Superpixel Filtering for Mean Field Inference in CRF’s Integrated with CNN’s

Jack Hunt

Background
- Convolutional Neural Networks have been used in recent years to learn abstract representations of image features[1].
- Networks consisting solely of convolutions have been used to perform semantic segmentation[1].
- Mean Field Inference for Conditional Random Fields has been used to refine coarse segmentations[2].

CRF’s and CNN’s
- Recent work has seen Mean Field Inference for CRF’s reformulated as Recurrent Neural Networks, allowing end to end training[3].
- End to end training with CNN’s can be performed[3].
- Message passing can be reformulated as filtering, which can be expensive.

Approximating Filters
Superpixel segmentations may be used to facilitate the parallel approximation of filter outputs for Mean Field Inference in unified CNN-CRF frameworks, as demonstrated.

Technical Background
Message passing involves computing the following linear combination of bilateral and spatial kernels for each pair of nodes in the densely connected CRF[2]:-

\[
\begin{align*}
&= w_1 \exp \left( -\frac{|p_i - p_j|^2}{2\theta_1^2} - \frac{|I_i - I_j|^2}{2\theta_2^2} \right) \\
&+ w_2 \exp \left( -\frac{|p_i - p_j|^2}{2\theta_3^2} \right)
\end{align*}
\]

Where \( \theta_{\alpha,\beta,\gamma} \) are the kernel standard deviations, \( p_n \) and \( I_n \) are the \( n \)th pixels locations and channel values, respectively. The above application of filters is in practice performed as a convolution operation.

Simple Linear Iterative Clustering (SLIC) is a special case of K-Means that segments the given image in to approximately evenly spaced clusters of a given size, known as superpixels[4]. For example:-

![Figure: Example Superpixel Segmentation](image)

The above spatial and bilateral filters can then be applied in parallel over superpixels.

Results
- Semantic Segmentation task.
- VOC2012 reduced validation set used.
- Initialised from a pretrained Fully Connected Network.
- Evaluated in terms of Intersection over Union score.
- Average 70.51 IoU.
- Preliminary results comparable to CRF-RNN implementation using Permutohedral Lattice for filtering.

Conclusion
- Superpixel filtering has shown efficacy for end to end learning in CRF-RNN models.
- Suitable replacement for the Permutohedral Lattice.
- Despite being an approximation, training is stable.
- Higher order relationships may be efficiently modelled in future work.

References