

# 小样本情况基于深度学习的水下目标识别研究\*

梁 红 金磊磊 杨长生

(西北工业大学航海学院 西安 710072)

摘要:水下自动目标识别一直是具有挑战性的任务.针对海洋环境下目标图像数据获取困难,样本数量不足以训练深层神经网络这一问题,提出小样本情况下基于深度学习的水下图像识别方法.利用提出的改进中值滤波器抑制水下小样本集图像的脉冲噪声;然后,采用 ImageNet 图像数据集对搭建的深度卷积神经网络模型进行预训练;使用水下降噪图像对经过预训练的神经网络进行参数微调.利用海洋鱼类图像数据集对完成训练的卷积网络性能进行验证,取得 85.08% 的正确识别率.

关键词:水下目标;图像识别;深度学习;小样本;卷积神经网络

中图分类号:TP391.41

doi:10.3963/j.issn.2095-3844.2019.01.002

## 0 引 言

受海洋环境复杂性的影响,水下摄像机等设备采集到的水下图像往往边缘和细节模糊,目标与背景对比度低,同时图像易受噪声污染<sup>[1-2]</sup>,这使得水下图像识别成为一项非常具有挑战性的工作.传统识别方法通常使用人工提取的数据特征,配合不同的分类器实现目标识别<sup>[3-4]</sup>.这些人为的特征提取手段对某些特定的识别任务有较好的性能,但是提取有效特征需要专业的领域知识,并且这些方法大多泛化能力有限.

深度学习是模式识别领域一个新的研究热点.借助深度学习,多处理层组成的神经网络模型通过多层抽象来自动学习数据表征,避免了人工提取特征的困难<sup>[5]</sup>.自 2006 年以来,借助于大数据,卷积神经网络(convolutional neural networks, ConvNets)在空气中的光学图像识别方面迅速发展,已经取得了巨大成功<sup>[6-7]</sup>.但是,基于深度学习的水下图像识别研究还远远不够<sup>[8]</sup>.其中一个重要原因,是因为当数据集训练样本较少时,深层神经网络很容易发生过拟合<sup>[9]</sup>,网络性能

严重下降.而水下图像采集通常需要耗费大量人力物力,使得很难获得足够的水下图像数据用于深度神经网络的训练.

针对水下图像识别中,图像受噪声污染及图像数据量不足等问题,提出改进中值滤波图像降噪算法以及小样本情况下基于深层卷积神经网络的水下图像识别方法.更具体地,利用水下摄像机在开放水域捕捉到的鱼类图像数据集,通过一套有效的算法来解决水下鱼类识别问题.

## 1 图像降噪与图像识别理论

### 1.1 中值滤波器

实验表明,相机拍摄的图像通常受到脉冲噪声和高斯噪声的影响.噪声的存在给图像后续处理带来困难,对图像分割、特征提取、图像识别等都有直接影响.

中值滤波是基于排序统计理论的一种能够有效抑制脉冲噪声的非线性信号处理技术<sup>[10]</sup>.标准中值滤波器一般使用滑动窗口,从图像左上角开始依次获取图像的局部像素值,然后用局部像素中值替代滑窗中心位置处的像素值,直到遍历整

收稿日期:2018-11-20

梁红(1969—),女,博士,教授,主要研究领域为水下目标探测

\* 国家自然科学基金项目资助(61379007,61771398)

张图像. 该算法相对简单, 易于实现.

以上给出了中值滤波器针对单通道灰度图像的处理算法, 彩色 RGB(三通道)图像的处理算法与之类似, 即对每一通道分别作中值滤波处理.

## 1.2 卷积神经网络

ConvNets 是一种有效的分类识别模型, 近年来广泛应用于图像识别任务中. 典型的 ConvNets 结构是多个基本模块的多层堆叠, 包括卷积层、池化层和全连接层等. 输入原始数据, 每个模块依次将上一个层次的数据表征转换为更高层次和更抽象层次的描述, 实现了原始数据特征的自动提取. 对于识别任务, 高级别的表征更能反映目标类别的本质特征.

受到生物视觉机制启发, 结合自然图像特点, ConvNets 模型的设计采用以下核心理念: 局部连

接, 权重共享, 池化和多网络层的使用. 因此, 与相同层数的全连接神经网络相比, ConvNets 内部具有更少的连接和参数, 相对更易于训练<sup>[11]</sup>. 此外, 可以通过改变 ConvNets 的深度和宽度来调整模型的性能.

一个典型的 ConvNets 模型是 AlexNet, 见图 1. 该网络共有八层, 其中前五层执行卷积运算, 后三层是全连接层. 最后一层的输出将被送到 Softmax 分类器中, 得到输入数据在 1000 个类别上的概率分布情况. 在 2012 届 ImageNet 大型视觉识别挑战赛(ILSVRC)中, Alex 团队利用该模型取得图像识别错误率 15.3% 的成绩, 而排名第二的传统模型识别错误率达 26.2%. 结果表明, 深层卷积神经网络能够通过监督学习方式, 在大数据集上实现较高准确率的物体识别.

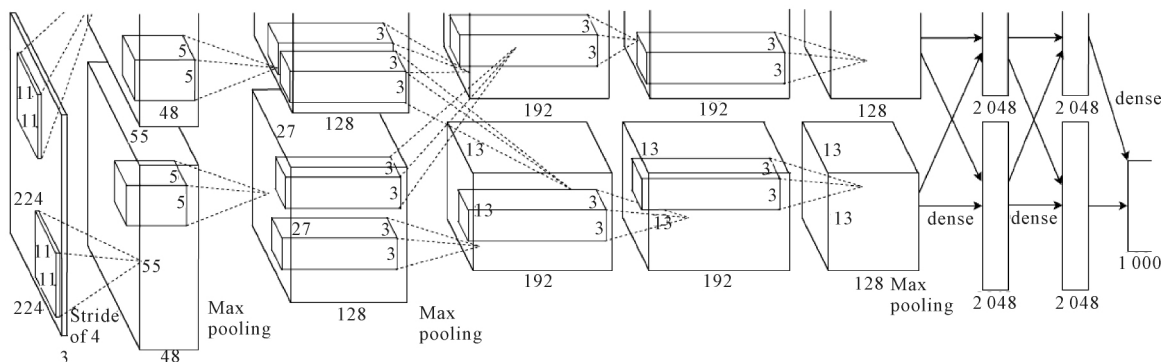


图 1 AlexNet 整体结构

## 2 小样本图像识别

### 2.1 改进的中值滤波器

标准中值滤波器会对原始图像中的每个像素点, 都进行中值滤波处理. 这样一来, 图像处理结果很大程度上将会受到设定的滤波器窗口大小的影响. 具体来说, 一个小的滤波窗口可以更好地保护原始图像的细节信息, 但是噪声抑制能力将受到限制; 一个大的滤波窗口可以增强噪声抑制能力, 但是原始图像的细节将变模糊. 为了尽量减小滤波窗口大小设定对图像处理结果的影响, 提出改进中值滤波器. 不同于标准中值滤波器处理图像中的所有像素, 改进中值滤波器通过设定阈值, 只处理被认为受脉冲噪声污染的像素. 减少滤波窗口处理的像素数量, 也就减少了滤波窗口大小设定对图像处理结果的影响. 以灰度图像为例, 改进中值滤波算法分为两个主要步骤.

**步骤 1** 根据受污染像素与其周围未受噪声污染像素的值相差较大这一先验信息来检测被脉

冲噪声污染的图像像素.

1) 脉冲噪声主要将像素真值改变为 0 或 255, 所以我们只关心那些值为 0 或 255 的像素, 其他像素被认为未受脉冲噪声污染.

2) 使用  $5 \times 5$  标准中值滤波器对整个原始图像进行处理, 得到粗略降噪图像. 求原始图像与粗略降噪图像对应像素值的差值, 将全部差值的均值作为阈值来决定我们感兴趣的像素是否为噪声污染像素. 具体而言, 如果原始图像与粗略降噪图像对应像素点的差值大于阈值, 则该像素被认为受到噪声污染, 否则认为该像素未受到噪声污染.

**步骤 2** 分别处理受脉冲噪声污染的像素点. 即对原始图像, 使用  $5 \times 5$  中值滤波窗口依次处理受污染的像素, 保留未受污染像素, 得到精细降噪图像.

以上给出了改进中值滤波器针对单通道灰度图像的处理算法, 该改进算法同样可以应用于彩色 RGB 图像的降噪中, 只需要对彩色图像的三个通道分别作改进中值滤波处理即可.

为定量评价降噪效果, 使用式(1)计算 RGB

图像的峰值信噪比(peak signal to noise ratio, PSNR):

$$PSNR = 10 \lg \left[ \frac{255^2}{(3 |X|)^{-1} \sum_{c=R,G,B} \sum_{x \in X} (y_c(x) - \hat{y}_c(x))^2} \right] \quad (1)$$

式中:  $x \in X$  为图片中像素点的空间坐标, 下标  $c \in \{R, G, B\}$  表示 RGB 图像的三个通道;  $y$  为原始图像像素值;  $\hat{y}$  为降噪图像像素值. PSNR 的值越大, 表示图像降噪效果越好.

## 2.2 卷积神经网络训练策略

深层神经网络通常有数百万的连接和权值. 虽然 ConvNets 采用了局部连接和权值共享理念, 仍然有大量的权值需要通过训练进行调整. 当 ConvNets 应用于大型数据集时, 能够取得很好的结果. 然而, 训练集样本数量较少时, 网络易发生过拟合. 所谓过拟合, 即卷积网络对训练集数据能取得较高的识别准确率, 但是对测试集数据的识别准确率很低. 对于水下图像识别任务, 一个实际问题是无法获得足够的样本图像来训练神经网络. 借鉴深度信念网络(deep belief networks, DBNs)的训练过程<sup>[12]</sup>, 将卷积神经网络的训练同样分为预训练和微调两步, 以解决小样本情况网络过拟合的问题, 从而提出一种小样本情况下基于卷积神经网络的水下图像识别方法. 方法分为三个主要阶段.

**步骤 1** 使用大数据集预先训练 ConvNets. ConvNets 不仅是分类器, 也是特征提取器, 其隐含层以利于预测目标类别的方式来学习并表示网络的输入, 经过预训练的 ConvNets 将对图像的

颜色、纹理和边缘等信息敏感. 自然图像具有一定的共性, 使用与我们期望识别的类别不一样的图像来预训练网络是合理的. ImageNet 图像数据集为我们提供了一个这样的大数据集, 它有超过 1 500 万个带标签的高分辨率图像, 属于大约 22 000 个类别<sup>[13]</sup>.

**步骤 2** 使用获取的小样本图像数据集对第一阶段训练完的 ConvNets 微调. 多层架构的卷积网络可以通过随机梯度下降优化算法来进一步训练, 实现参数的微调整. 此时, 分类器的输出个数需要修改为小样本图像数据集的目标类别数目.

**步骤 3** 最后一个阶段是对微调后的 ConvNets 识别准确率进行测试. 如果准确率达到应用要求, 则可将训练好的 ConvNets 用于该图像识别任务中.

## 3 水下图像识别实验

### 3.1 鱼类数据集

采用 Fish4Knowledge 项目<sup>[14]</sup> 在开放海域采集的鱼类图像数据进行小样本情况水下图像识别方法实验验证, 海洋鱼类图像见图 2, 在 10 类代表性鱼类图片的下面分别标有类别名称和相应的样本数量. 图 2 的图像是较为理想的, 许多图像受噪声污染质量较差. 该数据集是非常不平衡的, 10 类数据中, 样本数最多的有 450 张图像, 而样本数最少的只有 90 张图像. 从样本总数来看, 很难直接利用深度学习方法获得高的图像识别率.

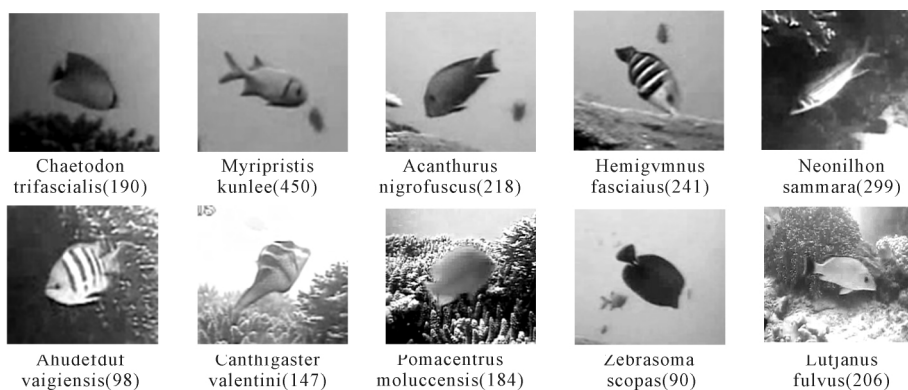


图 2 10 类海洋鱼类图像

### 3.2 图像降噪

图像降噪可以为后续的鱼类识别任务减少干扰. 为了克服标准中值滤波器的缺点, 采用本文提出的改进中值滤波器对图像进行噪声抑制. 图 3

分别是采用标准中值滤波器和改进中值滤波器对含噪鱼类图像的处理结果.

由图 3 可知, 标准中值滤波器的图像处理结果直接受滤波器窗口大小的影响. 相比于标准中

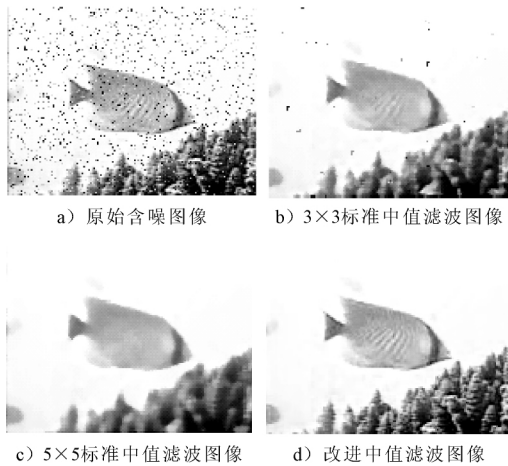


图 3 标准中值滤波与改进中值滤波图像降噪结果

值滤波算法,改进中值滤波算法不仅可以有效地抑制脉冲噪声,同时很好地保留原始图像细节。

为了进行定量比较,我们根据式(1)计算了三种不同方法处理后的图像的 PSNR,见表 1。通过比较最后一列数值与前两列数值,可以看到改进算法对图像峰值信噪比的提高是实质性的,相比于标准中值滤波器高出 2 dB 左右。

表 1 标准及改进中值滤波降噪图像峰值信噪比

降噪方法	3×3 中值滤波	5×5 中值滤波	改进中值滤波
PSNR/dB	22.40	21.25	24.36

### 3.3 卷积神经网络模型

ConvNets 的架构包含四个关键思想:局部连接、共享权值、池化以及使用多个层。本文实验使用的 ConvNets 模型的总体结构见图 4,近似于 AlexNet 模型。图 4 中标注了不同层的定义,前五层是卷积层,其余三层是全连接层,输出层的输出类别数目在不同的训练阶段不相同。采取零均值,标准差为 0.01 的高斯分布来初始化每个卷积层中最初的权重。

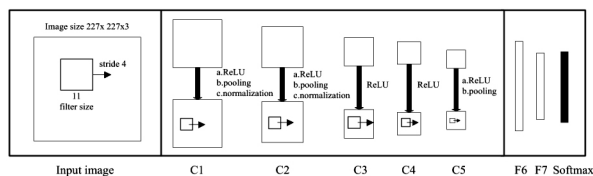


图 4 本文使用的 ConvNets 模型

### 3.4 卷积神经网络训练和测试

文中卷积神经网络的训练分为两个阶段,利用 Caffe 深度学习框架<sup>[15]</sup>完成网络训练和测试。在第一阶段,使用从 ImageNet 数据集获取的图像对网络进行预训练,预训练数据集包含 1 000 类物体,每个类别有 1 000 张图像。经过预训练,ConvNets 的网络权重等参数实现了恰当的初始化,网络初步具备了对图像信息的表达能力。

在第二阶段,将降噪预处理后的鱼类图像尺寸调整为  $227 \times 227$ ,并将整个鱼类数据集分为三个子集:训练集 500 张图像(10 个类别各 50 张图像),验证集 200 张图像(10 个类别各 20 张图像),其余的 1 420 张图像作测试。

最后一层输出神经元数目从阶段一的 1 000 改为 10,并将输出结果馈送到 10 路输出的 Softmax 分类器中,得到输入图像在 10 个类别上的概率分布情况。采用小批量梯度(mini-batch gradient descent, MBGD)下降算法(批大小设置为 40),通过对分类误差的反向传播来微调经过预训练的神经网络参数。基础学习率设为 0.001,同时设置 0.005 的权重衰减和 0.9 的动量系数。网络前七层的学习速率相等,最后一层的学习率设置为它们的 10 倍。每训练迭代 1 000 次,整体学习率下降为之前的一半。

经过 5 000 次迭代,完成 ConvNets 的微调。最终,在验证集上取得 85.50% 的平均识别准确率,在测试集上达到 85.08% 的平均识别准确率。图 5 为微调阶段训练损失随训练迭代次数的变化曲线,图 6 为验证准确率随训练迭代次数的变化曲线,证明了所提卷积神经网络训练策略对于小样本情况下的图像识别是有效的。

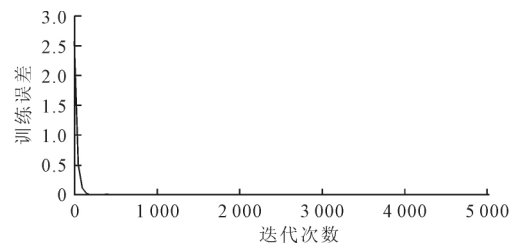


图 5 微调阶段训练误差-迭代次数曲线

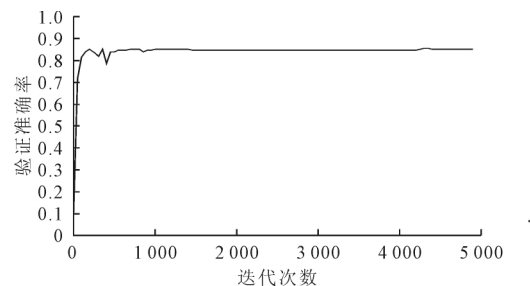


图 6 微调阶段验证准确率-迭代次数曲线

## 4 结束语

提出了一种改进的中值滤波器以及一种有效的小样本情况下基于卷积神经网络的水下目标识别方法,并通过鱼类图像数据集进行了实验验证。不同于传统水下目标识别方法,目标特征是从训

练数据中通过监督学习自动得到的,因此不需要专业的关于鱼类或图像特征提取的有关知识. 本文方法对解决其他水下目标识别问题提供了重要参考,对深度学习理论应用于水下目标识别具有促进作用,并且在海洋生物调查和水产养殖等方面具有广阔的应用前景.

### 参考文献

- [1] 乔曦. 基于水下机器视觉的海参实时识别研究[D]. 北京: 中国农业大学, 2017.
- [2] 汤春瑞. 水下目标声图像多分辨率分析及识别研究[D]. 哈尔滨: 哈尔滨工程大学, 2009.
- [3] FEI T, KRAUS D, ZOUBIR A M. Contributions to automatic target recognition systems for underwater mine classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2014, 53(1): 505-518.
- [4] 石敏, 徐袭. 小波包熵在水下目标识别中的应用研究[J]. 计算机工程与应用, 2014, 50(1): 215-217.
- [5] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015(2): 436-444.
- [6] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes[C]. IEEE Conference on Computer Vision and Pattern Recognition, Washington D C, 2014.
- [7] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[J]. Nature, 2016(1): 55-59.
- [8] QIN H, LI X, LIANG J, et al. Deep fish: accurate underwater live fish recognition with a deep architecture[J]. Neurocomputing, 2016(1): 49-58.
- [9] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [10] ATAMAN E, AATRE V K, WONG K M. Some statistical properties of median filters[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1981, 29(5): 1073-1075.
- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems, London, 2012.
- [12] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006(2): 504-507.
- [13] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR2009, Miami, 2009.
- [14] BOOM B J, HUANG P X, HE J, et al. Supporting ground-truth annotation of image datasets using clustering[C]. 21st International Conference on Pattern Recognition, Tsukuba, 2012.
- [15] JIA Y Q, SHELHAMER L. Caffe: convolutional architecture for fast feature embedding[J]. Eprint Arxiv, 2014(2): 675-678.

## Research on Underwater Target Recognition Based on Depth Learning with Small Sample

LIANG Hong JIN Leilei YANG Changsheng

(School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

**Abstract:** Automatic underwater target recognition has always been a challenging task. Aiming at the problem that it is difficult to acquire target image data in marine environment and the number of samples is not enough to train deep neural networks, an underwater image recognition method based on depth learning in the case of small samples is proposed. Firstly, the proposed improved median filter was used to suppress the impulse noise of underwater small sample set images. Then, ImageNet image data set was used to pre-train the built deep convolution neural network model. Finally, the pre-trained neural network was fine-tuned with underwater noise reduction images. The performance of the trained convolution network was verified by using the marine fish image data set, and the correct recognition rate was 85.08 %.

**Key words:** underwater targets; image recognition; in-depth learning; small sample; convolutional neural networks