

迁移学习实现猫狗图像分类

邓心一, 姜百淳, 刘姜旺

(北京邮电大学 信息与通信工程学院, 北京 100876)

摘 要: 不少于 200 字, 应完整概括出文章的目的、方法、结果及结论; 简洁, 排除常识内容, 避免重复题目; 独立, 不得引用文中参考文献号、图号和公式号; 具体, 尽量用具体数字来说明该项工作取得的进展或成效, 例如某项性能指标提高了百分之多少, 避免“效果很好”这类的含糊其辞; 便于收录, 摘要、题目中避免包含公式、上下标等, 以方便 EI 等文摘和题录数据库收录文本数据。高质量的摘要有利于文摘被国际权威数据库收录, 及引起同行的重视。用第 3 人称, 建议采用“对……进行了研究”、“报告了……现状”、“进行了……调查”等记述方法, 不必使用“本文”、“作者”等作为主语。缩写词应提供中文全称。

关 键 词: 迁移学习; 图像分类; 卷积神经网络; 模型融合

中图分类号: V221+.3; TB553

文献标识码: A

文章编号: 1001-5965 (XXXX) XX-XXXX-XX

由于卷积神经网络在机器视觉任务中的优秀表现, Kaggle 竞赛中与图像相关的题目, 绝大多数参赛者采用的是深度学习的解决方案。本小组选择了经典的图像分类任务, 在问题的解决过程中, 熟悉了一般 CNN 网络模型的构建方式, 尝试运用一些数据挖掘的技巧, 提升模型的表现, 最终模型分类的准确率令人满意, 也在一定程度上减少了模型训练的代价。挖掘任务的详细解决过程陈述如下。

1 任务描述

1.1 竞赛题目

竞赛题目选用 Kaggle 竞赛中, 类型为 Playground 的 Dogs vs. Cats Redux: Kernels Edition。题目提供已标注的训练集和未标注的测试集, 要求参赛者提交测试集中每一张图片为狗的概率。提交的结果用公式 (1), 即对数损失来衡量。

$$Loss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1)$$

式中, n 是测试集的图片数量, \hat{y}_i 是图片为狗的预测概率, 图片实际为猫时, y_i 取 0, 否则取 1。对数损失越小, 预测结果越好。

1.2 数据集

数据集分为训练集和测试集, jpg 格式。训练集包括 12500 张猫的图片 and 12500 张狗的图片, 大小为 543MB, 分辨率不等, 来自某个宠物领养的网站, 推测为程序自动爬取获得, 分类标签存在一定程度的错误。测试集包括 12500 张猫与狗的图片, 大小为 271MB, 来源与训练集相同。

2 问题定义

赛题是一个典型的图像分类问题, 即根据一定的分类规则将图像自动分到一组预定义类别中。图像分类在多领域有着广泛的实际应用。实际应用中, 图像分类任务会受到视角的变化, 大小的变化, 物体的形变, 遮挡, 光照条件, 背景干扰和多种子类型等各方面因素^[1] 的挑战。对于传统的数据挖掘算法, 如 K 近邻算法 (K-Nearest

Neighbor, KNN), 支持向量机 (Support Vector Machine, SVM), 图像分类任务是非常复杂甚至难以胜任的。卷积神经网络, 以其在图像识别方面的优异表现, 从诞生起就备受瞩目。因此, 本小组选择 CNN 作为图像分类算法的核心部分, 再利用迁移学习的思想, 减少计算量和过拟合的风险。最终将数据挖掘任务划分为以下三个子任务: (1) 图像数据的预处理; (2) 数据特征的提取; (3) 特征处理, 对提取出的特征进行分类。

3 解决方案

经过多次实验和迭代, 本小组最终提出了对图像仅进行简单预处理, 以多个卷积神经网络作为特征提取器, 训练深度前馈网络 (Feedforward Neural Network, FNN) 作为分类器, 最后进行模型融合的解决方案。

3.1 方案提出

较深的卷积神经网络的训练和计算, 需要相当大的计算能力。Google 团队 2016 年提出的 Xception 模型, 使用了 60 块 Nvidia K80 GPUs, 在 ImageNet 数据集上训练了 3 天^[2], 达到 79% 的 Top1 准确率。小组最终采用的模型为 Inception 系列的三个深度均达到 100 层以上的 CNN, 而拥有的计算资源仅仅是笔记本上的入门级 GPU GTX 1050m, 这对计算资源提出了不小的挑战。

此外, 虽然相对于小组拥有的计算资源, 题目所给训练集数据量较大, 但是对于深度神经网络的训练, 提供的数据又稍显不足, 容易出现过拟合的情况。针对这些问题, 在查阅相关资料后, 决定采取迁移学习的方式解决。在深度学习的分类器之外, 小组也采用了经典的 SVM 分类器与 FNN 作出对比。由于采用提取特征向量的方式, 需要数据集相对固定, 因此难以在短时间内尝试多种数据预处理方式。本小组把模型调整的重心放在网络的后半部分, 节约出训练网络的时间来对模型做出优化。因此, 形成了简单预处理 → CNN 提取特征向量 → FNN 作为分类器 → 模型融合的方案结构。

以下将从 CNN 网络、迁移学习和模型融合三个方面来进行算法的介绍。

3.2 Inception 系列网络

方案选取了在 ImageNet 数据集中, 表现优秀的 Inception 系列网络, 作为特征提取部分的模型。Inception 系列网络是 Google 团队提出的一系列不断改进的 CNN 模型, 名字来源于电影盗梦空

间 (Inception) 中的一句台词: We need to go deeper。经典的 CNN 网络, 由负责特征提取的卷积层和负责降采样的池化层交替堆叠而成。Inception 系列网络在经典 CNN 结构的基础上, 改进了卷积层的结构, 取得了一定的性能提升。

CNN 中的卷积核大小, 很大程度上影响着提取到的特征图像。一般来说, 卷积核越大, 感受野越大, 越能发现全局的图像特征; 卷积核越小, 对图像局部的特征更加敏感。传统的 CNN, 每一层采用的是一组大小相同的卷积核。为了加强网络的泛化能力, 对不同尺度的目标都能良好地识别, InceptionV1 提出在同一层采用多种大小的卷积核, 将卷积得到的特征向量进行拼接, 作为下一层的输入。这就给下一层网络提供了不同尺度上的特征向量信息, 使网络能够自由选择, 提高了网络的泛化能力。

多个卷积核带来了性能的提升, 也带来了极大的计算量。为了减少网络的计算, InceptionV1 提出了 1×1 的卷积核。 1×1 的卷积核并不会改变特征图像在空间上的相对关系, 而是把同一空间位置上的多个通道融合在一起, 通过减少输出特征图像的通道数量, 达到减少计算量的目的。

为了进一步减少神经网络中的计算, Inception 用两个 3×3 的卷积核代替了一个 5×5 的卷积核, 网络参数由 25 个下降到 18 个, 还进一步增加了网络的深度。更重要的是, 卷积核的感受野是不变的, 仍然能够包含 5×5 的特征图像区域。

此外, $1 \times n$ 和 $n \times 1$ 的网络也被引入到 Inception 网络结构中。

Xception 网络则在另一个角度上对传统 CNN 的结构作出了改进。

3.3 迁移学习

迁移学习

3.4 模型融合

模型融合

4 数据处理

数据处理

5 结果分析

结果分析

6 正文

6.1 量、单位、公式

6.1.1 公式编排

《北京航空航天大学学报》一般不编排单独的符号表,对于公式中的变量含义需要说明的,请在公式后的段落中,采用“式中: A 为某某; B 为某某; ……”的方式加以说明。

$$p_1(h) = \frac{n_{\text{He}}RT}{V} - \rho_{\text{He}}gh \quad (2)$$

式中: n_{He} 和 ρ_{He} 分别为艇囊内部氦气的物质的量和氦气在温度 T 时的平均密度; $V=36893.426\text{ m}^3$ 为艇囊体积; $T=216.65\text{ K}$ 为艇囊内稳态温度; h 为距离艇囊中心轴线 (x 轴) 的垂直高度. 请使用 **Mathtype** 编辑. 公式中字体的定义尺寸为 10 磅, 上标/下标 68%, 次下标/上标 42%, 符号 150%, 次符号 100%(设置方法: **Mathtype**-尺寸-定义). 长公式如需转行, 应在记号 $=$, $+$, $-$ 等之后断开, 而在下一行开头不再重复这一记号。

6.1.2 量和单位

有关记号的使用应符合国家标准,例如: \sin^{-1} 应为 \arcsin , ctg 应为 \cot , tg 应为 \tan , 不要使用非国家法定单位, 如 ppm 等表示法已要求停止使用 (rpm 应写为 r/min); 除 Re , Ma (其中 e , a 不是下标) 等几个特征数外, 变量应使用单个字母表示, 可以带上标和下标 (否则由多个字母表示单个变量, 易被误解为多个变量相乘)。

6.1.3 字体

矩阵、向量请用粗斜体表示, 变量用一般斜体表示; 下标字母若为说明性的 (如英文缩写) 则用正体表示, 若为量和变动性数字及坐标轴的符号则用一般斜体表示 (设置方法: **Mathtype**-样式-定义-高级)。

所有文中出现的符号请另附文档说明其是变量、向量等, 并说明各变量上下标的含义, 以便编辑确定它们应采用的排版字体 (变量符号说明表)。

请作者对易于混淆的字母和数字, 如数字 0 和字母 o, 英文 a 和希腊字母 α , O, P, S, C 等的大小写, 批注“英大” (代表英文大写)、“数字 0”、“希小” (代表希腊字母小写) 等。

6.2 图、表

图、表需给出中英文图题、表题 (子图也需给出图题), 但图表中图例、线型说明等一律用中文。

图表一般不超过 7.7 cm 宽。金相图和计算机云图, 其中的比例尺等字编辑过程中都不再重贴, 按照照片处理, 如有这两类图请保证美观清晰, 字体用 times new roman。

6.2.1 图片

对于函数曲线图, 采用全框图, 并注意检查以下各项:

- 1) 横纵坐标的标目 (即变量名), 尽量使用国标变量符号, 变量名要在正文中交待, 且与正文中符号一致; 若正文中无, 也可使用中文名称。
- 2) 坐标轴标目的量纲, 对于无量纲化或无单位的, 请注明“无单位”。
- 3) 坐标轴上的刻度线朝内, 刻度值完整 (坐标轴始末点均应有完整刻度值)。
- 4) 不同线型或图符是否有说明。
- 5) 是否矢量图格式, 从软件中输出或拷贝矢量图格式直接插入文档中, 避免用拷屏办法插图图片, 否则后期无法编辑。
- 6) 类似图片尺寸尽量相同。

《北航学报》自 2014 年起可提供彩版印刷, 如有彩印需求请作者在“出版工作单”中注明。若不需彩印, 请作者作图时注意用可区分的线形或符号区分不同曲线, 以保证黑白图清晰可分辨。

图中文字均用中文或变量名称表示!

图片样例见图 1 和图 2 (目前是位图格式, 不

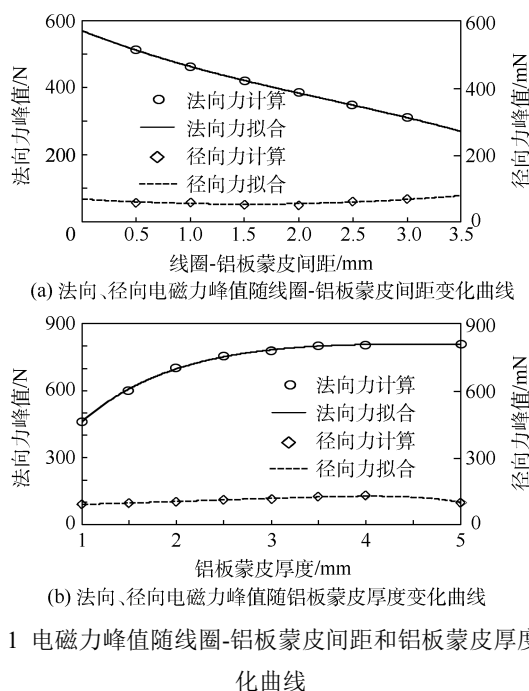


图 1 电磁力峰值随线圈-铝板蒙皮间距和铝板蒙皮厚度变化曲线

Fig. 1 Curves of electromagnetic force peak changing with coil-aluminum-plate gap and thickness of aluminum plate

能编辑，作者应提供可编辑的矢量图)。

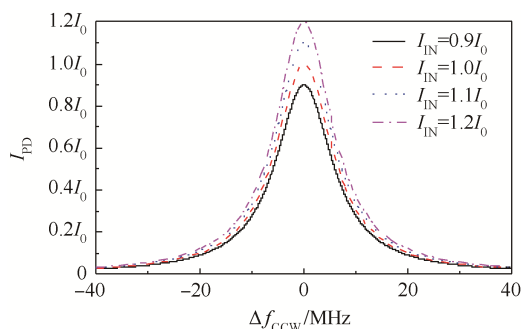


图2 谐振腔输入光强波动对谐振曲线的影响

Fig. 2 Fluctuation influence of the resonator's input intensity on resonance curve

6.2.2 表格

请使用三线表。选中表格，点右键打开“边框和底纹”，可对表格的边框等格式进行编辑，三线表的一般格式见表1。

表1 传输线积冰条件

Table 1 Icing conditions of transmission line

编号	直径/m	静温/K	时间/min
4	0.0349	268.15	30
5	0.01905	268.15	30

6.2.3 计算、实验

文章以数值计算为主要内容的，应给出所求解的方程、重要的计算参数、初始或边界条件、难点问题的处理等，应对方法的适用性和计算精度估计有所说明；文章以实验为主要内容的，应说明实验设备、实验条件，对实验误差的估计等。便于同行重复再现所报道的内容，由于保密原因不便公开某些内容的，应向责任编辑说明。

7 参考文献

- 1) 引用文献应遵循“最新、关键、必要和亲自阅读过”的原则；
- 2) 参考文献应是公开出版物；
- 3) 应在正文中顺次引述（按在正文中被提及的先后来排列各篇参考文献的序号，所有参考文献均应在正文中提及）；
- 4) 文献条数15条以上，且有适量近两年文献；
- 5) 参考文献中作者为3人或少于3人应全部列出，3人以上只列出前3人，后加“等”或“et al”；

6) 参考文献中外国人名书写时一律姓前，名后，姓用全称大写，名缩写为首字母(大写)，不加缩写点；

7) 为便于国际交流，对外文文献按外文著录；对于中文文献首先按中文著录，同时提供英文对照，并在其后注“(in Chinese)”注意对中文期刊刊名应使用其标准译法（通常在文章首页页眉可以找到）。

具体样例详见文后参考文献部分。

表2 文献类型和标志代码

Table 2 Reference type and identification code

参考文献 类型	文献类型 标识	参考文献 类型	文献类型 标识
专著	M	学位论文	D
会议录	C	报告	R
期刊	J	标准	S
报纸	N	专利	P
汇编	G	档案	A
计算机程序	CP	电子公告	EB
数据库	DB	美图	CM
数据集	DS	其他	Z

表3 电子文献载体和标志代码

Table 3 Electronic literature and identification code

载体 类型	磁带 (magnetic tape)	磁盘 (disk)	光盘 (CD-ROM)	联机网络 (online)
标志 代码	MT	DK	CD	OL

8 其他有关事项说明

- 1) 文章应着重撰写创新性、关键性内容，并以一般专业人员看得懂为原则
- 2) 返回时间：修改稿一般应在10天内返回，或以责任编辑的要求为准。如作者不能按时返回，请向责任编辑说明情况
- 3) 返回文件（请从系统上传）：
 - ① 论文电子版（修改部分用不同颜色标识）
 - ② 论文修改说明（写明对专家及编辑部所提意见如何修改）
 - ③ 变量符号说明表（模板见下载园地）
 - ④ 稿件出版工作单（word版，模板见下载园地）；“稿件出版工作单”中有关事项请认真填

写，联系电话最好有手机。后期编辑及发行过程中，会根据作者填写的信息与作者联系解决稿件问题，联系方式及寄刊地址有变更的，请及时通知责任编辑

稿件修改期间请对修改稿仔细审读、精加工，一经排版，一般不允许做大的改动

4) 出版过程：责任编辑在编辑修改稿过程中常有疑问请作者答复补正，请作者配合及时答复；稿件修改符合要求后，责任编辑将根据文章页码经电子信箱发送缴纳版面费通知单，作者应根据通知单要求及时缴纳版面费；编辑部有权对文章进行文字性修改，使之符合出版体例、规范要求和篇幅限制；责任编辑在编完稿件后，将其转至总编辑处，按来稿先后顺次发表；文章出版后，免费向作者提供样刊和抽印本，每篇文章1本样刊及5本抽印本，如作者需要可另购样刊，刊款可随版面费一并缴纳

5) 提前发表：本刊一般发表周期为1年，作者若有特殊情况确实需要提前发表的，请提前向学术编辑联系及说明情况，编辑部可根据实际情况适当安排

9 结论

分点总结，列出具体的结论，其他背景、方法都不必赘述。不与摘要和前言重复。具体样例如下：

1) 算法可实现较为优异的检索性能，例如返回10张结果条件下算法检索正确率83.15%，召回率8.42%，在60张下正确率39.33%，召回率24.61%。

2) 算法提出单张图片的引入不会造成原图片库的特征向量集和主题概率分配发生严重畸变的两个假设在一定范围（待检索图片与原图库特征类似）内是成立的。

3) 算法的预备工作使检索范围由原先整个库缩小至某个子类中，虽使召回率有所损失，但检索时间得到较大的缩短。

4) 可预估对于特征较接近的图片库，比如人脸库，图片预备工作会产生较大的分类误差，且可能进一步影响检索性能。

为使本文提出的算法能处理各种类型的图片，仍需要优化预备工作和检索实现过程的各项参数。

10 模板中一些问题

1) 所有间距都是手动设置，可能与word模板有些差别。包括正文行间距、各级节标题前后行间距、文本字与字间距、页面设置（页边距）、双栏间距、公式前后间距、图表（标题）前后间距、页眉页脚间距等等

2) 字体设置：正文中文、英文均是五号字（10.5pt），而公式中设置为10pt，所以公式中数字会小于普通文本数字，如 $x = 5$ 和5；带单位的量采用siunitx生成的话也有这个问题，如速度为5 m/s和5 m/s。公式中上下标看起来与word版稍有差别；公式中 g 与word版 g 也不同，默认公式字体可能并不是times new roman，本模板里未设置。

3) 双栏设置，采用的是article模板twocolumn选项；multicol对浮动图表支持要差一些；twocolumn也有些问题，比如首页跨两栏的脚注，没找到更好的办法，这里使用了`\enlargethispage{}`预留出脚注位置，然后用tikz手动调node的位置。还有跨两栏的图表灵活性稍差，`{figure*}`。

4) 图表中英文题注，使用ccaption得到。公式中向量矩阵粗斜体可以使用`\bm`得到。

5) 参考文献，为了自动排序，引用方便，使用BibTex，但是参考文献格式不属于标准的，所以所有参考文献只使用misc这个entry，而且只用到misc中note这一个field，也就是把整条参考文献都放到note里了。工作量与word差不多，但是引用、增删排序更方便些。

6) 变量符号说明表，里面加了一列符号所在位置，需要用到本文件生成的辅助文件，里面有可以引用的label信息。

致谢 感谢某某……注意：首页注明基金项目后，文末不必再致谢。

参考文献 (References)

- [1] <http://cs231n.github.io/classification/>.
- [2] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[J]. arXiv preprint arXiv:1610.02357, 2016.

作者简介:

姓名 性别, 学历, 职称。主要研究方向: xxxxxxxxxxxx。

张某某 男, 博士研究生。主要研究方向: xxxxxxxxxxxx。

李某某 男，博士，教授，博士生导师。主要研究方向：
XXXXXXXXXX。

附录 A:

若确有特殊需要设附录的, 附录部分置于作者简介后, 标题为“附录 A:”、“附录 B:”……。公式用大写字母和数字顺序编号, 例如“(A1)”,“(A2)”。

Transfer Learning for Image Classification of Cats and Dogs

DENG Xinyi, JIANG Baichun, LIU Jiangwang

(School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: (与中文摘要内容对应, 英文摘要字数 150~200 个单词) 英文摘要应和中文摘要对应, 并请导师或专业人士把关, 保证摘要质量, 高质量的摘要有利于文摘被国际权威数据库收录, 及引起同行的重视。如果英文摘要比中文摘要更详细, 应另提供一份英文摘要的中文副本, 以便于本刊英文编辑检查英文。首次出现英文缩写时应注意写明全称。

英文摘要的撰写规范请参考本刊网站“下载园地”中的《Ei 文摘要求》。

Key words: transfer learning; image classification; convolutional neural network; model ensembling

Received: 2018-01-01; **Accepted:** 2018-01-01

URL: (None)

Foundation item: (None)

Corresponding author. Tel.: 010-8231xxxx E-mail: ____@bupt.edu.cn