doi:10.6041/j.issn.1000-1298.2019.04.020

基于深度学习的群猪图像实例分割方法

高　云[1,2]　郭继亮[1]　黎　煊[1,2]　雷明刚[2,3]　卢　军[4]　童　宇[1,2]

(1.华中农业大学工学院,武汉 430070;2.生猪健康养殖协同创新中心,武汉 430070;
3.华中农业大学动物科技学院动物医学院,武汉 430070;4.华中农业大学理学院,武汉 430070)

摘要: 群养饲喂模式下猪群有聚集在一起的习性,特别是躺卧时,当使用机器视觉跟踪监测猪只时,图像中存在猪体粘连,导致分割困难,成为实现群猪视觉追踪和监测的瓶颈。根据实例分割原理,把猪群中的猪只看作一个实例,在深度卷积神经网络基础上建立 PigNet 网络,对群猪图像尤其是对粘连猪体进行实例分割,实现独立猪体的分辨和定位。PigNet 网络采用 44 层卷积层作为主干网络,经区域候选网络(Region proposal networks,RPN)提取感兴趣区域(ROI),并和主干网络前向传播的特征图共享给感兴趣区域对齐层(Region of interest align,ROIAlign),分支通过双线性插值计算目标空间,三分支并行输出 ROI 目标的类别、回归框和掩模。Mask 分支采用平均二值交叉熵损失函数计算独立猪体的目标掩模损失。连续 28 d 采集 6 头 9.6 kg 左右大白仔猪图像,抽取前 7 d 内各不同时段、不同行为模式群养猪图像 2 500 幅作为训练集和验证集,训练集和验证集的比例为 4∶1。结果表明,PigNet 网络模型在训练集上总分割准确率达 86.15%,在验证集上准确率达 85.40%。本文算法对不同形态、粘连严重的群猪图像能够准确分割出独立的猪个体目标。将本文算法与 Mask R-CNN 模型及其改进模型进行对比,准确率比 Mask R-CNN 模型高 11.40 个百分点,单幅图像处理时间为 2.12 s,比 Mask R-CNN 模型短 30 ms。

关键词: 群养猪;图像分割;实例分割;卷积神经网络;深度学习;粘连猪体

中图分类号: TP391　　**文献标识码:** A　　**文章编号:** 1000-1298(2019)04-0179-09

Instance-level Segmentation Method for Group Pig Images Based on Deep Learning

GAO Yun[1,2]　GUO Jiliang[1]　LI Xuan[1,2]　LEI Minggang[2,3]　LU Jun[4]　TONG Yu[1,2]

(1. *College of Engineering*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*
2. *The Cooperative Innovation Center for Sustainable Pig Production*, *Wuhan* 430070, *China*
3. *College of Animal Science and Technology*, *College of Animal Medicine*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*
4. *College of Science*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*)

Abstract: With the development of intelligence and automation technology, people pay more attention to use it to monitor pig welfare and health in modern pig industry. Since the behaviors of group pigs present their healthy status, it is necessary to detect and monitor behaviors of group pigs. At present, machine vision technology with advantages of low price, easy installation, non-invasion and mature algorithm has been preferentially utilized to monitor pigs' behaviors, such as drinking, eating, farrowing behavior of sow, and detect some of pigs' physiological indices, such as lean yield rate. Feeding pigs at group level was used the most in intensive pig farms. Owing to normally happened huddled pigs showing in group-pig images, it was challenging to utilize traditional machine vision technique to monitor the behaviors of group pigs through separating adhesive pig areas. Thus a new segmentation method was introduced based on deep convolution neural network to separate adhesive pig areas in group-pig images. A PigNet network was built to solute the problem of separating adhesive pig areas in group-pig images. Main part of the PigNet network was established on the structure of the Mask R-CNN network. The Mask R-CNN network was a deep convolution neural network, which had a backbone network with a branch of FCN from classification layer and regression layer to mask the region of interest. The PigNet network used 44 convolutional layers of backbone network of Mask R-CNN network as its main network. After the main

收稿日期:2018-10-17　修回日期:2019-01-29

基金项目: 国家重点研发计划项目(2016YFD0500506)和中央高校自主创新基金项目(2662018JC003、2662018JC010、2662017JC028)

作者简介: 高云(1974—),女,副教授,博士,主要从事农业智能检测与控制研究,E-mail:angelclouder@mail.hzau.edu.cn

network, the output feature image was fed to the next four convolutional layers with different convolution kernels, which was the resting part of the main network and produced binary mask for each pig area. As well, the output feature image was fed into two branches, one was the region proposal networks (RPN), the other was region of interest align (ROIAlign) processing. The first branch outputted the regions of interest, and then the second one aligned each pig area and produced class of the whole pig area and the background area and bounding boxes of each pig regions. A binary cross entropy loss function was utilized to calculate the loss of each mask to correct the class layer and the location of ROIs. Here, the ROIAlign was used to align the candidate region and convolution characteristics through the bilinear difference, and which would not lose information by quantization, making the segmentation more accurate, and FCN of the mask branch used average binary cross entropy as the loss function to process each mask, which avoided the competition among pig masks. After all, the ROI was labeled with different colors. Totally 2000 images captured from previous five days of a 28-day experiments were taken as the training set, and 500 images from the next 6th to 7th day were validation set. The results showed that the accuracy of the PigNet on training set was 86.15% and on validation set was 85.40%. The accuracies on each data set were very close, which showed that the model had effective generalization performance and high precision. The cooperation between the PigNet, Mask R-CNN (ResNet101-FPN) and its improvement showed the PigNet surpassed the other two algorithms in accuracy. Meanwhile, the PigNet run faster than the Mask R-CNN. However, the times of three algorithms spent on 500 samples of the validation set were similar. The algorithm can be used to separate individual pig from group-pig images with different behaviors and severe adhesion situation. The PigNet network model adopted the GPU operation mode, and used the three branches of class, regression box and mask to compute parallel processing time, which made the processing time of single image quick, only 2.12 s. To a certain degree, the PigNet could reduce convolution parameters and simplify the network structure. The research provided a new segmentation method for adhesive group-pig images, which would increase the possibility of group-pig tracing and monitoring.

Key words: group pig; image segmentation; instance segmentation; convolution neural network; deep learning; adherent pig body

0 引言

机器视觉技术是检测猪养殖状况、评价猪福利的重要技术手段之一,具有易安装、无侵入、设备平价等优点。目前,机器视觉技术较多应用于单体猪只的监测,主要用于识别猪只饮水行为[1]及监测猪只的饮水量[2-3]、监测母猪的分娩行为[4]、检测生猪的瘦肉率[5]、监测猪只蓝耳病疫情[6]等。对于群体饲喂阶段,由于猪群有聚集在一起的习性,特别是躺卧时,当使用机器视觉跟踪监测猪只时,图像中存在较多的粘连猪体,且分割困难。目前要对群养猪个体进行追踪和监测,尚未有较为成熟的应用。但群养模式无论是在种猪场还是商品肉猪场均占有较大的比重[7-8],对群养猪行为进行监测有利于全面改进猪场养殖管理水平,提高猪场的生产效率,从而提高猪场的经济效益。因此,解决群养猪图像中粘连猪体的分割问题是实现群猪视觉追踪和监测的关键步骤。

为了分割群养猪图像中的粘连猪体,采用传统数字图像处理结合模式识别的方法已有较多研究[9-13]。卷积神经网络(Convolutional neural network,CNN)方法[14-20]能够在背景知识不清晰,推理规则不明确情况下,通过训练,自动提取图像中目标特征,避免了通过人为观察来设计算法提取目标特征的过程,得到的算法模型普适性较强。GIRSHICK 等[21]提出了将 R-CNN(Regions with CNN)应用到候选区域(Region proposal)的策略,训练模型用于定位目标物和图像分割,其分割的精度、效率均不高。为了提高分割精度,LONG 等[22]对 CNN 进行改进,提出了全卷积网络(Fully convolutional networks,FCN)的图像语义分割算法,主要用来理解场景中存在的感兴趣对象,实现像素级别的分割,有研究人员也将该方法应用到母猪个体识别和分割方面[23]。语义分割可以分割出同一类物体,但无法分割出同一类物体中的不同个体,不适用于分割群养猪中的粘连猪个体。

实例分割主要用来分割同类物体的不同个体。SILBERMAN 等[24]在 2014 年用实例分割来分析室内场景。LI 等[25]实现了全卷积端到端实例分割模型(Fully convolutional instance-aware semantic segmentation,FCIS)。实例分割是一个多任务架构网络,通过卷积神经网络先提取初级特征,并提出感兴趣区域(Region of interest,ROI)。对于每个 ROI 区域,通过对 ROI 池化层处理提取出对应的特征。

所提取特征共享给检测目标任务的网络，完成目标检测、图像分割与图像分类任务。实例分割将图像中不同类别和个体的感兴趣对象分别进行分割和标记[26]。HE 等[27]构建的 Mask R－CNN 网络将实例分割推上了一个新的高度，对各种不同的粘连物体表现出了很好的分割性能。ZHANG 等[28]对道路两侧粘连汽车图像进行了实例分割研究。WARD 等[29]采用深度学习网络 Mask R－CNN 实现了植物叶片的实例分割。实例分割框架在语义分割的基础上，将相互粘连的同一类对象分割为不同的实例个体，为猪群中粘连个体的分割提供了可能性。目前还未见将实例分割用于粘连猪体分割的研究。

本文采用深度学习方法，对实例分割网络进行修改，并将其用于对群猪图像中的粘连猪体进行分割。通过对网络进行训练，得到有效的分割模型，并通过试验验证算法的可行性与有效性。

1　试验材料与方法

1.1　图像采集

于 2018 年 3 月 12 日—4 月 8 日在华中农业大学试验猪场进行图像数据采集，6 头 9.6 kg 左右的大白保育猪养殖在环境多变量控制养殖箱内。环境多变量控制养殖箱[30]是课题组按集约化猪舍设计标准设计的，已申请专利。根据养殖猪头数，按集约化猪场的猪只需占空间面积，设计猪栏的长、宽为 2.0 m×2.0 m。猪舍内的料槽和鸭嘴式饮水器为商用猪舍产品。猪舍底部为全漏缝地板，地板下有水泡粪粪池，按照水泡粪的模式，水面超出10 cm，粪池容量设计为一周抽粪一次。箱内用 LED 灯照明，猪生活区域照度根据猪场光照标准控制在50 lx 左右[31]。1080 p 高清摄像头安装于箱顶部正中间，垂直向下，采集 RGB 图像，镜头距漏缝地板高度为 180 cm，如图 1 所示。采集速率为 3 帧/s，图像为 1 920 像素×1 080 像素的 RGB 彩色图像，存储为 JPG 格式。

1.2　基于深度卷积神经网络的群养猪分割模型

已有研究采用 FCN 对粘连情况复杂的图像进行分割。FCN 是一种深度学习的全卷积神经网络模型，其处理分割问题时性能较好。FCN 网络的优势[32]包括：①去掉了末端的全连接层，用卷积层代替，实现端到端的卷积网络训练。②对图像的每个像素都进行预测分类，实现像素级分割。但对于粘连严重的猪群图像，FCN 网络结构[32]上采样采取简单的反卷积方式，这种方式的上采样使得图像失去了许多细节特征，导致最终分割图中猪体轮廓模糊，难以对粘连严重的猪群进行很好的分割。

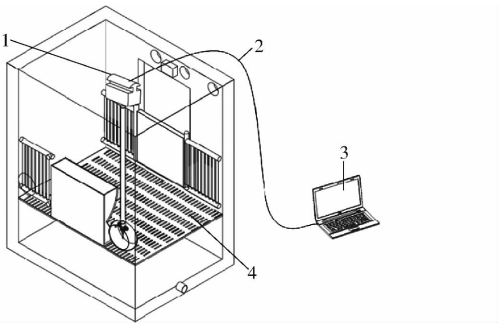

图 1　图像采集平台

Fig. 1　Image acquisition platform

1. 高清摄像头　2. 数据线　3. 计算机　4. 环境多变量控制养殖箱

Mask R－CNN 网络是一种实例分割（Instance segmentation）[27]方法，其检测出每个目标，并用边界框、标签和不同颜色掩模对目标个体进行分割描述。Mask R－CNN 网络结构基于 REN 等[33]提出的 Faster R－CNN 框架，该框架将特征抽取、方案（Proposal）提取、bounding box 回归、分类整合在同一个网络中。Mask R－CNN 网络结构是在 Faster R－CNN 框架分类和回归层之后增加了一个针对每一个 ROI 的 FCN 网络分支，实现对图像目标进行实例分割。Mask R－CNN 网络采用 ROIAlign 操作替换了 ROIPool 操作，保留目标空间位置的精度。ROIAlign 是一个修正空间量化特征提取错位的层[27]，利用双线性插值计算输入对应 ROI bin 上的 4 个坐标值，作用是保留其网络输入和输出之间的像素到像素的空间位置精度。Mask R－CNN 网络将判断类别（Class）和输出掩模（Mask）这两个任务解耦合，用平均二值交叉熵损失对每个目标独立地预测一个二值掩模，避免引入类别间竞争，大大提高了分割性能。本文中的每个目标即为猪群中的每头猪只区域，对每头猪只预测一个二值掩模。

本文在 Mask R－CNN 网络结构的基础上，对 Mask R－CNN 网络框架的网络深度和宽度进行优化调整，在已经有一定训练基础的参数上，进行迁移学习，用粘连群养猪图像作为训练集，比较不同层数和不同卷积核的网络模型的分割准确率，确定较优的分割粘连猪体图像的网络参数和模型。

最终，本文确定优化模型为 PigNet，其网络较 Mask R－CNN 网络做了两点优化改进：①针对本文试验场景中类别单一的不同实例对象，将 Mask R－CNN 主干网络第 4 阶段卷积的 69 层卷积层结构调整为 12 层（PigNet），这样不仅防止了特征损耗，而且减少了卷积运算量。②Mask R－CNN 网络的 Mask 分支最后一层卷积层通道数 80 修改为 2（PigNet），代表类别数为两类（pig 类和 background 类）。PigNet 具体网络结构如图 2 所示。

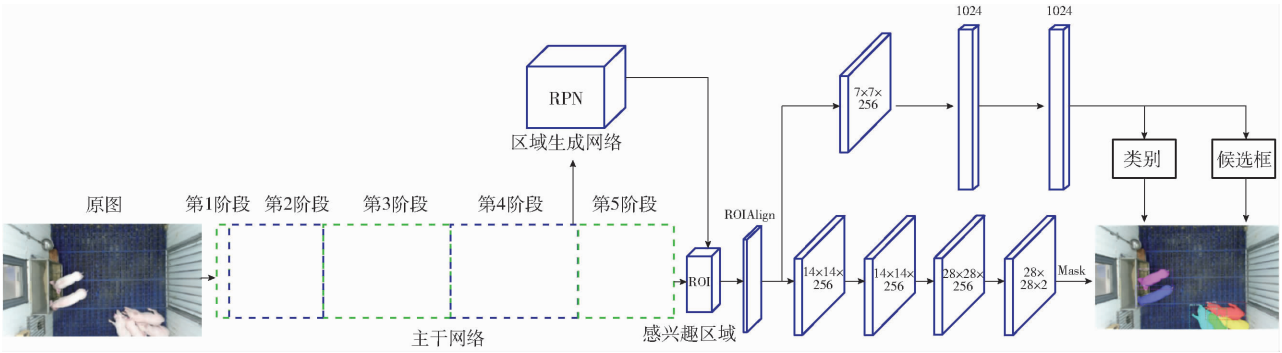

图 2 PigNet 网络结构

Fig. 2 PigNet network structure

PigNet 主干网络采用 44 层卷积层，有 5 个卷积阶段，每 1 阶段的卷积均采用残差学习结构(Building block)[34]，如图 3 所示，蓝色弧线包含的 3 个卷积层结构中，3×3×256 层表示卷积核为 3×3、通道数为 256 的卷积层(图 3 中类似的结构，依此类推)，1 024 表示向量维数为 1 024 的全连接层。残差学习结构的首端和末端采用 1×1 卷积核，中间采用 3×3 卷积核。残差学习结构的优点在于：其极大地降低了参数的数目，减少计算量的同时保持了精度，如第 2 阶段的第 1 个 1×1 的卷积将 256 维通道降到 64 维，然后再通过 1×1 卷积恢复到 256 维通道。残差学习结构有很多的旁路即蓝色弧线，将输入直接连接到后面的层，一定程度上解决特征损耗、梯度消失等训练问题。第 1 阶段卷积层后连接有 BatchNorm 层、Scale 层、ReLU 激活层，之后连接有池化层，采用卷积核为 3×3 的 Max pooling 操作。BatchNorm 层和 Scale 层作用是批量归一化，ReLU 激活层有利于训练收敛更快。后面的每层卷积层的滑动步长(Stride)为 2 个像素，经过一个残差学习结构的卷积、池化把尺寸缩小至原来的 1/4，PigNet 主干网络滑动步长不停的缩减尺寸，输出通道数持续增加，达到 2 048。

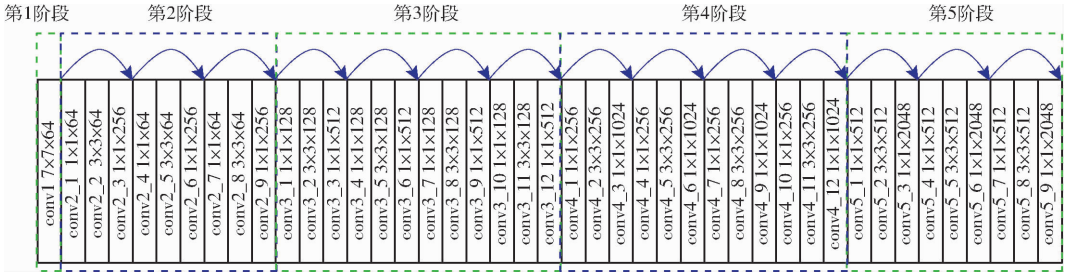

图 3 PigNet 主干网络结构

Fig. 3 Backbone network structure of PigNet

主干网络的特征提取分两方面：一方面是网络模型的 conv4_12 卷积层输出的特征图(Feature map)经过区域候选网络(Region proposal networks, RPN)计算处理，提供计算后提取的特征信息；另一方面是继续向前传播，生成特征图。主干网络所提取的特征图由 ROIAlign 层处理。RPN 的作用是更快地选取 ROI。ROIAlign 从 ROI 提取特征并保留输入和输出之间的像素的精准空间位置。从 ROI 分出来两个分支，一个分支经过 7×7 卷积核、通道数 256 的卷积层，再经过两个 1 024 维特征向量的全连接层，最后完成预测类别和回归框任务；另一 Mask 分支经过 5 次 14×14 卷积核、通道数为 256 的卷积层运算，再进行上采样生成 28×28 的特征图，最后使用 28×28 卷积核、通道数为 2 的卷积层对每个 ROI 目标(本文为每头猪只)预测输出一个掩模。

PigNet 实例分割网络的损失函数包括 3 部分，公式[27]为

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

式中 L——损失函数

L_{cls}——分类误差

L_{box}——检测误差

L_{mask}——分割误差

L_{cls}、L_{box}利用全连接层处理，预测出每个 ROI 的所属类别和回归框坐标值。L_{mask}对每个 ROI 的目标进行分割，并赋予掩模表示。Class 分支、Box 分支、Mask 分支共享卷积计算得出 ROI。对于每个 ROI，L_{cls}负责预测该目标的类别，若预测感兴趣区域 ROI 为 pig 类，则在分割该 ROI 过程的损失时只使用 pig 类的相对熵误差作为误差值进行计算，其他类别(如 background 类)的相对熵误差不参与到该 ROI

的损失函数中，避免了类别间竞争。L_{box}用来控制该目标的回归框的位置坐标。L_{mask}用来控制该目标生成精准的掩模。Class 分支预测感兴趣区域 ROI 为 pig 类，则 L_{mask}仅仅单独考虑 ROI 目标的回归框的 pig 类别相对熵误差来预测该感兴趣区域属于 pig 类别的像素点，精准地分割出目标轮廓，分割得到不同层深度的精准轮廓位置坐标信息，分割图中使用不同颜色的掩模来表示不同层深度的目标。对于本文测试图像，PigNet 网络模型卷积计算得 6 个感兴趣区域，L_{cls}预测每个感兴趣区域类别为 pig，L_{box}预测 6 个目标的回归框的位置坐标，L_{mask}采用平均二值交叉熵损失并使用 Sigmoid 函数对 6 个回归框的位置均独立地预测一个二值掩模，6 个目标分割的目标轮廓信息放置在 6 个不同层深度，在分割图中用 6 种不同颜色的掩模来表示 6 个目标。对于数量更多的猪群图像，PigNet 模型对每个猪体分割生成一个对应的二值掩模。

1.3　训练参数设置

本文网络激活函数采用修正线性单元（Rectified linear units，ReLU）和基于批量随机梯度下降法（Mini-batch stochastic gradient descend）的优化模型，使用 GPU 运算。批量训练的方法将数据集数据分为多个批次，设置迭代次数为 10，每个批次训练系数为 100，随着迭代次数的增加，网络损失函数损失收敛，训练集和验证集的准确率趋于稳定，调整动量因子参数为 0.9，正则化权重衰减系数为 0.000 1，网络权重参数的基础学习率为 0.001。

2　群养猪图像分割试验

2.1　试验硬件及软件

本文分割算法处理平台为台式计算机，处理器为 Inter Core i5－7500，主频为 3.40 GHz，32 GB 内存，2TB 硬盘，显卡为 NVIDA GeForce GTX1060。软件环境为 Windows 10 64 位系统，编程语言为 Python，采用 TensorFlow 深度学习开源框架，使用 Microsoft visual studio 软件编写深度神经网络层。

2.2　试验方法

连续采集 28 d 图像，包含了群养猪日常生活的吃食、饮水、躺卧、运动等多种行为的图像样本。为了确保训练样本的多样性，本文抽取试验前 5 d 中全天各时段不同行为状态下的群养猪图像 2 000 幅作为训练集。为了确保验证集与训练集之间无重叠图像数据，抽取第 6、7 天的各个时段的群养猪图像 500 幅作为验证集，总计 2 500 幅群养猪图像样本。

使用 Labelme 软件制作训练数据集。标注时，将每头猪标注为一个独立的连通域。试验对象为 6 头保育猪，因此每幅图像中标出 6 头猪个体连通域，即 6 个目标，人工每采集一个猪体目标轮廓就给该猪体目标赋予具体的标签。6 个猪体目标标签依次命名为 pig1、pig2、…、pig6。标注完一幅图像后得到一个对应的. json 文件。解析. json 文件，提取中位深度为 24 位的 label. png 标签图像，再将 24 位 label. png 图像转换为 8 位的 label. png 图像，完成一幅图像数据的制作。

若将 6 个猪体目标直接标签标注为 pig，则 6 个目标会放置在同一个通道层，被识别为同一目标对象。若把 6 个猪体目标命名为不同的标签，将会被识别为 6 个不同对象。因此在图像中用 6 种不同颜色表示，对应 Mask R－CNN 网络对不同的实例目标需要放在不同层中的输入要求。但是上述标注后 6 个不同的标签代表 6 个类别，而在主网络训练中要求 6 个猪体目标是属于 1 个类别标签，即 pig 大类别标签。本文采用 $w \times h \times n$ 三维数组将在 6 个深度层的 6 个类别转换为在 6 个深度层的 1 个类别，作为 Mask R－CNN 网络的输入。w 和 h 分别为输入图像的长和宽，n 为输入图像中实例目标的猪头数，本文设为 6。读取 6 个猪体目标的类别标签为 pig。

由于本研究的目是分割粘连的群养猪，准确识别出独立的猪个体目标位置和个数，而图像中猪个体区域的像素级别的差别，不在本文的研究范围内。因此分割算法将图中的自然猪个体分割成独立的连通域，即判为分割正确。本文采用准确率指标[35]评价 PigNet 模型分割效果，准确率为分割模型正确分割猪个体的图像数占验证集图像数的百分比。

使用已在 COCO（Common objects in context）数据集上训练好的 Mask R－CNN 模型权重对群养猪图像分割网络 PigNet 进行初始化（Fine-tuning）[36]。COCO 数据集由微软提供，数据集中含有类别、位置信息等图像目标的标注信息，是大多数深度神经网络图像分割算法性能评价的常用数据集。COCO 数据集不包含猪，但包含了猫、狗、马、羊、牛等四肢行走动物作为训练对象，训练集包含数十万幅图像。训练好的 Mask R－CNN 模型其权重已包含了动物的常见特征，在此基础上使用迁移学习方法来训练群养猪图像分割网络 PigNet，可以更快地学习猪只特征，大大地缩短了训练模型的时间。

为了进行效果对比，使用 PigNet 模型与基于 ResNet101 网络的 Mask R－CNN 模型及其改进模型对图像进行分割，对比分析其分割结果。

2.3　试验流程

群养猪图像的具体分割试验流程如图 4 所示，具体步骤如下：

(1)从采集的群养猪图像数据中抽取训练集,其中训练图像尽量包含群养猪日常 4 种不同形态图像,且验证集与训练集无交叉重复数据。

(2)使用 Labelme 软件对训练集图像进行标注,对应标签图像,完成训练集的制作。

(3)将制作好的数据集作为深度卷积网络的输入。

(4)搭建群养猪分割卷积网络 PigNet 网络。

(5)以 Mask R-CNN 网络训练好的模型作为群养猪图像分割网络 PigNet 的初始化(Fine-tuning),经训练得 PigNet 模型。

(6)使用验证集图像测试群养猪图像分割模型——PigNet 模型,得分割图像。

(7)采用准确率指标评价 PigNet 模型的分割结果。

(8)根据评价结果调整 PigNet 模型的参数。

(9)重新训练 PigNet 模型,重复步骤(5)~(8),直至训练集指标与验证集指标接近。

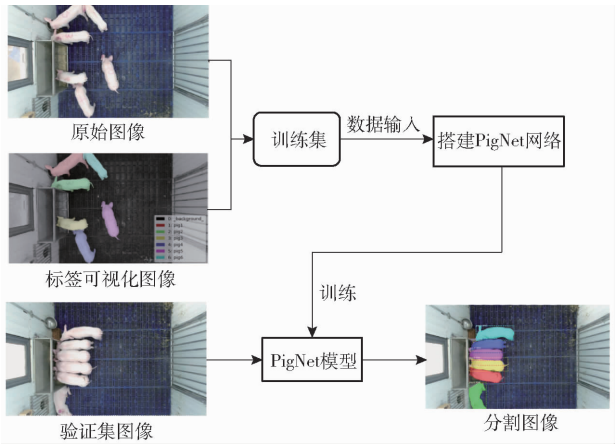

图 4 群养猪图像分割试验流程图

Fig. 4 Flow chart of group pig image segmentation test

3 结果与分析

3.1 PigNet 特征图可视化分析

分析 PigNet 模型的识别过程,模型的部分卷积层的可视化输出如图 5 所示。从图 5 中可以看出,每层卷积提取出的特征各有侧重,使 PigNet 模型学习猪个体颜色特征、形状特征、纹理特征等特征。随着卷积层层数的增多,PigNet 模型学习的特征信息越来越趋于抽象,使 PigNet 模型学习了更为深度的猪个体特征。

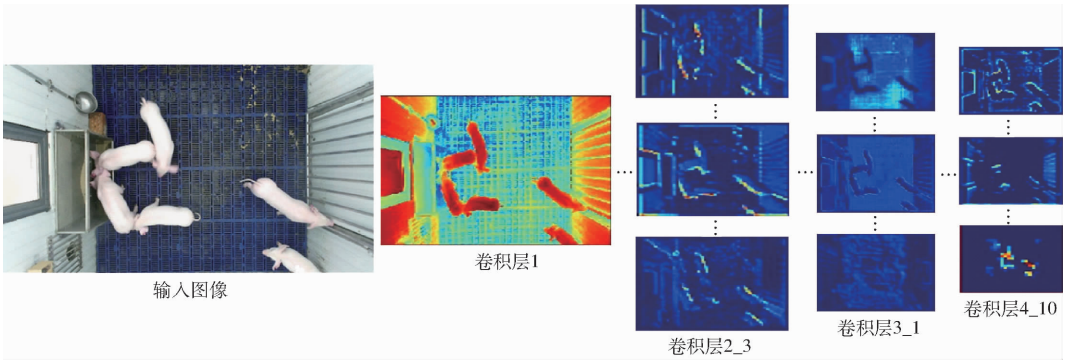

图 5 群养猪图像分割网络算法的卷积特征图可视化

Fig. 5 Convolutional characteristic map visualization of pig image segmentation network algorithms

3.2 PigNet 模型性能分析

已训练的 PigNet 模型在 2 000 幅训练集、500 幅验证集上的分割结果用评价指标分析,结果如表 1 所示。其中,训练集 4 种行为图像各 500 幅,验证集 4 种行为图像各 125 幅。由表 1 可知,PigNet 模型在训练集和验证集上均取得较好的分割结果,总分割准确率均大于 85.00%,训练集为 86.15%,验证集为 85.40%。在各行为样本中,PigNet 模型对吃食、饮水、运动等行为图像分割性能较好,训练集和验证集准确率均不小于 86.40%。验证集对运动行为图像分割准确率高达 92.00%。躺卧样本图像中猪体粘连情况最为严重,所有的样本图像中都有严重的猪体粘连现象,其它行为样本中,粘连现象都不同程度地出现,如图 6 所示。算法的总体准确率及每种行为的验证集准确率与训练集近似,较训练集略低,表明 PigNet 模型泛化性能较好。

表 1 PigNet 分割评价指标分析

Tab. 1 Analysis of PigNet segmentation evaluation index

数据集	行为状态	图像数量/幅	分割正确数量/幅	分割准确率/%
训练集	吃食	500	439	87.80
	饮水	500	436	87.20
	运动	500	465	93.00
	躺卧	500	383	76.60
	总计	2 000	1 723	86.15
验证集	吃食	125	111	88.80
	饮水	125	108	86.40
	运动	125	115	92.00
	躺卧	125	93	74.40
	总计	500	427	85.40

PigNet 模型对分割不同形态的粘连群养猪具有较好的分割性能。如图 6 所示，不同行为下，猪体有保持直线、弯曲、多猪粘连等情况，并未影响分割准确率。由于摄像头并未将整个猪栏区域完全覆盖，有些猪体超出图像范围，算法对拍摄不完整的猪体分割也较为准确，如图 6a、6c、6d 所示。图 6d 的分割难度最大，所有的猪只都聚集在一起，造成某些猪的部分猪体被其它猪只遮挡，导致无法拍摄部分猪只完整轮廓，算法也可准确将其分开，说明 PigNet 网络模型对粘连猪群能有效分割。

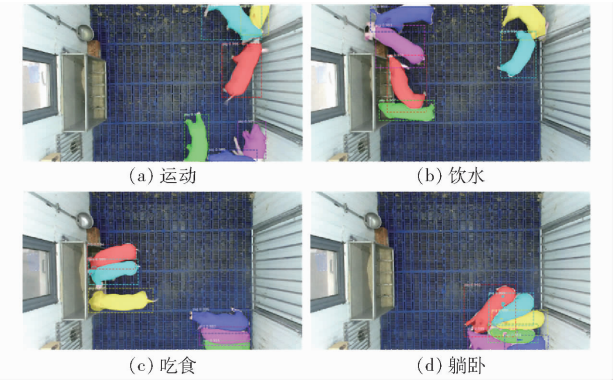

图 6　不同行为测试图像的分割结果

Fig. 6　Segmentation results of diversity behavioral images

在用本文算法完成分割的样本中，两类行为模式的样本出现少量错误，如群养猪的运动和吃食图像，如图 7 所示。猪只个体存在头部与头部、头部与尾部粘连严重或跨爬行为，造成图像中的单个猪体被分为两部分，从而引起分割错误。

图 7　分割错误图像

Fig. 7　Split error images

3.3　改进模型参数设置

为了得到较优的分割粘连猪体图像的网络参数和模型，本文分别对 Mask R－CNN 主干网络的卷积层数和宽度进行了调整。调整后的网络结构模型的参数设置、分割准确率如表 2 所示。从表 2 可以看出，网络模型的深度对分割准确率有明显的影响。改进模型 1 的准确率最高，从最终图像分割的质量来看，改进模型 4～8 的过分割现象明显，故较 101 层和 200 层卷积深度，44 层卷积层深度是较为适合猪群分割的层数。同时，改进模型的第 1 阶段的 conv1 卷积层采用不同卷积核，影响了网络提取的猪体特征，最终体现在了网络模型的分割准确率。由表 2 可看出，改进模型 1、4、6 的分割准确率均高于同一网络深度的其他改进模型。如改进模型 1 较改进模型 3，其分割准确率高出 0.6 个百分点。从表 2 整体来看，卷积层数对网络分割准确率影响较大，卷积核对网络分割准确率影响较小。本文采用改进模型 1 为分割粘连猪体图像的模型，即确定 PigNet 模型卷积层数为 44，第 1 阶段的 conv1 卷积层卷积核为 7×7。

表 2　改进网络的参数设置和分割准确率

Tab. 2　Parameter setting and segmentation accuracy of improved network

模型		卷积核	主网络卷积层数分布					准确率/%
			第 1 阶段	第 2 阶段	第 3 阶段	第 4 阶段	第 5 阶段	
Mask R－CNN 模型		7×7	1	9	12	69	9	74.00
改进模型	1	7×7	1	9	12	12	9	85.40
	2	9×9	1	9	12	12	9	85.20
	3	11×11	1	9	12	12	9	84.80
	4	9×9	1	9	12	69	9	73.40
	5	11×11	1	9	12	69	9	73.20
	6	7×7	1	39	75	75	9	73.40
	7	9×9	1	39	75	75	9	72.80
	8	11×11	1	39	75	75	9	70.40

3.4　不同分割模型对比分析

将本文算法模型与 Mask R－CNN 模型的分割效果进行对比。采用同样的 2 000 个训练样本和 500 个验证样本，训练 Mask R－CNN(ResNet101－FPN)模型。两种网络模型在 500 个验证样本上的分割性能比较如表 3 所示。

表 3　不同分割模型的性能比较

Tab. 3　Comparison of different segmentation networks performances

方法	准确率/%	单幅图像分割时间/s
Mask R－CNN(ResNet101－FPN)	74.00	2.15
本文算法	85.40	2.12

本文算法模型相较 Mask R－CNN(ResNet101－FPN)模型,其准确率高 11.40 个百分点,说明主干网络的卷积层不是越深越好,合适的卷积层深度有利于分割性能的提高。本研究的目的是分割粘连的群养猪,获得独立连通域的猪个体。增加网络卷积层运算和粘连猪体的权重因子,有利于提高分割准确率,但同时也容易导致分割性能的过拟合,将猪体部分区域分割为一个猪个体,造成同一目标猪个体的过分割,如图 7 所示。本文算法模型相较 Mask R－CNN(ResNet101－FPN)模型较好地权衡了分割性能和模型过拟合,一定程度上减少了卷积参数,精简了模型,同时使模型分割性能略有提高。

本文算法模型比 Mask R－CNN(ResNet101－FPN)模型的单幅图像分割处理时间缩短了 30 ms。本文算法模型的层结构较 Mask R－CNN(ResNet101－FPN)模型,精简了卷积层结构,也减少了单幅图像的分割处理时间。

4 结论

(1)提出了基于实例分割框架的粘连猪体图像分割算法 PigNet 网络模型,有效地解决群养猪俯视图像中难以解决的猪体粘连问题。PigNet 网络模型在 Mask R－CNN 主干网络基础上建立 PigNet 主干网络,经过区域候选网络提取的特征和主干网络前向传播提取的特征均被 ROIAlign 所共享。ROIAlign 操作从 ROI 提取特征并保留输入和输出之间的像素精准空间位置。从 ROI 分出两个分支,一个分支完成预测类别和回归框任务,另一 Mask 分支对每个 ROI 的目标预测输出一个掩模。

(2)PigNet 模型在训练集和验证集上均取得较好的结果,总分割准确率均大于 85.00%,训练集为 86.15%,验证集为 85.40%。与 Mask R－CNN 模型及其改进模型进行分割效果对比,PigNet 网络模型验证集的准确率较 Mask R－CNN(ResNet101－FPN)模型高 11.40 个百分点,且分割运算时间较 Mask R－CNN(ResNet101－FPN)模型短 30 ms。本文算法模型对不同形态、粘连严重的群养猪分割效果较好,能准确地分割粘连严重的猪群图像,可用于各种群养行为模式下的粘连猪体分割。

参 考 文 献

[1] 杨秋妹,肖德琴,张根兴.猪只饮水行为机器视觉自动识别[J/OL].农业机械学报,2018,49(6):232－238.
YANG Qiumei, XIAO Deqin, ZHANG Genxing. Automatic pig drinking behavior recognition with machine vision[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 232－238. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20180627&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.06.027. (in Chinese)

[2] KASHIHA M A, BAHR C, HAREDASHT S A, et al. The automatic monitoring of pigs water use by cameras[J]. Computers and Electronics in Agriculture, 2013, 90(1): 164－169.

[3] KASHIHA M A, BAHR C, OTT S, et al. Automatic monitoring of pig locomotion using image analysis[J]. Livestock Science, 2014, 159(1): 141－148.

[4] 刘龙申,沈明霞,柏广宇,等.基于机器视觉的母猪分娩检测方法研究[J/OL].农业机械学报,2014,45(3):237－242.
LIU Longshen, SHEN Mingxia, BO Guangyu, et al. Sows parturition detection method based on machine vision[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(3): 237－242. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20140339&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2014.03.039. (in Chinese)

[5] 张萌,钟南,刘莹莹. 基于生猪外形特征图像的瘦肉率估测方法[J].农业工程学报,2017,33(12):308－314.
ZHANG Meng, ZHONG Nan, LIU Yingying. Estimation method of pig lean meat percentage based on image of pig shape characteristics[J]. Transactions of the CSAE, 2017, 33(12): 308－314. (in Chinese)

[6] 周丽萍,陈达,陈志,等.基于图像处理的生猪耳部颜色异常检测技术[J/OL].农业机械学报,2017,48(4):166－172.
ZHOU Liping, CHEN Da, CHEN Zhi, et al. Pig ear abnormal color detection on image processing techniques[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(4): 166－172. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20170422&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.04.022. (in Chinese)

[7] 郁厚安,高云,黎煊,等.动物行为监测的研究进展——以舍养商品猪为例[J].中国畜牧杂志,2015,51(20):66－70.
YU Houan, GAO Yun, LI Xuan, et al. Research review of animal behavior monitoring technologies: commercial pigs as realistic example[J]. Chinese Journal of Animal Science, 2015, 51(20): 66－70. (in Chinese)

[8] 王德福,黄会男,张洪建,等.生猪养殖设施工程技术研究现状与发展分析[J/OL].农业机械学报,2018,49(11):1－14.
WANG Defu, HUANG Huinan, ZHANG Hongjian, et al. Analysis of research status and development on engineering technology of swine farming facilities[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(11): 1－14. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20181101&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.11.001. (in Chinese)

[9] NAVARRO-JOVER J M, ALCAÑIZ-RAYA M, GÓMEZ V, et al. An automatic colour-based computer vision algorithm for tracking the position of piglets[J]. Spanish Journal of Agricultural Research, 2009, 7(3): 535－549.

[10] SHAO B, XIN H. A real-time computer vision assessment and control of thermal comfort for group-housed pigs[J]. Computers and Electronics in Agriculture, 2008, 62(1): 15 – 21.

[11] 肖德琴,冯爱晶,杨秋妹,等. 基于视频追踪的猪只运动快速检测方法[J/OL]. 农业机械学报, 2016, 47(10): 351 – 357, 331.
XIAO Deqin, FENG Aijing, YANG Qiumei, et al. Fast motion detection for pigs based on video tracking[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(10): 351 – 357, 331. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20161045&journal_id = jcsam. DOI: 10.6041/j.issn.1000-1298.2016.10.045. (in Chinese)

[12] 马丽,纪滨,刘宏申,等. 单只猪轮廓图的侧视图识别[J]. 农业工程学报, 2013, 29(10): 168 – 174.
MA Li, JI Bin, LIU Hongshen, et al. Differentiating profile based on single pig contour[J]. Transactions of the CSAE, 2013, 29(10): 168 – 174. (in Chinese)

[13] 高云,郁厚安,雷明刚,等. 基于头尾定位的群养猪运动轨迹追踪[J]. 农业工程学报, 2017, 33(2): 220 – 226.
GAO Yun, YU Houan, LEI Minggang, et al. Trajectory tracking for group housed pigs based on locations of head/tail[J]. Transactions of the CSAE, 2017, 33(2): 220 – 226. (in Chinese)

[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems, 2012: 1097 – 1105.

[15] REN S, HE K, GIRSHICK R, et al. Faster R – CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137 – 1149.

[16] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// European Conference on Computer Vision, 2014: 818 – 833.

[17] SOOMRO M, FAROOQ M A, RAZA R H. Performance evaluation of advanced deep learning architectures for offline handwritten character recognition[C]// International Conference on Frontiers of Information Technology. IEEE, 2018: 362 – 367.

[18] HU H, GUAN Q, CHEN S, et al. Detection and recognition for life state of cell cancer using two-stage cascade CNNs[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2017, 99: 1 – 7.

[19] CASTRO J D B, FEITOZA R Q, ROSA L C L, et al. A comparative analysis of deep learning techniques for sub-tropical crop types recognition from multitemporal optical/SAR image sequences[C]// 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images, 2017: 382 – 389.

[20] MORI K, MATSUGU M, SUZUKI T. Face recognition using SVM fed with intermediate output of CNN for face detection[C]// IAPR Conference on Machine Vision Applications, 2009: 410 – 413.

[21] GIRSHICK R, DONAHUE J, TREVOR D, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1 – 8.

[22] LONG J, SHELHAMER E, TREVOR D. Fully convolutional networks for semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431 – 3440.

[23] 杨阿庆,薛月菊,黄华盛,等. 基于全卷积网络的哺乳母猪图像分割[J]. 农业工程学报, 2017, 33(23): 219 – 225.
YANG Aqing, XUE Yueju, HUANG Huasheng, et al. Lactating sow image segmentation based on fully convolutional networks [J]. Transactions of the CSAE, 2017, 33(23): 219 – 225. (in Chinese)

[24] SILBERMAN N, SONTAG D, FERGUS R. Instance segmentation of indoor scenes using a coverage loss[M]. Computer Vision – ECCV 2014. Springer International Publishing, 2014: 505 – 520.

[25] LI Yi, QI Haozhi, DAI Jifeng, et al. Fully convolutional instance-aware semantic segmentation[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2359 – 2367.

[26] SILBERMAN N, SONTAG D, FERGUS R. Instance segmentation of indoor scenes using a coverage loss[J]. Computer Vision, 2014, 8689: 616 – 631.

[27] HE K, GKIOXARI G, DOLLAR P, et al. Mask R – CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 1 – 12.

[28] ZHANG Z, FIDLER S, URTASUN R. Instance-level segmentation for autonomous driving with deep densely connected MRFs [J]. Computer Science, 2016, 12: 312 – 329.

[29] WARD D, MOGHADAM P, HUDSON N. Deep leaf segmentation using synthetic data[J]. arXiv: 1807.10931, 2018.

[30] 高云,郭继亮,陈震撼,等. 环境多变量控制养殖试验箱: 201720309723.9 [P]. 2017 – 03 – 28.

[31] ASAE EP344.4 – 2014 lighting systems for agricultural facilities[S]. 2014.

[32] KAMNITSAS K, LEDIG C, NEWCOMBE V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Medical Image Analysis, 2017, 36: 61 – 78.

[33] REN S, HE K, GIRSHICK R, et al. Faster R – CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 – 1149.

[34] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.

[35] BIRD C, NAGAPPAN N, GALL H, et al. Putting it all together: using socio-technical networks to predict failures[C]// IEEE International Conference on Software Reliability Engineering, 2009: 109 – 119.

[36] ZAVAREZ M V, BERRIEL R F, OLIVEIRASANTOS T. Cross-database facial expression recognition based on fine-tuned deep convolutional network[C]// 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images, 2017: 405 – 412.