Few-Shot Learning

翁家翌 2016011446

2018.6

Contents

1 Basic Idea

在课堂上，崔老师提到对于 Few-Shot Learning[5] 的一些看法，他指出人脑的学习速度会随着知识量的增加而加快，而现有的网络结构（如 AlexNet[6]、VGG[9]、ResNet[4] 等等）都是由数据驱动学习出来的，与人类学习方式不符合，因此提出了这个大作业。

顺着这个思路，我们组继续想下去：

首先，学习的过程对于一个人而言是十分重要的，他会从这个过程中学习到一些学习的方法，而这个学习学习的过程（Meta Learning[2]）也被广泛运用于许多神经网络的任务中，来更好地增加泛化能力。

其次，对于该任务而言，我们有的是一个在 ImageNet[8] 上 pretrain 好的 AlexNet[6] 模型，它对于图像的特征提取已经有了很强的能力，只不过没有对新类别的分类参数。考虑到图像的特征提取对于每个神经网络模型都是通用的，因此我们只决定修改最后一层的分类器。

如果只是简单的修改最后一层，将 $[4096, 1000]$ 的分类器改成 $[4096, 50]$，那和 fine-tune 没区别，并且也不符合人类学习的方式。因此我们考虑，这些参数是否能够被学习出来，这样一来也能够具有更好的泛化能力。

并且，一个类别中的图片总是对全连接层之前的某些激活节点产生很强烈的响应，举个例子，如果图像是一辆车，那么表示门、窗户、轮子的这些激活节点会产生强烈响应；如果在产生对车的预测概率时候，把门、窗户、轮子的激活节点对应的权重增加，那么是会有益于对车的分类准确性的。因此可以猜想，一张图经过神经网络得到的特征 a_y，对应某个类别 c 的权重 $w_{c,y}$，如果 $a_y \cdot w_{c,y}$ 足够大，那么分类效果越好。更抽象一点，权重 w 与激活节点 a 是存在某种对应关系的。

因此我们实现了这个方法，实验结果表明它比 fine-tune 的 CNN 效果会好一些。为什么好不太多呢？我觉得是数据集之间有 overlap，Caltech256[3] 其实和 ImageNet 的 Data Distribution 挺接近的，因此才会导致 baseline 太高，而传统的 Few-Shot Learning 任务定义是新数据集与原数据集完全没有 overlap。我个人觉得在 Caltech256 上面做这个实验不太具有说服力。

2 Algorithm

算法的核心思想是学习一个映射函数 $\phi : a_y \to w_{c,y}$，根据每个类别中激活函数的统计值来学习这个映射关系。

整个 forward 的流程如下：

1. 使用 pretrained AlexNet 对训练集图像提取特征；
2. 对于每个类别，将其所属的图像的 activation map 取平均，作为这个类别的 activation map $\overline{a}_y$；
3. 使用训练好的映射函数 ϕ（大小为 $[4096, 4096]$）得出当前最后一层的权重 $w_{c,y}$；
4. 使用该权重（大小为 $[4096, 50]$）对图像进行最后的预测；

理论上 ϕ 能够在 ImageNet 数据集上直接学出来，然而助教说不让用，并且我们使用的是 PyTorch 而不是 TensorFlow，无法直接使用 feature.npy 文件，故而只在给出的图片中训出 ϕ 的参数。我相信如果加入 ImageNet 的部分数据，这种方法能够有更出色的表现。

3 Result

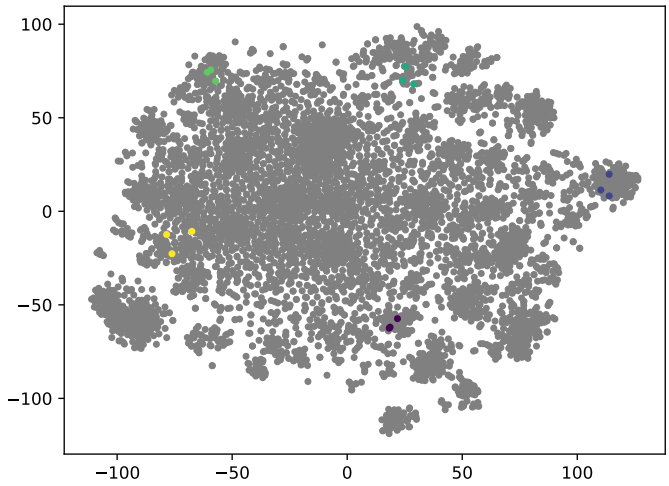

Figure 1: 使用 t-SNE[7] 对测试数据的特征进行可视化，图中灰色的点为测试集，非灰色的点为训练集中带 label 的图片，相同颜色代表相同类别。

我们采用了 t-SNE[7] 对测试集数据提特征之后进行可视化，如图1所示，可以直观地看出，相同的类别在特征空间中的距离十分近。因此猜想最后一层的线性分类器理论上应该有能力将这些特征分类。为了验证这个猜想，我们拿这些 500 张带 label 的数据来训练 CNN 的最后一层分类器，发现能够很快的 overfit，因此得出结论：在数据规模不大的时候，单层线性分类器理论上是有能力达到 100% 的分类准确率的。这个结论是我们实现方法的一个有力支撑。

表1是实验结果的对比数据表格。从训练图片数量对结果的影响趋势来看，训练图片越多，准确率越高。因此我们最终提交的模型是用 10 张图训练，不加任何测试集，训练 100 个 epoch，每 30 次之后 learning rate 降低至原来的 0.1 倍，初始的 learning rate 设置成 10^{-4}。

Method	One-shot	Three-shot	Five-shot
Nearest Nerghbour	44.0%	56.4%	59.6%
Fine-tune CNN	37.2%	54.8%	64.4%
Our Result	42.4%	59.6%	67.6%

Table 1: 实验结果

我们随机在测试集中选了一部分图片进行标注，结果显示我们的最终准确率大约在 70% 左右。

4 Contribution

在本次大作业中，我担任组长，负责统筹规划任务分配，check 进度，以及模型的训练与测试，和 baseline 的代码编写。我们比较了 fine-tune baseline 和 nearest neighbour in cosine distance 两种方法，这两部分的代码以及整体的代码架构都是我写的。以及在代码文件夹下面还有一些其他文件，比如：

merge.py 模型 ensamble 所用；

xgb.py XGBoost[1] 对于该任务的表现，只有 57% 左右的准确率，因此不是很适合该任务；

tsne.py 使用 AlexNet 提取测试集图片的 feature 之后将其可视化到二维平面上；

以上代码均由我完成。

高天宇和杨帆主要负责核心算法的实现、数据处理、参数调整、PPT 制作和最后的展示。还有一位同学胡钧，至今没有联系上，因此实际上我们组只有三个人。

References

[1] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785--794. ACM, 2016.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *arXiv preprint arXiv:1703.03400*, 2017.

[3] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. 2007.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770--778, 2016.

[5] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097--1105, 2012.

[7] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579--2605, 2008.

[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211--252, 2015.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.