课程论文题目

肖康

摘要

与相机拍摄的自然图像不同，屏幕内容图像（SCI）是一种包括文本和图形区域的合成图像。不同的特性导致了图像质量评估（IQA）中的许多困难。大多数基于卷积神经网络（CNN）的现有模型将大的图片分割成图像块，以增加 CNN 训练的训练样本。这带来了两个问题：（1）单个图像块不能代表整个图像的质量，特别是在 SCI 的 IQA 中；（2）由相同失真类型和强度退化的整个图像的 SCI 块可能具有显著不同的质量。此外，这些模型采用预测质量和主观差异平均意见得分（DMOS）之间的均方误差（MSE）来训练 CNN，而不考虑不同 SCI 之间的质量排名。所复现的文章提出了一种基于神经网络（CNN）的无参考（NR）IQA 模型。我们算法的贡献可以总结如下：（1）考虑到 SCI 中不同区域存在较大差异，利用多区域局部特征生成的伪全局特征进行质量评估，这比每个图像块的局部特征更好地反映图像质量；（2）噪声分类任务被用作辅助任务，辅助质量分数预测任务以提高表示能力；（3）孪生网络用于预测两个不同 SCI 的质量分数，并提出了一种新的排名损失来对预测的分数进行排名，旨在增强模型在质量方面对图像进行排名的能力。实验结果验证了我所复现出来的模型在屏幕内容图像质量评估数据库（SIQAD）上具有一定竞争力。

关键词：图像质量评价；屏幕内容图片

1　引言

随着多媒体，社交网络的快速发展，屏幕内容图片（screen content images）在我们的日常生活中变得愈加普遍。如在线新闻、电子杂志、电子商务、云游戏和云计算，尤其是在后疫情时代，很多会议都改为线上举行。由于在图像通信系统中可以引入各种失真，因此图像的视觉质量下降。因此，图像质量评价 (IQA) 在图像处理领域起着非常重要的作用。客观 IQA 模型具有动态监控、性能评估和参数优化的功能，在图像和视频系统中充当着” 监督员”、” 裁判” 和” 教练” 的角色。例如，在图像分类、物体跟踪、图像压缩、图像修补和图像超分辨等应用中，往往需要 IQA 模型进行性能评估和参数优化，从而使被人们所感知到的图片质量更优。

2　相关工作

之前的工作主要针对自然图片作质量评价，深度学习在 IQA 中的成功并不容易从自然图像转移到屏幕内容图片中，这是因为屏幕内容图片是由图片和文字组成的，而自然图像主要是由图片组成的。这导致了屏幕内容图片（SCI）中图片区域和文本区域在自然度统计上的巨大差异。也就是说不能直接照搬之前的针对自然图片的方法。所以之后有一些工作是专门为了屏幕内容图片而设计的客观质量评价方法。比如 Yue 等人[1]利用全参考模型来生成屏幕内容图片的质量分数，以此来作为标签训练一个卷积神经网络（CNN），然后再在基础数据集上对该网络权重进行微调。Chen 等人[2]提出了一个自然化模块，旨在通过这个模块把屏幕内容图片转换成更具有自然图片特征的图片。但是，它们的方法或多或少存在着一些缺陷。Yue 等人的方法采用了整张图片作为网络的输入，这样缺少了训练数据，导致训练效果不好。Chen 等人的方法把作为衡量整张图片的质量的 DMOS 用于该图片分割后所得的

多个部分当中，这是不合理的，会引入偏差。为了解决这些问题，作者首先把一张图片分割成四个区域，再在每个区域随机裁剪一块图像块，这样我们会在一张图片中得到 4 个图像块，分别提取它们的特征，最后再把特征拼接起来得到了伪全局特征。用这个特征来预测质量分数，从而避免了单个图像块就代表了整个图像的质量的问题。另一方面，不同区域的图像块的结合增加了训练样本的数量。

3 本文方法

3.1 本文方法概述

此部分对本文将要复现的工作进行概述，复现论文所提出的无参考 IQA 网络架构如图 1所示。它由四个模块组成：局部特征提取模块、伪全局特征生成模块、多任务训练模块和孪生网络模块。局部特征提取模块和伪全局特征生成模块用于提取用于质量评估的伪全局特征。多任务训练模块有两个任务：噪声分类任务和质量分数预测任务。噪声分类任务被用作辅助任务，辅助质量分数预测任务。孪生网络模块用于根据质量对 SCI 进行排名。

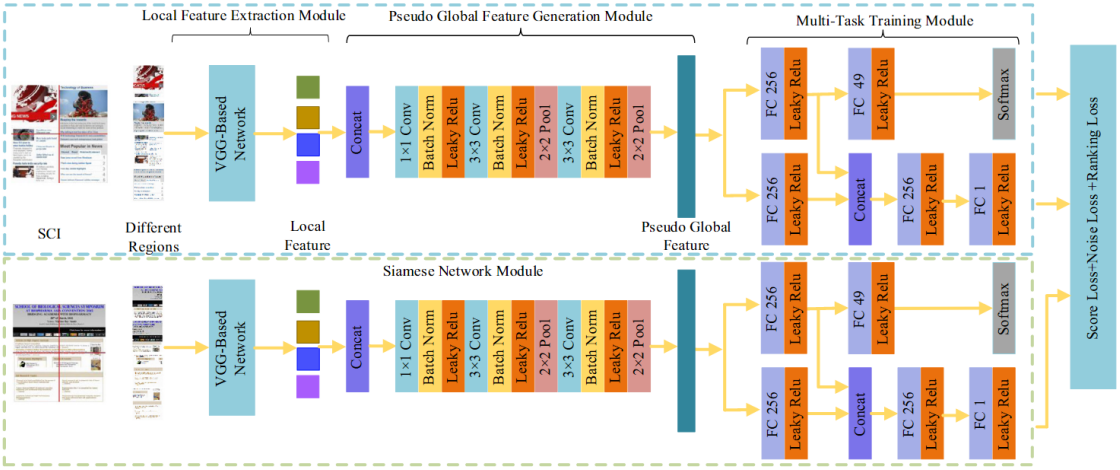

图 1: NR-IQA 网络架构示意图

3.2 各模块详解

局部特征提取模块：局部特征提取模块基于 VGGnet[3]，包括 Conv5-32、Conv5-32、Maxpool2、Conv3-64、Conv3-64、Maxpoor2、Conv3-128、Conv3-128、Maxpool 2、Conv3-256、Conv3-206、Maxpool2，如图 2所示。Conv3-64 表示大小为 3×3、卷积核数量为 64 的卷积层，Maxpool2 表示大小为 2×2 的池化层。将输入 SCIs 划分为多个区域，并从每个区域提取一个大小为 128×128 的图像块作为每个区域的代表性样本。局部特征提取模块用于分别提取每个区域的局部特征。在基于 vgg 的网络中，2 个 3 × 3 卷积层和 1 个池化层的组合具有更大的视图，以更少的数据提取局部特征，具有出色的特征提取能力。前两个卷积层采用 5 × 5 的大规模卷积核，从整个输入图像块中获取局部特征的一般信息。采用共享权值提取多区域特征的结构，可以确定该模块从每个区域提取相同的属性特征，并利用这些特征来预测图像质量。

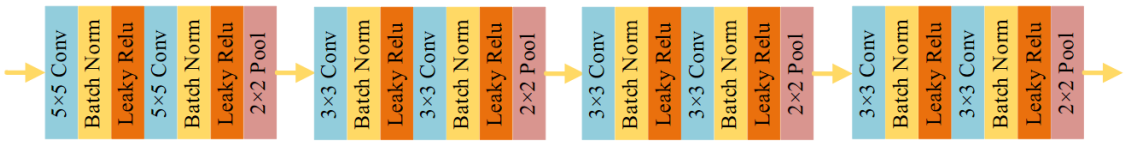

图 2: VGG-based 网络架构示意图

伪全局特征生成模块：伪全局特征生成模块融合局部特征生成伪全局特征。伪全局特征生成模块由串联层（concatenate layer）、Conv1-128、Conv3-256、Maxpool2、Conv3-256 和 Maxpool2 组成。使用串联层对多个区域的特征进行融合，使用 1 × 1 卷积层对融合特征的尺寸进行调整。采用 2 个 3 × 3 卷积层和 1 个池化层生成伪全局特征。

多任务训练模块：在获得伪全局特征后，将这些特征用于训练多任务学习模型。多任务训练模块包括两个分支：噪声分类任务和质量分数预测任务。噪声分类任务的分支由两个全连接层和一个 Softmax 层组成。由于图像的主观质量取决于噪声类型、噪声强度和图像内容，因此该分支采用分类网络来提取噪声类型和强度的特征。质量分数预测任务的另一个分支由三个 FC 层和一个串联层（concatenate layer）组成。具体配置如图 1所示。注意，将噪声分类任务的特征向量与质量分数预测任务的特征向量连接起来，生成新的质量分数预测任务的特征向量。通过学习的新特征向量来预测质量分数。

孪生网络模块：为了增强模型对 SCI 质量指标进行排序的能力，引入孪生网络来提取不同 SCI 的特征，并共享模型的权重。将两个不同的 SCI 同时输入到一个模型中，然后得到两个预测分数作为模型的输出，提出了一种新的排序损失，通过比较两个输入 SCI 的预测分数来对 SCI 进行排序。关键是要学习模型，准确预测 SCI 质量和不同 SCI 之间的质量差异。

3.3 损失函数定义

对于噪声分类任务的 loss 函数如下，为经典的交叉熵损失函数：

$$\text{noise loss} = -\sum_{i=1}^{N}\sum_{j=1}^{C} p_j^{(i)} \log \hat{p}_j^{(i)} \left(X^{(i)}\right)$$

其中 N 为一个批次中样本数量，C 为噪声的种类数，$p_j^{(i)}$ 是一个标签向量，$\hat{p}_j^{(i)}$ 是模型输出的 softmax 向量。

质量分数预测 loss 函数，其目的是使得模型预测的图片的质量分数更加接近 MOS 值（ground truth）：

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } x < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

$$\text{score loss} = \frac{1}{N}\sum_{i=1}^{N} \text{smooth}_{L_1}(\hat{q} - q)$$

其中 $\hat{q}$ 是网络的输出，q 是对应的 ground truth。

为了使该模型具有对图像进行排序的能力，利用孪生网络对具有共享网络权重的不同 SCI 进行特征提取，并提出了一种新的排序损失用于 IQA 对 SCI 进行排序。排名损失由以下定义：

$$\text{ranking loss} = \frac{1}{N}\sum_{i=1}^{N} \text{smooth}_{L_1}\left((\hat{q}_1 - \hat{q}_2) - (q_1 - q_2)\right)$$

最后将上述的 loss 加权再加和即可得总损失函数：

$$\text{total loss} = \text{noise loss} + \text{score loss} + \beta \cdot \text{ranking loss}$$

优化目标为：

$$\min\left\{\text{total loss} + \frac{\alpha}{2N}\sum w^2\right\}$$

其中 α 为惩罚因子，式中后面一项是 L2 正则化。

4 复现细节

本次复现是基于 pytorch 来实现的，遵循了经典的训练框架。具体来说，分别实现了数据导入模块，论文模型搭建，损失函数模块，训练模块，测试模块。这些模块的代码见文件夹中，原理已经在前面部分阐述清楚，所以在此就不再另外粘贴代码。还有一些关于训练的细节：输入的 SCI 被平均分配到多个区域。将每个区域提取的 128×128 个图像块组合起来作为模型的输入。采用 batch=32 的 Adam 优化算法进行训练。L2 正则化的惩罚因子 − 为 1×e-5。学习速率从 1×e−4 变化到 1×e−8 每间隔 10 个 epoch。

4.1 与已有开源代码对比

本报告没有参考任何相关源代码，在此明确申明。

5 实验结果分析

我所复现出来的结果如图 3所示：

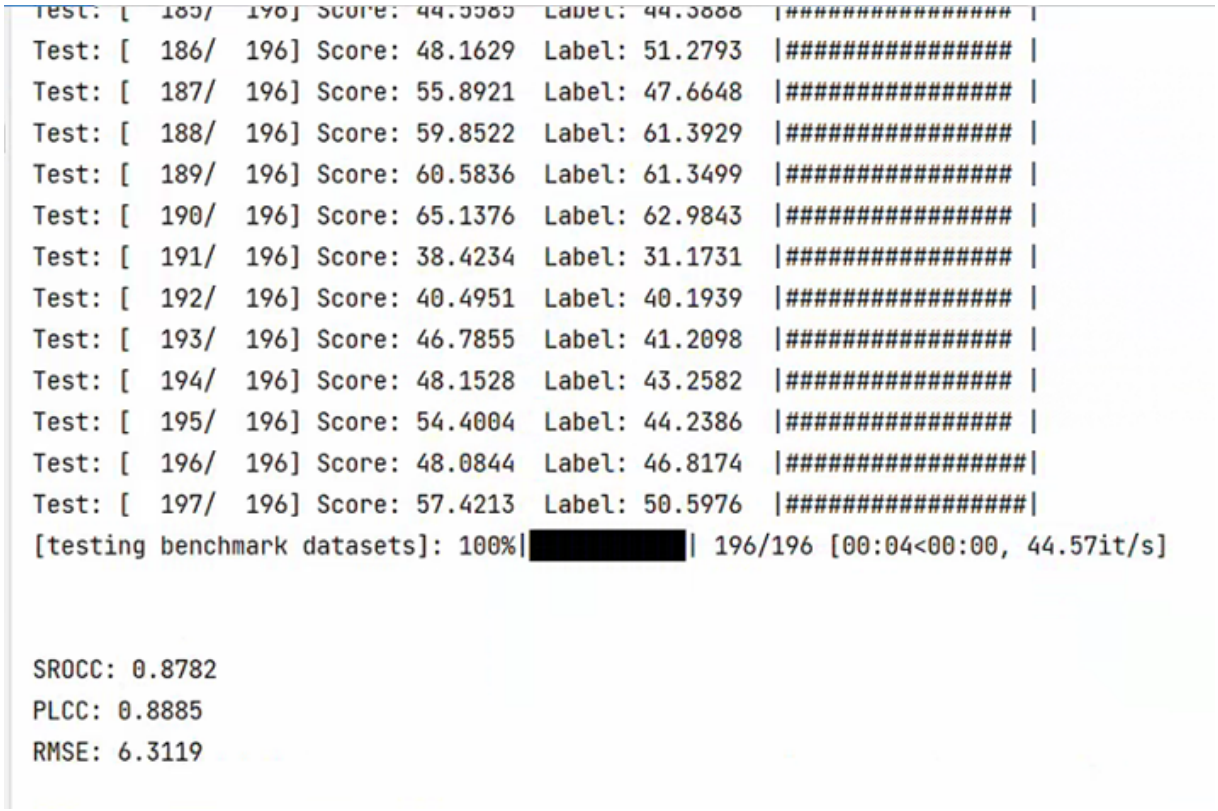

图 3: 复现实验结果示意

原论文中的结果如图 4所示：

Table 2
Experimental results of the proposed method and other existing NR methods on SIQAD database.

Method	BLINDS-II [6]	BRISQUE [8]	NRLT [36]	HRFF [27]	CNN-Kang [16]	RankIQA+FT [18]	WaDIQaM-NR [17]	BQACNN-Yue [22]	PICNN [23]	RIQA
PLCC	0.7255	0.7708	0.8442	0.852	0.8487	0.8776	0.8594	0.8834	0.896	**0.9107**
SRCC	0.6813	0.7237	0.8202	0.832	0.8091	0.8513	0.8522	0.8634	0.897	**0.9002**
RMSE	9.4991	8.1342	7.5957	7.415	7.4472	7.0046	7.0570	-	6.790	**5.8803**

图 4: 原论文实验结果示意

通过对比我们两张图的结果，可以发现结果并没有达到原论文中的那样好的结果，但是相差也不是很远，说明本次复现工作还是比较成功的。至于为什么表现得不错是因为本模型融合了多个 loss，首先，ranking loss 会起到提升 SROCC，KROCC 的作用，因为他们都是基于排序的，而 ranking loss 会使得图片之间的质量分数的顺序保持一定的一致。其次，quality score loss 会使得模型预测出来的质量分数尽量的接近 MOS，也就是真实值。最后，添加了附加的噪声分类任务，这样能辅助图像质量预测这个阶段的进行，因为图像的质量分数是和图像的噪声种类有关的。综上，这些都是导致最后效果还不错的原因。

6 总结与展望

本部分对整个文档的内容进行归纳并分析目前实现过程中的不足以及未来可进一步进行研究的方向。本文提出了一种基于多区域特征学习网络的 SCI NR-IQA 模型。为了提高特征表示能力和模型预测能力，引入了噪声分类任务的辅助任务和排序孪生网络。在 SIQAD 上对训练后的模型进行了性能评估，验证了多区域特征提取模块、多任务模块和排序模块设计的有效性。实验结果表明，所提出的 RIQA 模型优于目前最先进的 SCI 无参考 IQA 模型。此外，本复现工作还存在一些不足：性能并没有原论文中的那么好，我分析是采用的初始化方式不同，抑或是对数据集的划分只进行了一次，不足以说明自己的复现方案不行。对于未来可以进行的进一步的研究方向，我认识到大多数 IQA 模型不能同时评估自然图像和 SCIs 的质量的巨大挑战，我计划在未来设计一个统一的基于 cnn 的 NR IQA 模型，适应差异化内容类型。

参考文献

[1] YUE G, HOU C, YAN W, et al. Blind quality assessment for screen content images via convolutional neural network[J]. Digital Signal Processing, 2019, 91: 21-30.

[2] CHEN J, SHEN L, ZHENG L, et al. Naturalization Module in Neural Networks for Screen Content Image Quality Assessment[J]. IEEE Signal Processing Letters, 2018, 25(11): 1685-1689. DOI: 10.1109/LSP.2018.2871250.

[3] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [C]//International Conference on Learning Representations. 2015.