Computer Vision

第十一周　图像分割

庞彦
yanpang@gzhu.edu.cn

01 Introduction of Image Segmentation

图像分割介绍

Image Segmentation

Pixel-wise image segmentation is a well-studied problem in computer vision.

Image segmentation is the task of classifying each pixel in an image from a predefined set of classes.

Image Segmentation

Input

segmented

1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5
4	4	4	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4

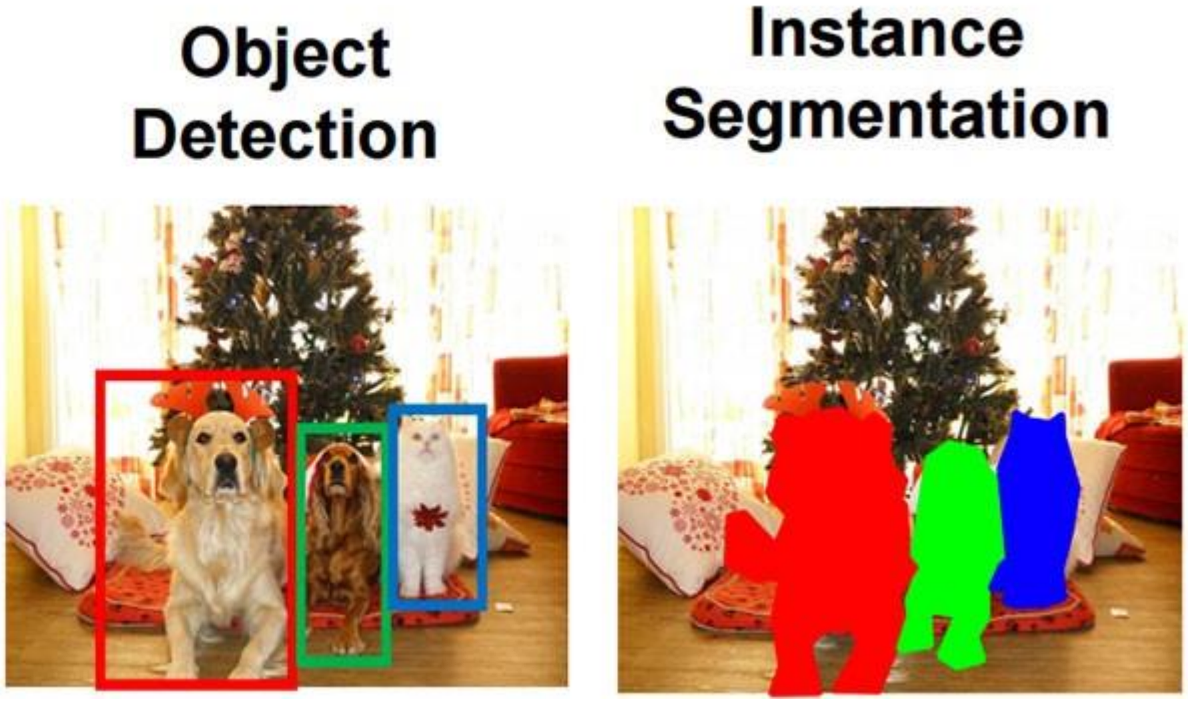
Semantic Labels

In particular, the goal is to take an image of size $W \times H \times 3$ and generate a $W \times H$ matrix containing the predicted class ID's corresponding to all the pixels.

Image Segmentation

Semantic segmentation is different from object detection as it does not predict any bounding boxes around the objects.

Bounding -Boxes



Bounding -Boxes

Image Segmentation

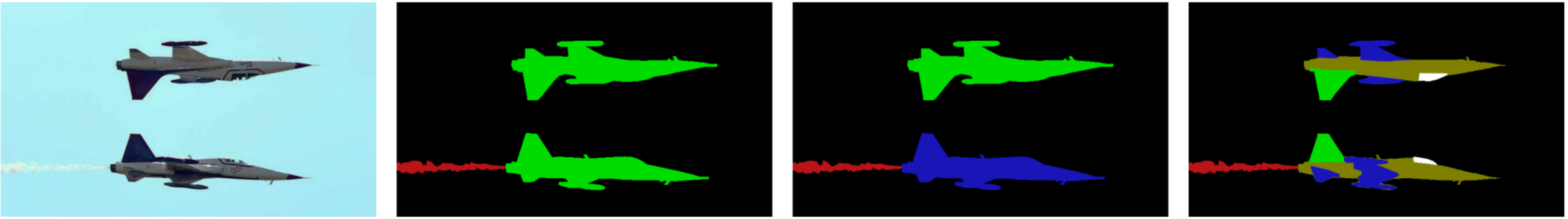


Figure 1: A sample image and its annotation for object, instance and parts segmentations separately, from left to right.

Semantic	Segmentation
Instance	Segmentation
Parts	Segmentation

Benchmarks

Semantic Segmentation

Trend	Dataset	Best Model	Paper Title	Paper	Code	Compare
	Cityscapes test	HRNet-OCR (Hierarchical Multi-Scale Attention)	Hierarchical Multi-Scale Attention for Semantic Segmentation			See all
	PASCAL VOC 2012 test	EfficientNet-L2+NAS-FPN (single scale test, with self-training)	Rethinking Pre-training and Self-training			See all
	PASCAL Context	CAA + Simple decoder (Efficientnet-B7)	Channelized Axial Attention for Semantic Segmentation			See all
	ADE20K val	Focal-L (UperNet, ImageNet-22k pretrain)	Focal Self-attention for Local-Global Interactions in Vision Transformers			See all
	Cityscapes val	HRNet-OCR	Hierarchical Multi-Scale Attention for Semantic Segmentation			See all
	ADE20K	Focal-L (UperNet, ImageNet-22k pretrain)	Focal Self-attention for Local-Global Interactions in Vision Transformers			See all
	PASCAL VOC 2012 val	EfficientNet-L2+NAS-FPN (single scale test, with self-training)	Rethinking Pre-training and Self-training			See all
	S3DIS	PointTransformer	Point Transformer			See all
	ScanNet	VMVF	Virtual Multi-view Fusion for 3D Semantic Segmentation			See all
	Semantic3D	RandLA-Net	RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds			See all

https://paperswithcode.com/task/semantic-segmentation

Benchmarks

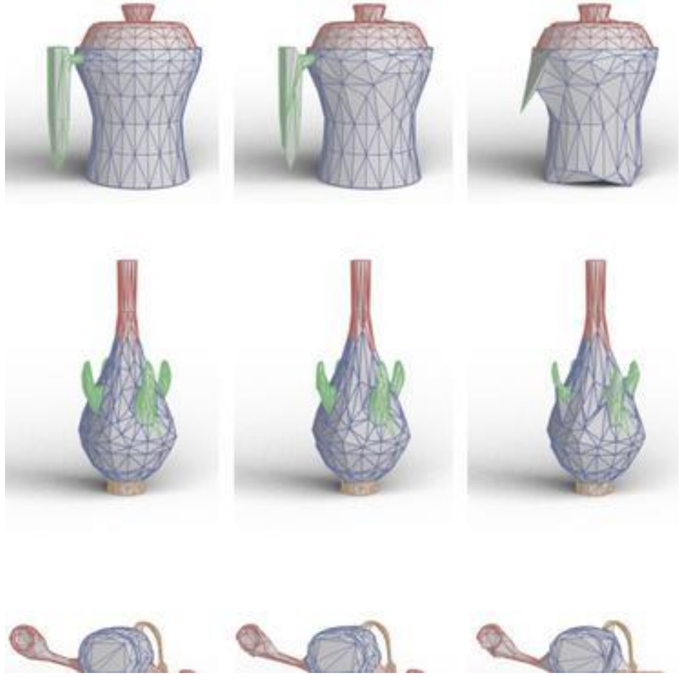
Instance Segmentation

Trend	Dataset	Best Model	Paper Title	Paper	Code	Compare
	COCO test-dev	Dual-Swin-B (HTC, multi-scale)	CBNetV2: A Composite Backbone Network Architecture for Object Detection			See all
	COCO minival	Focal-L (HTC++, multi-scale)	Focal Self-attention for Local-Global Interactions in Vision Transformers			See all
	Cityscapes test	PolyTransform	PolyTransform: Deep Polygon Transformer for Instance Segmentation			See all
	iSAID	PANet++	iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images			See all
	LVIS v1.0	PointRend (MaskR-CNN, ResNet-50-FPN)	PointRend: Image Segmentation as Rendering			See all
	NYU Depth v2	SGPN-CNN	SGPN: Similarity Group Proposal Network for 3D Point Cloud Instance Segmentation			See all
	Cityscapes val	GAIS-Net	Geometry-Aware Instance Segmentation with Disparity Maps			See all
	KINS	BCNet	Deep Occlusion-Aware Instance Segmentation with Overlapping BiLayers			See all
	nuScenes	TraDeS	Track to Detect and Segment: An Online Multi-Object Tracker			See all

https://paperswithcode.com/task/instance-segmentation

Benchmarks

3D Part Segmentation



Trend	Dataset	Best Model	Paper Title	Paper	Code	Compare
	ShapeNet-Part	🏆 Feature Geometric Net (FG-Net)	FG-Net: Fast Large-Scale LiDAR Point Clouds Understanding Network Leveraging Correlated Feature Mining and Geometric-Aware Modelling			See all

https://paperswithcode.com/task/3d-part-segmentation

Benchmarks

Instance Video Semantic Segmentation

Trend	Dataset	Best Model	Paper Title	Paper	Code	Compare
	Cityscapes val	🏆 TMANet-50	Temporal Memory Attention for Video Semantic Segmentation			See all
	CamVid	🏆 TMANet-50	Temporal Memory Attention for Video Semantic Segmentation			See all

https://paperswithcode.com/task/video-semantic-segmentation

Datasets

2D images

2.5D RGB-D

3D images

2D Datasets

PASCAL Visual Object Classes (VOC)

PASCAL Context

Microsoft Common Objects in Context (MS COCO)

Cityscapes

KITTI

...

2.5D Datasets

NYU-D V2

SUN RGB-D

UW RGB-D Object Dataset

ScanNet

…

3D Datasets

Stanford 2D-3D

ShapeNet Core

Sydney Urban Objects Dataset

…

Performance Evaluation

Accuracy:

ROC-AUC

Pixel Accuracy

Intersection over Union

Precision-Recall Curve

Dice

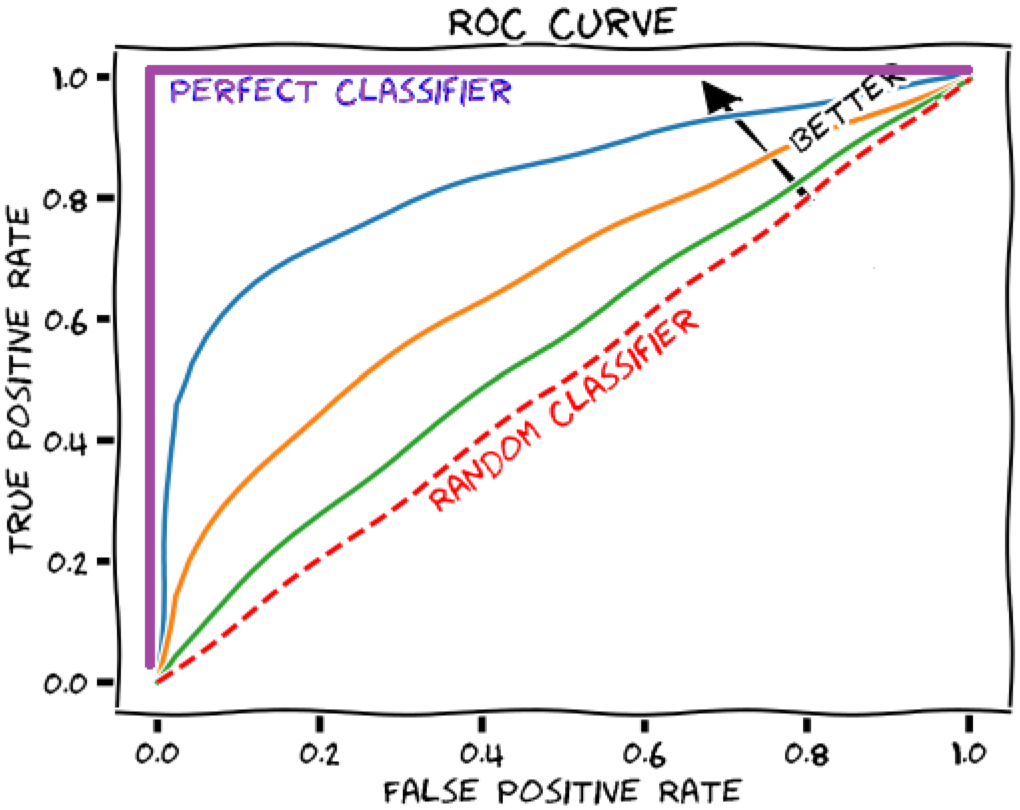
Performance Evaluation

Accuracy：

ROC-AUC

ROC summarizes the trade-off between true positive rate and false-positive rate for a predictive model using different probability thresholds;

AUC stands for the area under this curve, which is 1 at maximum.



Performance Evaluation

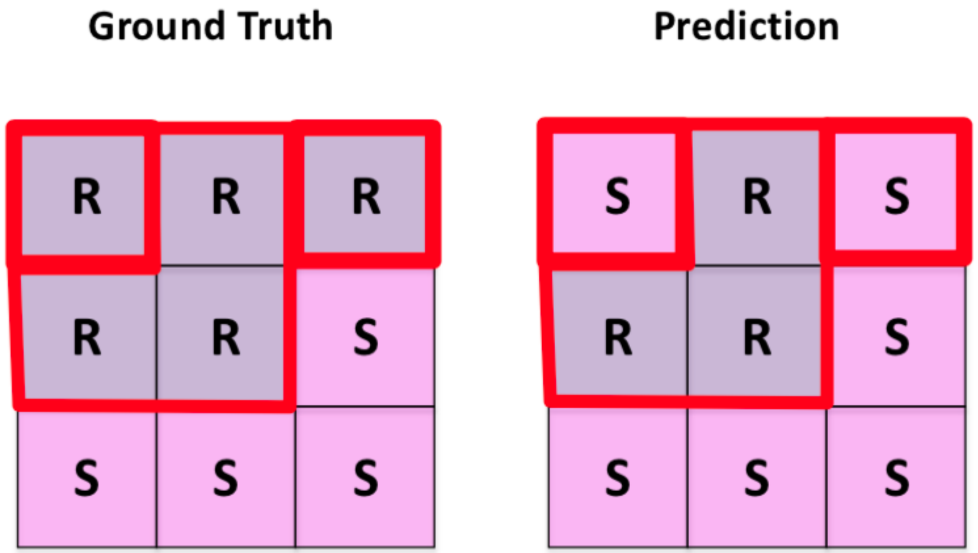
Accuracy：

Pixel Accuracy

$$PA = \frac{\sum_{j=1}^{k} n_{jj}}{\sum_{j=1}^{k} t_j}, \qquad mPA = \frac{1}{k}\sum_{j=1}^{k}\frac{n_{jj}}{t_j}$$

Pixel Accuracy: calculates the ratio between the amount of properly classified pixels and their total number.

Mean pixel accuracy (mPA) computes the ratio of correct pixels on a per-class basis.



Performance Evaluation

Accuracy：

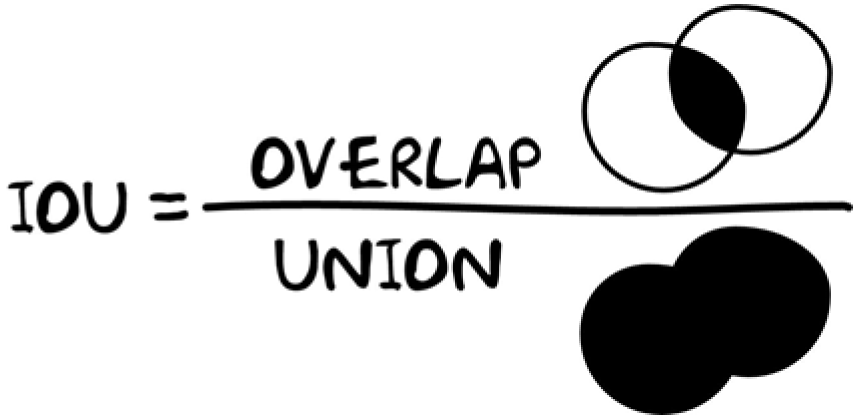
Intersection over Union

$$IoU = \frac{\sum_{j=1}^{k} n_{jj}}{\sum_{j=1}^{k}(n_{ij} + n_{ji} + n_{jj})}, \qquad i \neq j$$

IoU: the ratio of the intersection of the pixel-wise classification results with the ground truth, to their union

$$mIoU = \frac{1}{k}\sum_{j=1}^{k}\frac{n_{jj}}{n_{ij} + n_{ji} + n_{jj}}, \qquad i \neq j$$

mIoU: Mean Intersection over Union is the class-averaged IoU



Performance Evaluation

Accuracy：

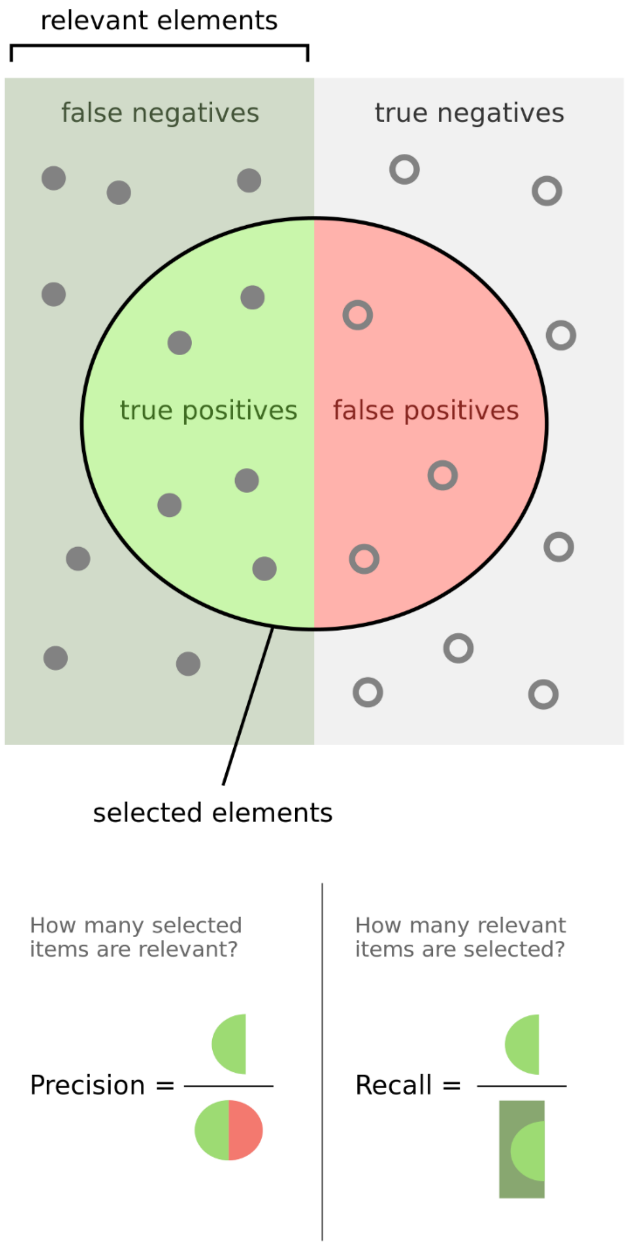
Precision-Recall Curve

Precision: ratio of hits over a summation of hits and false alarms

Recall: ratio of hits over a summation of hits and misses

$$Prec. = \frac{n_{jj}}{n_{ij} + n_{jj}}, \quad Recall = \frac{n_{jj}}{n_{ji} + n_{jj}}, i \neq j$$

$$F_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$



History

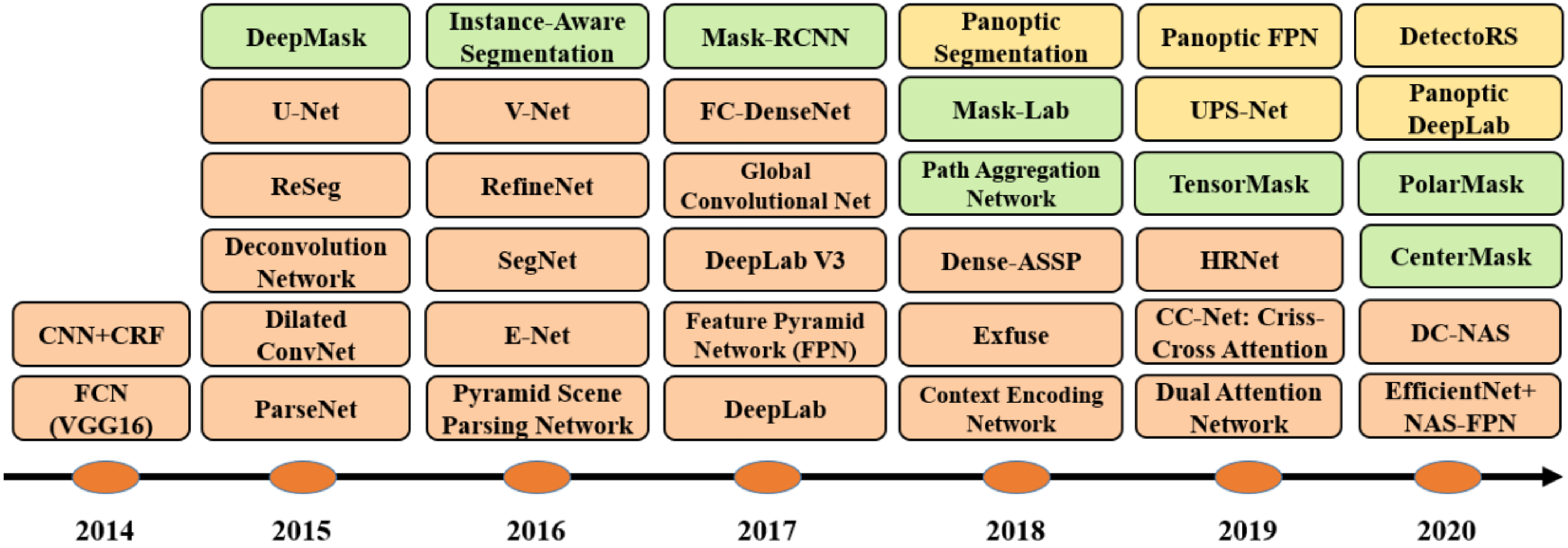
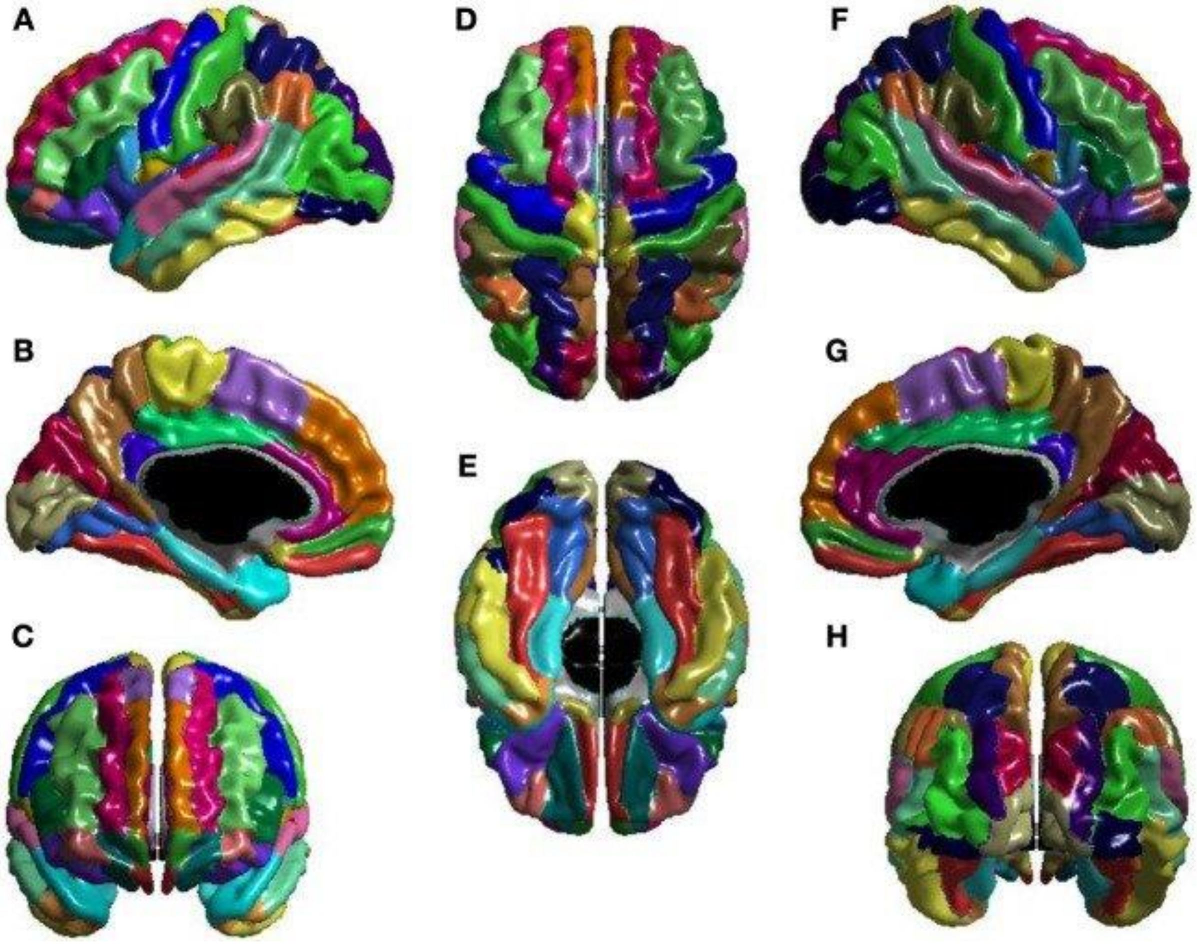


Fig. 32. The timeline of DL-based segmentation algorithms for 2D images, from 2014 to 2020. Orange, green, andn yellow blocks refer to semantic, instance, and panoptic segmentation algorithms respectively.

Applications

Medical Image



Applications

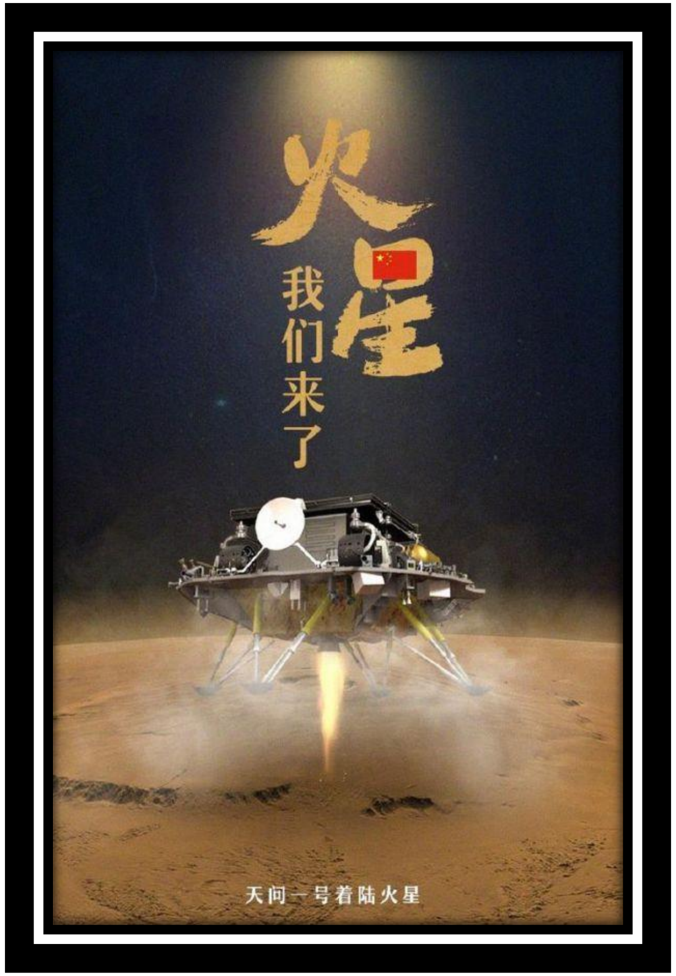
Autonomous Vehicles

Applications

Satellite Image Analysis

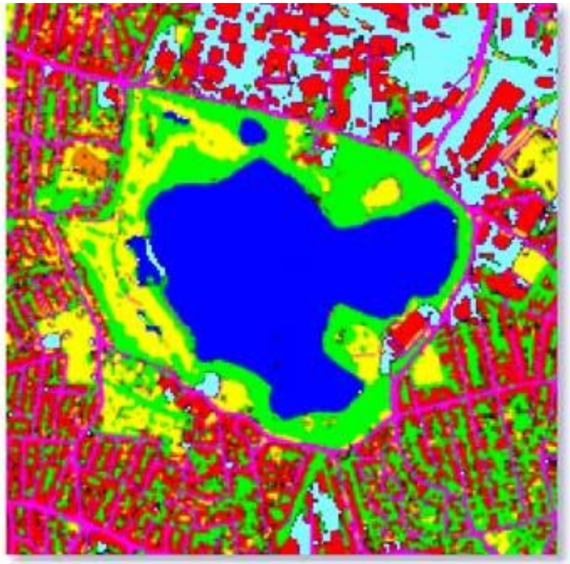
黄帝纪年第4718年

2021年5月15日，祖国的天问一号着陆巡视器成功着陆于火星乌托邦平原南部预选着陆区，我国首次火星探测任务着陆火星取得圆满成功。



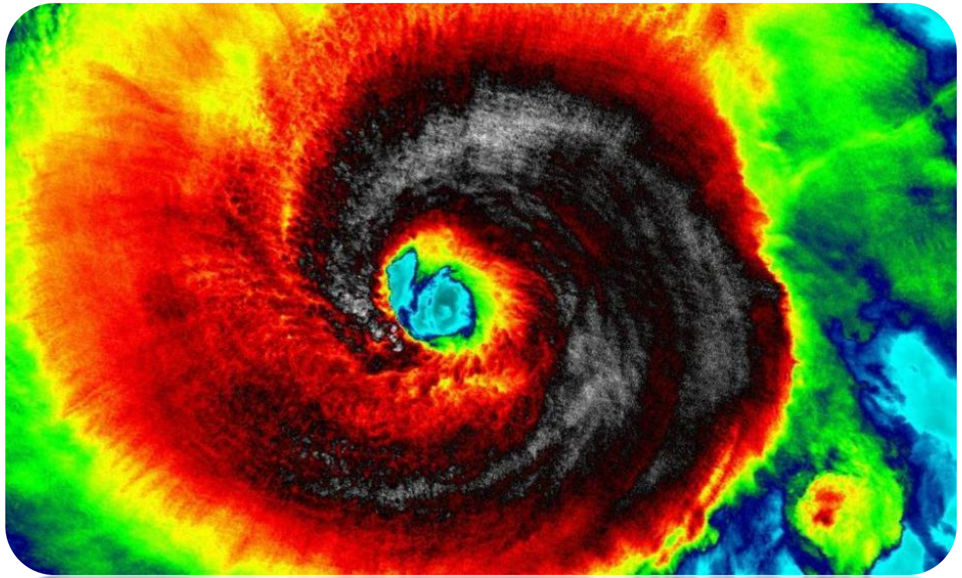
Applications

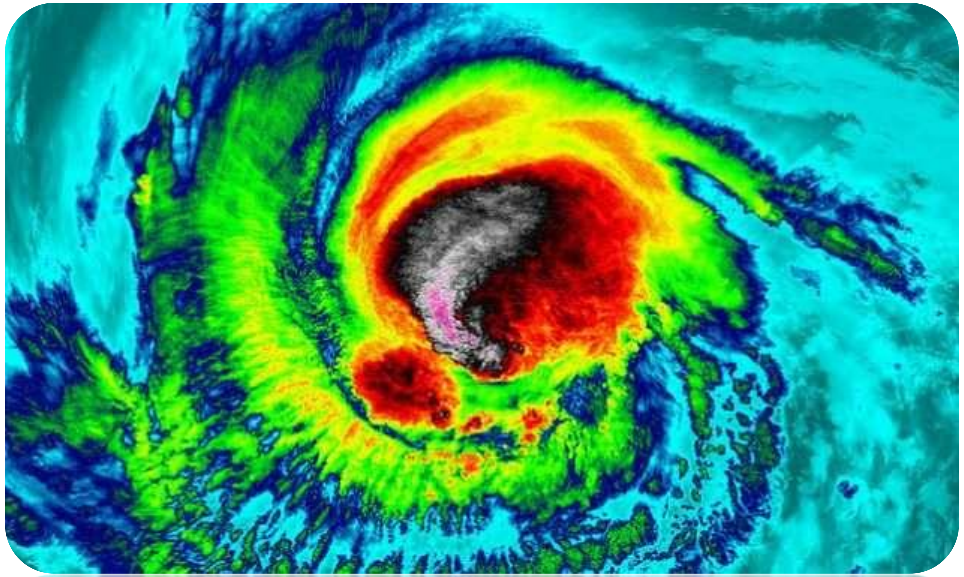
Satellite Image Analysis



Applications

Climate Analysis





02

Image Segmentation

图像分割核心算法

Summary

Fully Convolutional Networks

Encoder-Decoder Based Models

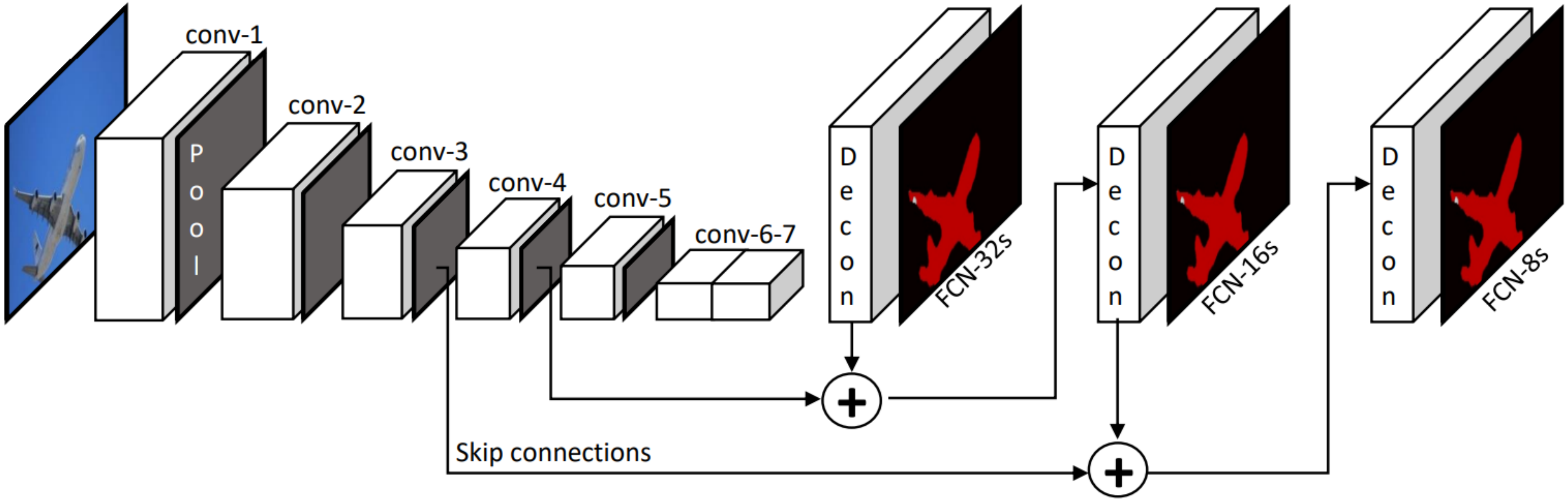
Multi-Scale and Pyramid Network Based Models

R-CNN Based Models (for Instance Segmentation)

Recurrent Neural Network Based Models

More …

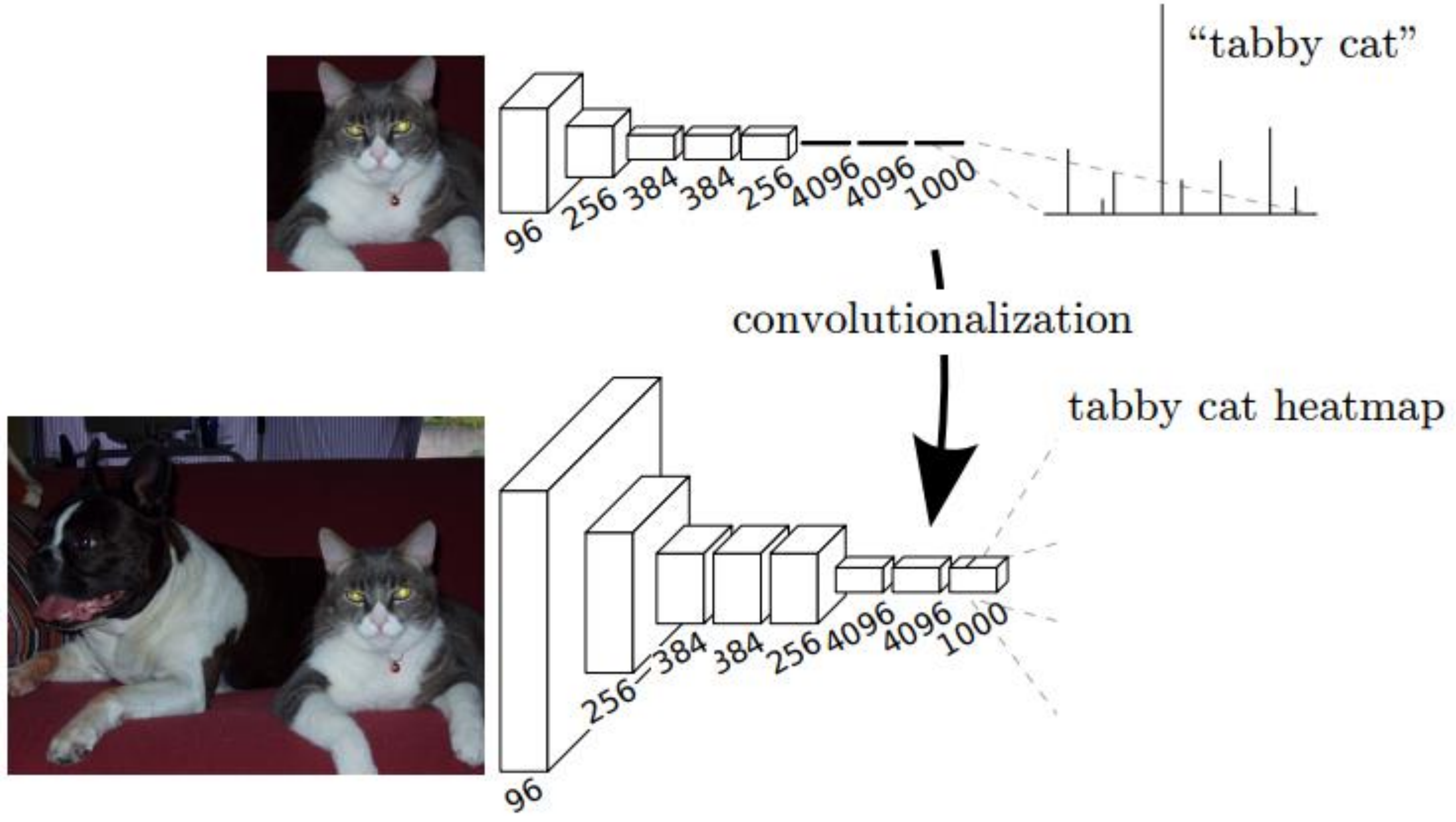
Fully Convolutional Networks



FCNs are trained end-to-end and are designed to make dense predictions for per-pixel tasks like semantic segmentation.

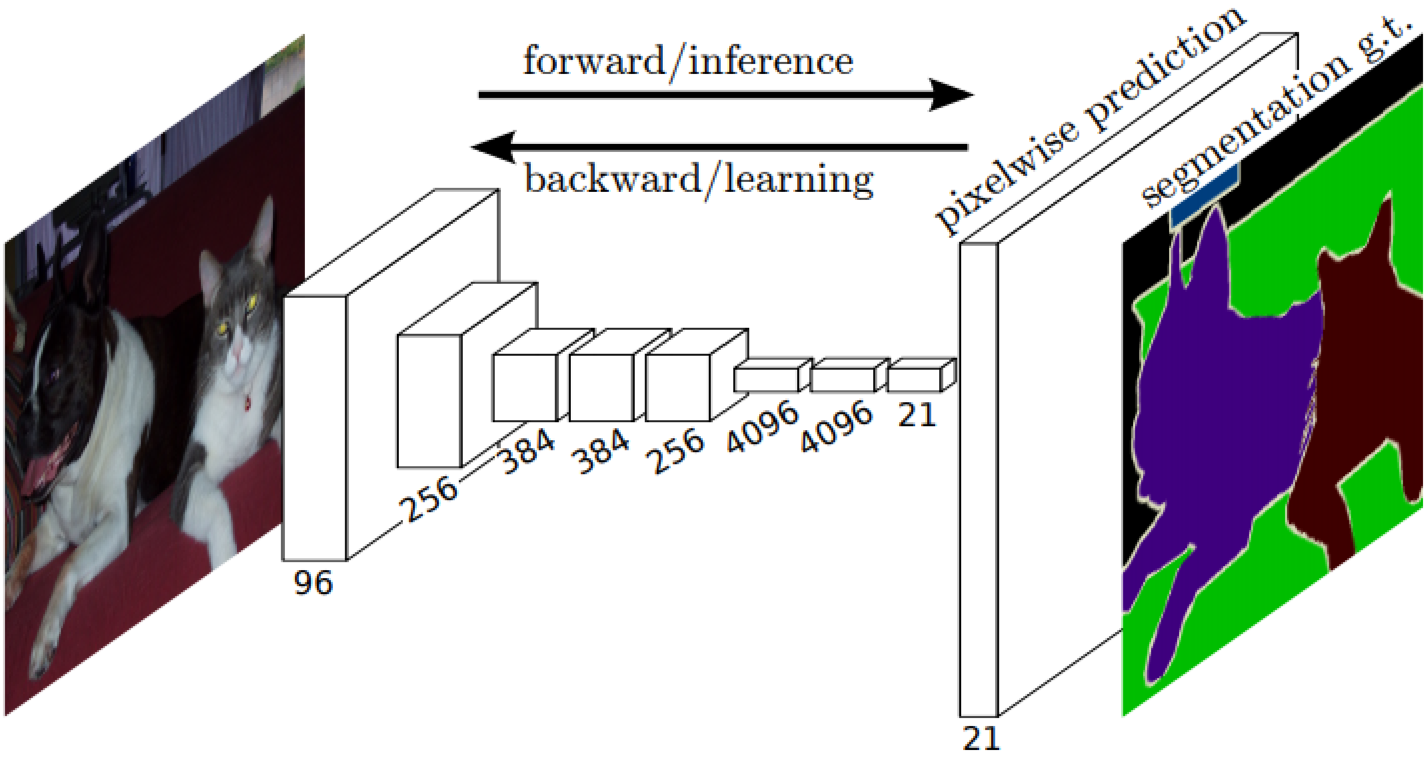
Fully Convolutional Networks

Transforming fully connected layers into convolution layers enables a classification net to output a heatmap.



Fully Convolutional Networks

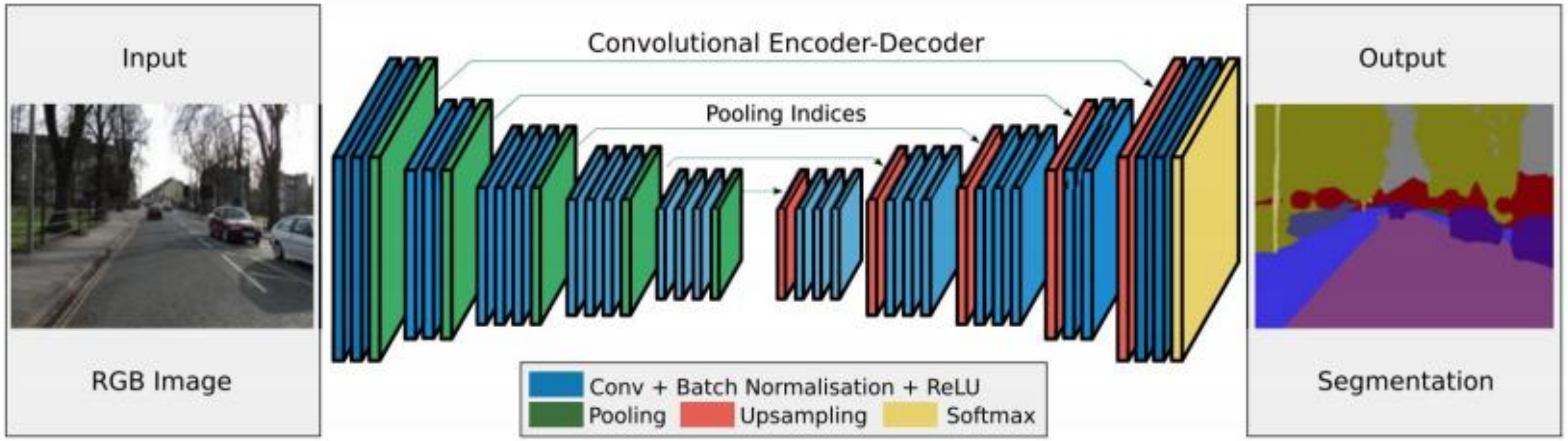
Transforming fully connected layers into convolution layers enables a classification net to output a heatmap.



Encoder-Decoder Based Models

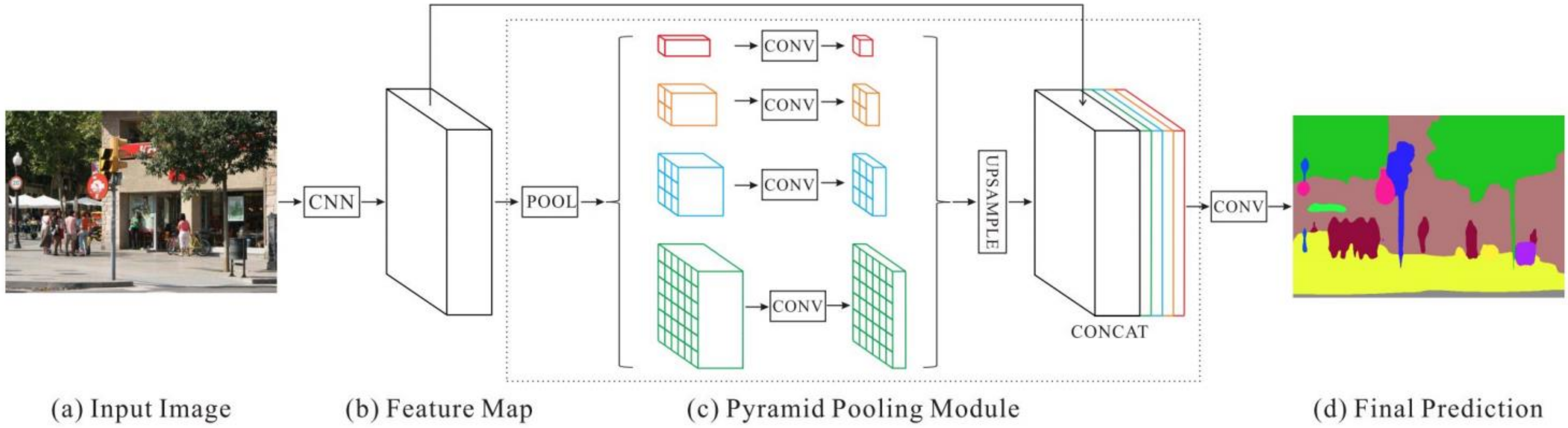
Another popular family of deep models for image segmentation is based on the convolutional **encoder-decoder** architecture.

SegNet



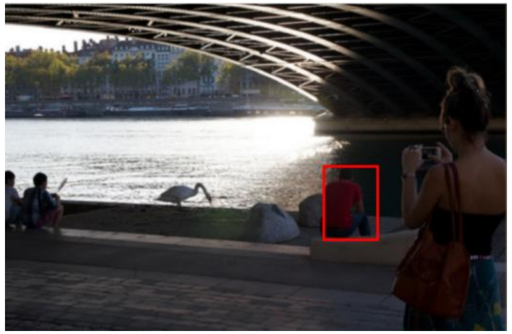
Multi-Scale and Pyramid Network Based Models

PSPN: Pyramid Scene Parsing Network

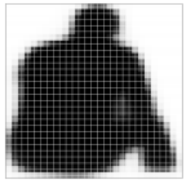


R-CNN Based Models (for Instance Segmentation)

Mask R-CNN

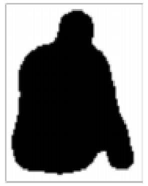


28x28 soft prediction from Mask R-CNN
(enlarged)



Soft prediction resampled to image coordinates
(bilinear and bicubic interpolation work equally well)

Final prediction (threshold at 0.5)



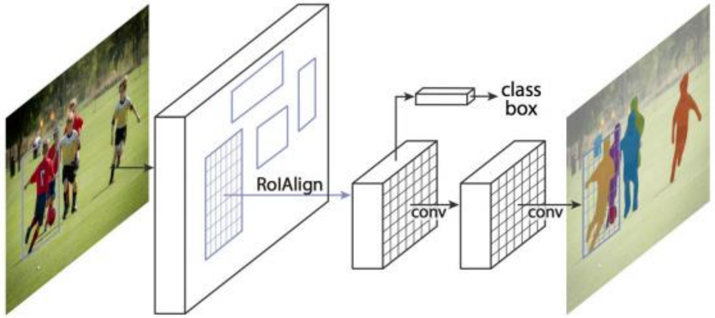


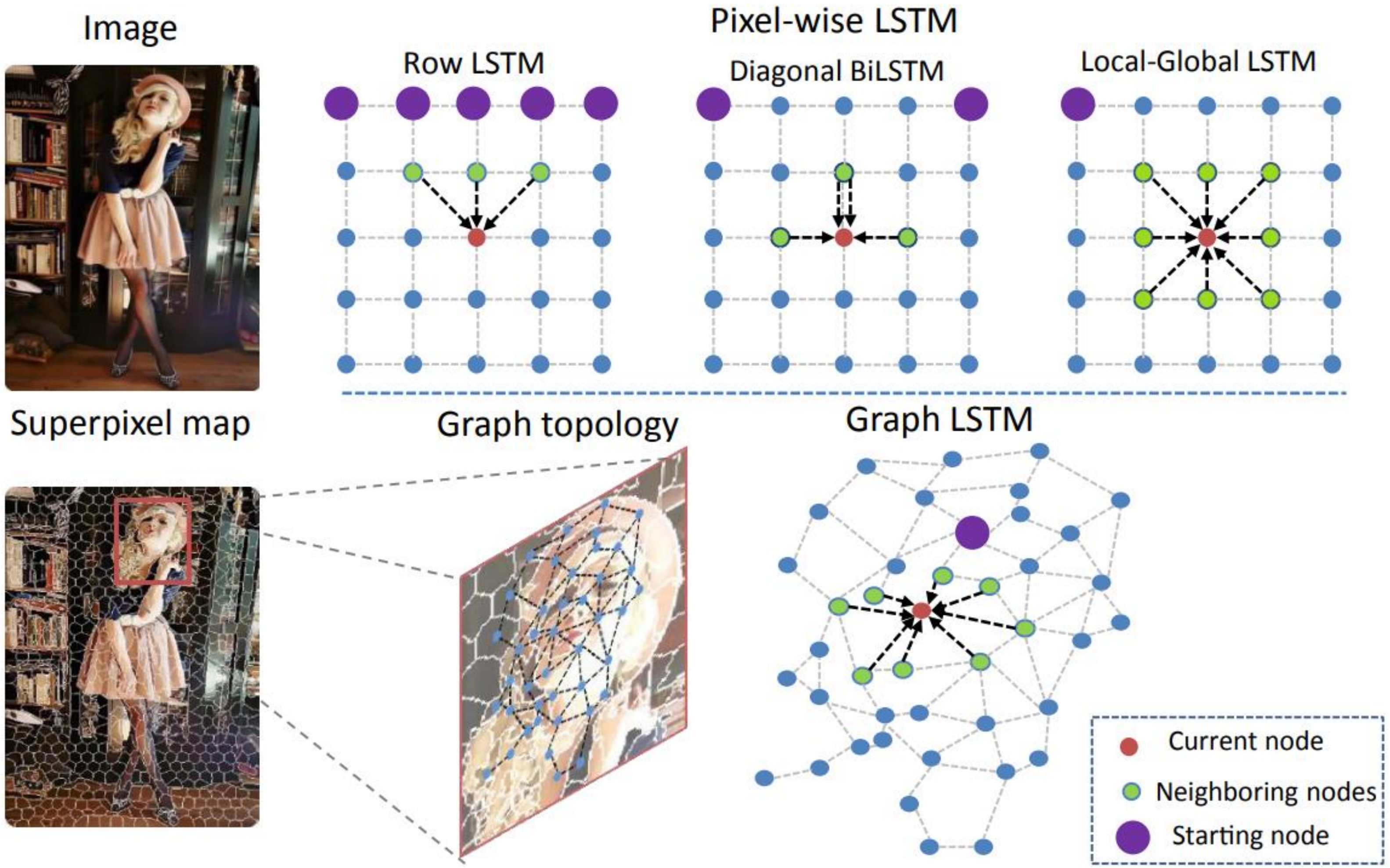
Fig. 17. Mask R-CNN architecture for instance segmentation. From [64].

Fig. 18. Mask R-CNN results on sample images from the COCO test set. From [64].

Recurrent Neural Network Based Models

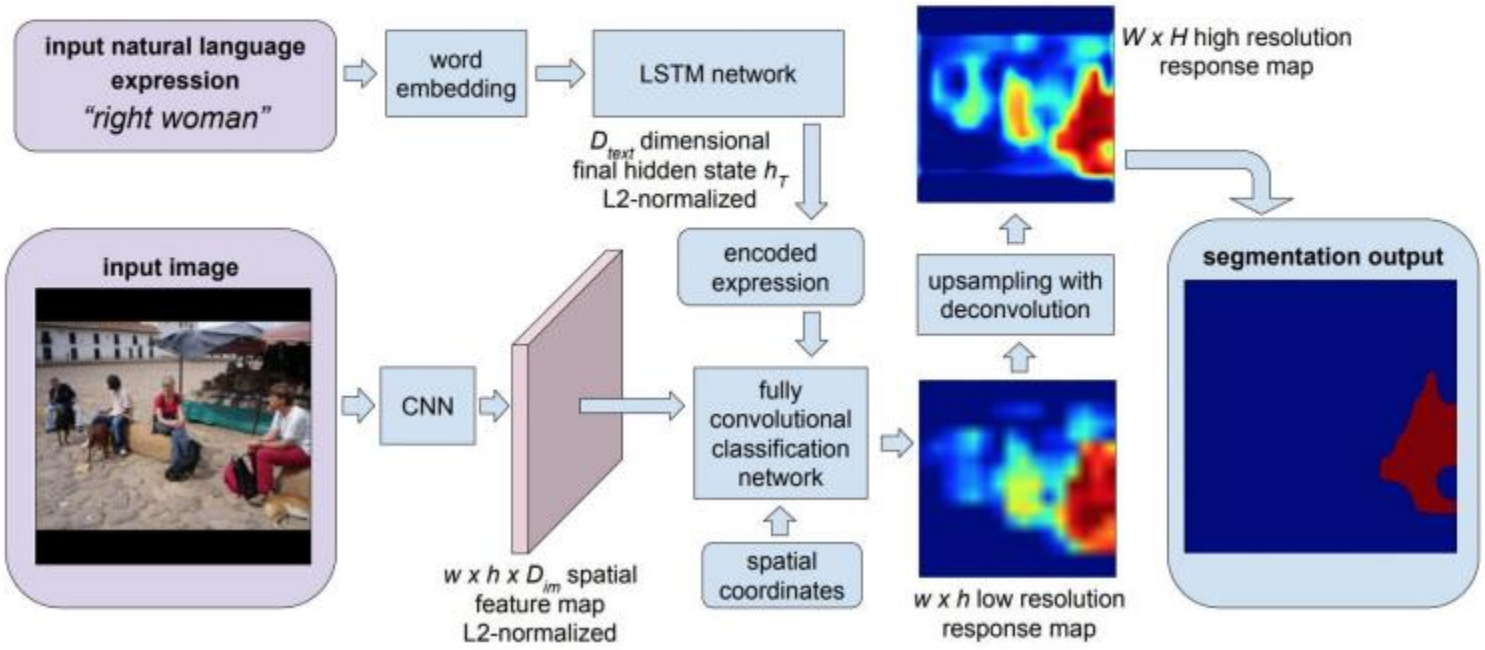
pixel-wise RNN model

graph-LSTM model



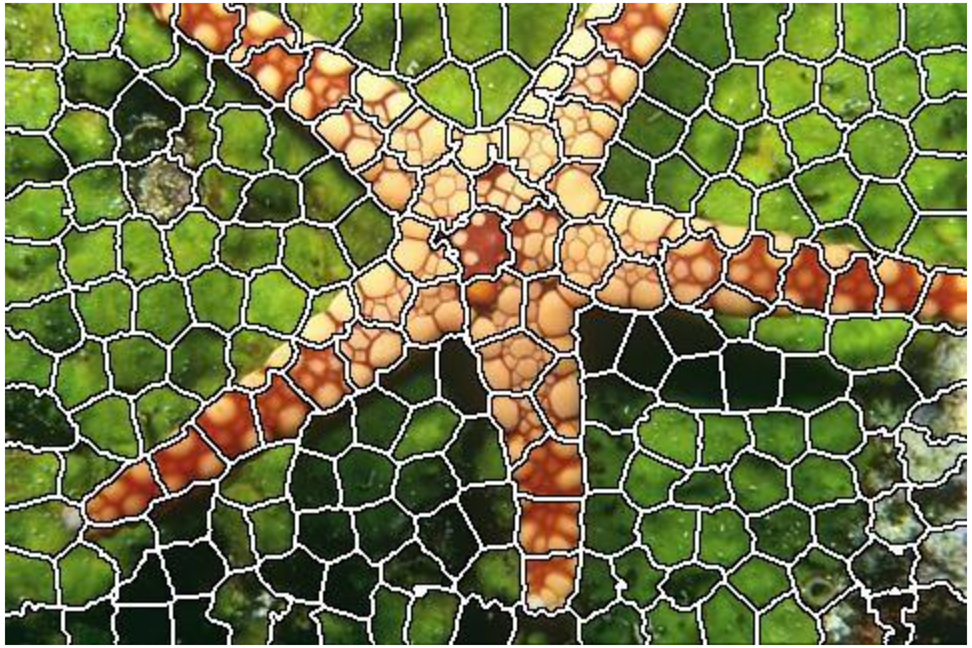
Recurrent Neural Network Based Models

CNN + RNN:
CNN to encode the image and LSTM to encode the Natural langage description.



Recurrent Neural Network Based Models

A **super-pixel** can be defined as a group of pixels that share common characteristics (like pixel intensity).



Q&A