

# 基于DCNN的马匹图像分割算法研究

张泽宇, 郭斌, 张太红\*

(新疆农业大学 计算机与信息工程学院, 新疆 乌鲁木齐 830052)

**摘要:**随着深度学习的发展,卷积神经网络被用于解决各类计算机视觉问题。物体的分割是图像处理的基础,传统的分割算法在特定强光噪声场景下对图片的理解能力及工作效率欠佳。针对传统马匹图像分割算法中存在的分割边缘不清晰、强光照射下马匹轮廓分割模糊,以及分割前需要进行手工标注等问题,提出一种基于全卷积神经网络并加入条件随机场的马匹图像分割算法。建立马匹图像数据集进行训练测试,网络中加入带有空洞卷积的空间金字塔池化模型(ASPP)、空洞卷积与深度分离卷积使得网络在参数不变的情况下扩大感受野从而有效地分割出马匹。在深度卷积神经网络(DCNN)模型上进行迁移学习,并加入全连接条件随机场(CRF)优化分割出来的马匹边缘轮廓。实验结果表明,该方法在分割精度上优于传统的马匹图像分割算法,平均交并比(MIOU)达到了92.8%。

**关键词:**马图像分割;语义分割;条件随机场;深度分离卷积;迁移学习

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1673-629X(2020)10-0210-06

doi:10.3969/j.issn.1673-629X.2020.10.037

## Research on Horse Image Segmentation Algorithm Based on DCNN

ZHANG Ze-yu, GUO Bin, ZHANG Tai-hong\*

(School of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

**Abstract:** With the development of deep learning, convolutional neural network is used to solve many computer vision problems. Image segmentation is the basis of all image processing. Traditional segmentation algorithm has poor understanding of images and work inefficiency under certain high-light noise scene. In view of the problems existing in the traditional horse image segmentation algorithm, such as unclear segmentation edge, fuzzy segmentation of horse contour under strong illumination, and manual annotation before segmentation, we propose a horse image segmentation algorithm based on full convolutional neural network and conditional random field. By establishing horse image dataset to train and adding the atrous spatial pyramid pooling (ASPP), the atrous convolution and the depthwise separable convolution, it makes the full convolutional neural network model enable the network to expand the receptive field with the parameters unchanged and segment the horse image effectively. The horses are migrated on the deep convolutional neural network (DCNN), and the horse edge contours segmented are added by the fully connected condition random field. The experiment shows that the proposed method is superior to the traditional horse image segmentation algorithm in terms of segmentation accuracy. The mean intersection over union (MIOU) reaches 92.8%.

**Key words:** horse image segmentation; semantic segmentation; conditional random field; depthwise separable convolution; transfer learning

## 0 引言

随着人工智能技术的飞速发展,人类社会正在从工业社会向智能社会迈进,信息智能化是当今世界发展的趋势,其将会对新世纪全球的经济产生深远的影响。新疆是国内的养马地区之一,有着悠久的养马历史及浓厚的马文化底蕴。在马匹登记测量检测过程中,由于马匹体型庞大,性格不如牛、羊等家畜温和,使得测量工作难度加大。

随着深度学习在计算机视觉领域的应用,使得基于深度学习的图像分析处理技术逐步深入到畜牧养殖的各个领域。

其中通过图像来识别及测量动物的体尺数据成为当今研究的热点,而拍摄图像分割结果的优劣是一切通过图像进行动物行为识别、体尺数据测量、体重检测、健康状况分析的基础,图像分割效果的优劣决定了后续工作的可靠性。

收稿日期:2019-11-22

修回日期:2020-03-23

基金项目:新疆维吾尔自治区重大科技专项(2017A01002-5)

作者简介:张泽宇(1994-),男,硕士研究生,研究方向为计算机视觉;张太红,通讯作者,教授,研究方向为数据库技术。

## 1 概述

基于深度学习的图像分割方法是当今计算机视觉领域的研究热点。2014 年 Long 等人<sup>[1]</sup>提出了 FCN (fully convolutional networks) 方法用于图像语义分割,此算法使用反卷积进行上采样并将全连接网络换为卷积网络,使得网络模型可以输入任意尺寸大小的图片,并且输出与原图像相同大小的像素级分类结果。FCN 开启了利用全卷积神经网络进行图像分割的热潮。2015 年 Olaf Ronneberger 等人<sup>[2]</sup>提出了 U-Net 算法,此算法广泛应用于生物医学图像分割,算法中应用编码层解码层,对卷积加深处理,对数据的增强使得带标记数据需求量减少,使得算法在训练时间方面得以缩短优化。2018 年 Liang-Chieh Chen 等人<sup>[3-6]</sup>提出的 Deeplab 系列算法,引入空洞卷积 (atrous convolution),在不改变图像的全局信息下,增加了感受野,在特征图缩小到相同倍数的情况下获取图像更多的全局信息,大大提高了算法的精准度。使得在复杂多类图片上的语义分割效果大幅提升。2018 年 He Kaiming 等人<sup>[7]</sup>提出 Mask R-CNN 算法,此算法对物体进行定位与分割,结合了目标检测与语义分割。在国内,白明月等<sup>[8]</sup>提出基于 YCbCr 空间改进 CV 主动轮廓模型的分割方法,对具有复杂背景的羊体图像进行分割,此算法在分割之前需要手工勾画,然后通过模型进行分割,分割物体边缘有待优化。胡志伟等<sup>[9]</sup>提出基于 VGG16 与 U-NET 的全卷积神经网络模型,以 30 个视频文件为原始数据,从视频中截取图像作为数据集进行训练验证测试,实现了生猪轮廓提取,且分割像素精度 (PA) 达到 94.32%,平均交并比 (MIOU) 达到 86.60%。李林葳等<sup>[10]</sup>使用改进的遗传算法对 OTSU 算法进行优化,实现了生猪目标提取,较原算法在计算时间上提升了 5%。高云等<sup>[11]</sup>提出 PigNet 网络,在环境多变控制养殖箱中拍摄图片作为数据集,解决粘连猪体的分割问题,实现猪体的实例分割。以上算法在有强光照噪声图像中,还存在分割边缘不清晰、轮廓模糊的问题。张新峰等<sup>[12]</sup>使用全卷积神经网络结合全连接条件场,在预训练 Pascal voc2012 数据集生成的模型参数基础上,训练舌图像,细化了分割舌体边缘,由于在预训练数据集中并没有舌图像,使得模型参数过多,增加了模型文件存储大小。

基于以上研究进展,该文提出一种结合深度神经网络 (DCCN)<sup>[13]</sup>与全连接条件随机场<sup>[14]</sup>的分割算法,实现马匹在不同光照环境下的语义分割。该模型以 Xception<sup>[15]</sup>结构为模型主干,使用两种类型的神经网络;加入空洞卷积的空间金字塔模块 (ASPP)<sup>[16]</sup>和编码解码架构 (encoder-decoder)<sup>[17]</sup>进行语义分割,结合全连接条件随机场优化分割边缘。通过空间金字塔模

块在不同分辨率上的图片进行池化操作以捕获丰富的上下文信息,编码解码结构实现物体的粗略边界,高层特征提供语义信息,解码层恢复马匹边界信息。全连接条件随机场实现全局边缘优化,去除分割边缘的噪声,对图像中所有的像素点建立边缘势能,从而实现细化分割效果。

## 2 相关模型

### 2.1 空洞卷积

传统的分类网络模型通过连续池化处理或其他下采样层来整合多尺度上下文信息,这种方式会损失部分像素。该文使用空洞卷积进行多尺度上下文信息聚合而不降低特征图大小,并且空洞卷积支持感受野的指数增长。空洞卷积的结构如图 1 所示,空洞卷积在二维运算中的表达式如下:

$$y[i] = \sum_k x[i + r \cdot k] \cdot \omega[k] \quad (1)$$

其中,  $y[i]$  为输出特征图,  $\omega$  为卷积核,  $x[i]$  为输入特征图,参数  $r$  为扩张率,用来控制空洞卷积的卷积间隔。当参数  $r$  为 1 时,此时的卷积核为  $1 \times 1$  的标准卷积核,可以通过更改扩张率来自适应地修改卷积的感受野。

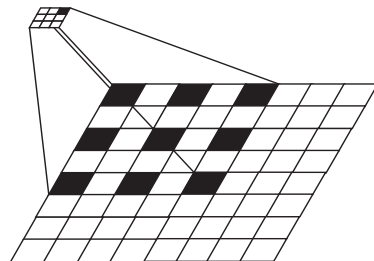


图 1 空洞卷积结构

### 2.2 深度分离卷积

深度分离卷积将一个完整的卷积运算分解为两步进行:(1)空间卷积 (depthwise convolution);(2)逐点卷积 (pointwise convolution)。空间卷积针对每个输入通道独立执行对空间的卷积,逐点卷积用于组合空间卷积的输出。深度分离卷积结构与常规卷积操作类似,可以用来提取特征,但是相比较于常规卷积操作,参数量与运算成本降低。深度分离卷积原理如图 2 所示。

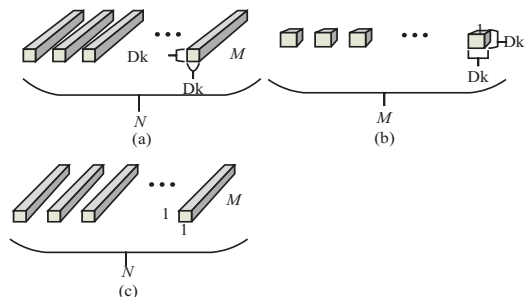


图 2 深度分离卷积原理

传统卷积如图 2(a) 所示, 设输入的特征图大小为  $Df \times Df \times M$ , 输出特征图大小为  $Df \times Df \times N$ , 卷积核大小为  $Dk \times Dk$ , 则传统卷积的计算量  $C_1$  为:

$$C_1 = Df \times Df \times M \times Df \times Df \times N$$

深度可分离卷积的计算量为空间卷积 (depthwise convolution) 与  $1 \times 1$  卷积 (pointwise convolution) 的计算量之和, 表示为  $C_2$ :

$$C_2 = Dk \times Dk \times M \times Df \times Df + M \times N \times Df \times Df$$

对应图 2(b)、(c)。

### 2.3 结合空洞卷积的空间金字塔模型

目前流行的 CNN 都需要固定输入图片的尺寸, 所以基本上都是通过剪裁 (crop) 和缩放 (wrap)。这种处理方式存在三个弊端: ①剪裁出的图片可能不包含整个物体; ②图像的缩放会导致物体变形; ③当图像中物体大小改变后, 预定义的比例 (scale) 可能不适合物体的变化。

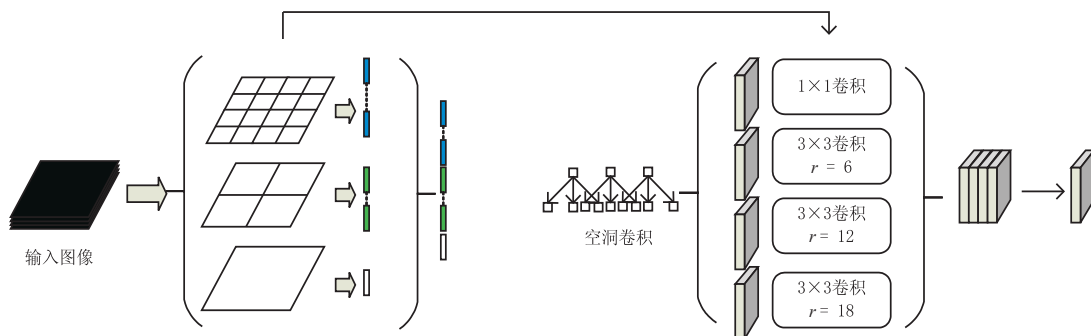


图 3 加入空洞卷积的空间金字塔池化模型结构

### 2.4 全连接条件随机场

DCNN 的固有特性包括定位的精度及目标的分类, 拥有多个最大池化层的深度网络模型在目标分类中的效果是显著的, 但是顶层节点的不变性及较大的感受野只能解决平滑的响应。DCNN 分数图 (score maps) 只能预测粗略的物体位置, 却不能精细地描绘出物体的边缘轮廓。

条件随机场 (CRF) 的作用是平滑分割图片的噪声。此模型通过耦合相邻节点, 以促进空间上相邻节点的像素获得相同的标签, 当然这些短距离 CRF 的主要功能是消除通过人工标记形成具有局部特征弱分类器的错误预测。

该文使用的深度全卷积神经网络不同于较弱的分类器, 在分割预测中使用的架构产生了分数图 (score maps) 以及语义标签的预测值。为了更好地细化边缘轮廓, 引入全连接条件随机场, 将 DCNN 模型与全连接条件随机场结合, 在 DCNN 分割的结果上进一步优化分割结果的边缘。全连接条件随机场的能量函数表达式如式 (2):

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \quad (2)$$

CNN 网络对于固定输入的要求, 主要在全连接的分类器层, 而特征提取层可以通过控制子采样比例和卷积核大小来调节, 从而接受各种比例和尺寸的输入, 得到固定的特征输出。该文引入空间金字塔层 (SPP), 通过 SPP 来移除 CNN 对于固定输入尺寸大小的要求, SPP 操作类似深层的信息“融合”。空间金字塔模型首先是输入层, 尺寸大小可以是任意的, 然后进行卷积运算, 到最后一个卷积层输出该层的特征图 (feature map), 之后进入由三个并行排列的 3 种不同参数的卷积核所在的 SPP 层, 最终将 3 种不同参数卷积核产生的特征合并, 形成一个固定大小的输出。使用加入空洞卷积的空间金字塔模型, 将上述三种不同参数卷积核换为并行排列的 4 种不同扩张率的空洞卷积, 对图像中的物体进行多尺度特征的提取, 每个通道提取到的特征进行融合, 再使用大小为  $1 \times 1$  的卷积核进行卷积操作, 结构如图 3 所示。

其中,  $x$  是像素的标签, 一元势能定义为  $\theta_i(x_i) = -\log P(x_i)$ , 其中  $P(x_i)$  表示由 DCNN 预测出的像素为  $i$  的马匹标签值概率。二元势能定义如式 (3):

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[ \omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2}\right) - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right] + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right) \quad (3)$$

其中当  $x_i = x_j$  时,  $\mu(x_i, x_j) = 0$ , 否则为 1。在波茨模型 (Potts model) 中, 表示只惩罚具有不同标签值的节点。余下表达式在不同的特征空间中使用两个高斯核, 第一个核是像素位置  $P$  和 RGB 颜色  $I$  之间的双向核, 第二个核是依赖像素位置处理平滑的单核,  $\omega$  为权重, 超参数  $\sigma_\alpha, \sigma_\beta, \sigma_\gamma$  控制高斯核的尺度。

### 2.5 马匹图像分割算法

算法网络结构如图 4 所示。该算法由 2 部分组成: (1) 使用加入空洞卷积的空间金字塔池化模型和融合深度分离卷积的深度卷积网络; (2) 全连接条件随机场。

首先将马匹图像输入到深度卷积网络中进行特征

提取,使用空洞卷积,在不增加参数的情况下,扩大感受野大小,捕捉更多的尺度信息,并控制前向传播过程中被计算的特征分辨率,加入深度分离卷积,将空洞卷积与深度分离卷积结合,构成空洞分离卷积 (atrous separable convolution),很大程度上减少了计算复杂度,利用 ASPP 模块通过结合多个不同扩张率 (dilation rate) 的深度分离卷积提取多尺度信息。之后利用全

连接条件随机场,对分割边缘进行优化,算法模型的步骤分为:(1) 图像预处理;(2) 建立马匹图像数据集;(3) 人工标注训练集、验证集、测试集;(4) 在 DCNN 上进行迁移学习;(5) 生成语义分割模型并用测试集测试;(6) 全连接条件随机场优化分割边缘;(7) 马匹图像分割结果评价。

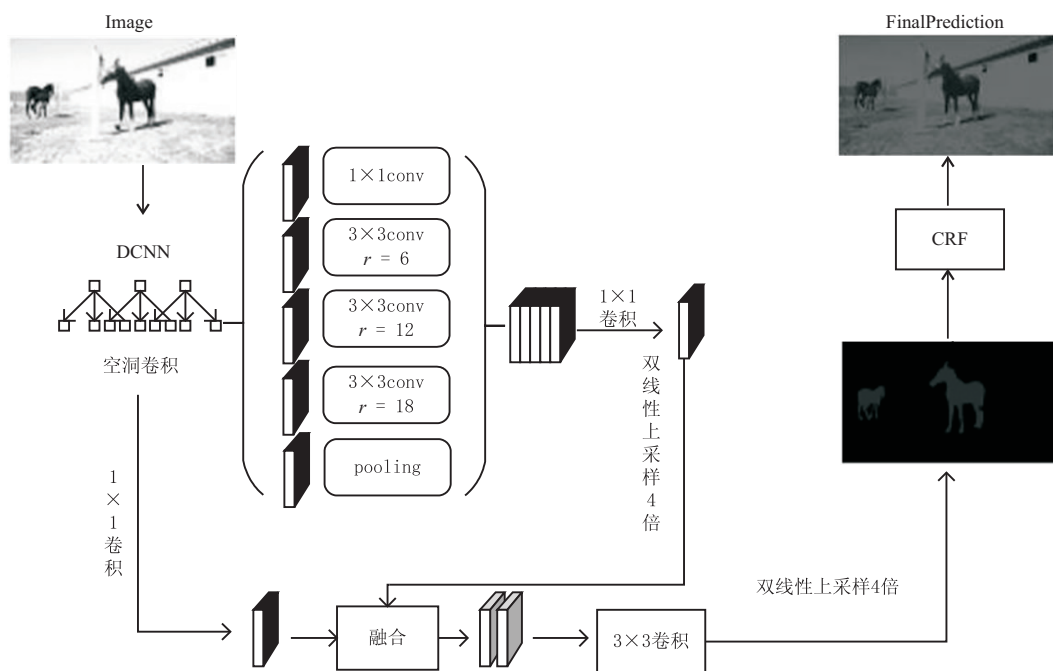


图 4 马匹图像分割算法结构

### 3 实验

#### 3.1 实验数据集

实验数据拍摄于新疆伊犁昭苏县昭苏马场及昭苏种马场,采集设备为索尼相机 NEX-5T,图像分辨率为  $3\,658 \times 2\,000$ ,马匹图像分早中晚不同时间段拍摄,按照采集的天气条件可以分为光照充足晴天、阴天及阴雨天。景深  $13.3\text{ m} (\pm 0.5)$ ,拍摄高度  $1.5\text{ m}$ ,曝光率 F5.6。马匹数据集包含训练集 2 000 张、测试集 100 张,验证集 50 张。原始图像分别于马匹的正面、左右前侧、左右正面、左右后侧面共 7 个方向拍摄。



图 5 图像标记标签图

预处理:原拍摄图片使用 Mac 软件 Pixelmator 进行图像翻转用于数据集扩充,以保证不同拍摄角度的训练样本对算法准确度的影响,然后使用 TinyPNG 进行图片的压缩,之后使用 MIT 开源软件 Labelme<sup>[18]</sup>进行图像标记,人工每标记一匹马的轮廓就赋予具体的

马匹标签,标记完成后生成 .json 文件,之后使用 Python 脚本生成灰度图。数据集原始图像与标记图像如图 5 所示。

#### 3.2 实验设计

图像分割模型在极客云 GPU 深度学习服务器上进行训练,主机配置为:显卡使用 Titan RTX,显存为 24 G,处理器为英特尔 E5-2680,内存 32 G。训练集、验证集、测试集的所有图像实验前全部采用 MIT 开源标记软件 Labelme 进行人工标注。训练过程中网络参数分别设置为:基础学习率 (base\_learning\_rate) 设置为 0.000 1,学习率的衰减率 (learning\_rate\_decay\_rate) 设置为 0.1,训练循环次数 (training\_steps) 设置为 12 000,批处理数据量大小 (batch\_size) 设置为 4,训练裁剪尺寸 (train\_crop\_size) 设置为  $513 \times 513$ 。

本数据集中分割案例为马匹图像分割,属于 2 分类问题,其中背景占了很大比例,故设置的比例为 1 : 10。Pascal voc 图像语义分割数据集上有 21 个类别,其中就包括马、牛等动物的数据。因此该文基于 DCNN 在 Pascal Voc 上的训练模型,之后加入马匹图像数据集进行迁移学习,使得网络可以更准确地学习马匹特征,大幅度减少了训练模型的时间。



### 3.3 实验评价标准

评价分割结果指标分别为像素精度 (pixel accuracy, PA)、均像素精度 (mean pixel accuracy, MPA) 与均交并比 (mean intersection over union, MIOU), 各项指标的定义式如下:

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (4)$$

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (5)$$

$$MIOU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ij} - p_{ii}} \quad (6)$$

其中, PA 表示正确标记的像素占总像素的比, 反映出分割像素的准确性; MIOU 表示真实值与预测像素之间的交集并集之比, 能充分反映模型分割位置的准确性及分割区域的完整性。上述指标为衡量图像语义分割的标准。

### 3.4 实验结果分析

网络分割效果受不同扩张率参数的空洞卷积影

响。对比结果如表 1。实验中卷积核尺寸大小为  $3 \times 3$ , ASPP 模块中每一个分支的空洞卷积扩张率参数  $r$  不同, 其中 ASPP-8 代表扩张率  $r$  分别为 (1, 8, 16, 24), ASPP-6 代表扩张率  $r$  为 (1, 6, 12, 18), ASPP-4 代表扩张率  $r$  为 (1, 4, 8, 12), ASPP-2 代表扩张率  $r$  为 (1, 2, 4, 6)。在使用加入空洞卷积的空间金字塔模块中, ASPP-6: (1, 6, 12, 18) 的 MIOU 最高, 在经过全连接条件随机场处理后, MIOU 值达到 92.8%

表 1 不同扩张率参数对算法分割的影响

名称	不同扩张率大小	MIOU/%
ASPP-8	(1, 8, 16, 24)	90.7
ASPP-6	(1, 6, 12, 18)	91.7
ASPP-4	(1, 4, 8, 12)	89.5
ASPP-2	(1, 2, 4, 6)	88.7

DCNN 在不同网络基础 (backbone) 上的分割结果对比图, 其中包括 Mobilenet<sup>[19]</sup> 和 Xception<sup>[15]</sup>, 以及提出的算法的分割结果, 如图 6 所示。

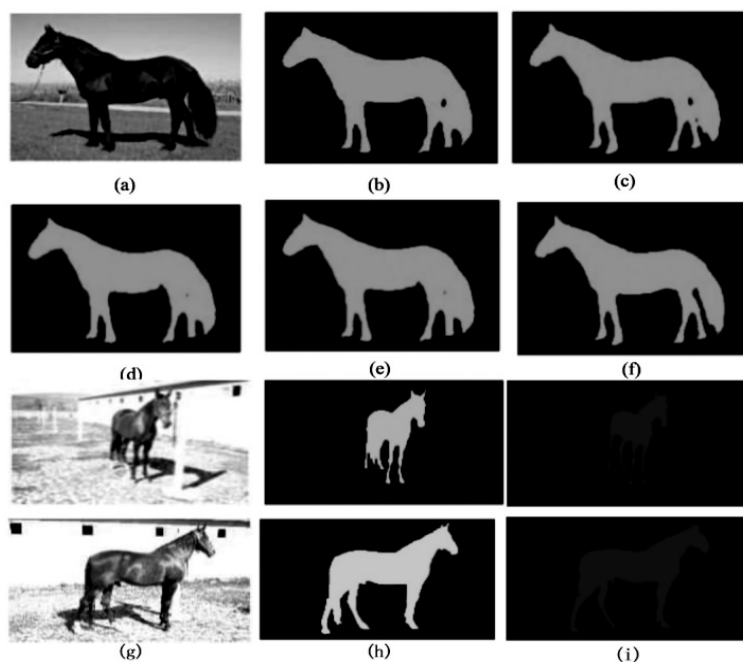


图 6 分割结果

(a) 原图像 (b)(c) Mobilenetv2 网络基础上的分割结果 (d)(e) Xception 网络基础上的分割结果  
(f) 迁移学习后的分割结果 (g) 原图像 (h) 文中模型分割结果 (i) Ground truth

## 4 结束语

提出了结合深度卷积神经网络和全连接条件随机场算法, 结合空洞卷积、深度分离卷积和空间金字塔池化模型, 在深度卷积神经网络上通过自己建立的马匹图像数据集进行迁移学习, 最后使用全连接条件随机

场完成最后的分割边缘优化, MIOU 值达到 92.8%。解决了强光照下分割效果不清晰、效果不好和分割前需要人工手动标注分割候选区等问题。实现马匹在不同光照环境下的语义分割。在原模型上分割精度提升 3.8%。

该方法具有稳定有效的特性, 模型分割的马匹个

体轮廓清晰,从而为深度卷积神经网络技术在马匹上的应用提供了思路,为通过计算机视觉估测马匹体尺体重提供了技术支持。

#### 参考文献:

- [1] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.
- [2] 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. Munich: Springer, 2015:234-241.
- [3] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40(4):834-848.
- [4] LIU C, CHEN L C, SCHROFF F, et al. Auto-deeplab: hierarchical neural architecture search for semantic image segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2019:82-92.
- [5] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European conference on computer vision (ECCV). Munich: [s. n.], 2018:801-818.
- [6] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 40(4):834-848.
- [7] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of the IEEE international conference on computer vision. Venice: IEEE, 2017:2961-2969.
- [8] 白明月, 薛河儒, 姜新华, 等. 改进 C-V 模型在 YCbCr 空间下的羊体图像分割 [J]. 中国农业大学学报, 2018, 23(10):137-143.
- [9] 胡志伟, 杨 华, 娄甜田, 等. 基于全卷积网络的生猪轮廓提取 [J]. 华南农业大学学报, 2018, 39(6):111-119.
- [10] 李林葳, 刘振宇, 李砚禾, 等. 基于机器视觉的生猪目标提取算法设计与实现 [J]. 山西农业大学学报:自然科学版, 2019, 39(5):101-105.
- [11] 高 云, 郭继亮, 黎 焯, 等. 基于深度学习的群猪图像实例分割方法 [J]. 农业机械学报, 2019, 50(4):179-187.
- [12] 张新峰, 郭宇桐, 蔡轶珩, 等. 基于 DCNN 和全连接 CRF 的舌图像分割算法 [J]. 北京航空航天大学学报, 2019, 45(12):2364-2374.
- [13] SAINATH T N, MOHAMED A, KINGSBURY B, et al. Deep convolutional neural networks for LVCSR [C]//2013 IEEE international conference on acoustics, speech and signal processing. Vancouver: IEEE, 2013:8614-8618.
- [14] SUTTON C, MCCALLUM A. An introduction to conditional random fields [J]. Foundations and Trends in Machine Learning, 2012, 4(4):267-373.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Venice: IEEE, 2017:1251-1258.
- [16] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2015, 37(9):1904-1916.
- [17] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495.
- [18] RUSSELL B C, TORRALBA A, MURPHY K P, et al. LabelMe: a database and web-based tool for image annotation [J]. International Journal of Computer Vision, 2008, 77(1-3):157-173.
- [19] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//IEEE international conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018:4510-4520.