STEVENS INSTITUTE OF TECHNOLOGY

Learning the Distribution of Errors in Stereo Matching for Joint Disparity and Uncertainty Estimation

Liyan Chen, Weihan Wang, Philippos Mordohai

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Email: lchen39@stevens.edu

Poster Section： THU-AM-072

Poster Location： West Building Exhibit Halls ABC 072

Implementation: https://github.com/lly00412/SEDNet

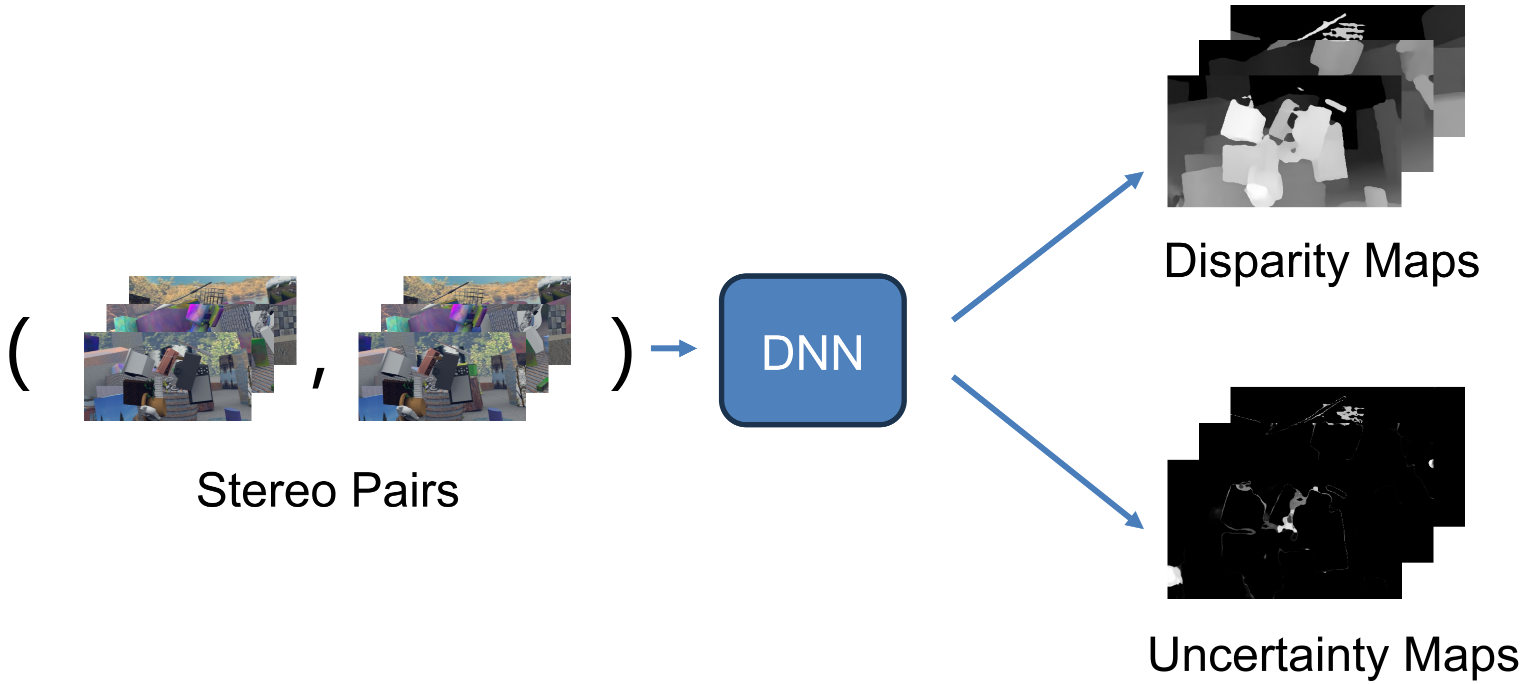
Contributions

- A novel joint estimation network, **SEDNet** *(Stereo Error Distribution Network)* predicts **disparity** as well as the **aleatoric uncertainty**.

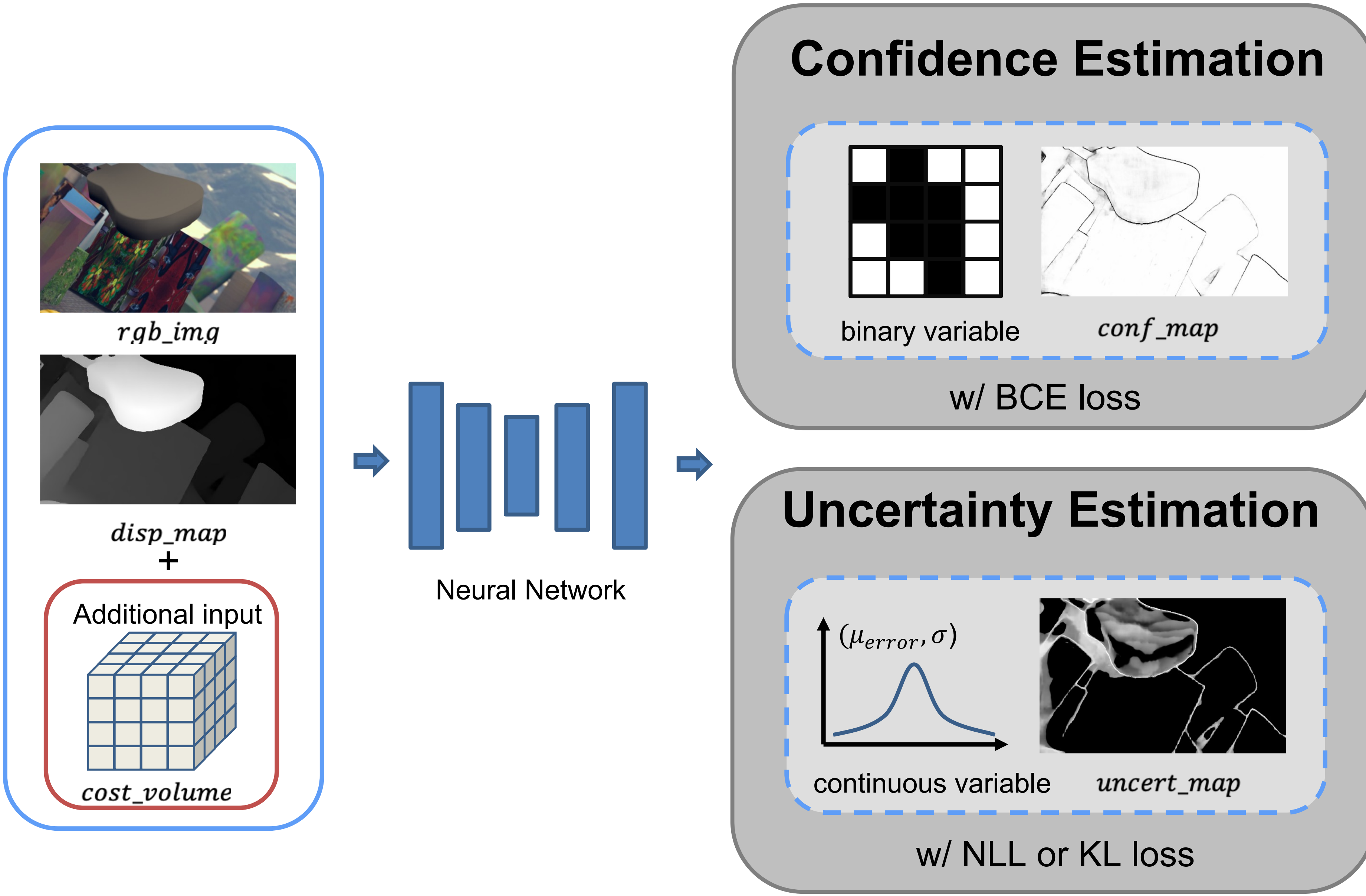
- A **differentiable soft-histogramming** technique used to **approximate the distributions** of disparity errors and estimated uncertainties.

- A **matching error loss** based on KL divergence applied on histograms obtained with the above technique to **improve the precision of uncertainty estimation**.

Introduction

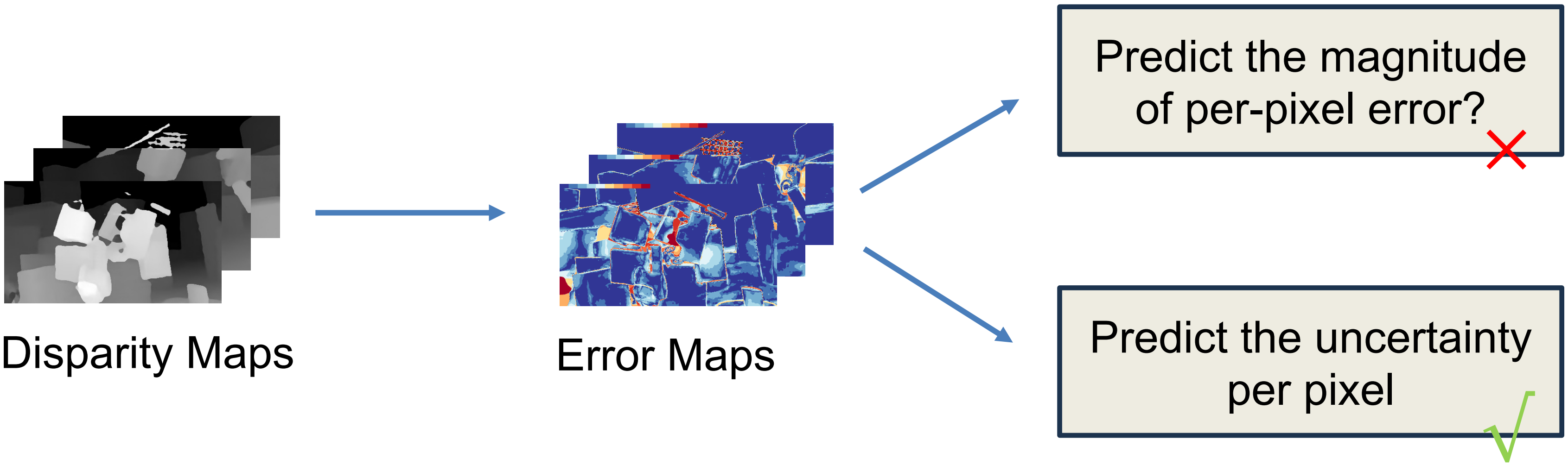


Joint estimation of disparity and uncertainty / confidence benefits both tasks due to **multi-task learning**.

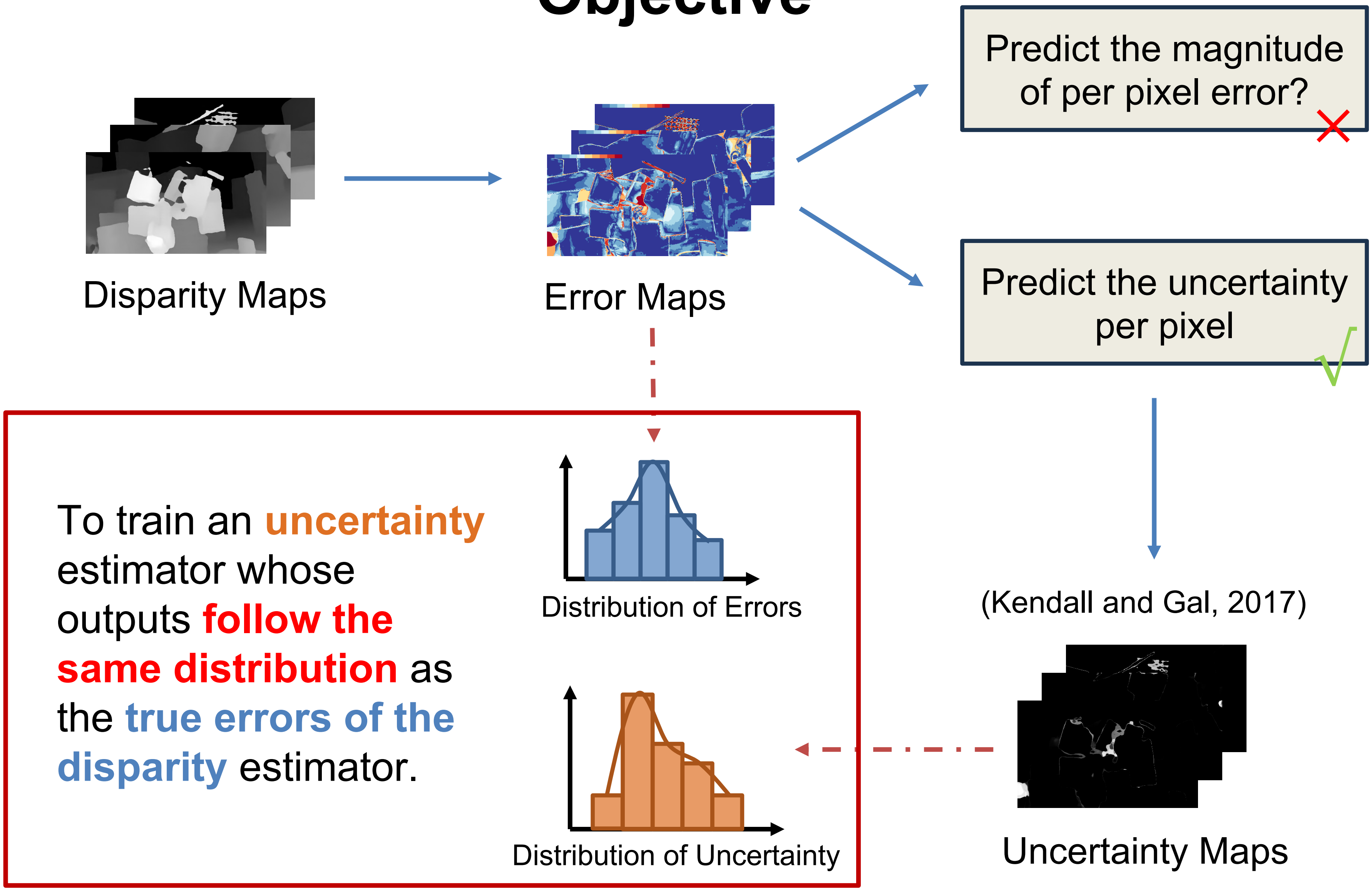


rgb_img
disp_map
+
Additional input
cost_volume
Neural Network
Confidence Estimation
binary variable
conf_map
w/ BCE loss
Uncertainty Estimation
(μ_{error}, σ)
continuous variable
uncert_map
w/ NLL or KL loss

Problem


Disparity Maps
Error Maps
Predict the magnitude of per-pixel error?
Predict the uncertainty per pixel

Objective


Disparity Maps
Error Maps
Predict the magnitude of per pixel error?
Predict the uncertainty per pixel
(Kendall and Gal, 2017)
Uncertainty Maps
To train an uncertainty estimator whose outputs follow the same distribution as the true errors of the disparity estimator.
Distribution of Errors
Distribution of Uncertainty

Aleatoric Uncertainty Estimation

(Kendall and Gal, 2017) & (Ilg et al., 2018)

Minimizing **NLL loss** per pixel

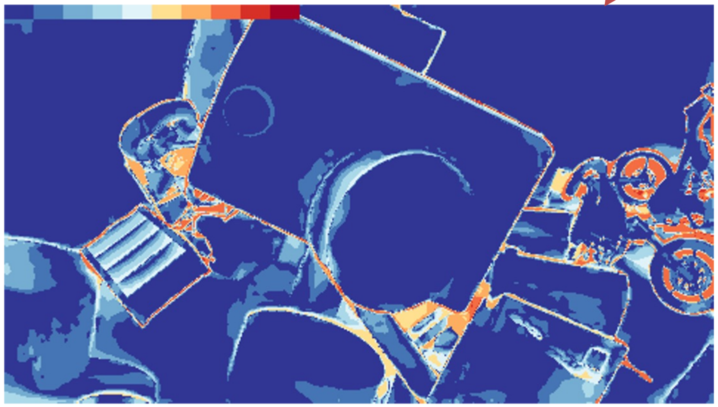
$$\mathcal{L}_{log} = \frac{1}{n}\sum_{i=1}^{n} \frac{|\hat{d}^{(i)} - d^{(i)}|}{\exp(s^{(i)})} + \frac{1}{n}\sum_{i=1}^{n} s^{(i)}$$

Aleatoric Uncertainty Estimation

(Kendall and Gal, 2017) & (Ilg et al., 2018)

Minimizing **NLL loss** per pixel

$$\mathcal{L}_{log} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\boxed{\hat{d}^{(i)} - d^{(i)}}\right|}{\exp(s^{(i)})} + \frac{1}{n}\sum_{i=1}^{n} s^{(i)}$$



Error Map

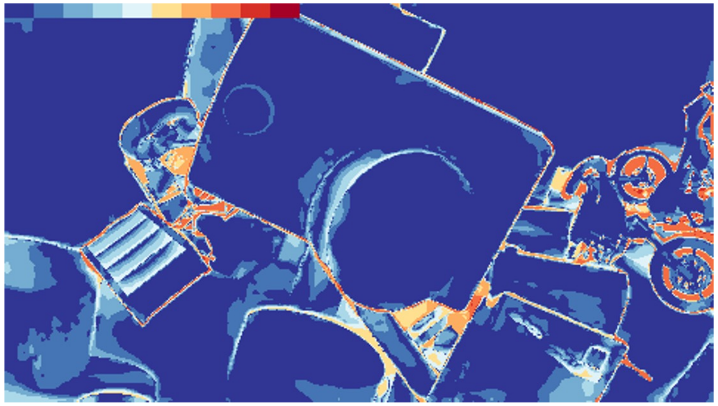
Aleatoric Uncertainty Estimation

(Kendall and Gal, 2017) & (Ilg et al., 2018)

Minimizing **NLL loss** per pixel

$$\mathcal{L}_{log} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{d}^{(i)} - d^{(i)}|}{\boxed{\exp(s^{(i)})}} + \frac{1}{n}\sum_{i=1}^{n} s^{(i)}$$

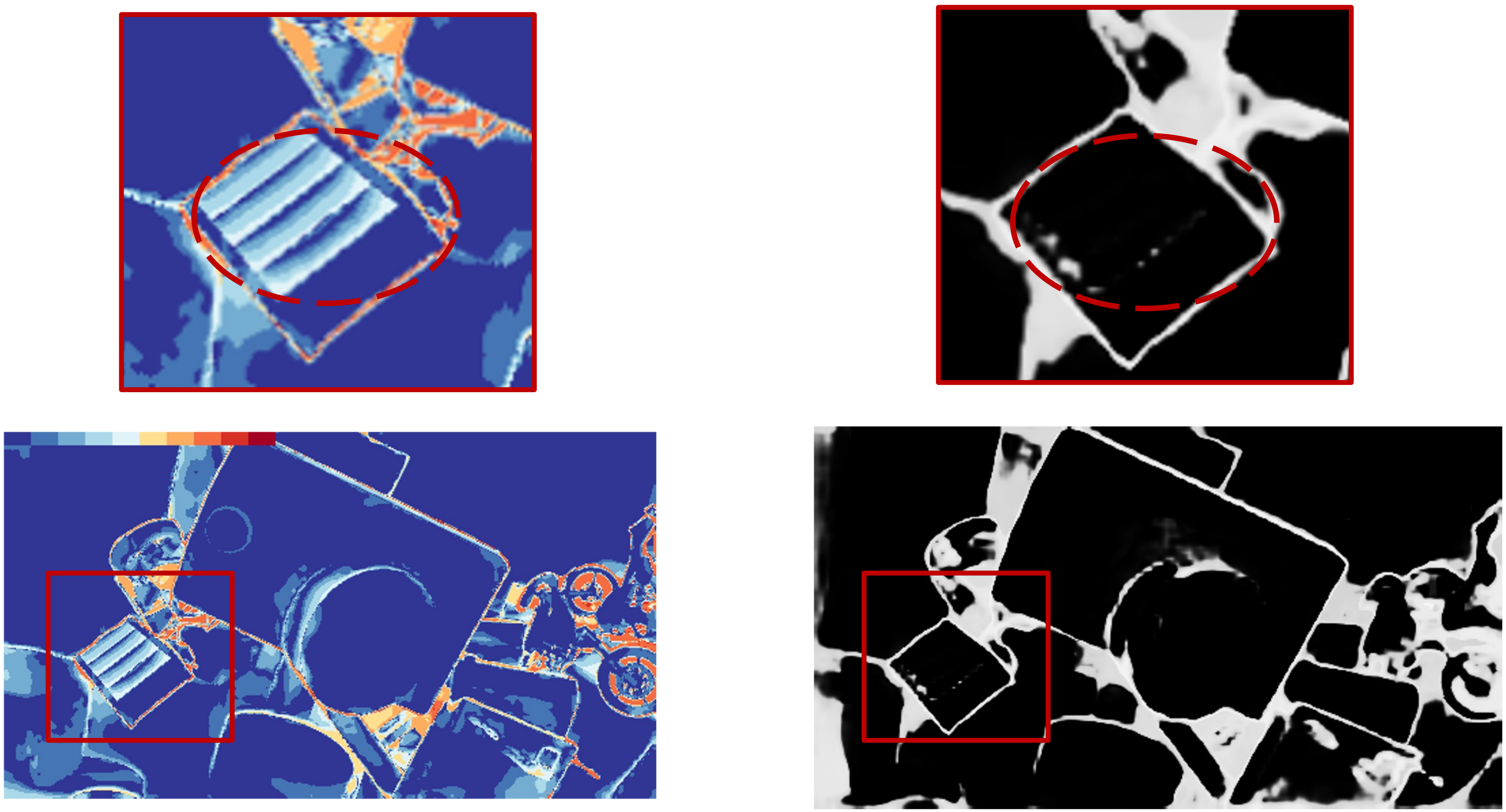
$i.e.\ \sigma^{(i)}$



Error Map

Uncertainty Map

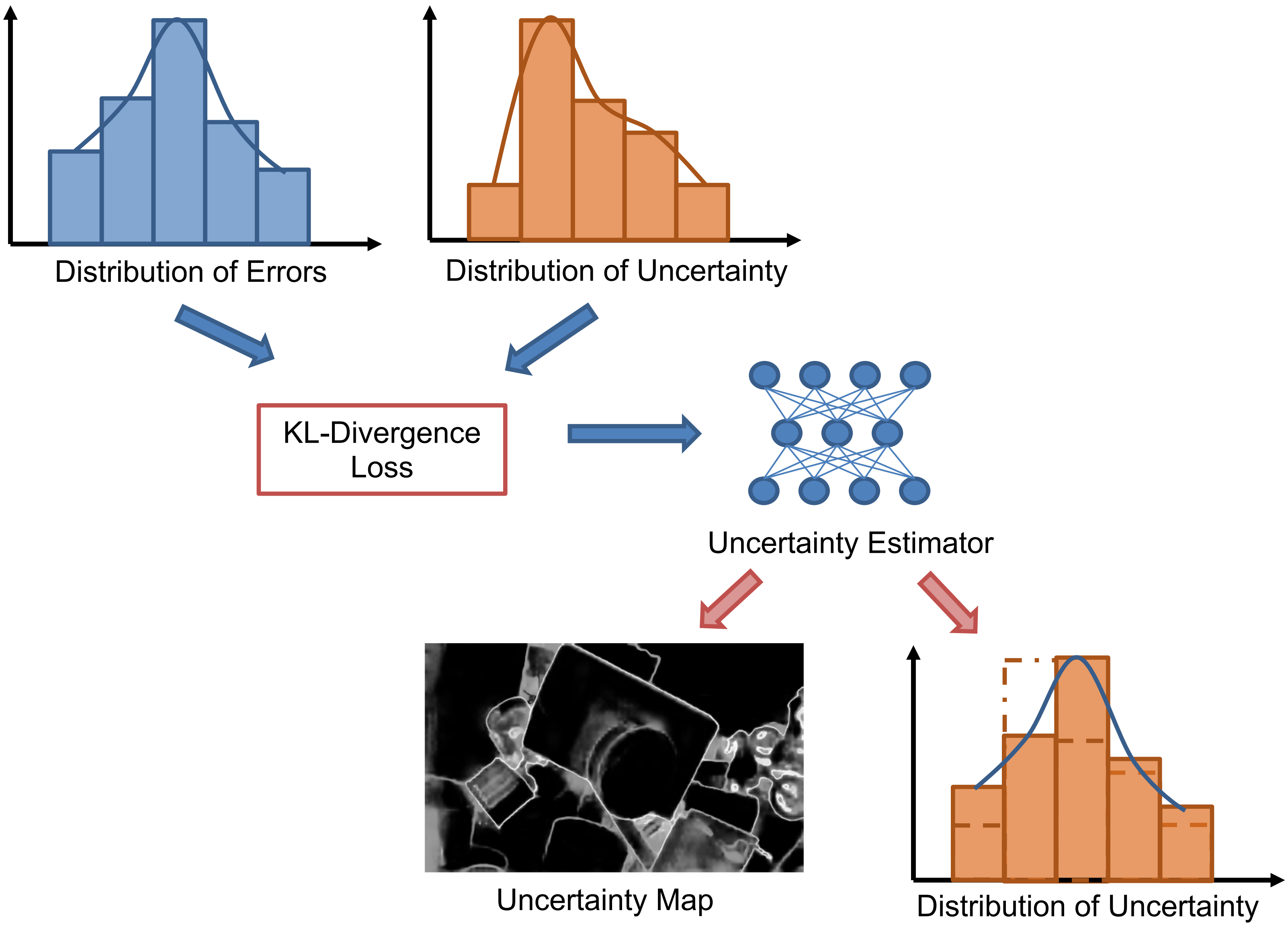
Mismatch Between Error and Uncertainty



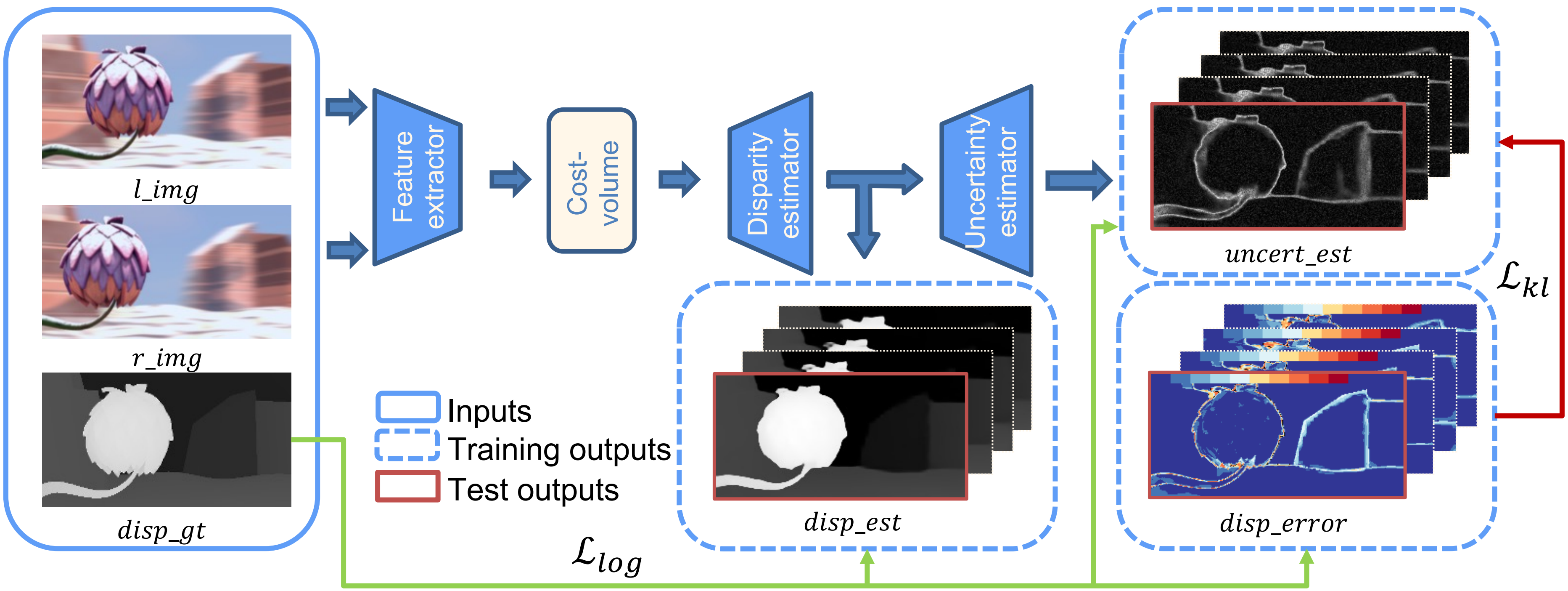
Error Map

Uncertainty Map

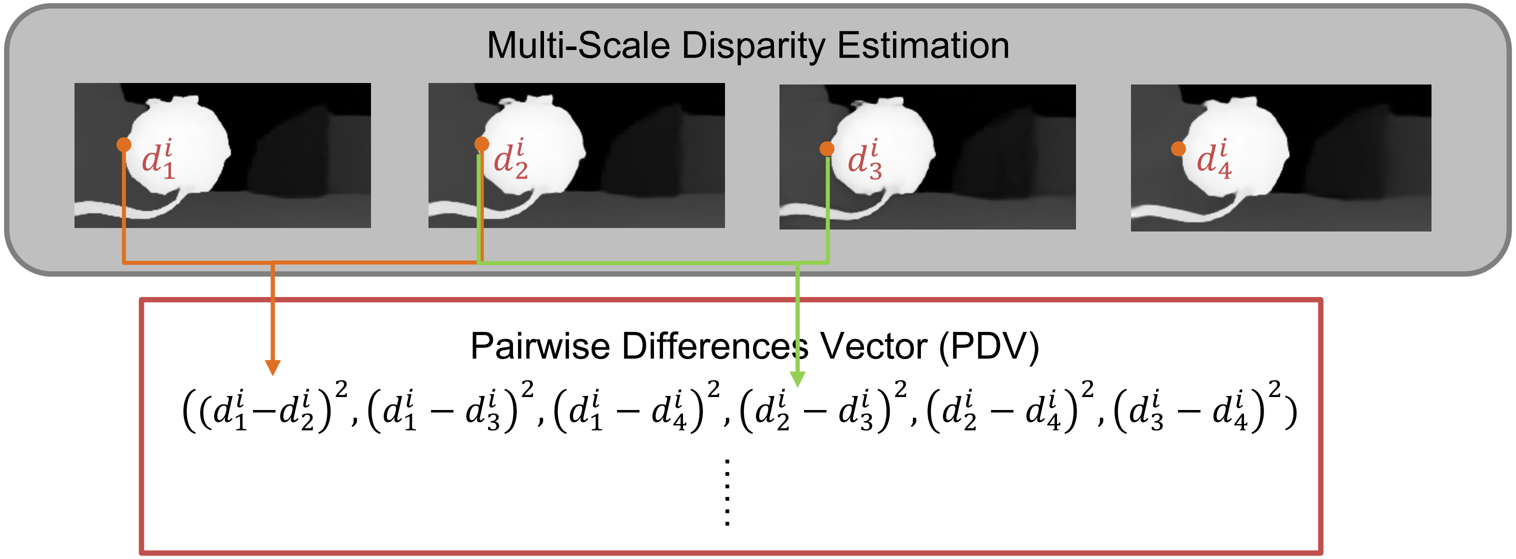
Learning the Distribution of Errors



SEDNet - Pipeline Overview

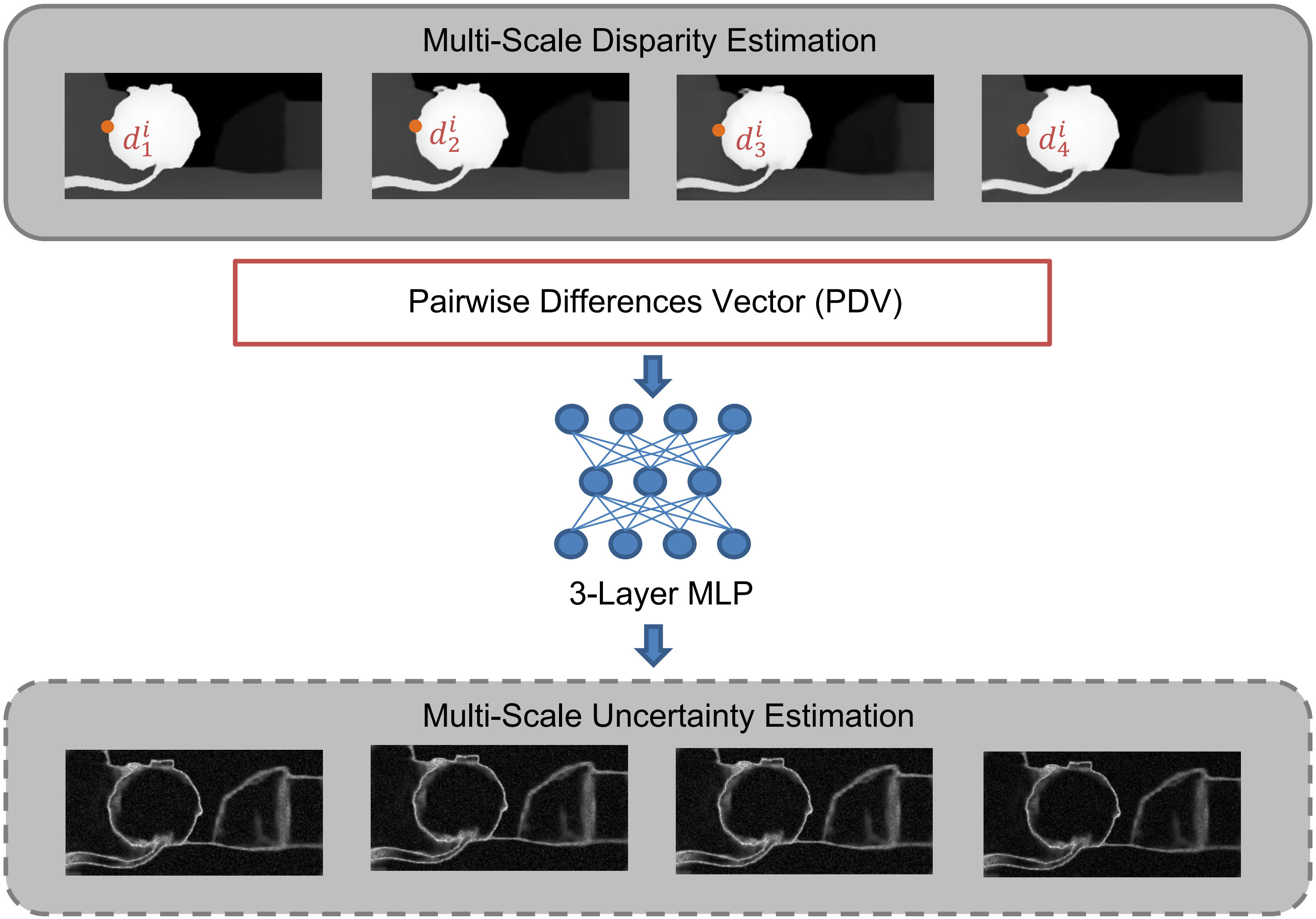


SEDNet – Uncertainty Estimator

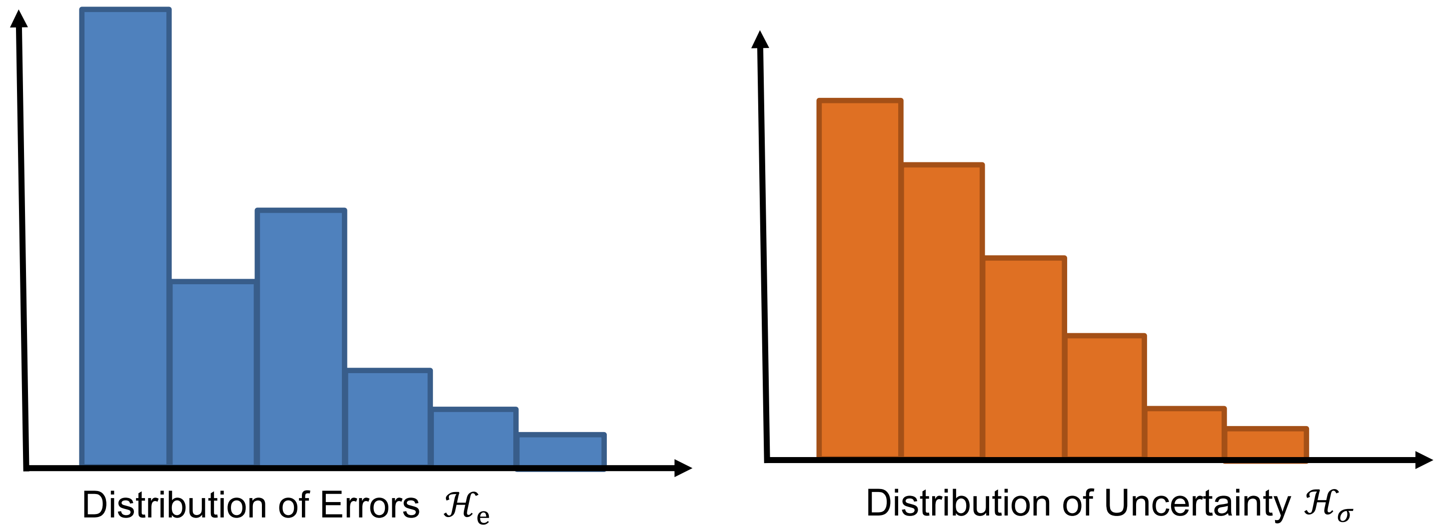


Squared differences between disparity estimates at different scales

SEDNet – Uncertainty Estimator

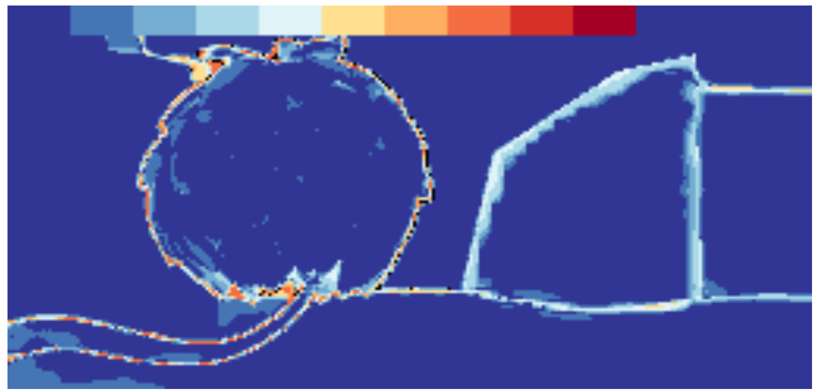


Matching the Distribution of Error

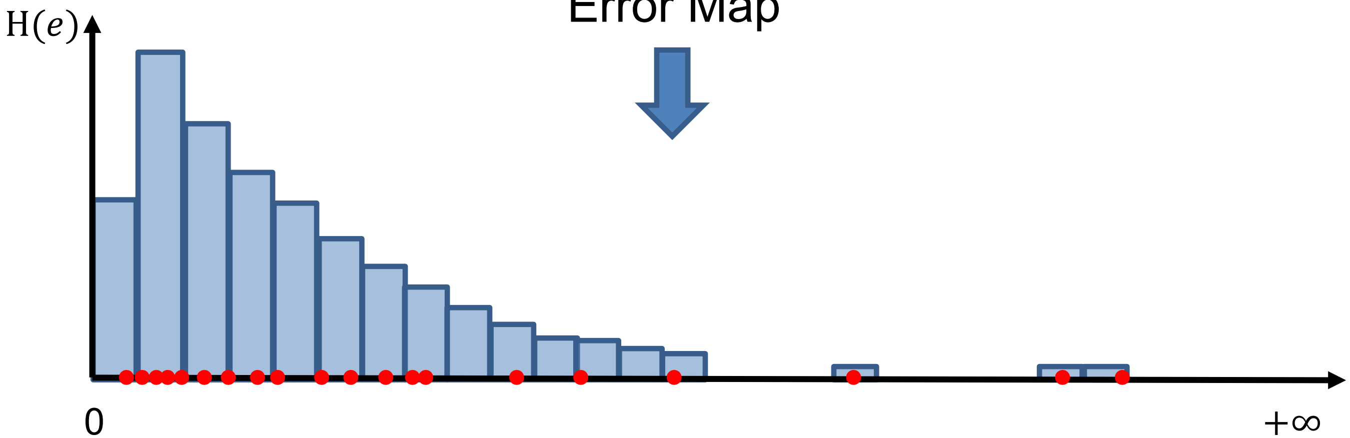


To prepare **the inputs to the matching error loss**, we need to build the distributions.

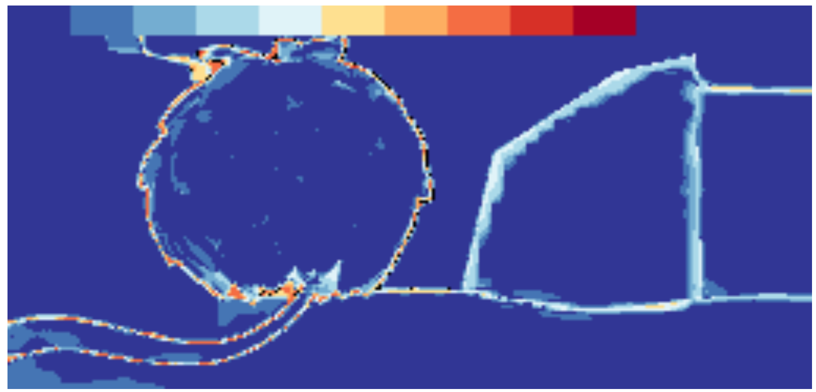
Convert Error Map to Distribution



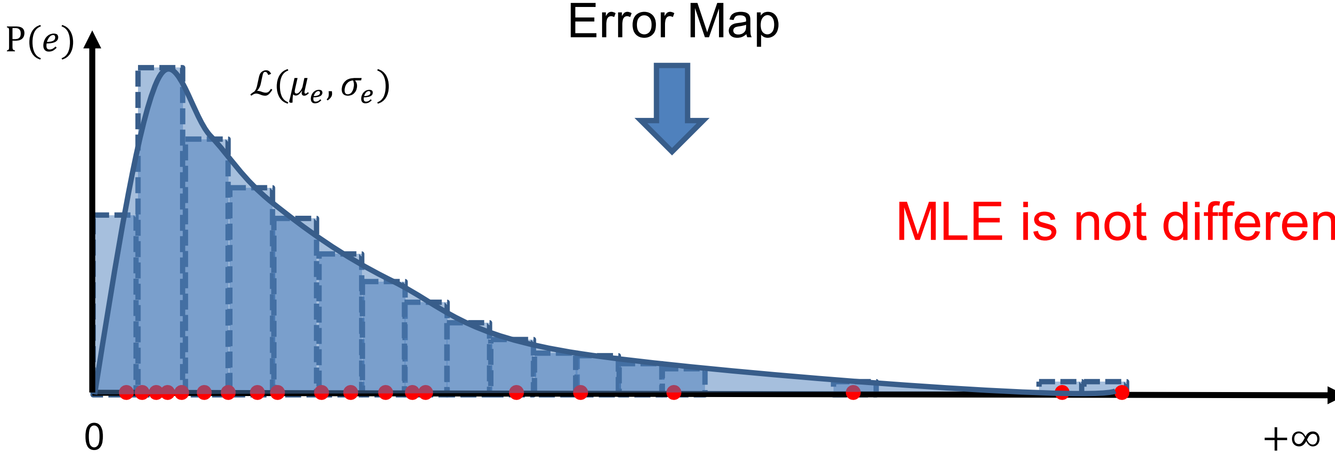
Error Map



Convert Error Map to Distribution

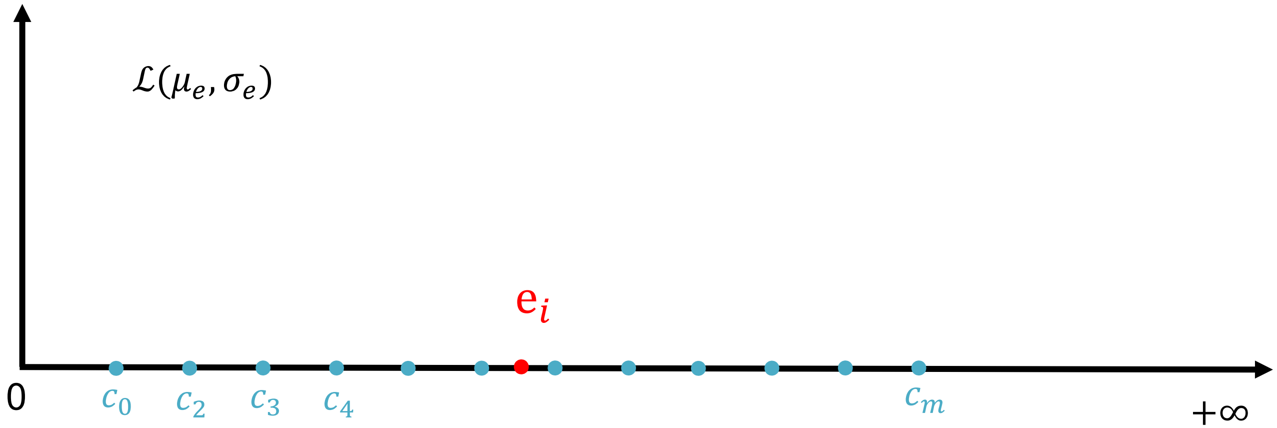


Error Map



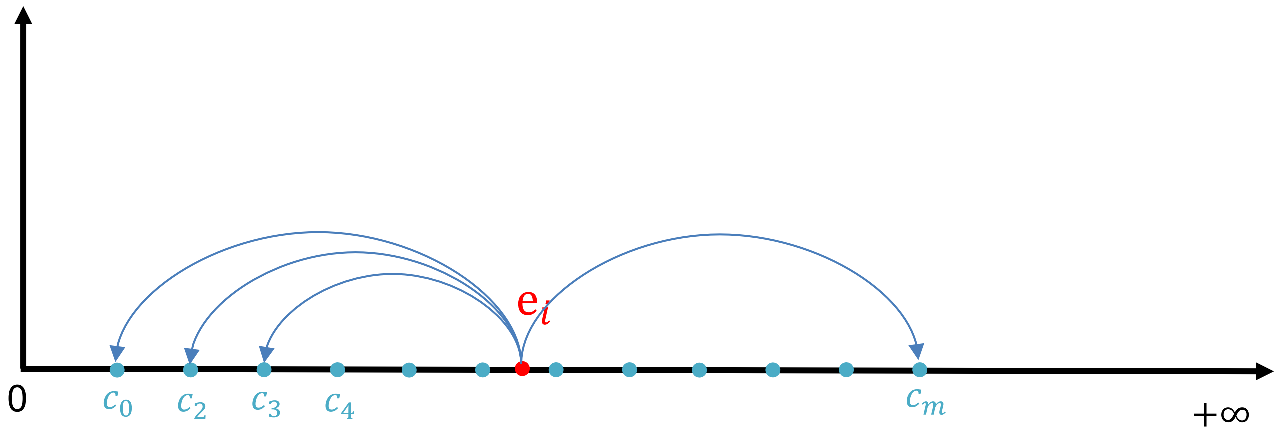
MLE is not differentiable !

Soft Histogramming



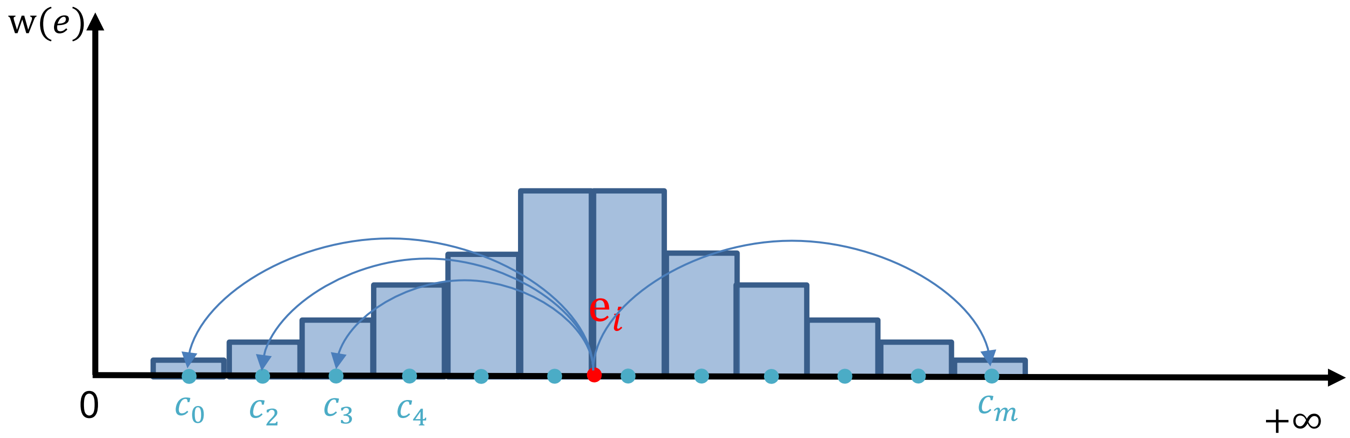
1. Pick m bin centers, where $c_1 = \mu_e, c_m = \mu_e + m\sigma_e$

Soft Histogramming



1. Pick m bin centers, where $c_1 = \mu_e$, $c_m = \mu_e + m\sigma_e$
2. Compute $L2$ distances between e_i and bin centers: $d_{i,1}, d_{i,2}, \dots, d_{i,m}$

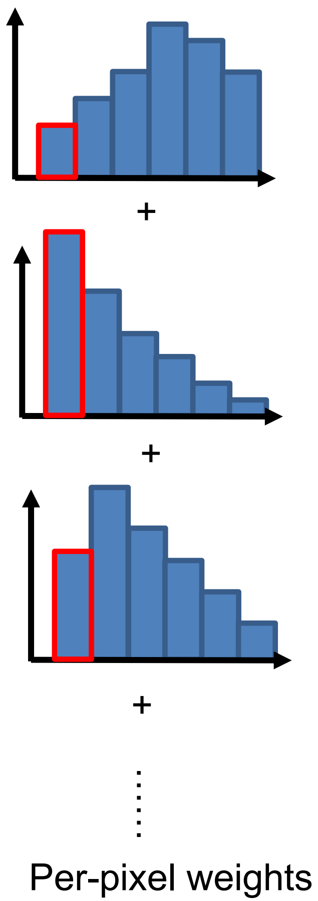
Soft Histogramming

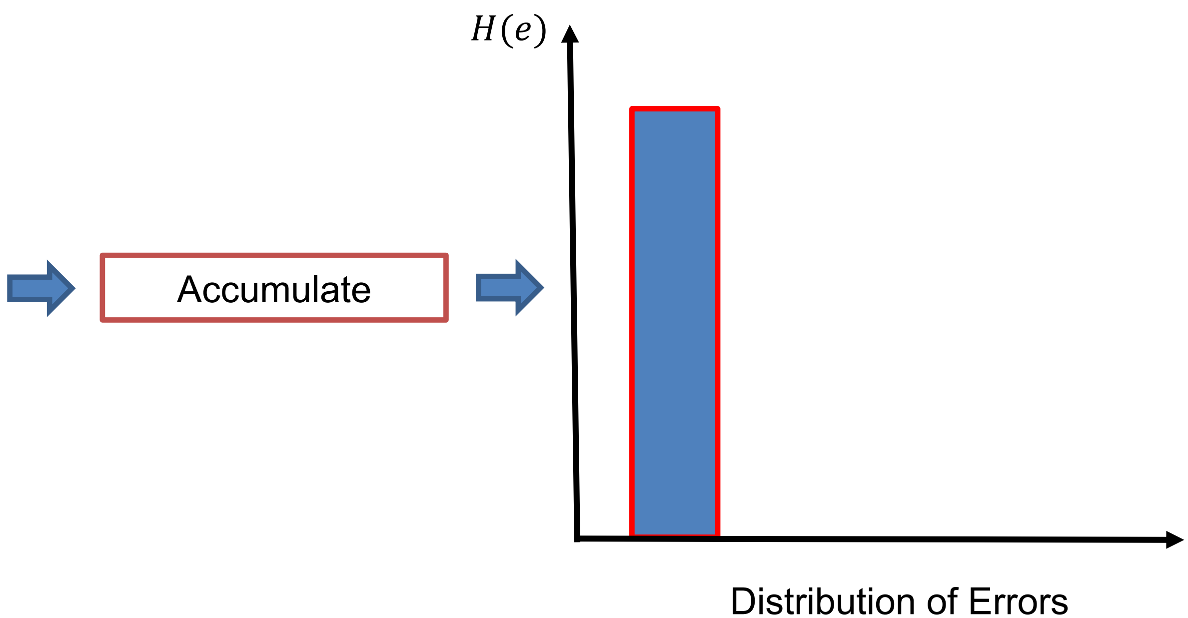


1. Pick m bin centers, where $c_1 = \mu_e$, $c_m = \mu_e + m\sigma_e$
2. Compute $L2$ distances between e_i and bin centers: $d_{i,1}, d_{i,2}, \dots, d_{i,m}$
3. Convert the distances $d_{i,j}$ to pixel weights

$$w_j(e_i) = \lambda_1 \exp\left(-\frac{d_{i,m}}{\lambda_2}\right)$$

Soft Histogramming

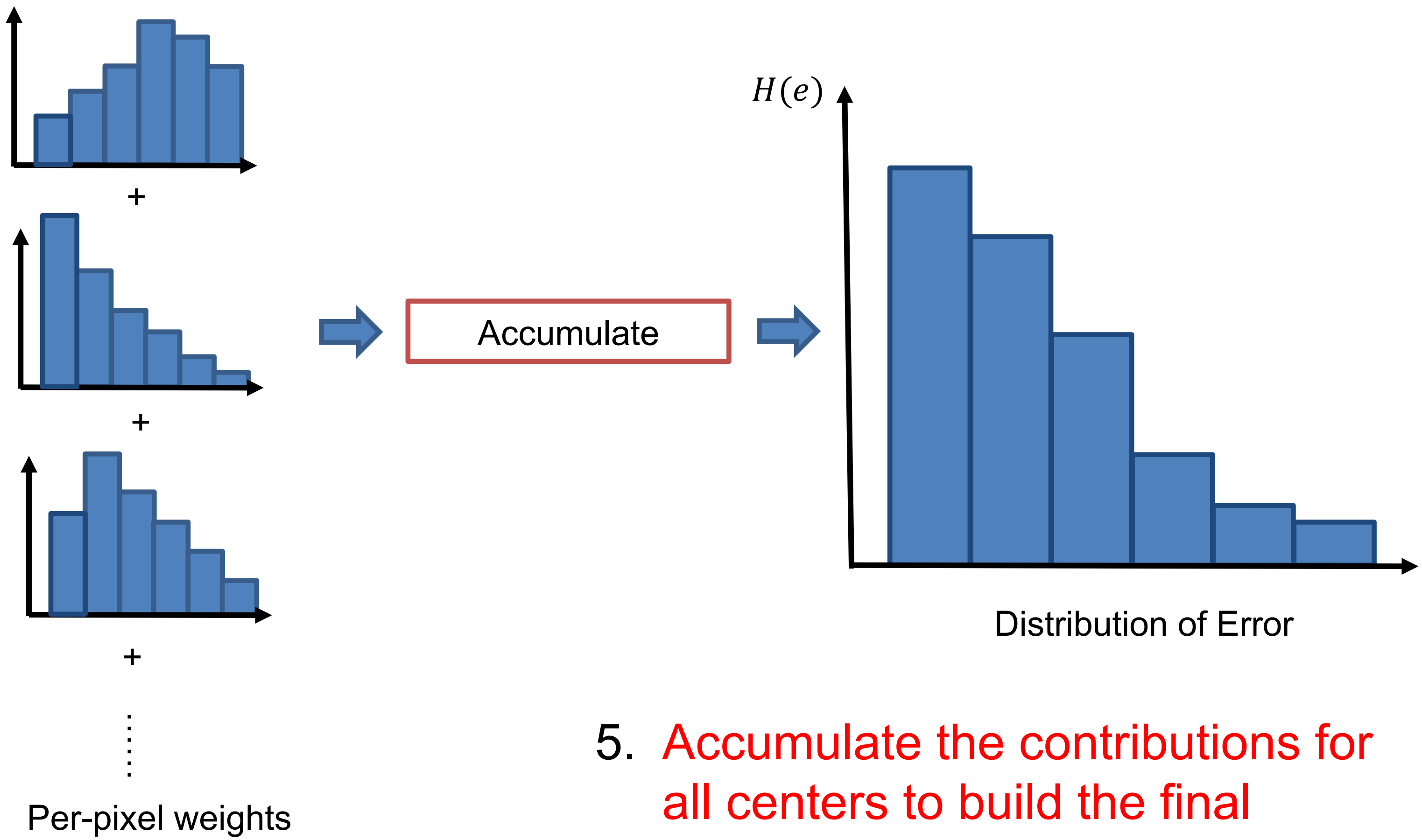




4. Using softmax to convert all pixel weights to be the contributions to the bin centers

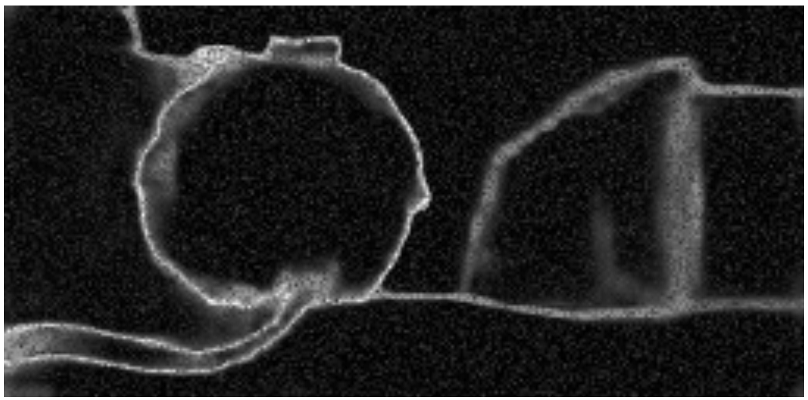
$$H_j(e_i) = \frac{1}{n}\sum_{i=0}^{n} \frac{w_j(e_i)}{\sum_0^{j=m} w_j(e_i)}$$

Soft Histogramming

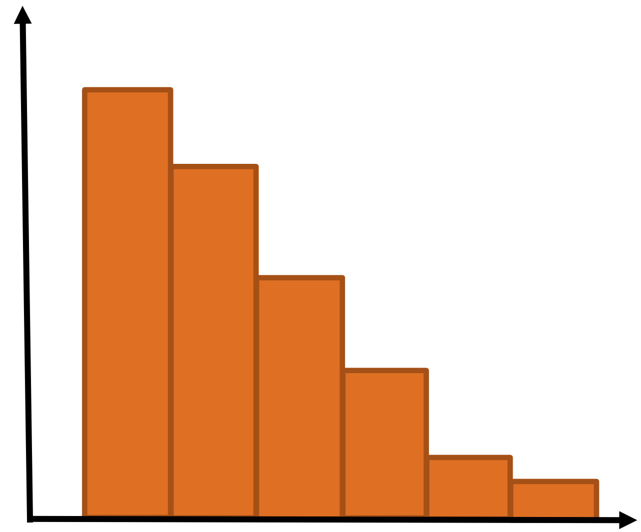


5. Accumulate the contributions for all centers to build the final distribution (histogram)

Convert Uncertainty Map to Distribution



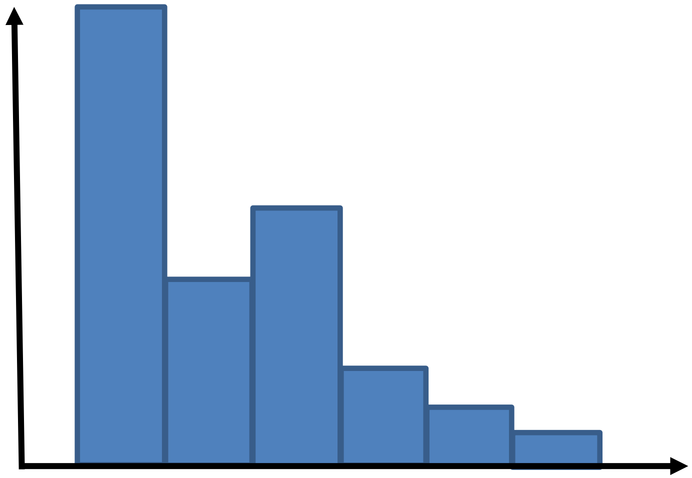
Uncertainty Map



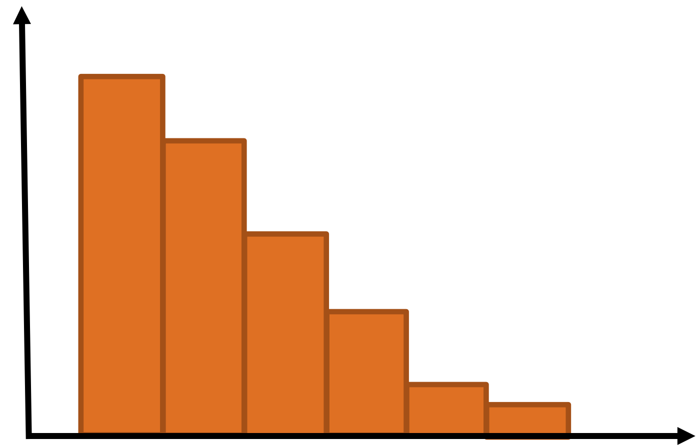
Distribution of Uncertainty

Using **the same bin centers** as for the errors!

Matching the Distribution of Error



Distribution of Errors $\mathcal{H}_{\mathrm{e}}$



Distribution of Uncertainty $\mathcal{H}_{\sigma}$

Using KL-divergence loss $\mathcal{L}_{KL} = \mathcal{D}_{KL}(\mathcal{H}_{\mathrm{e}}||\mathcal{H}_{\sigma})$

Loss Function

$$\mathcal{L} = \sum_{k=1}^{K} c^k \cdot (\mathcal{L}_{log}^k + \mathcal{L}_{KL}^k)$$

- All disparity and uncertainty maps are **upsampled to the highest resolution**
- c^k denotes **the coefficients** for the k^{th} resolution level
  (We use K = 4 in our experiments)
- **Sum up** the two loss across all resolution

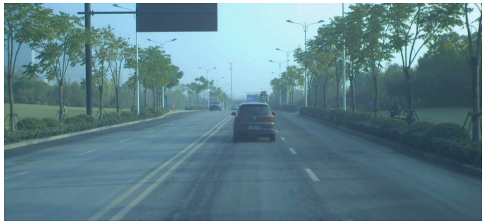
Datasets & Baselines

Datasets

Scene Flow

Virtual KITTI 2

Synthetic Data



DrivingStereo

Real Data

Baselines

- Pick *strong baselines* according to recent survey (Poggi et al., 2021)
- GwcNet (Guo et al., 2019) + L1
- LAF-Net (Kim et al., 2019) + BCE *[Confidence Network]*
- GwcNet + $\mathcal{L}_{log}$ (Kendall and Gal., 2017)

Primary Results – Disparity & Uncertainty Estimation

In domain: train on VK2, test on VK2

Dataset	Method	EPE(↓)	MAPE(↓)	AUC Opt.(↓)	AUC Est.(↓)
Virtual KITTI 2	GwcNet +$\mathcal{L}_{log}$	0.3899	0.4136	4.6872	12.5320
	SEDNet	**0.3236**	**0.3561**	**4.2767**	**9.9843**

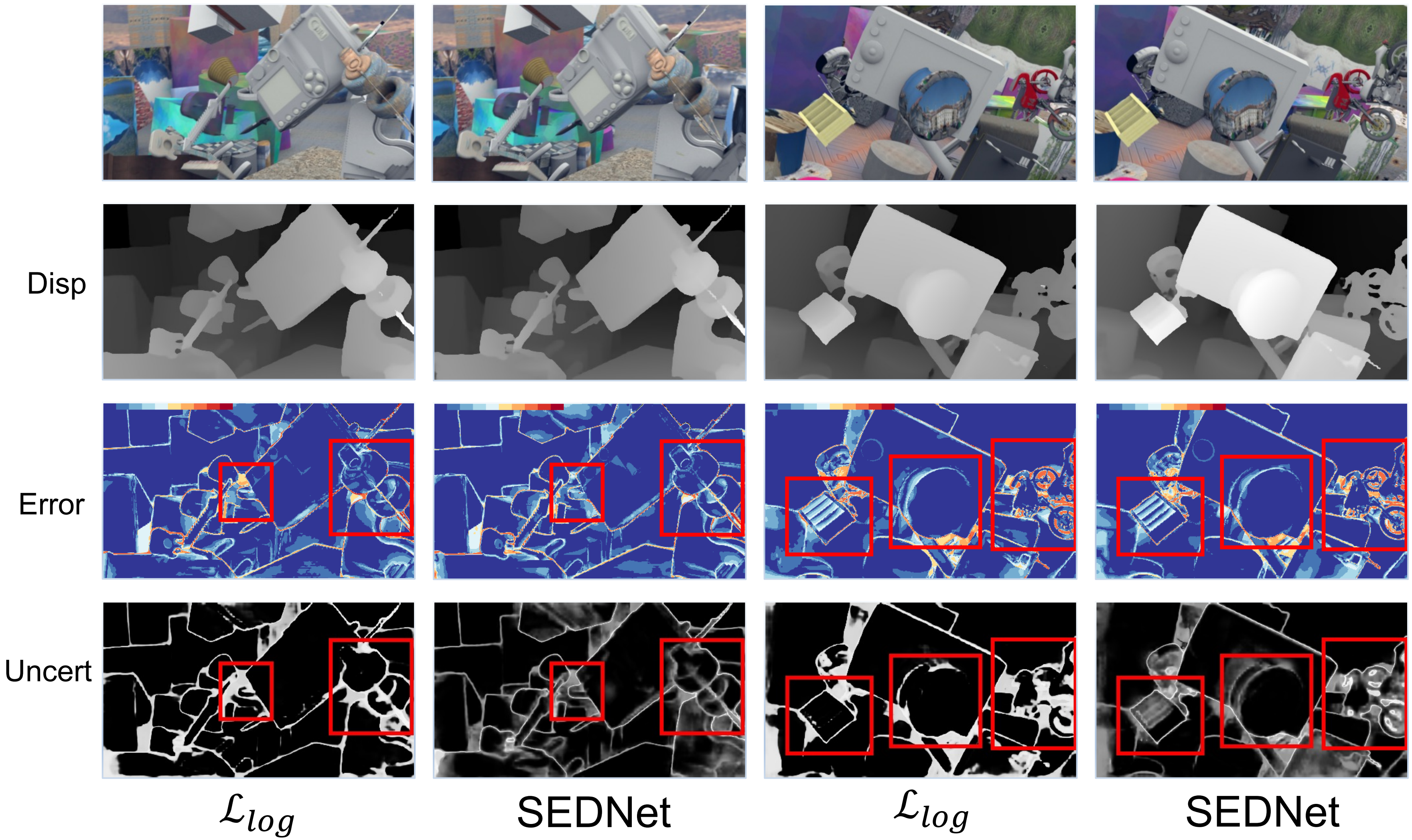
Cross domain: train on VK2, test on DS-Weather

Dataset	Method	EPE(↓)	MAPE(↓)	AUC Opt.(↓)	AUC Est.(↓)
DS-Weather	GwcNet +$\mathcal{L}_{log}$	2.3944	2.1443	41.1909	95.4264
	SEDNet	**1.7051**	**1.5842**	**39.8057**	**87.1882**

- **MAPE** is the mean L1 distance between the error and the observation uncertainty scalar, σ.
- Please see the paper for more results.

Qualitative Results – Scene Flow

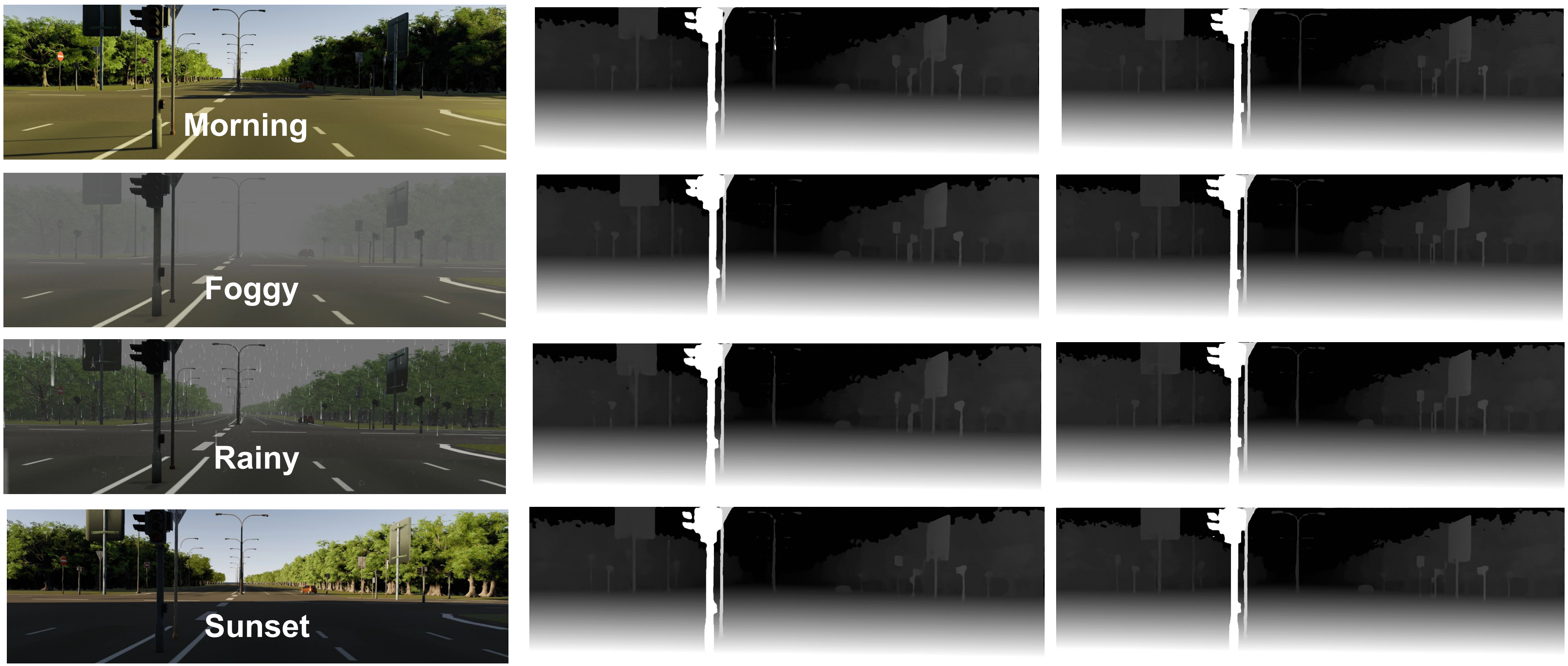
Synthetic stereo pairs for flying objects



SEDNet has smaller errors and more accurate uncertainty.

Qualitative Results – Virtual KITTI 2

Synthetic stereo pairs for driving in different weather

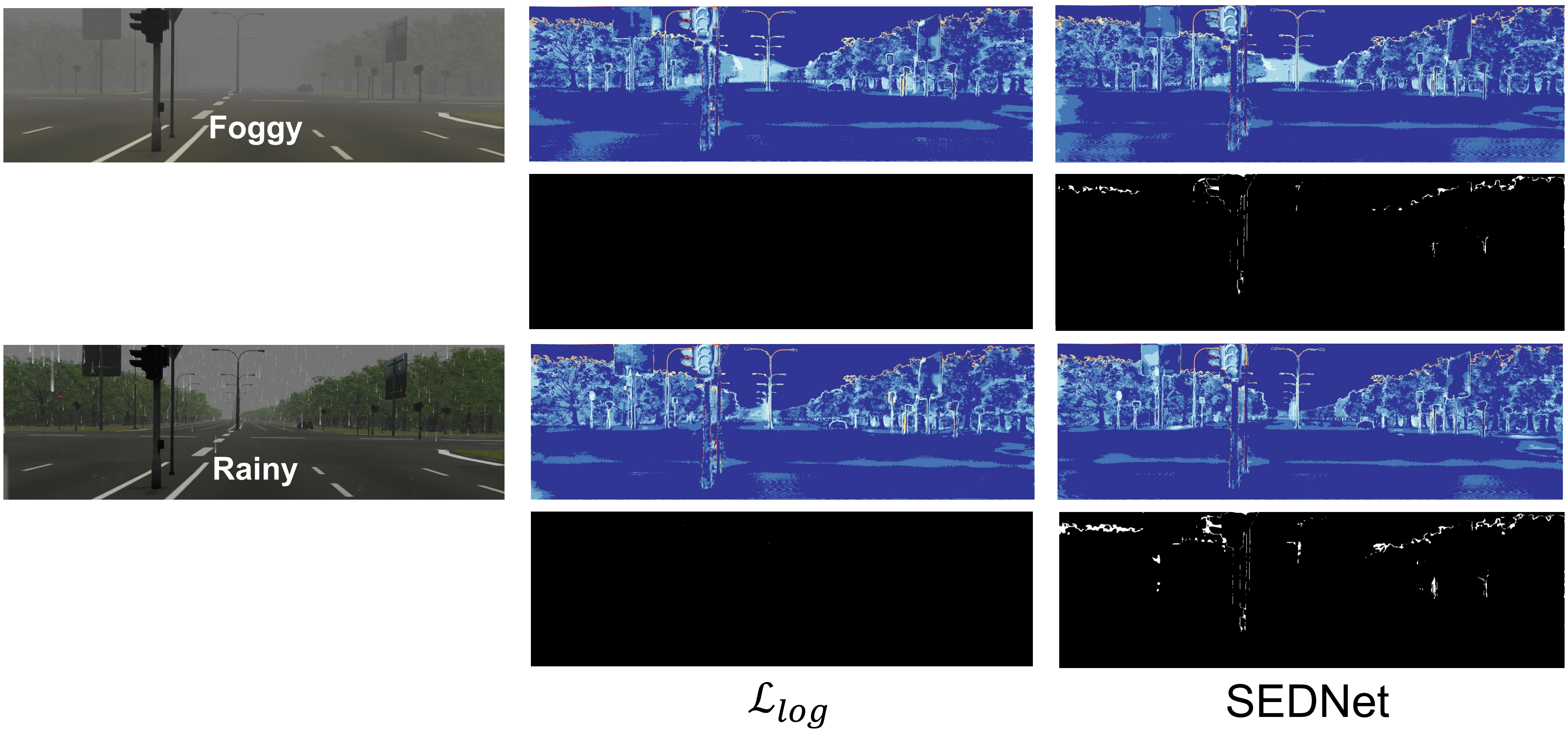


$\mathcal{L}_{log}$ SEDNet

SEDNet has better disparity estimation in different weather.

Qualitative Results – Virtual KITTI 2

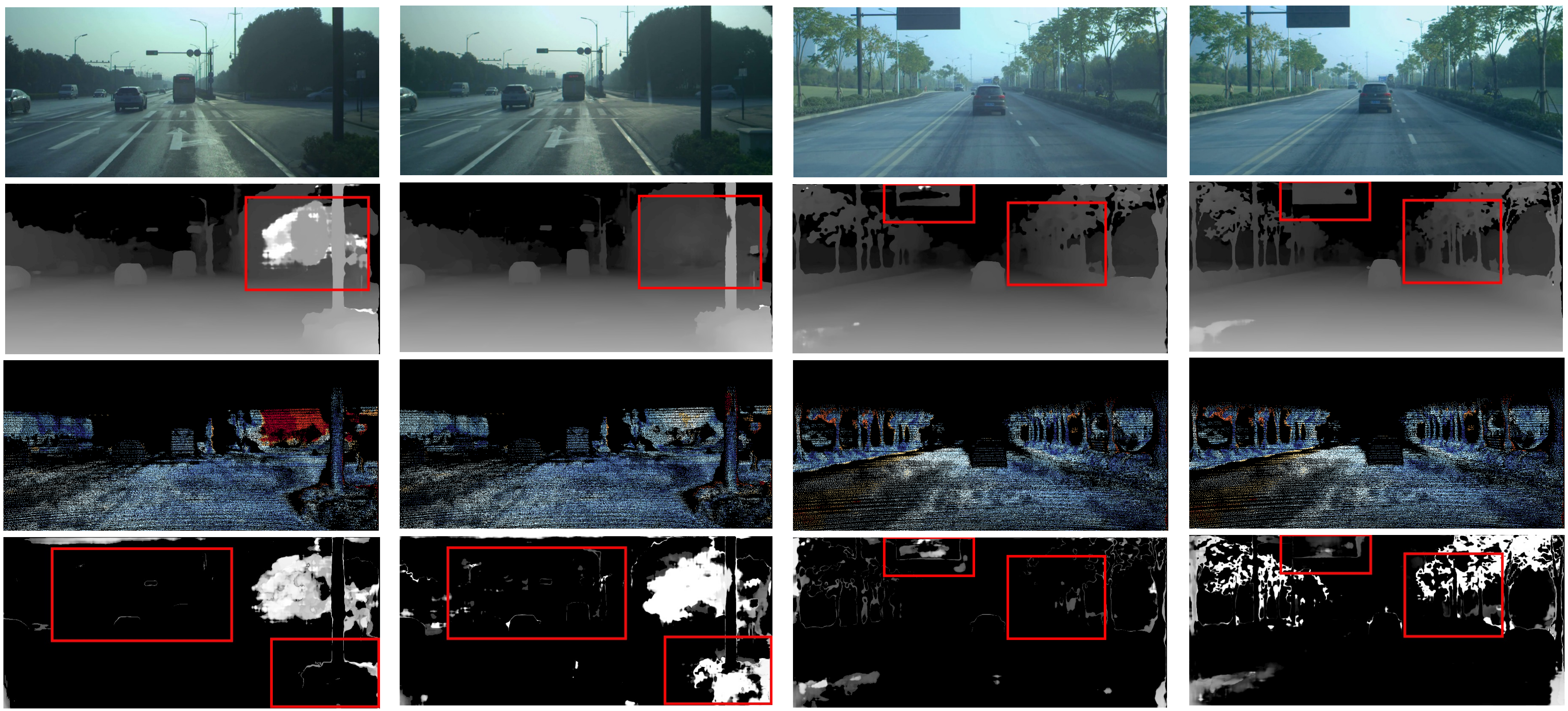
Comparing the challenging samples



The improvement of SEDNet on disparity and uncertainty estimation is more visible especially under bad weather.

Qualitative Results – DrivingStereo

Real stereo pairs for foggy weather

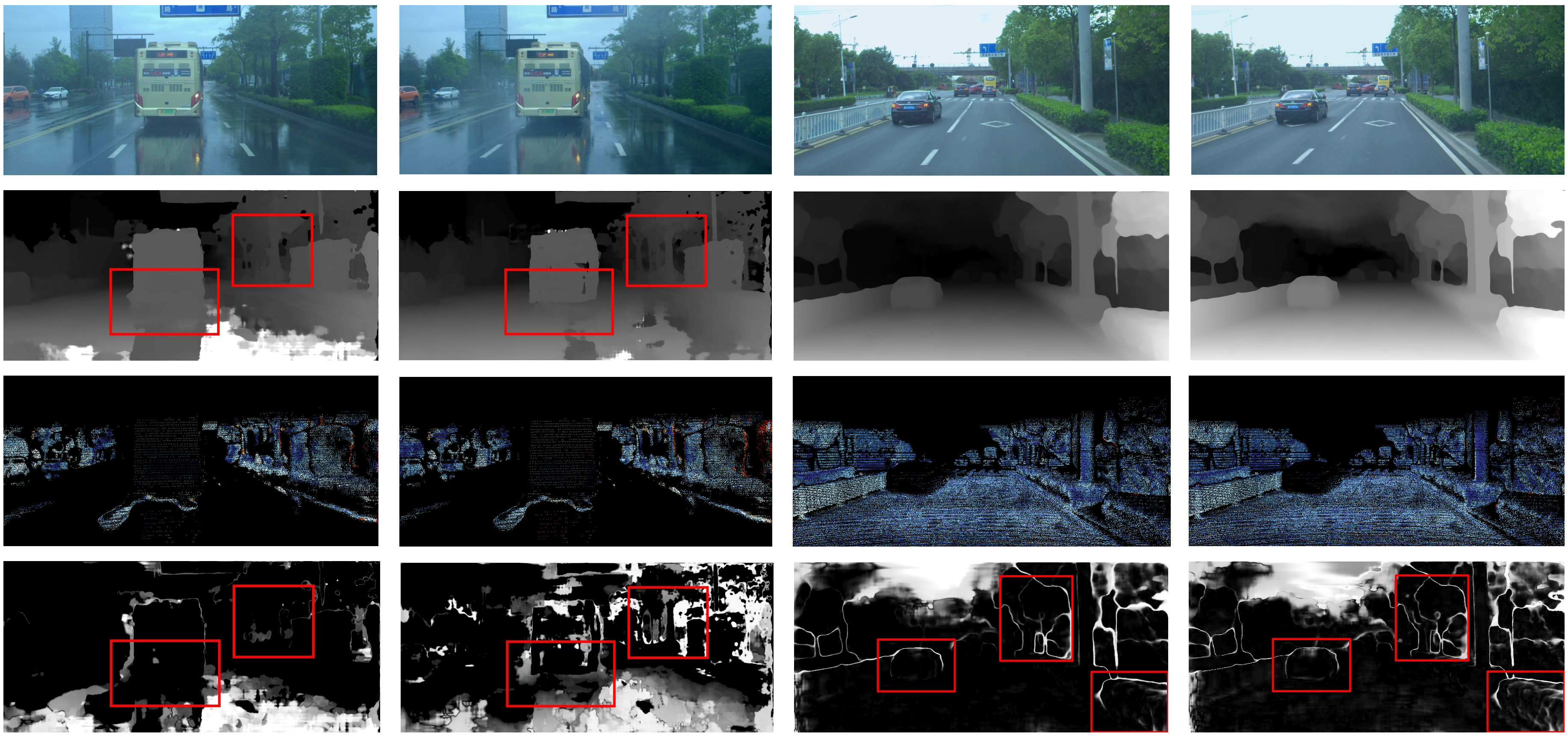


$\mathcal{L}_{log}$ SEDNet $\mathcal{L}_{log}$ SEDNet

- The foggy day are usually **very dark**, which makes distinguishing objects in the shadow difficult.
- SEDNet still performs better in predicting the uncertainty of the objects **far from the camera** and in the bottom right **dark corner**.

Qualitative Results – DrivingStereo

Real stereo pairs for rainy weather



$\mathcal{L}_{log}$ SEDNet $\mathcal{L}_{log}$ SEDNet

- Rainy images suffer from **poor illumination**, also face challenges due to **reflections in the water**.
- SEDNet captures **more faithful details** in both disparity and uncertainty maps.

STEVENS
INSTITUTE OF TECHNOLOGY

Code available at

https://github.com/lly00412/SEDNet

Poster Section： THU-AM-072

Thank you！