Boosting the Generalization Capability in Cross-Domain Few-shot Learning via Noise-enhanced Supervised Autoencoder

2021 ICCV OCTOBER 11-17 VIRTUAL

Hanwen Liang[*1] Qiong Zhang[*†2] Peng Dai[1] Juwei Lu[1]

[1]Huawei Noah's Ark Lab, Canada. [2]Department of Statistics, University of British Columbia, Vancouver, Canada. [*]Equal contribution with alphabetical order. [†]Work done as an intern in Noah's Ark Lab.

Problem & Motivation

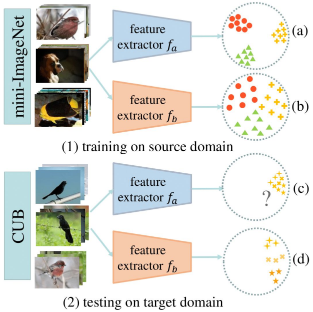
Problem: cross-domain few shot classification.

- Two domains: source and target domain.
- Enough data on source domain.
- Limited data on the target domain.
- Goal: borrow information from source domain to help the recognition on the target domain.

General approach: transfer learning

Observation: trade-off between the accuracy & generalization capability of the model trained on the source domain.

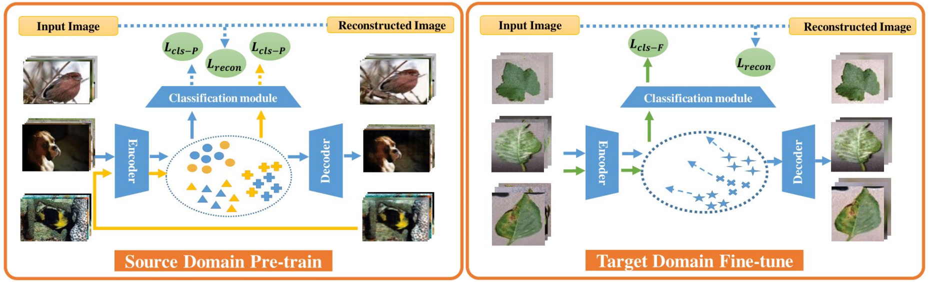
- A competent feature extractor f_a fits the source domain very well (a) yet fails when transferred to different classes in target domain (c).
- A less perfect feature extractor f_b on the source dataset (b) may have stronger generalization capability and perform well in target domain (d).



Motivation:

- increase generalization capability of the classifier trained on the source domain.

Method



An overview of the proposed pipeline. A noise-enhanced supervised autoencoder (NSAE) is pre-trained with source dataset on the source domain to improve the generalization capability. The fine-tuning on the target domain is a two-step procedure that first performs reconstruction task on novel dataset, and then the encoder is fine-tuned for classification.

Pre-train on the source domain

- Noise-enhanced autoencoder jointly predicts the labels of inputs and reconstructs the inputs
- Feed the reconstructed images as noisy inputs into encoder for classification.

$$\mathcal{L}_{\text{NSAE}}(\phi,\psi;\mathcal{D}_s)=\frac{1}{M}\sum_{m=1}^{M}\mathcal{L}_{\text{SAE}}^{\lambda_1,\text{cls-P}}(\phi,\psi;\mathbf{x}_m^s,y_m^s)+\frac{\lambda_2}{M}\sum_{m=1}^{M}\mathcal{L}_{\text{cls-P}}(\boldsymbol{\theta};f_\phi(\hat{\mathbf{x}}_m^s),y_m^s)$$

$$\mathcal{L}_{\text{SAE}}^{\lambda,cls}(\phi,\psi;\mathbf{x},y)=\mathcal{L}_{cls}(\tilde{\mathbf{x}},y)+\lambda\mathcal{L}_{\text{REC}}(\phi,\psi;\mathbf{x})$$

Fine-tune on the target domain

- Step 1: reconstruct the images for a few epochs by minimizing the reconstruction loss.
- Step 2: only train the encoder by minimizing the classification loss during fine-tune.

Classification loss functions:

- Classification loss during pre-train: cross entropy (CE), batch spectral regularization (BSR).
- Classification loss during fine-tune: CE, distance based loss for few-shot learning (D).
- The combination leads to 4 loss functions: CE+CE, BSR+CE, CE+D, BSR+D

Network architecture:

- Conv4 and ResNet10 as encoder
- Self-designed decoder that is roughly symmetric to the encoder.

Experiment setting:

- 5-way K-shot (K=5,20,50)
- 600 repetitions

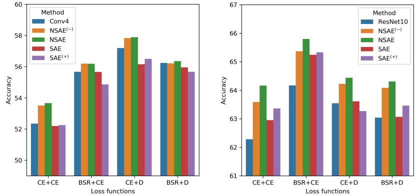
Experiment

Dataset:

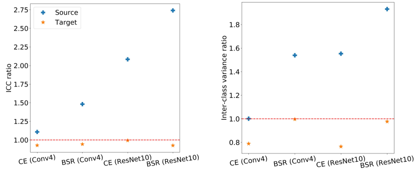
- Source: miniImageNet.
- Target: CropDisease, EuroSAT, ISIC, ChestX, Car, CUB, Plantae, Places

Ablation study and statistical analysis:

- A single encoder on source domain(Encoder).
- One-step fine-tune based on NSAE (NSAE-).
- Train with the proposed method (NSAE).
- Do not feed in reconstructed image during pre-train (SAE).
- Double the weight on the classification loss of original images without classifying the reconstructed image during pre-train(SAE(+)).



- Use intra-class correlation(ICC) to measure the discriminability of feature embeddings.
- Encoder from NASE is not as discriminative as traditional training on source domain, but it generalizes better on target domain.



- We can achieve the SoTA. See full experiment results in our paper.