Gaussian Conditional Random Fields

\[ p(x) = \frac{1}{Z} \exp \left( -\frac{1}{2} x^T \Theta x + \theta^T x \right) \]

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

\[ E(x) = \frac{1}{2} x^T (A + \lambda I) x - B x \]  

If \((A + \lambda I)\) is symmetric positive definite, unique global minimum at.

\[ (A + \lambda I) x = B \]  

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) \(L\)
- Gradient expressions:
  \[ \frac{\partial}{\partial A} \frac{\partial}{\partial B} \]  
- Gradient computed analytically by solving a system of linear equations

Potts Type Model with Shared Pairwise Terms

Notation: \( A_{p,k}(l_i, l_j) \) is the pairwise energy term for pixels \( p \), \( p \) taking the labels \( l_i \) and \( l_j \). Per-class scores and unaries are denoted by \( x_k \) and \( b_k \), where \( k \in \{1, \ldots, L\} \).

\[ A_{p,k}(l_i, l_j) = \begin{cases} 0 & l_i = l_j, \\ A_{p,k} & l_i \neq l_j \end{cases} \]  

Fewer parameters \((P \times P)\) 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
  \[ (A + (L-1)A) \sum_{l_i} x_i = \sum b_k, \]  
- \[ (A - A)x_k = b_k - \sum x_i \]  
- Training 3x faster, inference 6x faster than general setting on VOC Pascal

Gaussian Conditional Random Fields

\[ p(x) = \frac{1}{Z} \exp \left( -\frac{1}{2} x^T \Theta x + \theta^T x \right) \]

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

\[ E(x) = \frac{1}{2} x^T (A + \lambda I) x - B x \]  

If \((A + \lambda I)\) is symmetric positive definite, unique global minimum at.

\[ (A + \lambda I) x = B \]  

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) \(L\)
- Gradient expressions:
  \[ \frac{\partial}{\partial A} \frac{\partial}{\partial B} \]  
- Gradient computed analytically by solving a system of linear equations

Potts Type Model with Shared Pairwise Terms

Notation: \( A_{p,k}(l_i, l_j) \) is the pairwise energy term for pixels \( p \), \( p \) taking the labels \( l_i \) and \( l_j \). Per-class scores and unaries are denoted by \( x_k \) and \( b_k \), where \( k \in \{1, \ldots, L\} \).

\[ A_{p,k}(l_i, l_j) = \begin{cases} 0 & l_i = l_j, \\ A_{p,k} & l_i \neq l_j \end{cases} \]  

Fewer parameters \((P \times P)\) 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
  \[ (A + (L-1)A) \sum_{l_i} x_i = \sum b_k, \]  
- \[ (A - A)x_k = b_k - \sum x_i \]  
- 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

- Congruent Gradient > other algorithms
- Caffe-based implementation using efficient CUDA Sparse, Blue routines
- General Inference time \(\sim 0.02s\)
- Potts-type inference time \(\sim 0.003s\)
- 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 \(QOP\) Potts-type shared pairwise terms
- \(QO^{res}\): One \(QO\) per resolution \(QO^{res+}\): Multi-resolution \(QO\)

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