文章编号：2095-4980(2021)05-0922-07

基于一种卷积神经网络的目标快速跟踪技术

陈 斌[a,b]，王 磊[*a]

(中国工程物理研究院 a.应用电子学研究所；b.研究生院，四川 绵阳 621999)

摘 要：特定动态目标的快速检测及跟踪，是计算机视觉领域重要的课题。改变特征图在 YOLOv3 卷积神经网络中的选取位置，通过收集相关网络数据(类似模式分析、统计建模和计算学习视觉对象类别数据集合，即 PASCAL Visual Object Classes 数据集)构建自定义数据集合进行训练，使用面积的交并比完成辅助类别的联合，构建了能够实时检测特定目标在相关可视对象类检测数据集合上 mAP@75 达到 47.41 的检测器。联合卡尔曼滤波和匈牙利算法，通过将面积信息加入到匈牙利算法的代价矩阵中，改善了使用原方法产生大量 ID 切换(ID switch)的问题。该方法满足快速识别与跟踪的要求，在使用一张 NVIDIA GeForce GTX 1060 6GB GPU 条件下，平均速度能达到 0.109 7 s/帧。

关键词：卷积神经网络；目标检测；目标跟踪；匈牙利算法；卡尔曼滤波

中图分类号：TN914.42 **文献标志码：**A **doi：**10.11805/TKYDA2019562

Track and detection for specified moving object based on Convolutional Neural Networks

CHEN Bin[a,b]，WANG Lei[*a]

(a.Institute of Applied Electronic；b.Graduate University，China Academy of Engineering Physics，Mianyang Sichuan 621999，China)

Abstract：The detection and track of specified moving small object is an important subject in Computer Vision. By changing the position of the feature maps for fusion in the YOLOv3, building the custom database including three classes, and completing the combination of classes by using Intersection Over Union(IOU), a detector is created, which is able to detect the specified moving small object and makes mAP@75 reach 47.41 in the test customer's data set. Combining Kalman Filter and Hungarian method, and putting the scale information of predicted bounding box and ground bounding box, the detector can track the object and reduce the ID switch caused by camera's fast movement. The whole system's speed reaches up to 0.109 7 s/frame using one NVIDIA GeForce GTX 1060 6GB GPU.

Keywords：Convolutional Neural Networks；object detection；object track；Hungarian method；Kalman Filter

目标类别检测任务传统方法主要使用尺度不变特征变换(Scale-Invariant Feature Transform，SIFT)、方向梯度直方图(Histogram of Oriented Gradient，HOG)等为选取的特征描述算子来提取特征，然后使用支撑向量机(Support Vector Machine，SVM)等分类器来完成。其中较为著名的方法由 Dalal 等提出，通过使用方向梯度直方图描述子和支撑向量机出色地完成了行人检测任务[1]，这种方法与其他传统方法类似，主要依靠先验知识，依据具体任务，人为选取合适的特征描述子来描述待测物体的特征，再使用支持向量机等分类器对特征分类检测来完成任务，所以传统方法准确率和速度常常不能满足实时应用的需求。卷积神经网络(CNNs)的前身为反向传播(Back Propagation，BP)神经网络，BP 神经网络可以逼近非线性函数，能够解决任意形式的分类问题，在通信模式干扰识别等多方面应用[2]，由于其不具有卷积神经网络的权值共享、感受野(receptive field)约束等特点，不适合处理图片类型的数据。

在 2012 年，Krizhevsky 等通过展示卷积神经网络在图像网络大规模视觉识别挑战(ImageNet Large Scale Visual

收稿日期：2019-12-24；修回日期：2020-04-02

*通信作者：王 磊 email:wanglei839@sina.com

Recognition Challenge，ILSVRC)上的高准确率，使得人们再次对于卷积神经网络产生兴趣。卷积神经网络系列的网络结构，相比于之前的传统的图片分类算法，具有较少的前处理过程，并且不需要人为的设计特征(比如SIFT)和先验知识，使用卷积神经网络来对目标进行目标的特征提取，渐渐成为主流方法。其中较为著名的有R–CNN，R–CNN 是一种候选区域(region proposal)和卷积神经网络联合的方法[3]，其通过产生与类别无关的候选区域，在候选区域上使用卷积神经网络对区域进行定长的特征提取，最后使用指定类别的线性支持向量机(Support Vector Machine，SVM)进行分类来完成图片上的目标检测。R–CNN 相较于传统的基于 SIFT 或者 HOG 方法的分类器，其在准确度方面有了较大提高，由于其在候选区域中要卷积神经网络，使其检测的时间较长，无法用于快速分析视频数据和实时检测的系统，之后出现了各种改进的方法，比如：Fast R–CNN,Faster R–CNN，虽然速度相较于 R–CNN 都有较大的提高，但是都无法为实时或者准实时系统使用。Redmon J,Farhadi A 等在 2018 年提出的 YOLOv3(You Only Look Once v3)方法[4]，因其速度快，精确度高，受到广泛关注，然而大多是针对图片或静态物体的检测，对视频中动态目标的检测和跟踪方面的研究较少。

本文主要研究飞机这一特定动态目标的快速检测和跟踪问题，使用空中常见的鸟类作为干扰类别，固定翼飞机发动机作为亚类别。研究中采用针对小目标改进的 YOLOv3 卷积神经网络作为网络模型，收集相关网络数据(类似模式分析、统计建模和计算学习视觉对象类别数据集合，即 The PASCAL Visual Object Classes 数据集)构建自定义数据集合进行训练，构造了对于固定翼飞机具有较高准确率的快速检测器，再结合卡尔曼滤波算法和改进的匈牙利算法，完成动态特定目标的快速跟踪过程。在使用 NVIDIA GeForce GTX 1060 6GB GPU 的硬件条件下平均能够达到约 0.109 7 s/帧的速度，取得了较好的检测与跟踪效果。

1 研究方法

实验总体流程如图 1 所示，主要由 3 部分组成：

1) 使用改进的 YOLOv3 网络来计算检测当前帧的特定目标的边框位置和类别分数(class scores)。

2) 根据当前帧的检测结果使用卡尔曼滤波器来预测下一次检测的结果。卡尔曼滤波器轨迹的状态由 7 个值来描述，分别为：边框(bounding box)的中心坐标(x,y)、边框的面积(scale)、边框的宽高比(aspect ratio)、中心坐标对时间的导数、面积对时间的导数。卡尔曼滤波器将其作为动态系统进行建模。

3) 使用匈牙利算法完成新检测到的对象和第 2 步预测的结果对比和分配。如果满足分配条件，则将检测到的对象分配到预测结果所属的存在的轨迹中，将其认定为同一个物体；如果不满足分配条件，检测到的对象使用新的轨迹和 ID。

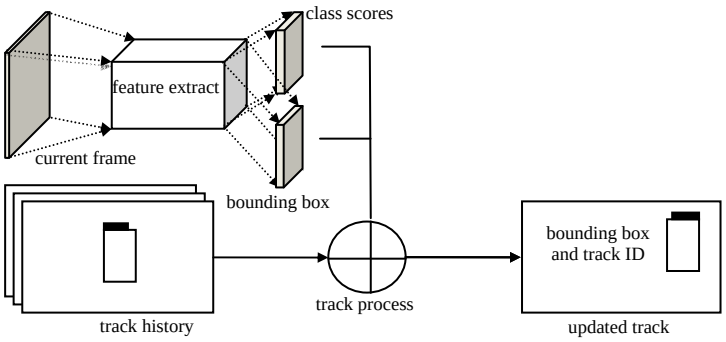

Fig.1 Experiment process
图 1 实验过程

1.1 构建基于卷积神经网络的检测器

VOC 挑战是可视物体类别识别与检测的基准，为视觉和机器学习社区提供标准的图像和注释数据集[5]，本文主要检测飞机、鸟类作为干扰类别，发动机作为亚类别。收集相关网络数据构建类似 VOC 自定义数据集合进行训练，在正样本中，飞机的实例个数为 1 162，鸟类的实例个数为 1 562，在使用相同数量的此数据集中不包含这 2 类的样本作为负样本，再对于包含飞机这一类别的图片中的涡轮发动机(实例个数为 644)进行标注，并创建新的发动机类别，完成数据集的创建。测试集中的正样本数量设置为飞机实例 123，鸟类实例 258，飞机引擎实例数量 50。视频测试图片集合由视频中每隔 10 帧抽取的

	type	filters	size	output
	convolutional	32	3×3	608×608
	convolutional	64	3×3/2	304×304
	convolutional	32	1×1	
1x	convolutional	64	3×3	
	residual			304×304
	convolutional	128	3×3/2	152×152
	convolutional	64	1×1	
2x	convolutional	128	3×3	
	residual			152×152
	convolutional	256	3×3/2	76×76
	convolutional	128	1×1	
8x	convolutional	256	3×3	
	residual			76×76
	convolutional	512	3×3/2	38×38
	convolutional	256	1×1	
8x	convolutional	512	3×3	
	residual			38×38
	convolutional	1024	3×3/2	19×19
	convolutional	512	1×1	
4x	convolutional	1024	3×3	
	residual			19×19
	avgpool		global	
	connected		1 000	
	softmax			

Fig.2 Darknet–53–608
图 2 Darknet–53–608

图片构成，固定翼飞机实例个数共为 177 个。

为了在追踪过程中尽早检测到目标，实验中使用针对小目标改进的 Darknet-53 网络结构来进行特征的提取。根据网络结构，特征提取在 3 个尺度上进行，其分别对输入图像的尺寸进行了 32,16,4 的下采样，所以输入图像的尺寸应该能被 32 整除，分辨力越高的图像包含越多的信息，但是会增长检测时间，考虑到主要的实验设备为 NVIDIA GeForce GTX 1060 6 GB GPU，本文使用 608×608 的输入尺寸。改进的 Darknet-53 的网络结构如图 2 所示。

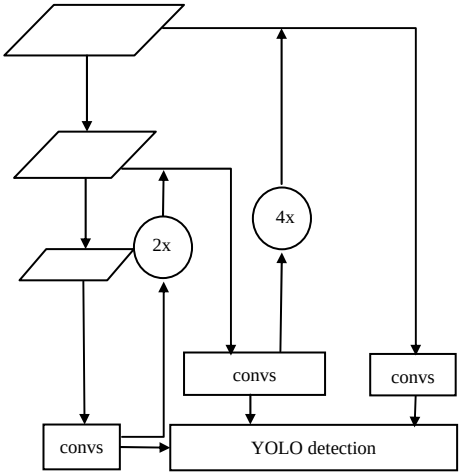

Fig.3 Fusion of feature map
图 3 特征图的融合过程

其主要由大量残差网络构成。大量的卷积过程，使网络获得的特征图的语义不断升高，然而分辨力却会不断下降。在网络结构中，第 3 残差层输出的特征图相较于输入图像的尺寸进行了步长为 4 的下采样，在输入图像的尺寸为 608×608 的情况下，其获得特征图的尺寸为 152×152，然而第 11 残差层输出的特征图相较于输入图像的尺寸则进行了步长为 8 的下采样，在输入图像的尺寸为 608×608 的情况下，其获得的特征图的尺寸为 76×76，相较于第 11 残差层的输出，第 3 残差层输出的特征图经历了更少卷积操作，图像的信息丢失较少，所以它具有更高的空间分辨力，更有利于小目标的检测。原 YOLOv3 算法使用第 11,19,23 残差层的特征图进行特征图的融合，为了在检测过程中尽早地检测到小目标，本文使用第 3,19,23 残差层的特征图进行融合。较高层的特征图拥有更高的分辨力，融合后的特征图兼顾语义和分辨力，其过程如图 3 所示，类似于特征金字塔网络(Feature Pyramid Network，FPN)[6]，图 3 中左边的 3 个特征图由上至下分别来自第 3、第 19 和第 23 残差层，网络检测在 3 个尺度上进行，在第 1 个尺度上，网络使用第 23 残差层输出的特征图，通过在其后加上一层 1×1 的卷积层完成检测；在第 2 个尺度上，通过对上一尺度的特征图进行 2x 的上采样，使其与第 19 残差层上输出的特征图的尺寸一致，之后将上采样过的特征图与第 19 残差层输出的特征图进行融合构成总的特征图，在获得的总的特征图之后加上一层 1×1 的卷积层完成检测；在第 3 个尺度上，通过对上一尺度的特征图进行 4x 的上采样使其与第 3 残差层上输出的特征图的尺寸一致，之后将上采样过的特征图与第 3 残差层输出的特征图进行融合构成总的特征图，在获得的总的特征图之后加上一层 1×1 的卷积层完成检测。

边框坐标预测使用锚框(anchor boxes)[7]方法，原理见图 4 和式(1)[8]，其中 b_x,b_y,b_w,b_h 表示边框最终中心位置和宽高预测值，t_x,t_y,t_ht_w 是在卷积神经网络检测出来的 4 个用来描述边框中心位置和宽高的参数，p_w,p_h 是由锚框给出的边框的初始值，c_x,c_y 表示由左上角(图像坐标系的坐标原点)到边框中心点所在的本地坐标原点之间的距离，$\sigma(t)$是逻辑的激活函数，它使得网络预测值落在 0~1 之间。为了使训练检测过程更快，锚框的初值通过在训练集上使用 k 均值丛聚获得，重新获得的初始值见表 1。

表 1 训练集的锚框

Table1 Anchor box's value of training dataset

anchor frame No	initial value of anchor frame
1	(34,36)
2	(76,63)
3	(89,155)
4	(168,114)
5	(181,327)
6	(331,184)
7	(319,421)
8	(535,275)

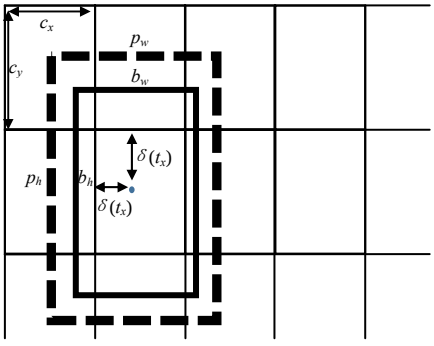

Fig.4 Bounding box's prediction
图 4 边框的预测图

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_h = p_h e^{t_h} \\ b_w = p_w e^{t_w} \end{cases} \tag{1}$$

训练过程中使用的图像通过先将原图保持比例进行缩放使其较长的边达到 608，在未达到 608 的较短边的维

度上在缩放后的图像两边分别填充(128,128,128)的数据使其也达到 608 的预处理方法来获得，该方法能有效防止由于图像缩放引起的振铃现象。之后使用 Darknet 神经网络框架进行训练。

在检测过程中加入新的规则：如果检测到物体的概率最大的类别是鸟类，那么将其边框与其余的类别为引擎的边框分别进行交并比(IOU)计算，如果有一个由计算获得的 i_{ou} 的值与直接使用引擎类别边框面积与鸟类别边框面积的比值相等，即如果 $i_{ou}=S_{引擎}/S_{鸟类}$，则更改该类别为飞机，不同边框的面积不同，获得的比值亦不同，但只要二者数值相等即可进行类别的更改。2 个边框的 IOU 使用式(2)进行计算。

$$i_{\text{out}}=\frac{S_{\text{overlap}}}{S_{\text{union}}} \tag{2}$$

式中 S_{overlap}，S_{union} 分别表示 2 个边框重叠部分的面积，以及 2 个边框联合所占的面积。

1.2 使用卡尔曼滤波器进行状态的预测

卡尔曼滤波[9]适用于含有不确定信息的任何动态系统中，其能够通过当下的状态，预测下一步的状态。卡尔曼滤波器的基本原理如图 5 所示。其中 $\hat{x}_{k-1}$ 为 k–1 帧的状态量，$\hat{x}_k$ 为 k 帧的状态量，$\boldsymbol{P}_{k-1}$ 为 k–1 帧的协方差矩阵，$\boldsymbol{F}_k$ 为预测矩阵，$\boldsymbol{Q}_k$ 为干扰矩阵，$\boldsymbol{P}_k$ 为 k 帧的协方差矩阵，R_k 为传感器的不确定性，$\boldsymbol{H}_k$ 为测量矩阵，$\bar{z}_k$ 为传感器测量值，K 为卡尔曼增益。根据图 5，使用卡尔曼滤波进行状态预测主要分为 2 步：

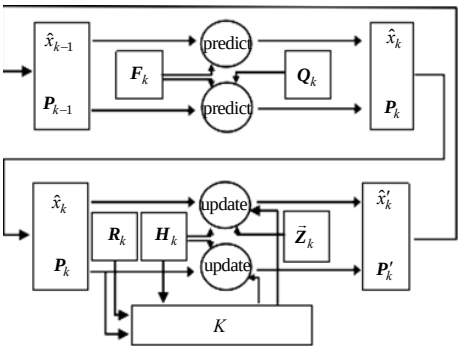

Fig.5 Process of Kalman filter
图 5 卡尔曼滤波过程

1) 通过当前状态进行下一帧的预测(predict)；

2) 进行状态的更新(update)。预测过程使用式(3)：

$$\begin{cases}\hat{x}_k=\boldsymbol{F}_k\hat{x}_{k-1}\\ \boldsymbol{P}_k=\boldsymbol{F}_k\boldsymbol{P}_{k-1}\boldsymbol{F}_k^{\mathrm{T}}+\boldsymbol{Q}_k\end{cases} \tag{3}$$

在实验中，系统的状态量为 $\hat{\boldsymbol{x}}=[u,v,s,r,u',v',s']^{\mathrm{T}}$，其中 u,v 为检测到的目标的中心位置坐标，s,r 分别为面积和边框(bounding box)的宽高比，u',v',s' 分别为其在匀速模型中变量变换的速度。通过式(3)可以从 k–1 帧的状态量得到 k 帧的状态量完成预测过程。系统使用匀速模型进行预测，预测矩阵使用 $\boldsymbol{F}$ 矩阵的值；在系统初始的时候，系统量之间的关系并不是很清楚，所以初始的协方差矩阵的值使用 $\boldsymbol{P}$ 矩阵的值；干扰矩阵为不可预知的未知干扰，使用 $\boldsymbol{Q}$ 矩阵的值。

$$\boldsymbol{F}=\begin{pmatrix}1&0&0&0&1&0&0\\0&1&0&0&0&1&0\\0&0&1&0&0&0&1\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{pmatrix};\quad \boldsymbol{P}_0=\begin{pmatrix}10^1&0&0&0&1&0&0\\0&10^1&0&0&0&1&0\\0&0&10^1&0&0&0&1\\0&0&0&10^1&0&0&0\\0&0&0&0&10^4&0&0\\0&0&0&0&0&10^4&0\\0&0&0&0&0&0&10^4\end{pmatrix};\quad \boldsymbol{Q}_k=\begin{pmatrix}1&0&0&0&1&0&0\\0&1&0&0&0&1&0\\0&0&1&0&0&0&1\\0&0&0&1&0&0&0\\0&0&0&0&10^{-2}&0&0\\0&0&0&0&0&10^{-2}&0\\0&0&0&0&0&0&10^{-4}\end{pmatrix}$$

状态更新使用式(4)：

$$\begin{cases}\hat{x}'_k=\hat{x}_k+\boldsymbol{K}'(\boldsymbol{Z}_k-\boldsymbol{H}_k\hat{x}_k)\\ \boldsymbol{P}'_k=\boldsymbol{P}_k-\boldsymbol{K}'\boldsymbol{H}_k\boldsymbol{P}_k\\ \boldsymbol{K}'=\boldsymbol{P}_k\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_k\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k)^{-1}\end{cases} \tag{4}$$

在得到 k 帧的状态预测值和 k 帧的协方差矩阵之后，由于实验中需要检测到状态量的前 4 个值，即边框的中心位置、面积和宽高比，使用 3×7 的矩阵 $\boldsymbol{H}$ 作为测量矩阵，4×4 的矩阵 $\boldsymbol{R}$ 代表传感器的不确定性。将预测到的 k 帧状态值和协方差矩阵带入式(4)中，完成状态更新(update)过程。将更新过的状态值和协方差矩阵再次带回式(3)可以对 k+1 帧的状态量和协方差进行预测和升级。

$$\boldsymbol{R}_k=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&10&0\\0&0&0&10\end{pmatrix};\quad \boldsymbol{H}_k=\begin{pmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\end{pmatrix}$$

1.3 使用匈牙利算法进行轨迹和检测的分配

匈牙利算法主要用于解决指派问题[10]，是一种用于求解最大匹配的算法。卡尔曼滤波器使用匀速模型进行速度的预测[11]。在一般拍摄条件下，相机的移动会引起 ID 切换(ID switch，IDsw)。为了改善这个问题，基于预测的边框与检测到的同一个物体边框之间面积变化不大这一思想，使用交并比和边框的面积联合来构成匈牙利算法的代价矩阵，见式(5)：

$$\begin{cases} O=\lambda_1 O_1+(1-\lambda_1)O_2 \\ O_1=\dfrac{S_{\text{overlap}}}{S_1+S_2-S_{\text{overlap}}} \\ O_2=\begin{cases} 1-\dfrac{|S_2-S_1|}{S_1}, & |S_2-S_1|\leqslant S_1 \\ 0, & |S_2-S_1|>S_1 \end{cases} \end{cases} \tag{5}$$

式中：S_1,S_2 为两边框的面积；O 为代价矩阵中对应元素的值；O_1 为两边框 IOU 部分对 O 的贡献；O_2 为两边框面积关系对 O 的贡献，其通过比较二者的边框的面积并且将其映射到[0~1]区间得到。

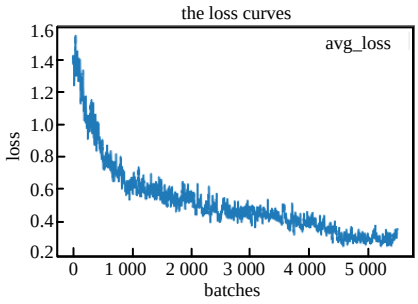

Fig.6 Process of train improved or dot object

图 6 针对小目标提升的网络训练过程的 loss 收敛过程

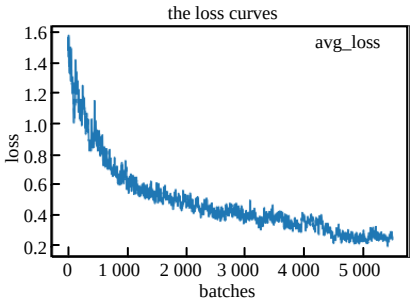

Fig.7 Process of train for normal object

图 7 适用于一般目标的网络训练过程的 loss 收敛过程

实验中通过调整二者的权重来改善 IDsw 的情况，在相机移动的情况较多时交并比部分减小，在相机移动的情况较少时交并比的权重增大。

2 实验结果

实验使用基于针对小目标改进的 YOLOv3 的网络，通过使用辅助类别、卡尔曼滤波和针对 ID 切换问题改进的匈牙利算法完成了固定翼飞机的检测和跟踪。构建的训练集以及测试集在标定时遵循以下原则：a) 尽早标定；b) 标记框尽量紧凑；c) 标注辅助的类别。实验的跟踪视频使用从网上下载的有相机移动的固定翼飞机视频来进行，通过每隔 10 帧抽取的视频帧并标定构成视频检测数据集合，在其上进行用以表征检测效果的 mAP 计算。实验中使用 NVIDIA GeForce GTX 1060 6 GB GPU 训练与检测追踪。

表 2 2 种网络训练过程中的 mAP

Table2 mAP of two nets during training

batches	improved network/mAP	unimproved network/mAP
4 000	34.34	41.39
5 000	44.28	47.35
6 000	47.41	44.90
7 000	47.27	–
8 000	40.20	–

表3 针对小目标改进的网络对常见尺寸的检测结果

Table3 mAP of improved net for normal size objects

detection category	AP_{75}	AP_{50}
aircraft	60.11	62.43
bird	42.21	39.44
engine	39.90	40.18

实验对比了针对小目标改进的 YOLOv3 网络和针对一般尺寸的 YOLOv3 算法网络在测试集合上的检测效果，2 个网络结构都使用同样的锚框的初值，具体值见表 1。改进的网络结构通过使用第 6 000 个循环获得的权重值，能够在测试集合上获得最大的 mAP，一般的网络结构通过使用第 5 000 个循环获得的权重值，能够在测试集上获得最大的 mAP，具体数值见表 2，训练过程的损失(loss)下降过程见图 6~图 7。使用第 6 000 个循环获得的权重值作为改进的网络的最终的权重值，构建检测器。其 3 种类别的检测效果见表 3。实验结果表示使用针对小目标改进的网络结构在训练过程中的损失最终能够达到 0.3，且二者在一般尺寸的目标检测方面没有明显的区别，并不会对一般尺寸目标的检测过程产生很大影响。改进的网络结构的 mAP 能够达到 47.41，获得较好

的检测效果。

由于追踪部分算法属于逐帧跟踪(track-by-frame)的方法，检测结果的好坏直接影响最终跟踪结果的好坏，本文利用检测类别内在关系，联合亚类—飞机发动机，进行固定翼飞机的辅助判断，在视频序列集合上，使得固定翼飞机的 AP_{50} 成功提高 1.27，具体结果见表 4。再通过改进的匈牙利算法的代价(cost)矩阵，通过调整式(5)中 λ 的权重，在有 2 次相机快速移动的情况下，成功将较难目标—飞机发动机的 ID 切换由 2 降到 0，增加了系统的稳定性和鲁棒性，具体见表 4~表 5，效果见图 8~图 9。

表 4 2 种网络训练过程中的 mAP

Table2 mAP of two nets during training

	with joint judgment	without joint judgment
AP_{50}	76.39	75.12

表 5 改进的匈牙利代价矩阵对 ID 切换的改善

Table5 IDsw with improved Hungarian cost matrix

λ	0.5	0.6	0.7	0.8	0.9	1.0
ID_{sw}	0	0	2	2	2	2

Fig.8 Situation of ID_{sw} with λ **错误!未找到引用源。**=0.7

图 8 在 λ=0.7 时 ID_{sw} 的情况

Fig.9 Situation of ID_{sw} with λ=0.6

图 9 在 λ=0.6 时 ID_{sw} 改善的情况

整个系统的运行时间具体见表 6。由表可见，系统的时间大多耗费在检测部分，其占用的时间约为总时间的 93%，追踪部分只用 7%。检测器部分基于卷积神经网络，其可以使用更好的 GPU 硬件进行并行加速，本实验使用的是常用的 NVIDIA GeForce GTX 1060 6 GB GPU 来进行加速，如果使用计算能力更强的 GPU 其速度可以进一步提高。

表 6 检测和追踪所用时间

Table6 Time of detecting and tracking

average time of detection process/s	average tracking time/s	total average time of system operation/s
0.109 7	0.001 577	0.102 6

3 结论

本文基于改进的 YOLOv3、卡尔曼滤波器和匈牙利算法，通过使用类别之间的关系进行辅助判断，增加了视频检测过程的稳定性和可靠性，并且使用视频序列中同一物体边框的内在关系，在保证追踪速度的情况下，有效减少了由于相机快速移动，导致同一物体的预测框和实际检测到的边框的交并比过小而产生 ID 切换(ID switch)的现象，增加了视频跟踪过程的鲁棒性。整个流程在 NVIDIA GeForce GTX 1060 6GB GPU 条件下，平均速度能达到 0.109 7 s/帧，追踪过程速度能达到 0.001 577 s/帧，能够满足快速识别与跟踪的要求。

参考文献：

[1] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]// Computer Vision and Pattern Recognition. San Diego,CA,USA:IEEE, 2005:886–893.

[2] 张智博,樊雅玄,孟骁. 基于谱图和神经网络的通信干扰模式识别方法[J]. 太赫兹科学与电子信息学报, 2019,17(6):959–963. (ZHANG Zhibo,FAN Yaxuan,MENG Xiao. Pattern recognition method of communication interference based on power spectrum density and neural network[J]. Journal of Terahertz Science and Electronic Information Technology, 2019, 17(6):959–963.) doi:10.11805/TKYDA201906.0959.

[3] GIRSHICK R B,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Computer Vision and Pattern Recognition. Columbus,OH,USA:IEEE, 2014:580–587.

[4] REDMON J,FARHADI A. YOLOv3:an incremental improvement[J/OL]. arXiv:Computer Vision and Pattern Recognition, 2018.

[5] EVERINGHAM M,VAN GOOL L,WILLIAMS C K I. The PASCAL Visual Object Classes(VOC) challenge[J]. International Journal of Computer Vision, 2010,88(2):303-338.
[6] LIN T Y,DOLLAR P,GIRSHICK R. Feature pyramid networks for object detection[C]// In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu,HI,USA:IEEE, 2017:2117-2125.
[7] REN S,HE K,GIRSHICK R B,et al. Faster R-CNN:towards real-time objection detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(6):1137-1149.
[8] REDMON J,FARHADI A. YOLO9000:better faster stronger[C]// Computer Vision and Pattern Recognition. Beijing:IEEE, 2017:7263-7271.
[9] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960,82(1): 35-45.
[10] KUHN H W. The hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955,2(1):83-97.
[11] BEWLEY A,GE Z,OTT L,et al. Simple online and realtime tracking[C]// International Conference on Image Processing. Phoenix,AZ,USA:IEEE, 2016.

作者简介：

陈　斌(1993-)，男，甘肃省嘉峪关市人，在读硕士研究生，主要研究方向为图像处理与模式识别. email:chenbin17@ gscaep.ac.cn.

王　磊(1962-)，男，研究员，硕士生导师，主要研究方向为模式识别.

(上接第 904 页)

[15] AKHMEDZHANOV Rinat,GUSHCHIN Lev,NIZOV Nikolay,et al. Microwave-free magnetometry based on cross-relaxation resonances in diamond nitrogen-vacancy centers[J]. Physical Review A, 2017,96(1):013806-1-013806-6.
[16] GLENN David R,FU Roger Rennan,KEHAYIAS Pauli,et al. Micrometer-scale magnetic imaging of geological samples using a quantum diamond microscope[J]. Geochemistry Geophysics Geosystems, 2017,18(8)3254-3267.
[17] YANG Bo,HE Wenhao,GU Bangxing,et al. Precision all-optical EMC test technique of integrated circuits[C]// 2019 12th International Workshop on the Electromagnetic Compatibility of Integrated Circuits(EMC Compo). Hangzhou,China: 2019: 243-245.
[18] 刘颖,董明明,胡振忠,等. 全光学非破坏微波场分布成像[J]. 微波学报, 2019,35(4):86-91. (LIU Ying,DONG Mingming, HU Zhenzhong,et al. All-optical non-destructive microwave field imaging[J]. Journal of Microwaves, 2019,35(4):86-91.)
[19] HU Zhenzhong,YANG Bo,DONG Mingming,et al. Optical sensing of broadband RF magnetic field using a micrometer-sized diamond[J]. IEEE Transactions on Magnetics, 2019,55(3):65003041-65003044.
[20] DONG Mingming,HU Zhenzhong,LIU Ying,et al. A fiber based diamond RF B-field sensor and characterization of a small helical antenna[J]. Applied Physics Letters, 2018,113(13):131105-1-131105-5.

作者简介：

顾邦兴(1996-)，男，在读硕士研究生，主要研究方向为芯片精密测量. email:gubangxing1996@163.com.