華中科技大學
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

图像分割
Image segmentation

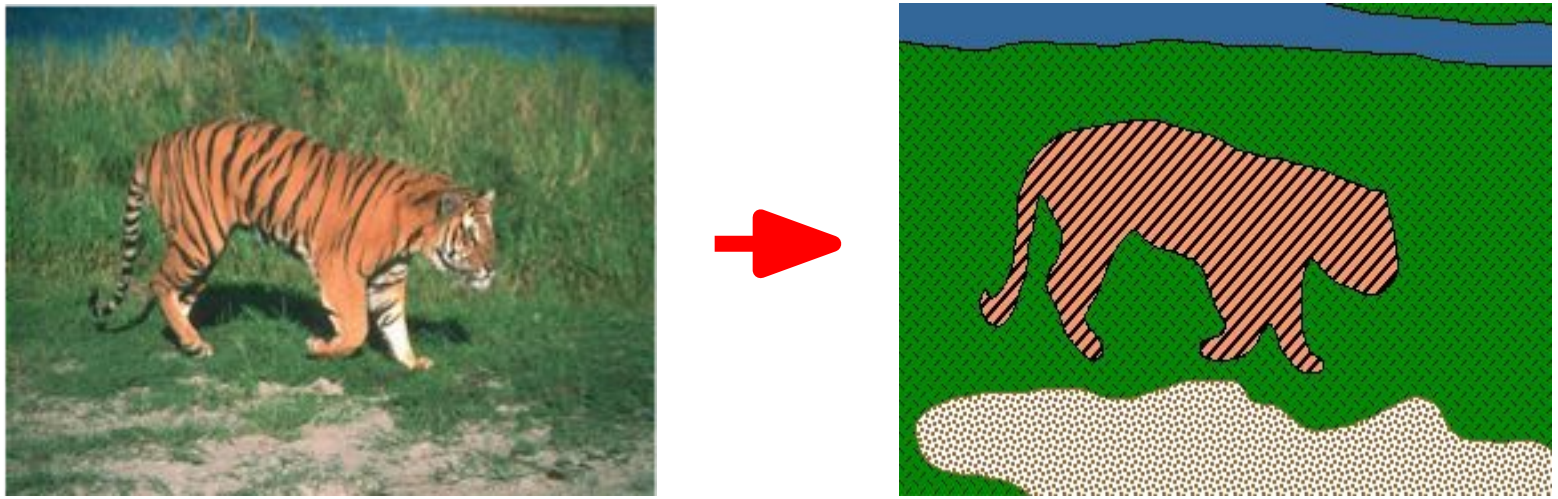
王兴刚
华中科技大学 电信学院

Outline

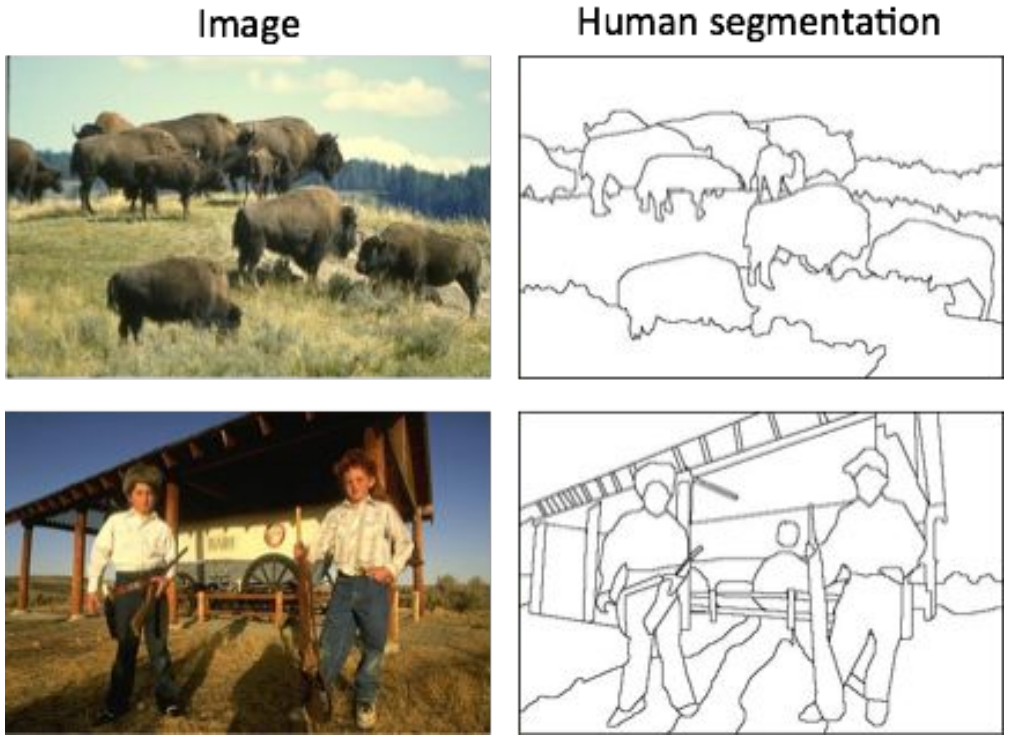
1. Introduction
2. Challenges
3. Some representative works
4. Perspective

Definition

Goal: identify groups of pixels that go together



Partition into coherent “objects”



Segmentation for efficiency

Group together similar-looking pixels for efficiency of further processing

[Felzenszwalb and Huttenlocher 2004]

[Achanta et al. 2005]

Segmentation as a result

"GrabCut" [Rother et al. 2004]

Scene parsing

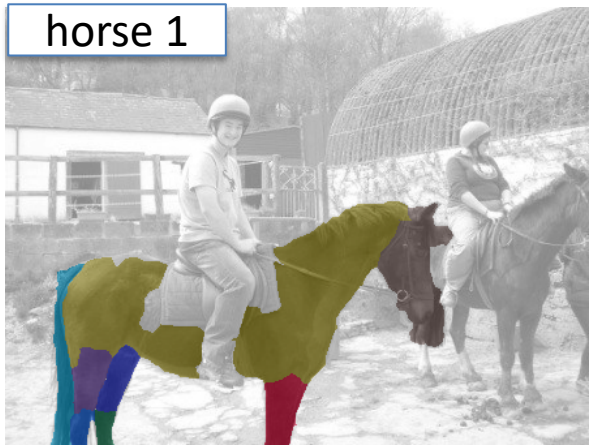
Label each pixel with a category label

Semantic segmentation

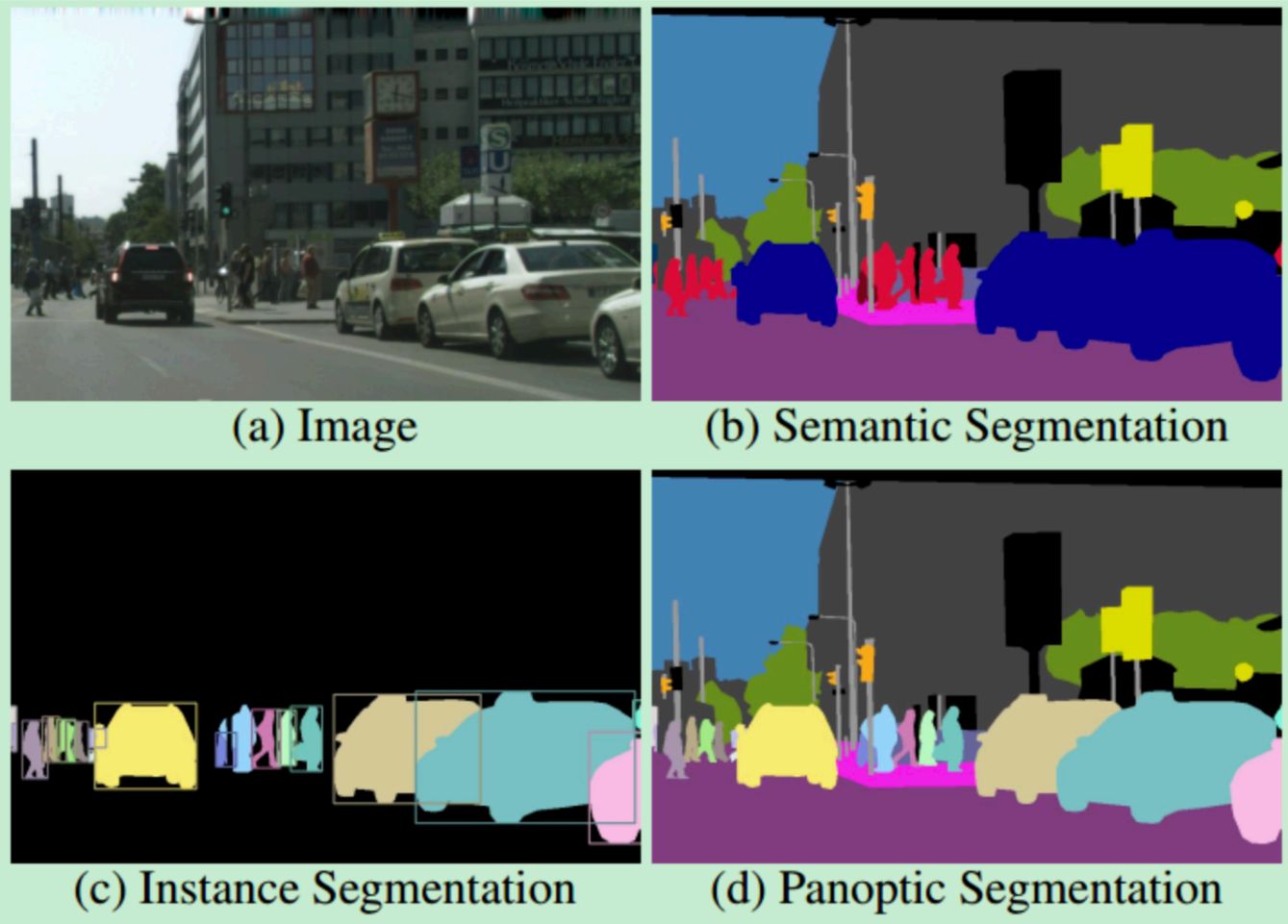
Label each pixel with an object category label

Segmentation and part labeling

Detect and segment every instance of the category in the image and label its parts



Panoptic segmentation



Evaluation measures and datasets

Evaluation measures

- Variation of information [Arbelaez et al., PAMI, 2011]
- Rand index [Arbelaez et al., PAMI, 2011]
- **Segmentation covering** $\mathcal{O}(R,R') = \frac{|R \cap R'|}{|R \cup R'|}$ $\mathcal{C}(S' \to S) = \frac{1}{N} \sum_{R \in S} |R| \cdot \max_{R' \in S'} \mathcal{O}(R,R')$ [Arbelaez et al., PAMI, 2011]
- **F-measure** = 2(P+R)/(P*R) for objects and parts [Pont-TUSET & Marques, PAMI, 2015]
- **AP/IoU** for semantic or instance segmentation

Popular datasets

- BSDS500 Dataset: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html
- PASCAL dataset: http://host.robots.ox.ac.uk/pascal/VOC/
- Cityscapes dataset: https://www.cityscapes-dataset.com/
- MVD dataset: https://eval-vistas.mapillary.com/

Outline

1. Introduction
2. Challenges
3. Some representative works
4. Perspective

Challenges

- Variant scenes and shapes
- Complex backgrounds and weak boundaries
- Textures
- Semantic segmentation and instance segmentation
- Panoptic segmentation

Outline

1. Introduction
2. Challenges
3. Some representative works
4. Perspective

Early merging and clustering methods

From 1980s to 2000s

- Segmentation using histograms and region merging [Beveridge et al., IJCV, 1989]
- Seeded region growing [Adams & Bischof, PAMI, 1994]
- Segmentation with geometric parametirc models [Leonardis et al., IJCV, 1995]
- Region competition [Zhu & Yuille, PAMI, 1996]
- **Mean shift [Comaniciu & Meer, PAMI, 2002]**
- …

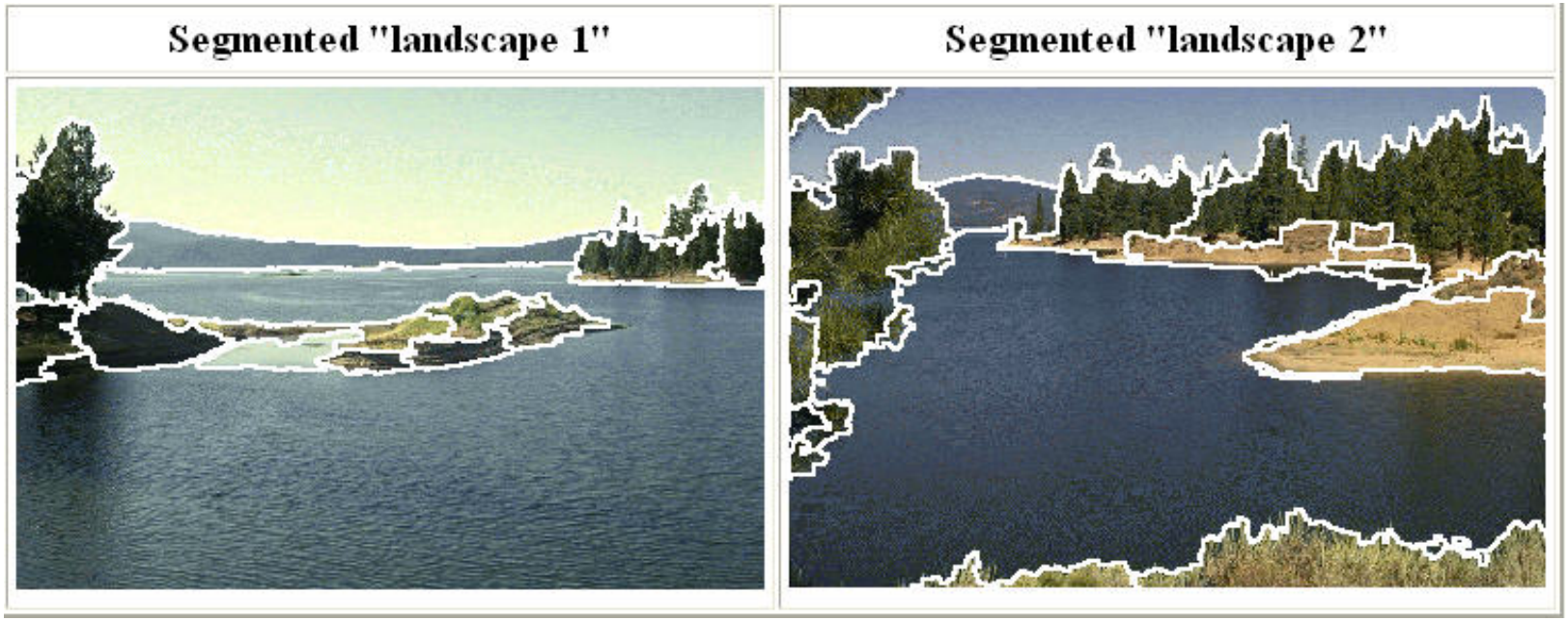
Key problems

1. Region model + merging criterion + merging order
2. Clustering method + number of clusters

Mean shift segmentation

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, PAMI 2002.

- Versatile technique for clustering-based segmentation

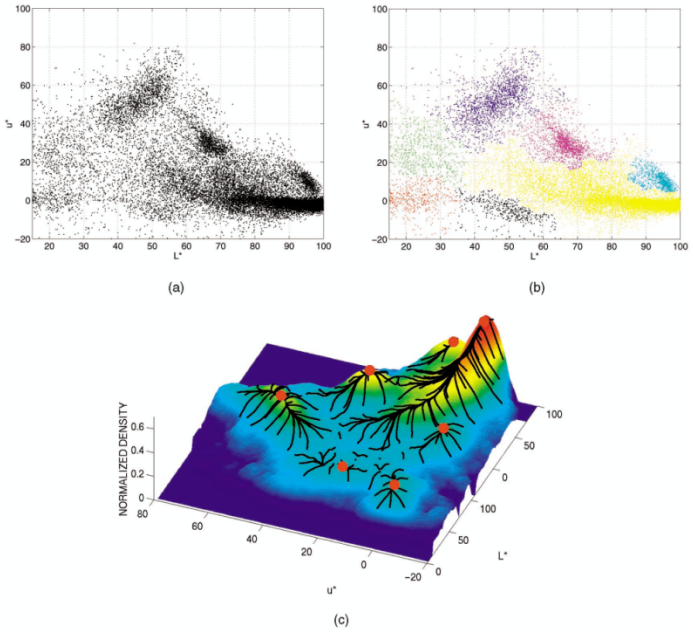


Mean shift clustering

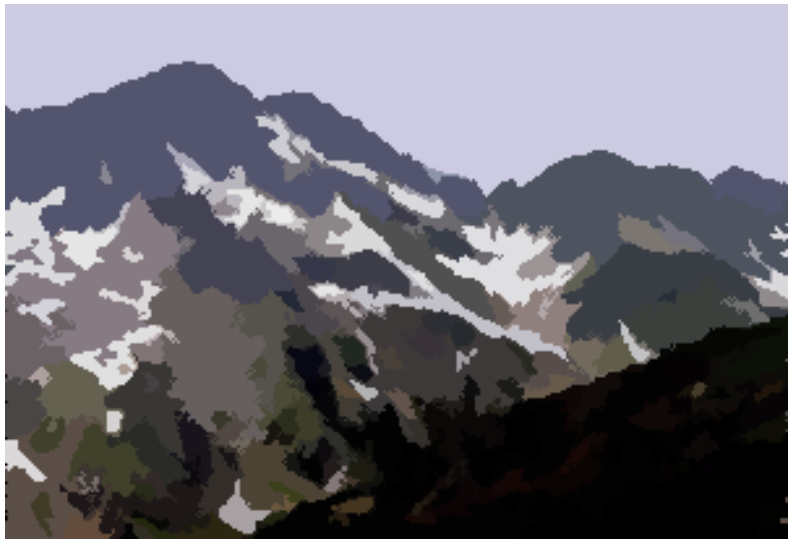
- The mean shift algorithm seeks *modes* of the given set of points
  1. Choose kernel and bandwidth
  2. For each point:
     a) Center a window on that point
     b) Compute the mean of the data in the search window
     c) Center the search window at the new mean location
     d) Repeat (b,c) until convergence
  3. Assign points that lead to nearby modes to the same cluster

Segmentation by mean shift

- Compute features for each pixel (color, gradients, texture, etc)
- Set kernel size for features K_f and position K_s
- Initialize windows at individual pixel locations
- Perform mean shift for each window until convergence
- Merge windows that are within width of K_f and K_s



Mean shift segmentation



Active contours

From 1980s to 2000s $E^*_{\text{snake}} = \int_0^1 E_{\text{int}}(\mathbf{v}(s)) + E_{\text{image}}(\mathbf{v}(s)) + E_{\text{con}}(\mathbf{v}(s))\, ds$

- Snakes: Active contour models [Kass et al., IJCV, 1988]
- On active contour models and balloons [Cohen, CVGIP, 1991]
- Geodesic active contours [Caselles et al., IJCV, 1997]
- Snakes, shapes, and gradient vector flow [Xu & Prince, ITIP, 1998]
- **Active contours without edges [Chan & Vese, ITIP, 2001]**
- …

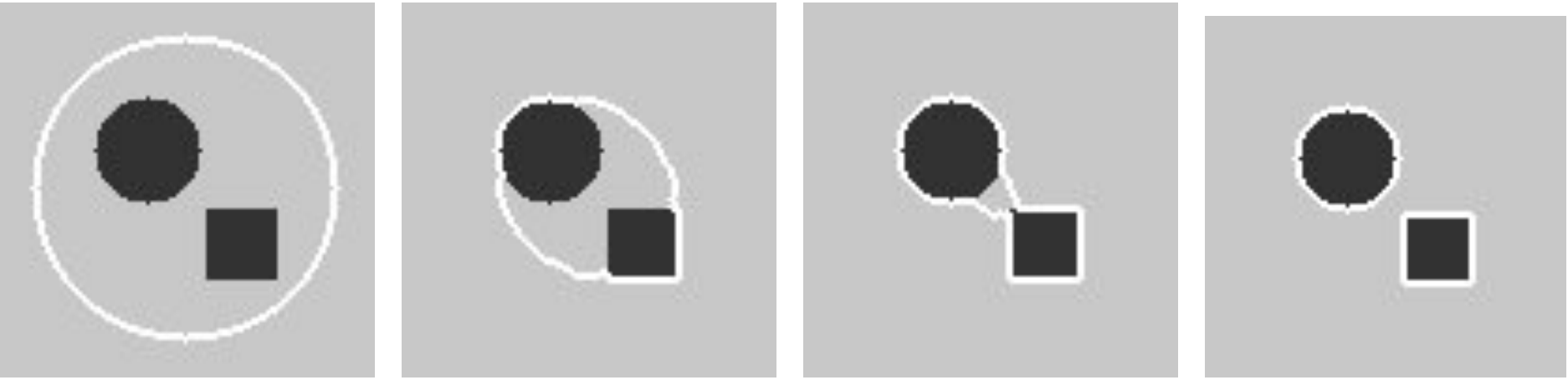
Key problems

1. Model definition + energy function
2. Initialization
3. Energy minimization

Active contour

- giving an image $u_0 : \Omega \to \Re$
- evolve a curve C to detect objects in u_0
- the curve has to stop on the boundaries of the objects

Initial Curve Evolutions Detected Objects



An active contour model without edges

Active contours without edges [Chan & Vese, ITIP, 2001]

Fitting + Regularization terms (length, area)

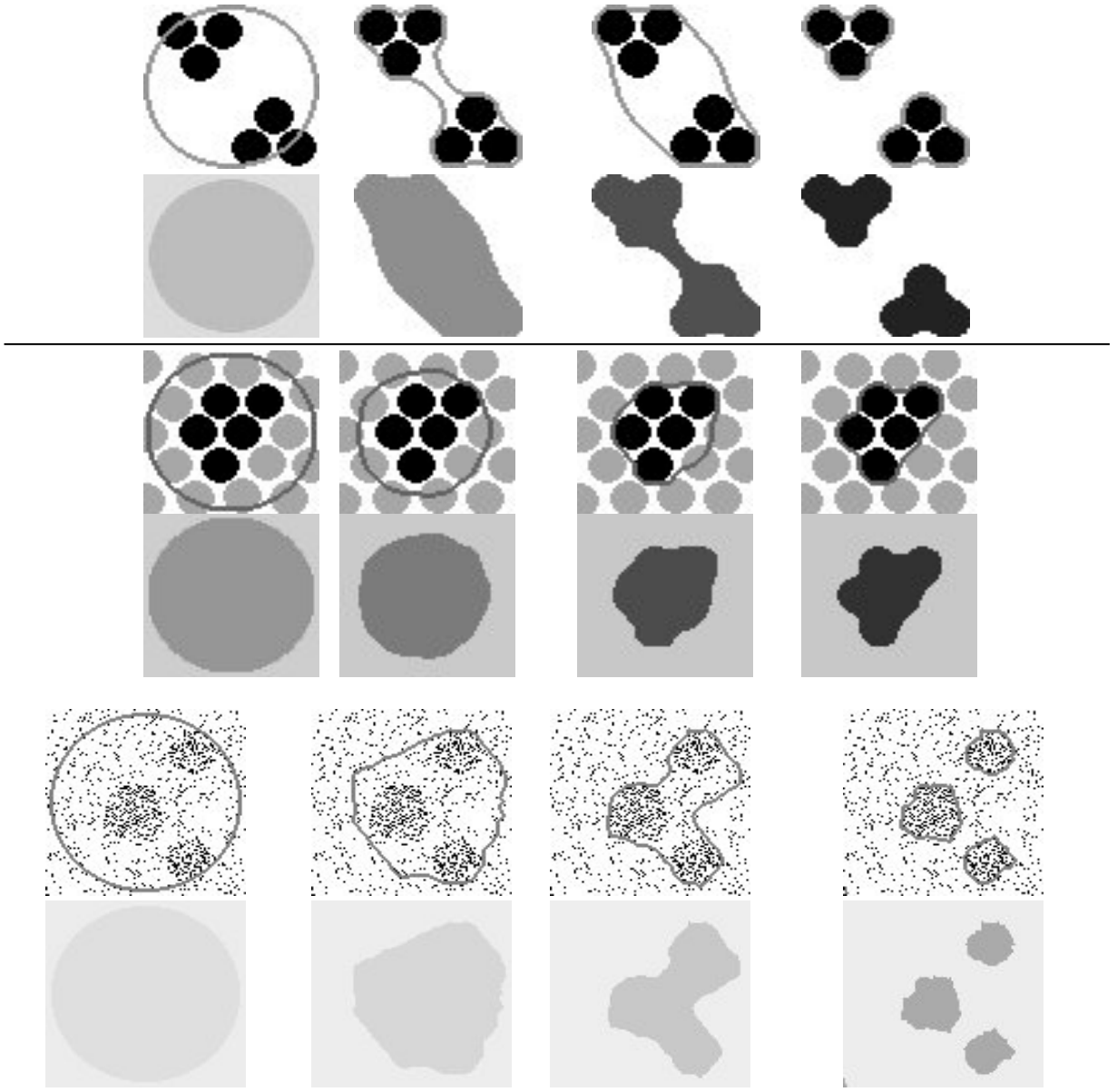
$$\inf_{c_1,c_2,C} F(c_1,c_2,C) = \mu\cdot|C| + \nu\cdot Area(inside(C))$$

$$+\lambda \int_{inside(C)} |u_0 - c_1|^2\, dxdy + \lambda \int_{outside(C)} |u_0 - c_2|^2\, dxdy$$

C = boundary of an open and bounded domain

$|C|$ = the length of the boundary-curve C

Some results



Variational approaches

From 1980s to 2000s $\mathcal{F}(u,C) = \int_{\Omega}(u-u_0)^2 dx + \mu \int_{\Omega\setminus C} |\nabla(u)|^2 dx + \nu|C|$

- **Mumford-Shah model** **[Mumford & Shah, CPAM, 1989]**
- Variational Methods in Image Segmentation [Morel & Solimini, BOOK, 1995]
- Level set methods and fast marching methods [Sethian., BOOK, 1999]
- Total variation segmentation [Bertelli et al., ICCV, 2009]
- …

Key problems

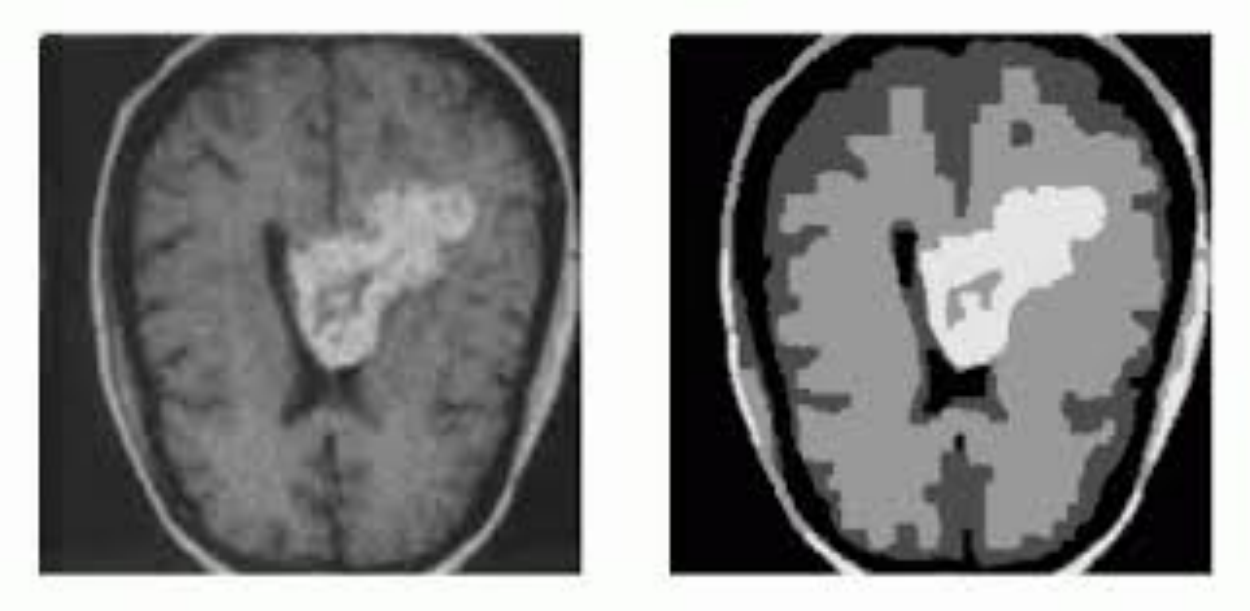
1. Model definition + energy function
2. Global energy minimization

Mumford-Shah Segmentation

$$F(f,K) = \frac{1}{2}\int_{\Omega}(f-g)^2 dA + \beta\int_{\Omega/K}|\nabla f|^2 dA + \alpha\int_K d\sigma$$

fidelity to image | gradients within segments | total edge length

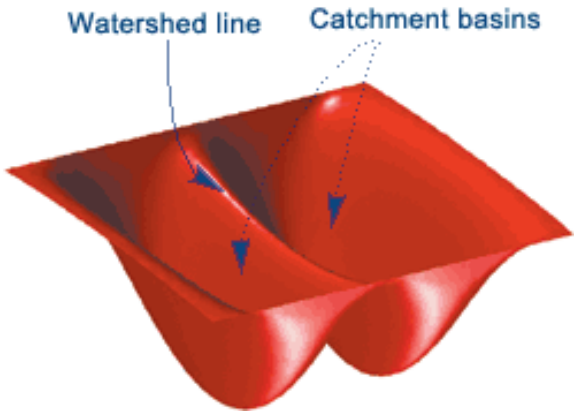
Ω: image domain K: edge set f: segmented image g: observed image



Watersheds

From 1991s to 2010s

- Watersheds in digital spaces [Vincent & Soille., PAMI, 1991]
- **Geodesic watershed [Najman & Schmitt, PAMI, 1996]**
- Topological watershed [Couprie et al., JMIV, 2005]
- Watershed cuts [Cousty et al., PAMI, 2009]
- Power watershed [Couprie et al., PAMI, 2011]
- gpb-owt-ucm [Arbelaez et al., PAMI, 2011]
- …



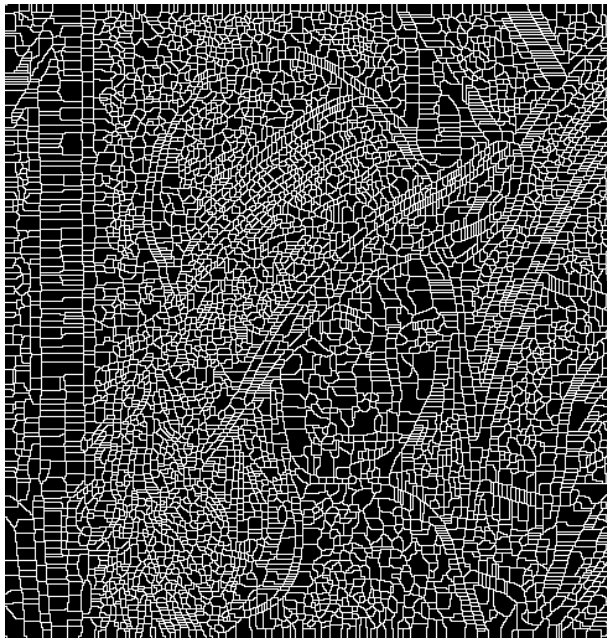
Key problems

1. Leakage problem in gradient image
2. Too many local minima => too many regions

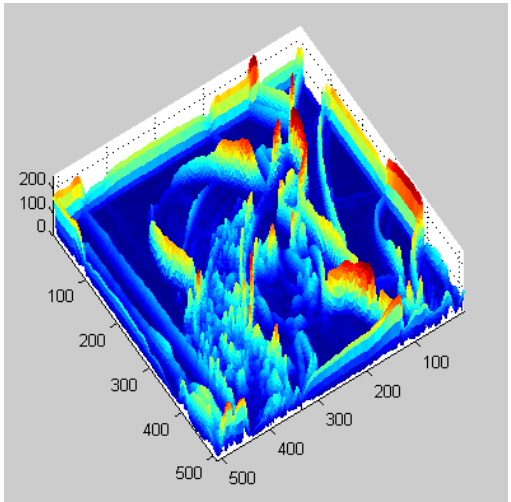
Watershed segmentation

Image

Gradient



Watershed boundaries



Meyer's watershed segmentation

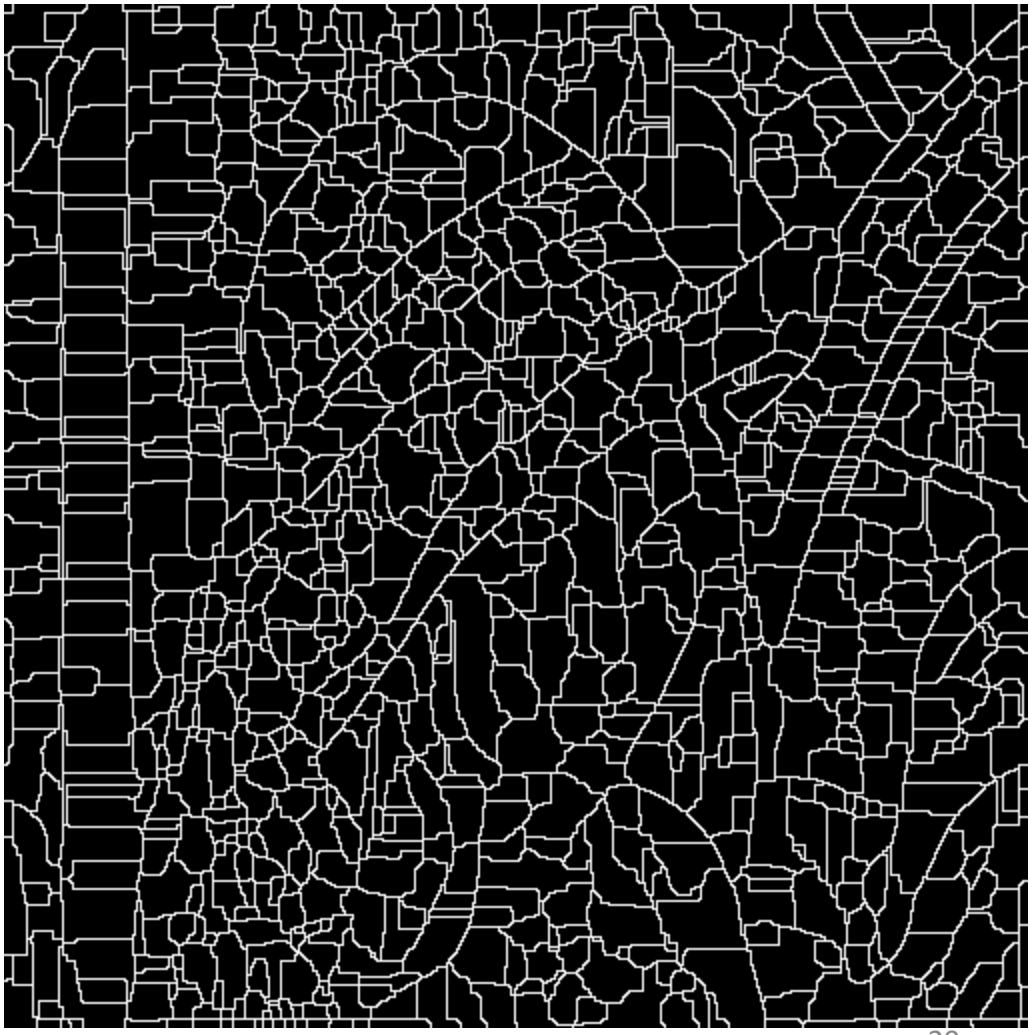
1. Choose local minima as region seeds
2. Add neighbors to priority queue, sorted by value
3. Take top priority pixel from queue
    1. If all labeled neighbors have same label, assign that label to pixel
    2. Add all non-marked neighbors to queue
4. Repeat step 3 until finished (all remaining pixels in queue are on the boundary)

Matlab: seg = watershed(bnd_im)

Meyer 1991

Simple trick

- Use Gaussian or median filter to reduce number of regions



Watershed usage

- Use as a starting point for hierarchical segmentation
  - Ultrametric contour map (Arbelaez 2006)

- Works with any soft boundaries
  - Pb (w/o non-max suppression)
  - Canny (w/o non-max suppression)
  - Etc.

Watershed pros and cons

- Pros
  - Fast (< 1 sec for 512x512 image)
  - Preserves boundaries
- Cons
  - Only as good as the soft boundaries (which may be slow to compute)
  - Not easy to get variety of regions for multiple segmentations
- Usage
  - Good algorithm for superpixels, hierarchical segmentation

Segmentation with graphical models

From 2000s to 2010s

$$E(f) = \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q) + \sum_{p \in \mathcal{P}} D_p(f_p)$$

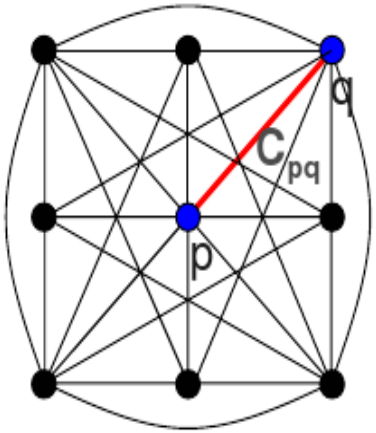
- **Normalized cuts [Shi & Malik, PAMI, 2000]**
- MRF for brain MR image segmentation [Zhang et al., ITMI, 2001]
- CRF for image segmentation [Lafferty et al., ICML, 2001]
- Graph cuts [Boykov et al., PAMI, 2001]
- GrabCut [Rother et al., TOG, 2004]
- Random walks for image segmentation [Grady, PAMI, 2006]
- …

Key problems

1. Definition of potential functions
2. Energy minimization

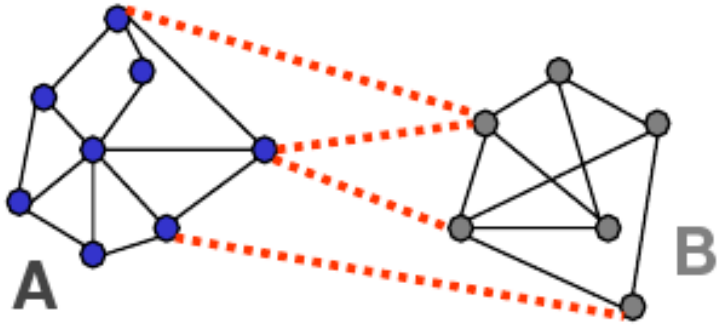
Normalized-Cuts [Jianbo Shi & Malik, PAMI, 2000]

- Image is modelled as a fully connected graph
- Each link between nodes (pixels) associated with a cost C_{pq} measures similarity inversely proportional to difference in feature



Find Cut that minimizes the cost function

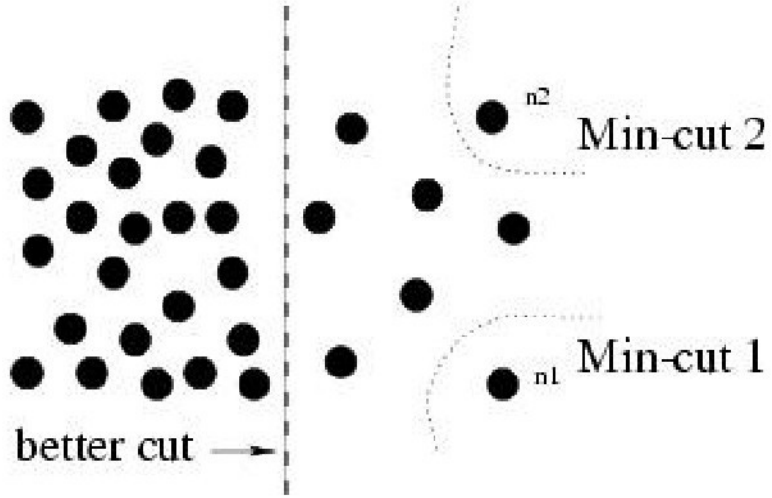
$$cut(A,B) = \sum_{p \in A, q \in B} c_{p,q}$$



However large segments are penalized, so fix by normalizing for size of segments

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}$$

$$assoc(A,V) = \sum_{u \in A, t \in V} c(u,t)$$



Learning-based segmentation

From 2000s to 2010s

- CRF for image segmentation [Lafferty et al., ICML, 2001]
- TextonBoost [Shotton et al., IJCV, 2009]
- **gpb-owt-ucm [Arbelaez et al., PAMI, 2011]**
- …

Key problems

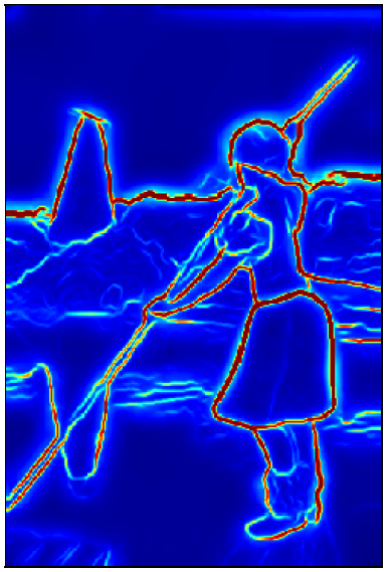
1. Integrating low-level features
2. Optimization

Contour detection and hierarchical image segmentation

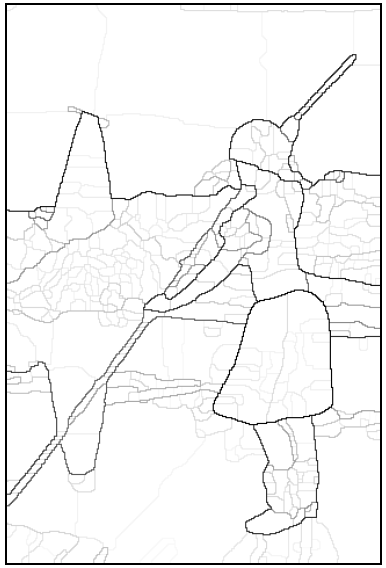
[Arbelaez et al., PAMI, 2011]

- Goal
  - Contour Detection
  - Hierarchical Segmentation from Contours

Original Image

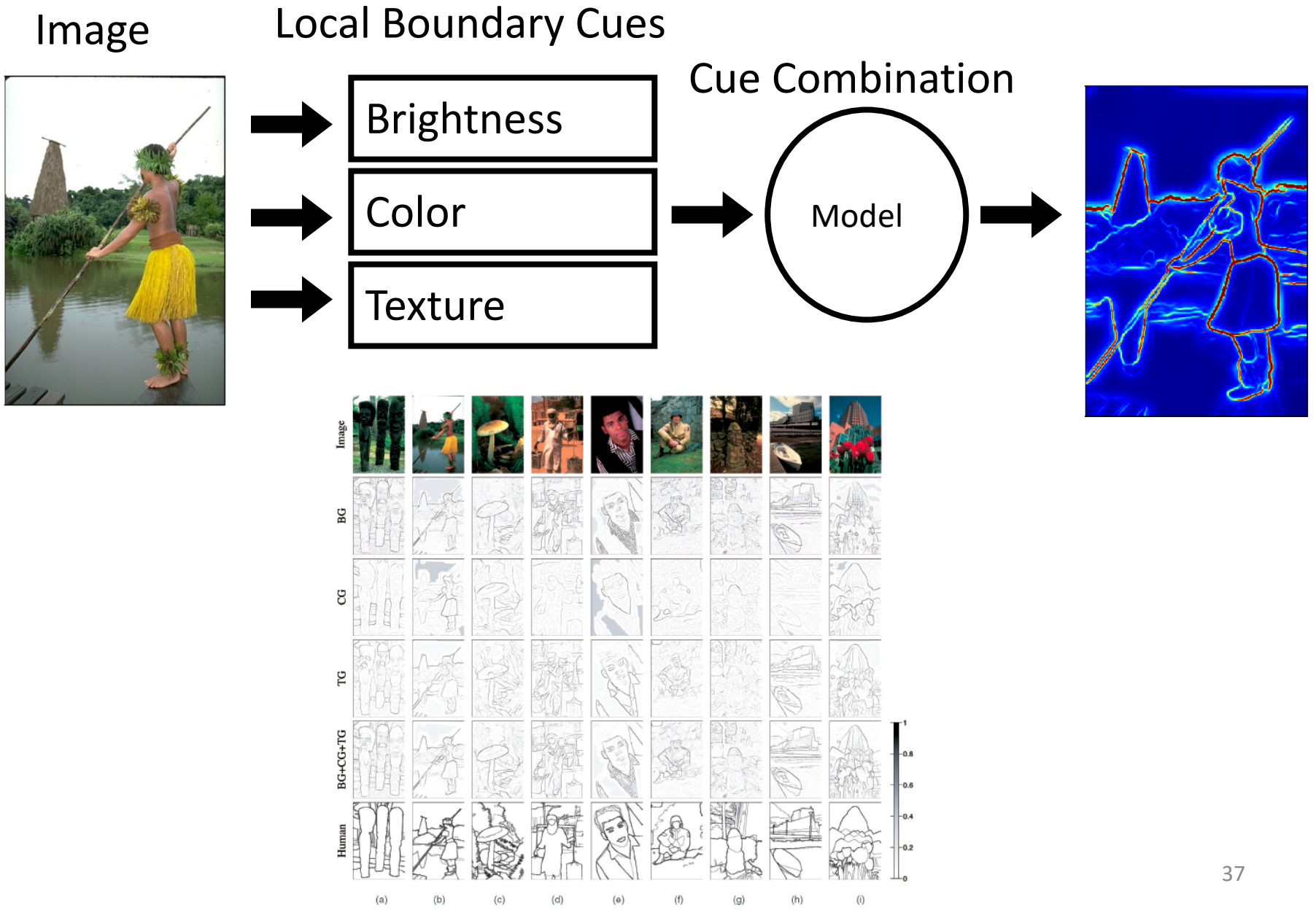


Contour



Segmentation

Contour Detection

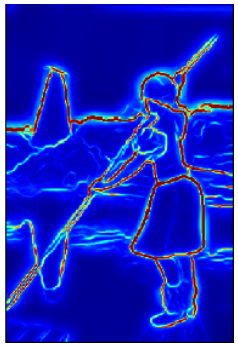


Pipeline of hierarchical Segmentation from Contours

Original Image

- Local cues
- Global cues

Oriented Gradient of histograms



Contour

Oriented Watershed Transform

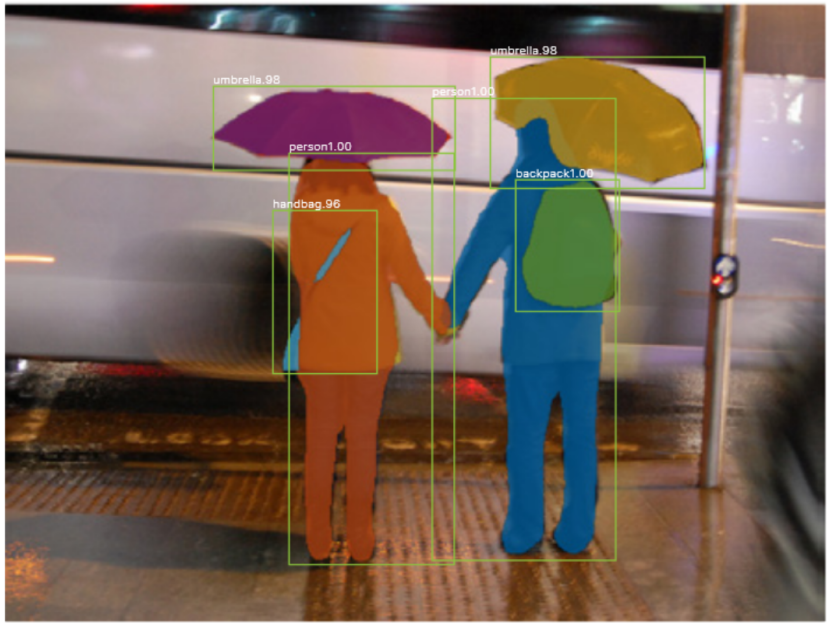
Hierarchical Segmentation

Iterative Merging

Some results

Limitations of previous methods

- Semantically meaningful object detection
  - Requires object-level (high level) information



So far: Image Classification

This image is CC0 public domain

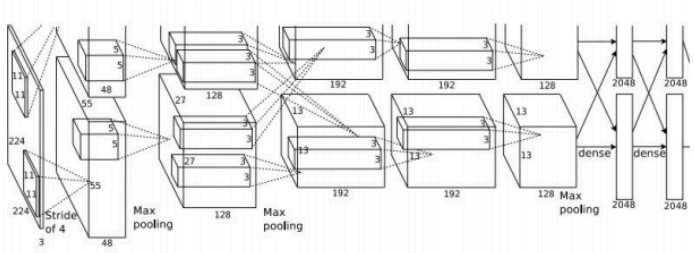


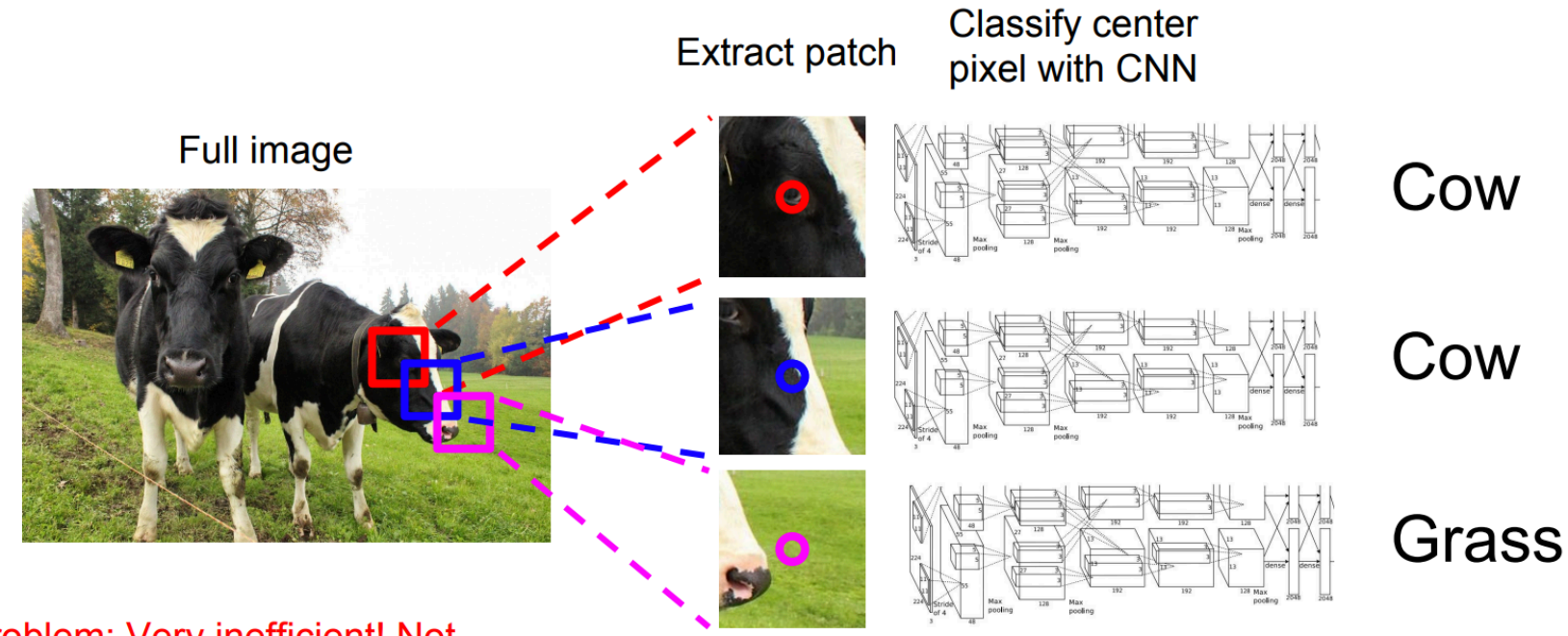
Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Vector:
4096

Fully-Connected:
4096 to 1000

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Semantic Segmentation Idea: Sliding Window



Problem: Very inefficient! Not reusing shared features between overlapping patches

Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

Semantic Segmentation Idea: Fully Convolutional

Design a network as a bunch of convolutional layers to make predictions for pixels all at once!

Input:
3 x H x W

Conv

Conv

Conv

Conv

argmax

Scores:
C x H x W

Predictions:
H x W

Convolutions:
D x H x W

Problem: convolutions at original image resolution will be very expensive ...

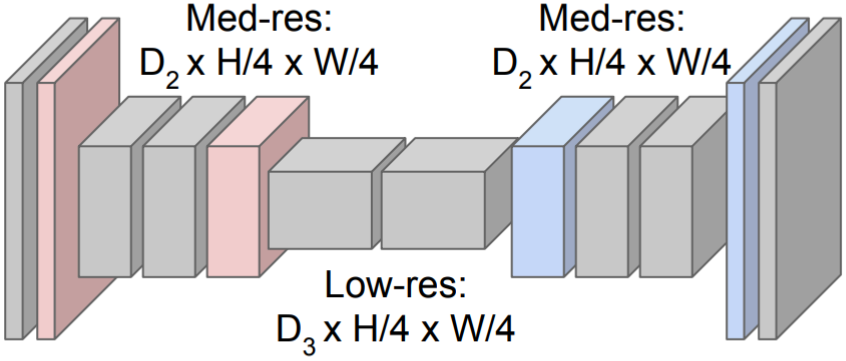
Semantic Segmentation Idea: Fully Convolutional

Downsampling:
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling:
???

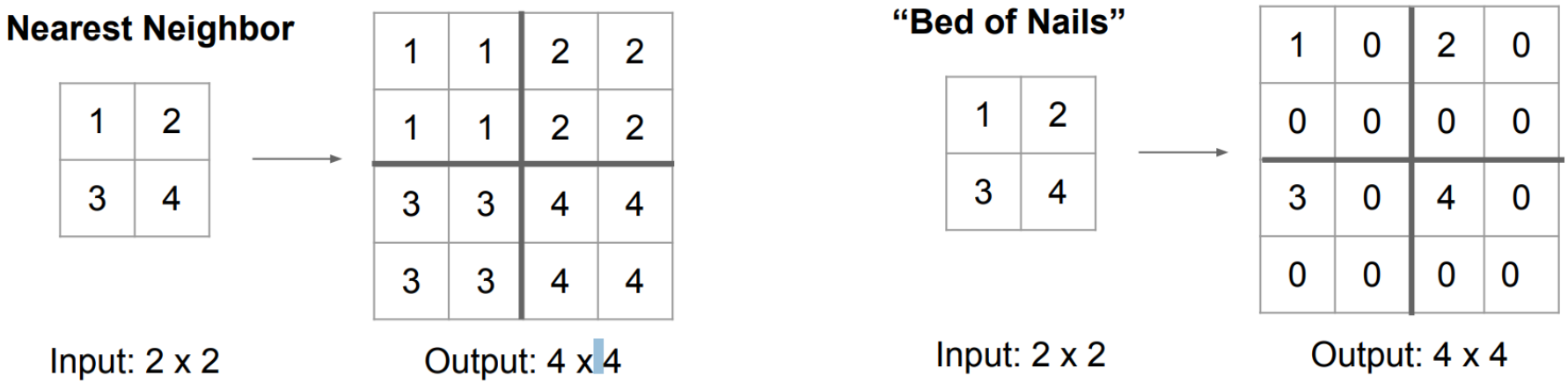
Input:
3 x H x W



Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

In-Network upsampling: “Unpooling”



In-Network upsampling: “Max Unpooling”

Max Pooling
Remember which element was max!

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8

Input: 4 x 4

5	6
7	8

Output: 2 x 2

…

Rest of the network

Max Unpooling
Use positions from pooling layer

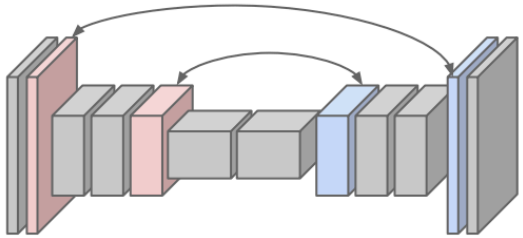
1	2
3	4

Input: 2 x 2

0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4

Output: 4 x 4

Corresponding pairs of downsampling and upsampling layers



Recent CNN based methods

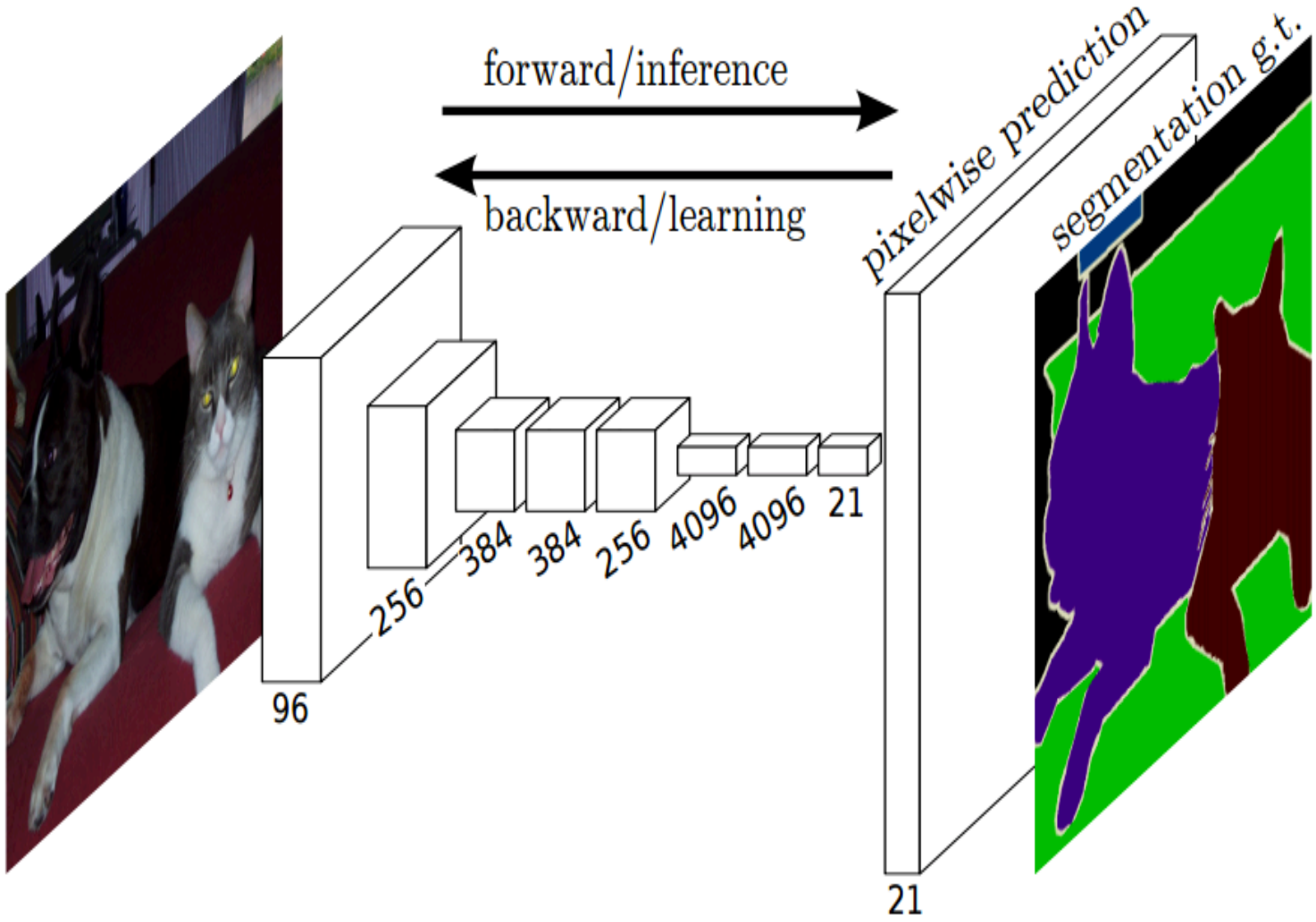
Manly since 2013

- Learning hierarchical features for scene labeling [Farabet et al., PAMI, 2013]
- **FCN for semantic segmentation [Long et al., CVPR, 2015]**
- Hypercolumns for object segmentation [Hariharan et al., CVPR, 2015]
- **DeepLab [Chen et al., ARXIV, 2016]**
- SegNet [Badrinarayanan et al., PAMI, 2017]
- **Mask R-CNN [He et al., ARXIV, 2017]**
- …

Key problems

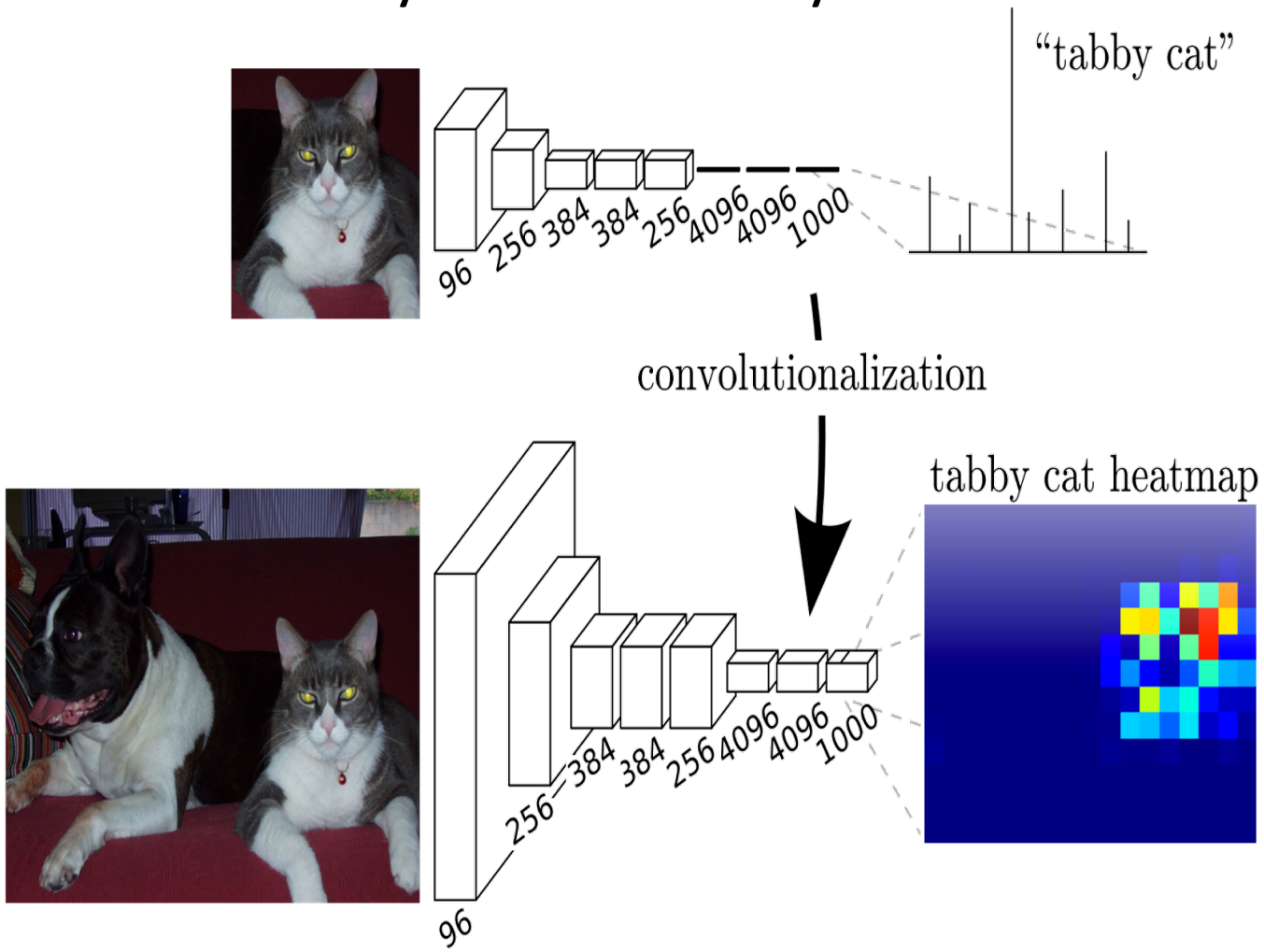
1. Network architecture
2. Adapted loss function

Fully Convolutional Networks [Long et al., CVPR, 2015]



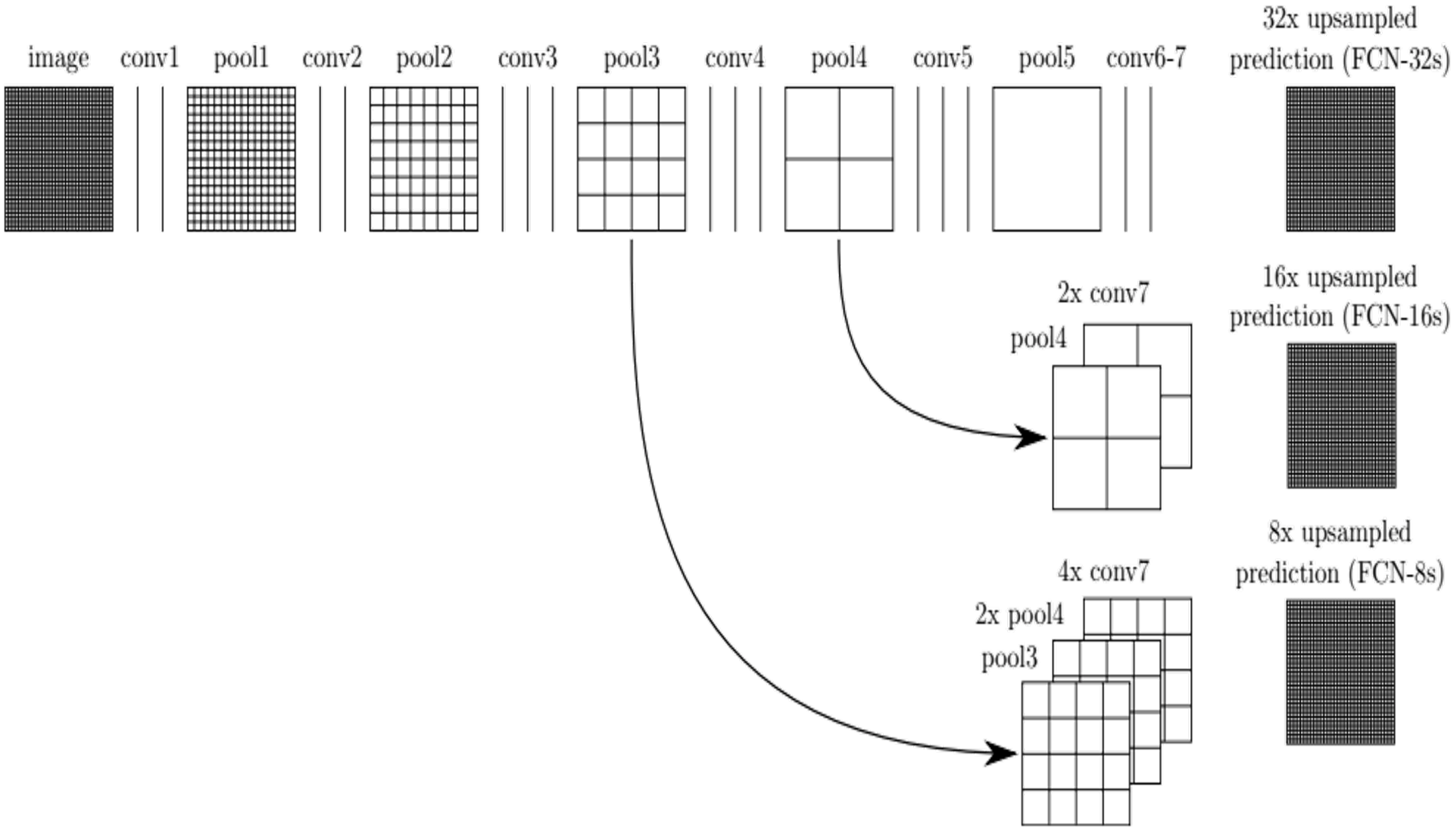
Fully Convolutional Networks [Long et al., CVPR, 2015]

- Can convert fully-connected layers to convolutional.

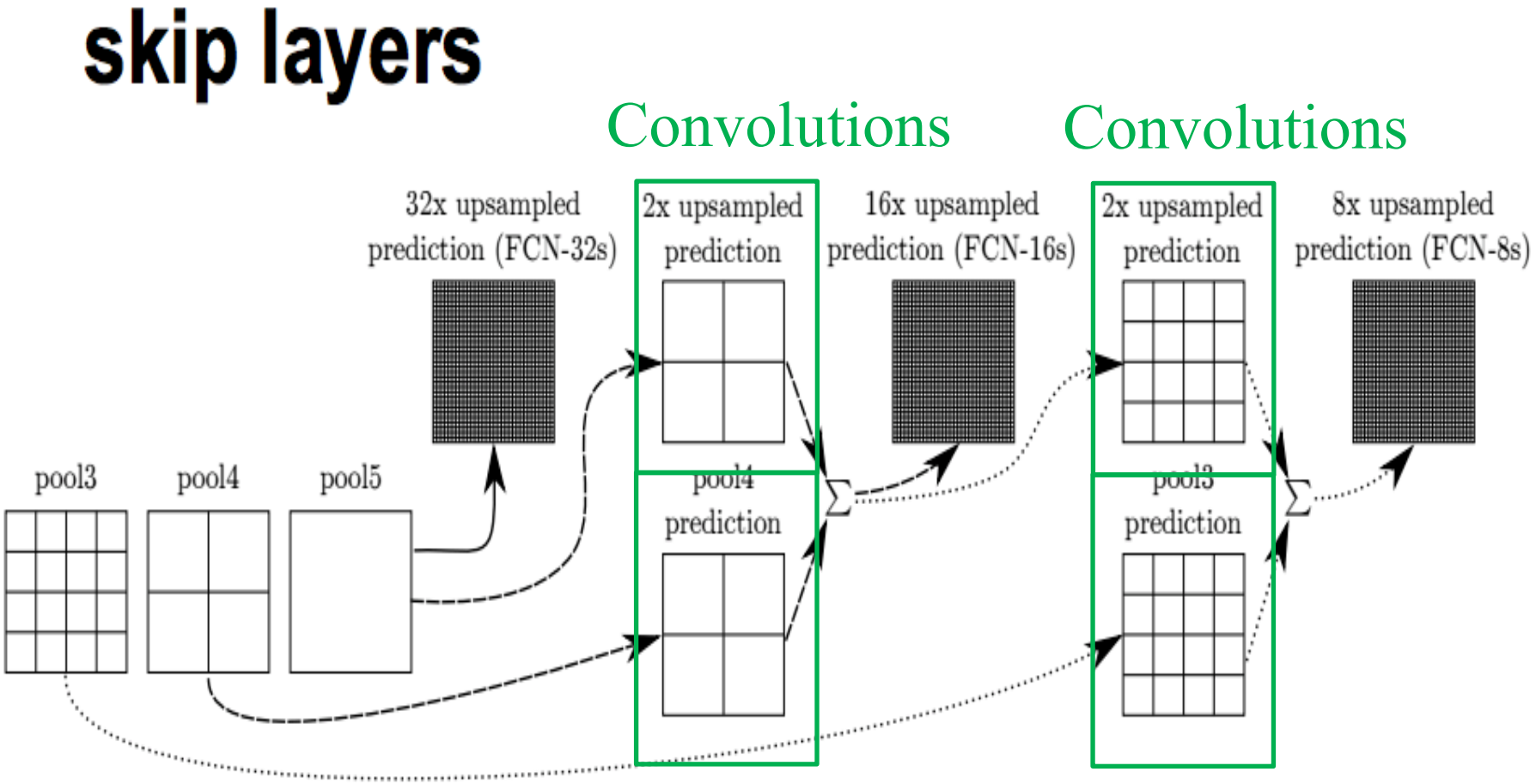


Fully Convolutional Networks [Long et al., CVPR, 2015]

Deconvolution



Fully Convolutional Networks [Long et al., CVPR, 2015]



Fully Convolutional Networks [Long et al., CVPR, 2015]

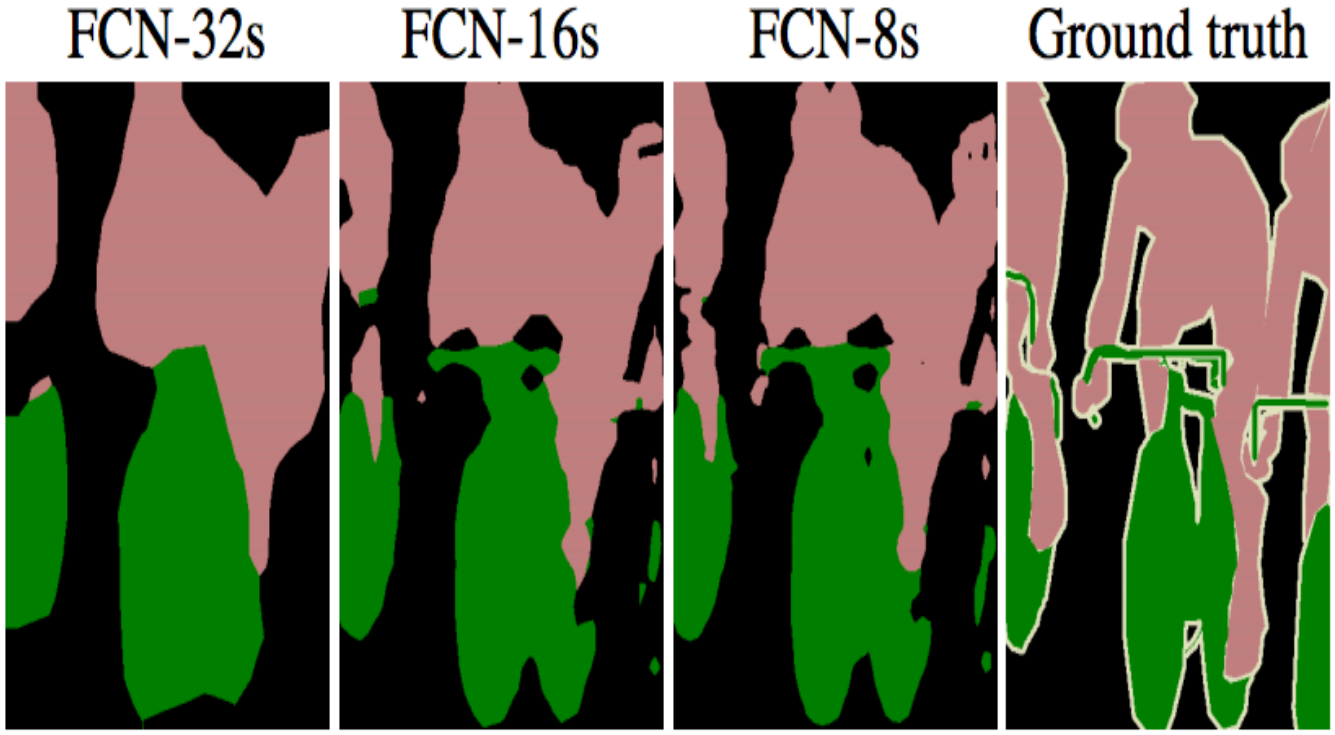
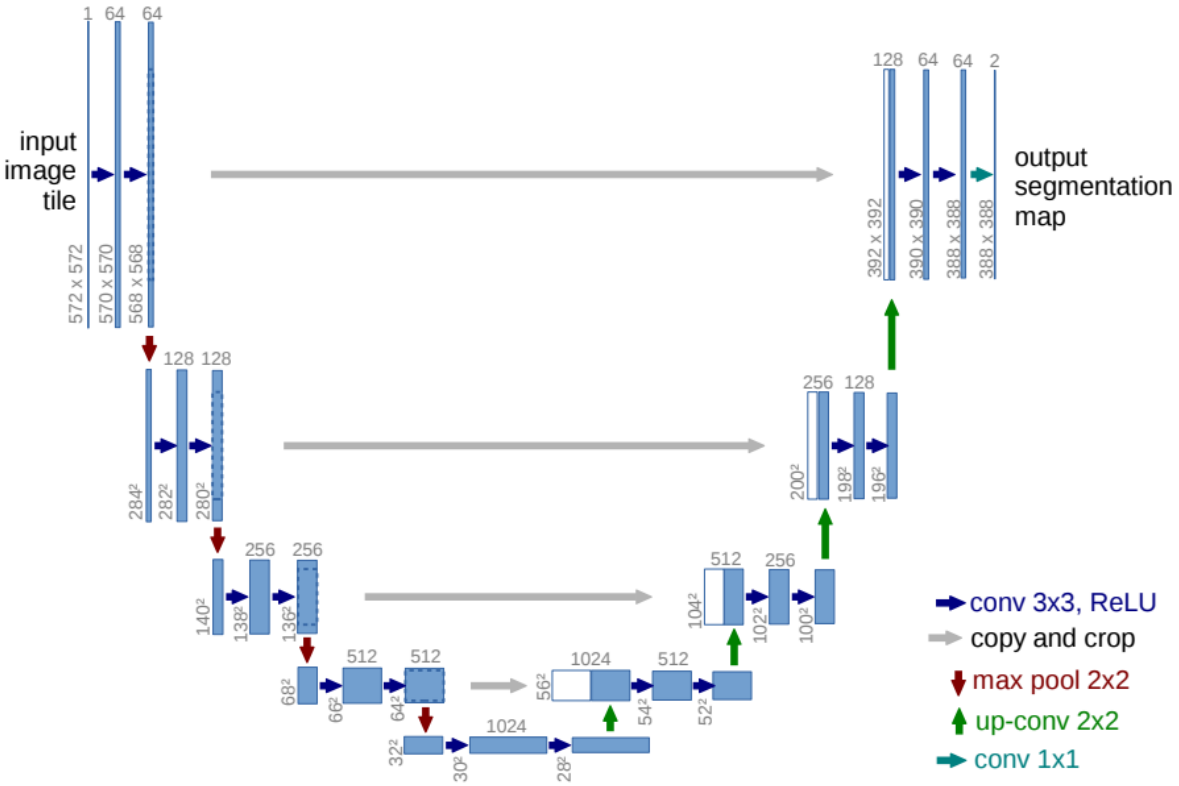


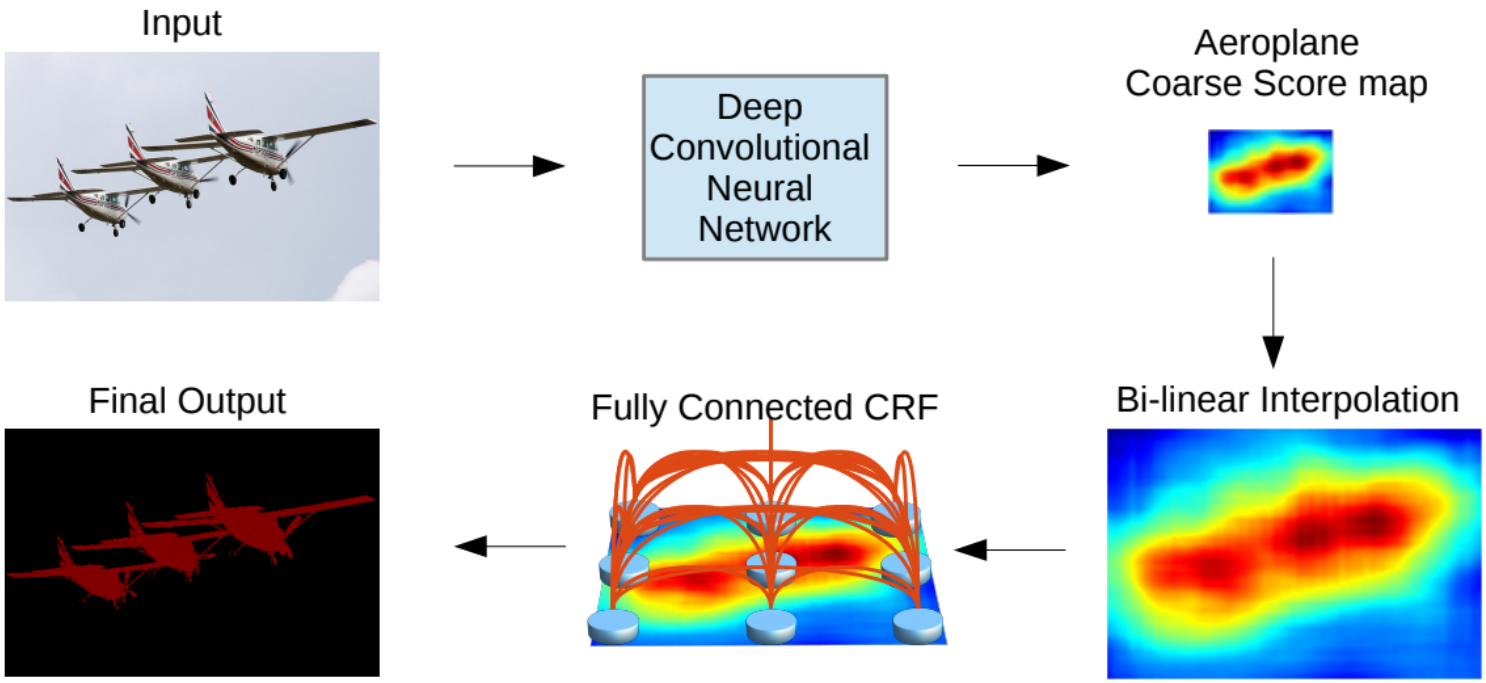
Figure 4. Refining fully convolutional nets by fusing information from layers with different strides improves segmentation detail. The first three images show the output from our 32, 16, and 8 pixel stride nets (see Figure 3).

UNet [Ronneberger et al., MICCAI, 2015]



The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization.

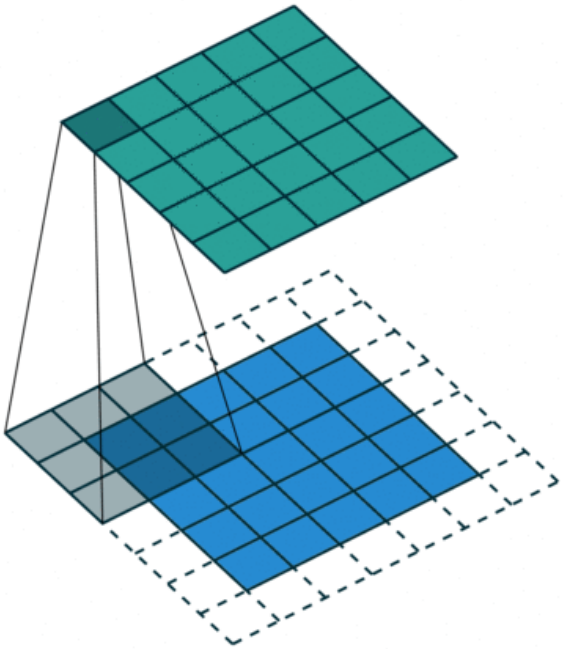
DeepLab v1 [Chen et al., ICLR, 2015]



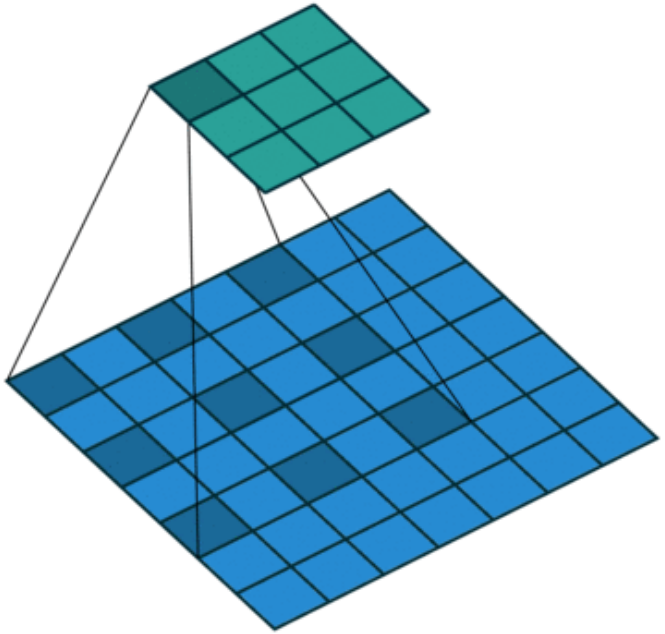
Model Illustration
DCNN for classification
Refine prediction with conditional random field(CRF)

DeepLab v1 [Chen et al., ICLR, 2015]

Dilated Convolution: control receptive field



Kernel Size: 3
Dilation Rate: 1
Stride: 1
Padding: 1

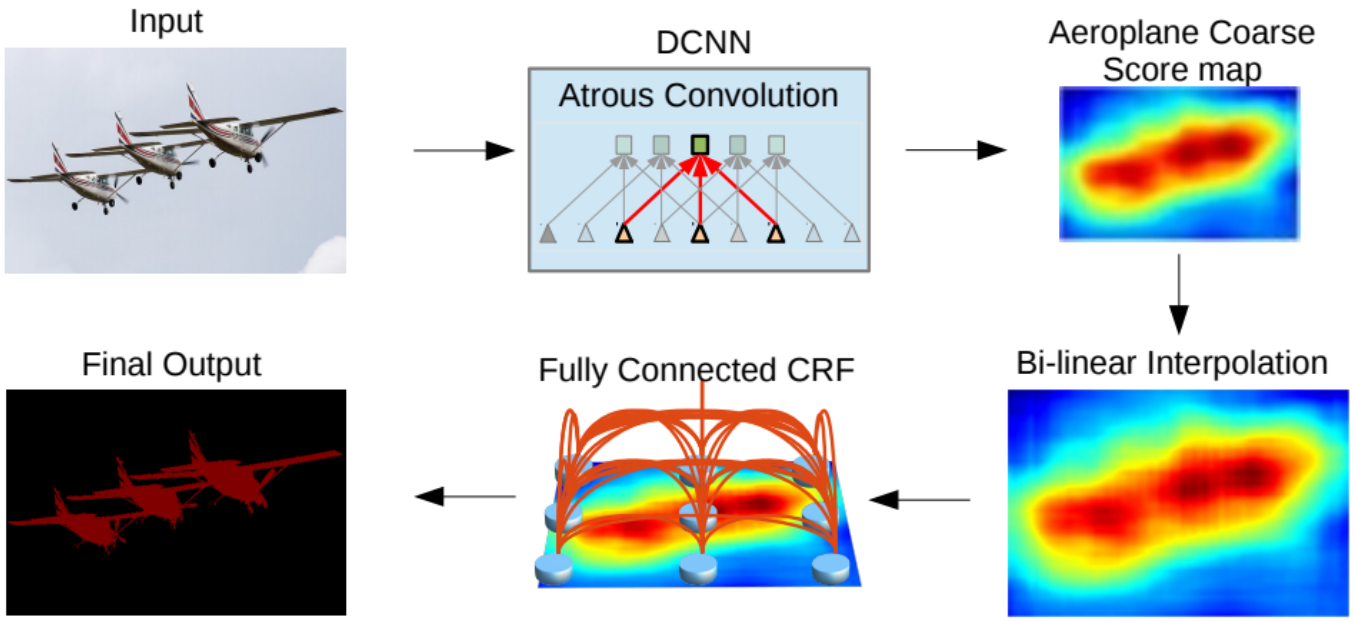


Kernel Size: 3
Dilation Rate: 2
Stride: 1
Padding: 0

DeepLab v2 [Chen et al., TPAMI, 2017]

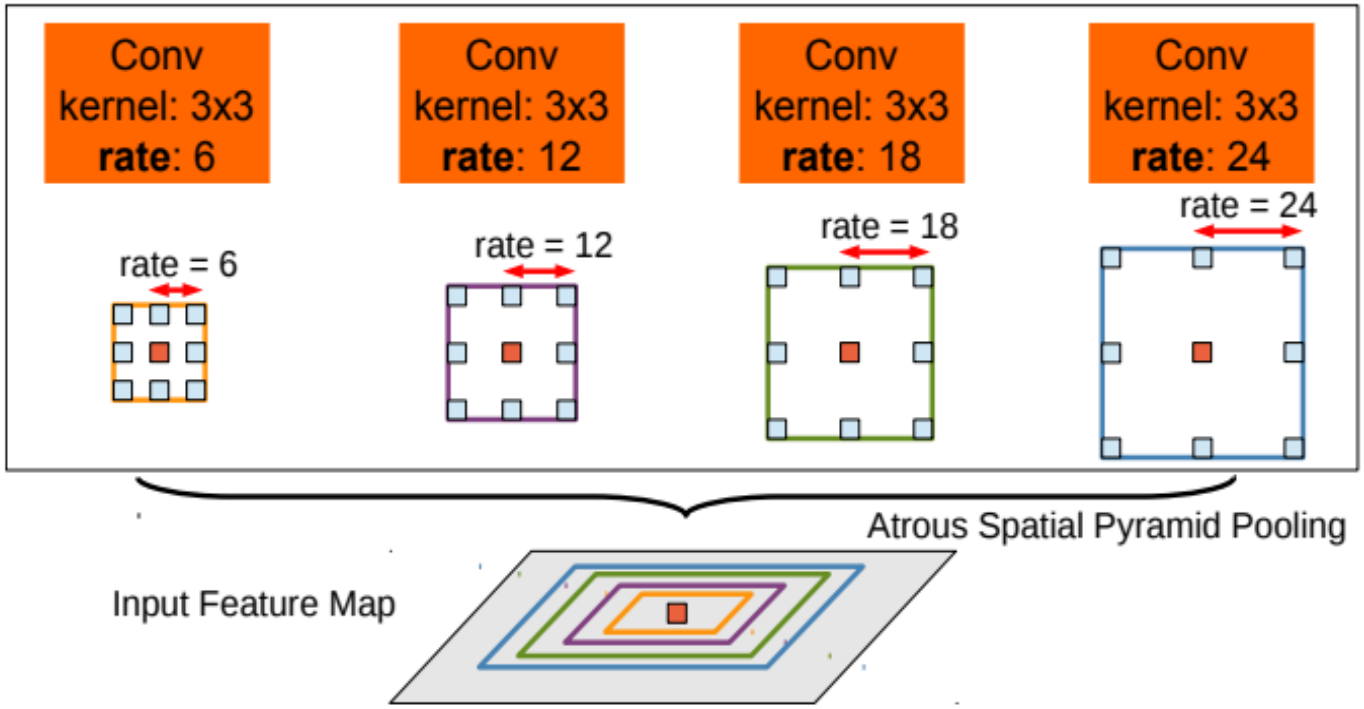
Improvements compared to DeepLab v1:

- Better segmentation of objects at multiple scales (using ASPP)
- Adapting ResNet image classification DCNN
- Learning rate policy



DeepLab v2 [Chen et al., TPAMI, 2017]

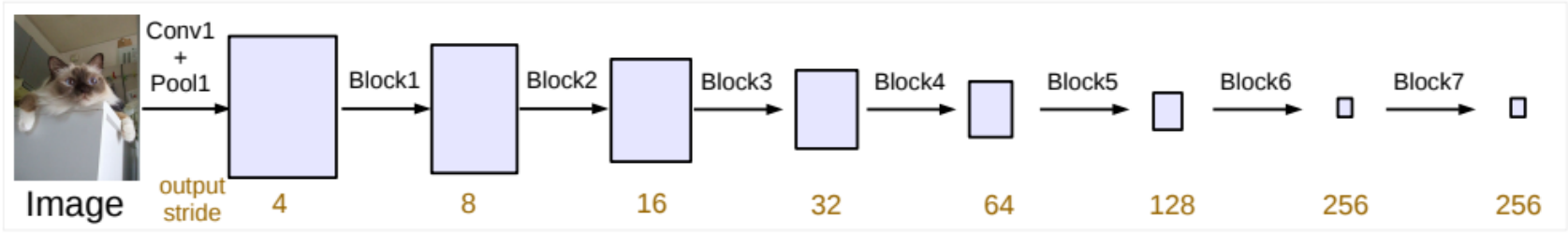
Atrous Spatial Pyramid Pooling (ASPP)



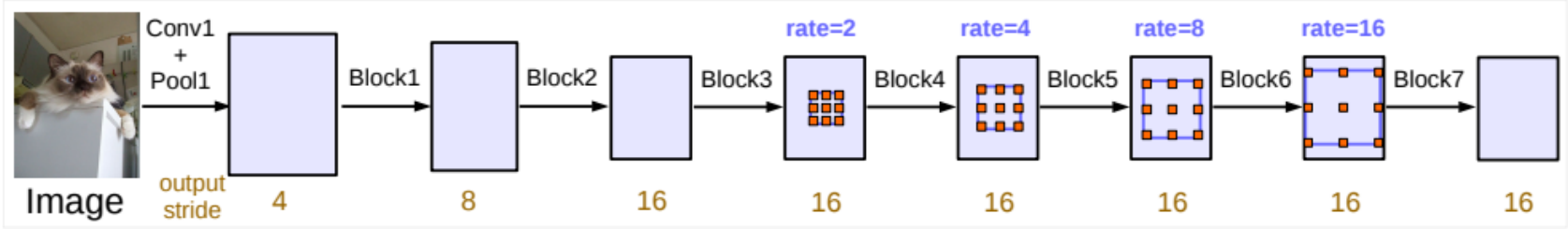
DeepLab v3 [Chen et al., ARXIV, 2017]

Changes compared to DeepLab V1 & DeepLab V2:

- The proposed framework is general and could be applied to any network
- Batch normalization is included within ASPP
- Copy last ResNet block, and arrange in cascade
- CRF is not used



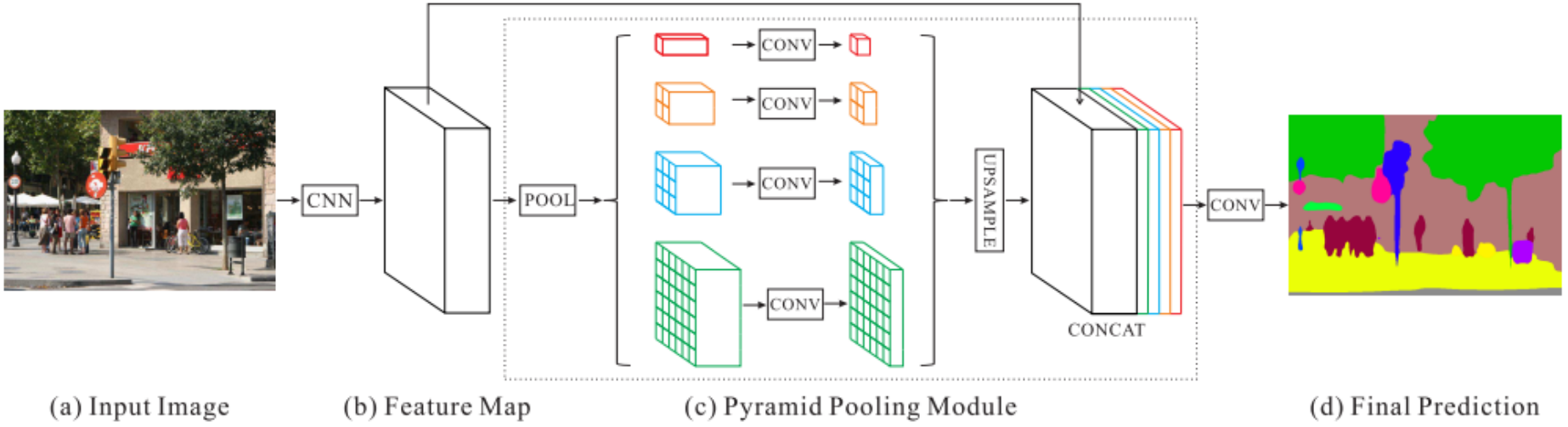
(a) Going deeper without atrous convolution.



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.

PSPNet [Zhao et al., ARXIV, 2017]



a pyramid scene parsing network to embed difficult scenery context features in an FCN based pixel prediction framework.

DeepLab v3 [Chen et al., ARXIV, 2017]

ASPP module:

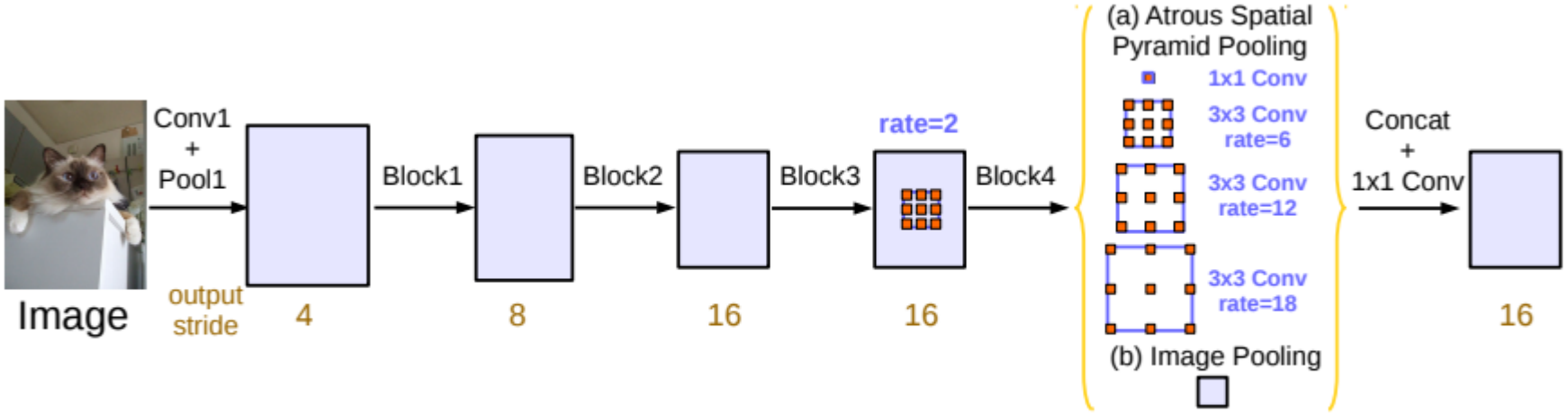


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

Best result includes:

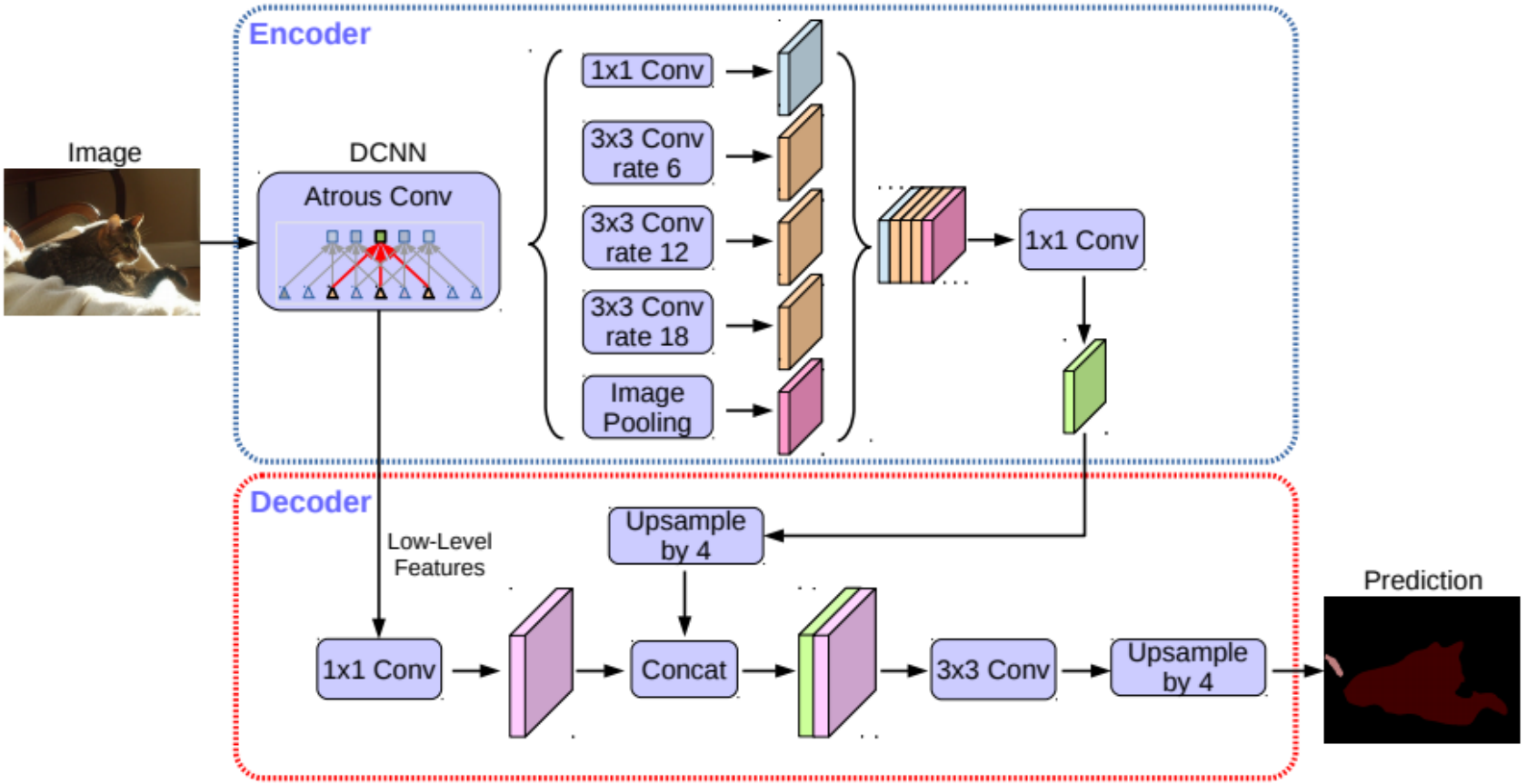
- ASPP
- Output stride of 8
- Flip and rescale augmentation

Outperforms DeepLab v2(77.69%)

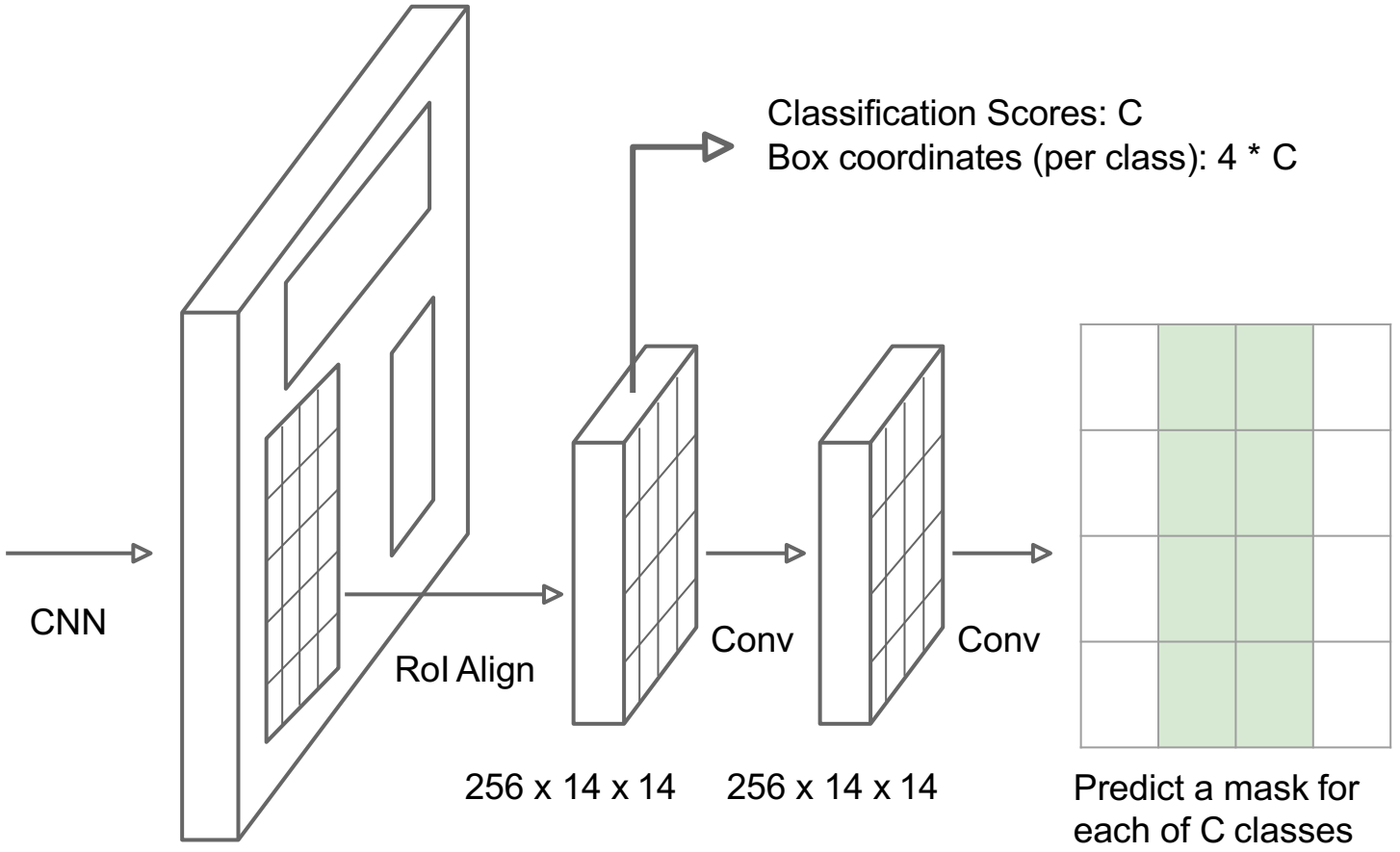
Method	OS=16	OS=8	MS	Flip	mIOU
MG(1, 2, 4) +	✓				77.21
ASPP(6, 12, 18) +		✓			78.51
Image Pooling		✓	✓		79.45
		✓	✓	✓	79.77

DeepLab v3 Plus [Chen et al., ECCV, 2018]

Extend DeepLab v3 with a simple yet effective decoder

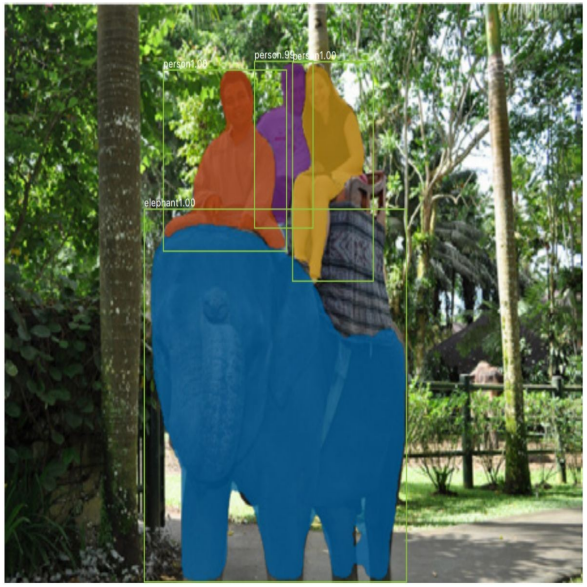


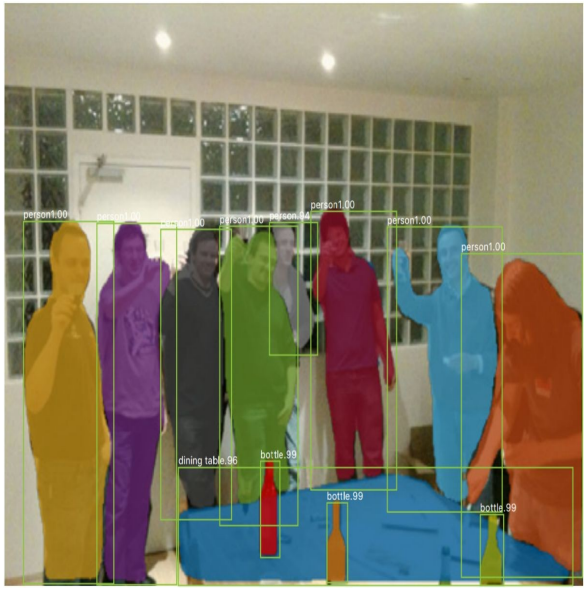
Mask R-CNN



He et al, “Mask R-CNN”, arXiv 2017

Mask R-CNN: Very Good Results!





He et al, “Mask R-CNN”, arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.
Reproduced with permission.

State-of-the-art results

on BSDS500 [Maninis et al., PAMI, 2017]

Method	Regions - F_{op} ODS	OIS	AP
COB (Ours)	**0.419**	**0.478**	**0.343**
LEP [47]	0.417	0.468	0.334
MCG [17]	0.380	0.433	0.271
ISCRA [48]	0.352	0.418	0.275
UCM [16]	0.348	0.385	0.235
MShift [50]	0.229	0.292	0.122
NCuts [41]	0.213	0.270	0.096
EGB [49]	0.158	0.240	0.080

On Pascal VOC 2012 [Hayder et al., CVPR, 2017]

VOC 2012 (val)	mAP (0.5)	mAP (0.7)	time/img (s)
SDS [14]	49.7	25.3	48
PFN [22]	58.7	42.5	~ 1
Hypercolumn [15]	60.0	40.4	>80
InstanceFCN [6]	61.5	43.0	1.50
MNC [8]	63.5	41.5	0.36
MNC-new	65.01	46.23	0.42
BAIS - insideBBox (ours)	64.97	44.58	0.75
BAIS - full (ours)	**65.69**	**48.30**	0.78

on Cityscapes dataset [He et al., ARXIV, 2017]

	training data	AP [val]	AP	AP_{50}	person	rider	car	truck	bus	train	mcycle	bicycle
InstanceCut [23]	fine + coarse	15.8	13.0	27.9	10.0	8.0	23.7	14.0	19.5	15.2	9.3	4.7
DWT [4]	fine	19.8	15.6	30.0	15.1	11.7	32.9	17.1	20.4	15.0	7.9	4.9
SAIS [17]	fine	-	17.4	36.7	14.6	12.9	35.7	16.0	23.2	19.0	10.3	7.8
DIN [3]	fine + coarse	-	20.0	38.8	16.5	16.7	25.7	20.6	30.0	23.4	17.1	10.1
Mask R-CNN	fine	31.5	26.2	49.9	30.5	23.7	46.9	22.8	32.2	18.6	19.1	16.0
Mask R-CNN	fine + COCO	**36.4**	**32.0**	**58.1**	**34.8**	**27.0**	**49.1**	**30.1**	**40.9**	**30.9**	**24.1**	**18.7**

Outline

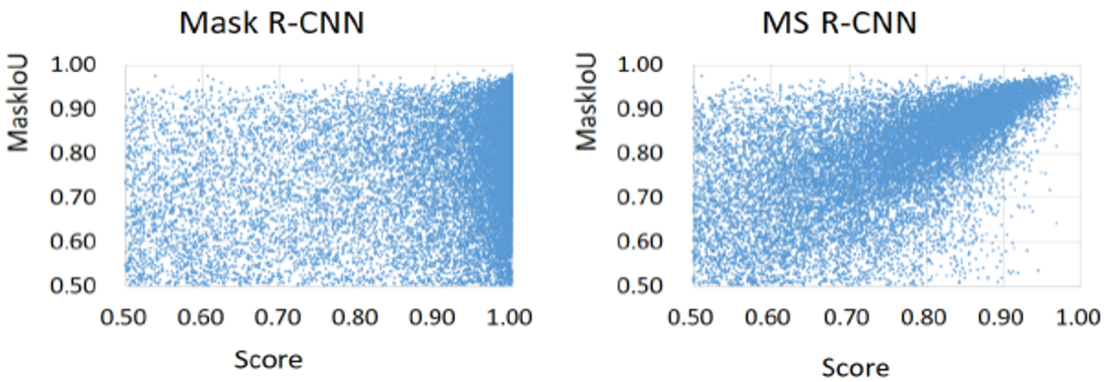
1. Introduction
2. Challenges
3. Some representative works
4. Perspective

Perspective

- Instance segmentation
- Deep learning for general image segmentation
- Weakly superwised segmentation
- Medical image segmentation
- 3D image segmentation and video segmentation

Mask Scoring R-CNN

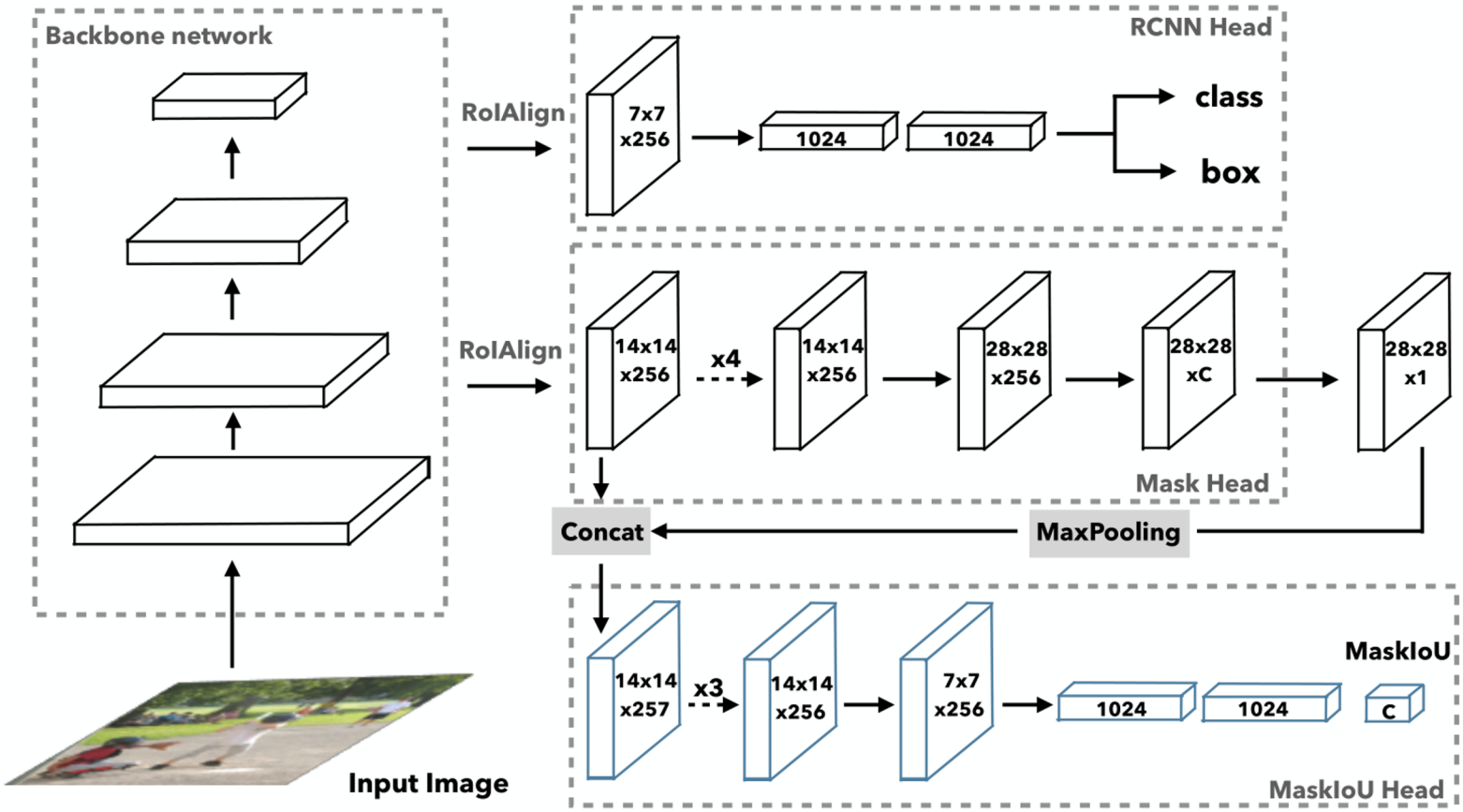
The mask quality, IoU between the predicted mask and its ground truth mask (**called MaskIoU**), is usually **not well correlated** with the mask sc



Learning a calibrated mask score according to MaskIoU for every detection hypothesis.

Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, Xinggang Wang. **Mask Scoring R-CNN. CVPR**, 2019. **Oral (5.6% acceptance rate)** arXiv:1903.00241

Mask Scoring R-CNN



The network - extending Mask R-CNN by adding a branch called MaskIoU head for predicting MaskIoU.

Mask Scoring R-CNN

Backbone	MaskIoU head	AP_m	AP_m@0.5	AP_m@0.75	AP_b	AP_b@0.5	AP_b@0.75
ResNet-18 FPN		27.7	46.9	29.0	31.2	50.4	33.2
	✓	29.3	46.9	31.3	31.5	50.8	33.5
ResNet-50 FPN		34.5	55.8	36.7	38.6	59.2	42.5
	✓	36.0	55.8	38.8	38.6	59.2	42.5
ResNet-101 FPN		36.6	58.6	39.0	41.3	61.7	45.9
	✓	38.2	58.4	41.5	41.4	61.8	46.3

Method	Backbone	AP	AP@0.5	AP@0.75	AP_S	AP_M	AP_L
MNC [7]	ResNet-101	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [23]	ResNet-101	29.2	49.5	-	-	-	-
FCIS+++ [23]	ResNet-101	33.6	54.5	-	-	-	-
Mask R-CNN [15]	ResNet-101	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN [15]	ResNet-101 FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN [15]	ResNeXt-101 FPN	37.1	60.0	39.4	16.9	39.9	53.5
MaskLab [3]	ResNet-101	35.4	57.4	37.4	16.9	38.3	49.2
MaskLab+ [3]	ResNet-101	37.3	59.8	36.6	19.1	40.5	50.6
MaskLab+ [3]	ResNet-101 (JET)	38.1	61.1	40.4	19.6	41.6	51.4
Mask R-CNN	ResNet-101	34.3	55.0	36.6	13.2	36.4	52.2
MS R-CNN		35.4	54.9	38.1	13.7	37.6	53.3
Mask R-CNN	ResNet-101 FPN	37.0	59.2	39.5	17.1	39.3	52.9
MS R-CNN		38.3	58.8	41.5	17.8	40.4	54.4
Mask R-CNN	ResNet-101 DCN+FPN	38.4	61.2	41.2	18.0	40.5	55.2
MS R-CNN		39.6	60.7	43.1	18.8	41.5	56.2

Mask Scoring R-CNN

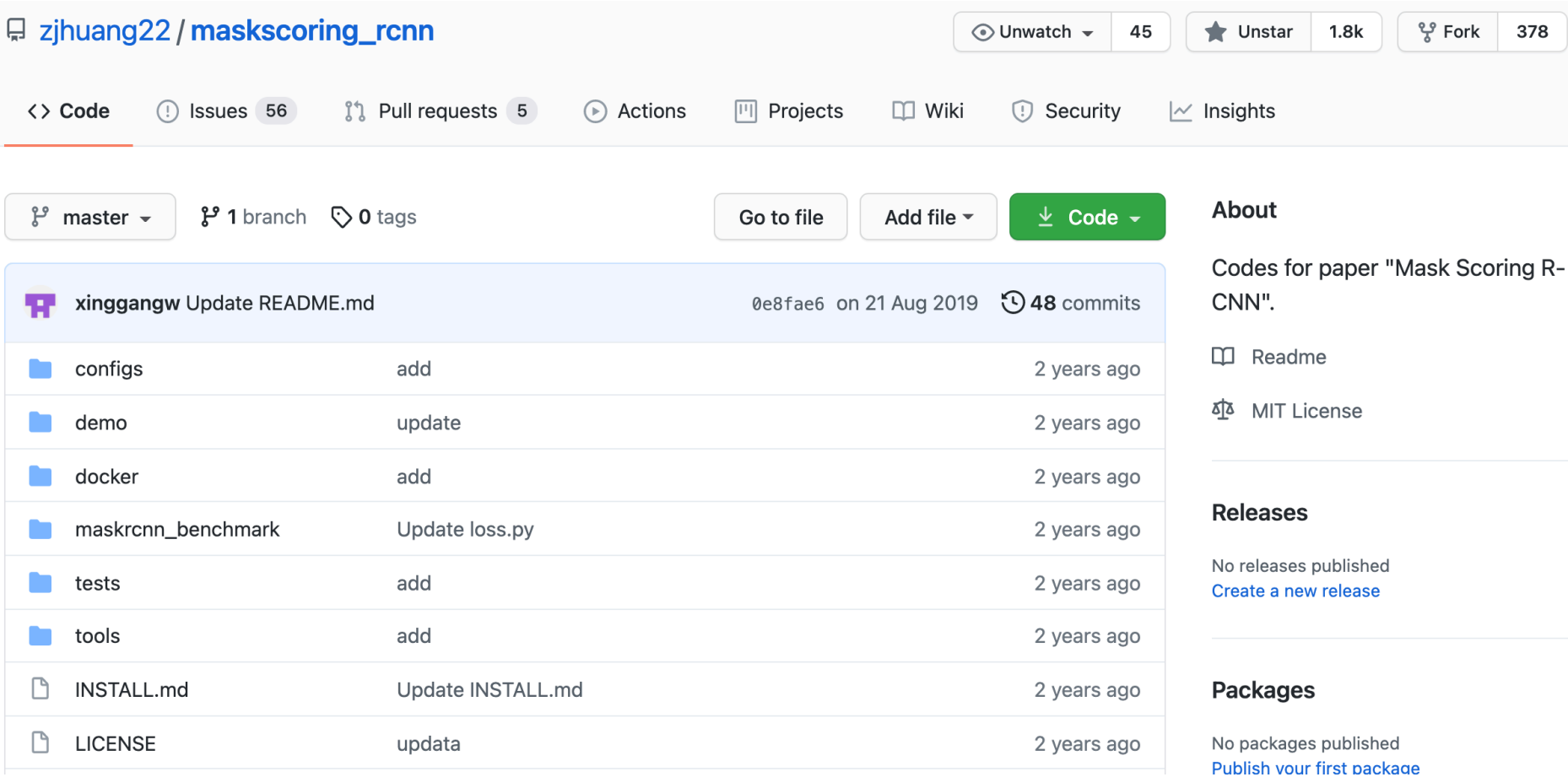
- **Computation cost**: the testing speed (sec./image) of MS R-CNN and Mask R-CNN is almost the same. ResNet-18 FPN: both about 0.132. ResNet-101 DCN+FPN: both about 0.202.

- The upp

Method	Backbone	AP
Mask R-CNN	ResNet-18 FPN	27.7
MS R-CNN		29.3
MS R-CNN*		31.5
Mask R-CNN	ResNet-101 DCN+FPN	37.7
MS R-CNN		39.1
MS R-CNN*		41.7

CN.

Mask Scoring R-CNN



https://github.com/zjhuang22/maskscoring_rcnn

性能超越何恺明Mask R-CNN！华科硕士生开源图像分割新方法

夏乙　　2019-03-05 17:13:43　　来源：量子位

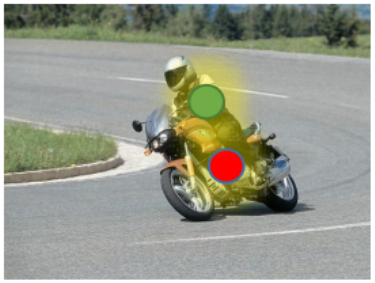
实习生又立功了

安妮 乾明 发自 凹非寺

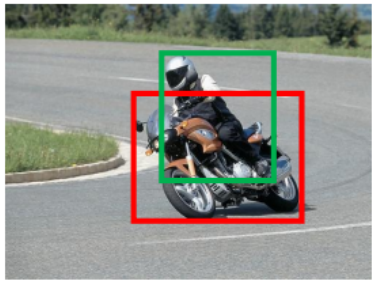
量子位 报道 | 公众号 QbitAI

Annotation time of manual supervision

{motorbike,person}



{motorbike (point), person (point)}



{motorbike (b-box), person (b-box)}

{motorbike (pixel labels), person (pixel labels)}

Annotation time:

1

2.4

10

78

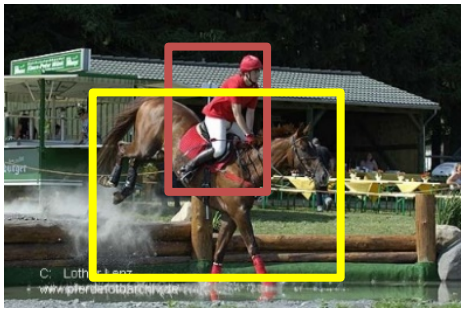
second per instance

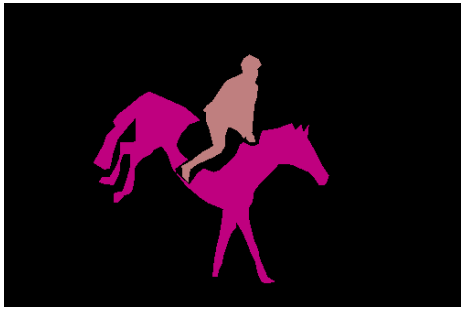
Berman et al., What's the Point: Semantic Segmentation with Point Supervision, ECCV 16

Slide credit: Hakan Bilen

Image labels

Person, Horse



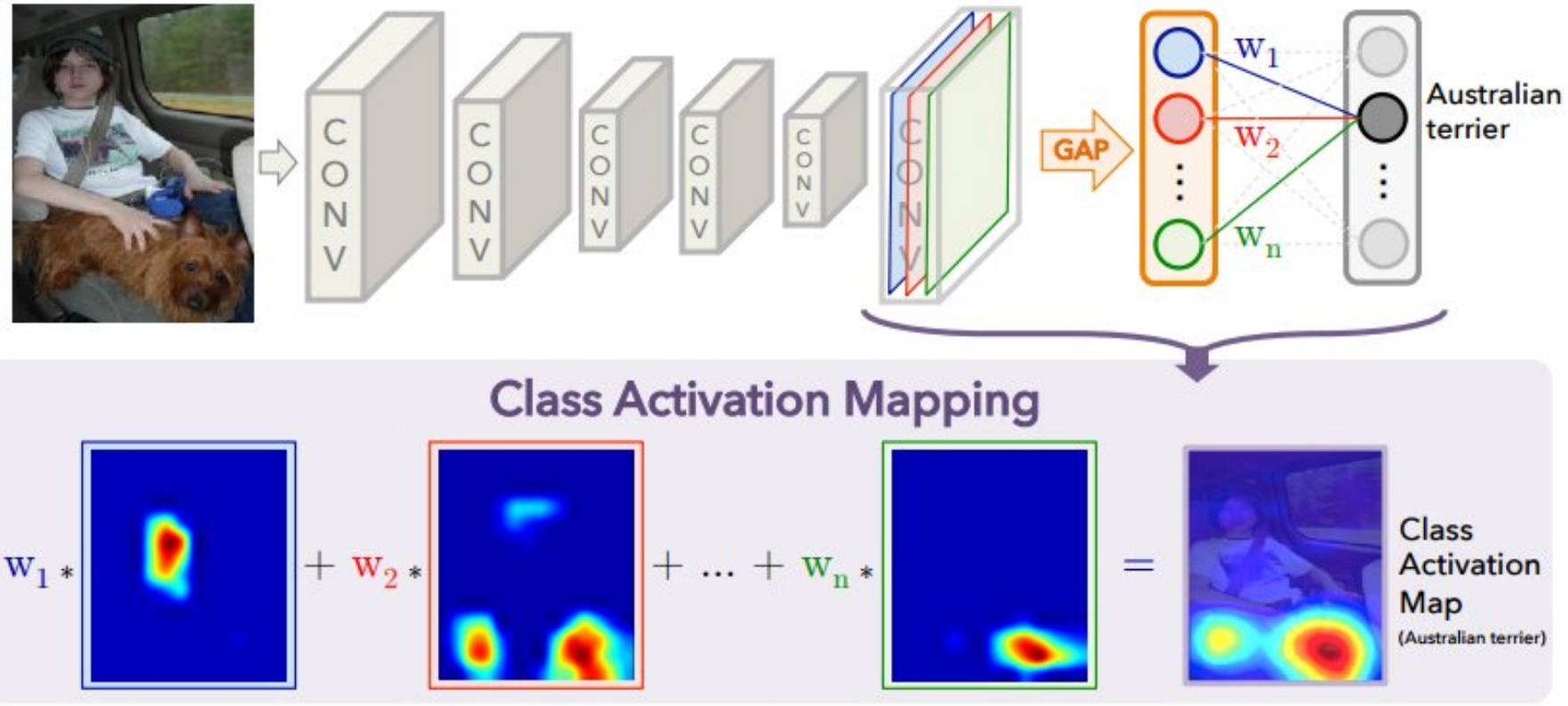


- Supervision: image (category) labels
- Target: Object detection, semantic segmentation etc

[Verbeek CVPR 07, Pendey, ICCV 11, Cinbis CVPR 14, Wang ECCV 14, Papandreou ICCV 15, Belien CVPR 15, Tang CVPR 17, Wei CVPR 17, Singh ICCV 17, Huang CVPR 18 etc.]

Class activation maps

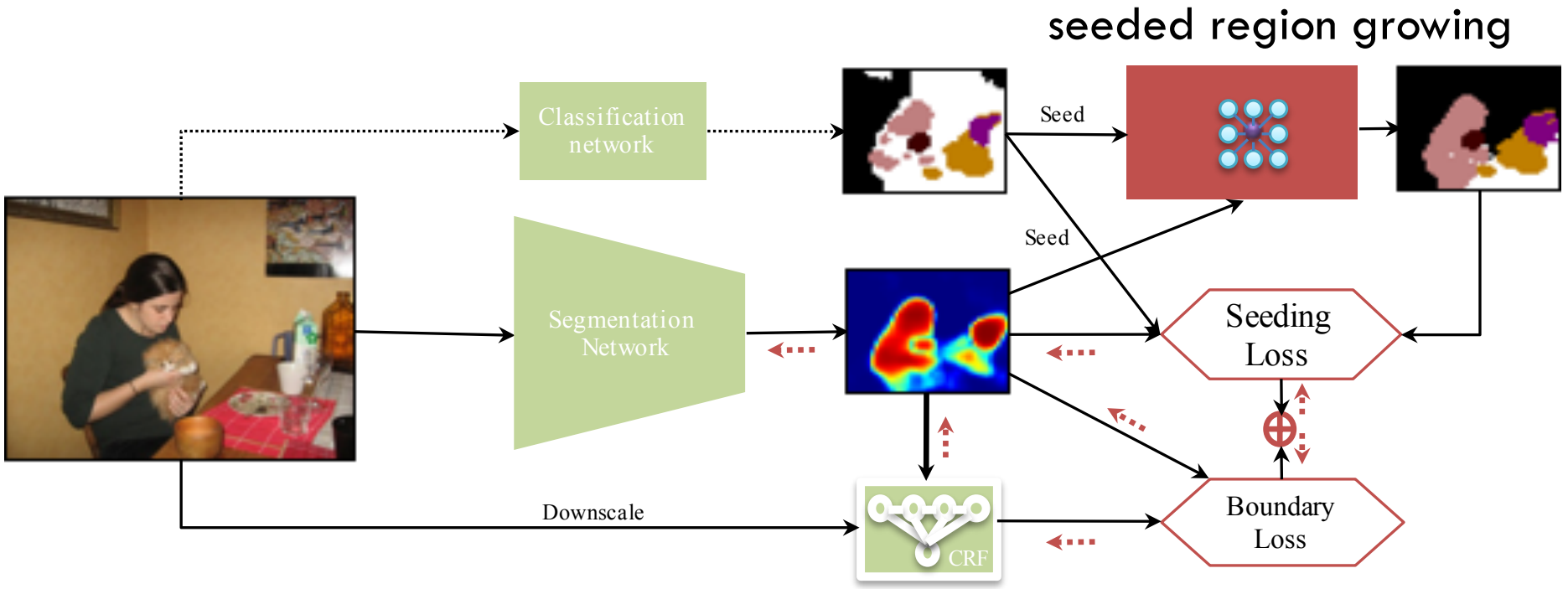
[Zhou CVPR 16]



☺ Finding discriminative regions by Global Average Pooling in a CNN trained using image labels

☺ A very insightful work for understanding CNN

Deep seeded region grown (DSRG) network

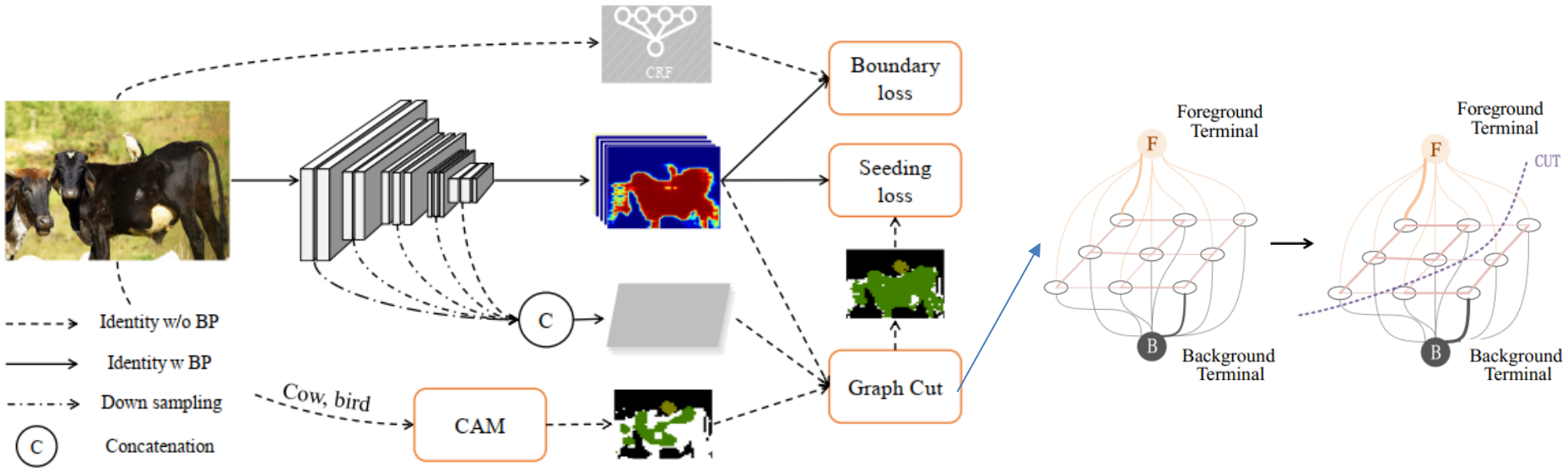


☺ Region growing for complete and dense object regions

☺ A segmentation network generates new pixel labels by itself

Zilong Huang, Xinggang Wang*, Jiasi Wang, Wenyu Liu, Jingdong Wang. **Weakly-Supervised Semantic Segmentation Network with Deep Seeded Region Growing**. **CVPR**, 2018. （人工智能顶级会议，Google scholar Top publications rank 10）

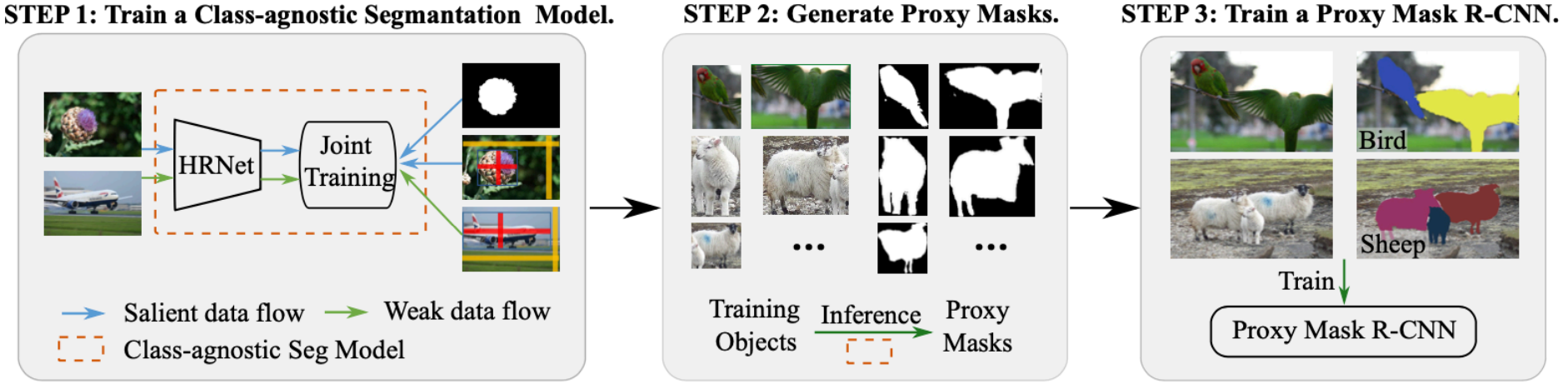
Deep Graph-cut network



☺ Formulating WSSS as a semi-supervised learning problem
☺ Classical graph-cut algorithm for better pseudo labeling

Jiapei Feng, Xinggang Wang*, Wenyu Liu. **Deep Graph Cut Network for Weakly-supervised Semantic Segmentation**. SCIENCE CHINA Information Sciences (**SCIS**)

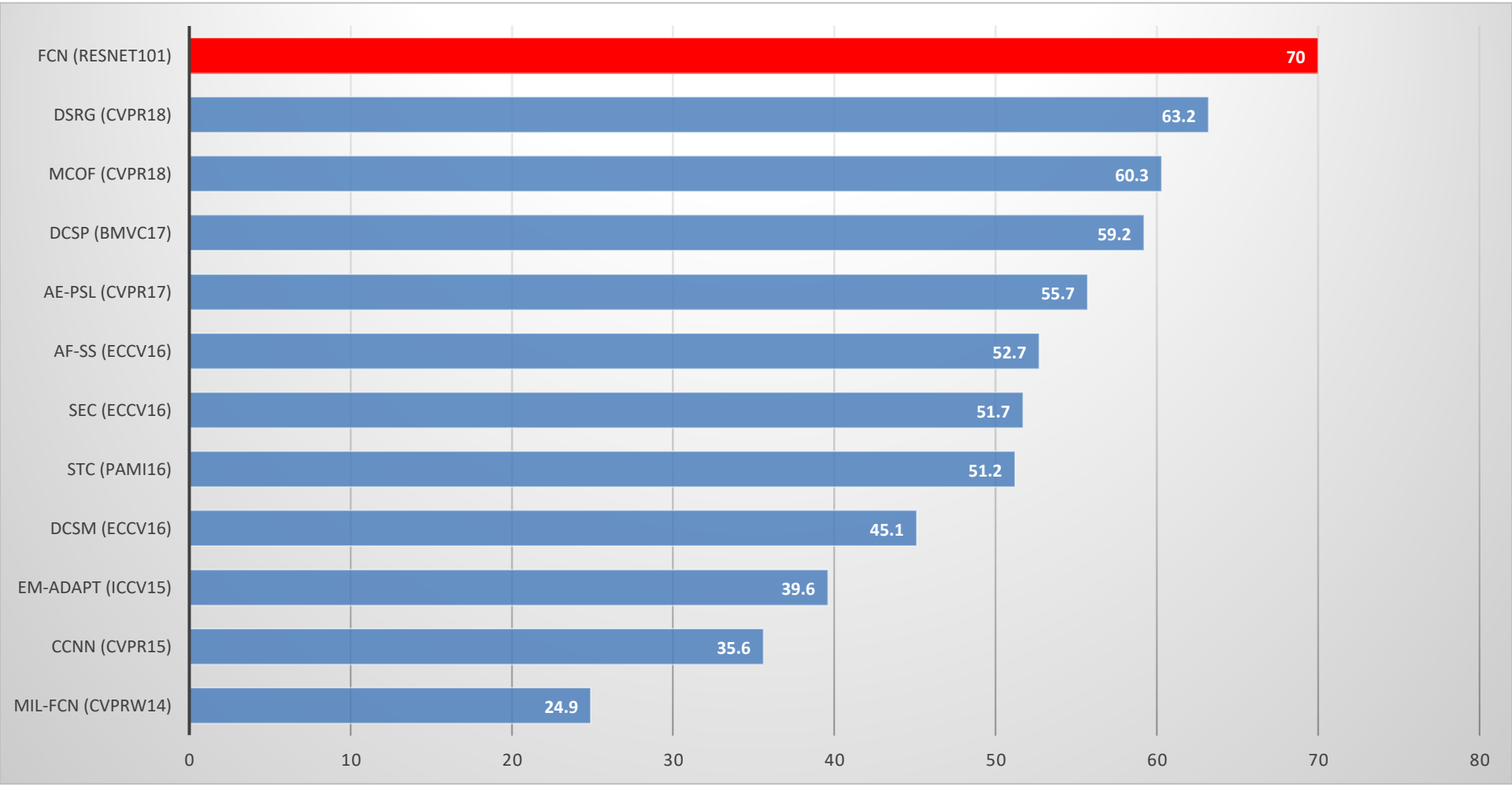
Weak-supervised semantic segmentation with salient priors



Method	Supervision	Backbone	AP_{25}	AP_{50}	AP_{70}	AP_{75}
Mask R-CNN (Oracle) [13]	M	ResNet-101	76.7	67.9	52.5	44.9
PRM [54]	I	ResNet-50	44.3	26.8	-	9.0
IAM [55]	I	ResNet-50	45.9	28.8	-	11.9
Label-PEnet [11]	I	ResNet-50	49.1	30.2	-	12.9
CountSeg [6]	I	ResNet-50	48.5	30.2	-	14.4
WISE [25]	I	ResNet-50	49.2	41.7	-	23.7
IRN [1]	I	ResNet-50	-	46.7	23.5	-
LIID [33]	I	ResNet-50	-	48.4	-	24.9
SDI [20]	B	ResNet-101	-	46.4	-	18.5
BBTP [16]	B	ResNet-101	75.0	58.9	30.4	21.6
AnnoCoIn [2]	B	ResNet-101	73.8	58.2	34.3	32.1
Ours	B+S	ResNet-101	77.7	67.6	49.4	42.4

在pascal voc数据集上，采用显著性先验+弱监督标记实现全监督的性能

Performance



WSSS performance (mIoU on PASCAL VOC 2012 test)

Resources

- CVPR tutorial: **Weakly supervised learning for computer vision**, by Hakan Belen, Rodrigo Benenson, Jasper Uijlings, https://hbilen.github.io/wsl-cvpr18.github.io

- Source codes
  - WSSDN: https://github.com/hbilen/WSDDN
  - CAM: http://cnnlocalization.csail.mit.edu
  - OICR/PCL: https://github.com/ppengtang/oicr/tree/pcl
  - SEC: https://github.com/kolesman/SEC
  - DSRG: https://github.com/speedinghzl/DSRG

Weakly-supervised medical image segmentation 81

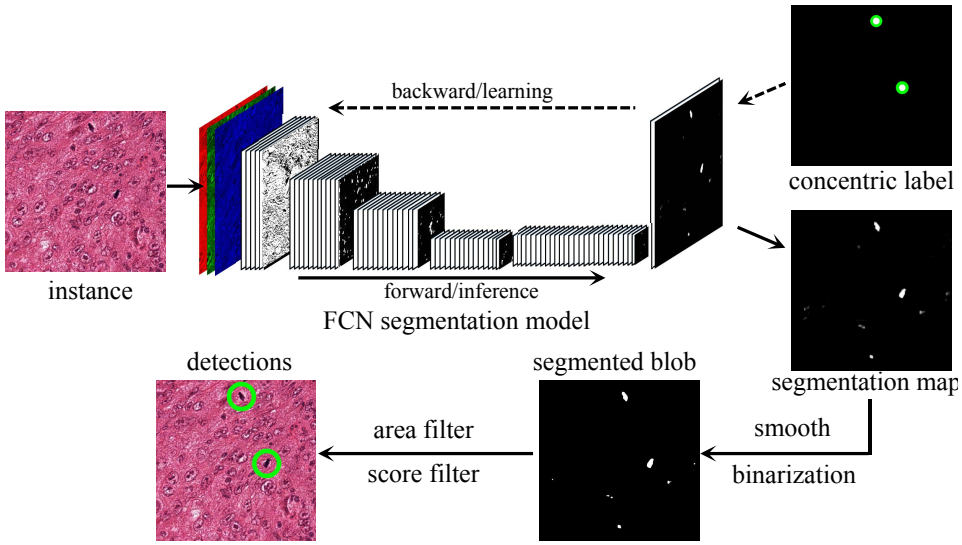


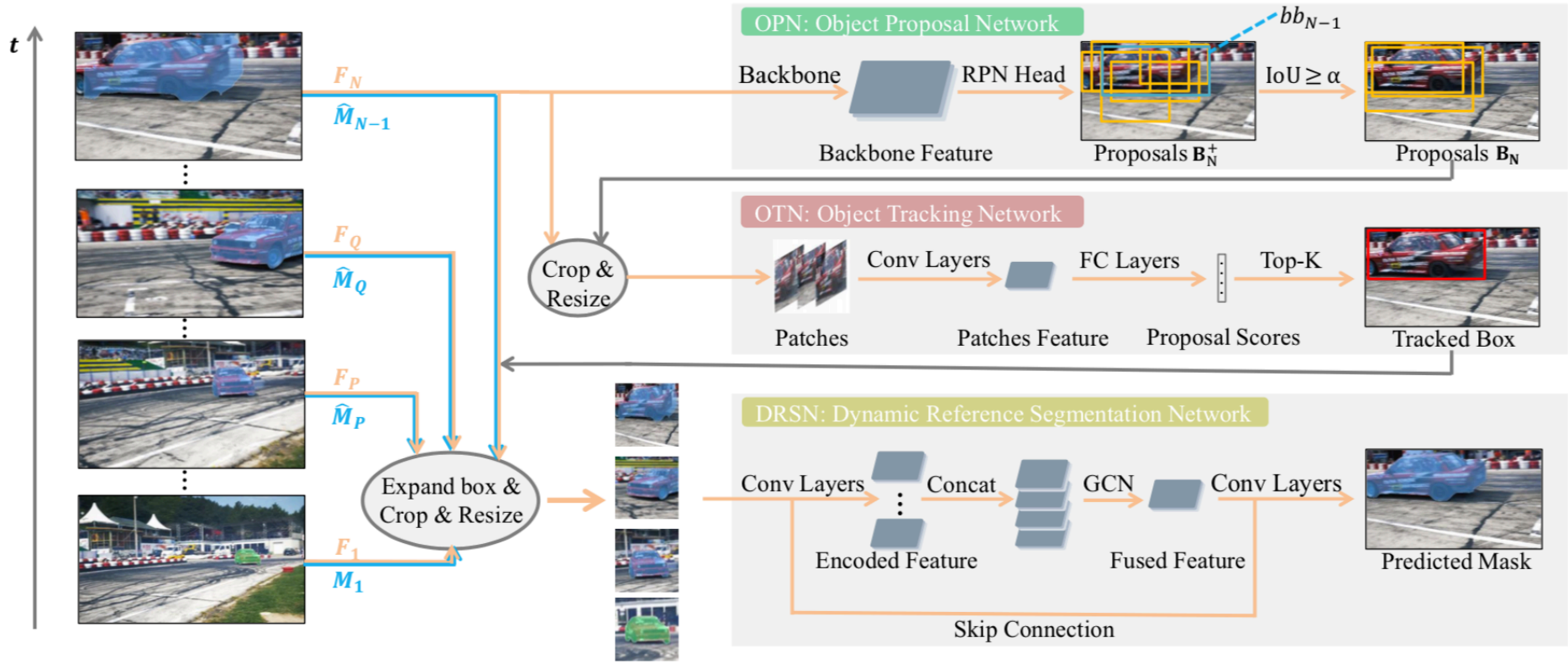
Table 5: Performance of different approaches on 2014 MITOSIS test set. "-" denotes the results which are not released.

Method	Precision(%)	Recall(%)	F-score(%)
STRASBOURG	-	-	2.4
YILDIZ	-	-	16.7
MINES-CURIE-INSERM	-	-	23.5
CUHK	44.8	30.0	35.6
DeepMitosis (Li et al., 2018)	43.1	44.3	43.7
CasNN(single) (Chen et al., 2016b)	41.1	47.8	44.2
CasNN(average) (Chen et al., 2016b)	46.0	50.7	48.2
SegMitos-r12	46.88	**51.72**	49.18
SegMitos-r12R24	62.25	46.31	53.11
SegMitos-r15R30	59.43	51.23	55.03
SegMitos-random	**63.75**	50.25	**56.20**

Chao Li, Xinggang Wang, Wenyu Liu, Longin Latecki, Bo Wang, Junzhou Huang. **Weakly Supervised Mitosis Detection in Breast Histopathology Images using Concentric Loss**. **Medical Image Analysis**. (2019)
（IF=8.88）

Chao Li, Xinggang Wang, Wenyu Liu, Longin Latecki. **DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks**. **Medical Image Analysis**. Volume 45, Pages 121-133, (2018)
（IF=8.88）

One-shot video object segmentation



2nd Place in the 1st Large-scale Video Object Segmentation Challenge Workshop in conjunction with ECCV 2018, Munich, Germany. Sep. 2018.

Qiang Zhou, Zilong Huang, Lichao Huang, Yongchao Gong, Han Shen, Chang Huang, Wenyu Liu, Xinggang Wang. Proposal, Tracking and Segmentation (PTS): A Cascaded Network for Video Object Segmentation. arXiv:1907.01203v2 (2019)

Computation efficient object detection and segmentation

- Deep neural networks for object detection/segmentation are inherently computation expensive
  - The networks contain a huge number of parameters
  - Object detection/segmentation requires high-resolution input image and feature maps
- While, real applications requires high-speed and low-power consumption.
- Efficient deep networks for object detection/segmentation
  - Neural architecture search, an AutoML solution
  - Efficient attention networks
  - Efficient solution for modeling temporal information in video

Image segmentation: summary

- Main concept of segmentation approaches before the era of deep learning
- Early merging and clustering methods
- Active contours
- Variational approaches
- Watershed families
- Classical learning-based approaches
- Trend towards deep learning based approaches
- FCN and DeepLab methods
- Mask R-CNN

Course project

链接：https://pan.baidu.com/s/1hRb8a9-tPJkLZOpPFOyoiQ 提取码：mw3l

Task of this project: TinySeg

- TinySeg dataset contains 6624 images (6000 for training & 624 for validation)
- A small segmentation network (ResNet18 + PSPNet) with input size 128*128 is given.

Requirements:

- Run the code, train for at least 3 epochs and Calculate pixel accuracy and mIoU at the validation set.
- Implementing DeepLabv3 (**Atrous spatial pyramid pooling**) using ResNet-18 in this framework.
- Comparing the results between DeepLabv3 vs PSPNet.
- **Testing some images in your own phone using the trained model and visualize the results in your slides.**

Additional credits:

- The higher testing mIoU/Dice the better