基于卷积神经网络的立体图像舒适度客观评价

李素梅，常永莉*，段志成
天津大学电气自动化与信息工程学院，天津 300072

摘要 基于卷积神经网络模型，提出一种立体图像舒适度评价方法。该方法无须提前根据特定的任务从图像中人工提取具体的特征，而是模拟人脑处理机制对图像进行层次化的抽象处理，自主提取特征。该方法采用三通道卷积神经网络结构，分别对原始图像进行主成分分析，以及 32×32、256×256 两种尺度的分块处理得到三条通道的输入数据集，根据输入数据设计每条通道的网络结构。采用两种尺寸分块处理得到不同尺寸的图像块特征信息，采用主成分分析降维处理得到原始图像的整体信息。此外，通过随机丢弃、局部响应归一化等方法提升算法的评价性能。实验结果表明，以修正线性单元为激活函数、输出层用 Softmax 分类器，对天津大学 TJU 立体图像数据库中 400 幅不同舒适度等级的立体图像样本进行测试，等级分类率正确达 94.52%，优于极限学习机、支持向量机算法。

关键词 图像处理；立体图像舒适度；客观评价；卷积神经网络；主成分分析；多尺度分块
中图分类号 TN911.73 **文献标识码** A **doi**：10.3788/AOS201838.0610003

Objective Assessment of Stereoscopic Image Comfort Based on Convolutional Neural Network

Li Sumei, Chang Yongli, Duan Zhicheng
School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, *China*

Abstract We propose a new method for stereoscopic image comfort assessment based on convolutional neural network, which does not need to extract specific manual features from images in advance according to specific tasks, but simulates hierarchical abstract processing mechanism of human brain to extract image features autonomously. This method adopts three channel convolutional neural network structure, and the input data sets of the three channel are obtained by reducing the dimension of the original data samples through principal component analysis, and chopping the original data samples into two size image patches (32×32, 256×256), respectively. The network structure of each channel is designed according to the input data sets. In addition, the classification accuracy of this method is improved by introducing dropout and local response normalization, etc. With rectified linear unit as the activation function and Softmax as the classifier in the output layer, experiment results on 400 stereo image samples in TJU database with different comfortable levels show that, the correct classification rate of this method is 94.52%, which is higher than that of the extreme learning machine and support vector machine.

Key words image processing; stereoscopic image comfort; objective assessment; convolutional neural network; principal component analysis; multi-scale blocking
OCIS codes 100.6890; 110.3000; 110.2970; 330.5000

1 引 言

立体图像能够带给人们身临其境的视觉体验，但立体图像从产生到呈现于人眼的整个过程需要经过采集、压缩、编码、存储及显示等步骤，难免引入噪声，导致观看者视觉不舒适。因此，如何实时有效地评估立体图像的舒适度已成为立体成像领域的关键问题之一。立体图像舒适与否的判定与通常所说的

收稿日期：2017-12-01；**收到修改稿日期**：2018-01-03
基金项目：国家自然科学基金(61520106002,161471262)
作者简介：李素梅(1975—)，女，博士后，副教授，硕士生导师，主要从事立体信息处理和计算机视觉方面的研究。E-mail：tjnklsm@163.com
* **通信联系人**。E-mail：cyl920611@163.com

质量等级相对一致，下文不妨将立体图像舒适度等级的判定称为立体图像质量评价。立体图像的质量评价通常分为主观评价和客观评价：主观评价方法是组织许多观测者参与图像评价实验后给出对应的主观质量分数，评价结果更加接近于人的真实感受，但该方法费时费力，不易操作；客观评价方法是通过客观模型给出立体图像的质量分数，能够有效克服主观评价方法的不足。因此，建立一套可以准确反映人眼主观感受的立体图像质量客观评价机制具有重要的意义[1]。

近年来，相关研究机构对立体图像质量评价算法进行了较为深入的研究。早期是直接将平面图像质量评价的相关指标应用于立体图像质量评价上，如峰值信噪比[2]、均方误差[3]、结构相似度[4]等。然而，立体图像相较于平面图像包含更多的深度信息，直接将平面图像质量评价算法应用于立体图像质量评价，不符合人眼的主观感受[5]，须将平面图像质量评价算法与立体图像中的一些立体信息相结合，进而提升评价效果。由于人眼视觉系统比较复杂，单纯结合一些立体信息的平面评价算法与人眼的主观感受仍不完全相符。一些研究人员尝试采用能够模拟人类大脑的神经网络进行立体图像质量评价，取得了良好的效果。文献[6]使用独立成分分析提取立体图像的有效特征，基于二叉树的支持向量机算法提出一种应用于立体图像质量客观评价的分类器，能够分类识别立体图像的质量；文献[7]考虑到传统神经网络学习速度慢、泛化能力差等缺点，首先通过主成分分析(PCA)对原始图像进行预处理，之后引入极限学习机(ELM)[8]对立体图像质量进行客观评价，但由于 ELM 网络的初始参数，即输入权重和阈值随机给定，导致网络的性能不稳定。

近几年，深度学习成为机器学习领域的研究热点。与传统的机器学习相比，深度学习更能模拟人脑深层次处理数据的方式，使得原始数据的内部结构和关系得到很好的层次化特征表示。通过深度学习方法自动提取的特征更加符合人脑的处理机制，大幅提升了网络模型的稳定性和泛化能力。卷积神经网络(CNN)是深度学习的一种典型网络，已被广泛应用于图像分类、语义识别等任务。CNN 在图像分类任务中以图像作为输入，并将特征学习和训练合为一体，能够高效地学习复杂的非线性关系[9]。文献[10]将尺寸较小的图像输入构建的 CNN 进行交通信号平面图像分类，取得良好的效果。文献[11]先将大尺寸输入图像切割成相同尺寸的图像块，然后把得到的图像块送入到构建的神经网络模型中得到质量分数。然而，将原始大图切割成小尺寸的图像块可能破坏原始图像的结构信息，进而影响立体图像质量评价。针对以上问题，提出基于三通道 CNN 的立体图像质量评价模型。三条通道中：一条的输入数据集是通过对原始图像进行 PCA 降维处理得到，旨在修改图像尺寸的同时保持图像的结构信息；另外两条通道的输入数据集分别通过将原始图像切割成 32 pixel×32 pixel、256 pixel×256 pixel 尺寸的图像块而得到，以避免单一尺寸图像分块对模型稳健性的影响。通过构建的三通道网络模型提取图像特征，之后将提取的特征输入 Softmax 分类器完成立体图像的质量评价，最后通过大量对比实验验证本文算法的有效性。

2　特征预处理

CNN 主要通过卷积和池化操作提取图像特征，若输入的图像尺寸太大，网络模型会很复杂，训练难度加大。因此，应对尺寸较大的输入图像预处理，本文采用图像分块和 PCA 降维两种方式。

2.1　图形分块预处理

图像分块预处理可将一幅大尺寸图像变成若干幅小尺寸图像。本文对原始图像进行两种尺寸的分块预处理，得到尺寸分别为 32×32、256×256 的图像块(这两种尺寸通过实验仿真得来)。处理的步骤如下。

1) 图像分块。假设图像尺寸为 $M\times N$，图像块的尺寸设置为 $p\times p$，那么分块后得到的图像块数量为

$$P=(M/p)(N/p)。\tag{1}$$

2) 图像块归一化处理。为了产生相似数量级像素值的图像块，需要对图像块进行归一化处理，处理规则为

$$\hat{I}(x,y)=\frac{I(x,y)-\mu(x,y)}{\sigma(x,y)+c},\tag{2}$$

$$\mu(x,y)=\frac{1}{m\times n}\sum_{(i,j)\in\Omega}I(x+i,y+j),\tag{3}$$

$$\sigma(x,y)=\frac{1}{m\times n}\sqrt{\sum_{(i,j)\in\Omega}[I(x+i,y+j)-\mu(x,y)]^2},\tag{4}$$

式中 $I(x,y)$ 表示 (x,y) 处的像素值，$\mu(x,y)$ 表示区域块的均值，$\sigma(x,y)$ 表示区域块的标准差，c 表示

任意的一个极小正数，以防止分母为 0，Ω 表示计算均值与方差的局部区域，i 和 j 表示区域块中坐标点移动的幅度，m 和 n 分别表示局部区域 Ω 的长和宽，$\hat{I}(x,y)$ 表示归一化后的像素值。经过归一化处理，原始图像块变成均值为 0、方差为 1 的图像块。

2.2　PCA 算法

通过 PCA 算法对实验图像进行降维预处理，能够提取立体图像的整体有效信息，减少计算量及噪声等因素对实验的影响[12]。本质上，PCA 算法是一种线性映射算法，不会丢失图像的结构信息。本文对原始图像降维后的维度设置为 400，之后将 400 维特征转换成尺寸为 20×20 的图像，算法步骤如下。

给定 l 个样本，每个样本的尺寸为 $m'\times n'$，样本矩阵 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_l)^{\mathrm{T}}$，其中 $\boldsymbol{x}_i(i=1,2,\cdots,l)$ 为第 i 个样本构成的 $m'\times n'$ 维向量。

1）对样本矩阵 $\boldsymbol{X}$ 作中心化处理：

$$\boldsymbol{p}=\frac{1}{l}\sum_{i=1}^{l}\boldsymbol{x}_i, \tag{5}$$

$$\boldsymbol{d}_i=\boldsymbol{x}_i-\boldsymbol{p}, \tag{6}$$

式中 $\boldsymbol{p}$ 为样本数据的均值，$\boldsymbol{d}_i$ 为零均值向量，则零均值矩阵 $\bar{\boldsymbol{X}}=[\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_l]$。

2）计算样本数据的协方差矩阵：

$$\boldsymbol{C}=\frac{1}{l}\bar{\boldsymbol{X}}\cdot\bar{\boldsymbol{X}}^{\mathrm{T}}。 \tag{7}$$

3）利用奇异值分解定理，获得 $\bar{\boldsymbol{X}}\cdot\bar{\boldsymbol{X}}^{\mathrm{T}}$ 的特征值 λ_i，以及特征向量 $\boldsymbol{\xi}_i$。

4）对得到的特征值从大到小排序，选取前 k 个特征值（k 值设置为 400）及其与之对应的特征向量，并计算贡献率 r。贡献率 r 表示所定义的主成分在整个数据分析中所占的比重，因此选取特征值之和与所有特征值之和的比值来表示：

$$r=\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{l}\lambda_i}, \tag{8}$$

式中 λ_i 表示特征值向量 $\boldsymbol{\lambda}$ 中的第 i 个特征值。

5）将各个样本数据投影到所选择的特征向量组成的有效子空间中。

3　三通道 CNN 算法

CNN 类似于多层感知机，具有良好的并行处理、自学习和泛化能力[13]。典型的 CNN 结构由卷积层、下采样层和全连接层组成，如图 1 所示。原始图像首先在卷积层与滤波器进行卷积，得到若干特征图后，通过下采样层对特征进行模糊，逐层提取特征后，通过全连接层输出，用以识别图像的特征。

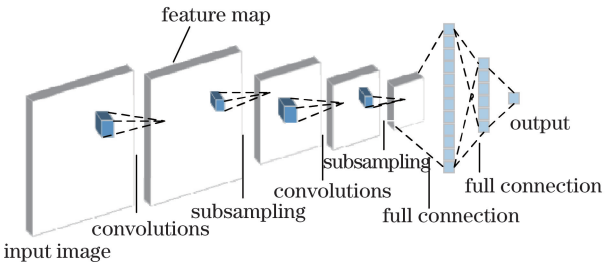

图 1　CNN 结构[14]

Fig. 1　Architecture of CNN

基于 CNN 构建三通道 CNN 模型，如图 2 所示。三通道 CNN 中的三条通道从上到下依次对应的输入数据集分别为 PCA 降维数据集和 32×32、256×256 数据集。每条通道的结构由卷积层和下采样层组成，之后在全连接层合并，最后使用 Softmax 分类器输出立体图像的质量分数。其中，下采样层使用最大值池化，所有卷积层和全连接层都配有修正线性单元（ReLU）激活函数[15]。为了避免过拟合，卷积层都配有局部响应归一化层（LRN）[16]，全连接层都配有随机丢弃（Dropout）层（丢失概率设置为 0.5）[17]。

三通道 CNN 算法的具体过程如下。

1）将样本分成训练集和测试集两部分，并将其读取成图像矩阵。

2）对训练数据集和测试数据集进行 PCA 降维与分块预处理，分别得到 PCA 降维数据集，以及尺寸为 32×32 和 256×256 的图像块数据集。

3）将预处理后的训练集数据送入构建的三通道 CNN 中进行训练，训练集训练好网络模型后，完成模型建立。

4）将测试集送入训练好的网络模型中，得出对应的图像质量分数。

4　实验结果与分析

4.1　实验环境及数据库

本文实验的模型基于 Caffe 深度学习框架搭建而成，仿真服务器硬件配置：英特尔 E5-2620 CPU，RAM 64G，GPU 为 NVIDIA Titan X。软件环境：Ubuntu 14.04 系统，Matlab R2014a，Caffe 深度学习框架。

实验中选取的原始立体图像数据均来自于天津

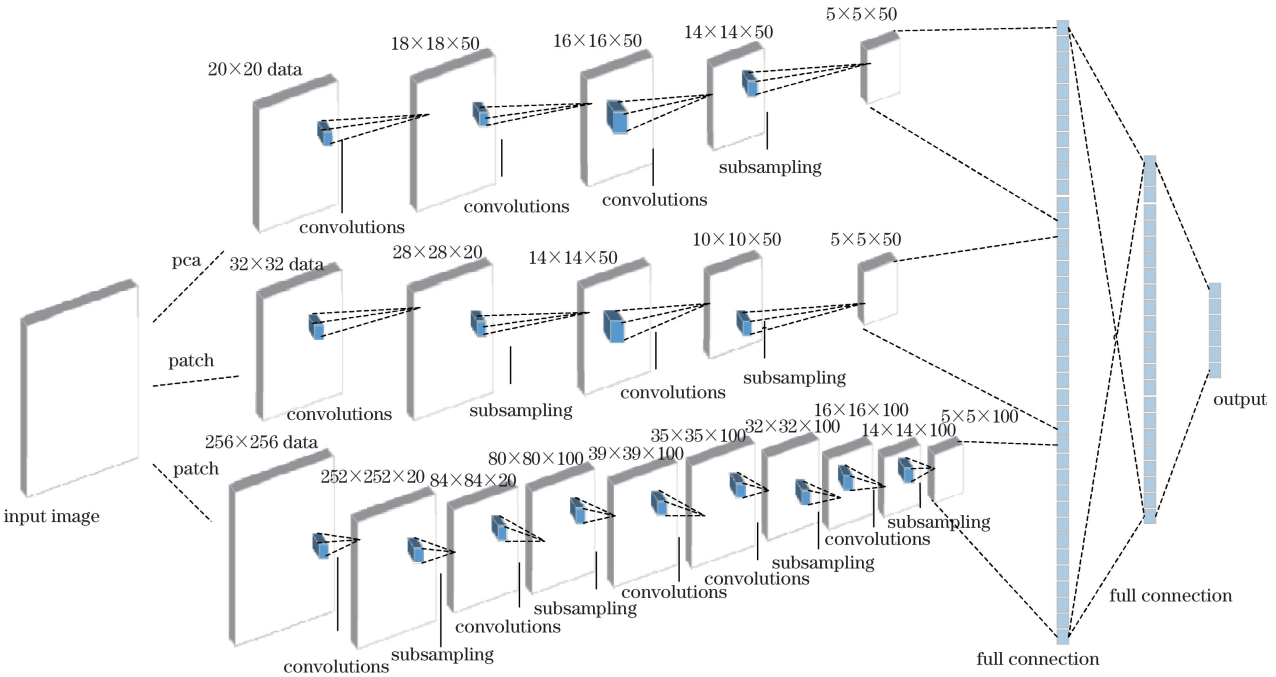

图 2　三通道的 CNN 结构

Fig. 2　Architecture of three-channel CNN

大学立体成像研究所提供的 TJU 立体图像数据库。数据库内立体图像的质量等级，根据国际电信联盟对立体图像质量的主观评价建议划分，如表 1 所示。将所有的立体图像质量分为 5 个等级：极好、好、一般、差、非常差，分别对应 5 、4、3 、2 、1 分。

表 1　立体图像质量划分建议

Table 1　Suggestions on quality division of stereoscopic images

Grade	Criteria for judging image damage	Degree of comfort
5	Almost no distortion	Excellent
4	Slightly distorted but not repugnant	Good
3	General distortion and a little repugnant	Fair
2	Obviously distorted but not disgusting	Poor
1	Serious distorted and disgusting	Bad

选择 TJU 立体图像数据库中包含上述 5 个质量等级的 400 幅立体图像进行实验。其中，样本图像的分辨率为 2560 pixel × 1024 pixel，并对这 400 幅立体图像进行镜像处理，扩充为 800 幅。立体图像数据库中的图像是通过对 4 幅源立体图像（无失真图像）进行不同程度的 JPEG 压缩、加噪、叠加失真及模糊失真处理得到，如图 3(a)～(d)所示。数据库中的部分失真图像如图 3(e)～(h)所示，分别是对图 3(a)进行模糊处理的 4 分立体图像、对图 3(b)经 JPEG 压缩处理的 1 分立体图像、对图 3(c)进行加噪处理的 2 分立体图像、对图 3(d)进行叠加失真的 3 分立体图像。从 800 幅立体图像样本中选取合适的 400 幅图像作为实验的训练数据集，剩余的 400 幅图像作为测试数据集。

图 3　(a)～(d)源图像和(e)～(h)失真图像

Fig. 3　(a)-(d) Source images and (e)-(h) distorted images

4.2　网络参数设置

三通道 CNN 模型的参数设置如表 2 所示：Pca_net表示 PCA 降维处理后数据集对应的通道网络；Patch_32_net 表示 32×32 图像块数据集对应的通道网络；Patch_256_net 表示 256×256 图像块数据集对应的通道网络；Conv 表示卷积层，“-”后数字表示该层卷积核的个数，括号中的数字表示卷积核的大小，如 Conv-100(3×3)表示该层是卷积层，卷积核的个数为 100，每个大小为 3×3；同理，Max 表示最大值池化，括号中表示池化核的大小；Fc 表示全连接层，“-”后数字表示全连接层的输出维度。

表 2　网络模型参数设置

Table 2　Parameter setting of network model

Patch_256_net	Patch_32_net	Pca_net
Conv-20(5×5)	Conv-20(5×5)	Conv-50(3×3)
Max(3×3)	Max(2×2)	Conv-50(3×3)
Conv-100(5×5)	Conv-50(5×5)	Conv-50(3×3)
Max(4×4)	Max(2×2)	Max(3×3)
Conv-100(5×5)	-	-
Conv-100(4×4)		
Max(3×3)		
Conv-100(3×3)		
Max(3×3)		
FC-2500		
FC-600		
FC-5		

4.3　结果分析

通过实验，不仅验证了一般分块预处理 CNN 算法与所提三通道 CNN 模型的性能，而且还分析了 Dropout 层与 LRN 层对所提模型的影响，最后比较本文算法和 ELM、支持向量机(SVM)在立体图像质量评价中的优劣。

本文模型是结合相应的理论知识、经验，以及大量实验而来。表 3 列出了不同网络结构对立体图像质量评价的影响，一种网络结构代表一种 CNN 算法。其中：Patch_32_net 表示图像块为 32×32 数据集的网络模型；Patch_256_net 表示图像块 256×256 数据集的网络模型；Patch_32_256_net 表示图像块 32×32、256×256 数据集相结合的网络模型；PCA_net 表示 PCA 数据集的网络模型；PCA_32_net表示 PCA 数据集网络与图像块 32×32 数据集网络相结合的网络模型；PCA_256_net 表示 PCA 数据集网络与图像块 256×256 数据集网络相结合的网络模型；PCA_32_256_net 表示 PCA 数据集网络与图像块 32×32、256×256 数据集网络相结合的网络模型(三通道 CNN 算法)。网络仿真参数如表 2 设置，激活函数设置为 ReLU，模型使用 Dropout 及 LRN 层。从表 3 可以看出，PCA_net 比 Patch_32_net、Patch_256_net 的性能更好，这是由于单纯分块得到的数据不能完全体现整幅立体图像的舒适度分数，局部图像块的舒适度分数不一定与整幅图像的舒适度相一致，而 PCA 算法是对一整幅图像进行处理，其数据大致能够表示整幅图像，得到的舒适度分数与整幅图像的舒适度分数基本一致。PCA_32_256_net 网络的识别率最高，这是由于该算法结合了 PCA 提取的特征和两种尺寸图像块所包含的特征，使得图像特征提取得更加充分，从而网络学习得更好。

表 3　不同网络结构下 CNN 模型对所有测试样本的识别率

Table 3　Recognition rates of test samples with different structures of CNN model

Structure of CNN model	Recognition rate/%
Patch_32_net	51.67
Patch_256_net	55.00
Patch_32_256_net	66.25
PCA_net	76.75
PCA_32_net	84.50
PCA_256_net	88.25
PCA_32_256_net	94.52

表 4 列出了 Dorpout 与 LRN 层对 PCA_32_256_net 网络性能的影响。从表 4 可以看出，在网络模型中同时加入 Dropout 与 LRN 层的性能更好。这是由于 Dropout 减少了神经元的复杂共适应性，增强了网络的稳健性，能够有效防止过拟合，加入 LRN 后，对网络下一层的输入进行了局部归一化，使得较大响应值变得相对更大，提高了模型的泛化能力。

表 4　Dropout 与 LRN 层对 PCA_32_256_net 识别率的影响

Table 4　Influence of Dropout and LRN layer on PCA_32_256_net model recognition rate

Optimization method		Recognition rate /%
Dropout layer	LRN layer	
No	Yes	93.20
Yes	No	93.40
Yes	Yes	94.52

表 5 列出了本文算法与 SVM[6]、ELM[7]针对 TJU 立体图像库的分类识别准确率。可以看出，本文算法分类识别正确率达 94.52%，高于其他两种算法。CNN 算法的训练过程是通过梯度下降算法不断自动调整网络参数，耗时较长（约 2.2 h），但提取的特征更加符合人脑的处理机制。但需要指出的是，本文方法训练时间的长短与仿真时使用的 GPU 性能和个数有关，并且在测试时，只需调用已经训练好的网络模型，完成一次前馈过程即可得到图像的舒适度分数，该过程耗时很短，在实际应用中并不影响用户体验。虽然 SVM 与 ELM 仿真所需时间短，但是二者在手动选择特征时亦需耗费大量时间，且无法保证手动调整参数后网络提取特征的质量。综上，本文算法对立体图像质量进行客观评价的结果与实际主观评价分数基本相符，能有效评价立体图像质量。

表 5　不同算法关于立体图像质量评价的识别率

Table 5　Recognition rates of test algorithms

Algorithm	Recognition rate /%	Train time /s	Test time /s
Proposed	94.52	7920.0000	0.1470
SVM[6]	92.50	20.2700	0.0047
ELM[7]	93.85	0.0025	0.0037

5　结　论

提出一种基于三通道 CNN 的立体图像舒适度评价算法，取得了较好的分类效果。CNN 模型通过使用 PCA 降维和分块两种预处理方式处理数据，并以三通道的形式进行网络优化，弥补了图像分块后结构信息被破坏，以及不同分块尺寸对系统稳健性的影响，使得 CNN 模型能够更好地提取图形特征。另外，通过使用 Dropout 和 LRN 层，提高了模型的评价准确性。同时，CNN 人工参与很少，为立体图像舒适度客观评价及系统的推广提供了有效途径。实验结果表明，本文算法在立体图像舒适度评价精度上优于 ELM 和 SVM，具有可行性。下一步的研究内容将重点考虑更加简单、高效的立体图像预处理方法，或者采用卷积稀疏字典等算法来提取特征，以获得更优的评价效果。

参　考　文　献

[1] Hou C P, Ma T T, Yue G H, *et al*. Multiply-distorted image quality assessment based on high-order phase congruence[J]. Laser & Optoelectronics Progress, 2017, 54(7): 071001.
侯春萍, 马彤彤, 岳广辉, 等. 基于高阶相位一致性的混合失真图像质量评价[J]. 激光与光电子学进展, 2017, 54(7): 071001.

[2] Yang J C, Hou C P, Shen L L, *et al*. Objective evaluation method for stereo image quality based on PSNR[J]. Journal of Tianjin University, 2008, 41(12): 1448-1452.
杨嘉琛, 侯春萍, 沈丽丽, 等. 基于 PSNR 立体图像质量客观评价方法[J]. 天津大学学报, 2008, 41(12): 1448-1452.

[3] Zhu Q S, Zhi L O, Liu R, *et al*. Research on image conversion from planar into stereo[J]. Computer Science, 2007, 34(7): 225-228.
朱庆生, 支丽欧, 刘然, 等. 平面图像立体化关键技术研究[J]. 计算机科学, 2007, 34(7): 225-228.

[4] Wang Z, Bovik A C, Sheikh H R, *et al*. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[5] Russo F, de Angelis A, Carbone P. A vector approach to quality assessment of color images[C]// Instrumentation and Measurement Technology Conference Proceedings, 2008: 814-818.

[6] Chen J C. Application of ICA and BT-SVM in stereo image quality assessment system[D]. Tianjin: Tianjin University, 2012: 41-45.
程金翠. ICA 和 BT-SVM 在立体图像质量评价系统中的应用[D]. 天津: 天津大学, 2012: 41-45.

[7] Wang G H, Li S M, Zhu D, *et al*. Application of extreme learning machine in objective stereo scopic image quality assessment[J]. Journal of Optoelectronics · Laser, 2014, 25(9): 1837-1842.
王光华, 李素梅, 朱丹, 等. 极端学习机在立体图像质量客观评价中的应用[J]. 光电子 · 激光, 2014, 25(9): 1837-1842.

[8] Bai J J, Sun Q, Jing S B, *et al*. Robust extreme learning machine and its application in analysis of near infrared spectroscopy data[J]. Laser & Optoelectronics Progress, 2015, 52(10): 103002.
白俊健, 孙群, 井诗博, 等. 稳健极限学习机及其在近红外光谱分析中的应用[J]. 激光与光电子学进展, 2015, 52(10): 103002.

[9] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Massachusetts, USA: The MIT Press, 2016: 331-339.

[10] Ciresan D, Meier U, Masci J, *et al*. Multi-column deep neural network for traffic sign classification[J]. Neural Networks, 2012, 32(1): 333-338.

[11] Lv Y, Yu M, Jiang G, *et al*. No-reference stereoscopic image quality assessment using binocular self-similarity

and deep neural network[J]. Signal Processing Image Communication, 2016, 47: 346-357.

[12] Cheng L Y, Mi G Y, Li S, *et al*. Quality diagnosis of joints in laser brazing based on principal component analysis: support vector machine model[J]. Chinese Journal of Lasers, 2017, 44(3): 0302004.
程力勇, 米高阳, 黎硕, 等. 基于主成分分析-支持向量机模型的激光钎焊接头质量诊断[J]. 中国激光, 2017, 44(3): 0302004.

[13] Li S M, Lei G Q, Fan R. Depth maps super-resolution reconstruction based on convolutional neural networks[J]. Acta Optica Sinica, 2017, 37(12): 1210002.
李素梅, 雷国庆, 范如. 基于卷积神经网络的深度图超分辨率重建[J]. 光学学报, 2017, 37(12): 1210002.

[14] Lecun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[15] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[C]// Proceedings of 30th International Conference on Machine Learning, 2013, 30(1): 3.

[16] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[17] Srivastava N, Hinton G, Krizhevsky A, *et al*. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.