- 分类网络
  - ✓ LeNet
  - ✓ AlexNet
  - ✓ VGGNet
  - ✓ GoogLeNet/Inception
  - ✓ MobileNet
  - ✓ ResNet
- 目标检测网络
  - ✓ RCNN
  - ✓ YOLO

分类网络

分类网络

- LeNet
- AlexNet
- VGGNet
- GoogLeNet/Inception
- MobileNet
- ResNet

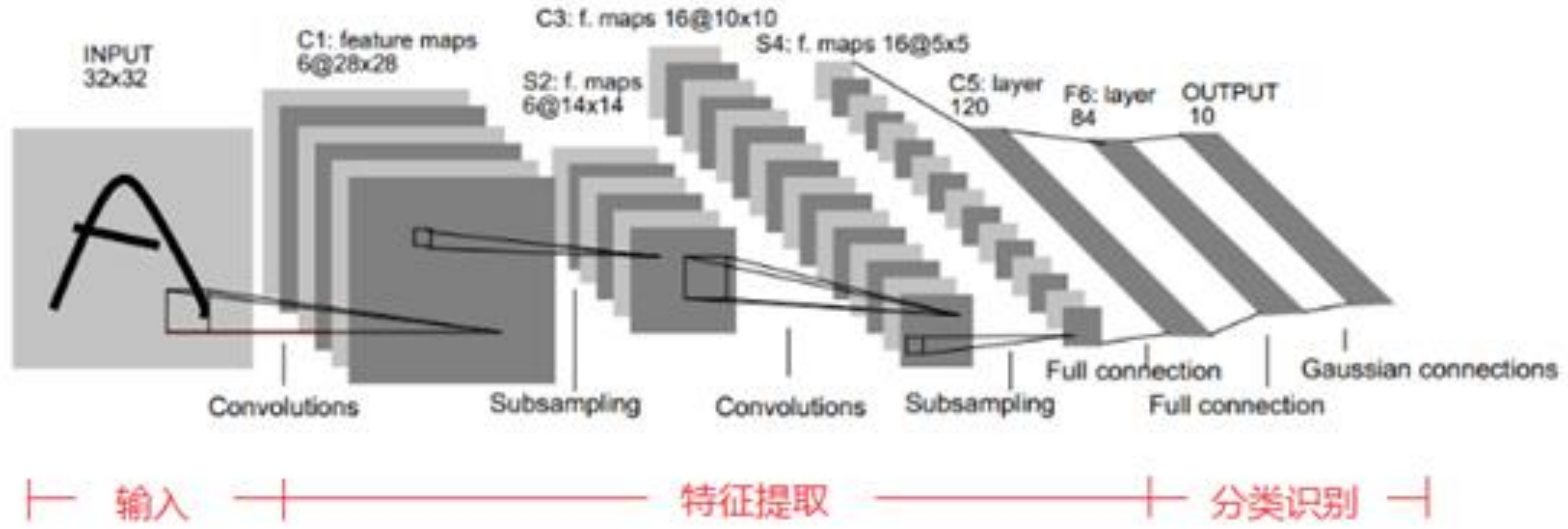
Tricks大集合

分类网络

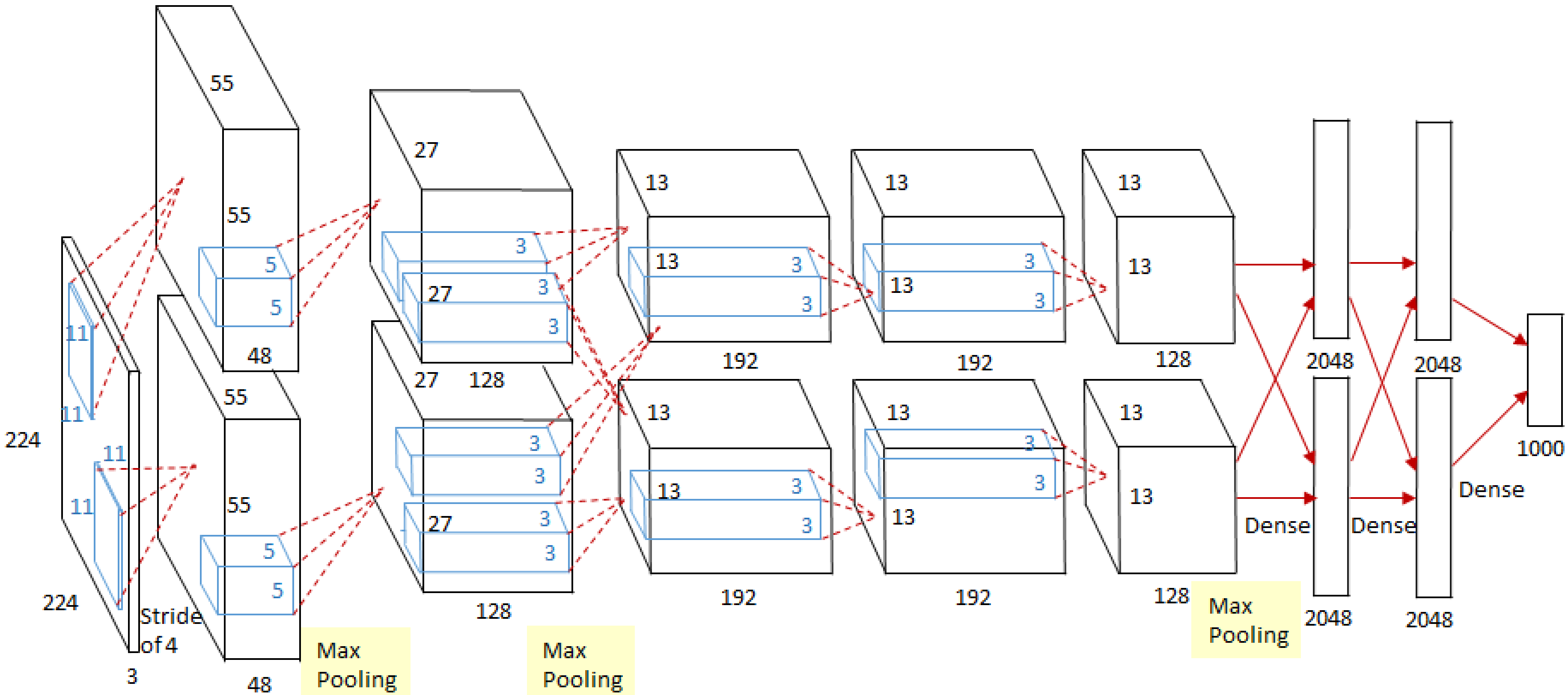
- LeNet
- AlexNet：ReLU, 分组卷积，Dropout, 数据增强
- VGGNet：3 X 3 小卷积核代替之前的5 X 5、7 X 7
- GoogLeNet/Inception：瓶颈结构，全局平均池化层
- MobileNet：深度可分卷积，批归一化BN
- ResNet：残差块

LENET

手写字体识别模型LeNet 诞生于1994年

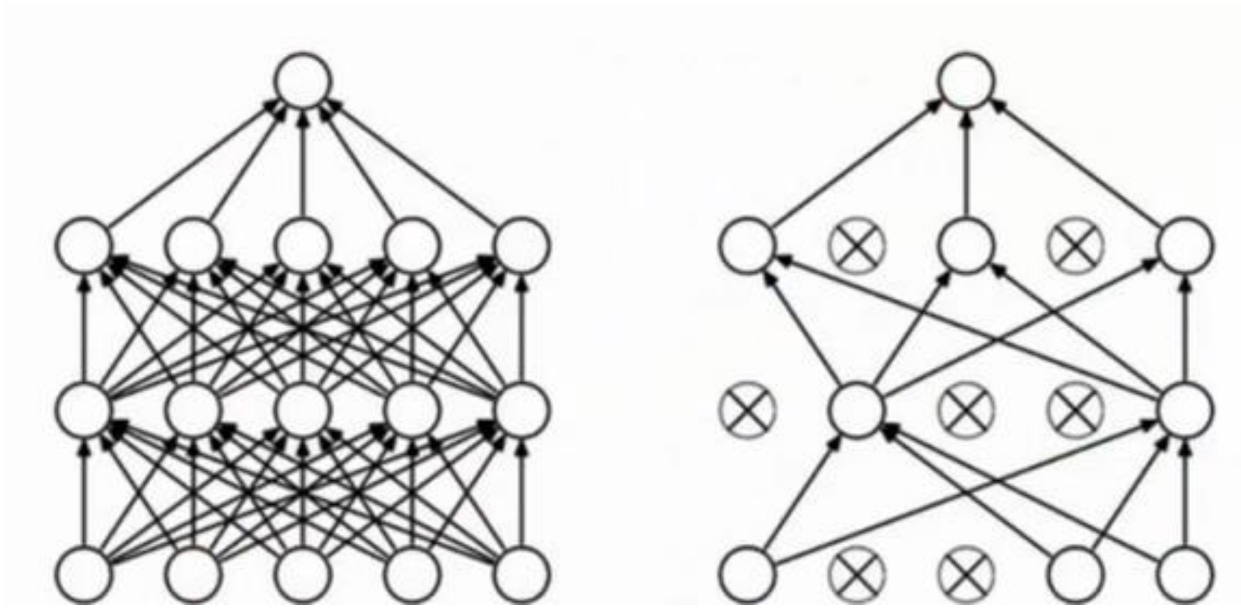


ALEXNET



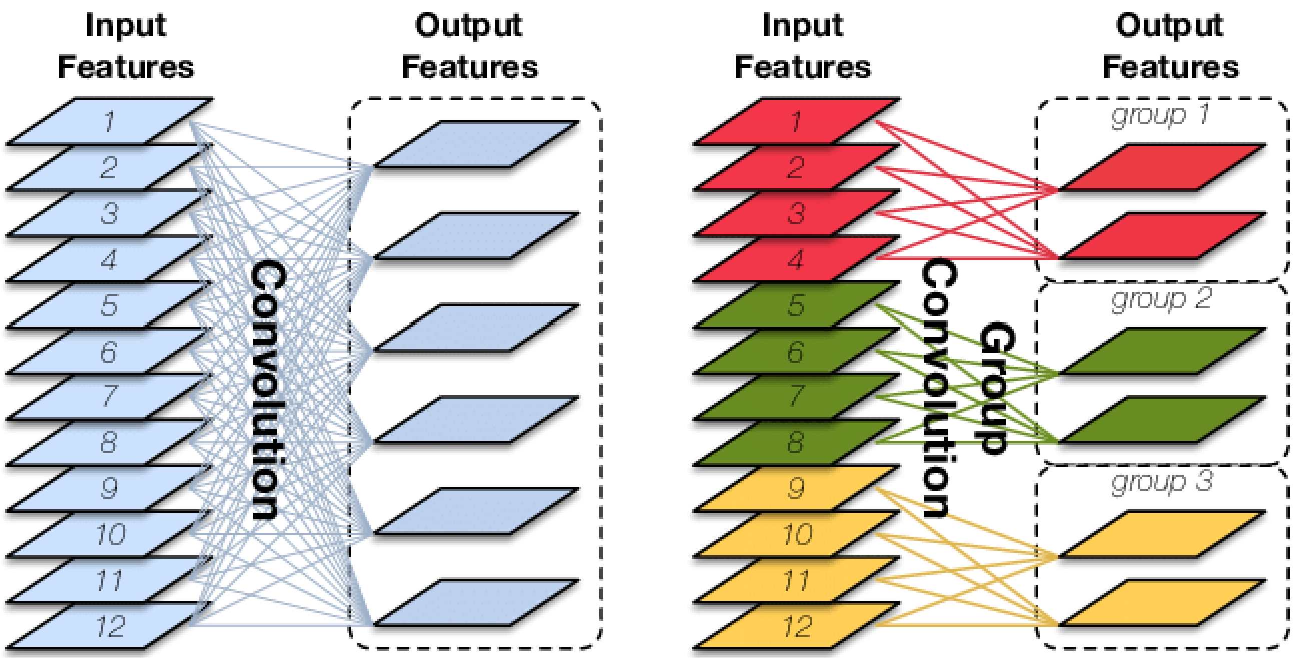
- 采用ReLU作为激活函数（替代了之前常用的Sigmoid函数），缓解了深层网络训练时的梯度消失问题。
- 应用了Dropout和数据增强（data augmentation）技术来提升训练结果。
- 用分组卷积突破当时GPU的显存瓶颈。

丢弃 /DROPOUT



- ✓ 防止过拟合的一大利器
- ✓ 在训练数据上防止了单元之间共同起作用
- ✓ 隐层的单元不再依赖于其他单元
- ✓ 促使每一个单元学习到有效的特征

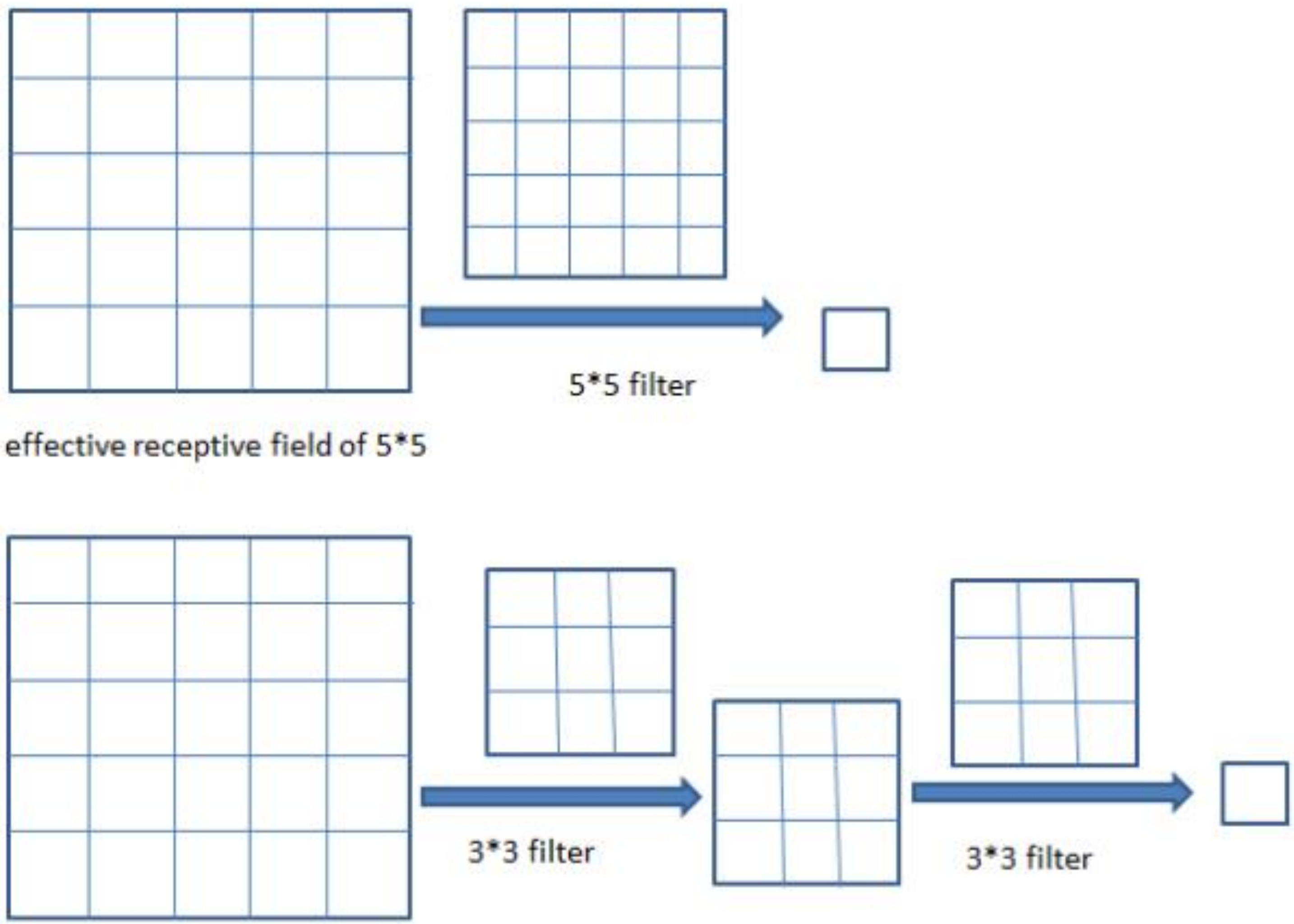
ALEXNET 分组卷积



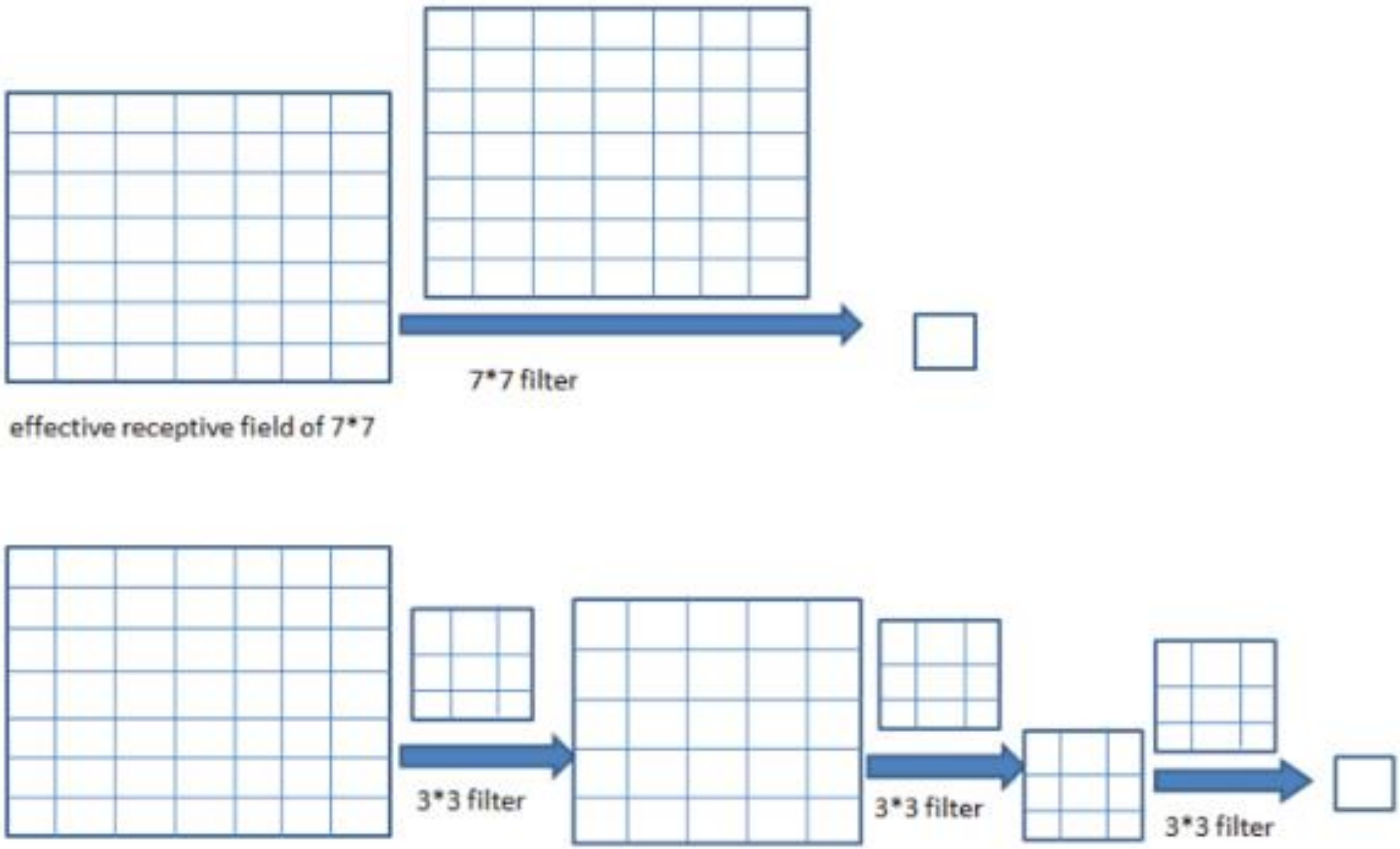
分组卷积：如果g为输入/输出通道所分的组数，则分组卷积能够将卷积操作的参数量和计算量都降低为普通卷积的1/g。

用多个3 X 3小卷积核代替之前的5 X 5、7 X 7等大卷积核，这样可以在更少的参数量、更小的计算量下，获得同样的感受野以及更大的网络深度。

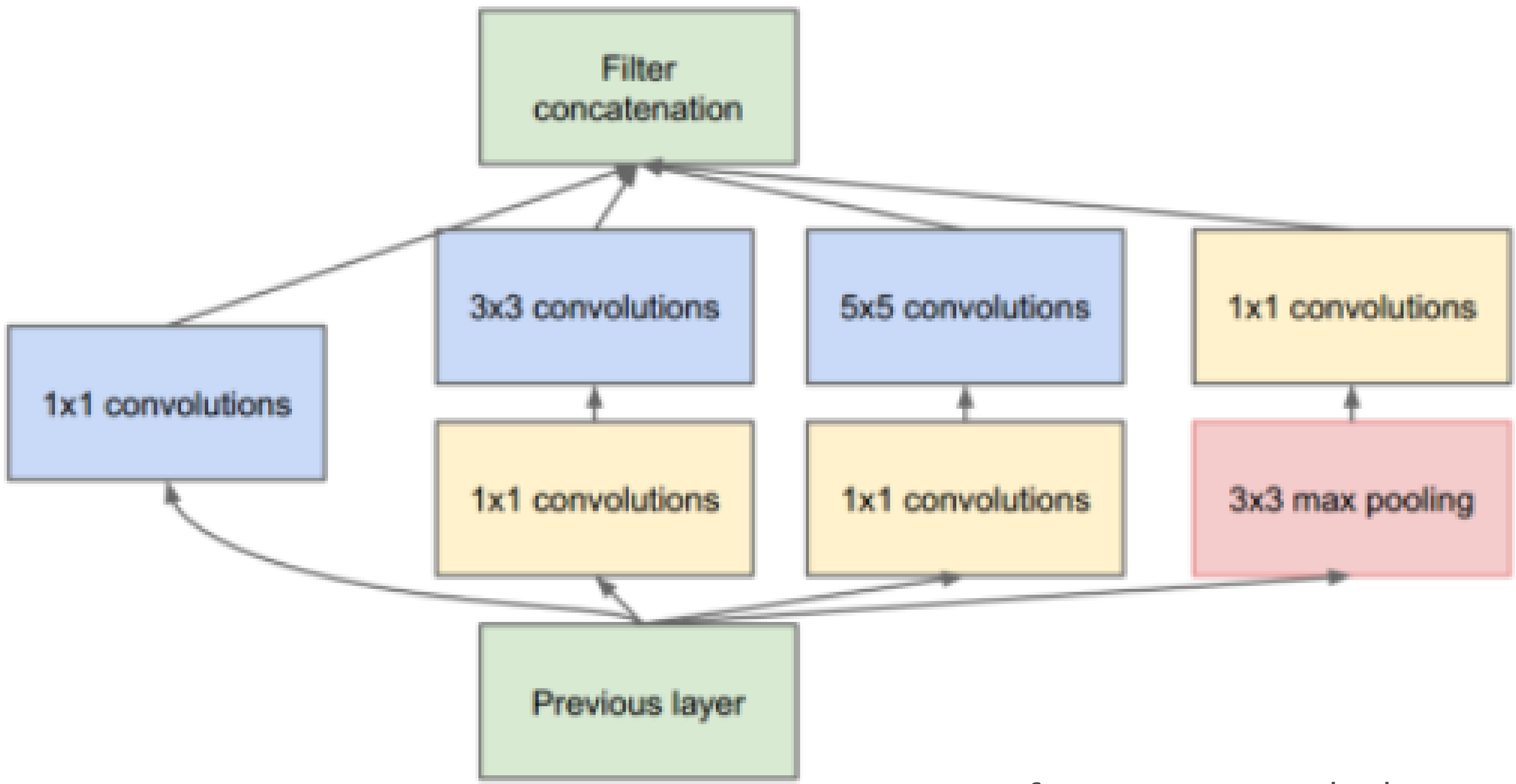
VGGNET


5*5 filter
effective receptive field of 5*5
3*3 filter
3*3 filter

VGGNET


7*7 filter
effective receptive field of 7*7
3*3 filter
3*3 filter
3*3 filter

GOOGLENET/INCEPTION



feature map: M*N*10
Filter：1个1*1*10
output：？ M*N*1

- 瓶颈（ bottleneck）结构

瓶颈（ **bottleneck**）结构，即在计算比较大比较厚的卷积层之前，先使用**1x1**卷积对其通道进行压缩以减少计算量（在较大卷积层完成计算之后，根据需要有时候会再次使用**1x1**卷积将其通道数复原）。

GOOGLENET

修改了之前VGGNet等网络在网络末端加入多个全连接层进行分类的做法，转而将第一个全连接层换成全局平均池化层（Global Average Pooling）。

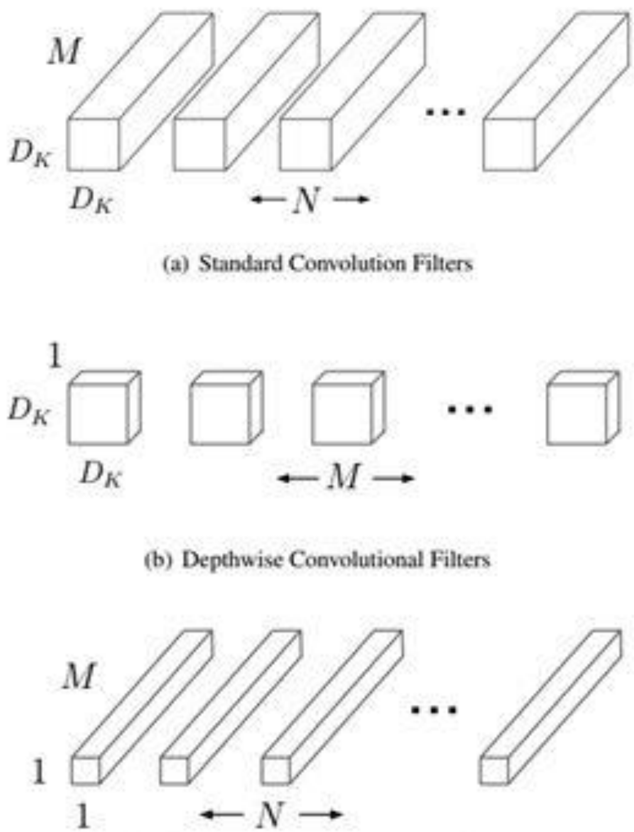
- 对每个特征图一整张图片进行全局均值池化，每张特征图都可以得到一个输出
- 采用平均池化，不需要参数，可以大大减小网络参数，避免过拟合
- 每张特征图相当于一个输出特征，增加可解释性

用于分类任务的卷积神经网络的最后几层一般是什么层？可以有什么变化？

卷积层：局部信息
全连接：(多层感知机) 空间位置的相关性，语义信息的相关性，全局信息
通常搭配：卷积层+全连接

全局平均池化
1）参数量，计算量降低
2）避免过拟合
3）较好的解释性

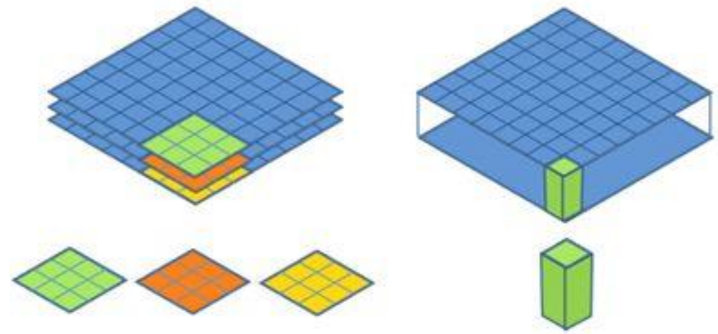
Depthwise separable convolution 深度可分离卷积



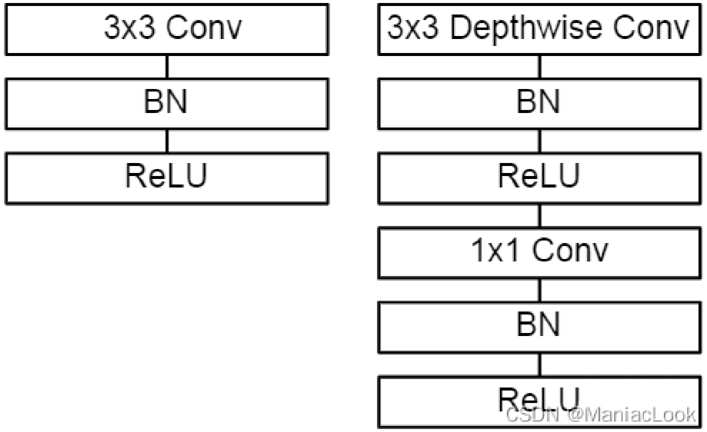
(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution



Depthwise Convolutional Filters　　Pointwise Convolutional Filters



卷积核 D_k,M,N, 输出的长宽是D_f，乘积的次数是多少？

Input Volume (+pad 1) (7x7x3)

x[:,:,0]

0	0	0	0	0	0	0
0	0	1	1	0	2	0
0	2	2	2	2	1	0
0	1	0	0	2	0	0
0	0	1	1	0	0	0
0	1	2	0	0	2	0
0	0	0	0	0	0	0

x[:,:,1]

0	0	0	0	0	0	0
0	1	0	2	2	0	0
0	0	0	0	2	0	0
0	1	2	1	2	1	0
0	1	0	0	0	0	0
0	1	2	1	1	1	0
0	0	0	0	0	0	0

x[:,:,2]

0	0	0	0	0	0	0
0	2	1	2	0	0	0
0	1	0	0	1	0	0
0	0	2	1	0	1	0
0	0	1	2	2	2	0
0	2	1	0	0	1	0
0	0	0	0	0	0	0

Filter W0 (3x3x3)

w0[:,:,0]

-1	1	0
0	1	0
0	1	1

w0[:,:,1]

-1	-1	0
0	0	0
0	-1	0

w0[:,:,2]

0	0	-1
0	1	0
1	-1	-1

Bias b0 (1x1x1)

b0[:,:,0]

1

Filter W1 (3x3x3)

w1[:,:,0]

1	1	-1
-1	-1	1
0	-1	1

w1[:,:,1]

0	1	0
-1	0	-1
-1	1	0

w1[:,:,2]

-1	0	0
-1	0	1
-1	0	0

Bias b1 (1x1x1)

b1[:,:,0]

0

Output Volume (3x3x2)

o[:,:,0]

6	7	5
3	-1	-1
2	-1	4

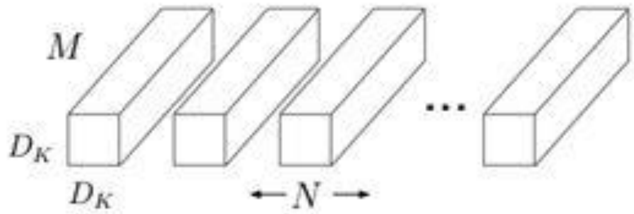
o[:,:,1]

2	-5	-8
1	-4	-4
0	-5	-5

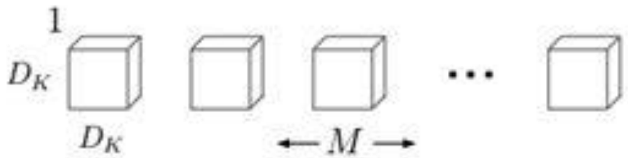
toggle movement

$D_k=3, M=3, N=2, D_f=3$

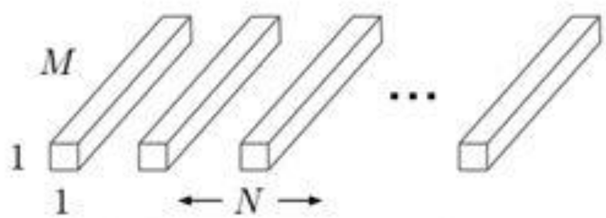
Depthwise separable convolution



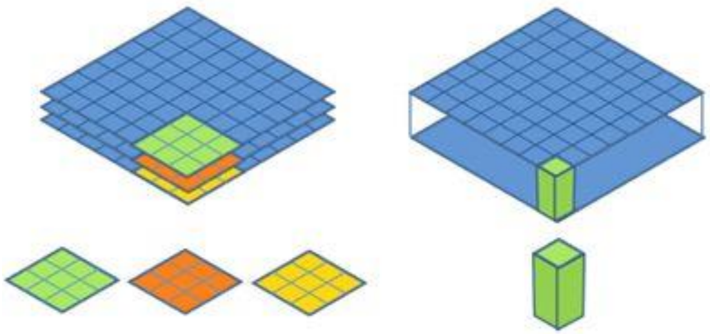
(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



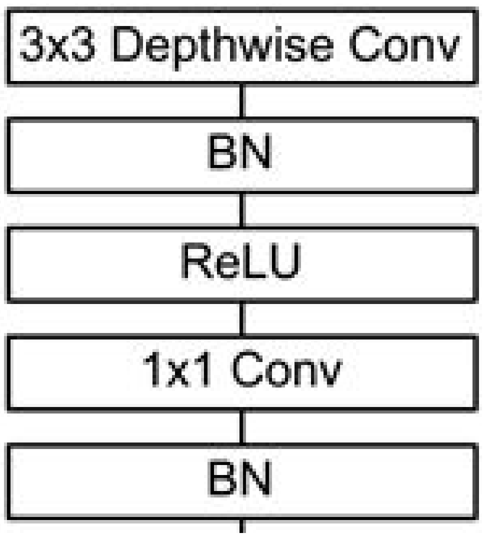
(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

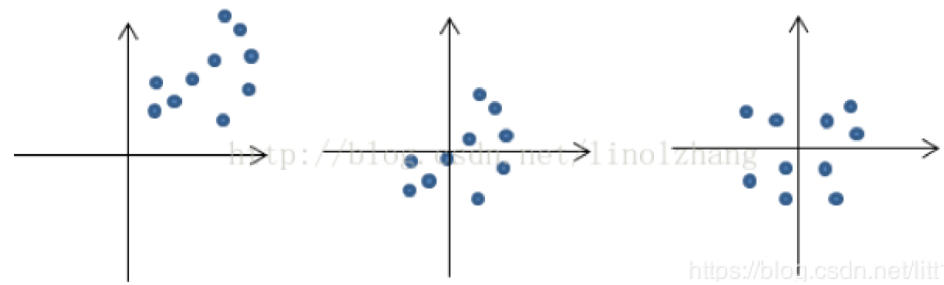


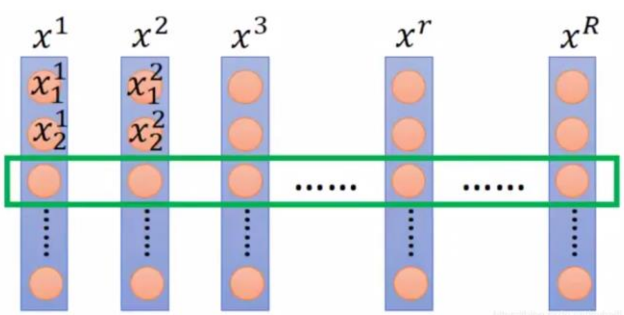
Depthwise Convolutional Filters　Pointwise Convolutional Filters

$$\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2}$$

Batch Normalization







$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

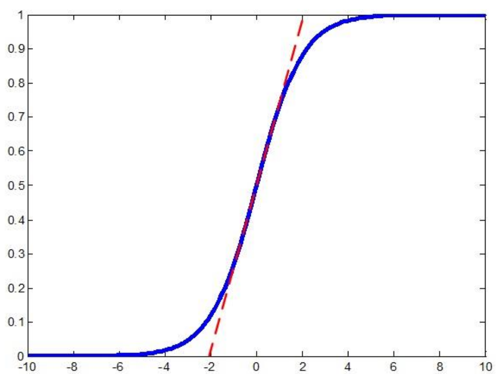
$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

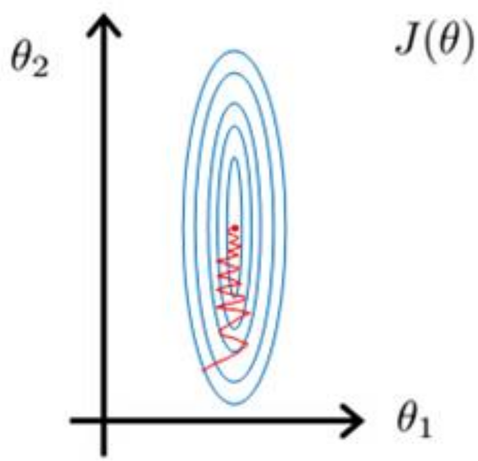
$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

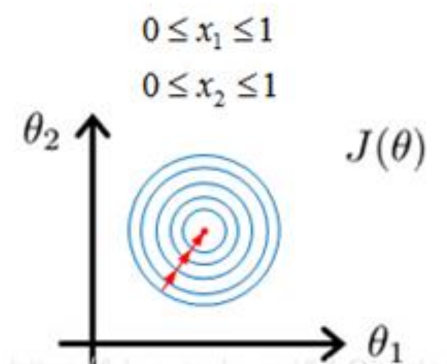
BATCH NORMALIZATION (BN)

作用：

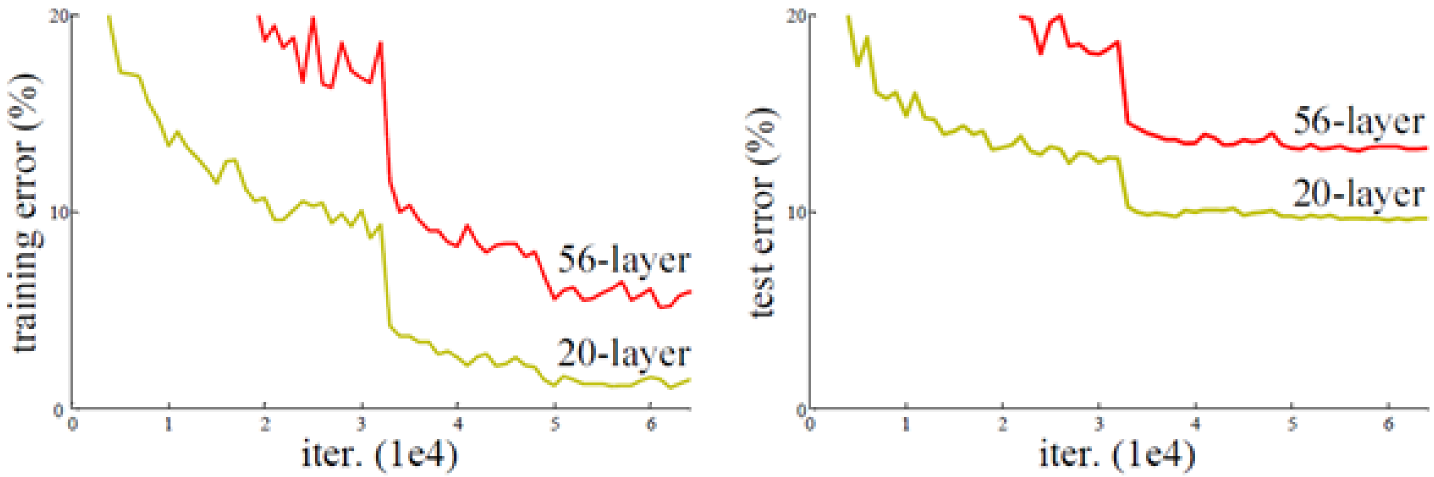
- 加快收敛速度：每层数据分布一致
- 防止梯度消失
- 防止过拟合：mini batch中所有样本都被关联在了一起，相互影响，消弱单个噪声的作用。







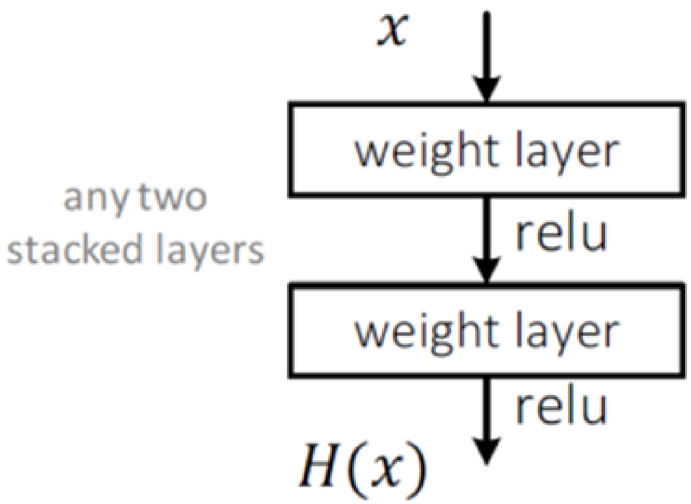
深度网络的退化问题（Degradation problem）



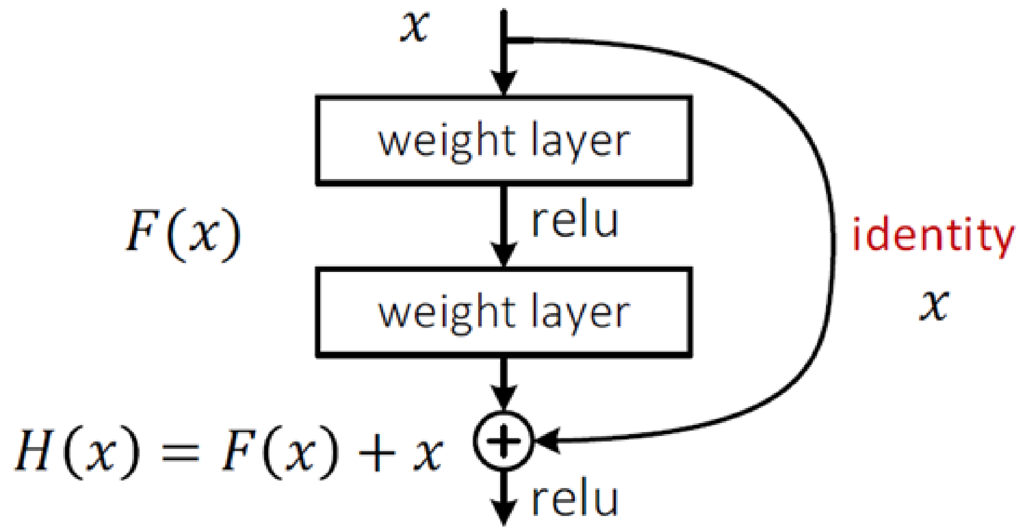
随着网络层数的加深，网络的训练误差和测试误差都会上升。

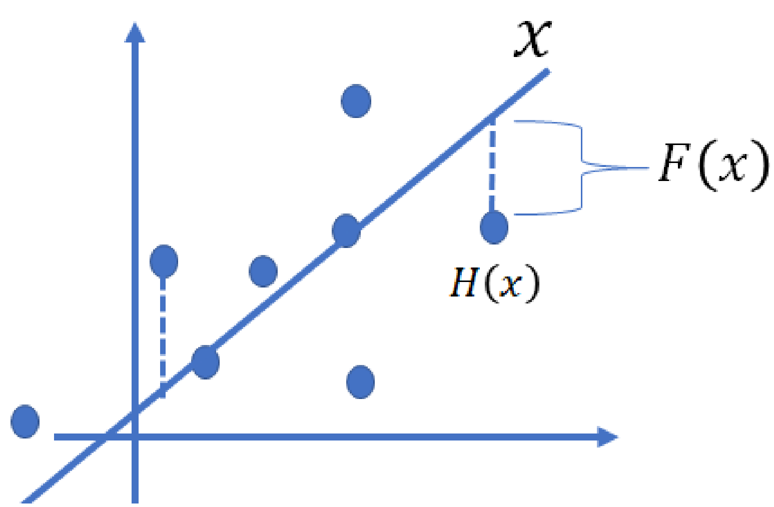
RESNET

- Plaint net



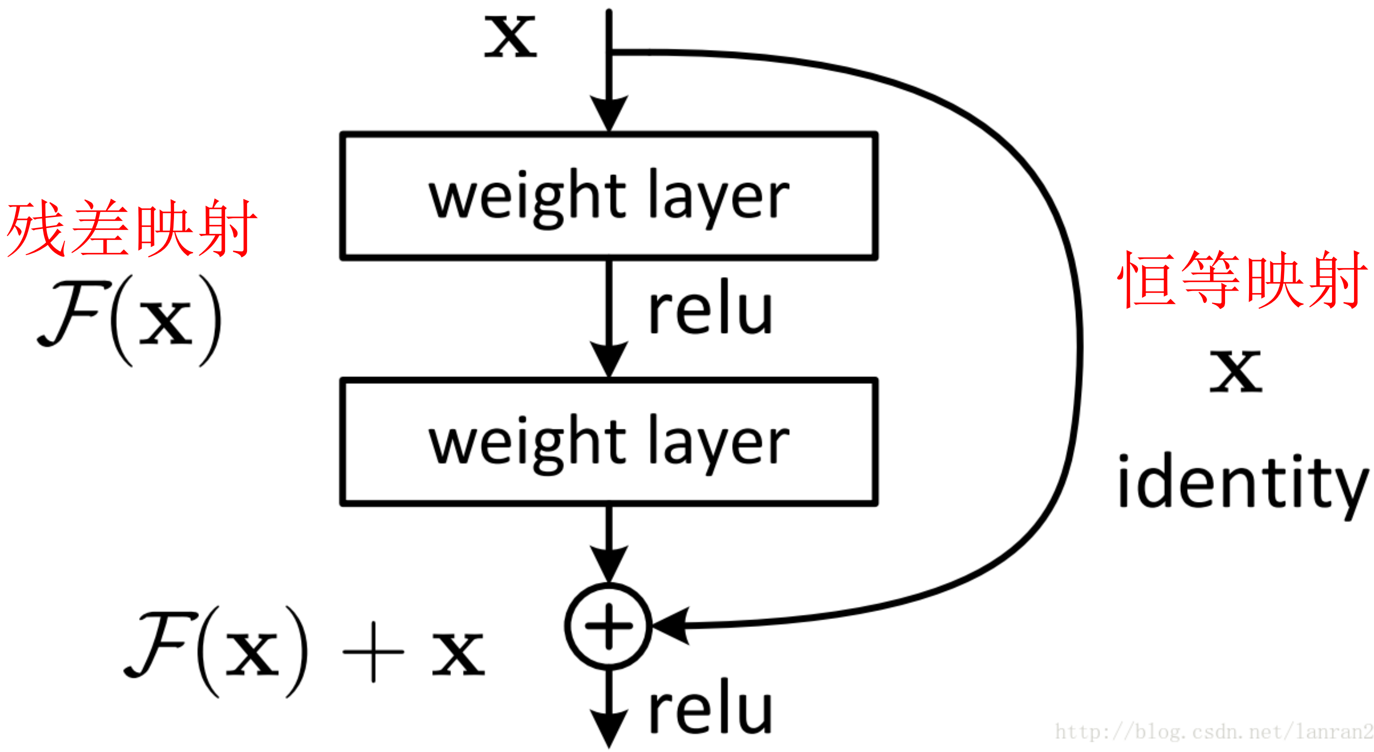
- Residual net





RESNET

Shortcut connection

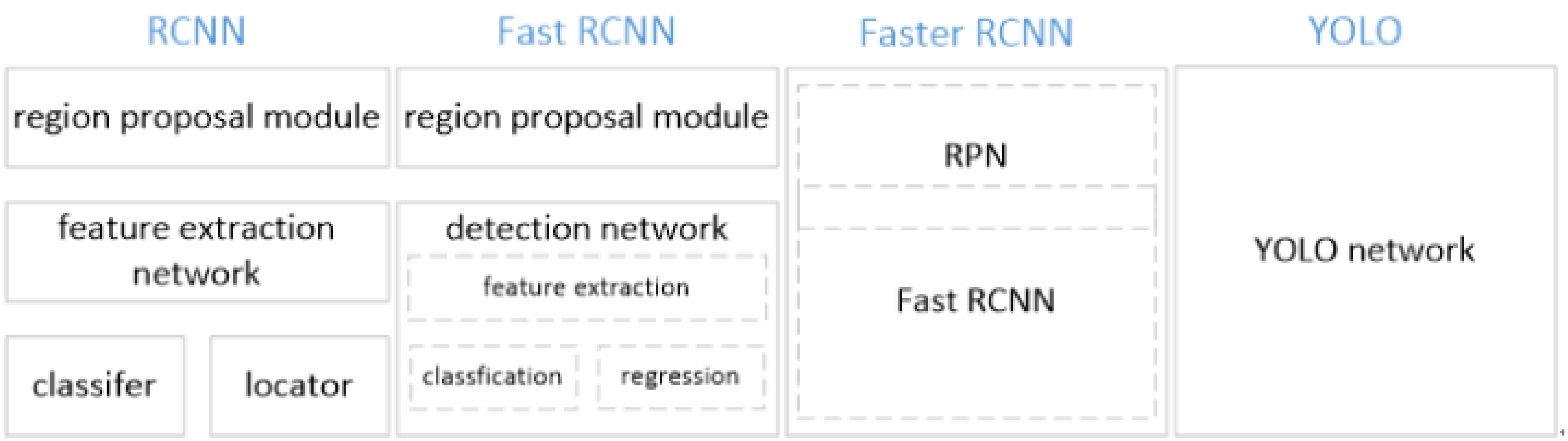


$$\mathcal{F} = W_2\sigma(W_1\mathbf{x}) \qquad \mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

OBJECT DETECTION

OBJECT DETECTION

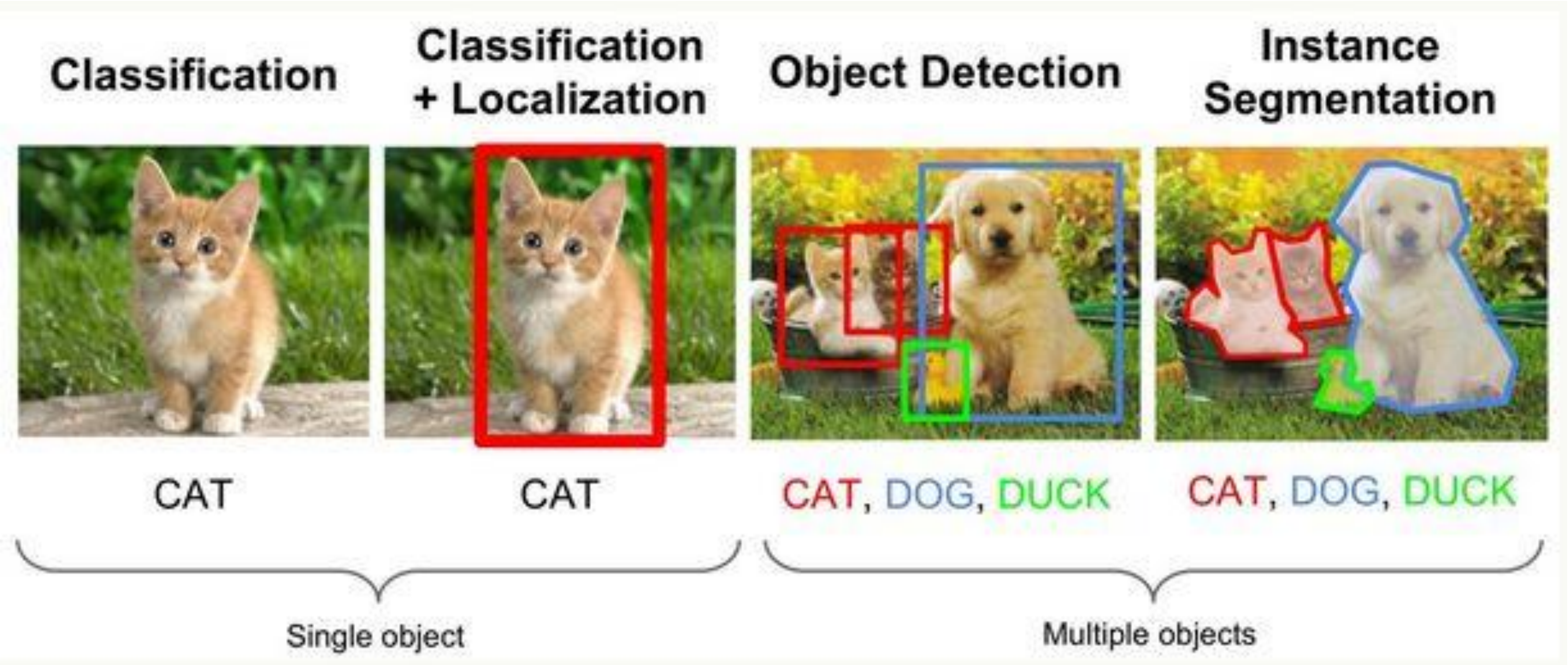
- 计算机视觉领域基础知识
- 二阶段RCNN (Region CNN) 系列
- 一阶段YOLO(You only look once)系列



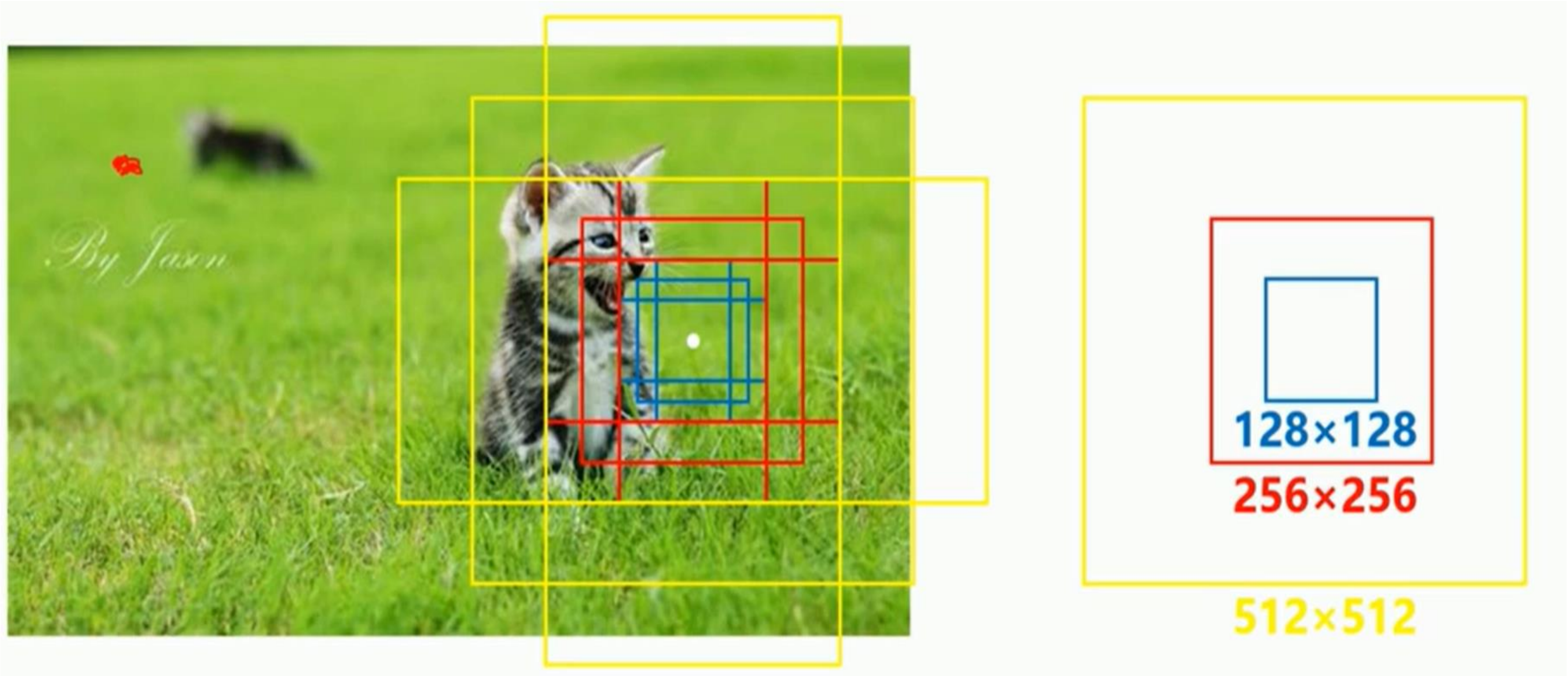
计算机视觉领域基础知识

计算机视觉领域的三大任务

a. 分类 b. 目标检测 c. 分割

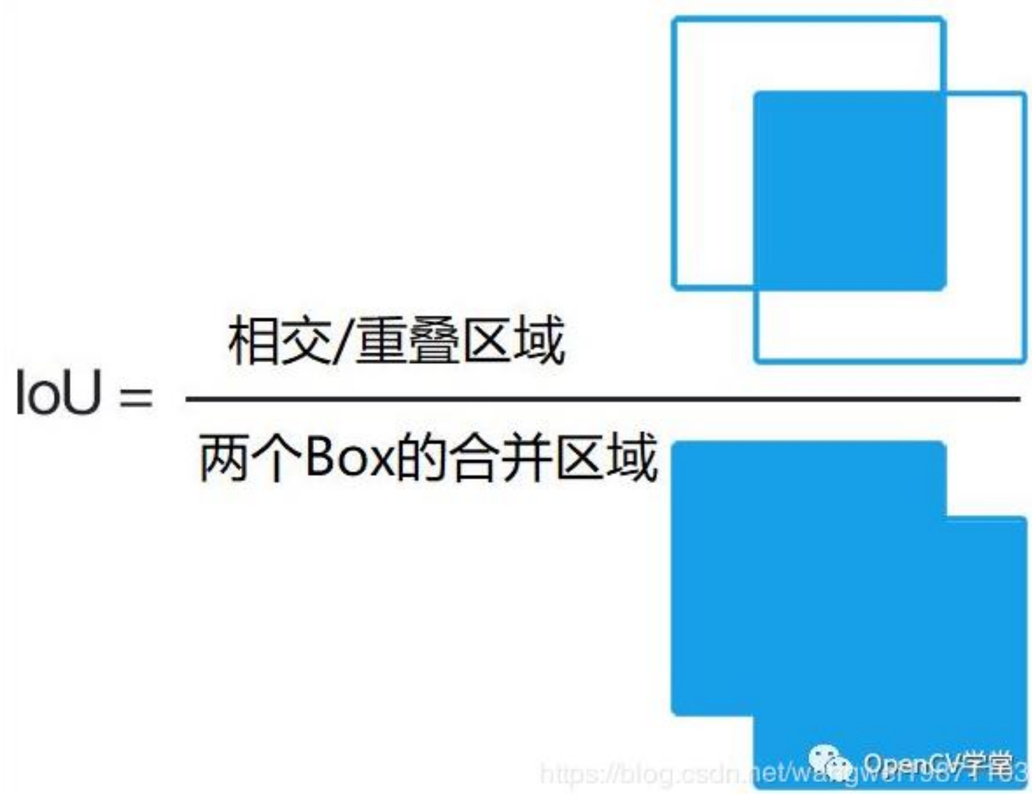


锚框

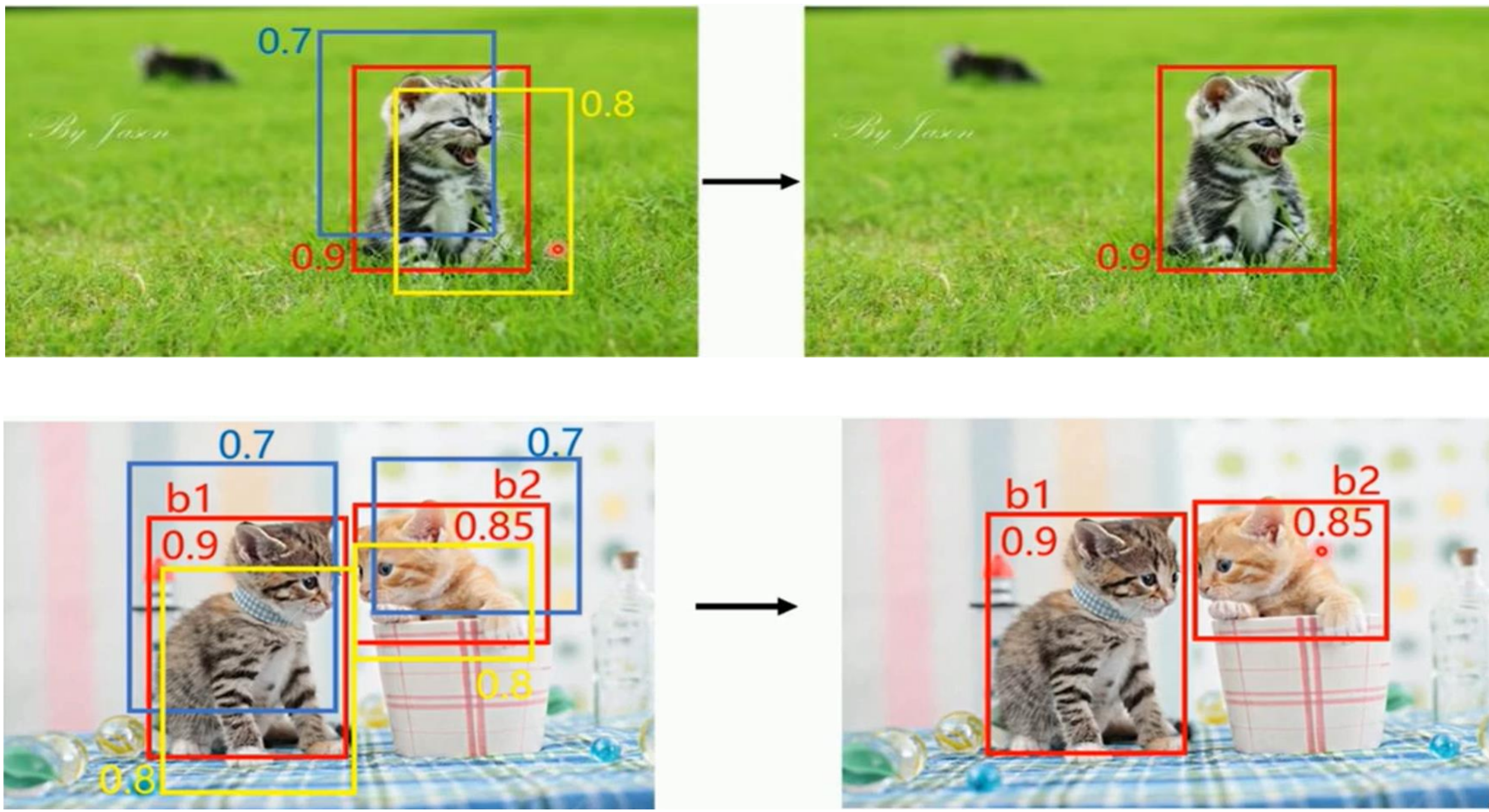


IOU(Intersection Over Union)

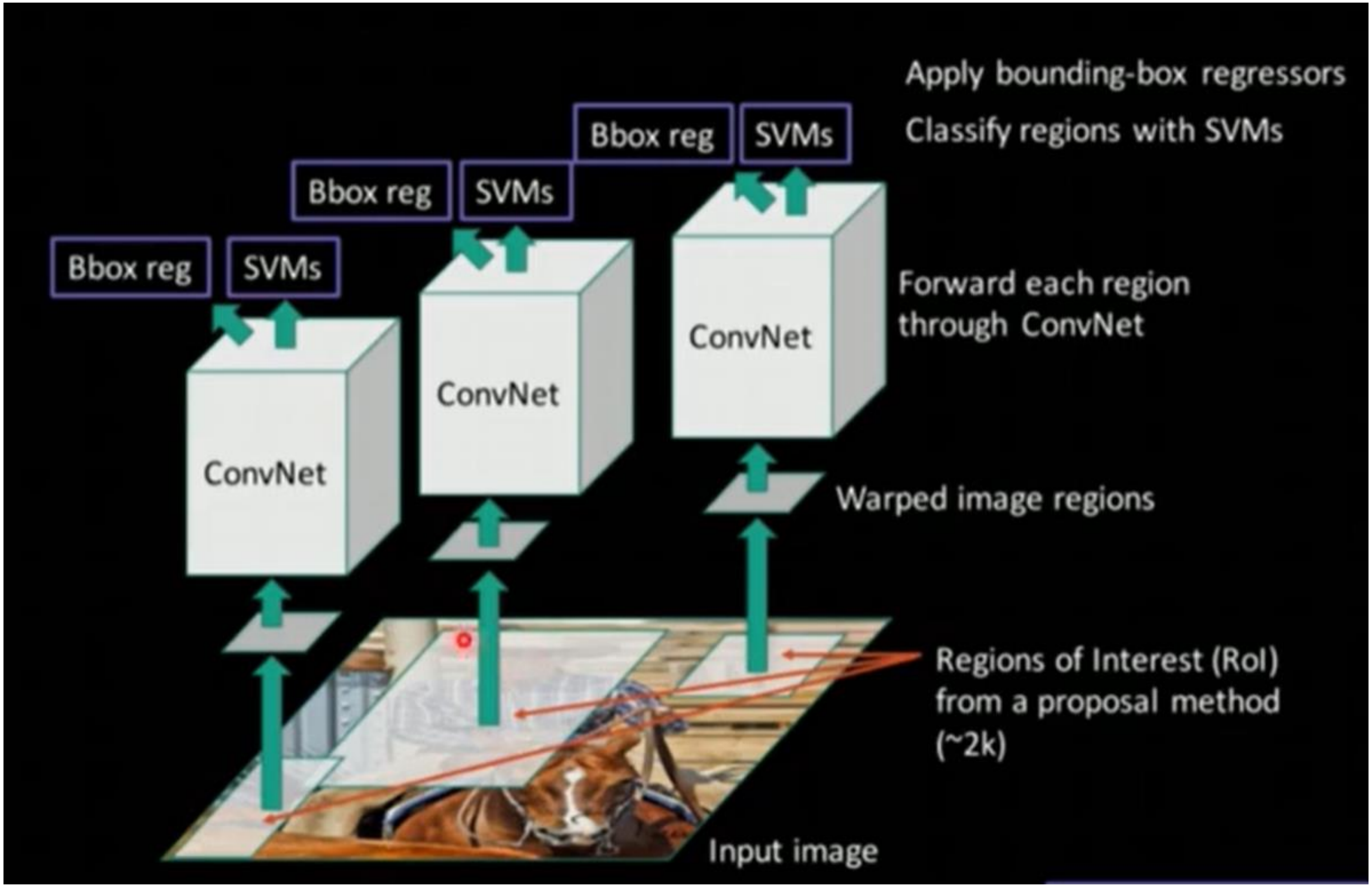
预测框和人工标注框(Groud Truth)的重合度



非最大值抑制 (NMS, Non-Maximum Suppression)

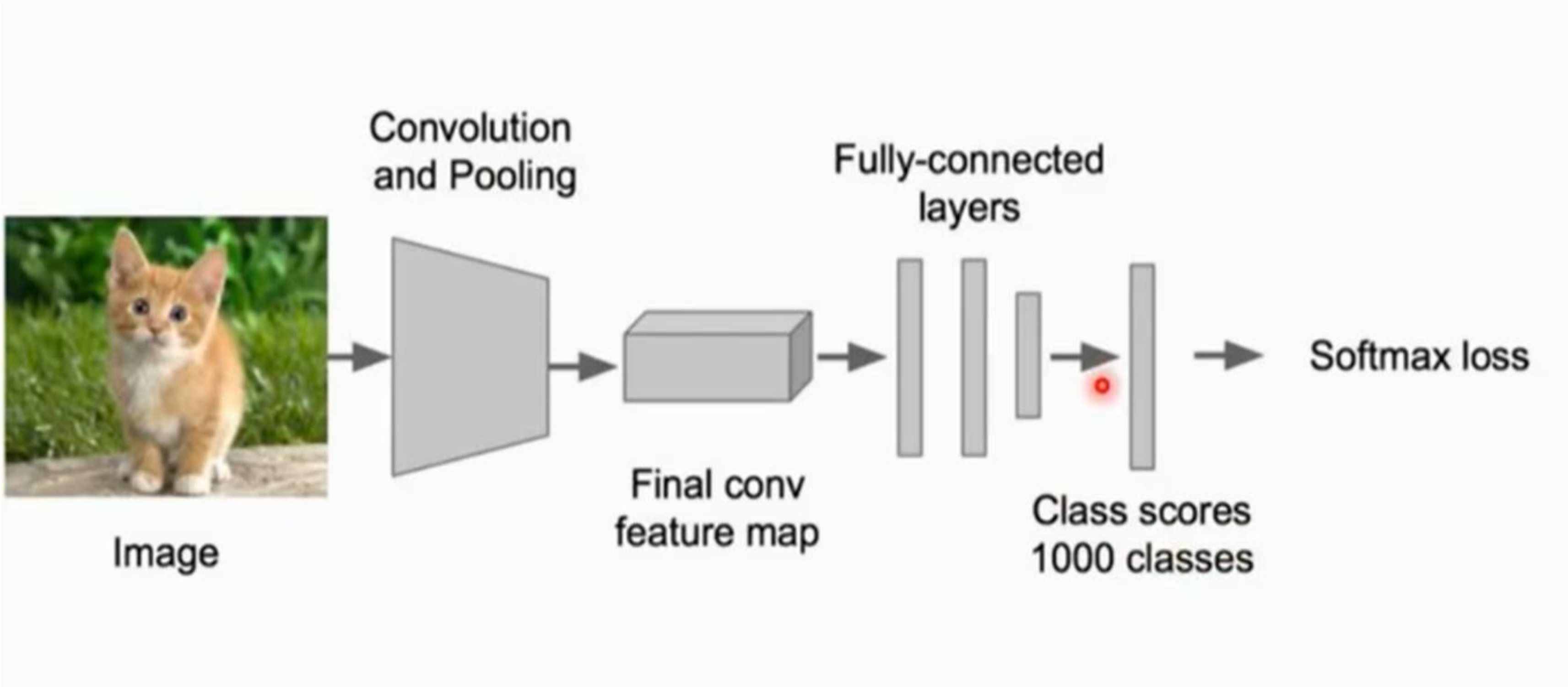


RCNN算法整体流程


Apply bounding-box regressors
Classify regions with SVMs
Bbox reg
SVMs
Bbox reg
SVMs
Bbox reg
SVMs
ConvNet
ConvNet
ConvNet
Forward each region through ConvNet
Warped image regions
Regions of Interest (RoI) from a proposal method (~2k)
Input image

RCNN (REGION CNN)

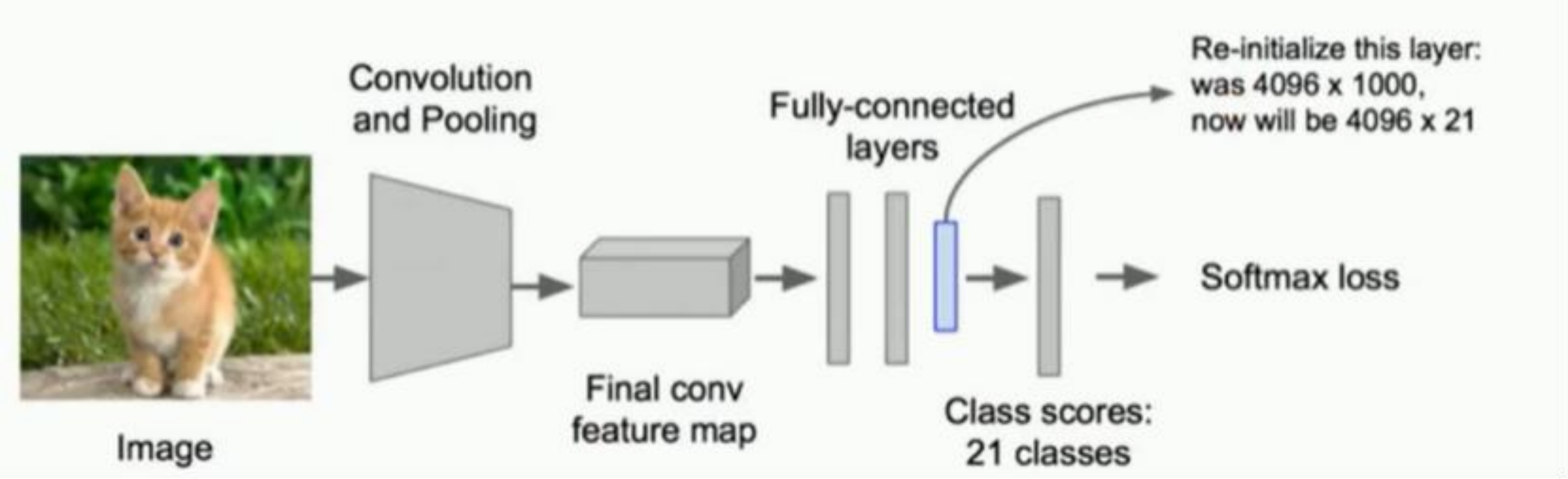
- 预训练一个分类模型



RCNN (REGION CNN)

- Fine-tuning微调

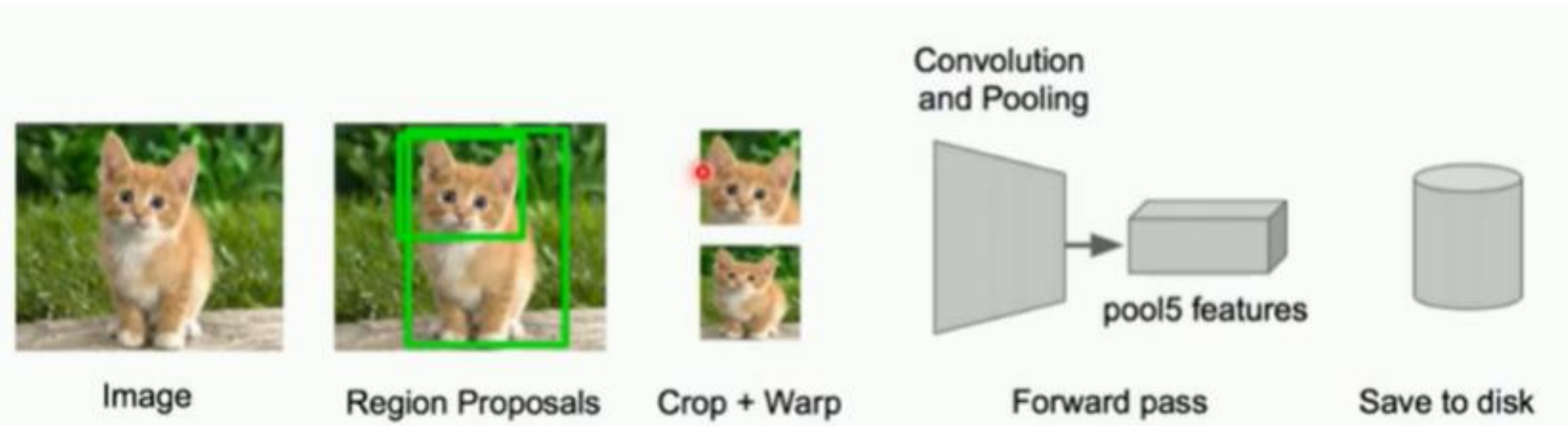
1. 去掉最后一个全连接层。
2. 将分类数从1000改为(N+1)。对于VOC，N=20；对于ILSVRC2013，N=200.
3. 对该模型做fine-tuning



RCNN (REGION CNN)

- 特征提取

1. 对每一个候选区域进行特征提取；
2. Resize区域大小，然后做一次前向运算，候选框提取到的特征保存。

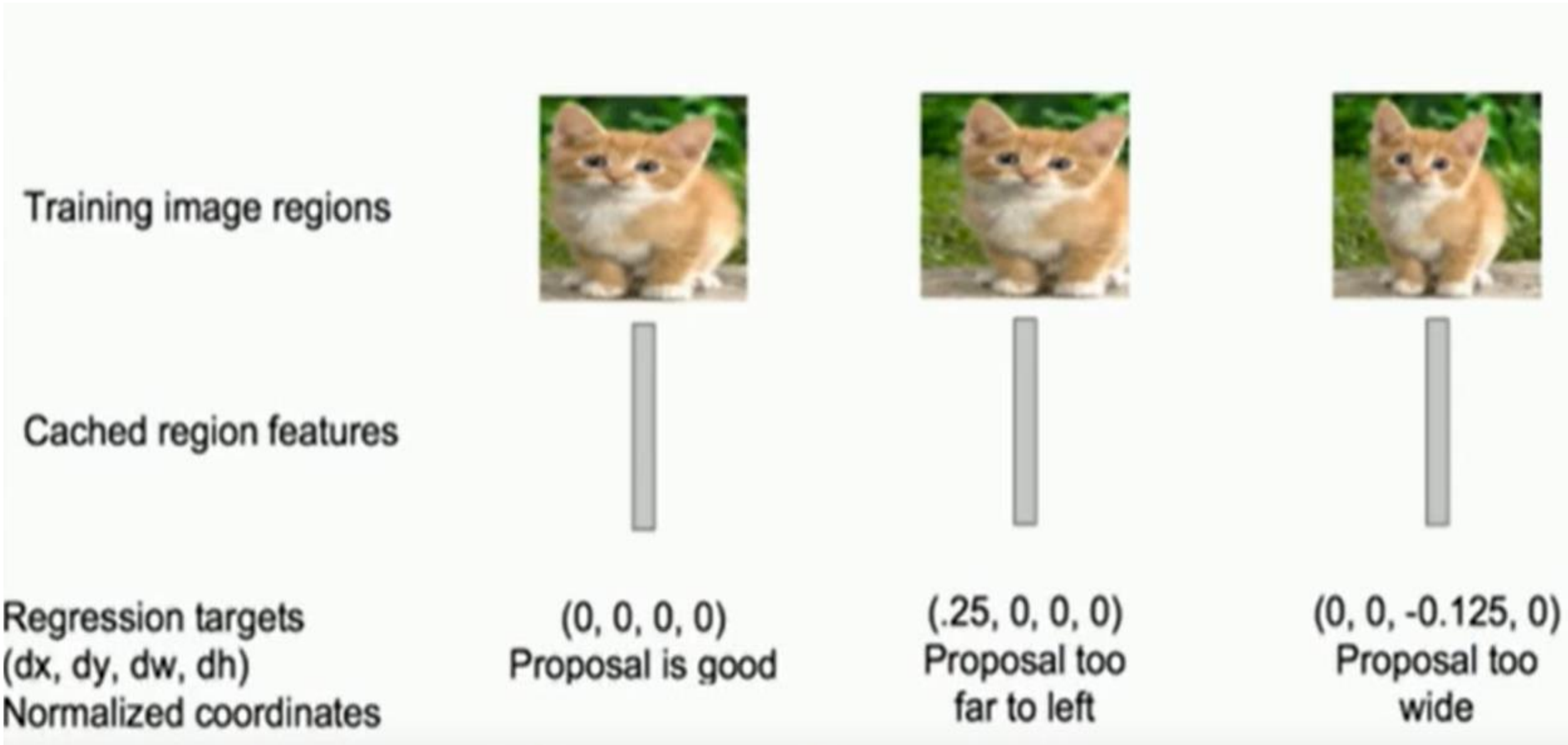


RCNN (REGION CNN)

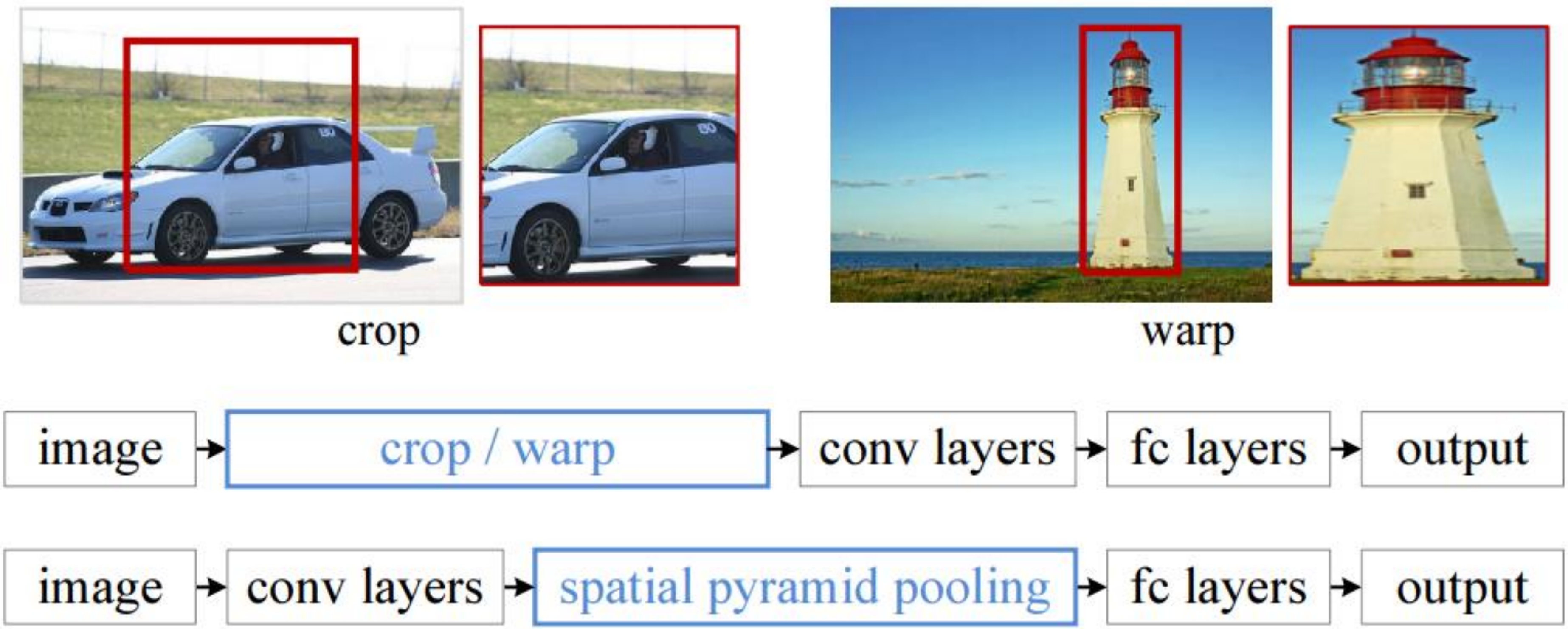
- Warp Resize

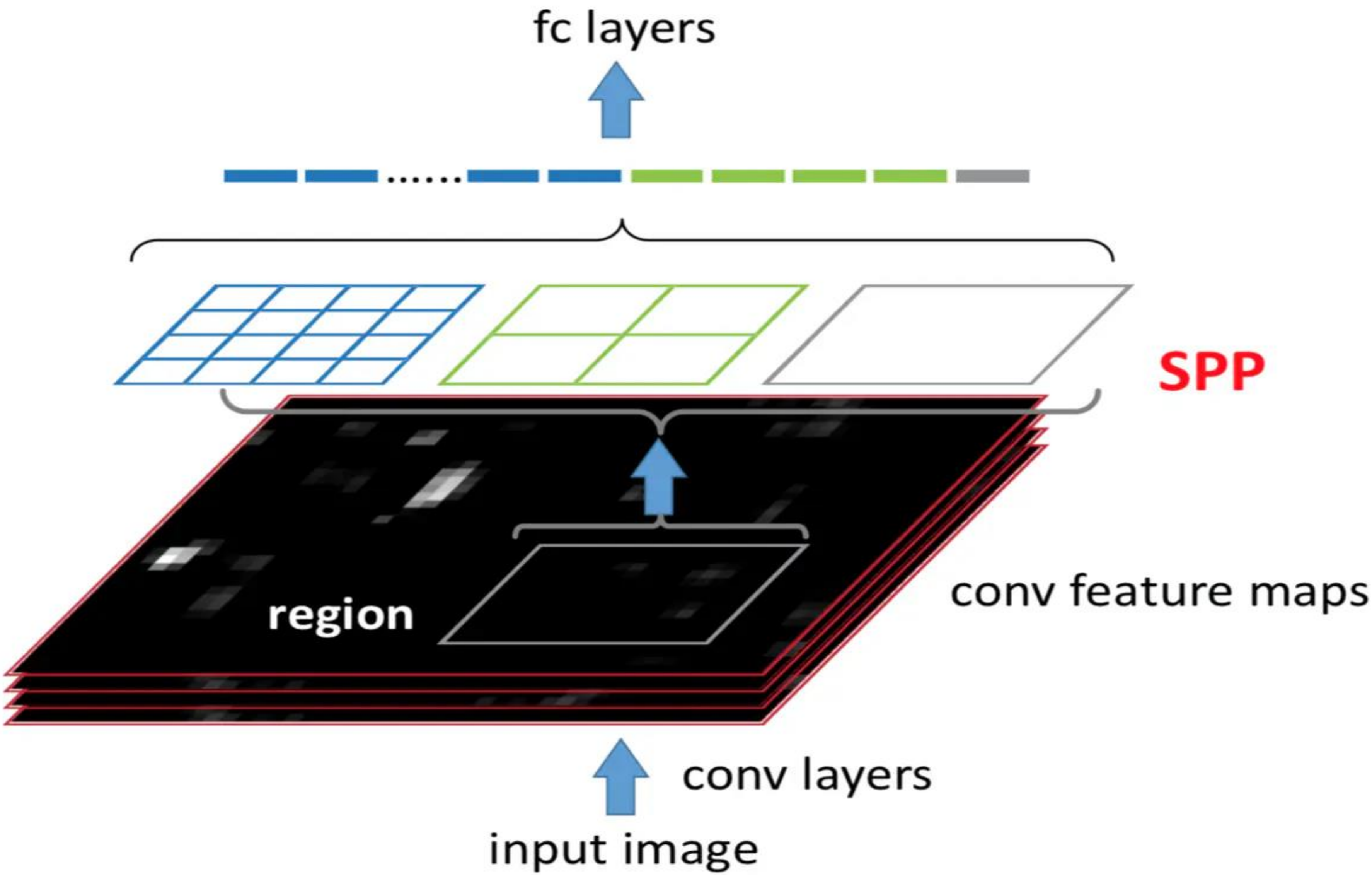
RCNN (REGION CNN)

- 使用回归器精细修正候选框位置

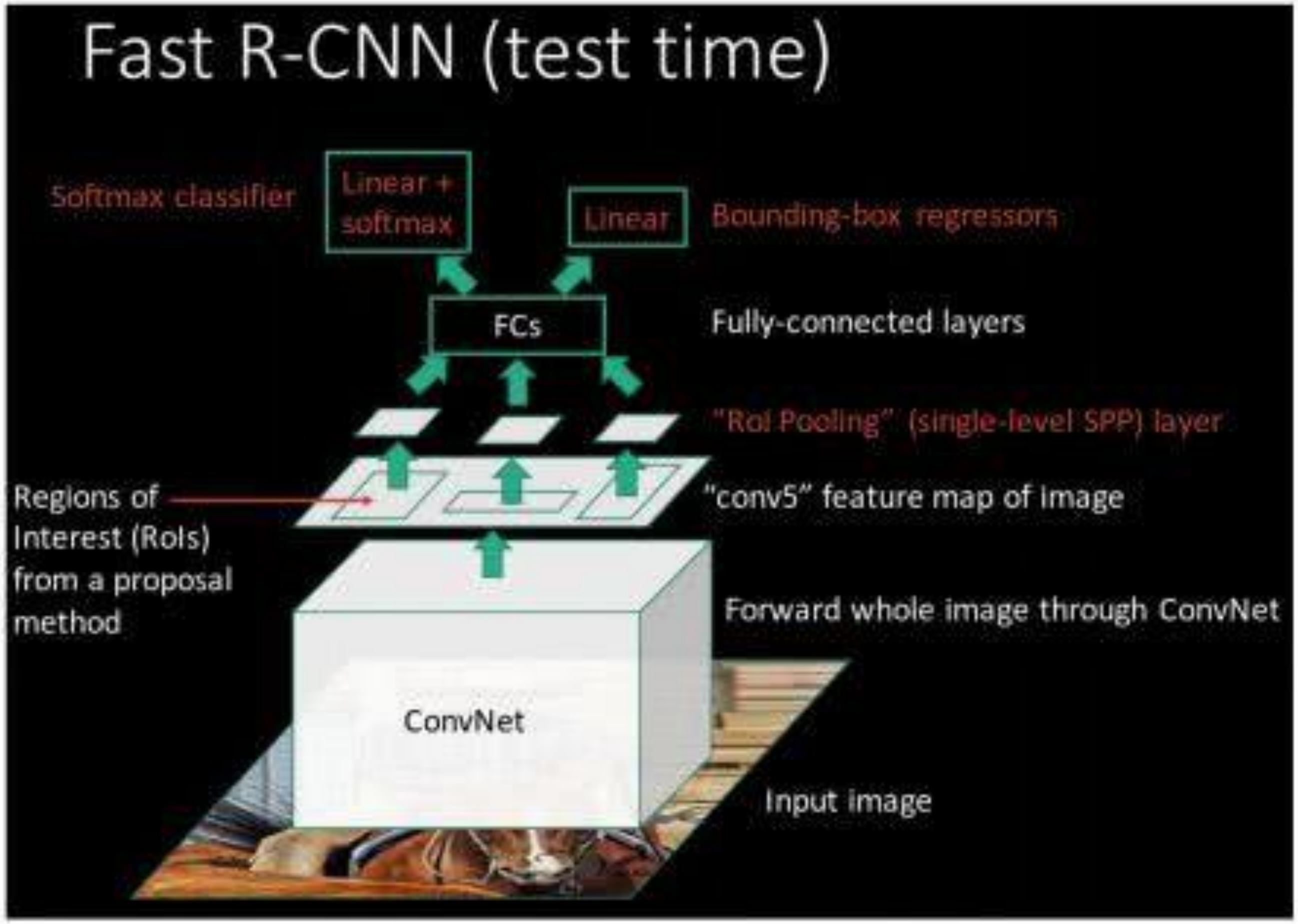


SPATIAL PYRAMID POOLING NET (SPP NET): 空间金字塔池化




fc layers
SPP
region
conv feature maps
conv layers
input image

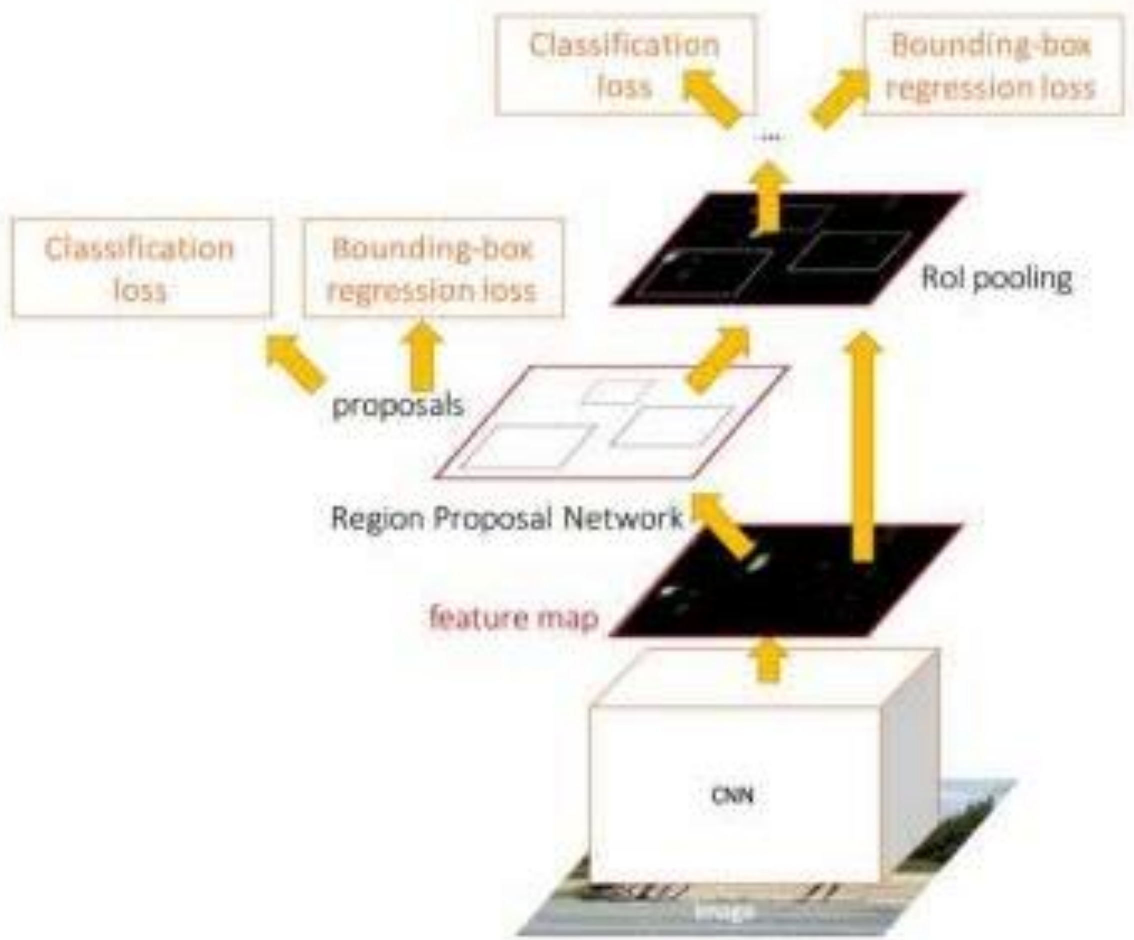
FAST R-CNN



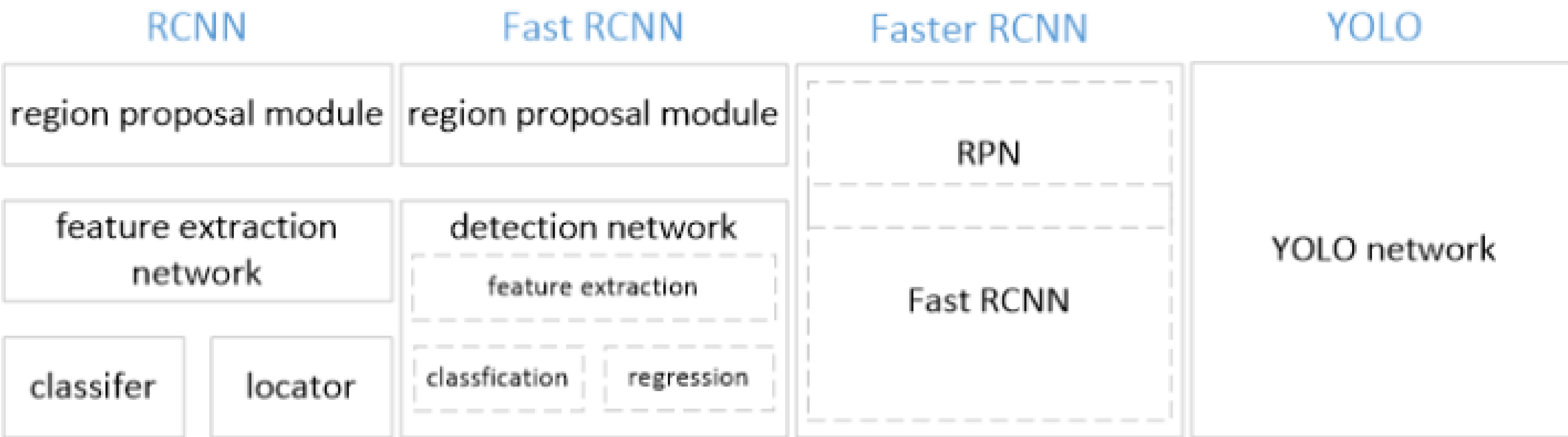
R-CNN Problem #1:
Slow at test-time due to independent forward passes of the CNN

Solution:
Share computation of convolutional layers between proposals for an image

Faster RCNN 网络结构

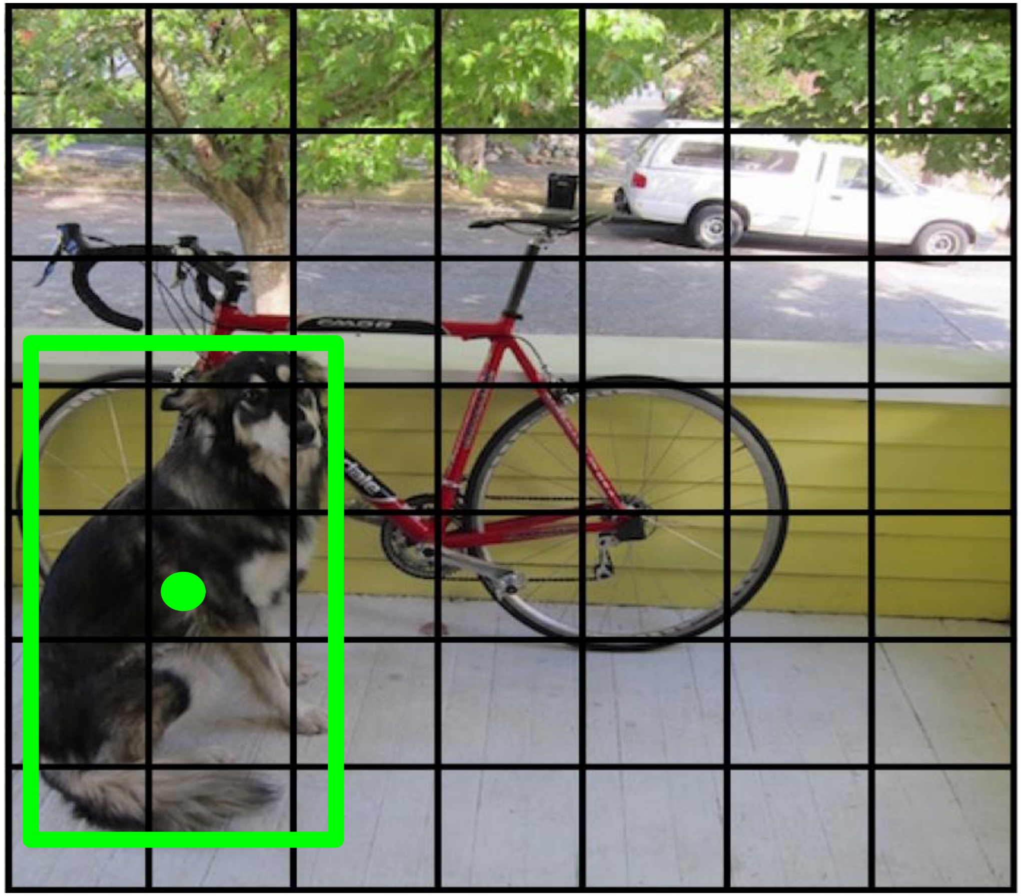


YOLO



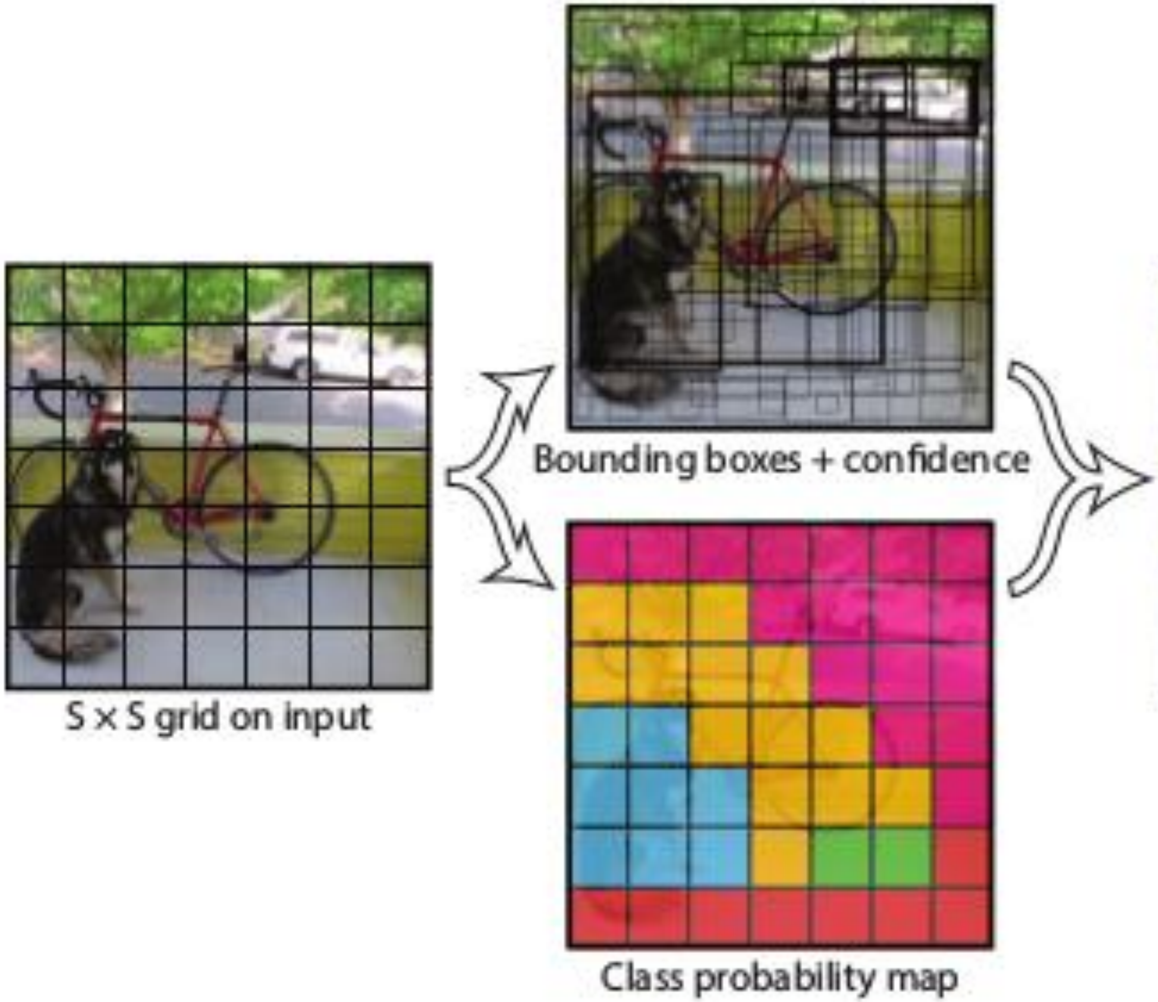
YOLO将物体检测作为一个**回归**问题进行求解

YOLO V1模型



Ground Truth

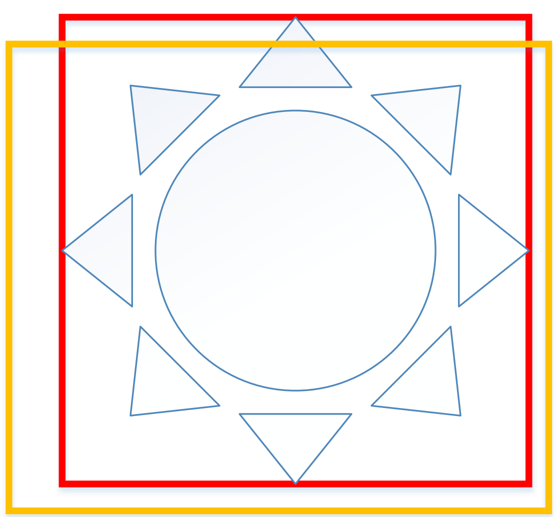
YOLO V1模型


Bounding boxes + confidence
S × S grid on input
Final detections
Class probability map

每个网格单元预测B个边界框bbox和这些框的**置信度得分**。得分反映了bounding box中有object的确定度。得分用 Pr * IOU 表示。若无物体，则得分为零（Pr 为零）。若有物体被识别，则得分为IOU。

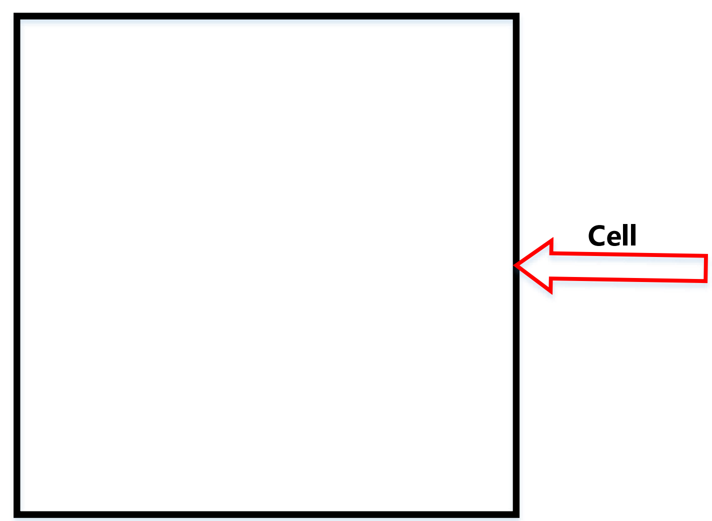
$$\Pr(\text{Object}) * \text{IOU}^{\text{truth}}_{\text{pred}}$$

Pr = 1



Pr * IOU = 1 * 0.8 = 0.8

Pr = 0



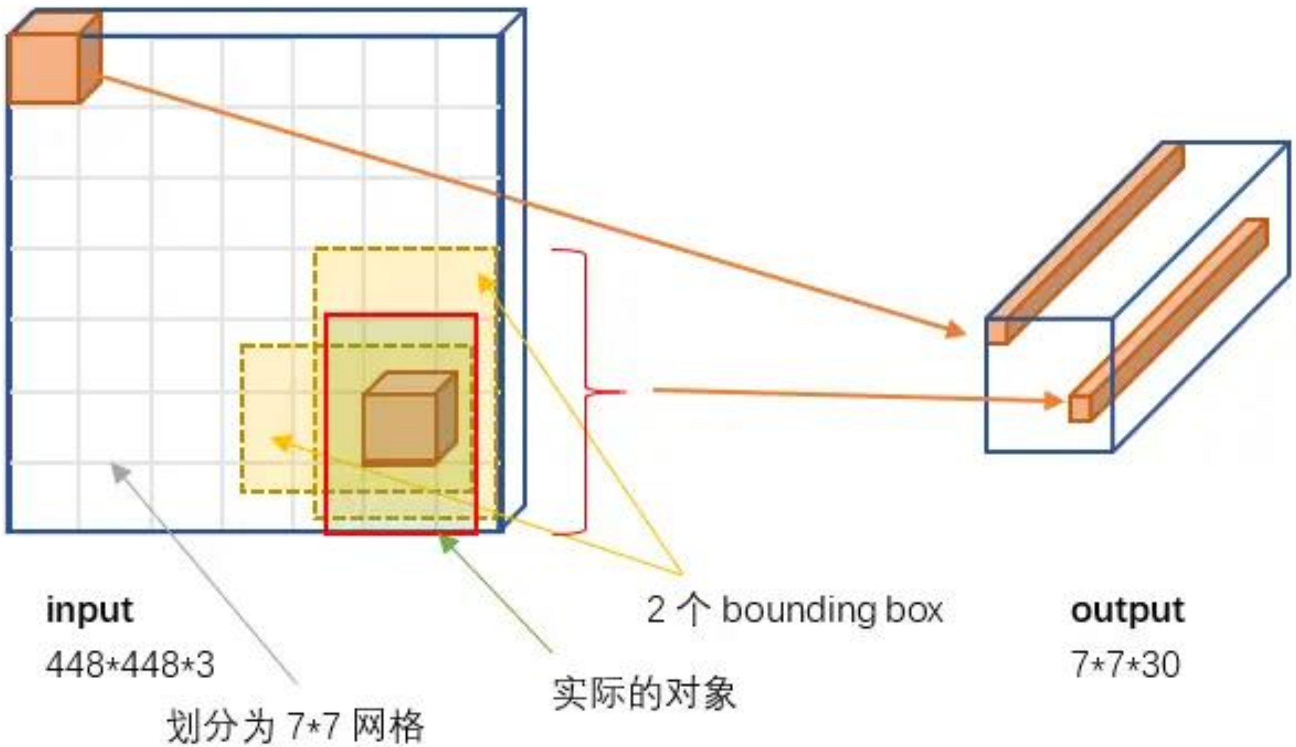
Pr * IOU = 0

YOLO 模型参数

SxS是grid cell网格的数量，B为每个cell中bounding box的数量，C为类别数；

YoloV1直接输出最终得到的结果，结果维度为SxSx(Bx5+C)tensor；

论文中C=20，B=2，S=7



YOLO LOSS FUNCTION

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

坐标误差

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

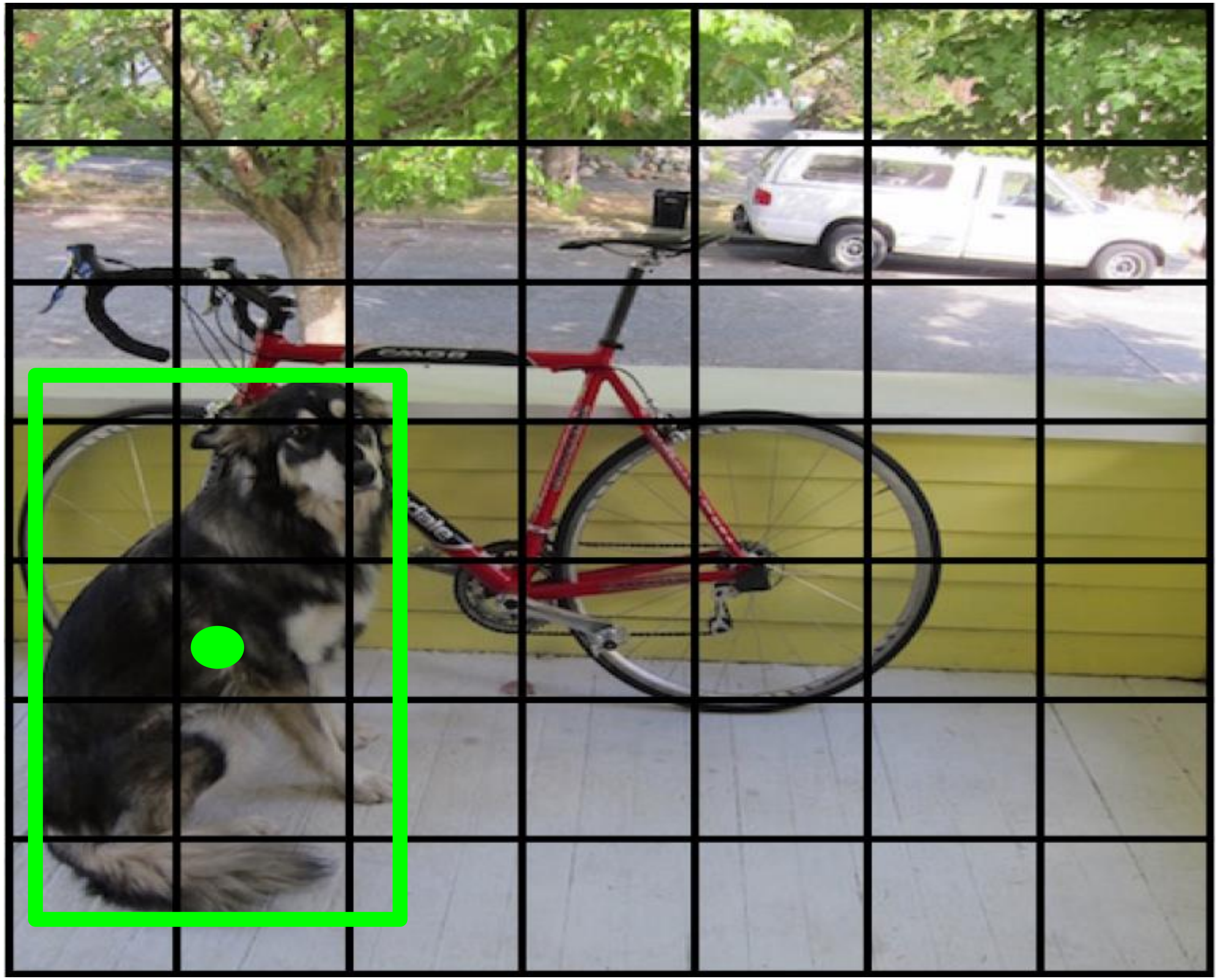
IOU误差

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

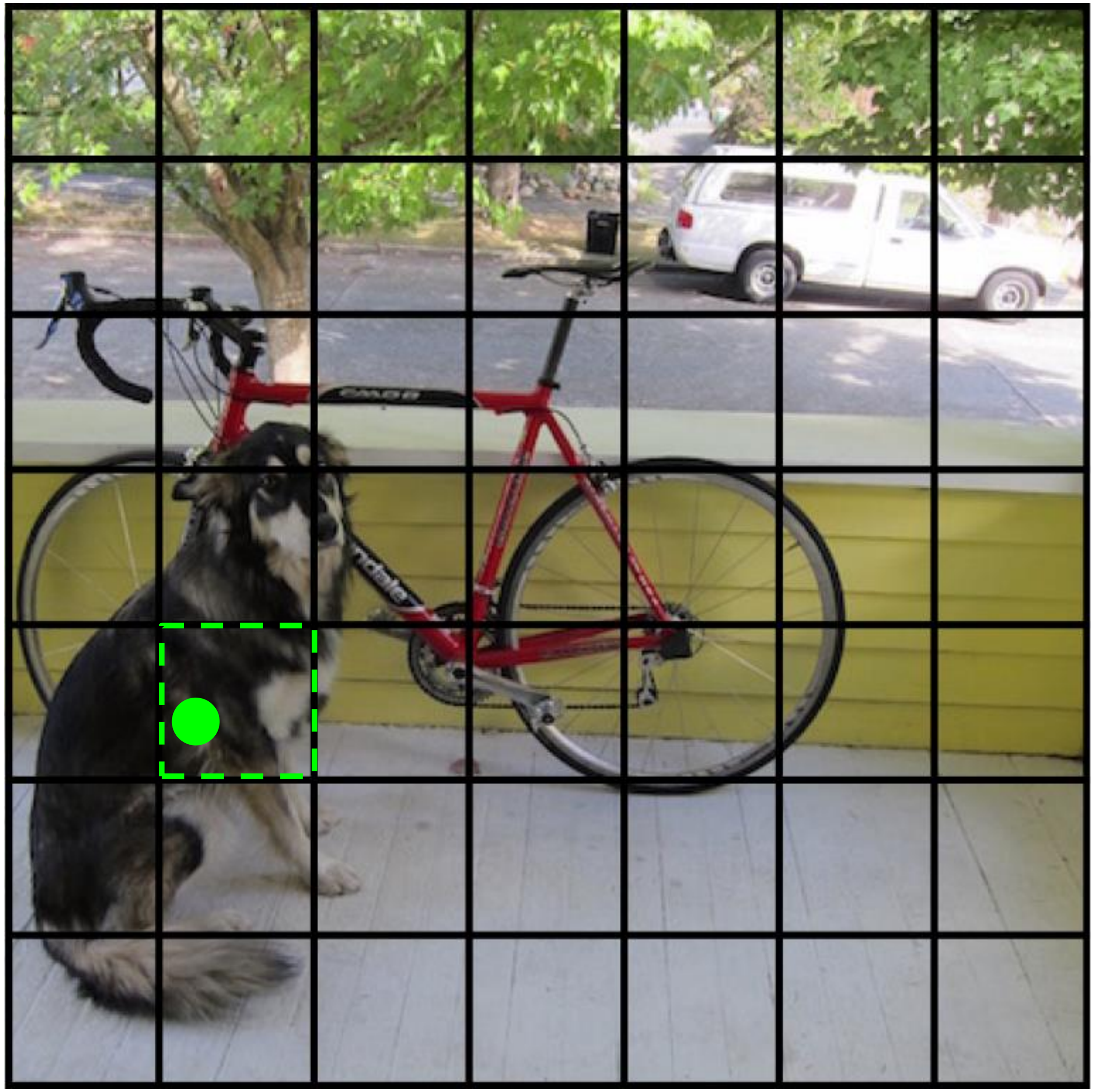
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

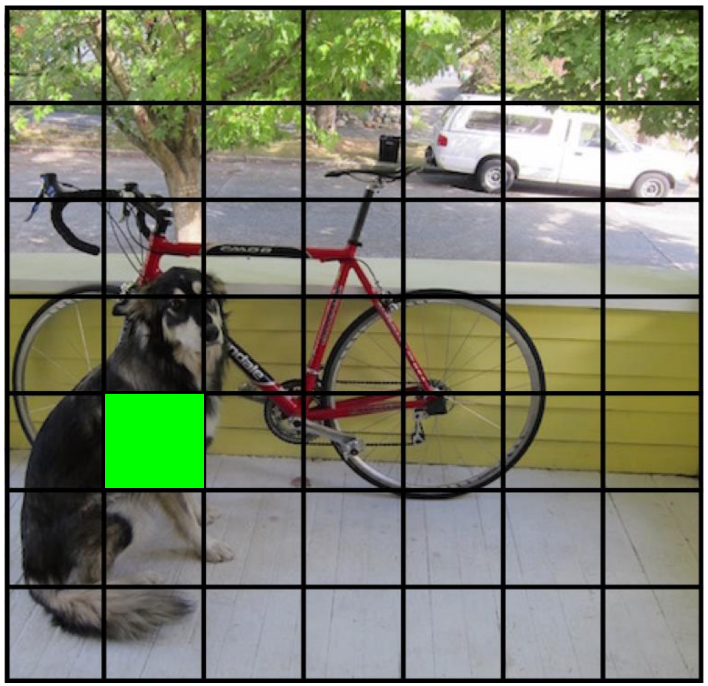
DURING TRAINING, MATCH EXAMPLE TO THE RIGHT CELL



DURING TRAINING, MATCH EXAMPLE TO THE RIGHT CELL



ADJUST THAT CELL’S CLASS PREDICTION



$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

坐标误差

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

IOU误差

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

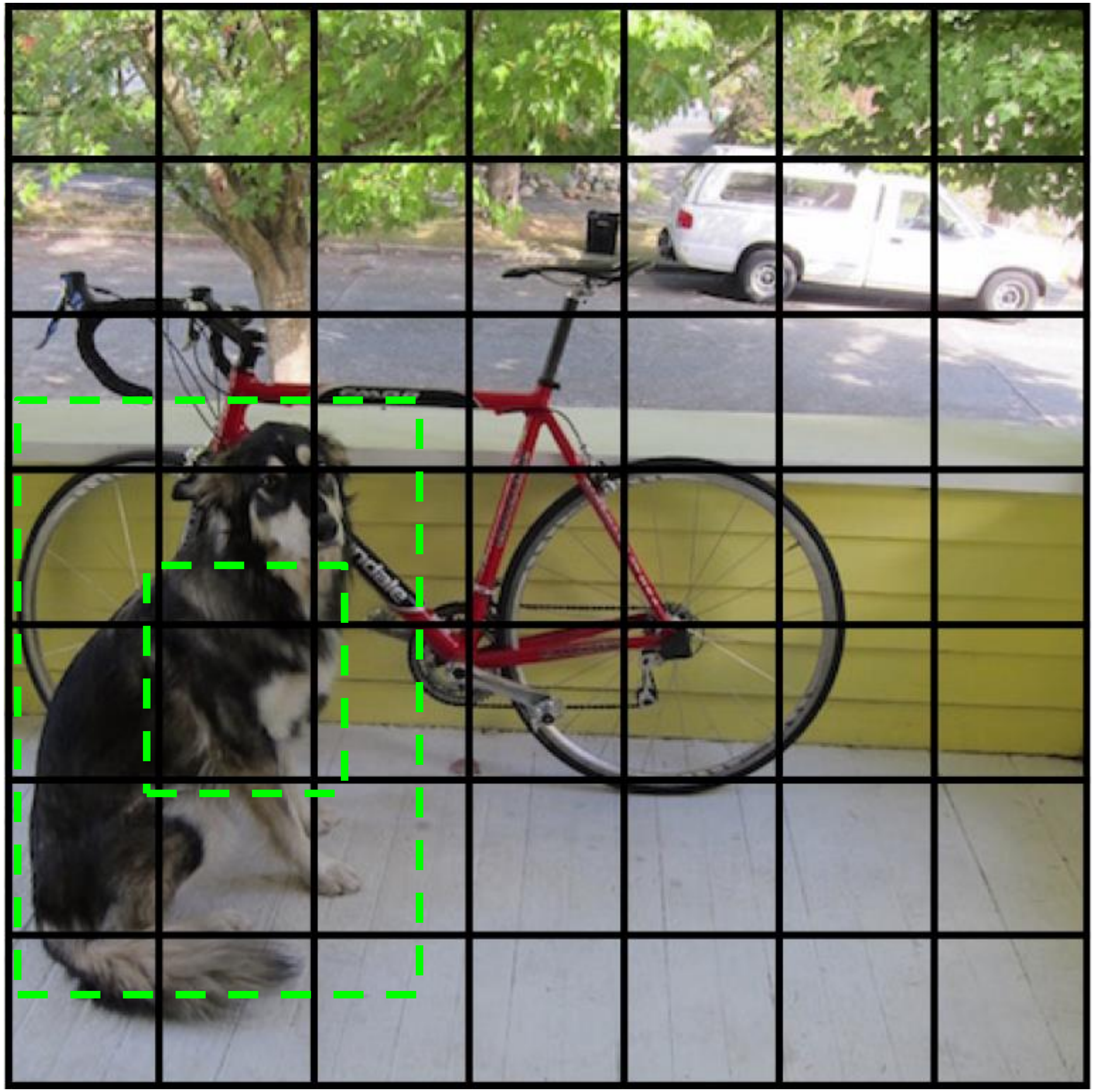
Dog = 1

Cat = 0

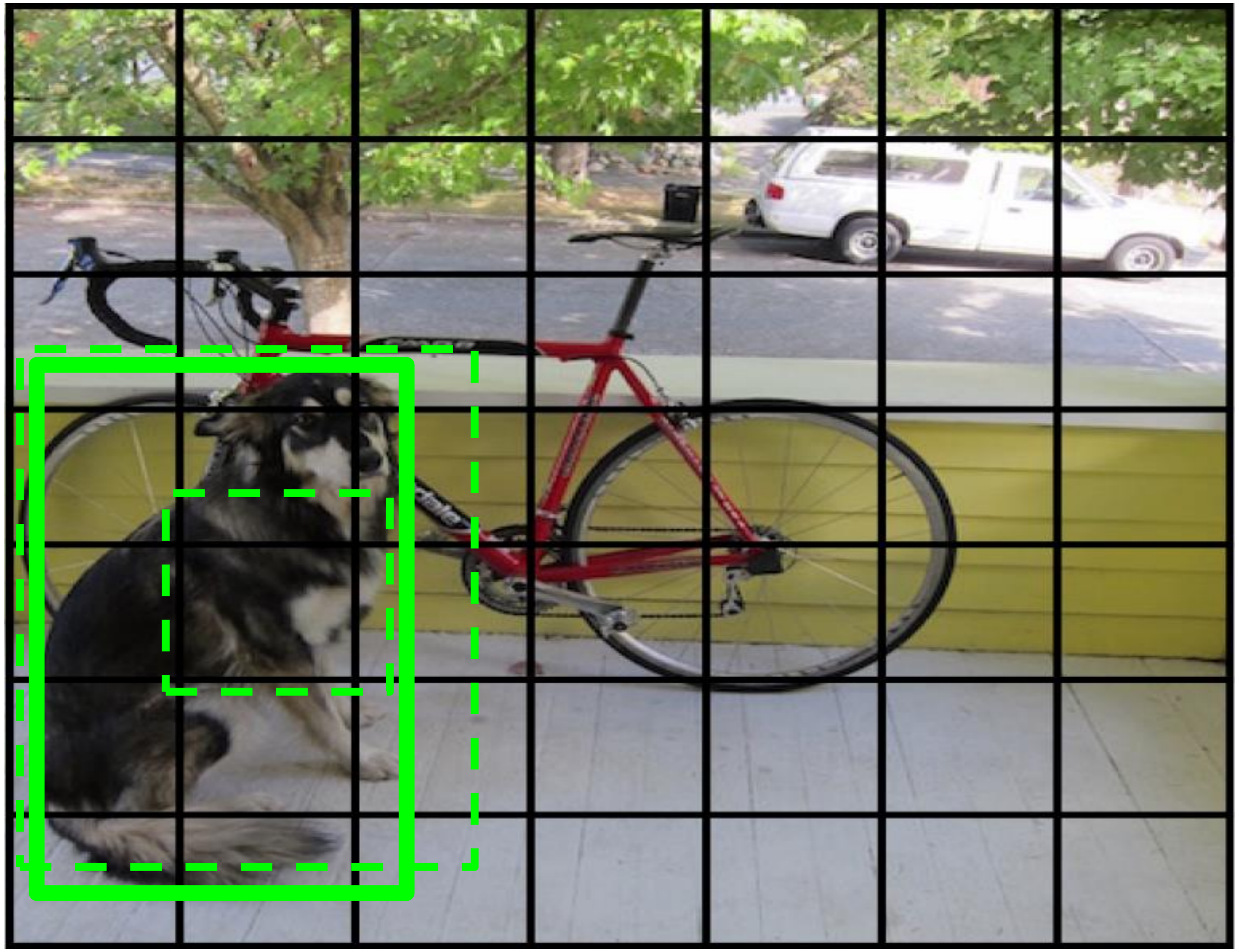
Bike = 0

...

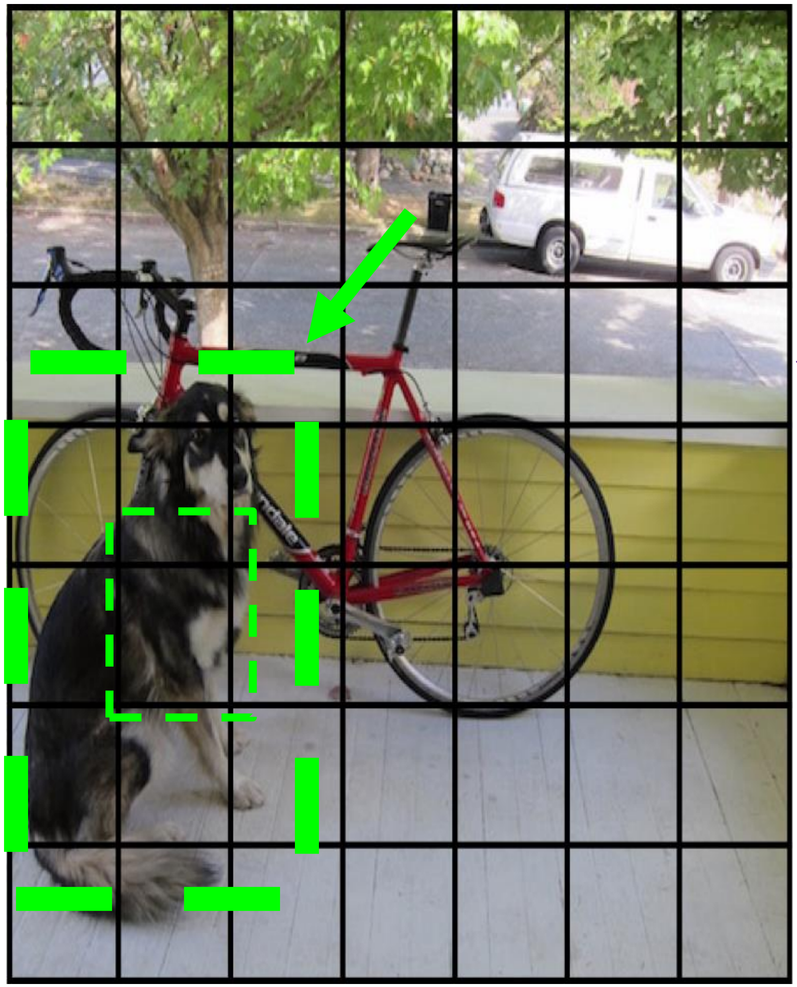
LOOK AT THAT CELL' S PREDICTED BOXES



FIND THE BEST ONE, ADJUST IT



FIND THE BEST ONE, INCREASE THE CONFIDENCE, ADJUST THE BBOX



$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

坐标误差

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

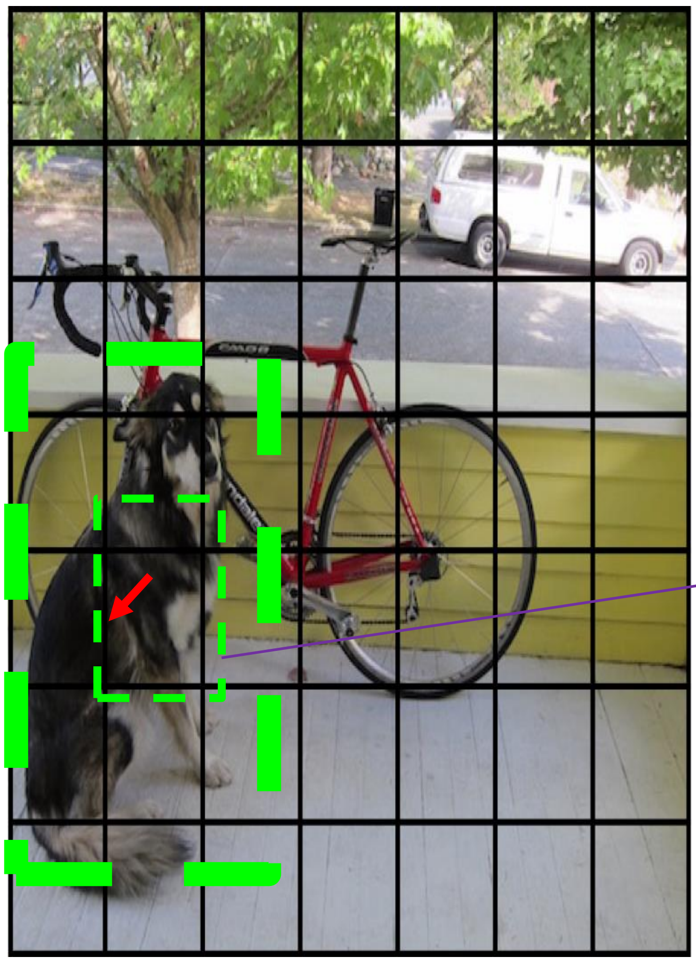
IOU误差

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

DECREASE THE CONFIDENCE OF THE OTHER BOX



$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

坐标误差

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

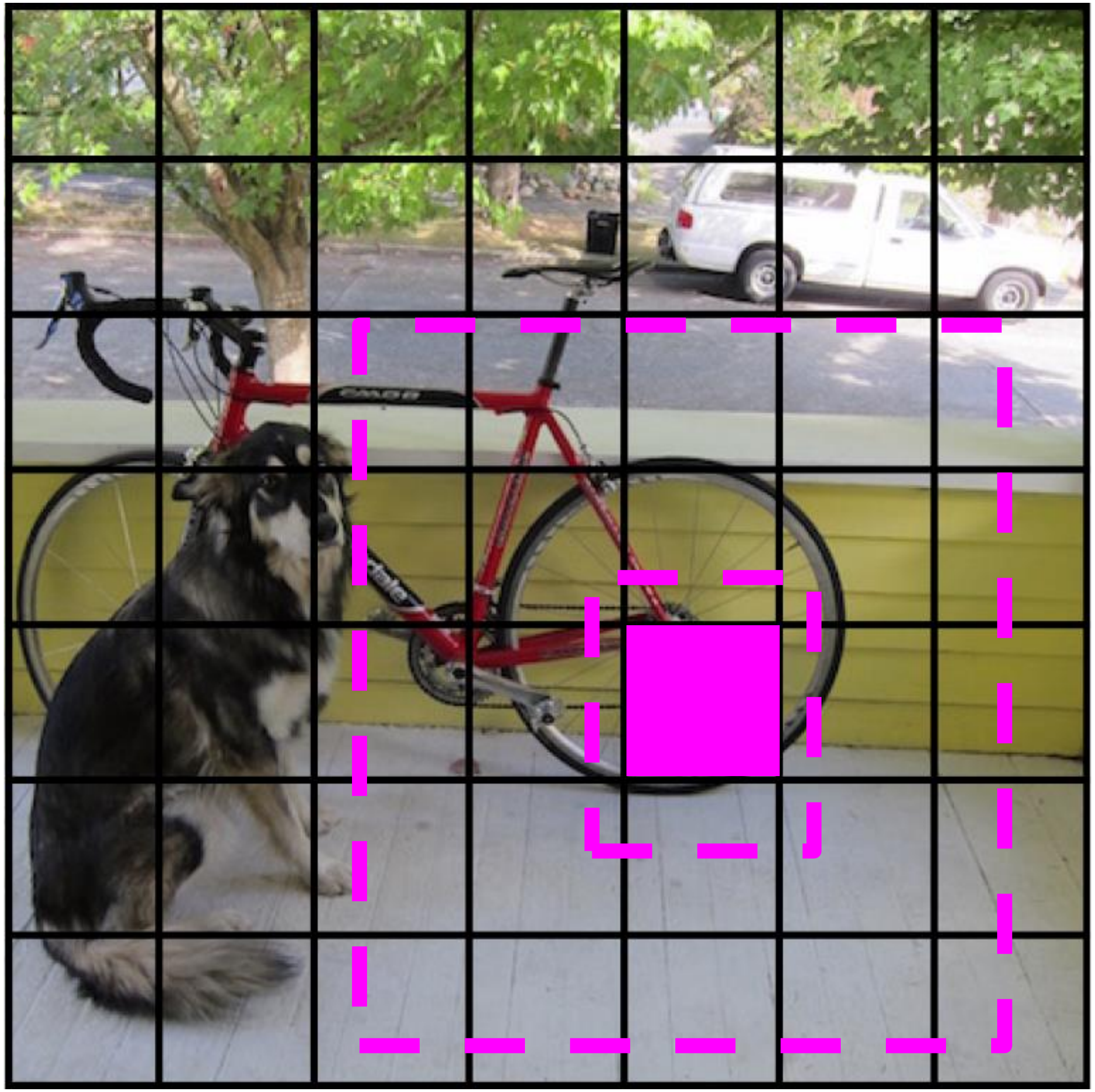
$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

IOU误差

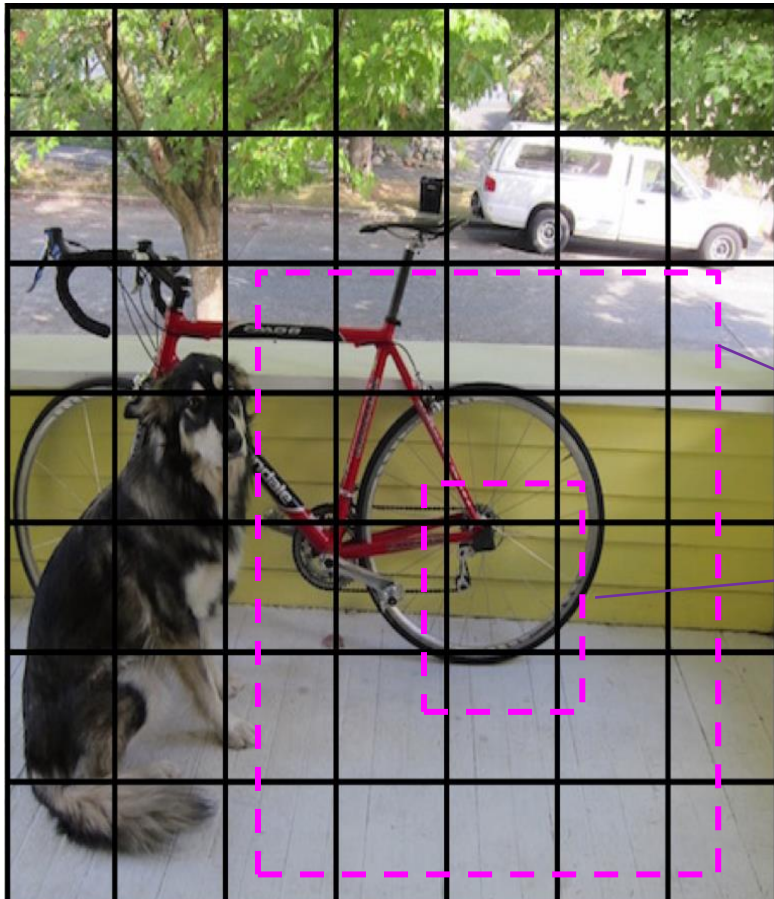
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

SOME CELLS DON’ T HAVE ANY GROUND TRUTH DETECTIONS



DECREASE THE CONFIDENCE OF THESE BOXES



$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

坐标误差

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

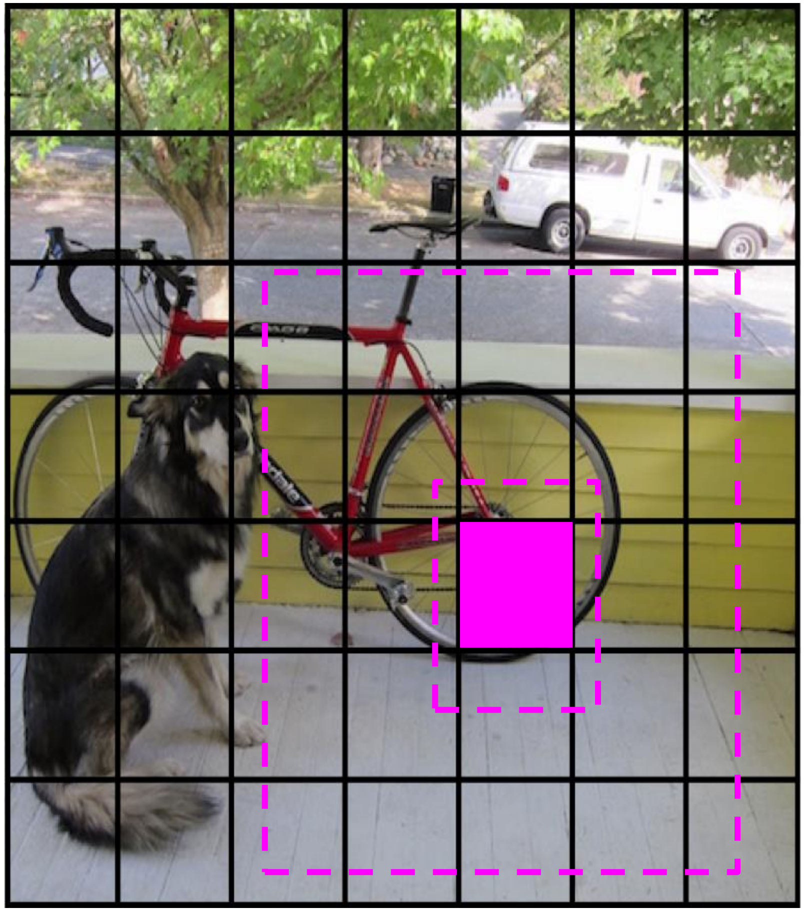
IOU误差

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

DON’ T ADJUST THE CLASS PROBABILITIES OR COORDINATES



$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

坐标误差

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

IOU误差

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

LOSS FUNCTION

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

坐标误差

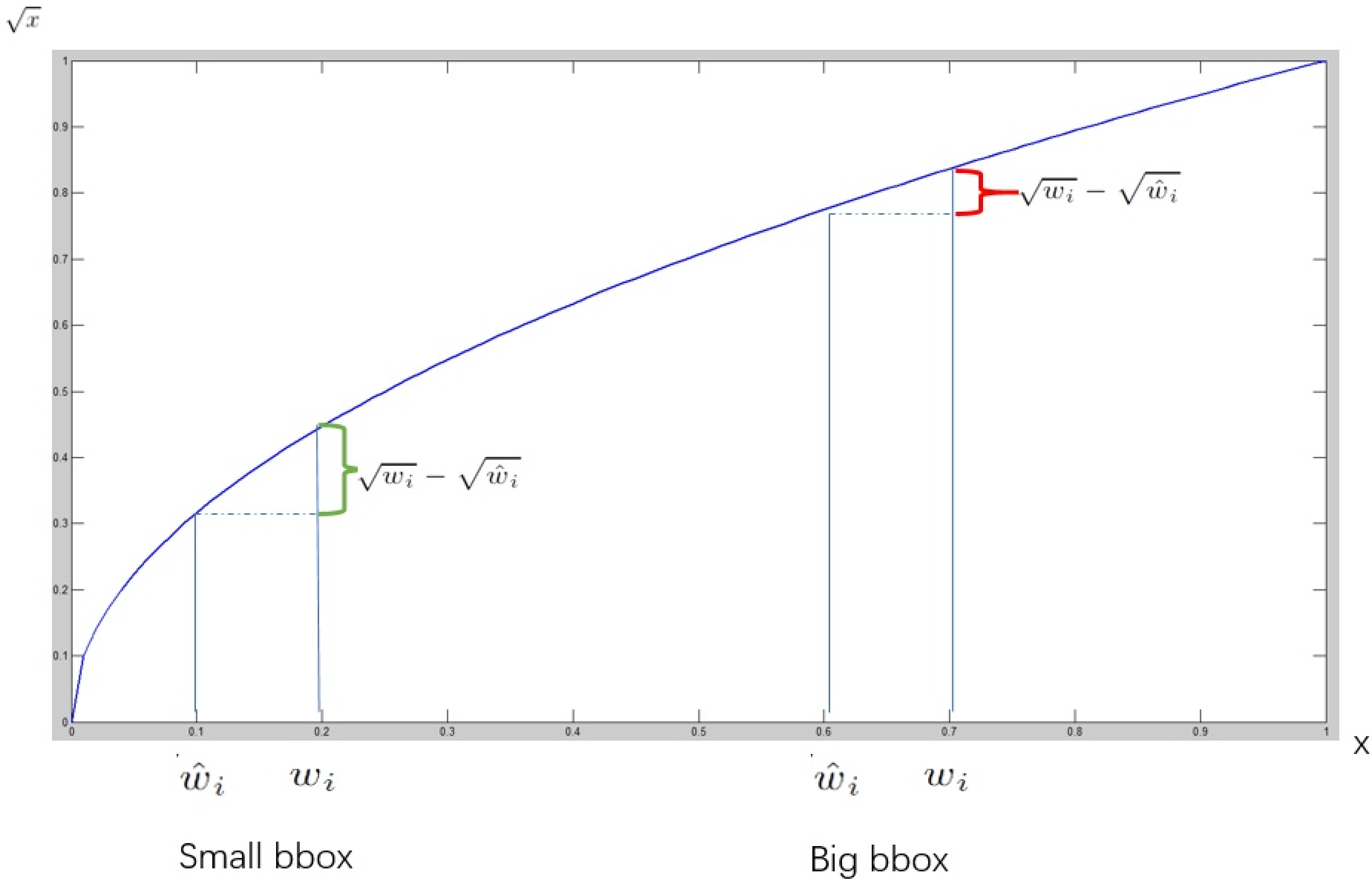
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

IOU误差

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

分类误差

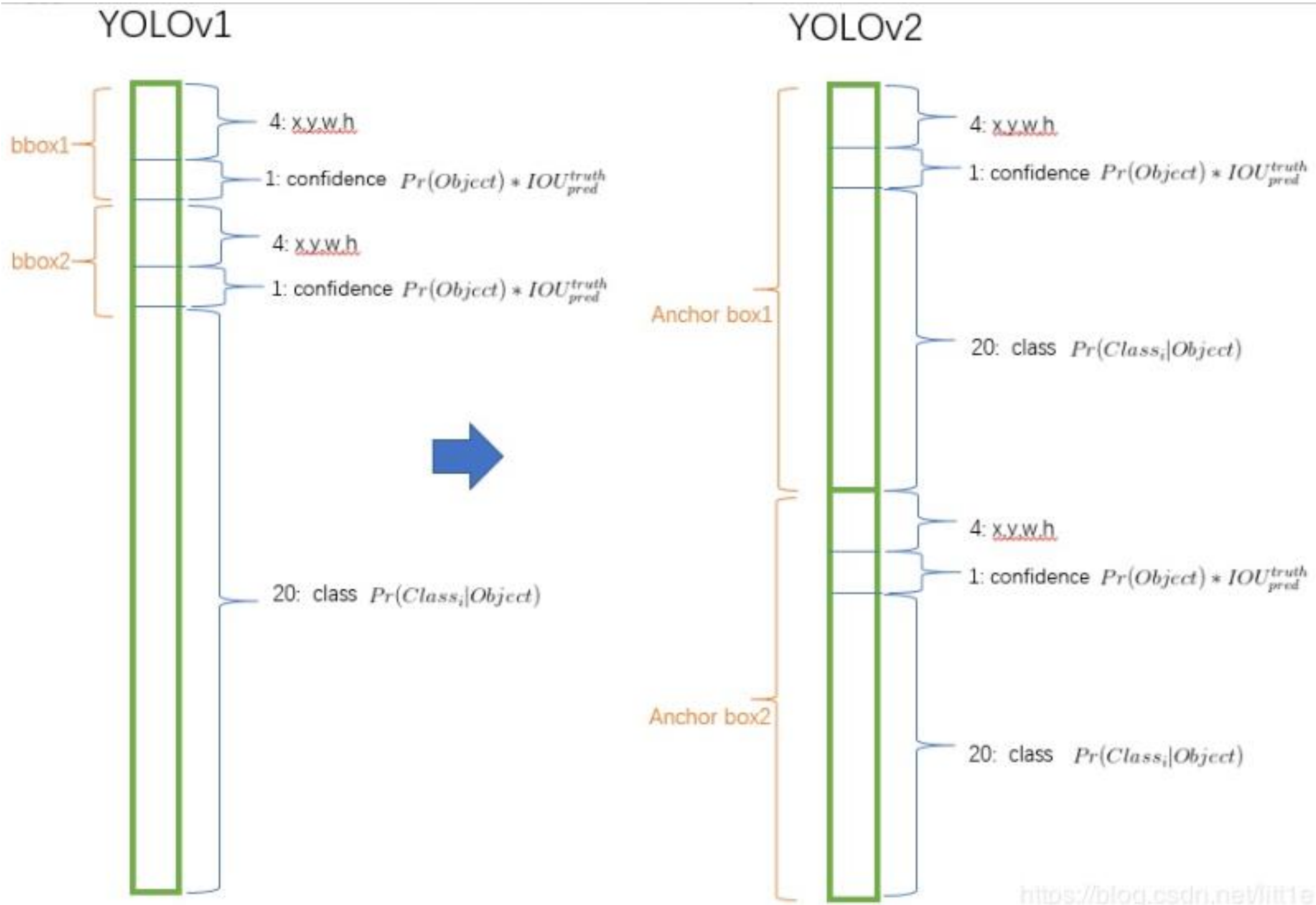


$\sqrt{x}$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
$\sqrt{w_i} - \sqrt{\hat{w}_i}$
$\sqrt{w_i} - \sqrt{\hat{w}_i}$
X
$\hat{w}_i$
w_i
$\hat{w}_i$
w_i
Small bbox
Big bbox

YOLO V1 局限

- 漏检，YOLO对相互靠近的物体（挨在一起且中点都落在同一个格子上的情况），还有很小的物体检测效果不好，这是因为一个网格中只预测了两个框，仅一类。
- 由于损失函数的问题，定位误差是影响检测效果的主要原因。
- 测试图像中，当同一类物体出现的不常见的长宽比和其他情况时泛化能力偏弱。
- 输入尺寸固定。

CONVOLUTION WITH ANCHOR BOXES



作业：

手写数字识别