

基于改进YOLOv5模型的口罩佩戴检测方法

汇报人 廖依璇

CONTENTS



01 YOLOv5架构

02 模型改进

03 实验结果

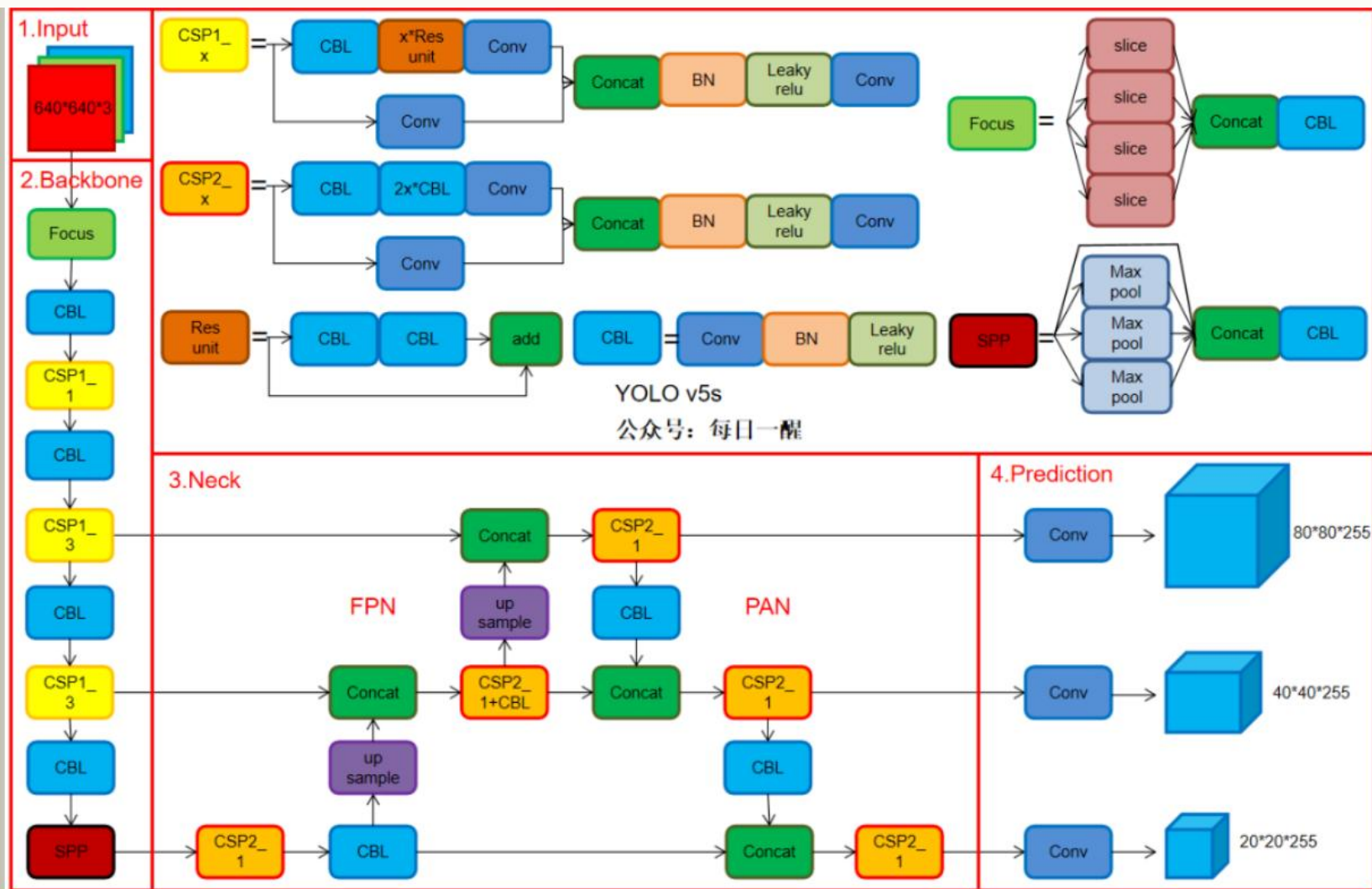


YOLOv5架构



- (1) 考虑了邻域的正样本anchor匹配策略，增加了正样本。
- (2) 通过灵活的配置参数，可以得到不同复杂度的模型。
- (3) 通过一些内置的超参优化策略，提升整体性能。
- (4) 和yolov4一样，都用了mosaic增强，提升小物体检测性能。

YOLOv5架构



Input: Mosaic数据增强、自适应锚框计算、自适应图片缩放

Backbone: Focus结构, CSP结构

Neck: FPN+PAN结构

Prediction: GIoU_Loss



YOLOv5架构-Input

(1) Mosaic数据增强

Yolov4中使用的Mosaic是参考2019年底提出的CutMix数据增强的方式，但CutMix只使用了两张图片进行拼接，而Mosaic数据增强则采用了4张图片，随机缩放、随机裁剪、随机排布的方式进行拼接。



原始图片



CutMix-2019年



Mosaic-2020年



YOLOv5架构-Input

(2) 自适应锚框:

在YOLO V5 中锚定框是基于训练数据自动学习的，此功能嵌入到代码中，每次训练时，自适应的计算不同训练集中的最佳锚框值。

Yolov5 中默认保存了一些针对 coco数据集的预设锚定框，在 yolov5 的配置文件*.yaml 中已经预设了 640×640 图像大小下锚定框的尺寸（以 yolov5s.yaml 为例）：

```
# anchors
anchors:
  - [10,13, 16,30, 33,23] # P3/8
  - [30,61, 62,45, 59,119] # P4/16
  - [116,90, 156,198, 373,326] # P5/32
```



YOLOv5架构-Input

```
# anchors
anchors:
  - [10,13, 16,30, 33,23] # P3/8
  - [30,61, 62,45, 59,119] # P4/16
  - [116,90, 156,198, 373,326] # P5/32
```

anchors参数共有三行，每行9个数值；且每一行代表应用不同的特征图；

- 1、第一行是在最大的特征图上的锚框
- 2、第二行是在中间的特征图上的锚框
- 3、第三行是在最小的特征图上的锚框；

在目标检测任务中，一般希望在大的特征图上去检测小目标，因为大特征图才含有更多小目标信息，因此大特征图上的[anchor](#)数值通常设置为小数值，而小特征图上数值设置为大数值检测大的目标。

在网络训练中，网络在初始锚框的基础上输出预测框，进而和真实框groundtruth进行比对，计算两者差距，再反向更新，迭代网络参数。



(3) 自适应图片缩放

在项目实际使用时，很多图片的长宽比不同，因此缩放填充后，两端的黑边大小都不同，而如果填充的比较多，则存在信息冗余，影响推理速度。

Yolov5的代码中datasets.py的letterbox函数中进行了修改，对原始图像自适应的添加最少的黑边。图像高度上两端的黑边变少了，在推理时，计算量也会减少，即目标检测速度会得到提升。

主要步骤为：计算缩放比例、计算缩放后的尺寸、计算黑边填充数值

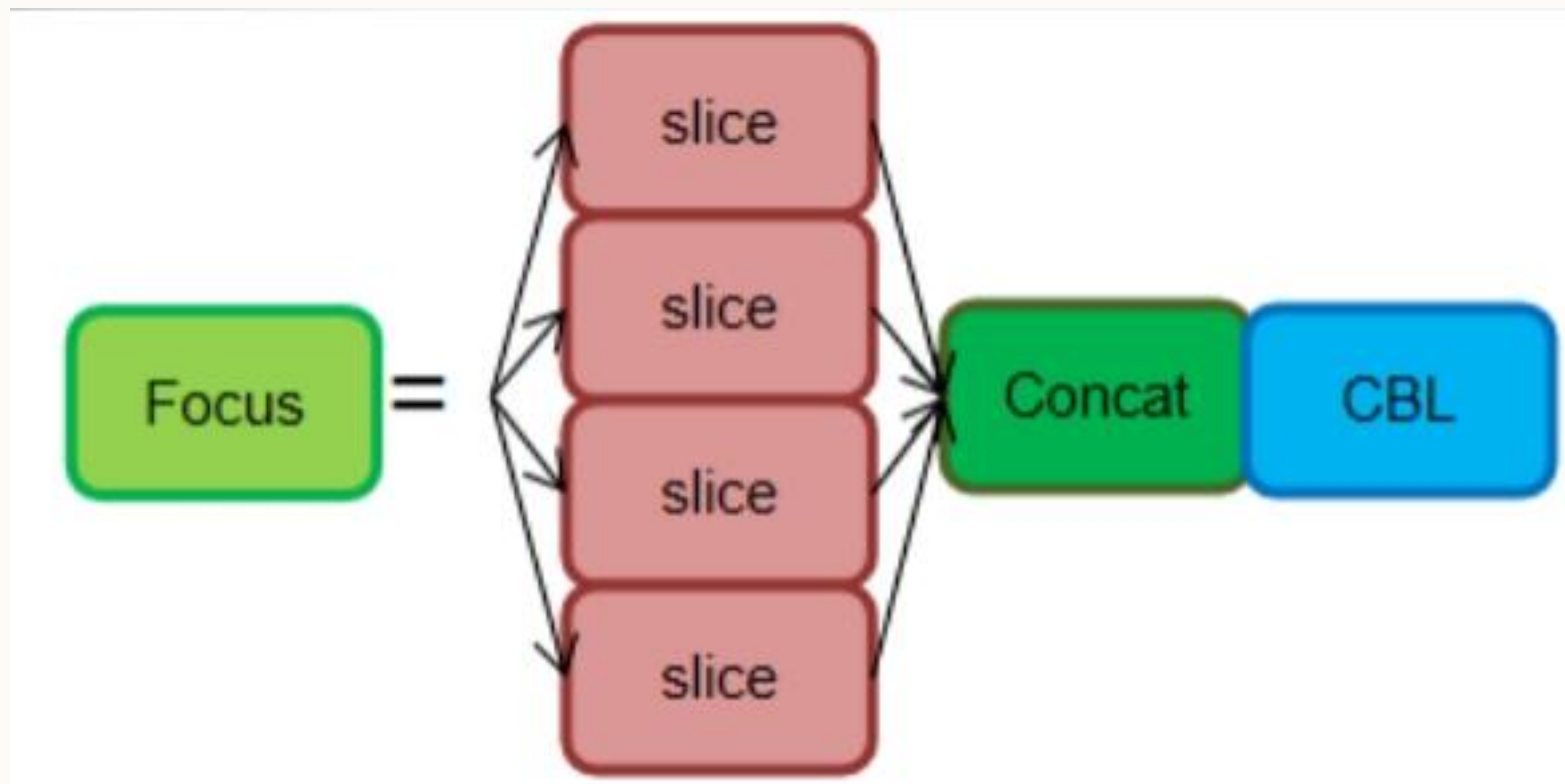
长*宽: 800*600	第一步:	$\begin{cases} 416/800=0.52 \\ 416/600=0.69 \end{cases}$	长*宽: 416*416
			知乎 @江大白
长*宽: 800*600	第二步:	$\begin{cases} 800*0.52=416 \\ 600*0.52=312 \end{cases}$	长*宽: 416*312
			知乎 @江大白
长*宽: 800*600	第三步:	$\begin{cases} 416-312=104 \\ \text{np.mod}(104,32)=8 \\ 8/2=4 \end{cases}$	长*宽: 416*320 (312+8)
			知乎 @江大白



YOLOv5架构-Backbone

Focus:

Focus模块先采用切片操作将一张特征图切成四张特征图，然后将切分的特征图按通道维度进行拼接。最后使用n个卷积核对拼接的特征图进行卷积操作。其最大好处是可以最大程度的减少信息损失而进行下采样操作。



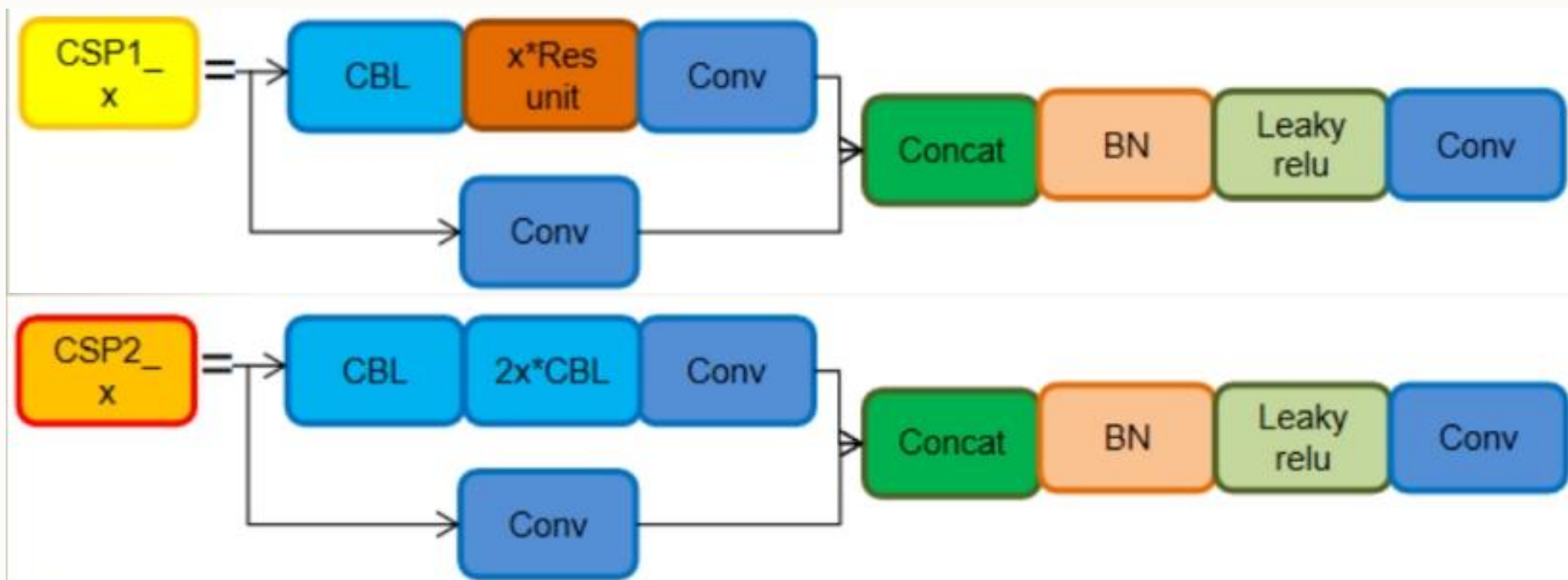


YOLOv5架构-Backbone

CSP (跨阶段局部网络)

跨阶段局部网络缓解以前需要大量推理计算的问题。

以Yolov5s网络为例，CSP1_X结构应用于Backbone主干网络，另一种CSP2_X结构则应用于Neck中。



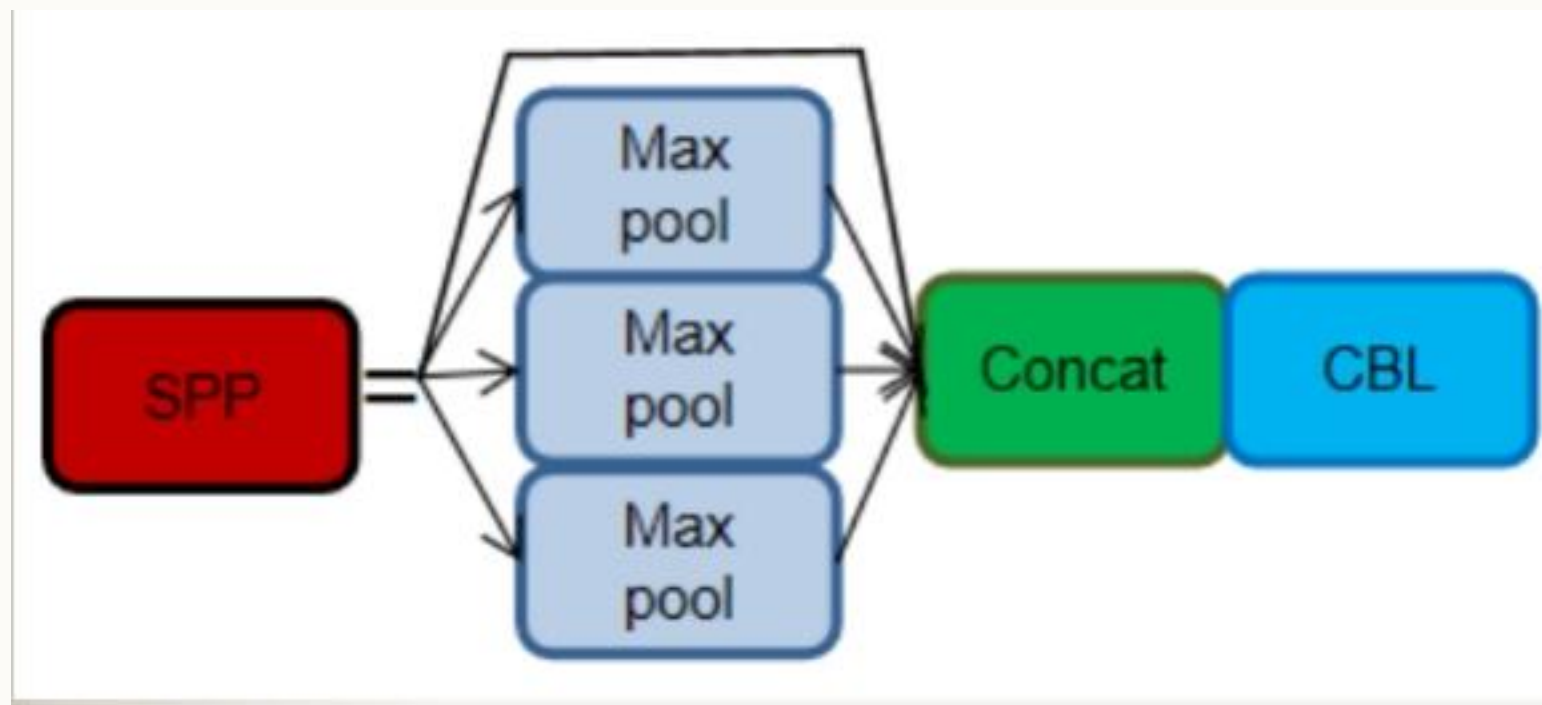


YOLOv5架构-Backbone

SPP:

SPP模块（空间金字塔池化模块），分别采用5、9、13的最大池化，再进行concat融合，提高感受野。

SPP的本质就是多层maxpool，只不过为了对于不同尺寸大小 $a \times a$ 的 featur map 生成固定大小 $n \times n$ 的输出。





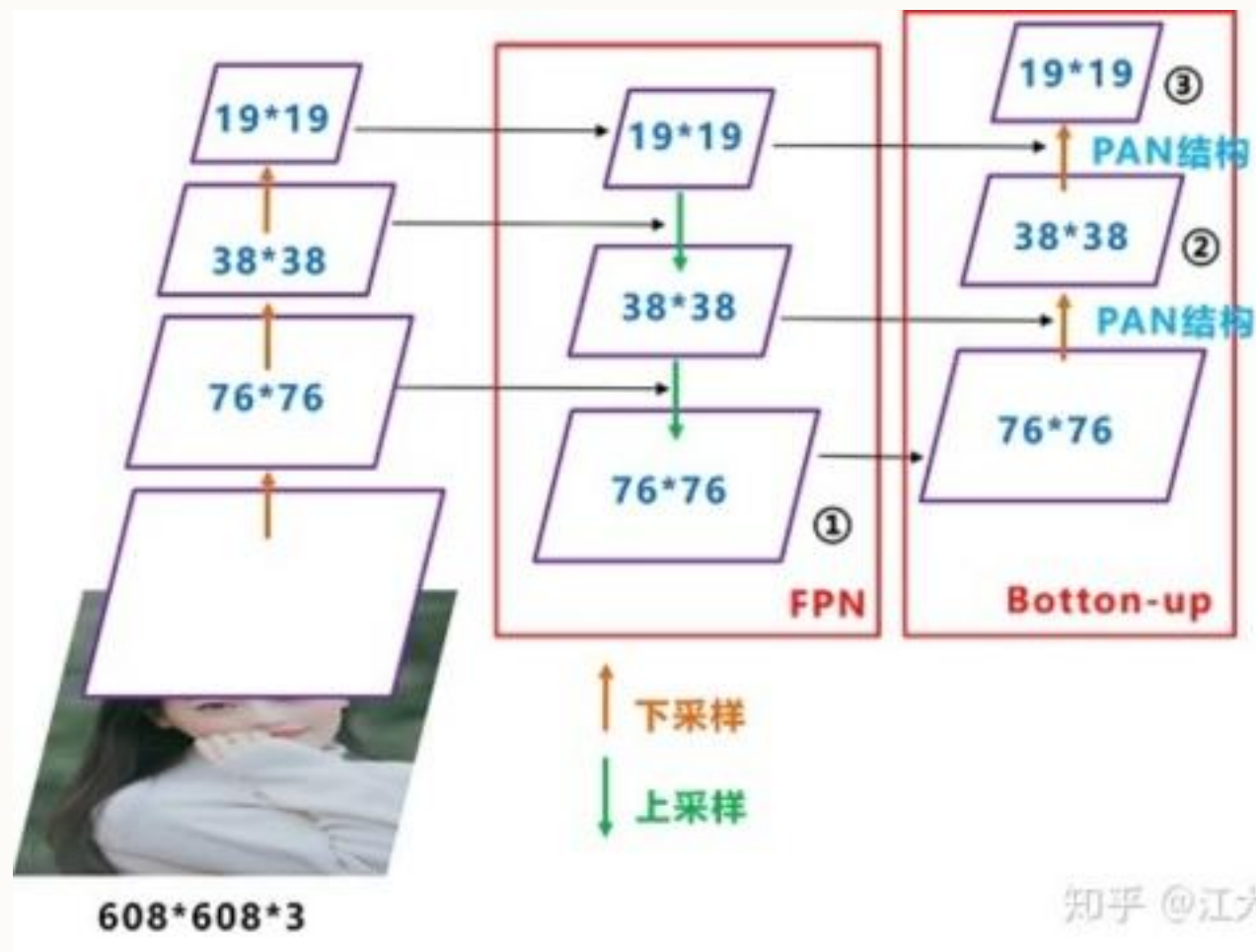
Neck是目标检测框架中承上启下的关键环节。它对Backbone提取到的重要特征，进行再加工及合理利用，有利于下一步head的具体任务学习，如分类、回归、keypoint、instance mask等常见的任务。

Yolov5现在的Neck和Yolov4中一样，都采用FPN+PAN的结构。Yolov5的Neck结构中，采用借鉴CSPnet设计的CSP2结构，加强网络特征融合的能力。



YOLOv5架构-Neck

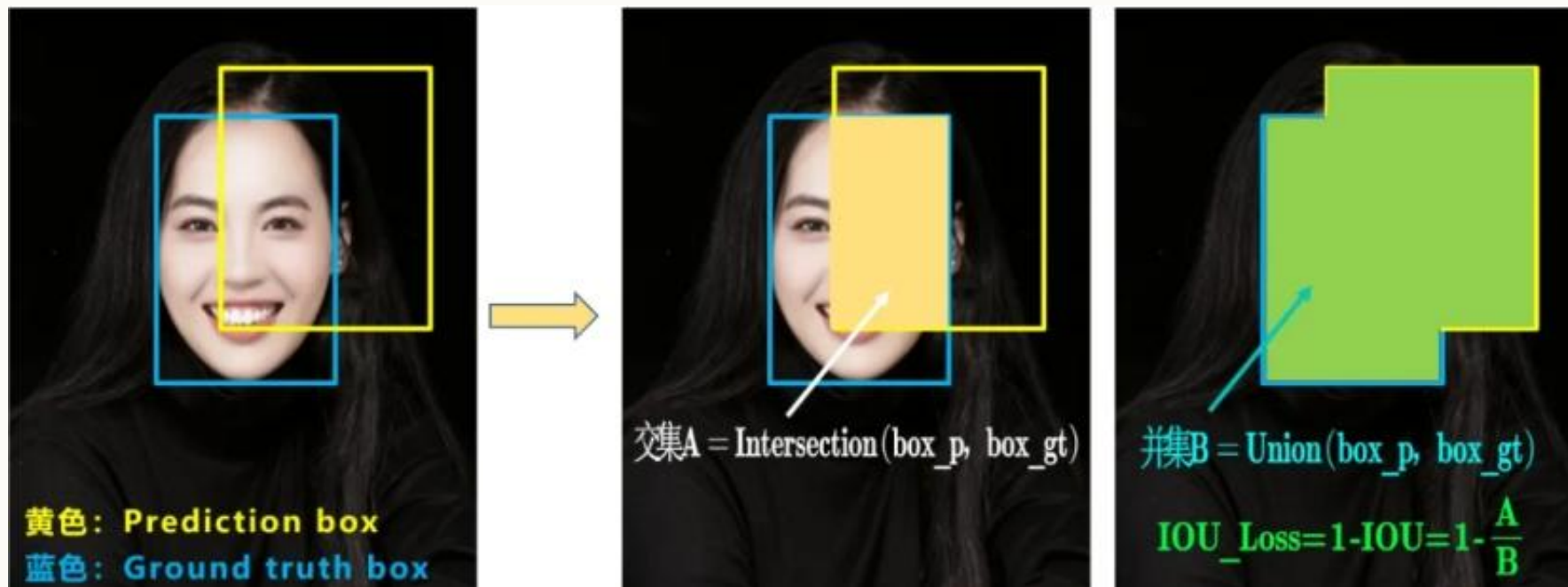
FPN层自顶向下传达强语义特征，而特征金字塔则自底上传达强定位特征，两两联手，从不同的主干层对不同的检测层进行参数聚合。





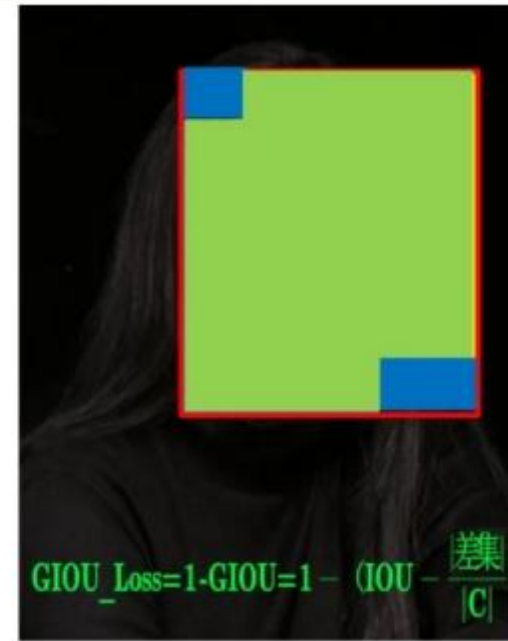
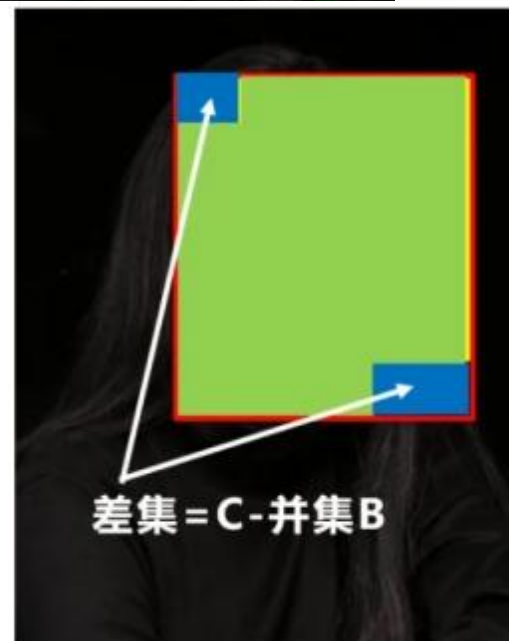
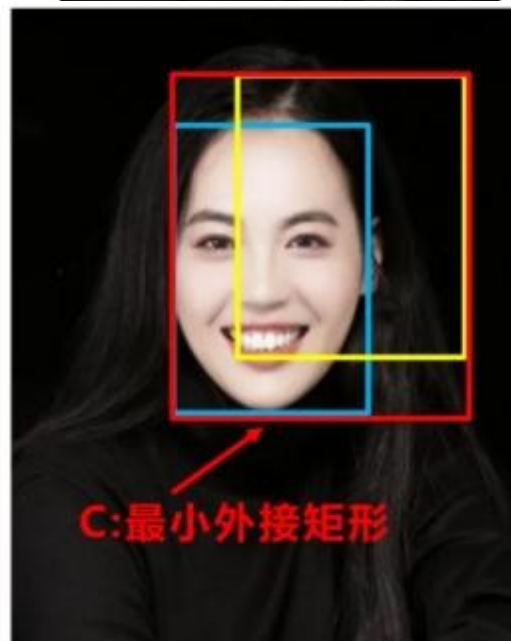
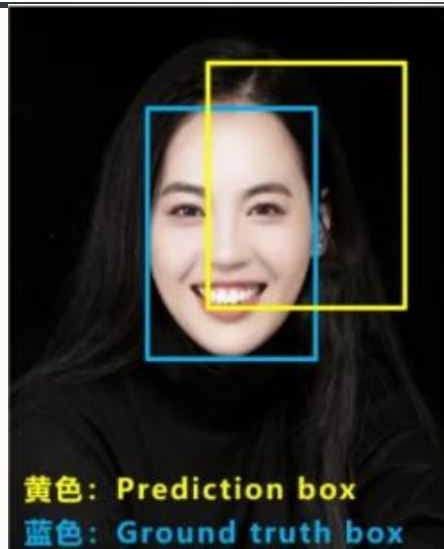
YOLOv5架构-Prediction:GIOW_Loss

Yolov5采用GIOW_Loss做Bounding box的损失函数，使用 二进制交叉熵(BCE) 和 Logits 损失函数 计算类概率和目标得分的损失。不相交时， $IOU=0$ ，两个框距离变换， IOU loss不变，改进为GIOW。GIOW Loss则是在 IOU 的基础上引入了预测框和真实框的最小外接矩形。





YOLOv5架构-Prediction:GIOW_Loss

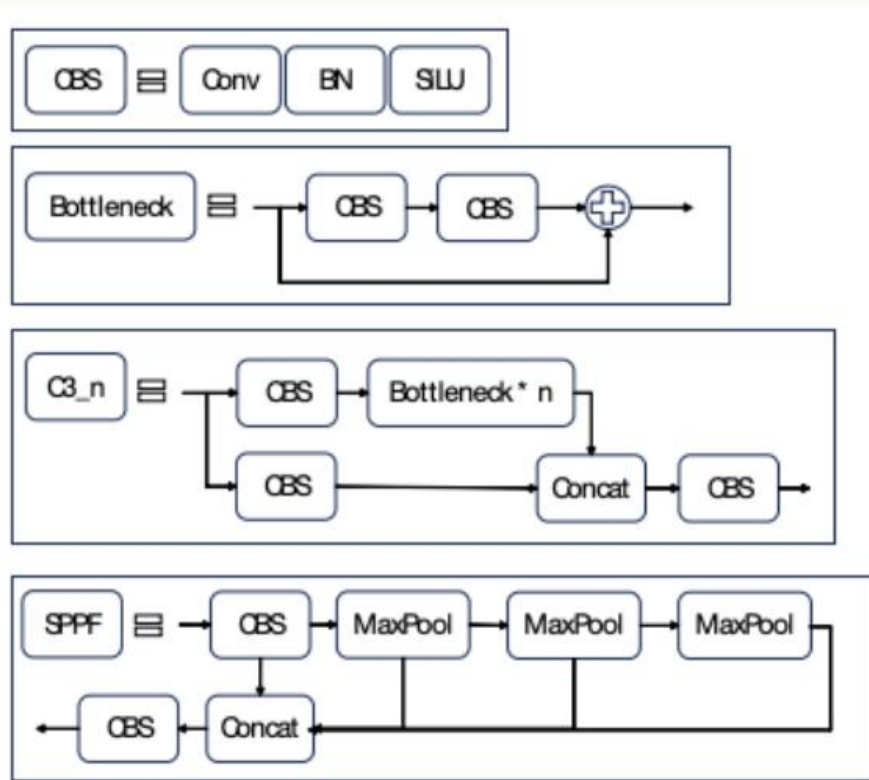




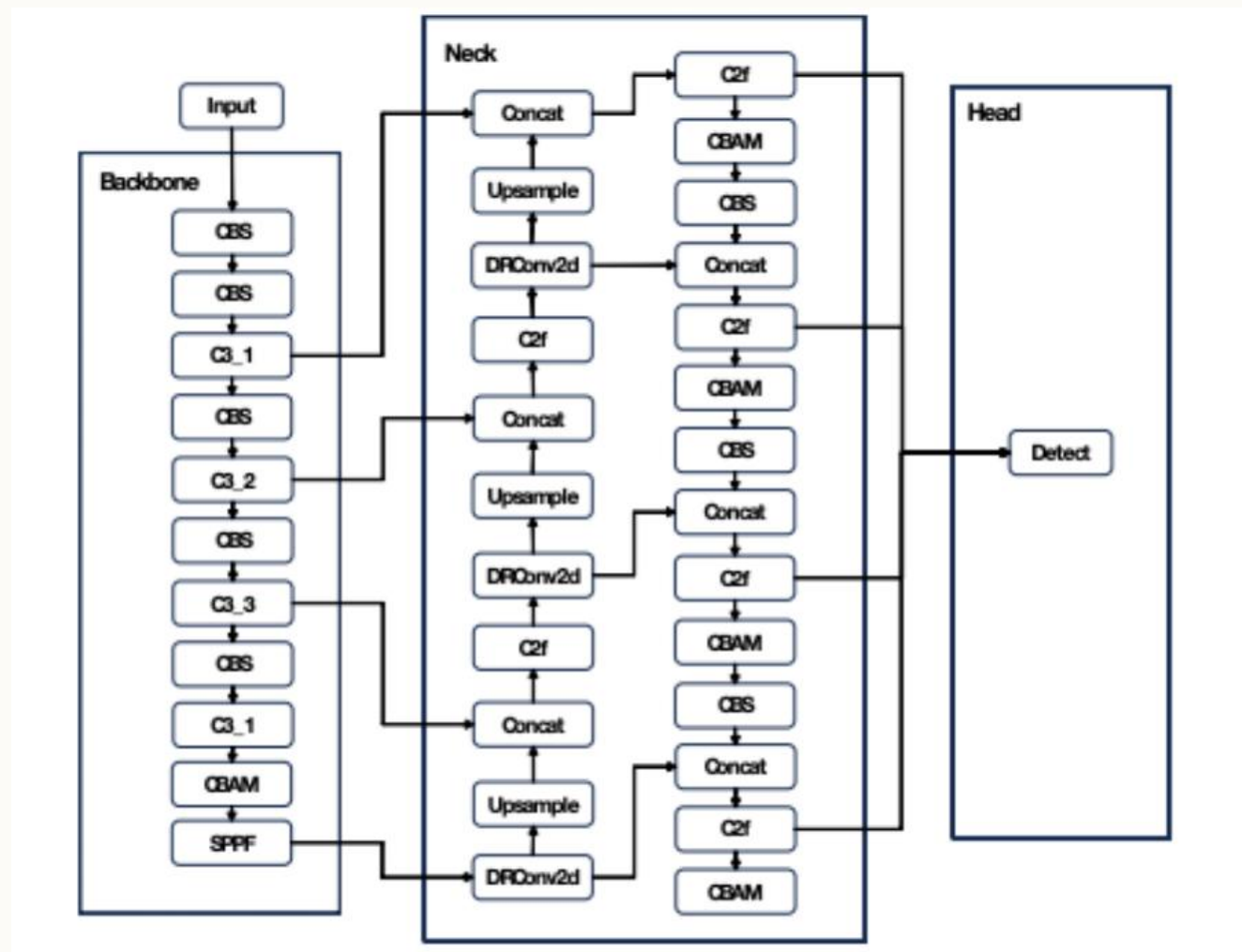
模型改进



模型改进



网络结构中原本有各模块的结构图



改进后口罩佩戴检测模型的网络结构图



模型改进

- 随机选取4张图片进行马赛克数据增强
- 计算符合当前数据集真实框的最佳锚框
- 在默认锚框中增加了一个尺寸比默认锚框中原最小锚框更小的锚框，用于更大特征图上的检测
- 将头部网络的检测头数目增加到4个，能够接受4个不同尺度的特征图，更好的检测较小的目标
- 在原始YOLOv5网络中的第17层与18层之间增加小目标检测层
- 在SPPF层前面和颈部网络中C3层后面添加CBAM，增强模型对小目标的关注，减少误检漏检
- 将MLP输出的特征图进行逐元素相加，再经过Sigmoid激活函数得到通道注意力特征
- 用DRConv替换原网络结构中颈部网络中部分普通卷积模块
- 用C2f模块替换原颈部网络中的C3模块



模型改进

- 使用SIOU损失函数计算样本损失

角度损失定义为: $\Lambda = 1 - 2 * \sin^2 \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) = \cos \left(2 * \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \right).$

距离损失定义为: $\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t} \right) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y}$

形状损失定义为: $\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta$

最终得到SIOU的结果如下: $Loss_{SIOU} = 1 - IoU + \frac{\Delta + \Omega}{2}.$



实验结果

实验结果

- 实验数据集
 - ①WIDER FACE数据集与MAFA数据集中筛选的人脸图片 (WIDER FACE & MAFA) ；
 - ②网络收集的开源人脸数据集 (FACE MASK DETECTION) 。

数据集	训练集		验证集		测试集	
	No-mask	Mask	No-mask	Mask	No-mask	Mask
WIDER FACE & MAFA	5626	2920	3283	1637	3709	2497
FACE MASK DETECTION	4000	1424	1235	439	1478	527

- 超参数设置

数据集	Lr0	Lrf	Weight_decay	Epoch	Batchsize	Iou_t
WIDER FACE & MAFA	0.01	0.1	0.0005	300	16	0.20
FACE MASK DETECTION	0.01	0.01	0.0005	300	8	0.20

- 评价标准 mAP@0.5: 在IoU0.5时, 佩戴与未佩戴口罩两类的平均精度均值。



实验结果-与其他目前前沿的目标检测方法进行对比

数据集	方法	Precision (%)	Recall (%)	mAP@0.5 (%)	Params (M)	GFLOPs	FPS	模型大小 (MB)
WIDER FACE & MAFA	YOLOv5s	93.2	83.0	91.1	7.0	15.8	68	13.7
	YOLOv4	83.4	79.8	82.7	64.4	60.5	41	244
	YOLOv3	95.4	87.4	86.1	61.5	154.6	18	117
	YOLOv3-tiny	93.8	81.9	78.8	8.7	12.9	38	16.6
	SSD	63.9	77.8	74.4	24.1	274	17	91.1
	YOLOv5-EMT	94.6	87.8	93.2	8.5	21.8	55	17.1
FACE MASK DETECTION	YOLOv5s	78.5	76.5	80.1	7.0	15.8	79	13.7
	YOLOv4	83.5	65.3	69.8	64.4	60.5	41	244
	YOLOv3	80.5	84.5	76.7	61.5	155.3	19	117
	YOLOv3-tiny	79.7	73.8	67.6	8.7	13	35	16.5
	SSD	59.2	74.9	63.6	24.1	274	17	91.1
	YOLOv5-EMT	80.9	78.8	82.0	8.5	21.8	59	17.1



实验结果-消融实验

小目标检测层	DRConv	C2f	CBAM	SIoU	WBF	No-mask (%)	Mask (%)	mAP@0.5 (%)
×	×	×	×	×	×	86.2	96.0	91.1
√	×	×	×	×	×	87.8	97.5	92.6
×	√	×	×	×	×	88.1	97.0	92.5
×	×	√	×	×	×	87.9	97.2	92.5
×	×	×	√	×	×	88.1	96.7	92.4
×	×	×	×	√	×	86.4	96.6	91.5
×	×	×	×	×	√	86.1	96.0	91.0
√	√	×	×	√	×	88.3	97.2	92.8
√	×	√	×	√	×	88.6	96.2	92.4
√	×	×	√	√	×	88.1	95.6	91.8
√	×	×	×	√	√	88.9	97.1	93.0
√	√	√	×	√	×	87.2	96.8	92.0
√	√	×	√	√	×	87.8	96.4	92.1
√	×	√	√	√	×	88.7	96.4	92.5
√	√	√	√	√	×	88.9	97.3	93.1
√	√	√	√	√	√	89.2	97.2	93.2

感谢观看
请批评指正!