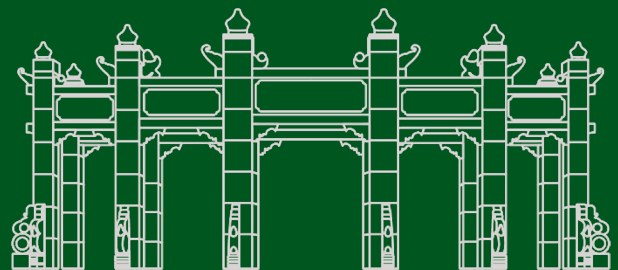
SYSU 2023
PRESENTATION SLIDE

中山大學

SUN YAT-SEN UNIVERSITY

# HiXray & LIM

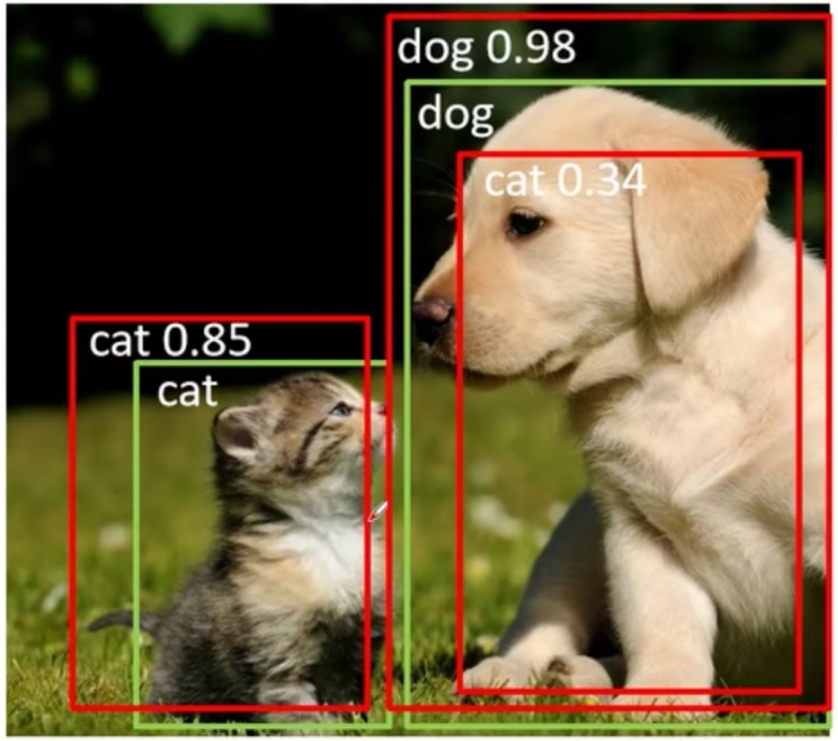


# 5.1. Experiment Setting Details

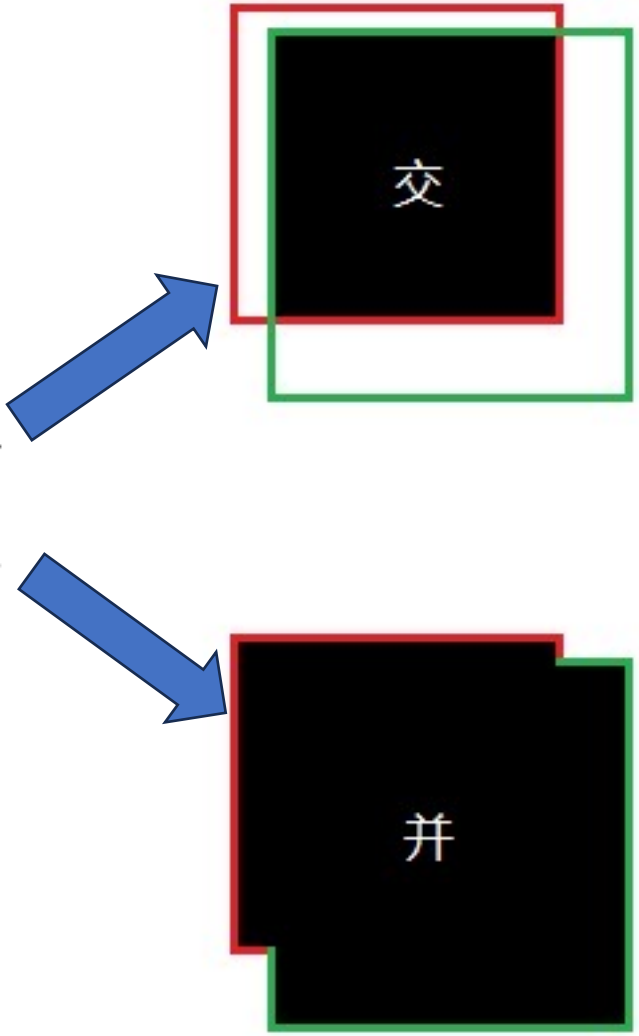
**Backbone Networks:** The backbone networks of SSD, FCOS and YOLOv5 are VGG16 [34], ResNet50 [10] and CSPNet [37] respectively. For each backbone network, we modify the corresponding network architecture to implement LIM mechanism.

**Parameters:** All experiments of LIM and baselines are optimized by the SGD optimizer and the initial learning rate is set to 0.0001. The momentum and weight decay are set to 0.9 and 0.0005 respectively. The batch size is set to 32 with shuffle strategy while training. We evaluate the mean Average Precision (mAP) of the object detection to measure the performance of all models fairly. Besides, the IOU threshold measuring the accuracy of the predicted bounding box against the ground-truth is set to 0.5.

# IOU



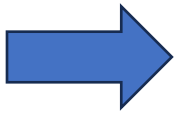

$$IOU = \frac{S_{交}}{S_{并}}$$



# mAP

| | 正样本（positive） | 负样本（negative） |
|---|---|---|
| one-stage 框架 | 匹配到目标的候选框（十几个或几十） | 没匹配到的候选框（10^4~10^5 个） |
| two-stage 框架 | 匹配到目标的候选框（十几个或几十） | 没匹配到的候选框（1~2k个） |

TP（True Positives）意思就是被分为了正样本，而且分对了。
TN（True Negatives）意思就是被分为了负样本，而且分对了，
FP（False Positives）意思就是被分为了正样本，但是分错了。
FN（False Negatives）意思就是被分为了负样本，但是分错了。



TP（True Positive）比如IOU>0.5的检测框数量
FP（False Positive）比如IOU<=0.5的检测框数量
FN（False Negative）没有检测到的GT的数量

# mAP

TP（True Positive）比如IOU>0.5的检测框数量
FP（False Positive）比如IOU<=0.5的检测框数量
FN（False Negative）没有检测到的GT的数量

precision（精确度/查准率）：分类器认为是正类并且确实是正类的部分占所有分类器认为是正类的比例，即模型预测的所有目标中预测正确的比例

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

recall（召回率/查全率）：分类器认为是正类并且确实是正类的部分占所有确实是正类的比例，即所有真实目标中模型预测正确的目标比例

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

# mAP

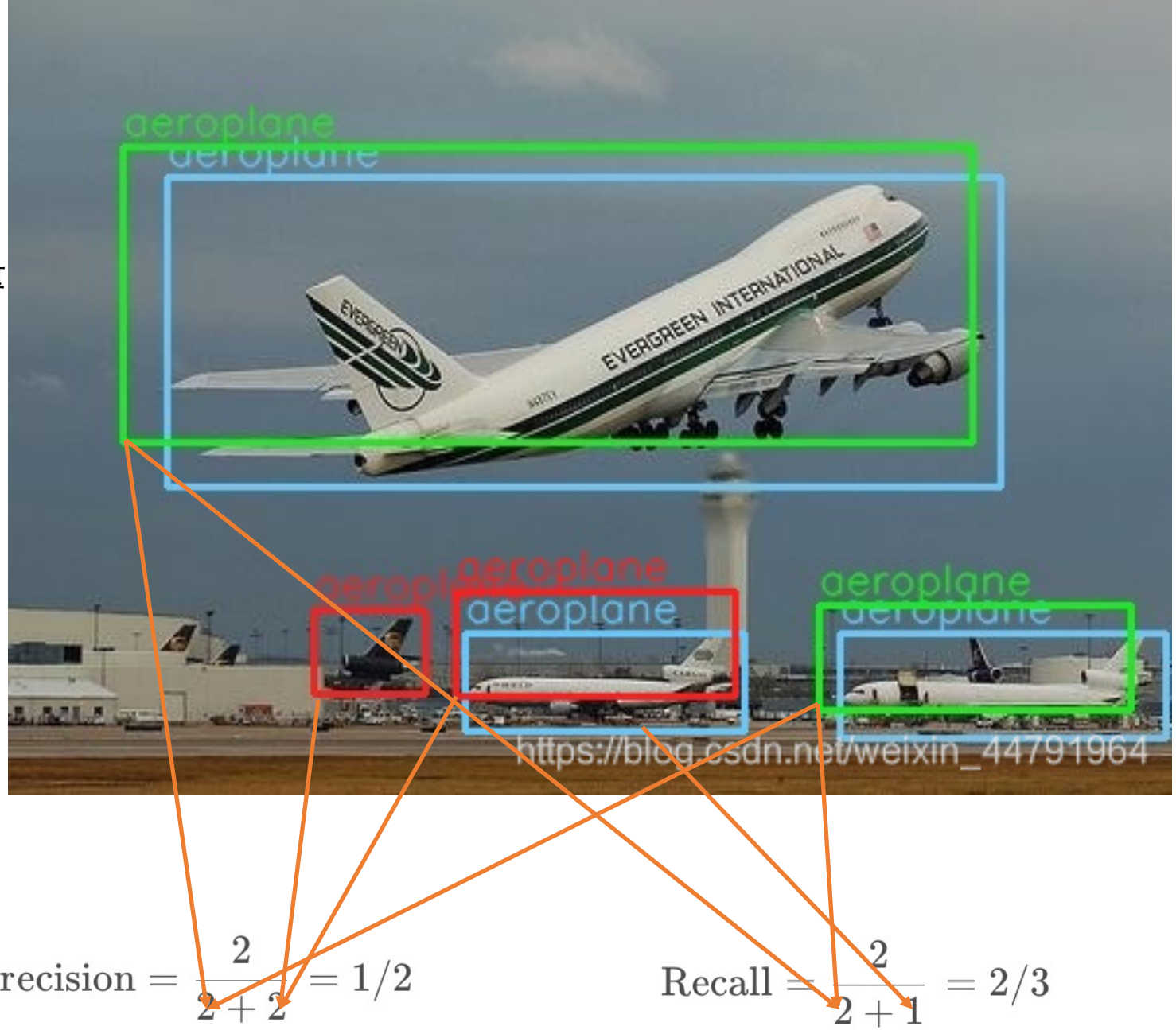
TP（True Positive）比如IOU>0.5的检测框数量
FP（False Positive）比如IOU<=0.5的检测框数量
FN（False Negative）没有检测到的GT的数量

precision（精确度/查准率）：模型预测的所有目标中预测正确的比例

$$\mathrm{Precision} = \frac{TP}{TP + FP}$$

recall（召回率/查全率）：所有真实目标中模型预测正确的目标比例

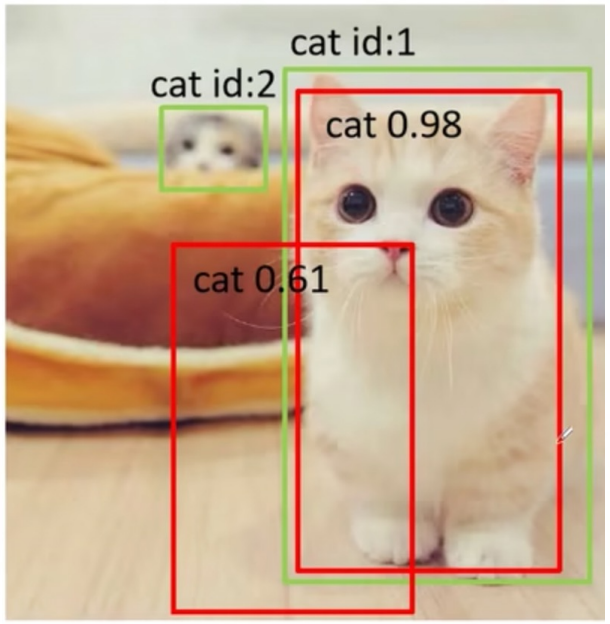
$$\mathrm{Recall} = \frac{TP}{TP + FN}$$

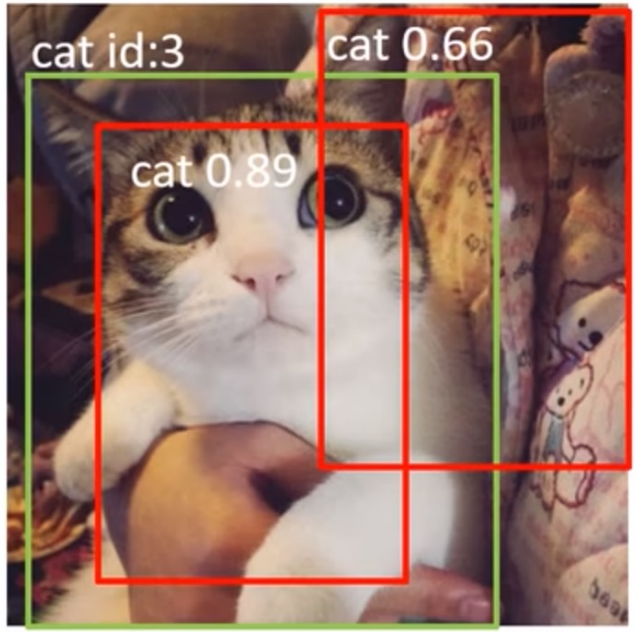


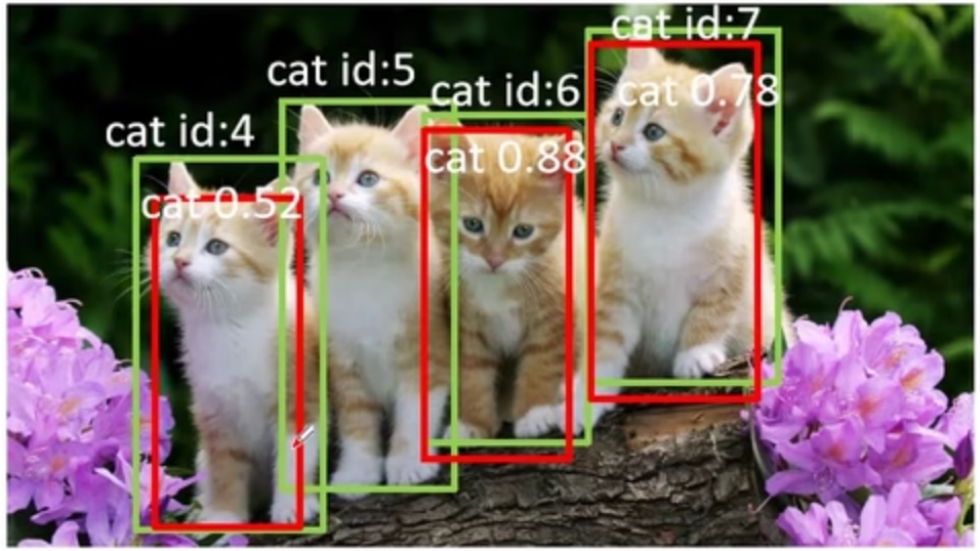

$$\mathrm{Precision} = \frac{2}{2+2} = 1/2$$

$$\mathrm{Recall} = \frac{2}{2+1} = 2/3$$

# mAP




| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

# mAP

| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

TP=1； FP=0； FN=6

| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

TP=2； FP=0； FN=5

| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

TP=3； FP=0； FN=4

TP（True Positive）比如IOU>0.5的检测框数量
FP（False Positive）比如IOU<=0.5的检测框数量
FN（False Negative）没有检测到的GT的数量

TP（True Positive）比如IOU>0.5的confidence>0.9的检测框数量
FP（False Positive）比如IOU<=0.5的confidence>0.9的检测框数量
FN（False Negative）confidence<0.9的检测框数量

| Rank | Precision | Recall |
|---|---|---|
| 1 | 1.0 | 0.14 |
| 2 | 1.0 | 0.28 |
| 3 | 1.0 | 0.42 |

# mAP

| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

TP=3； FP=0； FN=4

| Rank | Precision | Recall |
|---|---|---|
| 1 | 1.0 | 0.14 |
| 2 | 1.0 | 0.28 |
| 3 | 1.0 | 0.42 |

| GT ID | Confidence | OB(IOU=0.5) |
|---|---|---|
| 1 | 0.98 | True |
| 3 | 0.89 | True |
| 6 | 0.88 | True |
| 7 | 0.78 | True |
| 3 | 0.66 | False |
| 1 | 0.61 | False |
| 4 | 0.52 | True |

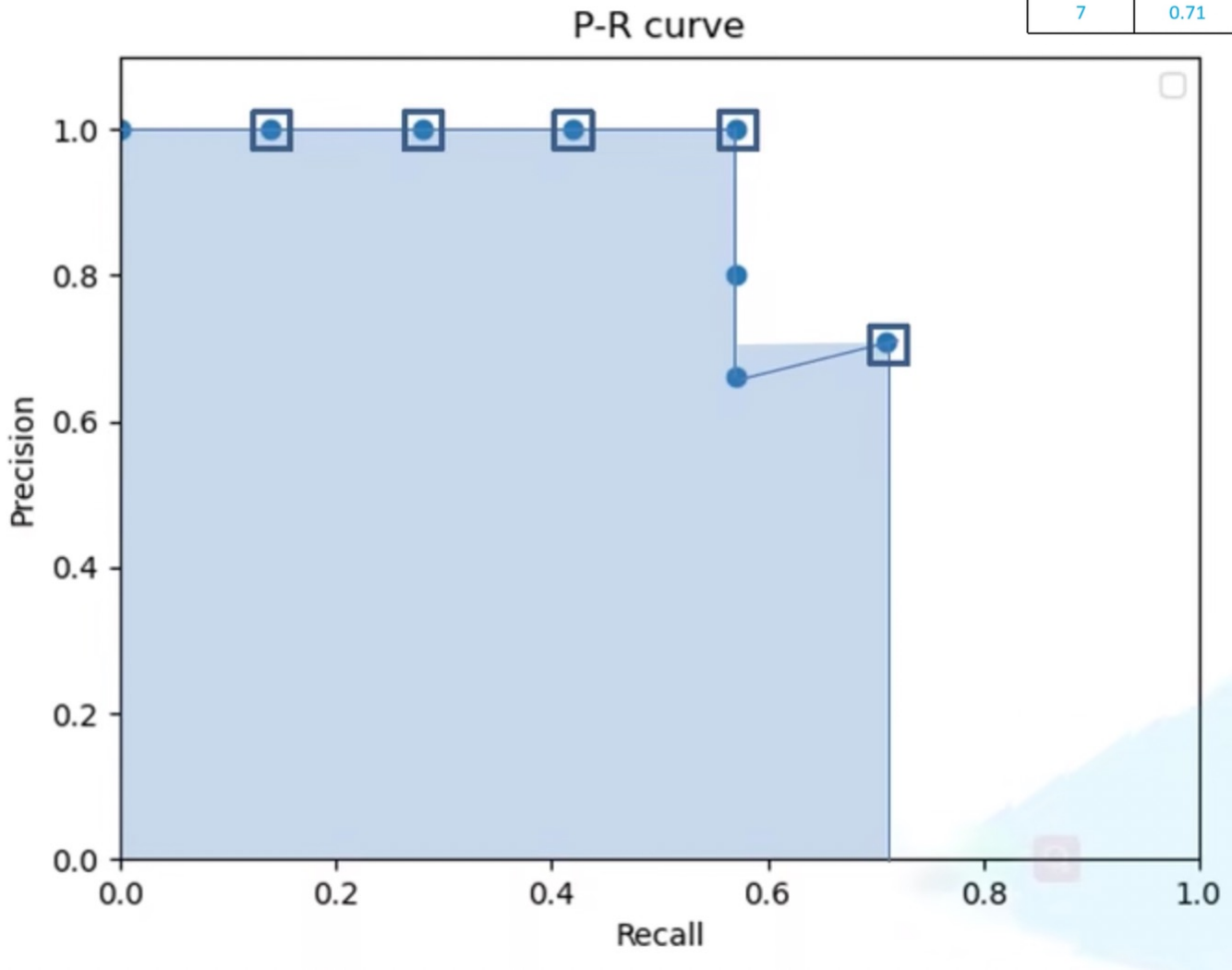
TP=5； FP=2； FN=2

| Rank | Precision | Recall |
|---|---|---|
| 1 | 1.0 | 0.14 |
| 2 | 1.0 | 0.28 |
| 3 | 1.0 | 0.42 |
| 4 | 1.0 | 0.57 |
| 5 | 0.80 | 0.57 |
| 6 | 0.66 | 0.57 |
| 7 | 0.71 | 0.71 |

# mAP

| Rank | Precision | Recall |
|---|---|---|
| 1 | 1.0 | 0.14 |
| 2 | 1.0 | 0.28 |
| 3 | 1.0 | 0.42 |
| 4 | 1.0 | 0.57 |
| 5 | 0.80 | 0.57 |
| 6 | 0.66 | 0.57 |
| 7 | 0.71 | 0.71 |


| Rank | Precision | Recall |
|---|---|---|
| 1 | 1.0 | 0.14 |
| 2 | 1.0 | 0.28 |
| 3 | 1.0 | 0.42 |
| 4 | 1.0 | 0.57 |
| ~~5~~ | ~~0.80~~ | ~~0.57~~ |
| ~~6~~ | ~~0.66~~ | ~~0.57~~ |
| 7 | 0.71 | 0.71 |

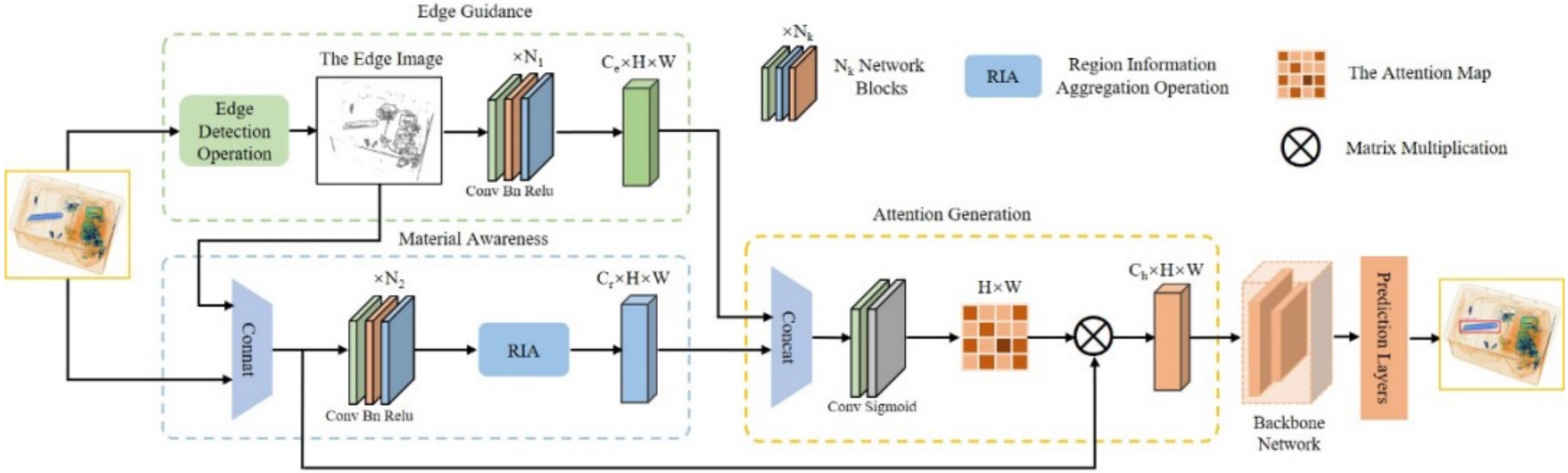



# 5.2. Comparing with SOTA Detection Methods

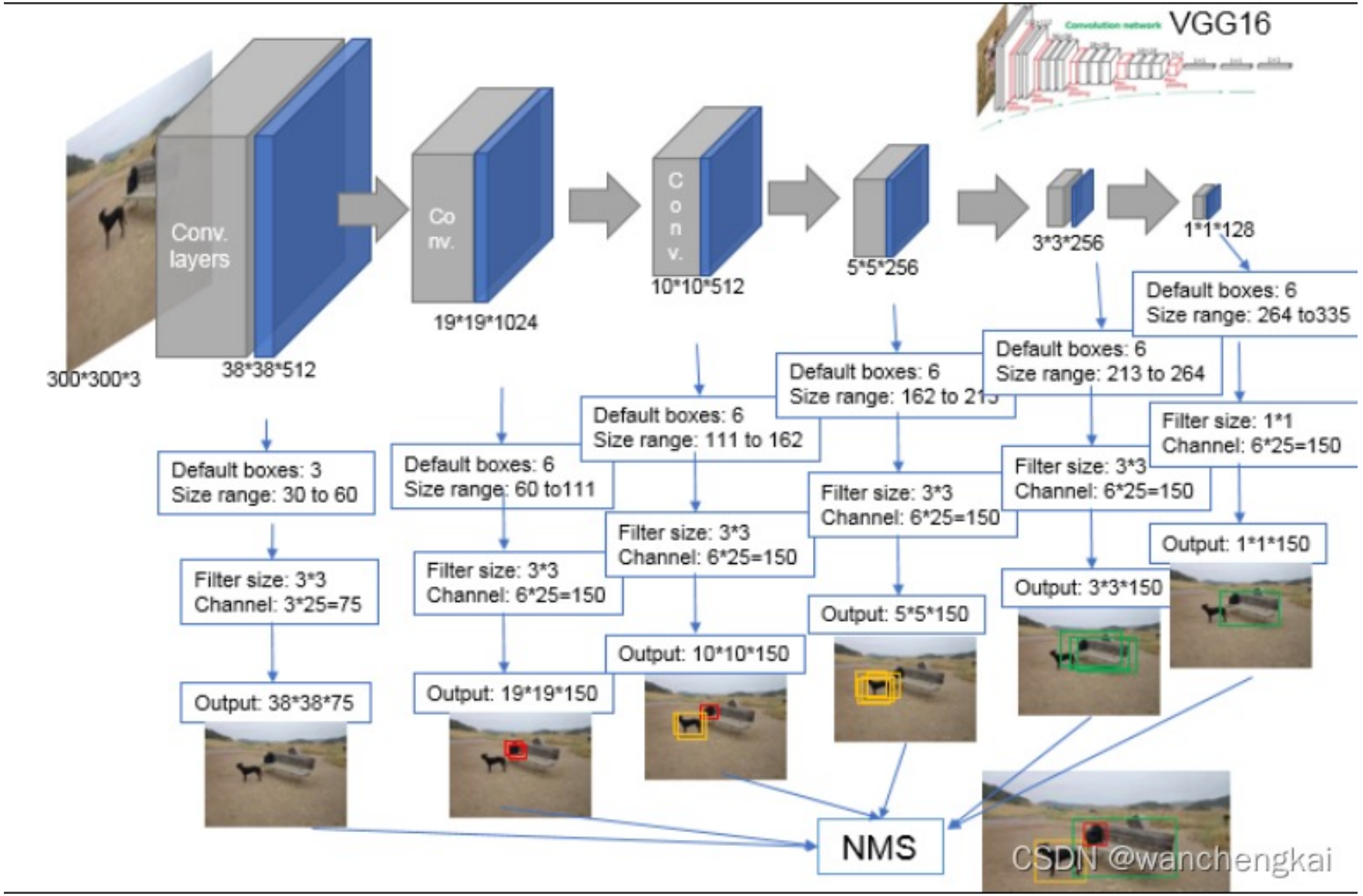
| Method | HiXray Dataset (**Ours**) | | | | | | | | | OPIXray Dataset [40] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL | **AVG** | FO | ST | SC | UT | MU |
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 | 70.9 | 76.9 | 35.0 | 93.4 | 65.9 | 83.3 |
| SSD+DOAM [40] | 72.1 | 88.6 | 82.9 | 83.6 | 97.5 | 94.1 | 92.1 | 38.2 | 0.01 | 74.0 | 81.4 | 41.5 | 95.1 | 68.2 | **83.8** |
| **SSD+LIM** | **73.1** | **89.1** | **84.3** | **84.0** | **97.7** | **94.5** | **92.4** | **42.3** | **0.1** | **74.6** | **81.4** | **42.4** | **95.9** | **71.2** | 82.1 |
| FCOS [35] | 75.7 | 88.6 | 86.4 | 86.8 | 89.9 | 88.9 | 88.9 | 63.0 | 13.3 | 82.0 | 86.4 | 68.5 | 90.2 | 78.4 | 86.6 |
| FCOS+DOAM [40] | 76.2 | 88.6 | 87.5 | 87.8 | 89.9 | 89.7 | 88.8 | 63.5 | 12.7 | 82.4 | 86.5 | 68.6 | 90.2 | 78.8 | **87.7** |
| **FCOS+LIM** | **77.3** | **88.9** | **88.2** | **88.3** | **90.0** | **89.8** | **89.2** | **69.8** | **14.4** | **83.1** | **86.6** | **71.9** | **90.3** | **79.9** | 86.8 |
| YOLOv5 [14] | 81.7 | 95.5 | 94.5 | 92.8 | 97.9 | 98.0 | 94.9 | 63.7 | 16.3 | 87.8 | 93.4 | 67.9 | 98.1 | 85.4 | 94.1 |
| YOLOv5+DOAM [40] | 82.2 | 95.9 | 94.7 | 93.7 | 98.1 | 98.1 | 95.8 | 65.0 | 16.1 | 88.0 | 93.3 | 69.3 | 97.9 | 84.4 | **95.0** |
| **YOLOv5+LIM** | **83.2** | **96.1** | **95.1** | **93.9** | **98.2** | **98.3** | **96.4** | **65.8** | **21.3** | **90.6** | **94.8** | **77.6** | **98.2** | **88.9** | 93.8 |

# 5.2. Comparing with SOTA Detection Methods

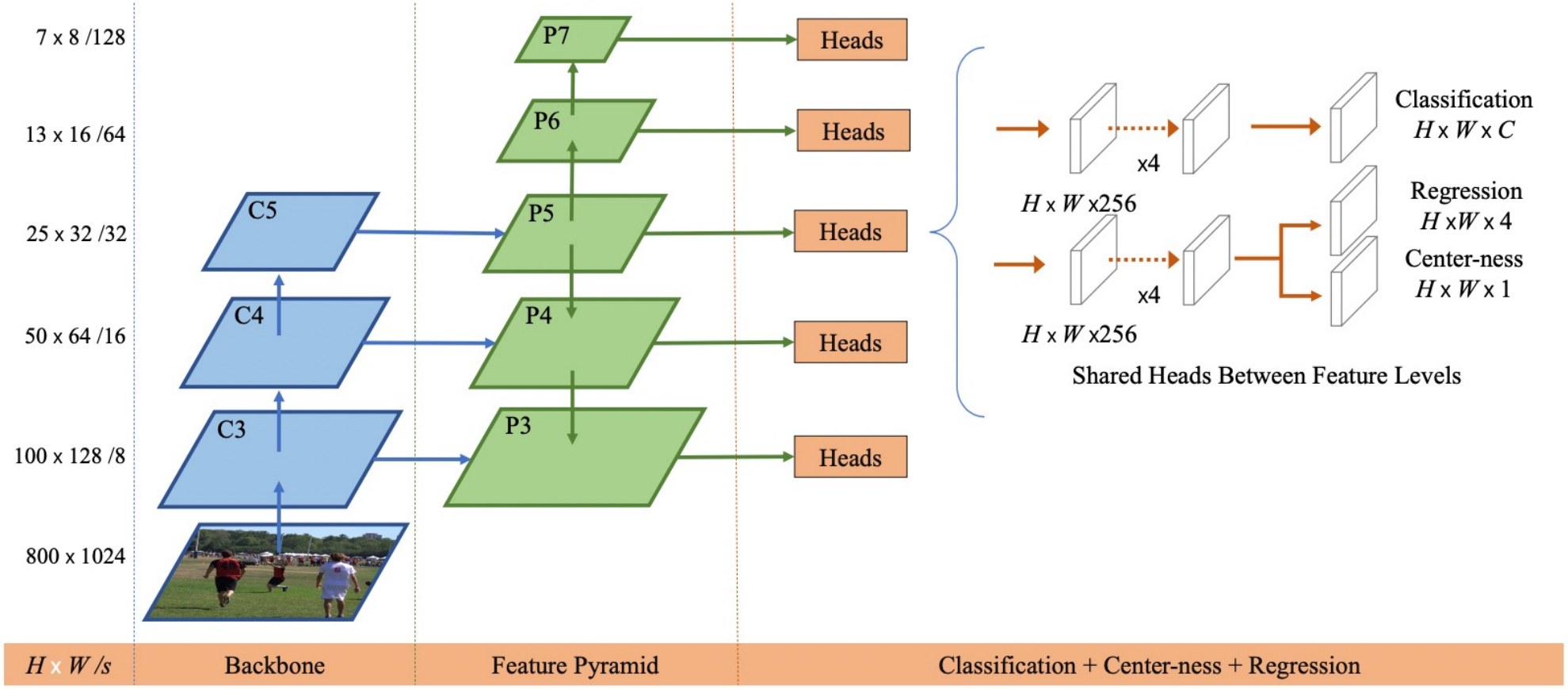
## DOAM



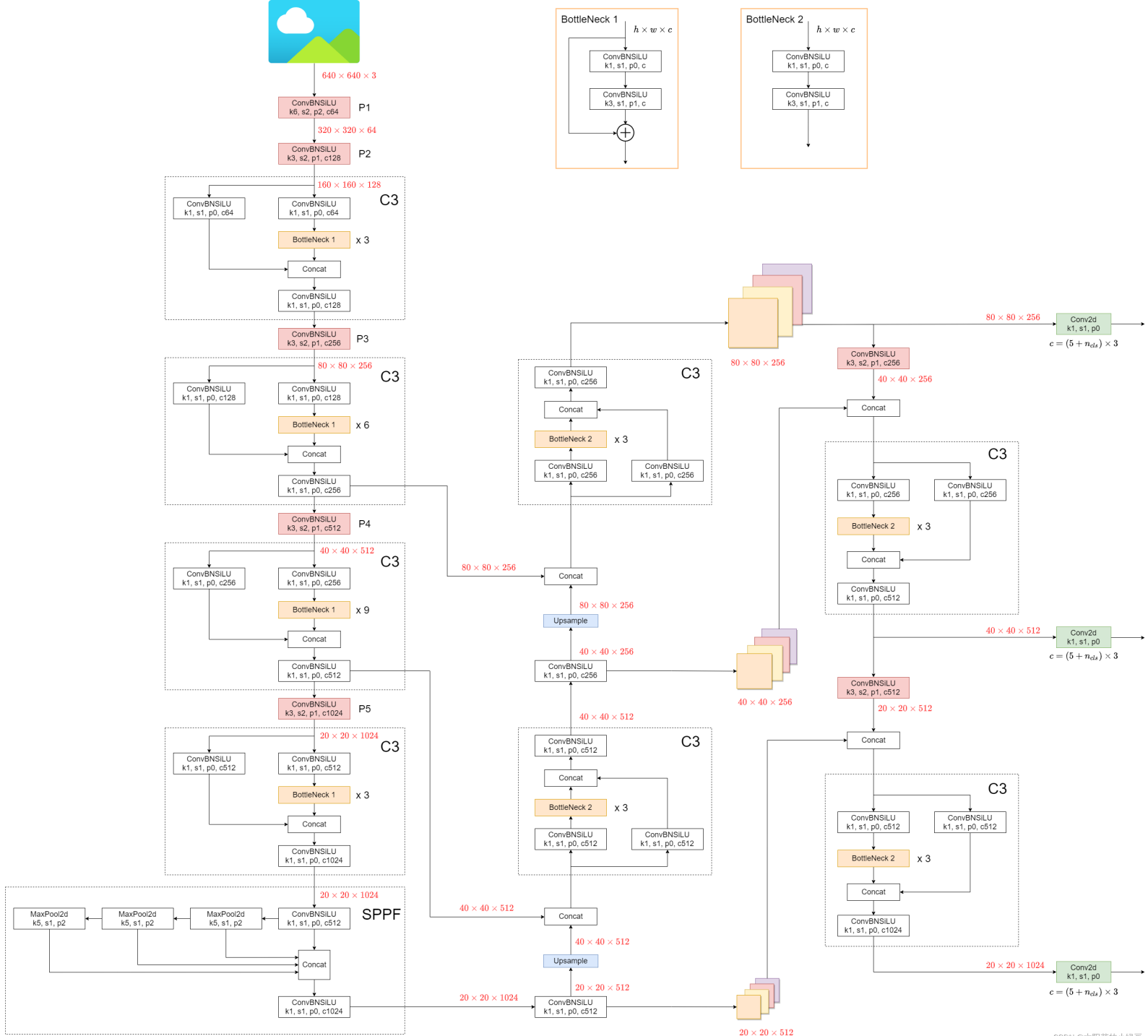
# SSD


VGG16
Conv. layers
300*300*3
38*38*512
Co nv.
19*19*1024
C o n v.
10*10*512
5*5*256
3*3*256
1*1*128
Default boxes: 6
Size range: 264 to335
Default boxes: 6
Size range: 213 to 264
Default boxes: 6
Size range: 162 to 213
Default boxes: 6
Size range: 111 to 162
Default boxes: 6
Size range: 60 to111
Default boxes: 3
Size range: 30 to 60
Filter size: 1*1
Channel: 6*25=150
Filter size: 3*3
Channel: 6*25=150
Filter size: 3*3
Channel: 6*25=150
Filter size: 3*3
Channel: 6*25=150
Filter size: 3*3
Channel: 3*25=75
Output: 1*1*150
Output: 3*3*150
Output: 5*5*150
Output: 10*10*150
Output: 19*19*150
Output: 38*38*75
NMS
CSDN @wanchengkai

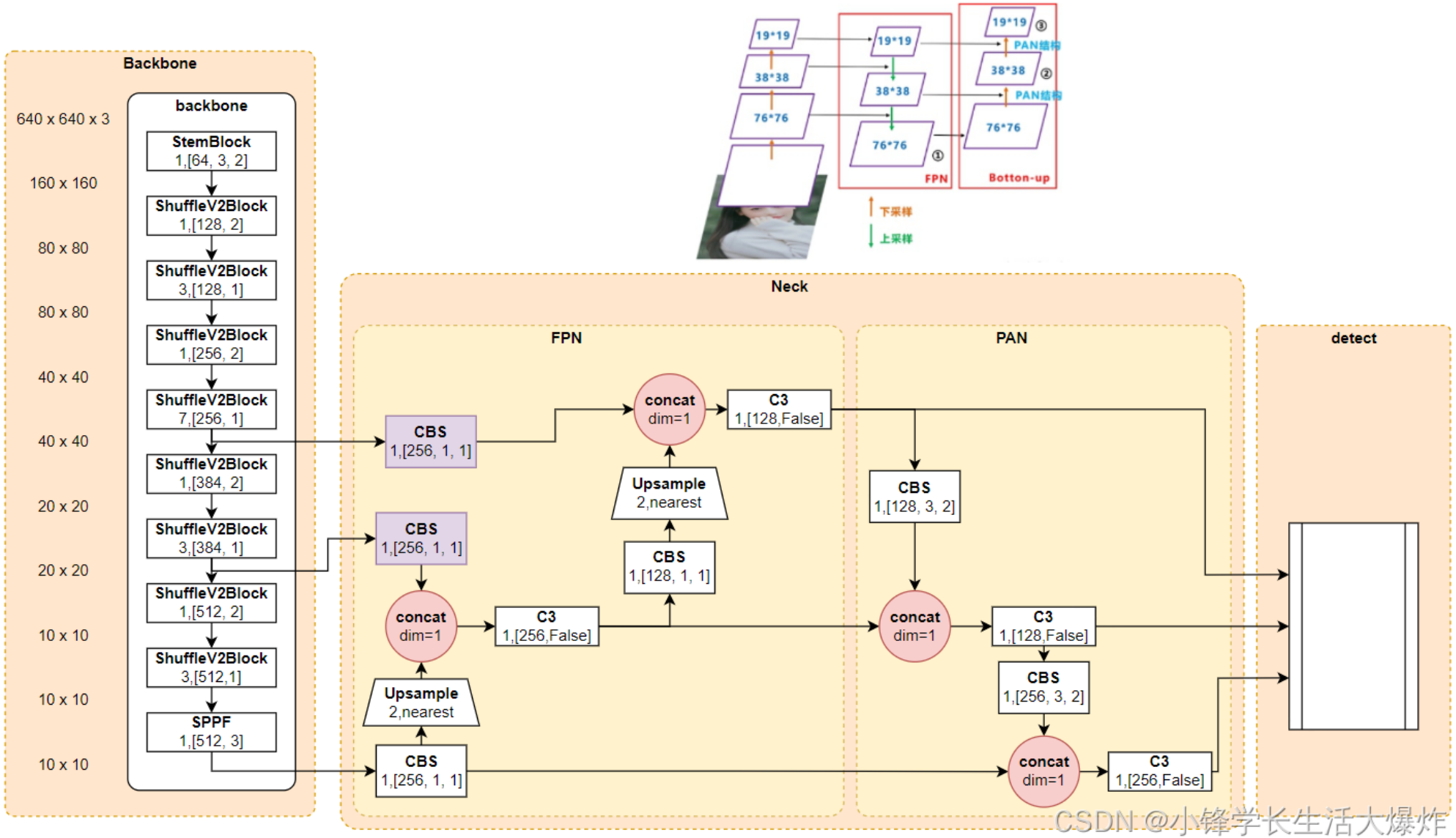
# FCOS


7 x 8 /128
13 x 16 /64
25 x 32 /32
50 x 64 /16
100 x 128 /8
800 x 1024
C5
C4
C3
P7
P6
P5
P4
P3
Heads
Heads
Heads
Heads
Heads
x4
H x W x256
x4
H x W x256
Classification
H x W x C
Regression
H xW x 4
Center-ness
H x W x 1
Shared Heads Between Feature Levels
H x W /s
Backbone
Feature Pyramid
Classification + Center-ness + Regression

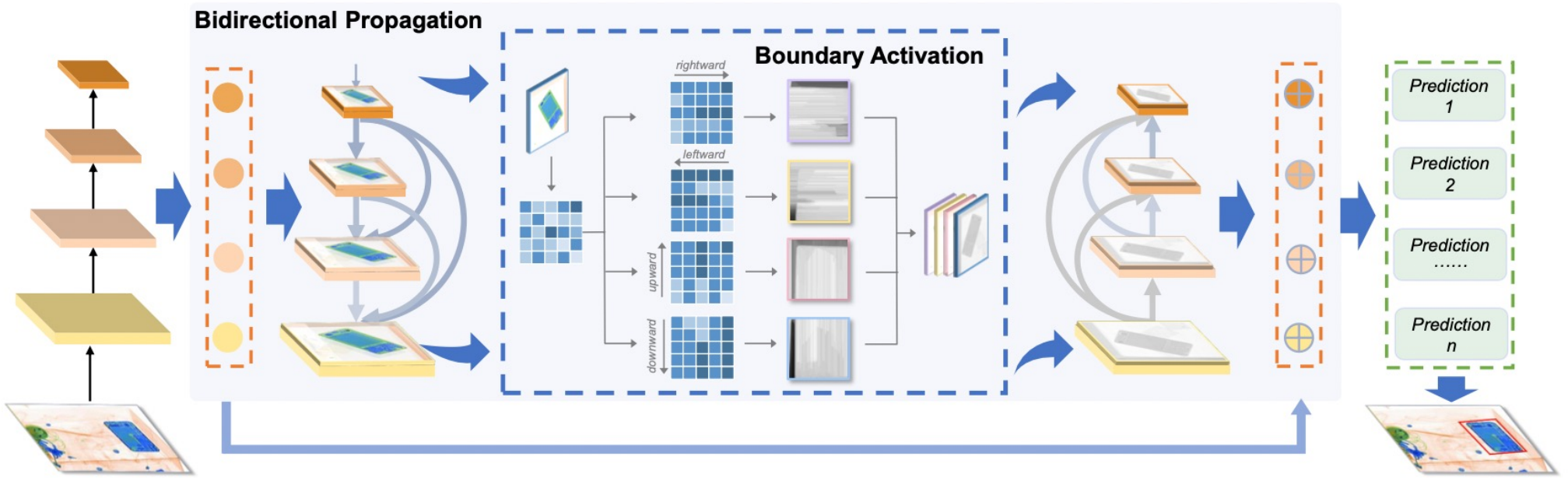
# YOLOv5

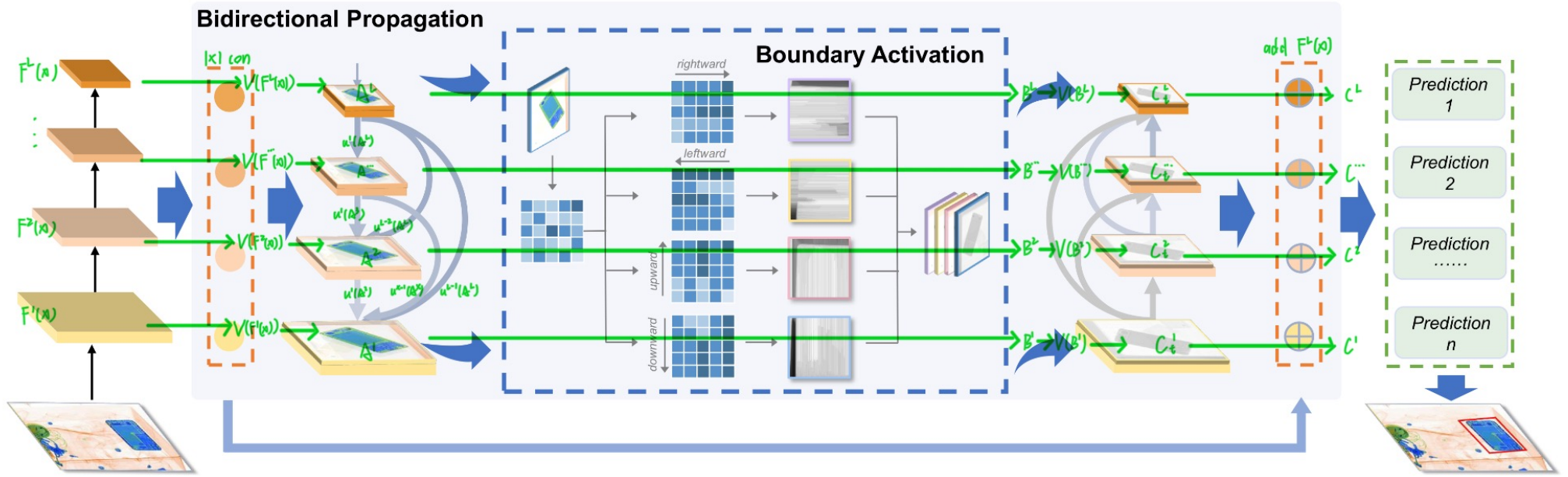


# YOLOv5



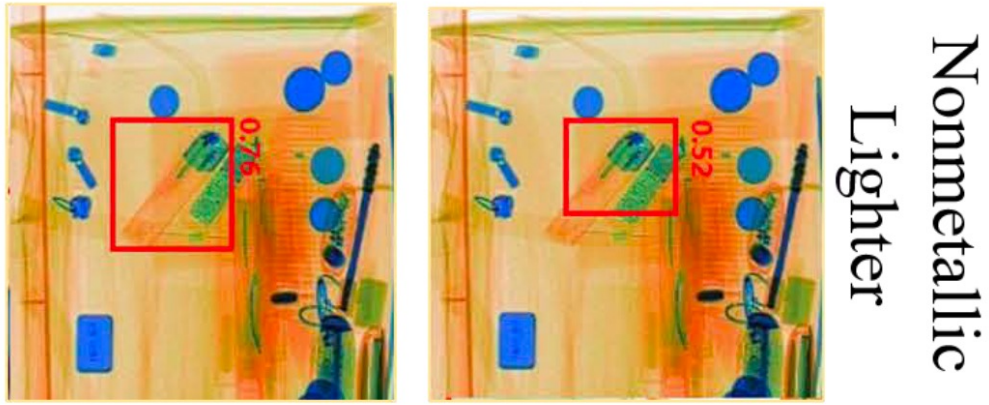
# LIM


Bidirectional Propagation
Boundary Activation
rightward
leftward
upward
downward
Prediction 1
Prediction 2
Prediction .....
Prediction n


Bidirectional Propagation
Boundary Activation
rightward
leftward
upward
downward
Prediction 1
Prediction 2
Prediction .....
Prediction n

# 5.2. Comparing with SOTA Detection Methods

| Method | HiXray Dataset (**Ours**) | | | | | | | | | OPIXray Dataset [40] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL | **AVG** | FO | ST | SC | UT | MU |
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 | 70.9 | 76.9 | 35.0 | 93.4 | 65.9 | 83.3 |
| SSD+DOAM [40] | 72.1 | 88.6 | 82.9 | 83.6 | 97.5 | 94.1 | 92.1 | 38.2 | 0.01 | 74.0 | 81.4 | 41.5 | 95.1 | 68.2 | **83.8** |
| **SSD+LIM** | **73.1** | **89.1** | **84.3** | **84.0** | **97.7** | **94.5** | **92.4** | **42.3** | **0.1** | **74.6** | **81.4** | **42.4** | **95.9** | **71.2** | 82.1 |
| FCOS [35] | 75.7 | 88.6 | 86.4 | 86.8 | 89.9 | 88.9 | 88.9 | 63.0 | 13.3 | 82.0 | 86.4 | 68.5 | 90.2 | 78.4 | 86.6 |
| FCOS+DOAM [40] | 76.2 | 88.6 | 87.5 | 87.8 | 89.9 | 89.7 | 88.8 | 63.5 | 12.7 | 82.4 | 86.5 | 68.6 | 90.2 | 78.8 | **87.7** |
| **FCOS+LIM** | **77.3** | **88.9** | **88.2** | **88.3** | **90.0** | **89.8** | **89.2** | **69.8** | **14.4** | **83.1** | **86.6** | **71.9** | **90.3** | **79.9** | 86.8 |
| YOLOv5 [14] | 81.7 | 95.5 | 94.5 | 92.8 | 97.9 | 98.0 | 94.9 | 63.7 | 16.3 | 87.8 | 93.4 | 67.9 | 98.1 | 85.4 | 94.1 |
| YOLOv5+DOAM [40] | 82.2 | 95.9 | 94.7 | 93.7 | 98.1 | 98.1 | 95.8 | 65.0 | 16.1 | 88.0 | 93.3 | 69.3 | 97.9 | 84.4 | **95.0** |
| **YOLOv5+LIM** | **83.2** | **96.1** | **95.1** | **93.9** | **98.2** | **98.3** | **96.4** | **65.8** | **21.3** | **90.6** | **94.8** | **77.6** | **98.2** | **88.9** | 93.8 |



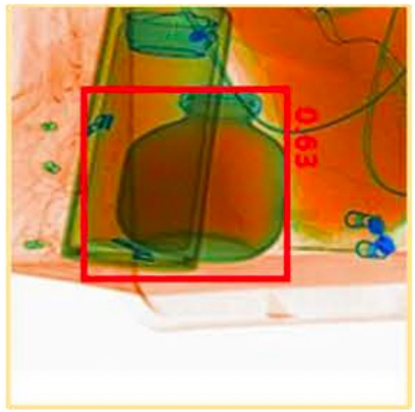

For NL, it is very small in size and composed of a small piece of iron and a plastic body. The plastic shown under X-ray appears orange, which almost blends in with the background.

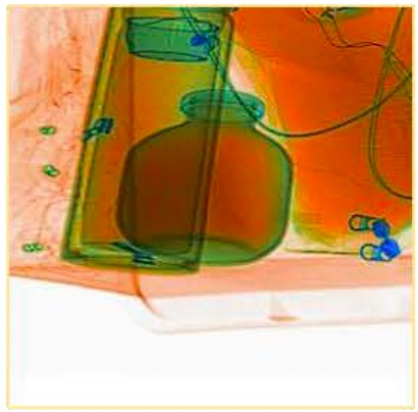
# 5.2. Comparing with SOTA Detection Methods

| Method | HiXray Dataset (**Ours**) | | | | | | | | | OPIXray Dataset [40] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL | **AVG** | FO | ST | SC | UT | MU |
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 | 70.9 | 76.9 | 35.0 | 93.4 | 65.9 | 83.3 |
| SSD+DOAM [40] | 72.1 | 88.6 | 82.9 | 83.6 | 97.5 | 94.1 | 92.1 | 38.2 | 0.01 | 74.0 | 81.4 | 41.5 | 95.1 | 68.2 | **83.8** |
| **SSD+LIM** | **73.1** | **89.1** | **84.3** | **84.0** | **97.7** | **94.5** | **92.4** | **42.3** | **0.1** | **74.6** | **81.4** | **42.4** | **95.9** | **71.2** | 82.1 |
| FCOS [35] | 75.7 | 88.6 | 86.4 | 86.8 | 89.9 | 88.9 | 88.9 | 63.0 | 13.3 | 82.0 | 86.4 | 68.5 | 90.2 | 78.4 | 86.6 |
| FCOS+DOAM [40] | 76.2 | 88.6 | 87.5 | 87.8 | 89.9 | 89.7 | 88.8 | 63.5 | 12.7 | 82.4 | 86.5 | 68.6 | 90.2 | 78.8 | **87.7** |
| **FCOS+LIM** | **77.3** | **88.9** | **88.2** | **88.3** | **90.0** | **89.8** | **89.2** | **69.8** | **14.4** | **83.1** | **86.6** | **71.9** | **90.3** | **79.9** | 86.8 |
| YOLOv5 [14] | 81.7 | 95.5 | 94.5 | 92.8 | 97.9 | 98.0 | 94.9 | 63.7 | 16.3 | 87.8 | 93.4 | 67.9 | 98.1 | 85.4 | 94.1 |
| YOLOv5+DOAM [40] | 82.2 | 95.9 | 94.7 | 93.7 | 98.1 | 98.1 | 95.8 | 65.0 | 16.1 | 88.0 | 93.3 | 69.3 | 97.9 | 84.4 | **95.0** |
| **YOLOv5+LIM** | **83.2** | **96.1** | **95.1** | **93.9** | **98.2** | **98.3** | **96.4** | **65.8** | **21.3** | **90.6** | **94.8** | **77.6** | **98.2** | **88.9** | 93.8 |

非金属打火机

化妆品





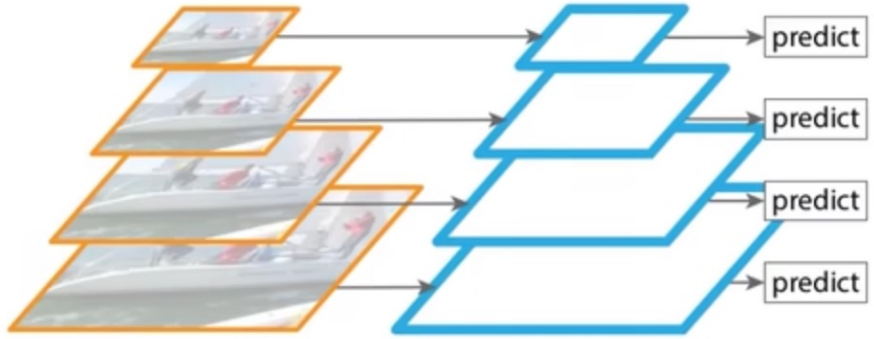
Cosmetics

For CO, the main reason is that there are big differences in the shapes of cosmetics, such as round and square, which are easy to be confused with other kinds of items.

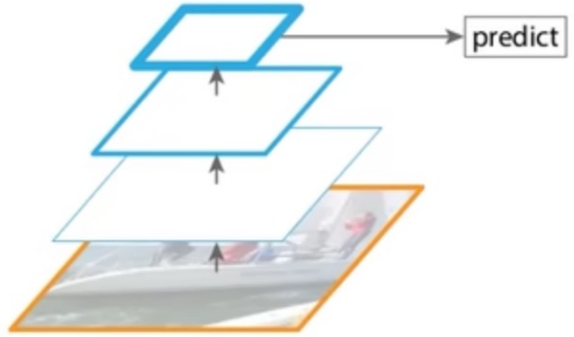
# 5.3. Comparing with Feature Pyramid Mechanisms

| Method | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL |
|---|---|---|---|---|---|---|---|---|---|
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 |
| +FPN [17] | 72.0 | 87.4 | 81.5 | 83.2 | 97.9 | 93.9 | 92.2 | 40.3 | 0.02 |
| +PANet [39] | 72.0 | 88.3 | 83.2 | 82.8 | **97.9** | 93.8 | **92.6** | 37.3 | 0.01 |
| **+LIM** | **73.1** | **89.1** | **84.3** | **84.0** | 97.7 | **94.5** | 92.4 | **42.3** | **0.1** |

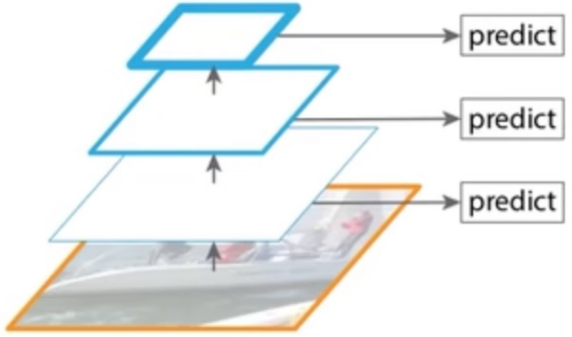
# FPN



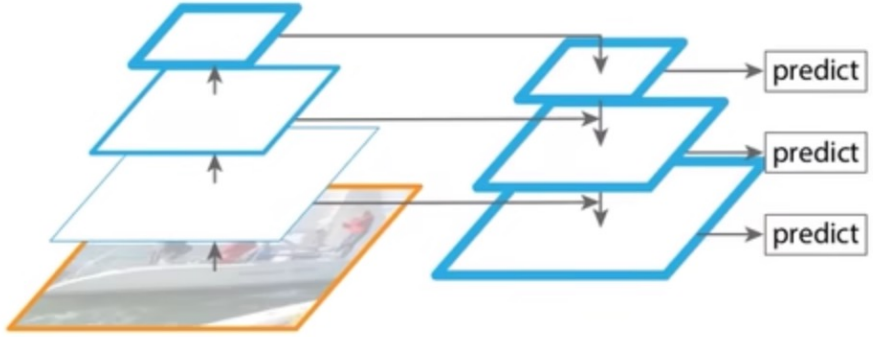

(a) Featurized image pyramid




(b) Single feature map

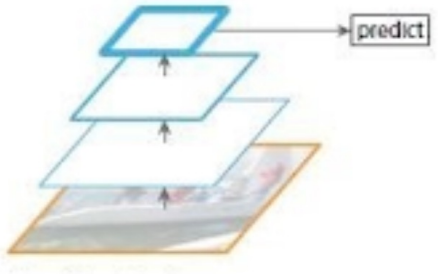



(c) Pyramidal feature hierarchy



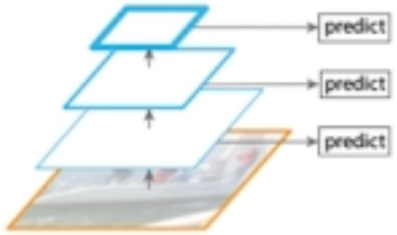

(d) Feature Pyramid Network

# FPN

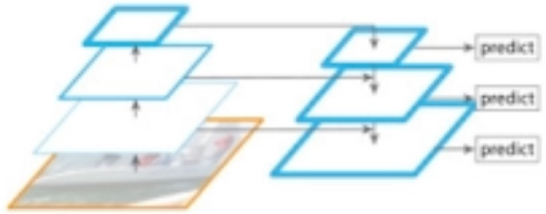



1.基础CNN网络结构


2.图片金字塔




3.多尺度特征融合




4.FPN（Feature Pyramid Networks）

基础cnn结构：
自底向上卷积，然后使用最后一层特征图进行预测，即仅采用网络最后一层的特征
**应用案例：分类器**

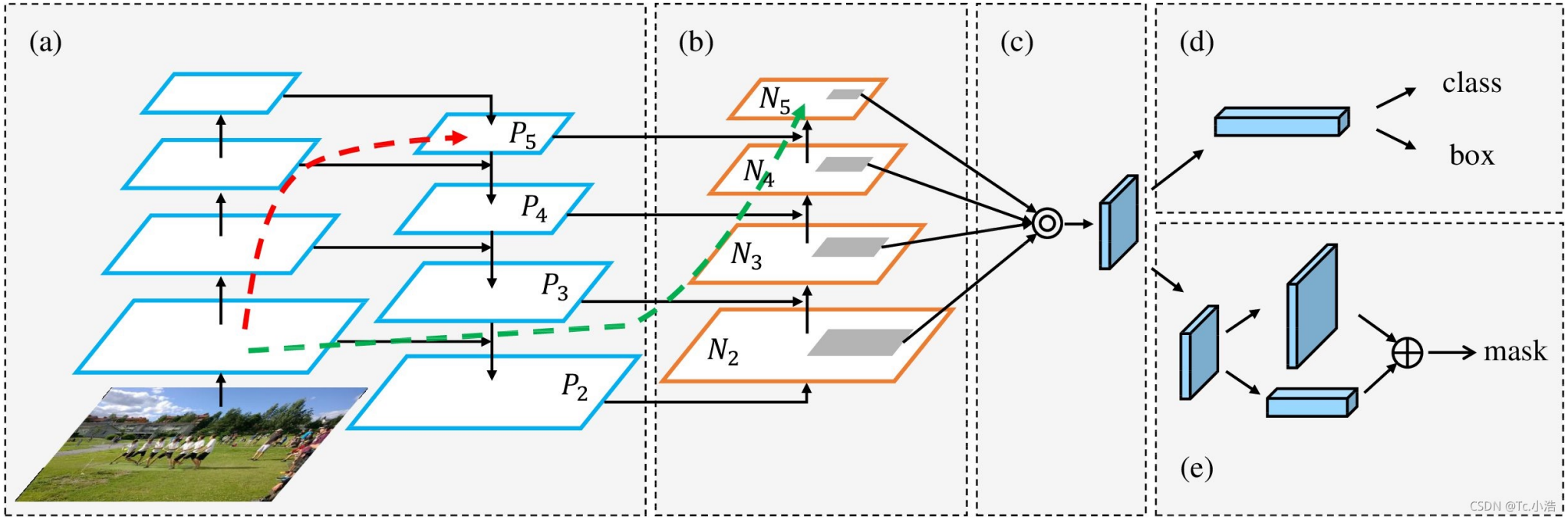
图片金字塔：
将图像做成不同的scale，然后不同scale的图像生成对应的不同scale的特征，将图片缩放成多个比例，每个比例单独提取特征图进行预测
**应用案例：Mtcnn的pnet； 测试增广（TTA）**

多尺度特征融合：
从网络不同层抽取不同尺度的特征做预测，这种方式不会增加额外的计算量

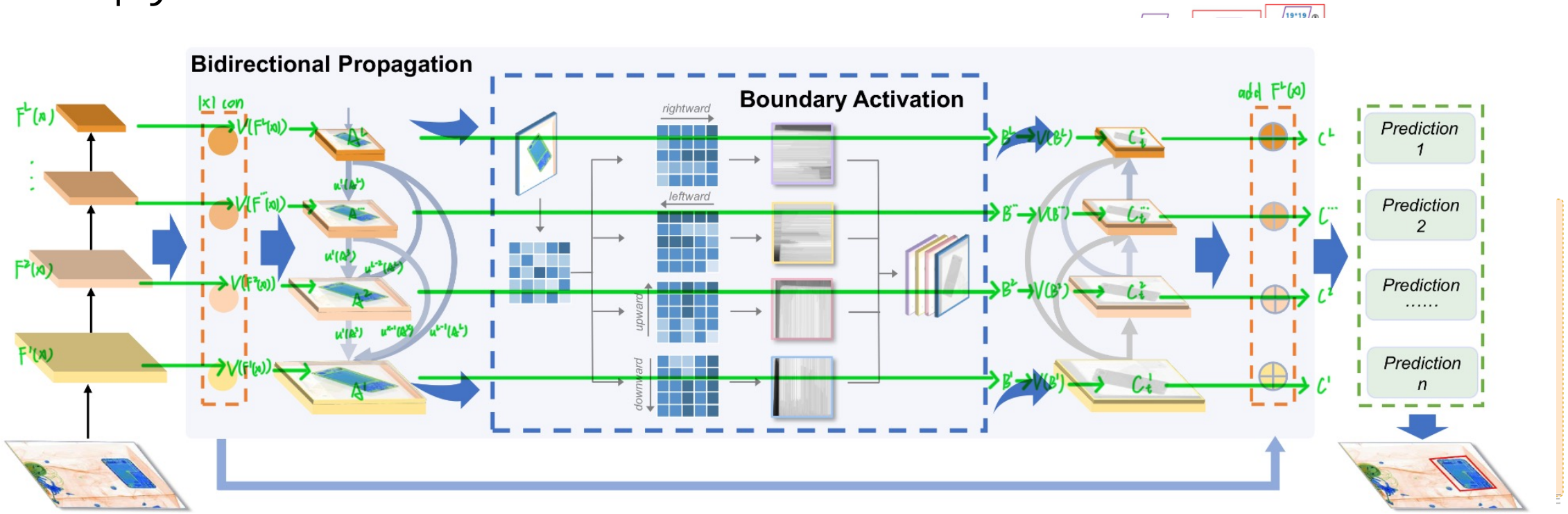
**应用案例：SSD(Single Shot Detector)目标检测算法**

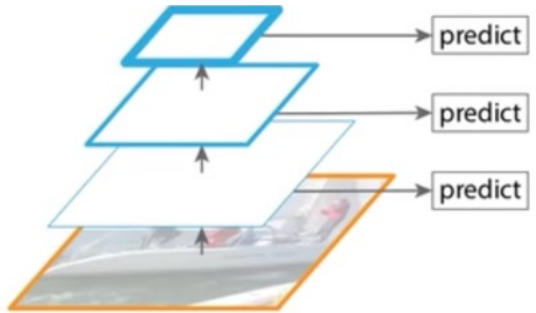
FPN：
与多尺度特征融合的方式相比较，增加了通过高层特征进行上采样和低层特征进行自顶向下的连接，然后进行下一步预测
**应用案例：Retinanet**

# PANet

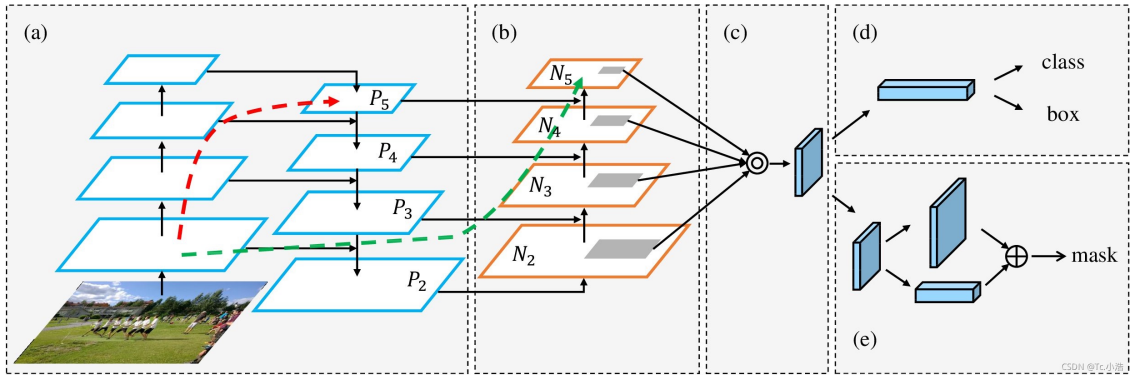


# feature pyramid




(c) Pyramidal feature hierarchy



# 5.3. Comparing with Feature Pyramid Mechanisms

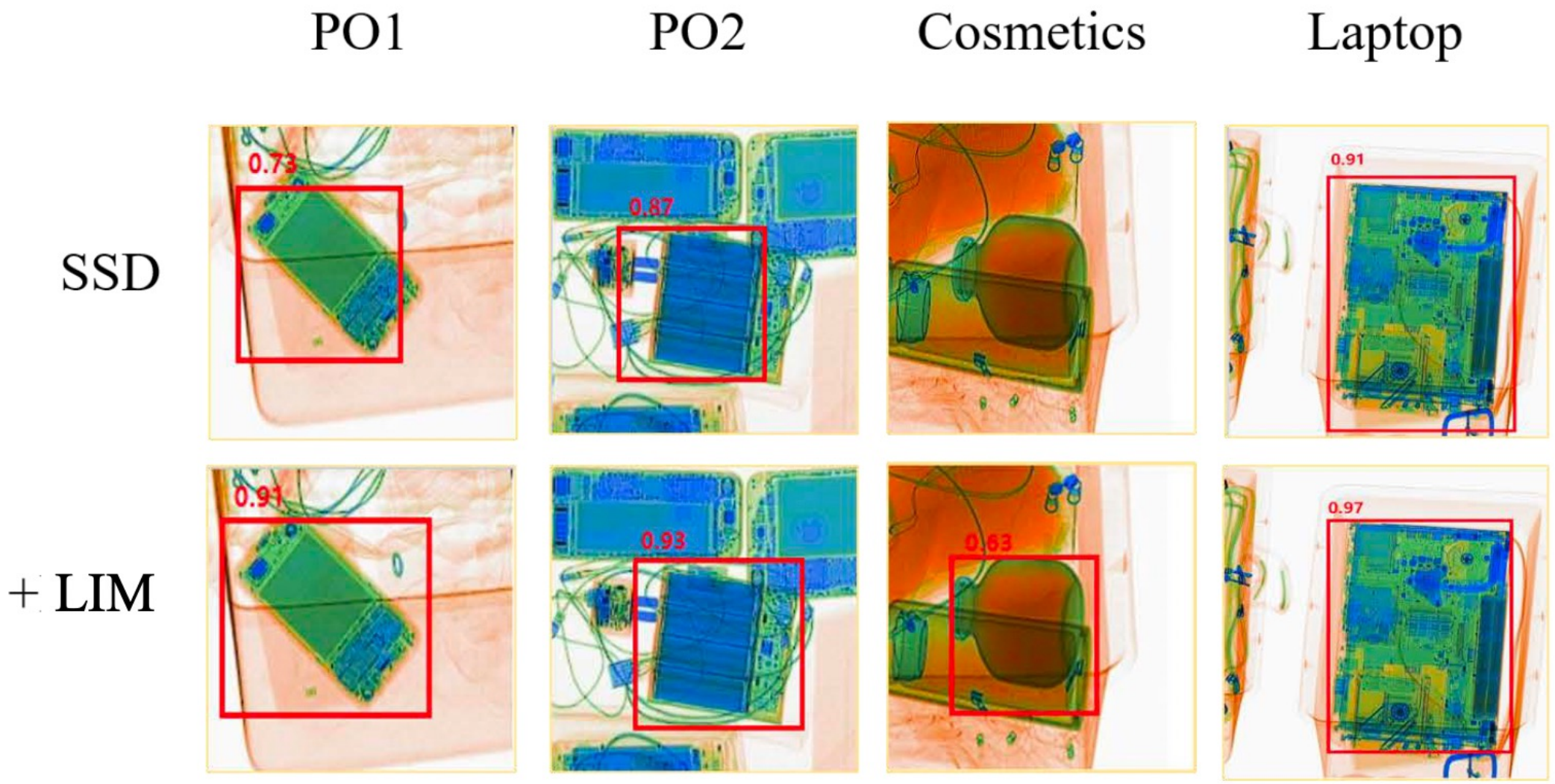
| Method | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL |
|---|---|---|---|---|---|---|---|---|---|
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 |
| +FPN [17] | 72.0 | 87.4 | 81.5 | 83.2 | 97.9 | 93.9 | 92.2 | 40.3 | 0.02 |
| +PANet [39] | 72.0 | 88.3 | 83.2 | 82.8 | **97.9** | 93.8 | **92.6** | 37.3 | 0.01 |
| **+LIM** | **73.1** | **89.1** | **84.3** | **84.0** | 97.7 | **94.5** | 92.4 | **42.3** | **0.1** |

# 5.4. Ablation Study

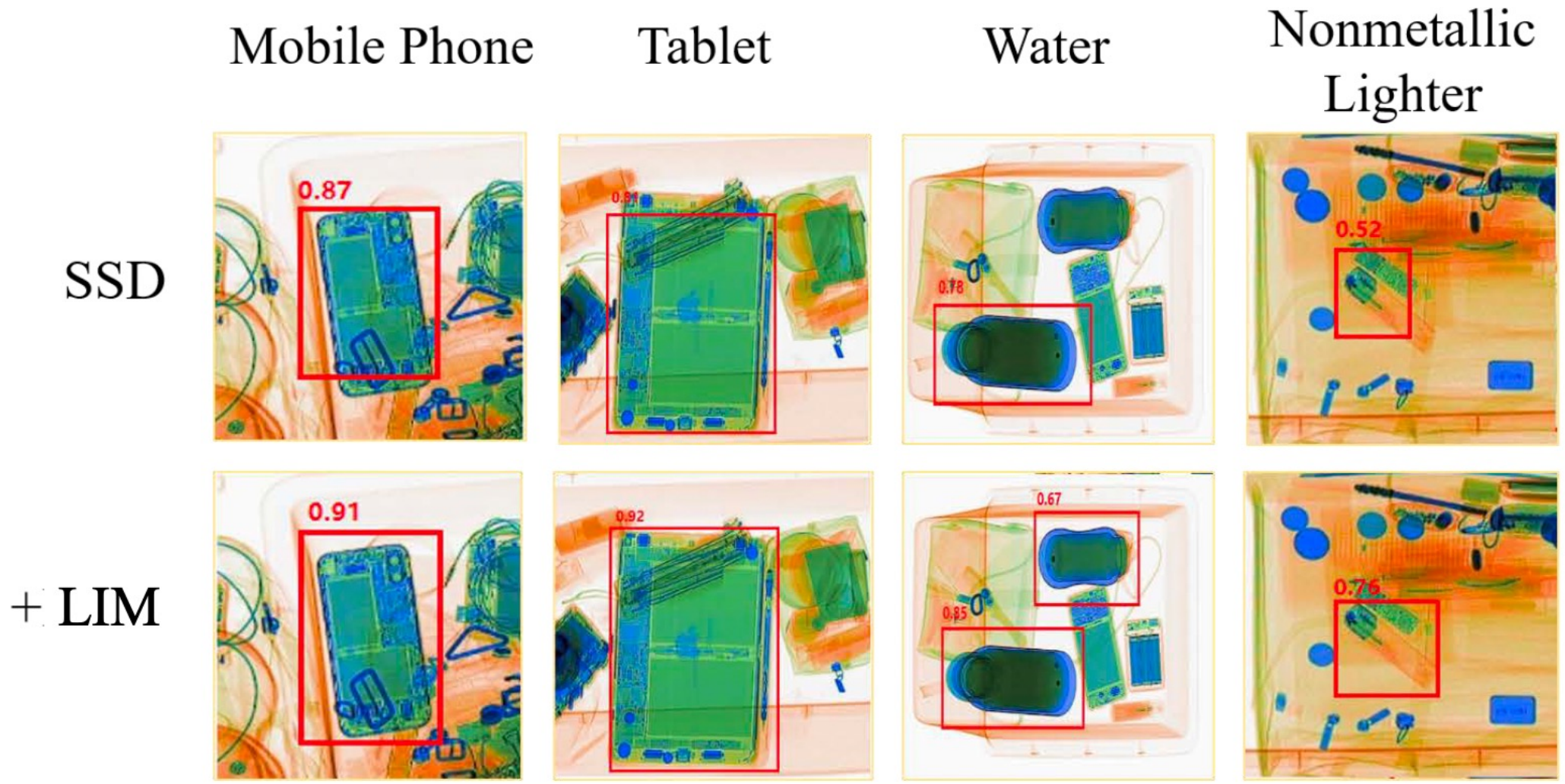
| Method | **AVG** | PO1 | PO2 | WA | LA | MP | TA | CO | NL |
|---|---|---|---|---|---|---|---|---|---|
| SSD [20] | 71.4 | 87.3 | 81.0 | 83.0 | 97.6 | 93.5 | 92.2 | 36.1 | 0.01 |
| +SP | 72.1 | 87.9 | 82.3 | 83.8 | 97.9 | 92.4 | 92.6 | 38.8 | 0.63 |
| +BP | 72.6 | 88.1 | 83.4 | 83.9 | **97.8** | 93.8 | **92.8** | 40.3 | 0.03 |
| +BP+BA | **73.1** | **89.1** | **84.3** | **84.1** | 97.7 | **94.5** | 92.4 | **42.3** | **0.1** |

# 5.5. Visualization

Portable Charger



# 5.5. Visualization



# 5.5. Visualization

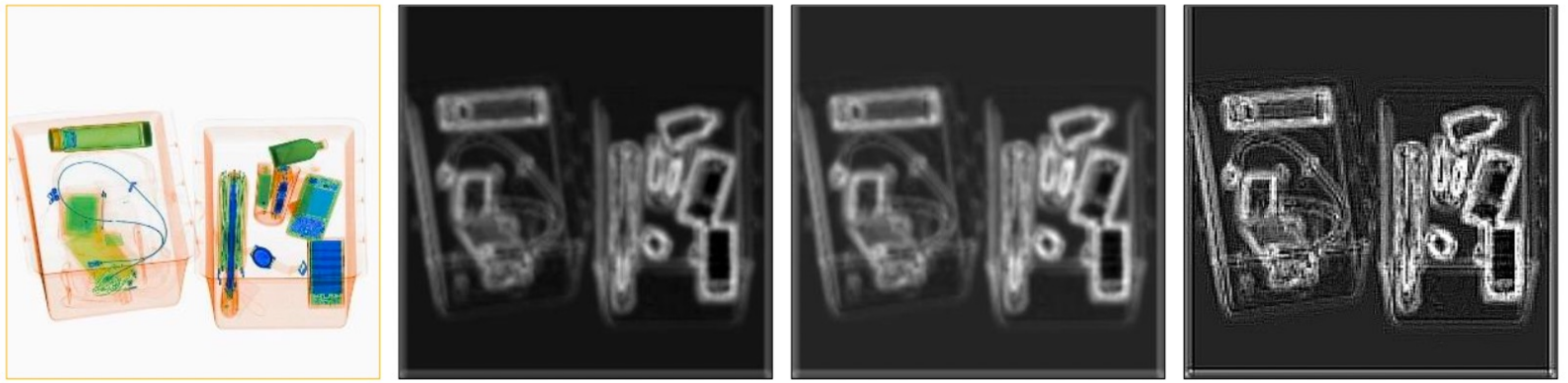



Figure 6 illustrated the effectiveness of our LIM and traditional boundary-enhanced methods, including DOAM [40], EEMEFN [44], etc.