

## 无人机小样本条件下遮挡和混淆目标识别方法\*

吴立珍, 李宏男, 牛轶峰

(国防科技大学 智能科学学院, 湖南 长沙 410073)

**摘要:**针对无人机对地目标识别过程中的小样本问题以及目标存在的遮挡和混淆情况,提出了一种融合自注意力机制的小样本目标识别模型。在利用元学习思想获取小样本学习能力的基础上,将自注意力机制学习目标内部各部分之间的上下文依赖关系引入模型,从而增强目标表征能力,以解决遮挡和混淆情况下有效特征不足的难题。为验证模型效果,通过对基准数据集和无人机航拍数据进一步加工,构建了遮挡和混淆目标数据集,设置了不同的遮挡程度和背景混淆率。通过在不同数据集上的验证,并与深度学习模型对比,证明提出的模型具有更高的学习效率和识别正确率。

**关键词:**无人机;小样本学习;自注意力;遮挡;混淆;目标识别

**中图分类号:**V279 **文献标志码:**A **文章编号:**1001-2486(2022)04-013-09

## Occlusion and confusion targets recognition method for UAV under small sample conditions

WU Lizhen, LI Hongnan, NIU Yifeng

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Aiming at the problem of occlusion and confusion targets recognition for UAV (unmanned air vehicle) under small sample conditions, a target recognition model integrating self-attention mechanism and few-shot learning was proposed. On the basis of using the idea of meta learning to obtain the ability of few-shot learning, the self-attention mechanism to learn the context dependence between the internal parts of the target was introduced into the model, so as to enhance the target representation ability and solve the problem of insufficient effective features in the case of occlusion and confusion. In order to verify the effect of the model, the occlusion and confusion target datasets were constructed by further processing the reference datasets and UAV aerial photography data, and different occlusion degrees and background confusion rates were set. Through the verification on different datasets and compared with the deep learning model, the proposed model is proved to possess higher learning efficiency and recognition accuracy.

**Keywords:** unmanned air vehicle; few-shot learning; self-attention; occlusion; confusion; target recognition

侦察无人机系统常用于未知场景下未知目标的识别。场景的复杂性和目标的特殊性使得这类任务相比于传统目标识别更具有挑战。困难一般源于两个方面:一是目标往往只能获取少量观测样本,其描述存在一定的模糊性和不确定性,导致现有依赖大量训练数据的算法识别效果不太理想;二是目标通常会经受一定程度的伪装处理,主要表现为物理遮挡以及背景混淆,这使得目标可提取的有效特征显著减少,更加剧了这类问题的处理难度。

针对遮挡和混淆目标识别,传统特征提取方法大多基于部件思想<sup>[1-5]</sup>,通过整合多个部件检测器的方式构建分类模型,计算复杂度相对较高,

速度成为该类算法的主要瓶颈。近年来,深度学习方法在这类问题中取得了显著效果。文献[6]利用卷积神经网络(convolutional neural networks, CNN)来学习和整合部件检测器,大大提高了计算的速度,然而弱监督学习过程限制了其效果。文献[7]提出了一种多标签学习过程,在提高部件检测器的可信度和减少计算量方面取得了一定进展。针对目标相互遮挡的情况,文献[8]提出了一种引力-斥力模型,建立目标预测位置、真实位置以及其他目标真实位置之间的距离关系,使目标检测能够在与真实值更逼近的同时减少与其他目标的重叠,提高了目标检测精度。针对目标被其他物体遮挡的情况,文献[9]提出了一种

\* 收稿日期:2021-11-05

基金项目:国家自然科学基金资助项目(61876187)

作者简介:吴立珍(1982—),男,江苏徐州人,助理研究员,博士,E-mail:lzww@nudt.edu.cn;

牛轶峰(通信作者),男,教授,博士,博士生导师,E-mail:niuyifeng@nudt.edu.cn

姿态引导特征对齐 (pose-guided feature alignment, PGFA) 算法, 利用注意力映射驱使模型更关注于未被遮挡的部分, 进而在匹配的过程中尽可能地剔除干扰特征。文献 [10] 提出了一种双边界框方法, 能够对遮挡部分进行估计, 同时提取出目标整体位置和目标可见区域。混淆目标识别的研究对象大多数为动态目标, 原理上主要利用了目标相邻帧之间的动态信息。文献 [11] 受自然界狩猎过程的启发, 提出一种基于搜索识别网络 (search identification network, SINet) 的伪装目标检测框架, 包括搜索模块和识别模块两个部分, 在自然场景应用中取得了较好的效果。

针对小样本条件下的目标识别, 主流方法包括样本扩充、空间映射以及策略搜索等<sup>[12]</sup>。样本扩充方法通过对原有样本数据扩充, 以达到增加样本数量的目的, 进而将小样本识别问题转化为常规目标识别问题处理。数据扩充生成对抗网络<sup>[13]</sup> (data augmentation generative adversarial networks, DAGAN) 是该类方法的典型代表; 空间映射方法通过学习, 将原始样本数据映射到一个更易于分类的空间, 从而降低分类难度, 一般包括空间映射、特征表示以及关系度量等三个环节。现有方法大多选择其中的一个或者几个进行研究, 如孪生网络模型<sup>[14]</sup>、原型网络<sup>[15]</sup>、关系网络<sup>[16]</sup>等。策略搜索方法的原理是利用已有的知识或经验, 提升能够快速进行学习的能力。其中, 比较有代表性的是元学习方法。元学习以任务作为学习的基本单位, 利用先验知识加速模型的学习效率, 能够在泛化性较强的初始网络基础上快速适应新的任务。此外, 基于统计学习的方法在近年来也取得了一些进展。比如 Si 等<sup>[17]</sup> 提出的与或图模型, Lake 等<sup>[18]</sup> 提出的基于与或图模型的贝叶斯学习框架, George 等<sup>[19]</sup> 提出的组合式生成模型递归皮层网络 (recursive cortical network, RCN) 等。

尽管小样本学习方法已经在目标识别领域取得一定进展, 但在面对一些复杂情况, 比如目标遮挡或混淆时, 识别效果仍然难以让人满意。2017 年, Google 机器翻译团队关于自注意力机制在自然语言处理领域的研究<sup>[20]</sup> 受到了广泛关注, 自注意力机制开始逐渐成为神经网络的研究热点。由于自注意力机制能够通过捕捉全局信息实现对上下文信息的建模, 因此在计算机视觉领域具有重要的潜在应用价值。受非局部均值的启发, 文献 [21] 构造了一种深度神经网络中的非局部操作, 可以有效捕获视频图像中时域和空域上的长距离依赖。双重注意力网络<sup>[22]</sup> (dual attention

network, DANet) 同时关注空间和通道两种方式的自注意力机制, 将得到的特征进行融合, 成功用于语义分割问题。目前, 自注意力机制在语义识别以及传统的图像识别领域都得到了一定的应用, 但是还未用于小样本条件下伪装目标的识别问题。

## 1 概述

相对于完整目标的目标识别问题, 遮挡和混淆目标识别的难点在于如何提取出足够的有效特征。自注意力机制能够学习目标部分与部分之间的依赖关系, 通过建立目标观测数据的局部关联性, 达到增加有效特征的目的, 能够有效提高模型的识别正确率。基于这一思想, 提出一种融合自注意力机制的小样本目标识别模型。该模型采用元学习构建目标识别的基础框架, 以获取小样本识别能力, 并融合自注意力机制形成一种新型学习网络, 提升处理遮挡和混淆问题的能力。

为了验证模型性能, 构建了两种遮挡和混淆目标数据集: 一种是对小样本学习领域常用基准数据集的二次制作, 随机添加了 5% 和 10% 两种程度的遮挡; 二是基于无人机航拍图像构建了涵盖 4 类目标、5 级混淆以及 10% ~ 30% 遮挡率的典型目标数据集, 可用于遮挡或混淆条件下的分类、检测、分割等多种视觉任务。

## 2 融入自注意力机制的元学习方法

### 2.1 元学习基础网络框架

基于模型无关自适应学习 (model-agnostic meta-learning, MAML) 模型<sup>[23]</sup> 构建基础网络框架。MAML 模型学习过程如图 1 所示。

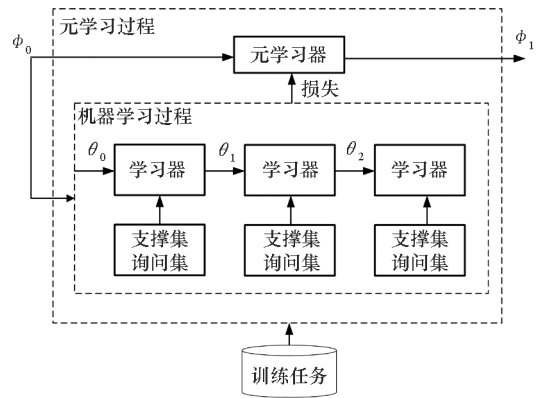


图 1 模型无关自适应学习过程  
Fig. 1 Learning process of MAML

图 1 中,  $\phi$  表示元学习器参数,  $\theta$  表示在不同任务中学习到的参数。MAML 主要关注如何设置

模型初始化参数的问题,它通过元学习方法进行训练,其特点是可以采用不同的网络结构用同一种训练方式进行训练。在该模型中,学习器利用元学习器设计的初始化参数,根据训练数据学习一个具体的分类器模型。MAML 具有的网络结构适应性和小数据集快速训练能力,为自注意力机制模块的融入以及小样本目标的识别提供了良好基础。

## 2.2 自注意力机制模块结构设计

构建自注意力机制网络模块如图2所示。该模块关注嵌入空间中的所有位置并取其加权平均值,按顺序计算某一位置的响应。

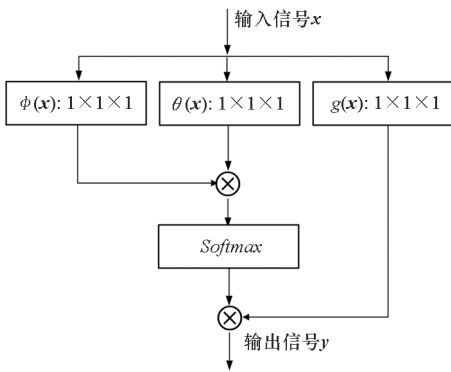


图2 自注意力机制模块网络结构

Fig.2 Network structure of self-attention mechanism module

模块的计算可表示为:

$$y_i = \frac{1}{C(\mathbf{x})} \sum_j f(x_i, x_j) \cdot g(x_j) \quad (1)$$

式中: $\mathbf{x}$ 表示输入信号,即输入图像; $\mathbf{y}$ 表示输出信号,与 $\mathbf{x}$ 具有相同尺度; $x_i$ 和 $x_j$ 分别代表图像位置*i*处和*j*处的图块或像素; $y_i$ 即 $x_i$ 经过自注意力机制处理之后的响应信号。函数 $f(x_i, x_j)$ 表示计算两个图块 $x_i$ 和 $x_j$ 之间的关系,能够基于图像信息获取图像中存在的大范围依赖关系,对于目标样本图像则意味着隐含的目标各部分之间的关系;函数 $g(x_j)$ 表示计算输入图块 $x_j$ 的信息变换;函数 $C(\mathbf{x})$ 是归一化因子。该模块代表的含义是,对于位置*i*,逐一遍历图像中的所有位置*j*,并且计算两者之间的关系以及输入图像在位置*j*的表达,之后进行加和以及归一化。

实现过程中,函数 $f(x_i, x_j)$ 和 $g(x_j)$ 可结合神经网络进行设定。这里, $g(x_j)$ 采用 $1 * 1$ 卷积,即线性函数:

$$g(x_j) = W_g x_j \quad (2)$$

式中: $W_g$ 参数通过学习得到。

函数 $f(x_i, x_j)$ 计算两个位置的相关性,采用嵌入式高斯函数:

$$f(x_i, x_j) = \exp[\boldsymbol{\theta}(x_i)^T \boldsymbol{\phi}(x_j)] \quad (3)$$

$$C(\mathbf{x}) = \sum_j f(x_i, x_j) \quad (4)$$

式中: $\boldsymbol{\theta}(x_i) = W_\theta x_i$ , $\boldsymbol{\phi}(x_j) = W_\phi x_j$ 。 $W_\theta$ 和 $W_\phi$ 是需要学习的参数。此式等价为一个Softmax函数:

$$\mathbf{y} = \text{Softmax}(\mathbf{x}^T W_\theta^T W_\phi \mathbf{x}) g(\mathbf{x}) \quad (5)$$

需要指出的是,函数 $f(x_i, x_j)$ 的选择多样,可以使用高斯函数、点乘函数、串联函数等。

## 2.3 融入自注意力机制的元学习模型

采用残差连接方式,将自注意力机制模块的输出与MAML基础网络相融合,实现对MAML网络结构的改进,形成的完整网络结构如图3所示。

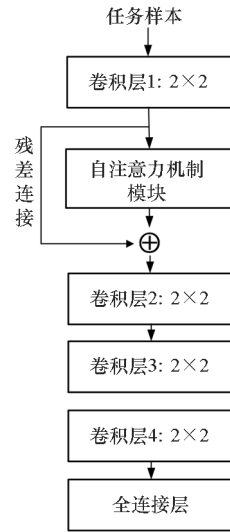


图3 融入自注意力机制模块的学习网络

Fig.3 Learning network integrated with self-attention module

残差连接可以表示为:

$$Z_i = W_z Y_i + x_i \quad (6)$$

残差连接方式可以非常方便地在已设计好的基础网络模型中插入新的模块。值得注意的是,图3中将自注意力机制模块插入MAML基础网络的第一层卷积层中,其目的是使自注意力机制能够对输入的任务样本进行目标内部关系建模,更高效地提升模型的特征提取能力。自注意力机制模块的卷积层是 $1 * 1$ 的卷积核,其他的几个卷积层都是 $2 * 2$ 的卷积核。

## 3 数据集构建

目前,遮挡和混淆目标识别的研究缺乏大型基准数据集支持。为验证提出的融入自注意力机制的元学习模型在无人机遮挡和混淆目标识别中的有效性,分别基于公开的小样本学习基准数据集和无人机航拍数据构建了两种遮挡和混淆目标数据集。

### 3.1 miniImageNet 数据集遮挡处理

miniImageNet 在 ImageNet 数据集基础上提取而成<sup>[24]</sup>, 是小样本学习领域常用的基准数据集之一。miniImageNet 包含 100 种类别, 每类有 600 个彩色图像样本。其样本数据涵盖了实际场景中常见的目标类别, 因此在目标检测和识别的算法对比研究中应用广泛。为了实现遮挡效果, 对 miniImageNet 数据集人为添加随机遮挡, 遮挡程度分为 5% 面积遮挡和 10% 面积遮挡两种。遮挡采用随机位置和随机颜色的方式进行, 添加遮挡后的数据集如图 4 和图 5 所示。

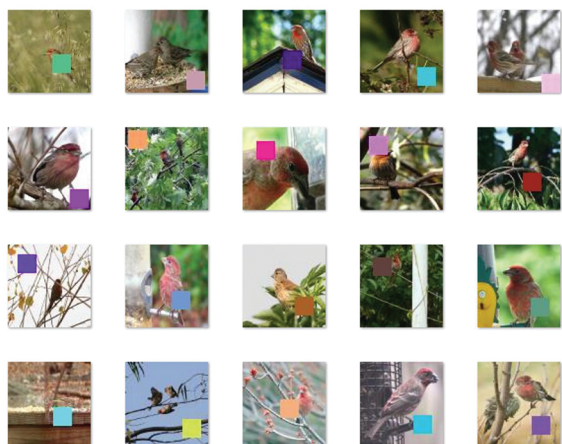


图 4 5% 遮挡率的 miniImageNet 数据集

Fig. 4 miniImageNet with 5% occlusion rate

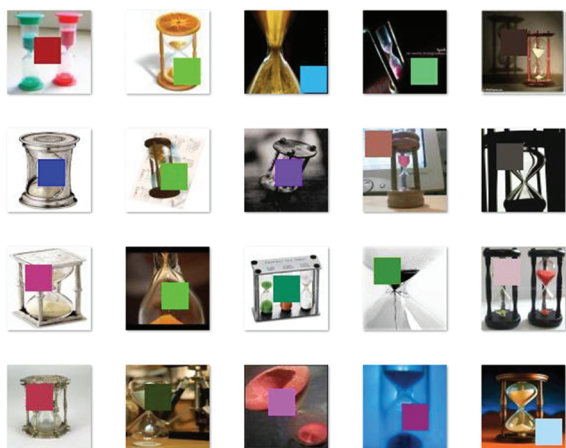


图 5 10% 遮挡率的 miniImageNet 数据集

Fig. 5 miniImageNet with 10% occlusion rate

### 3.2 无人机航拍数据集构建

为了验证模型在无人机应用场景中的识别效果, 选择无人机对地观测图像构建了航拍数据集, 分为遮挡目标数据集和混淆目标数据集两类。

遮挡目标数据集图像包括了行人、建筑、车辆、飞机、坦克和舰艇 6 类目标, 如图 6 所示。制

造遮挡的方式为在背景中随机裁剪出部分图像, 将其遮挡在目标上。随机遮挡面积的大小为目标面积的 10% 和 30% 之间。



图 6 航拍遮挡目标数据集

Fig. 6 Dataset of occluded targets in aerial images

混淆目标数据集包括飞机、坦克和舰艇 3 类目标, 目标局部覆盖有与背景近似的伪装网。以伪装处理的面积作为混淆程度的度量标准, 将混淆程度分为 0 ~ 4 共 5 个级别, 代表了从无混淆处理到完全混淆处理的 5 阶变化, 如图 7 所示。

数据采集采用无人机视频拍摄方式进行, 并按照每 10 帧取 1 帧的方式进行采样。裁剪的目标图像统一为 150 像素 × 150 像素大小。为保持数据集中各类别样本数目的均衡性, 在各个混淆程度下的每个类别中都采样 100 个样本。

## 4 实验与分析

在构建的遮挡和混淆目标数据集上进行目标识别实验。模型训练过程利用公开数据集 miniImageNet 中的训练集和验证集。测试过程分别利用加工处理后的 miniImageNet 数据集的测试集以及无人机航拍数据集, 以验证算法的效果。测试过程中目标类别在 miniImageNet 训练集和验证集中均未出现过。

实验选择 3 分类任务, 每一类包含的支撑集样本为 1 个或者 5 个, 即为 3 - way 1 - shot 任务或者 3 - way 5 - shot 任务。

### 4.1 模型训练过程

在模型训练过程中, 设置模型训练次数为 60 000 次, 每个任务进行 5 次梯度下降。

为了考察训练过程, 分别计算每一次训练过程中第 1 次迭代之后训练集的正确率以及第 5 次迭代之后训练集的正确率变化曲线, 如图 8 所示。





图7 航拍混淆目标数据集

Fig.7 Dataset of confusion targets in aerial images

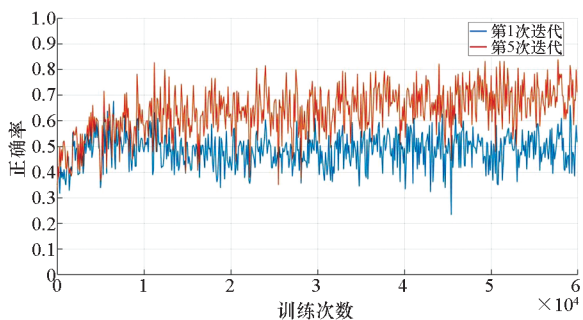


图8 模型训练过程中的正确率变化

Fig.8 Correct rate in model training

3 分类任务中,每一个小样本学习任务训练之前的正确率应当近似于随机猜想。可以看出,当经过少数次迭代后,蓝色曲线和橙色曲线代表的正确率均逐步上升,说明模型逐渐学到一个较好的初始化参数。当面对新的学习任务时,经过快速迭代就能够有比较好的表现。

#### 4.2 遮挡目标识别的实验结果

为了验证模型在遮挡目标识别任务中的表现,首先在 miniImageNet 数据集上进行验证。区

分无遮挡、遮挡率 5% 以及遮挡率 10% 三种情况, 3 - way 1 - shot 和 3 - way 5 - shot 任务的测试过程中正确率变化曲线如图 9 和图 10 所示。

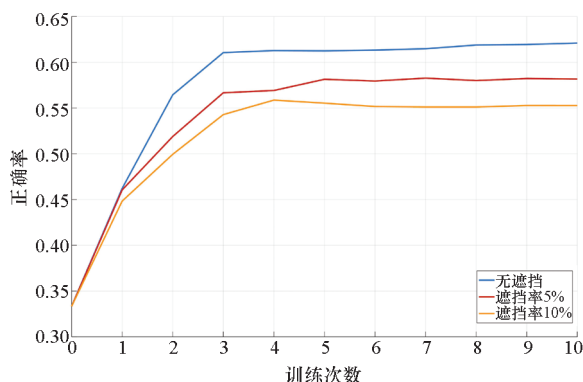


图9 3-way 1-shot 任务不同遮挡条件下的识别正确率

Fig.9 Recognition rate in 3-way 1-shot task under different occlusion conditions

从图 9 和图 10 可以看出,模型在每类只有 1 个或 5 个支撑集样本的情况下,经过 3 至 5 次迭代,即可达到比较好的识别正确率。随着遮挡程

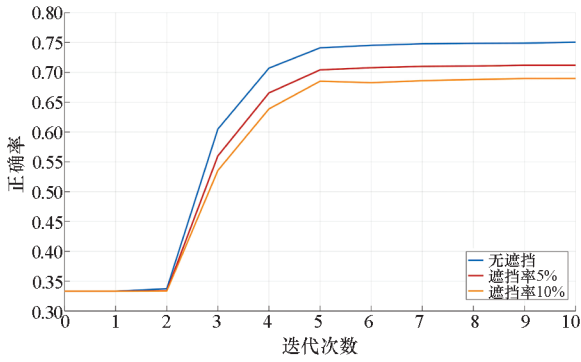


图 10 3-way 5-shot 任务不同遮挡条件下的识别正确率

Fig. 10 Recognition rate in 3-way 5-shot task under different occlusion conditions

度的升高,模型识别效率会小幅下降,这是因为遮挡程度变大后,样本识别难度不断增加。识别正确率如表 1 所示。

表 1 miniImageNet 数据集上遮挡目标的识别正确率  
Tab. 1 Recognition rate in miniImageNet

任务	无遮挡	遮挡率 5%	遮挡率 10%
3-way 1-shot	62.09	58.17	55.27
3-way 5-shot	75.05	71.19	68.97

利用 ResNet18<sup>[25]</sup> 模型和 MAML 模型进行对比实验。由于 ResNet18 模型在 3-way 1-shot 任务中表现近似于随机猜想,因此选择 5 样本条件下进行不同模型比较,结果如图 11 和图 12 所示。可以看出,本文模型相比 ResNet18 模型在小样本条件下的识别正确率显著提升,相比 MAML 模型表现效果相差不多,但是具有更高的训练效率,经过 5 次迭代的识别正确率如表 2 所示。

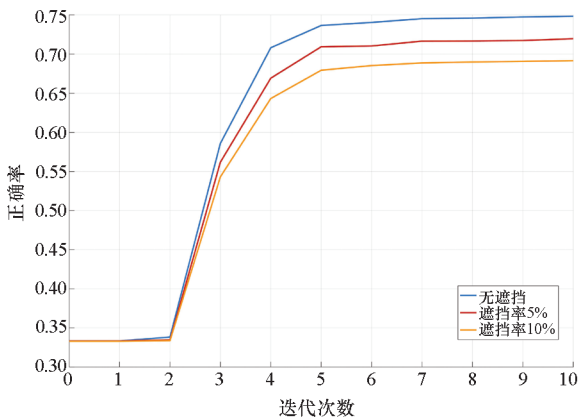


图 11 MAML 模型 3-way 5-shot 任务识别正确率

Fig. 11 Recognition rate in 3-way 5-shot task of MAML

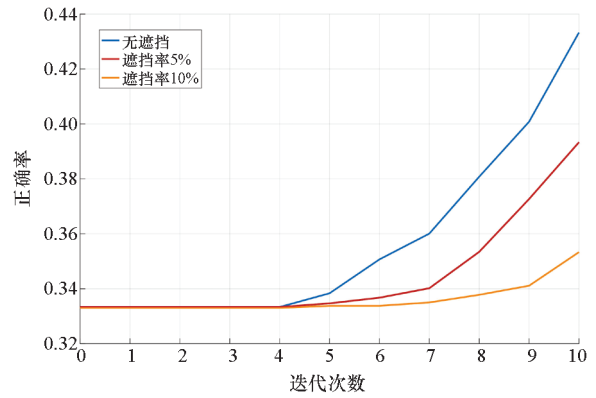


图 12 ResNet18 模型 3-way 5-shot 任务识别正确率

Fig. 12 Recognition rate in 3-way 5-shot task of ResNet18

表 2 miniImageNet 数据集上 5 次迭代的识别正确率

Tab. 2 Recognition rate of 5 iterations in miniImageNet

模型	无遮挡	遮挡率 5%	遮挡率 10%
本文模型	74.11	70.42	68.52
MAML	73.65	70.93	67.93
ResNet18	33.83	33.47	33.38

针对无人机航拍数据集,本文提出的模型识别正确率如图 13、图 14 所示。

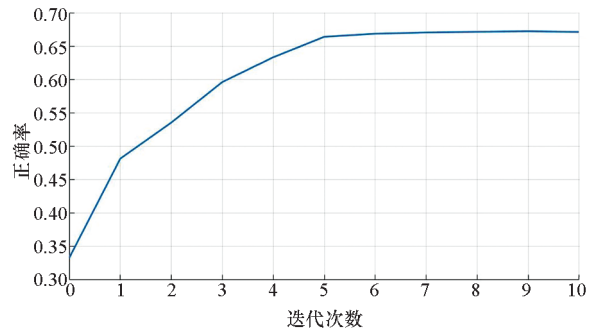


图 13 遮挡目标 3-way 1-shot 任务中识别正确率

Fig. 13 Recognition rate of occluded targets in 3-way 1-shot task

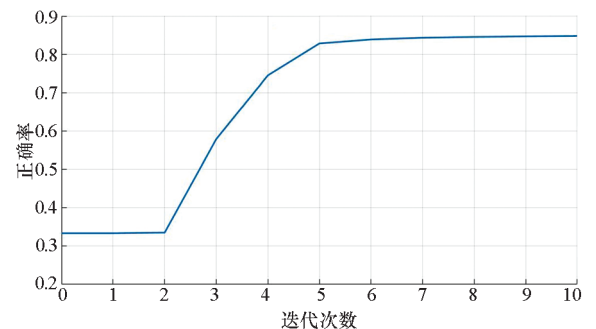


图 14 遮挡目标 3-way 5-shot 任务中识别正确率

Fig. 14 Recognition rate of occluded target in 3-way 5-shot task

进一步地,利用 ResNet18 模型和 MAML 模型进行对比实验。设计 ResNet18 训练过程中迭代次数为 10。同时,为了考察模型在不同样本数量下的识别正确率,按照 1、5、10、20、30、40、50 的梯度设置样本数量,不同模型的识别正确率变化如图 15 和表 3 所示。其中,本文模型和 MAML 模型只在小样本条件下进行测试。可以看出,相比于 ResNet18 模型,融合自注意机制的元学习模型具有更高的学习效率,只需要大约 1/8 的样本数量即可达到与之相当的识别效果,并表现出其在小样本条件下的性能优势。在 5 样本条件下,相比于 MAML 模型,本文算法的识别正确率由 64.78% 提升到 84.77%。由此可见,在元学习基础网络框架中嵌入自注意机制,对遮挡目标的识

别正确率提升具有显著的作用。

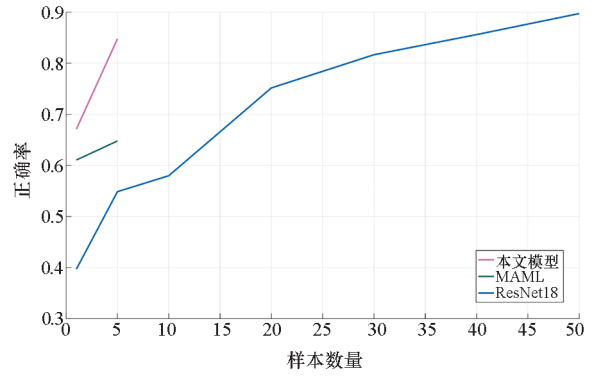


图 15 航拍数据集遮挡目标识别正确率对比  
Fig. 15 Comparison of recognition accuracy of occluded targets in aerial data sets

表 3 遮挡目标的识别正确率对比

Tab. 3 Recognition rates comparison of occluded target

模型	样本数目							%
	1	5	10	20	30	40	50	
本文模型	67.11	84.77	—	—	—	—	—	—
MAML <sup>[23]</sup>	61.06	64.78	—	—	—	—	—	—
ResNet18 <sup>[25]</sup>	39.70	54.86	57.98	75.17	81.67	85.61	89.73	—

### 4.3 混淆目标识别的实验结果

针对混淆目标,测试本文模型在两类任务以及不同混淆程度下的识别正确率,结果分别如图 16 和图 17 所示,其中不同颜色的曲线表示不同的混淆程度。模型在测试过程中每一个任务迭代 10 次。可以看出,模型在少量迭代后可以迅速达到较高的正确率,这说明模型能够快速适应新

的任务,具有了较强的学习能力。同时,随着样本混淆程度的增加,模型的识别正确率会有所下降。相同条件下,3-way 5-shot 任务的正确率相比于 3-way 1-shot 任务有很大提高,这说明模型在样本数目少量增加后会有比较大的性能提升。

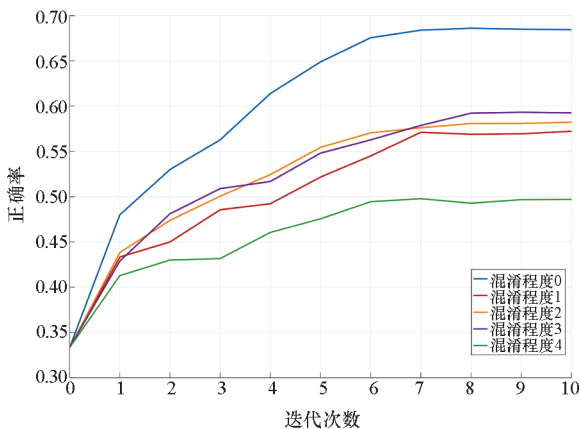


图 16 3-way 1-shot 任务中不同混淆程度下识别正确率

Fig. 16 Recognition rate of different levels of confusion in 3-way 1-shot task

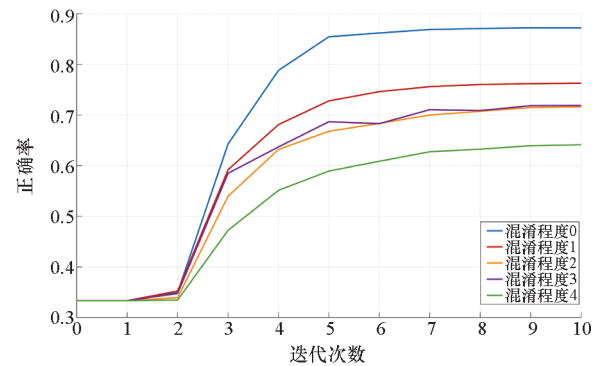


图 17 3-way 5-shot 任务中不同混淆程度下识别正确率

Fig. 17 Recognition rate of different levels of confusion in 3-way 5-shot task

进一步开展对比实验,计算 ResNet18 模型对混淆目标的识别正确率如表 4 所示。可以看出,该模型在样本数量较少时,表现不够理想,该模型难以应对小样本识别问题。

表 4 ResNet18 模型对混淆目标的识别正确率  
Tab.4 Confused target recognition rates of ResNet18 model

混淆程度	样本数目						
	1	5	10	20	30	40	50
0	38.15	49.58	44.48	73.26	83.40	88.89	90.67
1	35.08	39.16	44.61	58.54	76.87	78.28	87.00
2	38.35	42.11	46.74	61.46	71.76	72.67	80.67
3	35.08	44.47	45.67	57.58	71.19	74.67	80.60
4	33.77	40.74	45.87	61.04	67.14	72.28	74.87

表 5 展示了 3 - way 1 - shot 任务中不同模型的混淆目标识别正确率。可以看出,相比于未引入自注意力机制的 MAML 模型,本文模型在混淆程度较低的情况下识别效果有着一定的提升。由于

ResNet18 模型在单样本识别任务中效果较差,所以未列入比较。3 - way 5 - shot 任务中不同模型的混淆目标识别正确率变化情况如图 18 和表 6 所示。在没有伪装处理的情况下,本文模型的正确率

表 5 3 - way 1 - shot 任务中不同模型的混淆目标识别正确率对比  
Tab.5 Recognition rates comparison of confused target in 3 - way 1 - shot task

模型	混淆程度				
	0	1	2	3	4
本文模型	<b>68.60</b>	<b>57.20</b>	58.20	59.30	<b>49.80</b>
MAML <sup>[23]</sup>	60.50	50.89	<b>58.94</b>	<b>60.39</b>	48.67

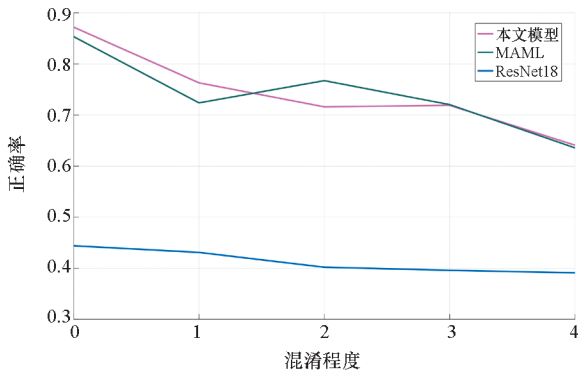


图 18 航拍数据集混淆目标识别正确率对比  
Fig. 18 Comparison of recognition accuracy of confused targets in aerial data sets

可以达到 87.2% 以上,自注意力机制的融入对于低混淆程度下的目标识别具有明显的提升效果。

综合分析上述实验结果可知,融入自注意力机制的元学习模型借助于元学习基础网络框架,使得其在小样本条件下的目标识别正确率显著优于 ResNet18 模型。同时,嵌入其中的自注意机制有效提升了对目标内部上下文依赖关系的描述能力,相比传统 MAML 模型能够更好地实现目标表征,有利于解决遮挡和混淆等情况带来的目标有效特征不足问题,在多数情况下识别速度和识别准确率都有所提升。

表 6 3 - way 5 - shot 任务中不同模型的混淆目标识别正确率对比  
Tab.6 Recognition rates comparison of confused target in 3 - way 5 - shot task

模型	混淆程度				
	0	1	2	3	4
本文模型	<b>87.20</b>	<b>76.30</b>	71.60	71.90	<b>64.10</b>
MAML <sup>[23]</sup>	85.34	72.38	<b>76.73</b>	<b>72.04</b>	63.57
ResNet18 <sup>[25]</sup>	44.40	43.10	40.20	39.60	39.10



## 5 结论

本文针对小样本条件下遮挡和混淆目标识别问题,构建了一种融合小样本学习和自注意力机制的目标识别模型。主要结论如下:

1)模型基于元学习基础网络框架构建,保留了小样本学习的优势,在面对新任务时能够迅速迭代。同时,考虑到遮挡和混淆条件,引入自注意力机制学习目标内部上下文依赖关系,提升了模型表征能力。

2)针对遮挡和混淆目标识别研究缺乏基准数据集支持的问题,在 miniImageNet 数据集和无人机航拍图像基础上,构建了涵盖不同遮挡程度和背景混淆率的目标数据集,可用于多种视觉任务。

3)完成了多种条件下遮挡和混淆目标识别实验,并展开了与 MAML 模型、ResNet18 模型的比较分析,证明了本文模型的有效性和先进性。相比于深度学习方法,本文模型在相同条件下只需要很少的样本数量即可达到相当的识别正确率,学习效率具有显著优势;与传统的小样本学习相比,通过融入自注意力机制模块,在遮挡和混淆目标识别上能够获取更高的正确率。

## 参考文献 (References)

- [1] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346–359.
- [2] DUAN G Q, AI H Z, LAO S H. A structural filter approach to human detection [M]//*Computer Vision—ECCV 2010*, 2010: 238–251.
- [3] SHET V D, NEUMANN J, RAMESH V, et al. Bilattice-based logical reasoning for human detection[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 1–8.
- [4] WU B, NEVATIA R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors [C]//*Proceedings of Tenth IEEE International Conference on Computer Vision Volume 1*, 2005: 90–97.
- [5] ZHOU C L, YUAN J S. Learning to integrate occlusion-specific detectors for heavily occluded pedestrian detection[C]//*Proceedings of Asian Conference on Computer Vision*, 2017: 305–320.
- [6] OUYANG W L, WANG X G. Joint deep learning for pedestrian detection[C]//*Proceedings of IEEE International Conference on Computer Vision*, 2013: 2056–2063.
- [7] ZHOU C L, YUAN J S. Multi-label learning of part detectors for heavily occluded pedestrian detection[C]//*Proceedings of IEEE International Conference on Computer Vision*, 2017: 3506–3515.
- [8] WANG X L, XIAO T T, JIANG Y N, et al. Repulsion loss: detecting pedestrians in a crowd[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7774–7783.
- [9] MIAO J X, WU Y, LIU P, et al. Pose-guided feature alignment for occluded person re-identification [C]//*Proceedings of IEEE/CVF International Conference on Computer Vision*, 2019: 542–551.
- [10] ZHOU C, YUAN J. Bi-box regression for pedestrian detection and occlusion estimation [C]//*Proceedings of the European Conference on Computer Vision*, 2018: 135–151.
- [11] FAN D P, JI G P, SUN G, et al. Camouflaged object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 2777–2787.
- [12] WANG Y Q, YAO Q M. Few-shot learning: a survey [EB/OL]. (2019–4–10)[2021–10–10]. <https://arxiv.org/abs/1904.05046v1>.
- [13] ANTONIOU A, STORKEY A, EDWARDS H. Data augmentation generative adversarial networks [EB/OL]. (2017–11–12)[2021–10–10]. <https://arxiv.org/abs/1711.04340>.
- [14] ZHANG R, CHE T, GHAHRAMANI Z, et al. MetaGAN: an adversarial approach to few-shot learning [C]//*Proceedings of 32nd Conference on Neural Information Processing Systems*, 2018.
- [15] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.
- [16] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for few-shot learning [C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1199–1208.
- [17] SI Z Z, ZHU S C. Learning AND-OR templates for object recognition and detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(9): 2189–2205.
- [18] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B. Human-level concept learning through probabilistic program induction [J]. *Science*, 2015, 350(6266): 1332–1338.
- [19] GEORGE D, LEHRACH W, KANSKY K, et al. A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs [J]. *Science*, 2017, 358(6368): eaag2612.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.
- [21] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7794–7803.
- [22] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation [C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3141–3149.
- [23] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//*Proceedings of the 34th International Conference on Machine Learning*, 2017: 1126–1135.
- [24] VINIYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning [J]. *Advances in Neural Information Processing Systems*, 2016: 3637–3645.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.