

基于 CPNet 的相对图像质量评估

李凯文 徐琳 陈强

南京理工大学计算机科学与工程学院 南京 210094

(kevinleekaiwen@foxmail.com)



摘要 对于两幅不同质量的图像,人眼视觉系统(Human Visual System,HVS)能够比较容易地区分两幅图像间的质量差异,因此通过模拟 HVS 来判断两幅图像的相对质量比给出图像的绝对质量分数更加准确。文中提出了一种用于评估图像间相对质量的 CPNet(Compare-net)模型,该模型是一种分数无关类型的算法,利用图像组合的形式解决数据量的限制,相比绝对质量分数标签,提出的相对质量标签以及相对质量顺序标签具有更广阔的应用场景,并且获取方式更加方便、准确。首先,通过分析卷积神经网络结构相关参数对网络性能的影响,来构建合理的网络基础结构;其次,以双通道输入网络和设计特征求差的方式得到两幅图像的质量差异特征,并结合图像对相对质量标签来完成分类学习;最后,通过在公共数据库上的实验证明了该算法的精度优于其他算法。所提算法在相同参考图像类型实验中分别取得了 0.971 和 0.947 的最优精度;在不同参考图像类型实验上也取得了很有竞争力的精度,分别为 0.926 和 0.860。另外,设计了三通道网络并进行实验来探究将所提算法扩展到多通道的可能性。

关键词: 图像质量评估;CPNet;相对质量顺序;质量差异;卷积网络

中图分类号 TP391.4

Relative Image Quality Assessment Based on CPNet

LI Kai-wen, XU Lin and CHEN Qiang

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract For two images with different quality, the human visual system (HVS) can easily distinguish their quality difference. Thus, it is more accurate to judge the relative quality of two images by simulating HVS than to give the absolute quality score of images. A CPNet (Compare-net) model for evaluating the relative quality between images is proposed in this paper. It is a score-independent algorithm that uses the form of image combination to solve the limitation of data volume. Compared with the absolute quality score label, the proposed relative quality label and relative quality order label have a broader application scenario than the absolute quality score label and are more convenient and accurate to obtain. Firstly, by analyzing the influence of convolutional neural network structure related parameters on network performance, a reasonable network infrastructure is constructed. Secondly, the quality difference characteristics of two images are obtained by the methods of two-channel input network and the feature differentiation, and the classification learning is completed by combining the relative quality labels of the image pairs. Finally, experiments on public database show that the accuracy of the proposed algorithm is better than that of other algorithms. CPNet achieved the best accuracy of 0.971 and 0.947 in the same reference image experiment, and also achieved a very competitive accuracy in different reference image experiments, 0.926 and 0.860 respectively. In addition, a three-channel network is designed and experiments are carried out to explore the possibility of extending the proposed algorithm to multiple channels.

Keywords Image quality assessment, CPNet, Relative quality order, Quality difference, Convolutional network

1 引言

图像作为传递信息的主要方式之一,在人们的生活和工作中发挥着重要作用;并且随着科技的发展,图像的获取途径越来越多,图像在生活中已经随处可见。但是,图像在获取、处理、传输等一系列操作中,存在着不同类型、不同程度的失

真,这些失真导致了图像质量的降低。为了对这些失真的图像进行处理并评估其失真程度,图像质量评估算法应运而生。

图像质量评估是图像处理、计算机视觉等领域中的一项基础而关键的任务,在许多场景中发挥着重要的作用,如图像压缩^[1]、图像增强或复原^[2-3]等。由于图像质量主观评估在实际应用中存在难度较高且客观条件受限等缺点,其应用场景

收稿日期:2019-09-06 返修日期:2019-11-21 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61671242)

This work was supported by the National Natural Science Foundation of China (61671242).

通信作者:陈强(chen2qiang@njust.edu.cn)

较为狭窄。因此,图像质量评估算法的研究基本聚焦于通过模拟人眼视觉系统进行自动、客观的评估。

根据算法对参考图像的依赖性,客观图像质量评估算法可以分为3类:1)全参考图像质量评估算法(Full-Reference Image Quality Assessment, FR-IQA);2)半参考图像质量评估算法(Reduced-Reference Image Quality Assessment, RR-IQA);3)无参考图像质量评估算法(No-Reference Image Quality Assessment, NR-IQA)。其中,全参考图像质量评估算法研究得最早,主要通过失真图与原始图的逼真度得出评估结果,如最早的基于均方误差的峰值信噪比(Peak Signal to Noise Ratio, PSNR)^[4],以及 Zhang 等以结构相似度(Structural Similarity Index, SSIM)为基础提出的 FSIM 方法^[5]。其中,FSIM 方法从相位一致性、梯度幅值和色彩变化3方面对失真图像的质量进行评估。但是,不管是全参考图像还是半参考图像,都没有摆脱参考图像的限制,导致其应用范围相对狭窄。无参考图像质量评估算法在没有包含原始图像的基础上直接对失真图像进行质量评估,该方法摆脱了对原始参考图像的依赖性,因此具有最为广阔的应用场景,近年来成为了研究的热点。

根据特征计算方法,可以将无参考图像质量评估算法分为4类。1)统计特定失真类型特征的算法。该类算法主要根据某一类失真给图像带来的特性统计计算其失真程度,从而得到质量分数^[6-11]。但是,该类算法需要假定失真类型且仅含一种失真类型,因此对应的使用范围较为局限。2)基于机器学习的方法^[12-14]。这类算法主要使用各类空域中的特征表征图像质量,根据这些特征汇总得到高维向量,然后使用机器学习算法对这些特征进行训练,预测图像质量。3)基于自然场景统计特性(Natural Scene Statistics, NSS)的评估算法。对自然图像进行标准化后可以发现其存在一定的统计规律,该规律符合广义高斯分布(Generalized Gaussian Distribution, GGD),而失真会影响 GGD 的形状参数,通过对这类形状参数进行统计对比可以得到失真程度。根据建立模型时是否使用平均主观得分(Mean Opinion Score, MOS),这类方法可以分为分数相关(Opinion-Aware, OA)^[15-23]和分数无关(Opinion-Unaware, OU)^[24-25]两类。4)基于深度学习的评估方法^[26-32]。近年来,深度学习得到了迅猛的发展,其中卷积神经网络在图像质量评估领域也取得了重大的进展。卷积神经网络的特点在于将特征提取和网络回归良好地结合在一起,并且在网络框架中采用反向误差传播的方法将整个网络作为一个整体进行训练,这样就可以在设计的网络模型中加入一些新的技术,如各类激活函数。

但是,上述各种算法都是直接输出图像的绝对质量分数,而对于两幅不同的图像,HVS能够比较容易地区分两幅图像间的质量差异,如图1所示,右边图像显然比左边图像的质量好。因此,HVS判断两幅图像的相对质量相比给出图像的绝对质量分数来说更加准确。但是,无参考图像质量评估算法只能将两幅图像当作单独的个体,分别进行绝对质量预测后才能得到两幅图像的相对质量,这与HVS的工作机制并不相同。而全参考图像质量评估方法需要参考图像的先验知识,HVS在不需要参考图像的前提下就能对两幅失真图像进

行相对质量的判断。因此,图像相对质量评估与全参考图像质量评估和无参考图像质量评估有着本质区别。



(a)具有相同参考图像的两幅图像



(b)具有不同参考图像的两幅图像

图1 相对图像质量评估示例图

Fig. 1 Example diagram of relative image quality evaluation

除此之外,无参考图像质量评估算法,尤其是分数相关的无参考图像质量评估算法,在完成图像质量评估任务时,需要的客观图像质量评估数据库均是采用主观评估方法对数据库中的图像进行质量分数标注。但是,当使用主观评估方法来构建客观图像质量评估数据库时,对于同一幅图像,因为主观评估个体之间存在着质量感知差异,如生理差异、心理差异或者个人喜好等,不同个体往往会给出不同的主观评估分数。也就是说,使用分数来进行无参考图像质量评估时存在着一定的误差。后续用无参考图像质量评估算法直接输出预测分数再进行比较,就进一步扩大了两幅图像的相对质量预测误差;同时,当使用主观评估方法来构建客观图像质量评估数据库时,采集主观评估意见往往会耗费大量的时间和人力/物力资源,并且其过程和结果不可重现。以上这些缺点在很大程度上限制了主观质量评估方法在实际场景中的应用。

在生活中,可以比较快速、准确地获取用于相对质量评估的相对质量标签。人眼视觉系统可以快速、准确地分辨出两幅图像之间的相对质量,相比主观评估方法中通过打分并收集意见分数去得到一幅图像的质量分数,其节省了大量的人力和物力;同时,在实际应用中,相对质量评估适用于某些很难获取主观评分的场景。

本文通过探究图像间差异性在网络结构的体现,提出了一种基于卷积神经网络的多输入相对图像质量评估算法。首先,对双通道网络进行研究和设计,通过在网络内设计特征求差的方式得到两幅图像的质量差异特征,结合图像间相对质量标签完成分类学习,并利用公共数据库进行验证。其次,将双通道网络扩展为三通道,同样通过特征求差的方式得到3幅图像两两间的差异,并结合图像对相对质量顺序标签完成分类学习。最后,通过实验验证了算法的有效性。

本文算法的优势在于:该算法是一种分数无关类型的算法,即不需要分数标签,而是采用相对质量标签或者相对质量顺序标签来进行网络的训练;同时,利用图像组合的方式能够摆脱数据量的限制,如含有982幅图像的LIVE数据库,通过两两组合可以得到 982×982 对的两幅图像的相对质量数据库;相对于输出绝对分数再进行比较得到相对质量标签或者

相对质量顺序标签,直接输出图像组合之间的相对质量标签或者相对质量顺序标签更加准确且更加方便;除此之外,遇到一些很难获取主观评分的场景时,采用本文方法只需要获得图像组合之间的相对质量或者相对质量顺序就可以进行训练,不需要花费大量的人力和物力去对图像进行打分;同时,因为人眼视觉特性,直接得到图像组合的相对质量或者相对质量顺序相比获取图像的绝对质量分数来说更加准确且快速。

2 相关工作

早期的无参考图像质量评估算法通常只能对某一特定类型的失真图像进行评估。例如,文献[6-7]计算 JPEG 压缩失真图像质量,文献[8-9]计算块效应失真图像质量,文献[10-11]计算模糊失真和噪声失真图像质量。后来的无参考图像质量评估算法则是针对多种失真类型。基于自然场景统计特性(NSS)来进行无参考图像质量评估是该领域内应用范围最广泛的一种算法,如基于离散余弦变换后的相位谱和频域谱特征^[17-19]、基于小波变换后的各频段特征^[20]、基于图像空间域的相关特征^[23]来进行拟合得到 NSS 特征。

近年来,深度学习开始成为人工智能领域的研究热点,且在计算机视觉领域取得了很大的成就。Kang 等^[26]于 2014 年首次将卷积神经网络应用到无参考图像质量评估中,其以图像块为网络输入,并使用 ReLU 和 Dropout 等技术对网络进行优化,最后通过对得到的属于一幅图像的所有图像块的局部质量分数估计值求平均值得到该图像的全局质量分数。此后出现了很多基于卷积神经网络的无参考图像质量评估算法。Kim 等^[27]提出了 BIECON 算法,该算法仿照全参考图像质量评估方法来生成局部图像块的质量分数,即在预处理中使用全参考方法来为不同图像块赋予不同的权值。Li 等^[29]通过在 ImageNet 数据库中进行预训练来解决数据量的限制。Bosse 等^[30]提出使用 RGB 图像块作为网络输入,以增加颜色信息。Gu 等^[31]将可学习池化引入到无参考图像质量评估中,其核心是以数据驱动的方式学得既有效又符合视觉注意力的池化方法。

目前也有一些通过图像质量排序来实现无参考图像质量评估的方法。Gao 等^[32]将无参考图像质量评估转换为一种利用偏好图像对的分类学习问题,采用特征融合和多核学习进行图像质量评估。Liu 等^[33]提出了一种数据增广方法,通过采用不同的失真方法并结合不同的失真强度使原始图像失真,从而得到大量失真图像,将这些图像两两组合后传入 Siamese 网络进行训练;预训练后,取 Siamese 网络的一支并使用 fine-tuning 方法来进行微调,以提升质量分数的预测精度。这些方法通过使用排序的思想来解决无参考图像质量评估存在的问题,本质上还是输出图像的预测质量分数。本文提出了一种基于图像质量差异的 CPNet(Compare-net)用于图像相对质量的评估,其核心为输出相对质量标签或者相对质量顺序,即根据输入的图像对得到图像间的相对质量排序结果,而不是图像的绝对预测质量分数,与上述提到的方法有显著不同。后续的实验也证明,相比直接输出图像的质量分数再通过比较得到图像组合之间的相对质量或者相对质量顺序,本文算法更加准确且更加方便。

3 CPNet

3.1 双通道 CPNet

双通道 CPNet 网络是一个二分类的网络,其输入和输出都是相对质量标签而不是绝对质量分数,即判断两幅图像中哪幅图的质量更好。双通道 CPNet 的网络结构如图 2 所示。

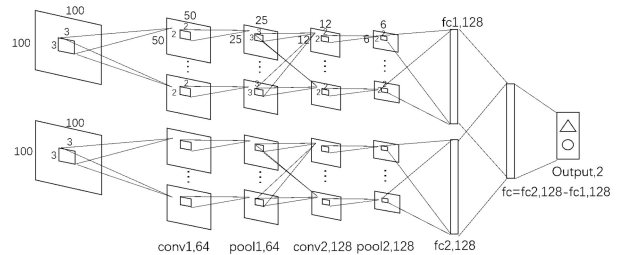


图 2 双通道 CPNet 的结构

Fig. 2 Architecture of dual-channel CPNet

3.1.1 网络设计

在双通道 CPNet 中,先按照通道数将两幅待评估图像分开,每幅图像均经过 2 个卷积层、2 个池化层和 1 个全连接层,将两个全连接层相减后即可得到两幅图像的特征差异,再通过一个全连接层得到预测的标签,最后使用 softmax 损失函数计算误差。其中,两个卷积层的卷积核尺寸都为 3×3 ,步长为 2,池化层为 2×2 的最大池,因此双通道 CPNet 的输出结构为 $100 \times 100 - 50 \times 50 - 25 \times 25 - 12 \times 12 - 6 \times 6 - 128 - 2$ 。两层 3×3 的卷积层可以较好地提取图像特征,由于第二层卷积提取的图像特征比第一层更抽象、复杂,并且根据对其他网络结构的研究,卷积核的个数一般随着卷积层的增加而增加,因此本文在第一个卷积层设置 64 个卷积核,在第二个卷积层设置 128 个卷积核。

Dropout、L1/L2 权重衰减和 batch normalization 批量标准化都是目前较流行的用于防止过拟合的正则化方法,但是本文在实验过程中发现是否添加 Dropout 和 batch normalization 对实验结果并没有太大影响,因此本文选择 L2 权重衰减的正则化策略。

3.1.2 输入样本的预处理

将两幅不同的待评估图像以及相应的相对质量标签作为输入样本,两幅图像以两个通道的形式作为一个整体,若第一通道的图像质量较好,则样本标签为 0,反之则为 1。文献[26]将输入尺寸设置为 32×32 ,并认为由于仿真图像失真均匀,因此每个图像块的质量都是相同的;但是当图像块过小时,图像块质量的这种赋值方式容易引入误差。本文认为图像块尺寸的增大,也可以在一定程度上减小这种误差,因此本文选择的输入样本尺寸为 100×100 ,并且后续实验中也证明了这个尺寸是合理的。

图像数据的预处理在深度学习技术中是非常重要的,其有助于更快到达全局最优值,如数据标准化就是一种预处理技术,其可以给更多算法带来性能上的提升。本文采用如下的局部归一化方法:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (1)$$

$$\mu(i, j) = \sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} I(i+p, j+q) \quad (2)$$

$$\sigma(i, j) = \sqrt{\sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} (I(i+p, j+q) - \mu(i, j))^2} \quad (3)$$

其中, $P=Q=5$; $I(i, j)$ 表示图像 I 上 (i, j) 处的灰度值; C 是一个正常量, 避免除数为 0。相比全局归一化, 失真对局部归一化的统计特性的影响更显著, 因此, 目前较流行的基于卷积神经网络的图像质量评估算法都对图像进行了局部归一化。

3.1.3 网络训练

本文输入的是尺寸为 100×100 的图像对及对应的相对质量标签。在 LIVE 数据库中, 根据两幅图像的平均主观得分差异 (Differential Mean Opinion Score, DMOS) 值可以得到对应的平均主观意见分 (Mean Opinion Score, MOS)。MOS 值越高, 图像质量就越好, 从而可以得到该图像对的相对质量标签。在测试阶段, 对图像对中所有的 100×100 大小的图像块对进行相对质量预测, 并统计图像对 0 和 1 的个数, 将个数较多的标签赋给该图像对, 即若质量较好的图像块超过一半, 则该图像的质量就高于另一幅图像。最后, 预测正确的图像对占所有测试集的比例即为算法的正确率。

3.2 三通道 CPNet

三通道 CPNet 是一个多分类的网络, 其输入和输出都是相对质量顺序标签, 即标签表示 3 幅图像的质量高低顺序。三通道 CPNet 的网络结构如图 3 所示。

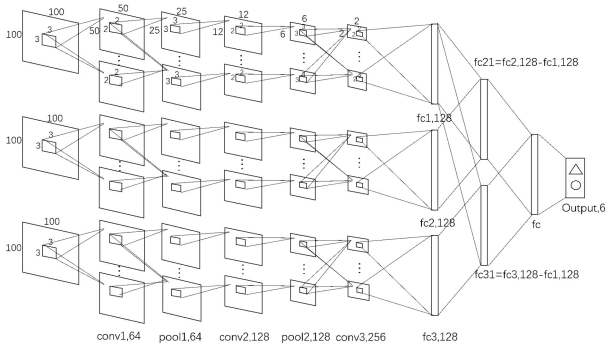


图 3 三通道 CPNet 的结构

Fig. 3 Architecture of three-channel CPNet

3.2.1 网络设计

与双通道 CPNet 网络相似, 在三通道 CPNet 网络中, 先按照通道数将 3 幅待评估图像分开, 每幅图像均经过 3 个卷积层、2 个池化层和 1 个全连接层, 将后两个全连接层分别与第一个全连接层相减后即可得到后两幅图像与第一幅图像的特征差异; 分别对两个特征差异进行全连接, 之后进行拼接, 并再次经过一个全连接层得到预测的标签; 最后使用 softmax 损失函数计算误差。三通道 CPNet 的每个通道中, 3 个卷积层的卷积核尺寸都为 3×3 , 步长为 2, 池化层为 2×2 的最大池, 因此三通道 CPNet 的输出结构为 $100 \times 100 - 50 \times 50 - 25 \times 25 - 12 \times 12 - 6 \times 6 - 2 \times 2 - 128 - 128 - 10$ 。因为 3 幅图像之间的特征差异较 2 幅图像之间的特征差异更复杂, 所以使用三层 3×3 的卷积层可以更好地提取图像特征, 在第一个卷积层设置 64 个卷积核, 第二个卷积层设置 128 个卷积核, 第三个卷积层设置 256 个卷积核。同样, 选择 L2 权重衰减的正则化策略。

3.2.2 输入样本的预处理

将 3 幅不同的待评估图像以及相应的相对质量顺序标签作为输入样本, 选择的输入样本尺寸为 100×100 。预处理时, 采用双通道 CPNet 的设置。相对质量顺序标签的设置如表 1 所列 (1, 2, 3 分别代表 3 幅图像, 相对质量顺序从左到右表示图像质量分数依次降低)。

表 1 相对质量顺序与对应标签

Table 1 Relative quality order and corresponding label

相对质量顺序	123	132	213	231	312	321
标签	1	2	3	4	5	6

3.2.3 网络训练

本文输入的是 100×100 尺寸的图像对及对应的相对质量顺序标签。在 LIVE 数据库中, 根据 3 幅图像的 DMOS 值可以得到 3 幅图像的 MOS 值, 进而可以得到该图像对的相对质量顺序标签。在测试阶段, 对图像对中所有的 100×100 大小的图像块对进行相对质量顺序预测, 然后统计图像对标签的个数, 并将个数较多的标签赋给该图像对。最后, 预测正确的图像对占所有测试集的比例即为算法的正确率。

4 实验与分析

为验证本文算法的性能, 在 LIVE 数据库上进行实验, 并将其与 6 种无参考图像质量评估算法进行性能对比。参与对比的算法包括 4 种机器学习算法 (BRISQUE^[23], DIVINE^[21], SSEQ^[16], BIQI^[20]) 和两种基于卷积神经网络的算法 (Kang^[26], RankIQ^A^[33])。

LIVE 数据库中共有 29 幅参考图像, 每个参考图像都有相应的失真图像。实验过程中, 随机选择 70% 的参考图像及其对应的失真图像, 将这些图像组成图像对以作为训练集, 10% 的参考图像作为验证集, 其他参考图像和对应的失真图像组成的图像对作为测试集; 其他 4 种方法的训练样本是 80% 的图像和对应的 MOS 值。将测试集中的数据组成图像对, 对图像对中的 2 幅图像或者 3 幅图像分别进行质量分数预测。根据 2 个或者 3 个质量分数得到相对质量标签或相对质量顺序标签, 即可得到该算法的准确率。

实验将根据图像对的构成, 从 4 个维度对算法进行验证, 包括相同参考图像、相同失真类型, 相同参考图像、不同失真类型, 不同参考图像、相同失真类型, 以及不同参考图像、不同失真类型。

4.1 双通道 CPNet 的对比实验

4.1.1 相同参考图像类型实验

本小节中的实验分为两个部分: 1) 图像对的两幅图像具有相同参考图像和相同失真类型; 2) 图像对的两幅图像具有相同参考图像和不同失真类型。

实验 1) 的结果如表 2 所列, 最后一列是每种算法的平均准确率。从表 2 可以看出, 3 种基于卷积神经网络的方法拥有比传统的机器学习方法更好的性能, 体现了卷积神经网络在图像质量评价领域的优势; 而本文算法优于 Kang^[26] 和 RankIQ^A^[33] 算法, 证明通过对两幅图像的特征做差可以学习到两幅图像之间的质量差异。实验 2) 的结果如表 3 所列。可以看出, 本文提出的基于 CPNet 的相对图像质量评估算法获得了最高的准确率。

表 2 各算法对相同参考图像、相同失真类型的两幅图像对的准确率

Table 2 Accuracy of each algorithm for two-image pair with same reference image and distortion type

算法	JP2K	JPEG	WN	BLUR	FF	Aver
SSEQ	0.914	0.885	0.969	0.956	0.915	0.928
DIIVINE	0.922	0.874	0.974	0.980	0.917	0.933
BRISQUE	0.929	0.893	0.973	0.987	0.936	0.944
BIQI	0.915	0.885	0.973	0.965	0.917	0.931
Kang	0.951	0.925	0.990	0.988	0.948	0.960
RankIQA	0.972	0.939	0.996	0.979	0.861	0.949
CPNet	0.967	0.940	0.997	0.995	0.955	0.971

表 3 各算法对相同参考图像、不同失真类型的两幅图像对的准确率

Table 3 Accuracy of each algorithm for two-image pair with same reference image and different distortion type

算法	准确率	算法	准确率
SSEQ	0.874	Kang	0.932
DIIVINE	0.888	RankIQA	0.917
BRISQUE	0.907	CPNet	0.947
BIQI	0.893	—	—

表 2 和表 3 表明:本文算法优于其他算法,对所有参考图像都有较好的结果。图 4 给出了本文算法预测成功的两组图像对。通过观察预测错误的图像对并分析其错误的原因发现,在所有被预测错误的图像对中,两幅图像的 MOS 值的差异都不超过 5,最小的差异甚至还不到 0.01,如图 5 所示,这个范围内的差异也是人眼很难分辨的。因此,本文的双通道 CPNet 算法在图像质量差异的学习能力上接近于 HVS。

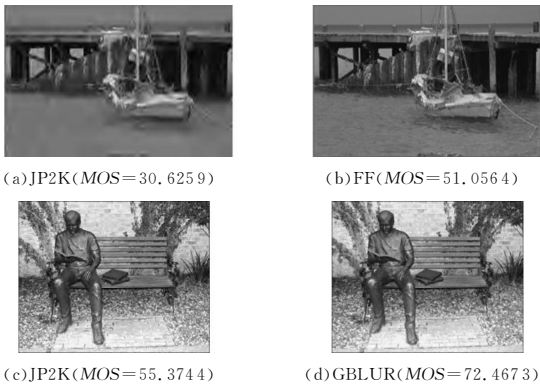


图 4 预测成功的图像对示例图

Fig. 4 Successful prediction examples of image pair

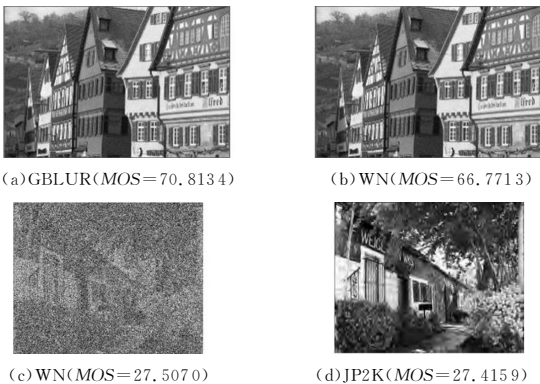


图 5 预测失败的图像对示例图

Fig. 5 Failure prediction examples of image pair

4.1.2 不同参考图像类型实验

本节的实验样本包括两种:1)组成图像对的两幅图像具有相同的失真类型,但是参考图像不同;2)数据库中随机两幅图像组成图像对,并且两幅图像既不是相同的失真类型,也没有相同的参考图像。

实验 1)的结果如表 4 所列,其中第 7 列为各算法的平均准确率。实验 2)的结果如表 5 所列。

表 4 各算法对不同参考图像、相同失真类型的两幅图像对的准确率

Table 4 Accuracy of each algorithm for two-image pair with different reference images and the same distortion types

算法	JP2K	JPEG	WN	BLUR	FF	Aver
SSEQ	0.832	0.865	0.905	0.919	0.840	0.872
DIIVINE	0.836	0.894	0.931	0.917	0.829	0.881
BRISQUE	0.855	0.873	0.922	0.929	0.837	0.883
BIQI	0.842	0.884	0.927	0.903	0.837	0.879
Kang	0.909	0.927	0.958	0.939	0.852	0.917
RankIQA	0.915	0.913	0.950	0.915	0.816	0.902
CPNet	0.931	0.944	0.962	0.936	0.859	0.926

表 5 各算法对不同参考图像、不同失真类型的两幅图像对的准确率

Table 5 Accuracy of each algorithm for two-image pair with different reference image and distortion type

算法	准确率	算法	准确率
SSEQ	0.842	Kang	0.821
DIIVINE	0.863	RankIQA	0.868
BRISQUE	0.861	CPNet	0.860
BIQI	0.846	—	—

由表 1—表 5 可知,双通道 CPNet 对具有相同参考图像或者相同失真类型的图像对的效果比传统的机器学习算法更好,但是在随机两幅图像的质量对比中效果不理想。这符合人眼的视觉特性,即一幅图像如果受到相同类型、不同程度的失真,人眼可以很容易地分辨出来,但是如果受到不同类型的失真,人眼也很难判断失真程度。

4 种基于机器学习的无参考图像质量评估算法在相同参考图像对实验中的性能也有一定程度的下降(见表 2 和表 3)。在相同参考图像、相同失真类型实验中,每幅参考图像被仿真到某个失真类型中的若干幅图像都具有差异较大的失真程度,因此图像对中的两幅图像质量的差异都比较明显,相对质量的预测准确率也会相应提高。相比相同参考图像、不同失真类型的图像对,不同参考图像、不同失真类型的图像对存在更多质量差异不明显的情况,因此这 4 种方法在不同参考图像、不同失真类型的图像对实验中的准确率也有所下降。

图 6 展示了不同参考图像对实验中 CPNet 对于两幅图像的第二个卷积层的卷积核权重。图 6(a)是来自于不同参考图像但是失真类型都为白噪声(White Noise, WN)的图像对的相对质量评估实验中的权重图;图 6(b)中的图像对的失真类型与参考图像都不同。从图 6 可以看出,相同失真类型下的两幅卷积核权重图相似,因此做差能够学习到两者之间的质量特征差异;而不同失真类型下的两幅卷积核权重图较混乱且不相似,因此实验效果并不理想。

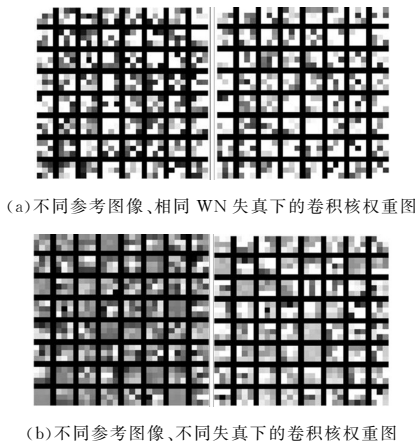


图6 不同维度实验中卷积核权重图的对比

Fig. 6 Comparison of convolution kernel weights in experiments of different dimensions

4.2 三通道 CPNet 对比实验

为了说明随着通道数的增加,提取的特征差异越加复杂,进行了2层卷积层双通道 CPNet (CPNet-1)、3层卷积层双通道 CPNet (CPNet-2)、2层卷积层三通道 CPNet (CPNet-3)、3层卷积层三通道 CPNet (CPNet-4)之间的对比。其中,2层双通道 CPNet 和 3层双通道 CPNet 的实验均是在三通道 CPNet 的训练集和测试集上进行,将一组三通道 CPNet 的数据分3次输入双通道 CPNet 来进行训练和测试。例如:一组三通道 CPNet 数据是(1,2,3),则将(1,2),(1,3)和(2,3)输入双通道 CPNet。

4.2.1 相同参考图像类型实验

本小节中的实验分为两个部分:1)图像对的3幅图像均具有相同参考图像和相同失真类型;2)图像对的3幅图像具有相同的参考图像,但是失真类型不全部相同。实验1)的结果如表6所列,其中最后一列是各算法的平均准确率;实验2)的结果如表7所列。

表6 各算法对相同参考图像、相同失真类型的3幅图像对的准确率

Table 6 Accuracy of each algorithm for three-image pairs with same reference image and distortion type

算法	JP2K	JPEG	WN	BLUR	FF	Aver
SSEQ	0.889	0.844	0.950	0.894	0.733	0.862
DIIVINE	0.841	0.803	1	0.837	0.681	0.832
BRISQUE	0.832	0.794	1	0.847	0.676	0.830
BIQI	0.851	0.749	0.832	0.914	0.686	0.806
Kang	0.936	0.878	1	1	0.589	0.881
RankIQA	0.940	0.835	1	0.917	0.558	0.850
CPNet-1	0.940	0.863	1	1	0.739	0.908
CPNet-2	0.924	0.871	1	1	0.719	0.903
CPNet-3	0.938	0.875	1	1	0.754	0.913
CPNet-4	0.932	0.868	1	1	0.711	0.902

由表6和表7可知,CPNet在整体上优于其他算法。同样,通过观察被预测错误的图像对可以发现,在几乎所有被预测错误的图像对中,3幅图像中至少存在两幅图像的MOS值差异不超过5,这与双通道CPNet中的错误预测分析相同。

表7 各算法对相同参考图像、不同失真类型的3幅图像对的准确率

Table 7 Accuracy of each algorithm for three-image pairs with same reference image and different distortion type

算法	准确率	算法	准确率
SSEQ	0.780	RankIQA	0.743
DIIVINE	0.769	CPNet-1	0.807
BRISQUE	0.776	CPNet-2	0.812
BIQI	0.699	CPNet-3	0.783
Kang	0.624	CPNet-4	0.788

图7给出了实验2)的3个测试示例。



图7 测试示例图

Fig. 7 Diagram of test examples

针对图7中的3个测试示例,各算法的预测结果如表8所列。其中,1,2,3分别代表3幅图像,相对质量顺序从左到右表示图像质量分数依次降低,加粗表示预测成功。除了示例3,本文算法在都预测正确,而其他算法在上述3个示例中都预测错误。示例3中的第一幅和第二幅图像的质量差异很小且属于不同的失真类型,人眼很难区别相对质量差异,因此各算法都未能正确预测。

表8 各算法在图7所示的3个示例上的预测情况

Table 8 Prediction of each algorithm on three examples in Fig. 7

算法	示例1	示例2	示例3
实际相对顺序	132	123	123
SSEQ ^[16]	123	123	231
DIIVINE ^[21]	123	132	213
BRISQUE ^[23]	123	132	213
BIQI ^[20]	321	132	213
Kang ^[26]	321	132	213
RankIQA ^[33]	123	132	213
CPNet-1	123	123	213
CPNet-2	132	123	213
CPNet-3	132	123	213
CPNet-4	132	123	213

4.2.2 不同参考图像类型实验

本节的实验包括如下两类:1)组成图像对的3幅图像具有相同的失真类型,但是参考图像不同;2)从数据库中随机选3幅图像组成图像对,3幅图像具有不同的失真类型和不同的参考图像。

实验1)和实验2)的结果分别如表9和表10所列。

表 9 各算法对不同参考图像、相同失真类型的 3 幅图像对的准确率

Table 9 Accuracy of each algorithm for three-image pairs with different reference images and same distortion types

算法	JP2K	JPEG	WN	BLUR	FF	Aver
SSEQ	0.746	0.758	0.822	0.797	0.760	0.777
DIIVINE	0.688	0.729	0.871	0.750	0.778	0.763
BRISQUE	0.682	0.738	0.865	0.749	0.786	0.764
BIQI	0.708	0.722	0.722	0.846	0.646	0.729
Kang	0.553	0.636	0.821	0.651	0.611	0.656
RankIQA	0.764	0.707	0.825	0.754	0.557	0.721
CPNet-1	0.735	0.690	0.856	0.663	0.739	0.736
CPNet-2	0.750	0.711	0.858	0.706	0.739	0.753
CPNet-3	0.7456	0.715	0.851	0.681	0.735	0.746
CPNet-4	0.763	0.724	0.856	0.699	0.754	0.759

表 10 各算法对不同参考图像、不同失真类型的 3 幅图像对的准确率

Table 10 Accuracy of each algorithm for three-image pairs with different reference image and distortion type

算法	准确率	算法	准确率
SSEQ	0.768	RankIQA	0.657
DIIVINE	0.775	CPNet-1	0.717
BRISQUE	0.771	CPNet-2	0.728
BIQI	0.696	CPNet-3	0.707
Kang	0.493	CPNet-4	0.711

由表 6—表 10 可以看出,三通道 CPNet 网络对于具有相同参考图像的图像对的效果相比传统的机器学习算法和其他两个卷积网络算法更好,但是在不同参考图像对的对比实验中效果不理想,相比双通道也有下降。这符合人眼的视觉特性,即对于具有相同参考图像的一组图像,人眼可以比较容易地区分图像间的质量差异;但对于不同参考图像的一组图像,随着该组图像数越来越多,人眼就越来越难进行准确的质量排序。因此,在具有相同参考图像的实验中,CPNet 有扩展到多路的可能性;而在具有不同参考图像的实验中,随着图像对中图像数量的增加,基于 CPNet 的图像质量排序的准确率会下降。

本文提出的 CPNet 整体上优于其他两种基于卷积网络的图像质量评估算法^[26,33],说明本文提出的 CPNet 能够通过做差的方式学习到图像间的质量差异,这比通过网络直接输出质量分数,然后对比得到相对图像质量的方法更有效。其他两种卷积网络输出分数后,尽管根据分数可以立马得到一组图像的排序结果,但是机器则需要两两对比 3 次才能够得到一组图像的排序结果,这增大了预测误差;而本文提出的 CPNet 则是直接将排序结果作为标签来进行训练,说明本文提出的 CPNet 在一定程度上更符合人眼的视觉特性。

表 6—表 10 表明:相同参考图像对实验中,两层卷积层 CPNet 的预测准确率高于三层卷积层 CPNet 的预测准确率;而在不同参考图像对实验中,两层卷积层 CPNet 的预测准确率低于三层卷积层 CPNet 的预测准确率。该结果说明,在相同参考图像对实验中,做差得到的特征差比较简单,浅层的网络就能够很好地进行区分;而在不同参考图像对实验中,做差得到的特征差就相对复杂,深层的网络能够更好地提取差异特征并进行区分。

4.3 网络参数实验

本节主要通过实验介绍双通道 CPNet 的设计,选择与 Kang^[26]类似的网络结构,在 LIVE 数据库使用相同的训练集和测试集进行实验。通常采用算法结果与主观分数之间的相关性来判断一种无参考图像质量评估算法的效果,相关性越高说明算法的效果越好。常见的相关性表示方法包括皮尔逊线性相关系数(Pearson's Linear Correlation Coefficient, LCC)和斯皮尔曼等级次序相关系数(Spearman's Rank Ordered Correlation Coefficient, SROCC)。

LCC 表征的是线性相关系数的结果,如式(4)所示:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \quad (4)$$

其中, \mathbf{X} 和 \mathbf{Y} 表示两个向量,分子表示两个向量的协方差,分母为两个向量的标准差的积; μ_X 和 μ_Y 表示向量 \mathbf{X} 和 \mathbf{Y} 的均值, σ_X 和 σ_Y 表示向量 \mathbf{X} 和 \mathbf{Y} 的标准差。

SROCC 主要表示的是两个向量对应的数据顺序的相关性,即当一个变量为另外一个变量的单调函数时表示两个变量的 SROCC 是最高的,如式(5)所示:

$$\rho_{X,Y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5)$$

其中, x_i 和 y_i 分别表示向量 \mathbf{X} 和 \mathbf{Y} 排序后的第 i 个元素, $i \in \{1, \dots, N\}$, N 表示向量的长度。

LCC 和 SROCC 对应的结果值域都是 $[-1, 1]$,其中 ± 1 均表示相关性最高,1 表示正相关,而 -1 表示负相关。如果对应的系统绝对值更接近于 0,那么对应的相关性就较低。

4.3.1 卷积核个数

固定网络的结构为一个卷积层,输入样本尺寸为 32×32 ,卷积核大小为 5×5 ,在卷积层后有一层尺寸与卷积层产生的特征图像大小相同的 Max-pooling 层,全连接层的参数设置与 Kang^[26]相同。卷积核个数选择为:16,32,48,64,80,96,112,128。实验结果如图 8 所示。

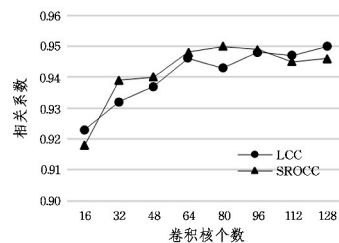


图 8 卷积核个数对 LCC 和 SROCC 的影响

Fig. 8 Influence of number of convolution kernels on LCC and SROCC

由图 8 可知,卷积核个数小于 64 时,随着个数的增加,性能显著提升;卷积核个数超过 64 后,卷积核个数的增加并没有带来很大的性能提升,而且卷积核个数越多,网络中需要训练的参数也就越多,反而导致更多的内存占用和更长的训练时间。因此,本文选择第一层卷积核的个数为 64。

4.3.2 卷积核尺寸

固定网络结构为包含一层卷积层的网络和包含两层卷积

层的网络。前者的卷积核个数为 64。根据对其他网络的研究,卷积核的个数一般随着卷积层的增加而增加,因此后者中第一层卷积层的卷积核个数为 64,第二层卷积核个数为 128。包含一层和包含两层卷积层的卷积核尺寸设置为: 3×3 , 5×5 , 7×7 和 9×9 。实验结果如图 9 所示。

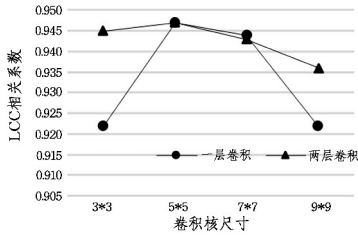


图 9 卷积核尺寸对 LCC 的影响

Fig. 9 Influence of convolution kernel size on LCC

由图 9 可知,在包含一层卷积层的网络中,卷积核尺寸过小或过大都会降低实验效果。卷积核尺寸过小时,无法学习到图像的结构特征;过大时,会丢失图像细节信息。而在包含两层卷积层的网络中, 3×3 尺寸的效果比包含一层卷积层的网络效果好得多。通过分析认为,叠加后的感受野相当于一个 5×5 尺寸的卷积层,因此如果网络的卷积层个数超过一层,则 3×3 尺寸的卷积核更合适。因为本文的卷积层超过一层,因此设置双通道 CPNet 的卷积核尺寸为 3×3 。

4.3.3 卷积层个数

固定网络中卷积核的个数为 64,除了最后一个 Max-pooling 层的尺寸为整幅特征图,每个卷积层的后面都有一个 2×2 尺寸大小的 Max-pooling 层,设置卷积层的层数为 1, 2, 3, 4, 5, 实验结果如图 10 所示。

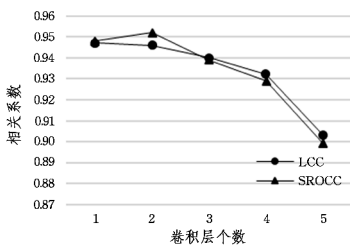


图 10 卷积层个数对 LCC 和 SROCC 的影响

Fig. 10 Influence of number of convolutional layers on LCC and SROCC

可以看出,当卷积层个数超过 3 后就出现了过拟合现象,并且随着层数的增加,参数数量和训练时间也大幅增加。因此,设置双通道 CPNet 的卷积层个数为 2。

4.3.4 输入样本尺寸

本节通过实验说明输入样本尺寸的选择以及选择根据通道数将输入样本分开的原因,然后分别经过卷积层和池化层,在全连接层对其进行特征求差。如果不根据通道数将输入样本分开,则输入样本中的两幅图像的特征在卷积层就融合到了一起,无法确定网络是否学习到图像之间的质量差异。但是因为不同的输入尺寸间的差异较大,所以需要根据输入尺寸来对网络参数进行微调,使得结果更合理。实验结果如图 11 所示。

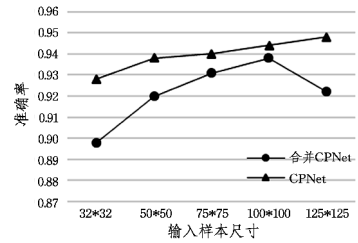


图 11 网络结构差异性和样本尺寸对算法准确率的影响

Fig. 11 Influence of network structure difference and sample size on algorithm accuracy

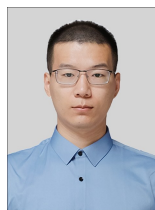
由图 11 可以看出,将两幅图像分开卷积后在全连接层进行做差可以帮助网络更好地学习到图像质量差异的特征;而且算法的性能随着样本尺寸的增大而升高,但是样本尺寸越大,带来的算法时间复杂性就越高,需要训练的参数量更多,同时会导致数据量减少,因此本文选择样本输入的尺寸为 100×100 。

结束语 本文提出了一种基于 CPNet 的相对图像质量评估方法。该方法是一种分数无关类型的算法,在一定程度上避免了主观评分带来的误差,同时利用图像组合的形式解决了数据量的限制。除此之外,本文提出的相对质量标签或者相对质量顺序标签具有更广阔的实际应用场景。相对于绝对质量分数标签来说,所提算法不必花费大量人力、物力以及时间来获取训练标签,同时更加准确。在 LIVE 数据库上的实验结果表明,本文方法更加贴近于人眼视觉特性,而且在相同参考图像实验上,本文提出的 CPNet 取得了比传统无参考算法以及其他卷积网络算法更好的结果,同时也具有扩展到多通道的可能性。而在不同参考图像实验上,随着通道数的增加,CPNet 预测的准确率呈现下降的趋势。未来需要进一步提高网络的预测性能,提高其在不同参考图像实验上的预测准确率。

参考文献

- [1] ALAKUIJALA J, OBRYK R, STOLIARCHUK O, et al. Guetzli: Perceptually guided JPEG encoder [EB/OL]. arXiv: arxiv/abs/1703.04421.
- [2] CHANNAPPAYYA S S, BOVIK A C, CARAMANIS C, et al. Design of linear equalizers optimized for the structural similarity index [J]. IEEE Transactions on Image Processing, 2008, 17(6): 857-872.
- [3] GU K, TAO D C, QIAO J F, et al. Learning a no-reference quality assessment model of enhanced images with big data [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(4): 1301-1313.
- [4] ZHOU W, BOVIK A C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures [J]. IEEE Signal Processing Magazine, 2009, 26(1): 98-117.
- [5] ZHANG L, ZHANG L, MOU X Q, et al. FSIM: A feature similarity index for image quality assessment [J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378.
- [6] VENKATESH B R, PERKIS A. An HVS-based no-reference perceptual quality assessment of JPEG coded images using neu-

- ral networks [C]//Proceedings of the IEEE International Conference on Image Processing. Genova: IEEE Press, 2005, 1: 1-433.
- [7] BRANDÃO T, QUELUZ M P. No-reference image quality assessment based on DCT domain statistics [J]. *Signal Processing*, 2008, 88(4): 822-833.
- [8] PAN F, LIN X, RAHARDJA S, et al. Measuring blocking artifacts using edge direction information [C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Taipei: IEEE Press, 2004: 1491-1494.
- [9] VLACHOS T. Detection of blocking artifacts in compressed video [J]. *Electronics Letters*, 2000, 36(13): 1106-1108.
- [10] CHEN M J, BOVIK A C. No-reference image blur assessment using multiscale gradient [C]//2009 International Workshop on Quality of Multimedia Experience. IEEE, 2011.
- [11] CIANCIO A, COSTA A L, SILVA E A, et al. No-reference blur assessment of digital pictures based on multifeature classifiers [J]. *IEEE Transactions on Image Processing*, 2010, 20(1): 64-75.
- [12] LI C F, BOVIK A C, WU X J. Blind image quality assessment using a general regression neural network [J]. *IEEE Transactions on Neural Networks*, 2011, 22(5): 793-799.
- [13] YE P, DOERMANN D. No-reference image quality assessment using visual codebooks [J]. *IEEE Transactions on Image Processing*, 2012, 21(7): 3129-3138.
- [14] ZHANG Y, WANG C, MOU X Q. SPCA: A no-reference image quality assessment based on the statistic property of the PCA on nature images [J]. *Digital Photography IX*, 2013, 8660(1): 311-316.
- [15] ZHOU L Y, ZHANG Z B. No-reference image quality assessment based on noise, blurring and blocking effect [J]. *Optik-International Journal for Light and Electron Optics*, 2014, 125(19): 5677-5680.
- [16] LIU L X, LIU B, HUANG H, et al. Noreference image quality assessment based on spatial and spectral entropies [J]. *Signal Processing: Image Communication*, 2014, 29(8): 856-863.
- [17] SAAD M A, BOVIK A C, CORMACK L. A DCT statistics-based blind image quality index [J]. *IEEE Signal Processing Letter*, 2010, 17(6): 583-586.
- [18] SAAD M A, BOVIK A C, CHARRIER C. DCT statistics model-based blind image quality assessment [C]//IEEE International Conference on Image Processing. Brussels: IEEE Press, 2011: 3093-3096.
- [19] SAAD M A, BOVIK A C, CHARRIER C. Blind image quality assessment: a natural scene statistics approach in the DCT domain [J]. *IEEE Transactions on Image Processing*, 2012, 21(8): 3339-3352.
- [20] MOORTH A K, BOVIK A C. A two-step framework for constructing blind image quality Indices [J]. *IEEE Signal Processing Letter*, 2012, 17(5): 513-516.
- [21] MOORTH A K, BOVIK A C. Blind image quality assessment: From natural scene statistics to perceptual quality [J]. *IEEE Transactions on Image Processing*, 2011, 20(12): 3350-3364.
- [22] TANG H X, JOSHI N, KAPOOR A. Learning a blind measure of perceptual image quality [C]//Proceedings of the IEEE International Conference on Image Processing. Brussels: IEEE Press, 2011: 305-312.
- [23] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain [J]. *IEEE Transactions on Image Processing*, 2012, 21(12): 4695-4708.
- [24] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a 'completely blind' image quality analyzer [J]. *IEEE Signal Processing Letters*, 2013, 20(3): 209-212.
- [25] MITTAL A, SOUNDARARAJAN R, MURALIDHAR G S, et al. Blind image quality assessment without training on human opinion scores [C]//Proceedings of the SPIE Human Vision and Electronic Imaging XVIII. Florida: IEEE Press, 2013.
- [26] KANG L, YE P, LI Y, et al. Convolutional neural networks for no-reference image quality assessment [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1733-1740.
- [27] KIM J, LEE S. Fully Deep Blind image quality predictor [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(1): 206-220.
- [28] LI J, ZOU L, YAN J, et al. No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks [J]. *Signal Image and Video Processing*, 2016, 10(4): 609-616.
- [29] LI Y M, PO L M, FENG L T, et al. No-reference image quality assessment with deep convolutional neural networks [C]//Proceedings of the IEEE International Conference on Digital Signal Processing. London: IEEE Press, 2017: 685-689.
- [30] BOSSE S, MANIRY D, MÜLLER K R, et al. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment [J]. *IEEE Transactions on Image Processing*, 2016, PP(99): 1-1.
- [31] GU J, MENG G F, XIANG S M, et al. Blind image quality assessment via learnable attention-based pooling [J]. *Pattern Recognition*, 2019, 91: 332-334.
- [32] GAO F, TAO D C, GAO X B, et al. Learning to rank for blind image quality assessment [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(10): 2275-2290.
- [33] LIU X L, WEIJER J V D, BAGDANOV A D. RankIQA: Learning from Rankings for No-Reference Image Quality Assessment [C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE Press, 2017: 1040-1049.



LI Kai-wen, born in 1994, postgraduate. His main research interests include image quality evaluation and so on.



CHEN Qiang, born in 1979, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include image processing and analysis, pattern recognition and machine learning.