基于小样本学习的医学影像分割

王浩

2021/06/20

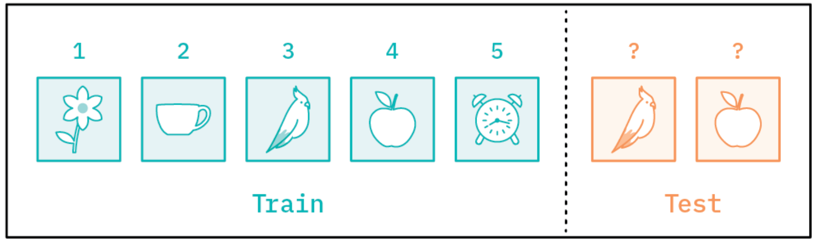
小样本学习

小样本学习（Few-shot learning）

- 小样本数据的学习
- 学习目标任务需要的知识
- 符合自然认知过程

深度学习模式

- 学习一种模式，目标是基于输入数据，得到目标输出
- 数据量要求较高，小样本难以实现效果

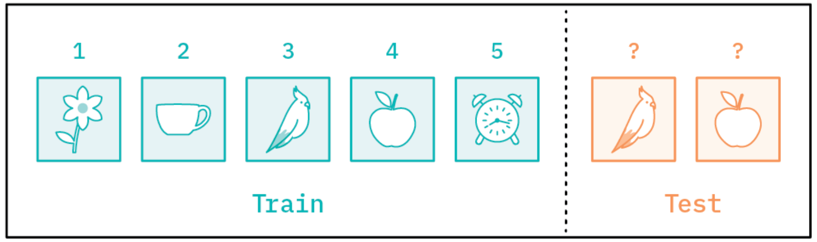


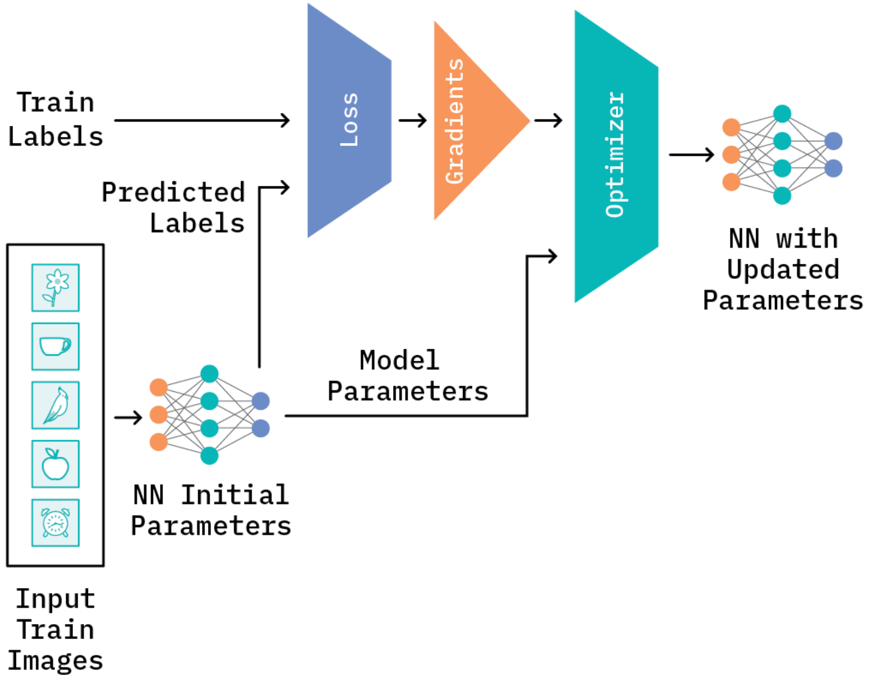
Deep Learning

https://meta-learning.fastforwardlabs.com/

深度学习模式

- 学习一种模式，目标是基于输入数据，得到目标输出
- 数据量要求较高，小样本难以实现效果



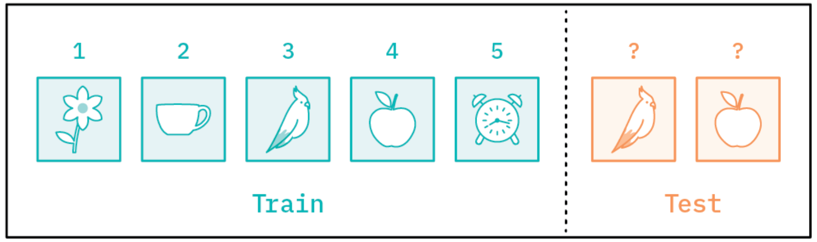


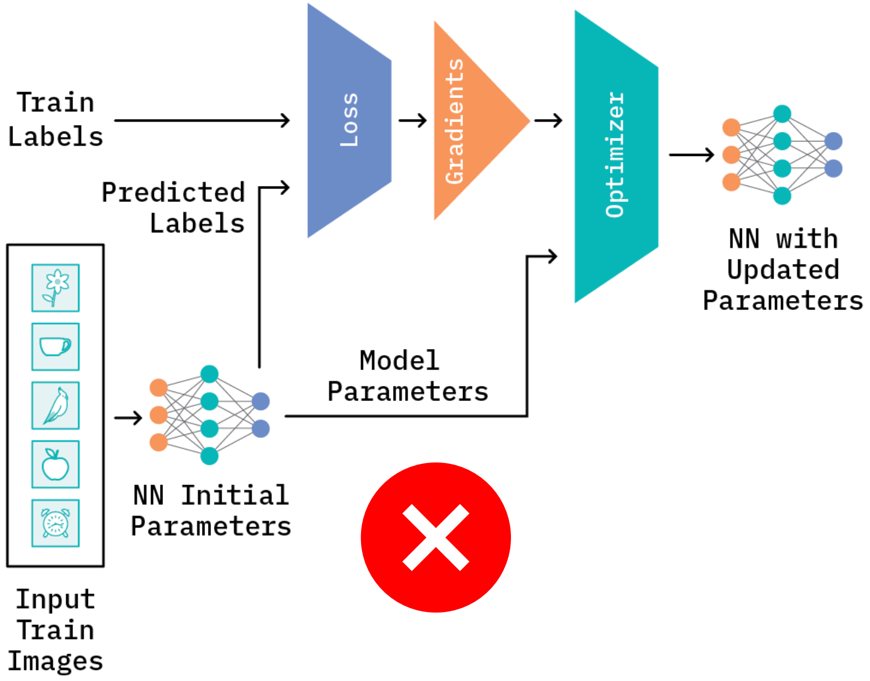
Deep Learning

https://meta-learning.fastforwardlabs.com/

深度学习模式

- 学习一种模式，目标是基于输入数据，得到目标输出
- 数据量要求较高，小样本难以实现效果



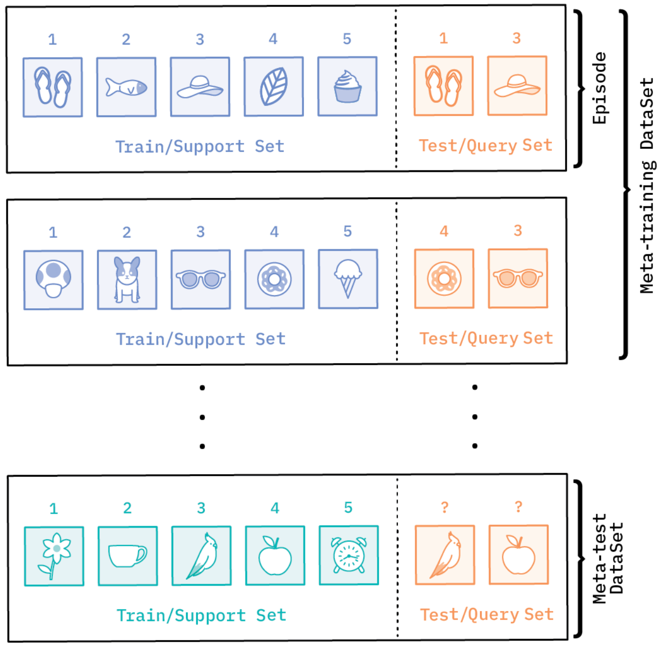


Deep Learning

https://meta-learning.fastforwardlabs.com/

元学习模式

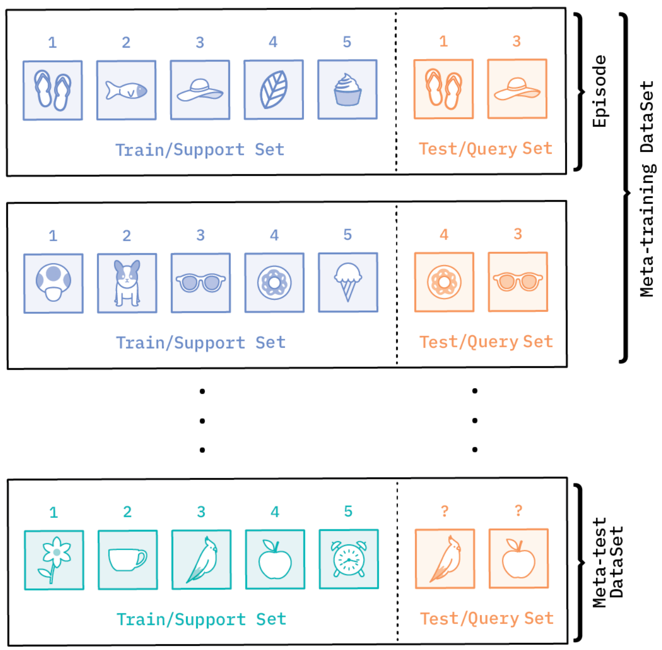
- 小样本学习与元学习存在交集
- 元学习Meta Learning——学习“如何学习”leaning to learn

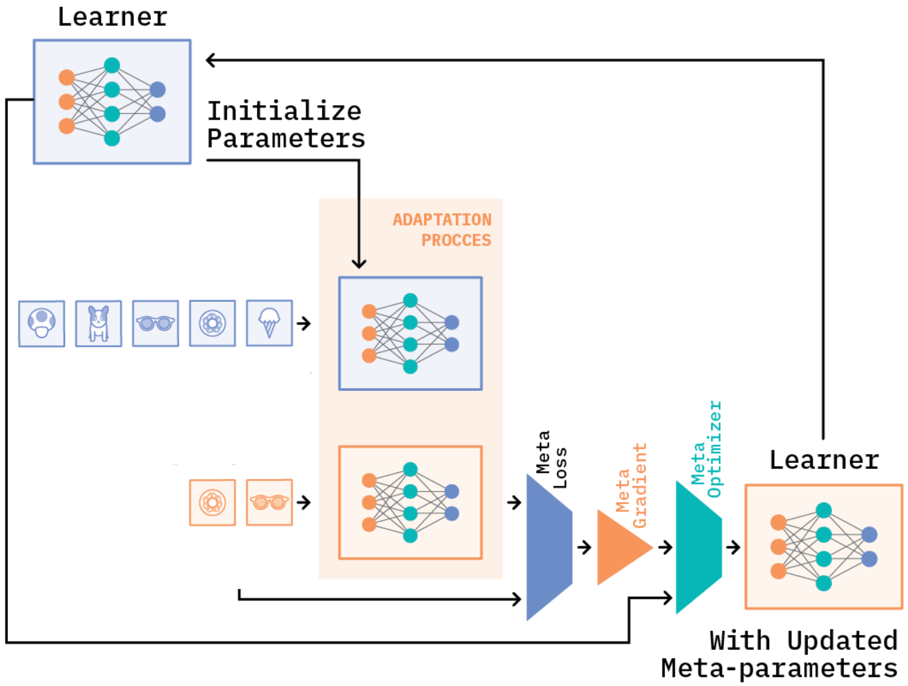


Meta Learning

元学习模式

- 小样本学习与元学习存在交集
- 元学习Meta Learning——学习“如何学习”leaning to learn

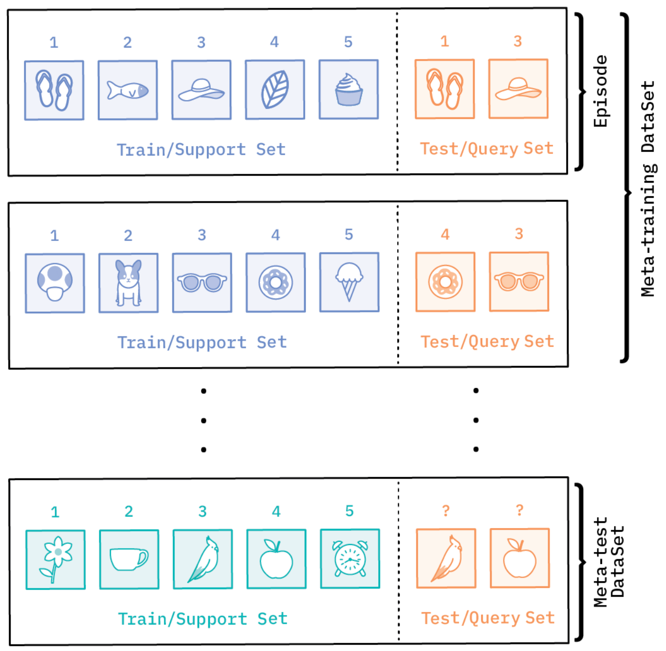


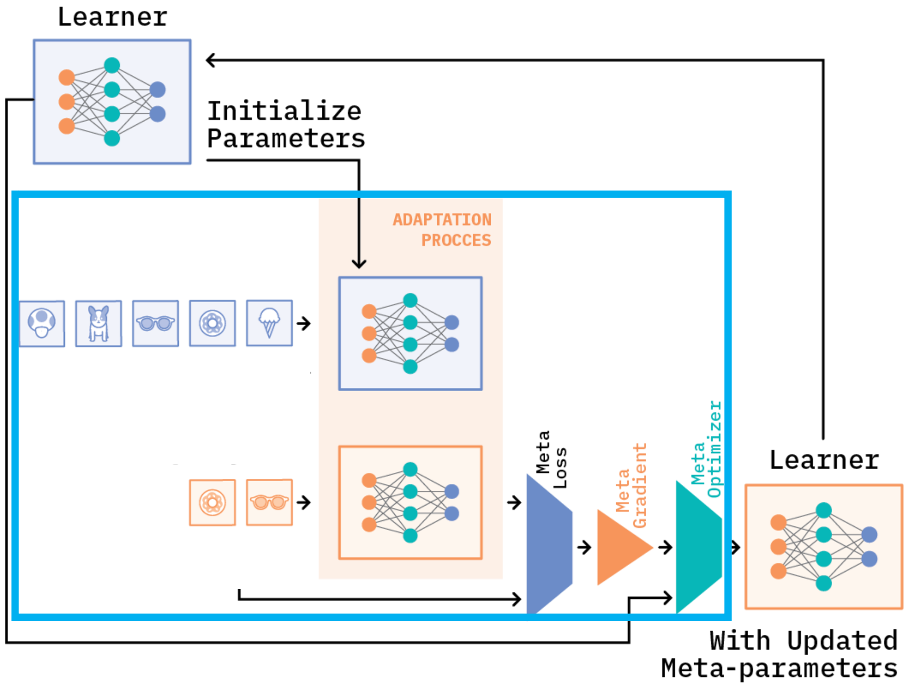


Meta Learning

元学习模式

- 小样本学习与元学习存在交集
- 元学习Meta Learning——学习“如何学习”leaning to learn

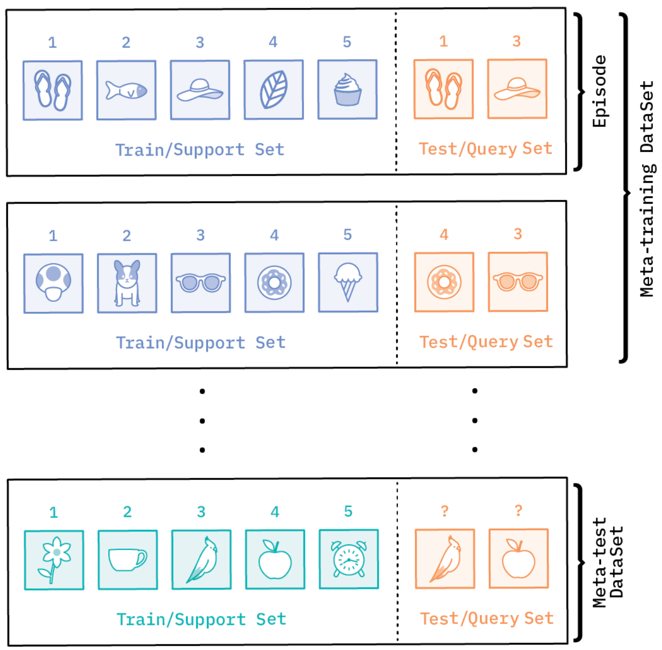


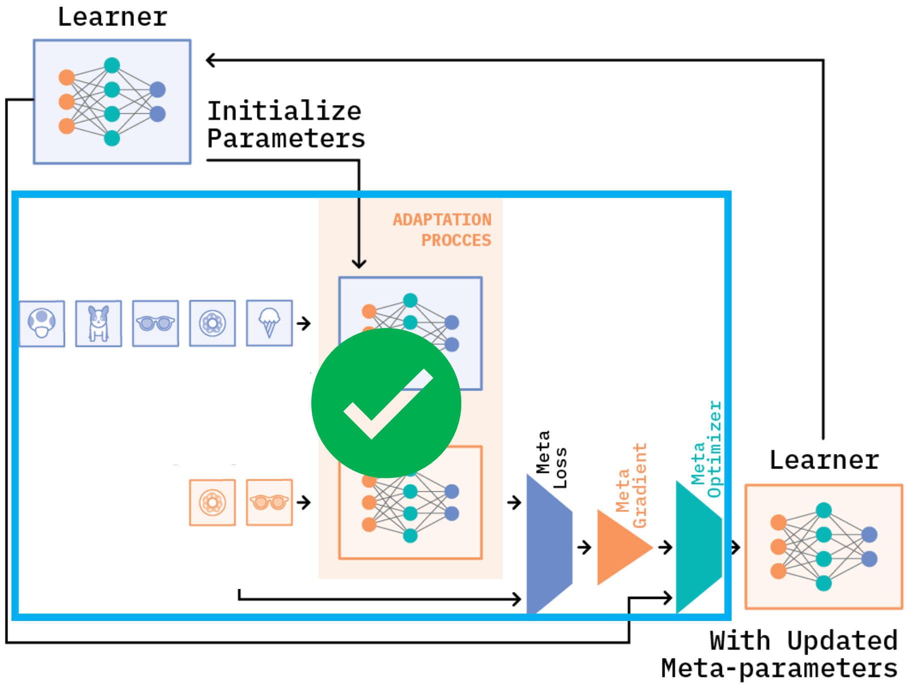


Meta Learning

元学习模式

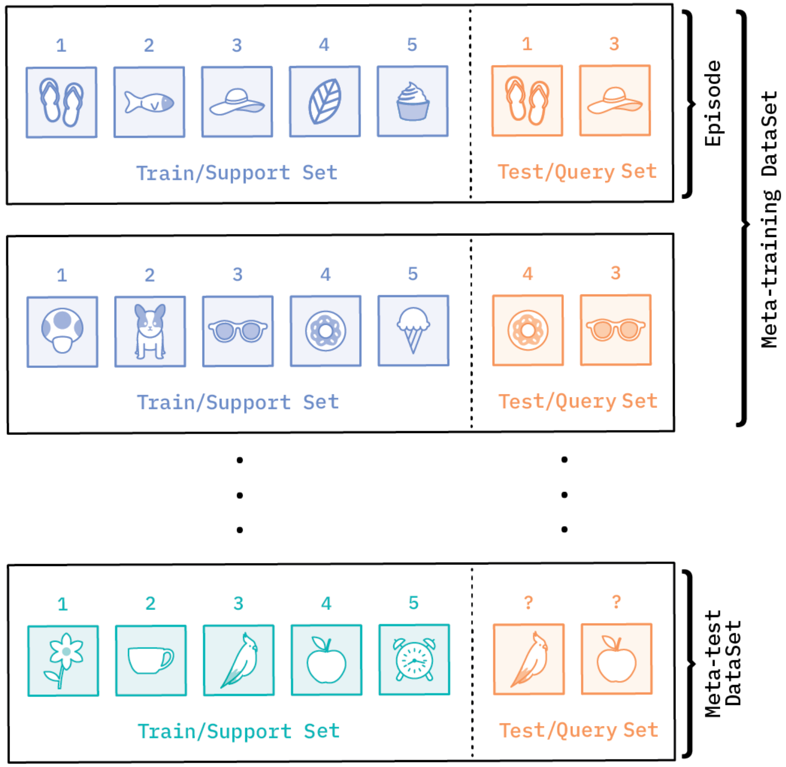
- 小样本学习与元学习存在交集
- 元学习Meta Learning——学习“如何学习”leaning to learn





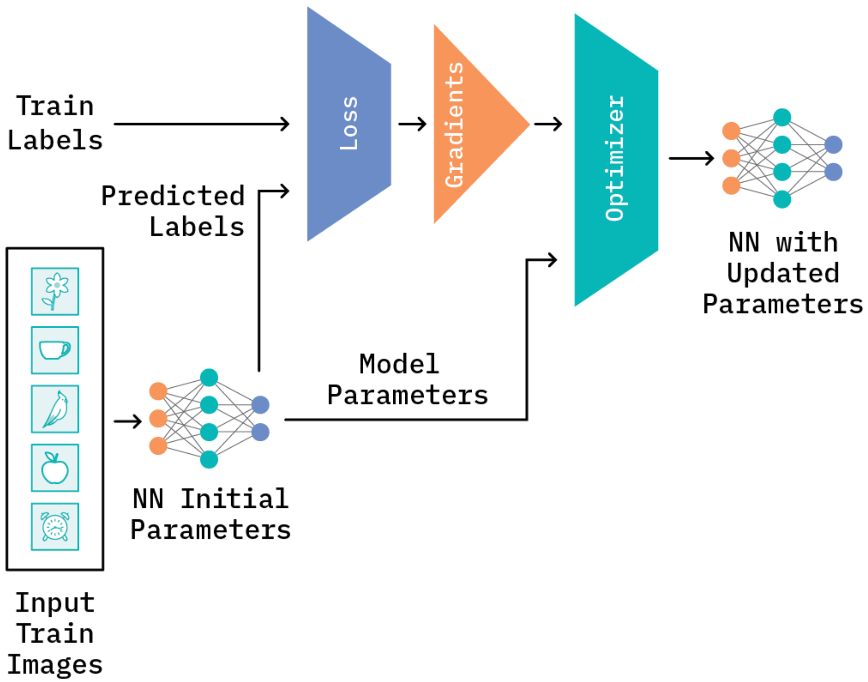
Meta Learning

元学习

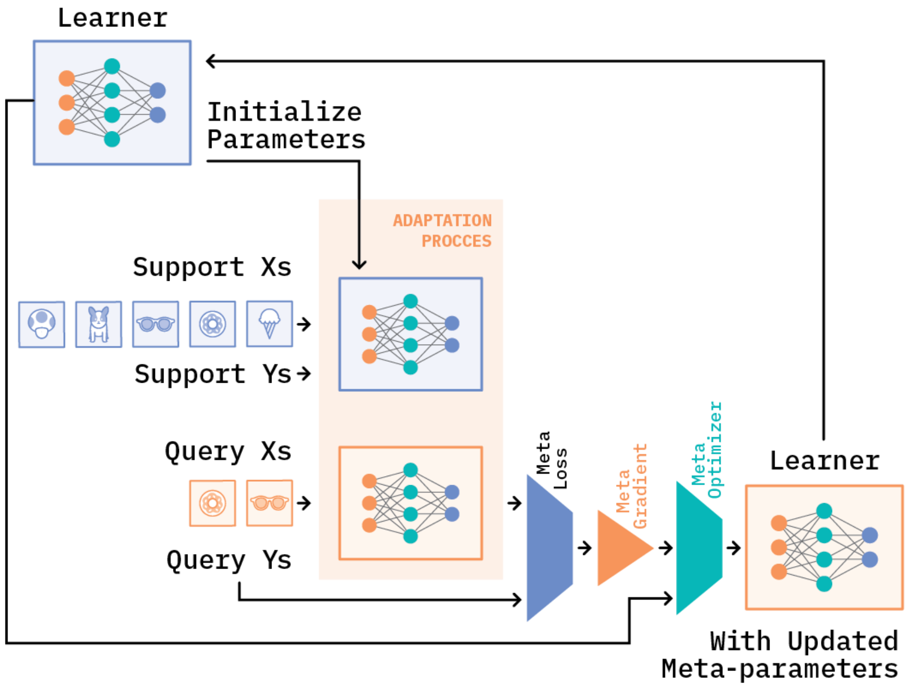


- 给定的新样本被称之为 Support set
- 相应的测试集合被称之为 Query set
- 每训练一次被称为一个 Episode
- 用来训练 F 的 Task 被称之为 Training task,验证F有效性的被称为 Testing task

对比



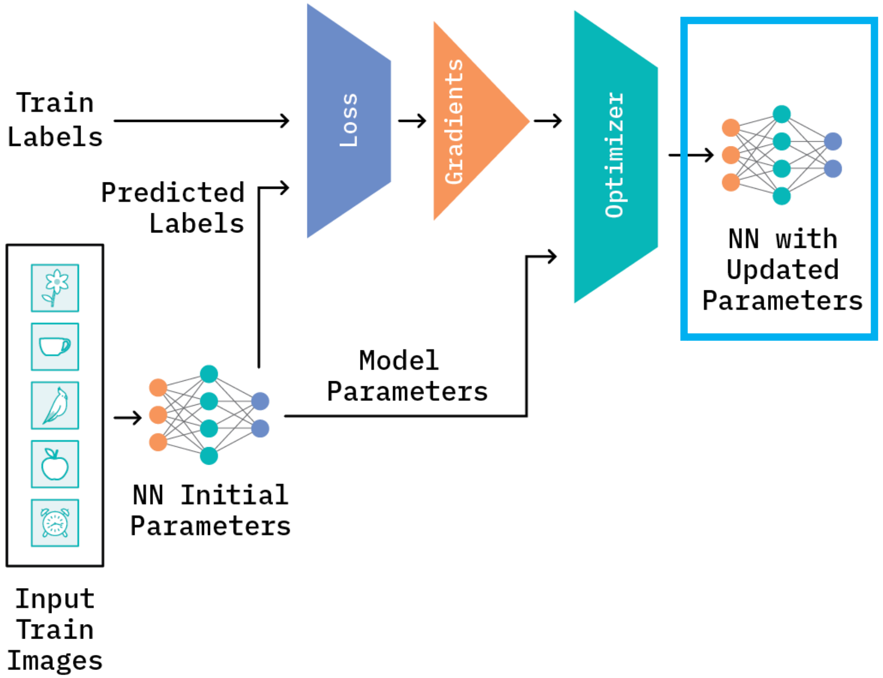
Deep Learning



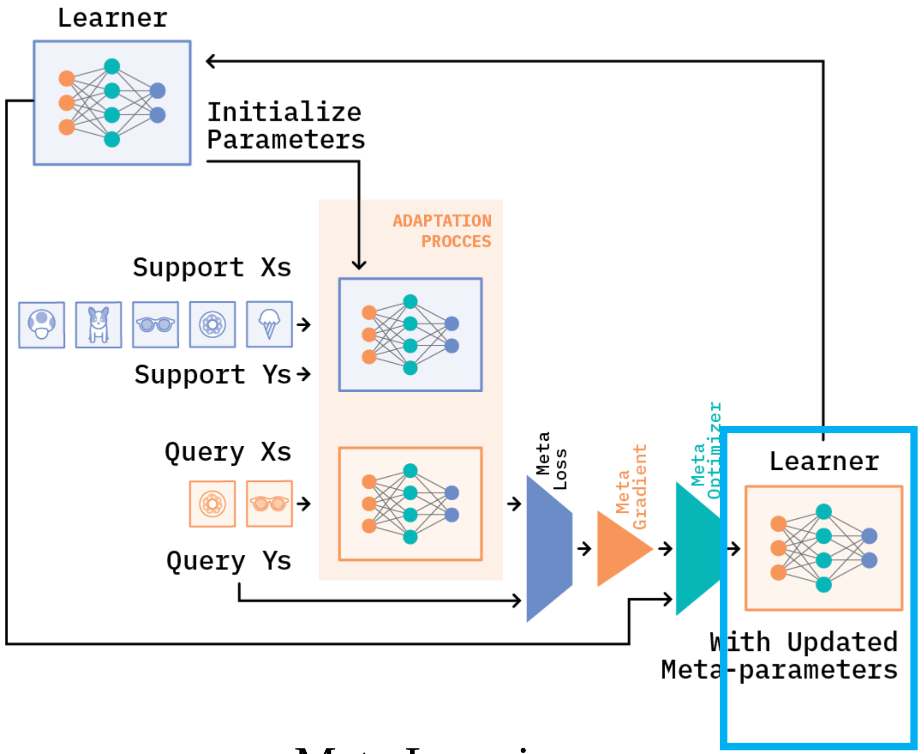
Meta Learning

https://meta-learning.fastforwardlabs.com/

对比



Deep Learning

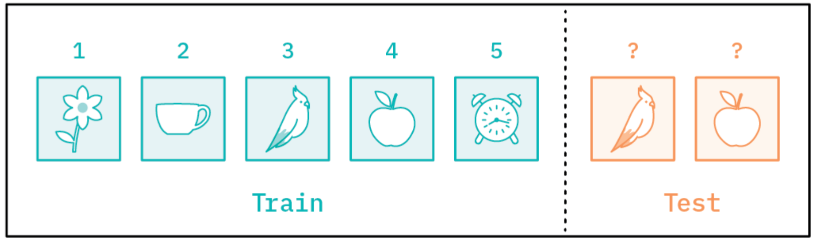


Meta Learning

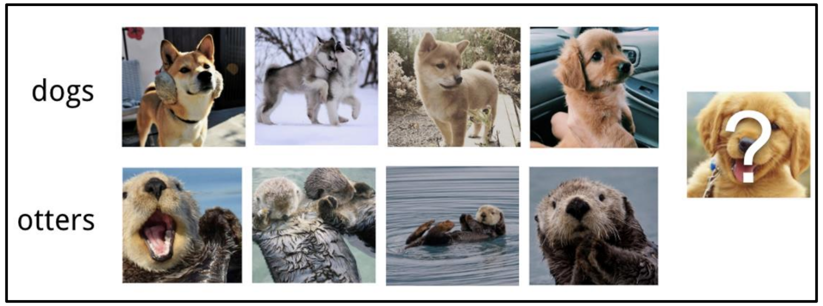
https://meta-learning.fastforwardlabs.com/

N-Way-M-Shot

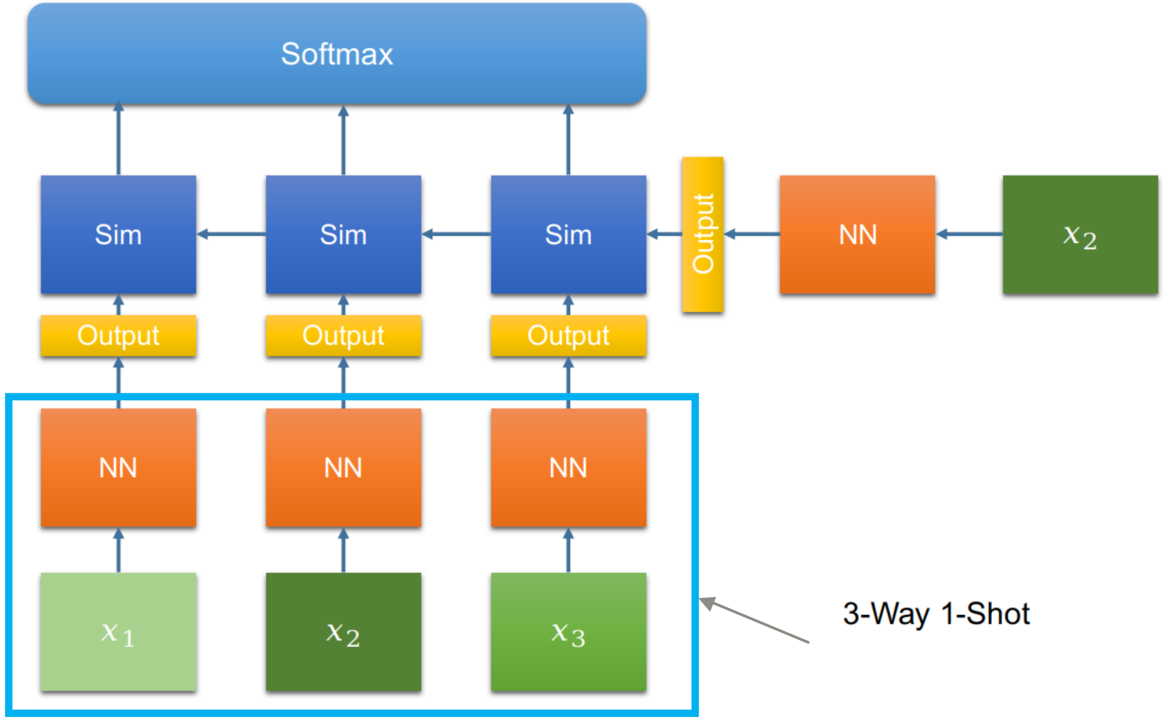
- 元学习基于一个Task 的支持集中给出的N 个新类(Way) M 个实例（shot）中学到一个认识模型F



5-Way-1-Shot



2-Way-4-Shot



- x1, x2, x3表示三个类别的数据，每类别只有一个，x2表示测试数据

元学习分类

- 基于度量的方法：Prototypical Network、 Matching Network
- 基于模型的方法：MANN、Meta Networks
- 基于优化的方法：MAML、Reptile

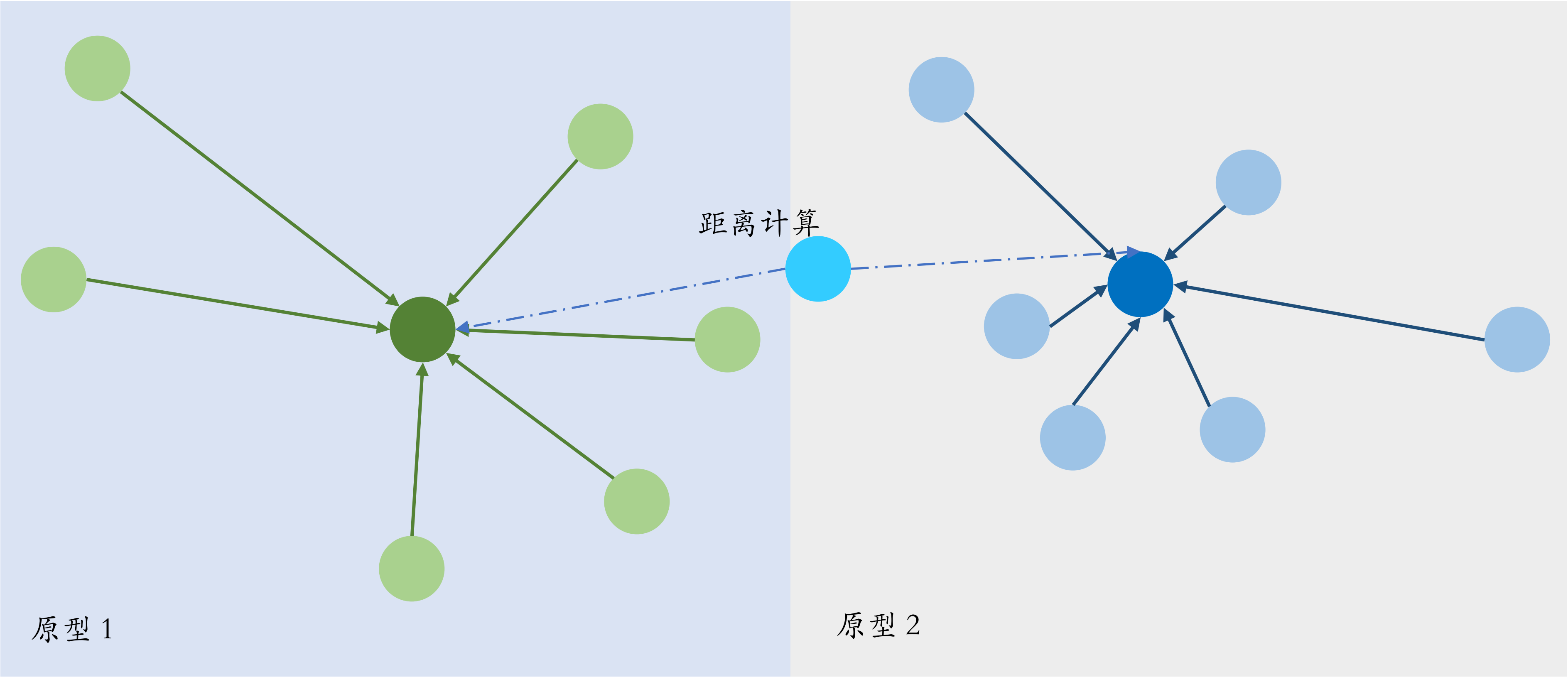
1. Chang J R, Chen Y S. Pyramid stereo matching network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5410-5418.
2. Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning[J]. arXiv preprint arXiv:1703.05175, 2017.
3. Santoro A, Bartunov S, Botvinick M, et al. Meta-learning with memory-augmented neural networks[C]//International conference on machine learning. PMLR, 2016: 1842-1850.
4. Munkhdalai T, Yu H. Meta networks[C]//International Conference on Machine Learning. PMLR, 2017: 2554-2563.
5. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning. PMLR, 2017: 1126-1135.
6. Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms[J]. arXiv preprint arXiv:1803.02999, 2018.

元学习分类

- 基于度量的方法：**Prototypical Network**、Matching Network
- 基于模型的方法：MANN、Meta Networks
- 基于优化的方法：MAML、Reptile

1. Chang J R, Chen Y S. Pyramid stereo matching network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5410-5418.
2. Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning[J]. arXiv preprint arXiv:1703.05175, 2017.
3. Santoro A, Bartunov S, Botvinick M, et al. Meta-learning with memory-augmented neural networks[C]//International conference on machine learning. PMLR, 2016: 1842-1850.
4. Munkhdalai T, Yu H. Meta networks[C]//International Conference on Machine Learning. PMLR, 2017: 2554-2563.
5. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning. PMLR, 2017: 1126-1135.
6. Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms[J]. arXiv preprint arXiv:1803.02999, 2018.

原型网络



对于一个类别k而言，所有属于k类别的数据的集合记作 S_k

Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 4080-4090.

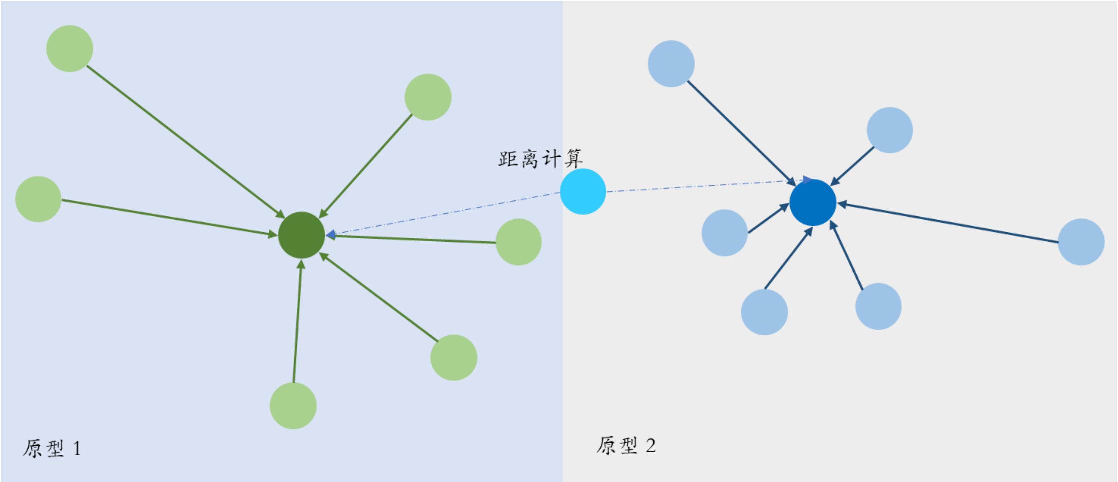
算法细节

原型中心点：

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i)$$

x属于原型k的概率，只要求$d(x)$可导：

$$p_\phi(y=k|x) = \frac{\exp(-d(f_\phi(x), c_k))}{\sum_{k'} \exp(-d(f_\phi(x), c_{k'}))}$$



https://blog.waynehfut.com/2020/11/02/prototypical_network_for_few_shot_learning/

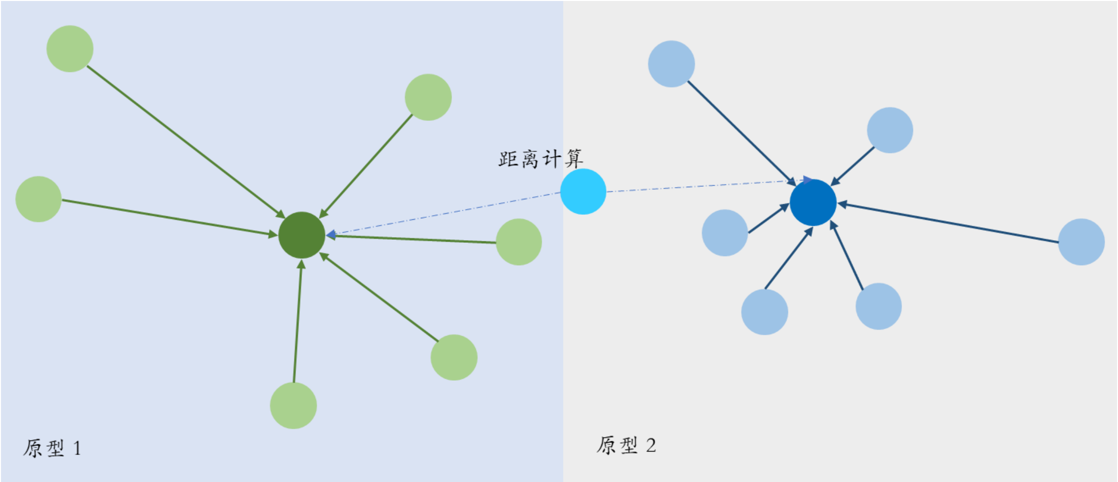
算法细节

原型中心点:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} \boxed{f_\phi(x_i)}$$

x属于原型k的概率，只要求$\boxed{d(x)}$可导:

$$p_\phi(y = k|x) = \frac{\exp(-d(f_\phi(x), c_k))}{\sum_{k'} \exp(-d(f_\phi(x), c_{k'}))}$$



https://blog.waynehfut.com/2020/11/02/prototypical_network_for_few_shot_learning/

算法细节

Algorithm 1 Training episode loss computation for Prototypical Networks. N is the number of examples in the training set, K is the number of classes in the training set, $N_C \leq K$ is the number of classes per episode, N_S is the number of support examples per class, N_Q is the number of query examples per class. $\textsc{RandomSample}(S, N)$ denotes a set of N elements chosen uniformly at random from set S, without replacement.

Input: Training set $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, where each $y_i \in \{1, \ldots, K\}$. $\mathcal{D}_k$ denotes the subset of $\mathcal{D}$ containing all elements $(\mathbf{x}_i, y_i)$ such that $y_i = k$.

Output: The loss J for a randomly generated training episode.

```
V ← RandomSample({1, ..., K}, N_C)                          ▷ Select class indices for episode
for k in {1, ..., N_C} do
    S_k ← RandomSample(D_{V_k}, N_S)                          ▷ Select support examples
    Q_k ← RandomSample(D_{V_k} \ S_k, N_Q)                    ▷ Select query examples
    c_k ← (1/N_S) Σ_{(x_i,y_i)∈S_k} f_φ(x_i)                  ▷ Compute prototype from support examples
end for
J ← 0                                                        ▷ Initialize loss
for k in {1, ..., N_C} do
    for (x, y) in Q_k do
        J ← J + 1/(N_C N_Q) [d(f_φ(x), c_k)) + log Σ_{k'} exp(−d(f_φ(x), c_{k'}))]   ▷ Update loss
    end for
end for
```

$$\mathbf{c}_k \leftarrow \frac{1}{N_S} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i)$$

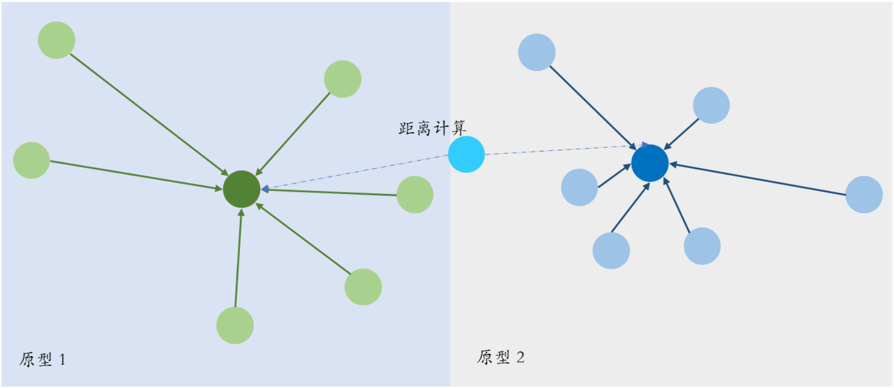
$$J \leftarrow J + \frac{1}{N_C N_Q}\left[d(f_\phi(\mathbf{x}), \mathbf{c}_k)) + \log \sum_{k'} \exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_{k'}))\right]$$

算法执行时，对于每个类别而言首先会为当前 episode 从 $\{1,\ldots,K\}$ 选择 N_C 个类，称之为 N-Ways。接着在选定的 N_C 个类别中，分别遍历每个类别，对于特定的类别 k，支持集 S_k 是从当前 episode 的数据集中所有类别为 k 的数据中选取 N_S 个数据，类别 k 下有 n 个数据，称之为 N-Shot。接着在当前 eposide 剩余的数据（$D_{V_k} \setminus S_k$）中选取 N_Q 个数据作为 Query Set。

算法细节

当前支持集所对应的原型：

$$c_k \leftarrow \frac{1}{N_S} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i)$$



该类所有对于Q_k数据的损失：

$$J \leftarrow J + \frac{1}{N_C N_Q}[d(f_\phi(x), c_k) + \log \sum_{k'} \exp(-d(f_{\phi(x)}, c_{k'}))]$$

推导：

$$J_i^k = -\log(\frac{\exp(-d(f_\phi(x_i), c_k))}{\sum_{j=1}^{N_Q} \exp(-df_\phi(x_i), c_j)}) = d(f_{\phi(x_i)}, c_k) + \log \sum_{j=1}^{N_Q} \exp(-d(f_\phi(x_i), c_j))$$

原理

- 原型网络实际上是数据分布密度估计
- 度量两个不同分布的距离

$$d_\varphi(z,z')=\varphi(z)-\varphi(z')-(z-z')\nabla\varphi(z')$$ ，$\varphi(z)$ 必须是凸函数

- 已有文献证明原型的最小距离是Bergman散度

$$p_\psi(z|\theta)=\exp\{z^T\theta-\psi(\theta)-g_\psi(z)\}=\exp\{-d_\varphi(z,\mu(\theta))-g_\varphi(z)\}$$

- 对于给定的指数族分布 $\Gamma=\{\theta_k,\pi_k\}_{k=1}^K$：

$$p(z\,|\,\Gamma)=\sum_{k=1}^{K}\pi_k p_\psi(z\,|\,\theta_k)=\sum_{k=1}^{K}\pi_k \exp(-d_\varphi(z,\mu(\theta_k))-g_\varphi(z))$$

- 对于给定的 Γ，对于张量z属于类别k的概率可记作：

$$p(y=k\,|\,z)=\frac{\pi_k\exp(-d_\varphi(z,\mu(\theta_k)))}{\sum_{k'}\pi_{k'}\exp(-d_\varphi(z,\mu(\theta_k)))}$$

覃含章，如何理解Bregman divergence? https://www.zhihu.com/question/22426561/answer/209945856
Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. Journal of Machine Learning Research, 6(Oct):1705–1749, 2005.

原理

- 对于距离度量
- 使用欧氏距离时:

$$d(z,z')=\|z-z'\|$$

$$-d(f_\phi(x),c_k) \longrightarrow -\|f_\phi(x)-c_k\|^2=-f_\phi(x)^\top f_\phi(x)+2c_k^\top f_\phi(x)-c_k^\top c_k$$

$$2c_k^\top f_\phi(x)-c_k^\top c_k=w_k^\top f_\phi(x)+b_k \quad , w_k=2c_k,\ b_k=-c_k^\top c_k$$

- 元学习的就是一个图嵌入

原理

- 对于距离度量
- 使用欧氏距离时:

$$d(z,z')=\|z-z'\|$$

$$-d(f_\phi(x),c_k) \longrightarrow -\|f_\phi(x)-c_k\|^2=\underbrace{\boxed{-f_\phi(x)^\top f_\phi(x)}}_{\text{无关}}+2c_k^\top f_\phi(x)-c_k^\top c_k$$

$$2c_k^\top f_\phi(x)-c_k^\top c_k=w_k^\top f_\phi(x)+b_k \quad , w_k=2c_k,\ b_k=-c_k^\top c_k$$

- 元学习的就是一个图嵌入

A Location-Sensitive Local Prototype Network For Few-Shot Medical Image Segmentation

Yu, Qinji, Kang Dang, Nima Tajbakhsh, Demetri Terzopoulos, and Xiaowei Ding

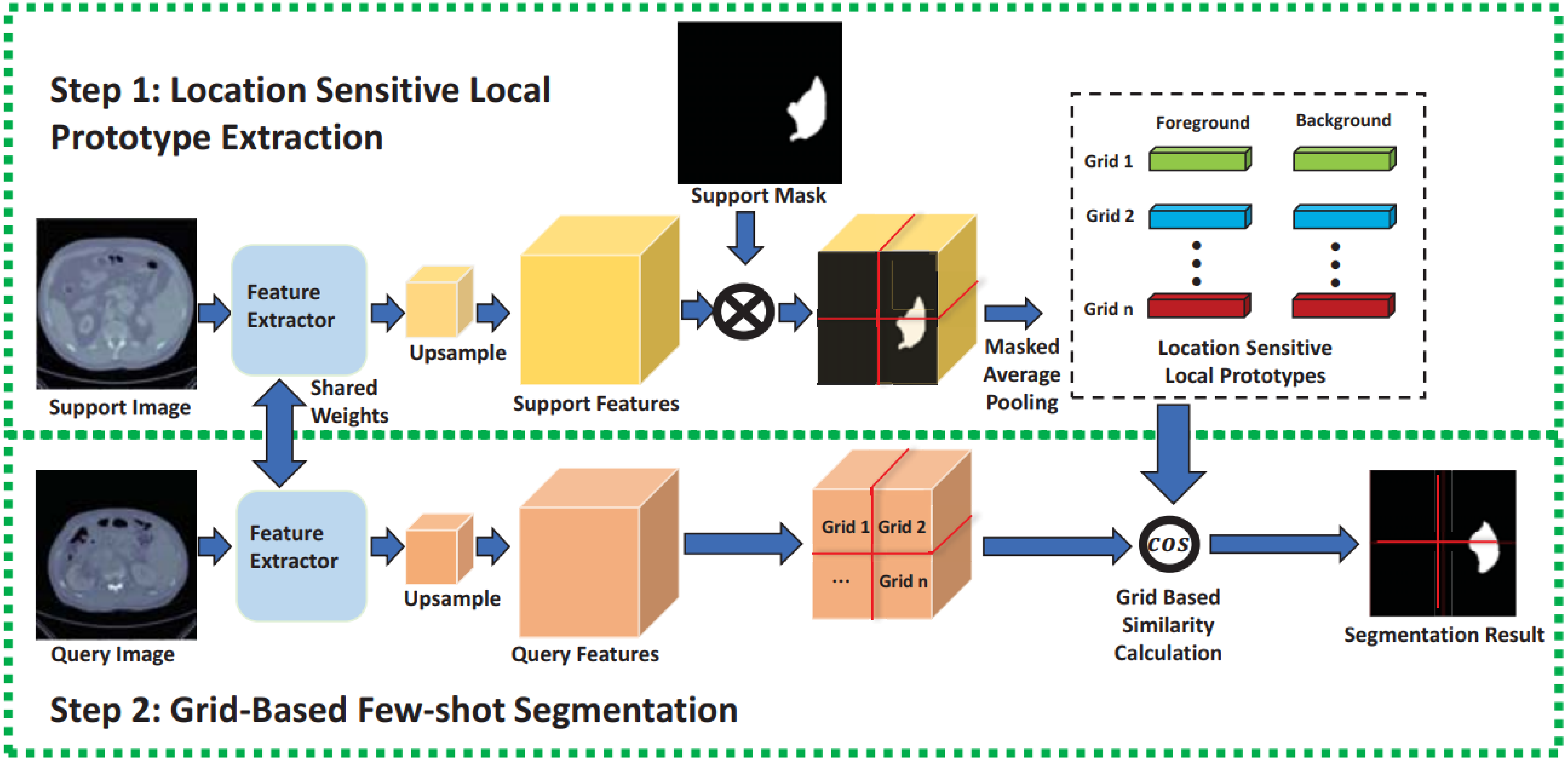
In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY
UCLA

医学影像中的小样本学习

- 典型影像数据数量少
- 精确标注成本过高
- 较少的数据辅助保护隐私

基本流程



Yu, Qinji, Kang Dang, Nima Tajbakhsh, Demetri Terzopoulos, and Xiaowei Ding. A Location-Sensitive Local Prototype Network For Few-Shot Medical Image Segmentation. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 262 – 66, 2021.

局部信息敏感的原型提取

- 原型提取:

$$p_{c,g_m}^{s}=\frac{1}{k}\sum_{i=1}^{k}\frac{\sum\limits_{(x,y)\in g_m}F_i^s(x,y)M_{i,c}^s(x,y)}{\sum\limits_{(x,y)\in g_m}M_{i,c}^s(x,y)}$$

其中$M_{i,c}^s(x,y)$表示平均池化,$F_i^s(x,y)$表示提取器的特征图,p_{c,g_m}^s与位置相关

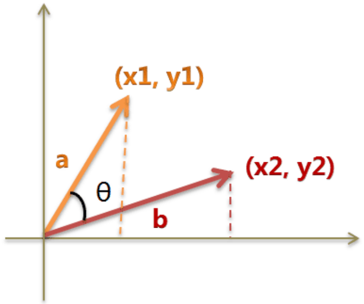
基于网格的小样本分割

- 概率计算:

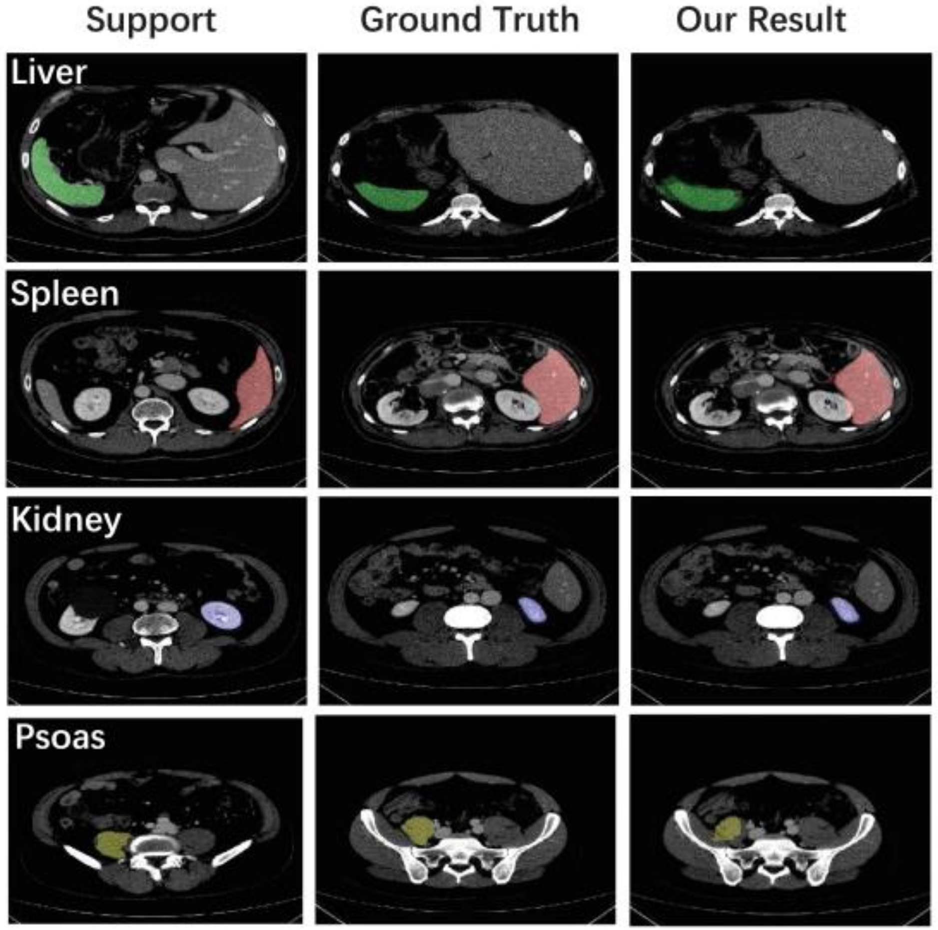
$$P_{j,c}^{q}(x,y)=Softmax(\mathrm{sim}(F_j^q(x,y),\mathcal{P}))$$,其中$\mathcal{P}$表示支持集的原型

- 相似度计算:

$$\mathrm{sim}(F_j^q(x,y),\mathcal{P})=\max_{g_m\in\Omega}(\cos(F_j^q(x,y),p_{c,g_m}^{s}))$$



实验效果



Method	liver	spleen	kidney	psoas	Mean
PANet [7]	59.4	24.1	23.7	15.6	30.7
SENet [10]	70.0	60.7	46.4	49.9	56.7
Our Result	77.9	71.5	67.5	49.9	66.7
w/ Added Classes	**79.3**	**73.3**	**76.5**	**52.4**	**70.3**
ALPNet [16]	78.3	70.9	72.1	-	-

1-Way-1-Shot

Reference

- Lilian Weng. Meta-Learning: Learning to Learn Fast. https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html. 2018-11-30
- Cyprien NIELLY. Few-shot Learning with Prototypical Networks. https://towardsdatascience.com/few-shot-learning-with-prototypical-networks-87949de03ccd
- Daisukelab. Prototypical Networks as a Fine Grained Classifier. https://www.kaggle.com/c/humpback-whale-identification/discussion/81085
- Hung-yi Lee. Meta Learning https://www.youtube.com/watch?v=EkAqYbpCYAc
- Chelsea Finn. Meta-Learning https://meta-learning.fastforwardlabs.com/

Q&A

Hao Wang
https://blog.waynehfut.com/2020/11/02/prototypical_network_for_few_shot_learning/