

Semantic Image Segmentation via Deep Parsing Network

Ziwei Liu*, Xiaoxiao Li*, Ping Luo, Chen Change Loy, Xiaoou Tang



Multimedia Lab, The Chinese University of Hong Kong

Problem



Problem



Previous Attempts



SVM



SVM + MRF

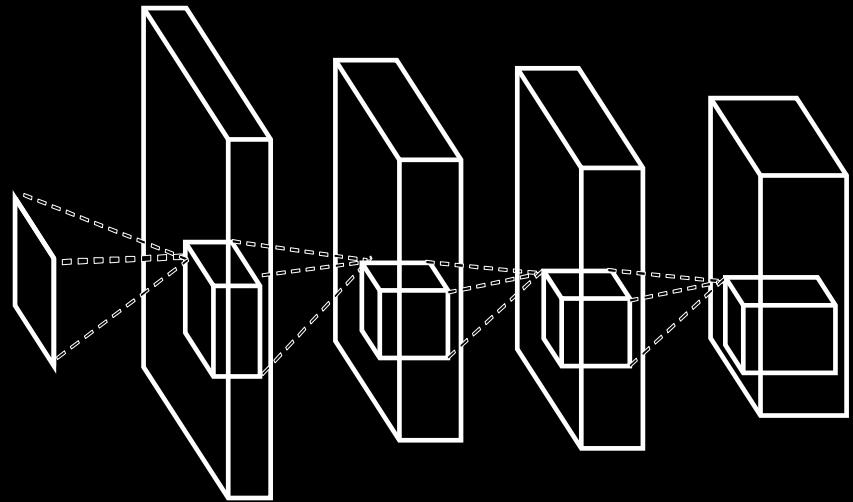


CNN



CNN + MRF ?

State-of-the-arts

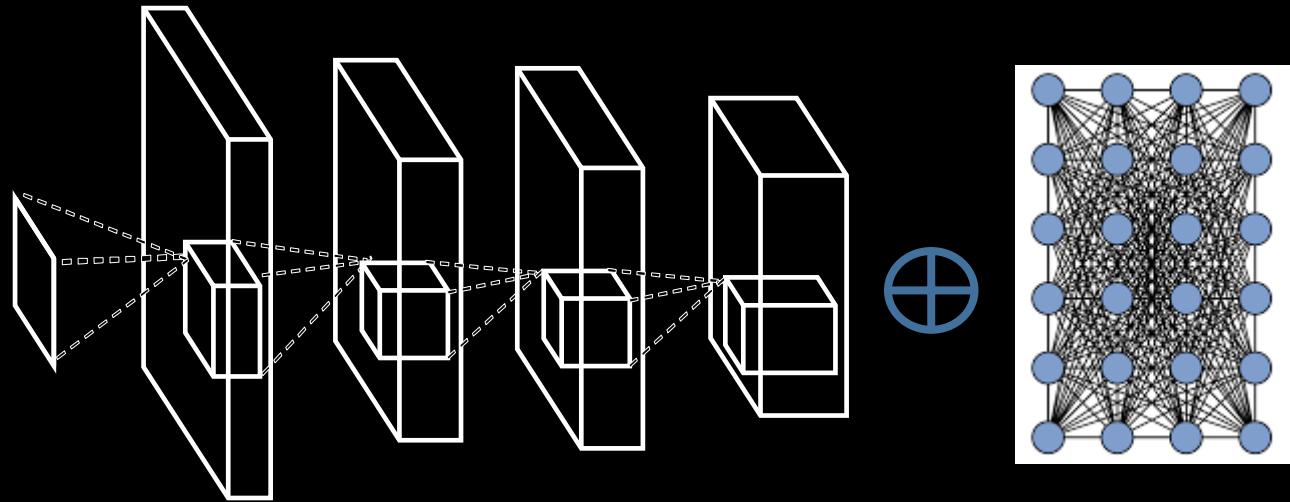


Fully Convolutional Network

[Long et al. CVPR 2015]

Learned Features	✓
Pairwise Relations	✗
Joint Training	-
# Iterations	-

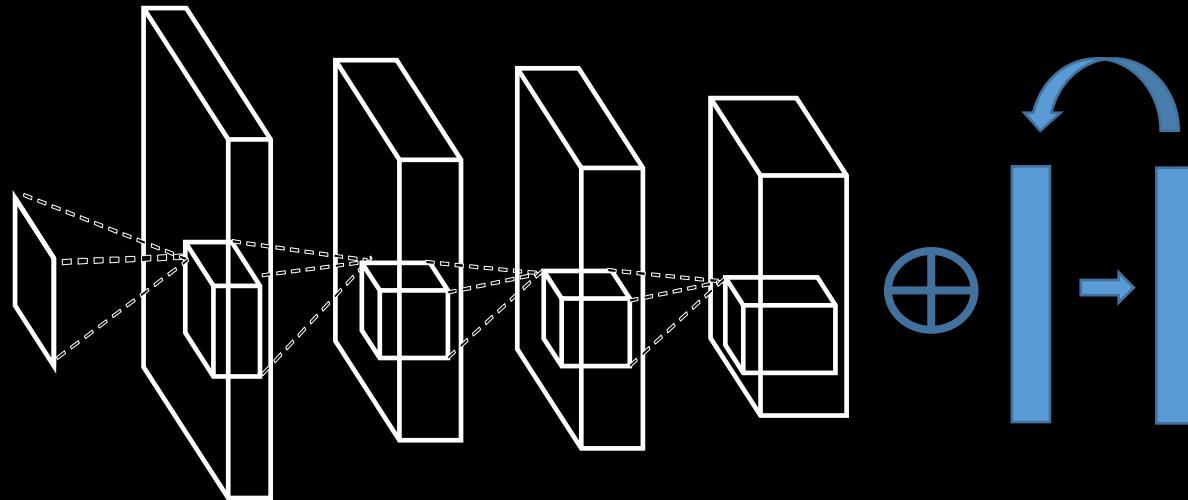
State-of-the-arts



DeepLab
[Chen et al. ICLR 2015]

Learned Features	✓
Pairwise Relations	✓
Joint Training	✗
# Iterations	10

State-of-the-arts

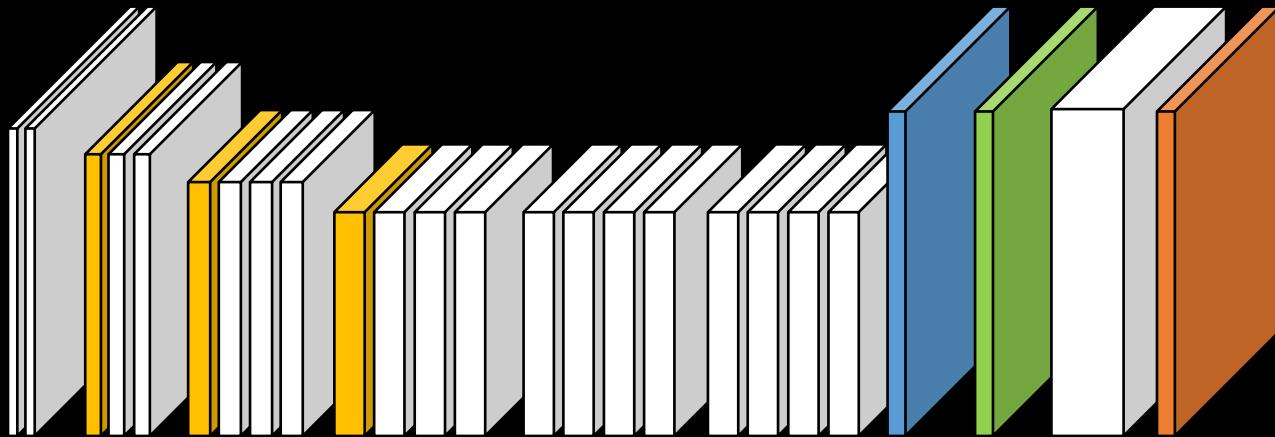


CRF as RNN

[Zheng et al. ICCV 2015]

Learned Features	✓
Pairwise Relations	✓
Joint Training	✓
# Iterations	10

State-of-the-arts



Deep Parsing Network (DPN)

Learned Features	✓
Pairwise Relations	✓
Joint Training	✓
# Iterations	1

Contributions

- Extend MRF to incorporate richer relationships
- Formulate mean field inference of high-order MRF as CNN
- Capable of joint training and one-pass inference

Revisit MRF

$p_i(\text{label} = 'table') = 0.8$



Energy Function

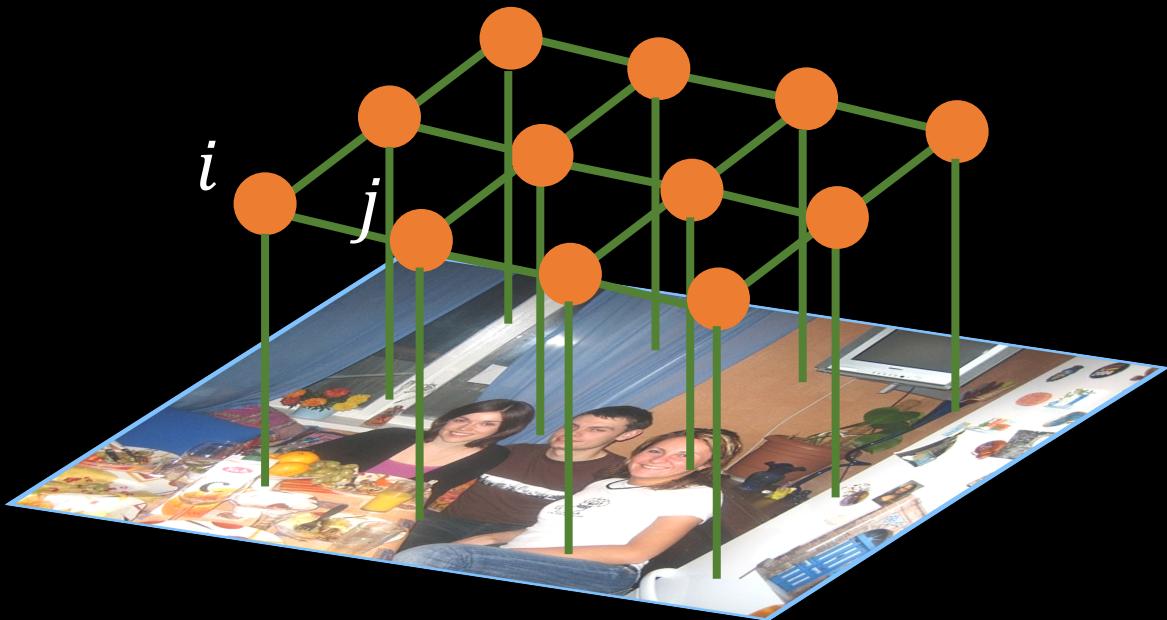
$$\min E = \text{Unary} + \text{Pair}$$

Unary Term

$$\text{Unary} = - \sum_i \ln p_i(\text{label})$$

Revisit MRF

$$diss(i, j) = (\text{Image } i, \text{ Image } j) = 0.8$$



Appearance Consistency

Energy Function

$$\min E = Unary + Pair$$

Unary Term

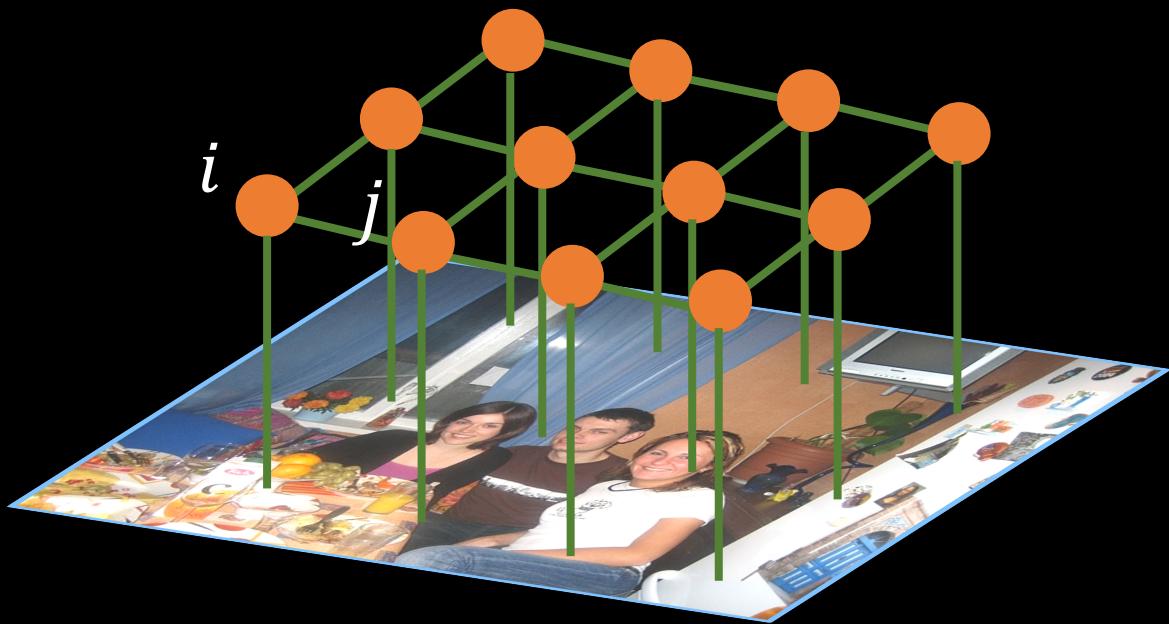
$$Unary = - \sum_i \ln p_i(label)$$

Pairwise Term

$$Pair = \sum_{i,j} cost(i) * diss(i,j)$$

Revisit MRF

$$cost(i; label = 'table') = 0.1$$



Label Consistency

Energy Function

$$\min E = Unary + Pair$$

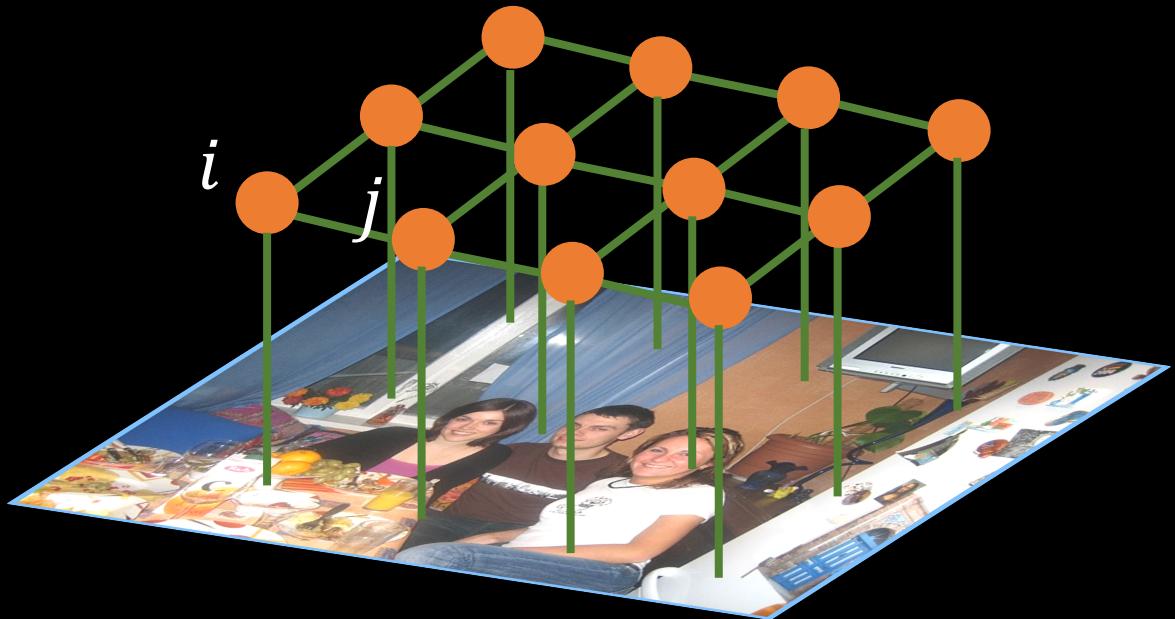
Unary Term

$$Unary = - \sum_i \ln p_i(label)$$

Pairwise Term

$$Pair = \sum_{i,j} cost(i) * diss(i,j)$$

Richer Relationships in DPN



Energy Function

$$\min E = \text{Unary} + \text{Pair}$$

Unary Term

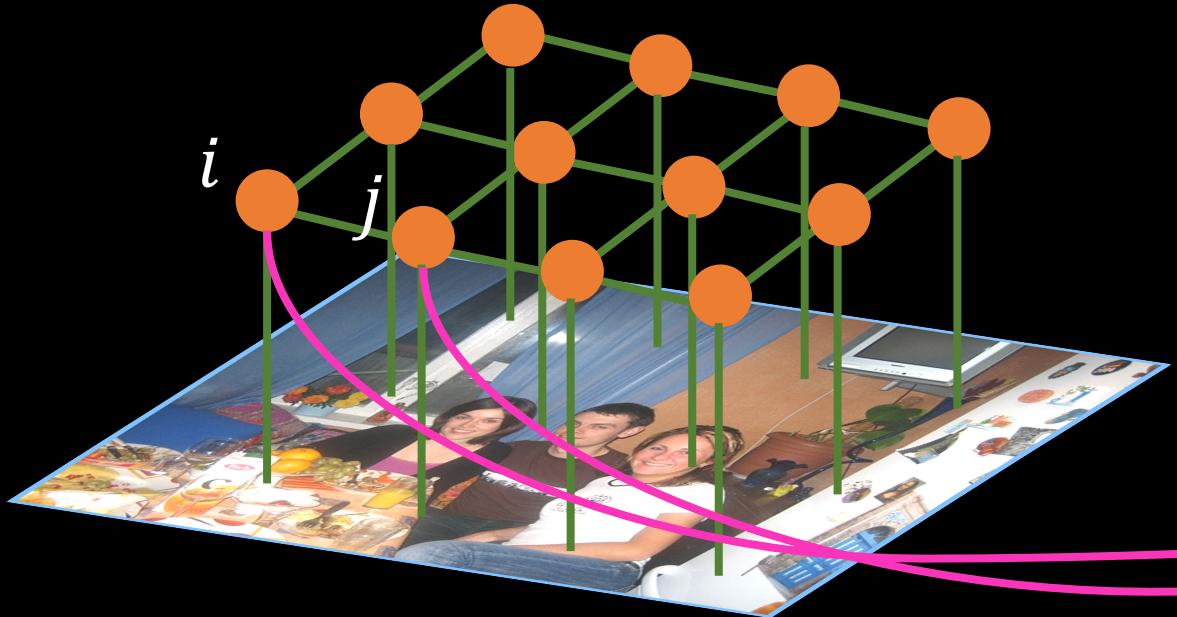
$$\text{Unary} = - \sum_i \ln p_i(\text{label})$$

Pairwise Term

$$\text{Pair} = \sum_{i,j} \text{cost}(i) * \text{diss}(i,j)$$

Richer Relationships in DPN

Triple Penalty



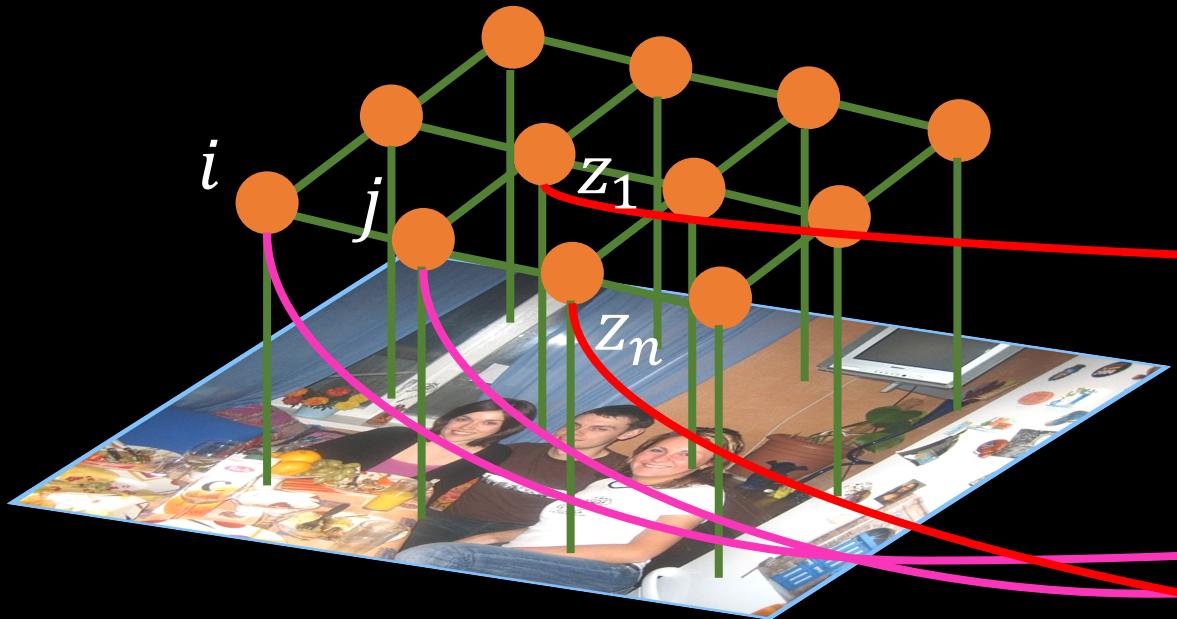
Pairwise Term

$$Pair = \sum_{i,j} cost(i) * diss(i,j)$$



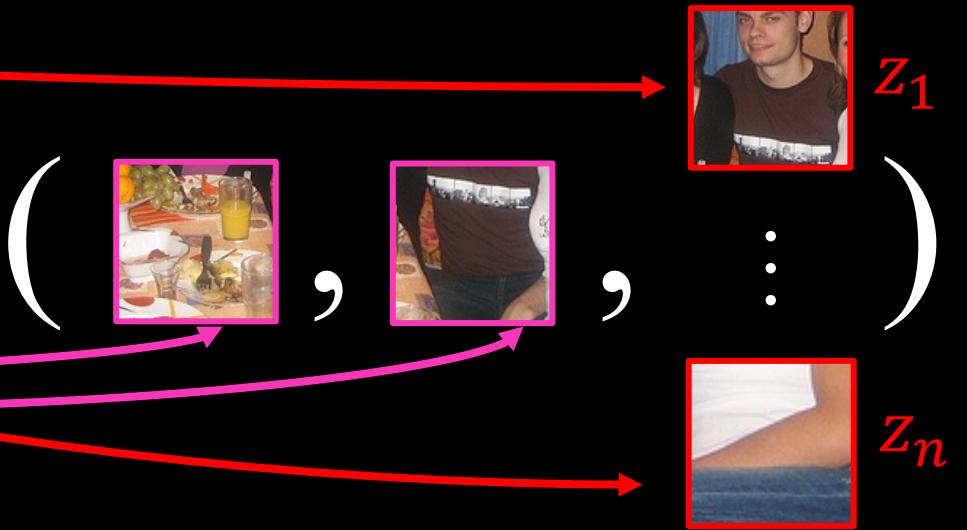
Richer Relationships in DPN

Triple Penalty



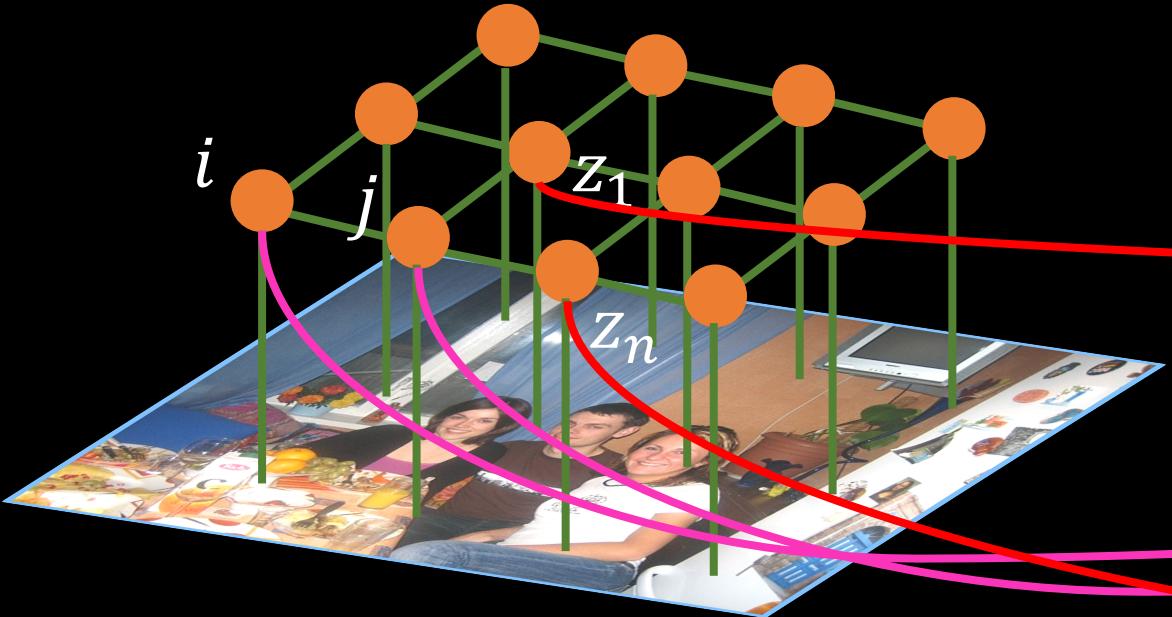
Pairwise Term

$$Pair = \sum_{i,j} cost(i) * diss(i,j)$$



Richer Relationships in DPN

Triple Penalty



Pairwise Term

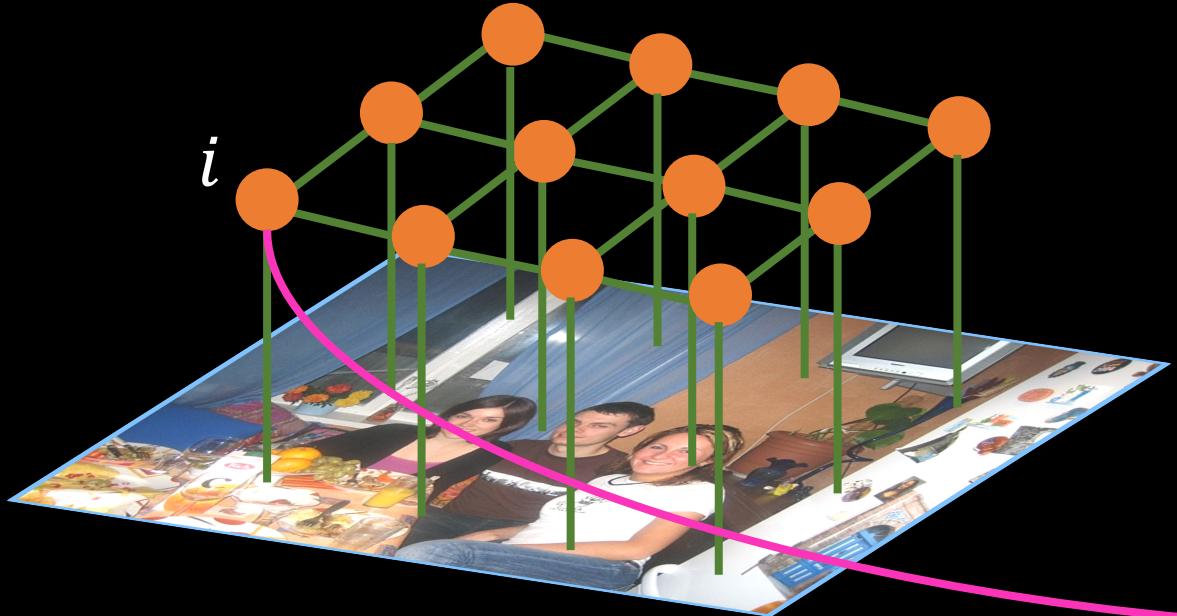
$$Pair = \sum_{i,j} cost(i) * \sum_z diss(i,j; z)$$



Triple Penalty

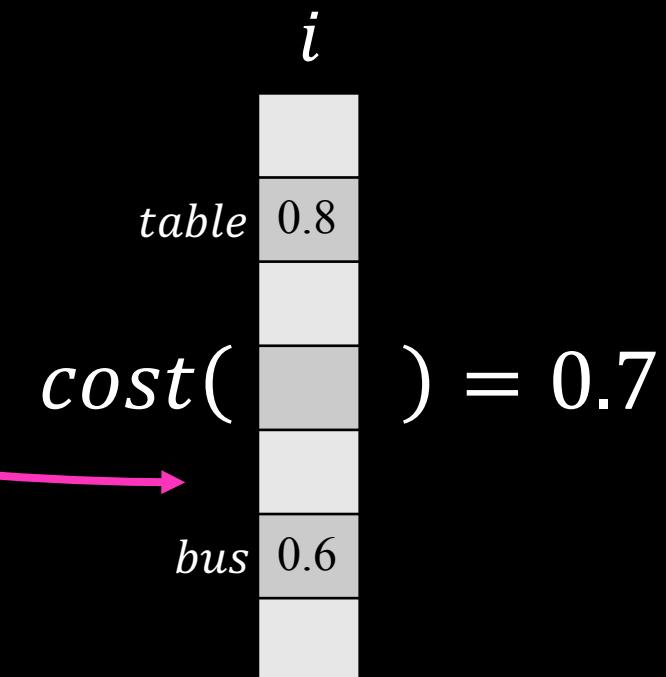
Richer Relationships in DPN

Mixture of Label Contexts



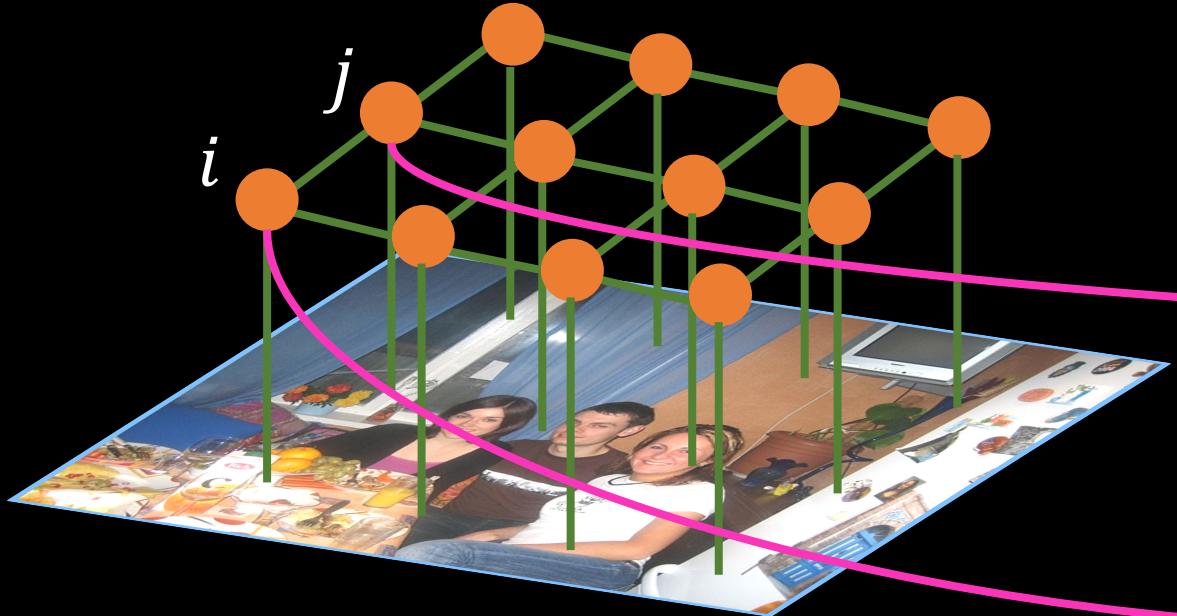
Pairwise Term

$$Pair = \sum_{i,j} [cost(i)] * \sum_z diss(i,j; z)$$



Richer Relationships in DPN

Mixture of Label Contexts



Pairwise Term

$$Pair = \sum_{i,j} cost(i, \textcolor{violet}{j}) * \sum_z diss(i, j; \textcolor{red}{z})$$

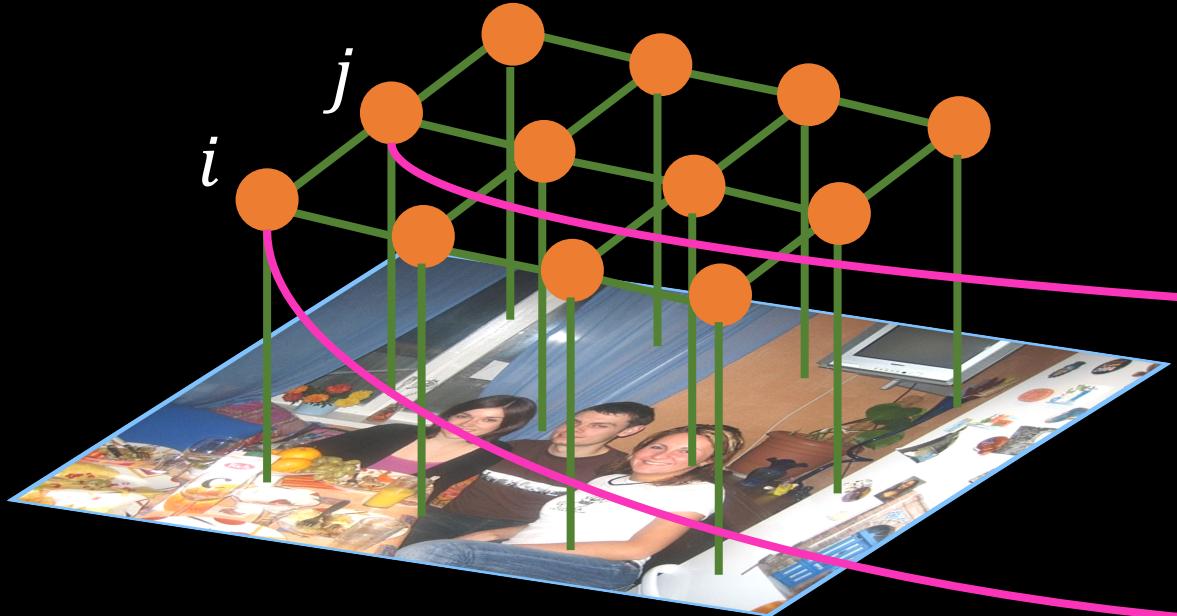
i 

j 

$cost(\textit{table}, \textit{person}, \dots) = 0.2$

Richer Relationships in DPN

Mixture of Label Contexts



Pairwise Term

$$Pair = \sum_{i,j} \boxed{cost(i,j)} * \sum_z diss(i,j; z)$$

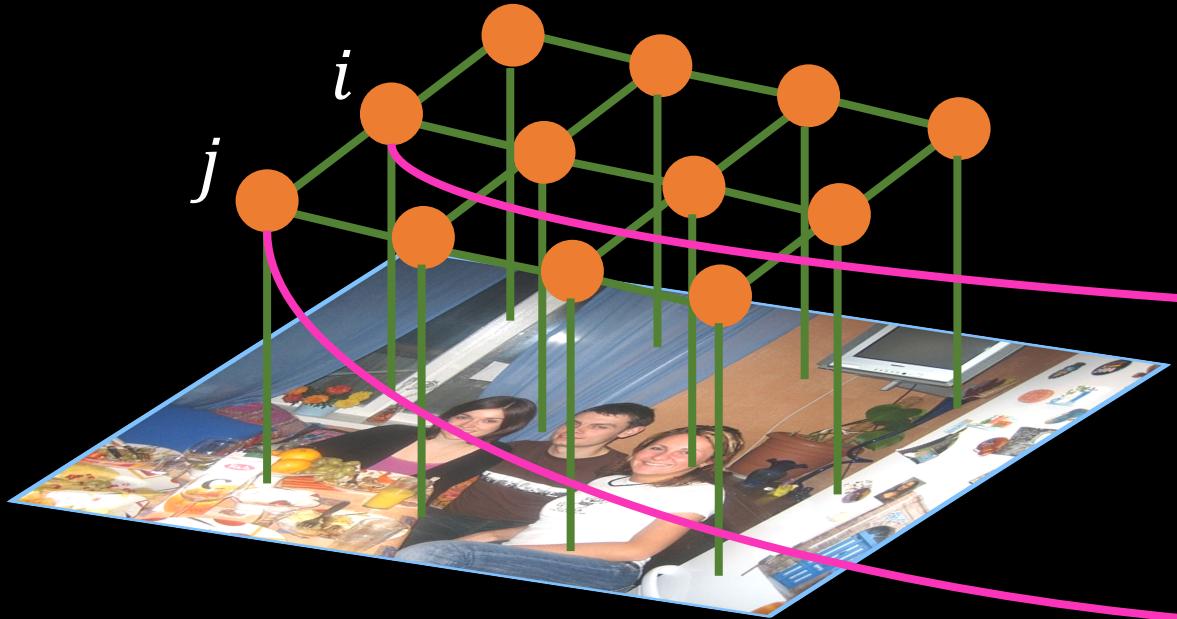
i *j*

The diagram illustrates a sequence of tokens and their associated costs. It consists of two vertical columns of boxes. The left column contains the tokens *table*, *cost*(, and), each with a cost of 0.8. The right column contains the tokens , *person*, and), each with a cost of 0.6. A pink arrow points from the first box in the left column to the first box in the right column, indicating a transition or comparison between these specific tokens.

<i>table</i>	0.8		
		<i>person</i>	0.6
<i>cost</i> (,)
)			= 0.2

Richer Relationships in DPN

Mixture of Label Contexts



Spatial Order

Pairwise Term

$$Pair = \sum_{i,j} cost(i,\textcolor{violet}{j}) * \sum_{\textcolor{red}{z}} diss(i,j; \textcolor{red}{z})$$

j

i

person
cost(

table

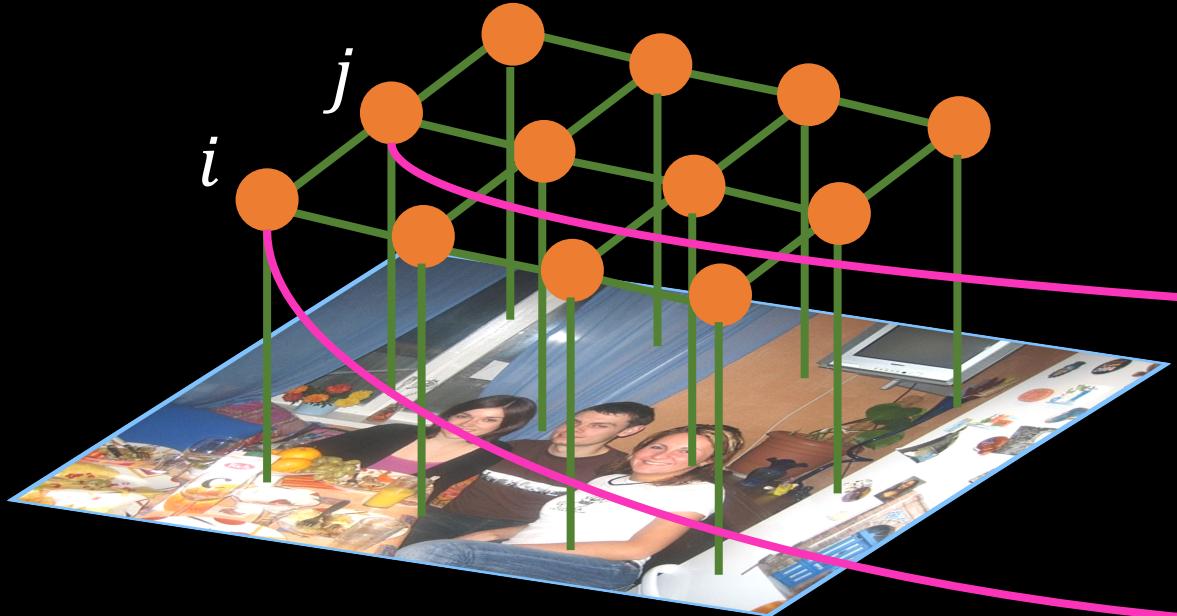
1

0.8

) = 0.8

Richer Relationships in DPN

Mixture of Label Contexts



Pairwise Term

$$Pair = \sum_{i,j} cost(i, \textcolor{violet}{j}) * \sum_z diss(i, j; \textcolor{red}{z})$$

i 

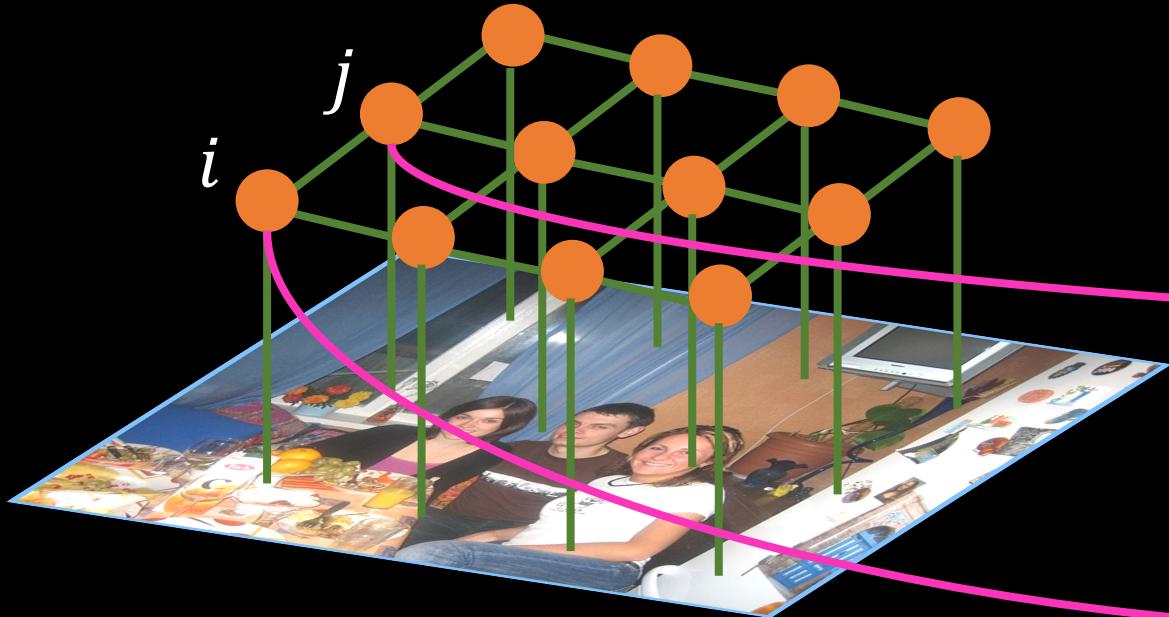
j 

$cost(\text{table}, \text{person}) = 0.2$

table	0.8
person	0.6
other	0.0

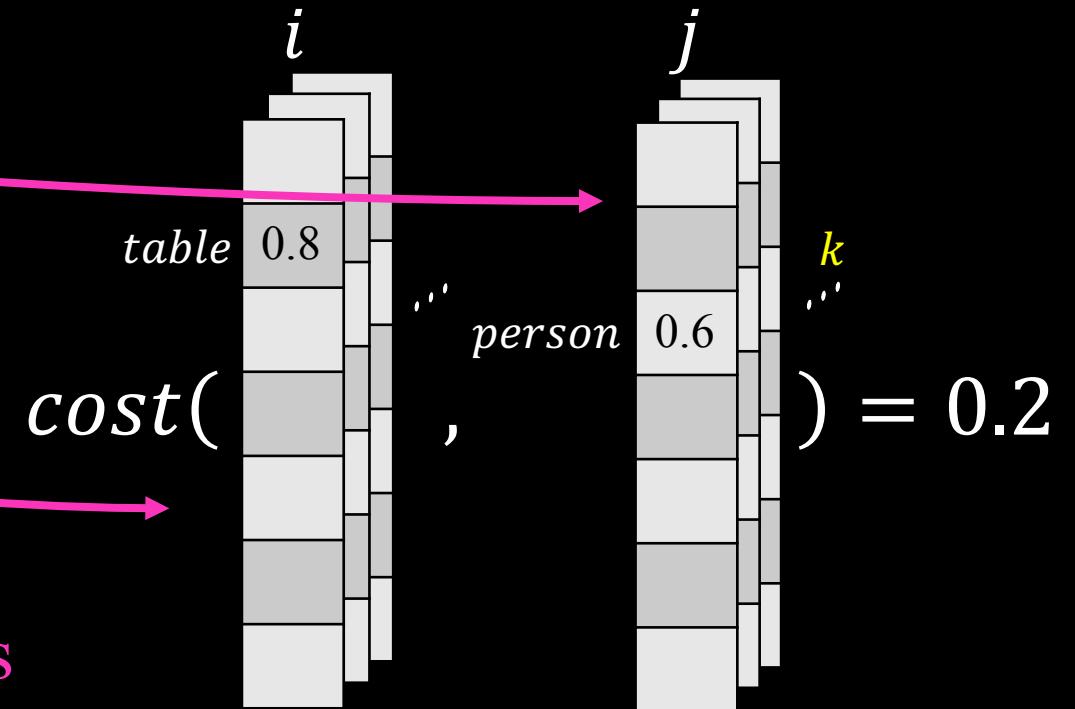
Richer Relationships in DPN

Mixture of Label Contexts



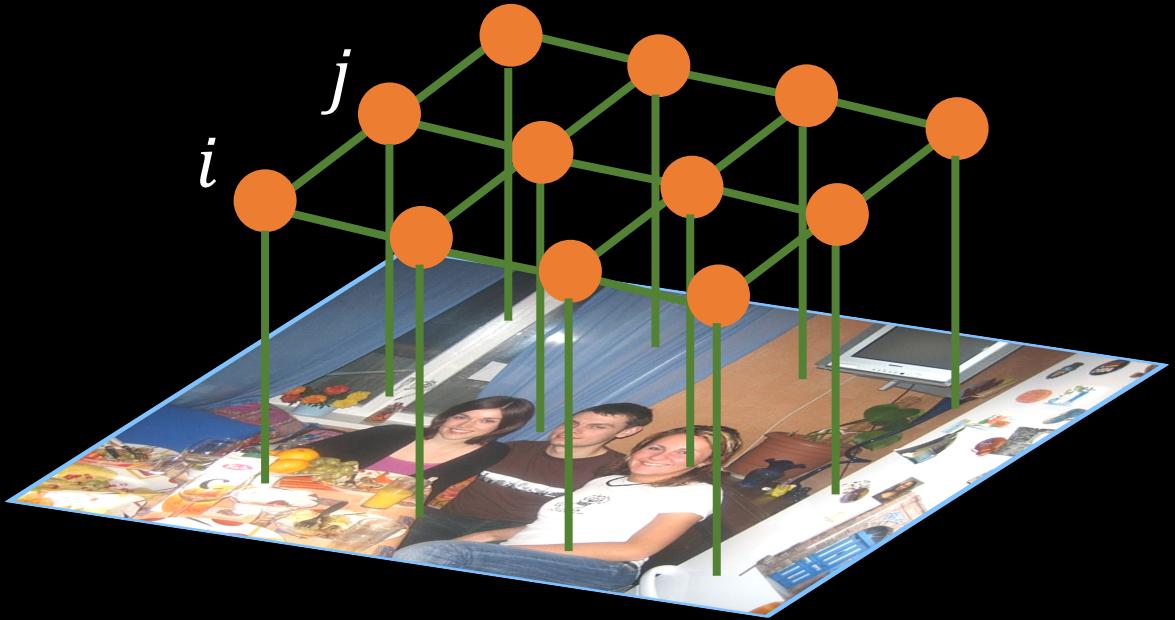
Pairwise Term

$$Pair = \sum_{i,j} \sum_k cost_k(i,j) * \sum_z diss(i,j; z)$$



Mixture of Label Contexts

Solve High-order MRF as Convolution



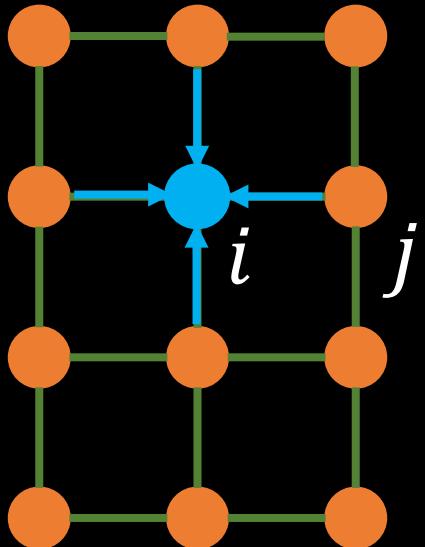
Pairwise Term

$$Pair = \sum_{i,j} \sum_k cost_k(i, j) * \sum_z diss(i, j; z)$$

Mean Field Solver

$$p_i \propto \exp \left\{ - \left(Unary_i + \sum_j Pair_{i,j} * p_j \right) \right\}$$

Solve High-order MRF as Convolution



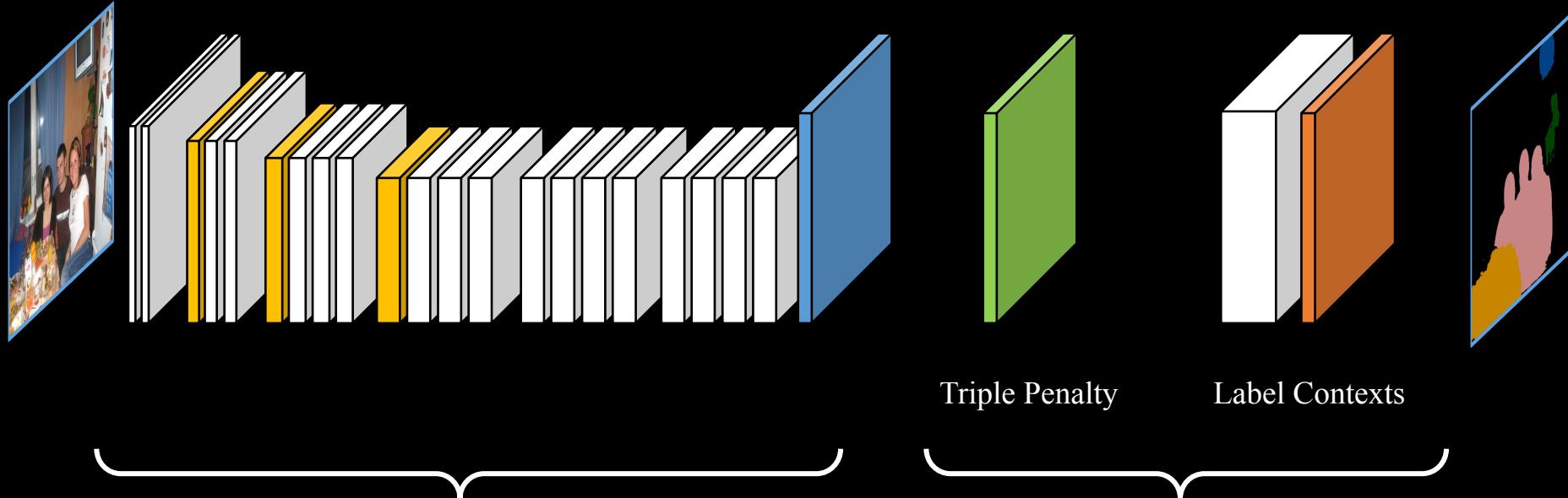
Iterative Updating Formula

$$p_i \propto \exp \left\{ - \left(\boxed{\text{Unary}_i} + \boxed{\sum_j \text{Pair}_{i,j} * p_j} \right) \right\}$$

Summation Convolution

$\text{Pair}_{i,j}$: Different Types of
Local and Global Filters

Deep Parsing Network



Convolution

Max Pooling

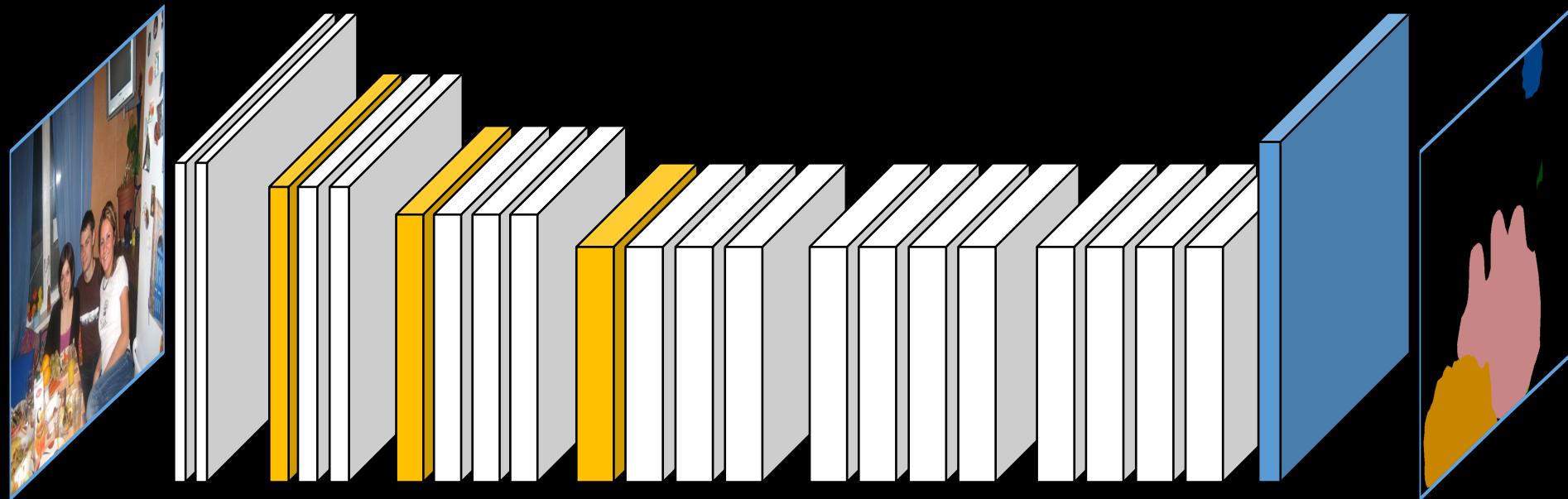
Deconvolution

Min Pooling

Local Convolution

Deep Parsing Network

Unary Term



Fine-tuned VGG-16 Network

Convolution

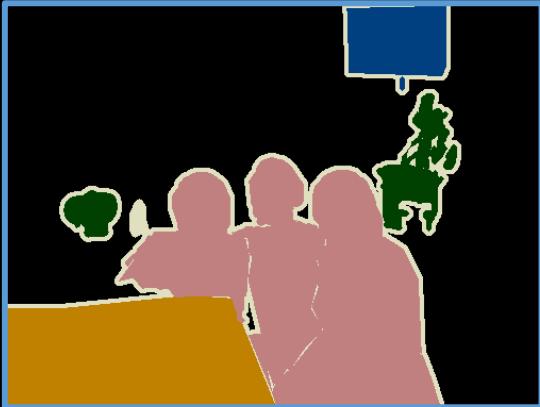
Max Pooling

Deconvolution

Deep Parsing Network



Original Image

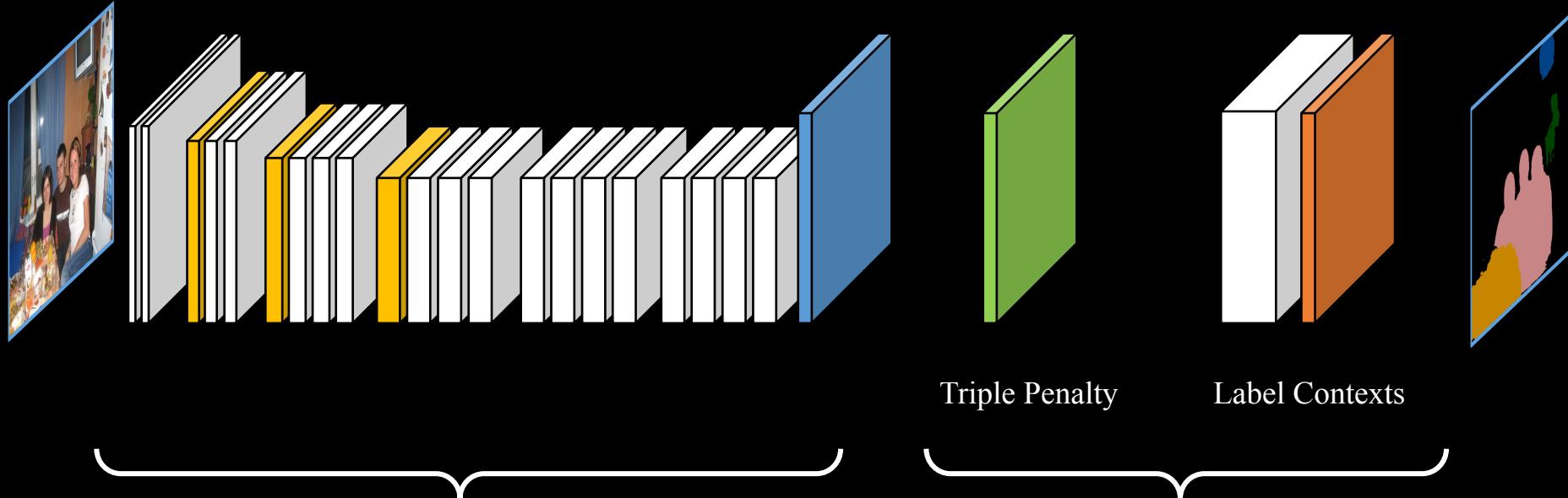


Ground Truth



Unary Term

Deep Parsing Network



Convolution

Max Pooling

Deconvolution

Min Pooling

Local Convolution

Deep Parsing Network

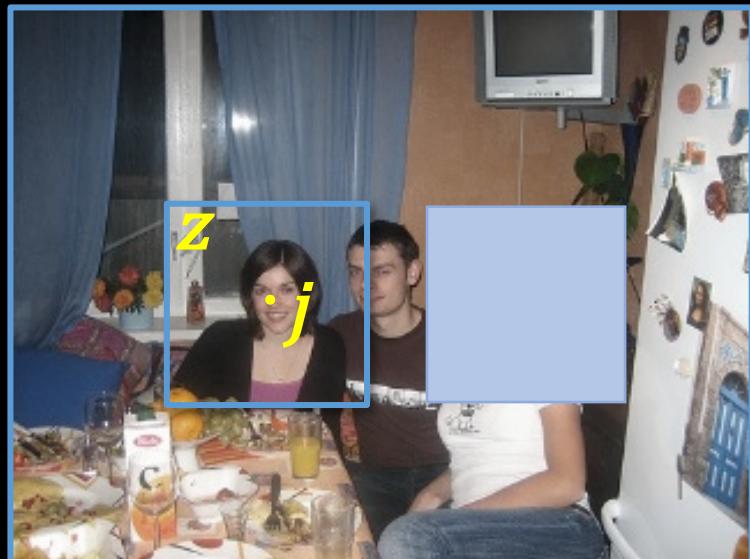
Triple Penalty

$$Pair = \sum_{i,j} \sum_k cost_k d_{i,j}(j) * \sum_z *diss(i,j; z) * p_z$$

Deep Parsing Network

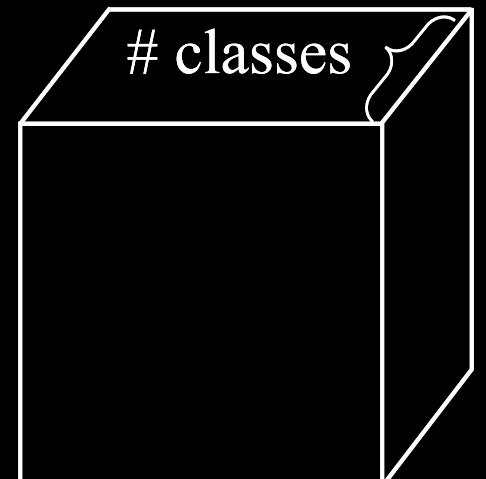
Triple Penalty

$$\sum_z diss(j; z) * p_z$$



$\cdot j$

Local Conv

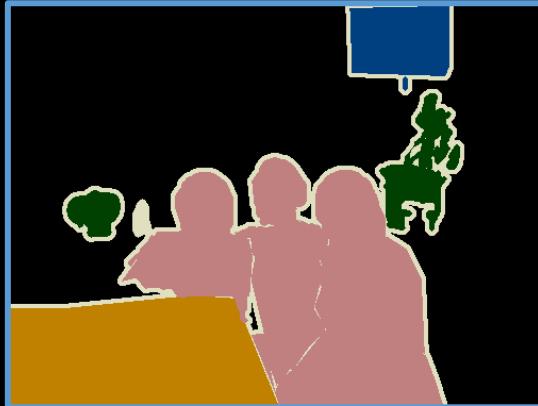


Unary Term

Deep Parsing Network



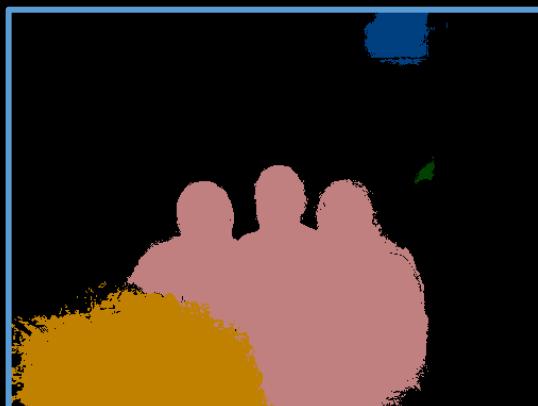
Original Image



Ground Truth

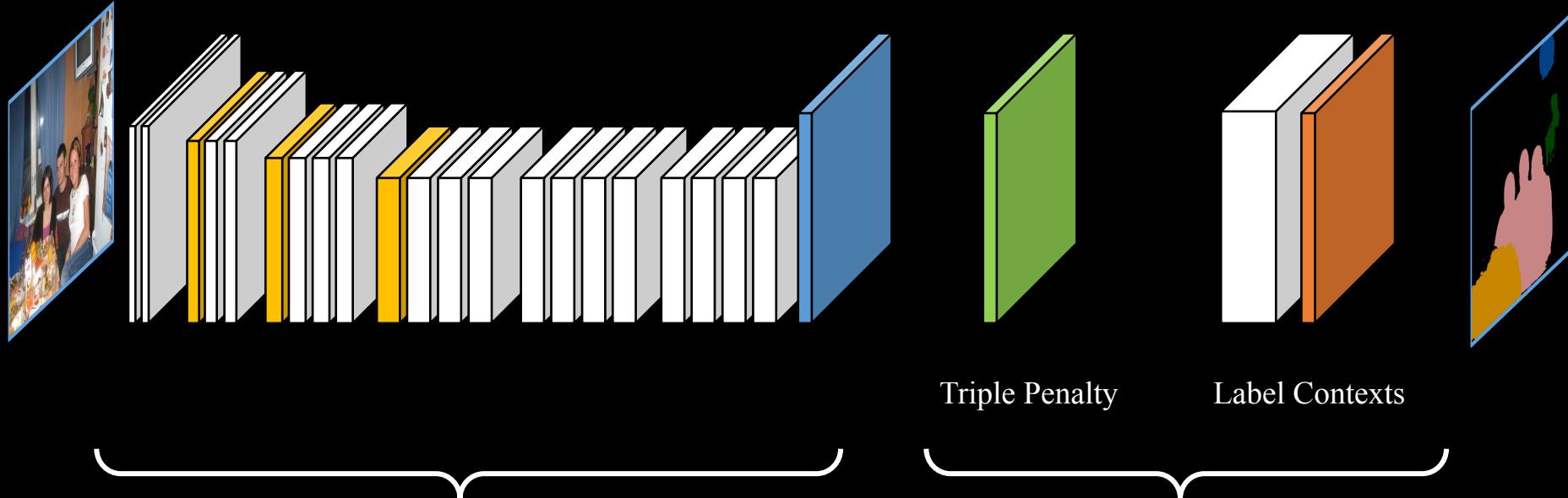


Unary Term



Triple Penalty

Deep Parsing Network



Convolution

Max Pooling

Deconvolution

Min Pooling

Local Convolution

Deep Parsing Network

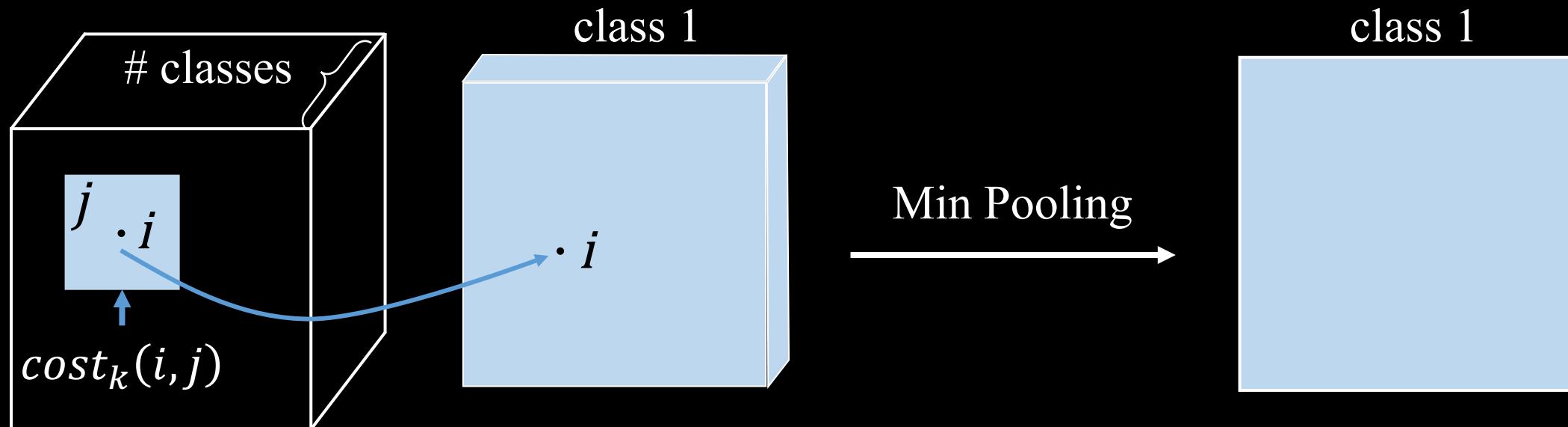
Mixture of Label Contexts

$$Pair = \sum_{i,j} \sum_k cost_k(i,j) * \sum_z dist(i,j; z) * p_z$$

Deep Parsing Network

Mixture of Label Contexts

Triple Penalty Result



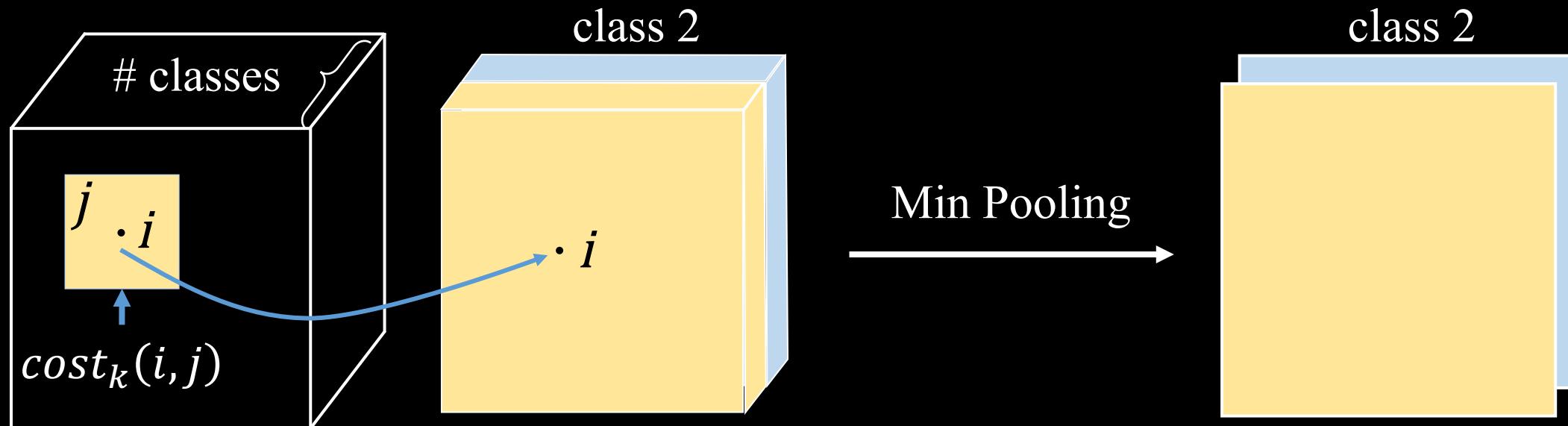
$$tri \quad \sum_j cost_k(i, j) * tri(j)$$

$$\sum_j \sum_k cost_k(i, j) * tri(j)$$

Deep Parsing Network

Mixture of Label Contexts

Triple Penalty Result



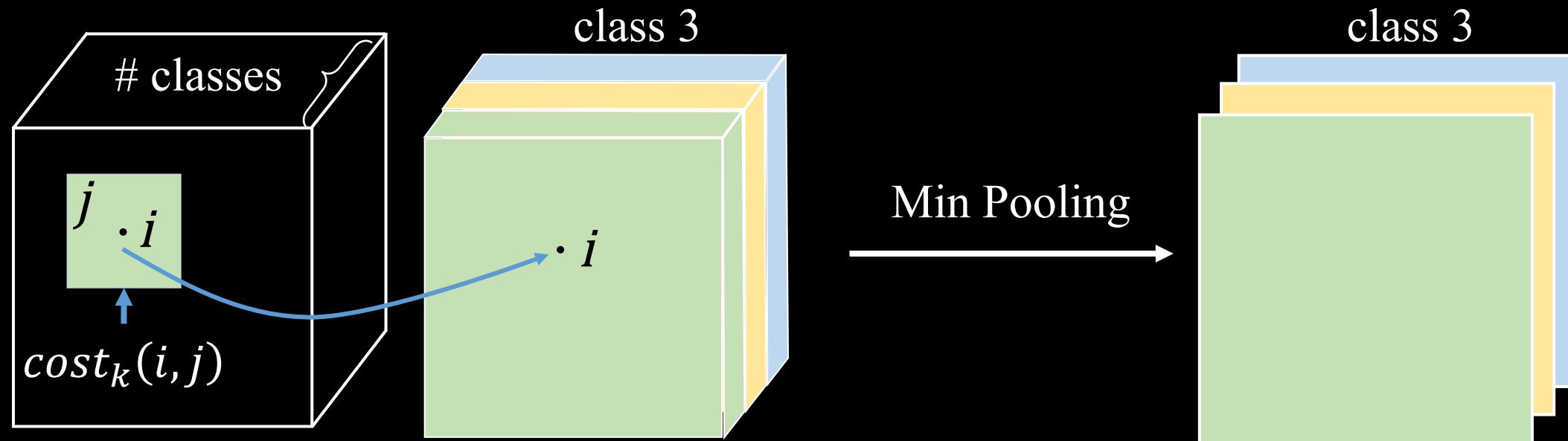
$$tri \sum_j cost_k(i, j) * tri(j)$$

$$\sum_j \sum_k cost_k(i, j) * tri(j)$$

Deep Parsing Network

Mixture of Label Contexts

Triple Penalty Result

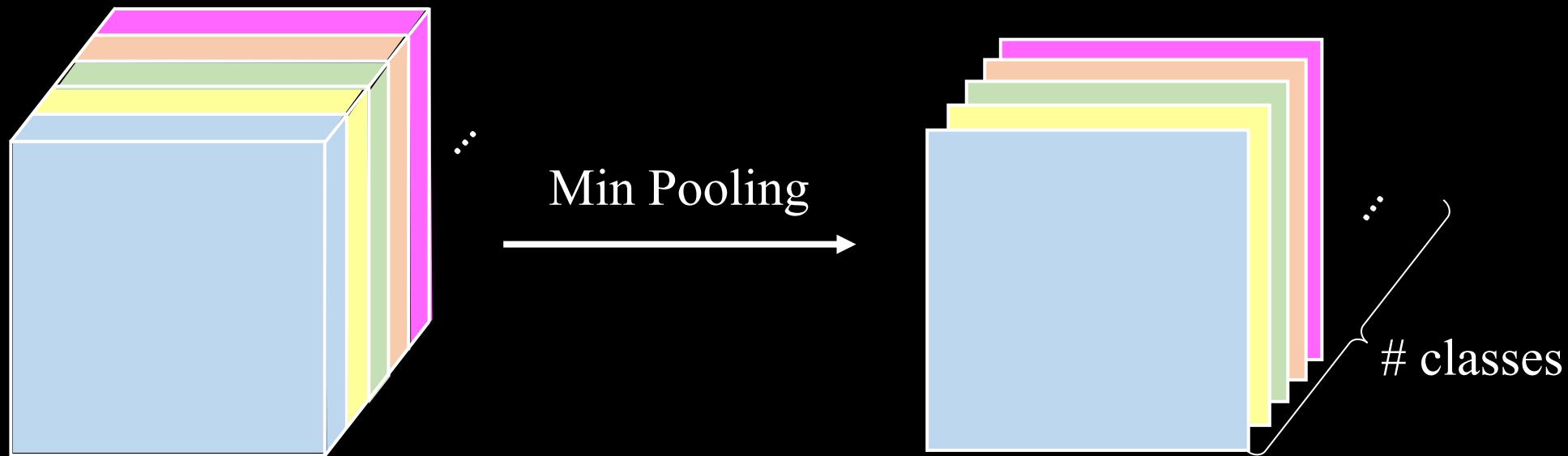


$$tri \quad \sum_j cost_k(i, j) * tri(j)$$

$$\sum_j \sum_k cost_k(i, j) * tri(j)$$

Deep Parsing Network

Mixture of Label Contexts

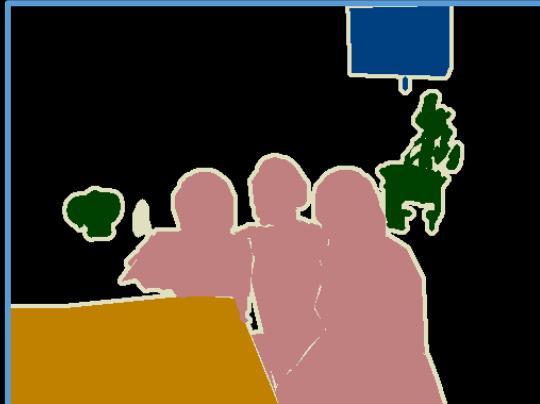


$$\sum_j \sum_k cost_k(i,j) * tri(j)$$

Deep Parsing Network



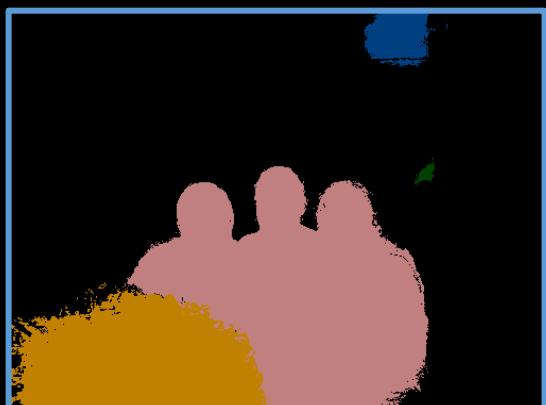
Original Image



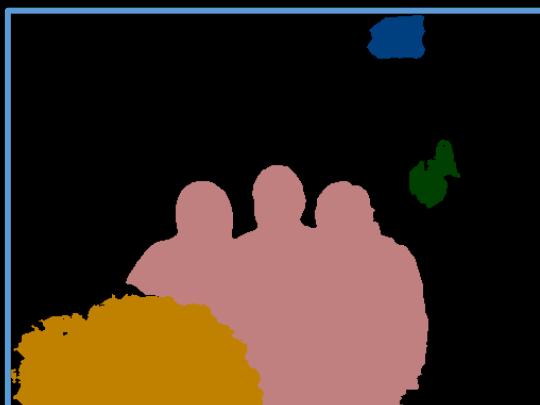
Ground Truth



Unary Term



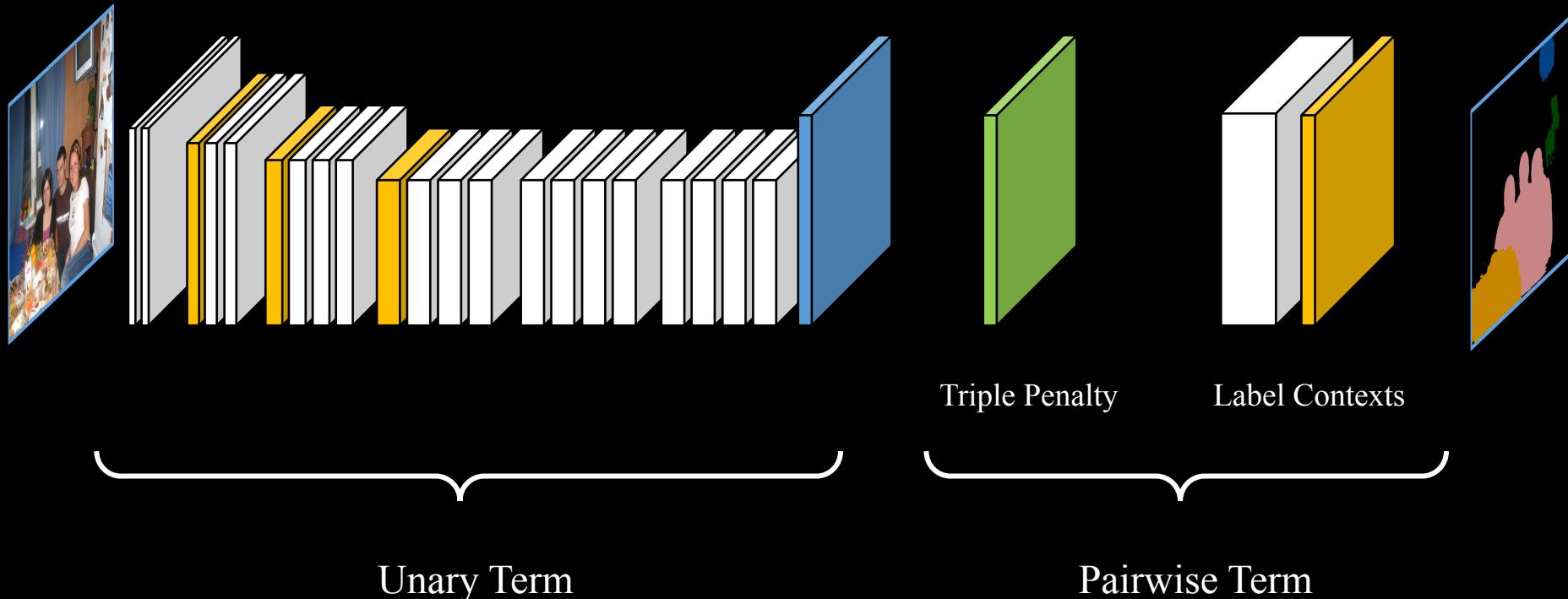
Triple Penalty



Label Contexts

Deep Parsing Network

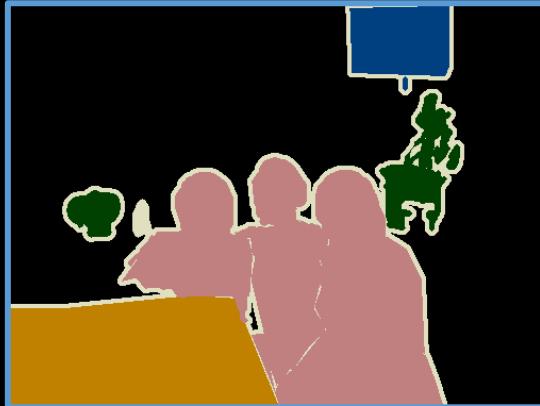
Joint Tuning



Deep Parsing Network



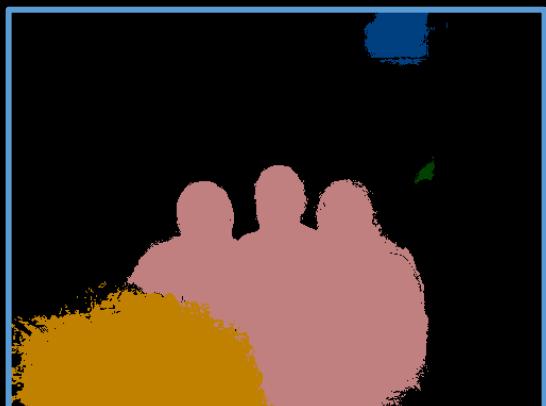
Original Image



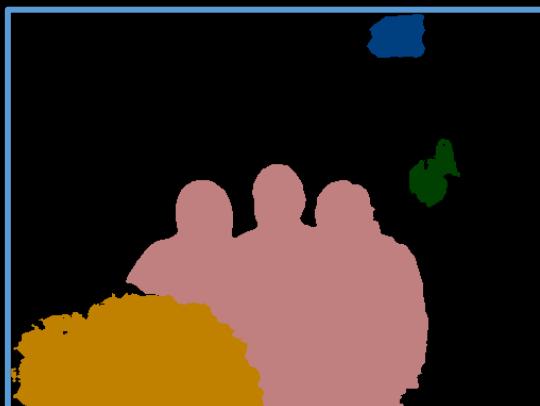
Ground Truth



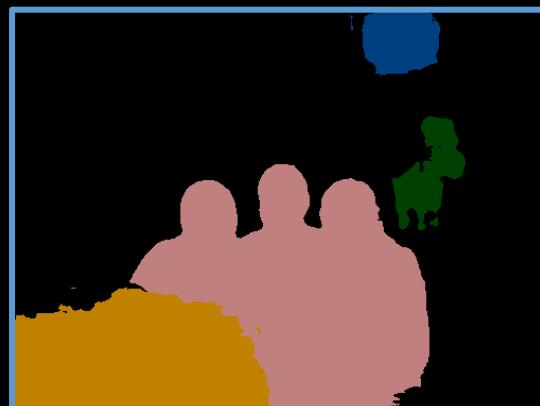
Unary Term



Triple Penalty



Label Contexts



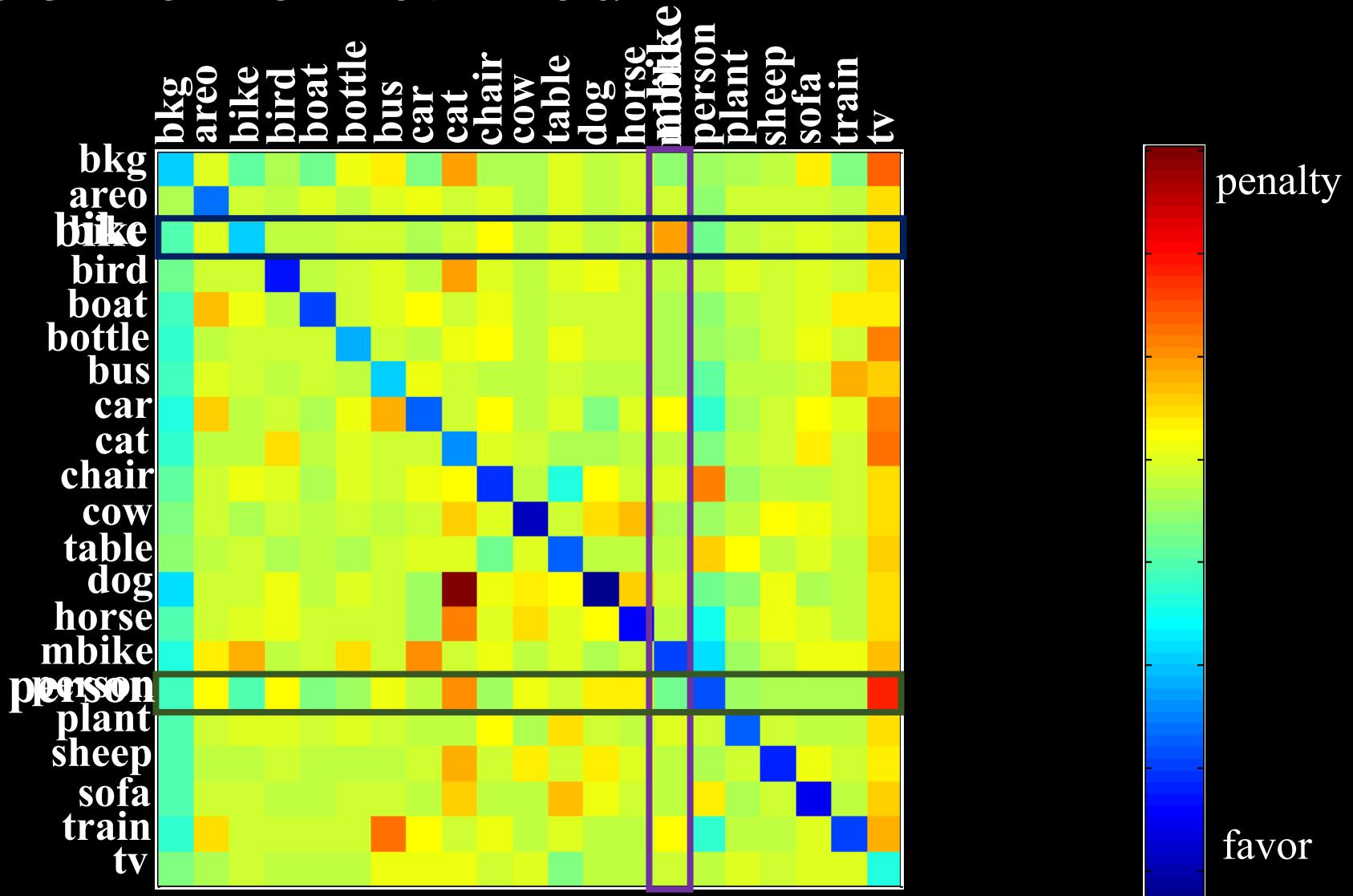
Joint Tuning

Overall Performance (Published Results)

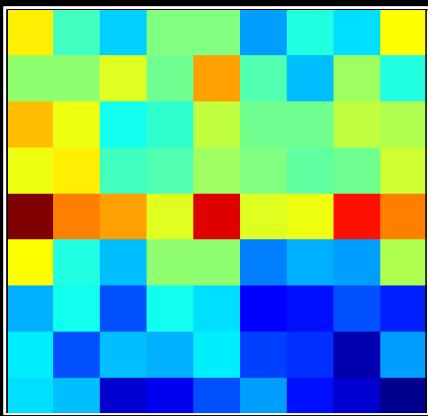
FCN	62.2
DeepLab [†]	73.9
CRFasRNN [†]	74.7
BoxSup [†]	75.2
DPN[†]	77.5

(PASCAL VOC 2012 Challenge test set)

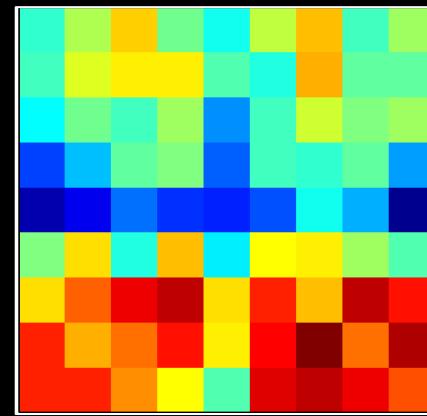
Label Contexts Learned



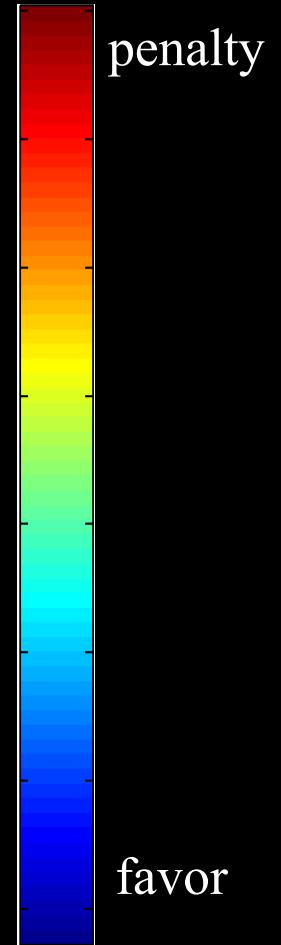
Label Contexts Learned



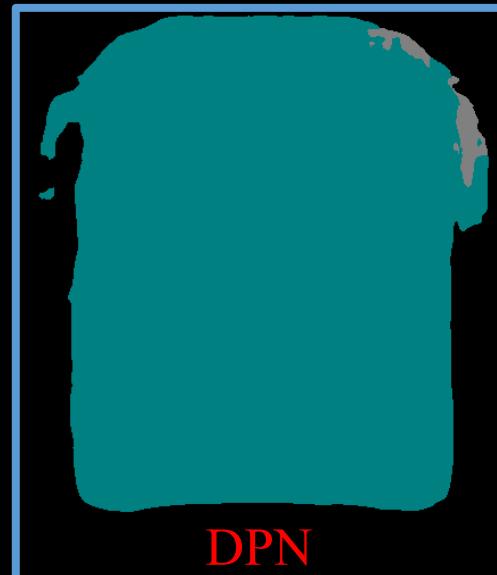
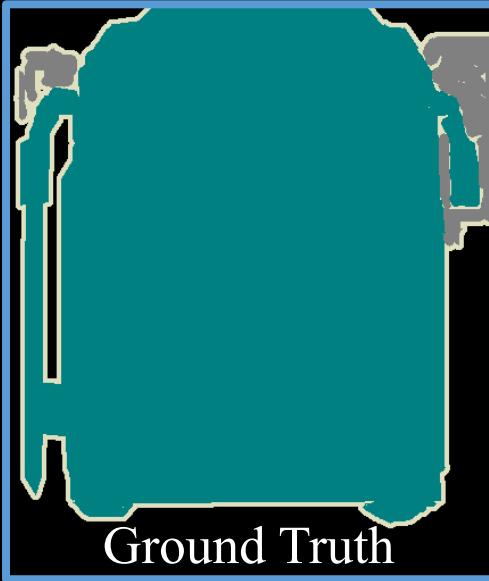
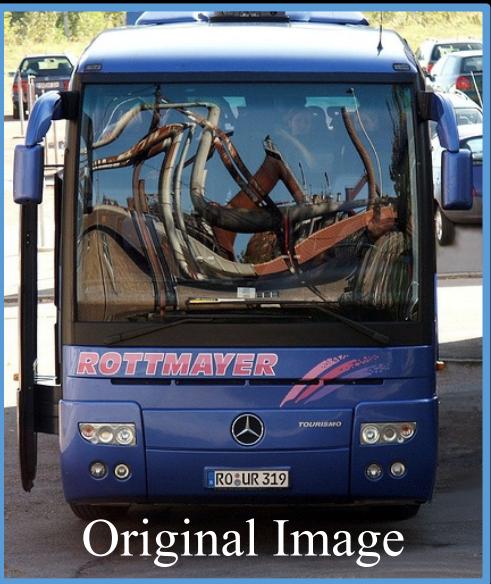
person : mbike



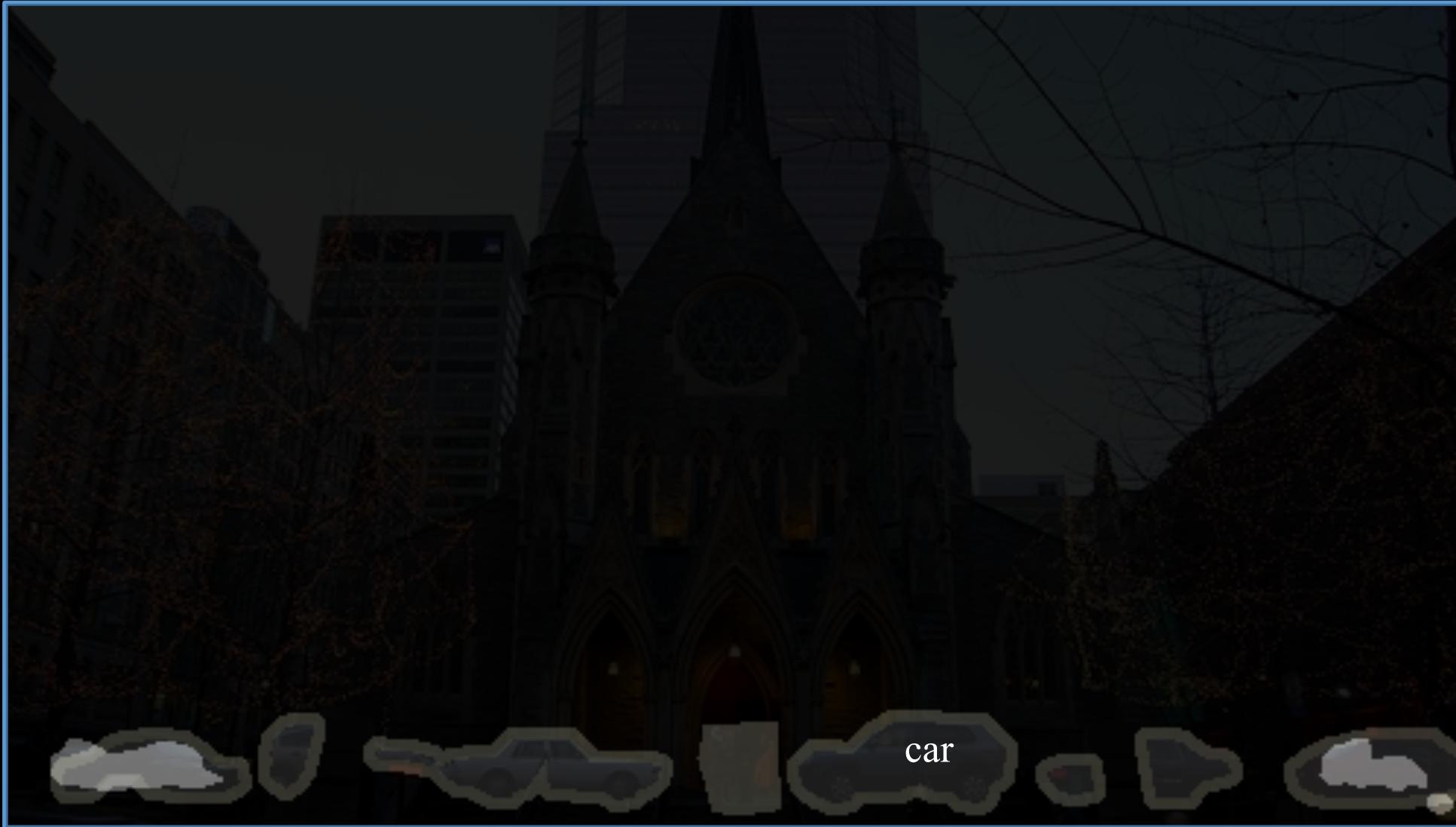
chair : person



Challenging Case



Failure Case



Cognitive Trust

Conclusions

- General framework of one-pass CNN to model high-order MRF
- Various types of pairwise terms are formulated as local and global filters
- High performance and easy to be speeded up

Thanks!

Semantic Image Segmentation via Deep Parsing Network

Project Page: <https://liuziwei7.github.io/projects/DPN>