基于多种数据集的自监督学习的少样本图像分类

左宗侑 李丹
四川大学锦城学院 四川 成都 610065

【摘 要】少样本图片分类的目的是用有限的标记样本对没有标记的类别进行分类。由于是对每个任务的样本数量限制，元学习的初始嵌入网络是元学习的重要组成部分并且在实际应用中会对元学习的性能产生很大的影响。因此，许多预先训练的方法被人们提了出来，其中大部分分类都是在监督的方式下进行训练的。在本文中，我们提出训练一个更为广义的嵌入网络与通过从数据本身学习来为下游任务提供表示的自监督学习 (SSL)。我们通过在 1-shot 和 5-shot 的参数下对多种数据集训练的结果进行比较来评估我们的工作。

【关键词】自监督学习；少样本学习；嵌入网络；图像分类

1 介绍

近年来，深度学习技术在很多领域都有显著的突破和成就。其中很重要的一点在于可以从大量标记的数据集中保存模型。这在某种程度上违背了人类的学习行为——一个人可以通过有限的先验知识从几个例子中轻松地对对象进行分类。如何对这种行为进行计算建模是近年来对少样本学习的研究，重点是如何使模型适应于实例数量有限的新数据。

少样本分类的一个常用解决方案是应用元学习过程，将整个图片数据集划分为不相同的元任务子集，学习如何按照任务变换对模型进行调整。但元学习很容易导致过拟合，因为每个类只有很少的样本数。针对这一问题，我们提出了一种通过学习嵌入标记样本来对未标记样本进行良好分类的机制。通过对类和其中的相关数据进行子抽样来模拟少样本任务。近期的元学习方法注重于从数据集中检索可转移的嵌入，以及图像和它们的类描述之间的关系。这是通过将训练分解成两个阶段来完成的，即 :(1) 学习鲁棒的、可转移的嵌入，(2) 对学习的嵌入进行细化微调，用于下游分类任务。这些研究表明，一个鲁棒的预先训练的嵌入网络对于少样本图像分类任务是至关重要的。

在本文中，我们应用了一个更大的具有自监督学习 (SSL) 的嵌入网络，与元学习相结合。在评价中，在多种数据集下，该方法可以比基线方法显著提高了少样本图像的分类性能。

2 相关工作

少样本学习作为一个活跃的研究课题，已经得到了广泛的研究。而一种常见的少样本学习策略是通过元学习 (也称为学习 – 学习) 和多辅助任务。其关键是如何有力地在不受训练数据过度拟合影响的情况下，加速学习网络的进展。基于元学习的另一种方法，试图学习一种能够有效地将输入样本投射到特定特征空间的深度嵌入模型。然后利用距离函数如余弦距离、欧氏距离等对样本进行最近邻分类。Koch 等人提出 Siamese 网络从输入图像中提取嵌入特征，并从同一类图像中聚合图像。匹配网络使用增广神经网络进行特征嵌入，形成度量学习的基础。在本文利用了 ProtNet 网络作为特征嵌入的网络。

自监督学习（SSL）是从没有类标签的数据本身中学习一种健壮性的表示。这里的主要任务是设置复杂的下游任务，以实现对解决后续任务有用的表示。这与少样本学习中的预训练嵌入网络的目的是相同的。AMDIM 学习从各方面的视图摘取的特征，这各方面的视图由重复运用在输入图形的数据扩增中产生。

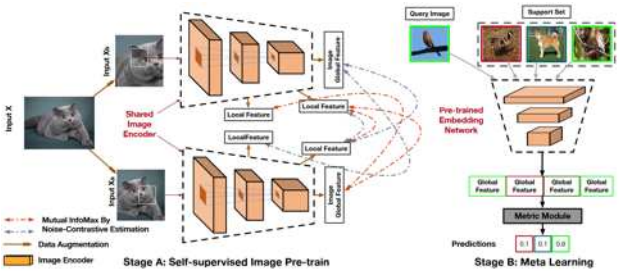

图 1

如图 1 所示。我们方法的总体架构。左 : 列车嵌入网络的自监督学习 (AMDIM)。下游任务的设计是为了最大化通过数据扩增生成相同的图像的两个视图之间的相互信息。右图 : 情景任务的元学习 (3–way, 1–shot 例子)。

对于每个任务，嵌入网络对训练样本和查询样本进行编码。将查询样本嵌入与训练样本嵌入的质心进行比较，并进行进一步预测。

3 方法

少样本学习是在有限的训练数据，需要验证未见类的数据性能。对于少样本学习分类问题的一个长用的解决方案是在预学习嵌入网络的基础上应用元学习。目前大多数方法主要集中在第二阶段，即元学习阶段。在遵循这两个阶段的同时，利用自我监督学习来训练一个大规模嵌入网络会作为我们的基础。

3.1 自监督学习阶段

我们的目标是学习增强特征泛化的表示。在我们的方法中，我们将 AMDIM 作为我们的自监督模型。下游任务旨是实现从共享上下文的多个视图中提取的特征之间的交互信息的最大化。交互信息 (MI) 度量两个随机变量 X 和 Y 之间共享的信息，定义为边值的乘积和关节之间的 KullbackLeibler (KL) 离散度。

$$I(X,Y) = D_{KL}\left(p(x,y)||p(x)p(y)\right) = \sum\sum p(x,y)\log\frac{p(x|y)}{p(x)} \tag{1}$$

P(x，y) 是联合分布，P(x) 和 P(y) 是 X 和 Y 的边际分布。由于我们只有样本，而不能直接获得下推分布，因此估算 MI 很有挑战性。证明了基于负采样的噪声对比估计 (NCE) 损失最小化可以使交互信息的下界最大化。

AMDIM 的核心概念是从同一幅图像的两个视图 (xa,xb) 最大化全局特征与局部特征之间的交互信息。具体来说，就是最大化相互信息调剂〈fg(xa),f5(xb)〉，〈fg(xa),f7(xb)〉和〈f5(xa),f5(xb)〉。其中 fg 是全局特征，f5 是编码器的 5x5 局部特征映射，f7 是编码器的 7x7 特征映射。例如 fg(xa) 和 f5(xb) 之间的 NCE 损失定义如下：

$$\mathcal{L}_{amdim}\left(f_g(x_a), f_5(x_b)\right) = -\log\frac{\exp\{\phi(f_g(x_a), f_5(x_b))\}}{\sum_{\widetilde{x_b}\in\mathcal{N}_x\cup x_b}\exp\{\phi(f_g(x_a), f_5(\widetilde{x_b}))\}} \tag{2}$$

Nx 为图像 x 的负样本，ϕ 为距离度量函数。最后，xa 与 xb 之间的总损失如下：

$$\mathcal{L}_{amdim}(x_a,x_b) = \mathcal{L}_{amdim}\left(f_g(x_a), f_5(x_b)\right) + \mathcal{L}_{amdim}\left(f_g(x_a), f_7(x_b)\right) + \mathcal{L}_{amdim}\left(f_5(x_a), f_5(x_b)\right) \tag{3}$$

在图 1 阶段，给出了 AMDIM 自监督学习方法的概述。红色和蓝色的线表示两个视图 xa 和 xb 之间的局部和全局特征。编码器网络的详细信息定义在表 1 中：

表 1

Layers	Output Size	ConvBlocks (kernel, output_channels, stride, pad)
conv1	62 × 62	[5 × 5, ndf, 2, 2] [3 × 3, ndf, 1, 0]
conv2_x	30 × 30	[4 × 4, 2 × ndf, 2, 0] [1 × 1, 2 × ndf, 1, 0] × (2 × ndepth − 1)
conv3_x	14 × 14	[4 × 4, 4 × ndf, 2, 0] [1 × 1, 4 × ndf, 1, 0] × (2 × ndepth − 1)
conv4_x	7 × 7	[2 × 2, 8 × ndf, 2, 0] [1 × 1, 8 × ndf, 1, 0] × (2 × ndepth − 1)
conv5_x	5 × 5	[3 × 3, 8 × ndf, 1, 0] [1 × 1, 8 × ndf, 1, 0] × (2 × ndepth − 1)
conv6_x	5 × 5	[3 × 3, 8 × ndf, 1, 0] [1 × 1, 8 × ndf, 1, 0] × (2 × ndepth − 1)
conv7	1 × 1	[3 × 3, nrkhs, 1, 0] [1 × 1, nrkhs, 1, 0]
# params.		198M(ndf=192, nrkhs=1536, ndepth=8)
# FLOPs.		10.96 GFLOPs (ndf=192, nrkhs=1536, ndepth=8)

表 1 为 AmdimNet 的模型架构。ndf 是网络的输出通道参数。ndepth 控制模型的深度。nrkhs 为嵌入维度。conv2_x 到 conv6_x 中的每个卷积块包含 2 个 n 深度卷积槽。

3.2 元学习阶段

给出了一个嵌入网络，应用元学习对其进行微调，以适应小样本分类的类变化要求。一个典型的元学习可以被认为是一个多任务的 K-way C-shot 情景分类问题。对于每个分类任务 t，有 K 个类，每个类有 C 个样本。整个训练数据集可以由 D={(x1,y1)，...,(xN,yN)} 其中 N 为 D 中的总类数，对于特定任务 T,V={yili= 1,...,K} 表示其中关联的类标签。这里 K 是单个训练任务的支持集的类数。支持集和查询集通常可以从 D 中随机选择，D: (a) 任务 T 的支持集可以由 S={(xi,yi)|i= 1,...,m} 表示，其中 m = C × K (K-way C-shot);(b) 查询集为 Q={(xi,yi)|j=1，..，n}，n 为待 meta 测试的样本数。

最近流行的框架如 Snell 等人能够学习一个嵌入函数，将所有输入样本映射到描述空间中的一个均值向量 c，以表示每个类。对于类 k，用训练样本嵌入特征的质心表示，可得：

$$c_k = \frac{1}{|S|}\sum_{(x_i,y_i)\in S} f_g(x_i), \tag{4}$$

其中 fg(xi) 为嵌入函数。基于度量学习的方法：

基于度量学习的方法，我们使用了距离函数 d，并在所有类上生成了一个分布，并给出了来自查询集 Q 的查询样本 q:

$$p(y=k|q) = \frac{\exp(-d(f_g(q), c_k))}{\sum_{k'}\exp(-d(f_g(q), c_{k'}))} \tag{5}$$

我们选用欧氏距离作为我们的距离函数 d, 如等式 5 所示，分布是基于样本嵌入 (在查询集中) 和类的重构特

征之间的 softmax。多元学习阶段的损失可以看到：

$$\mathcal{L}_{meta} = d(f_g(q), c_k) + \log \sum_{k'} d(f_g(q), c_{k'}) \quad (6)$$

4 实验结果

我们先介绍了我们评估中使用的数据集和训练过程，然后展示了不同数据集在 ProtNet 网络以及不同的监督模型下的测试结果，最后我们对算法进行了评估。

4.1 数据集

17flower 数据集，全部由花类组成，有 17 类不同的花，每个类有 80 个图片；17fish 数据集是 17 类的不同鱼的图片，每类有 80 张图片。

Cifa 数据集共有 60000 张彩色图像，这些图像是 32*32，分为 10 个类，每类 6000 张图。这里面有 50000 张用于训练，构成了 5 个训练批，每一批 10000 张图；另外 10000 用于测试，单独构成一批。测试批的数据里，取自 10 类中的每一类，每一类随机取 1000 张。抽剩下的就随机排列组成了训练批。注意一个训练批中的各类图像并不一定数量相同，总的来看训练批，每一类都有 5000 张图，我们是分别将训练集和测试集作为了我们的数据进行训练。

Caltech 是加州理工学院的图像数据库，包含 Caltech101 和 Caltech256 两个数据集。该数据集是由 Fei-FeiLi, Marco Andreetto, Marc 'Aurelio Ranzato 在 2003 年 9 月收集而成的。Caltech101 包含 101 种类别的物体，每种类别大约 40 到 800 个图像，大部分的类别有大约 50 个图像。Caltech256 包含 256 种类别的物体，大约 30607 张图像；其中我们只取了 40 个类作为我们的训练集；5 个类作为测试集；5 个类作为验证集，其中每个类包含 40 张图片，有少部分类是灰度图。

4.2 训练细节

最近的一些研究表明，一个典型的训练过程可以包括一个预先训练的网络或采用联合训练来对特征的嵌入。这可显著提高分类的准确率。我们采用 AMDIM [11] 自监督学习训练框架对 ProtNet 网络行了预训练。AmdimNet(ndf-256, ndepth=10, nrkhs=2048) 被用于所有数据集。选择 Adam 作为优化器，学习率为 0.0002。我们使用 128 x 128 作为这些数据集之间的输入分辨率。

表 2

	Embedding Net	1-shot 5-way	5-shot 5-way
17flower	4Conv	40.35%	62.26%
	ResNet12	40.63%	60.02%
	AmdimNet	42.79%	65.25%
cifa_train	AmdimNet	43.98%	66.27%
cifa_test	AmdimNet	42.85%	67.32%
caletch	AmdimNet	39.75%	68.93%
17fish	AmdimNet	43.55%	62.35%

表 2 为多种数据集在在三种模式下（主要是 AmdinNet 自监督网络）的 1-shot 5-way、5-shot 5-way 任务上的少样本分类精度结果。

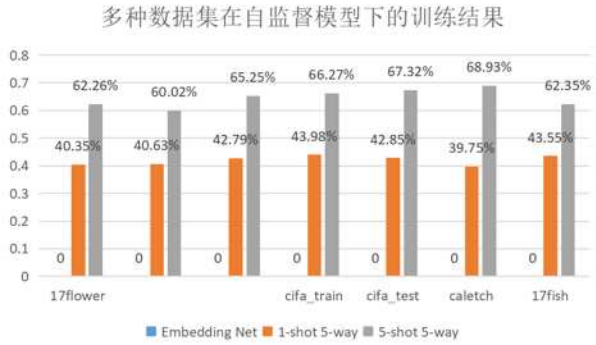

图 2

图 2 是用具体的柱状图来展示训练结果，前三组柱状图代表 17flower 数据集在三种模型下的 1-shot 5-way 以及 5-shot 5-way 的训练结果，其余组默认为对 1-shot/5-shot 5-way 在 AmdimNet 模型下的训练结果。

4.3 定量比较

对于多种数据集，我们评估了我们的方法在两种常见的少样本学习任务，即 1-shot、5-way task 和 5-shot、5-way task 对 ProtNet 网络。如表 2 所示，我们将不同的数据集在 ProtNet 网络和 Embedding Net 上的 1-shot 5-way 和 5-shot 和 5-way 的实验结果；明显可以看出 5-shot 5-way 的准确率明显提高；扩大类对准确率也有提高；数据集的大小对结果的影响也不是很大；类的不同对结果影响也不大；训练次数对结果的影响很大。

5 结束语

在本文中，我们提出利用自监督学习来有效地训练一个鲁棒的嵌入网络来进行少样本图像分类。通过元学习过程进行微调后，基于两种常用的少样本分类数据集的定量结果，该方法明显优于其他方法。利用本文的关于自监督学习的少样本图像分类的方法，在处理实际的图像分类时，特别是样本数量以及样本标签较少的图片特征来说，是一种很好的分类方法；在评估中我们可以发现对于不同的类和相同的类区分度不大，这在一定程度上做到了普遍分类，在准确率方面，最终结果的原因在于训练次数和模型参数的大小，本次训练因为设备原因都将参数和次数都调小了，我们认为将次数调到更多，参数调到更大，训练结果将有很大的提升。

【参考文献】

[1] OriolVinyals,CharlesBlundell,TimothyLillicrap,DaanWierstra, et al., “Matching networks for one shot learning,” in NeurIPS,

2016, pp. 3630 - 3638.

[2] Sachin Ravi and Hugo Larochelle, "Optimization as a model for few-shot learning," in ICLR, 2017.

[3] Jake Snell, Kevin Swersky, and Richard Zemel, "Prototypical networks for few-shot learning," in NeurIPS, 2017, pp. 4077 - 4087.

[4] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr,andTimothyMHospedales, "Learningtocompare:Relationnet workforfew-shotlearning," inCVPR,2018,pp.1199 - 1208.

[5] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell, "Meta-learning with latent embedding optimization," arXiv preprint arXiv:1807.05960, 2018.

[6] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille, "Fewshot image recognition by predicting parameters from activations," in CVPR, 2018, pp. 7229 - 7238.

[7] XiangJiang,MohammadHavaei,FarshidVarno,GabrielChartrand,NicolasChapados,andStanMatwin, "Learningtolearn with conditional class dependencies," in ICLR, 2019.

[8] BorisOreshkin,PauRodr´ıguezL´opez,andAlexandreLacoste, "Tadam: Task dependent adaptive metric for improved fewshot learning," in NeurIPS, 2018, pp. 719 - 729.

[9] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Modelagnostic meta-learning for fast adaptation of deep networks," in ICML. JMLR, 2017, pp. 1126 - 1135.

[10] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov, "Siamese neural networks for one-shot image recognition," in ICML Deep Learning Workshop, 2015, vol. 2.

[11] Philip Bachman, R Devon Hjelm, and William Buchwalter, "Learning representations by maximizing mutual information across views," arXiv preprint arXiv:1906.00910, 2019.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," inCVPR. IEEE, 2016, pp. 770 - 778.

[13] Bernaschi M , Castiglione F , Ferranti A. ProtNet: A tool for stochastic simulations of protein interaction networks dynamics[J]. BMC Bioinformatics, 2007, 8 Suppl 1(Suppl 1):S4.

[14] CodingFish，自己搭建一个神经网络进行识别分类，https://zhuanlan.zhihu.com/p/98981710,(2020.7.1).

[15] 杨卫红 . 数据库编程与图像处理 [J]. 电脑编程技巧与维护 , 2013(16):44-46.