

# 基于卷积神经网络的立体图像质量评价

张爽爽,周武杰

(浙江科技学院 信息与电子工程学院,杭州 310023)

**摘要:** 随着立体图像的广泛应用,迫切需要一个具有通用性的工具来评估立体图像的视觉质量,因此提出一种基于卷积神经网络的无参考立体图像质量评价算法。首先使用平面图像数据集对算法框架的主体结构质量图生成网络进行训练;然后使用训练好的网络预测立体图像融合视点图像的质量;最后使用一种加权融合方法得到最终的立体图像质量分数。试验结果表明,算法框架具有相对较好的准确性和鲁棒性。

**关键词:** 图像质量评价;卷积神经网络;加权融合;融合图像

中图分类号: TP389.1      文献标志码: A      文章编号: 1671-8798(2020)01-0026-06

## Stereoscopic image quality evaluation based on convolutional neural network

ZHANG Shuangshuang, ZHOU Wujie

(School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, Zhejiang, China)

**Abstract:** With the prevailing application of stereoscopic images, a universal tool is urgently needed to evaluate the visual quality of those images. Therefore, a non-reference stereoscopic image quality evaluation algorithm was proposed on the basis of convolutional neural network. Firstly, the planar image dataset was used to train the generation network of main structure quality map of the algorithm framework. Then, the well-trained network was used to predict the quality of the fusion viewpoint images. Finally, a weighted fusion method was used to obtain the final quality score of stereoscopic images. The experimental results show that the algorithm framework boasts fairly sound accuracy and robustness.

**Keywords:** image quality evaluation; convolutional neural network; weighted fusion; fused image

---

收稿日期: 2019-05-21

基金项目: 国家自然科学基金项目(61502429);浙江省自然科学基金项目(LY18F020012)

通信作者: 周武杰(1983—),男,浙江省临海人,副教授,博士,主要从事人工智能和视觉大数据研究。E-mail: wujiezhou@163.com。

立体图像的相关应用越来越多地出现在人们的日常生活中。在立体图像的收集、传输、处理和显示过程中,可能会发生多种失真。因此,建立一种高性能的立体图像质量评价方法具有重要的现实意义。立体图像质量评价分为客观评价方法和主观评价方法两种方式。由于主观评价方法耗时、费力,在实时场景中的应用成为一大难题,因此,客观评价方法的应用更加广泛,它包括全参考评价、部分参考评价和无参考评价3种主要方法。全参考评价方法利用原始参考图像与失真图像进行对比,从而获得它们之间的差异性;部分参考评价方法使用了部分未失真原始参考图像信息;无参考评价方法则完全没有使用未失真原始参考图像。由于未失真的原始参考图像在实际应用中很难得到,所以无参考评价方法具有更高的实际应用价值。

在对客观立体图像质量评价的研究中,Chen等<sup>[1]</sup>提出一种传统的立体图像质量评价方法,即使用二维图像拓展到立体图像的评价方法。随着深度学习方法的发展,Lv等<sup>[2]</sup>提出了一种传统算法与深度学习算法结合的方法——基于双目自相似性(binocular self-similarity,BS)和深层神经网络(deep neural network,DNN)的无参考立体图像质量评价方法;Zhang等<sup>[3]</sup>提出了一种基于深度学习端到端的网络框架进行训练测试。此外,还有一些比较经典的传统质量评价方法,如Sun等<sup>[4]</sup>基于双目视觉机制的质量评价方法;Jiang等<sup>[5]</sup>不仅利用双目相互作用作为图像质量评价特征,还利用非负矩阵因子分解的颜色视觉特征;Hachicha等<sup>[6]</sup>提出基于小波分解和统计模型联合的无参考立体图像质量评价方法;Yang等<sup>[7]</sup>提出基于梯度字典颜色视觉特征学习的立体图像质量评价方法。2019年立体图像质量评价领域依旧呈快速发展的状态<sup>[8-12]</sup>,Wu等<sup>[8]</sup>提出了一种基于层级结构的无参考图像质量评价方法,一种从局部结构到深度语义的方法。为了准确评价多种失真图像质量,Jiang等<sup>[9]</sup>和Liu等<sup>[10]</sup>提出针对一种或多种图像失真的质量评价方法,前者通过学习单目和双目局部视觉基元(monocular and binocular local visual primitives,MB-LVPs)来表征视觉皮层的局部感受野特征,从而进行图像质量评价,后者提出了一种基于深度学习网络框架进行分类和预测的立体图像质量评价方法。

上述研究虽然在立体图像评价问题上取得了一定的成果,但均没有考虑通用性,也没有将图像显著性作为研究对象,而且很少虑及立体图像左右视点权重分配问题,因此,在上述研究的基础上本文提出了一种基于卷积神经网络的无参考立体图像质量评价方法,并通过试验分析验证了算法的有效性和鲁棒性。

## 1 理论基础

### 1.1 融合图生成原理

立体图像左右视点的融合图采用文献[1]中的方法,又叫中央眼图。融合图像的生成公式如下:

$$\text{Merge}(x,y) = W_L(x,y)I_L(x,y) + W_R(x+d,y)I_R(x+d,y)。 \quad (1)$$

式(1)中:Merge为立体图像左右视点的融合图像; $W_L$ 和 $W_R$ 分别为左右视点的权值; $I_L$ 和 $I_R$ 分别为立体图像的左视点图像和右视点图像; $(x,y)$ 表示图像某一像素点位置; $d$ 为立体图像左右视点的视差值。 $W_L$ 和 $W_R$ 是由归一化Gabor滤波器幅值响应得到的,其公式如下:

$$W_L(x,y) = \frac{G_L(x,y)}{G_L(x,y) + G_R(x+d,y)}; \quad (2)$$

$$W_R(x,y) = \frac{G_R(x+d,y)}{G_L(x,y) + G_R(x+d,y)}。 \quad (3)$$

式(2)~(3)中: $G_L$ 和 $G_R$ 分别为立体图像左右视点的卷积响应和。滤波器的形式如式(4)所示,是一个复数Gabor滤波器:

$$G(x,y,\sigma_x,\sigma_y,\zeta_x,\zeta_y,\theta) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(1/2)[(R_1/\sigma_x)^2 + (R_2/\sigma_y)^2]} e^{i(x\zeta_x + y\zeta_y)}。 \quad (4)$$

式(4)中: $\sigma_x$ 和 $\sigma_y$ 分别为椭圆高斯包络线 $x$ 轴和 $y$ 轴的标准差; $\zeta_x$ 和 $\zeta_y$ 分别为 $x$ 方向和 $y$ 方向的空间频率; $\theta$ 为滤波器方向; $R_1 = x\cos\theta + y\sin\theta$ , $R_2 = -x\sin\theta + y\cos\theta$ 。

## 1.2 深度学习原理

深度学习是一种机器学习技术,教计算机执行人类与生俱来的活动:从示例中学习。在深度学习中,计算机模型直接从图像、文本或声音中学习执行分类任务或其他任务。深度学习模型可以达到相对最优的精确度,有时甚至超过人类的表现。该模型使用大量的标签化数据及包含许多层的神经网络架构训练得到。深度学习广泛应用于自动驾驶、航空、国防、医学、工业自动化、电子学等领域。大多数深度学习方法使用神经网络架构,这也是深度学习模型通常被称为深度神经网络的原因。深度一词通常是指神经网络中的隐藏层数。传统神经网络只包含 2~3 个隐藏层,而深度网络可能多达 150 个隐藏层。深度学习模型通过使用大量的标签化数据进行训练,而神经网络架构直接通过数据学习特征,不需要手动提取特征。卷积神经网络(convolutional neural network, CNN)是最流行的深度神经网络类型之一。CNN 使用二维卷积层学习输入数据的特征,这样的方式很适合处理二维数据(例如图像)。CNN 通过直接从图像提取特征来运作,这种自动化的特征提取使深度学习模型能够为计算机视觉任务(如图像质量评价)提供高精度。

## 2 模型方法

基于卷积神经网络的图像质量评价方法整体结构如图 1 所示,将立体图像的融合图像输入质量图生成网络中,得到相应的质量评分,然后加权融合得到最终的质量评分。质量图生成模块和加权融合模块是基于卷积神经网络的图像质量评价方法的主要组成部分。

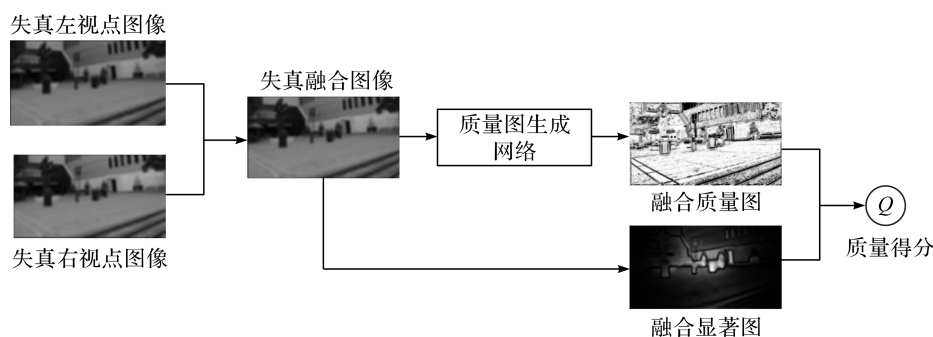


图 1 基于卷积神经网络的图像质量评价方法的整体结构

Fig. 1 Overall structure diagram of image quality evaluation method based on convolutional neural network

### 2.1 质量图生成模块

质量图生成网络模型主要采用 SegNet 网络<sup>[13]</sup>,其主要组成层为卷积层,在编码器部分,基本结构是两层卷积加一层池化作为一个小模块,输入失真图像的大小为  $w \times h \times 3$ ,输出质量图的大小为  $w \times h$ ,卷积核的大小都是  $3 \times 3$ ,激活函数采用线性修正单元(ReLU)函数。第 1 层具有  $C$  个滤波器的卷积层,其第  $i$  个滤波器的输出

$$y_i^l = s \left( \sum_{j=1}^{C^{l-1}} f_{i,j}^l y_j^{l-1} + b^l \right). \quad (5)$$

式(5)中:  $s$  为激活函数;  $f_{i,j}^l$  为第  $l$  层卷积第  $i$  个滤波器,并与第  $l-1$  层的第  $j$  个特征图相连接;  $b^l$  为第  $l$  层偏置向量。

网络中使用的监督标签是基于图像结构相似度(feature similarity, FSIM)的质量图,局部求 FSIM 质量图的效果要好于全局<sup>[14]</sup>。因此,本文中的 FSIM 质量图是局部映射矩阵。

### 2.2 加权融合模块

由于人眼对不同的立体图像的视觉关注度不一样,关注度越高的地方如果失真程度越大,则会严重影响立体视觉的整体质量,因此本文采用视觉显著的方法对质量图进行加权,得到最终的立体图像视觉质量

$$Q = \frac{\sum_x Q_{\text{map}}(x) S_{\text{map}}(x)}{\sum_x S_{\text{map}}(x)}. \quad (6)$$

式(6)中:  $Q_{\text{map}}$  为立体图像融合视点图像的预测质量图;  $S_{\text{map}}$  为立体图像融合视点图像的视觉显著图;  $x$  为立体图像融合视点图像的尺寸。

### 3 试验结果与分析

#### 3.1 试验数据库

将三维图像质量评价数据库 LIVE Phase I<sup>[15]</sup>和 LIVE Phase II<sup>[16]</sup>用于试验结果对比分析。

LIVE Phase I: 20 对未失真的原始参考立体图像, 365 对失真立体图像; 每个视点图像的大小为 640 pixel×360 pixel; 包含 JPEG 压缩失真、JPEG2000 压缩失真、加性高斯白噪声(white noise, WN)失真、快速衰减(fast fading, FF)通道失真和高斯模糊(gaussian blur, GB)失真 5 种不同程度的失真; 每对失真图像均为对称失真, 且都有一个差分均值意见得分(different mean opinion score, DMOS)。

LIVE Phase II: 8 对未失真的原始参考立体图像, 360 对失真立体图像; 每个视点图像的大小为 640 pixel×360 pixel; 与 LIVE Phase I 一样, 包含 5 种不同程度的失真; 针对每种失真类型, 每对原始立体图像生成 3 组对称失真立体图像对和 6 组非对称失真立体图像对, 非对称表示立体图像的左右视点为不同类型的失真或不同的失真等级; 每对立体图像都有一个 DMOS 值。

#### 3.2 试验训练阶段

试验在操作系统 64 位的计算机上实现, 使用 NVIDIA GTX 1080 TI 显卡。深度融合网络框架是在以 TensorFlow 为后端的 Keras 深度框架上运行的。在训练深度神经网络时, 使用的优化器是 Adam。在迭代过程中, 批量大小(batch size)为 16, 次数(epochs)为 100, 损失(loss)为均方误差(mean square error, MSE)。

#### 3.3 试验结果分析

利用皮尔森线性相关系数(Pearson linear correlation coefficient, PLCC)、斯皮尔曼秩相关系数(Spearman rank order correlation coefficient, SROCC)、均方根误差(root mean square error, RMSE)3 个评价指标对比模型的精确度。用 PLCC 度量模型的预测准确性, 若 PLCC 相关系数的值越接近 1, 则说明该图像质量评价算法的性能越好。由于我们创建的平面数据库只包含 JPEG、JPEG2000、WN 和 GB 4 种失真类型, 所以试验中只对这些失真进行测试分析。表 1 和表 2 分别为用 PLCC 和 SROCC 两个性能指标将本文方法与一些经典评价方法进行对比的结果。对比的 6 种方法分别是 SSIM、FSIM、GMSD、

表 1 本文方法与其他方法在 PLCC 评价指标上的结果对比

Table 1 Result comparison among the method of this paper and several other methods regarding PLCC evaluation index

数据库	失真类型	图像质量评价方法						
		SSIM	FSIM	GMSD	Chen	Bensalma	Shao	本文方法
LIVE Phase I	JPEG2000	0.868 3	0.936 6	0.928 3	0.855 3	0.848 4	0.872 3	0.910 7
	JPEG	0.495 5	0.600 7	0.651 8	0.475 7	0.376 3	0.597 2	0.612 9
	WN	0.937 8	0.931 4	0.947 4	0.953 3	0.913 7	0.916 1	0.939 4
	GB	0.912 0	0.933 4	0.938 1	0.938 5	0.915 7	0.923 3	0.921 5
	所有失真	0.898 9	0.935 7	0.943 2	0.928 4	0.894 9	0.899 4	0.912 9
LIVE Phase II-Symmetric	JPEG2000	0.816 2	0.818 3	0.874 9	0.670 4	0.689 7	0.903 2	0.895 3
	JPEG	0.677 0	0.845 6	0.844 3	0.601 3	0.551 4	0.873 2	0.891 7
	WN	0.974 9	0.963 0	0.961 3	0.946 3	0.935 9	0.917 0	0.982 4
	GB	0.832 5	0.863 8	0.927 9	0.917 8	0.952 7	0.977 3	0.962 8
	所有失真	0.732 6	0.830 1	0.924 5	0.813 5	0.823 2	0.911 9	0.921 9
LIVE 3D Phase II-Asymmetric	JPEG2000	0.675 5	0.784 6	0.868 0	0.722 0	0.619 4	0.789 3	0.876 5
	JPEG	0.684 5	0.796 3	0.869 0	0.563 6	0.630 5	0.705 2	0.689 7
	WN	0.823 0	0.941 0	0.916 0	0.944 9	0.932 5	0.923 5	0.964 9
	GB	0.840 3	0.887 9	0.741 1	0.691 8	0.862 1	0.854 7	0.952 9
	所有失真	0.749 7	0.677 5	0.653 3	0.633 7	0.743 2	0.565 1	0.766 1

表 2 本文方法与其他方法在 SROCC 评价指标上的结果对比

Table 2 Result comparison among the method of this paper and several other methods regarding SROCC evaluation index

数据库	失真类型	图像质量评价方法						
		SSIM	FSIM	GMAD	Chen	Bensalma	Shao	本文方法
LIVE 3D Phase I	JPEG2000	0.866 9	0.900 8	0.904 8	0.871 2	0.817 1	0.900 3	0.876 2
	JPEG	0.455 5	0.562 7	0.609 6	0.435 0	0.328 3	0.607 3	0.594 7
	WN	0.937 8	0.930 1	0.946 6	0.938 7	0.905 5	0.926 5	0.931 2
	GB	0.898 6	0.925 2	0.937 2	0.920 9	0.917 9	0.923 5	0.861 1
	所有失真	0.882 2	0.912 5	0.922 0	0.882 2	0.840 2	0.894 1	0.891 0
LIVE 3D Phase II-Symmetric	JPEG2000	0.726 1	0.824 3	0.866 9	0.661 7	0.607 8	0.904 3	0.882 2
	JPEG	0.718 0	0.840 6	0.838 0	0.630 4	0.547 5	0.910 2	0.883 1
	WN	0.945 2	0.936 5	0.926 9	0.907 0	0.924 3	0.936 5	0.939 6
	GB	0.770 4	0.849 5	0.835 6	0.845 0	0.846 1	0.911 3	0.649 8
	所有失真	0.700 3	0.908 6	0.910 2	0.837 2	0.804 6	0.896 6	0.893 1
LIVE 3D Phase II-Asymmetric	JPEG2000	0.723 7	0.804 6	0.853 6	0.722 0	0.619 4	0.789 3	0.871 9
	JPEG	0.714 4	0.805 0	0.875 8	0.635 9	0.677 9	0.696 1	0.691 1
	WN	0.882 1	0.952 1	0.936 6	0.929 2	0.940 9	0.923 5	0.949 1
	GB	0.806 8	0.850 1	0.887 7	0.691 2	0.840 2	0.803 1	0.841 5
	所有失真	0.719 3	0.661 0	0.642 0	0.610 8	0.696 8	0.524 4	0.709 5

Chen<sup>[1]</sup>、Bensalma<sup>[11]</sup>和 Shao<sup>[12]</sup>,前 3 种为平面全参考图像评价方法,后 3 种为立体图像质量评价方法。所有失真包含 JPEG2000、JPEG、WN、GB 4 种失真类型。PLCC 系数显示,本文方法在整体上占优势;SROCC 系数对比结果表明,本文方法应用于非对称失真立体图像会有更好的效果。由于本文为无参考评价方法,本文方法可能略低于全参考评价方法,但相比较于无参考方法,本文方法确实达到了一定的预期。

#### 4 结 语

本研究基于自制的失真平面图像数据集提出了一种有效的立体图像质量评价方法。首先以自制平面图像数据集及其相应的高性能全参考图像质量评价方法求得的质量图为标签,对质量图生成网络进行训练,利用深度卷积神经网络能够较好地模拟人类视觉系统神经元的生理功能,使得网络的预测与人类主观得分保持高度一致;然后使用训练好的网络对融合视点进行质量图预测;最后经加权融合得到立体图像的质量得分。在此基础上,将本文方法在两个知名的三维 LIVE 质量评价数据库进行试验分析,结果表明我们提出的算法在一定程度上提高了图像质量评价的性能。

#### 参考文献:

- [1] CHEN M J, SU C C, KWON D K, et al. Full-reference quality assessment of stereopairs accounting for rivalry[J]. Signal Processing: Image Communication, 2013, 28(9): 1143.
- [2] LV Y Q, YU M, JIANG G Y, et al. No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network[J]. Signal Processing: Image Communication, 2016, 47: 346.
- [3] ZHANG W, QU C F, MA L, et al. Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network[J]. Pattern Recognition, 2016, 59: 176.
- [4] SUN G M, DING Y, DENG R Z, et al. Stereoscopic image quality assessment by considering binocular visual mechanisms[J]. IEEE Access, 2018, 6: 2169.
- [5] JIANG G Y, XU H Y, YU M, et al. Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions[J]. Journal of Visual Communication and Image Representation, 2017, 46: 269.
- [6] HACHICHA W, KAANICHE M, BEGHDAI A, et al. No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models[J]. Signal Processing: Image Communication, 2017, 54: 107.

- [7] YANG J L, AN P, MA J, et al. No-reference stereo image quality assessment by learning gradient dictionary-based color visual characteristics[C]//International Symposium on Circuits and Systems(ISCAS). Florence: IEEE,2018:1.
- [8] WU J J, ZENG J C, DONG W S, et al. Blind image quality assessment with hierarchy: degradation from local structure to deep semantics[J]. Journal of Visual Communication and Image Representation,2019,58:353.
- [9] JIANG Q P, SHAO F, GAO W, et al. Unified no-reference quality assessment of singly and multiply distorted stereoscopic images[J]. IEEE Transactions on Image Processing,2019,28(4):1866.
- [10] LIU T J, LIN C T, LIU H H, et al. Blind stereoscopic image quality assessment based on hierarchical learning[J]. IEEE Access,2019,7:8058.
- [11] BENSALMA R, LARABI M C. A perceptual metric for stereoscopic image quality assessment based on the binocular energy[J]. Multidimensional Systems and Signal Processing,2013,24(2):281.
- [12] SHAO F, LIN W S, WANG S S, et al. Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook[J]. IEEE Transactions on Broadcasting,2015,61(2):154.
- [13] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(12):2481.
- [14] ZHANG L, ZHANG L, MOIU X, et al. FSIM: a feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing,2011,20(8):2378.
- [15] MOORTHY A K, SU C C, MITTAL A, et al. Subjective evaluation of stereoscopic image quality[J]. Signal Process: Image Communication,2013,28(8):870.
- [16] CHEN M J, CORMACK L K, et al. No-reference quality assessment of natural stereopairs[J]. IEEE Transactions on Image Processing,2013,22(9):3379.

(上接第18页)

用一对圆柱齿轮连接,使得传动机构结构简单、稳定可靠、工作寿命长。3)圆盘运动箱椭圆弧槽利用椭圆长短轴之差的特性来控制开合螺母与丝杠啮合状态的结构,在智能窗户设计中具有新颖性,能为防护装置的设计提供参考。

#### 参考文献:

- [1] 玫昆仑.“儿童坠楼身亡”事件的警示[N].中国妇女报,2018-06-10(2).
- [2] 罗晓薇.合肥一5岁女童爬窗坠亡事发时父母不在家 5月已发生2起儿童坠楼[EB/OL]. [2017-05-01][2019-04-22]. <http://365jia.cn/news/2017-05-07/452305E64B27DFCD.html>.
- [3] 李思杰,段蕾蕾.儿童跌落伤害预防研究进展[J].中国健康教育,2010,26(11):873.
- [4] 胡晓静,刘芳,沈国妹,等.住院患儿跌落发生率的监测与分析[J].中国护理管理,2017,17(7):962.
- [5] 来天皓,陈森林.多功能窗户控制系统的设计和应用[J].电子科技,2015,28(5):147.
- [6] 胡钰,杨应洪,刘欣,等.基于STM32智能窗户的设计与实现[J].机械工程与自动化,2019,49(2):105.
- [7] 李彦青.基于平移式窗户的智能启闭器的创新设计[J].中国新技术新产品,2018,23(3):114.
- [8] 张黎娜.防止儿童坠楼,防护网和监护都应加密[J].安全与健康,2013,20(8):32.
- [9] TUN T T, ELARA M R, KALIMUTHU M, et al. Glass facade cleaning robot with passive suction cups and self-locking trapezoidal lead screw drive[J]. Automation in Construction,2018,96:180.
- [10] 宗心南,李辉,武华红.中国7岁儿童重量指数的生长规律及参照标准建立[J].中国循证儿科杂志,2011,6(2):120.
- [11] 荆丽丽,弓美桃,赵鹏宇.基于传感器与PLC综合设计型实验探究:以银行一米线外报警提示系统为范例[J].集宁师范学院学报,2015,37(3):110.
- [12] 万文,吴敏.TTL三态非门电路的问题与改进[J].电子技术与软件工程,2018,7(21):69.
- [13] 王中宝.基于虚拟仪器的时序逻辑电路仿真实验平台设计[J].南方农机,2019,50(6):20.
- [14] 周梦茜,周川琦.执行机构限位开关装置的改进[J].自动化与仪器仪表,2016,23(8):123.