



机器之心 翻译

2019/04/22 14:50

机器之心编辑部  
编译

# 谷歌大脑提出NAS-FPN：这是一种学会自动架构搜索的特征金字塔网络

目标检测也能用神经架构搜索，这是一个自动搜索的特征金字塔网络。

神经架构搜索已经在图像识别上展现出很强的能力，不论是可微架构搜索的速度，还是基于强化学习搜索的准确度，很多时候自动架构搜索已经超越了我们手动设计的版本。与此同时，学习视觉的特征表示是计算机视觉中的一个基本问题。不论是图像分类还是目标检测，抽取图像特征才是最首要的。

在过去几年里，用于图像分类和目标检测的深度卷积网络在架构搜索方面取得了很大进展。与预测图像类别概率的图像分类不同，目标检测在大范围的尺度和位置上检测和定位多个目标时存在自身的挑战。为了解决这个问题，很多现代目标检测器普遍使用金字塔特征表示，它表示具有多尺度特征层的图像。

在这篇论文中，谷歌大脑的 Quoc V. Le 等研究者提出了一种新的架构搜索方法 NAS-FPN。他们希望借助神经架构搜索的优势，并学习更好的目标检测特征金字塔网络架构。目前 NAS-FPN 实现了优于当前最佳目标检测模型的准确率和延迟权衡

## 目标检测中的神经架构搜索

特征金字塔网络（FPN）是目标检测中生成金字塔形状特征表示的代表性模型架构之一。它采用通常为图像分类设计的骨干模型，然后通过自上而下和横向连接，并组合不同的特征层来构建特征金字塔。FPN 取代了检测器（如 Faster R-CNN）中的特征提取器，并生成更高质量的金字塔特征图。



图注：特征金字塔网络，选自 arXiv: 1612.03144。

FPN 由自下而上和自上而下路径组成。其中自下而上的路径是用于提取特征的常用卷积网络，空间分辨率自下而上地下降。当空间分辨率下降，且检测到更高层的结构时，每层的语义值增加。

设计特征金字塔架构的挑战在于其巨大的设计空间。组合不同尺度的可能连接数量随着网络层数的增加呈指数级增长。最近，神经架构搜索算法在巨大的搜索空间中有效地发现了用于图像分类的高性能架构。为了实现这个领域的结果，Quoc V. Le 等研究者提出生成金字塔表示的可扩展架构搜索空间。

### NAS-FPN：一种灵活的目标检测新方法

本文的主要贡献是设计搜索空间，覆盖所有可能的跨尺度连接，已生成多尺度特征表示。在搜索过程中，研究者的目标是发现具有相同输入和输出特征级别并且可以被重复应用的微粒架构。模块化搜索空间使得搜索金字塔架构变得易于管理。模块化金字塔架构的另一个好处是可以随时检测目标（即「early exit」），虽然这种「early exit」方法已经被尝试过，但手动设计这种架构依旧相当困难。

研究者构建的架构，即 NAS-FPN，在构建目标检测架构方面具有很大的灵活性。NAS-FPN 与各种骨干模型配合得很好，如 MobileNet、ResNet、AmoebaNet。它为移动端模型和高准确率模型在速度和准确率方面提供了更好的权衡。

骨干模型结合，NASFPN 在单个测试规模中达到了 48.3 的 AP 单模型准确率。其检测准确率超过了 Mask RCNN，同时使用的推理时间更少。几种模型的具体结果如图 1 所示。

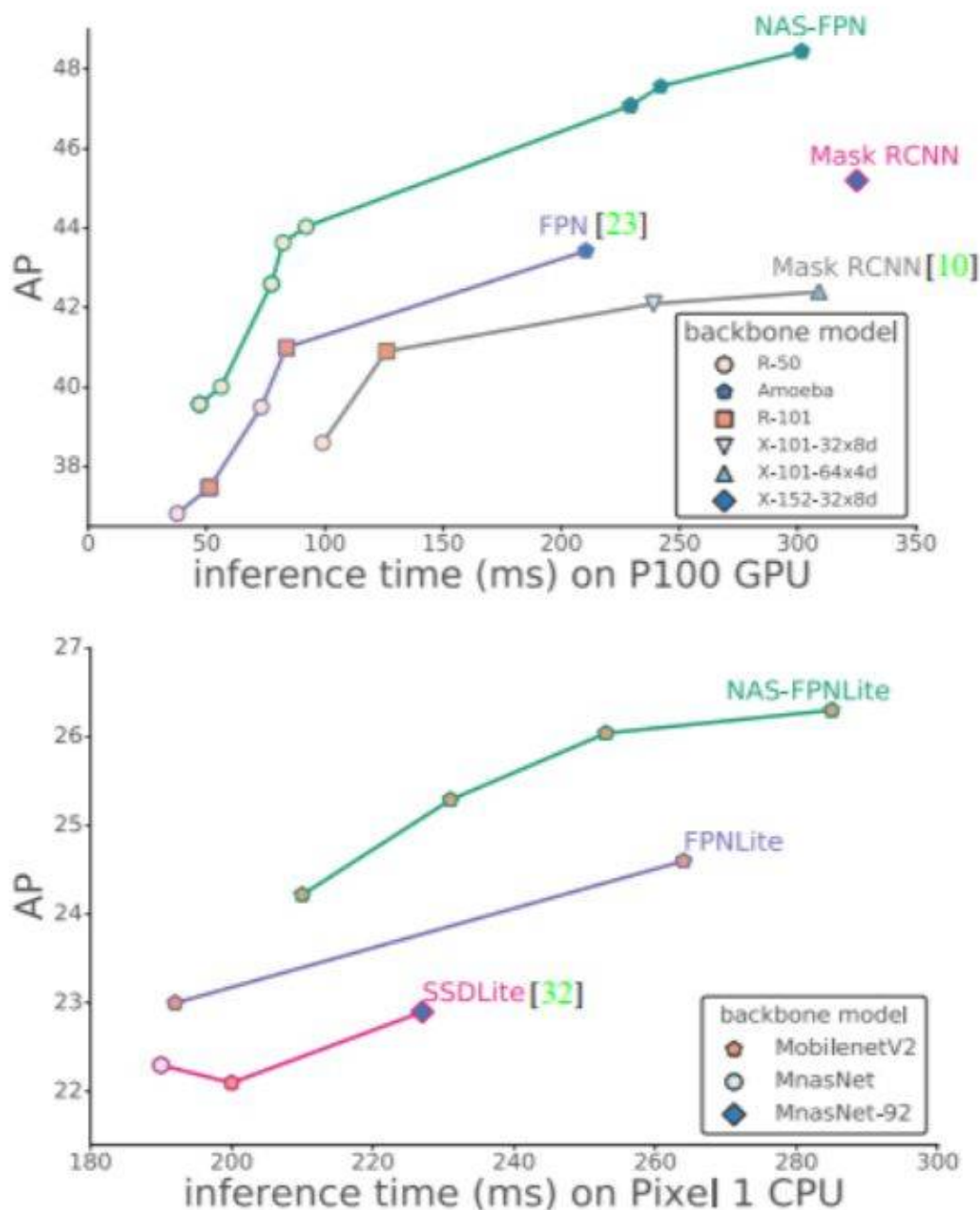


图1：移动设备上准确型模型（上）和快速型模型（下）的平均精度与推断时间关系。绿色折线是 NASFPN 与 RetinaNet 相结合的结果。详情请看图9。



## for Object Detection

Golnaz Ghaisi Tsung-Yi Lin Ruoming Pang Quoc V. Le

Google Brain

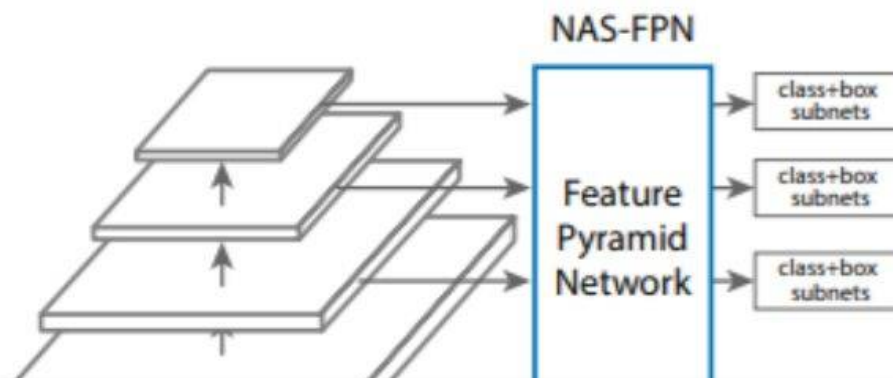
{golnazg, tsungyi, rpang, qvl}@google.com

论文地址：<https://arxiv.org/pdf/1904.07392.pdf>

**摘要：**当前最先进的目标检测卷积架构都是人工设计的。在这项工作中，我们的目标是学习更好的目标检测特征金字塔网络架构。我们采用了神经架构搜索，在一个包含所有跨尺度连接的新的可扩展搜索空间中发现了一个新的特征金字塔架构。这个名为 NAS-FPN 的架构包含自上而下和自下而上的连接，以融合各种尺度的特征。NAS-FPN 与 RetinaNet 框架中的若干骨干模型相结合，实现了优于当前最佳目标检测模型的准确率和延迟权衡。该架构将移动检测准确率提高了 2 AP，优于 [32] 中的当前最佳模型——与 MobileNetV2 相结合的 SSDLite，达到了 48.3 AP，超越了 Mask R-CNN [10] 的检测准确率，且计算时间更少。

## 方法

本文中的方法基于 RetinaNet 框架 [23]，因为该框架简单、高效。RetinaNet 框架有两个主要的组成部分：一个骨干网络（通常是当前最优的图像分类网络）和一个特征金字塔网络（FPN）。本文算法的目标是为 RetinaNet 框架发现更好的 FPN 架构。图 2 所示为 RetinaNet 架构。





堆叠多次，以获得更高的准确率。

为了找到更好的 FPN，研究者利用 Quoc Le 等人在「Neural architecture search with reinforcement learning」中提出的神经架构搜索（NAS）框架。NAS 利用强化学习训练控制器在给定的搜索空间中选择最优的模型架构。控制器利用子模型在搜索空间中的准确度作为奖励信号来更新其参数。因此，通过反复试验，控制器逐渐学会了如何生成更好的架构。

研究者还为 FPN 设计了一个搜索空间来生成特征金字塔表征。为了实现 FPN 的可扩展性，研究者强制 FPN 在搜索过程中重复 N 次，然后连接到一个大型架构中。他们将这一特征金字塔架构命名为 NAS-FPN。

## 实验

这一部分描述了学习一个 RNN 控制器来发现 NAS-FPN 架构的神经架构搜索实验。然后，研究者证明了他们发现的 NAS-FPN 在不同的骨干模型和图像大小下都能很好地工作。在金字塔网络中，通过改变叠加层数和特征维数，可以很容易地调整 NAS-FPN 的容量。此外，作者还在实验中展示了如何构建准确、快速的架构。

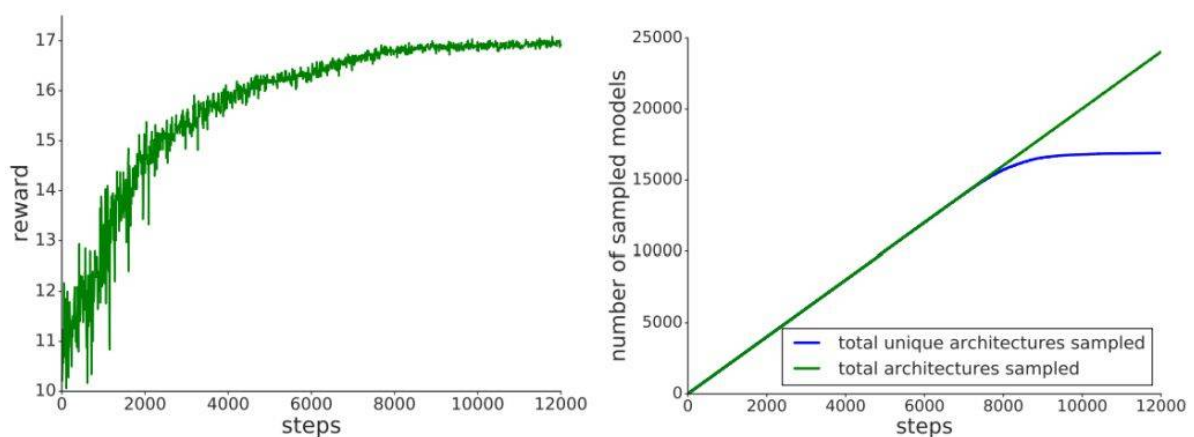


图5：左：强化学习训练的奖励。计算奖励的方法为在代理任务上采样架构的AP。右：采样的独特架构数与架构总数。随着控制器逐渐收敛，控制器会采样到越来越多的相同架构。



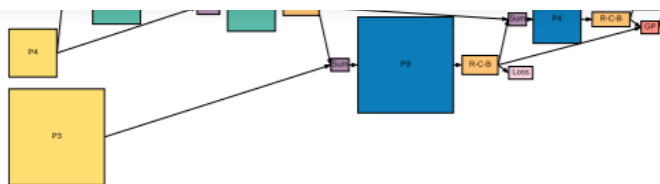


图6：5 个输入层（黄色）和5 个输出特征层（蓝色）的NAS-FPN 中发现的7-merging-cell 金字塔网络架构。GP：全局池化；R-C-B：ReLU-Conv-BatchNorm。

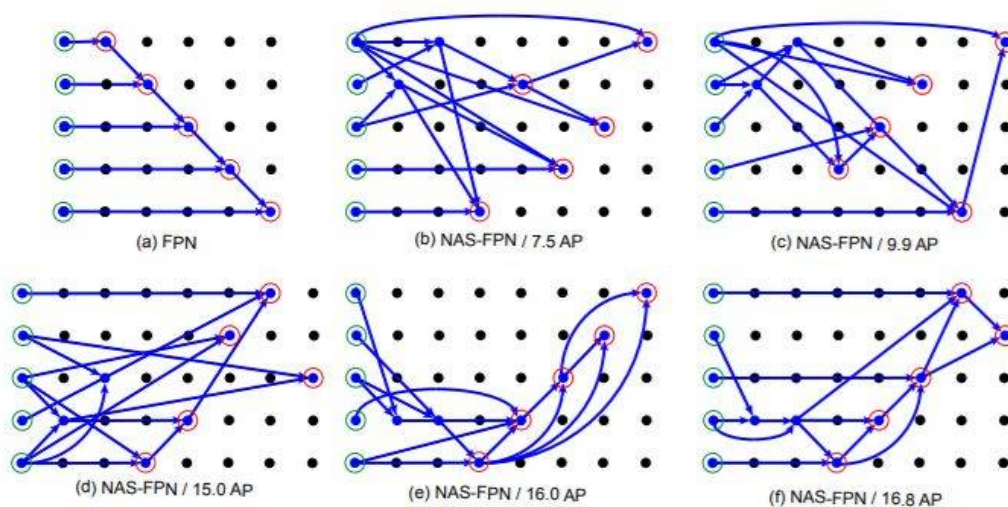
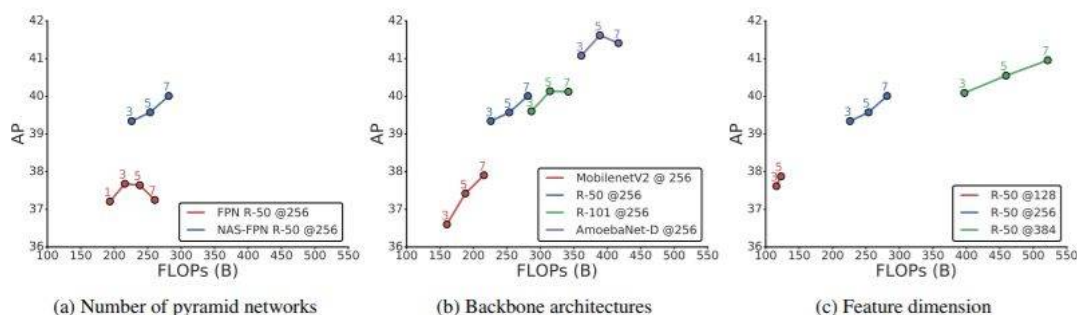


图7：NAS-FPN 的架构图。每个点代表一个特征层，同一行的特征层具有相同的分辨率，分辨率由下往上递减。箭头表示内层之间的连接，该图的结构是输入层位于左侧。金字塔网络的输入用绿色圆圈标记，输出用红色圆圈标记。(a) 基线 FPN 架构。(b~f) 在 RNN 控制器的训练中通过神经架构搜索发现的 7-cell NAS-FPN 架构。(f) 实验中最后收敛得出的 NAS-FPN。



在 NAS-FPN 中金字塔网络的数量。

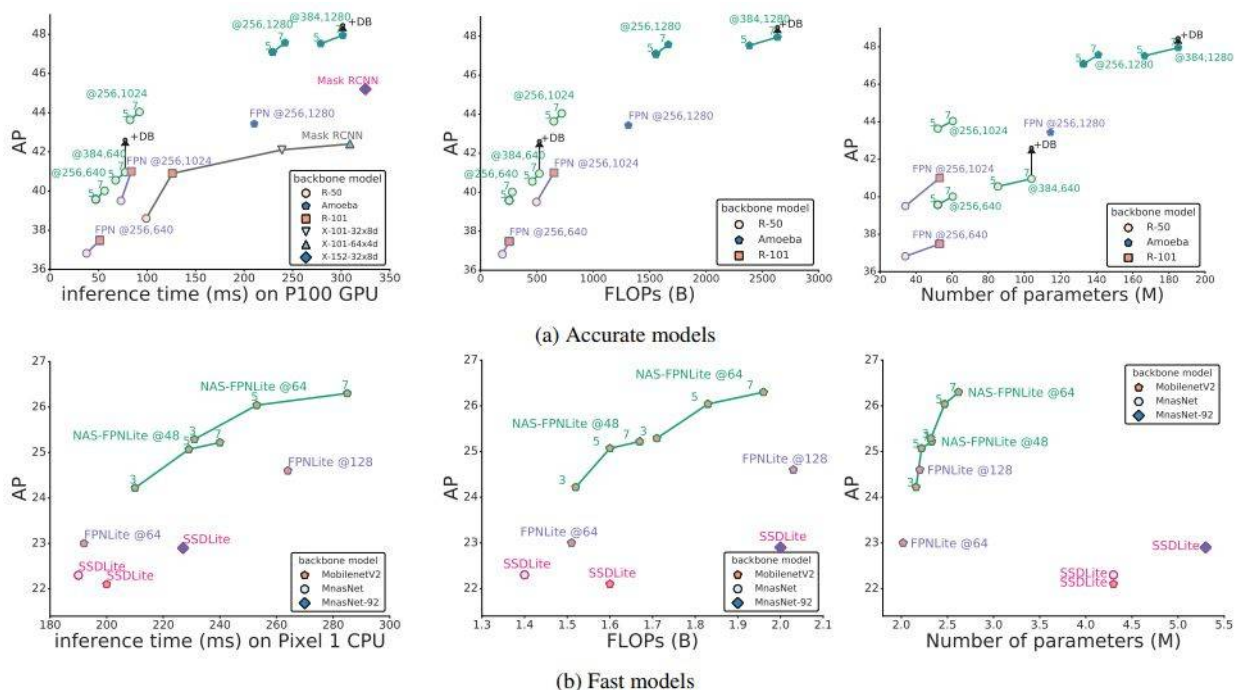


图9：检测准确率和推理时间的折衷（左）、浮点数（中）、参数（右）。（a）研究者将其与其他高准确率模型进行了对比。所有模型的推理时间都是在一个搭载 P100 GPU 的设备上计算出来的。绿色折线显示的是拥有不同骨干架构的 NAS-FPN 结果。该标记上的数字表示在 NAS-FPN 中金字塔网络的重复次数。在每个数据点旁边都表明了 NAS-FPN/FPN 的特征维数和输入图像大小。（b）研究者将自己的模型与其他快速模型进行了对比，其中所有模型的输入图像大小为 320x320，推理时间是在 Pixel 1 CPU 上计算的。本文中的模型是用 MobileNetV2 的轻型模型训练的。



MnasNet + SSDLite [36]	320 × 320	1.4B	4.3M	190 (Pixel 1 CPU)	22.3
MnasNet-92 + SSDLite [36]	320 × 320	2.0B	5.3M	227 (Pixel 1 CPU)	22.9
FPNLite MobileNetV2 @ 64	320 × 320	1.51B	2.02M	192 (Pixel 1 CPU)	22.7
FPNLite MobileNetV2 @ 128	320 × 320	2.03B	2.20M	264 (Pixel 1 CPU)	24.3
NAS-FPNLite MobileNetV2 (3 @ 48)	320 × 320	1.52 B	2.16 M	210 (Pixel 1 CPU)	24.2
NAS-FPNLite MobileNetV2 (7 @ 64)	320 × 320	1.96 B	2.62 M	285 (Pixel 1 CPU)	25.7
YOLOv3 DarkNet-53 [30]	608 × 608	140.69 B	-	51 (Titan X)	33.0
CornerNet Hourglass [18]	512 × 512	-	-	244 (Titan X)	40.5
Mask R-CNN X-152-32x8d [11]	1280 × 800	-	-	325 (P100)	45.2
RefineDet R-101 [41]	832 × 500	-	-	90 (Titan X)	34.4
FPN R-50 @256 [23]	640 × 640	193.6B	34.0M	37.5 (P100)	37.0
FPN R-101 @256 [23]	640 × 640	254.2B	53.0M	51.1 (P100)	37.8
FPN R-50 @256 [23]	1024 × 1024	495.8B	34.0M	73.0 (P100)	40.1
FPN R-101 @256 [23]	1024 × 1024	651.1B	53.0M	83.7 (P100)	41.1
FPN AmoebaNet @256 [23]	1280 × 1280	1311 B	114.4 M	210.4 (P100)	43.4
NAS-FPN R-50 (7 @ 256)	640 × 640	281.3B	60.3M	56.1 (P100)	39.9
NAS-FPN R-50 (7 @ 256)	1024 × 1024	720.4B	60.3M	92.1 (P100)	44.2
NAS-FPN R-50 (7 @ 256)	1280 × 1280	1125.5B	60.3M	131.9 (P100)	44.8
NAS-FPN R-50 (7 @ 384)	1280 × 1280	2086.3B	103.9 M	192.3 (P100)	45.4
NAS-FPN R-50 (7 @ 384) + DropBlock	1280 × 1280	2086.3B	103.9M	192.3 (P100)	46.6
NAS-FPN AmoebaNet (7 @ 384)	1280 × 1280	2633 B	166.5 M	278.9 (P100)	48.0
NAS-FPN AmoebaNet (7 @ 384) + DropBlock	1280 × 1280	2633 B	166.5 M	278.9 (P100)	48.3

表1：NAS-FPN 和其他当前最优检测器在 COCO 测试数据集上的性能表现。

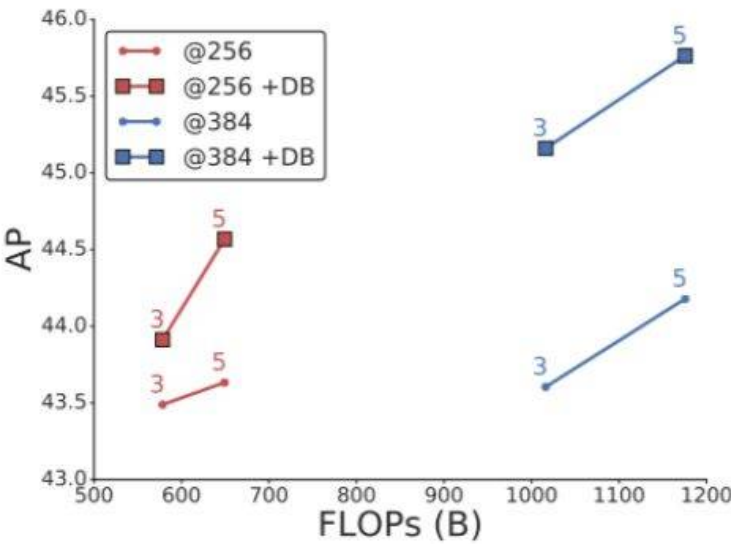


图10：特征维数为 256 或 384 的 NAS-FPN 在训练时有无 DropBlock (DB) 的性能对比。模型和 ResNet-50 骨干模型在大小为 1024x1024 的图像上训练。当我们在金字塔网络中增加特征维数时，添加 DropBlock 变得更重要。



推荐文章



CVPR 2017论文解读：特征金字塔网络FPN

 机器之心

 4



一文概览2017年Facebook AI Research的计算机视觉...

 路雪

 4



克服预训练的缺陷，清华&旷视提出专用于目标检测...

 路雪

 4

登录 后评论



暂无评论~



关于我们 寻求报道 商务合作 加入我们 服务条款

©2019 机器之心（北京）科技有限公司

京 ICP 备 12027496

- 全球人工智能  
信息服务
- 友情链接:
- Synced  
Global
- 机器之心  
Medium 博  
客
- PaperWeekly
- 网易智能
- 动脉网
- 硬蛋网
- 达观数据
- 品途商业评  
论
- 艾耕科技

