Modeling and Simulation 建模与仿真, 2023, 12(3), 3224-3237
Published Online May 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.123297

Hans汉斯

基于超分辨率重建的小目标检测算法研究

苏继贤[1,2]

[1]兰州交通大学光电技术与智能控制教育部重点实验室，甘肃 兰州
[2]兰州交通大学国家绿色镀膜技术与装备工程技术研究中心，甘肃 兰州

收稿日期：2023年3月29日；录用日期：2023年5月24日；发布日期：2023年5月31日

摘　要

近年来，随着深度神经网络的不断发展，使得目标检测算法对大型目标以及中型目标的检测已经具有较高的准确率，然而由于小目标在图像中面积占比较少，像素较低以及检测网络可利用特征较少等原因，导致小目标的检测存在严重的分类不准确和定位不精确的情况。为解决上述问题，本文将基于生成对抗网络的超分辨率重建技术和SPD-Conv模块融合到YOLOv5目标检测网络中。实验表明，VisDrone2019数据集上对比原始YOLOv5网络mAP@0.5提升了3.73个百分点。最后经过消融实验证明本文提出的两个模块对小目标检测效果均有一定提升。

关键词

目标检测，SPD-Conv，生成对抗网络，超分辨率重建

Study on Small Target Detection Algorithm Based on Super-Resolution Reconstruction

Jixian Su[1,2]

[1]Key Laboratory of Optoelectronic Technology and Intelligent Control of Ministry of Education, Lanzhou Jiaotong University, Lanzhou Gansu
[2]National Engineering Research Center for Technology and Equipment Technology of Environmental Deposition, Lanzhou Jiaotong University, Lanzhou Gansu

Received: Mar. 29th, 2023; accepted: May 24th, 2023; published: May 31st, 2023

Abstract

In recent years, with the continuous development of deep neural networks, target detection algorithms have achieved high accuracy in detecting large and medium-sized targets. However, due to

文章引用: 苏继贤. 基于超分辨率重建的小目标检测算法研究[J]. 建模与仿真, 2023, 12(3): 3224-3237.
DOI: 10.12677/mos.2023.123297

the relatively small area of small targets in the image, low pixels, and less available features of the detection network, there are serious situations of inaccurate classification and inaccurate positioning in the detection of small targets. In order to solve the above problems, this paper integrates the super-resolution reconstruction technology based on generating confrontation networks and the SPD-Conv module into the YOLOv5 target detection network. Experiments have shown that compared to the original YOLOv5 network, there is an improvement of 3.73 percent in mAP@0.5 on the VisDrone2019 dataset. Finally, the ablation experiment proves that the three modules proposed in this paper have a certain improvement in the detection effect of small targets.

Keywords

Target Detection, SPD-Conv, Generate Adversarial Networks, Super-Resolution Reconstruction

1. 引言

人们通过眼睛获取信息，大脑对信息进行处理从而得出结论。与人类处理数据的方式类似，计算机模拟生物视觉，通过大量的图像或者视频来学习包含在图像或者视频里的信息从而能够感知环境，来解决相应的问题。计算机视觉有图像的分类[1]、检测[2]、分割[3]这三大任务。小目标的检测一直是目标检测问题中一个难以解决的困难点，该技术的目的是准确地检测出图像中可视化特征信息非常少的小目标，通常指 32*32 像素以下的目标。与常规尺寸的目标物体相比，小目标在目标检测任务中常常存在几个困难，比如小目标可利用特征少、对定位精度要求高、数据集中小目标物体少、目标检测网络未针对小目标物体优化等问题。2012 年卷积神经网络开始兴起，目标检测任务进入快速发展期。基于卷积神经网络的目标检测算法主要有两条发展方向既：无锚框和基于锚框的方法，其中基于锚框的方法又包含了 One-stage (一阶段检测算法)和 Two-stage (二阶段检测算法)。两阶段算法的开端是由 Girshick 提出的 R-CNN [4]算法，它的主要思想是使用卷积神经网络(CNN)来提取图像中的特征，然后使用选择性搜索算法来提取图像中的候选区域，最后使用 CNN 来分类和定位这些候选区域。此后在 R-CNN 的基础上 Girshick 又提出 Fast-RCNN [5]算法。Fast-RCNN 只需要在原始图像上提取一次特征，从而大大缩短了训练和推理时间，降低了计算量。一阶段算法的提出则彻底解决了两阶段算法很难解决实时性这个问题从而真正的实现了端到端检测的目的。一阶段算法目前主要包括 YOLO (You Only Look Once) [6]算法和 SSD (Single Shot Multibox Detector) [7]算法，此后的算法大部分是在这两个算法的基础上衍生出来的。随着生成对抗网络(Generative Adversarial Network, GAN) [8]的兴起，使用 GAN 对特征图进行上采样，扩大特征图中小目标的尺寸，以此来提升检测网络对小目标母体的检测精度成为研究热点。Bai [9]与 Rabbi [10]等人分别提出了 SOD-MTGAN 模型和在遥感小目标检测领域使用超分辨率重建技术，针对小目标的边缘进行增强。此后，越来越多的学者投入到使用超分辨率重建技术来提升小目标检的研究中。

本文在上述问题的基础上，提出了使用基于生成对抗网络的超分辨率算法以及 SPD-Conv 模块来改进 YOLOv5 目标检测模型。实验表明，在 VisDrone2019 数据集上对比原始 YOLOv5 网络 mAP@0.5 提升了 3.73 个百分点。最后经过消融实验证明本文提出的两个模块对小目标检测效果均有一定提升。

2. 改进 ESRGAN 超分辨率模型

2.1. 改进 RRDB 基本块

为获得更好的感知质量和主观判断更好的超分图像，Wang 等人在 2018 提出了 ESRGAN [11]。算法。ESRGAN 算法使用 RRDB 作为基本块，如图 1 所示，它结合了多级残差网络和密集连接，与 SRGAN [12]。超分辨率模型相比，RRDB 的基本块具有更深的结构和更复杂的残差连接方式。此外，ESRGAN 在残差值添加到主路径之前乘以一个[0，1]的常数来缩小残差值，以防止不稳定。通过观察发现，当初始参数方差变小时，残差结构会变得更加容易训练。然而在处理极低分辨率的小图像时 ESRGAN 网络会出现模糊现象，同时 ESRGAN 算法也存在平滑图像细节丢失的问题，在处理复杂场景图片，诸如具有复杂色彩和复杂边缘信息的图像时会出现纹理细节丢失的现象。为此本文提出了一种新的网络架构，用一个新的基本块来代替 ESRGAN 使用的基本块，此外，向生成器网络中引入高斯噪声，利用随机变化来产生纹理细节更加真实的图像。

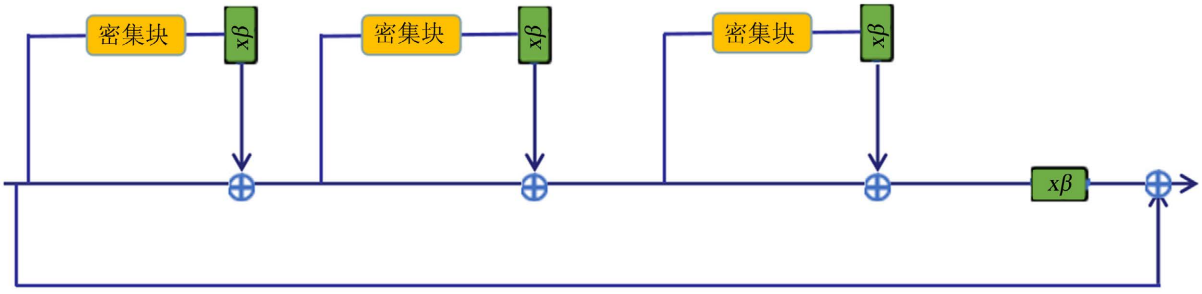
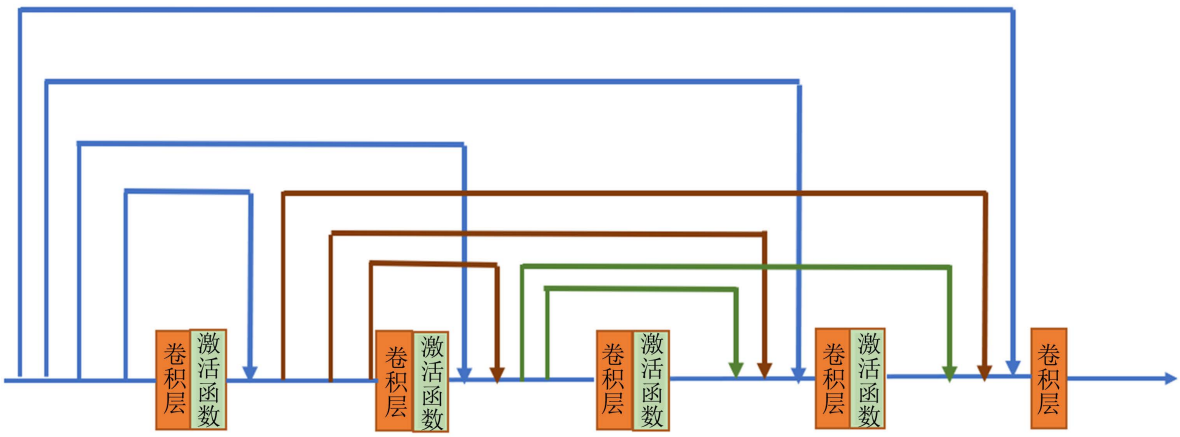
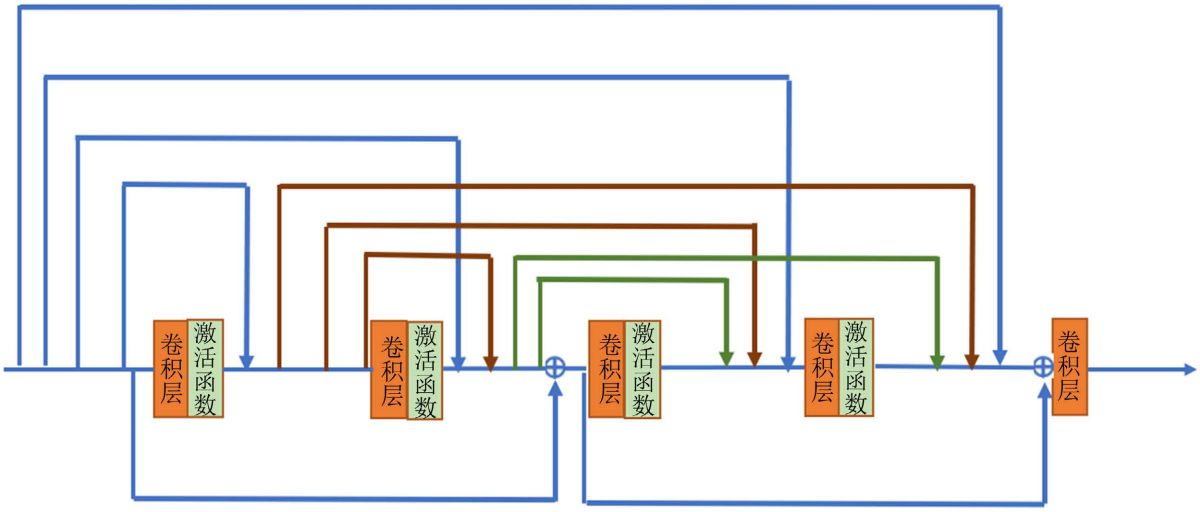


Figure 1. RRDB basic block structure diagram
图 1. RRDB 基本块结构图

原密集块 RRDB 结构如图 2(a)所示 RRDB 块融合了多级残差网络的思想也融合了密集连接的思想。在图 2(a)提供的密集块基础上添加额外的残差学习层，在不增加复杂性的同时增加模型的理解力，以达到更好的鲁棒性，然后在每个密集块的每两层中再添加一个残差。使用改进 RRDB 块的新模块生成的图像，在视觉质量方面要比原密集块的视觉质量高很多。在 Chen [13]等人提出的论文中也正印证了这样的观点：残差网络可以重复使用特征，密集网络可以找到新的特征。因此这种改进架构可以充分利用和探索特征，从而产生主观视觉质量更高的图像，改进后的网络结构如图 2(b)所示。



(a) RRDB 基本块



(b) 改进 RRDB 基本块

Figure 2. Improved RRDB basic block structure diagram
图 2. 改进后 RRDB 基本块结构图

2.2. 引入高斯噪声

往生成器中加入噪声常被应用于生成人脸的网络中，可见加入噪声对图像纹理细节的提升有很大的帮助。为了产生纹理更加真实的图像，将高斯噪声添加到每个密集残差块的输出中，其网络结构如图 3 所示。

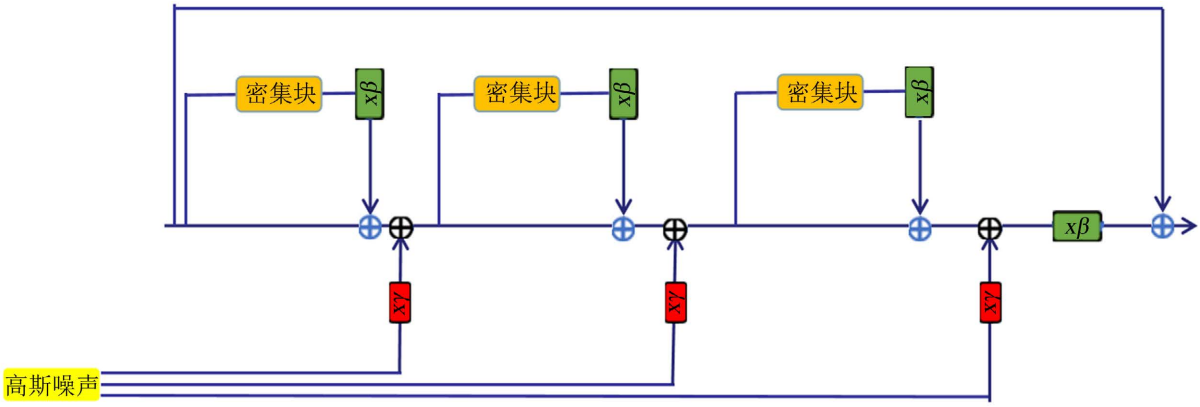


Figure 3. Network structure after adding Gaussian noise
图 3. 添加高斯噪声后的网络结构

2.3. 损失函数设计

本文判别器的设计使用相对判别的方法，判别器需要同时输入一组图像，其任务变成了相对其中一张图像另一张图像为真实的概率。相对判别器的输入是两张图像，公式表示为：

$$D_{Ra}\left(x_r, x_f\right)=\sigma\left(C\left(x_r\right)-\mathbb{E}_{x_f}\left[C\left(x_f\right)\right]\right) \tag{1}$$

其中 $x_r \sim P_{HR}, x_f \sim P_{SR}$。$C(.)$ 表示的是判别器未做 sigmoid 前的输出。$E_x(\cdot)$ 表示的是某一分布在判别器输出的期望。

则生成器的对抗损失为：

$$L_G^{Ra} = -\mathbb{E}_{x_r}\left[\log\left(1-D_{Ra}\left(x_r, x_f\right)\right)\right] - \mathbb{E}_{x_f}\left[\log\left(D_{Ra}\left(x_f, x_r\right)\right)\right] \tag{2}$$

由生成器损失可知，真实图像和生成图像都能够为生成器贡献损失，因此生成器可以从两者中获益。采用此方法可以训练生成器，生成纹理细节更丰富的图像。

使用感知损失的目的是在特征空间中，约束生成的图像尽可能和真实图像保持一致，这样得到的结果不像像素空间那么模糊。在 SRGAN 中使用 VGG 特征层激活函数后的特征图做约束。ESRGAN 指出激活后的特征是很稀疏的，监督能力很小。以 VGG19 网络为例，在第 5 个池化层前的第 4 个卷积层(VGG19-54)后激活函数后的信息仅有 11.17%，以该结果做监督，监督力度很弱，这致使网络效果不佳。因此本文采用 ESRGAN 中的方法，使用激活函数前的特征计算损失，依次训练网络，这样监督信息不是稀疏的，监督力度更强。VGG 损失的定义为，生成图像 $G_{\theta_G}\left(I^{LR}\right)$ 的特征与参考图像 I^{HR} 特征之间的欧氏距离：

$$L_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\varnothing_{i,j}\left(I^{HR}\right)_{x,y} - \varnothing_{i,j}\left(G_{\theta_G}\left(I^{LR}\right)\right)_{x,y}\right)^2 \tag{3}$$

式中，$W_{i,j}$，$H_{i,j}$ 为 VGG 网络中各个特征映射的维度。

综上，生成器的总损失为：

$$L_G = L_{VGG/i,j} + \lambda L_G^{Ra} + \eta L_1 \tag{4}$$

其中 $L_1 = E_{x_i}\left\|G\left(x_i\right) - y\right\|_1$ 为生成图像与参考图像之间的 L_1 损失。

2.4. 优化器设置与实验环境.

本文选用 Adam 优化器，其中 β_1 设置为 0.9，β_2 设置为 0.999。在网络收敛前，交替训练生成器 G 和判别器 D，以此来更新两个网络的参数。本文实验在阿里云平台进行，实验的深度学习环境和使用的框架如表 1 所示。

Table 1. Experimental environments
表 1. 实验环境

配置名称	配置参数
操作系统	Linux Ubuntu 18.04
CPU	Intel(r) Xeon(R) Platinum 8163
GPU	NVIDIA Tesla V100 32 GB
内存	100 GB
软件	Anaconda、Visual Studio Code
深度学习框架	Pytorch1.10.2
GPU 加速库	CUDA11.3

2.5. 实验结果分析

本文将优化的最终模型在 BSDS [14]数据集上与当前流行的超分模型(包括 SRGAN 和 ESRGAN)进行对比，来证明本文提出的改进 ESRGAN 中 RRDB 模块的残差连接以及引入相对判别器对超分辨率重建任务是相当有益的。为了展示本文所提出的网络的超分辨率细节纹理的生成能力。本文从 BSDS 数据集

的原图中裁剪出一块 96 × 96 的子图像，利用双三次插值法对子图像进行 4 倍下采样处理得到低分辨率图像。本文从生成结果中选取了几组具有代表性的结果进行展示。图 4(a)是测试数据集上的原始图像，图 4(b)的第一列是从测试集的原始图像中裁剪出的 96 × 96 的子图像，第二列是使用双三次插值法对子图像进行 4 倍下采样处理得到低分辨率图像。第三列是使用 ESRGAN 算法对低分辨率图像超分后的结果。第四列是使用本文所提出的算法对低分辨率图像超分后的结果。从本次的实验结果中可以看出，本文提出的改进 ESRGAN 算法比原始的 ESRGAN 算法生成的图像更加清晰，生成边缘和纹理细节更加丰富，从而可以证明加入密集残差块和引入对抗损失对超分辨率重建任务是相当有益的。

(a) 数据集中的三张原始图片

HR　　LR　　ESRGAN　　改进 ESRGAN

(b) 实验生成效果对比

Figure 4. Comparison between the improved ESRGAN algorithm and the original ESRGAN algorithm
图 4. ESRGAN 的改进算法与原始 ESRGAN 算法效果对比图

3. 基于超分辨率重建的 YOLOv5 改进算法

3.1. SPD-Conv

YOLO 系列目标检测模型对原图进行了缩放，检测网络需要由特征层使用分类和回归操作进行下采

样，下采样后小目标特征的感受野需要映射回原图，然而经过上述操作后，映射到原图上的小目标在原图中的尺寸可能小于原本的尺寸，造成检测效果变差。小目标往往更依赖浅层特征，因为浅层特征有更高的分辨率，其次对于深度卷积网络，在深度的特征图提取过程中小目标信息可能已经丢失。在常见的目标检测模型中需要使用主干网络对目标图像进行下采样处理，这种下采样处理会使面积较小的目标在特征图上的映射只有几个像素点的大小，因此分类器很难对目标进行分类。受大目标影响，检测网络会将重心转移到对大目标的检测上导致网络对小目标的检测效果变差。

深度卷积体系结构中有一个对于小目标来说的设计缺陷，既 CNN 架构中常采用跨步卷积或池化层，这导致了细粒度信息的丢失和较低效率的特征表示的学习。为此在 2022 年 8 月 7 日来自于 Missouri 大学的 Raja Sunkara and Tie Luo 提出了一种新的 CNN 模块，称为 SPD-Conv [15]。SPD-Conv 完全替代(从而消除)卷积步长和池化层。SPD-conv 是一个空间到深度层，后面跟着一个无步长卷积层对特征映射 X 进行下采样，但保留了通道维度中的所有信息，因此没有信息丢失。

3.2. 空间到深度

如图 5，SPD-conv 由一个空间到深度层和一个非跨步卷积层组成。考虑任何大小为 $S\times S\times C_1$ 的中间特征图 X，切出一系列子图为：

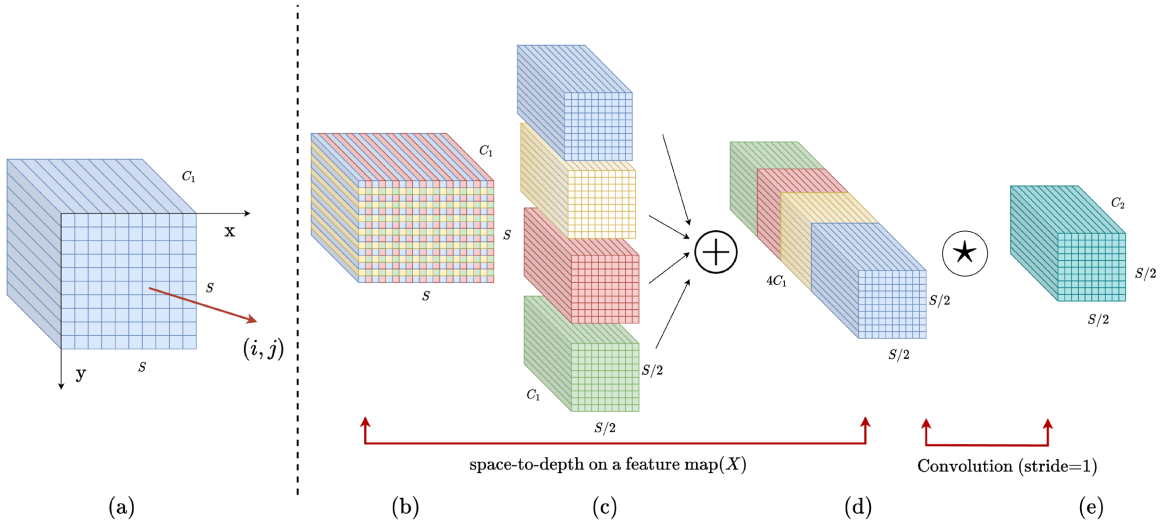


Figure 5. SPD structure [15]
图 5. SPD 结构[15]

$$f_{0,0}=X\left[0:S:scale,0:S:scale\right],f_{1,0}=X\left[1:S:scale,0:S:scale\right],\cdots, \tag{5}$$

$$f_{scale-1,0}=X\left[scale-1:S:scale,0:S:scale\right]; \tag{6}$$

$$f_{0,1}=X\left[0:S:scale,1:S:scale\right],f_{1,1},\cdots, \tag{7}$$

$$\dot{f}_{scale-1,1}=X\left[scale-1:S:scale,1:S:scale\right]; \tag{8}$$

$$f_{0,scale-1}=X\left[0:S:scale,scale-1:S:scale\right],f_{1,scale-1},\cdots, \tag{9}$$

$$f_{scale-1,scale-1}=X\left[scale-1:S:scale,scale-1:S:scale\right]. \tag{10}$$

一般来说，给定任何(原始)特征映射 X，子映射 $f_{x,y}$ 由所有特征映射组成的特征图 $X\left(i,j\right)$，$i+x$ 和

$j+y$ 可以按比例因子对 X 进行下采样。如图 5 给出了比例因为为 2 时的例子，得到 4 个子图 $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ 每个都具有形状 $(S/2, S/2, C_1)$ 并将 X 下采样 2 倍。接下来，沿着通道维度连接这些子特征图，从而获得一个特征图 X'，它的空间维度减少了一个比例因子，通道维度增加了一个比例因子。换句话说，SPD 将特征图 $X(S, S, C_1)$ 转换为中间特征图 $X'(S/scale, S/sacle, scale^2 C_1)$。

3.3. 将 SPD 模块嵌入到 YOLOv5 中

基于 SPD-Conv 的特性和 YOLOv5 目标检测的网络结构，只需要用 SPD-Conv 结构替换 YOLOv5 中随同 *stride*-2 的卷积层。在 YOLOv5 的主干网络中，使用了五个步长为 2 的卷积层来将特征图缩小 32 倍。YOLOv5 目标检测的网络在颈部使用了 2 个 *stride*-2 的卷积层。综上将 SPD-Conv 模块应用到 YOLOv5 中需要替换 7 处，分别为主干网络 5 处，颈部 2 处。YOLOv5 颈部的每个跨步卷积之后都有一个连接层，卷积层将得到保留，全连接层将保持在 SPD 和 Conv 之间。改进后的 SPD-YOLOv5 主干网络架构如图 6(a)所示，颈部网络如图 6(b)所示。

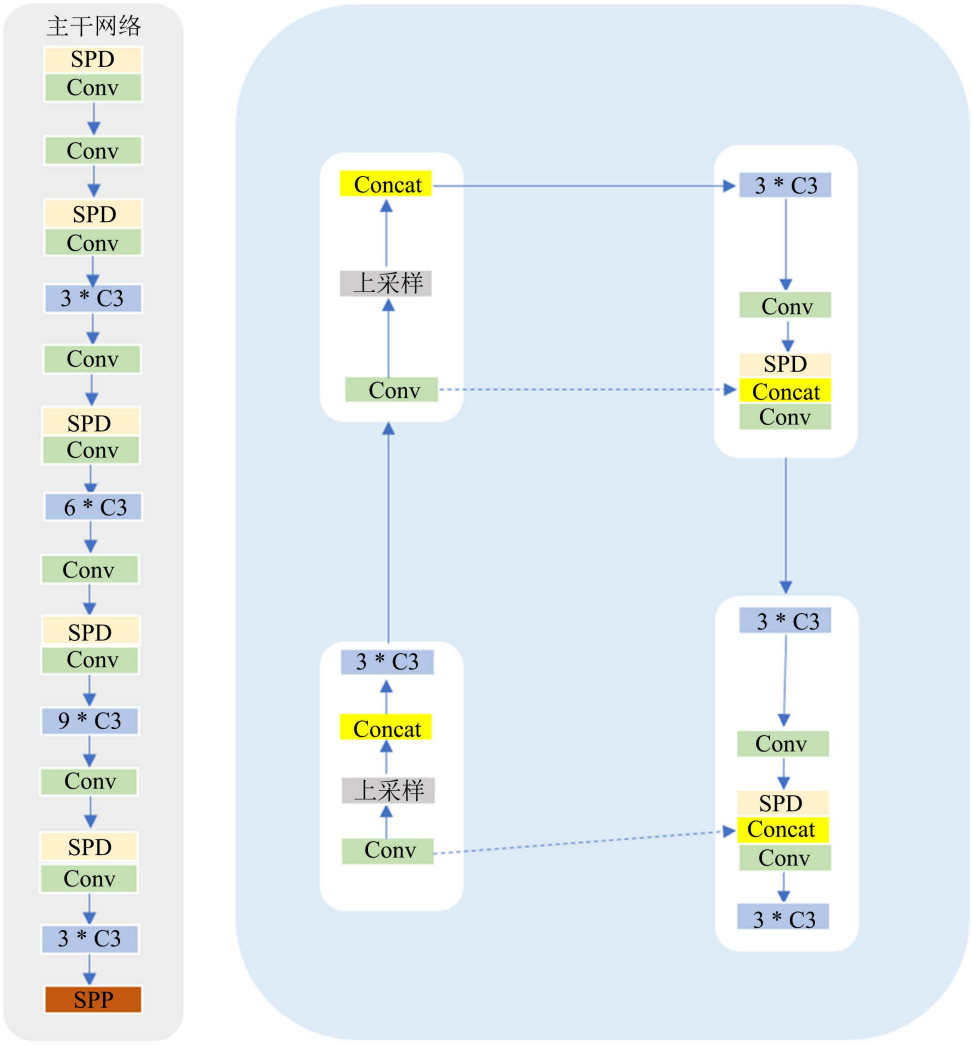


Figure 6. SPD Structure diagram of SPD embedded in YOLOv5
图 6. SPD 嵌入到 YOLOv5 中的结构图

3.4. 基于超分辨率重建的目标检测模型

本文将改进版 ESRGAN 超分辨率网络嵌入到 SPD-YOLOv5 中对图像进行上采样，丰富特征信息，以提高小目标检测任务的精确度。在本文中，为了减少模型的计算量，只对 YOLOv5 主干网络提取的特征图进行超分辨率重建。其基本网络结构如图 7 所示。

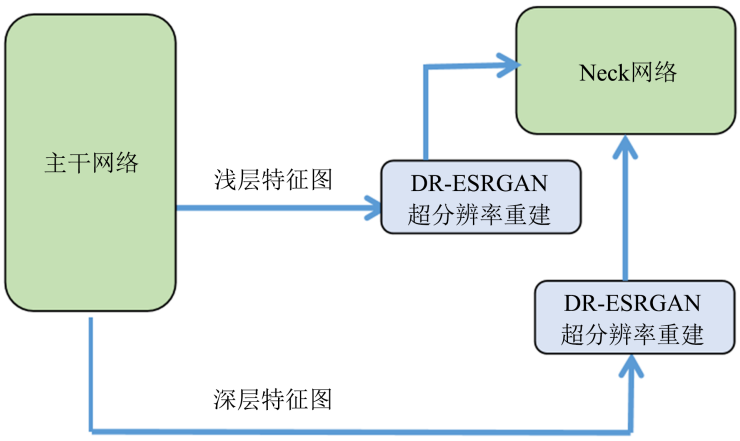
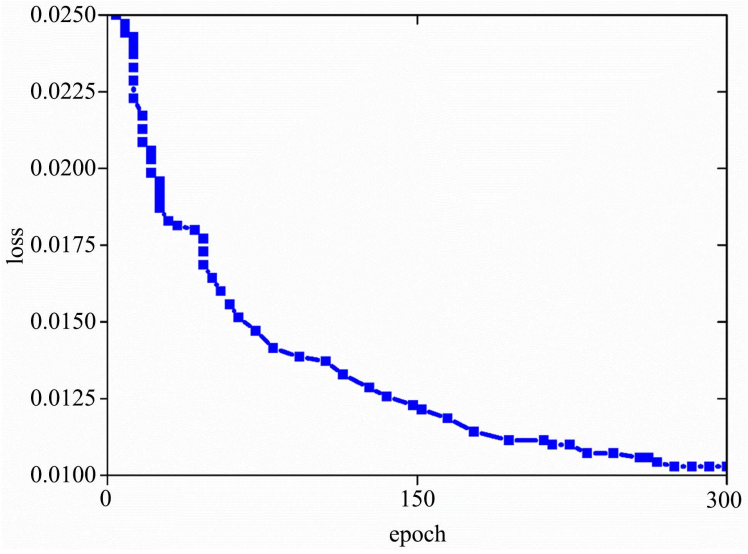


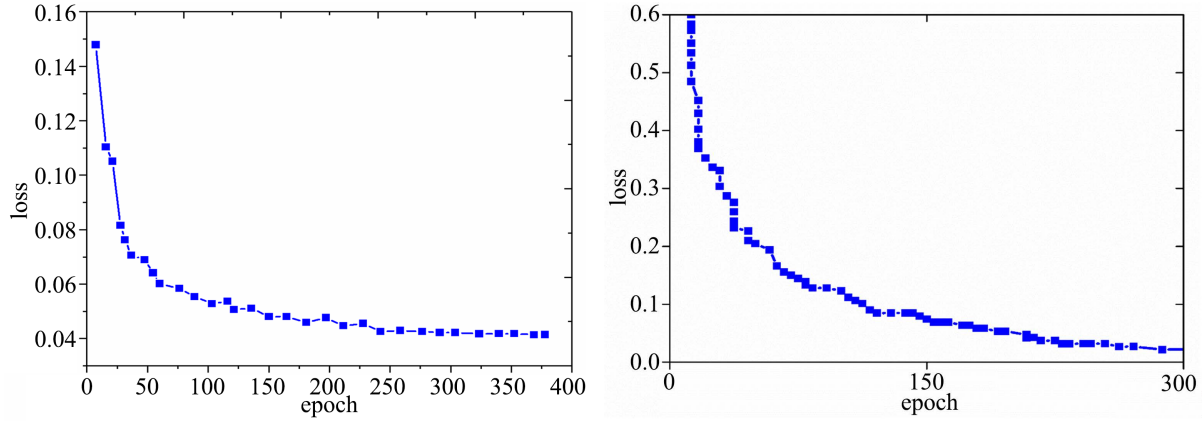
Figure 7. Improved version of ESRGAN + SPD-YOLOv5 simple diagram
图 7. 改进版 ESRGAN + SPD-YOLOv5 简易图

4. 实验分析

本文将训练的 batchsize 设置为 16，使用 SGD 随机梯度下降法作为本网络的优化器。为了获得本网络最优的超参数，本文采用基于遗传算法的超参数进化算法来对网络的超参数进行优化。为了减少训练时间，本文挑选 1000 张图片作为进化过程中的数据集，evolve 设置为 50，epoch 设置为 100，最终在第 30 代之后训练精度收敛于最大值，从而选择这组超参数作为本文网络的超参数。超参数具体设置如下：初始学习率设置为 0.01392，学习率动量设置为 0.96783，权重衰减系数何止为 0.002，图像 Mosaic 的概率设置为 1.0。最终三中损失函数收敛情况如图 8 所示。



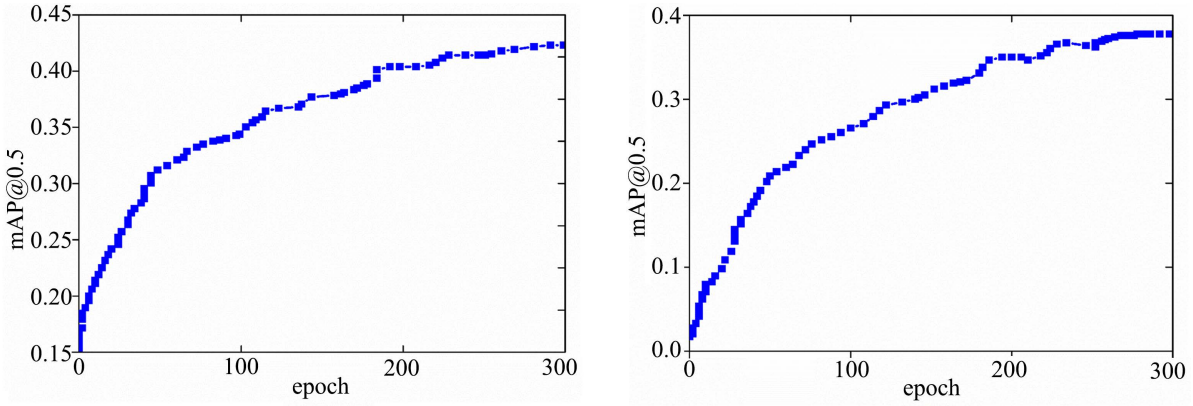
(a) $Loss_{obj}$



(b) $Loss_{box}$ (c) $Loss_{cls}$

Figure 8. Loss of function convergence
图 8. 损失函数收敛情况

预测精度mAP随epoch的变化如图9所示。



(a) 本文算法 (b) YOLOv5

Figure 9. mAP curve
图 9. mAP 曲线

4.1. VisDrone 数据集实验分析

为了让本文的实验更加周密和方便对比，本文借鉴了文献[16]所做的实验设置，验证 YOLOv3、YOLOv5 和本文所提出的模型在 VisDrone2019 数据集上的实验结果，与此同时也将 BetterFPN 和 RRNet 加入实验对比，最终结果如表 2 所示。由表 2 可以看出，本文所提出的模型*AP*评价指标均高于 YOLOv3、RRNet、BettrtFPN 等模型。比原始 YOLOv5 模型提高了 3.73 个百分点。实验也表明本文提出的模型在小目标数据集上效果优秀以及在不同的数据集上也具有较强的鲁棒性。本文模型在 VisDrone2019 数据集上训练出的混淆矩阵如图 10 所示。

Table 2. VisDrone dataset experimental results
表 2. VisDrone 数据集实验结果

模型	*mAP* (%)
YOLOv3	20.41

Continued

RRNet	29.13
BettrtFPN	28.55
YOLOv5	39.73
DR-ESRGAN + SPD + CA-YOLOv5	43.46

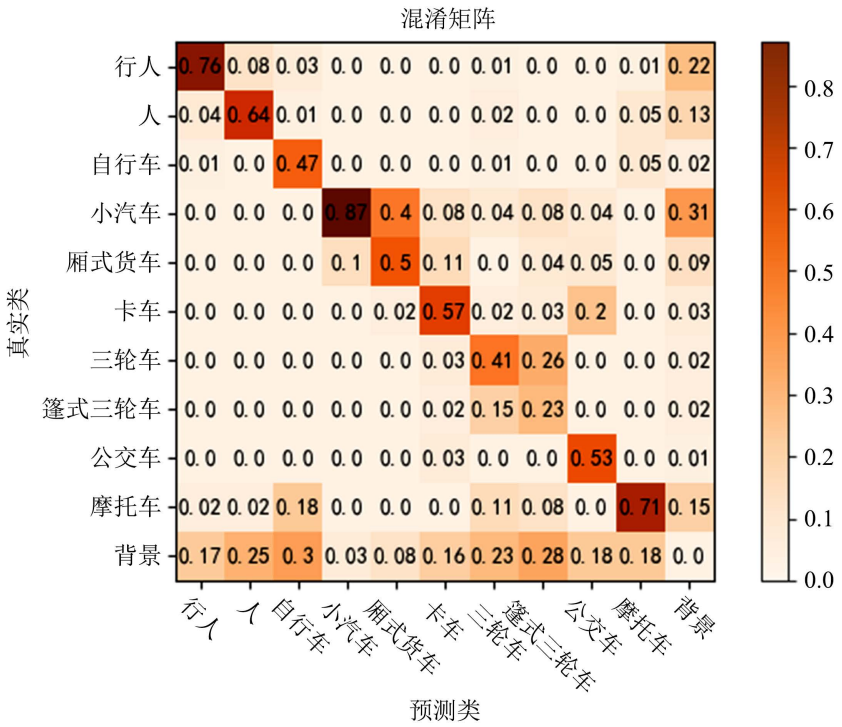


Figure 10. VisDrone dataset experimental confusion matrix
图 10. VisDrone 数据集实验混淆矩阵

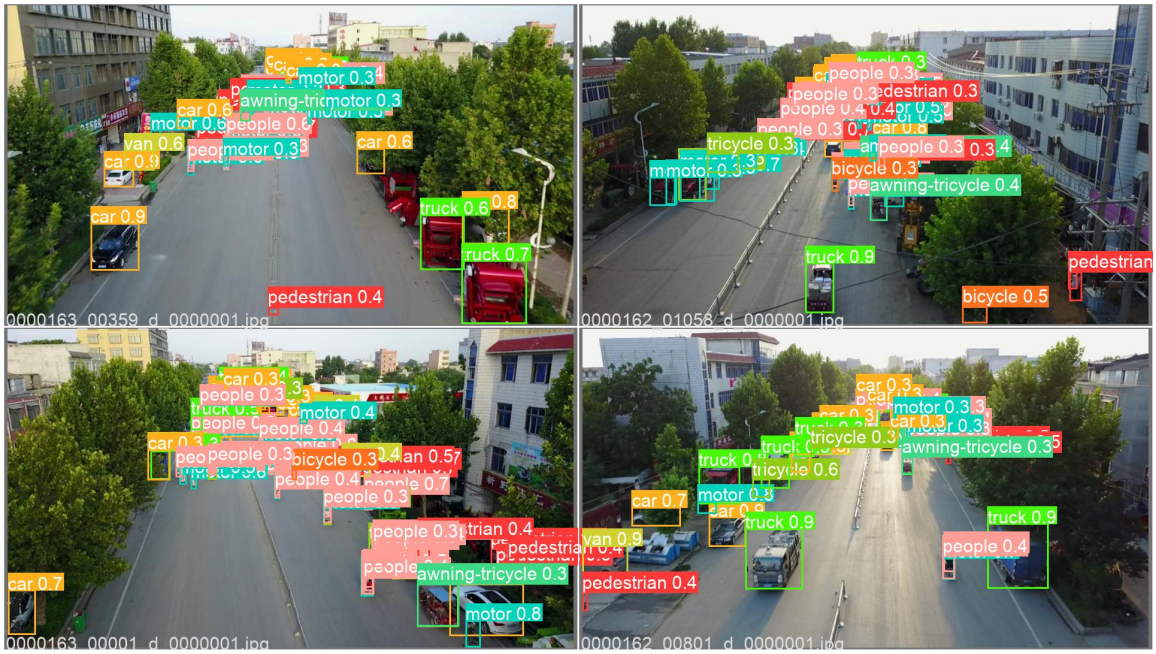
4.2. 消融实验

为了探究本文提出或改进模块对于整个模型的提升效果，本节使用VisDrone2019数据集进行验证，并进行消融实验，在原始YOLOv5的基础是逐一增加各模块，并比较实验结果以评估每个模块的贡献。表3列出了消融实验的结果。

Table 3. Results of ablation experiment
表 3. 消融实验结果

模型	*mAP* (%)
YOLOv5	39.73
YOLOv5 + DR-ESRGAN	41.65
YOLOv5 + DR-ESRGAN + CA	41.97
YOLOv5 + DR-ESRGAN + SPD	42.73
DR-ESRGAN + SPD + CA-YOLOv5	43.46

由表 3 可知，增加超分辨生成网络和 SPD 模块对于检测网络 *mAP* 的提升贡献相对较大，分别提升了 1.92 个百分点和 1.08 个百分点，说明对特征图进行超分辨重建和取消跨步卷积能有效的增强模型对于微小目标的检测能力。同时可以发现单独增加 CA 注意力模块对于检测网络 *mAP* 提升效果比较一般，但是将 CA 模块嵌入到 DR-ESRGAN + SPD-YOLOv5 中时提升较大为 0.73 个百分点，说明注意力模块能够结合其它提升模块能够发挥更好的效果。图 11 为本文模型在 VisDrone 数据集上的部分检测效果。



(a)

(b)

Figure 11. The partial detection effect of this model on the VisDrone dataset
图 11. 本文模型在 VisDrone 数据集上的部分检测效果

5. 总结

本文针对 YOLOv5 目标检测模型对小目标检测精度较差的问题，首先对 ESRGAN 超分辨率网络进行改进，在 RRDB 基本块的基础上添加额外的残差学习层，并且在每个密集块的每两层中再添加一个残差。在不增加复杂性的同时增加模型的理解力，以达到更好的鲁棒性，同时引入高斯噪声，在生成图像的某些局部方面随机化生成，能够使模型在更高级的层面提供更细致的纹理细节。接着引入 SPD-Conv 模块来取消跨步卷积，提升目标检测网络对微小物体的特征感知能力。最后对融合超分辨率重建、SPD-Conv 模块的 YOLOv5 目标检测网络进行实验分析，实验证明本文提出的网络对小目标物体的检测性能优越，对比原始 YOLOv5 模型 *mAP* (%)提高了 3.73 个百分点。最后进行消融实验，目的是验证本文提出的各种模块对小目标检测性能的贡献。

参考文献

[1] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR*'05), San Diego, 20-25 June 2005, 886-893. https://doi.org/10.1109/CVPR.2005.177

[2] Szegedy, C., Toshev, A. and Erhan, D. (2013) Deep Neural Networks for Object Detection. In: Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K.Q., Eds., *Advances in Neural Information Processing Systems* 26 (*NIPS* 2013), Curran Associates Inc., Red Hook.

[3] Ding, S. and Zhao, K. (2018) Research on Daily Objects Detection Based on Deep Neural Network. *IOP Conference Series*: *Materials Science and Engineering*, **322**, Article ID: 062024. https://doi.org/10.1088/1757-899X/322/6/062024

[4] Zhang, N., Donahue, J., Girshick, R. and Darrell, T. (2014) Part-Based R-CNNs for Fine-Grained Category Detection. In: Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T., Eds., *Computer Vision—ECCV* 2014. *Lecture Notes in Computer Science*, Vol. 8689, Springer, Cham, 834-849. https://doi.org/10.1007/978-3-319-10590-1_54

[5] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R., Eds., *Advances in Neural Information Processing Systems* 28 (*NIPS* 2015), Curran Associates Inc., Red Hook.

[6] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[7] Liu, W., Anguelov, D., Erhan, D., *et al*. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV* 2016. *Lecture Notes in Computer Science*, Vol. 9905, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[8] Creswell, A., White, T., Dumoulin, V., *et al*. (2018) Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, **35**, 53-65. https://doi.org/10.1109/MSP.2017.2765202

[9] Bai, Y., Zhang, Y., Ding, M. and Ghanem, B. (2018) SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018. *Lecture Notes in Computer Science*, Vol. 11217, Springer, Cham, 206-221. https://doi.org/10.1007/978-3-030-01261-8_13

[10] Rabbi, J., Ray, N., Schubert, M., Chowdhury, S. and Chao, D. (2020) Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network. *Remote Sensing*, **12**, Article No. 1432. https://doi.org/10.3390/rs12091432

[11] Wang, X., Yu, K., Wu, S., *et al*. (2019) ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: Leal-Taixé, L. and Roth, S., Eds., *Computer Vision—ECCV* 2018 *Workshops*. *Lecture Notes in Computer Science*, Vol. 11133, Springer, Cham, 63-79. https://doi.org/10.1007/978-3-030-11021-5_5

[12] Ledig, C., Theis, L., Huszár, F., *et al*. (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 105-114. https://doi.org/10.1109/CVPR.2017.19

[13] Chen, Y., Li, J., Xiao, H., *et al*. (2017) Dual Path Networks. Advances in Neural Information Processing Systems. In: Guyon, I., *et al.*, Eds., *Advances in Neural Information Processing Systems* 30 (*NIPS* 2017), Curran Associates Inc., Red Hook.

[14] Martin, D., Fowlkes, C., Tal, D. and Malik, J. (2001) A Database of Human Segmented Natural Images and Its Appli-

cation to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *Proceedings* 8*th IEEE International Conference on Computer Vision*. *ICCV* 2001, Vancouver, 7-14 July 2001, 416-423.

[15] Sunkara, R. and Luo, T. (2023) No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. In: Amini, MR., Canu, S., Fischer, A., Guns, T., Kralj Novak, P. and Tsoumakas, G., Eds., *Machine Learning and Knowledge Discovery in Databases*. *ECML PKDD* 2022. *Lecture Notes in Computer Science*, Vol. 13715, Springer, Cham, 443-459. https://doi.org/10.1007/978-3-031-26409-2_27

[16] Zhu, P., Wen, L., Du, D., *et al*. (2021) Detection and Tracking Meet Drones Challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 7380-7399. https://doi.org/10.1109/TPAMI.2021.3119563