

doi: 10.7690/bgzdh.2021.06.014

## 基于深度卷积神经网络的舰载机轮廓关键点检测算法

朱兴动<sup>1</sup>, 田少兵<sup>2</sup>, 范加利<sup>3</sup>, 王正<sup>3</sup>

(1. 海军航空大学岸防兵学院, 山东 烟台 264001; 2. 中国人民解放军 91851 部队 21 分队, 辽宁 葫芦岛 125000;  
3. 海军航空大学(青岛校区)舰面航空保障与场站管理系, 山东 青岛 266041)

**摘要:** 为解决航母舰载机已有关键点检测算法的检测性能不高且对遮挡关键点检测效果差的问题, 提出一种基于深度残差网络和特征金字塔网络的舰载机轮廓关键点检测算法。通过提取舰载机关键点深层图像特征及对不同尺度的特征进行融合, 实验分析目标检测算法、特征提取网络和输入图像大小等因素对关键点检测算法性能的影响, 并与其他关键点检测算法进行实验对比。结果表明, 该算法能取得最优效果。

**关键词:** 深度学习; 关键点检测; 深度残差网络; 特征金字塔网络; 舰载机

**中图分类号:** TJ85 **文献标志码:** A

## Detection Algorithm of Contour Key Point of Shipboard Aircraft Based on Deep Convolutional Neural Network

Zhu Xingdong<sup>1</sup>, Tian Shaobing<sup>2</sup>, Fan Jiali<sup>3</sup>, Wang Zheng<sup>3</sup>

(1. College of Coastal Defense, Navy Aviation University, Yantai 264001, China; 2. No. 21 Team, No. 91851 Unit of PLA, Huludao 125000, China; 3. Department of Ship Surface Aviation Support & Station Management, Qingdao Campus, Navy Aviation University, Qingdao 266041, China)

**Abstract:** In order to solve the problem that the existing key point detection algorithm of carrier aircraft has low detection performance and poor detection effect on key point occlusion, a key point detection algorithm of carrier aircraft contour based on deep residual network and feature pyramid network is proposed. By extracting the deep image features of the key point of the carrier aircraft and fusing the features of different scales, the experiment analyzes the influence of the object detection algorithm, the feature extraction network and the size of the input image on the performance of the key point detection algorithm, and compares it with other key point detection algorithms perform by experimental comparison. The results show that the algorithm can achieve the optimal results.

**Keywords:** deep learning; key point detection; deep residual network; feature pyramid network; shipboard aircraft

### 0 引言

现代航母是以舰载机为主要作战武器的大型水面舰艇, 舰载机在航母舰面作业过程中的安全保障问题至关重要。当前, 计算机视觉技术发展迅速, 已广泛应用于国防军事领域。在智能监控领域, 对舰载机关键点进行检测以形成舰载机外轮廓, 有助于对舰载机进行辅助监控和碰撞预警, 以保障舰载机装备安全。

传统的关键点检测算法大多利用特征算子提取图像关键点特征, 再利用识别算法进行识别检测。文献[1]利用形状索引的高斯函数差分提取脸部关键点特征, 再使用高斯回归树的方法检测关键点。文献[2]提出了一种随机森林与外形搜索的关键点定位算法, 极大地提高了运行速度和准确率, 但两者都过于依赖特征的选取, 模型泛化能力不足。当

前, 深度卷积神经网络在计算机视觉领域已取得重大突破, 关键点检测已成为视频监控、人机交互和行为分析的主要研究方向。深度卷积神经网络关键点检测算法主要分为自底向上和自顶向下的检测算法, 前者是先检测出图像中所有目标的关键点, 再分配到每个目标上面, 通常检测速度快, 但准确率不高。后者是先检测出单个目标, 再从单个目标框里面检测关键点, 准确率高, 但检测速度较慢<sup>[3-4]</sup>。Wei S E 等<sup>[5]</sup>提出的卷积姿态机(convolutional pose machine, CPM)模型, 使用顺序化的卷积结构来表达空间信息和纹理信息, 对同一个卷积架构使用多个尺度处理输入的特征, 检测准确度得到了进一步提高, 但在有遮挡、复杂的背景下检测效果较差。A.NEWLL 等<sup>[6]</sup>提出的堆叠沙漏网络则是设计一种与反卷积结合的沙漏网络, 充分利用多尺度特征, 堆叠多次来对目标的关键点位置进行更为准确

收稿日期: 2021-02-25; 修回日期: 2021-03-28

基金项目: 军队科研基金(舰载机自动牵引关键技术研究)

作者简介: 朱兴动(1967—), 男, 海南人, 博士, 教授, 从事武器装备信息化保障、IETM 交互式电子技术手册研究。

E-mail: zxdong223@163.com。

的预测，但检测耗时较长。

深度残差网络 ResNet 因其利用拟合残差函数的思想<sup>[7]</sup>，保证了网络可进行更深层次的训练，从而提高模型检测的准确率。特征金字塔网络 (feature pyramid networks, FPN) 具有强大的特征融合能力，可提高模型对局部、全局和环境等因素的抗干扰能力<sup>[8]</sup>。笔者针对已有的基于深度学习的关键点检测算法对部分关键点遮挡及检测准确率不高的问题，受自顶向下关键点检测算法的启发，提出了基于深度残差网络和特征金字塔网络的舰载机关键点检测算法。

### 1 算法概述

基于舰载机关键点检测的实时性和准确性，采用自顶向下的关键点检测算法对舰载机关键点进行检测。首先采用 YOLO v3 算法<sup>[9-11]</sup>对图像中的所有舰载机目标进行检测，得到多个舰载机目标框，然后将每个目标框中的舰载机图像输入到关键点检测网络中进行关键点检测定位。

通过对舰载机图像数据集标注框进行聚类分析，得到适合本文中舰载机目标检测的 anchor 来代替原 YOLO v3 算法的 anchor，提高了训练效率和检测精度。在检测到的舰载机目标框内进行关键点检测任务，为了进行算法的公平比较，本文中舰载机关键点检测的实验部分采用相同的目标检测结果。

### 2 舰载机关键点检测算法

#### 2.1 深度残差网络原理

卷积神经网络能够提取图像中各层次的特征信息，随着网络的层数增多，提取细节与抽象能力越丰富，语义信息越完整；但实际上网络层次越多，误差反向传播时易出现梯度爆炸或消失，造成模型准确率下降。He K M 等<sup>[7]</sup>提出的深度残差网络 ResNet，利用拟合残差函数的思想，在保证网络结构足够深的同时，实现了训练过程的优化。

如图 1 所示，残差网络的残差模块在输入和输出之间建立了一个直接连接，这样新增的层仅需在原输入层的基础上学习新的特征。定义最后输入为  $H(x)=F(x)+x$ ，可转化为学习一个残差函数  $F(x)=H(x)-x$ 。残差结构可表示为

$$x_{l+1} = x_l + F(x_l, W_l) \tag{1}$$

式中： $x_{l+1}$  为  $l+1$  层输出； $x_l$  为  $l+1$  层输入； $F(x_l, W_l)$  为  $l$  层残差模块学习到的图像特征。对于一个  $L$  层残差网络，其与  $l$  层的关系可表示为：

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i) \tag{2}$$

可以看出  $L$  层为一个比它浅的  $l$  层和它们之间的残差之和，根据反向传播中使用的导数链式法则，损失函数  $\varepsilon$  关于  $x_l$  的梯度可表示为

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \left( 1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i) \right) = \frac{\partial \varepsilon}{\partial x_L} + \frac{\partial \varepsilon}{\partial x_L} \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i) \tag{3}$$

由于在整个训练过程中， $\frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i)$  不可能为 -1，故残差网络中不会出现梯度消失的问题。

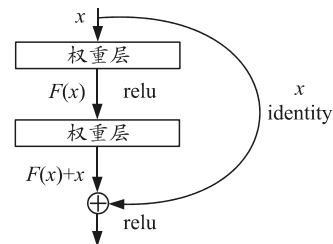


图 1 残差模块

如图 2 所示，ResNet 网络是由“bottleneck”类似结构的基本残差单元组成，以处理 256 维特征的“bottleneck”残差单元结构示例。“bottleneck”残差单元结构将输入的 256 维特征经过  $1 \times 1 \times 64$  卷积降到 6 维，最后再通过  $1 \times 1 \times 256$  卷积恢复到 256 维，降低了参数数量，从而加快了训练速度。

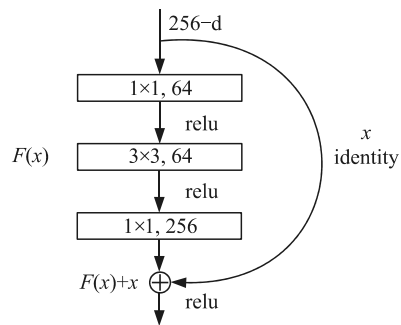


图 2 “bottleneck”残差单元结构

#### 2.2 特征金字塔网络原理

由于深层的语义特征比较丰富，常见的检测方法都是采用深层次的特征进行预测，预测结果比较粗略。反观浅层特征却具有图像信息准确的优点，但语义信息较少。受特征金字塔网络的启发，融合不同尺度的特征层，使网络同时保持较高的浅层图像信息和深层的语义信息。

图 3 为特征金字塔网络多尺度融合思想<sup>[12]</sup>，左侧为自下而上的特征提取网络，右侧为自顶向下的

预测网络。假设左侧最上端得到的特征图大小为  $A \times A$ ，传入右侧顶端的特征图大小也是  $A \times A$ 。之后经过 2 倍上采样，第 2 层与第 3 层的特征图变为  $2A \times 2A$  与  $4A \times 4A$ ，将这 2 层的特征与左侧得到的浅层特征进行融合，提升关键点预测位置精度。

左侧自下而上网络是图像空间分辨率依次降低但特征语义信息逐渐增强的过程，经过右侧自顶向下网络的特征融合过程，使网络同时保持了较高的空间分辨率和较高的语义信息，从而使网络由粗到细地回归关键点。

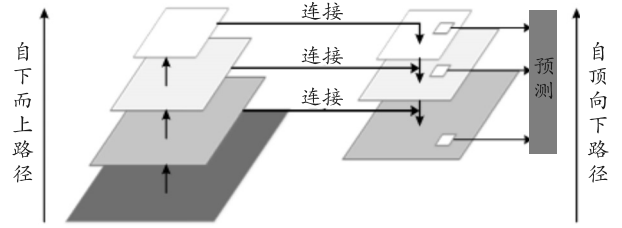


图 3 特征金字塔网络

### 2.3 舰载机关键点检测算法

舰载机关键点检测算法采用 ResNet 网络提取舰载机图像各尺度的特征，网络结构组成如图 4。

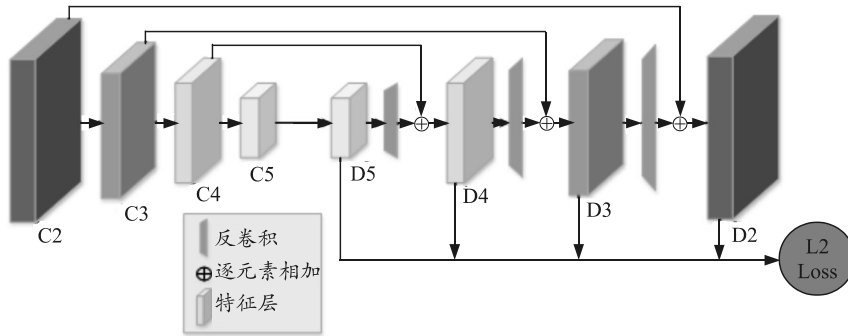


图 4 ResNet 网络结构

将 ResNet 网络的第 2 到第 5 个卷积层的最后残差块表示为  $C_2$ 、 $C_3$ 、 $C_4$ 、 $C_5$ 。 $C_2$  和  $C_3$  这样的浅层特征在舰载机关键点预测时具有较高的空间分辨率，但语义信息较低。另一方面，由于跨卷积和池化， $C_4$  和  $C_5$  这样较深的特征层具有更多的语义信息，但空间分辨率较低。借鉴特征金字塔网络的思想， $D_5$  特征层经过带有批量标准化和激活函数的反卷积网络后，转化到与  $C_4$  特征层相同的尺度大小并与  $C_4$  特征层进行逐元素相加形成  $D_4$  特征融合层，依次对  $D_4$ 、 $D_3$  特征层进行同样操作以形成多尺度特征融合网络。

分别在  $D_5$ 、 $D_4$ 、 $D_3$ 、 $D_2$  特征层后添加  $1 \times 1$  卷积层以生成预测舰载机关键点的热图，通过计算每层的预测热图与标注热图之间的  $L_2$  损失来进行误差反向传播。将舰载机关键点坐标回归转化为关键点热图回归时，需将每个关键点生成 2 维关键点热图  $C_{k(p)}^*$  作为标签进行训练，即每张热图中像素的数值代表该架舰载机的第  $k$  个关键点存在于像素点  $p \in \mathbf{R}^2$  的概率。假设  $X_k$  为已标注的舰载机关键点坐标，则  $C_{k(p)}^*$  中像素点  $p \in \mathbf{R}^2$  的数值可表示为：

$$C_{k(p)}^* = \exp\left(-\left(\|p - X_k\|_2^2 / \sigma^2\right)\right) \quad (4)$$

则损失函数可表示为：

$$L = \sum_{i=1}^N \sum_{j=1}^K \|C_{j(p)} - C_{j(p)}^*\|_2^2 \quad (5)$$

式中： $C_{j(p)}$  为预测的舰载机第  $k$  个关键点的热图； $K$  为舰载机关键点个数； $N$  为预测特征层的个数。

## 3 实验及分析

### 3.1 实验数据集、平台和性能评价指标

利用实验室条件下自采集包含不同角度和不同目标尺度大小的 2 500 多张航母舰面舰载机图像数据，对舰载机目标位置和关键点位置及其可见性进行标注得到舰载机关键点数据集。分别选取 100 个数据样本作为训练验证集和测试集。

在 Ubuntu16.04 系统上搭建实验环境，CPU 为 Inter(R) Xeon Silver 4110，GPU 为 NVIDIA GeForce RTX 2080Ti，深度学习框架为 Tensorflow。

选取目标关键点相似度 (object keypoint similarity, OKS) 来衡量预测舰载机关键点与真实关键点之间的相似度。其计算公式定义如下：

$$OKS_p = \frac{\sum_i \exp\{-d_{pi}^2 / 2S_p^2\sigma_i^2\} \delta(v_{pi} = 1)}{\sum_i \delta(v_{pi} = 1)} \quad (6)$$

式中： $p$  为舰载机的 ID； $i$  为舰载机关键点的 ID； $d_{pi}$  为第  $p$  架舰载机预测的第  $i$  个关键点与真实标注的关键点之间的欧式距离； $S_p^2$  为第  $p$  架舰载机目标

边界框所占的像素面积； $\sigma_i$ 为第  $i$  个关键点归一化因子； $v_{pi}$ 为第  $p$  架舰载机的第  $i$  个关键点是否可见； $\delta$ 为选择函数。

平均准确率 (average precision, AP) 定义为给定阈值  $s$ , 预测的舰载机关键点结果在整个测试集上的平均准确率, 可由测试集所有图片的 OKS 指标计算得到:

$$AP@s = \sum_p \delta(OKS_p > s) / \sum_p 1. \quad (7)$$

总体评价指标定义为平均准确率的均值 (mean average precision, mAP), 定义为

$$mAP = \text{mean}\{AP@(0.50:0.05:0.95)\}. \quad (8)$$

指定的阈值 0.5 到 0.95 按照 0.05 的步长依次计算 AP 后, 再取平均值。

### 3.2 训练及测试

数据扩充对于模型学习尺度不变性和旋转不变性至关重要。在训练模型的过程中应用随机水平、垂直翻转, 随机旋转范围为  $\pm 45^\circ$ , 随机缩放图像比例为  $\pm 30\%$ 。通过以上数据增强办法, 增加了数据集中样本的复杂度, 使模型具有更强的泛化能力, 也可避免模型的过拟合。

训练进行 100 轮迭代, 初始学习率为 0.000 5, 每训练 20 轮后学习率降低为原来的 0.5 倍。训练批量大小设置为 8, 权重衰减系数为 0.000 5。训练过程中学习率和损失函数曲线如图 5 所示, 曲线在进行 30 000 次批量训练后趋于收敛。

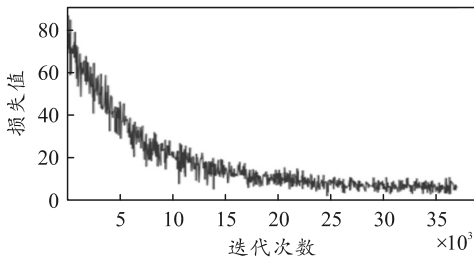


图 5 损失函数曲线

为使预测的舰载机关键点误差最小, 在预测的热图上应用高斯滤波器。对于每一个舰载机目标, 同时预测了其翻转图像的关键点, 并对两者的热图求平均以获得舰载机关键点定位, 从关键点热图的最大高斯响应到第二大高斯响应偏移 1/4, 用来获取最终的舰载机关键点定位。

### 3.3 影响关键点检测性能的因素分析

#### 3.3.1 目标检测算法对性能的影响

目标检测的准确率对提高关键点检测的准确率

起着至关重要的作用。将 Faster R-CNN<sup>[13]</sup>、SSD<sup>[14]</sup>和 YOLO v3 3 种目标检测算法分别作为算法的目标检测算法, 实验对比目标检测算法性能对后续关键点检测性能的影响, 结果如表 1 所示。

表 1 目标检测算法影响对比

目标检测算法	准确率(目标)/%	mAP(关键点)/%	速率/(帧/s)
SSD	70.3	71.0	10.6
YOLO v3	73.6	71.9	11.1
Faster R-CNN	78.4	73.5	8.7

由上表可知, 随着目标检测算法性能的提高, 舰载机关键点检测准确率也相应提高, Faster R-CNN 检测舰载机目标的准确率最高, 其相对应的舰载机关键点检测准确率也最高。原因在于目标检测准确率越高, 目标框内的舰载机图像目标信息越完整, 为后续关键点检测提供的图像信息也越全面。

#### 3.3.2 特征提取网络与输入图片大小对性能的影响

为探究特征提取主干网络和输入图片大小对关键点检测模型性能的影响, 实验设置如表 2 所示。

表 2 特征提取主干网和输入图片大小影响对比

方法	特征提取主干网	输入图片大小	mAP/%
a	ResNet-50	192×192	71.2
b	ResNet-50	256×256	71.9
c	ResNet-50	384×384	72.6
d	ResNet-101	256×256	73.1

由 a、b、c 方法的实验结果可知, 当输入关键点检测网络的图像越大, 算法检测性能越高。这是因为随着图像大小的增加, 更多的舰载机关键点的位置细节被送入到网络中, 从而大大提升了性能。比较 b、d 方法的实验结果可知, 当 ResNet 网络层数越深时, 由于学习到了图像更深层次的信息, 模型检测性能提升了约 1.2%。

### 3.4 不同算法比较

将本文中算法与文献[5-6]的关键点检测算法在自建舰载机测试图像集上进行实验比较。实验结果如表 3 所示, 从关键点检测准确率和速度方面评价算法的性能。

表 3 不同算法性能比较

算法	AP@0.5/%	AP@0.75/%	mAP/%	速率/(帧/s)
参考文献[5]	83.6	67.0	64.9	13.4
参考文献[6]	87.7	72.3	66.3	9.3
本文中	89.5	73.7	71.9	11.1

实验结果表明, 本文中提出的舰载机关键点检测方法准确性最优, mAP 达到了 71.9%, 比文献[5]

提高了 7%，比文献[6]提高了约 5.6%，说明笔者提出的算法在关键点检测方面的优越性，检测速度相比于文献[5]有所降低，但能满足实际应用需求。结

合图 6 的测试结果，表明笔者提出的改进算法对遮挡关键点具有较好的检测效果，算法实时性和鲁棒性较高。



图 6 测试结果

## 4 结论

为了检测航母舰面舰载机关键点以形成舰载机外轮廓从而应用于舰载机碰撞预警及辅助监控，笔者提出了基于深度残差网络和特征金字塔网络的舰载机轮廓关键点检测算法，通过提取舰载机关键点深层特征以及多尺度特征融合，提升了算法的性能，对一些遮挡的舰载机关键点也具有较好的检测效果。分析了影响舰载机关键点检测算法性能的因素，通过与其他关键点检测算法进行实验对比，该算法的性能最优，证明了算法的有效性和先进性。

## 参考文献：

- [1] 赵康. 基于形状索引的 DoG 特征结合 GPRT 的人脸关键点检测算法[J]. 计算机测量与控制, 2019, 27(5): 203-206.
- [2] 王丽婷, 丁晓青, 方驰. 基于随机森林的人脸关键点精确定位方法[J]. 清华大学学报(自然科学版), 2009, 49(4): 543-546.
- [3] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. Deeppercut: A deeper, stronger, and faster multiperson pose estimation model[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 34-50.
- [4] YANG W, LI S, OUYANG W, et al. Learning feature pyramids for human pose estimation[C]//The IEEE International Conference on Computer Vision (ICCV), Oct, 2017, Venice, Italy, New York: IEEE, 1290-1299.
- [5] WEI S E, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 4724-4732.
- [6] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 483-499.
- [7] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas NV USA. New York: IEEE, 2016: 770-778.
- [8] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 2117-2125.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once unified real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLO 9000 better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [11] REDMON J, FARHADI A. YOLO v3 an incremental improvement[EB/OL]. (2018-04-08). <https://arxiv.org/abs/1804.02767>.
- [12] 李锸, 姚麟倩, 关欣. 基于级联卷积神经网络的服饰关键点定位算法[J]. 天津大学学报(自然科学与工程技术版), 2020, 53(3): 229-236.
- [13] REN S Q, HE K M, 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.
- [14] LIU W, ANGELOV D, ERHAN D, et al. SSD: single shot multibox detection[C]//Proc of European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.