

小样本图像分类方法整理总结

一、小样本学习介绍

小样本学习可以理解成在样本量较少的情况下的深度学习。由于样本量较少的时候，往往深度学习模型会过拟合，所以一般不使用深度学习的模型，而小样本学习就是研究如何克服或者缓解这个问题而提出的。

举例说明，假设我们已经通过除狗以外的些许动物的样本训练出来了模型，现在想让模型识别新的类——狗。这需要将源知识（source）迁移到包含狗的知识领域来（target）。为了完成识别新类（狗）的问题，需要输入一张标好标签的样本用来泛化模型（如：牧羊犬），那么再来测试样本（如：哈士奇），我们就可以把这个哈士奇的样本的标签标出来为狗，达到分类的目的。这种学习叫做 one-shot learning，即单样本学习。同样的，如果刚才来的是一堆标好 label 的样本（除了牧羊犬，可能还有京巴、吉娃娃、泰迪等做支撑集），这种学习叫做 few-shot learning，即小样本学习，一般小样本学习标签数量不会大于 20。

小样本在图像分类上的应用主要分为以下几个方面：基于模型的小样本学习（Mode Based），通过模型结构的设计快速的在少量样本上更新参数，直接建立输入 x 和预测值 P 的映射函数；基于度量的小样本学习（Metric Based），通过度量训练集中的样本和测试集中样本的距离，借助最近邻的思想完成分类；基于元学习的小样本学习（Meta Based），该方法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。下面本文将对整理到的论文在这几个方面进行分析讨论。

二、基于模型的小样本学习（Mode Based）

2019 年提出的方法，《Unsupervised Meta-Learning for Few-Shot Image Classification》，是一种针对少样本的分类学习问题，提出的无监督式的元学习模型。本文提出一般的少样本学习方法都容易在数据分布的原生域（source domain）到目标域（target domain）的映射过程中出现局部偏移，从而导致学习效果不尽如人意。文章借此提出一种新的视觉结构限制来提升映射函数的泛化性，从而避免上述提到的偏移缺点，文章采用了新的训练策略，应用了提出的限制模块，在标准数据集上取得了不错的效果。

2019 年提出的方法，《Self-Supervised Learning For Few-Shot Image Classification》，同样是一种无监督式的学习方法。由于每个任务的样本数量有限，用于元学习的初始嵌入网络成为一个必不可少的组成部分，在实践中会在很大程度上影响性能。为此，人们提出了许多预训练方法，它们大多采用有监督的方式训练，对看不见的类转移能力有限。本文提出用自监督学习 (SSL) 来训练一个更通用的嵌入网络，它通过学习数据本身来为下游任务提供缓慢而健壮的表

示。

2019 年提出的方法，《Deep Kernel Transfer in Gaussian Processes for Few-shot Learning》，从另一个角度对少样本学习进行了分析，作者认为人类解决新问题的方式是做出远远超出现有信息的推论，重复使用之前学到的知识，在面对不确定性时权衡不同的选择。因此提出了一种简单而有效的深度核学习变体，其中内核在任务之间转移，我们称之为深度核转移。该方法实现简单，提供不确定性量化，并且不需要估计特定于任务的参数。

2018 年提出了一种小样本学习方法，《Learning Embedding Adaptation for Few-Shot Learning》，该方法来使嵌入模型适用于目标分类任务，从而产生特定于任务且具有区分性的嵌入。为此，本文使用了一种称为 Transformer 的自我注意机制，通过关注从测试实例到可见和不可见类中的训练实例的关联，将嵌入从任务不可知转换为特定于任务。在标准的小样本分类基准测试和四个扩展的小样本学习环境下（即跨域、导向式、泛化的少镜头学习和大规模的低镜头学习）验证了该模型的有效性。

2019 年提出了一种关于手写字符集的识别和分类方法，《TextCaps : Handwritten Character Recognition with Very Small Datasets》，它借助胶囊网络（Capsule Networks, CapsNets）解决了标注数据集太小的问题。仅通过操纵实例化参数，利用了 CapsNet 增强数据的能力。在本文的例子中，CapsNet 不仅识别了字符图像，还学习了它的属性。该架构以 Sabour 等人提出的 CapsNet 架构为基础，由胶囊网络和全连接解码器网络组成的。研究人员用（deconvolutional network）代替了解码器网络，同时还对胶囊网络做了一些小改动。通过给表征实体属性的实例化参数加入一些可控噪声，研究人员转换实体以表征现实中发生的实际变化。这样就产生了一种全新的数据生成技术，这种技术生成的数据会比基于仿射变换生成的增强数据更加逼真。

三、基于度量的小样本学习（Metric Based）

匹配网络：《Matching Networks for One Shot Learning》。2016 年提出的匹配网络采用了基于深度神经特征的度量学习和用外部记忆增强神经网络的最新进展的想法。该框架学习了一个网络，将一个小的有标签的支持集和一个无标签的示例映射到它的标签，从而消除了微调以适应新的类型的需要。在不改变网络模型的前提下能对未知类别生成标签，其主要创新体现在建模过程和训练过程上。对于建模过程的创新，文章提出了基于 memory 和 attention 的匹配网络，使得可以快速学习。

孪生网络：《Siamese neural networks for one-shot image recognition》，2015 年提出的孪生网络，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。双路的神经网络训练时，通过成对组合不同类的样本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。使用卷积结构，能够获得比其他深度学习模型更强的结果，在单次分类任务上具有接近最先进的性能。

原型网络：《Prototypical Networks for Few-shot Learning》。2017 年提出的原型网络是学习一个度量空间，在该度量空间中，可以通过计算到每个类的原型表示的距离来执行分类。与其他小样本学习的方法相比，它们反映了一种更简单的归纳偏向，在这种有限的制度下是有益的，并取得了很好的结果。

关系网络：《Learning to Compare: Relation Network for Few-Shot Learning》。2018 年 CVPR 上提出了关系网络，该网络能够通过计算图像与每个新类别的少数示例之间的关系分数来对新类别的图像进行分类，而无需进一步更新网络。

2018 年的 CVPR 提出的《Dynamic Few-Shot Visual Learning without Forgetting》，先用训练集训练出一个特征提取器，然后对于新的 few-shot 训练数据，通过一个 few-shot 分类权重生成器来生成对应的参数权重。这个权重生成器还把基本权重作为输入，使用余弦相似度来计算最后的概率输出，同时使用 attention 注意力机制来选择对应的基本权重。

2019 年提出的方法，《SimpleShot: Revisiting Nearest-Neighbor Classification for Few-Shot Learning》，本文发现简单的特征变换就足以获得具有竞争力的少样本学习精度。例如，作者发现与均值相减和 L2 归一化相结合使用的最近邻分类器在 MiniImageNet 数据集上的五个设置中有三个的性能优于先前的结果。

2018 年提出的方法，《TADAM: Task dependent adaptive metric for improved few-shot learning》，采用原型网络的思想，构造类别代表器来进行度量计算，文章做出如下三点改进方式：1、Metric Scaling: 在 similarity metric 的基础上学习一个 scaling factor，这样能够更好的使得输出的 metric 在合适的范围内；2、Task Conditioning: 构造一个 TEN network，通过输入样本数据来得到 task representation，并利用此作为 condition 来改变 feature extractor 的输出。简单的说就是使得每一个 task 的 feature extractor 都不一样，具备 adaptation 的能力；3、Auxiliary task co-training: 作者采用前面分析的几篇文章的做法，也使用所有训练数据来训练 feature extractor，并把这个作为一个辅助任务，这样能够提升特征提取能力。总的来说，这篇文章主要在细节上做了改进，但确实取得了挺不错的效果，非常值得借鉴的思想还是 task conditioning，这意味着我们应该构建更加 meta 的神经网络，使得最终得到的分类器的自适应能力足够强，这样更有希望使得效果更好。

2019 年 CVPR 上提出了一种图神经网络的方法，《Edge-labeling Graph Neural Network for Few-shot Learning》，少样本学习的难点通常在于可用样本有限，所以样本之间潜在的关联性在学习过程中就变得非常重。通用的前传网络很难捕捉到样本之间丰富的关联信息，图神经网络正好可以弥补这一点：图网

络在节点之间构建的丰富的连接,使得其能够利用节点之间的信息传递机制得到邻节点的信息并且进行聚合,通过一定复杂程度的图网络,就能够表达数据个体之间丰富的关联互动特征,在分类任务上实现了大幅度的性能提升。

2019 年的 CVPR 提出了一种在原型网络,图卷积的基础上进行改进的基于度量的方法,《Revisiting Local Descriptor based Image-to-Class Measure for Few-shot Learning》,该方法对于图像的每个局部特征都计算一个与 support set 图像的相似性,利用 CNN 提取特征,得到 query 和 support 的图像的特征 $\{x_1, x_2, x_3, \dots, x_m\}$,对于每个局部特征 x_i ,找到与 support set 图像所有局部特征最近邻的 k 个特征,计算与这 k 个特征的相似性,平均得到这个局部特征的相似性,所有局部特征相似性平均得到这个图像对于这类的相似性,利用原型 loss 优化整个网络。

四、基于元学习的小样本学习 (Meta Based)

元学习算法最经典的模型是 2017 年提出的 MAML,《Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks》,该算法训练模型使得它在各种任务都能取得较好的效果,然后再针对我们特定的任务进行 fine tuning。该方法的基础是:训练模型初始参数,从而使模型在经过新任务的少量数据的梯度步骤更新后有最好的表现。学习器会在元学习阶段学习一系列的任务,从而可以在新任务中快速适应。

此后有很多算法在 MAML 的基础上进行改进,2018 年提出的《Reptile:On First-Order Meta-Learning Algorithms》就是其中之一。与 MAML 类似,Reptile 是一种寻找神经网络初始化参数的元学习方法,在新任务上只需要少量数据进行 fine-tune。与 MAML 不同的是,Reptile 只需要简单地每个任务上执行 SGD 而不需要像 MAML 一样计算两次微分。这使得 Reptile 消耗的计算量和内存更少。在 2020 年的 ICLR 中进一步的对 MAML 进行了分析,《Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML》,作者通过对快速学习和特征重用的讨论和分析,发现测试任务的有效性完全取决于学习特征的质量,元学习的初始化过程如果包含了高质量的特征,则可以去除 MAML 的某些过程,从而提出了一个简化的算法。

2018 年提出了一种基于 MAML 的改进方法,《Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace》,该方法不但迁移了参数作为初始值,同时迁移了每层的激活。这个思路应该是借鉴了知识蒸馏 (knowledge distillation) 的相关工作。更重要的是其中关于相似度度量的定义,该方法是直接学习一个新的适合各个任务进行比较的度量空间。本文证明了这个学习子空间的维度反映了特定于任务的学习者适应任务的复杂性,并且与以前的基于梯度的元学习方法相比,我们的模型对初始学习率的选择不那么敏感。在小样本分类和回归任务上获得了最先进或相当的性能。

2019 年的 CVPR 同样有一篇基于 MAML 的改进方法,《Meta-Transfer Learning for Few-Shot Learning》, 本文在 MAML 的基础上使用了一个较深的网络, 不同于直接在预训练的模型上进行全局参数的 fine tuning, 本文在预训练模型的参数固定的前提下, 对预训练模型的每层参数重新学习一个 scale 和 shift, 在保证预训练模型不损失 general 的特性的前提下, 重新训练了参数, 减小了模型参数, 使得预训练模型可以 transfer 到新任务上, 同时, 比 fine tuning 的方式有更少的参数, 可以避免过拟合。

2018 年提出的方法也是基于 MAML 的改进, 《A Simple Neural Attentive Meta-Learner》, 该方法提出了一类简单而通用的元学习器体系结构, 它使用时间卷积和软注意的新组合; 前者从过去的经验中聚集信息, 后者定位特定的信息片段。在迄今为止最广泛的一组元学习实验中, 我们在几个重基准任务上评估了由此产生的简单神经注意力学习者。在所有任务中, 无论是在监督学习还是强化学习中, 该方法都以相当大的优势获得了最先进的性能。

2019 年提出的基于 MAML 的改进的方法, 《Meta-Learning with Implicit Gradients》, 该方法利用隐式微分, 提出了隐式 MAML 算法, 该算法只依赖于内层优化的解, 而不依赖于内环优化器所走的路径。这有效地将亚梯度计算与选择内循环优化器解耦。因此, 我们的方法与内循环优化器的选择无关, 并且可以优雅地处理许多梯度步骤, 而不会消除梯度或内存限制。

2019 年的 ICCV 会议提出了一种新型的小样本学习方法, 《Few-Shot Learning with Global Class Representations》, 在小样本学习问题中, 作者假设有一组基类, 以及一组新类。每个基类具有足够的训练样本, 而每个新类只有少量标记样本。小样本学习的目的是通过从基类转移知识来学习识别具有少量标注样本的新类别。本文提出一种基于全局类别表征的小样本学习方法, 称作广义小样本学习方法。传统小样本学习的方法, 通常只使用基类数据进行学习。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, 这一点在广义小样本问题中尤为突出。通过在训练阶段引入新类的数据, 我们同时对基类和新类学习全局类别表征, 并利用样本生成策略解决类别不均衡问题, 有效防止训练模型在基类数据中出现过拟合的现象, 从而提高了模型泛化到新类的能力。具体思路是将每个类 (包括基类和新类) 表示为特征空间中的一个高维向量 (称为类别表征), 然后将测试图片到各个类的表征的距离进行比较, 来对测试图片进行分类。我们同时使用基类与新类的 “所有训练样本” 来学习这种类别表征, 因此称这种表征为全局类别表征。

2019 年提出的方法, 《Charting the Right Manifold: Manifold Mixup for Few-shot Learning》, 也是独辟蹊径, 从另一个角度去考虑小样本学习问题。作者分析当前小样本学习算法的目标是学习模型参数, 使其能够仅借助几个标记的样本就能适应看不见的类。最近在其他领域有一种正则化技术——流形混合 (Manifold Mix-up), 专注于学习一种通用的表示法, 对数据分布中的微小变化

具有鲁棒性。由于小镜头学习的目标与鲁棒表示学习密切相关，因此作者在这个问题背景下研究了流形混合(Manifold Mixup)。利用自我监督和正则化技术研究学习相关特征流形在小样本任务中的作用。通过自我监督技术丰富的特征流形的正规化，以及流形混合显著地改善了小样本学习的性能。

2019 年的 CVPR 提出的元学习方法，《Meta-Learning With Differentiable Convex Optimization》，对于基于元学习方法的少样本学习，文中研究线性分类器作为基学习器；给定标注的训练样本集，线性 SVM 被用来学习分类器，泛化误差由相同任务的新样本集计算。由于最小化不同任务的泛化误差的元学习目标，要求在内部以循环优化的方式来训练线性分类器，因此可计算性是一个关键的挑战；但是线性模型的目标函数是凸的，可以有效地被解决。在少量样本环境下，凸性允许有效的元学习，它引出两个额外属性：（优化的）隐式可微和（分类器的）低秩特性。第一个属性允许使用离线凸优化估计最优值，并隐式地微分最优性条件或 KKT 条件来训练嵌入模型。第二个属性意味着对少量样本学习而言，对偶形式中待优化的变量数量远远小于特征维数。

五、相关数据集上的 SOTA 方法

在 Mini-ImageNet - 1-Shot Learning 和 Mini-ImageNet - 5-Shot Learning 数据集上，2019 年 CVPR 上提出的基于度量的《Self-Supervised Learning For Few-Shot Image Classification》实现了最佳的效果，前者达到了 76.82% 的准确率，比第二名领先了 6.51%，后者达到了 90.98% 的准确率，比第二名领先了 3.28%。

在 CUB 200 5-way 1-shot 和 CUB 200 5-way 5-shot 数据集上，2019 年提出的基于模型的《Charting the Right Manifold: Manifold Mixup for Few-shot Learning》实现了最佳的效果，前者达到了 80.68% 的准确率，比第二名领先了 3.6%，后者达到了 90.85%，比第二名提高了 1.67%。

在 OMNIGLOT - 1-Shot Learning 数据集上，2018 年 ICML 上提出的基于元学习方法的《Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace》实现了最佳的效果，达到了 99.5% 的准确率，比第二名领先了 0.34%。

在 OMNIGLOT - 5-Shot Learning 数据集上，2017 年 ICML 上提出的基于元学习方法的《Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks》实现了最佳的效果，达到了 99.9% 的准确率，比第二名领先了 0.16%。

在 CUB-200 - 0-Shot Learning 数据集上，2017 年 NeurIPS 提出的基于度量学习的方法《Prototypical Networks for Few-shot Learning》实现了最佳的效果，达到了 54.6% 的准确率，比第二名提高了 4.5%。

在 OMNIGLOT - 1-Shot, 20-way 数据集上，2019 年提出的 ICCV 提出的基于元学习方法的《Few-Shot Learning with Global Class Representations》实

现了最佳的效果，达到了 99.63% 的准确率，比第二名提高了 16.3%。

在 OMNIGLOT - 5-Shot, 20-way 数据集上，2019 年提出的 ICCV 提出的基于元学习方法的《Few-Shot Learning with Global Class Representations》实现了最佳的效果，达到了 99.32% 的准确率，比第二名提高了 5.2%。

在 mini-ImageNet - 100-Way 数据集上，2019 年提出的 ICCV 提出的基于元学习方法的《Few-Shot Learning with Global Class Representations》实现了最佳的效果，达到了 39.14% 的准确率，是该数据集上测试的唯一算法。

在 Fewshot-CIFAR100 - 10-Shot Learning 数据集上，2019 年 CVPR 上提出的基于元学习方法的《Meta-Transfer Learning for Few-Shot Learning》实现了最佳的效果，达到了 63.4% 的准确率，是该数据集上测试的唯一算法。

在 Fewshot-CIFAR100 - 1-Shot Learning 数据集上，2019 年 CVPR 上提出的基于元学习方法的《Meta-Transfer Learning for Few-Shot Learning》实现了最佳的效果，达到了 45.1% 的准确率，是该数据集上测试的唯一算法。

六、小样本分类方法的比较

2019 年 ICLR 提出了一篇对当前小样本学习方法进行性能比较和阐释的文章，《A Closer Look at Few-shot Classification》，文章指出在小样本学习方面，近几年各种方法层出不穷，模型结构和学习算法也越来越复杂。然而这些方法之间没有统一的框架下进行比较。这篇文章针对几个关键问题，如数据集对小样本学习方法的影响、网络深度对这些方法的性能的影响以及领域漂移下这些方法的表现，对当前的小样本学习方法进行性能比较和阐释。具体地，本文有以下三个主要贡献：1、对集中已有的方法做了比较，特别地，实验了基础模型的能力对这些方法的性能的影响，结果显示，在领域差异比较小的情况下（如 CUB 这种细粒度分类任务），随着基础即特征提取神经网络的能力的提高（从四层的 CNN 提升到 resnet），这些方法之间的性能差异越来越小；相反地，在领域差异比较大的情况下（如 miniImageNet），随着基础即特征提取神经网络的能力的提高，这些方法的性能差异越来越大。这一点在第五章 SOTA 的展示中也能看出来；2、文章建立了两个普通简单的 baseline，发现在 CUB 和 miniImageNet 上的性能足以和当前最先进的基于元学习的方法媲美；3、基于 1 的结果和一些已有的研究，作者特别强调了小样本学习任务中的领域自适应问题，并且设计实验显示当前这些 state-of-the-art 的小样本学习方法在领域漂移的情况下表现相当不好，没有 baseline 表现好，提醒之后的研究人员对这个方向多关注。

本文没有提出什么先进的模型，但是却进行了一些有意义的实验，提出了几个大家容易忽略的问题，对小样本分类这一问题有了更加深刻的认识，也指出了一些存在的问题，因此是比较有价值的。特别地，文章强调了特征提取网络的能力、数据集的差异性以及领域自适应问题对小样本学习任务的影响，特别强调了应该关注小样本学习中的领域自适应问题。这一研究也启示我们在研究过程中应

该全面考虑问题，自己得到的模型性能也很大程度上受限于数据集、基础模型等问题的影响，在其他条件下，可能最简单的模型最具有竞争力。

七、分析与总结

以上总共是从 2017 年至 2020 年近三年的 25 篇关于小样本图像分类的文章，其中有 5 篇是基于模型的小样本学习方法，有 9 篇是基于度量的小样本学习方法，有 10 篇是基于元学习的小样本学习方法，有 1 篇是方法的比较和讨论。从年份上来看，2019 年共有 13 篇关于小样本图像分类的文章，说明该方向在近年来得到了国内外研究人员的广泛关注。

基于模型的小样本学习都是基于具体问题，借助其他方法和思想，对原有的模型进行改进或者提出新的模型直接建立输入 x 和预测值 P 的映射函数。可借鉴的方法不多，不建议直接入手进行学习。

基于度量的方法近几年在原型网络、孪生网络、匹配网络、关系网络的基础上，进一步的研究更好的度量方法去增加分类的准确率。事实证明，它对于解决少样本分类任务非常有用：度量学习算法不必在支持集（少量的带标签图像）上进行微调，而是通过将查询图像与带标签图像进行比较来对其进行分类。度量方法可以在聚类等其他领域中获取灵感，从而应用于小样本分类方面，该类方法可以做更进一步的研究，可能会有一些突破。

至于基于元学习的方法，大多是在最基础的模型无关的元学习方法 MAML 上进行改进，它基本上是最纯粹的元学习形式，通过神经网络进行两级反向传播。在元训练期间，MAML 学习初始化参数，这些参数允许模型快速有效地适应新的少样本任务，其中这个任务有着新的、未知的类别。相比于度量学习算法，该模型的训练难度很大，超参数搜索更为复杂。此外，mate 的反向传播意味着需要计算梯度的梯度，因此你必须使用近似值来在标准 GPU 上进行训练。但是，模型无关的元学习的模型是不可知的。这意味着它几乎可以应用于任何神经网络，适用于任何任务。掌握 MAML 意味着只需少量样本就能够训练任何神经网络以快速适应新的任务。MAML 的作者 Chelsea Finn 和 Sergey Levine 将其应用于有监督的少样本分类，监督回归和强化学习。但是通过想象和努力研究，可以用它把任何一个神经网络转换成少样本有效的神经网络，这也是此后基于 MAML 的改进算法百花齐放的原因。

参考文献

[见 EXCEL 附件]