

文章编号: 1007-2780(2024)06-0813-09

## 遥感图像目标检测的相似目标替换增广算法

孙得耀<sup>1,2</sup>, 朱明<sup>1,2\*</sup>, 王佳荣<sup>1</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

**摘要:** 现今主流遥感图像目标检测算法依赖于深度学习技术的发展。在模型训练时,对训练集数据进行增广处理是增强模型泛化能力的重要方法。当前遥感图像目标检测算法的图像增广方法主要沿用通用的目标检测增广方法,亟需发展针对遥感目标特点的增广方法。本文提出基于遥感图像的相似目标替换增广方法,设计了相似目标替换的增广流程,即通过建立样本库、划分相似类、查询相似样本及替换样本预处理等步骤对数据集中的目标进行相似目标分类。首先,建立所有训练集样本的样本库,统计样本类别信息;其次,对训练集中的目标类别划分相似类;然后,针对数据集中样本数量不均衡的问题,设计了少数样本补偿机制,通过控制采样样本的概率平衡样本数据的训练比例;最后,设计了替换样本预处理机制,针对不同目标类别的特点,使用适宜的变换方法作为样本预处理方法。在 DOTA 数据集上的实验表明,使用相似目标替换增广算法的 DCL 检测算法检测结果的 mAP 值相较于 baseline 提高了 1.34,模型对相似类别的预测准确率上升,对样本数量较少类别的目标检测能力得到提高。

**关键词:** 遥感图像;目标检测;数据增广;相似目标替换

中图分类号: TP391 文献标识码: A doi: 10.37188/CJLCD.2023-0195

## Similar target replacement for remote sensing object detection data augmentation

SUN Deyao<sup>1,2</sup>, ZHU Ming<sup>1,2\*</sup>, WANG Jiarong<sup>1</sup>

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Nowadays remote sensing image object detection algorithms are highly relied on the development of deep learning technology. Data augmentation to dataset images is an important way to enhance model's generalization ability. Current augmentation methods of remote sensing object detection algorithms still use general object detection methods. There is an urge need to develop methods that focus on the properties of remote sensing objects. This paper raised a data augmentation method named Similar Targets Replacement (STR) that based on remote sensing images and designed the STR process. First, the sample library is built to collect statistics of samples' categories. Second, categories of the dataset is divided into several

收稿日期: 2023-05-26; 修订日期: 2023-06-01.

基金项目: 吉林省科技厅重点研发项目 (No. 20210201137GX)

Supported by Science and Technology Department of Jilin Province, China (No. 20210201137GX)

\*通信联系人, E-mail: zhu\_mingca@163.com

similar target categories. Then, to solve the problem of disbalanced sample amounts, minority sample compensation is designed to balance the proportion of samples from different categories by controlling the probability of input samples. Finally, sample replacing preprocess mechanics is designed by using appropriate transforms for each different categories as the preprocess methods. Experiments on DOTA dataset shows the mAP of DCL detection algorithm using STR augmentation raises 1.34 compared to baseline. Model's detection accuracy to similar categories raises and ability to categories with fewer number of samples is strengthened.

**Key words:** remote sensing images; object detection; data augmentation; similar target replacement

## 1 引言

基于深度学习的目标检测算法性能高度依赖数据集的数据质量。深度学习算法的本质是对各种情况下真值样本的拟合,因此训练集中数据的代表性越高,算法可参考的对象就越接近真实的检测场景,算法的泛化能力和鲁棒性越好。目前最先进的基于遥感图像的目标检测算法都以深度学习算法作为基础,遥感图像数据集的预处理成为检测算法性能的重要支撑。在遥感图像数据集的制作过程中,最为耗费时间和精力环节是数据的标注。由于遥感图像背景尺度大、背景复杂、目标多样性高等特点,在标注时需要耗费相当大的时间和精力。基于现有的数据集进行数据增广是一种替代的方法,可以在不引入新数据的情况下,丰富现有样本的信息,提高数据的多样性。现有的遥感图像目标检测数据增广方法大多沿用通用数据集的增广方法,并未对遥感图像目标本身的特点进行针对性的处理,亟需一种基于遥感图像本身特点进行增广的算法。

## 2 现有的遥感图像目标检测样本增广方法

### 2.1 仿射变换法

在几何中,一个向量空间进行一次线性变换并接上一个平移变换得到另一个向量空间的变换称作仿射变换<sup>[1]</sup>。通常图像中的仿射变换包含尺度变换、平移、旋转、翻转等操作。利用图像的仿射变换进行数据增广的方法广泛应用于深度学习领域。

仿射变换可以用公式(1)表示:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (1)$$

其中: $x'$ 、 $y'$ 表示变换后的像素点坐标, $x$ 、 $y$ 为变换前的像素点坐标, $A$ 为仿射变换矩阵。

尺度变换<sup>[2]</sup>是指将原始图片的尺寸以一定的比例进行缩放。尺度变换是基于深度学习的目标检测算法中重点关注的问题。在深度卷积神经网络中,每一次进行最大池化的特征深度提取都使得卷积核的感受野范围扩大,同时降低了尺度敏感性。越深的特征中的向量越不能体现较小尺度的特征。因此,对输入图像进行一定比例的尺度变换可以使网络学习到多种尺度的特征。缩放比例为 $\alpha$ 时,图像缩放的变换矩阵为:

$$A_{\text{scale}} = \begin{bmatrix} \alpha x & 0 & 0 \\ 0 & \alpha y & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

对图片进行平移操作也是常用的数据增广方法<sup>[3]</sup>。卷积运算从理论上是平移等变的,换句话说输入图片平移一定距离也会使卷积运算的结果平移对应的距离。但是在实际的运算过程中,由于浮点数的运算精度有限,在经过若干层的深度网络运算后,平移等变性并不一定能够保持,因此仍然有必要对输入的样本进行平移特征的学习。设水平平移和垂直平移的距离分别为 $dx$ 和 $dy$ ,图像平移的变换矩阵为:

$$A_{\text{shift}} = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

对图片进行随机旋转<sup>[4-5]</sup>是非常重要的—种增广手段。由于CNN的卷积核不具有旋转等变性,一个旋转后的目标输入到网络中所输出的特征网络并不能够识别出来。所以对于旋转后的目标,即使网络之前学习过目标特征,仍然要重

新学习。如果训练样本中目标的旋转角度多样性不足,在验证时旋转目标的效果就会非常不好。在训练时加入充足的旋转目标样本是保证模型对旋转目标检测效果的必要条件。图像旋转变换的矩阵为:

$$A_{\text{rotate}} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

镜像翻转<sup>[2]</sup>是常用的图像增广手段。与旋转变换相似,CNN的卷积核也不具有镜像变换的等变性。在数据增广过程中常用水平或者垂直方向上的镜像翻转提高模型的泛化能力。镜像翻转操作实现简单,只需要对图像像素进行重新排列即可。水平镜像和垂直镜像的变换矩阵分别为:

$$A_{\text{flip},H} = \begin{bmatrix} -x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_{\text{flip},V} = \begin{bmatrix} x & 0 & 0 \\ 0 & -y & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

仿射变换增广方法是对图像目标的几何变换。在通用目标检测任务中,目标尺寸通常占整幅图像的尺寸比例较大,简单的仿射变换可以增加目标的特征多样性而不引入过多的虚假上下文信息。然而在遥感图像中,目标尺寸与整幅图像相比尺寸比例较小,简单的仿射变换所增加的目标多样性有限。

## 2.2 色彩空间变换法

色彩空间变换法通过对图像本身的特征进行变换增加样本的丰富程度。通过对图像的亮度、对比度、饱和度、色调等色彩特征进行各种调制变换,模拟图像中的样本受到光照、遮挡、阴影等的影响。

在RGB空间中进行图像变换,可以通过对每个颜色通道加上一定的随机扰动值来实现<sup>[6]</sup>。这样图像中的样本会产生随机的颜色偏差,从而提高样本的多样性。

HSV空间中的变换是一种非常有效的数据增广手段,应用在各种先进的目标检测算法如YOLO X<sup>[7]</sup>中。HSV的3个表示色彩的维度分别是:色调(Hue)、饱和度(Saturation)、明度(Value)。色度用角度度量,表示不包含亮度在内的颜色的性质。饱和度表示颜色接近光谱色的程度,饱和度越低颜色越接近白光,饱和度越低颜色越接近光谱色。明度表示颜色明亮的程度,通常明度越

低颜色越暗淡,明度越高颜色越明亮。在HSV空间中进行随机变换,可以模拟各种不同颜色的目标、被阴影遮挡的目标,或者在不同光照强度下的目标等。

色彩空间变换法虽然能在一定程度上丰富图像的多样性,但对于某些颜色作为显著特征的目标来说,色彩变换反而使目标失去了真实的特征信息,对模型学习目标特征产生负面效果。

## 2.3 基于目标插值的增广方法

基于目标插值的增广方法近年来取得了非常优秀的效果,基本的实现方式是对图片上的一些图像块进行裁剪,然后在本张图片或者数据集中其他图片选取另外的经过变换的图像块粘贴到被裁剪的区域。

随机裁剪是比较容易实现的基于插值的方法,有Cutout<sup>[8]</sup>、Random Erasing<sup>[9]</sup>等实现方式。Cutout方法是随机选取图像中一块正方形区域,将其中的像素值移除并替换为0。而Random Erasing是选取矩形区域进行擦除,并且在区域中的像素填充随机数或者区域像素均值。

Mixup<sup>[10]</sup>是一种在深度学习领域有着广泛应用的增广方法,其主要思想是从每个批次中随机选取样本,并以一定比例混合生成新的样本。生成混合图像的公式如式(6)所示:

$$\begin{cases} x' = \lambda x_i + (1 - \lambda) x_j \\ p' = \lambda p_i + (1 - \lambda) p_j \end{cases}, \quad (6)$$

其中: $x'$ 、 $p'$ 为混合样本及其对应的标签, $\lambda$ 是服从beta分布的随机数(取值区间为 $[0,1]$ ), $(x_i, p_i)$ 与 $(x_j, p_j)$ 表示随机选取的两个样本及其标签。

Cutmix<sup>[11]</sup>结合Cutout和Mixup的特点,首先随机选择一幅图像中的一块区域,然后将这块区域粘贴到另一幅图像上,计算两幅图像中样本的重叠率,对真值标签进行对应的调制计算。Cutmix克服了Cutout方法中被裁剪掉区域中信息丢失的问题,而且相对于Mixup的全图混合,在生成的混合图像中既具有两幅源图像的混合特征,也具有背景图像的原始特征。

马赛克(Mosaic)增广<sup>[12]</sup>随机选取训练集中4幅图片,分别按照随机的比例缩放,将这4幅图片摆放在一张生成的空白图片的左上、右上、左下、右下位置,并使这4幅图片紧贴在一起,保证

生成的混合图片的中间没有空白区域。图片的缩放比例可以在 $[0.1, 2.0]$ 区间内自由选择,其中小于1的缩放倍数可以增加图片中小目标的数量,大于1的缩放倍数弥补了马赛克算法带来的大样本数量的减少。在训练过程中选择以一定的概率使用马赛克增广,在其余情况下仍然可以使用标准的增广方式如仿射变换法和色彩空间变换法等,进行样本的多样化增广处理。

目前的主流数据增广方式取得了较优的效果。在训练过程中,一般会进行上述几种方法的选择与整合。但是这些方法仍是对于包括背景与目标的原图像整体进行处理,并没有基于局部的目标及目标本身类别特征进行适应性的增广变换。

### 3 相似目标替换增广算法

基于插值的目标检测数据增广技术中,并没有对基于具体目标样本的特性进行增广处理,都是对基于水平包围框的目标进行变换。在遥感图像中,对图像进行水平矩形的剪切、粘贴处理会丢失目标信息或引入冗余的背景信息,因此目标以带有角度信息的斜框表示法为最优。受Cutmix方法启发,本文提出一种基于DOTA数据集<sup>[13]</sup>的带有目标旋转特性的遥感图像目标增广方法相似目标替换(Similar Targets Replacement, STR)。

#### 3.1 相似目标替换算法的流程

为了提高模型对目标纹理及背景变化的鲁棒性,对训练时输入图片中的目标进行相似替换。替换算法的流程如下:

(1) 建立样本库:对所有训练集图像中的有效样本进行统计,按照类别储存样本信息,DOTA数据集中的目标类别如表1所示。

(2) 划分相似类:依据训练集中的目标类别,将目标特征、背景上下文信息类似的类别合并为相似类,作为相似替换的依据。

(3) 查询相似样本:在每一步训练前,准备训练样本。对每一个输入的样本真值,以概率 $p_e$ 选择是否进行后续的样本替换处理。对被选中的目标,在样本库中查询满足相似替换条件的另一个样本作为被选中目标的替换者;而对于未被选中的目标,则直接输入网络进行训练。在查询相似样本时,依据样本真值所在的相似类。另外,在查询相似替换样本时,依据相似类中的各个类别的样本数量,使用少数样本补偿机制调整查询相似类样本的概率,使查询到各个类别样本的概率相等。

(4) 替换样本预处理:将原始目标斜框从图像中擦除,然后根据替换样本的类别对照表2进行一系列仿射变换预处理,最后将相似类别替换样本填充至原始目标的中心位置。

其中 $p_e$ 是替换概率系数,取值不大于0.5。虽然数据增广对于提高样本的多样性有重要作用,但是经过增广的样本毕竟不是真实采集的样本,而原始样本中目标和背景之间的上下文信息在增广过程中会有所损失。为了保留一定的原始目标信息,设定了超参数替换概率系数 $p_e$ 。对于样本库中的样本有 $p_e$ 的概率进行后续的替换增广预处理,而其余的样本则不处理直接以样本的原始形态输入网络训练,以确保有足够多的真实样本和上下文信息输入模型进行训练。

表1 DOTA数据集中的目标类别及其英文缩写

Tab. 1 Categories in DOTA dataset and abbreviations

缩写	PL	HC	TC	SBF	GTF	BC	BD	SH
类别名	飞机	直升机	网球场	足球场	田径场	篮球场	棒球场	船只
缩写	SV	LV	HA	ST	RA	SP	BR	
类别名	小型车	大型车	码头	储油罐	环岛	游泳池	桥梁	

在随机抽选时,已经被分为同一组相似类别的样本相互之间可以替换,如飞机和直升机的样本互相之间可以替换,小型车辆和大型车辆可以

互相替换等。这样在训练时同一组相似类别的上下文信息就可以共享,提高了样本数较少的类别目标的多样性。

### 3.2 相似类的划分

对于相似样本的定义,根据数据集和具体类别的不同,可以进行不同的设置。对于DOTA数据集的15个类别,相似样本的划分如表2所示,其中包括3组相似目标类别,分别是飞机与直升机;网球场、足球场和田径场;小型车辆和大型车辆。划分的依据是:飞机与直升机在遥感图像中的背景相近,一般都停放在机场之上。网球场、足球场和田径场的尺度相似,形状也近似矩形,

背景基本都位于城市区域。篮球场与前面的几类相比,虽然背景相似,但目标的尺度相对较小,因此独自归为一类。棒球场的形状及纹理与之前的几种体育场地都不相同。船只样本绝大多数都位于海上背景。小型车辆和大型车辆可以归为相似的样本,因为车辆的背景基本都位于道路、停车场等地点,与车辆的尺寸无关。码头、储油罐、环岛、游泳池、桥梁等其余类别的目标在纹理和背景上没有近似之处,因此不作为相似类。

表 2 相似替换算法对DOTA数据集中各个类别目标的预处理方法及相似目标类别划分

Tab. 2 Preprocess methods of STR for each category in DOTA dataset and grouping of similar objects

类别	PL HC	TC SBF GTF	BC	BD	SH	SV LV	HA	ST	RA	SP	SBR	LBR
尺度变换	√	√	√	√	√	√	×	×	√	√	√	×
旋转变换	√	√	√	√	×	×	×	×	×	√	×	×
镜像变换	√	√	√	√	√	√	×	×	√	√	×	×

注:√表示采用对应的预处理方法,×代表不采用

### 3.3 少数样本补偿机制

如图1所示,在DOTA数据集中存在样本数量不均衡的问题,有的样本数量级接近 $10^5$ ,而有的样本数量级只有 $10^3$ 。因此在进行相似样本替换时,对于相似类加入了少数样本补偿机制,即对样本数量较少的类以更高的概率选中。在训练过程中有更多的这一类样本,这样不同数量样本类对于网络模型的贡献可以更为接近,改善了分类结果。

少数样本补偿机制的具体设置为:设 $x_i$ 为某

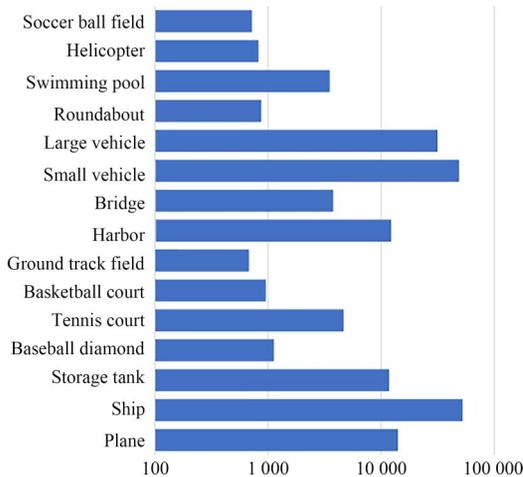


图1 DOTA数据集中各个类别目标的样本数量

Fig. 1 Sample amount in DOTA dataset for each category

一组相似样本中的一类,该类样本数量为 $c_i$ , $N_i$ 为这一组相似样本的类数。选中该类作为替换样本的概率为 $p_i$ 。由上述讨论可知,希望选取到第 $i$ 类样本的概率 $p_i$ 正比于 $1/c_i$ 。如果期望对于每一组相似类中的所有类别的样本选取概率相等,那么对概率 $p_i$ 进行归一化可得:

$$p_i = \frac{1/c_i}{\sum_{i=0}^{N_i-1} 1/c_i} \quad (7)$$

在训练样本预处理时,假设某一相似目标类的总样本数为 $N$ ,那么对于第 $i$ 类相似样本,送入模型训练的样本数为:

$$N'_i = p_i \times c_i = \frac{1}{\sum_{i=0}^{N_i-1} 1/c_i} \quad (8)$$

因此在同一个相似类的每一类所产生的训练样本数相等,这样样本数较少的类别也可以获得与其他类别相等的机会得到训练。

### 3.4 替换样本的预处理

为了进一步增加样本的多样性,使用几种基本变换对替换样本进行预处理:

(1) 尺度变换:对替换样本进行随机比例的缩放,以增加模型对目标尺寸变化的鲁棒性。在本文中缩放比例取 $[0.5, 1.3]$ 区间内的随机值;

(2) 旋转变换:对替换样本施加 $(-90^\circ, 90^\circ]$

随机角度的旋转变换；

(3) 镜像变换:对替换样本以 0.5 的概率施加横向或纵向的镜像变换。

尺度变换的缩放比例值参考了马赛克变换的缩放比例选值[0.5,2.0]。由于DOTA数据集中样本排布相对紧凑,过大的缩放比例会引起样本的相互重叠,且会破坏样本类别的上下文信息,因此选择1.3为缩放比例上限。旋转变换中对目标进行左右各90°范围的随机旋转,如此可以保持目标的头尾不倒转。镜像变换的概率为0.5,这样经过镜像对称处理的样本和原始样本各占1/2。

DOTA数据集中的目标类别跨度较大,对于不同纹理形态特征的目标应当采用不同的预处理变换手段。车辆和船只目标常以密集分布的形式出现,如果对这些目标进行旋转替换,那么很容易和周围的目标发生重叠,干扰深度学习算法的特征提取效果。同时车辆、船只的目标在数据集中的样本数量较多,其多样性有保证,不需要额外增加旋转样本数量。因此对于车辆和船只,不采取旋转变换预处理。

对于如储油罐、环岛等圆形的目标类别,旋转变换不能改变目标的纹理特征,但是会改变标注框的信息,不仅对训练效果无帮助,而且会对角度分类预测带来干扰,因此对这些类别不采用旋转变换进行处理。储油罐在遥感图像中常密集分布,采用尺度变换容易造成目标重叠,反而使模型难以辨认目标的纹理;环岛目标经过缩放之后,与背景中道路的连接变得不自然,破坏了目标的上下文信息,因此也不能进行尺度变换。最后,由于这些目标形状为圆形,镜像变换并不会增加目标特征的丰富度,因此也不采用。

对于码头目标来说,其背景一边为陆地,另一边为水域,采用旋转变换会破坏背景与目标之间的上下文关系,因此不使用旋转变换,而只使

用尺度和镜像变换进行目标增广。

对桥梁这种具有大宽高比且尺度跨度范围大的类别,需要谨慎选择预处理手段。小尺度的桥梁和大尺度的桥梁,其图像特征具有很大差别。因此将桥梁样本作为两种不同的类别进行处理,其区分依据为宽高比和样本尺寸。规定当桥梁样本的宽高比( $w/h$ ) $>3$ 或者样本的面积( $w \times h$ ) $>(60 \times 60)$ 时,样本归类于长桥梁(LBR),否则归类于短桥梁(SBR)。对于长桥梁,由于样本尺寸较大,模型预测较为容易,因此仿射变换预处理的意义并不大,故不使用仿射变换作为预处理手段。对于短桥梁,有限度的仿射变换有利于提高数据集中的样本丰富度,因此预处理手段适用于这类目标。

最后,由于形状比较规则,对于网球场、篮球场、棒球场等目标使用几何变换进行增广并不会产生负面作用,因此几种几何变换都可以应用于这些目标上。在相似目标替换方法中各种类别的目标所对应的预处理方法如表2所示。

对样本进行的预处理变换与对整幅图片的变换效果不同。对样本周围区域进行变换,样本与背景之间的位置关系发生了改变,样本多样性相比对整幅图片进行变换更加丰富。

## 4 实验与结果

### 4.1 实验设定

本文采用基于角度分类法预测遥感图像中的斜框目标检测算法DCL<sup>[14]</sup>在DOTA数据集中进行数据集增广效果的验证。DCL算法对遥感目标的斜框预测可以更准确地定位目标位置和姿态,较好地验证数据增广的效果。

DOTA数据集<sup>[13]</sup>的图像来自于Google Earth等多个不同的平台,其中包含了采集自多种传感器、多种分辨率的遥感图像。DOTA数据集中包含了2806张图像,其中的目标种类具有很大的代表性。图像的宽高尺寸从2k到4k不等。数据集共有15类目标,目标类别及其简称分别为:飞机-PL、棒球场-BD、桥梁-BG、田径场-GTF、小型车辆-SV、大型车辆-LV、船只-SH、网球场-TC、篮球场-BC、储油罐-ST、足球场-SBF、环岛-RA、码头-HA、游泳池-SP、直升机-HC等。DOTA中全部标注的目标有188282个样本,样本以有序



图2 相似目标替换后的输入样本示例

Fig. 2 Input samples of similar target replacement

四边形法标注。训练集、测试集、验证集比例分别为原始数据集的1/2、1/3、1/6。

在本文的实验中,将图片分割为 $600 \times 600$ 大小的图像块,同时为了不丢失图像边缘的样本,设置重叠区域为150像素,最后缩放到 $800 \times 800$ 的大小。在测试集中使用原始大小的图像,以模拟验证集的检测条件。由此获得的20 673张训练图片和6 341张测试图片分别作为训练集和测试集。实验硬件平台为基于2张Nvidia Titan XP显卡的服务器,每批次训练2张图片,每步共训练4张图片。超参数设置中,动量为0.9,学习

率设置为0.0005。上述超参数使用的是DCL<sup>[13]</sup>算法的超参数设置,保证训练模型能够正常产出预测模型。在主要实验中,相似目标替换概率系数 $p_s$ 取0.3。本文分别使用无增广方法、马赛克增广方法、局部马赛克增广方法和相似目标替换增广方法实验验证各个增广算法的效果(图3)。其中,无增广方法在网络训练时对输入样本不进行增广,马赛克增广对所有输入样本都进行马赛克增广处理,局部马赛克增广将20%的输入样本进行马赛克增广预处理,相似目标替换增广使用上一节介绍的算法进行增广处理。

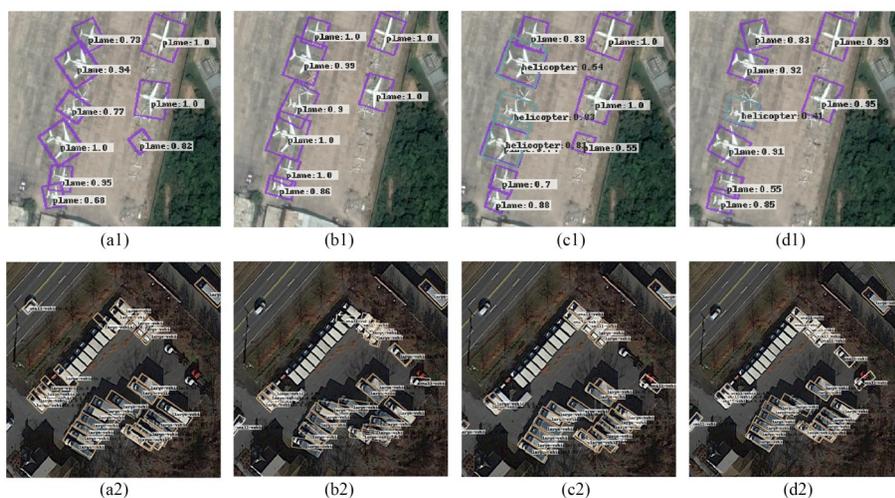


图3 使用DCL算法<sup>[13]</sup>在DOTA上应用无增广(a)、马赛克增广(b)、局部马赛克增广(c)和相似目标替换增广(d)的实验结果示例。

Fig. 3 Examples of experiment results of DCL algorithm<sup>[13]</sup> on DOTA applying no augmentation (a), mosaic augmentation (b), partial mosaic augmentation (c) and similar target replacement (d).

## 4.2 实验结果

如表3所示,相似目标替换法在大多数类别上的检测效果都超过无增广、马赛克增广和局部马赛克增广的方法。在飞机、小型车辆、大型车辆、船只上,几种算法的性能差距不大,这是因为这些类别在数据集中有足够多的样本,基准预测准确率较高,增广效果提升有限。相似目标替换法相对于baseline,主要的性能优势体现在几组相似目标中,如在网球场、足球场、田径场这一组相似目标,其增广效果均有所提高。对于桥梁目标,相似目标替换法的性能优势体现了划分大尺度桥梁和小尺度桥梁进行不同预处理手段的重要性。

马赛克增广法的性能相对无增广和相似目

标替换增广并没有体现出太大优势,这主要是因为DOTA数据集中大尺寸的目标较多,只使用马赛克增广处理训练样本将会使大尺寸目标变小,使网络丢失部分目标特征信息。局部马赛克增广的训练数据是由80%不经过预训练的样本数据和20%的马赛克增广样本组成,这样模型保留了模型的样本尺度特征信息,同时增加了对目标上下文特征提取能力,训练效果强于全部样本使用马赛克增广的效果,但仍然比相似目标替换法的mAP低了0.89。

对于3组相似样本,相较于baseline,在直升机和飞机类别中,两种马赛克增广方法对直升机的提升效果有限,而相似目标替换将直升机的AP提高了5.87。由于直升机的样本数相对飞机较少,

表 3 DCL 算法<sup>[13]</sup>在 DOTA 数据集上分别使用无增广(1)、马赛克增广(2)、局部马赛克增广(3)和相似目标替换增广(4)的结果

Tab. 3 Result of DCL<sup>[13]</sup> using no augmentation (1), mosaic augmentation (2), mix-mosaic augmentation (3) and STR (4) in DOTA dataset

算法	PL	HC	TC	SBF	GTF	BC	BD	SH	SV	LV	HA	ST	RA	SP	BR	mAP50
1	88.52	56.27	90.78	55.60	65.24	79.13	75.20	74.10	67.20	56.02	53.82	77.95	61.90	66.18	39.78	67.02
2	88.77	57.18	90.10	51.81	64.17	74.57	73.05	74.53	68.19	58.37	58.86	76.17	60.25	61.34	38.24	66.37
3	89.01	58.29	90.02	53.15	66.07	76.15	74.28	74.37	68.45	59.63	59.19	77.81	61.87	63.52	40.35	67.47
4	89.03	62.14	89.35	57.16	68.02	78.40	74.19	75.58	66.08	61.15	54.13	78.02	62.55	68.41	41.12	68.36

实验结果说明了模型使用相似目标替换对少数样本类别的检测效果有明显的提升。在网球场、足球场和田径场这组相似样本中,这3类目标的样本数量相近,相似目标替换对这3种目标的检测效果均有提升。对于小型车辆、大型车辆这组相似样本,其中小型车辆的 AP 相较于 baseline、马赛克增广和混合马赛克增广分别下降了 1.12、2.11 和 2.37,大型车辆的 AP 反而分别上升了 5.13、2.78 和 1.52,原因是小型车辆的密度较大,目标剪切容易影响到目标的上下文信息;而大型车辆受到的影响相对较小,增广带来的多样性对模型性能的提高抵消了上下文信息缺失带来的损失。

总体来看,相似目标替换法中提出的几种改进在实验结果中都有所体现。首先是相似目标中检测结果都有所提高;其次样本数较少的目标得到了更多的训练机会,检测效果也更好;其三对一些类别的选择性预处理方式规避了目标的互相干扰和训练难度的增加,保证了这些类别的检测效果不出现下滑。

#### 4.3 消融实验

为了研究替换概率系数  $p_e$  的取值对数据集

增广效果的影响,设计了消融实验。实验设置与上一节中的主要实验相同,只是在参数上改变替换概率系数  $p_e$  的值,分别取 0、0.3 和 0.5。 $p_e$  取值过低,相似目标替换的样本数就会过少,对训练结果的影响与不增广相差不大; $p_e$  取值过高,则会在网络模型中引入过多的生成信息,稀释了原始样本中目标与背景的上下文信息。为了验证  $p_e$  的合适取值,选取 0.3 和 0.5 进行消融实验,使用 DCL 算法在 DOTA 上使用不同的  $p_e$  进行实验验证相似目标替换增广的算法性能。

由表 4 可以看出, $p_e$  取 0.3 与 0.5 时,对几组相似目标的预测结果基本上都高于 baseline。 $p_e$  取 0.5 时,个别类别如田径场、棒球场、桥梁等分类比  $p_e$  取 0.3 时检测结果好。不过,在篮球场、码头、储油罐、环岛、游泳池等类别中,检测结果相较  $p_e$  取 0.3 时有所下降。对实验结果分析认为, $p_e$  取值过高会减少真实目标数目,丢失真实目标与背景之间的特征关系,算法模型学习到的“假目标”更多,降低了对真实场景的判断能力。因此  $p_e$  应适当选取。实验证明  $p_e$  取 0.3 时模型的训练结果优于 0.5。

表 4 相似目标替换算法取不同替换概率系数  $p_e$  的结果比较

Tab. 4 Results comparison of STR with different  $p_e$

$p_e$	PL	HC	TC	SBF	GTF	BC	BD	SH	SV	LV	HA	ST	RA	SP	BR	mAP50
0	88.52	56.27	90.78	55.60	65.24	79.13	75.20	74.10	67.20	56.02	53.82	77.95	61.90	66.18	39.78	67.02
0.3	89.03	62.14	89.35	57.16	68.02	78.40	74.19	75.58	66.08	61.15	54.13	78.02	62.55	68.41	41.12	68.36
0.5	88.95	60.51	90.53	55.33	68.46	76.86	75.65	74.01	66.48	60.28	52.77	76.53	61.10	64.98	41.51	67.60

## 5 结 论

本文根据遥感图像目标的特点提出了一种相似目标替换增广方法。通过随机选取目标替

换为相似类别的其他目标,丰富了背景和前景目标的组合。使用少数样本补偿机制,增加了数据集中数量较少的样本数量。对替换目标进行的多仿射变换增加了样本形态的多样性。经过

实验验证,DCL算法使用相似目标替换增广方法,在替换概率系数取0.3时效果最好,相较于局部马赛克增广方法的mAP50提高了0.89,相较于baseline提高了1.34。模型对相似类别的预测

准确率上升,并且对样本数量较少的目标类别检测能力有所提高。对于不同类别的针对性预处理较少地引入了虚假的上下文信息,丰富了样本的多样性。

## 参 考 文 献:

- [1] STEARNS C C, KANNAPPAN K. Method for 2-D affine transformation of images: US, 5475803 [P]. 1995-12-12.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//*Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe: NIPS, 2012: 1106-1114.
- [3] HOWARD A G. Some improvements on deep convolutional neural network based image classification [C]. 2nd International Conference on Learning Representations. Banff: ICLR, 2014.
- [4] DIELEMAN S, WILLETT K W, DAMBRE J. Rotation-invariant convolutional neural networks for galaxy morphology prediction [J]. *Monthly Notices of the Royal Astronomical Society*, 2015, 450(2): 1441-1459.
- [5] SHARIF RAZAVIAN A, AZIZPOUR H, SULLIVAN J, *et al*. CNN features off-the-shelf: an astounding baseline for recognition [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Columbus: IEEE, 2014: 512-519.
- [6] WU R, YAN S G, SHAN Y, *et al*. Deep image: scaling up image recognition [J/OL]. *arXiv*, 2015: 1501.02876.
- [7] GE Z, LIU S T, WANG F, *et al*. YOLOX: exceeding yolo series in 2021 [J/OL]. *arXiv*, 2021: 2107.08430.
- [8] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout [J/OL]. *arXiv*, 2017: 1708.04552.
- [9] ZHONG Z, ZHENG L, KANG G L, *et al*. Random erasing data augmentation [C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020: 13001-13008.
- [10] ZHANG H Y, CISSE M, DAUPHIN Y N, *et al*. Mixup: beyond empirical risk minimization [C]. 6th International Conference on Learning Representations. Vancouver: ICLR, 2018.
- [11] YUN S, HAN D, CHUN S, *et al*. CutMix: regularization strategy to train strong classifiers with localizable features [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 6022-6031.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J/OL]. *arXiv*, 2020: 2004.10934.
- [13] XIA G S, BAI X, DING J, *et al*. DOTA: a large-scale dataset for object detection in aerial images [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 3974-3983.
- [14] YANG X, HOU L P, ZHOU Y, *et al*. Dense label encoding for boundary discontinuity free rotation detection [C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 15814-15824.

## 作者简介:



孙得耀,男,硕士研究生,2014年于哈尔滨工业大学获得学士学位,主要从事机器视觉与人工智能方面的研究。  
E-mail:576923447@qq.com



朱明,男,硕士,研究员,1991年于中国科学院长春光学精密机械与物理研究所获得硕士学位,主要从事成像目标检测与跟踪、图像处理基础理论以及计算机视觉方面的研究。E-mail: zhu\_mingca@163.com