Instance Segmentation

Min Bai

University of Toronto

April 4, 2016

Segmentation Review

- Task in computer vision
- Assign object class label to each pixel in image

Source: *Taegyu Lim*

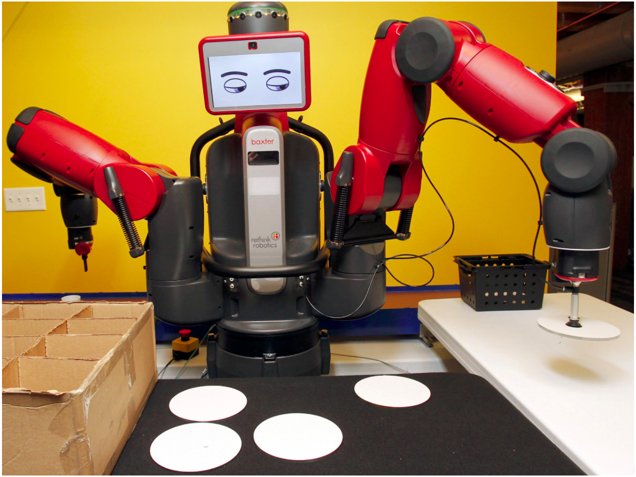
What is instance segmentation?

- Problem:
  - How many cows are there?
  - How many cars are there?

Source: *Taegyu Lim*

Why do we care?

- Richer information about world
- Object localization, tracking
- Interactions with objects



Current Approaches

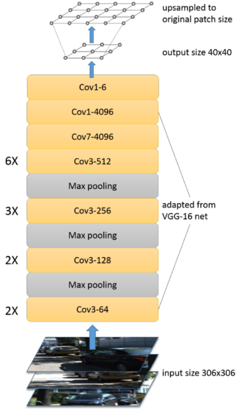
- Convolutional Neural Networks + Conditional Random Fields for depth ordering
  - Z. Zhang et al, Monocular Object Instance Segmentation and Depth Ordering with CNNs (ICCV 2015)
  - Z. Zhang et al, Instance-Level Segmentation with Deep Densely Connected MRFs (CVPR 2016)
- Recurrent Neural Networks
  - Parades et al, Recurrent Instance Segmentation

Monocular Object Instance Segmentation and Depth Ordering with CNNs

- Instance Segmentation and Depth Ordering Network
- MRF for Patch Merging
  - Reasons about a globally consistent depth ordering instances

Monocular Object Instance Segmentation and Depth Ordering with CNNs

- Instance Segmentation and Depth Ordering Network



Source: *Zhang et al*

MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p)\right) \\
&+ \sum_{p,p':\mathcal{C}(p)\neq\mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p'\in\mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

Source: *Zhang et al*

MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p) \right) \\
&+ \sum_{p,p':\mathcal{C}(p)\neq\mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p'\in\mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

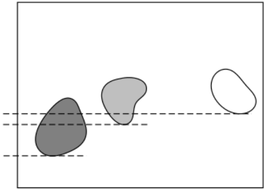
- First term: Global ordering should always $\geq$ ordering within patch seen by CNN

MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p) \right) \\
&+ \sum_{p,p' : \mathcal{C}(p) \neq \mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p' \in \mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

- Second term: Connected components are ordered vertically. Depth label should $\geq$ vertical ordering

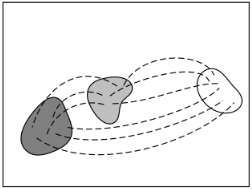


MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p) \right) \\
&+ \sum_{p,p':\mathcal{C}(p)\neq\mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p'\in\mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

- Third term: depth labeling for pixels belonging to different connected components should be different
- Sparse, random connectivity

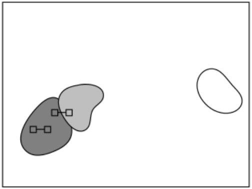


MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p) \right) \\
&+ \sum_{p,p' : \mathcal{C}(p) \neq \mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p' \in \mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

- Fourth term: depth labeling of neighboring pixels in same connected component should be the same



MRF Patch Merging

- Total energy function to be minimized

$$
\begin{aligned}
E(\mathbf{y}) &= \sum_{p} \left(E_{\text{CNN},p}(y_p) + E_{\text{CCO},p}(y_p)\right) \\
&+ \sum_{p,p':\mathcal{C}(p)\neq\mathcal{C}(p')} E_{\text{long},p,p'}(y_p, y_{p'}) \\
&+ \sum_{p,p'\in\mathcal{N}(p)} E_{\text{short},p,p'}(y_p, y_{p'}),
\end{aligned}
$$

- Minimized via quadratic pseudo-boolean optimization and graph cut

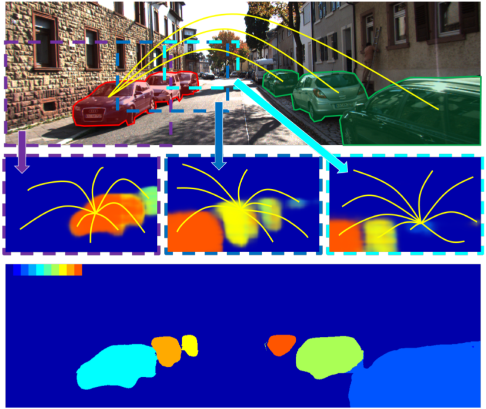
MRF Patch Merging

- Results - Left: input, middle: ground truth, right: result

Source: *Zhang et al*

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Extends previous project
- Instead of locally connected MRF, uses fully connected MRF within each patch and between connected components of different patches



Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Similar instance segmentation and depth ordering network
- Densely connected MRF
    - Does not reason about depth ordering - only instance identities
    - Longer range smoothness
    - Innovative method for MRF energy minimization

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Complete MRF energy term

$$E(\mathbf{y}) = E_{\text{smo}}(\mathbf{y}) + E_{\text{cnn}}(\mathbf{y}) + E_{\text{icc}}(\mathbf{y})$$

Source: *Zhang et al*

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Complete MRF energy term

$$E(\mathbf{y}) = E_{\text{smo}}(\mathbf{y}) + E_{\text{cnn}}(\mathbf{y}) + E_{\text{icc}}(\mathbf{y})$$

  - First term: encourages smoothness of instance label assignments

$$E_{\text{smo}}(\mathbf{y}) = \sum_{z} \sum_{i,j:i,j\in\mathcal{P}_z,i<j} \varphi_{\text{smo}}^{(z,i,j)}(y_i, y_j)$$

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Complete MRF energy term

$$E(\mathbf{y}) = E_{\text{smo}}(\mathbf{y}) + E_{\text{cnn}}(\mathbf{y}) + E_{\text{icc}}(\mathbf{y})$$

  - Second term: encourages global instance assignments of pixels to be same if CNN assigns them to be the same, and different otherwise

$$E_{\text{cnn}}(\mathbf{y}) = \sum_{z} \sum_{i,j:i,j \in \mathcal{P}_z, i<j} \varphi_{\text{cnn}}^{(z,i,j)}(y_i, y_j)$$

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Complete MRF energy term

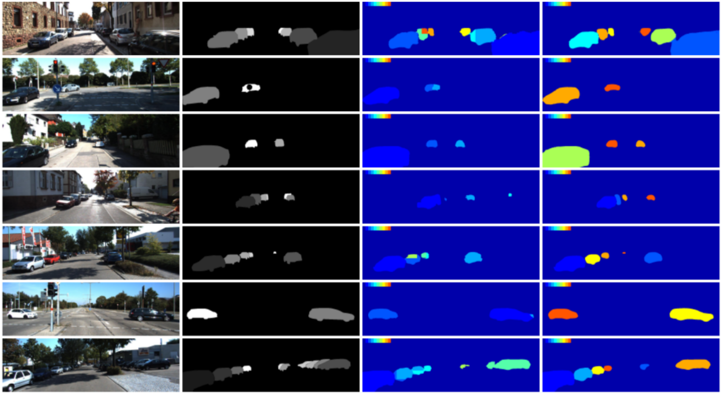
$$E(\mathbf{y}) = E_{\text{smo}}(\mathbf{y}) + E_{\text{cnn}}(\mathbf{y}) + E_{\text{icc}}(\mathbf{y})$$

  - Third term: encourages assignments of labels to pixels belonging to different inter-connected-components to be different

$$E_{\text{icc}}(y) = \sum_{m,n:m<n} \sum_{i,j:i\in\mathcal{C}_m, j\in\mathcal{C}_n} w_{\text{icc}}\mu_{\text{icc}}(y_i, y_j)$$

Instance-Level Segmentation with Deep Densely Connected MRFs, Zhang et al

- Results, from left to right: input, ground truth, previous paper, this paper



Source: *Zhang et al*