# DOMAIN-GENERALIZED FEW-SHOT CLASSIFICATION VIA CROSS-DOMAIN EPISODIC META-LEARNING

[1] *Yuan-Chia Cheng* (鄭元嘉), [2] *Yu-Chiang Frank Wang* (王鈺強), [3] *Chiou-Shann Fuh* (傅楸善)

[1 2] Graduate Institute of Communication Engineering,
National Taiwan University, Taiwan,
[3] Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan
E-mail: [1] r08942154@ntu.edu.tw, [2] ycwang@ntu.edu.tw , [3] fuh@csie.ntu.edu.tw

## ABSTRACT

Few-shot classification aims to carry out classification given only few labeled examples for the categories of interest. While several approaches have been proposed to address the few-shot learning tasks (FSL), most assume that training and test data (i.e., base and novel classes) are drawn from the same domain. When it comes to recognizing novel-class data in a different and unseen domain, this is considered as a domain generalized few-shot classification task. In this paper, we present a unique learning framework for domain-generalized few-shot classification, where base classes from homogeneous multiple source domains are available during training, while novel classes in an unseen target domain need to be recognized. By advancing meta-learning strategies, our learning framework exploits data across multiple source domains to capture domain-invariant features, while FSL ability is preserved by observing the sampled support and query data. We conduct extensive experiments to verify the effectiveness of our proposed learning framework. Moreover, from the experimental results, we are able to provide insights into data and backbones of models for domain-generalized few-shot classification.

***Keywords:*** *Few-shot learning, Domain generalization, Meta-learning.*

## 1. INTRODUCTION

Recent development of deep learning technologies brings significant improvements on a variety of real-world applications. However, on the other hand, its high expenditure of collecting and labeling data would limit its feasibility in real-world scenarios. Few-shot learning is among one of the challenges, in which the amount of data of particular classes (i.e., novel classes) is limited during training. Broadening the applicability, few-shot classification has been among active research topics in computer vision and various approaches [9, 10, 26, 14, 1, 41, 39, 31, 33] have been subsequently proposed. In particular, metric-based meta-learning methods [39, 31, 33] draw most attention by its simplicity and generalization ability to recognize novel classes.

Though few-shot classification has been well studied, most existing FSL methods assume that base and novel classes are both presented in the same data domain. In the practical setting of domain generalization, the learning model is trained using data observed from source domains, while it is applied to unseen target domains for recognition tasks. In this challenging scenario, the difference of data distributions between source and target domains exist and cannot be easily modeled beforehand as domain adaptation works did (e.g., [11, 37, 4]). In other words, how to model and handle the unknown domain shift poses a frustrating obstacle if recognizing cross-domain novel-class data is desirable. As indicated in [13, 5], most previous methods related to few-shot classification would be prone to a significant performance drop under the domain generalization setting. To tackle this unseen domain shift issue, a number of works [22, 24, 32] are proposed to apply on different problems such as classification and person re-identification. However, few efforts are targeted at the cross-domain or domain-generalized few-shot classification.

In this paper, we tackle the task of domain-generalized few-shot classification, which recognizes data of novel classes in unseen domains. As illustrated in Figure 1, the learning of our proposed framework requires training data from multiple homogeneous source domains. Note that homogeneous source-domain data indicate training data of base classes sharing the same label set. Based on metric-based meta-learning, the proposed learning framework aims to derive domain-invariant features by recognizing support and query data from distinct source domains, with such a learning mechanism can be generalized to data in unseen domains.

Moreover, inspired by [5, 34], we perform extensive quantitative analyses and provide insights into few-shot classification under domain generalization settings. We assess the robustness of different backbone models (e.g., ProtoNet [31] and RelationNet [33]), and evaluate their generalization to image data with distinct domains and various varieties (e.g., PACS [20], Omniglot [18], Mini-ImageNet [29] and CUB [40]) for domain-generalized few-shot learning. Surprisingly, deeper backbones do not provide generalization guarantees, while fine-grained novel image categories do not benefit from the cross-domain classification settings.

The contributions of our work are highlighted below:

1. We address the task of domain-generalized few-shot classification, which learns domain-invariant features with domain generalization guarantees to recognize image data of novel categories in unseen domains.
2. We advance meta-learning and propose a unique learning scheme, which tackles the matching of support and query image data across distinct domains, with flexibility of utilizing state-of-the-art metric-learning models as the backbones.
3. From extensive experiments on multiple image datasets, we are able to conclude that deeper backbones would not be desirable for recognizing data of unseen categories/domains, while novel yet fine-grained categories would be more challenging to handle. This is very different from the experiences observed for image classification in a single domain.

## 2. RELATED WORK

### 2.1. Few-Shot Classification

By observing a sufficient amount of training examples of seen classes (i.e., base classes), few-shot classification [19] aims to classify unlabeled examples of unseen classes (i.e., novel classes), given only few labeled examples for each novel class. To address this topic, a large quantity of approaches have been proposed, including initialization based [9, 10, 26], hallucination based [14, 1, 41] and metric-learning based [16, 39, 31, 33, 30, 12, 28] methods. Initialization based methods focus on learning proper initialization of model parameters, so that they could be fine-tuned to novel-class data through a small number of optimization steps. Hallucination based methods introduce a learnable generator to generate more image data for novel classes, which alleviates the few-shot learning setting. Metric-learning based methods, with the goal of "learning to compare", apply a deterministic or learnable distance metric to evaluate the similarity between images in the derived feature space.

Among the above approaches, metric-based meta-learning methods have drawn extensive attention due to their simplicity. In general, metric-based methods comprise two main components, a feature encoder and a metric-learning module. Most state of the art approaches adopt CNN as the feature encoder and focus on designing proper metric functions. For instance, MatchingNet [39] employs recurrent-based model to further encode the feature in pursuit of full context embeddings, and computes image similarity by cosine similarity measures. ProtoNet [31] adopts Euclidean distance, RelationNet [33] applies CNN modules and GNN [30] uses graph convolution modules as the metric functions. It is exploited by [27] that metric scaling further improves the recognition performances.

Though many efforts and progresses have been made, most existing works did not address the obstacle in few-shot classification due to the domain shift between base and novel class data [5]. Tseng et al. [36] thereby turn their attention to cross-domain scenarios and propose learnable feature-wise transformation layers inserted into the model in training stage which enhance the generalization ability to unseen domains.

### 2.2. Domain Adaptation and Generalization

How to alleviate the domain shift problems has been an active topic in the learning communities. Recent domain adaptation methods [11, 37, 6] have been proposed, which are built upon adversarial training [11] for aligning the data distributions between source and target domains. And, some works [35, 7, 4] choose to align such distributions at pixel levels. Nevertheless, domain adaptation requires the presence of target-domain data during training, which would be intractable in some real-world scenarios (i.e., road scene segmentation across different cities).

Domain generalization, on the other hand, tackles the task of generalizing the learning from multiple source domains to unseen target domains. A plethora of deep learning approaches [3, 25, 20, 23] have been proposed to introduce such generalization abilities. Generally, existing domain generalization approaches are built upon the three core ideas. The first [42] is to decompose the model into domain-specific components and estimate the relevance to make use of the corresponding components to make predictions. The second [25, 23] tends to learn a domain-invariant feature space, while the third [20] can be deemed as a mix-up which decomposes the model into domain-specific and domain-invariant components and utilizes both to make predictions.

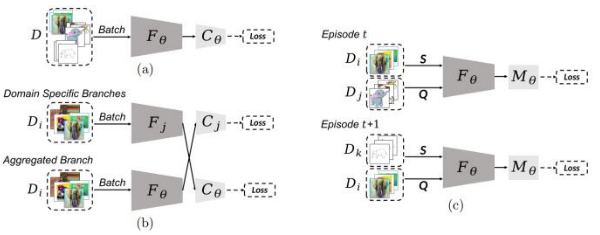


Figure 2: Standard classification, domain-generalized classification, and domain-generalized few-shot classification. (a) Naive baseline: do not distinguish between different source domains, and simply train feature encoder $F_\theta$ and classifier $C_\theta$ using data from source domains. (b) Episodic-learning for domain generalization [22]: with feature encoder $F_j$ and classifier $C_j$ pre-trained on each source domain $D_j$, train $F_\theta$ and $C_\theta$ via the associate paths to recognize test data in unseen domains (sharing the same label set). (c) Cross-domain episodic meta-learning (x-EML): $F_\theta$ and $M_\theta$ (metric learner) are trained in N-way K-shot fashions, with support S and query Q sampled from distinct domains. $F_\theta$ and $M_\theta$ can thus be applied to recognize novel classes in unseen domains.

Recent research efforts on domain generalization applies meta-learning [9] for deriving the associated learning models or features [2, 21, 22, 24]. MLDG [21] carries out modification on MAML [9] to adapt the learning scheme to domain generalization settings. [24] introduces a learnable auxiliary loss into the meta-learning scheme and expects it to guide learning in the meta-training stage. [22] proposes an episodic training procedure where domain-specific feature extractors and classifiers are crossly trained to simulate interacting with a badly tuned partner and in turn brings about robustness.

As of now, most works on domain generalization target themselves at standard classification problems, and did not consider the few-shot learning settings. Although [22, 24] address heterogeneous domain generalization, their settings still require a sufficient amount of data for novel classes in the target domains. In this work, we deal with a more challenging but practical scenario of domain-generalized few-shot classification. Inspired by [25, 23], our model aims at deriving domain-invariant features in a meta-learning fashion, with generalization guarantees to recognition of unseen target-domain data.

## 3. PROPOSED METHOD

### 3.1. Notations and Problem Formulation

For the sake of completeness, we first define the notations used in this paper. To tackle domain-generalized few-shot classification, we follow [43] and consider base classes from multiple homogeneous source domains D = $\{D_1, D_2, ..., D_{N_s}\}$ ($N_s$ is the

**Algorithm 1:** Cross-Domain Episodic Meta-Learning (x-EML)

**Data:** Training data in homogeneous source domains $D$,
Test data in target domain $D^*$ under few-shot settings (support set $\mathbf{S}^*$ and query set $\mathbf{Q}^*$)
Define loss function $L$ and optimizer $A$;
Initialize feature encoder $F_\theta$ and metric learner $M_\theta$ in terms of the parameters $\theta$;
**while** *meta-training* **do**
  Randomly sample two distinct domains $D_i$, $D_j$ from $D$;
  Randomly sample a shared label set $Y$ of $N$ classes;
  Construct pseudo support set $\mathbf{S}$ by sampling $K$ shot per class in $Y$ from $D_i$;
  Construct pseudo query set $\mathbf{Q}$ by sampling $m$ images per class in $Y$ from $D_j$;
  Update $\theta$ by $\nabla_\theta \sum_{(x_q, y_q) \in \mathbf{Q}} L(y_q, M_\theta(Y_s, F_\theta(\mathbf{X}_s), F_\theta(x_q)))$ with $A$;
**end**
**while** *meta-testing* **do**
  Given $(\mathbf{X}^*_s, \mathbf{Y}^*_s) \in \mathbf{S}^*$ and $\mathbf{X}^*_q \in \mathbf{Q}^*$;
  **for** $x^*_q \in \mathbf{X}^*_\mathbf{q}$ **do**
    Predict the label of $x^*_q$ by $M_\theta(\mathbf{Y}^*_s, F_\theta(\mathbf{X}^*_s), F_\theta(x^*_q))$;
  **end**
**end**

Algorithm 1: Pseudo code for proposed training framework x-EML.

number of source domains observed), which share the same label space. The novel classes are in a target domain $D^*$, which is not available during training. In our meta-learning scheme, D is utilized during meta-training, and only the meta-testing stage has access to $D^*$. For each training episode in the meta-training stage, we consider a standard N-way K-shot classification by constructing a support set S = $\{(X_s, Y_s)\}$ and a query set Q = $\{(X_q, Y_q)\}$, while S and Q are sampled from distinct domains of D. As for meta-testing, we only have the access to support and query set data from $D^*$. That is, our model observes the support set $S^* = \{(X^*_s, Y^*_s)\}$ (also in a N-way K-shot fashion), and it is required to predict the labels for the query set $Q^* = \{X^*_q\}$.

### 3.2. Cross-Domain Episodic Meta-Learning

Metric-learning based meta-learning has been widely applied for solving few-shot classification tasks [39, 31, 33]. However, as pointed out in Sect. 2, they generally assume that both base and novel classes are from the domain, and thus lack the ability to generalize across data domains. While one can naively include the data of base categories from multiple multiple source domains D for training, as shown in Figure 2(a), there is no guarantee that the learned model would be able to handle unseen target domain data. As depicted in Figure 2(b), an episodic learning scheme was recently proposed by [22], which utilizes local feature extractors and classifiers pre-trained in each data domain, guiding the learned global ones to generalize to unseen data domains. However, they require the training and test data (from seen and unseen domains) share exactly the same label space. Therefore, while promising domain generalization ability was observed, such episodic learning schemes would still not be able to recognize novel class in unseen domains.

To address domain-generalized few-shot classification, we propose Cross-Domain Episodic Meta-Learning (x-EML) for deriving domain-invariant yet semantics-discriminative representations, which can be applied to tackle novel class data in unseen target domains. As

Table 1: Performance comparisons on Omniglot, mini-ImageNet, and CUB using ProtoNet as the backbone models. Note that Single Domain-S indicates the training of ProtoNet with support and query sets always from the same domain S, and x-EML* denotes the training of our model using support and query sets sampled from the same domain in each episode. The averaged accuracy is reported along with the 95% confidence interval.

| | Omniglot | Mini-ImageNet | CUB |
|---|---|---|---|
| Single Domain-S | 86.35 | 42.34 | 38.27 |
| Single Domain-C | 86.67 | 45.99 | 41.61 |
| Single Domain-A | 87.02 | 48.77 | 42.77 |
| LFT [36] | 85.43 | 43.04 | 38.65 |
| x-EML* | 88.55 | 48.38 | 42.90 |
| x-EML (Ours) | **89.62** | **49.00** | **44.26** |

Table 2: Performance comparisons on Omniglot, mini-ImageNet, and CUB using RelationNet as the backbone models. Note that Single Domain-S indicates the training of RelationNet with support and query sets always from the same domain S, and x-EML* denotes the training of our model using support and query sets sampled from the same domain in each episode. The averaged accuracy is reported along with the 95% confidence interval.

| | Omniglot | Mini-ImageNet | CUB |
|---|---|---|---|
| Single Domain-S | 70.02 | 29.90 | 31.76 |
| Single Domain-C | 73.36 | 33.00 | 32.66 |
| Single Domain-A | 61.02 | 34.91 | 31.90 |
| LFT [36] | **79.17** | 35.95 | 35.68 |
| x-EML* | 74.56 | 36.54 | **36.27** |
| x-EML (Ours) | 77.89 | **39.15** | 34.94 |

depicted in Figure 2(c), the meta-training stage of our proposed framework follows the N-way K-shot scenario. To construct S and Q for each episode, we first randomly select N classes from a source domain $D_i$ from D, with K examples sampled for each class. On the other hand, to form Q, we sample m examples from a different source domain $D_j$ for each of the N classes. Thus, for each episode in meta-training, we have a total of N × K examples in S along with N × m examples in Q. Such support and query set samples are fed into the encoder $F_\theta$ for feature extraction, followed by a proper metric-learning module $M_\theta$ to predict class labels for each query example. We note that, for the metric-learning module $M_\theta$, one can apply any existing metric-based algorithms or networks, comparing the similarity between $F_\theta(X_s)$ and $F_\theta(x_q)$ and assign labels (out of N) to $F_\theta(x_q)$. Moreover, we calculate the loss for the episode by loss function L and update the parameters $\theta$ via conventional back-propagation and gradient descent strategies. Overall, the objective function for training the x-EML framework in each episode can be formulated as follows:

$$\underset{\theta}{\operatorname{argmin}} \sum_{(x_q,\, y_q)\in Q} L(y_q, M_\theta(Y_s, F_\theta(X_s), F_\theta(x_q)))$$
$$S \in D_i,\, Q \in D_j,\, i \neq j$$

It is worth noting that, the key of our proposed x-EML framework lies in that S and Q sampled from different source domains but share the same label space. In other words, the N-way K-shot meta-training scheme is realized on data across distinct domains, which thus results in learning semantics-discriminative yet domain-generalized features. This is the reason why our proposed learning framework, while applicable to any existing meta-learning modules, can address the challenging task of domain-generalized few-shot classification. The pseudo code for meta-training and meta-testing stages of our x-EML is summarized in Algorithm 1. Later in the experiments, we will further elaborate the derivation of domain-invariant features using different modules as $F_\theta$ and $M_\theta$ . We will explain why learning with deeper backbones does not necessarily provide better domain generalization guarantees, while fine-grained novel image categories would be more difficult to recognize under the domain generalization setting.

## 4. EXPERIMENTAL RESULTS

### 4.1. Settings

To evaluate the performance of the proposed model on domain-generalized few-shot classification, we re-purpose the benchmark PACS dataset [20] as the source domain training data, which consists of four homogeneous data domains of Photo, Art-Painting, Cartoon and Sketch, sharing identical 7 categories with 9991 images in total. The use of PACS for training allows us to ensure the homogeneity of multiple source domains, with significant domain shifts preserved for domain-generalized classification. Such source-domain homogeneity is not considered in [36], as it takes Mini-ImageNet [29] as the major source domain data, with four datasets CUB, Cars [17], Places [44] and Plantae [38] applied in leave-one-out fashion for domain-generalized few-shot learning. Finally, we note that we exclude Photo (P) domain from training our model in this work; this is to ensure the domain shift between PACS and target domains like Mini-ImageNet and CUB.

Table 3: Effects of backbones with varying depths on domain generalized few-shot classification. We consider ProtoNet as the backbone of our x-EML framework, and Conv-n indicates n ConvBlocks. Note that x-EML is trained on PACS in 3-way 5-shot settings, with results presented on PACS, Omniglot, mini-ImageNet, and CUB. It can be seen that, use of deeper backbones is not necessarily preferable for domain generalization due to possible overfitting of source domains.

| | PACS → PACS | Omniglot | Mini-ImageNet | CUB |
|---|---|---|---|---|
| Conv-2 | 62.88 | 92.33 | 55.41 | 52.73 |
| Conv-3 | 64.55 | **94.74** | 60.11 | 57.55 |
| Conv-4 | 64.78 | 94.17 | **61.38** | **58.59** |
| Conv-5 | 64.92 | 92.33 | 60.47 | 57.78 |
| Conv-6 | 65.62 | 90.99 | 58.92 | 56.52 |

As for the target-domain data to be recognized, we consider Omniglot [18], Mini-ImageNet [29] and CUB [40]. Omniglot is a hand-written character dataset containing 1623 different hand-written character images from 50 different alphabets. Mini-ImageNet consists of a subset of 100 classes from the ImageNet dataset [8], and includes 600 images for each class. CUB, a bird species dataset holding 200 classes and 11,788 images in total, which widely serves as a fine-grained benchmark. Since the target-domain data are to be recognized, we do not split them into separate subsets as [39, 29, 15] did. Instead we take each target-domain dataset for testing purposes.

With the standard N-way K-shot protocol in few-shot classi-fication, all experiments in this paper are conducted in 1-shot or 5-shot over 5-way classification. Following [5], we randomly sample N classes from the target domain, with K and 16 images sampled for each class to form support and query sets, respectively. We present the quantitative results by averaging the accuracy of 1000 runs for Omniglot [31], and 600 runs for Mini-ImageNet and CUB [5].

In our work, we employ two most commonly used metric-based deep learning models, ProtoNet [31] and RelationNet [33], as the backbones. For ProtoNet, the model architecture contains a feature encoder and a metric method which is built upon Euclidean distance. The feature encoder backbone comprises stacked ConvBlocks (i.e., convolutional blocks) where a ConvBlock consists of a convolutional layer followed by a batch normalization layer, an ReLU activation layer and an optional max pooling layer. For RelationNet, the model architecture also contains only a feature encoder and a metric method which is a learnable relation module outputting similarity scores between images. The relation module contains 2 ConvBlocks followed by 2 linear layers which is identical to [33], and the feature encoder backbone comprises stacked ConvBlocks as described above. More specifically, to achieve fair comparisons all feature encoders consist of 5 ConvBlocks, denoted as Conv-5, if not explicitly assigned, and the channel size of ConvBlock is capped to 512. Moreover, we obeyed the following training procedure in all experiments. All images were resized to 64×64 as inputs while no data augmentation was applied, and we performed training via the SGD optimizer with a learning rate of 0.0005 and a warm-up of 3000 steps.

### 4.2. Evaluation and Analysis

#### *4.2.1. Quantitative Evaluation*

To quantitatively evaluate the performance of our proposed model, we present few-shot classification results on unseen target domains of Omniglot, Mini-ImageNet and CUB. Tables 1 and 2 list the performances using ProtoNet and RelatiionNet, respectively. For comparison purposes, we first consider baseline methods of using training data of a single domain in PACS for training the few-shot learning models (i.e., Single Domain-S, C, A in the above tables). For state-of-the-arts domain-generalized FSL method, we consider LFT [36]. We note that, since the experiment protocol in [36] is not identical to ours (LFT always observes mini-ImageNet plus additionally data sampled from other source domains during training), we adapt their released code for experiments.

From the results shown in Tables 1 and 2, we observe that our x-EML was able to produce promising recognition results on novel classes in unseen test target domains, with improvements over baseline and LFT observed. We note that, in order to verify the effectiveness of cross-domain meta-learning strategy presented in x-EML, we consider a controlled version (denoted as x-EML*) which collects support and query sets from the same domain (i.e., $D_i = D_j$) in each episode during training. Although satisfactory results were achieved by this controlled version, our full model x-EML is preferable and exhibits more excellent domain generalization ability in few-shot classification.

#### *4.2.2. Analysis of backbone choices of metric-learning models*

We now discuss the choice of metric-learning models for the task of domain-generalized few-shot classification. As noted in 3.2, our x-EML is applicable to various metric-based algorithms. In Tables 1 and 2, we list and compare the performances using ProtoNet [31] and RelationNet [33] as the backbone models. While these two network models share the same feature encoder, ProtoNet applies Euclidean distance and RelationNet utilizes a learnable relation module for matching purposes.

From the above tables, we observe that RelationNet suffered from remarkable performance drops on all three test target domains. We feel that the reason is due to the use and learning of the relation module in RelationNet. The relation module is deployed to compare in few-shot settings, aiming at introducing abilities to recognize novel class data. However, the learning of this module assumes that base and novel class data are from the same domain [33]. In other words, there is no guarantee that such relations can be generalized to matching unseen target domain data. A recent work of [5] also observes this property when conducting few-shot classification, which reported a severe performance drop if domain difference between base and novel classes exists. Next, we will further investigate how the architecture of the feature encoder affects under our settings.

#### *4.2.3. Effects of backbone depth on domain generalization*

In [5], it is observed that deeper model backbone is able to improve the performances in few-shot classification tasks under the single domain setting. In other words, when base and novel classes come from the same domain, one generally expect the use of deeper backbones would be preferable (if not overfitting the dataset). In this work, we are intrigued to observe if the same property holds under the domain-generalized settings.

To assess this issue, we carry out extensive experiments by incrementally stacking ConvBlocks to build up feature encoder backbones. As the results shown in Table 3, we observe that deeper backbones did not necessarily ensure better generalization (e.g., Conv-3 and Conv-4 were preferable while the performances declined generally as the backbone grew deeper). This is very different from what was observed in the standard single-domain few-shot classification tasks (as concluded in [5]). By comparing the first column in Table 3 (i.e., PACS → PACS) and those in other columns (i.e., results on unseen domains), we see that features learned via deeper backbones tend to better model the training data and its domain, which limits its domain generalization capability. However, if domain generalization ability is not of interest, we still see that deeper backbones would still be preferable (if sufficient training data). Nevertheless, the above experiments verify and explain why in cross-domain meta-learning schemes, training models with deeper backbones would not necessarily increase the generalization ability. Note that for all the experiments in Table 3 (for both meta-training and meta-testing stages) are conducted under 3-way 5-shot settings (instead of 5-way 5-shot), since PACS contains only 7 classes in total.

#### *4.2.4. Target Domain Characteristics*

To uncover how the characteristics of the target domain might affect the performances of domain-generalized few-shot classification, we now take a closer look at Table 1. With the same *N*-way *K*-shot setting, we see that the results of mini-ImageNet were much lower than those of Omniglot (by a performance drop of about 30%). Since mini-ImageNet contains real-world images of various objects while Omniglot holds hand-written character images consisting of simple black strokes against white background, recognition novel classes in mini-ImageNet would be much more challenging than that using Omniglot. On the other hand, when comparing the performances between mini-ImageNet and CUB, the results of CUB were generally lower than those of mini-ImageNet. Since CUB is a fine-grained dataset containing real-world images of different bird species, discriminating between CUB images would be viewed as a more difficult task, and thus the above trend can be expected. Conclusively, we found that recognizing real-world fine-grained novel classes poses a great challenge to domain-generalized few-shot classification, which would require future research efforts for further improvements

## 5. CONCLUSIONS

In this paper, we proposed a novel learning framework, Cross-Domain Episodic Meta-Learning (x-EML), to tackle the task of domain-generalized few-shot classification. In order to recognize novel classes in unseen target domains, our learning framework aims at deriving domain-invariant yet category-discriminative representations from multiple source domains. Instead of fitting the label spaces in the source domains, the presented meta-learning scheme ensures the derived features to be generalized to data domains and object categories not observed during training. Our experimental results quantitatively verify the effectiveness of our x-EML and its superiority over baselines and state-of-the-arts methods. In addition, we are able to provide insights into the choice of backbone models and their depths, explaining why deeper backbones do not necessarily generalize across distinct domains, and why fine-grained datasets would pose a challenge to domain-generalized few-shot classification.

## REFERENCES

[1] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. arXiv, 2017.

[2] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In NeurIPS, 2018.

[3] Gilles Blanchard, Gyemin Lee, and Clayton Scott. Generalizing from several related classification tasks to a new unlabeled sample. In NeurIPS, 2011.

[4] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-

level domain adaptation with generative adversarial networks. In CVPR, 2017.

[5] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In ICLR, 2019.

[6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In CVPR, 2018.

[7] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In CVPR, 2019.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009.

[9] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In ICML, 2017.

[10] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning. In NeurIPS, 2018.

[11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain adversarial training of neural networks. JMLR, 2016.

[12] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In CVPR, 2018.

[13] Yunhui Guo, Noel CF Codella, Leonid Karlinsky, John R Smith, Tajana Rosing, and Rogerio Feris. A new benchmark for evaluation of cross-domain few shot learning. arXiv, 2019.

[14] Bharath Hariharan and Ross Girshick. Low-shot visual recognition by shrinking and hallucinating features. In ICCV, 2017.

[15] Nathan Hilliard, Lawrence Phillips, Scott Howland, Artëm Yankov, Courtney D Corley, and Nathan O Hodas. Few-shot learning with metric-agnostic conditional embeddings. arXiv, 2018.

[16] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In ICML workshop, 2015.

[17] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In ICCV workshops, 2013.

[18] Brenden Lake, Ruslan Salakhutdinov, Jason Gross, and Joshua Tenenbaum. One shot learning of simple visual concepts. In CogSci, 2011.

[19] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. Science, 2015.

[20] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In ICCV, 2017.

[21] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In AAAI, 2018.

[22] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In ICCV, 2019.

[23] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In CVPR, 2018.

[24] Yiying Li, Yongxin Yang, Wei Zhou, and Timothy M Hospedales. Feature-critic networks for heterogeneous domain generalization. In ICLR, 2019.

[25] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In ICML, 2013.

[26] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. arXiv, 2018.

[27] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In NeurIPS, 2018.

[28] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In CVPR, 2018.

[29] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In ICLR, 2017.

[30] Victor Garcia Satorras and Joan Bruna Estrach. Few-shot learning with graph neural networks. In ICLR, 2018.

[31] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In NeurIPS, 2017.

[32] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In CVPR, 2019.

[33] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In CVPR, 2018.

[34] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, and Hugo Larochelle. Meta-dataset: a dataset of datasets for learning to learn from few examples. In ICLR, 2020.

[35] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In CVPR, 2018.

[36] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. In ICLR, 2020.

[37] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In CVPR, 2017.

[38] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In CVPR, 2018.

[39] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In NeurIPS, 2016.

[40] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. Technical report, California Institute of Technology, 2011.

[41] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In CVPR, 2018.

[42] Zheng Xu, Wen Li, Li Niu, and Dong Xu. Exploiting low-rank structure from latent domains for domain generalization. In ECCV, 2014.

[43] Sicheng Zhao, Bo Li, Xiangyu Yue, Yang Gu, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. Multi-source domain adaptation for semantic segmentation. In NeurIPS, 2019.

[44] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. TPAMI, 2017.