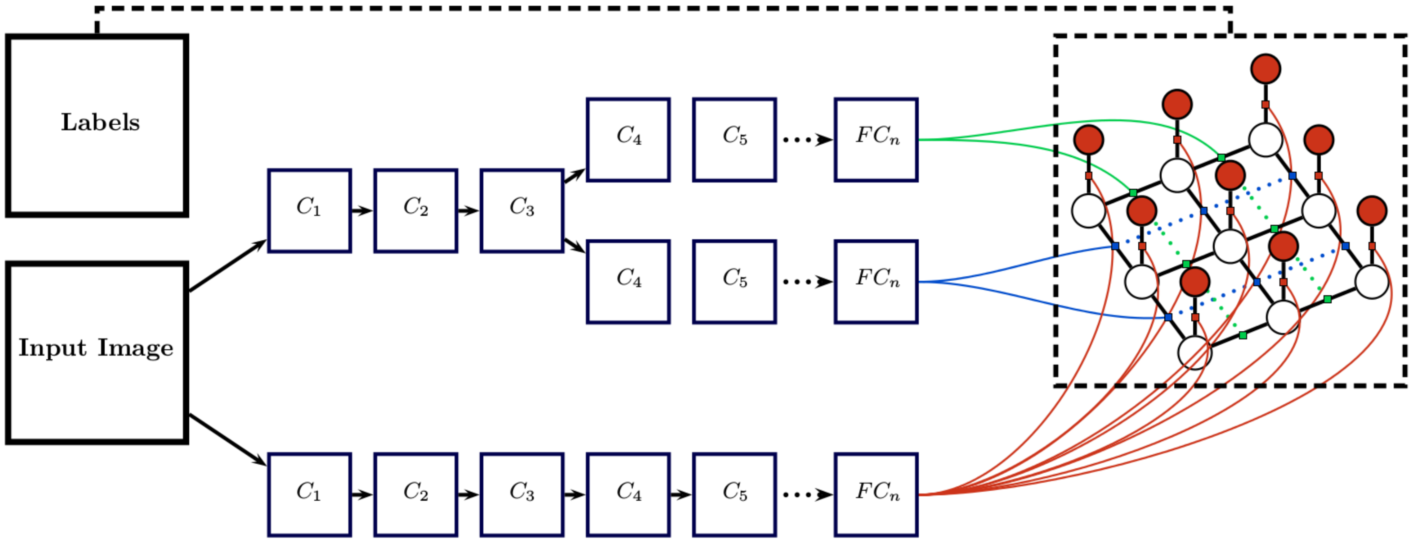
Integrating Structural Information in Deep Convolutional Neural Networks for Low- and High-Level Vision



Iasonas Kokkinos

Center for Visual Computing
CentraleSupelec

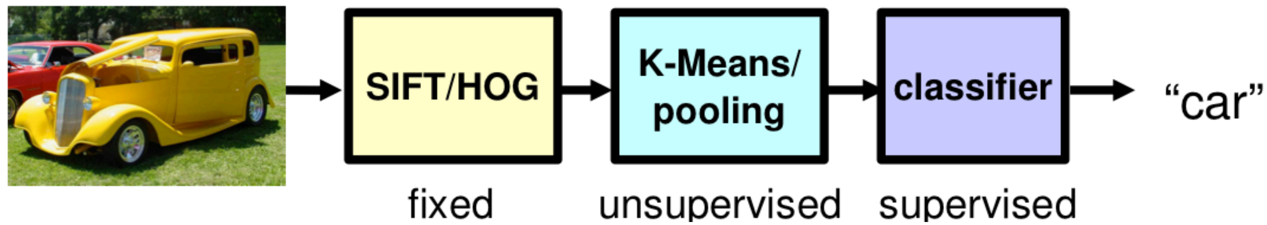
Galen Group
INRIA-Saclay

31 March, 2016
Center for Machine Perception, Prague

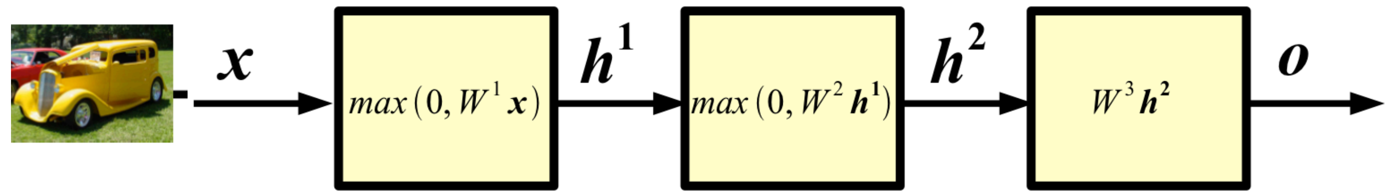
Deep Learning and Computer Vision

1980's pixels → edge → texton → motif → part → object

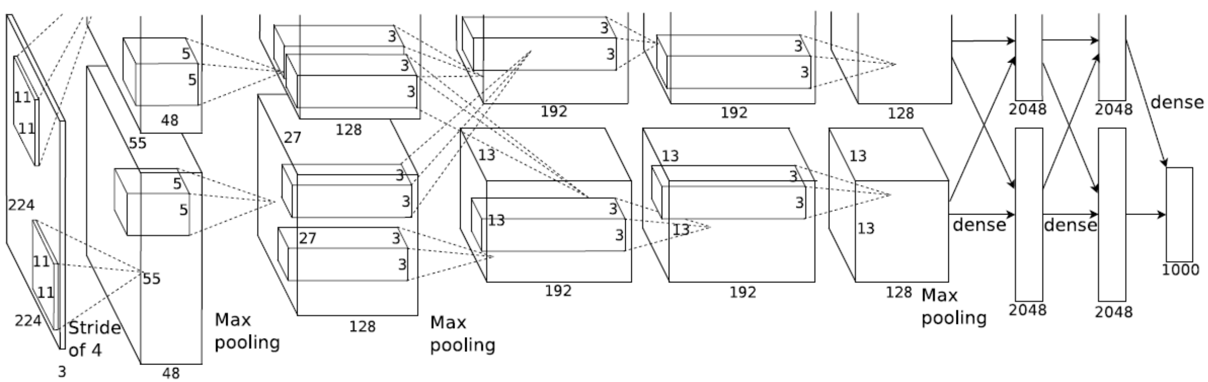
2000-2010



2010+

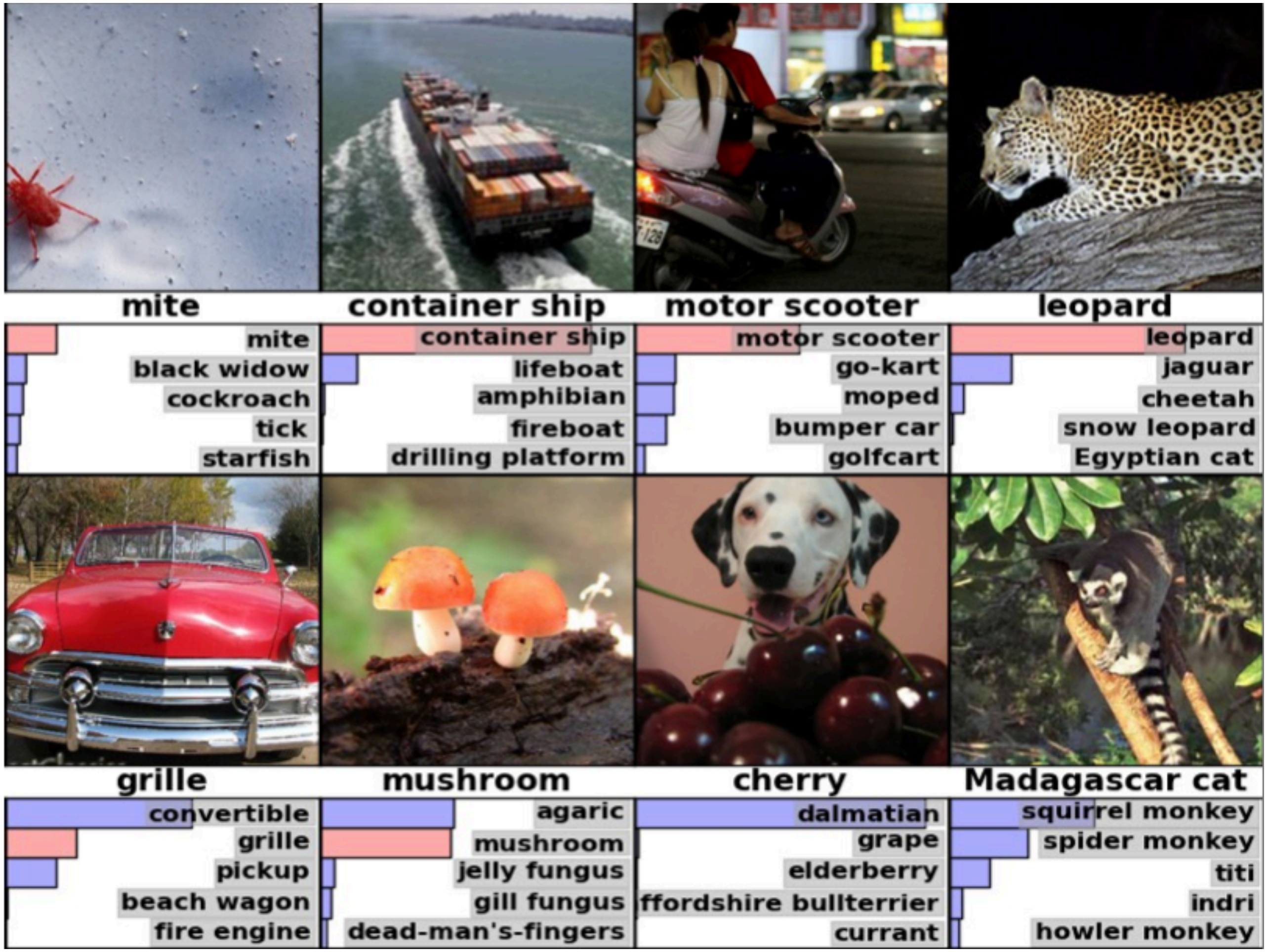


Breakthrough: Imagenet 2012



A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

Imagenet top-5 error rates



Humans:5.4%

A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13 [18%] (best shallow competitor: 36%)

K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, http://arxiv.org/abs/1502.01852, 2015. [4.5%]

S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, http://arxiv.org/abs/1502.03167, 2015. [4.5%]

K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, Arxiv, 2015 [3.6%]

DCNNs and Vision

2012 onwards: all about DCNNs

if [all] you have [is] a hammer, you treat everything like a nail

Today:

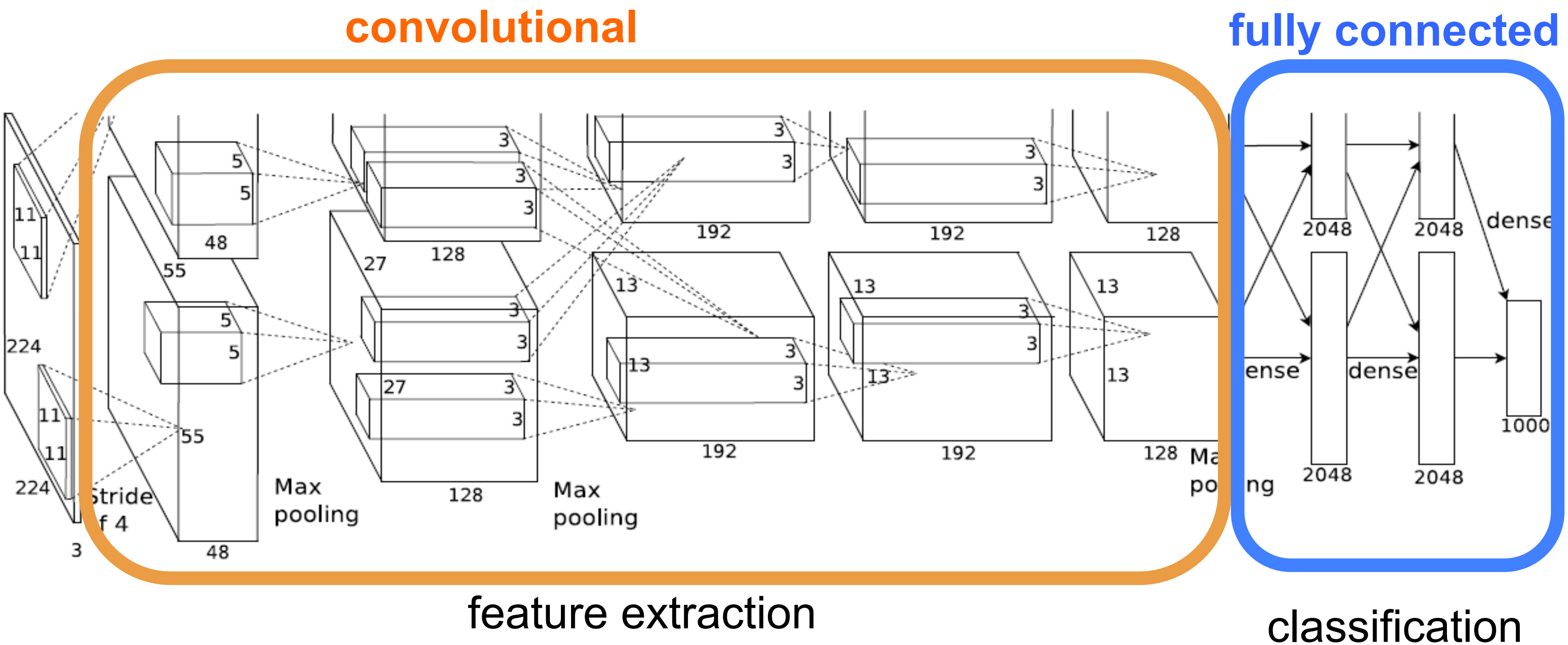
- -Classification & Detection
- -Semantic Segmentation
- -Boundary Detection
- -Feature Descriptors

2014 onwards: structured prediction and DCNNs

trust is good, but control is better!

This talk: controlling DCNNs for low- and high- level tasks

Convolutional/Fully Connected DCNN layers



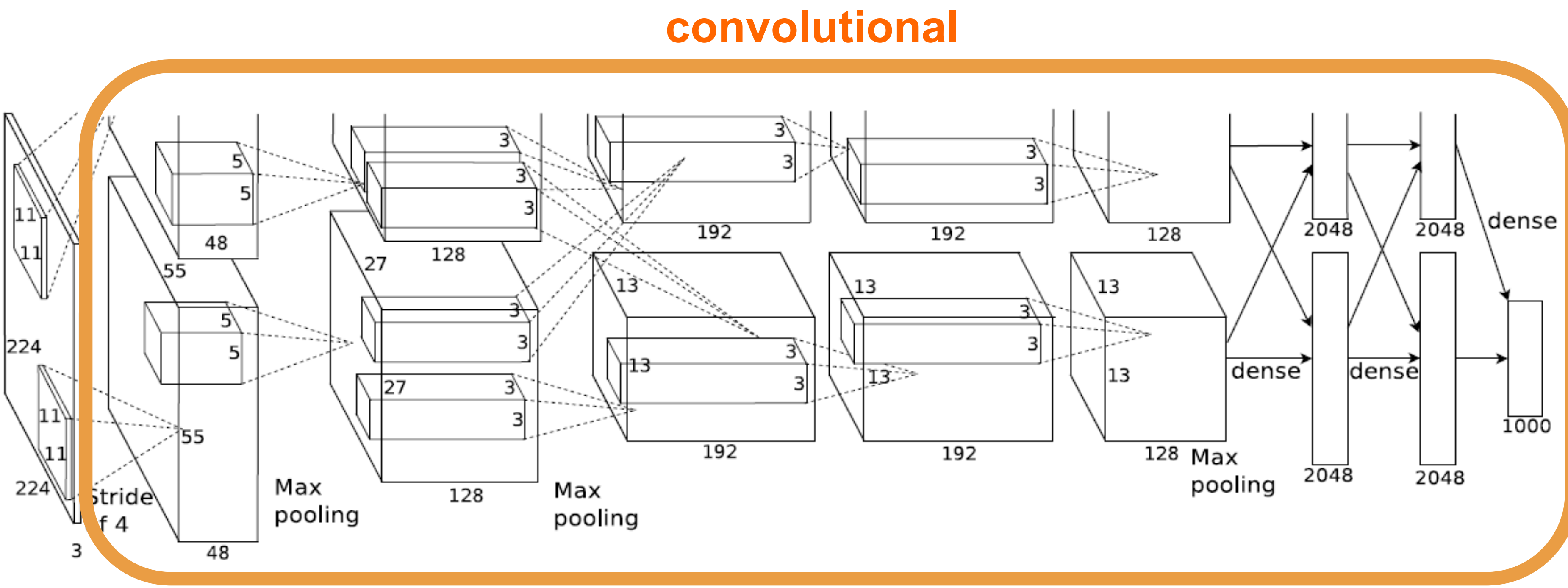
AlexNet

A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. NIPS13

VGG network

K. Simonyan and A. Zisserman. Very deep CNNs for large-scale image recognition, ICLR 2015

Fully convolutional neural networks



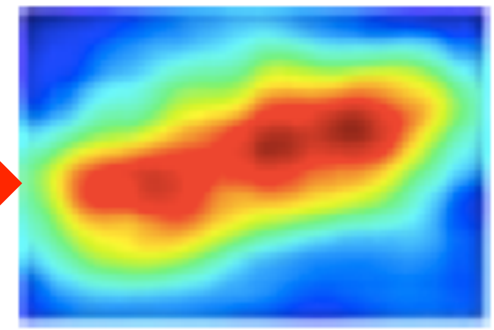
Fully connected layers: 1x1 spatial convolution kernels

“FCNNs” (2015) or "Space Displacement Neural Nets” (1998)

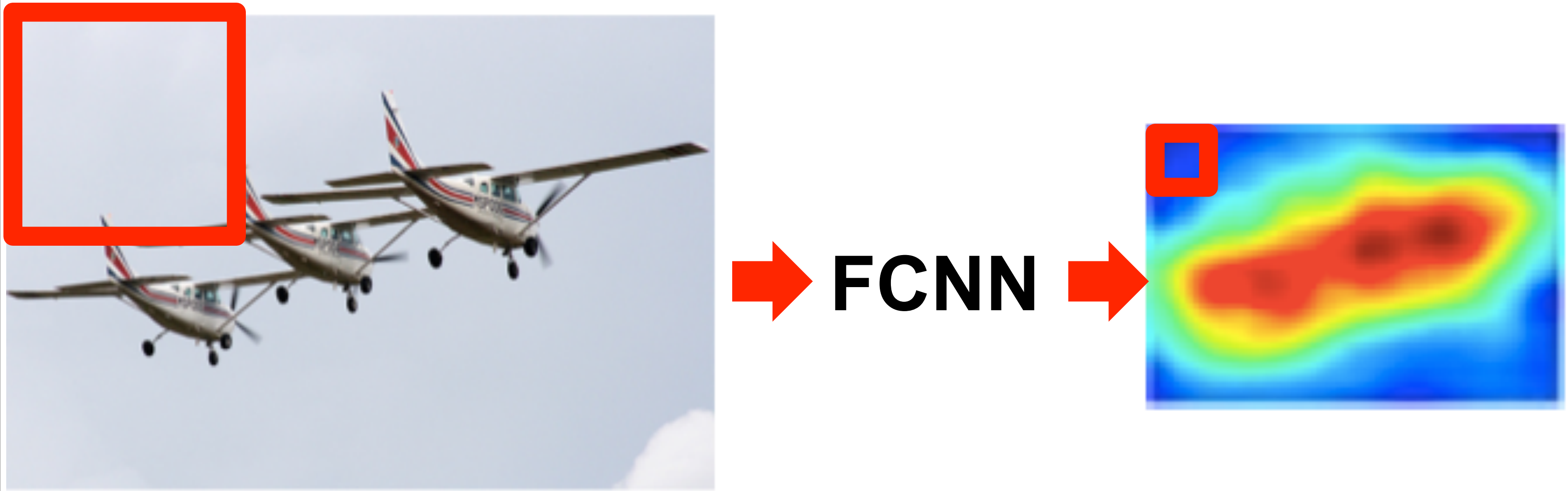
Y. LeCun, et al, Gradient-Based Learning Applied to Document Recognition, Proc. IEEE 1998
J. Long, et al., Fully Convolutional Networks for Semantic Segmentation, CVPR 2015

Fully convolutional neural networks

FCNN



Fully convolutional neural networks



Fully convolutional neural networks

Fully convolutional neural networks

Fully convolutional neural networks

Fast (shared convolutions)
Simple (dense)

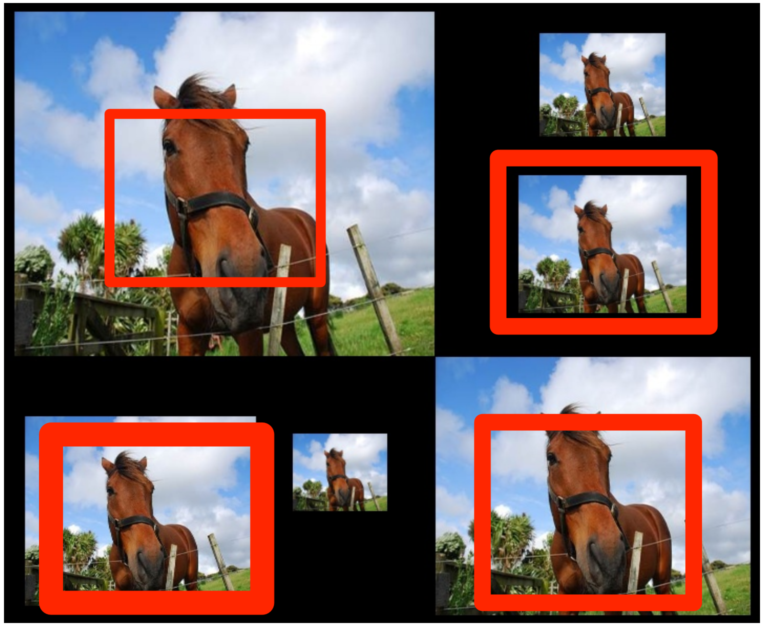
This talk: controlling DCNNs for low- and high- level tasks

-Classification & Detection

-Semantic Segmentation

-Boundary Detection

-Feature Descriptors



G. Papandreou

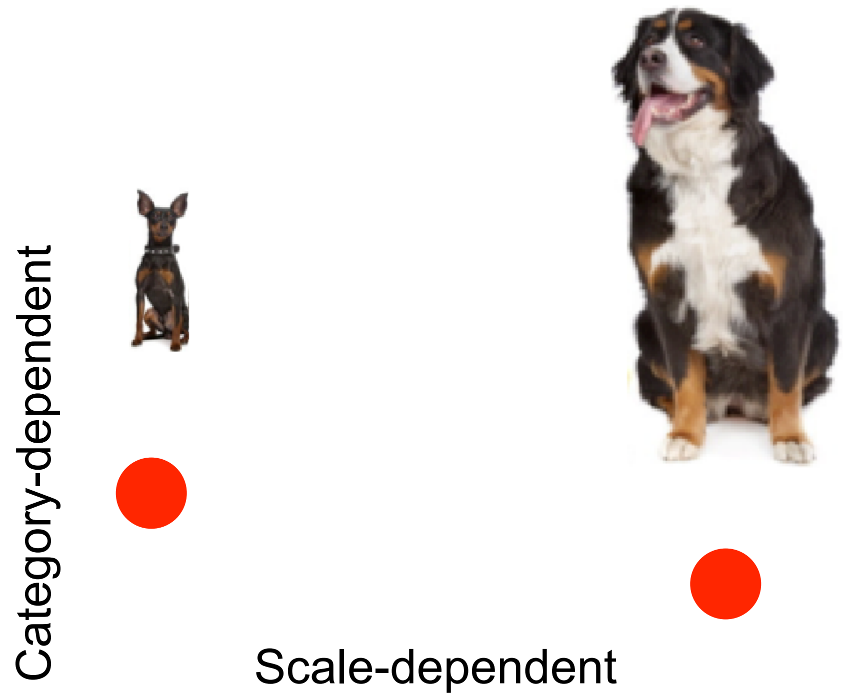
P.-A. Savalle

S. Tsogkas

G. Papandreou, P. A. Savalle, I. Kokkinos Modeling Local and Global Deformations in Deep Learning: Epitomic Convolution, MIL, and Sliding Window Detection, CVPR 2015

P.-A. Savalle, S. Tsogkas, G. Papandreou, I. Kokkinos. Deformable Part Models with CNN features (ECCVW 2014)

Scale-Invariant classification



$$\boxed{x} \rightarrow \boxed{\{x_{s_1}, \ldots, x_{s_K}\}}$$

MIL: 'bag' of features

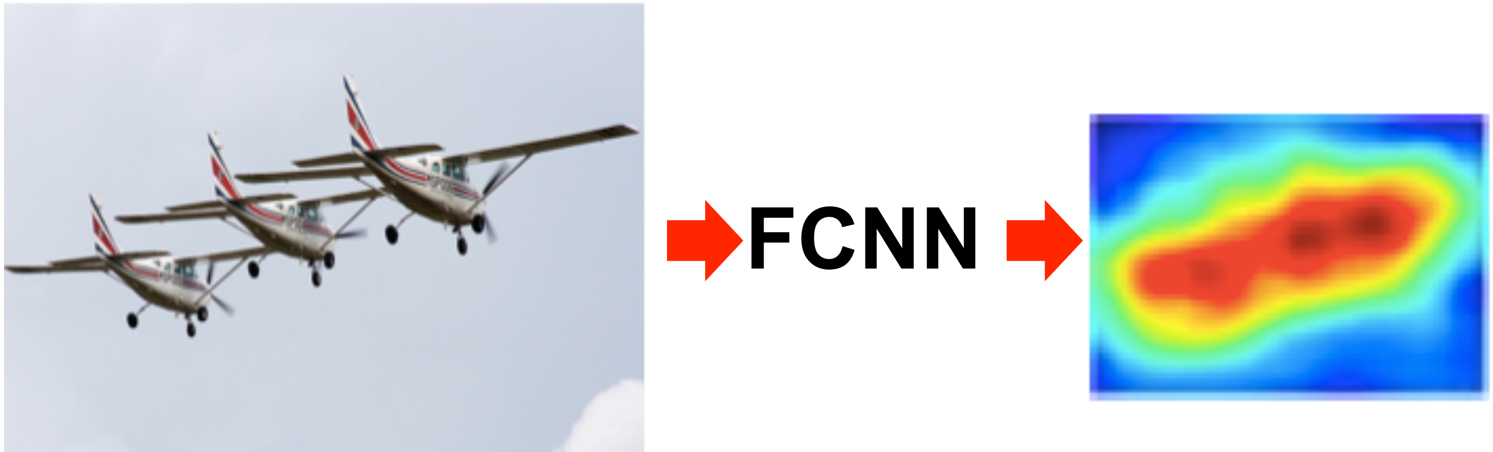
$$F(x) \rightarrow \{F(x_{s_1}), \ldots, F(x_{s_K})\}$$

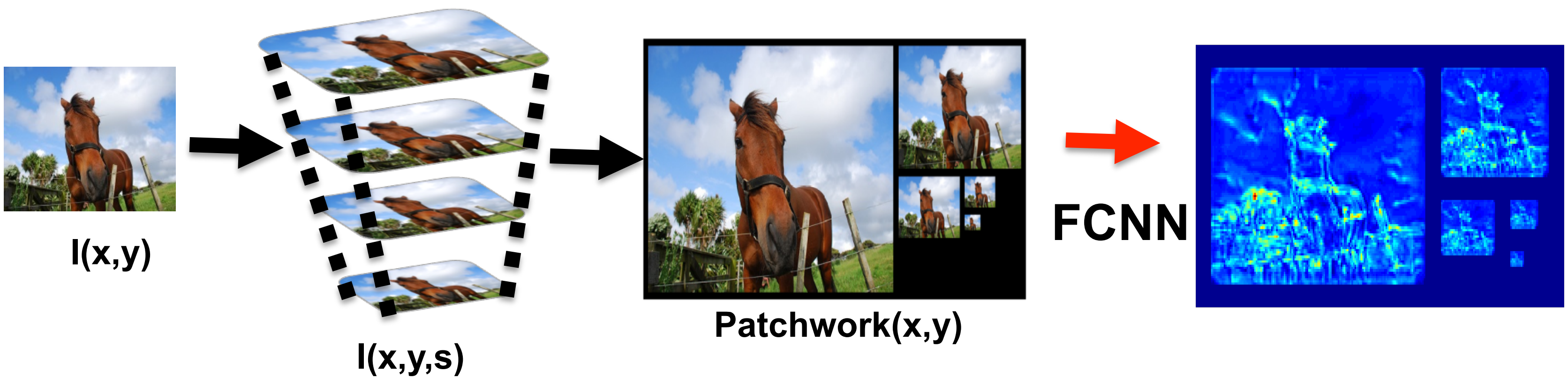
$$F'(x) = \frac{1}{K}\sum_{k=1}^{K} F(x_{s_k})$$

This work: $F'(x) = \max_k F(x_{s_k})$

A. Howard. Some improvements on deep convolutional neural network based image classification, 2013.
K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.
T. Dietterich et al. Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence, 1997

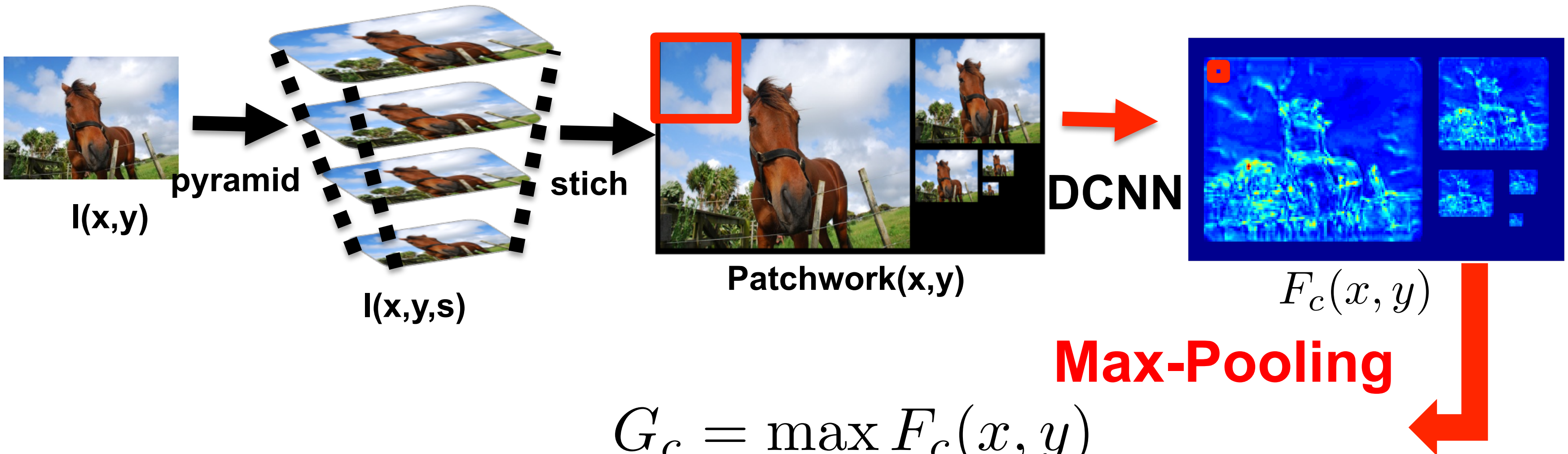
Position and Scale evaluation in `batch mode'





Dubout, C., Fleuret, F.: Exact acceleration of linear object detectors. ECCV 2012
Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K.: Densenet. arXiv 2014

Explicit Scale/Position Search + MIL Training



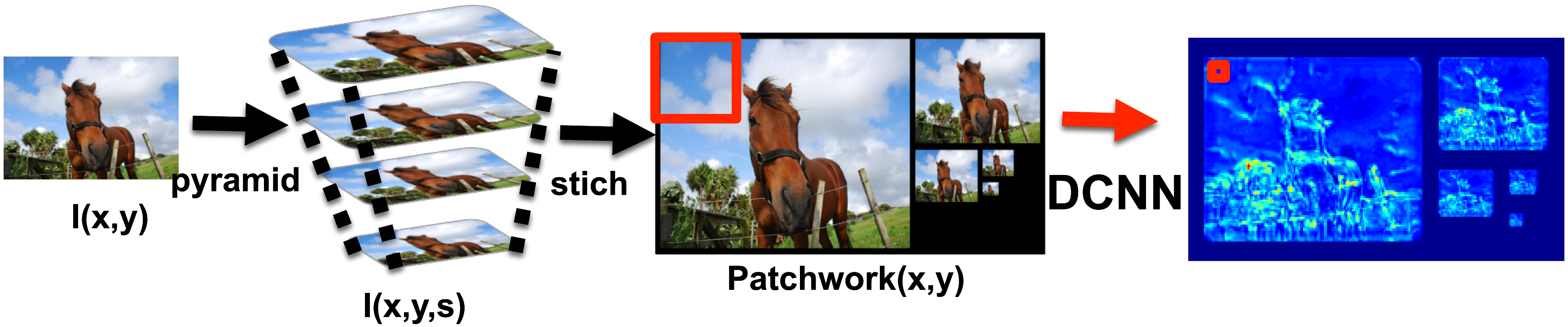
$$G_c = \max_{(x,y)} F_c(x,y)$$

MIL: Explicit position & scale search during both training and testing

(0) Baseline: max-pooled net		(1) epitomic DCNN		(2) epitomic DCNN + search
13.0%	→ ~1% gain	11.9%	→ ~2% gain	10.0%

Bonus: Vanilla argmax yields 48% localization error in Imagenet

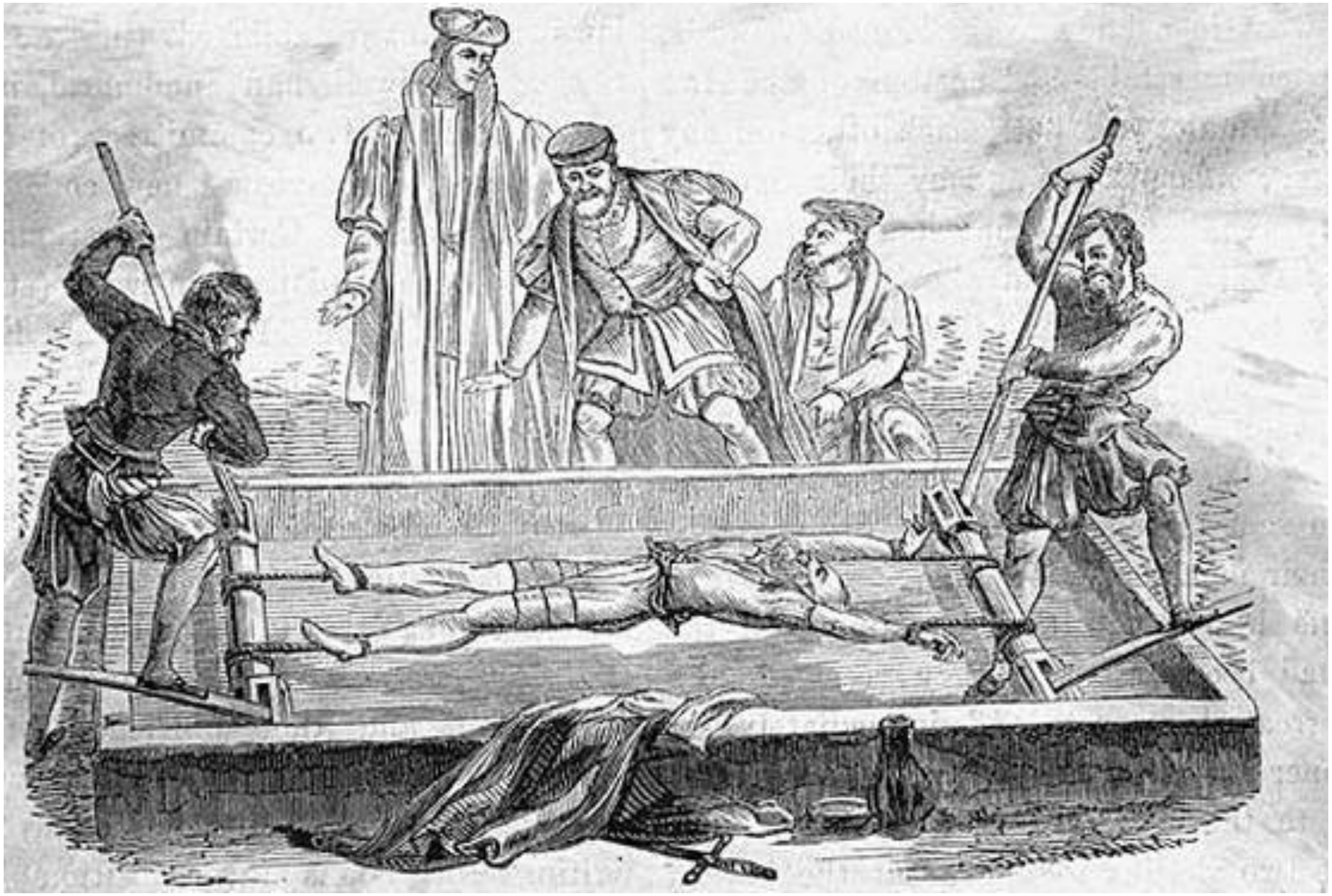
Towards Object Detection



Search over position and scale: done!

Missing: aspect ratio

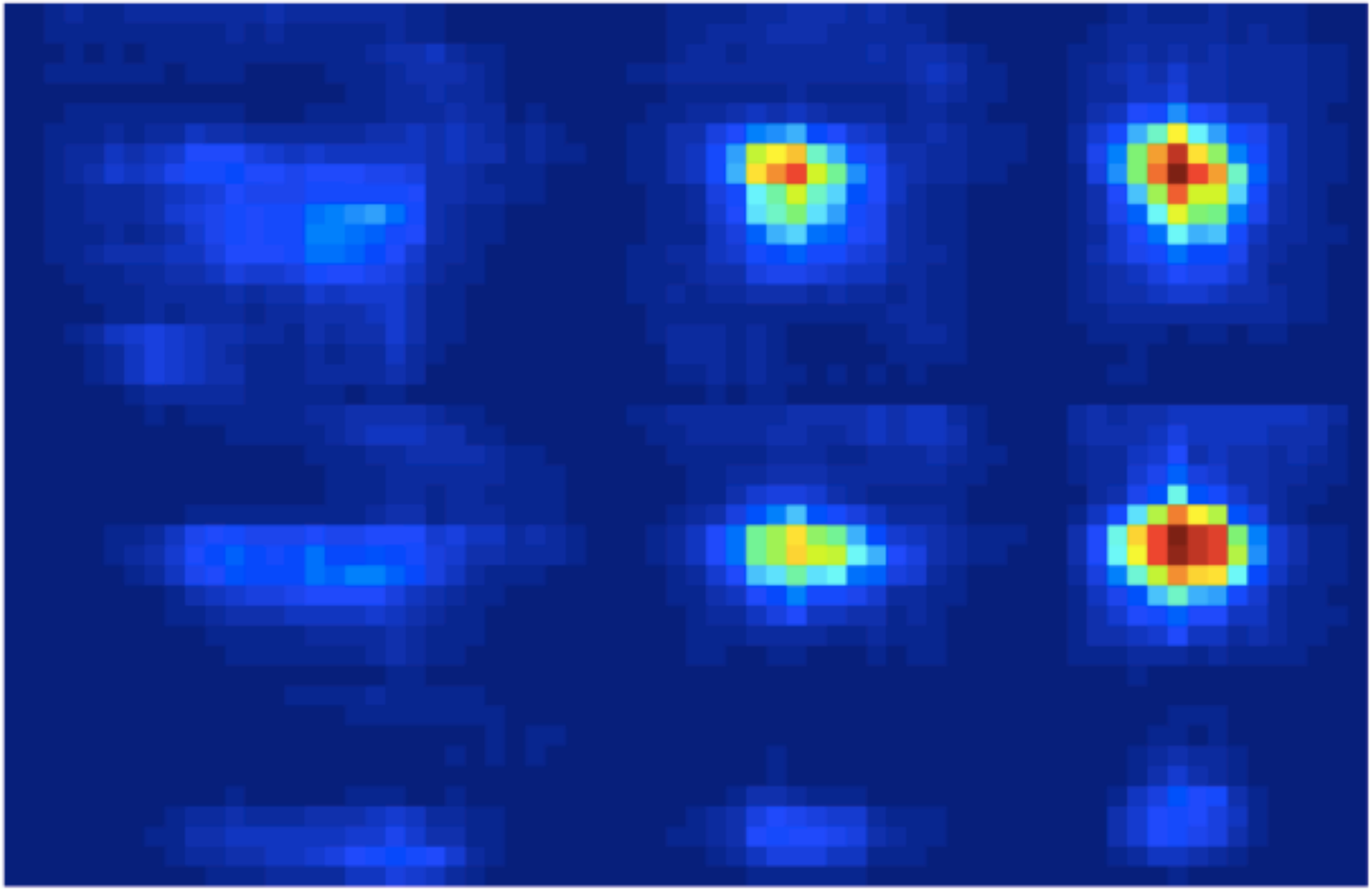
Procrustes Alignment: The Greeks did it first!



F.L. Bookstein, *Morphometric tools for landmark data*, Cambridge University Press, (1991).
T.F. Cootes and C.J. Taylor and D.H. Cooper and J. Graham (1995). "Active shape models - their training and application". *Computer Vision and Image Understanding* (61): 38–59
M.-M. Cheng, Z. Zhang, W.-Y. Lin, P. Torr, BING. CVPR, 2014.
R. Girschick, Donahue, Darrell, Malik, RCNN, CVPR 2014

Explicit search over aspect ratio, scale & position

Explicit search over aspect ratio, scale & position



See also: Region Proposal Networks (RPN) Faster-RCNN, 2016

Pascal VOC 2007: Best sliding-window detector

sliding windows

CNN-DPM [1]	MP-DPM [2]	EE-DPM [3]	Ours
43.4%	46.5%	46.9%	58.6%

~10 sec / image

region proposals

RCNN [4]

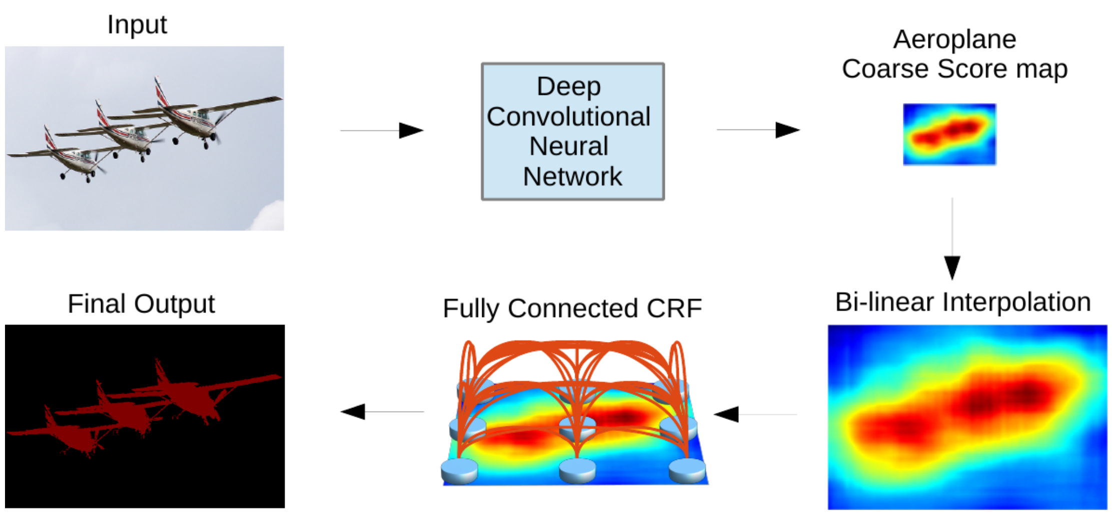
~50 sec / image

62.2%

[1] CNN-DPM: PA Savalle, S. Tsogkas, G. Papandreou, I. Kokkinos. DPM with CNN features (ECCVW 2014)
[2] MP-DPM: R. Girshick, F. Iandola, T. Darrell, and J. Malik. DPMs are CNNs (CVPR 15)
[3] EE-DPM: L. Wan, D. Eigen, R. Fergus. End-to-end integration of CNN, DPM, NMX (CVPR 15)
[4] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation (CVPR 2014)

This talk: controlling DCNNs for low- and high- level tasks

- -Classification & Detection
- -Semantic Segmentation
- -Boundary Detection
- -Feature Descriptors



L-C. Chen

G. Papandreou

A. Yuille

K. Murphy

S. Chandra

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille, Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, ICLR 2015

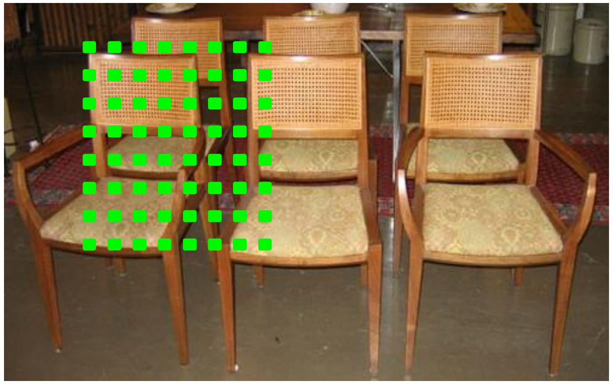
S. Chandra, I. Kokkinos, Fast, Exact and Multi-Scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs, arXiv:1603.08358

Semantic segmentation task

Repurposing DCNNs for semantic segmentation

- Accelerate CNN evaluation by ‘hard dropout’ & finetuning
    - In VGG: Subsample first FC layer 7x7 → 3x3

- Decrease score map stride (32->8) with ‘atrous’ (w. holes) algorithm

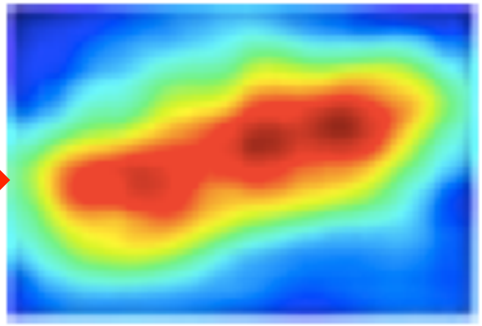


→ 8 FPS

M. Holschneider, et al, A real-time algorithm for signal analysis with the help of the wavelet transform, *Wavelets, Time-Frequency Methods and Phase Space,* 1989.

FCNN for semantic segmentation: results

FCNN



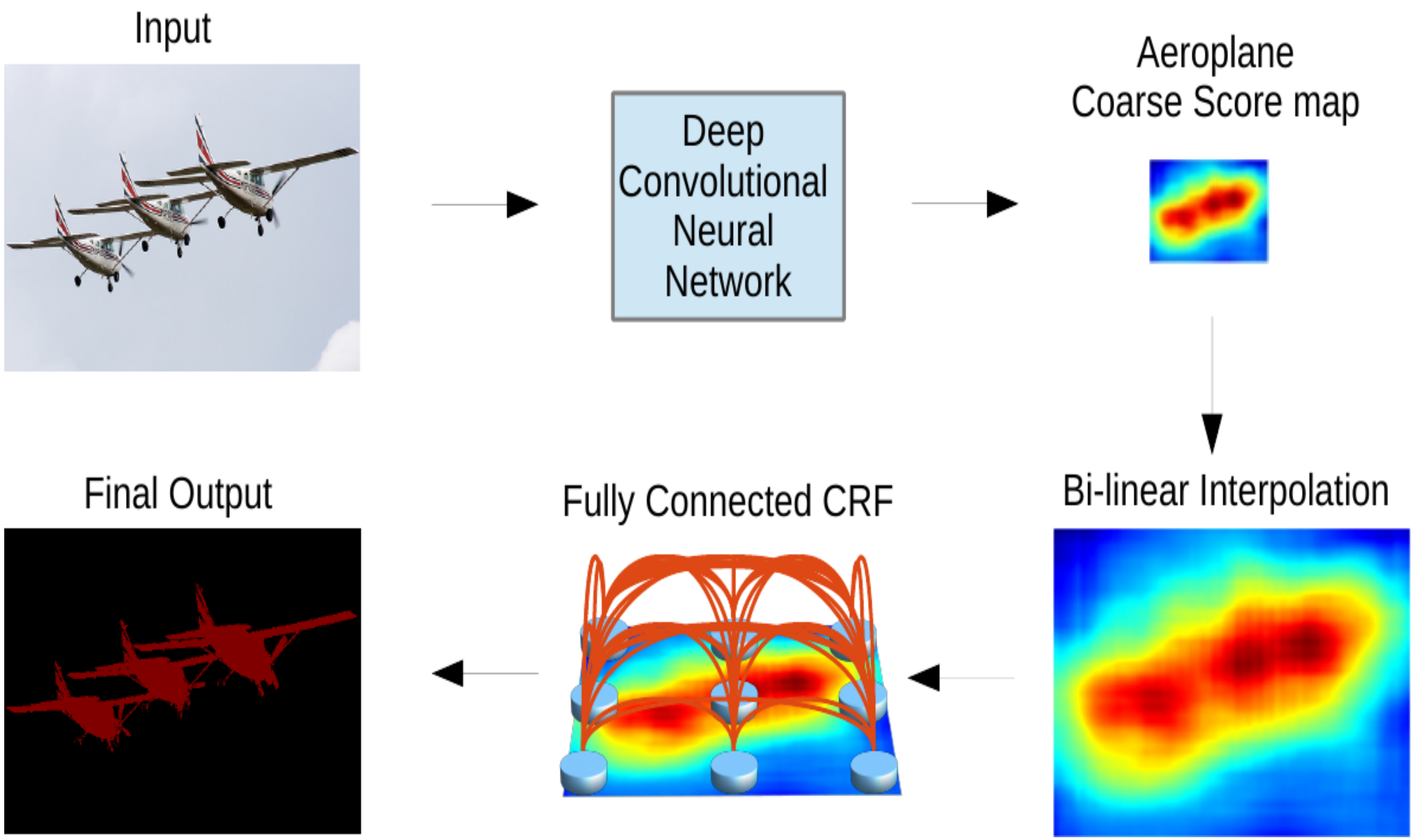
OK classification-wise, rather poor segmentation-wise

- Large CNN receptive field:
  + good accuracy
  - worse performance near boundaries

J. Long, E. Shelhamer, T. Darrell, FCNNs for Semantic Segmentation, CVPR 15

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille, Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, ICLR 2015

FCNN-DenseCRF: Accurate & Sharp

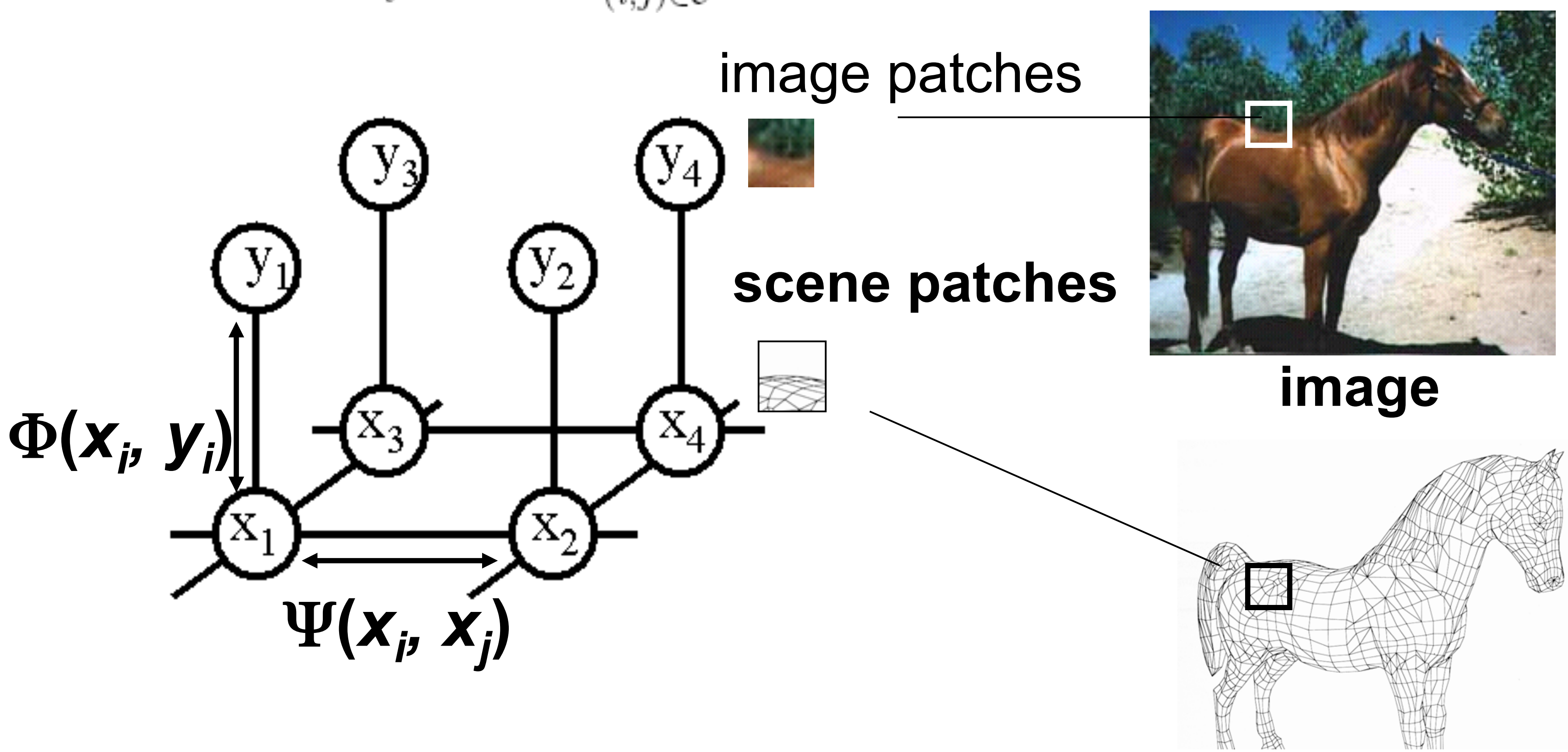


P. Krähenbühl and V. Koltun, Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, NIPS 2011

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille, Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, ICLR 2015

Markov Random Fields in Vision

$$P(X,Y) \;=\; \frac{1}{Z}\prod_i \Phi(Y_i,X_i)\prod_{(i,j)\in\mathcal{C}}\Psi(X_i,X_j)$$



$$P(X|Y) \;=\; ?$$

Mean Field Inference for the Ising Model

Variational Inference: $\mathbf{q}^* \quad = \quad \text{argmin}_{\mathbf{q}\in\mathcal{Q}} KL(\mathbf{q}||\mathbf{p})$

where: $KL(\mathbf{q}||\mathbf{p}) \quad = \quad \sum_{\mathbf{x}} \mathbf{q}(\mathbf{x}) \log \frac{\mathbf{q}(\mathbf{x})}{\mathbf{p}(\mathbf{x})}$, and $\mathcal{Q}$ simplifies minimization

Naïve mean field: $\mathcal{Q} : \{\mathbf{q} : \; \mathbf{q}(\mathbf{x}) = \prod_n \mathbf{q}_n(x_n)\}$

Ising model:

$$\mathbf{p}(\mathbf{x}) = \frac{1}{Z} \exp\left(-E(\mathbf{x})\right)$$

$$E(\mathbf{x}) = \sum_n \sum_{m\in\mathcal{N}_n} J_{m,n}|\mathbf{x}_m - \mathbf{x}_n| \qquad \mathbf{x}_n \in \{-1, 1\}$$

Mean Field equations: $\mathbf{q}_n(1) = \tanh\left(\sum_m J_{n,m}\mathbf{q}_m(1)\right)$

Dense CRF: smart choice of pairwise term

$$\psi_{i,j}(l,l') = \mu(l,l') \sum_{m=1}^{M} w_m k_m(\mathbf{f}_i, \mathbf{f}_j)$$

$$= [l \neq l'] \left[w_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_a^2} - \frac{\|I_i - I_j\|^2}{2\sigma_b^2} \right) + w_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right) \right]$$

Potts model ‘Bilateral kernel’ Spatial proximity

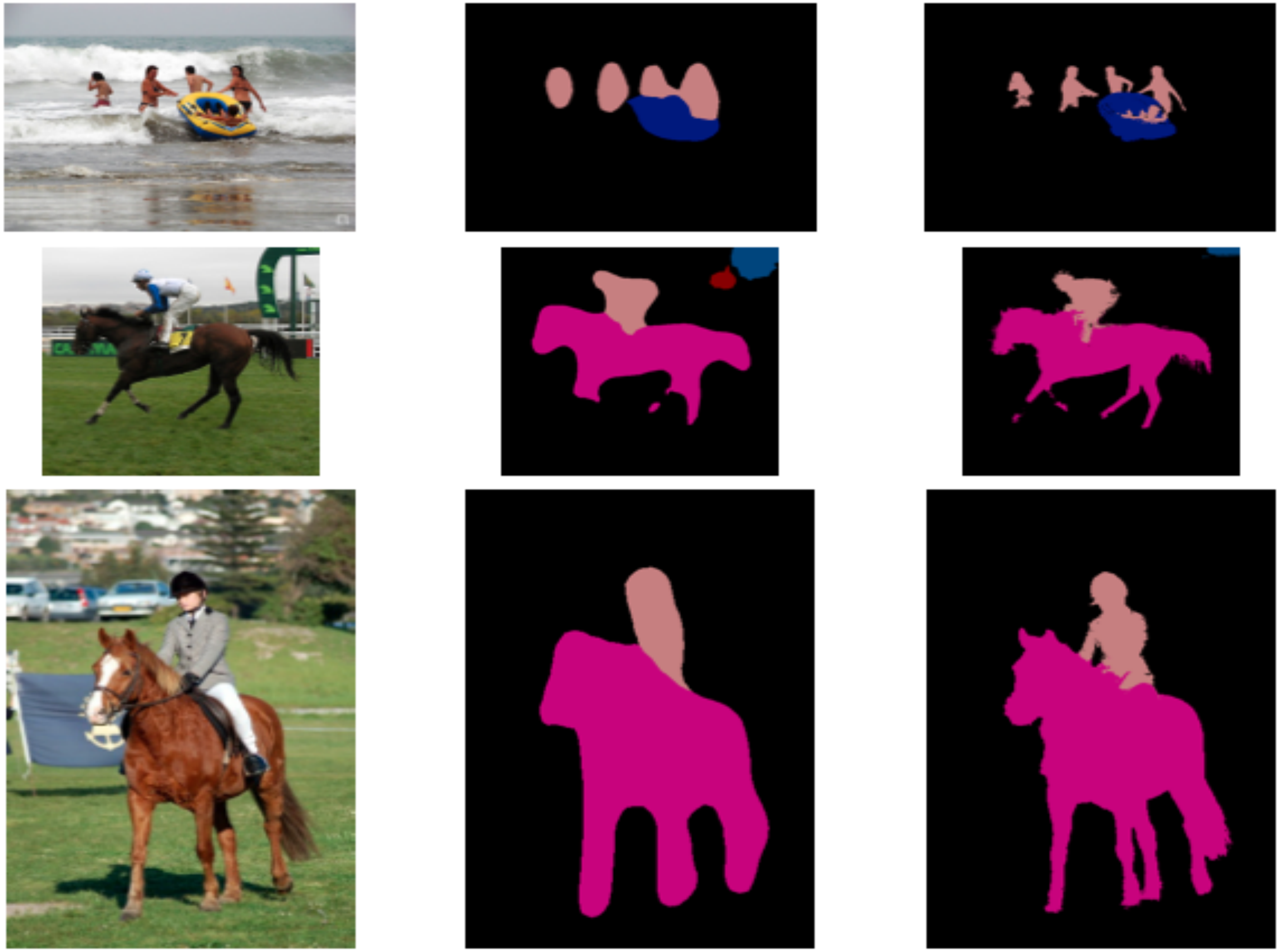
Mean Field Updates:

$$Q_i(l) = \frac{1}{Z_i} \exp\left\{ -\psi_i(l) - \sum_{l'} \mu(l,l') \sum_{m=1}^{M} w_m \sum_{j \in \mathcal{N}_i} k_m(\mathbf{f}_i, \mathbf{f}_j) Q_j(l') \right\}$$

Efficient high-dimensional convolutions using the Permutohedral Lattice

Philipp Krähenbühl and Vladlen Koltun, Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, NIPS 2011

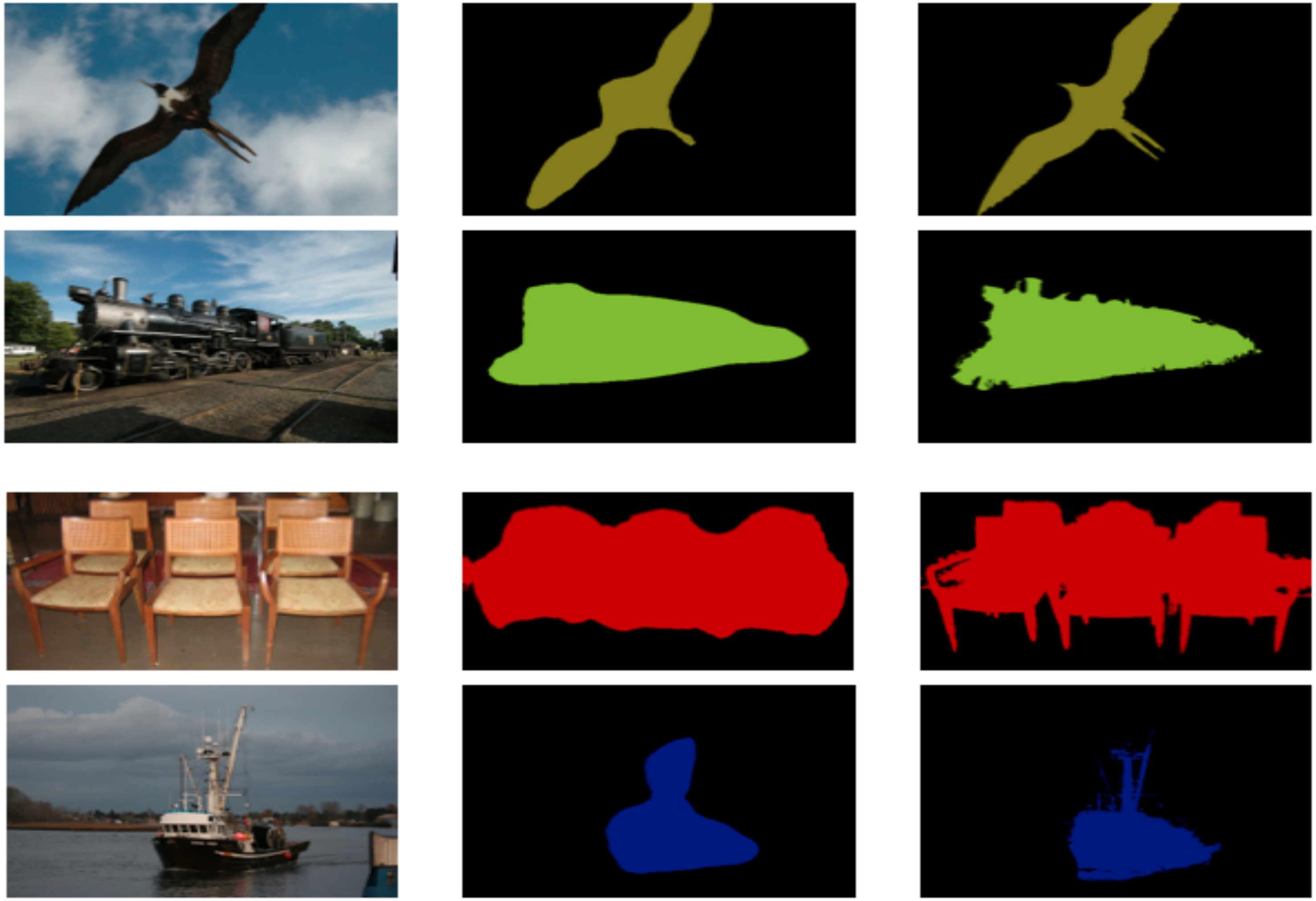
Qualitative Results



FCNN

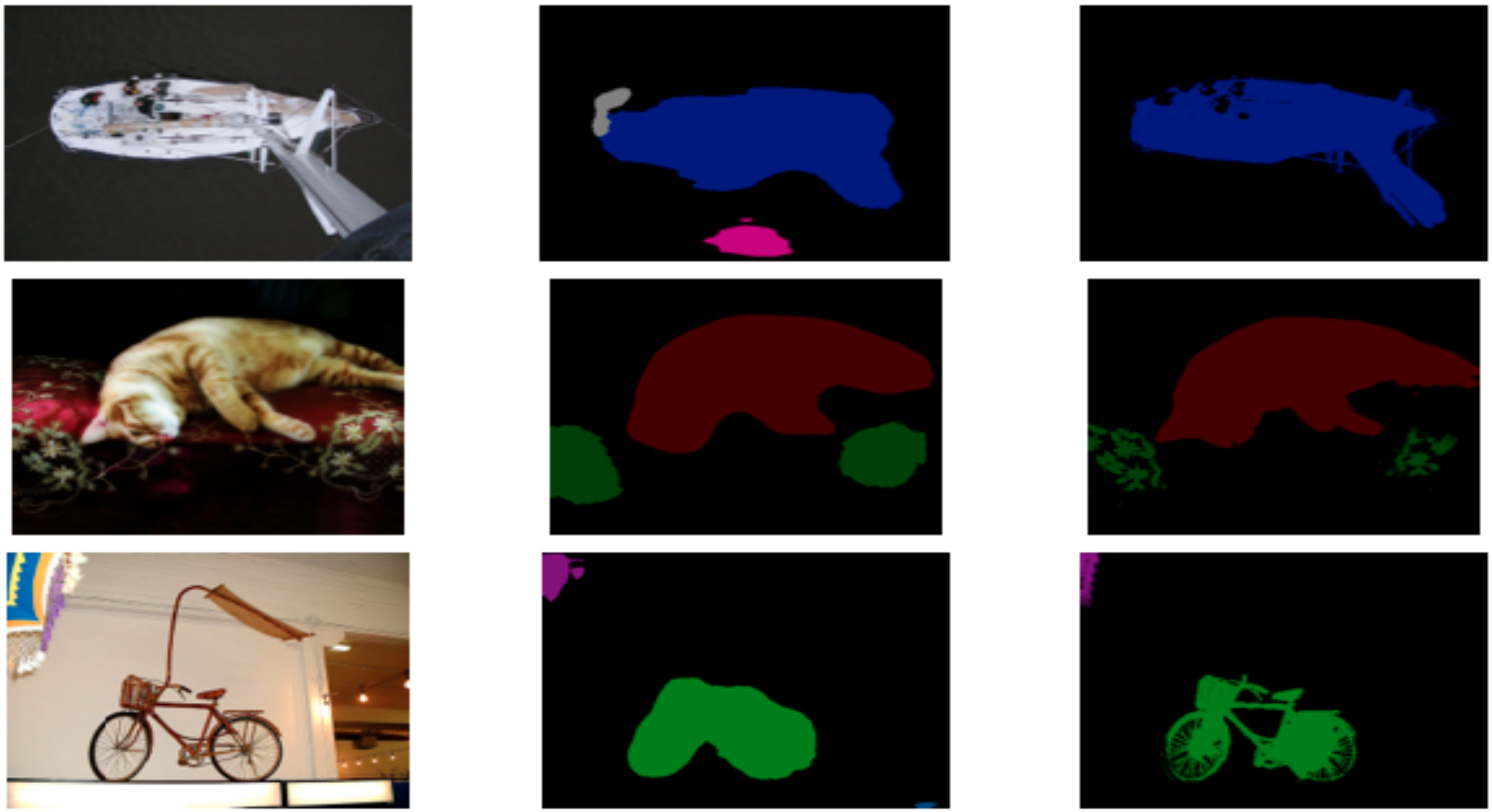
FCNN-DCRF

Qualtiative Results



FCNN FCNN-DCRF

Qualitative Results



FCNN

FCNN-DCRF

Indicative Results

FCNN

FCNN-DCRF

Comparison to state-of-the-art (Pascal VOC test)

Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF (our)	66.4
DeepLab-MSc-CRF (our)	67.1

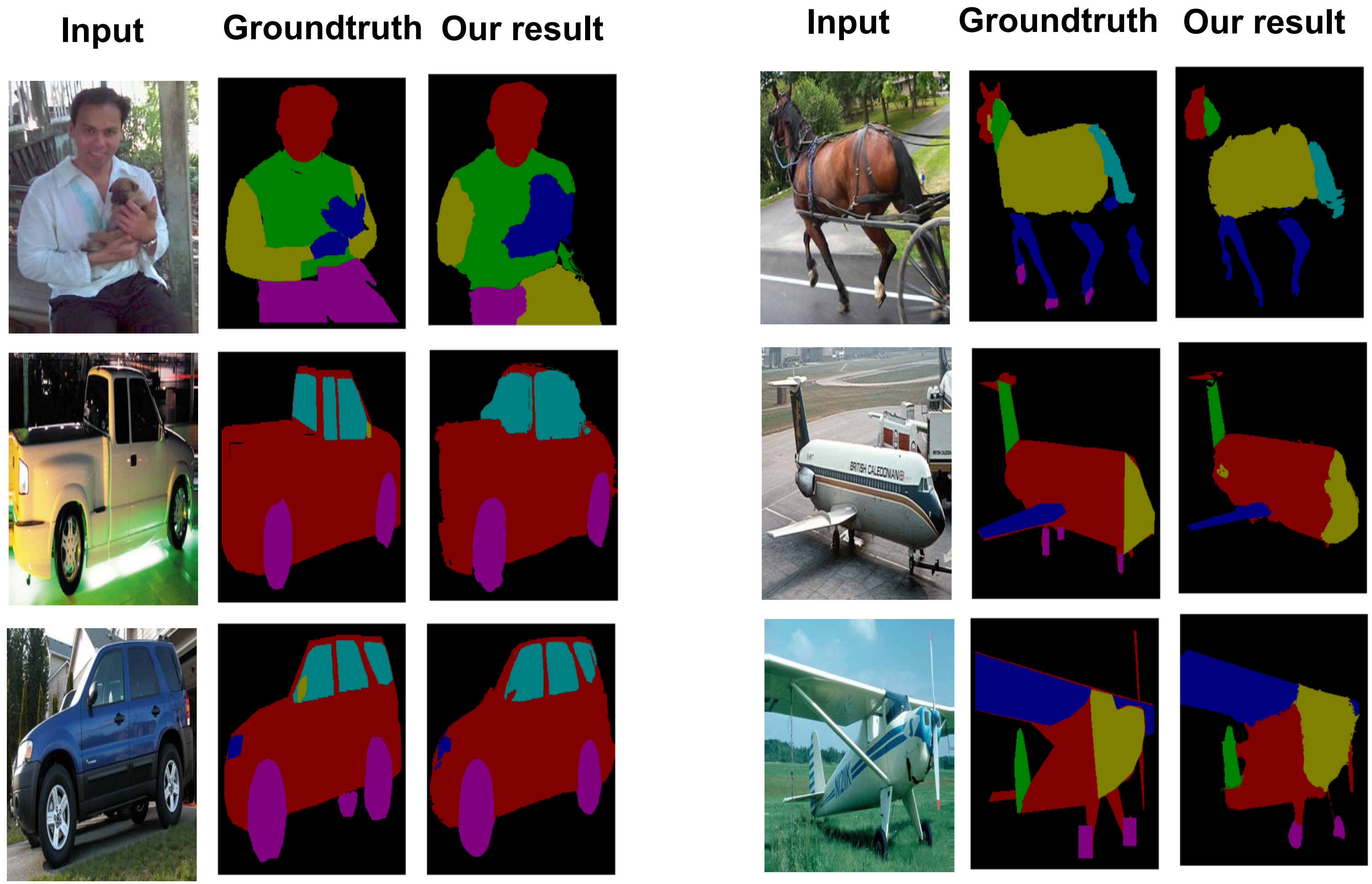
Pre-CNN: Up to 50% → **CNN: 60-64%** → **CNN + CRF: >67%**

Pascal Train: 67% → **Coco + Pascal 71%**

G. Papandreou, et al, Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation, arxiv 2015

Current: 74.7 end-to-end S. Zheng, et al. CRFs as recurrent neural networks. In ICCV, 2015.

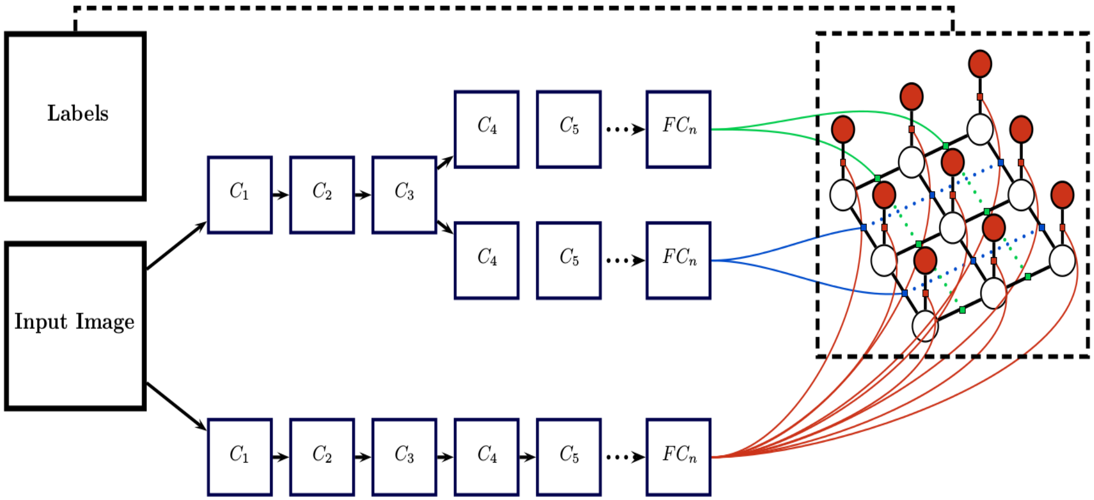
Semantic Part Segmentation

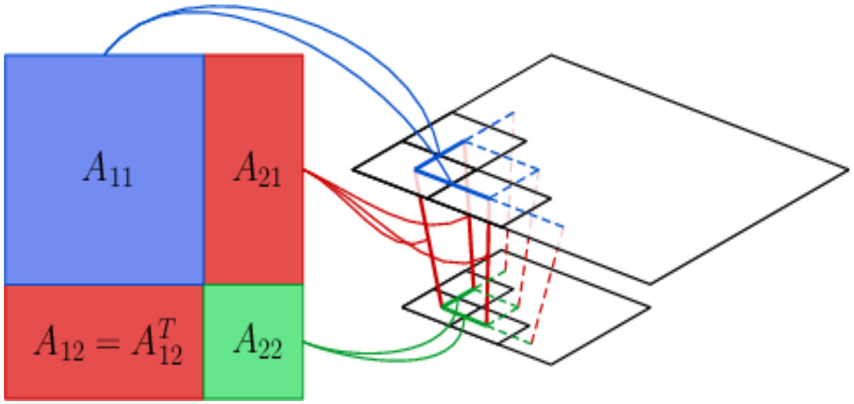


S. Tsogkas, G. Papandreou, I. Kokkinos, and A. Vedaldi, Semantic Part Segmentation using high-level guidance, Arxiv, 2015

Fast, Exact, and Multi-Scale Inference for FCNN-CRF

S. Chandra

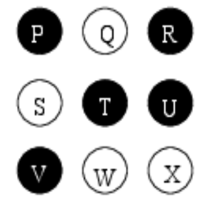




S. Chandra, I. Kokkinos, Fast, Exact and Multi-Scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs, arXiv:1603.08358

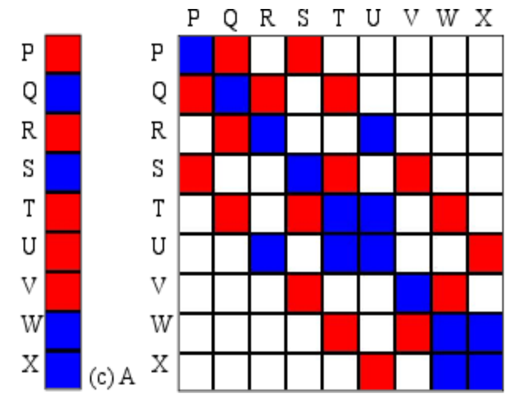
Gaussian Random Fields: Random Fields for dummies

$$\pi(\mathbf{x}) = \frac{1}{Z}\exp\left(-\mathbf{x}^T\Theta\mathbf{x} + \theta^T\mathbf{x}\right)$$



$$\Theta\mathbf{x}^* = \theta$$

$\mathbf{x}^*$



θ Θ

Maximum-A-Posteriori inference =
Minimum Mean-Squared Error inference =
solution of linear system

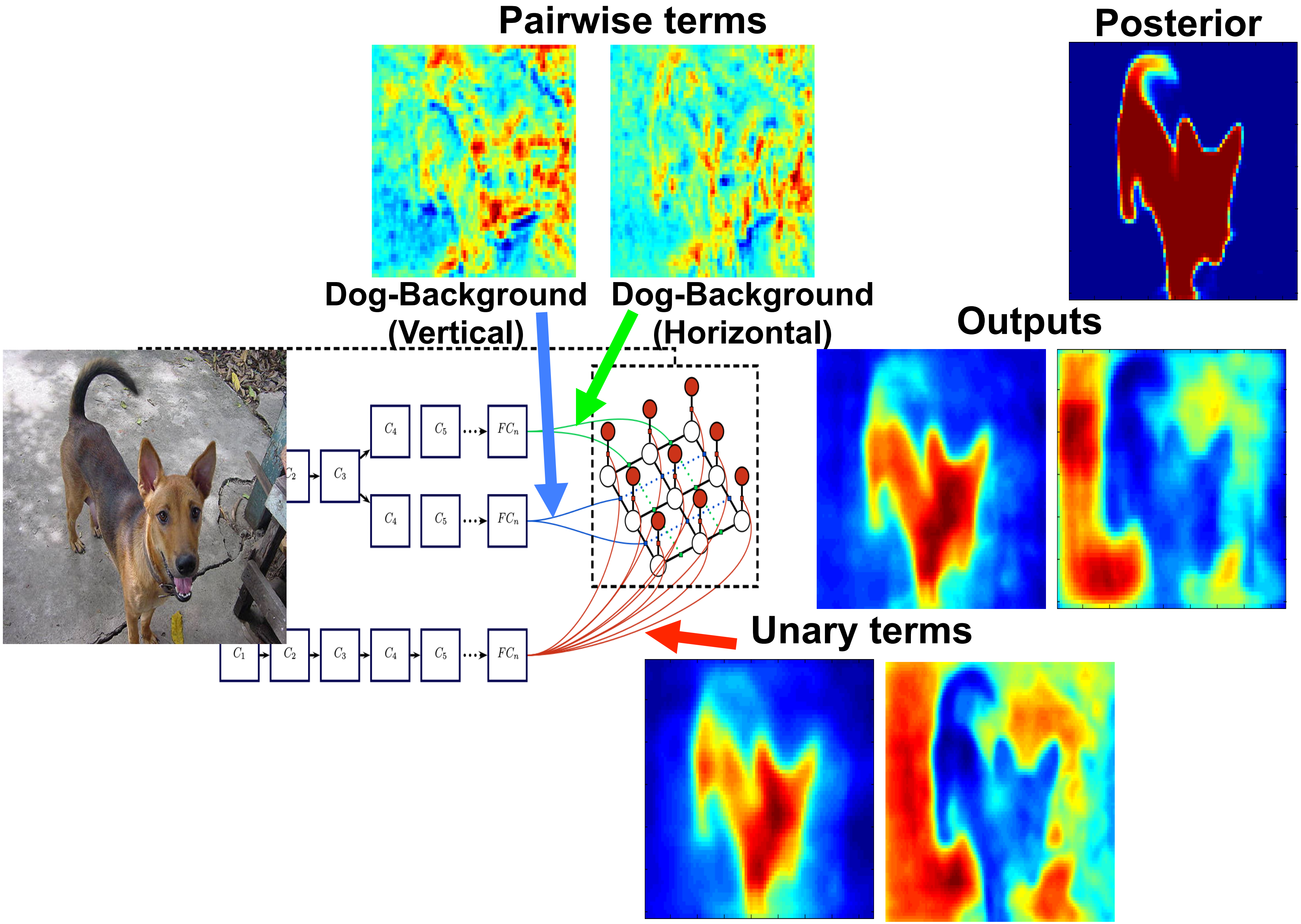
Gaussian MRF: blurry samples (hard to have outliers)

Gaussian CRF: image-based pairwise terms (e.g. discontinuity -preserving)

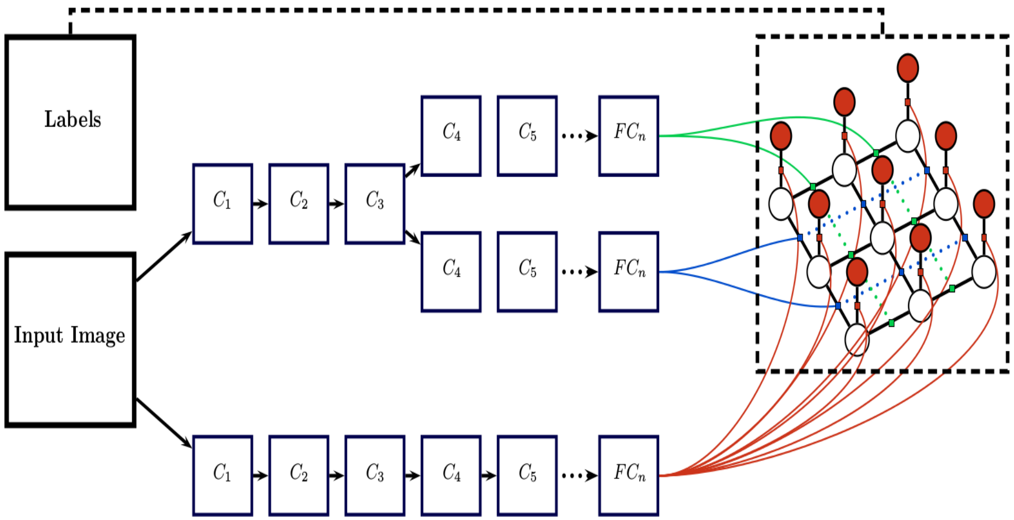
Jancsary, Nowozin, Sharp & Rother, Regression Tree Fields, CVPR12

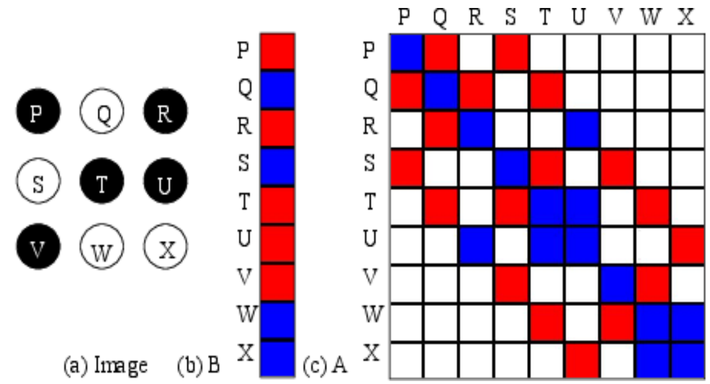
Tappen, Liu, Adelson & Freeman, Learning Gaussian CRFs for low-level vision, CVPR07

Deep Gaussian Conditional Random Field



Deep Gaussian Conditional Random Field vs. DenseCRF





	Deep Gaussian CRF	Dense CRF
Variables	continuous	discrete
Inference	exact (linear system)	approximate (mean-field)
Learning	exact (linear system)	BackProp on mean-field
Unary terms	CNN-based	CNN-based
Pairwise terms	CNN-based	parametric (Gaussian form)

Linear systems & Gaussian CRFs

$$A\mathbf{x} = B$$

$$\Theta \mathbf{x}^* = \theta$$

Gauss-Seidel:

sequential Mean-Field

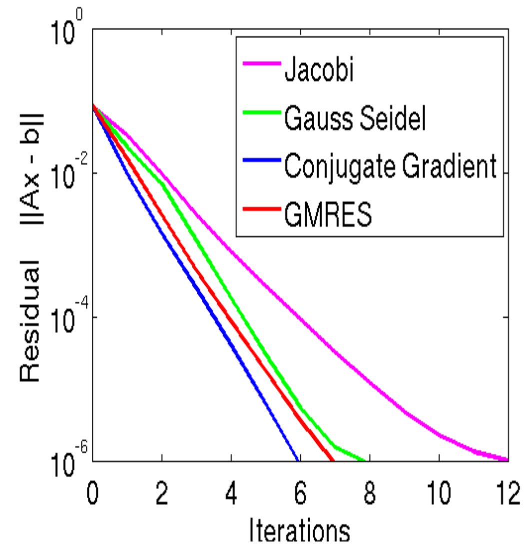
$$x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left\{ b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x_j^{(k)} \right\}$$

Jacobi:

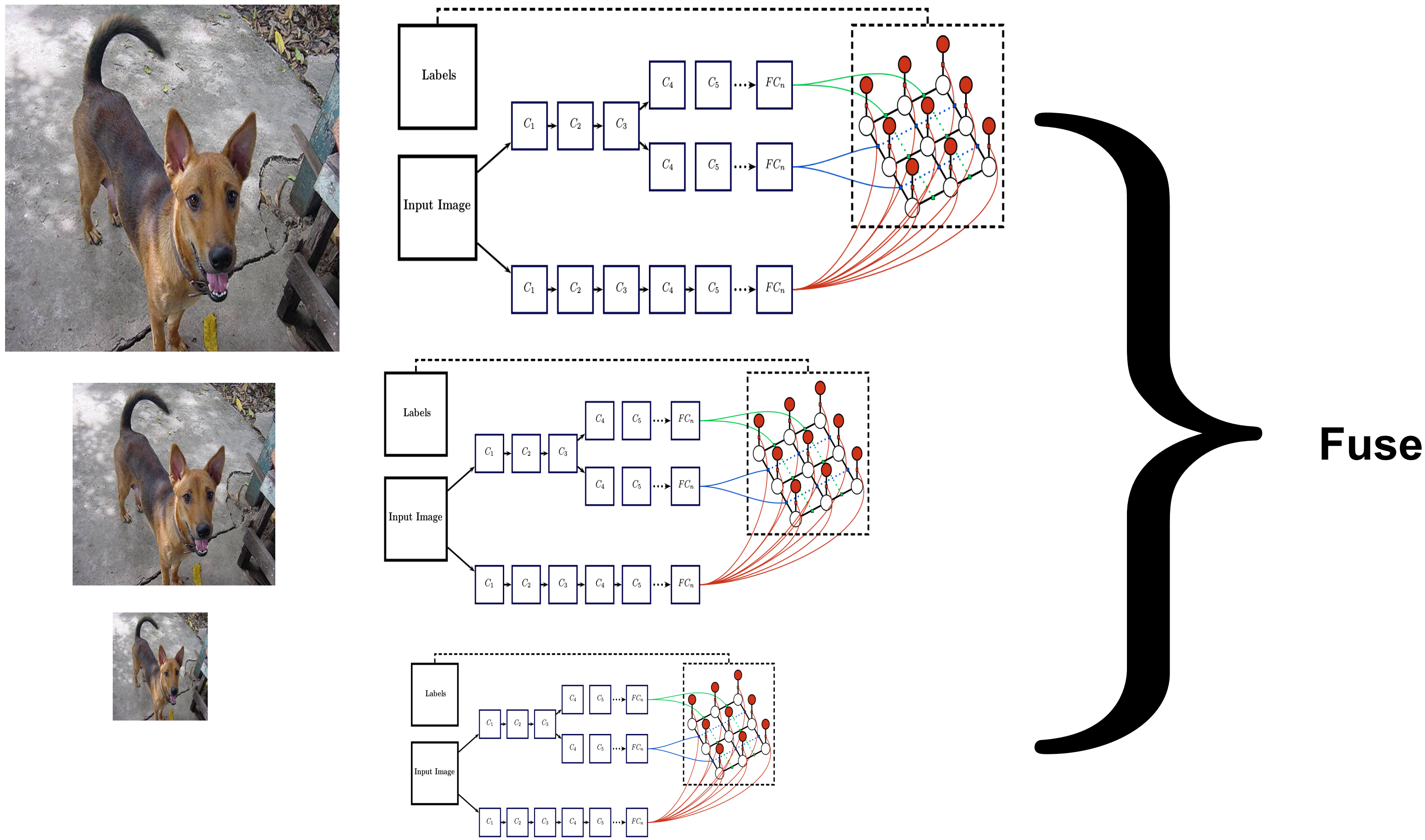
parallel Mean-Field

$$x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left\{ b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right\}$$

Conjugate gradients: 2x faster!

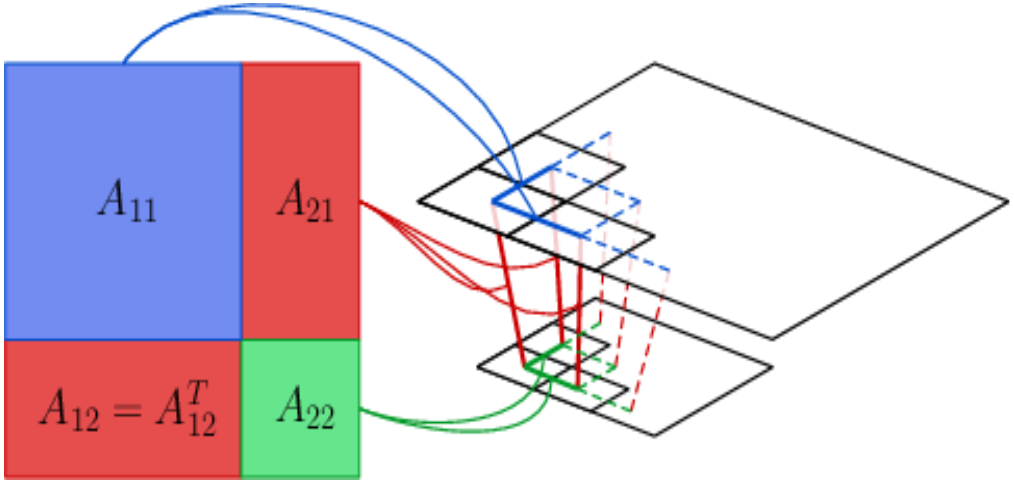


Naïve Multi-Resolution Semantic Segmentation



L.-C. Chen, Y. Yang, J. Wang, W. Xu and A. Yuille, 'Attention to Scale: Scale-aware Semantic Image Segmentation, CVPR 2016
I. Kokkinos, Pushing the Boundaries of Boundary Detection using Deep Learning, ICLR 2016

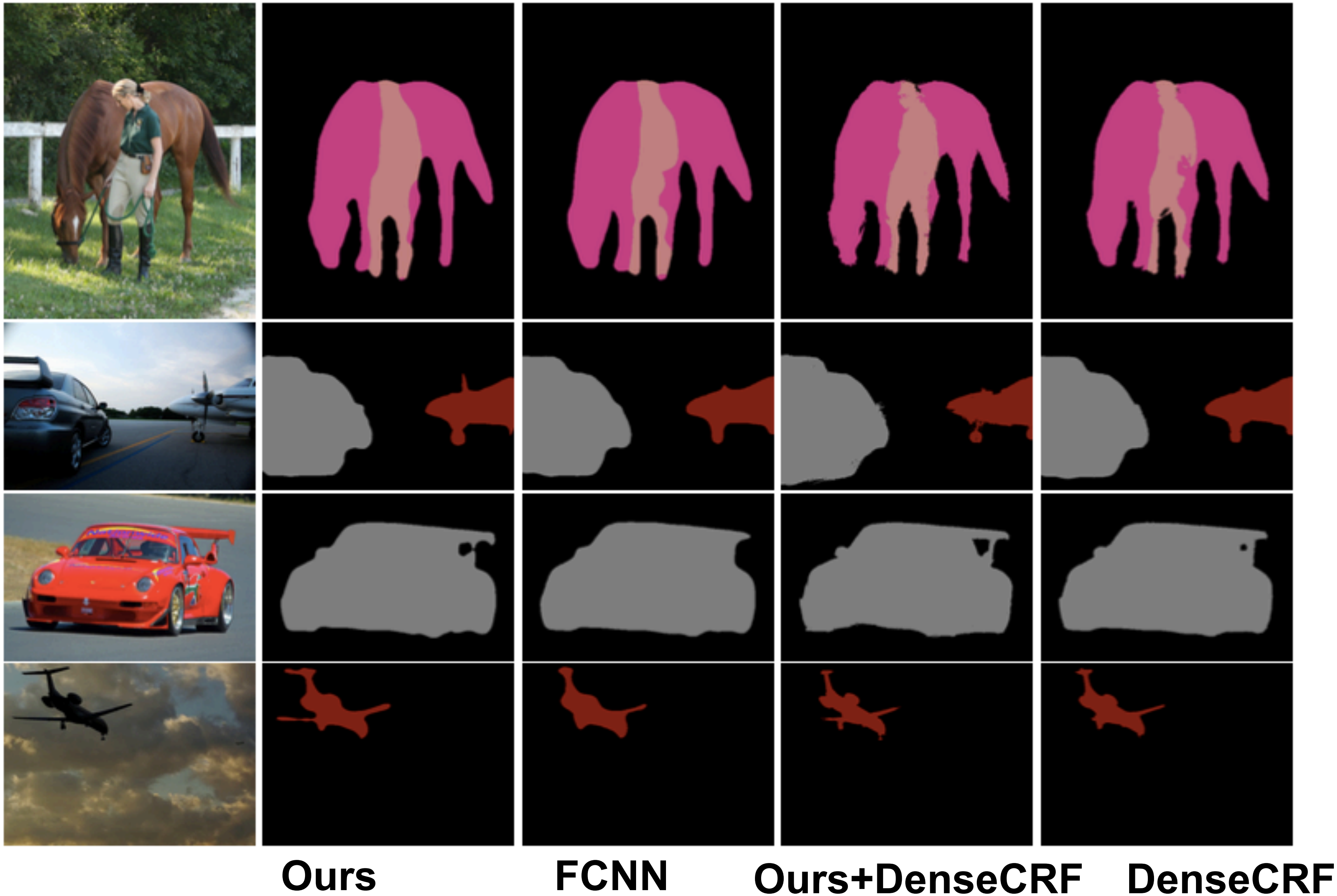
Linear systems & Multi-resolution CRFs



Learn to enforce coupling of different results

Consistently better results than decoupled learning!

Improvements/Complementarity with DenseCRF



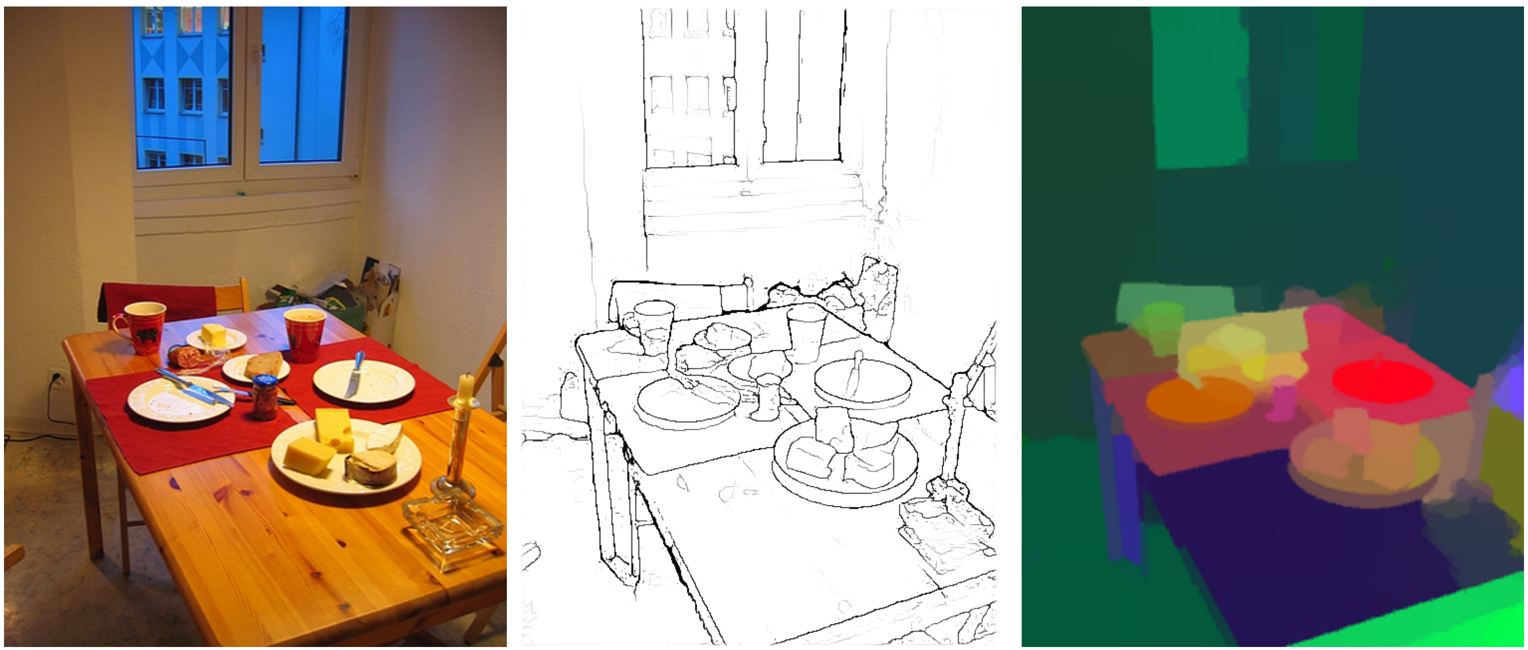
Quantitative Results

Method	IoU	IoU after *dense CRF*
Basenet	72.72	73.78
QO_4	73.41	75.13
QO_4^{mres}	73.86	75.46

Method	mean IoU (%)
DeepLab-CRF (Chen et al., 2014)	66.4
DeepLab-MSc-CRF (Chen et al., 2014)	67.1
DeepLab-CRF-7x7 (Chen et al., 2014)	70.3
DeepLab-CRF-LargeFOV (Chen et al., 2014)	70.3
DeepLab-MSc-CRF-LargeFOV (Chen et al., 2014)	71.6
Deeplab-Cross-Joint (Chen et al., 2015a)	73.9
CRFRNN (Zheng et al., 2015)	74.7
Adelaide Context (Lin et al., 2016)	77.8
Deep Parsing Network (Liu et al., 2015)	77.4
Ours (QO_4^{mres})	75.5

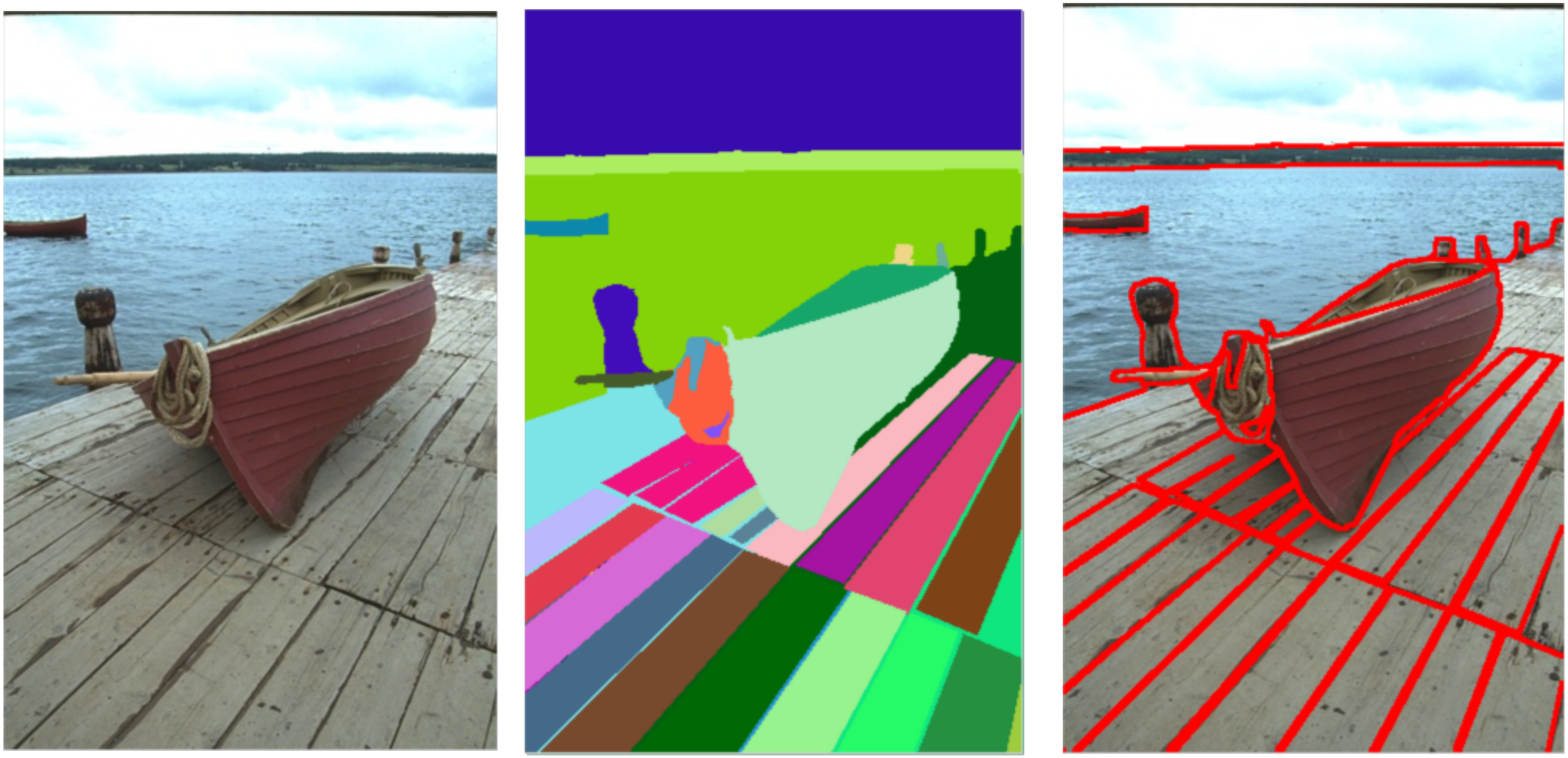
This talk: controlling DCNNs for low- and high- level tasks

- -Classification & Detection
- -Semantic Segmentation
- -Boundary Detection
- -Feature Descriptors



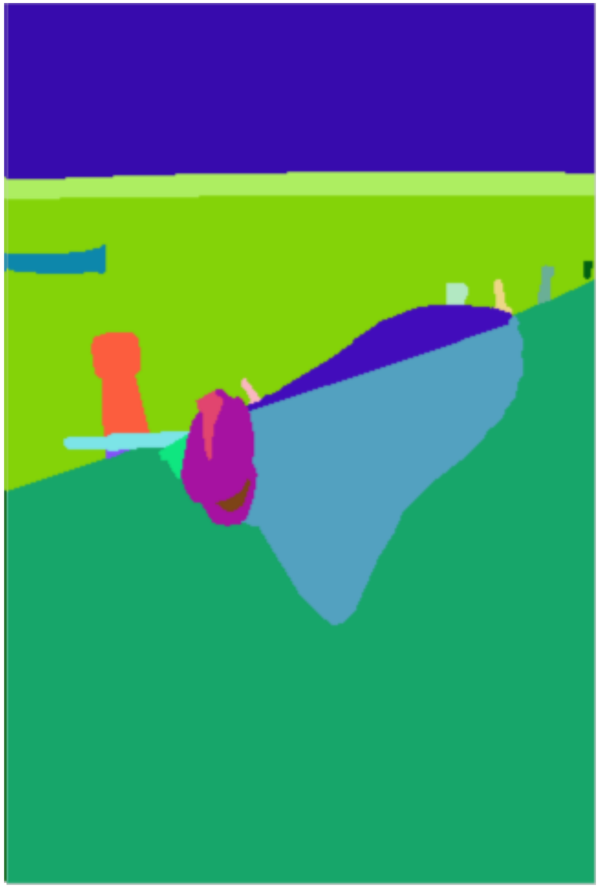
I. Kokkinos, Pushing the Boundaries of Boundary Detection using Deep Learning, ICLR 2016
(earlier title: ‘Surpassing Humans in Boundary Detection’)

Can humans do it?



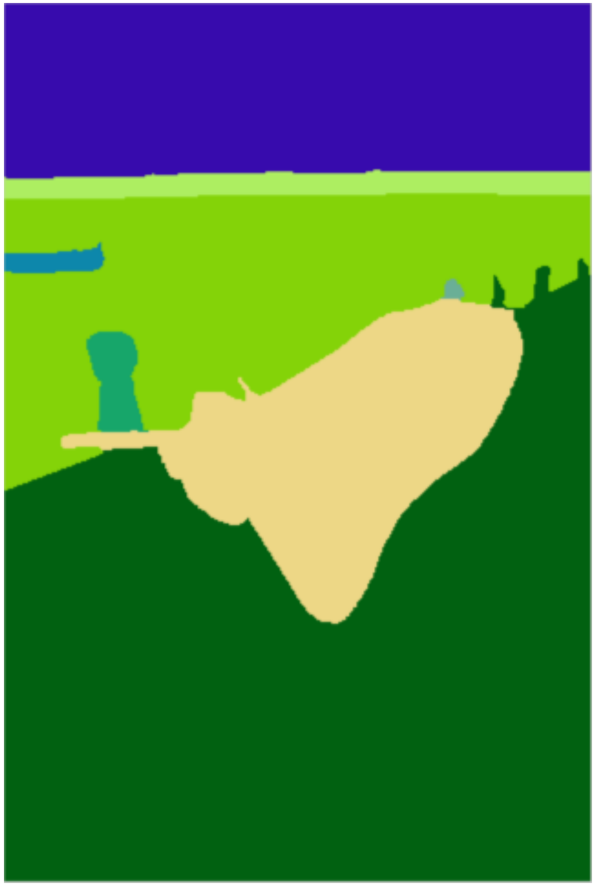
Segmentation: task-agnostic, ill-posed

Can humans do it?



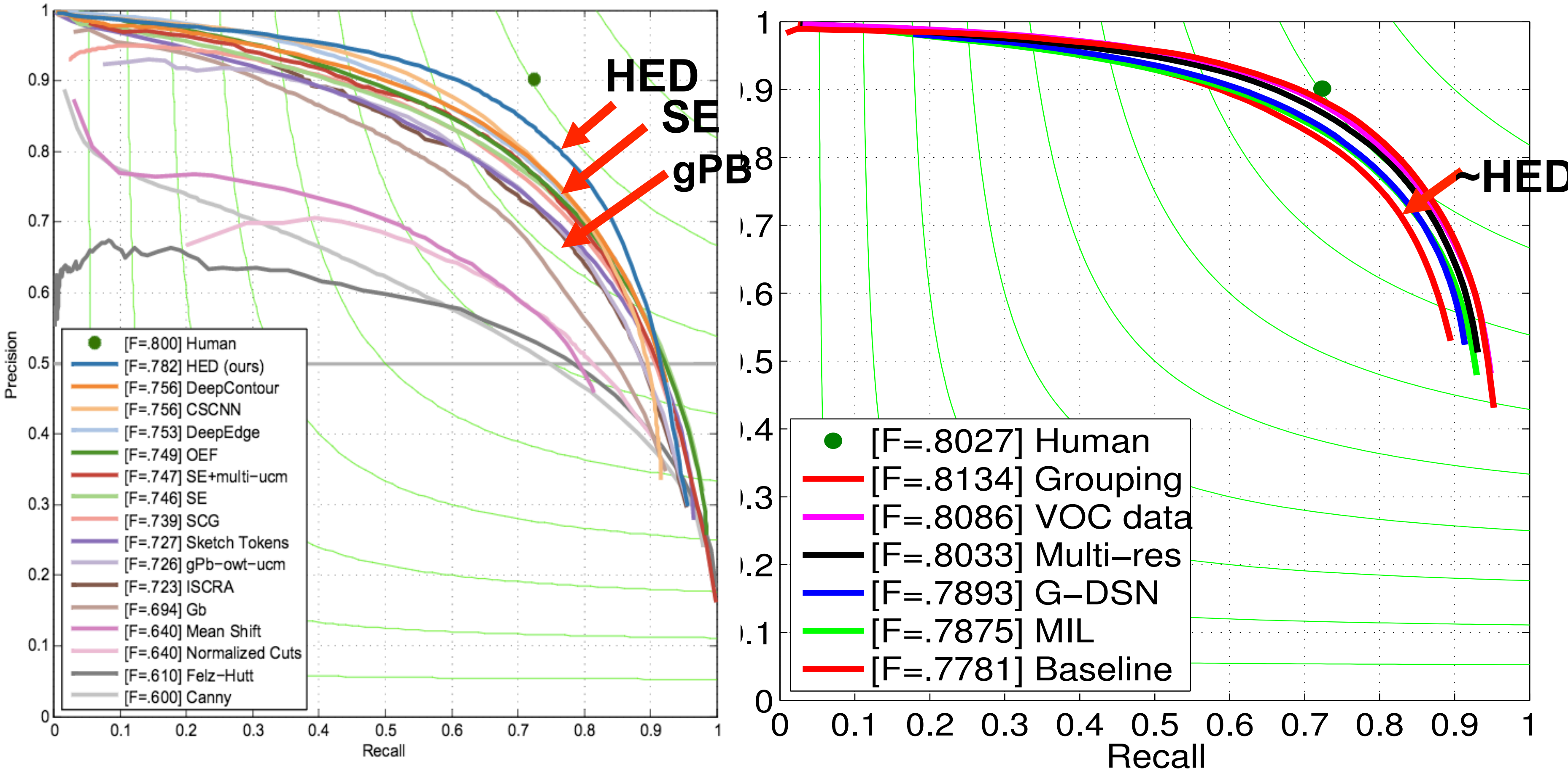
Segmentation: task-agnostic, ill-posed

Can humans do it?


Segmentation: task-agnostic, ill-posed

30 years of boundary detection



S. Xie and Z. Tu, Holistically-Nested Edge Detection, ICCV 2015

I. Kokkinos, Pushing the boundaries of boundary detection using deep learning, ICLR 2016

This work

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

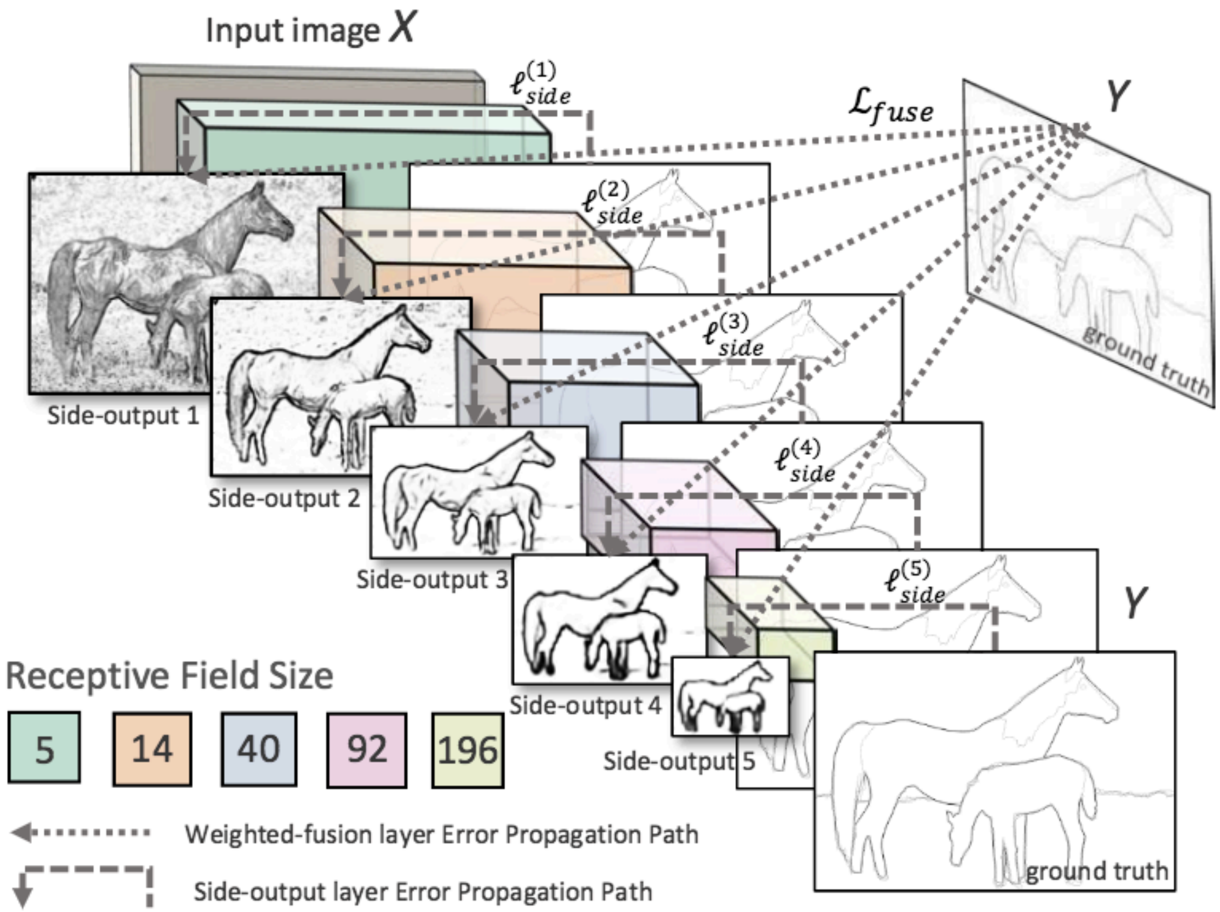
Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

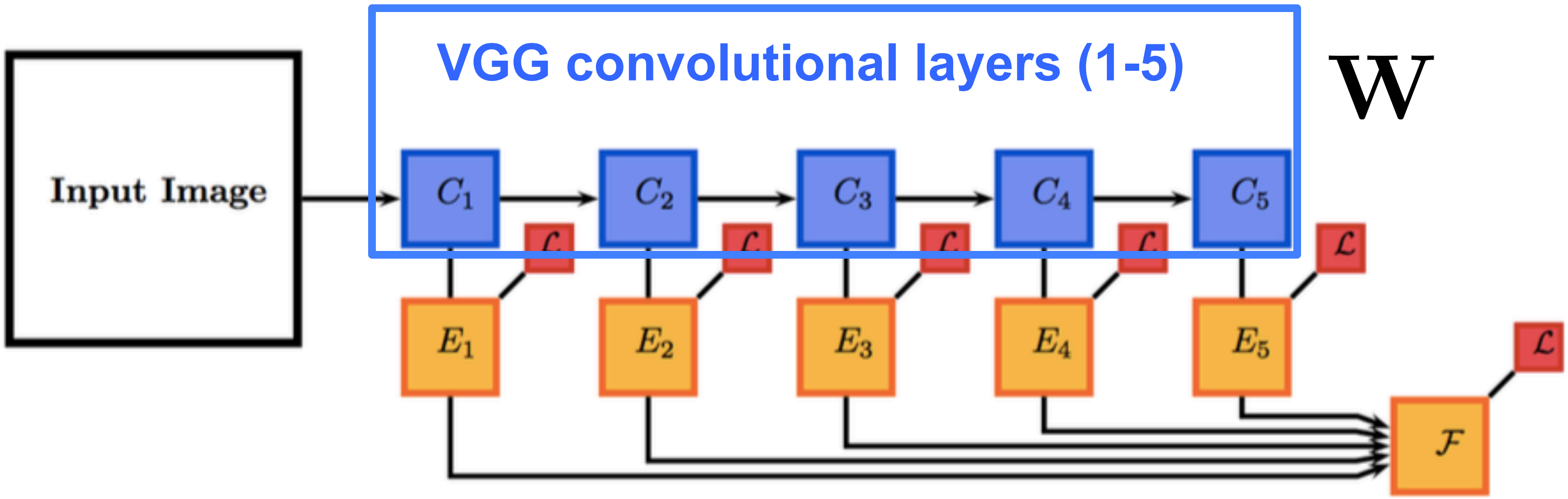
Network Architecture:

Tied Multi-Scale Networks
Grouping in DCNNs

Holistically-Nested Edge Detection network

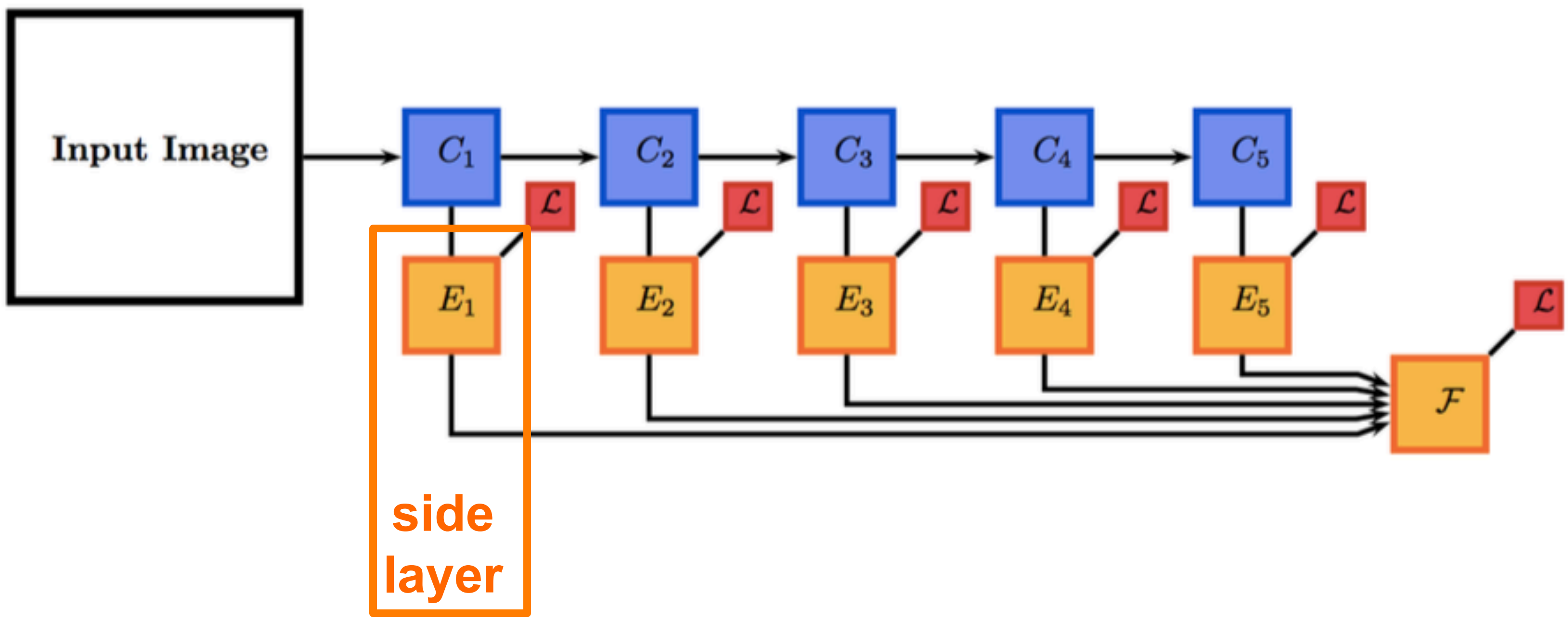


HED network



Outputs: $\mathbf{f}^m, \quad m = 1, \ldots, 5$

HED network

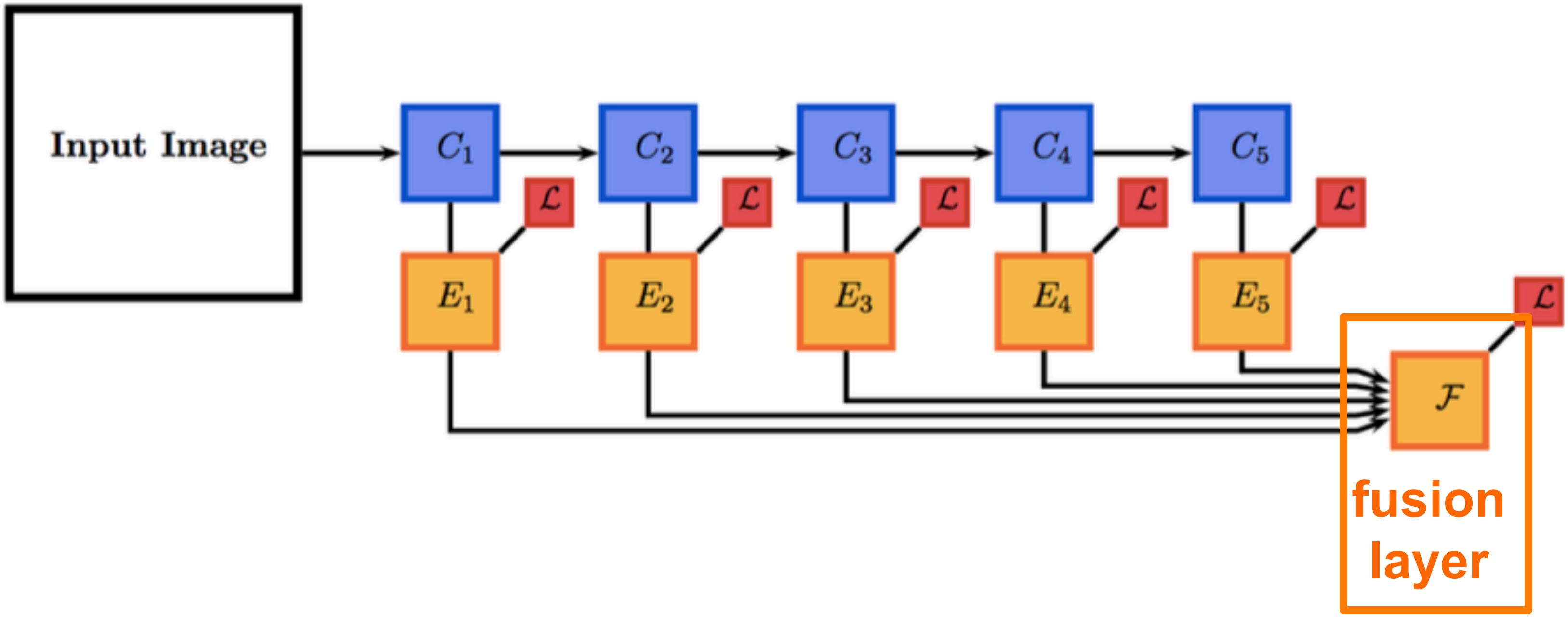


Parameters: $\mathbf{w}^m$

Inputs: $\mathbf{f}^m$

Outputs: $\mathbf{s}^m = \langle \mathbf{w}^m, \mathbf{f}^m \rangle \qquad m = 1, \ldots, 5$

HED network

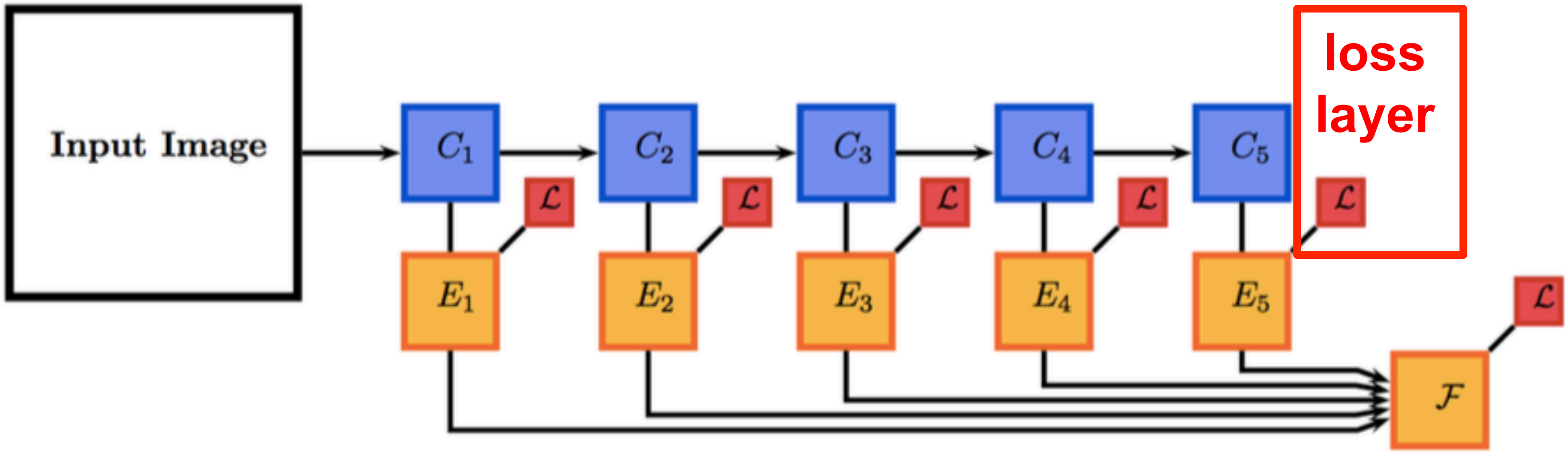


Parameters: $(\alpha_1, \ldots, \alpha_5)$

Inputs: $\mathbf{s}^1, \ldots, \mathbf{s}^5$

Outputs: $f = \sum_{m=1}^{5} \alpha_m \mathbf{s}^m$

HED network



$$l^m(\mathbf{W}, \mathbf{w}^{(m)}) \doteq \sum_{j \in Y} w_{\hat{y}_j} S(\hat{y}_j, s_j^m) \qquad s_j^m = \langle \mathbf{w}^{(m)}, \mathbf{f}_j \rangle$$

$$\mathcal{L}_{side}(\mathbf{W}, \mathbf{w}) = \sum_{m=1}^{M} \alpha_m l^m(\mathbf{W}, \mathbf{w}^{(m)})$$

$$\mathcal{L}_{HED}(\mathbf{W}, \mathbf{w}, \mathbf{h}) = \mathcal{L}_{side}(\mathbf{W}, \mathbf{w}) + \mathcal{L}_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{h})$$

This work in a nutshell

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

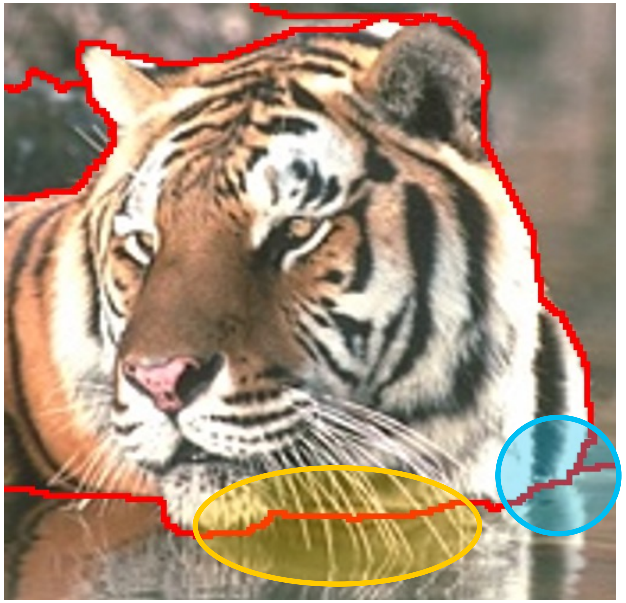
Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

Network Architecture:

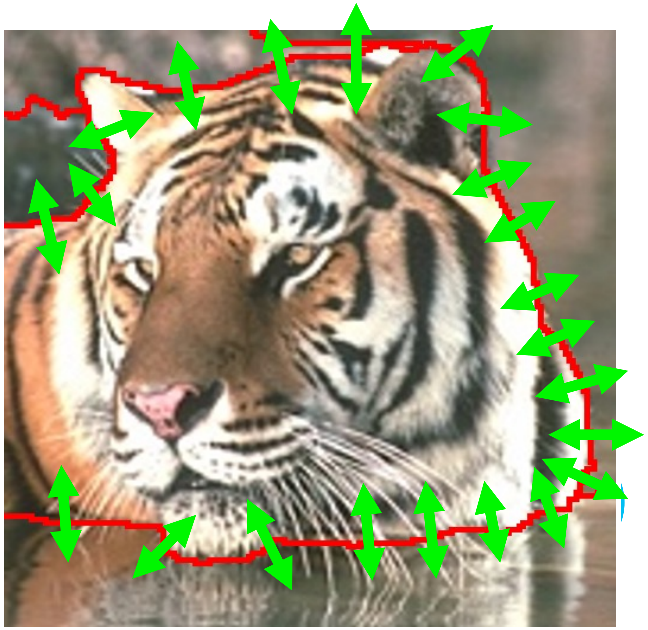
Tied Multi-Scale Networks
Grouping in DCNNs

Ambiguity in boundary annotations



Common interpretation, but different position information!

Ambiguity in boundary annotations



Solution: take into account annotator inaccuracies

Ambiguity in boundary annotations

$$(x_j, y_j) \rightarrow (\{x_b\}, y_j), b \in \mathcal{B}_j$$

$$l(y_j, s_j) \rightarrow l(y_j, \max_{b \in \mathcal{B}_j} s_b)$$

For every positive point, gather set of locations that can `support' it

False negative if no such point leads to a positive decision

Method	Baseline	MIL	G-DSN	M-Scale	VOC	Grouping
ODS	0.7781	0.7863	0.7892	0.8033	0.8086	0.8134
OIS	0.7961	0.8083	0.8106	0.8196	0.8268	0.8308
AP	0.804	0.802	0.789	0.8483	0.861	0.866

This work in a nutshell

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

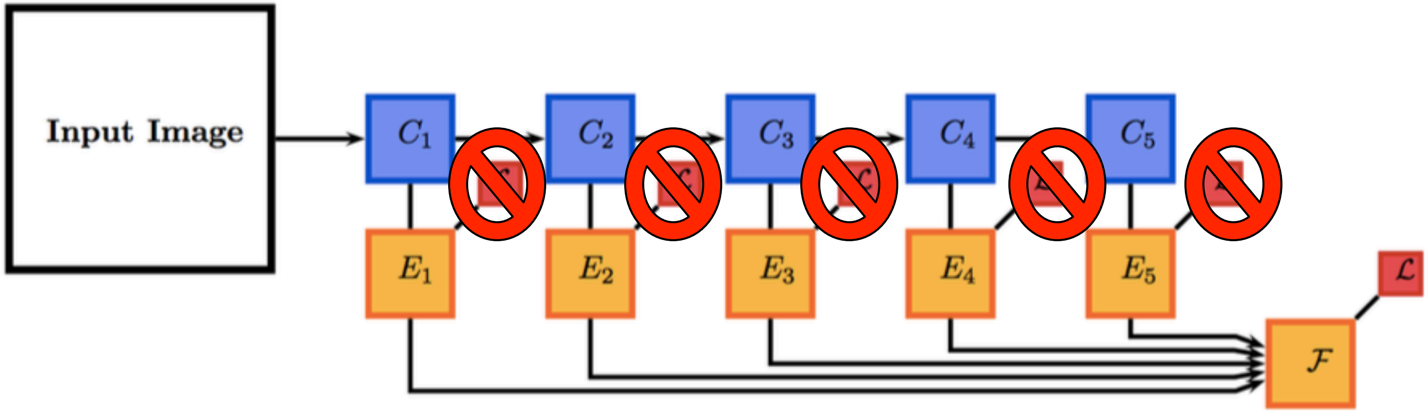
Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

Network Architecture:

Tied Multi-Scale Networks
Spectral Clustering in DCNNs

Holistically-Nested Edge Detection Training



$$\mathcal{L}(\mathbf{W}, \mathbf{w}, \mathbf{h}) = \mathcal{L}_{side}(\mathbf{W}, \mathbf{w}) + \mathcal{L}_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{h})$$

DSN's side losses: steer network parameters to correct values

$$\mathcal{L}^{(t)}(\mathbf{W}, \mathbf{w}, \mathbf{h}) = (1 - \frac{t}{T})\mathcal{L}_{side}(\mathbf{W}, \mathbf{w}) + \mathcal{L}_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{h})$$

Graduated DSN: remove side losses as training progresses

Method	Baseline	MIL	G-DSN	M-Scale	VOC	Grouping
ODS	0.7781	0.7863	0.7892	0.8033	0.8086	0.8134
OIS	0.7961	0.8083	0.8106	0.8196	0.8268	0.8308
AP	0.804	0.802	0.789	0.8483	0.861	0.866

This work in a nutshell

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

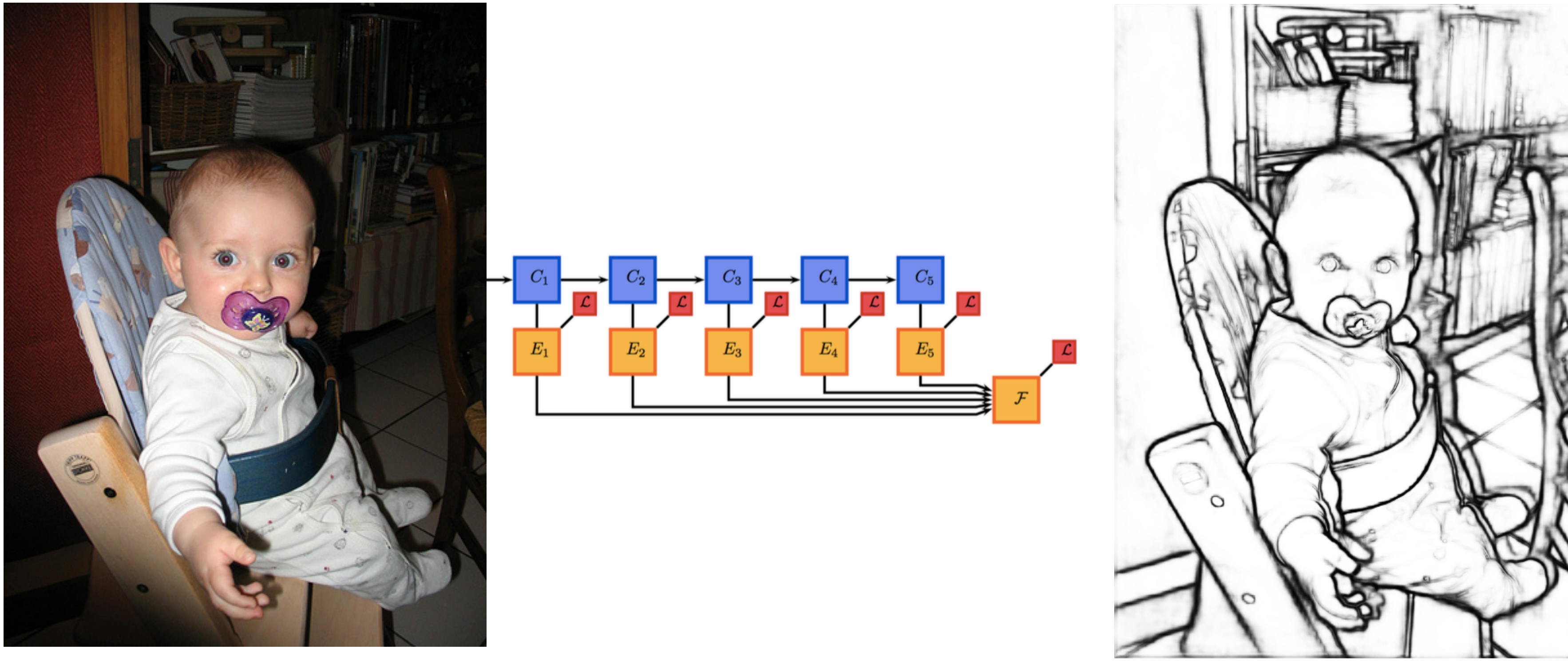
Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

Network Architecture:

Tied Multi-Scale Networks
Grouping in DCNNs

Boundary CNN scale-space

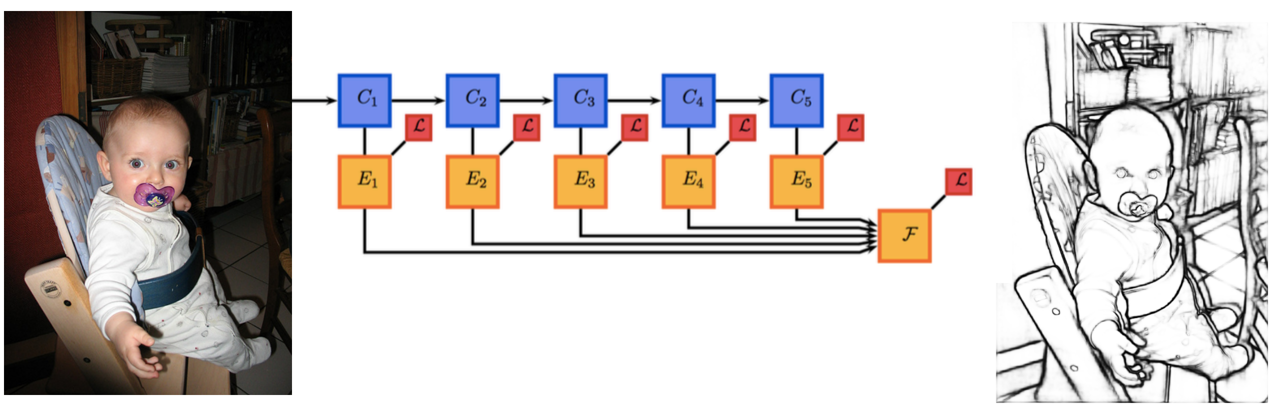


Boundary CNN scale-space

$\downarrow \frac{1}{2}$

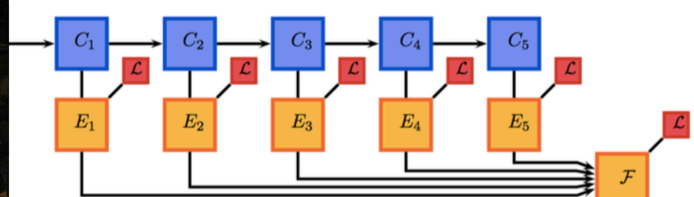
Boundary CNN scale-space

$$\downarrow \frac{1}{2}$$



Boundary CNN scale-space

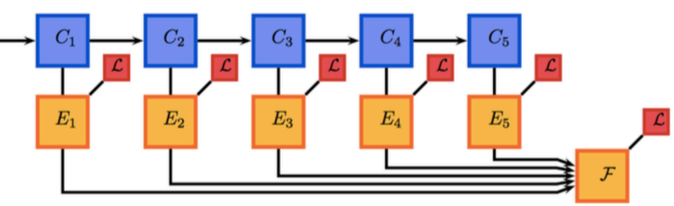
$\downarrow \frac{1}{2}$

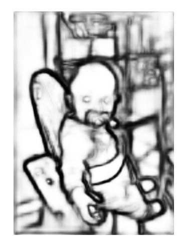


$\uparrow 2$

Boundary CNN scale-space

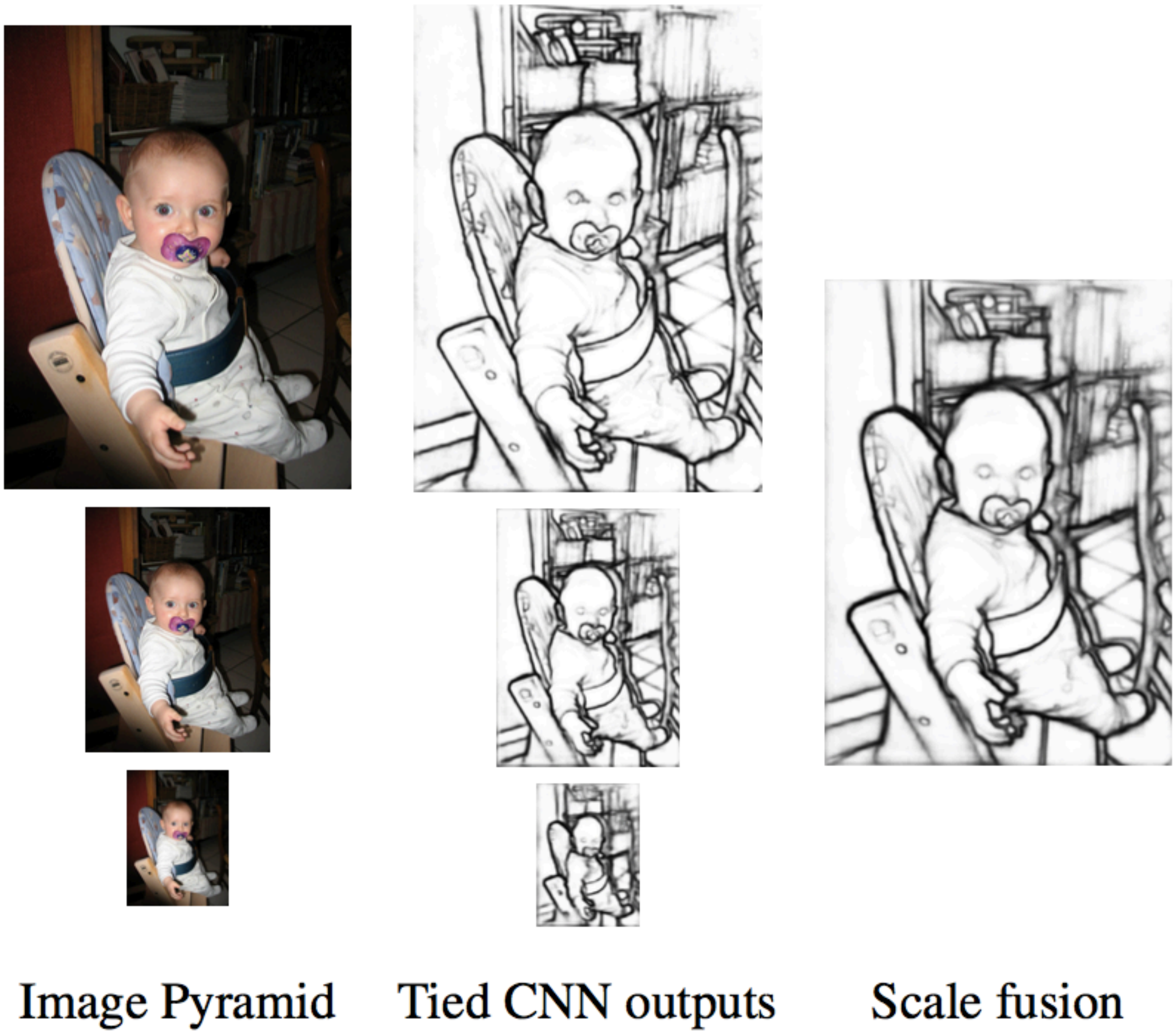
$\downarrow \frac{1}{4}$



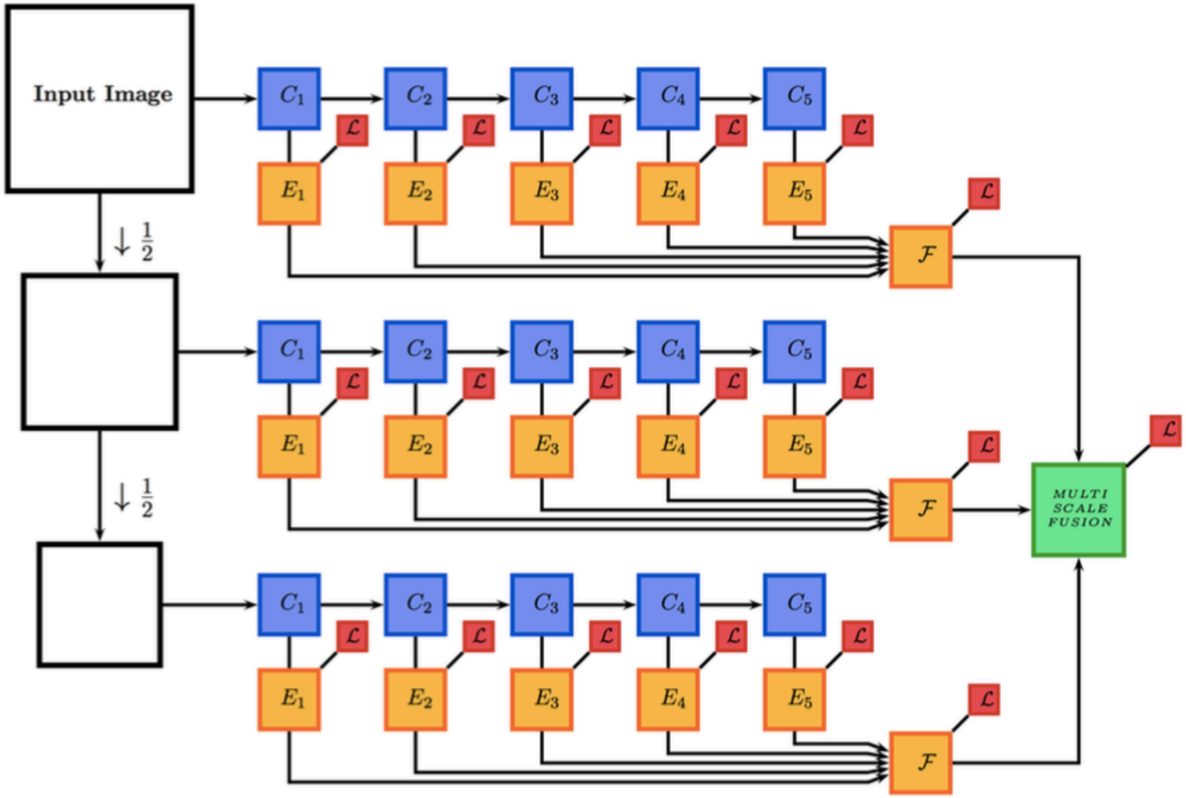


$\uparrow 4$

Multi-Scale DSN



Multi-Scale DSN



-tied weights

-end-to-end training

Method	Baseline	MIL	G-DSN	M-Scale	VOC	Grouping
ODS	0.7781	0.7863	0.7892	0.8033	0.8086	0.8134
OIS	0.7961	0.8083	0.8106	0.8196	0.8268	0.8308
AP	0.804	0.802	0.789	0.8483	0.861	0.866

Pascal Context Dataset

The Role of Context for Object Detection and Semantic Segmentation in the Wild , R. Mottaghi, et al, CVPR 2014

-tied weights -end-to-end training -more data ☺

Method	Baseline	MIL	G-DSN	M-Scale	VOC	Grouping
ODS	0.7781	0.7863	0.7892	0.8033	0.8086	0.8134
OIS	0.7961	0.8083	0.8106	0.8196	0.8268	0.8308
AP	0.804	0.802	0.789	0.8483	0.861	0.866

This work in a nutshell

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

Network Architecture:

Tied Multi-Scale Networks
Grouping in DCNNs

This work in a nutshell

Starting point:

Holistically-Nested Edge Detection,
S. Xie and Z. Tu, ICCV 2015

Learning Techniques:

Multiple Instance Learning for Boundary Detection
Graduated Deep Supervised Networks

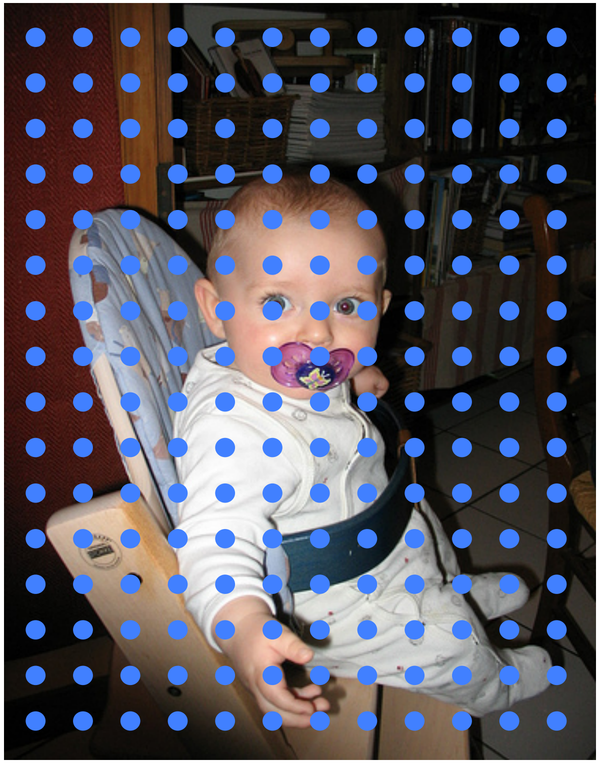
Network Architecture:

Tied Multi-Scale Networks
Grouping in DCNNs

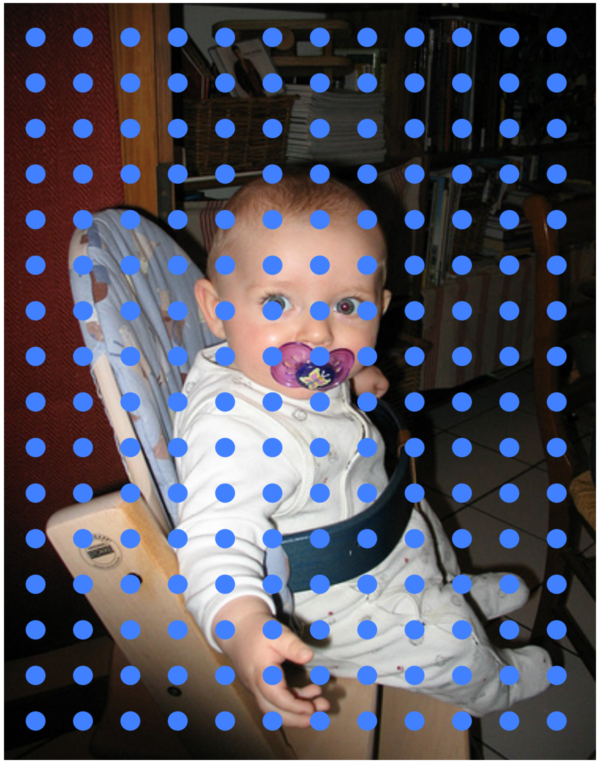
Shi & Malik, Normalized Cuts and Image Segmentation. PAMI 2000
Arbelaez, et al, Contour Detection and Hierarchical Image Segmentation. PAMI 2011
C. Ionescu et al, Matrix Backpropagation for Training Deep Networks with Structured Layers, ICCV 2015
Catanzaro et. al.: Efficient, high-quality image contour detection. ICCV 2009

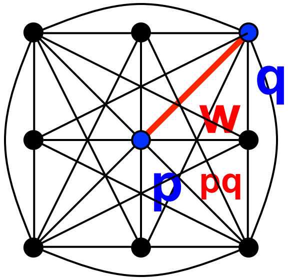
FCNNs + Spectral Clustering

FCNNs + Spectral Clustering



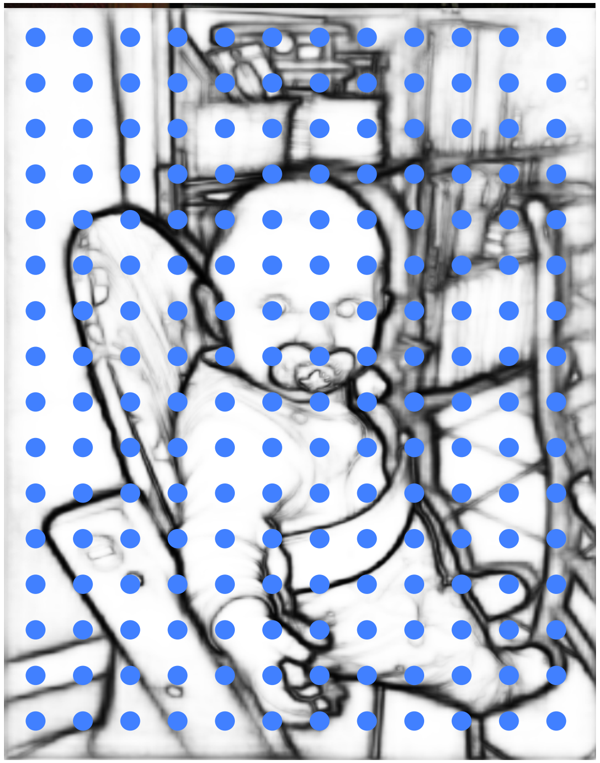
FCNNs + Spectral Clustering

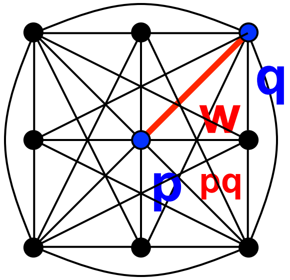




$$(\mathbf{D}-\mathbf{W})y=\lambda \mathbf{D}y$$

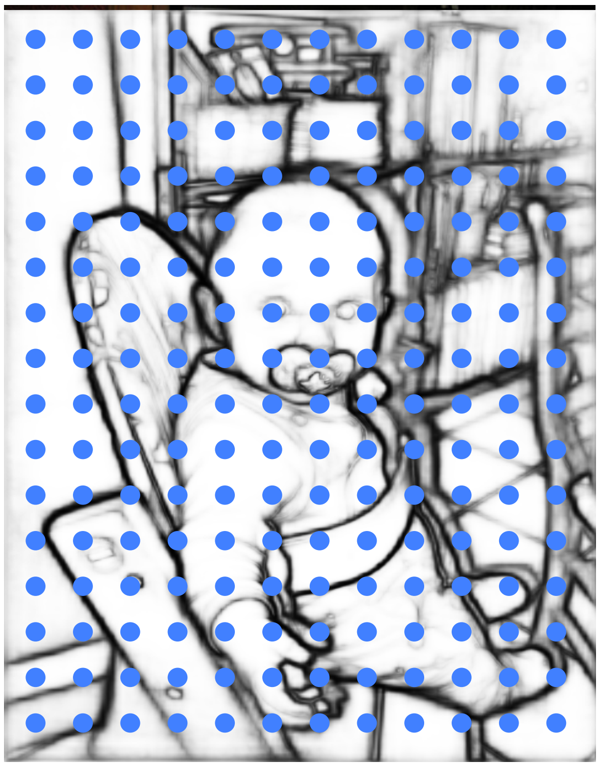
FCNNs + Spectral Clustering

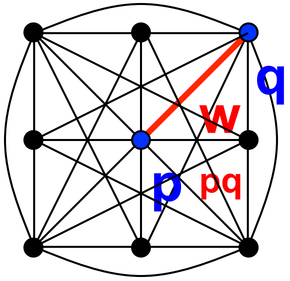




$$(\mathbf{D} - \mathbf{W})y = \lambda \mathbf{D} y$$

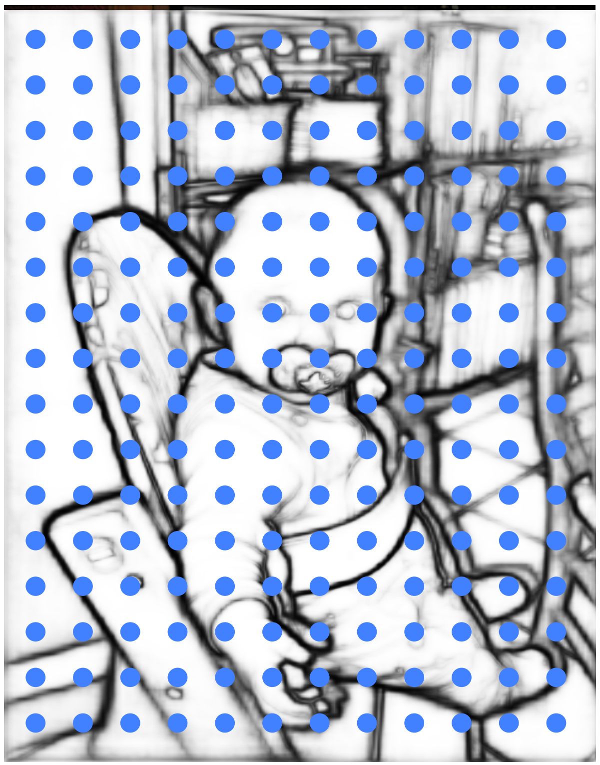
FCNNs + Spectral Clustering

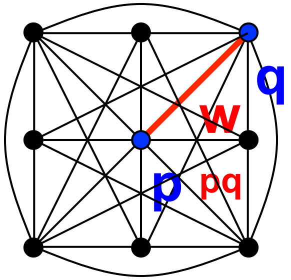




$$(\mathbf{D} - \mathbf{W})y = \lambda \mathbf{D} y.$$

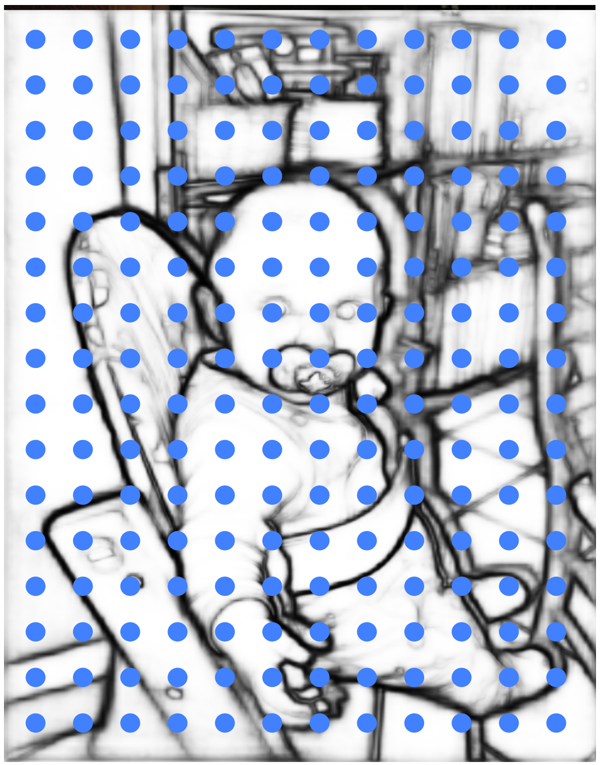
FCNNs + Spectral Clustering

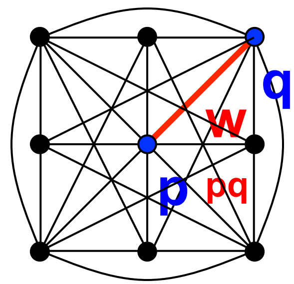




$$(\mathbf{D} - \mathbf{W})y = \lambda \mathbf{D}y.$$

FCNNs + Spectral Clustering





$$(\mathbf{D} - \mathbf{W})y = \lambda \mathbf{D} y$$

Catanzaro et. al.: Efficient, high-quality image contour detection. ICCV 2009

-Global Pb: ~60 seconds (CPU) -spectralPb layer: 0.2 seconds (GPU)

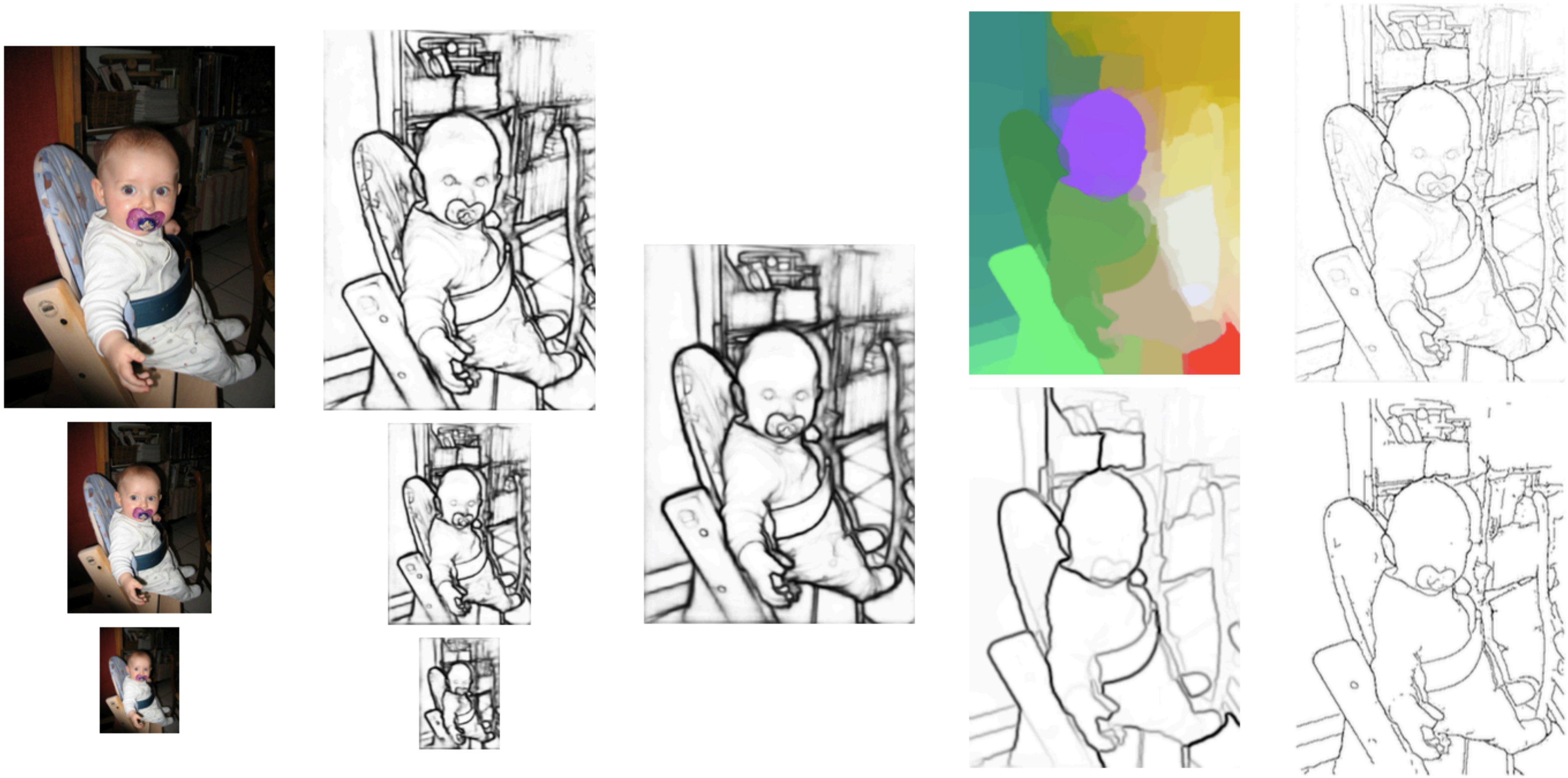
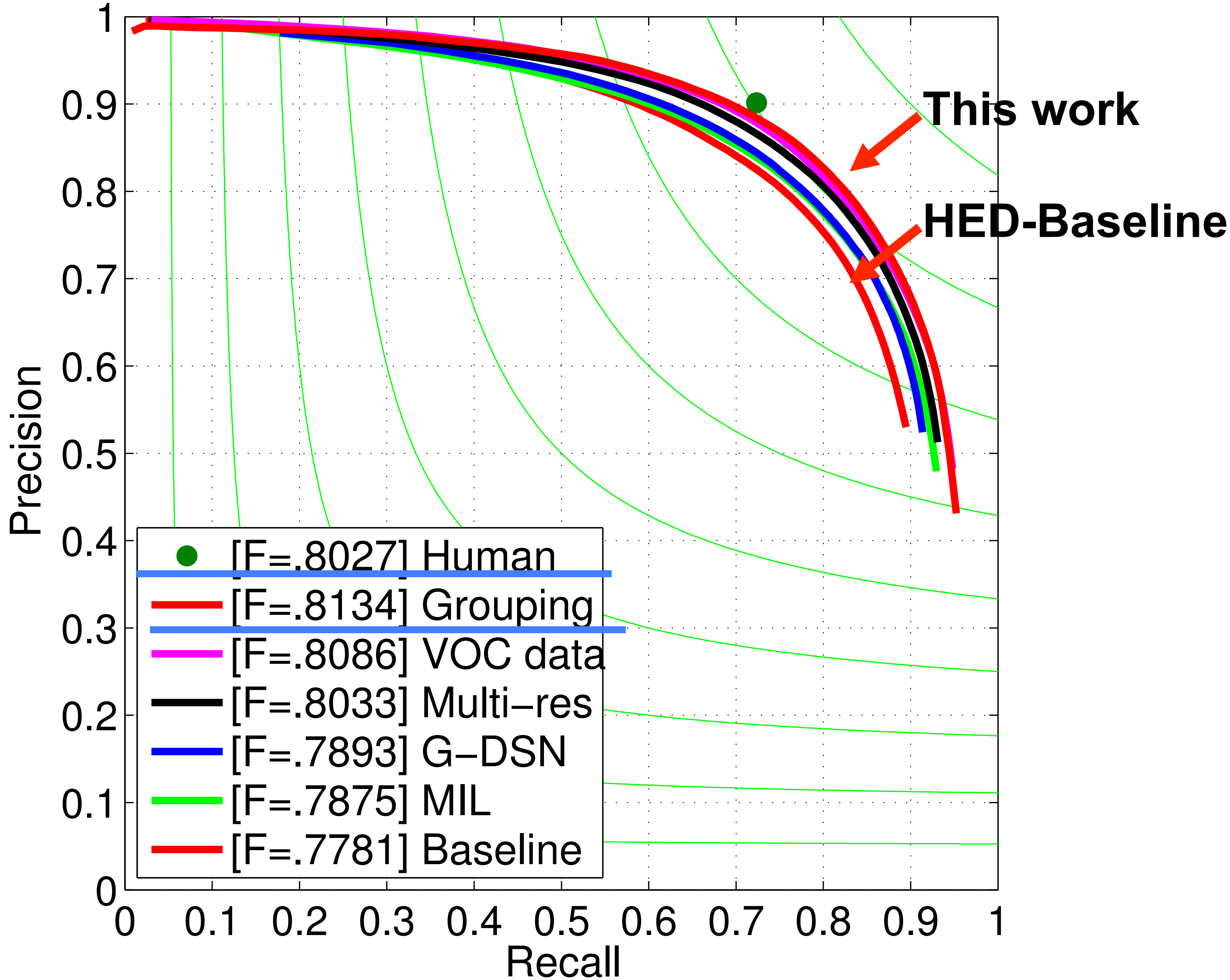


Image Pyramid Tied CNN outputs Scale fusion NCuts & boundaries Final outputs

All-in-one caffe network, ~1 second per frame

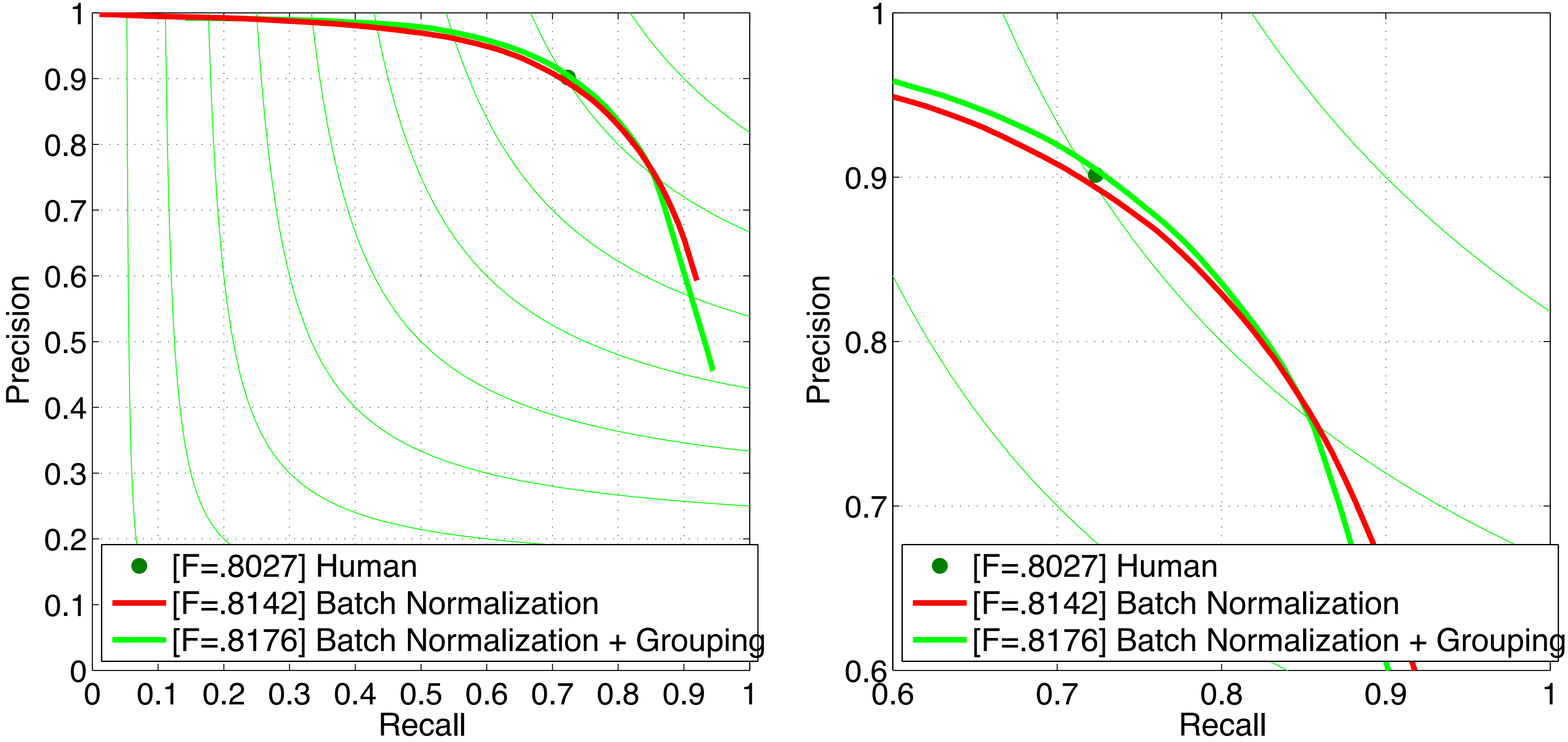
Progress in edge detection



I. Kokkinos, Pushing the boundaries of boundary detection using deep learning, ICLR 2016

One last trick!

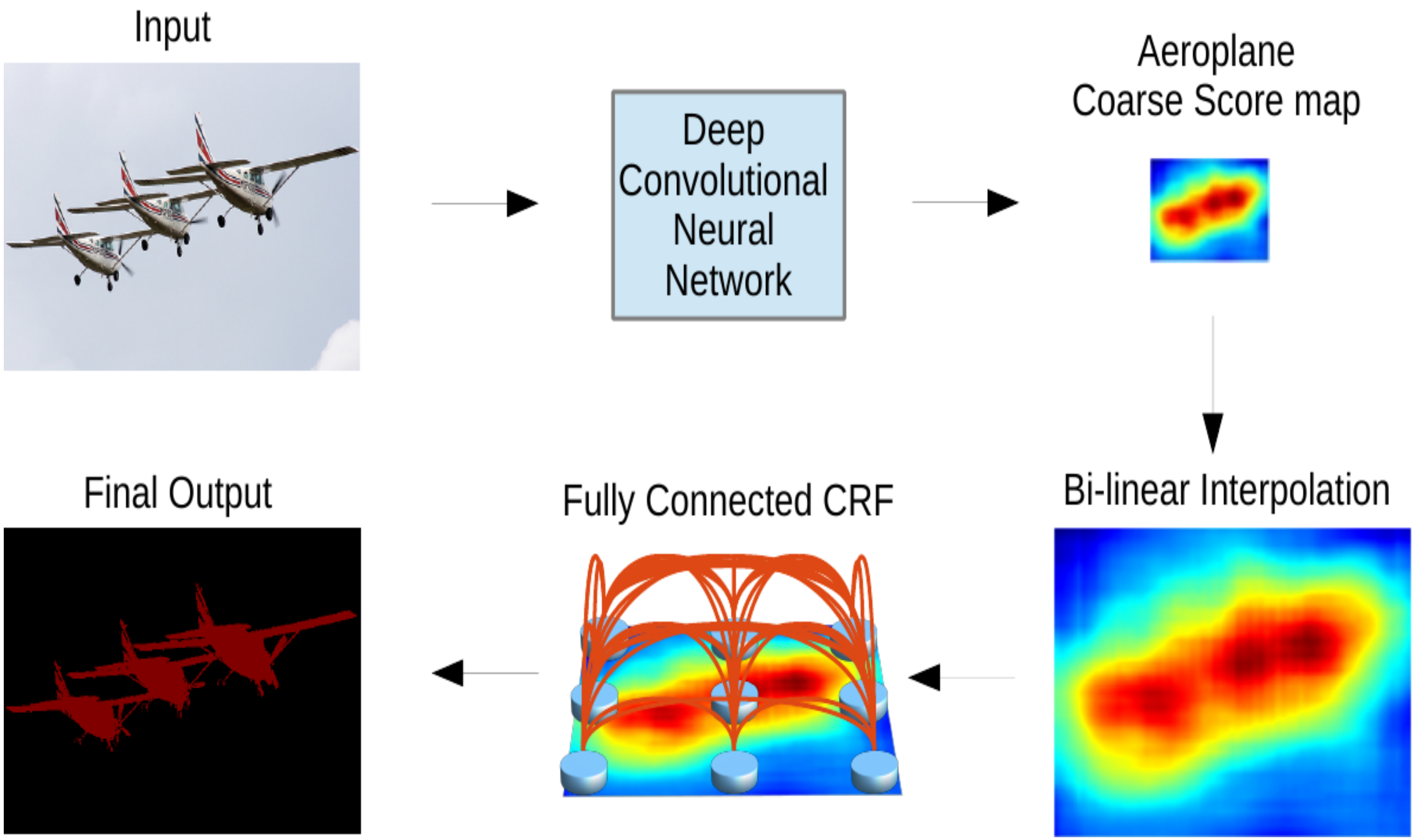
Batch normalization: stable & faster training



Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, S. Ioffe, C. Szegedy

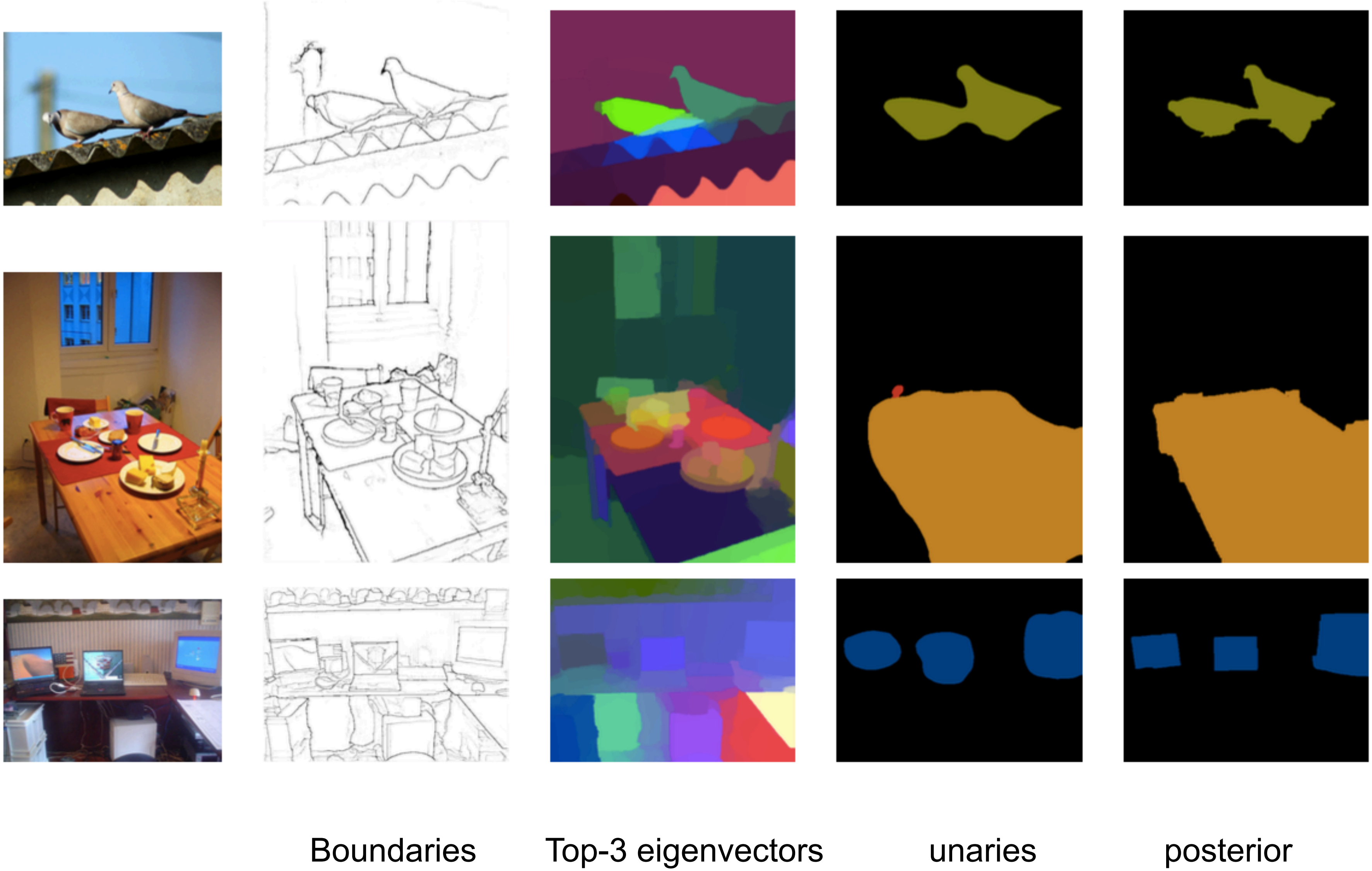
I. Kokkinos, Pushing the boundaries of boundary detection using deep learning, ICLR 2016

2015: Deeplab: FCNNs + DenseCRF

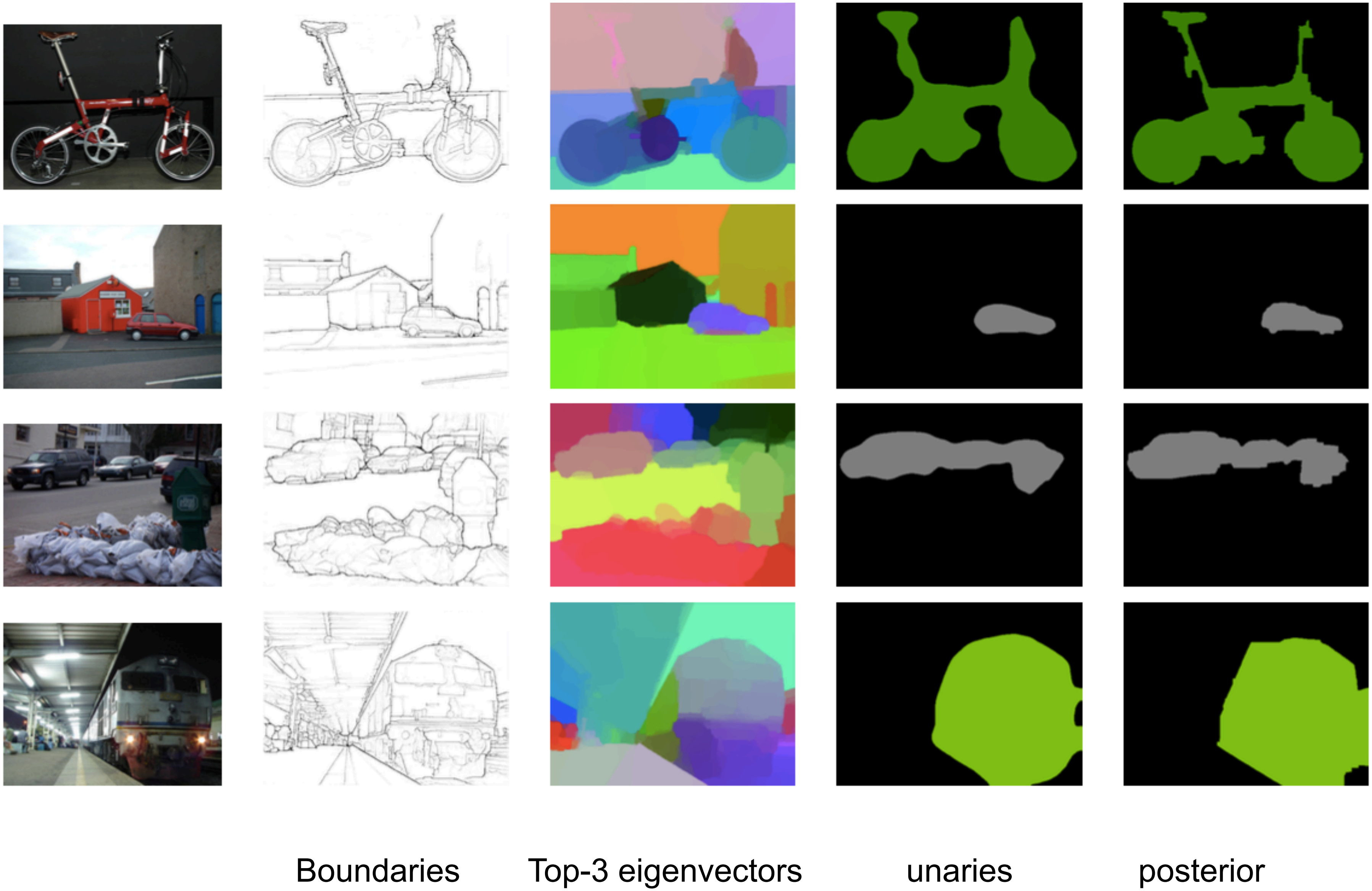


L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille, Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, ICLR 2015

2016: Combine with spectral embedding



2016: Combine with spectral embedding

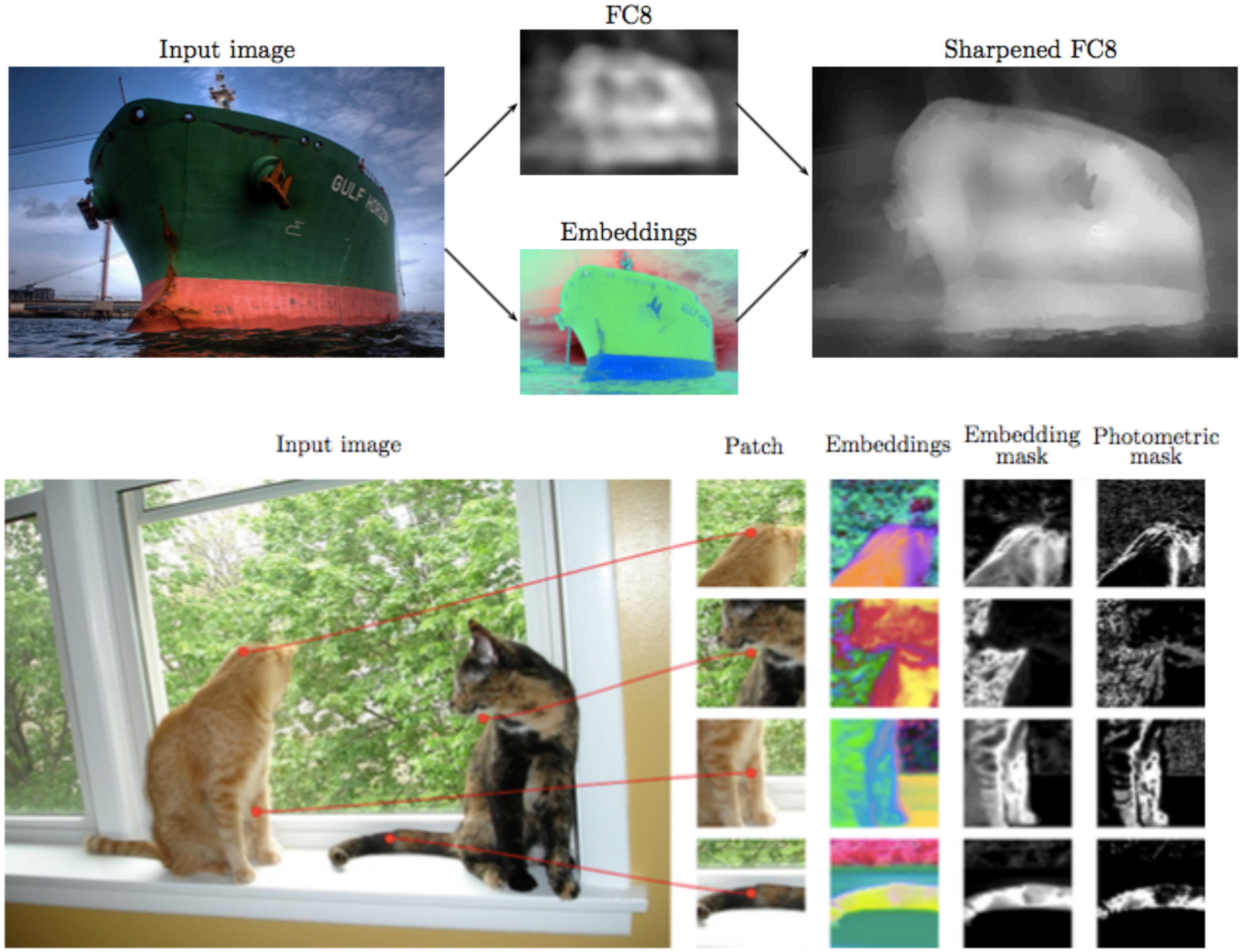


Spectral embedding + DenseCRF

Method	mAP %
Adelaide-Context-CNN-CRF-COCO (Lin et al., 2015)	77.8
CUHK-DPN-COCO (Liu et al., 2015)	77.5
Adelaide-Context-CNN-CRF-COCO (Lin et al., 2015)	77.2
MSRA-BoxSup (Dai et al., 2015)	75.2
Oxford-TVG-CRF-RNN-COCO (Zheng et al., 2015)	74.7
DeepLab-MSc-CRF-LF-COCO-CJ (Chen et al., 2015)	73.9
DeepLab-CRF-COCO-LF(Chen et al., 2015)	72.7
Multi-Scale DeepLab	72.1
Multi-Scale DeepLab-CRF	74.8
Multi-Scale DeepLab-CRF-Embeddings	75.4
Multi-Scale DeepLab-CRF-Embeddings-GraphCuts	75.7

I. Kokkinos, Pushing the boundaries of boundary detection using deep learning, ICLR 2016

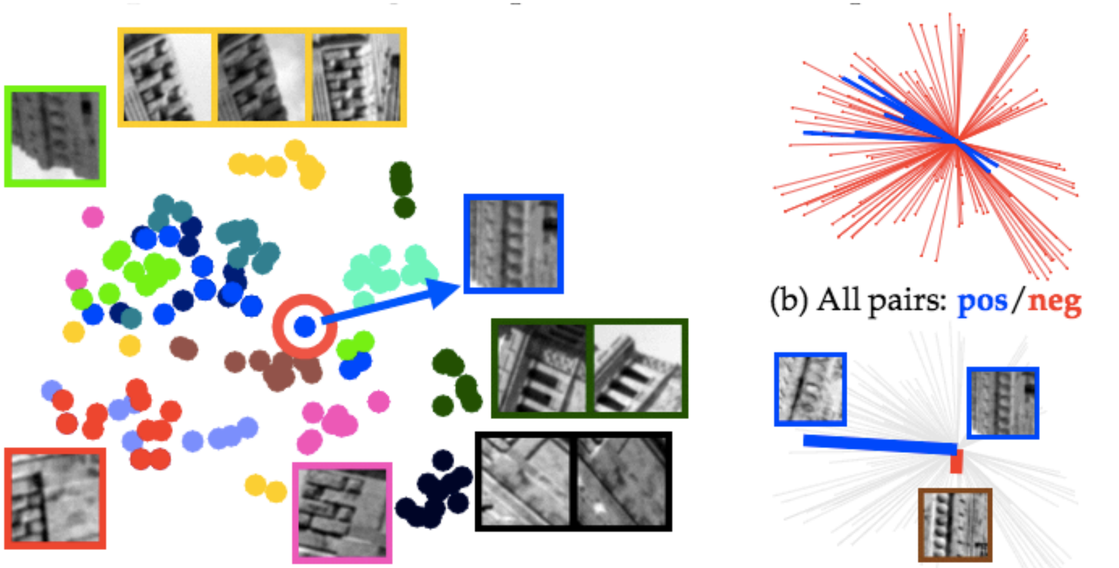
Bottom-up alternative: metric learning



A. Harley, I. Kokkinos, and K. Derpanis, Learning Dense Convolutional Embeddings for Semantic Segmetnation, ICLR workshops 2016

This talk: controlling DCNNs for low- and high- level tasks

- -Classification & Detection
- -Semantic Segmentation
- -Boundary Detection
- -Feature Descriptors



(a) 12 points/132 patches with t-SNE [8]

(b) All pairs: pos/neg

(c) "Hard" pairs: pos/neg

E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, F. Moreno-Noguer, Discriminative Learning of Deep Convolutional Descriptors, ICCV15

Advertisement #1

ICCV 2015

Discriminative learning of Deep Convolutional Feature Point Descriptors

Edgar Simo-Serra, Eduard Trulls, Luis Ferraz,
Iasonas Kokkinos, Pascal Fua, Francesc Moreno-Noguer

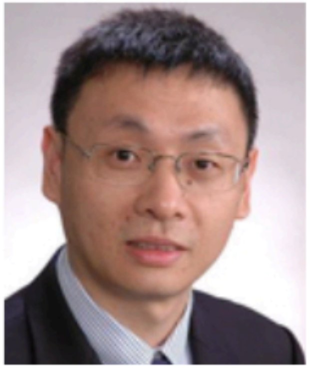
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Institut de Robòtica i Informàtica Industrial
WASEDA University
CSIC Consejo Superior de Investigaciones Científicas
EPFL École Polytechnique Fédérale de Lausanne
catchoom
Centrale Paris

https://github.com/cvlab-epfl/deepdesc-release

Advertisement #2

CVIU Special Issue on Deep Learning for CV

Submission deadline: April 16, 2016

Conclusion

2012 onwards: all about DCNNs

if [all] you have [is] a hammer, you treat everything like a nail

- -Classification & Detection
- -Semantic Segmentation
- -Boundary Detection
- -Feature Descriptors

2014 onwards: incorporating structure in DCNNs

trust is good, but control is better!

even better are results!

Thanks!

CentraleSupélec

informatics mathematics

SEVENTH FRAMEWORK PROGRAMME

NVIDIA