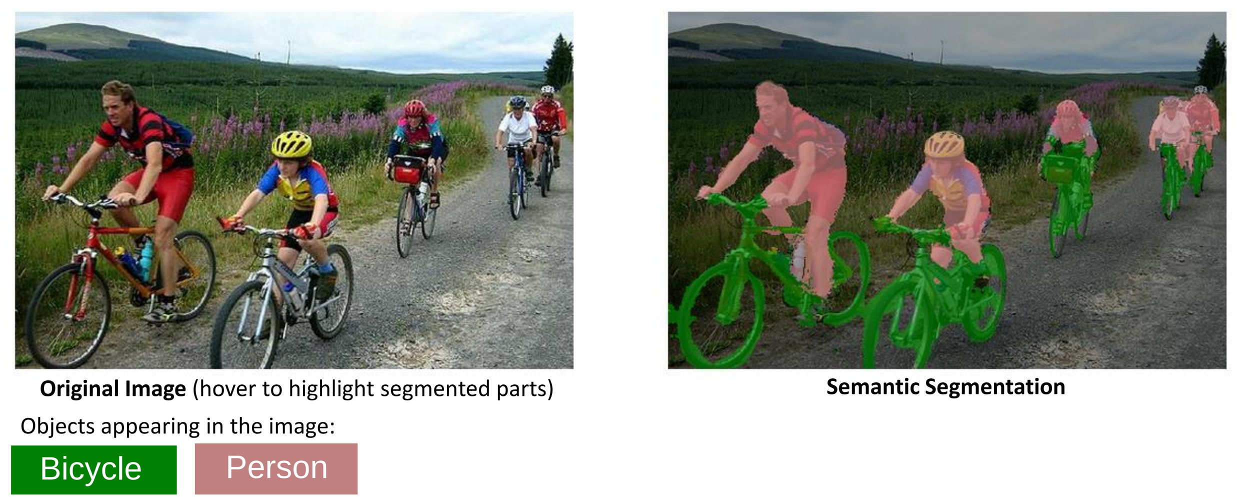
Conditional Random Fields as Recurrent Neural Networks

Shuai Zheng[1] Sadeep Jayasumana[1] Bernardino Romera-Paredes[1] Vibhav Vineet[1,2]
Zhizhong Su[3] Dalong Du[3] Chang Huang[3] Philip H. S. Torr[1]
[1]Torr-Vision Group, University of Oxford [2]Stanford University [3]Baidu IDL

UNIVERSITY OF OXFORD

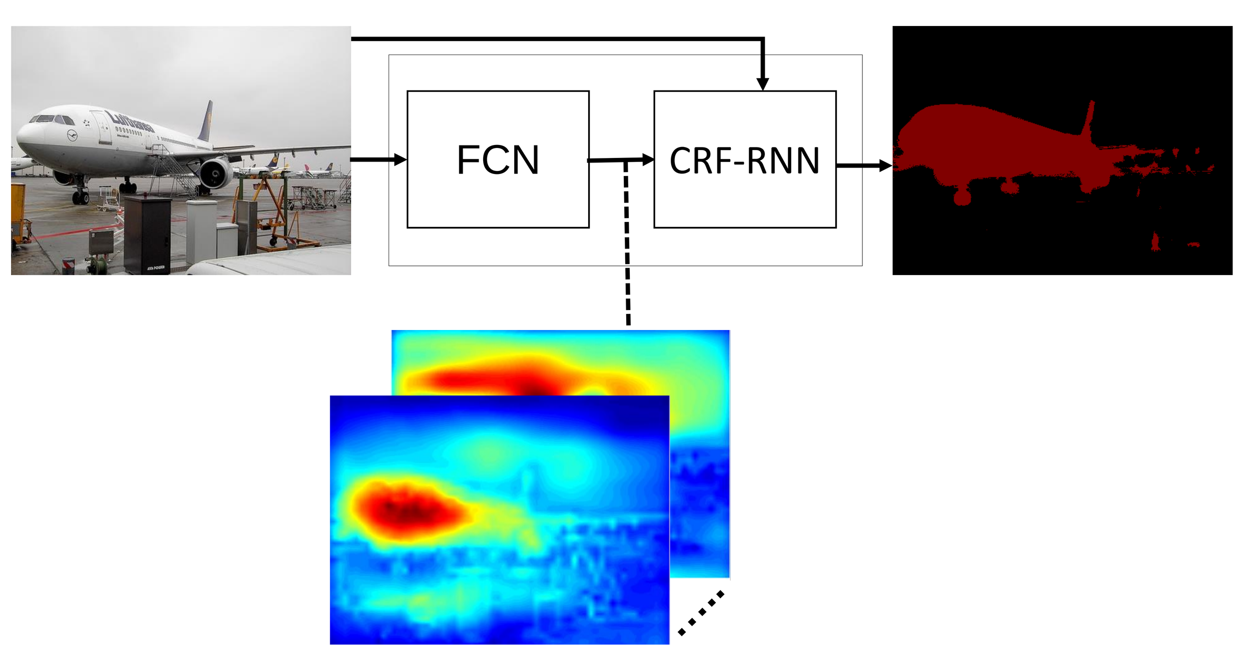
Semantic Image Segmentation

- Semantic image segmentation aims to assign labels (e.g. bicycle, person, etc) to each pixel.
- Recent approaches [4, 2] harness the capabilities of deep learning for image recognition to tackle pixel-wise labeling problems.
- One central issue in this methodology is the limited capacity of deep learning techniques to delineate visual objects.



End-to-end Trainable Network

- Deep Convolutional Neural Networks (CNNs) are successful at learning a good representation of the visual inputs.
- A Conditional Random Fields (CRF) can account for contextual information in the image (e.g. color consistency, spatial consistency, etc).
- Our approach CRF-RNN allows an end-to-end trainable neural network that combines both convolutional neural networks and probabilistic graphical models.



CRF-RNN

We formulate the fully-connected CRF with Gaussian pairwise potentials as a recurrent neural network (RNN) which can refine coarse outputs from a traditional CNN in the forward pass, while passing error differentials back to the CNN during training.

Fully-connected CRF

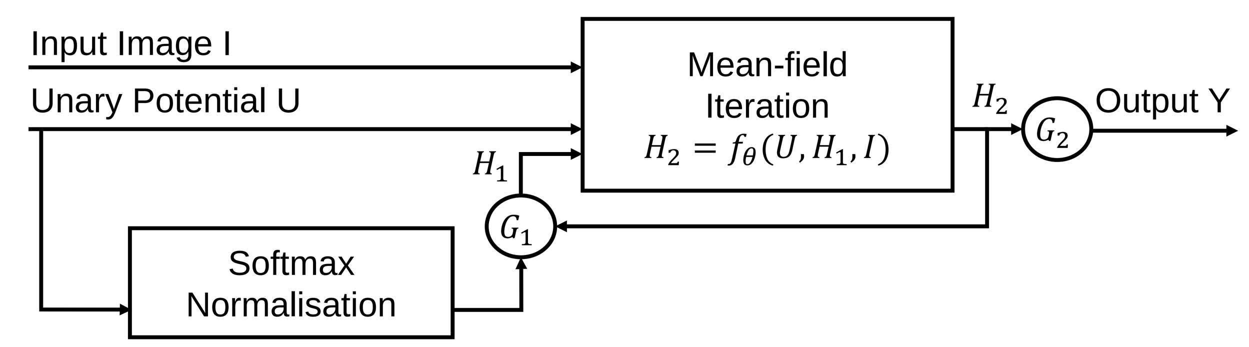
The energy of a label assignment x is given by:

$$E(x) = \sum_i \phi_U(x_i) + \sum_{i<j} \phi_P(x_i, x_j), \qquad (1)$$

- Unary energy $\phi_U(x_i)$: measures the cost if the label assignment disagrees with the initial classifier.
- Pairwise energy $\phi_P(x_i, x_j)$: measures the cost if two similar pixels (e.g. neighbor pixels or the pixels have similar color) take different label.

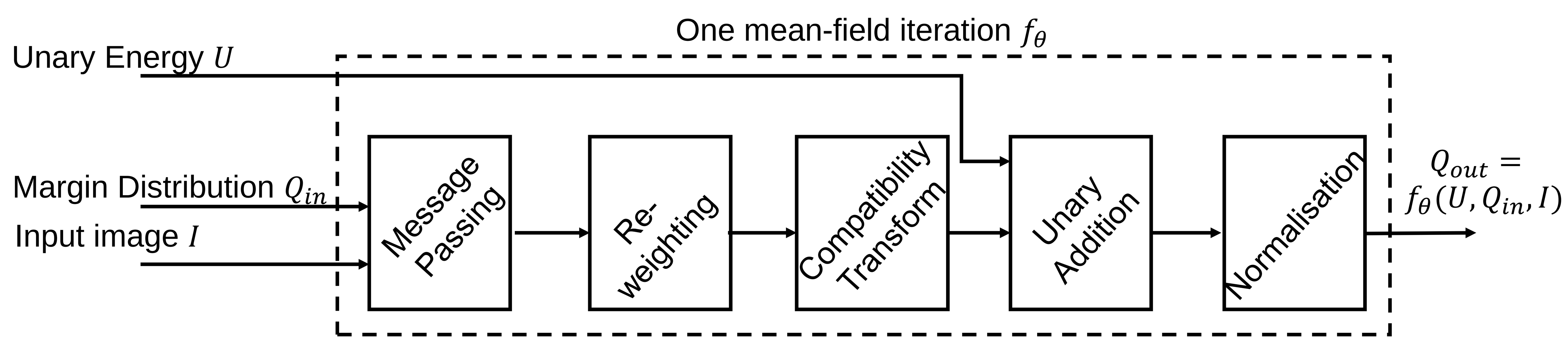
CRF-RNN network

- Multiple mean-field inference iterations are implemented by repeating the stack of CNN layers.
- This is equivalent to treating the mean-field inference as a recurrent neural network.



A mean-field inference iteration as a stack CNN layers

- The fully-connected CRFs with Gaussian pairwise [3] can be reformulated as RNNs.
- The filtering-based approximated mean-field inference can be broken into a series of CNN atomic operations.



Method	FCN [3]	DeepLab [1]	CRF-RNN
Mean IOU	68.3	69.5	72.9

Comparison results on VOC 12 reduced-validation set

Method	O_2P	CFM	FCN-8s	CRF-RNN
Mean IOU	18.1	34.4	37.78	39.28

Results on Pascal-context validation set

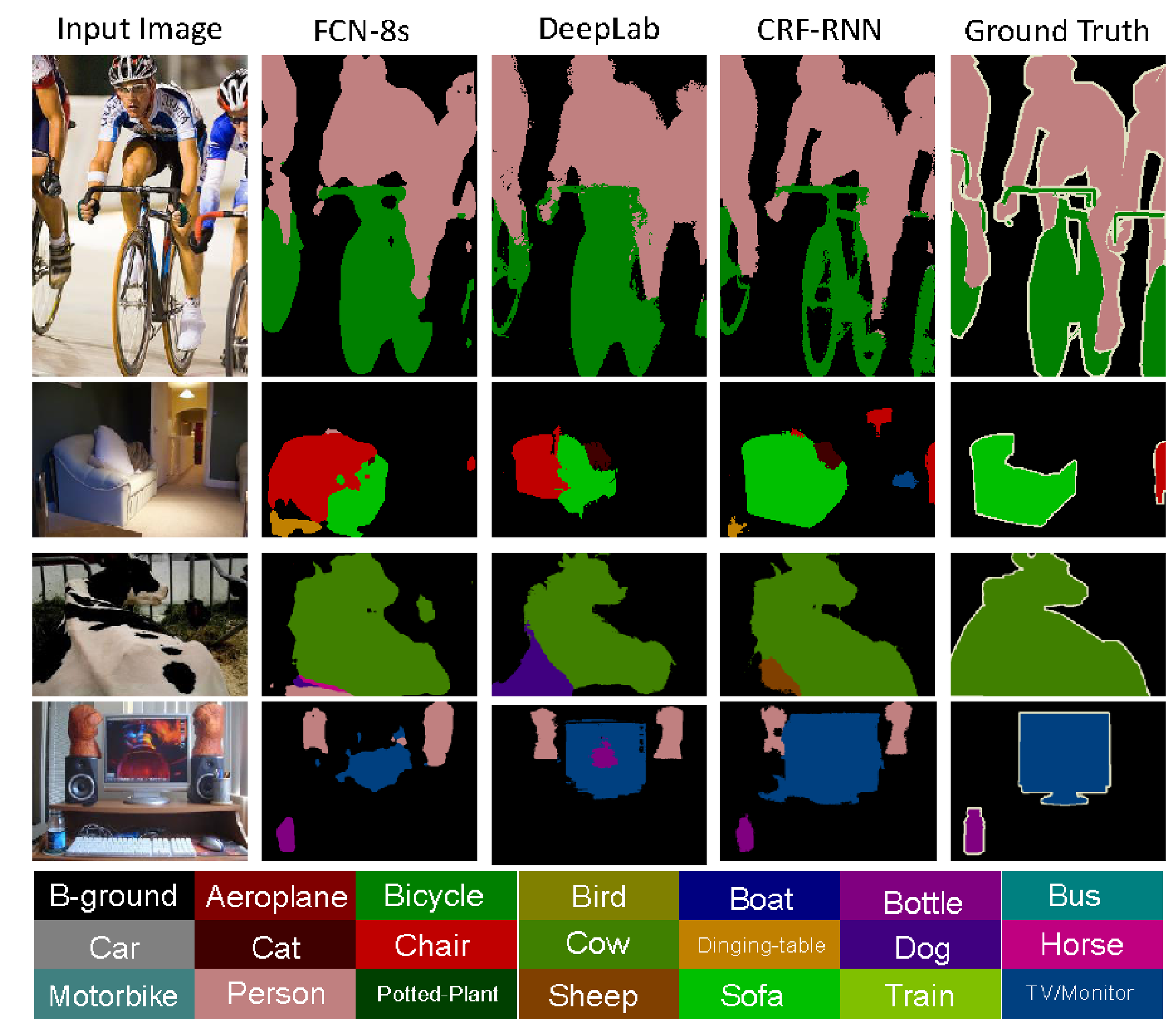
Summary

- CNNs yield a coarse prediction on pixel-wise labeling tasks.
- CRFs improve the results by accounting for the contexture information.
- Learning the CNNs and CRFs in an end-to-end pipeline significant improves the results.
- In subsequent work [1], we added higher order potentials to achieve a state-of-the-art performance of 77.9%.

Experiments

PASCAL VOC

We achieved state-of-the-art comparable performance (mean intersection-over-union 74.7%) in PASCAL VOC 2012 semantic image segmentation benchmark, Further work [1] achieved 77.9%. CRF-RNN GPU version takes less than 0.4 seconds for processing an image with resolution 500×500.



Live Demo & Source Code

Demo/Code: http://crfasrnn.torr.vision

Examples

Acknowledgement

This work is supported by Oxford ARC, EPSRC (EP/I001107/2), and ERC- 2012-AdG 321162-HELIOS.

References

[1] Anurag Arnab, Sadeep Jayasumana, Shuai Zheng, and Philip Torr. Higher order conditional random fields in deep neural networks. In *arXiv preprint arXiv:1511.08119*, 2015.

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *ICLR*, 2015.

[3] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 2011.

[4] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.