

水下视频中华鲟检测方法研究

周温晖, 贾永红, 焦 阳

武汉大学遥感信息工程学院, 湖北 武汉

收稿日期: 2022年7月18日; 录用日期: 2022年8月18日; 发布日期: 2022年8月24日

摘 要

针对水下拍摄的中华鲟视频, 本文研究并改进了YOLOv5检测算法, 通过加入了标签平滑与SENet注意力机制, 改进后的YOLOv5算法用于水下视频中华鲟检测。试验结果表明, 改进后YOLOv5算法的检测平均精度(mAP)比原YOLOv5算法提升6.6%以上, 可用于水下视频中华鲟检测。

关键词

YOLOv5, 中华鲟, 标签平滑, 注意力机制

Study on Underwater Video Detection Method of Chinese Sturgeon

Wenhui Zhou, Yonghong Jia, Yang Jiao

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan Hubei

Received: Jul. 18th, 2022; accepted: Aug. 18th, 2022; published: Aug. 24th, 2022

Abstract

Aiming at the underwater video of Chinese sturgeon, in this paper, the YOLOv5 detection algorithm was studied and improved by adding tag smoothing and SENet attention mechanism. The improved YOLOv5 algorithm is used for underwater video Chinese sturgeon detection. The experimental results show that the average detection accuracy (mAP) of the improved YOLOv5 algorithm is 6.6% higher than that of the original YOLOv5 algorithm, and can be used for underwater video detection of Chinese sturgeon.

Keywords

YOLOv5, Chinese Sturgeon, Label Smoothing, Attention Mechanism

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中华鲟是硬骨鱼纲、鲟科的鱼类，是中国长江中最大的鱼，故有“长江鱼王”之称。由于中华鲟是地球上最古老的脊椎动物，其自身所具有的一系列原始特征具有重要的科学研究价值，同时研究其生态、社会和经济价值也有很大的前景。随着长江生态环境的恶化，以及人类对中华鲟的过度捕捞行为，使得中华鲟在栖息地不断减少，其自然种群规模也急剧缩小，中华鲟种群延续正面临严峻考验。

国内外学者基于计算机视觉对鱼类检测与识别进行了诸多的研究。传统的鱼类目标检测方法是机助提取鱼类的形状、大小、颜色、纹理等特征，然后把特征向量输入到分类器中进行分类。但是传统方法使用的特征较少，很难从大数据量的视频进行更丰富的特征提取，存在较大的局限性。随着深度学习的不断发展，众多科研机构将经典深度学习检测方法应用于水下目标识别[1]。2016年 LiXiu 等[2]将 Fast R-CNN 应用于水下鱼类图像检测，该方法不仅平均检测精度比传统 R-CNN 算法提高了 10%以上，而且检测速度更是提高了一个数量级。XuC 等[3]利用 Faster R-CNN 检测算法评估视频中鱼的多样性和密度，在水下目标的数据集上分类准确率达到 94.1%。2018年 XinS 等[4]就提出了一种 CNN 知识传递框架，这种框架用于水下目标识别，有效解决了低对比度图像鉴别特征的提取问题。ChunleiX 等[5]利用 YOLOv2 模型实现了水下视频海参目标检测，提高了检测效率。沈军宇等[6]利用 CNN 和非极大值抑制 (Non-Maximum Suppression, NMS)处理，基于 YOLO 算法提出了一种端到端鱼群检测算法，获得了良好的检测精度，提升了检测速度。LabaoA 等[7]提出了一种基于深度网络结构的鱼类检测系统，该系统由一组卷积神经网络通过长、短期记忆网络串接组成，具有很好的鲁棒性。朱世伟等[8]提出了一种基于类加权 YOLO 网络的水下目标检测方法，引入了目标框自适应维度聚类算法来提升召回率，并用类加权损失来权衡不同样本的难易程度，提高了平均准确率。张明华等[9]提出了一种基于改进 Cascade R-CNN 目标检测模型和 MSRCF 图像增强的夜晚水下鱼类目标检测方法，实现对夜晚水下鱼类的精准检测，与其他目标检测方法相比，提高了检测精度和检测效率。2022年郑宗生等[10]基于 YOLOv4 网络模型，在 CIoU 损失函数的基础上引入新的损失项，改进 YOLOv4 模型的平均精度均值(mean Average Precision, mAP)比原模型有较大提升，但同时也存在对模糊图像的识别结果较差的问题。

为开展水下视频中华鲟的检测与识别，针对水下视频图像目标检测精度较低问题，研究并改进了 YOLOv5 检测算法，加入标签平滑与 SENet 注意力机制，以期提高检测精度和效率，从而为后续对中华鲟的行为分析与保护工作提供技术支持。

2. 研究方法

目前 YOLOv5 因其快速、准确、轻量等诸多优点，被广泛使用诸多方面。本文主要集中在原始的 YOLOv5 算法上使用一些技巧，用来提高检测精度和效率。鉴于 YOLOv5 的特点。在此介绍 YOLOv5 网络及其改进用于水下视频中华鲟的检测与识别方法。

2.1. YOLOv5 网络

YOLOv5 检测网络共有 YOLOv5x、YOLOv5l、YOLOv5m、YOLOv5s 四个版本。其中，YOLOv5

系列中深度最小, 特征图宽度最小的网络是 YOLOv5s 网络, 其它的 3 种都是在此基础上不断加深, 不断加宽。以 YOLOv5s 为例, 其网络结构如图 1。

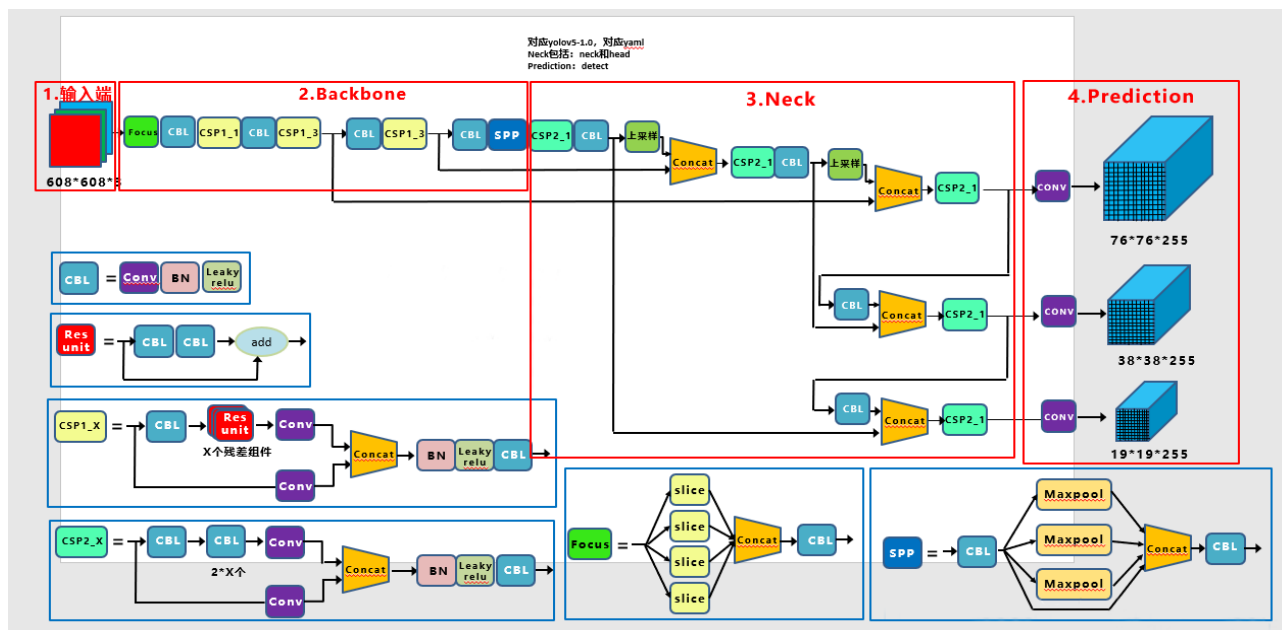


Figure 1. YOLOv5s Network structure

图 1. YOLOv5s 网络结构

YOLOv5 的结构分为输入端、骨干、颈部、预测输出端四个部分。

2.1.1. 输入端

1) 马赛克数据增强: YOLOv5 的输入端继承了 YOLOv4 的马赛克数据增强方式。马赛克数据增强方法可以很好地丰富检测物体的背景, 而且减少了本身对 batch size 的依赖, 对小目标的检测有较好的提升效果。

2) 自适应锚框计算: 这是 YOLO 中比较关键的一环, 由于不同的数据集有不同的格式和大小, 因此需要根据不同数据集来设定特定长宽的锚点框。在网络训练模型的初始阶段, 模型首先根据初始锚点框对目标的大概位置进行预测, 并产生对目标位置的预测框, 然后计算其与真实框之间的差距进行调整, 通过反向更新的操作来更新使整个网络的参数。不同于 YOLOv3 和 YOLOv4 中, 对于不同数据集的锚点框的获取是通过单独的程序运行来获得, YOLOv5 中将此功能嵌入到代码中, 每次训练, 根据数据集的名称自适应地计算出最佳的锚点框。

3) 自适应图片缩放: 在目标检测算法中, 常常会遇到图片尺寸差异问题, 面对图片长宽都不相同的情况, 在将其送入检测网络之前, 需要将原始图片处理成一个统一的标准尺寸。在实际应用中, 由于不同图片长宽比不同, 经缩放填充一系列统一化操作后, 不同图像两边会存在不同的填充黑边。如果黑边较大, 则信息冗余明显会影响推理速度。因此 YOLOv5 对原始图像自适应地添加最少的黑边, 从而减少信息的冗余, 提升检测的速度。

2.1.2. 骨干

1) 聚焦结构: 这一结构关键环节为切片操作。切片操作类似于邻近下采样对图片进行隔点抽取像素值, 从而一张图片就变成了互补的四张图片, 四张图片内容相似且无信息丢失。将四张图片拼接起来后,

得到了一张 12 个通道的新图片,最后将得到的新图片再经过卷积操作,最终得到了没有信息丢失情况下的二倍下采样特征图。

2) 跨阶段局部网络(Cross Stage Partial Networks, CSPNet)结构:YOLOv5s 的 CSPNet 结构是将原输入分成两个分支,分别进行卷积操作使得通道数减半,然后一个分支进行多个残差结构操作,另一个分支直接进行卷积;然后两个分支进行组合,使得 BottleneckCSP 的输入与输出是一样的大小,这样是为了让模型学习到更多的特征。YOLOv5 在骨干网络应用了 CSP1_X 结构,在颈部应用了 CSP2_X 结构。

2.1.3. 颈部

YOLOv5 采用改进版本的 CIOU_Loss 做候选框的损失函数,使得预测框回归的速度和精度更高一些。在 YOLOv5 的颈部结构中,CSP2 结构加强了其网络特征融合的能力。

2.1.4. 预测输出端

NMS 非极大值抑制:目标检测中利用 NMS 操作来进行很多目标框的筛选,是目标检测后续处理过程中比较重要的一个环节。因为 CIOU_Loss 中包含影响因子 v ,涉及正确真实的标签信息,而测试推理时,是没有正确真实的标签信息的。YOLOv5 采用加权 NMS 的方式,大多数状态下效果差不多,但在不增加计算成本的情况下,这种改变对于一些遮挡重叠的目标,确实会有改进。

2.2. YOLOv5 算法改进

2.2.1. 标签平滑(Label Smoothing)

过拟合和概率校准是训练深度学习模型时出现的两个问题。深度学习中有许多正则化技术可以解决过拟合问题,而标签平滑是其中一种正则化技术,它扰动目标变量,使模型对其预测的确定性降低。它之所以被视为一种正则化技术,是因为它限制了 softmax 函数的最大概率,使最大概率不会比其他标签大得多,即过度自信。其公式如下:

$$q'_i = (1 - \varepsilon)q_i + \frac{\varepsilon}{K} \quad (1)$$

式中 q 表示真实标签, ε 是一个非常小的常数, K 代表分类的类别数。经过标签平滑后能够减少模型过度依赖标签的问题,有效改善标签准确性不高的情况。

标签平滑方法用于 YOLOv5 的预测输出层。

2.2.2. SE 注意力机制

在 YOLOv5 的支柱中引入 SE 注意力机制是根据 CVPR2017 的 Squeeze-and-Excitation Networks [11] 一文中提到的注意力模块 SENet,如图 2 所示。该模块旨在通过使网络能够动态调制各通道的权重,从而重新校准特征来提高网络的表示能力。

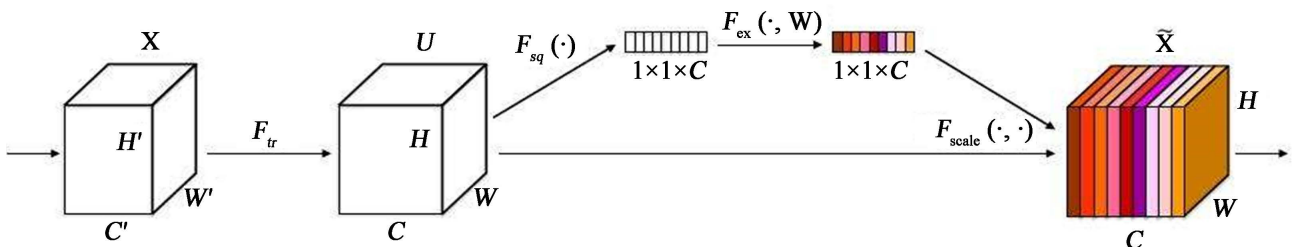


Figure 2. SE module working principle

图 2. SE 模块工作原理

图 2 中给定一个输入 x ，其特征通道数为 c_1 ，通过一系列卷积等一般变换后，得到一个特征通道数为 c_2 的特征。首先通过 Squeeze 操作将每个二维的特征通道变成一个实数，这个实数某种程度上具有全局的感受视野，并且输出的维度和输入的特征通道数相匹配。其次是 Excitation 操作，通过参数来为每个特征通道生成权重，这个过程参考 Resnet 中的瓶颈结构。

3. 试验与分析

3.1. 环境搭建

表 1 为基于 YOLOv5 的水下视频中华鲟检测与识别所用软硬件配置。

Table 1. Hardware resource configuration and software environment

表 1. 硬件资源配置和软件环境

Name	Type
操作系统	Windows 10 $\times$ 64
GPU	Intel(R)Core(TM)i5-7300HQ
CPU	NVIDIA GeForce GTX1050
内存	8.00 GB
CUDA	10.2.89
CUDNN	7.6.5
Pytorch	1.8.0
Anaconda3	4.10.1
Python	3.7.3

3.2. 试验分析

为进行水下视频中华鲟的检测与识别，视频采集获得包含中华鲟的 5000 张图像，对样本数据使用 Labellmg 标注工具进行人工标注，标签文件保存为 xml 形式文件，主要内容包含图像的宽、高、维度以及所标注的标签的类别、左上角坐标、右下角坐标等信息。图 3 为标注的示意图。将制作的样本集按照 8:1:1 分成训练样本集、验证集和测试集。上述工作完成后，开始网络训练。通过几次实验对比，将训练 epoch 时的训练周期设置为 130。



Figure 3. Sample labeling
图 3. 样本标注

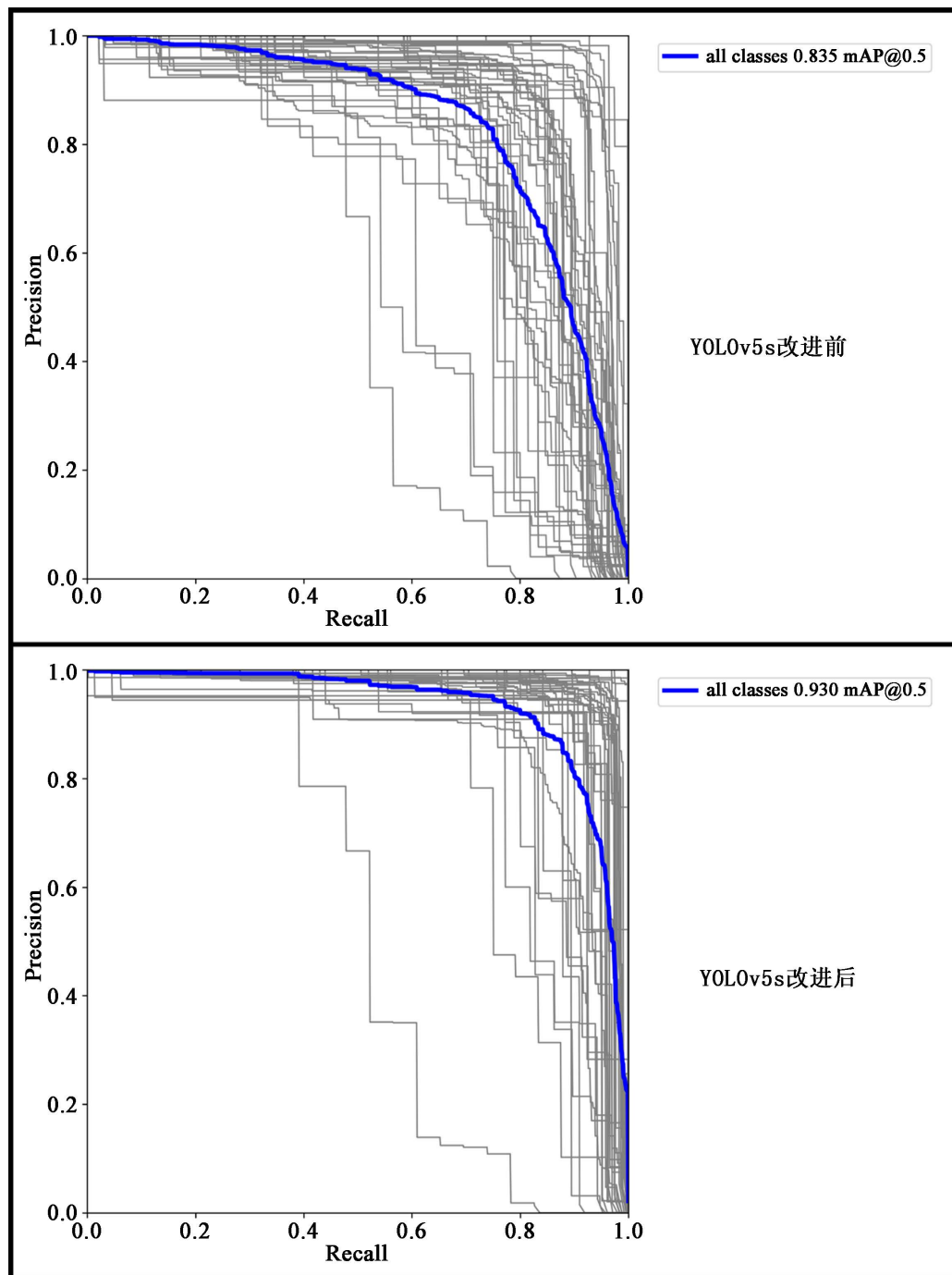


Figure 4. Comparison of PR curves before and after YOLOv5s model improvement

图 4. YOLOv5s 模型改进前后 PR 曲线对比

为了评价训练好的网络监测水下视频中华鲟的效果,采用 PR 曲线和 mAP 进行评价。PR 曲线中的 P 代表精准率(Precision), R 代表召回率(Recall)。将 R 设置为横坐标, R 设置为纵坐标, PR 曲线反映了精准率与召回率的关系。

以 YOLOv5s 模型为例,该算法改进前后 PR 曲线图如图 4。从图 4 中可以清楚地看到,算法改进后的 PR 曲线可以完全包住改进前的 PR 曲线,这说明改进后的算法精度有着较为明显的提高。

Table 2. Model detection accuracy of YOLOv5 algorithm
表 2. YOLOv5 算法模型检测精度

改进项	模型	YOLOv5s		YOLOv5l	
	Label	否	是	否	是
	SE	否	是	否	是
	mAP@0.5%	83.5	93.0	84.6	91.3

文中采用了 YOLOv5 四个模型中的 YOLOv5s 和 YOLOv5l 模型进行中华鲟检测试验。表 2 给出了不同模型进行中华鲟检测的 mAP 检测精度。表 2 中的 Label 代表 Label Smoothing 标签平滑, SE 表示 SENet 注意力机制, “是”、“否”代表在模型中是否采用该类方法。从表 2 来看, 对于 YOLOv5s 和 YOLOv5l 均进行标签平滑和注意力机制改进用于中华鲟检测, 相比原 YOLOv5s 和 YOLOv5l, mAP 分别提高了 9.5% 和 6.6%。可见, 本文所采用的改进算法能有效提高水下视频中华鲟的检测精度。

4. 结语

在深入了解 YOLOv5 算法的基础上, 对水下视频中的中华鲟检测提出了优化策略, 引入标签平滑以解决标签过度自信的问题, 提高模型的泛化性能; 加入 SENet 注意力机制, 增强网络的特征提取能力。试验结果表明, 改进后 YOLOv5 算法的检测精度有明显提高, 实现了水下视频中华鲟目标识别。但对于一些有差异的复杂水下环境还是会发生误检与漏检情况, 仍需要探索对特征提取网络作进一步优化的可能性, 同时扩大数据集, 使训练出的模型具有更好的实用性与通用性。

基金项目

该项成果获得“中华鲟产卵场重建”项目、武汉大学创新创业项目资助。

参考文献

- [1] 佟鑫. 基于深度学习的海洋生物目标检测方法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2020. <https://doi.org/10.27060/d.cnki.ghbcu.2020.002239>
- [2] Li, X., Shang, M., Qin, H., et al. (2016) Fast Accurate Fish Detection and Recognition of Underwater Images with Fast R-CNN. *OCEANS 2015-MTS/IEEE Washington*, Washington, 19-22 October 2015, 1-5.
- [3] Cao, X., Zhang, X., Yru, Y., et al. (2016) Deep Learning Based Recognition of Underwater Target. 2016 *IEEE International Conference on Digital Signal Processing*, Beijing, 16-18 October 2016, 89-93. <https://doi.org/10.1109/ICDSP.2016.7868522>
- [4] Sun, X., Shi, J.Y., Liu, L.P., et al. (2018) Transferring Deep Knowledge for Object Recognition in Low-Quality Underwater Videos. *Neurocomputing*, **275**, 897-908. <https://doi.org/10.1016/j.neucom.2017.09.044>
- [5] Xia, C.L., Fu, L.W., Liu, H. and Chen, L.X. (2018) *In Situ* Sea Cucumber Detection Based on Deep Learning Approach. 2018 *OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, Kobe, 28-31 May 2018, 1-4.
- [6] 沈军宇, 李林燕, 夏振平, 等. 一种基于 YOLO 算法的鱼群检测方法[J]. 中国体视学与图像分析, 2018, 90(2): 54-60.
- [7] Labao, A.B. and Naval Jr, P.C. (2019) Cascaded Deep Network Systems with Linked Ensemble Components for Underwater Fish Detection in the Wild. *Ecological Informatics*, **52**, 103-121. <https://doi.org/10.1016/j.ecoinf.2019.05.004>
- [8] 朱世伟, 杭仁龙, 刘青山. 基于类加权 YOLO 网络的水下目标检测[J]. 南京师大学报(自然科学版), 2020, 43(1): 129-135.
- [9] 张明华, 龙腾, 宋巍, 等. 基于改进 Cascade R-CNN 和图像增强的夜晚鱼类检测[J]. 农业机械学报, 2021, 52(9): 179-185.
- [10] 郑宗生, 李云飞, 卢鹏, 等. 改进 YOLOv4 模型在鱼类目标检测上的应用研究[J]. 渔业现代化, 2022, 49(1):

82-88+96.

- [11] Lin, T.Y., Dollar, P., Girshick, R., *et al.* (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 936-944.
<https://doi.org/10.1109/CVPR.2017.106>