

## 训练过程中使用支持集信息的单样本学习算法

辛守宇, 郑蕊蕊, 周 瑜, 刘文鹏, 贺建军

(大连民族大学信息与通信工程学院, 辽宁大连 116000)

**摘要:** 单样本学习的目的是利用一个包含大量训练样本的源类别数据集以及每个类别只包含一个训练样本的目标类别数据集来构建一种学习算法, 使得算法能够对目标类别空间中的样本进行准确分类. 已有的单样本学习算法主要是先利用源类别数据集来训练模型, 然后在测试时将目标类别训练数据作为支持集来实现对未标注样本的分类, 因此在训练时没有有效地利用支持集的信息. 为此提出一种在训练阶段和测试阶段同时利用支持集信息的单样本学习算法, 基本思想是利用孪生神经网络构建模型并在训练时加入支持集信息, 即让不同类别的支持集样本之间的相似度尽可能小. 在 Omniglot 数据集和满文识别问题上的实验结果表明, 该算法能取得较好的识别准确率.

**关键词:** 单样本学习; 孪生神经网络; 度量学习

**中图分类号:** TP391.1 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2020.08.020

**引用格式:** 辛守宇, 郑蕊蕊, 周瑜, 等. 训练过程中使用支持集信息的单样本学习算法[J]. 中国科学技术大学学报, 2020, 50(8):1187-1192.

XIN Shouyu, ZHENG Ruirui, ZHOU Yu, et al. A one-shot learning algorithm using support set information during training[J]. Journal of University of Science and Technology of China, 2020, 50(8):1187-1192.

### A one-shot learning algorithm using support set information during training

XIN Shouyu, ZHENG Ruirui, ZHOU Yu, LIU Wenpeng, HE Jianjun

(College of Information and Communication Engineering, Dalian Minzu University, Dalian 116000, China)

**Abstract:** The purpose of one-shot learning is to use a source category dataset containing a large number of training samples and a target category dataset containing only one training sample per category to construct a learning algorithm that enables accurate classification of samples in the target category space. The existing one-shot learning algorithm mainly uses the source category data to train the model, and then uses the training data of the target category as the support set to realize the classification of the unlabeled samples during the test. Therefore, it fails to effectively utilize the information of the support set during the training. Here, a one-shot learning algorithm using support set information in both the training and test stages is established. The basic idea is to use Siamese neural networks to build models and add support set information during training, that is, to make the similarity between different types of support set samples as small as possible. Experimental results on Omniglot data set and Manchu recognition show that the proposed algorithm can achieve better recognition accuracy.

**Key words:** one-shot learning; siamese neural networks; metric learning

## 0 引言

深度学习模型在图像识别和检测任务中取得了优异的成果<sup>[1-4]</sup>. 这些有监督的神经网络模型需要大量有标注的样本来训练其参数, 严重限制了模型的适用性. 于是, 学者开始关注小样本学习问题. 在近些年的研究中, 针对小样本学习的策略主要集中于零样本学习 (zero-shot learning)、单样本学习 (one-shot learning) 和小样本学习 (few-shot

learning), 其中最具代表性的是单样本学习. 单样本学习任务是指在对测试实例进行测试前, 我们只能观察每个可能类的单个示例, 以此为限制条件对测试样本进行分类, 这也是本文工作所关注的焦点.

随着深度学习方法在许多分类任务上的成功, 人们产生了将这种方法推广到单样本学习领域的兴趣. 早期关于单样本学习的算法主要基于概率生成方法, 如 Li 等<sup>[5]</sup>提出的基于贝叶斯生成框架的单样本学习算法, Lake 等<sup>[6]</sup>提出的将分层隐马尔可夫

收稿日期: 2020-06-03; 修回日期: 2020-07-09

基金项目: 国家自然科学基金(61702081)资助.

作者简介: 辛守宇, 男, 1995年生, 硕士研究生. 研究方向: 深度学习. Email: xsy2268745113@163.com

通讯作者: 贺建军, 博士/副教授. Email: jianjunhe@live.com

模型与贝叶斯推理过程相结合的单样本学习算法. 在实际问题的处理中, 因数据非常复杂, 概率生成模型无法取得很好效果. 现在人们关于单样本学习的研究开始关注其他方法, 如度量学习和元学习等. 在这些方法中, 人们选择把复杂的多分类问题看作二分类问题处理, 只关注样本相似或非相似的问题. 如 Koch 等<sup>[7]</sup> 提出了用于单样本图片识别的孪生神经网络 (siamese neural network), 它先通过一个有监督的基于度量学习的孪生神经网络来进行训练, 然后用训练完成的网络来自动发现新样本上泛化的特征, 最后通过比较支持集样本和测试样本是否同类进行单样本图像识别任务. Oriol 等<sup>[8]</sup> 提出了匹配网络 (matching networks), 其训练数据和测试数据都以“集 (episode)”的形式进行组织.  $N$ -way-1-shot 的任务集是指由  $N$  种类别的单个样本组成支持集, 并且由来自这  $N$  种类别的其他样本作为测试样本的一次训练或测试任务. 此外, 在研究单样本学习中, 人们更多地使用元学习的方法. 如 Adam 等<sup>[9]</sup> 提出了一种记忆增强神经网络以快速学习元学习机制中的新数据, 再利用这些数据对一个新样本进行准确预测. Chen 等<sup>[10]</sup> 提出了一种基于元学习的新型图像变形框架以解决单样本学习问题也取得了很好的效果.

这些有关单样本学习的研究方法虽然取得了很好的效果, 但没有考虑在训练时利用目标类别中有标注的单个样本, 只是在测试时把这些有标注的单个样本作为支持集来实现对未标注样本的分类. 本文建立了基于度量学习的孪生网络模型, 并重新设计了孪生网络的损失函数, 在训练时针对支持集样本训练模型加大其不同类别间的相似性度量. 不

仅将这些有标注的单个样本作为支持集进行单样本识别任务, 而且在训练模型时利用支持集的信息加强模型对当前支持集中不同类别样本的区分能力. 我们在 Omniglot 数据集和满文数据集上验证了算法的性能. 实验不但进行了在 20-way 1-shot 的场景下的实验, 为验证算法在类别数量较多的复杂场景下依然有较好的识别率, 还进行了在 200-way 1-shot 场景下的实验. 实验的结果表明, 本文提出的算法在这两种场景下都有较好的识别能力.

### 1 相关工作

早期关于单样本识别方法的研究大都是利用概率生成模型进行的. 如 Li 等<sup>[11]</sup> 引入了贝叶斯生成框架并将先验知识表示为模型参数上的概率密度函数, 通过贝叶斯方法实现单样本识别. Lake 等<sup>[12]</sup> 提出了一个基于组合因果关系的分层贝叶斯模型, 可以学习多种物体的视觉概念, 并且能从一个图像中进行图像特征的推广. Ruslan 等<sup>[13]</sup> 提出了一个分层贝叶斯模型, 该模型能够将目标类别分组为有意义的超类别, 给定目标类别的单个示例, 模型可以有效地推断出目标类别属于哪个超类别, 从而实现单样本学习. 这些方法在最初为单样本学习提供了一个很好的思路, 但是实际问题中的数据往往更复杂, 仅依靠概率生成模型则无法更好地解决单样本识别的实际问题.

单样本学习一直和元学习关系紧密. 元学习的目标是通过大量的训练任务学习到内在的元知识, 从而能够快速处理新的同类任务, 这和单样本学习的目标设定是一样的. 这些元学习的算法在单样本或小样本的场景下也都取得了很好的效果. 如

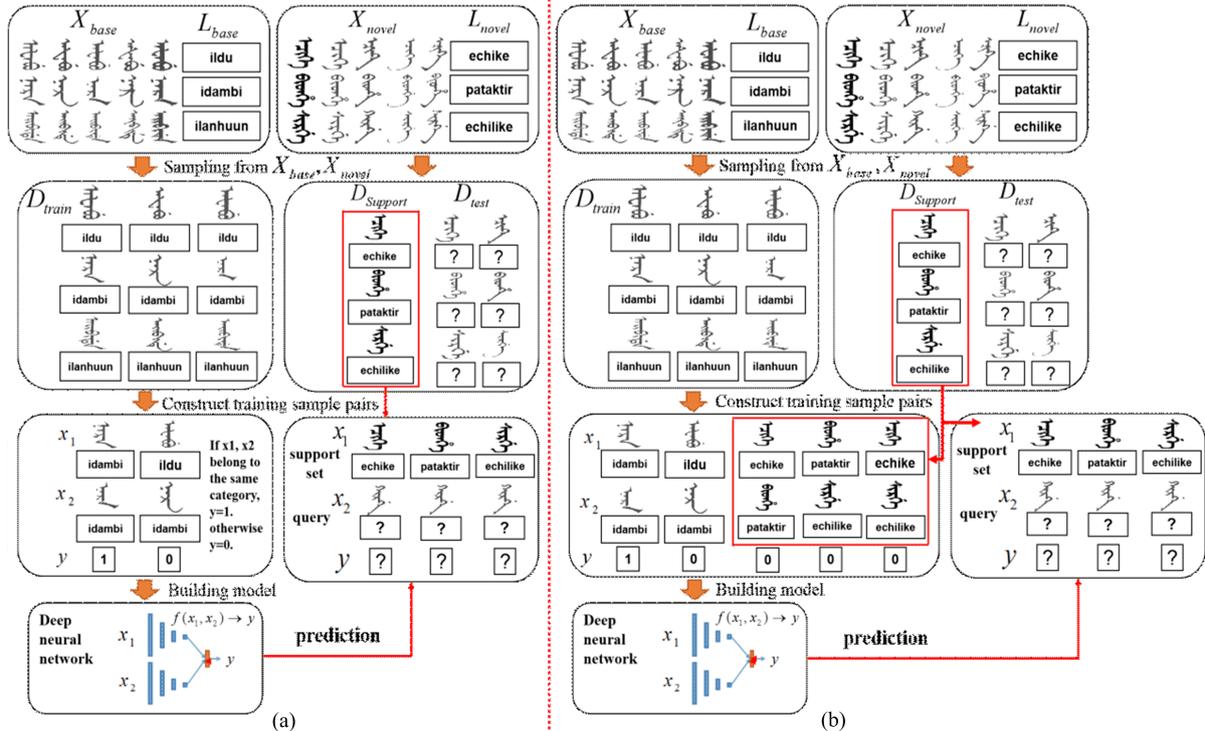


图 1 (a) 为传统算法训练过程, (b) 为本文算法的训练过程

Fig. 1 (a) is the training process of traditional algorithms, (b) is the training process of the proposed algorithm

Mishra 等<sup>[14]</sup>提出了一类简单而通用的元学习架构,它创新性地使用了时间卷积和软注意力的组合,取得了很好的效果. Ravi 等<sup>[15]</sup>提出一个基于 LSTM 的元学习模型来学习一个优化方法,用来实现小样本分类器. Kim 等<sup>[16]</sup>提出了一种新的基于贝叶斯的与模型无关的元学习方法. Finn 等<sup>[17]</sup>提出的 MAML 只学习参数的初始值并且只使用 SGD 进行优化,它还适用于监督学习和强化学习任务. Reptile<sup>[18]</sup>类似于 MAML 但仅使用一阶梯度进行优化. Amirreza 等<sup>[19]</sup>提出了一种新的基于双分支方法来实现单样本语义图像分割.

度量学习在单样本学习领域的应用大多是在元学习的框架下构造一种端到端训练的小样本分类器并对样本间的距离进行建模,使得同类样本靠近,异类样本远离. 例如 Oriol 等<sup>[8]</sup>提出了用以解决单样本识别问题的匹配网络,作者使用了 attention 机制来度量测试样本与支持集样本的匹配度,并用类似于加权求和的方式计算测试样本的标签,取得了很好的效果. Cai 等<sup>[20]</sup>在匹配网络的基础上提出了记忆匹配网络以解决单样本分类问题,该网络通过记忆模块有效利用全部支持集的信息,从得到更好的识别效果. Jake 等<sup>[21]</sup>提出了原型网络 (prototypical networks),该方法通过学习一个度量空间,在度量空间中每个类别都存在一个聚集在某个原型表达周围的特征. 以这种形式把分类问题看作是度量空间中的最近邻问题. Sun 等<sup>[22]</sup>提出了一种双分支关系网络 (RN),该网络不再使用单一且固定的距离度量方式,而是训练一个网络来学习距离的度量方式. 相较于早期的概率生成方法,度量学习的方法目前更受青睐.

## 2 学习算法

本节将详细介绍本文提出的训练过程中使用支持集信息的单样本学习算法. 2.1 节给出了算法的基本思想 and 与传统模型的对比. 2.2 节描述了孪生网络的网络结构. 2.3 节描述了算法的损失函数. 2.4 节描述了算法训练过程中的一些细节. 2.5 节给出了关于算法训练过程的伪代码.

### 2.1 基本框架

设  $X_{\text{base}}$  为源类别样本空间,  $X_{\text{novel}}$  为目标类别样本空间. 利用有标签的数据集  $X_{\text{base}}$  进行训练,实现对  $X_{\text{novel}}$  中无标签样本的识别. 本文从  $X_{\text{base}}$  采样训练集  $D_{\text{train}}$ , 从  $X_{\text{novel}}$  采样测试集  $D_{\text{test}}$  和支持集  $D_{\text{support}}$ . 训练集给定为  $D_{\text{train}} = \{(x_i, T_i) \mid 1 \leq i \leq N\}$ , 其中  $x_i$  为第  $i$  类的样本,  $T_i$  为第  $i$  类样本所对应的标签. 测试集  $D_{\text{test}}$  由  $X_{\text{novel}}$  中无标签样本组成, 支持集由  $X_{\text{novel}}$  中有标签样本组成, 支持集给定为  $D_{\text{support}} = \{(z_k, S_k) \mid 1 \leq k \leq m\}$ , 其中  $z_k$  为目标类别中已标注的单个样本,  $S_k$  为样本  $z_k$  所对应的标签. 与以往有关单样本学习的研究相比, 本文不仅将支持集用于分类目标类别中未标注的样本, 还将支持集样本加入孪生网络模型的训练中, 以增加模型对支持集中不同类别样本间的区分能力. 本文在构建网络时利用孪生网络进行建模, 分别取训

练集样本和支持集样本进行训练, 具体如图 1 所示. 在训练时训练集样本和支持集样本都被组织成样本对的形式进行训练, 并使训练集样本和支持集样本分批次进行训练.

### 2.2 网络结构

关于网络结构, 孪生神经网络中的两个相同的卷积神经网络和参数共享机制既能保证对不同类别图片可以较好地实现区分, 又能保证两个相同类别的图像不会被映射到特征空间中相距较远的位置, 从而准确地完成识别. 单个卷积神经网络由 4 个具有不同大小卷积核和固定步长为 1 的卷积层构成, 具体如图 2 所示. 在每个卷积层中我们经由 ReLU 激活函数来输出特征映射, 接着在每个卷积层后使用滤波器大小和步长都为 2 的最大池化层. 在最后一个卷积操作完成后, 用一个全连接层将卷积产生的张量变为一个特征向量, 孪生神经网络中的所有全连接层使用的激活函数都为 Sigmoid 激活函数. 样本  $x_{1_i}, x_{2_i}$  经过两个相同的卷积神经网络得到特征向量  $h_{1_i}^1, h_{1_i}^2$ , 而后再将  $h_{1_i}^1, h_{1_i}^2$  输入基于  $L_1$  距离的度量层来计算样本对的相似性, 并通过节点数为 1 带有 Sigmoid 激活函数的全连接层将输出映射到区间  $[0, 1]$  之内. 样本对  $x_{1_i}, x_{2_i}$  经孪生网络的输出越接近 1, 则他们是相同类别样本的可能性越大. 孪生神经网络的输出为  $p = \sigma(W | h_{1_i}^1 - h_{1_i}^2 | + b)$ , 其中  $\sigma$  为 Sigmoid 激活函数,  $W$  为最后一个全连接层的权重,  $b$  为其偏置.

### 2.3 损失函数

针对二分类问题本文使用带正则项的交叉熵损失函数作为优化目标, 将样本对  $x_{1_i}, x_{2_i}$  输入孪生神经网络进行训练. 为避免出现过拟合, 本文为损失函数添加了  $l_2$  正则化项,  $l_2$  正则化直接在损失函数的基础上加权重参数的平方和, 其中  $\lambda$  是正则化参数.  $l_2$  正则化作用在模型的权值上, 可以使模型尽可能的简单, 并在一定程度上避免过拟合. 具体损失函数为

$$L_{\text{train}} = -\frac{1}{M} \sum_{i=1}^{M_1} [y(x_i^1, x_i^2) \log p(x_i^1, x_i^2) + (1 - y(x_i^1, x_i^2)) \log(1 - p(x_i^1, x_i^2))] + \lambda^T |W|^2 \quad (1)$$

式中,  $y(x_i^1, x_i^2)$  是样本对标签, 当  $x_i^1$  和  $x_i^2$  是来自同一个类别图片时  $y(x_i^1, x_i^2) = 1$ , 当  $x_i^1$  和  $x_i^2$  不是同一个类别的图片时  $y(x_i^1, x_i^2) = 0$ , 其中  $x_i^1, x_i^2 \in D_{\text{train}}$ .

本文额外利用支持集样本来进行模型的训练. 训练时先将支持集中不同类别样本组成为样本对, 然后将其输入孪生神经网络来进行训练. 针对支持集训练孪生神经网络加大其不同类别样本间的度量, 以这种方式提高模型在当前支持集下进行单样本识别任务时的精确度. 针对支持集样本的损失函数为

$$L_{\text{support}} = -\frac{1}{M_2} \sum_{i=1}^{M_2} \log(1 - p(z_i^1, z_i^2)) + \lambda^T |W|^2 \quad (2)$$

式中,  $z_{1_i}, z_{2_i} \in D_{\text{support}}$ .

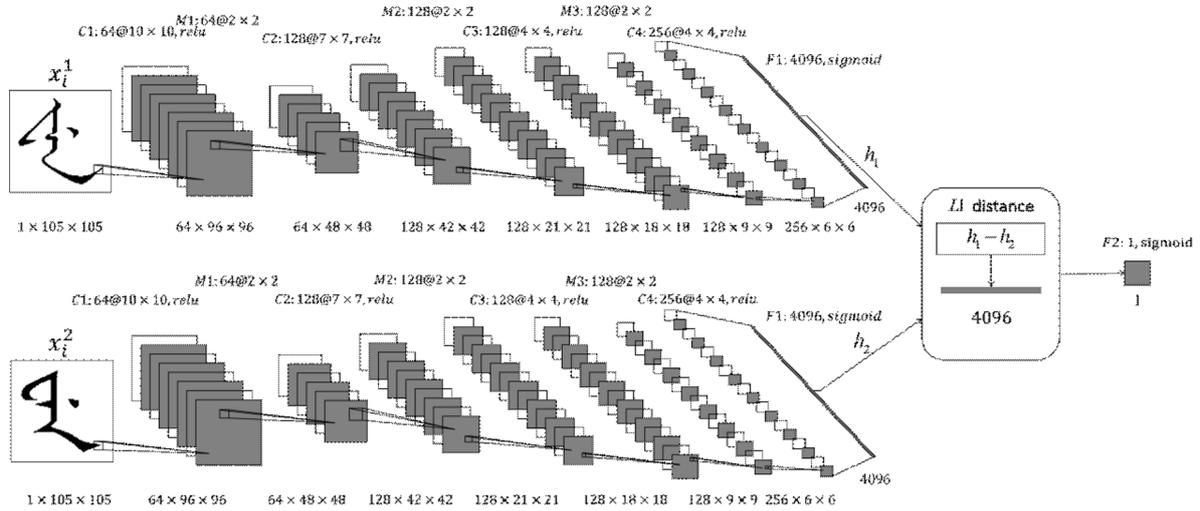


图 2 孪生网络结构

Fig. 2 The structure of the Siamese neural network

在训练模型时,我们分批次取训练集和支持集两个数据集的样本交叉的进行模型的训练.首先选取批次大小为  $M_1$  的训练集样本进行模型的训练,再选取批次大小为  $M_2$  大小的支持集样本进行训练.以这种方式迭代训练模型,总的损失为针对训练集样本的损失加上支持集样本的损失,即

$$L_{\text{total}} = L_{\text{train}} + L_{\text{support}} \quad (3)$$

#### 2.4 训练细节

本节对算法训练过程中的细节进行描述.首先是权值的初始化,初始化模型中所有卷积层中的权重为符合零均值和标准差为  $10^{-2}$  的正态分布.偏置的初始化也符合正态分布,但其平均值为 0.5,标准差为  $10^{-2}$ .我们对模型进行了 100000 次的迭代训练,每 500 次迭代对当前模型的参数进行验证.此外模型的验证是在 20-way 1-shot 的场景下进行的.如果当前模型的识别率大于上一次验证的识别率,就将当前模型参数保存下来,否则就继续进行训练,这样在 100000 次迭代完成后可得到准确率较高的模型参数.如果模型在经历 20000 次迭代后仍没有更新参数,则代表模型已经找到最优参数,模型将自动停止训练.在训练中模型使用 Adam 优化器进行优化并设置学习率为 0.00006.实验中在学习率为 0.00001 到 0.00009 之间分别训练模型,实验结果证明,学习率对模型影响不大.

#### 2.5 训练过程

在模型的训练过程中,分别取支持集和训练集的样本进行训练,利用支持集提供的信息加大模型对支持集中不同类别样本间的度量.训练和参数更新过程见算法 2.1.

##### 算法 2.1 算法训练过程中参数更新的伪代码

孪生网络训练中的参数更新过程.其中  $D_{\text{train}} = \{(x_1, y_1), \dots, (x_N, y_N)\}$  为训练集,  $D_{\text{support}} = \{(z_1, L_1), \dots, (z_s, \dots, L_s)\}$  为支持集;  $N$  是训练的迭代轮数;  $\text{RANDOMSAMPLE}(A, B)$  是从  $A$  中随机采样数量为  $B$  的样本;  $\text{ALLSAMPLE}(A)$  是将  $A$  中的全部采样并打乱顺序;  $M((x_1, x_2), \theta)$  是  $x_1, x_2$  两个样本经参数为  $\theta$  的孪生神经网络

计算后得到的值.

Input: Training set  $D_{\text{train}}$ , Support set  $D_{\text{support}}$ ;

Output: Parameters of Siamese network  $\theta$ ;

$J \leftarrow 0$

for  $k$  in  $N$  do:

$B_{\text{train}} \leftarrow \text{RANDOMSAMPLE}(D_{\text{train}}, m)$ .

$B_{\text{train}} = \{(x_1^i, x_2^i), y_i \mid 1 \leq i \leq m\}$ . If  $x_1^i$  and  $x_2^i$  are of the same category,  $y = 1$ . Otherwise  $y = 0$  \* /

$B_{\text{support}} \leftarrow \text{ALLSAMPLE}(D_{\text{support}})$ .

$B_{\text{support}} = \{(z_1^i, z_2^i), 0 \mid 1 \leq i \leq s\}$ .  $z_1^i$  and  $z_2^i$  are of the different category \* /

for  $(x_1, x_2, y)$  in  $B_{\text{train}}$  do:

$J = y \log M((x_1, x_2), \theta) + (1 - y) \log(1 - M((x_1, x_2), \theta)) + \lambda^T |W|^2$

$\theta \leftarrow \eta \nabla_{\theta} J((x_1, x_2), \theta)$  \* /  $\eta$  is the update step. \* /

end for

for  $(z_1, z_2)$  in  $B_{\text{support}}$  do:

$J = \log(1 - M((z_1, z_2), \theta)) + \lambda^T |W|^2$

$\theta \leftarrow \eta \nabla_{\theta} J((z_1, z_2), \theta)$

end for

end for

### 3 实验与结果

本节使用本文提出的方法构建利用支持集信息的单样本学习算法,并且在 2 个数据集 (Omniglot 数据集和满文数据集) 上进行验证.

#### 3.1 实验设置

我们将算法在 Omniglot 数据集和满文数据集上进行验证并与单样本学习领域内一些有代表性的算法进行比较.在 Omniglot 数据集和满文数据集上的实验都是以  $N$ -way 1-shot 的形式进行,即利用目标类别中每类样本仅有的一个带标注的样本组成有  $N$  类样本的支持集,并以其他未标注的样本进行测试.最后比较本文提出的算法与其他算法的结果.在进行训练时对孪生网络<sup>[7]</sup>,匹配网络<sup>[8]</sup>这两个算法的参数均按照其原文中给定的默认值设定.关于数据集的划分如表 1 所示,所有算法均采用相同的划分方法进行测试.

表 1 数据集划分说明  
Tab. 1 Description of data set partition

Data set	The number of categories and samples in the train set	The number of categories and samples in the test set
Omniglot	1200(category)	423(category)
	24000(sample)	8460(sample)
Manwen	366(category)	300(category)
	4133(sample)	3600(sample)

3.2 在 Omniglot 数据集上的实验结果

Omniglot 数据集是小样本学习领域中十分受欢迎的基准数据集之一. Omniglot 数据集包含 50 个字符表的例子, 这些例子包含拉丁文、韩文等国际语言以及国内一些地方方言, 此外还包含一些虚拟的字符集等. Omniglot 数据集中每个字符表中的字符类数在 15 个字符到 40 个字符之间. 我们重新划分了原本的数据集, 将其分为训练集和测试集. 其中训练集中包含 1200 类字符, 测试集包含 423 类字符. 训练集用于训练模型学习特征映射. 测试集样本用于测试模型的单样本识别性能. 在构建测试集时在每类字符中选取一个样本作为支持集样本构建支持集并将测试集中每类字符随机打乱, 使结果更有随机性.

测试时随机选取支持集样本以外的未标注样本作为测试样本, 将测试样本和当前支持集中每类样本分别组成样本对进行一次查询, 每次测试从测试集产生 1000 个 episodes, 并求出其识别率. 最后进行 10 组训练, 每组训练重新选取支持集样本并求出算法精度的平均值作为最终的结果. 将本文提出的算法与两种代表性的方法进行比较, 其结果如表 2 所示. 对于匹配网络<sup>[9]</sup>在 200-way 1-shot 场景下的实验, 我们在该场景下重新训练匹配网络模型, 但是产生了模型不收敛的问题, 所以在此处数据为空.

表 2 在 Omniglot 数据集上的识别精度对比  
Tab. 2 Comparison of recognition accuracy on Omniglot dataset

Algorithm	20-way 1-shot	200-way 1-shot
Siamese Neural Networks <sup>[7]</sup>	88.2%	53.8%
Matching Networks <sup>[8]</sup>	93.8%	—
Ours	94.2%	61.2%

3.3 在满文数据集上的实验结果

满文数据集中的所有数据分别从《满语 365 句》和《新满汉大辞典》两本书中扫描得到, 共计 666 个满文单词, 其中每个满文单词都有多种字体, 每个满文单词称其为一类. 具体在满文数据集中有 259 类满文单词有 11 种字体, 另外 407 类满文单词有 12 种字体, 其中 11 种字体来自《满语 365 句》, 其余来自《新满汉大辞典》. 其具体构成如图 3 所示, 其为

测试集中某一类满文单词的所有样本. 满文数据集中的所有字符图片均为 105×105 大小, 对于部分满文单词图像宽或高大于 105 的情况, 需要把字符图像的宽和高按照图像原比例缩放放在 105 像素以内. 满文数据集是一个典型的小样本数据集, 每类字符的样本数量极少, 然而整个数据集中字符的类别很多. 在这种场景下想要较好地完成字符识别工作, 以往的深度学习算法是不适用的, 而小样本学习算法能给我们解决这个问题带来帮助. 本文使用利用支持集信息的孪生网络算法完成满文字符的识别工作.



图 3 同一类别下满文单词的不同字体的图像示例  
Fig. 3 Image examples of different fonts of Manchu words in the same category

在训练时, 进一步将数据集划分为训练集和测试集. 训练集包含 366 类满文单词, 其中具有 11 种字体的满文单词为 259 类, 有 12 种字体的满文单词为 107 类, 训练集共有 4133 个图片样本. 训练时具有 11 种字体的 259 类满文单词用于训练, 具有 12 种字体的 107 类满文单词用于验证. 其余 300 类满文单词作为测试集, 测试集中每类满文单词包含 12 个图片样本对应 12 种字体, 测试集样本总数为 3600 个. 测试时取《新满汉大辞典》中字体的图片样本为支持集, 其余用作测试图片来测试算法性能, 具体测试方法与 Omniglot 数据集相同. 我们进行了 10 组测试求出识别率的均值, 在满文数据集上依然与上述 2 种方法进行比较, 结果如表 3 所示. 本文提出的方法在 20-way 1-shot 的场景下有 97.2% 的识别率. 为验证在满文字符类别较多的场景下算法的性能, 我们在 200-way 1-shot 的场景下也进行了实验, 结果显示算法依然有 87.4% 的识别率. 在 200-way 1-shot 的场景下对匹配网络<sup>[9]</sup>进行训练产生了与上述 Omniglot 实验中相同的问题, 所以此处数据为空.

表 3 在满文数据集上的识别精度对比  
Tab. 3 Comparison of recognition accuracy on Manchu dataset

Algorithm	20-way 1-shot	200-way 1-shot
Siamese Neural Networks <sup>[7]</sup>	89.6%	77.4%
Matching Networks <sup>[8]</sup>	95.7%	—
Ours	97.2%	87.4%

## 4 结论

本文建立了一种在训练过程中使用支持集信息的单样本识别算法,我们利用孪生神经网络构建模型并在训练时加入支持集的信息,加大支持集样本类别之间的相似性度量.在 Omniglot 数据集上的实验结果表明,以这种方法训练模型能得到更好的识别效果,并且在支持集样本类别较多时能使识别率有更多的提升.本文还将此算法应用在满文识别上,实验结果证明了本文中提出的算法在满文数据集中应用的可行性,特别是在 200-way 1-shot 的复杂场景下,本文提出的方法识别率能达到 87.4%,模型依然有较好的泛化能力.

### 参考文献(References)

- [1] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C]// Proceedings of the International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 764-773.
- [2] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [3] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely connected convolutional networks [C]// Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, United states: IEEE, 2017: 2261-2269.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet classification with deep convolutional neural networks[J]. Communications of The ACM, 2017, 60(6): 84-90.
- [5] LI F F, ROB F, PIETRO P. One-shot learning of object categories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (4): 594-611.
- [6] LAKE B M, LEE C, GLASS J, et al. One-shot learning of generative speech concepts[J]. Cognitive Science, 2014, 36(36): 803-808.
- [7] GREGORY K, RICHARD Z, RUSLAN S. Siamese neural networks for one-shot image recognition [C/OL]// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR: W&CP, 2015. [2020-05-18]. <https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf>
- [8] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning [C]// Advances in Neural Information Processing Systems 29 - Proceedings of the 2016 Conference. Barcelona, Spain: Neural information processing systems foundation, 2016, 0(1): 3637-3645.
- [9] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory-augmented neural networks [C]// 33rd International Conference on Machine Learning. New York, United states: International Machine Learning Society, 2016, 4(4): 2740-2751.
- [10] CHEN Z, FU Y, WANG Y, et al. Image deformation Meta-networks for one-shot learning [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Long Beach, CA, United states: IEEE, 2019: 8672-8681.
- [11] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B. Human-level concept learning through probabilistic program induction[J]. Science, 2015, 350 (6266): 1332-1338.
- [12] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B, et al. One-shot learning by inverting a compositional causal process [C]// Neural Information Processing Systems 26, NIPS 2013. Lake Tahoe, United states: Neural information processing systems foundation, 2013: 1-9.
- [13] SALAKHUTDINOV R, TENENBAUM J, TORRALBA A, et al. One-shot learning with a hierarchical nonparametric Bayesian model [C]// Workshop on Unsupervised and Transfer Learning. Washington: IEEE, 2012: 195-207.
- [14] MISHRA N, ROHANINEJAD M, CHEN X, et al. A simple neural attentive meta-learner [C]// 6th International Conference on Learning Representations. Vancouver, Canada: IEEE, 2018: 1-17.
- [15] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning [C]// 5th International Conference on Learning Representations. Toulon, France: IEEE, 2017: 1-11.
- [16] YOON J, KIM T, DIA O, et al. Bayesian model-agnostic Meta-learning [C]// Proceedings of the Conference on Advances in Neural Information Processing Systems. Montreal, Canada: Neural information processing systems foundation, 2018, arXiv:1806.03836v4.
- [17] FINN C, ABBEEL P, LEVINE S, et al. Model-agnostic Meta-learning for fast adaptation of deep networks [C]// 34th International Conference on Machine Learning. Sydney, Australia: IMLS, 2017, 3 (8): 1856-1868.
- [18] NICHOL A, ACHIAM J, SCHULMAN J, et al. On first-order Meta-learning algorithms. [J/OL]. 2018, arXiv: abs/1803.02999.
- [19] SHABAN A, BANSAL S, LIU Z, et al. One-shot learning for semantic segmentation [C]// British Machine Vision Conference 2017. London, United Kingdom: BMVA Press, 2017, arXiv:1709.03410.
- [20] CAI Q, PAN Y, YAO T, et al. Memory matching networks for one-shot image recognition [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Salt Lake City, United states: IEEE Computer Society, 2018: 4080-4088.
- [21] SNELL J, SWERSKY K, ZEMEL R S, et al. Prototypical networks for few-shot learning [C]// Advances in Neural Information Processing Systems. Long Beach, United states: IEEE, 2017: 4078-4088.
- [22] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few-shot learning [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, United states: IEEE Computer Society, 2018: 1199-1208.