



(12)发明专利申请

(10)申请公布号 CN 110175986 A
(43)申请公布日 2019.08.27

(21)申请号 201910327556.4

(22)申请日 2019.04.23

(71)申请人 浙江科技学院

地址 310023 浙江省杭州市西湖区留和路
318号

(72)发明人 周武杰 吕营 雷景生 张伟
何成 王海江

(74)专利代理机构 宁波奥圣专利代理事务所
(普通合伙) 33226

代理人 周珏

(51)Int.Cl.

G06T 7/00(2017.01)

G06T 7/50(2017.01)

G06N 3/04(2006.01)

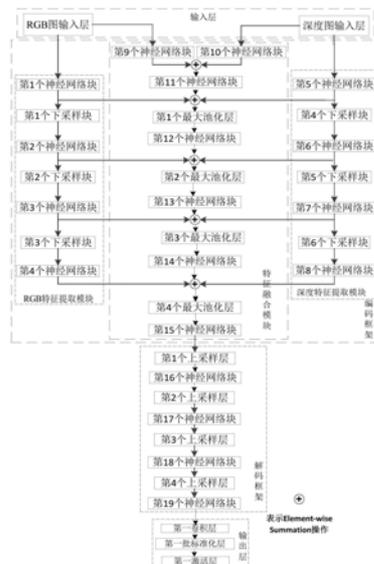
权利要求书7页 说明书15页 附图1页

(54)发明名称

一种基于卷积神经网络的立体图像视觉显著性检测方法

(57)摘要

本发明公开了一种基于卷积神经网络的立体图像视觉显著性检测方法,其构建卷积神经网络,包含输入层、隐层、输出层,输入层包括RGB图输入层和深度图输入层,隐层包括编码框架和解码框架,编码框架由RGB特征提取模块、深度特征提取模块和特征融合模块组成;将训练集中的每幅立体图像的左视点图像和深度图像输入到卷积神经网络中进行训练,得到训练集中的每幅立体图像的显著性图像;计算训练集中的每幅立体图像的显著性图像与真实人眼注视图像之间的损失函数值,重复执行多次后得到卷积神经网络训练模型;待测试的立体图像的左视点图像和深度图像输入到卷积神经网络训练模型中,预测得到显著性预测图像;优点是具有较高的视觉显著性检测准确性。



CN 110175986 A

1. 一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于包括训练阶段和测试阶段两个过程;

所述的训练阶段过程的具体步骤为:

步骤1_1:选取N幅宽度为W且高度为H的原始的立体图像;然后将选取的所有原始的立体图像及所有原始的立体图像各自的左视点图像、深度图像和真实人眼注视图像构成训练集,将训练集中的第n幅原始的立体图像记为 $\{I^n(x,y)\}$,将 $\{I^n(x,y)\}$ 的左视点图像、深度图像和真实人眼注视图像对应记为 $\{L_{RGB}^n(x,y)\}$ 、 $\{D^n(x,y)\}$ 、 $\{G_{real}^n(x,y)\}$;其中,N为正整数, $N \geq 300$,W和H均能够被2整除,n为正整数,n的初始值为1, $1 \leq n \leq N$, $1 \leq x \leq W$, $1 \leq y \leq H$, $I^n(x,y)$ 表示 $\{I^n(x,y)\}$ 中坐标位置为(x,y)的像素点的像素值, $L_{RGB}^n(x,y)$ 表示 $\{L_{RGB}^n(x,y)\}$ 中坐标位置为(x,y)的像素点的像素值, $D^n(x,y)$ 表示 $\{D^n(x,y)\}$ 中坐标位置为(x,y)的像素点的像素值, $G_{real}^n(x,y)$ 表示 $\{G_{real}^n(x,y)\}$ 中坐标位置为(x,y)的像素点的像素值;

步骤1_2:构建卷积神经网络:该卷积神经网络包含输入层、隐层、输出层,输入层包括RGB图输入层和深度图输入层,隐层包括编码框架和解码框架,编码框架由RGB特征提取模块、深度特征提取模块和特征融合模块三部分组成,RGB特征提取模块由第1个至第4个神经网络块、第1个至第3个下采样块组成,深度特征提取模块由第5个至第8个神经网络块、第4个至第6个下采样块组成,特征融合模块由第9个至第15个神经网络块、第1个至第4个最大池化层组成,解码框架由第16个至第19个神经网络块、第1个至第4个上采样层组成;输出层由第一卷积层、第一批标准化层和第一激活层组成,第一卷积层的卷积核大小为 3×3 、步幅大小为1、卷积核个数为1、填充为1,第一激活层的激活方式为“Sigmoid”;

对于RGB图输入层,其输入端接收一幅训练用左视点图像,其输出端输出训练用左视点图像给隐层;其中,要求训练用左视点图像的宽度为W且高度为H;

对于深度图输入层,其输入端接收RGB图输入层的输入端接收的训练用左视点图像对应的训练用深度图像,其输出端输出训练用深度图像给隐层;其中,训练用深度图像的宽度为W且高度为H;

对于RGB特征提取模块,第1个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第1个神经网络块的输出端输出64幅宽度为W且高度为H的特征图,将输出的所有特征图构成的集合记为 P_1 ;第1个下采样块的输入端接收 P_1 中的所有特征图,第1个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_1 ;第2个神经网络块的输入端接收 X_1 中的所有特征图,第2个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_2 ;第2个下采样块的输入端接收 P_2 中的所有特征图,第2个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_2 ;第3个神经网络块的输入端接

收 X_2 中的所有特征图,第3个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_3 ;第3个下采样块的输入端接收 P_3 中的所有特征图,第3个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_3 ;第4个神经网络块的输入端接收 X_3 中的所有特征图,第4个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_4 ;

对于深度特征提取模块,第5个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第5个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_5 ;第4个下采样块的输入端接收 P_5 中的所有特征图,第4个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_4 ;第6个神经网络块的输入端接收 X_4 中的所有特征图,第6个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_6 ;第5个下采样块的输入端接收 P_6 中的所有特征图,第5个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_5 ;第7个神经网络块的输入端接收 X_5 中的所有特征图,第7个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_7 ;第6个下采样块的输入端接收 P_7 中的所有特征图,第6个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_6 ;第8个神经网络块的输入端接收 X_6 中的所有特征图,第8个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_8 ;

对于特征融合模块,第9个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第9个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_9 ;第10个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第10个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{10} ;对 P_9 中的所有特征图和 P_{10} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 E_1 ;第11个神经网络块的输入端接收 E_1 中的所有特征图,第11个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{11} ;对 P_1 中的所有特征图、 P_5 中的所有特征图和 P_{11} 中的所有

特征图进行Element-wise Summation操作,Element-wise Summation操作后输出64幅宽度为W且高度为H的特征图,将输出的所有特征图构成的集合记为E₂;第1个最大池化层的输入端接收E₂中的所有特征图,第1个最大池化层的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为Z₁;第12个神经网络块的输入端接收Z₁中的所有特征图,第12个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为P₁₂;对P₂中的所有特征图、P₆中的所有特征图和P₁₂中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为E₃;第2个最大池化层的输入端接收E₃中的所有特征图,第2个最大池化层的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为Z₂;第13个神经网络块的输入端接收Z₂中的所有特征图,第13个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为P₁₃;对P₃中的所有特征图、P₇中的所有特征图和P₁₃中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为E₄;第3个最大池化层的输入端接收E₄中的所有特征图,第3个最大池化层的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为Z₃;第14个神经网络块的输入端接收Z₃中的所有特征图,第14个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为P₁₄;对P₄中的所有特征图、P₈中的所有特征图和P₁₄中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为E₅;第4个最大池化层的输入端接收E₅中的所有特征图,第4个最大池化层的输出端输出512幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为Z₄;第15个神经网络块的输入端接收Z₄中的所有特征图,第15个神经网络块的输出端输出1024幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为P₁₅;

对于解码框架,第1个上采样层的输入端接收P₁₅中的所有特征图,第1个上采样层的输

出端输出1024幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 S_1 ;第16个神经网络块的输入端接收 S_1 中的所有特征图,第16个神经网络块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_{16} ;第2个上采样层的输入端接收 P_{16} 中的所有特征图,第2个上采样层的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 S_2 ;第17个神经网络块的输入端接收 S_2 中的所有特征图,第17个神经网络块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_{17} ;第3个上采样层的输入端接收 P_{17} 中的所有特征图,第3个上采样层的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 S_3 ;第18个神经网络块的输入端接收 S_3 中的所有特征图,第18个神经网络块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_{18} ;第4个上采样层的输入端接收 P_{18} 中的所有特征图,第4个上采样层的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 S_4 ;第19个神经网络块的输入端接收 S_4 中的所有特征图,第19个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{19} ;

对于输出层,第一卷积层的输入端接收 P_{19} 中的所有特征图,第一卷积层的输出端输出一幅宽度为 W 且高度为 H 的特征图;第一批标准化层的输入端接收第一卷积层的输出端输出的特征图;第一激活层的输入端接收第一批标准化层的输出端输出的特征图;第一激活层的输出端输出一幅训练用左视点图像对应的立体图像的显著性图像;其中,显著性图像的宽度为 W 且高度为 H ;

步骤1_3:将训练集中的每幅原始的立体图像的左视点图像作为训练用左视点图像,并将训练集中的每幅原始的立体图像的深度图像作为训练用深度图像,输入到卷积神经网络中进行训练,得到训练集中的每幅原始的立体图像的显著性图像,将 $\{I^n(x,y)\}$ 的显著性图像记为 $\{G_{pre}^n(x,y)\}$;其中, $G_{pre}^n(x,y)$ 表示 $\{G_{pre}^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值;

步骤1_4:计算训练集中的每幅原始的立体图像的显著性图像与真实人眼注视图像之间的损失函数值,将 $\{G_{pre}^n(x,y)\}$ 与 $\{G_{real}^n(x,y)\}$ 之间的损失函数值记为

$Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$, $Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$ 采用均方误差损失函数获得;

步骤1_5:重复执行步骤1_3和步骤1_4共 V 次,得到卷积神经网络训练模型,并共得到 $N \times V$ 个损失函数值;然后从 $N \times V$ 个损失函数值中找出值最小的损失函数值;接着将值最小的

损失函数值对应的权值矢量和偏置项对应作为卷积神经网络训练模型的最优权值矢量和最优偏置项,对应记为 W^{best} 和 b^{best} ;其中, $V>1$;

所述的测试阶段过程的具体步骤为:

步骤2_1:令 $\{\hat{I}(x',y')\}$ 表示待测试的宽度为 W' 且高度为 H' 的立体图像,将 $\{\hat{I}(x',y')\}$ 的左视点图像和深度图像对应记为 $\{\hat{L}_{RGB}(x',y')\}$ 和 $\{\hat{D}(x',y')\}$;其中, $1\leq x'\leq W'$, $1\leq y'\leq H'$, $\hat{I}(x',y')$ 表示 $\{\hat{I}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值, $\hat{L}_{RGB}(x',y')$ 表示 $\{\hat{L}_{RGB}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值, $\hat{D}(x',y')$ 表示 $\{\hat{D}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值;

步骤2_2:将 $\{\hat{L}_{RGB}(x',y')\}$ 和 $\{\hat{D}(x',y')\}$ 输入到卷积神经网络训练模型中,并利用 W^{best} 和 b^{best} 进行预测,得到 $\{\hat{I}(x',y')\}$ 的显著性预测图像,记为 $\{\hat{G}_{pre}(x',y')\}$;其中, $\hat{G}_{pre}(x',y')$ 表示 $\{\hat{G}_{pre}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值。

2.根据权利要求1所述的一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于所述的步骤1_2中,第1个至第8个神经网络块的结构相同,其由依次设置的第一空洞卷积层、第二批标准化层、第二激活层、第一残差块、第二空洞卷积层、第三批标准化层构成,第一空洞卷积层的输入端为其所在的神经网络块的输入端,第二批标准化层的输入端接收第一空洞卷积层的输出端输出的所有特征图,第二激活层的输入端接收第二批标准化层的输出端输出的所有特征图,第一残差块的输入端接收第二激活层的输出端输出的所有特征图,第二空洞卷积层的输入端接收第一残差块的输出端输出的所有特征图,第三批标准化层的输入端接收第二空洞卷积层的输出端输出的所有特征图,第三批标准化层的输出端为其所在的神经网络块的输出端;其中,第1个和第5个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为64,第2个和第6个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为128,第3个和第7个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为256,第4个和第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为512,第1个至第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核大小均为 3×3 、步幅均为1、空洞均为2、填充均为2,第1个至第8个神经网络块各自中的第二激活层的激活方式均为“ReLU”;

第9个和第10个神经网络块的结构相同,其由依次设置的第二卷积层和第四批标准化层构成,第二卷积层的输入端为其所在的神经网络块的输入端,第四批标准化层的输入端接收第二卷积层的输出端输出的所有特征图,第四批标准化层的输出端为其所在的神经网络块的输出端;其中,第9个和第10个神经网络块各自中的第二卷积层的卷积核个数均为3、卷积核大小均为 7×7 、步幅均为1、填充均为3;

第11个和第12个神经网络块的结构相同,其由依次设置的第三卷积层、第五批标准化层、第三激活层、第四卷积层、第六批标准化层构成,第三卷积层的输入端为其所在的神经网络块的输入端,第五批标准化层的输入端接收第三卷积层的输出端输出的所有特征图,

第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图,第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图,第四卷积层的输入端接收第三激活层的输出端输出的所有特征图,第六批标准化层的输入端接收第四卷积层的输出端输出的所有特征图,第六批标准化层的输出端为其所在的神经网络块的输出端;其中,第11个神经网络块中的第三卷积层和第四卷积层的卷积核个数为64,第12个神经网络块中的第三卷积层和第四卷积层的卷积核个数为128,第11个和第12个神经网络块各自中的第三卷积层和第四卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1;第11个和第12个神经网络块各自中的第三激活层的激活方式为“ReLU”;

第13个至第19个神经网络块的结构相同,其由依次设置的第五卷积层、第七批标准化层、第四激活层、第六卷积层、第八批标准化层、第五激活层、第七卷积层、第九批标准化层构成,第五卷积层的输入端为其所在的神经网络块的输入端,第七批标准化层的输入端接收第五卷积层的输出端输出的所有特征图,第四激活层的输入端接收第七批标准化层的输出端输出的所有特征图,第六卷积层的输入端接收第四激活层的输出端输出的所有特征图,第八批标准化层的输入端接收第六卷积层的输出端输出的所有特征图,第五激活层的输入端接收第八批标准化层的输出端输出的所有特征图,第七卷积层的输入端接收第五激活层的输出端输出的所有特征图,第九批标准化层的输入端接收第七卷积层的输出端输出的所有特征图,第九批标准化层的输出端为其所在的神经网络块的输出端;其中,第13个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为256,第14个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为512,第15个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为1024,第16个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为512、512、256,第17个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为256、256、128,第18个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为128、128、64,第19个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为64,第13个至第19个神经网络块各自中的第五卷积层、第六卷积层和第七卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1,第13个至第19个神经网络块各自中的第四激活层和第五激活层的激活方式均为“ReLU”。

3. 根据权利要求2所述的一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于所述的步骤1_2中,第1个至第6个下采样块的结构相同,其由第二残差块构成,第二残差块的输入端为其所在的下采样块的输入端,第二残差块的输出端为其所在的下采样块的输出端。

4. 根据权利要求3所述的一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于所述的第一残差块和所述的第二残差块的结构相同,其包括3个卷积层、3个批标准化层和3个激活层,第1个卷积层的输入端为其所在的残差块的输入端,第1个批标准化层的输入端接收第1个卷积层的输出端输出的所有特征图,第1个激活层的输入端接收第1个批标准化层的输出端输出的所有特征图,第2个卷积层的输入端接收第1个激活层的输出端输出的所有特征图,第2个批标准化层的输入端接收第2个卷积层的输出端输出的所有特征图,第2个激活层的输入端接收第2个批标准化层的输出端输出的所有特征图,第3个卷积层的输入端接收第2个激活层的输出端输出的所有特征图,第3个批标准化层的输入端接收第

3个卷积层的输出端输出的所有特征图,第1个卷积层的输入端接收的所有特征图和第3个批标准化层的输出端输出的所有特征图相加,再经过第3个激活层后第3个激活层的输出端输出的所有特征图作为所在的残差块的输出端输出的所有特征图;其中,第1个和第5个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为64,第2个和第6个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为128,第3个和第7个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为256,第4个和第8个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为512,第1个至第8个神经网络块各自中的第一残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第8个神经网络块各自中的第一残差块中的第2个卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1,第1个和第4个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为64,第2个和第5个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为128,第3个和第6个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为256,第1个至第6个下采样块各自中的第二残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第6个下采样块各自中的第二残差块中的第2个卷积层的卷积核大小均为 3×3 、步幅均为2、填充均为1,3个激活层的激活方式均为“ReLU”。

5. 根据权利要求1至4中任一项所述的一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于所述的步骤1_2中,第1个至第4个最大池化层的池化窗口的大小均为 2×2 、步幅均为2。

6. 根据权利要求5所述的一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征在于所述的步骤1_2中,第1个至第4个上采样层的采样模式均为双线性插值、比例因子均为2。

一种基于卷积神经网络的立体图像视觉显著性检测方法

技术领域

[0001] 本发明涉及一种视觉显著性检测技术,尤其是涉及一种基于卷积神经网络的立体图像视觉显著性检测方法。

背景技术

[0002] 视觉显著性是近年来神经科学、机器人技术、计算机视觉等多个领域的热门研究课题。关于视觉显著性检测的研究可以分为两大类:眼球注视预测和显著性目标检测。前者是预测人在观看自然景物时的若干注视点,后者是准确提取感兴趣的物体。一般来说,视觉显著性检测算法可以分为自顶向下和自底向上两类。自顶向下的方法是任务驱动的,需要监督学习。而自底向上的方法通常使用低层线索,如颜色特征、距离特征和启发式显著性特征。最常用的启发式显著性特征之一是对比度,例如基于像素或基于块的对比度。以往对视觉显著性检测的研究大多集中在二维图像上。然而发现,首先,三维数据代替二维数据更切合实际应用;其次,随着视觉场景变得越来越复杂,仅利用二维数据来提取突出的物体是不够的。近年来,随着诸如Time-of-Flight传感器和Microsoft Kinect等三维数据获取技术的进步,推动了结构有限元法的采用,提高了外观相似的不同物体之间的识别能力。深度数据易于捕获,与光线无关,还可以提供几何线索,提高视觉显著性预测。由于RGB数据与深度数据的互补性,提出了大量将RGB图像与深度图像成对组合用于视觉显著性检测的方法。之前的工作主要集中在利用领域特定的先验知识来构建低层次的显著性特征,例如人类倾向于更关注更近的对象,然而这种观察很难推广到所有的场景。在以往的大部分工作中,多模态融合问题都是通过直接序列化RGB-D通道来解决的,或者独立处理每种模态,然后结合两种模态的决策。虽然这些策略取得了很大的改进,但它们很难充分探索跨模式互补。近年来,随着卷积神经网络(CNNs)在学习RGB数据判别特征方面的成功,越来越多的工作利用CNNs探索更强大的有效多模态组合的RGB-D表示。这些工作大多基于两流的体系结构,其中RGB数据和深度数据在一个独立的自底向上的流中学习,并在早期或后期结合特征进行联合推理。作为最流行的解决方案,双流架构比基于手工RGB-D特性的工作实现了显著的改进,然而,存在最关键问题:如何有效的利用自底向上过程中的多模态互补信息。因此,有必要对RGB-D图像视觉显著性检测技术进行进一步研究,以提高视觉显著性检测的准确性。

发明内容

[0003] 本发明所要解决的技术问题是提供一种基于卷积神经网络的立体图像视觉显著性检测方法,其具有较高的视觉显著性检测准确性。

[0004] 本发明解决上述技术问题所采用的技术方案为:一种基于卷积神经网络的立体图像视觉显著性检测方法,其特征不在于包括训练阶段和测试阶段两个过程;

[0005] 所述的训练阶段过程的具体步骤为:

[0006] 步骤1_1:选取N幅宽度为W且高度为H的原始的立体图像;然后将选取的所有原始的立体图像及所有原始的立体图像各自的左视点图像、深度图像和真实人眼注视图像构成

训练集,将训练集中的第 n 幅原始的立体图像记为 $\{I^n(x,y)\}$,将 $\{I^n(x,y)\}$ 的左视点图像、深度图像和真实人眼注视图像对应记为 $\{L_{RGB}^n(x,y)\}$ 、 $\{D^n(x,y)\}$ 、 $\{G_{real}^n(x,y)\}$;其中, N 为正整数, $N \geq 300$, W 和 H 均能够被2整除, n 为正整数, n 的初始值为1, $1 \leq n \leq N$, $1 \leq x \leq W$, $1 \leq y \leq H$, $I^n(x,y)$ 表示 $\{I^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值, $L_{RGB}^n(x,y)$ 表示 $\{L_{RGB}^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值, $D^n(x,y)$ 表示 $\{D^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值, $G_{real}^n(x,y)$ 表示 $\{G_{real}^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值;

[0007] 步骤1_2:构建卷积神经网络:该卷积神经网络包含输入层、隐层、输出层,输入层包括RGB图输入层和深度图输入层,隐层包括编码框架和解码框架,编码框架由RGB特征提取模块、深度特征提取模块和特征融合模块三部分组成,RGB特征提取模块由第1个至第4个神经网络块、第1个至第3个下采样块组成,深度特征提取模块由第5个至第8个神经网络块、第4个至第6个下采样块组成,特征融合模块由第9个至第15个神经网络块、第1个至第4个最大池化层组成,解码框架由第16个至第19个神经网络块、第1个至第4个上采样层组成;输出层由第一卷积层、第一批标准化层和第一激活层组成,第一卷积层的卷积核大小为 3×3 、步幅大小为1、卷积核个数为1、填充为1,第一激活层的激活方式为“Sigmoid”;

[0008] 对于RGB图输入层,其输入端接收一幅训练用左视点图像,其输出端输出训练用左视点图像给隐层;其中,要求训练用左视点图像的宽度为 W 且高度为 H ;

[0009] 对于深度图输入层,其输入端接收RGB图输入层的输入端接收的训练用左视点图像对应的训练用深度图像,其输出端输出训练用深度图像给隐层;其中,训练用深度图像的宽度为 W 且高度为 H ;

[0010] 对于RGB特征提取模块,第1个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第1个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_1 ;第1个下采样块的输入端接收 P_1 中的所有特征图,第1个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_1 ;第2个神经网络块的输入端接收 X_1 中的所有特征图,第2个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_2 ;第2个下采样块的输入端接收 P_2 中的所有特征图,第2个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_2 ;第3个神经网络块的输入端接收 X_2 中的所有特征图,第3个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_3 ;第3个下采样块的输入端接收 P_3 中的所有特征图,第3个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_3 ;第4个神经网络块的输入端接收 X_3 中的所有特征图,第4个神

经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_4 ;

[0011] 对于深度特征提取模块,第5个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第5个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_5 ;第4个下采样块的输入端接收 P_5 中的所有特征图,第4个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_4 ;第6个神经网络块的输入端接收 X_4 中的所有特征图,第6个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_6 ;第5个下采样块的输入端接收 P_6 中的所有特征图,第5个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_5 ;第7个神经网络块的输入端接收 X_5 中的所有特征图,第7个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_7 ;第6个下采样块的输入端接收 P_7 中的所有特征图,第6个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_6 ;第8个神经网络块的输入端接收 X_6 中的所有特征图,第8个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_8 ;

[0012] 对于特征融合模块,第9个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第9个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_9 ;第10个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第10个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{10} ;对 P_9 中的所有特征图和 P_{10} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 E_1 ;第11个神经网络块的输入端接收 E_1 中的所有特征图,第11个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{11} ;对 P_1 中的所有特征图、 P_5 中的所有特征图和 P_{11} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 E_2 ;第1个最大池化层的输入端接收 E_2 中的所有特征图,第1个最大池化层的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 Z_1 ;第12个神经网络块的输入端接收 Z_1 中的所

有特征图,第12个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_{12} ;对 P_2 中的所有特征图、 P_6 中的所有特征图和 P_{12} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 E_3 ;第2个最大池化层的输入端接收 E_3 中的所有特征图,第2个最大池化层的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 Z_2 ;第13个神经网络块的输入端接收 Z_2 中的所有特征图,第13个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_{13} ;对 P_3 中的所有特征图、 P_7 中的所有特征图和 P_{13} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 E_4 ;第3个最大池化层的输入端接收 E_4 中的所有特征图,第3个最大池化层的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 Z_3 ;第14个神经网络块的输入端接收 Z_3 中的所有特征图,第14个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_{14} ;对 P_4 中的所有特征图、 P_8 中的所有特征图和 P_{14} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 E_5 ;第4个最大池化层的输入端接收 E_5 中的所有特征图,第4个最大池化层的输出端输出512幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为 Z_4 ;第15个神经网络块的输入端接收 Z_4 中的所有特征图,第15个神经网络块的输出端输出1024幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为 P_{15} ;

[0013] 对于解码框架,第1个上采样层的输入端接收 P_{15} 中的所有特征图,第1个上采样层的输出端输出1024幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 S_1 ;第16个神经网络块的输入端接收 S_1 中的所有特征图,第16个神经网络块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_{16} ;第2个上

采样层的输入端接收 P_{16} 中的所有特征图,第2个上采样层的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 S_2 ;第17个神经网络块的输入端接收 S_2 中的所有特征图,第17个神经网络块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_{17} ;第3个上采样层的输入端接收 P_{17} 中的所有特征图,第3个上采样层的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 S_3 ;第18个神经网络块的输入端接收 S_3 中的所有特征图,第18个神经网络块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_{18} ;第4个上采样层的输入端接收 P_{18} 中的所有特征图,第4个上采样层的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 S_4 ;第19个神经网络块的输入端接收 S_4 中的所有特征图,第19个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{19} ;

[0014] 对于输出层,第一卷积层的输入端接收 P_{19} 中的所有特征图,第一卷积层的输出端输出一幅宽度为 W 且高度为 H 的特征图;第一批标准化层的输入端接收第一卷积层的输出端输出的特征图;第一激活层的输入端接收第一批标准化层的输出端输出的特征图;第一激活层的输出端输出一幅训练用左视点图像对应的立体图像的显著性图像;其中,显著性图像的宽度为 W 且高度为 H ;

[0015] 步骤1_3:将训练集中的每幅原始的立体图像的左视点图像作为训练用左视点图像,并将训练集中的每幅原始的立体图像的深度图像作为训练用深度图像,输入到卷积神经网络中进行训练,得到训练集中的每幅原始的立体图像的显著性图像,将 $\{I^n(x,y)\}$ 的显著性图像记为 $\{G_{pre}^n(x,y)\}$;其中, $G_{pre}^n(x,y)$ 表示 $\{G_{pre}^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值;

[0016] 步骤1_4:计算训练集中的每幅原始的立体图像的显著性图像与真实人眼注视图像之间的损失函数值,将 $\{G_{pre}^n(x,y)\}$ 与 $\{G_{real}^n(x,y)\}$ 之间的损失函数值记为 $Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$, $Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$ 采用均方误差损失函数获得;

[0017] 步骤1_5:重复执行步骤1_3和步骤1_4共 V 次,得到卷积神经网络训练模型,并共得到 $N \times V$ 个损失函数值;然后从 $N \times V$ 个损失函数值中找出值最小的损失函数值;接着将值最小的损失函数值对应的权值矢量和偏置项对应作为卷积神经网络训练模型的最优权值矢量和最优偏置项,对应记为 w^{best} 和 b^{best} ;其中, $V > 1$;

[0018] 所述的测试阶段过程的具体步骤为:

[0019] 步骤2_1:令 $\{\hat{I}(x',y')\}$ 表示待测试的宽度为 W' 且高度为 H' 的立体图像,将

$\{\hat{I}(x',y')\}$ 的左视点图像和深度图像对应记为 $\{\hat{L}_{RGB}(x',y')\}$ 和 $\{\hat{D}(x',y')\}$ ；其中， $1 \leq x' \leq W'$ ， $1 \leq y' \leq H'$ ， $\hat{I}(x',y')$ 表示 $\{\hat{I}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值， $\hat{L}_{RGB}(x',y')$ 表示 $\{\hat{L}_{RGB}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值， $\hat{D}(x',y')$ 表示 $\{\hat{D}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值；

[0020] 步骤2_2: 将 $\{\hat{L}_{RGB}(x',y')\}$ 和 $\{\hat{D}(x',y')\}$ 输入到卷积神经网络训练模型中, 并利用 w^{best} 和 b^{best} 进行预测, 得到 $\{\hat{I}(x',y')\}$ 的显著性预测图像, 记为 $\{\hat{G}_{pre}(x',y')\}$ ；其中， $\hat{G}_{pre}(x',y')$ 表示 $\{\hat{G}_{pre}(x',y')\}$ 中坐标位置为 (x',y') 的像素点的像素值。

[0021] 所述的步骤1_2中, 第1个至第8个神经网络块的结构相同, 其由依次设置的第一空洞卷积层、第二批标准化层、第二激活层、第一残差块、第二空洞卷积层、第三批标准化层构成, 第一空洞卷积层的输入端为其所在的神经网络块的输入端, 第二批标准化层的输入端接收第一空洞卷积层的输出端输出的所有特征图, 第二激活层的输入端接收第二批标准化层的输出端输出的所有特征图, 第一残差块的输入端接收第二激活层的输出端输出的所有特征图, 第二空洞卷积层的输入端接收第一残差块的输出端输出的所有特征图, 第三批标准化层的输入端接收第二空洞卷积层的输出端输出的所有特征图, 第三批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第1个和第5个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为64, 第2个和第6个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为128, 第3个和第7个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为256, 第4个和第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为512, 第1个至第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核大小均为 3×3 、步幅均为1、空洞均为2、填充均为2, 第1个至第8个神经网络块各自中的第二激活层的激活方式均为“ReLU”；

[0022] 第9个和第10个神经网络块的结构相同, 其由依次设置的第二卷积层和第四批标准化层构成, 第二卷积层的输入端为其所在的神经网络块的输入端, 第四批标准化层的输入端接收第二卷积层的输出端输出的所有特征图, 第四批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第9个和第10个神经网络块各自中的第二卷积层的卷积核个数均为3、卷积核大小均为 7×7 、步幅均为1、填充均为3；

[0023] 第11个和第12个神经网络块的结构相同, 其由依次设置的第三卷积层、第五批标准化层、第三激活层、第四卷积层、第六批标准化层构成, 第三卷积层的输入端为其所在的神经网络块的输入端, 第五批标准化层的输入端接收第三卷积层的输出端输出的所有特征图, 第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图, 第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图, 第四卷积层的输入端接收第三激活层的输出端输出的所有特征图, 第六批标准化层的输入端接收第四卷积层的输出端输出的所有特征图, 第六批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第11个神经网络块中的第三卷积层和第四卷积层的卷积核个数为64, 第12个神经网络块中的第三卷

积层和第四卷积层的卷积核个数为128,第11个和第12个神经网络块各自中的第三卷积层和第四卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1;第11个和第12个神经网络块各自中的第三激活层的激活方式为“ReLU”;

[0024] 第13个至第19个神经网络块的结构相同,其由依次设置的第五卷积层、第七批标准化层、第四激活层、第六卷积层、第八批标准化层、第五激活层、第七卷积层、第九批标准化层构成,第五卷积层的输入端为其所在的神经网络块的输入端,第七批标准化层的输入端接收第五卷积层的输出端输出的所有特征图,第四激活层的输入端接收第七批标准化层的输出端输出的所有特征图,第六卷积层的输入端接收第四激活层的输出端输出的所有特征图,第八批标准化层的输入端接收第六卷积层的输出端输出的所有特征图,第五激活层的输入端接收第八批标准化层的输出端输出的所有特征图,第七卷积层的输入端接收第五激活层的输出端输出的所有特征图,第九批标准化层的输入端接收第七卷积层的输出端输出的所有特征图,第九批标准化层的输出端为其所在的神经网络块的输出端;其中,第13个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为256,第14个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为512,第15个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为1024,第16个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为512、512、256,第17个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为256、256、128,第18个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为128、128、64,第19个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为64,第13个至第19个神经网络块各自中的第五卷积层、第六卷积层和第七卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1,第13个至第19个神经网络块各自中的第四激活层和第五激活层的激活方式均为“ReLU”。

[0025] 所述的步骤1_2中,第1个至第6个下采样块的结构相同,其由第二残差块构成,第二残差块的输入端为其所在的下采样块的输入端,第二残差块的输出端为其所在的下采样块的输出端。

[0026] 所述的第一残差块和所述的第二残差块的结构相同,其包括3个卷积层、3个批标准化层和3个激活层,第1个卷积层的输入端为其所在的残差块的输入端,第1个批标准化层的输入端接收第1个卷积层的输出端输出的所有特征图,第1个激活层的输入端接收第1个批标准化层的输出端输出的所有特征图,第2个卷积层的输入端接收第1个激活层的输出端输出的所有特征图,第2个批标准化层的输入端接收第2个卷积层的输出端输出的所有特征图,第2个激活层的输入端接收第2个批标准化层的输出端输出的所有特征图,第3个卷积层的输入端接收第2个激活层的输出端输出的所有特征图,第3个批标准化层的输入端接收第3个卷积层的输出端输出的所有特征图,第1个卷积层的输入端接收的所有特征图和第3个批标准化层的输出端输出的所有特征图相加,再经过第3个激活层后第3个激活层的输出端输出的所有特征图作为所在的残差块的输出端输出的所有特征图;其中,第1个和第5个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为64,第2个和第6个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为128,第3个和第7个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为256,第4个和第8个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为512,第1个至第8个神经网络块各自中的第

一残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第8个神经网络块各自中的第一残差块中的第2个卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1,第1个和第4个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为64,第2个和第5个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为128,第3个和第6个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为256,第1个至第6个下采样块各自中的第二残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第6个下采样块各自中的第二残差块中的第2个卷积层的卷积核大小均为 3×3 、步幅均为2、填充均为1,3个激活层的激活方式均为“ReLU”。

[0027] 所述的步骤1_2中,第1个至第4个最大池化层的池化窗口的大小均为 2×2 、步幅均为2。

[0028] 所述的步骤1_2中,第1个至第4个上采样层的采样模式均为双线性插值、比例因子均为2。

[0029] 与现有技术相比,本发明的优点在于:

[0030] 1) 本发明方法通过构建的卷积神经网络中提出的编码框架对RGB图像和深度图像分别训练一个模块(即RGB特征提取模块和深度特征提取模块)来学习不同级别的RGB和深度特征,而且提出了一个专门融合RGB和深度特征的模块,即特征融合模块,从低级到高级融合这两种特征,有利于充分利用跨模态信息,形成新的判别特征,提高了立体视觉显著性预测的准确性。

[0031] 2) 本发明方法构建的卷积神经网络中的RGB特征提取模块和深度特征提取模块中的下采样块利用步幅为2的残差块来代替以往工作中所用的最大池化层,有利于模型自适应地选取特征信息,避免了由于最大池化操作而丢失重要的信息。

[0032] 3) 本发明方法构建的卷积神经网络中的RGB特征提取模块和深度特征提取模块中引入了前后带有空洞卷积层的残差块,扩大了卷积核的接受域,有利于构建的卷积神经网络更关注全局信息,学到更为丰富的内容。

附图说明

[0033] 图1为本发明方法构建的卷积神经网络的组成示意图。

具体实施方式

[0034] 以下结合附图实施例对本发明作进一步详细描述。

[0035] 本发明提出的一种基于卷积神经网络的立体图像视觉显著性检测方法,其包括训练阶段和测试阶段两个过程。

[0036] 所述的训练阶段过程的具体步骤为:

[0037] 步骤1_1:选取N幅宽度为W且高度为H的原始的立体图像;然后将选取的所有原始的立体图像及所有原始的立体图像各自的左视点图像、深度图像和真实人眼注视图像构成训练集,将训练集中的第n幅原始的立体图像记为 $\{I^n(x, y)\}$,将 $\{I^n(x, y)\}$ 的左视点图像、深度图像和真实人眼注视图像对应记为 $\{L_{RGB}^n(x, y)\}$ 、 $\{D^n(x, y)\}$ 、 $\{G_{real}^n(x, y)\}$;其中,N为正整数, $N \geq 300$,如取 $N = 600$,W和H均能够被2整除,n为正整数,n的初始值为1, $1 \leq n \leq N$, $1 \leq x \leq$

$W, 1 \leq y \leq H, I^n(x, y)$ 表示 $\{I^n(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值, $L_{RGB}^n(x, y)$ 表示 $\{L_{RGB}^n(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值, $D^n(x, y)$ 表示 $\{D^n(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值, $G_{real}^n(x, y)$ 表示 $\{G_{real}^n(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值。

[0038] 步骤1_2:构建卷积神经网络:如图1所示,该卷积神经网络包含输入层、隐层、输出层,输入层包括RGB图输入层和深度图输入层,隐层包括编码框架和解码框架,编码框架由RGB特征提取模块、深度特征提取模块和特征融合模块三部分组成,RGB特征提取模块由第1个至第4个神经网络块、第1个至第3个下采样块组成,深度特征提取模块由第5个至第8个神经网络块、第4个至第6个下采样块组成,特征融合模块由第9个至第15个神经网络块、第1个至第4个最大池化层组成,解码框架由第16个至第19个神经网络块、第1个至第4个上采样层组成;输出层由第一卷积层、第一批标准化层和第一激活层组成,第一卷积层的卷积核大小为 3×3 、步幅大小为1、卷积核个数为1、填充为1,第一激活层的激活方式为“Sigmoid”。

[0039] 对于RGB图输入层,其输入端接收一幅训练用左视点图像,其输出端输出训练用左视点图像给隐层;其中,要求训练用左视点图像的宽度为 W 且高度为 H 。

[0040] 对于深度图输入层,其输入端接收RGB图输入层的输入端接收的训练用左视点图像对应的训练用深度图像,其输出端输出训练用深度图像给隐层;其中,训练用深度图像的宽度为 W 且高度为 H 。

[0041] 对于RGB特征提取模块,第1个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第1个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_1 ;第1个下采样块的输入端接收 P_1 中的所有特征图,第1个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_1 ;第2个神经网络块的输入端接收 X_1 中的所有特征图,第2个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_2 ;第2个下采样块的输入端接收 P_2 中的所有特征图,第2个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_2 ;第3个神经网络块的输入端接收 X_2 中的所有特征图,第3个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_3 ;第3个下采样块的输入端接收 P_3 中的所有特征图,第3个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_3 ;第4个神经网络块的输入端接收 X_3 中的所有特征图,第4个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的

集合记为 P_4 。

[0042] 对于深度特征提取模块,第5个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第5个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_5 ;第4个下采样块的输入端接收 P_5 中的所有特征图,第4个下采样块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 X_4 ;第6个神经网络块的输入端接收 X_4 中的所有特征图,第6个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_6 ;第5个下采样块的输入端接收 P_6 中的所有特征图,第5个下采样块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 X_5 ;第7个神经网络块的输入端接收 X_5 中的所有特征图,第7个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_7 ;第6个下采样块的输入端接收 P_7 中的所有特征图,第6个下采样块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 X_6 ;第8个神经网络块的输入端接收 X_6 中的所有特征图,第8个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_8 。

[0043] 对于特征融合模块,第9个神经网络块的输入端接收RGB图输入层的输出端输出的训练用左视点图像,第9个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_9 ;第10个神经网络块的输入端接收深度图输入层的输出端输出的训练用深度图像,第10个神经网络块的输出端输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{10} ;对 P_9 中的所有特征图和 P_{10} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出3幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 E_1 ;第11个神经网络块的输入端接收 E_1 中的所有特征图,第11个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{11} ;对 P_1 中的所有特征图、 P_5 中的所有特征图和 P_{11} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 E_2 ;第1个最大池化层的输入端接收 E_2 中的所有特征图,第1个最大池化层的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 Z_1 ;第12个神经网络块的输入端接收 Z_1 中的所有特征图,第12个神经网络块的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_{12} ;对 P_2 中的所有特征图、 P_6 中的所有特征图和 P_{12} 中的所有

特征图进行Element-wise Summation操作,Element-wise Summation操作后输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 E_3 ;第2个最大池化层的输入端接收 E_3 中的所有特征图,第2个最大池化层的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 Z_2 ;第13个神经网络块的输入端接收 Z_2 中的所有特征图,第13个神经网络块的输出端输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_{13} ;对 P_3 中的所有特征图、 P_7 中的所有特征图和 P_{13} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出256幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 E_4 ;第3个最大池化层的输入端接收 E_4 中的所有特征图,第3个最大池化层的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 Z_3 ;第14个神经网络块的输入端接收 Z_3 中的所有特征图,第14个神经网络块的输出端输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_{14} ;对 P_4 中的所有特征图、 P_8 中的所有特征图和 P_{14} 中的所有特征图进行Element-wise Summation操作,Element-wise Summation操作后输出512幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 E_5 ;第4个最大池化层的输入端接收 E_5 中的所有特征图,第4个最大池化层的输出端输出512幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为 Z_4 ;第15个神经网络块的输入端接收 Z_4 中的所有特征图,第15个神经网络块的输出端输出1024幅宽度为 $\frac{W}{16}$ 且高度为 $\frac{H}{16}$ 的特征图,将输出的所有特征图构成的集合记为 P_{15} 。

[0044] 对于解码框架,第1个上采样层的输入端接收 P_{15} 中的所有特征图,第1个上采样层的输出端输出1024幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 S_1 ;第16个神经网络块的输入端接收 S_1 中的所有特征图,第16个神经网络块的输出端输出256幅宽度为 $\frac{W}{8}$ 且高度为 $\frac{H}{8}$ 的特征图,将输出的所有特征图构成的集合记为 P_{16} ;第2个上采样层的输入端接收 P_{16} 中的所有特征图,第2个上采样层的输出端输出256幅宽度为 $\frac{W}{4}$ 且

高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 S_2 ;第17个神经网络块的输入端接收 S_2 中的所有特征图,第17个神经网络块的输出端输出128幅宽度为 $\frac{W}{4}$ 且高度为 $\frac{H}{4}$ 的特征图,将输出的所有特征图构成的集合记为 P_{17} ;第3个上采样层的输入端接收 P_{17} 中的所有特征图,第3个上采样层的输出端输出128幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 S_3 ;第18个神经网络块的输入端接收 S_3 中的所有特征图,第18个神经网络块的输出端输出64幅宽度为 $\frac{W}{2}$ 且高度为 $\frac{H}{2}$ 的特征图,将输出的所有特征图构成的集合记为 P_{18} ;第4个上采样层的输入端接收 P_{18} 中的所有特征图,第4个上采样层的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 S_4 ;第19个神经网络块的输入端接收 S_4 中的所有特征图,第19个神经网络块的输出端输出64幅宽度为 W 且高度为 H 的特征图,将输出的所有特征图构成的集合记为 P_{19} 。

[0045] 对于输出层,第一卷积层的输入端接收 P_{19} 中的所有特征图,第一卷积层的输出端输出一幅宽度为 W 且高度为 H 的特征图;第一批标准化层的输入端接收第一卷积层的输出端输出的特征图;第一激活层的输入端接收第一批标准化层的输出端输出的特征图;第一激活层的输出端输出一幅训练用左视点图像对应的立体图像的显著性图像;其中,显著性图像的宽度为 W 且高度为 H 。

[0046] 步骤1_3:将训练集中的每幅原始的立体图像的左视点图像作为训练用左视点图像,并将训练集中的每幅原始的立体图像的深度图像作为训练用深度图像,输入到卷积神经网络中进行训练,得到训练集中的每幅原始的立体图像的显著性图像,将 $\{I^n(x,y)\}$ 的显著性图像记为 $\{G_{pre}^n(x,y)\}$;其中, $G_{pre}^n(x,y)$ 表示 $\{G_{pre}^n(x,y)\}$ 中坐标位置为 (x,y) 的像素点的像素值。

[0047] 步骤1_4:计算训练集中的每幅原始的立体图像的显著性图像与真实人眼注视图像之间的损失函数值,将 $\{G_{pre}^n(x,y)\}$ 与 $\{G_{real}^n(x,y)\}$ 之间的损失函数值记为 $Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$, $Loss(\{G_{pre}^n(x,y)\},\{G_{real}^n(x,y)\})$ 采用均方误差损失函数获得。

[0048] 步骤1_5:重复执行步骤1_3和步骤1_4共 V 次,得到卷积神经网络训练模型,并共得到 $N \times V$ 个损失函数值;然后从 $N \times V$ 个损失函数值中找出值最小的损失函数值;接着将值最小的损失函数值对应的权值矢量和偏置项对应作为卷积神经网络训练模型的最优权值矢量和最优偏置项,对应记为 w^{best} 和 b^{best} ;其中, $V > 1$,如取 $V = 50$ 。

[0049] 所述的测试阶段过程的具体步骤为:

[0050] 步骤2_1:令 $\{\hat{I}(x',y')\}$ 表示待测试的宽度为 W' 且高度为 H' 的立体图像,将 $\{\hat{I}(x',y')\}$ 的左视点图像和深度图像对应记为 $\{\hat{L}_{RGB}(x',y')\}$ 和 $\{\hat{D}(x',y')\}$;其中, $1 \leq x' \leq$

$W', 1 \leq y' \leq H'$, $\hat{I}(x', y')$ 表示 $\{\hat{I}(x', y')\}$ 中坐标位置为 (x', y') 的像素点的像素值, $\hat{L}_{RGB}(x', y')$ 表示 $\{\hat{L}_{RGB}(x', y')\}$ 中坐标位置为 (x', y') 的像素点的像素值, $\hat{D}(x', y')$ 表示 $\{\hat{D}(x', y')\}$ 中坐标位置为 (x', y') 的像素点的像素值。

[0051] 步骤2_2: 将 $\{\hat{L}_{RGB}(x', y')\}$ 和 $\{\hat{D}(x', y')\}$ 输入到卷积神经网络训练模型中, 并利用 w^{best} 和 b^{best} 进行预测, 得到 $\{\hat{I}(x', y')\}$ 的显著性预测图像, 记为 $\{\hat{G}_{pre}(x', y')\}$; 其中, $\hat{G}_{pre}(x', y')$ 表示 $\{\hat{G}_{pre}(x', y')\}$ 中坐标位置为 (x', y') 的像素点的像素值。

[0052] 在此具体实施例中, 步骤1_2中, 第1个至第8个神经网络块的结构相同, 其由依次设置的第一空洞卷积层、第二批标准化层、第二激活层、第一残差块、第二空洞卷积层、第三批标准化层构成, 第一空洞卷积层的输入端为其所在的神经网络块的输入端, 第二批标准化层的输入端接收第一空洞卷积层的输出端输出的所有特征图, 第二激活层的输入端接收第二批标准化层的输出端输出的所有特征图, 第一残差块的输入端接收第二激活层的输出端输出的所有特征图, 第二空洞卷积层的输入端接收第一残差块的输出端输出的所有特征图, 第三批标准化层的输入端接收第二空洞卷积层的输出端输出的所有特征图, 第三批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第1个和第5个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为64, 第2个和第6个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为128, 第3个和第7个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为256, 第4个和第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核个数为512, 第1个至第8个神经网络块各自中的第一空洞卷积层和第二空洞卷积层的卷积核大小均为 3×3 、步幅均为1、空洞均为2、填充均为2, 第1个至第8个神经网络块各自中的第二激活层的激活方式均为“ReLU”。

[0053] 第9个和第10个神经网络块的结构相同, 其由依次设置的第二卷积层和第四批标准化层构成, 第二卷积层的输入端为其所在的神经网络块的输入端, 第四批标准化层的输入端接收第二卷积层的输出端输出的所有特征图, 第四批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第9个和第10个神经网络块各自中的第二卷积层的卷积核个数均为3、卷积核大小均为 7×7 、步幅均为1、填充均为3。

[0054] 第11个和第12个神经网络块的结构相同, 其由依次设置的第三卷积层、第五批标准化层、第三激活层、第四卷积层、第六批标准化层构成, 第三卷积层的输入端为其所在的神经网络块的输入端, 第五批标准化层的输入端接收第三卷积层的输出端输出的所有特征图, 第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图, 第三激活层的输入端接收第五批标准化层的输出端输出的所有特征图, 第四卷积层的输入端接收第三激活层的输出端输出的所有特征图, 第六批标准化层的输入端接收第四卷积层的输出端输出的所有特征图, 第六批标准化层的输出端为其所在的神经网络块的输出端; 其中, 第11个神经网络块中的第三卷积层和第四卷积层的卷积核个数为64, 第12个神经网络块中的第三卷积层和第四卷积层的卷积核个数为128, 第11个和第12个神经网络块各自中的第三卷积层和第四卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1; 第11个和第12个神经网络块

各自中的第三激活层的激活方式为“ReLU”。

[0055] 第13个至第19个神经网络块的结构相同,其由依次设置的第五卷积层、第七批标准化层、第四激活层、第六卷积层、第八批标准化层、第五激活层、第七卷积层、第九批标准化层构成,第五卷积层的输入端为其所在的神经网络块的输入端,第七批标准化层的输入端接收第五卷积层的输出端输出的所有特征图,第四激活层的输入端接收第七批标准化层的输出端输出的所有特征图,第六卷积层的输入端接收第四激活层的输出端输出的所有特征图,第八批标准化层的输入端接收第六卷积层的输出端输出的所有特征图,第五激活层的输入端接收第八批标准化层的输出端输出的所有特征图,第七卷积层的输入端接收第五激活层的输出端输出的所有特征图,第九批标准化层的输入端接收第七卷积层的输出端输出的所有特征图,第九批标准化层的输出端为其所在的神经网络块的输出端;其中,第13个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为256,第14个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为512,第15个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为1024,第16个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为512、512、256,第17个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为256、256、128,第18个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数对应为128、128、64,第19个神经网络块中的第五卷积层、第六卷积层和第七卷积层的卷积核个数为64,第13个至第19个神经网络块各自中的第五卷积层、第六卷积层和第七卷积层的卷积核大小均为 3×3 、步幅均为1、填充均为1,第13个至第19个神经网络块各自中的第四激活层和第五激活层的激活方式均为“ReLU”。

[0056] 在此具体实施例中,步骤1_2中,第1个至第6个下采样块的结构相同,其由第二残差块构成,第二残差块的输入端为其所在的下采样块的输入端,第二残差块的输出端为其所在的下采样块的输出端。

[0057] 在此具体实施例中,第一残差块和所述的第二残差块的结构相同,其包括3个卷积层、3个批标准化层和3个激活层,第1个卷积层的输入端为其所在的残差块的输入端,第1个批标准化层的输入端接收第1个卷积层的输出端输出的所有特征图,第1个激活层的输入端接收第1个批标准化层的输出端输出的所有特征图,第2个卷积层的输入端接收第1个激活层的输出端输出的所有特征图,第2个批标准化层的输入端接收第2个卷积层的输出端输出的所有特征图,第2个激活层的输入端接收第2个批标准化层的输出端输出的所有特征图,第3个卷积层的输入端接收第2个激活层的输出端输出的所有特征图,第3个批标准化层的输入端接收第3个卷积层的输出端输出的所有特征图,第1个卷积层的输入端接收的所有特征图和第3个批标准化层的输出端输出的所有特征图相加,再经过第3个激活层后第3个激活层的输出端输出的所有特征图作为所在的残差块的输出端输出的所有特征图;其中,第1个和第5个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为64,第2个和第6个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为128,第3个和第7个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为256,第4个和第8个神经网络块各自中的第一残差块中的各个卷积层的卷积核个数为512,第1个至第8个神经网络块各自中的第一残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第8个神经网络块各自中的第一残差块中的第2个卷积层的卷积核大小均为 3×3 、

步幅均为1、填充均为1,第1个和第4个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为64,第2个和第5个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为128,第3个和第6个下采样块各自中的第二残差块中的各个卷积层的卷积核个数为256,第1个至第6个下采样块各自中的第二残差块中的第1个卷积层和第3个卷积层的卷积核大小均为 1×1 、步幅均为1,第1个至第6个下采样块各自中的第二残差块中的第2个卷积层的卷积核大小均为 3×3 、步幅均为2、填充均为1,3个激活层的激活方式均为“ReLU”。

[0058] 在此具体实施例中,步骤1_2中,第1个至第4个最大池化层的池化窗口的大小均为 2×2 、步幅均为2。

[0059] 于此具体实施例中,步骤1_2中,第1个至第4个上采样层的采样模式均为双线性插值、比例因子均为2。

[0060] 为了验证本发明方法的可行性和有效性,进行实验。

[0061] 在此,采用中国台湾交通大学提供的三维人眼跟踪数据库(NCTU-3DFixation)来分析本发明方法的准确性和稳定性。这里,利用评估视觉显著提取方法的4个常用客观参量作为评价指标,即线性相关系数(Linear Correlation Coefficient,CC)、Kullback-Leibler散度系数(Kullback-Leibler Divergence,KLD)、AUC参数(the Area Under the receiver operating characteristics Curve,AUC),标准化扫描路径显著性(Normalized Scanpath Saliency,NSS)。

[0062] 利用本发明方法获取中国台湾交通大学提供的三维人眼跟踪数据库中的每幅立体图像的显著性预测图像,并与三维人眼跟踪数据库中的每幅立体图像的主观视觉显著图即真实人眼注视图像(三维人眼跟踪数据库中存在)进行比较,CC、AUC和NSS值越高、KLD值越低说明本发明方法获得的显著性预测图像与主观视觉显著图的一致性越好。反映本发明方法的显著提取性能的CC、KLD、AUC和NSS相关指标如表1所列。

[0063] 表1利用本发明方法得到的显著性预测图像与主观视觉显著图的准确性和稳定性

[0064]

性能指标	CC	KLD	AUC(Borji)	NSS
性能指标值	0.7583	0.4868	0.8789	2.0692

[0065] 从表1所列的数据可知,按本发明方法得到的显著性预测图像与主观视觉显著图的准确性和稳定性是很好的,表明客观检测结果与人眼主观感知的结果较为一致,足以说明本发明方法的可行性和有效性。

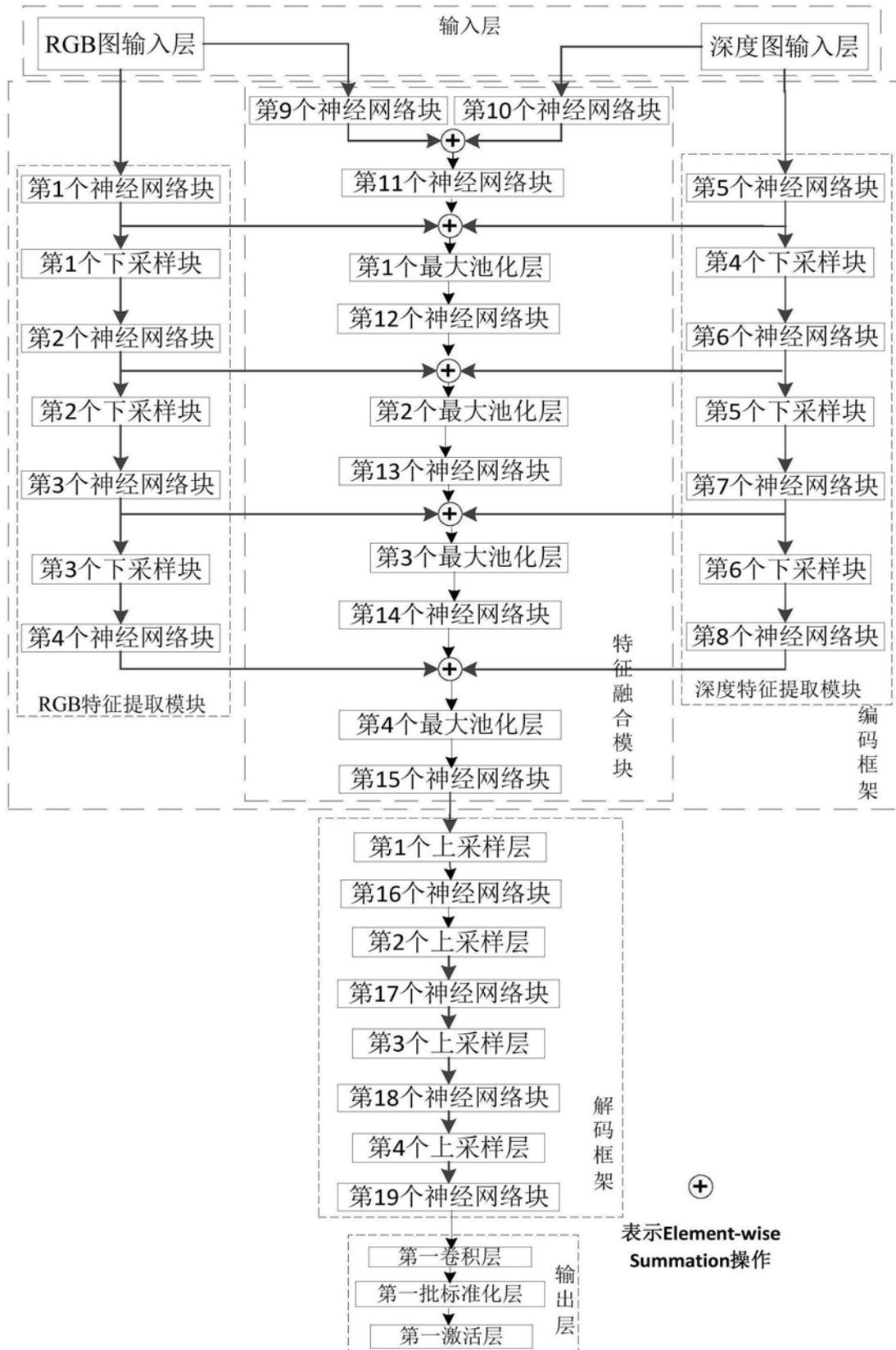


图1