

# 基于深度学习的海底观测视频中鱼类的识别方法

张俊龙<sup>1</sup>, 曾国荪<sup>1\*</sup>, 覃如符<sup>2,3</sup>

(1. 同济大学 电子与信息工程学院, 上海 200092; 2. 同济大学 海洋与地球科学学院, 上海 200092;

3. 海洋地质国家重点实验室(同济大学), 上海 200092)

(\* 通信作者电子邮箱 zjl\_ouc@163.com)

**摘 要:** 针对海底环境恶劣、海底观测视频品质差导致视频中的海洋鱼类识别难的问题, 提出一种基于深度学习的海洋鱼类识别方法。首先, 将海底观测视频分解为图片, 由于海底观测视频中存在较大比例的空白数据, 使用背景差分法过滤不包含鱼类的图片, 缩短处理全部数据的时间; 然后, 考虑到海底拍摄环境亮度低、场景模糊的实际情况, 对图片基于暗通道先验算法进行预处理提高品质; 最后, 以卷积神经网络(CNN)为基础构建深度学习模型, 并且提出了权重化特征的卷积过程, 提高模型的鲁棒性。实验结果表明: 面对较差品质的海底观测视频图片, 在深度学习模型结构相同的条件下, 与普通卷积神经网络模型相比, 使用权重化卷积作为隐层并且加入预处理过程后, 对海洋鱼类识别准确率的提升幅度达到 23%, 有助于实现对海底观测视频图片中海洋鱼类的精准识别。

**关键词:** 海底观测; 视频图片; 图片品质; 深度学习; 鱼类识别

**中图分类号:** TP391.4      **文献标志码:** A

## Fish recognition method for submarine observation video based on deep learning

ZHANG Junlong<sup>1</sup>, ZENG Guosun<sup>1\*</sup>, QIN Rufu<sup>2,3</sup>

(1. College of Electronic and Information, Tongji University, Shanghai 200092, China;

2. School of Ocean and Earth Science, Tongji University, Shanghai 200092, China;

3. State Key Laboratory of Marine Geology (Tongji University), Tongji University, Shanghai 200092, China)

**Abstract:** As it is hard to recognize marine fishes occurred in submarine observation videos due to the bad undersea environment and low quality of the video, a recognition method based on deep learning was proposed. Firstly, the video was split into pictures, and as this type of video contains a large proportion of useless data, a background subtraction algorithm was used to filter the pictures without fish to save the time of processing all data. Then, considering the undersea environment is blurring with low bright, based on the dark channel prior algorithm, the pictures were preprocessed to improve their quality before recognition. Finally, a recognition deep learning model based on Convolutional Neural Network (CNN) was constructed with weighted convolution process to improve the robustness of the model. The experimental results show that, facing submarine observation video frames with poor quality, compared with traditional CNN, the method with preprocessing and weighted convolution as hidden layer can increase the recognition accuracy by 23%, contributing to the recognition of marine fishes in submarine observation video.

**Key words:** submarine observation; video picture; picture quality; deep learning; fish recognition

## 0 引言

近年来, 世界各国对于海洋权益、海洋资源的重视程度与日俱增, 海洋资源中重要的一项就是渔业资源, 目前我国的渔业资源呈衰退趋势<sup>[1]</sup>。对代表性海洋鱼类的习性或分布状况的研究对于指导海洋捕捞业可持续发展、保护海洋生态平衡有着重要的意义。

对于海洋鱼类的分布情况, 传统的研究方法以出海捕捞为主, 使用延绳钓探捕、拖网探捕等常用海洋捕捞技术。随着科技的进步, 不仅渔具等硬件设备有了很大的改善<sup>[2]</sup>, 而且伴随着自动化技术的应用, 捕捞的效率也有了很大提升。这种传统调研方式固然有效, 但是对于相关科研人员来说, 调研过程耗时耗力, 一旦出海就会对研究进度造成较大影响; 并且

由于调研过程耗费资源太多, 不能频繁实施, 研究人员大多数情况下只能在历史调研数据的基础上作研究, 这会严重影响研究成果的质量。针对上述问题, 国家于“十二五”期间提出将全面推动国家海底观测平台的建设, 其中就包括通过部署水下摄像器材来实时监控关键海洋生物的重要任务, 这样就可以通过分析观测视频来代替出海调研, 而且保证了获得数据的实时性, 能够极大地提高科研人员的研究效率。传统的鱼类识别研究多基于声呐系统, 对鱼体回波信号进行处理分析, 从声学信号中提取适于分类的特征, 如许枫等<sup>[3]</sup>的工作, 但是这类方法已不适用于当前的视频数据。由于利用水下摄像器材观测鱼类(特别是海洋鱼类)的工作是最近几年才开始推进的, 因此关于识别视频中鱼类的研究较少, 且研究对象多以环境条件好的淡水养殖场场景下的淡水鱼为主。张志强等<sup>[4]</sup>

收稿日期: 2018-07-02; 修回日期: 2018-09-05; 录用日期: 2018-09-12。

基金项目: 国家社会科学基金资助项目(17BQ1086); 同济大学实验教改项目(0800104214)。

作者简介: 张俊龙(1994—), 男, 山东淄博人, 硕士研究生, 主要研究方向: 可信搜索、深度学习; 曾国荪(1964—), 男, 江西吉安人, 教授, 博士, 主要研究方向: 智能搜索、可信网络软件、并行分布计算; 覃如符(1979—), 男, 广西柳州人, 副教授, 博士, 主要研究方向: 地理信息系统、海洋信息技术、海底观测系统。

提取鱼类图像中的各个颜色分量及长短轴之比作为分类特征,该方法提取的特征过于一般化,只适用于识别类别单一的淡水鱼。姚润璐等<sup>[5]</sup>则从图像中分割出鱼背、鱼尾等部位的图像块,以此为基础提取相关性更强的特征;该方法提取的特征与鱼类相关性较强,但是提取过程较为复杂,需要人工制定鱼类各个部位的匹配规则才能分割出对应图像块,同样不适用于分析海量数据。林明旺<sup>[6]</sup>利用特征提取能力强大、可以充分利用海量数据的深度学习来识别鱼类图像;但是仅仅构建了一个卷积神经网络(Convolutional Neural Network, CNN)模型,没有结合实际应用环境考虑拍摄环境的特点、数据本身的特点等因素。

综上所述,过于泛化的特征质量低下,过于精确的特征提取过程困难,而能够充分利用海量数据、自动提取高质量的特征的深度学习方法是目前最佳的解决方案。本文使用深度学习模型作为分类模型。同时,现有工作研究的主要是淡水鱼类,且不关心鱼类图像的拍摄环境等影响因素,在本文的应用场景中,针对海底环境恶劣导致拍摄图像品质差的问题,本文还将采用预处理方法降低环境因素对识别结果的影响。

## 1 海底观测视频鱼类识别系统

### 1.1 系统功能

海底观测平台的系统结构如图1所示:位于海面以下的水下接驳器统一接收来自各个传感器的采集数据,包括水下摄像器材采集的视频数据,然后将数据传输至位于陆地上的岸基站;岸基站接收并缓存所有来自海底的数据,按约定的协议和规则转发给大数据中心;大数据中心由多个子系统构成,负责对不同类型数据的转化、存储、处理、分析,其中包括本文将要研究的海洋鱼类识别系统,负责对海底观测视频的处理分析。本文对视频的处理过程以图片为单位,在接收到海底观测视频数据后首先将视频分解成帧,然后针对每一帧进行必要的预处理,使其能够满足识别过程对于图片质量的需求,之后使用基于深度学习的方法对图片进行识别。此外,基于鱼类在海洋中分布稀疏这一基本事实,为了简化分析,本文假设一张图片中至多存在一条鱼类。

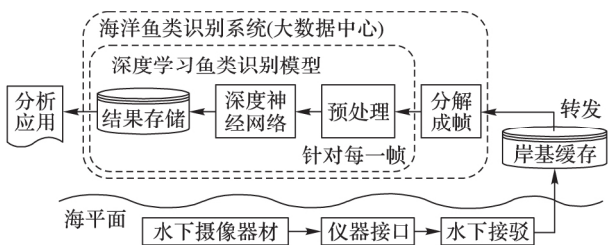


图1 海底观测平台

Fig. 1 Submarine observation platform

### 1.2 海底观测视频数据特征

海底观测视频拍摄于海洋中,视频分辨率为 $1280 \times 720$ ,即宽 $W = 1280$ 、高 $H = 720$ ,帧率为50 fps。由于本文对视频的处理分析以图片为单位,采用开源计算机视觉工具软件OpenCV来读取视频数据,并且将其分解为图片,每一张图片的本质是 $1280 \times 720$ 的彩色图像。

在计算机中,图像的本质是 $W \times H$ 个像素点构成的像素点阵,每个像素点的灰度值决定了该点的颜色,对应的计算机内部表示就是一个 $H$ 行 $W$ 列的矩阵。在本文中,用一个矩阵 $I_{720 \times 1280}$ 表示一张图片,矩阵中每个元素 $p_{i,j}$ 的值代表位于 $(i,$

$j)$ 位置的像素点的灰度值。在本文中,图片为彩色,像素点的灰度值由3个字节表示,从高位到低位的每个字节分别对应了红、绿、蓝三个颜色通道,每个字节称为一个“灰度单元”。由于下文将以灰度单元为单位对图片进行处理,因此,将 $p_{i,j}$ 视为长度为3的数组,即 $p_{i,j}[k](k = 1, 2, 3)$ , $p_{i,j}[k]$ 表示其中一个灰度单元。

## 2 视频图片的预处理

### 2.1 过滤空白数据

通常,海底观测视频中存在较大比例的没有鱼类的空白数据,对空白数据的识别浪费时间和计算资源,因此有必要使用计算量小、速度快的算法对这些没用的、无价值的空白数据先行过滤。本文采用背景差分法<sup>[7]</sup>,选取一空白图片作为参考图片,对应矩阵 $I^*$ ,对于之后待处理的每一张图片,对应矩阵 $I$ ,都计算与参考图片之间的差分 $d = \sum_{i=1}^{1280} \sum_{j=1}^{720} \sum_{k=1}^3 |p_{i,j}[k] - p_{i,j}^*[k]|$ ,若 $d$ 不超过指定阈值 $\tau$ ,则判断为空白图像并丢弃,否则保留。显然,该方法过滤效果的关键在于阈值 $\tau$ 的选择,若 $\tau$ 取值过低则会导致错误地保留了空白数据,即假阳性错误;反之,则会导致错误地丢弃了有效数据,即假阴性错误。

一方面,时长1 s的有效视频段就可以分解为50张有效图片,有效图片有着足量的冗余,对假阴性错误的容忍度较高;另一方面,为了提高对海量数据的处理速度,要尽可能减少需要识别的图片的数量,所以对假阳性错误的容忍度较低。因此,本文将使用较高的阈值。对于空白参考图片,选取100张不包含鱼类的空白图片,对应的矩阵为 $I^1, I^2, \dots, I^{100}$ ,进行平均计算,即 $p_{i,j}^*[k] = \frac{1}{100} \sum_{l=1}^{100} p_{i,j}^l[k]$ 。同样,使用了100张包含完整鱼类的图片分别与参考帧计算差分 $d$ ,以平均值作为 $\tau$ 值,即 $\tau = \frac{1}{100} \sum_{i=1}^{100} d_i$ 。显然,使用这样的 $\tau$ 值会使得一部分包含不完整鱼类的有效图片也被过滤掉,但是这种假阴性错误是可以容忍的。因为对每一张图片过滤的时间远小于接下来的处理过程的时间总和,所以先对空白数据进行过滤,可以大大缩短处理全部数据的时间。

### 2.2 提高图片品质的预处理

海底观测视频拍摄于海底,拍摄环境光照条件差、整体亮度低,水中杂质对光的散射作用也会使画面模糊不清,这些因素都会影响识别的效果。因此,本文先使用提高亮度、提高清晰度这两步预处理方法,尽可能消除亮度低和模糊带来的影响。

对于提高亮度的方法,本文选择伽马校正法<sup>[8]</sup>,使用指数函数调整每个灰度单元,计算公式如下:

$$p_{i,j}[k] = p_{i,j}[k]^\gamma \quad (1)$$

亮度的高低由作为指数的参数 $\gamma$ 来控制, $\gamma$ 的取值视应用环境的整体亮度而定,将在实验部分说明。

对于提高清晰度的方法,本文参考的是基于暗通道先验算法去雾的研究<sup>[9]</sup>,该研究认为雾天拍摄图片的模糊是由空气中的杂质对光的散射造成的。本文中,水下拍摄图片的模糊也恰好是由水中的杂质对光的散射造成的,于是在该算法的基础上,针对水下环境的特点作了改进。算法对每个灰度单元进行优化,计算公式如下:

$$p_{i,j}[k] = \frac{p_{i,j}[k] + a[k]}{\max(t_{i,j}, \varepsilon) + a[k]} \quad (2)$$

式中:  $t_{i,j}$  为矩阵  $T_{720 \times 1280}$  中  $(i, j)$  位置的元素, 矩阵  $T$  表示光的穿透能力;  $a[k]$  为数组  $a[3]$  的第  $k$  个元素, 数组  $a[3]$  表示环境的亮度;  $\varepsilon$  为一个接近于 0 的正数, 避免分母为 0 的情况出现。显然, 等号右侧的矩阵  $T$  与数组  $a$  都是需要计算的未知量。

文献[9]提出了通过计算图像的暗通道  $D$ , 可以有效地求解未知量  $T$  和  $a$ 。其中  $D$  是由满足  $\min_{1 \leq k \leq 3} (p_{i,j}[k])$  的灰度单元组合而成的矩阵经由最小值滤波后得到的, 详细计算过程可参考文献[9]。对于未知量  $T$ , 文献[9]使用平滑后的  $D$  计算。对于未知量  $a$ , 由于陆地上自然光的亮度分布均匀, 所以文献[9]认为  $D$  中最亮点(最小值点)的值可以反映出整个环境的光值。但是, 本文的水下环境与陆地环境不同, 自然光的亮度分布不均, 还会受照明器材的干扰而存在多处亮斑,  $D$  中的最亮点可能只属于场景中某处极小的亮斑, 不能反映出整个环境的光值, 所以选取前 10% 个最亮点。本文中共 92 160 个点, 得到坐标集合  $\{(x_1, y_1), (x_2, y_2), \dots, (x_{92160}, y_{92160})\}$ , 然后令  $a[k] = \frac{1}{92160} \sum_{i=1}^{92160} p_{x_i, y_i}[k]$ 。综合上述两步分别提高了亮度和清晰度过程, 预处理过程的算法如下:

算法 1 预处理 Preprocess。

输入 过滤得到的图片集合  $F$ ;

输出 提高品质后的图片集合  $F'$ 。

```
Preprocess ( F )
{
  forall  $f \in F$ 
  {
     $f \leftarrow \text{enhanceBrightness}(f, \gamma)$ ; //使用式(1)提高亮度
     $D \leftarrow \text{computeDark}(f)$ ;
     $C \leftarrow \text{getMaxBrightCoords}(D, 0.1)$ ; //10%个最亮点坐标
     $a \leftarrow \text{computeA}(f, C)$ ;
     $D \leftarrow \text{smooth}(D)$ ;
     $T \leftarrow \text{computeT}(D)$ ;
     $f \leftarrow \text{enhanceDefinition}(f, a, T)$ ; //使用式(2)提高清晰度
     $F' \leftarrow F' + \{f\}$ ;
  }
  return  $F'$ ;
}
```

### 2.3 海洋鱼类图片数据的收集

本文拟采用深度学习开展对图片中海洋鱼类的识别工作, 采用有监督学习的方式训练深度学习模型。然而, 目前我国海底观测平台项目刚刚起步, 还没有积累下来有效的、可用的、真实的海洋生物图片数据。因此, 为了获取海量的学习训练数据, 本文使用爬虫技术, 从互联网上搜索下载海洋鱼类的图片数据, 从而解决学习训练数据“从无到有”的问题。理论上, 所有海洋中的鱼类都可作为识别对象, 本文以东海常见鱼类作为识别对象, 也可以推广至其他海域的鱼类。综合文献[10]和文献[11]的调研结果, 东海中常见的鱼类共有 30 种左右, 为了简化下文的论述, 不妨选取其中 10 类作为本文的识别对象: 海鳗、带鱼、龙头鱼、鲐鱼、马鲛、鲟、马面鲀、棕斑腹刺鲀、黄条鲷、短尾大眼鲷。本文以 Google Images 为目标网站, 使用网站提供的应用程序接口(Application Programming Interface, API)进行爬取。

搜索结果中存在内容上无关的或重复的图片, 在剔除这些无用数据后, 最终保证每种鱼类收集 700 张有效图片。此

外, 还收集了 700 张其他鱼类的图片, 这是为了使模型在无法确定图片中的鱼类属于上述 10 个类别时, 也能合理地将识别为“其他”类别。

## 3 基于深度学习的海洋鱼类识别

### 3.1 卷积神经网络

自深度学习被广泛研究以来, 应对不同任务的不同结构的神经网络模型层出不穷, 其中 CNN 适用于图像处理, 它模仿了人类大脑皮层中的视觉细胞, 使用卷积核对图像进行卷积<sup>[12]</sup>。使用  $N$  个卷积核意味着提取  $N$  种特征, 得到  $N$  个特征图层。CNN 还具有参数共享的特点, 其参数量仅与卷积核尺度  $K$  和数量  $N$  有关, 参数总个数为  $N \times (K \times K + 1)$ 。CNN 的特点决定了它非常适合应用到图像处理的任任务中, 本文将以 CNN 为基础搭建深度学习模型。

### 3.2 海洋鱼类识别深度学习模型

#### 3.2.1 模型的输入与输出

在构建具体的模型结构前, 有必要明确模型的输入与输出。本文图像的分辨率为  $1280 \times 720$ , 为了使模型中间结果的数据格式更为规整, 这里将图像统一缩放为  $1024 \times 1024$  的分辨率, 于是模型输入实质便是表示图像的  $1024 \times 1024$  的矩阵  $I$ , 意味着模型的输入层有  $1024 \times 1024$  个变量, 每个变量对应矩阵  $I$  中的一个元素  $p_{i,j}$ 。

对于模型的输出, 常用的分类器有支持向量机(Support Vector Machine, SVM)、Softmax 等, 本文选择适合多分类任务的 softmax 分类器。因为本文中有 11 个类别: 10 个鱼类、1 个其他类, 因此, Softmax 分类器将读入一个 11 维的向量  $v$ , 通常  $v$  是模型最后一层全连接的输出。之后, 使用公式  $c_i = \frac{e^{v_i}}{\sum_{j=1}^{11} e^{v_j}}$

计算同样是 11 维的结果向量  $c$ ,  $c$  中每个分量满足  $c_i > 0$ , 且  $\sum_{i=1}^{11} c_i = 1$ , 符合概率分布的定义, 其中每一个分量  $c_i$  的含义是模型“判断”输入图像的种类是第  $i$  类的概率。令第 1 类为其他类, 令第 2 类至第 11 类分别为: 海鳗、带鱼、龙头鱼、鲐鱼、马鲛、鲟、马面鲀、棕斑腹刺鲀、黄条鲷、短尾大眼鲷。显然,  $c$  中的最大分量对应的类别就是预测类别。举例来说, 假设结果向量  $c = (0.00, 1.00, 0.00, \dots, 0.00)$ , 其中最大概率为  $c_2 = 1.00$ , 于是预测类别就是第 2 类海鳗。

#### 3.2.2 权重化特征的卷积

在图片输入深度学习模型前, 虽然已通过预处理提高了图片品质, 但是由于真实情况难以预测, 实际环境可能很极端, 导致视频图片品质很差, 预处理方法不足以将图片品质提高到对识别结果没有影响的水平, 因此, 有必要增强深度学习模型本身的鲁棒性。

在卷积中, 使用卷积核提取特征, 这些特征可以是与鱼类相关的特征, 如: 鱼鳍的位置、鱼鳍的个数、鱼尾的形状、体色等; 也可以是与鱼类无关的特征, 如: 背景颜色、环境亮度等。通常, 每一种特征对分类结果都有贡献, 贡献的强弱是在模型的训练过程中通过优化每一个参数的参数值而自发确定的。本文对卷积过程加以改进, 进一步提高极端环境下关键特征的贡献, 抑制非关键特征的贡献, 放大二者贡献度之间的差距, 提高极端环境下卷积结果的质量, 从而提高模型对极端环境的鲁棒性。主要思想是对提取到的特征赋以权重, 具体做

法如下: 当对输入数据卷积得到特征图层之后, 先基于特征图层生成特征权重, 再将特征图层与对应的特征权重相乘后作为输出。这样一来, 在训练过程中, 模型可以不断学习, 直至为每个特征生成最为合适的权重。假设输入是  $M$  个特征图层, 本质是  $M$  个  $W \times H$  的矩阵。经  $N$  个卷积核卷积后得到  $N$  个特征图层, 本质是  $N$  个  $W \times H$  的矩阵。使用全局平均池化 (取每个特征图层中所有数值的平均值作为输出) 得到池化结果, 本质是  $N$  维向量。然后经过 2 层  $N$  个神经元的全连接, 对池化结果进行充分的非线性化映射, 最终得到  $N$  个特征权重, 本质是  $N$  维向量。将  $N$  个特征权重与  $N$  个特征图层一一对应相乘, 即可实现特征权重化。整个过程如图 2 所示。

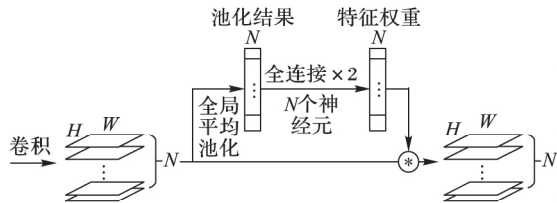


图2 权重化特征的卷积

Fig. 2 Convolution with weighted features

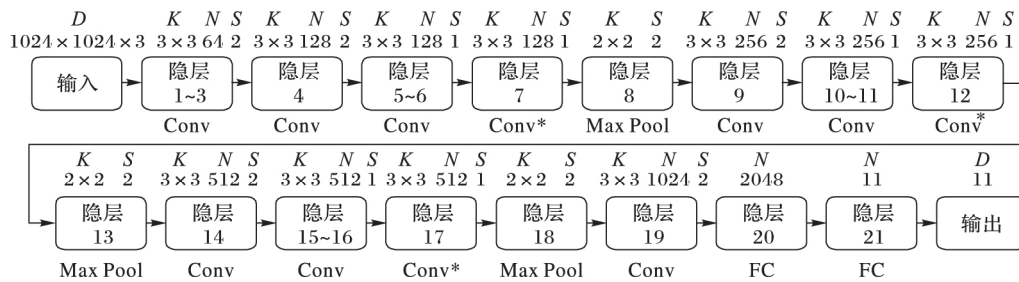


图3 深度学习鱼类识别模型结构

Fig. 3 Deep learning model structure for fish recognition

### 3.2.4 损失函数与参数更新方法

模型学习的过程是根据反馈不断更新参数的过程, 反馈是通过损失函数来量化的, 更新参数的依据是损失函数在对应参数上的偏导数。

本文使用 Softmax 分类器, 输出为 11 维的结果向量  $c$ 。为了计算预测值与真实值之间的差异, 要将每张训练图片对应的真实类别转换成 11 维的向量  $t$ , 其中对应真实类别的分量为 1, 其余分量为 0。对于损失函数, 使用的是交叉熵损失函数, 该函数衡量的是两个概率分布之间的差异, 计算公式如下:

$$loss = - \sum_{i=1}^{11} t_i \ln c_i \quad (3)$$

式(3)中, 损失值  $loss$  越大, 意味着预测概率分布与真实概率分布的差异越大, 预测结果越不准确; 反之, 则意味着预测概率分布与真实概率分布的差异越小, 预测结果越准确。

在得到损失值后, 根据反向传播算法, 逐层计算损失函数在各个权重上的偏导数  $\frac{\partial l}{\partial w}$ 。选择动量法<sup>[13]</sup>作为参数更新方法, 该方法仿照物理学上物体存在摩擦力情况下的运动规律来模拟寻找最优解。假设参数更新的速度是  $v$ , 摩擦力就是偏导数  $\frac{\partial l}{\partial w}$ , 使用受摩擦力影响后的  $v$  更新参数, 计算公式如下:

$$v = m \times v - r \times \frac{\partial l}{\partial w} \quad (4)$$

其中的关键是 2 层全连接, 它们基于特征图层的池化结果生成对应的特征权重。在训练过程中, 它们将通过调整神经元中的参数, 为不同的特征生成合适的权重, 于是随着层数的增加, 非关键特征的贡献会越来越弱, 直至对分类结果没有任何影响。但是, 全连接会引入大量参数, 大大增加整个模型的训练计算量, 因此, 本文将混合使用卷积与权重化特征的卷积。

### 3.2.3 模型的结构

确定了模型的输入与输出后, 就可以确定模型结构。模型结构如图 3 所示, 使用圆角矩形表示一层神经网络, 类型标注在圆角矩形下方, 关键参数标注在圆角矩形上方。使用 Conv 表示卷积, 使用 Conv\* 表示权重化特征的卷积, 使用 FC 表示全连接, 使用 Max Pool 表示最大值池化。对于输入、输出, 关键参数  $D$  表示数据的维度; 对于卷积, 关键参数  $K$  表示卷积核尺度,  $N$  表示卷积核数量,  $S$  表示卷积核移动步长; 对于权重化特征的卷积, 关键参数及含义均与卷积相同; 对于全连接, 关键参数  $N$  表示神经元个数; 对于最大值池化, 关键参数  $K$  表示池化核尺度,  $S$  表示池化核移动步长。

式中:  $m$  的物理学意义是动量, 表示以速度  $v$  前进的趋势;  $r$  为学习率, 是一个小于 1 的正数, 确保  $v$  的变化幅度不至于太大, 使参数  $w$  可以逐渐接近最优解。

## 4 实验结果及分析

实验时, 使用的计算机为 Dell PowerEdge 6850, 其中 CPU 为 3.00 GHz dual-core 64 bit Intel Xeon, 内存为 32 GB。软件方面, 预处理阶段的相关程序使用 Python 语言配合开源的计算机视觉库 OpenCV 完成编写。深度学习模型使用 Tensorflow<sup>[14]</sup> 开源框架面向 Python 语言的应用程序接口来完成模型的构建、训练与使用各个过程的相关程序编写。

### 4.1 预处理实验分析

以原始图片与经预处理后提升品质的图片为一组, 本文共选择了 3 组图片, 展示预处理方法对图片品质的提升效果。取式(1)中的参数  $\gamma$  为 0.65, 处理结果如图 4 所示。图 4(a) 从左到右对应的原始图片品质逐渐变差, 从图 4(b) 可以看出, 本文的预处理方法可以比较有效地提高图片的亮度以及清晰度, 但是面对品质极差的图片 (最右图片示例), 预处理方法效果有限, 无法将其品质提高到鱼类各项关键特征都清晰的程度。

### 4.2 权重化卷积实验分析

本文按照图 5 所示的模型结构构建深度学习模型  $model_1$ ; 同时, 保持图 5 中各个参数不变, 将所有的权重化卷积层都换为普通卷积层, 得到对照模型  $model_2$ 。



对于收集到的每个类别 700 张的海洋鱼类图片数据集,按照 8:2 的比例划分为训练数据集(560 张)与测试数据集(140 张)。不过这样的训练数据量难以满足需要海量数据的深度学习模型,面对训练数据不足的现实情况,本文对  $model_1$  与  $model_2$  均使用迁移学习<sup>[15]</sup>的方法,先将模型在 ImageNet 数据集<sup>[16]</sup>(约 120 万张图片)上训练得到中间模型,然后以中间模型的参数值为初始值,在鱼类图片训练集上训练,得到最终的模型参数值。在鱼类图片训练集上训练时,两个模型都取如下训练参数值:学习率  $r = 0.001$  动量  $m = 0.9$ 。训练过程结束后,两个模型在测试集上得到的准确率如表 1 所示。

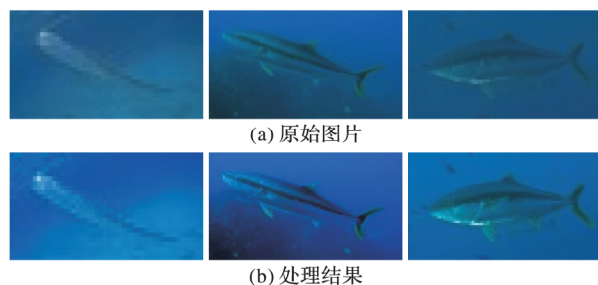


图 4 预处理结果

Fig. 4 Pre-processed results

为了同时展示预处理方法对识别准确率的辅助性提高作用,在表 1 中又加入了一项对照,即加入了预处理过程的  $model_1$ 。此外,在训练数据集与测试数据集之外,本文额外地为每种鱼类分别收集了 10 张清晰图片、10 张中等品质图片和 10 张较差品质图片。表 1 同时展示了三种模型在不同品质数据集上的识别准确率。

表 1 模型在各个数据集上的准确率 %

Tab. 1 Accuracy of models on different datasets %

实验模型	测试图片集	清晰图片	中等品质图片	较差品质图片
$Model_1$	91	97	91	84
$Model_2$	85	88	80	73
$Model_1$ + 预处理	—	97	94	90

分析表 1 可知,训练结束后,使用了权重化卷积的  $model_1$  在测试集上比没有使用权重化卷积的  $model_2$  的识别准确率

更高,说明权重化卷积提取的特征质量更高。在面对中等品质图片以及较差品质图片时,  $model_1$  与  $model_2$  的识别准确率虽然都有所下降,但  $model_1$  的识别准确率下降幅度更小,对低图片品质的鲁棒性更强,进一步说明了权重化卷积提取的特征质量更高。

同时,对比  $model_1$  与加入了预处理过程后的  $model_1$  可知,预处理提高了图片品质,可以一定程度上提高分类模型的准确率,而且图片品质越差这种提升效果越明显。

在 4.3 节中,举例展示了加入预处理过程后的  $model_1$  面对较差品质图片时的输出结果。

#### 4.3 识别结果展示与分析

表 2 展示了 10 种鱼类较差品质图片(如图 5)的识别结果。较差品质图片亮度很低且场景十分模糊,从示例图片上可以明显看出,鱼类特征已被严重破坏,即使是人也需要反复观察才能区分形态相似的不同鱼类。在这样恶劣的条件下,模型出现了明显的分类错误,将 1 号海鳗的图片分类为带鱼,并且所有结果向量中的最大概率都比较低,平均在 0.76 左右。这是因为较差品质图片中的部分关键特征遭到了破坏,即使经过了预处理也难以恢复,使得形态相似的鱼类之间尤其难以分辨。本实验表明,本文提出的方法能够基本有效地识别较差品质的图片;但是也同时说明了本文方法存在不足,基于暗通道先验算法的预处理过程性能欠佳,面对极差品质图片时,难以将图片品质提高至对分类结果没有影响的程度。当然,从另一方面来说,深度学习模型对极差品质图片的鲁棒性也有待提高。

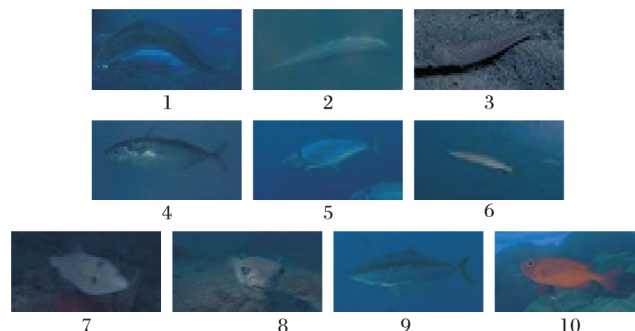


图 5 品质较差图片示例

Fig. 5 Poor quality picture examples

表 2 较差品质图片识别结果

Tab. 2 Recognition results of poor-quality pictures

输入示例图片序号	结果向量	最大概率	预测类别
1	(0.10 0.27 0.63 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00)	$c_3 = 0.63$	带鱼
2	(0.06 0.19 0.75 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00)	$c_3 = 0.75$	带鱼
3	(0.02 0.09 0.07 0.82 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00)	$c_4 = 0.82$	龙头鱼
4	(0.00 0.01 0.00 0.03 0.71 0.00 0.07 0.02 0.00 0.16 0.00 0.00)	$c_5 = 0.71$	鲈鱼
5	(0.03 0.00 0.00 0.03 0.13 0.68 0.04 0.00 0.00 0.09 0.00 0.00)	$c_6 = 0.68$	马鲛
6	(0.00 0.01 0.00 0.12 0.03 0.05 0.77 0.00 0.00 0.02 0.00 0.00)	$c_7 = 0.77$	鲟
7	(0.03 0.00 0.00 0.00 0.00 0.00 0.00 0.87 0.04 0.02 0.04 0.00)	$c_8 = 0.87$	马面鲀
8	(0.01 0.00 0.00 0.00 0.04 0.03 0.01 0.05 0.86 0.00 0.00 0.00)	$c_9 = 0.86$	刺鲀
9	(0.00 0.00 0.00 0.00 0.09 0.04 0.03 0.01 0.04 0.79 0.00 0.00)	$c_{10} = 0.79$	黄条鲷
10	(0.05 0.00 0.00 0.00 0.07 0.04 0.01 0.00 0.00 0.05 0.78 0.00)	$c_{11} = 0.78$	短尾大眼鲷

## 5 结语

海底观测视频存在大量空白数据,并且亮度低、场景模糊,导致视频图片品质差,直接使用深度学习方法识别视频中的海洋鱼类效果不好。本文提出了过滤空白数据、预处理提高图片品质、鲁棒性强的深度学习模型结构等方法,实现对较差品质图片的准确识别。对于空白数据,使用背景差分法过滤,根据应用场景对错误丢弃有效数据容忍度高的特点,选用了较高的阈值,大大缩短了处理全部数据的耗时。针对亮度低和场景模糊的问题,在预处理时首先使用伽马校正法提高了图片的亮度,然后参考基于暗通道先验算法的去雾研究针对水下环境的特点作了改进,提高了图片的清晰度。由于目前还没有积累下来有效、真实的海洋生物图片数据,本文使用爬虫技术,从互联网上收集了海洋鱼类图片数据集。在构建深度学习模型时,提出了权重化特征的卷积,激活关键特征、抑制非关键特征,提高了模型的鲁棒性。从实验结果可以看出,本文所提方法能以高达90%的准确率识别较差品质的图片,验证了该方法的有效性。然而,本文的方法还存在不足,其中的关键在于海底环境的特殊性,提高图片品质的预处理方法还没有完全结合海底环境的特点,提高图片品质的能力有限。下一步的研究工作将是深入分析、挖掘海底环境的特点,提出针对性更强的、更有效的预处理方法。

### 参考文献:

- [1] 李继龙,曹坤,丁放,等.基于渔获物统计的中国近海鱼类营养级结构变换及其与捕捞作业的关系[J].中国水产科学,2017,24(1):109-119.(LI J L, CAO K, DING F, et al. Changes in trophic-level structure of the main fish species caught by China and their relationship with fishing method[J]. Journal of Fishery Sciences of China, 2017, 24(1): 109-119.)
- [2] 岳冬冬,王鲁民,张勋,等.我国海洋捕捞装备与技术发展趋势研究[J].中国农业科技导报,2013,15(6):20-26.(YUE D D, WANG L M, ZHANG X, et al. The development trends of marine fishing equipment and technology in China[J]. Journal of Agricultural Science and Technology, 2013, 15(6): 20-26.)
- [3] 许枫,张乔,张纯,等.Walsh变换对鱼类特征识别的研究[J].应用声学,2015,34(5):465-470.(XU F, ZHANG Q, ZHANG C, et al. Walsh transform for fish identification[J]. Applied Acoustics, 2015, 34(5): 465-470.)
- [4] 张志强,牛智有,赵思明.基于机器视觉技术的淡水鱼品种识别[J].农业工程学报,2011,27(11):388-392.(ZHANG Z Q, NIU Z Y, ZHAO S M. Identification of freshwater fish species based on computer vision[J]. Transactions of the Chinese Society of Agricultural Engineering, 2011, 27(11): 388-392.)
- [5] 姚润璐,桂咏雯,黄秋桂.基于机器视觉的淡水鱼品种识别[J].微型机与应用,2017,36(24):37-39.(YAO R L, GUI Y W, HUANG Q G. Recognition of freshwater fish species based on machine vision[J]. Microcomputer and Applications, 2017, 36(24): 37-39.)
- [6] 林明旺.深度学习在鱼类图像识别与分类中的应用[J].数字技术与应用,2017,63(4):96-97.(LIN M W. Application of deep learning in fish image recognition and classification[J]. Digital Technology and Application, 2017, 63(4): 96-97.)
- [7] 沈瑜,王新新.基于背景减法和帧间差分法的视频运动目标检测[J].自动化与仪器仪表,2017(4):122-124.(SHEN Y, WANG X X. Video moving target detection method based on background subtraction and interframe difference method[J]. Automation and Instrumentation, 2017(4): 122-124.)
- [8] 蒋明敏.基于FPGA的LCD伽马校正研究[D].南京:南京林业大学,2016:25-27.(JIANG M M. Research on LCD Gamma correction based on FPGA[D]. Nanjing: Nanjing Forestry University, 2016: 25-28.)
- [9] HE K, SUN J, TANG X. Single image haze removal using dark channel prior[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341-2353.
- [10] 张波,唐启升,金显仕.东海高营养层次鱼类功能群及其主要种类[J].中国水产科学,2007,14(6):939-949.(ZHANG B, TANG Q S, JIN X S. Functional groups of fish assemblages and their major species at high trophic level in the East China Sea[J]. Journal of Fishery Sciences of China, 2007, 14(6): 939-949.)
- [11] 宋超,侯俊利,赵峰,等.春、秋季东海大桥海上风电场水域鱼类群落结构[J].海洋科学,2017,41(6):34-40.(SONG C, HOU J L, ZHAO F, et al. Fish community structure in the off-shore wind farm of Donghai Bridge in spring and autumn[J]. Ocean Science, 2017, 41(6): 34-40.)
- [12] RAWAT W, WANG Z. Deep convolutional neural networks for image classification: a comprehensive review[J]. Neural Computation, 2017, 29(9): 2352-2449.
- [13] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//Proceedings of the 30th International Conference on Machine Learning: Vol. 28. Atlanta, GA: JMLR, 2013: III-1139-III-1147.
- [14] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: a system for large-scale machine learning[C]//Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Berkeley, CA: USENIX Association, 2016: 265-283.
- [15] 李彦冬,郝宗波,雷航.卷积神经网络研究综述[J].计算机应用,2016,36(9):2508-2515.(LI Y D, HAO Z B, LEI H. Survey of convolutional neural network[J]. Journal of Computer Applications, 2016, 36(9): 2508-2515.)
- [16] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2009: 248-255.

This work is partially supported by the National Social Science Foundation of China (17BQT086), the Experimental Teaching Reform Project of Tongji University (0800104214).

**ZHANG Junlong**, born in 1994, M. S. candidate. His research interests include trusted search, deep learning.

**ZENG Guosun**, born in 1964, Ph. D., professor. His research interests include intelligent search, trusted Web software, parallel and distributed computing.

**QIN Rufu**, born in 1979, Ph. D., associate professor. His research interests include geographic information system, marine information technology, submarine observation system.