

## 基于改进 P-Unet 模型的岩屑颗粒识别\*

万川<sup>1</sup>, 王正勇<sup>1</sup>, 何海波<sup>2</sup>, 滕奇志<sup>1</sup>, 何小海<sup>1</sup>

(1. 四川大学 电子信息学院, 四川 成都 610065; 2. 成都西图科技有限公司, 四川 成都 610065)

**摘要:** 提出了一种基于改进 P-Unet 模型的岩屑颗粒识别方法。该方法基于 Unet 模型结构, 运用金字塔池化模块聚合不同区域的上下文特征信息, 以充分利用全局信息。改进 P-Unet 模型采用了残差网络 ResNeXt101, 在提高岩屑颗粒识别准确率前提下, 减少了超参数数量。该模型采用了焦点损失函数, 在一定程度上解决岩屑颗粒类别不平衡的问题, 同时运用深度可分离卷积代替传统卷积, 较大程度减少了网络的参数以及预测的时间。实验结果表明, 改进 P-Unet 模型得到的识别准确率对比同类先进算法有一定的提升, 对岩屑颗粒识别的结果更加准确。

**关键词:** 岩屑颗粒识别; Unet 网络; P-Unet 网络; 金字塔池化模块; 深度可分离卷积; 焦点损失函数; 残差网络

中图分类号: TP391.4

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2020.11.009

引用格式: 万川, 王正勇, 何海波, 等. 基于改进 P-Unet 模型的岩屑颗粒识别[J]. 信息技术与网络安全, 2020, 39(11): 56-61.

### Identification of cuttings particle based on improved P-Unet model

Wan Chuan<sup>1</sup>, Wang Zhengyong<sup>1</sup>, He Xiaobo<sup>2</sup>, Teng Qizhi<sup>1</sup>, He Xiaohai<sup>1</sup>

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China;

2. Chengdu Xitu Technology Co., Ltd., Chengdu 610065, China)

**Abstract:** This paper proposes a model based on improved P-Unet cuttings particle identification method. The method is based on the Unet model structure, and uses pyramid pooling module to aggregate contextual feature information in different regions to make full use of global information. The improved P-Unet model uses the residual network ResNeXt101, which reduces the number of super parameters on the premise of improving the accuracy of debris particle recognition. In this model, the focus loss function is used to solve the problem of the unbalance of cuttings particles. At the same time, the deep separable convolution is used to replace the traditional convolution, which greatly reduces the parameters of the network and the prediction time. The experimental results show that the recognition accuracy of the improved P-Unet model is improved compared with similar advanced algorithms, and the recognition results of cuttings particles are more accurate.

**Key words:** identification of cuttings particles; Unet network; P-Unet network; pyramid pooling module; deep separable convolution; focus loss function; residual network

### 0 引言

岩屑是在钻井过程中产生的岩石碎块, 其种类繁多, 按照岩屑的组成成分可以将岩石大致分成沉积岩、变质岩和岩浆岩。岩屑是岩屑录井的直观材料, 其中岩屑颗粒识别工作是岩屑录井的主要工作。对岩石岩性的识别工作可以反映出地段地质的地

层特性, 从而能帮助勘探工作者分析地层的具体特性。岩性准确识别是地质人员研究地层特征和地质建模的基础。

目前国内对岩屑颗粒的识别鉴定传统的做法是采用人工标记的方式对岩屑颗粒进行采集分析。这需要专业人员来识别, 并且工作量巨大。人工方式的标记对岩屑提取的信息与分析的结论存在误

\* 基金项目: 国家自然科学基金(61372174)

差,对地质专家分析地质存在较大影响。绝大多数在岩屑录井上采用数字图像识别和机器学习等技术手段,运用数字图像处理技术分析岩屑颗粒图像,从而得到岩屑的纹理、颜色、空洞等特征,来用作区分岩屑颗粒岩性的识别分类。虽然利用数字图像处理等方法解决了传统特定方法中<sup>[1-2]</sup>人为标记岩性的问题,但是由于仍需要在分类时人为设置相应的特征,在某些程度上限制了识别的准确度。随着神经网络广泛被应用于计算机视觉领域,且不同的实验场景推出了不同深度神经网络模型<sup>[3-5]</sup>,通过将较低分辨率图片的特征映射到像素级尺度对图片中的每个像素进行识别,从而大幅度提升了图像识别的水平。但神经网络模型复杂度提升,增加了岩屑识别所耗费时长,对于岩屑颗粒的识别准确率有很大的提升空间。

为了提高岩屑颗粒识别算法准确率和识别效率,本文提出了一种基于改进 P-Unet 岩屑颗粒识别模型,在文献[6]基础上做出了如下改进:

(1)在不改动 Unet 模型对称结构的情况下,将金字塔池化模块<sup>[7]</sup>融合到 Unet 中,从而得到了本文模型 P-Unet 网络结构。

(2)运用了多分类焦点损失函数<sup>[8]</sup>代替传统的交叉熵损失函数。

(3)将深度残差网络 ResNet<sup>[9]</sup>的升级版 ResNeXt<sup>[10]</sup>的网络结构用于 P-Unet 下采样的结构。

(4)用深度可分离卷积<sup>[11-12]</sup>代替传统卷积。

## 1 传统 Unet 网络模型

传统的 Unet 模型呈 U 形,具体结构如图 1 所示,图中每一层数字表示卷积核个数,结构前半部

分是下采样,经过不同的卷积层卷积,提取图像深层次的特征。结构后半部分是上采样,实现方式利用反卷积的方式实现,输出指定类别数量的特征图。有学者也称这样的结构为编码器-解码器结构。

## 2 改进的 Unet 网络

传统 Unet 网络在上采样时会直接拼接上一层的特征和对应下采样得到的特征信息,结果容易出现特征丢失等问题。为了有效解决上述问题,本文提出了改进的 P-Unet 网络。本文模型采用了 Unet 网络模型编码和解码过程,同时在上采样时将对应的下采样得到的特征经过金字塔池化模块之后和上一层特征拼接,拼接完之后再上进行上采样。本文还用金字塔池化模块替换了最后一层下采样与上采样的过程。将金字塔池化模块运用在特征拼接过程和最后一层的上下采样过程,能充分利用上下层卷积层所提取的特征信息。同时运用了残差网络 ResNeXt101 代替传统残差网络 ResNet,并且将 P-Unet 网络中的传统卷积方式替换为深度可分离卷积,以及采用焦点损失函数代替传统的损失函数,具体 P-Unet 结构如图 2 所示。

### 2.1 残差网络

传统方式一般通过加深或者加宽网络提高网络准确率,但同时伴随着网络参数和网络计算时间的增加。本文 P-Unet 网络采用了升级版的残差网络 ResNeXt,此结构利用了 Inception<sup>[11]</sup>的多支路的思想,保持了 ResNet 可移植性的优点,在改变模型复杂度的情况下增加一定的准确率。相比较 ResNet 结构,ResNeXt 可以在不增加网络参数复杂度的情况下,同时减少了超参数的数量。ResNeXt 除采用了

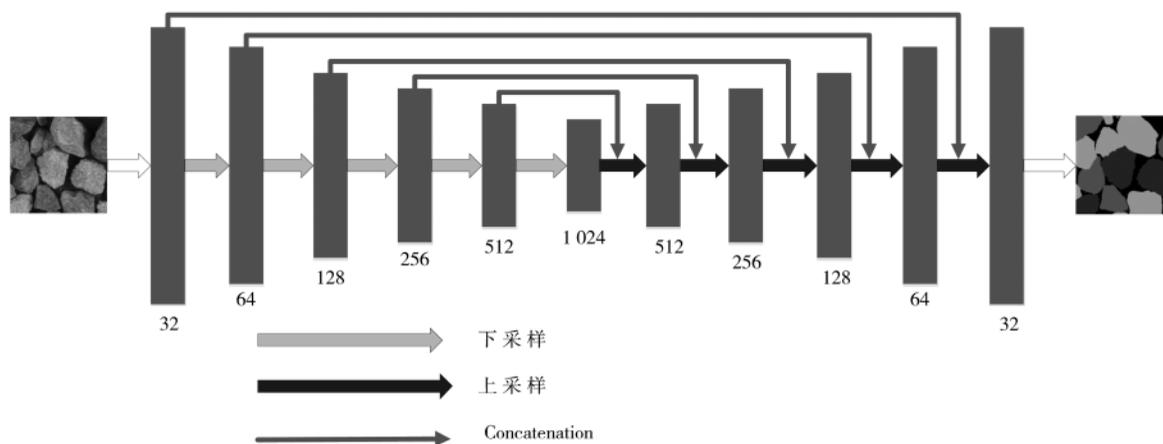


图 1 Unet 网络结构

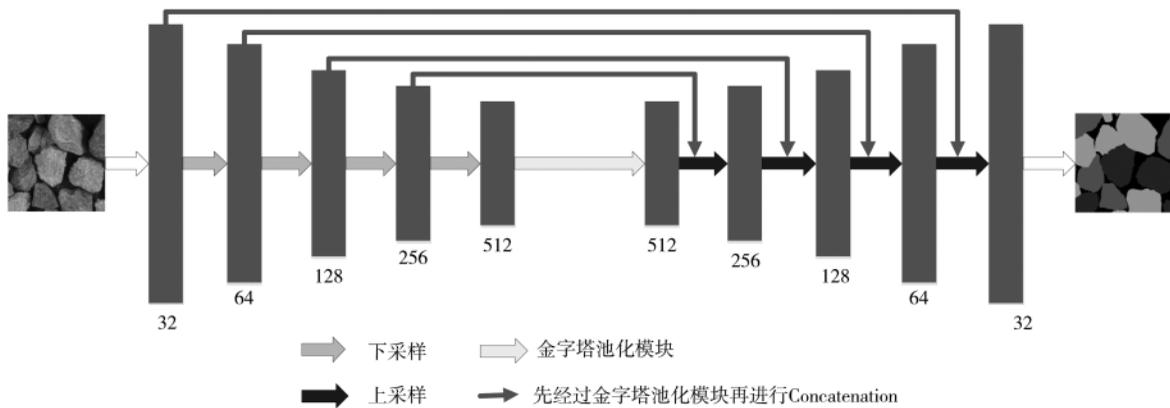


图 2 P-Unet 网络结构

VGG 网络堆叠的思想,同时还引用了 Inception 中多支路的思想。图 3(a)显示了 ResNet 结构中的基本块结构,图 3(b)显示 ResNeXt 结构中对应 ResNet 结构的基本块结构,图 3 长方形中的第一个数字表示参加卷积操作的输入通道数,长方形中的第二个表达式中相同的数字表示卷积操作卷积核的大小,第三个数字表示卷积的通道数。ResNeXt 结构引入了一个名叫基数(cardinality)的超参数,指的是独立路径的数量,通过这个参数可以调整模型的容量。在图 3(b)中基数为 32,通过扩大基数,网络模型的准确率能得到一定的提升。

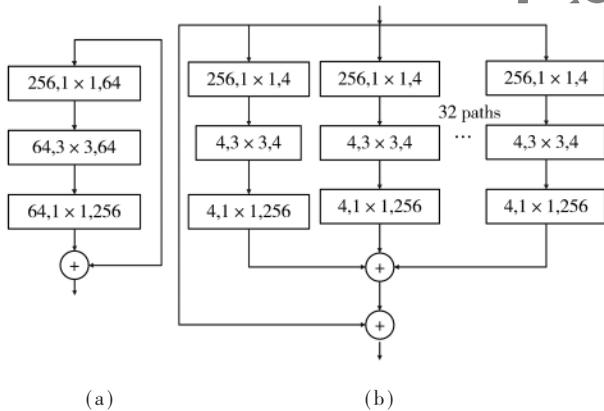


图 3 ResNet 基本块结构和 ResNeXt 基本块结构

本文改进 P-Unet 网络融合了 ResNeXt101 网络,具体网络结构如表 1 所示(表中 C 表示的是基数)。与 ResNet 相比较,在相同参数个数条件下,一个 101 层 ResNeXt 网络和 200 层 ResNet 网络准确度差不多,但前者计算量只有后者的一半。ResNeXt 网络采用的是 Inception 网络结构的思想,但残差网络全新的架构相比 Inception 网络更加能适应新的数据。

表 1 ResNeXt101 结构与 ResNet101 的结构对比

Stage	Output	ResNet101	ResNeXt101
Conv1	512×512	7×7, 64, stride 2 3×3 max pool, stride 2	7×7, 64, stride 2 3×3 max pool, stride 2
Conv2	256×256	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64, C=32 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3	128×128	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128, C=32 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4	64×64	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1\ 024 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256, C=32 \\ 1 \times 1, 1\ 024 \end{bmatrix} \times 23$
Conv5	32×32	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2\ 048 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512, C=32 \\ 1 \times 1, 2\ 048 \end{bmatrix} \times 3$
	1×1	Pool, softmax	Pool, softmax

## 2.2 金字塔池化模块

金字塔池化模块最初应用于 PSPnet 网络<sup>[7]</sup>,用于聚合不同区域的上下文特征信息,从而提高获取全局特征的能力。金字塔池化模块结构如图 4 所示,模块中融合了 4 种不同金字塔尺度的特征,Level1 显示了最粗略的全局池化后产生的层级,该层是单个特征输出。Level2、Level3 和 Level4 是不同尺度池化后的特征,能获得不同的全局范围的特征。为了获取全局特征的权重,假如金字塔总共有  $n$  个不同的级别,则在池化之后产生的每一个级别分别进行卷积将每个级别的通道数变为原来的  $1/n$ 。金字塔池化模块每个级别的池化核大小可人为设定,池化核大小不同对应不同的级别,池化核大小与输入的尺寸有关。

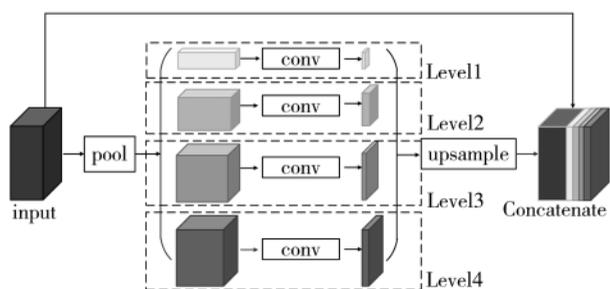


图4 金字塔池化模块结构

本文 P-Unet 网络模型融合金字塔池化模块结构对于像素级上的识别能充分获取上下文不同区域场景下的特征,对岩屑颗粒识别的准确率有一定的提升。

### 2.3 焦点损失函数

焦点损失函数(Focal Loss)是在交叉熵损失函数<sup>[12]</sup>基础上进行改进得到的损失函数,焦点损失函数表示为:

$$FL(p) = -\alpha * (1-p)^\gamma * \log(p) \quad (1)$$

从式(1)可以看到焦点损失函数相比交叉熵损失函数引入了两个超参数  $\alpha$  和  $\gamma$ ,  $\alpha$  是用来平衡样本数量,即可以平衡正负样本本身的比例不均衡的问题,  $\gamma$  是用来调节简单样本和复杂样本所占的权重,即调节简单样本权重降低的速率,当  $\gamma$  增加时,调整因子的影响也在增加。当  $\gamma=0$  且  $\alpha=1$  即为交叉熵损失函数,对于交叉熵损失函数,正样本的输出概率越高损失越小,负样本输出概率越小损失越小。本文岩屑颗粒识别是多分类问题,因此焦点损失函数中  $\alpha$  是一个数组,数组大小为岩屑颗粒类别个数,分别代表着每一个类别对应的权重。普通交叉熵损失函数在大量简单样本的迭代过程中存在着比较缓慢而且可能无法达到最优的问题。焦点损失函数能有效地解决交叉熵损失函数所不能解决的样本比例不平衡的问题以及上述问题。

### 2.4 深度可分离卷积

文中采用残差网络虽然能提升岩屑识别的准确率,但是随着网络特征提取能力的增强,网络复杂度也随之增高,网络参数也增加,使得预测和训练的时间增长。为了减少网络参数,本文用深度可分离卷积<sup>[13]</sup>代替常规的卷积方式极大减少网络参数。

深度可分离卷积主要是将传统的卷积分为两部分,一部分是深度卷积,另一部分为卷积核大小  $1 \times 1$  的卷积,如图5所示。

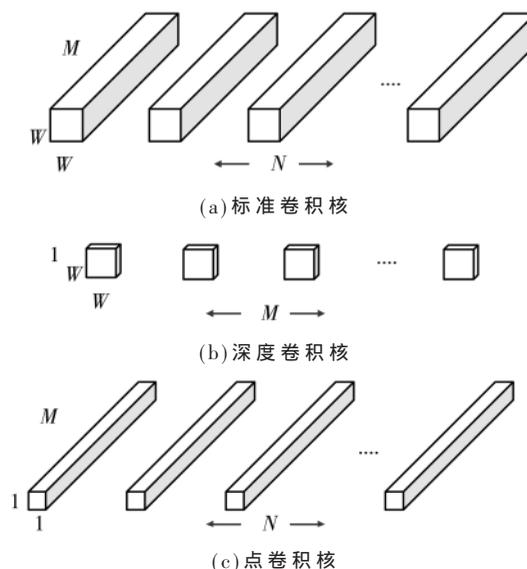


图5 传统卷积核、深度卷积核、点卷积核

图5中(a)、(b)、(c)表示分别为传统的卷积、深度卷积和  $1 \times 1$  的卷积。传统卷积对特征图的所有通道同时卷积,无论多少个通道输出一个数。深度可分离卷积将正常的卷积过程分为两步:假设输入有  $N$  个通道,第一步是用  $N$  个卷积对  $N$  个通道分别做卷积,这样可以得到  $N$  个数;第二步是将第一步得到的特征图通过  $1 \times 1 \times N$  的卷积核进行卷积。

假设输入的特征图的尺寸为  $W \times W \times M$ , 输出特征图的大小为  $W \times W \times N$ , 卷积核的大小为  $H \times H$ , 则通过传统卷积的性质可知该计算量为:

$$H \times H \times M \times N \times W \times W \quad (2)$$

深度可分离卷积的计算量为上述两步的计算量之和,第一步深度卷积的计算量为:

$$H \times H \times M \times W \times W \quad (3)$$

第二步  $1 \times 1$  卷积的计算量为:

$$W \times W \times M \times N \quad (4)$$

根据上述的计算量得到深度可分离卷积和传统卷积之比,如式(5)所示:

$$\frac{H \times H \times M \times W \times W + W \times W \times M \times N}{H \times H \times M \times N \times W \times W} = \frac{1}{N} + \frac{1}{H \times H} \quad (5)$$

假设卷积核大小为  $C \times C$ , 可以得到传统卷积的计算量大致是深度可分离卷积的  $C^2$  倍。

根据以上性质可知,用深度可分离卷积代替传统卷积可以大量减少网络的参数,同时减少预测时间。

## 3 实验结果及分析

### 3.1 实验数据

本文的数据集来源于地质录井公司提供的岩

屑样本,先经过筛选岩屑样本得到 6 类比较多的岩屑颗粒,再通过实验室采集软件采集得所需样本图片。本文将采集得到的样本进行裁剪得到初步样本。再将样本经过旋转、镜像、平移等数据增强操作的方式得到最终所需的数据集,总共得到 570 张 512×512 的岩屑颗粒的训练集样本,以及通过随机选取得到的 60 张 512×512 的岩屑颗粒的测试集样本。岩屑样本中总共分为 7 类,将不同类别的岩屑用不同 RGB 值标记,其中本文将背景(背景指的是装岩屑颗粒的盒子)也当成 1 类,背景的 RGB 值设置为全 0。

### 3.2 实验环境

本文的实验在 Windows 7 64 位操作系统下进行,实验使用的硬件资源为 Intel Core i5-7500 CPU。实验中使用的深度学习框架是以 TensorFlow 作为其后端,Keras 作为其前端,同时利用 Python3.6 作为编程语言进行实验。实验中采用的是传统收敛速度较快的 Adam 优化器优化损失,其中 epochs 为 100, batch\_size 为 1。

### 3.3 结果分析

将自制的训练数据集输入改进的 P-Unet 网络,用改进的网络进行训练,采用客观评价指标<sup>[14]</sup>对训练得到的结果与一些算法进行了对比。传统准确率的公式如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (6)$$

式中:TP 表示预测图和标签图中像素值不为 0 的像素点的总数量;FP 表示预测图中像素值不为 0 并且标签图中像素值为 0 的像素点的总数量;FN 表示预测图中像素值为 0 并且标签图中像素值不为 0 的像素点的总数量;TN 表示预测图和标签图中像素值为 0 的像素点的总数量;可以看出 FN、FP 是预测错误的像素,TP、TN 是预测正确的像素,本

文准确率公式简化为:

$$\text{Acc} = \frac{\text{PN}}{\text{PN} + \text{FN}} \quad (7)$$

式中:PN 表示预测值与标签值相同的像素;FN 表示预测值与标签值不同的像素。

从表 2 看出:改进 P-Unet 网络准确率相比 Linknet<sup>[15]</sup>网络、传统的 Unet 网络分别上升了 2.31%、2.65%。同时对比了采用不同残差网络得到的准确率,运用了 ResNeXt101 的 Unet 网络准确率比运用了 ResNet101 的 Unet 网络准确率上升了 0.16%。改进的模型可以更加准确地预测出岩屑颗粒的种类。

表 2 本文模型与其他模型性能对比

方法	准确率/%
Unet 网络	93.72
Linknet 网络+ResNet101	94.06
Unet+ResNet101	95.47
Unet+ResNeXt101	95.63
P-Unet(传统卷积)	96.78
改进的 P-Unet	96.37

根据表 2 和表 3 可知使用传统卷积的 P-Unet 网络准确率高,但是模型的复杂度高。本文中使用了深度可分离卷积替换传统卷积之后,虽然准确率下降了 0.41%,但使用深度分离卷积的 P-Unet 网络训练参数大约是使用传统卷积的 P-Unet 网络训练参数的 1/3 倍,网络的复杂度也相应大大降低。

## 4 结论

本文提出了一种改进 P-Unet 网络模型的岩屑颗粒识别方法,利用传统 Unet 网络的上采样和下采用的技术,融合了金字塔池化模块,采用了焦点损失函数作为网络的损失函数,以及 ResNeXt 残差网络结构等一系列的技术,本文改进的模型在岩屑颗粒识别准确性上得到进一步的提升。实验结果表明,

表 3 本文模型与其他模型的参数量

方法	总参数	可训练参数量	不可训练参数量
Linknet 网络+ResNet101	47 821 553	47 723 755	97 798
Linknet 网络+ResNeXt101	47 497 713	47 359 851	137 862
Unet+ResNet101	105 343 577	105 343 577	97 798
Unet+ResNeXt101	105 019 737	104 881 875	137 862
Unet(传统卷积+ResNet101)	161 024 089	160 926 291	97 798
P-Unet(传统卷积+ResNeXt101)	160 700 249	160 562 387	137 862
P-Unet(深度可分离卷积)	58 340 084	58 202 222	137 862

本文的方法有效改进了岩屑颗粒识别的准确率,在岩屑颗粒识别领域上有一定的参考价值。

#### 参考文献

- [1] MALLADI S R S P, RAM S, RODRIGUE J J. Super-pixels using morphology for rock image segmentation[C]. 2014 Southwest Symposium on Image Analysis and Interpretation. San Diego, 2014: 145-148.
- [2] YANG G Q, WANG H G, XU W L, et al. Ore particle image region segmentation based on multilevel strategy[J]. Chinese Journal of Analysis Laboratory, 2014, 35(24): 202-204.
- [3] SHELHAMER E, LONG J, DARRELL T. Fully convolution networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4): 640-651.
- [4] NOH H, HONG S, & HAN B. Learning deconvolution network for semantic segmentation[J]. IEEE International Conference on Computer Vision (ICCV), 2015: 1520-1528.
- [5] SIMONYAN K, ZISSERMAN A. Very Deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014: 1-14.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, 2015: 234-241.
- [7] ZHAO H, SHI J, QI X, WANG X, et al. Pyramid scene parsing network[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230, 239.
- [8] LIN T, GOYAL P, GIRSHICK R B, et al. Focal loss for dense object detection[J]. IEEE International Conference on Computer Vision (ICCV), 2017: 2999-3007.
- [9] SZEGEDY C, IOFFE S, VANHOUCKE V. Inception-v4 inception-resnet and the impact of residual connections on learning[J]. (2016-02-xx). <https://arxiv.org/pdf/1602.07261.pdf>.
- [10] XIE S, GIRSHICK R B, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5987-5995.
- [11] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [12] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of Operation Research, 2005, 134(1): 19-67.
- [13] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1800-1807.
- [14] ORLANDO J F, PROKOFVEVA E, BLASCHKO M B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(1): 16-27.
- [15] CHAURASIA A, CULURCIELLO E. Linknet: exploiting encoder representations for efficient semantic segmentation[C]. IEEE Visual Communications and Image Processing, 2017.

(收稿日期: 2020-08-05)

#### 作者简介:

万川(1994-),男,硕士研究生,主要研究方向:智能系统与设计。

王正勇(1969-),通信作者,女,博士,副教授,主要研究方向:图像处理与模式识别。E-mail: wangzheng@scu.edu.cn。

何海波(1968-),男,高级工程师,主要研究方向:图像处理与计算机应用。

# 版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部  
中国电子信息产业集团有限公司第六研究所