



基于坐标注意力机制融合的反无人机系统图像识别方法

薛珊 陈宇超 吕琼莹 曹国华

Image recognition method of anti drone system based on coordinate attention mechanism

Xue Shan, Chen Yuchao, Lv Qiongying, Cao Guohua

在线阅读 View online: <https://doi.org/10.3788/IRLA20211101>

您可能感兴趣的其他文章

Articles you may be interested in

基于卷积神经网络的反无人机系统图像识别方法

Image recognition method of anti UAV system based on convolutional neural network

红外与激光工程. 2020, 49(7): 20200154 <https://doi.org/10.3788/IRLA20200154>

无人机红外辐射建模与图像仿真

UAV infrared radiation modeling and image simulation

红外与激光工程. 2017, 46(6): 628002 <https://doi.org/10.3788/IRLA201746.0628002>

基于DRFP网络的无人机对地车辆目标识别算法

Vehicle target recognition algorithm for UAV image based on DRFP

红外与激光工程. 2019, 48(S2): 125 <https://doi.org/10.3788/IRLA201948.S226001>

点源红外诱饵干扰下环境复杂度量化建模

Quantification modeling for environmental complexity under point source infrared decoy interference

红外与激光工程. 2020, 49(2): 0204003 <https://doi.org/10.3788/IRLA202049.0204003>

融合通道互联空间注意力的Siamese网络跟踪算法

Siamese networks tracking algorithm integrating channel-interconnection-spatial attention

红外与激光工程. 2021, 50(3): 20200148 <https://doi.org/10.3788/IRLA20200148>

基于SAE与底层视觉特征融合的无人机目标识别算法

UAV target recognition algorithm based on fusion of SAE and bottom visual feature

红外与激光工程. 2018, 47(S1): 197 <https://doi.org/10.3788/IRLA201847.S126004>

基于坐标注意力机制融合的反无人机系统图像识别方法

薛 珊^{1,2*}, 陈宇超¹, 吕琼莹^{1*}, 曹国华²

(1. 长春理工大学机电工程学院, 吉林 长春 130022;
2. 长春理工大学重庆研究院, 重庆 400000)

摘要: 反无人机系统是识别和打击“黑飞”无人机的有效手段, 图像识别无人机是反无人机系统的关键之一。针对采集的无人机样本属于小样本、提取特征不够多, 识别准确率不够高的问题, 提出了一种基于迁移学习、密集卷积网络和坐标注意力机制融合的反无人机系统图像识别方法。首先, 运用自制设备采集了多种无人机在不同背景下的图片, 建立数据样本; 其次, 设计针对无人机小样本识别的基于迁移学习、坐标注意力机制和密集卷积网络融合的网络 TL-CA4-DenseNet-121、基于通道注意力机制融合的网络 TL-SE4-DenseNet-121 等网络, 运用设计的网络对小样本进行识别, 并进行对比, 然后分别进行了基于不同位置和不同个数的坐标注意力模块和通道注意力模块的网络识别实验; 最后, 将识别效果最优的网络与经典卷积神经网络模型进行对比实验。实验结果表明, 提出的 TL-CA4-DenseNet-121 网络识别效果优于其他网络, 识别的平均准确率为 97.93%, F1-Score 为 0.9826, 网络训练时间为 6832 s。结果表明了该网络在识别小样本无人机方面的优越性和可行性。

关键词: 无人机; 图像识别; 坐标注意力机制; 密集卷积网络

中图分类号: TP391 文献标志码: A DOI: 10.3788/IRLA20211101

Image recognition method of anti drone system based on coordinate attention mechanism

Xue Shan^{1,2*}, Chen Yuchao¹, Lv Qiongying^{1*}, Cao Guohua²

(1. College of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China;
2. Chongqing Research Institute, Changchun University of Science and Technology, Chongqing 400000, China)

Abstract: Anti drone system is an effective way to identify and attack the "black flying" drone. Image recognition drone is one of the keys of anti drone system. Aiming at the problems that the samples collected from drones are small samples, the features are not enough and the recognition accuracy is not high enough, an image recognition method of anti drone system based on transfer learning, dense convolutional network and coordinate attention mechanism was proposed. Firstly, a variety of drone images in different backgrounds were collected by using self-made device, and data samples were set up; Secondly, the network TL-CA4-DenseNet-121 based on transfer learning, coordinate attention mechanism and dense convolutional network, the network TL-SE4-DenseNet-121 based on channel attention mechanism were designed to identify small samples. The designed network was used to identify small samples and compare. The network recognition experiment of coordinate attention module and channel attention module based on different positions and different numbers were carried out respectively; Finally, the network with the best recognition effect was compared with the classical

收稿日期: 2021-12-20; 修订日期: 2022-02-10

基金项目: 吉林省科技厅重点科技研发项目 (20180201058 SF); 吉林省教育厅科学技术研究项目 (JJKH20210812 KJ)

作者简介: 薛珊, 女, 副教授, 硕士生导师, 博士, 主要研究方向为现代检测理论与技术。

通讯作者: 吕琼莹, 男, 教授, 博士生导师, 博士, 主要研究方向为现代检测理论与技术。

convolutional neural network models. The experimental results show that the proposed TL-CA4-DenseNet-121 network has better recognition effect than other networks, and the average accuracy of recognition is 97.93%, F1-Score is 0.9826 and training time is 6832 s. It shows the superiority and feasibility of this network in identifying small sample drones.

Key words: drone; image recognition; coordinate attention mechanism; dense convolutional network

0 引 言

随着无人机的普及和发展,无人机的“黑飞”行为也越来越威胁到公共安全,尤其在大型公园和游乐场等公共场所,抵制“黑飞”的反无人机系统研发迫在眉睫^[1]。识别无人机是反无人机系统的首要 and 关键问题之一,运用图像识别无人机是重要的识别方法之一。图像识别的关键在于图像特征的提取,使用显著对象作为图像内容的表示。根据特征提取的方式不同,分为传统图像特征手动提取方法和神经网络自动提取特征两种图像识别方法^[2]。

随着技术的发展,手动提取特征已经被神经网络自动提取所取代。图像识别的神经网络模型发展很快。2012年, Krizhevshy 等人提出的 AlexNet^[3] 模型使 ImageNet 数据集的分类准确率得到明显提高; 2014年,由 Google 团队提出的 GoogLeNet^[4] 模型通过引入 Inception 模块增加网络宽度,从而提高模型的表达能力; 同年由 Simonyan 等人提出的 VGG^[5] 模型通过多个小感受野卷积核堆叠的方式代替大感受野的卷积核,从而减少网络参数; 2015年,由 He 等人提出的 ResNet^[6] 模型通过引入残差块结构解决了网络因为层数增加导致的梯度消失和梯度爆炸现象。目前,大部分的卷积神经网络对大样本的识别效果好,准确率高; 样本越充分,效果越好^[7]。但是对于小样本效果

不佳,准确率不高。如何设计一种网络,针对无人机小样本,能够提取更多的特征,提高识别准确率成为了亟待解决的问题。

文中提出了一种基于迁移学习、密集卷积网络和坐标注意力机制融合的图像识别方法。运用迁移学习节省训练时间,并提高准确率; 使用密集卷积网络识别小样本,增多样本的提取通道,增多了特征的提取; 由于采用了密集卷积网络,在增加特征提取的同时,也增加了无效通道; 运用坐标注意力机制突出有效通道,抑制无效通道,突出有效位置,抑制无效位置,从而提高准确率。

1 无人机样本建立及预处理

文中选取“大疆 御 Mavic 2”、“大疆 御 Mavic Air”无人机以及一种自制四旋翼无人机作为研究对象,通过可见光相机与转台集成的实验设备采集无人机原始图片。同时考虑无人机飞行空域的实际情况,文中从加利福尼亚理工学院鸟类数据库-2011(Caltech-UCSD-Birds-200-2011)^[8] 中选取部分鸟类图片作为图像数据集的负样本。由于拍摄的无人机原始图片的像素尺寸不规则,为了使其能够输入卷积神经网络,还需要对这些原始图片做图像预处理操作,图像预处理效果如图 1 所示。通过灰度处理、边缘检测、图像滤

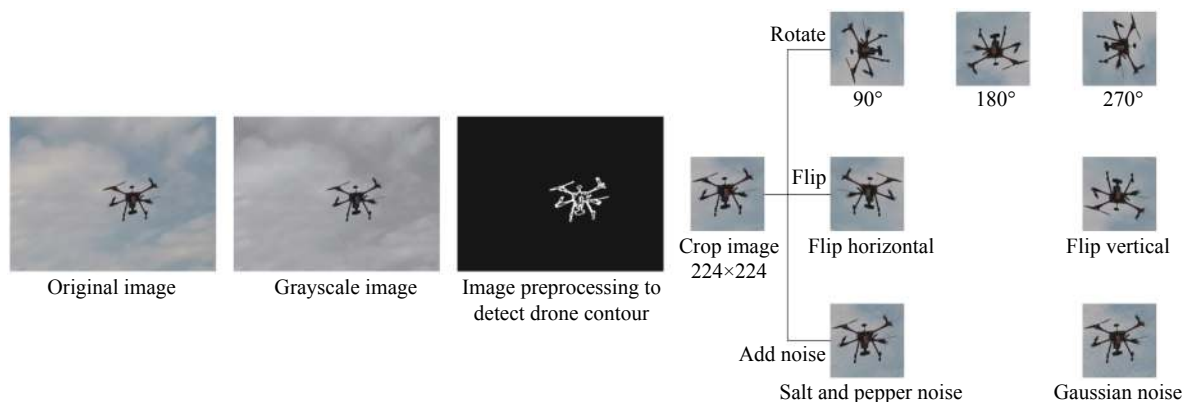


图 1 图像预处理效果图

Fig.1 Effect drawing of image preprocessing

波、图像增强和图像形态学处理等一系列图像处理方
法检测原始图片中的无人机轮廓并裁剪出来;再通过
尺寸归一化操作统一裁剪图片的像素尺寸,由于文
中使用迁移学习方法迁移在 ImageNet 数据集上学习好
的网络预训练参数,因此需要将裁剪图片的像素尺寸
修改为 224×224 pixel,使无人机图片像素尺寸与预训

练网络的输入图片像素尺寸一致;最后通过旋转、翻
转以及添加噪声等图像扩增方式扩充无人机图像数
据集,防止网络出现过拟合现象。

为了更符合实际情况,除了蓝天白云等简单背景
外,还获取包括楼房、树木、操场等复杂背景下的无
人机图片,部分图片展示如图 2 所示。



图 2 无人机图像数据集图片展示

Fig.2 Images display of drone image dataset

2 反无人机系统图像识别网络设计

2.1 基于迁移学习的密集卷积网络设计

文中建立的无人机样本属于小样本,特征提取相对
较少,识别准确率不够高。密集卷积网络 (DenseNet)^[9]
具有密集块 (Dense Block),结构见图 3,该结构使密集
块中的每个卷积层均直接与前面层相连,这样每个层
的输入会接收前面所有层的输出作为额外输入,所有
的输入特征图通过通道级联的方式沿通道方向进行

拼接,使每个层输入的特征图更多,具有更多的图像
特征,实现了特征的重复使用^[10],提高了网络的计算
效率,相对增多了特征的提取。密集块的具体实现为:

$$x_l = H([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

式中: x_l 表示密集块中第 l 层输出的特征图; x_0, x_1, \dots, x_{l-1} 分别表示原始输入、第 1 层输出以及第 $l-1$ 层输出的特征图; $H(\cdot)$ 表示在第 l 层中进行的一系列非线性变换,在密集块中 $H(\cdot)$ 由 BN+ReLU+1×1 Conv+BN+

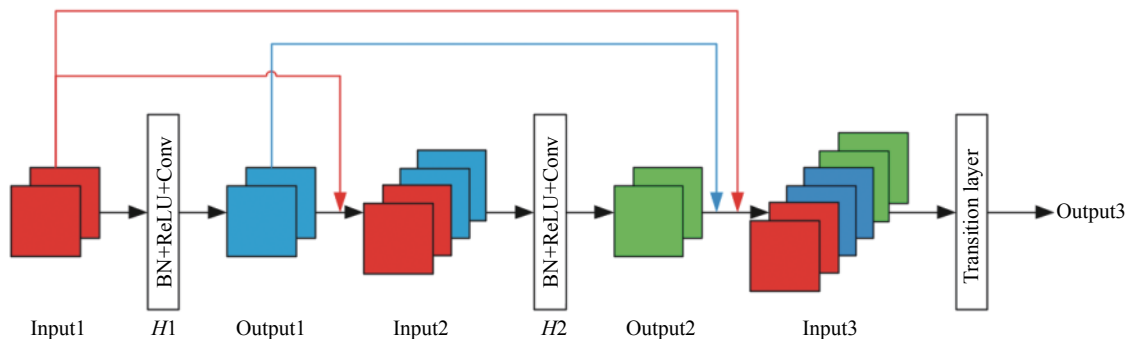


图 3 密集块结构示意图

Fig.3 Diagram sketch of Dense Block

ReLU+3×3 Conv 组成。

迁移学习 (Transfer Learning) 能够将在大型数据集 ImageNet 上学习到的图像分类知识应用于不同但相关的图像分类任务中。迁移学习的网络与训练随机初始化的新网络相比速度更快、构建更容易, 节省训练时间; 而且迁移学习的网络已经学习到足够丰富的图像特征, 与训练随机初始化的新网络相比识别准确率更高; 同时迁移学习的网络不需要大量的数据进行训练, 解决了目标领域中仅有少量样本的学习问题^[11-13]。

文中设计的基于迁移学习的密集卷积网络为 TL-

DenseNet-121, 其结构和各层输出结果如图 4 所示。其中每个特征图下方均标注了特征图的尺寸, 即通道数 $C \times$ 高度 $H \times$ 宽度 W , 特征图通道数 C 用红色标注。输入的无人机图片在 RGB 颜色空间下的通道数为 3, 经过 TL-DenseNet-121 网络中的一系列卷积层、池化层以及 4 个 Dense Block 的特征提取和特征拼接后, 输入进全局平均池化层 (Global Average Pooling) 的特征图由 3 通道变为 1024 通道。通道数的增加证明了 TL-DenseNet-121 可以在无人机图像数据集样本数量较少的情况下增加提取到的图像特征, 使后续的分类结果更准确。

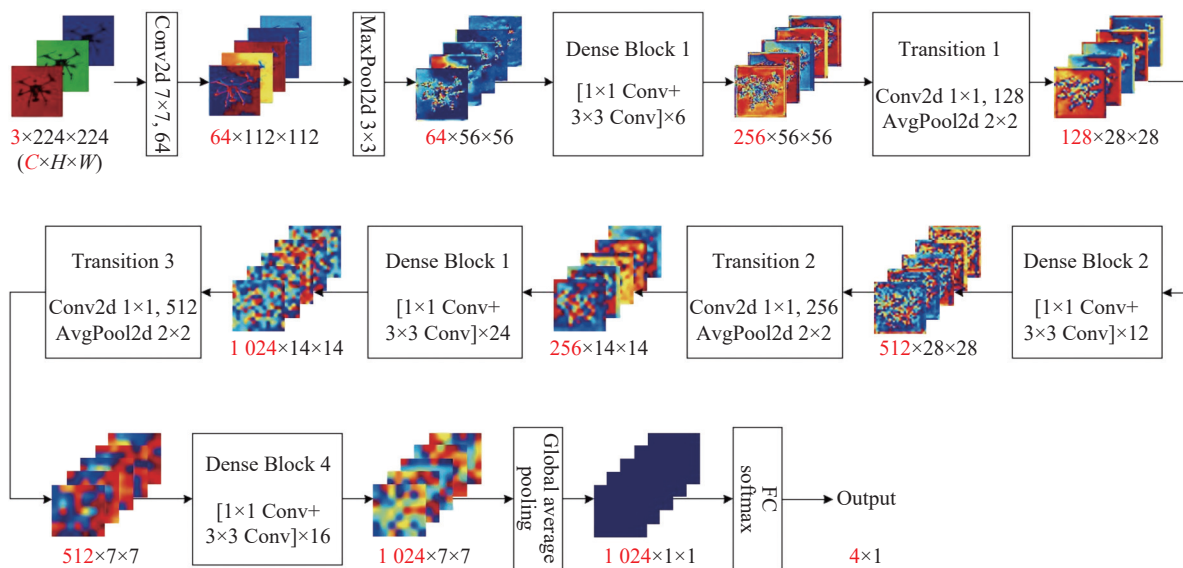


图 4 TL-DenseNet-121 结构图

Fig.4 Structure diagram of TL-DenseNet-121

2.2 基于注意力机制融合的密集卷积网络设计

注意力机制 (Attention Mechanism) 是一种通过模仿人脑来实现从大量输入信息中重点处理有用信息并忽略无关信息的机制^[14], 通过突出模型中对机器学习任务有益的特征信息实现模型的优化。目前, 注意力机制已被广泛应用于图像、语音及文本识别等深度学习任务中^[15]。例如, Liu 等通过使用具有注意力机制和卷积层的双向 LSTM 实现文本分类任务^[16]; 张宇等通过使用注意力机制对输入层不同时刻的语音特征进行加权, 从而提高文中设计的基于多任务学习框架的 LSTM 声学模型识别远场语音的准确率^[17]。因此, 文中对基于迁移学习的密集卷积网络 TL-DenseNet-121 进行改进, 在网络中加入不同的注意力机制模块,

使网络识别无人机图像数据集的准确率更高。

2.2.1 基于通道注意力机制融合的密集卷积网络设计

由于密集卷积网络采用通道级联方式沿深度方向增加了特征图的通道数, 增加了特征提取, 提取了更多的特征, 但是同时也增加了无效的通道, 必须要去除。根据图 3 可以看出, 不断地进行通道级联会使密集块之后的过渡层 (Transition Layer) 的输入特征图通道数达到当前密集块中所有特征图通道数的最大值, 每个过渡层输入的特征图通道数分别为 256、512、1024 和 1024。因此文中设计了基于通道注意力机制融合的密集卷积网络 TL-SE1-DenseNet-121 和 TL-SE4-DenseNet-121。它们是在 TL-DenseNet-121 网络中加入通道注意力机制 SE Block 来降低冗余通道对特征

提取以及网络识别产生的不好影响,使提取的特征更为准确。

通道注意力模块 SE Block^[18] 的结构如图 5(a) 所示,该模块主要包含压缩 (Squeeze) 和激励 (Excitation) 两部分。其中 Squeeze 操作通过一个全局平均池化将输入特征图 $C \times H \times W$ 压缩为 $C \times 1 \times 1$; Excitation 操作通过全连接层和激活函数生成通道权重;最后通过 Scale

操作将通道权重与特征图对应通道的二维矩阵相乘,实现通道加权操作。

文中为了对比融合不同个数和位置通道注意力模块的网络识别效果,设计了融合不同个数和不同位置通道注意力模块的密集卷积网络。分别为融合一个 SE Block 的网络 TL-SE1-DenseNet-121 和融合四个 SE Block 的网络 TL-SE4-DenseNet-121。其结构图如图 6

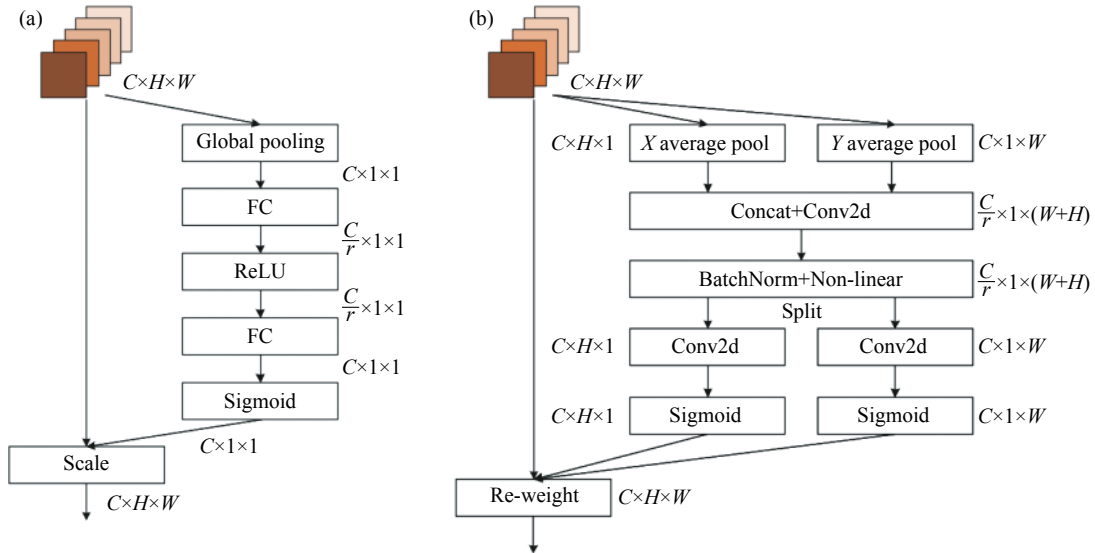


图 5 不同注意力机制模块。(a) SE Block; (b) CA Block

Fig.5 Different attention mechanism modules. (a) SE Block; (b) CA Block

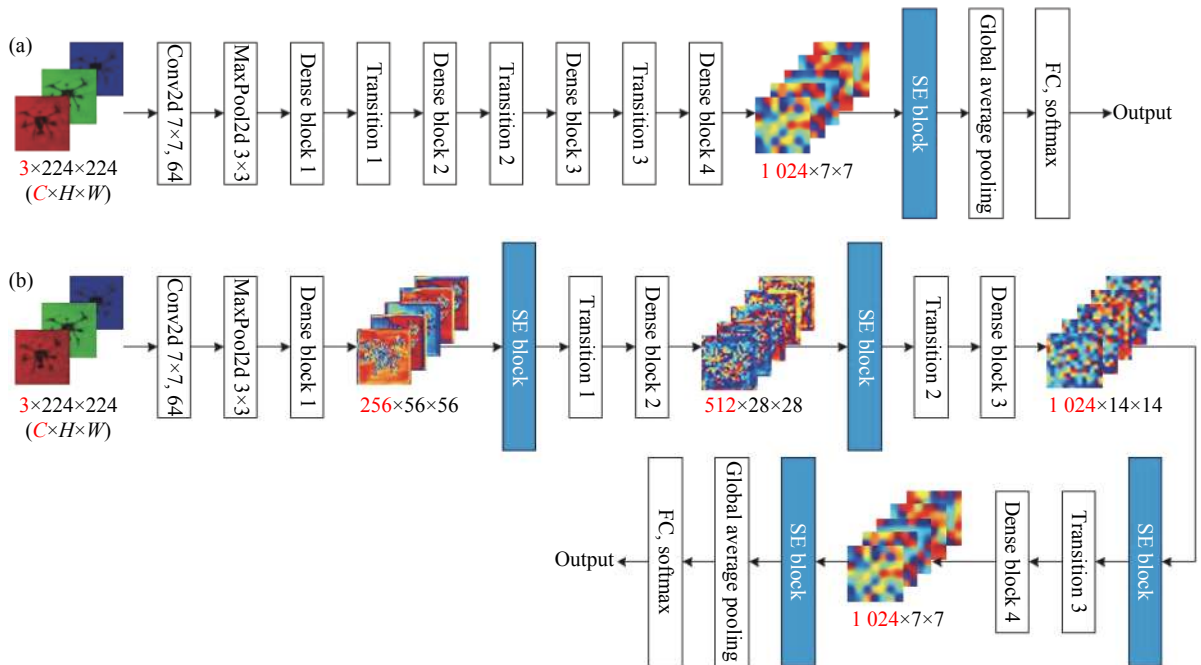


图 6 融合不同个数、不同位置通道注意力模块的密集卷积网络。(a) TL-SE1-DenseNet-121; (b) TL-SE4-DenseNet-121

Fig.6 Dense convolutional network incorporating different number of SE Blocks in different locations. (a) TL-SE1-DenseNet-121; (b) TL-SE4-DenseNet-121

所示。图中, SE Block 用蓝色标出, 它的位置和个数明显可见。融合不同个数和位置的 SE Block 将会对网络的识别效果造成不同影响, 在后面的实验体现。

2.2.2 基于坐标注意力机制融合的密集卷积网络设计

为了减少运用密集卷积网络带来的无效通道的问题, 并且能够使识别准确率更高, 文中设计了融合坐标注意力模块的密集卷积网络 TL-CA1-DenseNet-121 和 TL-CA4-DenseNet-121。它们不仅能降低冗余通道对特征提取以及网络识别产生的不好影响, 还能关注空间信息, 提高识别准确率。坐标注意力模块 CA Block^[19] 的结构如图 5(b) 所示, 它不仅考虑了特征通道之间的关系, 同时还考虑了特征的空间位置信息, 注意的特征信息更为全面。该模块包括了 Coordinate 信息嵌入以及 Coordinate Attention 生成两个步骤。其中 Coordinate 信息嵌入操作通过使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核分别沿水平坐标以及垂直坐标对特征

图进行编码, 该方法可以避免 SE Block 中 Squeeze 操作将全局空间信息压缩至通道描述符从而难以保存位置信息的问题; Coordinate Attention 生成操作通过 Concatenate 操作将不同方向感知的特征图融合, 使用卷积变换改变特征图通道数, 之后将特征图沿空间维数分解为两个单独的张量, 最后通过卷积变换以及激活函数生成两个权重, 将两个权重与原特征图相乘实现加权操作。

文中为了对比融合不同个数和位置坐标注意力模块的网络识别效果, 设计了融合不同个数和不同位置坐标注意力模块的密集卷积网络。分别为融合一个 CA Block 的网络 TL-CA1-DenseNet-121 和融合四个 CA Block 的网络 TL-CA4-DenseNet-121。其结构图如图 7 所示。如图所示 CA Block 用蓝色标出, 它的位置和个数明显可见。不同的识别效果在实验部分体现。

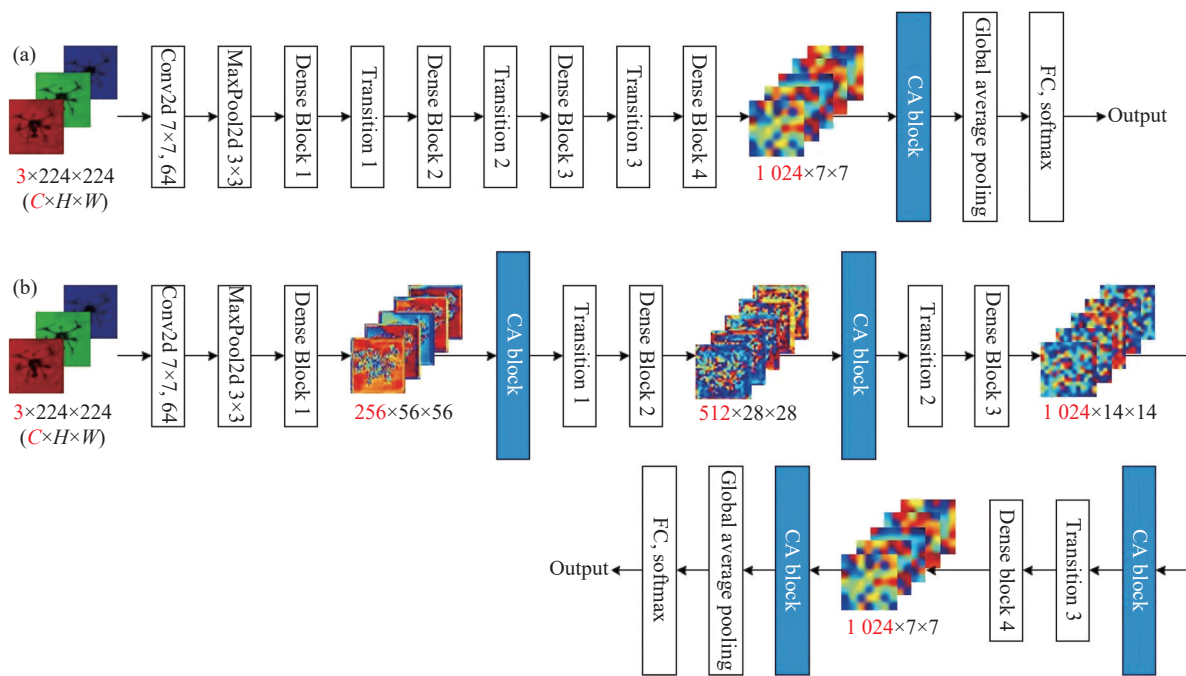


图 7 融合不同个数、不同位置坐标注意力模块的密集卷积网络。(a) TL-CA1-DenseNet-121; (b) TL-CA4-DenseNet-121

Fig.7 Dense convolutional network incorporating different number of CA Blocks in different locations. (a) TL-CA1-DenseNet-121; (b) TL-CA4-DenseNet-121

3 实验

3.1 无人机图片采集

笔者课题组设计的反无人机系统是针对广场、公园等公共场合, 在 1000 m 范围内识别无人机。无人

机图像采集在学校的广场进行, 如图 8 所示, 采集 1000 m 范围内的各种背景下的图片。包括纯天空的背景、有楼的背景、有树的背景。采集的无人机图片经过图像预处理后组成实验所需的图像数据集, 无人机图像数据集分布如表 1 所示。实验中这些图片按照 7 : 3 的

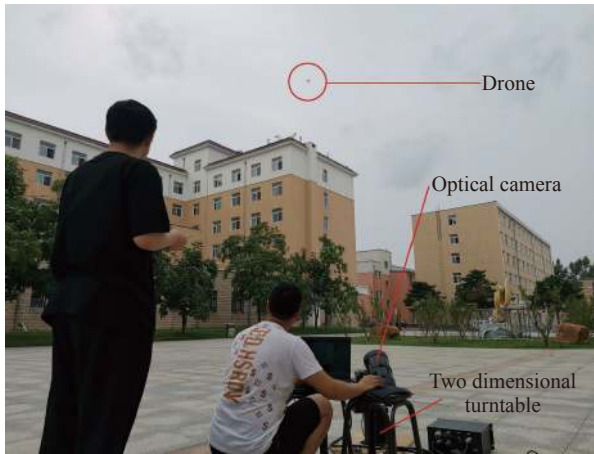


图 8 实验现场图片

Fig.8 Experimental scene pictures

表 1 无人机图像数据集分布情况

Tab.1 Distribution of drone image dataset

	Simple background	Complex background
DJI mavic 2	1 256	960
DJI mavic air	1 088	960
Homemade Quad-Rotor drone	976	624
Caltech-UCSD-Birds-200-2011		2400
Total		8264

比例划分为训练集和验证集。

3.2 实验环境

文中的硬件环境为 Windows 10 系统; Intel(R) Core(TM) i9-10900 F CPU @ 2.80 GHz 2.81 GHz, 32 G RAM; NVIDIA GeForce RTX 2060 GPU, 6 G 显存; 软件环境为 Python 3.7、Sublime、PyCharm; 使用的深度学习框架为 Pytorch。

3.3 基于迁移学习的密集卷积网络识别实验

文中使用 TL-DenseNet-121 网络对无人机图像数据集进行识别实验。实验首先对网络中自定义的全连接层参数进行初始训练, 实验过程中将每批次训练样本数量大小 batch size 设置为 16、历元数 epoch 设置为 50。经过 50 个 epoch 的初始训练后得到 TL-DenseNet-121, 识别结果如图 9 所示。

选取初始训练中验证集准确率最高的 epoch 所对应的模型参数并保存, 根据图 9 可知, 此时 TL-DenseNet-121 在验证集上的最高准确率为 96.77%。

调用初始训练保存的网络模型, 在此基础上对网

```

205 epoch_loss = running_loss / len(dataloaders[phase].dataset)
206 epoch_acc = running_corrects.double() / len(dataloaders[phase]
207
208
209 time_elapsed = time.time() - since
210 print('Time elapsed {:.0f}m {:.0f}s'.format(time_elapsed // 60,
211 print('{} Loss: {:.4f} Acc: {:.4f}'.format(phase, epoch_loss,
212
213 # 得到最好那次的模型
214 if phase == 'valid' and epoch_acc > best_acc:
215     best_acc = epoch_acc
216     best_model_wts = copy.deepcopy(model.state_dict())
217     state = {
218         'state_dict': model.state_dict(),
219
AP: 0.954
tensor([[602., 45., 1., 3.],
        [ 22., 562., 1., 2.],
        [ 40., 0., 478., 2.],
        [ 0., 0., 0., 713.]])
Optimizer learning rate : 0.0010000

Epoch 50/50
-----
Time elapsed 25m 39s
train Loss: 0.0565 Acc: 0.9791
Time elapsed 25m 48s
valid Loss: 0.1808 Acc: 0.9427
AP: 0.954
tensor([[559., 16., 0., 0.],
        [ 45., 581., 0., 1.],
        [ 56., 12., 480., 1.],
        [ 4., 7., 0., 718.]])
Optimizer learning rate : 0.0010000

Training complete in 25m 48s
Best val Acc: 0.967742
-----> 0.967 742
Epoch 1/50
-----
Time elapsed 3m 1s
train Loss: 0.0979 Acc: 0.9656
    
```

图 9 TL-DenseNet-121 的初始训练结果图

Fig.9 Initial training results of TL-DenseNet-121

络中的所有卷积层和全连接层参数做全训练, 网络超参数设置与初始训练相同。经过 50 个 epoch 的全训练后得到 TL-DenseNet-121 的识别结果如图 10 所示。实验选取全训练中 50 个 epoch 的验证集平均准确率以及网络训练时间作为评价指标, 根据图 10 可知, 此时 TL-DenseNet-121 的验证集平均准确率为 95.24%、网络训练时间为 4985 s。

```

Epoch 50/50
-----
Time elapsed 56m 47s
train Loss: 0.0521 Acc: 0.9803
Time elapsed 56m 56s
valid Loss: 0.1594 Acc: 0.9500
AP: 0.952
tensor([[562., 12., 0., 1.],
        [ 54., 599., 0., 2.],
        [ 47., 5., 480., 2.],
        [ 1., 0., 0., 715.]])
Optimizer learning rate : 0.0010000

Training complete in 56m 56s
Best val Acc: 0.963306
[Finished in 4985.4s]
    
```

图 10 TL-DenseNet-121 的全训练结果图

Fig.10 Complete training results of TL-DenseNet-121

为了充分说明模型识别无人机图像数据集的效果, 实验还获取了 TL-DenseNet-121 网络在全训练中验证集准确率最高时的混淆矩阵, 混淆矩阵如公式 (2) 所示, 此时的验证集准确率为 96.33%。混淆矩阵在对角线上的值越大, 表明模型的分类效果越好^[20]。

$$\text{Confusion Matrix} = \begin{bmatrix} 613 & 33 & 1 & 1 \\ 25 & 581 & 2 & 1 \\ 26 & 2 & 477 & 0 \\ 0 & 0 & 0 & 718 \end{bmatrix} \quad (2)$$

通过混淆矩阵可以计算出模型的精确率 (Precision) 和召回率 (Recall), 再根据这两个参数可以计算出 F1-Score, F1-Score 综合了精确率和召回率的结果, 它是评价分类模型的重要指标。其取值范围为 [0, 1], 值越大代表网络输出结果越好。计算公式如下所示:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1 - Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

式中: TP 表示预测为正的样本; FP 表示预测为正的负样本; FN 表示预测为负的正样本。

以“大疆御 Mavic 2”为例计算该情况下的 F1-Score, 将四分类问题合并为二分类问题可得此时 TP=613、FP=35、FN=51, 将数据代入公式 (3)~(5) 可得 F1-Score 为 0.9345, 结果表明 TL-DenseNet-121 能够识别构建的无人机图像数据集。

为了验证迁移学习方法对网络识别效果以及训练速度的影响, 文中还使用未使用迁移学习的训练初始化的 DenseNet-121 网络对无人机图像数据集进行识别实验。实验过程中将每批次训练样本数量大小 batch size 设置为 16, 同时为了对应 TL-DenseNet-121 网络的初始训练和全训练次数, 将历元数 epoch 设置为 100。经过 100 个 epoch 的训练后得到 DenseNet-121 的识别结果如图 11 所示。实验选取训练中 100 个 epoch 的验证集平均准确率以及网络训练时间作为评价指标, 根据图 11 可知此时 DenseNet-121 的验证集平均准确率为 87.64%、网络训练时间为 7258 s。

根据 DenseNet-121 网络在训练中验证集准确率最高时的混淆矩阵可以计算出此时 DenseNet-121 网络的 F1-Score 为 0.8793。通过对比 DenseNet-121 和 TL-DenseNet-121 网络识别无人机图像数据集的实验结果可知: (1) 识别平均准确率和 F1-Score 的不同是因为 TL-DenseNet-121 网络迁移了在大型数据集 ImageNet 上已经学习好的先验知识, 因此在识别不同但相关的图像分类任务时, 网络识别效果要优于训练初始化的

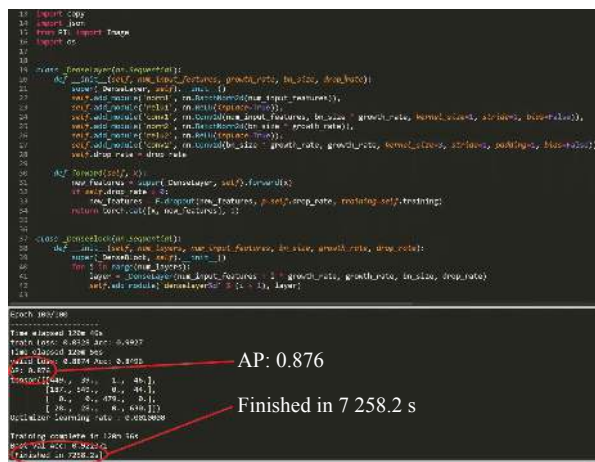


图 11 DenseNet-121 的训练结果图

Fig.11 Training results of DenseNet-121

DenseNet-121 网络; (2) 网络训练时间不同是因为 TL-DenseNet-121 网络使用初始训练和全训练相结合的训练策略, 在初始训练中仅对自定义的全连接层参数进行训练。而训练初始化的 DenseNet-121 网络需要始终对网络中的所有参数进行训练。因此 TL-DenseNet-121 网络的训练时间要比 DenseNet-121 网络短。

通过以上对比结果证明了迁移学习网络的快速性和准确性。

3.4 融合不同个数、不同位置、不同注意力机制模块的网络实验

文中融合不同个数、不同位置、不同注意力机制模块的多个网络对无人机图像数据集进行识别实验。训练策略和超参数设置均与 TL-DenseNet-121 一致。经过 50 个 epoch 的初始训练和 50 个 epoch 的全训练后, 取识别准确率最高的 TL-CA4-DenseNet-121 的识别结果如图 12 所示。

取多个网络在全训练中验证集的准确率曲线与 TL-DenseNet-121 进行对比, 绘制对比曲线如图 13 所示。

为直观反映各网络的识别效果, 以验证集的平均准确率、F1-Score 以及网络训练时间作为评价指标对实验数据进行整理, 实验数据对比结果如表 2 所示。

对比表 2 中的实验数据可知: (1) 无论是加入 SE Block 的 TL-SE1-DenseNet-121、TL-SE4-DenseNet-121 网络, 还是加入 CA Block 的 TL-CA1-DenseNet-121、TL-CA4-DenseNet-121 网络, 它们的识别平均准确率和 F1-Score 均高于未加入注意力机制的 TL-DenseNet-121 网络。但是由于加入注意力机制会使网络参数的

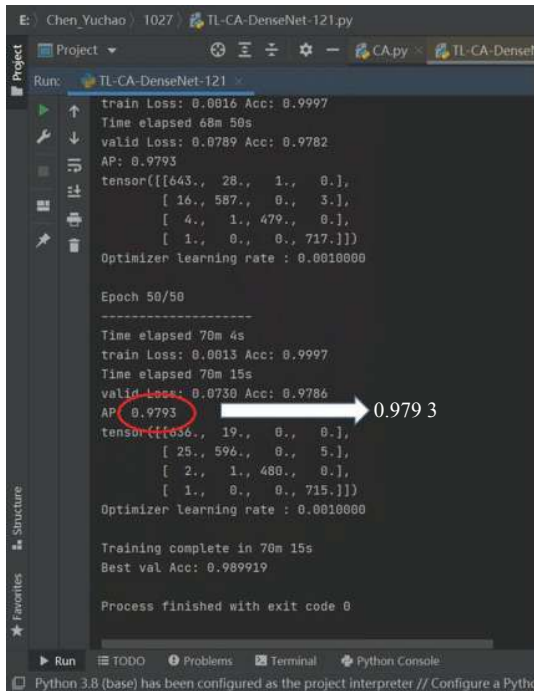


图 12 TL-CA4-DenseNet-121 的全训练结果图

Fig.12 Complete training results of TL-CA4-DenseNet-121

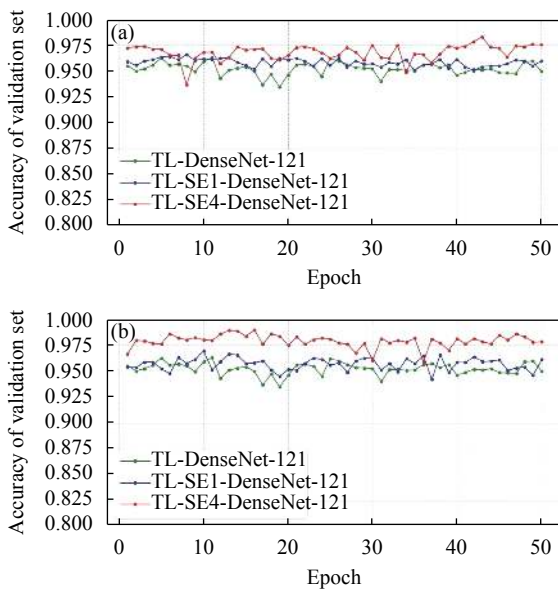


图 13 多种网络的对比曲线。(a) 加入 SE Block 的网络对比曲线；(b) 加入 CA Block 的网络对比曲线

Fig.13 Comparison curves of various networks. (a) Network comparison curves introduced by SE Block; (b) Network comparison curves introduced by CA Block

数量增加, 这些增加的随机初始化的网络参数需要从头开始训练, 因此加入注意力机制的网络训练时间要比未加入注意力机制的 TL-DenseNet-121 网络长。同

表 2 不同网络的对比结果

Tab.2 Comparison results of different networks

	Average accuracy	F1-Score	Training time/s
TL-DenseNet-121	95.24%	0.9345	4985
TL-SE1-DenseNet-121	95.85%	0.9375	4984
TL-SE4-DenseNet-121	96.84%	0.9686	6362
TL-CA1-DenseNet-121	95.70%	0.9490	5602
TL-CA4-DenseNet-121	97.93%	0.9826	6832

时引入注意力机制模块的数量越多, 网络的训练时间越长; (2) 在每个 Dense Block 之后分别加入注意力机制的网络的识别效果要高于只在 Dense Block 4 之后加入注意力机制的网络, 其中引入 CA Block 的 TL-CA4-DenseNet-121 网络的识别平均准确率和 F1-Score 达到最高。

综合对比全部实验数据, 文中选用识别效果最好的 TL-CA4-DenseNet-121 网络作为文中设计的反无人机系统图像识别算法。该网络识别无人机图像数据集的验证集平均准确率为 97.93%, F1-Score 为 0.9826, 网络训练时间为 6832 s。

3.5 多模型对比实验

为了验证文中设计的 TL-CA4-DenseNet-121 网络识别无人机图像数据集的优越性和可行性, 本节使用经典卷积神经网络模型 AlexNet、VGG-16、ResNet-152 和 EfficientNet-B0 对无人机图像数据集进行识别实验。训练策略和超参数设置均与 TL-CA4-DenseNet-121 一致。经过 50 个 epoch 的初始训练和 50 个 epoch 的全训练后, 取多个网络在全训练中验证集的准确率曲线与 TL-CA4-DenseNet-121 进行对比, 绘制对比曲线如图 14 所示。

为直观反映各模型的识别效果, 以验证集的平均准确率、F1-Score 以及网络的训练时间作为评价指标对实验数据进行整理, 实验数据对比结果如表 3 所示。

对比表 3 中的实验数据可知: (1) 与其他经典卷积神经网络模型相比, TL-CA4-DenseNet-121 网络的参数数量较多, 因此 TL-CA4-DenseNet-121 网络的训练时间长于其他对比网络模型; (2) 由于 TL-CA4-DenseNet-121 网络加入多个坐标注意力模块, 使特征图中的通道信息和位置信息更为突出, 因此 TL-CA4-DenseNet-121 网络的平均准确率和 F1-Score 高于其

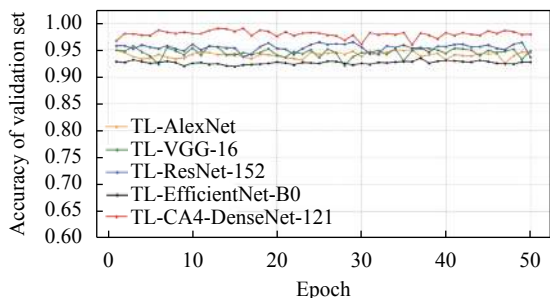


图 14 多模型的对比曲线

Fig.14 Comparison curves of various models

表 3 多模型的对比结果

Tab.3 Comparison results of different models

	Average accuracy	F1-Score	Training time/s
TL-AlexNet	94.06%	0.915 1	1 219
TL-VGG-16	94.49%	0.924 4	6 002
TL-ResNet-152	95.42%	0.934 5	9 023
TL-EfficientNet-B0	92.59%	0.885 3	3 388
TL-CA4-DenseNet-121	97.93%	0.982 6	6 832

他对比网络模型。

通过以上对比结果证明了文中设计的 TL-CA4-DenseNet-121 网络识别无人机图像数据集的优越性和可行性。

4 结 论

(1) 为了解决无人机样本小、特征提取少的问题, 以及提高反无人机系统中无人机图像识别的准确率, 提出了一种基于迁移学习、密集卷积网络和坐标注意力机制融合的反无人机系统图像识别方法。此识别方法采用密集卷积网络增加了图像的特征提取; 采用迁移学习方法提高网络的识别准确率; 采用坐标注意力机制抑制无效通道, 关注位置特征, 提高网络的识别准确率。

(2) 运用自制设备建立无人机图像数据集。设计了针对小样本的基于迁移学习的密集卷积网络 TL-DenseNet-121、设计了基于迁移学习的密集卷积网络分别和不同个数、不同位置、不同注意力机制融合的网络。针对设计的各种网络进行了识别对比实验, 同时将识别效果最好的 TL-CA4-DenseNet-121 网络与经典卷积神经网络模型进行识别了对比实验。实验结果表明了基于迁移学习和坐标注意力机制的密集卷

积网络的优越性和可行性。

参考文献:

- [1] Zhu Mengzhen, Chen Xia, Liu Xu, et al. Situation and key technology of tactical laser anti-UAV [J]. *Infrared and Laser Engineering*, 2021, 50(7): 20200230. (in Chinese)
- [2] Xue Shan, Zhang Zhen, Lv Qiongying, et al. Image recognition method of anti UAV system based on convolutional neural network [J]. *Infrared and Laser Engineering*, 2020, 49(7): 20200154. (in Chinese)
- [3] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [J]. *Advances in Neural Information Processing Systems*, 2012, 25: 1097-1105.
- [4] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions[C]//IEEE Computer Society, 2014.
- [5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv*, 2014: 1409.1556.
- [6] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [7] Zheng Yili, Zhang Lu. Plant leaf image recognition method based on transfer learning with convolutional neural networks [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(S1): 354-359. (in Chinese)
- [8] Wah C, Branson S, Welinder P, et al. The Caltech-UCSD birds-200-2011 dataset[C]//Computation & Neural Systems Technical Report, CNS-TR, 2011: 001.
- [9] Huang Gao, Liu Zhuang, Laurens V, et al. Densely connected convolutional networks[C]//IEEE Computer Society. IEEE Computer Society, 2016.
- [10] Zhu Siqi, Wang Jue, Cai Yufang. Low-dose CT denoising algorithm based on improved CycleGAN [J]. *Acta Optica Sinica*, 2020, 40(22): 2210002. (in Chinese)
- [11] Zhuang Fuzhen, Luo Ping, He Qing, et al. Survey on transfer learning research [J]. *Journal of Software*, 2015, 26(1): 26-39. (in Chinese)
- [12] Zhang Ruiqing, Li Zhangwei, Hao Jianjun, et al. Image recognition of peanut pod grades based on transfer learning with convolutional neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(23): 171-180. (in Chinese)
- [13] Gong Renjie, Zheng Zhihui, Cong Longjian, et al. Infrared target detection and recognition based on transfer learning with small

- samples [J]. *Journal of Northwestern Polytechnical University*, 2021, 39(S1): 84-88. (in Chinese)
- [14] Ren Huan, Wang Xuguang. Review of attention mechanism [J]. *Journal of Computer Applications*, 2021, 41(S1): 1-6. (in Chinese)
- [15] Zhu Zhangli, Rao Yuan, Wu Yuan, et al. Research progress of attention mechanism in deep learning [J]. *Journal of Chinese Information Processing*, 2019, 33(6): 1-11. (in Chinese)
- [16] Liu Gang, Guo Jiabao. Bidirectional LSTM with attention mechanism and convolutional layer for text classification [J]. *Neurocomputing*, 2019, 337: 325-338.
- [17] Zhang Yu, Zhang Pengyuan, Yan Yonghong. Long short-term memory with attention and multitask learning for distant speech recognition [J]. *Journal of Tsinghua University (Science and Technology)*, 2018, 58(3): 249-253. (in Chinese)
- [18] Hu Jie, Shen Li, Albanie S, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [19] Hou Qibin, Zhou Daquan, Feng Jiashi. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 13708-13717.
- [20] Li Xiukun, Xu Tianyang, Ji Shoucong. Posture recognition of underwater target based on deep learning method [J]. *Journal of Harbin Engineering University*, 2021, 42(10): 1503-1509. (in Chinese)