# FEW-SHOT FEW-SHOT LEARNING AND THE ROLE OF SPATIAL ATTENTION

Yann Lifchitz[1,2], Yannis Avrithis[1], Sylvaine Picard[2]
[1]Inria, Univ Rennes, CNRS, IRISA [2]Safran

# Few-shot learning

- **Definition:**
  - Access to a base dataset of images in class set C
  - Very small support set of images in class set C’
  - The goal is to classify queries in C’
  - C and C’ belong to the same domain

- **Motivation for a new setting:**
  - In domain data (base dataset) can be scarce
  - Does not take advantage of the large scale dataset available

# Overview

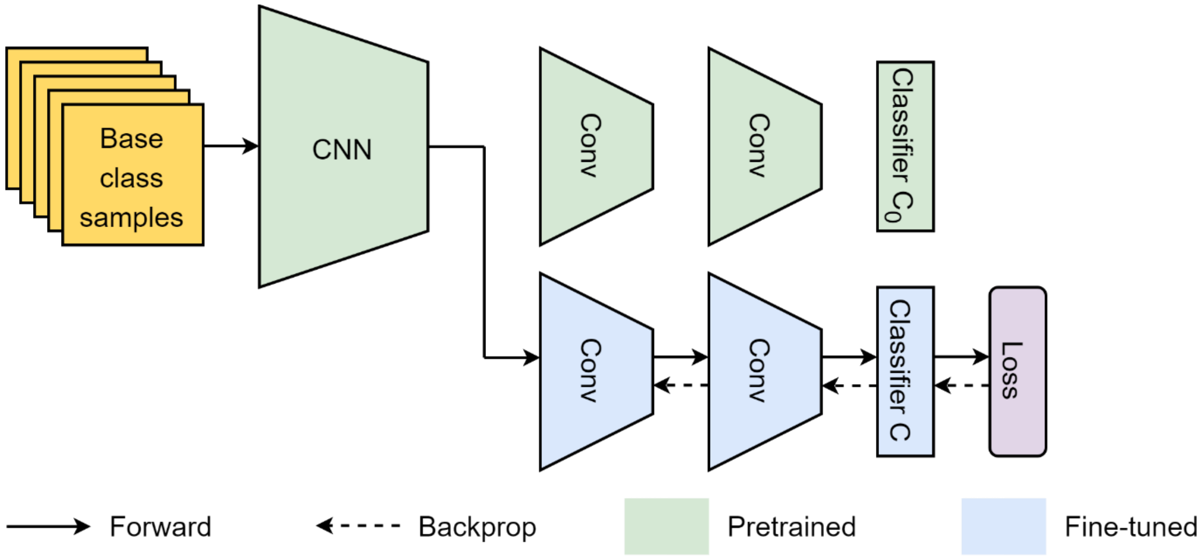
- **New few-shot paradigm:**
  - Model pretrained on a large-scale dataset, with classifier on a prior class set $C^0$
  - Only few or zero examples per base class $C$

- **3 stages:**
  - Domain adaptation (base class dataset)
  - Novel classes adaptation (few-shot novel data)
  - Classifying queries into novel classes $C'$

- **Introduction of a simple attention mechanism that improves classification in this setting**

# Base class training

- In case base class training data is available
- Copy of a pretrained network
- Fine-tuning of the few last layers with dense classification[1]






[1]Dense Classification and Implanting for Few-Shot Learning, Lifchitz et al., CVPR 2019

# Spatial Attention

- **For each pixel of the feature map is computed a prediction on the prior classes $C^0$**

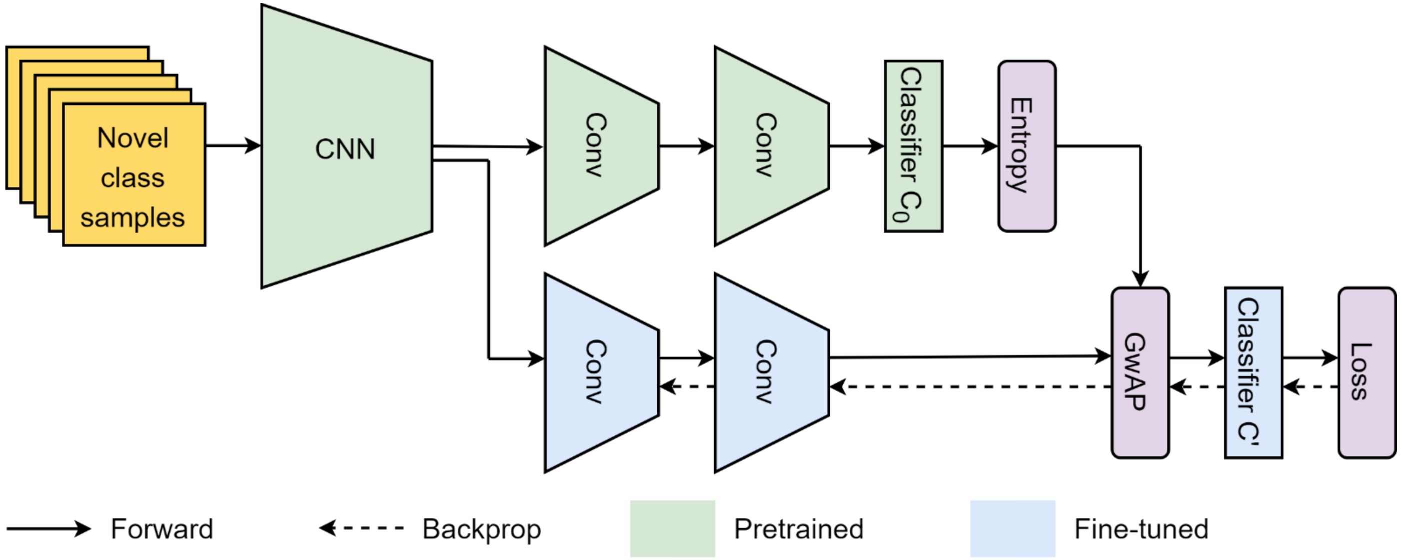
◆ No ground truth (ground truth in $C'$)

- **For each location we compute the entropy of the prediction**
- **Low entropy indicates → discriminative region**
- **High entropy → background**
- **Entropy is normalized to produce weights maps**
- **Pooling function is global weighted average pooling (GwAP)**

Novel classes images, overlaid with entropy-based spatial attention maps

# Novel class adaptation

- Original pretrained network used to produce attention weights
- Global average pooling is replaced by global weighted average pooling
- Same forward process at inference



# Results: CUB

- **Effect of base training**
  - Very important for large domain gap
- **Effect of novel class adaptation**
  - More significant for small base datasets
- **Effect of spatial attention**
  - More significant for small base datasets
  - Appears to be domain independent
- **Attention and adaptation can be combined for the best results**

| | NOVEL: $k' = 1$ | | | | NOVEL: $k' = 5$ | | | |
|---|---|---|---|---|---|---|---|---|
| Attention | | ✓ | | ✓ | | ✓ | | ✓ |
| Adaptation | | | ✓ | ✓ | | | ✓ | ✓ |
| BASE | PLACES | | | | | | | |
| $k = 0$ | 38.80±0.24 | 39.69±0.24 | 39.76±0.24 | 40.79±0.24 | 55.09±0.24 | 56.95±0.23 | 63.29±0.24 | 64.27±0.23 |
| $k = 1$ | 40.50±0.23 | 41.74±0.24 | 41.11±0.24 | 42.23±0.24 | 57.25±0.22 | 58.89±0.23 | 65.42±0.23 | 66.78±0.23 |
| $k = 5$ | 56.47±0.28 | 57.16±0.29 | 56.69±0.29 | 57.32±0.29 | 74.27±0.23 | 74.95±0.23 | 75.82±0.23 | 76.32±0.23 |
| ALL | 80.68±0.27 | 80.48±0.27 | 80.68±0.27 | 80.56±0.27 | 90.38±0.16 | 90.33±0.16 | 91.22±0.15 | 91.17±0.15 |
| BASE | RANDOMLY INITIALIZED | | | | | | | |
| $k = 1$ | 31.65±0.19 | - | 31.37±0.19 | - | 39.45±0.20 | - | 42.70±0.21 | - |
| $k = 5$ | 40.52±0.25 | - | 40.50±0.26 | - | 52.94±0.25 | - | 53.45±0.25 | - |
| ALL | 71.78±0.30 | - | 71.77±0.30 | - | 85.60±0.18 | - | 85.96±0.19 | - |
| Baseline++ [5] | 67.02±0.90 | - | - | - | 83.58±0.54 | - | - | - |
| ProtoNet [5] | 71.88±0.91 | - | - | - | 87.42±0.48 | - | - | - |
| Ensemble [7] | 68.77±0.71 | - | - | - | 84.62±0.44 | - | - | - |

## Thank you for your attention