目标检测

作者：Calvin
QQ: 179209347
Mail: 179209347@qq.com

介绍

笔记简介：

- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：

https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：

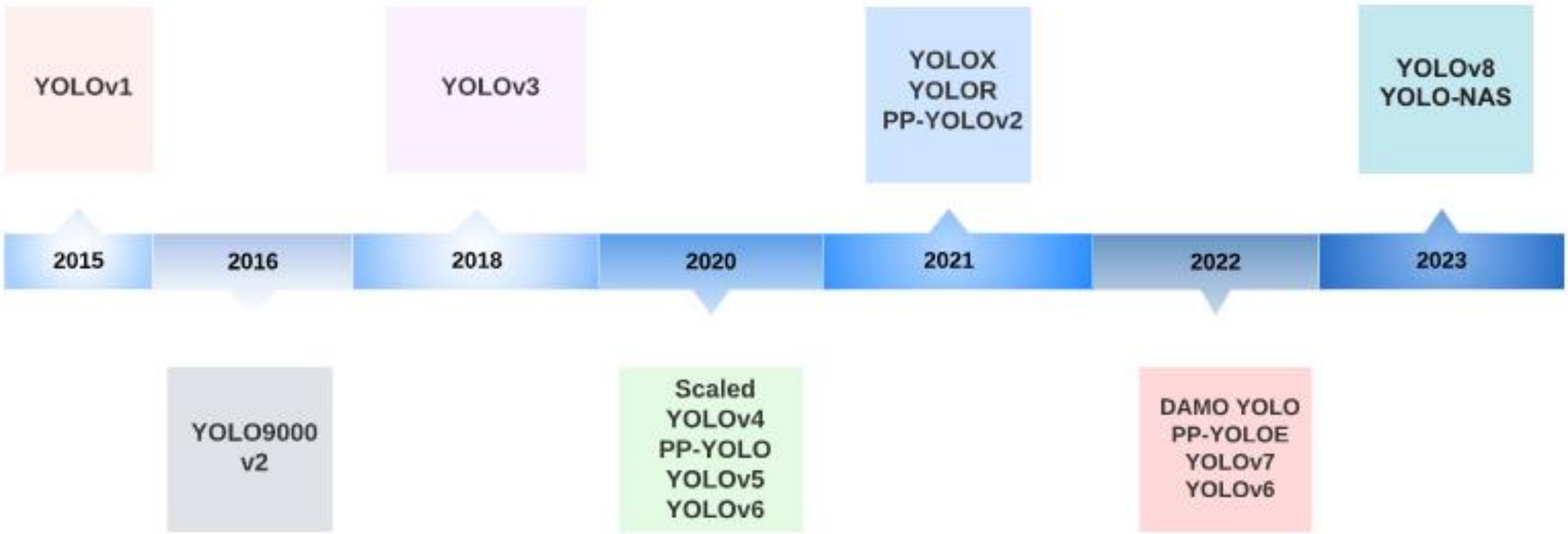
https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

*** 版权声明：**

- 仅限用于个人学习
- 禁止用于任何商业用途

单阶段目标检测 – YOLO系列

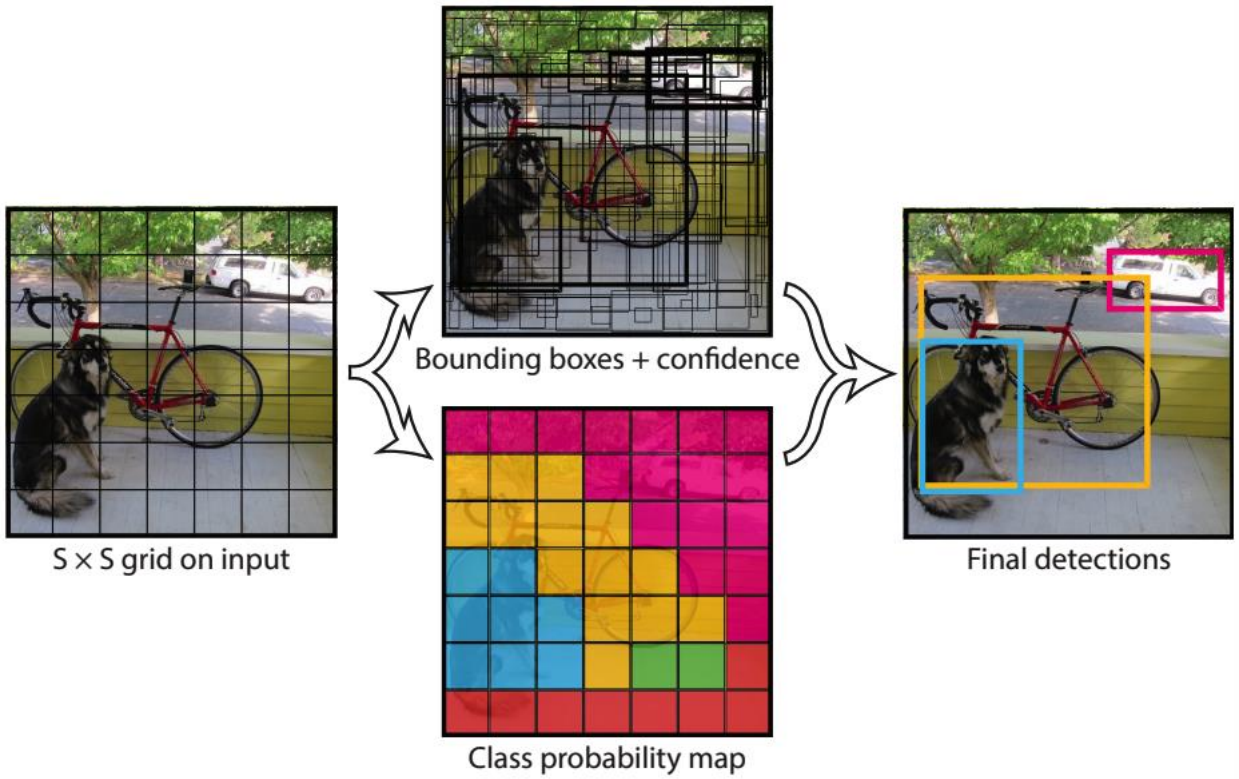


单阶段目标检测 – YOLOv1

YOLO（You Only Look Once）是一种流行的目标检测算法，YOLOv1是该系列算法的第一个版本。YOLO的主要特点是在单个神经网络中实现端到端的目标检测，相比传统的目标检测方法，YOLO具有更快的检测速度。

YOLOv1的一些关键特点：

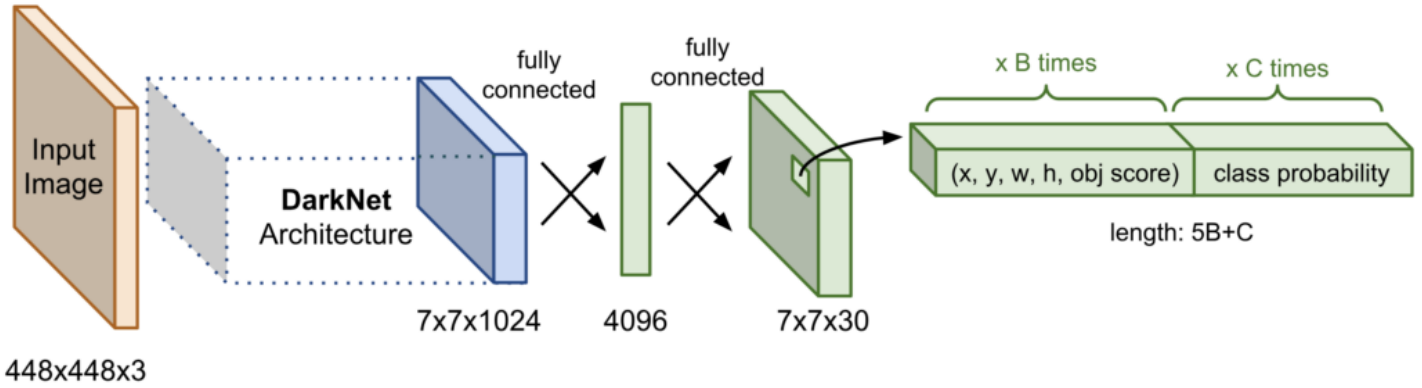
- **单阶段检测：**通过一个卷积神经网络直接预测图像中的目标类别和位置。
- **全图检测：**将输入图像分成网格，每个网格负责检测该网格中的目标。
- **多尺度预测：**使用多个尺度的特征图来检测不同尺寸的目标，这有助于提高检测的准确性。
- **损失函数：**综合考虑了位置误差和类别误差，以及目标置信度的误差。
- **非极大值抑制（NMS）：**使用NMS算法来筛选最终的检测结果。



单阶段目标检测 – YOLOv1

网络架构

YOLOv1架构受到GoogleNet架构的启发，有24个卷积层和两个全连接层。在这些层中，前二十层充当主干，其余层通向另外两个完全连接的层，充当检测头。



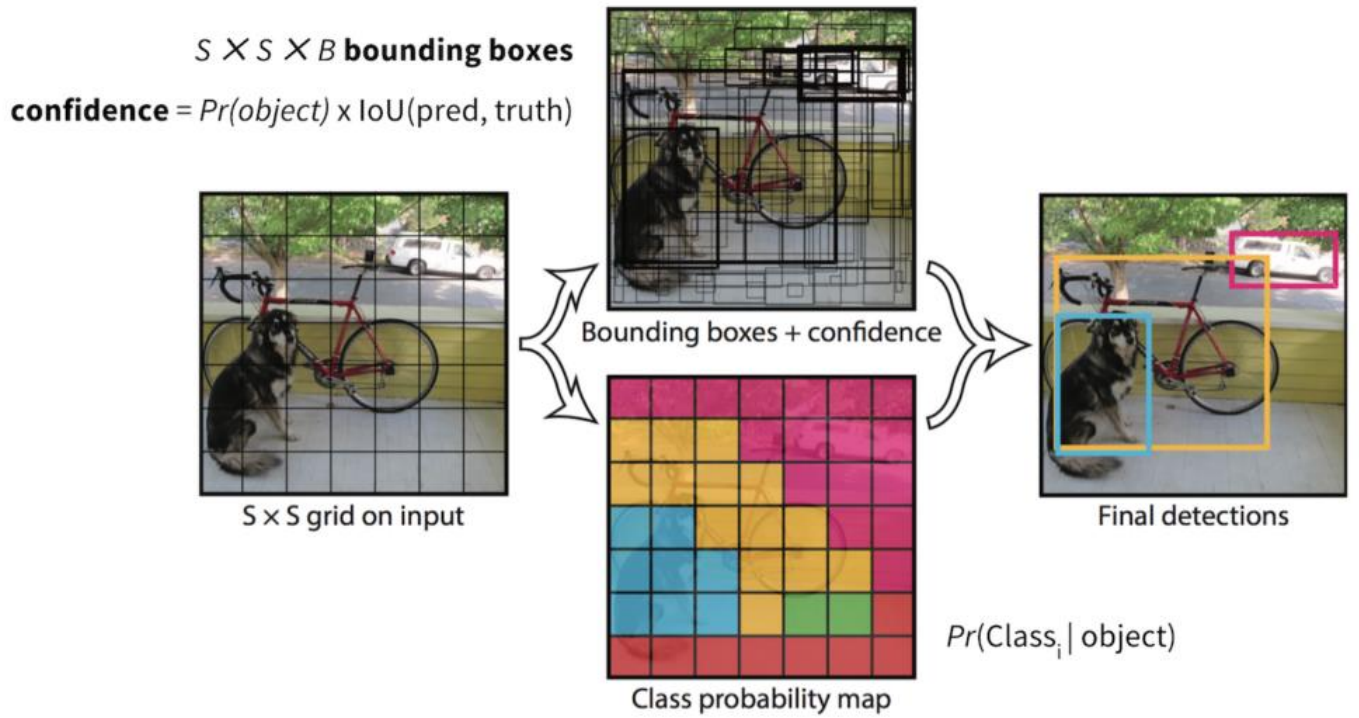
1. 前24个卷积层：

- 前20层：在ImageNet中以224×224的分辨率预训练；
- 其余四层在PASCAL VOC 2012中以448×448的分辨率进行了微调；
- 这增加了模型更准确地检测小物体的信息。

2. 检测头：

由于检测头需要预测边界框坐标、物体分数和物体类别，因此它们的损失函数由三个部分组成：定位损失、置信度损失和分类损失。

- 输入图像被划分为一个S×S网格
- 目标中心所在的单元格负责检测它
- 一个网格单元预测多个边界框
- 每个预测数组由5个元素组成：边界框的中心(x,y)、框的维度(w,h)以及置信度分数。



单阶段目标检测 – YOLOv2

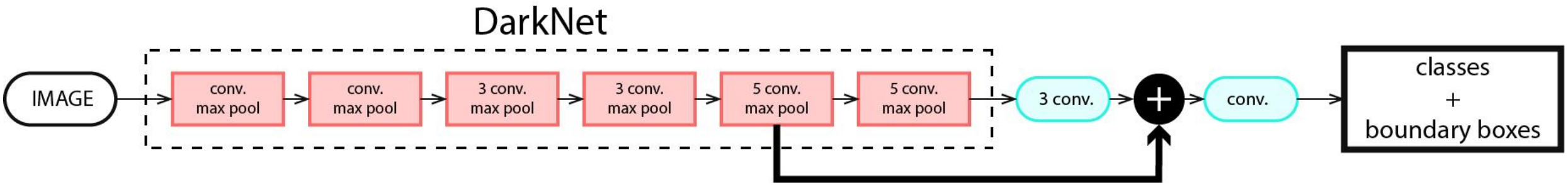
YOLO v2 和 YOLO 9000 是由 J. Redmon 和 A. Farhadi 于 2016 年在题为YOLO 9000: Better, Faster, Stronger的论文中提出的。他们稍微修改了 YOLOv1 架构并改进了训练过程，使 YOLOv2 更快、更准确。

与 YOLOv1 相比的架构变化：

1. 使用更深的神经网络：
YOLOv2用DarkNet-19作为主干网络架构;

2. 批量归一化 (Batch Normalization)：
通过在架构中添加批量归一化，我们可以提高模型的收敛性，从而加快训练速度。这也消除了应用其他类型的归一化（例如 Dropout）而不会过度拟合的需要。与基本 YOLO 相比， 单独添加批量归一化可以使 mAP 增加2% ;

3. 高分辨率分类器：
YOLOv1 在训练期间使用224 * 224 作为输入尺寸，但在检测时，它需要尺寸高达448 * 448的图像，导致 mAP 下降。
YOLOv2 版本在 ImageNet 数据上以更高分辨率(448 * 448)训练10 个 epoch，mAP 增加了4%。

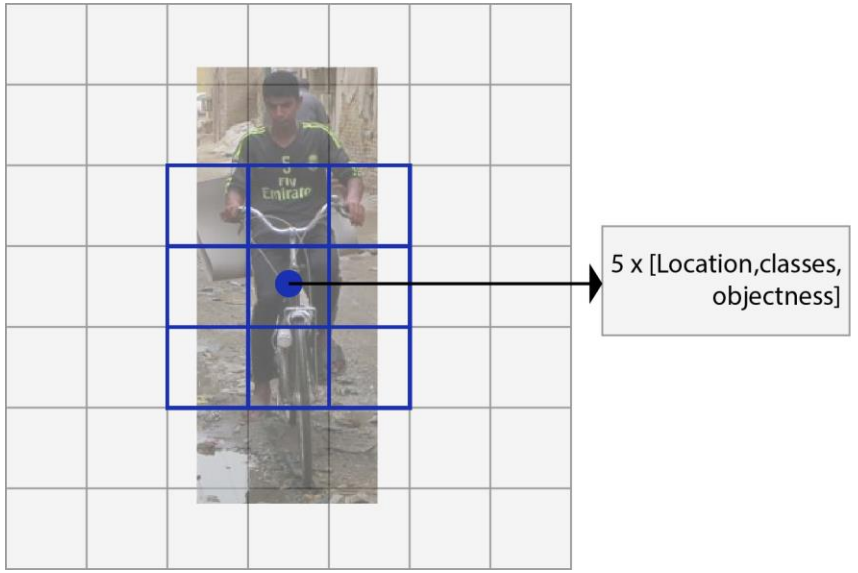


单阶段目标检测 – YOLOv2

与 YOLOv1 相比的架构变化：

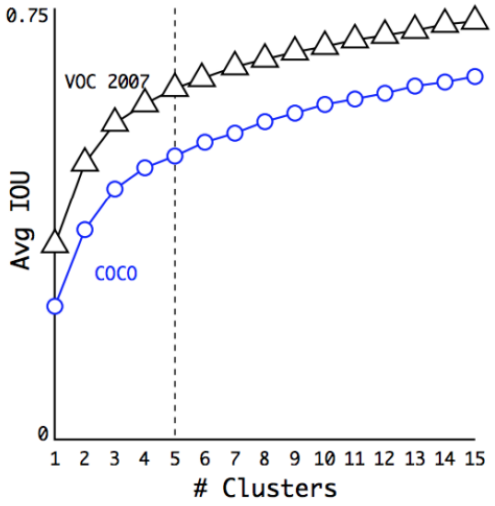
3. 使用锚框作为边界框：

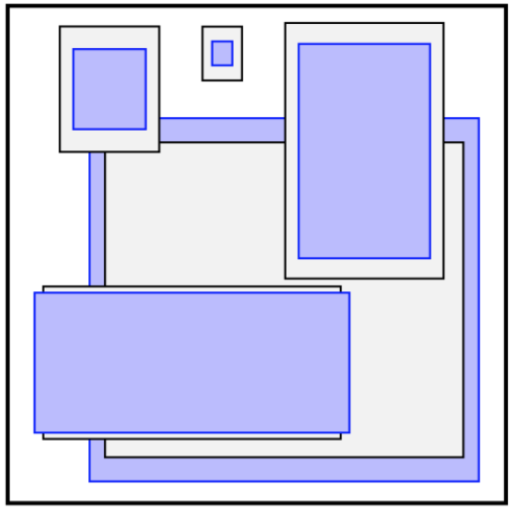
- YOLO 使用全连接层来预测边界框，而不是像 Fast R-CNN、Faster R-CNN 那样直接从卷积网络预测坐标。
- 在 YOLOv2 中，删除了全连接层，而是添加了锚框来预测边界框。
- 将输入的大小从448 * 448更改为416 * 416。当将其下采样 32 倍时，这会创建一个大小为13 * 13的特征图。这背后的想法是，物体很可能位于特征图的中心。
- 删除一个池化层以获得13 * 13空间网络而不是 7*7



4. 维度聚类：

- 需要确定生成的锚点（先验）的数量，识别具有最高准确度的图像的前 K 个边界框。使用 K 均值聚类算法。最大化 IOU 作为该算法的目标。
- YOLO v2 使用 K=5 是为了算法更好的权衡。可以从图中得出结论，随着 K=5 值的增加，精度不会发生显着变化。
  基于K = 5的 IOU 聚类给出的 mAP 为61%。





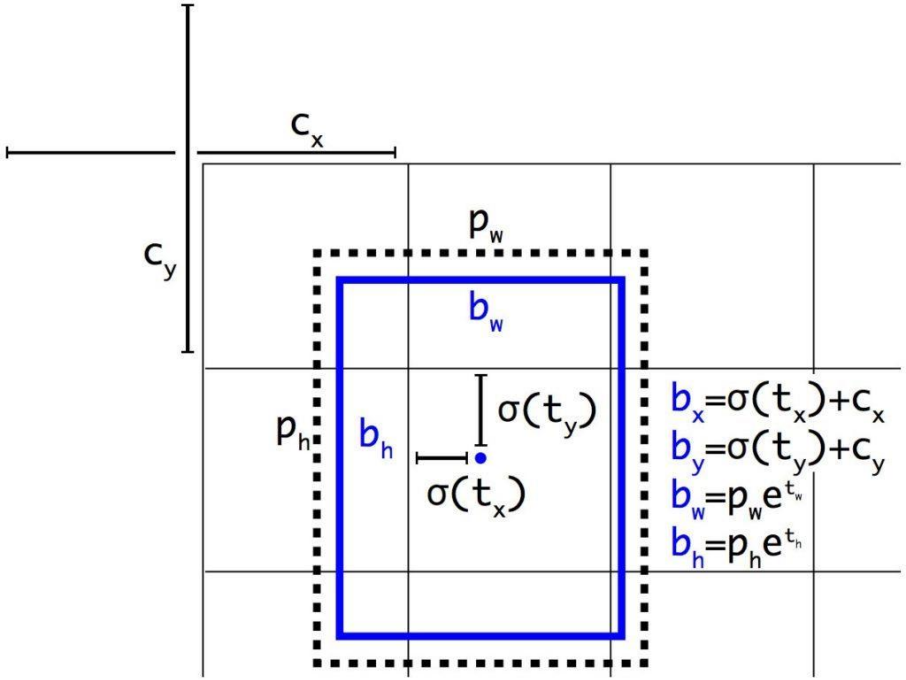
单阶段目标检测 – YOLOv2

与 YOLOv1 相比的架构变化：

5. 直接定位问题：
YOLO之前的版本没有对位置预测的约束，导致早期迭代不稳定。
YOLOv2 预测 5 个参数（tx、 ty 、 tw、 th 、to（客观性得分）），并应用 sigma 函数来约束其值落在 0 和 1 之间。 这种直接位置约束使 mAP 增加了5%。

6. 细粒度特征：
YOLOv2 生成13 * 13足以检测大物体。然而，如果想检测更精细的物体，可以修改架构，使前一层的输出 26 * 26 * 512 变为 13 * 13 * 2048，并与原始的 13 * 13 * 1024 输出层连接，使我们的输出层的大小。

7. 多尺度训练：
YOLO v2已使用32的步长对从 320 * 320 到 608 * 608 的
不同输入大小进行了训练。该架构每 10 个批次随机选择图像尺寸。可以在精度和图像大小之间进行权衡。例如，图像大小为288 * 288、90 FPS的YOLOv2提供的 mAP 与 Fast R-CNN 一样多。



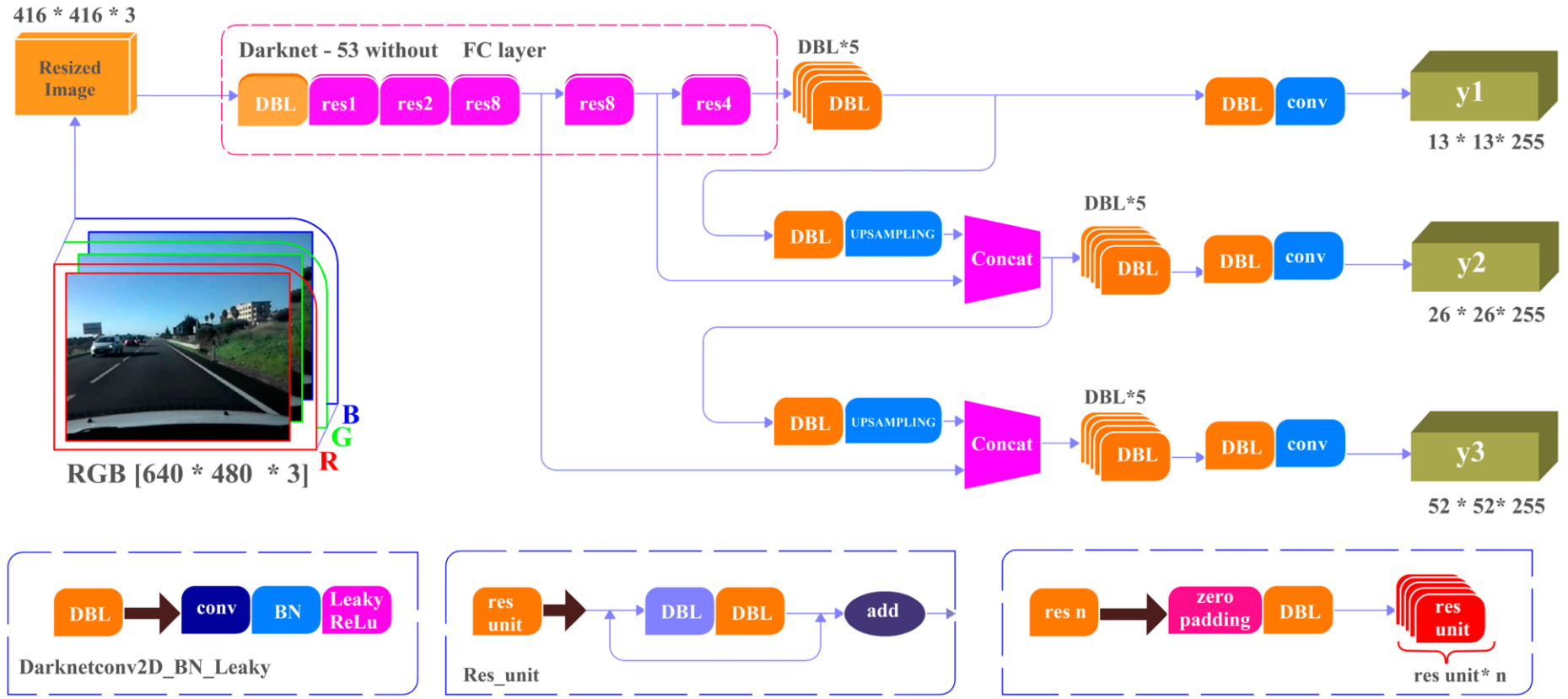
	YOLO							YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓
anchor boxes?				✓	✓			
new network?					✓	✓	✓	✓
dimension priors?						✓	✓	✓
location prediction?						✓	✓	✓
passthrough?							✓	✓
multi-scale?								✓
hi-res detector?								✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	**78.6**

单阶段目标检测 – YOLOv3

YOLOv3（You Only Look Once version 3）是一种流行的目标检测算法，由Joseph Redmon和Ali Farhadi开发。它是YOLO系列中的第三个版本，于2018年发布。用更大的Darknet-53网络替换了原始的Darknet-19特征提取网络。

关键特点和改进：

- 多尺度预测
- Darknet-53
- Bounding Box预测
- 非极大值抑制（NMS）
- Anchor Boxes

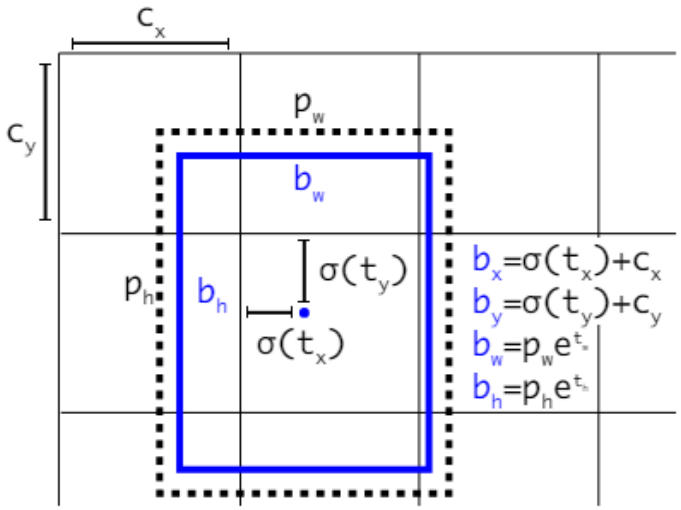


单阶段目标检测 – YOLOv3

与 YOLOv2 相比的架构变化：

- **Bounding Box预测：**在 YOLOv2 中，模型预测每个网格的对象类。但是，在 YOLOv3 中，模型为每个预测的边界框预测类别。该模型为每个网格预测 3 个边界框、客观性得分和类别预测。在输出侧，张量为 N x N x (3 x (4 + 1 + 80))。该模型输出三种不同尺度的边界框。

- **Darknet-53：**YOLOv3 使用DarkNet-53作为特征提取的主干。该架构具有替代的 1×1 和 3×3 卷积层以及受 ResNet 模型启发的跳跃/残差连接。他们还添加了FPN的想法，以利用网络早期所有先前计算和细粒度特征的优势。



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

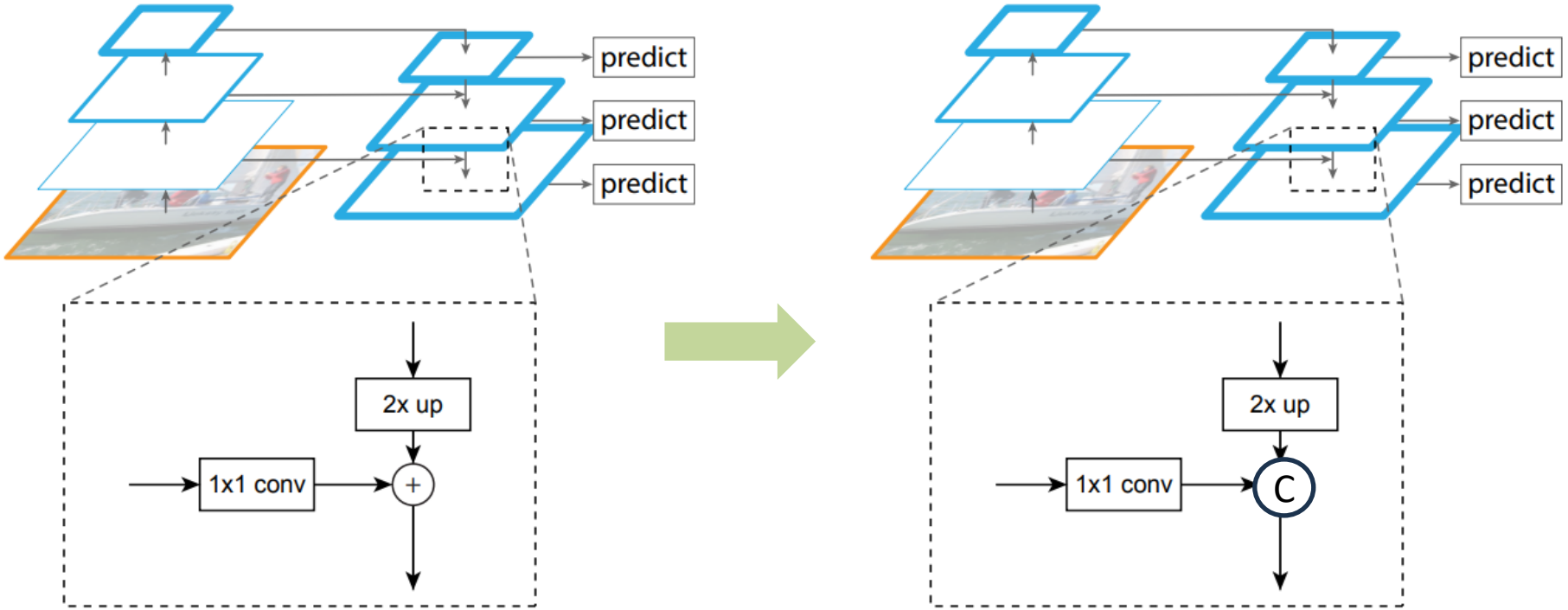
Darknet-53

单阶段目标检测 – YOLOv3 - FPN

- **YOLOv3的FPN：**

为了进一步降低模型的复杂度进而提升速度，YOLOv3选择了重用Backbone所提取的不同Level的特征图，主要是8倍、16倍以及32倍下采样的特征图，同时采用了FPN 的设计思想，分别对16倍、32倍以及各自上采样后的结果进行了融合，但是也对其进行了一定的改进，就是将特征融合的操作由 Add 改为了 Concat。

检测不同尺度的物体具有挑战性，特别是对于小物体。我们可以使用同一图像不同比例的金字塔来检测对象。
特征金字塔网络（FPN）是专为此类金字塔概念而设计的特征提取器，同时考虑到准确性和速度。它取代了 Faster R-CNN 等检测器的特征提取器，并生成多个特征图层（多尺度特征图）。

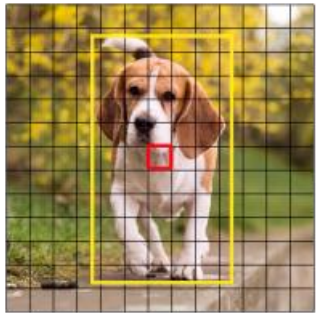


单阶段目标检测 – YOLOv3 - Anchor机制

- **Anchor Boxes，多尺度预测：**与 YOLOv2 类似，YOLOv3 也使用 k-means 来查找锚点之前的边界框。在这个模型中，他们使用了三个不同尺度的先验框，与 YOLOv2 不同。

  如果输入的是416×416的3通道图像，YOLOv3会产生3个尺度的特征图，分别为：13×13、26×26、52×52。对于每个Grid Cell，对应3个Anchor Box，于是，最终产生了(13×13+26×26+52×52)×3=10647个预测框。
  其中不同尺度特征图对应的预测框相对预测的目标大小规模也不一样，具体如下：
  - 13×13 预测大目标
  - 26×26 预测中目标
  - 52×52 预测小目标

Prediction Feature Maps at different Scales



13 x 13

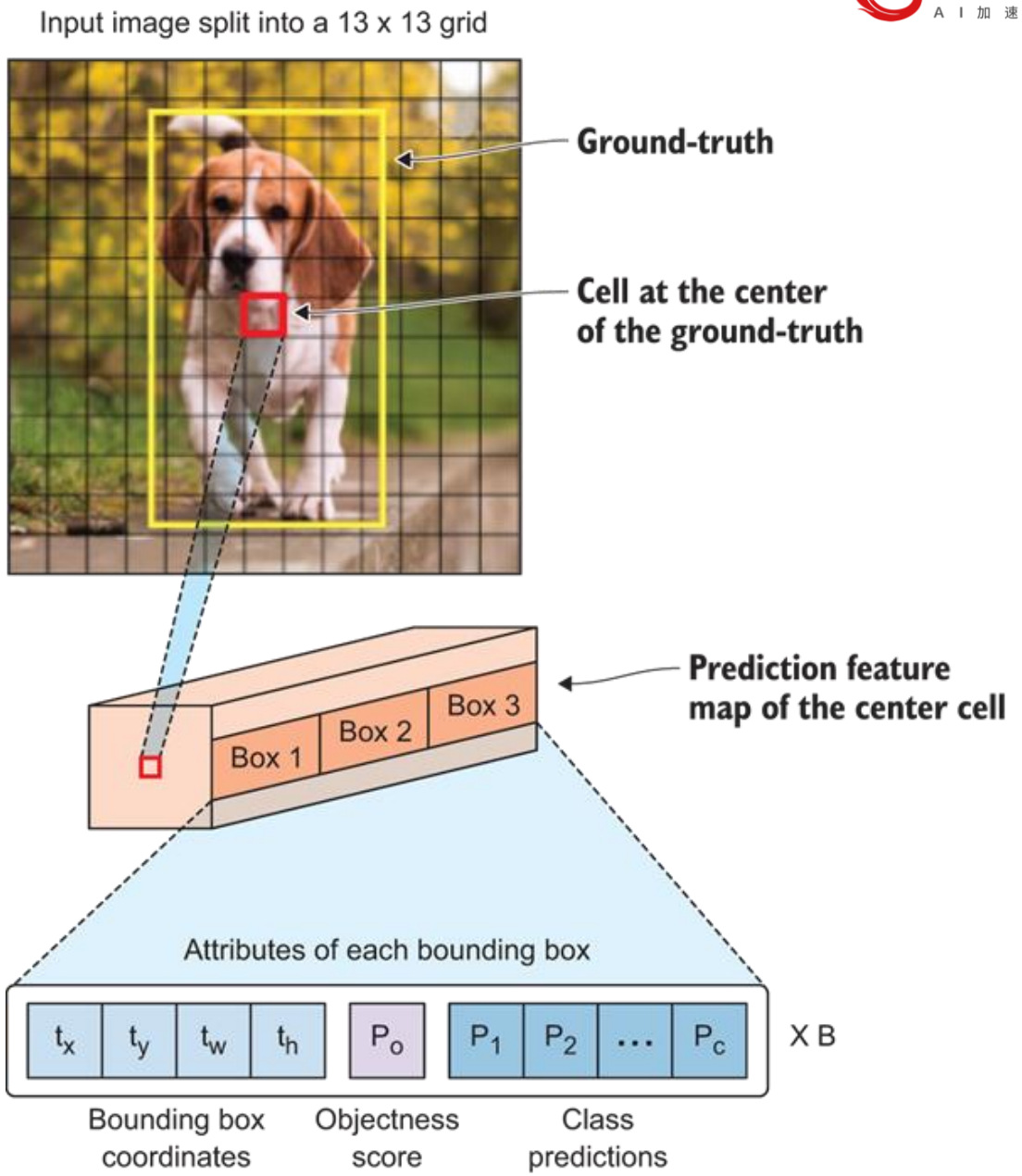
26 x 26

52 x 52

单阶段目标检测 – YOLOv3 - Anchor机制

对于每个 Grid Cell，其都对应一个85维度的数组：

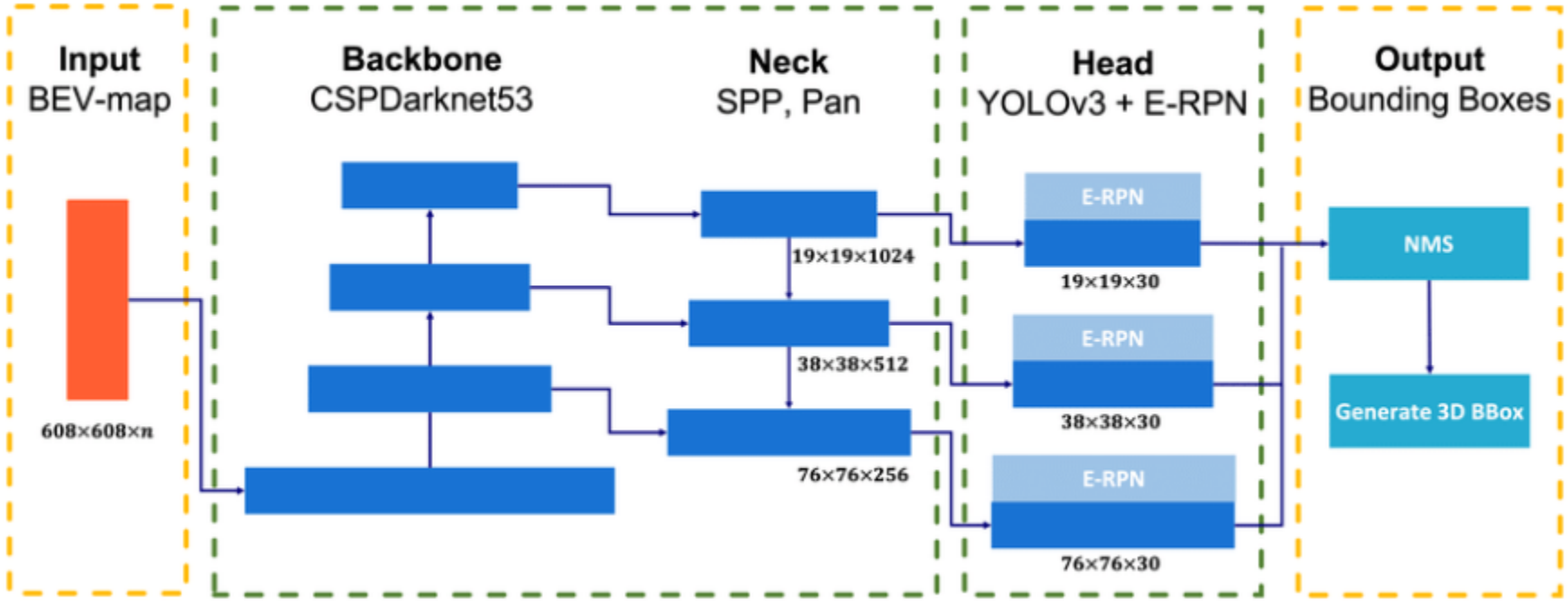
- 80+5=80+4+1也就是数据集的类别数量+坐标值+目标值
- 5=4+1 ： 中心点坐标、宽、高，置信度
- 80：80个类别的类别概率（COCO数据集的类别是80个）；



单阶段目标检测 – YOLOv4

Alexey Bochkovskiy 与 CSPNet（2019 年 11 月）的作者 Chien-Yao Wang 和 Hong-Yuan Mark Liao 合作开发了 YOLOv4。 YOLOv4 与其前身的唯一相似之处在于它是使用 Darknet 框架构建的。他们尝试了许多新想法，并随后将其发表在另一篇论文Scaled-YOLOv4中。

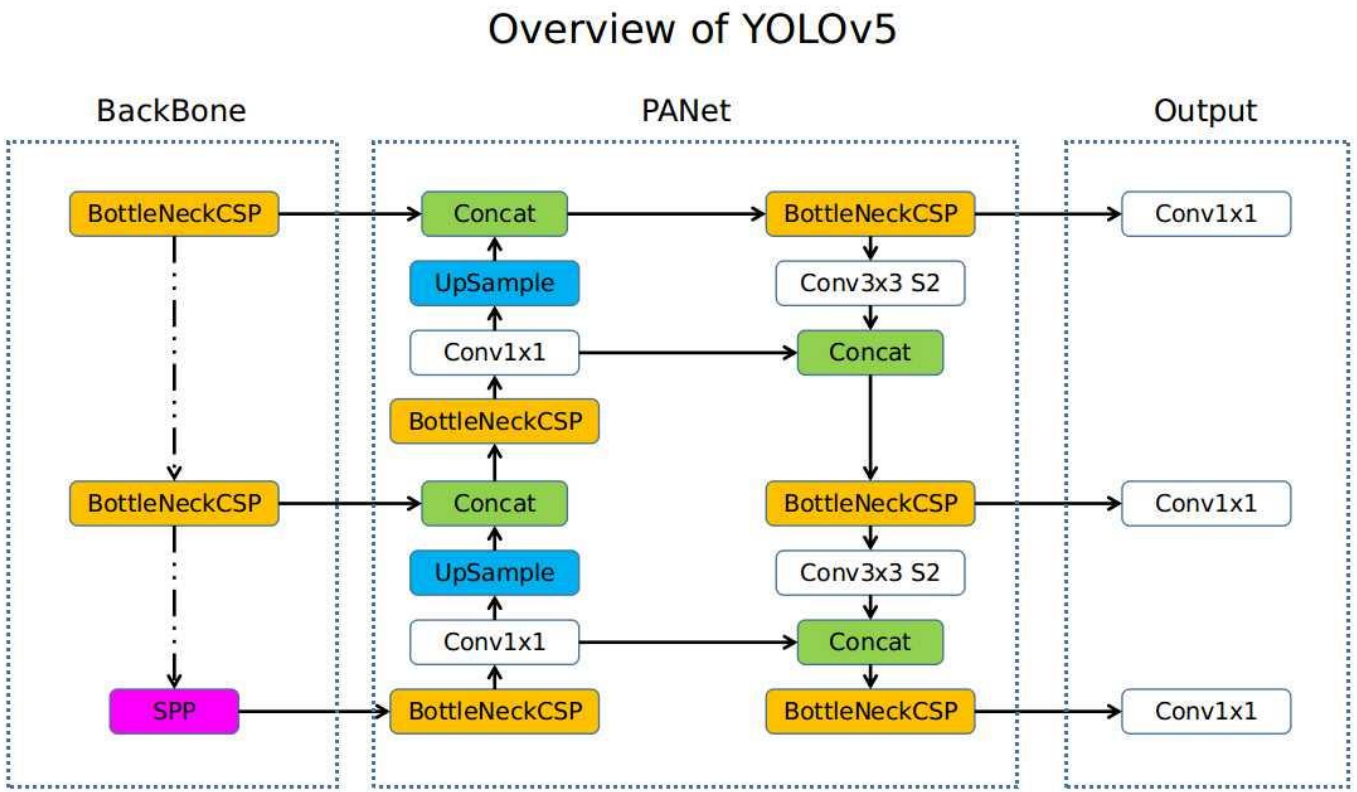
- YOLOv4使用了诸如数据增强技术、正则化方法、类标签平滑、CIoU-loss、Cross mini-Batch Normalization (CmBN)、自对抗训练、余弦退火算法等来改进其最终的性能。
- YOLOv4还使用了包括Mish激活函数、跨阶段部分连接（CSP）、SPP-Block、PAN路径聚合块、多输入加权残差连接（MiWRC）等。
- 使用遗传算法搜索超参数。



单阶段目标检测 – YOLOv5

YOLOv5模型分为3个部分，Backbone、Neck、Head。每个卷积之后都进行批量归一化 (BN) 和 SiLU 激活。以下是YOLOv5 模型架构概述：

- 骨干网：**CSP-DarkNet53**
- 颈部：**SPPF 和 CSP-PANet**
- 头部：**YOLOv3 头部**



单阶段目标检测 – YOLOv5

1. 骨干网络是一个修改过的CSPDarknet53：

- 它以一个具有较大窗口大小的步幅卷积层作为起始（称为Stem），以减少内存和计算成本；
- 随后是卷积层，从输入图像中提取相关特征。
- SPPF（空间金字塔池化快速）层和接下来的卷积层在不同尺度上处理特征，而上采样层增加特征图的分辨率。SPPF层旨在通过将不同尺度的特征汇集到固定大小的特征图中加速网络的计算。每个卷积后面跟着批归一化（BN）和SiLU激活。

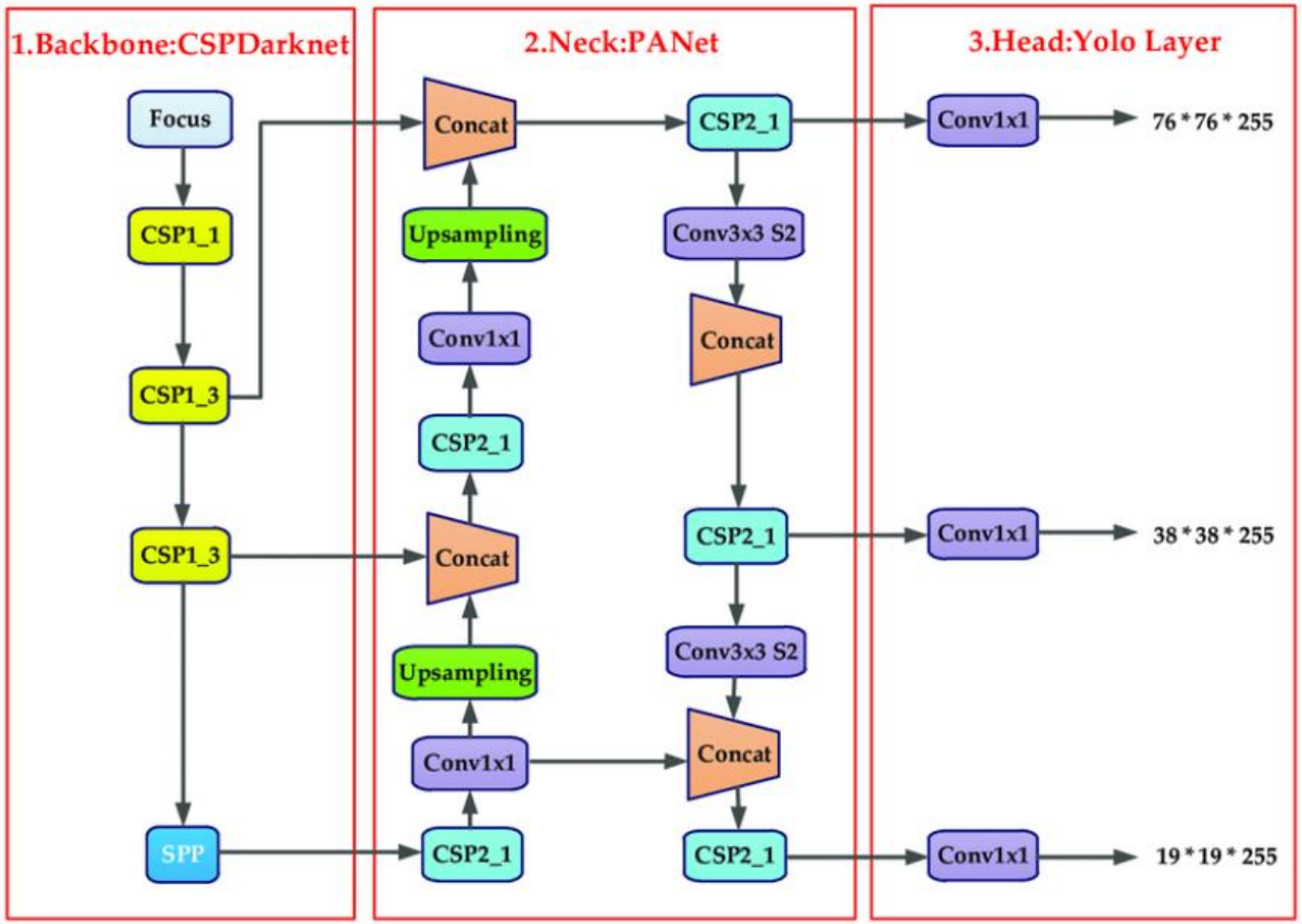
2. 颈部：

- 使用SPPF和修改过的CSP-PAN

3. 头部：

- 类似于YOLOv3。

YOLOv5使用多种数据增强技术，如Mosaic、copy paste、随机仿射、MixUp、HSV增强、随机水平翻转，以及来自albumentations包的其他增强技术。还改进了网格的敏感性，使其更稳定，避免梯度爆炸。

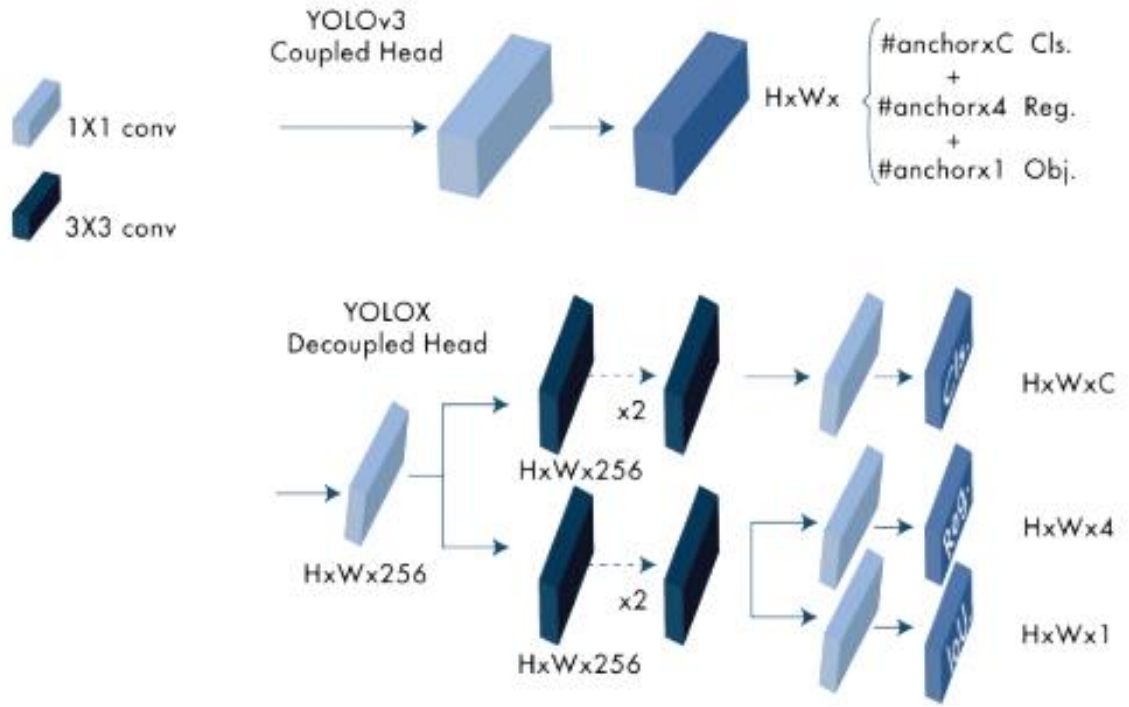


单阶段目标检测 – YOLOv6

YOLOv6 是由美团视觉人工智能部门于2022年9月发布的。该网络设计包括一个具有RepVGG或CSPStackRep块的高效主干网络，一个PAN拓扑结构的颈部，以及一个采用混合通道策略的高效解耦头部。此外，该论文引入了增强的量化技术，使用后训练量化和通道级蒸馏，从而实现更快速、更准确的检测器。总体而言，YOLOv6在准确性和速度等指标上优于先前的最先进模型，如YOLOv5、YOLOX和PP-YOLOE。下图展示了YOLOv6的详细架构。

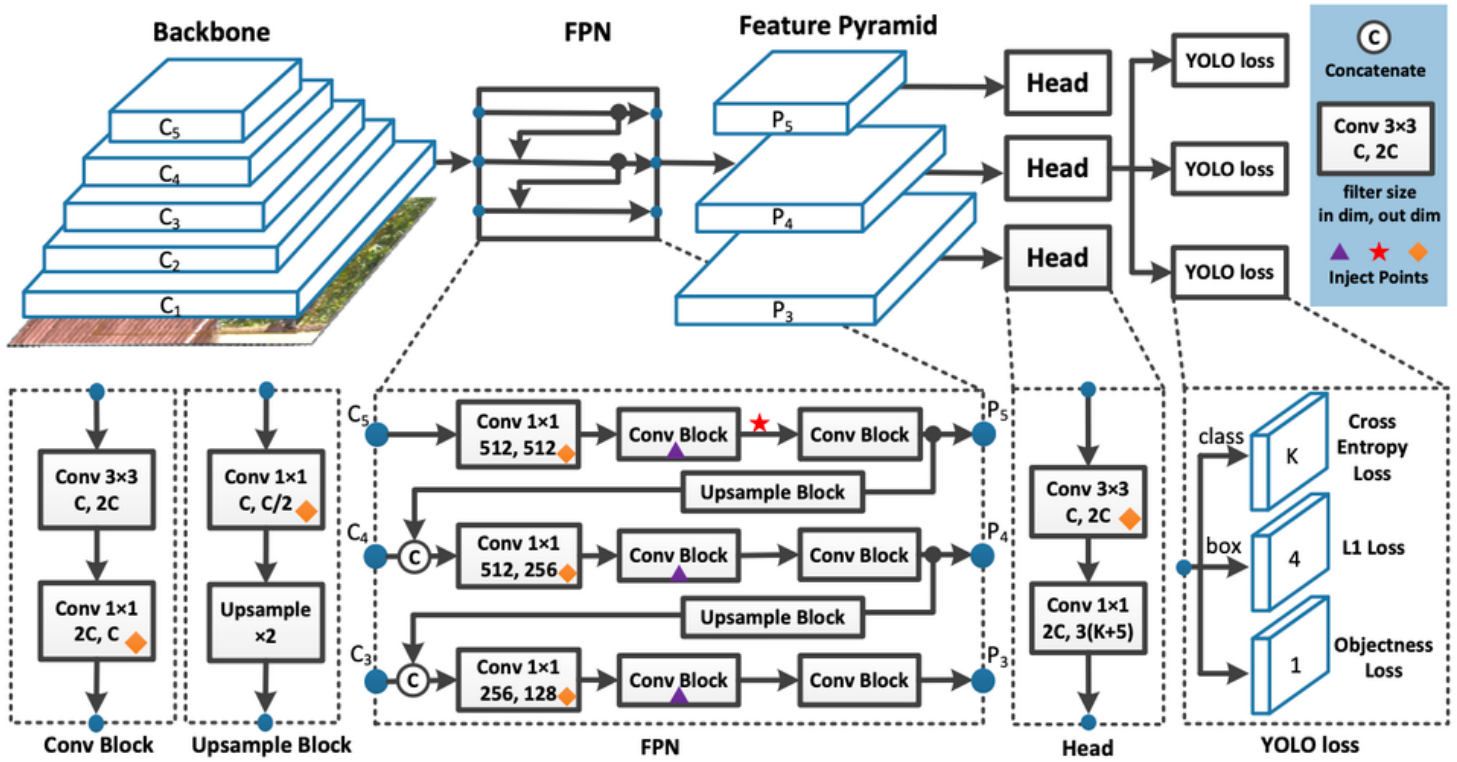
这个模型的主要创新点总结如下：

- 基于RepVGG 的新骨干，称为EfficientRep，比之前的YOLO骨干使用更高的并行性。对于neck部分，他们使用了PAN ，并增强了RepBlocks 或 CSPStackRep Blocks用于更大的模型。在YOLOX之后，他们开发了一个高效的解耦头部。
- 使用TOOD 中引入的任务对齐学习方法进行标签分配。
- 新的分类和回归损失。他们使用了分类VariFocal损失和SIoU/GIoU 回归损失。
- 用于回归和分类任务的自蒸馏策略。
- 使用RepOptimizer和通道级蒸馏的量化方案进行检测，帮助实现更快的检测器。



单阶段目标检测 – YOLOv7

YOLOv7 是 Scaled-YOLOv4 的延续。 YOLOv4 和 YOLOv7 由同一作者发表。 YOLOv7提出了一些架构变化，这些改进提高了准确性，而不影响推理速度，只增加了训练时间。

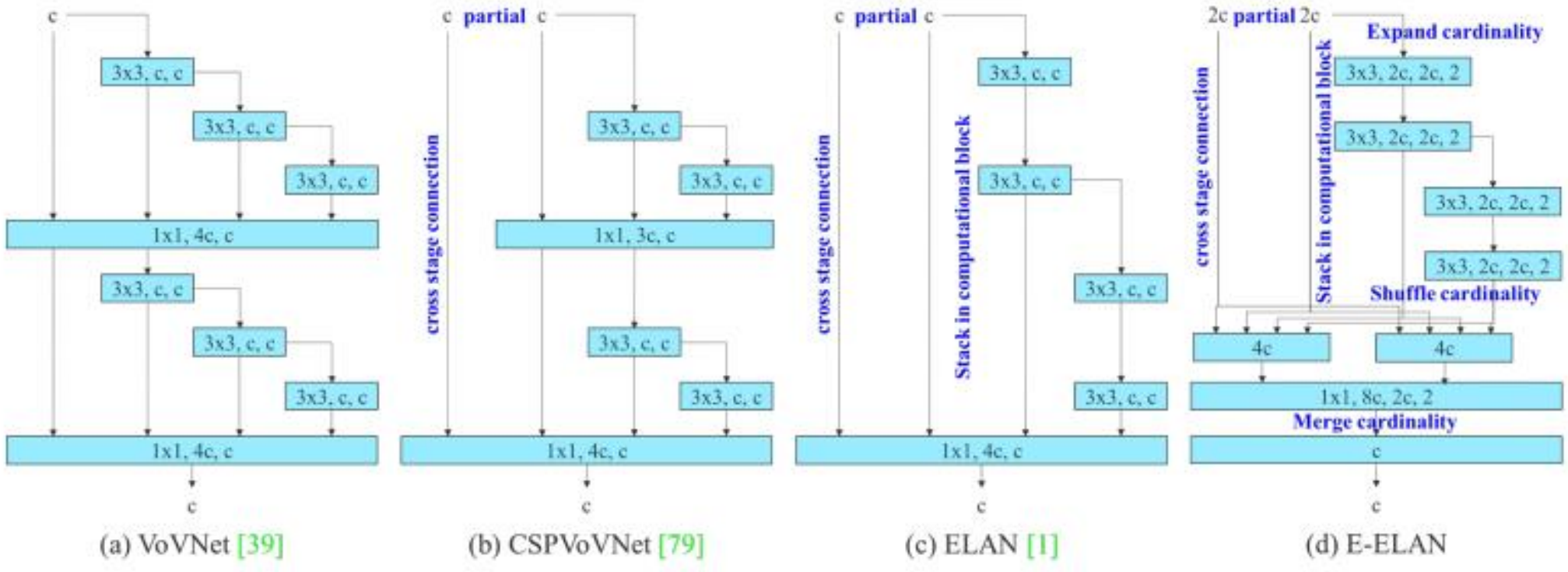


单阶段目标检测 – YOLOv7

YOLOv7 论文中的三个显着贡献：

1. 扩展高效层聚合网络（E-ELAN）

ELAN 是一种策略，通过控制最短最长梯度路径，使深度模型能够更有效地学习和收敛。YOLOv7提出了适用于具有无限堆叠计算块的模型的E-ELAN。E-ELAN通过混洗和合并基数来结合不同组的特征，以增强网络的学习能力，同时不破坏原始梯度路径。

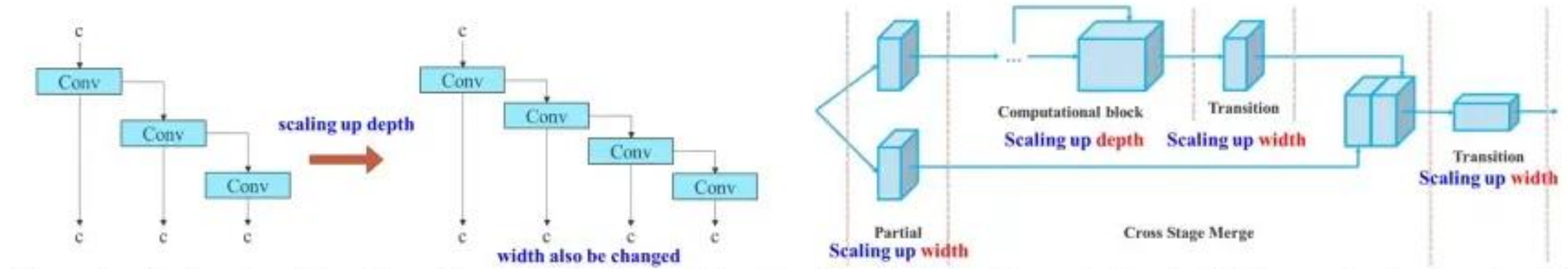


单阶段目标检测 – YOLOv7

YOLOv7 论文中的三个显着贡献：

2. 模型缩放技术

对象检测模型通常以一系列模型的形式发布，尺寸按比例放大或缩小，因为不同的应用程序需要不同级别的精度和推理速度。通常，对象检测模型会考虑网络的深度、网络的宽度以及网络训练的分辨率。在 YOLOv7 中，作者在将各层连接在一起的同时缩放网络深度和宽度。该技术可以在缩放不同尺寸的同时保持模型架构最佳。



(a) concatenation-based model (b) scaled-up concatenation-based model (c) compound scaling up depth and width for concatenation-based model

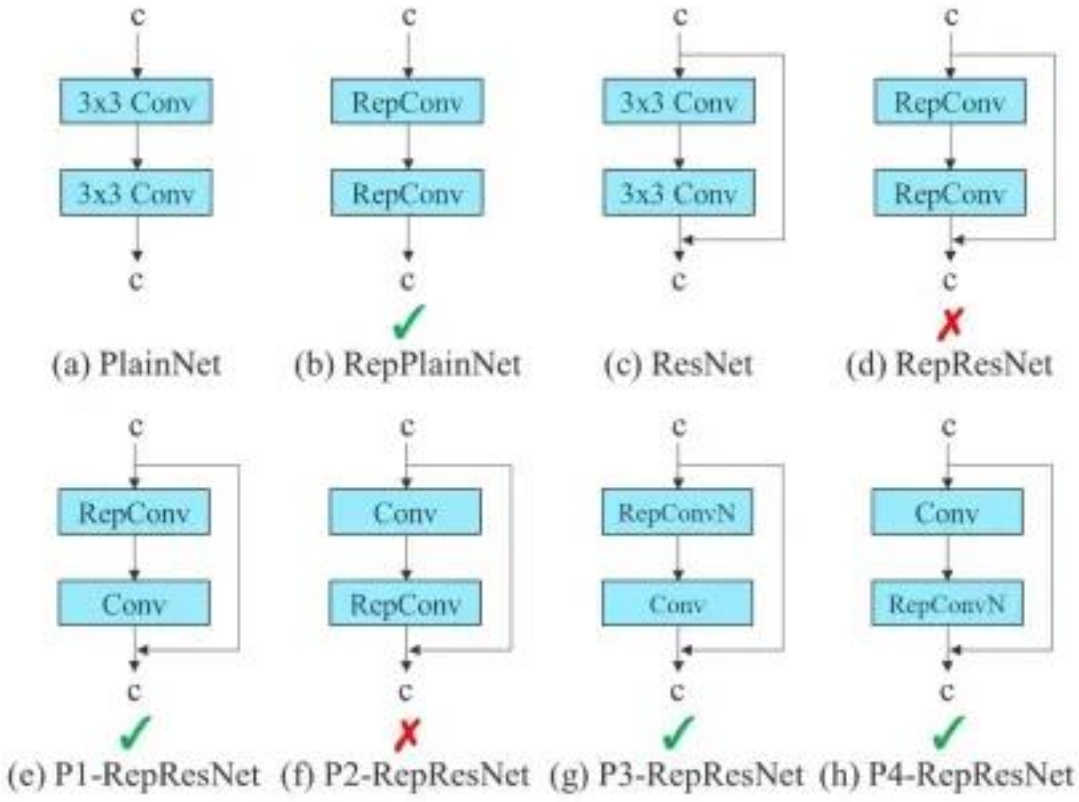
单阶段目标检测 – YOLOv7

YOLOv7 论文中的三个显着贡献：

3. 重新参数化

重新参数化技术涉及对一组模型权重进行平均，以创建一个对于尝试建模的一般模式更稳健的模型。在研究中，最近的重点是**模块**级重新参数化，其中网络的一部分有自己的重新参数化策略。

YOLOv7 作者使用梯度流传播路径来查看网络中的哪些模块应该使用重新参数化策略，哪些不应该使用。



单阶段目标检测 – YOLOv8

YOLOv8 是由开发YOLOv5的公司Ultralytics于2023年1月发布的。YOLOv8提供了五个不同规模的版本：YOLOv8n（nano）、YOLOv8s（small）、YOLOv8m（medium）、YOLOv8l（large）和YOLOv8x（extra large）。YOLOv8支持多种视觉任务，如目标检测、分割、姿态估计、跟踪和分类。

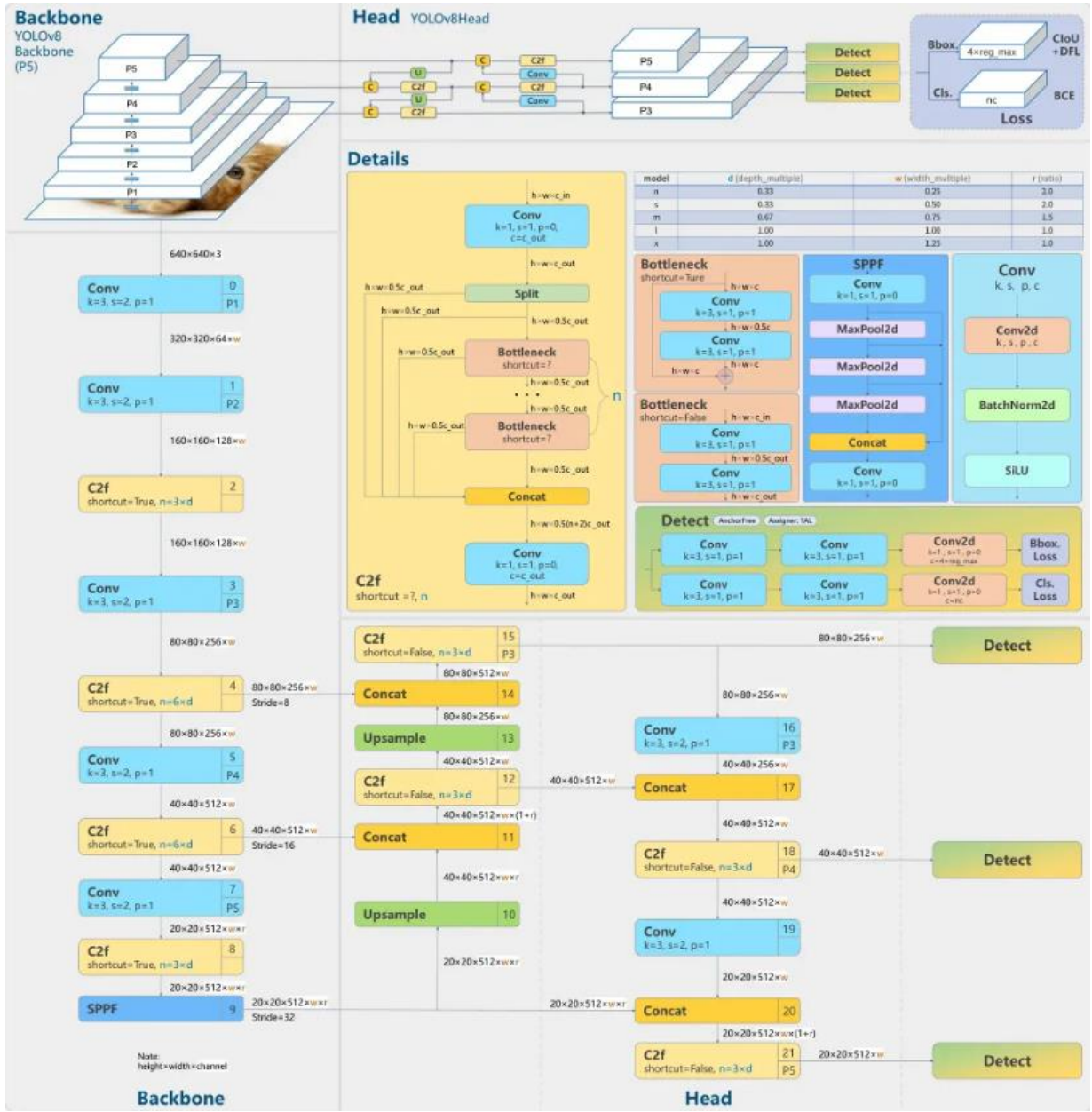
改进点：

- YOLOv8使用一种无锚点模型，具有分离的头部来独立处理目标性、分类和回归任务。这种设计使每个分支能够专注于自己的任务，并提高了模型的整体准确性。

- YOLOv8使用CIoU和DFL损失函数来处理边界框损失，以及二元交叉熵用于分类损失。这些损失改善了目标检测性能，特别是在处理较小目标时。

- YOLOv8还提供了一个称为YOLOv8-Seg的语义分割模型。骨干是CSPDarknet53特征提取器，后跟一个C2f模块，而不是传统的YOLO颈部架构。C2f模块后面是两个分割头，它们学习为输入图像预测语义分割掩模。

- 该模型具有类似于YOLOv8的检测头部，包括五个检测模块和一个预测层。




Thank
You