# Dynamic Distillation Network for Cross-Domain Few-Shot Recognition with Unlabeled Data

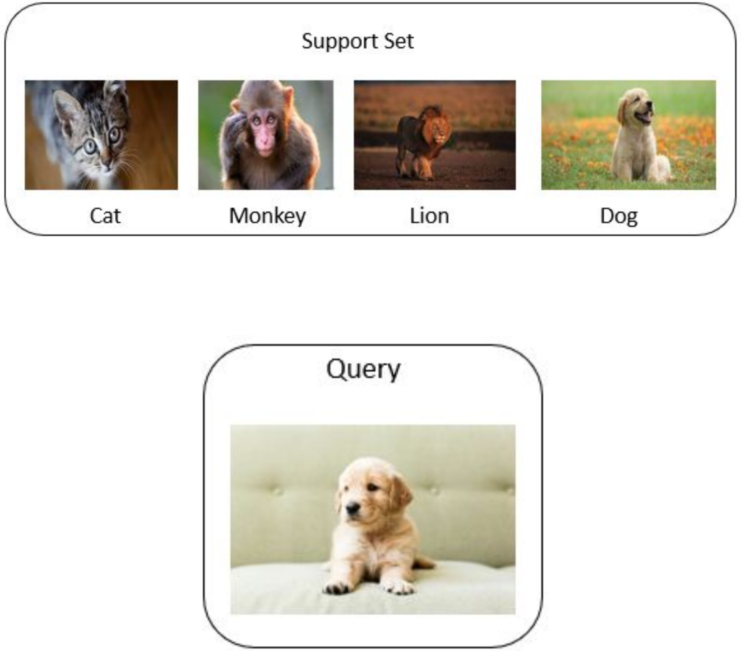
Presenter: Wenhan Yang

# What is few-shot learning

Definition: The problem of making predictions based on a limited number of samples.

Typical process:

- Meta-training: A base dataset (the training set) with labeled images for training.
- Meta-testing: Adapted to a set of novel classes with only a few examples per class (the support set) and evaluated on a set of test images from the same novel classes (the query set)
- Important to note: training set and the support/query set  are of disjoint classes, but same domain


Support Set
Cat
Monkey
Lion
Dog
Query

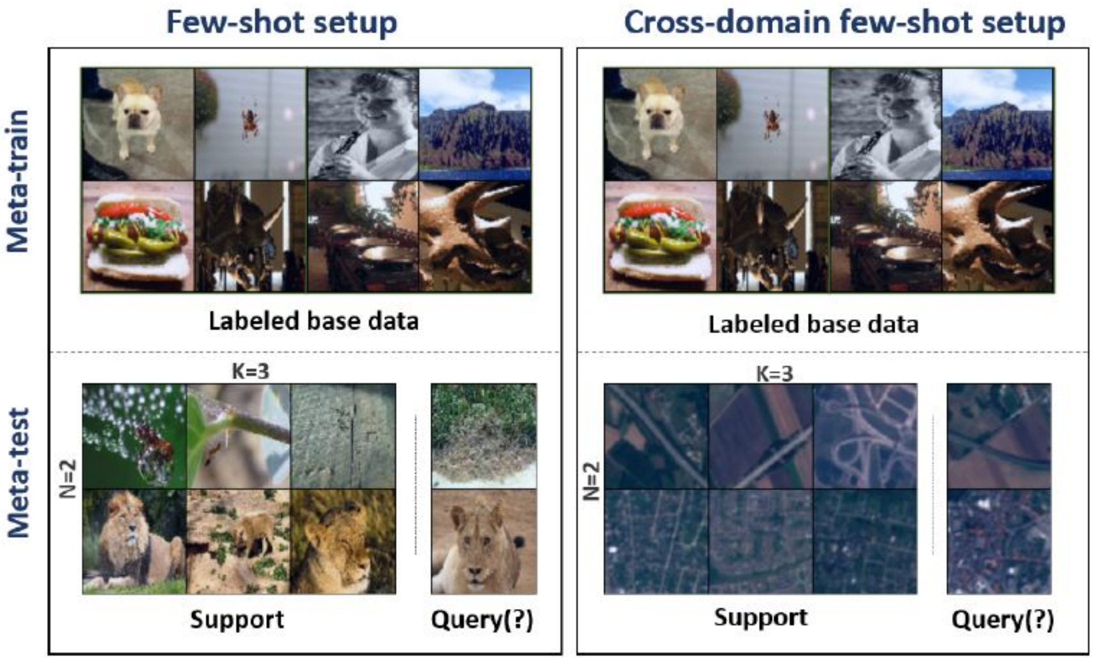
# Limitation of few-shot learning

Problem: training the model on a base dataset from the same domain as the target dataset is difficult and infeasible.

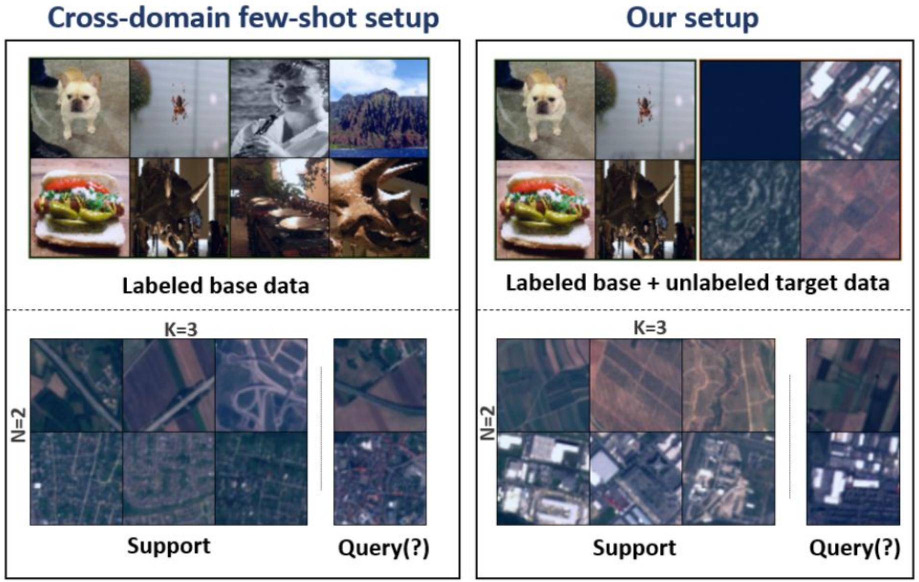
# Cross-domain few-shot learning

BSCD-FSL:

- Datasets from extremely different domains
- Source dataset and target dataset are of different domains
- Surprising result: Traditional pretraining and fine-tuning outperforms few-shot learning method


Few-shot setup
Cross-domain few-shot setup
Meta-train
Meta-test
Labeled base data
Labeled base data
K=3
N=2
Support
Query(?)
K=3
N=2
Support
Query(?)

# Proposed Method




Setup: additional unlabeled images during meta-training stage

# Proposed Method

Motivation for the new setup:

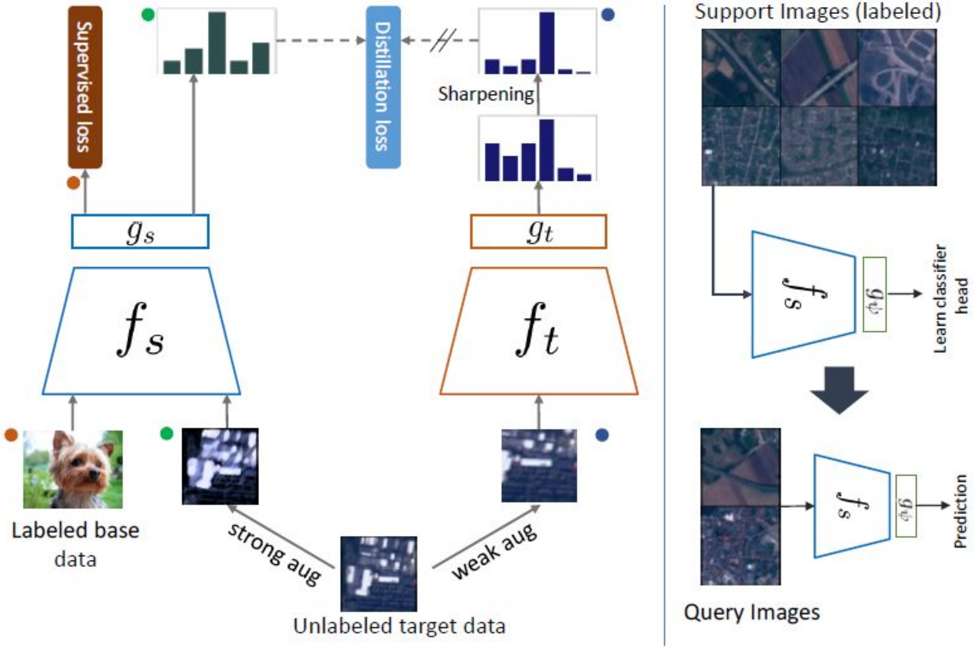
- Learn more target domain specific representations
- Combining supervised and unsupervised learning -> more transferable representation.

# Proposed Method

Goal: Train a feature extractor via dynamic distillation

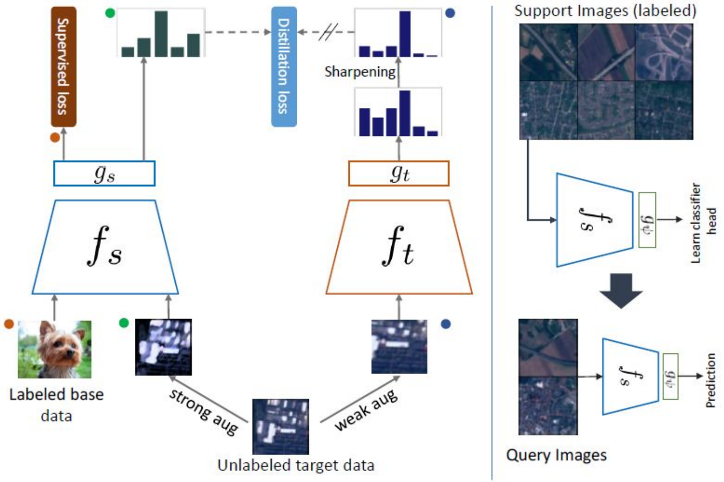
Architecture: encoder + classifier

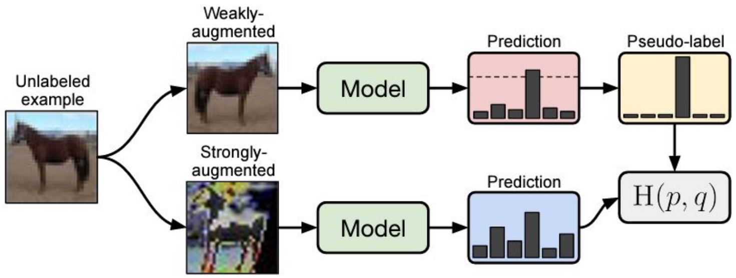
- Student network
  - labeled data: supervised cross-entropy loss
  - unlabeled data: strong augmentation
- Teacher network
  - unlabeled data: weak augmentation
  - produce pseudo labels
  - updated as a moving average of the student network



Supervised loss
Distillation loss
Sharpening
$g_s$
$g_t$
$f_s$
$f_t$
Labeled base data
strong aug
weak aug
Unlabeled target data
Support Images (labeled)
Learn classifier head
Prediction
Query Images

# Relationship with Fixmatch



# Differences with Fixmatch

1. Fixmatch does not assume cross domain
2. Fixmatch generate strongly/weakly augmented data with the same network
3. soft pseudo-labeling

## Proposed Method

- Encoder (supervised loss)

$$l_{\mathrm{CE}}(y, p) = H(y, p)$$

where $p = \mathtt{Softmax}(g_s(f_s(x)))$, and $H(a, b) = -a \log b$.

# Dynamic distillation

Consistency Regularization

$$p_i^s = \texttt{Softmax}(g_s(f_s(x_i^s))); \quad p_i^w = \texttt{Softmax}(g_t(f_t(x_i^w))/\tau)$$

$$l_U(p_i^w, p_i^s) = H(p_i^w, p_i^s)$$

# Teacher network update

$$\theta_t = m\theta_t + (1 - m)\theta_s$$

- m = 0: student network = teacher network
- 0 < m < 1:  teacher network is a moving average of the student network (dynamic distillation)
- m = 1: fixed teacher network

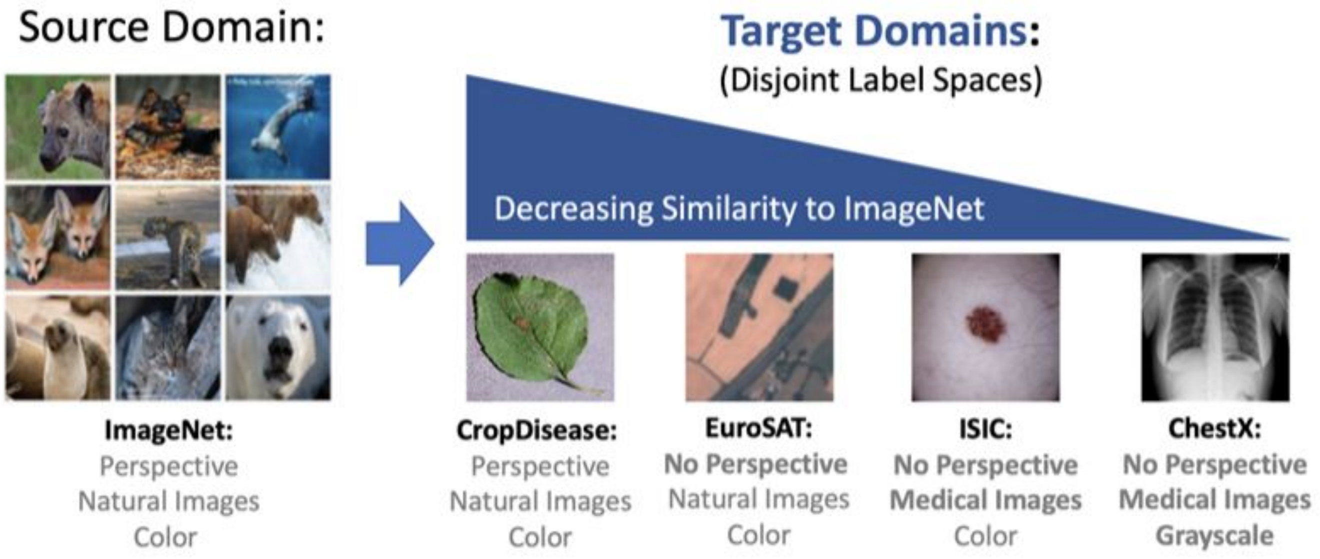
# Total loss function

$$\mathcal{L} = \frac{1}{N_S} \sum_{(x_i, y_i) \in \mathcal{D}_S} l_{\mathrm{CE}}(y, p) + \lambda \frac{1}{N_U} \sum_{x_i \in \mathcal{D}_U} l_U(p_i^w, p_i^s)$$

# Dataset

Base Dataset: miniImageNet (100 classes, each class has 600 images)

Novel Data: CropDisease, EuroSAT, ISIC, ChestX: Cover plant disease images, satellite images, dermoscopic images of skin lesions, and X-ray images, respectively.


Source Domain:
Target Domains:
(Disjoint Label Spaces)
Decreasing Similarity to ImageNet
ImageNet:
Perspective
Natural Images
Color
CropDisease:
Perspective
Natural Images
Color
EuroSAT:
No Perspective
Natural Images
Color
ISIC:
No Perspective
Medical Images
Color
ChestX:
No Perspective
Medical Images
Grayscale

## Some Implementation details

- ResNet-10 as the backbone network
- Two-Step pretraining:
    - train on only miniImageNet for 200 epochs
        - m = 0.9
    - train with both miniImageNet and unlabeled dataset
        - 20% of the original set as the unlabeled dataset
        - m = 0.99

# Some Implementation details

- Weak augmentation
  - random-resize-crop
  - horizontal flip
  - normalization augmentations
- Strong augmentation:
  - color jitter
  - gaussian blur
  - random grayscale transformations

# Result

| Model | 1-shot | | | | 5-shot | | | |
|---|---|---|---|---|---|---|---|---|
| | EuroSAT | CropDisease | ISIC | ChestX | EuroSAT | CropDisease | ISIC | ChestX |
| MAML* | - | - | - | - | 71.70±.72 | 78.05±.70 | 40.13±.58 | 23.48±.48 |
| ProtoNet* | - | - | - | - | 73.29±.71 | 79.72±.79 | 39.57±.57 | 24.05±1.01 |
| MetaOpt* | - | - | - | - | 64.44±.73 | 68.41±.73 | 36.28±.50 | 22.53±.91 |
| STARTUP† | 63.88±.84 | 75.93±.80 | 32.66±.60 | 23.09±.43 | 82.29±.60 | 93.02±.45 | 47.22±.61 | 26.94±.45 |
| ProtoNet | 55.32±.88 | 52.94±.81 | 29.58±.57 | 21.32±.37 | 76.92±.67 | 81.84±.68 | 42.49±.58 | 24.72±.43 |
| MatchingNet | 54.88±.90 | 46.86±.88 | 27.37±.51 | 20.65±.29 | 68.00±.68 | 63.94±.84 | 33.96±.54 | 22.62±.36 |
| Transfer | 58.14±.83 | 68.78±.84 | 32.12±.59 | 22.60±.39 | 80.09±.61 | 89.79±.52 | 43.88±.57 | 26.51±.43 |
| SimCLR(Base) | 58.28±.90 | 61.58±.88 | 32.43±.56 | 22.37±.42 | 80.83±.64 | 83.44±.61 | 44.04±.55 | 26.63±.46 |
| SimCLR | 62.63±.87 | 69.22±.93 | 31.45±.59 | 23.59±.44 | 82.76±.59 | 89.31±.53 | 42.18±.54 | **29.56±.49** |
| STARTUP | 64.32±.88 | 74.45±.86 | 31.73±.57 | 22.27±.41 | 83.58±.60 | 92.41±.47 | 45.73±.62 | 26.21±.46 |
| Transfer+SimCLR | 63.91±.83 | 70.35±.85 | 31.67±.55 | **23.72±.44** | 85.78±.51 | 91.10±.49 | 45.97±.54 | 29.45±.10 |
| Ours | **73.14±.84** | **82.14±.78** | **34.66±.58** | 23.38±.43 | **89.07±.47** | **95.54±.38** | **49.36±.59** | 28.31±.46 |

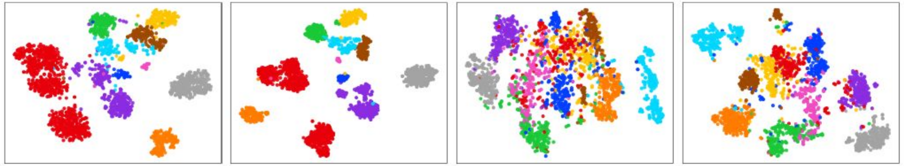
# Effect of dynamic distillation

1. Use pretrained model to extract features
2. Use KMeans to create clusters from the feature

# Effect of dynamic distillation

Table 4: **V-measure cluster score (%) [20] on the KMeans clustering of the extracted features with the ground-truth clustering**. The backbone is ResNet-10 pretrained on the miniImageNet dataset and/or the unlabeled target dataset.

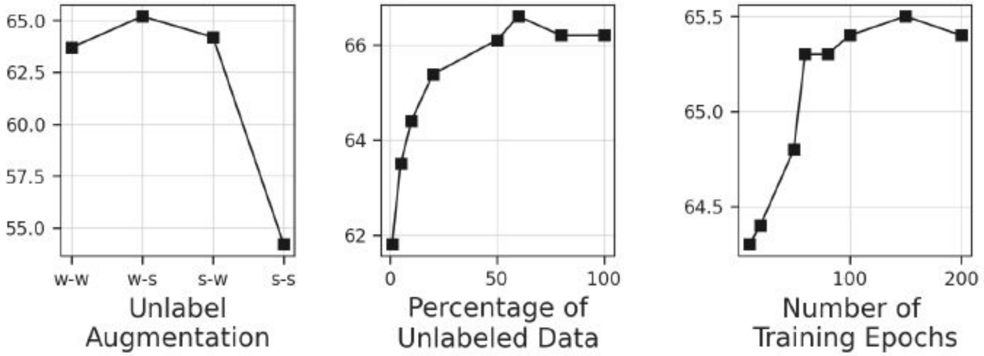
| | EuroSAT | CropDisease | ISIC | ChestX |
|---|---|---|---|---|
| Transfer | 57.01 | 62.58 | **14.67** | 2.45 |
| SimCLR | 60.06 | 62.02 | 12.12 | **3.84** |
| STARTUP | 62.02 | 69.50 | 14.05 | 2.71 |
| Ours | **69.58** | **73.27** | 14.32 | 3.32 |



(a) Transfer (b) Ours (c) Transfer (d) Ours

Figure 3: t-SNE plot of 10 classes from CropDisease (a & b) and EuroSAT (c & d) test sets with features obtained from Transfer and our method.

# Effect of data augmentation / longer training / # of unlabeled data




(a) (b) (c)

# Effect of momentum parameter (dynamic)

Table 13: **Effect of momentum parameter for teacher update.** 5-way 1-shot and 5-shot scores on the BSCD-FSL benchmark datasets for ResNet-10 backbone pretrained on miniImageNet dataset. The mean and 95% confidence interval of 600 runs are reported.

| | $m$ | EuroSAT | CropDisease | ISIC | ChestX |
|---|---|---|---|---|---|
| 5-way 1-shot | | | | | |
| Ours (fixed) | 0 | 69.99±.91 | 76.78±.81 | 35.99±.63 | 22.44±.43 |
| Ours (self) | 1 | 70.01±.87 | 82.27±.80 | 33.87±.59 | 22.98±.45 |
| Ours | 0.99 | **73.14±.84** | 82.14±.78 | **34.66±.58** | **23.38±.43** |
| 5-way 5-shot | | | | | |
| Ours (fixed) | 0 | 86.26±.53 | 93.24±.41 | 50.35±.60 | 26.56±.46 |
| Ours (self) | 1 | 88.17±.47 | 95.22±.37 | 48.45±.61 | 28.03±.47 |
| Ours | 0.99 | **89.07±.47** | **95.54±.38** | **49.36±.59** | **28.31±.46** |

# Comparison with Fixmatch

Essential difference:

- Momentum teacher(m)
  - Fixmatch under-performs the proposed method, but the accuracy improves with the additional m.
- Soft-pseudo-labelling (sharpening parameter)

Table 15: **Ablation on sharpening temperature.**

| $\tau$ | EuroSAT | CropDisease | ISIC | ChestX | Mean |
|---|---|---|---|---|---|
| 0.02 | 88.44 | 95.25 | 49.28 | 28.25 | 65.30 |
| 0.06 | 88.56 | 95.46 | 48.23 | 28.17 | 65.11 |
| 0.2 | 88.40 | 95.18 | 47.99 | 28.17 | 64.94 |
| 0.5 | 88.97 | 95.09 | 48.12 | 28.35 | 65.13 |
| 0.8 | 88.08 | 94.70 | 49.11 | 27.87 | 64.94 |
| 1 | 88.06 | 94.83 | 49.83 | 28.08 | 65.20 |
| 2 | 86.30 | 90.49 | 47.06 | 26.92 | 62.69 |

# Contribution

- A simple method for few-shot learning across extreme domain difference
- Dynamic distillation based approach that uses both labeled source data and unlabeled target data
- Outperforms the current state of the art