# [论文笔记] Few-Shot Learning with Global Class Representations

## 0. 写在前面的话

- 解决的问题
  - 由于基类和新类之间**存在严重的样本不均衡问题**，导致容易过拟合到基类数据
  - 训练模型的时候，使用来自novel class的样本。
- 新颖点
  - 在few-shot里面使用类原型。
- 训练数据包括来自novel class的样本。
- 面临的困难- base classes 与 novel classes 样本数量的不对等
  - 数据扩充
  - 元学习策略

项目代码以及开源->网址

## 1，概况

为了解决few-shot中由于基类和新类之间**存在严重的样本不均衡问题**，导致容易过拟合到基类数据的问题，作者将新类样本加入至训练集对模型进行训练。此外，作者将类原型的概念引入few-shot，但相比于之前学习阶段性类别表征的方法，本文使用的是**全局类别表征**直接与所有基类和新类训练样本进行比较，更具可分辨性。

## 2，具体框架介绍

### 2.1 Registration Module

#### 输入&输出

- 输入
  - 局部类原型$\{r_{cj}, c_j \in C_{train}\}$
  - all global class representations $G = \{g_{cj}, cj \in C_{total}\}$
- 输出
  - 一个矩阵，其中每一列向量-> $V_i = [v_i^{c_1}, ...., v_i^{c_N}]^T$

    代表 ith visual features 与 G 全局类原型的相似度

#### 具体步骤

$$
\begin{aligned}
v_i^{c_j} &= \frac{\exp\left(d_i^{c_j}\right)}{\sum_{c_j \in C_{total}} \exp\left(d_i^{c_j}\right)} \\
d_i^{c_j} &= -\left\|\theta(f_i) - \phi(g_{c_j})\right\|_2
\end{aligned}
\tag{1}
$$

where $\theta(\cdot)$ and $\phi(\cdot)$ are embeddings for visual feature of samples and global class representations, respectively.

## 2.2 Sample Synthesis Module

### 输入&输出

- 输入
  - support集合中对应novel类的visual 特征
- 输出
  - 代表这个类的局部类原型

### 具体步骤

- 首先通过随机裁剪、随机翻转和数据幻觉来将每一个novel class的数据扩充至$k_t$
- 对于每一个novel class, 从第一步得到的$k_t$ 样本中随机取$k_r$个，进行如下操作

$$
\begin{aligned}
&r_{c_j} = \sum_{i=1}^{k_r} \frac{\nu_i}{\sum_j \nu_j} f_i, \text{where} y_i = c_j \\
&\hat{k}_r \sim \mathcal{U}(0, k_t), k_r = \lceil \hat{k}_r \rceil, \nu_i \sim \mathcal{U}(0, 1),
\end{aligned}
\tag{3}
$$

where $r_{c_j}$ denotes the synthesized sample for novel class $c_j$. $\mathcal{U}(a, b)$ denotes a uniform distribution ranging from $a$ to $b$.

## 2.3 整体框架介绍

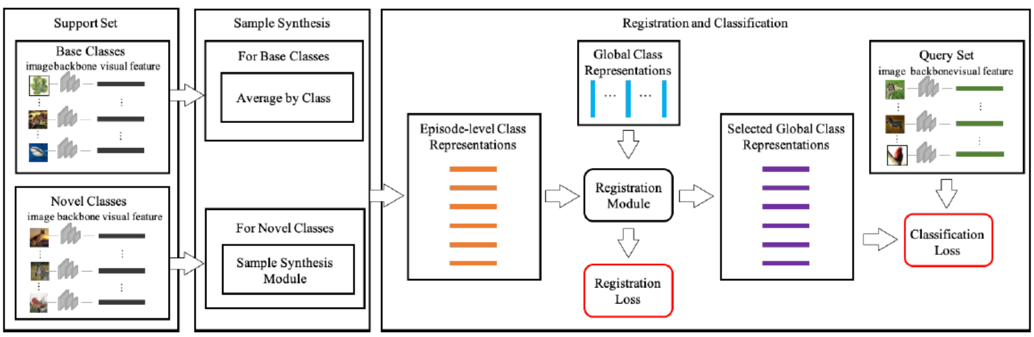


Figure 2. Overview of the whole framework. First, we propose a sample synthesis method to synthesize episodic representation for each class in the support set. Second, the registration module is leveraged to select global representation according to their episodic representation, and the selected global representations are then used to classify query images. The classification loss and registration loss are used to jointly optimize the global representations, the registration module, and the feature extractor. (Best viewed in color)

1. 首先用每一个类中所有样本的视觉特征的平均值初始化全局类原型G
2. 构造局部类原型$\{r_{cj}, c_j \in C_{train}\}$

    1. 对于base classes，直接取support集合中对应类样本视觉特征的平均值
    2. 对于novel classes, 使用Sample Synthesis Module生成
3. 把局部类原型当作输入，通过registration module 认领对应的全局类原型
    1. 其中为了可微行(梯度下降优化参数)，采取软优化，认领$\xi_i = V_i$ G,而非直接$argmax_{V_i}(G)$
4. 计算query集合的视觉特征到全局类原型的欧氏距离，最小者对应的类别为最终结果

**Algorithm 1** Training episode loss computation.

**Input:** Whole class set $C_{total}$, base class set $C_{base}$, novel class set $C_{novel}$, training set $D_{train}$, test set $D_{test}$, feature extractor $F$, registration module $R$ and global class representations $G = \{g_{c_j}, c_j \in C_{total}\}$.
**Output:** The loss for a randomly generated training episode.
1. Randomly sample $n_{train}$ classes from $C_{total}$ to form $C_{train}$;
2. Randomly sample $n_s$ images per class in $C_{train}$ to form a support set $S = \{(x_i, y_i), i = 1, .., n_s \times n_{train}\}$;
3. Randomly sample $n_q$ images per class in $C_{train}$ to form a query set $Q = \{(x_j, y_j), j = 1, .., n_q \times n_{train})\}$;
4. Compute visual features of images in $S$ by using the feature extractor $F$, and obtain visual features $\{f_i = F(x_i), i = 1, ..., n_q \times n_{train}\}$;
5. Construct episodic representations $\{r_{c_i}, c_i \in C_{train}\}$ by using the features within their own classes and the sample synthesis module.
6. Compute the similarity score vector $V_i = [v_i^{c_1}, ...., v_i^{c_N}]^T$ between each episodic representation $r_{c_i}$ and all global class representations $G = \{g_{c_j}, c_j \in C_{total}\}$ according to Equation 4;
7. Compute the registration loss according to Equation 4;
8. Select corresponding global class representation $\{\xi_i, i = 1, .., n_{train}\}$ by using $\xi_i = V_i G$;
9. Compute the classification loss of query images according to Equation 5;
10. Compute the total loss according to Equation 6.

## 3，损失

### registration loss $L_{reg}$

$$
\begin{aligned}
\mathcal{L}_{reg}(r_{c_i}) &= CE(c_i, V_i), \\
v_i^{c_j} &= \frac{\exp\left(d_i^{c_j}\right)}{\sum_{c_j \in C_{total}} \exp\left(d_i^{c_j}\right)} \\
d_i^{c_j} &= -\|\theta(r_{c_i}) - \phi(g_{c_j})\|_2
\end{aligned} \tag{4}
$$

CE(.) 代表交叉熵损失

### classifacation loss $L_{fsl}$

$$
\begin{aligned}
\mathcal{L}_{fsl}(x_k) &= CE(y_k, W_k), \\
w_k^i &= \frac{\exp\left(d_k^i\right)}{\sum_{i \in C_{train}} \exp\left(d_k^i\right)} \\
d_k^i &= -\|F(x_k) - \xi_i\|_2
\end{aligned} \tag{5}
$$

$$
\mathcal{L}_{total}(S, Q) = \sum_{c_i \in S} \mathcal{L}_{reg}(r_{c_i}) + \sum_{k \in Q} \mathcal{L}_{fsl}(x_k) \tag{6}
$$

# 4，部分实验效果

## 4.1 标准小样本学习

作者从 miniImageNet 中取 64 类用作 training，16 类用作 validation，20 类用作 testing。待分类的测试图片只从新类中选取。1 shot 表示在训练集中每个新类有 1 个样本，5 shot 表示每个新类有 5 个样本。每次预测需要从 5 个候选类中选择一个作为给定图片的分类。

| **Model** | **5 way Acc.** | | **20 way Acc.** | |
|---|---|---|---|---|
| | 1 shot | 5 shot | 1 shot | 5 shot |
| MN [36] | 97.9 | 98.7 | 93.5 | 98.7 |
| APL [22] | 97.9 | **99.9** | 97.2 | 97.6 |
| DLM [35] | 98.8 | 95.4 | 99.6 | 98.6 |
| PN [31] | 98.8 | 99.7 | 96.0 | 98.9 |
| MA [9] | 98.7±0.4 | **99.9**±0.1 | 95.8±0.3 | 98.9±0.2 |
| RN [32] | 99.6±0.2 | 99.8±0.1 | 97.6±0.2 | 99.1±0.1 |
| MMN [3] | 99.28±0.08 | 99.77±0.04 | 97.16±0.10 | 98.93±0.05 |
| MG [38] | 99.67±0.18 | 99.86±0.11 | 97.64±0.17 | 99.21±0.10 |
| Ours | **99.72**±0.06 | **99.90**±0.10 | **99.63**±0.09 | **99.32**±0.04 |

Table 1. Comparative results for FSL on the Omniglot dataset. The averaged accuracy (%) on 1,000 test episodes is given followed by the standard deviation (%).

## 4.2 广义小样本学习

z作者仍在 miniImageNet 上进行实验，具体设定和小样本学习的设定相同。唯一的区别在于测试样本同时从基类和新类抽取。每个新类包含 5 个训练样本。衡量指标有三个：

- **$acc_b$**：基类（base）测试样本的分类准确度；
- **$acc_n$**：新类（novel）测试样本的分类准确度；
- **$acc_a$**：：所有（all）测试样本的分类准确度。

| **Model** | $accu_a$ | $accu_b$ | $accu_n$ |
|---|---|---|---|
| MN [36] | 26.98 | 33.54 | 0.75 |
| PN [31] | 31.17 | 39.53 | 0.52 |
| RN [32] | 32.48 | 40.24 | 1.42 |
| Ours | **39.14** | **46.32** | **12.98** |

Table 3. Comparative results (%) on the miniImageNet dataset under the generalized FSL setting. In this setting, test examples are from both the base and novel classes and each approach has to predict labels from the joint label space. Notations: $acc_a$ – the accuracy of classifying the all test samples to all the classes (both base and novel). $acc_b$ – the accuracy of classifying the data samples from the base classes to all the classes. $accu_n$ – the accuracy of classifying the data samples from the novel classes to all the classes.