中国科学院大学
University of Chinese Academy of Sciences

深度学习应用（计算机视觉）

AI DISCOVERY

目录

目录

AI DISCOVERY

1 目标检测 两阶段方法、一阶段方法、最新进展

2 典型图像分析任务 图像分割、图像搜索、目标跟踪

3 特色图像分析任务 细粒度分类、风格迁移、标题生成、超分辨率

4 垂直应用与实践 医学影像分析、文字检测识别
实践：目标检测

AI DISCOVERY

目标检测

AI DISCOVERY

目标检测任务

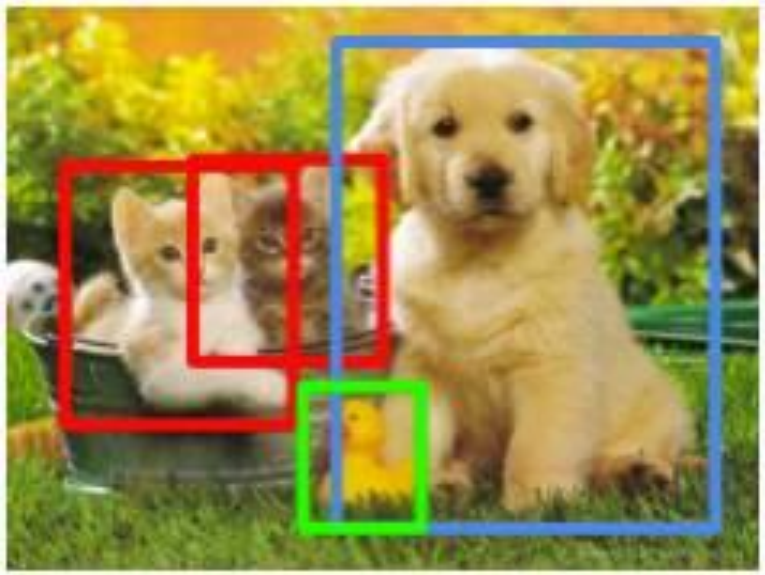
分类

Cat

定位

Cat

目标检测



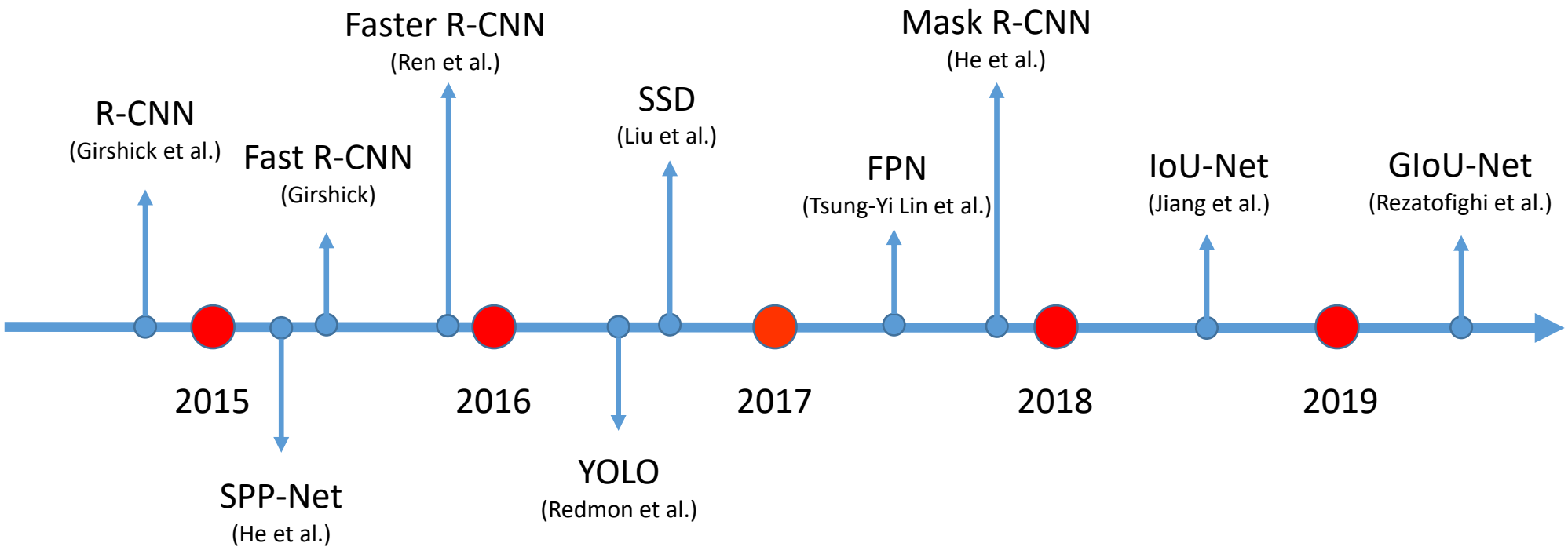
Cat dog duck

AI DISCOVERY

目标检测

AI DISCOVERY

目标检测发展历程



AI DISCOVERY

目标检测

AI DISCOVERY

两阶段方法	R-CNN、SPP-Net、Fast R-CNN、Faster-RCNN、FPN、Mask-RCNN
一阶段方法	YOLO、SSD
最新进展	IoU-Net、GIoU

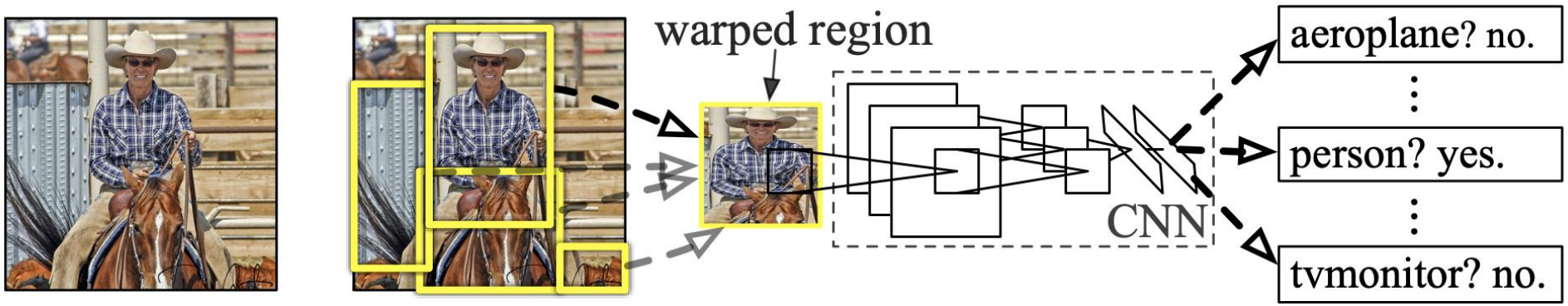
AI DISCOVERY

两阶段方法

AI DISCOVERY

1、R-CNN

R-CNN: *Regions with CNN features*



1. Input image

2. Extract region proposals (~2k)

3. Compute CNN features

4. Classify regions

输入图像

提取候选检测框（约2000个）

为每个候选检测框提取CNN特征

为每个候选检测框进行分类

AI DISCOVERY

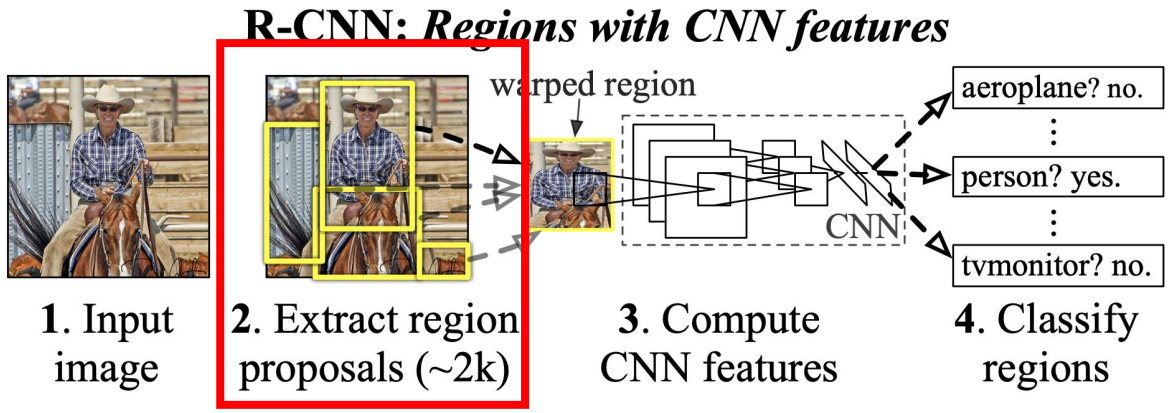
Ross B. Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” CVPR 2014

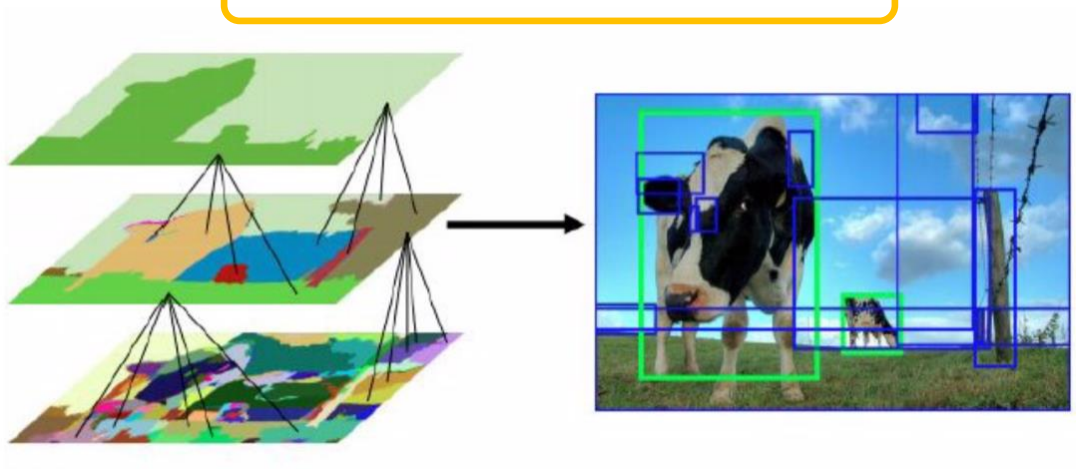
两阶段方法

AI DISCOVERY

1、R-CNN：候选区域生成

类似于一种层次聚类算法



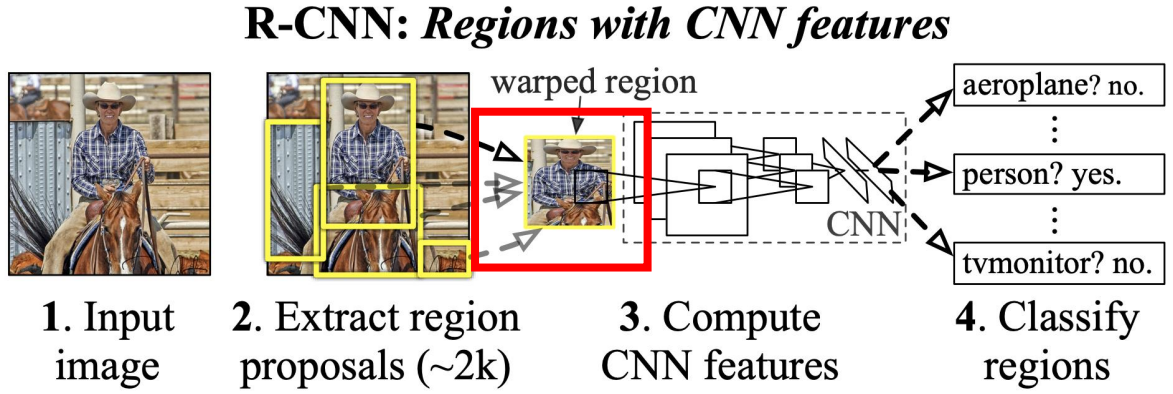


- Selective Search(SS，选择性搜索)：根据颜色、纹理、尺寸和空间交叠相似度提取约2000个region proposal(候选区域)。
- 存在问题：
  - ✓ 对于每张图像，还需要额外的步骤提取region proposal
  - ✓ 存储和重复的提取每个region proposal的特征花费大量的存储和计算资源

两阶段方法

AI DISCOVERY

1、R-CNN：统一尺寸



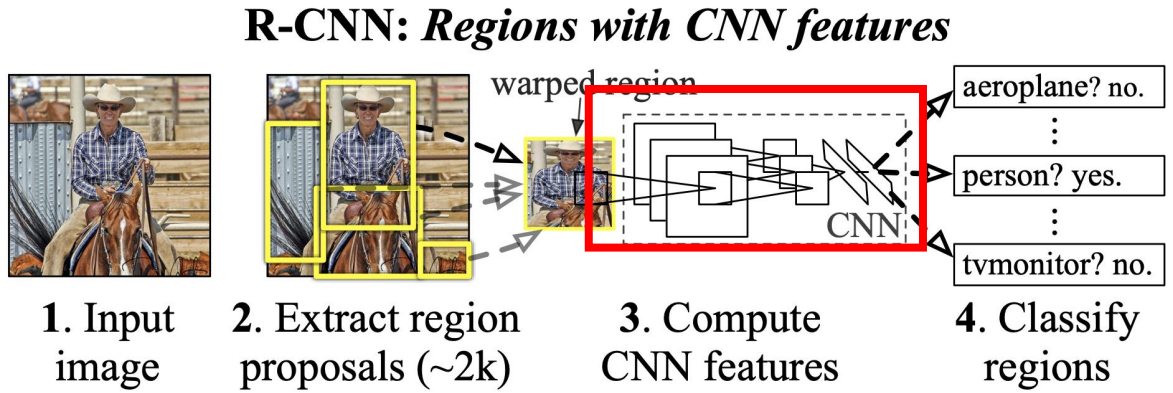
- Warped region (区域拉伸)：通过Selective Search产生的候选区域大小不一样，为了与CNN（AlexNet）兼容，R-CNN直接将所有的候选区域统一到227*227的尺寸。
- 存在问题：
  - ✓ 将每个region proposal统一成同样的尺寸，严重影响CNN提取特征的质量

AI DISCOVERY

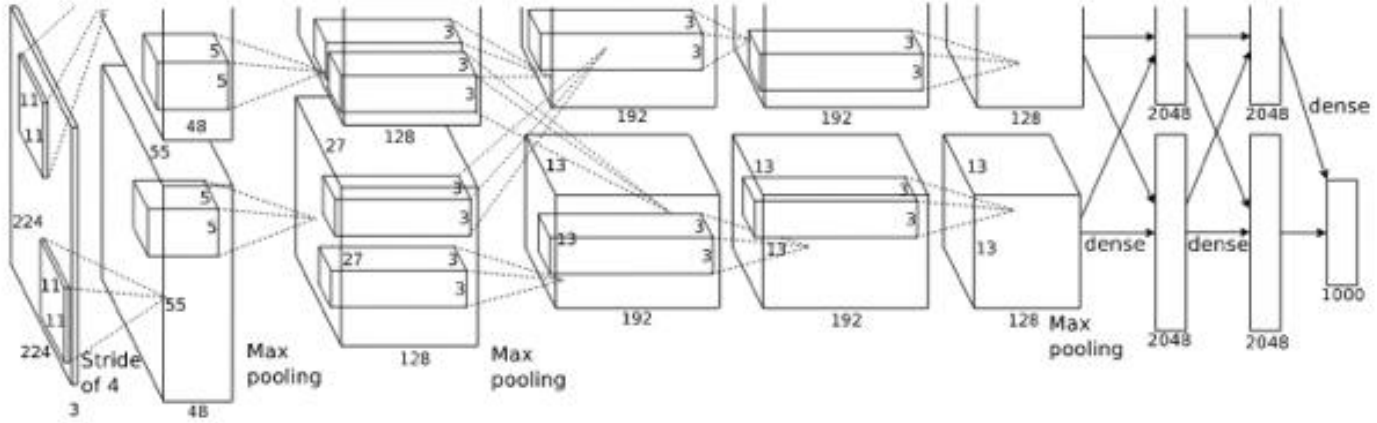
两阶段方法

AI DISCOVERY

1、R-CNN：特征提取



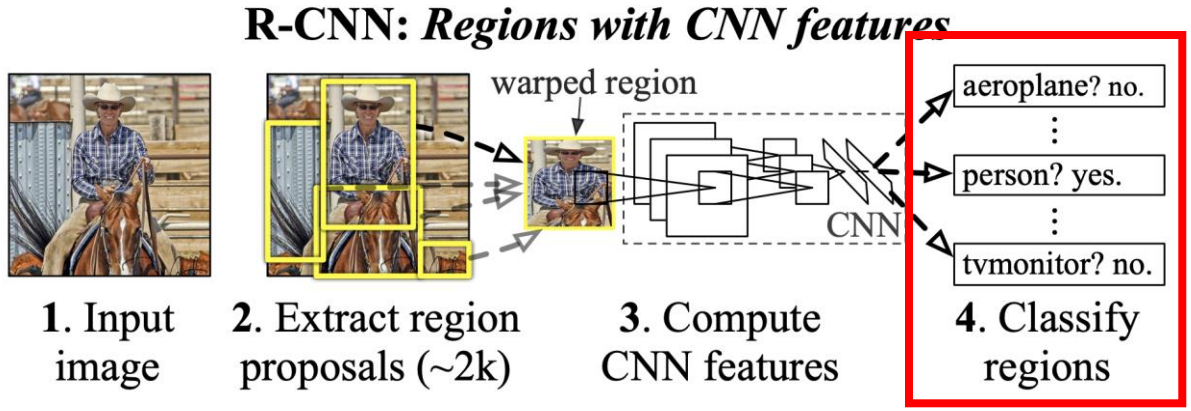
- 特征提取：
  - ✓ 通过卷积神经网络提取CNN特征，用于分类。
- 存在问题：
  - ✓ 保存所有的目标候选区域的特征大约占用了200G的空间。



两阶段方法

AI DISCOVERY

1、R-CNN：区域分类



- Classify regions(区域分类): 为每一个类（包括背景类）训练SVM。
- 存在问题：
  - ✓ SVM需要单独的训练，使网络训练上更加复杂，随着类别的增加训练SVM的个数也随之增加。此外，逐个执行SVM分类也将消耗较多的时间。

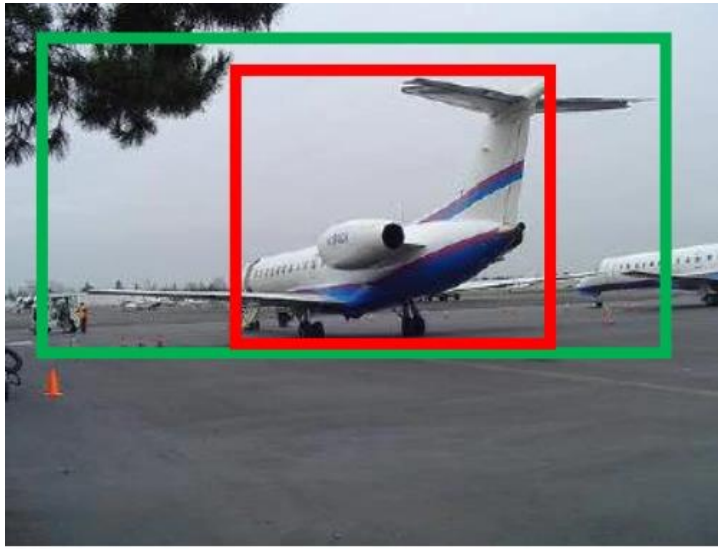
AI DISCOVERY

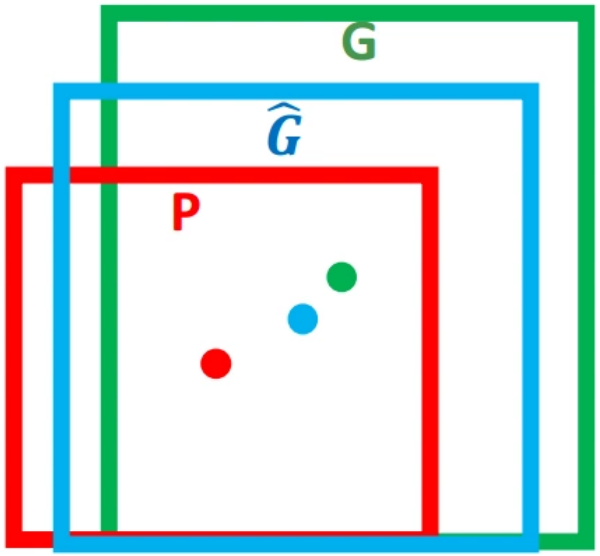
两阶段方法

AI DISCOVERY

1、R-CNN：边界框回归

通过一种学习一种映射关系，对目标候选的位置进行精化（Refine）：





- G：真实坐标
- P：RoI坐标（Selective Search获得）
- $\hat{G}$：修正后坐标

学习一种映射f，可最小化Region Proposal（p_x, p_y, p_w, p_h）和Ground Truth（G_x, G_y, G_w, G_h）的差异

$$f(p_x, p_y, p_w, p_h) = (\widehat{G_x}, \widehat{G_y}, \widehat{G_w}, \widehat{G_h}) \approx (G_x, G_y, G_w, G_h)$$

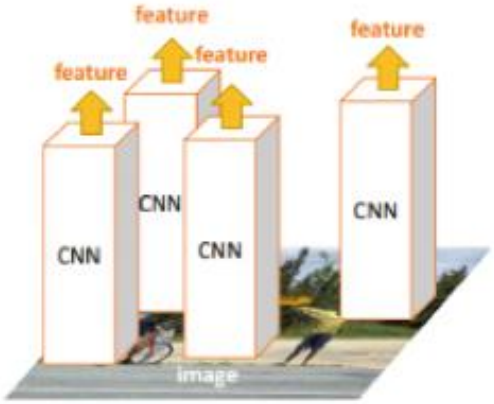
AI DISCOVERY

两阶段方法

AI DISCOVERY

2、SPP-Net

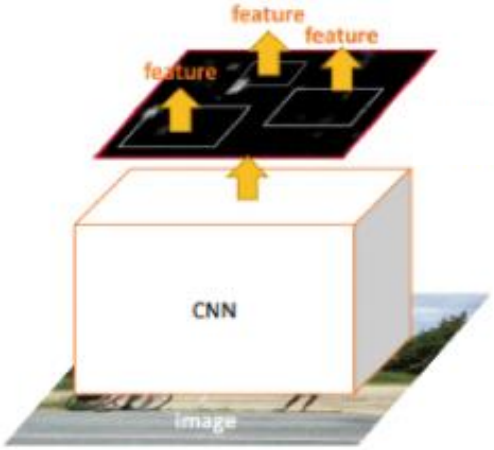
R-CNN缺点：为每个region proposal提取特征，花费大量的计算时间和存储空间。



R-CNN

2000 nets on image regions

逐个提取特征



SPP-NET

1 net on full image

一次性提取特征

SPP-Net不再单独提取每个候选区域的特征，而是一次性提取整个图像的特征，再在特征图上取出对应于不同region proposal的区域

减少了提取特征的时间，和用来存储特征的空间。

AI DISCOVERY

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Spatial pyramid pooling in deep convolutional networks for visual recognition.“, ECCV 2014, PAMI2015

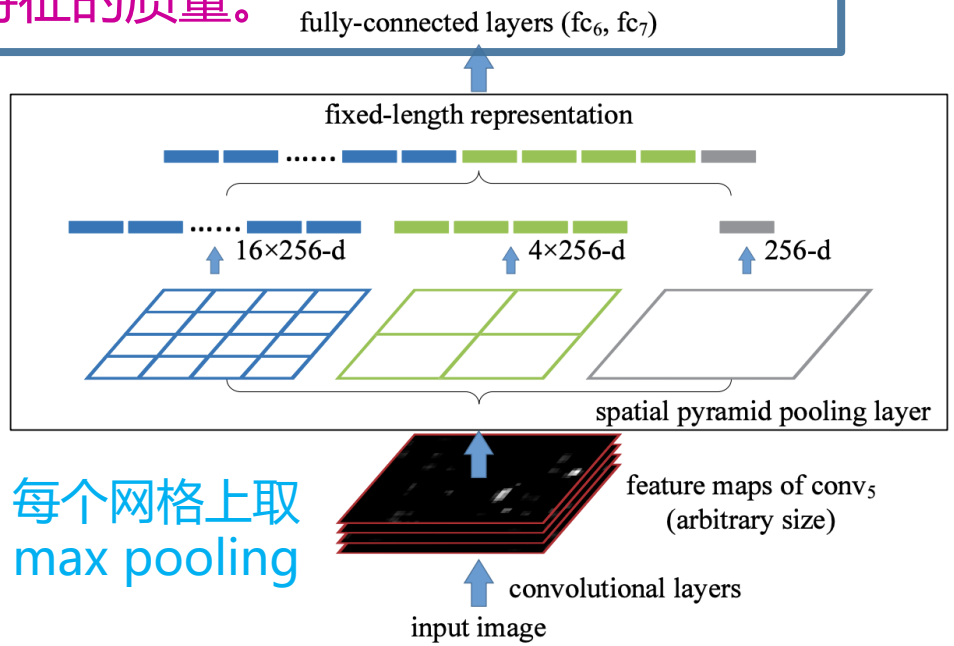
两阶段方法

AI DISCOVERY

2、SPP-Net

R-CNN缺点：使用Warp，为每个region proposal统一尺寸，严重影响CNN提取特征的质量。

丢失信息/形态



每个网格上取max pooling

通过spatial pyramid pooling将任意大小RoI特征统一成相同的尺寸

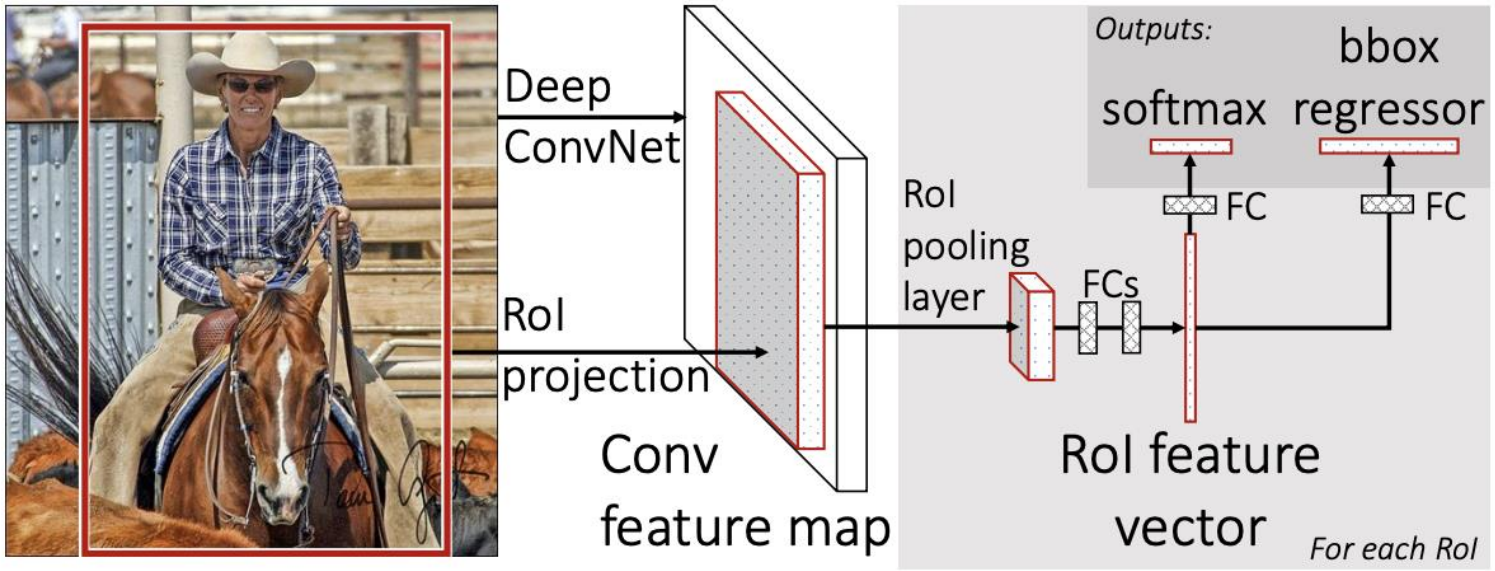
可以输入任意大小的候选区域，不再需要warp输入图像，提升了CNN提取的特征的质量，使特征更鲁棒

AI DISCOVERY

两阶段方法

AI DISCOVERY

3、Fast R-CNN：将分类损失和回归损失统一在同一个框架中



- ✓ 图像的warp->特征的warp
- ✓ 损失：svm+regressor->多任务损失

  （softmax+regressor）

通过SS在图像中提取RoI->卷积网络提取特征->RoI Pooling->全连接层->分类/边界框回归

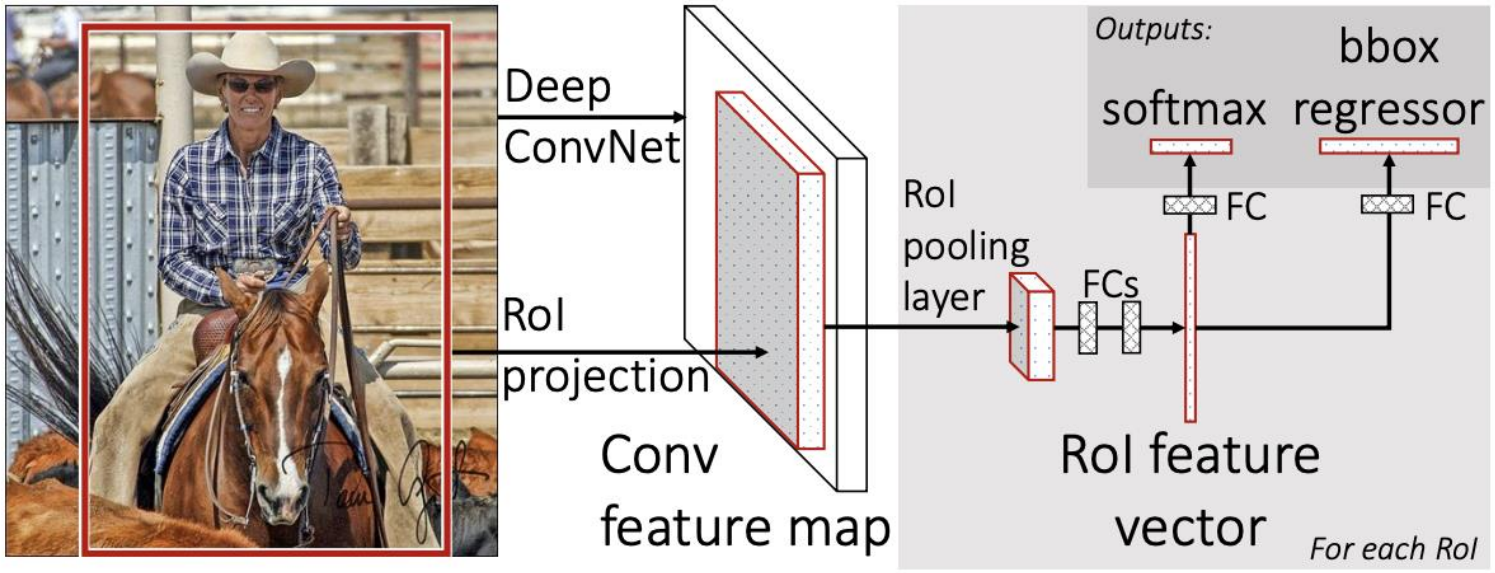
AI DISCOVERY

Ross B. Girshick, “Fast R-CNN,” ICCV 2015

两阶段方法

AI DISCOVERY

3、Fast R-CNN：将分类损失和回归损失统一在同一个框架中



特征比图像抽象，学习过程可更好的适应弯曲和裁剪变化

- ✓ 图像的warp->特征的warp

- ✓ 损失：svm+regressor->多任务损失

  (softmax+regressor)

通过SS在图像中提取RoI->卷积网络提取特征->RoI Pooling->全连接层->分类/边界框回归

AI DISCOVERY

两阶段方法

AI DISCOVERY

3、Fast R-CNN RoI Pooling

图像的warp->
特征的warp: RoI Pooling

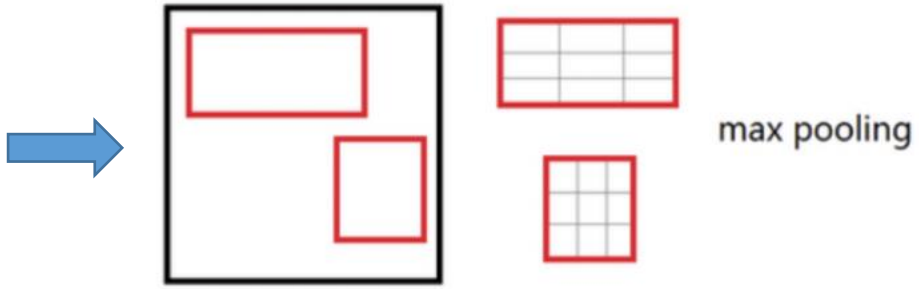
示例：

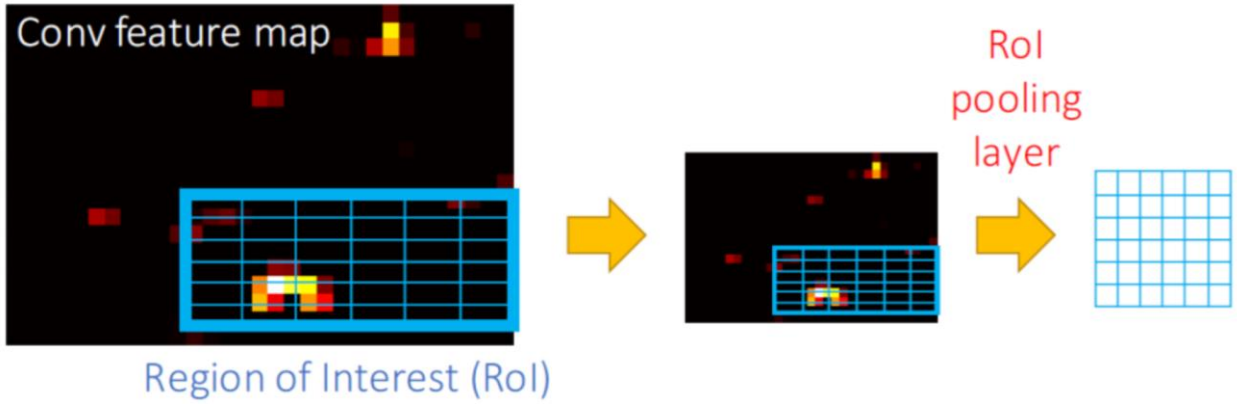
输入ROI特征图的大小：24*12　　max pooling: stride(4*2)

输入ROI特征图的大小：12*18　　max pooling: stride(2*3)

输出特征图：6*6

注：不能整除，则向下取整（丢弃小部分右侧和下侧像素）





◆与SPP-Net的异同：

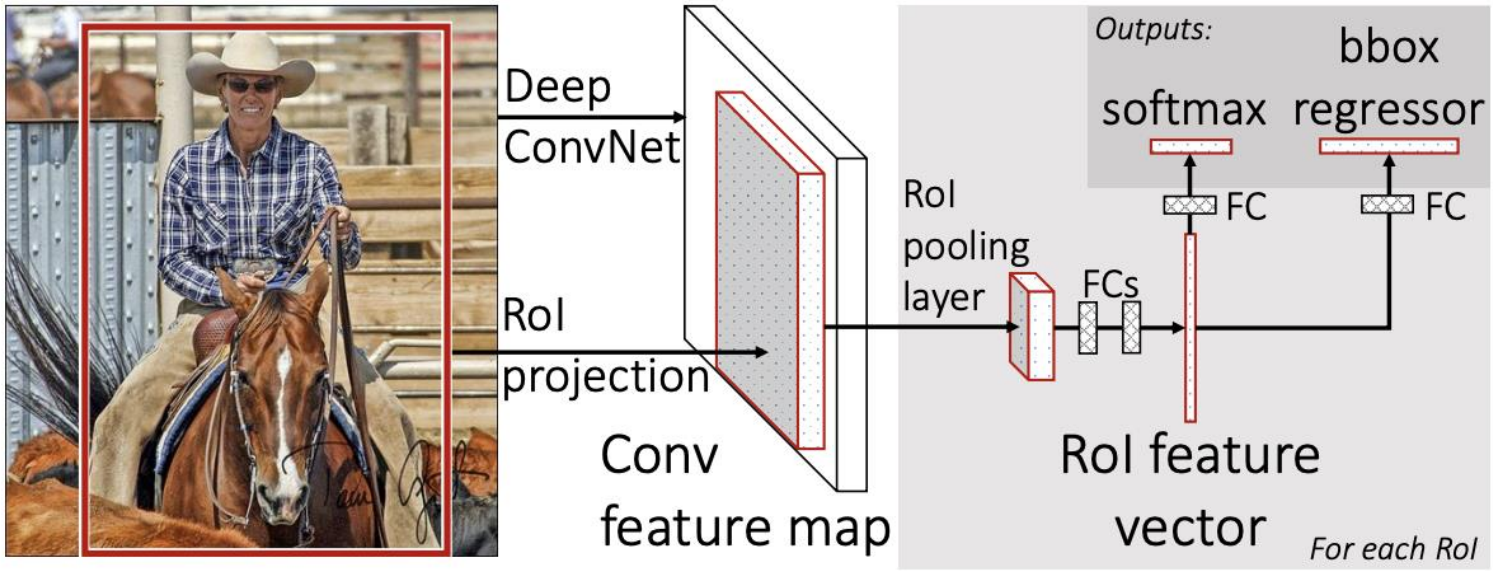
- ✓ 都是特征而不是图像的warp
- ✓ SPP-Net是空间金字塔池化
- ✓ Fast R-CNN是RoI pooling

AI DISCOVERY

两阶段方法

AI DISCOVERY

3、Fast R-CNN：将分类损失和回归损失统一在同一个框架中



- ✓ 图像的warp->特征的warp
- ✓ 损失：svm+regressor->多任务损失

  (softmax+regressor)

两个任务一起优化，互相促进和增强

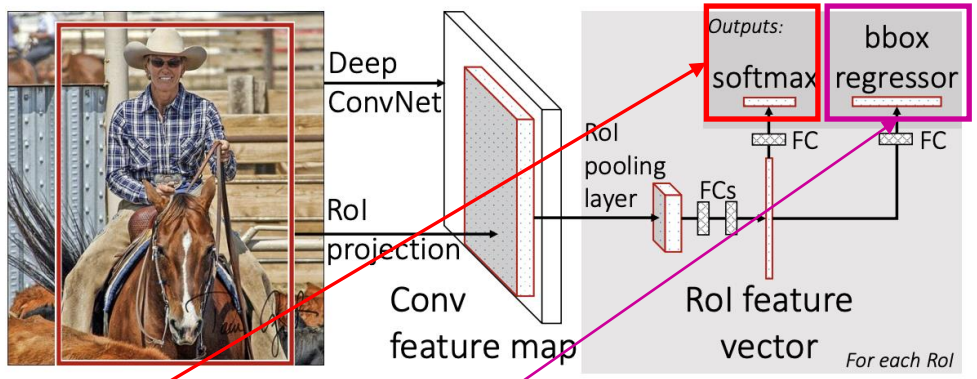
通过SS在图像中提取RoI->卷积网络提取特征->RoI Pooling->全连接层->分类/边界框回归

AI DISCOVERY

两阶段方法

AI DISCOVERY

3、Fast R-CNN



◆多任务损失：

真实边界框坐标（x,y,w,h）

预测的边界框坐标（$\hat{x}$, $\hat{y}$, $\hat{w}$, $\hat{h}$）

$$L(p,u,t^u,v) = L_{\text{cls}}(p,u) + \lambda[u \geq 1]L_{\text{loc}}(t^u,v)$$

真实类别

预测的类别得分

分类损失

边界框回归损失-L1 loss

u=0为背景类，不对应任何实际类别，无边界回归损失

$$L_{\text{loc}}(t^u,v) = \sum_{i\in\{\text{x,y,w,h}\}} \text{smooth}_{L_1}(t_i^u - v_i),$$

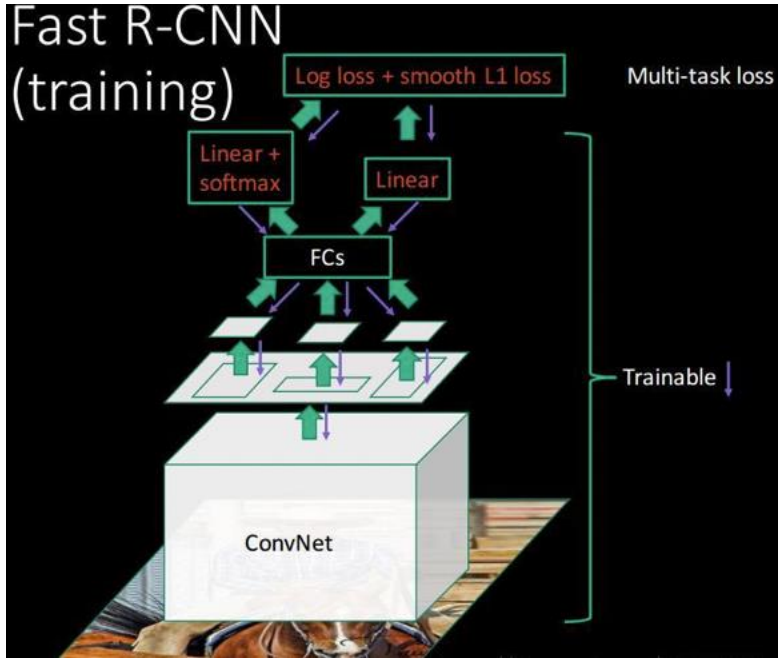
in which

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

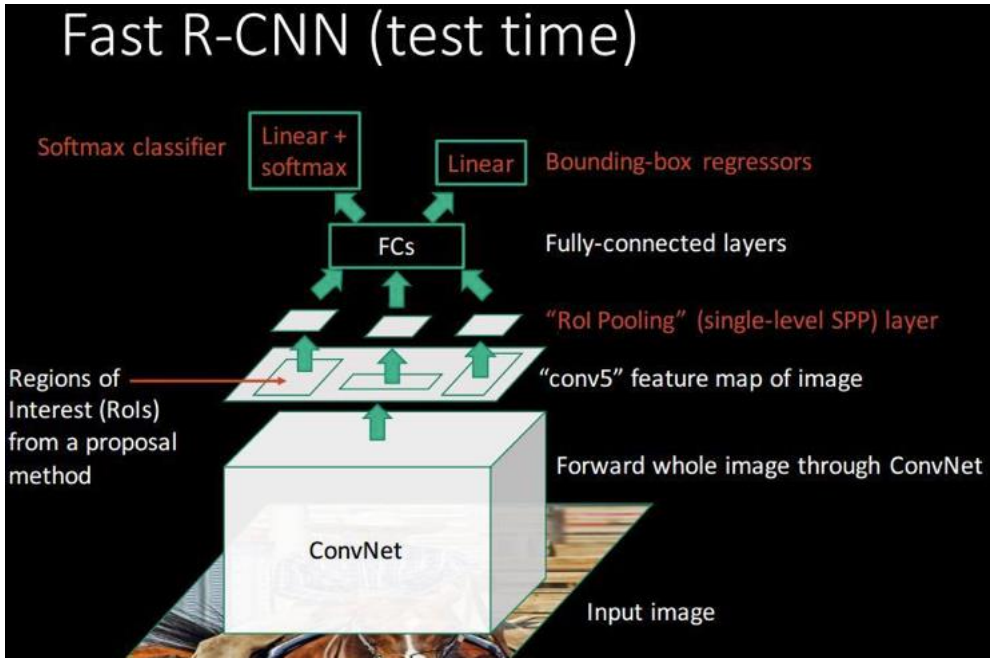
两阶段方法

AI DISCOVERY

3、Fast R-CNN：训练和测试过程小结



训练过程



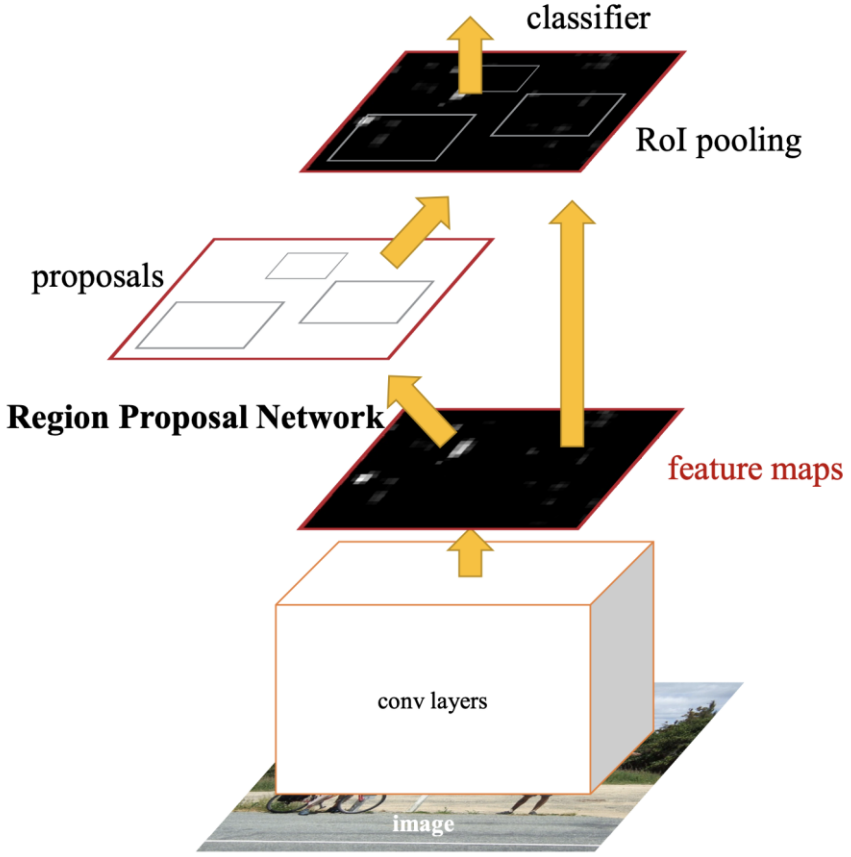
测试过程（无损失计算和梯度反传）

AI DISCOVERY

两阶段方法

AI DISCOVERY

4、Faster R-CNN：端到端检测网络，极大提升了检测速度



◆如何实现：

✓ 在最后一个卷积层后面添加Region Proposal Network (RPN)

✓ 候选区域由RPN网络直接生成，不再依靠额外的候选区域生成算法（如SS）

✓ 由RPN网络提取候选区域后，类似Fast RCNN，使用RoI Pooling，再使用两个分支分别计算类别和边界框回归

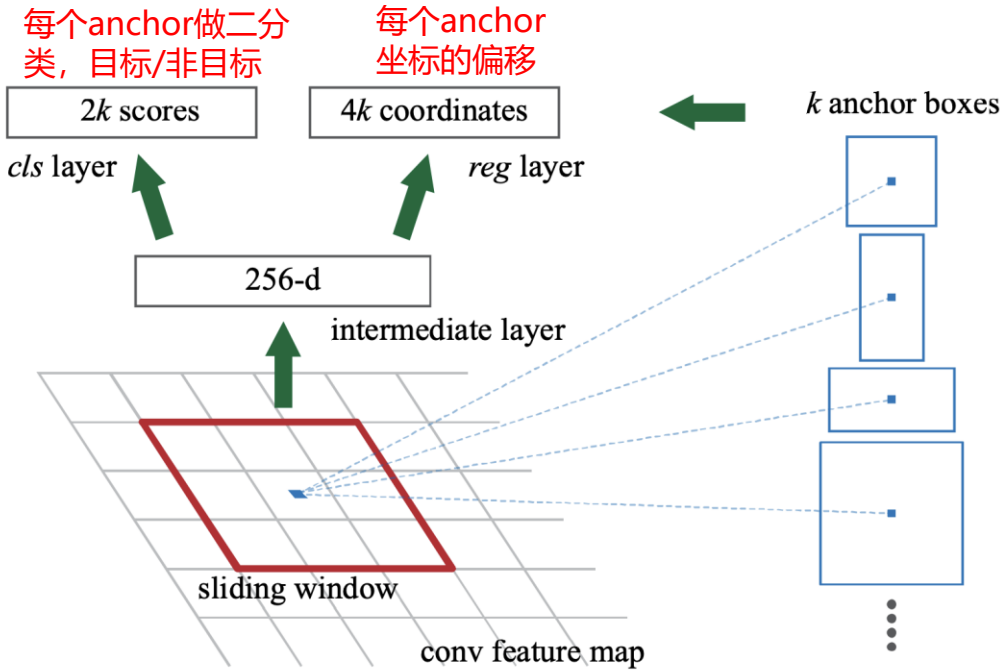
AI DISCOVERY

Shaoqing Ren, Kaiming He, Ross B. Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” NIPS 2015

两阶段方法

AI DISCOVERY

4、Faster R-CNN RPN网络：直接在卷积特征图上计算目标可能的位置



- ✓ 在最后一层特征图上使用滑动窗口
- ✓ 构建一个小的网络，来

（1）分类：判断每个anchor（锚点）区域内是否存在目标

（2）回归：确定anchor边界框的位置（不同形状的锚点框在边界框回归函数中体现）

3*3*(2k+4K)的卷积核在特征图上滑动

2k对应k个anchor的前景/背景分类

4k对应k个anchor的位置偏置

- ✓ 滑动窗口在特征图上的位置，对应于原图上的位置
- ✓ 边界框回归，使anchor位置更加精确
- ✓ 由RPN网络可以定位出原图中可能存在物体的候选区域

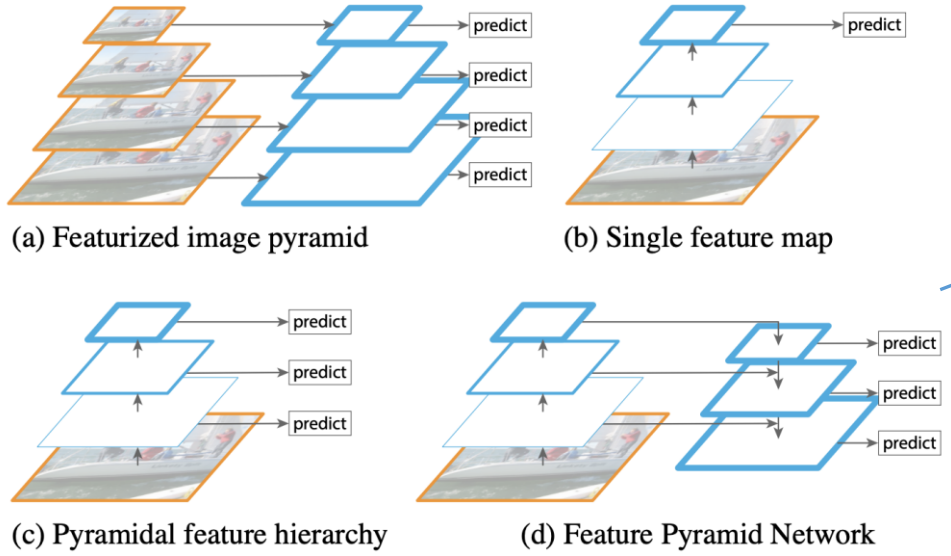
AI DISCOVERY

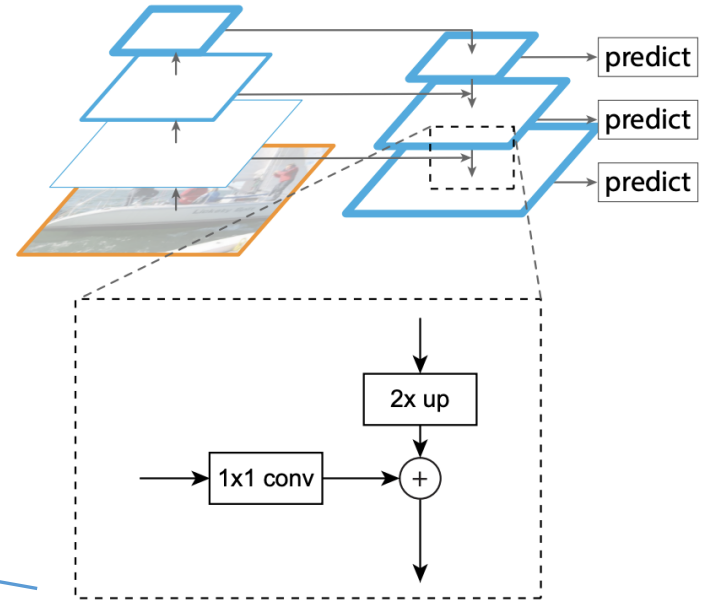
两阶段方法

蓝色越深代表语义信息越多
蓝色的框越大，特征图的分辨率

5、FPN：特征融合，多层预测以提升精度

- ✓ 越深层的特征图，包含的语义信息更强
- ✓ 越浅层的特征图，包含的上下文信息更强





FPN通过将深层特征与浅层特征相融合，并在多层预测：

- ✓ 加强了浅层特征图的语义，特征更加鲁棒，定位更准确
- ✓ 提高检测精度，尤其是对小目标提升比较大

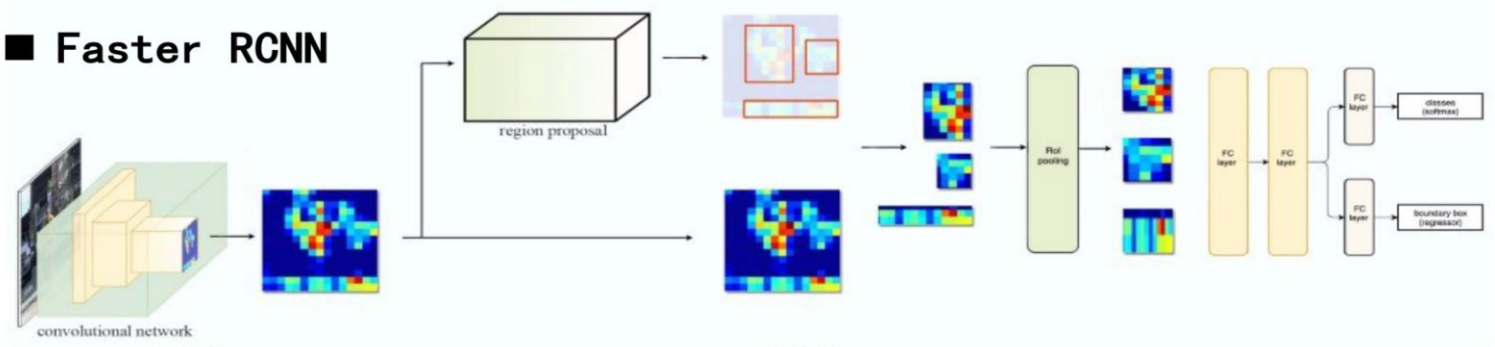
Lin, Tsung-Yi, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. "Feature Pyramid Networks for Object Detection." CVPR 2017.

两阶段方法

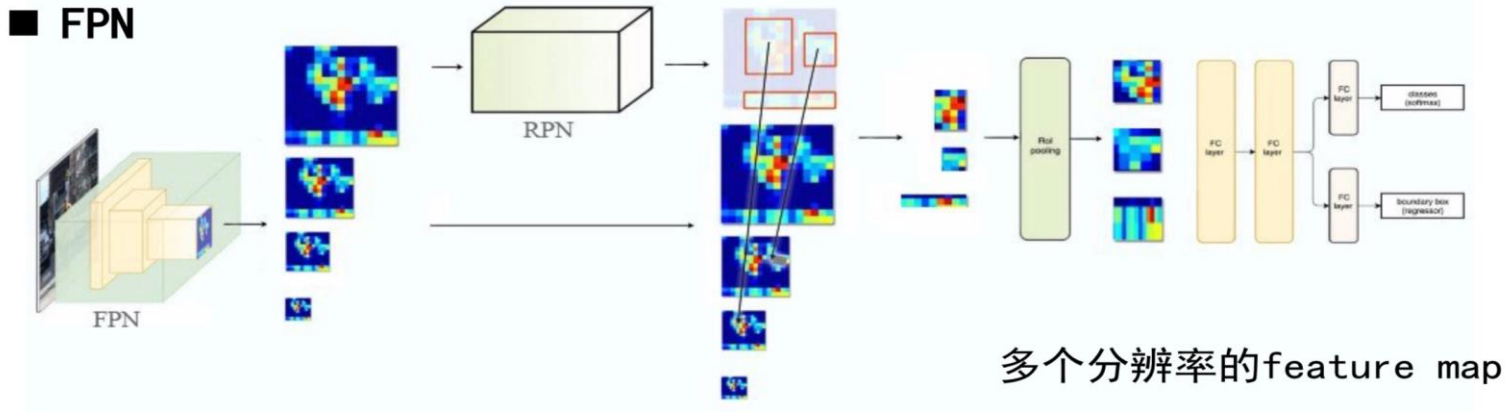
AI DISCOVERY

5、FPN：特征融合，多层预测以提升精度

◆ 网络结构对比：



Faster R-CNN在单一尺度特征图上使用RPN网络，提取region proposal，并使用单张特征图进行预测

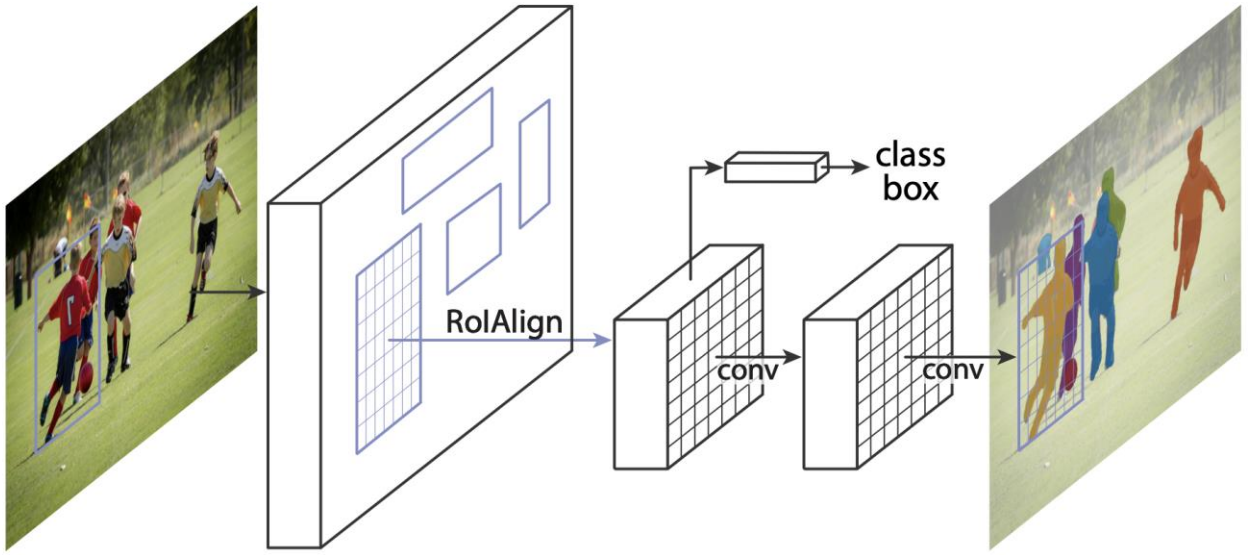


FPN在特征金字塔上使用RPN网络提取region proposal，并在特征金字塔的多个尺度上进行预测

两阶段方法

AI DISCOVERY

6、Mask-RCNN



◆ 主要贡献:

✓ 强化的基础网络：通过ResNet-101+FPN用作特征提取网络，达到state-of-the-art的效果。

✓ 采用RoI Align，解决misalignment

✓ 处理分割任务的分支使用全卷积网络

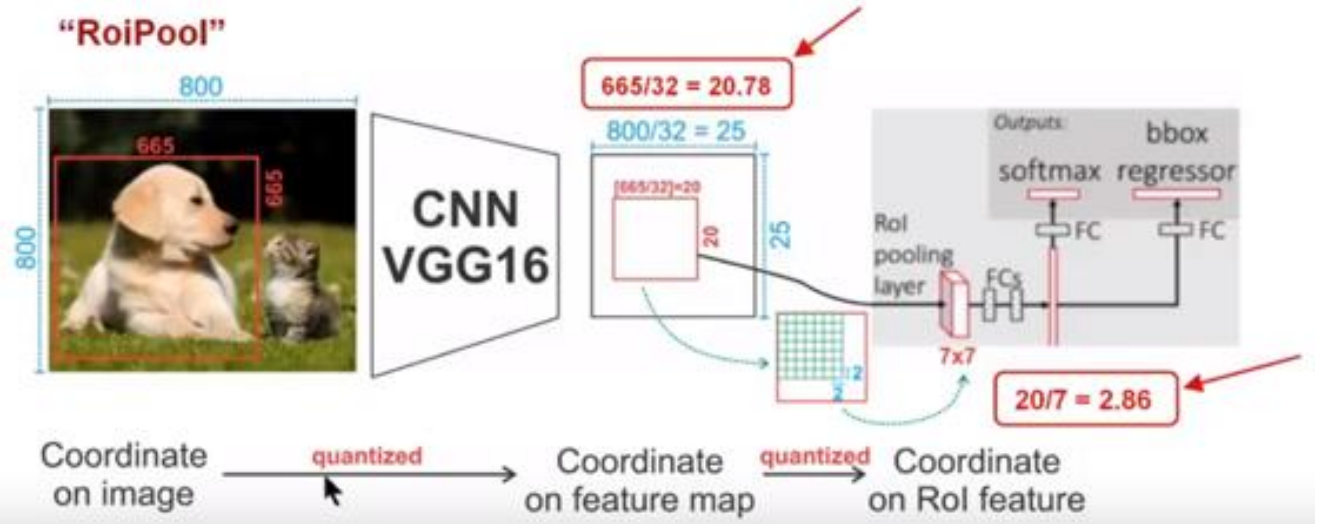
AI DISCOVERY

He K, Gkioxari G, Dollár P, et al. Mask r-cnn, Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

两阶段方法

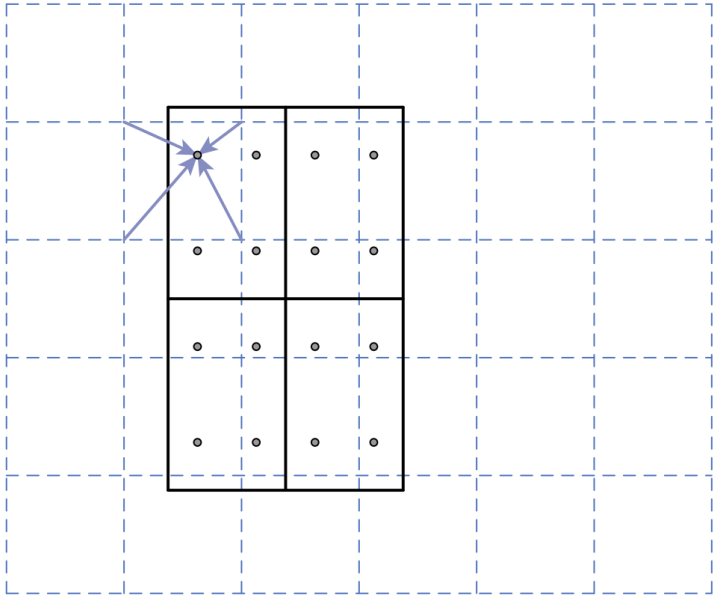
AI DISCOVERY

6、Mask-RCNN RoI Align



由于RoI pooling中的取整操作，经过RoI pooling，会使产生misalignment，导致回归坐标发生偏差，对于检测精度有影响，分割的影响更大

RoI Align:



RoI Align采用双线性差值，很好的解决了misalignment

AI DISCOVERY

目标检测

AI DISCOVERY

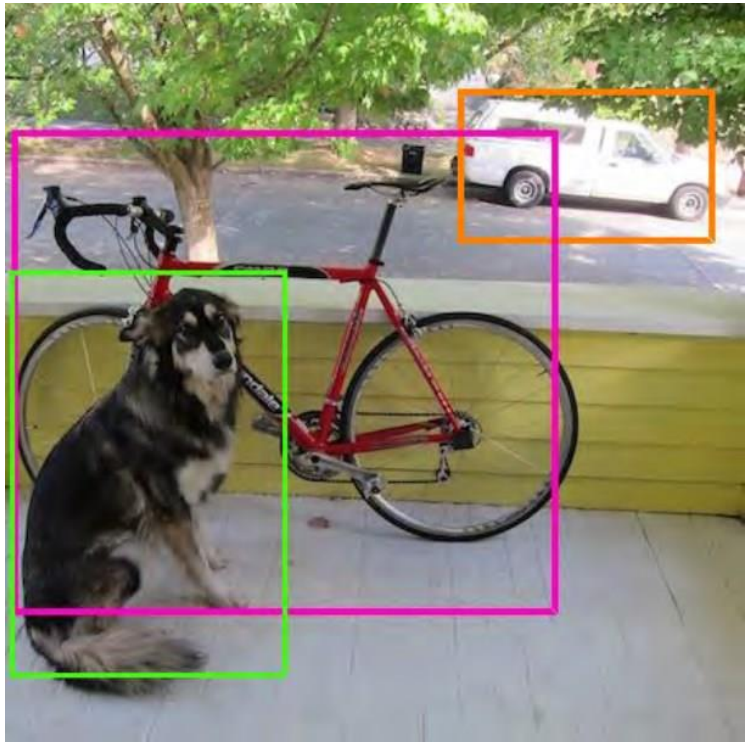
两阶段方法	R-CNN、SPP-Net、Fast R-CNN、Faster-RCNN、FPN、Mask-RCNN
一阶段方法	YOLO、SSD
最新进展	IoU-Net、GIoU

AI DISCOVERY

一阶段方法

You Only Look Once:
Unified, Real-Time Object Detection

1、YOLO：实时检测



- ✓ YOLO的检测速度非常快。在Titan X上，不需要经过批处理，标准版本的YOLO系统可以每秒处理45张图像；YOLO的极速版本可以处理150帧图像。
- ✓ YOLO在做预测的时候，使用的是全局图像。与sliding window和region proposal这类方法不同，YOLO一次“看”一整张图像，所以它可以将物体的整体（contextual）的类别信息（class information ）以及 外观信息（appearance information） 进行编码，背景错误分类为目标的概率低。
- ✓ YOLO学到物体更泛化的特征表示。当在自然场景图像上训练YOLO，再在艺术风格化图像的数据集上去测试YOLO时，YOLO的表现相比其他网络要好。YOLO模型能更好的适应新的领域。

Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” CVPR 2016

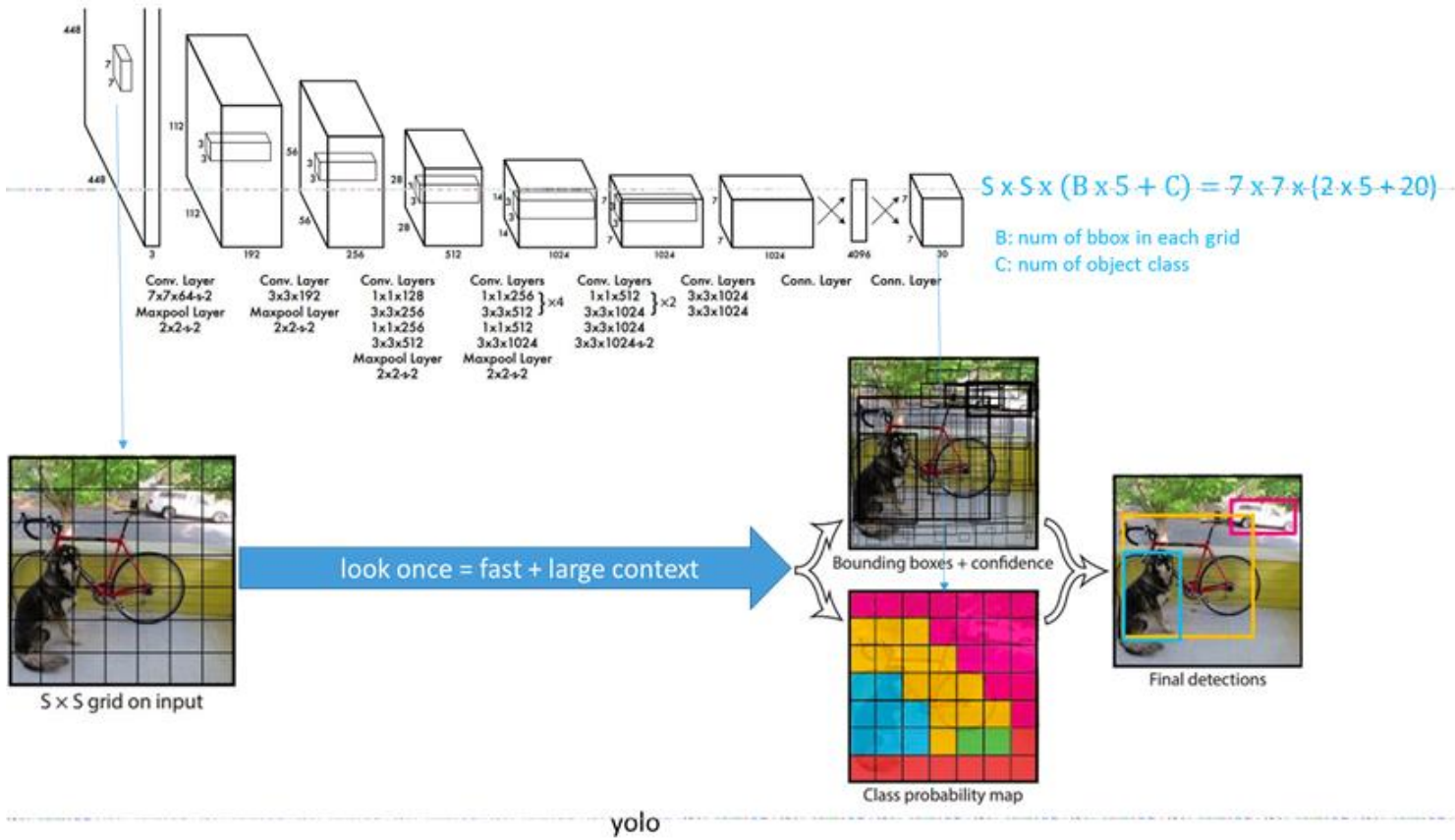
一阶段方法

AI DISCOVERY

1、YOLO：框架

◆算法流程

- ✓ 将输入的图像划分成S*S个网格（S=7）
- ✓ 每个网格预测B个边界框和这个边界框是物体的概率(Objectness);具体的，每个边界框会预测出5个值:x,y,w,h和置信度Pr(Object)*IoU(truth&pred)
- ✓ 每个网格预测C个类的概率



训练时，有物体时Pr(Object)=1，否则Pr(Object)=0

AI DISCOVERY

一阶段方法

AI DISCOVERY

1、YOLO：损失函数

表示由第i个网格中第j个边界框

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

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

边界框回归损失-均方误差

网格i，第j个边界框是否是物体，是物体为1

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

预测是目标的得分

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

预测是背景的得分

网格i，第j个边界框是否是物体，不是物体为1

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

预测物体类别

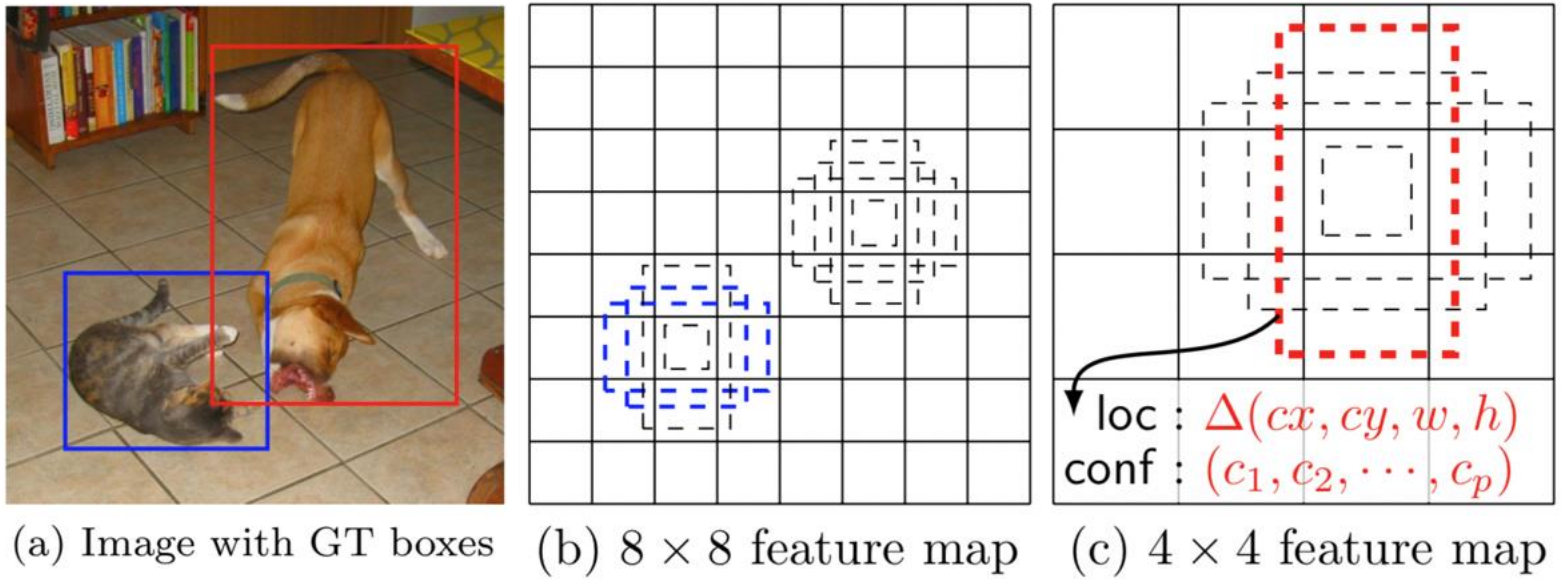
网格i中是否有目标，有目标为1

AI DISCOVERY

一阶段方法

AI DISCOVERY

1、YOLO default box：速度快，精度高



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

- ✓ 对于一张特征图，在每一个位置上提取预设数量的default box。
- ✓ 直接在特征图上密集的提取proposal进行预测，使网络不需要先提取候选目标区域，速度大幅度提升。

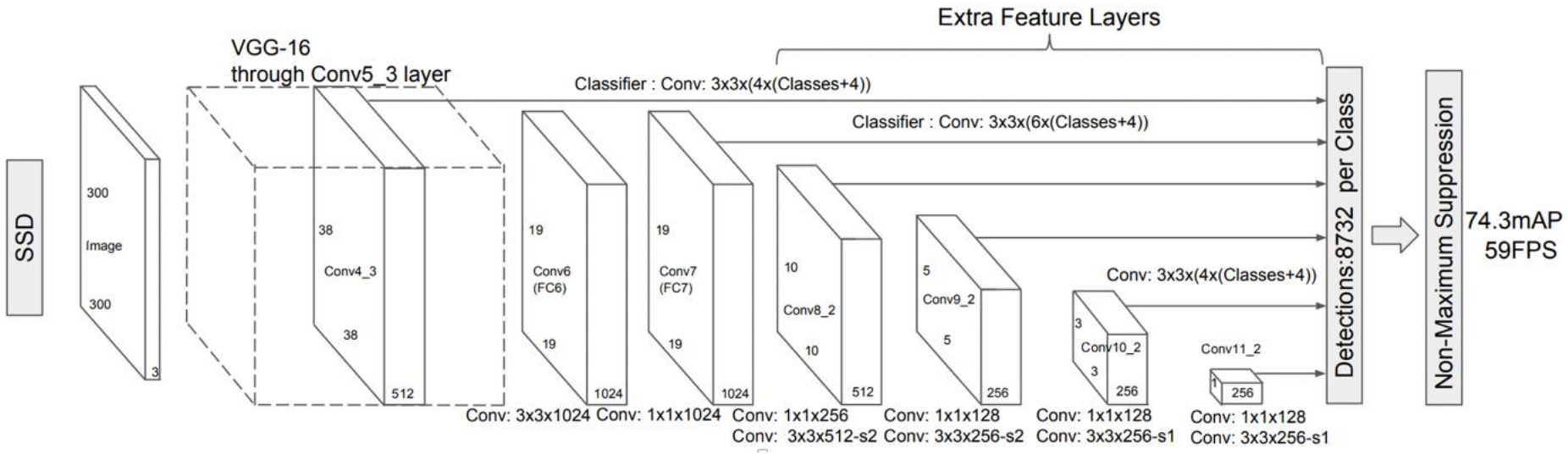
AI DISCOVERY

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, Alexander C. Berg, “SSD: Single Shot MultiBox Detector,” ECCV 2016

一阶段方法

AI DISCOVERY

2、SSD 网络结构：速度快，精度高



- ✓ 在不同尺度的特征图，直接提取预设数目default box，进行预测
- ✓ 提高检测精度（尤其在小目标有提升）

AI DISCOVERY

一阶段方法

AI DISCOVERY

2、SSD 损失函数

分类的置信度

Ground truth

$$L(x,c,l,g)=\frac{1}{N}\left(L_{conf}(x,c)+\alpha L_{loc}(x,l,g)\right)$$

预测框的坐标

分类损失

定位损失

$$L_{loc}(x,l,g)=\sum_{i\in Pos}^{N}\sum_{m\in\{cx,cy,w,h\}}x_{ij}^{k}\,\mathrm{smooth}_{\mathrm{L1}}(l_i^m-\hat{g}_j^m)$$

第i个预测框与第j个真实框关于类别k是否匹配：0，1

预测框i与真实框j关于类别p的匹配，预测为p类的概率越高，则损失越小

$$L_{conf}(x,c)=-\sum_{i\in Pos}^{N}x_{ij}^{p}log(\hat{c}_i^p)-\sum_{i\in Neg}log(\hat{c}_i^0)$$

预测框其实没有物体，则预测为背景类概率越高，损失越小

$$\hat{c}_i^p=\frac{\exp(c_i^p)}{\sum_p\exp(c_i^p)}$$

概率通过softmax生成

AI DISCOVERY

目标检测

AI DISCOVERY

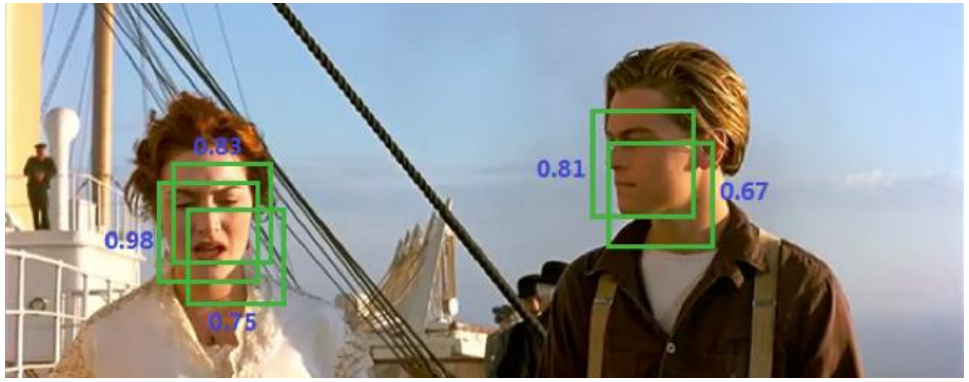
两阶段方法	R-CNN、SPP-Net、Fast R-CNN、Faster-RCNN、FPN、Mask-RCNN
一阶段方法	YOLO、SSD
最新进展	IoU-Net、GIoU

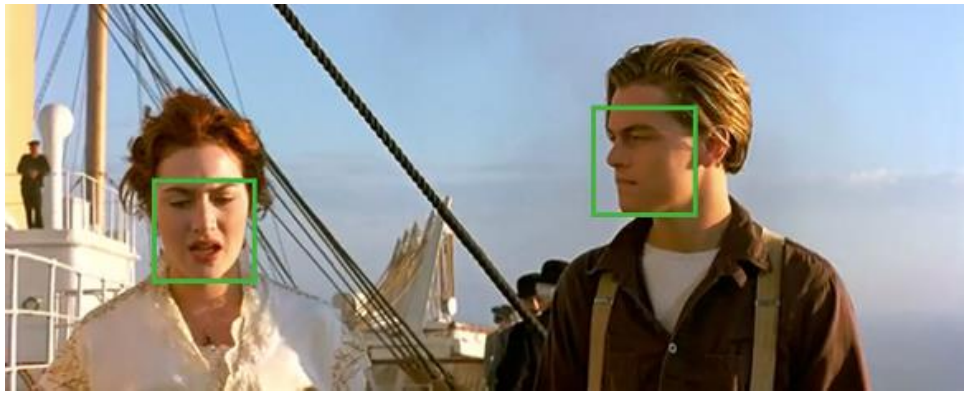
AI DISCOVERY

最新进展

AI DISCOVERY

1、IoU-Net





◆什么是NMS（非极大值抑制）

- ✓ 将同一类的所有的检测框，按照分类置信度排序
- ✓ 将与分类置信度最高的检测框的重叠面积（IoU）大于一定阈值的检测框删除
- ✓ 从未处理的框中，再选出一个分类置信度最高的检测框，重复上一步操作

通过非极大值抑制，消除冗余的检测框

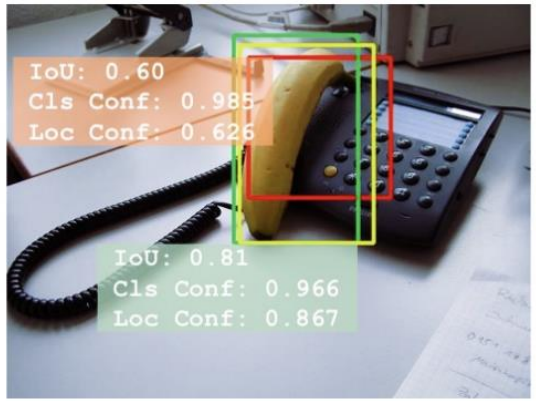
AI DISCOVERY

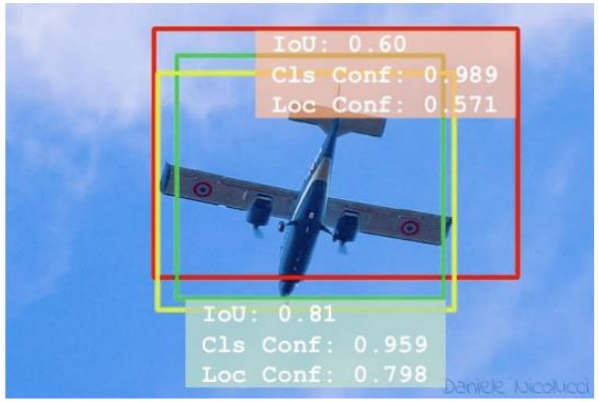
Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, Yuning Jiang “Acquisition of Localization Confidence for Accurate Object Detection” ECCV 2018

最新进展

AI DISCOVERY

1、IoU-Net





- 真实边界框
- 预测边界框
- 预测边界框

红色的预测框，分类置信度高，绿色的预测框，定位的置信度高，如果使用分类置信度来进行NMS，红色的预测框会被保留，绿色的预测都会被舍弃，很明显，在当前的例子中，绿色的预测框的效果要好。

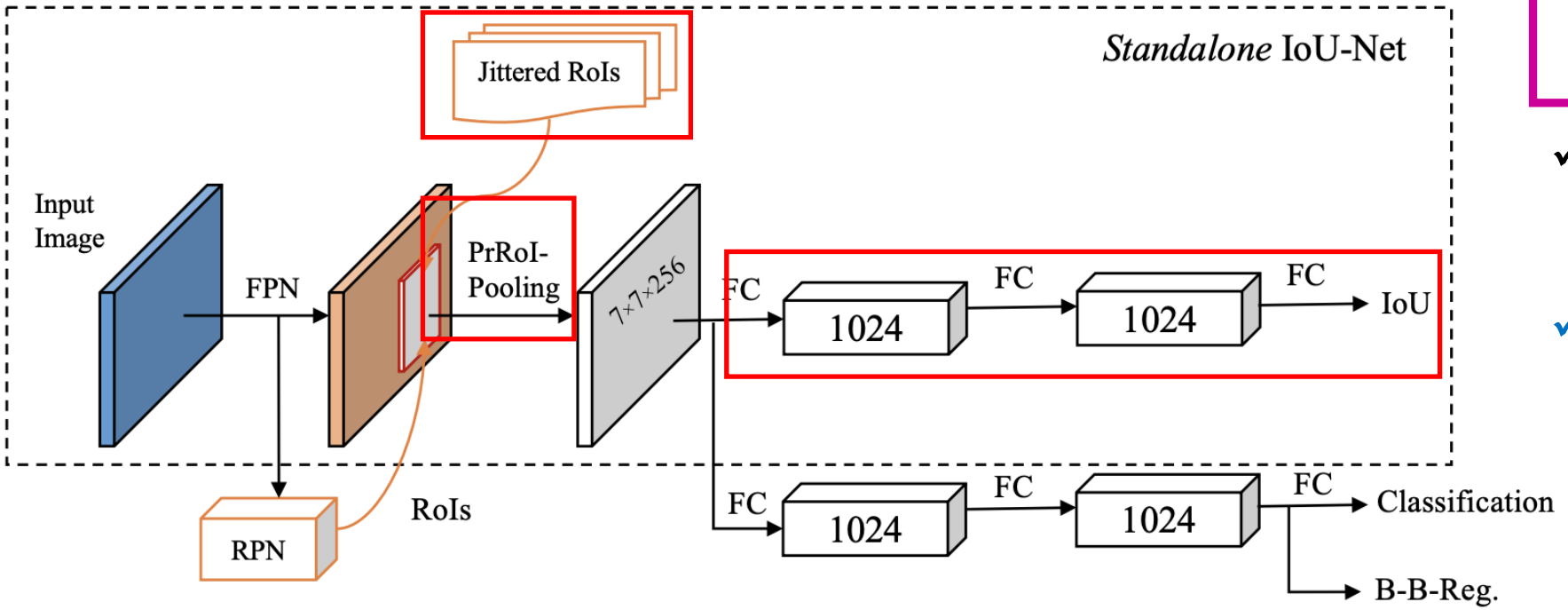
分类置信度和定位准确度不对齐：

仅依靠分类的置信度来作为NMS中的关键依据，可能会导致大量高质量的边界框丢失

最新进展

AI DISCOVERY

1、IoU-Net：网络结构



- ✓ 提出PrRoI-Pooling，来更好的解决misalignment
- ✓ 网络训练IoU分支，提取每个边界框的定位置信度
- ✓ 训练时，使用Jittered RoIs提取RoI，以更好的训练IoU分支

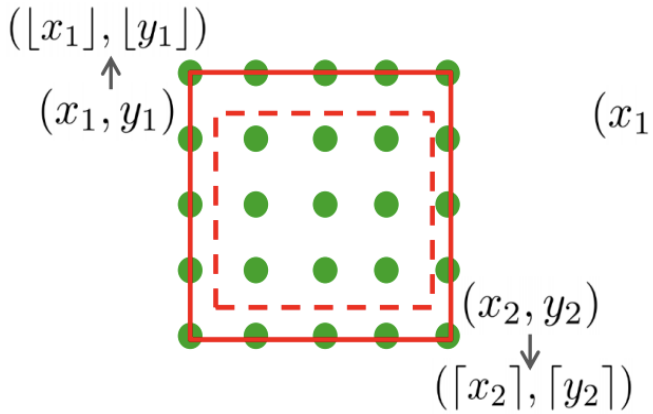
AI DISCOVERY

最新进展

AI DISCOVERY

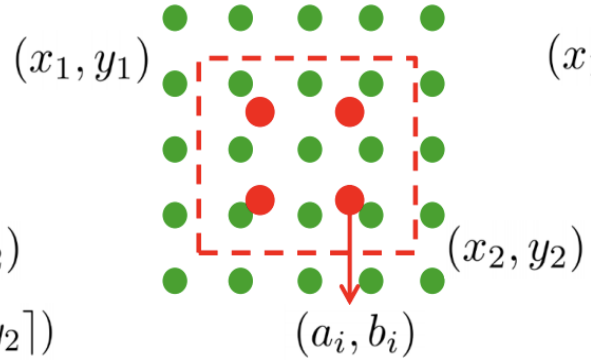
1、IoU-Net：PrRoI-Pooling

1. RoI Pooling



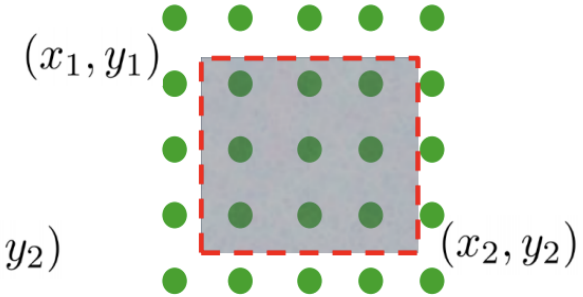
$$\frac{\sum_{i=\lfloor x_1 \rfloor}^{\lceil x_2 \rceil} \sum_{j=\lfloor y_1 \rfloor}^{\lceil y_2 \rceil} w_{i,j}}{(\lceil x_2 \rceil - \lfloor x_1 \rfloor + 1) \times (\lceil y_2 \rceil - \lfloor y_1 \rfloor + 1)}$$

2. RoI Align



$$\sum_{i=1}^{N} f(a_i, b_i)/N$$

3. PrRoI Pooling



$$\frac{\int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y)\, dxdy}{(x_2 - x_1) \times (y_2 - y_1)}$$

RoI Pooling取整运算时，会一定程度丢失位置信息（misalignment）

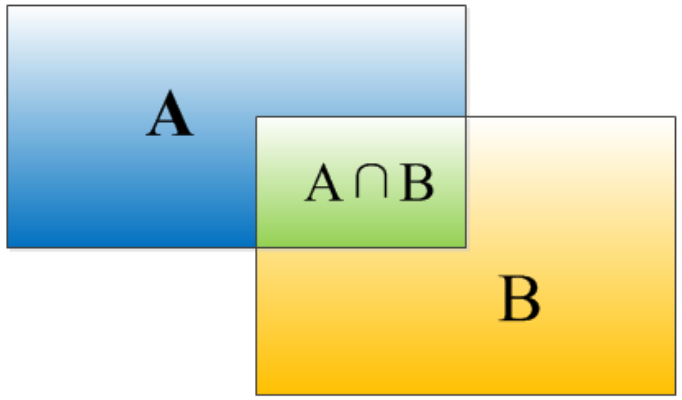
在RoI Align中采用双线性运算，需要预设N的个数

PrRoI Pooling不需要，直接使用积分取均值

AI DISCOVERY

最新进展

2、GIoU



$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

◆IoU存在的问题：

- ✓ IoU与常用的损失没有强相关性
- ✓ 如果两个对象不重叠，则IoU值将为零，并且不会反映两个形状彼此之间的距离
- ✓ IoU无法正确区分两个对象的对齐方式。

Hamid Rezatofighi,Nathan Tsoi, JunYoung Gwak,Amir Sadeghian,Ian Reid2 Silvio Savarese “Generalized Intersection over Union: A Metric and A Loss for Bounding Box

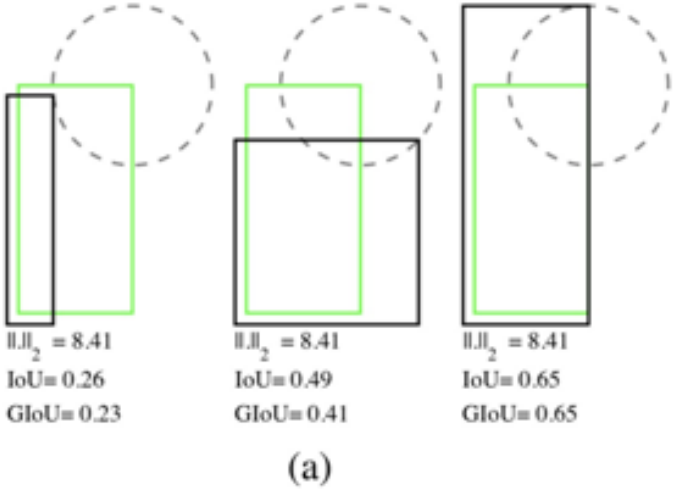
最新进展

常用的边界框回归损失，通常基于1-范式和2-范式（L1-smooth，MSE）

2、GIoU：损失函数

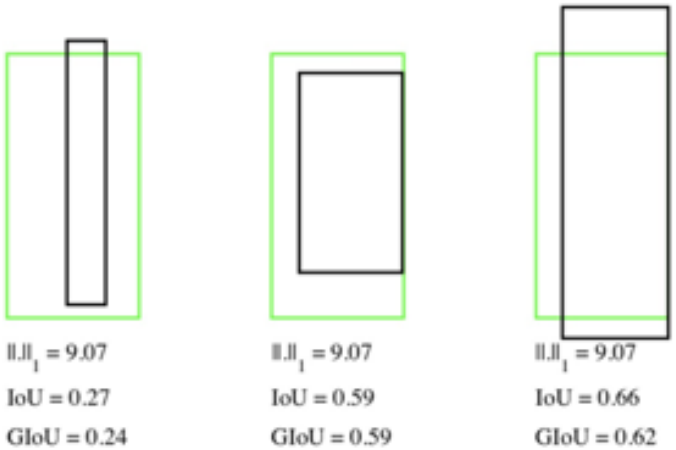
IoU与常用的损失没有强相关性

2-范式距离相同
IoU/GIoU不同：



(a)

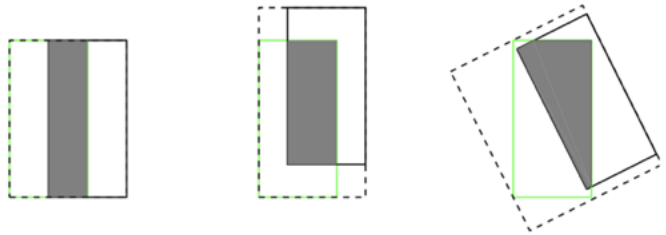
1-范式距离相同
IoU/GIoU不同：



(b)

如果两个对象不重叠，则IoU值将为零，并且不会反映两个形状彼此之间的距离

IoU无法正确区分两个对象的对齐方式。



最新进展

AI DISCOVERY

2、GIoU：损失函数

Algorithm 1: Generalized Intersection over Union

input : Two arbitrary convex shapes: $A, B \subseteq \mathbb{S} \in \mathbb{R}^n$

output: $GIoU$

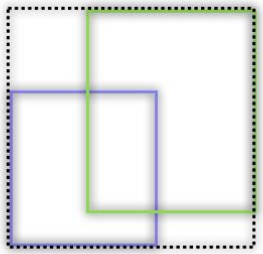
1 For A and B, find the smallest enclosing convex object C, where $C \subset \mathbb{S} \in \mathbb{R}^n$

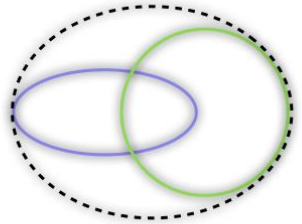
2 $IoU = \frac{|A \cap B|}{|A \cup B|}$

3 $GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}$

那么C是怎么得到的呢?

两个区域的最小外接凸多边形（或圆形）





$$\mathcal{L}_{IoU} = 1 - IoU,$$

IoU-Loss

$$\mathcal{L}_{GIoU} = 1 - GIoU.$$

GIoU-Loss

实验表明，只需要将边界框回归分支的损失改为GIoU-Loss，即可获得2%-14%的检测性能提升

AI DISCOVERY

目录

AI DISCOVERY

1 目标检测

两阶段方法、一阶段方法、最新进展

2 典型图像分析任务

图像分割、图像搜索、目标跟踪

3 特色图像分析任务

细粒度分类、风格迁移、标题生成、超分辨率

4 垂直应用与实践

医学影像分析、文字检测识别
实践：目标检测

AI DISCOVERY

典型图像分析任务

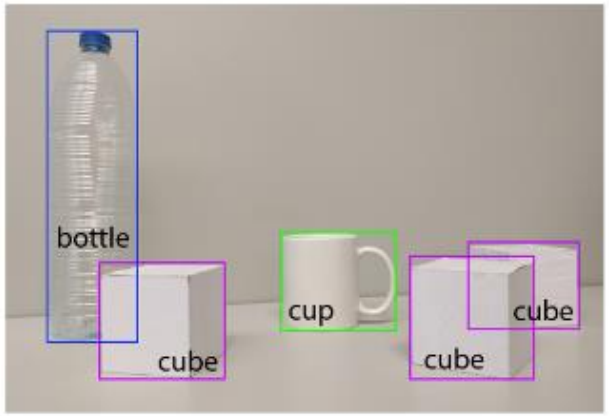
图像分割

图像搜索

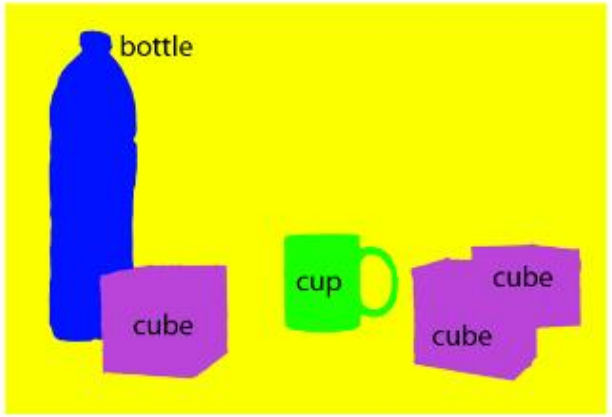
目标跟踪

图像分割

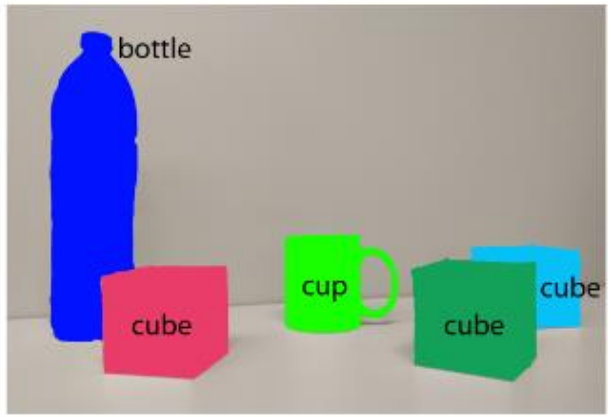
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation

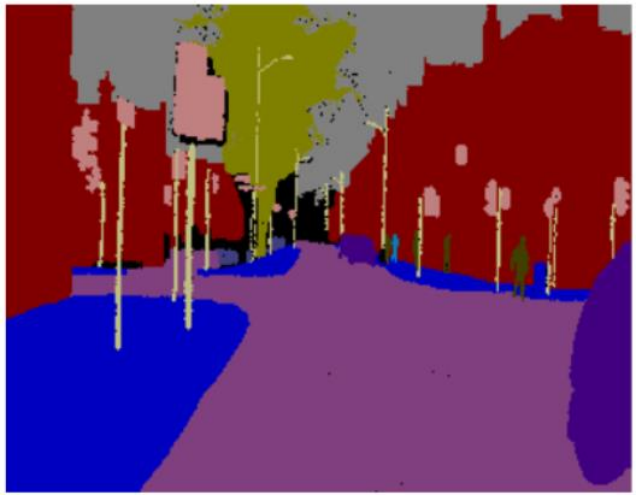
图中展示了四种图像分析任务，其中：

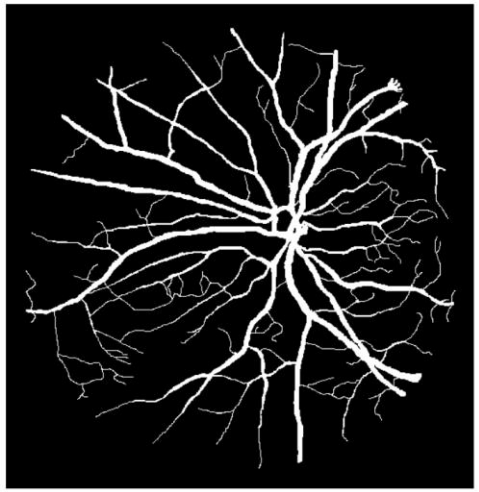
(c)表示语义分割（semantic segmentation）.

(d)表示实例分割（instance segmentation）.

语义分割是典型的图像分割问题，而实例分割就是语义分割与目标检测的结合。本节我们将以语义分割为例，了解图像分割的基本内容。

图像分割





自动驾驶

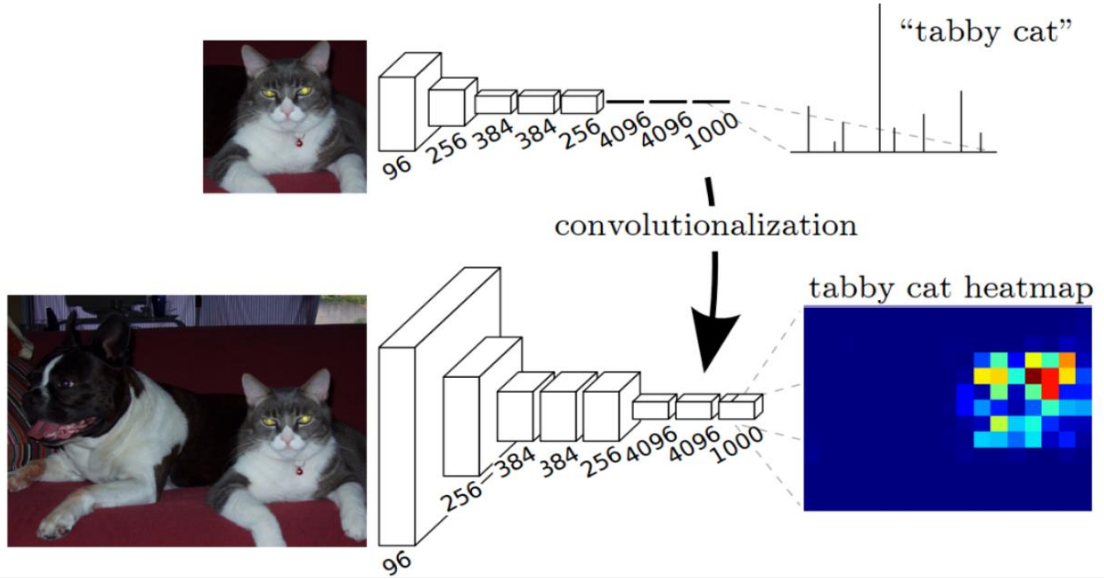
医学诊断

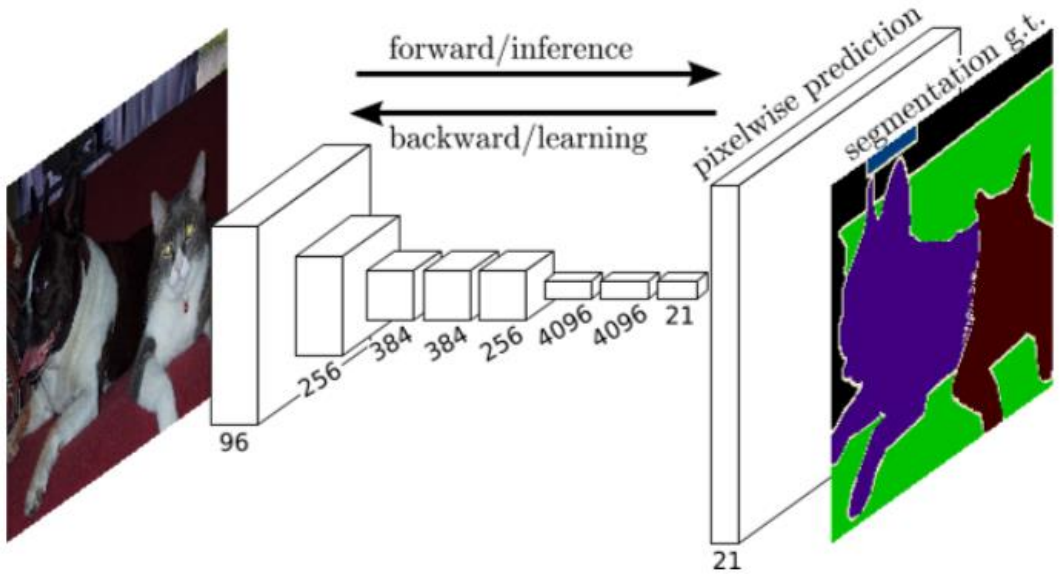
图像分割是场景理解（complete scene understanding）的关键技术，在自动驾驶，人机交互，图像搜索，增强现实和医学诊断等应用场景中具有重要意义，是计算机视觉领域的核心问题。

深度学习在图像分割中的应用

AI DISCOVERY

深度学习在图像分割领域的里程碑--------**FCN**（Fully Convolution Network）



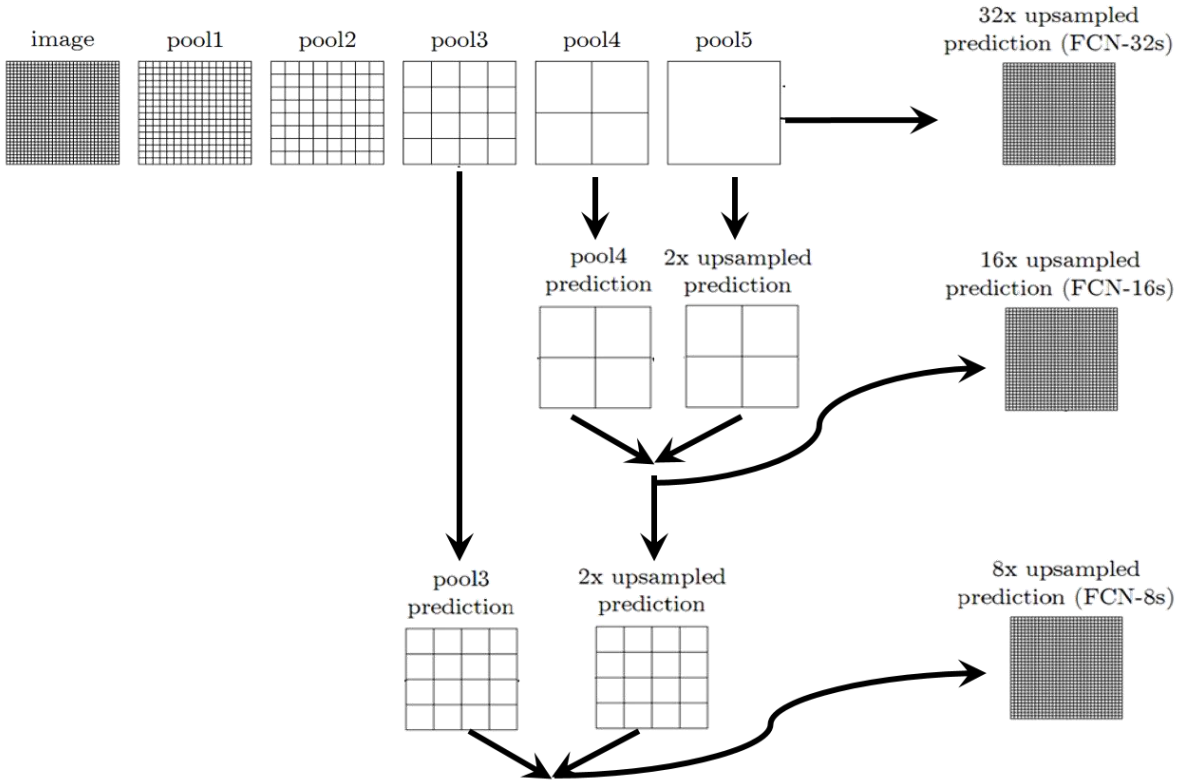


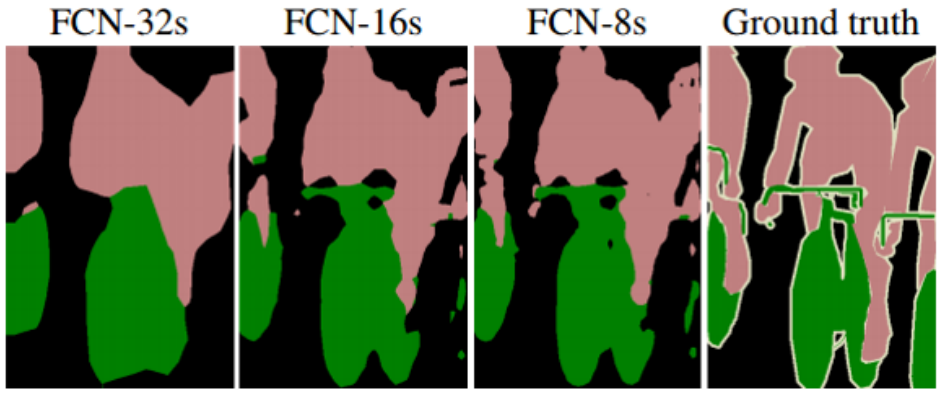
在VGG-16的基础上将全连接层替换为卷积层，**输出空间映射**而不是分类分数。这些映射由小步幅卷积上采样（又称反卷积）得到，来产生密集的像素级别的预测。

AI DISCOVERY

Jonathan Long, Evan Shelhamer, Trevor Darrell "Fully Convolutional Networks for Semantic Segmentation

深度学习在图像分割中的应用





该工作被视为**里程碑式的进步**，它阐释了CNN如何可以在语义分割问题上被端对端的训练，而且高效的学习了如何基于任意大小的输入来为语义分割问题产生像素级别(pixel-wise)的预测。

更加优良的分割网络

AI DISCOVERY

◆ 基于FCN的主流优化策略：

✓ 上下文信息整合（Integrating Context Knowledge）

DeepLab

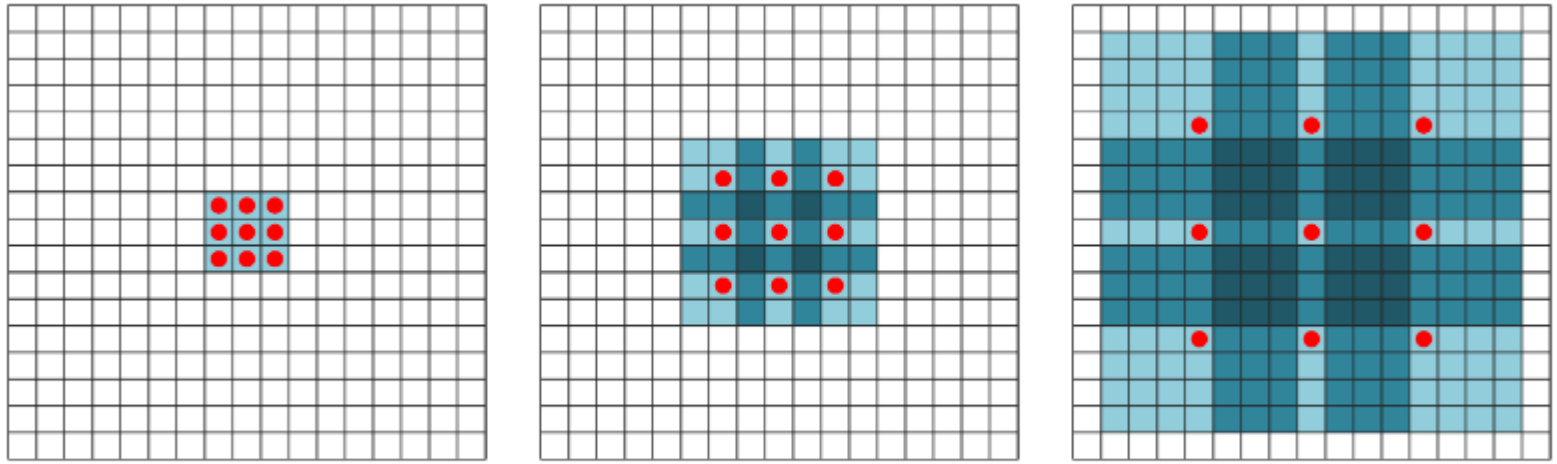
✓ 解码器变种 （Decoder Variants）

SegNet U-net

DeepLab 系列网络*

AI DISCOVERY

◆扩张卷积（Dilated Convolutions）



(a) 1-dilated　(b) 2-dilated　(c) 3-dilated

dilated参数与感受野的关系：

$F_i=2^{i+1}-1$

扩张卷积又称为空洞卷积(a-trous convolutions)，根据扩张率扩充其尺寸，为空元素位置补零。这种卷积核可以指数级地扩大感受野而不丢失分辨率，这意味着扩张卷积可以在任意分辨率图片上高效地提取密集特征。

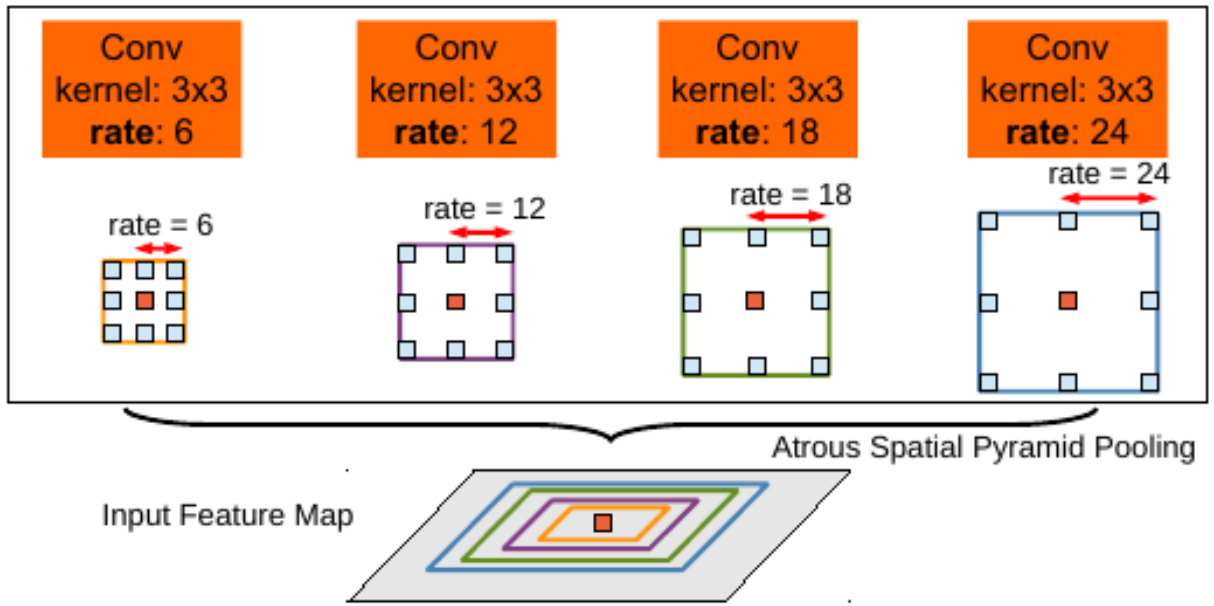
AI DISCOVERY

*：DeepLab 是一个系列网络，目前为止一共有四个结构分别为：v1，v2，v3，v3+。我们主要关注其针对上下文整合（integrating context knowledge）采用的技术。

DeepLab 系列网络

AI DISCOVERY

◆ Atrous Spatial Pyramid Pooling (ASPP)



rate的定义与上一页有所不同，这里定义为：

K+(K-1)*(rate-1)

K为卷积核的大小

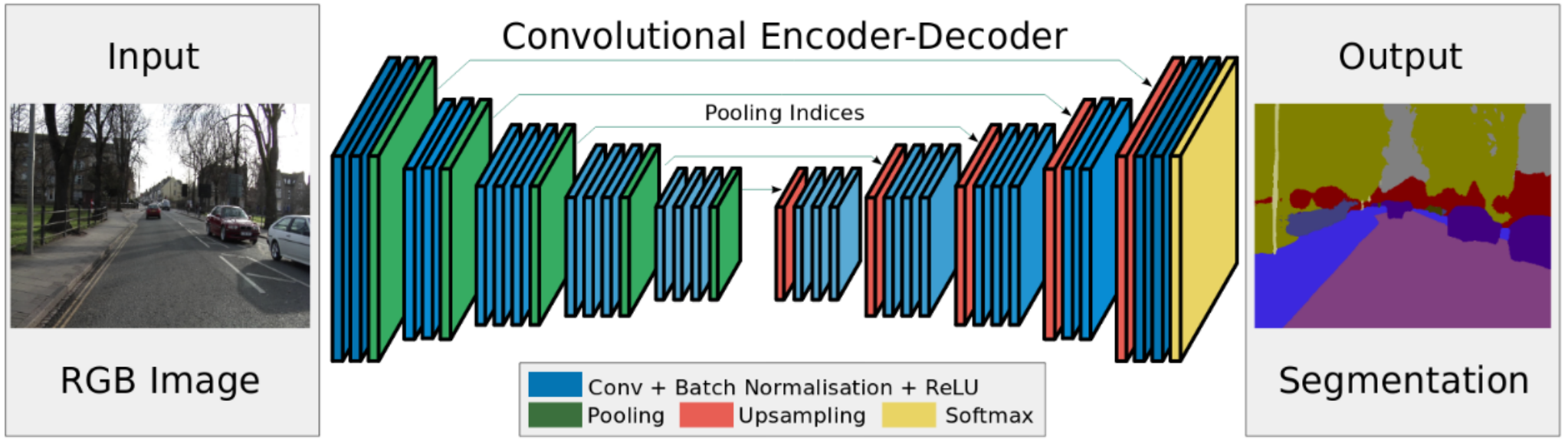
DeepLab使用了ASPP结构，在多个尺度的卷积上并行执行特征提取，并最后进行融合，以达到上下文信息整合的作用。

AI DISCOVERY

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation" ECCV 2018

实时分割的SegNet

AI DISCOVERY

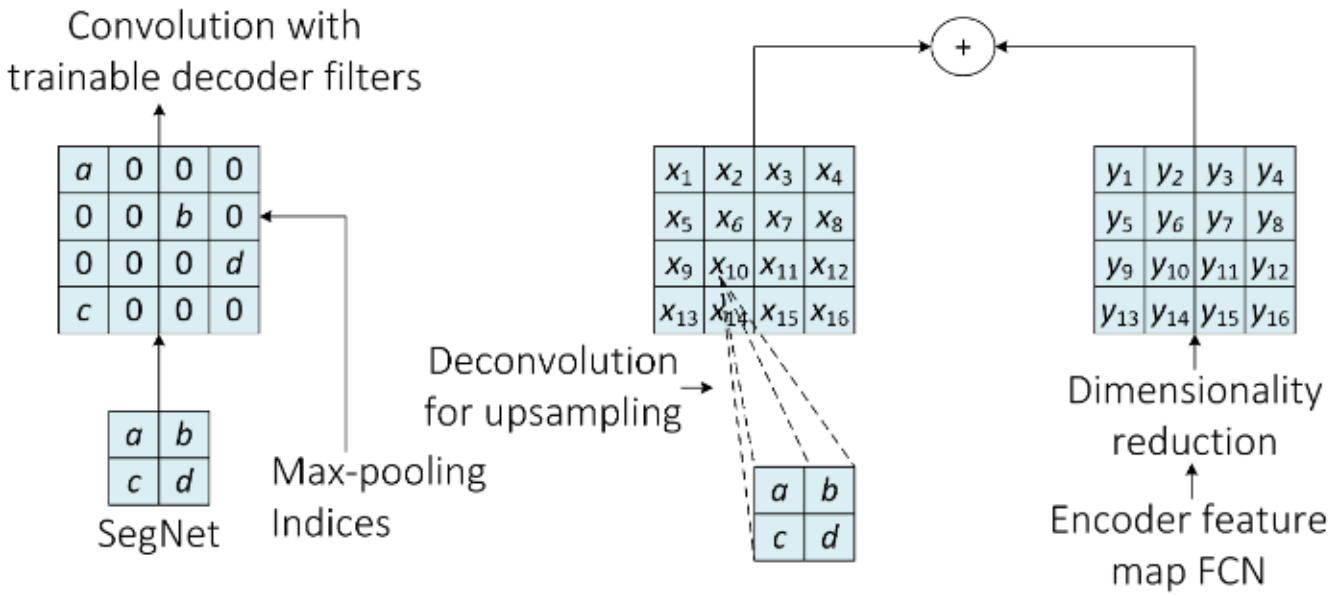


SegNet使用了经典的编码器-解码器（encoder-decoder）结构。SegNet专注于道路场景下的语义分割，由于自动驾驶对于分割速度的要求非常高，SegNet在解码器部分进行了优化，使得其能够满足实时语义分割（real-time semantic segmentation）的需求。

AI DISCOVERY

Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation" IEEE Transactions on PAMI (2017)

实时分割的SegNet

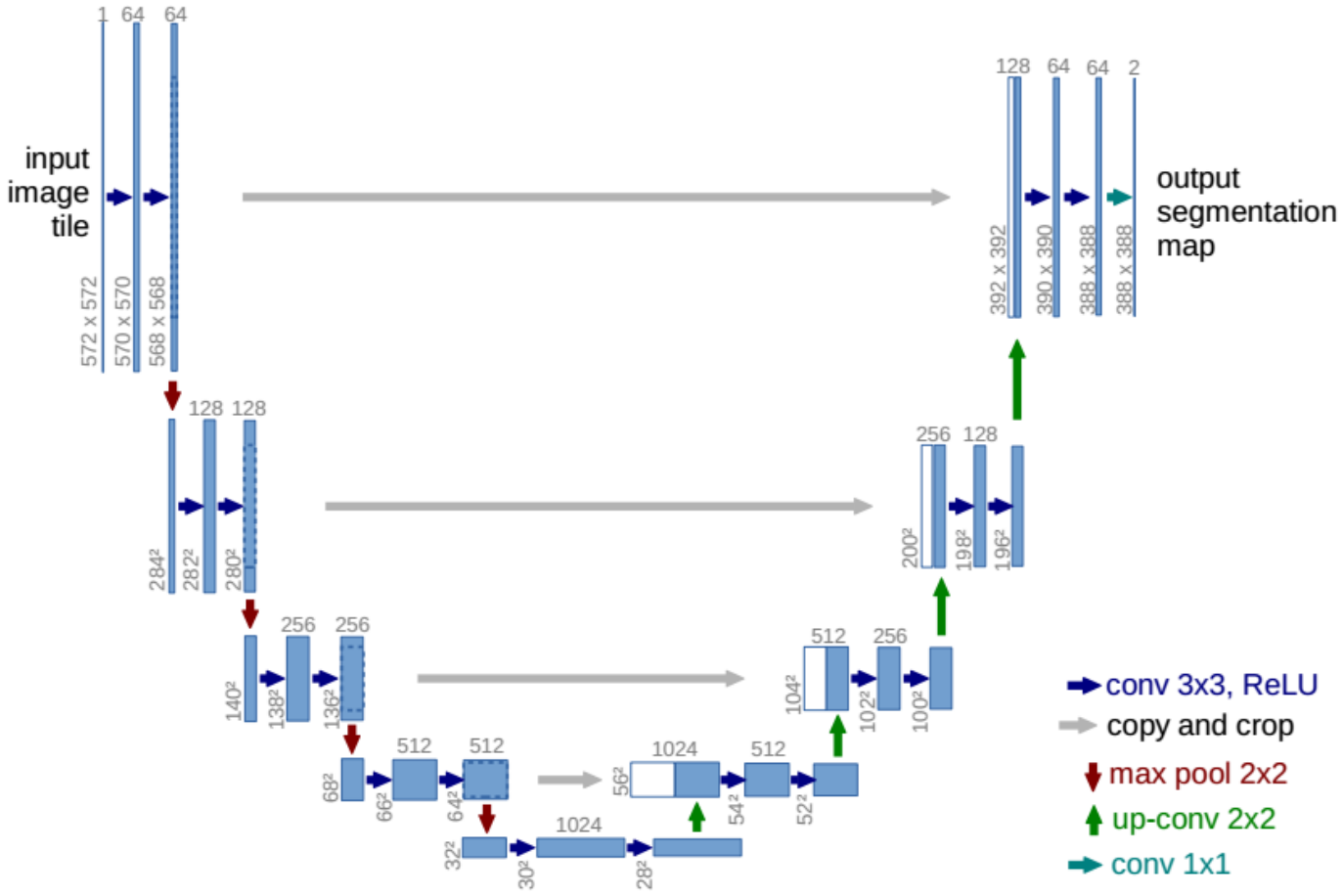


解码器对比：SegNet（左）FCN（右）

- ✓ 池化运算时，SegNet使用对应编码器记录的位置信息进行上采样，其他位置为零。这样的方法加快了网络的运行速度，满足了实时分割的需求。
- ✓ FCN使用反卷积滤波器进行上采样，然后与对应编码器的特征映射相加。

医学影像分割的重要网络U-net

AI DISCOVERY



- ◆ U-net也是解码器变种的一个经典网络
- ✓ 解码器采用了反卷积滤波器进行上采样，并增加对应编码器的特征。
- ✓ 在解码器部分，并没有进行单纯的上采样操作，每次上采样后伴随着两次卷积。
- ✓ 与SegNet不同，医学图像对于实时性的要求不高，但对图像的精度要求苛刻。

AI DISCOVERY

Olaf Ronneberger, Philipp Fischer, Thomas Brox "U-Net: Convolutional Networks for Biomedical Image Segmentation" MICCAI 2015

典型图像分析任务

图像分割

图像搜索

目标跟踪

基于内容的图像检索

AI DISCOVERY

- ✓ 随着各种社交网络的兴起，网络中图片数据每天都以惊人的速度增长。如何有效地从巨大的图像数据库中检索出用户需要的图片，成为信息检索领域研究者感兴趣的一个研究方向。
- ✓ 基于内容语义的图像检索技术是指根据图片的颜色、纹理及图片包含的**物体、类别**等信息检索图片。

(a)光照变化　(b)尺度变化　(c)视角变化　(d)遮挡　(d)背景杂乱

相同物体图像检索面临的挑战

(a)类内变化巨大(湖泊)

(b)类间相似性干扰

相似类别图像检索面临的挑战

目标：

越来越快！

越来越准！

AI DISCOVERY

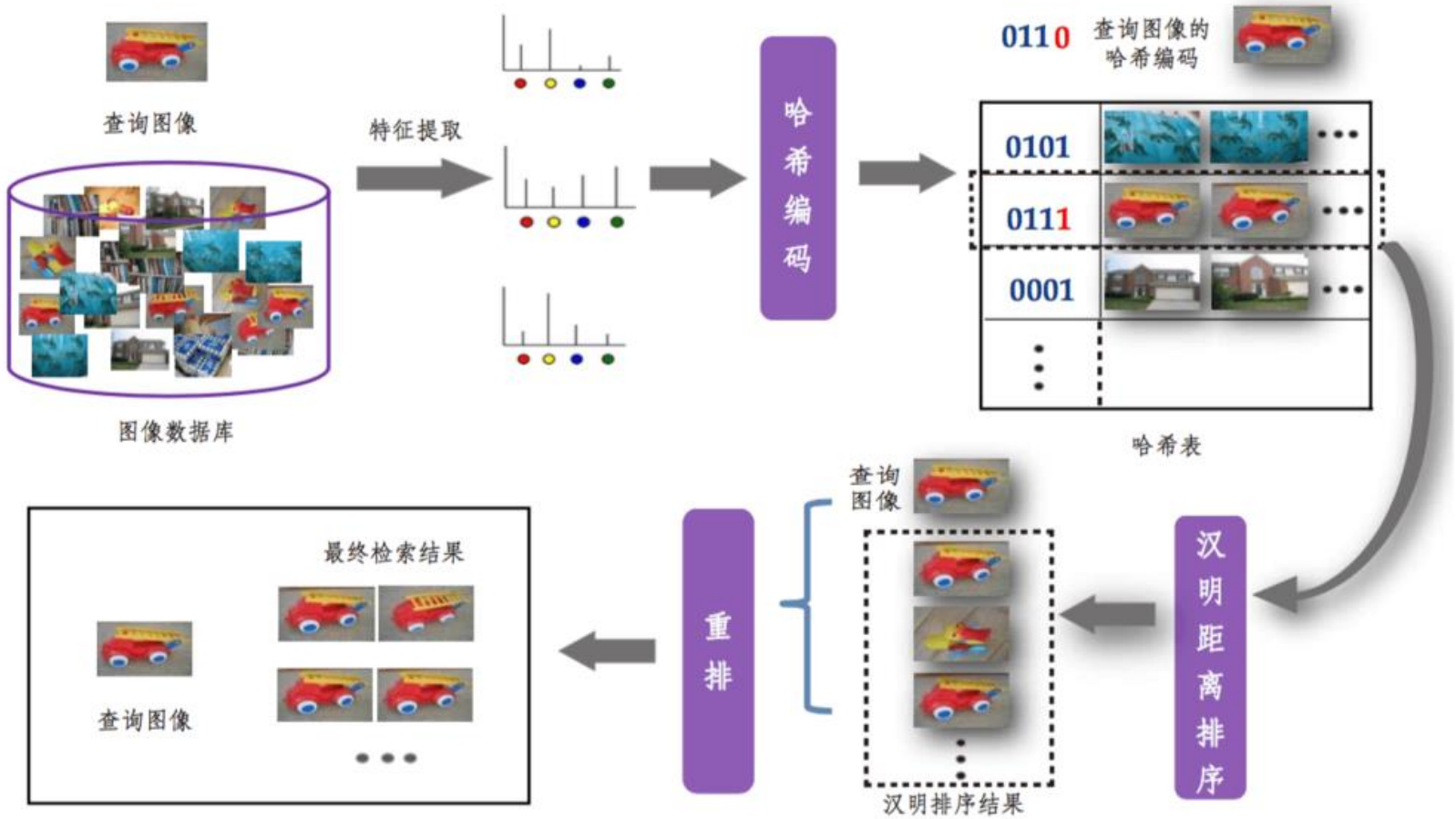
图 1.3　相同物体图像检索和相同类别图像检索面临的挑战

图像检索框架

AI DISCOVERY

◆基于哈希的图像检索架构：

二进制编码：存储小，匹配计算速度快



AI DISCOVERY

图像特征表示

Hand crafted features (颜色特征、形状特征、纹理特征、SIFT、SURF等)

2012

CNN based features (AlexNet、GoogleNet、ResNet等)

AlexNet

- 2012年ImageNet分类冠军
- 第一次证明了CNN在复杂模型下的有效性，并且GPU使得训练在可接受的时间范围内得到结果。

GoogleNet

- 2014年ImageNet分类冠军
- 证明了用更多的卷积，更深的层次可以得到更好的结构。

ResNet

- 2015年ImageNet分类冠军
- GoogleNet的进一步推广，将模型推向更深。

哈希编码学习——方法分类

AI DISCOVERY

◆ **什么是哈希编码?**

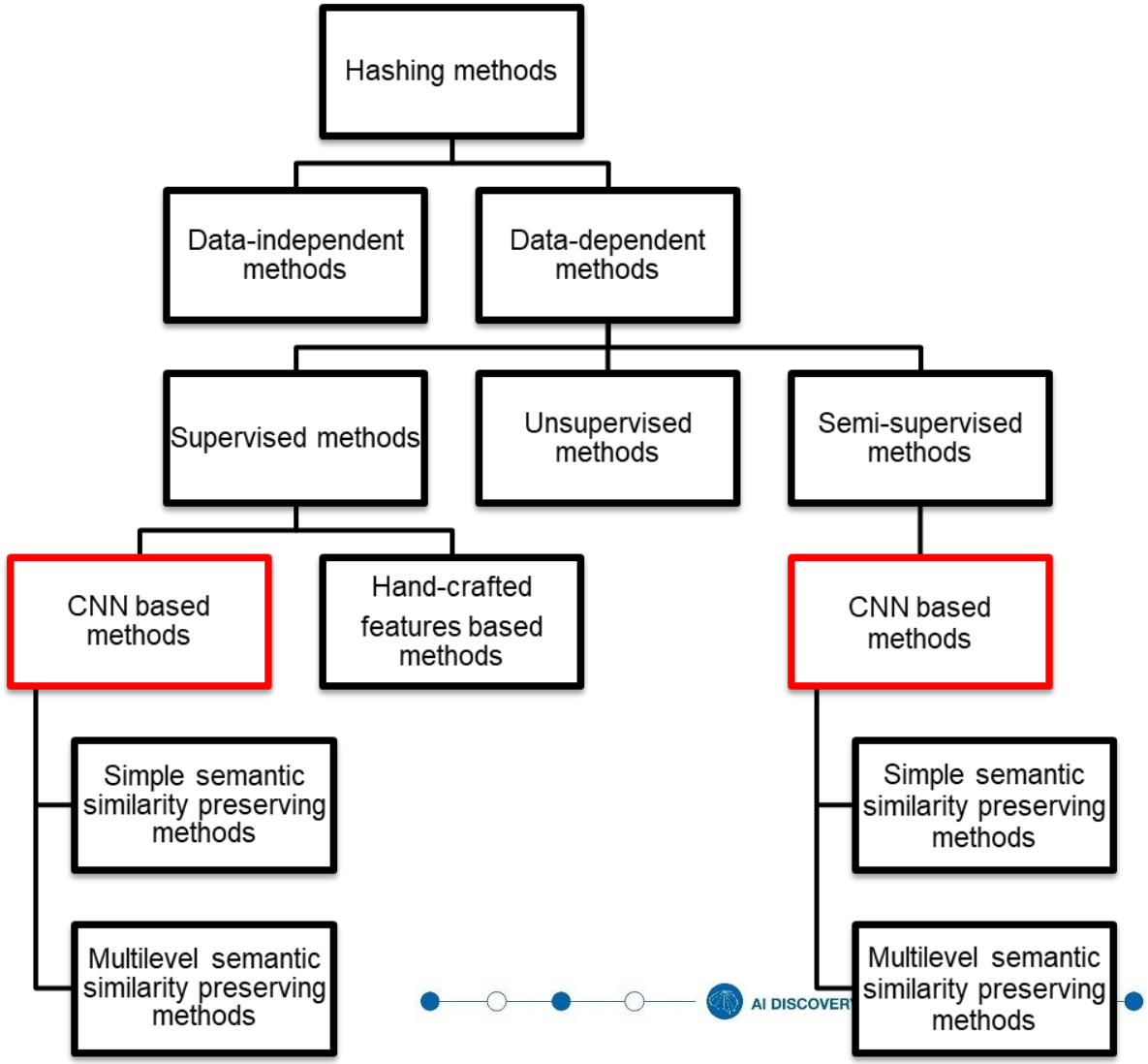
将图片的高维内容特征映射到汉明空间（二值空间）中，生成一个低维的哈希序列来表示一幅图片。

◆ **哈希编码有什么好处?**

降低了图像检索系统对计算机内存空间的要求，提高了检索速度，能更好的适应海量图片检索的要求。

◆ **哈希编码有几个阶段?**

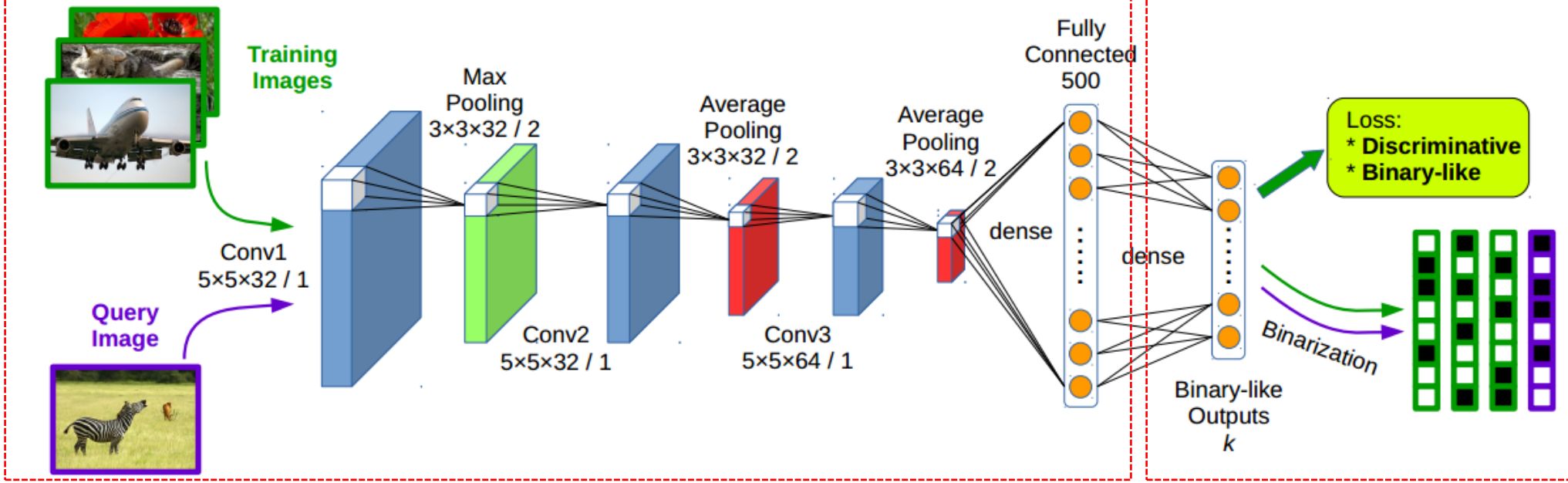
两个阶段：在哈希函数学习阶段，将特征库划分成训练集和测试集，在训练库上对构造的哈希函数集进行训练学习；在正式的哈希编码阶段，分别将原来的特征代入到学习得到的哈希函数集中，从而得到相应的哈希编码。



AI DISCOVERY

基于有监督的深度哈希编码学习

AI DISCOVERY



图像特征提取层　　哈希编码学习层

基于有监督的深度哈希编码学习框架主要包括图像特征提取层和哈希编码表示层，监督信息指导整个模型的训练。与传统哈希编码学习方法不同，深度哈希编码学习中的图像特征提取和哈希编码学习是同时进行的，并且互相影响。

典型图像分析任务

图像分割

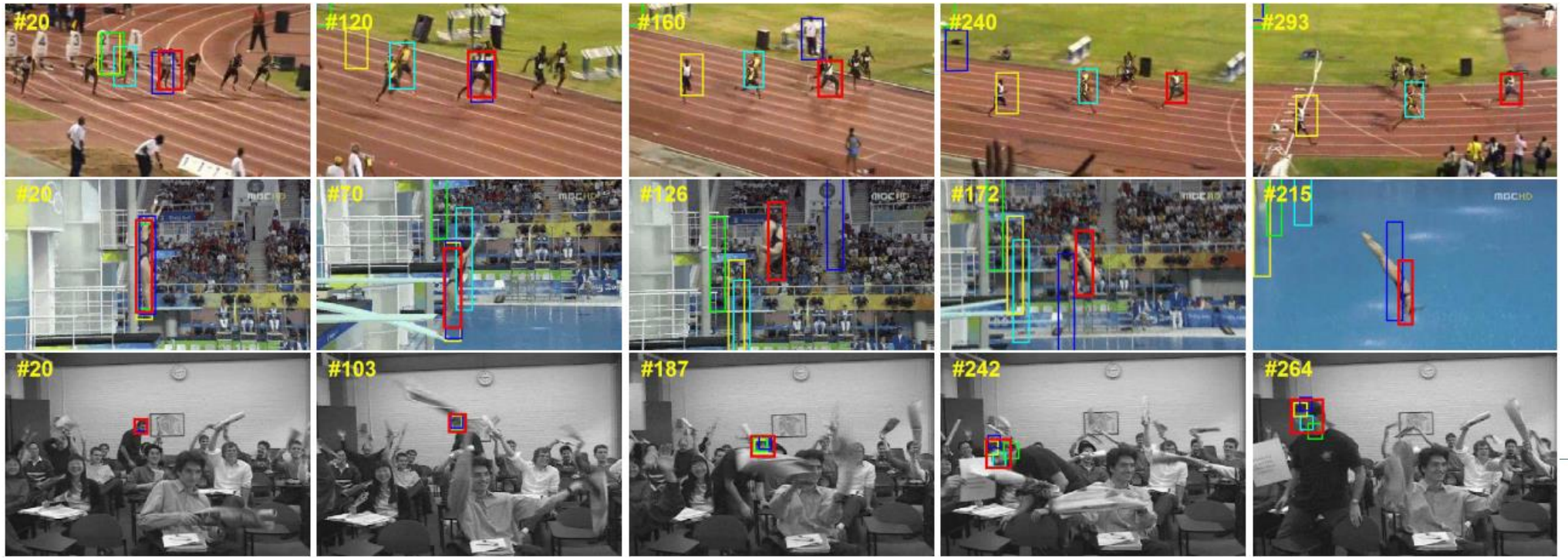
图像搜索

目标跟踪

目标跟踪

AI DISCOVERY

- ✓ 目标跟踪是计算机视觉研究领域的热点之一，并得到广泛应用。相机的跟踪对焦、无人机的自动目标跟踪等都需要用到了目标跟踪技术。
- ✓ 目标跟踪就是在连续的视频序列中，建立所要跟踪物体的位置关系，得到物体完整的运动轨迹。给定图像第一帧的目标坐标位置，计算在下一帧图像中目标的确切位置。



在线跟踪方法（MD-Net）

AI DISCOVERY

◆ 挑战：

- ✓ 不同跟踪序列跟踪目标不同，某类物体在一个序列中是跟踪目标，在另外一个序列中可能只是背景

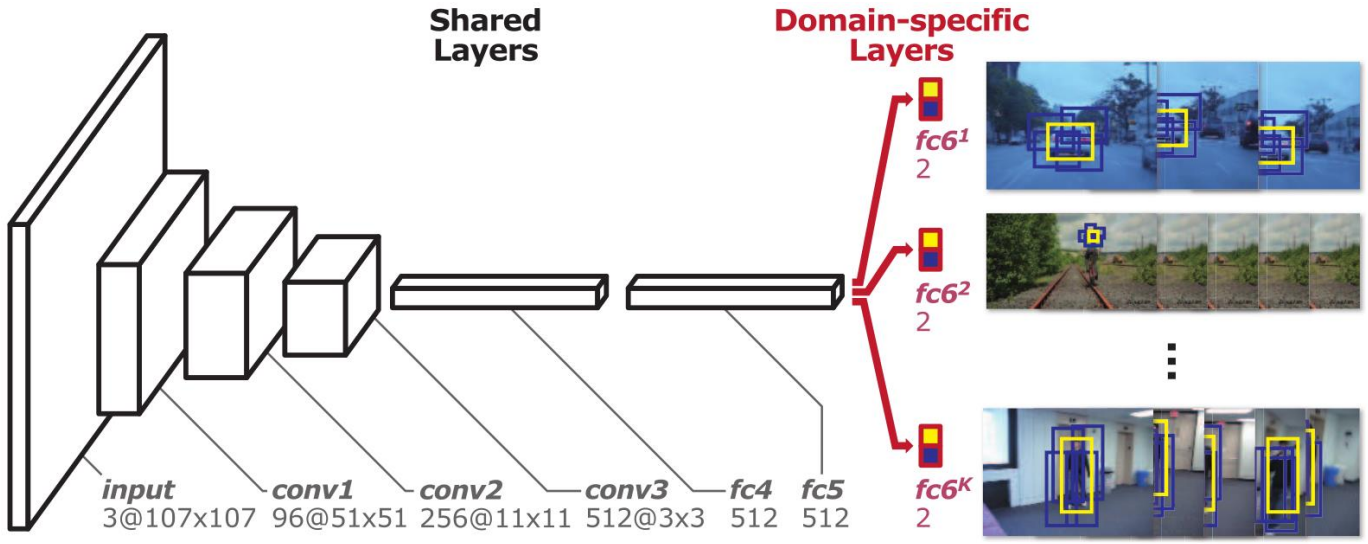
◆ 解决方案：

- ✓ 通过设计共享层和任务专有层，区分对待训练数据，构建具有通用目标表示能力的CNN主干网
- ✓ 每个训练序列被视为一个单独的任务，对应一个私有的二分类层（前景/背景）

◆ 训练阶段：

- ✓ 每个训练序列中提取的固定大小图像块，用来训练网络的共享层和自己私有的二分类层

在线表示在跟踪过程中，目标模板或网络参数会不断调整，离线则不会



◆ 在线跟踪阶段：

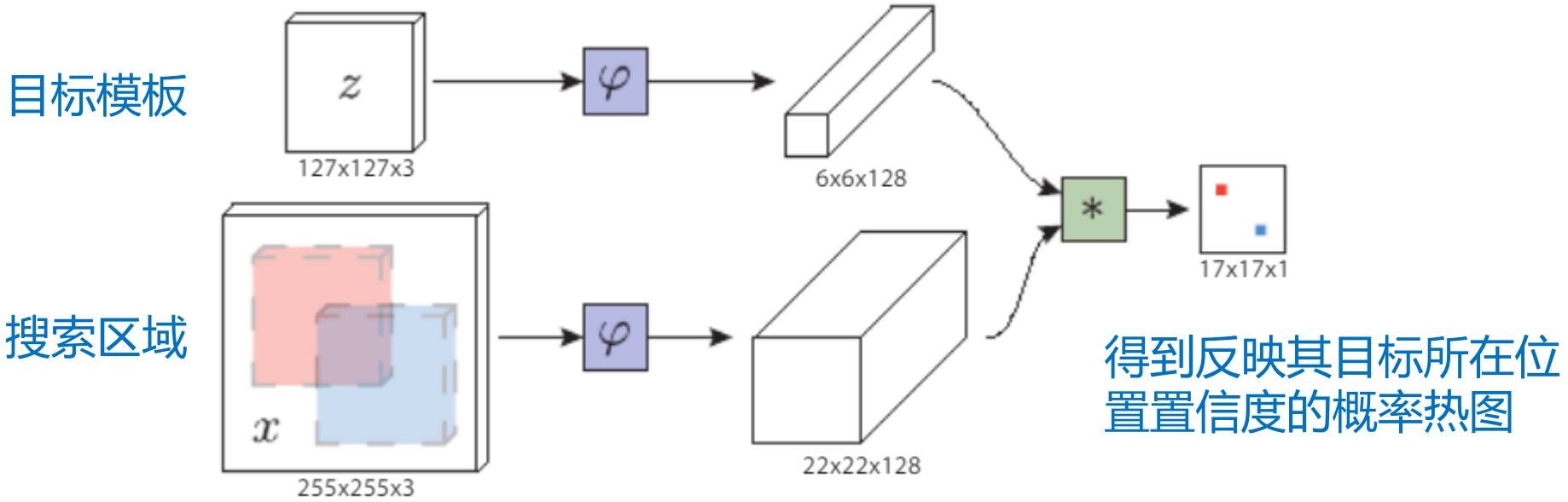
- ✓ 为任务新建一个fc6的二分类层
- ✓ 基于第一帧信息，在线更新全连接层（fc4-fc6）参数
- ✓ 提取多个候选跟踪区域输入网络，保留置信度最高的检测结果

AI DISCOVERY

Hyeonseob Nam, Bohyung Han. "Learning Multi-Domain Convolutional Neural Networks for Visual Tracking." CVPR 2016.

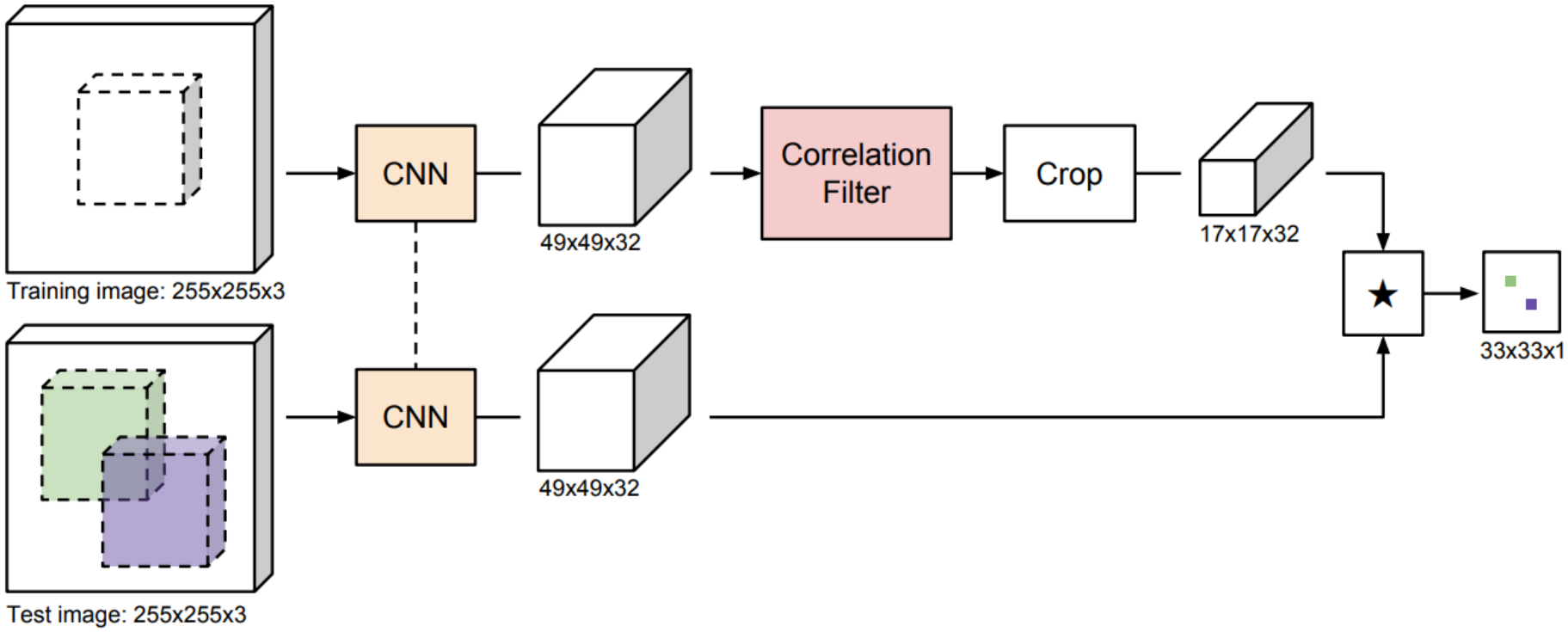
离线跟踪方法（Siamese Network）

AI DISCOVERY



- ✓ 算法本身是比较搜索区域与目标模板的相似度，最后得到搜索区域的概率热图
- ✓ z是模板图像，x是搜索图像（即当前帧图像，搜索当前帧图像里与z最匹配的区域），ϕ表示的是两个孪生的神经网络
- ✓ 使用6x6x128卷积核卷积22x22x128图像，得到一个关于区域的17x17的概率热图，选取得分最高的点作为目标区域的中心点

基于深度学习的目标跟踪CFnet



在**Siamese网络上加入了CF层（相关滤波）**，并且可以进行端到端的训练。

相对于在线方法，离线跟踪方法不需要一边测试一边finetune，速度快，但性能稍差，加入相关滤波则在保证速度的同时提升了性能。

目录

AI DISCOVERY

1 目标检测 — 两阶段方法、一阶段方法、最新进展

2 典型图像分析任务 — 图像分割、图像搜索、目标跟踪

3 特色图像分析任务 — 细粒度分类、风格迁移、标题生成、超分辨率

4 垂直应用与实践 — 医学影像分析、文字检测识别
实践：目标检测

AI DISCOVERY

特色图像分析任务

AI DISCOVERY

- 细粒度分类
- 风格迁移
- 标题生成
- 超分辨率生成

AI DISCOVERY

什么是细粒度图像分类?

AI DISCOVERY




Bird

Coarse-grained

Crested Auklet 海雀

Groove Billed Ani 沟嘴犀雀

Parakeet Auklet 长尾鹦鹉

Red Winged Blackbird 美洲红翼鸫

Fine-grained

AI DISCOVERY

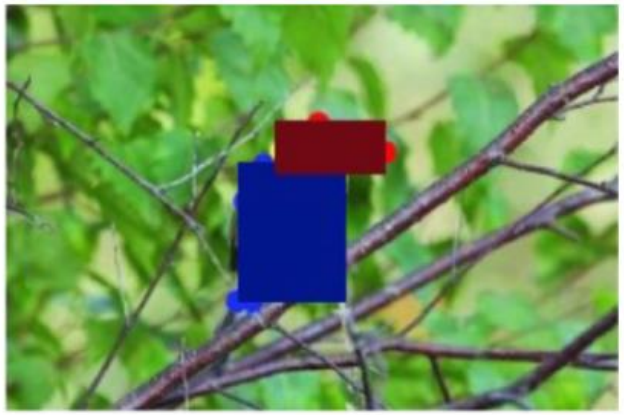
细粒度分类—Mask-CNN

AI DISCOVERY

鸟类数据集通常包含Part annotations用于标记特征点

Mask-CNN首先将Part annotations转换为bounding box形式（区分Head/Torso），并使用其作为GT mask，基于FCN预测Head/Torso的mask

(a) Part annotations



(b) Part rectangles

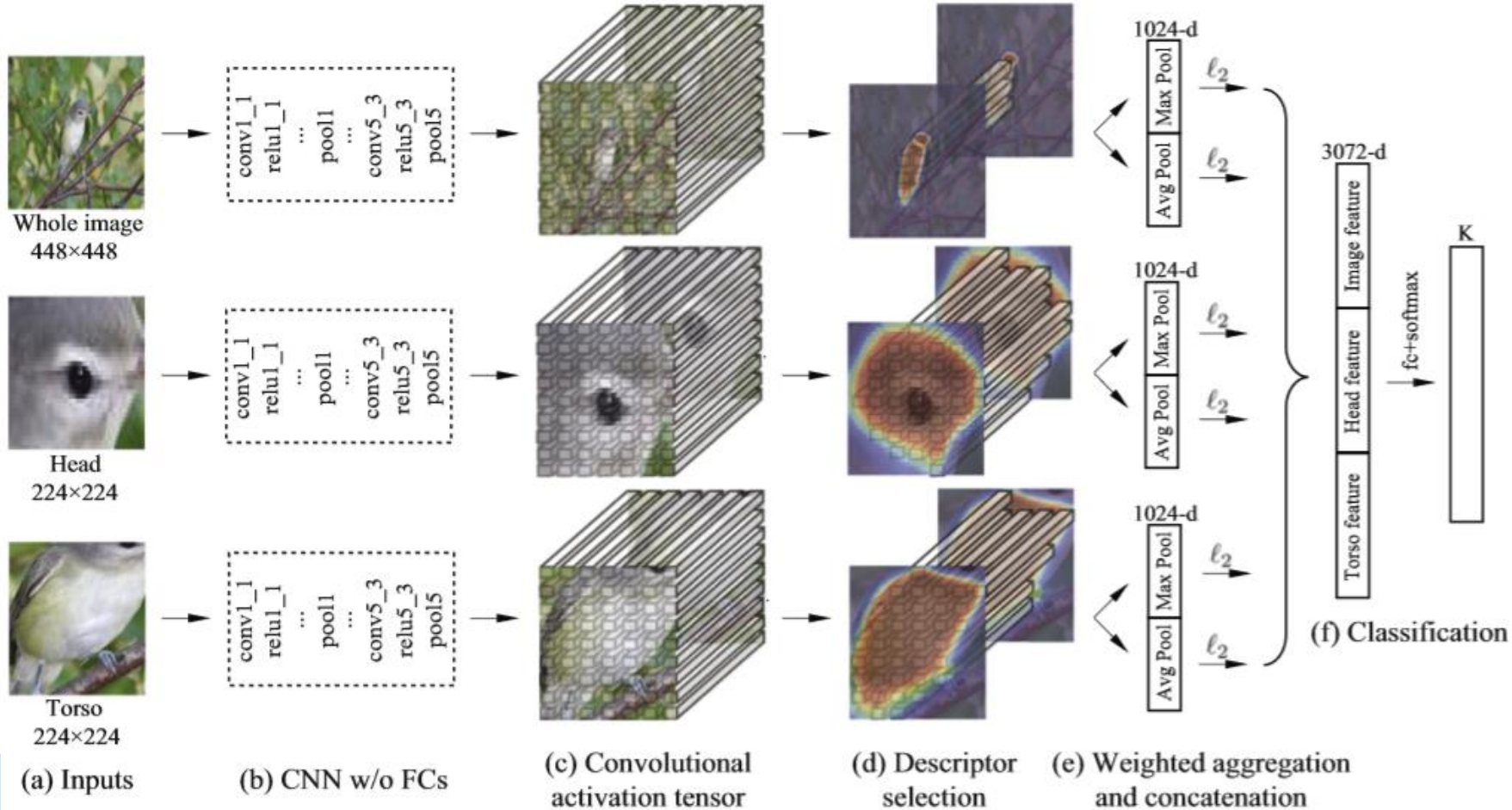
细粒度分类—Mask-CNN

AI DISCOVERY

Mask-CNN分为三个Stream，分别对应原图、Head、Torso区域图像。

不同的分支使用VGG分别提取卷积特征，并依据FCN预测出的Mask进行Descriptor selection，将保留下的特征进行池化和聚合，最终进行分类

需要额外的Part和Mask标注，不易获得，方法扩展性不高



AI DISCOVERY

Xiu-Shen Wei, Chen-Wei Xie, Jianxin Wu, Chunhua Shen “Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization” Pattern Recognition 2018

细粒度分类—RA-CNN

AI DISCOVERY

由于细粒度图像分类的标注代价高昂，RA-CNN使用弱监督的方法来逐层精化的挖掘注意力区域

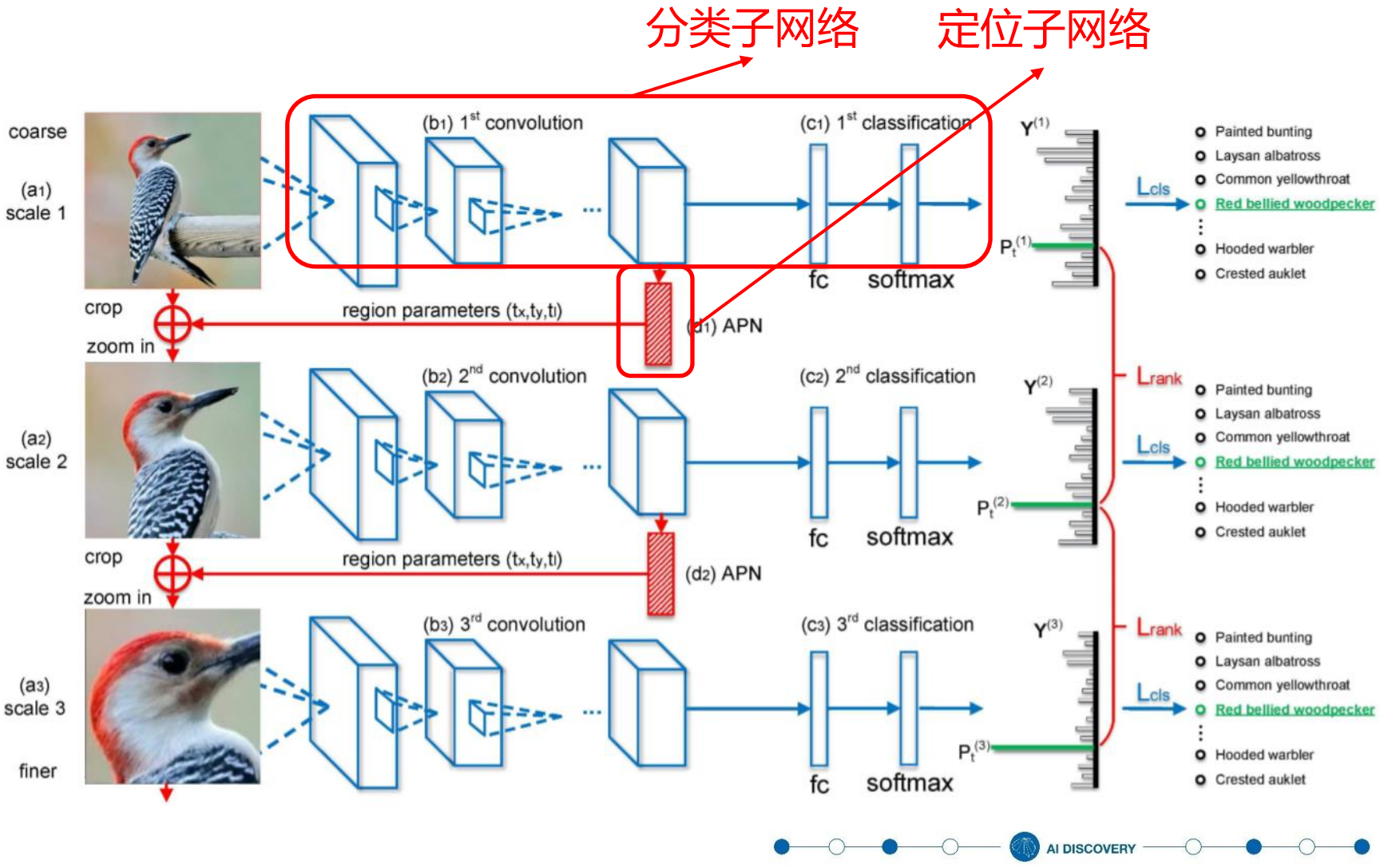
- Attention Proposal Network

tx,ty,tl表示注意力区域中心点及边长

- Rank Loss

$$L_{rank}\left(p_t^{(s)}, p_t^{(s+1)}\right) =$$

$$\max\{0, p_t^{(s)} - p_t^{(s+1)} + margin\}$$



AI DISCOVERY

Jianlong Fu, Heliang Zheng, Tao Mei “Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition” CVPR 2017

特色图像分析任务

细粒度分类

风格迁移

标题生成

超分辨率生成

图像风格迁移

AI DISCOVERY

那么对喜欢的绘画风格，怎么将其风格，搬到另外一张图片上呢？

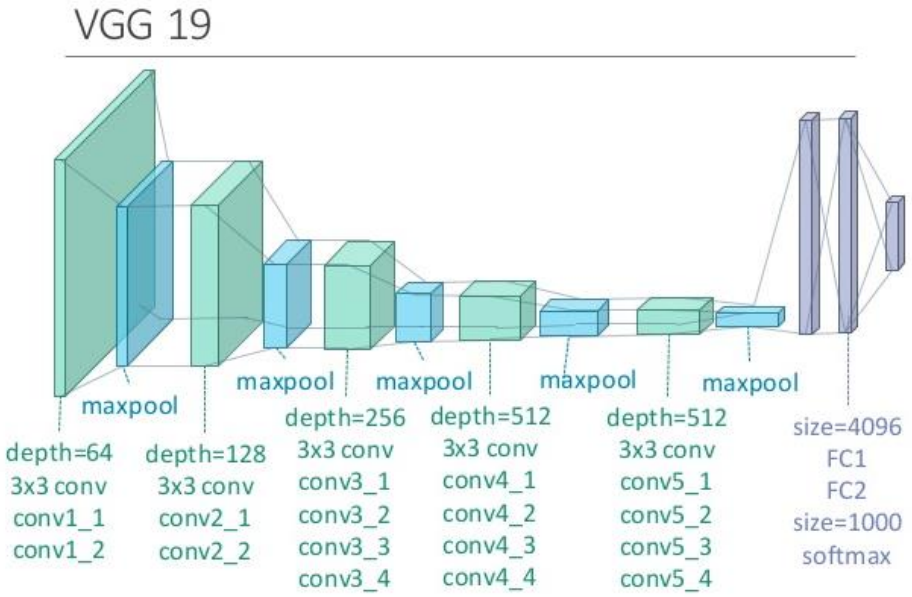
AI DISCOVERY

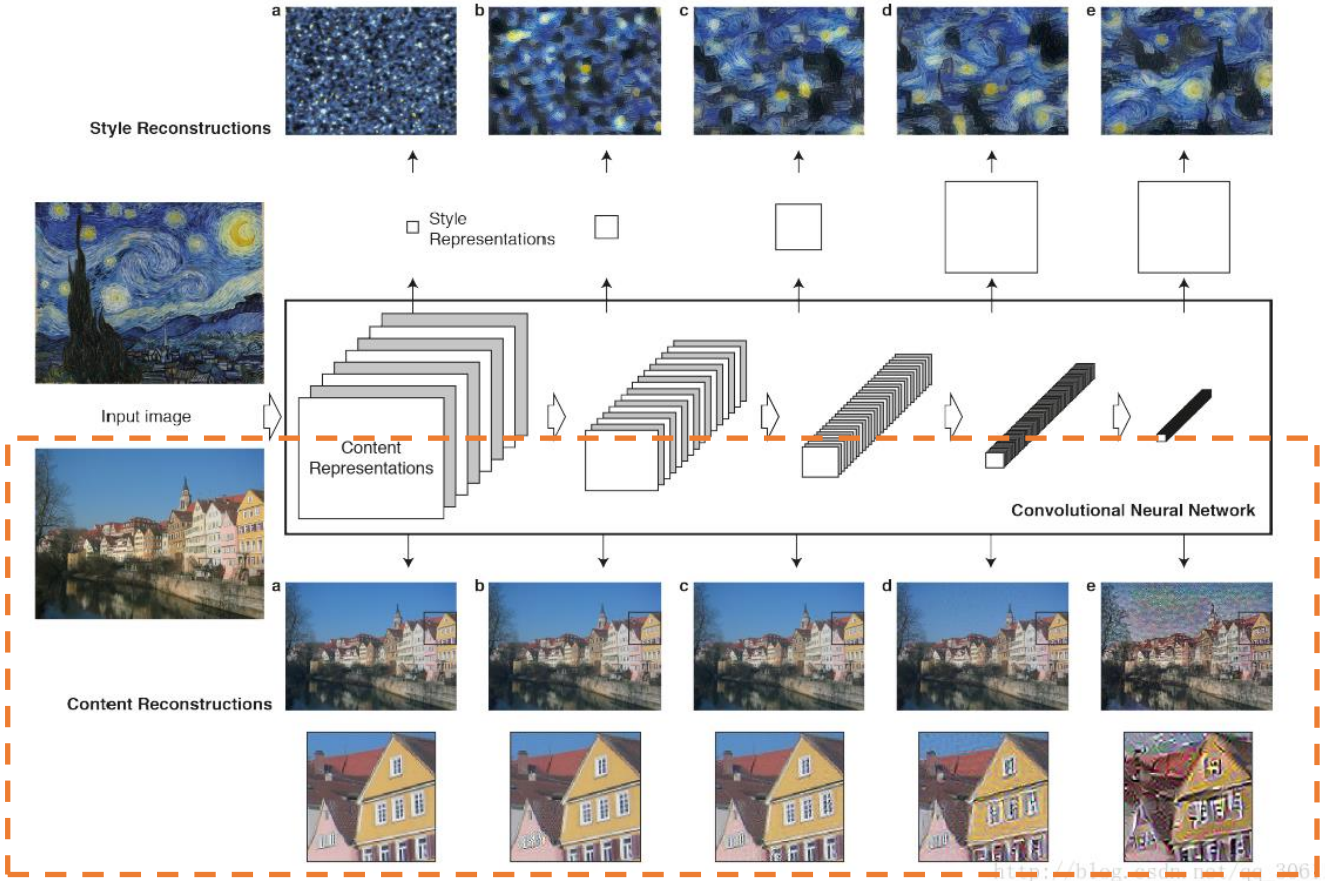
如何描述一张图的绘画内容

AI DISCOVERY

新生成的图要保留原本的内容大体不变，那就需要图片内容表示方法，并且定义图像间内容上的差距

feature-map本身就表示了图像在不同层上的内容，所以依然将其作为内容的表征形式





Leon A. Gatys,Alexander S. Ecker,Matthias Bethge “Image Style Transfer Using Convolutional Neural Networks” CVPR 2016

如何描述一张图的绘画风格

AI DISCOVERY

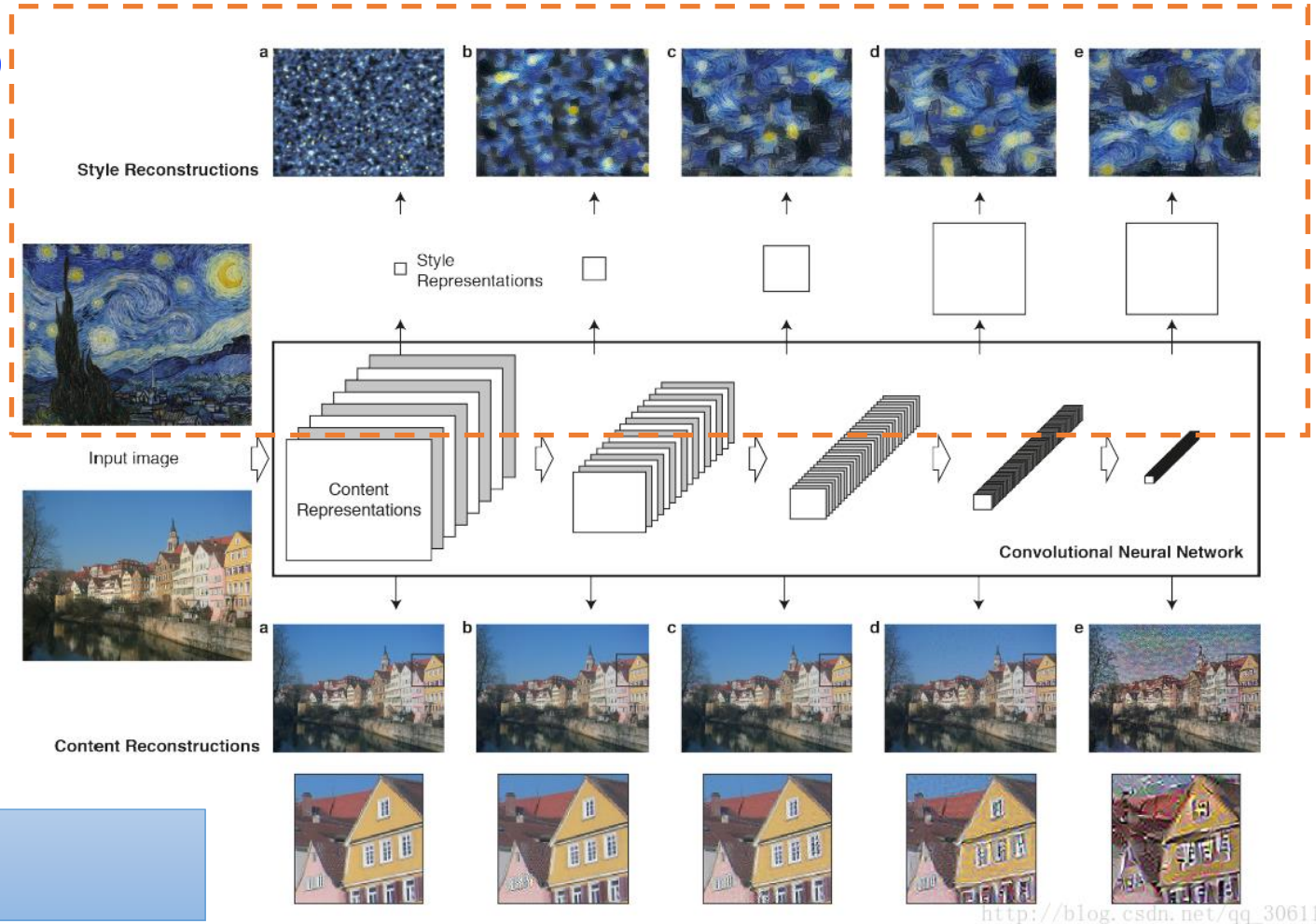
◆ texture representation

卷积网络的feature-map在分类任务中学习了图像的深层特征，这些feature-map如何表示图像的纹理特性呢？

对某一隐层，假设有N个channel的feature-map，每个map的size=height×width，那么每两个channel间都计算各inner-production，再累和作为本层的Texture表示矩阵，数学上称之为Gram matrix

$$G^l_{i,j} = \sum_k F^l_{i,k} F^l_{j,k}$$

*F*表示feature-map被展开成1-D的向量；*k*是向量index; *i,j*表示同层内的不同feature-map的channel-index；*l*表示当前所在的隐层



从左往右，层数越深，风格越明显。

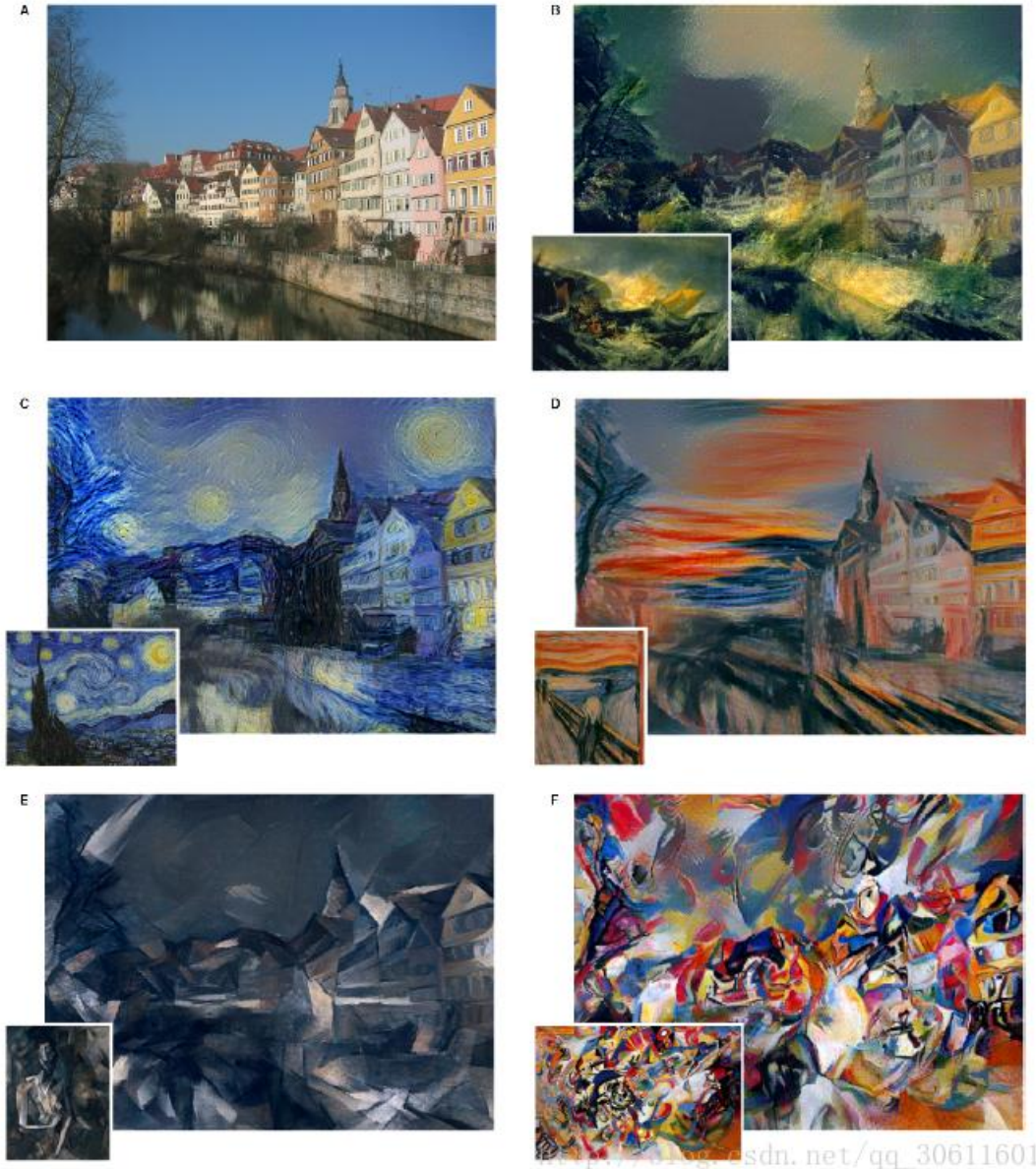
怎么权衡内容和风格

AI DISCOVERY

为了权衡内容和风格，整体优化目标如下：

$$L_{total} = \alpha L_{content} + \beta L_{style}$$

图像的内容和风格并不能被完全地分开。当风格与内容来自不同的两个图像时，被合成的新图像并不存在在同一时刻完美地符合了两个约束。但是，在图像合成中最小化的损失函数分别包括了内容与风格两者，所以，我们可以平滑地将重点既放在内容上又放在风格上。



如何得到风格转换后的图像

AI DISCOVERY

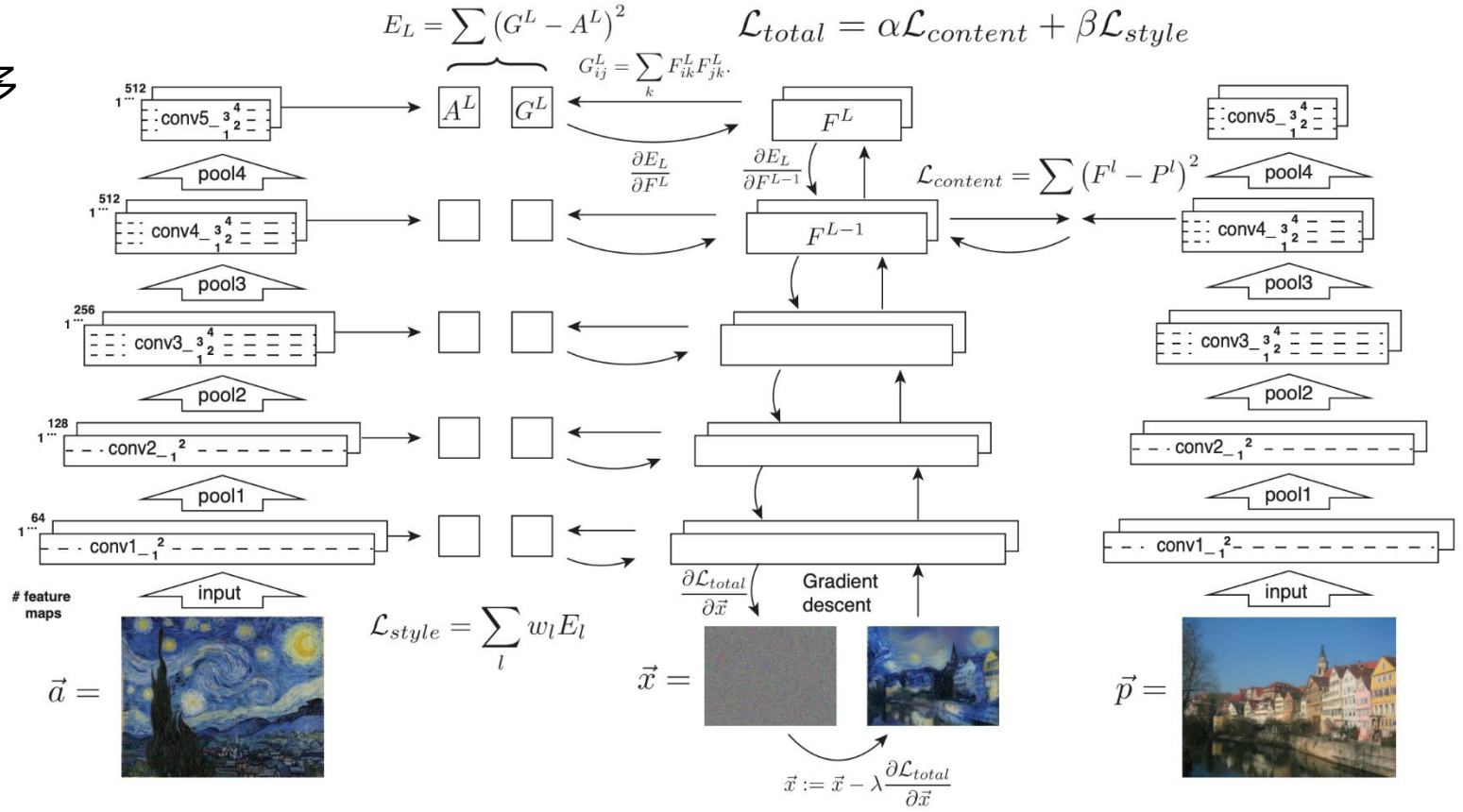
风格分支：

VGG的每个卷积层都可以得到很多feature map，每一层的特征都将被保存下来，与白噪声图像的相应feature map计算Gram matrix，得到风格损失

内容分支：

内容图像每个卷积层都会生成很多feature map，只保存conv4层的feature map，与白噪声图像计算两个feature map的均方误差，得到内容损失

从白噪声图像出发，生成了一张风格化图像！



AI DISCOVERY

特色图像分析任务

细粒度分类

风格迁移

标题生成

超分辨率生成

什么是图像标题生成

AI DISCOVERY

- ✓ 输入是一张图片，输出是一句对图片进行描述的文本，这就是图像标题生成
- ✓ 要为一张图片生成一句描述之前，首先要正确理解图片的主要内容和含义，在图片内容和语言文字之间建立语义关联

"Two people are walking down at river in a wooded area"

AI DISCOVERY

图像标题生成—最简版encoder-decoder

AI DISCOVERY

当前大多数的Image Caption方法基于encoder-decoder模型。其中encoder一般为卷积神经网络，利用最后全连接层或者卷积层的特征作作为图像的特征，decoder一般为递归神经网络，主要用于图像描述的生成。

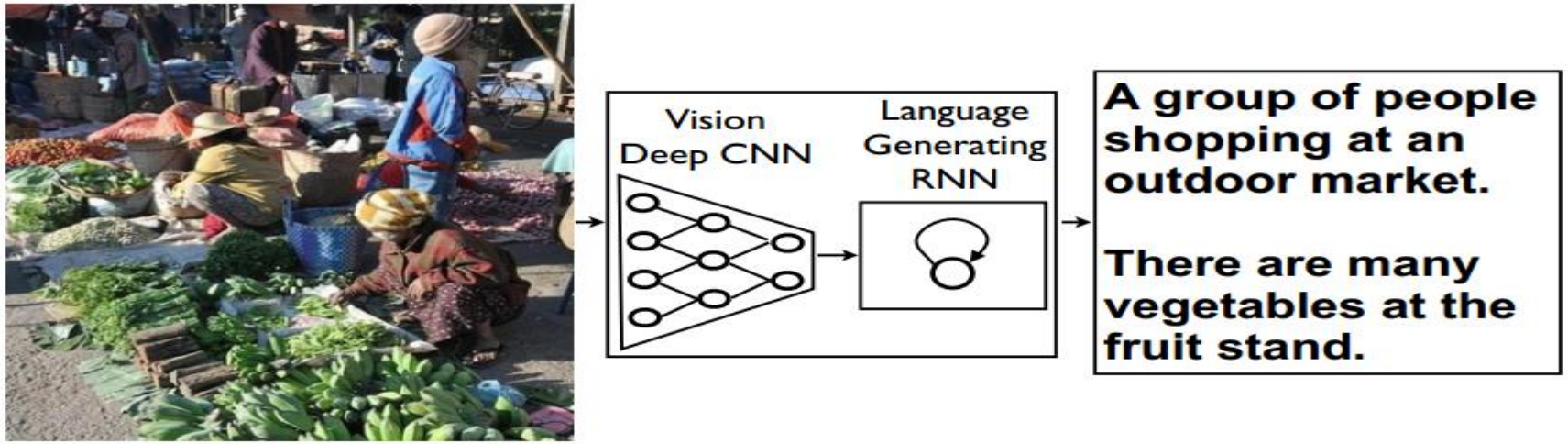


Figure 1. NIC, our model, is based end-to-end on a neural network consisting of a vision CNN followed by a language generating RNN. It generates complete sentences in natural language from an input image, as shown on the example above.

图像标题生成—MS Captivator

AI DISCOVERY

先通过目标检测+物体识别，把图像中的实体词都识别出来，实体词相关之间的连接词是构造完整句子的核心，然后采用Multiple Instance Learning(MIL)的弱监督方法进行造句。

步骤是：

- ✓ detect words：识别实体词
- ✓ generate sentences：生成句子
- ✓ re-rank sentences：重整句子结构

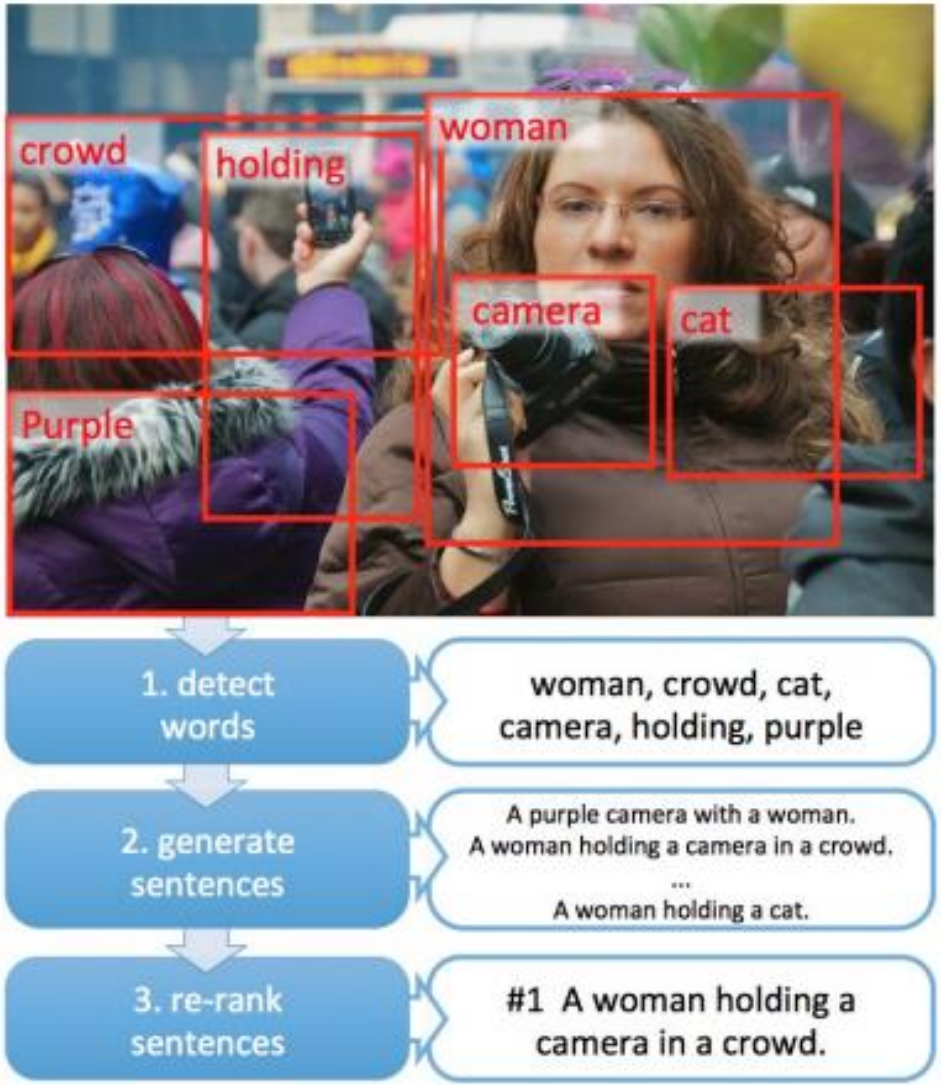


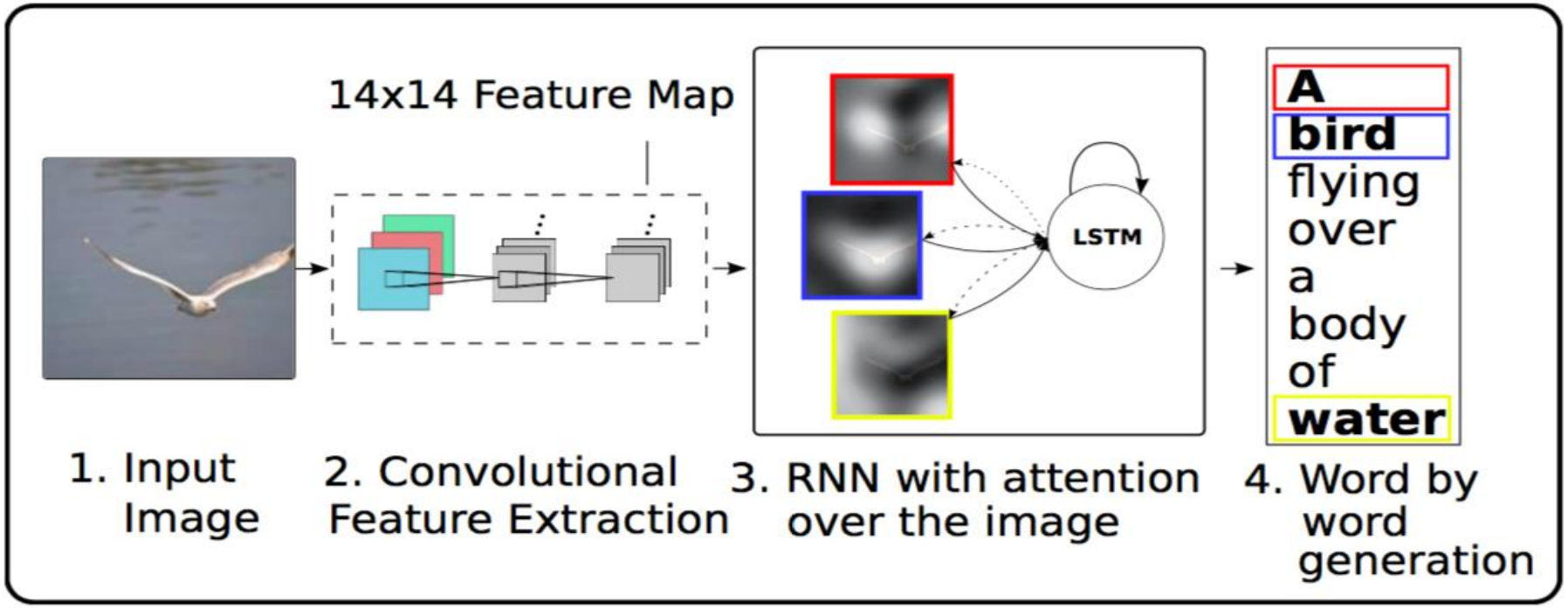
Figure 1. An illustrative example of our pipeline.

图像标题生成—基于注意力的模型

AI DISCOVERY

◆ 第一步：与直接通过全连接层提取特征不同，从某一个卷积层得到原始图片的表示，称为contexts；例如从VGG19的conv5_3拿到原始图片表示，shape为14*14*512，可以理解为将原始图片分成14*14共196个小块，每个小块对应一个512维的特征

◆ 第二步：根据contexts使用LSTM逐步生成单词，attention的加入，能够显著提高描述的性能。



特色图像分析任务

细粒度分类

风格迁移

标题生成

超分辨率生成

什么是图像超分辨率

AI DISCOVERY

bicubic
(21.59dB/0.6423)

SRResNet
(23.53dB/0.7832)

SRGAN
(21.15dB/0.6868)

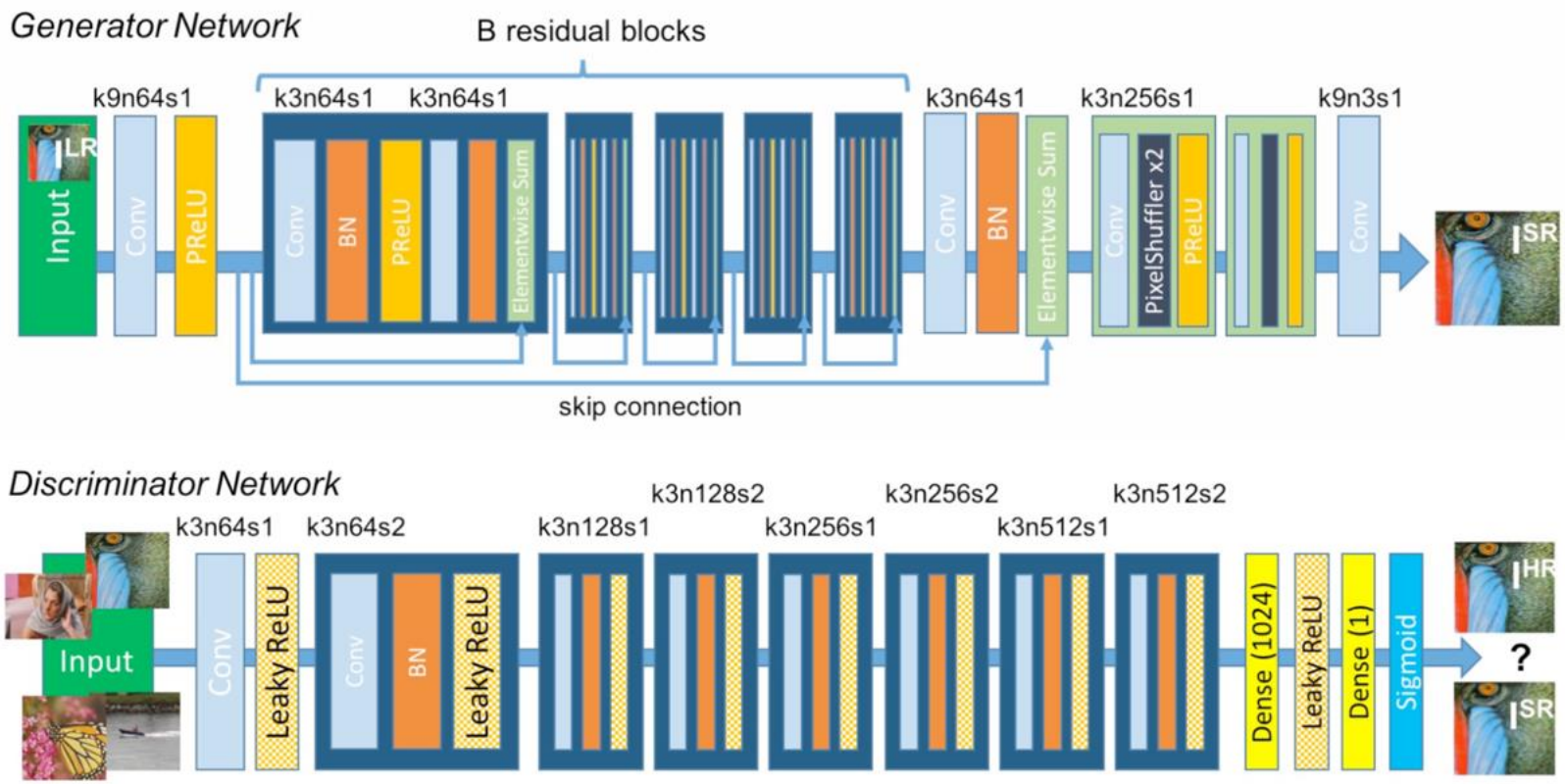
original

图像超分辨率是指由一幅低分辨率图像或图像序列恢复出高分辨率图像。

基于重建的图像超分辨率

AI DISCOVERY

SRGAN：



- ✓ 生成器将输入图像生成高分辨率图像。
- ✓ 判别器将判断输入图像是生成器生成的图像还是原始的高分辨率图像。
- ✓ 生成器和判别器通过交替迭代训练，最终使生成器生成更接近原始图像的超分辨率图像。

AI DISCOVERY

Christian Ledig, Lucas Theis, Ferenc Huszar “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network” CVPR 2017

基于重建的图像超分辨率

$$l_{MSE}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l_{VGG/i.j}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

$$\min_{\theta_G}\max_{\theta_D}\ \mathbb{E}_{I^{HR}\sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR}\sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))]$$

SRGAN：

生成器使用均方误差损失函数优化生成图像和高分辨率图像像素空间的最小均方差。

生成器同时使用感知损失（vgg）优化生成图像和高分辨率图像特征表示的欧氏距离

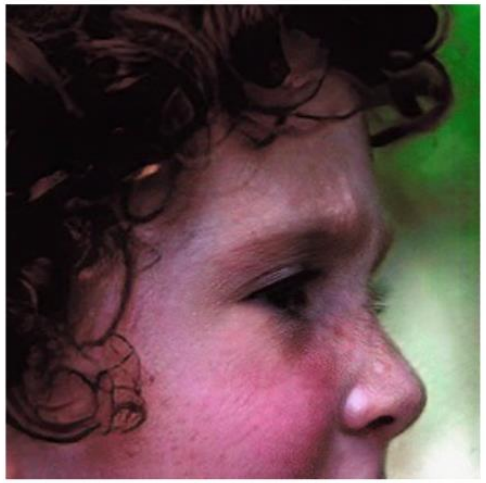
生成器和判别器使用对抗损失（生成器和判别器的交替迭代）训练整个网络。

基于重建的图像超分辨率

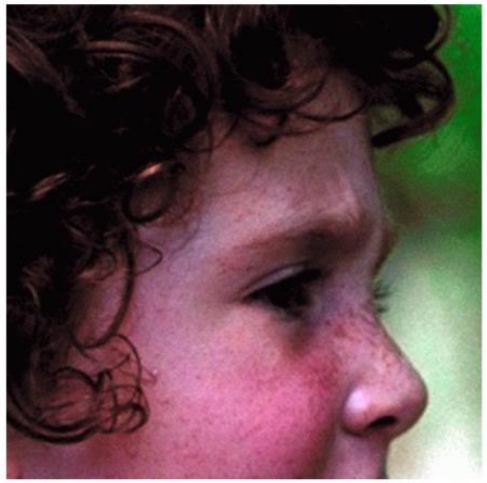
bicubic

SRResNet

SRGAN



original



相比于双三次插值、SRResNet等网络，SRGAN网络重建的超分辨率图像更好的视觉效果，更好的纹理细节。

目录

垂直应用与实践

AI DISCOVERY

医学影像分析

文字检测识别

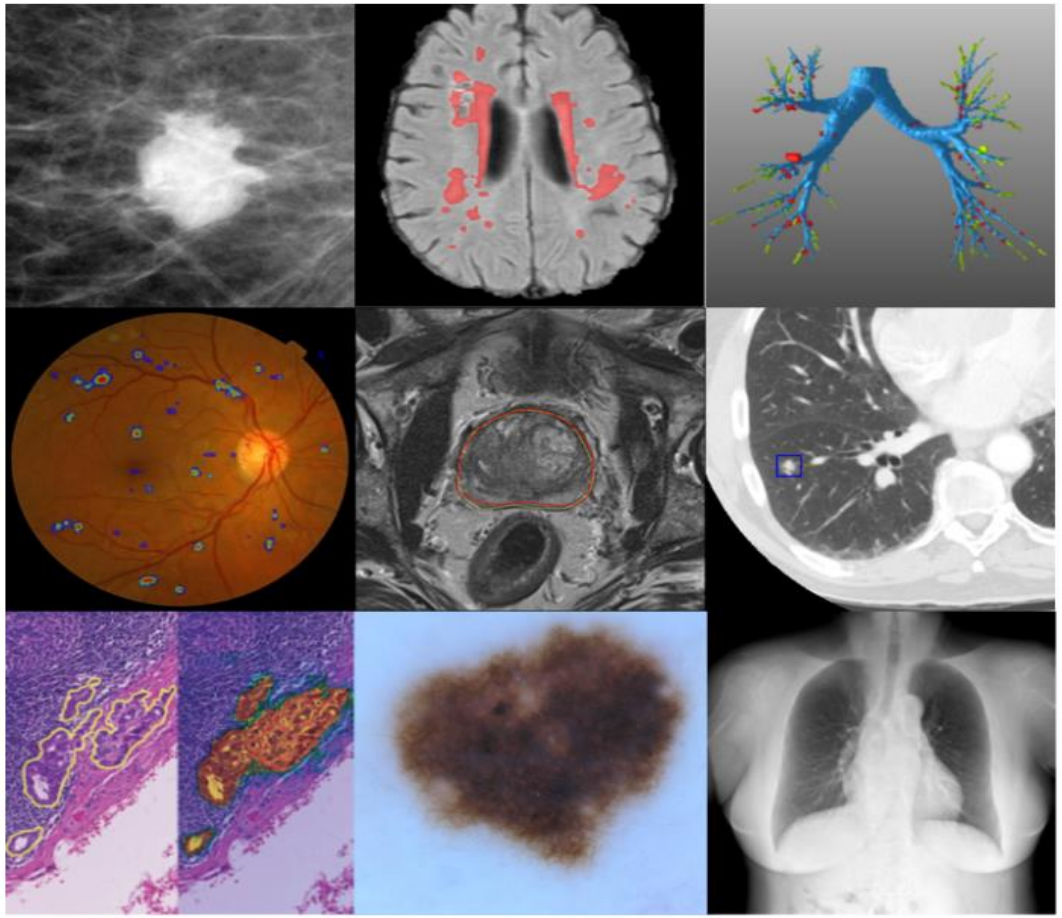
实践：目标检测

AI DISCOVERY

医学影像分析

AI DISCOVERY

- ✓ 右图展示了目前深度学习常规任务如分类、检测、分割在医学影像领域的应用。
- ✓ 例如，乳腺肿块分类任务，脑损伤部位的分割，气道树病变部位分割检测，糖尿病视网膜病变分类，前列腺分割，结节分类，淋巴结中的乳腺癌转移检测，皮肤病变分类，x射线骨髓抑制检测等。
- ✓ 由此可见，深度学习技术在医学影像分析领域应用十分广泛。



AI DISCOVERY

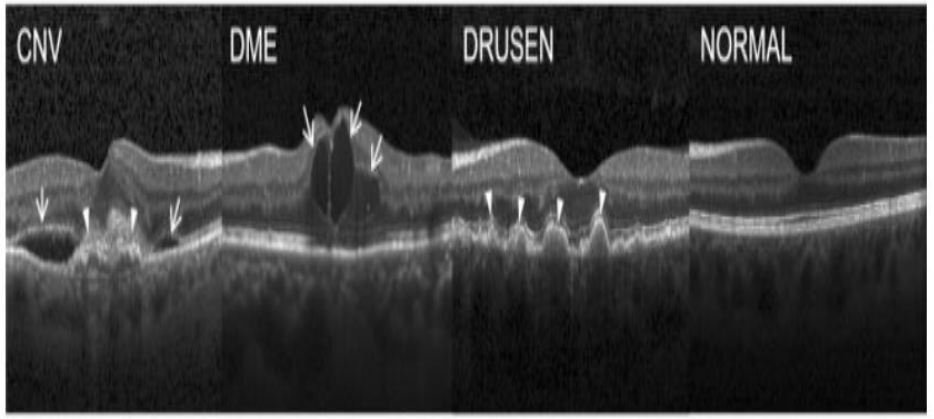
实例-眼底病变分类

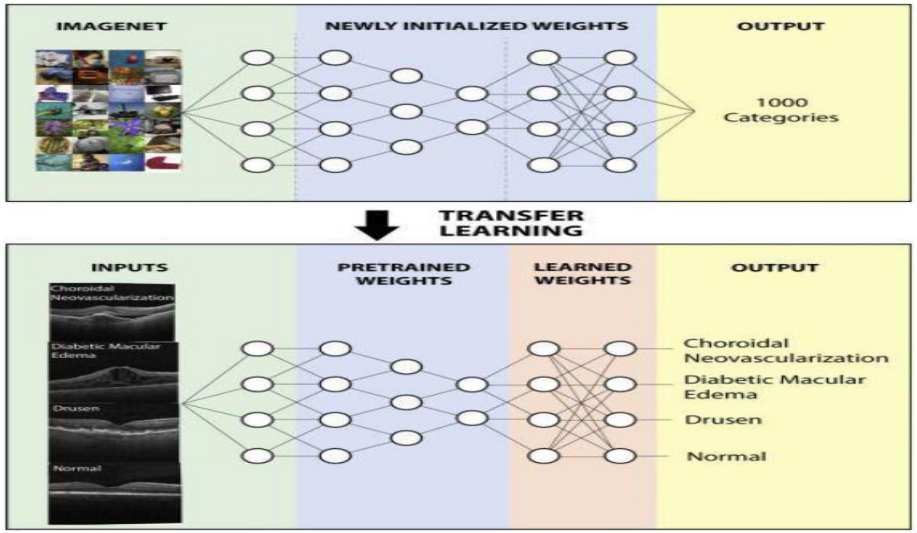
AI DISCOVERY

◆ 背景：

- ✓ 在美国，近1000万人患有年龄相关性黄斑变性（AMD）,近75万名年龄在40岁或以上的人患有糖尿病性黄斑水肿, 视网膜OCT图像对于指导治疗方案至关重要。上图所示是脉络膜新生血管（CNV）,糖尿病性黄斑水肿（DME）,和玻璃疣（DRRUSEN）与正常，四种情况下的OCT图像。

2018年，加州大学圣迭戈分校的张康课题组在顶级期刊Cell上发表了用于精确诊断致盲性视网膜疾病的论文，实现了图像精准分类并且荣登封面。





AI DISCOVERY

Kermany D S , Goldbaum M , Cai W , et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning[J]. Cell, 2018, 172(5):1122–1131.e9.

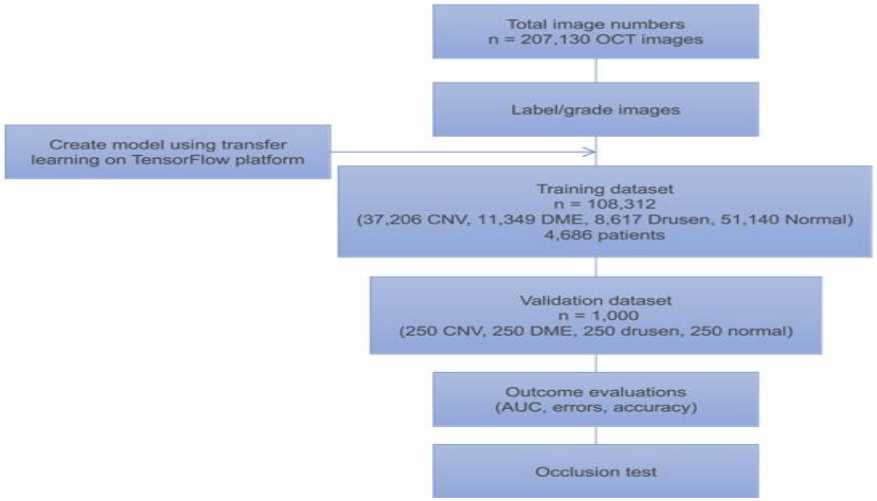
实例-眼底病变分类

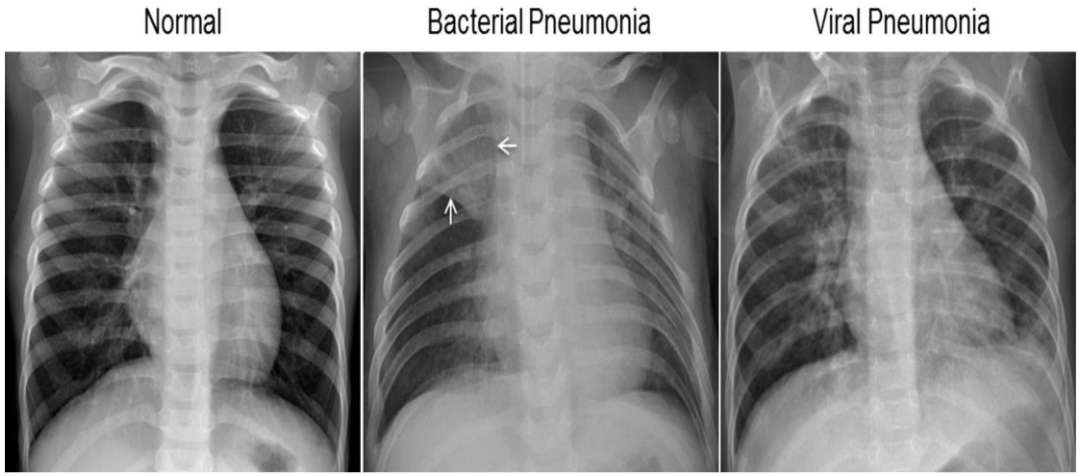
AI DISCOVERY

◆ 方法流程：

- ✓ 采集OCT图像数据，并请专家标注病变类型
- ✓ 训练CNN分类模型并且在验证集上计算相关测评指标。

将该方法与人类专家的观察结果进行对比，发现在诊断眼底病变分类的评估指标方面，文章提出的方法表现比人类专家更精确并且稳定。此外采用迁移学习方法，该方法对肺炎以及胸片的鉴别准确度也高达**90%以上**。





AI DISCOVERY

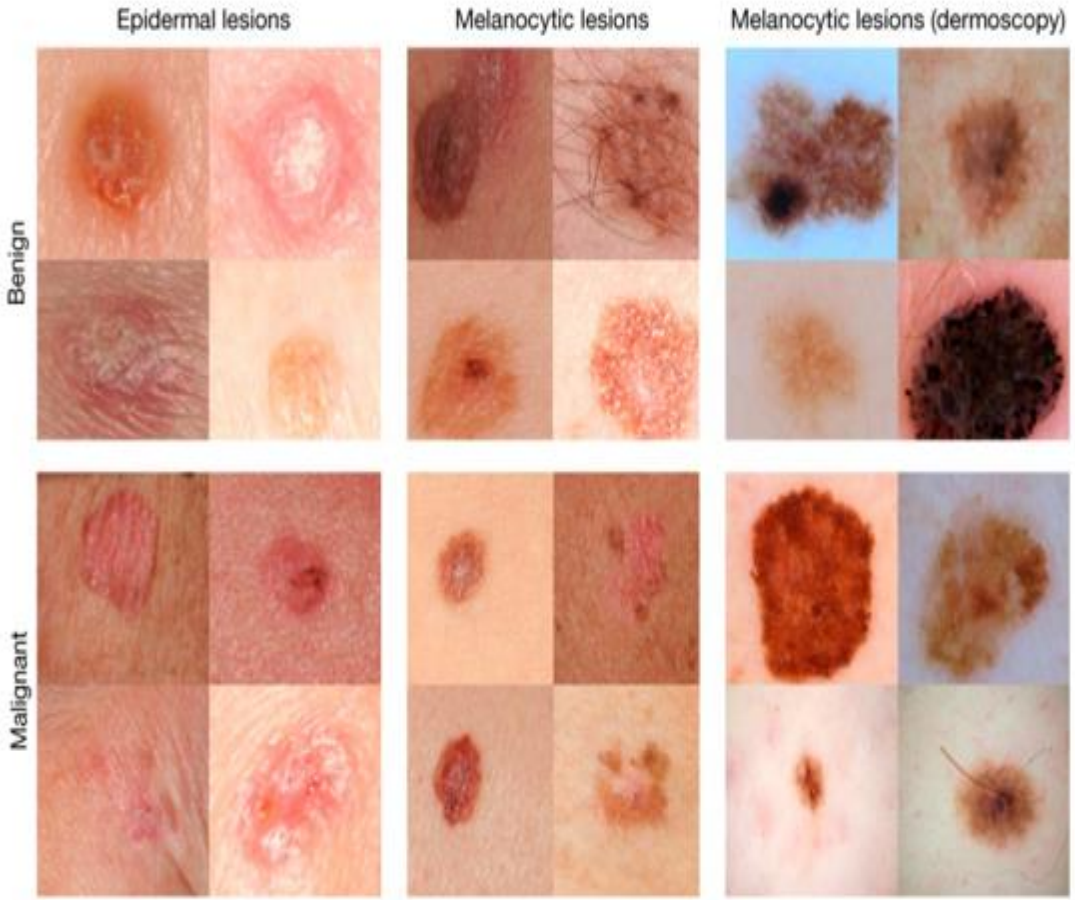
Kermany D S , Goldbaum M , Cai W , et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning[J]. Cell, 2018, 172(5):1122–1131.e9.

实例-皮肤病变分类

AI DISCOVERY

◆ 背景：

- ✓ 美国有540万新发皮肤癌病例2每年。五分之一的美国人将在其一生中被诊断出患有皮肤恶性肿瘤。虽然黑色素瘤占美国所有皮肤癌的不到5%，但它们占所有皮肤癌相关死亡的约75%，并且仅在美国每年造成10,000多人死亡。早期检测至关重要，因为如果在最早阶段检测到，黑素瘤的估计5年生存率从超过99%下降到最近阶段检测到的约14%。
- ✓ 右图中上半部分是良性病变下半部分是恶性病变，分别是上皮细胞的病变，黑色素细胞的病变以及皮肤镜下黑色素细胞的病变。Andre Esteva团队，基于 GoogleNet Inception v3 设计了分类器，其准确程度也要比皮肤科医生优越，论文发表在顶级期刊Nature上。

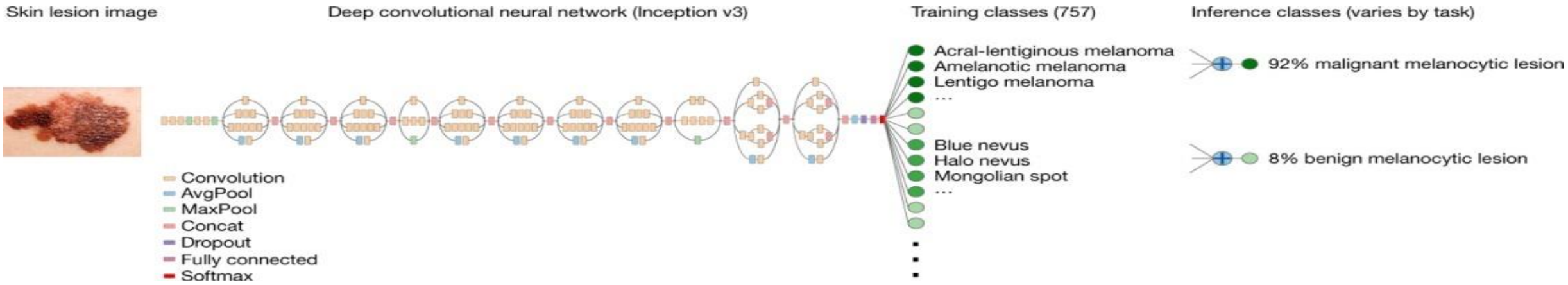


AI DISCOVERY

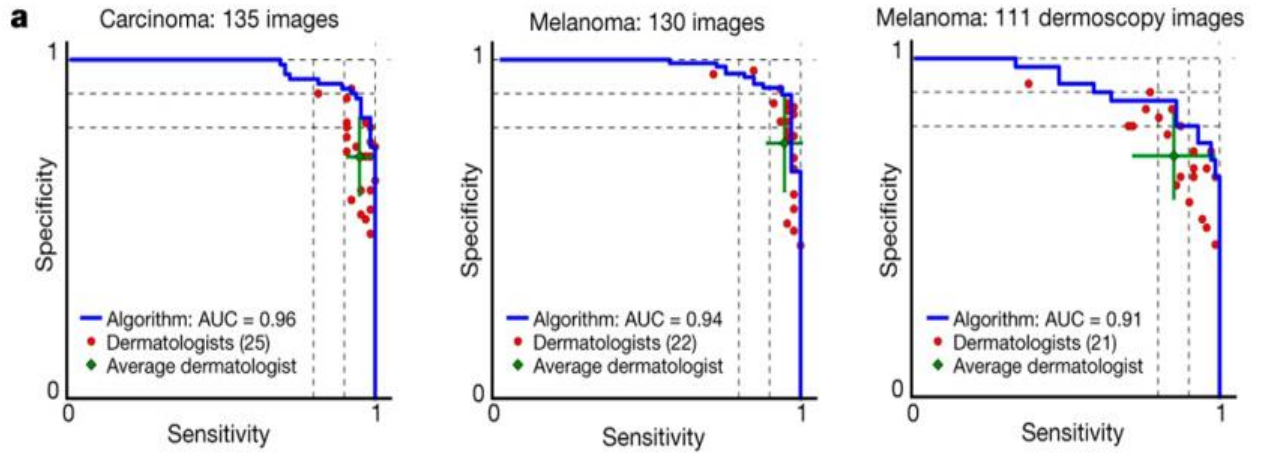
Esteva A , Kuprel B , Novoa R A , et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639):115-118.

实例-皮肤病变分类

AI DISCOVERY



AUC是CNN对性能的度量，最大值为1。如果皮肤科医生的敏感性-特异性点位于蓝色曲线以下(大多数情况下都是这样)，CNN的性能将优于皮肤科医生，可以直观看到文章提出的方法表现比人类专家更精确。



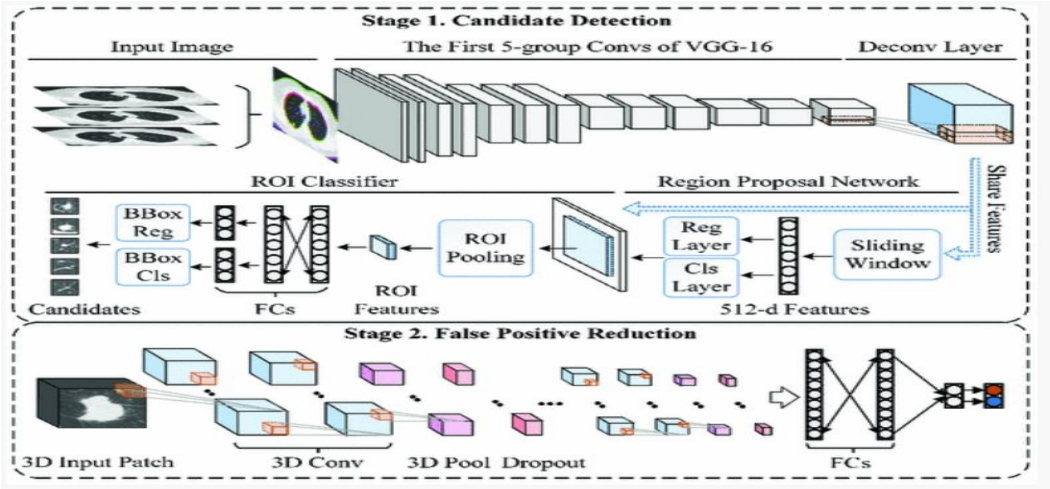
AI DISCOVERY

实例-肺结节检测

AI DISCOVERY

◆ 背景：

早期发现肺癌是提高患者生存机会的最有希望的方法。计算机断层扫描（CT）图像中准确的肺结节检测是诊断肺癌的关键步骤。北京大学王立威教授指导的LAB2112团队获得了决赛总冠军，并将他们的方法发表在在MICCAI。



◆ 方法简介：

贾丁等人在阿里天池肺结节检测比赛中提出了一种基于深层卷积神经网络（DCNN）的新型计算机辅助检测（CAD）系统，用于精确的肺结节检测。在所提出的CAD系统中，首先将反卷积结构引入faster-rcnn，用于轴向切片上的候选区域检测。然后，使用三维DCNN（3D DCNN）减少假阳性的出现。达到的效果与相关方法的对比也有明显的提升。

System	Sensitivity
ISICAD	0.856
SubsolidCAD	0.361
LargeCAD	0.318
M5L	0.768
ETROCAD	0.929
Baseline(w/o deconv)	0.817
Baseline(4 anchors)	0.895
Ours	**0.946**

AI DISCOVERY

Ding J , Li A , Hu Z , et al. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks[J]. 2017.

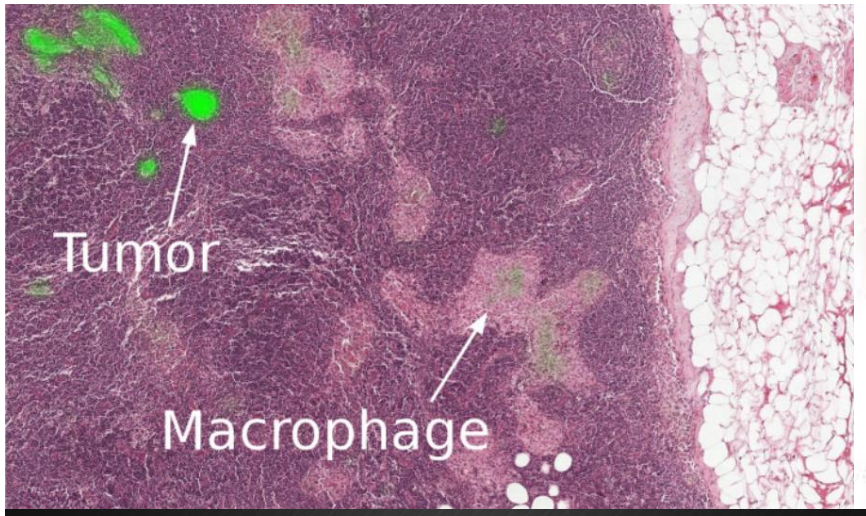
乳腺癌病理转移检测

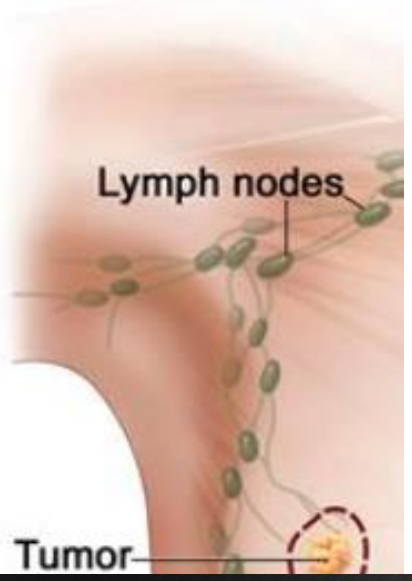
AI DISCOVERY

Camelyon16/17挑战赛

◆任务：

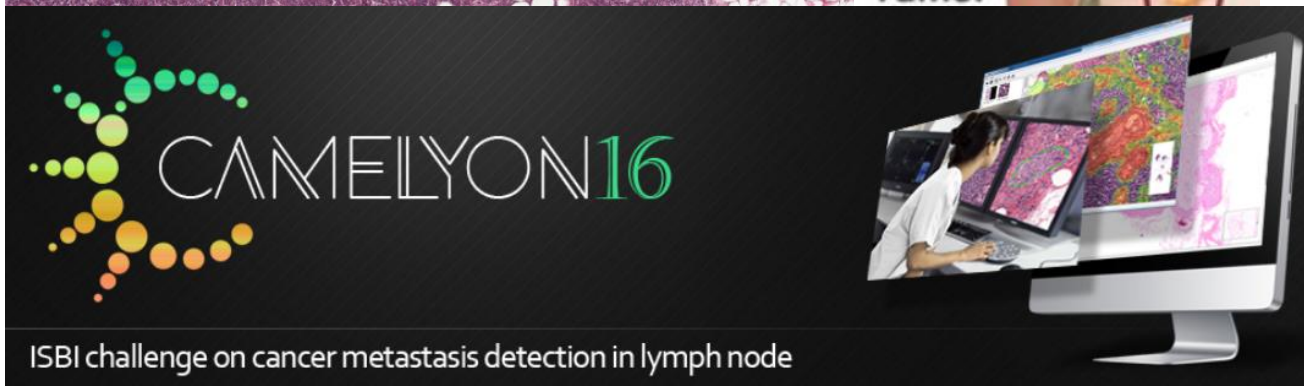
- ✓ 根据全片扫描图像H&E染色(WSI)，判断乳腺癌是否转移
- ✓ 定位出乳腺癌向邻近淋巴结扩散的位置





◆数据集：

- ✓ Camelyon16 数据集共 400张 WSI
  - 训练集：270张
  - 测试集：130张
  - 数据集大小：700GB
- ✓ Camelyon17 数据集共 500张 WSI
  - 训练集：50张
  - 测试集：500张
  - 数据集大小：2.25TB



一张图像非常大！

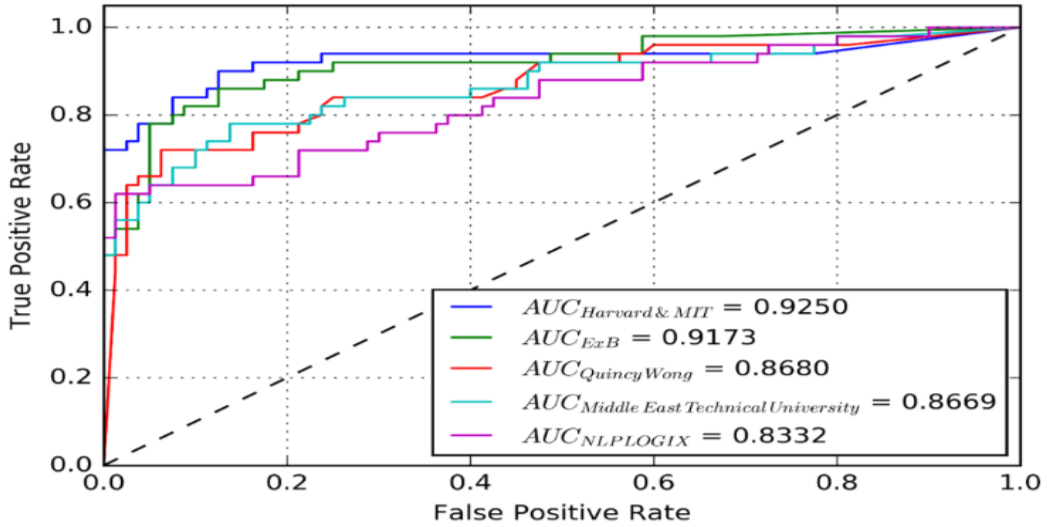
乳腺癌病理转移检测

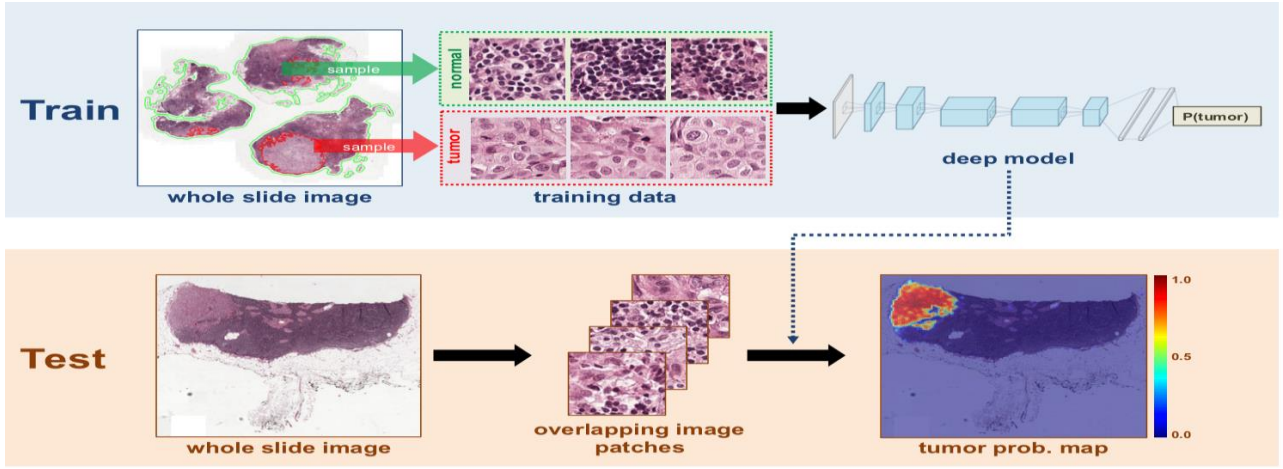
AI DISCOVERY

哈佛医学院采用基于深度学习的方法赢得冠军

◆方法：

- ✓训练卷积神经网络对小图像patch分类（tumor vs normal）
- ✓整合patch级预测结果构建肿瘤可能性热图
- ✓预测全片和定位肿瘤区域





将病理学家诊断和深度学习系统预测结果相结合，可以将病理学家的AUC得分提升至0.995，这意味着可以减少将近85%的人判别错误率

AI DISCOVERY

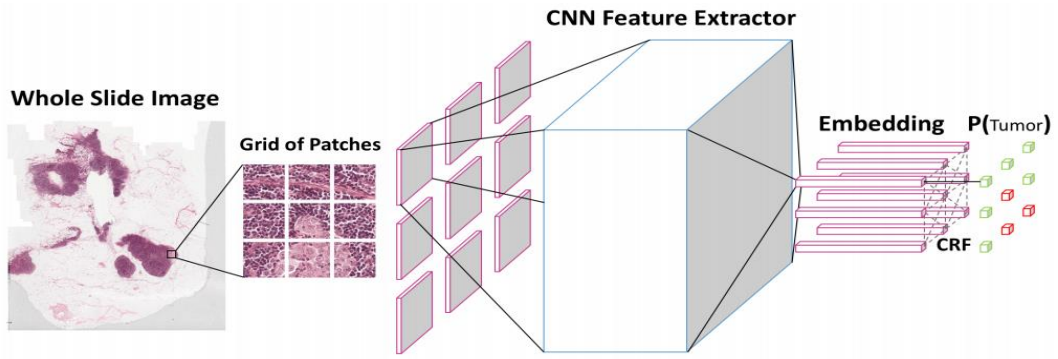
Dayong Wa et al, Deep Learning for Identifying Metastatic Breast Cancer, 2016

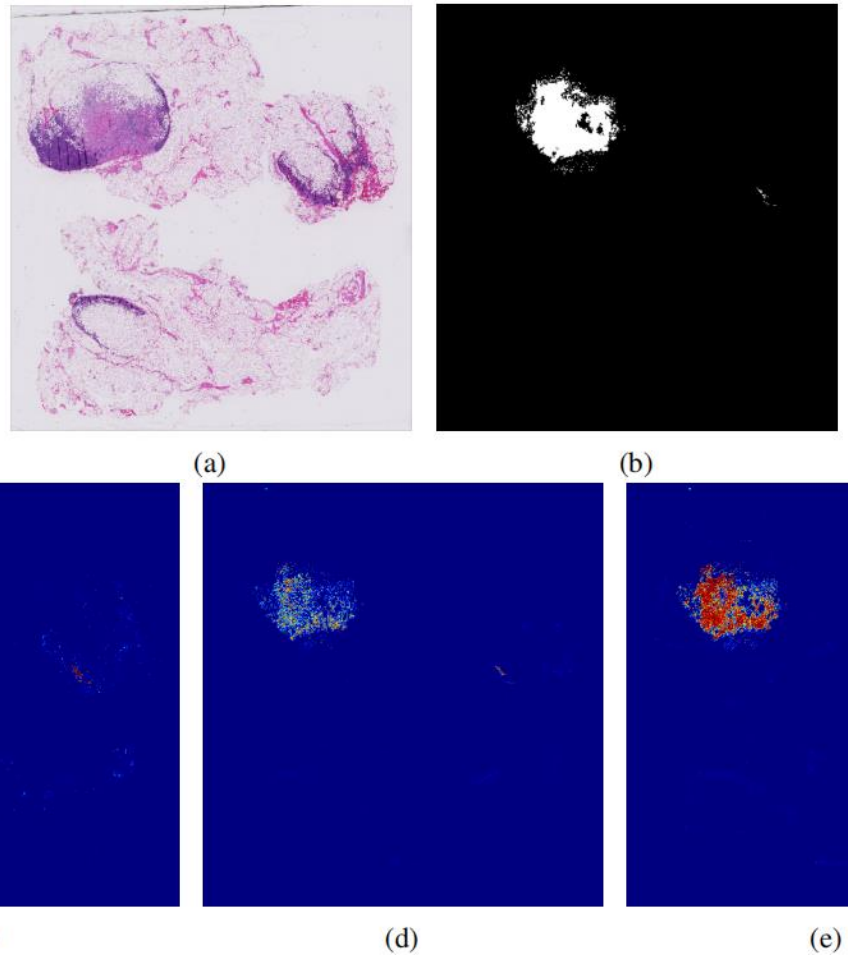
乳腺癌病理图像转移检测

AI DISCOVERY

◆2018年百度提出NCRF

- ✓基于resnet18和resnet34架构
- ✓将卷积神经网络和条件随机场（CRF）结合
- ✓ 以建模相邻patch之间的空间相关性





与不考虑空间相关性的基线方法相比，NCRF获得具有更好视觉质量的patch预测的可能性热力图。同时，NCRF方法在Camelyon16数据集上的癌症转移检测中优于baseline

AI DISCOVERY

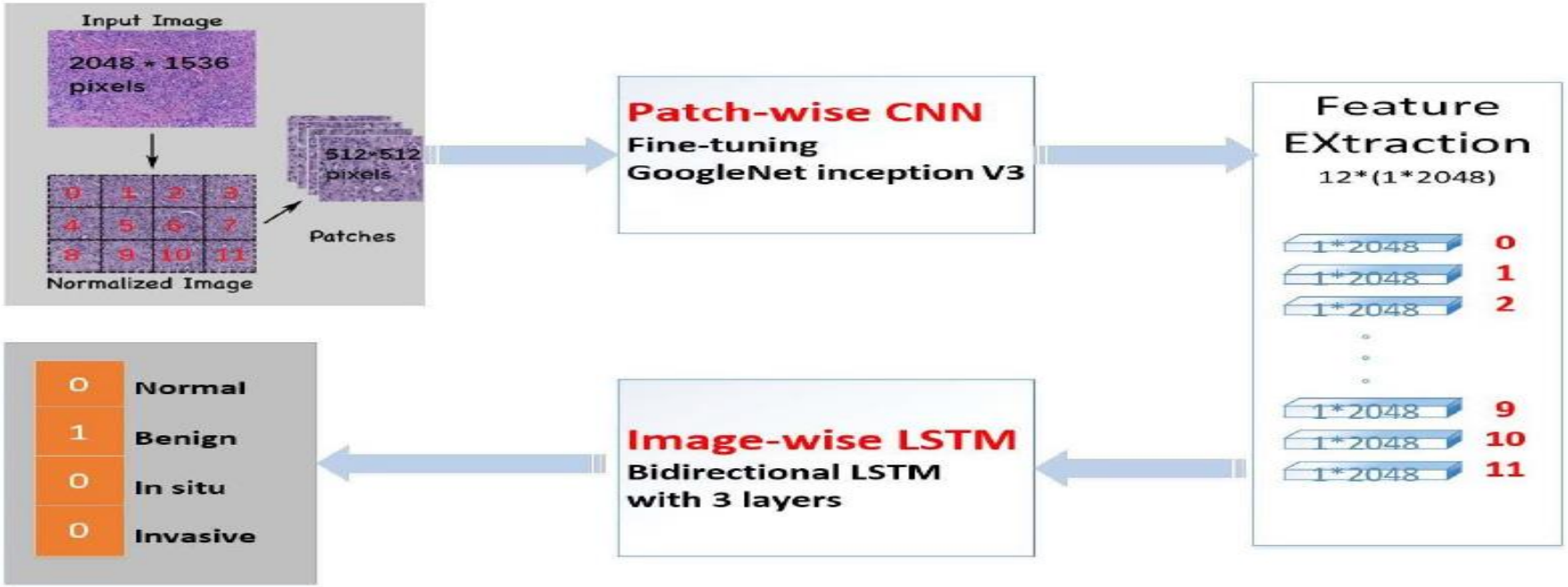
Li Y , Ping W . Cancer Metastasis Detection With Neural Conditional Random Field. 2018.

病理图像良恶性判别和转移检测

AI DISCOVERY

◆ 2018年中科院计算所提出CNN+LSTM的方案

- ✓提出混合Ineption-V3和LSTM的架构
- ✓将CNN提取的特征向量送入3层双向LSTM
- ✓同时发布了一个包含1568个乳腺癌病理图像　的更大数据集



提出的方法得到的四分类任务的平均准确率为90.5%

AI DISCOVERY

Rui Yan et al. A Hybrid Convolutional and Recurrent Deep Neural Network for Breast Cancer Pathological Image Classification. 2018.

垂直应用与实践

AI DISCOVERY

医学影像分析

文字检测识别

实践：目标检测

AI DISCOVERY

文字检测

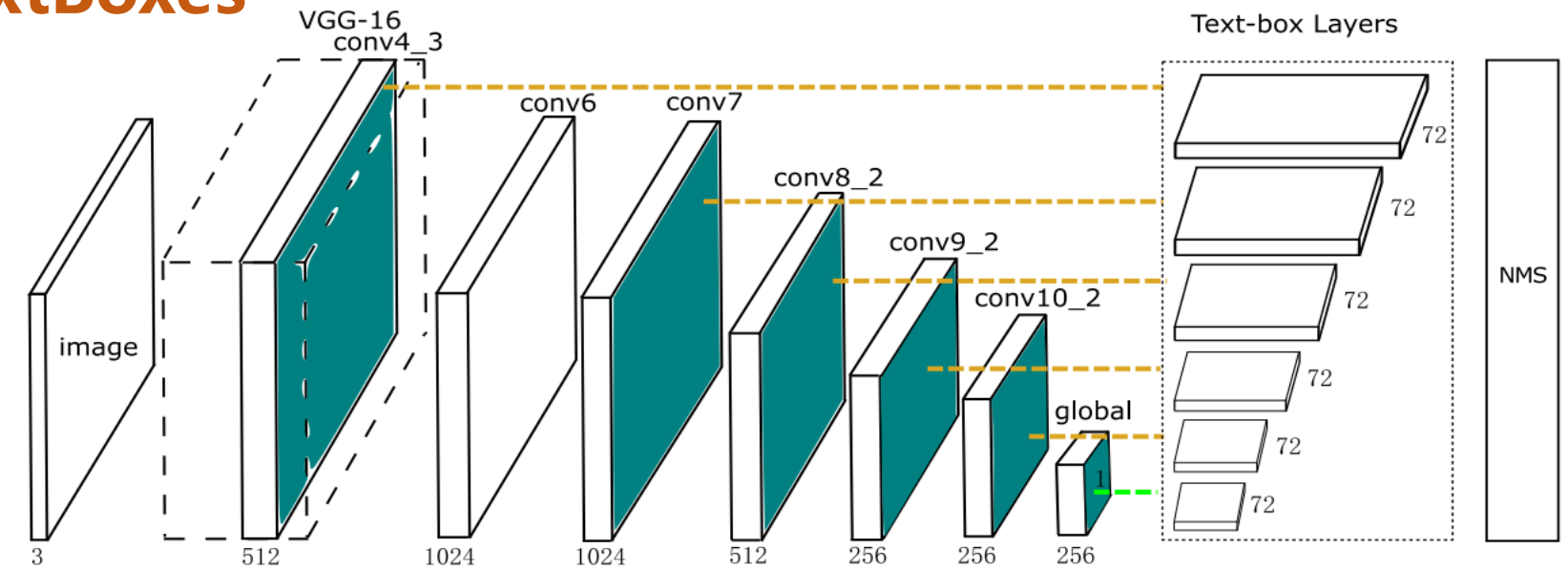
AI DISCOVERY

- ✓ 文字检测是文字识别的前提。
- ✓ **任务：**给定一张图片，找出这张图里文字出现的所有位置位置。
- ✓ 自然场景下的文字检测，非常具有**挑战性**，主要有以下几个难点：
  - 文本存在多种分布
  - 文本排布形式多样
  - 文本存在多个方向
  - 多种语言混合

文字检测

AI DISCOVERY

TextBoxes



TextBoxes的网络结构和SSD类似。主体网络继承自VGG-16，保留了conv1_1到conv4_3，用于做输入图片的特征提取。并通过一系列的卷积和池化下采样提取6个尺寸的特征图。并将这些特征图送入Text-box Layers，得到若干个可能存在文字的区域候选框，再跟一个NMS（非极大值抑制）层得出最终的预测的区域。

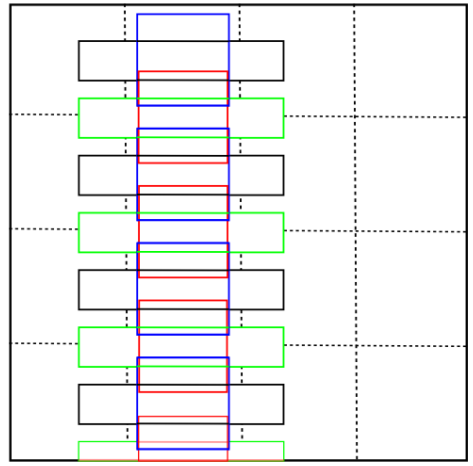
Minghui Liao, Baoguang Shi, et al, TextBoxes: A Fast Text Detector with a Single Deep Neural Network. AAAI 2017

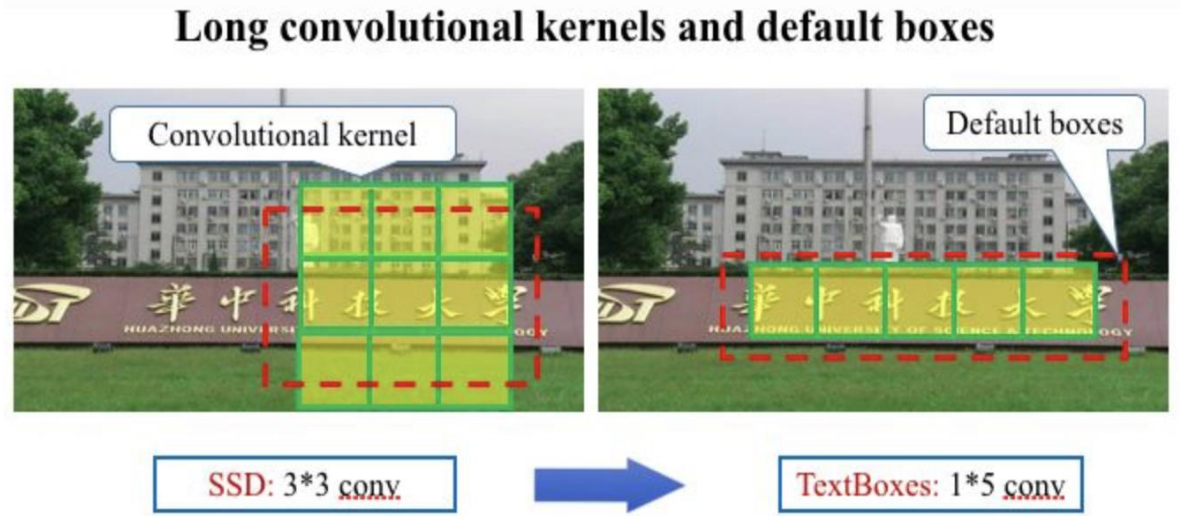
文字检测

AI DISCOVERY

Text-Box Layers

a. 输出由两部分构成，一部分是bounding box的位置，另一部分是该bounding box区域分别是文字和背景的概率。
b. 设计了长宽比分别是1、2、3、5、7、10的默认框，以匹配各种大小和和比例的文本。
c. 为了防止默认框在水平方向上排列紧密而垂直方向上排列稀疏从而造成检测失误的情况，将水平方向上的这些默认框全部向下平移半个区域的单位（右图中黑色与绿色，蓝色与红色）
d. 把原先的作为分类的卷积核3*3改成了1*5，以适应长文本。





AI DISCOVERY

文字检测

AI DISCOVERY

TextBoxes++

TextBoxes存在的问题

default box的水平的框，不能很好地检测出如图所示的倾斜文本。

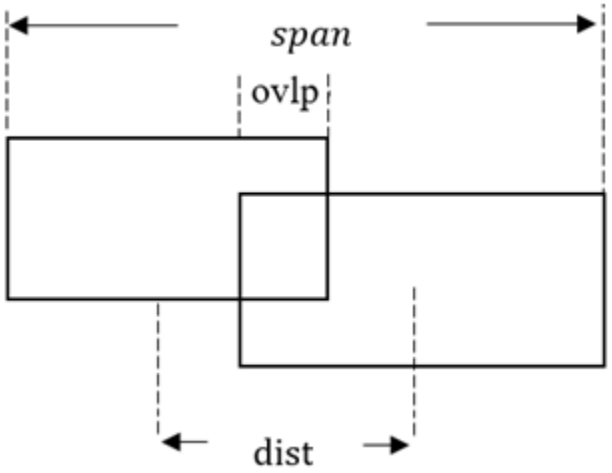
TextBoxes++作出的改进

a. 文本框的表示方式发生了变化：用4个点坐标8个数字（x1,y1,x2,y2,x3,y3,x4,y4）表示。
b. 级联NMS
c. 卷积大小改变：1*5 -> 3*5，以适应倾斜文本
d. 训练过程采用OHEM策略
e. 多尺度训练以适应不同尺度的目标

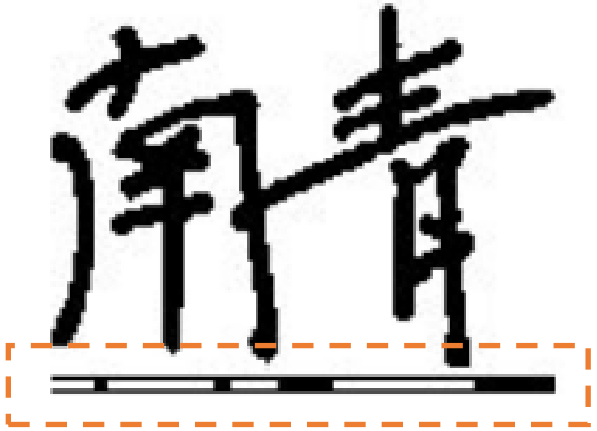
Minghui Liao, Baoguang Shi, et al, TextBoxes++: A Single-Shot Oriented Scene Text Detector, TIP, 2018

文字识别-过分割

- 第一步：连通域分析，寻找图像中的连通区域（即文字笔画）
- 第二步：重叠区域合并，通过计算标准重合程度nmovlp来合并重叠的连通区域
- 第三步：判断潜在的粘连区域，计算文本行平均高度LH，当$\mathrm{sw} > \theta_1 \cdot LH$时，认为该片段存在粘连
- 第四步：粘连区域切割，寻找粘连区域中的单一笔画区域，进行切割



$$\mathrm{nmovlp} = \frac{1}{2}\left(\frac{ovlp}{w_1} + \frac{ovlp}{w_2}\right) - \frac{dist}{span}$$



黑色区域即
单一笔画区域

文字识别-CNN单字识别

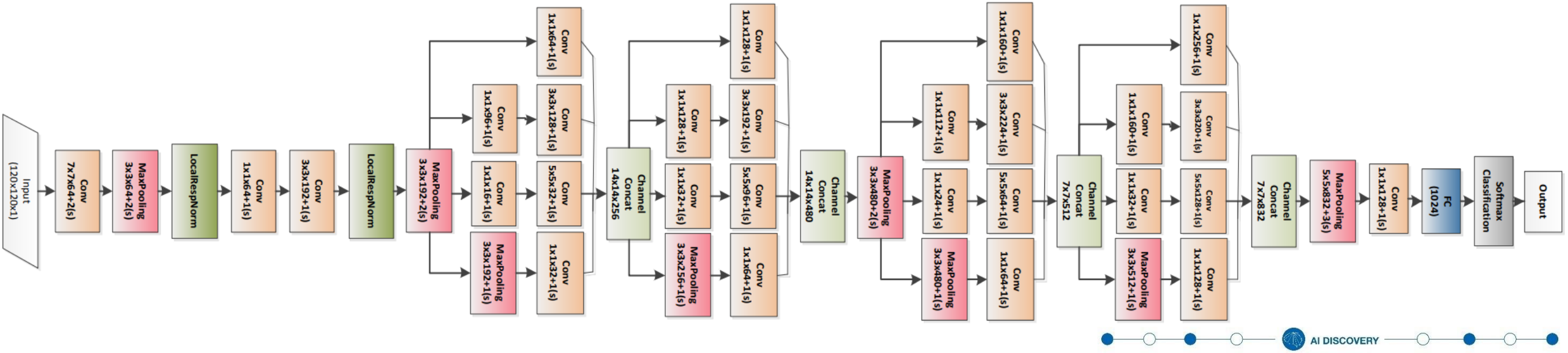
AI DISCOVERY

◆HCCR-GoogLeNet

在ICDAR2013离线手写汉字识别竞赛数据集上达到了96.35%的准确率

共19层（只考虑输入层、卷积层、池化层、Softmax输出层），由4个Inception模块组成

使用HCCR-GoogLeNet识别单个文字及文字片段，并获取识别分数



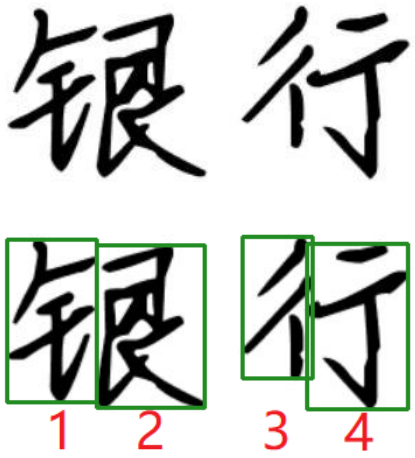
AI DISCOVERY

文字识别-集束搜索

AI DISCOVERY

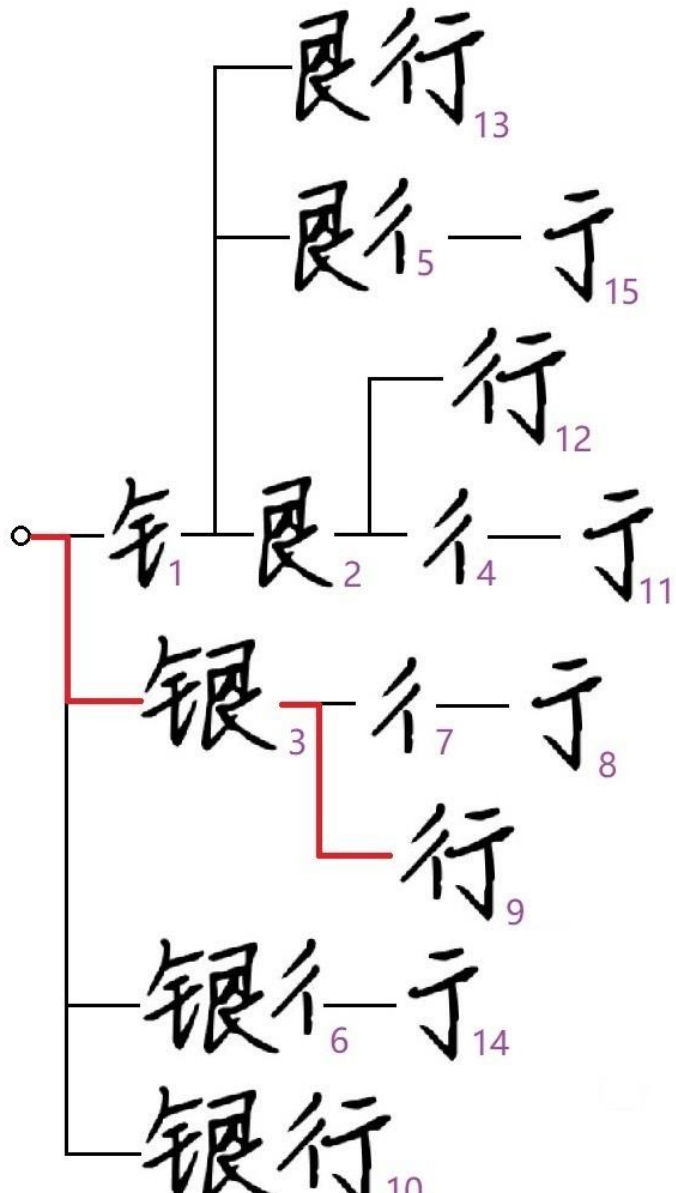
◆集束搜索

- ✓ 使用集束搜索，将过分割后的文字片段组合成正确的识别结果



◆搜索过程基于：

- ✓ 单个文字片段的识别分数
- ✓ 前后文字片段的语义关系
- ✓ 文字片段的最大组合数量



文字识别-无分割方法

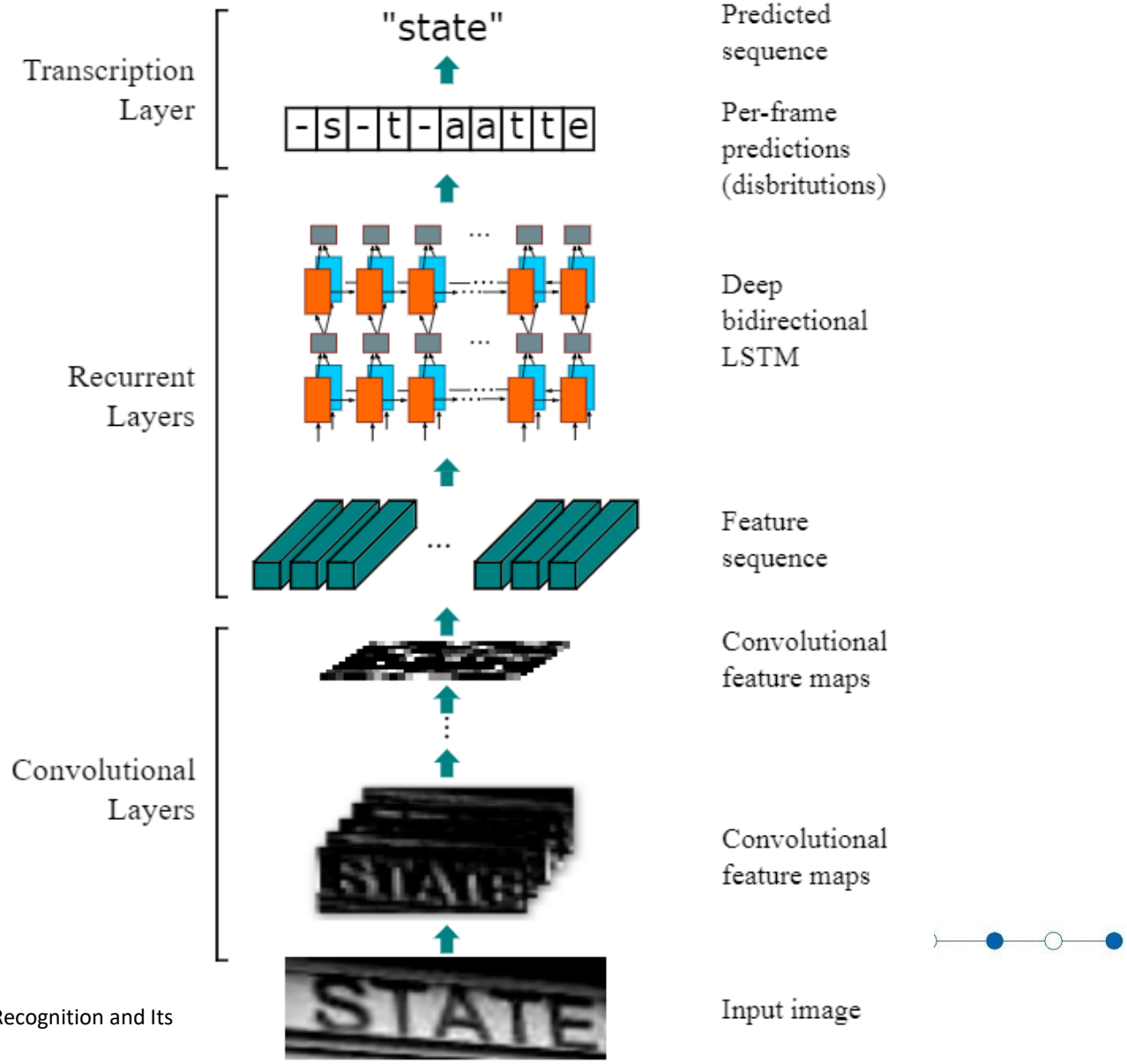
AI DISCOVERY

◆CRNN

- ✓ 端到端识别文本行图像，无需分割过程

◆ 步骤：

- ✓ 卷积层：从输入图像中提取特征序列表示
- ✓ 循环层：使用LSTM预测特征序列中每个特征向量的概率分布
- ✓ 转录层：使用CTC将预测的概率序列转换成最终识别结果



Baoguang Shi and Xiang Bai and Cong Yao, An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition, IEEE PAMI, 2017: 2298-2304

垂直应用与实践

AI DISCOVERY

医学影像分析

文字检测识别

实践：目标检测

AI DISCOVERY

课程实践

AI DISCOVERY

实践：目标检测