

3.7 *Emotion Label Prediction*

Finally, in our model, we use a matching network to compute cosine similarity, convert labels to one-hot encoding to generate predictions, and use the cross-entropy loss function to measure the difference between predicted and true labels. We successfully predict the labels of emotion words.

$$\mathrm{F.cr}oss\_entropy = -\log \frac{e^{-sim\left(x,P_y\right)}}{\sum_j e^{-sim\left(x,P_j\right)}} \tag{5}$$

where P_y represents the prototype vector corresponding to the true class of the sample, and j represents all classes. Emotion Label Prediction is as follows:

$$\mathrm{pred} = \mathrm{XQ} \cdot \mathrm{XS}^{\mathrm{T}} \cdot YS\_onehot \tag{6}$$

$$\mathrm{loss} = F.cross\_entropy\left(\left(XQ \cdot XS^{T}\right) \cdot YS\_onehot, YQ\right) \tag{7}$$

where XQ represents sample feature matrix in the query set, XS represents sample feature matrix in the support set, YS_onehot represents unique thermal coding of labels supporting centralised samples, pre$d$ represents predicted results, YQ represents the true labels of the samples in the query set. F.cr$oss_entropy$ represents the cross-entropy loss function.

4 Experiment

We conducted extensive experiments to validate the effectiveness of the proposed method. Section 4.1 provides a detailed overview of the experimental setup, including statistics on the datasets used, baseline comparisons, and implementation details. The performance of both the proposed method and baselines across various datasets is validated, with the analysis of the results presented in Section 4.2. Section 4.3 discusses the ablation study, while Section 4.4 presents the visualizations of the proposed approach.

4.1 *Experiment Setup*

4.1.1 Dataset

GoEmotions [44]: Proposed by Demszky et al. in 2021, this dataset contains 58,000 Reddit comments across 27 emotion categories, capturing a broad spectrum of emotional experiences, from basic emotions like joy, sadness, and anger to more nuanced ones like jealousy, surprise, and confusion. **YELP dataset:** Collected from the YELP platform, this dataset focuses on reviews within the food domain and includes 17 categories that address various aspects of food-related experiences. **DailyDialog [45]:** This dataset comprises over 10,000 manually annotated multi-turn dialogues gathered from an English learning website, covering topics such as daily life, culture and education, travel, and work. It includes annotations for 7 distinct emotion categories.

4.1.2 Baseline

Prototypical networks [46]: Designed to address few-shot learning challenges, prototypical networks aim to create an embedding space where each category is represented by a central prototype. This method employs a neural network to map input data into an embedding space, where the prototype of each category is defined as the mean of its support set within this space. Classification of a given query point is then performed by determining its closest category prototype. In our experiments, we followed the standard configurations typically associated with prototypical networks.

Induction networks [47]: Induction Networks are a type of inductive network designed for few-shot text classification. They capture relationships between samples and categories using dynamic routing and matrix transformations. In Induction Networks, samples are encoded as vectors through neural networks. A dynamic routing mechanism then interacts sample vectors with category vectors, assigning weights to each sample for each category. These weights are transformed into final category predictions through matrix transformations and nonlinear functions.

DS-FSL (Distributional Signatures Few-Shot Learning) [48]: DS-FSL is a text classification method that assesses word importance by weighting their frequencies and biases within specific documents in a collection. This helps identify word relevance for different categories, improving classification accuracy. Using limited labeled data, DS-FSL estimates word importance for target categories and refines this estimation through a meta-learning framework. By leveraging distributional features, DS-FSL enhances the model's generalization to new categories.

MLADA (Meta-Learning Adversarial Domain Adaptation Network) [17]: MLADA is a model designed for few-shot text classification. It combines adversarial domain adaptation with episode-based meta-learning techniques to enhance model performance with limited data. MLADA leverages domain-adversarial tasks to expand training data and extract transferable features through meta-learning. The model consists of several key components: a word representation layer that encodes each word as a vector, a domain discriminator to distinguish between source and target domain samples, and a meta-knowledge generator that uses bidirectional LSTM to create context embeddings. The interaction layer merges transferable features with sentence-specific features to form sentence embeddings, while the classifier, trained on support sets, produces the final classification results.

TART (Task-Adaptive Reference Transformation) [49]: TART is a method for enhancing few-shot text classification tasks. It improves model generalization by constructing task-adaptive metric spaces. TART uses linear transformation matrices to project category prototypes to fixed reference points for each category, enhancing differences between category prototypes in transformation space. Additionally, TART introduces a discriminative reference regularization method that maximizes differences between transformed category prototypes in the task-adaptive metric space, further improving performance.

DualAN (Dual Adversarial Network) [50]: primarily explores a meta-learning-based knowledge transfer method for addressing few-shot text classification problems. It begins by acquiring task-relevant domain knowledge and then utilizes word embeddings along with this domain knowledge to compute sentence representations. A domain discriminator is introduced to differentiate between knowledge from different domains, and a ridge regression classifier is employed for classification. During model training, implement a dual adversarial training strategy, simultaneously training the knowledge generator, classifier, and domain discriminator. This approach enables the model to learn how to transfer knowledge from the support set during each meta-training cycle, allowing for effective classification on the query set.

BiLSTM-Attention-CNN [51] is a model that combines Bidirectional Long Short-Term Memory networks (BiLSTM), attention mechanisms, and Convolutional Neural Networks (CNN), and is commonly used in text classification and emotion recognition tasks. It utilizes methods such as word2vec to convert words in the text into word vectors, preserving semantic information. The word vectors are then input into the BiLSTM to generate contextual representations of the words. The attention mechanism is used to compute the context vector for each word, extracting key information. The output from the attention mechanism is fed into the CNN to extract salient topic features. Finally, a fully connected layer is used to classify the extracted features, resulting in the categorization of the text.

4.1.3 Experiment Detail

The BERT encoder is a lightweight version based on BERT proposed in the Hugging Face code repository, used as a text encoder. When using BERT as the encoder, we selected a model pretrained with specific strategies as the pretrained model. We strictly limited the number of BERT's output layers, choosing only the outputs from layers 10, 11, and 12 as the final encoder output.

During the model training process, we experimented with four datasets: Y(Yelp) → D(Daily Dialog), Y(Yelp)→ G(GoEmotions), G(GoEmotions) → D(DailyDialog), and G(GoEmotions) → Y(Yelp). We used the Adam optimizer for parameter optimization with a learning rate set to 0.00001. To prevent overfitting, we randomly partitioned the training, testing, and validation sets, and used an early stopping strategy to prevent performance degradation on the validation set. Specifically, training was stopped if the accuracy on the validation set did not improve over 10 epochs to avoid overfitting.

During meta-training, we performed 100 training tasks per epoch to assess model performance. We evaluated the model's classification accuracy and scenario standard deviation using a test set containing 1200 samples, verifying the model's stability across different domains. This experimental design comprehensively evaluates the model's generalization ability and performance across various scenarios.

All experiments were conducted on Nvidia Geforce RTX 6000 GPUs.

4.2 Experiment Result and Analysis

The experimental results of the six methods on the four datasets are shown in Table 1 and Fig. 2. Based on these results, we made the following three observations.

Table 1: The experimental results of the four methods for the Y → D, Y → G, G → D, and G → Y transfer tasks are presented. Bold and underlined entries indicate the best and second-best results, respectively

Model	Y→D		Y→G		G→D		G→Y	
	3-way 1-shot	3-way5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot
Prototypical	46.04 ± 1.13	58.28 ± 1.02	47.31 ± 1.23	59.14 ± 1.17	46.84 ± 1.13	61.67 ± 1.03	58.16 ± 1.41	71.79 ± 1.23
Induction	44.77 ± 1.23	52.74 ± 1.10	49.24 ± 1.52	62.03 ± 1.28	48.99 ± 1.26	55.20 ± 1.10	63.74 ± 1.61	75.14 ± 1.32
DS-FSL	45.44 ± 0.85	58.39 ± 0.84	44.07 ± 0.84	57.19 ± 0.92	45.45 ± 0.88	58.42 ± 0.83	48.77 ± 1.02	64.60 ± 1.10
MLADA	45.07 ± 0.91	56.42 ± 0.87	48.87 ± 1.02	65.29 ± 1.07	43.80 ± 0.87	58.81 ± 0.89	53.75 ± 1.17	67.79 ± 1.16
TART	47.09 ± 1.10	62.20 ± 0.88	50.38 ± 1.19	64.83 ± 0.96	44.92 ± 0.97	60.96 ± 0.89	54.98 ± 1.24	73.24 ± 1.15
DualAN	47.92 ± 1.23	56.54 ± 1.12	52.98 ± 1.34	64.95 ± 1.26	43.98 ± 1.12	56.75 ± 1.09	52.23 ± 1.34	63.62 ± 1.17
BiLSTM-Attention-CNN	45.34 ± 1.33	59.29 ± 1.19	51.34 ± 1.48	64.96 ± 1.37	50.04 ± 1.31	63.80 ± 1.18	63.35 ± 1.64	76.80 ± 1.34
DTCML	**51.53 ± 1.23**	**64.22 ± 1.10**	**59.16 ± 1.54**	**70.73 ± 1.10**	**54.77 ± 1.36**	**66.96 ± 1.06**	**65.86 ± 1.39**	**77.65 ± 1.33**
Model	**Y→D**		**Y→G**		**G→D**		**G→Y**	
	5-way 1-shot	5-way5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Prototypical	30.96 ± 0.75	42.49 ± 0.70	34.01 ± 0.94	44.70 ± 0.84	35.15 ± 0.80	46.00 ± 0.69	34.72 ± 0.63	56.46 ± 0.91
Induction	31.42 ± 0.77	39.98 ± 0.71	40.78 ± 1.18	49.37 ± 1.09	34.21 ± 0.85	43.72 ± 0.87	49.23 ± 1.16	61.86 ± 1.07
DS-FSL	32.12 ± 0.62	44.40 ± 0.61	30.78 ± 0.63	44.01 ± 0.74	31.72 ± 0.61	43.53 ± 0.79	35.32 ± 0.75	51.10 ± 0.88
MLADA	32.02 ± 0.66	44.24 ± 0.62	35.51 ± 0.75	54.04 ± 0.83	32.59 ± 0.69	46.05 ± 0.83	43.09 ± 0.93	59.32 ± 1.01
TART	30.52 ± 0.89	45.43 ± 0.62	37.39 ± 0.91	55.82 ± 0.84	33.54 ± 0.91	44.57 ± 0.59	43.53 ± 0.97	57.43 ± 0.97
DualAN	32.50 ± 0.85	45.31 ± 0.85	35.90 ± 0.97	54.84 ± 0.96	34.14 ± 0.72	44.31 ± 0.83	45.60 ± 0.91	57.36 ± 0.91

(Continued)

Table 1 (continued)

Model	Y→D		Y→G		G→D		G→Y	
	3-way 1-shot	3-way5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot
BiLSTM-Attention-CNN	31.04 ± 0.83	43.36 ± 0.82	40.33 ± 1.17	51.66 ± 1.12	36.24 ± 0.94	50.15 ± 0.86	49.45 ± 1.19	65.24 ± 1.14
DTCML	**37.94 ± 0.99**	**49.16 ± 0.78**	**44.89 ± 1.14**	**59.55 ± 1.00**	**38.41 ± 0.93**	**53.90 ± 0.88**	**51.05 ± 1.10**	**67.41 ± 1.05**

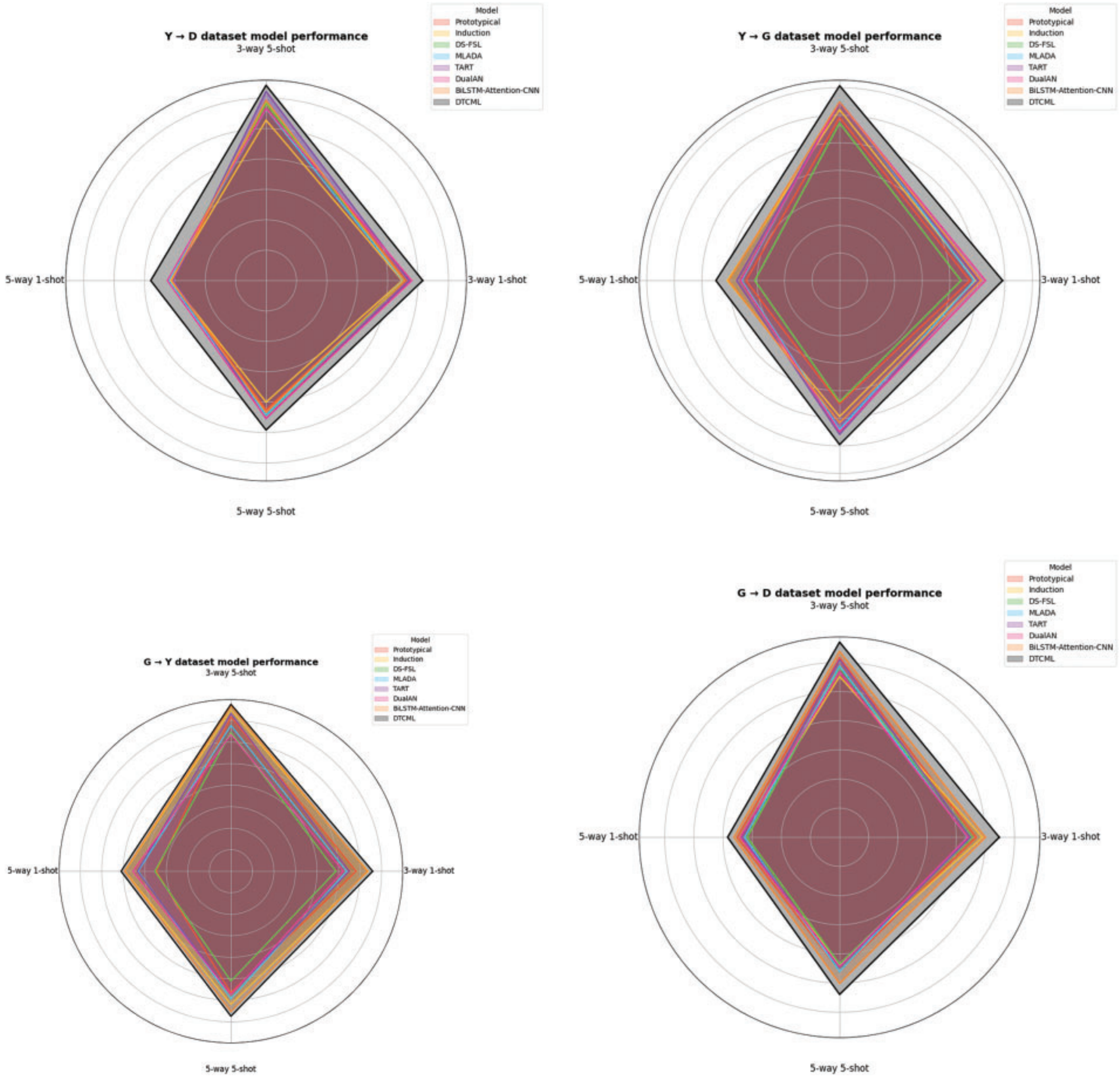


Figure 2: Experiment results of the six methods on three datassets

Firstly, under all datasets and settings, Dual-Task Contrastive Meta-Learning (DTCML) significantly outperforms all baseline methods. Compared to the second-best methods, DTCML improves performance by an average of 4.48%. For various cross-domain transfer tasks, DTCML shows average performance improvements of 5.88%, 12.04%, 8.49%, and 8.40% on Yelp→DailyDialog, Yelp→GoEmotions, GoEmotions→DailyDialog, and GoEmotions→Yelp, respectively. These results thoroughly demonstrate the effectiveness and superiority of the DTCML method. This outstanding performance is attributed to the synergistic combination of multi-domain datasets and contrastive learning, significantly enhancing the training quality of contrastive learning models. Additionally, by employing dual tasks where two tasks are simultaneously inputted in one episode, DTCML obtains more reliable supervision signals from different domains, further strengthening the model's robustness.

Secondly, in the field of cross-domain few-shot emotion recognition, our DTCML model demonstrates significant performance advantages. Compared to adversarial network-based approaches such as TART and MLADA, DTCML exhibits greater robustness when handling data from different domains, particularly in the learning of emotional features. By leveraging a dual-task contrastive learning mechanism, DTCML achieves performance improvements across various task settings. For instance, in the 3-way 1-shot task on the G→Y dataset, DTCML reaches an impressive performance of 65.86%, clearly outperforming other models. This enhancement in performance can be attributed to DTCML's ability to effectively capture emotional features within cross-domain data. Unlike traditional models that face challenges due to inconsistent feature distributions across different domains, DTCML promotes a deeper understanding of similar samples through contrastive learning, thereby better adapting to the emotional information in varying domains. Additionally, we also included a model based on the dual-task learning framework, DualAN, and the BiLSTM-Attention-CNN baseline model. The experimental results show that both methods can achieve good performance. However, the proposed method still outperforms them. We analyze that DualAN effectively addresses the few-shot problem, but its performance in cross-domain few-shot emotion recognition is not ideal. This may be due to its inability to adequately solve the issue of domain generalization. The BiLSTM-Attention-CNN architecture is typically used for feature extraction from samples. However, in few-shot learning, the scarcity of samples limits the ability of the BiLSTM-Attention-CNN neural network combination to learn sufficient sample features, leading to overfitting.

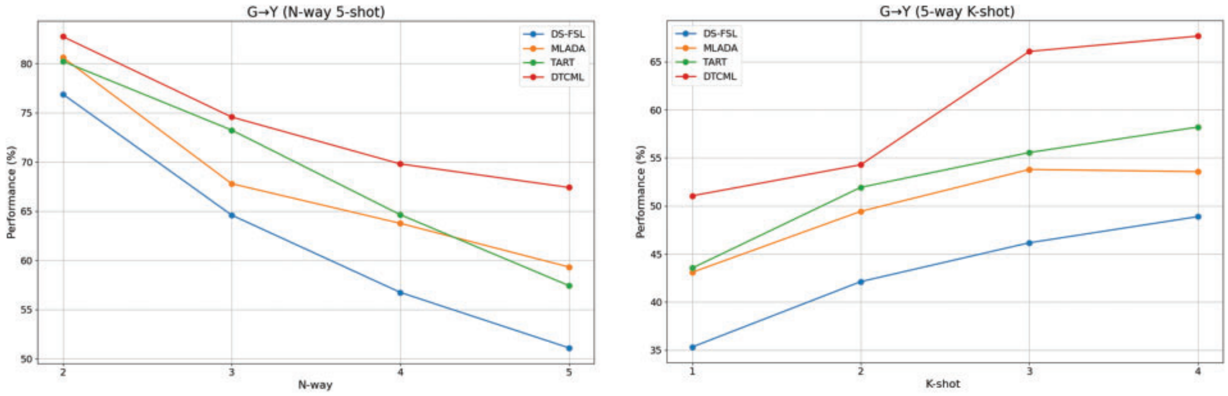
Thirdly, models in the realm of cross-domain few-shot text classification exhibit a greater sensitivity to sample quantity than to the number of methods employed. For example, the performance of models in a 3-way 1-shot setting is, on average, 14.75% higher than in a 5-way 1-shot scenario. Similarly, the average accuracy in 3-way 5-shot classification surpasses that of 5-way 5-shot classification by 15.26%. These findings suggest that in few-shot learning, both controlling the number of categories and increasing the sample size are crucial strategies for enhancing model performance. Reducing the number of categories can improve performance by minimizing confusion among fewer categories, thereby simplifying the classification task for the model. This is particularly important during domain transfer between training and testing data, where the model may become significantly confused, leading to a notable decline in performance. Additionally, this insight clarifies why methods that leverage supplementary data, such as DTCML, MLADA, and TART, frequently outperform other approaches. Their superiority largely stems from the effective mitigation of bias through the incorporation of more labeled data.

4.3 Stability Analysis

Given the scarcity of labeled data in each N-way-K-shot task, few-shot models demonstrate notable sensitivity to changes in the few-shot configurations. Therefore, experiments were carried out to evaluate the stability of the models across a spectrum of N-way-K-shot settings, as presented in Table 2 and Fig. 3.

Table 2: The stability of DTCML and baseline models was evaluated across various N-way-K-shot settings

Model	N-way-5-shot on G→Y			
	N = 2	N = 3	N = 4	N = 5
DS-FSL	76.89 ± 1.18	64.60 ± 1.10	56.74 ± 0.94	51.10 ± 0.88
MLADA	80.62 ± 1.22	67.79 ± 1.16	63.76 ± 1.02	59.32 ± 1.01
TART	80.23 ± 1.20	73.24 ± 1.15	64.64 ± 1.07	57.43 ± 0.97
DTCML	**82.74** ± **1.39**	**74.57** ± **1.23**	**69.80** ± **1.12**	**67.41** ± **1.05**
Model	5-way-K-shot on G→Y			
	K = 1	K = 2	K = 3	K = 4
DS-FSL	35.32 ± 0.75	42.12 ± 0.79	46.16 ± 0.85	48.88 ± 0.87
MLADA	43.09 ± 0.93	49.43 ± 0.90	53.79 ± 0.90	53.56 ± 0.86
TART	43.53 ± 0.97	51.93 ± 0.95	55.55 ± 0.92	58.20 ± 0.97
DTCML	**51.05** ± **1.10**	**54.28** ± **1.10**	**66.07** ± **1.14**	**67.66** ± **1.10**



(a) 5-way-K-shot on G→Y (b) N-way-5-shot on G→Y

Figure 3: Stability assessment of DTCML and baseline models across various N-way-K-shot settings

As the number of samples K increases, all models generally exhibit improved classification performance, as shown in Fig. 3a. In contrast, an increase in the number of categories N leads to a noticeable decline in performance, as illustrated in Fig. 3b. Notably, the DTCML model consistently maintains the highest performance across various experimental settings, demonstrating its robustness and consistency. This advantage stems from the model's effective use of dual-task contrastive learning, which enables it to draw insights from multiple N-way-K-shot tasks across diverse domains. Consequently, the DTCML model excels at accurately classifying data from different domains within each set. Through iterative training, the model enhances its adaptability to the unique features of these domains, resulting in significant performance gains.

4.4 Ablation Experiments

Table 3 displays the findings from the ablation study conducted on the two datasets. Fig. 4 illustrates how the number of N-way-K-shot tasks influences model performance. A dual-task network was introduced, where two tasks are processed within a single network and trained using contrastive learning techniques. Additionally, several variants were explored to investigate the underlying sources of effectiveness.

- W/1 Task : Without dual-task training, only one task is accepted for training.
- W/2 Task: Elimination of instance contrastive learning, accepting only prototype contrastive loss.
- W/3 Task: Elimination of prototype contrastive learning, accepting only instance contrastive loss.

From the results presented in Table 3 and Fig. 4, three key conclusions can be drawn. First, our method consistently outperforms others across all datasets and settings. Second, the performance of the three individual tasks does not exceed that of our method. The dual-task mechanism acts as a regularization strategy during the model training process, facilitating the learning of more generalized feature representations. By incorporating dual tasks, the model is compelled to develop the ability to adapt to multiple tasks concurrently, enhancing its overall generalization performance. Consequently, removing the dual-task component from DTCML may lead to a decline in performance on specific tasks, as the model loses the mechanism for joint optimization across multiple tasks, thereby restricting its learned feature representations.

Table 3: The ablation studies on the four datasets

Model	Y→G		G→Y		Y→G		G→Y	
	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot	3-way 1-shot	3-way 5-shot
W/1 task	50.53 ± 1.25	65.78 ± 1.25	51.13 ± 1.35	73.91 ± 1.35	40.59 ± 1.12	54.50 ± 0.99	48.71 ± 1.10	60.20 ± 1.07
W/2 task	56.77 ± 1.49	62.68 ± 1.28	65.02 ± 1.52	74.51 ± 1.33	37.44 ± 0.95	57.46 ± 1.06	51.10 ± 1.15	64.42 ± 1.09
W/3 task	53.95 ± 1.38	68.86 ± 1.25	65.14 ± 1.56	74.94 ± 1.36	36.05 ± 0.92	57.26 ± 1.07	51.09 ± 1.14	62.12 ± 1.12
DTCML	**59.16 ± 1.54**	**70.73 ± 1.10**	**67.49 ± 1.53**	**78.55 ± 1.34**	**44.89 ± 1.14**	**59.55 ± 1.00**	**54.77 ± 1.21**	**65.17 ± 1.13**

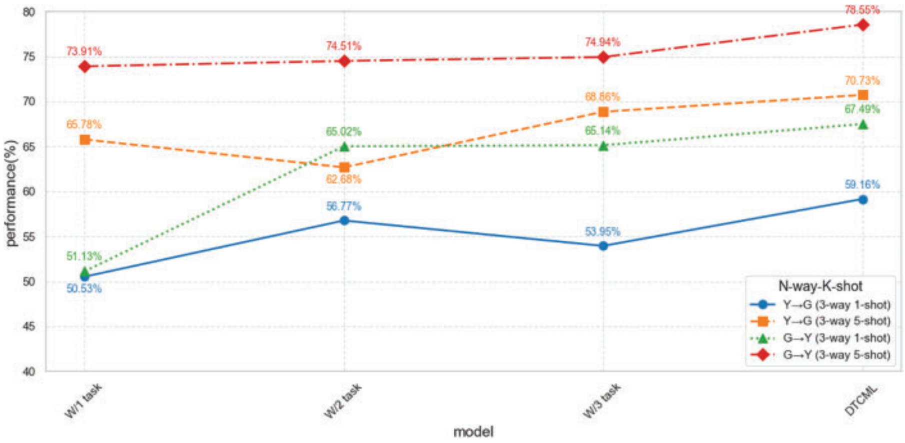


Figure 4: The impact of the number of N-way-K-shot tasks on model performance

Similarly, instance contrastive learning establishes relationships among samples by comparing their similarities, which aids the model in acquiring more discriminative feature representations. Without prototype contrastive learning, the model struggles to adequately identify similarities between samples, resulting in diminished differentiation among various categories. Lastly, prototype contrastive learning maps samples into a prototype space, allowing for the learning of category prototypes at the task level, which fosters better generalization to new tasks. Therefore, the absence of prototype contrastive learning hinders the model's ability to effectively learn shared features across tasks, leading to decreased performance on novel tasks. As indicated in Table 3, the performance of both tasks experiences a significant drop after the removal of contrastive learning, demonstrating that models trained without this mechanism face challenges in handling data from different domains. This further underscores the importance of dual-contrastive learning in enhancing domain adaptation for few-shot models.

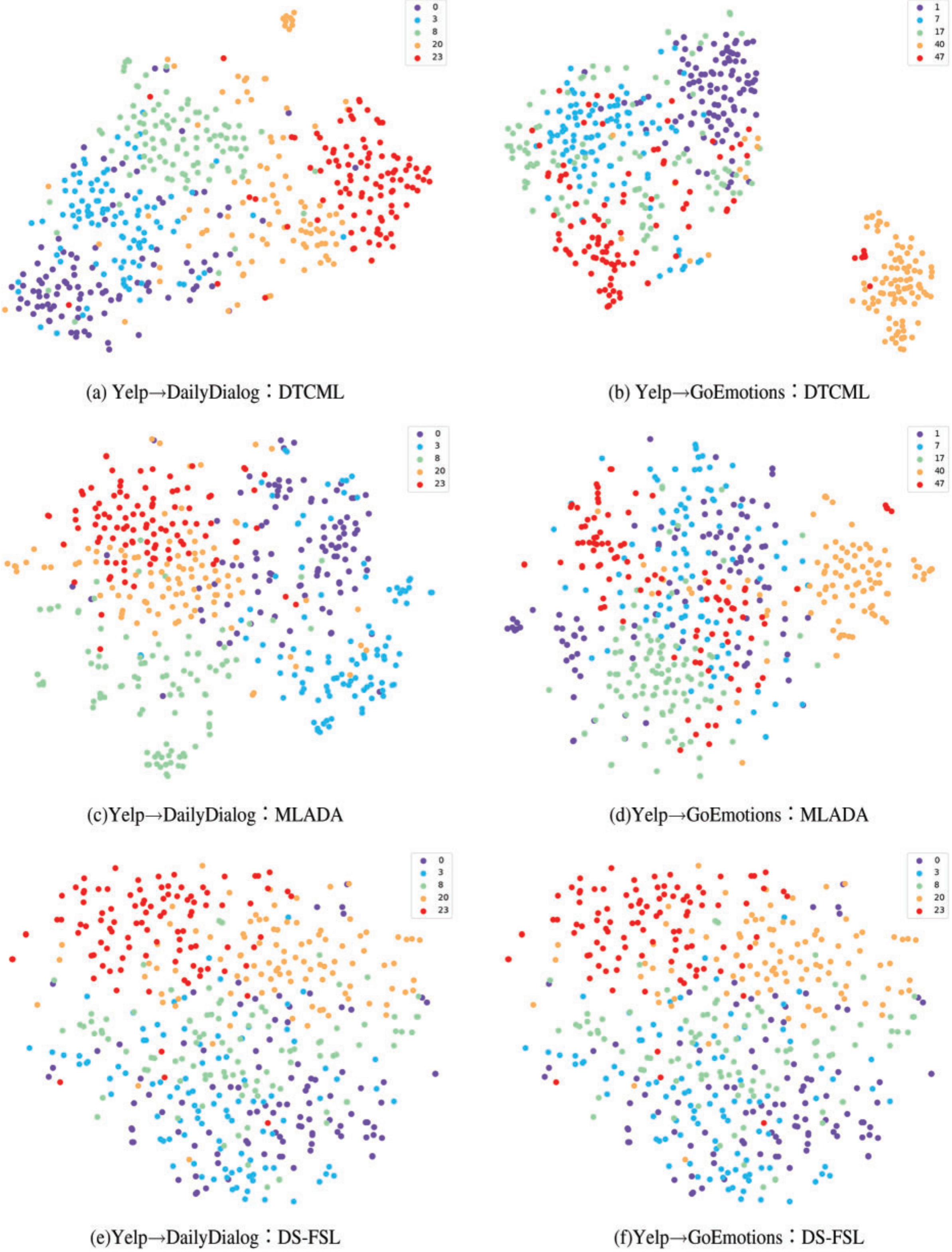
4.5 Visualization

To visually demonstrate the effectiveness of the DTCML model in distinguishing emotional categories, we employed T-SNE (T-Distributed Stochastic Neighbor Embedding) to visualize the latent space. We conducted 5-way 1-shot model training on the Yelp→DailyDialog and Yelp→GoEmotions datasets, randomly selecting five categories from the test set, with each category containing 100 samples. The visualization results are shown in Fig. 5. These t-SNE plots illustrate how the model clusters and separates different emotional categories in the latent space.

The figures clearly show that our method, DTCML, produces feature spaces with more distinct and separable representations across both datasets, whereas the feature distributions of other baseline models are more mixed. In Fig. 5b, for instance, samples of the same class are tightly clustered, especially for class 40, which effectively pushes apart representations from different classes. This clustering of similar samples demonstrates the effectiveness of DTCML in enhancing class separation in cross-domain emotion recognition. Due to the nature of contrastive learning, DTCML creates a latent space where samples from the same class are grouped closely together, while those from different classes are pushed farther apart. This approach improves the discriminative power of the model for emotional features. In contrast, the MLADA and DS-FSL models show more overlap between class distributions, with less defined boundaries between classes, further highlighting the superior performance of DTCML in distinguishing emotional categories. These visualization results align with the quantitative findings from our experiments, confirming that DTCML not only improves performance across various cross-domain emotion recognition tasks but also generates more discriminative feature vectors, leading to better emotion classification accuracy.

4.6 Error Analysis

To demonstrate that our method can maintain good accuracy even under random data sampling conditions, we have introduced a new error analysis experiment. The data for this experiment was randomly sampled from 100 episodes in the test set. The prediction results for the query set are shown in the Fig. 6.



(a) Yelp→DailyDialog：DTCML

(b) Yelp→GoEmotions：DTCML

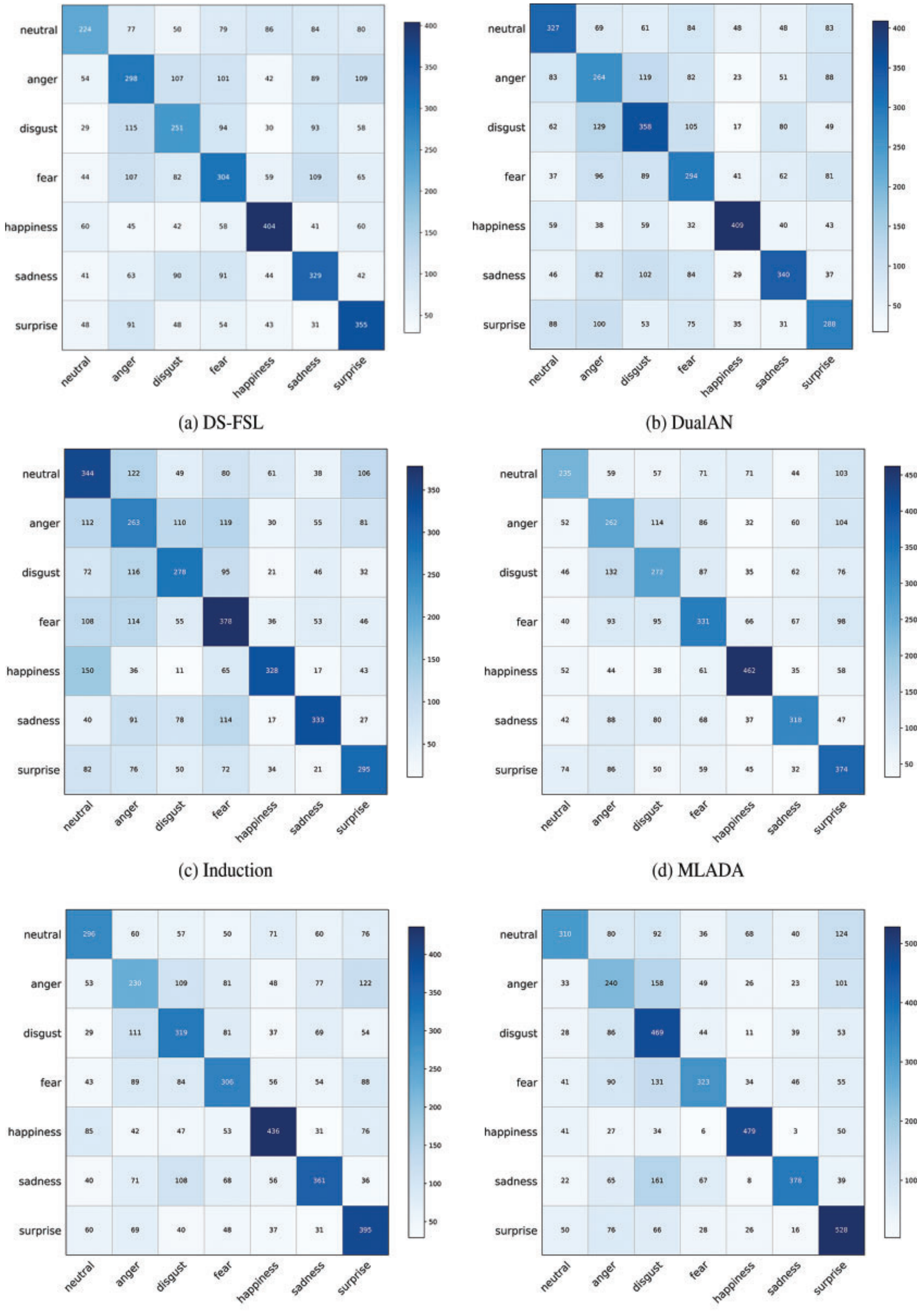
(c)Yelp→DailyDialog：MLADA

(d)Yelp→GoEmotions：MLADA

(e)Yelp→DailyDialog：DS-FSL

(f)Yelp→GoEmotions：DS-FSL

Figure 5: Visualization of the latent space for DTCML, MLADA, and DS-FSL using t-SNE under 5-way 1-shot on the Yelp→DailyDialog and Yelp→GoEmotions datasets



(a) DS-FSL

(b) DualAN

(c) Induction

(d) MLADA

(e) TART

(f) DTCML

Figure 6: Error analysis

The figure clearly demonstrates that our proposed method outperforms other models in terms of correctly predicted categories, with fewer misclassifications. In Fig. 6f, the values along the diagonal are significantly higher than those of the baseline models, and the color intensity is also more pronounced. For example, in the "happiness" category, the surrounding values are all single digits, which strongly indicates a lower error rate in our model. This improvement can be attributed to our use of contrastive learning, which effectively differentiates complex emotions within the samples. Specifically, when compared to the DualAN model in Fig. 6b, our method shows a substantial increase in correct predictions, further emphasizing the effectiveness of the dual-task approach in improving the model's adaptability across domains.

5 Conclusion

Recent research emphasizes the difficulties in transferring emotional features effectively across various domains and capturing semantic characteristics in textual contexts during cross-domain sentiment analysis. To tackle these issues, we introduce a method called Dual-Task Contrastive Meta-Learning (DTCML). This method constructs a training framework using a multi-domain dataset to improve the model's ability to adapt to different domains and learn features that are invariant across them. Once the model has undergone meta-training, it can generalize to previously unseen domains without the need for additional retraining or fine-tuning. Additionally, by combining dual-task networks with contrastive learning techniques, we utilize multi-domain training data to enhance the model's flexibility in adapting to a range of domains. We performed comprehensive experiments to assess the effectiveness of our proposed method, and the results indicate that DTCML surpasses all other methods across four datasets, achieving an average performance boost of 4.48% compared to the second-best approach. These findings substantiate the efficacy of DTCML. Our experimental analysis indicates that effectively addressing few-shot problems hinges on reducing few-shot bias. DTCML accomplishes this by employing two tasks within each task set, which significantly diminishes bias and enhances model performance. Additionally, we performed ablation experiments to identify the sources of the DTCML model's effectiveness, revealing that the inclusion of two tasks from different domains within each task set contributes to the model's ability to generate improved domain knowledge.

Despite certain advantages, DTCML has some limitations. First, the method heavily relies on diverse multi-domain datasets, which may not always be available in practical applications. This can impact the model's transferability, especially when there are significant domain gaps. Second, although DTCML has achieved certain success in capturing complex or implicit emotional expressions, further exploration is needed for more subtle emotional differences. Lastly, the complexity of the dual-task framework may increase computational costs, potentially affecting scalability in larger datasets or real-time applications.

Acknowledgement: We are deeply grateful to Qian Wei and the reviewers for their proofreading and revisions of this article. Without their support, it would not have been possible to submit this article in its current form. We also express our sincere thanks to Computers, Materials & Continua for their recognition of this work. Additionally, we would like to extend our gratitude to each of the funding bodies for their strong support of this article. Without their support, it would have not been possible to submit this in the current form.

Funding Statement: This work was supported by the Scientific Research and Innovation Team Program of Sichuan University of Science and Technology (No. SUSE652A006) and Sichuan Key Provincial

Research Base of Intelligent Tourism (ZHYJ22-03). In addition, it is also listed as a project of Sichuan Provincial Science and Technology Programme (2022YFG0028).

Author Contributions: Conceptualization, methodology and validation, Yujiao Tang; resources, data curation, and visualization, Yujiao Tang, Yadong Wu; writing original draft preparation, Yujiao Tang, Yuanmei He; writing review and editing, formal analysis, Yujiao Tang, Yadong Wu, Yuanmei He, Jilin Liu, Weihan Zhan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] M. Xu *et al.*, "Adversarial domain adaptation with domain mixup," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, pp. 6502–6509, 2020. doi: 10.1609/aaai.v34i04.6123.

[2] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 1180–1189.

[3] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, 2010.

[4] Y. Dai, J. Liu, X. Ren, and Z. Xu, "Adversarial training based multi-source unsupervised domain adaptation for sentiment analysis," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 5, pp. 7618–7625, 2020. doi: 10.1609/aaai.v34i05.6262.

[5] J. Chen, R. Zhang, Y. Mao, and J. Xu, "ContrastNet: A contrastive learning framework for few-shot text classification," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 10, pp. 10492–10500, 2022. doi: 10.1609/aaai.v36i10.21292.

[6] P. Sun, Y. Ouyang, W. Zhang, and X. Dai, "MEDA: Meta-learning with data augmentation for few-shot text classification," in *Proc. Thirtieth Int. Joint Conf. Artif. Intell.*, 2021, pp. 3929–3935.

[7] L. Yan, Y. Zheng, and J. Cao, "Few-shot learning for short text classification," *Multimed. Tools Appl.*, vol. 77, no. 22, pp. 29799–29810, 2018. doi: 10.1007/s11042-018-5772-4.

[8] X. Wu, C. Gao, M. Lin, L. Zang, Z. Wang and S. Hu, "Text smoothing: Enhance various data augmentation methods on text classification tasks," 2022, *arXiv:2202.13840*.

[9] B. Raskutti, H. Ferra, and A. Kowalczyk, "Combining clustering and co-training to enhance text classification using unlabelled data," in *Proc. Eighth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2002, pp. 620–625.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2009.

[11] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," in *Advances in Data Science and Information Engineering*. Cham: Springer, 2021, pp. 877–894.

[12] G. Csurka, "Domain adaptation for visual applications: A comprehensive survey," 2017, *arXiv:1702.05374*.

[13] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," *Adv. Neural Inf. Process. Syst.*, vol. 19, 2006.

[14] L. Torrey and J. Shavlik, "Transfer learning," in *Handb. Res. Mach. Learn. Appl. Trends: Algorithms, Methods, Tech.*, IGI Global, 2010, pp. 242–264.

[15] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[16] B. Bozorgtabar, D. Mahapatra, and J. -P. Thiran, "ExprADA: Adversarial domain adaptation for facial expression analysis," *Pattern Recognit.*, vol. 100, 2020, Art. no. 107111.

[17] C. Han, Z. Fan, and D. Zhang, "Meta-learning adversarial domain adaptation network for few-shot text classification," 2021, *arXiv:2107.12262*.

[18] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, 2020.

[19] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," 2022, *arXiv:2203.04291*.

[20] K. Parvaiz, M. Azam, F. Nasim, S. Noor, and K. Ayub, "Cross-domain sentiment analysis: A multi-task learning approach with shared representations," *J. Comput. Biomed. Inf.*, vol. 7, no. 2, 2024.

[21] X. Li, Z. Sun, J. H. Xue, and Z. Ma, "A concise review of recent few-shot meta-learning methods," *Neurocomputing*, vol. 456, pp. 463–468, 2021. doi: 10.1016/j.neucom.2020.05.114.

[22] J. Xu and Q. Du, "Learning transferable features in meta-learning for few-shot text classification," *Pattern Recognit. Lett.*, vol. 135, no. 6266, pp. 271–278, 2020. doi: 10.1016/j.patrec.2020.05.007.

[23] H. Zhang, X. Zhang, and H. Huang, "Prompt-based meta-learning for few-shot text classification," in *Proc. 2022 Conf. Empir. Methods Natural Lang. Process.*, 2022, pp. 1342–1357.

[24] S. Tobias, "Interest, prior knowledge, and learning," *Rev. Educ. Res.*, vol. 64, no. 1, pp. 37–54, 1994. doi: 10.3102/00346543064001037.

[25] P. Kitcher, "A priori knowledge," *Philos. Rev.*, vol. 89, no. 1, pp. 3–23, 1980. doi: 10.2307/2184861.

[26] Y. Qin *et al.*, "Prior-knowledge and attention based meta-learning for few-shot learning," *Knowl. Based Syst.*, vol. 213, no. 1, 2021, Art. no. 106609. doi: 10.1016/j.knosys.2020.106609.

[27] R. Zhang, H. F. Guo, Z. X. Xu, Y. X. Hu, M. M. Chen and L. P. Zhang, "MGFKD: A semi-supervised multi-source domain adaptation algorithm for cross-subject EEG emotion recognition," *Brain Res. Bull.*, vol. 208, 2024, Art. no. 110901. doi: 10.1016/j.brainresbull.2024.110901.

[28] X. Cai, J. Yuan, R. Zheng, L. Huang, and K. Church, "Speech emotion recognition with multi-task learning," *Interspeech*, vol. 2021, pp. 4508–4512, 2021.

[29] R. Xia and Y. Liu, "A multi-task learning framework for emotion recognition using 2D continuous space," *IEEE Trans. Affect. Comput.*, vol. 8, no. 1, pp. 3–14, 2015. doi: 10.1109/TAFFC.2015.2512598.

[30] D. Nguyen *et al.*, "Joint deep cross-domain transfer learning for emotion recognition," 2020, *arXiv:2003.11136*.

[31] Y. Fan, X. Mi, and Y. Nie, "Cross-domain discriminative subspace classification algorithm for review text sentiment recognition oriented e-commerce platforms," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3455–3463, 2024. doi: 10.1109/TCE.2024.3372503.

[32] Q. Wang, Z. Wen, K. Ding, B. Liang, and R. Xu, "Cross-domain sentiment analysis via disentangled representation and prototypical learning," *IEEE Trans. Affect. Comput.*, pp. 1–13, 2024. doi: 10.1109/TAFFC.2024.3431946.

[33] S. Wang, J. Zhou, Q. Chen, Q. Zhang, T. Gui and X. Huang, "Domain generalization via causal adjustment for cross-domain sentiment analysis," 2024, *arXiv:2402.14536*.

[34] T. Xiao, X. Wang, A. A. Efros, and T. Darrell, "What should not be contrastive in contrastive learning," 2020, *arXiv:2008.05659*.

[35] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, and P. Isola, "What makes for good views for contrastive learning?" *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6827–6839, 2020.

[36] X. Wang and G. -J. Qi, "Contrastive learning with stronger augmentations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5549–5560, 2022. doi: 10.1109/TPAMI.2022.3203630.

[37] P. Khosla *et al.*, "Supervised contrastive learning," *Adv. Neural Inf. Process Syst.*, vol. 33, pp. 18661–18673, 2020.

[38] C. Azuma, T. Ito, and T. Shimobaba, "Adversarial domain adaptation using contrastive learning," *Eng. Appl. Artif. Intell.*, vol. 123, no. 6, 2023, Art. no. 106394. doi: 10.1016/j.engappai.2023.106394.

[39] H. Pashler, "Dual-task interference in simple tasks: Data and theory," *Psychol. Bull.*, vol. 116, no. 2, pp. 220.

[40] T Strobach, "Cognitive control and meta-control in dual-task coordination," *Psychon. Bull. Rev.*, vol. 31, no. 4, pp. 1445–1460, 2024. doi: 10.3758/s13423-023-02427-7.

[41] Y. Zhang, H. Zhang, L. -M. Zhan, X. -M. Wu, and A. Lam, "New intent discovery with pre-training and contrastive learning," 2022, *arXiv:2205.12914*.

[42] J. Devlin, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv:1810.04805*, 2018.

[43] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sens. J.*, vol. 19, no. 6, pp. 2266–2274, 2018. doi: 10.1109/JSEN.2018.2883497.

[44] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade and S. Ravi, "GoEmotions: A dataset of finegrained emotions," 2020, *arXiv:2005. 00547*.

[45] Y. Li, H. Su, X. Shen, W. Li, Z. Cao and S. Niu, "DailyDialog: A manually labelled multi-turn dialogue dataset," 2017, *arXiv:1710.03957*.

[46] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 4080–4090, 2017.

[47] R. Geng, B. Li, Y. Li, X. Zhu, P. Jian, and J. Sun, "Induction networks for few-shot text classification," 2019, *arXiv:1902.10482*.

[48] Y. Bao, M. Wu, S. Chang, and R. Barzilay, "Few-shot text classification with distributional signatures," 2019, *arXiv:1908.06039*.

[49] S. Lei, X. Zhang, J. He, F. Chen, and C. -T. Lu, "TART: Improved few-shot text classification using task-adaptive reference transformation," 2023, *arXiv:2306.02175*.

[50] X. Wang *et al.*, "Dual adversarial network with meta-learning for domain-generalized few-shot text classification," *Appl. Soft Comput.*, vol. 146, no. 2, 2023, Art. no. 110697. doi: 10.1016/j.asoc.2023.110697.

[51] J. Deng, L. Cheng, and Z. Wang, "Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification," *Comput. Speech Lang.*, vol. 68, no. 6, 2021. Art. no. 101182. doi: 10.1016/j.csl.2020.101182.