BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detectors

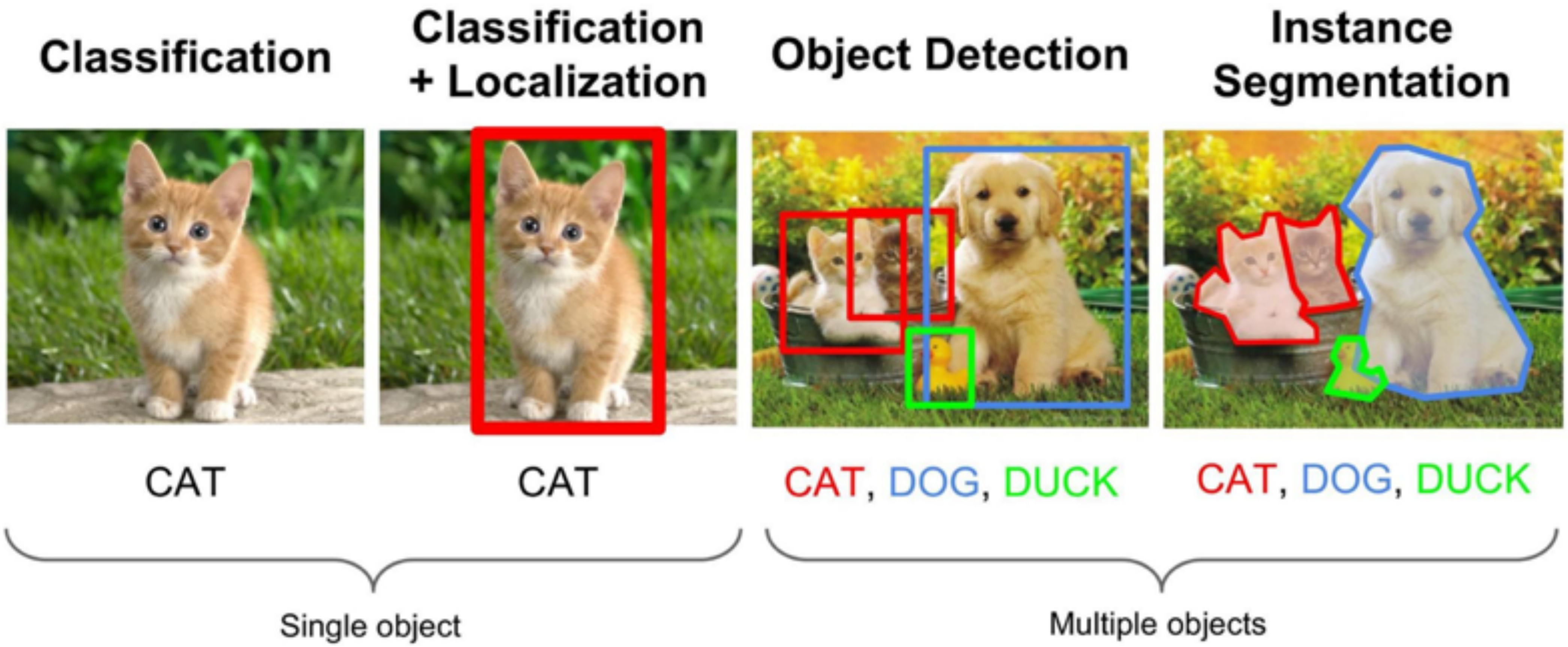
Ali Harakeh, Michael Smart, Steven Waslander

Presented by Liam Paull at Mila Robotics Reading group
July 19, 2019

Outline

- Some review of deep object detection
- Probabilistic object detection
- Some review of uncertainty estimation and bayesian inference
- BayesOD

Deep Object Detection

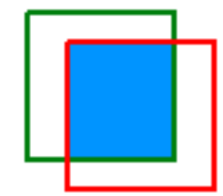


- 1 stage vs. 2 stage

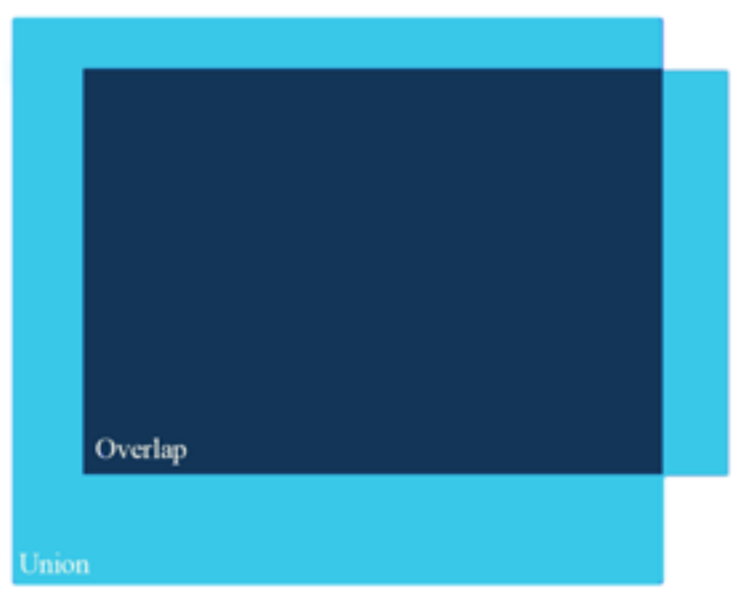
Metrics

- Intersection over union

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$



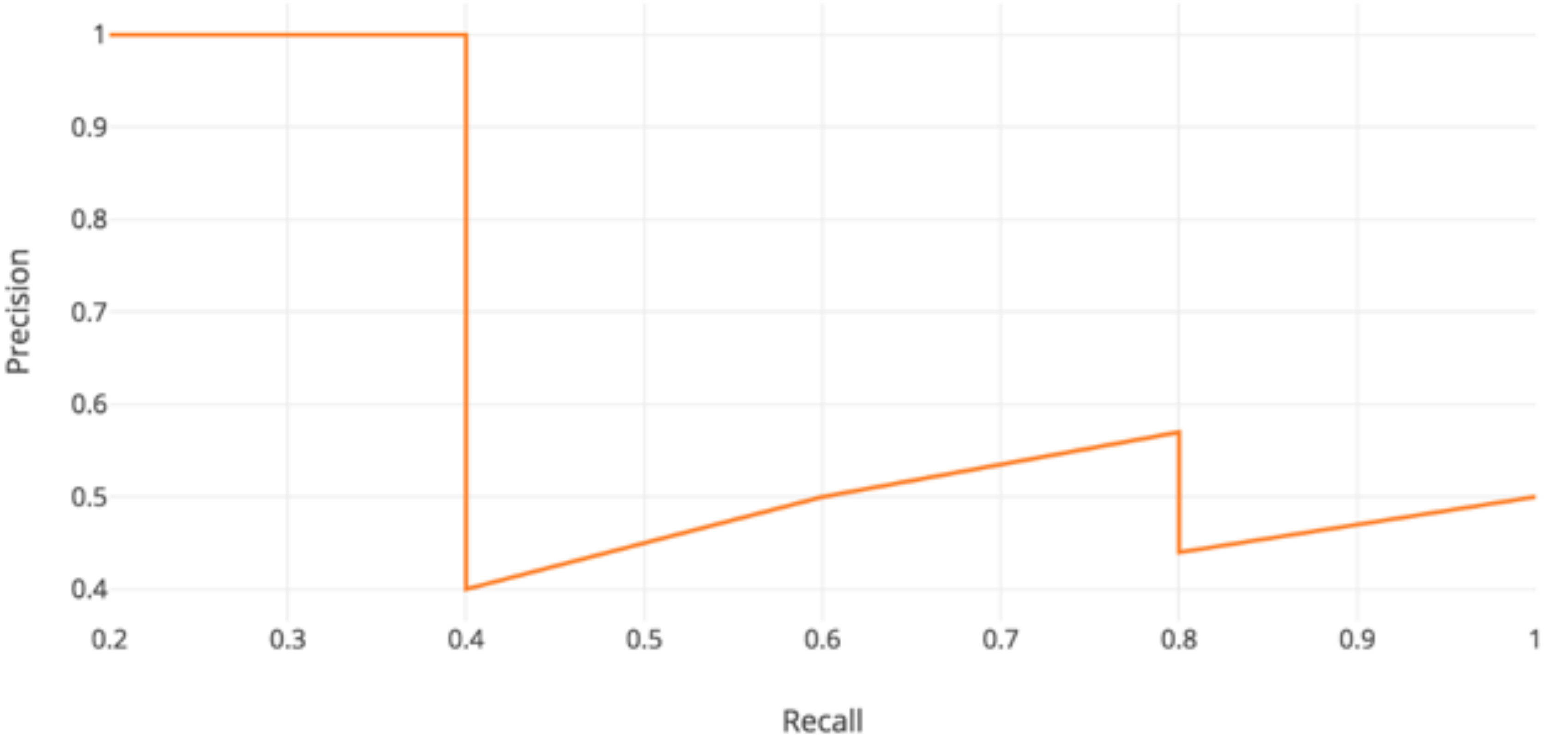
If IOU > threshold
Object detected!

Metrics

- Precision-recall

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}}$$



Average precision is area under the PR curve

$$\mathrm{AP} = \int_0^1 p(r)dr$$

Approximate as: $$AP = \frac{1}{11} \times \left(AP_r(0) + AP_r(0.1) + \ldots + AP_r(1.0)\right)$$

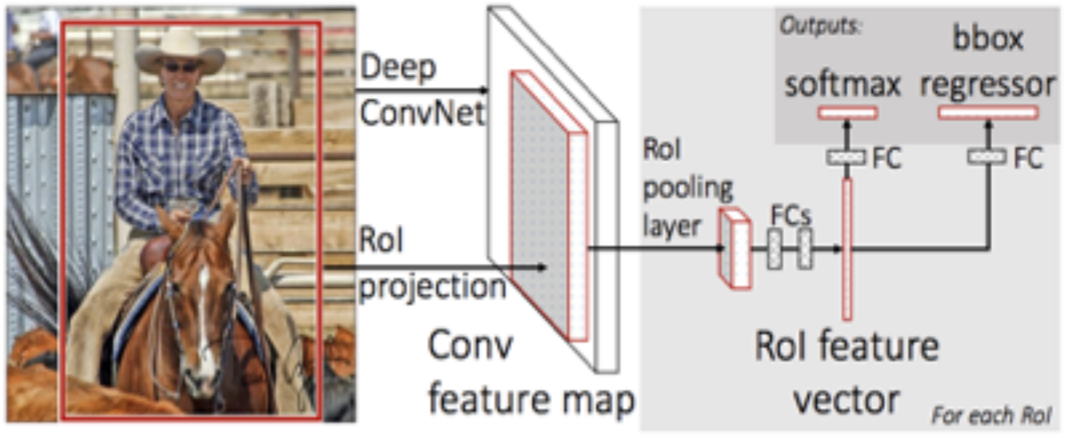
Metrics

- Mean average precision (mAP)
  - Take the mean AP over all output classes
  - Or take the mean AP over a range of IOU thresholds

2-stage Object Detectors

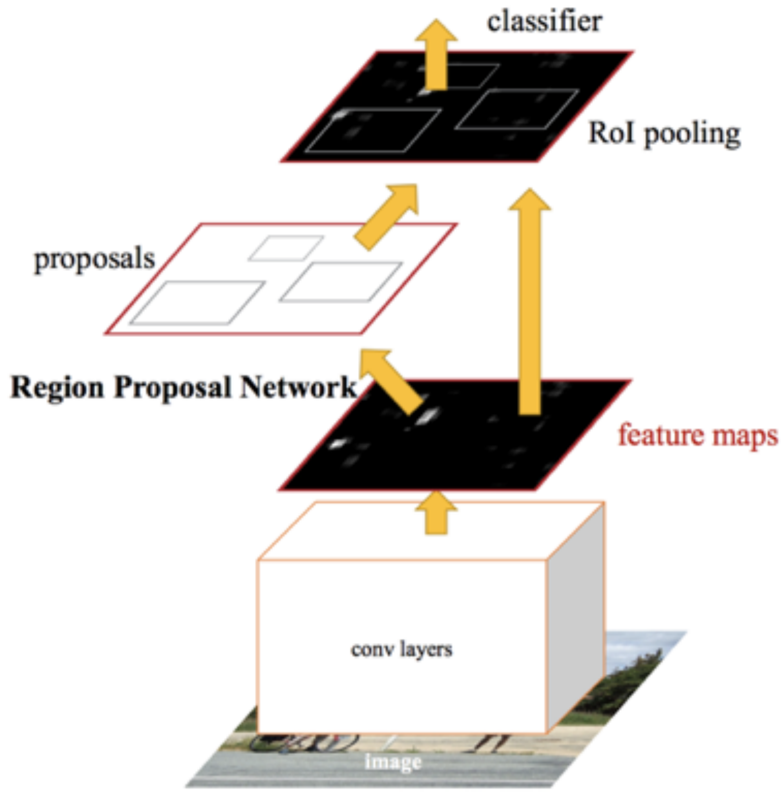
- R-CNN

- Fast R-CNN

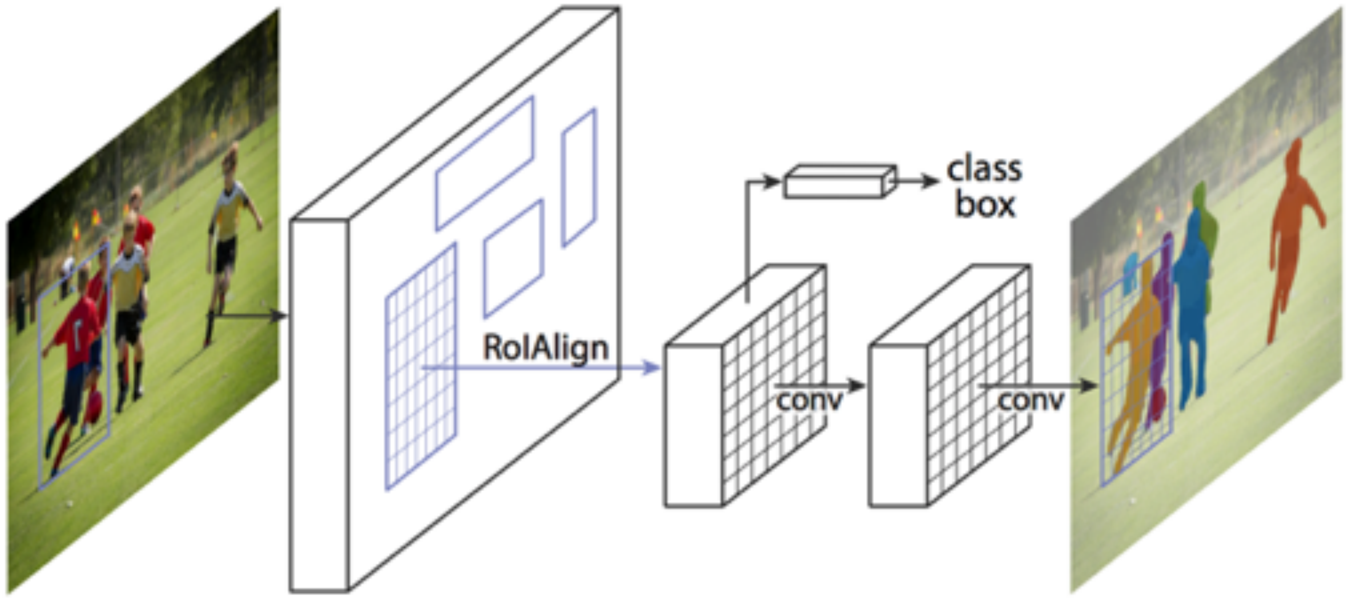


2-stage Object Detectors

- Faster R-CNN

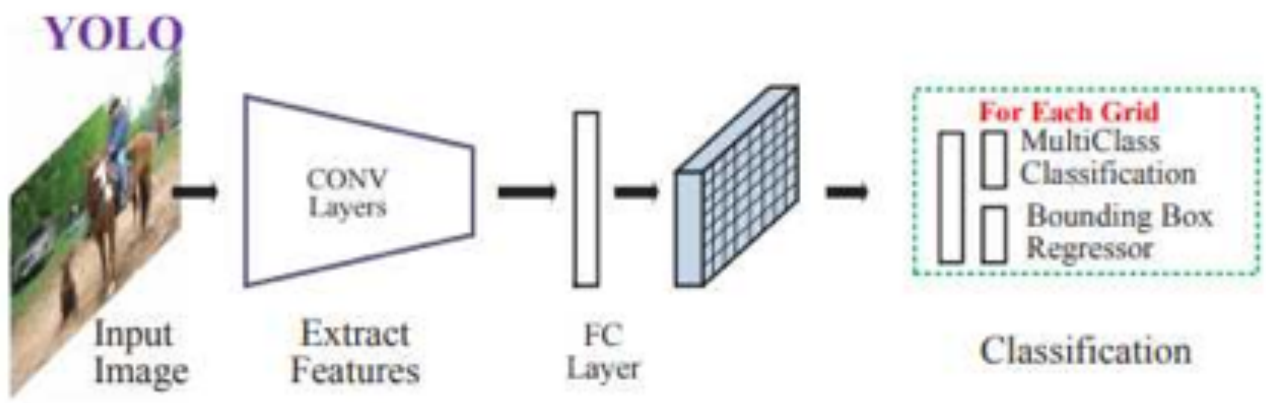


- Mask R-CNN

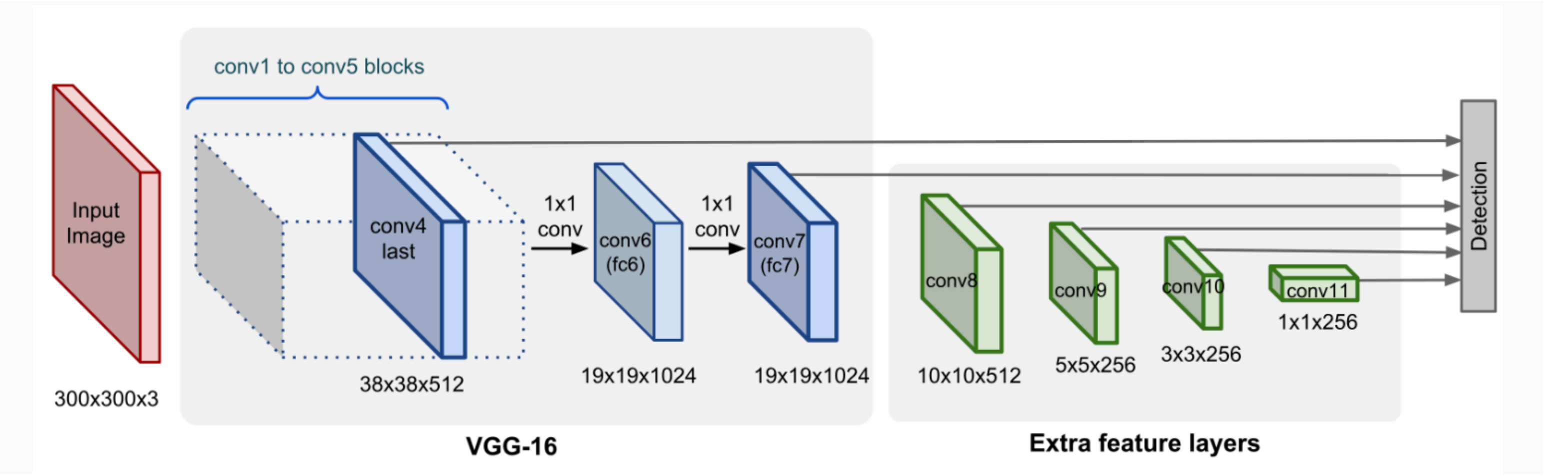


1-stage Object Detectors

- YOLO v1



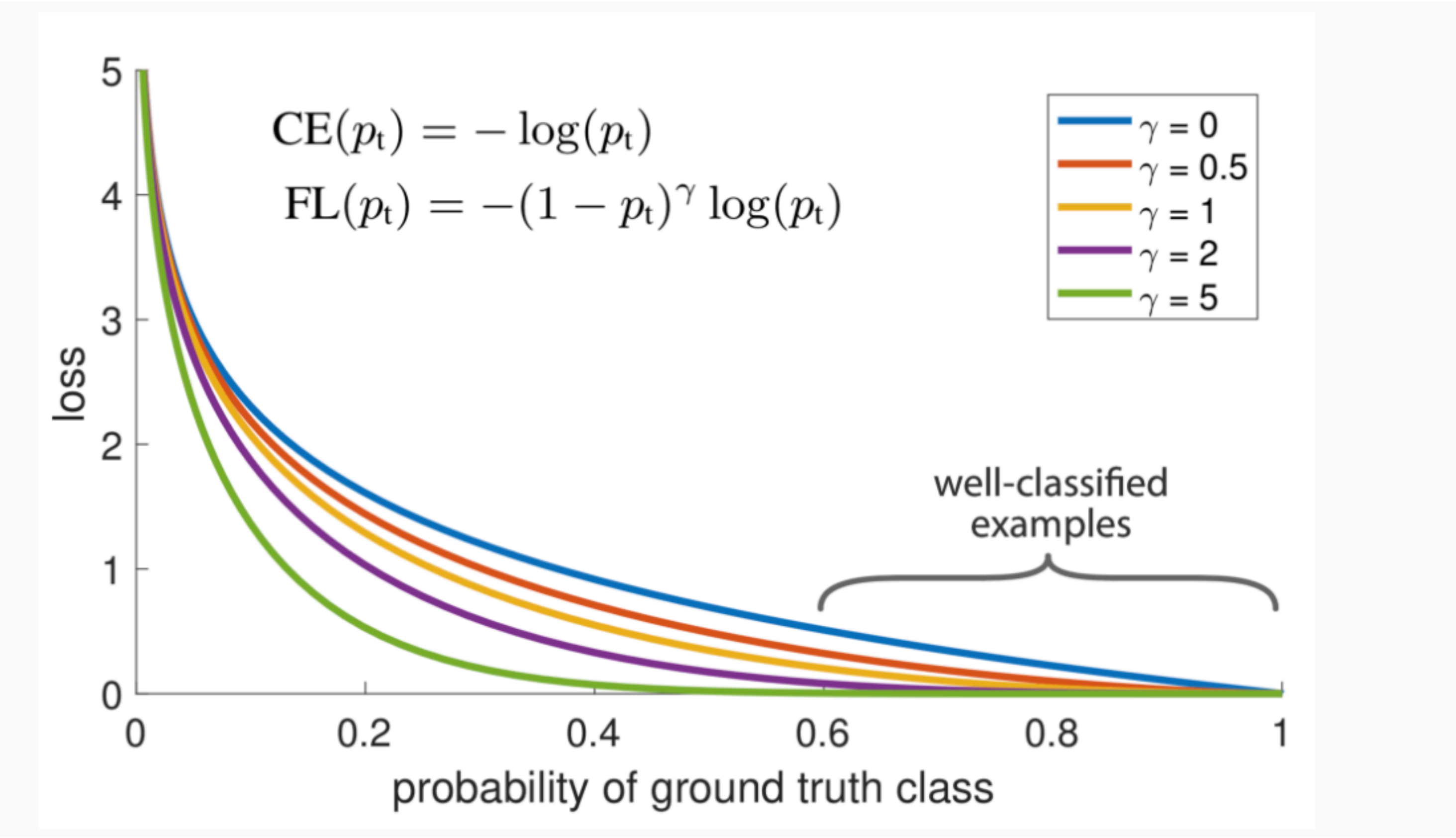
- SSD



$$\mathscr{L} = \frac{1}{N}\left(\mathscr{L}_{\text{cls}} + \alpha \mathscr{L}_{\text{loc}}\right)$$

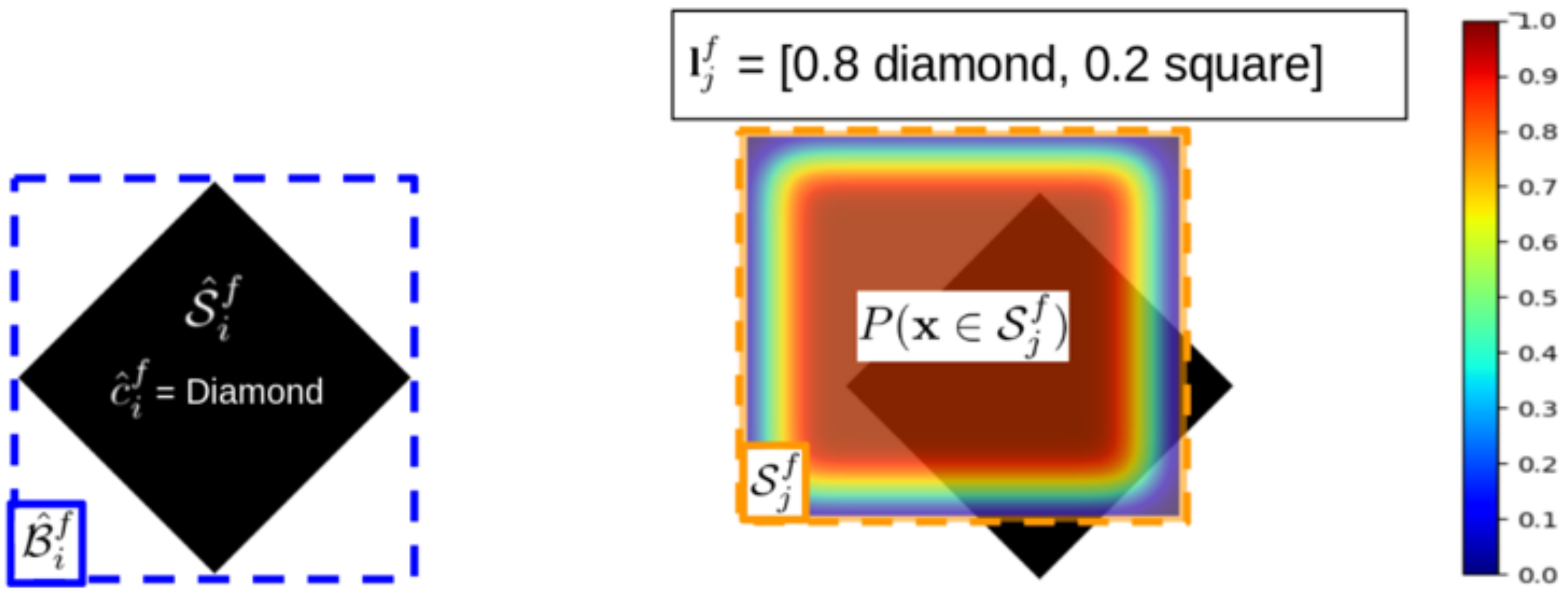
1-stage Object Detectors

- RetinaNet



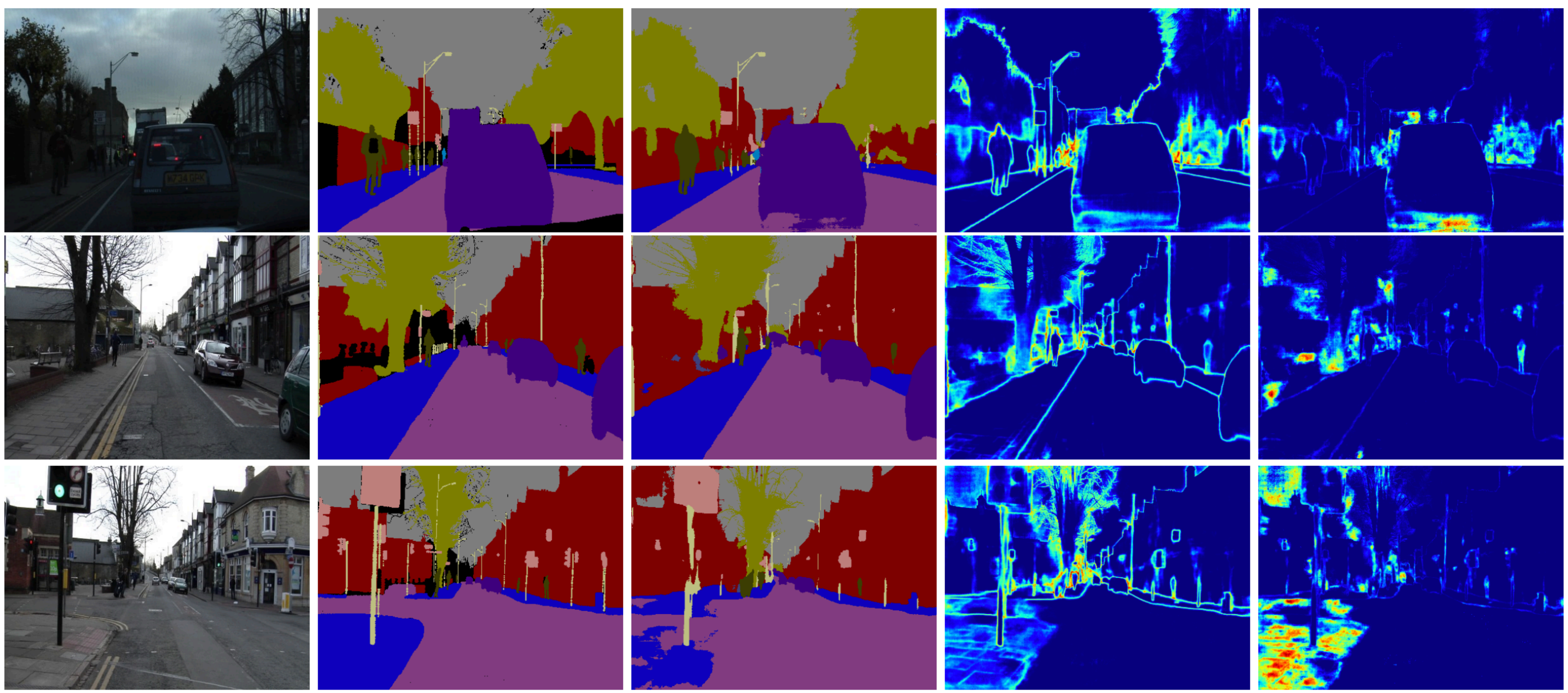
$$\mathrm{FL}\left(p_t\right) = -\left(1 - p_t\right)^\gamma \log p_t$$

Probabilistic Object Detection Challenge



Uncertainty

- Aleatoric vs. Epistemic Uncertainty



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Uncertainty

- Regression vs. Classification Uncertainty
  - In both cases this can be evaluated via entropy

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}\left(x_i\right) \log_b \mathrm{P}\left(x_i\right)$$

- This can be calculated from the *sufficient statistics*:
  - In the case of a Gaussian (regression) we need the mean and variance
  - In the case of a Categorical distribution (classification) we need the probabilities for each class

Variance Regression

$$L_{reg} = \frac{1}{N}\sum_{i=1}^{N} L\left(\mathbf{x}_i\right)$$

$$L\left(\mathbf{x}_i\right) = \frac{1}{2\sigma\left(\mathbf{x}_i\right)^2} \left\| \mathbf{y}_i - f\left(\mathbf{x}_i\right) \right\| + \frac{1}{2}\log\sigma\left(\mathbf{x}_i\right)^2$$

Loss function defined in Kendall and Gal NIPS 2017

- Regress the variance

Monte Carlo Dropout

- Do dropout during training and testing
- Since dropout performed at train and test time enables us to interpret the output as a sample from a distribution

BayesOD

Problem Formulation

Per anchor:

$$p(\mathcal{S}|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta})$$
$$p(\mathcal{B}|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta})$$

Joint:

$$p(\mathcal{S}|\mathcal{X}, \mathcal{D}, \boldsymbol{\theta})$$
$$p(\mathcal{B}|\mathcal{X}, \mathcal{D}, \boldsymbol{\theta})$$

Usually get from per anchor to joint via NMS

Estimating Epistemic Uncertainty

Want: $p(\boldsymbol{\theta} \mid \mathscr{D})$

$$p\left(\hat{\boldsymbol{y}}_i \mid \mathbf{x}_i, \mathscr{D}\right) = \int_\theta p\left(\hat{\boldsymbol{y}}_i \mid \mathbf{x}_i, \mathscr{D}, \boldsymbol{\theta}\right) p(\boldsymbol{\theta} \mid \mathscr{D}) d\boldsymbol{\theta}$$

Bounding Box Regression:

$$\boldsymbol{\mu}(\mathbf{x}_i) = \tfrac{1}{T} \textstyle\sum_{t=1}^{T} f(\mathbf{x}_i, \boldsymbol{\theta}_t)$$
$$\Sigma_e(\mathbf{x}_i) = \tfrac{1}{T} \left(\textstyle\sum_{t=1}^{T} f(\mathbf{x}_i, \boldsymbol{\theta}_t) f(\mathbf{x}_i, \boldsymbol{\theta}_t)^\top \right) - \boldsymbol{\mu}(\mathbf{x}_i)\boldsymbol{\mu}(\mathbf{x}_i)^\top$$

Categorical distribution parameter estimation:

$$\hat{p}_k = \frac{1}{T} \sum_{t=1}^{T} \mathrm{SoftMax}\left(g\left(\mathbf{x}_i, \boldsymbol{\theta}_t\right) \right)_k$$

Use entropy of these distributions as a proxy for epistemic uncertainty

Aleatoric Uncertainty

Slightly modified loss function:

$$L_{reg} = \frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} L\left(\mathbf{x}_i\right) + \frac{1}{N_{neg}} \sum_{i=1}^{N_{neg}} \frac{1}{\sigma\left(\mathbf{x}_i\right)^2}$$

Since we are doing MC Dropout need to average:

$$\Sigma_a(\mathbf{x}_i) = \mathrm{diag}\big(\big[\sigma^1, \ldots, \sigma^n(\mathbf{x}_i)\big]\big)$$
$$\sigma^j(\mathbf{x}_i) = \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \sigma^j(\mathbf{x}_i, \boldsymbol{\theta}_t)$$

Total uncertainty: $\Sigma\left(\mathbf{x}_i\right) = \Sigma_e\left(\mathbf{x}_i\right) + \Sigma_a\left(\mathbf{x}_i\right)$

Bayesian Inference

$$p\left(\mathscr{B} \,|\, \mathbf{x}_i, \mathscr{D}, \hat{\mathscr{B}}_i\right) \propto p\left(\hat{\mathscr{B}}_i \,|\, \mathbf{x}_i, \mathscr{D}, \mathscr{B}\right) p\left(\mathscr{B} \,|\, \mathbf{x}_i\right)$$

Bounding Box Regression:

$$\Sigma'(\mathbf{x}_i) = \left(\Sigma_0^{-1} + \Sigma(\mathbf{x}_i)^{-1}\right)^{-1}$$
$$\boldsymbol{\mu}'(\mathbf{x}_i) = \Sigma'(\mathbf{x}_i)\left(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma(\mathbf{x}_i)\boldsymbol{\mu}(\mathbf{x}_i)\right)$$

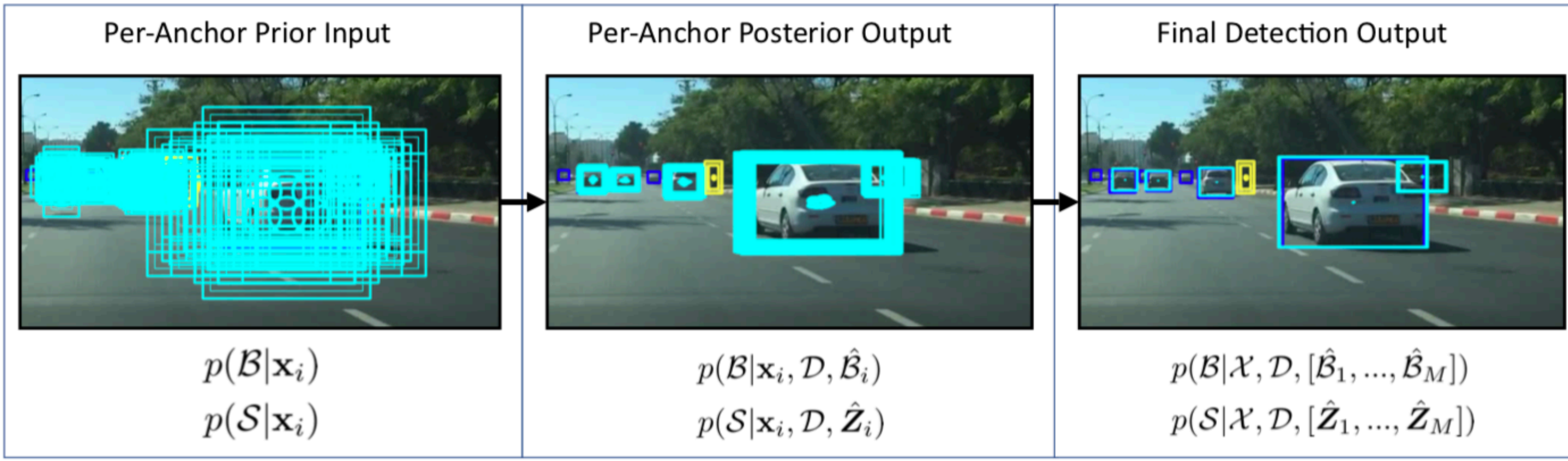
Categorical sufficient statistics estimation:

$$p\left(\mathscr{P} \,|\, \mathbf{x}_i, \mathscr{D}, \hat{\mathbf{Z}}_i\right) \propto p\left(\hat{\mathbf{Z}}_i \,|\, \mathbf{x}_i, \mathscr{D}, \mathscr{P}\right) p\left(\mathscr{P} \,|\, \mathbf{x}_i\right)$$

$$p\left(\mathscr{P} \,|\, \mathbf{x}_i, \mathscr{D}, \hat{\mathbf{Z}}_i\right) \propto \prod_{k=1}^{K} p_k^{\alpha_k - 1} \prod_{f=1}^{F} \prod_{k=1}^{K} p_k^{1\left[\hat{z}_{fk} = 1\right]}$$
$$= \mathrm{Dir}\left(\alpha'_1, \ldots, \alpha'_K\right)$$

$$p\left(\mathscr{S} \,|\, \mathbf{x}_i, \mathscr{D}, \mathbf{Z}_i\right) = \mathrm{Cat}\left(\left[p'_1, \ldots, p'_K\right]\right) \qquad p'_k = \frac{\alpha'_k}{\sum_{j=1}^{K} \alpha'_j}$$

Non-informative Priors



Bayesian Inference for NMS

- Per-anchor outputs from the neural network are **clustered** using spatial affinity
- Cluster center chosen greedily (pick anchor with largest non-background score)
- Add any anchor with IOU > 0.5 to the cluster
- Result: H anchor clusters, A, each containing an anchor set [a1,..aM] (M can vary with cluster)
- a1is the cluster centre, all others are considered as **measurements**

$$p\left(\mathcal{B} \mid \mathcal{X}, \mathcal{D}, \left[\hat{\mathcal{B}}_1, \ldots, \hat{\mathcal{B}}_M\right]\right) \propto p\left(\mathcal{B} \mid \mathbf{x}_1, \mathcal{D}, \hat{\mathcal{B}}_1\right) \prod_{i=2}^{M} p\left(\hat{\mathcal{B}}_i \mid \mathbf{x}_i, \mathcal{D}, \mathcal{B}\right)$$

Replacement for NMS

$$p\left(\mathscr{B} \mid \mathscr{X}, \mathscr{D}, \left[\hat{\mathscr{B}}_1, \ldots, \hat{\mathscr{B}}_M\right]\right) = \mathscr{N}\left(\boldsymbol{\mu}''(\mathscr{X}), \Sigma''(\mathscr{X})\right)$$

$$= \begin{array}{l} \Sigma''(\mathcal{X}) = \left(\sum_{i=1}^{M} \Sigma'(\mathbf{x}_i)^{-1}\right)^{-1} \\ \boldsymbol{\mu}''(\mathcal{X}) = \Sigma''(\mathcal{X})\left(\sum_{i=1}^{M} \Sigma'(\mathbf{x}_i)^{-1} \boldsymbol{\mu}'(\mathbf{x}_i)\right) \end{array}$$

$$p\left(\mathscr{P} \mid \mathbf{x}_i, \mathscr{D}, \left[\hat{\mathbf{Z}}_1, \ldots, \hat{\mathbf{Z}}_M\right]\right) \propto p\left(\mathscr{P} \mid \mathbf{x}_1, \mathscr{D}, \hat{\mathbf{Z}}_1\right) \prod_{i=2}^{M} p\left(\hat{\mathbf{Z}}_i \mid \mathbf{x}_i, \mathscr{D}, \mathscr{P}\right)$$

$$= \mathrm{Dir}\left(\alpha_1'', \ldots, \alpha_K''\right)$$

$$p\left(\mathscr{S} \mid \mathscr{X}, \mathscr{D}, \left[\hat{Z}_1, \ldots, \hat{Z}_M\right]\right) = \mathrm{Cat}\left(p_1'', \ldots, p_K''\right)$$

$$p_k'' = \frac{\alpha_k''}{\sum_{j=1}^{K} \alpha_j''}$$

Uncertainty Metric

- Minimum Uncertainty Error (MUE): determine the ability of an uncertainty measure to discriminate between true positives and false positives (true = IOU>0.5 and same category)

$$UE(\delta) = 0.5\frac{|TP > \delta|}{|TP|} + 0.5\frac{|FP \leq \delta|}{|FP|}$$

- MUE is the best uncertainty error achievable at the best value of delta

Results

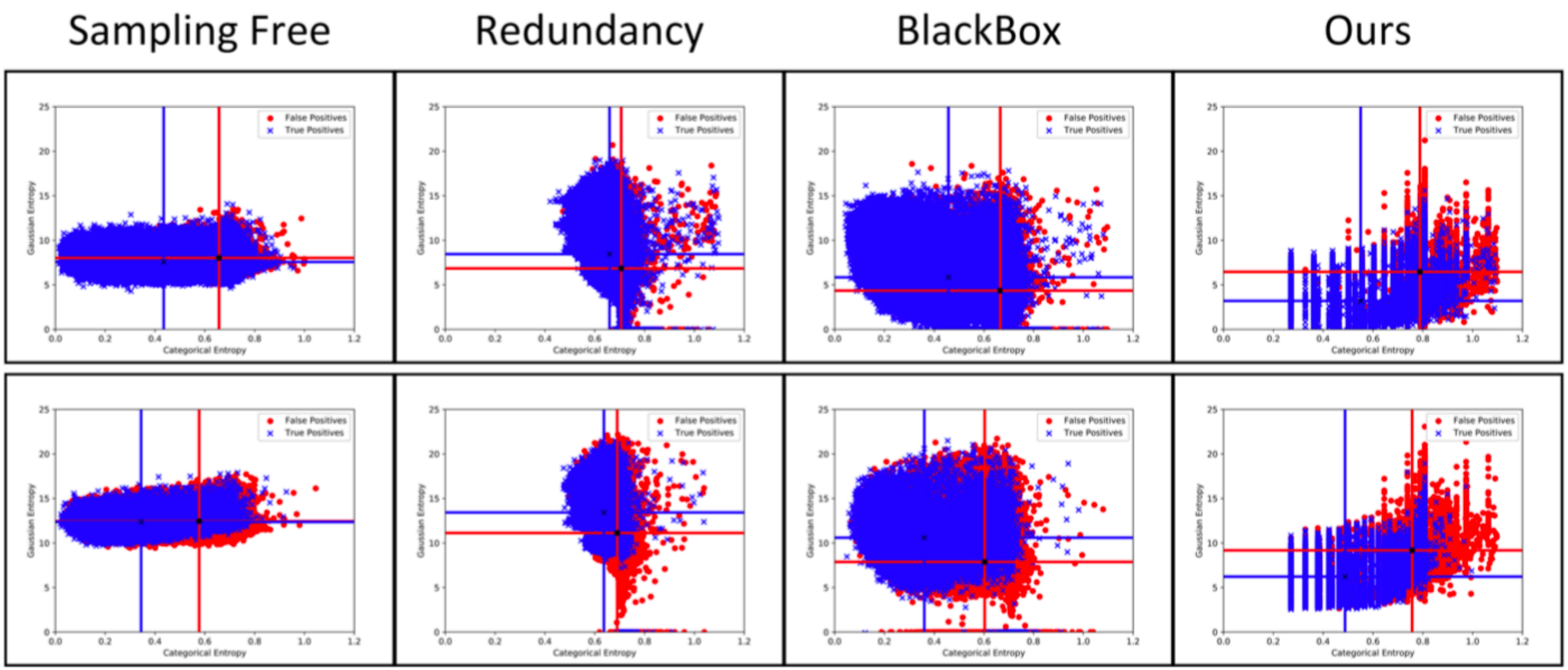


Fig. 3. Plots of the entropy of the Gaussian distribution representing the state $\mathcal{B}$, vs the entropy of the categorical distribution representing the category state $\mathcal{S}$ of true positives (in Blue) and false positives (in Red), combined over the *car* and *pedestrian* categories. The mean of both types of entropy is shown for the true positives and the false positives, as vertical and horizontal lines extending from the respective axes, and colored accordingly. The results of testing on the BDD dataset are shown in the top row, while those of testing on the KITTI dataset are shown in the bottom row. BayesOD is shown to be the only method that provides a substantially lower entropy of the state $\mathcal{B}$ for the true positives over the false positives on both datasets.

Results

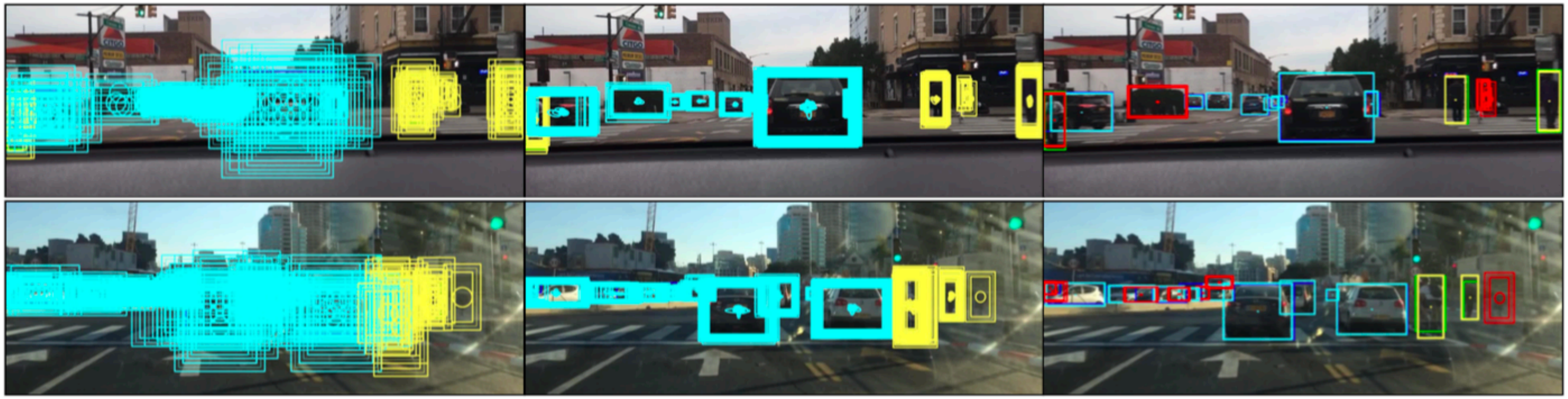


Fig. 4. Qualitative results using the same color scheme as Fig. 2, showing the progression of object state along BayesOD. **Left:** Object state non-informative priors represented by the anchor grid. Although the category state is uniform, prior boxes have been colored according to estimated category to help with correlation. **Middle:** The per-anchor posterior states, updated with the output from the neural network. **Right:** The final detection output, updated with information from spatially clustered anchors. Bounding boxes shown in *red* are rejected by the neural network for having a high entropy for both states.