Deep Learning for Semantic Segmentation

Shenlong Wang

CSC2523 Course Presentation

Jan 19

Outline

- Introduction
- Previous work
- Conditional Random Field
- DeepLab
- Discussion

Segmentation

- Goal: Partitioning an image into multiple groups

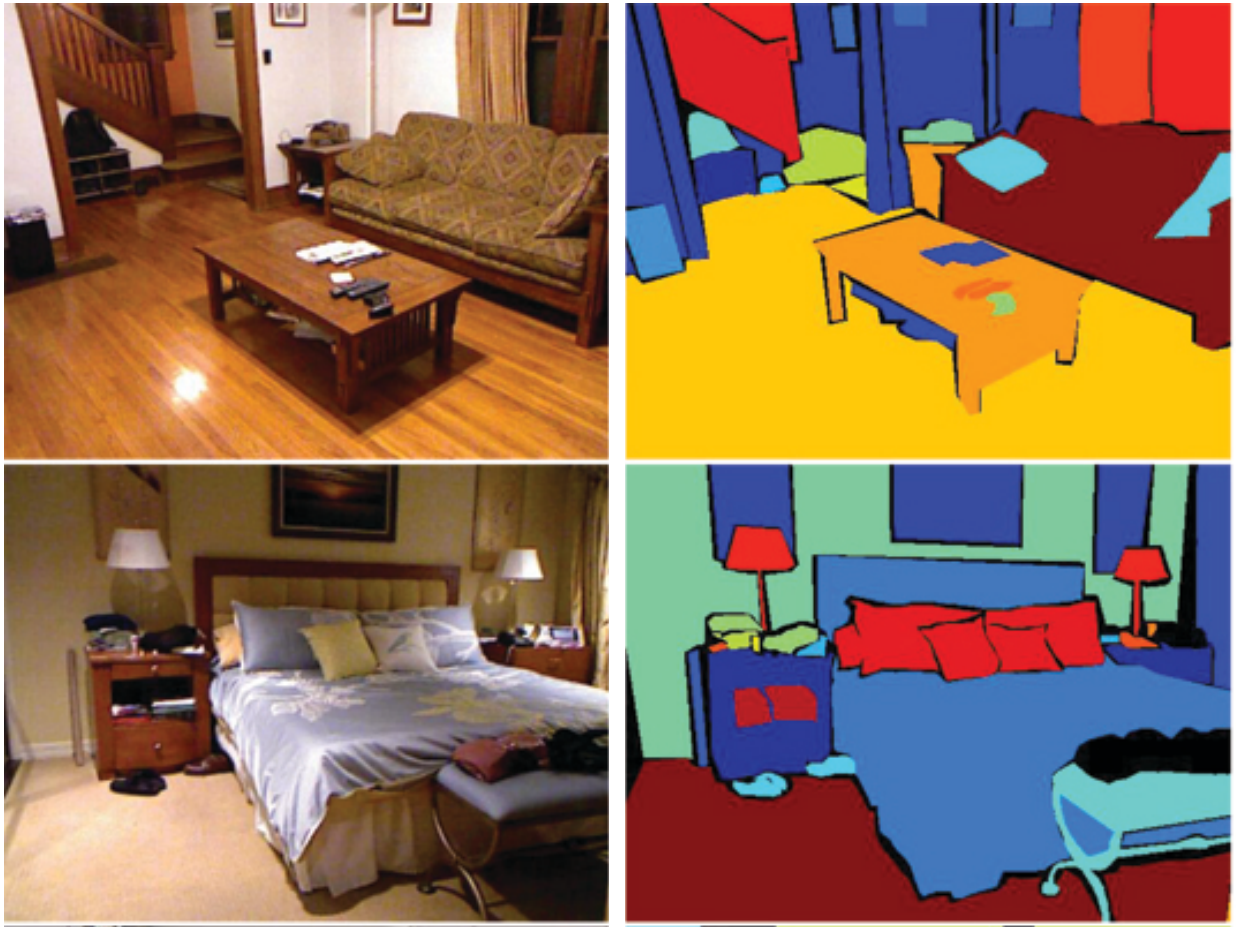


Image credit: Silberman et al.

Foreground Segmentation

- Goal: Extract foreground from the image

Image credit: Rother et al

Unsupervised Segmentation

- Goal: Grouping pixels based similarity

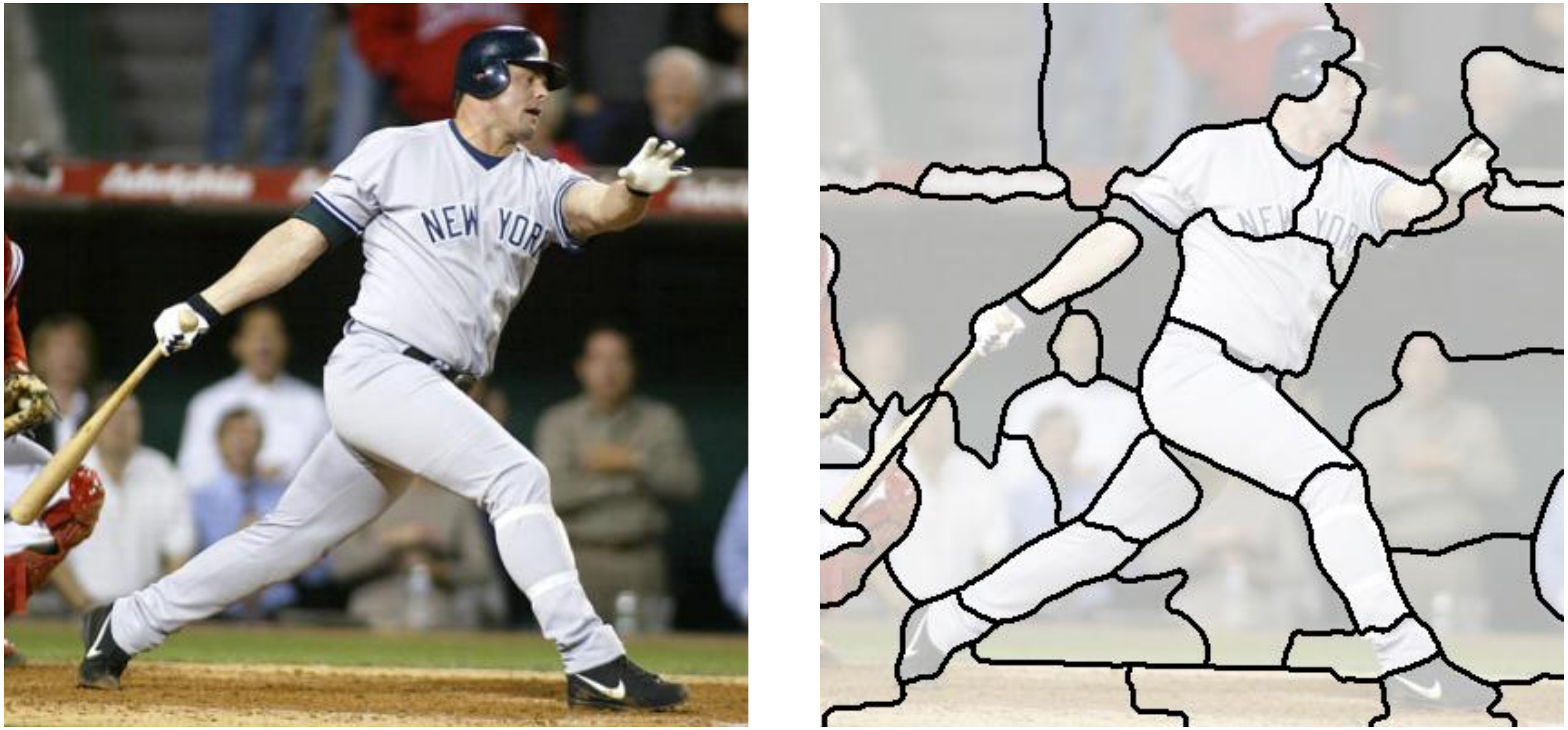


Image credit: Shi et al.

Cosegmentation

- Goal: Segmenting common objects from multiple images

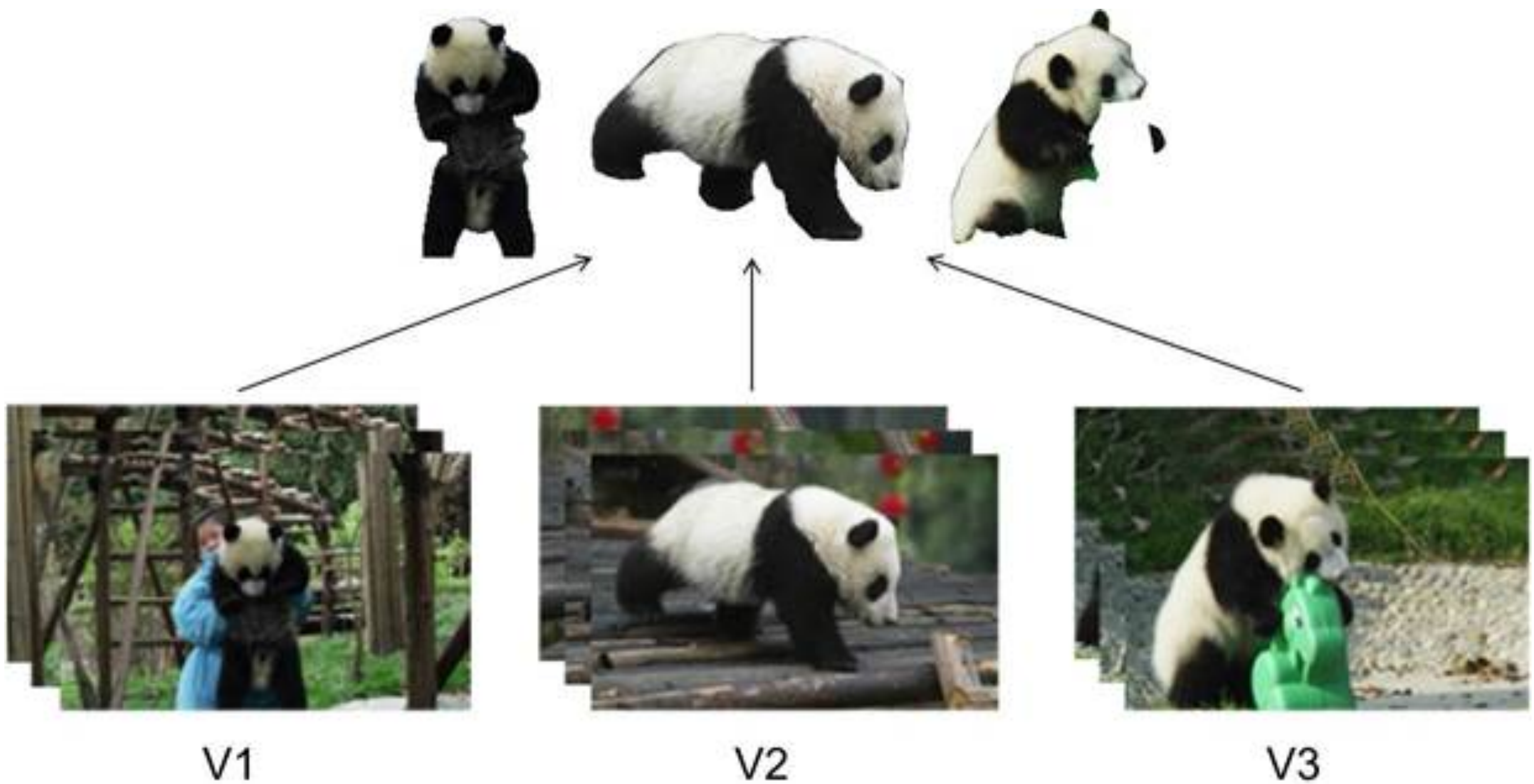


Image credit: Guo et al.

Instance Segmentation

- Goal: Assign each pixel an object instance

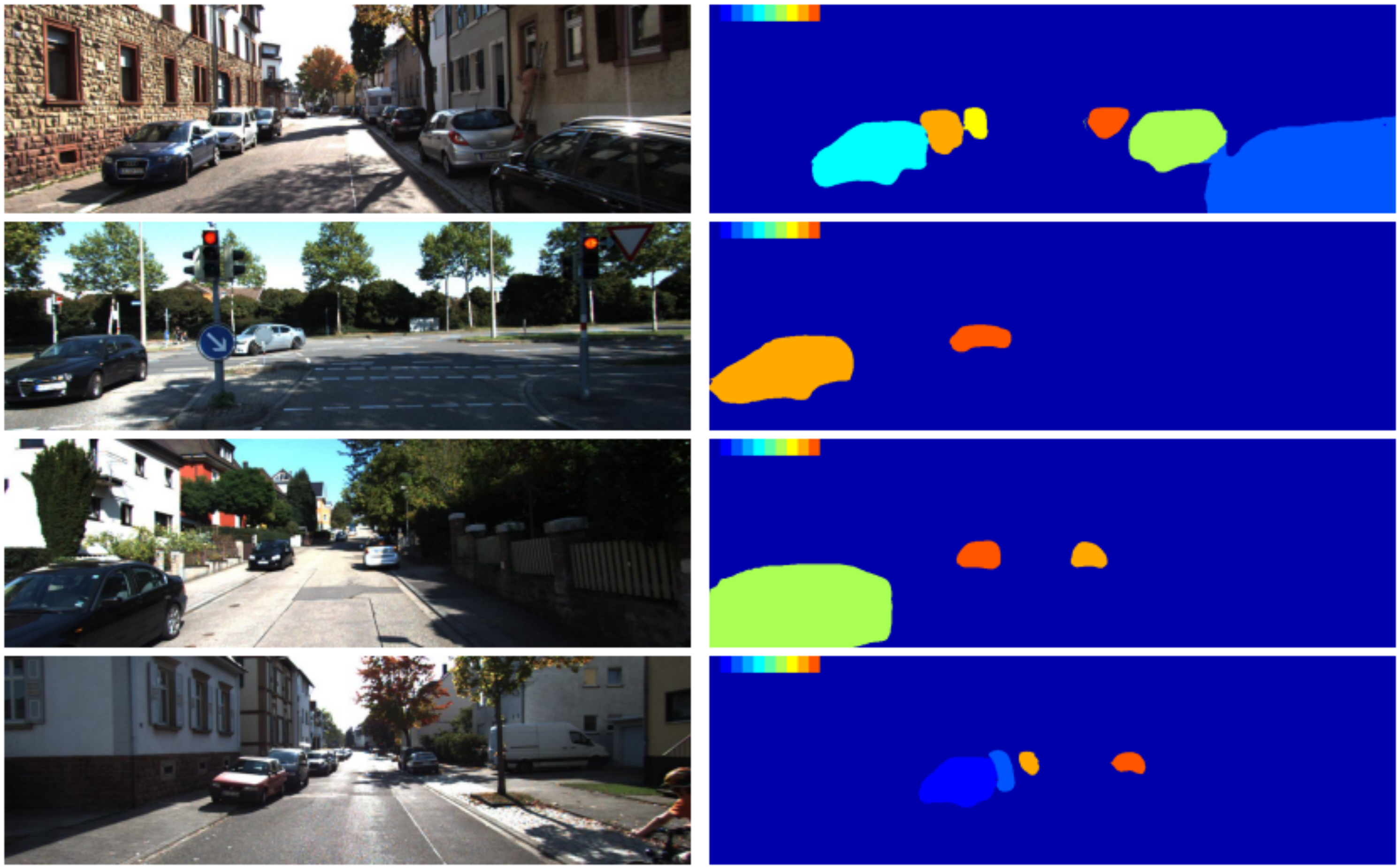


Image credit: Zhang et al.

Semantic Segmentation

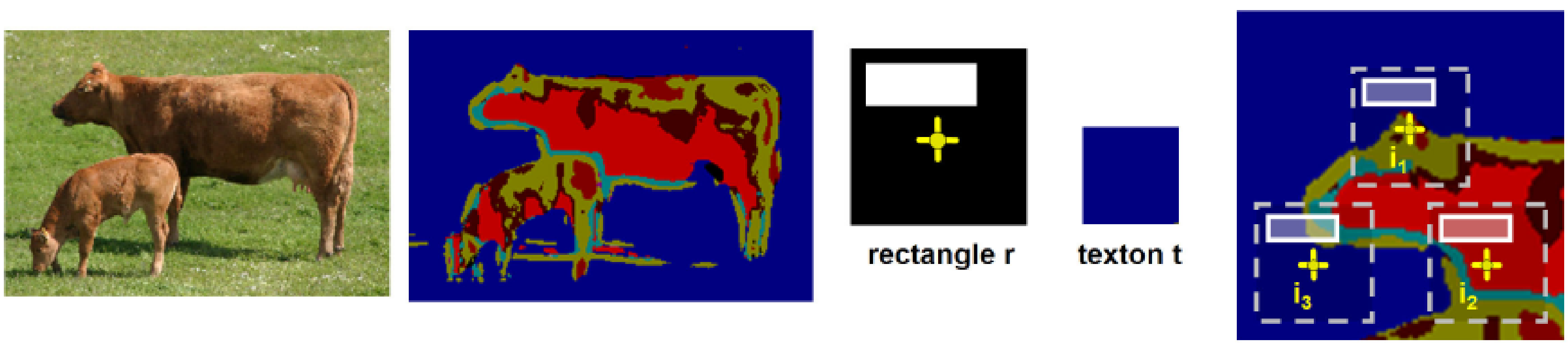
- Goal: Assign a class label to each pixel in the image


Why semantic segmentation?

Image credit: PASCAL VOC

TextonBoost

- Texton, Location, Color Features
  - Texton: Clusters of filter-bank responses
- Joint Boosting
  - Different classes share features
  - Weak classifier is based counting features



(a) Input image (b) Texton map (c) Feature pair = (r,t) (d) Superimposed rectangles

Image credit: Rother et al.

Decision Forest

- Texton, Location, Color Features
  - Encoded in a hierarchical way through decision forest
- Decision Forest
  - Pass from root to leaf through decisions
  - Each leaf node maintains a class distribution
  - Weak classifier is very simple

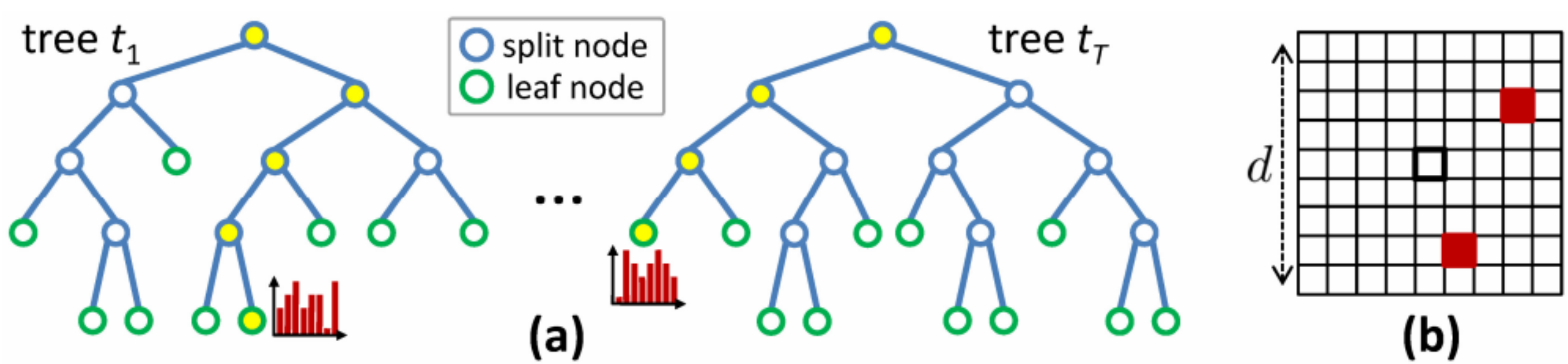


Image credit: Shotton et al.

Labeling Transfer

- Find pixel-wise correspondence (SiftFlow)
- Transfer labels

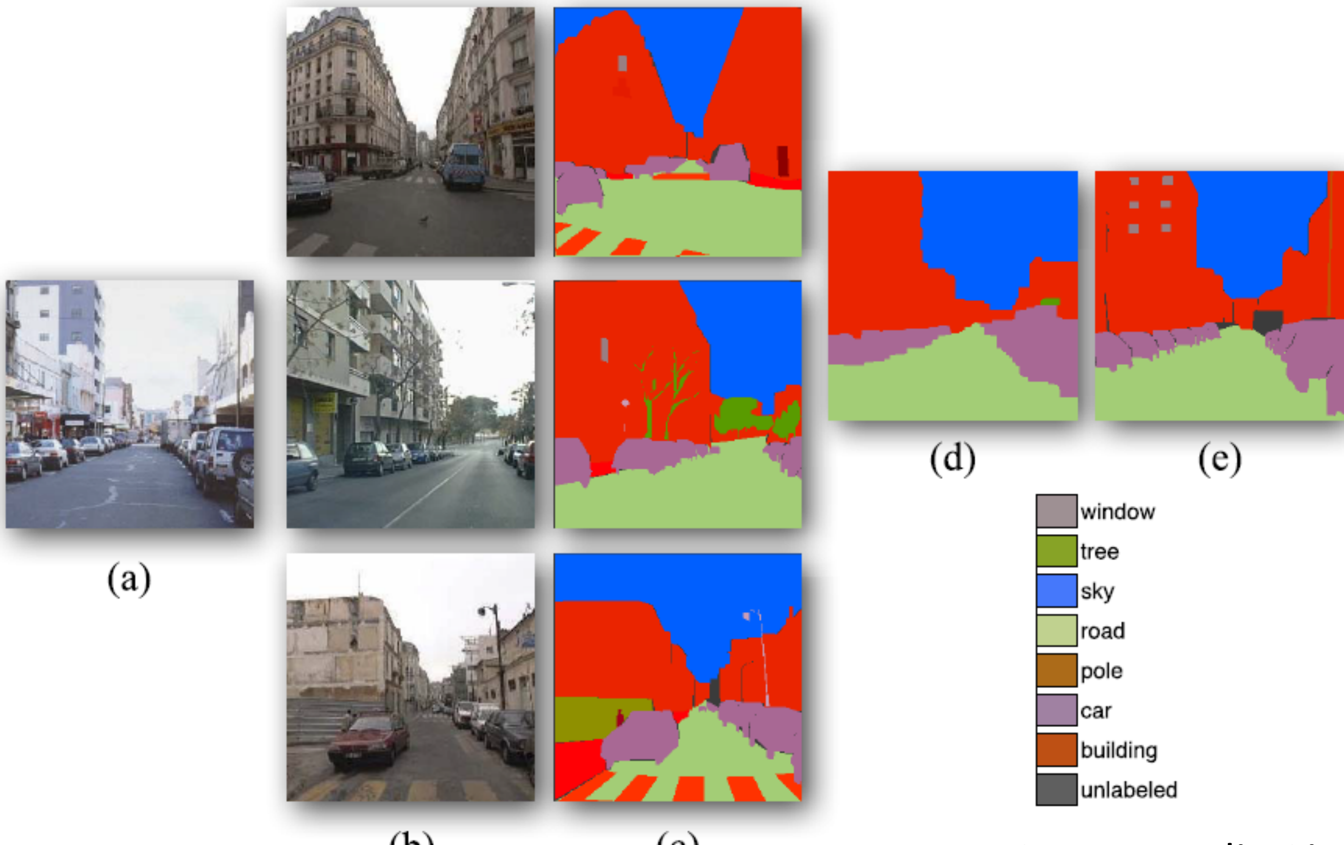
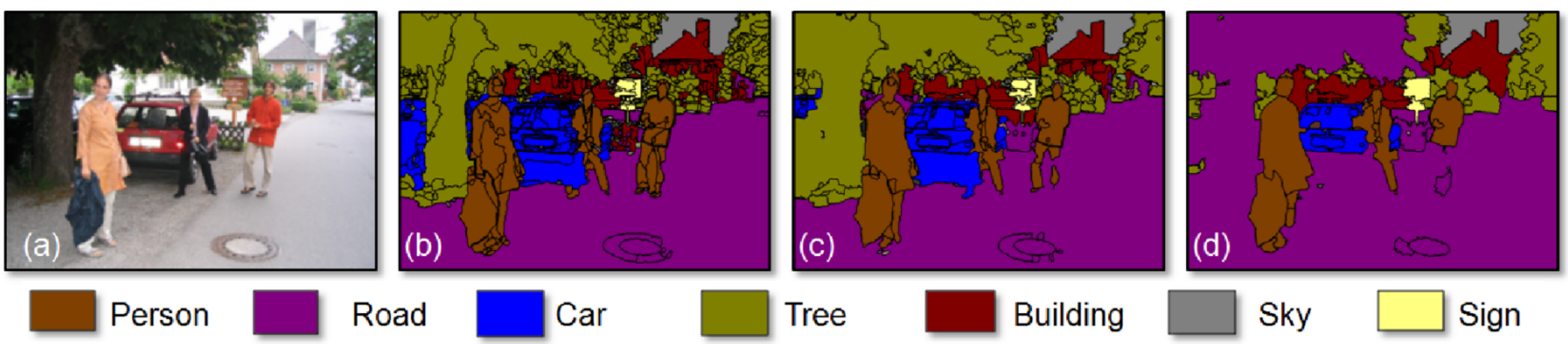


Image credit: Liu et al.

Super-pixel Methods

- Do over-segmentation
- Extract local descriptors for each superpixel
- Consider each superpixel as a sample to classify
- Used for scene labeling



Pro & Con?

Image credit: Ladicky et al.

Region Methods

- Sample object region proposals
- Extract local descriptors for each proposal
- Consider each region as a sample to classify
- Usually used for object segmentation/detection

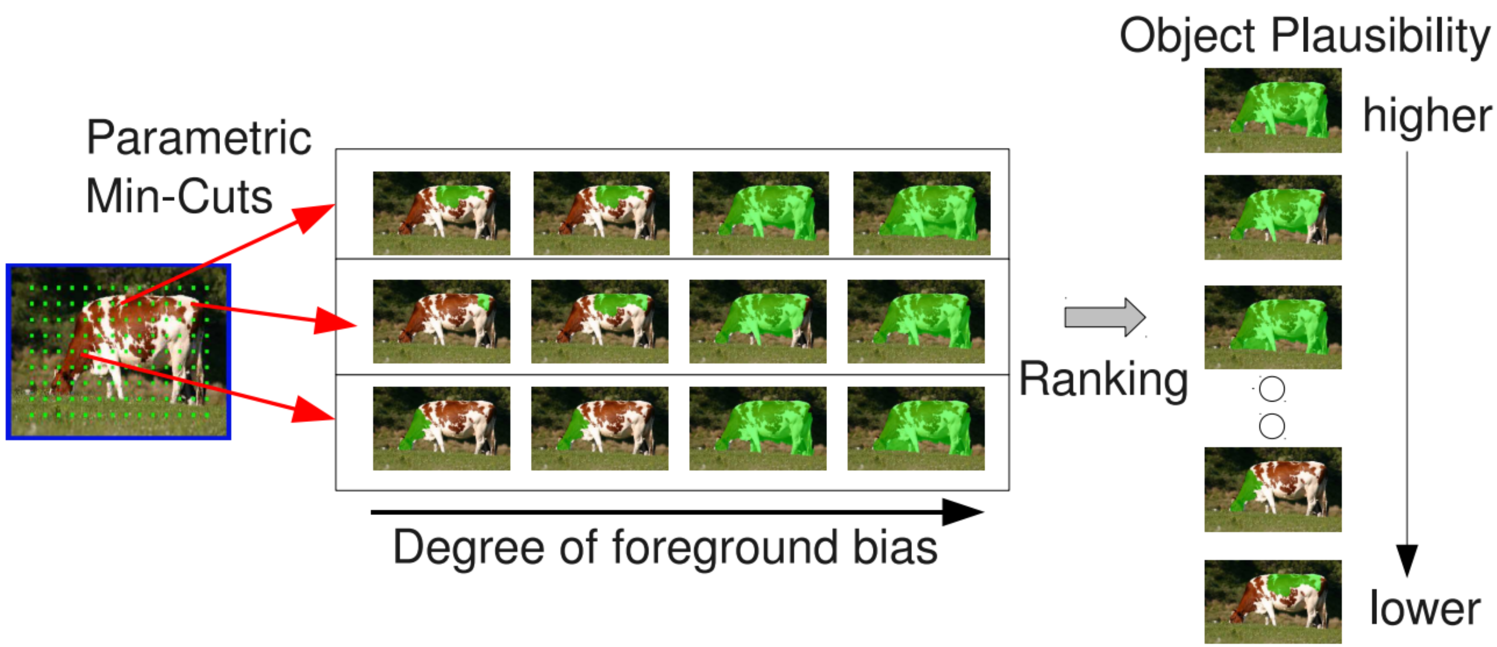


Image credit: Carreira et al.

Random Field Model

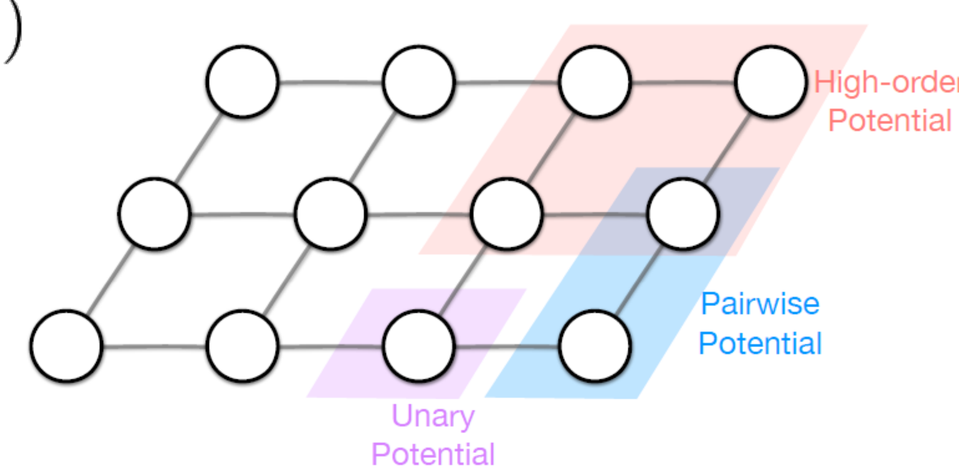
$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in N} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) + \sum_\alpha \phi_\alpha(\mathbf{x}_\alpha)$$

Unary term
Pairwise term
High-order term

- Probabilistic interpretation:

$$P(\mathbf{x}) = \frac{1}{Z} \exp(-E(\mathbf{x}))$$

- Map Inference
  - Mostly likely labeling
  - Has lowest energy
  - In general, NP-Hard



Random Field Model

- All the methods I cover today use random fields

Why?

CRF for Semantic Segmentation

- A discriminative MRF
  - Node: usually pixels
  - Edge: interactions between pixels
- Unary term
  - Local classifiers
- Pairwise term
  - Neighboring pixels tend to have similar labels
  - Usually weighted by color similarity

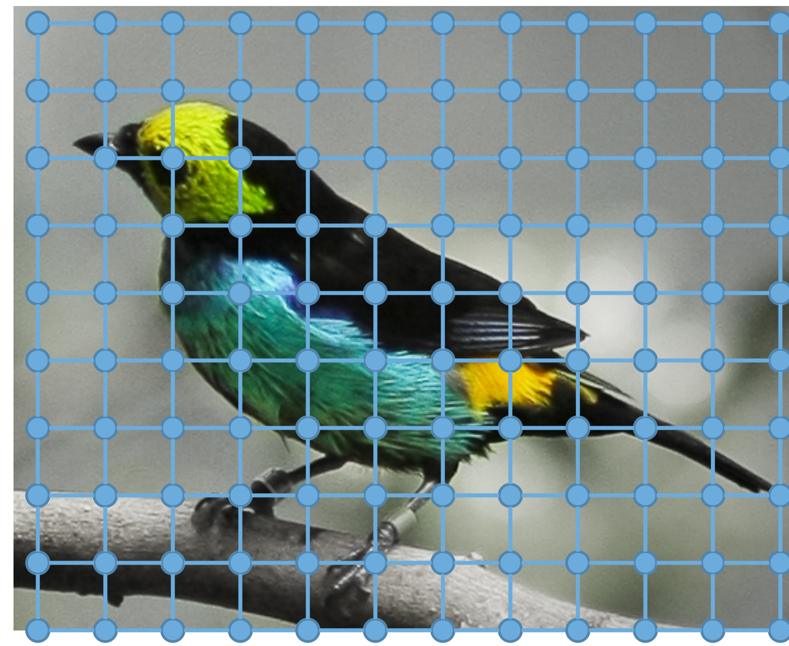


Image credit: Shi et al.

CRF for Semantic Segmentation

- A discriminative MRF
  - Node: usually pixels
  - Edge: interactions between pixels
- Unary term
  - Local classifiers
- Pairwise term
  - Neighboring pixels tend to have similar labels
  - Usually weighted by color similarity

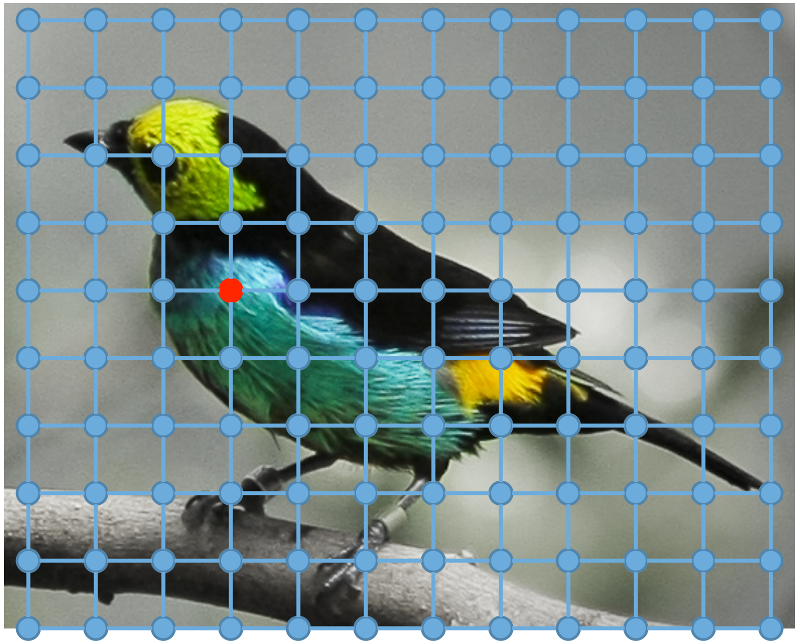


Image credit: Shi et al.

CRF for Semantic Segmentation

- A discriminative MRF
  - Node: usually pixels
  - Edge: interactions between pixels
- Unary term
  - Local classifiers
- Pairwise term
  - Neighboring pixels tend to have similar labels
  - Usually weighted by color/texture/brightness similarity

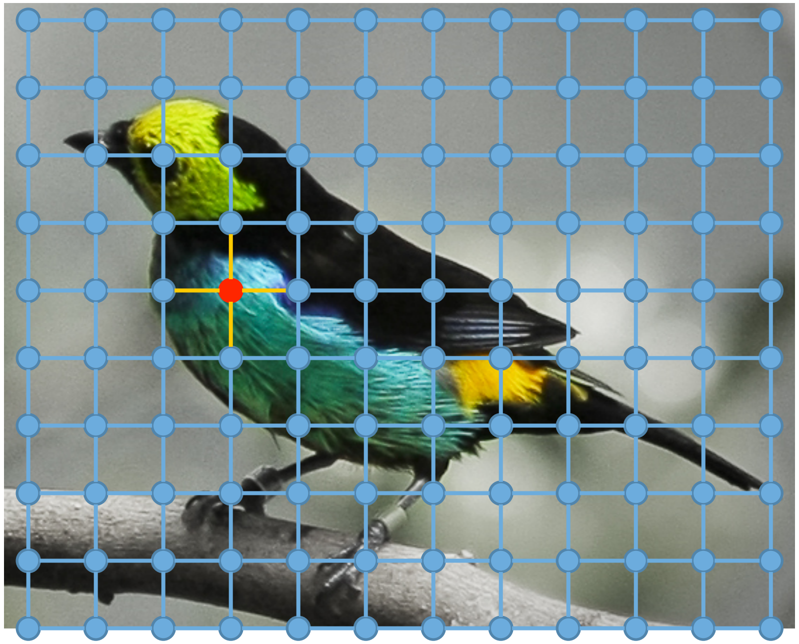


Image credit: Shi et al.

CRF for Semantic Segmentation

- Location interactions

$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in N} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

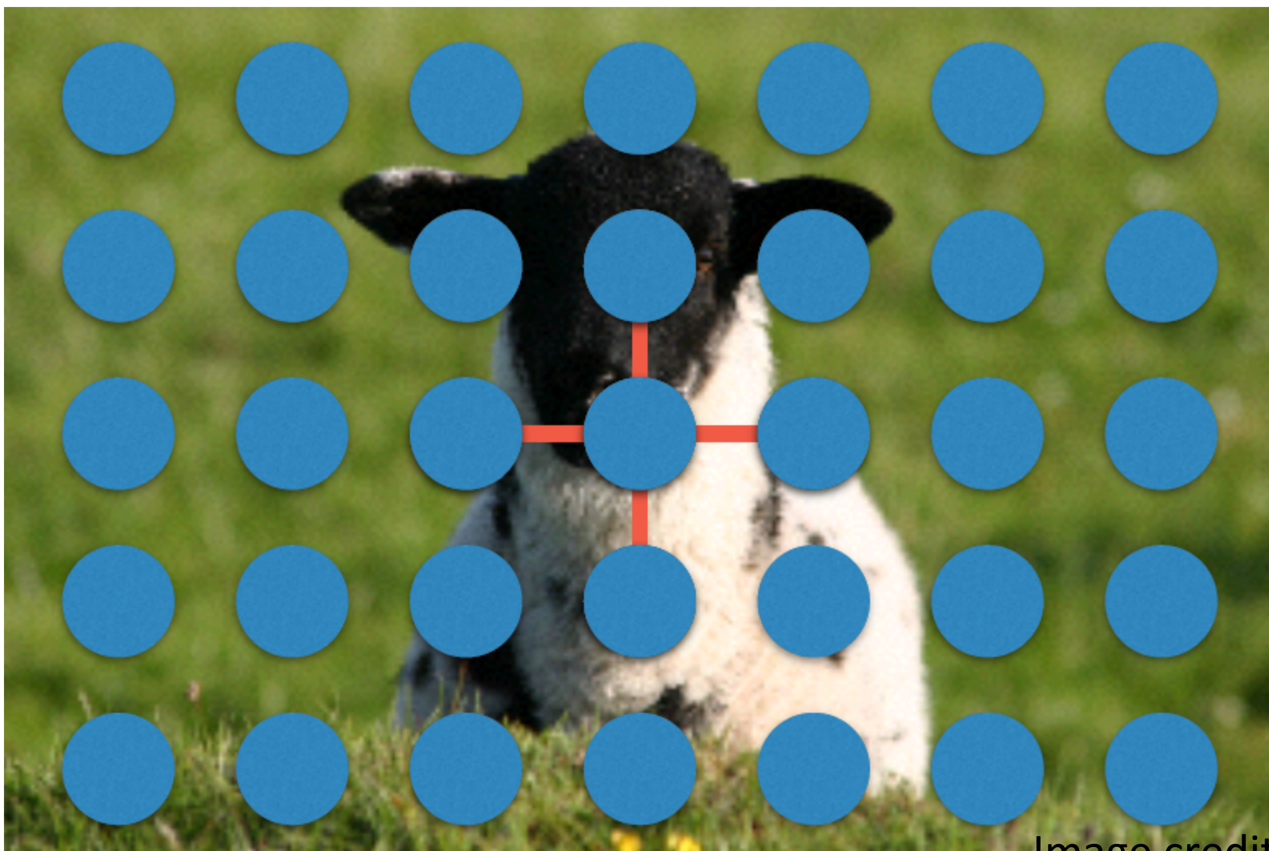


Image credit: Krähenbühl et al.

Cons of adjacent connectivity

- Shrinking bias and limited propagation

$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in N} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

Unary

Image credit: Krähenbühl et al.

Cons of adjacent connectivity

- Shrinking bias and limited propagation

$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in N} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

CRF

Image credit: Krähenbühl et al.

Cons of adjacent connectivity

- Shrinking bias and limited propagation

$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in N} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

CRF

Krähenbühl et al.

Densely-connected CRF

- Pixels are densely connected

$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \cancel{\in \mathcal{N}}} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

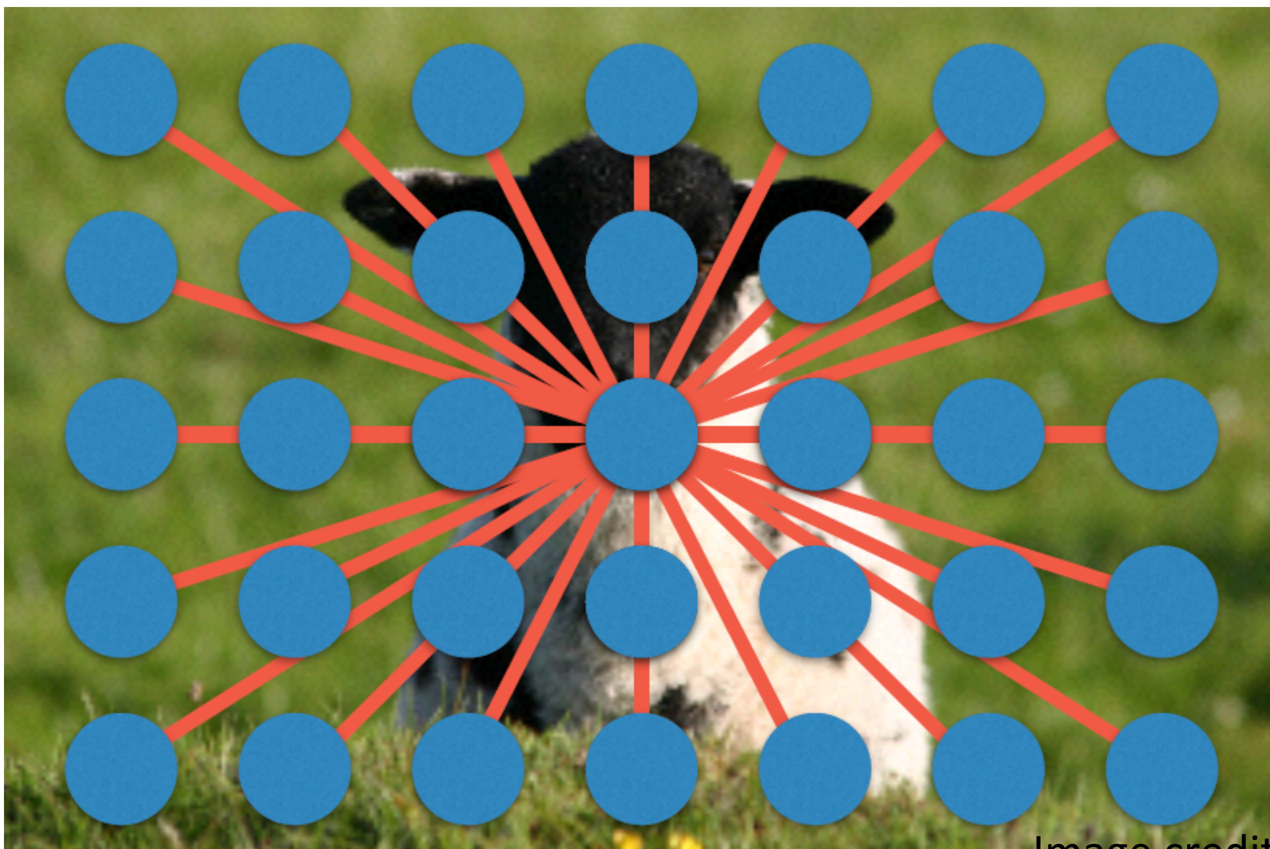


Image credit: Krähenbühl et al.

Densely-connected CRF

- Bilateral Pairwise Term

$$\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \mu(\mathbf{x}_i, \mathbf{x}_j)\exp\left(-\frac{|s_i - s_j|^2}{2\sigma_\alpha^2} - \frac{|c_i - c_j|^2}{2\sigma_\beta^2}\right)$$

Label compatibility

Location smoothness

Appearance smoothness

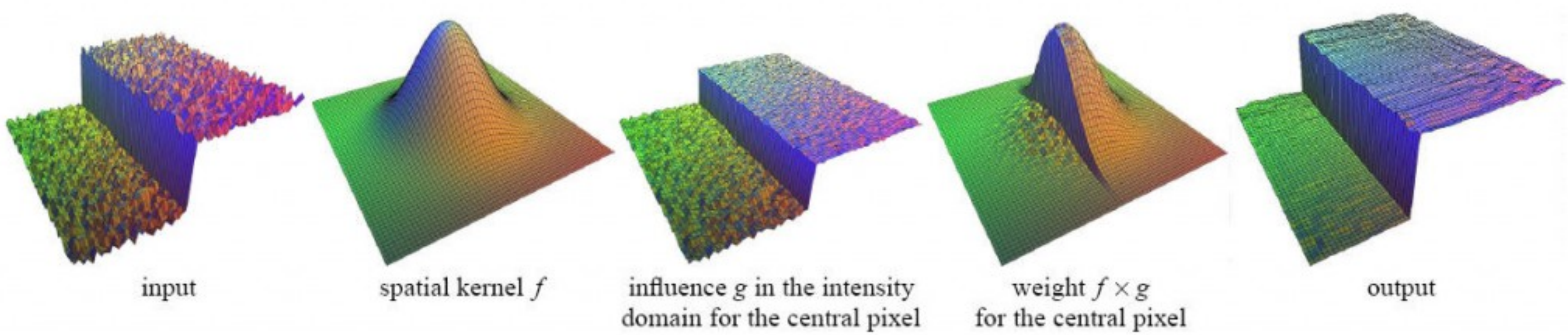


Image credit: Krähenbühl et al.

Outline

- Introduction
- Previous work
- Conditional Random Field
- DeepLab
- Discussion

Densely-connected CRF

$$E(\mathbf{x}) = \sum_{i} \phi_i(\mathbf{x}_i) + \sum_{i,j \cancel{\in \mathcal{N}}} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

- Pros
  - Long range interactions, overcome shrinking bias
- Cons
  - Billions of pairwise terms sound scary for inference

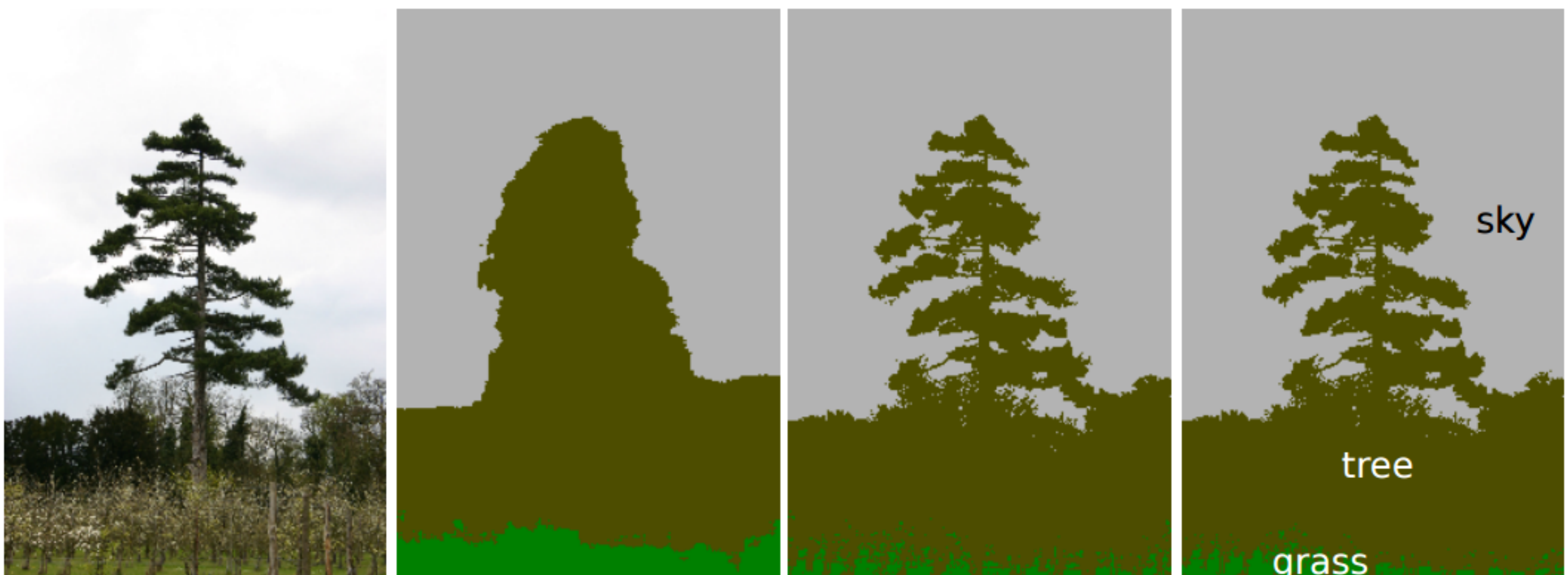
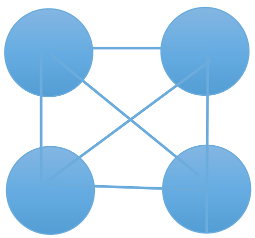


Image credit: Krähenbühl et al.

Warning

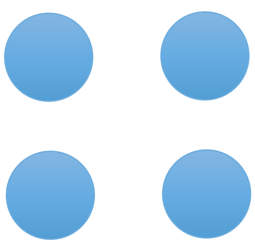
A huge wave of Math is approaching!

Mean-field Approximate Inference



$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in \cancel{\mathcal{N}}} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

$$\mathbf{x} = \arg\max_{\mathbf{x}} P(\mathbf{x}) \quad \text{where} \quad P(\mathbf{x}) = \frac{1}{Z} \exp(-E(\mathbf{x}))$$



$$\mathbf{x} = \arg\max_{\mathbf{x}} Q(\mathbf{x})$$

$$Q(\mathbf{x}) = \prod_i Q_i(\mathbf{x}_i) \;\; \text{close to} \;\; P(\mathbf{x}) \quad \text{in terms of } D(Q||P)$$

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

Why we want such a factorized Q?

Mean-field Approximate Inference

Algorithm 1 Mean field in fully connected CRFs

Initialize Q

while not converged **do**

O(N^2) $\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l)$ for all m Message Passing

O(N) $\hat{Q}_i(x_i) \leftarrow \sum_{l \in \mathcal{L}} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l)$ Compatibility Transform

O(N) $Q_i(x_i) \leftarrow \exp\{-\psi_u(x_i) - \hat{Q}_i(x_i)\}$ Update

normalize $Q_i(x_i)$

end while

Mean-field Approximate Inference

$$\mathbf{x} = \arg\max_{\mathbf{x}} Q(\mathbf{x})$$

$$Q(\mathbf{x}) = \prod_i Q_i(\mathbf{x}_i) \quad \text{close to} \quad P(\mathbf{x}) \quad \text{in terms of } D(Q||P)$$

Algorithm 1 Mean field in fully connected

Initialize Q

while not converged **do**

O(N) $\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} \underline{k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)}Q_j(l)$ for all m — Efficient Convolution

O(N) $\hat{Q}_i(x_i) \leftarrow \sum_{l \in \mathcal{L}} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l)$ — Compatibility Transform

O(N) $Q_i(x_i) \leftarrow \exp\{-\psi_u(x_i) - \hat{Q}_i(x_i)\}$ — Update

normalize $Q_i(x_i)$

end while

permutohedral lattice

Image credit: Wikipedia

Since Nov. 2014...

State-of-the-art: mAP 77.8

	mean	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor	submission date
Adelaide_Context_CNN_CRF_COCO [?]	77.8	92.9	39.6	84.0	67.9	75.3	92.7	83.8	90.1	44.3	85.5	64.9	87.3	88.8	84.5	85.5	68.1	89.0	62.8	81.2	71.4	06-Nov-2015
CUHK_DPN_COCO [?]	77.5	89.0	61.6	87.7	66.8	74.7	91.2	84.3	87.6	36.5	86.3	66.1	84.4	87.8	85.6	85.4	63.6	87.3	61.3	79.4	66.4	22-Sep-2015
Adelaide_Context_CNN_CRF_COCO [?]	77.2	92.3	38.8	82.9	66.1	75.1	92.4	83.1	88.6	41.8	85.9	62.8	86.7	88.4	84.0	85.4	67.4	88.8	61.9	81.9	71.7	13-Aug-2015
CentraleSuperBoundaries++ [?]	76.0	91.1	38.5	90.9	68.7	74.2	89.9	85.3	89.1	34.4	82.5	65.6	83.1	82.9	85.7	85.4	60.6	84.5	59.9	80.2	69.9	13-Jan-2016
Oxford_TVG_HO_CRF [?]	75.9	91.2	56.2	88.9	68.0	70.7	89.5	83.8	87.2	33.6	81.0	66.4	82.4	83.1	87.8	82.3	59.8	83.5	53.4	79.5	71.1	08-Jan-2016
CentraleSuperBoundaries [?]	75.7	90.3	37.9	89.6	67.8	74.6	89.3	84.1	89.1	35.8	83.6	66.2	82.9	81.7	85.6	84.6	60.3	84.8	60.7	78.3	68.3	01-Dec-2015
Adelaide_Context_CNN_CRF_VOC [?]	75.3	90.6	37.6	80.0	67.8	74.4	92.0	85.2	86.2	39.1	81.2	58.9	83.8	83.9	84.3	84.8	62.1	83.2	58.2	80.8	72.3	30-Aug-2015
MSRA_BoxSup [?]	75.2	89.8	38.0	89.2	68.9	68.0	89.6	83.0	87.7	34.4	83.6	67.1	81.5	83.7	85.2	83.5	58.6	84.9	55.8	81.2	70.7	18-May-2015
POSTECH_DeconvNet_CRF_VOC [?]	74.8	90.0	40.8	84.2	67.3	70.7	90.9	84.8	87.4	34.8	83.0	58.7	82.3	87.1	86.9	82.4	64.5	84.6	54.9	77.5	64.1	18-Aug-2015
Oxford_TVG_CRF_RNN_COCO [?]	74.7	90.4	55.3	88.7	68.4	69.8	88.3	82.4	85.1	32.6	78.5	64.4	79.6	81.9	86.4	81.8	58.6	82.4	53.5	77.4	70.1	22-Apr-2015
DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint [?]	73.9	89.2	46.7	88.5	63.5	68.4	87.0	81.2	86.3	32.6	80.7	62.4	81.0	81.3	84.3	82.1	56.2	84.6	58.3	76.2	67.2	26-Apr-2015
DeepLab-CRF-COCO-LargeFOV [?]	72.7	89.1	38.3	88.1	63.3	69.7	87.1	83.1	85.0	29.3	76.5	56.5	79.8	77.9	85.8	82.4	57.4	84.3	54.9	80.5	64.1	18-Mar-2015
POSTECH_EDeconvNet_CRF_VOC [?]	72.5	89.9	39.3	79.7	63.9	68.2	87.4	81.2	86.1	28.5	77.0	62.0	79.0	80.3	83.6	80.2	58.8	83.4	54.3	80.7	65.0	22-Apr-2015
Oxford_TVG_CRF_RNN_VOC [?]	72.0	87.5	39.0	79.7	64.2	68.3	87.6	80.8	84.4	30.4	78.2	60.4	80.5	77.8	83.1	80.6	59.5	82.8	47.8	78.3	67.1	22-Apr-2015
DeepLab-MSc-CRF-LargeFOV [?]	71.6	84.4	54.5	81.5	63.6	65.9	85.1	79.1	83.4	30.7	74.1	59.8	79.0	76.1	83.2	80.8	59.7	82.2	50.4	73.1	63.7	02-Apr-2015
MSRA_BoxSup [?]	71.0	86.4	35.5	79.7	65.2	65.2	84.3	78.5	83.7	30.5	76.2	62.6	79.3	76.1	82.1	81.3	57.0	78.2	55.0	72.5	68.1	10-Feb-2015
DeepLab-CRF-COCO-Strong [?]	70.4	85.3	36.2	84.8	61.2	67.5	84.6	81.4	81.0	30.8	73.8	53.8	77.5	76.5	82.3	81.6	56.3	78.9	52.3	76.6	63.3	11-Feb-2015
DeepLab-CRF-LargeFOV [?]	70.3	83.5	36.6	82.5	62.3	66.5	85.4	78.5	83.7	30.4	72.9	60.4	78.5	75.5	82.1	79.7	58.2	82.0	48.8	73.7	63.3	28-Mar-2015
TTI_zoomout_v2 [?]	69.6	85.6	37.3	83.2	62.5	66.0	85.1	80.7	84.9	27.2	73.2	57.5	78.1	79.2	81.1	77.1	53.6	74.0	49.2	71.7	63.3	30-Mar-2015
DeepLab-CRF-MSc [?]	67.1	80.4	36.8	77.4	55.2	66.4	81.5	77.5	78.9	27.1	68.2	52.7	74.3	69.6	79.4	79.0	56.9	78.8	45.2	72.7	59.3	30-Dec-2014
DeepLab-CRF [?]	66.4	78.4	33.1	78.2	55.6	65.3	81.3	75.5	78.6	25.3	69.2	52.7	75.2	69.0	79.1	77.6	54.7	78.3	45.1	73.3	56.2	23-Dec-2014
CRF_RNN [?]	65.2	80.9	34.0	72.9	52.6	62.5	79.8	76.3	79.9	23.6	67.7	51.8	74.8	69.9	76.9	76.9	49.0	74.7	42.7	72.1	59.6	10-Feb-2015
TTI_zoomout_16 [?]	64.4	81.9	35.1	78.2	57.4	56.5	80.5	74.0	79.8	22.4	69.6	53.7	74.0	76.0	76.6	68.8	44.3	70.2	40.2	68.9	55.3	24-Nov-2014
Hypercolumn [?]	62.6	68.7	33.5	69.8	51.3	70.2	81.1	71.9	74.9	23.9	60.6	46.9	72.1	68.3	74.5	72.9	52.6	64.4	45.4	64.9	57.4	09-Apr-2015
FCN-8s [?]	62.2	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	12-Nov-2014
MSRA_CFM [?]	61.8	75.7	26.7	69.5	48.8	65.6	81.0	69.2	73.3	30.0	68.7	51.5	69.1	68.1	71.7	67.5	50.4	66.5	44.4	58.9	53.5	17-Dec-2014
** SegNet ** [?]	59.9	73.6	37.6	62.0	46.8	58.6	79.1	70.1	65.4	23.6	60.4	45.6	61.8	63.5	75.3	74.9	42.6	63.7	42.5	67.8	52.7	10-Nov-2015
TTI_zoomout [?]	58.4	70.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SDS [?]	51.6	63.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
NUS_UDS [?]	50.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TTIC-divmbest-rerank [?]	48.1	62.7	25.6	46.9	43.0	54.8	58.4	58.6	55.6	14.6	47.5	31.2	44.7	51.0	60.9	53.5	36.6	50.9	30.1	50.2	46.8	15-Nov-2012

The best without CNN: mAP 48.1

DeepLab

- Outline

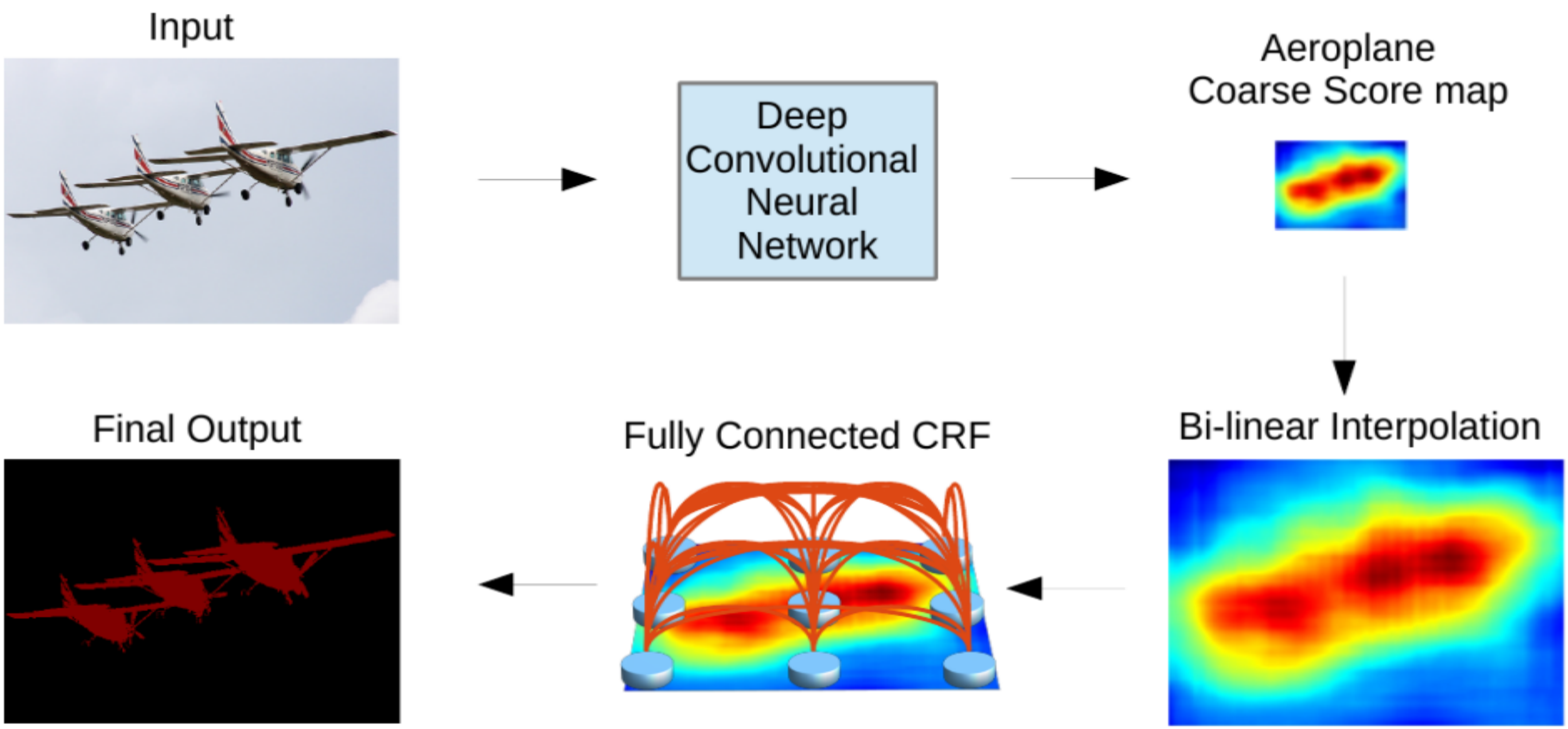


Image credit: Chen et al.

DeepLab

- Convert a detection architecture for segmentation
  - A variant of VGG-16
  - Initial weight from ImageNet for classification.

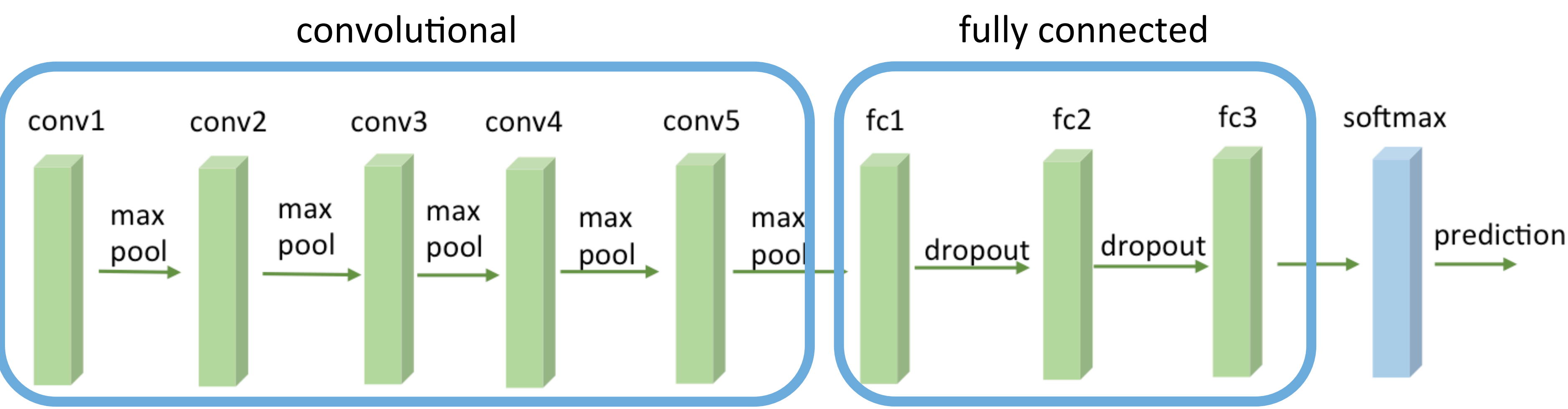


Image credit: Simonyan et al.

DeepLab

- Convert a detection architecture for segmentation
  - A variant of VGG-16
  - Initial weight from ImageNet for classification.

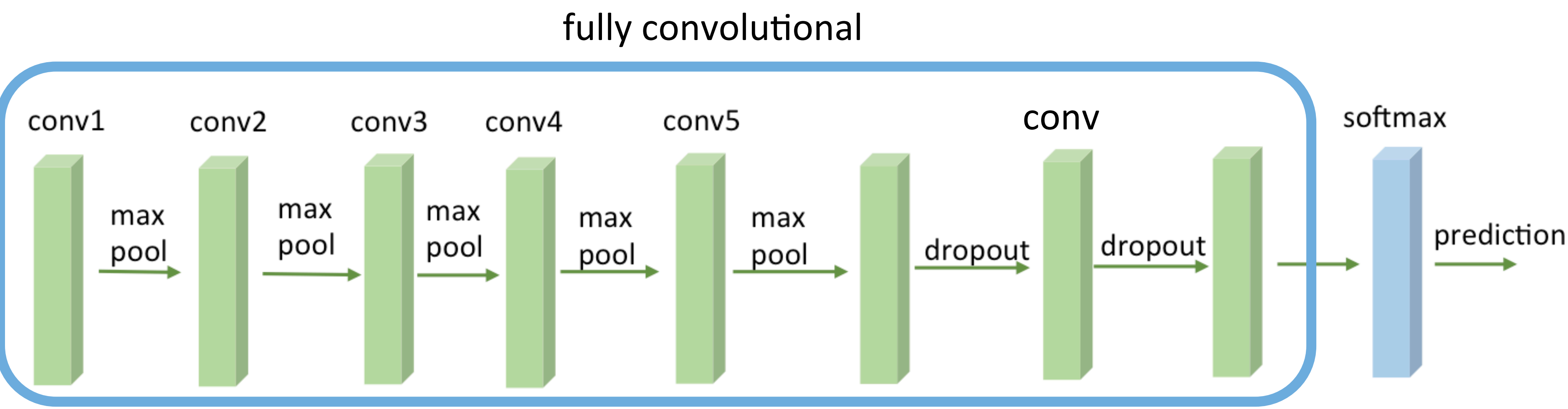


Image credit: Simonyan et al.

DeepLab

- Convert a detection architecture for segmentation
  - Resolution is a problem (x32)
  - You could either avoid some down-sampling (DeepLab)
  - Or add additional deconvolution/up-sampling layers afterwards (FCN8s, CRFasRNN)

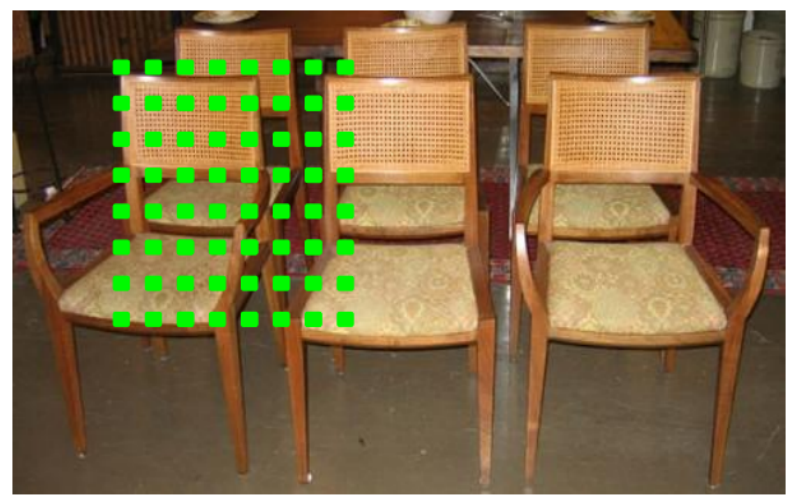
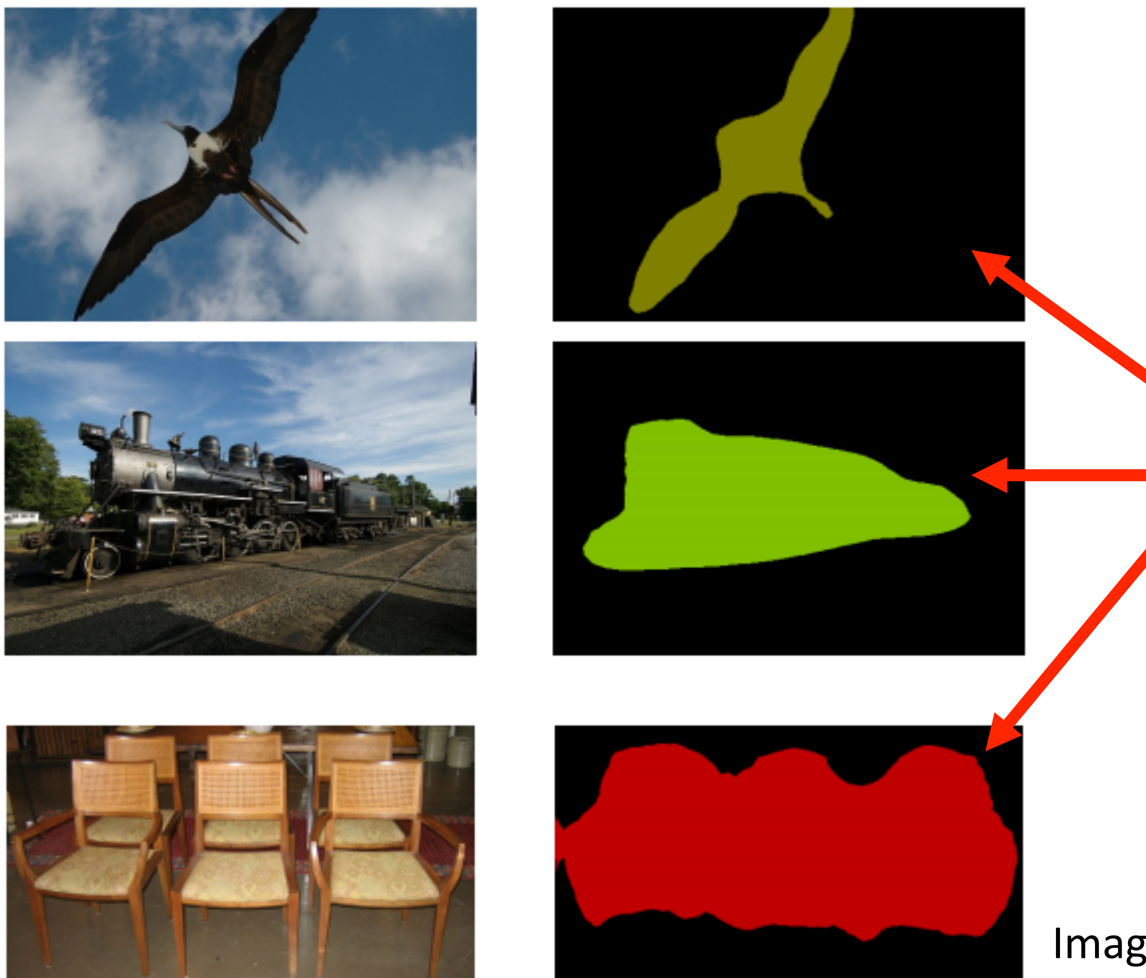


Image credit: Chen et al.

DeepLab

- Generate score map (8x)
- Interpolation to image size



Boundary not aligned.

Image credit: Chen et al.

DeepLab

- Densely-Connected CRF
  - Sharpen boundaries using image-based info
  - Gaussian spatial pairwise + Bilateral pairwise potential
  - Grid-search hyper-parameters over validation set

$$\underline{w_1} \exp\left(-\frac{||p_i - p_j||^2}{2\underline{\sigma_\alpha^2}} - \frac{||I_i - I_j||^2}{2\underline{\sigma_\beta^2}}\right) + \underline{w_2} \exp\left(-\frac{||p_i - p_j||^2}{2\underline{\sigma_\gamma^2}}\right)$$

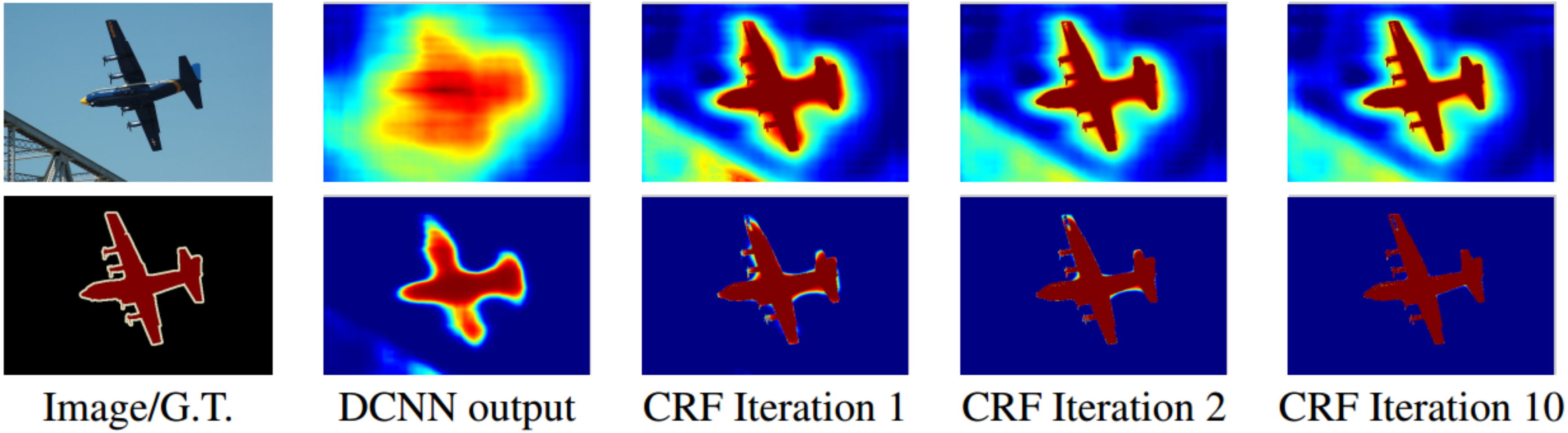


Image credit: Chen et al.

DeepLab

- Experimental Result

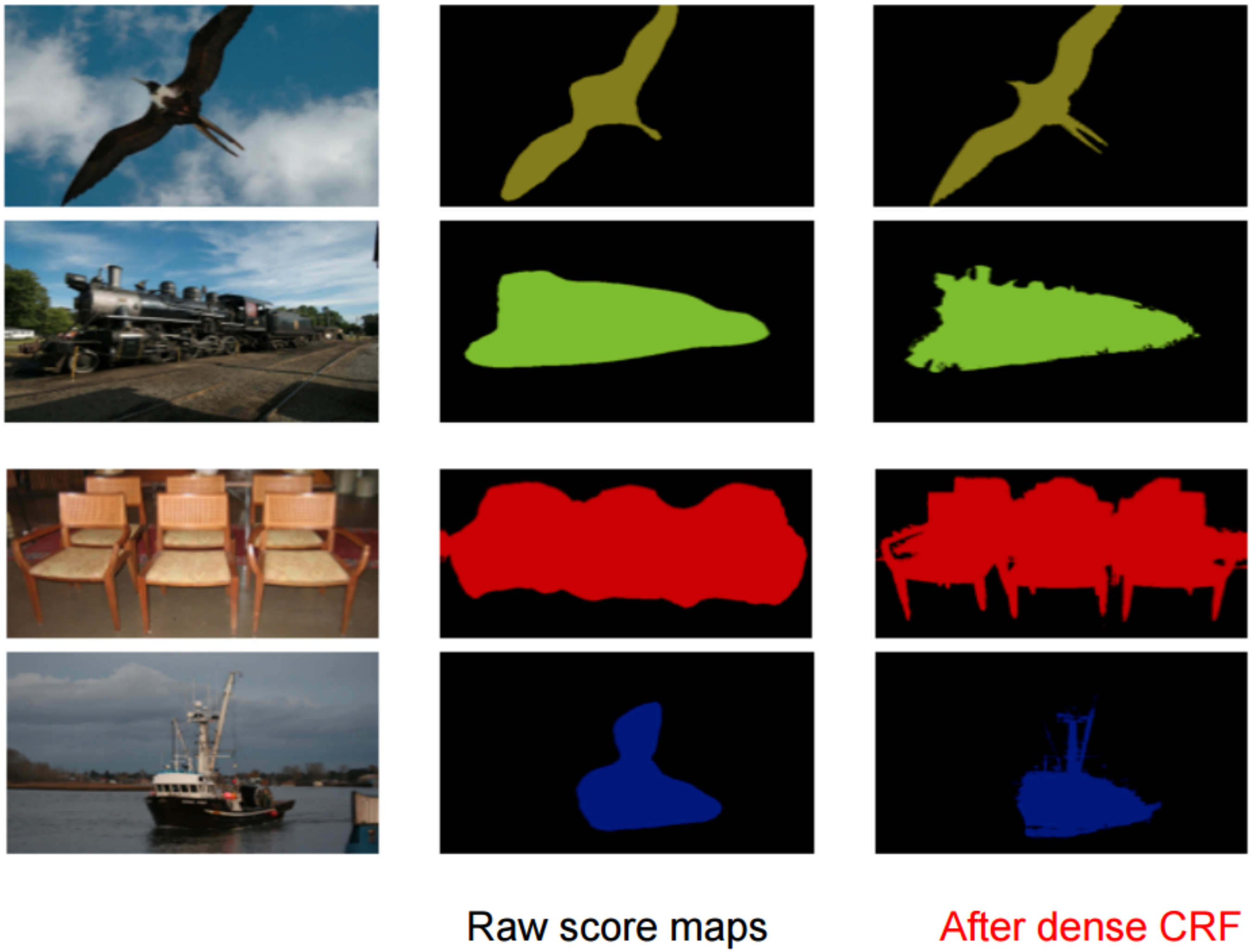


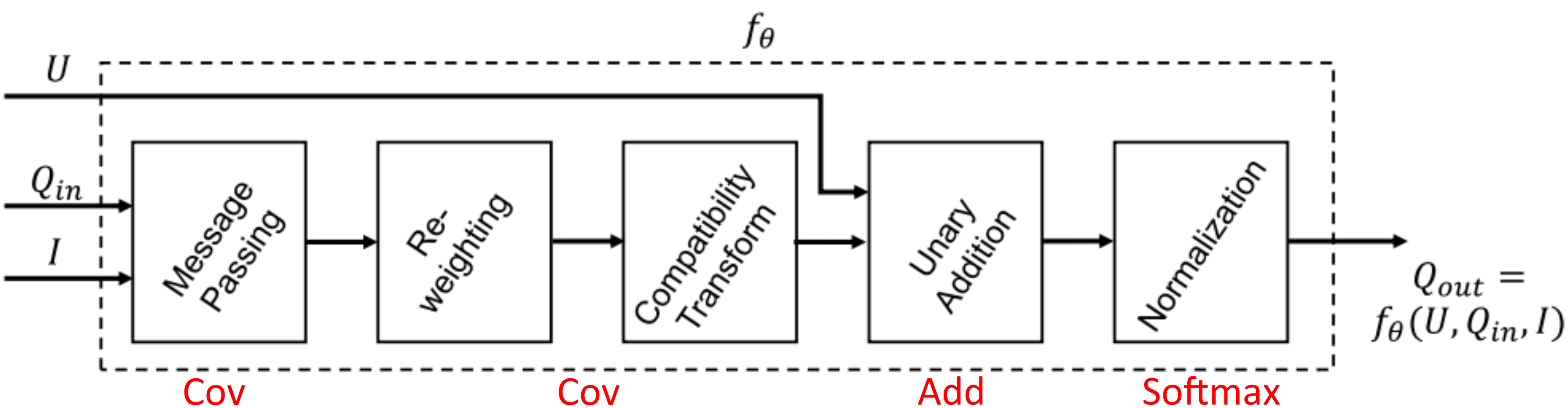
Image credit: Chen et al.

CRF as RNN

- Jointly learning the parameters of CNN and CRF

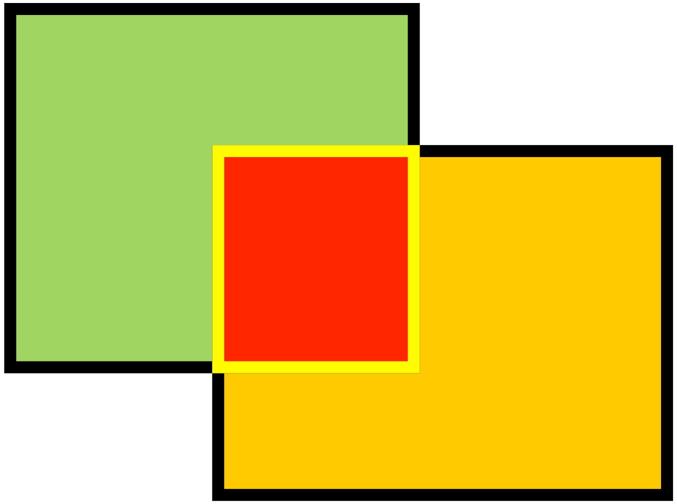
Algorithm 1 Mean field in fully connected CRFs

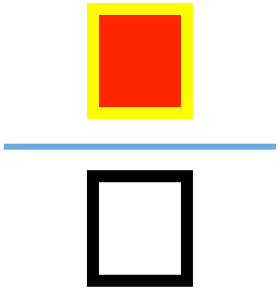
Initialize Q
while not converged **do**
  $\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l)$ for all m
  $\hat{Q}_i(x_i) \leftarrow \sum_{l \in \mathcal{L}} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l)$
  $Q_i(x_i) \leftarrow \exp\{-\psi_u(x_i) - \hat{Q}_i(x_i)\}$
  normalize $Q_i(x_i)$
end while



Quantitative Results

- Intersection-Over-Union





$$\text{segmentation accuracy} = \frac{\text{true positives}}{\text{true positives} + \text{false positives} + \text{false negatives}}$$

Quantitative Results on Pascal VOC 2012

- Improvements
  - Pre-CNN (< 50%)
  - CNN (60-64%)
  - CNN + CRF (67%)
  - Data Augmentation (71%)
    - Pretraining using other dataset (COCO ←→ Pascal)
    - Weakly supervision using bounding boxes
  - Learning the parameters for CRF and CNN jointly (74%)
  - Learning the pairwise label compatibility (77%)

Demo

- Notebook Example: FCN_8s + DenseCRF
- Online Demo: CRF-RNN

Summary

- Fully Convolutional Network
- Fully connected Conditional Random Field

Dataset

- MSRC
- SiftFlow
- Stanford
- LabelMe

Classic but might not large
enough for data-hungry models

- NYU
- Sun RGBD

RGBD-benchmark, large, indoor
Good for CNNs

- Pascal VOC
- Microsoft COCO

Object labeling, large, object seg
Good for CNNs, good for multi-task

- KITTI
- CamVid
- Cityscapes

Autonomous driving, video, scene
Enough for fine-tuning CNNs
Good for multi-task
Good for taking two courses

Reference List: Before CNN

- Shotton, Jamie, et al. "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation." *Computer Vision–ECCV 2006*. Springer Berlin Heidelberg, 2006. 1-15.
- Russell, Chris, Pushmeet Kohli, and Philip HS Torr. "Associative hierarchical crfs for object class image segmentation." *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009.
- Carreira, Joao, and Cristian Sminchisescu. "Constrained parametric min-cuts for automatic object segmentation." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.
- Ladicky, Lubor, et al. "Graph cut based inference with co-occurrence statistics." *Computer Vision–ECCV 2010*. Springer Berlin Heidelberg, 2010. 239-253.
- Liu, Ce, Jenny Yuen, and Antonio Torralba. "Nonparametric scene parsing: Label transfer via dense scene alignment." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- Ren, Xiaofeng, Liefeng Bo, and Dieter Fox. "Rgb-(d) scene labeling: Features and algorithms." *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012.
- Tighe, Joseph, and Svetlana Lazebnik. "Superparsing." *International Journal of Computer Vision* 101.2 (2013): 329-349.
- Shotton, Jamie, Matthew Johnson, and Roberto Cipolla. "Semantic texton forests for image categorization and segmentation." *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008.
- Gu, Chunhui, et al. "Recognition using regions." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- Kohli, Pushmeet, and Philip HS Torr. "Robust higher order potentials for enforcing label consistency." *International Journal of Computer Vision* 82.3 (2009): 302-324.
- Krähenbühl, Philipp, and Vladlen Koltun. "Efficient inference in fully connected crfs with gaussian edge potentials." *arXiv preprint arXiv:1210.5644* (2012).

Reference List: After CNN

- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*(2014).
- Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *arXiv preprint arXiv:1411.4038* (2014).
- Farabet, Clement, et al. "Learning hierarchical features for scene labeling."*Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.8 (2013): 1915-1929.
- Chen, Liang-Chieh, et al. "Semantic image segmentation with deep convolutional nets and fully connected crfs." *arXiv preprint arXiv:1412.7062*(2014).
- Liu, Ziwei, et al. "Semantic image segmentation via deep parsing network."*Proceedings of the IEEE International Conference on Computer Vision*. 2015.
- Schwing, Alexander G., and Raquel Urtasun. "Fully connected deep structured networks." *arXiv preprint arXiv:1503.02351* (2015).
- Zheng, Shuai, et al. "Conditional random fields as recurrent neural networks."*arXiv preprint arXiv:1502.03240* (2015).
- Lin, Guosheng, Chunhua Shen, and Ian Reid. "Efficient piecewise training of deep structured models for semantic segmentation." *arXiv preprint arXiv:1504.01013* (2015).