

one shot learning

zhangzihao

- One shot learning是一种对象分类问题，出现于计算机视觉领域。尽管大多数基于机器学习的对象分类算法需要对数百或数千个样本/图像和非常大的数据集进行训练，但是一次性学习旨在**从一个或仅少数训练样本/图像中学习关于对象类别的信息**。
- Few shot learning 同样是用少量训练样本训练模型。

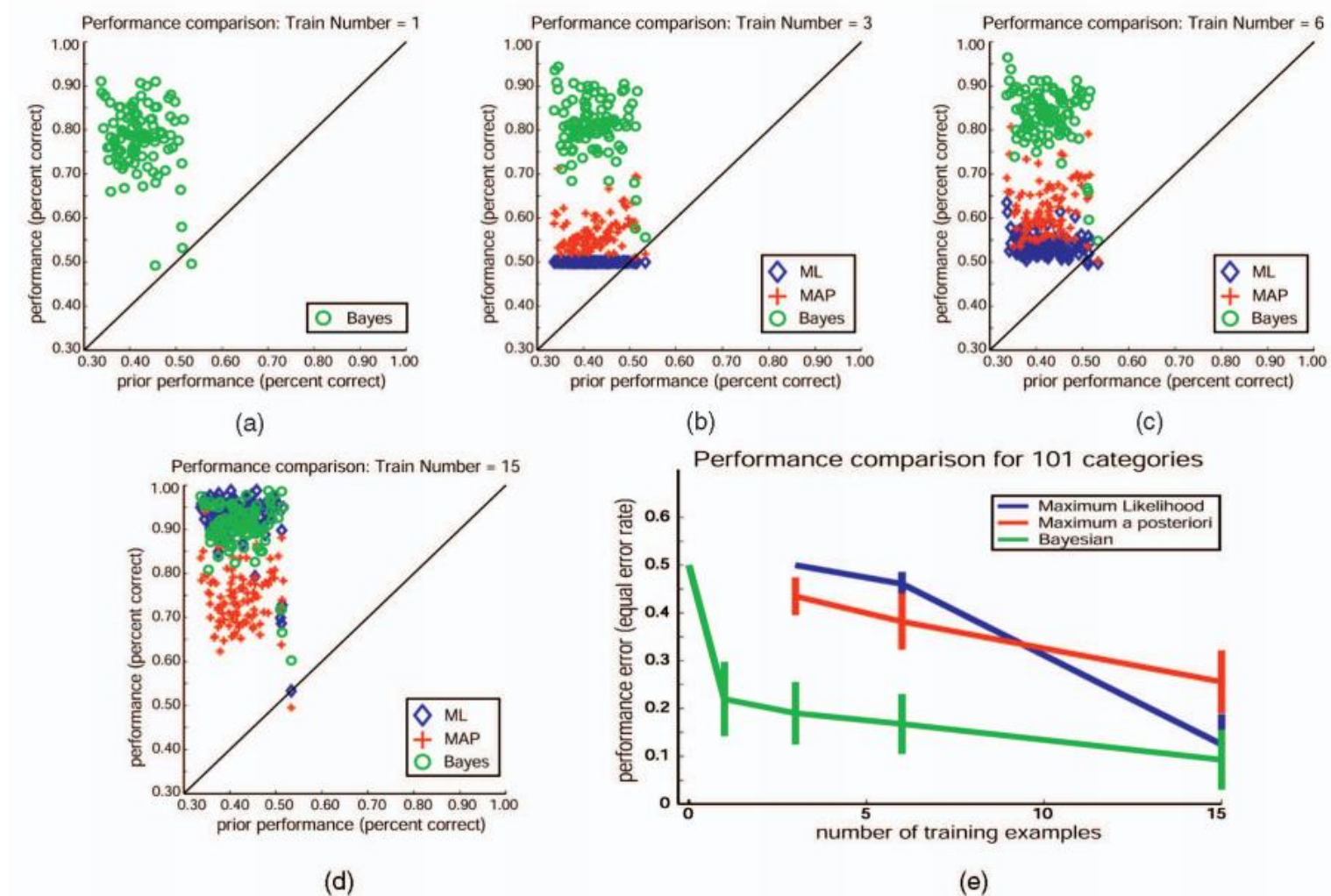
Zero shot learning

- Zero Shot Learning (零样本学习) 属于迁移学习的一个分支, 先由李飞飞在 One-Shot Learning of Object Categories TPAMI 2006 提出 One Shot Learning, 通过少量样本做目标分类, 然后由 Bengio 等在 Zero-data Learning of New Tasks AAAI 2008 提出无样本目标分类。
- 在 Zero Shot Learning 中**训练集和测试集类别无交集**。比如训练集中有马, 老虎, 测试时要预测斑马。
- Zero Shot Learning **的样本类别需要提供一些类别的描述信息**, 包括 Attribute (属性)、或者 Word Vector (词向量), 所有类别共有。简要描述 Zero Shot Learning 的流程则是: 使用训练集训练一个 CNN 模型 (在有些方法里则是直接使用 IMAGENET 上的预训练模型提取特征, 然后使用属性或词向量在训练一个网络结合特征向量进行预测, 例如在 DEM CVPR 2017 就是一个这样的过程。), 数据集的标注是属性或者词向量, 计算输出与每个类别之间的距离之后最近的则是预测的类别。

2006LiFeiFei:One-Shot Learning of Object Categories

- 从头开始学习，还不如利用以前学习过的类别中的知识，不管这些类别可能多么不同。本文讨论了该思想的贝叶斯实现。对象类别由概率模型表示。先验知识表示为这些模型的参数的概率密度函数。根据一个或多个观测值对先验进行更新，得到目标类别的后验模型。文章在101个不同类别的数据库上测试。通过贝叶斯方法学习的类别模型与通过最大似然 (ML) 和最大后验 (MAP) 方法学习的模型进行比较。我们发现在101数据集中，当训练实例数量太少而其他方法无法成功运行时，贝叶斯方法可以生成信息模型。

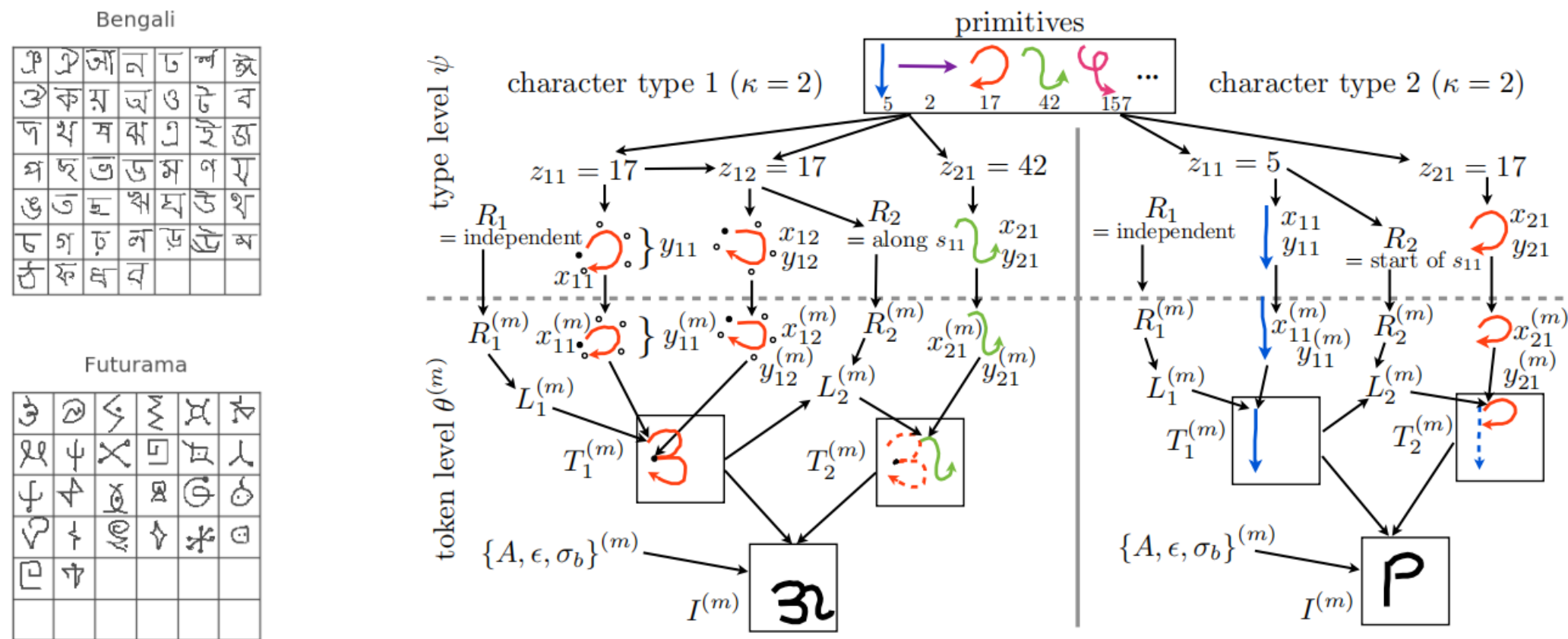
One-Shot Learning of Object Categories



2013 HBPL: One-shot learning by inverting a compositional causal process

- 提出了一个基于组合性和因果关系的分层贝叶斯模型HBPL，可以学习各种自然（虽然简单）的视觉概念，从一个图像中以类似人的方式进行推广。我们在具有挑战性的一次性分类任务中评估了性能，其中我们的模型实现了人类错误率，同时大大超过了两个深度学习模型。我们还在另一个概念性任务上测试了该模型，通过使用“视觉图灵测试”来生成新的示例，以表明我们的模型产生类似人类的表现
- 使用Omniglot数据集。有50种文字，1623类手写字符。对于每类字符仅有20个样本，每个样本分别由不同的人绘制而成，分辨率为 105×105 。HBPL达到95.2%。

2013HBPL:One-shot learning by inverting a compositional causal process



度量学习

- 1、度量(Metric)
- 在数学中，一个度量（或距离函数）是一个定义**集合中元素之间距离的函数**。一个具有度量的集合被称为度量空间。
- 2、度量学习
- K-means、K近邻方法、SVM等算法需要给定好的度量来反映数据间存在的一些重要关系。这一问题在无监督的方法（如聚类）中尤为明显。假设要计算图像之间的相似度。如果我们的目标是识别人脸，那么就需要构建一个距离函数去强化合适的特征（如发色，脸型等）；为了处理各种各样的特征相似度，我们可以在特定的任务通过选择合适的特征并手动构建距离函数。然而这种方法会需要很大的人工投入，也可能对数据的改变非常不鲁棒。**度量学习作为一个理想的替代，可以根据不同的任务来自主学习出针对某个特定任务的度量距离函数。**

度量学习CVPR2018: Learning to Compare: Relation Network for Few-Shot Learning

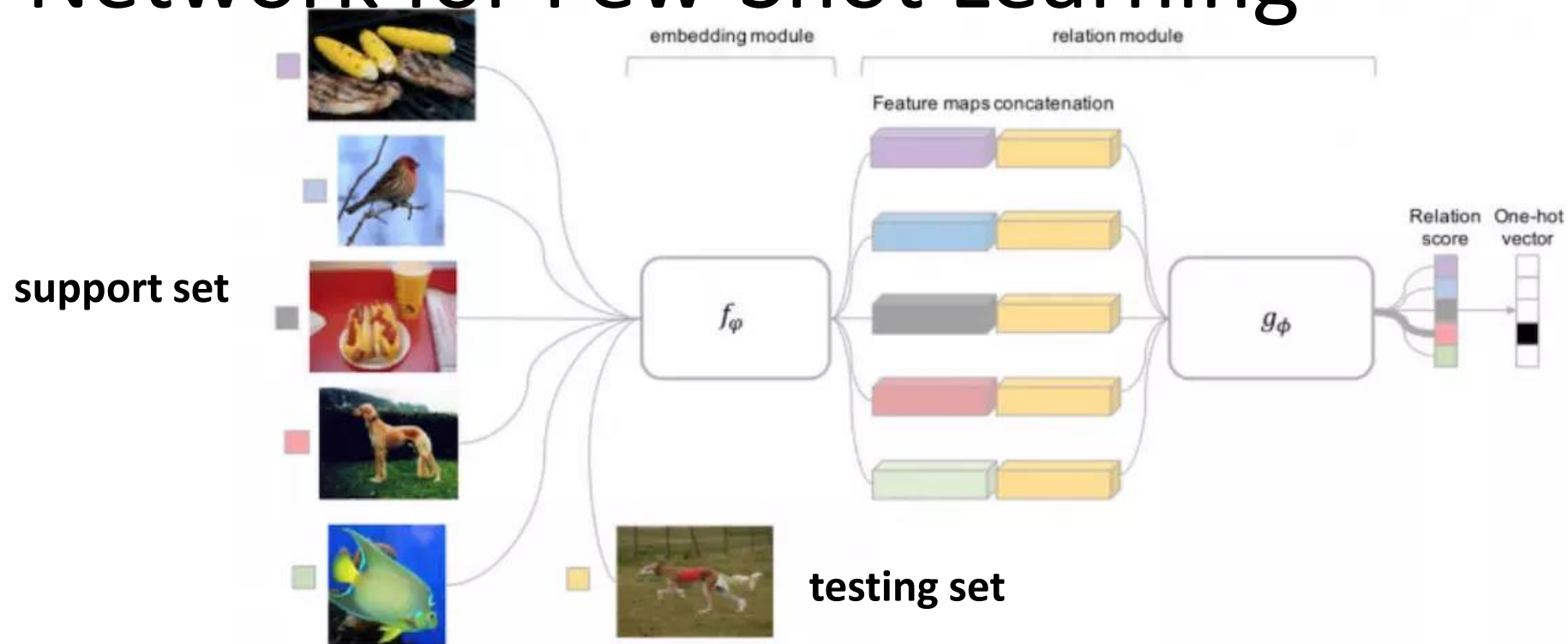
- 本文提出关系网络 (Relation Network (RN)) , 它能够学会比较, 从而实现小样本学习 (Few-Shot Learning) , 并且稍加改进, 就能够实现从少样本学习到零样本学习。该方法在五个基本测试集中都取得了很好的效果。
- 为什么人可以看到 iPhone X 一眼然后就可以认出它? 丑陋的齐刘海, 后双摄像头是竖着的。人之所以能够识别一个新的东西, 在于我们人的视觉系统天生的能够对任意物体提取特征, 并会和其他物体进行比较。
- 模型提出 (Relation Network) : 将该任务定义为小样本分类器学习, 将数据集分成 training set、support set和testing set, 其中support set和testing set共享相同的label空间, training set有自己的label空间。

<https://www.jianshu.com/p/49b02c40000a>

<https://arxiv.org/pdf/1711.06025.pdf>

https://github.com/floodsung/LearningToCompare_FSL

CVPR2018: Learning to Compare: Relation Network for Few-Shot Learning



- 整个关系网络由嵌入模块和关系模块组成，support set和testing set都需要经过嵌入模块提取特征，然后将特征联合起来输送到关系模块进行比较，然后我们根据比较的结果（relation score）来判断这个测试图片到底属于哪一个类。这很好理解，比如上图中测试图片是狗，那么它跟训练样本中狗的图片相似度比较高，那么我们就认为这个新的图片是狗。整个过程模拟人去识别一个物体的过程。

CVPR2018: Learning to Compare: Relation Network for Few-Shot Learning

少样本学习实验: Omniglot

Model	Fine Tune	5-way Acc.		20-way Acc.	
		1-shot	5-shot	1-shot	5-shot
MANN [32]	N	82.8%	94.9%	-	-
CONVOLUTIONAL SIAMESE NETS [20]	N	96.7%	98.4%	88.0%	96.5%
CONVOLUTIONAL SIAMESE NETS [20]	Y	97.3%	98.4%	88.1%	97.0%
MATCHING NETS [39]	N	98.1%	98.9%	93.8%	98.5%
MATCHING NETS [39]	Y	97.9%	98.7%	93.5%	98.7%
SIAMESE NETS WITH MEMORY [18]	N	98.4%	99.6%	95.0%	98.6%
NEURAL STATISTICIAN [8]	N	98.1%	99.5%	93.2%	98.1%
META NETS [27]	N	99.0%	-	97.0%	-
PROTOTYPICAL NETS [36]	N	98.8%	99.7%	96.0%	98.9%
MAML [10]	Y	98.7 \pm 0.4%	99.9 \pm 0.1%	95.8 \pm 0.3%	98.9 \pm 0.2%
RELATION NET	N	99.6 \pm 0.2%	99.8 \pm 0.1%	97.6 \pm 0.2%	99.1 \pm 0.1%

少样本学习实验: Minilmagenet

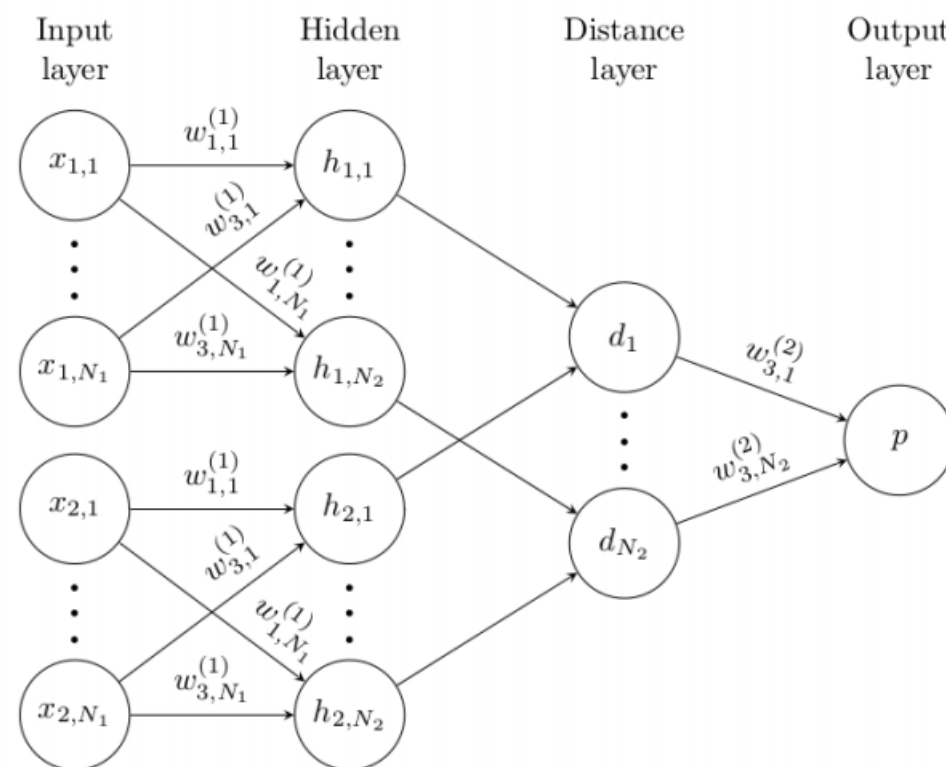
Model	FT	5-way Acc.	
		1-shot	5-shot
MATCHING NETS [39]	N	43.56 \pm 0.84%	55.31 \pm 0.73%
META NETS [27]	N	49.21 \pm 0.96%	-
META-LEARN LSTM [29]	N	43.44 \pm 0.77%	60.60 \pm 0.71%
MAML [10]	Y	48.70 \pm 1.84%	63.11 \pm 0.92%
PROTOTYPICAL NETS [36]	N	49.42 \pm 0.78%	68.20 \pm 0.66%
RELATION NET	N	50.44 \pm 0.82%	65.32 \pm 0.70%

零样本学习实验结果

Model	F	SS	AwA	CUB
			10-way 0-shot	50-way 0-shot
SJE [3]	F_G	A	66.7	50.1
ESZSL [31]	F_G	A	76.3	47.2
SSE-RELU [46]	F_V	A	76.3	30.4
JLSE [47]	F_V	A	80.5	42.1
SYNC-STRUCT [6]	F_G	A	72.9	54.5
SEC-ML [5]	F_V	A	77.3	43.3
PROTO. NETS [36]	F_G	A	-	54.6
DEVISE [11]	N_G	A/W	56.7/50.4	33.5
SOCHER <i>et al.</i> [37]	N_G	A/W	60.8/50.3	39.6
MTMDL [43]	N_G	A/W	63.7/55.3	32.3
BA <i>et al.</i> [25]	N_G	A/W	69.3/58.7	34.0
DS-SJE [30]	N_G	A/D	-	50.4/ 56.8
SAE [21]	N_G	A	84.7	61.4
DEM [45]	N_G	A/W	86.7/78.8	58.3
RELATION NET	N_G	A	84.5	62.0

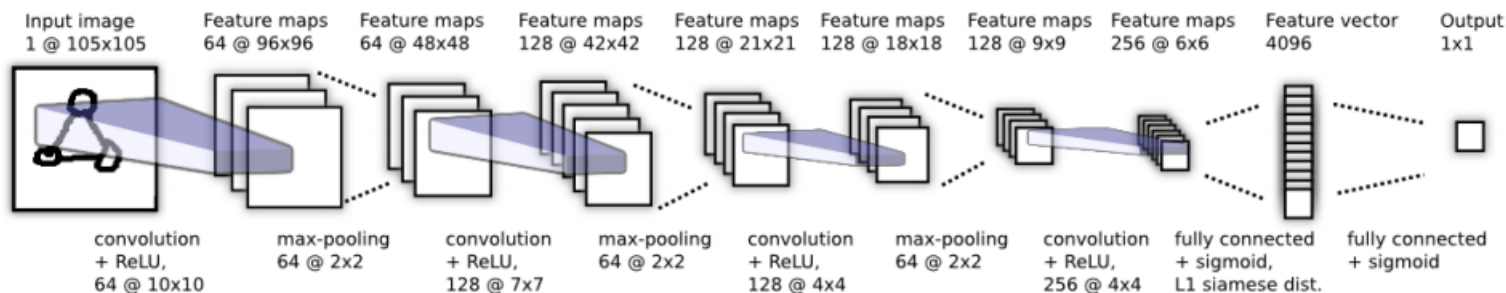
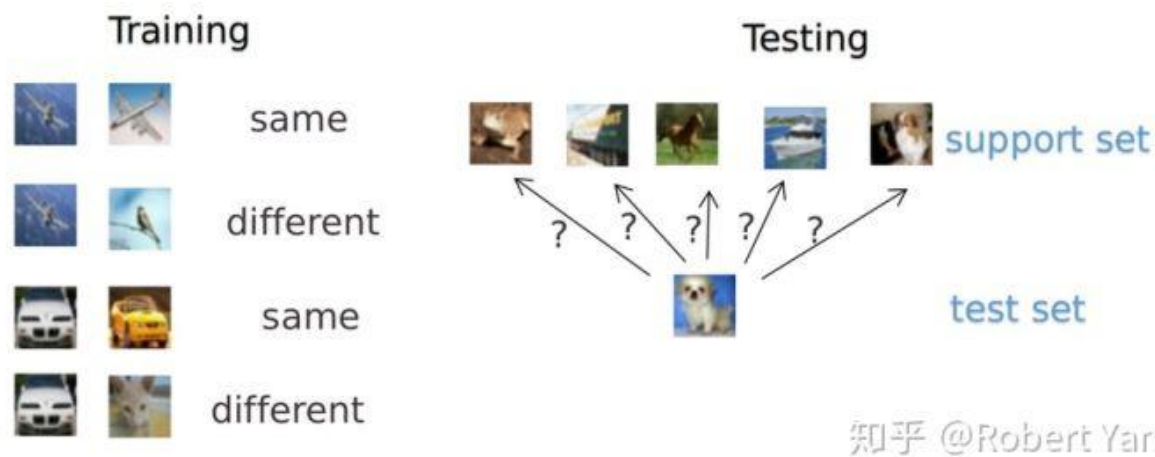
孪生网络Siamese neural networks for one-shot image recognition(2015)

- 对输入的结构进行限制并自动发现可以从新样本上泛化的特征。通过一个有监督的基于孪生网络的度量学习来训练，然后重用那个网络所提取的特征进行one/few-shot学习。
- 它是一个双路的神经网络，训练时，通过组合不同类的样本成对，同时输入网络进行训练，在最上层通过一个距离的交叉熵进行loss的计算。



孪生网络Siamese neural networks for one-shot image recognition(2015)

- 在预测的时候，以5way-5shot为例，从5个类中随机抽取5个样本，把这个mini-batch=25的数据输入网络，最后获得25个值，取分数最高对应的类别作为预测结果。
- 网络结构如图4所示，是一个8层深度卷积孪生网络，图中只展示了其中一路计算，在网络的4096维的全连接层后执行component-wise的L1距离计算，产生一个4096维的特征向量，并通过sigmoidal激活获得一个0到1的概率作为两个输入样本是否相似的结果。



孪生网络Siamese neural networks for one-shot image recognition(2015)

- 列表中还包含一个只包括两个全连层的非卷积神经网络。我们卷积模型达到92%的准确率，比HBPL差点。但是我们的模型没有关于字符和笔画的任何先验知识，这是模型的主要优势。

- 直接使用Omniglot数据集上的训练模型测试MNIST的分类效果

Table 2. Comparing best one-shot accuracy from each type of network against baselines.

Method	Test
Humans	95.5
Hierarchical Bayesian Program Learning	95.2
Affine model	81.8
Hierarchical Deep	65.2
Deep Boltzmann Machine	62.0
Simple Stroke	35.2
1-Nearest Neighbor	21.7
Siamese Neural Net	58.3
Convolutional Siamese Net	92.0

Table 3. Results from MNIST 10-versus-1 one-shot classification task.

Method	Test
1-Nearest Neighbor	26.5
Convolutional Siamese Net	70.3

匹配网络Matching Networks for One Shot Learning(2016)

- 本文受到利用深度神经网络来学习图像、文本特征的启发，通过拓展的外部信息加强记忆神经网络的效果，提出了一种针对one-shot学习的神经网络模型-Matching Net (MN).该模型可以将一个小的标注集合以及一个未标注的测试样例映射到它对应的标签，在这个过程中避免了对于新的标签类别进行调整的需求。
- 该方法在ImageNet数据集的one-shot准确率从87.6%提升到了93.2%，在Omniglot数据集上的效果从88.0%提升到了93.8%！

匹配网络Matching Networks for One Shot Learning

Task:

训练过程: 给定一个有 k 个样本的支撑集 $S = \{(x_i, y_i)\}_{i=1}^k$, 对测试样本 \hat{x} 分类 (归类, 因为 \hat{x} 的类别是和 S 中的某个或几个样本类别相同的) , 类别是 $C_S(\hat{x})$ 。

定义 $S \rightarrow C_S(\hat{x})$ 这一映射为 $P(\hat{y}|\hat{x}, S)$, 这里的 P 的参数是通过神经网络学习到的, 映射方式就是我们最后学到的模型。

因此, 在测试过程中: 给定一个新的支撑集 S' , 我们可以用之前学到的模型对每个测试样本 \hat{x} 得到他们可能的label \hat{y} 。

比如, 在训练时给定一张暹罗猫的图片 and 一张柯基的图片作为 S , 对于一张新的哈士奇的图片模型可以将其分类为狗; 在测试时拿来一张熊二的图片和一张兔八哥的图片, 又拿来一张小熊维尼的图片问机器这个新图片是属于哪一类的, 机器就会告诉你这个熊 (跟那个熊二是一个类的)

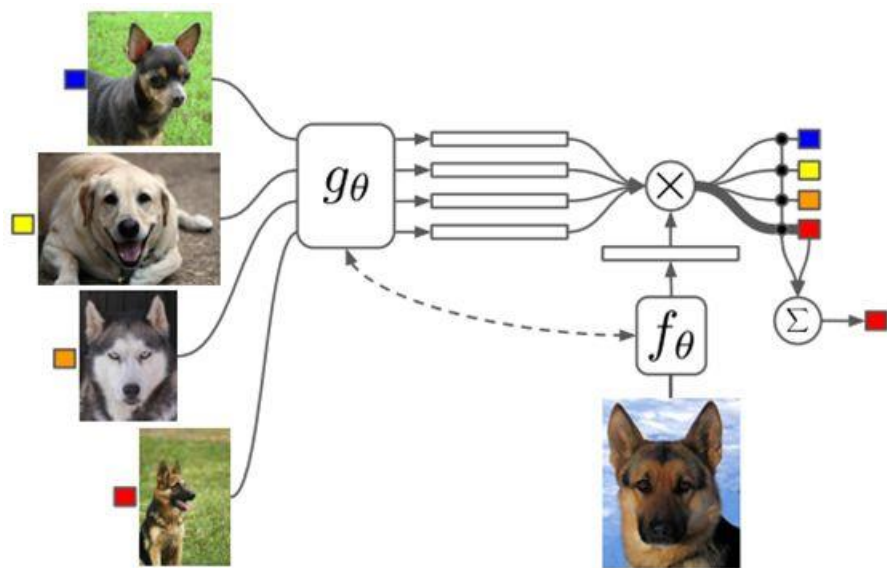


Figure 1: Matching Networks architecture

- 左边4个图片形成一组, 称为 support set; 右下1个狗, 称为test example。全部5个图片称为1个task。

匹配网络Matching Networks for One Shot Learning

OmniGlott 是一个类似 MNIST 的数据集，一共有1623个characters，每个里面有20个样本。

Model	Matching Fn	Fine Tune	5-way Acc		20-way Acc	
			1-shot	5-shot	1-shot	5-shot
PIXELS	Cosine	N	41.7%	63.2%	26.7%	42.6%
BASLINE CLASSIFIER	Cosine	N	80.0%	95.0%	69.5%	89.1%
BASLINE CLASSIFIER	Cosine	Y	82.3%	98.4%	70.6%	92.0%
BASLINE CLASSIFIER	Softmax	Y	86.0%	97.6%	72.9%	92.3%
MANN (NO CONV) [21]	Cosine	N	82.8%	94.9%	—	—
CONVOLUTIONAL SIAMESE NET [11]	Cosine	N	96.7%	98.4%	88.0%	96.5%
CONVOLUTIONAL SIAMESE NET [11]	Cosine	Y	97.3%	98.4%	88.1%	97.0%
MATCHING NETS (OURS)	Cosine	N	98.1%	98.9%	93.8%	98.5%
MATCHING NETS (OURS)	Cosine	Y	97.9%	98.7%	93.5%	98.7%

Table 1: Results on the OmniGlott dataset.

作者还新定义了一个数据集 miniImageNet —— 一共有100个类别，每个类有600个样本。其中80个类用于训练20个类用于测试。

Table 2: Results on miniImageNet.

Model	Matching Fn	Fine Tune	5-way Acc	
			1-shot	5-shot
PIXELS	Cosine	N	23.0%	26.6%
BASLINE CLASSIFIER	Cosine	N	36.6%	46.0%
BASLINE CLASSIFIER	Cosine	Y	36.2%	52.2%
BASLINE CLASSIFIER	Softmax	Y	38.4%	51.2%
MATCHING NETS (OURS)	Cosine	N	41.2%	56.2%
MATCHING NETS (OURS)	Cosine	Y	42.4%	58.0%
MATCHING NETS (OURS)	Cosine (FCE)	N	44.2%	57.0%
MATCHING NETS (OURS)	Cosine (FCE)	Y	46.6%	60.0%

在ImageNet 上的实验：

In the rand setup：在训练集中随机去除了118个label的样本，并将这118个标签的样本用于之后的测试。

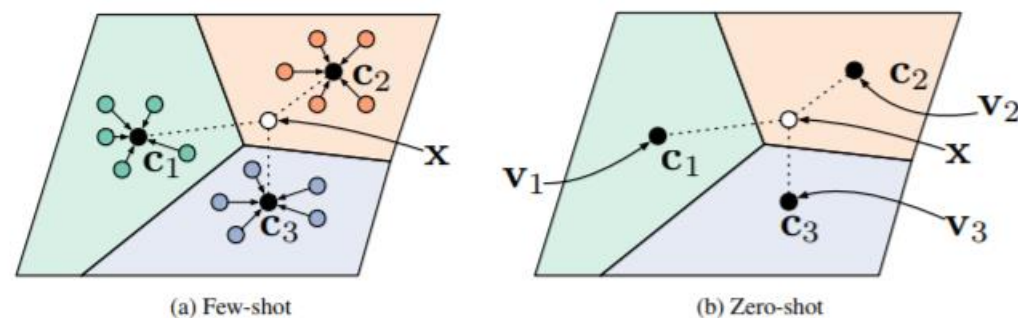
For the dogs setup：移除了所有属于狗这一大类的样本（一共118个子类），之后用这118个狗的子类样本做测试。

Table 3: Results on full ImageNet on rand and dogs one-shot tasks. Note that $\neq L_{rand}$ and $\neq L_{dogs}$ are sets of classes which are seen during training, but are provided for completeness.

Model	Matching Fn	Fine Tune	ImageNet 5-way 1-shot Acc			
			L_{rand}	$\neq L_{rand}$	L_{dogs}	$\neq L_{dogs}$
PIXELS	Cosine	N	42.0%	42.8%	41.4%	43.0%
INCEPTION CLASSIFIER	Cosine	N	87.6%	92.6%	59.8%	90.0%
MATCHING NETS (OURS)	Cosine (FCE)	N	93.2%	97.0%	58.8%	96.4%
INCEPTION ORACLE	Softmax (Full)	Y (Full)	$\approx 99\%$	$\approx 99\%$	$\approx 99\%$	$\approx 99\%$

原型网络2017Prototypical networks for few-shot learning

- 对于分类问题，原型网络将其看做在语义空间中寻找每一类的原型中心。针对Few-shot的任务定义，原型网络训练时学习如何拟合中心。学习一个度量函数，该度量函数可以通过少量的几个样本找到所属类别在该度量空间的原型中心。测试时，用支持集（Support Set）中的样本来计算新的类别的聚类中心，再利用最近邻分类器的思路进行预测。本文主要针对 Few-Shot / Zero-Shot任务中过拟合的问题进行研究，将原型网络和聚类联系起来，和目前的一些方法进行比较，取得了不错的效果。



原型网络2017Prototypical networks for few-shot learning

- 本文对few-shot任务在Omniglot和miniImageNet version of ILSVRC-2012 图像数据集上进行了实验。对zero-shot任务在Caltech UCSD bird dataset (CUB-200 2011) 进行了实验。
- 在mini ImageNet数据集上作者还对距离函数的选择、每个训练单元中包含的类的数量 (5/20) , 对测试结果 (5-way) 的影响。作者发现训练时每个episode中包含的类多一些效果会比较好。

- Zero-Shot本文也达到了目前最好的效果

Table 1: Few-shot classification accuracies on Omniglot.

Model	Dist.	Fine Tune	5-way Acc.		20-way Acc.	
			1-shot	5-shot	1-shot	5-shot
MATCHING NETWORKS [29]	Cosine	N	98.1%	98.9%	93.8%	98.5%
MATCHING NETWORKS [29]	Cosine	Y	97.9%	98.7%	93.5%	98.7%
NEURAL STATISTICIAN [6]	-	N	98.1%	99.5%	93.2%	98.1%
PROTOTYPICAL NETWORKS (OURS)	Euclid.	N	98.8%	99.7%	96.0%	98.9%

Table 2: Few-shot classification accuracies on *miniImageNet*. All accuracy results are averaged over 600 test episodes and are reported with 95% confidence intervals. *Results reported by [22].

Model	Dist.	Fine Tune	5-way Acc.	
			1-shot	5-shot
BASELINE NEAREST NEIGHBORS*	Cosine	N	28.86 \pm 0.54%	49.79 \pm 0.79%
MATCHING NETWORKS [29]*	Cosine	N	43.40 \pm 0.78%	51.09 \pm 0.71%
MATCHING NETWORKS FCE [29]*	Cosine	N	43.56 \pm 0.84%	55.31 \pm 0.73%
META-LEARNER LSTM [22]*	-	N	43.44 \pm 0.77%	60.60 \pm 0.71%
PROTOTYPICAL NETWORKS (OURS)	Euclid.	N	49.42 \pm 0.78%	68.20 \pm 0.66%

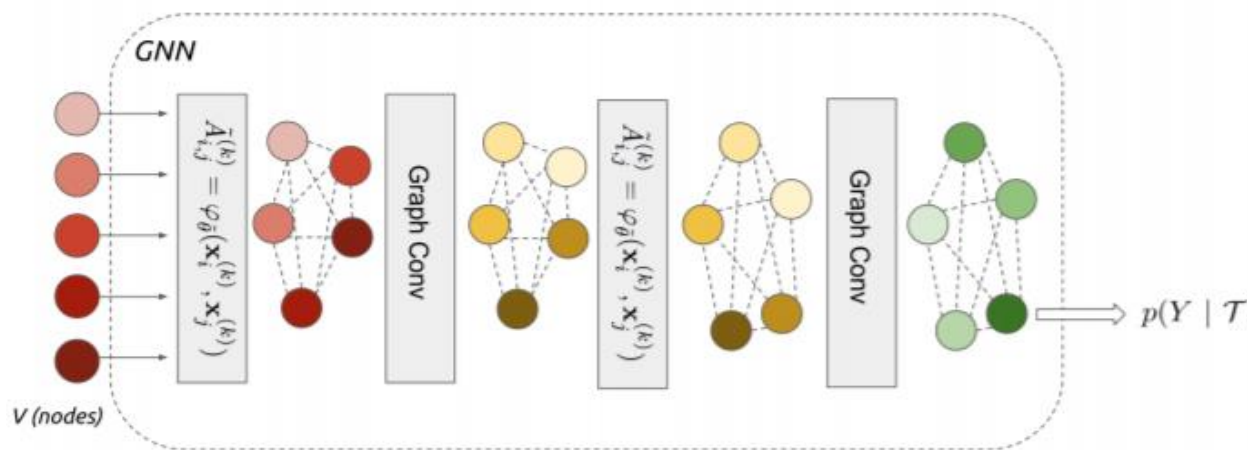
Table 3: Zero-shot classification accuracies on CUB-200.

Model	Image Features	50-way Acc. 0-shot
ALE [1]	Fisher	26.9%
SJE [2]	AlexNet	40.3%
SAMPLE CLUSTERING [17]	AlexNet	44.3%
SJE [2]	GoogLeNet	50.1%
DS-SJE [23]	GoogLeNet	50.4%
DA-SJE [23]	GoogLeNet	50.9%
PROTO. NETS (OURS)	GoogLeNet	54.6%

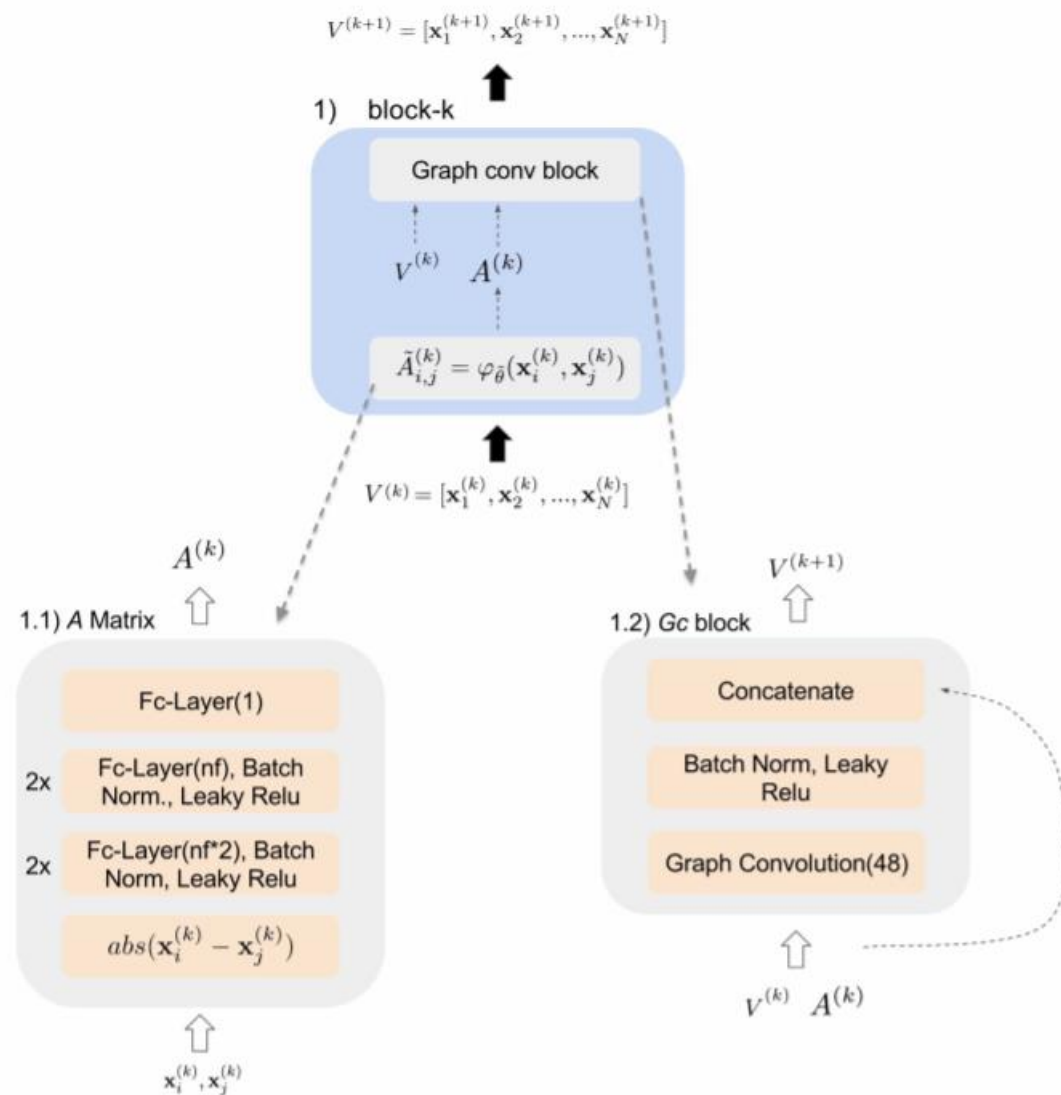
基于图2018Few-shot learning with graph neural networks

- 作者定义了一个图神经网络框架，端到端地学习消息传递的“关系”型任务。在这里，每个样本看成图的节点，该方法不仅学习节点的embedding，也学习边的embedding。如图7，在网络第一层5个样本通过边模型A ~ 构建了图，接着通过图卷积（graph conv）获得了节点的embedding，然后在后面的几层继续用A ~ 更新图、用graph conv更新节点embedding, 这样便构成了一个深度GNN，最后输出样本的预测标签。

基于图2018Few-shot learning with graph neural networks



- 在构建边模型时，先采用一个4层的CNN网络获得每个节点特征向量，然后将节点对 x_i, x_j 的差的绝对值过4层带Batch Norm和Leaky Relu的全连接层，从而获得边的embedding，如右图的左侧。随后，我们将节点的embedding和边的embedding一起过图卷积网络，从而获得更新后的节点的embedding，如右图的右侧。



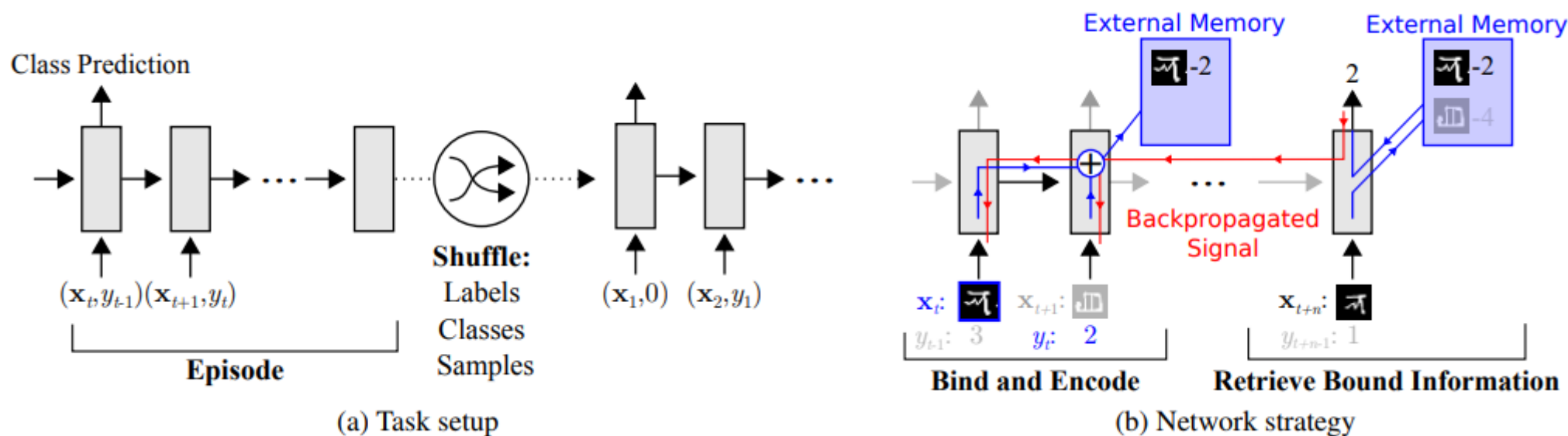
元学习

- 我们希望它能够从之前的经验快速地学习新的技能，而不是把新的任务孤立地考虑。这个方法，我们称为元学习（learning to learn,或meta learning），使得我们的系统在它的整个生命周期中可以持续地学习各种各样的任务。
- 通过使用不同类型的元数据，如学习问题的属性，算法属性（如性能测量）或从之前数据推导出的模式，可以选择、更改或组合不同的学习算法，以有效地解决给定的学习问题。

递归记忆模型2016Meta-Learning with Memory-Augmented Neural Networks

- 通过权重更新来调节bias，并且通过学习将表达快速缓存到记忆中来调节输出。然而，利用循环神经网络的内部记忆单元无法扩展到需要对大量新信息进行编码的新任务上。因此，我们需要让存储在记忆中的表达既要稳定又要是元素粒度访问的，前者是说当需要时就能可靠地访问，后者是说可选择性地访问相关的信息；另外，参数数量不能被内存的大小束缚。
- 文章基于神经网络图灵机（NTMs），通过外部存储（external memory）进行短时记忆，通过缓慢权值更新来进行长时记忆，NTMs可以学习将表达存入记忆的策略，并用这些表达来进行预测。由此，文章方法可以快速准确地预测那些只出现过一次的数据。文章基于LSTM等RNN的模型，将数据看成序列来训练，在测试时输入新的类的样本进行分类。

递归记忆模型Meta-Learning with Memory-Augmented Neural Networks



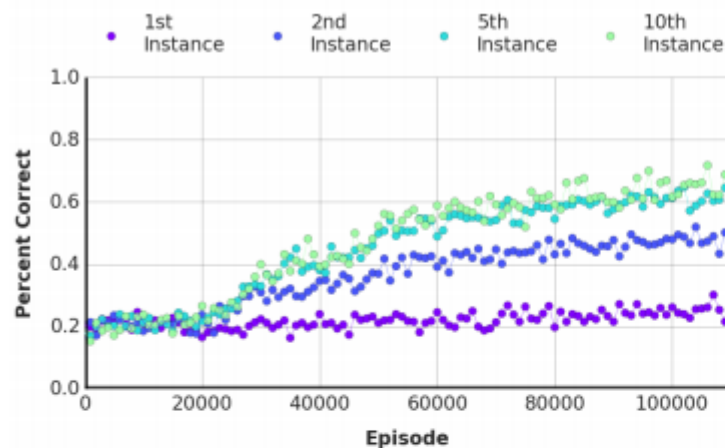
网络的输入把上一次的y (label)也作为输入，并且添加了external memory存储上一次的x输入，这使得下一次输入后进行反向传播时，可以让y (label)和x建立联系，使得之后的x能够通过外部记忆获取相关图像进行比对来实现更好的预测。这里的RNN就是meta-learner。

递归记忆模型Meta-Learning with Memory-Augmented Neural Networks

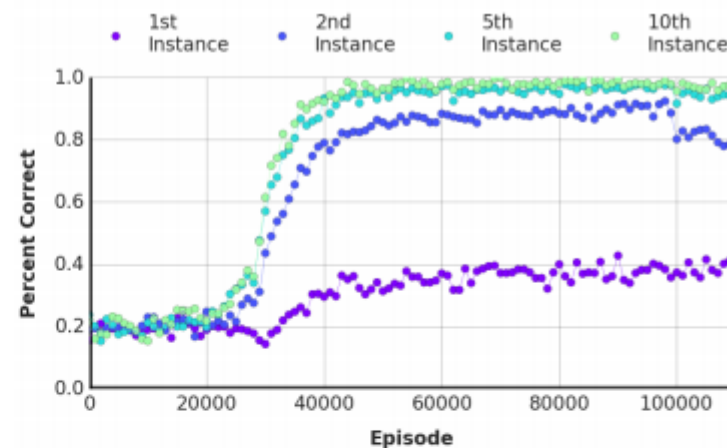
本文分别进行了Omniglot数据集分类任务以及采样于高斯过程的回归任务。

对于分类任务，本文设置的对照组为人类、使用前馈神经网络、LSTM以及MANN；同时每一个对照组分别测试了在训练好后，分别第一次、第二次、第五次以及第十次遇见未见过样本使的分类结果。

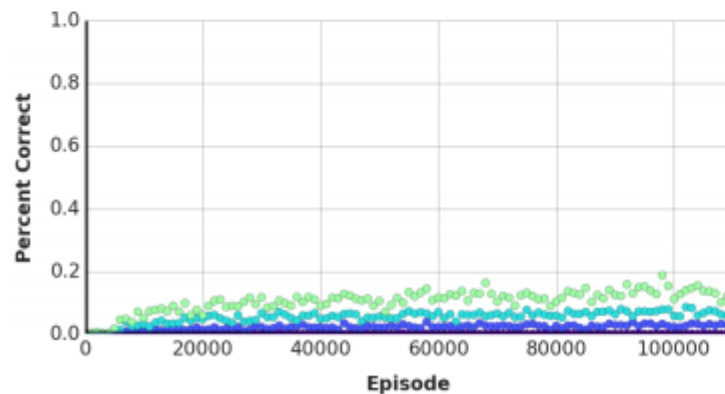
MANN第二次碰到未见过样本使就可以达到较高分类准确率，而LSTM处理该任务效果较差。也部分验证了作者对于适合于元学习的记忆增强神经网络模块的要求。MANN能够学到碰到与内存单元不相似样本时，降低访存概率，将输出猜对标记的输出概率提高技巧。



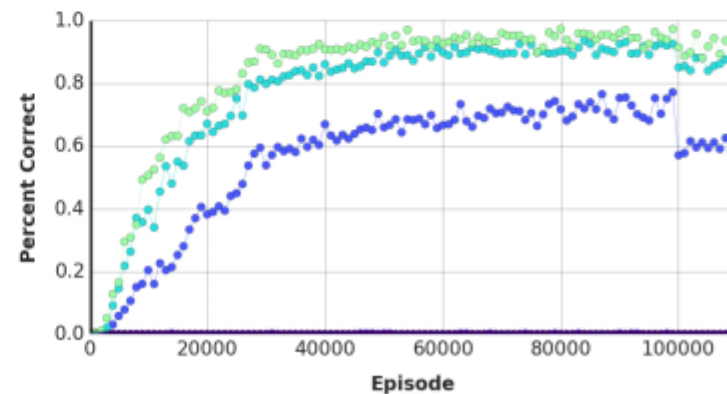
(a) LSTM, five random classes/episode, one-hot vector labels



(b) MANN, five random classes/episode, one-hot vector labels



(c) LSTM, fifteen classes/episode, five-character string labels

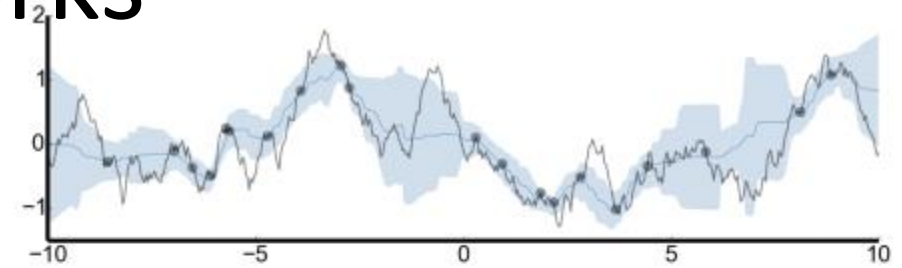


(d) MANN, fifteen classes/episode, five-character string labels

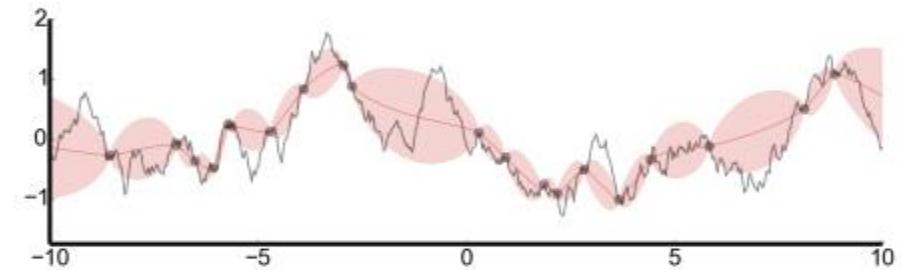
递归记忆模型Meta-Learning with Memory-Augmented Neural Networks

Table 1. Test-set classification accuracies for humans compared to machine algorithms trained on the Omniglot dataset, using one-hot encodings of labels and five classes presented per episode.

MODEL	INSTANCE (% CORRECT)					
	1 ST	2 ND	3 RD	4 TH	5 TH	10 TH
HUMAN	34.5	57.3	70.1	71.8	81.4	92.4
FEEDFORWARD	24.4	19.6	21.1	19.9	22.8	19.5
LSTM	24.4	49.5	55.3	61.0	63.6	62.5
MANN	36.4	82.8	91.0	92.6	94.9	98.1



(a) MANN predictions along all x -inputs after 20 samples



(b) GP predictions along all x -inputs after 20 samples

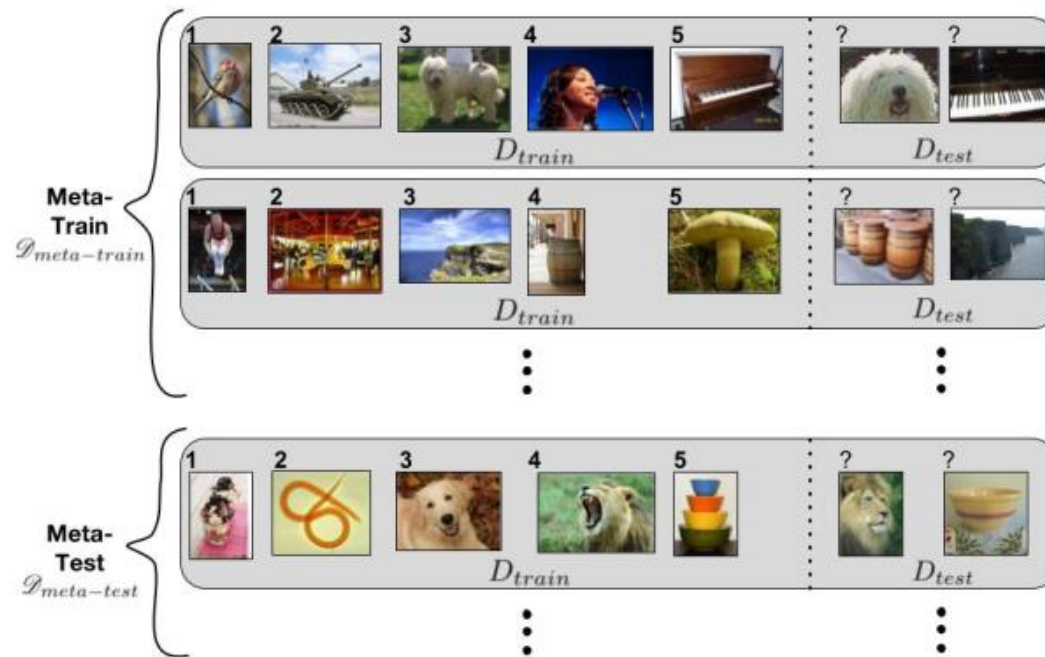
Figure 5. Regression. The network received data samples that were x -values for a function sampled from a GP with fixed hyperparameters, and the labels were the associated function values. (a) shows the MANN's predictions for all x -values after observing 20 samples, and (b) shows the same for a GP with access to the same hyper-parameters used to generate the function.

优化器学习2017 Optimization as a model for few-shot learning

- 本文提出一种基于LSTM的meta-learner模型，用于学习将会作用于另一个任务的分类器的最优优化算法。
- 作者发现LSTM的更新规则和一般的梯度下降算法更新规则非常类似，无法直接用于meta learning，因此提出一种新的LSTM更新规则。
- 元学习有很多参数，为了防止参数爆炸，使用参数共享。

优化器学习 Optimization as a model for few-shot learning

- 如图是一个元学习，元学习的数据集(Meta sets)包括元训练集 (Meta-train set) 和元测试集 (Meta-test set)。其中包含有若干个灰色矩形框分别表示每一条样本，而矩形框内部又被虚线划分为左右两个部分，虚线左侧的包含有训练集 D_{train} ，而虚线的右侧包含有测试集 D_{test} 。该示例所展示的是一个5种分类别的类任务，然后测试集中则包含有2个样例用于评估模型。图片下半部分的表示元测试集 (Meta-test set)， $D_{meta-test}$ 包含有 $D_{meta-train}$ 中没有覆盖到的类别。



优化器学习 Optimization as a model for few-shot learning

Model	5-class	
	1-shot	5-shot
Baseline-finetune	$28.86 \pm 0.54\%$	$49.79 \pm 0.79\%$
Baseline-nearest-neighbor	$41.08 \pm 0.70\%$	$51.04 \pm 0.65\%$
Matching Network	$43.40 \pm 0.78\%$	$51.09 \pm 0.71\%$
Matching Network FCE	$43.56 \pm 0.84\%$	$55.31 \pm 0.73\%$
Meta-Learner LSTM (OURS)	$43.44 \pm 0.77\%$	$60.60 \pm 0.71\%$

Table 1: Average classification accuracies on Mini-ImageNet with 95% confidence intervals. Marked in bold are the best results for each scenario, as well as other results with an overlapping confidence interval.

模型无关自适应2017Model-agnostic meta-learning for fast adaptation of deep networks

- **摘要：**我们提出了一种与模型无关的（model-agnostic）元学习算法，它能匹配任何使用梯度下降算法训练的模型，并能应用于各种不同的学习问题，如分类、回归和强化学习等。元学习的目标是在各种学习任务上训练一个模型，因此我们就可以只使用少量的梯度迭代步来解决新的学习任务。在我们的方法中，模型的参数能精确地进行训练，因此少量的梯度迭代步和训练数据样本能在该任务上产生十分优秀的泛化性能。实际上，我们的方法可以很简单地对模型进行参数微调（fine-tune）。我们证明，MAML 方法在两个小规模图像分类基准上有最优秀的性能，在少量样本回归中也产生了非常优秀的性能，同时它还能通过神经网络策略加速策略梯度强化学习（policy gradient reinforcement learning）的微调。

模型无关自适应2017Model-agnostic meta-learning for fast adaptation of deep networks

- 作者认为存在某些内在的特征能够广泛地适用于由任务分布 $P(\tau)$ 得来的任务，而不是仅仅适用于单个任务。因此，作者希望通过改变梯度下降的方向（如图），找到对于任务分布来说敏感的参数，这些参数细微的改动能引起loss函数很大的变化。因此在使用少量的训练数据时，这些参数的梯度下降能使得loss快速下降，达到很好的fine-tune效果。

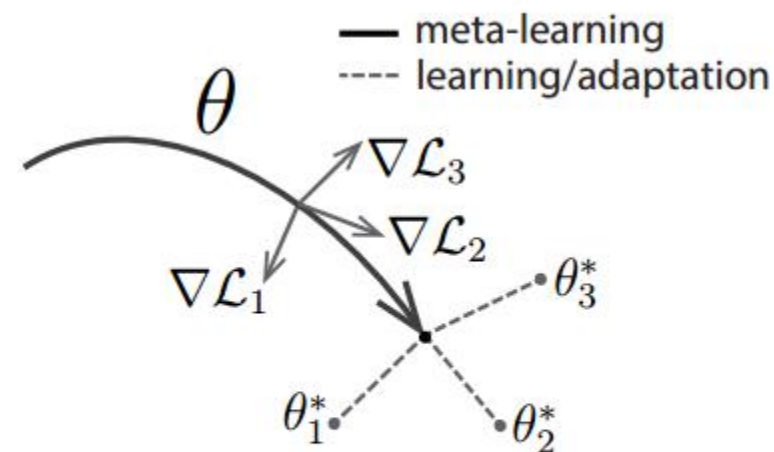


Figure 1. Diagram of our model-agnostic meta-learning algorithm (MAML), which optimizes for a representation θ that can quickly adapt to new tasks.

模型无关自适应2017Model-agnostic meta-learning for fast adaptation of deep networks

- 在作者给出的代码中，外层循环的次数由超参数 `metatrain_iterations` 指定。对于每次循环，从 $\delta_{\text{meta-train}}$ 中采样 `meta_batch_size` 个数的任务作为一个 batch。图中是有内层循环，对于每个任务单独计算一个 θ'_i 。在实际代码中，论文作者把它 K 个任务化作一个 batch，同时计算 θ_i 以及对应的 loss。值得注意的是：在 Meta-Train 阶段，更新得到 θ'_i 时，使用的是每个任务的 train 数据集；而在外层更新新的 θ 时，loss 函数是在 test 集合上的 loss，对原 θ （注意不是 θ' ）进行梯度下降。

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks

Require: α, β : step size hyperparameters

```
1: randomly initialize  $\theta$ 
2: while not done do
3:   Sample batch of tasks  $\mathcal{T}_i \sim p(\mathcal{T})$ 
4:   for all  $\mathcal{T}_i$  do
5:     Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  with respect to  $K$  examples
6:     Compute adapted parameters with gradient descent:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$ 
7:   end for
8:   Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$ 
9: end while
```

Thanks

张子昊

CVPR2017: Learning a Deep Embedding Model for Zero-Shot Learning

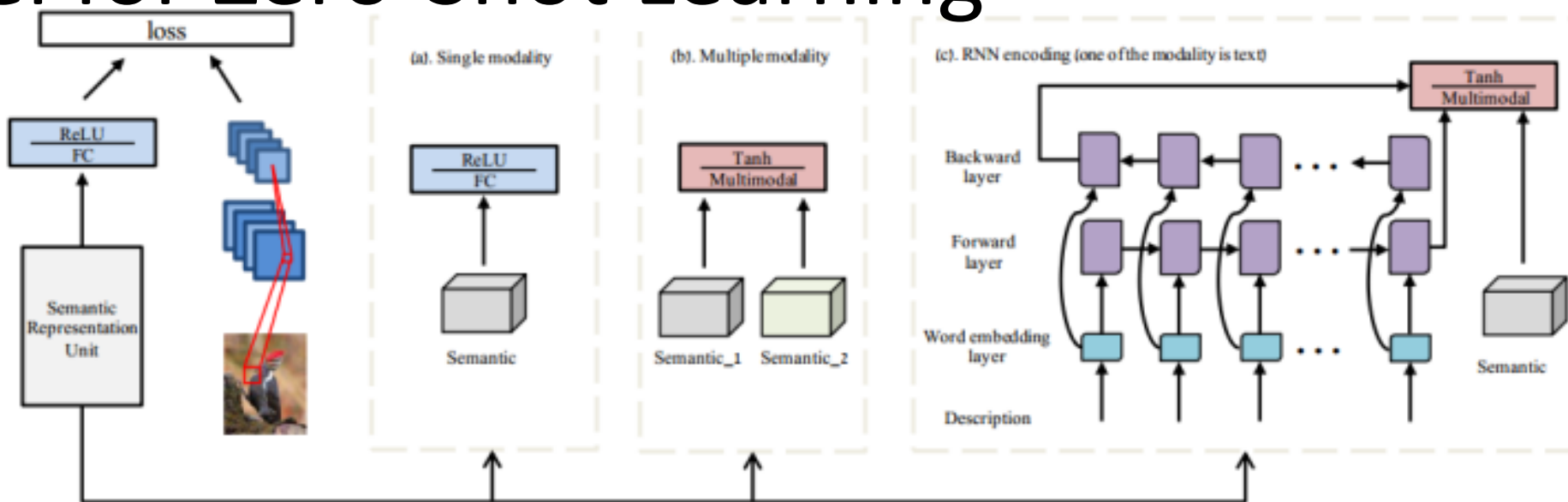
- 贡献：1、本文提出了一种新的端对端的深度嵌入模型来完成Zero-shot learning (ZSL) 工作；2、能够很自然地衍生到多模态的融合算法中；3、在多个数据集上取得了最好的效果
- 利用端对端的模型学习一个深度嵌入模型的好处：1、它能够更好地学习一个嵌入空间；2、为基于神经网络的联合嵌入模型提供了灵活性，能够解决多种迁移性问题；3、可以很自然地多模态的数据进行fusing。
- 枢纽度问题 (hubness problem)：由于嵌入空间是一个高维空间，所以很容易出现hubness problem。该问题是指：在高维空间中，一部分测试集类别可能会成为很多数据点的K近邻(KNN)，但其类别之间却没什么关系。当我们使用语义空间(semantic space)作为嵌入空间时，需要将视觉特征映射到语义空间中，这样会使得空间发生萎缩，点与点之间更加稠密，从而加重hubness problem。

<https://zhuanlan.zhihu.com/p/28378592>

http://openaccess.thecvf.com/content_cvpr_2017/html/Zhang_Learning_a_Deep_CVPR_2017_paper.html

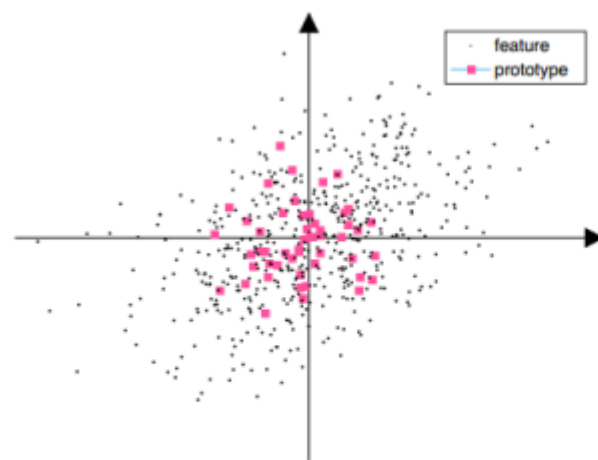
https://github.com/dragen1860/DeepEmbeddingModel_ZSL-Pytorch

CVPR2017: Learning a Deep Embedding Model for Zero-Shot Learning

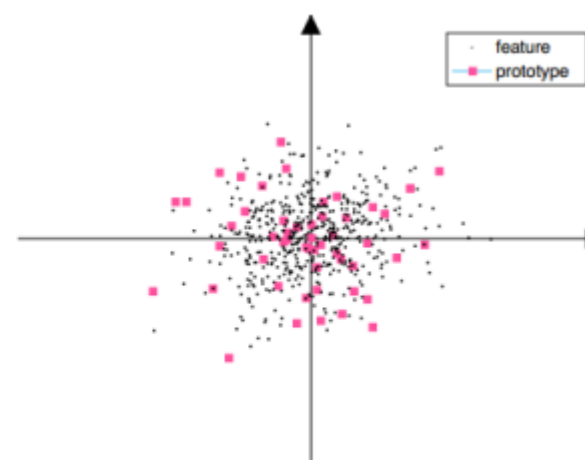


- 图中最左边的图为整个算法的模型，当中的语义表示单元(Semantic Representation Unit)又可以分别利用右边的三个结构进行代替，它们分别是单模态形式、多模态形式和RNN词嵌入形式。

CVPR2017: Learning a Deep Embedding Model for Zero-Shot Learning



(a) $S \rightarrow V$



(b) $V \rightarrow S$

- s 表示语义空间， v 表示视觉特征空间。当将语义特征映射到视觉特征空间中时（就如本算法），hubness problem不会加重，但当映射为 $V \rightarrow S$ 时，hubness problem会明显加重，feature点之间变更稠密了。

CVPR2017: Learning a Deep Embedding Model for Zero-Shot Learning

- 在数据集AwA和CUB上的结果如表，其中FO表示overfeat，FG表示GoogleNet，FV表示VGGnet，NG表示Inception-V2；A为attribute表示，W为特征的词嵌入表示，D为句子描述（sentence description）。

Model	hit@5
ConSE	28.5
DeViSE	31.8
Mensink <i>et al.</i>	35.7
Rohrbach	34.8
PST	34.0
AMP	41.0
Ours	46.7
Gaussian Embedding	45.7
PDDM	48.2
Ours	60.7

Table 2. Comparative results (%) on ILSVRC 2010

Model	F	SS	AwA	CUB
AMP	F_O	A+W	66.0	-
SJE	F_G	A	66.7	50.1
SJE	F_G	A+W	73.9	51.7
ESZSL	F_G	A	76.3	47.2
SSE-ReLU	F_V	A	76.3	30.4
JLSE	F_V	A	80.5	42.1
SS-Voc	F_O	A/W	78.3/68.9	-
SynC-struct	F_G	A	72.9	54.5
SEC-ML	F_V	A	77.3	43.3
DeViSE	N_G	A/W	56.7/50.4	33.5
Socher <i>et al.</i>	N_G	A/W	60.8/50.3	39.6
MTMDL	N_G	A/W	63.7/55.3	32.3
Ba <i>et al.</i>	N_G	A/W	69.3/58.7	34.0
DS-SJE	N_G	A/D	-	50.4/56.8
Ours	N_G	A/W(D)	86.7/78.8	58.3/53.5
Ours	N_G	A+W(D)	88.1	59.0

CVPR2019 Revisiting Local Descriptor based Image-to-Class Measure for Few-shot Learning

- 图像分类中的few shot learning旨在学习分类器，以便在每个类只有少数训练样例时对图像进行分类。最近的工作已经取得了很好的分类性能，**其中通常使用基于图像级特征的度量。在本文中，我们认为，鉴于少数学习中的例子稀缺，在这样的水平上的衡量可能不够有效。相反，我们认为应该采用基于本地描述符的图像到类测量，这是受其在本地图像不变特征的全盛时期的惊人成功所启发。**具体来说，在最近的情节训练机制的基础上，**我们提出了一个深度最近邻神经网络（简称DN4）并以端到端的方式训练它。**它与文献的主要区别在于**通过基于局部描述符的图像到类测量来替换最终层中基于图像级特征的测量。该测量是通过k-最近邻搜索在线性特征图的深度局部描述符在线进行的。**所提出的DN4**不仅学习了图像到类测量的最佳深度局部描述符，而且在示例稀缺的情况下也利用了这种测量的更高效率，这要归功于图像中的视觉模式的可交换性。**类。我们的工作为简单的学习提供了一个简单，有效且计算有效的框架。对基准数据集的实验研究一致表明其优于相关的现有技术，最大的绝对改进率为17%。
源代码可以从<https://github.com/WenbinLee/DN4.git>获得 **(未公开代码)**

CVPR2019Rethinking Knowledge Graph Propagation for Zero-Shot Learning

- 最近，图卷积神经网络显示出了Zero shot learning任务的巨大潜力。这些模型具有高度的样本效率，因为图结构中的相关概念共享统计强度，允许在面对缺乏数据时对新类进行泛化。然而，需要将知识传播到图中的远程节点所需的多层体系结构通过在每一层执行大量拉普拉斯平滑来稀释知识，从而因此降低性能。**为了仍然享受图结构带来的好处，同时防止远距离节点的知识被稀释，我们提出了一种密集图传播（DGP）模块，在远端节点之间有精心设计的直接链接。DGP允许我们通过附加连接利用知识图的层次图结构。根据节点与其祖先和后代的关系添加这些连接。加权方案还用于根据到节点的距离来权衡它们的贡献，以改善图中的信息传播。结合两阶段训练方法中的表示的微调，我们的方法优于最先进的零射击学习方法。**

CVPR2019 Rethinking Knowledge Graph Propagation for Zero-Shot Learning

- 本文在论文 1 的基础上进行了改进，包括以下几个方面：
 - (1) 更少的 GCN 层数，论文 1 中使用了 6 层神经网络进行训练，考虑到模型参数的优化问题，本文只使用了 2 层神经网络进行计算，即 GPM；
 - (2) 减少层数的同时，一些较远节点将不被考虑在内，为了解决这个问题，作者将一些节点的祖先节点/子孙节点直接与该节点相连，生成了更密集的图，即 DGPM；同时，这些直接相连的边按照距离的远近，加入 attention 机制进行了加权计算，即 ADGPM；
 - (3) 作者还提出了在 CNN 部分根据 graph 信息进行 fine tune 的计算方式，使得提取图片特征的卷积网络可根据一些新出现的 class 进行更新。
- 作者使用了与论文 1 中相同的数据集，即 ImageNet 2012 1k class (seen class) 作为训练，ImageNet 2-hops/3-hops/all (unseen class) 三个测试集作为测试。对比结果如下，其中 GCNZ 代表论文 1 中的方法，GPM、DGPM、ADGPM 分别表示上述优化的 (1)(2) 方面，GPM(f)、DGPM(f)、ADGPM(f) 表示 finetune 的结果，同样地，“2-hops+1k” 表示 generalized ZSL。

论文1: Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs

CVPR2019Rethinking Knowledge Graph Propagation for Zero-Shot Learning

Test set	Model	Hit@k (%)				
		1	2	5	10	20
2-hops	ConSE*	8.3	12.9	21.8	30.9	41.7
	SYNC*	10.5	17.7	28.6	40.1	52.0
	EXEM [†]	12.5	19.5	32.3	43.7	55.2
	GCNZ [‡]	19.8	33.3	53.2	65.4	74.6
	SGCN (ours)	26.2	40.4	60.2	71.9	81.0
	DGP (ours)	26.6	40.7	60.3	72.3	81.3
3-hops	ConSE*	2.6	4.1	7.3	11.1	16.4
	SYNC*	2.9	4.9	9.2	14.2	20.9
	EXEM [†]	3.6	5.9	10.7	16.1	23.1
	GCNZ [‡]	4.1	7.5	14.2	20.2	27.7
	SGCN (ours)	6.0	10.4	18.9	27.2	36.9
	DGP (ours)	6.3	10.7	19.3	27.7	37.7
All	ConSE*	1.3	2.1	3.8	5.8	8.7
	SYNC*	1.4	2.4	4.5	7.1	10.9
	EXEM [†]	1.8	2.9	5.3	8.2	12.2
	GCNZ [‡]	1.8	3.3	6.3	9.1	12.7
	SGCN (ours)	2.8	4.9	9.1	13.5	19.3
	DGP (ours)	3.0	5.0	9.3	13.9	19.8

Test set	Model	Hit@k (%)				
		1	2	5	10	20
2-hops+1K	DeViSE ^{††}	0.8	2.7	7.9	14.2	22.7
	ConSE ^{‡‡}	0.3	6.2	17.0	24.9	33.5
	ConSE [‡]	0.1	11.2	24.3	29.1	32.7
	GCNZ [‡]	9.7	20.4	42.6	57.0	68.2
	SGCN (ours)	11.9	27.0	50.8	65.1	75.9
	DGP (ours)	10.3	26.4	50.3	65.2	76.0
3-hops+1K	DeViSE ^{††}	0.5	1.4	3.4	5.9	9.7
	ConSE ^{‡‡}	0.2	2.2	5.9	9.7	14.3
	ConSE [‡]	0.2	3.2	7.3	10.0	12.2
	GCNZ [‡]	2.2	5.1	11.9	18.0	25.6
	SGCN (ours)	3.2	7.1	16.1	24.6	34.6
	DGP (ours)	2.9	7.1	16.1	24.9	35.1
All+1K	DeViSE ^{††}	0.3	0.8	1.9	3.2	5.3
	ConSE ^{‡‡}	0.2	1.2	3.0	5.0	7.5
	ConSE [‡]	0.1	1.5	3.5	4.9	6.2
	GCNZ [‡]	1.0	2.3	5.3	8.1	11.7
	SGCN (ours)	1.5	3.4	7.8	12.3	18.2
	DGP (ours)	1.4	3.4	7.9	12.6	18.7

2018One-Shot Relational Learning for Knowledge Graphs

- 知识图 (KG) 是各种自然语言处理应用的关键组成部分。为了进一步扩大知识图的覆盖范围, 先前对知识图完成的研究通常需要为每个关系提供大量训练实例。然而, 我们观察到长尾关系实际上在知识图更常见, 而那些新增的关系往往没有很多已知的训练三元组。在这项工作中, 我们的目标是在具有挑战性的环境中预测新事实, 其中只有一个训练实例可用。我们提出了一次性关系学习框架, 该框架利用通过嵌入模型提取的知识, 并通过考虑学习嵌入和单跳图结构来学习匹配度量。根据经验, 我们的模型比现有的嵌入模型产生了相当大的性能改进,

2018One-Shot Relational Learning for Knowledge Graphs

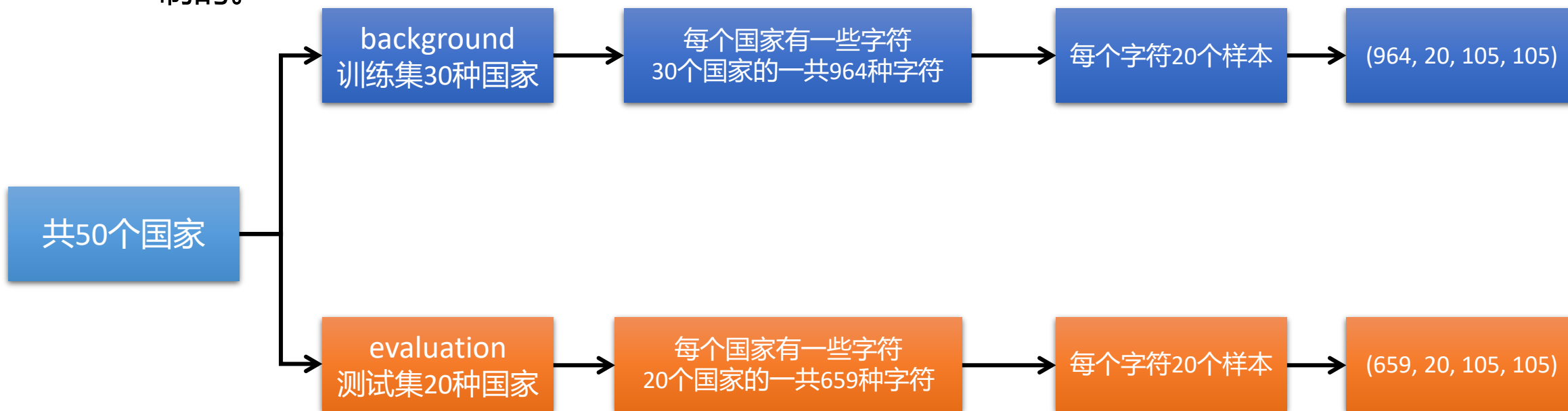
- 在知识图谱补全 (Knowledge Graph Completion) 任务中, 对实体间的关系分类是非常重要的的一环。目前的方法假设每种关系都存在大量训练样本进行有监督训练, 但是现实中的知识图谱训练集中普遍存在大量长尾的关系类别, 使训练效果无法达到预期; 而且对于任意新增关系, 在没有充足训练数据的前提下无法进行预测。针对这些问题, 本文提出了一种one-shot知识谱图关系学习框架, 利用embedding-based的方法学习匹配度量函数, 在新增关系或者长尾关系只有非常少量的训练数据的前提下也可以进行很好的预测, 同时取得了不错的效果。

2018 One-Shot Relational Learning for Knowledge Graphs

<http://tongtianta.site/paper/2790>

Omniglot

- Omniglot数据集旨在开发更像人类的学习算法。它包含来自50种不同字母的1623种不同的手写字符。1623个字母中的每一个都是由20位不同的人通过亚马逊的机械土耳其人在线绘制的。



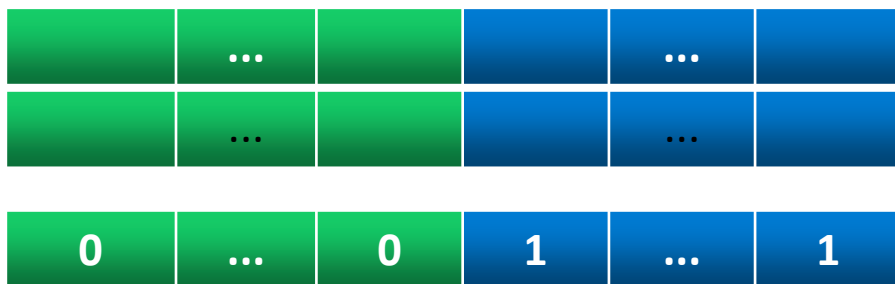
孪生网络Siamese neural networks for one-shot image recognition(2015)

训练:

input1(32, 105, 105, 1)

input2(32, 105, 105, 1)

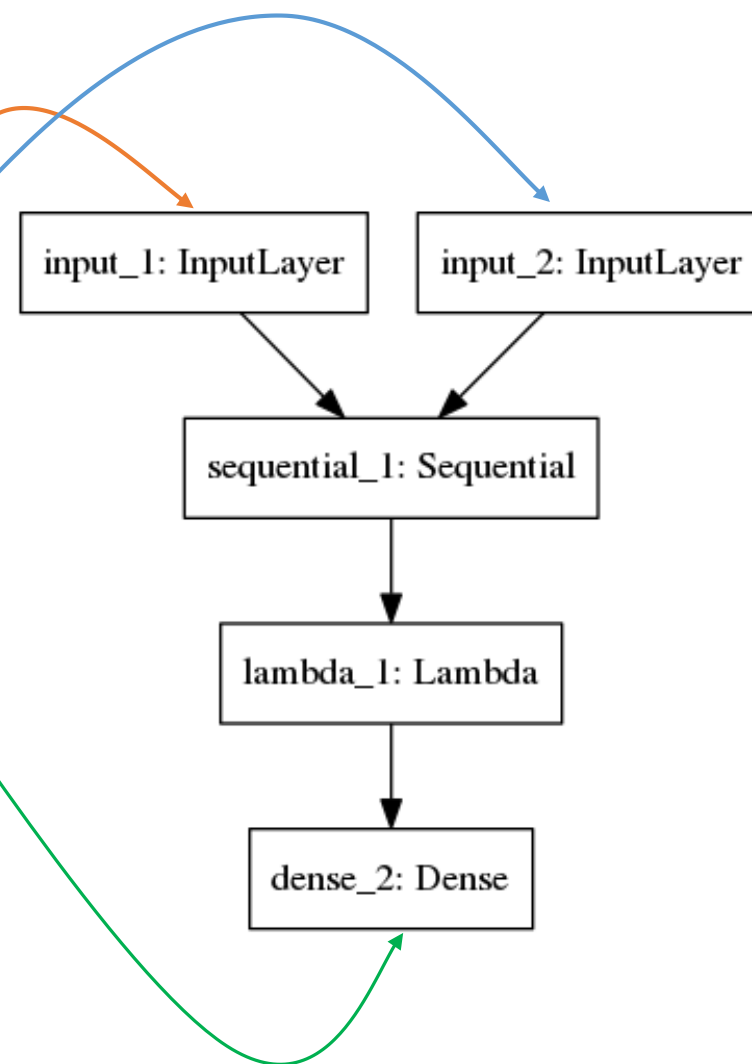
label(32,)



前16个类别不同

后16个类别相同

一个batch共32个数据



孪生网络Siamese neural networks for one-shot image recognition(2015)

验证:

