

参考图像生成式的图像质量评价算法

罗 鹏 刘 佳

武警工程大学 陕西西安 710086

摘 要: 针对目前图像质量评价算法都是依据特定特征而设置的, 为了满足对未知失真类型进行准确评价, 本文借助 GANs 在图像生成和图像补全领域的优异性能, 提出了一种新的用于图像质量评价的深度学习网络: IQA-GAN。实验结果表明, 设计的模型评价准确性与已有的无参考图像质量评价方法处于同一量级, 但其不需要原始图像作为参考, 具有明显优势。

关键词: 图像质量评价; 深度学习; 卷积神经网络;

Reference image generation algorithm for image quality evaluation

Peng Luo, Jia Liu

Engineering University of PAP Xi' an, Shaanxi 710086

Abstract: At present, image quality evaluation algorithms are set according to specific features. To accurately evaluate unknown distortion types, this paper proposes a new deep learning network for image quality evaluation, IQA-GAN, based on the excellent performance of GANs in the field of image generation and image completion. The experimental results show that the evaluation accuracy of the designed model is in the same order as that of the existing non-reference image quality evaluation methods, but it does not require the original image as a reference, which has obvious advantages.

Keywords: IQA ; Deep Learning; CNN

现代社会由于数字图像获取的高度便利性, 其数量呈现出爆发式增长, 如何在海量数字图像中筛选出高质量的资源就显得十分紧迫且必要, 图像质量评价算法作为衡量图像质量的重要手段, 越来越受到研究人员的关注。随着图像质量评价算法的不断发展, 有大量的方案被提出。这些方案按照评价主体的不同, 可以分为主观图像质量评价 (Subjective Assessment) 和客观质量评价 (Objective Assessment)^[1]。主观图像质量评价方案起初是依据主观欣赏体验而设计的, 其技术路线是: 测试者根据自己的实际视觉感受对评价图像的量化评判, 得到一个评价分数, 实验中由于主观评价收到干扰因素较多, 因此为了保证主观评价具有较高准确度和可信性, 通常需要几个甚至几十个测试者对同一图像进行打分, 最终产生平均分数, 这就是主观平均分数 MOS (Mean Opinion Scores) 和差分主观平均分数的 DMOS (Differential Mean Opinion Scores) 的由来。显而易见, 主观质量评价算法最能真实地反映出人的感受, 也是图像质量评价最直接的解决办法。但是主观质量评价方法的也有不可忽略的缺陷, 一是耗时耗力, 二是不能即时处理。因此研究者的算法创新主要集中在如何运用客观手段对图像质量进行评价, 客观评价算法性能的评价标准主要依据其与主观评价分数的拟合程度。

目前, 客观图像质量评价大致有三类: 无参考图像质量评价方法, 半参考图像质量评价方法, 全参考图像质量评价方法。但在具体实践中, 往往缺少无失真的参考图像, 因此无参考图像质量评价算法 (NR-IQA) 是客观图像质量评价领域的研究重点。

由于不同失真水平和不同失真类型的图像和其质量之间的联系不是线性的, 很难去找到一个统一的合适的公式去表达, 因此研究人员多采用机器学习的办法去研究 NR-IQA。通常的做法是通过在频域提取图像特征, 如基于自然场景统计的图像关键特征, 输入到回归模型或者贝叶斯模型, 对图像质量进行量化估计。这类算法涌现出很多优秀的方案, 由于传统的机器学习模型性能的限制, 导致这类算法也出现了效果瓶颈。

加拿大多伦多大学教授 Geoffrey Hinton 于 2006 年提出深度学习, 为图像质量评价提供了新的思路。深度学习对于模型训练方法的改进突破了反向传播算法神经网络 (Backpropagation algorithm, BP) 发展的瓶颈, 其通过建立、仿真人脑进行分析学习的神经网络来解释数据。因其强大的建模和表达能力, 一经提出便引起广泛关注。深度学习网络中一个重要的模型是卷积神经网络 (Convolutional Neural Network, CNN), 其在各种数字图像处理的任务中都有优异的表现。CNN 的复杂神经元结

构可以将输入和输出进行有效映射。因此,卷积神经网络的思想传统的图像质量评价方法与卷积神经网络的思想一经结合,就出现了很多新型的图像质量评价算法。

本文在 CNN 的基础上提出了一种新的基于深度学习的图像质量评价算法 IQA-GAN, 首先利用生成对抗网络的基本思想, 得到一张接近原始图像的生成图像, 以其替代原始参考图像, 最终将无参考的图像质量评价算法转化成全参考的图像质量评价算法。

一、生成对抗网络

生成式对抗网络 (Generative adversarial networks, GANs)^[2] 是一种新型生成式模型, Goodfellow 等于 2014 年提出。GANs 在结构上受博弈论中的二人零和博弈思想的启发, 整个网络结构由两部分组成, 由一个生成器网络 (Generator) 和一个判别器网络 (Discriminator)。生成器网络根据真实数据样本的潜在分布进行采样, 并生成新的数据样本; 通过不断训练, 使得判别器能够在一定置信度下区分真实样本还是生成的数据。其中, 生成器和判别器均可以采用目前研究火热的卷积神经网络的某种变形。实际上, GANs 的优化过程是一个极大极小博弈 (Minimax Game) 问题, 优化目标是达到纳什均衡^[8], 使生成器生成的样本分布尽可能接近真实数据样本的分布。在人工智能发展的如火如荼的当下, GANs 的提出为许多领域的研究提供了新的思路, 同时也为这些领域的不断发展注入了新的动力。从 GANs 的提出到现在, 短短五年的时间, 业已成为人工智能领域一个非常热门的研究方向, 甚至连图灵奖得主 LeCun 将其称为“过去十年间机器学习领域最让人激动的点子”。当前, GANs 主要应用于数字图像和机器视觉领域, 可以完成图像超分辨率重建、图像风格迁移、图像翻译、图像去噪和图像补全等。本文主要利用 GAN 的图像补全功能, 来输出接近原始图像的生成样本。

二、IQA-GAN 网络结构

本文算法提出的 IQA-GAN 网络结构也是有两个网络分支, 一个生成器 (Generator), 另一个是判别器 (Discriminator)。生成器的输入端是一个局部归一化后的图像块, 之后是为卷积层, 他有 8 个大小为 3×3 的卷积核, 卷积时的滑动步长为 1, 后面是残差模块, 主要是解决多层网络梯度消失, 最后两层卷积层将得到一个三通道的生成样本; 对于鉴别网络而言, 首先是一层卷积层和生成网络一样的参数设置, 接下来也是残差模块, 最后是全连接层输出判别结果。为了加速训练, 使得模型收敛更快, 达到更好的效果, 我们采用了 VGG 模型作为预训练模型, 在其基础上进行进一步训练。

在本文 IQA-GAN 模型中, 为了提高训练效果首先对图像局部亮度指数进行归一化预处理, 第一步把图像从 RGB 变换到 Lab 空间, 原因是 Lab 颜色更符合人类视觉系统成像原理, 其中 L 分量与人类亮度感知高度关联。

为了提高算法效果, 本文在模型训练之前对图像进行归一化处理, 目的是将特征值大小调整到相近的范围, 如果不归一化处理, 当特征值较大时, 梯度值也会较大, 特征值较小时, 梯度值也会较小。在模型反向传播时, 梯度值更新与学习率一样, 当学习率较小时, 梯度值较小会导致更新缓慢, 当学习率较大时, 梯度值较大导致模型不易收敛, 因此为了使模型训练收敛平稳, 对图像进行归一化操作, 把不同维度的特征值调整到相近的范围内, 就可以采用统一的学习率加速模型训练。归一化之后, 再把图像分成 32×32 图像块, 输入 IQA-GAN。在网络的每个卷积层后都有池化层, 用于降低卷积特征图的维度。

本方案采用 Wasserstein 距离作为损失函数, 由于它相对 KL 散度 (相对熵) 与 JS (交叉熵) 散度具有优越的平滑特性, 理论上可以解决梯度消失问题。接着通过数学变换将 Wasserstein 距离写成可求解的形式, 利用一个参数数值范围受限的判别器神经网络来最大化这个形式, 就可以近似 Wasserstein 距离。在此近似最优判别器下优化生成器使得 Wasserstein 距离缩小, 就能有效拉近生成分布与真实分布。WGAN 既解决了训练不稳定的问题, 也提供了一个可靠的训练进程指标, 而且该指标确实与生成样本的质量高度相关。

三、实验结果及分析

在上述生成网络和判别网络的训练过程中, 图像源于图像质量评价领域广泛使用的 LIVE 数据库, 它由 LIVE 是美国德克萨斯大学奥斯汀分校的电气与计算机工程系与心理学系联合建立, 包含 29 个无失真的原始图像产生 982 幅失真图像。数据库的建立者为每张照片挑选了 25 个人进行测试, 最终得到了图像的 DMOS (Differential Mean Opinion Scores) 值, 在 0 ~ 100 之间, 并且包含在数据库内。由于数据库的样本较小, 为了避免过拟合、提升模型鲁棒性、增加训练数据, 提高模型泛化能力、避免样本不均衡, 我们采用了基于几何变换的数据增强方法^[9] (翻转, 旋转, 裁剪, 缩放, 平移, 抖动), 本质是人为地引入人视觉上的先验知识, 提升模型的性能。

为了准确客观估计本文提出的 IQA-CNN 客观图像质量评价算法性能, 采用线性相关系数做为评价指标进行衡量并和其他算法进行对比, 其计算方法如式 (1) 所示:

$$C_{LC} = (N \sum x_i y_i - \sum x_i \sum y_i) / \{ [N \sum x_i^2 - (\sum x_i)^2] [N \sum y_i^2 - (\sum y_i)^2] \}^{1/2}$$

(1) 式中: N 为样本数; \bar{x} 为主观评价均值; \bar{y} 为客观评价均值。C_{LC} 值越大, 算法性能越好。

表 1 给出了本文算法与五种较新全参考算法的比较。从表中可以看出本文方法能较好地评估五种失真, 综合评价结果表明基于 IQA-GAN 的方法接近部分经典的全参考图像质量评价方法的评价水平。

表 1 不同方法在 LIVE 数据库上的 C_{LC} 值比较

方案 / 攻击 类型	Jpeg2000 压缩	JPEG 压缩	高斯白 噪声	高斯模 糊	快衰落
文献 [3]	0.922	0.921	0.988	0.923	0.888
文献 [4]	0.934	0.920	0.936	0.927	0.896
文献 [5]	0.922	0.935	0.966	0.945	0.903
文献 [6]	0.947	0.933	0.963	0.950	0.942
文献 [7]	0.949	0.964	0.977	0.954	0.917
OURS	0.812	0.821	0.824	0.813	0.802

四、结论

本文把生成对抗网络的思想引入到图像质量评价方案的设计过程当中，提出了一新的生成对抗网络 IQA-GAN，并应用于无参考图像质量预测，达到了接近全参考的效果。为了能够帮助网络模型有效学习到图像质量特征，本文在预处理阶段，用亮度系数归一化手段对图像做了局部处理。实验结果显示本文所提出的算法能对五种常用的图像失真进行较为准确地评价，整体性能与其他经典评价方法处于一个量级水平，但是本文所提的方案无需原始图像，实现了的无参考质量评价。

参考文献:

[1] 高敏娟, 党宏社, 魏立力, 刘国军, 张选德. 全参考图像质量评价回顾与展望 [J]. 电子学报, 2021, 49(11):

2261-2272.

[2] 王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望 [J]. 自动化学报, 2017, 43(03): 321-332.

[3] Temel D, AlRegib G. Perceptual image quality assessment through spectral analysis of error representations[J]. Signal Processing: Image Communication, 2019, 70: 37 - 46.

[4] Kim W, Nguyen A D, Lee S, et al. Dynamic receptive field generation for full-reference image quality assessment[J]. IEEE Transactions on Image Processing, 2020, 29: 4219 - 4231.

[5] Ling W Y, Hu Y. Machine learning to design full-reference image quality assessment algorithm[J]. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2013, 11(6): 3439 - 3444.

[6] Narwaria M, Lin W S. SVD-based quality metric for image and video using machine learning[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(2): 347 - 364.

[7] Liu T J, Lin W S, Kuo C C J. Image quality assessment using multi-method fusion[J]. IEEE Transactions on Image Processing, 2013, 22(5): 1793 - 1807.

[8] 汪美琴, 袁伟伟, 张继业. 生成对抗网络 GAN 的研究综述 [J]. 计算机工程与设计, 2021, 42(12): 3389-3395. DOI:10.16208/j.issn1000-7024.2021.12.012.

[9] 阙云, 季雪, 蒋子平, 戴伊, 王叶飞, 陈嘉. GAN 数据增强下路面裂缝语义分割算法 [J/OL]. 吉林大学学报 (工学版): 1-13[2022-09-03]. DOI:10.13229/j.cnki.jdxgbxb 20220003.