

基于 U-net 网络改进的线谱检测方法^{*}裴 明¹ 陈 阳¹ 邓林红²

(1.常州大学 微电子与控制工程学院 常州 213164; 2.常州大学 生物医学工程与健康科学学院 常州 213164)

摘 要: 针对被动声纳系统中不规则波动线谱检测准确率低的问题,提出一种基于 U-net 网络改进的线谱检测方法。以 U-net 作为网络主体框架,引入残差结构,增加网络的深度,加强模型特征学习能力。同时在编码器部分引入特征通道注意力机制,使模型学习到通道之间的不同特征的重要程度,从而提升模型的特征表达能力。最后,在解码器部分采用 DUpsampling 上采样方法,利用分割标签空间中的冗余能力,更准确的恢复线谱像素级预测。将改进模型与 HMM 模型和 CEM 模型在线谱检测效果上进行比较。实验结果表明,在信噪比为 $-24 \sim -20$ dB 下,改进模型线谱检测准确率为 $0.314 \sim 0.526$,优于 HMM 模型和 CEM 模型。

关键词: 线谱检测;U-net;残差结构;特征通道注意力机制;DUpsampling

中图分类号: TN911.7 **文献标识码:** A **国家标准学科分类代码:** 140.2050

An improved spectrum line extraction method based on U-net network

Pei Ming¹ Chen Yang¹ Deng Linhong²

(1. School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213164, China;

2. School of Biomedical Engineering and Health Sciences, Changzhou University, Changzhou 213164, China)

Abstract: A line spectrum detection method improved by U-net network is proposed, which is aiming at the problem of low detection accuracy of the irregular fluctuating line spectrum in passive sonar system. The system framework is based on U-net network, and the residual structure is introduced to increase the depth of the network and strengthen the learning ability of the model feature. Meanwhile, the feature channel attention mechanism is brought in the encoder part which can help the model learn the importance of different features between channels, thus improving the feature expression ability of the model. Finally, the DUpsampling up-sampling method is used in the decoder part, and the redundancy capability in the segmented label space is utilized which can accurately restore the pixel-level prediction. The improved model is compared with the HMM model and CEM model on the line spectrum detection effect. Under the SNR of $-24 \sim -20$ dB, the accuracy of line spectrum detection of the improved model is $0.314 \sim 0.526$, which is better than that of HMM model and CEM model.

Keywords: line spectrum extraction; U-net; residual structure; the feature channel attention mechanism; DUpsampling

0 引 言

被动声纳领域中,谱图包含了丰富的信息。不仅含有水下运动目标的速度或航迹等变化情况,而且还有助于水下运动目标的检测和识别^[1]。因此谱图可以作为水下运动目标分析的可靠特征。所以,频谱的检测、提取和分析具有重要的意义。实际应用中,频谱分析通常是通过 LOFAR 谱图进行的。LOFAR 谱图是利用短时傅里叶变换(short-time Fourier transform, STFT)将时域和频域的处理过程结合起来形成的^[2]。LOFAR 谱图的横轴和纵

轴分别对应频率和时间,每个像素的强度值是用给定信号在相应频率和时间的功率或能量来测量的。通过观察 LOFAR 谱图,被动声纳用户可以快速探测和识别目标船检。

低信噪比(signal noise ration, SNR)和水下目标运动引起的频率变化是线谱检测的一大挑战^[3]。文献[4]中率先使用动态规划进行单频线跟踪。把数据当作一段含有噪声的时间信号序列,此时可以通过维特比算法估计频率线位置。文献[5]提出一种基于马尔科夫模型(hidden Markov models, HMM)的频线跟踪方法。以位置和速度

收稿日期:2021-05-10

^{*} 基金项目:声纳技术国防科技重点实验室基金(6142109180206)项目资助

二维状态向量构建滤波器,并在HMM中引入谱功率概率积分,提高了线谱提取质量^[6]。但是,在低信噪比条件下该滤波器得到的特征是不可靠的。Lampert等^[7]提出一种用于谱图轨迹检测的主动轮廓能量最小化(contour energy minimisation,CEM)算法。该算法定义能量最小化过程,此过程受到内部能量的约束,内部能量根据曲率和连通性等条件约束线谱形状,使得线谱形状和其局部能量梯度变化一致。该方法将结构信息添加到决策过程中,提高了低信噪比条件下线谱检测率^[7]。但是,低信噪比下不规则波动线谱检测准确率依然较低。

随着深度卷积神经网络在图像分割方面取得了一系列突破。深度卷积神经网络把输入转换为抽象的高级特征,从而放大重要方面,同时抑制无关的变化^[8]。随着网络层数的加深,特征图包含的图像信息会逐层减少,同时梯度扩散会使网络难以训练^[9]。2015年,由He等^[10]提出残差结构,残差结构在一定程度上解决了这种退化现象。基于编码器和解码器结构的卷积网络在图像像素预测上有着更好的效果,Long等^[11]提出了全卷积神经网络(full connected network,FCN)用于图像的分割。该方法在采样过程中会损失部分信息,从而得到更为重要的特征,然后通过上采样操作可以得到更精确的分割边界。但此过程是不可逆转的,有时甚至会导致图像分辨率低以及细节丢失。FCN方法是一种反卷积采样结构,其核心思想是经过采样扩大像素,再进行卷积并通过学习获得权值。其优点是可以接受任意大小的输入图像,并保留其空间信息,解决了图像语义级分割问题。但由于每个像素的上采样都是单独分开执行的,因此没有考虑像素之间的关系和缺乏空间一致性,从而导致部分结果模糊以及对图像细节处理不敏感。Ronneberger等^[12]提出U型网络(U-net),它主要由收缩路径(特征提取)和扩展路径(上采样)两部分组成。收缩路径部分用于捕获图像中的上下文信息,而上采样部分则用于恢复图像位置信息。但是,该上采样方法不考虑每个像素预测之间的相关性,导致它在精确恢复像素预测方面的能力有限。最近,由Tian等^[13]提出一种新的上采样方法:Dupsampling,该方法用分割标签空间中的冗余能力,使得编码器生成的粗糙特征也能获得精确的分割结果,比其他上采样方法更好。

本文提出一种基于U-net网络改进的模型提高不规则波动线谱检测准确率。模型以U-net作为网络主体框架,引入残差结构,在编码器部分加入残差单元,增加网络的深度,增强模型特征学习能力。同时在残差单元后引入特征通道注意力机制^[14],使模型学习到线谱特征图通道之间的不同特征的重要程度,从而提升模型的特征表达能力。最后,在解码器部分采用DUpsampling上采样方法,DUpsampling利用了分割标签空间中的冗余优势恢复编码路径中丢失的细粒度信息,增大特征图的分辨率,可以更准确地恢复线谱像素级预测。

1 网络结构

本文网络的主体结构如图1所示,整体分为编码器和解码器两个部分,编码器向下采样输入以提取特征,而解码器向上采样输入以恢复分辨率。网络使用1个 224×224 的灰度图片作为编码器的输入,编码器以残差单元(residual unit)作为主体提取特征,并加入特征通道注意力模块。特征通道注意力模块对提取的特征在通道维度上进行重新标定。然后Max pool下采样,采样之后的输出通道数为原来通道数的2倍。在解码器部分用DUpsampling上采样代替传统的双线性插值上采样,上采样后输出的通道数为输入通道数的一半。再经过2个 3×3 卷积层作用,每个卷积层之后都有规范化处理(batch normalization, BN)和ReLU激活函数。每层之间都有一个跳跃连接,以便解码时结合相应的编码特征。

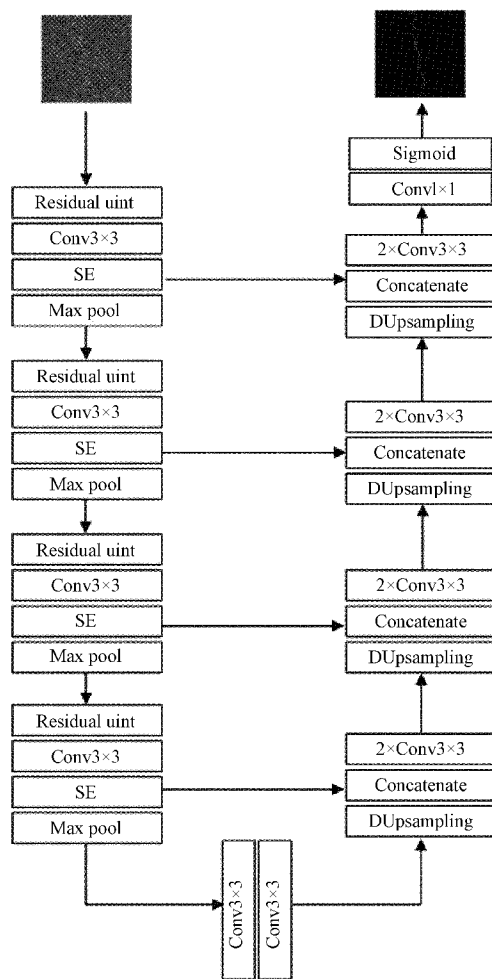


图1 本文网络结构

2 编码模块

2.1 残差单元

如图2所示,假设 x 是输入,输出是 $H(x)=F(x)+x$,

$H(x)$ 是一个复杂的潜在映射,要想学习这样的模型,训练的难度较大。如图 2 所示,通过“捷径连接”,直接把输入 x 作为结果的一部分输出。当 $F(x)=0$ 时, $H(x)=x$,这就是恒等映射。此时学习的目标已经变为 $F(x)$, $F(x)=H(x)-x$,即所谓的残差,训练的目标是使残差趋向于 0。残差网络通过将输入与多个卷积层输出结果相加的方式,加深了网络深度,减少了训练误差,获得了更多特征语义信息。

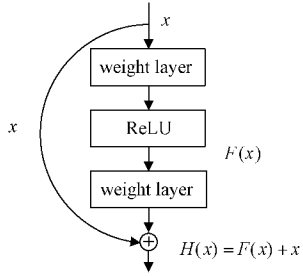


图 2 残差结构

图 3 为本文残差单元模块,每个 3×3 卷积层之后都有规范化操作和 ReLU 激活函数。

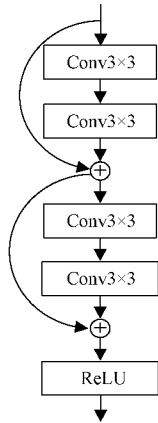


图 3 残差单元

2.2 特征通道注意力模块

注意力模块能让网络选择聚焦位置,产生更具分辨性的特征表示^[15]。普通的卷积无法提取不同特征的重要程度,本文引入特征通道注意力模块,在编码器 3×3 卷积之后增加压缩激励(sequeeze and excitation, SE)模块。如图 4 所示,SE 模块在训练过程中学习到卷积层提取到的不同特征对于线谱提取的重要性,再通过每个特征通道特征重标定,从而增强模型对有用特征的响应,抑制不重要特征,提升模型的特征表示能力。

卷积层计算得出特征 $U = [u_1, u_2, \dots, u_c]$, u_c 表示任意一个特征通道的特征图。采用全局平均池化将每个通道上的特征图压缩为一个有全局感受野的实数,假设压缩后的特征 $z \in \mathbf{R}^c$, $z = [z_1, z_2, \dots, z_c]$, z_c 表示任意一个通道上的全局特征,计算公式如下:

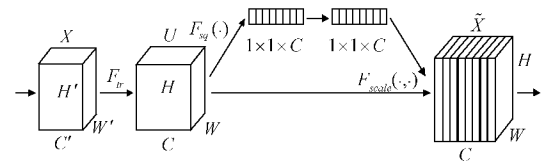


图 4 SE 模块

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

如图 5 所示,经过压缩之后的特征图通过 2 个全连接层 FC, ReLU 激活函数,最后通过 Sigmoid 激活函数得到归一化特征权重 s 。

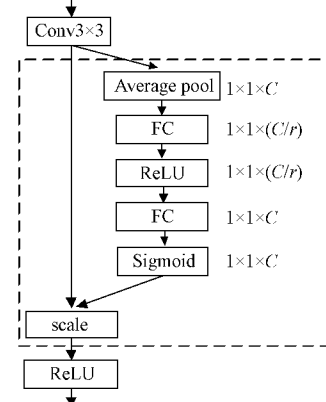


图 5 SE 模块内部结构

s 计算公式如下:

$$s = F(s, W) = \sigma(g(z, w)) = \sigma(W_2 \delta(W_1 z)) \quad (2)$$

式中: σ 和 δ 分别表示 Sigmoid 激活函数和 ReLU 激活函数。 $W_1 \in \mathbf{R}^{\frac{c}{r} \times c}$, $W_2 \in \mathbf{R}^{\frac{c}{r} \times c}$, r 为超参数。最后,对卷积层输出特征 U 进行权重重标定,把各个通道的权重 s 与特征 U 做权重相乘,如下计算:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (3)$$

输出特征 $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_c]$, $\tilde{x}_c$ 表示任意一个通道输出特征图。对于各个通道输出特征,权重 s 的值可以调节通道之间的相关性。在网络训练过程中,对于与线谱提取的有用特征,通过 SE 模块,加大对对应特征通道的权重值,以增强相应特征响应;同时,对于无用和干扰特征,减小对应特征通道的权重值,抑制相应特征。通过引入注意力机制,让网络提取特征的指向性更强,从而进一步增强模型的线谱提取能力。

3 解码模块

在解码模块部分,网络采用的上采样方法不是 U-net 的双线性插值法,而是选择 DUpsampling,这是一种依赖数据的上采样结构。双线性插值法不考虑每个像素预测之间的相关性,导致它在精确恢复像素预测方面的能力有限。DUpsampling 利用了分割标签空间中的冗余能力,能从相

对粗糙的卷积网络输出中准确的恢复像素级预测,通过编码器生成的粗糙特征也能获得精确的分割结果。

DUpsampling 的采样过程如图 6 所示, H 、 W 、 C 分别是特征图的高度、宽度、特征通道数, $1 \times C$ 表示输入特征图中每个像素的特征维度, 将其乘上维度为 $C \times N$ 的矩阵 W , 得到一个 $1 \times N$ 的特征表示。最后将其 Rearrange 为 $2 \times 2 \times N/4$ 的特征表示, 上采样过程就完成了。

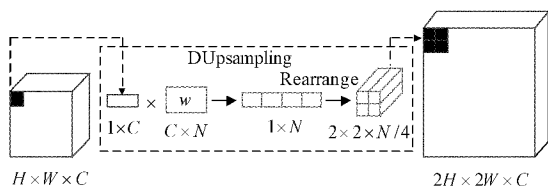


图6 DUPsampling 过程

上面的矩阵 W 是根据已知的训练标签得到的, 是真实的结果。然后对每个分割图进行矩阵变换, 其过程如下:

$$x = Pv \quad (4)$$

$$\tilde{v} = Wx \quad (5)$$

其中, v 表示的是真正分割结果中的区域, 矩阵 P 是矩阵 W 的逆变换。通过矩阵 P 对 v 进行压缩。 P 矩阵和矩阵 W 可以通过最小化 v 和 $\tilde{v}$ 误差得到。过程如下:

$$P^*, W^* = \arg \min_{P, W} \sum_v \|v - \tilde{v}\|^2 = \arg \min_{P, W} \sum_v \|v - PWv\|^2 \quad (6)$$

解码器模块有 4 次下采样, 对应的解码器模块有 4 次 DUPsampling 上采样, 并且编码器和解码器之间有跳跃连接, 使得解码器可以获得编码的特征信息, 减少解码过程中的信息丢失。通过 4 次上采样, 可以有效地恢复线谱特征图信息。

4 实验和分析

4.1 实验数据

LOFAR 谱图制作流程如图 7 所示, 假设系统输入信号为 $s(t)$:

$$s(t) = a(t) \sin(2\pi f(t) + \varphi) + \eta(t) \quad (7)$$

式中: t 是时间; f 是随机变化的频率, $\eta(t)$ 是均值为 0, 方差为 σ^2 高斯白噪声。首先, $s(t)$ 把输入信号分割成 k 帧, 第 k 帧和第 $k-1$ 帧可以部分重叠。通过短时傅里叶变换(STFT)得到每帧的功率谱图, 按照时间顺序把每帧功率谱连接起来, 形成一幅 LOFAR 谱图^[16]。信噪比(signal to noise ratio, SNR)定义如下:

$$SNR = 10 \log \frac{P_s}{P_\eta} = 10 \log \frac{a^2}{2\sigma^2} \quad (8)$$

式中: P_s 表示信号功率, P_η 表示噪声功率, a 是信号幅值, σ^2 为噪声方差。根据式(8)可以得到指定信噪比的 LOFAR 谱图。

4.2 训练设置

LOFAR 谱图训练集为 2 000 张, 测试集为 100 张。如

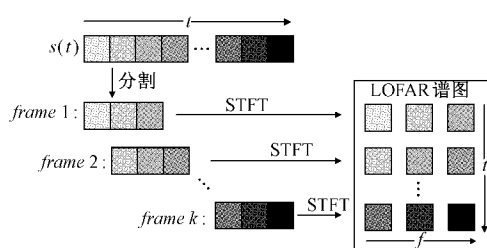


图7 LOFAR 谱图制作过程

图 8 所示, 对训练集进行垂直翻转, 水平翻转数据增强操作, 防止网络过拟合。所有训练集和测试集大小为 224×224 , 并对其进行灰度化和归一化。先用 SNR 为 $-18 \sim -21$ dB 的 LOFAR 谱图预训练网络, 学习率为 0.001。然后再用 $-21 \sim -24$ dB 的 LOFAR 谱图训练网络, 学习率调整为 0.000 1。采用 Adam 优化器进行优化, 动量参数设置为 0.9, batch-size 设置为 2, 网络训练迭代 60 次。

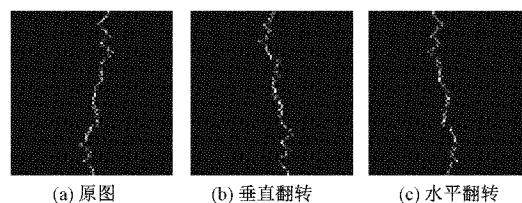


图8 数据增强

4.3 评价指标

本文采用一种用于 LOFAR 谱线检测效果的评估指标来衡量网络的性能。由 Pratt^[9] 提出的线位置准确率(line location accuracy, LLA), 评价指标定义如下:

$$LLA = \frac{1}{\max(|P_t|, |D_t|)} \times \sum_{(l,m) \in D_t} \frac{1}{1 + \lambda \min_{(i,j) \in P_t} (\| [l,m] - [i,j] \|^2)} \quad (9)$$

式中: P_t 表示预测谱线, D_t 是谱线标签, $\| [l,m] - [i,j] \|^2$ 表示检测到的谱线与实际谱线的欧氏距离。根据经验, 把 λ 设置为 1。

4.4 实验结果和分析

在信噪比为 $-24 \sim -20$ dB 下蒙特卡罗试验次数 100 次。将本文模型与 HMM 模型和 CEM 模型在其上进行测试, 并计算线谱检测准确率平均值。实验结果如图 9 所示。

信噪比为 $-20 \sim -24$ dB 时, HMM 模型的线谱检测准确率为 0.184 ~ 0.319, CEM 模型的线谱检测准确率为 0.194 ~ 0.358, 本文模型线谱检测准确率为 0.314 ~ 0.526。实验结果表明, 本文模型在低信噪比下线谱检测效果优于 HMM 模型和 CEM 模型。

图 10 是本文模型与 HMM 模型、CEM 模型在信噪比 -20 、 -22 、 -24 dB 下 LOFAR 谱图线谱检测结果比较。图中第 1 列分别是一 -20 、 -22 、 -24 dB 的待检测 LOFAR 谱图。第 2 列是 HMM 对应的 3 种信噪比下 LOFAR 谱图

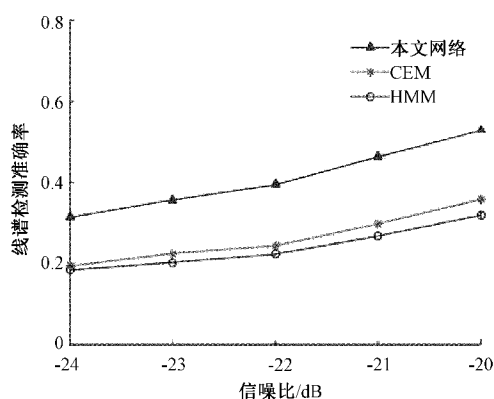


图9 不同模型的线谱位置准确率

检测结果,第3列是CEM对应的3种信噪比下LOFAR谱图检测结果,第4列是本文模型对3种信噪比下LOFAR谱图检测结果。实验的真值图如图8(a)所示。对比观察图中,CEM和HMM在信噪比为-20 dB时,谱线轮廓开始模糊。在-22 dB时,看不到谱线的变化趋势。而本文模型信噪比为-20和-22 dB时,线谱轮廓较为明显,谱线整体变化趋势依然和真值一致,而在信噪比为-24 dB时,由于背景噪声太强,网络已经难以提取到足够特征,谱线轮廓开始模糊。但是,本文模型已经远远优于HMM模型和CEM模型,提升了低信噪比下不规则波动谱线的检测准确率。

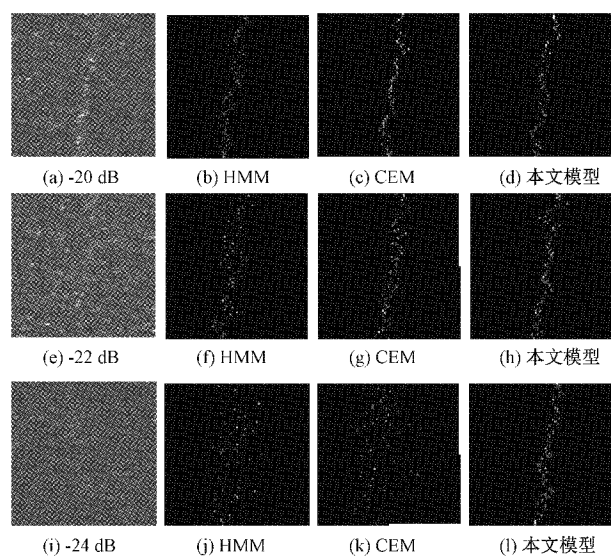


图10 不同信噪比 LOFAR 线谱检测效果比较

5 结 论

线谱检测是水声信号分析的重要手段之一,低信噪比下提取不规则波动的线谱对水声信号分析具有重大意义。为了有效检测低信噪比下不规则波动线谱,本文以 U-net 作为网络框架,结合残差结构,特征通道注意力机制,提出一种改进的深度卷积神经网络模型。该模型通过使用多个残差单元,SE 注意力机制和 DUpsampling 上采样操作改

进传统编码-解码模型,对低信噪比下的 LOFAR 谱图谱线具有很好的特征提取能力。在对比试验中,本文模型的表现比已有模型具有更好的线谱检测效果和泛化能力。下一步,考虑到水声 LOFAR 谱图数据集有限,打算利用无监督学习训练网络,进一步提升低信噪比下不规则波动谱线检测准确率。

参考文献

- [1] 袁延艺,蒋小勇,杜选民. 多级背景均衡在多波束 LOFAR 处理中的应用[J]. 声学技, 2012, 31(3): 300-303.
- [2] 宋振宇,丁勇鹏,赵秀丽,等. 基于 LOFAR 谱图的水下目标识别方法[J]. 海军航空工程学院学报, 2011, 26(3): 283-286.
- [3] 王朋云,廖育荣,倪淑燕. 低信噪比条件下的信噪比估计算法[J]. 电子测量技术, 2017, 40(9): 151-154.
- [4] 李山,王德俊,王海斌. 一种基于线谱特征函数提取 LOFAR 图线谱的方法[J]. 声学技术, 2016, 35(4): 373-377.
- [5] LUO X, SHEN Z. A sensing and tracking algorithm for multiple frequency line components in underwater acoustic signals[J]. Sensors, 2019, DOI:10.3390/s19224866.
- [6] LAMPERT T A, O'KEEFE S E M. A detailed investigation into low-level feature detection in spectrogram images[J]. Pattern Recognition, 2011, 44(9): 2076-2092.
- [7] LAMPERT T A, O'KEEFE S E M. On the detection of tracks in spectrogram images [J]. Pattern Recognition, 2013, 46(5): 1396-1408.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [9] 马燕,余海军,钟发生,等. 基于残差编解码网络的 CT 图像金属伪影校正[J]. 仪器仪表学报, 2020, 41(8):162-171.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.
- [12] 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, 2015: 234-241.

- [13] TIAN Z, HE T, SHEN C, et al. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3126-3135.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [15] 赵斐. 基于金字塔注意力机制的遥感图像语义分割[J]. 国外电子测量技术, 2019, 38(8): 156-160.
- [16] 白旭, 胡辉. 基于快速傅里叶变换和互相关的多频微弱信号重构法[J]. 电子测量与仪器学报, 2019, 33(3): 168-175.

作者简介

裴明, 硕士生, 主要研究方向为水声信号处理。

E-mail: pciming_234@163.com

陈阳(通信作者), 副教授, 主要研究向为水声阵列信号处理。

E-mail: chenyang_hcu@163.com

邓林红, 教授, 主要研究方向为呼吸医疗技术装备。

E-mail: dlh@cczu.edu.cn