

分类号：

密级：

U D C：

学号： 406100210077

南 昌 大 学 硕 士 研 究 生

学 位 论 文

基于联合教学知识蒸馏与特征相关性的 艺术风格分类研究

Research on Art Style Classification Based on Collaborative
Teaching Knowledge Distillation and Feature Correlation

刘金

培养单位（院、系）：数学与计算机学院计算机科学与技术系

指导教师姓名、职称：刘丽 副教授

申请学位的学科门类：工学

学科专业名称：计算机科学与技术

论文答辩日期：2024 年 6 月 1 日

答辩委员会主席：_____

评阅人：_____ 盲审

_____ 盲审

2024 年 6 月 1 日

摘 要

随着人们对精神和文化生活质量的需求日益增强,绘画、建筑、摄影等视觉艺术形式已经渗透到日常生活的每一个角落,成为不可或缺的重要组成部分。艺术图像风格分类不仅增强公众对不同艺术流派特征的认识,还推动艺术教育和研究的发展,同时为数字艺术和在线艺术平台的个性化推荐提供有力支撑。因此,运用计算机视觉技术来辅助非专业人士理解与鉴赏视觉艺术作品的风格变得愈发重要。然而,艺术图像风格分类存在很多挑战。首先,艺术图像作品的风格不仅体现在颜色、线条或形状等直观元素上,更体现在作品的深层次意象和情感表达中,这要求模型能够捕捉并理解艺术作品的深层语义。此外,由于艺术图像数据集的数量庞大且分辨率高,对计算资源和存储空间的要求较高,在资源受限的设备上难以部署。

为了解决上述问题,本文使用了一种基于联合教学知识蒸馏与特征相关性的艺术风格分类模型。该模型具有初始教师,专家教师 and 自适应加权格拉姆矩阵等三个核心部分,这些模块的协同工作使得模型能够有效的识别艺术风格特征。专家教师网络采用注意力图机制指导学生网络,帮助学生网络学习如何有效地处理和解析输入数据。同时,初始教师网络在与学生网络的共同训练中通过其阶段性输出逐渐引导学生网络,使学生网络拥有更优的学习路径,并通过 L2 约束防止学生网络过度模仿教师网络,进而保持学生网络的独立性和灵活性。此外,自适应加权格拉姆矩阵通过结合风格特征间的相关性,有效提高了模型对细微风格差异的捕捉能力,提高艺术图像风格分类的性能。

本文在三个公开的艺术风格数据集上进行了系统的实验,结果表明本文所提出的方法在艺术风格识别任务上展现出优越的性能。消融实验结果显示,模型中的每个模块都在提升整体性能方面发挥了重要作用。在 WikiArt、MultitaskPainting100k 和 Pandora18k 等三个公开艺术图像风格数据集上,本文所使用的方法的分类准确率分别达到了 73.48%、72.26%和 66.83%,均超过了现有的艺术图像风格分类方法,证明基于联合教学知识蒸馏与自适应加权格拉姆的艺术风格分类模型的有效性。

关键词：艺术图像风格分类，联合教学知识蒸馏，特征相关性，自适应加权格拉姆矩阵

ABSTRACT

With the increasing demand for spiritual and cultural quality of life, visual art forms such as painting, architecture and photography have penetrated into every corner of daily life and become an indispensable and important part. Art image style classification not only enhances the public's understanding of the characteristics of different art genres, but also promotes the development of art education and research, and at the same time provides a strong support for personalized recommendation of digital art and online art platforms. Therefore, it has become increasingly important to use computer vision technology to assist non-specialists in understanding and appreciating the styles of visual art works. However, there are many challenges in categorizing art image styles. First, the style of art image works is not only reflected in intuitive elements such as colors, lines or shapes, but also in the deeper imagery and emotional expression of the works, which requires the model to be able to capture and understand the deeper semantics of art works. In addition, due to the large number and high resolution of art image datasets, it requires high computational resources and storage space, which is difficult to be deployed on resource-constrained devices.

In order to solve the above problems, this paper uses an art style classification model based on collaborative teaching knowledge distillation and feature correlation. The model has three core components such as initial teacher, expert teacher and adaptive weighted Gram matrix, and the synergy of these modules enables the model to recognize art style features effectively. The expert teacher network uses the attention graph mechanism to guide the student network and help the student network learn how to process and parse the input data effectively. Meanwhile, the initial teacher network gradually guides the student network through its stage-by-stage outputs during the co-training with the student network, so that the student network has a better learning path, and prevents the student network from over-imitating the teacher network through the L2 constraints, which in turn maintains the independence and flexibility of the student network. In addition, the adaptive weighted Gram matrix effectively improves the model's ability to capture subtle style differences by combining the correlation between

style features, and improves the performance of art image style classification.

In this paper, systematic experiments are conducted on three publicly available art style datasets, and the results show that the proposed method demonstrates superior performance on the art style recognition task. The results of the ablation experiments show that each module in the model plays an important role in improving the overall performance. On three publicly available art image style datasets such as WikiArt, MultitaskPainting100k, and Pandora18k, the classification accuracies of the method used in this paper reach 73.48%, 72.26%, and 66.83%, respectively, which are all superior to existing art image style classification methods, demonstrating that based on the co-taught knowledge distillation and self Adaptive Weighted Gram's art style classification model is effective.

Keywords: art image style classification, collaborative teaching knowledge distillation, feature correlation, adaptive weighted Gram matrix

目 录

第 1 章 引言.....	1
1.1 研究背景与意义.....	1
1.2 研究现状与趋势.....	2
1.2.1 艺术图像风格表示.....	2
1.2.2 艺术图像风格分类.....	4
1.2.3 知识蒸馏.....	6
1.3 本文研究内容.....	9
1.4 论文章节安排.....	9
第 2 章 相关理论及技术	11
2.1 主干网络介绍.....	11
2.1.1 ResNet.....	11
2.1.2 WideResNet.....	12
2.2 空间注意力机制.....	14
2.3 知识蒸馏.....	16
2.3.1 知识蒸馏的基本框架.....	17
2.3.2 基于中间特征的知识蒸馏.....	19
2.3.3 基于助推器网络的知识蒸馏.....	20
2.3.4 基于空间注意力图的知识蒸馏.....	22
2.4 本章小结.....	23
第 3 章 基于联合教学知识蒸馏与特征相关性的风格分类	25
3.1 基于联合教学知识蒸馏与特征相关性的模型结构图	25
3.2 自适应加权格拉姆矩阵介绍	26
3.3 联合教学知识蒸馏模型介绍	30
3.4 本章小结.....	34
第 4 章 实验设置及结果分析	35
4.1 数据集.....	35
4.2 实验环境及参数设置.....	37
4.2.1 评价指标.....	37
4.2.2 实验环境.....	38
4.2.3 实验参数设置.....	38
4.3 实验结果分析.....	39
4.3.1 超参数 α, β 对网络性能的影响	39
4.3.2 知识蒸馏温度相关实验.....	40
4.3.3 不同空间注意力图效能对比实验.....	42

4.3.4 不同相关性权重矩阵的效果评估实验	43
4.3.5 不同宽度因子的知识蒸馏方法性能比较	47
4.3.6 消融实验	50
4.3.7 混淆矩阵与误差分析	51
4.3.8 与现有方法的对比实验	57
4.4 本章小结	60
第 5 章 结论与展望	61
5.1 结论	61
5.2 展望	61
参考文献	63

第 1 章 引言

1.1 研究背景与意义

在最近几年，随着物质生活的快速发展，人们开始更加注重精神和文化生活的提升。这种趋势体现在人们对于博物馆和画廊的浓厚兴趣上。艺术品，如雄伟的建筑、多姿多彩的服装和精美的手工艺品，已成为日常生活的重要组成部分。这些艺术品不仅反映了人们的个人审美，也是对美的不断追求的体现。人类对艺术的热爱可以追溯到远古时期，那时人们已在岩石上绘制壁画。从古至今，参与和欣赏艺术活动，如摄影，已成为人们日常生活中放松和陶冶情操的方式。随着数字艺术平台和在线美术馆的兴起，网络成为了众多艺术爱好者获取知识的新途径。在这种趋势下，利用计算机科技对大量艺术创作进行深度处理和细致分析变得尤为关键，这不仅可以促进公众对艺术作品的深刻理解，还能增强他们的审美体验。通过这样的技术手段，能够揭示艺术作品背后的历史脉络、艺术风格的演变以及艺术家的创作意图，从而为广大艺术爱好者提供一个更加丰富和立体的艺术学习平台。此外，这种技术的应用还可以推动艺术教育的创新，使艺术欣赏和教学活动更加生动和互动，进一步拓宽人们接触和理解艺术的渠道。这种技术的运用不仅使艺术作品更易于被大众接触，也为艺术教育和研究开辟了新的路径。

在评价艺术创作时，分析专家通常会深入探讨包括线性构图、色彩对比度、光影效果以及画面的整体布局等关键视觉因素。例如，著名的瑞士艺术理论家沃尔夫林，在其开创性著作《艺术风格学》^[1]中，依据这些视觉属性，提出了辨识文艺复兴与巴洛克艺术风格的五大原则。同样，艺术历史学者贡布里希在《艺术的故事》^[2]一书中探讨 16 世纪初期意大利北部和威尼斯地区的艺术时，特别注重了作品中光与色的运用和表现力。此外，不同的艺术流派在选材和主题内容上亦有着鲜明的区别；例如，文艺复兴和巴洛克时代的画作常涵盖人像、宗教和叙事主题，而浪漫主义与现实主义作品则倾向于描绘自然风光。在目前对视觉艺术品进行分类与分析的研究中，学者们倾向于集中于这类视觉属性的分析。他们经常采用自定义的视觉描述工具或依托于深度学习技术的模型框架，以捕获艺术

图像中的关键视觉特征。这些分析工具，包括视觉描述符和卷积神经网络中的基础特征，能够洞察到图像的线条、色彩和纹理等细节，而网络中的高阶特征则有助于解析作品的深层次意义和语义内容。

与此同时，为了解决独特风格的分类问题，人们提出了更深更广的架构，但这也伴随着增加了存储和计算时间的代价。然而，这些庞大的模型却无法对模型效率的迫切需求。因此，模型压缩已经成为追求更高效模型的基本主题，旨在获得性能可接受的轻量级模型。在这个领域，人们已经尝试了多种技术，包括过滤器剪枝^[3]、知识蒸馏^[4]、低秩分解^[5]、参数量化^[6]等。其中，知识蒸馏是一项被广泛探讨的技术，因为它具有硬件友好的特性且无需使用特殊的库。在知识蒸馏方法中，通过训练轻量级网络（被称为学生网络）来模仿大型网络（被称为教师网络）。与没有指导的训练相比，在教师网络的指导下，学生网络能够获得更出色的性能。

1.2 研究现状与趋势

1.2.1 艺术图像风格表示

对于视觉艺术作品的风格分析与一般的图像精细分类任务相比，存在根本的区别。在传统的图像分类中，主要关注于图像中心的显著物体，而背景元素通常被忽略。这种方法在处理日常物体时效果显著，但在分析艺术作品时却显得不足。在艺术画作中，不论是显著还是非显著区域，甚至包括背景，都是艺术家精心构思的，每个细节都蕴含着艺术家的个性和时代特色。因此，在处理视觉艺术图像时，不能简单地忽视任何部分。

在 Peng 等人^[7]的研究中，提出了一个假设：艺术作品中分割出的每个部分都应该保留其原始的艺术家和风格标签。这一假设强调了艺术作品中每个元素的重要性，即使是那些看似不显眼的部分。这种对艺术图像进行全面分析的方法不仅对艺术历史和鉴赏学产生深远影响，同时也为计算机视觉和人工智能领域带来了新的挑战和发展机遇。通过深入探究艺术图像的每个角落，可以更全面地理解艺术作品的复杂性，同时也可能促进图像分析和识别技术的进步。

当前，许多研究者指出图像的纹理特性与其艺术风格之间存在密切的关系。在采纳手工设计的视觉特征技术成为主流之前，已有研究开始探讨纹理信息在

辨识视觉艺术图像方面的重要作用^[8]。此外, Khan 等人^[9]也将 LBP 技术^[10]应用于视觉艺术图像风格的分类中。

在艺术画风格转化领域,目标是在保持原图内容的基础上,实现风格的迁移。这一目标通常通过独立处理图像内容和风格来实现,纹理合成在风格转换中扮演着关键角色。传统纹理合成分为基于像素和基于纹理片段的方法。如 Efros 等人^[11]开发的无参数技术,像素方法从单点出发,逐像素扩展以模拟纹理。纹理片段方法则拼接多个小图像片段构建纹理。针对拼接不连贯问题, Wu 等人^[12]提出了基于曲线特性分析和结构相似性测量的匹配对齐解决方案,从而有效缓解了纹理片段间的断裂, Xie 等人^[13]则进一步提出了一种特征驱动的纹理合成新算法,成功地实现了图像艺术风格的转换。

随着深度学习领域的快速发展,一些学者开始采用深度特征的相关性分析来描述图像的纹理信息,进而应用于艺术风格的转换。2015 年, Gatys 等人^[14]通过卷积神经网络(CNN)推出了一种创新的纹理合成方法,该方法通过分析网络中特定层级的特征图相互间的相关性来抽取纹理信息。他们进一步展示了通过跨多个层级提取纹理特征可以更加深入地揭示图像的细节,从而生成更高质量的图像。继而在 2016 年, Gatys 等人^[15]基于此前的工作,提出了一种革命性的神经风格迁移技术。该技术独立处理图像的风格与内容,之后再将它们融合,创造出一幅既反映了原始内容又融合了特定艺术风格的新图像。在风格定义上,他们利用了一个基于网络各层滤波器反应构建的特征空间,以捕获纹理信息。通过计算不同层级的滤波器反应生成的 Gram 矩阵,实现了风格特征的相关性分析。这一突破性的技术及其在艺术风格迁移方面的应用随后激发了更多学者进行深入的研究与探索。

在艺术风格转换研究的影响下, Chu 等人^[16]在执行视觉艺术图像的分类任务中,引入了纹理特征作为表现艺术风格的新途径。他们依靠分析特征图相互之间的关联度来描绘图像的纹理属性,并首先利用卷积神经网络捕获图像的深层特征,随后采用多种相关性分析技术,将这些深层特征转换为代表风格的向量,并进行了详尽的效能评价。他们进一步探索了多个相关性分析技术的结合使用,并评估了其在分类效率上的差别。借鉴于风格迁移领域的方法,他们也采用了 Gram 矩阵进行实验,以此验证了 Gram 矩阵在对视觉艺术图像进行分类上的功效,并对比了基于不同深度层次的特征图计算得出的相关性特征在性能上的不同。

在近期的艺术图像风格表示领域中, Ruta 等人^[17]致力于探索更加高效和准

确的风格转换技术。其中一项具有代表性的研究是通过利用最新的神经风格转换（NST）技术，来实现对视觉艺术风格更加解耦的表示学习。这项技术的核心是创建一个合成的数据集，其中风格保持一致，而内容的语义变化广泛。这样的方法对于训练模型来说至关重要，因为它可以有效地区分图像的风格和内容。在这项工作中，他们提出了一种新颖的方法，使用快速的前馈神经风格转换方法动态生成图像，从而在不需要大量存储空间的前提下，最大化风格的多样性。通过对比损失来促进风格学习，实现了一种自监督的学习框架，强调了风格信息而非图像的语义内容。

1.2.2 艺术图像风格分类

识别艺术图像风格经常涉及分析其核心视觉元素，比如色彩、线性表达和画面布局，这些因素吸引了大量计算机视觉领域研究者的关注^[18-20]。在这一研究领域，研究人员常利用视觉描述工具或采用深度学习技术来捕捉这些关键特征，进而实现艺术图像风格的有效分类和深入分析。

在计算机视觉及多媒体研究领域，利用计算技术对视觉艺术图像按风格进行系统的分类和详尽分析已成为一个重要的研究方向。多种研究手段已被开发，采用手工定义的视觉特征来解决这个问题。例如，Sablatnig 等人^[21]开发了一个系统，专为分析肖像画中艺术家的绘画风格而设计，该系统通过分析颜色特征、脸部形状和笔触结构等多层次信息来辅助分类。另外，Zujovic 等人^[8]依据艺术流派，利用颜色、纹理和边缘等显著信息对艺术图像进行分类，并且针对不同的应用场景使用了多分辨率数据。Shen 等人^[22]提出了一个整合多视觉特征的框架，并通过一种编码方案来融合子分类器的分数，以支持大规模西方画作的自动识别。颜色直方图在分类中被证明有效，因为相似风格或艺术家的作品中常用颜色较为一致^[23]。尺度不变特征变换（SIFT）技术^[24]被广泛应用于识别图像中的局部特点，它对于图像的扭曲和光照的变化表现出了显著的鲁棒性，这一点在视觉艺术图像的分类任务中得到了充分的验证。另一方面，空间包络特征（GIST）技术^[25]专注于提取图像的整体属性，这使得它在场景理解以及艺术作品的辨识中发挥了重要作用。这两种技术各自针对图像的局部和全局特征提供了有效的分析手段，为艺术图像的深入研究和识别提供了强有力的工具。Guo 等人^[26]对局部二值模式用于描述纹理特征的方法进行了改进，以捕获局部差异特征。Khan

等人^[9]对包括词袋模型、局部二值模式（LBP）、GIST 在内的多种特征提取技术在艺术家识别和艺术风格分类上的效用进行了探究，他们的研究成果表明，将这些视觉特征综合运用能显著增强分类的准确度。另一方面，Xu 等人^[27]通过采用多项式潜在对数回归模型来对建筑风格的图像进行分类，展示了这种方法在艺术领域分类问题上的应用潜力。这些研究不仅拓展了视觉特征应用的边界，也为艺术作品和建筑风格的自动识别提供了新的视角和方法，这种方法适用于处理多类别问题。Chen 等人^[28]通过提取绘画的局部颜色特征，并使用 K-均值聚类 and 词袋模型来训练分类器，这种方法适合小规模训练样本。

随着深度学习技术在图像识别领域的显著进展，卷积神经网络(CNN)在视觉艺术图像的分类和分析中变得日益流行。Karayev 等人^[23]提出了一种分类方法，该方法依据图像在视觉风格上的多个维度进行区分，包括摄影方法、布局设计、情感表达、艺术流派以及描绘的场景等。他们运用了标准的视觉描述符以及深度学习提取的特征来执行风格的区分工作，并对比了依靠单一特征与综合多种特征进行分类的效能。这一研究不仅拓宽了对图像风格分类的认识，也为如何有效结合不同类型的特征提供了深入的见解。Hicsonmez 等人^[29]利用深度学习模型成功辨识了儿童图书插画中的绘画风格，这一成果凸显了深度网络在抽取绘画中复杂语义信息方面的出色能力。Elgammal 等人^[30]探讨了利用卷积神经网络(CNN)进行艺术作品分类的效果，并指出了网络分类结果与艺术史上下文之间的联系，这一发现为使用 CNN 进行艺术鉴赏提供了更多的解释性依据。同时，Lecoutre 等人^[31]采用深度残差网络来鉴定不同的画作风格，展现了深度学习技术在艺术风格认识上的应用潜力。这些研究表明，深度学习不仅能够识别艺术作品的风格，还能深化对艺术历史和文化背景的理解。Tan 等人^[32]开发了基于模糊量空间的深度压缩方法，以减少 CNN 模型的内存需求并应用于艺术分类，展示了从 CNN 多层结构提取的特征融合在艺术分类中的有效性。

Chu 等人^[16]强调了纹理特征在视觉艺术风格分类中的重要性，通过 CNN 提取特征并采用不同距离度量进行图像分类。Peng 等人^[7]开发了一种创新的研究方法，基于卷积神经网络（CNN），他们认为艺术作品的风格具有传承性。为了分类这些风格，他们采用了多尺度特征提取技术，并通过这种方式生成了大量的训练数据，有效缓解了样本稀缺的挑战。此外，他们还利用了这些多尺度特征来识别艺术作品中的关键区域，并结合费舍尔向量编码以及支持向量机（SVM）技术完成了作品风格的精确识别。在此基础上，Sethi 等人^[33]利用 CNN 技术构建了

一个开源的网络平台,旨在对艺术作品的风格进行分类与深度分析,并且专门设计了用于辅助风格辨识的艺术特征工具。受到深度学习理念的启发, Gultepe 等人^[34]采纳了 K-均值聚类方法来进行无监督的特征学习,通过对艺术作品中的边缘、线条和色彩特征进行聚类分析,他们探讨了不同艺术风格之间的相互关系,开拓了艺术风格研究的新视角。为了在风格识别中有效利用卷积特征的相关性, Chen 等人^[35]提出了一种端到端的网络结构,该网络通过编码不同特征映射之间的相关性并计算卷积特征的内积,来实现一个自适应且跨层的模型。这种方法旨在提取图像的纹理信息,以同时捕捉高层的抽象和低层的细节信息,与仅关注物体的细粒度分类任务不同。通过在每个空间位置上对特征向量进行自适应加权,这种方法能更准确地对视觉艺术图像进行分类。

Joshi 等人^[36]提出了一种利用自训练的自编码变换集合进行艺术风格分类的新方法。这种方法通过自编码变换作为正则化网络,学习不同图像变换下的鲁棒特征,以提高对变换图像的标签预测一致性。通过这种方式,研究团队成功地应用了半监督学习的范式,不仅提升了分类的准确性,还增强了模型对艺术图像细节结构的捕捉能力,这在艺术风格分类的领域中是一个重要的进步。Chen 等人^[37]针对中国艺术画作的艺术风格分类问题,利用卷积神经网络(CNN)模型提出了一种新的分类方法。通过细调预训练的 VGG-F 模型来提取中国画的图像特征,并引入互信息理论到嵌入式机器学习中,提出了一种基于互信息的嵌入式分类算法。这种方法考虑了特征选择和特征重要性的影响,展现了深度学习技术在中国艺术画作风格分类中的应用潜力,并为中国艺术画作的数字化分类与研究提供了新的技术路径。Akter 等人^[38]探索了使用浅层卷积神经网络(SCNN)自动识别绘画艺术风格的新方法。他们提出的模型包含两个连续的卷积层用于提取绘画特征,以及一个全连接层用于识别艺术风格。这种模型设计尝试通过减少网络深度来避免过拟合和欠拟合问题,同时保持高准确率。该研究显示了深度学习技术在艺术风格自动识别方面的应用潜力,并为数字艺术分类和识别领域贡献了新的视角和工具。

1.2.3 知识蒸馏

在近年来,深度学习技术已经在计算机视觉(CV)和自然语言处理(NLP)等领域取得了显著成就,并且这些成就已经在商业环境中得到了广泛应用。这些

技术的成功很大程度上得益于计算设备的性能提升和存储能力的增加。然而，现有的深度学习模型往往过分强调提高性能指标，如准确性，而忽视了模型规模的膨胀问题，这导致了模型变得庞大而复杂，难以在有限资源的设备上进行有效部署。特别是在网络流量的分类和识别等应用场景中，确保服务质量（QoS）不仅需要模型具有高的准确率和低的误报率，还要求模型能够快速计算、响应迅速、并且运行稳定。因此，如何实现模型的轻量化、降低参数数量、提升计算效率，已成为工业界面临的关键挑战。面对这一挑战，研究者们提出知识蒸馏的优化策略，它通过将复杂模型中的知识转移到更简单高效的模型中，实现了模型的加速和轻量化，为深度学习模型的优化和应用开辟了新的道路。这种方法不仅促进了模型在资源受限环境中的应用，也为深度学习技术的未来发展提供了新的思路 and 工具。

Bucilua 等人^[39]是最早探索模型压缩技术的先驱，他们通过 MUNGE 技术实现了大模型知识向小模型的有效转移。紧接着，Ba 等人^[40]通过利用深度网络生成的 logits 以及应用 L2 正则化策略，成功地训练了一个性能优异的浅层网络。他们认为，深层网络由于其能够从复杂任务中提取高阶特征的能力，其生成的 logits 包含了丰富的信息，比如不同类别之间以及类别内部的相似性。随后，Hinton^[41]及其同事首次引入了知识蒸馏这一术语，强调了通过模仿教师模型，学生模型能够更有效地学习复杂知识并展现出出色的泛化能力。在 Yang 等人^[42]的工作中，他们提出了一种从 logits 选择置信度最高的几个类别，并假设这些类别与输入图像有着密切的关联，通过计算这些顶级置信类别与其他类别的损失，来优化网络的训练过程。为了更精确地理解和传递教师模型中的知识，Kim 等人^[43]提出了一个包含编码器和解码器的卷积模型架构。其中编码器在无监督条件下训练，用于捕获教师模型中的关键信息，而解码器则负责将这些信息转换为学生模型可以理解的格式。通过这种辅助训练方法，学生模型的性能得到了显著提升。这些进展共同构成了基于模型响应的知识蒸馏研究的框架。

在深度学习领域，中间层的特征表示被认为更为丰富和具有高度代表性，这促使了基于特征图的知识蒸馏方法的发展，并已经证明了其有效性。为了捕获教师模型中深层的特征表示，Romero 等人^[44]开发了一种名为 Fitnets 的框架，该框架通过直接采用教师模型的激活层特征来训练学生模型。Mishra 等人^[45]使用知识蒸馏（KD）从零开始联合训练一个教师网络（全精度）和一个学生网络（低精度）。Zagoruyko 等人^[46]则采用了一种创新的方法，通过生成原始特征图的注

意力图来更精确地传达知识。鉴于不同特征图和数据样本所包含的信息差异，Zhou 等人^[47]引入了助推网络来指导轻量级网络的学习过程，利用加速器网络在整个训练过程持续指导轻量级网络的学习。Passalis 等人^[48]则通过将数据样本间的相互关系转化为概率分布，有效地揭示了特征空间中的数据样本相关性，这不仅克服了知识蒸馏仅限于分类任务的限制，还能深入学习特征空间的几何结构。Park 等人^[49]采纳了一种多教师网络输出单元的方法，替代了传统的单一教师模型。这种方法基于的思想是，不同的教师网络能够掌握多样化的知识，通过整合多教师网络，学生模型得以从中学习到更加广泛和多元的知识。

在探索知识蒸馏的不同途径中，除了采用结构不同的教师和学生模型进行学习的传统方法外，Zhang 等人^[50]引入了自蒸馏的概念来增强卷积神经网络的效能。通过他的研究发现，当教师和学生模型拥有相同架构时，采用实时蒸馏策略能够有效地将深层特征知识传递给浅层神经元，促进了更加精确的特征提取，并提高了模型训练和推理的速度。Xu 等人^[51]也通过元网络学习软目标进行自蒸馏策略，专注于学习统一的全局特征分布和后验概率，它不需要额外的教师模型。相反，它通过模型本身或模型的不同版本来实现知识的蒸馏。这种方法的模型在训练过程中同时扮演教师和学生的角色，这种方式有助于提升模型的性能，尤其是在有限数据情况下。Li 等人^[52]提出的研究着眼于通过层级知识蒸馏优化深度神经网络的学习。他们开发了一种层选择性学习（LSL）框架，该框架侧重于使用大模型来指导小模型的训练过程，并通过精选特定的中间层进行知识传递，以此来提升模型性能。该方法突出了在深度学习模型压缩过程中，如何有效地在不同层级之间传递和利用知识的重要性，从而达到性能提升的目的。

Zhao 等人^[53]提出了一种新的协作教学知识蒸馏（CTKD）策略，旨在通过从大型预训练教师模型传递知识来训练一个紧凑的学生网络，以解决深度神经网络在资源受限设备上部署时面临的高存储和计算成本问题。他们的方法使用两种特殊的教师：一位是从头开始训练的“初学教师”，它逐步使用其临时输出辅导学生，强迫学生朝着最终的高准确率逻辑输出优化；另一位是预训练好的“专家教师”，指导学生关注任务中更有用的关键区域。这两位教师提供的知识结合起来，可以显著提高学生网络在知识蒸馏中的性能。Tudor 等人^[54]引入了一种在线知识蒸馏框架，该框架利用创新的注意力机制来有效整合轻量级网络群体的预测，构建了一个强大的集成体作为蒸馏信号。这项研究的重点在于探索如何通过注意力机制增强学生网络群体的性能，使之在图像分类等任务中取得更好的

效果。此外，该框架还注重于轻量级模型的部署，旨在边缘设备上实现高效的深度学习模型部署，这对于资源受限设备的深度学习应用至关重要。Liu 等人^[55]提出了一种动态可蒸馏性和稀疏性学习框架（DDSL）用于模型压缩。它动态地整合了知识蒸馏（KD）和结构稀疏性剪枝，并通过评估模型的可蒸馏性和稀疏性来调整联合教学过程。与“教师-学生”框架不同，所提出的 DDSL 可以被描述为一个“在学校学习”的框架，它由三个部分组成：教师、学生和院长。教师像以前一样教学生。院长控制学习强度和学习方式。根据教师和学生当前的状态，院长网络评估学生网络的可蒸馏性和稀疏性，然后动态平衡和控制 KD 和结构稀疏性剪枝的监督强度。

这些研究不仅拓宽了知识蒸馏技术的应用范围，也为提升模型效率和性能提供了新的思路。

1.3 本文研究内容

为了应对艺术图像风格分类的挑战，本文使用了一个基于联合教学知识蒸馏与特征相关性技术的模型框架。通过知识蒸馏策略，模型不仅减少了推理时间，同时提高了艺术风格识别的性能，在该过程中引入的初始教师和专家教师网络，分别从不同角度指导学生网络，优化学习路径，增强模型对关键风格特征的捕捉能力。此外，通过自适应特征格拉姆矩阵分析风格特征间的相关性，提高了模型对细微风格差异的识别精度。通过在三个公开艺术风格数据集上的实验验证，模型展现了在艺术风格分类任务上的出色性能，证明了所提方法的有效性。

1.4 论文章节安排

本文探讨了基于联合教学知识蒸馏和特征相关性技术的艺术风格分类方法。通过深入分析与系统性研究，论文构建了一个高效的艺术风格识别框架，有效地提高了分类精度和模型的实用性。本研究的组织结构安排如下：

第 1 章，引言。本章节首先介绍了研究背景，阐述了艺术风格分类的重要性及其在数字化艺术领域的应用前景。随后，详细探讨了知识蒸馏与艺术图像风格表示及分类的研究现状与发展趋势。

第 2 章，相关理论及技术。本章详细介绍了构建本文模型所依赖的关键理论

和技术，包括主干网络架构、空间注意力机制以及知识蒸馏的原理和应用。这为理解后续章节的模型构建与实验验证奠定了理论基础。

第 3 章，模型构建与实现。详细阐述了基于联合教学知识蒸馏与特征相关性的艺术风格分类模型的设计思路、网络结构及实现方法。通过对模型组件的细致描述，展现了该模型如何有效捕获艺术作品的风格特征。

第 4 章，实验设置及结果分析。本章记录了模型在多个公开艺术风格数据集上的实验设置、参数优化过程以及实验结果。通过与现有技术的对比分析，验证了所提出模型的性能优势和应用价值。

第 5 章，结论与展望。总结了本研究的主要发现和贡献，并对可能的未来研究方向进行了展望，指出了该领域的研究潜力和挑战。

第 2 章 相关理论及技术

2.1 主干网络介绍

在本节中将简单介绍本次毕业设计中使用的主干网络架构，包括残差网络 ResNet^[56]和宽残差网络 WideResNet^[57]。

2.1.1 ResNet

增加卷积神经网络的深度被认为能提高模型性能，但超过一定深度后，性能可能下降，称为网络退化。ResNet 通过引入残差学习解决这一问题，允许深层网络维持性能，通过残差块实现特征的增量学习和无损信息传递，保持梯度稳定，简化训练过程。图 2.1 展示了 ResNet 通过短接线的跳跃连接过程：

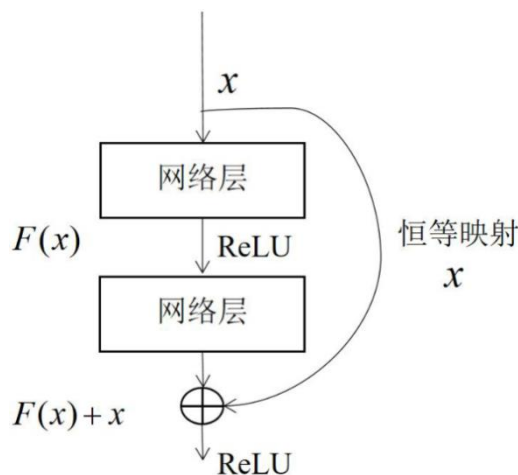


图 2.1 ResNet 的跳跃连接过程

ResNet 的几个变体，包括 ResNet-18、ResNet-34、ResNet-50、ResNet-101 和 ResNet-152 等，展现了不同深度的网络在实际应用中的性能，其中的数字表示模型中含参数的层的数量，涵盖了卷积层和全连接层等不同类型的层。数字代表的是模型中参数层的总数，这包括了卷积层、全连接层等。不同的版本让研究者和开发者可以根据具体的需求和资源限制选择最合适的模型，这在实际应用中是非常有价值的。通过这种方式，ResNet 不仅解决了深层网络训练的难题，还提供了一种灵活的框架来探索深度学习的极限。

ResNet-34 是残差网络 (Residual Network) 的一种, 其特点是使用残差块来跳过一些层, 从而能够构建更深的网络结构而不会出现训练难度增大的问题。ResNet-34 模型的结构简介如表 2.1 所示:

表 2.1 ResNet-34 网络模型框架

组块名	输出尺寸	层结构
Conv1	112×112	[7×7, 64], stride=2, padding=3
MaxPool	56×56	[3×3], stride=2, padding=1
Conv2	56×56	3×[3×3, 64]
Conv3	28×28	4×[3×3, 128]
Conv4	14×14	6×[3×3, 256]
Conv5	7×7	3×[3×3, 512]
AvgPool	1×1	[7×7] Average Pooling
FC	1×1	[512, 1000]
Softmax	-	[1000]

2.1.2 WideResNet

WideResNet(WRN) 提出了一种对标准残差网络 ResNet 进行改良的策略, 结构和 ResNet 相似, 通过减少网络的深度同时增加其宽度来优化性能。这一结构通过引入宽度因子 m 来扩展原有残差模块的卷积核数量, 即通过增加每个卷积层的输出维度 (即通道数) 来实现。与传统的深度优先策略不同, 即便是深度残差网络 ResNet 通过层叠更多的残差块以加深网络结构, 也未必能带来预期的性能提升, 因为每增加的层都不一定能为网络的学习能力带来实质性的增长。

ResNet 网络采用两种主要的结构单元, 即基础块 (basicblock) 和瓶颈块 (bottleneck), 它们的结构分别如图 2.2(a)和图 2.2(b)所展示。在此基础上, WideResNet 对这些残差块进行了改进, 其改进后的结构如图 2.2(c)和图 2.2(d)所示。改进的关键在于所谓的“宽残差块”, 这种结构通过在残差块内部显著增加卷积核的数量来扩展特征通道。具体来说, 与传统残差块的两层卷积结构不同, 每层含有较少卷积核, WideResNet 中的每个宽残差块同样包含两层卷积, 但卷积核的数目较大, 相当于将特征通道的数目提高了数倍。

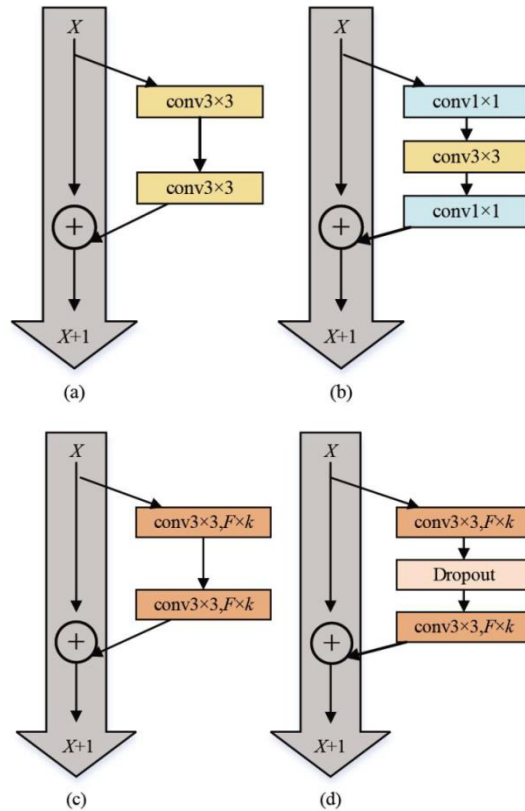


图 2.2 ResNet 的残差块和 WideResNet 的宽残差块对比

相较于单纯增加网络深度，扩展残差块的宽度在提升网络性能方面显现出更为有效的策略。这种方法虽然减少了网络的层数，但却保持了模型的参数数量不变，同时能够提升模型的计算效率。在实际应用中，宽残差网络（例如 WRN-40）与更深的残差网络（如 ResNet-101）在性能上相似，参数数量也相近，但是宽残差网络的训练速度却快了约八倍。这种设计不仅减轻了深度模型训练时的计算负担，也提供了一种有效途径来提高深度学习模型的训练和推理效率。

在 WideResNet 网络中，卷积层 Conv1 采用了标准尺寸（ 3×3 ），而引入的宽度系数旨在调整三个不同的残差块结构的维度。不同于传统的 ResNet 网络，此处引入了宽度因子 k ，其通过扩展或缩减残差块的尺寸来增强模型的灵活性，其中原始的 ResNet 可视为宽度因子 m 等于 1 的特例。此外， N 代表每个群组中残差块的数量。通过 $W(M)$ 来标记残差块的构造，其中 M 代表该残差块内卷积核尺寸的数组。例如， $W(3,3)$ 描述的是一个包含两个 3×3 卷积层的残差块。WideResNet 的基本结构详细地展示在了表 2.2 中。

表 2.2 WideResNet 网络的通用框架

组块名	输出尺寸	残差块结构 W (3,3)
Conv1	32×32	$[3 \times 3, 16]$
Conv2	32×32	$\begin{bmatrix} 3 \times 3, 16 \times k \\ 3 \times 3, 16 \times k \end{bmatrix} \times N$
Conv3	16×16	$\begin{bmatrix} 3 \times 3, 32 \times k \\ 3 \times 3, 32 \times k \end{bmatrix} \times N$
Conv4	8×8	$\begin{bmatrix} 3 \times 3, 64 \times k \\ 3 \times 3, 64 \times k \end{bmatrix} \times N$
Avg-pool	1×1	8×8

2.2 空间注意力机制

注意力机制^[58]的核心是将查询 (Query, Q) 映射到一系列键值对 (Key-Value, K-V)。这种机制在很多神经网络架构中, 尤其是在 Encoder-Decoder 框架中得到广泛应用。在这种框架下, 注意力机制通过计算查询与每个键之间的相似度, 然后用这些相似度作为权重, 来对所有值进行加权求和, 从而得到对应于给定查询的输出。

具体而言, 注意力机制可以被视为一个将输入序列映射到输出序列的过程, 其中每个输出元素都是输入元素的加权组合。这种加权组合的方式确保了模型能够关注 (即 “注意”) 到与当前任务最相关的输入部分。

在 Encoder-Decoder 架构中, 注意力机制允许解码器针对每个输出步骤, 专门关注到编码器的不同部分, 从而提高了信息的利用率和模型的性能。这种机制尤其在处理长序列数据时显示出其强大的能力, 比如在机器翻译、文本摘要、语音识别等任务中。其基本结构如图 2.3 所示:

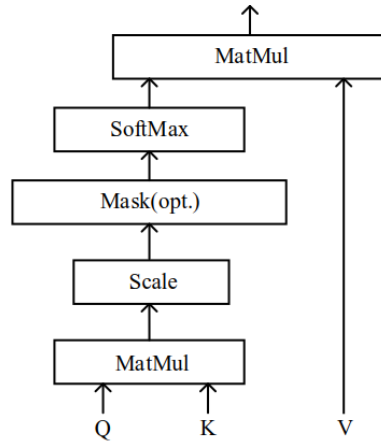


图 2.3 注意力机制结构

在注意力机制中，首先将输入的三维张量 $\chi \in \mathbf{R}^{H \times W \times C}$ 通过形状变换转换成二维矩阵 $X \in \mathbf{R}^{(HW) \times C}$ ，这一步骤简单地重塑了数据的维度。接下来，通过可训练的变换（通常是线性映射或卷积操作），将 X 映射成查询（ Q ）、键（ K ）和值（ V ）的表示形式，分别表示为 $Q = XW_Q$ 、 $K = XW_K$ 、和 $V = XW_V$ ，其中 W_Q 、 W_K 、 W_V 是相应的权重矩阵或卷积核。

接着，计算查询 Q 与键 K 之间的相似度，生成权重矩阵 $S = QK^T$ ，代表了不同组件间的关系强度。然后，将权重矩阵通过 `softmax` 函数进行归一化，确保权重的总和为一，这一步生成了最终的注意力权重。最后，将这些权重应用于值 V ，通过加权求和得到最终的输出 $Y = \text{softmax}(S)V$ 。这个过程使得模型能够集中关注最重要的部分，忽略不相关的信息，从而实现了输入特征的动态加权和聚焦。

空间注意力机制^[59]作为注意力机制领域的一个分支，在计算机视觉任务中的应用体现了一种高度先进的模拟人类视觉注意力分配技术。其主要思想在于对图像中各个空间位置进行重要性评估，并据此赋予不同的权重，使得深度学习模型能够优先处理对当前任务最为关键的信息区域。这样的机制不只是提升了模型在图像处理过程中的效率与准确性，还加强了其对图像细节及上下文信息的理解和捕捉能力。

在实现空间注意力机制的过程中，主要通过构建一个与输入特征图维度相同的注意力图来实现。这个注意力图通过一系列计算得到，其目的是为图像中的每一个像素或区域分配一个权重，这些权重反映了各自在当前任务中的重要性。

生成注意力图的函数可以是简单的卷积网络，也可以是更复杂的结构，关键在于能够有效地从输入特征中学习空间位置的重要性分布。

一旦生成了注意力图，就会使用逐元素乘法将其与原始的输入特征图相乘，得到最终的加权特征图。这一步骤实质上是在调整输入特征的分布，使得模型的“视线”自然而然地集中于那些被认为更重要的区域。在许多情况下，为了使注意力分配更加精细和合理，会对注意力图应用 softmax 函数等归一化操作，以确保所有空间位置的权重加起来为 1，从而在保持总体信息量不变的前提下，有效地突出重要特征。

在深入探讨空间注意力机制的核心原理和应用过程中，可以通过一个简洁的公式来概括其基本运作机制。空间注意力机制的核心是通过动态地分配不同的重要性权重于输入图像的各个空间位置，以促进模型更加关注于对当前任务更为关键的区域。这一过程可以被简化并表示为下列公式 2.1：

$$Y = (A \odot X) \quad (2.1)$$

其中， Y 表示加权后的特征图，用于后续的处理或决策任务； A 是通过某种形式的映射函数 f 从输入特征图 X 中得到的注意力图，即 $A=f(X)$ ，它为图像中的每个空间位置分配了一个权重； \odot 代表逐元素乘法操作，这一操作实现了将注意力权重应用到原始输入特征图上，从而调整其空间分布，使模型能够集中关注于重要区域。

通过上述公式，空间注意力机制实现了对输入特征的动态重塑，有效地增强了模型在处理图像识别、目标检测、场景理解等计算机视觉任务时的性能。这种机制通过精确调控模型的注意力焦点，不仅优化了信息处理流程，还大幅提升了模型对复杂视觉场景的理解能力。因此，空间注意力机制作为一种提高计算机视觉模型性能的关键技术，其研究和应用将继续是视觉计算领域的热点话题。

2.3 知识蒸馏

知识蒸馏，作为一种模型简化技术，因其适应性强且效果出色，在深度学习领域受到重视。该方法采用教师-学生框架，教师模型结构复杂、参数多，学生模型则更轻量、参数少。知识蒸馏通过固定教师模型来训练学生模型，使其不仅学习数据标签，还学习教师的深层知识，超越传统数据训练。这样，学生模型在

小体积下保留了教师的高级特征，平衡了效率和性能。

2.3.1 知识蒸馏的基本框架

Hinton^[41]在知识蒸馏领域的开创性工作引入了一个关键概念：软目标（soft targets）。软目标是通过将教师模型输出的 logits（即 softmax 函数之前的模型输出）应用温度参数 T 来获得的。这里，温度参数 T 起到了一个非常重要的作用：它调节了类别概率分布的平滑程度。温度参数 T 在知识蒸馏过程中扮演了一个非常重要的角色，它用于调整教师模型和学生模型输出概率分布的平滑程度。温度参数的主要作用是控制软目标的平滑性，从而帮助学生模型更好地捕捉和学习教师模型的知识。温度参数 T 被引入到 softmax 函数中，以调整输出概率分布的平滑度。下图 2.4 展示了知识蒸馏的基本框架：

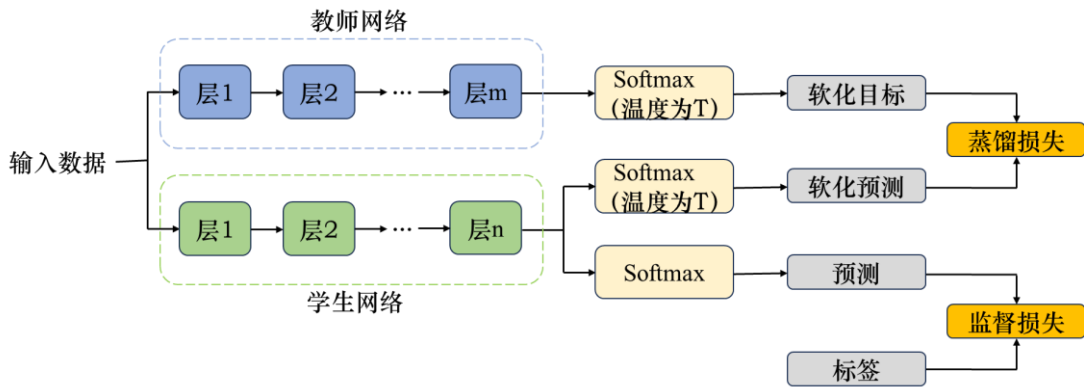


图 2.4 知识蒸馏基本框架

具体来说，在标准的 softmax 函数中，logits 被直接用来计算每个类别的概率。但在知识蒸馏中，通过将 logits 除以一个温度参数 T （通常 $T>1$ ），可以得到所谓的软目标。具体来说，对于每个类别 i ，其软目标的计算公式如式 2.2 所示：

$$p_i = \frac{e^{\frac{z_i}{T}}}{\sum_j e^{\frac{z_j}{T}}} \quad (2.2)$$

其中 p_i 是经过温度调节后的教师模型对类别 i 的预测概率，用于表示给定输入属于类别 i 的概率。 T 是温度参数，它是一个正实数。当 $T=1$ 时，上式等同于传统的 softmax 公式。当 $T>1$ 时，输出概率分布变得更加平滑。 z_i 是模型对类别 i 的原始输出（logit），分母中的 $\sum_j e^{z_j/T}$ 是一个归一化因子，确保经过 softmax 函数转换后的所有类别概率之和为 1。这个归一化过程涵盖了所有可能的类别，通过

对每个类别的 $\text{logit}z_j$ 应用指数函数并除以温度 T 然后对所有类别进行求和来完成。在上述公式中, z_i 和 z_j 是本质上代表同一组变量, 即模型对不同类别的原始输出, 也就是 logits 。它们之间的区别仅在于它们指向的是不同的类别。这表示对于特定类别 i 的 logit 。

总结来说, z_i 和 z_j 都是模型输出的 logits , 它们代表了模型对各个类别的评分或信心水平。区别在于, 当谈论 z_i 时, 通常是指与特定输出类别 i 相关联的 logit 。而 z_j 则更多用于表示涉及所有类别的计算或求和过程中, 作为一个通用索引。在 softmax 函数中使用这两个变量是为了将模型对特定类别的评分 z_i 转换成一个概率值 p_i 同时保证所有类别概率之和为一。

另外在知识蒸馏框架中, 监督损失 (Supervised Loss) 和蒸馏损失 (Distillation Loss) 是两个核心组成部分, 它们共同指导学生模型的训练过程。

(1) 监督损失: 监督损失用于确保学生模型可以正确地预测训练数据的标签, 就像在传统的监督学习任务中一样。这通常使用交叉熵损失 (Cross-Entropy Loss) 来实现, 计算公式如式 2.3 所示:

$$L_{\text{supervised}} = - \sum_{i=1}^N y_i \log(p_i) \quad (2.3)$$

其中 N 是类别总数, y_i 是真实标签的 one-hot 表示, p_i 是学生模型预测第 i 类的概率。

(2) 蒸馏损失: 蒸馏损失是知识蒸馏特有的损失, 它的目的是使学生模型的输出尽可能接近教师模型的输出。这通常通过计算学生模型输出和教师模型输出之间的某种距离或差异来实现。最常见的方法是使用带温度调节的软目标 (Soft Targets) 来计算两者的交叉熵, 蒸馏损失的公式如式 2.4 所示:

$$L_{\text{distillation}} = T^2 \times KL(\sigma(\frac{z_T}{T}) \parallel \sigma(\frac{z_S}{T})) \quad (2.4)$$

其中 $\sigma(\cdot)$ 表示 softmax 函数。 z_T 和 z_S 分别表示教师和学生模型的 logits (即 softmax 函数之前的层的输出)。 T 是一个大于 1 的温度参数, 用于平滑概率分布, T^2 是用来调整蒸馏损失在总损失中的比重, 确保它在梯度下降时有适当的贡献。 KL 表示 Kullback-Leibler 散度, 式 2.5 展示了 KL 散度的相关公式:

$$KL(P \parallel Q) = \sum_x P(x) \log(\frac{P(x)}{Q(x)}) \quad (2.5)$$

其中, P 和 Q 分别是离散概率分布, 是在这两个分布上的事件。对于连续概率分

布，求和（ \sum ）会被积分（ \int ）替代。

（3）总损失：在实际应用中，学生模型的总损失通常是这两种损失的加权和，式 2.6 展示了知识蒸馏中总损失的相关公式：

$$L_{total} = (1 - \alpha)L_{supervised} + \alpha L_{distillation} \quad (2.6)$$

其中 α 是一个介于 0 到 1 之间的权重参数，用于平衡监督损失和蒸馏损失。

这种组合方法使得学生模型能够同时学习到教师模型的知识（通过蒸馏损失）和直接从数据中学习的知识（通过监督损失），从而在保持模型大小和计算成本较低的同时，提高其性能。

2.3.2 基于中间特征的知识蒸馏

FitNet^[44]是一种基于特征的知识蒸馏方法。不同于传统的知识蒸馏，它主要关注于教师模型和学生模型之间的中间层特征，而不仅仅是输出层。FitNet 旨在帮助学生模型更好地学习和模仿教师模型的中间表示，这被认为可以帮助学生模型更有效地捕获和利用输入数据的信息，进而提升性能。

在传统的知识蒸馏方法中，通常只使用教师模型和学生模型的输出层来进行知识传递。然而，FitNet 提出了一种新的思路，即不仅仅蒸馏输出层的知识，还蒸馏中间层（即隐藏层）的特征表示。这种方法基于这样一种假设：教师模型的中间层特征包含了关于数据的重要信息和内部表示，这些信息如果能够被学生模型学习，将有助于学生模型更好地理解数据，从而提高其性能。

在 FitNet 的训练框架中，首先进行的是“Hint Training”。在这个阶段，选定学生模型中的一个中间层作为“学生隐藏层”，并从教师模型中选定一个相应的中间层作为“hint layer”。同样的，选择学生网络的一个隐藏层称为 guided layer，来学习教师网络的 hint layer。学生隐藏层的目标是学习模仿教师模型的这一层。为了达到这个目的，通常会定义一个损失函数，来衡量学生隐藏层输出与教师层输出之间的差异。通过最小化这个损失，学生模型可以学习到与教师模型相似的中间层表示。

由于教师模型和学生模型的结构可能不同（特别是在它们的中间层），直接比较这两个模型的相应层的特征可能是不可行的。例如，它们可能有不同数量的特征通道或不同的空间维度。为了解决这个问题，FitNet 引入了映射函数，通常是一组可学习的参数（例如，1x1 卷积层），用于将学生模型的特征转换成与教

师模型相同的空间。这样，即使学生和教师模型在结构上有所不同，它们的特征也可以在相同的空间中进行有效比较。

FitNet 通过在中间层级引入知识蒸馏，使得学生模型能够不仅学习从数据到输出标签的映射，还能学习如何通过中间表示层来理解和处理数据。这种方法有助于学生模型更全面地理解数据结构和特征，特别是在学生模型比教师模型结构上要简单得多时。通过这种方式，FitNet 提高了学生模型的性能，使其能够在保持较低复杂度的同时，更接近教师模型的性能。

2.3.3 基于助推器网络的知识蒸馏

Zhou 等人^[47]提出了一个基于助推器网络的知识蒸馏框架，这个框架的目的是在保持响应时间和准确性要求的同时，降低网络模型的复杂度和计算成本。这对于需要实时响应的应用场景，如点击率（CTR）预测模型等，尤为重要。

核心思想是利用一个复杂的助推器网络（booster net）来指导轻量级网络（light net）的训练，其中助推器网络的作用类似于教师网络，而轻量级网络则类似于学生网络。与传统的教师-学生模型不同，轻量级网络和助推器网络在训练过程中是同时训练的，而不是先训练一个已经确定的教师网络然后再训练学生网络。这种同步训练方式可以让轻量级网络更直接地从助推器网络学习，并在整个训练过程中不断调整和优化。

该模型的轻量级网络在某些层上与助推器网络共享参数，这有助于轻量级网络直接从助推器网络获取知识。并且轻量级网络和助推器网络一起训练，而不是分开训练。这样可以缩短总的训练时间，并让轻量级网络更有效地从助推器网络中学习。同时通过最小化提示损失（Hint Loss）促进轻量级网络模仿助推器网络的特征表示。此外，还采用了梯度块（Gradient Block）技术，技术是为了防止助推器网络在训练过程中由于轻量级网络反馈造成的性能下降。在实践中，这通常意味着在反向传播过程中阻止从提示损失到助推器网络特定参数的梯度更新。这确保了助推器网络可以独立于轻量级网络的性能改进来优化其任务。这些特点结合在一起，使得轻量级网络在减少计算成本的同时，仍能保持较高的性能表现。图 2.5 展示了基于联合教学的知识蒸馏框架流程图。

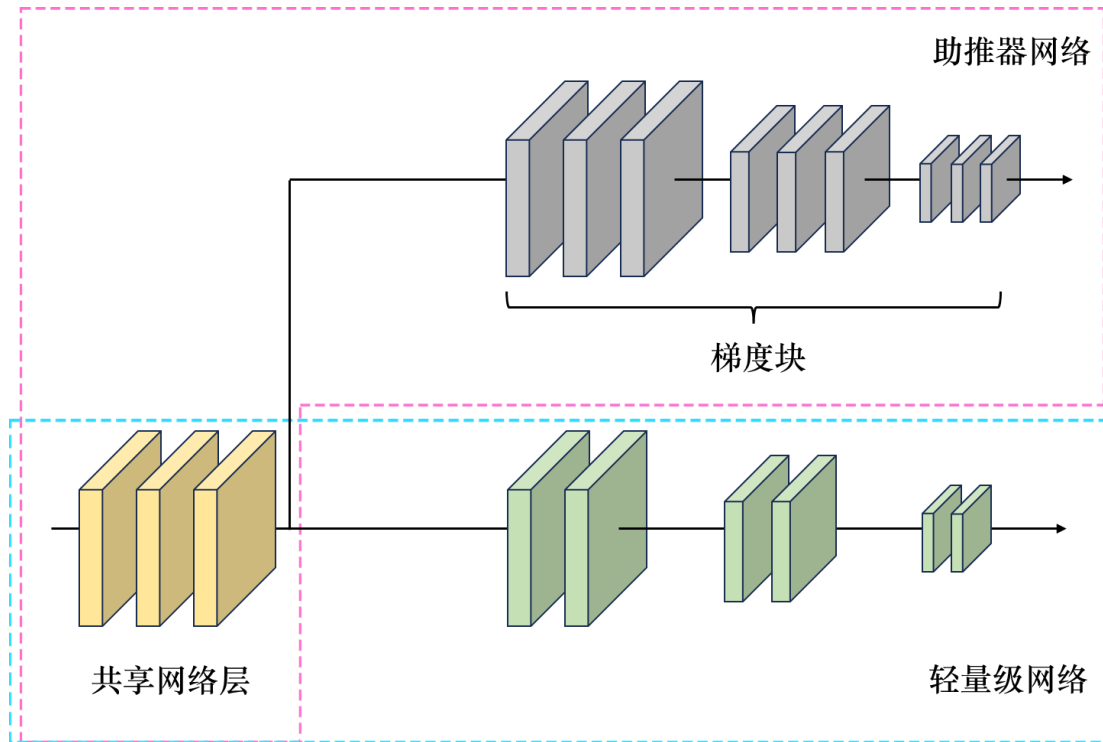


图 2.5 基于联合教学的知识蒸馏框架图

提示损失用于量化轻量级网络和助推器网络之间的知识差异。通常，这通过计算两个网络输出特征之间的均方误差（MSE）来实现。如果假设轻量级网络的某层输出为 f_{light} ，而对应的助推器网络层的输出为 $f_{booster}$ 。式 2.7 展示了提示损失的损失函数公式：

$$L_{hint} = \frac{1}{N} \sum_{i=1}^N \|f_{light}^{(i)} - f_{booster}^{(i)}\|^2 \quad (2.7)$$

其中 N 是批次中样本的数量， $\|\cdot\|^2$ 表示欧式距离的平方，对于一个 N 维空间中的两个点，若 P 和 Q 的坐标可以表示为 $P(p_1, p_2, p_3, \dots, p_N)$ 和 $Q(q_1, q_2, q_3, \dots, q_N)$ ，那么 P 和 Q 的欧式距离的相关公式如式 2.8 所示：

$$d(P, Q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_N - p_N)^2} \quad (2.8)$$

轻量级网络的总损失函数由两部分组成：传统的交叉熵损失和提示损失。交叉熵损失确保模型可以准确预测实际的标签，而提示损失则帮助轻量级网络学习助推器网络的特征表示。总的损失函数可以表示为下式 2.9：

$$L_{total} = \alpha L_{cross-entropy} + (1 - \alpha) L_{hint} \quad (2.9)$$

其中 $L_{cross-entropy}$ 是轻量级网络对真实标签的交叉熵损失, L_{hint} 是提示损失, α 是用来平衡两种损失的权重参数。

通过这个框架,轻量级网络可以在保持较低复杂度的同时,达到与更复杂网络相似的性能。该论文中还详细介绍了在不同数据集上的实验结果,验证了该方法的有效性。

2.3.4 基于空间注意力图的知识蒸馏

在 Zagoruyko 等人^[46]的研究中,空间注意力图被用作一种工具,以改善卷积神经网络(CNN)的性能。通过定义空间注意力,研究表明可以显著提高 CNN 在各种任务上的表现。具体来说,这项技术侧重于通过教师网络的空间注意力图,来指导学生网络学习,从而达到性能提升的目的。图 2.6 展示了教师网络利用空间注意力图来计算注意力蒸馏损失并指导学生网络的过程示意图。

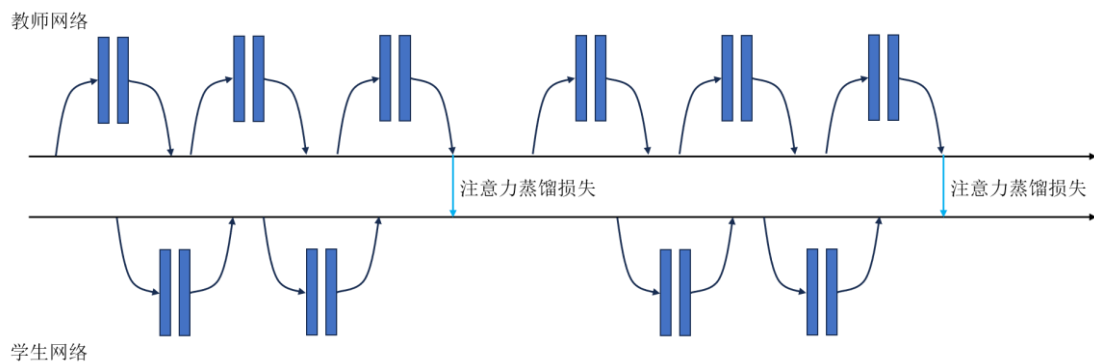


图 2.6 基于空间注意力图的知识蒸馏方法指导过程

空间注意力图体现了网络对输入图像特定区域的关注程度,可以从不同的层级捕捉低级到高级的表示信息。Zagoruyko 等人提出了两种类型的空间注意力图:基于激活的(activation-based)和基于梯度的(gradient-based)。

对于基于激活的空间注意力图,定义如下:给定一个 CNN 层的激活张量 $A \in \mathbb{R}^{C \times H \times W}$,其中 C 、 H 、 W 分别代表通道数、高度和宽度。空间注意力图可以通过不同的方式计算得到,例如绝对值的和(F_{sum})或者绝对值的幂次和($F_{sum,p}$),可以得到一个空间维度上的注意力图,该图反映了网络在做决策时最关注图像的哪些区域。此外,还可以使用绝对值的最大值的幂次(F_{maxp})来获取更加集中的注意力区域。具体的计算方式如下:

(1)基于激活的注意力图求和 F_{sum} : 对于给定的 CNN 层的激活张量 $A \in$

$\mathbf{R}^{C \times H \times W}$ ，该方法通过计算所有通道上激活值的绝对值之和来构建空间注意力图，公式表示为：

$$F_{sum}(A) = \sum_{i=1}^C |A_i| \quad (2.10)$$

其中， A_i 表示激活张量在第*i*个通道上的值。

(2)基于激活的注意力图求和的幂次 $F_{sum,p}$ ：在 F_{sum} 的基础上，引入幂次 p 来进一步强调更高激活值的重要性，公式表示为：

$$F_{sum,p}(A) = \sum_{i=1}^C |A_i|^p \quad (2.11)$$

其中， $p > 1$ ，用以加权高激活值对总和的贡献，使模型更关注于那些激活值高的区域。

(3)基于激活的注意力图最大值的幂次： $F_{max,p}$ 与 $F_{sum,p}$ 类似，但是在每个空间位置上只考虑最大的激活值，公式表示为：

$$F_{max,p}(A) = \max_{i=1}^C (|A_i|^p) \quad (2.12)$$

这种方法使得注意力图仅仅聚焦于最为显著的特征，忽略了其他较小的激活值，可能有助于突出最关键的视觉特征。

通过上述方法获得的空间注意力图，教师网络可以指导学生网络使其注意力分布尽可能接近教师网络，进而提高学生网络的性能。这不仅增强了学生网络对关键特征的捕捉能力，还有助于其在各种数据集和架构上的泛化能力。

此方法的优势在于它直接作用于网络的内部表示，而不仅仅是输出层。这允许学生网络学习到更加细致和复杂的特征表示，进而在各种任务上达到或超越教师网络的性能。通过这种空间注意力的知识蒸馏策略，Zagoruyko 等人的研究不仅展示了其在视觉任务上的有效性，也为未来研究提供了新的方向，即通过内部表示的迁移和优化来提升深度学习模型的性能。

2.4 本章小结

本章首先介绍所使用的主干网络及其相关结构，然后介绍了注意力机制中的空间非局部注意力机制相关原理。随后，本章详述了四种知识蒸馏的理论技术，

在讨论这四种知识蒸馏策略时，分析了各种方法的理论基础，并详细介绍了它们在传递知识过程中采用的方法和公式。

第3章 基于联合教学知识蒸馏与特征相关性的风格分类

3.1 基于联合教学知识蒸馏与特征相关性的模型结构图

本文使用了一种基于融合了双教师知识蒸馏策略与自适应加权格拉姆矩阵模块的图像风格分类方法。自适应加权格拉姆矩阵通过对特征间的相互关系进行自适应权重调整，有效地提取了图像的风格特征。

在知识蒸馏方面，本模型采用了双教师结构，即一个与学生网络同步训练的初始教师网络和一个预训练的专家教师网络。这两个教师网络共同对学生网络进行指导，精炼并传递重要的知识，促进学生网络获得更高效的学习成果。通过这种结构，学生网络能够在两个教师网络的共同辅助下，更加精准地识别和学习复杂的风格特征，从而显著提高了图像风格分类任务的性能。基于联合教学知识蒸馏与特征相关性的风格分类模型整体架构图如图 3.1 所示

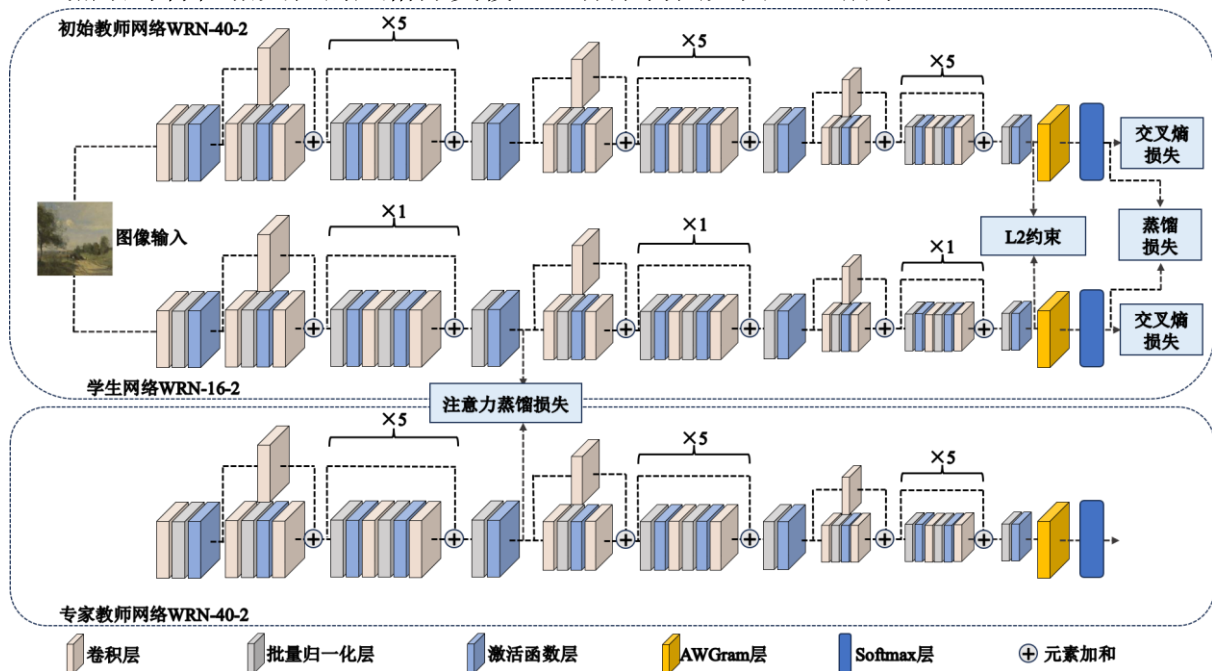


图 3.1 基于联合教学知识蒸馏与特征相关性的风格分类模型整体架构图

3.2 自适应加权格拉姆矩阵介绍

研究视觉艺术图像风格分类是艺术认知领域的一个关键课题，它需要从色调、笔触、布局等多维要素中提炼出艺术作品的风格特征。这一过程不仅加深了对艺术作品视觉属性的理解，而且对于艺术历史的解释、艺术作品的归档和检索以及艺术鉴赏教育均有着重要价值。相比传统的图像分类，艺术风格分类需要捕捉那些超越图像主体不变特性的细微差别，如画作中的情感氛围、时代背景的印记以及艺术家的个人特色。此外，艺术图像风格分类的挑战来自于艺术作品风格的主观性和复杂性。不同的艺术流派和时期，即使在相似的文化和地理背景下，也会展现出独特的风格，这些风格往往难以用传统的计算机算法精确划分。

图 3.2 是对一幅原始艺术画作的不同类型风格的展示，揭示了不同风格对艺术作品的影响。



图 3.2 不同风格对艺术画作产生的影响

在艺术图像风格分类的任务中，格拉姆矩阵(Gram Matrix)是一个重要的工具，用于捕捉和描述图像的风格属性。这是通过使用深度卷积神经网络（例如 VGG 网络）处理图像，并从网络的不同层级提取特征来实现的。这些层级能够捕获图像的多个维度，如纹理、形状和颜色。

对于网络中的每一层，通过将该层的特征图的向量化形式与它的转置相乘来计算 Gram 矩阵。该矩阵的每个元素代表特征间的相关性，从而编码了图像的风格特征。比较不同图像的 Gram 矩阵可以用于它们风格的分类和分析。

在风格迁移中，Gram 矩阵也发挥着核心作用。风格迁移任务的目标是将一

种图像风格应用到另一幅保留内容结构的图像上。这通常通过最小化参照风格图像和目标生成图像的 Gram 矩阵之间的距离来实现。具体而言,这个距离通过计算两个 Gram 矩阵之间的 Frobenius 范数来度量。因此, Gram 矩阵不仅在风格分类中作为一种工具来描述和表征图像中的风格特征,也在风格迁移中作为风格的量化和转移工具。下图 3.3 展示了 Gram 矩阵关于特征图之间的计算过程的示意图:

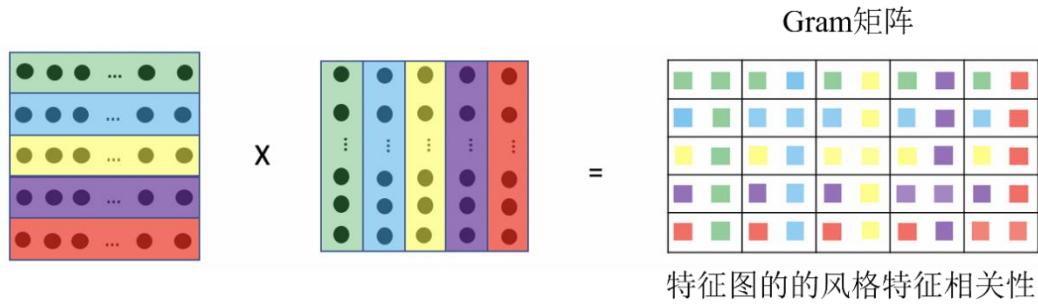


图 3.3 Gram 矩阵对特征图之间的计算过程示意图

总之, Gram 矩阵在风格分类和风格迁移的应用中,提供了一种有效的方式来量化和操作图像的风格特征。式 3.1 展示了 Gram 矩阵的基本定义:

$$Gram_{ij}^{AB} = \sum_{n=1,2,\dots,m} M_{ij}^A(n) M_{ij}^B(n) \quad (3.1)$$

其中 $Gram_{ij}^{AB}$ 表示两个特征图 M_A 和 M_B 之间的内积,具体地,矩阵元素 $M_{ij}^A(n)$ 和 $M_{ij}^B(n)$ 代表在特征图上第 (i, j) 位置上,第 n 个过滤器的激活值,而 m 代表特征图 M_A 和 M_B 的可以进行 Gram 矩阵内积操作的维度。

但公式 3.1 中的 Gram 矩阵存在一定局限性,它假设特征图中所有空间位置的激活值均具有等同的重要性。为了优化此模型,本文引入了一个自适应特征权重矩阵 w_{ij} ,用于调整改进后公式中各空间位置的重要性。在改进的方案中,通过赋予不同空间位置以不同的权重,能够更精细地控制各个位置对于总体特征表示的贡献,从而提高模型的表达能力和灵活性。式 3.2 展示了自适应加权格拉姆矩阵(AWGram 矩阵)的相关公式:

$$AWGram_{ij}^{AB} = \sum_{n=1,2,\dots,m} w_{ij} M_{ij}^A(n) M_{ij}^B(n) \quad (3.2)$$

AWGram 矩阵通过使用一种基于特征相关性的加权机制,提升了 Gram 矩阵提取风格特征的能力,这一方法显著提升了艺术图像风格分类的准确度。下面将

介绍六种常见的特征相关性测量方法，这些特征相关性方法在统计分析和深度学习领域中得到了广泛应用，它们各自基于独特的计算框架和理论基础，对于揭示和理解数据间的复杂关系具有至关重要的作用。

(1) 皮尔森相关性 (Pearson correlation):

定义：皮尔森相关性度量了两个变量之间的线性关系强度和方向。其值范围从-1 到 1，其中 1 表示完全正相关，-1 表示完全负相关，而 0 表示没有线性相关。式 3.3 展示了皮尔森相关性的有关公式：

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]} \quad (3.3)$$

(2) 斯皮尔曼相关性 (Spearman correlation):

定义：斯皮尔曼相关性度量两个变量的等级(排名)之间的相关性，是一种非参数相关性度量。式 3.4 展示了斯皮尔曼相关性的有关公式：

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (3.4)$$

(3) 协方差 (Covariance):

定义：协方差表示了两个变量共同变化的趋势。式 3.5 展示了协方差的有关公式，其中 \bar{x} 和 \bar{y} 分别是 X 和 Y 的样本均值：

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3.5)$$

(4) 切比雪夫距离 (Chebyshev distance):

定义：切比雪夫距离在多维上衡量最大的绝对值差异。在许多情况下，这个度量表明了最坏情况下的相异度。式 3.6 展示了切比雪夫距离的有关公式：

$$D_{cheb}(x, y) = \max_i(|x_i - y_i|) \quad (3.6)$$

(5) 欧几里得距离 (Euclidean distance):

定义：欧几里得距离或欧几里得度量是两点之间“直线”距离，有时简称为欧式距离。在多维空间中，这是从一点到另一点的距离。式 3.7 展示了欧几里得距离的有关公式：

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.7)$$

(6) 余弦相似度 (Cosine similarity):

定义: 余弦相似度度量了两个向量在方向上的相似性, 忽略它们的大小。它计算两个向量的夹角的余弦值。式 3.8 展示了余弦相似度的有关公式, 其中 A, B 一般是一维向量:

$$\text{Cos}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (3.8)$$

本文的 AWGram 矩阵选取上面的六种相关性中的余弦相似度来进行权重矩阵的设计, 具体来说, 在模型训练阶段, 首先计算卷积层中不同特征图的余弦相似度, 首先将不同特征图矩阵展平成一个向量, 然后计算这两个长向量之间的余弦相似度, 接着对这些值进行归一化处理来生成相应的权重 w_{ij} , 余弦相似度原本的范围是 $[-1, 1]$, 转换为一个新的范围 $[0, 1]$ 来将余弦相似度转化为相似度度量。相似度度量越高, 两个实体就被认为越“接近”。下图 3.4 直观展示了利用自适应加权格拉姆 Gram 矩阵来提取风格特征进行风格分类的过程:

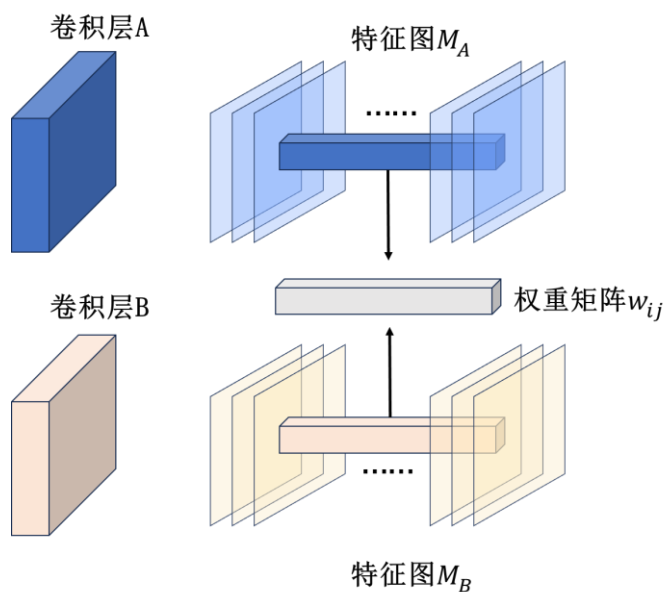


图 3.4 自适应加权 Gram 矩阵提取风格特征的示意图

通过采用这种自适应权重的特征表示方法, 能够更准确地捕获并表达艺术

图像的风格特征。在完成特征提取后,通过 $L2$ 正则化方法对得到的 $AWGram^{AB}$ 进行标准化处理,以保证模型的泛化能力,并通过一个分类层(一般是 softmax 层)来进行图像风格的预测。

3.3 联合教学知识蒸馏模型介绍

艺术风格分类数据集往往体积庞大,高昂的存储和计算成本使得深度神经网络在资源受限的设备上的部署变得更加困难。因此,考虑采用知识蒸馏技术^[41],其主要目标是通过从体量较大的预训练教师模型中提取并转移知识,以训练出一个体积更小、结构更紧凑的学生模型。此方法在缩减模型尺寸的同时,能基本保持甚至提升模型的性能,通过精炼和传递关键信息,实现知识的高效转化和应用。

知识蒸馏技术中,教师网络的不同层次的特征或输出作为要传递给学生网络的知识。通常情况下,较高层的特征更直接相关于执行特定任务的对象部分,而较低层的特征则常常是典型的基础特征,如边缘和角点。因此,可以采用深度神经网络较低部分生成的特征作为中间提示。所有这些特征包含的有价值的“隐知识”,可以转移到指导学生网络的训练过程中。

本文使用并改进了一种联合教学知识蒸馏模型^[53],提升了其在艺术图像风格分类任务中的适用性和效能。该方法利用两种不同的教师网络共同训练学生网络,与学生网络共同训练的教师网络(即初始教师)通过其临时输出逐步辅助学生网络,使学生网络能够朝着高准确度的最终逻辑(logits)接近最优路径。另一位预训练好的教师网络(即专家教师)则指导学生网络关注对任务更有用的关键区域。通过结合这两位教师的知识,可以显著提高学生网络在知识蒸馏中的性能。

本文使用 x 和 y 分别代表深度神经网络(DNN)的输入和相应的标签。初始教师网络的 softmax 输出记作 $P_{T1}=\text{softmax}(z_{T1})$,这里 P_{T1} 是通过将 softmax 函数应用于未归一化的对数概率值 z_T 上得到的。同样地,学生网络接收同一图像作为输入,以获取其预测 $P_S=\text{softmax}(z_S)$ 。在这里, z_{T1} 和 z_S 分别是教师网络和学生网络的未归一化的对数概率值,也被称为网络的“logits”。简单来说,它们是深度神经网络在应用 softmax 函数之前的最后一层输出。

用 P_{T1}^T 和 P_S^T 表示经过温度调整的初始教师网络和学生网络的 softmax 输出。

P_{T1}^T 的公式如式 3.9 所示, P_S^T 的公式具有相似的形式:

$$P_{T1}^T(i) = \frac{e^{\frac{z_{T1i}}{T}}}{\sum_j e^{\frac{z_{T1j}}{T}}} \quad (3.9)$$

其中, 温度参数 T 用于控制 softmax 函数的平滑程度。较高的 T 值会使输出的概率分布更加平滑, 这有助于传递教师网络的“隐知识”到学生网络。 z_{T1i} 表示初始教师网络输出对于类别 i 的 logit 值, 分母是对所有可能类别的同样处理的 logit 值进行归一化的结果, 确保了输出是一个概率分布。

通过这种方式, 可以获得一个相对于原始输出更加平滑的概率分布, 这有助于学生网络学习到初始教师网络的深层次和细微的特征。

学生模型输出与真实标签之间的损失, 通常不需要通过温度参数 T 进行软化。这部分损失旨在直接优化学生模型的性能, 使其尽可能准确地预测正确的标签, 因此使用标准的交叉熵损失函数即可, 无需使用温度参数 T 进行软化处理。交叉熵损失的计算公式如式 3.10 所示:

$$L_{CE} = - \sum_{i=1}^N p(x_i) \log(q(x_i)) \quad (3.10)$$

其中 $p(x_i)$ 表示真实分布, $q(x_i)$ 表示预测分布。

针对艺术图像风格分类任务的复杂性和特殊性, 可以对原有的联合教学知识蒸馏方法进行了一定的优化。原方法主要通过交叉熵损失来评估学生网络与初始教师网络输出之间的差异。然而, 艺术风格的识别通常需要对图像中微妙差异的敏感捕获, 这一过程要求学生网络能够深入理解并精确模仿教师网络对不同艺术风格的识别。为了应对这一挑战, 采用了根据温度调节的 KL 散度损失函数, 并通过引入系数 T^2 进行了改进, 从而增强学生网络学习教师网络知识的能力。

通过引入 KL 散度作为损失函数, 本文的方法能够更精细地衡量学生网络与教师网络输出概率分布之间的差异。与简单的交叉熵损失相比, KL 散度为学生网络提供了更为详尽的指导, 使其能够更准确地复制教师网络的预测行为。这在艺术风格分类的任务中尤为关键, 因为区分不同艺术风格往往依赖于对图像细节的精确解读。

此外, 通过在 KL 散度前乘以 T^2 , 对由于温度提高所导致的梯度缩放问题进行了有效的补偿, 确保了整个学习过程的稳定性和效率。这种调整策略不仅保留

了通过温度调节带来的优势，同时也避免了潜在的学习不稳定问题，确保学生网络能够从教师网络中高效、稳定地获取并利用知识。

当对学生网络 S 和初始教师网络 $T1$ 进行联合训练时，损失函数如式 3.11 所示。该损失函数的第一部分 $L_{CE}(y_{true}, P_S) + L_{CE}(y_{true}, P_{T1})$ 表示的是学生网络和初始教师网络的预测输出和实际标签之间的交叉熵损失，其中 y_{true} 是实际的标签分布， P_S 和 P_{T1} 是学生网络和初始教师网络的 softmax 输出。第二部分 $KL(P_{T1}^T, P_S^T)$ 表示初始教师网络 $T1$ 和学生网络 S 输出的软化概率分布之间的蒸馏损失，该蒸馏损失使用 KL 散度计算。

$$L_{KD}(S, T1) = L_{CE}(y_{true}, P_S) + L_{CE}(y_{true}, P_{T1}) + T^2 KL(P_{T1}^T \| P_S^T) \quad (3.11)$$

其中 T^2 在蒸馏项前用作调整因子，确保当调整温度时，损失的相对大小保持一致。

然而，在初始教师网络和学生网络共同训练的过程中，随机初始化的初始教师网络可能会对学生网络提供不正确的指导。另一个问题是，在没有引入中间层约束的情况下，学生网络难以模仿教师网络复杂的决策边界和数据表示，利用知识蒸馏训练更深的学生网络可能会遇到困难。

首先通过引入 L2 约束^[60]，该约束在训练过程中持续对学生网络进行约束，并且仅影响学生网络的反向传播过程，防止学生网络过度模仿教师网络，确保学生网络能够在保留初始教师网络有用信息的同时，还能保持一定的自主学习能力，这种策略有助于保持学生网络的独立性和灵活性。L2 约束的公式如式 3.12 所示：

$$L_2(S, T2) = \|z_S - z_{T2}\|_2^2 \quad (3.12)$$

然后通过引入空间注意力图^[46]作为中间层约束，该约束指导学生网络在训练过程中效仿专家教师网络的空间关注模式，使学生网络可以学习到如何以类似于教师网络的方式处理和理解输入数据。这种约束不仅提高了学生网络对数据深层次理解的能力，还强化了其对关键特征的识别和响应，从而促进了学生模型性能的整体提升。图 3.5 展示了教师网络通过注意力图指导学生网络的过程：

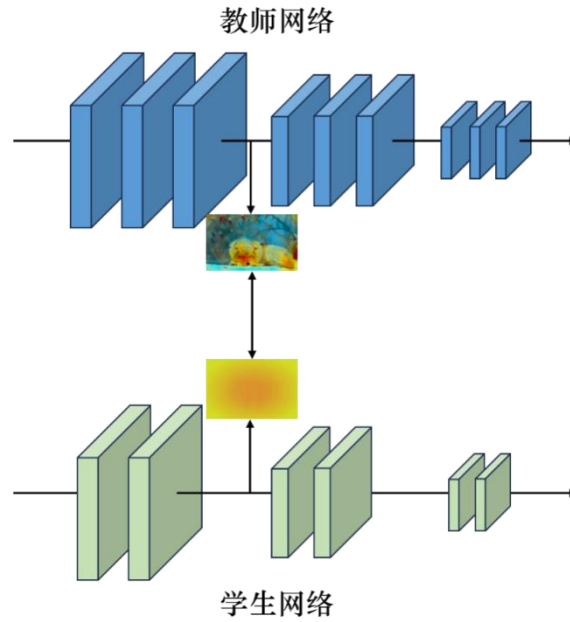


图 3.5 教师网络利用注意力图指导学生网络的示意图

空间注意力图是通过分析卷积层的激活值来生成的。对于给定的卷积层，其激活张量 $A \in \mathbb{R}^{C \times H \times W}$ ，然后定义了一个激活映射函数 F ，该函数将激活张量映射到一个二维的空间注意力图 $F(A) \in \mathbb{R}^{H \times W}$ 上。空间注意力图反映了网络在各个空间位置的关注度，从而为知识的传递提供了依据。

对于不同的任务，选用合适的注意力图激活函数是至关重要的。针对本文研究的艺术图像风格分类任务，本文所使用的空间注意力图激活函数为 $F_{sum,2}(A)$ ，其计算方式为对所有通道的激活值取绝对值并平方后再进行求和，这种方法有效强调了对艺术风格识别关键的视觉特征，通过加权显著特征的方式，增强了模型对风格细节的敏感性和泛化能力，从而在精确捕捉艺术风格的细微差异方面表现出优越性，提升了分类准确率。激活函数 $F_{sum,2}$ 的公式如式 3.13 所示：

$$F_{sum,2}(A) = \sum_{i=1}^C |A_i|^2 \quad (3.13)$$

其中 A_i 是激活张量 A 的第 i 个通道， C 表示总通道数。

该约束方法的关键在于使学生网络生成的注意力图与教师网络的尽可能相似。这一步骤通过计算学生网络 S 和专家教师网络 $T2$ 的注意力图之间的差异，并将这个差异作为额外的损失来实现，促使学生网络在学习过程中也学会关注专家教师网络所关注的区域。具体的约束方式如下所式 3.14 所示：

$$L_{AT}(S, T2) = \sum_{i=1}^N \left\| \frac{F(A_S^j)}{\|F(A_S^j)\|_2} - \frac{F(A_{T2}^j)}{\|F(A_{T2}^j)\|_2} \right\|_2 \quad (3.14)$$

其中, $\|\cdot\|_2$ 是 L2 范数, 用于计算特征图的注意力图的欧几里得长度, $F(A_S^j)$ 和 $F(A_{T1}^j)$ 是学生网络和专家教师网络 $T2$ 的空间注意力图。除以各自的 L2 范数的作用是用于规范化注意力图, 确保比较是在同一尺度上进行。

综上所述, 联合教学知识蒸馏的总损失函数如式 3.15 所示:

$$L_{total}(S, T1, T2) = L_{KD}(S, T1) + \alpha L_2(S, T2) + \beta L_{AT}(S, T2) \quad (3.15)$$

其中 α 和 β 是平衡因子, 用于平衡不同损失项对总体训练目标的贡献。

该损失函数不仅促进了学生网络在直接任务性能上的优化, 也通过引入专家教师网络和初始教师网络的辅助监督, 实现了知识的深层次转移。

3.4 本章小结

本章节提供了一个完整的框架介绍, 详细解析了两项关键技术——基于联合教学的知识蒸馏和自适应加权格拉姆模块的理论基础和实施方法。本章首先探讨了利用 Gram 矩阵加权内积来提取图像风格特征的 AWGram 矩阵方法。其次, 本章描述了通过融合双教师网络的知识的策略来优化学生网络学习的知识蒸馏方法, 详细解释了其涉及的原理及其应用的具体公式。

第4章 实验设置及结果分析

4.1 数据集

在本节中将详细介绍本研究所采用的数据集，并探讨其在风格分类及知识蒸馏实验中的应用和意义。本文选取了三个独特的数据集：Wiki Art、MultitaskPainting100k 以及 Pandora18k，这些数据集在艺术领域内各具特色，涵盖了从古典到现代的广泛风格，以及不同的艺术流派和画家作品。通过详细了解这些数据集的基本情况，提供了一个清晰的实验背景，更好地理解本文的研究目的和实验设计。

4.1.1 Wiki Art 数据集

第一个数据集是大规模的 Wiki Art^[75]，由非营利性绘画艺术网站 Wiki Art 的全球志愿者共享和标注。随着时间的推移，网站画作数量增多，不同研究因采用不同筛选方法导致数据差异。本文采用了 Tan 等人^[76]整理的，收集自 Wiki Art 的 81444 幅作品的数据集，拥有 27 种艺术风格。下图 4.1 展示了 Wiki Art 数据集的部分图像。

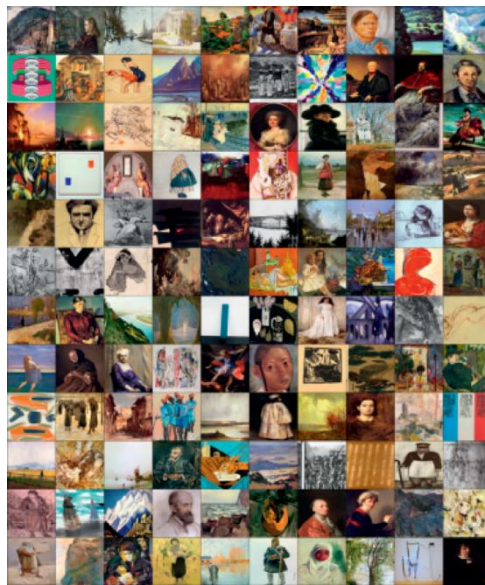


图 4.1 Wiki art 数据集

4.1.2 MultitaskPainting100k 数据集

MultitaskPainting100k(<https://www.payititi.com/opendatasets/show-26603.html>), 它源自于 Kaggle 竞赛 “Painter by Numbers”, 主要用于预测一对图像是否出自同一艺术家之手。该数据集包含 99816 幅图像, 覆盖了 125 种不同的艺术风格。这种方法通过增加多样性来提升模型的泛化能力, 并为解决复杂的分类问题提供了丰富的实验条件。下图 4.2 展示了 MultitaskPainting100k 数据集中的部分图像。



图 4.2 MultitaskPainting100k 数据集

4.1.3 Pandora18k 数据集

Pandora18k (http://imag.pub.ro/pandora/Download/Pandora_18k.zip), 是一个艺术风格分类数据集, 包含了由艺术专家严格标注的 18 种风格和 19,320 张图像。它主要用于分类和识别不同的艺术运动。下图 4.3 展示了 Pandora18k 数据集中的部分图像。

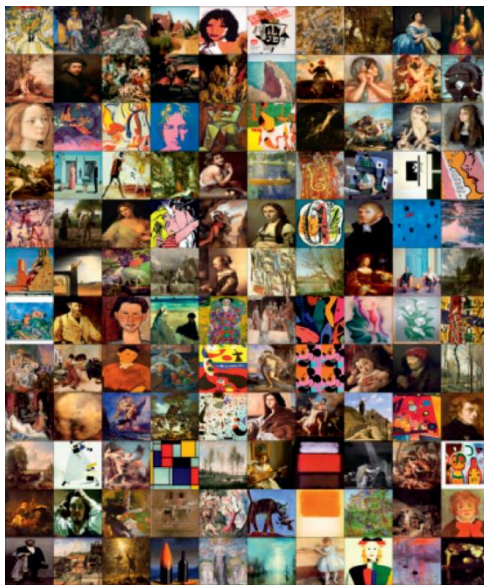


图 4.3 Pandora 18k 数据集

表 4.1 展示了按艺术风格分类的图像数据集及其属性信息，其中包括风格类别数量、图片总量、训练集数量和测试集数量。

表 4.1 艺术风格图像数据集及其属性信息

数据集	风格类别数量	图片数量	训练集数量	测试集数量	验证集数量
Wiki art	27	81444	57022	12211	12211
MultitaskPainting100k	125	99816	69820	14998	14998
Pandora 18k	18	19320	13524	2898	2898

4.2 实验环境及参数设置

4.2.1 评价指标

在艺术图像风格分类实验中，本文采用总体准确率作为主要评价指标。总体准确率衡量模型在整个测试集上正确分类图像的比例，直接反映了模型的整体性能。

4.2.2 实验环境

本文的实验采用 Python3.6 作为开发语言，并在 Pytorch1.10.2 框架下进行，该框架配置了 CUDA11.2 和 cuDNN8.6.0 以充分利用硬件加速。硬件环境包括一颗 Intel(R) i7-10700K CPU @ 3.80GHz、32GB 的内存，以及一块配备 12GB 显存的 NVIDIA GeForce RTX 3080ti GPU，确保了数据处理和模型训练的高效率。本文所用的实验环境配置见表 4.2 所示。

表 4.2 实验环境配置清单

配置名称	配置说明
开发语言	Python3.6
深度学习框架	Pytorch 框架
框架环境配置	CUDA11.2、cuDNN8.6.0、Pytorch1.10.2
处理器	Intel(R) i7-10700K CPU @ 3.80GHz
内存	32GB
显卡 GPU	NVIDIA GeForce RTX 3080ti
显存	12GB

4.2.3 实验参数设置

本文实验使用大小为 128 的小批量进行训练。此外，在每个小批量之前使用水平翻转和随机裁剪进行数据增强。学习率从 0.001 开始，在第 60、120 轮进行衰减，衰减系数为 0.2，使用随机梯度下降（SGD），动量固定在 0.9，训练 200 轮。下表 4.3 展示了本节所进行的实验所使用的参数。

表 4.3 实验参数设置

参数名称	参数值
学习率	0.01
Batch size	128
Epoch	200
优化器	SGD
momentum	0.9

4.3 实验结果分析

本文使用了一种改进的知识蒸馏模型，该模型通过结合“初始教师”和“专家教师”两种教师模型的联合教学，实现了对学生网络的有效指导。初始教师使学生模型能够逐步学习，而专家教师提供关于任务关键区域的深入指导。此外，通过引入自适应加权格拉姆矩阵，并将其融入知识蒸馏框架，目的是通过更准确地捕获和理解艺术风格特征，提升风格分类的准确性。

4.3.1 超参数 α 、 β 对网络性能的影响

本文所提出联合教学知识蒸馏深度学习网络在训练过程中，采用了一个多部分构成的损失函数，总损失函数如式 3.15 所示，其中超参数 α 、 β (beta)被用来平衡总损失函数中 L2 损失和注意力蒸馏损失的权重。本小节的实验探究了不同的 α 、 β 设置如何影响所提网络模型的性能，初始教师网络和专家教师网络选取 WRN-40-2，学生网络选取 WRN-16-2。表 4.4 展示了不同的超参数 α 、 β 组合在 MultitaskPainting100k 数据集上的性能变化：

表 4.4 MultitaskPainting100k 数据集上基于 α 和 β 超参数优化的模型准确率（%）比较

$\alpha \backslash \beta$	0.5	1.0	1.5	2.0	2.5
0.0	64.75	65.47	66.16	65.73	64.92
0.1	64.67	65.43	66.33	65.66	64.88
0.2	64.97	65.73	66.83	65.84	64.68
0.3	64.82	65.75	66.36	65.67	64.66
0.4	64.98	65.47	66.04	65.75	64.96

从表 4.4 可知，超参数 α 和 β 分别设置为 1.5 和 0.2 时模型的整体性能最好。

当 α 值设定为 1.5 时，模型在主分类损失与 L2 损失之间寻求平衡。这种平衡有助于确保模型既不会过度依赖于主分类损失，也不会完全被 L2 损失主导，从而促使学生网络模型对初始教师进行有效学习。

同时, β 值被设置为 0.2, 这意味着注意力蒸馏损失在整个损失函数中占据较小的比重。较低的 β 值指示模型应更多关注于主分类损失和 L2 损失, 而非过分强调注意力蒸馏损失。这有利于防止模型去过度学习特定的、可能不具代表性的特征, 同时也避免了专家教师横向分支的过度优化。

总的来说, 当超参数 α 和 β 分别设置为 1.5 和 0.2 时, 有利于在模型的主要学习任务 and 辅助学习任务之间建立一个有效的平衡。这种平衡有助于模型更全面地学习图像特性, 同时促进主干网络和横向分支的相互补充, 从而提高模型在处理复杂数据时的整体准确率和泛化能力。合理的超参数设置有助于提升模型性能, 实现对不同特征的有效学习和提取。

4.3.2 知识蒸馏温度相关实验

本节对知识蒸馏模块的蒸馏温度进行相关实验。初始教师网络和专家教师网络选取 WRN-40-2, 学生网络选取 WRN-16-2, 选取不同的温度参数 T , 分析学生网络在 Pandora18k、WikiArt 和 MultitaskPainting100k 数据集上的性能变化, 实验结果如图 4.4、图 4.5 和图 4.6 所示。

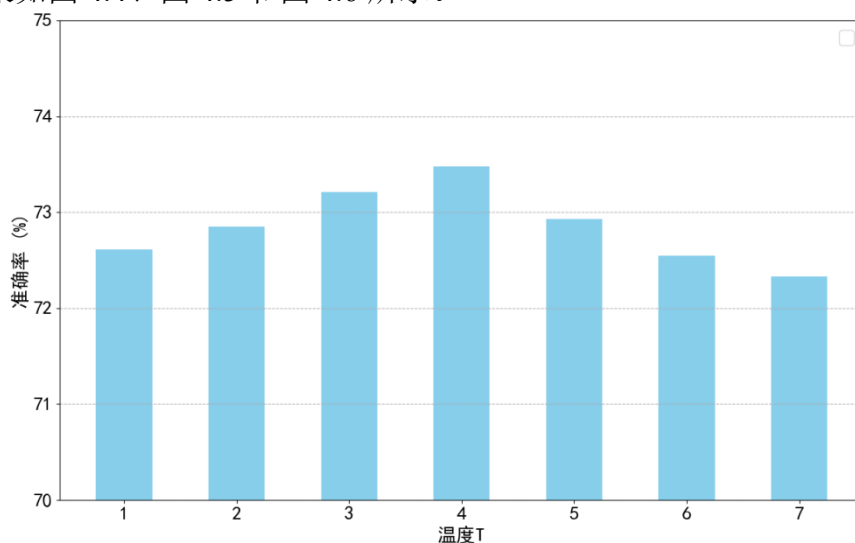


图 4.4 Pandora18k 数据集上不同蒸馏温度的学生网络性能

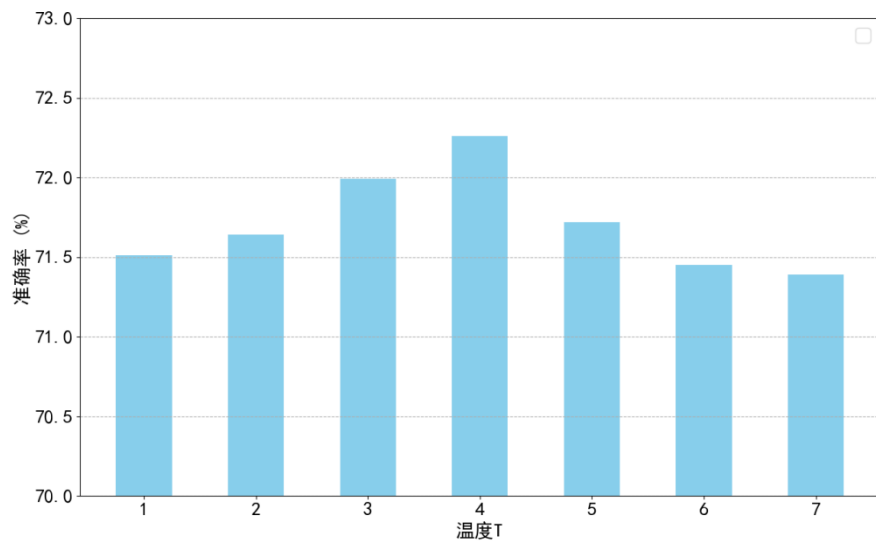


图 4.5 WikiArt 数据集上不同蒸馏温度的学生网络性能

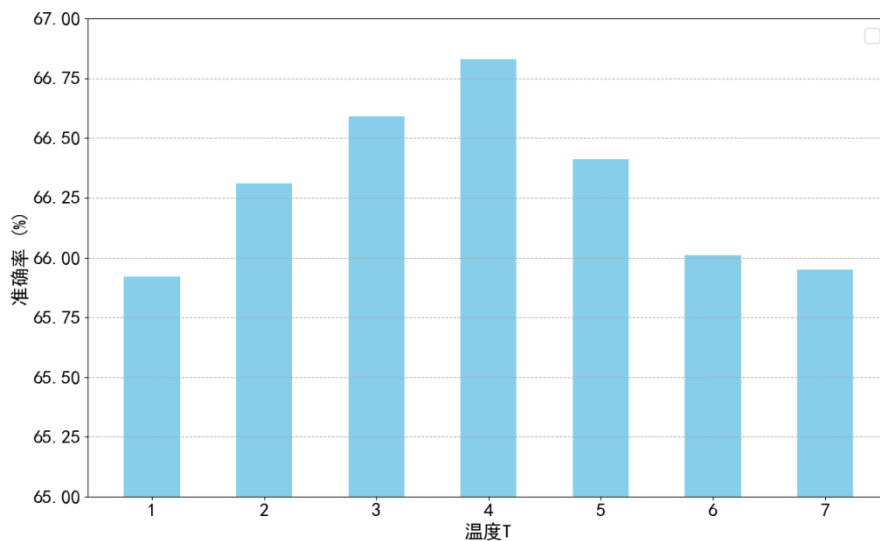


图 4.6 MultitaskPainting100k 数据集上不同蒸馏温度的学生网络性能

在本小节的实验中，研究了蒸馏温度 τ 在知识蒸馏过程中的作用，特别是它如何影响教师网络（WRN-40-2）向学生网络（WRN-16-2）传递知识的效率。蒸馏温度是调节教师网络输出概率分布平滑度的关键参数，它在决定学生网络能否有效捕获并模仿教师网络的复杂决策边界方面扮演着至关重要的角色。

通过在多个数据集上（Pandora18k、WikiArt 和 MultitaskPainting100k）进行实验，可以观察到在蒸馏温度为 4 时，学生网络表现出了最佳的性能。这一现象可以从两个方面进行理解。首先，适中的蒸馏温度有助于产生既不过于平滑也不过于尖锐的概率分布，使学生网络能够更好地理解和模仿教师网络的行为。过低

的蒸馏温度可能导致概率分布过于尖锐，限制了学生网络学习更为细微差别的能力；而过高的温度则可能使得概率分布过于平滑，减少了类别间的区分度，影响学生网络的准确性。其次，选择蒸馏温度为4，可能与数据集的特性和学生网络的容量相匹配，使得学生网络能够在模仿教师网络的同时，还能保持一定的泛化能力。

综上所述，选择蒸馏温度为4的目的是为了平衡概率分布的平滑度与学生网络的学习效率。实验结果强调了在进行知识蒸馏时，适当选择蒸馏温度的重要性，以及它对于最终模型性能的影响。

4.3.3 不同空间注意力图效能对比实验

本实验评估了知识蒸馏中使用不同注意力图转移方法对学生模型性能的影响。通过引入预训练的专家教师网络和初始教师网络（Wide ResNet-40-2）对学生网络（Wide ResNet-16-2）的训练进行指导，可以分析激活基（ F_{sum} 、 $F_{sum,2}$ 、 $F_{sum,4}$ ）和梯度基（ $F_{max,2}$ ）等四种注意力图映射函数在 MultitaskPainting100k 数据集上的性能变化。这四种注意力图映射函数的公式如公式 2.10、公式 2.11 和公式 2.12 所示。表 4.5 展示了不同注意力图映射函数在 MultitaskPainting100k 数据集上的风格分类准确率。

表 4.5 不同注意力图映射函数在 MultitaskPainting100k 数据集上的风格分类准确率

注意力图映射函数	准确率(%)
F_{sum}	65.98
$F_{sum,2}$	66.83
$F_{sum,4}$	66.34
$F_{max,2}$	66.61

通过对实验结果的分析，可以观察到 $F_{sum,2}$ 映射函数在 MultitaskPainting100k 数据集上表现出较高的准确率。这表明该方法能够更有效地促进学生模型从教师模型中学习到重要的特征，并将其应用于图像识别任务中。对于图像风格分类任务而言，这种注意力图方法具有明显的优势。通过引导学生模型关注教师模型在识别不同风格方面的关键区域， $F_{sum,2}$ 映射函数有助于提高学生模型在艺术品风格分类等任务中的性能。在实际应用中，图像风格的识别是一个具有挑战性的

任务, 因为不同的艺术家和时期可能会产生截然不同的风格特征。通过利用注意力图转移方法, 特别是采用了 $F_{sum,2}$ 映射函数, 能够更精确地捕捉到图像中与风格相关的特征。图 4.7 展示了数据集中部分样本基于不同空间注意力图的效果展示。

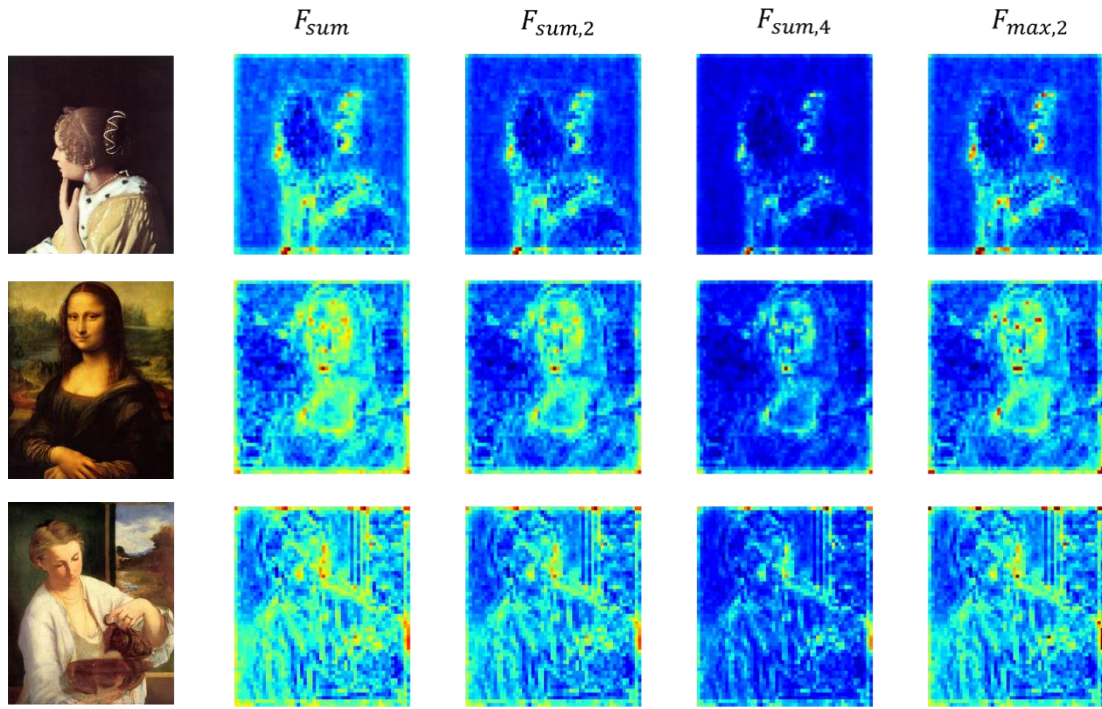


图 4.7 部分样本的不同空间注意力图效果展示

4.3.4 不同相关性权重矩阵的效果评估实验

在本小节的实验中利用相关性度量来形成自适应特征相关性 AWGram 矩阵, AWGram 的公式为 $AWGram^{AB} = w \odot M_A * M_B^T$ 。其中 w 是由相关性度量所构成的权值矩阵。

例如, 若使用特征图 M_A 和 M_B 之间的余弦相似度来定义权值 w , 首先令 $w = \text{Cos}(A, B) = A \cdot B / \|A\| \cdot \|B\|$, 此过程包括将每个特征图从其原始的多维形态 (通常是三维: 通道、高度、宽度) 降维至一维向量。这一步骤使得能够将图像特征转化为标准的线性格式, 从而有效地计算余弦相似度。在此, 引入了“ \odot ”符号, 表示两个矩阵的乘法操作, 即哈达马积 (逐元素乘积)。通过这种方式, 构建了基于余弦相似度的 AWGram 矩阵。其余相关性构成 AWGram 矩阵的方式是

类似的。

在本节中进行的所有实验中，采用了宽残差网络 WRN-40-2 作为初始教师网络和专家教师网络，采用宽残差网络 WRN-16-2 作为学生网络，以学生网络的准确率性能进行评估。表 4.6、表 4.7 以及表 4.8 展示了基于不同相关性度量构建的 AWGram 矩阵在 Pandora18k、WikiArt 以及 MultitaskPainting100k 数据集上的性能表现。其中用“相关性名称+AWGram”来表示基于对应的相关性所构成的 AWGram 矩阵方法。

表 4.6 不同的相关性所构成的 AWGram 矩阵在 Pandora18k 数据集上的准确率

不同的相关性所构成的 AWGram 矩阵	准确率(%)
皮尔森相关性 AWGram	65.84
斯皮尔曼相关性 AWGram	67.61
协方差 AWGram	69.04
切比雪夫距离 AWGram	70.23
欧几里得距离 AWGram	71.75
余弦相似度 AWGram	73.48

表 4.7 不同的相关性所构成的 AWGram 矩阵在 WikiArt 数据集上的准确率

不同的相关性所构成的 AWGram 矩阵	准确率(%)
皮尔森相关性 AWGram	63.71
斯皮尔曼相关性 AWGram	65.29
协方差 AWGram	67.69
切比雪夫距离 AWGram	68.19
欧几里得距离 AWGram	70.03
余弦相似度 AWGram	72.26

表 4.8 不同的相关性所构成的 AWGram 矩阵在 MultitaskPainting100k 数据集上的准确率

不同的相关性所构成的 AWGram 矩阵	准确率(%)
-------------------------	--------

皮尔森相关性 AWGram	59.31
斯皮尔曼相关性 AWGram	60.95
协方差 AWGram	62.38
切比雪夫距离 AWGram	63.77
欧几里得距离 AWGram	65.11
余弦相似度 AWGram	66.83

从表 4.6、表 4.7 和表 4.8 的实验结果可以看出在 Pandora18k、WikiArt 以及 MultitaskPainting100k 三个数据集上, 基于余弦相似度的 AWGram 矩阵均表现出了最高的准确率, 分别为 73.48%, 72.26%, 和 66.83%。这表明余弦相似度作为权值矩阵 w 的计算方法, 对于提升图像识别任务的准确率具有显著效果。

这些实验结果指出, 选择合适的相关性度量对于构建有效的 AWGram 矩阵至关重要。余弦相似度通过考虑特征向量的方向而非其大小, 能够更准确地捕捉特征之间的关联性, 从而提高图像识别的准确性。相反, 基于大小或分布的相关性度量, 如欧几里得距离或切比雪夫距离相关性, 虽然能够提供有用的信息, 但在捕捉深层特征的细微差异方面可能不如余弦相似度敏感。通过选择合适的基于特征图之间相关性的度量方法, 可以显著提高基于深度学习模型的图像识别任务的性能。

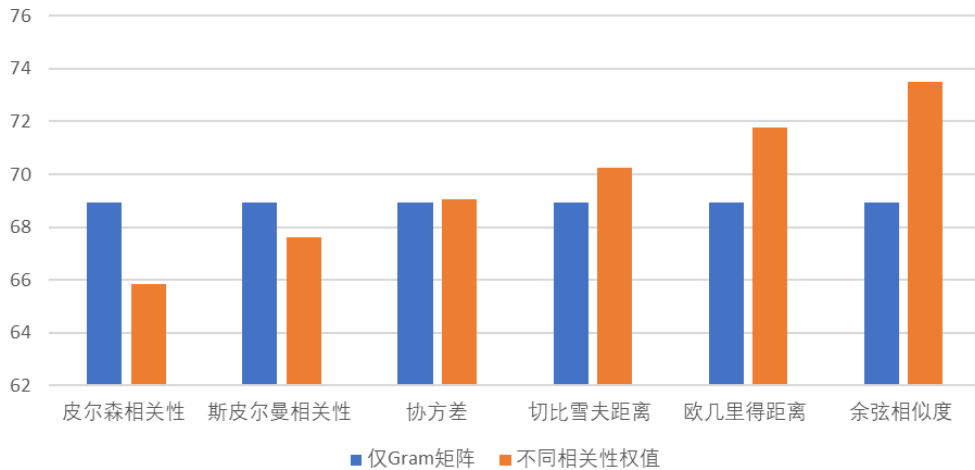


图 4.8 不同相关性 AWGram 与标准 Gram 方法在 Pandora18k 数据集上的效果对比

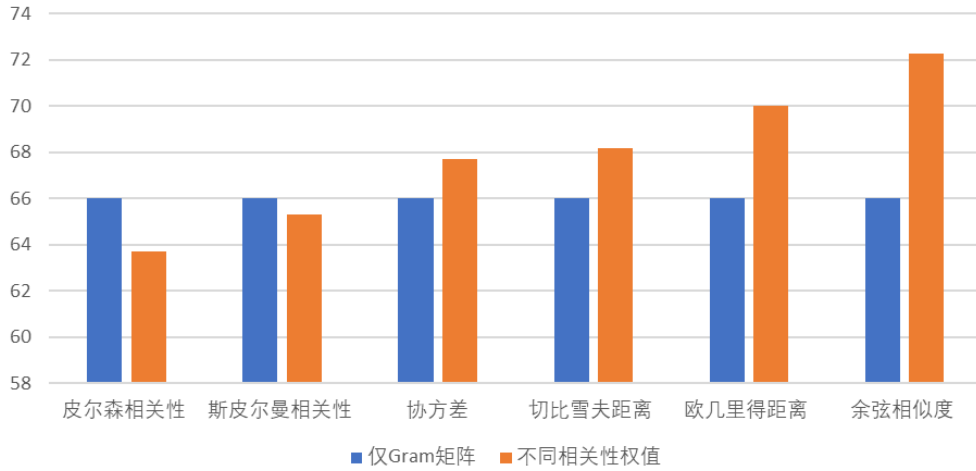


图 4.9 不同相关性 AWGram 与标准 Gram 方法在 WikiArt 数据集上的效果对比

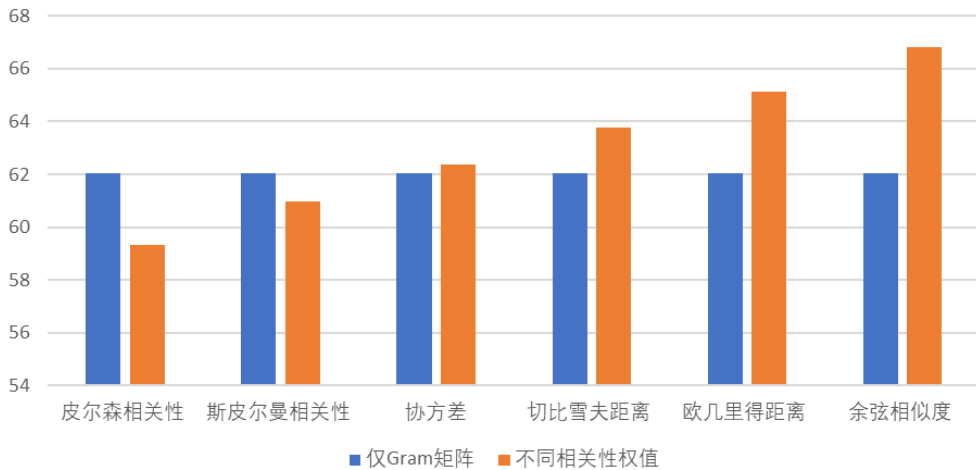


图 4.10 不同相关性 AWGram 与标准 Gram 方法在 MultitaskPainting100k 数据集的效果对比

在图 4.8、图 4.9 和图 4.10 中可以看出本实验将不同相关性 AWGram 与标准 Gram 矩阵方法进行了对比，从图中可以更加直观地观察到，基于皮尔森相关性和斯皮尔曼相关性的 AWGram 矩阵相比于仅使用标准 Gram 矩阵的情况，其准确率表现出了一定程度的下降。这说明了在特定的艺术风格分类任务中，将相关性度量融入特征表示可能并非总是能够提升模型性能的有效策略。

具体而言，皮尔森相关性度量的使用假设了特征间存在线性关系，其计算依赖于变量的均值及标准差。然而，在艺术作品的图像特征与风格属性之间，存在的关系往往是非线性、复杂及高度抽象的。这一特性使得皮尔森相关性度量难以准确捕捉特征间的实际相互作用，进而可能导致分类性能的降低。

斯皮尔曼相关性度量，则是基于变量的等级（或排序）之间的相关性，而非其直接的数值大小，使其对于非参数和非线性关系表现出更好的鲁棒性。尽管如此，斯皮尔曼相关性度量主要适用于单调关系的分析，而艺术风格分类中的图像特征关系可能远比简单的单调变化更加复杂，包括非单调的、多样化的模式和结构。因此，斯皮尔曼相关性度量在这一应用背景下同样存在局限。

值得注意的是，实验结果也表明，采用余弦相似度作为相关性度量构建的 AWGram 矩阵，在三个数据集上均显示出较其他度量方法更高的准确率。这一发现突显了余弦相似度在捕捉艺术图像风格特征复杂关系方面的优势，相较于基于线性假设或依赖等级相关性的其他度量方法，余弦相似度为艺术风格分类任务提供了更为精确和有效的特征相似度度量。

综上所述，本研究的分析强调了在选择特定问题的特征相似度度量方法时，必须考虑到数据的特性和任务的具体需求。虽然皮尔森和斯皮尔曼相关性度量在众多领域中已被证明有效，但它们在处理复杂特征关系的艺术风格分类任务中的局限性亦不可忽视。此外，本研究的发现证实了利用余弦相似度构成自适应加权格拉姆矩阵在艺术风格分类任务中的有效性。

4.3.5 不同宽度因子的知识蒸馏方法性能比较

在本小节的实验中，采用了不同宽度因子的 WRN-40 网络（如 WRN-40-1、WRN-40-2 以及 WRN-40-4）作为初始教师网络和专家教师网络，并采用不同宽度因子的 WRN-16 网络（如 WRN-16-1、WRN-16-2 以及 WRN-16-4）作为学生网络。宽度因子的变化直接导致网络的参数量随之增加，例如，当宽度因子从 1 增加到 4 时，网络的参数量会显著增长，这种参数量的增加，实际上提供了一个探索网络在不同参数规模下准确率表现的机会。通过这种方式，能够更深入地理解宽度因子对网络性能的影响，并评估在不同参数量下网络学习能力的变化。初始教师网络和学生网络是随机初始化的，专家教师网络是预训练的。本实验的关键在于比较本文提出的基于联合教学知识蒸馏与特征相关性的方法与现有部分知识蒸馏方法在 MultitaskPainting100k 数据集上的表现，同时还关注了不同宽度因子对于这些知识蒸馏方法性能的影响。表 4.9 展示了不同宽度因子的 WRN 模型的参数量。

表 4.9 不同宽度因子的 WRN 模型的参数量

模型	参数量(单位: 百万)
WRN-16-1	0.17
WRN-16-2	0.69
WRN-16-4	2.79
WRN-40-1	0.56
WRN-40-2	2.20
WRN-40-4	8.90

表 4.10 宽度因子为 1 的 WRN 模型上不同知识蒸馏方法性能对比

方法	基准模型	准确率(%)
KD ^[41]	WRN-16-1	60.14
ATKD ^[46]	WRN-16-1	61.31
RLKD ^[47]	WRN-16-1	62.87
CTKD ^[53]	WRN-16-1	64.41
本方法	WRN-16-1	64.92
初始教师网络	WRN-40-1	65.88

表 4.11 宽度因子为 2 的 WRN 模型上不同知识蒸馏方法性能对比

方法	基准模型	准确率(%)
KD	WRN-16-2	62.04
ATKD	WRN-16-2	63.21
RLKD	WRN-16-2	64.58
CTKD	WRN-16-2	65.25
本方法	WRN-16-2	66.83
初始教师网络	WRN-40-2	67.57

表 4.12 宽度因子为 4 的 WRN 模型上不同知识蒸馏方法性能对比

方法	基准模型	准确率(%)
KD	WRN-16-4	62.28
ATKD	WRN-16-4	63.47
RLKD	WRN-16-4	64.88

CTKD	WRN-16-4	66.37
本方法	WRN-16-4	66.95
初始教师网络	WRN-40-4	68.20

从表 4.10、表 4.11 和表 4.12 的实验数据中可以看到不同知识蒸馏 (KD) 技术在宽残差网络 (WRN) 上的应用效果, 并比较了不同知识蒸馏策略对模型性能的影响。通过在 WRN-16-1、WRN-16-2 和 WRN-16-4 这三种不同宽度因子的模型上进行实验, 可以发现所有知识蒸馏方法均能有效提升模型的准确率, 其中本文使用的方法的性能提升最为显著。尤其是在参数量相同的条件下, 本文的方法能够使学生网络的准确率接近复杂的教师网络。这一结果不仅展示了知识蒸馏在模型压缩和性能提升中的有效性, 也强调了在设计轻量级模型时, 选择适当的知识蒸馏策略的重要性。

进一步分析不同宽度因子模型上的实验结果, 可以注意到随着宽度因子的增加, 模型的性能确实得到了提升, 但是性能增益的幅度逐渐减小。例如, 从 WRN-16-1 到 WRN-16-2, 本文的方法将模型性能提升了约 2.7 个百分点; 而从 WRN-16-2 到 WRN-16-4, 性能增益降为约 2.1 个百分点。此外, 与初始教师网络相比, 即使是在较高宽度因子的设置下, 通过本文方法优化的学生网络仍然能够在显著减少参数量的情况下, 达到与教师网络相近的准确率。这一发现表明, 在特定的参数量限制下, 适度增加模型的宽度能够更高效地利用模型容量, 而过度增加宽度因子可能会导致性能收益的边际递减。

综上所述, 本文的方法中选用宽度因子为 2 的 WRN-16-2 模型作为学生网络。这一选择基于几方面的考量: 首先, 宽度因子为 2 的模型在参数量和计算成本之间提供了一个平衡点, 既不会像宽度因子为 1 的模型那样参数过少, 性能受限, 也不会因为过度增加宽度而导致边际性能增益减小。在此宽度因子下进行知识蒸馏能够有效提升模型性能, 同时保持模型的轻量级特性。

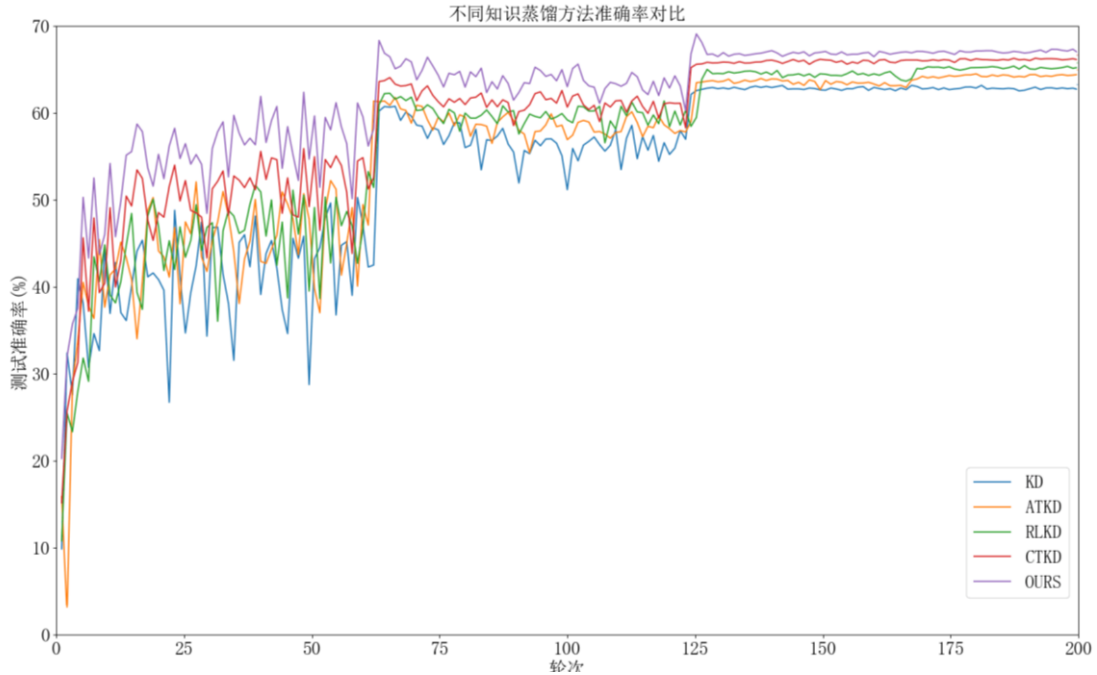


图 4.11 不同知识蒸馏方法在 MultitaskPainting100k 数据集的测试准确率对比

从图 4.11 中可以看出，在评估 MultitaskPainting100K 艺术风格数据集的准确性方面，本文使用的方法展现了一定的优势，其准确率超越了其他四种对比方法。这一成果不仅验证了本文使用的方法对训练数据的处理效果，更重要的是，它揭示了该方法在处理未知数据集时具备更优越的泛化性能。这一现象可能源于本方法在执行知识蒸馏过程中采纳的更为精细化与目标化的策略，从而能够更有效地识别、捕捉并传递关键的特征信息，加强了模型对复杂数据特征的理解和应用能力。实验结果也清晰地展示了本文使用的方法在艺术风格分类任务上的显著优势。通过在知识蒸馏中引入自适应加权格拉姆技术，能够更深入地挖掘教师模型和学生模型之间的潜在联系，实现更高效的知识转移。

4.3.6 消融实验

本文使用了一种基于自适应加权格拉姆矩阵（AWGram）与知识蒸馏的艺术风格分类模型。该模型结合了自适应特征提取、深度知识蒸馏以及特征融合三个核心部分，旨在通过深入学习艺术作品的风格特征以提升分类的准确性。消融实验是评估各个模块对模型整体性能贡献大小的有效方法。通过逐步添加各个部分，并观察分类准确率的变化，可以明确每个组件的作用与重要性。表 4.13 展

示了不同模块在 MultitaskPainting100K 数据集上的消融实验。

表 4.13 不同模块在 MultitaskPainting100K 数据集上的消融实验

方法/配置	准确率 (%)
基线方法	60.76
基线方法+专家教师	62.04
基线方法+专家教师+注意力图	63.21
基线方法+专家教师+注意力图+初始教师	64.71
基线方法+专家教师+注意力图+初始教师+特征相关性矩阵	66.83

(1)基线方法：实验首先采用了基线方法，即不包含专家教师、注意力图、初始教师和特征相关性矩阵的 WRN-16-2 模型，准确率为 60.76%。这一结果作为对照，表明了单纯使用基础模型在艺术风格分类任务中的效能。

(2)添加专家教师：在基线方法基础上引入预训练的专家教师 WRN-40-2，准确率提高到 62.04%，表明专家教师的引入能够有效地提升模型性能。这符合知识蒸馏理论，即通过模仿复杂的教师网络，学生网络能够学习到更加丰富和深入的特征表示。

(3)引入注意力图：进一步加入空间注意力图 $F_{sum,2}$ 后，准确率增至 63.21%。注意力机制能够帮助模型集中学习艺术作品中最具代表性的特征，进一步提升了模型的性能。

(4)结合初始教师：当同时使用专家教师、注意力图和初始教师，准确率提高到 64.71%。这说明初始教师的引入为模型提供了额外的指导信息，有助于模型更好地理解艺术风格的复杂性。

(5)整合 AWGram：最终，将自适应加权格拉姆矩阵(AWGram)融入模型中，准确率达到 66.83%。AWGram 通过捕捉艺术作品不同层之间的风格信息，为模型提供了全面的风格特征理解，显著提升了分类的准确性。

通过这一系列的消融实验，可以看出，每个部分对提升模型性能都有积极的贡献，其中 AWGram 的加入带来了最明显的性能提升。这些实验结果不仅验证了所提出模型各个组成部分的有效性，也展现了在艺术风格分类任务中，深度知识蒸馏与特征相关性技术的结合使用的优越性。

4.3.7 混淆矩阵与误差分析

在本节中的实验中对 Pandora18k 和 WikiArt 数据集进行了混淆矩阵和误差分析，目的是为了深入理解本文提出的方法在艺术风格分类任务上的性能及其局限性。通过分析这两个不同的艺术风格数据集，能够评估模型在面对不同风格和复杂度的艺术作品时的鲁棒性和准确度。混淆矩阵提供了一种直观的方式来查看模型对于每个艺术风格的分类效果，包括它在哪些类别上表现良好，以及它最容易混淆哪些风格。此外，误差分析进一步识别了模型性能不足的原因，无论是数据集的不平衡、特定风格之间的相似性，还是模型自身的限制。通过在这两个数据集上的详细分析，可以优化本文的分类模型，提高其在复杂和多样化的艺术风格识别任务中的表现。

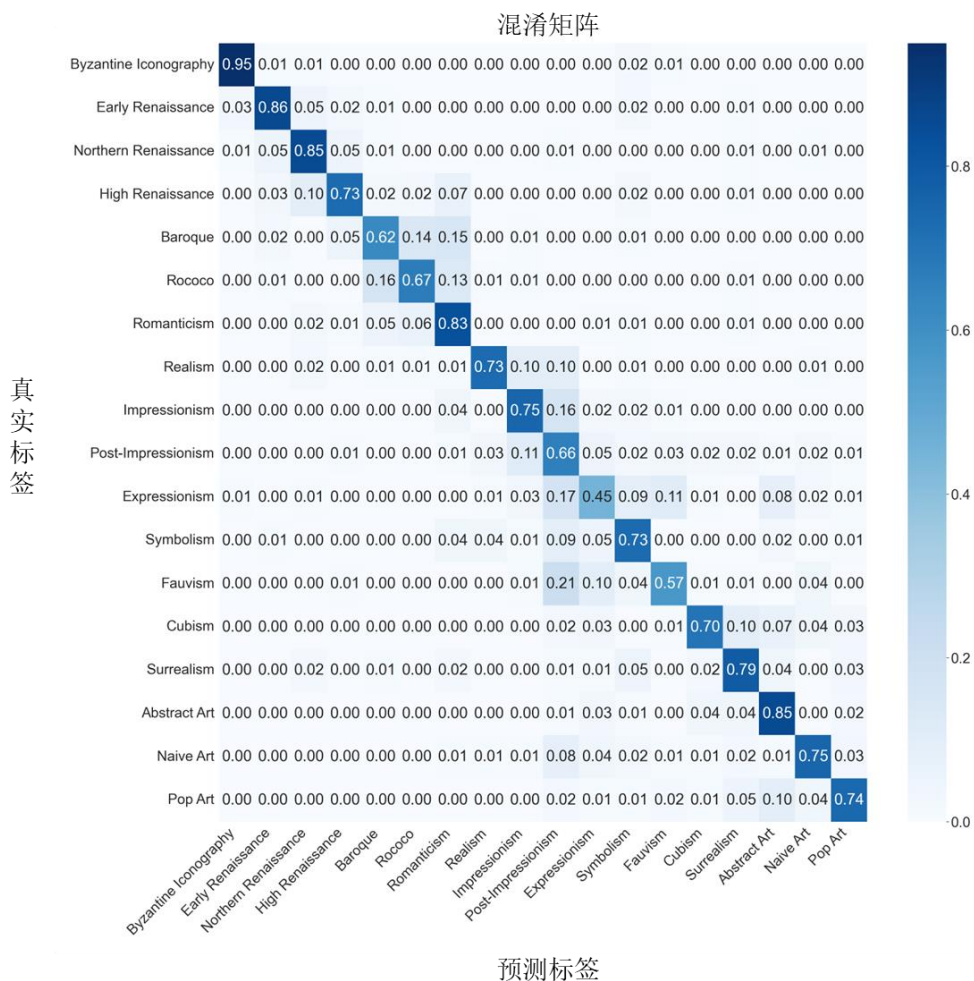


图 4.12 Pandora18k 的准确率混淆矩阵

根据图 4.12 所展示的混淆矩阵，可以分析 Pandora18k 数据集中不同艺术风格的分类性能。例如，拜占庭图标画风格的识别准确率达到了 95%，这反映了其

独特的宗教象征和黄金背景在视觉上的突出性，大幅提升了模型的识别精度。然而，对于巴洛克风格的识别准确率仅为 62%，这暗示了模型在区分巴洛克与洛可可风格的识别上存在困难，可能由于两者在豪华装饰和动态形式上的相似性所致；模型在区分巴洛克与浪漫主义风格时也遇到了挑战，可能源于这两种风格都倾向于利用鲜明的光影对比来营造深度和动态感。图 4.13(a)和图 4.13(b)展示了巴洛克风格被误分至洛可可风格和浪漫主义风格的样本示例。



(a) 巴洛克风格被误分至洛可可风格的样本示例



(b) 巴洛克风格被误分至浪漫主义风格的样本示例

图 4.13 Pandora18k 数据集上巴洛克风格误分类样本示例

再观察印象主义，其准确率为 75%，说明模型能较为有效地辨识其典型的散点刷法和光影效果。然而，与后印象主义的混淆率显示，模型在处理这两种风格的细微区别，如色彩运用和主题表达上，仍存在挑战。立体主义的分类准确率为 70%，表明模型在捕捉到立体主义的几何解构和抽象表现上具有一定的准确性，但与构成主义的混淆表明了识别使用几何形态进行构图的风格时的限制。

抽象艺术的极高分类准确率（85%）凸显了模型对于非具象表现的高度敏感性，这可能归因于抽象艺术在形式和色彩上的显著差异。流行艺术的准确率为 74%，其与抽象艺术和现实主义风格的混淆较少，反映了模型对于其独有的商业

文化和大众媒体元素的辨识能力。

此外，浪漫主义和超现实主义的较高准确率（分别为 83%和 79%）表明模型能有效区分这两种风格的情感表达和梦幻场景。尽管两者都具有强烈的情感载体，但它们在视觉表现上的差异被模型成功捕捉。最终，象征主义的识别准确率为 73%，这表明尽管其主题深邃且多样，模型仍能在一定程度上区分其与其他风格的视觉和意象差异。

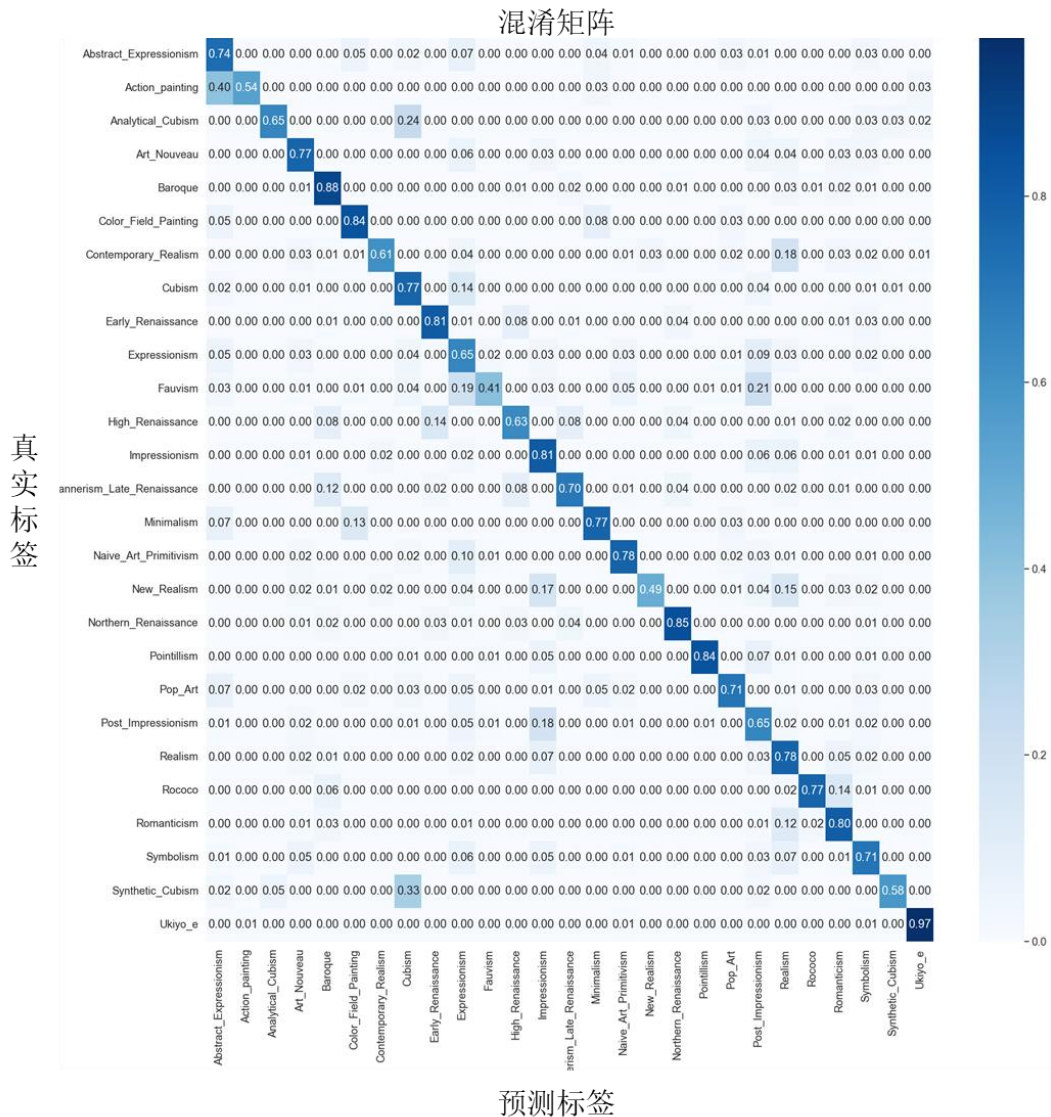


图 4.14 WikiArt 数据集风格分类准确率的混淆矩阵

如图 4.14 所示，从 WikiArt 数据集风格分类准确率的混淆矩阵中，可以对

WikiArt 数据集上的各艺术风格识别性能进行深入分析。首先,值得注意的是抽象表现主义,其自分类准确率为 74%,显示了这一风格在数据集中的鲜明特性被较好地捕捉。这种风格于 20 世纪中期在美国发展,强调自由表达和个人情感的表达,其特有的无规则笔触和色彩应用使得其与其他风格区分开来。

分析立体主义和合成立体主义,分别具有 65%和 70%的自分类准确率,它们相对易于被归类为立体主义。这两种风格虽然属于立体主义的不同阶段,但它们共享立体主义的基本特征,如对传统视角的解构和对几何形状的强调,使得它们在视觉上相互关联。在浮世绘方面,该风格的分类准确率高达 97%,这是由于浮世绘具有独特的风格特征,如线条的使用、平面的色彩以及独特的主题和构图方式,这使得浮世绘在数据集中容易被识别。

然而,行动绘画风格,作为抽象表现主义的一个分支,其与抽象表现主义的混淆较大,54%的作品被正确分类,而有相当一部分作品被误分类为抽象表现主义,这反映了在某些视觉特征上,行动绘画与抽象表现主义之间有较大的相似性。行动绘画强调画笔动作的直观表达,这可能与抽象表现主义共享某些视觉语言。图 4.15 展示了行动绘画风格被误分至抽象表现主义风格的样本示例。



图 4.15 WikiArt 数据集上行动绘画风格被误分至抽象表现主义风格的部分样本示例

此外,印象派的自分类准确率为 81%,印象派风格以其独特的光影处理和色彩运用闻名,这些特征在数据集中得到了较好的识别。然而,与之类似的后印象派风格,其自分类准确率为 65%,表明在细节和色彩表达方面,后印象派与印象派之间存在一定的视觉相似性,导致一些混淆。图 4.16 展示了后印象主义风格被误分至印象主义风格的部分样本示例。

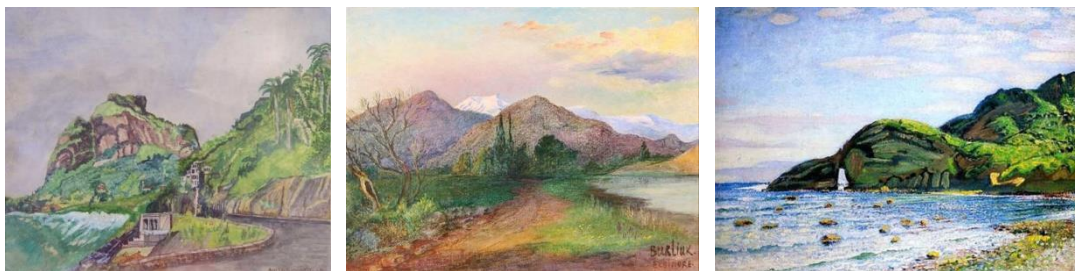


图 4.16 WikiArt 数据集上后印象主义风格被误分至印象主义风格的部分样本示例

考虑到 MultitaskPainting100K 数据集覆盖了多达 125 个广泛的艺术风格类别，展示一个完整的混淆矩阵将会难以直观的理解。因此，本文未展示 MultitaskPainting100K 的完整混淆矩阵，而是对易混淆的类别进行一定的分析，目的是揭示模型在艺术风格识别上遇到的一些挑战。

表 4.14 MultitaskPainting100K 数据集上识别准确率最低的四类及其主要误分类情况

风格种类	准确率(%)	最常误分类为	误分类概率 (%)
抽象表现主义	49%	非形式主义	19%
洛可可	54%	巴洛克	21%
象征主义	57%	浪漫主义	11%
野兽派	57%	表现主义	18%

例如，野兽派与表现主义这两个艺术流派在强调个体情感表达的同时，均通过使用鲜明的色彩和自由奔放的笔触来实现这一艺术追求。野兽派艺术家倾向于利用色彩的强烈对比和纯粹性来捕捉观者的视觉注意力，从而传达一种直接而强烈的视觉冲击力。与之相较，表现主义则通过色彩和形式的扭曲来表达更深层次的个人情绪和社会批判，反映了艺术家对内心世界和外部现实的主观感受。尽管这两种风格在意图和重点上存在明显差异——野兽派更多地聚焦于色彩本身的表达力，而表现主义则深入探讨了情感的复杂性与多样性——但它们在视觉表现手法上的共通之处，尤其是在对色彩和线条的大胆运用上，可能导致在视觉识别上的相互混淆。图 4.17 展示了野兽派风格被误分至表现主义风格的部分样本示例。

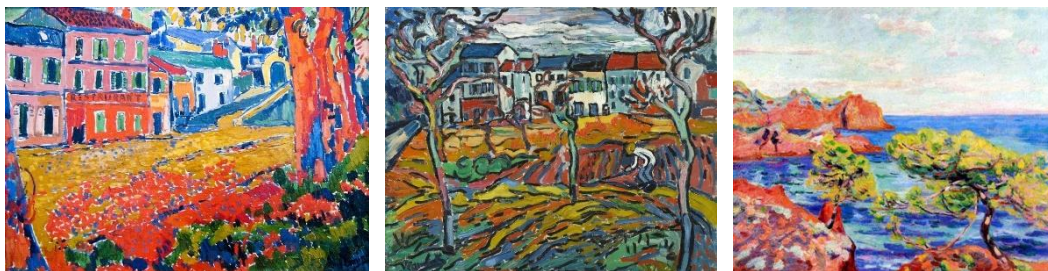


图 4.17 MultitaskPainting100K 数据集上野兽派风格被误分至表现主义风格的部分样本示例

通过上述分析可以总结出,虽然本文提出的方法在识别某些独特和明显的艺术风格上表现出色,但在处理一些风格上的细微差别及历史相近风格时,仍存在挑战。这些发现不仅揭示了本文提出的方法的强项和弱点,也为未来的研究方向和数据集的改进提供了宝贵的见解。

4.3.8 与现有方法的对比实验

为了本方法的有效性,本文在 Pandora18k、Wiki Art 和 MultitaskPainting100k 等三个数据集上将提出的方法中学生模型的准确率与其他研究方法进行了比较。

4.3.8.1 Pandora18k 数据集的现有方法准确率对比

表 4.15 展示了本文的方法以及其他分类方法在 Pandora18k 数据集上对艺术风格进行分类的准确率实验结果比较:

表 4.15 不同方法在 Pandora18k 数据集上的实验结果

方法	准确率 (%)
Florea 等人 ^[61]	50.12
Gieseke 等人 ^[62]	63.50
Rodriguez 等人 ^[63]	70.20
Menis 等人 ^[64]	72.47
本方法	73.48

从表 4.15 中可以观察到, Florea 等人的方法在该数据集上获得了 50.12% 的准确率,这可能反映出他们的模型对艺术风格特征的捕获能力有限。而 Gieseke 等人的方法准确率提升至 63.50%,这一改进可能源于他们在模型中采用了更高级的特征提取技术。

Rodriguez 等人的方法通过利用迁移学习和加权图像块,将准确率进一步提

升至 70.20%，表明了预训练模型在艺术风格识别任务中的有效性。Menis 等人通过实施深度集成学习策略，使准确率达到 72.47%，这一结果突出了不同模型集成带来的性能提升。

本文方法以 73.48% 的最高准确率领先于其他方法，通过结合知识蒸馏技术与自适应加权格拉姆矩阵，本文方法不仅减少了模型的计算负担，同时保持了学生网络的高分类精度。此外，本文方法的成功实现了对艺术风格特征的深层次理解和有效分类，这一点在 Pandora18k 数据集上的表现尤为突出。

4.3.8.2 WikiArt 数据集的现有方法准确率对比

表 4.16 展示了本文的方法以及其他分类方法在 WikiArt 数据集上对艺术风格进行分类的准确率实验结果比较：

表 4.16 不同方法在 WikiArt 数据集上的实验结果

方法	准确率 (%)
Tan 等人 ^[65]	54.52
Cetinic 等人 ^[66]	56.43
Mo 等人 ^[67]	59.53
Lecoutre 等人 ^[31]	62.80
Wu 等人 ^[68]	64.44
Sandoval 等人 ^[69]	66.71
Zhao 等人 ^[70]	69.97
Chu 等人 ^[71]	70.99
Jiang 等人 ^[72]	71.24
本方法	72.26

从表 4.15 中可以看出，Tan 等人的方法获得了 54.52% 的准确率，这可能是由于他们的深度卷积网络在处理细腻的艺术风格特征时存在一定局限性。Cetinic 等人通过对卷积神经网络进行微调，将准确率提升至 56.43%，这表明了微调技术在艺术分类任务中的有效性。Mo 等人进一步将准确率提升到 59.53%，可能是因为他们采用了层次分类方法，更好地处理了不同艺术风格间的差异性。

Lecoutre 等人的研究结果显著提升至 62.80%，他们可能采用了更复杂的特征提取和分类策略，以自动识别绘画中的艺术风格。Wu 等人的方法通过稳定扩散技术进行内容-风格分离，在艺术分析方面获得了 64.44% 的准确率，突出了将

生成艺术与传统艺术风格识别结合的重要性。

Sandoval 等人采用了两阶段深度学习方法对细分艺术品进行分类,准确率达到 66.71%,这进一步证实了深度学习在细粒度艺术品分类中的潜力。Zhao 等人利用模型比较在艺术分类中取得了 69.97%的准确率,展现了不同模型在捕捉艺术风格差异方面的效力。

Chu 等人通过学习深度相关特征,使得图像风格分类的准确率达到 70.99%,显示了深层特征在捕获艺术特性方面的有效性。Jiang 等人的方法通过大规模迁移学习,准确率进一步提升至 71.24%,说明了利用大数据预训练模型在艺术风格分类中的有效性。

最终,本文方法在所有对比的方法中表现最佳,达到了 72.26%的最高准确率。这一成果归功于本文方法在特征提取、模型训练和知识蒸馏等方面的创新应用,证明了本文方法在处理艺术风格分类任务时的高效性。

4.3.8.2 MultitaskPainting100k 数据集的现有方法准确率对比

表 4.17 展示了本文的方法以及其他分类方法在 MultitaskPainting100k 数据集上对艺术风格进行分类的准确率实验结果比较:

表 4.17 不同方法在 MultitaskPainting100k 数据集上的实验结果

方法	准确率 (%)
Bianco 等人 ^[73]	57.2
Zhao 等人 ^[70]	63.15
Jiang 等人 ^[72]	65.82
本方法	66.83

从表 4.16 中的数据可以看出,Bianco 等人的方法在该数据集上取得了 57.2%的准确率,这可能反映出他们的模型在捕获艺术风格特征方面存在一定的局限性。而 Zhao 等人通过引入更复杂的模型架构和优化的特征融合技术,将准确率提升到 63.15%,这种提升突显了在模型设计中考虑艺术风格的多样性和复杂性的重要性。接着,Jiang 等人利用大规模迁移学习和精细化训练方法,进一步将准确率提高至 65.82%,显示了在艺术风格识别任务中利用预训练模型的有效性。本文提出的方法以 66.83%的准确率超越了其他方法,通过融合知识蒸馏技术和自适应加权格拉姆矩阵,既降低了模型的计算需求,又保证了模型对艺术风格特征的敏感度和分类精度。此外,本文方法在艺术风格的深层次理解和有效分类方

面展现了显著优势，特别是在处理复杂多样的艺术风格数据集时，显示了良好的泛化能力和分类效果。

通过对多个公开艺术风格数据集进行的一系列实验，本文的方法展现出了一定的性能优势。相比于传统方法，本文提出的模型不仅能够更准确地识别和分类不同的艺术风格，而且在面对数据集规模变化时显示出更好的适应性和稳定性。

4.4 本章小结

在本章中，详细探讨了基于自适应加权格拉姆矩阵与联合教学知识蒸馏的艺术风格分类模型，并对其进行了系统的实验验证。通过在三个公开的艺术风格数据集上进行的一系列实验，本研究不仅展示了所提模型中各模块的有效性，还通过与现有方法的比较，证明了本文提出的方法的优越性。

首先，通过超参数对比实验，本文精确确定了最适宜的损失函数超参数以及知识蒸馏的最优温度参数，随后通过实验比较确定了最佳的注意力图映射函数。通过分析不同相关性权重构成的自适应加权格拉姆矩阵对风格分类性能的提升效果，本研究识别出了最优的特征相关性配置。此外，通过对比分析不同知识蒸馏策略，本研究验证了所提出的联合教学知识蒸馏技术的显著优越性，该技术不仅显著降低了计算负荷，同时保持了模型对艺术风格的高度敏感性。实验结果充分证明，所提模型在艺术风格分类任务中实现了显著的性能提升。

其次，消融实验深入验证了模型各个组件对整体性能的贡献，确立了自适应加权格拉姆矩阵与双教师知识蒸馏技术在提高艺术风格分类准确率方面的关键作用。此外，通过对模型性能的准确率混淆矩阵进行分析，本研究深入探讨了所提方法在不同艺术风格识别任务中的优势与潜在局限，揭示了模型在特定风格分类上的表现优劣，从而为未来的优化策略提供了重要指引。

最后，通过与现有方法的对比实验，本文的方法在 WikiArt、MultitaskPainting100k 和 Pandora18k 数据集上均展现出优于现有技术的性能，验证了结合自适应加权格拉姆矩阵与知识蒸馏技术的艺术风格分类模型的有效性与实用性。

第5章 结论与展望

5.1 结论

本文提出了一种基于联合教学知识蒸馏与特征相关性技术的艺术风格分类模型。模型的核心在于融合了自适应特征提取与联合教学知识蒸馏的策略，为艺术风格的自动识别和分类提供了一种新的视角。

通过引入自适应特征相关性矩阵，模型能够在不同的层次上捕捉到艺术作品的细粒度特征，从而实现对艺术风格更为全面的理解。这种自适应特征提取机制不仅加深了对艺术品风格特征的认识，而且提升了模型的分类精确度和泛化能力。与此同时，深度知识蒸馏技术的应用，进一步降低了模型在实际应用中的计算负担，同时确保了对艺术风格的高敏感性。

本文通过在三个不同的公开艺术风格数据集上进行系统的实验验证，证明了所提出模型在艺术风格识别任务上的有效性和实用性。结果表明，与现有的艺术风格分类方法相比，本文的模型不仅在分类精度上有一定提高，而且在处理不同规模数据集时表现出了良好的适应性和鲁棒性。特别是在面对大规模艺术作品数据集时，本文的模型能够有效地识别和分类不同的艺术风格，显示了其在艺术领域应用的广泛潜力。

此外，本研究在特征提取和融合方面的创新，为艺术风格分类领域提供了新的研究方向。通过自适应地调整特征提取策略，本文的模型不仅提高了对复杂艺术风格的识别能力，而且增强了对图像不同层面信息的捕获能力，从细节到整体都进行了深入分析。这一方法论的成功实施，为未来在艺术领域中深入探索高效且灵活的特征提取和融合技术奠定了坚实的基础。

5.2 展望

尽管本论文在艺术风格分类领域取得了显著的进步，但未来的研究仍有以下几个方向：首先，当前的数据集虽然已经包含了丰富的样本，但在规模和多样性上还有提升的空间。未来可以通过扩大数据集的规模和多样性，进一步提升模型的泛化能力和鲁棒性。其次，当前模型的训练与优化过程中还可以引入更多创

新的策略，如进一步优化知识蒸馏的过程，以及探索更有效的特征融合方法。此外，未来的研究还可以探索模型在其他艺术相关任务上的应用，如艺术作品的生成、编辑和恢复等。最后，考虑到艺术风格的主观性和多样性，未来的工作可以探索结合艺术理论和历史背景，提高模型对艺术风格深层次特征的理解和识别能力。这不仅可以促进计算机视觉与艺术史学科的交叉融合，也能为艺术品的自动分类与分析提供更加深入和准确的技术支持。

参考文献

- [1] Wölfflin H. Principles of art history[M]. Principles of art historyCourier Corporation, 2012.
- [2] 范景中, 杨成凯. 《艺术的故事》[J]. 收藏与投资, 2014(12): 124-124.
- [3] Wen W, Wu C, Wang Y, et al. Learning structured sparsity in deep neural networks[J]. Advances in Neural Information Processing Systems, 2016, 29.
- [4] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. 计算机学报, 2022, 45(3): 624-653.
- [5] Zhang X, Zou J, He K, et al. Accelerating very deep convolutional networks for classification and detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(10): 1943-1955.
- [6] Hubara I, Courbariaux M, Soudry D, et al. Quantized neural networks: Training neural networks with low precision weights and activations[J]. Journal of Machine Learning Research, 2018, 18(187): 1-30.
- [7] Peng K-C, Chen T. A framework of extracting multi-scale features using multiple convolutional neural networksA framework of extracting multi-scale features using multiple convolutional neural networks[C]. 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015: 1-6.
- [8] Zujovic J, Gandy L, Friedman S, et al. Classifying paintings by artistic genre: An analysis of features & classifiersClassifying paintings by artistic genre: An analysis of features & classifiers[C]. 2009 IEEE International Workshop on Multimedia Signal Processing, 2009: 1-5.
- [9] Khan F S, Beigpour S, Van De Weijer J, et al. Painting-91: A large scale database for computational painting Categorization[J]. Machine Vision and Applications, 2014, 25(6): 1385-1397.
- [10] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [11] Efros A A, Leung T K. Texture synthesis by non-parametric samplingTexture synthesis by non-parametric sampling[C]. Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999: 1033-1038.
- [12] Wu Q, Yu Y. Feature matching and deformation for texture Synthesis[J]. ACM Transactions on Graphics, 2004, 23(3): 364-367.
- [13] Xie X, Tian F, Seah H S. Feature Guided Texture Synthesis (FGTS) for artistic style TransferFeature Guided Texture Synthesis (FGTS) for artistic style transfer[C]. Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts, Perth Australia, 2007: 44-49.
- [14] Gatys L, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[J].

- Advances in Neural Information Processing Systems, 2015, 28.
- [15] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networksImage style transfer using convolutional neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2414-2423.
- [16] Chu W-T, Wu Y-L. Deep Correlation Features for Image Style ClassificationDeep Correlation Features for Image Style Classification[C]. Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam The Netherlands, 2016: 402-406.
- [17] Ruta D, Tarres G C, Black A, et al. ALADIN-NST: Self-supervised disentangled representation learning of artistic style through Neural Style Transfer[J]. ArXiv, 2023(2023-08-17).
- [18] Bianconi F, Bello-Cerezo R. Evaluation of visual descriptors for painting categorisationEvaluation of visual descriptors for painting categorisation[C]. IOP Conference Series: Materials Science and Engineering, 2018: 012037.
- [19] Garcia N, Vogiatzis G. How to read paintings: semantic art understanding with multi-modal retrievalHow to read paintings: semantic art understanding with multi-modal retrieval[C]. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018: 0-0.
- [20] Nanni L, Ghidoni S, Brahnem S. Handcrafted vs. non-handcrafted features for computer vision classification[J]. Pattern Recognition, 2017, 71: 158-172.
- [21] Sablatnig R, Kammerer P, Zolda E. Hierarchical classification of paintings using face-and brush stroke modelsHierarchical classification of paintings using face-and brush stroke models[C]. Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170), 1998: 172-174.
- [22] Shen J. Stochastic modeling western paintings for effective classification[J]. Pattern Recognition, 2009, 42(2): 293-301.
- [23] Karayev S, Trentacoste M, Han H, et al. Recognizing Image StyleRecognizing Image Style[C]. Proceedings of the British Machine Vision Conference 2014, 2014: 122.1-122.11.
- [24] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [25] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145-175.
- [26] Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification[J]. IEEE transactions on Image Processing, 2010, 19(6): 1657-1663.
- [27] Xu Z, Tao D, Zhang Y, et al. Architectural Style Classification Using Multinomial Latent Logistic Regression[M]Architectural Style Classification Using Multinomial Latent Logistic Regression//Computer Vision – ECCV 2014. Cham: Springer International Publishing, 2014: 600-615.
- [28] Chen L, Chen J, Zou Q, et al. Multi-View Feature Combination for Ancient Paintings Chronological Classification[J]. Journal on Computing and Cultural Heritage, 2017, 10(2): 1-15.

- [29] Hicsonmez S, Samet N, Sener F, et al. DRAW: Deep Networks for Recognizing Styles of Artists Who Illustrate Children's BooksDRAW: Deep Networks for Recognizing Styles of Artists Who Illustrate Children's Books[C]. Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Bucharest Romania, 2017: 338-346.
- [30] Elgammal A, Liu B, Kim D, et al. The shape of art history in the eyes of the machineThe shape of art history in the eyes of the machine[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [31] Lecoutre A, Negrevergne B, Yger F. Recognizing art style automatically in painting with deep learningRecognizing art style automatically in painting with deep learning[C]. Asian Conference on Machine Learning, 2017: 327-342.
- [32] Tan W R, Chan C S, Aguirre H E, et al. Fuzzy qualitative deep compression network[J]. Neurocomputing, 2017, 251: 1-15.
- [33] Sethi R J, Buell C A, Seeley W P, et al. An Open Access Platform for Analyzing Artistic Style Using Semantic Workflows[M]An Open Access Platform for Analyzing Artistic Style Using Semantic Workflows//Web Services – ICWS 2018. Cham: Springer International Publishing, 2018: 166-180.
- [34] Gultepe E, Conturo T E, Makrehchi M. Predicting and grouping digitized paintings by style using unsupervised feature learning[J]. Journal of cultural heritage, 2018, 31: 13-23.
- [35] Chen L, Yang J. Recognizing the Style of Visual Arts via Adaptive Cross-layer CorrelationRecognizing the Style of Visual Arts via Adaptive Cross-layer Correlation[C]. Proceedings of the 27th ACM International Conference on Multimedia, Nice France, 2019: 2459-2467.
- [36] Joshi A, Agrawal A, Nair S. Art Style Classification with Self-Trained Ensemble of AutoEncoding Transformations[J]. ArXiv, 2020(2020-12-06).
- [37] Chen B. Classification of Artistic Styles of Chinese Art Paintings Based on the CNN Model[J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [38] Akter M, Akther Mst R, Khaliluzzaman Md. Recognizing Art Style Automatically in Painting Using Convolutional Neural Network[M]Recognizing Art Style Automatically in Painting Using Convolutional Neural Network//Computational Intelligence in Machine Learning. Singapore: Springer Nature Singapore, 2022: 221-236.
- [39] Bucilua C, Caruana R, Niculescu-Mizil A. Model CompressionModel compression[C]. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia PA USA, 2006: 535-541.
- [40] Ba J, Caruana R. Do deep nets really need to be deep?[J]. Advances in Neural Information Processing Systems, 2014, 27.
- [41] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network[J]. ArXiv, 2015(2015-03-09).
- [42] Yang C, Xie L, Qiao S, et al. Knowledge Distillation in Generations: More Tolerant Teachers Educate Better Students[J]. ArXiv, 2018(2018-09-07).
- [43] Kim J, Park S, Kwak N. Paraphrasing complex network: Network compression via factor

- transfer[J]. Advances in Neural Information Processing Systems, 2018, 31.
- [44] Adriana R, Nicolas B, Ebrahimi K S, et al. Fitnets: Hints for thin deep nets[J]. Proc. ICLR, 2015, 2(3): 1.
- [45] Mishra A, Marr D. Apprentice: Using Knowledge Distillation Techniques To Improve Low-Precision Network Accuracy[J]. ArXiv, 2017(2017-11-15).
- [46] Zagoruyko S, Komodakis N. Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer[J]. ArXiv, 2017(2017-02-12).
- [47] Zhou G, Fan Y, Cui R, et al. Rocket launching: A universal and efficient framework for training well-performing light netRocket launching: A universal and efficient framework for training well-performing light net[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [48] Passalis N, Tefas A. Learning deep representations with probabilistic knowledge transferLearning deep representations with probabilistic knowledge transfer[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 268-284.
- [49] Park W, Kim D, Lu Y, et al. Relational knowledge distillationRelational knowledge distillation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3967-3976.
- [50] Zhang L, Song J, Gao A, et al. Be your own teacher: Improve the performance of convolutional neural networks via self distillationBe your own teacher: Improve the performance of convolutional neural networks via self distillation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3713-3722.
- [51] Xu T-B, Liu C-L. Data-distortion guided self-distillation for deep neural networksData-distortion guided self-distillation for deep neural networks[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 5565-5572.
- [52] Li H-T, Lin S-C, Chen C-Y, et al. Layer-level knowledge distillation for deep neural network learning[J]. Applied Sciences, 2019, 9(10): 1966.
- [53] Zhao H, Sun X, Dong J, et al. Highlight every step: Knowledge distillation via collaborative teaching[J]. IEEE Transactions on Cybernetics, 2020, 52(4): 2070-2081.
- [54] Borza D-L, Darabant A S, Ileni T A, et al. Effective online knowledge distillation via attention-based model ensembling[J]. Mathematics, 2022, 10(22): 4285.
- [55] Liu Y, Cao J, Li B, et al. Learning to explore distillability and sparsability: a joint framework for model compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(3): 3378-3395.
- [56] He K, Zhang X, Ren S, et al. Deep residual learning for image recognitionDeep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [57] Zagoruyko S, Komodakis N. Wide Residual Networks[J]. ArXiv, 2017(2017-06-14).
- [58] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. ArXiv, 2016(2016-05-19).
- [59] Cordonnier J-B, Loukas A, Jaggi M. On the Relationship between Self-Attention and

- Convolutional Layers[J]. ArXiv, 2020(2020-01-10).
- [60] Alkhulaifi A, Alsahli F, Ahmad I. Knowledge distillation in deep learning and its applications[J]. PeerJ Computer Science, 2021, 7: e474.
- [61] Florea C, Condorovici R, Vertan C, et al. Pandora: Description of a painting database for art movement recognition with baselines and perspectivesPandora: Description of a painting database for art movement recognition with baselines and perspectives[C]. 2016 24th European Signal Processing Conference (EUSIPCO), 2016: 918-922.
- [62] Florea C, Gieseke F. Artistic movement recognition by consensus of boosted SVM based experts[J]. Journal of Visual Communication and Image Representation, 2018, 56: 220-233.
- [63] Rodriguez C S, Lech M, Pirogova E. Classification of style in fine-art paintings using transfer learning and weighted image patchesClassification of style in fine-art paintings using transfer learning and weighted image patches[C]. 2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS), 2018: 1-7.
- [64] Menis-Mastromichalakis O, Sofou N, Stamou G. Deep ensemble art style recognitionDeep ensemble art style recognition[C]. 2020 International Joint Conference on Neural Networks (IJCNN), 2020: 1-8.
- [65] Tan W R, Chan C S, Aguirre H E, et al. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classificationCeci n'est pas une pipe: A deep convolutional network for fine-art paintings classification[C]. 2016 IEEE International Conference on Image Processing (ICIP), 2016: 3703-3707.
- [66] Cetinic E, Lipic T, Grgic S. Fine-tuning convolutional neural networks for fine art classification[J]. Expert Systems with Applications, 2018, 114: 107-118.
- [67] Mohammadi M R, Rustae F. Hierarchical classification of Fine-art paintings using deep neural Networks[J]. Iran Journal of Computer Science, 2021, 4(1): 59-66.
- [68] Wu Y, Nakashima Y, Garcia N. Not Only Generative Art: Stable Diffusion for Content-Style Disentanglement in Art AnalysisNot Only Generative Art: Stable Diffusion for Content-Style Disentanglement in Art Analysis[C]. Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, Thessaloniki Greece, 2023: 199-208.
- [69] Sandoval C, Pirogova E, Lech M. Two-stage deep learning approach to the classification of fine-art paintings[J]. IEEE Access, 2019, 7: 41770-41781.
- [70] Zhao W, Zhou D, Qiu X, et al. Compare the performance of the models in art classification[J]. Plos One, 2021, 16(3): e0248414.
- [71] Chu W-T, Wu Y-L. Image style classification based on learnt deep correlation features[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2491-2502.
- [72] Zhao W, Jiang W, Qiu X. Big transfer learning for fine art classification[J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [73] Bianco S, Mazzini D, Napoletano P, et al. Multitask painting categorization by deep multibranch neural network[J]. Expert Systems with Applications, 2019, 135: 90-101.