

doi: 10.7690/bgzdh.2023.08.013

## 基于 VGG 网络的轻量化小型水下构筑物探伤模型

冯家乐, 陆伟, 陈玮, 杨雪

(南京工程学院创新创业学院, 南京 211100)

**摘要:** 针对水下构筑物伤痕形态随机多变, 特征提取困难, 导致水下探伤识别精度较低的问题, 提出一种基于视觉几何组 (visual geometry group, VGG) 网络的轻量化小型 VGG (lite small visual geometry group, LSVGG) 模型。采用经典 VGG 网络结构, 减少卷积层和提高特征数量的方法在保证识别精度的前提下降低运算时间和系统开销。实验结果表明: 该 LSVGG 模型可以部署在小型无缆水下机器人 (autonomous underwater vehicle, AUV) 上, 具有较高水下构筑物探伤识别精度; 与传统模型相比, 水下构筑物探伤识别精度提高了近一倍, 识别准确率高达 99.7%。

**关键词:** VGG 网络; 轻量化; 水下构筑物探伤

**中图分类号:** TP183 **文献标志码:** A

## Flaw Detection Model of Lightweight Small Underwater Structures Based on VGG Network

Feng Jiale, Lu Wei, Chen Wei, Yang Xue

(School of Innovation and Entrepreneurship, Nanjing Institute of Technology, Nanjing 211100, China)

**Abstract:** Aiming at the problem of low recognition accuracy of underwater flaw detection due to the random and changeable shape of underwater structures and the difficulty of feature extraction, a lightweight small visual geometry group (LSVGG) model based on visual geometry group (VGG) network is proposed. The classical VGG network structure is adopted to reduce the convolution layer and improve the number of features, which can reduce the operation time and system overhead on the premise of ensuring the recognition accuracy. The experimental results show that the LSVGG model can be deployed on the small autonomous underwater vehicle (AUV) and has high flaw detection and identification accuracy for underwater structures. Compared with the traditional model, the recognition accuracy of underwater structure flaw detection is nearly doubled, and the recognition accuracy is as high as 99.7%.

**Keywords:** VGG network; lightweight; crack detection of underwater structure

### 0 引言

水坝是重要的水利基础设施, 在水资源管理、防洪减灾中发挥着至关重要的作用。大坝基础施工中, 易受内部环境影响, 如机械、设备、人为、材料等<sup>[1]</sup>, 导致大坝出现不同程度变形, 内部出现裂缝、渗漏等, 需针对具体情况监测诊断, 及时解决, 提高大坝整体安全性, 避免一系列灾难发生。我国目前的水下结构物探伤主要为人工定期检测, 国内外学者提出了一系列诸如高密度电阻率法、面波勘探法、瞬变电流法、超声波检测法、振动法和传感器检测法等方法<sup>[2]</sup>, 但都存在测量深度浅、无法完整探测整体情况以及效率低、误差大、适应性较弱等缺点。与传统测量方法相比, 数字图像处理凭借其成本低、处理精度高、适用性广泛的优势, 正应用于各方面解决多种复杂问题<sup>[3]</sup>。

近年来, 国内外研究人员对基于光学图像的水

下目标探测关键技术进行了大量研究<sup>[4]</sup>, Han 等<sup>[5]</sup>总结了一系列水下图像增强算法, C.Fabbri 等<sup>[6]</sup>提出了用于水下图像增强的生成对抗网络 UGAN, 王龔<sup>[7]</sup>对暗通道先验算法在水下图像复原进行研究, S.Anwar 等<sup>[8]</sup>提出用于水下图像增强的卷积神经网络 UIENet、DUIENet, 这些算法很好地将水下图像还原为非水下图像。研究人员还提出水下目标检测算法, Tsou 等<sup>[9]</sup>将霍夫变换用于识别物体的中心特征, M.Pedersen 等<sup>[10]</sup>对 YOLOv2 算法进行了调整, 并用于水下目标检测。

水下构筑物伤痕有着图像模糊、数据集少、裂缝伤痕形态多变、特征提取困难、不能进行标准化描述和表征等特点, 导致识别精度低。VGG16 网络<sup>[11]</sup>是一款经典的使用卷积神经网络进行识别分类的模型, 其有着识别准确率高、模型搭建较为简单等优点, 夏坚等<sup>[12]</sup>使用迁移学习与 VGG16 网络

收稿日期: 2023-04-19; 修回日期: 2023-05-20

基金项目: 江苏省大学生创新训练计划 (202111276010Z; 202011276007Z)

作者简介: 冯家乐 (2001—), 男, 安徽人。

在建筑物裂缝检测方面取得了一定进步；但检测速度慢、时间长、模型规模大无法在嵌入式设备上部署等缺点使其无法在水下构筑物探伤方面得到有效的大规模利用。针对以上问题，笔者提出一个轻量化的小型视觉几何组 (VGG) 模型，使其在满足对水下构筑物裂缝识别精度的条件下，有更小的参数规模、更快的识别速度，并自主设计搭载了 Jetson Nano 嵌入式设备的无缆水下机器人 (AUV) 系统平台。

## 1 神经网络模型结构设计

### 1.1 VGG 网络概述

VGG16 模型相比传统模型，主要在卷积层中提出了使用连续的小卷积核取代大卷积核，多个小卷

积层和一个池化层形成一个卷积组，在输出层提出了“全连接转卷积”的做法，使训练出的模型可以对任意大小的图像进行处理。VGG16 网络是一个巨大的网络，整体参数达到 1.4 亿多个，不管是训练还是推理对于算力的要求都非常高。网络结构如图 1 所示。

### 1.2 LSVGG 模型结构

针对 VGG16 模型对算力要求过高的问题，笔者设计了轻量化小型 VGG 网络 (LSVGG)。该网络采用了 VGG 网络提出的使用堆砌的小卷积核取代大卷积核的思想，实现了一个 9 层网络的二分类模型。模型结构如图 2 所示。

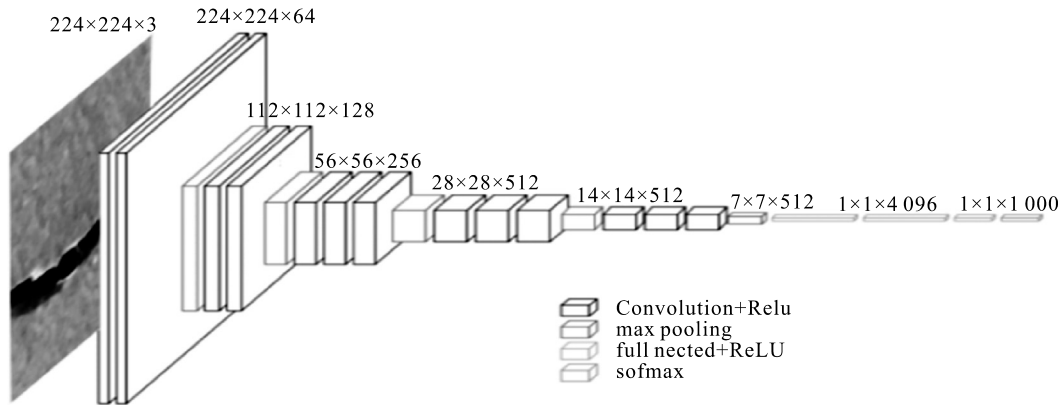


图 1 VGG16 网络结构

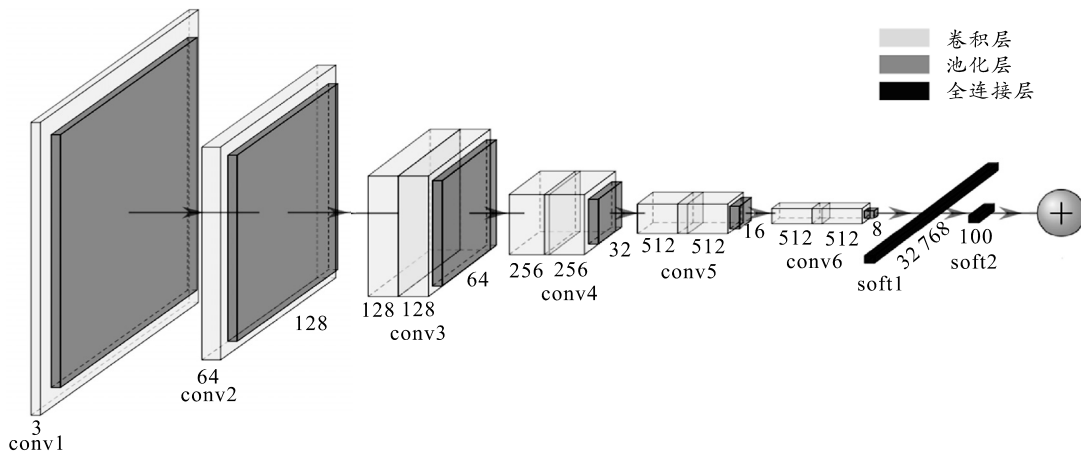


图 2 LSVGG 模型结构

LSVGG 网络包括输入层、5 个卷积层、2 个全连接层和输出层。与 VGG 模型不同，LSVGG 输入层接受 256x256 的 RGB 三通道图像，既提高了图像的清晰度与质量，又提高了提取到的特征数量。

卷积层是 LSVGG 网络的核心，为减少计算量，提高计算速度，LSVGG 在卷积层上相比于 VGG16 网络每个卷积组里减少一个 3x3 的卷积核，使用连续的 2 个 3x3 的卷积核进行特征提取，设置步长为

1，采用边缘补零的方式对图像进行卷积计算，每个卷积核计算的结果都要通过 ReLU 激活函数，ReLU 函数为：

$$f(x)=\max(0, \omega^T+b) \tag{1}$$

式中： $\omega$  为卷积核； $x$  为输入图像； $b$  为偏差。将结果通过最大值池化，使用 2x2 的滑动窗口，滑动步长为 2，采用对图像边缘不补 0 的方法，将计算出的输出图像长与宽缩短为原来的一半，最终输出为

64 个  $128 \times 128$  的特征图。输入图像每经过一次卷积-激活卷积-激活-最大池化这样的卷积组后,特征图数量翻倍、大小减小一半。每层卷积层还使用了参数为 0.25 的 Dropout 来引入随机噪声以防止模型过拟合。

全连接层首先会将前一层卷积层输出的 2 维特征图扁平化为含有 32 768 个特征的 1 维向量,将这些特征作为全连接层的输入,得到含有 100 个特征的 1 维向量。将这 100 个特征 1 维向量再次进行全连接,得到只含有 1 个特征的 1 维向量。全连接层本质为矩阵乘法:

$$y=(Wx^T)^T. \quad (2)$$

式中:  $W$  为参数权重矩阵;  $x$  为 1 维特征向量;  $y$  为输出特征向量。

输出层将前一层全连接层得到的 1 个 1 维特征向量通过 Sigmoid 函数进行图像最终分类。Sigmoid 函数有着平滑、易于求导等优点,可实现将一个实数进行 0~1 的映射,用于二分类问题有着非常好的效果, Sigmoid 函数  $S(x)$  及其导数  $S'(x)$  为:

$$S(x)=1/(1+e^{-x}); \quad (3)$$

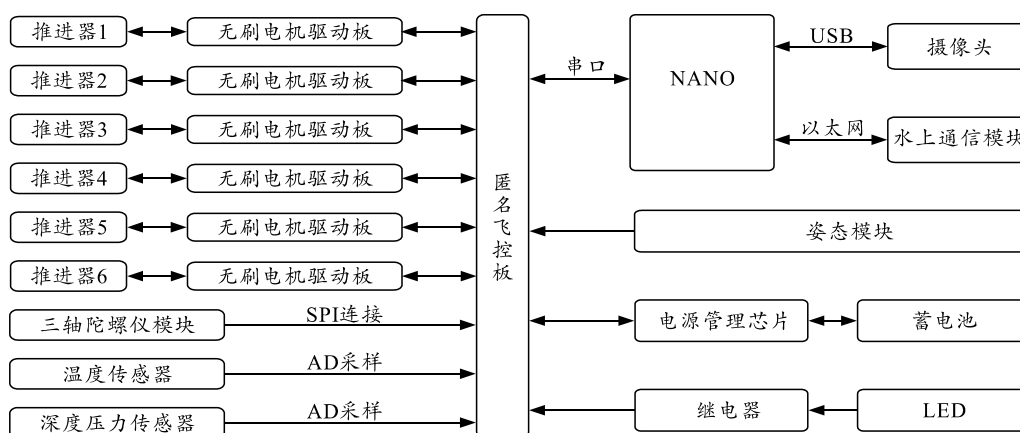


图 3 系统结构

该款 AUV 主要包括 Jetson Nano 嵌入式设备、匿名飞控、陀螺仪模块、温度传感器模块、深度压力传感器模块、姿态控制模块、电源管理模块和电机驱动模块。采用六推进器驱动,可使用电脑和手机 APP 通过网络进行图像传输与机器人运动控制。

Jetson Nano 是英伟达公司开发的嵌入式计算设备,具有体积小、算力强、硬件接口丰富、易于开发等优点。Nano 可以安装 linux 嵌入式操作系统,同时英伟达公司为 Nano 开发了 CUDA 和 PyTorch 环境支持,无需对训练好的模型进行任何转换,可以直接运行。

$$S'(x)=S(x)(1-S(x)). \quad (4)$$

式中  $x$  为图像最终的 1 个 1 维特征向量。与 VGG 网络不同,LSVGG 的输出层都没有使用  $1 \times 1$  的卷积核进行“全连接转卷积”,主要是考虑到图像输入大小固定以及对训练好的模型添加  $1 \times 1$  的卷积核时对参数的微调十分困难。同时,由于 LSVGG 网络的输出结果为二分类网络,大大降低了全连接层的参数量,提高了运行速度。由于在最后使用了 Sigmoid 函数进行结果输出,LSVGG 在训练时采用交叉熵损失函数来计算 loss,交叉熵损失函数适用于单标签二分类网络,对于一批包含  $N$  个样本的数据,其计算式为:

$$\text{loss} = \frac{1}{N} \sum_{n=1}^N l_n; \quad (5)$$

$$l_n = -[y_n \log x_n + (1 - y_n) \log(1 - x_n)]. \quad (6)$$

式中:  $x_n$  为模型预测输出;  $y_n$  为数据实际标签。

## 2 模型训练与评估

### 2.1 实验测试条件

本模型部署在自主设计的 AUV 上,如图 3 所示。

PyTorch 是一款开源的、专门针对 GPU 加速的神经网络框架,拥有简洁、快速、易用等特点。与 Tensorflow 的静态计算图不同,PyTorch 的计算图是动态的,可根据计算需要实时改变计算图,PyTorch 为 Python 语言使用者提供了舒适的写代码选择。

虽然 AUV 已使用算力较强的 Jetson Nano,但由于嵌入式设备的种种限制,Nano 的算力及显存相较于普通电脑 GPU 依然很低,无法运行大型网络。

### 2.2 模型评估指标

二分类模型对于预测结果分为 4 种情况:1)  $TP$ ,数据实际有裂缝,模型预测有裂缝;2)  $FN$ ,数据实

际有裂缝，模型预测无裂缝；3) *FP*，数据实际没有裂缝，模型预测有裂缝；4) *TN*，数据实际没有裂缝，模型预测没有裂缝。基于这 4 种情况，对于模型的评估可以使用精确率 *P*、召回率 *R* 和准确度 *Accuracy* 3 个指标。精确率 *P* 表示预测为真的样本总数占实际为真的样本总数的占比：

$$P=TP/(TP+FP)。(7)$$

召回率 *R* 又叫查全率，表示在所有实际为真的样本中，预测为真的样本数量占比：

$$R=TP/(TP+FN)。(8)$$

准确度 *Accuracy* 是最常见的评价指标，表示预测正确样本的占比：

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)。(9)$$

### 2.3 模型训练与结果分析

训练数据集来自 Kaggle 开源数据集 Surface Crack Detection。该数据集共有 40 000 张图片，其中正类与负类各占一半，部分数据集如图 4 所示。为更好地训练与评估模型，将数据集随机打乱后，按照 6:2:2 的比例划分为训练集、交叉验证集、测试集。

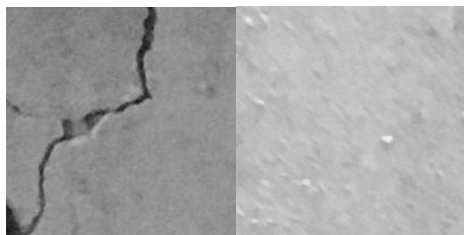


图 4 数据集演示

对于模型的训练和超参数的选择策略为对包括 epoch、adam 优化器的 learning rate、b1、b2 等参数进行排列组合，通过训练集的全部数据进行训练，获得一批模型，选取在训练集上表现较好的模型，通过交叉验证集进行测试，选取其中表现最好的模型，以此作为最终结果，并在测试集上验证模型的效果。经过调试，最终对于各参数的选择为 learning rate=0.000 2，b1=0.5，b2=0.999，最终模型的准确率 *P* 和召回率 *R* 均达到 99%，如表 1 所示。

表 1 模型在不同数据集上的表现 %

数据集类型	Precision	Recall	Accuracy
训练集	100.00	100.00	95.95
交叉验证集	99.77	99.85	99.81
测试集	99.67	99.82	99.75

为验证模型的有效性，在相同的数据集上使用一些经典模型进行训练，得到的效果均略逊于笔者提出的模型，如表 2 所示。

表 2 不同模型效果对比 %

模型类型	Precision	Recall	Accuracy
LeNet	89.58	12.90	55.70
传统 CNN	88.76	20.27	58.64
VGG16	99.70	99.76	99.85
LSVGG	99.67	99.82	99.75

可见，相比传统的 LeNet 与常见的卷积神经网络，在相同的参数初始化设置下，对于准确率和召回率均有所提升，而与 VGG16 相比，LSVGG 在各项指标上均与 VGG16 网络表现相差无几，考虑到水下构筑物损伤的严重危害，能够提升损伤识别精度对于经济发展有着重要的贡献。此外，对不同模型在 Nano 上运行相同图片所耗费的时间和巅峰时内存的占用进行了计算比较，如表 3 所示。

表 3 不同模型运行单张图片耗时

模型类型	单张图片耗时/s	内存占用率/%
LeNet	0.002	57.90
传统 CNN	0.006	64.46
VGG16	2.780	100.00
LSVGG	0.120	81.57

可见，LSVGG 相比于 VGG16 网络在运行速度上有了显著提升，每秒可处理 10 帧图像，基本满足对于视频流处理的实时性要求。由于 Nano 内存限制，VGG16 网络会占满内存，系统使用 swap 虚拟内存交换，导致运算速度急剧下降。LeNet 虽然运算速度快、内存占用率低，但识别精度偏低。同时，考虑到识别的精确度和运算的速度，LSVGG 模型无疑是最优的选择。

### 3 结束语

笔者提出一种适用于水下构筑物探伤的基于 VGG 网络的轻量化小型模型。模型中的改进都是基于图像特点与嵌入式设备部署构建的，具体表现为：使用连续的小卷积核堆积，使用 Dropout 以防止过拟合，减少模型层数以减少计算量。实验结果表明：与传统模型相比，笔者提出的模型有更高的识别准确率。在后续研究中，将扩充更多的数据集，对于水下图像采用预处理以提高图像质量，并对模型进行优化，进一步提高模型的识别精度。

### 参考文献：

[1] 牛燕雄, 陈梦琪, 张贺. 基于尺度不变特征变换的快速景象匹配方法[J]. 电子与信息学报, 2019, 41(3): 626-631.  
 [2] 肖飞. 大坝安全监测技术探讨[J]. 低碳世界, 2021, 11(1): 70-71.