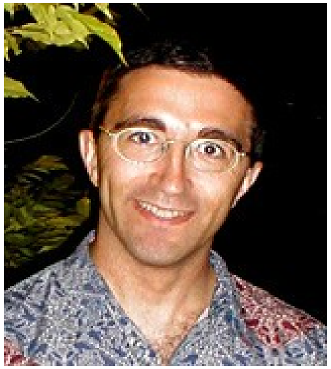
# Towards Urban Semantization

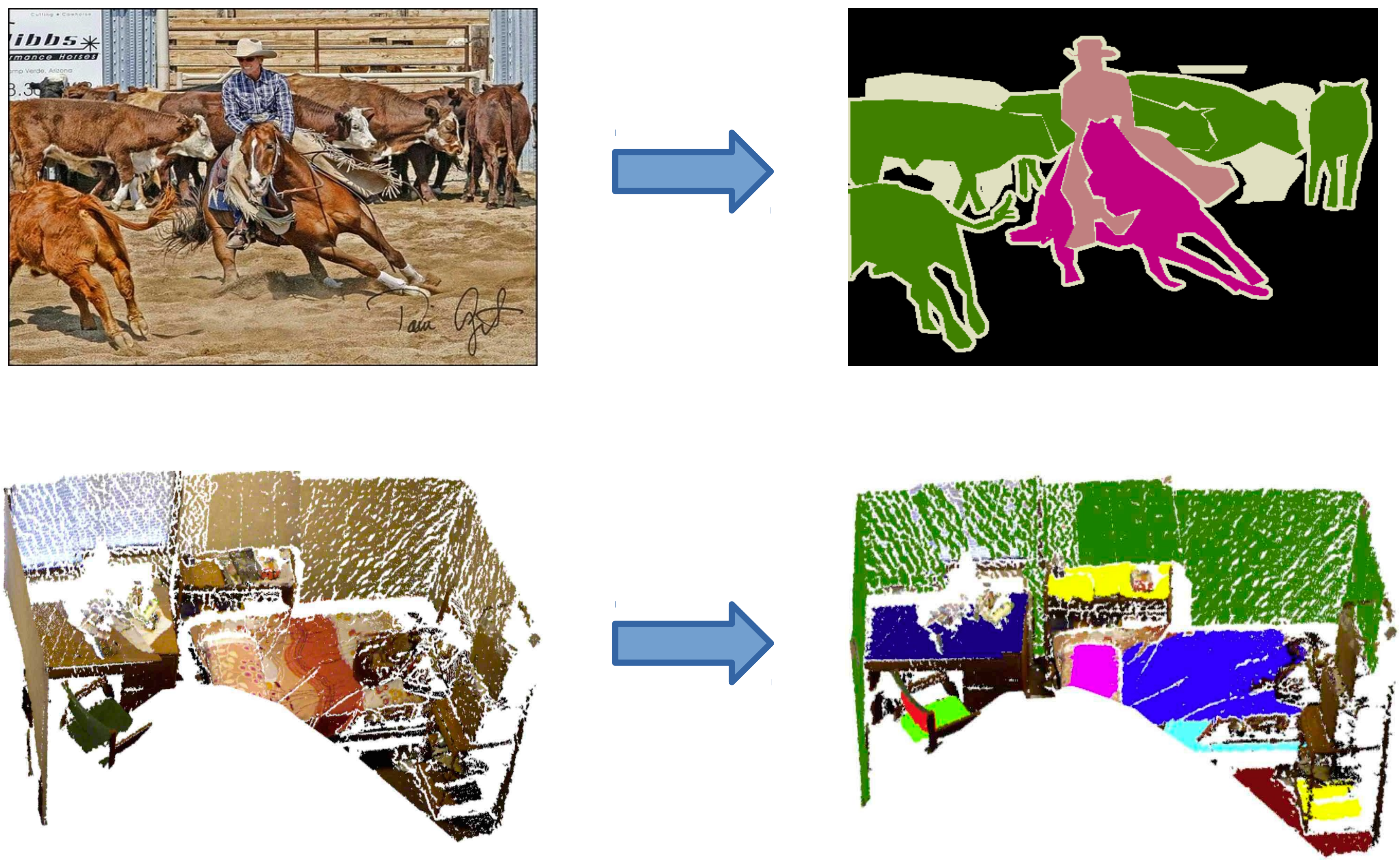
Shell Xu Hu, Guillaume Obozinski, Renaud Marlet,
Mathieu Aubry and Nikos Komodakis

# Schedule

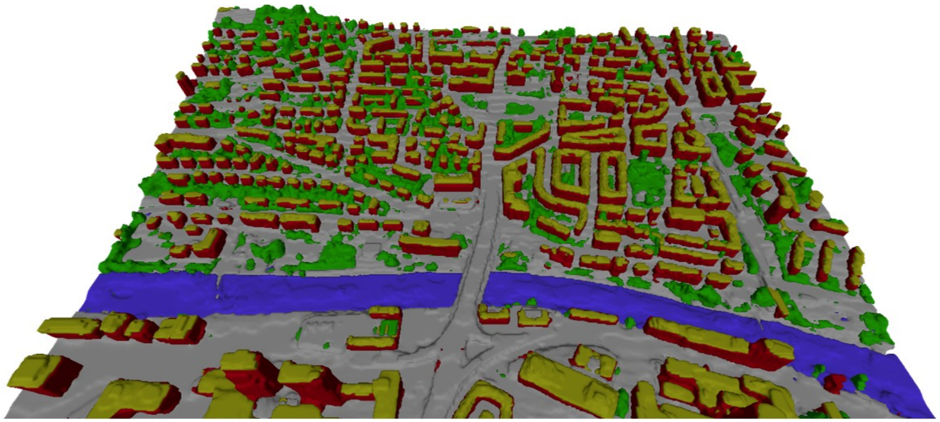
- The problem and applications.
- An introduction to discrete CRF.
- Faster inference and learning for CRF.
- Learning feature representations by CNN.
- Results and Demo.

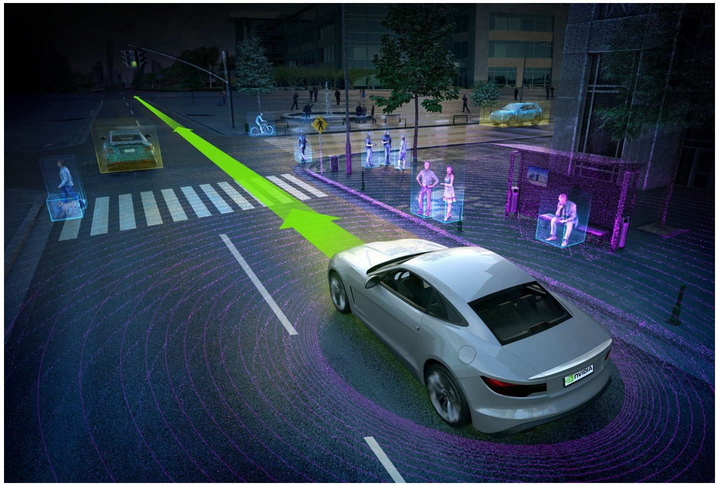
# Problem: Semantic Segmentation



# Applications

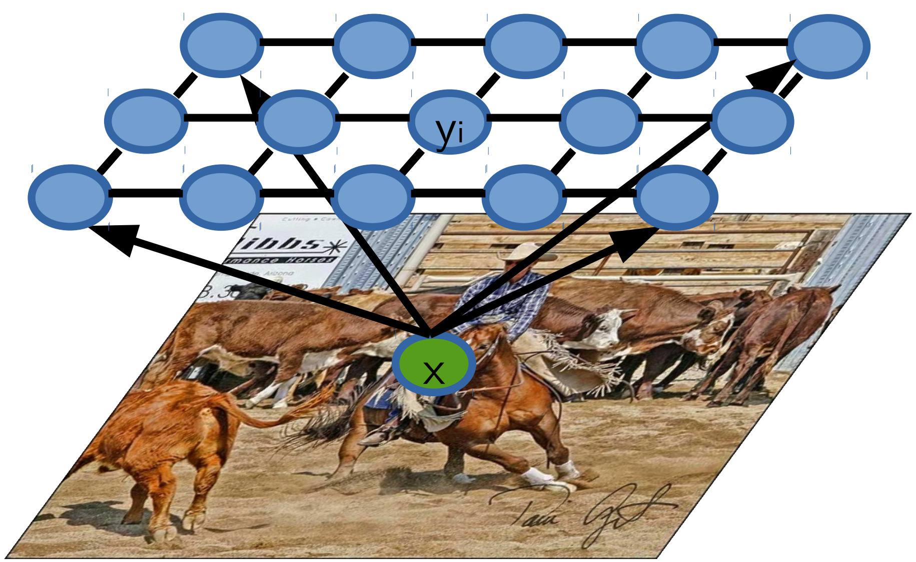
- Additional constraints for 3D reconstruction.
- Self-driving cars.
- 3D semantic maps.





# Discrete CRF

- Conditional random field is an undirected graphical model with discrete random variables.



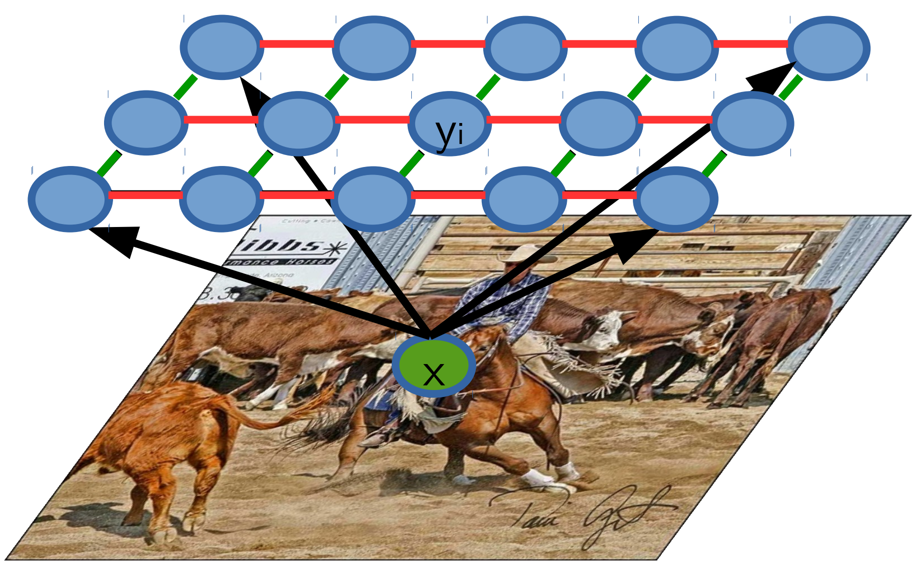

$y_i$ is a discrete random variable

x is a random vector in a high dimensional image manifold

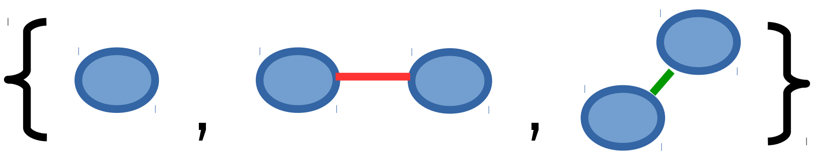
# Discrete CRF

- Definition: A conditional distribution

$$p(y|x;w) = \frac{1}{Z(x;w)} \exp\left( \sum_{a \in \mathcal{A}} \sum_{c \in G_a} \theta_c(y_c, x; w_a) \right)$$



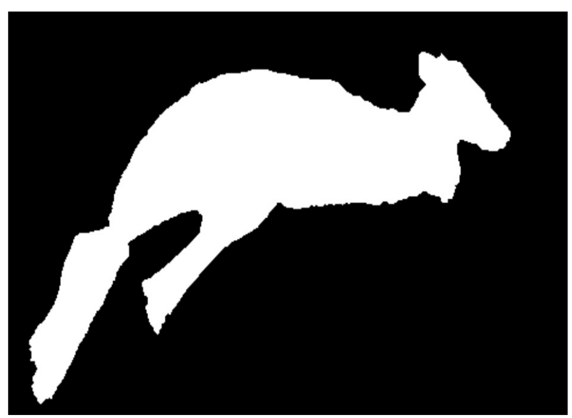

Clique types:

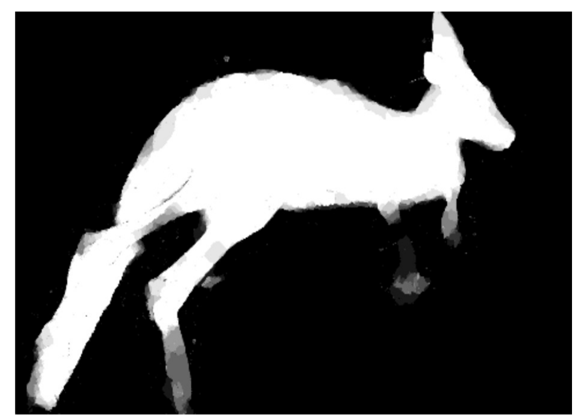


# Inference of CRF

- Maximum a posterior inference: $\text{argmax}_y p(y|x)$
- Probabilistic/marginal inference: $Z(x)$ $p(y_i|x)$


image


MAP prediction


marginals

# Inference of CRF

- Maximum a posterior inference: $\operatorname{argmax}_y p(y|x)$
- Probabilistic/marginal inference: $Z(x)$ $p(y_i|x)$


# Parameter Estimation in CRF

- Inverse problem: Given samples of (x, y), to estimate model parameters w.
  - Maximum likelihood estimation

    $$\max_w \mathbb{E}_{\text{data}}[p(Y|X;w)]$$

    Issue: computing gradients need marginal inference
  - Max-margin learning

    $$\min_w \mathbb{E}_{\text{data}}[\max_y \big(\theta_w(y,X) + \ell(Y,y)\big) - \theta_w(Y,X)]$$

    Issue: computing gradients need MAP inference

# Inference-Free Parameter Estimation

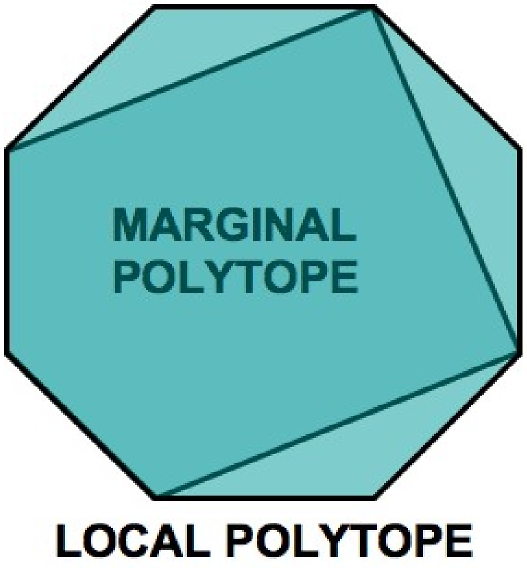
- Can we learn parameters without performing inference at each iteration?
- Yes! Working on **dual**. For MLE, we assume
  - linear function w.r.t. w: $\theta_w(y, x; w) = \psi(y, x)^T w$

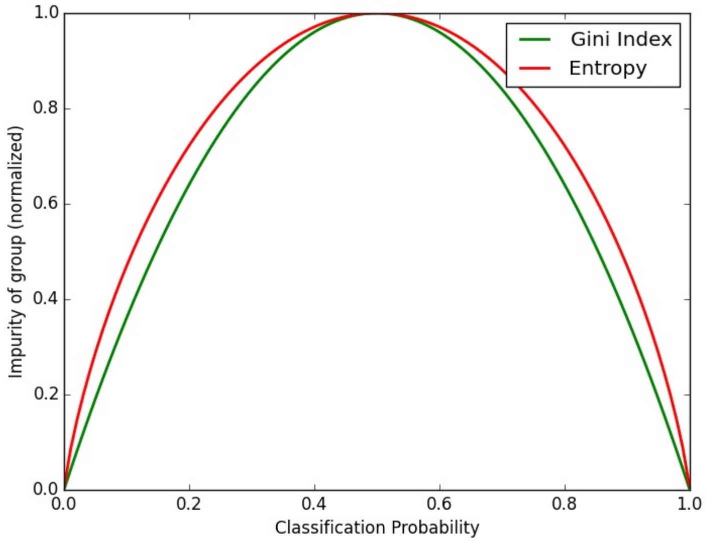
    then $\theta(w) = [\theta_w(y, x; w) - \theta_w(y^*, x; w)]_y = \Psi^T w$
  - $\Phi(\theta(w)) = \log \int_y \exp(\langle \theta(w), y \rangle) = \max_{\mu \in \mathcal{M}} \mu^T \theta(w) + H(\mu)$

    is well approximated by its **variational relaxation**.

# Variational Relaxation

$$\max_{\mu \in \boxed{\mathcal{M}}} \mu^T \theta + \boxed{H_{\text{Shannon}}}(\mu) \quad \Rightarrow \quad \max_{\mu \in \boxed{\mathcal{L}}} \mu^T \theta + \boxed{H_{\text{Gini}}}(\mu)$$



# Relaxed CRF Learning

- It's equivalent to work on the dual.

Primal:

$$\min_{w} \max_{\mu \in \mathcal{L}} \mu^T \theta(w) + H_{\text{Gini}}(\mu) + \frac{\lambda}{2} \|w\|_2^2$$

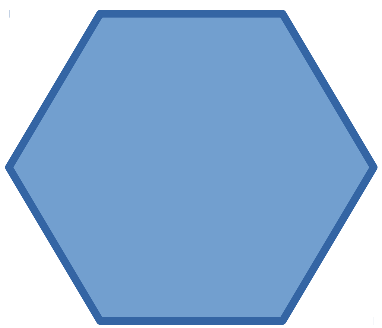
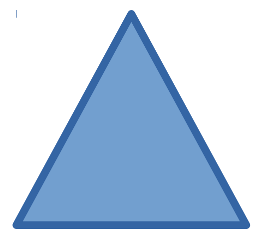
Dual:

$$\max_{\mu \in \mathcal{L}} H(\mu) - \frac{1}{2\lambda} \|\Psi \mu\|_2^2$$

N.B.: consider all graphs as a single graph with multiple connected components.

# Relaxed CRF Learning

- The local polytope can be decomposed as a product of simplices and hyperplanes.

 $=\Big\{$  $\Big\} + \Big\{ \; A\mu = 0 \; \Big\}$

- The dual augmented Lagrangian factor over cliques:

$$\min_{\xi} \max_{\mu \in \Delta^{\#\mathrm{cliques}}} H(\mu) - \frac{1}{2\lambda}\|\Psi\mu\|_2^2 + \langle \xi, A\mu \rangle - \frac{1}{2\rho}\|A\mu\|_2^2$$

# Relaxed CRF Learning with Block Proximal Methods

- The relaxed CRF learning can be solved by proximal block coordinate method of multipliers.

| Method | MLE / Max-Margin | Primal / Dual | Convergence | Inference Oracle |
|---|---|---|---|---|
| Blend. (Meshi 10) | Max-Margin | Primal | O(1/eps) | Graph-wise MAP (10 iters) |
| Blend. (Hazan 10) | MLE | Primal | O(1/eps) | Graph-wise Marg. (10 iters) |
| BCFW (Lacoste-Julien 12) | Max-Margin | Dual | O(1/eps) | Graph-wise MAP |
| BCFW (Tang 16) | MLE | Dual | O(1/eps) | Graph-wise Marg. |
| BCFW (Meshi 15) | Max-Margin | Dual | O(1/eps) | Clique-wise MAP |
| Prox-BCMM / Prox-SDCA (ours) | MLE | Dual | O(log 1/eps) | Clique-wise Marg. |

# Relaxed CRF Learning with Block Proximal Methods

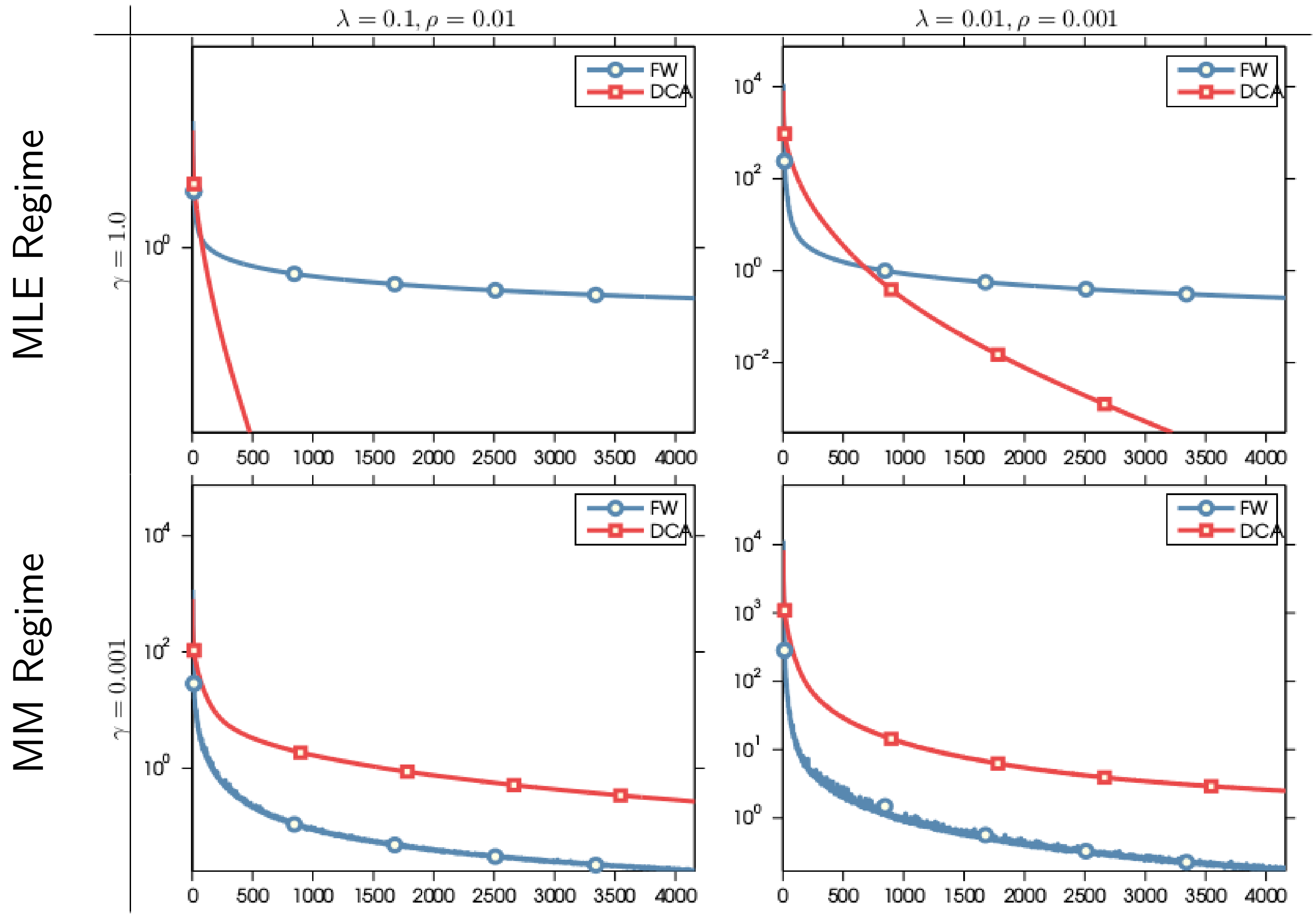



Figure 1: Semantic segmentation: duality gap (second).

# Relaxed CRF Learning with Block Proximal Methods

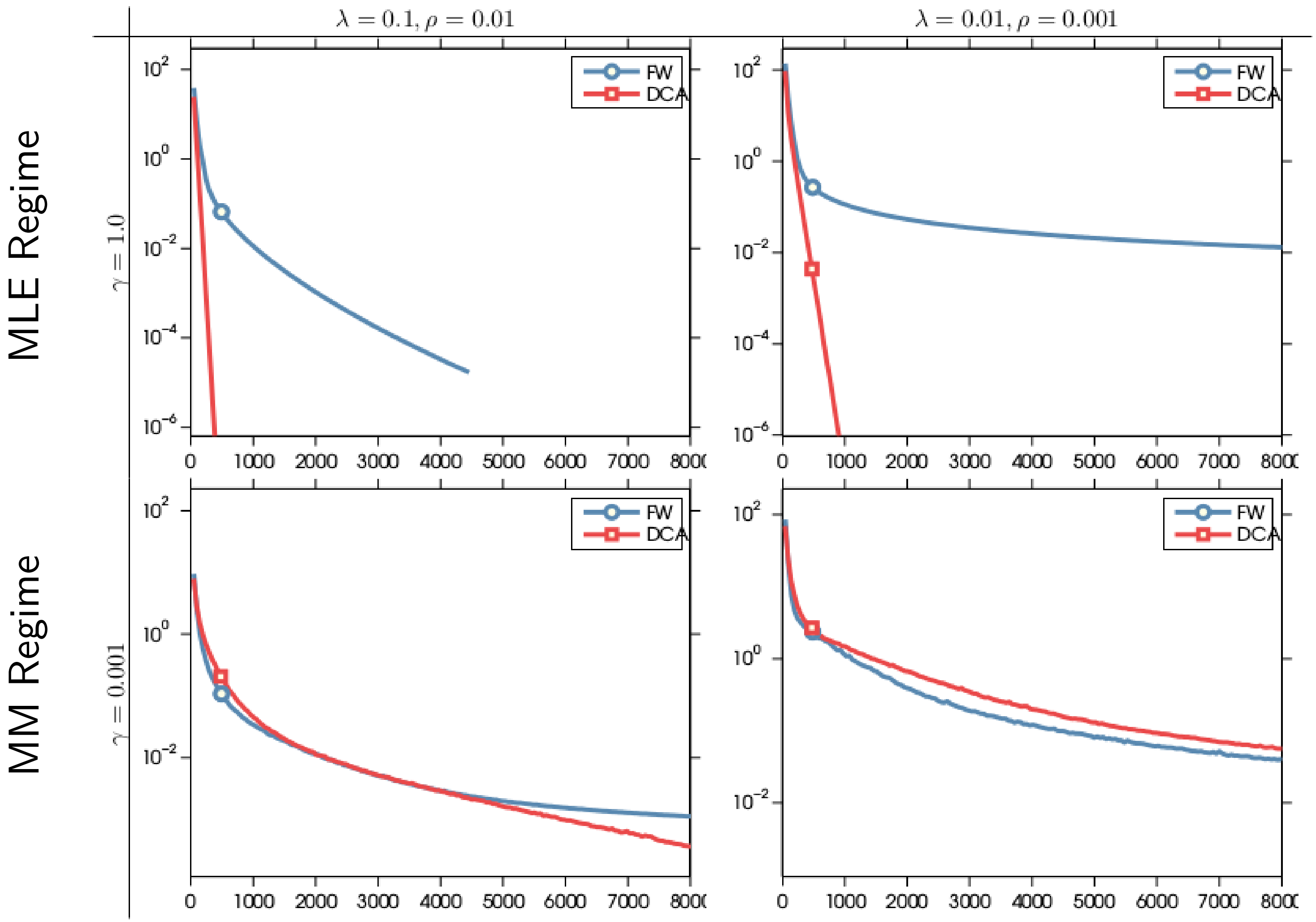



Figure 3: multilabel: duality gap (second).

# Relaxed CRF Learning with Block Proximal Methods

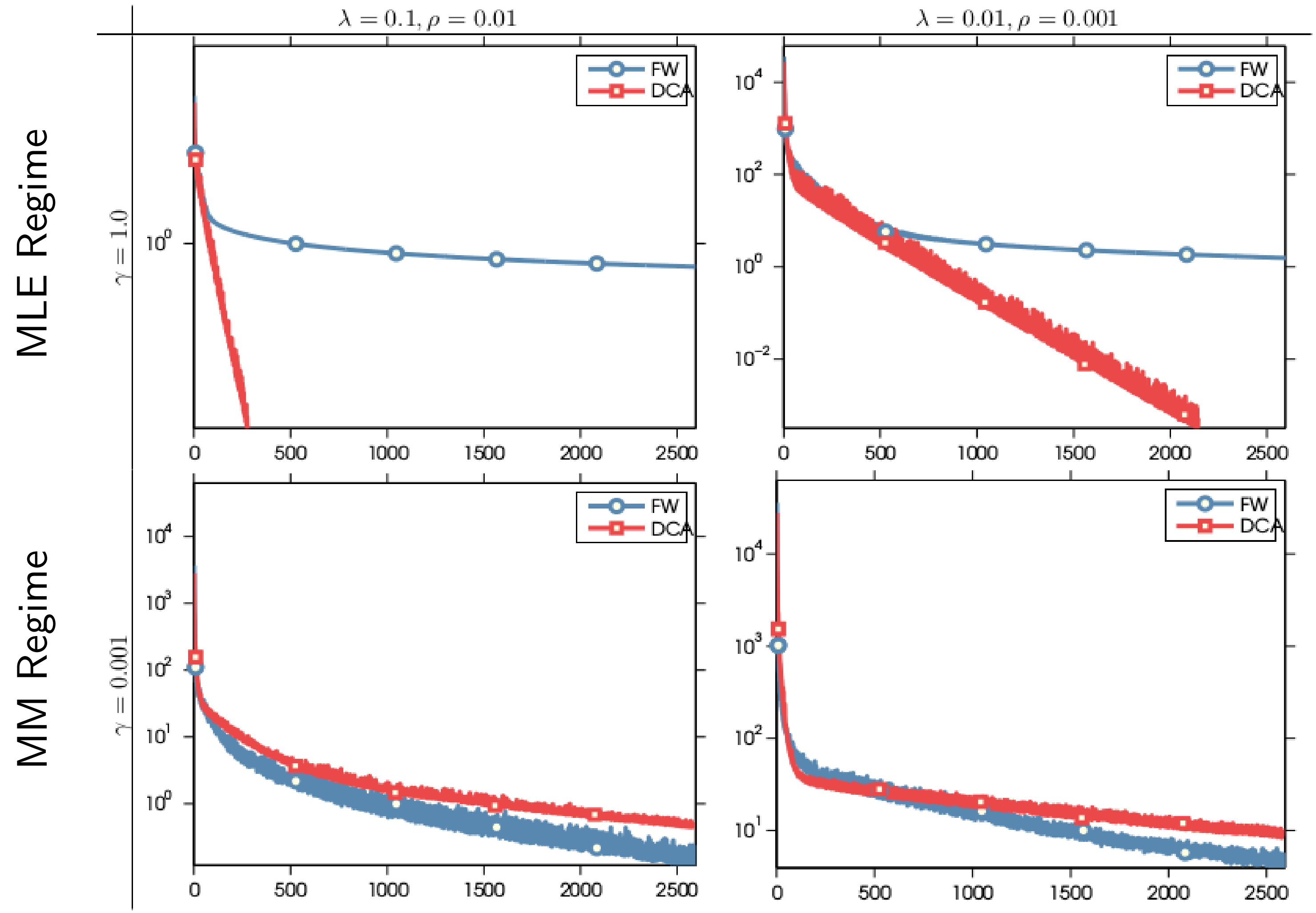


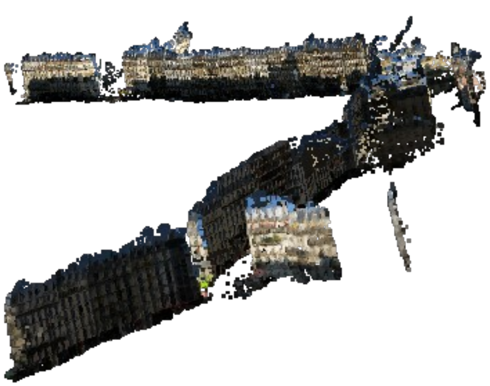

Figure 5: GMM Potts: duality gap (second).

# Relaxed CRF Learning with Block Proximal Methods

- Take-home messages:
  - If inference is expensive, try relaxed CRF.
  - If your problem cares about marginals, use MLE with Prox-BCMM;
  - If MAP inference is the goal, use max-margin with block-coordinte Frank-Wolfe algorithm.

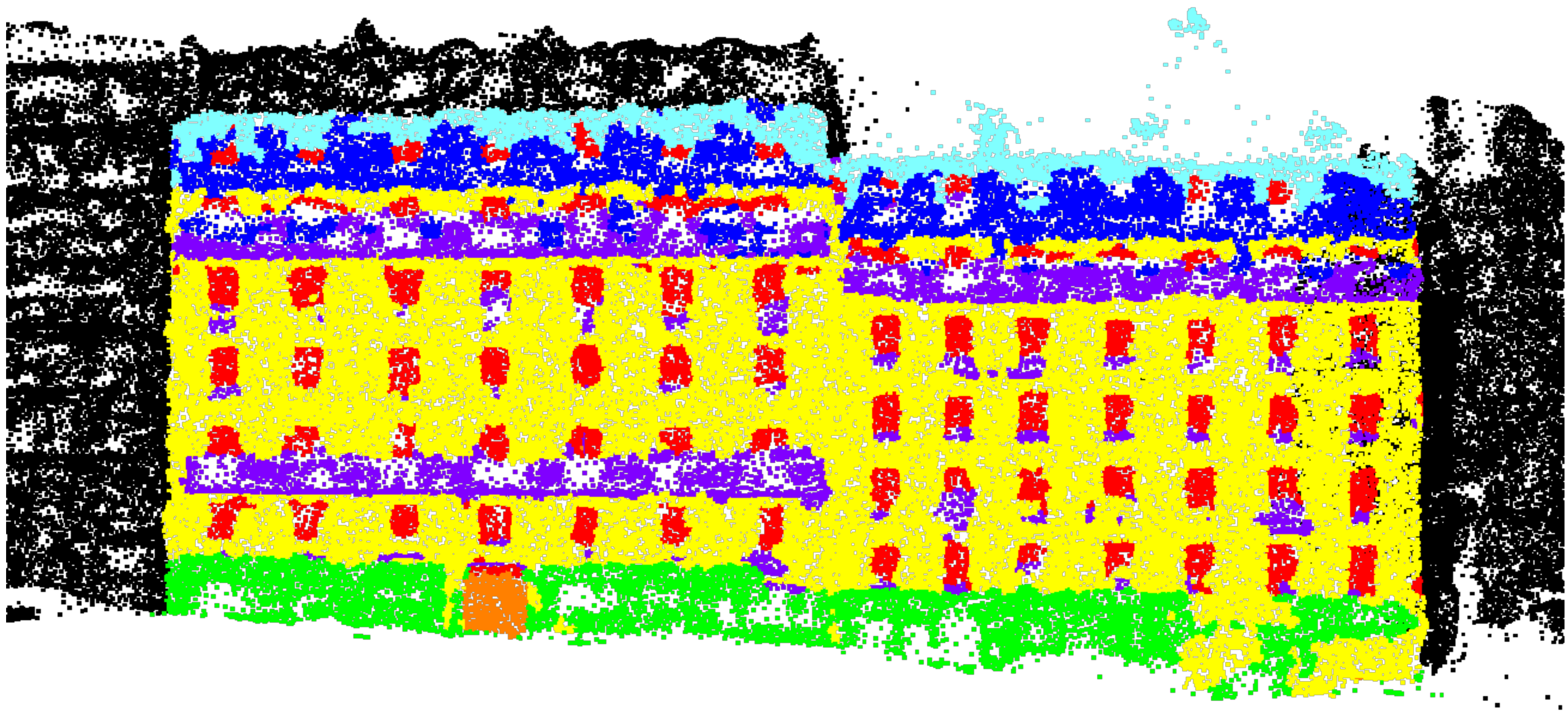
# Experiments on Rue-Monge Dataset

- 290196 points for training, 276529 points for testing.
- 7 classes: window, wall, balcony, door, roof, sky and shop.
- Features: RGB + Normal + Height + Depth + Spin image.



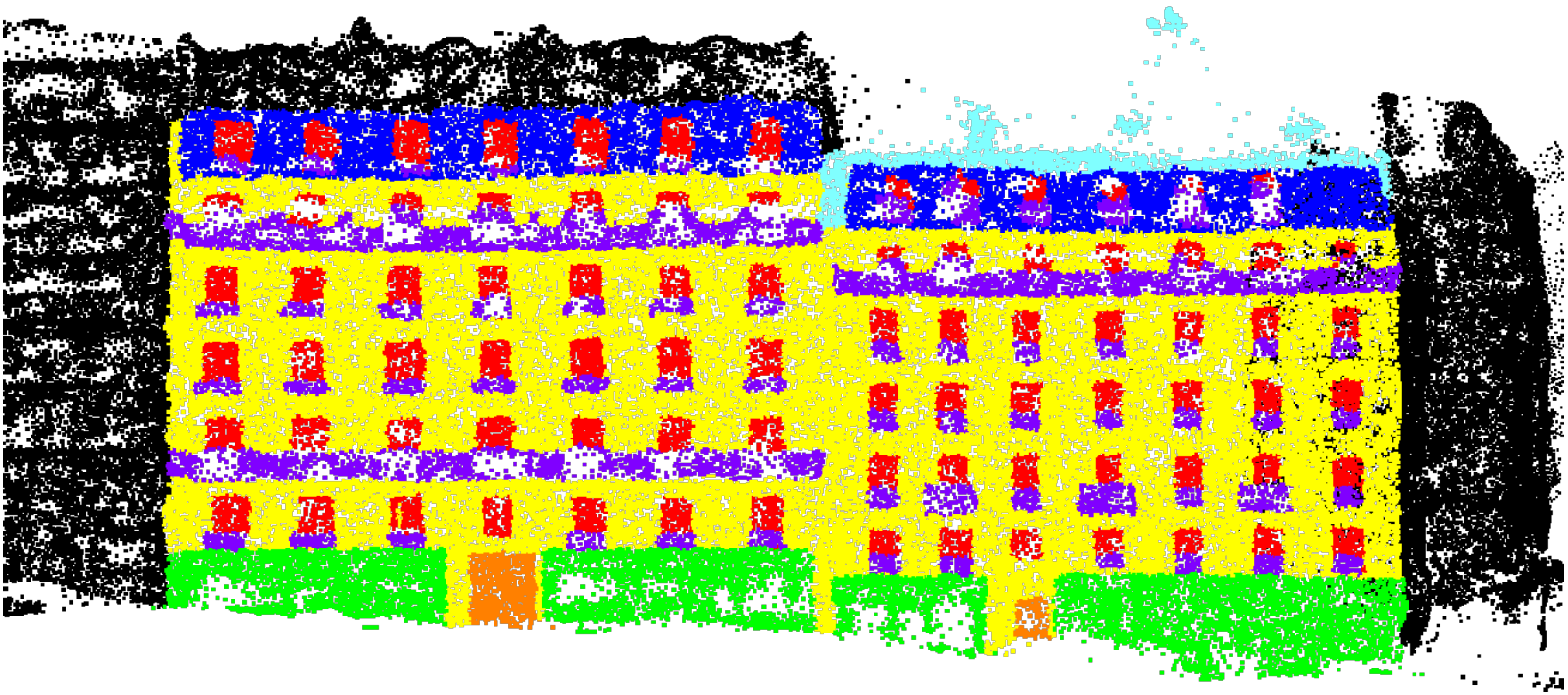
# Experiments on Rue-Monge Dataset

MAP prediction IoU: 59.2%

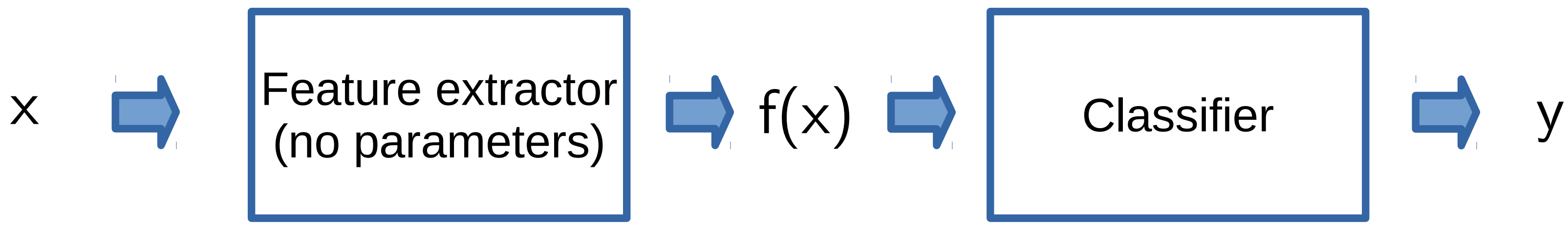


# Experiments on Rue-Monge Dataset

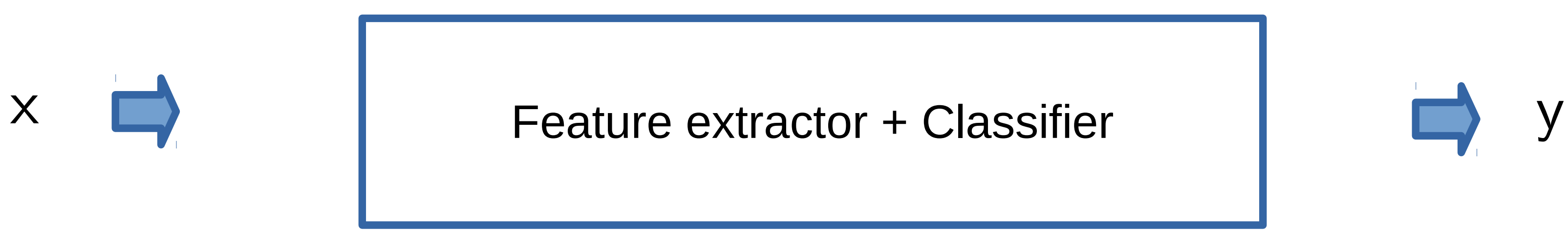
Ground truth



# Learning Feature Representations: A Deep Learning Approach



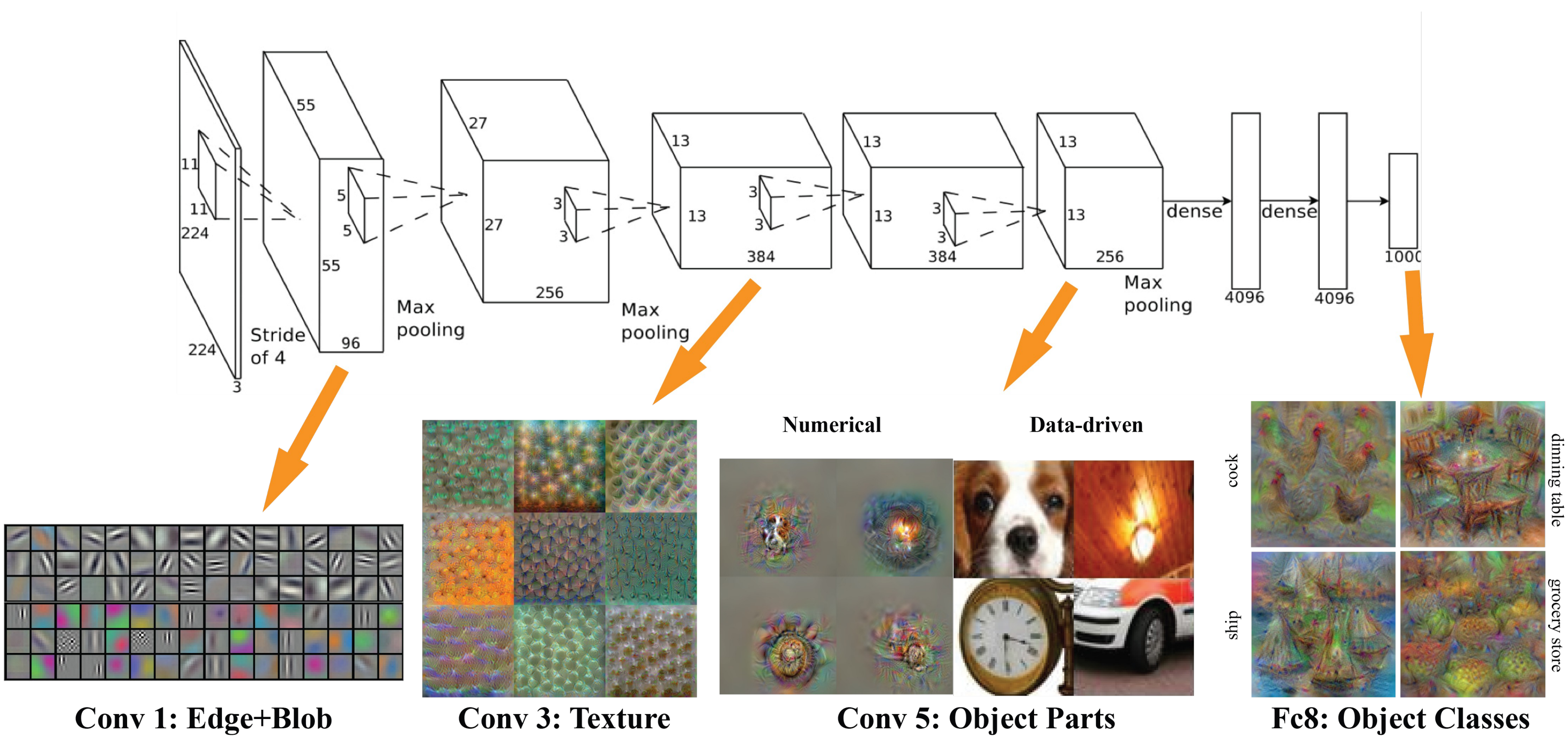

Classical machine learning




Deep machine learning

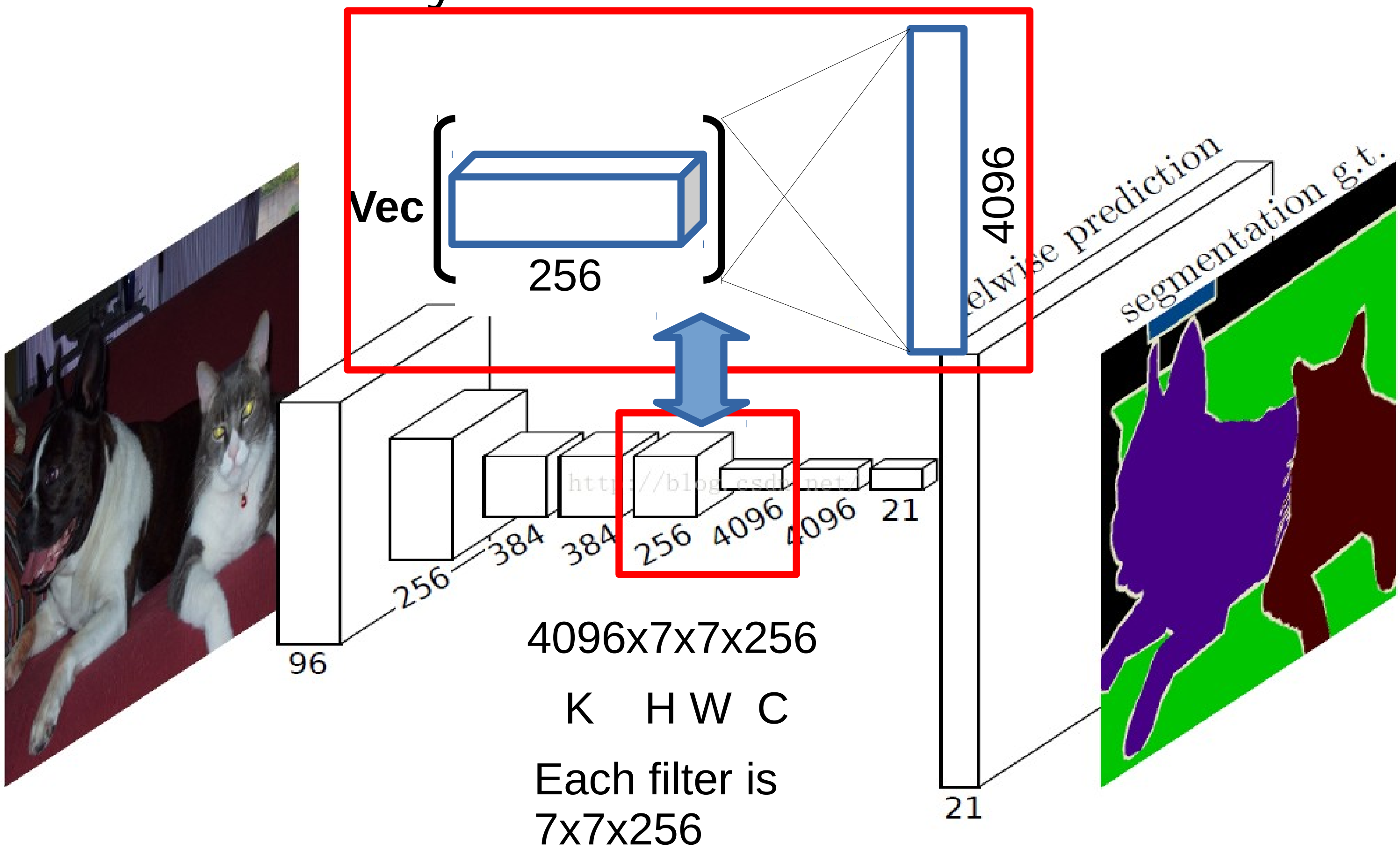
# Convolutional Neural Networks (CNN)

- A case study: VGG16-Net



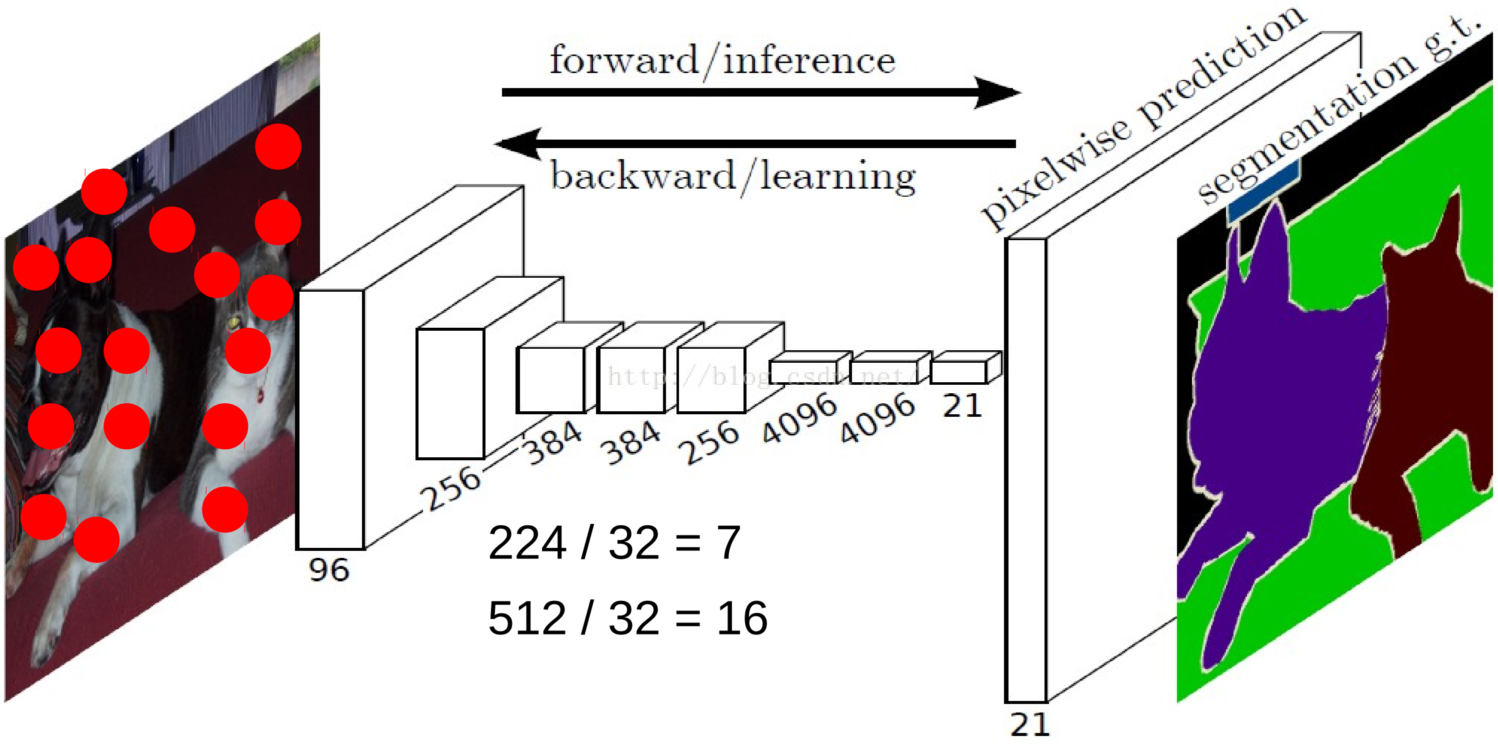
# Fully Convolutional Networks (FCN)

- Trick: Treat dense connection layers as convolutional layers.



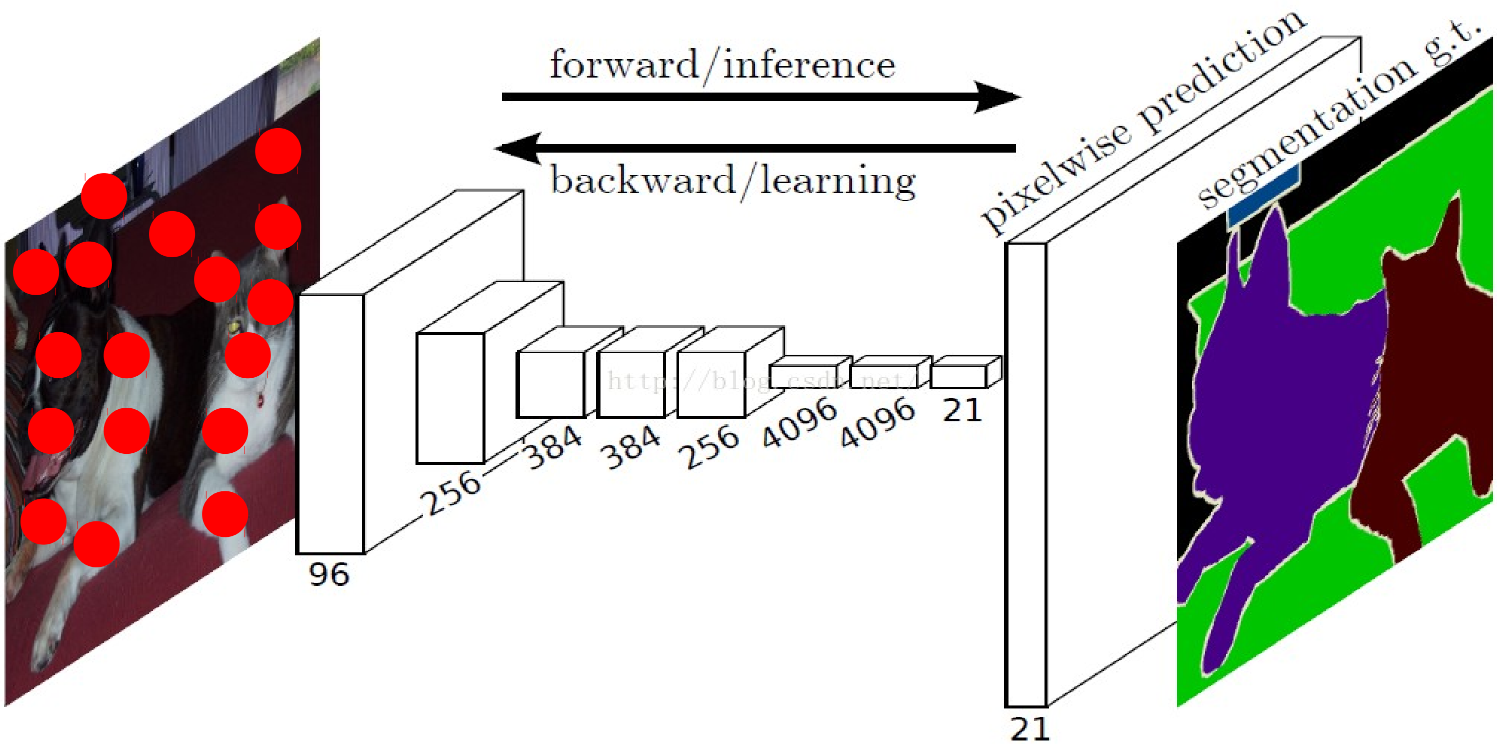
# FCN

- Computation sharing: If input image is 512x512, the output is 10x10 (due to 5 max poolings, 16-7+1). It classifies 100 patches in a single feedforward pass.



# FCN

- The 100 patches are chosen by max poolings, which give high activations. Pixelwise prediction is obtained by upsampling (e.g. bilinear interpolation or deconvolution).

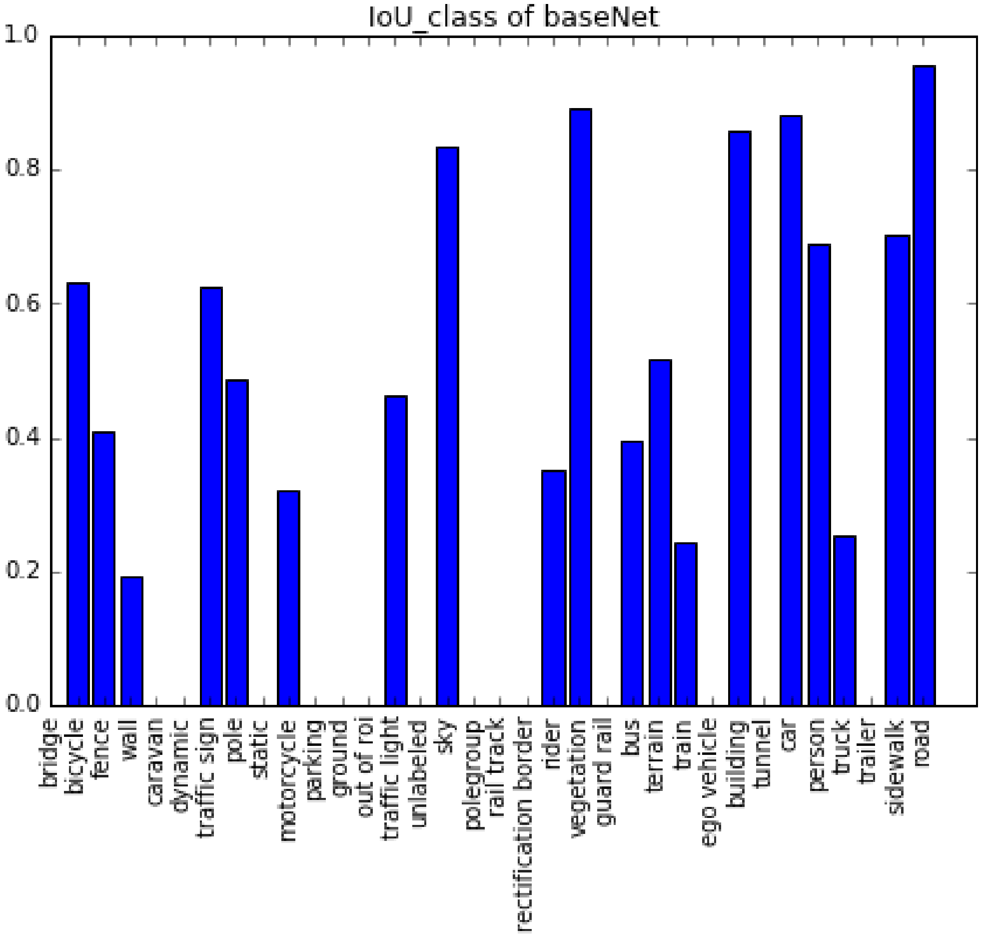


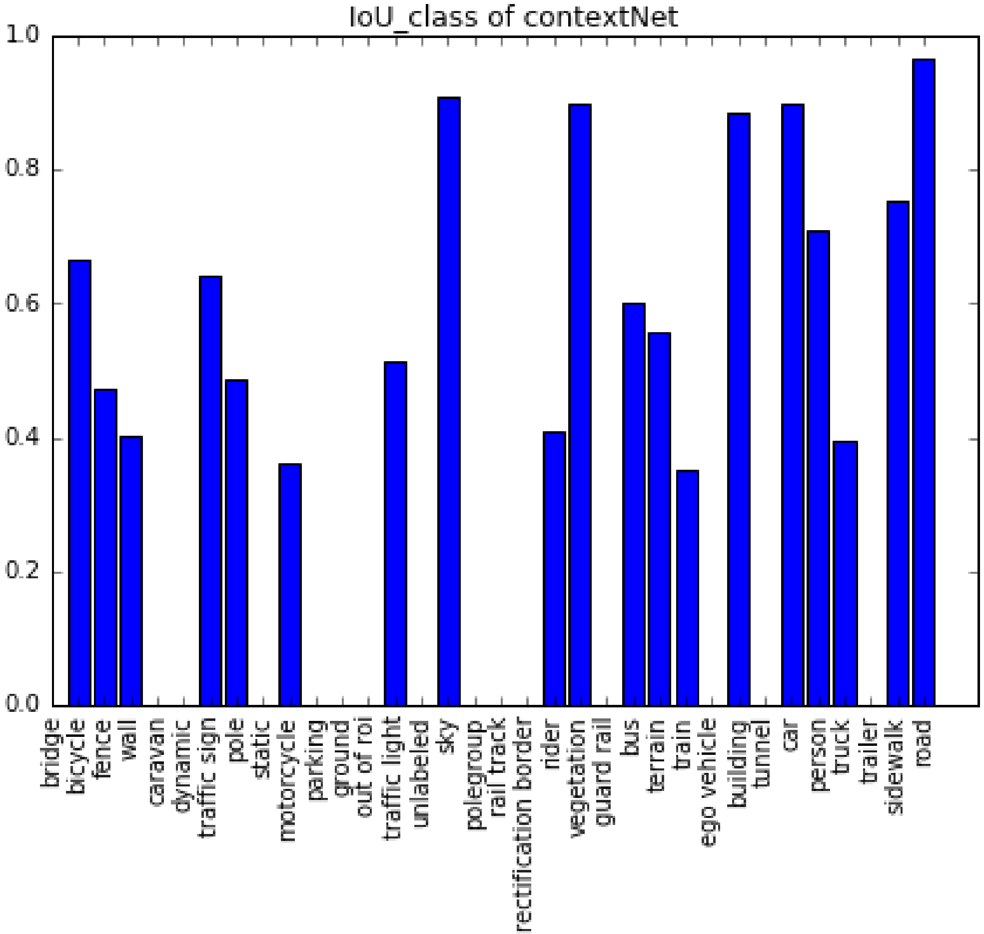
# Results on CityScapes Dataset

- Street images from 50 cities. 19 classes involved. 2975/500/1525 images for train/val/test.
- Baseline: FCN8s
- Our: FCN8s + additional convolutions on top of the pixelwise prediction to capture context information. A simple experiment to test our higher order CRF model.

# Results on CityScapes Dataset

- IoU: FCN8s 56.3%; ContextNet 62.5%



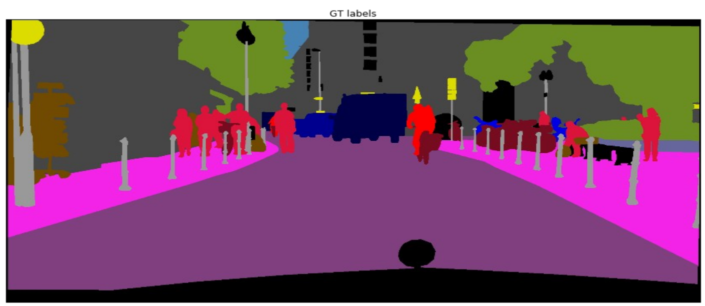



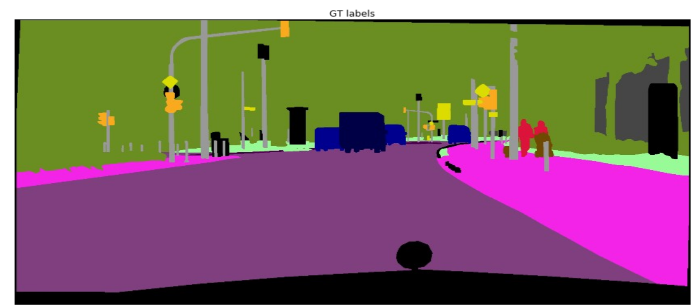
# Results on CityScapes Dataset

Image



GT

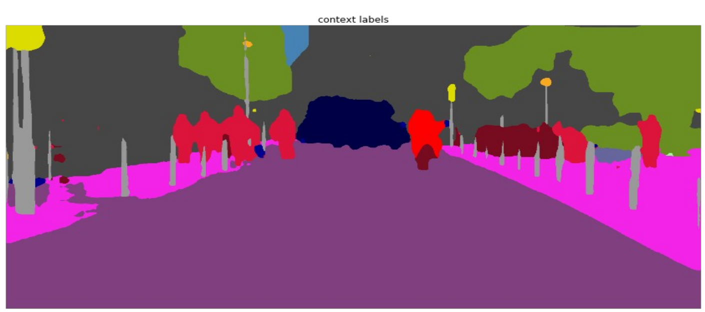



Base



Context




# Results on CityScapes Dataset

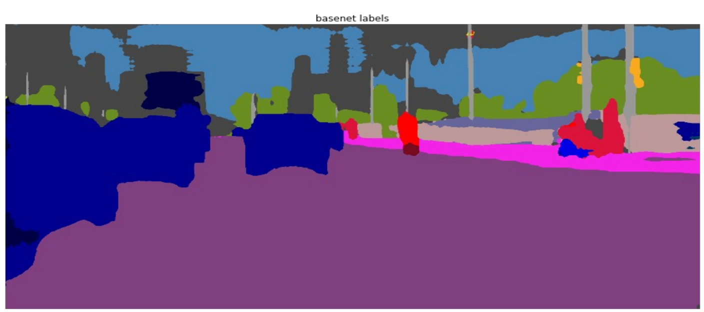
Image

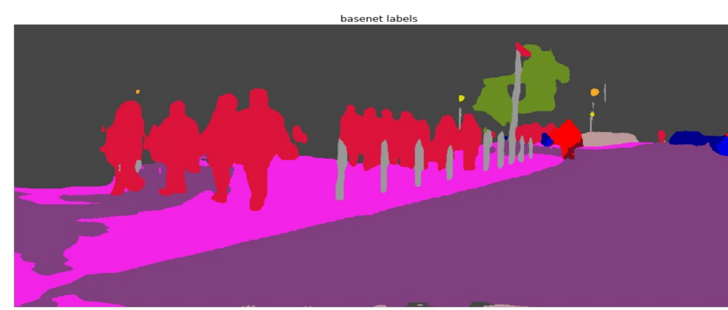


GT

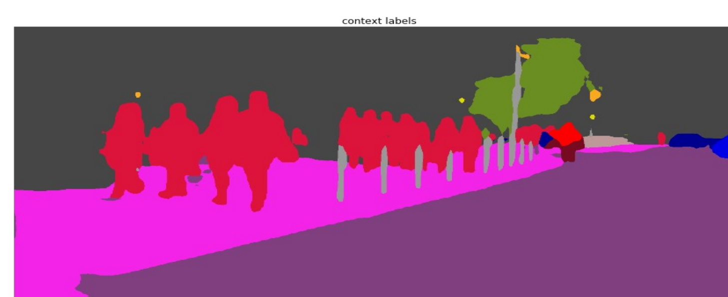


Base




Context

# Brainstorming

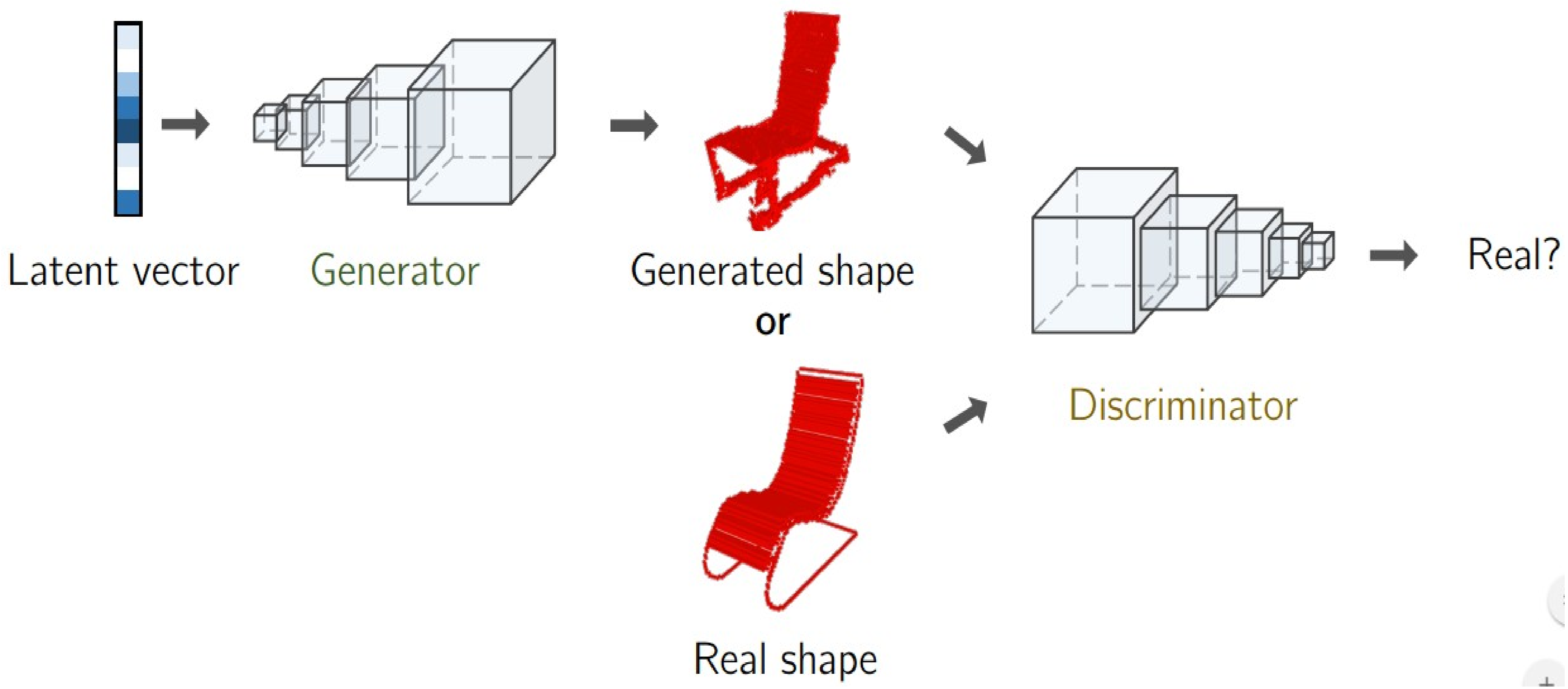



Image credit: Jiajun Wu

# Demo

- Rue-Monge 3D results
- Cityscapes video (trained with LRR by Golnaz 16)

# Thank you!