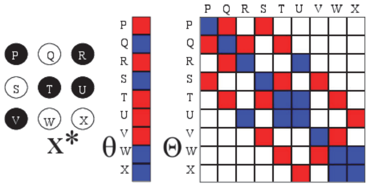
Fast, Exact & Multi-Scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs

Siddhartha Chandra & Iasonas Kokkinos

INRIA GALEN & Centrale Supélec Paris

Gaussian Conditional Random Fields



$$\pi(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{1}{2}\mathbf{x}^T\Theta\mathbf{x} + \theta^T\mathbf{x}\right)$$
$$\Theta\,\mathbf{x}^* = \theta$$

Unique and exact global optimum, pairwise interactions discovered from the data via **end-to-end** deep learning, and **fast** inference via efficient implementation.

Quadratic Energy Optimization

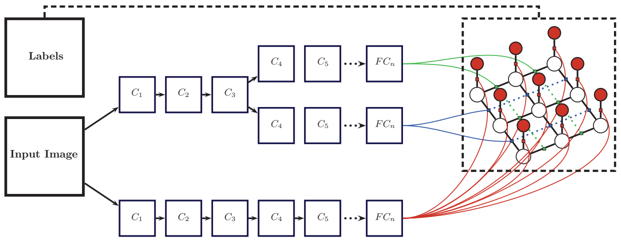
$$E(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T(A + \lambda\mathbf{I})\mathbf{x} - B\mathbf{x} \qquad (1)$$

If $(A + \lambda\mathbf{I})$ is symmetric positive definite, **unique global minimum** at,

$$(A + \lambda\mathbf{I})\mathbf{x} = B. \qquad (2)$$

Inference involves solving a **system of linear equations**.

Quadratic Optimization in Deep Learning



- Network populates unary and pairwise terms
- QO module proposes scores after inference
- Model parameters learnt end-to-end for arbitrary global loss (objective) $\mathcal{L}$
- **Gradient expressions:**
  - $$(A + \lambda\mathbf{I})\frac{\partial\mathcal{L}}{\partial B} = \frac{\partial\mathcal{L}}{\partial\mathbf{x}}, \qquad (3) \qquad\qquad \frac{\partial\mathcal{L}}{\partial A} = -\frac{\partial\mathcal{L}}{\partial B}\otimes\mathbf{x}. \qquad (4)$$
- Gradient computed **analytically** by solving a **system of linear equations**

Potts Type Model with Shared Pairwise Terms

Notation: $\hat{A}_{p_i,p_j}(l_i, l_j)$ is the pairwise energy term for pixels p_i, p_j taking the labels l_i, and l_j. Per-class scores and unaries are denoted by $\mathbf{x}_k$, and $\mathbf{b}_k$, where $k \in \{1, \cdots, L\}$.

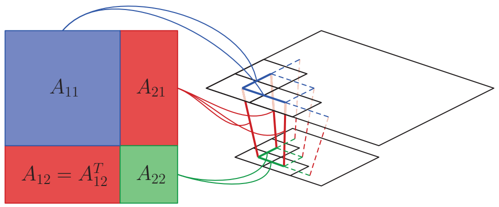
$$A_{p_i,p_j}(l_i, l_j) = \left\{\begin{matrix} 0 & l_i = l_j \\ A_{p_i,p_j} & l_i \neq l_j. \end{matrix}\right\} \qquad (5)$$

- Fewer parameters ($P \times P$ terms) compared to general setting ($PL \times PL$ terms) for P pixels, L labels. Reduction factor of **441** for VOC Pascal
- Algebraic simplifications enable us to infer scores for each label independently

$$\left(\lambda\mathbf{I} + (L-1)\,\hat{A}\right)\sum_i \mathbf{x}_i = \sum_i \mathbf{b}_i, \quad (6) \qquad\qquad (\lambda\mathbf{I} - \hat{A})\mathbf{x}_k = \mathbf{b}_k - \hat{A}\sum_i \mathbf{x}_i. \quad (7)$$

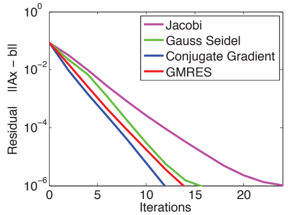
- Training **3x** faster, inference **6x** faster than general setting on VOC Pascal

Multi-Resolution Architecture



- Model parameters can capture pairwise terms between pixels across scales
- Information flow across scales
- Two kinds of interactions
  - Pairwise constraints between pixels at each resolution (Blue and Green)
  - Pairwise constraints between the same image region at different resolutions
- Inter-resolution constraints encourage pixels to share labels across resolutions

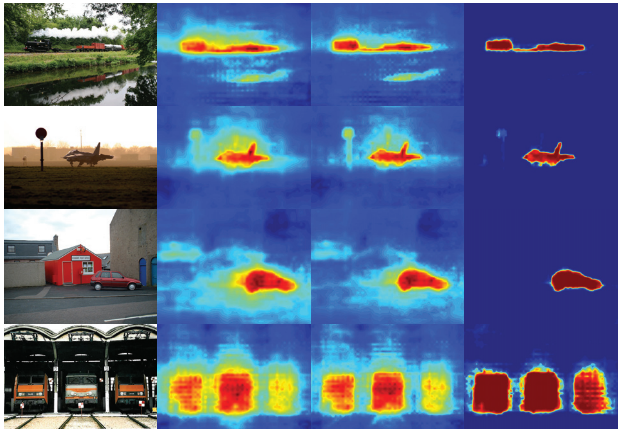
Implementation Details and Efficiency



- Congugate Gradient > other algorithms
- Caffe based implementation using efficient CUDA Sparse, Blas routines
- General Inference time ~ 0.02s
- Potts-type inference time ~ 0.003s
- Code available at `https://github.com/siddharthachandra/gcrf`.

Experimental Setup

- All methods use *VOC PASCAL 2012* image segmentation benchmark
- Basenet is a 3−resolution variant of **Deeplab-LargeFOV**
- We experiment with 4 variants of our method
- **QO**: General pairwise terms **QO^s**: Potts-type *shared* pairwise terms
- **QO^{res}**: One *QO* per resolution **QO^{mres}**: Multi-resolution *QO*



(a) image (b) unary (c) output (d) probability

Results

- Comparision with Basenet:

Method	IoU (%)	IoU after *Dense CRF* (%)
Basenet	72.72	73.78
QO	73.41	75.13
QO^s	73.20	75.41
QO^{mres}	73.86	75.46

- Comparision with *previously published* approaches:

Method	mean IoU (%)
Deeplab-Cross-Joint (Chen et al. ICCV 2015)	73.9
CRFRNN (Zheng et al. ICCV 2015)	74.7
Basenet	73.8
QO	75.1
QO^s	75.4
QO^{mres}	75.5
Deeplab-V2 (Chen et al. Arxiv 2016) *new*	79.7
QO^s + Deeplab-V2 *new*	80.2

(a) Image (b) Basenet (c) Basenet + DCRF (d) QO^{mres} (e) QO^{mres} + DCRF

Acknowledgements: This work has been funded by the EU Projects MOBOT FP7-ICT-2011-600796 and I-SUPPORT 643666 #2020.