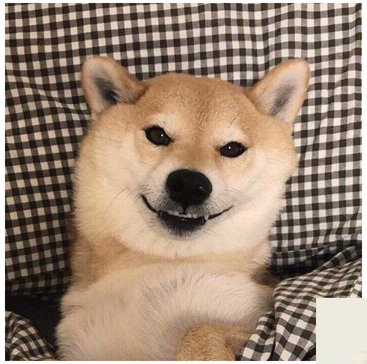
模型无关元学习

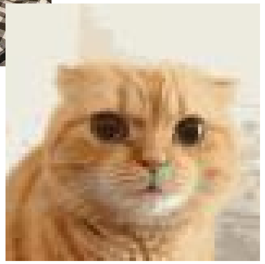
在小样本学习上的应用

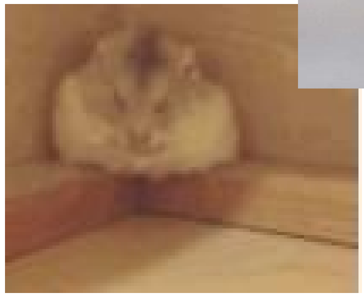
论文：Model-Agnostic Meta-Learning for Fast Adaption of Deep Network

什么是元学习？

传统的深度学习方法训练模型



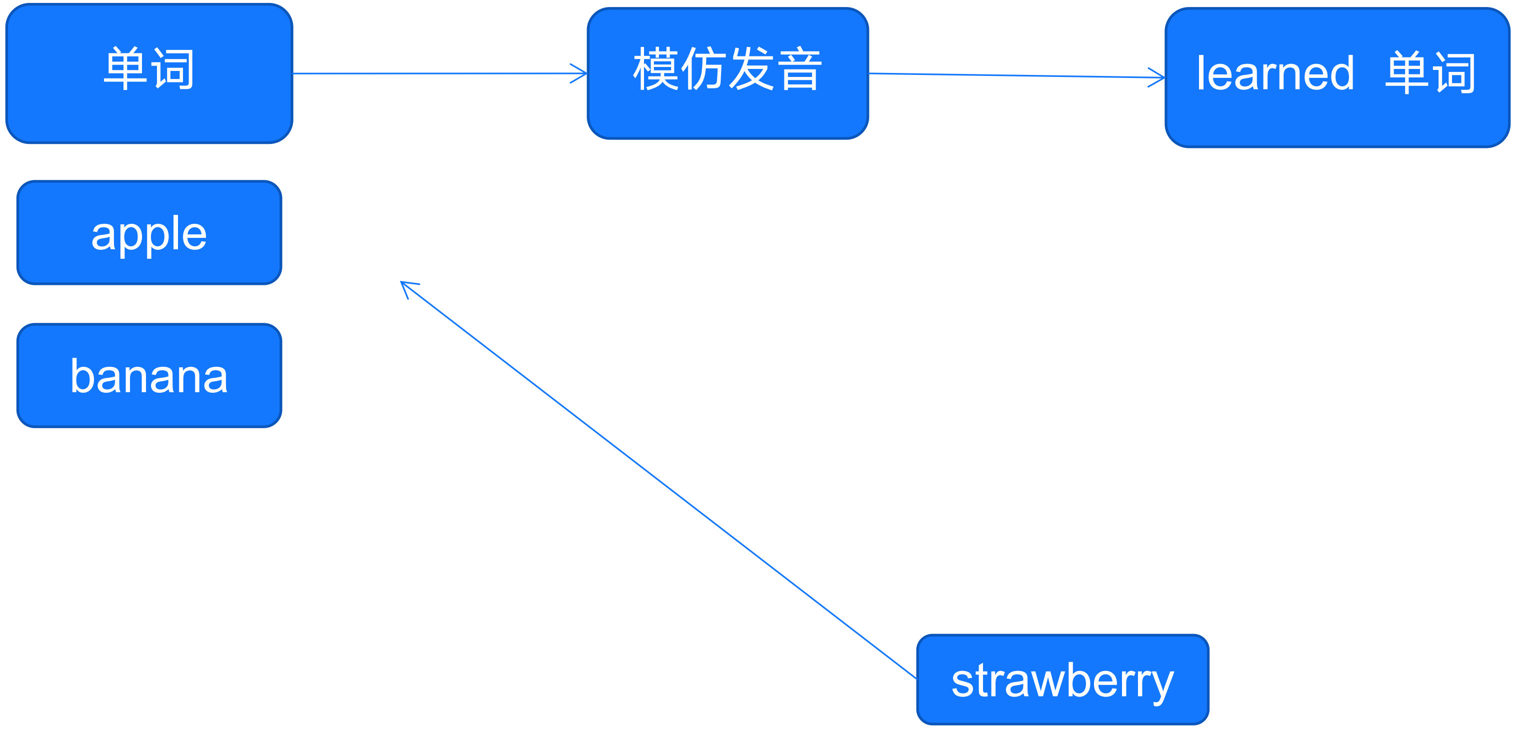


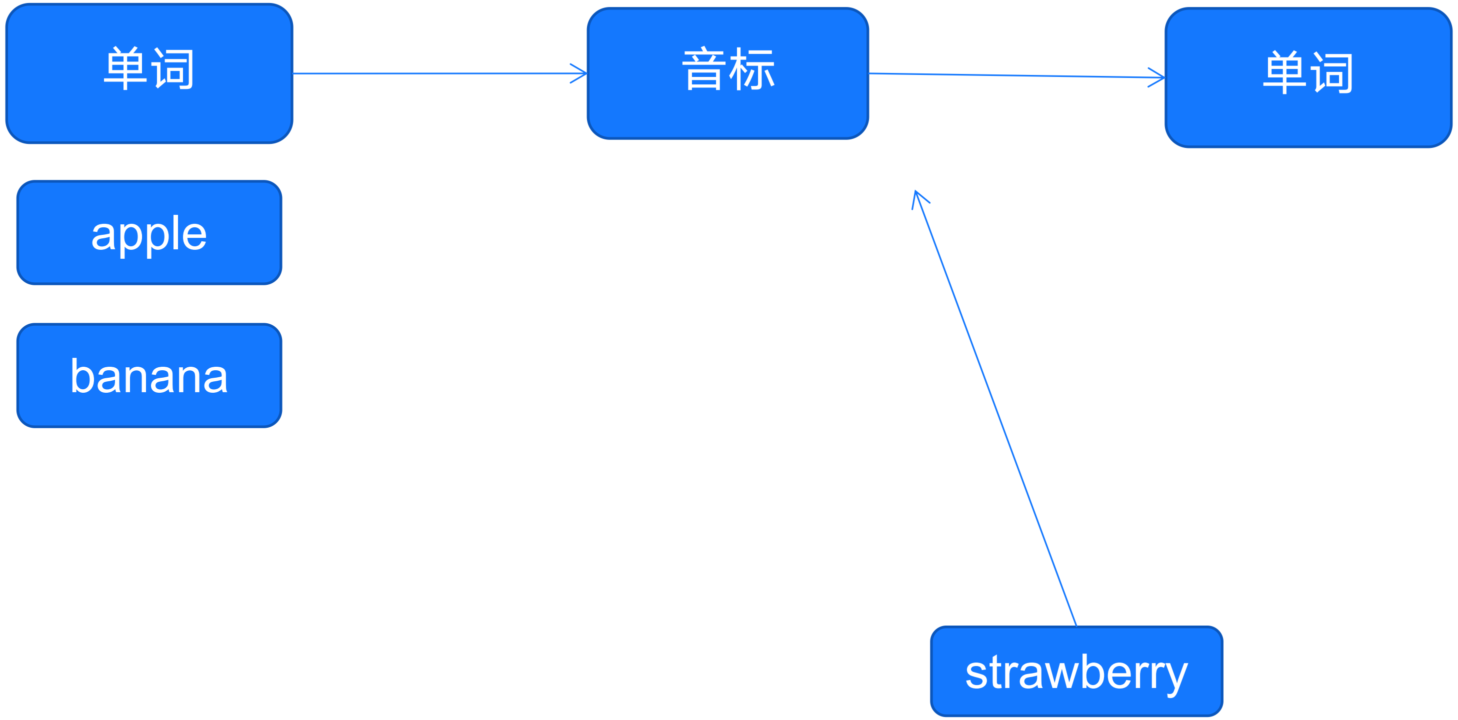


如果想获得一个会多个技能的AI系统，

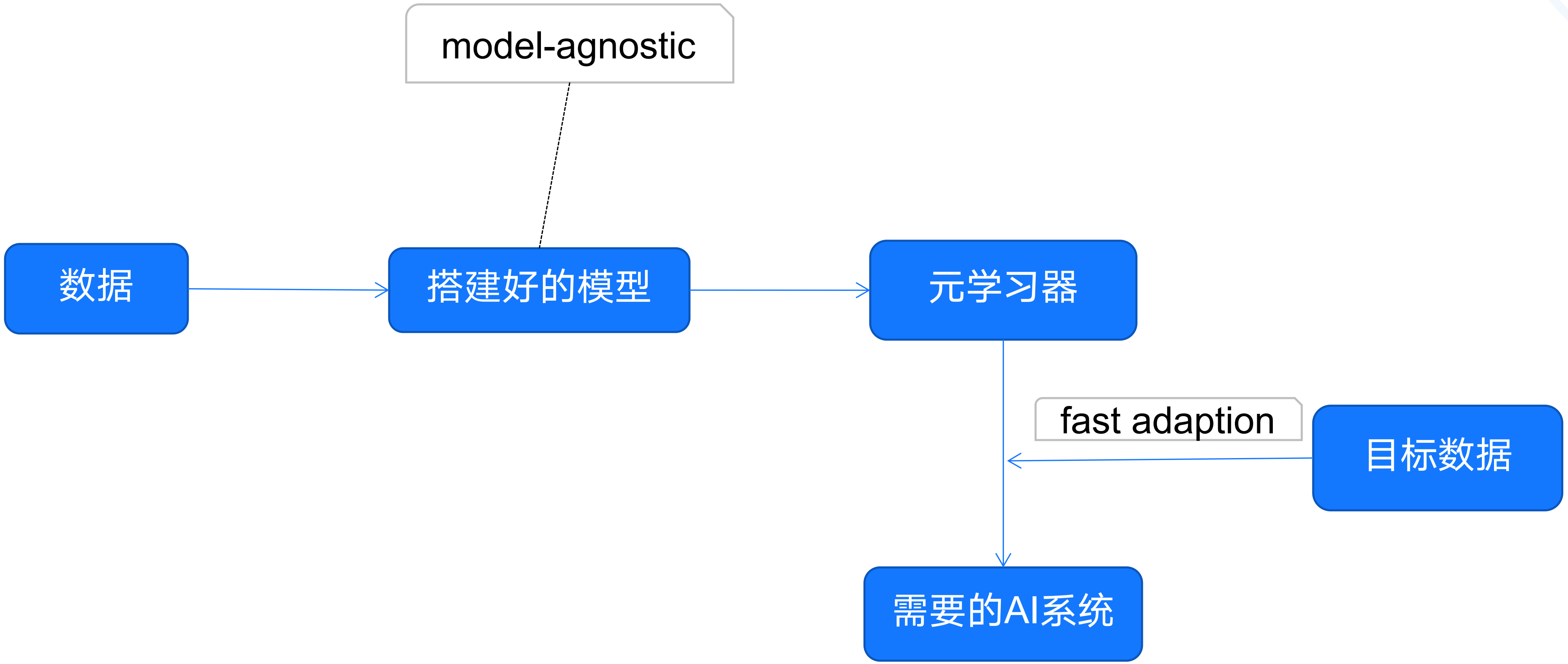
是否每个技能都要从头开始训练？

例子




单词
音标
单词
apple
banana
strawberry

训练元学习器

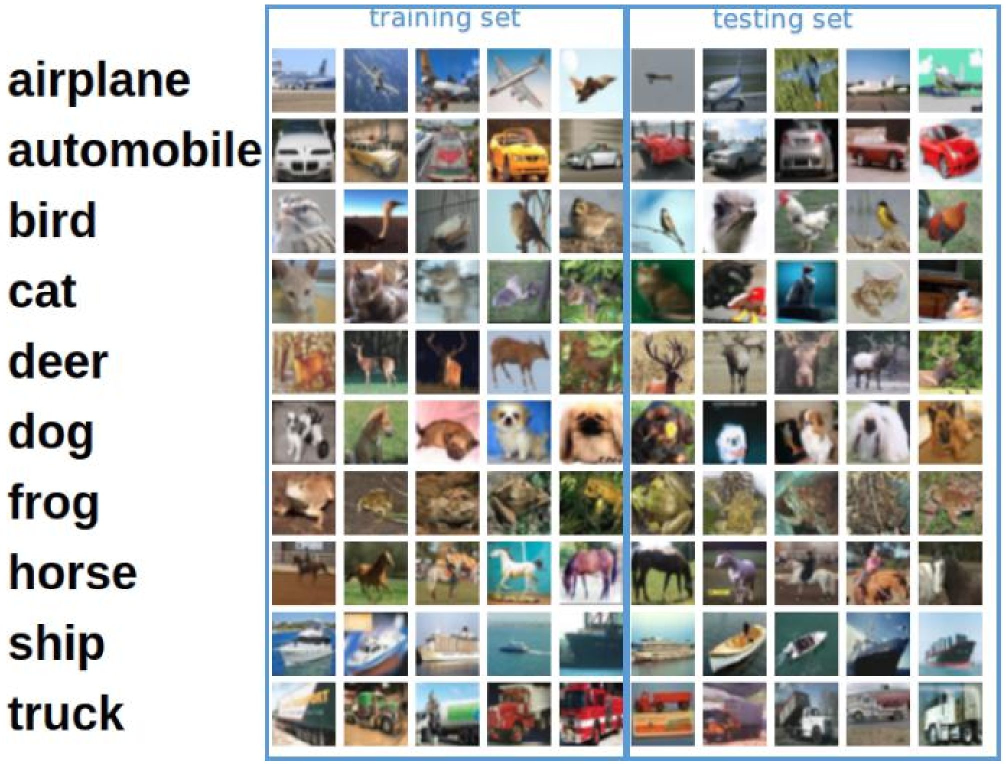


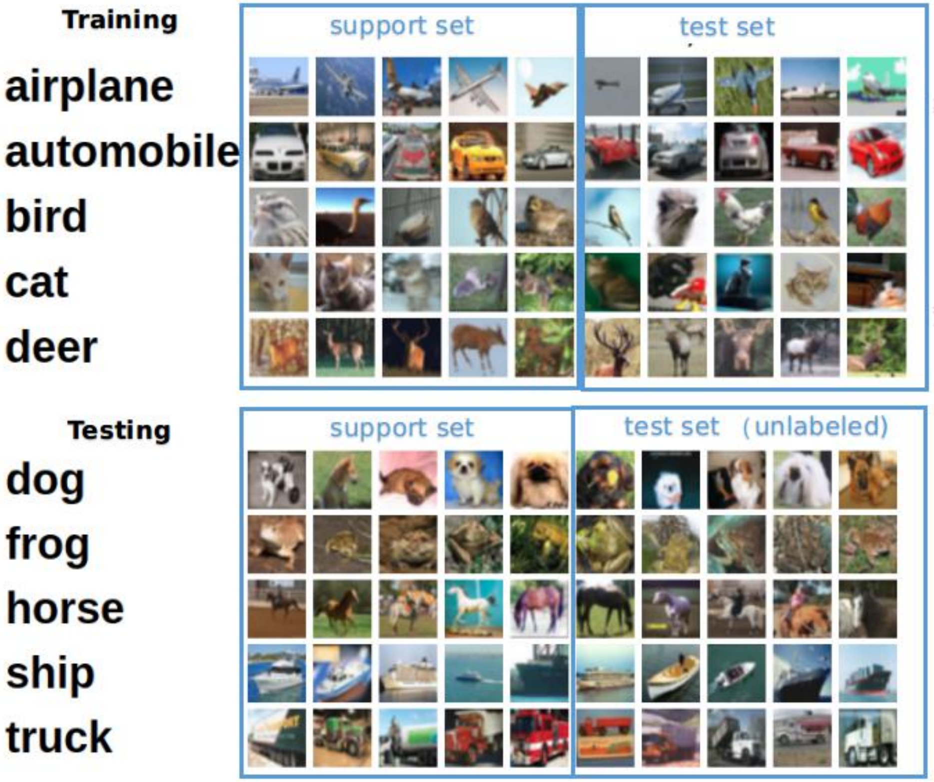
MAML：模型无关元学习

- 目标：
  - 从多个不同的学习任务中，学习到一个模型，这个模型能够快速学习如何解决一个只含有少量训练样本的新任务。
- 核心思想：
  - 寻找一个模型的初始值，使得该模型能在新任务的少量训练数据上进行快速学习，获得一个较好的效果。

什么是小样本学习？

为什么要进行小样本学习？


training set
testing set
airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck


Training
support set
test set
airplane
automobile
bird
cat
deer
Testing
support set
test set (unlabeled)
dog
frog
horse
ship
truck

N-way K-shot

- N-way 指随机抽取训练数据集中 N 个类别
- K-shot 指每个类别用于训练的标记样本数量
- 目标是要求模型从 N*K 个数据中学会如何区分这 N 个类别。

Task

- T <support set, qurey set>
  - NK shot for support set --> train set
  - NK' shot for qurey set --> test set

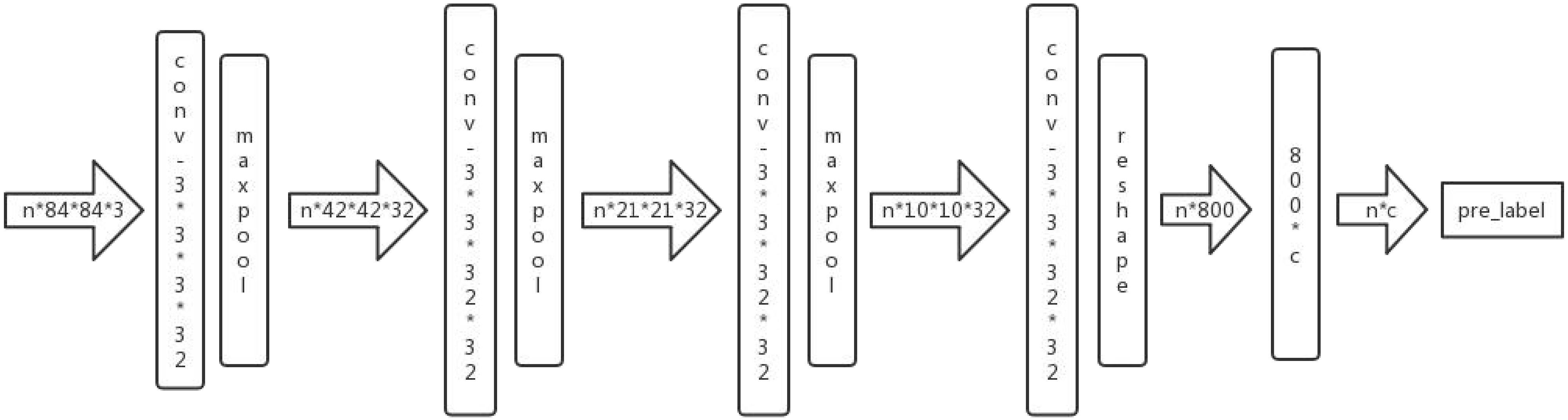
例子

- 假设场景

$M_{base-learner}$ $P_1 \sim P_5$ （20个样本）

M_{meta} $C_1 \sim C_{10}$ （30个样本）

Net



n*84*84*3
conv-3*3*3*32
maxpool
n*42*42*32
conv-3*3*32*32
maxpool
n*21*21*32
conv-3*3*32*32
maxpool
n*10*10*32
conv-3*3*32*32
reshape
n*800
800*c
n*c
pre_label

计算细节

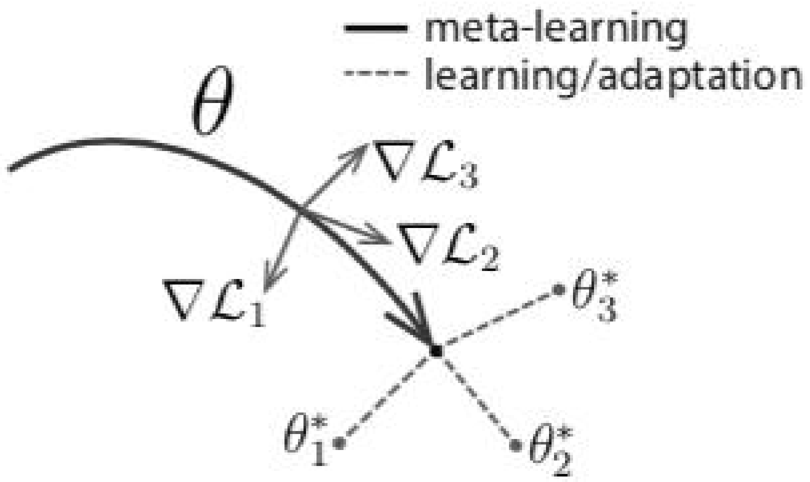
- T <support set, qurey set>

NK个样本　　NK'个样本

$$T_1 \rightarrow \theta_1' \rightarrow loss_1$$
$$T_2 \rightarrow \theta_2' \rightarrow loss_2$$
$$T_3 \rightarrow \theta_3' \rightarrow loss_3$$

θ ← tf.global_variables_initializer().run()

$\sum_{T_i} loss_i \xrightarrow{\text{梯度下降}} update\ \theta$



Algorithm

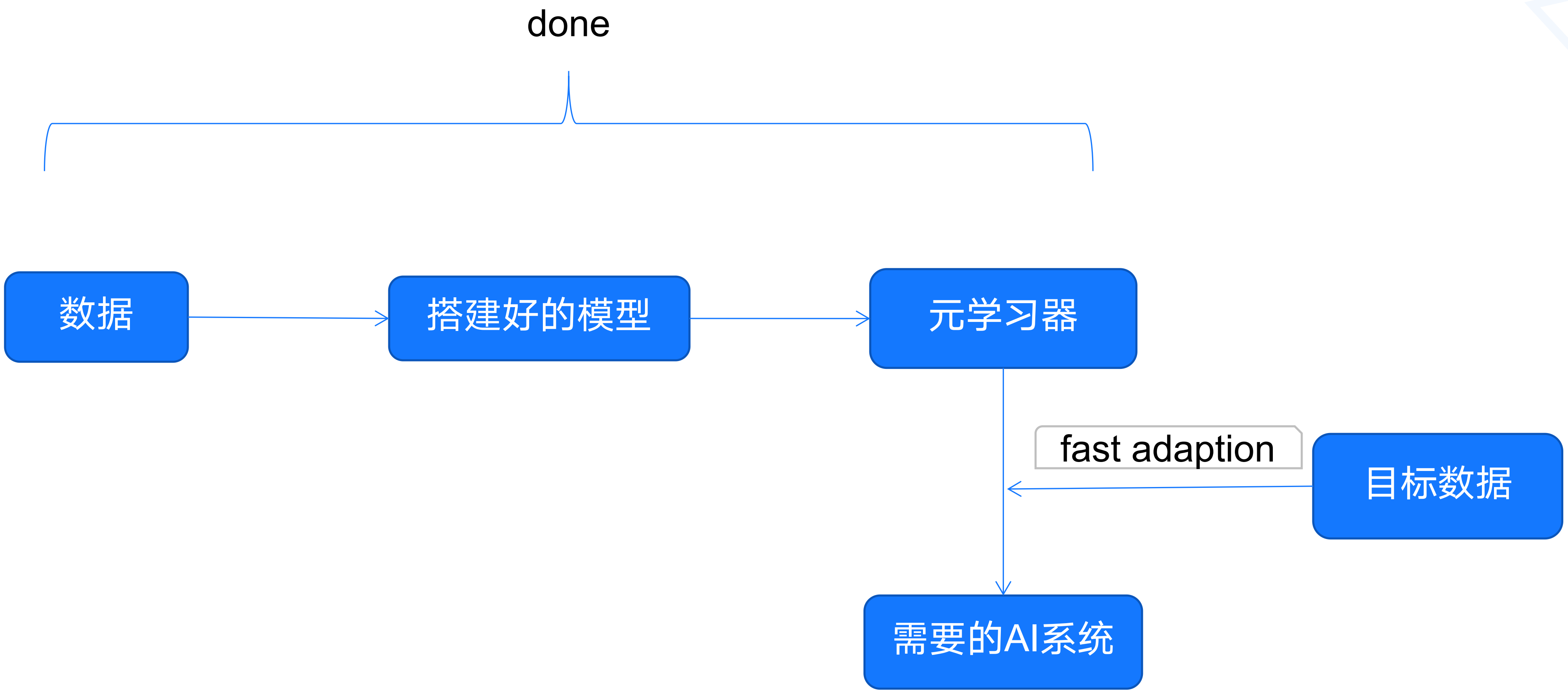
meta learner

Algorithm 2 MAML for Few-Shot Supervised Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters

randomly initialize θ
while not done **do**
  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
  for all $\mathcal{T}_i$ **do**
    Sample K datapoints $\mathcal{D} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$
    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation (2) or (3)
    Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
    Sample datapoints $\mathcal{D}_i' = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$ for the meta-update
  end for
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$ using each $\mathcal{D}_i'$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation 2 or 3
end while

训练base learner



实验

Omniglot (Lake et al., 2011)	5-way Accuracy 1-shot	5-way Accuracy 5-shot	20-way Accuracy 1-shot	20-way Accuracy 5-shot
MANN, no conv (Santoro et al., 2016)	82.8%	94.9%	–	–
MAML, no conv (ours)	**89.7 ± 1.1%**	**97.5 ± 0.6%**	–	–
Siamese nets (Koch, 2015)	97.3%	98.4%	88.2%	97.0%
matching nets (Vinyals et al., 2016)	98.1%	98.9%	93.8%	98.5%
neural statistician (Edwards & Storkey, 2017)	98.1%	99.5%	93.2%	98.1%
memory mod. (Kaiser et al., 2017)	98.4%	99.6%	95.0%	98.6%
MAML (ours)	**98.7 ± 0.4%**	**99.9 ± 0.1%**	**95.8 ± 0.3%**	**98.9 ± 0.2%**

MiniImagenet (Ravi & Larochelle, 2017)	5-way Accuracy 1-shot	5-way Accuracy 5-shot
fine-tuning baseline	28.86 ± 0.54%	49.79 ± 0.79%
nearest neighbor baseline	41.08 ± 0.70%	51.04 ± 0.65%
matching nets (Vinyals et al., 2016)	43.56 ± 0.84%	55.31 ± 0.73%
meta-learner LSTM (Ravi & Larochelle, 2017)	43.44 ± 0.77%	60.60 ± 0.71%
MAML, first order approx. (ours)	**48.07 ± 1.75%**	**63.15 ± 0.91%**
MAML (ours)	**48.70 ± 1.84%**	**63.11 ± 0.92%**

实验

- **5 way 5 shot**
- 训练 meta learner Accuracy
  - miniImageNet train classes A组support set 20%，B组 query set 72%
- 训练 base learner Accuracy
  - miniImageNet test classes 62.8%
  - cifar-10 test 48%
- **5 way 10 shot**
- 训练 meta learner Accuracy
  - miniImageNet train classes A组 support set 20%， B组 query set 78%
- 训练 base learner Accuracy
  - miniImageNet test classes 68%
  - cifar-10 test 50%

谢谢大家！