

# 基于注意力机制的个性化图像美学质量评估

江佳俊<sup>1,2</sup>, 蒋 旻<sup>1,2\*</sup>, 杨晓雨<sup>1</sup>, 郭 嘉<sup>1,2</sup>

(1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;

2. 智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学), 湖北 武汉 430065)

**摘 要:** 图像美学质量评估在图像推荐和图像编辑美学等行业具有良好的应用前景。大部分美学质量评估方法存在两个问题: 第一, 只建立了通用美学评估模型, 而忽略了个性化审美, 这样的评估方法只适用于大众审美的评估, 无法区分出不同用户之间的差异; 第二, 往往只从图像本身特征和整体布局进行评估, 没有考虑主体的突出性和位置合理性, 由于主体特征对评估结果有较大的影响, 以致忽视主体的做法会使得评估结果具有较差的完备性。针对这些问题, 文中设计了一种基于注意力机制的个性化图像美学质量评估方法。该方法在个性化残差网络的基础之上引入了注意力机制, 既保持了个人在美学质量评估的主观偏好, 又将主体的显著性因素加入评分过程。在 Flickr 数据集上的实验中, 该算法的平均 SROCC 相关系数表现为 0.667, 相较于单一的图像美学质量评估在 Flickr 数据集上的表现, 性能提升了 3%。

**关键词:** 图像美学质量评估; 注意力机制; 残差网络; 个性化; 深度学习

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2021)10-0056-07

doi: 10.3969/j.issn.1673-629X.2021.10.010

## Research on Evaluation of Personalized Image Aesthetic Quality Based on Attention Mechanism

JIANG Jia-jun<sup>1,2</sup>, JIANG Min<sup>1,2\*</sup>, YANG Xiao-yu<sup>1</sup>, GUO Jia<sup>1,2</sup>

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;

2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System

(Wuhan University of Science and Technology), Wuhan 430065, China)

**Abstract:** Image aesthetic quality assessment has a good application prospect in image recommendation and image editing aesthetics. There are two problems in most aesthetic quality assessment methods. On the one hand, only the general aesthetic assessment model is established, while the individual aesthetic assessment is ignored. Such an assessment method is only applicable to the public aesthetic assessment and cannot distinguish the differences between users. On the other hand, the evaluation is usually conducted only from the features of the image itself and the overall layout, without considering the prominence and rationality of the subject. Since the subject features have a great impact on the evaluation results, the practice of ignoring the subject will make the evaluation results have poor completeness. Aiming at these problems, we design a personalized image aesthetic quality evaluation method based on attention mechanism. This method introduces the attention mechanism on the basis of the personalized residual network, which not only maintains the subjective preference of the individual in the aesthetic quality assessment, but also adds the significance factor of the subject into the scoring process. In experiments on Flickr data sets, the average SROCC correlation coefficient of this algorithm is 0.667, which is 3% better than the performance of a single image aesthetic quality assessment on Flickr data sets.

**Key words:** image aesthetics quality evaluation; attention mechanism; residual network; personalized; deep learning

## 0 引言

图像美学质量评估是利用计算机模拟人类对美的感知和认知, 自动评价图像的“美感”<sup>[1]</sup>。图像美学质量评估是计算美学与计算机视觉、心理学、虚拟现实等

领域的学科交叉新方向, 其在图像推荐、图像编辑美学等行业具有良好的应用前景。

在图像美学质量评估的研究历史中, 大部分美学图像评估方法都试图用量化指标对图像美感进行评

收稿日期: 2020-11-26

修回日期: 2021-03-26

基金项目: 国家自然科学基金项目(41571396, 61702385); 2020年国家级创业训练计划项目(202010488007)

作者简介: 江佳俊(1997-), 男, 硕士研究生, 研究方向为计算机视觉、深度学习; 通信作者: 蒋 旻(1975-), 女, 教授, 博士, 研究方向为计算机视觉、机器人自动导航。

估,根据量化指标的不同,大部分美学质量评估方法可以被划分成以下三类:(1)基于分类的方法,图像美学质量评估结果被分成“高质量”、“低质量”和“普通质量”三种类别;(2)基于评分的方法,图像评估结果为一个量化的分值,通常在 1~10 之间;(3)基于分布的方法,评估结果用各美学评分分值的概率分布来表达。由于近年来提出的大部分美学评分方法采用第二种方式,所以文中也选用基于评分的方式评估图像美学。

早期较传统的图像美学评价方法是根据与摄影相关的视觉特征(例如对称性、构图布局、颜色、纹理、清晰度、对比度等)来进行图像美学质量评估。这些特征往往是用一些手工设计的视觉特征描述子来表达。虽然手工设计的特征可设计性范围很广,但是由于设计特征时需要具备摄影中的美学规则的相关经验,计算量较大、很少考虑特征之间的关系和很难涵盖所有的美学规则等因素的存在,所以基于手工设计的特征的方法往往效率较低且输出结果不稳定。

为了改善传统方法中手工设计特征的低效和鲁棒性差的缺陷,近年来,研究人员逐渐将深度学习技术引入审美评估,但是大部分基于深度学习的审美评估方法存在一些共同的缺陷:

(1)一幅图像除了考虑构图布局外,往往存在需要突出的主体(主题)部分。大部分基于深度学习的方法往往只从图像整体角度进行评估,没有考虑主体

的突出性和位置合理性。

(2)对于同一幅图像,不同人的美学评价可能不尽相同,甚至相去甚远。通用的美学质量评估方案一般没有考虑不同用户审美间的差异。如何将个人审美主观偏好建立在评估系统中,实现个性化美学质量评估,成为一个有挑战性的问题。

文中以个性化图像评估为研究对象,利用视觉显著性技术将主体的突出性和位置合理性等因素引入审美评估过程,提出了一种新的基于注意力机制的个性化美学质量评估方法。该方法在 Rodriguez-Pardo 等人提出的个性化残差网络 PAA<sup>[2]</sup> (personalised-aesthetic-assessment-using-residual-adapters) 的基础之上引入了注意力机制,既保持了图像内容个性化偏好(个人美学质量评估的主观偏好),又将个人对图像构图的偏好(即主体的突出性)因素加入评分过程。在 Flickr<sup>[3]</sup> 数据集上的实验表明,该方法相较于 PAA 方法在评估准确率上提升了约 3%。

## 1 相关工作

### 1.1 注意力机制

注意力机制的本质在于让网络能够学会忽略无关信息而关注重点信息。从作用角度而言,注意力机制主要分为空间注意力和时间注意力两类,文中采用的是空间注意力非局部操作 non\_local 模块。

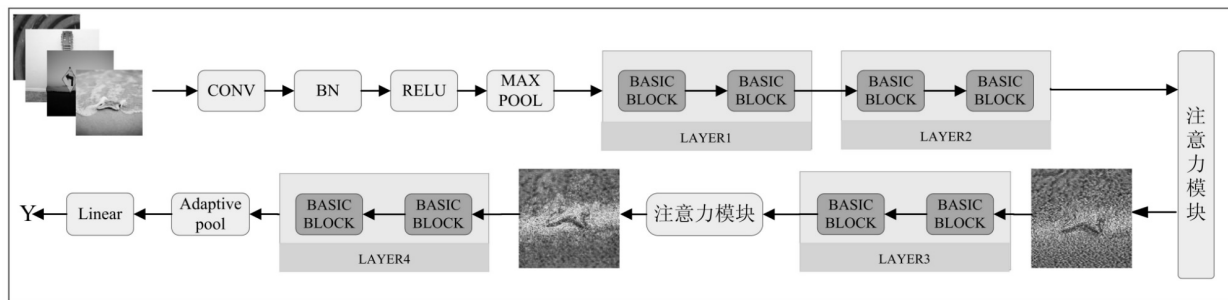


图 1 文中算法网络结构

### 1.2 残差网络

残差网络主要用于解决深层网络退化问题和梯度爆炸问题。一条主路径和跳层连接组成残差块,若干个残差块组成残差网络。

### 1.3 研究背景

近些年来,图像美学质量评估成为了一个研究热点。传统的美学质量评估方法一般是通过图像的对称性、三分法则和物体显著性等原则进行图像评估,这类方法往往需要大量的人工标定和专业化手工设计特征。Yan Ke 等人<sup>[4]</sup>提出了一种评估方法,包含图片简洁性、清晰度、颜色、对比度和平均亮度(曝光度)等特征描述,实现了基于分类的美学质量评估。Datta 等人在文献<sup>[5]</sup>中将底层特征(颜色、纹理、形状、图片大小

等)和高层特征(景深、三分法则、区域对比度)结合起来作为图像美学特征,得到了较好的效果。Luo 等人在文献<sup>[6]</sup>中提出将前景和背景进行分离,得到前-背景对比度特征,该方法最后采用的美学特征包括清晰度、对比度、亮度对比度、颜色的简洁性、和谐度和三分法则的符合程度等。其在 DPChallenge 数据集上进行了实验,具有 93% 的分类准确率。Wong 等人<sup>[7]</sup>在 Luo 方法的基础上进行了改进,采用基于显著度的方法提取前景,还融合了 Yan Ke<sup>[4]</sup>和 Datta<sup>[5]</sup>等人提出的一些特征,丰富了特征内容,取得了较好的实验结果。Marchesotti 等人<sup>[8]</sup>直接用尺度不变特征变换(SIFT)、Fisher 变换和局部颜色描述来进行美学图像评估,其在 MRSA 数据集上的实验表明,该方法评估

性能显著提升。

上述基于手工设计特征进行图像质量评估的方法,由于需要专业化手工设计特征,存在因手工设计特征不完备性造成评估效率低且鲁棒性不高等缺陷。为了改善传统方法的缺陷,许多学者提出了一些基于深度学习的图像质量评估方法。与传统手工特征提取方法不同的是,这类方法不使用手工设计特征,而是采用基于监督学习的神经网络模型来进行质量评估,从实验效果来看,与基于手工特征的方法相比,这类方法普遍提高了评估的准确度。Kong S 等人<sup>[9]</sup>提出了一种基于深度卷积神经网络图像美学排名方法,该方法的评分模型中包含了图像属性和内容,可以预测 1 到 5 级美学的属性和等级。该方法在 AVA<sup>[10]</sup>数据集上的实验取得了较好的分类效果。Talebi 等人<sup>[11]</sup>提出了一种深度 CNN 模型(NIMA)用于图像美学质量评估。NIMA 模型没有采用基于分类或者基于评分的美学量化指标,而是根据美学质量评估模型评估任意给定的图像的评分等级分布。NIMA 给每幅图像在 1 到 10 的范围内每一个可能的分数的可能性都进行赋值,输出一个分数分布直方图。这种方法能更好地表达美学评分问题本身具有的模糊性,所以在某种程度上提高了图像美学质量评估的精确度,实验表明,该算法在公共数据集 AVA 上获得了良好的性能测试结果,在 AVA 数据集中的极端美学品质图像表现更优。

考虑到图像的整体布局和内容相关性,Liu D 等人<sup>[12]</sup>认为图像中视觉元素之间存在相互依赖关系,将图像合成信息建模为其局部区域的相互依赖更为合理,所以应该将图像划分为局部区域的组合,并在这些区域上计算美学特征。基于这种思想,他们提出了一种基于组合感知的网络,该网络通过神经网络来学习视觉元素之间的相互依赖关系,在 AVA 数据集上的测试得到了较好的结果。

面对同一幅图像,不同的人可能会有不同的美学评价。图像美学质量评估往往带有大量的个人感情色彩,大部分美学质量评估方法一般是试图建立一个通用美学模型,没有考虑个人主观偏好。但是近年来个性化图像美学也逐渐受到学者们的关注<sup>[2,13]</sup>。Rodriguez-Pardo 等人<sup>[2]</sup>采用一种基于卷积神经网络模型的方法对具有不同类型的用户特定偏好进行建模,并利用该模型 PAA (personalized aesthetic assessment using residual adapters) 进行个性化的美学质量评估。该模型使用了残差适配器对特定用户的偏好进行建模。残差网络的加入不仅在一定程度上解决了网络退化问题和梯度弥散问题,而且大大加快了网络的效率。在 Flickr 数据集<sup>[3]</sup>上的实验体现了该方法基本能够实现个性化美学质量评估的功能。J Ren 等

人<sup>[13]</sup>提出了一种主动学习算法,并使用一个由所有者评定的真实个性相册的小数据集来优化网络。J Ren 提出的网络主要包括三部分:(1)通用美学模型;(2)图像属性特征网络;(3)图像内容属性网络。将(2)、(3)提出的图像特征进行融合,作为美学图像评估的偏置来调整网络,实验表明,该方法能有效的自主学习个性化偏好,且在定量比较方面优于现有方法。由于真实个性化数据集难以收集,文中最终决定在 PAA 网络基础上进行改进。

一幅图像在进行美学质量评估时,除了需要考虑到图像本身的属性特点和结构外,还需要考虑图像的主体内容对图像评估结果的影响,但大部分评估方法只从图像本身特征和整体布局进行评估,没有考虑到图像主体的突出性及其位置的合理性对美学质量评估的影响。一般来说,视觉显著性最高的区域或者显著物体对应着图像中最重要的内容,即图像主体内容,所以显著性分析也被用于美学质量评估。Tu Y 和 Niu L 等人在文献[14]中使用一个全卷积网络来生成能同时感知照片构图和物体显著性的美学评分图,其模型在实验中展现出了较好的性能。但是,该方法在进行图像显著性分析时,只考虑到了视觉显著性最高的区域,而忽略了其他区域,导致评估结果不够全面。

综上,文中提出了一种基于注意力机制的个性化图像美学质量评估方法。该方法在个性化残差网络的基础之上加入注意力机制来进行显著性分析,按显著性的高低对图像区域分配对应权重,这样不仅考虑了图像主体的突出性(显著性),而且还保留了图像显著性低的区域的影响力,使得各个区域对评估结果影响更加合理。文中提出的方法既考虑了图像内容偏好(个人美学质量评估的主观偏好),又考虑到对图像构图的偏好(即主体的突出性),在 Flickr 数据集上实验表明,与 Rodriguez-Pardo 提出的 PAA 方法<sup>[2]</sup>相比,文中方法取得了更好的评估准确率。

## 2 基于注意力机制和个性化的网络模型

假设存在  $N$  个用户,其序号分别为  $1, 2, \dots, N$ 。给定一个训练数据集  $K = \{D_i\}_{i=1}^N$ , 其中  $D_i$  表示第  $i$  个用户的训练样本集合,  $D_i = \{(x_{i,1}, s_{i,1}), \dots, (x_{i,N_i}, s_{i,N_i})\}$ ,  $N_i$  表示第  $i$  个用户的样本个数,  $s_{i,j}$  表示第  $i$  个用户对其第  $j$  个样本图像  $x_{i,j}$  的美学质量评分(真实), 训练数据集  $K$  的大小为  $\sum_{i=1}^N N_i$ 。算法的目标是:通过数据集  $K$ , 训练生成一个模型  $M$ , 使得对任意图像  $x'$ , 能根据  $M$  估计出用户  $m(m \in [1, N])$  对  $x'$  的美学质量评分  $s'_m$ 。

为了实现上述目标,文中对 Rodriguez-Pardo 提出



的 PAA 方法<sup>[2]</sup>进行了扩展,提出一种基于注意力机制的个性化残差网络。算法模型中融入个性化模块和注意力模块。网络结构如图 1 所示。该网络以 ResNet18 为基线<sup>[15]</sup>,除了在每个 Basic block 都加入个性化模块

外,还在 layer2 和 layer3 之间,以及 layer3 和 layer4 之间加入了注意力模块。每个 Basic Block 的结构相同,如图 2 所示,均是在 ResNet18 原始 Basic block 模块中的基础模块  $C_i$  上加入了个性化(建模)模块。

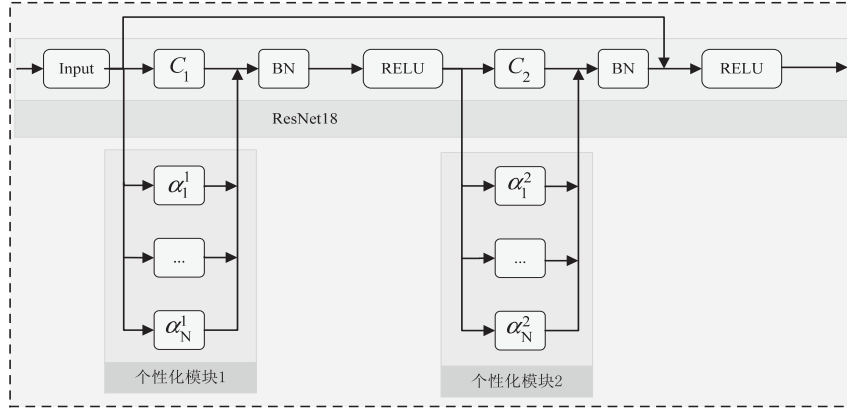


图 2 含个性化建模的 Basic Block 模块的结构

图 2 中  $C_i$  为基础模块,对第  $i$  层通用审美规则进行建模,其在个性化训练过程中保持权重不变。而  $\alpha_j^i$  层建模第  $i$  个用户的第  $j$  层审美规则偏移( $i \in [1, N], j \in [1, M]$ ,  $M$  可以根据基线网络规模进行调整,文中取值为 2),在用户  $i$  的个性化训练中优化  $\alpha_j^i$  的参数。 $C_i(i \in [1, 2])$  采用的是大小为  $3 \times 3$ 、步长为 1 的卷积核,  $\alpha_j^i$  采用大小为  $1 \times 1$ 、步长为 1 的卷积核。

网络训练过程分为两个阶段,第一阶段用训练数据集训练基线网络。第二阶段进行个性化训练。用户

$i$  的个性化数据集  $D_i$  是由用户  $i$  评估过的所有图像及其标签组成。当第  $i$  个用户进行个性化训练时,其对应的  $\alpha_i^k(k \in [1, 2])$  层通道开启,其他个性化层通道关闭,网络数据经过  $\alpha_i^k$  层和  $C_i$  层后进行融合,作为之后层的输入,这样训练之后每个  $\alpha_i^k$  就保留了第  $i$  个用户的个人偏好特征。当有新的用户加入模型参与训练时,只需保留 ResNet18 基础模块不变,新增  $\alpha_i^k$  模块即可,网络结构灵活,具有较好的可拆卸性和可扩展性。

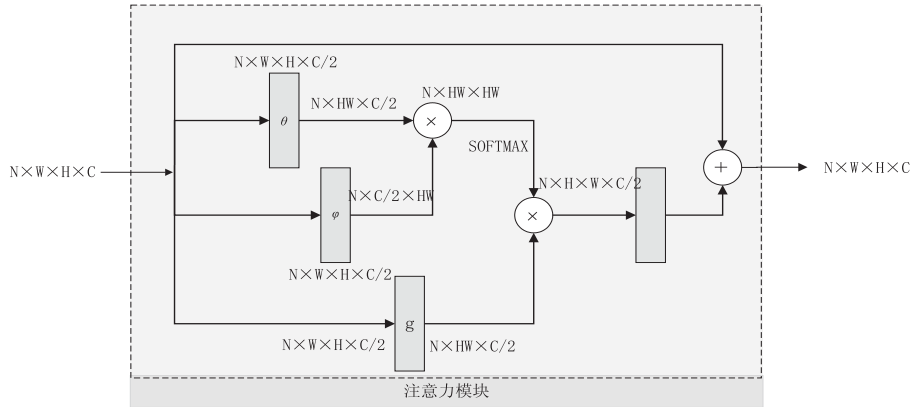


图 3 注意力机制模块内部结构

文中用非局部操作 non\_local<sup>[16]</sup> 模块实现注意力模块,其内部结构如图 3 所示。引入注意力机制主要是为了增强图像构图对最终美学质量评估的影响。考虑到深层网络主要包含图像高层语义信息,位置信息量较少,而最浅层网络图像信息较为驳杂稀疏,所以文中设置两个注意力模块,分别放置在较浅层卷积层(layer2 和 layer3)之后。注意力模块将输入特征信息进行分块处理,采用特征矩阵逐乘的方式,得到每个位置像素的权重,从而实现权重的重新分配,对图像评估影响较大的位置区域的像素分配较大的权重,影响力

较小的位置分配较小的权重。经过该模块处理后的特征如图 1 中的热力图所示,经第二个注意力模块处理得到的热力图信息更为详细,主体信息更为明确。

在 layer4 后网络又经过三个全连接层和两个正则化层,在最后一层全连接层后输出归一化后的美学审美评分。其中穿插了随机丢弃(dropout)部分神经元的方法,在一定程度上降低了过拟合的几率。此外,文中在训练过程中将卷积与其紧接的批标准化(batch normalization)进行折叠封装。由于一般卷积后接 BN 层是没有偏差(bias)的,这样将两部分折叠,加快了前

向传播的速度,也不会影响最终的实验结果。文中网络使用的损失函数是均方误差 MSE,定义如下:

$$\text{MSE}(s, \hat{s}) = \frac{\sum_{i=1}^n (s_i - \hat{s}_i)^2}{n} \quad (1)$$

式中,  $s_i$  表示在一个 batch 中第  $i$  个样本的真实美学质量评分值,  $\hat{s}_i$  表示网络对第  $i$  个样本估计得到的美学质量评分。通过该损失来评估网络的性能。在训练中,选用 Adam 优化器,用动量和自适应学习率来加快收敛速度。

### 3 实验

#### 3.1 数据集

为了对图像进行美学质量评估,前人构造了许多美学质量评估数据集,包括 FLICKR<sup>[3]</sup>、AVA<sup>[10]</sup>、CUHKPQ<sup>[17]</sup>等,这些数据集中的图片大多数来自于网络上的图像分享网站 DPChallenge、Photo. Net 等。文中使用 Flickr 数据集进行相关测试。Flickr 数据集属于社交数据集,包含 1 亿+张图片,其中除了图像的评分外,还包含各个用户之间的关系和共属的兴趣小组等信息,在进行个性化建模前需要将数据集按用户进行整理,整理得出多个个性化数据集,每个个性化数据集包括某用户 id、该用户评估的所有图像 id 及其评分分值。考虑到图像的大小可能并未统一规格,需对数据集进行统一处理,将图像缩放至 256×256 大小,然后对图像进行随机水平翻转,并将图像随机裁剪为网络模型输入大小 224×224,然后进行正则化处理,随机裁剪和翻转增加了数据集的多样性,可有效避免过拟合。为了加快训练过程中数据集的读取速度以及防止数据集散列存放占用大量内存空间,将数据集处理好之后以 TFRecord 格式进行存储。TFRecord 格式采用二进制编码,占用空间小(只占用一个内存块),加载数据集时,只需要一次性加载这个二进制文件即可,简单、快速。对 Flickr 这种大型数据进行训练时,可以将数据分成多个 TFRecord 文件,来提高处理效率。

#### 3.2 网络结构和参数设置

网络输入初始向量为 224×224,紧接着进行卷积,卷积核大小为 7×7,步长为 2,之后是 4 个 layer 层,每个 layer 层包含两个 Basic Block 模块,每个 Basic Block 是一个 bottleneck,其中使用到两个卷积:大小 3×3、步长为 1 的卷积核 conv 和大小 1×1、步长为 1 的卷积核 Rconv。Basic Block 模块中的通用美学评估模块  $C_i$ ,通过使用专家评估的图像及标签集合对其进行预训练并在之后的训练中保持权重不变;Basic Block 模块中的个性化模块  $\alpha_j^i$  则通过使用 Flickr 数据集中收集到的个人用户的评估数据集对其进行训练。layer2 和

layer3 后的注意力机制模块为非局部操作 non\_local 模块,对构图偏好进行特征的加权总和计算,能扩大构图偏好部分对最终评估结构的影响,增强置信度。特征图  $f$  进入注意力 attention 块,该模块通过 1×1×1 的卷积对输入的特征图  $f$  进行线性映射,压缩通道数,得到  $\theta$ 、 $\varphi$ 、 $g$  特征,如图 3 所示, $\theta$  与  $\varphi$  进行矩阵点乘得到自相关特征  $f_i$  ( $f_i$  表示每帧中每个像素对其他所有帧中所有像素的关系),然后对自相关特征进行 softmax 得到自注意系数,最后将自注意系数对应乘回特征矩阵  $g$  中,扩展通道数,与原输入特征图  $f$  做残差运算,得到非局部操作 non\_local 模块的输出。实验对注意力模块在网络中不同位置带来的性能差异进行了测试。测试发现 non\_local 模块加载到 layer2 层和 layer3 层后得到的结果最佳,所以后续实验网络中在 layer2 层和 layer3 层后加入注意力模块。

另外考虑到图像评估阶段中个性化的问题,在图像特征提取阶段中加入了残差块,通过残差块保存少量的用户偏好信息,将瓶颈层(bottleneck)与残差块相融合,从而达到个性化评估的效果。

#### 3.3 实验设计

文中使用 torch 在具有 6 GB GPU 内存的 NVIDIA GeForce GTX 1060Ti 上的 Python 环境中训练和测试该算法,使用 Adam 优化器来进行参数学习以及网络优化,Batchsize 设置为 30,epochs 设置为 200,总共迭代次数为 22 200 次。在 Flickr 数据集上进行实验,分别对残差块和注意力机制模块的有效性进行对比评估。Flickr 数据集总共包含 40 500 幅图像,210 个用户参与评分。文中选取 37 个用户作为测试集,共 4 739 幅图像,剩下的所有数据作为训练集。为保证实验过程中单一变量原则,所有实验中的学习率统一设置为 0.1,并在每训练 10 轮 epochs 后降低 90%。在训练开始时,使用较大的学习率能使得网络快速收敛,随着训练的进行,逐渐降低学习率有助于找到最优解。为了得到稳定的模型,并兼顾效率,文中选择三折交叉验证进行个性化的实现。以第  $i$  个用户为例,第  $i$  个用户的个性化数据集  $D_i$  会在实验中依据三折交叉验证原则进行划分,分为训练集合和测试集合。

文中使用的美学质量评估指标为 SRCC<sup>[18]</sup> (Spearman 等级相关系数) 范围为  $[-1, 1]$ , 定义如下:

$$\rho = 1 - \frac{6 \times \sum_{i=1}^N (r_i - \hat{r}_i)^2}{N^2(N-1)} \quad (2)$$

通过该指标来展现出模型预测值和真实值的相近程度,  $\hat{r}_i$  表示第  $i$  幅图像的预测分数,  $r_i$  表示第  $i$  幅图像的实际分数,SRCC 值( $\rho$ )越大,表明正相关程度越

大,模型的性能越好。

### 3.4 实验结果与分析

对比实验在 Flickr 数据集上进行,文中与 PAA (personalized aesthetics with residual adapters) 网络进行实验对比,结果如表 1 所示。

表 1 与 PAA 网络在 Flickr 数据集实验结果对比

算法	SRCC( mean)	SRCC( std)
PAA	0.631	0.115
Ours	0.659	0.110

由表 1 可以看出,在同一数据集 Flickr 上,PAA 网络取得 SRCC 均值为 0.631 的结果。文中以 ResNet18 网络为基线网络,在该网络基础上融入了残差块并集成了注意力机制,并通过选取 ResNet18 预训练模型前四层权重对网络进行训练拟合,引入注意力机制增强了特征提取过程中提取特征的有效程度,相当于权重的重新分配。通过对图像评估影响较大的区域分配较大权重,从而提高了美学质量评估的准确度,同时文中网络相较于 PAA,稳定性略有增加。

另外进行了消融实验,文中在 Flickr 数据集上依次验证了残差块和注意力机制对最终的美学质量评估的改进,对比实验结果如表 2 所示。

表 2 在 Flickr 数据集消融实验结果对比

算法	SRCC( mean)	SRCC( std)
Baseline	0.561	0.223
Baseline+Residual	0.593	0.205
Baseline+Attention	0.631	0.183
Ours	0.659	0.110

文中的基线网络使用的是 ResNet18 网络,使用 ResNet18 预训练模型对模型进行训练后,在 AVA 上展现的性能为 SRCC 均值为 0.561。

在基线网络上加入残差块后,由于残差块用来学习用户的特定偏好,而 Flickr 数据集中用户量较大,每个人的偏好不尽相同,这就造成了虽然评估效果有了一定的改善,但是 SRCC 值波动依旧较大。另外考虑到图像构图本身(主体的位置、布局等)对图像质量评估有影响,为了在特征提取过程对美学质量评估影响较大的部分特殊关注,在基线网络上加入注意力机制模块,为这部分区域分配较大的权重,提高此区域对最终美学质量评估的影响力,以便提升结果的可信度。从在 Flickr 数据集上的表现来看,评估效果有了较大的提高。最终,文中在基线网络上融合了注意力机制和残差块,在 Flickr 数据集上取得了 SRCC 值为 0.659 的结果,通过 SRCC( std) 来看该模型有了比较稳定的表现。

由在 Flickr 数据集上的消融实验表明,残差块和

注意力机制的引入在一定程度上改善了图像质量评估效果。

表 3 与 NIMA 网络在 AVA 数据集上实验结果对比

算法	SRCC( mean)	SRCC( std)
NIMA	0.612	0.218
Ours	0.637	0.159

另外文中网络模型在 AVA 数据集上也做了测试,虽然 AVA 数据集中缺乏用户与评估图像间的关联信息,文中网络无法发挥出其个性化优势,所以将文中算法作为通用美学评估方法和最流行的通用美学评估方法 NIMA 进行了对比(如表 3 所示)。从 AVA 数据集上的测试结果显示,文中方法取得了 SRCC 均值为 0.637 的结果,比 AVA 稍好。从网络评估的稳定性而言,文中网络在 AVA 数据集上的表现比 NIMA 网络在 AVA 数据集上的表现来说要稳定。整体上看,文中网络表现稍优于 NIMA 的网络。

综合以上两个实验对比,文中在基线网络 ResNet18 上融合注意力机制和残差块后,网络性能有了明显的改善。

## 4 结束语

文中提出了一种基于注意力机制和个性化的图像美学质量评估网络,该网络主要以 ResNet18 为基线网络,在其基础上进行改进,在每个 layer 层中融入了残差块,并在第二个 layer 层和第三个 layer 层后加入了注意力机制模块,每个 layer 层的输出都是经过当前层处理(包括残差块)后与当前 layer 层输入相加后的结果,这样能更好地保证特征提取的合理性。实验结果表明,模型在对图像进行质量评估时保留了个人偏好,对个人的审美有了较大的改善,其次注意力机制的引入使得图像评估指标有了较大提升,相比于现有方法,文中提出的模型性能稍强,仍有很大的改进空间。文中只是在基线网络 ResNet18 上做了简单的改进,可以考虑将基线网络替换为 Inception 网络或者是其他网络,ResNet18 只是对最终的结果进行了激活,而 Inception 在每次卷积后都进行了激活,这样提高了神经网络对模型的表达能力,与此同时,Inception 卷积后对通道数进行了叠加,提升了模型的拟合能力。另外可考虑扩充数据集,现有的数据集缺乏对个人偏好的整理,可自己去收集整理出有关个人偏好的数据集。

### 参考文献:

- [1] 金鑫,周彬,邹冬青,等. 图像美学质量评价技术发展趋势[J]. 科技导报,2018,36(9):36-45.
- [2] RODRÍGUEZ-PARDO C, BILEN H. Personalised aesthetics with residual adapters[C]//Iberian conference on pattern

- recognition and image analysis. Madrid, Spain; Springer, 2019;508–520.
- [3] LIU D, HUA X, YANG L, et al. Tag ranking[C]//Proceedings of the international conference on world wide web. Madrid, Spain; [s. n.], 2009;351–360.
- [4] KE Y, TANG X, JING F. The design of high-level features for photo quality assessment [C]//IEEE computer society conference on computer vision and pattern recognition (CVPR'06). New York, NY, USA; IEEE, 2006;419–426.
- [5] DATTA R, JOSHI D, LI J, et al. Studying aesthetics in photographic images using a computational approach [C]//Computer vision – ECCV 2006. Graz, Austria; Springer, 2006;288–301.
- [6] LUO Y, TANG X. Photo and video quality evaluation; focusing on the subject [C]//European conference on computer vision. Berlin; Springer, 2008;386–399.
- [7] WONG L K, LOW K L. Saliency-enhanced image aesthetics class prediction [C]//IEEE international conference on image processing (ICIP). Berkeley, CA, USA; IEEE, 2009.
- [8] MARCHESOTTI L, PERRONNIN F, LARLUS D, et al. Assessing the aesthetic quality of photographs using generic image descriptors [C]//IEEE international conference on computer vision. Barcelona, Spain; IEEE, 2011;1784–1791.
- [9] KONG S, SHEN X, LIN Z, et al. Photo aesthetics ranking network with attributes and content adaptation [C]//Computer vision – ECCV 2016. Amsterdam, The Netherlands; Springer, 2016;662–679.
- [10] PERRONNIN F. AVA: a large-scale database for aesthetic visual analysis [C]//IEEE conference on computer vision & pattern recognition. Providence, RI, USA; IEEE, 2012;2408–2415.
- [11] TALEBI H, MILANFAR P. NIMA: neural image assessment [J]. IEEE Transactions on Image Processing, 2018, 27(8): 3998–4011.
- [12] LIU D, PURI R, KAMATH N, et al. Composition-aware image aesthetics assessment [C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV). Snowmass Village, CO, USA; IEEE, 2020;3559–3578.
- [13] REN J, SHEN X, LIN Z, et al. Personalized image aesthetics [C]//IEEE international conference on computer vision. Venice, Italy; IEEE, 2017;638–647.
- [14] TU Y, NIU L, ZHAO W, et al. Image cropping with composition and saliency aware aesthetic score map [C]//Proceedings of the AAAI conference on artificial intelligence. New York, USA; AAAI, 2020;12104–12111.
- [15] YU X, WANG S H. Abnormality diagnosis in mammograms by transfer learning based on ResNet18 [J]. Fundamenta Informaticae, 2019, 168(2–4): 219–230.
- [16] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Salt Lake City, Utah; IEEE, 2018;7794–7803.
- [17] 罗伟, 王晓刚. 基于内容的照片质量评估 [C]//2011 年国际计算机视觉会议. 巴塞罗那; 出版者不详, 2011;2206–2213.
- [18] BELIAKOV G, YEARWOOD J, KELAREV A. Application of rank correlation, clustering and classification in information security [J]. Journal of Networks, 2012, 7(6): 935–945.