BayesOD: Bayesian Inference for Fusing Epistemic and Aleatoric Uncertainty in Deep Object Detectors

Ali Harakeh, and Steven L. Waslander

Toronto Robotics and Artificial Intelligence Laboratory

University of Toronto Institute for Aerospace Studies

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

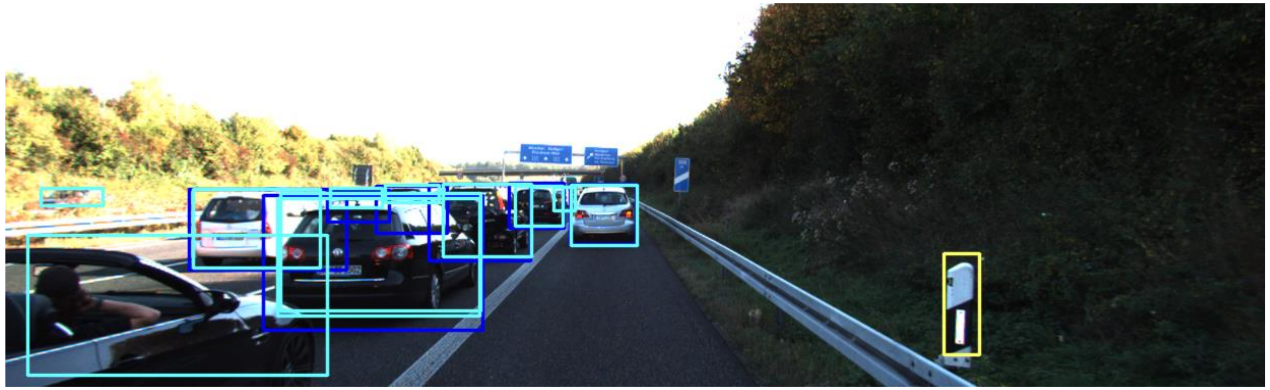
TRAIL
Toronto Robotics and AI Laboratory

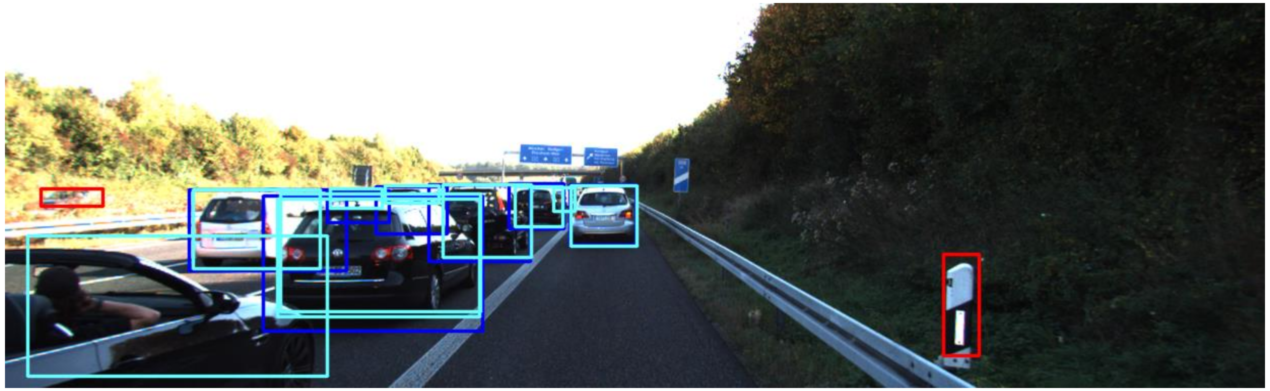
Goals:

- Provide uncertainty estimates for the **category** and **bounding box** states associated with detected object instances.

- Uncertainty estimates should be **meaningful,** and well correlated to the correctness of a detection.

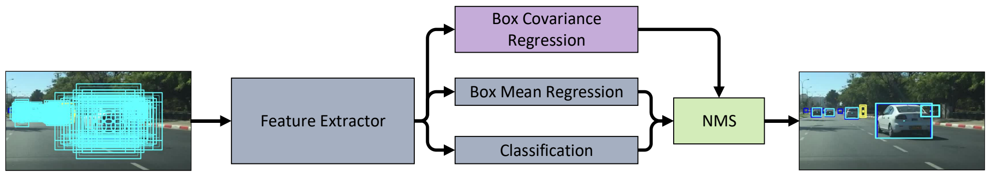
- Allow the treatment of the neural network as a sensor within the robot sensor suite.





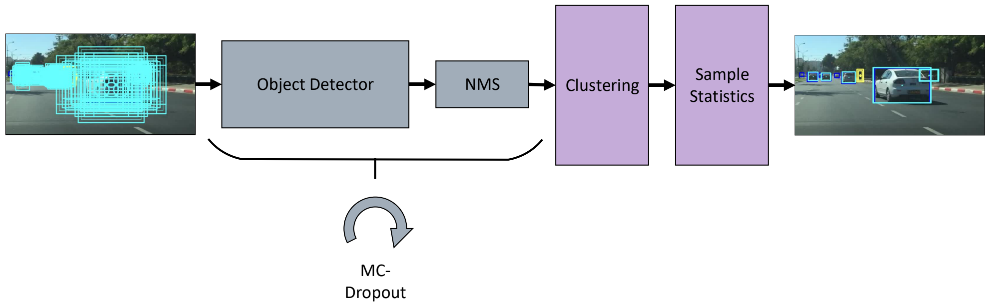
Related Work

Sample-Free [Feng, Rosenbaum, Dietmeyer 2018], [Le, Diehl, Brunner, Knol 2018]



- A single instance detector with box covariance regression

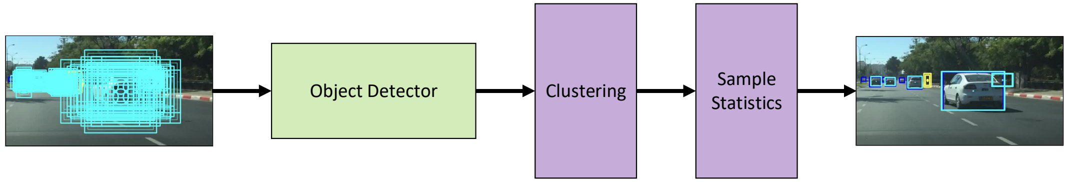
Black Box [Miller, Nicholson, Dayoub, Sünderhauf 2018], [Miller, Dayoub, Milford, Sünderhauf 2018]



- MC-dropout with T instances
- Cluster after NMS for each instance and estimate sample mean and covariance

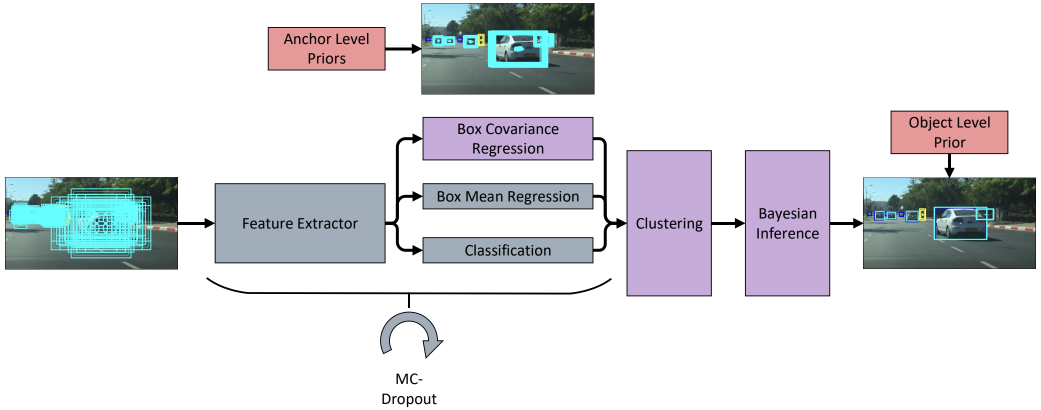
Related Work

Redundancy [Le, Diehl, Brunner, Knoll 2018]



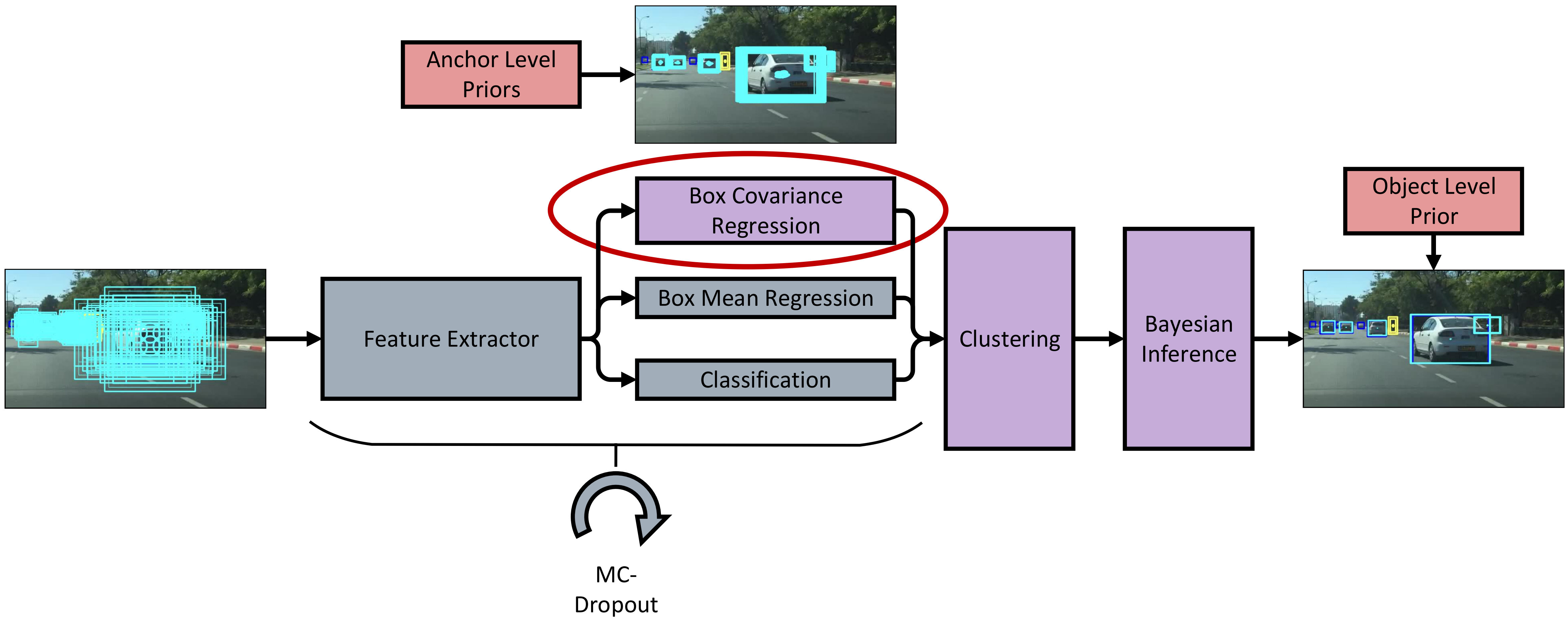
- Replace NMS with anchor output clustering and sample mean and covariance estimation

BayesOD [Harakeh, Waslander 2019]



- Combine box covariance regression, MC dropout and clustering for joint aleatoric and epistemic uncertainty estimation
- Incorporate Dirichlet and Gassian priors on classification and regression of anchors and objects
- Perform Bayesian inference over anchor and MC dropout clusters, replacing NMS

Bayes OD – Box Covariance Regression



Estimating Per-Anchor Aleatoric Uncertainty:

Given

$$f(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbb{E}[p(\hat{\mathcal{B}}_i | \mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta})]$$

$$p(\hat{\mathcal{S}}_i | \mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) = Cat(\hat{p}_1, \dots, \hat{p}_K)$$

Required

$$p(\hat{\mathcal{B}}_i | \mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) = \mathcal{N}(\mu(\mathbf{x}_i, \boldsymbol{\theta}), \Sigma_a(\mathbf{x}_i, \boldsymbol{\theta}))$$

$$p(\hat{\mathcal{S}}_i | \mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) = Cat(\hat{p}_1, \dots, \hat{p}_K)$$

- Aleatoric uncertainty already captured by the parameters of the categorical distribution.
- The only requirement is to estimate the covariance matrix of the Gaussian distribution describing objects' bounding boxes.

Learning to Estimate a Multivariate Covariance Matrix:

- Extension to stable multi-variate covariance training through LDL matrix decomposition:

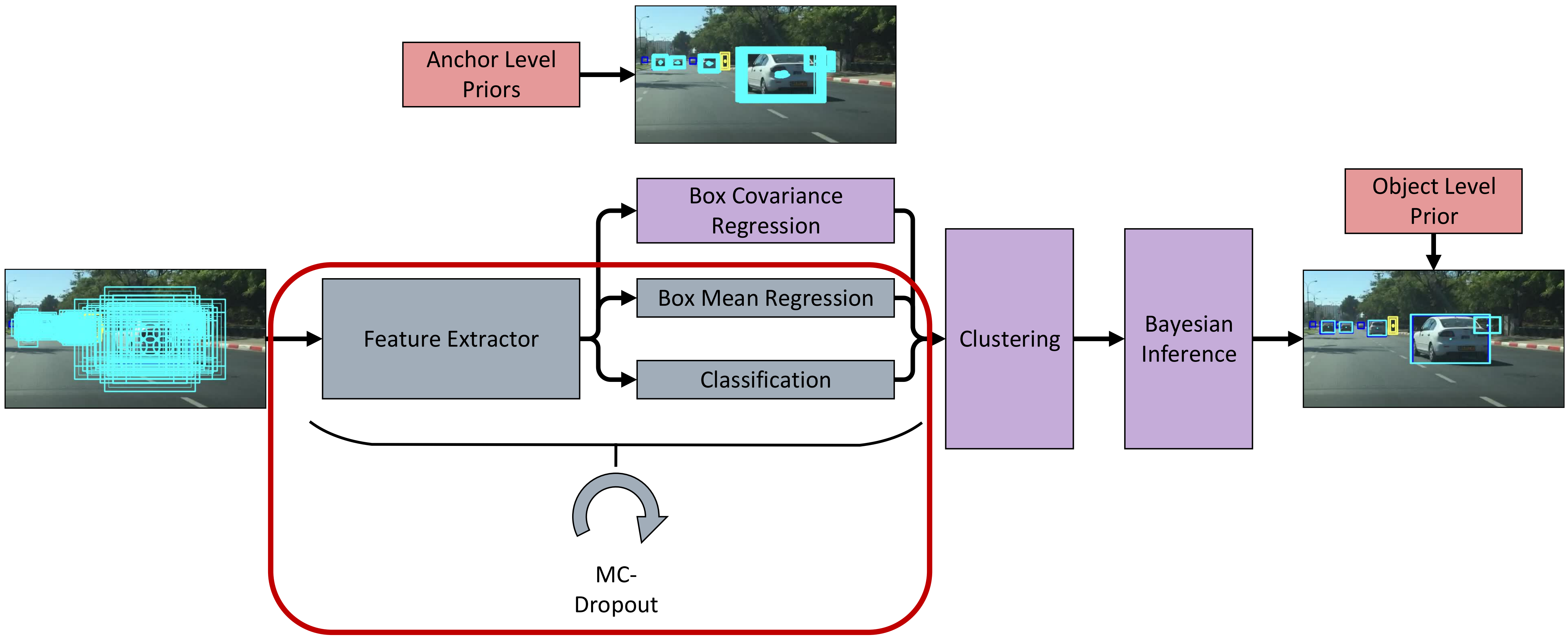
$$\Sigma_a(\mathbf{x}_i, \boldsymbol{\theta}) = L(\mathbf{x}_i, \boldsymbol{\theta}) D(\mathbf{x}_i, \boldsymbol{\theta}) L(\mathbf{x}_i, \boldsymbol{\theta})^\intercal$$

$$\begin{matrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{matrix} \qquad\qquad \begin{matrix} e & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & g \end{matrix}$$

- Learning the components through cross-(differential) entropy loss function:

$$\min\ L(\mathbf{x}_i, \boldsymbol{\theta}) = \frac{1}{2}\left[|| D(\mathbf{x}_i, \boldsymbol{\theta})^{-1/2} \mathrm{adj}(L(\mathbf{x}_i, \boldsymbol{\theta})) \mu(\mathbf{x}_i, \boldsymbol{\theta}) - \mu_{gt} ||_2^2 + \mathrm{Tr}(\log(D(\mathbf{x}_i, \boldsymbol{\theta}))) \right]$$

Bayes OD – Epistemic Uncertainty



Estimating Per-Anchor Epistemic Uncertainty

Given

$$p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) = \mathcal{N}(\mu(\mathbf{x}_i, \boldsymbol{\theta}), \Sigma_a(\mathbf{x}_i, \boldsymbol{\theta}))$$

$$p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) = Cat(\hat{p}_1, \dots, \hat{p}_K)$$

Required

$$p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}) = \mathcal{N}(\mu(\mathbf{x}_i), \Sigma(\mathbf{x}_i))$$

$$p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}) = Cat(\hat{p}_1, \dots, \hat{p}_K)$$

- Use Bayesian Neural Networks to eliminate the dependence on a point estimate of the parameters theta:

$$p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}) = \int_\theta p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}$$

$$p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}) = \int_\theta p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}$$

Estimating Per-Anchor Epistemic Uncertainty

- Monte-Carlo Dropout as variational Bayesian approximation of the integral:

Regression

$$p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}) = N(\mu(\mathbf{x}_i), \Sigma(\mathbf{x}_i))$$

$$\mu(\mathbf{x}_i) = \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_i, \boldsymbol{\theta}_t)$$

$$\Sigma(\mathbf{x}_i) = \frac{1}{T}\left(\sum_{t=1}^{T} f(\mathbf{x}_i, \boldsymbol{\theta}_t) f(\mathbf{x}_i, \boldsymbol{\theta}_t)^\intercal\right) - \mu(\mathbf{x}_i)\mu(\mathbf{x}_i)^\intercal + \frac{1}{T}\sum_{t=1}^{T} \Sigma_a(\mathbf{x}_i, \boldsymbol{\theta}_t)$$

Classification

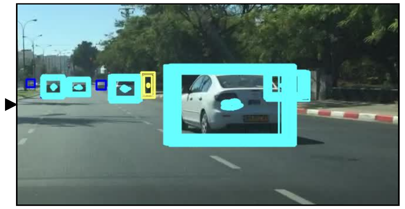
$$p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}) = Cat([\hat{p}_1, \dots, \hat{p}_K])$$

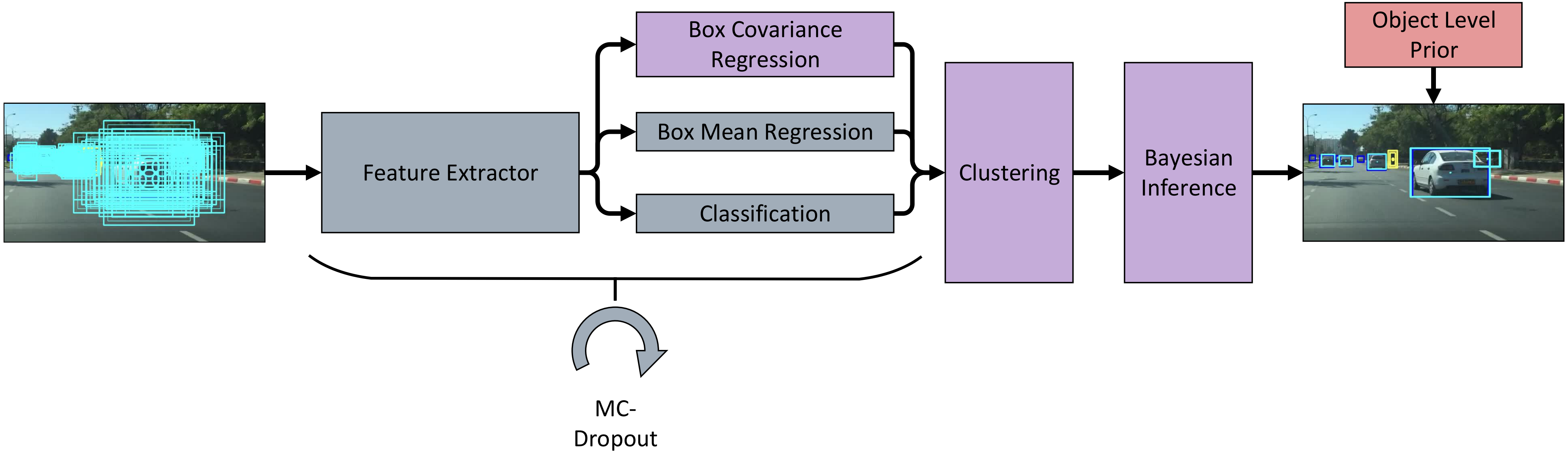
$$\hat{p}_k = \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{x}_i, \boldsymbol{\theta}_t)$$

Bayes OD - Per-Anchor Probability Distributions

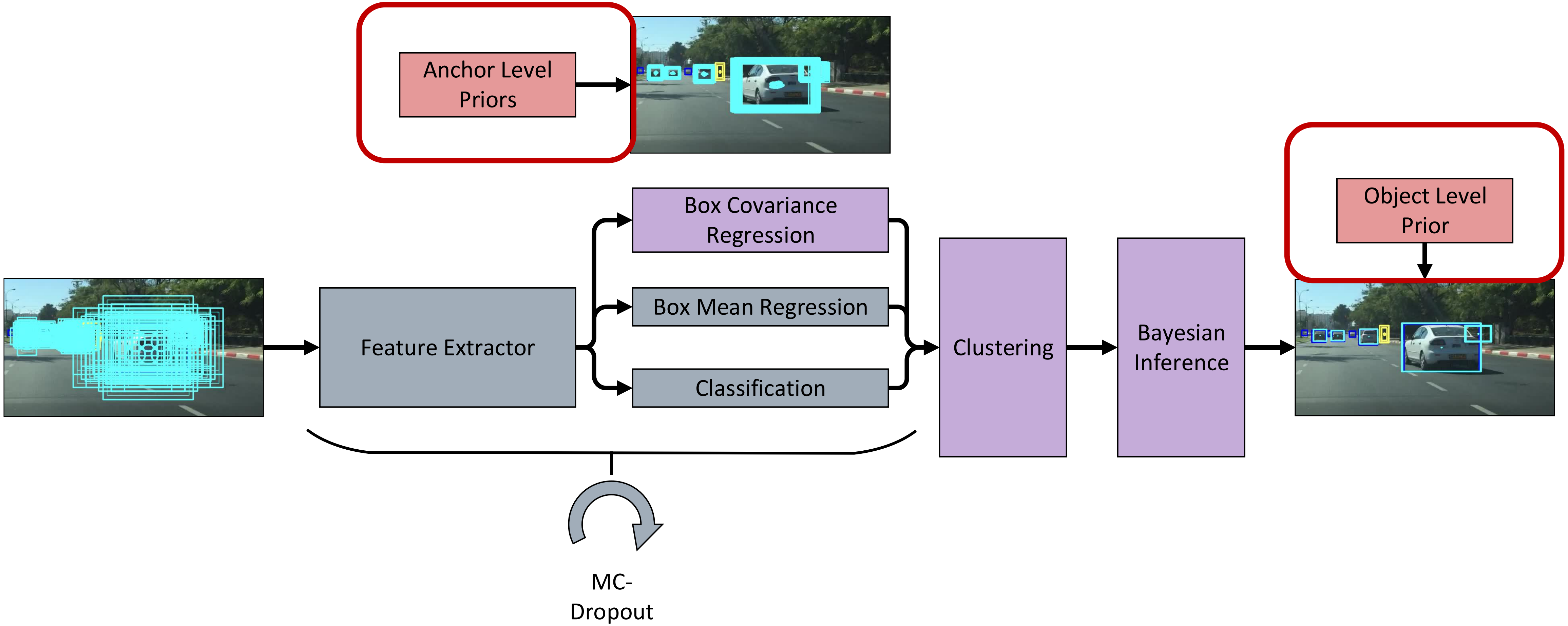
$$p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}) = \mathcal{N}(\mu(\mathbf{x}_i), \Sigma(\mathbf{x}_i))$$

$$p(\hat{\mathcal{S}}_i|\mathbf{x}_i, \mathcal{D}) = Cat(\hat{p}_1, \ldots, \hat{p}_K)$$





Bayes OD – Incorporating Priors



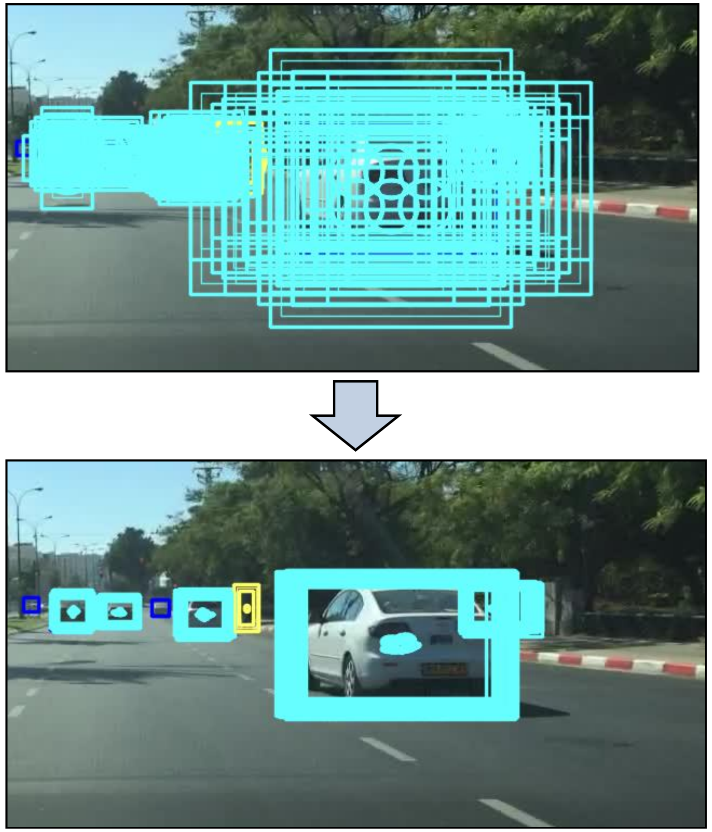
Bounding Box Per-Anchor Posterior

- The bounding box per anchor posterior distribution can be written as:

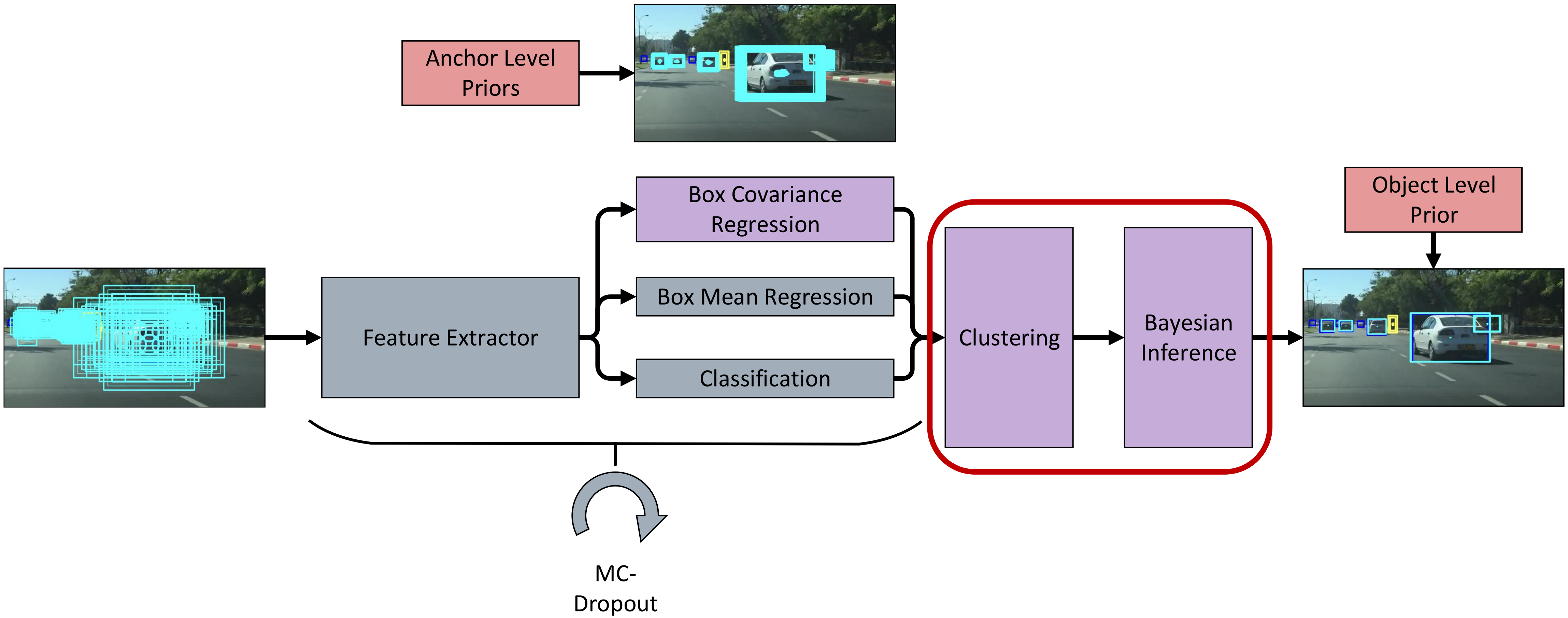
$$p(\mathcal{B}|\mathbf{x}_i, \mathcal{D}, \hat{\mathcal{B}}_i) \propto \underbrace{p(\hat{\mathcal{B}}_i|\mathbf{x}_i, \mathcal{D}, \mathcal{B})}_{N(\boldsymbol{\mu}(\mathbf{x}_i), \Sigma(\mathbf{x}_i))} \underbrace{p(\mathcal{B}|\mathbf{x}_i)}_{\mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)}$$

- Can be computed in closed form as:

$$p(\mathcal{B}|\mathbf{x}_i, \mathcal{D}, \hat{\mathcal{B}}_i) = \mathcal{N}(\boldsymbol{\mu}'(\mathbf{x}_i), \Sigma'(\mathbf{x}_i))$$
$$\Sigma'(\mathbf{x}_i) = (\Sigma_0^{-1} + \Sigma(\mathbf{x}_i)^{-1})^{-1}$$
$$\boldsymbol{\mu}'(\mathbf{x}_i) = \Sigma'(\mathbf{x}_i)(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma(\mathbf{x}_i)\boldsymbol{\mu}(\mathbf{x}_i)).$$



Bayes OD – Incorporating Priors



Bayesian Inference Over Object Clusters

- **Step 1:** Cluster detections using your favorite clustering algorithm [Mill,er Dayoub, Milford, Sünderhauf, 2019].

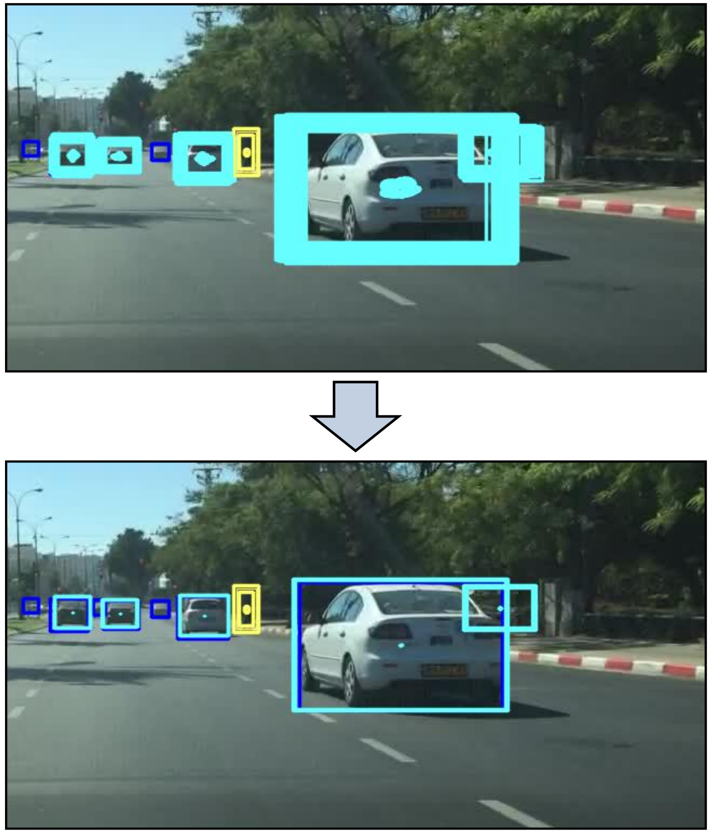
- **Step 2:** Use cluster members to update the states of the center of the cluster. **Assume independent measurements.**

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

$$= \mathcal{N}(\boldsymbol{\mu}''(\mathcal{X}),\Sigma''(\mathcal{X}))$$

$$\Sigma''(\mathcal{X}) = \left(\sum_{i=1}^{M}\Sigma'(\mathbf{x}_i)^{-1}\right)^{-1}$$

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



- **Alternative Step 2:** Inverse Covariance Intersection to estimate correlation between anchor measurements

Results - MUE and AP

- We used RetinaNet as 2D object detector
- Trained on BDD, tested on BDD, KITTI
- Minimum Uncertainty Error Evaluation:
  - Categorical Minimum Uncertainty Error (CMUE): Lowest possible uncertainty error using a threshold on the categorical entropy.
  - Gaussian Minimum Uncertainty Error (GMUE): Lowest possible uncertainty error using a threshold on the gaussian entropy.

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

Results - MUE and AP

- Bayesian inference treats anchors as separate measurements, and fuses them
- Leads to better estimates and higher AP
- Significant improvement in both Gaussian and Categorical Minimum Uncertainty Error

		Car			Pedestrian		
Test Dataset	Method	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓
BDD [21]	Sampling Free [16, 15]	55.16	38.99	21.96	37.64	47.49	30.55
	Black Box [13, 14]	57.34	49.75	21.71	41.54	49.86	29.43
	Redundancy [16]	56.43	49.71	24.80	40.43	49.96	38.56
	Ours (Diagonal)	**60.98**	25.76	16.67	**42.97**	**26.68**	**22.72**
	Ours (Full Covar)	60.79	**25.64**	**16.50**	42.05	27.25	23.02

Results - Inverse Covariance Intersection

- Bayesian inference treats anchors as separate measurements, and fuses them
- Leads to better estimates and higher AP
- Significant improvement in both Gaussian and Categorical Minimum Uncertainty Error
- Inverse Covariance Intersection leads to reduced AP, increased GMUE

		Car			Pedestrian		
Test Dataset	Method	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓
BDD [21]	Sampling Free [16, 15]	55.16	38.99	21.96	37.64	47.49	30.55
	Black Box [13, 14]	57.34	49.75	21.71	41.54	49.86	29.43
	Redundancy [16]	56.43	49.71	24.80	40.43	49.96	38.56
	Ours (Diagonal)	**60.98**	25.76	16.67	**42.97**	**26.68**	**22.72**
	Ours (Full Covar)	60.79	**25.64**	**16.50**	42.05	27.25	23.02
	Ours (Full Covar + ICI)	60.63	36.04	16.83	41.65	40.00	23.76

PDQ Score

- PDQ results show some complex behaviours
  - PDQ spatial quality does not seem to be positively correlated to minimum uncertainty error or AP performance in these comparisons.
  - ICI has a strong positive influence on spatial quality

Method	Score ↑	TP ↑	FP ↓	FN ↓	Spatial Quality(%) ↑	Label Quality(%) ↑	Overall Quality(%) ↑
Sampling Free	**32.98**	71597	**12608**	36897	45.47	**81.78**	**55.79**
Black Box	27.80	69274	13709	39220	39.85	78.47	49.04
Redundancy	28.24	70323	13916	38171	46.57	63.44	49.17
Ours (Diag)	17.92	73757	31881	35059	24.16	77.82	34.20
Ours (Full Covar)	17.84	73453	31977	35342	24.16	77.78	34.20
Ours (Full Covar + ICI)	31.20	**76207**	25464	**32588**	**46.93**	77.52	54.97

		Car			Pedestrian		
Test Dataset	Method	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓	AP(%) ↑	GMUE(%) ↓	CMUE(%) ↓
BDD [21]	Sampling Free [16, 15]	55.16	38.99	21.96	37.64	47.49	30.55
	Black Box [13, 14]	57.34	49.75	21.71	41.54	49.86	29.43
	Redundancy [16]	56.43	49.71	24.80	40.43	49.96	38.56
	Ours (Diagonal)	**61.35**	**25.53**	16.96	**43.62**	**26.15**	**23.56**
	Ours (Diagonal + ICI)	60.59	36.94	16.87	41.28	41.68	23.60
	Ours (Full Covar + ICI)	60.63	36.04	**16.83**	41.65	40.00	23.76

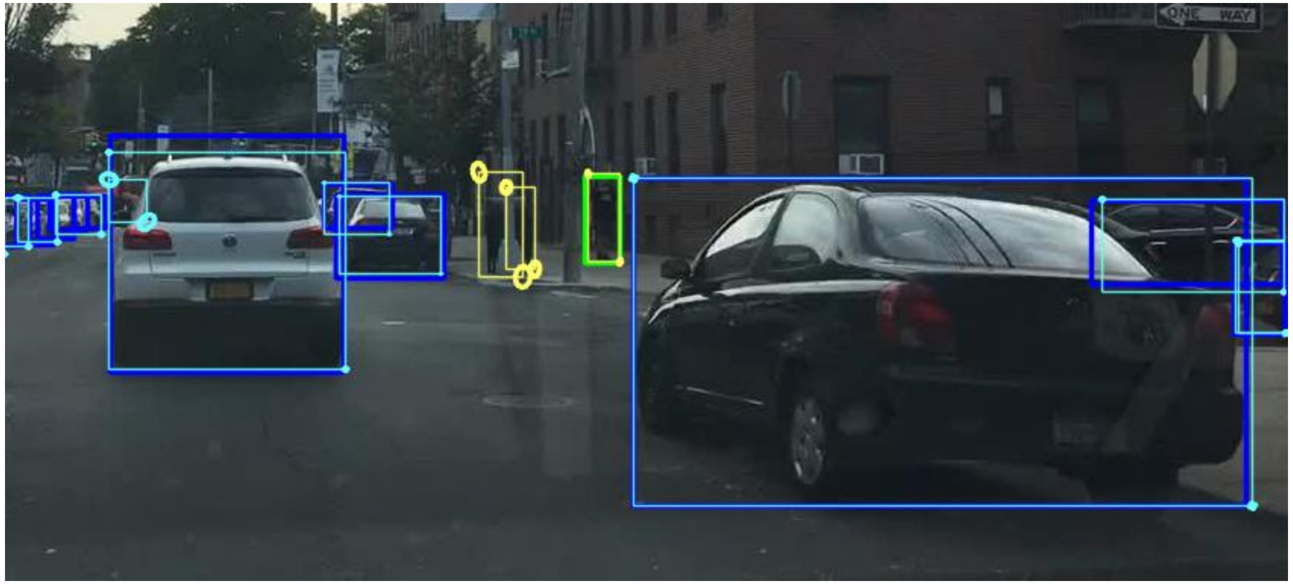
PDQ Score

- Can achieve 100% label quality by replacing predicted category probability vector with a one hot vector.
- We have a few ideas to tackle this issue, mainly by incorporating the predicted probability of false positives into the label quality.

Method	Score ↑	TP ↑	FP ↓	FN ↓	Spatial Quality(%) ↑	Label Quality(%) ↑	Overall Quality(%) ↑
Sampling Free	**35.79**	71197	**12503**	37297	45.62	100.00	60.83
Black Box	30.45	68999	13524	39495	39.99	100.00	53.86
Redundancy	34.84	69987	13730	38507	46.73	100.00	60.85
Ours (Full Covar + ICI)	34.60	**75897**	25784	**32898**	**47.11**	100.00	**61.36**

Next Steps

- Further investigate the effect of correlated measurements.
- Uninformative priors only the first step, dataset based anchor priors and tracking based object priors to be exploited.
- Alternatives to MC Dropout for epistemic uncertainty estimation



Extra Slides

Quantitative Representations

Quantitative Representations

Quantitative Representations (Kitti Dataset)