Simultaneous Edge Alignment and Learning

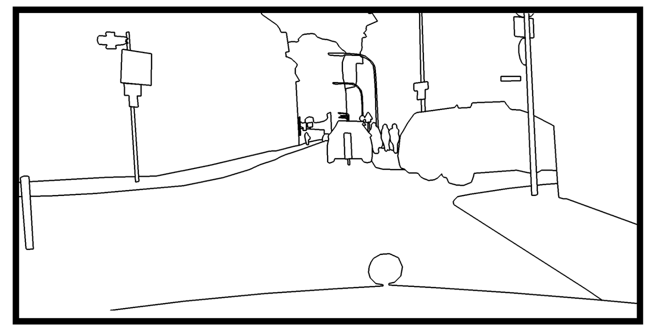
Zhiding Yu, Weiyang Liu, Yang Zou, Chen Feng, Srikumar Ramalingam, B. V. K. Vijaya Kumar, Jan Kautz
{zhidingy, jkautz}@nvidia.com, wyliu@gatech.edu, yzou2@andrew.cmu.edu, cfeng@nyu.edu, srikumar@cs.utah.edu

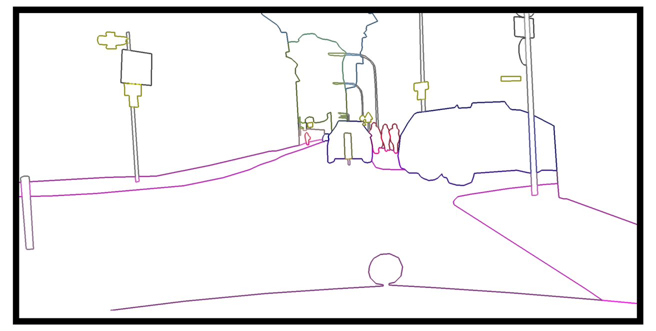
Introduction

Several existing edge/boundary detection problems:


Original Image


Perceptual Edges

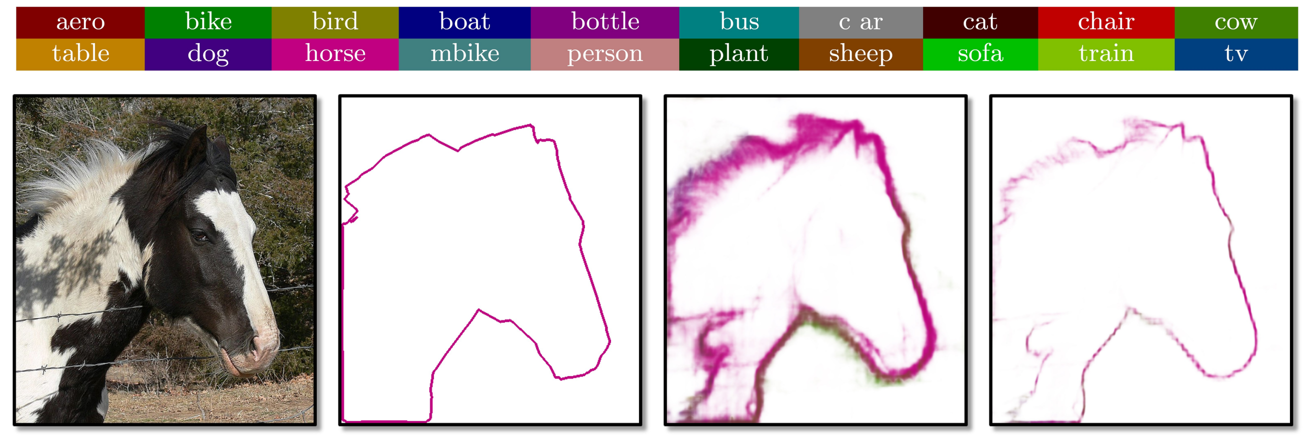

Semantic Edges


Category-Aware SEs

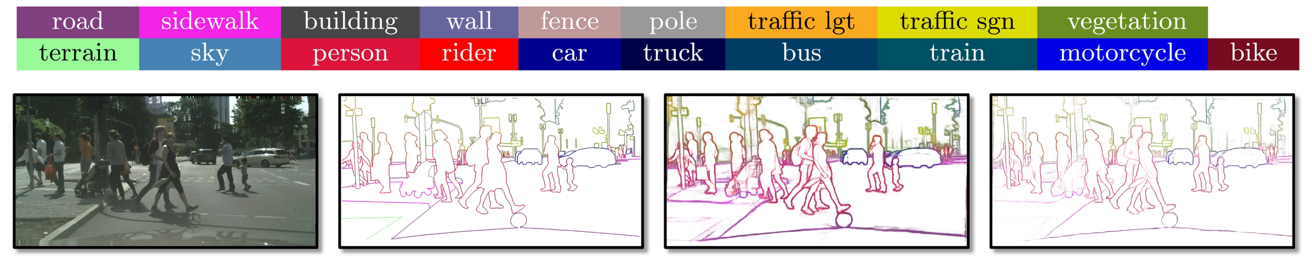
Research Motivation

Noisy label learning: Automatic alignment of edge labels

Learning to directly predict sharp/crisp boundaries



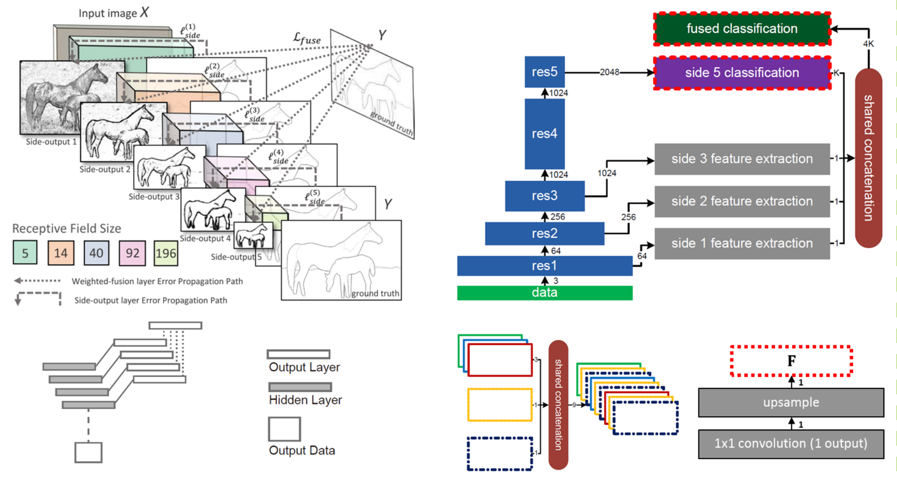
(a) Original image (b) Ground truth (c) CASENet (d) SEAL



(e) Original image (f) Ground truth (g) CASENet (h) SEAL

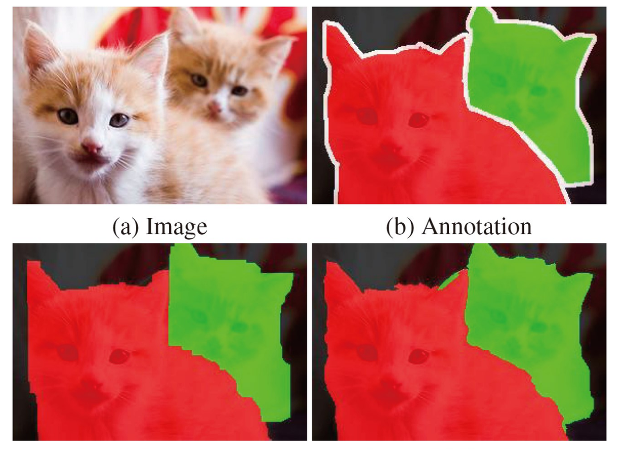
Previous Related Work

CNN-based Edge Detection



[1] Xie et al., "Holistically-nested edge detection," *ICCV15.*
[2] Yu et al., "CASENet: Deep category-aware semantic edge detection," *CVPR17.*

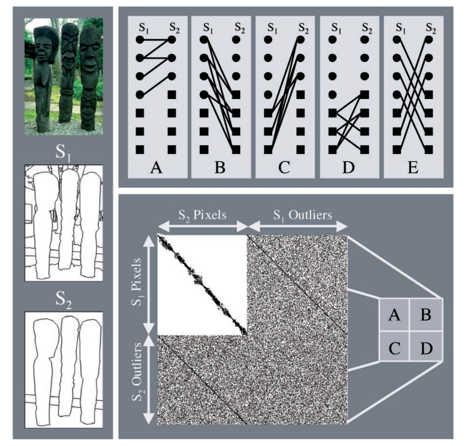
Mask Refinement via CRF



(a) Image (b) Annotation
(c) GraphCut refinement (d) DenseCRF refinement

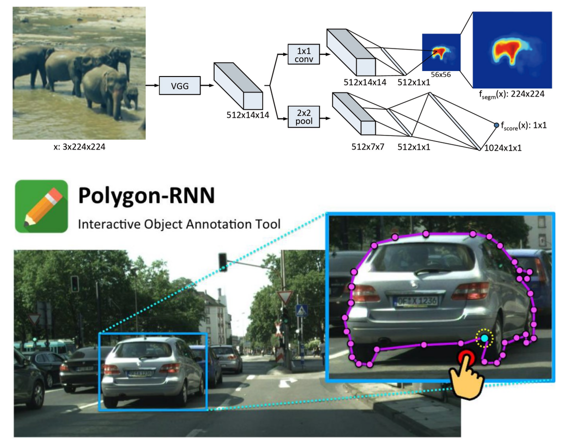
[3] Yang et al., "Object contour detection with a fully convolutional encoder-decoder network," *CVPR16.*

Edge Matching



[4] Martin et al., "Learning to detect natural image boundaries using local brightness, color, and texture cues" *IEEE TPAMI 04.*

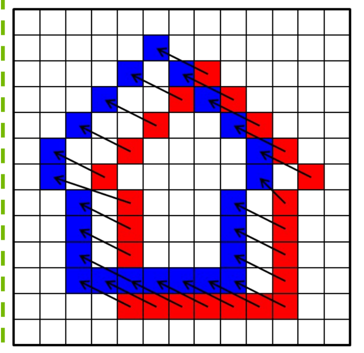
Instance Mask Learning



[5] Pinheiro et al., "Learning to segment object candidates," *NIPS15.*
[6] Castrejon et al., "Annotating object instances with a Polygon-RNN", *CVPR17.*

The Vanilla SEAL Framework

Notations & Fundamentals



Traditional edge learning:

$$\max_{\mathbf{W}} \mathcal{L}(\mathbf{W}) = P(\mathbf{y}|\mathbf{x};\mathbf{W})$$

Simultaneous edge alignment & learning:

$$\max_{\hat{\mathbf{y}},\mathbf{W}} \mathcal{L}(\hat{\mathbf{y}},\mathbf{W}) = P(\mathbf{y},\hat{\mathbf{y}}|\mathbf{x};\mathbf{W}) = P(\mathbf{y}|\hat{\mathbf{y}})P(\hat{\mathbf{y}}|\mathbf{x};\mathbf{W})$$
$$= \prod_k \boxed{P(\mathbf{y}^k|\hat{\mathbf{y}}^k)}\boxed{P(\hat{\mathbf{y}}^k|\mathbf{x};\mathbf{W})}$$

Edge prior Network likelihood

$\mathbf{p} = (x_p, y_p), \mathbf{q} = (x_q, y_q)$: Pixel index
$k \in \{1, ..., K\}$: Semantic class index
$\mathbf{y} = \{y_{\mathbf{q}}^k \in \{0,1\}\}$: Human annotation
$\hat{\mathbf{y}} = \{\hat{y}_{\mathbf{p}}^k \in \{0,1\}\}$: Aligned edge label
$y_{\mathbf{q}}^k = 1$ $\hat{y}_{\mathbf{p}}^k = 1$ $\nwarrow$ $m(\mathbf{q}) - \mathbf{q}$

Edge Prior/Net Likelihood

Assumption: Edge in $\mathbf{y}^k$ is generated from those in $\hat{\mathbf{y}}^k$ through one-to-one assignment.

$$\mathcal{M}(\mathbf{y}^k,\hat{\mathbf{y}}^k) = \{m(\cdot)|\forall \mathbf{u},\mathbf{v} \in \{\mathbf{q}|y_{\mathbf{q}}^k = 1\} : \hat{y}_{m(\mathbf{u})}^k = 1, \hat{y}_{m(\mathbf{v})}^k = 1, \mathbf{u} \neq \mathbf{v} \Rightarrow m(\mathbf{u}) \neq m(\mathbf{v})\},$$

$$m(\cdot) \sim E_m = \{(\mathbf{p},\mathbf{q})|\hat{y}_{\mathbf{p}}, y_{\mathbf{q}} = 1, m(\mathbf{q}) = \mathbf{p}\}$$

$$\boxed{P(\mathbf{y}^k|\hat{\mathbf{y}}^k)} \propto \sup_{m\in\mathcal{M}(\mathbf{y}^k,\hat{\mathbf{y}}^k)} \prod_{(\mathbf{p},\mathbf{q})\in E_m} \exp\left(-\frac{\|\mathbf{p}-\mathbf{q}\|^2}{2\sigma^2}\right)$$
$$= \exp\left(-\inf_{m\in\mathcal{M}(\mathbf{y}^k,\hat{\mathbf{y}}^k)} \sum_{(\mathbf{p},\mathbf{q})\in E_m} \frac{\|\mathbf{p}-\mathbf{q}\|^2}{2\sigma^2}\right)$$

Edge prior

$$\boxed{P(\hat{\mathbf{y}}^k|\mathbf{x};\mathbf{W})} = \prod_{\mathbf{p}} P(\hat{y}_{\mathbf{p}}^k|\mathbf{x};\mathbf{W})$$
$$= \prod_{\mathbf{p}} \sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W})^{\hat{y}_{\mathbf{p}}^k}(1-\sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W}))^{(1-\hat{y}_{\mathbf{p}}^k)}$$

Net likelihood

Learning

$$\log \mathcal{L}(\hat{\mathbf{y}},\mathbf{W}) = \sum_k \Big\{ -\inf_{m\in\mathcal{M}(\mathbf{y}^k,\hat{\mathbf{y}}^k)} \sum_{(\mathbf{p},\mathbf{q})\in E_m} \frac{\|\mathbf{p}-\mathbf{q}\|^2}{2\sigma^2}$$
$$+ \sum_{\mathbf{p}} \Big[\hat{y}_{\mathbf{p}}^k \log \sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W}) + (1-\hat{y}_{\mathbf{p}}^k)\log(1-\sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W}))\Big]\Big\}$$

Step 1: Update network parameters.

$$\min_{\mathbf{W}} -\log \mathcal{L}_N(\mathbf{W}) = \sum_k \sum_{\mathbf{p}} -\Big[\hat{y}_{\mathbf{p}}^k \log \sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W}) + (1-\hat{y}_{\mathbf{p}}^k)\log(1-\sigma_k(\mathbf{p}|\mathbf{x};\mathbf{W}))\Big]$$

Step 2: Update aligned edge labels.

$$\min_{\hat{\mathbf{y}}^k} -\log \mathcal{L}_E(\hat{\mathbf{y}}^k) = \inf_{m\in\mathcal{M}(\mathbf{y}^k,\hat{\mathbf{y}}^k)} \sum_{(\mathbf{p},\mathbf{q})\in E_m} \frac{\|\mathbf{p}-\mathbf{q}\|^2}{2\sigma^2}$$
$$- \sum_{\mathbf{p}} \Big[\hat{y}_{\mathbf{p}}^k \log \sigma_k(\mathbf{p}) + (1-\hat{y}_{\mathbf{p}}^k)\log(1-\sigma_k(\mathbf{p}))\Big]$$

s.t. $|\hat{\mathbf{y}}^k| = |\mathbf{y}^k|$

Problem Transform

Definition: Let $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}||\hat{\mathbf{y}}| = |\mathbf{y}|\}$, then:

$$\mathbf{M} = \{m|m \in \mathcal{M}(\mathbf{y},\hat{\mathbf{y}}), \hat{\mathbf{y}} \in \hat{\mathbf{Y}}\}$$

The constrained optimization problem in **Step 2** can be transformed into a bipartite graph min-cost assignment problem, and solved by a standard solver:

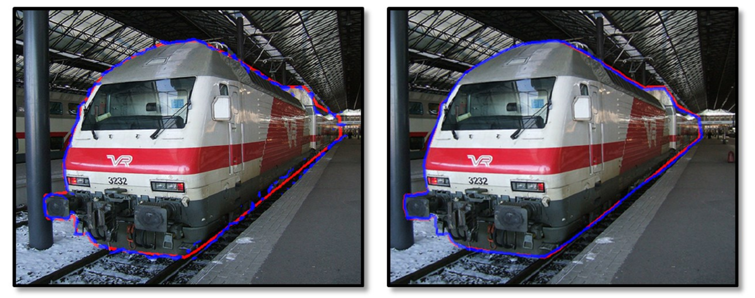
$$\min_{m\in\mathbf{M}} \sum_{(\mathbf{p},\mathbf{q})\in E_m} \Big[\boxed{\frac{\|\mathbf{p}-\mathbf{q}\|^2}{2\sigma^2}} + \boxed{\log(1-\sigma(\mathbf{p})) - \log\sigma(\mathbf{p})}\Big]$$

Edge prior Network guidance

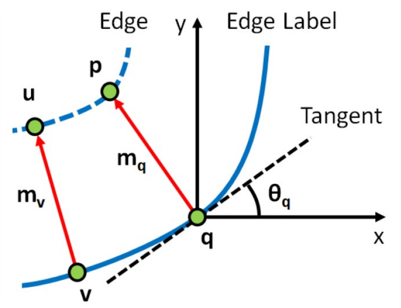
Theorem: The minimizer of the assignment problem is also a minimizer of the constrained optimization problem in **Step 2**.

Biased Gaussian Kernel and Markov Prior

Motivation and Formulation



Vanilla SEAL SEAL with BGK and MP



Graphical illustration

The new edge prior with biased Gaussian kernel and Markov prior can be written as:

$$P(\mathbf{y}|\hat{\mathbf{y}}) \propto \sup_{m\in\mathcal{M}(\mathbf{y},\hat{\mathbf{y}})} \prod_{(\mathbf{p},\mathbf{q})\in E_m} \boxed{\exp(-\mathbf{m}_{\mathbf{q}}^\top \mathbf{\Sigma}_{\mathbf{q}} \mathbf{m}_{\mathbf{q}})} \prod_{\substack{(\mathbf{u},\mathbf{v})\in E_m,\\ \mathbf{v}\in\mathcal{N}(\mathbf{q})}} \boxed{\exp(-\lambda\|\mathbf{m}_{\mathbf{q}} - \mathbf{m}_{\mathbf{v}}\|^2)}$$

Biased Gaussian kernel Markov prior

where $\mathbf{m}_{\mathbf{q}} = \mathbf{p} - \mathbf{q}$, $\mathbf{m}_{\mathbf{v}} = \mathbf{u} - \mathbf{v}$ and

$$\mathbf{\Sigma}_{\mathbf{q}} = \begin{bmatrix} \frac{\cos(\theta_{\mathbf{q}})^2}{2\sigma_x^2} + \frac{\sin(\theta_{\mathbf{q}})^2}{2\sigma_y^2} & \frac{\sin(2\theta_{\mathbf{q}})}{4\sigma_y^2} - \frac{\sin(2\theta_{\mathbf{q}})}{4*\sigma_x^2} \\ \frac{\sin(2\theta_{\mathbf{q}})}{4\sigma_y^2} - \frac{\sin(2\theta_{\mathbf{q}})}{4\sigma_x^2} & \frac{\sin(\theta_{\mathbf{q}})^2}{2\sigma_x^2} + \frac{\cos(\theta_{\mathbf{q}})^2}{2\sigma_y^2} \end{bmatrix}$$

Optimization for Assignment with Pairwise Cost

Given the new edge prior model, the aligned edge label update steps can be reformulated as solving the following assignment problem with the cost consisting a unary term and a pairwise term:

$$\min_{m\in\mathbf{M}} \mathcal{C}(m) = \mathcal{C}_{Unary}(m) + \mathcal{C}_{Pair}(m)$$
$$= \sum_{(\mathbf{p},\mathbf{q})\in E_m} \Big[\mathbf{m}_{\mathbf{q}}^\top \mathbf{\Sigma}_{\mathbf{q}} \mathbf{m}_{\mathbf{q}} + \log((1-\sigma(\mathbf{p}))/\sigma(\mathbf{p}))\Big] + \lambda \sum_{(\mathbf{p},\mathbf{q})\in E_m} \sum_{\substack{(\mathbf{u},\mathbf{v})\in E_m,\\ \mathbf{v}\in\mathcal{N}(\mathbf{q})}} \|\mathbf{m}_{\mathbf{q}} - \mathbf{m}_{\mathbf{v}}\|^2$$

Solving the assignment with pairwise dependencies is difficult. As a result we decouple the pairwise term as follows:

$$\mathcal{C}_{Pair}(m,m') = \sum_{(\mathbf{p},\mathbf{q})\in E_m} \sum_{\substack{(\mathbf{u},\mathbf{v})\in E_{m'},\\ \mathbf{v}\in\mathcal{N}(\mathbf{q})}} \|\mathbf{m}_{\mathbf{q}} - \mathbf{m}_{\mathbf{v}}\|^2$$

We then propose an iterative-conditional mode like algorithm to approximate the solution:

Iterative Approximation Algorithm

Initialize: $m^{(0)} = \arg\min_{m\in\mathbf{M}} \mathcal{C}_{Unary}(m)$

Assign: $m^{(t+1)} = \arg\min_{m\in\mathbf{M}} \mathcal{C}_{Unary}(m) + \mathcal{C}_{Pair}(m, m^{(t)})$

Update: $\mathcal{C}_{Pair}(m, m^{(t)}) \rightarrow \mathcal{C}_{Pair}(m, m^{(t+1)})$

Repeat **Assign** and **Update** steps multiple times.

Baselines & Benchmarks

CASENet: Network from [1] trained with instance-sensitive edge labels.
CASENet-S: CASENet with loss replaced to regular sigmoid cross-entropy loss.
CASENet-C: CASENet-S trained on labels preprocessed by dense-CRF following [3].
SEAL: The proposed learning framework trained with a CASENet backbone.
MF: Maximum F-measure (ODS). **Thin:** Evaluation with thinned GT and prediction.
Raw: Evaluation with unthinned GT (same width as train label) and original prediction.

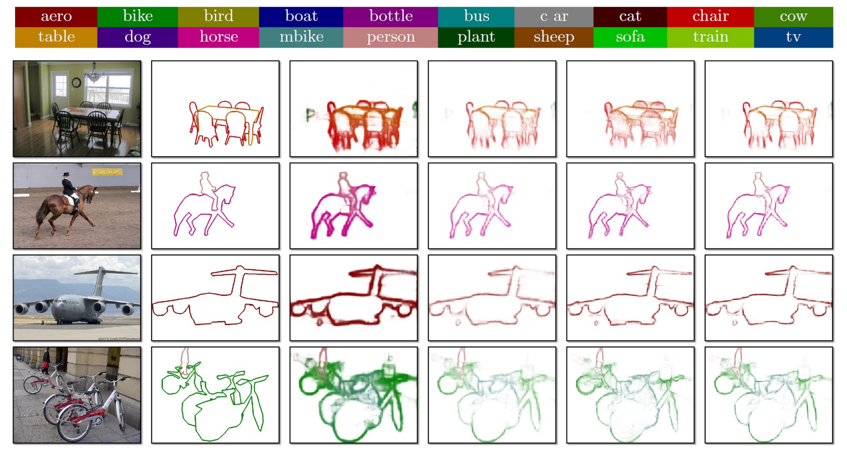
Experimental Results (SBD)

MF scores on the SBD test set. Results are measured by %.

Metric	Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
MF (Thin)	CASENet	83.6	75.3	82.3	63.1	70.5	83.5	76.5	82.6	56.8	76.3	47.5	80.8	80.9	75.6	80.7	54.1	77.7	52.3	77.9	68.0	72.3
	CASENet-S	**84.5**	**76.5**	**83.7**	**65.3**	71.3	**83.9**	**78.3**	84.5	**58.8**	76.8	50.8	81.9	**82.3**	**77.2**	82.7	**55.9**	78.1	54.0	**79.5**	69.4	**73.8**
	CASENet-C	83.9	71.1	82.5	62.6	71.0	82.2	76.8	83.4	56.5	**76.9**	49.2	81.0	81.1	75.4	81.4	54.0	**78.5**	53.3	77.1	67.0	72.2
	SEAL	**84.5**	**76.5**	**83.7**	64.9	**71.7**	83.8	78.1	**85.0**	**58.8**	76.6	**50.9**	**82.4**	82.2	77.1	**83.0**	55.1	78.4	**54.4**	79.3	**69.6**	**73.8**
MF (Raw)	CASENet	71.8	60.2	72.6	49.5	59.3	73.3	65.2	70.8	51.9	64.9	41.2	67.9	72.5	64.1	71.2	44.0	71.7	45.7	65.4	55.8	62.0
	CASENet-S	75.8	65.0	78.4	56.2	64.7	76.4	71.8	75.2	55.2	68.7	45.8	72.8	77.0	68.1	76.5	47.1	75.5	49.0	70.2	60.6	66.5
	CASENet-C	80.4	67.1	79.9	57.9	65.9	77.6	72.6	79.2	53.5	72.7	45.5	76.7	79.4	71.2	78.3	**50.8**	77.6	50.7	71.6	61.6	68.5
	SEAL	**81.1**	**69.6**	**81.7**	**60.6**	**68.0**	**80.5**	**75.1**	**80.7**	**57.0**	**73.1**	**48.1**	**78.2**	**80.3**	**72.1**	**79.8**	50.0	**78.2**	**51.8**	**74.6**	**65.0**	**70.3**

MF scores on the re-annotated SBD test set. Results are measured by %.

Metric	Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
MF (Thin)	CASENet	74.5	59.7	73.4	48.0	67.1	78.6	67.3	76.2	47.5	69.7	36.2	75.7	72.7	61.3	74.8	42.6	71.8	48.9	71.7	54.9	63.6
	CASENet-S	75.9	62.4	75.5	52.0	66.7	79.7	**71.0**	79.0	**50.1**	70.0	39.8	77.2	74.5	65.0	77.0	47.3	72.7	51.5	72.9	57.3	65.9
	CASENet-C	**78.4**	60.9	74.9	49.7	64.4	75.8	67.2	77.1	48.2	71.2	40.9	76.1	72.9	64.5	75.9	**51.4**	71.3	51.6	68.6	55.4	64.8
	SEAL	78.0	**65.8**	**76.6**	**52.4**	**68.6**	**80.0**	70.4	**79.4**	50.0	**72.8**	**41.4**	**78.1**	**75.0**	**65.5**	**78.5**	49.4	**73.3**	**52.2**	**73.9**	**58.1**	**67.0**
MF (Raw)	CASENet	65.8	51.5	65.0	43.1	57.5	68.1	58.2	66.0	45.4	59.8	32.9	64.2	65.8	52.6	65.7	40.9	65.0	42.9	61.4	47.8	56.0
	CASENet-S	68.9	55.8	70.9	47.4	62.0	71.5	64.7	71.2	48.0	64.8	37.3	69.1	68.9	58.2	70.2	44.3	68.7	46.1	65.8	52.5	60.3
	CASENet-C	**75.4**	57.7	73.0	48.7	62.1	72.2	64.4	74.3	46.8	68.8	38.8	73.4	71.4	**62.2**	72.1	**50.3**	69.8	48.4	66.1	53.0	62.4
	SEAL	75.3	**60.5**	**75.1**	**51.2**	**65.4**	**76.1**	**67.9**	**75.9**	**49.7**	**69.5**	**39.9**	**74.8**	**72.7**	62.1	**74.2**	48.4	**72.3**	**49.3**	**70.6**	**56.7**	**64.4**

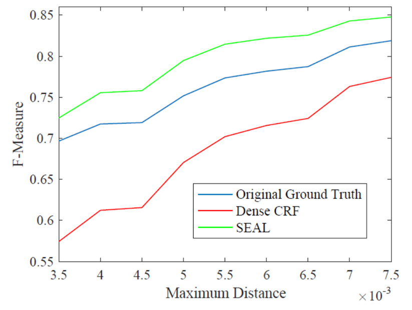


Comparison of GT, CASENet, CASENet-S, CASENet-C and SEAL.

SBD test set re-annotation.

Instance-insensitive MF scores.

Mode	CNet	CNet-S	CNet-C	SEAL
Thin	63.6	66.4	64.7	**66.9**
Raw	56.1	60.6	62.1	**64.6**

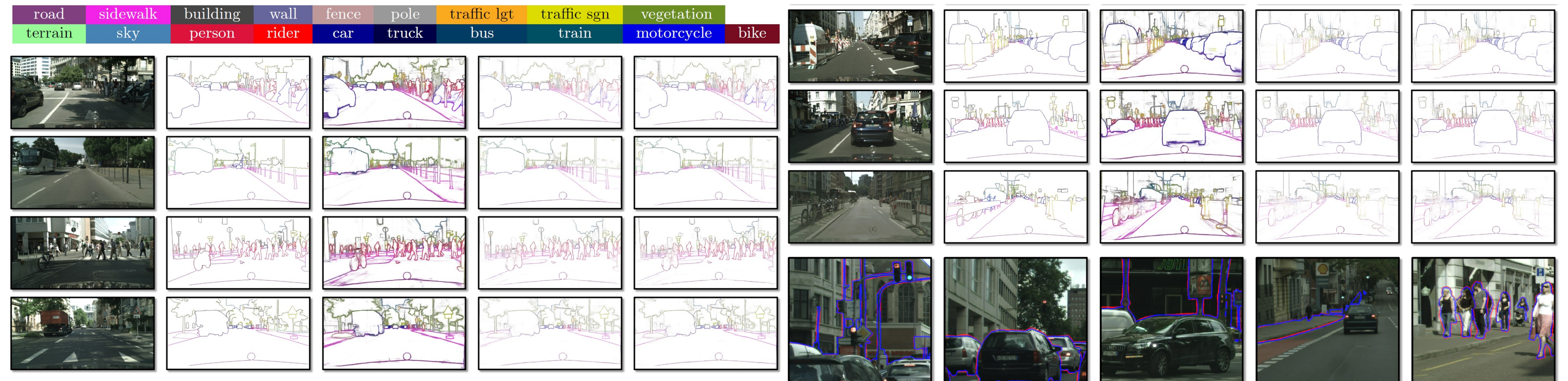


GT refinement comparison.

Experimental Results (Cityscapes)

MF scores on the Cityscapes validation set. Results are measured by %.

Metric	Method	road	sidewalk	building	wall	fence	pole	t-light	t-sign	veg	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mean
MF (Thin)	CASENet	86.2	74.9	74.5	47.6	**46.5**	72.8	70.0	73.3	79.3	57.0	86.5	80.4	66.8	88.3	49.3	64.6	**47.8**	**55.8**	71.9	68.1
	CASENet-S	**87.6**	77.1	**75.9**	**48.7**	46.2	**75.5**	**71.4**	75.3	80.6	59.7	86.8	81.4	68.1	**89.2**	**50.7**	**68.0**	42.5	54.6	72.7	**69.1**
	SEAL	**87.6**	**77.5**	**75.9**	47.6	46.3	**75.5**	71.2	**75.4**	**80.9**	**60.1**	**87.4**	**81.5**	**68.9**	88.9	50.2	67.8	44.1	52.7	**73.0**	**69.1**
MF (Raw)	CASENet	66.8	64.6	66.8	39.4	40.6	71.7	64.2	65.1	71.1	50.2	80.3	73.1	58.6	77.0	42.0	53.2	39.1	46.1	62.2	59.6
	CASENet-S	79.2	70.8	70.4	42.5	42.4	73.9	66.7	68.2	74.6	54.6	82.5	75.7	61.5	82.7	46.0	59.7	39.1	47.0	64.8	63.3
	SEAL	**84.4**	**73.5**	**72.7**	**43.4**	**43.2**	**76.1**	**68.5**	**69.8**	**77.2**	**57.5**	**85.3**	**77.6**	**63.6**	**84.9**	**48.6**	**61.9**	**41.2**	**49.0**	**66.7**	**65.5**



Qualitative comparison of GT, CASENet, CASENet-S, and SEAL, and the visualization of edge alignment on Cityscapes.