

基于多层次注意力融合的图像美学重定向

于明, 张吉俊, 郭迎春*, 张萌, 王丹

河北工业大学人工智能与数据科学学院, 天津 300401

摘要 随着 5G 的发展和多种显示终端的出现, 图像重定向算法倍受关注。目前大多数的算法没有考虑到重定向后图像的美学分布, 因而影响了人类的视觉美学感知。鉴于此现状, 提出了一种基于多层次注意力融合的图像美学评价网络, 通过对不同细粒度的特征的提取, 并根据注意力机制自适应地融合后获得美学信息, 将学习到的美学信息与图像的显著图、梯度图和直线特征图融合作为具有美学分布的重要度图, 以此来指导多操作的图像重定向算法。通过实验验证了生成的重要度图能够很好地保护图像重定向中的美学信息, 图像重定向结果相较于主流方法具有更好的视觉感知效果。

关键词 图像处理; 图像重定向; 美学评价; 注意力机制; 特征融合

中图分类号 TP3

文献标志码 A

doi: 10.3788/LOP202158.2410002

Image Aesthetics Retargeting Algorithm Based on Multi-Level Attention Fusion

Yu Ming, Zhang Jijun, Guo Yingchun*, Zhang Meng, Wang Dan

School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

Abstract With the development of 5G and the emergence of multiple display terminals, image retargeting algorithms have received extensive attention. Most existing algorithms do not consider the aesthetic distribution of the image during retargeting, thus affecting the human visual aesthetic perception. In view of this situation, we propose an image aesthetic evaluation network based on multi-level attention fusion. The aesthetic information is obtained by extracting different fine-grained features and adaptively fusing them according to the attention mechanism. Then, the learned aesthetic information is combined with the saliency map, gradient map, and linear feature map of the image as the importance map to guide the multi-operation image retargeting algorithm. Experimental results show that the generated importance maps can well protect aesthetic information, and the obtained retargeting images has a better visual perception than the state-of-the-art methods.

Key words image processing; image retargeting; aesthetic evaluation; attention mechanism; feature fusion

OCIS codes 100.2960;110.3000

1 引言

随着 5G 和数字视频技术的飞速发展, 不同分辨率及尺寸比例的数字显示设备的种类越来越多, 例如各种不同尺寸的手机、平板电脑、笔记本电脑和扩展显示器, 在人们的生活中占据了重要地位。由于不同显示设备在生活中的应用场景不同、显示的纵横比也各不相同, 当一张固定分辨率的图像在不

同电子设备上显示时, 就会出现显示比例不匹配的问题, 因而如何提高显示设备的利用率和人眼的视觉感知效果成为一个亟待解决的问题。

图像重定向旨在改变同一张图像的纵横比的同时保留完整的图像主体信息以及图像的视觉美感。根据图像重定向算法的技术特点, 大致可以分为 4 类(基于离散的接缝裁剪方法^[1-4]、基于连续的图像变形方法^[5-7]、多操作方法^[8-11]和基于深度学习的方

收稿日期: 2020-12-17; 修回日期: 2021-02-07; 录用日期: 2021-02-12

通信作者: *gyc@scse.hebut.edu.cn

法^[12-16])。这些方法的主要目标一是改变输入图像的纵横比的同时保持主体区域特征不变,二是背景区域不产生扭曲,图像整体结构不发生变化。这两个目标是两个互相矛盾的质量度量,当图像主体较大,或者图像背景区域结构复杂时,经常会出现由于过量的变形,导致重定向后的图像不能满足人类的视觉美学感知。

针对以上问题,本文提出了一种基于注意力机制的图像美学重定向方法,构建了美学评价网络,以提取多层级的深度特征,并基于注意力机制自适应融合来获得符合人类美学感知的美学特征,可以准确地预测图像的美学评分。同时,提取出的美学特征与图像的显著图(E_{sal})、梯度和直线信息融合作为重要度图,引导图像的重定向,使图像在满足目标长宽比的同时还能保留较高的视觉美感。

2 相关工作

图像重定向的基本思路是根据图像中各个区域像素的重要度的不同对它们进行不同的操作,以尽可能地保留具有高重要度的区域,而具有低重要度的区域在不引入严重视觉伪影的前提下进行删除或复制的操作,并最终满足目标尺寸要求。随着内容感知图像重定向方法的不断研究发展,重定向算法的性能有了极大的提升,其主要研究可以分为两类,第一类是基于重要度图的重定向方法,第二类是结合图像美学评价的重定向方法。

2.1 基于重要度图的重定向方法

Avidan 和 Shamir^[1]在 2007 年的 SIGGRAPH 会议上首次提出了基于内容感知图像重定向算法,该算法是一种离散的重定向方法,将图像的梯度图作为重要度图,利用动态规划算法计算累积重要度值,通过移除或复制累积能量和最小的八连接接缝来进行裁剪。Mansfield 等^[2]提出了一种关于主体保护的缝裁剪改进算法,将缝裁剪算法的主体位置像素重要度设置为无穷大,所以没有任何接缝可以穿过该主体区域,但是当主体过大时会发生明显的变形。为了更精确地保护图像的显著对象,吴加莹等^[3]提出融合显著和深度信息构建图像重要度图,但该方法对于背景复杂的区域保护不佳。针对背景失真和面积较小的显著物体被当作背景裁剪掉的情况,郭迎春等^[4]提出一种基于分块的线裁剪算法,根据显著图的平均累加能量将图像分为保护区域和非保护区域,再根据不同区域的面积来分配裁剪线的数量,该方法降低了小面积显著主体被裁剪掉的可

能性,但是在显著性算法检测效果不好的图像中,主体区域会发生严重变形。

连续型图像重定向算法以变形算法为主^[5-7],相比于接缝裁剪,变形算法能够有效地减少图像的锯齿型变形,避免了接缝裁剪过程中容易产生的结构断裂的问题,即图像变形算法对图像的直线特征保护较差,同时算法的时间复杂度很高,算法效率低。

现有基于重要度的重定向方法大多采用多操作算法。Rubinstein 等^[8]首次提出多操作算法的概念,统计了用户对于各种图像重定向操作、结合多操作以及单一方法选择的结果,得到了多操作图像重定向算法一般优于单一操作的结论。该类方法的实质是权衡结构信息破坏、主体对象变形、产生人工噪声、图像内容丢失等对图像整体效果的影响程度,然后折中选择,此算法的主要难点在于确定最佳的操作顺序。Cui 等^[9]提出一种图像变形检测算法,通过检测每个接缝去除后图像变形的概率分布选择操作算子。Zhou 等^[10]提出将多操作算子公式化为马尔可夫决策过程,利用强化学习在重定向过程中自适应地选择操作算子。Hwang 等^[11]提出了一种基于视觉模型的图像重定向方法,首先对图像进行线裁剪,当最小能量接缝大于一定的阈值时即采用均匀缩放方法将调整到目标尺寸,但该方法采用的是单个接缝的能量而不是图像全局特征作为终止条件,算法的稳定性和鲁棒性不能被保证。

随着深度学习的发展,研究人员将卷积神经网络(CNN)^[17]应用于图像重定向方向^[12-16],端到端地学习原始图像到重定向图像之间的映射,并定义不同的损失函数来控制网络的反向传播,不断优化以生成重定向目标图像。

2.2 结合美学评价的重定向方法

图像美学评价方法旨在通过学习人类感知和摄影规则来获得图像的美学分类信息。初期的美学评价方法往往通过提取图像美学方面的低级特征来进行美学预测。Datta 等^[18]最先将图像的美学理解转化为二元分类问题。Datta 等结合了通常用于图像检索的低级和高级特征,并训练支持向量机(SVM)分类器在美学质量方面对图像进行二进制分类。该方法在描述图像的美学特征方面有较好的效果,但由于不同的人对同一张图片的美学感知是不一样的,所以通过计算来模拟人类美学感知仍是图像美学评价的一项重大挑战。随着新提出来的大规模美学数据集 AVA^[19]统计了大量的图像美感评分,许多基于深度学习的美学评价方法^[20-22]开始提取图

像中的深度特征,进行符合大多数人美学感知的图像美学预测。Kao 等^[20]提出一个多任务框架,通过两个网络共享参数,同时训练美学预测和美学分类两个任务。Zhang 等^[21]基于图像的语义标签将图像基于语义进行美学质量的评估。但是由于这些方法的输入图像都限制了大小,因此,图像在预处理过程中就不可避免地损失了很多美学信息。针对此问题,Hosu 等^[22]提出一种图像输入大小可变的图像美学识别方法,可以针对任意大小的高质量图像进行精确地预测。

结合美学评价的图像重定向的方法大多应用于图像的剪切,Kao 等^[23]提出了一种基于美学图和梯度能量图的自动图像裁剪技术。该技术主要使用基于滑动窗口的方法获得候选窗口,窗口大小受限且费时。为了克服滑动窗口裁剪区域固定与耗时长的问题,Li 等^[24]提出了 A2-RL 模型,该模型可以快速完成裁剪过程,得到任意形状的裁剪结果。Wang 等^[25]采用深度学习的方法,利用注意力预测和美学评估,找到一个裁剪窗口,使输入图像的重要部分尽可能保留下来。然而,仅使用裁剪技术实现的重定向效果是非常有限的,当目标尺寸与原始图像尺寸相差较大时,往往会破坏或丢失主体对象。

3 本文算法

受 Hosu 等^[22]启发,本文提出将图像不同细粒度的深度特征进行自适应融合,以学习符合人类美学感知的图像美学特征,通过全局平均池化和全连接生成图像的美学评价来预测模型。此外,根据所提取的多层级美学特征生成美学重要度图(E_{aes}),并和原始 E_{sal} 、梯度图和直线检测图加权融合,共同指导图像的重定向操作,使重定向结果较好地保留了美学信息。图 1 是本文提出的整体流程框架,编码器充当多层级特征提取器,将输入图像转化为多层次的美学特征表示,而基于注意力的美学评价网络从这些美学特征中进行准确的美学评价预测。获取到的美学注意力图通过上采样生成美学图,并与显著图和梯度直线图共同指导图像的重定向。多层级特征提取模块从编码器网络中提取不同尺度的深度特征,通过全局池化并连接生成 $a \times a \times d$ 大小的特征堆栈,经过注意力模块加权融合后,由全局平均池化和全连接层获得图像的美学预测。生成的注意力特征图经过上采样生成美学图,与显著性图和梯度直线图融合共同指导多操作图像重定向方法。

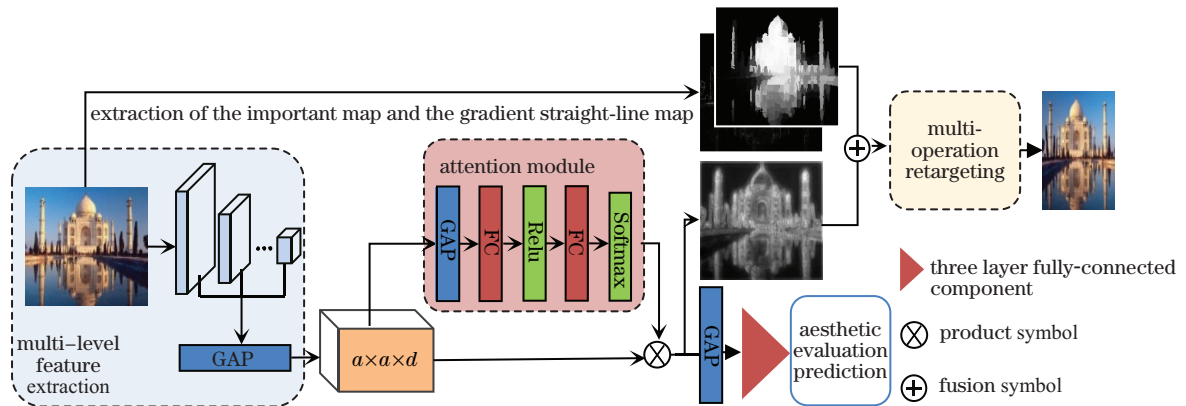


图 1 基于注意力融合的图像美学重定向主体框架。

Fig. 1 Flowchart of image aesthetic retargeting based on attention fusion

3.1 美学评价网络

本文提出的美学评价网络由两部分组成,第一部分是基于编码器网络的多层级特征提取模块,用于提取图像中不同细粒度的美学特征。第二部分是基于注意力机制的美学评价预测模块,对提取到的美学特征基于注意力机制自适应地融合后进行美学评价预测。

3.1.1 多层级特征提取

多层级美学特征提取编码器是基于预训练的

Inception ResNet-v2 网络。经过 ImageNet 数据集预训练的 Inception ResNet-v2 网络在分类和分割任务中具有卓越的性能,可以生成精确丰富的语义信息,因此通过获取 Inception ResNet-v2 网络的前 m 层输出来生成不同层级的美学特征信息,然后统一经过池化操作生成固定大小的美学特征图。所描述的过程可以表示为

$$C_{map}^{a \times a \times d} = \bigcup_{i=1}^m P^{a \times a \times d_i} [F_i(I; \omega_i)], \quad (1)$$

式中: \cup 表示特征图的连接操作; I 和 C_{map} 分别为输入图像和输出美学特征; $F_i(\ast)$ 表示编码器网络第 i 层的卷积操作; P 函数表示全局平均池化; m 是编码器网络的层数; 输出的美学特征大小为 $a \times a \times d$, 其中 a 是特征堆栈的大小, $d = \sum_{i=1}^m d_i$ 是通道数; ω_i 和 d_i 分别表示第 i 层卷积的权重参数和通道数。此特征堆栈具有多层次语义信息。值得注意的是, 特征大小由池化操作实施, 对输入图像的大小没有固定限制。

3.1.2 基于注意力机制美学评价预测

现有的美学评价模型^[20-22]大多在提取图像的深度特征后训练美学 SVM 分类器或全连接用于美学分类和排名, 虽然编码器的最深层具有丰富的高级特征, 但仅仅依靠高级特征进行图像美学预测会丢失具有更高细粒度的浅层特征信息。因此, 多层次特征编码器提取的特征可以有效地结合全局和局部美学信息。然而, 如何分配不同尺度的美学特征对预测准确性的权重仍然是一个重要的研究内容。本文提出的基于注意力机制的美学预测模型通过自适应的注意力模型将提取的美学特征按其美学贡献能力不同, 分配不同的权重进行预测。如图 1 所示, 注意力模型由一个全局平均池化层、两个全连接层组成和两个激活层组成。大小为 $a \times a \times d$ 的特征堆栈 C_{map} 经全局平均池化生成长度仍为 d 的特征向量, 然后经过两层全连接学习通道注意力的特征权重 (Q), 其中设置的两层激活层用于加入非线性

因素。最后特征堆栈 C_{map} 与 Q 相乘获得美学注意力图 (C_{attn}), 表示为

$$C_{\text{attn}} = C_{\text{map}} \times Q. \quad (2)$$

所获得的美学注意力图通过全局平均池化降维后与三层全连接块相连, 直接预测得出图像的美学评分。使用均方误差 (MSE) 损失 (L) 来优化整个网络, 其定义为

$$L = \sum_{i=1}^n \frac{1}{n} [f(I_i) - y_i]^2, \quad (3)$$

式中: y_i 表示图像 (I_i) 的平均标注得分; f 函数表示美学评价预测函数; n 为样本数量。此外, 为了减少网络的过拟合, 在每一层全连接之后加入了 Dropout 来随机地失活网络中隐藏的神经元。

3.2 融合美学特征的重要度图

通过美学评价预测网络获得的美学注意力图带有丰富的美学特征, 经过上采样将其重构为原图大小获得 E_{acs} , 如图 2(b) 所示。所获得的 E_{acs} 具有丰富的美学特征信息, 包含了主体和背景区域的美学特征。但是, 单独使用 E_{acs} 对图像中的显著物体缺乏良好的保护, 同时对直线以及物体的边缘刻画不明显, 因此需要与 E_{sal} 、梯度和直线检测图 (E_{grad}) 融合, 共同指导图像的重定向。重要度图为

$$E = E_{\text{acs}} + \lambda_1 E_{\text{sal}} + \lambda_2 E_{\text{grad}}, \quad (4)$$

式中: λ_1 和 λ_2 为超参数, 用来平衡 E 的融合权重, 在本文中, $\lambda_1 = 0.5$ 、 $\lambda_2 = 1.0$ 时达到最佳效果。重要度图融合效果如图 2(e) 所示。

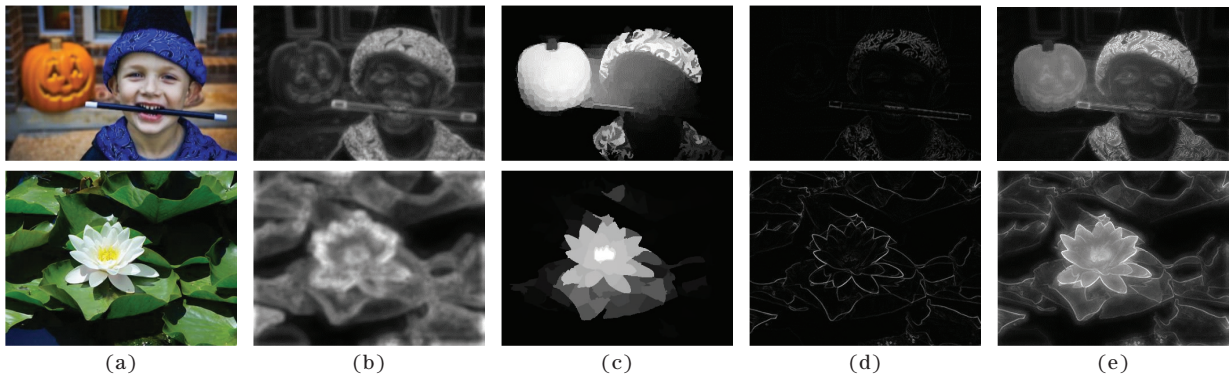


图 2 重要度图的融合。(a) 原始图像; (b) 美学特征图; (c) E_{sal} ; (d) E_{grad} 融合结果; (e) 重要度图

Fig. 2 Fusion of importance maps. (a) Original images; (b) aesthetic feature maps; (c) E_{sal} ; (d) fusions of E_{grad} ; (e) importance maps

3.2.1 显著图

图像的视觉显著性检测可以模拟人类的视觉感知, 获取图像中的主体对象。将视觉显著图作为图像重定向中的重要度图用于保护主体对象不发生变形。最初的视觉显著性模型是 Itti 等^[26]提出的以

生物认知体系为指导的视觉关注度计算模型。随后 Schölkopf 等^[27]提出了基于图论的视觉显著度模型 (GBVS), 引入了马尔可夫链来改进 Itti 等的模型, 能够更为准确地识别出图像中的主体对象, 但该模型同样忽略了图像中的结构信息。越来越多的算法

采用背景优先或边界优先原则,其中基于背景先验的显著性检测算法(RBD)^[28]使用边界连通度的概念来估计背景概率,然后对局部对比度进行加权,能准确地检测出图像中的显著目标,算法性能得到较大提升。从图 2(c)中可以看到,生成的显著图与人类视觉感知相一致。但是,单独使用显著图对物体的边界信息分割不明显,因此, E_{grad} 的融合仍十分必要。

3.2.2 梯度图

Avidan 等^[1]提出用梯度图作图像重定向的重要度图,图像中梯度大的像素将被赋予高重要度值,梯度小的像素被赋予低重要度值,因此, E_{grad} 可以很好地描述图像中的边缘信息梯度图。计算公式为

$$E_{grad} = |\partial I / \partial x| + |\partial I / \partial y|, \quad (5)$$

式中: I 为图像; x 、 y 分别水平和垂直方向。梯度图能够反映图像的边缘特征,同时由于人类视觉对具有直线性质的物体扭曲失真的感知敏感,本文算法将图像中的直线检测融合到梯度图中来描述图像中的重要区域,如图 2(d)所示。

3.2.3 重要度图融合

根据(4)式将美学特征图、显著图、 E_{grad} 融合而形成重要度图如图 2(e)所示,显著图可以将图像中的主体区域提取出来, E_{grad} 能够很好地反映图像的结构边缘信息,由美学注意力图生成的 E_{aes} 可以完整地突出图像中的高美学区域。所获得的重要度图既可以保护图像中的结构信息,又可以保留图像中的美学特征。

3.3 美学特征指导图像多操作重定向

为了获得更好的重定向结果,本文使用结合线裁剪算法和均匀缩放算法^[29]的多操作方法进行图像重定向。线裁剪算法可以在一定范围内很好地保护图像的主体区域,但是当图像主体对象过大,或者重定向比例过大时,线裁剪算法往往会破坏图像的

结构完整性,因此,当线裁剪进行到一定程度后,使用均匀缩放算法继续完成后续重定向操作是最佳选择。

线裁剪算法在重要度图的指导下迭代地移除或复制最低能量接缝,当原始图像主体区域较大或存在背景较复杂时,大量接缝出现在重要度较低的背景区域,就会导致背景区域发生明显扭曲,甚至丢失全部的背景区域信息。针对该问题,本文采用能量转移规则尽可能地分散图像中的接缝发生的位置,保护部分背景区域。能量转移规则过程如图 3 所示。当确定了图像的重要度图和最佳接缝后,通过计算能量比,向相邻的像素转移一定的能量值。不失一般性,仅以水平图像重定向为例,若要对高度为 H 宽度为 W 的图像进行重定向,只需对图像进行水平方向缩放图像数据转置即可。对能量转移规则为

$$\begin{cases} E'(i, j - 1) = E(i, j) + r_{left} \times E(i, j - 1) \\ E'(i, j + 1) = E(i, j) + r_{right} \times E(i, j + 1) \end{cases}, \quad (1 \leq i \leq H), \quad (6)$$

式中: $E(i, j)$ 表示图像在坐标 (i, j) 处的能量值; $E'(i, j)$ 表示更新后的能量值; r_{left} 和 r_{right} 分别表示最佳能量接缝的左右邻域能量变化比,则,

$$\begin{cases} r_{left} = \sum_{i=1}^W E(i, j - 1) / \sum_{i=1}^W E(i, j) \\ r_{right} = \sum_{i=1}^W E(i, j + 1) / \sum_{i=1}^W E(i, j) \end{cases}, \quad (1 \leq j \leq W). \quad (7)$$

在利用能量转移规则更新重要度图的基础上,线裁剪算法迭代地去除最小能量接缝,当线裁剪算法进行到一定程度后,转为均匀缩放以达到目标尺寸。本文通过计算重定向前后图像的客观相似性分数来衡量线裁剪对图像造成的失真程度。分别将原图像和线裁剪后图像进行分块处理,然后统计两个图像中各

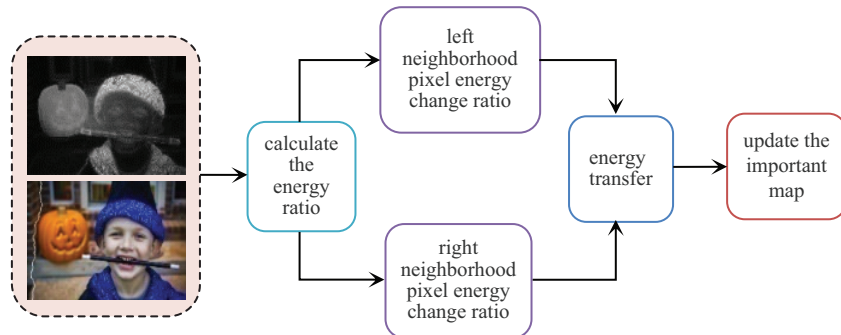


图 3 基于能量转移的图像重定向过程

Fig. 3 Image retargeting process based on the energy transfer

子块的纵横比变化来衡量图像的形变,并设定形变阈值来确定裁剪后的图像主体区域是否发生了形变。定义原图像被划分的子块大小为 $c \times c$ 的像素,那么原图像被分成的块数为 $n_1 \times n_2$,计算公式为

$$n_1 = \text{floor}(W/c), n_2 = \text{floor}(H/c), \quad (8)$$

式中:floor 是向下取整函数。原图像中所有子块对图像变形的重要程度不一致,第 (p, q) 个子块的重要度值占比 $[B(p, q)]$ 为

$$B(p, q) = \frac{\sum_{u=c \times (p-1)+1}^{c \times p} \sum_{v=c \times (q-1)+1}^{c \times q} E(u, v) / e_{L_sum}}{(1 \leq p \leq n_1, 1 \leq q \leq n_2)}, \quad (9)$$

式中: e_{L_sum} 为原图像所有像素重要度总和; (u, v) 为位于子块内部的坐标。位于图像主体位置的子块重要度占比较高,而位于背景区域的子块占比略低。

计算客观相似性分数的另一重要因素是各个分块的变化相似性,即每个子块经过重定向后与原图对应子块的相似性衡量。定义图像中第 (p, q) 个子块的相似性为 $S(p, q)$,计算公式为

$$S(p, q) = \frac{2 \times w(p, q) \times h(p, q) + \beta}{w(p, q)^2 + h(p, q) + \beta} \times \exp\left\{-\alpha \left[\frac{w(p, q) + h(p, q)}{2} - 1\right]^2\right\}, \quad (10)$$

式中: α 为大于零的参数,通过调整信息损失惩罚程度来平衡信息损失和视觉损失, $\alpha = 0.3$; β 为小的常数,避免分母为零,用于增加稳定性, $\beta = 10^{-6}$; $w(p, q)$ 和 $h(p, q)$ 分别表示第 (p, q) 个子块的宽度和高度变化率,在水平方向的图像重定向中,子块的高度不发生变化,所以 $h(p, q) = 1$,宽度变化率为

$$w(p, q) = l/c, \quad (11)$$

式中: l 表示重定向后子块的宽度。当 S 接近 1 时,说明该小块内容相对于原图得到了很好的保留,而 S 接近 0 时,说明小块存在严重的变形。该公式量化了线裁剪算法对图像局部的变形。

图像的整体相似性评价是结合每个小块的相似性分数以及小块的重要度值占比计算得出,

$$s = \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} B(p, q) \times S(p, q), \quad (12)$$

$$(1 \leq p \leq n_1, 1 \leq q \leq n_2).$$

将移除最佳接缝之后得到的图像与原图像进行相似性比较,每移除一条接缝,计算裁剪后图像和原图像的客观相似性分数(s),并且与实验设置阈值 T_h (经验值 $T_h = 0.98$) 进行比较。当 $s < T_h$ 时,转到均匀缩放方法得到所需目标尺寸的目标图像。

4 实 验

4.1 实验数据集

本文实验基于 AVA 准数据库、MIT RetargetMe 数据库和 MSRA 数据库进行。AVA 数据集是图像美学评价的最大的公开数据集,共包含约 250000 张图片,每张图片都具有美学评价标注,由于图像的美学评价是一个较为主观的指标,不同的人对同一张图片可能有不同的评分,AVA 数据集的每张图片收集了大约 200 人的评分,最终由它们的平均值作为图像的美学质量标签。评分的范围是 1~10,10 代表最高美学评价。

MIT RetargetMe 数据集是基于内容感知图像重定向公认的权威数据集,该数据集包含了 80 张原始图像以及 976 张缩放结果图,图像质量较高。

MSRA 数据集中图像分辨率相对于 MIT RetargetMe 数据集较低,但包括的图像种类很多,具体包括植物、动物、人、室内外等,具有很好的代表性。

4.2 美学性能评估

基于注意力机制的图像美学评价网络在 AVA 数据集上训练,其中 20000 张图片用于测试,剩余 230000 张图片用于训练。训练 batch-size 设置为 32,全部训练数据迭代次数 epoch 设置为 60,初始学习率设置为 10^{-4} ,每 20 个 epoch 减少为其 1/10,最低学习率为 10^{-6} ,在获得最低损失时停止训练。

AVA 数据集现有的性能评估方法是二元分类精度,以 5 为分界线将图像分为高质量和低质量两类。为了评估图像美学评价的准确率,本文分别采用斯皮尔曼秩相关系数(SRCC)和皮尔逊线性相关系数(PLCC)作为补充评估标准。与主流美学评价模型进行比较,实验结果如表 1 所示,只有 Talebi

表 1 本文方法与主流方法对比

Table 1 Comparison of between proposed method and mainstream methods

Method	SRCC	PLCC	Accuracy / %
Ref. [19]	—	—	66.70
Ref. [20]	—	—	71.42
Ref. [31]	—	—	74.46
Ref. [32]	—	—	75.76
Ref. [33]	—	—	77.40
Ref. [34]	0.558	—	77.33
Ref. [35]	—	—	81.70
Ref. [30]	0.612	0.636	81.51
Ref. [22]	0.752	0.755	81.61
Ref. [22]	0.756	0.757	81.72
Proposed algorithm	0.755	0.757	82.28

等^[30]和 Hosu 等^[22]同时采用了 SRCC 和 PLCC 作为评估标准衡量美学评价的准确性,其中,由于 Hosu 等提出的方法对数据进行了大规模的扩充,其 SRCC 结果稍高于本文算法,但准确度结果低于本文算法,而未进行数据扩充的实验结果各项性能皆低于本文算法。

4.3 重定向实验结果

本文对比了均匀缩放(US)^[29]、线裁剪(SC)^[1]、

网格变形(SNS)^[5]、分块线裁剪(BSC)^[4]、内部生成对抗网络(InGAN)^[16]和双循环重定向方法(Cycle-IR)^[15]。前四种方法基于传统方法,后两种基于深度学习。将两个数据集中的图像在水平方向上缩小为原始图像的 0.6,部分图像重定向效果如图 4 所示,其中图 4(a)和图 4(b)从 MSRA 数据库中选取,图 4(c)和图 4(d)从 RetargetMe 数据库中选取。



图 4 针对不同类型图片的七种重定向结果。(a)具有单个较大主体;(b)具有复杂背景;(c)具有多个主体;(d)主体对象复杂
Fig. 4 Seven retargeting results for different type of images. (a) With a single large subject; (b) with a complex background; (c) with multiple subjects; (d) with complex subject objects

如图 4 所示,US 算法在重定向过程中不考虑像素的视觉主次,会导致图像的主体比例发生明显变化。SC 算法虽然保护了图像的主体区域,但受梯度图限制无法全面地识别图像信息,会对图像结构造成严重破坏,如图 4(d)中南瓜发生明显形变。SNS 算法采用网格变形的方式进行缩放,对主体相对大小的改变较严重,例如图 4(c)中两艘船的大小发生了改变。BSC 算法接缝倾向于在相同区域选取,会导致背景区域语义信息丢失,例如图 4(a)中沙滩被全部裁剪,广告牌也有语义信息丢失的问题。InGAN 算法学习单个图片的分块信息后重构生成重定向图片,图片的亮度和色彩会有明显下降,且容易发生像素块丢失的问题,例如图 4(c)两艘船体中间区域像素丢失,图片质量有明显降低。Cycle-IR 算法由双循环网络学习图像结构,但对具有多个主

体或较大主体的图像保护不佳,例如图 4(a)中广告牌形状发生扭曲,图 4(d)中南瓜被过度拉伸。本文提出的基于美学特征的重定向算法在保护图像主体的同时,对背景中具有丰富美学特征的区域也能够完整的保留。同时,经过能量转移后的接缝选取可以避免对图像结构造成明显破坏。生成的重定向图片既保留了原始图像的所有语义内容,又能够保持较高的视觉美感。

这里选择利用 Liu 等^[36]提出的客观质量评估方法(IQA)评估了图 4 中展示的四幅图像使用 7 种方法的重定向效果,这是用于图像重定向质量评估方法中最权威的方法之一。评估指标的范围是 0~1,数值越大,重定向结果越好。客观评估结果统计如表 2 所示,本文所提出的基于美学特征的图像重定向方法获得了更高的评分。

表 2 四幅图像的客观评价结果

Retargeting algorithm	Fig. 4(a)	Fig. 4(b)	Fig. 4(c)	Fig. 4(d)
US	0.59	0.60	0.68	0.60
SC	0.63	0.65	0.62	0.61
SNS	0.60	0.61	0.68	0.63
BSC	0.66	0.65	0.70	0.68
InGAN	0.40	0.61	0.49	0.62
Cycle-IR	0.60	0.68	0.66	0.70
Proposed algorithm	0.68	0.73	0.70	0.72

4.4 消融实验及定量评估

为了评估本文提出的美学特征图在图像重定向过程中的有效性,对实验图像分别基于融合美学特征图的重要度图和不具有美学特征的重

要度图进行重定向操作。选取部分图像结果如图 5 所示,图 5(b)中 Img1 由于缺乏图像整体结构信息,新郎右侧手臂和建筑(框内)发生了明显形变,而图 5(c)中 Img2 则很好地保护了这些细节。图 5(b)中 Img2 由于缺少美学整体特征信息,虽然对主体对象(画板)有一定保护,但背景区域(墙体)却发生了严重扭曲,图 5(c)中 Img2 则很好地保护了图像的整体结构性。本文提出的美学特征可以在图像重定向中很好地保护整体美学结构。

为了定量地评估本文算法的有效性,分别从主观和客观两个角度对实验结果进行评估。本文使用 RetargetMe 基准测试集(包括 80 张原始图片)和从 MSRA-1000 数据集中随机挑选了 120 张共计 200 张原始图片进行实验。

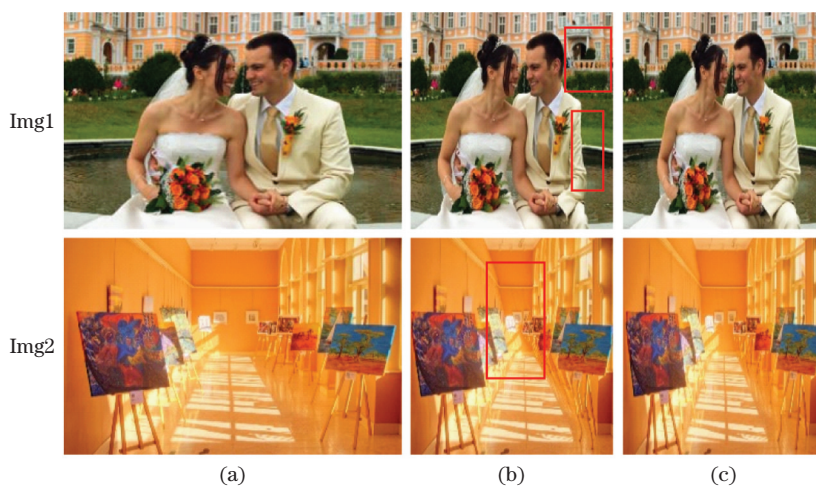


图 5 消融实验结果。(a)原始图像;(b)重要地图+梯度和直线地图;(c)审美特征图+重要图+梯度和直线图

Fig. 5 Results of ablation experiments. (a) Original images; (b) important map + gradient and straight-line map; (c) aesthetic feature map + important map + gradient and straight-line map

分别使用 8 种重定向方法将图像宽度缩小为原始图像的 0.6。并随机提供给本研究所 20 名进行图像处理研究的研究生进行评价打分,评价者只根据自身视觉感受选出缩放效果最好的图像。根据所有评价者选择结果统计出各个缩放算法的选中次数和选中率。对七种重定向算法和未融合美学特征图的结果的选中率统计结果如表 3 所示。本文算法的主观选取率高均于其他算法,生成的重定向图像更符合人类审美感知。

使用 IQA 客观评价方法评估了 8 种方法的重定向结果,统计了 200 张所选图片的客观评分,将平均值作为每种方法的最终得分。评估结果如表 4 所示,所提出的基于美学重定向算法与目前主流方法相比具有较好的排名和效果。

表 3 主观评价统计

Retargeting algorithm	Select number	Select rate /%
US	116	2.90
SC	212	5.30
SNS	407	10.18
BSC	489	12.22
InGAN	132	3.30
Cycle-IR	559	13.98
W/o aesthetic feature map	773	19.33
Proposed algorithm	1312	32.80

表 4 客观评价统计

Table 4 Objective evaluation statistics

Retargeting algorithm	Average score
US	0.60
SC	0.63
SNS	0.63
BSC	0.68
InGAN	0.59
Cycle-IR	0.68
W/o aesthetic feature map	0.68
Proposed algorithm	0.71

5 结 论

本文提出了一种基于注意力机制融合多层次特征的图像美学评价方法,通过编码器主干网络提取多层次的图像美学特征,然后由注意力机制自适应地融合学习图像的美学信息。在 AVA 数据集上进行的实验结果表明,美学评价预测结果在多个评估维度下均有所提升。此外所提取的美学特征信息与 E_{sal} 、 E_{grad} 融合,生成的重要度图指导多操作重定向算法。在 RetargetingMe 和 MSRA-1000 数据集上进行的实验结果表明,本文算法与当前主流重定向算法相比具有较大优势。

参 考 文 献

- [1] Avidan S, Shamir A. Seam carving for content-aware image resizing[J]. ACM Transactions on Graphics, 2007, 26(3): 10-18.
- [2] Mansfield A, Gehler P, van Gool L, et al. Scene carving: scene consistent image retargeting[M]// Daniilidis K, Maragos P, Paragios N. Computer vision-ECCV 2010. Lecture notes in computer science. Heidelberg: Springer, 2010, 6311: 143-156.
- [3] Wu J Y, Yang S, Du J, et al. Image retargeting with seam-carving combining saliency and depth information [J]. Acta Electronica Sinica, 2019, 47(7): 1547-1550. 吴加莹, 杨赛, 堵俊, 等. 融合显著与深度信息的缝切割重定向方法[J]. 电子学报, 2019, 47(7): 1547-1550.
- [4] Guo Y C, Liang Y H, Yu M, et al. An improved seam carving algorithm based on image blocking and optimized cumulative energy map [J]. Journal of Electronics & Information Technology, 2018, 40(2): 331-337. 郭迎春, 梁云鹤, 于明, 等. 基于图像分块和优化累积能量图的线裁剪算法[J]. 电子与信息学报, 2018, 40(2): 331-337.
- [5] Wang Y S, Tai C L, Sorkine O, et al. Optimized scale-and-stretch for image resizing[C]// ACM SIGGRAPH Asia 2008 papers on-SIGGRAPH Asia'08, December 10-13, 2008. Singapore. New York: ACM Press, 2008: 110-118.
- [6] Zhang G X, Cheng M M, Hu S M, et al. A shape-preserving approach to image resizing[J]. Computer Graphics Forum, 2009, 28(7): 1897-1906.
- [7] Lin S S, Yeh I C, Lin C H, et al. Patch-based image warping for content-aware retargeting[J]. IEEE Transactions on Multimedia, 2013, 15(2): 359-368.
- [8] Rubinstein M, Shamir A, Avidan S. Multi-operator media retargeting[J]. ACM Transactions on Graphics, 2009, 28(3): 1-11.
- [9] Cui J, Cai Q Q, Lu H J, et al. Distortion-aware image retargeting based on continuous seam carving model[J]. Signal Processing, 2020, 166: 107242.
- [10] Zhou Y, Chen Z B, Li W P. Weakly supervised reinforced multi-operator image retargeting[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(1): 126-139.
- [11] Hwang D S, Chien S Y. Content-aware image resizing using perceptual seam carving with human attention model[C]//2008 IEEE International Conference on Multimedia and Expo, June 23-April 26, 2008, Hannover, Germany. New York: IEEE Press, 2008: 1029-1032.
- [12] Cho D, Park J, Oh T H, et al. Weakly- and self-supervised learning for content-aware deep image retargeting[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4568-4577.
- [13] Lin J X, Zhou T K, Chen Z B. DeepIR: a deep semantics driven framework for image retargeting[C]// 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), July 8-12, 2019, Shanghai, China. New York: IEEE Press, 2019: 54-59.
- [14] Zhou Y Z, Zhang L M, Zhang C, et al. Perceptually aware image retargeting for mobile devices[J]. IEEE Transactions on Image Processing, 2018, 27(5): 2301-2313.
- [15] Tan W M, Yan B, Lin C M, et al. Cycle-IR: Deep cyclic image retargeting [J]. IEEE Transactions on Multimedia, 2020, 22(7): 1730-1743.
- [16] Shocher A, Bagon S, Isola P, et al. InGAN: capturing and retargeting the "DNA" of a natural image[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 4491-4500.

- [17] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [M] // Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8689: 818-833.
- [18] Datta R, Joshi D, Li J, et al. Studying aesthetics in photographic images using a computational approach [M] // Leomardis A, Bischof H, Pinz A. Computer vision-ECCV 2006. Lecture notes in computer science. Heidelberg: Springer, 2006, 3953: 288-301.
- [19] Murray N, Marchesotti L, Perronnin F. AVA: A large-scale database for aesthetic visual analysis [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2408-2415.
- [20] Kao Y Y, He R, Huang K Q. Deep aesthetic quality assessment with semantic information [J]. IEEE Transactions on Image Processing, 2017, 26 (3): 1482-1495.
- [21] Zhang L M, Yin J W, Li P, et al. Flickr image community analytics by deep noise-refined matrix factorization [J]. IEEE Transactions on Multimedia, 2020, 22(5): 1273-1284.
- [22] Hosu V, Goldlücke B, Sauppe D. Effective aesthetics prediction with multi-level spatially pooled features [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9367-9375.
- [23] Kao Y Y, He R, Huang K Q. Automatic image cropping with aesthetic map and gradient energy map [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 5-9, 2017, New Orleans, LA, USA. New York: IEEE Press, 2017: 1982-1986.
- [24] Li D B, Wu H K, Zhang J G, et al. A2-RL: Aesthetics aware reinforcement learning for image cropping [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8193-8201.
- [25] Wang W G, Shen J B, Ling H B. A deep network solution for attention and aesthetics aware photo cropping [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1531-1544.
- [26] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [27] Schölkopf B, Platt J, Hofmann T. Graph-based visual saliency [M] // Advances in neural information processing systems. Cambridge: The MIT Press, 2007: 545-552.
- [28] Zhu W J, Liang S, Wei Y C, et al. Saliency optimization from robust background detection [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2814-2821.
- [29] Ramponi G. Warped distance for space-variant linear image interpolation [J]. IEEE Transactions on Image Processing, 1999, 8(5): 629-639.
- [30] Talebi H, Milanfar P. NIMA: neural image assessment [J]. IEEE Transactions on Image Processing, 2018, 27(8): 3998-4011.
- [31] Lu X, Lin Z, Shen X H, et al. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 990-998.
- [32] Hii Y L, See J, Kairanbay M, et al. Multigap: Multi-pooled inception network with text augmentation for aesthetic prediction of photographs [C] // 2017 IEEE International Conference on Image Processing (ICIP), September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 1722-1726.
- [33] Mai L, Jin H L, Liu F. Composition-preserving deep photo aesthetics assessment [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 497-506.
- [34] Kong S, Shen X H, Lin Z, et al. Photo aesthetics ranking network with attributes and content adaptation [M] // Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 662-679.
- [35] Ma S, Liu J, Chen C W. A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 722-731.
- [36] Liu Y J, Luo X, Xuan Y M, et al. Image retargeting quality assessment [J]. Computer Graphics Forum, 2011, 30(2): 583-592.