Tensor feature hallucination for few-shot learning

Michalis Lazarou[1], Tania Stathaki[1], Yannnis Avrithis[2]

[1]Imperial College London [2]Athena RC

Motivation

Problem: Few-shot learning
- Limited labeled data make the problem fundamentally hard
- Difficult to train classifiers with very small amount of labeled data

Solution: Few-shot synthetic feature generation
- Exploit the spatial properties of tensor features to train a feature hallucinator
- Introduce a novel loss function that is simpler than other state of the art few-shot synthetic data generation methods

Problem definition

Pre-training:
- Base class data: $D_{\text{base}} := \{(x_i, y_i)\}_{i=1}^{I}$ where $y_i \in C_{\text{base}}$
- Train embedding network $f_\theta : \mathcal{X} \to \mathbb{R}^{d \times h \times w}$ on D_{base}
- $\bar{f}_\theta : \mathcal{X} \to \mathbb{R}^d$ denotes f_θ followed by global average pooling (GAP)

Inference stage
- Novel class data D_{novel} with C_{novel} disjoint from C_{base}
- Sample a support set S and a query set Q from D_{novel}
- S consists of N classes with K labeled examples per class
- Given S and f_θ, classify examples from Q

Representation learning

First stage: train the backbone network using standard cross-entropy:

$$L(D_{\text{base}}; \theta, \phi) := \sum_{i=1}^{I} L_{\text{CE}}(c_\phi(\bar{f}_\theta(x_i)), y_i) + R(\phi) \tag{1}$$

Second stage: adopt a self-distillation process:

$$L_{\text{KD}}(D_{\text{base}}; \theta', \phi') := \alpha L(D_{\text{base}}; \theta', \phi') + \beta \text{KL}(c_{\phi'}(\bar{f}_{\theta'}(x_i)), c_\phi(\bar{f}_\theta(x_i))) \tag{2}$$

Meta-training hallucinator

- Meta-training stage: sample episodes of S from D_{base}
- Pre-trained network $f_{\theta'} : \mathcal{X} \to \mathbb{R}^{d \times h \times w}$ maps images to tensors
- Class prototype tensors:

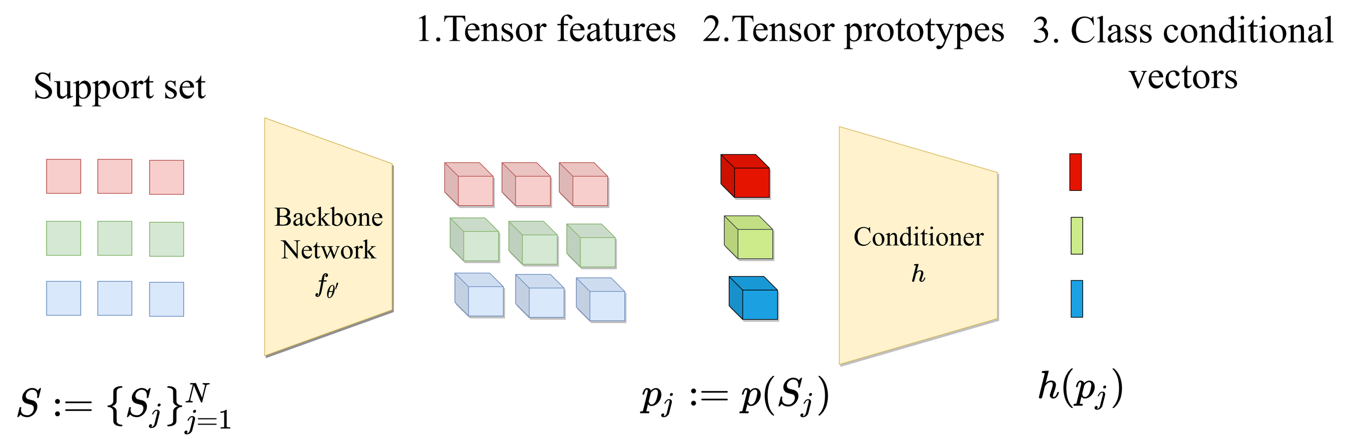
$$p_j := \frac{1}{K} \sum_{i=1}^{K} f_{\theta'}(x_i^j) \tag{3}$$

- Conditioner h maps prototypes to class-conditional vectors $s_j := h(p_j) \in \mathbb{R}^{d'}$
- Generator maps to $g(z; s_j) \in \mathbb{R}^{d \times h \times w}$ where $z \sim \mathcal{N}(\mathbf{0}, I_k)$ Jointly trained by:

$$L_{\text{hal}}(X; h, g) = \frac{1}{MN} \sum_{j=1}^{N} \sum_{m=1}^{M} \| g(z_m; h(p_j)) - p_j \|^2 \tag{4}$$

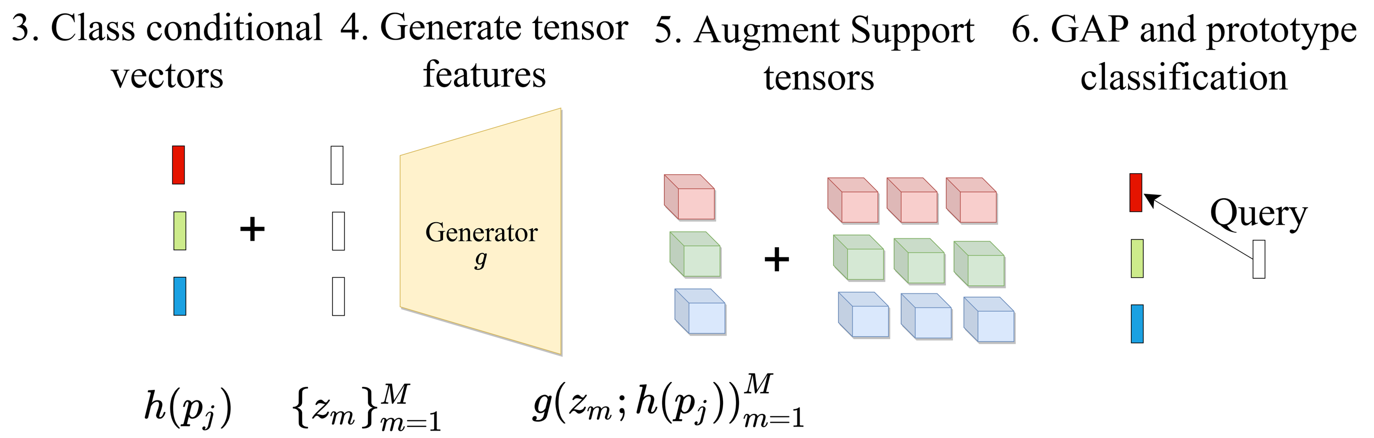
Inference stage: Tensor prototypes

- Map all images from S and Q to tensors using the pre-trained $f_{\theta'}$
- Calculate the class tensor prototype of each class using equation 3
- Conditioner h maps prototypes to class-conditional vectors $s_j := h(p_j) \in \mathbb{R}^{d'}$



inference stage: Synthetic tensor features and classification

- Generator maps to $g(z; s_j) \in \mathbb{R}^{d \times h \times w}$ where $z \sim \mathcal{N}(\mathbf{0}, I_k)$
- Use the augmented support set to train a prototypical classifier
- Classify each query from query set, Q



Implementation details

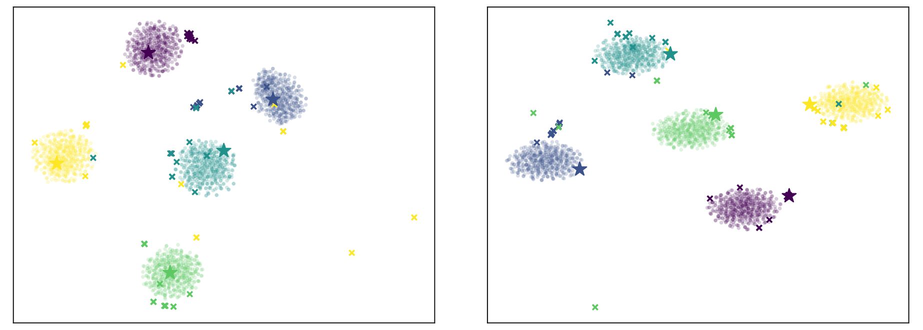
- Our implementations are based on PyTorch and scikit-learn

Ablation study

Method	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	CUB 1-shot	CUB 5-shot	CIFAR-FS 1-shot	CIFAR-FS 5-shot
Baseline (1)	ResNet-18	$56.81_{\pm0.81}$	$78.31_{\pm0.59}$	$67.14_{\pm0.89}$	$86.22_{\pm0.50}$	$65.71_{\pm0.95}$	$84.68_{\pm0.61}$
Baseline-KD (2)	ResNet-18	$59.62_{\pm0.85}$	$79.31_{\pm0.62}$	$70.85_{\pm0.90}$	$87.64_{\pm0.48}$	$69.15_{\pm0.94}$	$85.89_{\pm0.59}$
VFH (ours)	ResNet-18	$61.88_{\pm0.85}$	$79.63_{\pm0.61}$	$75.44_{\pm0.85}$	$87.82_{\pm0.47}$	$72.31_{\pm0.91}$	$85.64_{\pm0.64}$
TFH (ours)	ResNet-18	$64.49_{\pm0.84}$	$79.94_{\pm0.60}$	$75.66_{\pm0.85}$	$88.39_{\pm0.49}$	$73.77_{\pm0.85}$	$86.68_{\pm0.63}$
TFH-ft (ours)	ResNet-18	**65.07**$_{\pm0.82}$	**80.81**$_{\pm0.61}$	**75.76**$_{\pm0.83}$	**88.60**$_{\pm0.47}$	**74.77**$_{\pm0.90}$	**86.88**$_{\pm0.59}$

Standard few-shot classification. Baseline (1), Baseline-KD (2): prototypical classifier at inference, no feature generation. VFH: our vector feature hallucinator; TFH: our tensor feature hallucinator; TFH-ft: our tensor feature hallucinator followed by novel-task fine-tuning

Visualizations



(a) *mini*ImageNet, ResNet-18 (b) CUB, ResNet-18

t-SNE visualization of the augmented support feature set of an 1-shot task using both ResNet-18 and ResNet-12 backbones on *mini*ImageNet and CUB. Colors indicate different classes. $\star$: support features; $\bullet$: generated features; '$\times$': query features.

CUB original images (row 1) followed by images gen-erated from separately trained reconstructors using as input tensorfeatures (row 2) or vector features (row 3).

Experiments

State of the art comparisons:

Method	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	CUB 1-shot	CUB 5-shot
Dual TriNet [2]	ResNet-18	$58.80_{\pm1.37}$	$76.71_{\pm0.69}$	69.61	84.10
IDeMe-Net [1]	ResNet-18	$59.14_{\pm0.86}$	$74.63_{\pm0.74}$	–	–
AFHN [3]	ResNet-18	$62.38_{\pm0.72}$	$78.16_{\pm0.56}$	$70.53_{\pm1.01}$	$83.95_{\pm0.63}$
VI-Net [4]	ResNet-18	61.05	78.60	74.76	86.84
TFH (ours)	ResNet-18	$64.49_{\pm0.84}$	$79.94_{\pm0.60}$	$75.66_{\pm0.85}$	$88.39_{\pm0.49}$
TFH-ft (ours)	ResNet-18	**65.07**$_{\pm0.82}$	**80.81**$_{\pm0.61}$	**75.76**$_{\pm0.83}$	**88.60**$_{\pm0.47}$

Standard few-shot classification. Comparison of our TFH to SOTA few-shot data augmentation methods.

Cross-domain few-shot learning:

Method	mIN→CUB 1-shot	mIN→CUB 5-shot	mIN→CIFAR-FS 1-shot	mIN→CIFAR-FS 5-shot
Baseline (1)	$43.14_{\pm0.78}$	$62.20_{\pm0.70}$	$50.25_{\pm0.86}$	$69.43_{\pm0.74}$
Baseline-KD (2)	$44.40_{\pm0.82}$	**63.83**$_{\pm0.73}$	$51.54_{\pm0.89}$	$70.40_{\pm0.72}$
VFH (ours)	$44.77_{\pm0.79}$	$62.61_{\pm0.73}$	$50.36_{\pm0.87}$	$69.31_{\pm0.74}$
TFH (ours)	$45.67_{\pm0.80}$	$63.08_{\pm0.73}$	$51.82_{\pm0.89}$	$69.77_{\pm0.76}$
TFH-ft (ours)	**45.96**$_{\pm0.80}$	$63.64_{\pm0.74}$	**53.07**$_{\pm0.86}$	**71.29**$_{\pm0.75}$

Few-shot cross-domain classification. Comparison of our TFH to variants and baselines as defined in 2, using ResNet-18 trained on *mini*ImageNet (mIN).

Experiments

Alternative classifiers:

Method	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	CUB 1-shot	CUB 5-shot
Logistic regression					
Baseline-KD (2)	ResNet-18	$61.83_{\pm0.82}$	$79.27_{\pm0.61}$	$72.74_{\pm0.88}$	$87.71_{\pm0.49}$
TFH (ours)	ResNet-18	$64.03_{\pm0.84}$	$79.93_{\pm0.60}$	$75.08_{\pm0.85}$	**88.82**$_{\pm0.46}$
TFH-ft (ours)	ResNet-18	**64.83**$_{\pm0.82}$	**80.49**$_{\pm0.61}$	**75.43**$_{\pm0.85}$	$88.52_{\pm0.48}$
Support vector machine					
Baseline-KD (2)	ResNet-18	$60.21_{\pm0.84}$	$78.28_{\pm0.61}$	$71.23_{\pm0.89}$	$86.34_{\pm0.52}$
TFH (ours)	ResNet-18	$64.20_{\pm0.83}$	$79.62_{\pm0.60}$	$75.64_{\pm0.85}$	**88.74**$_{\pm0.45}$
TFH-ft (ours)	ResNet-18	**65.06**$_{\pm0.82}$	**80.33**$_{\pm0.60}$	**75.77**$_{\pm0.83}$	$88.22_{\pm0.46}$

Alternative classifiers. Our TFH, variants baseline-KD as defined in 2, using ResNet-18, where at inference, the prototypical classifier is replaced by logistic regression or SVM.

Alternative backbone:

Method	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	CUB 1-shot	CUB 5-shot
Logistic regression					
Baseline	ResNet-12	$61.83_{\pm0.85}$	$79.60_{\pm0.60}$	$68.49_{\pm0.49}$	$84.75_{\pm0.59}$
TFH (ours)	ResNet-12	$64.20_{\pm0.84}$	$80.21_{\pm0.58}$	$68.37_{\pm0.92}$	**85.12**$_{\pm0.55}$
TFH-ft (ours)	ResNet-12	**64.53**$_{\pm0.53}$	**80.25**$_{\pm0.62}$	**69.59**$_{\pm0.88}$	$85.08_{\pm0.55}$
Support vector machine					
Baseline	ResNet-12	$59.94_{\pm0.84}$	$78.43_{\pm0.60}$	$66.72_{\pm0.90}$	$83.29_{\pm0.60}$
TFH(ours)	ResNet-12	$64.62_{\pm0.83}$	$79.56_{\pm0.58}$	$68.75_{\pm0.90}$	$84.05_{\pm0.56}$
TFH-ft(ours)	ResNet-12	**64.64**$_{\pm0.86}$	**79.88**$_{\pm0.62}$	**69.42**$_{\pm0.88}$	**84.41**$_{\pm0.55}$

Alternative backbone network. Our TFH, variants and baselines as defined in 2, using the publicly available pre-trained ResNet-12 backbones provided by [5].

[1] Z. Chen, Y. Fu, Y.-X. Wang, L. Ma, W. Liu, and M. Hebert. Image deformation meta-networks for one-shot learning. In *CVPR*, 2019.

[2] Z. Chen, Y. Fu, Y. Zhang, Y. Jiang, X. Xue, and L. Sigal. Multi-level semantic feature augmentation for one-shot learning. *IEEE Transactions on Image Processing*, 2019.

[3] K. Li, Y. Zhang, K. Li, and Y. Fu. Adversarial feature hallucination networks for few-shot learning. In *CVPR*, 2020.

[4] Q. Luo, L. Wang, J. Lv, S. Xiang, and C. Pan. Few-shot learning via feature hallucination with variational inference. In *WACV*, 2021.

[5] C. Zhang, Y. Cai, G. Lin, and C. Shen. DeepEMD: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In *CVPR*, 2020.

Contact: michalis.lazarou14@imperial.ac.uk