

知网个人查重服务报告单(全文标明引文)

报告编号:BC202405021526569648106700

检测时间:2024-05-02 15:26:56

篇名: 文档图像分类方法研究与分析

作者: 梁奕宸

检测类型: 毕业设计

比对截止日期: 2024-05-02

检测结果

去除本人文献复制比: 16.9% 去除引用文献复制比: 15.2% 总文字复制比: 16.9%  
单篇最大文字复制比: 2.7% (基于机器视觉的常用生活物品识别研究)

重复字符数: [4931] 单篇最大重复字符数: [796] 总字符数: [29140]

0% (0)	0% (0)	中英文摘要等 (总1996字)
7.5% (417)	7.5% (417)	第一章绪论 (总5551字)
24.4% (1539)	24.4% (1539)	第二章图像处理相关知识 (总6320字)
20.3% (1529)	20.3% (1529)	第三章SwinTransformer模型的基本框架 (总7542字)
3.1% (134)	3.1% (134)	第四章基于SwinTransformer模型的文档图像分类实验 (总4342字)
0% (0)	0% (0)	第五章总结与展望 (总1050字)
54.8% (1070)	54.8% (1070)	第六章参考文献 (总1953字)
62.7% (242)	62.7% (242)	第七章致谢 (总386字)

(注释: 无问题部分 文字复制部分 引用部分)

1. 中英文摘要等 总字符数: 1996

相似文献列表

去除本人文献复制比: 0% (0) 去除引用文献复制比: 0% (0) 文字复制比: 0% (0)

原文内容

南开大学  
本科生毕业论文(设计)  
中文题目: 文档图像分类方法研究与分析  
外文题目: Research and analysis of document image classification methods  
学号: 2011631  
姓名: 梁奕宸  
年级: 2020 级  
专业: 软件工程  
系别: 软件工程  
学院: 软件学院  
指导教师: 张玉志  
完成日期: 2024年5月2日

关于南开大学本科毕业论文(设计)的声明

本人郑重声明: 所呈交的学位论文, 是本人在指导教师指导下, 进行研究工作所取得的成果。除文中已经注明引用的内容外, 本学位论文的研究成果不包含任何他人创作的、已公开发表或没有公开发表的作品内容。对本论文所涉及的研究工作做出贡献的其他个人和集体, 均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名:

年月日

本人声明：该学位论文是本人指导学生完成的研究成果，已经审阅过论文的全部内容，并能够保证题目、关键词、摘要部分中英文内容的一致性和准确性。

学位论文指导教师签名：

年月日

摘要

如今，随着信息化时代的到来，越来越多的文档被数字化，因此如何有效地管理、搜索和利用这些文档变成一个非常重要的问题。在这样的时代背景下，深度学习方法有了更广阔的前景，开展对于文档图像分类方法的研究与分析具有重要的理论价值和现实意义。文中首先对图像分类的相关知识和一些经典的深度学习框架进行了研究和对比，谈到了相比与传统机器学习方法的一些优势；介绍了注意力机制和移动窗口的引入，并深入研究了Swin Transformer模型是如何基于卷积神经网络和ViT的先验知识优化而来。

然后重点介绍了Swin Transformer模型中的关键模块，包括移动窗口内的自注意力机制和相对位置编码的部分；在初步了解模型后，我进行了三组对比实验来进一步研究影响Swin Transformer性能的因素，分别是数据集的分布、学习率调整策略和不同的预训练模型。

最后提出了模型未来的优化方向和应用场景，希望未来能够在更多的场景中看到更多性能更优秀的模型。

关键词：图像分类；深度学习；Swin Transformer；移动窗口机制；对比试验；

Abstract

With the advent of the information age, an increasing number of documents are being digitized, making it crucial to effectively manage, search, and utilize these documents. In this context, research and analysis of document image classification methods have significant theoretical value and practical significance. This paper first studies and compares relevant knowledge of image classification and some classical deep learning frameworks, highlighting the advantages compared to traditional machine learning methods. The introduction of attention mechanism and moving window is discussed, followed by an in-depth investigation into how the Swin Transformer model is optimized based on prior knowledge of convolutional neural networks and ViT.

The key modules of the Swin Transformer model are then emphasized, including the self-attention mechanism within the moving window and the part related to relative position encoding. After gaining a preliminary understanding of the model, three sets of comparative experiments are conducted to further study the factors affecting the performance of the Swin Transformer: dataset distribution, learning rate adjustment strategies, and different pre-training models.

Finally, future optimization directions and application scenarios of the model are proposed, hoping to see more superior-performing models in various scenarios in the future.

Keywords: Image classification; Deep learning; Swin Transformer; Moving window mechanism; Comparative experiments.

目录

第一章绪论	1
第一节研究背景和意义	1
第二节国内外研究现状	2
第三节研究的关键问题	2
1.3.1 传统机器学习分类方法的局限性	3
1.3.2 卷积神经网络在文档图像分类任务中的局限性	3
1.3.3 ViT (Vision Transformer) [2]和Swin Transformer	4
1.3.4 特殊文档数据集的针对性研究	5
第四节论文的研究目标	6
第五节论文的内容组织	7
第二章图像处理相关知识	8
第一节图像分类方法	8
2.1.1 卷积神经网络	8
2.1.2 图像分类经典框架	11
2.1.3 注意力机制 (Attention)	13
第二节常见预处理算法[10]	15
2.2.1 标准化 (Standardization)	15
2.2.2 数据增强 (Data Augmentation)	15
2.2.3 噪声过滤 (Noise Filtering)	15
2.2.4 正则化 (Regularization)	15
2.2.5 缺失值处理 (Missing Value Handling)	15
第三节常见图像特征提取算法	16
2.3.1 主成分分析 (Principal Component Analysis, PCA)	16
2.3.2 预训练模型的特征提取	17
第三章 Swin Transformer模型的基本框架	18
第一节 Swin Transformer模型基本结构	18

第二节 Swin Transformer重要原理.....20

3.2.1 SW-MSA与mask的实现.....20

3.2.2 相对位置编码.....24

第四章基于Swin Transformer模型的文档图像分类实验.....27

第一节实验环境.....27

第二节数据集来源和划分.....27

第三节数据预处理操作.....28

第四节对比试验.....28

4.4.1 数据集均衡对比.....28

4.4.2 动态学习率和固定学习率对比.....29

4.4.3 不同预训练模型对比.....30

第五节本章小结.....32

第五章总结与展望.....33

第一节主要工作.....33

第二节展望.....33

第六章参考文献.....35

第七章致谢.....37

2. 第一章绪论		总字符数：5551
相似文献列表		
去除本人文献复制比：7.5%(417)      去除引用文献复制比：6.1%(337)      文字复制比：7.5%(417)		
1	基于深度学习的脑外医学图像分类 唐安莹 - 《大学生论文联合比对库》- 2022-05-25	2.9% (160) 是否引证：否
2	2122_11_10008_080901_41824229_BS_001 BS - 《大学生论文联合比对库》- 2022-09-09	2.9% (160) 是否引证：否
3	基于人工神经网络的图像压缩与解压系统设计与实现 王甜 - 《大学生论文联合比对库》- 2022-05-31	2.4% (135) 是否引证：否
4	基于人工神经网络的图像压缩与解压系统设计与实现 王甜 - 《大学生论文联合比对库》- 2022-06-07	2.4% (135) 是否引证：否
5	1813039-阳铠行-基于半监督学习的牙齿单目深度估计 阳铠行 - 《大学生论文联合比对库》- 2022-05-05	1.4% (76) 是否引证：否
6	基于Transformer模型的肺癌病理图像二分类 张语嫣 - 《大学生论文联合比对库》- 2022-05-20	1.3% (73) 是否引证：否
7	201824110155_张语嫣_基于Transformer模型的肺癌病理图像二分类 张语嫣 - 《大学生论文联合比对库》- 2022-06-03	1.3% (73) 是否引证：否
8	基于深度学习的遥感影像超分辨率设计与实现 杨育瑾 - 《大学生论文联合比对库》- 2023-06-26	1.3% (72) 是否引证：否
9	基于深度学习的医学图像分割方法与应用 林嘉澍 - 《大学生论文联合比对库》- 2023-06-08	0.6% (33) 是否引证：否
10	201924100733_魏佳豪_基于RGB的动作描述生成研究 魏佳豪 - 《大学生论文联合比对库》- 2023-05-26	0.5% (30) 是否引证：否
11	基于卷积神经网络的艾德莱丝绸纹样的真假识别研究 吐玛热斯·阿力木 - 《大学生论文联合比对库》- 2023-06-06	0.5% (27) 是否引证：否
原文内容		

第一章绪论

第一节研究背景和意义

当我们谈论文档图像分类时，我们实际上是在谈论一种能够使我们的日常工作更加高效、智能化的技术。现代社会已经进入了数字化时代，大量的文档被电子化处理，这包括从合同、报告到表格等各种类型的文档。但是，这些文档的数量往往是庞大的，因此如何有效地管理、搜索和利用这些文档就成为了一个非常重要的问题。

首先，让我们考虑一个办公室的场景。办公室里可能会有大量的文件，包括合同、报告、会议记录等。通过文档图像分类

技术，我们可以自动将这些文档按照其内容或特征分类，比如将所有的合同放在一起，将所有的报告放在一起，这样就可以更方便地管理和查找这些文档了。而且，一旦文档被分类，就可以轻松地建立一个智能搜索系统，用户可以通过关键词或其他条件快速找到他们需要的文档，而不必手动查阅每一个文件。

除了办公室场景，文档图像分类技术还可以在许多其他领域发挥作用。比如，在图书馆或档案馆里，有大量的文献和档案需要管理和分类。通过文档图像分类技术，可以将这些文献和档案按照主题、作者、出版年份等信息进行分类，从而方便读者或研究人员查阅和利用这些资源。在金融领域，文档图像分类技术可以帮助银行或保险公司对大量的合同和文件进行自动化处理和管理，提高工作效率和准确性。在医疗领域，文档图像分类技术可以帮助医院或诊所对患者的病历和医疗记录进行分类和管理，支持临床决策和医疗服务。

除了提高工作效率和智能化服务外，文档图像分类技术还有助于信息安全和隐私保护。通过对文档中的敏感信息进行识别和分类，可以帮助组织和机构加强对这些信息的保护和管理，防止泄露和滥用。

综上所述，文档图像分类技术在促进信息化、提高工作效率、支持智能化服务等方面具有重要的研究价值和实际意义。随着科技的不断进步和应用场景的不断拓展，我们相信文档图像分类技术将会在未来发挥越来越重要的作用，并为我们的生活带来更多的便利和智能化体验。

## 第二节 国内外研究现状

图像分类领域，国内外的研究现状都在不断地发展和演进[1]。以下是一些国内外研究现状的主要方面：

传统方法：在早期，文档图像分类主要依靠传统的特征提取和机器学习方法，如SIFT、HOG等特征提取方法结合SVM、KNN等分类器进行分类。这些方法在一定程度上取得了一定的效果，但受限于特征的表达能力和泛化能力。

深度学习方法：随着深度学习[2]的兴起，特别是卷积神经网络（CNN）的成功应用，文档图像分类领域也开始使用[深度学习方法](#)。通过[使用卷积神经网络进行端到端的特征学习和分类](#)，取得了较好的效果。其中，一些经典的模型如AlexNet、VGG、ResNet等被广泛应用于文档图像分类任务。

迁移学习和预训练模型：迁移学习和预训练模型的出现为文档图像分类带来了新的机遇。研究者们发现，通过在大规模数据集上预训练的模型，如ImageNet上预训练的模型，可以在文档图像分类任务上进行微调，取得更好的效果。这种方法能够利用大规模数据集中学到的通用特征，提升文档图像分类的性能。

注意力机制和Transformer模型：最近，注意力机制和Transformer模型在文档图像分类领域也得到了广泛的关注。这些模型能够有效地捕捉文档图像中的长距离依赖关系和语义信息，提升了分类性能。Swin Transformer等新型模型的出现，进一步推动了文档图像分类的发展。

数据增强和对抗训练：数据增强和对抗训练等方法也被应用于文档图像分类任务中，以提升模型的泛化能力和抗干扰能力。通过在训练过程中引入随机扰动或对抗样本，可以增加模型对不同条件下的适应能力。

综上所述，国内外在文档图像分类领域的研究现状呈现出多样化和多方面的发展趋势，不断涌现出新的方法和技术，推动着文档图像分类技术的不断进步和完善。

## 第三节 研究的关键问题

作为当下深度学习领域的骨干网络，对Swin Transformer[1]模型具体结构的进一步了解是非常必要的，鉴于传统机器学习和卷积神经网络的先验知识，来研究Swin Transformer是如何在Vit上进行改进，在移动窗口内计算自注意力，并结合卷积神经网络的层次化结构，形成一个层次化的Transformer模型；除此之外，模型的准确性还涉及到很多超参数和数据集本身的问题，预训练模型涉及到的迁移学习的方法等，因此，为了研究不同因素对于模型性能的具体影响程度，我们需要进行多组的对比试验来进一步了解Swin Transformer模型。

### 1.3.1 传统机器学习分类方法的局限性

1. 传统方法通常依赖于手工设计的特征，例如颜色直方图、纹理特征和形状描述符等。这些特征需要领域专家花费大量时间和精力来设计，且往往难以捕捉到文档图像中的高级语义信息。

2. 传统方法的特征表示通常较为简单，缺乏对文档图像中复杂结构和内容的泛化能力。因此，当面对图像多样性和类别不平衡等挑战时，传统方法的性能往往受到限制。

3. 传统方法往往对图像的变换（如旋转、缩放、平移等）和噪声非常敏感，导致模型的鲁棒性较差。这使得传统方法在实际应用中往往难以处理现实场景中的复杂图像数据。

4. 传统方法通常采用基于统计学习的分类器（如支持向量机、随机森林等），在处理大规模数据时计算和存储成本较高，且模型的训练和推理速度较慢。

5. 传统方法中的参数通常需要手动调整，例如分类器的超参数和特征选择的阈值等。这需要领域专家具有丰富的经验和专业知识，且往往需要进行大量的实验来找到最优的参数设置。

### 1.3.2 卷积神经网络在文档图像分类任务中的局限性

数据需求量大：[CNN通常需要大量的标记数据进行训练，以学习有效的特征表示](#)。对于某些文档图像分类任务，可用的标记数据可能有限，这可能限制了CNN模型的性能。

对位置和尺度敏感：CNN是局部连接的神经网络，因此对于输入图像的位置和尺度变化非常敏感。这可能导致在不同分辨率或不同排列方式的文档图像上的性能下降。

可解释性差：CNN模型通常被认为是黑盒模型，难以解释其内部如何进行决策。在一些应用场景下，尤其是对于需要解释分类决策的任务，这可能是一个限制因素。

过拟合：对于小样本问题，CNN容易过拟合，特别是在训练数据不足时。这可能导致模型在未见过的数据上的泛化性能下降。

计算和存储成本高：CNN通常具有大量的参数和复杂的计算结构，因此在[训练和推理时需要大量的计算资源和存储空间](#)。这[可能对于资源受限的环境](#)来说是一个限制因素。

对输入数据质量敏感：CNN对输入数据质量的要求较高，例如噪声、模糊或失真可能会对其性能产生负面影响。

### 1.3.3 Vit (Vision Transformer) [2]和Swin Transformer

其实Swin Transformer就是一个使用了移动窗口的层级式的Vision Transformer，Swin来自于Shifted Window（移动窗口



), 希望 Vision Transformer像卷积神经网络一样, 也能够分成几个 block, 做层级式的特征提取, 从而导致提出来的特征有多尺度的概念。

下面我们对两种模型作一些关键性的对比:

在建模地尺度方面, ViT 将整个输入图像拆分成固定大小的图像块, 并将每个图像块视为序列输入, 然后通过 Transformer 模型进行处理。这种方法限制了模型对图像中局部结构的建模能力; 而Swin Transformer 引入了分层的注意力机制, 将输入图像分解为一系列的图像块, 并在不同阶段(stage)上交换局部和全局信息。这样, 模型能够更好地捕获图像中的局部和全局特征, 提高了模型对图像语义结构的建模能力。

在窗口地划分和自注意力地计算方面: ViT在整个图像上直接应用自注意力机制, 导致模型的计算复杂度较高, 尤其是在处理大规模输入图像时; Swin Transformer则引入了窗口式的自注意力机制, 将输入图像分割成一系列的小图像块, 并在每个窗口上独立地应用自注意力机制。这样, 只要窗口大小是固定的, 自注意力的计算复杂度就是固定的, 整张图的计算复杂度就会跟图片的大小而成的线性增长关系, 就是说图片增大了 $x$ 倍, 窗口数量也增大了 $x$ 倍, 计算复杂度也就乘以 $x$ , 而不是乘以 $x$ 的平方, 因此能够减少计算的复杂度, 使模型能够更高效地处理大规模的输入图像。

除此之外, 二者的区别还在于ViT 使用单一的 Transformer 编码器层来处理整个输入序列, 缺乏对不同尺度信息的有效建模能力; Swin Transformer则基于CNN的先验知识, 使用分层的特征表示, 将输入图像划分为多个stage, 每个stage包含多个 Transformer 编码器层。这样, 模型能够更好地处理不同尺度的信息, 并且有助于提高模型的泛化能力。

通过以上改进, Swin Transformer 在图像分类任务中能够更好地捕获图像的局部和全局特征, 提高了模型的性能和泛化能力, 并且在一些情况下能够使用更小的模型规模取得类似甚至更好的性能。

#### 1.3.4 特殊文档数据集的针对性研究

图片本次使用的数据集是由近代书籍中提取出来的页面图像, 相较于传统的公共数据集ImageNet[3]、CIFAR-10以及MINIST来说, 本次实验使用的数据集明显具有更大的规模, 图片的尺寸基本为 $1300 \times 2000$ 像素, 而更大尺寸的图片, 给实验使用的模型也提出了更高的要求:

计算资源要求: 更大尺寸的图片通常意味着更多的像素和更高的分辨率, 这会增加模型的计算负担。因此, 处理大尺寸的图片可能需要更多的计算资源, 包括更多的内存和更长的处理时间。

模型设计和参数调整: 需要设计具有更大容量的模型, 以便能够充分利用图像中的信息。此外, 可能需要对模型的参数进行调整, 以适应更大尺寸的输入数据。

内存和存储需求: 更大尺寸的图片需要更多的内存来存储和处理, 这可能会导致内存不足或存储空间不足的问题。因此, 处理更大尺寸的图片可能需要更大的内存和存储资源。

数据预处理和增强: 更大尺寸的图片可能需要更复杂的数据预处理和增强技术, 以确保模型能够有效地学习和泛化。这可能涉及到裁剪、缩放、旋转等操作, 以及其他数据增强技术。

模型的鲁棒性: 更大尺寸的图片可能包含更多的细节和复杂性, 这可能会增加模型的复杂性和过拟合的风险。因此, 需要设计具有更强鲁棒性的模型, 以应对更大尺寸图片带来的挑战。

除此之外, 数据集本身还存在各类型数据分布不均衡的问题, 从书籍中提出的页面图像, 正文页的占比显然是更高的, 这是在做文档图像分类时不可避免的事情。所以我们需要研究数据不均衡时对模型正确率的影响。

#### 第四节论文的研究目标

本文首先主要研究Swin Transformer模型本身, 与CNN以及Vision Transformer相比所提出的移动窗口内的自注意力机制, 以及类似于卷积神经网络的层级式结构, 其次就是在针对特殊数据集的针对性研究, 在不同情况下参数和数据本身的调整对模型分类准确率的影响。

在之前的研究中, 对于文档图像分类的主流方法仍是基于CNN模型(如ResNet、VGG等)对文档图像进行特征提取, 然后使用全连接层或其他分类器对提取的特征进行分类, 相较于之前的传统机器学习算法已经在准确率和性能上有了更好的表现。但自从2021 ICCV的最佳论文《Swin Transformer: Hierarchical Vision Transformer using Shifted Windows》的发表, 结合其之后在各大数据集上优良的表现, 都已经充分说明了Transformer在视觉领域上的强大潜力, 也将在未来各个领域取代卷积神经网络成为计算机视觉领域的骨干网络。如今我们的自有数据集中包含的都是大尺寸的图像, 而Swin Transformer模型中通过下采样层达到的对高分辨率图像较好的处理效果正好满足我们的实验需求, 所以此次实验我们也选用Swin Transformer来作为主干网络, 下面是本篇论文的主要研究内容:

1. 在原始论文以及相关文献的基础上, 了解模型的整体结构、自注意力机制和层归一化等每个组件的功能以及如何组合这些组件来构建完整的模型; 并基于模型架构图研究各个组件之间的连接方式和信息流动, 理解该模型是如何使用分层架构来处理大尺寸图像。

2. 通过对学习率的不同调整策略来优化模型训练过程, 以期望在比较稳定的情况下加快模型收敛的速度, 以更快速得到预测效果较好的模型。

3. 由于我们的数据集是由近代书籍中提取得来, 所以原始的页面类型比例显然是不均衡的, 正文页的比例将会远超其他类型的页面。因此为了更客观地评估模型的预测效果, 我引入常见的文档图像的开源数据集Tobacco3482与我们的自由数据集作预测结果上的对比; 此外也会对自由数据集本身作一些调整, 以研究数据集均衡与否对模型的准确率有影响。

4. 研究不同规模的预训练模型对预测准确率的影响。我们选用在同样规模数据集上进行训练的两个层数和尺寸不同的Swin Transformer模型来进行对比, 研究图片调整的尺寸大小和模型的层数对于模型性能的影响。

#### 第五节论文的内容组织

本文共分为六章, 各章的内容如下:

第一章绪论。介绍了本文的研究背景和意义, 要解决的问题和本文的主要工作。

第二章图像处理相关知识

第三章 Swin Transformer模型的基本框架

第四章基于Swin Transformer的文档图像分类网络

第五章总结与展望

### 3. 第二章图像处理相关知识

总字符数：6320

#### 相似文献列表

去除本人文献复制比：24.4%(1539)

去除引用文献复制比：20%(1261)

文字复制比：24.4%(1539)

1	基于时频结合空间注意力网络的3D骨架人体运动预测 张禄钧;何志权; - 《生物医学工程学进展》 - 2023-12-30	4.8% (303) 是否引证：否
2	基于CycleGAN的低光照图像增强算法设计与实现 李昊洋 - 《大学生论文联合比对库》 - 2023-06-07	4.7% (300) 是否引证：否
3	基于CycleGAN的低光照图像增强算法设计与实现 李昊洋 - 《大学生论文联合比对库》 - 2023-06-14	4.7% (300) 是否引证：否
4	上海汽车制造业在新冠疫情前后的统计研究 金涵洋 - 《大学生论文联合比对库》 - 2023-06-09	3.4% (214) 是否引证：否
5	动态注意力机制引导下的个性化动态血糖预测 刘浩 - 《大学生论文联合比对库》 - 2023-06-12	3.1% (193) 是否引证：否
6	EfficientNet：一种选取网络结构的最优策略 范淮宝 - 《大学生论文联合比对库》 - 2023-05-29	2.3% (145) 是否引证：否
7	1808104029_张志豪_基于LSTM的电缆局部放电检测方法研究 张志豪 - 《大学生论文联合比对库》 - 2022-05-30	1.8% (112) 是否引证：否
8	基于人工智能的胃癌医学图像分割和检查方法的研究 张禹铖 - 《大学生论文联合比对库》 - 2022-05-30	1.8% (111) 是否引证：否
9	2015040042809_张涵_基于机器视觉的输电线路检测 张涵 - 《大学生论文联合比对库》 - 2019-05-20	1.7% (107) 是否引证：否
10	基于CNN-LSTM耦合模型的典型小流域径流预报 刘盼盼 - 《大学生论文联合比对库》 - 2022-05-26	1.5% (95) 是否引证：否
11	基于机器视觉的城市轨道交通的异常人员的识别与特征研究 王毅豪 - 《大学生论文联合比对库》 - 2023-05-29	1.4% (88) 是否引证：否
12	基于CNN的人脸年龄估计的设计与实现 刘文豪 - 《大学生论文联合比对库》 - 2023-06-27	1.4% (86) 是否引证：否
13	基于深度学习的垃圾分类助手 钟一栋 - 《大学生论文联合比对库》 - 2020-04-28	1.3% (83) 是否引证：否
14	邹奕雯-3190431050-基于CNN和Transformer结合的阻塞性睡眠呼吸-计算机科学与技术-计191-石争浩 邹奕雯 - 《大学生论文联合比对库》 - 2023-06-02	1.2% (77) 是否引证：否
15	141_黄瑜锋_标签缺陷检测关键技术研究_黄瑜锋_标签缺陷检测关键技术研究 黄瑜锋 - 《大学生论文联合比对库》 - 2023-03-20	1.2% (75) 是否引证：否
16	RK房地产项目工期延误风险管理研究 贡觉朗卓 - 《大学生论文联合比对库》 - 2023-05-24	1.2% (73) 是否引证：否
17	基于YOLO的目标检测技术研究 李承泽 - 《大学生论文联合比对库》 - 2023-05-27	1.2% (73) 是否引证：否
18	基于前馈神经网络的人脸识别优化方法研究 李浩博 - 《大学生论文联合比对库》 - 2023-05-26	1.1% (71) 是否引证：否
19	基于刑事电子卷宗分析的罪名与刑期预测 胡鹤还 - 《大学生论文联合比对库》 - 2019-05-23	1.1% (69) 是否引证：否
20	基于YOLO9000的人脸识别的实战与分析 张妮 - 《大学生论文联合比对库》 - 2020-06-08	0.9% (56) 是否引证：否
21		0.8% (53)

244-3019209351-刘雪阳		
刘雪阳 - 《大学生论文联合比对库》 - 2023-05-31		是否引证：否
22	基于电商数据的线上手机销量预测系统	0.8% (52)
杨新元 - 《大学生论文联合比对库》 - 2023-06-05		是否引证：否
23	吸毒人脸特征分析及提取算法研究	0.8% (49)
王南 - 《大学生论文联合比对库》 - 2021-05-25		是否引证：否
24	基于深度多分支CNN的阿尔茨海默症早期诊断	0.7% (46)
万果鑫 - 《大学生论文联合比对库》 - 2023-05-26		是否引证：否
25	基于深度学习的通信干扰信号的识别与处理	0.6% (40)
牛彦淇 - 《大学生论文联合比对库》 - 2023-05-27		是否引证：否
26	基于主成分分析算法的静态人脸识别	0.6% (39)
鲁一澄 - 《大学生论文联合比对库》 - 2023-05-17		是否引证：否
27	基于生成式对抗网络的风格迁移平台设计	0.5% (33)
张世英 - 《大学生论文联合比对库》 - 2023-06-05		是否引证：否

原文内容

第二章图像处理相关知识  
第一节图像分类方法

2.1.1 卷积神经网络

卷积神经网络 (Convolutional Neural Networks, CNN) 作为一种优秀的深度学习模型，出于其良好的特征提取能力、参数共享和稀疏连接的优势、适应性强、泛化能力强等特点，在计算机视觉领域以及NLP领域都已经比较成熟且广泛的应用。

2.1.1.1 输入层

卷积神经网络的输入层用于接收数据的输入，由原始的图像或文本转化为可以被卷积神经网络处理的向量或是矩阵（由像素值大小组成）。在这个过程中，还包括了数据预处理、特征提取等多种功能，对后续的模型处理性能都会产生重要的影响。

2.1.1.2 卷积层

这是卷积神经网络非常重要的一个结构，也是对输入矩阵进行特征提取的关键一步。每个卷积层中都包含多个卷积核，通过卷积核在输入层传过来的矩阵上滑动来完成特征提取的工作；卷积核本身就是一个规模小于或等于输入图像矩阵，通过设置步长 (Stride) 和填充 (Padding) 操作来控制卷积核的移动，每移动到新的位置就做一次对应矩阵的乘积和求和，最后得到一个新的矩阵，我们叫做特征图 (Feature Map)，其中值的大小就代表着该局部区域内的特征。第一层的卷积核提取到的信息是有限的，往往只能得到比较低级的特征，但通过多层的迭代，就可以从低级的Feature Map中提取更复杂的特征。

2.1.1.3 激活层

激活层通过激活函数将输入值映射到非线性空间，使神经网络能够学习非线性的关系，对复杂的数据进行更好的建模和拟合，有些激活函数还可以缓解梯度消失的问题，从而使得神经网络的训练更加稳定和高效，具有更好的性能。

2.1.1.4 池化层

但为了充分提取特征，我们不得不接收非常多的特征图，然而得到的大量的特征可能并不是我们想要的，其中很多冗余的特征往往会带来过拟合和维度过高的问题，为了解决这个问题，我们引入了池化层以及下采样的概念。具体来讲，就是在卷积操作过后，对得到的Feature Map在进行特征提取，将其中最具代表性的特征提取出来，这个过程类似于卷积的操作，对划分好的区域中进行下采样从而只得到一个值。常见的池化操作有最大池化 (Max Pooling) 和平均池化 (Average Pooling)，区别就在于：最大池化是选择区域中的最大值作为最具代表性的特征，而平均池化是取所有值的平均值，充分考虑区域中每个位置的值对特征的影响。所以由上可知，池化层通常与卷积层是紧密相连的，对卷积层的结果进行降维，减少网络中的参数个数。

2.1.1.5 全连接层

再通过前面的多层结构提取出图像的特征后，我们需要做的就是如何利用这些特征来完成我们最终的任务，全连接层 (Fully Connected Layers) 的目的就是通过对上层特征进行连接，进行加权后用于将输入分为不同的类。全连接层的每一个节点都和上一层的每个节点全部相连，通过权重矩阵和输入层相乘并加上一个偏置 (bias)，以得到最后的全连接层的输出矩阵。

图二 - 1全连接层示意图

具体的计算公式如下：

(3.1)

其中表示的是全连接层的结果，表示的是全连接层的输入，表示的是权重大小，表示的是偏置量。

2.1.1.6 输出层

输出层是用来输出不同类别的概率分布，具体的概率数值是应用Softmax函数得到的，通过将输出值转换为范围在[0, 1]并且和为1的概率分布。其具体的计算公式如下：

(3.2)

其中是第i个类别的输出值，下面是Softmax的一个示例，从最后的输出结果可知，分类结果是类别1的可能性最大。

图二 - 2Softmax示例



## 2.1.2 图像分类经典框架

### 2.1.2.1 AlexNet

AlexNet[4]是由由 Alex Krizhevsky、Ilya Sutskever 和 Geoffrey Hinton 在 2012 年提出的一个经典的卷积神经网络模型，用于在ImageNet数据集上进行图像分类，是深度学习在计算机视觉领域的重要里程碑之一。它作为新一代的神经网络，下面是它一些比较重要的创新点：

图二 - 3 AlexNet结构示意图

如图所示，AlexNet是一个相对较深的卷积神经网络，包含8个卷积层和3个全连接层，其中包含的大量参数可以保证模型学习到更复杂的图像特征；与此同时，AlexNet使用ReLU作为激活函数，取代传统的Sigmoid函数，具有非线性特性和稀疏激活性的特点，有助于加速模型的收敛速度和避免梯度消失的问题；此外，模型引入Dropout 正则化技术，通过随机地将一部分神经原地输出设置为零，以减少模型地过拟合程度，以提高其泛化能力和性能。最后LRN层的引入，对神经元的输出进行了局部的归一化处理，进一步增强了模型的泛化能力，提高了模型的鲁棒性。但由于AlexNet包含大量的参数，训练过程十分耗时，因此为了加速训练，需要将训练数据分配到多个GPU上进行并行计算，这也在一定程度上增加了训练成本。

### 2.1.2.2 VGG

VGG[5]是Karen Simonyan和Andrew Zisserman在2014年提出的卷积神经网络模型，同样在ImageNet上完成了图像分类任务。相比于之前的网络，他的主要贡献则在于提出了一种简单有效的网络结构设计原则，即使用连续的 $3 \times 3$ 卷积核和池化层，以及全连接层，构建了一个具有较深层次的卷积神经网络。

图二 - 4 VGG结构示意图

在搭建过程中，卷积层都采用 $3 \times 3$ 的卷积核，池化层采用 $2 \times 2$ 的池化核，这样同一的网络结构设计被证明是更加简单有效的；相较于之前的神经网络，VGG包含16或19个卷积层，更深的层数使得VGG可以学习到更加复杂的图像特征；此外，用小卷积核代替较大的卷积核，可以增加网络的深度，并且减少参数的数量，多个小卷积核的组合可以学习到更复杂的特征表示。

### 2.1.2.3 ResNet

ResNet[6]是由Kaiping He等人在2015年提出的一个经典的深度卷积神经网络模型，其主要的创新点在于 引入了残差连接的概念，通过跨层的直接连接来解决梯度消失和梯度爆炸的问题，使得可以训练更深的网络结构。下面是一些ResNet的一些重要特点和创新点：

1. 残差连接：ResNet通过在每个卷积层的输入和输出之间添加一个残差连接来构建残差块（Residual Block）。这个残差连接将输入信号直接添加到了卷积层的输出之中，即  $\text{output} = \text{input} + F(\text{input})$ ，其中  $F(\cdot)$  是残差函数。这种跨层的直接连接可以有效地传递梯度信息，解决了深度网络中梯度消失和梯度爆炸的问题，使得可以训练更深的网络结构。

2. 深度：ResNet是一个非常深的卷积神经网络模型，可以堆叠成几十甚至上百层。这使得ResNet可以学习到非常复杂的特征表示，提高了模型的性能和泛化能力。

3. 批量归一化：ResNet在每个卷积层之后都采用了批量归一化操作，将每个特征通道的输入归一化为均值为0、标准差为1的分布。这有助于加速模型的收敛速度和稳定性，同时还可以减少模型对初始参数的依赖。

4.  $1 \times 1$ 卷积核：ResNet在残差块中 使用了 $1 \times 1$ 的卷积核来降低输入的通道数，然后再使用 $3 \times 3$ 的卷积核进行特征提取。这样可以减少模型的计算复杂度和参数数量，同时还可以增加网络的非线性表达能力。

5. 全局平均池化：ResNet在最后的卷积层之后采用了全局平均池化操作，将每个特征图的尺寸降低到 $1 \times 1$ ，然后通过一个全连接层进行分类。这种操作可以减少模型的参数数量，同时还可以提高模型的泛化能力。

如图所示，X是作为残差块的输入，F(X)是经过第一层线性变化并激活后的输出，在第二层进行线性变化后面的激活层之前，将F(x)与输入值X相加，然后再经过激活函数 relu后输出。在第二层输出值激活前加入X的这条路径称作shortcut connection。

图二 - 5 残差结构图

### 2.1.3 注意力机制（Attention）

深度学习中的注意力机制[7]（Attention Mechanism[8]）是一种模仿人类视觉和认知系统的方法，它允许神经网络在处理输入数据时集中注意力相关的部分。通过引入注意力机制，神经网络能够自动学习并选择性地关注输入中地重要信息，提高模型地性能和泛化能力。

下面可以用一个示例来帮助我们更好地理解注意力机制，展示了人类在看到一幅图时如何高效分配有限注意力资源的，其中未遮挡的区域“锦江饭店”招牌表明视觉系统更加关注的目标，从图中可以看出：人们会把注意力更多地投入到离我们更近、更大，颜色对比更明显的区域。

图二 - 6 人类视觉系统的注意力机制说明

注意力机制本质上讲和人类的选择性注意力机制类似，核心目标也是从众多的信息中选出对当前任务目标更加关键的信息。其中，最典型的注意力机制包括自注意力机制、空间注意力机制和时间注意力机制。这些注意力机制允许模型对输入序列的不同位置分配不同的权重，以便在处理每个序列元素时专注于最相关的部分。

由于我们要研究的是Swin Transformer模型，其中使用的是自注意力机制[9]（Self-Attention Mechanism），所以我们主要对自 注意力机制作一定的介绍。自注意力机制的基本思想是，在处理序列数据时，每个元素都可以与序列中的其他元素建立关联，而不仅仅是依赖于相邻位置的元素。它通过计算元素之间的相对重要性来自适应地捕捉元素之间地长程依赖关系。

具体而言，对于序列中的每个元素，自注意力机制计算其与其他元素之间的相似度，并将这些相似度归一化为注意力权重。然后，通过将每个元素与对应的注意力权重进行加权求和，可以得到自注意力机制的输出。

## 第二节 常见预处理算法

图像预处理[12]在图像处理和计算机视觉任务中扮演着至关重要的角色，无论是传统方法还是深度学习方法，其分类性能



都与输入的图像质量有紧密的关联。因此如何选择合适的预处理算法来提高图像的质量，对于我们的实验来说亚视需要考虑的重要因素，下面我们对于常用的预处理算法进行简单的介绍：

2.2.1 标准化 (Standardization)

将数据集中的特征按照一定的方式进行缩放，使得它们具有零均值和单位方差。具体一般是通过每个特征减去其均值，然后除以其标准差来实现。这有助于加速模型的训练过程，并且可以使不同特征之间得权重更加均匀，提高模型的稳定性和收敛速度。

2.2.2 数据增强 (Data Augmentation)

通过对训练数据进行随即变换，如旋转、翻转、裁剪、缩放、平移、添加噪声等。这些变换可以生成与原始数据相似但又略有不同的数据样本，增加数据的多样性，从而提高模型的泛化能力。

2.2.3 噪声过滤 (Noise Filtering)

对输入数据进行去噪处理，如高斯模糊、中值滤波等，可以减少数据中的噪声对模型的影响，提高模型的鲁棒性和性能表现。

2.2.4 正则化 (Regularization)

正则化是在模型训练过程中引入额外的约束条件，以减少模型的复杂度并防止过拟合。常见的正则化方法包括L1正则化、L2正则化等，这些方法可以惩罚模型中的大权重，从而促使模型更加简单和泛化能力更强。

2.2.5 缺失值处理 (Missing Value Handling)

缺失值处理是指对输入数据中的缺失值进行处理，以确保模型能够正确处理这些缺失值。常见的处理方法包括填充缺失值、删除包含缺失值的样本或特征、使用插值等。这些方法可以确保模型能够充分利用可用的信息，并且能够正确地处理缺失值。

第三节 常见图像特征提取算法

作为计算机视觉领域中至关重要的一步，特征提取方法也经历了多次的更迭，不断为后续任务的推进提供着动力。一个好的特征提取方法，直接影响着特征表示的质量；好的特征表示能够更好地捕捉图像的语义信息和抽象特征，减少输入图像受光照、视角等因素的影响，对这些变化提供一定鲁棒性的特征表示，进而直接影响着分类器的性能；除此之外，好的特征提取方法能将原始数据的高维特征空间有效地降低到一个更低维度的表示空间，减少特征的冗余，从而节省计算资源，提高分类器的泛化能力。

常见的图像特征提取算法[13]分为传统方法和深度学习方法。传统方法通常是通过手工设计特征提取算法来获取数据的有效表示，而基于深度学习的方法可以端到端地学习图像的特征表示，通常能够获得更高质量的特征表示。相较于传统方法，深度学习方法具有更强的泛化性，且能处理更复杂的数据，也不需要很强的专业性来完成参数的设计，显然具有更广泛的应用前景，因此在多数领域深度学习方法都已经逐渐成为主流，相比之下，PCA等传统方法在图像特征提取方面的应用已经相对较少，更多地被用于其他领域或者作为基准方法进行比较。

下面我们就对其中的一些常见方法作一些了解：

2.3.1 主成分分析 (Principal Component Analysis, PCA)

主成分分析[15]作为一种简单高效的降维技术，用于将高维数据集投影到低维空间，保留主要特征并去除噪声。下面是其主要的工作原理：

- (1) 标准化数据：对原始数据进行标准化处理，将每个特征的均值归零，方差归一化，以保证各个特征之间的量纲一致。
- (2) 计算协方差矩阵：计算标准化后的数据集的协方差矩阵，该矩阵描述了数据集中各个特征之间的线性关系。
- (3) 计算特征值和特征向量：对协方差矩阵进行特征值分解，得到特征值和对应的特征向量。特征向量表示新的特征空间的方向，特征值表示在每个方向上的方差大小。
- (4) 选择主成分：根据特征值的大小，选择最重要的特征向量，即主成分，以保留尽可能多的原始数据信息。通常选择方差最大的前k个特征向量作为主成分。
- (5) 投影到新的特征空间：将原始数据投影到由选定的主成分构成的新的特征空间中，得到降维后的数据。

2.3.2 预训练模型的特征提取

预训练模型的特征提取是指利用在大规模数据集上预训练的深度学习模型，将其作为特征提取器来获取输入数据的高层次特征表示。

由于我们自己用来实验的数据集往往不够大，也不希望花费太多的时间和资源进行训练，但又希望尽可能获得性能更好的模型，于是我们引入了预训练模型来学习到更丰富的特征表示。

首先需要选择一个在大规模数据集上预训练的深度学习模型，常见的包括 VGG、ResNet、Inception、EfficientNet 等。通常选择的模型应该具有良好的性能和泛化能力，并且在相似的图像分类任务上已经取得了良好的效果，

然后移除掉原模型中的全连接层，而保留卷积层部分，起到提取特征的作用；

在接收了大量的图像输入后，这些特征表示通常具有较高的语义信息和抽象能力，可以用于各种后续任务，如图像分类、目标检测等。根据实际任务的需要，我们可以选择不同层次的特征表示，一般来说，底层的特征更加局部化和具体化，而顶层的特征更加抽象和语义化。这些通过预训练模型得到的特征表示，既可以作为新任务的输入，也可以作为基础特征，进一步在新的任务上进行微调或训练新的分类器。

与传统的特征提取方法相比，预训练模型的方法具有很强的通用性和灵活性，能够快速、有效地获取输入数据的高层次特征表示，并且在很多视觉任务中都被证明切实能够取得更好的性能，在越来越多的计算机视觉任务中扮演者重要的角色。

去除本人文献复制比：20.3%(1529)		去除引用文献复制比：20.3%(1529)	文字复制比：20.3%(1529)
1	基于机器视觉的常用生活物品识别研究 彭程 - 《大学生论文联合比对库》 - 2022-05-22	10.6% (796)	是否引证：否
2	基于机器视觉的常用生活物品识别研究 彭程 - 《大学生论文联合比对库》 - 2022-05-27	10.6% (796)	是否引证：否
3	基于深度学习的人脸活体检测算法实现 杜珑衍 - 《大学生论文联合比对库》 - 2023-05-29	7.4% (559)	是否引证：否
4	基于Transformer的指节纹识别系统设计与实现 尚胜 - 《大学生论文联合比对库》 - 2023-06-07	3.9% (297)	是否引证：否
5	基于深度学习的盾构机渣片图像分割研究 with 实现 王宇杰 - 《大学生论文联合比对库》 - 2022-05-19	3.4% (256)	是否引证：否
6	基于YOLO算法的车辆识别 胡钰欢 - 《大学生论文联合比对库》 - 2023-05-26	2.8% (211)	是否引证：否
7	193897_郑森潇_电子信息工程(人工智能)_Transformer框架下图像分类算法的设计_论文定稿_1684832374874 郑森潇 - 《大学生论文联合比对库》 - 2023-05-24	2.5% (185)	是否引证：否
8	基于移位窗口自注意力机制的新生儿脑部多功能区域分割研究 王涛(导师：张永刚) - 《吉林大学硕士学位论文》 - 2023-03-01	2.2% (169)	是否引证：否
9	放射治疗肿瘤和器官的自动勾画研究 赵静宇(导师：常世杰) - 《中国医科大学硕士学位论文》 - 2023-03-01	2.1% (161)	是否引证：否
10	基于Swin Transformer的交通场景目标检测 李文轩 - 《大学生论文联合比对库》 - 2023-05-31	1.6% (118)	是否引证：否
11	基于Transformer模型的不确切目标分割方法设计与实现 王新元 - 《大学生论文联合比对库》 - 2022-05-28	1.4% (104)	是否引证：否
12	基于深度学习与信息融合技术的陶瓷辊道窑温度智能检测方法 朱永红;戴晨雨;李蔓华; - 《陶瓷学报》 - 2024-01-31 12:38	1.0% (72)	是否引证：否
13	基于视觉Transformer结构的医学图像配准算法研究 朱子衿 - 《大学生论文联合比对库》 - 2023-05-27	0.5% (34)	是否引证：否
14	基于视觉Transformer结构的医学图像配准算法研究 朱子衿 - 《大学生论文联合比对库》 - 2023-05-29	0.5% (34)	是否引证：否

原文内容

第三章 Swin Transformer模型的基本框架

通过之前的分析我们可以得知，因为需要针对大尺寸图片进行分类任务，所以我们选择Swin Transformer模型完成后续实验是比较理想可行的，下面我们就来详细介绍一下模型本身是由哪些模块构成的。

图三 - 1Swin Transformer网络结构图

第一节 Swin Transformer模型基本结构

在原文的模型中，首先将大小为 $H \times W \times 3$ 的图片输入到Patch Partition模块中进行分块，即每 $4 \times 4$ 相邻的像素为一个Patch，然后在channel方向进行flatten，若输入为RGB三通道图片，那么每个Patch就有 $4 \times 4 = 16$ 个像素，所以得到展平后的值为 $16 \times 3 = 48$ ，所以通过Patch Partition后图像的shape由 $[H, W, 3]$ 变为 $[H/4, W/4, 48]$ 。然后再通过Linear Embedding层对每个像素的channel数据做线性变换，由48变为C，即图像shape再由 $[H/4, W/4, 48]$ 变成了 $[H/4, W/4, C]$ 。

然后通过接下来的4个Stage构建不同大小的feature map，除了Stage1中先通过一个Linear Embedding层外，剩下的3个剩下三个stage都是先通过一个Patch Merging层进行下采样。最后就是重复堆叠两种不同结构的Swin Transformer Block，如图(b)中所示，这两种结构的不同之处仅在于一个使用了W-MSA结构，一个使用了SW-MSA结构。而且这两种结构是成对使用的，先使用一个W-MSA结构再使用一个SW-MSA结构。

最后对于分类网络，后面接上一个Layer Norm层、全局池化层以及全连接层得到最终输出，这些都是在源码中的实现方法。

下面我们对这个模型中的关键模块进行详细的研究：

1. Patch Merging：该模块的作用是在每个Stage开始前做下采样，用于缩小分辨率，调整通道数，进而形成层次化的设计，同时节省一定的运算量。由于每次下采样两倍，在行和列的方向上间隔2选取元素，然后拼接在一起作为整个向量：

—  $x = x.view(B, H, W, C)$ ：将输入特征向量重塑为四维向量，形状为 $[B, H, W, C]$ ，其中B是批大小，H和W是输入特征的高度和宽度，C是通道数。

二  $x_0 = x[:, 0::2, 0::2, :]$ 、 $x_1 = x[:, 1::2, 0::2, :]$ 、 $x_2 = x[:, 0::2, 1::2, :]$ 、 $x_3 = x[:, 1::2, 1::2, :]$ ：分别从输入特征向量中提取每个  $2 \times 2$  区域的4个像素位置的特征，形状均为  $H/2 \times W/2 \times C$ 。

三  $x = \text{torch.cat}([x_0, x_1, x_2, x_3], -1)$ ：将四个特征向量沿着通道维度进行拼接，通道维度方向拉长，得到形状为  $H/2 \times W/2 \times 4C$  的向量，相当于用空间上的维度换了更多的通道数。

四  $x = x.\text{view}(B, -1, 4 \times C)$ ：将拼接后的向量展平为二维向量，形状为  $[B, H/2 \times W/2, 4 \times C]$ ，这就是我们合并后最终得到的特征向量。

五通过上述操作，就把原来一个大的张量变小了，这就类似于卷积神经网络中的池化操作，所以为了和卷积神经网络保持一致，在池化降维后通道数仅仅是翻倍而不是变成4倍，所以在C的维度上进行一次卷积操作，把通道数降下来变成2C，通过这个操作把原来  $H \times W \times C$  的张量变成  $H/2 \times W/2 \times 2C$  的一个张量，空间大小减半，通道数乘2，这样就跟卷积神经网络完全对等起来了；综上，每经过一个Stage，下采样率两倍，H和W都会减半，通道数翻倍。

2. Window Attention和Shifted Window Attention：传统的 Transformer 都是基于全局来计算注意力的，因此计算复杂度十分高。而 Swin Transformer 则将注意力的计算限制在每个窗口内，进而减少了计算量。基于窗口计算注意力的方式（W-MSA (Windows Multi-head Self-Attention)）虽然可以很好地解决内存和计算量地问题，但是窗口和窗口之间没有通信，这样就达不到全局建模的效果，也会限制模型的能力，所以文章中提出了一种移动窗口（Shifted Window）的方式，也就是把原来的窗口往右下角移动一半窗口的距离，如果Transformer是上下两层连着做这种操作，先是window再是shifted window的话，就能起到窗口和窗口之间互相通信的目的。

3. Transformer Block：每次输入进来先做一次Layernorm和基于窗口的多头自注意力，然后再做一次Layernorm和MLP，第一个block就结束了；接下来是基于移动窗口的多头自注意力，最后同样通过MLP得到输出。两个block结合起来才是Swin Transformer的一个基本的计算单元，因此也就解释了为什么Stage1, 2, 3, 4中的block是depths=(2, 2, 6, 2)这样安排的，即无论有多少层block，其中的block的数字都是偶数，因为他始终需要两层的block连在一起作为一个基本单元，所以一定是2的倍数。

图三-2 Swin Transformer Block网络结构图

## 第二节 Swin Transformer重要原理

### 3.2.1 SW-MSA与mask的实现

首先在原论文中，为了解释窗口机制的引入，我们需要对W-MSA和MSA的计算复杂度在维度上进行一个估计：

(3.1) (3.2)

其中，公式 (3.1) 对应的是标准的多头自注意力的计算复杂度，每一个图片有  $h \times w$  个patch，C是特征的维度；公式 (3.2) 对应的是基于窗口的自注意力计算的复杂度，注意每个窗口并不是最小的计算单元，这里的M就是一个窗口的某条边上有多少个patch，所以整个窗口中就有  $M \times M$  个patch。综上，前者的复杂度和patch的数量成平方关系，后者在M固定时则与patch成线性关系，对于较大的h和w，基于全局的自注意力计算通常难以承受，而基于窗口的自注意力所需的计算就会减少很多。

如图三-3所示，如果仅仅是使用左图所示的窗口划分方法，那就仅仅计算到了块内的自注意力，而忽略了他相邻的其他块中的信息，为了解决这个问题，Swin Transformer团队引入了Shifted Windows Multi-Head Self-Attention (SW-MSA) 模块，也就是加上偏移的W-MSA。

图三-3shifted window中计算self-attention的示意图

具体操作就是先将窗口从左上角向右下各偏移个像素得到右侧的Layer1+1层，通过这样的操作，就使得多个窗口间的信息得到了交流，其中的C窗口沟通了原来上面两个窗口的信息，而中间的  $4 \times 4$  的窗口沟通了4个窗口的信息。但进行了窗口的偏移后由原来的4个窗口变成了9个窗口，窗口数量的变多，显然会增加MSA的计算量，所以如图三-4是作者提出的一种叫做Efficient batch computation for shifted configuration更高效的计算方法：

图三-4在shifted window中计算self-attention的高效批处理方法

如图三-5，将A和C两个窗口翻转到最下面，然后将A和B翻转到最右边，得到新的窗口并进行新的划分，移动完后，4是一个单独的窗口；将5和3合并成一个窗口；7和1合并成一个窗口；8、6、2和0合并成一个窗口。这样又和原来一样是4个  $4 \times 4$  的窗口了，所以能够保证计算量是一样的。

图三-5cyclic shift

但是这样子，就导致原本并不相邻的图片连在了一起，直接在窗口内计算自注意力，就会导致并不相关的信息反而产生了关系，并不是我们想要的结果。所以我们使用的是带蒙版(mask)的MSA，这样就能通过设置蒙版来隔绝不同区域的信息。具体来讲，为了保证shifted window self-attention计算的正确性，只能计算相同子窗口的self-attention，不同子窗口的self-attention结果要归0，在进行cyclic shift之前，需要给子窗口进行编码，如图表4，编码之后通过torch.roll对窗口进行滚动，在(5, 3) (7, 1) (8, 6, 2, 0)组成的新窗口中，只有相同编码的部分才能计算self-attention，不同编码位置间计算self-attention需要归0，根据self-attention公式，最后需要进行Softmax操作，不同编码位置间计算的self-attention结果通过mask加上-100，在Softmax计算过程中，Softmax(-100)无限趋近于0，达到归0的效果，所以对于像素0而言，实际上还是只和区域5内的像素进行MSA，在计算完成后还要把数据挪回到原来的位置。

通过设置合理的mask，让Shifted Window Attention在与Window Attention相同的窗口个数下，达到等价的计算结果。在我们实现的过程中，Mask是初始就给定的，即只有特征发生滚动去迎合Mask。

```
def create_mask(self, x, H, W):
    # calculate attention mask for SW-MSA
    # 1. 先计算输入图像的高度Hp和宽度Wp，使其能够被window_size整除，然后创建一 个形状为(1, Hp, Wp, 1)的全零张量
    img_mask, 用于存储图像的掩码信息。
    Hp = int(np.ceil(H / self.window_size)) * self.window_size
    Wp = int(np.ceil(W / self.window_size)) * self.window_size
    # 拥有和feature map一样的通道排列顺序，方便后续window_partition
```



```

img_mask = torch.zeros((1, Hp, Wp, 1), device=x.device) # [1, Hp, Wp, 1]
# 2. 定义两个切片对象h_slices和w_slices, 分别表示高度和宽度上的切片范围。这些切片范围将被用于将图像掩码赋值给img_mask张量的不同部分。
h_slices = (slice(0, -self.window_size),
            slice(-self.window_size, -self.shift_size),
            slice(-self.shift_size, None))
w_slices = (slice(0, -self.window_size),
            slice(-self.window_size, -self.shift_size),
            slice(-self.shift_size, None))
# 3. 使用两个嵌套的循环遍历h_slices和w_slices中的切片范围, 并未每个切片位置赋予一个递增的值cnt, 这样就可以将相应的索引值赋给img_mask张量, 以表示每个窗口的位置。
cnt = 0
for h in h_slices:
    for w in w_slices:
        img_mask[:, h, w, :] = cnt
        cnt += 1
# 4. 使用window_partition函数将img_mask张量划分为大小为window_size的窗口, 得到形状为(nW, window_size, window_size, 1)的张量mask_windows。将mask_windows重塑为形状为(nW, window_size*window_size)的二维张量。
mask_windows = window_partition(img_mask, self.window_size)
mask_windows = mask_windows.view(-1, self.window_size * self.window_size)
# 5. 通过计算mask_windows与自身转置之间的差异来计算注意力掩码(attn_mask), 将attn_mask中不为零的位置用-100.0来替换, 将为零的位置用0.0替换。这样可以为注意力机制建立一个掩码, 以自注意力计算的过程中过滤不相关的位置。
attn_mask = mask_windows.unsqueeze(1) - mask_windows.unsqueeze(2)
attn_mask = attn_mask.masked_fill(attn_mask != 0, float(-100.0)).masked_fill(attn_mask == 0, float(0.0))

```

### 3.2.2 相对位置编码

经Swin Transformer团队实验, 发现Swin-T网络在Attention计算中引入相对位置偏置机制, 其准确率能够提高1.2%~2.3%不等。

下面我们通过源码来具体了解其实现原理:

1. 定义一个相对位置偏置参数表:

```

self.relative_position_bias_table = nn.Parameter(
    torch.zeros((2 * window_size[0] - 1) * (2 * window_size[1] - 1), num_heads))
# [2*Mh-1 * 2*Mw-1, nH]

```

2. 定义不参与网络学习的变量的相对位置索引, 首先使用torch.arange函数创建两个张量coords\_h和coords\_w, 分别表示窗口大小的高度和宽度范围内的坐标:

```

coords_h = torch.arange(self.window_size[0])
coords_w = torch.arange(self.window_size[1])

```

然后, 利用torch.meshgrid函数将coords\_h和coords\_w组合起来, 生成一个二维坐标矩阵coords, 其中包含了所有可能的坐标组合。coords的shape是(2, Wh, Ww), 其中Wh和Ww分别代表窗口的高度和宽度的大小:

```

coords = torch.stack(torch.meshgrid([coords_h, coords_w]))

```

3. 接下来, 通过 torch.flatten 函数将 coords 展平为一维张量 coords\_flatten, shape 为 (2, Wh\*Ww), 表示所有坐标点的组合:

```

coords_flatten = torch.flatten(coords, 1)

```

计算相对坐标 relative\_coords, 这里通过广播操作实现了两两坐标之间的相对位置计算。计算结果的 shape 是 (2, Wh\*Ww, Wh\*Ww), 因为对每个坐标点, 需要计算它与所有其他坐标点的相对位置:

```

relative_coords = coords_flatten[:, :, None] - coords_flatten[:, None, :]
relative_coords = relative_coords.permute(1, 2, 0).contiguous()

```

根据 Swin Transformer 中相对位置索引的计算规则, 通过对相对坐标进行数据变换和求和操作, 得到最终的相对位置索引矩阵 relative\_position\_index, shape 为 (Wh\*Ww, Wh\*Ww)。

```

relative_coords[:, :, 0] *= 2 * self.window_size[1] - 1
relative_position_index = relative_coords.sum(-1)

```

4. 最后, 通过 self.register\_buffer 将计算得到的相对位置索引矩阵 relative\_position\_index 注册为网络的缓冲区(buffer), 以便在网络的前向传播过程中进行使用。

```

self.register_buffer("relative_position_index", relative_position_index)

```

5. 在前向计算中, 相对位置偏置(relative\_position\_bias)被计算为相对位置索引矩阵(relative\_position\_index)对应的预先计算好的相对位置偏置表格(relative\_position\_bias\_table)中的值:

首先, 利用 relative\_position\_index.view(-1) 将相对位置索引矩阵展平为一维张量, 并使用它来索引相对位置偏置表格 relative\_position\_bias\_table 中的值。这样得到的 relative\_position\_bias 是一个形状为 (window\_size\*window\_size, window\_size\*window\_size, nhead) 的三维张量。

接着, 通过 .permute(2, 0, 1) 操作对 relative\_position\_bias 进行维度交换, 将nhead维度移动到最前面, 得到新的

```
相对位置偏置张量。  
relative_position_bias=self.relative_position_bias_table[self.relative_position_index.view(-  
1)].view(self.window_size[0] * self.window_size[1], self.window_size[0] * self.window_size[1], -1)  
relative_position_bias = relative_position_bias.permute(2, 0, 1).contiguous()
```

5. 第四章基于SwinTransformer模型的文档图像分类实验

总字符数：4342

相似文献列表

去除本人文献复制比：3.1%(134)		去除引用文献复制比：3.1%(134)	文字复制比：3.1%(134)
1	基于CLDNN的音乐风格自动标识方法与系统 李锐 - 《大学生论文联合比对库》- 2023-06-05	1.6% (69)	是否引证：否
2	基于深度学习的铁路钢轨扣件缺陷无人机图像智能分析技术研究 崔京 - 《大学生论文联合比对库》- 2020-05-15	0.9% (41)	是否引证：否
3	基于视觉和语言变换器的图像分类方法 宋鑫 - 《大学生论文联合比对库》- 2022-05-25	0.9% (41)	是否引证：否
4	轻量化的基于CNN的人脸识别算法 李睿捷 - 《大学生论文联合比对库》- 2023-05-28	0.6% (24)	是否引证：否

原文内容

第四章基于Swin Transformer模型的文档图像分类实验  
第一节实验环境  
本文的算法主要是依托于深度学习主流开源框架PyTorch，Python版本是Python3.11.8，初始学习率设置为0.001，epoch设置为100，学习率调整算法使用的是余弦退火学习率调整器。

实验环境以及参数设置			
配置	型号	参数设置	内容
操作系统	80	迭代次数 (epoch)	100
CPU	Intel(R) Xeon(R) CPU E5-2673 v3	批次值 (batch_size)	16
GPU	4090-24G	学习率 (learning rate)	0.001
Python版本	3.11.8	损失函数	CrossEntropyLoss
CUDA版本	12.1	类别数目 (num_classes)	10

实验环境以及参数设置  
配置型号参数设置内容  
操作系统 80 迭代次数 (epoch) 100  
CPU Intel(R) Xeon(R) CPU E5-2673 v3 批次值 (batch\_size) 16  
GPU 4090-24G 学习率 (learning rate) 0.001  
Python版本 3.11.8 损失函数 CrossEntropyLoss  
CUDA版本 12.1 类别数目 (num\_classes) 10  
第二节数据集来源和划分  
本次实验的数据集来源是近代的300本书籍，排除了一些数量过少的类别，最后得到10个分类，图像尺寸基本分布在1300×2000像素左右，相较于传统的公共数据集，这样的尺寸显然是更大的，也就意味着有更多的特征需要提取，更多的细节需要把握，对模型的能力提出了更高的要求。

下面是数据集的划分：

页面类别 (10个)	训练集train/张	测试集test/张	数据集总数/张
BodyPage_Annotation	410	121	531
BodyPage_Common	865	214	1079
BodyPage_ImageTitle	113	23	136
BodyPage_PageFooter	308	71	379
BodyPage_PageHeader	409	103	512
BodyPage_Table	802	191	993
BodyPage_TableTitle	244	45	289
CatalogPage	305	63	368
CopyrightPage	136	38	174
CoverPage	216	41	257

页面类别 (10个) 训练集train/张测试集test/张数据集总数/张  
BodyPage\_Annotation 410 121 531  
BodyPage\_Common 865 214 1079  
BodyPage\_ImageTitle 113 23 136

BodyPage\_PageFooter 308 71 379  
BodyPage\_PageHeader 409 103 512  
BodyPage\_Table 802 191 993  
BodyPage\_TableTitle 244 45 289  
CatalogPage 305 63 368  
CopyrightPage 136 38 174  
CoverPage 216 41 257

### 第三节数据预处理操作

1. `transforms.Resize(448)`: 将图像的大小调整为 $448 \times 448$ 像素, 这样做可以统一输入图像的尺寸, 便于后续的处理。
2. `transforms.CenterCrop(384)`: 对图像进行中心裁剪, 裁剪出 $384 \times 384$ 像素的部分。这样做可以去除图像边缘的噪声, 集中在图像的主要内容上。
3. `transforms.RandomHorizontalFlip()`: 随机水平翻转图像。这样做可以增加数据的多样性, 提高模型的泛化能力。
4. `transforms.ToTensor()`: 将图像转换为向量格式, 由于深度网络模型的输入往往是向量格式的数据, 所以只有转化为向量才能用于神经网络的处理。
5. `transforms.Normalize([0.5, 0.5, 0.5], [0.5, 0.5, 0.5])`: 对图像进行标准化处理, 将图像的像素值缩放到 $[-1, 1]$ 的范围内。这样做可以加速模型的收敛过程。

### 第四节对比试验

作为科学实验的重要方法, 对比试验一直以来都能给我们带来丰富的实验结果, 尤其是针对比较复杂的深度学习模型, 大量的参数和层数的堆叠, 让我们对它的研究变得更加困难, 所以我们需要借助对比试验的方法, 来研究一些超参数以及其他可能影响模型预测结果的因素是否真的如我们所想。但同时要注意的是, 这样的实验结果也只能在一定程度上证明这些因素是有关联的, 但可能也仅限于当前的实验环境和数据集, 不是绝对的因果关系。下面是我进行的一些对比试验, 希望可以对Swin Transformer模型的理解有一些帮助。

#### 4.4.1 数据集均衡对比

前文讲过, 原始数据集是具有显著特征的, 区别于传统的数据集, 在没有进行实验之前, 不免会有一些想法, 大量的正文页比例或许会影响模型的准确率, 所以我们进行了一组对比试验来验证我们的想法, (从300本近代书籍中提取出来的数目最多的BodyPage\_PageHeader有36082张, 为了加快实验的速度且不影响实验的最终结果, 我们对数据集进行了两次削减, 第一次实验我们选择了大约一万张图片, 第二次和其余类别的数据作进一步的均衡, 留下了512张图片作为最终的选择) 下面是我们具体的实验结果

图三:6 不均衡数据集和均衡数据集的loss\_acc\_lr

由上述的实验结果可知, 当数据集分布不均衡时, 某个数量居巨大的类别将会很大程度上影响我们的分类结果, 如左图所示, 仅需50个epoch, 准去率就可以迅速收敛至89.80%左右, 这看起来是非常优秀的结果; 但实际上这样的情况并不能很好地反映出模型的真实性能, 因为无论实在训练还是测试集中, 那个不均衡的数据类别都会很容易得到正确的答案, 这就好比我们在完成选择题时, 遇到不会的都选择了相同的答案, 但恰好蒙对, 这样的话最终成绩显然是不能反映我们的真实实力的。而右边的结果就是我们在削减特殊类别的图片数量后进行的二次实验, 可以看到训练所需的epoch变多了, 预测的准确率也出现了明显的下滑, 但这才是模型的真实能力, 明显更加客观。因此, 在这样的结果上对模型进行优化才是有意义的, 我们希望的是模型具有很好的泛化能力, 而不是一味追求在特定数据集上的高准确率。

#### 4.4.2 动态学习率和固定学习率对比

学习率[16] (Learning Rate) 是神经网络中一个重要的超参数, 控制着模型在每次参数调整时的步长大小, 学习率设置的大小都会影响模型的训练过程和最终的性能。

如果学习率设置的过大, 可能导致模型的更新过大, 在参数空间中来回跳跃, 出现震荡或不稳定的情况, 更严重可能发生梯度爆炸 (Gradient Exploding) 和梯度消失 (Gradient Vanishing) 的问题, 导致模型无法收敛到最优解。相反, 如果学习率设置的过小, 模型的参数更新步长过小, 可能使模型的收敛速度缓慢, 不光会浪费更多的训练时间, 还有可能导致模型陷入局部最优解或鞍点, 更容易受到局部极小值的影响, 同样会难以找到全局的最优解。

综上, 合理的学习率调整策略可以帮助模型更快地收敛到最优解, 节省训练资源, 优化模型性能; 于是我们进行了一组对比试验, 以研究动态调整学习率和固定学习率对于模型训练过程的影响, 本次实验中选取的学习率调整算法是余弦退火学习率调整器, 我们可以通过下面的函数表达式来简单了解其原理, 其中的 $x$ 是当前的训练周期数,  $T$ 是总训练周期数,  $\text{args.lrf}$ 是一个超参数, 表示学习率的最低比例, 即余弦函数的最小值,  $\text{lrf}(x)$ 是当前训练周期下的学习率比例:

(4.1)

下面是动态学习率调整的训练结果:

图四 - 1 固定学习率和动态学习率loss\_acc\_lr曲线图

通过上述的loss\_acc\_lr曲线图可以看到, 固定学习率使得模型在100个epoch内无法收敛至一个稳定的准确率, 而动态调整学习率可以收敛至大约85.18%的转却屡, 因此动态调整学习率的方式对于模型的训练显然是积极作用的。

#### 4.4.3 不同预训练模型对比

Swin Transformer模型提出后, 团队人员在各种公开的大规模数据集上都进行了相应的实验, 并在github上传了训练过的一些模型来作为预训练模型, 作为迁移学习[17]的基础用来优化后续模型的使用。

预训练模型通过在大规模的数据集上进行训练, 学习到了丰富的数据特征, 可以作为新任务的良好参数初始化, 这样可以使得模型更容易地收敛到最优解, 减少了在小规模数据集上训练时的过拟合风险; 除此之外, 还可以起到迁移学习的效果, 可以将在大规模数据集上学习到的知识和特征迁移到目标任务中, 通过在少量标注数据上进行微调, 从而充分利用到大规模数据集上收集到的信息, 使得模型具有更强的泛化能力, 可以适用于各种领域和任务, 节省大量的训练时间和计算资源, 提高模型的训练效率。

我们本次实验使用的是在ImageNet-22k上进行图像分类的有监督的Swin Transformer预训练模型, ImageNet作为一个计算



机视觉系统识别项目，是目前世界上最大的图像识别数据库。ImageNet中目前共有14197122幅图像，总共分为21841个类别，作为一个超大规模的优质数据库，它已经成为用于训练和评估模型性能的基准。最终我们选用的分别是swin\_base\_patch4\_window7\_224\_22k和swin\_base\_patch4\_window12\_384\_22k这两个模型，从名字上不难看出两个模型的Patch都是4，window大小为7和12，输入图像大小分别是224\*224像素和384\*384像素。

Swin Transformer团队提供了四个不同大小的模型以供使用，下面是不同模型的参数C和Swin Transformer block的层数信息：

```
Swin-T:C=96, layerSwin-S:C=96, layer}
Swin-B:C=128, layer numbers={2, 2, 18, 2}
Swin-T:C=192, layer numbers={2, 2, 18, 2}
```

下面是我实验的结果：

图四 - 2swin\_base\_patch4\_window7\_224\_22k

由实验结果可知，保留更大尺寸的输入图像，选择更大的window\_size对于本次大尺寸的图像数据集的实验来说，显然具有更好的训练效果，当预训练模型为swin\_base\_patch4\_window12\_384\_22k时，实验准确率最终收敛于85.16%，当预训练模型为swin\_base\_patch4\_window7\_224\_22k时，模型最终收敛至84.18%，二者相差了1%，在深度学习的方法对比中，这已经足够说明问题。事实证明在本次的实验中，当loss都下降至稳定情况下，更大尺寸的输入尺寸的保留和窗口大小能够提取到更丰富的特征，也就能够在测试集中有更好的表现，具有更好的分类性能，这在以后的应用中将是重要的经验。

图四 - 3swin\_base\_patch4\_window12\_384\_22k

第五节本章小结

本章对于Swin Transformer模型进行了简单的介绍，并对其中比较重要的部分做更详细的研究学习，作为如今图像分类任务的骨干网络，这样的研究是必要的。然后通过对数据集均衡、学习率调整策略和预训练模型三个对比试验，来研究不同因素对实验模型的准确率影响，通过上述实验，我们对该模型以及深度学习都有了更充分的理解，对于以后的多场景下的应用都有重要的意义。

6. 第五章总结与展望	总字符数：1050
相似文献列表	
去除本人文献复制比：0%(0)      去除引用文献复制比：0%(0)      文字复制比：0%(0)	
原文内容	

第五章总结与展望

第一节主要工作

本文的选题为文档图像分类方法研究与分析，这也是当下比较热门的图像分类深度学习模型的一个具体应用。本文首先介绍了图像处理的相关知识，涉及到图像分类方法、图像预处理算法、常见的图像特征提取算法和深度学习的基础知识；之后对于我们本次实验的骨干网络Swin Transformer的基本框架以及其中的关键模块作一定的研究；最后就是对比实验的具体内容，通过三组的对比试验，结合控制变量法，来研究学习率、数据集和预训练模型对于模型预测效果的影响程度，以此进一步了解Swin Transformer在实际图像分类任务中应该注意的地方和可以调整的参数，从而使它更好地服务于我们的生活。

第二节展望

在结束了上述实验后，我已经对于图像分类任务以及Swin Transformer模型有了更深的理解，原论文中提出的移动窗口机制和局部位置编码让后来的研究者们受益匪浅，也有了也大致明白学习率的调整、数据集的分布和预训练模型的选取究竟是如何影响模型的性能。

但是针对Swin Transformer模型和其再实际生活中的应用，还有很多可以改进和优化的地方：

模型优化：针对 Swin Transformer 模型的局限性和不足之处，可以进一步进行模型优化和改进。例如，设计更加高效的注意力机制、更有效的参数初始化方法、更精细的模型剪枝策略等，以提高模型的性能和泛化能力。

多任务学习：将 Swin Transformer 应用于多任务学习场景，探索其在图像分类、目标检测、图像分割等不同任务中的应用。可以设计多任务联合训练的网络结构，充分利用模型的表示能力和泛化能力，提高多任务学习的效果。

跨领域应用：将 Swin Transformer 应用于其他领域和应用场景，如自然语言处理、语音识别、医疗影像分析等。可以根据不同领域的特点和需求，设计适用于 Swin Transformer 的新型任务和模型结构，拓展其应用范围。

模型解释性：提高 Swin Transformer 模型的解释性和可解释性，探索其内部机制和决策过程，从而更好地理解模型的工作原理和预测结果。可以设计可视化方法、注意力热力图等技术，帮助解释模型的预测结果和推理过程。

硬件加速：结合硬件加速技术，如 GPU、TPU 等，进一步优化 Swin Transformer 模型的训练和推理速度，提高模型的效率和实用性。可以探索模型在不同硬件平台上的部署和优化方法，实现更快速、更高效的模型运行。

7. 第六章参考文献	总字符数：1953
相似文献列表	
去除本人文献复制比：54.8%(1070)      去除引用文献复制比：46.9%(916)      文字复制比：54.8%(1070)	

1	基于双通道Transformer的地铁站台异物检测	27.4% (536)
	刘瑞康;刘伟铭;段梦飞;谢玮;戴愿; - 《计算机工程》 - 2023-07-31 12:51	是否引证: 否
2	10716180_佚名_基于深度图像特征的蛋白质亚细胞定位方法研究	24.3% (475)
	佚名 - 《大学生论文联合比对库》 - 2021-05-10	是否引证: 否
3	Targeted Fool: 一种实现有目标攻击的算法	24.2% (473)
	张华;高浩然;杨兴国;李文敏;高飞;温巧燕; - 《西安电子科技大学学报》 - 2020-12-08 1	是否引证: 否
4	可学习深度位置编码引导的车前图像道路可行驶区域检测	17.8% (347)
	韩汀;陈思宇;马津;蔡国榕;张吴明;陈一平; - 《武汉大学学报(信息科学版)》 - 2023-10-25 17:07	是否引证: 否
5	151003500226-白智远-临床影像学中的皮肤病变智能分类系统	17.2% (336)
	白智远 - 《大学生论文联合比对库》 - 2019-05-20	是否引证: 否
6	跨模态交互融合与全局感知的RGB-D显著性目标检测	14.6% (285)
	孙福明;胡锡航;武景宇;孙静;王法胜; - 《软件学报》 - 2023-06-15 12:15	是否引证: 否
7	基于机器学习的图像视频质量评价	9.1% (178)
	衡稳 - 《大学生论文联合比对库》 - 2015-05-28	是否引证: 否
8	基于深度神经网络的抽象画生成技术研究是实现	6.0% (118)
	陆志成 - 《大学生论文联合比对库》 - 2018-05-29	是否引证: 否
9	基于深度学习的医学图像分割	5.6% (110)
	王娟丽(导师: 张静) - 《兰州大学硕士论文》 - 2022-03-01	是否引证: 否
10	Prediction of pandemic risk for animal-origin coronavirus using a deep learning method	3.6% (70)
	Zheng Kou;Yi-Fan Huang;Ao Shen;Saeed Kosari;Xiang-Rong Liu;Xiao-Li Qiang; - 《贫困所致传染病(英文)》 - 2021-10-10 00:00	是否引证: 否
11	Faster R-CNN深度学习算法识别胸部X片图像的应用研究	2.2% (43)
	谢晨 - 《大学生论文联合比对库》 - 2019-05-21	是否引证: 否

原文内容

第六章参考文献

[1] 田永林,王雨桐,王建功,等. 视觉Transformer研究的关键问题:现状及展望[J]. 自动化学报, 2022, 48(04):957-979. DOI:10.16383/j. aas. c220027.

[2] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用, 2019, 55(12):20-36.

[3] Liu, Ze et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 9992-10002.

[4] Dosovitskiy, Alexey et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." ArXiv abs/2010.11929 (2020): n. pag.

[5] J. Deng, Li Fei-Fei et al. "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR. 2009. 5206848.

[6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[7] Simonyan, Karen and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." CoRR abs/1409.1556 (2014): n. pag.

[8] He, Kaiming et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 770-778.

[9] 邱锡鹏, 神经网络与深度学习, 机械工业出版社, <https://nndl.github.io/>, 2020.

[10] Bahdanau, Dzmitry et al. "Neural Machine Translation by Jointly Learning to Align and Translate." CoRR abs/1409.0473 (2014): n. pag.

[11] 朱张莉,饶元,吴渊,等. 注意力机制在深度学习中的研究进展[J]. 中文信息学报, 2019, 33(06):1-11.

[12] 周妮娜,王敏,黄心汉,等. 车牌字符识别的预处理算法[J]. 计算机工程与应用, 2003(15):220-221+232.

[13] 常昌. 图像特征提取方法研究及应用[D]. 华中科技大学, 2011.

[14] 陈成,耿晓中,刘柏进,等. 一种改进的Swin Transformer图像分类识别方法[J]. 软件工程, 2024, 27(01):19-22. DOI:10.19644/j. cnki. issn2096-1472. 2024. 001. 005.

[15] 赵蕾. 主成分分析方法综述[J]. 软件工程, 2016, 19(06):1-3.

[16] 蒋文斌, 彭晶, 叶阁焰. 深度学习自适应学习率算法研究[J]. 华中科技大学学报(自然科学版), 2019, 47(05):79-83. DOI:10.13245/j.hust.190515.

[17] 石祥滨, 房雪键, 张德园, 等. 基于深度学习混合模型迁移学习的图像分类[J]. 系统仿真学报, 2016, 28(01):167-173+182. DOI:10.16182/j.cnki.joss.2016.01.023.

8. 第七章致谢		总字符数: 386
相似文献列表		
去除本人文献复制比: 62.7%(242)		去除引用文献复制比: 62.7%(242) 文字复制比: 62.7%(242)
1	1110595-崔斌-信安 崔斌 - 《大学生论文联合比对库》- 2015-05-06	62.7% (242) 是否引证: 否
2	SQL与NOSQL混合型数据库架构设计方法研究 - 《大学生论文联合比对库》- 2016-05-04	62.2% (240) 是否引证: 否
3	5_杨帆_基于树莓派图像识别及远程监控系统的设计与实现_1312664_杨帆_基于树莓派图像识别及远程监控系统的设计与实现 基于树莓派图像识别及远程监控系统的设计与实现 - 《大学生论文联合比对库》- 2017-05-02	61.7% (238) 是否引证: 否
4	1110597-金浩-法学 金浩 - 《大学生论文联合比对库》- 2015-05-06	52.6% (203) 是否引证: 否
原文内容		

第七章致谢

值此论文顺利完成之际, 我首先要向我的导师\*\*\*老师表达深深的敬意感谢。本论文是在\*老师的悉心指导之下完成的。在课题的研究过程中, \*老师每次对我的疑问都给予细心的解答并给出建议, 使笔者顺利的解决了研究过程中遇到的各种困难。在撰写论文的过程中他还对我进行了细心的指导, 认真阅读和修改了我的论文, 使我顺利的完成了论文撰写。

感谢所有在大学四年期间, 帮助过我的老师们, 是你们引导我一步一步向前走, 让我不断成长。大学四年的时光, 是我们从家长的庇护下真正走向社会的过渡阶段, 你们的包容、教诲和对于我的耐心指导, 都指引着我在学习和科研的道路上走得更远。

感谢我所有的同学、朋友, 感谢你们四年的陪伴和宽容, 陪伴我走过四年的大学时光, 让我每一天的生活变得更加丰富, 更加充实。

再次感谢你们, 谢谢你们为我所做的一切, 希望在未来的日子里, 你我都能在不同的生活中闪闪发光!

- 说明: 1. 总文字复制比:被检测文献总重复字符数在总字符数中所占的比例
2. 去除引用文献复制比:去除系统识别为引用的文献后, 计算出来的重合字符数在总字符数中所占的比例
3. 去除本人文献复制比:去除系统识别为作者本人其他文献后, 计算出来的重合字符数在总字符数中所占的比例
4. 单篇最大文字复制比:被检测文献与所有相似文献比对后, 重合字符数占总字符数比例最大的那一篇文献的文字复制比
5. 复制比按照“四舍五入”规则, 保留1位小数;若您的文献经查重检测, 复制比结果为0, 表示未发现重复内容, 或可能存在的个别重复内容较少不足以作为判断依据
6. 红色文字表示文字复制部分;绿色文字表示引用部分(包括系统自动识别为引用的部分);棕灰色文字表示系统依据作者姓名识别的本人其他文献部分
7. 系统依据您选择的检测类型(或检测方式)、比对截止日期(或发表日期)等生成本报告
8. 知网个人查重唯一官方网站:<https://cx.cnki.net>