分 类 号： TP391.4 UDC： 510

盲审代码： 密级： 公开

**硕 士 学 位 论 文**

**基于深度学习的标签传播的图像分类研究**

|  |  |  |
| --- | --- | --- |
| 学科、专业： | **应用数学** |  |
| 研究方向： | **智能系统与控制** |  |
| 完成日期： | **2019年3月** |  |

**基于深度学习的标签传播的图像分类研究**

摘 要

在计算机视觉领域,图像分类任务一直都是热点问题.本文关注的半监督学习任务仅仅需要利用少量有标签数据和大量无标签数据便可训练出一些实用的模型, 因此获得了国内外学术界的广泛关注.

在当下,深度学习对于图像处理任务有着得天独厚的优势 ,本文先是通过卷积神经网络提取有标签数据和无标签数据的高级语义特征向量; 然后基于高级语义特征利用标签传播算法获得无标签数据的伪标签;将带有伪标签的无标签数据集和有标签数据集都放入卷积神经网络模型中进行深度模型的微调。至此，神经网络提取特征与半监督学习的标签预测通过神经网络的微调机制建立起了交互的功能。 交互的使用深度模型与标签传播模型不断的优化伪标签直到模型达到收敛.本文在 Cifar10 数据集上做了一些对比实验最终的实验结果表明:本文提出的算法 A, B, C 总是比单独使用半监督学习算法或者深度学习算法要好的多.其中,针对800 有标签数据,49200无标签数据集在算法 C达到了 0.91的准确率.

关键词:卷积神经网络,基于图的半监督分类,降维,模型微调

RESEARCH ON IMAGE CLASSIFICATION OF LABEL PROPAGATION BASED ON DEEP LEARNING

ABSTRACT

In the domain of computer vision, image Classification task has always been a hot issue. The semi-supervised learning task concerned in this thesis only needs to use a small amount of labeled data and a large amount of unlabeled data to train some practical models.

At present, deep learning is a unique to image processing tasks. Taking into account the advantages of semi-supervised learning and deep learning, in this thesis convolution neural network first extracts the advanced semantic eigenvectors with label data and unlabeled data. Then, based on the advanced semantic features, the label propagation algorithm is used to obtain the pseudo-label of unlabeled data. The non-label dataset with pseudo-label and the labeled dataset are put into the convolution neural network model to fine-tune the depth model. The interactive use depth model and the label propagation model constantly optimize the pseudo-label until the model reaches convergence. In this thesis, some comparative experiments are carried out on the Cifar10 dataset. The final results show that the algorithm A, B and C proposed in this thesis are always much better than semi-supervised learning algorithm or deep learning algorithm alone. Among them, for 800 tagged data and 49200 unlabeled datasets classification accuracy of algorithm C reached over 0.91.

KEYWORDS:convolution neural network, semi-supervised classification based on graph, descending dimension, fine-tuning of models

目录

[摘 要 I](#_Toc4964829)

[关键词 I](#_Toc4964830)

[ABSTRACT II](#_Toc4964831)

[第 1 章 绪论 1](#_Toc4964832)

[1.1 研究背景与意义 1](#_Toc4964833)

[1.2 国内外研究进展 3](#_Toc4964834)

[1.2.1 卷积神经网络简介 3](#_Toc4964835)

[1.2.2 k 近邻 4](#_Toc4964836)

[1.2.3 两种降维策略 5](#_Toc4964837)

[1.2.4 两种标签传播算法 7](#_Toc4964838)

[1.3 本文的主要工作及章节安排 9](#_Toc4964839)

[第 2 章 基于深度卷积提取特征的直推式标签传播算法 11](#_Toc4964840)

[2.1 本章算法简介 11](#_Toc4964841)

[2.2 数据集简介 13](#_Toc4964842)

[2.3 实验结果与分析 13](#_Toc4964843)

[2.3.1 固定每类有标签样本数目为50观察邻域范围对实验的影响 14](#_Toc4964844)

[2.3.2 固定邻域半径为18观察有标签样本的数目对实验的影响 15](#_Toc4964845)

[2.3.3 特征的维度对实验的影响 16](#_Toc4964846)

[2.4 本章小节 18](#_Toc4964847)

[第 3 章 利用深度特征进行标签传播并微调深度卷积网络 19](#_Toc4964848)

[3.1 算法引入 19](#_Toc4964849)

[3.2 基于 LGC 的深度卷积神经网络微调算法 19](#_Toc4964850)

[3.2.1 算法简介 19](#_Toc4964851)

[3.2.2 实验结果与分析 20](#_Toc4964852)

[3.3 本章小节 21](#_Toc4964853)

[第 4 章 总结与展望 22](#_Toc4964854)

[4.1 研究总结 22](#_Toc4964855)

[4.2 研究展望 22](#_Toc4964856)

[参考文献 24](#_Toc4964857)

# 第 1 章 绪论

## 1.1 研究背景与意义

本文仅仅考虑计算机视觉的核心任务之一的图像分类这一语义检索任务. 诸如对象识别、图像标注、行为识别等均可转换为图像分类问题[1]. 对于人类来说, 识别图 1-1是一只狗是一个十分简单的问题. 但是, 即使是顶级的程序员也无法直接编写出让计算机直接判定图 1-1是狗还是熊的程序.因为计算机仅仅可以识别像素级别的数字特征(对于 RGB 彩色图像, 它通常表示为三维数组, 其数据类型为 uint8.), 而无法直接解码图片的具体语义.

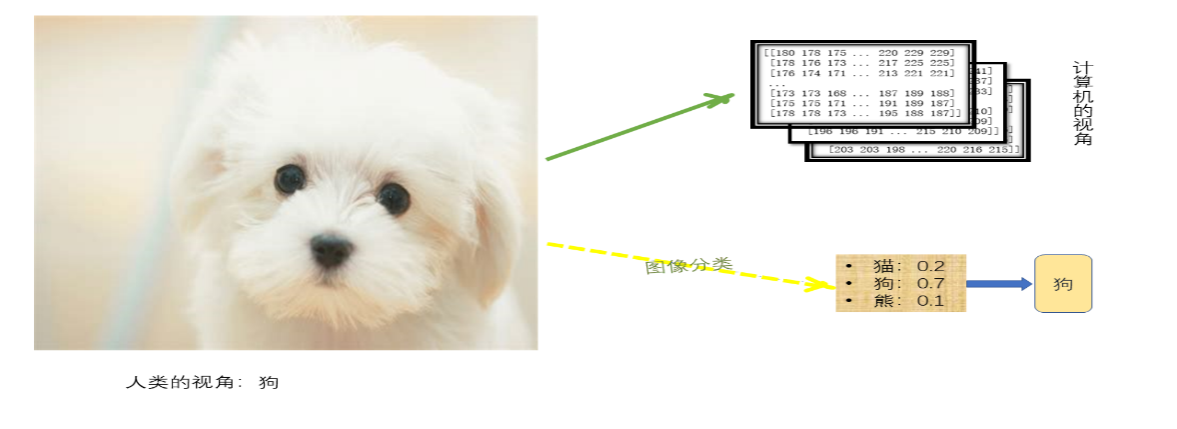


图1-1图像识别问题的一个示例

Figure 1-1 An example of an image recognition problem

为此, 人们逆向思考: 借助人类自己的判别能力优选出大量图片并对其手工标注, 得到一个带有标签的数据集 (一般称其为训练集), 然后, 设法去拟合图片的像素级别的特征和其标签的关系. 由此产生了大量的用**数据去编程**的优秀算法. 比如,对于图 1-1涉及的图片标注问题,可以这样来思考如何设计图像的分类模型: 思考如何从这一堆数字中识别出其属于猫, 狗, 熊的概率分布, 然后选择概率最高的作为模型的输出,即图片的标注信息.

一般地, 图像分类模型可以看作利用数据进行编程的模型, 即根据带有标签的训练集训练一个拟合模型, 然后对给定的新的预测样本通过已经训练好的该模型预测它属于哪一个分类标签 (或者给出属于一系列不同标签的可能性).识别图 1-1是否是一只 "狗" 是对人类来说是一个很简单的任务, 然而对于计算机来说就显得尤为困难, 计算机一般会将其解码为类似形式的数组, 每一个元素对应于其像素值. 受拍摄视角, 图片尺寸, 光照条件等各种因素干扰, 图片的像素值可能会发生改变, 哪怕是一点轻微的色调扰动都会增加计算机对图片的识别难度.

人工特征不仅仅需要人们耗费大量的精力和时间去研究,同时对人们的专业知识有着很严格的要求.尽管诸如支持向量机分类器 (SVM[2]) 和视觉词袋模型(bag of visual word[3]) 已经在许多数据集上取得了不错的效果, "语义鸿沟" (semantic gap[4]) 问题 (计算机表示的底层图像特征和人类所感知的高层语义信息之间的跨度问题) 却没有得到很好的解决. 幸运的是深度学习的出现, 为语义鸿沟问题的解决提供了一个比较好的解决策略.

深度学习的端到端的学习方式为提取图像语义特征带来了福音.深度学习的最大优点是可以自主地学习图像多级特征[5].对于图像的语义信息的提取, 卷积神经网络(CNN, convolutional neural network)[6]在深度学习邻域中扮演着很重要的角色[[1]](#footnote-1).深度卷积神经网络已经广泛应用到计算机视觉[7]、自然语言处理 [8]等多个领域.在工业界和学术界不仅掀起了卷积神经网络研究的浪潮, 同时也促进了人工智能的发展. 卷积神经网络可以直接从大量训练数据中自动学习特征的表示. 卷积神经网络还具有如下特性：局部连接[9]、权值共享[10]和池化 [11]操作通过这些特性可以有效降低模型的计算复杂度, 减少训练参数的数目, 使模型对平移、扭曲、缩放具有一定程度的不变性.目前, 深度卷积神经网络主要是通过增加网络的层数,使用更大规模的训练数据集以及改进现有神经网络的网络结构或训练学习算法等方法来模拟人脑复杂的层次化认知规律,拉近与人脑视觉系统的差距,使机器获得“抽象概念”的能力[12].

一般地, 人们先在 ImageNet[13]数据集上训练了许多的“基网络”并保存其训练后的参数, 接着依据深度卷积神经网络的逐层地、分级的提取特征的特性, 我们可以结合迁移学习[14]的思想, 利用“基网络”提取需要训练的数据集的高级特征 (即一个具体图像的纹理, 甚至语义信息的特征向量或者特征响应图), 然后再利用一些比较成熟的诸如标签传播算法进一步的研究图像的高级语义特征.

对于样本数目较少的数据集,借助卷积神经网络的微调技术可以取得不错的结果,如果同时引入无标签的数据的信息则可以进一步提升模型的泛化能力.即利用在 ImageNet 上训练好的卷积神经网络模型来提取目标数据集的深度特征,同时结合标签传播技术(本文仅仅考虑LGC(Local and Global Consistency)[15]或者GRF(Gaussian Random Field)[16]) 来优化模型.

考虑到深度特征也许会有冗余,本文采用的模型加入了降维处理的技术,同时本文也将kNN(k-nearest neighbors algorithm)[17]加入了讨论范围作为算法比较基准, 以此来对照实验的效果. 综合上述观点考虑,本文提出的模型对于半监督学习在图片分类上的研究有着很好的补充,是一个十分有意义的研究课题.

## 1.2 国内外研究进展

为了更好的说明本文的研究内容,对符号做如下约定:

假设原始数据由有标签数据(一般被视为训练集) 和无标签数据(一般被视为测试集或验证集) 组成. 其中

* 是图片的像素级特征(一般分为灰度图和彩色图片);
* (为图片所属类别的个数);

对于任意的像素级别的特征, 通过某一个特征提取器(可以是手工特征也可以是精心设计的特征) 提取其语义特征向量为, 且记

(1-1)

下面便可以通过来计算原始数据集的相似度矩阵[18], 记, 则可以计算对角矩阵以及拉普拉斯矩阵.

为了利用向量化编程来提高算法的运算效率, 还需要将向量化:

(1-2)

这样, 便可定义

(1-3)

定义满足

. (1-4)

其中, 表示属于各个类别的概率分布, 且

(1-5)

由于后面将要介绍的算法大都与"距离"相关, 为了防止数据的不同维度的尺度不一而造成一些算法不收敛等问题, 本文默认做了标准化处理.

### 1.2.1 卷积神经网络简介

卷积神经网络经历LeNet[19],AlexNet[20], VGG[21], GoogleNet[22] (Inception[23]), ResNet[24], Densnet[25]等变革, 卷积神经网络变得越来越深, 同时训练难度也不断变大. 考虑到计算设备, 数据集的规模等的制约, 迁移学习在计算机视觉中扮演的角色也越来越重要. 同时半监督学习对于图片分类任务也是至关重要的, 但是仅仅使用浅层机器学习进行标签传播远远无法满足当今社会的需求.

从技术角度来看, CNN 是一种至少包含一个卷积层 (有滤波器的作用) 的神经网络. 一般地, 一个神经元可以看作是线性映射与非线性映射的组合, 而多层神经网络是多个神经元按照分层结构组合而成的. 神经网络模型的优化方法大都是基于反向传播算法求得的数值解. 本文使用了何恺明等[24]人提出的残差网络(ResNet). ResNet网络的基本单元是残差块, 见图2.,

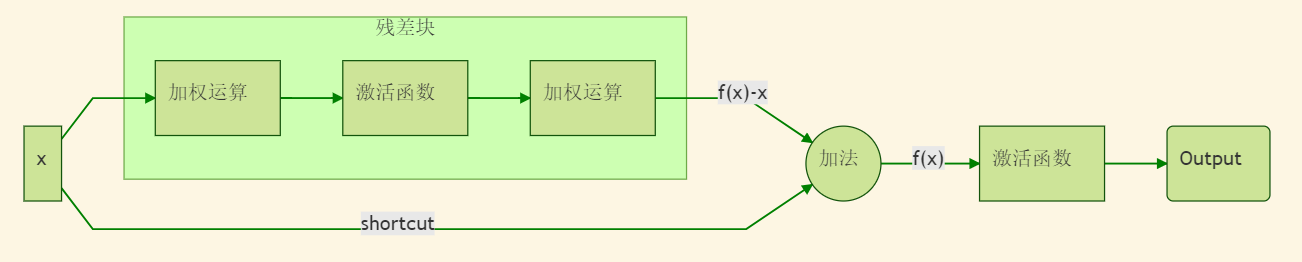


图1-2. 残差模块示意图

Figure 1-2. Diagram of residuals module

残差块通过跨层连接 (shortcut) 可以使得神经网络变得更深, 且一定程度上缓解了梯度消失和梯度爆炸所造成的一系列问题.

在ResNet中指出神经网络的训练存在退化问题, 即随着层数加深到了一定的深度之后, 越深的网络训练的效果越差, 但这并不是梯度消失和梯度爆炸引起的问题 (因为已经有许多方法来解决梯度传播的问题). 数据经过了两条路线, 一条与普通神经网络类似, 另一条实现了单位映射的路线 (被称为 shortcut). 这两条路线一般被称为"残差块". 在网络中使用残差块的神经网络被称为ResNet. ResNet很好的应对了网络退化问题, 可以使得网络变得很深. 由于 ResNet 的这么多优秀性质, 本文采用 ResNet50 这一结构来拟合半监督学习模型.

回到图像分类任务上, 神经网络模型具有训练时间长、测试时间短的优点, 只要它在一个好而大的数据集上训练好了参数, 该模型便具有提取比较好的特征的功能. 对此, 大多数人都选择在 ImageNet 数据集上进行模型训练[[2]](#footnote-2), 然后将训练好的模型保存下来, 通过其强大的特征提取能力, 迁移学习人们将会在新的数据集上快速的训练出满意的模型(这里的迁移学习特指深度模型的微调).虽然, CNN 提取的特征很好的代表了图片的高级语义特征,但是仅仅在训练集上进行模型的学习,极大的浪费了无标签的测试集.为了更好的利用无标签的数据的结构消息,人们设计了许多的基于度量(即”距离”)的算法来进一步提升模型的泛化能力,下面本文先介绍一种相对较简单度量学习策略: kNN算法.

### 1.2.2 k 近邻

k 近邻(k-nearest neighbors algorithm) 是一种非参数的算法, 它可以用于分类和回归任务. k 近邻的核心思想可以通过“物以类聚”来形象的说明, 利用训练数据集 来推断测试集 的标签. 下面以 中的样本 来说明 k 近邻算法的具体流程是:

* 1.1计算 中的样本到 中的样本的距离(可以采用欧式距离, 余弦距离, 汉明距离等);
* 1-2 将 中的点按与 的距离由近到远进行排列, 然后仅仅保留最近的前 个点 ;
* 1-3. 计算 对应的标签的众数 , 这样便将 标注为 .

对于测试集中的每个点都按照 1-1~1-3 进行操作, 便可完成对测试集 的打标任务.

虽然有许多的改进版的 kNN 算法, 但是, 对于图像分类任务, kNN 有点力不从心: 图片的维度一般都很高, 普通的提取特征的策略往往无法获得预期结果, 因为, 传统的特征提取方法并不能很好的去除图片冗余信息. 为了去除图片的冗余信息, 人们设法通过降维技术去学习图片的一个低维的嵌入.

### 1.2.3 两种降维策略

降维是指提取高维空间的关键信息, 并将高维空间的问题转换为易于计算的低维空间的问题进行求解的过程. 降维是十分有必要的, 因为它缓解了以下三种问题:

* 多重共线性: 预测变量之间可能会存在某种关联性, 这样会导致模型的解空间十分的不稳定性;
* 高维空间的稀疏性: 高维空间的数据往往是稀疏的, 增加了模型求解的难度;
* 过高的维度可能含有大量的冗余信息, 妨碍人们找到变量之间的真正的隐含信息.

本文主要涉及两种降维策略, 无监督的 PCA[26]和有监督的 LDA[27].下面逐一介绍.

#### 1.2.3.1 PCA

下面我们来探讨**主成分分析法**(Principal component analysis, PCA): 对于任意的, PCA 方法为了学习某一正交投影而将输入数据正交投影表示为, 同时还需要降维后的数据尽可能的保留原数据空间的信息. 令, 约束为. 其中为特征向量的维度. 这里被称为解码矩阵, 且被称为**投影矩阵**. PCA 算法的核心思想是: **最近重构性**, 即

(1-6)

和**最大可分性**, 即

. (1-7)

很容易证明式(1-6) 与(1-7)是等价的: 式(1-6) 的求解需要两步:固定,先求出,这样(1-5) 便转换为(1-7).而(1-7) 亦可以转换为.这样原问题可以转换为

(1-8)



其中, 是拉格朗日乘子. 这样, 将进行标准化处理, 那么可以看作是样本的协方差矩阵; 再对进行特征值分解, 并求得最大的个特征值所对应的特征向量 (在 PCA 算法中, 一般称其作主成分). , 则便可得到. 因而, 对于任意的,便可得到其低维表示



. (1-9)

PCA仅仅利用了数据的内在度量信息,而没有考虑其标签信息,为此, LDA 便派上用场了.

#### 1.2.3.2 LDA

线性判别分析 (Linear Discriminant Analysis, LDA) 的核心思想是设法将样本投影到一个子空间上, 且使得同类的样本投影后尽可能的接近,同时不同类的样本投影后尽可能的分开. LDA 与 PCA 十分相似, PCA 试图找到方差最大的几个主成分, 而 LDA 的目标是发现可以最优化分类的特征子空间. 与 PCA 类似, 需要将进行标准化处理, 接着将有标签的训练集数据按其类别划分为个子集类, 并计算各个子类所在集合的均值向量, 则由于不同类之间的样本尽可能分离等价于

(1-10)

而同类之间尽可能接近等价于

(1-11)

其中,类间散度矩阵以及类内散度矩阵分别为:

(1-12)

故而原问题便转换为以下优化问题:

(1-13)

因而,取的前大的特征值所对应的特征向量组成, 最后使用将样本投影到新的特征子空间中.虽然 LDA利用了数据的标签信息,并且去除了信息冗余,但是 LDA也忽视了无标签数据的信息. 标签传播算法便是同时考虑标签数据和无标签数据进一步优化模型.标签传播算法有许多不同的版本,下面仅仅重点介绍其中两个最常用的标签传播算法.

### 1.2.4 两种标签传播算法

由于获得大量无标签的数据很容易, 而获得有标签的数据需要花费大量的人力, 物力和时间. 考虑到打标的代价太大, 一些学者将目光转向了半监督学习: 尽可能利用少量有标签的数据和大量无标签的数据来优化模型, 使得无标签数据通过半监督学习推断出无标签数据的标签, 完成对无标签数据的打标的任务. 半监督学习的算法有很多, 本文仅仅考虑标签传播算法.

对于任意的, 构建一个图. 其中, 边集为, 可表示为一个相似度矩阵(affinity matrix, 矩阵中每个元素表示两两之间的相似性分数) .假定从图中学得一个映射, 其对应的分类规则是. 其中被称为标记矩阵.为了实现:相似的样本有相似的标记这一目标, 只需要最小化[28]



(1-14)

为了最小化本文考虑 LLGC和 GRFH这两种算法.

#### 1.2.4.1 LGC: 通过局部和全局一致性进行学习

LGC( Local and Global Consistency) 的相似度矩阵采用的构图方式为全连接图[29]:

(1-15)

有了便可以计算传播矩阵

(1-16)

其中

(1-17)

通过下面的迭代方程便可以完成标签传播. 其中, 便有迭代方程:

(1-18)

其中用来平衡原有的标记信息和从邻近样本传来的信息.考虑到模型的稳定性,正则化形式的LGC 标签传播公式

(1-19)

其中为正则化参数.

最终得到模型的解析解为:

(1-20)

既可以通过迭代方程(1-18) 通过迭代的方式求得,也可直接通过解析式(1-20) 来获得.获得后,通过公式(1-4) 便可得到无标签数据的标签.还有

#### 1.2.4.2 GRF: 高斯随机场和调和函数

GRF(Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions)是采用kNN的方式来构建相似度矩阵:

(1-21)

其中, (1-21) 中的表示的 k近邻.

采用矩阵分块的思想:

(1-22)

且将分块为:

(1-23)

且将分别分块为

(1-24)

(1-25)

故而有

(1-26)

其中表示矩阵或者向量的内积.

GRF实现方式, 直接考虑, 故而 (1-26) 可以转换为

(1-27)

因此，由, 可知

(1-28)

这样便可以使用来标记.

令, 称为**转移概率矩阵**. 类比的分块方式, 对做一样的处理, 则有

(1-29)

这样,可以通过迭代的形式来求取

可以通过迭代方程(1-29) 通过迭代的方式求得,也可以直接通过解析式 (1-28) 直接获取. 无标签数据的标签直接使用即可.

#### 1.2.4.3 其他基于图的标签传播算法

基于图的标签传播算法大都是以LGC与 GRF为基础进行改进的,比如LLNP(Linear Neighborhood Propagation and Its Applications)[30]的构图主要是利用了非负线性最小二乘去重构样本, 具体公式:

(1-30)

而RMGT(Robust Multi-Class Transductive Learning with Graphs)[31],首先采用的是普通的kNN图:

(1-31)

然后得到最后的图为:

(1-32)

需要注意的是:如果得到的图是一个非对称的矩阵，我们只需要将即可得到对称的矩阵[[3]](#footnote-3).

可以看出他们的改进思路大都集中在定义和损失函数上面花费时间和精力.还有学者结合kNN与 LGC提出一种应用于高光谱图像的算法[32], 白本督和范九伦结合稀疏算法以及 K-聚类算法与 LGC极大的降低了计算的时间耗费[33].

## 1.3 本文的主要工作及章节安排

在本硕士论文中,主要提出了基于深度卷积神经网络提取深度语义特征的标签传播算法,主要贡献是:

1. 探讨了使用 CNN与否对 1NN和 LGC以及 GRF的影响;(算法 A)
2. 探讨了先是使用 CNN提取高级语义特征向量再利用 PCA或者 LDA降维技术后对1NN和 LGC以及 GRF的影响; (算法 B)
3. 对 1和 2的模型进行在 CNN提取特征这一项上进行微调来观察 LGC算法的表现.

本论文的实验采用了Matlab(第2章) 与 Python(第3章) 两种编程语言,下面简要的介绍本论文的主要内容及章节安排:

1. 第1章的绪论部分介绍了传统的kNN(k最近邻算法) 与两种标签传播算法: LGC(局部与全局一致性), GRF(高斯随机场和调和函数), 同时还有两种降维策略 PCA (主成分分析算法) 与 LDA(线性判别分析算法).
2. 第2章主要介绍了如何依赖深度卷积神经网络提取图片的高级语义特征来做最近邻分类与基于图的标签传播分类.同时也介绍了如何引入 PCA与 LDA这两种降维的技术进一步优化模型的泛化能力.
3. 第3章主要介绍了通过深度卷积神经网络与基于图的标签传播之间的交互来获得更加优秀的模型.
4. 第4章主要总结了本文的主要工作并做一些展望.

# 第 2 章 基于深度卷积提取特征的直推式标签传播算法

## 2.1 本章算法简介

虽然通过 PCA, LDA 降维技术可以去除了图片低级特征的冗余信息, 并且保留了其大部分关键信息, 减少了模型的计算量. 但是, 第 1.2.3 节介绍的特征向量一般是图片的像素级别的特征拉直为一个向量而得到的, 这样做破坏了图片的部分邻域相关性和平移不变性. 基于此种情况, CNN 的强大的分级提取特征的功能恰恰为该难题的解决提供了强大的支持. 因而, 本文基于 CNN 提取特征, 即可得到特征向量, 然后再进行标签传播的算法操作. CNN 与标签传播算法相结合的两步训练模型的策略极大的提升了单独使用 CNN 或者标签传播算法进行训练的模型的泛化能力. 本文便将kNN也纳入考察范围, 以此对比传统的分类算法与本文的算法的优劣. 本文到考虑目标数据集的有标签样本量很小,如果直接在目标数据集上进行CNN特征的提取工作,得到的特征也并不会一定比手工特征要好,为此,本文考虑在 ImageNet 上先训练一个 CNN模型并将其视作**基模型,**然后再将基模型迁移到我们的目标数据集上进行特征提取.

ImageNet 几乎是深度学习领域的王牌数据集, 它可以用于图像分类, 目标检测, 对象识别, 检测, 定位等多个领域. ImageNet 有多万幅有标签的图片, 涵盖了万多个类别 (其中有超过百万的图片有明确的类别标注和图像中物体位置的标注)[[4]](#footnote-4). 因而 ImageNet 数据集可以处理绝大多数的计算机视觉任务. 基于此, 人们在该数据集上训练处理许许多多的优秀模型, 同时它也为人们进行迁移学习提供了很大的便利.

鉴于此, 本章所提出的方法是利用在ImageNet数据集上预训练好的深度学习模型,提取新图片数据集(目标数据集)中每幅图像的高级特征表示,这样的策略不断缩短数据类别标签与高级语义之间的"鸿沟", 同时通过降维技术进一步地去除高级特征表示的信息冗余, 最后对其使用半监督学习(标签传播算法)对无标签样本进行分类.

人们对于标签传播算法的研究大多数是基于“聚类假设”和“流形假设”, 但其本质是基于 “相似样本拥有相似的输出” 这一**平滑假设**. 而深度学习有着提取样本高级特征的特性, 本章假设: 相似样本的高级特征做标签传播仍然拥有相似的输出. 基于此假设, 本文提出基于深度卷积神经网络提取图片的特征进行标签传播的算法, 为了更好的解释问题,本文称该流程为**算法 A,** 具体算法流程如下 (对于任意的):

**算法 A**:

1. 找到一个在 ImageNet 数据集已经训练好的模型, 记作 (本文使用 ResNet50), 为迁移学习做准备. 一般情况下模型分为和两个部分, 分别代表**特征提取层**和**输出层**.
2. 对于每个图片像素级别的特征的提取其特征即, 为了和1.2节的约定相对应,我们将其记作.
3. 利用公式 (1-18) 或 (1-20) 和(1-28) 或(1-29) 对分别进行LGC标签传播和 GRF标签传播获得不同的.

通过**算法 A**得到无标签样本的预测类别标签来实现无标签图像数据的半监督分类.虽然算法 A提出的模型的泛化性能比较好,但是,CNN提取的特征向量仍然可能带有部分冗余信息,为此,**算法 B更进一步，**引入了先降维(PCA 或者 LDA) 再进行标签传播的思想,具体流程如下:

**算法B**:

1. 找到一个在 ImageNet 数据集已经训练好的模型, 记作 (本文使用 ResNet50), 为迁移学习做准备. 一般情况下模型分为和两个部分, 分别代表**特征提取层**和**输出层**.
2. 对于每个图片像素级别的特征的提取其特征即, 为了和1.2节的约定相对应,我们将其记作.
3. 得到的的特征维度仍然是比较高的,其中也存在着冗余和噪声,我们对其做降维处理 (本文对无标签数据采用 PCA, 对有标签数据采用 LDA). 降维的记作.
4. 利用公式 (1-18) 或 (1-20) 和(1-28) 或(1-29) 对分别进行LGC标签传播和 GRF标签传播获得不同的.

## 2.2 数据集简介

Cifar 10是图像分类问题的基准数据集, 它十分适合对模型做一些简单的预测和微调. Cifar10 分为训练集和测试集, 它有个类别. 其中训练集包含个样本, 而测试集包含个样本, 且尺寸为的彩色图片, 因而数据集足以应对一些中小型模型的需求. 类的标记为: airplane、automobile、bird、cat、deer、dog、frog、horse、ship、truck 这些类是完全互斥的, 相互之间没有重叠.图2-1展示了 cifar10 的其中 64张图片, Cifar10 数据集的每张图片的像素级别的特征维度是3072, 直接将其作为浅层分类模型的输入会有很多的冗余信息,本章借助 CNN提取其高级语义特征极大的减少了数据的冗余信息量.

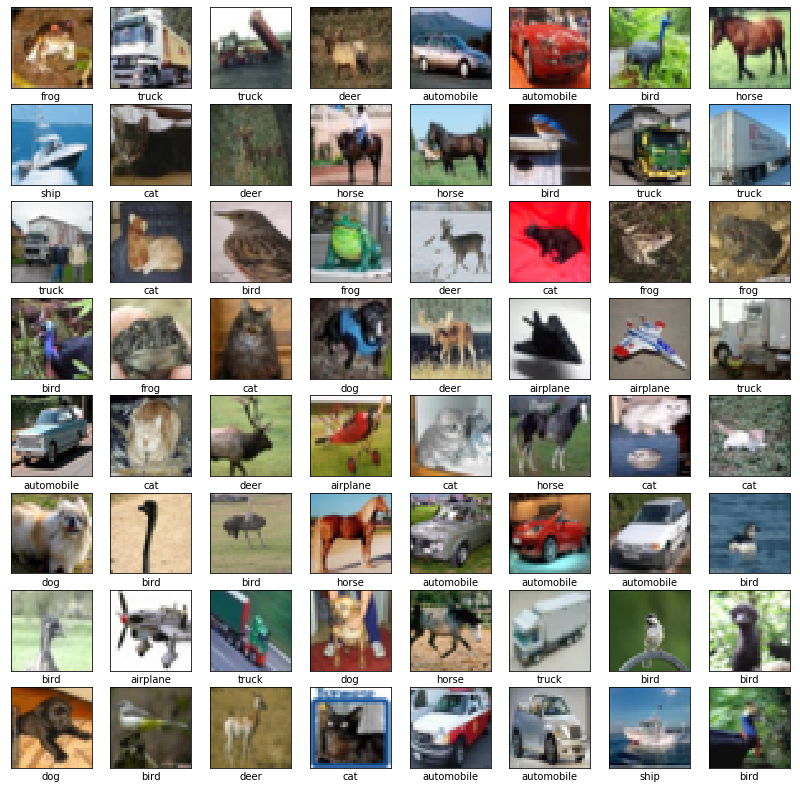


图2-1 Cifar10 部分图片展示

Figure 2-1 Cifar10 part of the picture display

## 2.3 实验结果与分析

图2-2以一张测试图片为例来说明算法A:

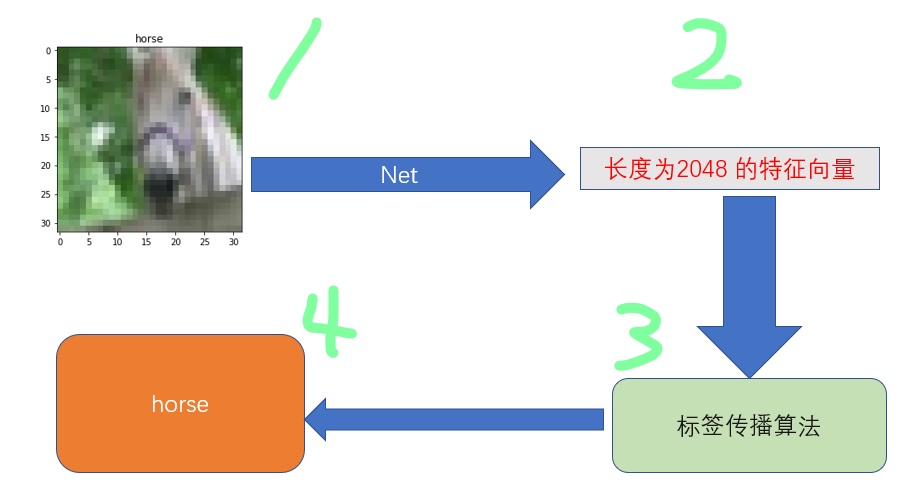


图2-2算法 A示意图

Figure 2-2 Algorithm A diagram

针对Cifar10数据集中的训练集中,我们随机选取6000个图像的样本构成的子集作为本章中的图像分类数据集.在这6000个样本中，我们假设部分数据存在类别标签,而剩余的其它数据无类别标签,这样便构成一个半监督学习需要的数据集.以分类精确度作为性能度量, 并且做了一些对比实验. 参与比较的方法有作为基准的最近邻分类法,基于原始图像像素特征的半监督学习方法,以及基于深度特征的半监督学习方法.固定6000个样本构成的本章的实验数据集后,我们每次随机选择有标签的样本,在下面每次实验设置中都要重复10次实验,所得到结果的均值和方差在图中均有体现.为了可视化需要,本章约定:

1. 1NN Feats:表示通过 CNN提取高级语义特征后再做最近邻分类;
2. 1NN + raw: 表示仅仅利用最近邻分类;
3. LGC + raw: 表示仅仅使用 LGC做标签传播;
4. LGC Feats:表示通过 CNN提取高级语义特征后再做LGC标签传播:
5. GRF + raw: 表示仅仅使用GRF 做标签传播;
6. GRF Feats:表示通过 CNN提取高级语义特征后再做GRF 标签传播

### 2.3.1 固定每类有标签样本数目为50观察邻域范围对实验的影响

我们首先固定每类数据中有标签样本数目为50,让半监督学习算法的参数:构图的邻域半径的范围从3变化到50,来测试所提出方法对该参数的鲁棒性.图2-3给出该实验关于邻域范围对结果的影响.

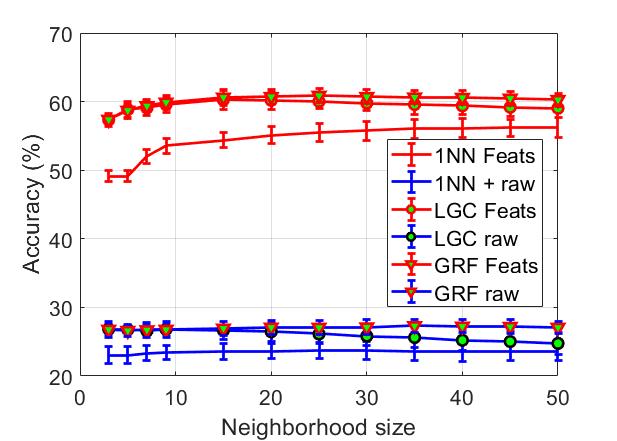


图2-3: 邻域范围变化的实验结果

Figure 2-3: Experimental results of neighborhood range changes

如图2-3所示，无论邻域范围如何改变,我们都可以看出:

1. 基于深度特征的算法的识别效果显著地好于基于图像的原始像素特征(相差准确率绝对值20%~30%);
2. 无论是对数据的原始特征还是深度特征,半监督学习方法的效果都好于有监督的近邻分类方法.
3. 无论是否利用 CNN提取高级语义特征,GRF都略高于 LGC;
4. 基于深度特征的标签传播算法(LGC与GRF)的与1NN之间的差距明显大于基于图像的原始像素特征的差距.

较好地最近邻参数选择范围也较广,比如18就是一个还不错的选择.因为图示的大多数算法当邻域尺寸超过18时,分类的准确度有所下降.

### 2.3.2 固定邻域半径为18观察有标签样本的数目对实验的影响

固定邻域半径k=18,我们让每类有标签样本的数目从10变化到250来观察有标签样本数目的变化对分类效果的影响.如图2-4所示,当每类有标签样本数目为10个时,半监督分类准确率达到高于50%;当每类有标签样本数目为250个时,半监督分类准确率达到高于接近70%.这充分说明了标签数量对图像分类效果的重要性.当每类样本有标签的数目是 50时半监督分类的准确度变化最快.

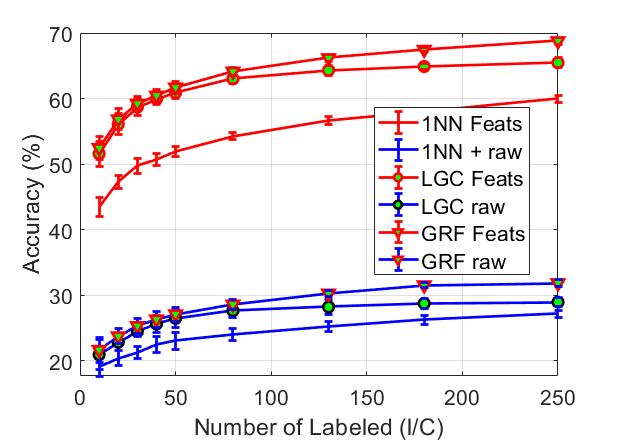


图2-4. 有标签样本的数目从10变化到250的实验效果

Figure 2-4. Experimental results with the number of labeled samples changing from 10 to 250

结合图2-3和图2-4可以看出: 深度学习技术提取特征后再做半监督学习得到的准确度明显比直接使用原始图像做半监督学习的准确度要高. 更具体是, 不管是深度特征还是原始特征使用 1NN (最近邻分类) 的分类准确度总是低于LGC和 GRF的分类准确度. 总体看来, 随着近邻的尺寸 (size) 或者有标签数据的样本数的增加, 分类的准确度均有一定的增幅.综上所述,算法 A的泛化性能相对传统的 1NN与 LGC, GRF 有着很大的优越性.

### 2.3.3 特征的维度对实验的影响

本章节测试了PCA, LDA 这两种降维方法在去除特征噪声,提高识别效果方面的性能(即算法B).固定每类数据中有标签样本数目为50,最近邻范围k=18, 图2-5给出了当PCA降维范围在10到2000时，半监督分类在深度特征上的效果.其中, 图2-6是图2-5的局部放大版.(同时也将1NN+Feats 加入对照实验中)

从图 2-5可以看出 1NN Feats与 LGCFeats, GRF Feats以及 LGC + PCA (先提取CNN深度语义特征,再降维之后再进行LGC 标签传播 )和 GRF + PCA (先提取CNN深度语义特征,再降维之后再进行GRF 标签传播)的准确度差距很大, 1NN Feats 总是比其他算法低 10个百分点左右.PCA降维的维度升高1NNFeats 的波动很小,而其他算法的波动相对较大.从图 2-6可以看出随着 PCA降低的维度的升高,没有使用 PCA降维的 LGCFeats或者GRFFeats在750维之前总是低于使用了 PCA降维技术的LGC+PCA或者 GRF+PCA,之后二者便趋于一致.

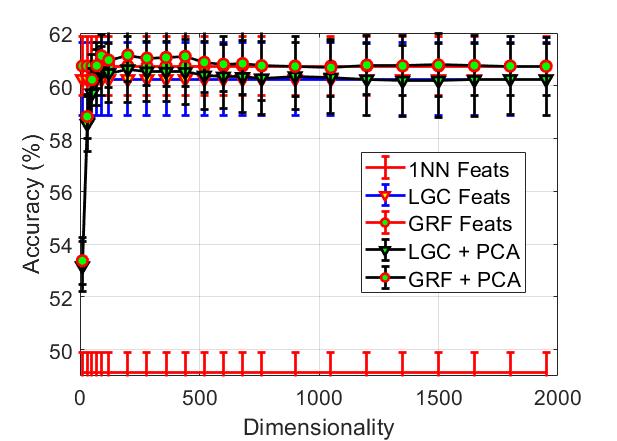


图2-5. PCA降维对基于深度特征的半监督算法效果

Figure 2-5. Effect of PCA descending dimension on semi-supervised algorithm based on depth feature

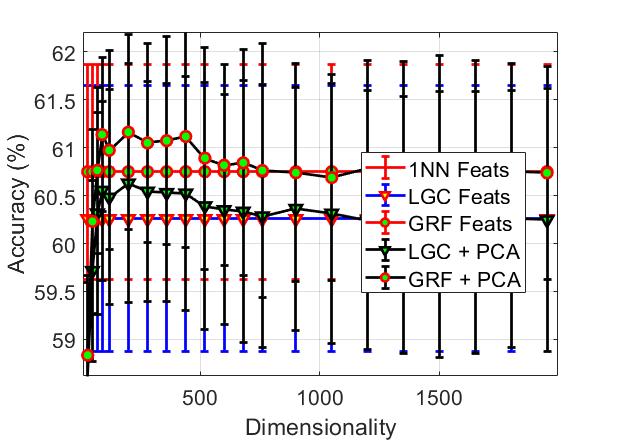


图2-6. 是图2-5的局部放大图，看起来更清晰

Figure 2-6. is a local amplification of figure 2-5which looks clearer.

换句话说,当PCA维数为200左右的时候,半监督算法得到比较高的准确率.而随着降维维数的升高,算法的性能反而开始下降,直到稳定在和不对深度特征降维的准确率相同.

图2-7给出的是基于有监督数据用LDA方法进行降维的效果图.注意LDA方法在维度是9维时能够实现最高的准确率.从图2-5和图2-7可以看出: 使用降维的技术可以提升分类的准确度左右. 从图2-7可以看出有监督的降维既可以得到更低的维度空间还可以更加有效的去除信息的冗余, 获得更高的分类准确度.

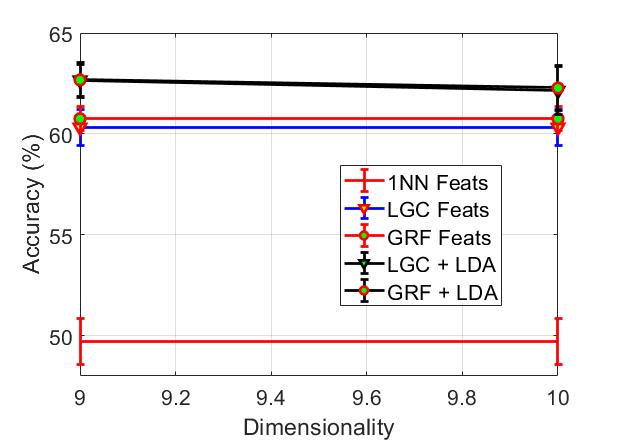


图2-7基于LDA方法降维的深度特征半监督学习效果图

Fig. 2-7 Depth feature semi-supervised learning effect diagram based on LDAmethod

## 2.4 本章小节

本章的实验结果是在Matlab软件环境中进行的,利用MATCONVNET（http://www.vlfeat.org/matconvnet/）实现的RESNET-50预训练CNN模型提取高级语义特征,再利用 1NN, LGC, GRFH 来优化图像分类模型的策略(简记作 CNN+)远远高于不使用 CNN的算法.如果从模型复杂度来考虑这是必然的,因为 CNN+的参数量远远高于不使用 CNN的算法.而对于1NN 算法使用 CNN虽然也极大的改善了算法模型的泛化能力,但是不是太明显,而借助无标签信息的基于图的标签传播算法(LGC 与 GRF) 提升的幅度要比 1NN要大得多,同时基于图的标签传播算法的泛化能力相对要比 1NN 高.换句话说,本章提出的算法 A对于图像分类的模型的泛化能力,相对于 1NN+raw, 1NN+Feats, LGC+raw, GRF+raw有着极为显著的改善. 同时算法 A的降维版本算法 B在PCA嵌入的维度相对较低

的情况下要明显优于算法 A.

# 第 3 章 利用深度特征进行标签传播并微调深度卷积网络

## 3.1 算法引入

虽然第2章介绍的算法 A与算法 B在图像分类任务上有着十分傲人的成绩,但是,考虑到它们是基于 ImageNet 数据集进行训练得到的深度卷积神经网络模型,该深度卷积神经网络模型并没有在我们的目标数据集上进行拟合,故而本章考虑对该深度模型进行微调.本章仍然采用第2章介绍的 Cifar10 数据集进行实验.

从第2章可以看出LGC与 GRF算法对于本文的图像分类任务而言差别不是很大,故而本章仅仅考虑 LGC标签传播算法,同时为了充分利用 GPU的并行计算的优势,本章的实验均使用 Python作为编程语言.

## 3.2 基于 LGC 的深度卷积神经网络微调算法

### 3.2.1 算法简介

本节提出方法的主体思想是,让半监督学习和深度CNN网络模型调优相互结合起来.让这两个工作互相促进,使得模型的性能更好的优化.微调(fine tuning): 利用在源数据集(一般是一个样本量比较大的数据集)上训练好的深度神经网络”基”模型来指导目标数据集(相对源数据集而言,样本量比较小,且和源数据集拥有相同或相似的数据分布相似)进行学习.具体的操作是复制基模型的全连接层之前的所有结构和参数,而修改基模型的全连接层以适用于目标数据集(当然,如若目标数据集也很大,可以将基模型的结构和参数冻结的更少些). 换言之,微调是一个有点儿类似于老师(基模型和源数据集)教学生(基模型的子模型与目标数据集)的一个优化过程.

本节提出的基于 LGC的深度卷积神经网络微调算法流程如下:

**算法C**:

1. 找到一个在 ImageNet 数据集已经训练好的模型, 记作 (本文使用 ResNet50), 为迁移学习做准备. 一般情况下模型分为和两个部分, 分别代表**特征提取层**和**输出层**.
2. 对于每个图片像素级别的特征的提取其特征即, 为了和1.2节的约定相对应,我们将其记作.
3. 通过2的方式获得有标签数据集的高级语义特征向量组,且在进行 LGC操作时需要将打乱;
4. 对于无标签数据集以2的同样方式获取高级语义特征向量,但是当无标签数据集的样本量比较大时,考虑将其按照某种方式进行划分而得到互不相交的个集合,且为所有无标签数据集的高级语义特征向量组成的集合.对于无标签样本量比较小的情况,只需为 1即可;
5. 对于,利用公式 (1-18) 或 (1-20) 和 (1-28) 或 (1-29) 对进行 LGC 标签传播操作, 得到对应的标签集;
6. 令, 且.将作为对应的无标签数据的伪标签;
7. 将拥有标签的数据集与带有无标签的数据集依次对基网络微调各一次;
8. 不断重复2-7直到模型收敛.（从第二步到第七步，不包括第一步）

### 3.2.2 实验结果与分析

在第2章仅使用了Cifar10数据集的训练集中的其中一部分数据(6 000个样本), 为了充分利用无标签数据集,同时考虑到有标签数据集样本数对模型泛化能力的影响,本小节将 Cifar10的全部训练集(50 000 个)均纳入考虑范围,且从其中均匀地选择800个样本作为有标签数据集,其余数据集当作无标签数据集.在训练的过程中为了观察 CNN在未知数据的泛化能力,本小节亦将 Cifar10的测试集(10 000个)作为验证集来避免 CNN出现过拟合的现象.

本节的实验是利用了Pytorch1.0.0来训练 CNN, 而利用scikit-learn 来训练 LGC模型(具体使用的是semi\_supervised. LabelSpreading, 即正则化形式的LGC 算法的实现).限于硬件设备的配置,对于 LGC算法来说,无标签数据集过于庞大,本节将其按照 20000为单位进行划分,得到三个子集分别记作 T1, T2, T3(数据集无需打乱).而800个有标签的数据集每次在标签传播的训练阶段都需要打乱,且记作 T0.按照算法 C的操作来拟合数据,最终的实验结果见图3-1.

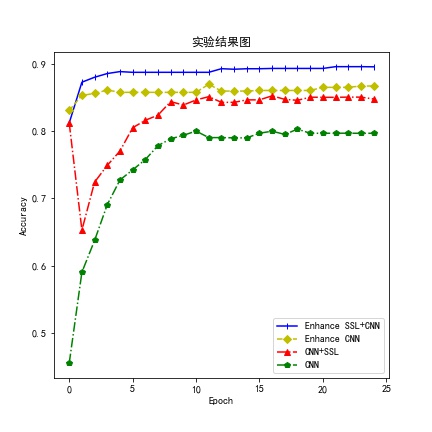


图3-1算法 C的实验结果图

Fig. 3-1 Experimental results diagram of algorithm C

图3-1的图示说明:

1. CNN: 表示仅仅使用 CNN在CIfar10的测试数据集上去预测模型的泛化能力;
2. CNN+SSL: 表示使用 CNN提取高级语义特征后,再利用 LGC做标签传播;
3. Enhance CNN: 表示使用 CNN提取高级语义特征后,再利用 LGC做标签传播得到无标签数据集的伪标签当作 CNN的输入的标签再次训练 CNN, 以此来微调 CNN模型的过程中在CIfar10的测试数据集上去预测模型的泛化能力;
4. Enhance SSL+CNN: 表示算法C 的模型.

通过实验分析发现, 本章提出的算法C的性能明显优于算法A的性能, 更加坚定的说明了深度学习技术相对于浅层机器学习有着很大的优越性,深度学习对标签噪声存在比较好的抗干扰能力. 通过上面的实验, 我们还可以看出: 在半监督学习任务中, 深度学习技术可以很好的提升分类的准确度.同时,我们从图3-1可以看出 CNN模型在未知数据集上的表现因加入了无标签信息, 故而其泛化能力也大大的得到了改善; 并且算法 C的在无标签数据集上的准确度也提升了 5个左右的百分点.

## 3.3 本章小节

算法 C不仅仅提升了LGC对无标签数据集的泛化能力,同时也提高了 CNN对未知数据的泛化能力.换句话说,基于 LGC的深度卷积神经网络微调算法有效的融合了LGC 和 CNN的优点,因而在图像分类任务上,它是一个十分有前景的具有实用价值的研究思路.

# 第 4 章 总结与展望

## 4.1 研究总结

本论文提出了3 个相互关联的算法,它们在不同场景将有不同的用途.针对LGC 在处理无标签数据的样本量很大的情况,算法 C提出随机打乱有标签数据,然后再将无标签数据集划分为几个不同的子集;接着将打乱后的有标签数据集与无标签数据集的子集合并,然后一起送入 LGC模型进行训练得到其无标签数据集的伪标签,以同样的方法获取所有无标签数据集的伪标签;将伪标签当作无标签数据集的标签和有标签数据一起送入 CNN进行微调来进一步修改伪标签;不断的迭代上述过程直至算法收敛,便可以得到一个相对不错的模型.

本文不仅仅是提出了算法 A, B, C, 它们是可以拓展出很大版本的算法的. 比如你可以更换基网络模型, 又或者你可以更换降维方法、改变权重矩阵的计算方法等.

## 4.2 研究展望

本文提出的算法 C还有很大的改进空间, 比如可以将降维技术也融合进算法 C进一步的去除信息的冗余. 除此之外, 本文的求取权重矩阵的思路也有很大的改进空间, 比如直接利用深度学习模型获取一个具有学习能力的权重矩阵.

本文提出的模型是以3个模块为基础来拟合模型的:

1. 依据距离度量求取权重矩阵;
2. 借助卷积神经网络提取图像的深度语义特征向量.
3. 标签传播算法借助这两个独立模块得到的权重矩阵和特征向量获取无标签数据集的标签.

其中, 1 和 2 是相互独立的过程，而3 是最为耗时间的过程. 为了提高模型的计算效率, 可以考虑如何将 1, 2, 3 这三个过程统一为一个过程, 实现真正的端到端的模型.

由于卷积神经网络模型具有分层抽象的功能, 我们也可以考虑在不同的卷积层提取不同的特征向量, 然后将这些向量拼接为一个更长的向量以此来提高模型的泛化能力. 具体流程见图 4-1:

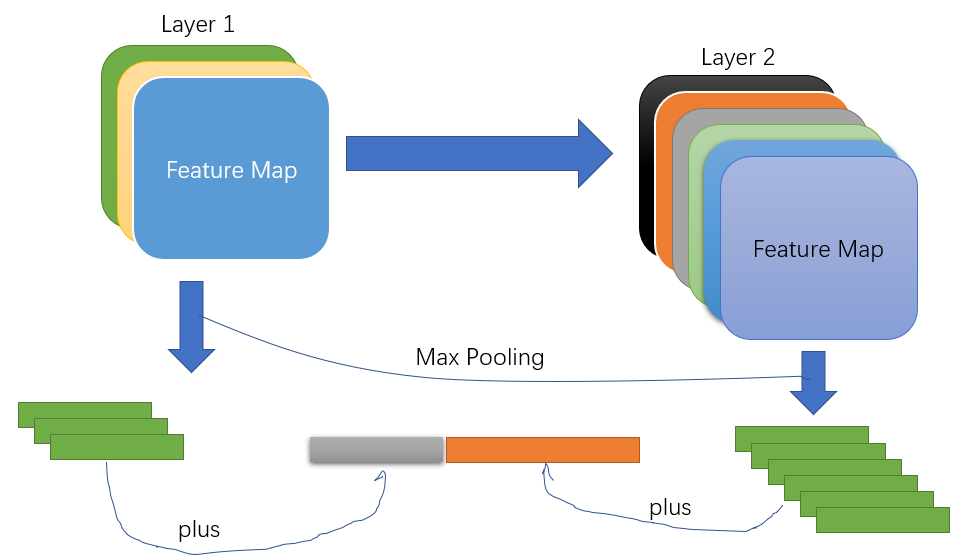


图 4-1: 多尺度特征提取

Figure 4-1: Multi-scale feature extraction

图 4-1 展示了两层卷积层提取不同尺度的特征的流程. 更一般地, 假设第 层的特征图 的宽度和高分别为: 和 , 先对 执行 的 Max Pooling(池化操作), 得到 的特征表示 . 接着我们将 按通道维执行加法运算, 便可以得到一个一维的张量 . 最后将所有的一维张量 进行拼接得到一个高维的一维张量 , 再对 执行标签传播算法.

这种多尺度的消息融合可能会进一步的提升模型的泛化能力.

# 参考文献

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, & S. Ma, et al., Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, 2015

[2] A. J. Smola and B. Schölkopf, On a kernel-based method for pattern recognition, regression, approximation, and operator inversion, Algorithmica[J] 1998, 22, 211-231

[3] Sivic, Zisserman. Video Google: a text retrieval approach to object matching in videos[C]. international conference on computer vision, 2003: 1470-1477

[4] Presenters Peter, G. B. Enser, Christine J. S, Paul H. Lewis and Jonathon S. Hare.The Reality of the Semantic Gap in Image Retrieval [J].1st International Conference on Semantic and Digital Media Technologies, 2006

[5] 郑浩,廖梦怡.一种改进半监督学习的图像语义自动识别方法[J].科技通报,2014,30(02):239-241

[6] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324

[7] 卢宏涛,张秦川.深度卷积神经网络在计算机视觉中的应用研究综述[J].数据采集与处理,2016,31(01):1-17

[8] Goldberg Y. A primer on neural network models for natural language processing[J]. Journal of Artificial Intelligence Research, 2016, 57(1): 345-420

[9] Qiang Gan, Yu Wei, Conrad, M. Locally-connected multilayer neural networks consisting of enzymatic neurons[P]. Circuits and Systems, 1991. Conference Proceedings, China., 1991 International Conference on,1991

[10] Wu J, Wang Y, Wang Z, et al. Deep k-Means: Re-Training and Parameter Sharing with Hard Cluster Assignments for Compressing Deep Convolutions[J]. international conference on machine learning, 2018: 5359-5368

[11] Nagi J, Ducatelle F, Caro G A, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition[C]. international conference on signal and image processing applications, 2011: 342-347

[12] 张顺,龚怡宏,王进军.深度卷积神经网络的发展及其在计算机视觉领域的应用[J/OL].计算机学报,2017:1-29

[13] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet Classification with Deep Convolutional Neural Networks[C]. neural information processing systems, 2012: 1097-1105

[14] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359

[15] Zhou D, Bousquet O, Lal T N, et al. Learning with Local and Global Consistency[C]. neural information processing systems, 2003: 321-328

[16] Zhu X, Ghahramani Z, Lafferty J D, et al. Semi-supervised learning using Gaussian fields and harmonic functions[C]. international conference on machine learning, 2003: 912-919

[17] Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician. 1992, 46 (3): 175–185. doi:10.1080/00031305.1992.10475879

[18] (美) 索利姆著. 朱文涛, 袁勇译. Python计算机视觉编程[M].北京: 人民邮电出版社, 2014.7

[19] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324

[20] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105)

[21] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

[22] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., &Anguelov, D. &Rabinovich, A.(2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9)

[23] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., &Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826)

[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778)

[25] Huang, G., Liu, Z., Weinberger, K. Q., & van der Maaten, L. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (Vol. 1, No. 2)

[26] Abdi H, Williams L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433-459

[27] Ueki K, Hayashida T, Kobayashi T, et al. Two-dimensional Heteroscedastic Linear Discriminant Analysis for Age-group Classification[C]. international conference on pattern recognition, 2006: 585-588

[28] 周志华.机器学习[M].北京: 清华大学出版社,2013.3

[29] 王秀秀. 基于稀疏图的半监督学习方法研究[D].西安电子科技大学,2013

[30] Wang J, Wang F, Zhang C, et al. Linear Neighborhood Propagation and Its Applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(9): 1600-1615

[31] Liu W, Chang S. Robust multi-class transductive learning with graphs[C]. computer vision and pattern recognition, 2009: 381-388

[32] 盛振国,王立国.改进的LLGC高光谱图像半监督分类[J].哈尔滨工程大学学报,2017,38(07):1086-1092

[33] 白本督,范九伦.基于稀疏分解的局部全局一致性学习算法[J].西安邮电大学学报,2015,20(03):65-70

1. [https://blog.csdn.net/qq\_23301703/article/details/79280977](mhtml:file://D:\API\A\Xwtrite\硕士毕业论文\查重2\基于深度学习的标签传播的图像分类研究（全文对照）.mht!https://blog.csdn.net/qq_23301703/article/details/79280977) [↑](#footnote-ref-1)
2. [https://bbs.csdn.net/topics/392382530](mhtml:file://D:\API\A\Xwtrite\硕士毕业论文\查重2\基于深度学习的标签传播的图像分类研究（全文对照）.mht!https://bbs.csdn.net/topics/392382530) [↑](#footnote-ref-2)
3. <http://www.mamicode.com/info-detail-1373869.html> [↑](#footnote-ref-3)
4. [https://blog.csdn.net/haoji007/article/details/82988514](mhtml:file://D:\API\A\Xwtrite\硕士毕业论文\查重2\基于深度学习的标签传播的图像分类研究（全文对照）.mht!https://blog.csdn.net/haoji007/article/details/82988514) [↑](#footnote-ref-4)