Augmented Bi-path Network for Few-shot Learning

Baoming Yan[1,2*], Chen Zhou[1*], Bo Zhao[3], Kan Guo[2], Jiang Yang[2], Xiaobo Li[2], Ming Zhang[1] and Yizhou Wang[1]

[1] Peking University, [2] Alibaba Group, [3] The University of Edinburgh

(* Indicates equal contribution)

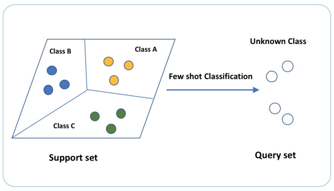
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

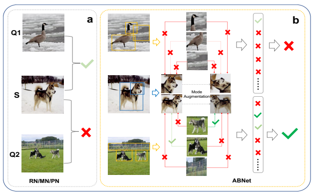
Problem Definition

Goal: Few-shot Learning (FSL) which aims to learn from few labeled training data and train a good classification model with few labeled data.

Definition:

For few-shot classification, the dataset contains a meta-train set and a meta-test set, which have disjoint categories ($C_{meta-train} \cap C_{meta-test} = \emptyset$). The meta-train set, in which each category contains many labeled samples, is used to train a generalized model. Then, it is evaluated on the meta-test set with only few labeled images in each category. In the popular setting, many testing episodes are performed during evaluation process. Each testing episode contains N categories and each category has K labeled images. These labeled data form the support set $S = \{(x_{11}, y_{11}), (x_{12}, y_{12}) \dots, (x_{N,K}, y_{N,K})\}$, where (x_{ij}, y_{ij}) is jth datum of ith class. The rest unlabeled testing data of these N categories form the query set Q=$\{(x_{1,K+1}, y_{1,K+1}), (x_{1,K+2}, y_{1,K+2}), \dots (x_{N,K+1}, y_{N,K+1}), (x_{N,K+2}, y_{N,K+2})\}$. Such setting is called N-way K-shot learning.



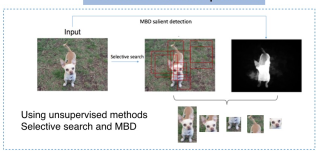


Motivation:

- Previous FSL method learns to compare the testing (query) image and training (support) image by simply concatenating the features of two images and feeding it into the neural network.
- With few labeled data in each class, the neural network has difficulty in learning or comparing the local features of two images. Such simple image-level comparison may cause serious mis-classification.
- For more precise comparison, we propose Augmented Bi-path Network (ABNet) to learn to compare both global and local features on multi-scales.

• Learn to augment

• Introduce salient patches



Key Contribution:

- We propose Augmented Bi-path Network (ABNet) to learn to compare both global and local features of support and query images, which includes two novel modules, namely "Learning to Augment" and "Learning to Compare".
- We evaluate our approach on three challenging Few-shot Learning benchmarks, miniImageNet, Caltech-256 and tieredImageNet. Our ABNet outperforms the state-of-the-art by a large margin.

Experiments and Results

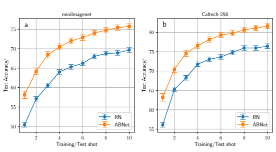
Results on three challenging Few-shot Learning benchmark:

Method		miniImageNet		Caltech-256		tieredImageNet	
		5way1shot	5way5shot	5way1shot	5way5shot	5way1shot	5way5shot
MatchingNet [16]	NIPS'16	43.56±0.84	55.31±0.73	45.59±0.77	54.61±0.73	54.02	70.11
MetaLSTM [21]	ICLR'17	43.44±0.77	60.60±0.71	-	-	-	-
MAML [20]	ICML'17	48.70±0.84	55.31±0.73	48.09±0.83	57.45±0.84	51.67±1.81	70.30±0.08
MetaNet [17]	ICML'17	49.21±0.96	-	-	-	-	-
ProtoNet [12]	NIPS'17	49.42±0.87	68.20±0.70	-	-	54.28±0.67	71.42±0.61
RelationNet [13]	CVPR'18	50.44±0.82	65.32±0.77	56.12±0.94	73.04±0.72	54.48±0.93	71.32±0.78
CTM [42]	CVPR'19	41.62	58.77	-	-	-	-
Spot&Learn [37]	CVPR'19	51.03±0.78	67.96±0.71	-	-	-	-
MetaOptNet [43]	CVPR'19	52.87±0.57	68.76±0.48	-	-	54.71±0.67	71.79±0.59
ABNet		**58.12±0.94**	**72.02±0.75**	**63.20±0.99**	**78.42±0.69**	**62.10±0.96**	**75.11±0.78**

Results of Resnet backbone:

Method		Backbone	miniImageNet	
			5way1shot	5way5shot
RelationNet [13]	CVPR'18	ResNet-18	58.21	74.29
CTM [42]	CVPR'19	ResNet-18	62.05±0.55	78.63±0.06
MetaOptNet [43]	CVPR'19	ResNet-12	62.64±0.61	78.63±0.46
ABNet		ResNet-18	**63.15±0.63**	**78.85±0.56**

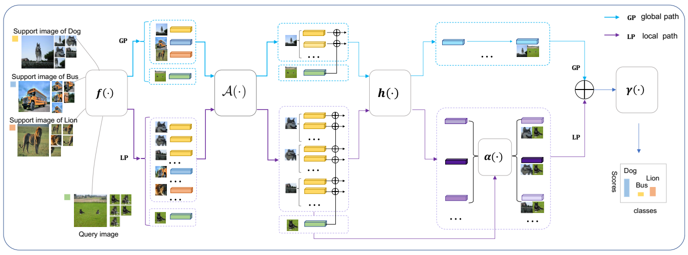
Results of different shots:



Method

To fully exploit the few labeled images, we propose Augmented Bi-path Network (ABNet) for Few-shot Learning, the framework is illustrated as follow.

Network Architecture:



Algorithm:

1. ***Salient Patch Extraction***: starts with the patches sampled by selective search method and selects **Top5 Salient Patches** measured by:

$$S(r_i) = \Pi(r_i) \cdot \frac{1}{K}\sum_{j=0}^{K} v(p_j)$$

$$v(p_j) = \min_{\pi \in \mathbb{S}}[\max_{t=0}^{T} I(\pi(t)) - \min_{t=0}^{T} I(\pi(t))]$$

2. ***Feature Embedding***: encode the whole image and salient patches by CNN backbone *f*.
3. ***Learning to Augment***:

$$\begin{pmatrix} x_i^a \\ y_i^a \\ z_i^a \end{pmatrix} = \mathcal{A}\begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = \begin{bmatrix} \mathcal{A}_{11}, & \mathcal{A}_{12}, & \mathcal{A}_{13}, & \mathcal{A}_{14} \\ \mathcal{A}_{21}, & \mathcal{A}_{22}, & \mathcal{A}_{23}, & \mathcal{A}_{24} \\ \mathcal{A}_{31}, & \mathcal{A}_{32}, & \mathcal{A}_{33}, & \mathcal{A}_{34} \\ 0, & 0, & 0, & 1 \end{bmatrix} \cdot \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}$$

4. ***Learning to Compare***:

a) Generating Similarity Maps
b) Learning to Re-weight:
c) Learning to Merge

$$o(s,q) = \gamma\Big([\bigoplus_{i,j=1}^{i,j \le N} \alpha(i,j) \cdot m_{s,q}^{i,j}] \oplus m_{s,q}\Big)$$

Loss Function:

Classification Loss: $$\mathcal{L}_{cls} = \frac{1}{B \times C}\sum_{i=1}^{B}\sum_{j=1}^{C}(P(i,j) - \mathbb{I}(y_i == y_j))^2$$

Attention Loss: $$\mathcal{L}_{att} = \frac{S_{att}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\alpha(i,j)|$$

Augmentation Loss: $$\mathcal{L}_{aug} = \frac{S_{aug}}{B \times C}\sum_{i=1}^{B}\sum_{j=1}^{C}\sum_{k=1}^{K}(f(I_q) - f_k(I_s))^2\mathbb{I}(y_i == y_j)$$

Total Loss: $$\mathcal{L} = \mathcal{L}_{cls} + \lambda_{att} \cdot \mathcal{L}_{att} + \lambda_{aug} \cdot \mathcal{L}_{aug}$$

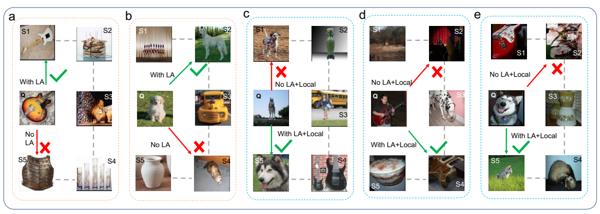
Ablation Studies

Quantitative comparison of four variants of our methods:

- Baseline: fixed global features
- Baseline + LA (Learn to Augment)
- Baseline + LA + SP (Salient Patch)
- Baseline + LA + SP + LR (Learn to Re-weight)

Model	miniImageNet	
	5way1shot	5way5shot
Baseline	52.44±0.91	66.50±0.77
Baseline+LA	54.27±0.91	68.02±0.77
Baseline+LA+SP	55.76±0.89	69.70±0.72
Baseline+LA+SP+LR	**58.12±0.94**	**72.02±0.75**

Visualization Comparison:



Conclusion

We propose a novel meta-learning based method, namely Augmented Bi-path Network, for Few-shot Learning. The proposed method extends the previous "learn-to-compare" based methods by introducing both global and local features on multi-scales, and learns to augment the features for better robustness. Experimental results show that our method significantly outperforms the state-of-the-art on three challenging datasets under all settings.