Few-Shot Few-Shot Learning and the role of Spatial Attention

Yann Lifchitz[1,2], Yannis Avrithis[1], Sylvaine Picard[2]

[1]Univ Rennes, Inria, CNRS, IRISA [2]Safran

Contributions

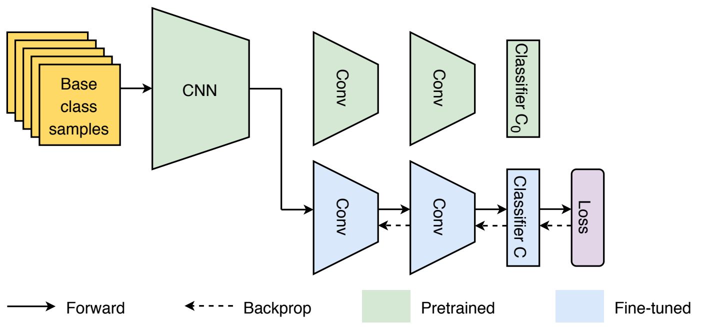
- Novel few-shot learning setting: **Few-shot few-shot learning**
- Few-shot accuracy improvement by using a pretrained network
- Novel domain independent spatial attention mechanism

Few-shot Few-shot

- Previous knowledge modeled as pretrained embedding network on a large scale dataset
- In domain labeled data can be hard to collect
- Base class data is limited to a few or even zero examples per class

Base class training

- If base class examples are available
- Fine-tuning of a copy of the last few layers of the network using [3]



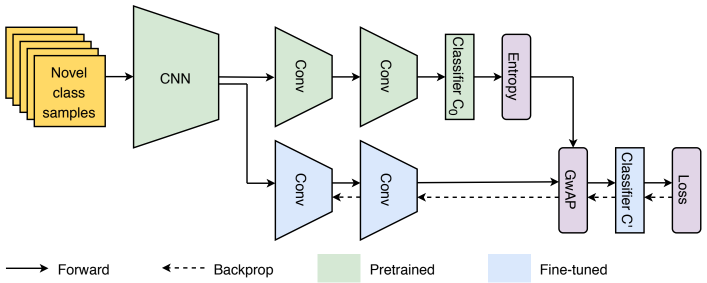
Spatial attention

- With few base class data, the network can't learn to focus on the relevant parts of the images
- Local features are classified using the pretrained classifier as in dense classification [3]
- Certainty of the prediction relates to the discriminative power of the region
- For a region with prediction f over the c_0 base classes, the corresponding weight is: $w := 1 - \frac{H(f)}{\log c_0}$, H being the entropy function
- Global average pooling is replaced by global weighted average pooling (GwAP)

Examples of images overlaid with spatial attention maps

Novel class adaptation

- Original pretrained network and classifier used to produce attention weights
- Global average pooling replaced by global weighted average pooling
- Spatial attention apply at inference after adaptation too



Results

- Experiments on fine-grained dataset: CUB and general classification: miniImageNet
- Modification of miniImageNet to remove overlap with pretraining
- Base class training is important for large domain gaps
- Novel class adaptation and spatial attention are more important when few base class data are available
- Combining the two leads to the best results

Attention		✓		✓
Adaptation			✓	✓
Base	Places			
$k = 0$	38.80±0.24	39.69±0.24	39.76±0.24	40.79±0.24
$k = 1$	40.50±0.23	41.74±0.24	41.11±0.24	42.23±0.24
$k = 5$	56.47±0.28	57.16±0.29	56.69±0.29	57.32±0.29
$k = 10$	62.83±0.30	64.32±0.30	62.97±0.30	64.41±0.30
All	80.68±0.27	80.48±0.27	80.68±0.27	80.56±0.27
Base	Randomly Initialized			
$k = 1$	31.65±0.19	-	31.37±0.19	-
$k = 5$	40.52±0.25	-	40.50±0.26	-
$k = 10$	48.25±0.28	-	48.61±0.29	-
All	71.78±0.30	-	71.77±0.30	-
Baseline++ [1]	67.02±0.90	-	-	-
ProtoNet [4]	71.88±0.91	-	-	-
Ensemble [2]	68.77±0.71	-	-	-

Average 5-way 1-shot novel class accuracy on CUB with ResNet-18 either pre-trained on Places or trained from scratch on k base class examples

References

[1] W. Chen, Y. Liu, Z. Kira, Y. F. Wang, and J. Huang. A closer look at few-shot classification. *ICLR*, 2019.

[2] N. Dvornik, C. Schmid, and J. Mairal. Diversity with cooperation: Ensemble methods for few-shot classification. *ICCV*, 2019.

[3] Y. Lifchitz, Y. Avrithis, S. Picard, and A. Bursuc. Dense classification and implanting for few-shot learning. *CVPR*, 2019.

[4] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. In *NIPS*, 2017.

Contact: yann.lifchitz@safrangroup.com