IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 March 2023, accepted 25 March 2023, date of publication 5 April 2023, date of current version 17 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264769

RESEARCH ARTICLE

Meta-Learning Based Tasks Similarity Representation for Cross Domain Lifelong Learning

MINGGE SHEN[1,2], DEHU CHEN[3,4], AND TENG REN[5]
[1]College of Intelligent Equipment and the Zhejiang College of Security Technology, Wenzhou, Zhejiang 325016, China
[2]Wenzhou Key Laboratory of Stereoscopic and Intelligent Monitoring and Warning of Natural Disasters, Wenzhou, Zhejiang 325016, China
[3]College of Architecture and Energy Engineering, Wenzhou University of Technology, Wenzhou 325035, China
[4]Wenzhou Key Laboratory of Intelligent Lifeline Protection and Emergency Technology for Resilient City, Wenzhou University of Technology, Wenzhou, Zhejiang 325035, China
[5]Dominican University, River Forest, IL 60305, USA

Corresponding author: Dehu Chen (chendehu01@163.com)

ABSTRACT Deep neural networks perform better in most specific single tasks than humans, but it is hard to handle a sequence of new tasks from different domains. The deep learning-based models always need to remember the parameters of the learned tasks to perform well in the new tasks and forfeit the ability to generalize from previous data, which is inconsistent with human learning. We propose a novel lifelong learning framework that can guide the model to learn new knowledge without forgetting the old knowledge through learning the similarity representation based on meta-learning. Specifically, we employ a cross-domain triplets network (CDTN) by minimizing the maximum mean discrepancy (MMD) between the current task and the knowledge base to learn the domain invariant similarity representation among tasks in different domains. Furthermore, we add a self-attention module to enhance the extraction of similarity features. Secondly, a soft attention network (SAN) which can assign different weights according to the learned similarity representation of tasks is proposed. In addition, a low-level feature enhancement module (LLEM) based on self-attention mechanisms is developed to capture domain-invariant similarity information. The experimental results show that our method effectively reduces catastrophic forgetting compared with the state-of-the-art methods when learning many tasks. Moreover, we show that the proposed method can hardly forget the old knowledge while continuously enhancing the performance of the old tasks, which is more in line with the human way of learning.

INDEX TERMS Lifelong learning, catastrophic forgetting, tasks similarity, cross domain triplets network.

I. INTRODUCTION

In the past few years, whether the traditional machine learning methods based on probability model and statistical model or the deep-learning methods such as deep convolution network (DCN) [1], transformers [2], and deep reinforcement learning (DRL) [3] is performed better than humans in some specific tasks, such as Go, image recognition and natural language process [4], [5], [6], etc. However, most existing deep models have a fatal flaw, which can not continuously learn tasks in sequence across domains like human beings. The deep models will always fall into catastrophic forgetting to perform better on new tasks because of the inherent optimization methods of the model. Because the data distribution of the new task and the old task is different, the optimal solution is different, so the weight of a trained model often changes when it is trained on the new task. This will inevitably fall into catastrophic forgetting. However, the way of human learning will not quickly forget the previous knowledge and instead accelerate the learning of new tasks based on what has been learned in previous tasks.

To prevent catastrophic forgetting, some approaches optimize parameter updates for new tasks in a space orthogonal to old tasks [8]. Others use rehearsal, adding a small number

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

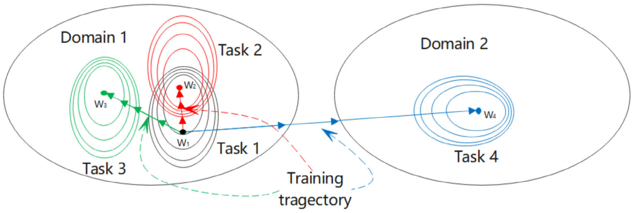


FIGURE 1. **The training process of lifelong learning (from task 1 to task 2, 3 and 4 respectively). Tasks 1, 2 and 3 come from the same domain, and task 4 comes from different domains. From the weights update tragectories, we can see that tasks similarity information, especially cross domain similarity information is very important for lifelong learning.**

of training samples from old tasks to new ones, mimicking human review behavior [9], [10]. Distillation [11] is a popular method for ensuring good performance across all tasks. Over-parameterization is leveraged in some methods [7], [12], [13] to activate or expand neurons for new tasks. Recently, Mahalanobis similarity was employed as a learning parameter to learn meaningful features while linearly increasing parameters as tasks increase [14]. However, most lifelong learning methods assume tasks are from the same distribution, ignoring the more general scenario of tasks from different domains.

Meta-learning [15], which is also referred to as ''learning to learn,'' involves training models to learn characteristics beyond the specifics of a task. For instance, models can learn to represent similarities between tasks, which allows for quick adaptation to new tasks. If two tasks are similar, their distance in the feature space is small, and vice versa. Despite these benefits, most of the current meta-learning approaches are limited by the fact that they are trained and tested on data from the same distribution. As a result, existing meta-learners are ineffective at learning essential similarity representations when tasks belong to different domains. This limitation is observed in the majority of existing meta-learning methods [16], [17], [18].

Fig. 1 shows the training process of lifelong learning (from task 1 to task 2, 3 and 4 respectively). Tasks 1, 2 and 3 come from the same domain, but task 4 comes from another domain. The black point is the weights W1 obtained by task 1. When the model continuously learns other tasks, the optimization trajectory of training weights is shown as the red, green and blue arrows in the Fig. 1. Due to the high similarity between task 1 and task 2, the weights quickly converge (short path) to W2, which may be better than original W1. Task 3 has lower similarity than task 2 with task 1, but it is still in the same domain. The updated weights W3 change greatly compared with W2, resulting in catastrophic forgetting to a certain extent. When learning task 4 is situated in another domain, the training trajectory shows the weights W4 change greatly compared with W3 to a great extent. It has led to catastrophic forgetting. Therefore, tasks similarity information, especially cross-domain similarity information, is very important for lifelong learning. The starting point of meta-learning used by us and other meta-learning methods like GPT-3 [2] is the same. Through pre-training, we can get a broader optimization space, which is convenient for other tasks to finetune. Reference [2] use a large network with about 17 billion parameters to get a better result. However, in this paper, the proposed cross-domain triplet network can cross-domain do meta-learning and learn the similarity between tasks, so there is no need to use complex network structure. To address these issues, we propose a lifelong learning method based on the similarity of tasks in the different domains and effectively reduce catastrophic forgetting. The main contributions of this paper are concluded as follows:

1) Our proposed cross-domain lifelong learning framework aims to prevent catastrophic forgetting, which sets it apart from previous lifelong learning algorithms. By utilizing meta-learning to understand task similarity and leveraging previously learned knowledge when acquiring new skills, which maximizes information retention.
2) A Triplets Network (CDTN) is designed to learn task similarity information, particularly when tasks belong to different domains. A low-level feature enhancement module (LLEM) based on self-attention mechanisms is developed to capture domain-invariant similarity information. Furthermore, based on the learned similarity information, we introduce a Soft Attention Network (SAN) that activates different neurons according to different tasks.
3) The experimental results show the performance of our proposed lifelong learning framework is greatly improved compared with the previous methods when there are a lot of tasks to learn in sequence. Moreover, the model can fully use the learned knowledge and improve the performance of previous tasks when learning new tasks.

This paper is organized as follows: Section II provides an overview of previous work on lifelong learning and meta-learning methods. Section III outlines the proposed method with detailed explanation. In Section IV, the experimental results are presented. Finally, Section V concludes the paper.

II. RELATED WORK

Lifelong learning is a method of training models in a way that prevents them from losing previously acquired knowledge as they learn new tasks. This is achieved by maintaining the consistency of the model's plasticity and stability, allowing it to continually absorb new information without forgetting old knowledge. Other related approaches of lifelong learning include continual learning [19], class incremental learning [20], task incremental learning [21].

Early approaches to lifelong learning were categorized into several main methods, such as minimizing representation overlap [22], [23], utilizing past samples or generated virtual samples, and implementing dual architectures [24], [25]. For instance, [26] used clustering techniques to assess the similarity and transfer invariance between tasks. Additionally, reinforcement learning methods and meta-learning technology have also shown promising results in lifelong learning tasks [29], [30], [31].

Due to limited resources, some early lifelong learning methods could only learn a small number of tasks sequentially using a specific shallow architecture. However, ELLA [32] was introduced as a general lifelong learning algorithm that operates within a multi-task learning framework, allowing the model to learn multiple basic learning models continuously. This led to the development of probability-based [33] and non-parametric Bayesian methods [34], which share information between tasks through linear combinations of basis vectors to enhance model performance. However, these methods have limitations in terms of the types of learning tasks they can handle. To overcome this limitation, GO-MTL [35] proposed a sparse shared model to address the multi-task learning problem.

However, these methods have limited applicability to simple tasks and are not suitable for handling a large number of complex tasks. With the recent growth in interest in deep neural networks, researchers pay more attention to avoiding catastrophic forgetting in lifelong learning and conducted empirical studies on the impact of dropout and activation functions on catastrophic forgetting, as well as exploring task incremental learning from a theoretical standpoint [36], [37], [38].

Deep lifelong learning methods can be broadly categorized into three categories. The first category is rehearsal-based methods, which are similar to human review. These methods take into account the impact of old tasks when the model learns new tasks, allowing it to better remember them and avoid catastrophic forgetting. Distillation technology is often used in rehearsal-based methods, which enables quick learning of new tasks using only a few samples. One example of a rehearsal-based method is the ICARL algorithm [9], which uses a teacher network and a student network to quickly converge all learned tasks with a small number of training samples. This approach allows for storing only a few samples from previous tasks when learning a new task, thereby reducing memory overhead.

A different approach to address the problem is the GEM method [41]. Instead of storing training samples, GEM stores the gradient of previous tasks, ensuring that the gradient update for new tasks is orthogonal to previous tasks. This reduces the interference of previous knowledge. Some GAN-based methods have also been proposed to generate high-quality images and model the data-generating distribution of previous tasks, allowing for retraining on generated examples [25], [42], [43], [44]. However, these methods require more calculations and additional resources, despite their ability to reduce storage space.

GAN-based methods provide storage space savings but require additional calculations, while other approaches like Continual Prototype Evolution (CPE) [39] combine the nearest-mean classifier approach with a more efficient reservoir-based sampling scheme. For more detailed experiments on rehearsal for lifelong learning, refer to [40].

One type of deep lifelong learning methods uses regularization to control parameter updates. These methods assign a weight to each parameter based on its importance for previous tasks and adjust it accordingly. LwF [8] limits the changes to parameters that are consistent with previous tasks. EWC [7] measures the significance of parameters using the Fisher information matrix from previous training. However, this method can restrict the network too much when there are many tasks and hinder new learning. Some methods like SI algorithm [45] solve this problem by considering the variation of parameters from previous to new tasks. However, this approach can lead to unstable results with random gradient descent. MAS [46] allows unsupervised estimation of parameter importance, making it suitable for specific data processing without supervision. VCL [47] applies a variational framework for continual learning. Other Bayesian-based methods [48], [45] estimate parameter importance online during task training. Aljundi et al. [46] proposed an unsupervised method for evaluating parameter importance, which can be adapted to different settings without supervision. This method was extended to the case of no task setting [49], [50]. However, these methods often have difficulty converging.

The third category of deep lifelong learning methods is neurons activation or expansion techniques. These methods use different parameters for different tasks or add extra parameters for new tasks if the network has spare parameters. However, this can quickly fill up the model parameters as the number of tasks grows. PackNet [12] ranks weights in the network based on their significance and only trains the current task with the first 50% of the selected weights. HAT [71] freezes the parameters of previous tasks or allocates a separate model for each task when learning new tasks. The network structure remains fixed, with specific components assigned to each task. During the training of a new task, the parameters of previous tasks are masked and converted into embeddings. After passing through these embeddings, the network transforms them into masks. HAT [71] uses sparsity as its loss function, making it more sophisticated. These methods usually need a task oracle to activate the relevant masks or task branches during prediction, thus limiting them to a multi-head setup and preventing them from handling a shared head between tasks. Expert Gate [51] solves this problem by learning an auto-encoder gate. In contrast to fixed network weight numbers, there are also methods such as progressive network [52], dynamic memory network [53], and DER [20] that increase the network structure. Whenever a new task is performed, suitable neurons are added to train it. However, these methods are restricted to small-scale task learning due to the constraints of parameter numbers.

Meta-learning, or learning to learn, is a machine learning method that aims to acquire higher-level data such as task-level and hyperparameter-level data. This higher-level data, known as meta-knowledge, helps the model to acquire new data more quickly and effectively. MAML [54] applies gradient descent to acquire the hyperparameters or initial parameters of the basic learner. [55] employs a read and write mechanism and merges it with soft attention and time convolution to access data from previous events. The siamese network [56], matching network [57], and prototype network [58] rely on the metric learning paradigm and use deep neural networks to acquire a mapping function from the input space to the feature space. The model acquires how to place examples that belong to the same category near each other and examples that belong to different categories far from each other, enabling it to classify new tasks effectively. Meta-learning has more robustness than traditional similarity measurement methods. Reference [59] apply meta-learning methods to acquire generalized parameters that are not specific to either old or new tasks to prevent catastrophic forgetting. Reference [68] introduced a differentiable Bayesian change point detection scheme to improve meta-learning methods for continuous learning tasks. With the help of the idea from MAML [54], [60] suggested an activation-gating function that selectively activates neurons, but these methods struggle to acquire cross-domain task similarity well.

To conclude, the existing methods for lifelong learning encounter major difficulties in terms of resource utilization and model performance when handling a large number of tasks. To address these difficulties, we suggest a meta-learning-based task similarity representation lifelong learning framework, which is a significant improvement over previous methods.

III. METHOD

The proposed new cross-domain lifelong learning framework, shown in Fig. 2, composed of two stages. In the first stage, a Cross Domain Triplets Network (CDTN) is trained to acquire a similarity representation of tasks across domains. In the second stage, a Soft Attention Network (SAN) uses this similarity data, combined with knowledge from a knowledge base, to obtain a task-specific attention map and assign specific weights to the network, enabling it to learn the current task effectively.

As illustrated in Algorithm 1, we start by initializing the two network parameters, W1 and W2, and setting the hyper-parameters. When a new task t arrives, we first train the CDTN by optimizing the weights W1 using meta-learning loss (MLL) and maximum mean discrepancy loss (MMDL) to obtain the similarity representation, b_l^t, as shown in lines 2 to 6. In lines 7 to 9, we adjust the number of channels to match the channels of the SAN by using 1 × 1 convolution based on the b_l^t and compute the attention map a using Eq. (8). To preserve prior knowledge, we calculate the gradients using Eq. (9), and update W2 with cross-entropy loss (CEL) for the classification task, all while keeping W1 fixed.

Algorithm 1 Proposed Cross Domain Lifelong Learning

```
Input: initial models weights: $W_1$,$W_2$ knowledge base samples and hyper-parameters $\sigma$, $\alpha$, $\beta$, $\lambda$, $\theta \ldots$, a set of training sets $\left\{\left\{\left(x_i^t, y_i^t\right)\right\}\right\}_{t=1}^{T}$, a knowledge base $\left\{\left(x_i', y_i'\right)\right\}$ learning rate $\delta$, number of epoches $E$.
Output: updated model weights: $W_1'$, $W_2'$
1:  Initialize network weights $W_1$ and $W_2$.
2:  for $t = 1, 2, 3, \ldots, T$ do
3:      Construct dataset $D_t$:$\left\{\left(x_i^t, y_i^t\right)\right\}_{t=1}^{T} \cup \left\{\left(x_i', y_i'\right)\right\}$
4:      for $e = 1, 2, 3, \ldots, E$ do
5:          Fix $W_2$ and Update $W_1$ with MLF loss. Eq. (5).
6:      end for
7:      for $e = 1, 2, 3, \ldots, E$ do
8:          Fix $W_1$ and adjust the number of channels of output $b_l^t$ use $1 \times 1$ convolution and compute attention maps $a$: $a_1, a_2, a_3, a_4$.
9:          Multiply attention maps $a$ by feature maps $F$ from SAN.
10:         Compute gradients use Eq. (9) and update $W_2$ use cross entropy loss (CEL) (for classification task).
11:     end for
12: end for
13: return $W_1'$, $W_2'$;
```

A. THE PROPOSED LIFELONG LEARNING PROBLEM SETTING

First, we have a set of labeled samples from various domains that serve as the knowledge base, which is different from the definition of knowledge graph and does not have a graph structure. These training samples can pre-train a base model so that they can better finetune in new tasks and better learn the similarity in different tasks, which agrees with the definition of meta-learning. The model faces a series of supervised learning tasks, denoted as Z1, Z2…Zt, where each task Zt = f(x,y), consists of data X(t) and Y(t), drawn randomly from different distributions D1, D2…Dt. X(t) represents a set of data samples for task t, while Y(t) represents the corresponding ground truth labels, typically either 1 or −1 for classification tasks or a real number for regression tasks. Each task t has N training samples. Unlike previous methods, we have a more flexible definition of lifelong learning tasks, where tasks don't have to come from the same distribution. T represents the total number of tasks encountered so far. At each time step, the model is given a batch of labeled data for a task t in T. After being trained on each batch. The model can predict on instances of any previous or current task without access to X(T) and Y(T) from the previous tasks.

B. CROSS DOMAIN TRIPLETS NETWORK FOR TASK SIMILARITY REPRESENTATION

In the lifelong learning scenario, where tasks come from different distributions, traditional meta-learning methods struggle to learn the similarity of tasks across domains.

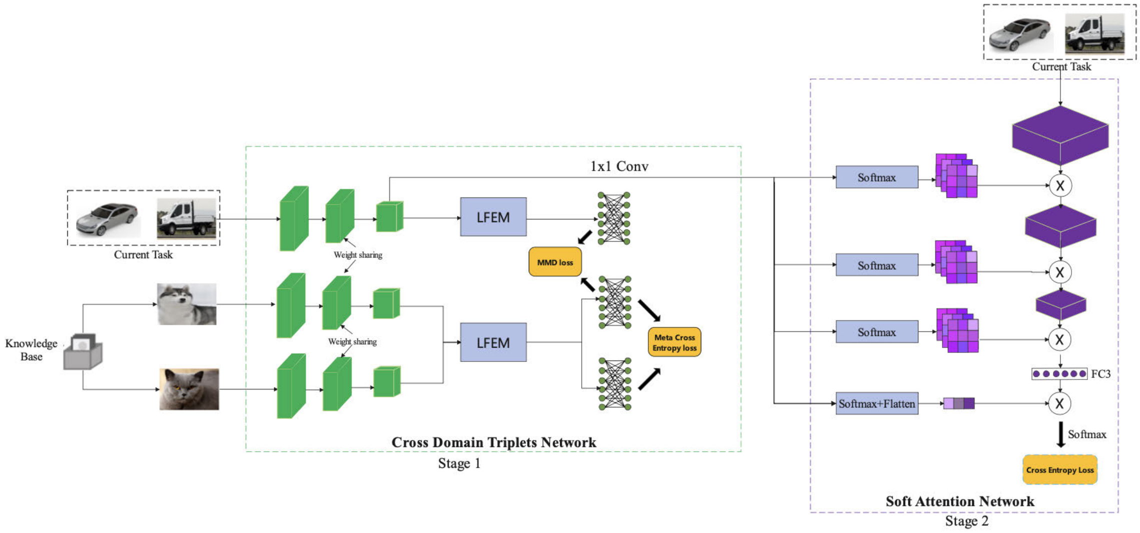


FIGURE 2. In the first stage of the proposed framework, a cross domain triplets network (CDTN) can learn the similarity representation of tasks not only in the same domain but also in the different domains. In the second stage, a soft attention network (SAN) is proposed to obtain the specific attention map of the task according to the similarity information of the tasks.

To overcome this issue, we introduce the Cross Domain Triplets Network, shown in the left half of Fig. 2. This network consists of a shared weight triplets network that is designed to learn the similarity representation of domain-invariant features. In conventional deep neural networks, the features become more specific from the last layer, causing a transferability gap that increases with regional differences. To tackle this problem, in the first step, we aim to reduce the inter-domain differences between samples and increase the differences within the same domain. This improves the diversity of the sample distribution, leading to better extraction of domain-invariant features. We measure the inter-domain distribution differences using maximum mean discrepancy (MMD) as described in Eq. (1) and (2).

$$L_{\mathrm{MMD}}\left(x_i, y_j\right) = \sup_{f \in \mathcal{F}} \left(\frac{1}{m} \sum_{i=1}^{m} f\left(x_i\right) - \frac{1}{n} \sum_{i=1}^{n} f\left(y_i\right) \right) \tag{1}$$

where x and y are the samples from knowledge base and current task respectively, f is the mapping function, here it refers to deep neural network, m and n are the number of samples of a batch from knowledge base and current task respectively.

$$L_{\mathrm{MMD}}\left(x_i, x_j\right) = \sup_{f \in \mathcal{F}} \left(\frac{1}{m} \sum_{i=1}^{m} f\left(x_i\right) - \frac{1}{m} \sum_{j=1}^{m} f\left(x_j\right) \right) \tag{2}$$

To ease calculation, we use the kernel embedding of distributions, the hidden layers related with the learning task in the convolutional neural networks (CNN) is mapped into the reproducing kernel Hilbert space (RKHS), and the distance between different domains is reduced by the multi-core optimization method. As shown in Eq. (3).

$$\begin{aligned} L_{\mathrm{MMD}}^{2}\left(x_i, y_j\right) &= \frac{1}{m(m-1)} \sum_{i \neq j}^{m} k\left(x_i, x_j\right) \\ &\quad + \frac{1}{m(m-1)} \sum_{i \neq j}^{m} k\left(y_i, y_j\right) - \frac{2}{m^2} \sum_{i,j=1}^{m} k\left(x_i, y_j\right) \end{aligned} \tag{3}$$

where k is Gaussian kernel function shown in Eq.(4).

$$k\left(x_i, y_j\right) = \exp\left(- \left\| x_i - y_j \right\|^2 / \left(2\sigma^2\right)\right) \tag{4}$$

The proposed CDTN has another purpose, which is to create similar representations. The inputs for the network are training samples from the knowledge base. The network creates these similarity representations by minimizing the meta-learning function (MLF), as demonstrated in Eq. (5). The distance between the network embeddings for two inputs will be minimized if they belong to the same class, and will be greater than a certain margin value "n" if they belong to different classes.

$$\begin{aligned} L_{\mathrm{MLF}}\left(x_i, x_j, x'\right) &= \frac{1}{2} x' \left\| f\left(x_i\right) - f\left(x_j\right) \right\|_2^2 \\ &\quad + \frac{1}{2}\left(1 - x'\right) \left\{ \max\left(0, n - \left\| f\left(x_i\right) - f\left(x_j\right) \right\|_2\right) \right\}^2 \end{aligned} \tag{5}$$

where x_i and x_j is image pairs and x' is equal 1 if the two images come from the same class. On the contrary, if the two image come from different classes the x is equal0,n is a margin parameter to balance the training loss.

Furthermore, to better obtain the cross domain similarity feature representation, we have introduced a low-level feature enhancement module (LFEM). This approach is based on the observation that shallow neural networks possess some degree of robustness to object features across domains.

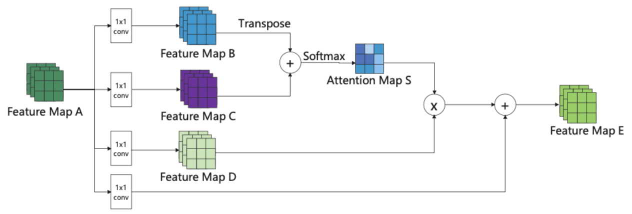


FIGURE 3. The detail of low-level feature enhancement module.

To extract domain-invariant information, we utilized a self-attention module, as depicted in Fig. 3. In this module, feature map A is first transformed into B, C, and D via three 1 × 1 convolution layers. Then, B and C are reshaped and multiplied to obtain the attention map S through the Softmax function. Finally, the feature map D is multiplied by S, and the resulting feature map is added to A to obtain the final feature map E, as shown in Eqs. (6) and (7).

$$s_{ji} = \frac{\exp\left(B_i \cdot C_j\right)}{\sum_{i=1}^{N} \exp\left(B_i \cdot C_j\right)} \tag{6}$$

$$E_j = \alpha \sum_{i=1}^{N} \left(s_{ji} D_i\right) + A_j \tag{7}$$

where sji measure the the influence degree of j position on i position. The homologous features will increase the corresponding, and enhance the ability of feature extraction. The value of each position of E is obtained by the weighted sum of the original features.

C. SOFT ATTENTION NETWORK

As depicted in the right half of Fig. 2, in the lifelong learning stage, to prevent catastrophic forgetting and retain previous task knowledge, we propose a Soft Attention Network (SAN) based on the soft attention mechanism to allocate specific parameters to new tasks. Moreover, we also factor in the similarity information between tasks, which can enhance the performance of related old tasks during new task learning. The attention mechanism is feature-wise, instead of channel-wise, allowing the model to learn as many tasks as possible without the need for additional hyperparameters to regulate network unit plasticity. The intermediate similarity feature from the trained Cross Domain Triplets Network is transformed into a channel matching feature map through four bottleneck layers, and the attention map is then generated via the Sigmoid function, as shown in Eq. (8). This attention map encompasses task-specific features and task similarity information. These attention maps are multiplied with the SAN to yield task-specific features, and the final classification result is obtained through a fully connected layer and Softmax.

We use cross-entropy loss and Stochastic Gradient Descent (SGD) to train the network. To ensure that information learned from previous tasks is preserved upon learning a new task, the gradients are conditioned based on the attention value. If the attention value of a feature is high, it indicates that it is beneficial to learning the task, and thus, its gradient update should be substantial. Conversely, if the attention value is low, it suggests that the feature is not valuable to the task, and its gradient update should be reduced, as demonstrated in Eq. (8).

$$g'_{l,t} = \delta_l^t \cdot \beta \left(g_{l,t} - \eta g_{l,t}\right) \tag{8}$$

where g is the gradients and δ is the learning rate. β is a hyperparameter control gradient update. A low β provide splasticity to the units and capacity of adaptation, but the network may easily forget what it learned. A high β prevents forgetting, but the network may have difficulties in adapting to new task.

D. LOSS FUNCTION

The loss function of the proposed framework is divided into two parts. In the CDTN, we minimize the MMD between the current task and knowledge base, which can learn the invariant feature of different domains. We minimize the MLF to learn the similarity information. The joint loss function is as Eq. (9):

$$L_{CDTR} = L^2_{\mathrm{MMD}}\left(x_i, y_i\right) - L^2_{\mathrm{MMD}}\left(x_i, x_j\right) + \theta L_{\mathrm{MLF}}\left(x_i, y_i\right) \tag{9}$$

where $L_{MMD}(x_i,y_i)$ and $L_{MMD}(x_i,x_j)$ are inter domain loss and intra domain loss respectively, L_{MLF} (x_i,y_j) is the meta cross entropy loss, θ is a hyperparameter to balance the two loss functions. In the SAN, our loss function is cross entropy loss for classification tasks.

IV. EXPERIMENT

We test the effectiveness of our lifelong learning approach with various experiments and analyze how each part contributes to the results and explain our method better.

A. DATASET AND TRAINING DETAILS

For our experiments, we have selected 8 image classification tasks [60], [61], [62], [63], [64], [65], [66], [67]. We randomly choose a portion of the data as the knowledge base. The datasets have class numbers ranging from 10 to 100, with 80% of the data used for training and 20% for testing. In our experiments, we use the ResNet-18 [69] architecture as our backbone and SGD as the optimization method. We train our network with the knowledge base in two stages with different learning rates. We reduce the learning rate if the validation loss does not improve and stop the training early if needed. We use 64 samples per batch and the same settings for all methods. We run our method on a computer with a CPU, a GPU, and 32 GB of RAM.

B. EVALUATION CRITERIA

To assess the generic performance of the model, we calculate the average accuracy (AA) on all the testing datasets from tasks t to T after training each task t. The AA is defined as in Eq. (10). A higher AA value indicates better performance of

TABLE 1. Results of the proposed method and other methods based on the average accuracy and average forgetting rate.

Methods	AA(%)	AF(%)
SI [45]	53.93	18.77
EWC [7]	62.43	10.51
LwF [8]	66.9	4.58
IMM [76]	66.79	6.08
GEM [41]	56.89	3.98
ICARL [9]	65.13	7.67
HAT [71]	63.57	0.15
PackNet [12]	66.68	0.1
AANet [78]	68.76	2.89
EWC [7]	67.12	0.5
DER [20]	68.69	0.09
Ours	70.21	-0.07

the model.

$$AA = \frac{1}{T}\sum_{t=1}^{T}(\frac{\mathrm{TP_t} + \mathrm{TN_t}}{\mathrm{P_t} + \mathrm{N_t}}) \times 100 \tag{10}$$

where TP and TN are the numbers of correctly classified samples. P_t and N_t are the number of positive and negative samples for task t. T is the total number of tasks.

We also use the average forgetting rate (AF) in Eq. (11) to measure how much the model forgets the old knowledge. The lower the AF, the better the model remembers the old knowledge. If the AF is negative, it means the model improves the old task when learning a new task.

$$AF = \frac{1}{T}\sum_{t=1}^{T}(A_\tau - A_t) \times 100 \tag{11}$$

where A_τ is the accuracy obtained when the task is the latest task, A_t is the accuracy of the previous tasks test on the current model obtained from the latest task. T is the number of tasks. Note that the forgetting rate of latest task is 0 and the smaller the average forgetting rate, the better the performance of the model.

C. PERFORMANCE EVALUATION

1) COMPARATIVE EXPERIMENTS

We compare the AA and AFR of different methods after 8 tasks in Table 1. Our method has the best AA of 70.21%. Other methods like DER [20] have good AA but more network parameters and computation. Methods like EWC [7] and IMM [72] have lower AA due to gradient update issues. Methods like GEM [41] and ICARL [9] have good performance but it drops as tasks increase and old samples decrease. GAN-based [73] methods can generate enough samples and have good results, but they need more resources.

"Our method has the lowest average forgetting rate AF (-0.07%), which means learning new tasks helps previous tasks perform better. This is because the tasks are similar and can fine-tune the parameters of earlier tasks. Methods based on weight regularization, such as EWC [7] and IMM [72], have trouble updating parameters with new tasks. As more tasks are learned, the gradient direction gets more restricted and forgetting increases. Methods based on rehearsal, like GEM [41] and ICARL [9], can lower forgetting, but they need more training samples for earlier tasks as the number of tasks grows. Otherwise, the average forgetting rate goes up. Some methods based on attention, such as PackNet and HAT [71], use fixed weight masks to avoid forgetting, but they ignore the similarities between tasks. So, learning new tasks does not improve previous tasks and they cannot handle many tasks."

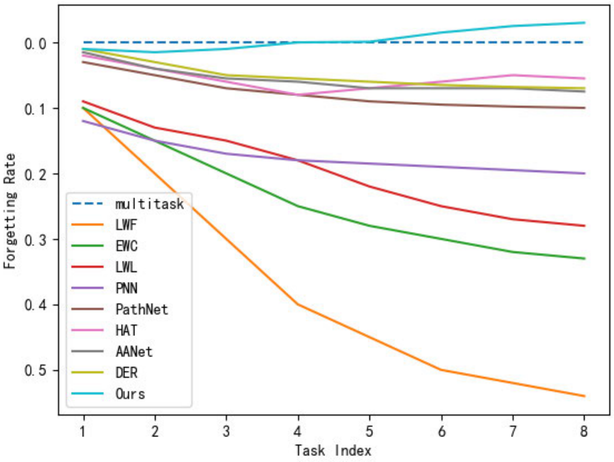


FIGURE 4. Average forgetting rate for the considered approaches when learning eight tasks.

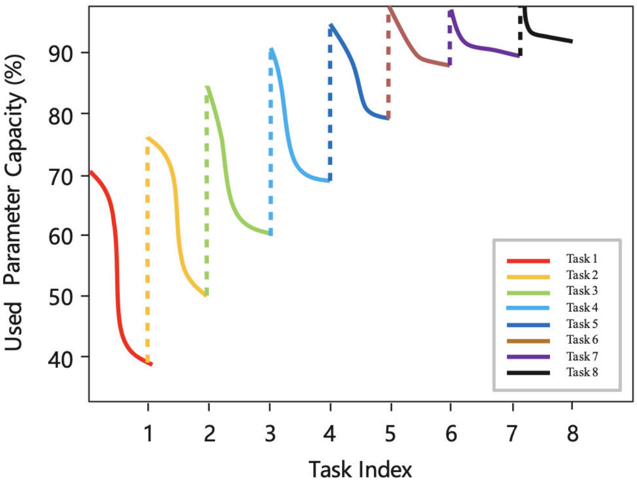


FIGURE 5. Network capacity usage with sequential task learning.

We tested how well our method can avoid catastrophic forgetting by learning eight tasks and measuring the forgetting rate against other methods (see Fig. 4). The multi-task learning method learns all tasks at once and never forgets. Our method improves previous parameters by using task similarity and performs better on new tasks that are related to old ones. This leads to a low forgetting rate. Other methods only try to reduce forgetting. PackNet [12] and HAT [71] have limited capacity and do worse on new tasks than our method. But they keep all the knowledge by locking task parameters with masks. EWC [7] and IMM [72] still forget over time

TABLE 2. The accuracy of the proposed model in a large number of tasks (10) (%).

Task Index	A(%)	A(%)	A(%)	A(%)	A(%)	A(%)	A(%)	A(%)	A(%)	A(%)
1	93.9									
2	94.0	96.1								
3	94.1	96.1	85.7							
4	94.1	96.1	85.7	89.4						
5	94.2	96.1	85.9	89.8	93.1					
6	94.2	96.1	85.9	89.8	93.1	92.5				
7	94.3	96.1	85.9	89.8	93.1	92.5	93.5			
8	94.3	96.1	85.9	89.8	93.1	92.5	93.5	89.3		
9	94.3	96.1	85.9	90.0	93.1	92.5	93.5	89.3	87.8	
10	94.3	96.1	85.9	90.0	93.1	92.5	93.5	89.3	87.8	90.3

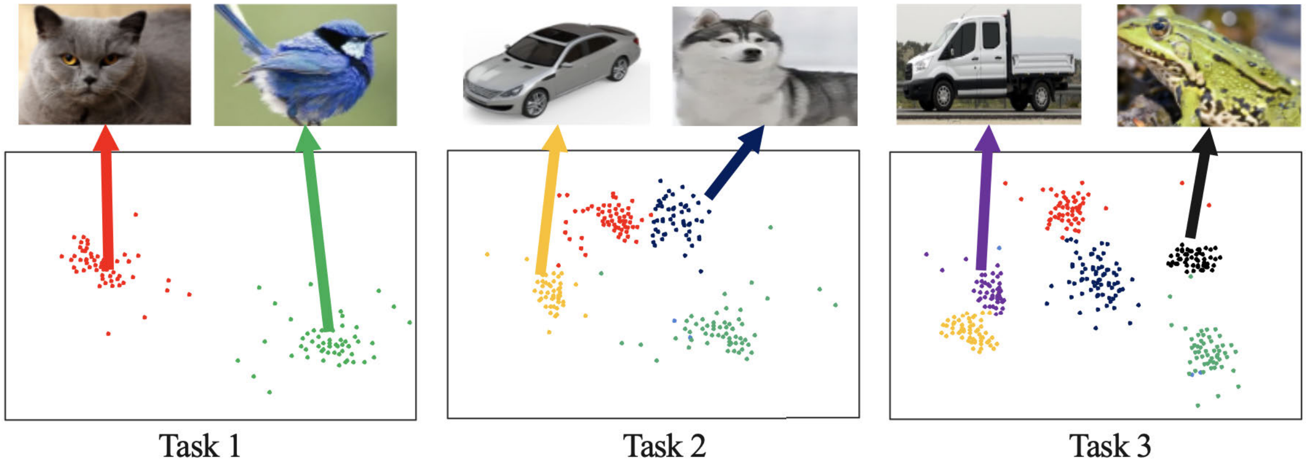


FIGURE 6. The visualization learning results of task similarity.

because they don not solve the forgetting problem completely. GEM [41] and ICARL [9] also forget little, but they need to store training samples for new tasks, which takes more space.

2) EFFECTS OF MODEL CAPACITY

Network capacity is important for lifelong learning. A model with high capacity can learn more tasks. Fig. 5 shows how the model uses its parameters. When a new task is learned, more weights are used. During training, the usage rate drops slowly at first, then faster until it stops. This means the network can be made smaller by 10% to 50%, depending on the task. When learning task 4, fewer new parameters are used because it is similar to task 2. The method uses the task similarity to improve learning. But when learning task 8, with no similar task before, the usage amount goes up by about 10% in top 5 tasks. The method uses less parameters than PackNet (25% to 80%) and HAT (15% to 70%) when learning similar tasks.

Table 2 shows how well the model can do multi-task classification. The accuracy stays the same even when learning 10 tasks in the CIFAR-100 dataset without forgetting. When more tasks are added, the old tasks get better. This is because the method uses the similarity between tasks and the sparsity from the loss function to learn many tasks continuously.

TABLE 3. Results of the ablation study for the proposed framework.

Module	1	2	3	5
CDTN	×	✓	×	✓
LLEF	×	×	✓	✓
AA(%)	63.57	65.87	67.61	70.21
AF(%)	0.15	0.04	0.12	-0.07

3) ABLATION STUDY

To evaluate the contribution of each module of the proposed method, an ablation study was conducted. In the study, one part of the method was omitted while the rest was kept intact. The average accuracy and average forgetting rate are shown in Table 3.

The results showed that the CDTN alone increased the average accuracy by about 4% and reduced the average forgetting rate by nearly 0.2%. This suggests that the task similarity information helps in learning new tasks. Moreover, the LLEF in the lifelong steps boosted the average accuracy by more than 2%, proving that the LLEF is very effective.

Fig. 6 illustrates the visualization results of similarity when learning three binary classification tasks. As more tasks are learned, CDTN can recognize the similarities among tasks, resulting in a small distance between similar tasks. This maximizes the utilization of parameters to ensure model accuracy.

V. CONCLUSION

We propose a new lifelong learning method that uses the task similarity to retain previous task information while learning new tasks. First, we use a Cross Domain Triplets Network (CDTN) to learn the similarity representation between tasks across different domains. Then, we use a Soft Attention Network (SAN) to assign different weights to different tasks in the deep network based on the similarity representation which effectively captures task similarity information. It significantly improves the lifelong learning performance compared to previous methods when dealing with a number of sequential tasks. In addition, our method avoids catastrophic forgetting and enhances the performance of old tasks when learning new tasks. In the future, we plan to add distillation technology to our framework to further improve our model.

REFERENCES

[1] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for Twitter sentiment analysis," *IEEE Access*, vol. 6, pp. 23253–23260, 2018.

[2] T. B. Brown, "Language models are few-shot learners," 2020, *arXiv:2005.14165*.

[3] M. Q. Mohammed, K. L. Chung, and C. S. Chyi, "Review of deep reinforcement learning-based object grasping: Techniques, open challenges, and recommendations," *IEEE Access*, vol. 8, pp. 178450–178481, 2020.

[4] N. D. Nguyen, T. Nguyen, and S. Nahavandi, "System design perspective for human-level agents using deep reinforcement learning: A survey," *IEEE Access*, vol. 5, pp. 27091–27102, 2017.

[5] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, "Deep learning in skin disease image recognition: A review," *IEEE Access*, vol. 8, pp. 208264–208280, 2020.

[6] Y. Tian, "Artificial intelligence image recognition method based on convolutional neural network algorithm," *IEEE Access*, vol. 8, pp. 125731–125744, 2020.

[7] M. Omar, S. Choi, D. Nyang, and D. Mohaisen, "Robust natural language processing: Recent advances, challenges, and future directions," *IEEE Access*, vol. 10, pp. 86038–86056, 2022.

[8] K. James, "Overcoming catastrophic forgetting in neural networks," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 13, pp. 3521–3526, Mar. 2017.

[9] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[10] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "ICaRL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2001–2010.

[11] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, and G. Wayne, "Experience replay for continual learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 350–360.

[12] K. Lee, K. Lee, J. Shin, and H. Lee, "Overcoming catastrophic forgetting with unlabeled data in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 312–321.

[13] A. Mallya and S. Lazebnik, "PackNet: Adding multiple tasks to a single network by iterative pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7765–7773.

[14] C. Fernando, D. Banarse, C. Blundell, Y. Zwols, D. Ha, A. A. Rusu, A. Pritzel, and D. Wierstra, "PathNet: Evolution channels gradient descent in super neural networks," 2017, *arXiv:1701.08734*.

[15] C. Simon, M. Faraki, Y.-H. Tsai, X. Yu, S. Schulter, Y. Suh, M. Harandi, and M. Chandraker, "On generalizing beyond domains in cross-domain continual learning," 2022, *arXiv:2203.03970*.

[16] S. Thrun and L. Pratt, "Learning to learn: Introduction and overview," in *Learning to Learn*. Cham, Switzerland: Springer, 1998, pp. 3–17.

[17] Y. Du, J. Xu, H. Xiong, Q. Qiu, X. Zhen, C. G. Snoek, and L. Shao, "Learning to learn with variational information bottleneck for domain generalization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 200–216.

[18] Y. Guo, N. C. Codella, L. Karlinsky, J. V. Codella, J. R. Smith, K. Saenko, T. Rosing, and R. Feris, "A broader study of cross- domain few-shot learning," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 124–141.

[19] H.-Y. Tseng, H.-Y. Lee, J.-B. Huang, and M.-H. Yang, "Cross-domain few-shot classification via learned feature-wise transformation," 2020, *arXiv:2001.08735*.

[20] M. Delange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "A continual learning survey: Defying forgetting in classification tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3366–3385, Jul. 2022.

[21] Y.-C. Hsu, Y.-C. Liu, A. Ramasamy, and Z. Kira, "Re-evaluating continual learning scenarios: A categorization and case for strong baselines," 2018, *arXiv:1810.12488*.

[22] S. Yan, J. Xie, and X. He, "DER: Dynamically expandable representation for class incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3014–3023.

[23] R. M. French, "Semi-distributed representations and catastrophic forgetting in connectionist networks," *Connection Sci.*, vol. 4, nos. 3–4, pp. 365–377, Jan. 1992.

[24] R. M. French, "Dynamically constraining connectionist networks to produce distributed, orthogonal representations to reduce catastrophic interference," in *Proc. 16th Annu. Conf. Cogn. Sci. Soc.* Evanston, IL, USA: Routledge, 2019, pp. 335–340.

[25] D. L. Silver and R. E. Mercer, "The task rehearsal method of lifelong learning: Overcoming impoverished data," in *Proc. Conf. Can. Soc. Comput. Studies Intell.* Cham, Switzerland: Springer, 2002, pp. 90–101.

[26] A. Robins, "Catastrophic forgetting, rehearsal and pseudo rehearsal," *Connection Sci.*, vol. 7, no. 2, pp. 123–146, 1995.

[27] S. Thrun and J. O'Sullivan, "Discovering structure in multiple learning tasks: The TC algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, vol. 96, 1996, pp. 489–497.

[28] R. S. Sutton, A. Koop, and D. Silver, "On the role of tracking in stationary environments," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, pp. 871–878.

[29] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka, and T. M. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 1306–1313.

[30] B. Ans and S. Rousset, "Avoiding catastrophic forgetting by coupling two reverberating neural networks," *Comp. Rendus de l'Académie des Sci. Ser. III Sci. de la Vie*, vol. 320, no. 12, pp. 989–997, Dec. 1997.

[31] J. Rueckl, "JumpNet: A multiple-memory connectionist architecture," in *Proc. 15 th Annu. Conf. Cognit. Sci. Soc.*, vol. 24, 1993, pp. 866–871.

[32] R. M. French, "Pseudo-recurrent connectionist networks: An approach to the 'sensitivity-stability' dilemma," *Connection Sci.*, vol. 9, no. 4, pp. 353–380, Dec. 1997.

[33] P. Ruvolo and E. Eaton, "ELLA: An efficient lifelong learning algorithm," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 507–515.

[34] J. Zhang, Z. Ghahramani, and Y. Yang, "Flexible latent variable models for multi-task learning," *Mach. Learn.*, vol. 73, no. 3, pp. 221–242, Dec. 2008.

[35] P. Rai and H. Daume III, "Infinite predictor subspace models for multitask learning," in *Proc. 13th Int. Conf. Artif. Intell. Statist. Workshop Conf.*, 2010, pp. 613–620.

[36] A. Kumar and H. Daume III, "Learning task grouping and overlap in multitask learning," 2012, *arXiv:1206.6417*.

[37] A. Pentina and C. Lampert, "A PAC-Bayesian bound for lifelong learning," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 991–999.

[38] P. Alquier, "Regret bounds for lifelong learning," in *Proc. Artif. Intell. Statist.*, 2017, pp. 261–269.

[39] A. Pentina and C. H. Lampert, "Lifelong learning with non-i.i.d. tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1540–1548.

[40] M. D. Lange and T. Tuytelaars, "Continual prototype evolution: Learning online from non-stationary data streams," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8250–8259.

[41] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, "Class-incremental learning: Survey and performance evaluation on image classification," 2020, *arXiv:2010.15277*.

[42] D. Lopez-Paz and M. Ranzato, "Gradient episodic memory for continual learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6467–6476.

[43] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[44] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay," 2017, *arXiv:1705.08690*.

[45] F. Ye and A. G. Bors, "Lifelong infinite mixture model based on knowledge-driven Dirichlet process," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, p. 10695.

[46] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3987–3995.

[47] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, "Memory aware synapses: Learning what (not) to forget," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 139–154.

[48] T. N. Nguyen, T. D. Ngo, and H. Nguyen-Xuan, "A novel three-variable shear deformation plate formulation: Theory and isogeometric implementation," *Comput. Methods Appl. Mech. Eng.*, vol. 326, pp. 376–401, Nov. 2017.

[49] H. Ahn, S. Cha, D. Lee, and T. Moon, "Uncertainty-based continual learning with adaptive regularization," 2019, *arXiv:1905.11614*.

[50] M. D. Lange, X. Jia, S. Parisot, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "Unsupervised model personalization while preserving privacy and scalability: An open problem," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, p. 14463.

[51] R. Aljundi, K. Kelchtermans, and T. Tuytelaars, "Task-free continual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11254.

[52] R. Aljundi, P. Chakravarty, and T. Tuytelaars, "Expert gate: Lifelong learning with a network of experts," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3366–3375.

[53] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive neural networks," 2016, *arXiv:1606.04671*.

[54] M. Perkonigg, J. Hofmanninger, C. J. Herold, J. A. Brink, O. Pianykh, H. Prosch, and G. Langs, "Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging," *Nature Commun.*, vol. 12, no. 1, pp. 1–12, Sep. 2021.

[55] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1842–1850.

[56] X. Dong and J. Shen, "Triplet loss in Siamese network for object tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 459–474.

[57] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5410–5418.

[58] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017, *arXiv:1703.05175*.

[59] J. Rajasegaran, S. Khan, M. Hayat, F. S. Khan, and M. Shah, "ITAML: An incremental task-agnostic meta-learning approach," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, p. 13588.

[60] J. Harrison, A. Sharma, C. Finn, and M. Pavone, "Continuous meta-learning without tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17571–17581.

[61] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.

[62] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 343–347.

[63] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.

[64] V.Bui and L.-C. Chang, "Deep learning architectures for hard character classification," in *Proc. Int. Conf. Artif. Intell. (ICAI). Steering Committee World Congr. Comput. Sci., Comput.*, 2016, p. 108.

[65] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[66] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," Tech. Rep., 2011.

[67] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: A multi-class classification competition," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2011, pp. 1453–1460.

[68] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25. Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

[69] J. Harrison, A. Sharma, C. Finn, and M. Pavone, "Continuous metalearning without tasks," in *Proc. Conf. Neural Inf. Process. Syst.*, 2020, pp. 1–11.

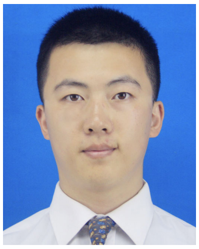
[70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[71] B. Pfulb and A. Gepperth, "Overcoming catastrophic forgetting with Gaussian mixture replay," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 4548–4557.

[72] J. Lee, D. Joo, H. G. Hong, and J. Kim, "Residual continual learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, pp. 4553–4560.

[73] C. Atkinson, B. McCane, L. Szymanski, and A. Robins, "Pseudo-recursal: Solving the catastrophic forgetting problem in deep neural networks," vol. 2, 2018, *arXiv:1802.03875*.

MINGGE SHEN received the master's degree in surveying and mapping from Tongji University, Shanghai, China, in 2015. She is currently a Lecturer with the College of Intelligent Equipment, and the Zhejiang College of Security Technology, Wenzhou, China. Her research interests include urban environment remote sensing, drone mapping, and surveying and target detection in remote-sensing images.



DEHU CHEN received the master's degree in surveying and mapping from Tongji University, Shanghai, China, in 2014. He is currently a Lecturer with the College of Architecture and Energy Engineering, Wenzhou University of Technology, Wenzhou, China. His research interests include engineering measurement, deformation monitoring of urban rail transit, and high-rise buildings.

TENG REN received the Ph.D. degree in information studies from Dominican University, River Forest, IL, USA, in 2021. He is currently the CEO of Qianxing Intelligence (Zhuhai) Science and Technology Ltd. Company. His research interests include AI, engineering operations, information technology, information management, and high-rise building technology.

• • •