

# 深度残差网络的无人机多目标识别

翟进有<sup>1</sup>, 代冀阳<sup>1,2</sup>, 王嘉琦<sup>1</sup>, 应进<sup>1</sup>

(1. 南昌航空大学信息工程学院, 江西 南昌 330063;  
2. 南昌航空大学无损检测技术教育部重点实验室, 江西 南昌 330063)

**摘 要:** 传统目标识别算法中, 经典的区域建议网络(RPN)在提取目标候选区域时计算量大, 时间复杂度较高, 因此提出一种级联区域建议网络(CRPN)的搜索模式对其进行改善。此外, 深层次的卷积神经网络训练中易产生退化现象, 而引入残差学习的深度残差网络(ResNet), 能够有效抑制该现象。对多种不同深度以及不同参数的网络模型进行研究, 将两层残差学习模块与三层残差学习模块结合使用, 设计出一种占用内存更小、时间复杂度更低的新型多捷联式残差网络模型(Mu-ResNet)。采用 Mu-ResNet 与 CRPN 结合的网络模型在无人机目标数据集以及 PASCAL VOC 数据集上进行多目标识别测试, 较使用 ResNet 与 RPN 结合的网络模型, 识别准确率提升了近 2 个百分点。

**关 键 词:** 无人机; 残差网络; 级联区域建议网络; 目标识别

中图分类号: TP 391.4

DOI: 10.11996/JGj.2095-302X.2019010158

文献标识码: A

文章编号: 2095-302X(2019)01-0158-07

## Multi-Objective Identification of UAV Based on Deep Residual Network

ZHAI Jin-you<sup>1</sup>, DAI Ji-yang<sup>1,2</sup>, WANG Jia-qi<sup>1</sup>, YING Jin<sup>1</sup>

(1. School of Information Engineering, Nanchang Hangkong University, Nanchang Jiangxi 330063, China;

2. Key Laboratory of Nondestructive Testing, Nanchang Hangkong University, Ministry of Education, Nanchang Jiangxi 330063, China)

**Abstract:** In traditional target recognition algorithms, the classical region proposal net (RPN) has large amount of computation and high complexity of time at extracting the target candidate region. Cascade region proposal network (CRPN) is proposed as a new search method for improving the performance of RPN, in which residual learning based deep residual network (ResNet) is also used effectively to suppress the degradation phenomenon in deep-level convolution neural networks. Aimed at the network models with different depths and parameters, a novel multi-strapdown residual network (Mu-ResNet) model, which is of less memory and lower time complexity, is designed by combining two-layer and three-layer residual learning modules. The combination model of Mu-ResNet and CRPN is used for multi-target recognition test by using the unmanned aerial vehicle (UAV) target data and PASCAL VOC data. The results have shown that nearly 2% of recognition accuracy is increased compared with the combination model of ResNet and RPN.

**Keywords:** unmanned aerial vehicle (UAV); residual network; cascade region proposal network; target recognition

收稿日期: 2018-05-23; 定稿日期: 2018-09-03

基金项目: 国家自然科学基金项目(61663030, 61663032)

第一作者: 翟进有(1994-), 男, 江西上饶人, 硕士研究生。主要研究方向为计算机视觉、无人机感知与规避及深度学习。E-mail: 710644534@qq.com

通信作者: 代冀阳(1966-), 男, 江西九江人, 教授, 博士。主要研究方向为先进控制理论及应用、飞行器控制设计。E-mail: 112357057@qq.com

卷积神经网络属于神经网络的一个分支, 目前国际上有关卷积神经网络的学术研究进行的如火如荼, 且该技术在计算机视觉、模式识别等领域成功得到应用<sup>[1-2]</sup>。卷积神经网络的优点非常明显, 其是一种多层神经网络, 通过卷积运算和降采样对输入图像进行处理, 其网络权值共享, 可以有效减少权值数目, 降低模型复杂度, 且该网络结构具有高度的尺度不变性、旋转不变性等多种仿射不变性。

近两年的 ILSVRC (imagenet large scale visual recognition challenge) 竞赛中, 在目标检测、分类、定位等项目上前 3 名的获奖者均使用了神经网络的算法, 如文献[3]。在自主目标识别领域, 基于机器学习的方法越来越展现出其强大的生命力, 学习能力是智能化行为的一个非常重要的特征。随着人工智能的发展, 机器学习的方法以其强大的泛化能力, 不断在目标识别领域创造突破, 也日渐成为计算机视觉各领域的研究主流。

深度卷积神经网络<sup>[4-5]</sup>为图像分类带来了许多突破<sup>[6-7]</sup>。深层神经网络以端到端的方式自然地集成了低、中、高级功能和类别<sup>[8]</sup>, 功能的“级别”可以通过堆叠层数(深度)来丰富。近年来, 在十分具有挑战性的 ImageNet 数据集上进行图像分类取得领先的结果都是采用了较深的网络。当更深的网络能够开始收敛时, 就会暴露出退化的问题, 随着网络深度的增加, 精度达到饱和(这并不奇怪), 然后迅速退化<sup>[9-11]</sup>。这种降级不是由过度设置引起的, 是因在适当深度的模型中添加更多的层而导致了更高的训练误差。

## 1 深度残差网络搭建

### 1.1 深度卷积神经网络

在 2012 年的 ILSVRC 图像分类大赛上, KRIZHEVSKY 等<sup>[4]</sup>提出了 AlexNet 这一经典的深度卷积神经网络模型。AlexNet 相较于浅层的卷积神经网络模型, 在性能上有了巨大的提升, 而深层的网络模型较浅层的网络模型具有更大的优势。VGG16 网络以及 GoogleNet 网络不断刷新着 ILSVRC 竞赛的准确率。从 LeNet, AlexNet 至 VGG16 和 GoogleNet, 卷积神经网络的层数在不断加深。伴随着网络层数的加深, 数据量以及运算量也在剧增。卷积模块的数学表达为

$$\begin{cases} y = \text{conv}(x, w) = (y(1), y(2), \dots, y(n-m+1)) \\ \quad \in R^{n-m+1} \\ y(t) = \sum_{i=1}^m x(t+i-1)w(i) \end{cases} \quad (1)$$

其中,  $x$  为输入信号;  $w$  为卷积核尺寸大小且  $w \in R^{s \times k}$ ,  $x \in R^{n \times m}$ ,  $t=1, 2, \dots, n-m+1$ , 且  $n > m$ 。卷积操作的核心是: 可以减少不必要的权值连接, 引入稀疏或局部连接, 带来的权值共享策略大大地减少参数量, 相对地提升了数据量, 从而可以避免过拟合现象的出现; 另外, 因为卷积操作具有平移不变性, 使得学习到的特征具有拓扑对应性和鲁棒性。

### 1.2 深度残差学习机制

当卷积神经网络层数过深, 容易出现准确率下降的现象。通过引入深度残差学习可以有效解决退化问题。残差学习并不是每几个堆叠层直接映射到所需的底层映射, 而是明确地让其映射到残差层中<sup>[12]</sup>。训练深层次的神经网络是十分困难的, 而使用深度残差模块搭建网络可以很好地减轻深层网络训练的负担(数据量)并且实现对更深层网络的训练<sup>[13]</sup>。残差网络的学习机制如图 1 所示。

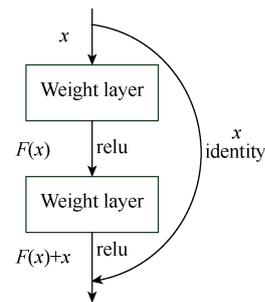


图 1 残差网络学习模块

残差学习模块也是深度卷积神经网络的一部分<sup>[12]</sup>。现假设残差学习模块的输入为  $x$ , 要拟合的函数映射即输出为  $H(x)$ , 那么可以定义另一个残差映射为  $F(x)$ , 且  $F(x)=H(x)-x$ , 则原始的函数映射  $H(x)=F(x)+x$ 。文献[12]通过实验证明, 优化残差映射  $F(x)$  比优化原始函数映射  $H(x)$  容易的多。 $F(x)+x$  可以理解为在前馈网络中, 主网络的输出  $F(x)$  与直接映射  $x$  的和。直接映射只是将输入  $x$  原封不动地映射到输出端, 并未加入任何其他参数, 不影响整体网络的复杂度与计算量。残差学习模块的卷积层使用 2 个  $3 \times 3$  卷积模块, 进行卷积运算<sup>[13-14]</sup>。引入了残差学习机制的网络依然可以使用现有的深度学习反馈训练模式求解。

另一种 3 层残差学习模块, 与 2 层不同的是第一个卷积层采用了  $1 \times 1$  的卷积核, 对数据进行降维, 再经过一个  $3 \times 3$  的卷积核, 最后在经过一个  $1 \times 1$  的卷积核还原。这样做既保存了精度, 又能够有效地减少计算量, 对于越深层次的网络效果往往更好。图 2 为 3 层残差学习模块, 由于其是先降维再卷积

再还原, 这种计算模式对于越深层次的网络训练效果越好。

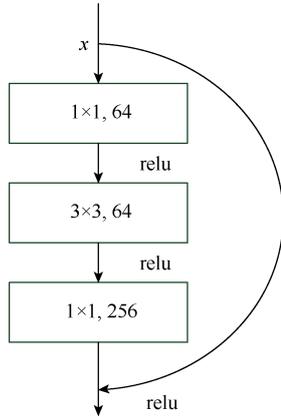


图 2 3 层残差学习模块

### 1.3 搭建深度残差网络模型

本文设计出一种包含多条捷径通道多捷联式残差学习模块, 结合文献[10]中 2 种残差模块的优势, 多捷联式残差模块使用了 5 层的堆栈, 由 1 个  $1 \times 1$  的卷积层进行降维, 再经过 3 个  $3 \times 3$  的卷积层, 最后经过一个  $1 \times 1$  的卷积层还原。在第一个  $1 \times 1$  的卷积层之前和第二个  $3 \times 3$  的卷积层之后使用一条捷径, 第三个  $3 \times 3$  的卷积层之前和最后一个  $1 \times 1$  的卷积层之后使用一条捷径, 再将第一个  $1 \times 1$  的卷积层之前和最后一个  $1 \times 1$  的卷积层之后使用一条捷径。通过设置多条捷径减轻网络的训练难度, 采用降维以卷积的方式减少训练的时间复杂度。多捷联式残差模块如图 3 所示。

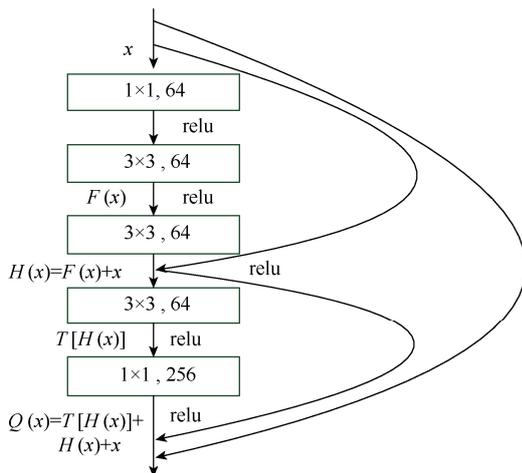


图 3 捷联式残差模块

前三层卷积层定义为多捷联式残差模块的第一级, 后两层网络定义为第二级。现假设多捷联式残差模块的输入为  $x$ , 第一级的残差映射为  $F(x)$ , 要拟合的函数映射即第一级的输出为  $H(x)$ , 则

$$F(x)=H(x)-x \quad (2)$$

第二级输入函数为  $H(x)=F(x)+x$ 。第二级残差映射为  $T[H(x)]$ , 第二级的输出  $Q(x)$ , 即

$$Q(x)=T[H(x)]+H(x)+x \quad (3)$$

$Q(x)$  为多捷联式残差学习模块的最终输出, 该模块设置了多条捷径, 在深层的网络中训练更易收敛。

针对输入为任意大小的多类别目标图像, 无人机(unmanned aerial vehicle, UAV)目标识别系统采用卷积神经网络和残差学习机制结合。采用五层残差学习模块搭建了多种不同深度的残差网络(multi-strapped ResNet, Mu-ResNet)。

## 2 无人机目标识别算法

### 2.1 级联区域搜索网络

在 Faster-RCNN 算法出现之前, 最先进的目标检测网络都需要先用区域选择建议算法推测目标位置, 例如 SPPnet 和 Fast-RCNN 虽然都已经做出相应的改进, 减少了网络运行的时间, 但是计算区域建议依然需要消耗大量时间。所以, REN 等<sup>[15]</sup>提出了区域建议网络(region proposal network, RPN)用来提取检测区域, 其能和整个检测网络共享所有的卷积特征, 使得区域建议的时间大大减少。

简而言之, Faster-RCNN 的核心思想就是 RPN, 而 RPN 的核心思想是使用 CNN 卷积神经网络直接产生 region proposal, 使用的方法就是滑动窗口(只在最后的卷积层和最后一层特征图上滑动), 即 anchor 机制<sup>[15]</sup>。RPN 网络需要确定每个滑窗中心对应视野内是否存在目标。由于目标的大小和长宽比不一, 使得需要多种尺度和大小的滑窗。Anchor 给出一个基准窗的大小, 按照不同的倍数和长宽比得到不同大小的窗<sup>[15]</sup>。例如, 文献[15]给出了 3 种面积  $\{128^2, 256^2, 512^2\}$  以及 3 种比例  $\{1:1, 1:2, 2:1\}$  共 9 种尺度的 anchor, 如图 4 所示。

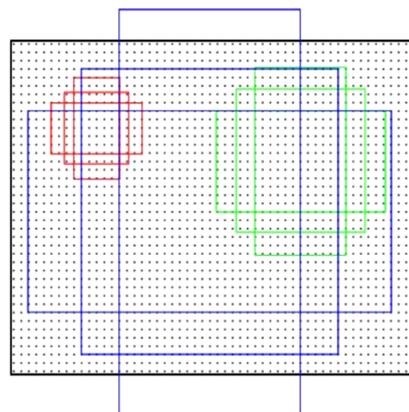


图 4 anchor 示意图

考虑到经典的 RPN 网络搜索区域过多, 提取目标候选区域时计算量大, 导致时间复杂度较高。本文提出一种级联区域建议网络(cascade region proposal network, CRPN)的搜索模式, 第一级 CNN 采用  $32 \times 32$  的卷积核, 步长为 4, 在  $W \times H$  的特征图上进行滑窗, 得到检测窗口。由于搜索区域尺度较小, 通过第一级 CRPN 可以过滤掉 90% 以上的区域窗口, 最后在采用非极大值抑制(non-maximum suppression, NMS)消除高重合率的候选区域, 可大大减少下级网络的计算量以及时间复杂度。

将  $32 \times 32$  网络最后得到的目标候选区域调整为  $64 \times 64$ , 输入第二级 CRPN, 再滤掉 90% 的候选区域窗口, 留下的窗口再经过 NMS 消除高重合率的候选区域。

最后将第二级 CRPN 输出的候选区域调整为  $128 \times 128$ , 经过同样的操作滤掉背景区域。最后通过 NMS 去除高度重合区域, 最终得到 300 个矩形候选区域。CRPN 模型如图 5 所示。

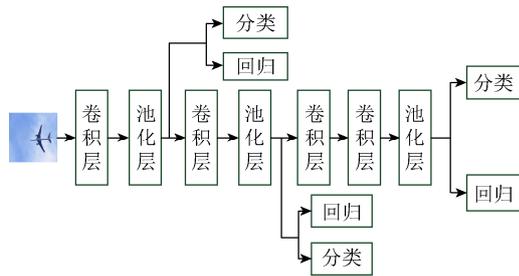


图 5 级联区域搜索网络

经过级联区域搜索网络输出的每一个位置上对应 3 种面积  $\{128^2, 256^2, 512^2\}$  以及 3 种比例  $\{1:1, 1:2, 2:1\}$  共 9 种尺度的 anchor 属于目标前景和背景的概率以及每个窗口的候选区域进行平移缩放的参数。最后的候选区域分类层和窗口回归层可采用尺寸为  $1 \times 1$  的卷积层实现。

## 2.2 目标识别网络数学模型

目标识别网络的第一部分是以 ResNet 为模型提取特征图作为 RPN 输入。经过 ResNet 共享卷积层厚的特征为

$$X_{\text{part1}} = \text{ConvNet}^{(\text{part1})}(x, \theta_{\text{part1}}) \in R^{u \times v \times r} \quad (4)$$

其中, 输入图像为  $x \in R^{n \times m}$ , 通过残差网络的共享卷积层实现特征提取, 提取的特征(输出)为  $X_{\text{part1}} \in R^{u \times v \times r}$ , 即有  $r$  个特征图, 其尺寸为  $u \times v$ ; 待学习的参数(权重和偏置)记为  $\theta_{\text{part1}}$ 。

通过共享卷积层提取特征图之后, 利用区域生

成网络实现候选目标区域的提取。对于区域生成网络的训练, 是将一幅任意大小的图像作为输入, 输出则是候选目标(矩形)区域的集合, 并且对每个(矩形)区域给出是否为目标的得分, 即

$$\begin{aligned} RP_x &= [RP_x^1 \in R^{2 \times S}, RP_x^2 \in R^{4 \times S}] \\ &= \text{RPNNet}(X_{\text{part1}}(x), \theta_{RP}) \end{aligned} \quad (5)$$

其中,  $x$  为输入的图像信息;  $RP_x$  为输出;  $S$  为候选目标区域(矩形)个数;  $RP_x^1$  为判断每个目标候选区域的得分, 即

$$RP_i = P\left(\frac{\text{class}_i}{\text{objects}}\right) P(\text{objects}) T_{\text{true}}^{\text{pred}} \quad (6)$$

其中,  $P(\text{objects})$  为该矩形框中是否有目标;  $P(\text{class}_i/\text{objects})$  为在有目标的情况下目标属于第  $i$  类的概率;  $T_{\text{true}}^{\text{pred}}$  为该预测矩形框与真实目标边框之间的交集与并集的比值。对于每一幅特征提取图, 选取  $RP_x^1$  得分大于 0.7 的  $k_n$  个候选区域(矩形)作为正样本, 小于 0.3 的作为负样本, 其余的全部舍弃。

将区域生成网络的输出结合特征提取图作为输入, 得到目标识别网络的第二部分, 即 Fast-RCNN 网络。在共享卷积层后得到

$$X_{\text{part2}} = \text{ConvNet}^{(\text{part2})}(X_{\text{part2}}, \theta_{\text{part2}}) \quad (7)$$

这一层卷积网络是利用特征图  $X_{\text{part1}}$  来生成更深一层特征  $X_{\text{part2}}$ , 其中待学习的网络参数为  $\theta_{\text{part2}}$ 。感兴趣区域(region of interest, ROI)的池化层输出为

$$X_{\text{part3}} = \text{ROI}^{\text{part3}}(X_{\text{part2}}, RP_x) \quad (8)$$

其中, 输入为区域生成网络得到的候选区域  $RP_x$  和式(8)中的  $X_{\text{part2}}$ , 但是因为目标候选区域的尺寸大小不同, 为了减少裁剪和缩放对信息造成的损失, 引入单层空域塔式池化(spatial pyramid pooling, SPP)来实现对不同尺寸图像的输出。最后经过全连接层输出为

$$\begin{aligned} y &= [RP_x^s, \text{Label}_x^s] \\ &= \text{FC}(X_{\text{part3}}, \theta_{\text{part3}}, C, \text{Refine}(RP_x)) \end{aligned} \quad (9)$$

输出  $y$  包含目标区域位置和目标区域类别 2 部分。其中输入为式(9)中的  $X_{\text{part3}}$  和目标的总类别个数  $C$  以及对每一类所对应的目标区域进行精修的参数  $\text{Refine}(RP_x)$ , 即目标矩形区域中左上标与长宽高等滑动窗的位移, 其中  $c=1, 2, \dots, C$ 。对于每一幅输入图像, 可能存在的目标区域个数为  $s=1, 2, \dots, S$ 。

### 3 目标检测实验

#### 3.1 Cifar10

Cifar10 数据集共有 60 000 张彩色图像, 每张图像大小为  $32 \times 32$ , 分为 10 个类, 每类 6 000 张图。数据集中有 50 000 张图像用于训练, 构成了 5 个训练集, 每一批包含 10 000 张图; 另外 10 000 张用于测试, 单独构成一批测试集。数据集和训练集图像的选取方法是从数据集 10 类图像中的每一类中随机取 1 000 张组成测试集, 剩下的图像就随机排列组成了训练集。训练集中各类图像数量可不同, 但每一类都有 5 000 张图。

本文以文献[10]中的残差网络形式搭建多种不同深度的经典残差网络模型, 而后又以本文提出的 5 层残差模块为基础, 搭建了 22 层、37 层和 57 层的残差网络模型, 最后将本文的不同 ResNet 模型网络的分类效果与经典残差网络模型进行对比。

图 6 是文献[10]与本文设计的网络性能对比图。图中 ResNet 网络表示文献[10]中的网络, Mu-ResNet 表示本文设计的网络。随着网络深度的增加, 训练误差均不断减小, 网络性能变的更优。Mu-ResNet 在 22 层时误差高于 ResNet, 到 37 层时误差低于 ResNet, 当继续加深网络至 57 层式, 误差下降近一个百分点。实验结果表明, 残差网络的收敛速度在迭代 2 k 以前较快, 随后减慢, 在迭代 3 k 左右准确率快速下降, 当到 32 k 时则开始收敛。

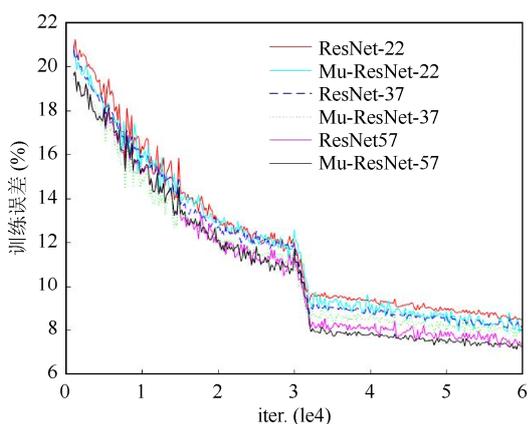


图 6 残差网络性能对比图

图 7 为本文搭建的 3 种不同深度的残差网络训练 cifar10 的准确率, 由图可知深层次的 Mu-ResNet-57 准确率比 Mu-ResNet-22 高出了 3 个百分点。本文设计的网络模型减少了内存、时间复杂度和模型大小, 使得同样深度的网络具有更好的性能, 且网络性能在更深层的优化越明显。

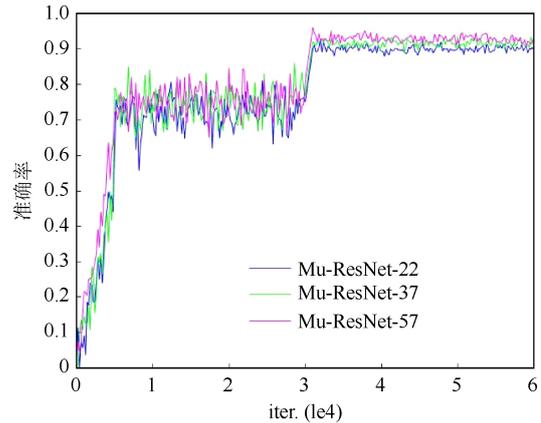


图 7 Mu-ResNet 准确率曲线

通过 cifar10 实验结果表明, 本文提出的 Mu-ResNet 在目标分类上具有较好的效果。相比于经典残差网络不仅减少了内存、时间复杂度, 在准确率上也有了一定的提升。从 22、37、57 层的 Mu-ResNet 准确率在不断提升(图 6), 且在训练的过程中未观察到退化现象, 因此可从增加深度上显著提高准确度。所有评估指标都体现了深度的好处。更深层的网络能够更好地提升 Mu-ResNet 的性能。

#### 3.2 无人机目标识别

UAV 执行目标识别任务主要包含飞机、行人、小汽车、大型公交车、船以及飞行的鸟等 10 类常见目标。数据源主要来源于 ImageNet, 各大开源数据库以及通过 UAV 航拍的图像共 100 000 张。其中将 70 000 张图像作为训练集, 30 000 张作为测试集。

为了研究不同性能参数的特征提取网络对识别效果的影响, 共搭建了 3 种不同深度的网络进行性能测试。22、37、57 层的 ResNet 分别为 Mu-ResNet-22, Mu-ResNet-37 以及 Mu-ResNet-57。

为了进行横向对比, 将以上网络与 VGG16, ZF 和经典残差网络结合经典区域搜索网络进行性能比较。通过在 UAV 目标数据集上进行训练, 测试得到各网络模型在测试集上的识别准确率以及检测每张图片的时间, 结果见表 1。

表 1 不同网络实验结果对比

网络	数据(s)	准确率(%)
ZF	0.035	79.2
VGG16	0.116	86.6
ResNet-57	0.107	89.0
Mu-ResNet-22	0.076	87.8
Mu-ResNet-37	0.089	88.2
Mu-ResNet-57	0.093	90.4

在表 1 中, Mu-ResNet-57 的识别准确率略高于 ResNet-57, 但检测时间更少; Mu-ResNet-22 和 Mu-ResNet-37 的整体性能也优于 VGG16 和 ZF 网

络。采用本文提出的 5 层残差学习模块搭建网络模型能够更有效地提高算法的性能, 在计算时间和识别精度上都提升。

将 ZF, VGG16, ResNet-57 和 Mu-ResNet-57 结合 CRPN, 得到在 UAV 目标数据集上的召回率曲线如图 8 所示。

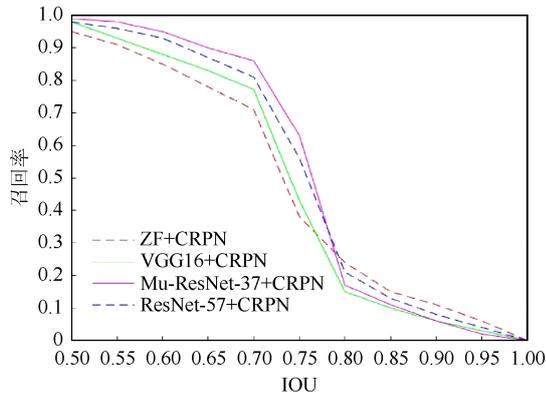


图 8 不同网络模型的召回率曲线

图 8 显示, 当目标候选区域更少时, IOU 更大,

ResNet-57 和 Mu-ResNet-57 的召回率较慢, 当 IOU 在 0.70~0.75 区间内 Mu-ResNet-57 效果最好。在 IOU 过小( $<0.60$ )时, 提取候选区域过多, 严重影响网络的速度, 在 IOU 过大( $>0.80$ )时, 提取候选区域过少, 导致召回率较低。

表 2 为在十万张图像的 UAV 数据集上, VGG16, ResNet-57 和 Mu-ResNet-57 与经典区域搜索网络和级联区域搜索网络相结合, 对每一类目标的检测准确率。RPN 的 proposal 参数采用文献[15]中效果最好的 2 000。本文提出的 CRPN 中 proposal 参数取 1 000。由表 2 可知, Mu-ResNet 对 airplane, human, automobile, truck 的分类效果要优于 VGG16 和 ResNet, 对 ship, bird, horse 的分类效果要差于 ResNet。而采用 CRPN 的性能整体要优于 RPN。

通过以上实验, 综合考虑检测每幅图像的时间以及检测的准确率, 采用 5 层残差学习模块搭建的 57 层 Mu-ResNet 结合 CRPN 在数据集上能达到 90.40% 的识别率且检测每张图像只需要 0.093 s, 故该网络模型较适合用于 UAV 目标识别。

表 2 不同网络模型对数据集各类别检测准确率(%)

目标	网模型络					
	VGG16+RPN	VGG16+CRPN	ResNet-57+RPN	ResNet-57+CRPN	Mu-ResNet-57+RPN	Mu-ResNet-57+CRPN
mAP	86.80	86.60	89.70	89.10	91.90	90.40
airplane	92.13	92.36	94.21	91.895	93.03	95.07
automobile	90.27	90.55	93.20	89.38	90.06	94.08
bird	89.35	90.10	92.05	89.14	88.31	89.30
balloon	94.78	96.48	95.11	93.75	95.17	96.14
cat	86.33	85.31	86.31	87.86	84.93	90.11
dog	93.27	93.17	95.07	93.15	95.02	96.22
human	90.30	93.09	94.30	91.81	93.34	94.87
horse	87.92	88.16	91.82	83.09	86.79	90.60
ship	90.12	91.07	91.74	88.92	90.88	93.17
truck	82.06	85.30	86.26	88.48	90.12	88.40

### 3.3 PASCAL VOC

PASCAL VOC 2007 数据集作为经典的开源数据集, 包含 5 k 训练集样本图像和 5 k 测试机样本图像, 共有 21 个不同对象类别。PASCAL VOC 2012 数据集包含 16 k 训练样本和 10 k 测试样本。本文使用 VOC 2007 的 5 k 和 VOC 2012 的 16 k 作为网络的训练集, 使用 VOC 2007 的 5 k 和 VOC 2012 的 10 k 作为网络的测试集。采用 ZF, VGG16, 文献[10]中提出的 ResNet 和本文中搭建的 Mu-ResNet-57 模型分别与文献[15]中的 RPN 和本文提出的 CRPN 结合, 来评估网络的平均检测精度, 在 VOC 2007+2012 数据集上检测结果见表 3。

将 Mu-ResNet-57 与 CRPN 结合在 VOC2007+2012 数据集上测试, 准确率达到 76.20%。采用较深层网络 ResNet-57 和 Mu-ResNet-57 模型提取特征时,

CRPN 效果较 RPN 有所提升。而采用浅层网络 ZF 和 VGG16 模型时, CRPN 与 RPN 对结果影响基本无区别。

表 3 不同网络测试的平均检测精度

网络模型	数据集	mAP(%)
ZF+RPN	VOC 2007+2012	58.6
ZF+CRPN	VOC 2007+2012	58.8
VGG16+RPN	VOC 2007+2012	71.5
VGG16+CRPN	VOC 2007+2012	72.1
ResNet57+RPN	VOC 2007+2012	74.6
ResNet57+CRPN	VOC 2007+2012	75.3
Mu-ResNet57+RPN	VOC 2007+2012	74.8
Mu-ResNet57+CRPN	VOC 2007+2012	76.2

图 9 是以 New-ResNet-57 结合 CRPN 进行测试的结果图。对于飞机、汽车、行人、鸟类等目标的识别具有良好的效果, 该网络模型对大小不同的物体识别效果均较好, 具有良好的适应性。



图9 目标识别效果图

## 4 结束语

针对 UAV 目标识别的实时性与准确性要求,以卷积神经网络为基础,搭建了多种框架的 ResNet 模型进行训练,并且通过比较各网络的性能优劣,最终找到最适用于 UAV 目标识别的网络。基于 ResNet 的目标识别算法,相较于传统的目标识别算法可以避免人为提取目标特征而带来的误差,在识别精度上有了巨大的提升。对于复杂背景下的目标,ResNet 的捷径反馈机制能够有效地降低网络训练的难度,使得更深层次的网络依然能够有效地训练使用。使用 ResNet 的目标检测应用于 UAV 目标识别中,通过准确率与检测时间的综合性能考虑要更优于传统的深度卷积神经网络。

然而本文所研究的目标数据集依然较少,在以后的研究中需要获取更大、更多的 UAV 目标数据集,并且搭建更深层次、性能更优的网络以及更高效特征提取来提高目标检测的准确率与检测时间。

## 参考文献

- [1] 李伟,张旭东. 基于卷积神经网络的深度图像超分辨率重建方法[J]. 电子测量与仪器学报, 2017, 31(12): 1918-1928.
- [2] 蒋帅. 基于卷积神经网络的图像识别[D]. 长春: 吉林大学, 2017.
- [3] CHANG L, DUARTE M M, SUCAR L E, et al. A Bayesian approach for object classification based on clusters of SIFT local features [J]. *Expert Systems with Applications*, 2012, 39(2): 1679-1686.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//*Advances in Neural Information Processing Systems*. Nevada: MIT Press, 2012: 1097-1105.
- [5] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. *Neural Computation*, 1989, 1(4): 541-551.
- [6] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [EB/OL]. [2018-03-23]. <https://arxiv.org/abs/1312.6229>.
- [7] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//*European Conference on Computer Vision*. Cham: Springer International Publishing, 2014: 818-833.
- [8] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks [EB/OL]. [2018-03-25]. <https://arxiv.org/abs/1505.00387>.
- [9] SAXE A M, MCCLELLAND J L, GANGULI S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks [EB/OL]. [2018-03-23]. <https://arxiv.org/abs/1312.6120>.
- [10] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]//*Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Press, 2015: 1026-1034.
- [11] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feed forward neural networks [EB/OL]. [2018-04-01]. <http://www.doc88.com/p-7738903804120.html>.
- [12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2016: 770-778.
- [13] 张珂,高策,郭丽茹,等. 非受限条件下多级残差网络人脸图像年龄估计[J]. 计算机辅助设计与图形学学报, 2018, 30(2): 346-353.
- [14] 陆永帅,李元祥,刘波,等. 基于深度残差网络的高光谱遥感数据霾监测[J]. 光学学报, 2017, 37(11): 314-324.
- [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 and Machine Intelligence*, 2015, 39(6): 1137-1149.