基于全卷积神经网络的多目标显著性检测

翟正利,孙 霞,周 炜,梁振明
(青岛理工大学 信息与控制工程学院,山东 青岛 266520)

摘 要:随着图像数据的爆炸性增长,图像处理变得越来越重要。显著性目标检测是图像处理的重要研究方向之一,目前已采用多种研究方法进行显著性目标检测,但是传统的显著性检测方法所使用的低级特征对于复杂场景并不健壮。全卷积神经网络在图像处理中表现出良好的性能,但存在目标显著性检测边界模糊等不足。为解决边界模糊等问题,该模型采用了一种具有跳跃连接的全卷积神经网络,以及5个不同膨胀率的空洞卷积按照一定规则组成的ESP模块,在全卷积神经网络的基础上采用ESP模块和不同的跳跃连接方式,以获取更多的低级特征来精确多目标显著对象的边界。实验中运用MIT Scene Parsing数据集训练和测试模型,通过与相关模型在精度和MIOU上的比较结果表明,在保证模型的处理时间未增加的同时,经过改进的全卷积神经网络的检测具有更高的准确度以及更精确的边界信息。

关键词:目标检测;空洞卷积;低级特征;全卷积神经网络;跳跃连接
中图分类号:TP302.7 **文献标识码**:A **文章编号**:1673-629X(2020)08-0034-06
doi:10.3969/j.issn.1673-629X.2020.08.006

Multi-objective Saliency Detection Based on Full Convolution Neural Network

ZHAI Zheng-li, SUN Xia, ZHOU Wei, LIANG Zhen-ming
(School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China)

Abstract: With the explosive growth of image data, image processing becomes more and more important. Saliency target detection is one of the important research directions in image processing. At present, many research methods have been used to detect saliency targets, but the low-level features used in traditional saliency detection methods are not robust for complex scenes. Fully convolutional network has excellent performance in image processing, but it has some shortcomings such as ambiguous edge of target saliency detection. In order to solve the problem of boundary fuzzy, a fully convolutional network with jump connection and ESP module composed of five hollow convolutions with different expansion rates according to certain rules are adopted in this model. On the basis of the fully convolutional network, the ESP module and different jump connection methods are used to obtain more low-level features to accurately define the boundary of multi-objective significant objects. In the experiment, we use the data set of MIT Scene Parsing to train and test the model. Compared with the related models in accuracy and MIOU, the results show that the improved fully convolutional network has higher accuracy and more accurate boundary information while ensuring that the processing time of the model does not increase.

Key words: object detection; void convolution; low-level feature; fully convolutional neural network; jump connection

0 引 言

显著性目标检测通过人类视觉和认知系统识别图像中最重要和最显著的对象区域,目前许多计算机视觉研究人员提出许多用于模拟人类视觉注意过程或识别显著对象的计算模型[1]。多目标检测可以帮助盲人知道避开什么样的障碍物,有助于他们提前做好心理准备。

在算法策略方面,显著性目标检测方法可以分为两类,一个是自上而下的数据驱动方法;另一个是自上而下的任务驱动方法。对于大多数自下而上的方法,使用低级特征来计算显著特征值,例如颜色、纹理和对比度。Itti等人[2]提出了中心-圆形方案,使用多尺度显著图的线性和非线性组合方法来提取颜色、强度、方向和纹理等低级特征[3]。然而这些方法仍然存在一些需要解决的问题,比如大多数全局对比度方法[4]在进行全局统计时依赖于颜色唯一性。Fareed等人[5]提出

收稿日期:2019-09-03 **修回日期**:2020-01-08
基金项目:国家自然科学基金(61502262)
作者简介:翟正利(1972-),男,博士(后),研究方向为深度学习、网络信息服务、Petri网理论与应用、智能信息处理等;通信作者:孙 霞(1994-),女,硕士研究生,研究方向为深度学习、模式识别。

了一种基于稀疏表示和图形排序的显著区域检测算法,它结合了高斯和贝叶斯过程,生成平滑精确的显著图。以上方法存在两个主要的问题:首先,这些方法主要依赖于手工制作的低级特征,这些特征可能无法用于目标与背景相似的图像以及具有复杂背景的图像;其次,这些方法可能不足以捕获自然图像中的语义特征,并且它们通常会忽略图像中不同部分之间的空间关系。

随着深度网络的迅速发展,已有研究人员将神经网络应用于计算机视觉中并取得成功。包括图像分类[6]、物体检测[7]、语义分析[8]等。利用深度神经网络可以直接通过网络训练模型获得物体特征以及相互之间的关系。Li 等人[9]提出了一种基于深度卷积神经网络的由粗到细显著性目标检测,首先利用完全卷积网络产生粗略显著图,然后将得到的粗显著图与原始灰度图加黑色图作为基于超像素 CNN 模型的输入,进行精确边界细化检测。Zhu D 等人[10]提出了一种基于深度残差网络的局部和全局显著性目标检测,该方法分别利用深度残差网络来提取显著目标的局部特征和全局特征,通过贝叶斯结合来获得最终显著性图。Cao F 等人[11]使用具有特征选择的卷积神经网络进行高效的显著性检测,他们提出了一种特征图选择方法用于选择有用的特征图并丢弃其余部分。为了解决显著性目标的边界不清晰问题,文中提出了一种具有跳跃连接的全卷积神经网络,通过结合低级特征来增强边界信息。实验表明基于跳跃连接的全卷积神经网络的目标检测方法明显优于之前的方法,在增强特征提取的同时还增强边界信息。

1　全卷积神经网络 FCN

近年来卷积神经网络一直在驱动着图像识别领域的进步。无论是整张图像的分类,还是物体检测等都在 CNN 的帮助下取得非常大的发展。语义分割是计算机视觉中的基本任务,也是计算机视觉的热点,在语义分割中需要将视觉输入分为不同的语义可解释类别。传统用卷积神经网络进行语义分割的方法是将像素周围的一个小区域作为卷积神经网络的输入,然后做训练和预测。但在用卷积神经网络进行语义分割的过程也存在一些需要进一步改进的地方,如下:

(1)占用内存比较大;

(2)存在过多的重复计算,导致计算效率比较低;

(3)区域大小如何确定,同时也限制了感受野的大小。

为了解决这些问题,Long[12]提出了全卷积网络 FCN 用于语义分割,FCN 能够对图像进行像素级的端到端的分类检测。与传统的 CNN 在卷积层之后使用全连接层得到固定长度的特征向量进行分类不同,FCN 可以接受任意尺寸的输入图像,并对最后一层卷积后得到的特征图采用反卷积层进行上采样操作,得到与原始输入图像相同尺寸大小的特征图,从而可以进行像素级的预测,同时还保留了原始输入图像中的空间信息进而解决了语义分割问题。

FCN 主要由三部分组成,分别是卷积层、池化层和上采样层。

1.1　卷积层

在卷积层的前向传播过程,输入特征映射 X 与具有偏置 b 的可学习内核 W 卷积,并且通过激活函数 Relu,形成输出特征映射。为了更加详细地理解卷积过程,首先对图像的每个像素进行编号,用 X_{ij} 表示图像的第 i 行第 j 列元素;对 filter 的每个权重进行编号,用 W_{mn} 表示第 m 行第 n 列权重,用 W_b 表示 filter 的偏置项;对 Feature Map 的每个元素进行编号,用 a_{ij} 表示 Feature Map 的第 i 行第 j 列元素;用 f 表示激活函数。然后,使用下列公式计算卷积:

$$a_{ij} = f\left(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n} x_{i+m,j+n} + w_b\right) \tag{1}$$

FCN 的卷积层以 VGG19 为基础,VGG19 的原始结构有 19 个卷积层,由 5 个池化层隔开,并使用完全连接层来进行分类[13]。19 个卷积层采用 3 × 3 的卷积核。为了使 VGG19 适应完全卷积网络,将全连接层转换为卷积层。第一个完全连接层被转换为具有 7 × 7 内核的卷积层,第二个完全连接层被转换为具有内核大小为 1 × 1 的卷积层,这两个卷积层之后进行 Dropout 操作。第三个完全卷积层被转换为具有内核大小为 1 × 1 的卷积层。如果没有完全连接层,输入图像的大小可以为任意尺寸,使得操作更容易,因此不需要像传统网络那样将输入调整为固定大小。在全卷积神经网络中,多个卷积运算采用的是全零填充方法,以保留图像边界信息,并且更容易控制输出图像的大小[14-15]。与完全连接层相比,卷积层的好处在于实现权重稀疏和权重共享,减少神经网络中的参数量。

1.2　池化层

池化层也称为欠采样或下采样,池化操作主要分为最大池化和平均池化。如图 1 所示,数据被分成相邻非重叠块,每个块由四个元素组成。平均池化操作是计算四个元素的平均值,类似的最大池化的操作是输出四个元素中的最大值。

在全卷积神经网络中,经过卷积层进行卷积之后,还需要对得到的特征图进行进一步处理。处理得到的特征图需要面临计算的挑战,并增加了网络中的参数量,同时也会增加过度拟合的风险。因此,在卷积层之后进行 2 × 2 的最大池化操作[16](即图像中的每个通

道的每个 2 × 2 非重叠区的最大值),并输出作为下一层的输入。对于大小为 ch × h × w 的输入,经过最大池化计算之后,其大小为 ch × h /2 × w /2。最大池化的定义如下:

$$f(x) = \max(0,x) \tag{2}$$

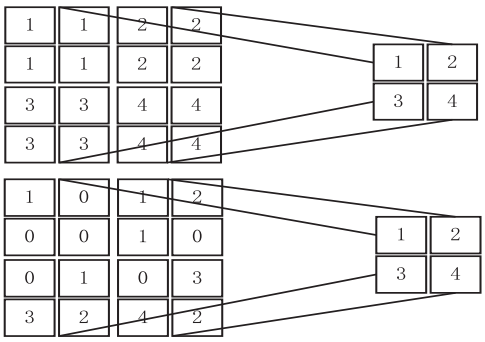

图 1 平均池化和最大池化

1.3 上采样层

上采样,也称为反卷积或者转置卷积,与卷积层相反。在 FCN 全卷积网络中,由于要对图像进行像素级的分割,需要将图像尺寸还原到原来的大小。当反卷积步幅>1 时,需要在中间填充图像。

首先,对卷积层的最后一层进行反卷积,输入为 ESP 模块的最后一层,输出层的形状为池化层第四层,反卷积核为4×4,使得特征图的大小扩大两倍。然后,将得到的结果作为下一次反卷积的输入。第二次反卷积的输出为池化层第三层的形状,反卷积核为 4×4,使得特征图的大小在第一次反卷积之后继续扩大两倍。第三次反卷积核的大小为 16×16,得到与原始图像相同大小的特征图。通过微调预训练 VGG19,可以用很少的训练数据有效地训练模型。

2 改进网络体系

2.1 高效金字塔卷积网络

高效金字塔网络是一种基于新的卷积模块的能有效对高分辨率图片进行语义分割的网络[17],其网络结构如图 2 所示。高效金字塔卷积模块主要包含两部分:第一部分是采用 1 × 1 的逐点卷积,通过逐点卷积操作可以减少计算量;第二部分是扩张卷积的空间金字塔,扩张卷积的空间金字塔首先通过空洞卷积对特征图进行不通膨胀率的进一步提取,以获得更大感受野,最后进行分层特征融合。

ESP 网络是基于 ESP 模块的,ESP 模块的核心是“reduce-split-transform-merge”,过程如下:

(1) Reduce (point-wise convolution):对于 $H \times W \times M$ 的输入,使用 d 个 $1 \times 1 \times M$ 的卷积核,将 M 维的输入特征图降到 d 维,得到的输出是 $H \times W \times d$ 。

(2) Split & Transform (spatial pyramid of dilated convolutions):使用 k 个不同膨胀率的卷积核对第一步的输出特征图并行进行卷积,输出将是 k 个相同大小的特征图。

(3) HFF (hierarchical feature fusion):分层特征融合,将第二步得到的 k 个相同大小的特征图按照一定规则进行拼接,得到最终的输出特征图。

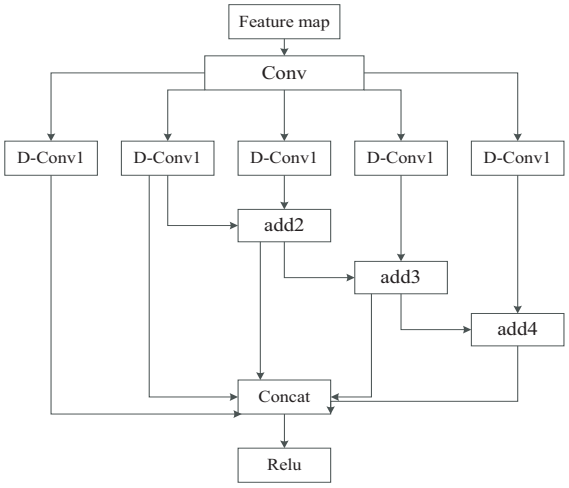

图 2 金字塔网络结构

在 ESP 网络中通过不同膨胀率的空洞卷积进行特征提取不仅获得更大的感受野,而且减少了参数量,因为首先第一步进行的降维使得通道数减少之后,每一个空洞卷积的参数也随之减少。拼合策略与普通空洞卷积特征融合的方法不同,这里为了避免 gridding artifacts 现象,采用了逐级相加的策略。

2.2 网络的改进

全卷积神经网络可以高效且有效地预测显著图,尤其是显著对象与背景类似时。但是显着对象的边界难以进行精确分割,因为池化过程通常忽略局部信息以便拥有更大的感受野。为了实现在拥有更大的感受野的同时更好地捕获对象边界,实验在 FCN 的基础上结合了空洞卷积以及跳跃连接方法。

首先采用 VGG19 的前 5 个卷积块来进行特征提取,然后将完全连接的层转换为 ESP 模块。ESP 模块首先将输入的特征图进行降维,然后使用 k 个不同膨胀率的卷积核对降维后的特征图并行进行卷积,实验中采用 5 个不同膨胀率的卷积核,分别为 1×1、2×2、3×3、5×5、16×16。然后将会得到 5 个相同大小的特征图。最后一步进行分层特征融合 HFF (hierarchical feature fusion),将得到的 5 个相同大小的特征图按照一定的规则进行拼接,得到模块最终的特征图。

目前主要采用的跳跃连接方法有两种,其主要是将池化层的浅层特征图与多层卷积之后的深度特征图进行结合,使其包含更多的边界细节。

(1) 反卷积+连接。

反卷积加连接是最终深度特征图通过反卷积扩大

特征图尺寸,选定的池化层通过反卷积扩大特征图尺寸,使通过反卷积得到的特征图具有相同的尺寸大小,然后把具有相同尺寸的特征图在通道级别上连接在一起。

(2)反卷积+连接+反卷积。

反卷积加连接加反卷积是首先对最后一层卷积层卷积得到的深度特征图进行反卷积,得到与池化层 pool3 相同尺寸大小的特征图,然后在像素级别上对应相加(即进行加和运算),然后进行下一步的反卷积,反卷积之后得到的特征图的形状与池化层 pool4 的相同,进而与池化层 pool4 进行加和运算。然后进行最后一次反卷积,得到与原始图像相同形状的特征图。

文中所采用的跳跃连接方法结合以上两种方法,如图 3 所示。

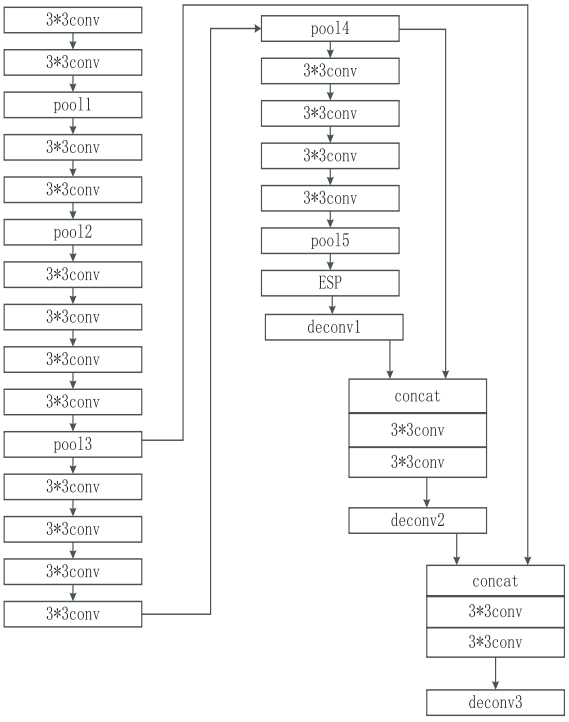

图 3　具有跳跃连接的 FCN 架构

2.3　网络模型内部各层具体设计

在本节中,将介绍网络中出现的各种中间层的设计。改进网络的各层参数如表 1 所示。将尺寸 3×224×224 图像(即具有 RGB3 个通道且边长为 224 个像素的彩色图像)输入到所构造的神经网络中。在计算之后,最终输出图像是 151×224×224(即有 151 个类别和边长为 224 个像素的图像)。通道数与分类数相同,对于每个像素而言,NUM_OF_CLASSESS 个通道中哪个数值最大,这个像素就属于哪个分类。

表 1 中列出了每个层的详细操作:

1. Conv 表示卷积层
2. Relu 表示 Relu 激活函数;
3. Pooling 表示最大池化层;
4. deConv 表示反卷积层;
5. Concat 表示两个图像在通道级别上连接为一个图像;
6. Softmax 表示 Softmax 激活函数。

以下详细介绍这些层及其相应的操作。

3　网络模型训练及测试结果

实验采用的处理器为英特尔 i7-8750H,运行内存为 8 GB,使用 GPU 加速,GPU 为 GTX1060,显卡为 6 G。同时采用 pycharm 开发工具进行训练测试实验。

3.1　实验数据集及分析

为了验证改进模型的有效性,选取常用于目标检测的 MITSceneParsing 数据集。MITSceneParsing 数据集中包含了 151 个物体类别,包括天空、道路、草地、人、车、床等离散对象,并且图像中所存在的对象是不均匀分布的,更自然地模拟了日常场景中所存在的对象。MITSceneParsing 数据集总共 22 210 张图片(包括用于训练模型的 20 210 张图片和用于验证模型的 2 000张图片)。在模型训练中,对数据集总共训练 11 000次,训练集每10次输出训练损失率,验证集每

表 1　改进网络的各层参数

层数	图像大小	操作	卷积核大小	卷积核个数
esp	512 * 7 * 7	DConv+Relu	512 * 3 * 3	64
T1	64 * 7 * 7	deConv1	64 * 4 * 4	512
	512 * 14 * 14	Concat	–	–
	1 024 * 14 * 14	Conv9+Relu	1 024 * 3 * 3	512
	512 * 14 * 14	Conv10+Relu	512 * 3 * 3	512
T2	512 * 14 * 14	deConv2	512 * 4 * 4	256
	256 * 28 * 28	Concat	–	–
	512 * 28 * 28	Conv11+Relu	512 * 3 * 3	256
	256 * 28 * 28	Conv12+Relu	256 * 3 * 3	256
T3	256 * 28 * 28	deConv3	256 * 16 * 16	NUM_OF_CLASSESS
OUT	NUM_OF_CLASSESS * 224 * 224			

500 次输出验证损失率。为了防止出现过拟合现象，在训练过程中运用参数正则化方法，以 0.85 的概率保留部分权重参数参与训练。

3.2 网络模型及训练

网络输入图像大小为 3 × 224 × 224，输出特征图大小为 224 × 224，实现像素级端到端的目标检测。文中使用 Adam 算法来优化模型，Adam 算法结合了 Momentum 算法和 RMSprop 算法的优点。Adam 算法也是基于梯度下降法，但 Adam 算法在每次迭代期间的参数变化都在一定范围内。由于在特定时间计算的梯度值，参数不会急剧变化，并且参数的值相对稳定。文中初始 η 为 0.000 1，一阶矩估计的指数衰减率为 $\beta_1 = 0.9$，并且二阶矩估计的指数衰减率为 $\beta_2 = 0.999$。具体在优化可训练变量时，采用的是 Adam 的改进，LazyAdam 是 Adam 的变体，可以更有效地处理稀疏更新，使其训练速度更快。原始的 Adam 算法为每个可训练变量维护两个移动平均累加器，累加器在每一步都会更新。而 LazyAdam 算法仅更新当前 batch 中出现的稀疏变量索引的移动平均累加器，而不是更新所有索引的累加器。

3.3 模型的性能评估

文中利用深度学习框架 Tensorflow 进行网络模型的训练和测试，实验采用交叉熵损失函数作为目标检测模型的评测指标。交叉熵是 Shannon 信息论中一个重要概念，在神经网络中用于度量模型中预测标记分布概率与真实标记分布概率之间的差异性信息。当交叉熵损失函数越小，模型中预测标记分布概率与真实标记分布概率之间的差异性越小，显著性目标检测越准确。

基于文中所改进的模型与全卷积神经网络模型在同一数据集上分别进行多次迭代实验，获得的迭代次数与交叉熵的变化曲线如图 4 所示。从实验结果观察可知，所改进的检测模型在相同的数据集上验证损失更少，达到相同的损失所需迭代次数更少。在时间成本上，模型进行批处理所消耗的时间如表 2 所示。所改进的模型虽然卷积层数增加，但所消耗的时间几乎与 FCN 基础模型所消耗的时间相同。所以，降低损失率的同时没有增加时间成本。

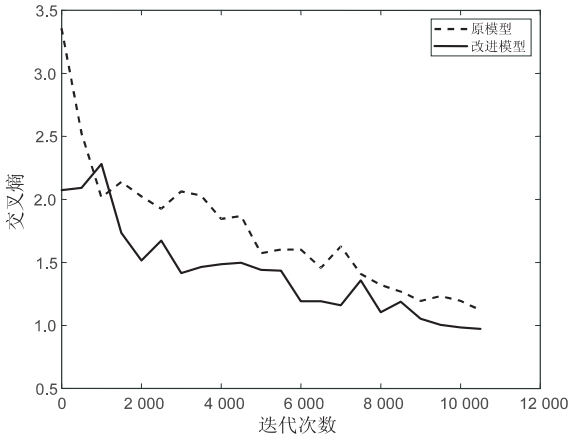

图 4 迭代次数与交叉熵的变化曲线

表 2 批处理所消耗的时间

时间成本	时间/(s/batch)
FCN	28.7
FCN+跳跃连接	32.2
FCN+ESP	31.1
文中方法	29

实验结果表明，在相同的数据集上文中所提出的目标检测模型具有更小的损失率，并且时间成本并没有显著增加，证明该方法具有更加良好的性能。

目标检测结果分别使用 MIOU 以及精确率的方式进行评价。表 3 给出了文中算法与相关算法在场景分析 ADE20K 数据集中的实验结果，在 MIOU 评价方式中，文中的目标定位稍微弱于其他相关算法，但相差数据并不大。例如平均目标定位相比于级联分割网络降低了 0.001 4。若单独使用 MIOU 等作为评估指标，由于真值是目标整体，评估结果具有较大的误差。文中通过像素精度和像素平均精度结合的评估方法，进一步验证算法的有效性。文中目标检测模型相比较于分割网络和空洞网络的像素精度和平均精度都有明显提高，在平均像素精度方面，文中方法比表现最好的级联空洞网络提升了 0.73%，比级联分割网络提升了 8.13%。表明文中算法的检测性能远优于其他相关算法。

表 3 场景分析数据库 ADE20K 验证集上的表现

Networks	Pixel Acc/%	Mean Acc/%	Mean IOU	Weighted IOU
FCN-8s	70.12	39.32	0.275 9	0.489 3
SegNet	71.00	30.74	0.213 1	0.535 7
DilatedNet	72.25	43.54	0.315 3	0.601 0
Cascade-SegNet	70.74	37.98	0.331 1	0.577 6
Cascade-DilatedNet	73.62	45.38	0.346 8	0.609 6
文中方法	73.83	46.11	0.329 7	0.590 1

4　结束语

提出一种改进的基于全卷积神经网络的多目标显著性检测模型。主要是对经典的全卷积神经网络的部分卷积过程和反卷积过程进行改进,即用 ESP 模块代替部分卷积层并且将深层特征图和浅层特征图在通道级别上进行连接并进一步进行特征提取,使得显著性目标特征图具有更多的边界信息。这样可以增强模型的鲁棒性和准确性,使得显著性目标更精确,并且模型的训练次数要求更低。虽然在损失上有改进,但是由于卷积层的增加,实验也存在着计算量大的缺点。由于硬件配置的原因,不能对模型进行更高迭代次数的训练及验证。模型的最终稳定状态需要进一步的探究。在未来的研究中,可以进一步减少计算成本和时间成本。

参考文献:

[1] 王　宇,张焕君,黄海新. 基于深度学习的图像语义分割算法综述[J]. 电子技术应用,2019,45(6):23-27.

[2] 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.

[3] LIU Tie,YUAN Zejian,SUN Jian,et al. Learning to detect a salient object[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(2):353-367.

[4] CHENG M,MITRA N J,HUANG X,et al. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(3):569-582.

[5] FAREED M M S,AHMED G,CHUN Q. Salient region detection through sparse reconstruction and graph-based ranking[J]. Journal of Visual Communication & Image Representation,2015,32:144-155.

[6] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM,2017,60(6):84-90.

[7] ERHAN D,SZEGEDY C,TOSHEV A,et al. Scalable object detection using deep neural networks[C]//2014 IEEE conference on computer vision and pattern recognition: 2014 27th IEEE conference on computer vision and pattern recognition (CVPR 2014). Columbus, Ohio: IEEE, 2014: 2155-2162.

[8] 王　旺,徐俊武,李颖先. 改进的基于卷积神经网络的图像超分辨率方法[J]. 计算机应用与软件,2019,36(3):214-218.

[9] LI Y,CUI F,XUE X,et al. Coarse-to-fine salient object detection based on deep convolutional neural networks[J]. Signal Processing:Image Communication,2018,64:21-32.

[10] ZHU D,YE L,LEI D,et al. Salient object detection via a local and global method based on deep residual network[J]. Journal of Visual Communication & Image Representation, 2018,54:1-9.

[11] CAO F,LIU Y,WANG D. Efficient saliency detection using convolutional neural networks with feature selection[J]. Information Sciences,2018,456:34-49.

[12] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,39(4):640-651.

[13] SHELHAMER E,LONG J,DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,39(4):640-651.

[14] 梁　博,于　蕾,李　爽. 基于卷积神经网络的多任务图像语义分割[J]. 无线电工程,2019,49(7):575-580.

[15] 施泽浩,赵启军. 基于全卷积网络的目标检测算法[J]. 计算机技术与发展,2018, 28(5):55-58.

[16] 易　阳. 基于深度学习的视觉显著性检测技术研究[D]. 北京:中国科学院大学,2018.

[17] 于　鹏. 基于无人机图像的识别算法研究[D]. 哈尔滨:哈尔滨工程大学,2018.