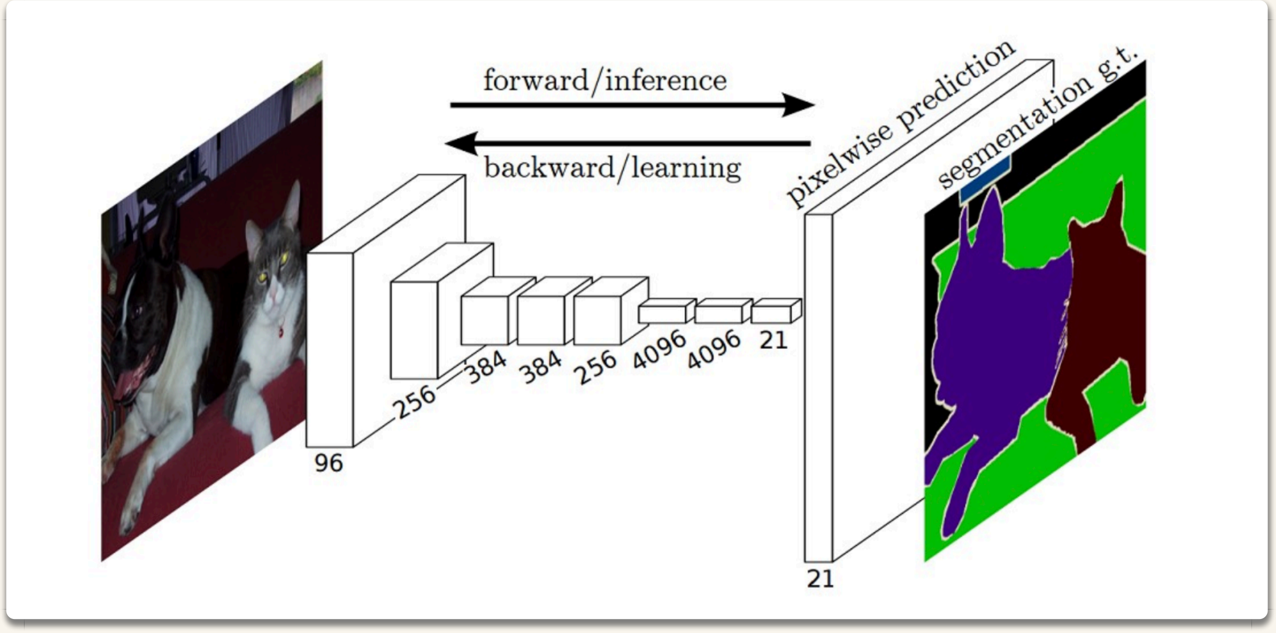
Object Segmentation Models

1. FCN

模型特点:

- 采用反卷积对最后一层的feature map进行上采样(up-sampling)使他恢复到与输入相同尺寸，保留了原输入图像的空间信息，最后在up sampling(反卷积deconvolutional) 的特征图上进行逐帧的像素分类--pixel wise **softmax** prediction (**softmax loss**)。
- 属于语义分割 (**Semantic Segmentation**)



2. U-Net

3. SegNet

4. RefineNet

5. PSPNet

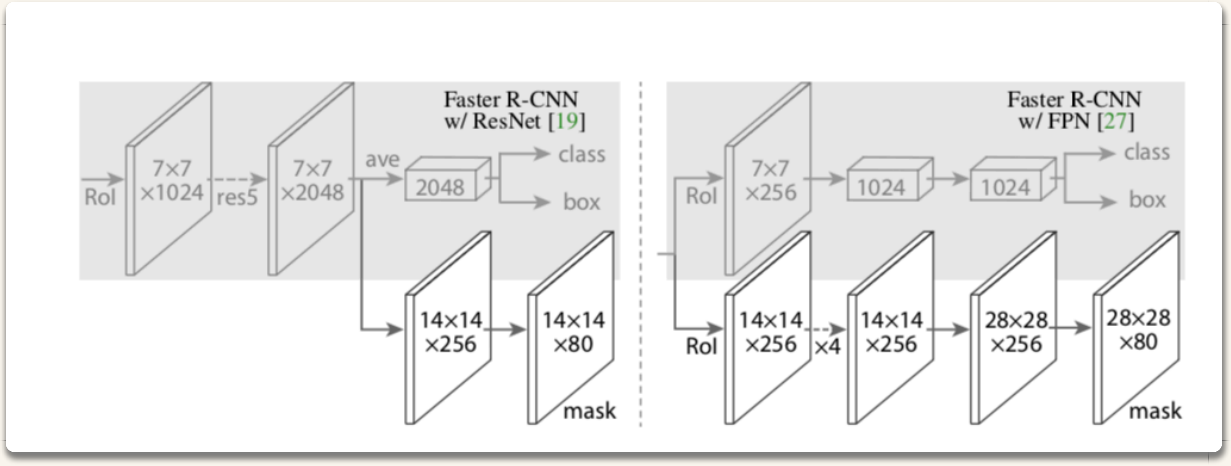
6.Mask-R-CNN

模型特点：

- Two-stage which is same as Faster-RCNN
  1. RPN proposes candidate object bounding boxes
  2. extreacts features using RoIPool from each candidate box and performs classification and bounding-box regression

- **Binary mask for each RoI**
  - $RoILostFunction : L = L_{cls} + L_{box} + L_{mask}$
  - Mask branch (**FCN** layers) has a Km^2 dimensional output for each RoI (resolution $m * m$), K for K classes
  - 通过**FCN**生成mask，然后再逐帧做**pixel-wise sigmoid**

- **RoIAlign**
  - **保留浮点数**，用除法将region proposal平均分成kxk个。
  - 不在pixel边界的点使用**双线性插值**计算得出。
  - 解决了misalignment的问题，该问题在分类问题中影响不大。但在pixel级别分割问题中存在较大误差，特别是针对小物体
  - Mask path可以嵌入各种**Head Architecture**



- Multinomail vs. Independent Masks
  - OvR分类的效果优于OvO的效果 (Sigmoid 属于二分类, 其他classes对loss不产生影响，binary loss)
  - softmax为概率loss

- Class-Specific vs. Class-Agnostic Masks
  - Class-Specific: one mxm mask per class
  - Class-Agnostic: single mxm output regardless of class

- Main Results

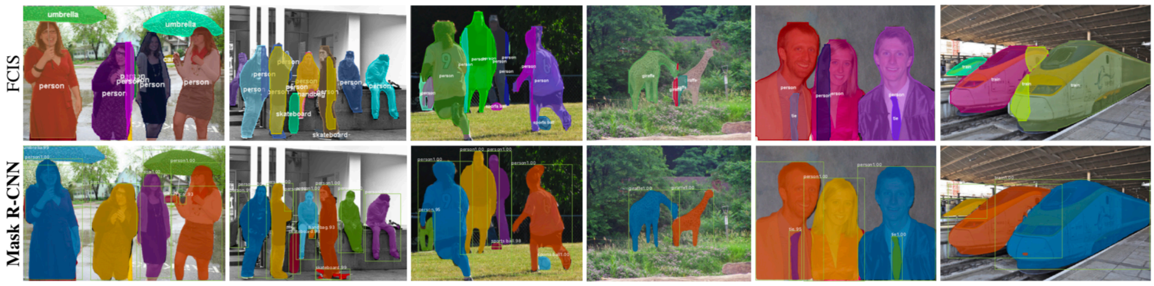


Figure 6. FCIS+++ [26] (top) *vs*. Mask R-CNN (bottom, ResNet-101-FPN). FCIS exhibits systematic artifacts on overlapping objects.

net-depth-features	AP	AP_{50}	AP_{75}
ResNet-50-C4	30.3	51.2	31.5
ResNet-101-C4	32.7	54.2	34.3
ResNet-50-FPN	33.6	55.2	35.3
ResNet-101-FPN	35.4	57.3	37.5
ResNeXt-101-FPN	**36.7**	**59.5**	**38.9**

(a) **Backbone Architecture**: Better backbones bring expected gains: deeper networks do better, FPN outperforms C4 features, and ResNeXt improves on ResNet.

	AP	AP_{50}	AP_{75}
softmax	24.8	44.1	25.1
sigmoid	**30.3**	**51.2**	**31.5**
	+5.5	*+7.1*	*+6.4*

(b) **Multinomial *vs.* Independent Masks** (ResNet-50-C4): *Decoupling* via per-class binary masks (sigmoid) gives large gains over multinomial masks (softmax).

	align?	bilinear?	agg.	AP	AP_{50}	AP_{75}
RoIPool [12]			max	26.9	48.8	26.4
RoIWarp [10]		✓	max	27.2	49.2	27.1
		✓	ave	27.1	48.9	27.1
RoIAlign	✓	✓	max	**30.2**	**51.0**	**31.8**
	✓	✓	ave	**30.3**	**51.2**	**31.5**

(c) **RoIAlign** (ResNet-50-C4): Mask results with various RoI layers. Our RoIAlign layer improves AP by ~3 points and AP_{75} by ~5 points. Using proper alignment is the only factor that contributes to the large gap between RoI layers.

	AP	AP_{50}	AP_{75}	AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}
RoIPool	23.6	46.5	21.6	28.2	52.7	26.9
RoIAlign	**30.9**	**51.8**	**32.1**	**34.0**	**55.3**	**36.4**
	+7.3	*+ 5.3*	*+10.5*	*+5.8*	*+2.6*	*+9.5*

(d) **RoIAlign** (ResNet-50-**C5**, *stride 32*): Mask-level and box-level AP using *large-stride* features. Misalignments are more severe than with stride-16 features (Table 2c), resulting in big accuracy gaps.

	mask branch	AP	AP_{50}	AP_{75}
MLP	fc: $1024\rightarrow1024\rightarrow80\cdot28^2$	31.5	53.7	32.8
MLP	fc: $1024\rightarrow1024\rightarrow1024\rightarrow80\cdot28^2$	31.5	54.0	32.6
FCN	conv: $256\rightarrow256\rightarrow256\rightarrow256\rightarrow256\rightarrow80$	**33.6**	**55.2**	**35.3**

(e) **Mask Branch** (ResNet-50-FPN): Fully convolutional networks (FCN) *vs.* multi-layer perceptrons (MLP, fully-connected) for mask prediction. FCNs improve results as they take advantage of explicitly encoding spatial layout.

Table 2. **Ablations**. We train on `trainval35k`, test on `minival`, and report *mask* AP unless otherwise noted.