第 46 卷 第 4 期 计 算 机 工 程 2020 年 4 月
Vol. 46 No. 4 Computer Engineering April 2020

·图形图像处理· 文章编号：1000-3428(2020)04-0266-07 文献标志码：A 中图分类号：TP751.1

基于视觉注意力分析的 3D 内容生成方法

蔡 凯，李新福，田学东

（河北大学 网络空间安全与计算机学院，河北 保定 071000）

摘 要： 由于在某些特殊场景中获取深度线索的难度较高，使得已有 3D 内容生成方法的应用受到限制。为此，以显著图代替深度图进行 2D-3D 转换，提出一种 3D 内容生成方法。使用全卷积网络（FCN）生成粗糙的显著图，通过条件随机场对 FCN 的输出结果进行优化。实验结果表明，该方法可以解决现有方法中因使用低等级特征进行视觉注意力分析而导致显著图质量不高的问题，且能够生成具有良好视觉效果的 3D 内容。

关键词： 3D 内容生成；显著性检测；显著图；全卷积网络；条件随机场

开放科学（资源服务）标志码（OSID）：

中文引用格式： 蔡凯，李新福，田学东. 基于视觉注意力分析的 3D 内容生成方法[J]. 计算机工程，2020，46(4)：266-272.

英文引用格式： CAI Kai，LI Xinfu，TIAN Xuedong. 3D content generation method based on visual attention analysis[J]. Computer Engineering，2020，46(4)：266-272.

3D Content Generation Method Based on Visual Attention Analysis

CAI Kai，LI Xinfu，TIAN Xuedong

（School of Cyber Security and Computer，Hebei University，Baoding，Hebei 071000，China）

【Abstract】 Due to the difficulty of obtaining depth cues in some special scenarios，the application of existing 3D content generation methods is limited. Therefore，by replacing depth map with saliency map for 2D to 3D conversion，this paper proposes a 3D content generation method. The rough saliency map is generated by using Fully Convolutional Network (FCN) and the output results are optimized by Conditional Random Field (CRF). Experimental results show that the proposed method can solve the problem of low quality of saliency map in existing methods caused by using low-level features for visual attention analysis，and it can generate 3D content with satisfying visual effect.

【Key words】 3D content generation；saliency detection；saliency map；Fully Convolutional Network (FCN)；Conditional Random Field (CRF)

DOI： 10.19678/j.issn.1000-3428.0056246

0 概述

目前，以 3D 技术为基础的 3D 电影[1]、立体医学影像[2-3]、虚拟现实（Virtual Reality，VR）[4]等应用在人们的日常生活中占据重要地位，这些应用所需的显示设备可能需要场景中的多个临近视角[5]，不同于立体显示设备中仅需 2 个视角。虽然可以通过多个相机同步拍摄得到视图，但事实证明，这种方法既繁琐又耗时。因此，越来越多的人使用基于深度图的 2D-3D 转换技术来获得视图。深度图由深度线索构成，深度线索是人类用来感知真实三维世界的最重要的信息。基于深度图的方法一般由 3 个步骤组成，首先通过计算或者深度传感器得到深度图，然后利用深度图和原图像合成新视图，最后将新视图与原图像相结合得到最终的 3D 图像。由于新视图与原图像存在视差，因此最终合成的 3D 图像能够在视觉上给予观察者以立体感。

目前，研究人员提出多种基于深度图的 3D 内容生成方法，这些方法大致可以分为三类，即纯人工转换方法、人工辅助转换（半自动转换）方法和全自动转换方法。纯人工转换方法将视频的每一帧进行区域/目标分割，然后再由“深度专家”将深度值分配给每一个分割出的区域/目标[6]，这种方法可以产生高质量的深度图，但是价格十分昂贵且非常费时。人工辅助转换方法在将 2D 图像转换为 3D 图像的过程中进行人工“手动”修正，虽然这种方法可以降低

基金项目： 河北省高等学校科学技术研究重点项目（ZD2017208，ZD2017209）。

作者简介： 蔡 凯（1993—），男，硕士研究生，主研方向为图像处理；李新福，教授、博士；田学东，教授、博士、博士生导师。

收稿日期： 2019-10-10 **修回日期：** 2019-11-28 **E-mail：** lampard666@sina.cn

时间消耗,但是仍然需要大量的人工成本[7-9]。为了将 2D-3D 转换技术在商业中进行更广泛的推广,需要其他方法来解决时间和人工成本问题。全自动转换方法先使用单幅图像生成深度图,然后再将 2D 图像转换为 3D 图像,全程无需人工干预。

一幅图像的显著性区域是人们在该图像中注意到的最重要的部分,因此,使显著性物体靠近人眼的同时将人们不感兴趣的目标远离人眼合乎情理。文献[10]通过视觉注意力分析进行 2D-3D 转换,并且证明了这种方法的可行性,但是,由于其使用低级别手工方式提取特征,因此所生成的显著图较为粗糙。本文使用显著图代替深度图生成 3D 内容,通过全卷积网络(Fully Convolutional Network,FCN)[11]估算粗糙的显著图并用条件随机场(Conditional Random Field,CRF)[12]进行优化,这种 FCN + CRF 的方法可以适用于大部分设备,包括个人计算机、智能手机等。

1　相关工作

1.1　显著性检测

文献[13]将视觉注意力问题用计算模型进行表达。在此之后,显著性检测领域出现了越来越多的新算法,这些算法可以分为两大类:基于任务驱动的自顶向下的算法[14-16]和由数据驱动的自底向上的算法[17-19]。自底向上的算法多数都将研究重点集中在低等级视觉特征上,例如中心偏移量[20]、对比度[17]和边缘[21]等。在显著性检测领域,从传统方法发展到深度学习方法之前,文献[22]对非深度学习方法做了充分的总结,一些表现良好的方法[20]在此阶段得到广泛应用。

随着深度学习技术的快速发展,越来越多的研究人员开始使用卷积神经网络(Convolutional Neural Network,CNN)进行显著性检测。近几年,人们更加倾向于使用端到端(end-to-end)的网络,这是因为由文献[11]提出的 FCN 可以接收各种尺寸的输入图像,缩短了人工预处理和后续处理的时间,给予模型更多的自动调节空间。为了使算法可以更充分地应用于人们的生产生活中,本文算法并未使用结构复杂的网络模型,而是使用对硬件要求相对较低的 FCN + CRF 检测模型。

1.2　基于深度线索的深度信息提取

在现实世界中,人类从外界接收到各种各样的深度线索,并利用深度线索来感知三维世界,这些深度线索可以分为两类:双目深度线索和单目深度线索。双目深度线索通过双眼对场景中感知到的 2 幅图像的差别来提供深度信息,单目深度线索通过用一只眼睛观察场景来提供深度信息。

单目深度线索又可分为图绘深度线索和运动深度线索。图绘深度线索是人类从 2D 图像中感知深度的最基本的线索,这些线索包括相对高度或尺寸、光线与遮蔽、大气散射、纹理梯度、对焦/去对焦等。在实践中,主要通过 2 种方法来从对焦/去对焦中提取深度:

1)从具有不同焦距特征的多个图像中提取模糊变化,这些变化可以转换为深度[23]。

2)通过测量每个相关像素的模糊量来从单个图像中提取模糊信息,然后将模糊量映射为该像素的深度[24-25]。

几何学相关的绘图深度线索包括相对高度或尺寸、纹理梯度等,但是,其中有些深度线索很难应用到实际中[26]。从颜色和亮度线索中提取的深度体现在颜色和亮度的变化上,这些深度线索包括大气散射、光线与遮蔽、局部对比度等。与绘图深度线索不同的是,视频序列提供运动视差作为额外的深度线索。

近年来,研究人员在使用深度学习技术进行深度图生成方面投入了大量精力,他们将一幅 2D 图像作为输入并学习如何预测出一幅与之相对应的深度图。但是,上述研究在收集高质量的图像-深度图对上遇到了很大困难,因此,数据集的缺失限制了深度图生成算法的发展。

文献[10]通过视觉注意力分析得到显著图,用显著图代替深度图来生成 3D 图像,这种方法可以取得较好的效果[25]。但是该方法中使用的是低等级视觉特征,使得算法性能十分依赖于所选取特征的优劣,因此,在大部分情况下无法获得较高的精度。为解决上述问题,本文使用 FCN 进行显著性检测,以提高检测精度并得到效果更好的 3D 图像。

2　算法设计

虽然本文方法与传统方法[5,8,24]在细节上有些差异,但是在整体流程(见图 1)上并没有太大的差别。因此,本节主要讨论本文方法与传统方法不同的部分(显著图生成阶段),视差计算与图像绘制部分可以参考文献[26]。

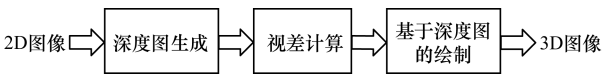

图 1　基于深度图的 2D-3D 转换流程

Fig. 1　2D to 3D conversion procedure based on depth map

2.1　显著图计算

FCN 是一种卷积神经网络,其通过一系列卷积层提取特征并使用池化层减少过拟合现象。FCN 将原始卷积网络中的全连接层直接替换为卷积层,从而使运算更便捷。最后一个卷积层的输出被称作热图,FCN 通过反卷积使热图恢复到与输入图像相同的尺寸,从而对每个像素都产生一个预测。如果只使用由热图反卷积而来的图像进行预测,得到的结果往往比较粗糙,FCN 通过定义跳跃结构来解决该问题。FCN 的整体结构如图 2 所示,卷积网络每一

层的输入数据都是三维张量,尺寸为 $h \times w \times d$,其中,h 和 w 为数据的空间维度,d 为数据的特征或者通道维度。FCN 的第一层为输入层,用来接收输入图像,输入图像的尺寸为 $h \times w$,d 为图像的颜色通道。在 CNN 中,决定某一层输出结果中一个元素所对应输入层的区域大小被称作感受野。

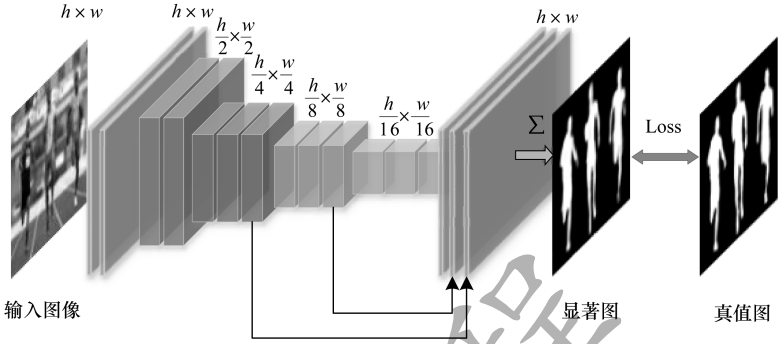

图 2 FCN 结构框架

Fig. 2 FCN framework

卷积网以平移不变性为基础,构成卷积网的组成部分(卷积、池化和激活函数)在对局部输入区域进行操作时仅与相对空间坐标有关。令 $\boldsymbol{x}_{ij}$ 为某一层中的数据向量,该向量位于坐标 (i,j) 处,$\boldsymbol{y}_{ij}$ 为下一层中的数据向量,则 $\boldsymbol{y}_{ij}$ 的计算公式为:

$$\boldsymbol{y}_{ij} = f_{ks}(\{\boldsymbol{x}_{si+\delta i, sj+\delta j}\}_{0 \leqslant \delta i, \delta j \leqslant k}) \tag{1}$$

其中,k 为算子尺寸,s 被称作步长或下采样算子,函数 f_{ks} 取决于该层类型。算子尺寸和步长遵循的变换规则如式(2)所示:

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', ss'} \tag{2}$$

随着卷积和池化操作次数的增多,图像的尺寸变得越来越小,此时可以通过反卷积操作对特征图进行尺寸还原。因此,FCN 可以利用反卷积对任意尺寸的输入进行操作,然后计算一个相应尺寸的输出。

一些常见的 CNN 只接收固定尺寸的输入图像,并且在输出中不包含空间信息,造成这种情况的原因在于网络中应用了全连接层,全连接层不能接收尺寸不同的输入张量,这也会导致空间信息丢失。然而,任何一个全连接层都可以被替换为卷积层,它们之间唯一的不同在于卷积层中的神经单元只与输入数据中的局部区域相连接,并且处在同一列卷积的神经单元共享参数。这种方法不仅可以使网络接收任意尺寸的输入,在计算的过程中保留空间信息,还可以提高前向传播的效率[11]。一种将传统卷积网络变换为全卷积网络的方法是:将原始卷积网络中的全连接层直接替换为卷积层,这时网络中只存在卷积层,全卷积网络的名称即由此得来。

经过多次卷积和池化操作之后,输入数据的空间尺寸会越来越小,当尺寸在网络中达到最小时,可将该图像称为热图,热图表示输入数据的高位特征。为了将热图恢复到与输入数据相同的尺寸,需要对热图进行上采样。FCN 通常使用一系列反卷积层和激活函数进行上采样,这是因为在网络学习的过程中,使用反卷积和激活函数的速度快、效率高[11]。

如果单独对热图进行上采样,得到的输出预测图会显得过于粗糙,一些细节无法恢复。随着层数越来越高,得到的特征会越来越抽象,细节丢失会越来越严重。FCN 使用跳跃结构对低层中的输出进行上采样并与最高层中的预测图相连接,这样可以兼顾局部和全局信息。需要注意的是,虽然使用跳跃结构可以得到更精细的预测图,但是所产生的开销会增加。

本文将 FCN 应用于显著性检测以生成显著图,但是得到的显著图精度并非很高,因此,需要使用 CRF 对结果进行优化。

2.2 基于 CRF 的结果优化

本节主要对 CRF 在图像标注任务中的应用进行概述。在像素级标签预测问题中,令输入图像为全局观测 $\boldsymbol{I}$,CRF 将像素标签作为随机变量,如果像素标签在全局观测的条件下可以构成马尔科夫随机场,则 CRF 将对像素标签进行建模。

令随机变量 $\boldsymbol{X}_i$ 是图像中像素 i 的标签,$\boldsymbol{L}$ 为预定义的标签集合,$\boldsymbol{L} = \{l_1, l_2, \cdots, l_k\}$,$k$ 为标签数。变量 $\boldsymbol{X}$ 由随机变量 $X_1, X_2, \cdots, X_N$ 组成,其中,N 为图像中像素的总个数。条件随机场符合吉布斯分布:

$$P(\boldsymbol{X} = x \mid \boldsymbol{I}) = \frac{1}{Z(\boldsymbol{I})}\exp(-E(x \mid \boldsymbol{I})) \tag{3}$$

其中,$E(x)$ 为 $x \in \boldsymbol{L}^N$ 的能量函数,$Z(\boldsymbol{I})$ 为规范化因子。为了方便起见,下文讨论中省略全局观测 $\boldsymbol{I}$。在全连接条件随机场中,$E(x)$ 的表达式为:

$$E(x) = \sum_i \Psi_c(x_i) + \sum_{i<j} \Psi_d(x_i, x_j) \tag{4}$$

其中,一元势函数 $\Psi_c(x_i)$ 测量将像素 i 标注为 x_i 的代价,即为上一小节中 FCN 的输出预测图。二元势函数 $\Psi_d(x_i, x_j)$ 同时将像素 i,j 标注为 x_i, x_j 的代价,用来描述像素点之间的关系,鼓励相似的像素分配相同的标签,差别较大的像素分属于不同的类别。二元势函数 $\Psi_d(x_i, x_j)$ 的表达式为:

$$\Psi_d(x_i, x_j) = \varphi(x_i, x_j) \sum_n^N w^{(n)} k_G^{(n)}(\boldsymbol{f}_i, \boldsymbol{f}_j) \tag{5}$$

其中,当 $n=1,2,\cdots,N$ 时,$k_G^{(n)}$ 为特征向量中的高斯核,$\boldsymbol{f}_i$ 为像素 i 的特征向量,其是从图像中提取出来的特征。$\varphi(.,.)$ 为标签兼容性函数,用来捕获不同标签之间的兼容性。

只需最小化 CRF 的能量函数 $E(x)$ 便可得到输入图像的最优预测图。因为直接最小化能量函数时计算量会非常大,所以本文使用平均近似方法[12]推断结果。

2.3　2D-3D 转换

在进行视差计算时,本文方法与传统方法差别不大。令图中像素点 (x,y) 处的视差值为 $R(x,y)$,则有:

$$R(x,y)=Z\times\left(1-\frac{S(x,y)}{128}\right) \tag{6}$$

其中,Z 为最大视差,计算公式为 $Z=\frac{G\times K}{G+D}$,K 为人双眼的距离,D 表示人眼与屏幕的间距,G 为最大深度,$S(x,y)$ 表示点 (x,y) 处的显著值(取值 0~255)。

本文方法在进行 2D-3D 转换时除了将深度图替换为显著图以外,在其他步骤上与 DIBR 方法无太大差别,且本文模型对硬件的要求较低,这 2 个优点有助于本文方法的推广。

3　实验结果与分析

3.1　实验设置

根据本文的网络结构,需要找到合适的基网络并添加一些预训练层来构成 FCN。因此,实验的第一步为找到合适的网络模型进行微调。目前,在数量庞杂的网络模型中,有多种卷积网络可供选择,比较流行的卷积网络有 AlexNet[27]、VGG16[28]、GoogLeNet[29]、InceptionV3[30]等。本文对改造后的 VGG16、AlexNet、GoogLeNet 进行了对比,使用的度量标准为:

$$\text{meanIU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{7}$$

其中,$k+1$ 为类别总数,p_{ij} 为本应预测为 i 类但预测为 j 类的像素个数(即预测错误的像素总数),p_{ii} 为预测正确的像素总数。3 种网络对比结果如表 1 所示。

表 1　3 种网络的 meanIU 与运行时间对比

Table 1　Comaprison of meanIU and runtime of three networks

网络	meanIU	预测时间/ms
FCN-VGG16	56.0	98
FCN-AlexNet	39.8	42
FCN-GoogLeNet	42.5	46

根据表 1 的结果,本文使用 VGG16 作为基网络,并利用目前较流行的 Keras 搭建显著性检测网络模型,Keras 是一种开源的深度学习框架。FCN 前 5 层的参数通过 VGGNet 初始化,其他层的参数通过标准差为 0.01、偏置为 0 的零均值高斯分布初始化。在训练开始之前,本文算法使用 SGD(Stochastic Gradient Descent)作为优化器,超参数的设置分别为:学习率 10^{-8},权重衰减 5^{-4},冲量 0.9。FCN 训练时的最大迭代次数为 100 000,每批次大小为 1,这是为了避免内存不足从而引发错误。

FCN 所使用的损失策略是,对最后一层输出图中的所有像素应用 softmax 损失函数后再求和。如果损失函数对最后一层中每个像素都进行损失计算,则将所有这些损失值求和,可由公式 $\xi(x;\theta)=\sum_{ij}\xi'(x_{ij};\theta)$ 表示,则 FCN 损失函数的梯度就等同于其空间域中每个像素的损失梯度。因此,在整幅图像中对 ξ 进行梯度下降与对所有 ξ' 进行随机梯度下降之后求和相同。

本文采用平均近似方法[12]优化 CRF,通过 PyDenseCRF(https://github.com/lucasb-eyer/pydensecrf)来实现。

3.2　数据集与度量标准

为了更好地验证本文模型的性能并与已有方法进行对比,本文使用 ECSSD[31]、PASCALS[32]和 DUT-OMRON[33]3 种数据集对网络模型进行训练与测试,这些数据集在显著性检测领域应用广泛,且每一个数据集都包含了大量的图像。ECSSD 拥有 1 000 张图像,其中多数图像具有丰富的语义并且结构复杂。PASCALS 中的图像全部是从 PASCAL VOC 2010 数据集中精心挑选而出。DUT-OMRON 数据集包含 5 000 张图像,用于进行大规模模型比较。为了保证对比的公平性,本文使用相同的训练集对所有对比模型进行训练,并且应用相同测试集进行测试。

在显著性检测领域,有 3 种被广泛认可的度量标准,分别是精度-召回率(Precision-Recall,PR)曲线、F 度量(F-measure)和平均绝对误差(Mean Absolute Error,MAE)。精度和召回率是由二值显著图和真值图做比较计算而出,假设显著图为 S,将 S 进行二值化后得到二值掩模 M,将 M 与真值 G 作比较得出:

$$P_{\text{precision}}=\frac{|M\cap G|}{|M|},R_{\text{recall}}=\frac{|M\cap G|}{|G|} \tag{8}$$

F 度量可以测量模型的整体性能,其定义为精度和召回率的加权平均值:

$$F_\beta=(1+\beta^2)\frac{P_{\text{precision}}\times R_{\text{recall}}}{\beta^2 P_{\text{precision}}+R_{\text{recall}}} \tag{9}$$

本文将 β^2 设置为 0.3,这样可以提升精度在测量中的重要性。MAE 表示显著图与真值中每个像素之间的绝对误差,计算公式为:

$$M_{\mathrm{MAE}}=\frac{1}{h\times w}\sum_{x=1}^{h}\sum_{y=1}^{w}\left|\bar{M}(x,y)-\bar{G}(x,y)\right| \quad (10)$$

其中,h 为显著图和真值图的高度,w 为显著图和真值图的宽度,$\bar{M}$ 和 $\bar{G}$ 是 M 和 G 经过标准化后所得,区间范围为[0,1]。

3.3 结果分析

本文将文献[10]方法、RC 方法[17]、DRFI 方法[20]作为对比方法。文献[10]方法使用基于对比度的方法进行显著性检测。作为传统显著性检测方法中的代表,DRFI 方法在各个数据集上都有较好的性能表现,其将显著性检测看作一个回归问题,通过监督学习的方法对图像进行分割,然后再将分割得到的区域特征进行整合得到最终的显著图。RC 结合了空间关系来获取显著图,但是其相比于文献[10]方法而言计算量更高。4 种方法的 F_β 和 MAE 对比结果如表 2 所示,其中,最优结果用加粗字体标出。

表 2 4 种方法在 3 个数据集上的性能对比结果

Table 2 Performance comparison results of four methods on three datasets

方法	PASCALS		ECSSD		DUT-OMRON	
	F_β	MAE	F_β	MAE	F_β	MAE
本文方法	**0.734**	**0.159**	**0.816**	**0.110**	**0.709**	**0.094**
DRFI 方法	0.679	0.221	0.787	0.166	0.665	0.155
RC 方法	0.640	0.225	0.741	0.187	0.599	0.189
文献[10]方法	0.387	0.441	0.460	0.331	0.382	0.310

从表 2 可以看出,本文深度学习方法在 F-measure 和 MAE 上的表现明显优于传统方法,主要原因在于深度学习方法强大的特征提取能力和抽象能力。本文选取了各模型在 DUT-OMRON 上的精度-召回率来绘制 PR 曲线,结果如图 3 所示,从图 3可以看出,本文深度学习方法性能同样优于传统学习方法。

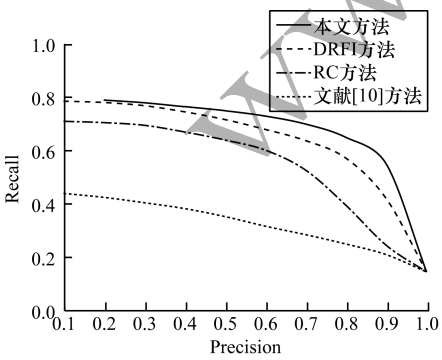

图 3 4 种方法在 DUT-OMRON 数据集上的 PR 曲线

Fig. 3 PR curves of four algorithms on DUT-OMRON dataset

图 4 所示为深度学习方法与传统方法的视觉效果对比,从图 4 可以看出,无论是在简单场景还是复杂场景下,深度学习方法所生成的显著图都要优于其他传统方法,这是因为无论在哪种场景下,手工选取的显著性特征都具有一定的局限性,而深度学习方法则可以通过自我选取特征和高度抽象特征的能力对场景进行分析,从而预测出更为精确的显著图。

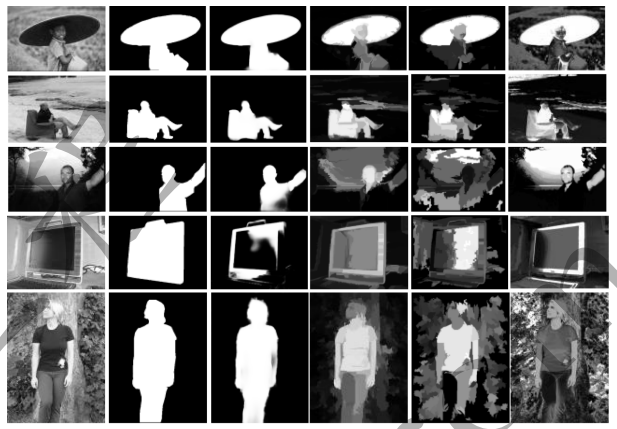

图 4 不同显著性检测方法的视觉效果对比

Fig. 4 Visual effect comparison of different saliency detection algorithms

对于所生成的 3D 内容而言,最重要的就是给予观察者舒适的视觉体验,因此,本文采用文献[34]中的立体图像质量评价方法对本文模型进行评估,该评价方法通过提取感兴趣区域的视差图,综合多种影响视觉舒适度的因素,建立一个多维度的 3D 内容视觉舒适度评价系统。评价的满分为 5 分,代表“舒适”,4 分表示“较舒适”,3 分为“效果一般”,2 分为“稍有不适”,1 分为“难以忍受”。本次实验选用文献[7]方法作为对比对象,该方法借助 L1 范数对异常数据的抵制,在一个统一框架下实现结构相关、具有容错能力的稀疏深度稠密插值。本文取 NJUD[35]数据集中的 15 种场景图像作为实验测试集,实验结果如图 5 所示。

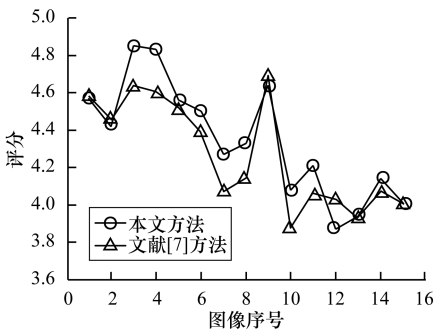

图 5 2 种 3D 内容生成方法在测试集上的图像质量评分

Fig. 5 Score of map quality of two 3D content generation methods on testing set

由图 5 可知,本文方法在 15 张图像中的平均评分为 4.348,高于文献[7]方法(4.272),主要原因在于,如果希望观察者得到满意的 3D 效果,那么将场景中最令人感兴趣的目标置于人眼舒适区至关重要[36],本文方法正是利用这一立体感知原理来生成 3D 内容。实验结果也证明了显著性检测方法可以给予观察者良好的立体感受。

图 6 所示为本文方法生成的 3D 视觉效果图。从中可以看出,本文方法在显著性较高的区域/物体上能取得良好的 3D 视觉效果,也论证了使用视觉注意力分析生成 3D 内容的可行性。

(a)原图像

(b)显著图

(c)红蓝立体图像

图 6　本文方法的 3D 视觉效果

Fig. 6　3D visutal effect of the proposed method

4　结束语

显著图通过高效的深度学习方法进行预测,并且对硬件要求较低,因此,本文使用显著图代替深度图进行 2D-3D 转换,提出一种基于视觉注意力分析的 3D 内容生成方法。实验结果表明,该方法在生成 3D 内容时能够取得良好的视觉效果。但是,本文方法在处理视频序列时仍然具有局限性,下一步考虑将时间序列信息融入到网络模型中,以提升该方法的实用性。

参考文献

[1] ZILLY F, MÜLLER M, EISERT P, et al. The stereoscopic analyzer——an image-based assistance tool for stereo shooting and 3D production[C]//Proceedings of 2010 IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2010: 4029-4032.

[2] FENG Haozhe, ZHANG Peng, XU Xinnan, et al. An unsupervised suggestive annotation algorithm for 3D CT image processing[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(2): 3-9. (in Chinese)
冯浩哲,张鹏,徐欣楠,等. 面向 3D CT 影像处理的无监督推荐标注算法[J]. 计算机辅助设计与图形学学报,2019,31(2):3-9.

[3] SONG Hongning, GUO Ruiqiang. Application and research progresses of three-dimensional printing based on medical imaging in diagnosis and treatment of cardiovascular diseases[J]. Chinese Journal of Medical Imaging Technology, 2017, 33(3): 375-380. (in Chinese)
宋宏宁,郭瑞强. 基于医学影像学的 3D 打印技术在心血管疾病诊疗中的应用现状及研究进展[J]. 中国医学影像技术,2017,33(3):375-380.

[4] LU Chihao, LI Haifeng, GAO Tao, et al. Virtual reality head-mounted display with large field of view based on stitching[J]. Acta Optica Sinica, 2019, 39(6): 150-157. (in Chinese)
陆驰豪,李海峰,高涛,等. 基于拼接的大视场虚拟现实头戴显示装置[J]. 光学学报,2019,39(6):150-157.

[5] HARMAN P V, FLACK J, FOX S, et al. Rapid 2D-to-3D conversion[EB/OL]. [2019-09-20]. https://www.ixueshu.com/document/fb7ae8da636b56a2318947a18e7f9386.html.

[6] HARMAN P. Home based 3D entertainment-an overview[C]// Proceedings of 2000 International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2000: 1-4.

[7] YUAN Hongxing, AN Peng, WU Shaoqun, et al. Error-tolerant semi-automatic 2D-to-3D conversion via L_1 optimization[J]. Acta Electronica Sinica, 2018, 46(2): 447-455. (in Chinese)
袁红星,安鹏,吴少群,等. 具有容错能力的 L_1 最优化半自动 2D 转 3D[J]. 电子学报,2018,46(2):447-455.

[8] LIE W N, HU C H, CHEN Y K, et al. Multi-layer background sprite model for 2D-to-3D video conversion [C]// Proceedings of 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Washington D. C., USA: IEEE Press, 2017: 270-274.

[9] ZHANG R, TSAI P S, CRYER J E, et al. Shape-from-shading: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690-706.

[10] KIM J, BAIK A, JUNG Y J, et al. 2D-to-3D conversion by using visual attention analysis[EB/OL]. [2019-09-20]. https://www.researchgate.net/publication/25273 3584_2D-to-3D_Conversion_by_Using_Visual_Attention_Analysis.

[11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3431-3440.

[12] KRÄHENBÜHL P, KOLTUN V. Efficient inference in fully connected CRFS with Gaussian edge potentials [C]// Proceedings of the 24th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2011: 109-117.

[13] 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.

[14] KRUTHIVENTI S S S, AYUSH K, VENKATESH BABU R. DeepFix:a fully convolutional neural network for predicting human eye fixations[J]. IEEE Transactions on Image Processing,2017,26(9):4446-4456.

[15] WANG W, SHEN J. Deep visual attention prediction[J]. IEEE Transactions on Image Processing, 2017, 27(5): 2368-2378.

[16] LI J, RAJAN D, YANG J. Locality and context-aware top-down saliency[J]. IET Image Processing, 2017, 12(3):400-407.

[17] CHENG M M, ZHANG G X, MITRA N J, et al. Global contrast based salient region detection[C]//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2011:569-582.

[18] ZHANG P, WANG D, LU H, et al. Amulet: aggregating multi-level convolutional features for salient object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017:202-211.

[19] WANG W, SHEN J, DONG X, et al. Salient object detection driven by fixation prediction[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018:1711-1720.

[20] JIANG H, WANG J, YUAN Z, et al. Salient object detection: a discriminative regional feature integration approach[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013:2083-2090.

[21] ROSIN P L. A simple method for detecting salient regions[J]. Pattern Recognition, 2009, 42(11):2363-2371.

[22] BORJI A, CHENG M M, JIANG H, et al. Salient object detection: a benchmark[J]. IEEE Transactions on Image Processing, 2015, 24(12):5706-5722.

[23] FAVARO P. Shape from focus and defocus: convexity, quasiconvexity and defocus-invariant textures[C]// Proceedings of 2007 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2007:1-7.

[24] TSAI T H, HUANG T W, WANG R Z. A novel method for 2D-to-3D video conversion based on boundary information[J]. EURASIP Journal on Image and Video Processing, 2018(1):2-5.

[25] GUO G, ZHANG N, HUO L, et al. 2D to 3D convertion based on edge defocus and segmentation[C]// Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D. C., USA: IEEE Press, 2008:2181-2184.

[26] ZHANG L, VAZQUEZ C, KNORR S. 3D-TV content creation: automatic 2D-to-3D video conversion[J]. IEEE Transactions on Broadcasting, 2011, 57(2):372-383.

[27] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Washington D. C., USA: IEEE Press, 2012:1097-1105.

[28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2019-09-20]. http://www.arxiv.org/pdf/1409.1556.pdf.

[29] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015:1-9.

[30] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016:2818-2826.

[31] YAN Q, XU L, SHI J, et al. Hierarchical saliency detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013:1155-1162.

[32] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Hypercolumns for object segmentation and fine-grained localization[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015:447-456.

[33] YANG C, ZHANG L, LU H, et al. Saliency detection via graph-based manifold ranking[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013:3166-3173.

[34] QUAN Wei, ZHAO Yunxiu, HAN Cheng, et al. Comfort evaluation model based on region of interest and contrast of stereo images[J]. Acta Photonica Sinica, 2018, 47(12):174-180. (in Chinese)
权巍,赵云秀,韩成,等. 基于立体图像感兴趣区域及对比度的舒适度评价模型[J]. 光子学报, 2018, 47(12):174-180.

[35] JU R, GE L, GENG W, et al. Depth saliency based on anisotropic center-surround difference[C]//Proceedings of 2014 IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2014: 1115-1119.

[36] WANG W, SHEN J, YU Y, et al. Stereoscopic thumbnail creation via efficient stereo saliency detection[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 23(8):2014-2027.

编辑 吴云芳