Learning Deconvolution Network for Semantic Segmentation

Hyeonwoo Noh, Seunghoon Hong, Bohyung Han
Department of Computer Science and Engineering, POSTECH, Korea

Presentation Outline
- Overview
- Background
- Algorithm
- Training
- Results
- Conclusions

Semantic Segmentation
Associating each pixel a pre-defined class label

Overview
Training a deconvolution network to perform semantic segmentation.

Presentation Outline
- Overview
- Background
- Algorithm
- Training
- Results
- Conclusions

CNN
CNN

**Pooling Layer**: non-linear down-sampling layer used to reduce spatial size.

\[
\begin{array}{cccc}
1 & 0 & 2 & 3 \\
4 & 6 & 6 & 8 \\
3 & 1 & 1 & 0 \\
1 & 2 & 2 & 4 \\
\end{array}
\rightarrow
\begin{array}{cc}
6 & 8 \\
3 & 4 \\
\end{array}
\]

**Rectified Linear Units Layer**: Applies the activation function.

\[
f(X) = \max(0, X)
\]

Deconvolution Neural Network

Reconstruction an image from the classification vector.

Deconvolution Neural Network

Reconstruction an image from the classification vector.

DCNN

**Unpooling Layer**: Reconstruct the original size activation.

DCNN

**Deconvolution Layer**: Densify the sparse activations obtained by unpooling.
**Fully Convolutional Network (FCN)**
A semantic segmentation using a single convolutional network.

**Conditional Random Fields (CRFs)**
Defining a conditional probability distribution over label sequences, rather than a joint distribution over both label and observation sequences.

**FCN**
A semantic segmentation using a single convolutional network.

**CRF**
\[ P(X, Y) \] – Joint distribution
\[ X \] – Observations (Pixels)
\[ Y \] – Labels
**CRF**

\[
P(Y \mid X) - \text{Conditional Distribution}
\]

\[
P(Y \mid X) = \frac{P(X, Y)}{P(X)} = \frac{1}{Z(X)} \prod_{c \in C} \Psi_c(X_c, Y_c)
\]

**Presentation Outline**

- Overview
- Background
- Algorithm
- Training
- Results
- Conclusions

**Algorithm**

[Hyeonwoo Noh et al. 2015]

**Algorithm Evolutions**

Several Evolutions:
- DeconvNet
- EDeconvNet
- DeconvNet+CRF
- EDeconvNet+CRF
Training is done on PASCAL 2012 dataset
- 2.9M images
- 250x250 images, 20 classes
- Training took 6 days on Nvidia TitenX

Training is a great challenge as the network’s depth leads to significant number of parameters.

Batch Normalization: Normalizing each input channel to standard Gaussian distribution.

Two Stage Training: To improve convergence rate, training would first be done with “easy” examples and than with “challenging” examples.
Conclusions

- A novel semantic segmentation algorithm by learning a deconvolution network.
- Ensemble approach of FCN + CRF.

References


