

双波段彩色融合图像色彩和谐性客观评价^①



高绍姝, 宋尚鸽, 倪 潇

(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)

通信作者: 宋尚鸽, E-mail: xyy699@163.com

摘 要: 针对现有的图像质量评价方法较少利用人眼视网膜和视觉皮层的颜色编码机制, 并且未能充分考虑图像色彩信息对图像质量的影响, 提出了一种基于多视觉特征的可见光(微光)与红外彩色融合图像色彩和谐性客观评价模型. 该模型在图像质量评估中融入了更多的颜色信息, 综合考虑多种人眼视觉特征包括视觉对立色彩特征、色彩信息波动特征和高级视觉内容特征, 经过特征融合和支持向量回归训练, 实现彩色融合图像的色彩和谐性客观评价. 采用3种典型场景融合图像数据库进行实验比较与分析. 实验结果表明, 与现有的8种图像质量客观评价方法相比, 所提出的方法与人眼主观感受更加一致, 具有较高的预测准确度.

关键词: 图像质量评价; 彩色融合图像; 视觉对立色彩; 色彩信息波动; 高级视觉内容

引用格式: 高绍姝, 宋尚鸽, 倪潇. 双波段彩色融合图像色彩和谐性客观评价. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9515.html>

Objective Assessment of Color Harmony in Dual-band Color Fused Images

GAO Shao-Shu, SONG Shang-Ge, NI Xiao

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: The currently available quality assessment methods for images rarely fully utilize the color coding mechanisms of the retina of human eyes and the visual cortex and fail to fully consider the influence of color information on image quality. In this study, an objective assessment model for the color harmony of visible light (dim-light) and infrared color fused images based on multiple visual features is proposed to address the above problems. This model incorporates more color information into image quality assessment by considering a variety of visual features of human eyes comprehensively, including the feature of visual contrast colors, the feature of color information fluctuation, and the feature of advanced visual content. Through feature fusion and support vector regression training, it achieves the objective assessment of the color harmony of color fused images. Experimental comparisons and analyses are conducted using databases of fused images in three typical scenes. The experimental results show that compared with the existing eight methods of objective image quality assessment, the proposed method is more consistent with the subjective perception of human eyes and has higher prediction accuracy.

Key words: image quality assessment; color fusion images; visual contrasting colors; color information fluctuation; advanced visual content

彩色融合图像是通过彩色数字图像处理的方法将可见光(微光)和红外图像融合获得, 广泛应用于车辆夜间辅助驾驶、森林火情监测、海上救援等领域^[1,2],

彩色融合图像的质量在完成这些视觉任务中起着决定性作用. 融合图像包含人为赋予的假彩色信息, 图像色彩和谐性是影响图像质量的关键因素(融合图像中不

^① 基金项目: 国家自然科学基金(61801517); 中央高校基本科研业务专项经费(19CX02029A, 19CX02027A)

收稿时间: 2023-12-13; 修改时间: 2024-01-10; 采用时间: 2024-01-18; csa 在线出版时间: 2024-04-07

同的颜色组合产生令人满意、愉快的效果,则认为该图像颜色组合具有协调性),然而现实应用中缺乏理想的参考图像,更加大了融合图像质量评价的难度.科学有效地评价彩色融合图像的颜色和谐性成为当前国内外彩色图像融合技术发展亟待解决的问题之一.

人类视觉系统(human visual system, HVS)是感知彩色(夜视)融合图像质量的最终决策者.大规模规范的主观评价是衡量融合图像色彩和谐性最准确的方法.但是,主观评价实验耗时费力且非常昂贵.因此,自动预测彩色融合图像色彩质量的无参考客观评价方法成为当前研究的焦点.

近年来出现了大量的无参考图像质量评价算法(no-reference quality assessment, NR-IQA),其中一个分支是以自然场景统计(natural scene statistics, NSS)理论为基础,当图像质量下降时,会出现图像统计特性的偏离.基于NSS的评价算法由3个基本步骤组成:提取图像特征、基于NSS理论建模和图像特征回归分析.例如,ContourletIQA^[3]将图像转化为CIELAB色彩空间,然后再转化为图像梯度图,最后在轮廓波(contourlet)域中提取不同尺度的自然场景统计特征,完成图像质量预测.Zhang等^[4]提出的算法包含图像色彩、结构和频率特征.Zhou等^[5]通过多结构统计表示对比度失真图像的质量.

另一方面,机器学习和深度学习广泛应用于NR-IQA算法.深度学习可以将图像特征学习提取和回归分析结合,并且形成一个统一的图像质量客观评价优化模型^[6].Zhang等^[7]提出一种用于无参考图像质量评估的深度卷积神经网络双线性模型,该模型适用于合成失真和真实失真的图像.Golestaneh等^[8]又提出新的模型来解决NR-IQA任务,通过卷积神经网络(CNN)捕捉图像局部结构信息和非局部图像特征,对提取的特征使用变形器(Transformer)网络模型实现图像质量预测.考虑到彩色图像的色彩影响因素,Gu等^[9]将图像对比度、清晰度和亮度等特征通过机器学习回归模块得到图像质量预测分数.Ma等^[10]保持原始图像的布局,考虑图像成像原理、色彩和谐和焦距参数,使用卷积神经网络(convolutional neural network, CNN)评估图像美学分数.Lu等^[11]提出了一种基于隐式狄利克雷分布(latent Dirichlet allocation, LDA)的色彩和谐模型,该模型通过色彩和谐之间的空间距离来约束LDA的训练.之后,他们使用深度学习获取原始图

像色块之间的相关性,将这些相关性和每个色块的特征嵌入所设计的条件随机场(conditional random field, CRF)^[12],从而测量整个图像的色彩和谐分数并完成图像美学评估任务.

此外,生物视觉皮层的计算处理机制由于与HSV具有较强的相似性,在图像质量评价领域受到越来越多的关注.Gu等^[13]提出无参考图像质量评价需要以自由能人脑理论和基于HSV启示的图像特征为基础.Wu等^[14]基于HSV的方向选择性机制提出一种方向选择视觉模式提取图像视觉内容.后来,Wu等又提出了分层特征退化的无参考图像质量评价方法^[15],使用初级视觉皮层的定向选择计算机制,提取图像局部低层视觉信息特征,通过深度神经网络提取多层高级语义特征,获得图像分层特征集,最后建立了图像质量退化与对应分层特征集的相关性.

总的来说,以往的很多NR-IQA算法研究是将原始图像转化为灰度图像,图像不可避免地缺失了色彩信息.与此同时,这些方法未充分使用视网膜神经节细胞的对比度编码机制以及视网膜和视觉皮层的颜色编码机制.因此,为了解决这些问题,本文提出基于视觉特征的红外与可见光(微光)波段彩色融合图像色彩和谐性客观评价模型,称为双波段彩色融合图像色彩和谐性(dual-band color fusion image color harmony, DFICH)模型.通过模拟视网膜细胞的颜色编码机制,从原始融合图像中提取低层视觉对立色彩特征,以不同半径计算图像每个像素周围的色彩信息波动(color information fluctuation, CIF),并获得4个不同检测方向的色彩信息波动图(color information fluctuation map, CIFM),然后从CIFM提取色彩波动特征以更加全面地描述图像色彩信息.同时,使用多尺度特征集成的深度残差神经网络提取图像高层视觉内容特征,建立图像低层色彩信息与高层内容信息的紧密联系.最后,通过特征融合和支持向量回归(support vector regression, SVR)训练DFICH模型.利用包含3种典型场景的彩色融合图像数据库^[16]进行实验比较与分析,结果表明,所提出的方法与人眼主观感受评估更一致.

1 DFICH模型的基本原理

本文提出了一种双波段彩色融合图像色彩和谐性(DFICH)模型,如图1所示.模型的总体思路为:基于视网膜理论研究,将彩色融合图像转换为 lab 色彩空间,

根据 $l\alpha\beta$ 色彩空间中融合图像色彩分布的统计特性提取视觉对立色彩特征. 其次, 在不同检测方向上计算原彩色融合图像每个颜色通道像素之间的色彩相关性, 提取每个像素的CIF并获得CIFM, 从而提取色彩信息波动特征. 然后, 使用多尺度深度神经网络提取融合图像高层视觉内容特征. 最后, 将低层视觉色彩特征和高层视觉内容特征进行特征组合, 形成综合特征集, 通过SVR分析彩色融合图像的色彩和谐性.

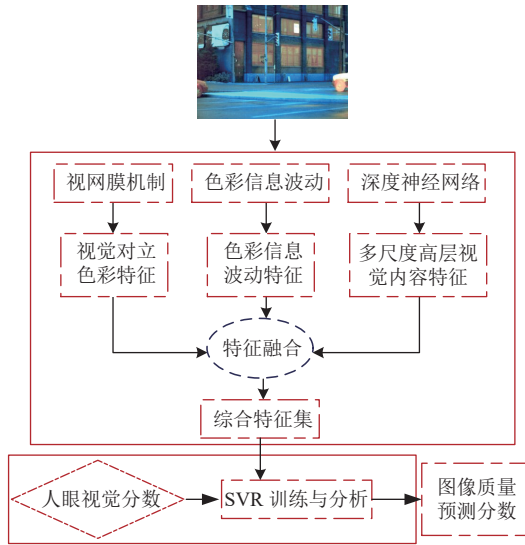


图1 DBICH模型框架图

1.1 视觉对立色彩特征

图像色彩特征是生物视觉系统所传达的低层次视觉特征, 是彩色融合图像色彩和谐性的主要影响因素. 根据视网膜研究^[17], 人类视网膜中的简单细胞将图像色彩编码为色彩对立差异. Daniel等提出的色彩空间^[18-20]包括3个对立颜色通道: 灰度对立通道、黄-蓝对立通道和红-绿对立通道, 它们与HVS的色彩对立是一致的. 同时, 色彩空间解除了红(R)、绿(G)、蓝(B)三通道之间的依赖关系, 便于单独处理图像的强度和图像的各个颜色通道. 因此, 从原始图像的色彩空间提取视觉对立色彩特征并用于DFICH.

图2中描绘了某一组海天场景8幅融合图像^[16], 并在图3中绘制了每一幅图像对应的黄-蓝对立通道(yellow-blue antagonism channel)和红-绿对立通道(red-green antagonism channel)的色彩分布. 根据图3的色彩分布图, 视觉对立颜色通道的图像色彩分布具有独特的统计属性. 因此, 在DFICH模型中考虑彩色融合图像色彩分布的变化.

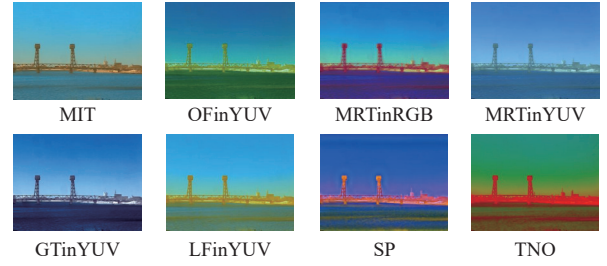
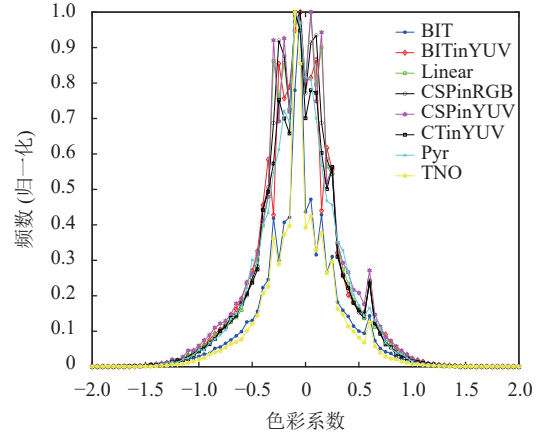
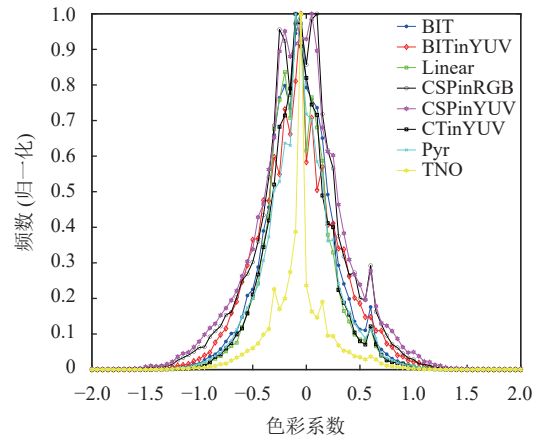


图2 某一组海天场景8幅融合图像



(a) 黄-蓝对立通道



(b) 红-绿对立通道

图3 图2对应融合图像的色彩分布

色彩分布使用非对称广义高斯分布(asymmetric generalized Gaussian distribution, AGGD)^[21]拟合:

$$f(x; \alpha, \beta_l^2, \beta_r^2) = \begin{cases} \frac{\alpha}{(\varphi_l + \varphi_r) \Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{|x|}{\varphi_l}\right)^\alpha\right), & x < 0 \\ \frac{\alpha}{(\varphi_l + \varphi_r) \Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{|x|}{\varphi_r}\right)^\alpha\right), & x \geq 0 \end{cases} \quad (1)$$

其中, x 是色彩像素.

$$\varphi_l = \beta_l \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \quad (2)$$

$$\varphi_r = \beta_r \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \quad (3)$$

$\Gamma(\cdot)$ 定义为:

$$\Gamma(\cdot) = \int_0^{\infty} t^{a-1} e^{-t} dt, \quad a > 0 \quad (4)$$

其中, α 控制分布形状, β_l 和 β_r 分别控制左右分布的扩散程度, 它们共同决定了色彩整体分布. 因此, 将这些参数作为视觉对立色彩特征用于 DFICH.

正如图3所示, 不同融合算法生成的融合图像的色彩分布偏差是显而易见的. 对于色彩和谐性较高的融合图像, 色彩分布向外大大扩展, 而色彩和谐性较低的色彩分布明显变窄. 因此, 使用黄-蓝对立通道和红-绿对立通道的视觉对立色彩特征作为 DFICH 特征是合理的. 差法与背景减法类似, 事实上它们是同一种运动目标检测算法的两种不同的变体, 它们的原理都是通过前景图像与背景图像做差运算, 从而检测出运动目标, 它们的优点是简单, 易于实现, 这个优点也使得它们具有较大的应用空间和较高的应用价值.

1.2 色彩信息波动特征

视觉对立色彩特征是建立在色彩空间的基础之上. 在色彩空间转化过程中, 融合图像的像素色彩信息会有部分丢失. 因此, 在不同的检测方向上计算每个颜色通道中相邻像素之间的色彩关系, 通过结合红 (R)、绿 (G)、蓝 (B) 三通道中像素强度的波动提取每个像素的 CIF. CIF 计算流程图如图4所示. 如图4(a)所示, 包含中心像素和一定半径的相邻像素用紫色填充, 4个检测方向如图4(b)所示, 它们分别是: 0° 、 45° 、 90° 、 135° . 根据4个检测方向综合分析中心像素以及相邻像素色彩的变化.

彩色融合图像其中一个像素在某一检测方向的 CIF 计算过程如图4(c)和图4(d)所示. 3个颜色通道的中心像素以及相邻像素的色彩向量表示为: $[V_l^R, V_l^G, V_l^B]$, $[V_c^R, V_c^G, V_c^B]$, $[V_r^R, V_r^G, V_r^B]$, 其中 R、G 和 B 是颜色通道, V_l^B 、 V_c^B 和 V_r^B 是蓝色通道中图像的中心像素和相邻像素的色彩分量. CIF 的计算方法如图4(d)所示, 其中 O 是三维坐标系的原点, O_l 和 O_r 的坐标是中心像素与相邻像素色彩分量之间的差异. $\overrightarrow{OO_l}$ 和 $\overrightarrow{OO_r}$ 之间的角度 ω 是当前检测方向上中心像素与相邻像素之间

的 CIF. 基于角度 ω 的取值范围是 $0-180^\circ$, 在 $0-180^\circ$ 之间是单调的, 因此使用 $\cos \omega$ 表示 CIF, 计算过程如下:

$$\begin{aligned} \cos \omega &= \frac{\overrightarrow{OO_l} \cdot \overrightarrow{OO_r}}{|\overrightarrow{OO_l}| \times |\overrightarrow{OO_r}|} \\ &= \frac{D_l^R \times D_r^R + D_l^G \times D_r^G + D_l^B \times D_r^B}{\sqrt{D_l^{R2} + D_l^{G2} + D_l^{B2}} \times \sqrt{D_r^{R2} + D_r^{G2} + D_r^{B2}}} \end{aligned} \quad (5)$$

$$D_s^{color} = V_s^{color} - V_c^{color} \quad (6)$$

其中, $color$ 是颜色通道, $color \in \{R, G, B\}$, s 用于标记相邻像素, $s \in \{l, r\}$.

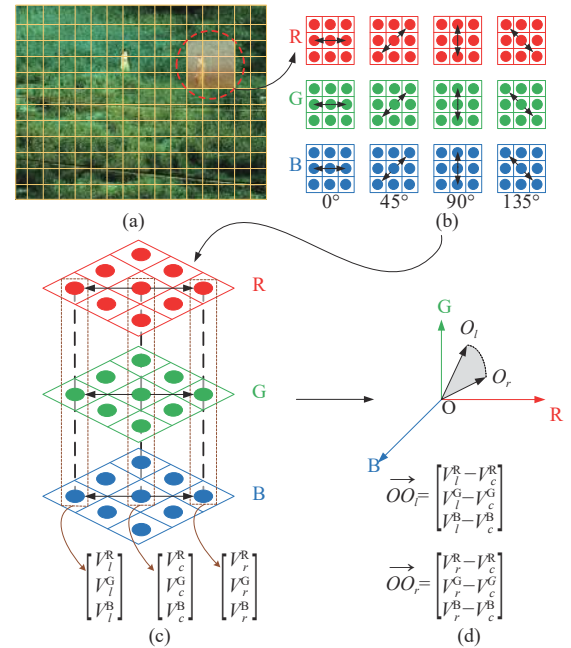


图4 色彩信息波动计算流程图

对于大小为 $W \times H$ 的彩色融合图像 I , 给定色彩像素 (i, j) , 那么 CIF 为:

$$Cifm_k(i, j) = Cif(I(i, j), k) \quad (7)$$

其中, $Cifm_k(i, j)$ 是图像 I 的 CIFM, k 是检测方向, $k \in \{1, 2, 3, 4\}$

图5中显示了数据库中的一组相同场景下的部分融合图像效果图. 图5(a)–(d)是城镇场景中4幅融合图像, 它们是该场景下人眼所标注的色彩和谐性较好的两幅融合图像和较差的两幅融合图像. 图5(a1)–(d1)、(a2)–(d2)、(a3)–(d3)、(a4)–(d4)分别对应不同检测方向上的 CIFM. 从图5可以看出, 同一场景下的不同融合图像具有不同结构的 CIFM, CIFM 可有效地表示不同检测方向的融合图像的色彩波动信息. 因此,

CIFM 可用于提取融合图像色彩质量感知特征。

1.3 高级视觉内容特征

色彩信息仅是图像低层视觉特征, 然而高层视觉特征在人类视觉感知中起着关键作用. 深度神经网络可以有效地表达高层视觉内容特征, 并以此进行视觉识别^[22]. 目前, 残差网络 (residual network, ResNet) 是其中一个功能强大且更深层的深度神经网络学习架构, 常常用于提取图像深层语义特征^[23]. 由于计算效率和复杂度, 本文选择 ResNet50^[23]的输出作为多尺度高层视觉内容特征, 网络框架如图 6 所示. 为了减少数据库大小对网络训练的影响, 本文使用的是经过 ImageNet 数据集^[24]初始化的 ResNet50.

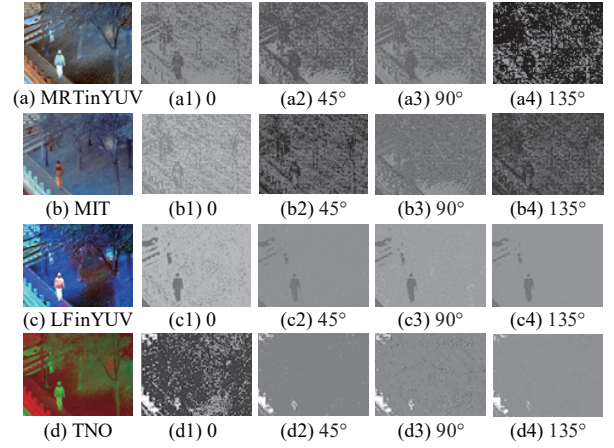


图 5 城镇场景融合图像及其 CIFM

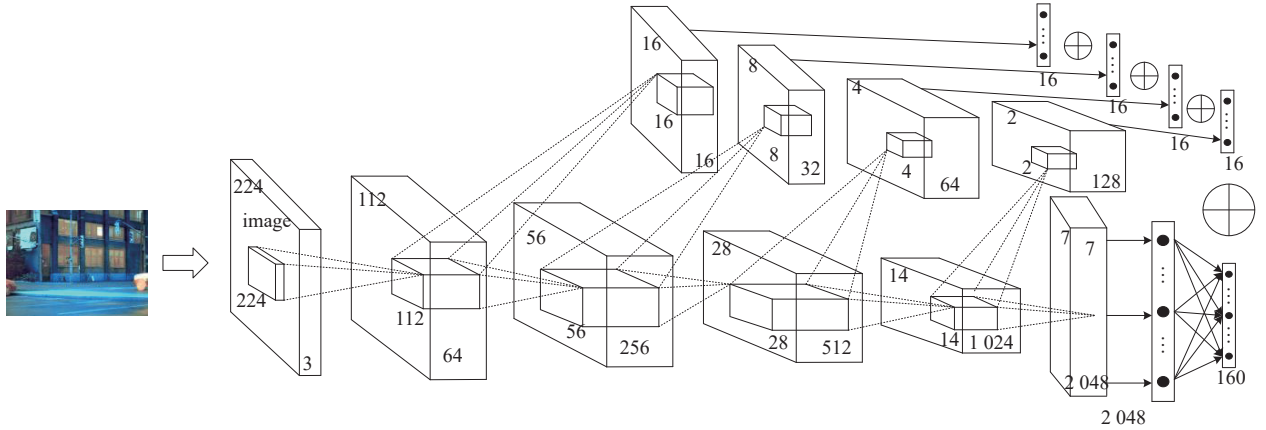


图 6 用于多尺度高级视觉特征提取的 50 层 ResNet 结构

1.4 DFICH 模型训练和质量预测

为了综合衡量彩色融合图像色彩和谐性, 首先, 将第 1.1、1.2、1.3 节中所提取的特征归一化以进行融合.

$$\tilde{f}_i(j) = \frac{f_i(j)}{\sqrt{\sum_n (f_i(n))^2}} \quad (8)$$

其中, f_i 可以是低层视觉对立色彩 (f_{lc})、色彩信息波动 (f_{cif}) 或高层视觉内容 (f_{hc}) 中的任意一个, $\tilde{f}_i(n)$ 是第 n 个归一化特征. 其次, 进行特征融合并获得综合特征集, 用于色彩和谐性分析. 最后, 通过支持向量回归 (support vector regression, SVR) 训练并学习彩色融合图像主观质量分数与综合特征集之间的映射关系.

$$DFICH_Model = SVR_train(F^{train}, Q^{train}) \quad (9)$$

其中, $DFICH_model$ 是双波段盲彩色融合图像质量评价模型, F^{train} 是训练图像的综合特征集, $F^{train} = \{F_1^{train}, F_2^{train}, \dots, F_{n_{training}}^{train}\}$, Q^{train} 是训练图像的主观评分集合,

$Q^{train} = \{Q_1^{train}, Q_2^{train}, \dots, Q_{n_{training}}^{train}\}$, $n_{training}$ 是训练集的图像数目. 对于待预测的彩色融合图像, 首先根据所提出的方法获取综合特征集 F^{test} , 然后通过 F^{test} 和 $DFICH_model$ 自动获得色彩和谐性分数, 方法如下:

$$q^{test} = SVR_predict(F^{test}, DFICH_model) \quad (10)$$

2 实验结果与分析讨论

本文所涉及的全部实验环境为: Windows server 2016, CPU Intel(R) Xeon(R) W-2145 CPU @ 3.70 GHz, 64 GB RAM, GPU 为 NVIDIA GeForce RTX 4000, Python 3.7, PyTorch 11.0, PyCharm 2022.

2.1 实验数据库与实验结果分析

本文是在 Jin 等^[16]研究基础上对彩色融合图像进行色彩和谐性客观评价. 实验数据库包括海天、植物和城镇等场景, 共 256 幅彩色融合图像, 每幅图像的像素均为 320×240 , 并由 52 人对图像进行色彩和谐性标

注. 这些彩色融合图像是选取了 32 组不同场景的红外图像与可见光 (微光) 图像, 对同一场景下的红外图像和可见光 (微光) 图像使用不同融合算法生成的, 其中包括 TNO (the Netherlands Organization for Applied Science Research)、MIT (Massachusetts Institute of Technology)、LFinYUV (linear fusion algorithm in YUV space)、OFinYUV (fusion algorithm based on opponent vision property in YUV space)、SP (steerable pyramid color fusion algorithm) 及 GTinYUV (global color transfer algorithm in YUV space)、MRTinRGB (color transfer algorithm based on multi-resolution in RGB space) 和 MRTinYUV (color transfer algorithm based on multi-resolution in YUV space, MRTinYUV), 8 种融合方法是目前实际使用的主要方法, 选择的原则为算法代码的公开和处理速度达到准实时以上. 图 7 给出典型场景植物的 1 组融合图像, 每组包含 1 幅可见光图像和 1 幅红外图像以及 8 幅彩色融合图像.

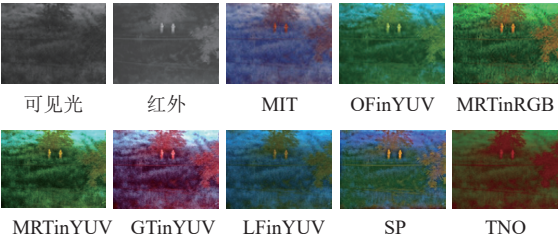


图 7 绿色植物场景中的 1 组源图像及 8 幅融合图像

由于数据库包含的融合图像较少, 在具体实验训练中, 随机在图像数据库选择 80% 的图像训练, 剩余 20% 的图像用于测试, 为了保证训练集和测试集的融合图像内容没有重复, 将数据库按照图像的场景进行分类, 多次重复实验, 并记录每一次实验图像质量预测值和人眼主观分数之间的斯皮尔曼秩相关系数 (Spearman rank order correlation coefficient, SROCC)、皮尔逊线性相关系数 (Pearson linear correlation coefficient, PLCC) 和均方根误差 (root mean square error, RMSE). 图 8 给出模型最终的测试效果图. 从图 8 中可以看出, 模型预测值与人眼评价结果拟合较好, 本文模型可以较好地反映人眼主观感知.

2.2 消融实验和性能比较

图 1 模型结构的各个组成部分对图像质量预测会产生不同的影响, 进行消融实验验证各自的不同影响程度, 并记录本文模型预测的图像色彩和谐性分数和

人眼主观分数的 SROCC、PLCC 和 RMSE 的均值. 如表 1 所示, 可以看出, 将视网膜机制、色彩信息波动和深度神经网络相结合可以获得理想的图像质量评价效果. 以往深度神经网络方法直接通过原始融合图像获取图像质量过程中会丢失很多图像信息, 比如色彩信息, 这使得模型训练无法取得最好的效果, 从而得到的图像质量分数会产生较大的偏差. 因此, 引入视网膜和视觉皮层的颜色编码机制或者融合图像色彩信息波动所获得的结果要明显优于仅使用深度神经网络的方法, 而相比于单独引入色彩信息, 将二者结合并且同时引入模型, 质量预测结果与人眼感受更加一致.

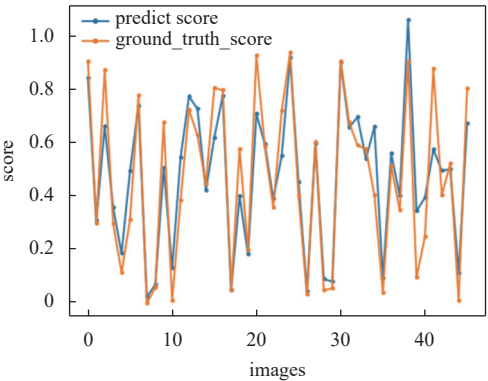


图 8 图像质量预测值和人眼主观分数的结果比较图

表 1 关于不同组成部分影响的消融实验

视网膜机制	色彩信息波动	深度神经网络	SROCC	PLCC	RMSE
—	—	√	0.802	0.821	0.178
—	√	√	0.842	0.854	0.186
√	—	√	0.850	0.866	0.181
√	√	√	0.903	0.898	0.142

将本文提出的 *DFICH* 模型与现有的 8 种图像质量评价方法进行性能对比实验, 其中包括 NRSL^[25], NCMQA^[5], MetaIQA^[26], HyperIQA^[27], DBCNN^[7], TReS^[8], MANIQA^[28]和 VCRNet^[29], 除了比较各个模型图像质量预测值和人眼主观分数之间的 SROCC, PLCC 和 RMSE, 还对模型预测时间性能 (平均时间 mean time, MT) 作了分析, 具体结果如表 2 所示.

从表 2 中可以看出, 本文方法与其他 8 种方法相比, 在预测时间上处于中等水平, 没有较明显的优势, 这是因为实验室过程中对输入图像的每个通道单独进行计算, 耗费了大量时间. 但是, 本文方法评价融合图像质量更准确, 质量预测值与人眼主观分数之间的 SROCC 可以达到 0.9 以上, 这表明了视觉特征更能准

确地描述融合图像质量, 评价结果与人眼感受具有更高的一致性。

表2 与现有8种图像质量评价方法的性能比较

方法	SROCC	PLCC	RMSE	MT
NRSL	0.729	0.630	0.194	2.583
NCMQA	0.699	0.726	0.218	1.917
MetalQA	0.893	0.890	0.154	0.125
HyperIQA	0.841	0.829	0.163	0.583
DBCNN	0.871	0.862	0.149	0.042
TRes	0.859	0.855	0.196	1.391
MANIQA	0.890	0.887	0.162	0.286
VCRNet	0.874	0.879	0.152	0.703
Ours	0.903	0.898	0.142	1.019

此外, 将模型在合成失真和真实失真数据集上进行对比实验. 和其他方法相比, 本文方法在TID2013、CSIQ、LIVE数据集上有一定优势, 结果如表3所示. 在TID2013和CSIQ数据集上, 本文方法有较大的优势. 虽然与文献[7]方法相比, 本文方法在LIVE数据集上还存在上升空间, 但整体上本文的方法是优于其他方法的。

表3 模型在不同失真数据库的比较结果

方法	TID2013		LIVE		CSIQ	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
NRSL	0.733	0.747	0.859	0.857	0.814	0.849
HOSA	0.814	0.849	0.843	0.845	0.774	0.702
DBCNN	0.838	0.896	0.912	0.903	0.883	0.868
HyperIQA	0.835	0.864	0.831	0.862	0.833	0.858
TIQA	0.826	0.792	0.781	0.805	0.821	0.849
Ours	0.901	0.897	0.882	0.895	0.909	0.912

3 结论与展望

本文提出的基于多视觉特征的双波段彩色融合图像色彩和谐性模型解决了现有方法中未充分利用视网膜和视觉皮层的颜色编码机制所带来的质量预测不准确的问题, 同时避免了在质量预测过程中缺失图像色彩信息的情况. 实验结果表明, 对于包含3种典型场景的彩色(夜视)融合图像数据库, 该方法的准确性和实用性较高, 与人眼主观感受评估更为一致. 由于该方法对融合图像每个通道进行处理, 运算量较大且耗时较长, 未来研究可以进一步探讨如何提高该模型的时间性能和推广场景应用范围. 同时如何进一步提高模型在自然图像领域的性能, 将是今后重点研究的内容之一。

参考文献

1 Luo YY, He KJ, Xu D, *et al.* Infrared and visible image

fusion based on visibility enhancement and hybrid multiscale decomposition. *Optik*, 2022, 258: 168914. [doi: [10.1016/j.ijleo.2022.168914](https://doi.org/10.1016/j.ijleo.2022.168914)]

2 Ma JY, Tang LF, Fan F, *et al.* SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(7): 1200–1217. [doi: [10.1109/JAS.2022.105686](https://doi.org/10.1109/JAS.2022.105686)]

3 Li CF, Guan TX, Zheng YH, *et al.* Blind image quality assessment in the contourlet domain. *Signal Processing: Image Communication*, 2021, 91: 116064. [doi: [10.1016/j.image.2020.116064](https://doi.org/10.1016/j.image.2020.116064)]

4 Zhang L, Zhang L, Bovik AC. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 2015, 24(8): 2579–2591. [doi: [10.1109/TIP.2015.2426416](https://doi.org/10.1109/TIP.2015.2426416)]

5 Zhou ZH, Lu W, Yang JC, *et al.* No-reference image quality assessment based on neighborhood co-occurrence matrix. *Signal Processing: Image Communication*, 2020, 81: 115680. [doi: [10.1016/j.image.2019.115680](https://doi.org/10.1016/j.image.2019.115680)]

6 Kang L, Ye P, Li Y, *et al.* Convolutional neural networks for no-reference image quality assessment. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 1733–1740.

7 Zhang WX, Ma KD, Yan J, *et al.* Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(1): 36–47. [doi: [10.1109/TCSVT.2018.2886771](https://doi.org/10.1109/TCSVT.2018.2886771)]

8 Golestaneh SA, Dadsetan S, Kitani KM. No-reference image quality assessment via transformers, relative ranking, and self-consistency. *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa: IEEE, 2022. 3989–3999.

9 Gu K, Tao DC, Qiao JF, *et al.* Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(4): 1301–1313. [doi: [10.1109/TNNLS.2017.2649101](https://doi.org/10.1109/TNNLS.2017.2649101)]

10 Ma S, Fan YY, Chen CW. Pose maker: A pose recommendation system for person in the landscape photographing. *Proceedings of the 22nd ACM International Conference on Multimedia*. Orlando: ACM, 2014. 1053–1056.

11 Lu P, Peng XJ, Yuan CX, *et al.* Image color harmony modeling through neighbored co-occurrence colors. *Neurocomputing*, 2016, 201: 82–91. [doi: [10.1016/j.neucom.2016.08.082](https://doi.org/10.1016/j.neucom.2016.08.082)]

- 2016.03.035]
- 12 Lu P, Yu JB, Peng XJ. Deep conditional color harmony model for image aesthetic assessment. Proceedings of the 24th International Conference on Pattern Recognition (ICPR). Beijing: IEEE, 2018. 2845–2850.
 - 13 Gu K, Zhai GT, Yang XK, *et al.* Using free energy principle for blind image quality assessment. IEEE Transactions on Multimedia, 2015, 17(1): 50–63. [doi: [10.1109/TMM.2014.2373812](https://doi.org/10.1109/TMM.2014.2373812)]
 - 14 Wu JJ, Lin WS, Shi GM, *et al.* Orientation selectivity based visual pattern for reduced-reference image quality assessment. Information Sciences, 2016, 351: 18–29. [doi: [10.1016/j.ins.2016.02.043](https://doi.org/10.1016/j.ins.2016.02.043)]
 - 15 Wu JJ, Zeng JC, Dong WS, *et al.* Blind image quality assessment with hierarchy: Degradation from local structure to deep semantics. Journal of Visual Communication and Image Representation, 2019, 58: 353–362. [doi: [10.1016/j.jvcir.2018.12.005](https://doi.org/10.1016/j.jvcir.2018.12.005)]
 - 16 Jin WQ, Jia XT, Gao SS, *et al.* Subjective evaluation of quality for color fusion images. Optics and Precision Engineering, 2015, 23(12): 3465–3471. [doi: [10.3788/OPE.20152312.3465](https://doi.org/10.3788/OPE.20152312.3465)]
 - 17 Kolb H. How the retina works. American Scientist, 2003, 91(1): 28–35. [doi: [10.1511/2003.11.28](https://doi.org/10.1511/2003.11.28)]
 - 18 Ruderman DL, Cronin TW, Chiao CC. Statistics of cone responses to natural images: Implications for visual coding. Journal of the Optical Society of America A, 1998, 15(8): 2036–2045. [doi: [10.1364/JOSAA.15.002036](https://doi.org/10.1364/JOSAA.15.002036)]
 - 19 Pridmore RW. A new transformation of cone responses to opponent color responses. Attention, Perception, & Psychophysics, 2021, 83(4): 1797–1803.
 - 20 Reinhard E, Adhikhmin M, Gooch B, *et al.* Color transfer between images. IEEE Computer Graphics and Applications, 2001, 21(5): 34–41.
 - 21 Lasmar NE, Stitou Y, Berthoumieu Y. Multiscale skewed heavy tailed model for texture analysis. Proceedings of the 16th IEEE International Conference on Image Processing. Cairo: IEEE, 2009. 2281–2284.
 - 22 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)]
 - 23 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778.
 - 24 Deng J, Dong W, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255.
 - 25 Li QH, Lin WS, Xu JT, *et al.* Blind image quality assessment using statistical structural and luminance features. IEEE Transactions on Multimedia, 2016, 18(12): 2457–2469. [doi: [10.1109/TMM.2016.2601028](https://doi.org/10.1109/TMM.2016.2601028)]
 - 26 Zhu HC, Li LD, Wu JJ, *et al.* MetaQA: Deep meta-learning for no-reference image quality assessment. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 14131–14140.
 - 27 Su SL, Yan QS, Zhu Y, *et al.* Blindly assess image quality in the wild guided by a self-adaptive hyper network. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 3664–3673.
 - 28 Yang SD, Wu TH, Shi SW, *et al.* MANIQA: Multi-dimension attention network for no-reference image quality assessment. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans: IEEE, 2022. 1190–1199.
 - 29 Pan ZQ, Yuan F, Lei JJ, *et al.* VCRNet: Visual compensation restoration network for no-reference image quality assessment. IEEE Transactions on Image Processing, 2022, 31: 1613–1627. [doi: [10.1109/TIP.2022.3144892](https://doi.org/10.1109/TIP.2022.3144892)]

(校对责编: 张重毅)