CASENet: Deep Category-Aware Semantic Edge Detection

Zhiding Yu*, Chen Feng*, Ming-Yu Liu, Srikumar Ramalingam

yzhiding@andrew.cmu.edu, cfeng@merl.com, mingyul@nvidia.com, Srikumar@cs.utah.edu

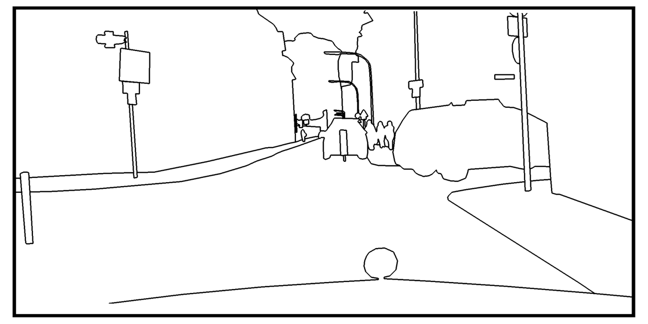
MITSUBISHI ELECTRIC *Changes for the Better*

Introduction

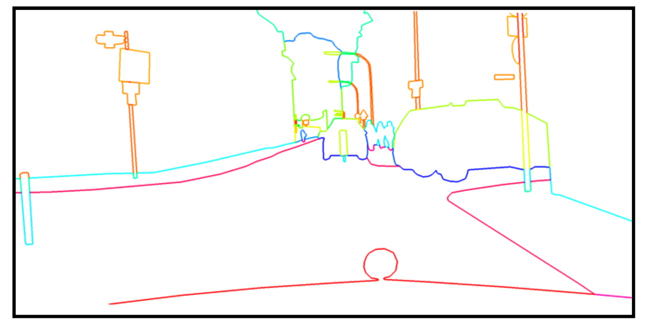
Several types of edge detection problems:

Original Image

Perceptual Edges



Semantic Edges



Category-Aware SEs

Category-aware semantic edge detection: To simultaneously detect and categorize the semantic object boundaries.

Research Motivation

Edge detection is a fundamental vision problem.

- Edge plays an important role in perceptual grouping.

Wide applications in scene understanding:

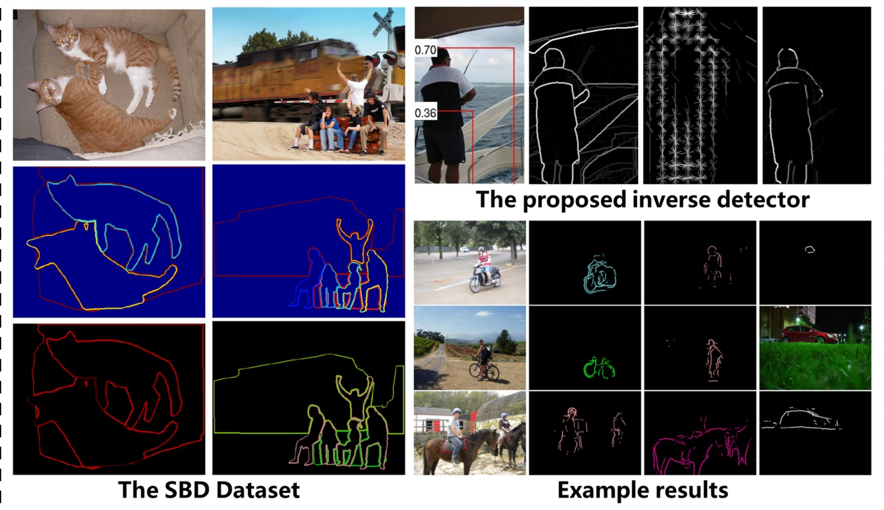
- Improve image segmentation
- Generate object proposal
- 3D shape recovery and 3D reconstruction
- Depth estimation and refinement

Formulation & Notation

- Input: RGB image **I**.
- Output: K edge probability maps $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_K\}$.
  - K: Number of object categories.
  - $\mathbf{Y}_k(\mathbf{p}) \in [0, 1]$: Predicted edge probability of pixel **p** on the k-th class.
- End to end trainable with a deep neural network

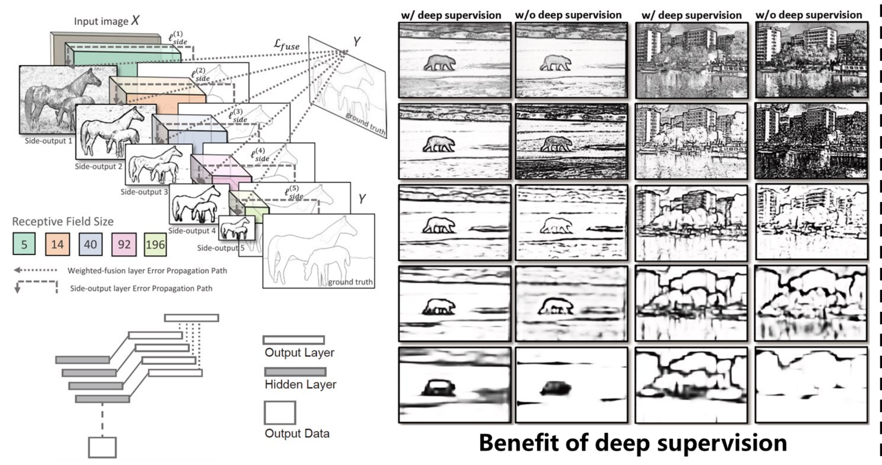
Previous Related Literature

Inverse Detector



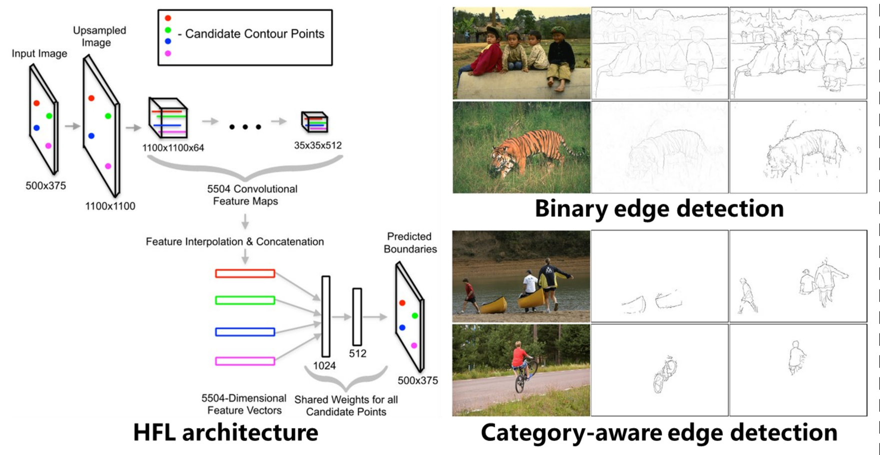
Bharath Hariharan et al., **Semantic Contours from Inverse Detectors**, *ICCV* 2011.

HED



Saining Xie et al., **Holistically-Nested Edge Detection**, *ICCV* 2015.

HFL

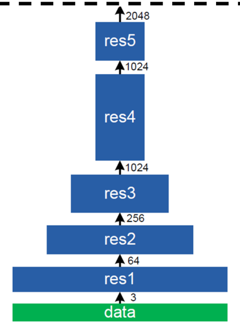


Gedas Bertasius et al., **High-for-Low and Low-for-High: Efficient Boundary Detection from Deep Object Features and its Applications to High-Level Vision**, *ICCV* 2015.

Deeper Base Network

Previous work indicate that recognition performance benefits from going deeper:

- Evolution of network architectures on ILSVRC (AlexNet->VGG->ResNet) with significant performance gains.
- Segmentation also benefitted from deeper networks (Eg. DeepLab v1 (VGG) -> DeepLab v2 (ResNet-101)).
- Deeper architecture leads to larger receptive field and more context, which is important.



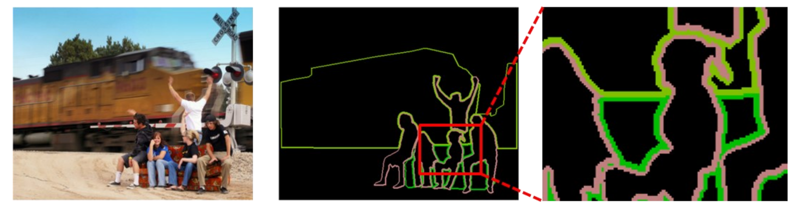
Based on ResNet-101 with certain changes

- Fully convolutional
  - Removed average pooling + FC
- Preserve low-level edge information
  - Stride 1 in res1 and res5 instead of Stride 2
  - Maintain receptive field with dilated convolution

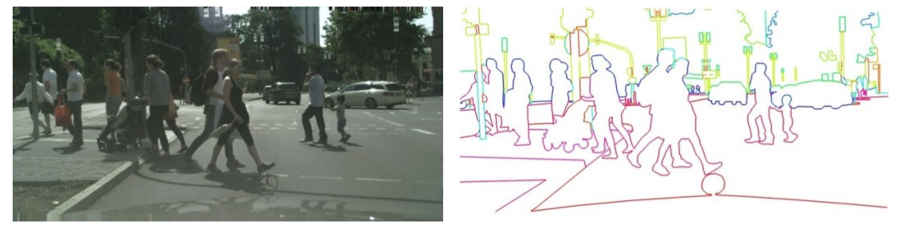
The Multi-Label Loss

Previous work often forms a multi-class problem

- SBD explicitly annotates with non-overlapping edges.
- HFL labels edge categories in a **mutually exclusive** way.



We argue that edge categories should be non-mutually exclusive.



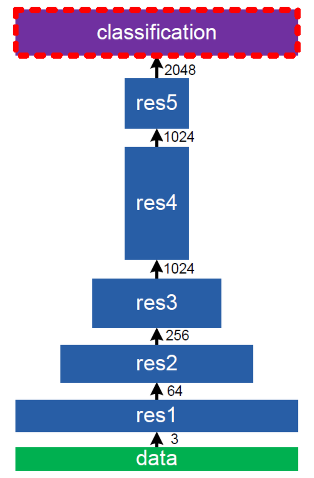
Model as a weighted multi-label cross-entropy loss:

- Sum of binary cross-entropy losses from each category.
- β: percentage of non-boundary pixels in the image.
- Able to handle heavy bias of positive/negative ratio.

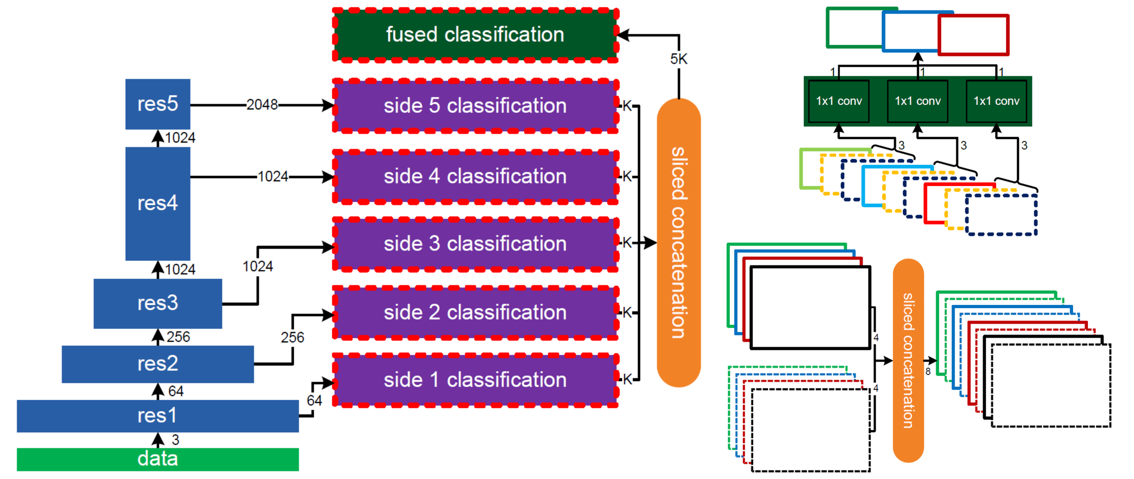
$$\begin{aligned}\mathcal{L}(\mathbf{W}) &= \sum_k \mathcal{L}_k(\mathbf{W}) \\ &= \sum_k \sum_{\mathbf{p}} \{-\beta \bar{\mathbf{Y}}_k(\mathbf{p}) \log \mathbf{Y}_k(\mathbf{p}|\mathbf{I};\mathbf{W}) \\ &\quad - (1-\beta)(1-\bar{\mathbf{Y}}_k(\mathbf{p})) \log(1-\mathbf{Y}_k(\mathbf{p}|\mathbf{I};\mathbf{W}))\}\end{aligned}$$

The CASENet Architecture

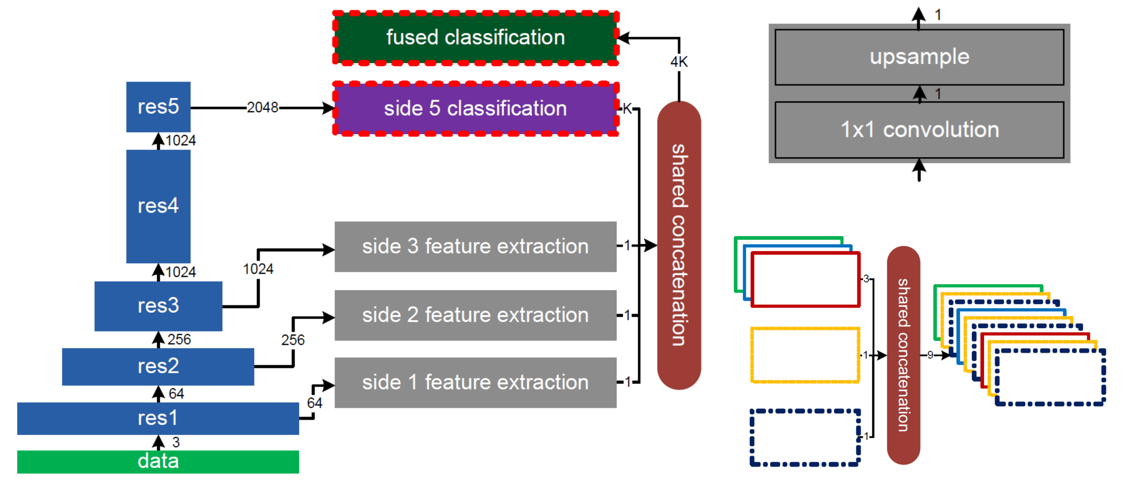
Basic



Deeply Supervised Net (DSN)



CASENet



Experimental Results (SBD)

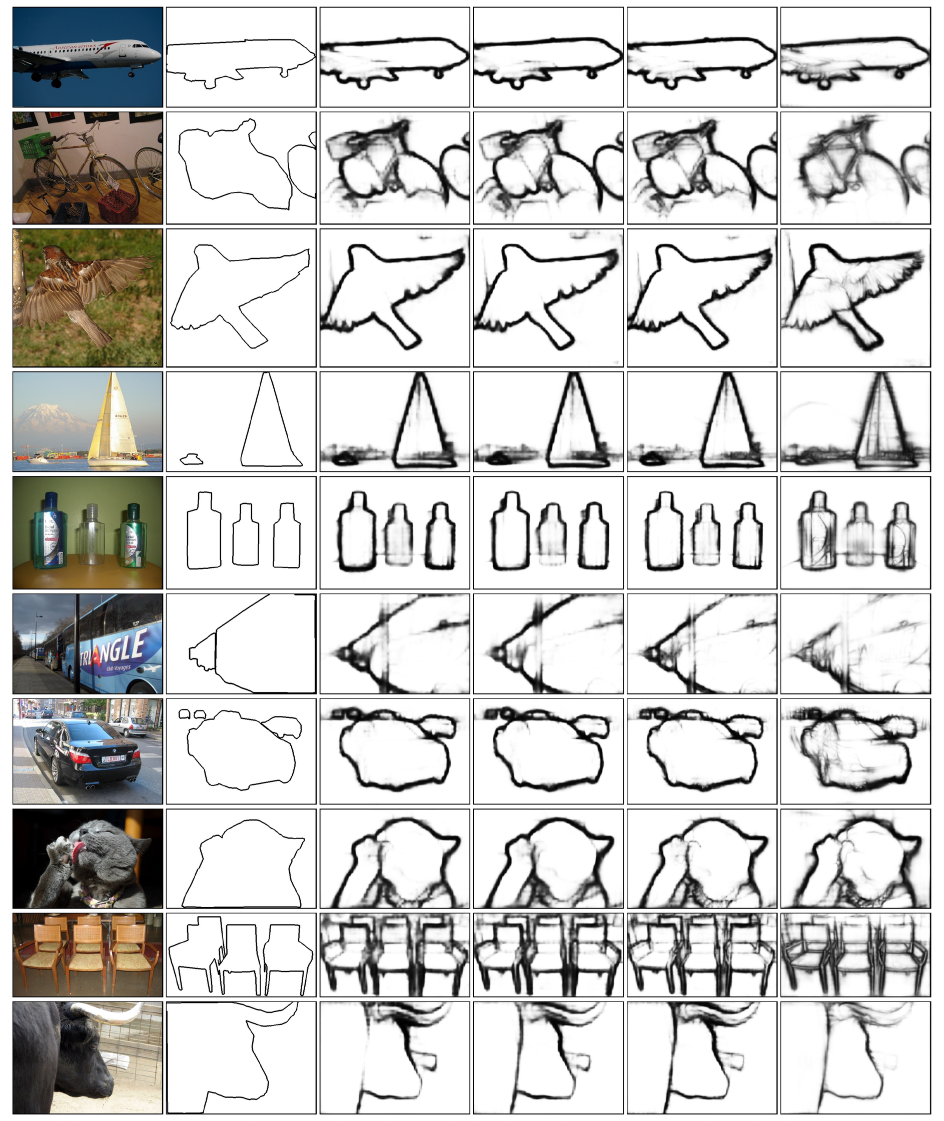


Figure 2. Class-wise prediction results of comparing methods on the SBD Dataset. Rows correspond to the predicted edges of "aeroplane", "bicycle", "bird", "boat", "bottle", "bus", "car", "cat", "chair" and "cow". Columns correspond to original image, ground truth, and results of Basic, DSN, CASENet and CASENet-VGG.

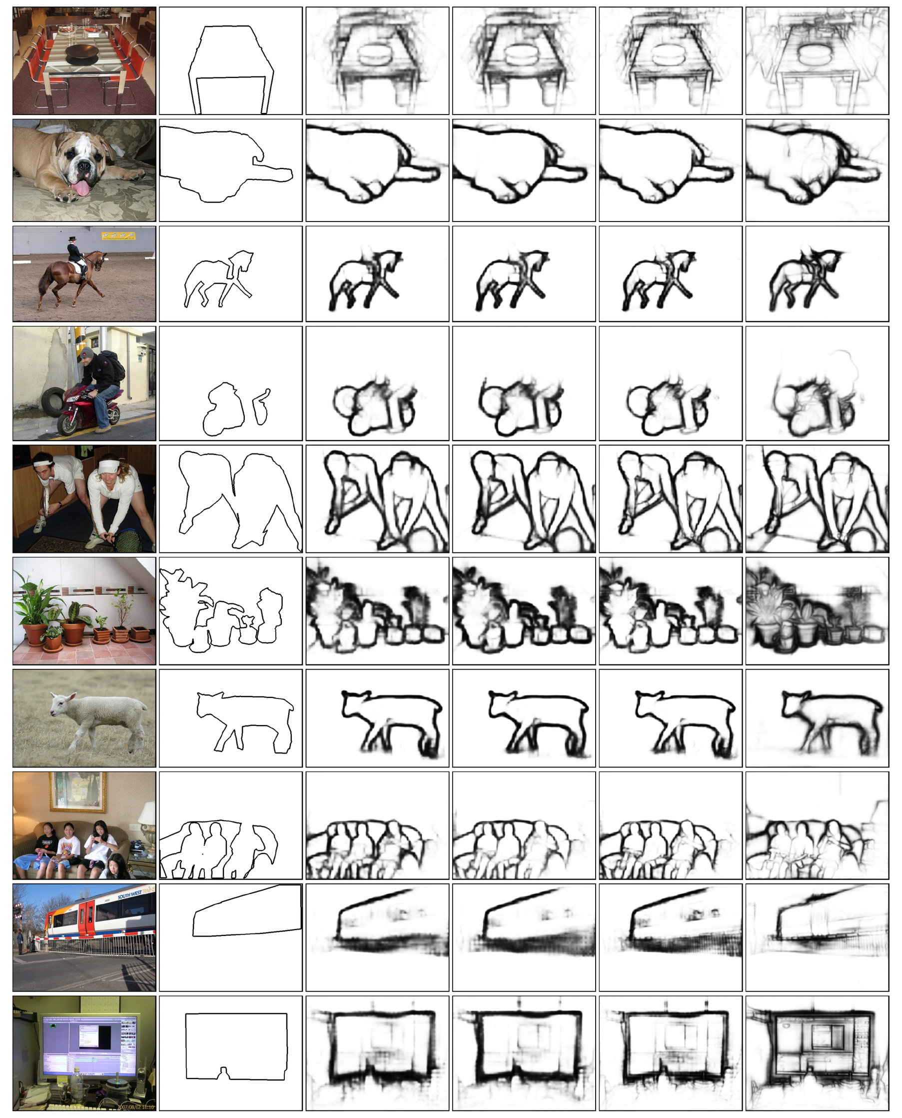
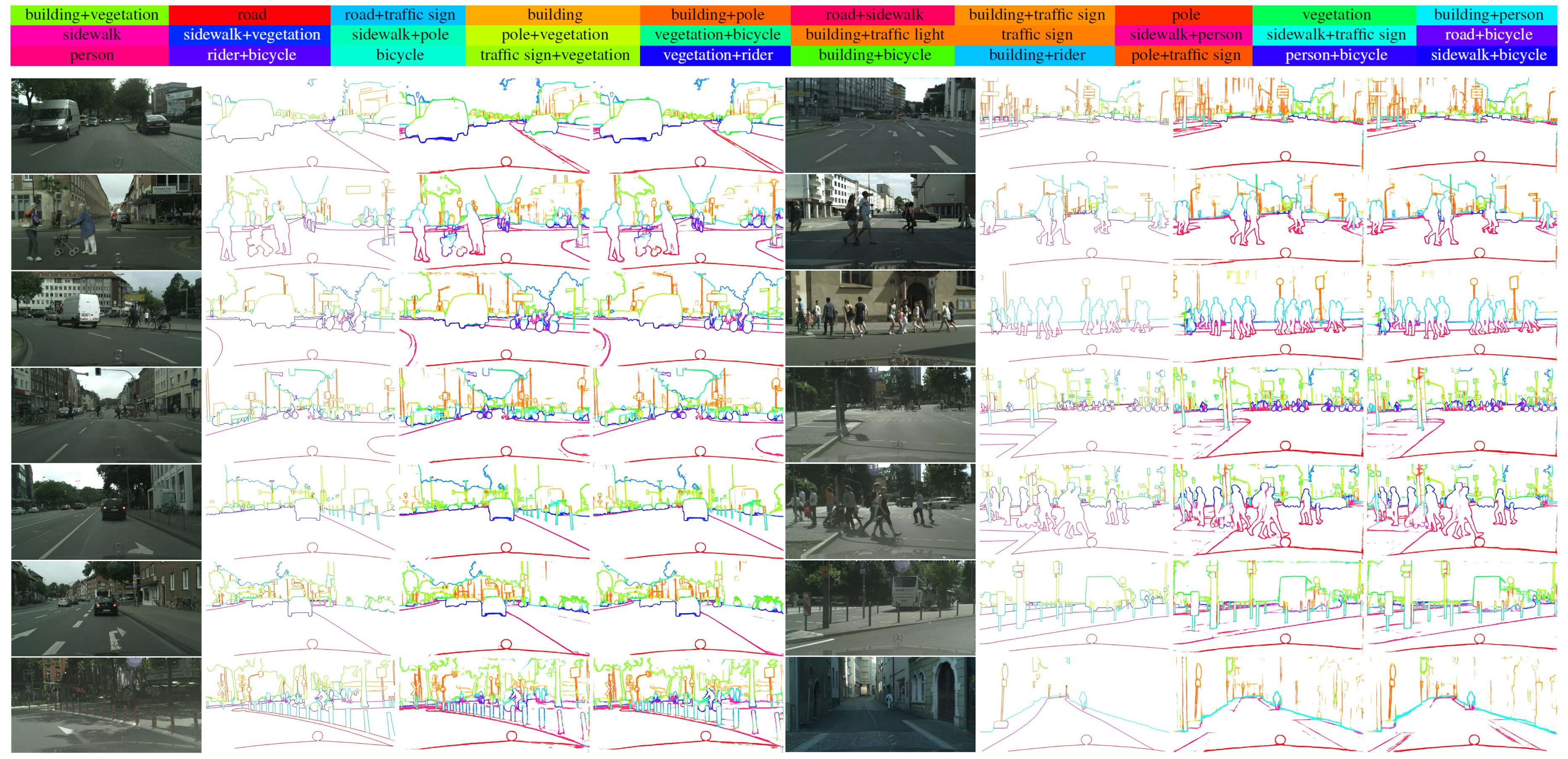


Figure 3. Class-wise prediction results of comparing methods on the SBD Dataset. Rows correspond to the predicted edges of "dining table", "dog", "horse", "motorbike", "person", "potted plant", "sheep", "sofa", "train" and "tv monitor". Columns correspond to original image, ground truth, and results of Basic, DSN, CASENet and CASENet-VGG.

Metric	Category	Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
MF (ODS)	Baseline	InvDet	41.5	46.7	15.6	17.1	36.5	42.6	40.3	22.7	18.9	26.9	12.5	18.2	35.4	29.4	48.2	13.9	26.9	11.1	21.9	31.4	27.9
		HFL-FC8	71.6	59.6	68.0	54.1	57.2	68.0	58.8	69.3	43.3	65.8	33.3	67.9	67.5	62.2	69.0	43.8	68.5	33.9	57.7	54.8	58.7
		HFL-CRF	73.9	61.4	74.6	57.2	58.8	70.4	61.6	71.9	46.5	72.3	36.2	71.1	73.0	68.1	70.3	44.4	73.2	42.6	62.4	60.1	62.5
	VGG	Basic-Softmax	67.6	55.3	50.4	44.9	42.3	64.6	61.0	63.9	37.4	43.1	25.3	57.9	57.1	60.0	72.0	33.0	53.5	30.9	54.4	47.7	51.1
		Basic	70.0	58.6	62.5	50.2	51.2	65.4	60.6	66.9	39.7	47.3	31.0	60.1	59.4	60.2	74.4	38.0	56.0	35.9	60.0	53.8	55.1
		CASENet	72.5	61.5	63.8	54.5	52.3	65.4	62.6	67.2	42.6	51.8	31.4	62.0	61.9	62.8	75.4	41.7	59.8	35.8	59.7	50.7	56.8
	ResNet	Basic-Softmax	74.0	64.1	64.8	52.5	52.1	73.2	68.1	73.2	43.1	56.2	37.3	67.4	68.4	67.6	76.7	42.7	64.3	37.5	64.6	56.3	60.2
		Basic	82.5	74.2	80.2	62.3	68.0	80.8	74.3	82.9	52.9	73.1	46.1	79.6	78.9	76.0	80.4	52.4	75.4	48.6	75.8	68.0	70.6
		DSN	81.6	75.6	78.4	61.3	67.6	**82.3**	74.6	82.6	52.4	71.9	45.9	79.2	78.3	76.2	80.1	51.9	74.9	48.0	76.5	66.8	70.3
		CASENet$^-$	83.0	74.7	79.6	61.5	67.7	80.7	74.1	82.8	53.3	**75.0**	44.5	79.8	**80.4**	76.2	80.2	53.2	**77.3**	47.7	75.6	66.3	70.7
		CASENet	**83.3**	**76.0**	**80.7**	**63.4**	**69.2**	81.3	**74.9**	**83.2**	**54.3**	74.8	**46.4**	**80.3**	80.2	**76.6**	**80.8**	**53.3**	77.2	**50.1**	**75.9**	**66.8**	**71.4**

Experimental Results (Cityscapes)



Example results on Cityscapes. Columns from left to right: Input, Ground Truth, DSN and CASENet. CASENet shows better detection qualities on challenging objects, while DSN shows slightly more false positives on non-edge pixels.

Metric	Method	road	sidewalk	building	wall	fence	pole	traffic lgt	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bike	mean
MF (ODS)	DSN	85.4	76.4	82.6	**51.8**	56.5	66.5	62.6	72.1	80.6	61.1	76.0	77.5	66.3	84.5	**52.3**	67.3	49.4	56.0	76.0	68.5
	CASENet	**86.6**	**78.8**	**85.1**	51.5	**58.9**	**70.1**	**70.8**	**74.6**	**83.5**	**62.9**	**79.4**	**81.5**	**71.3**	**86.9**	50.4	**69.5**	**52.0**	**61.3**	**80.2**	**71.3**
AP	DSN	**78.0**	76.0	83.9	47.9	53.1	67.9	57.9	75.9	79.9	60.2	75.0	75.4	61.0	85.8	**50.6**	67.8	42.5	51.4	72.0	66.4
	CASENet	77.7	**78.6**	**87.6**	**49.0**	**56.9**	**72.8**	**70.3**	**78.9**	**85.1**	**63.1**	**78.4**	**83.0**	**70.1**	**89.5**	46.9	**70.0**	**48.8**	**59.6**	**78.9**	**70.8**