

DOI:10.19651/j.cnki.emt.2209925

综合利用肤色模型和感知损失的偏色去除^{*}

桂冰洁 贾振堂

(上海电力大学电子与信息工程学院 上海 201306)

摘要: 由于现有的颜色恒常性算法在非均匀光照、场景复杂的情况下对于偏色去除表现不好,本文提出了一种综合利用肤色模型和感知损失去除偏色的深度学习算法。该算法综合利用了肤色模型和感知损失,在计算过程中可以识别且重点关注肤色信息,更注重对图像语义的理解,而不是简单的进行像素间的计算。同时,将肤色模型与注意力机制结合,更突出了肤色区域的作用。实验结果表明,本文提出的颜色恒常计算方法能在语义层面较为准确地消除单一光照和多光照场景下图像的偏色,该算法与其他算法相比,能获得更良好的效果。

关键词: 颜色恒常性;图像处理;图像恢复;感知损失

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Color cast removal using skin color model and perceptual loss

Gui Bingjie Jia Zhentang

(College of Electrical and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: Since the existing color constancy algorithms do not perform well for color cast removal in the case of non-uniform illumination and complex scenes, this paper proposes a deep learning algorithm that comprehensively uses the skin color model and perceptual loss to remove color casts. The algorithm integrates the skin color model and perceptual loss, so that can recognize and focus on skin color information in the calculation process, and pay more attention to the understanding of image semantics, rather than simple calculation between pixels. At the same time, the skin color model is combined with the attention mechanism, which highlights the role of the skin color area. The experimental results show that the color constancy calculation method proposed in this paper can accurately eliminate the color cast of images in single-illumination and multi-illumination scenes at the semantic level. Compared with other algorithms, this algorithm can achieve better results.

Keywords: color constancy; image processing; image restoration; perceived loss

0 引言

人眼依靠感光细胞观察事物,光线反射到人类双眼经由一系列处理,在大脑中形成图像,以识别物体、认识世界。尽管物体的表色常常会由于受到光照还有物体表面反射性的影响,发生不同程度的变化,但是经大脑处理以后,人们依然能够感知到物体的本质色泽,具有实现颜色恒常的特性。

摄像机记录的图像取决于3个因素:场景的物理内容、入射到场景中的照明以及摄像机的特性。但是成像设备无法像人类一样实时感知外界环境改变造成的色彩变换从而自动校正偏色,而是通过传感器的感光特性对光谱进行响应,最终“看到”物体。因此,在保证物体识别的鲁棒性前提

下,排除光照变化等外界干扰因素,研究得到与人类视觉特性相符的颜色自动校正算法是计算颜色恒常性的本质。

近年来主要的颜色恒常性算法可分为无监督和有监督两大类。无监督的颜色恒常性计算方法是指不依赖于其他先验知识,图像成像时的场景光照颜色是凭借图像的底层颜色特征估计生成^[1],但图像在无约束条件下的光照估计并不能得到很好的结果。无监督颜色恒常性算法起源于Retinex理论^[2],该理论认为人眼看到的物体颜色 C 决定于光照 L 和物体本身的反射属性 R ,即 $C = L \cdot R$ 。基于该理论提出了Grey-World假设、Shade of Grey算法以及White-Patch算法。这些无监督颜色恒常性算法的优点是计算难度低且便于实现。

有监督的颜色恒常性算法是指在有标签的情况下,通

收稿日期:2022-05-12

^{*} 基金项目:国家自然科学基金(62105196)项目资助

通过对图像场景光源颜色分析训练得到相应的颜色分布,然后再根据训练学习得到的模型或者统计规律估计光照颜色。其中,依据学习训练和分类的方法不同,可分为基于统计学习和基于机器学习两大类光照估计方法^[3]。

基于统计的方法通过对场景中的光照颜色的值进行数理统计,从而预测场景中的光照颜色。该方法处理的数据大部分服从高斯分布且采用极大似然进行参数估计。常见的基于统计算法有色域映射算法^[4]、基于贝叶斯推理的颜色恒常性计算^[5]、基于高阶语义信息算法^[6]等。基于学习的方法是通过数据集进行训练学习得到对应的光照估计模型,由构造的模型对后续测试数据进行处理,从而消除图像色偏。常见的方法有:神经网络估计场景照明色度的颜色恒常计算方法^[7]、基于支持向量机的颜色恒常计算方法^[8]、基于深度学习的 Retinex 分解方法^[9]等。

在处理颜色恒常性问题上,基于学习的方法比基于统计的方法准确度更高、效果更好。但是,目前在该领域的研究仍存在缺陷和不足,一是真实场景分为单光照场景与多光照场景,多光照下的颜色恒常性计算要比单光照下复杂很多。迄今为止,大多数先进的颜色恒常性方法都假定场景中存在单一均匀的照明,另外一些经典的针对多光照的算法也只是将现有的颜色恒常性算法从局部扩展到全局,但这一假设违背了现实世界的客观场景。

因为在现实生活中,面临处理的大多是多光照场景下的问题。因此,多光照场景下的颜色恒常问题仍需更深入探索。二是现有的估计光照颜色算法通常忽视了高层的语义信息,将关注点放在低层的像素层面,描述图像每个像素点之间颜色及其变化,然而面对复杂多变的现实环境时,其适应性明显不足。三是现如今并没有稳定的颜色恒常性算法既可以针对面部区域又不会排除环境因素的影响。

针对目前国内外颜色恒常性算法中无法适应复杂环境、忽略语义信息等问题,本文提出了一种综合利用肤色模型和感知损失去除偏色的深度学习算法。该算法将注意力机制和高斯肤色模型结合,使其在计算过程中重点关注图像中某个部分,更注重对图像语义的理解,而不是简单的进行像素间的计算。实验结果表明,本文提出的颜色恒常计算方法能在语义层面较为准确地消除单一光照和多光照场景下图像的色偏,同时能在较短的时间内获得较好的性能。

1 改进的基于注意力机制的残差卷积网络结构

1.1 残差神经网络

由于深度学习的高速发展,卷积神经网络模型越来越深。这也带来了一些问题,在能够收敛的神经网络里,随着网络深度增加,网络的表现先是逐渐增加至饱和,然后迅速下降。不仅如此,过深的网络结构会导致梯度消失、梯度爆炸甚至是网络退化。于是,残差网络(ResNet)的诞生可以很好的解决上述问题。

残差网络是通过给卷积层增加一条“捷径连接”

(shortcut connection)的方式来提高信息的传播效率。在一个深度网络中,我们期望一个非线性单元(可以为一层或多层的卷积层) $F(x)$ 去逼近一个目标函数 $H(x)$ 。目标函数可分成两部分:恒等函数 x 和残差函数 $H(x)-x$ 。根据通用近似定理,一个由神经网络构成的非线性单元有足够的的能力来近似逼近原始目标函数或残差函数,但实际中更容易学习后者^[10]。因此,原来的优化问题可以转换为:让非线性单元 $F(x)$ 去近似残差函数 $H(x)-x$,那么实际的目标函数 $H(x)$ 可表示为 $F(x)+x$ 。

ResNet 是由若干残差模块构建的,如图 1 所示。左图是普通的神经网络,右图为 ResNet 的基础结构。

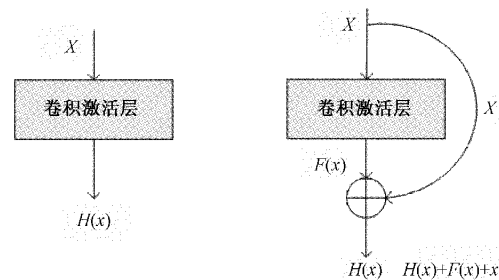


图 1 常规神经网络(左)残差网络基本单元(右)

从图 1(右)可以看出: x 为残差块的输入,通过卷积激活层得到 $F(x)$,然后 $F(x)$ 再与 x 叠加得到 $H(x)$ 。

1.2 注意力机制

近年来,随着 CNN 的兴起,很多结构新颖、有效的网络结构被提出,比如所熟知的是 VGG, ResNet, Inception, DenseNet 等模型,它们的效果已经被充分验证,而且被广泛应用在各类计算机视觉任务上。然而,经大量实验证明,可以通过在网络结构中引入注意力机制进一步提升模型的特征表达能力。注意力机制不止能让网络模型学会该注意什么,同时也能学会在哪个特定区域增强注意。

注意力机制的灵感来源于人类视觉,对于重要的信息分配更多权重,不那么重要的信息分配小一点的权重。人类视觉系统有“主次”这一概念,观察物体时更偏向于部分可以帮助判断的信息,并忽视相关性小的信息,所以才能轻松辨别事物本质。因此,引入的注意力机制模块能排除外界干扰因素,有效地提取到物体本质特征。

本文采用了 CBAM 模块^[11],这是一种轻量的注意力模块,该模块能够灵活的加入到目前的主流 CNN 网络结构且不会明显增加模型的计算量和参数量。其主要结构是在通道(Channel)和空间(Spatial)两个维度上引入 Attention,分别在这两个维度上强调学习“关注什么(what)”和“在哪儿关注(when)”。

本文深度学习模型主要采用轻量级的残差网络结构,以残差网络结构为基础,构建一种轻量级网络结构,不同于经典的 Resnet,该网络结构更简单,层数更少。如图 2 所示。

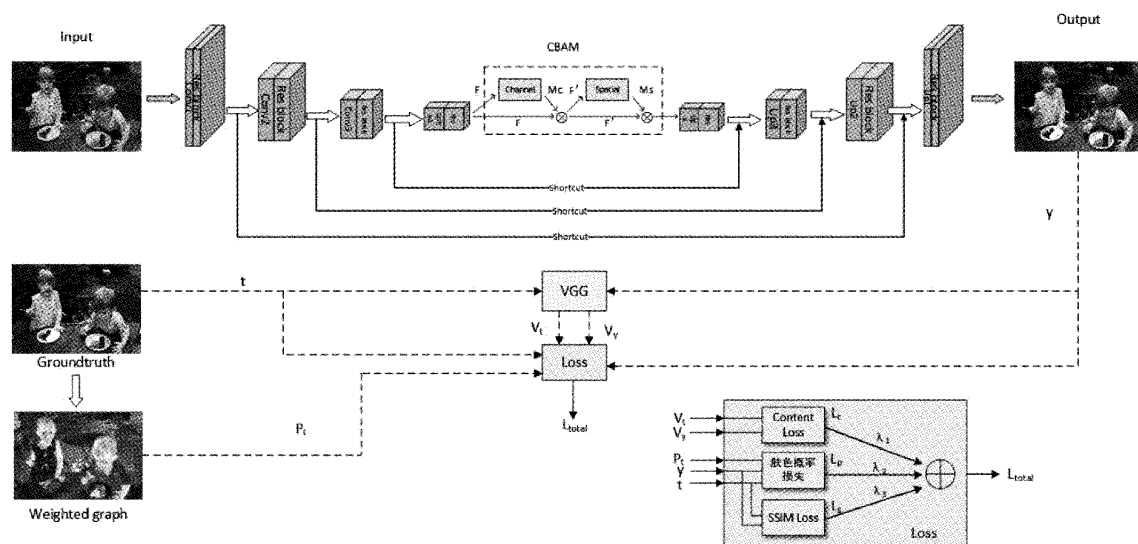


图 2 本文设计的轻量化残差网络结构图

改进的基于注意力机制的残差卷积网络不仅结构简单清晰,同时也具备着高效的提取特征能力,并且保证网络不损失大量精度的情况下降低网络参数和计算量。用 U-net 网络在训练集上的训练总时长为 9.7 h,而我们的网络的训练总时长为 4.8 h。两个网络分别训练得到的结果如图 5 所示,由此可见,我们的网络和 U-net 网络对比,效果更好,运行时间更快。

2 综合利用肤色模型和感知损失的偏色去除

2.1 颜色空间

在色彩学中,某一色彩可以用一维、二维、三维甚至四维空间坐标来表示。人们建立了多种可以用坐标系统能表达的彩色模型,即色彩空间^[12]。YCbCr 色彩空间主要应用于数字电视系统,它可以由 RGB 线性转换而得。该空间将亮度和色度分离,其中 Y 为颜色的亮度,Cb 和 Cr 分别表示蓝色和红色色度分量,这两个分量都可以相对独立表示图像的色度信息。

肤色是人的固有属性,面对光照变化仍趋于稳定。人脸肤色模型是基于色彩空间建立的一种高效识别人脸肤色信息的模型。经研究表明,肤色在 YCbCr 空间内聚类效果更好,更适合肤色检测。其中应用较为广泛的肤色模型有基于 YCbCr 色彩空间的简单阈值模型、高斯肤色模型及椭圆模型^[13]。本文选择基于 YCbCr 空间的高斯肤色模型,对其进行相似度计算,并将该模型与注意力机制原理结合得到肤色概率损失函数。

2.2 肤色概率损失函数

根据已建立的高斯肤色模型,肤色概率图是通过计算图像中所有像素点颜色与人体肤色之间的相似度得到,其取值范围为[0, 1],相似度计算公式如式(1)所示。

$$P(x) = \exp[-0.5(x - m)^T \sigma^{-1}(x - m)] \quad (1)$$

式中: m 表示肤色区域中心值, σ 表示方差, $x = [C_b, C_r]^T$ 为像素点在 $C_b C_r$ 空间中的向量。该式表示每个像素属于皮肤区域的概率,该值越大则属于肤色的可能性也越大,反之越小。

经过实验统计^[12],肤色高斯模型的均值 m 和方差 σ 分别为:

$$m = (156.559\ 9, 117.436\ 1)^T \quad (2)$$

$$\sigma = \begin{bmatrix} 299.457\ 4 & 12.143\ 0 \\ 12.143\ 0 & 160.130\ 1 \end{bmatrix} \quad (3)$$

在以前的研究中,也有针对面部区域的颜色恒常性算法,但是他们都是检测出面部区域之后,丢弃掉不属于肤色的那一部分,只对人脸区域进行检测。由于肤色倾向于在颜色空间中形成聚类,检测面部皮肤像素色域和皮肤标准色域之间的差异可以用于估计场景光源,通过局部光照差异估计出全局光照。但却忽视了一个至关重要的问题,实际的光照并不是均匀单一的,这种方法只存在于假定单一光照的情况。

本文设计的肤色概率损失可以通过肤色模型检测出肤色区域,给这部分区域分配更大的权重,使我们的网络在提取特征时更关注肤色特征,如式(4)。根据肤色高斯模型算出每个像素点属于肤色的概率,从而得到一张肤色概率的权重图,此权重图的每个像素点都与原图像素点相对应,将生成的权重图和待处理图像作面向像素的乘法,再将其与 L2 损失函数结合,达到了对人脸肤色区域的重点处理。这样既没有忽视周围环境因素,同时也达到了重点对人脸进行检测处理的效果,这种方法同样适用于多光照场景。

$$L_{skin} = \|(I'_i - I_h) \odot P\|_2^2 \quad (4)$$

其中, \odot 是面向像素(pixel-wise)的乘积, $\|\cdot\|_2^2$ 表示 L2 范数, $\|\cdot\|_2^2$ 表示输入偏色图像, I'_i 表示模型的输出

图像, I_h 表示地面真实。

2.3 感知损失

由 Retinex 理论得知, 一张由人眼观察到的图像实质上分为光照和反射两部分, 反射量是反映物体固有属性不变的量。由此可知, 处于偏色光照下的物体和白光下的物体的本质都是一样的, 于是本文模型的核心思想便是尽量去控制本质(反射量)不变, 去除变量(光照)的干扰便可达到去除偏色的效果。但是现有的颜色恒常性算法都是假设光照是均匀且单一, 即使存在多种光照也将其平均为一种光照^[4]。但是这种假设违背了现实世界的客观事实, 现实生活中主要面临的是多光照且非均匀的场景。

为了尽量克服这个局限性, 希望整个网络在学习的过程中更加关注重建图片和原始图片的语义特征差异, 而不是逐个像素之间的颜色亮度差异。由于人眼的视觉特性, 人眼对某一场景区域的感知并不是所有点都是一样的, 而是对处于视觉中心位置的感觉更明显^[14]。于是本文参考文献[15]的感知损失函数, 并对其改进得到了一个新的感知损失函数, 如式(5), 该损失函数可以很好的解决关于感知相关语义特征的问题。

将改进的感知损失表述为偏色图与地面真实之间内容损失、两者之间结构相似性损失之和:

$$L^{pe} = L_{content} + L_{SSIM} \quad (5)$$

式(5)中 $L_{content}$ 是指 I_h 和 I'_l 之间的内容损失, L_{SSIM} 是结构相似性损失函数, $L_{SSIM} = 1 - SSIM(I'_l, I_h)$, $SSIM(I'_l, I_h)$ 是结构相似性度量, I_l 表示输入偏色图像, I'_l 表示模型输出图, I_h 表示地面真实。

结构相似性(SSIM)是衡量两幅图像相似度的指标, SSIM 指数一般是通过分离照度对物体的影响来得到图像中的结构信息^[16]。所以我们用 SSIM 函数约束白光和偏光下物体的反射率, 将两者规范化为无限接近。

计算重建图像和原始图像差异的时候是直接在像素图像上用 MSE 准则进行比较, 但是这种方式只会过度的让模型去学习这些像素差异, 而忽略了原本图像物体的固有特征, 导致感知质量较差^[15]。因此, 实际的差异计算不仅只应该关注每个像素间的差异, 更应该关注特征差异。

本文受此启发设计了一个“内容损失”函数, 如式(6)。通过 VGG19 网络^[17]提取特征图, 并计算他们欧式距离, 如此可以实现在语义层面上进行两幅图像的计算, 而不是在低级像素层面的误差计算。

$$L_{content} = \|VGG(I'_l) - VGG(I_h)\|_2^2 \quad (6)$$

其中, $VGG(\cdot)$ 表示通过截取的 VGG19 模型提取得到的特征图, I_l 表示输入偏色图像, I'_l 表示本文模型的输出图像, I_h 表示地面真实。

3 实 验

3.1 实验数据集

本次实验用的数据集为 The Gallagher Collection Person Dataset; 是 Andrew Gallagher 的家庭相册。该数据集内容很接近实际应用的场景。

本文针对该数据集做了增强处理, 用不同程度的色温进行渲染, 使 589 张真实场景下的图像失去白平衡, 从而模拟在不同程度颜色光照下的场景。最终一生成 7 068 张非白平衡场景下的图像, 其中 2 356 张为多光照非白平衡图像。选取 6 200 张用于训练集, 868 张用于测试集, 并且每个训练图像都有一个与之相关联的地面真实图像。

3.2 实验结果及数据分析

采用主观评价和客观评价相结合的方法对测试结果进行了评价, 主观评价是通过人眼观察图像并真实反馈视觉感知特征, 客观评价则是利用角度误差这一指标来衡量, 这是颜色恒常性问题常用的评价指标。

将本文模型与 3 种经典的传统算法(Gray World、white patch 和动态阈值法)以及 3 种先进的深度学习算法(U-net、Retinex-net^[9]和 Kind-net^[18])在测试集上分别进行数据分析。与三种无监督传统算法对比的结果如图 3 和 4 所示, 其中(a)列为假定多光照下的输入偏色图像, 将本文计算方法(e)和 Grayworld 算法(b)、动态阈值算法(c)及 white patch 算法(d)进行对比, 图像右下角数字代表与地面真实间的角度误差值, 前 3 行为室外场景和后 3 行为室内场景。综合图 3 和 4 可以看出传统算法虽然在处理单一光照图像有一定的效果, 但处理多光照场景时有明显缺陷。由此可见, 本文提出的方法针对任意场景和光照图像都可以良好准确的去除色偏恢复白平衡。在单一光照和多光照场景下本文算法与 3 种有监督算法对比的结果如图 5 所示, 其中前两行为单光照, 后两行为多光照场景, (a)列为输入的偏色图, 将本文的网络模型(e)和 U-net 网络(b)、Kind-net(c)以及 Retinex-net(d)进行比较, 可以看出本文的算法比起别的深度学习算法, 也具有明显优势。

无监督和有监督颜色恒常性算法在不同场景下的角度误差对比结果分别如图 6 和 7 所示。结果表明, 在混合场景下, 本文提出方法明显取得更好的效果, 且该算法对于室外场景比室内场景效果更好。

除此, 通过客观评价指标对测试集数据进行量化分析, 将 3 种传统的无监督算法(Gray World、white patch 和动态阈值)和 3 种基于学习的有监督算法(U-net、Retinex-net^[9]和 Kind-net^[18])与本文方法在测试集上进行测试, 得到的结果分别与 groundtruth 进行角度误差计算, 并得出了他们的最小值、最大值及平均值。结果如表 1 所示, 从数字上可以看到本文的方法明显优于所有其他方法。

综上所述, 本文提出的方法在任意光照和任意场景

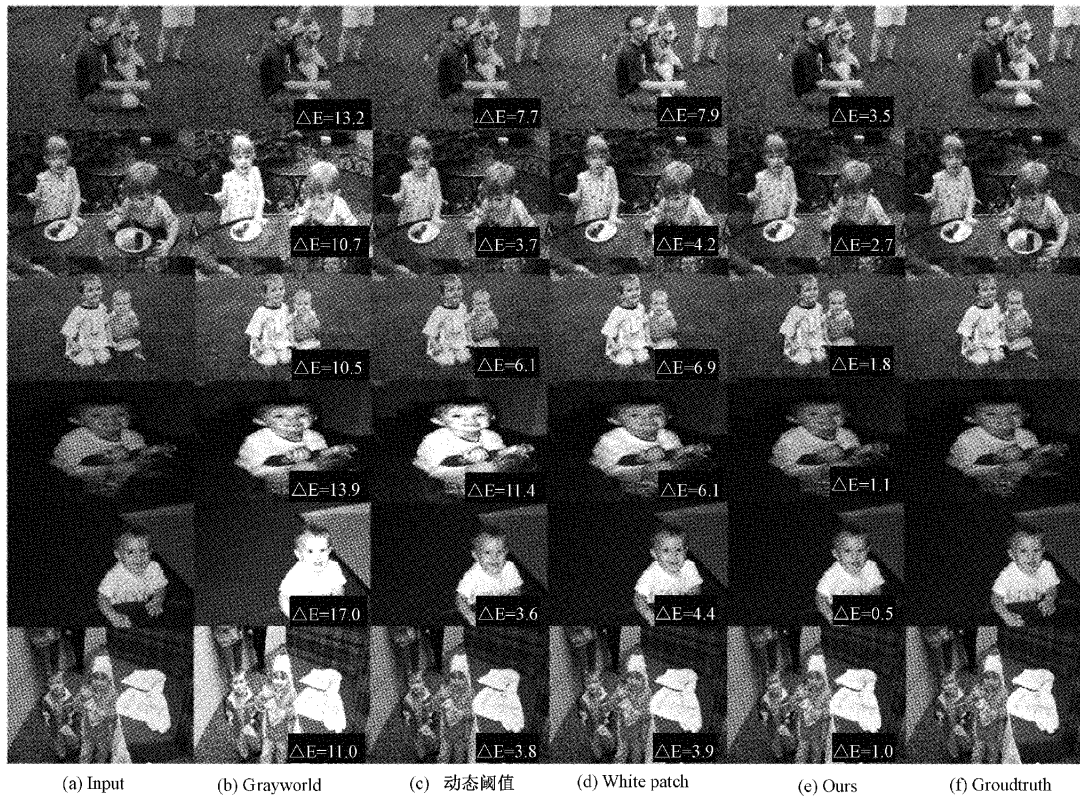


图 3 单一光照下各种无监督颜色恒常计算方法和本文方法的结果对比



图 4 假定多光照下无监督颜色恒常计算方法和本文方法的结果对比

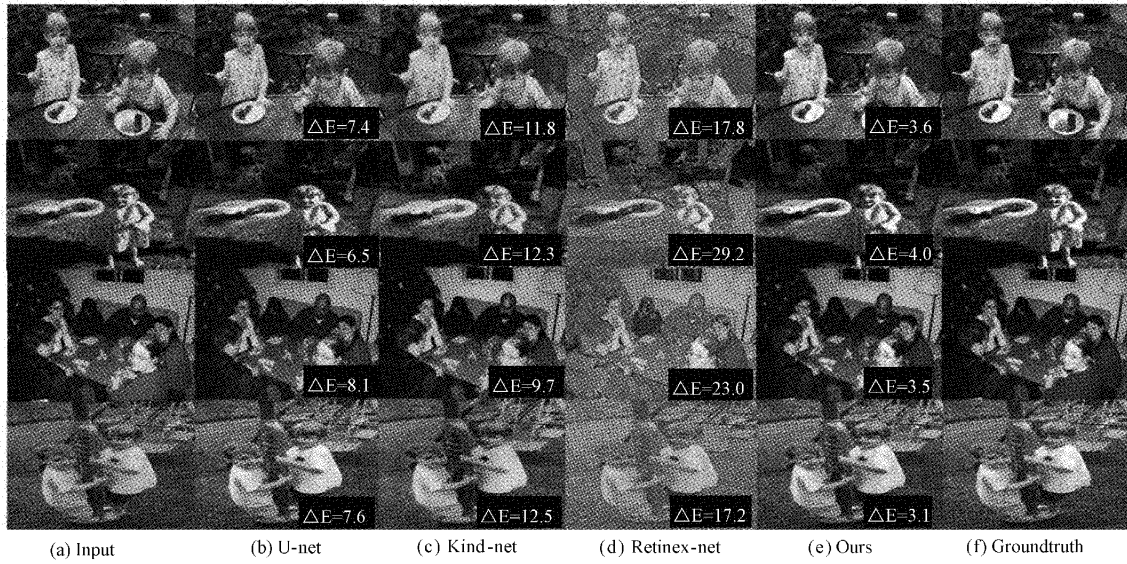


图 5 混合场景下有监督颜色恒常计算方法和本文方法的结果对比

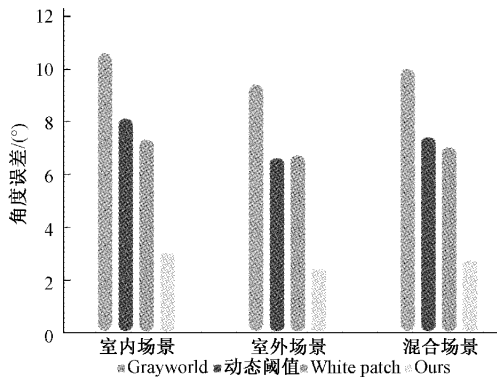


图 6 无监督颜色恒常性算法在不同场景下的角度误差对比

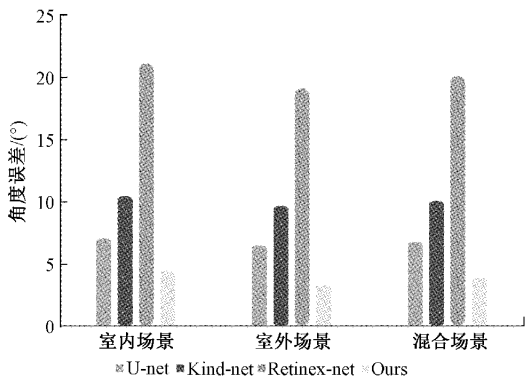


图 7 有监督颜色恒常性算法在不同场景下的角度误差对比

下,都能更大程度的去除偏色,结果也更接近地面真实。

表 1 各种方法在数据集上角度误差的结果

算法	平均值	最小值	最大值
无监督算法			
Grayworld	11.52	2.41	39.93
动态阈值	6.32	2.52	13.32
White patch	7.02	3.72	13.57
Kind-Net	10.06	5.80	15.39
有监督算法			
Retinex-Net	21.35	8.85	38.84
U-Net	7.67	3.14	19.63
本研究算法	5.25	1.28	11.02

4 结 论

本文提出了一种综合利用肤色模型和感知损失去除偏色的深度学习算法。该方法首次将残差网络和高斯肤色模型结合,设计了肤色概率损失函数。除此,引入了注意力机制模块,利于我们更加准确地提取到关于肤色的特征信息。同时我们采用感知损失函数,将其上升到了语义层面,使我们的模型能够高效准确地处理单光照和多光照场景下的图像。

将本文模型在公开的数据集 The Gallagher Collection Person Dataset 上,从主观和客观两个角度进行结果测试,并与 Gray-World、White-Patch 和动态阈值三种无监督算法和 Kind、Retinex-Net 两种基于学习的有监督算法进行性能比较,结果表明本文提出的方法能够更有效地去除偏色并且侧重关注肤色信息。

参考文献

- [1] 孙馨宇, 解凯, 何翊卿, 等. 颜色恒常性计算综述[J]. 北京印刷学院学报, 2020, 28(6):5.
- [2] 罗圣敏. 基于 Retinex 理论的图像增强算法研究[J].

- 电脑知识与技术, 2017(10):3.
- [3] 唐正, 刘宏哲, 袁家政. 单一光照颜色恒常性计算研究进展[J]. 计算机科学, 2016, 43(11):7.
- [4] FORSYTH D A. A novel algorithm for color constancy [J]. *International Journal of Computer Vision*, 1990, 5(1):5-35.
- [5] ROSENBERG C R, MINKA T P, LADSARIYA A. Bayesian color constancy with non-gaussian models [C]. *Advances in Neural Information Processing Systems*, 2017.
- [6] 徐亮山, 张霞, 范珮. 基于对象语义分割的改进颜色恒常性算法[J]. 数字印刷, 2022(2):8.
- [7] HU Y, WANG B, LIN S. FC4: Fully convolutional color constancy with confidence-weighted pooling [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
- [8] 程奥运. 基于视觉显著性的视频偏色检测的应用研究[D]. 武汉:武汉科技大学, 2017.
- [9] WEI C, WANG W, YANG W, et al. Deep retinex decomposition for low-light enhancement [J]. *ArXiv Preprint*, 2018, ArXiv:1808.04560.
- [10] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3-19.
- [12] 周利琴, 谷林. 基于高斯肤色模型的人脸区域及下巴检测[J]. 西安工程大学学报, 2015, 29(6):5.
- [13] 邹奉天. 几种皮肤检测算法的仿真[J]. 计算机与数字工程, 49(6):4.
- [14] 傅剑峰, 汪荣贵, 张新龙, 等. 基于人眼视觉特性的Retinex算法研究[J]. 电子测量与仪器学报, 2011, 25(1):29-37.
- [15] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C]. *European Conference on Computer Vision*. Springer, Cham, 2016: 694-711.
- [16] 朱新山, 姚思如, 孙彪, 等. 图像质量评价:融合视觉特性与结构相似性指标[J]. 哈尔滨工业大学学报, 2018, 50(5):8.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *ArXiv Preprint*, 2014, ArXiv:1409.1556.
- [18] ZHANG Y, ZHANG J, GUO X. Kindling the darkness: A practical low-light image enhancer [C]. *Proceedings of the 27th ACM International Conference on Multimedia*, 2019: 1632-1640.

作者简介

桂冰洁, 硕士, 主要研究方向为图像处理。

E-mail: 237363979@qq.com

贾振堂, 副教授, 硕士生导师, 主要研究方向为智能视频监控(涉及多模态深度学习、目标识别、人体姿态分析、立体视觉等)。

E-mail: 462458081@qq.com