What Uncertainties do we need in Bayesian Deep Learning for computer vision?

zhuofehuang

Paper Reading 2021/04/21

Outline

- Introduction and background of uncertainty:
  - What Uncertainties do we need in Bayesian Deep Learning for computer vision? Neurips 2017

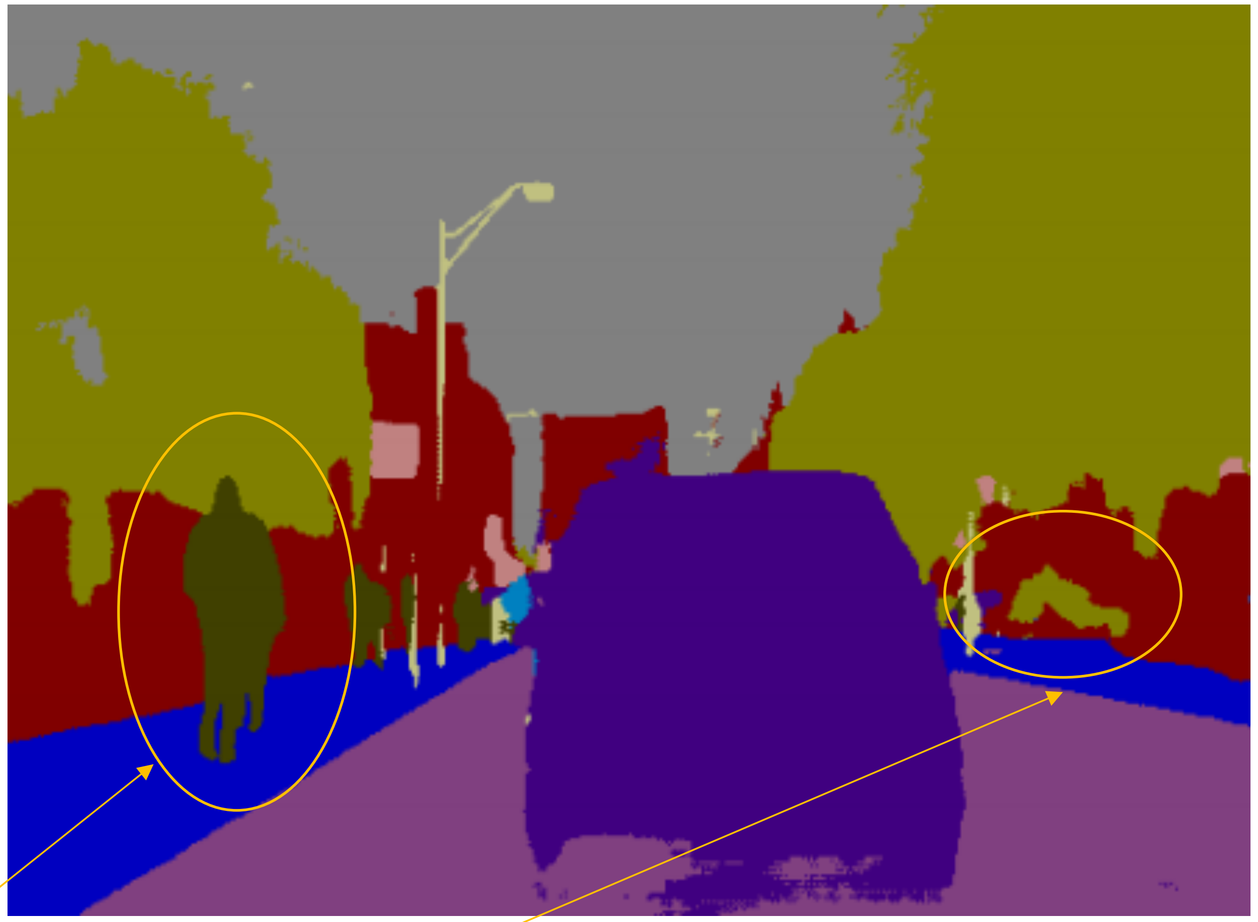
- Uncertainty applied on Unsupervised Depth Estimation
  - Learning SFM from SFM, ECCV 2018
  - D3VO, Nan Yang, CVPR 2020
  - Mono-uncertainty, Matteo Poggi, CVPR 2020

Why do we care about uncertainty?

Why do we care about uncertainty?

Cat?
Dog?
???

Why do we care about uncertainty?

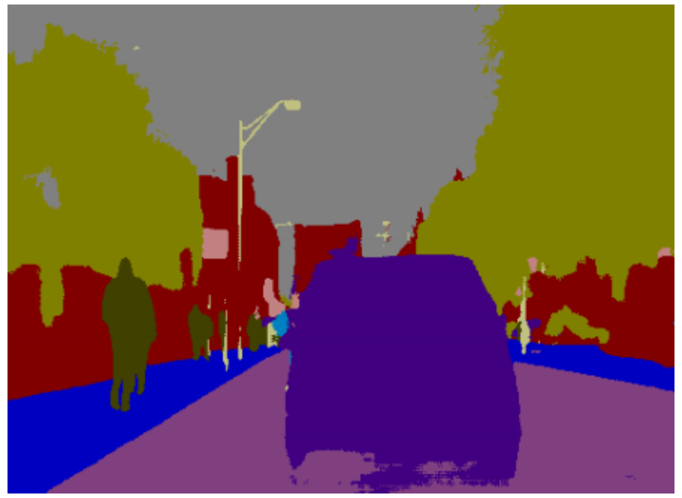


reliable?

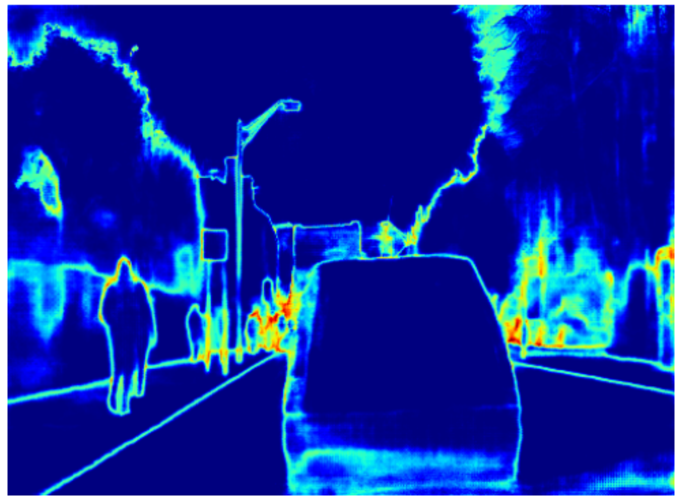
Background

- 我们希望模型能够在预测结果的同时，能够自己估计**uncertainty**，从而辅助决策。
- 对于错误的预测结果以及超出预期范围的(**Out-of-distribution, OoD**) 样本输入，我们希望模型能够给出较高的**uncertainty**。

input



prediction



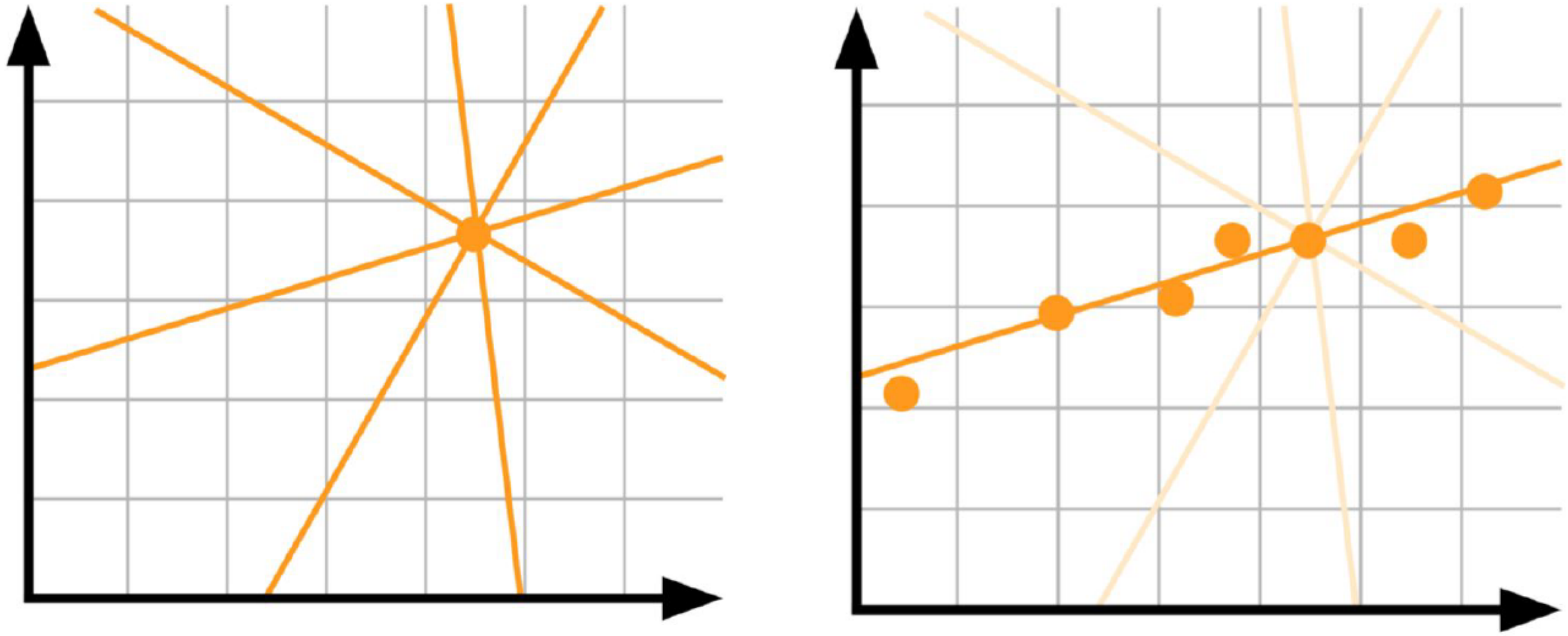
uncertainty map

Types of Uncertainty

- Epistemic uncertainty – 认知不确定性
- Aleatoric uncertainty – 偶然不确定性

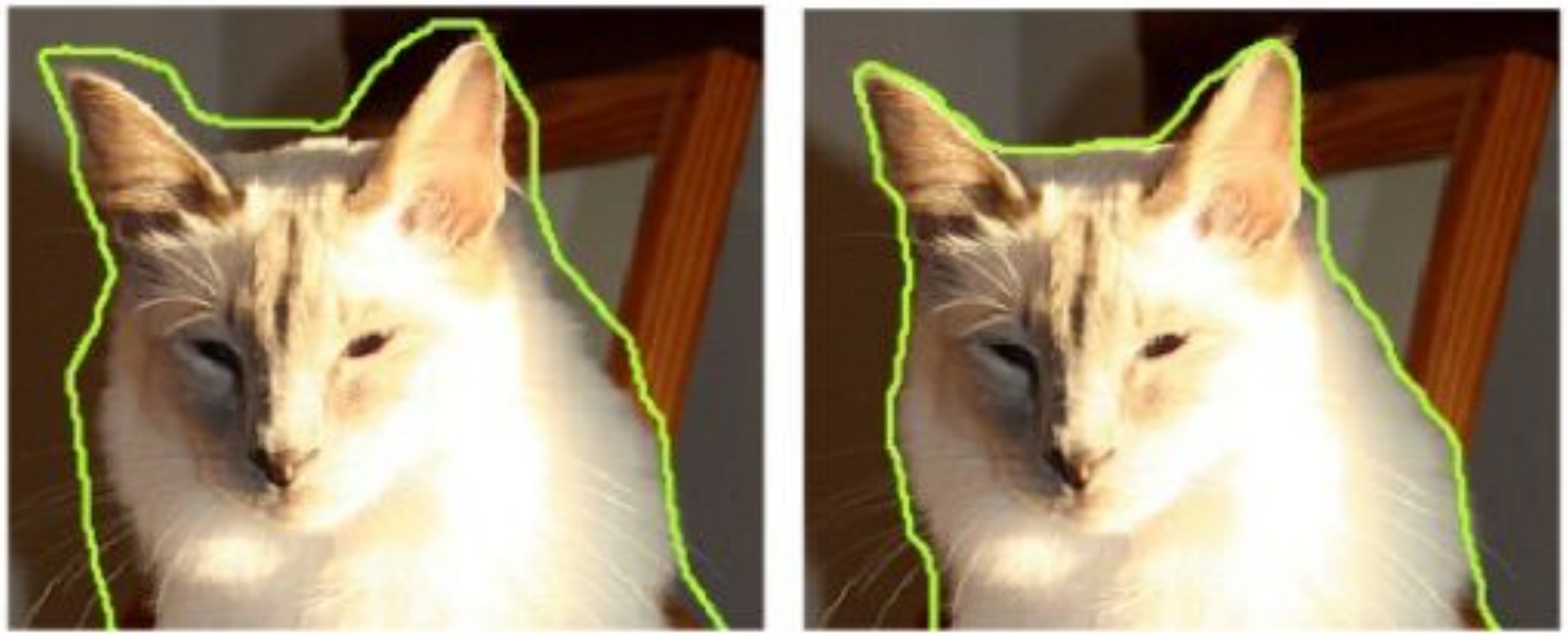
Types of Uncertainty

- Epistemic uncertainty (认知不确定性)
  - lack of knowledge, reducible
  - uncertainty in model (either model parameters or model structure)
  - more data helps (eg., data augment)



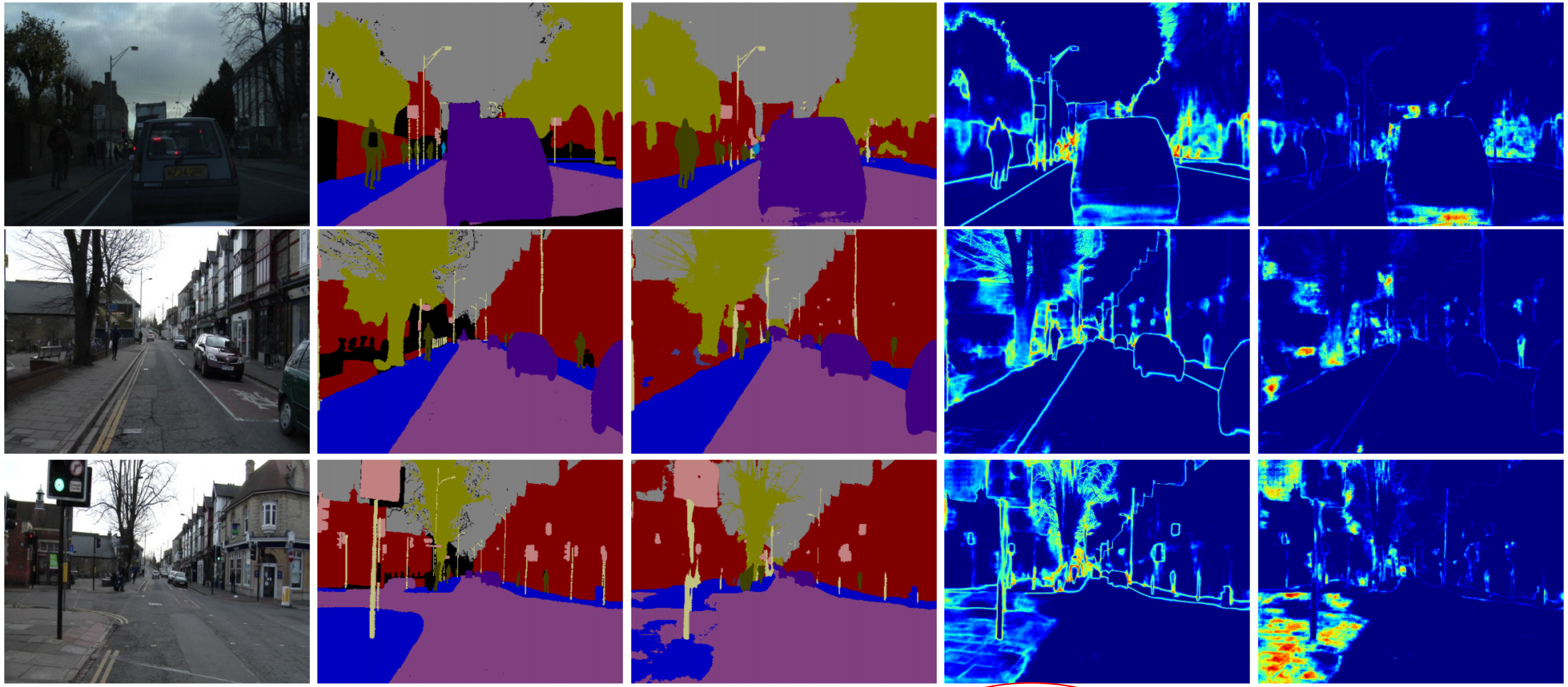
Types of Uncertainty

- Aleatoric uncertainty (偶然不确定性)
  - stochastic, irreducible
  - captures noise inherent in the observations
  - more data doesn't help






Neurips 2017



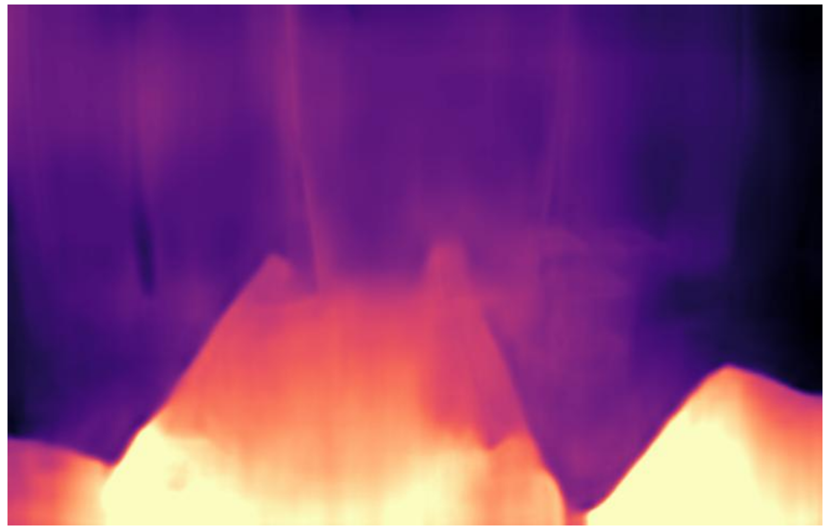
(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Start with an example

- Consider a depth estimation task:
  - Assuming depth gt is provided (Supervised Learning)
  - For each single image I
  - Prediction: $\tilde{d} = DepthNet(I)$

$$loss = |\tilde{d} - d|$$

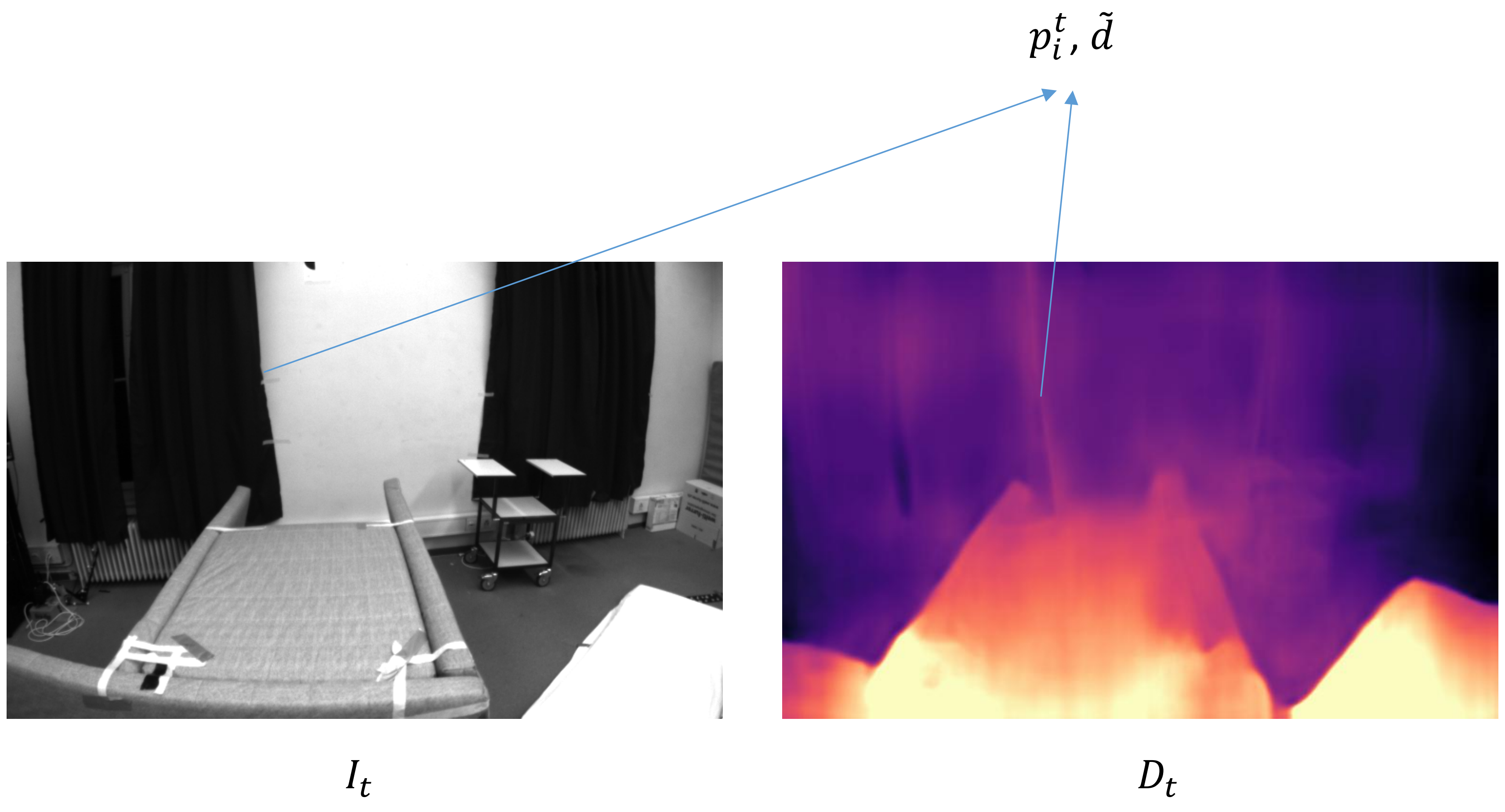
confidence level for $\tilde{d}$?



Start with an assumption

- For each timestamp t,
  - For each pixel p_i^t on frame I_t
    - Hints:
      - (1) distribution of prediction $\tilde{d}$ around ground-truth d
      - (2) $\tilde{d}$ will be far away from d with lower probability.

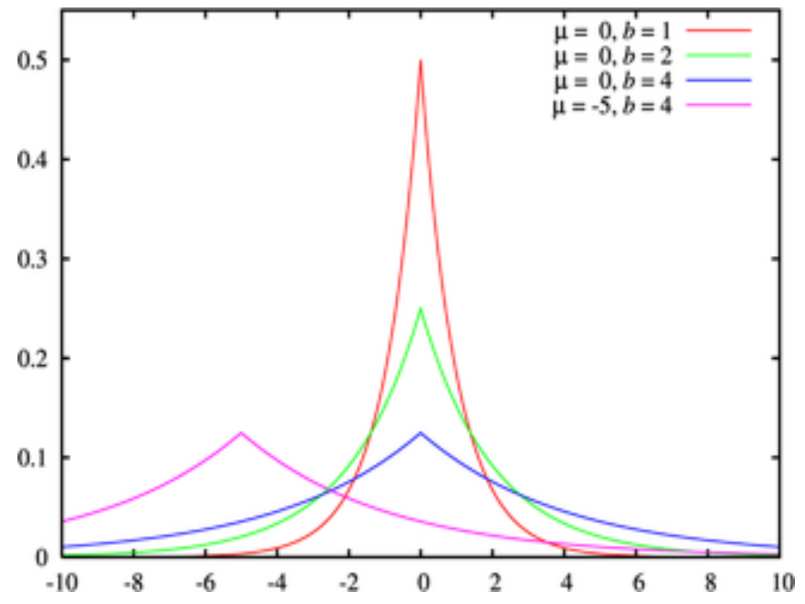
$$loss = |\tilde{d} - d|$$

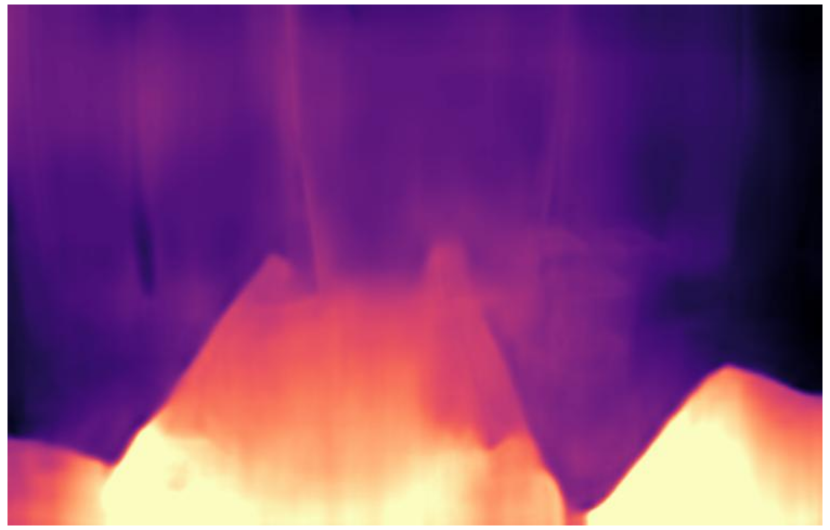


Laplace Distribution

- Key idea: predict a posterior probability distribution for each pixel parameterized with its mean as well as its variance $\mathrm{p}\left(d \middle| \tilde{d}, \sigma\right)$ over the ground-truth label d

$$f(x|\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right), \sigma > 0$$

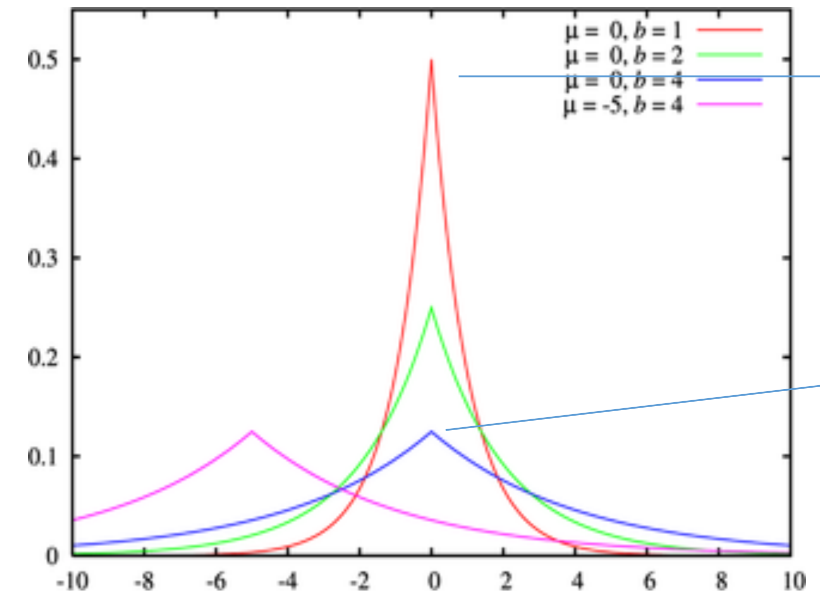




Laplace Distribution

- Key idea: predict a posterior probability distribution for each pixel parameterized with its mean μ as well as its variance σ: $\mathrm{p}\left(d \middle| \tilde{d}, \sigma\right)$

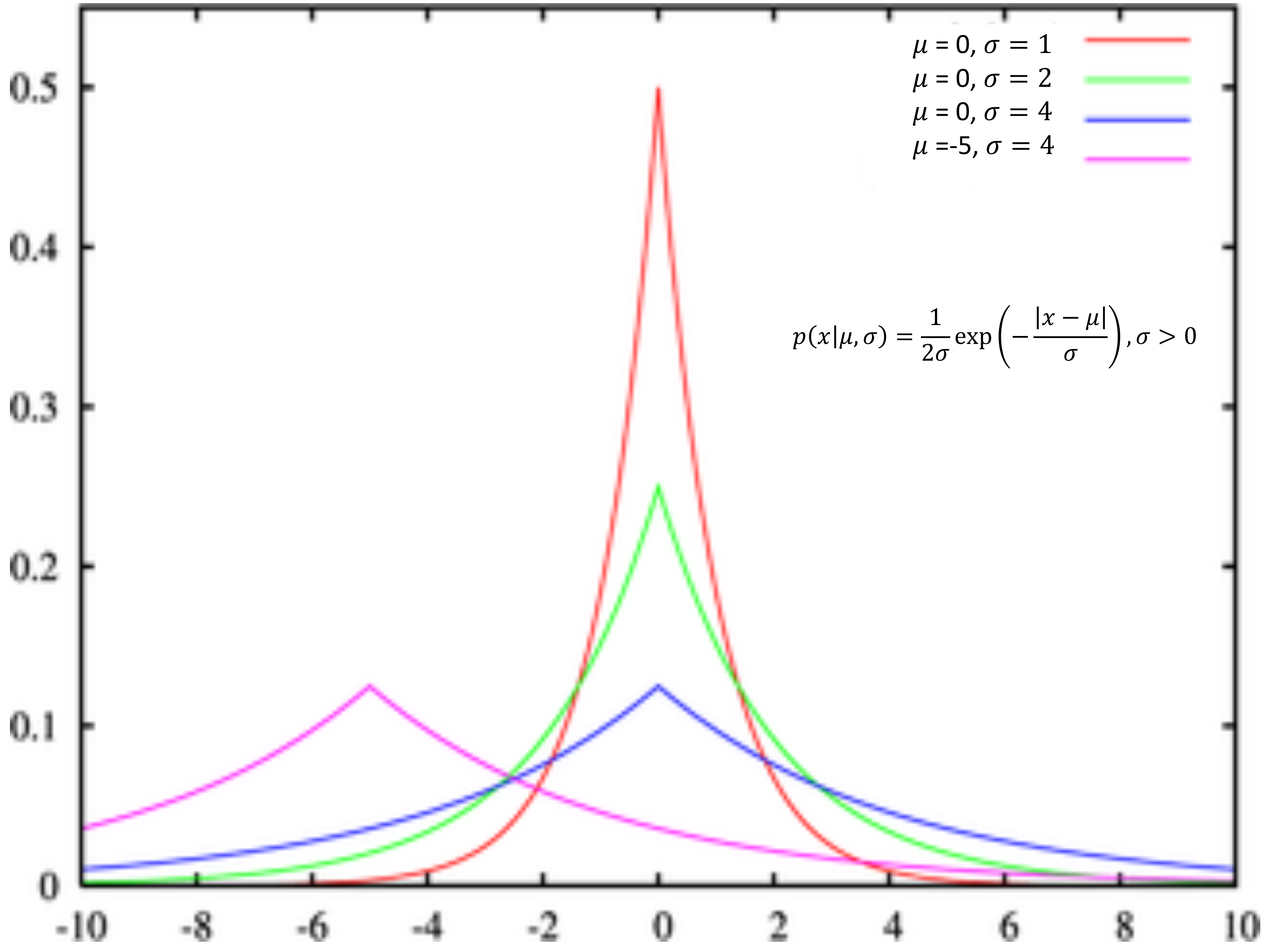
$$p\left(d \middle| \tilde{d}, \sigma\right) = \frac{1}{2\sigma} \exp\left(-\frac{\left|d - \tilde{d}\right|}{\sigma}\right)$$



Negative log-likelihood (NLL loss):

σ : uncertainty

$$-\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma) + const$$

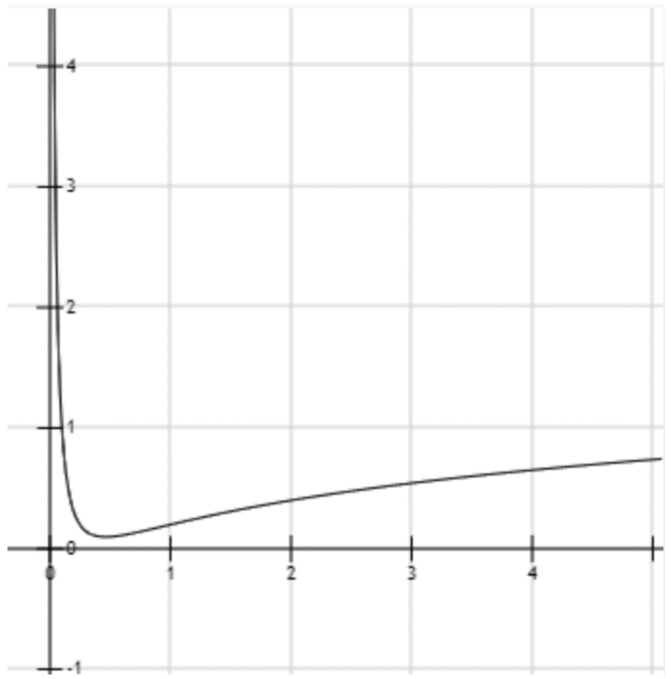

$\mu = 0, \sigma = 1$
$\mu = 0, \sigma = 2$
$\mu = 0, \sigma = 4$
$\mu = -5, \sigma = 4$
$p(x|\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right), \sigma > 0$
0.5
0.4
0.3
0.2
0.1
0
-10
-8
-6
-4
-2
0
2
4
6
8
10

NLL Loss

- Negative log-likelihood loss: (reds are unknowns)

$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma)$$

- Gradients:
  - $\frac{\partial L}{\partial \tilde{d}}$: $\tilde{d}$ approaches d
  - $\frac{\partial L}{\partial \sigma}$: σ approaches $\left|d - \tilde{d}\right|$



$y = \frac{a}{x} + \log(x)\,, a > 0$

$y' = 0 \quad \rightarrow \quad x = a$

NLL Loss

- Negative log-likelihood loss: (reds are unknowns)

$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma) + const$$

- Gradients:
  - $\frac{\partial L}{\partial \tilde{d}}$: $\tilde{d}$ approaches d $\rightarrow$ prediction result $\tilde{d}$ becomes more accurate as gt
  - $\frac{\partial L}{\partial \sigma}$: σ approaches $\left|d - \tilde{d}\right|$ $\rightarrow$ areas with inaccurate predictions takes higher σ

Mathematical View

- Negative log-likelihood loss: (reds are unknowns)

$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma) + const$$

- Ground truth d is not reliable, down-weight the loss $l = \left|d - \tilde{d}\right|$ as $\frac{l}{\sigma}$
- However, just minimize $\frac{l}{\sigma}$ will cause degeneration: $\sigma \rightarrow +\infty$
- So add another regularization term: $\log(\sigma)$

Properties

- Negative log-likelihood loss: (reds are unknowns)

$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma) + const$$

- Attention:
  - σ will be predicted by our model, just by expanding output channel
  - σ do not need groundtruth label, and it will improves the robustness of the model to noisy data or erroneous labels

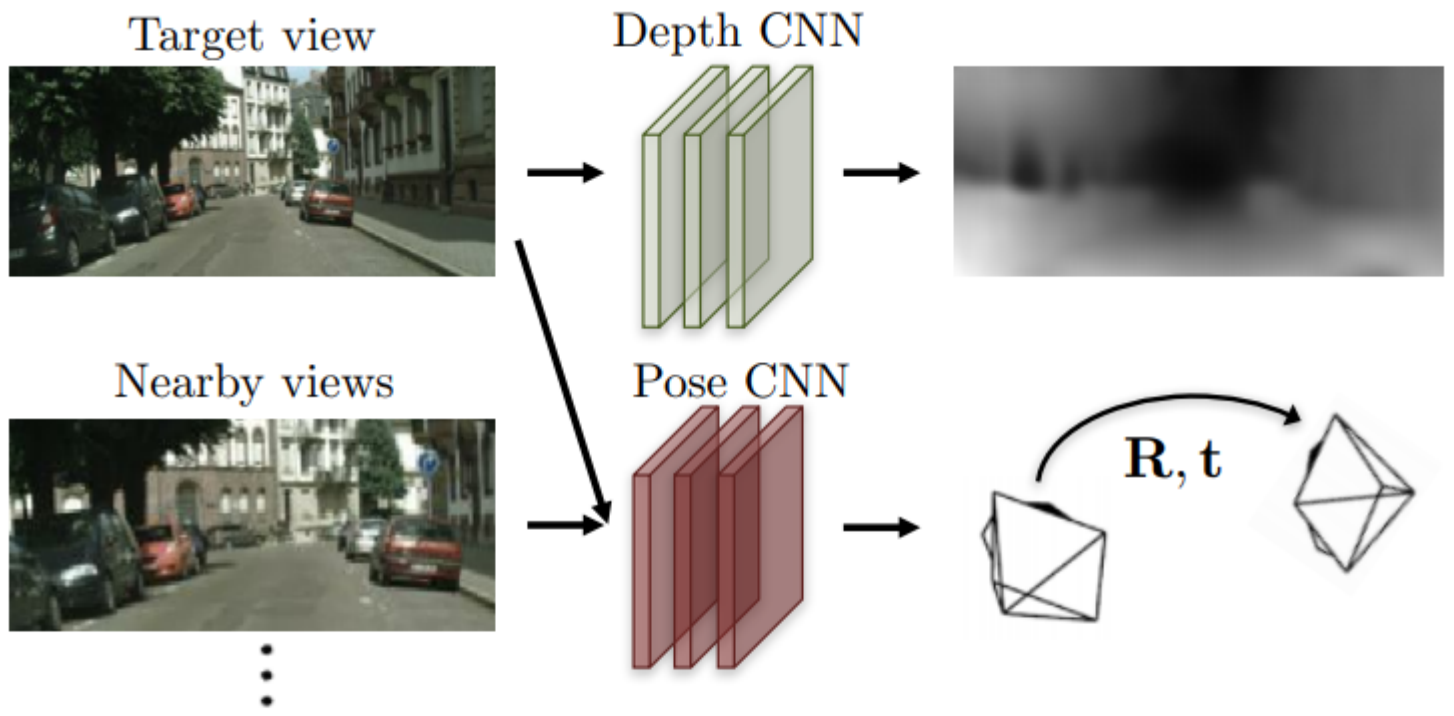
Unsupervised?

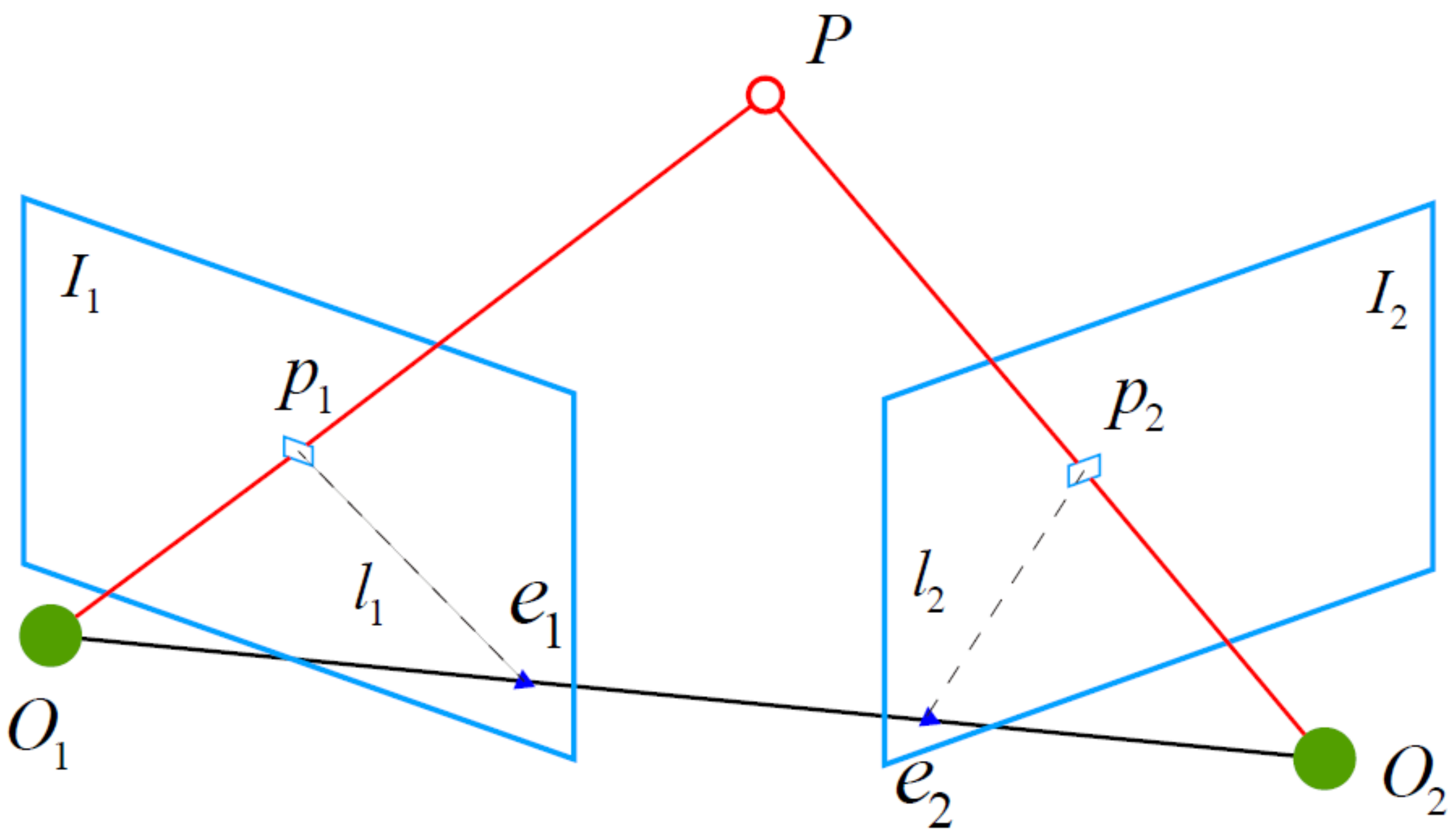
- What happens if depth gt NOT provided ?

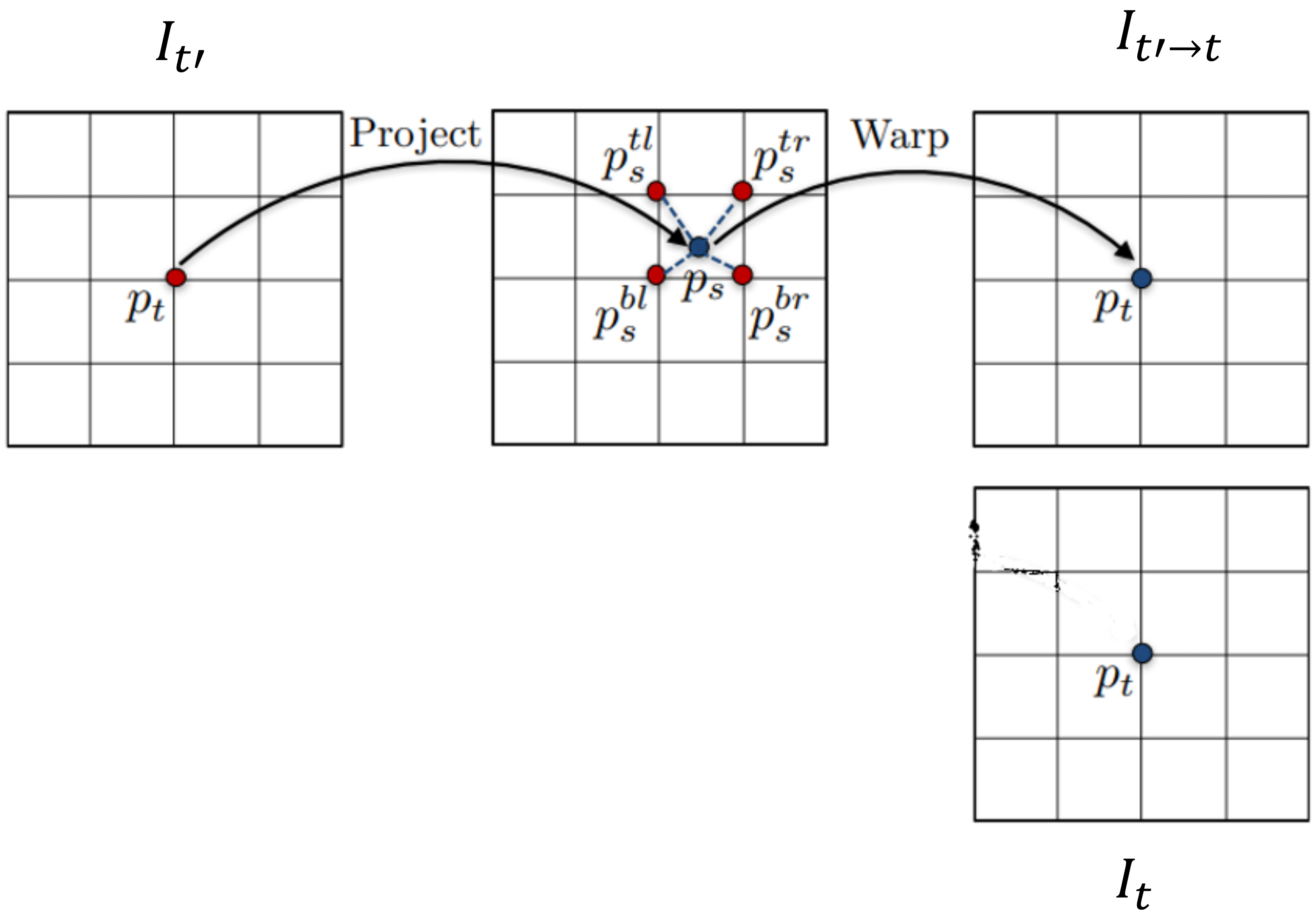
$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma) + const$$

- Recall what we have done in unsupervised task

Recall Monodepth2 (ICCV 2019)





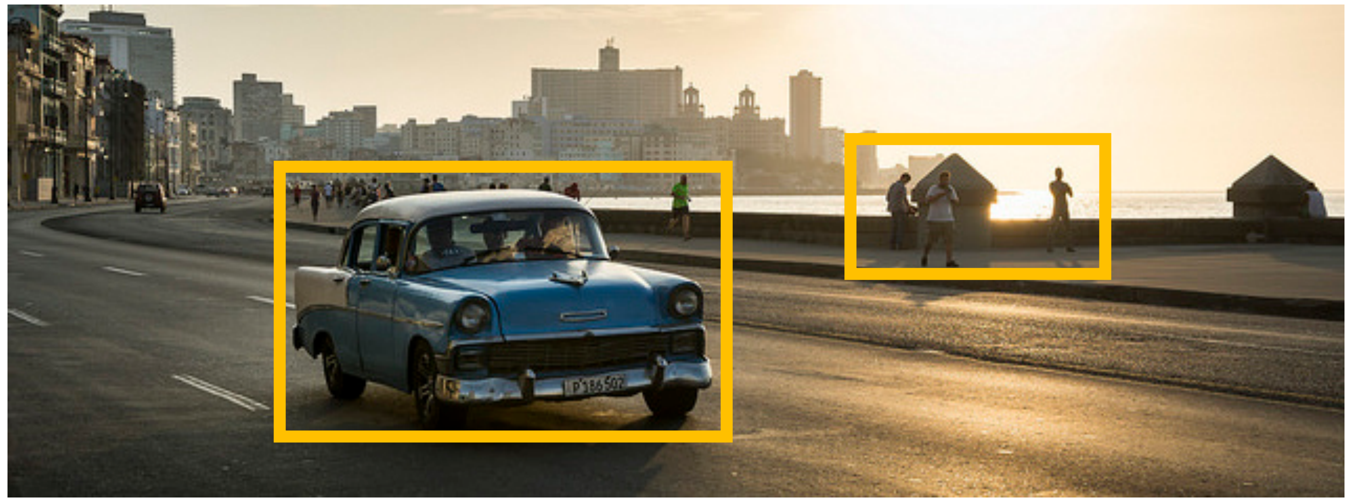


$$L1\ loss = L_{proj} = |I_{t' \to t} - I_t|$$

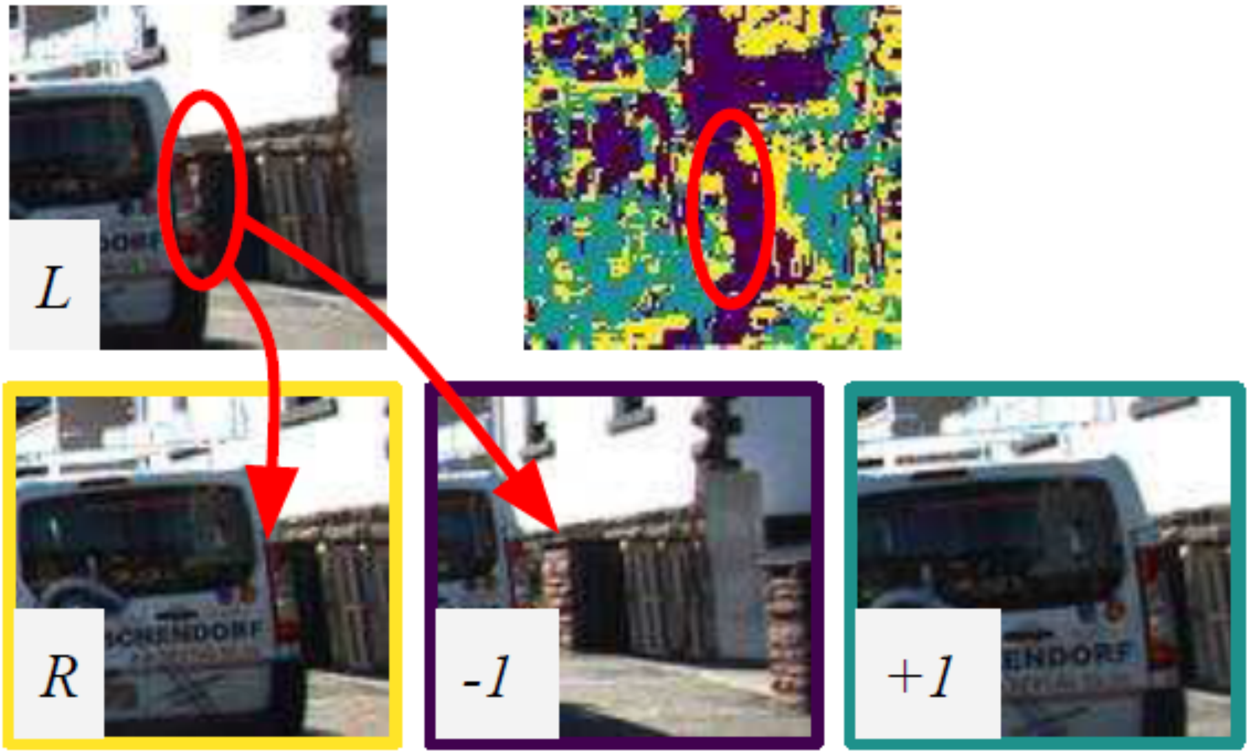
Uncertainty cases

- Brightness constant assumption?
  - non-Lambertian surfaces,
  - reflective materials

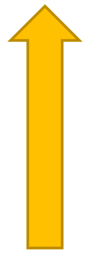
  - moving objects



  - occlusion

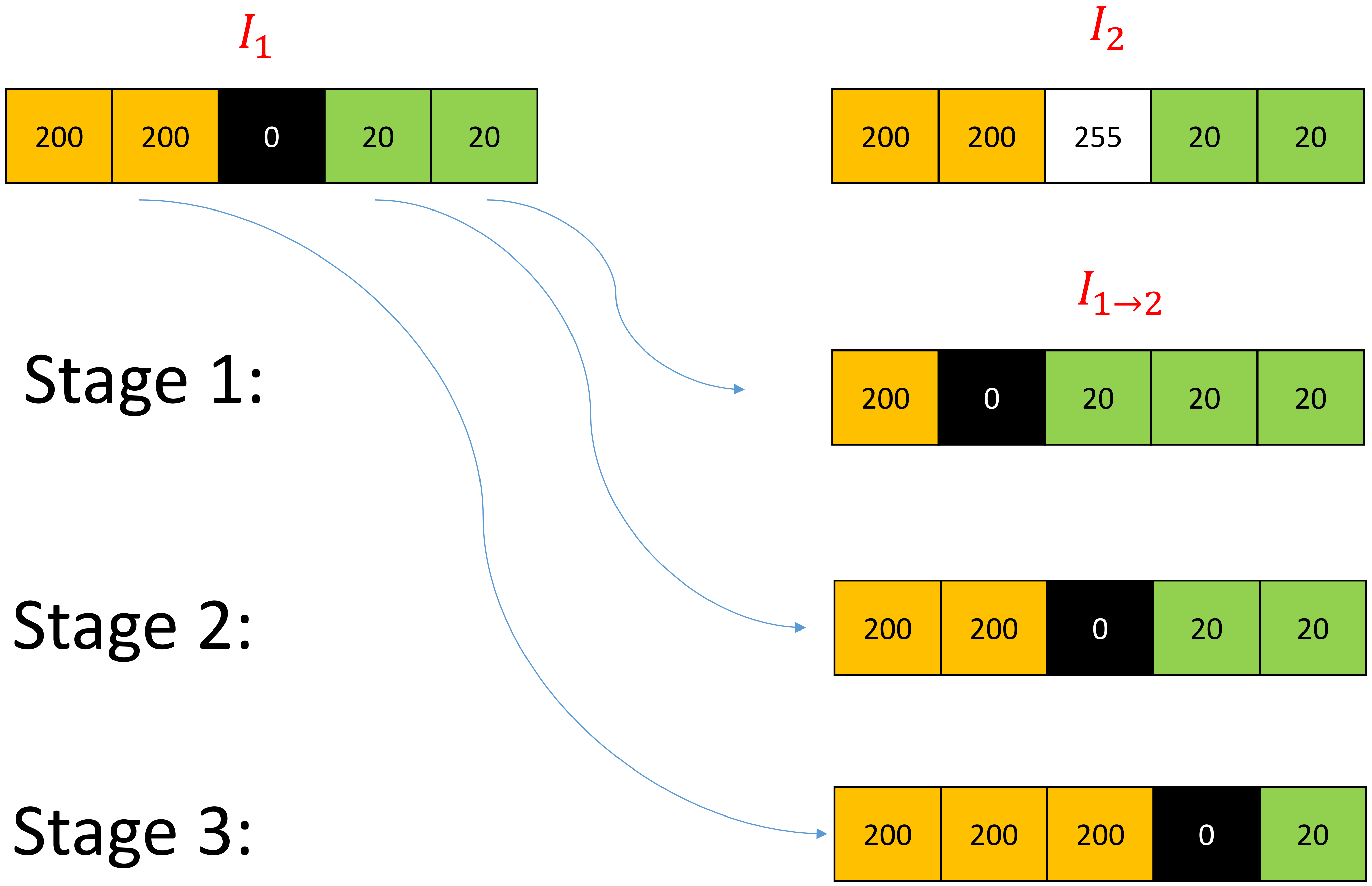


$$L1\ loss = L_{proj} = |I_{t' \to t} - I_t|$$



Uncertainty cases

- How will occlusion cases affect performance?



$$L_{proj} = |I_{1\to 2} - I_2|$$

$$L_{proj} = 255 - 20 + 200 = 435$$

$$L_{proj} = 255,\ \text{gt!}$$

$$L_{proj} = 255 - 200 + 20 = 75,$$

lower loss…

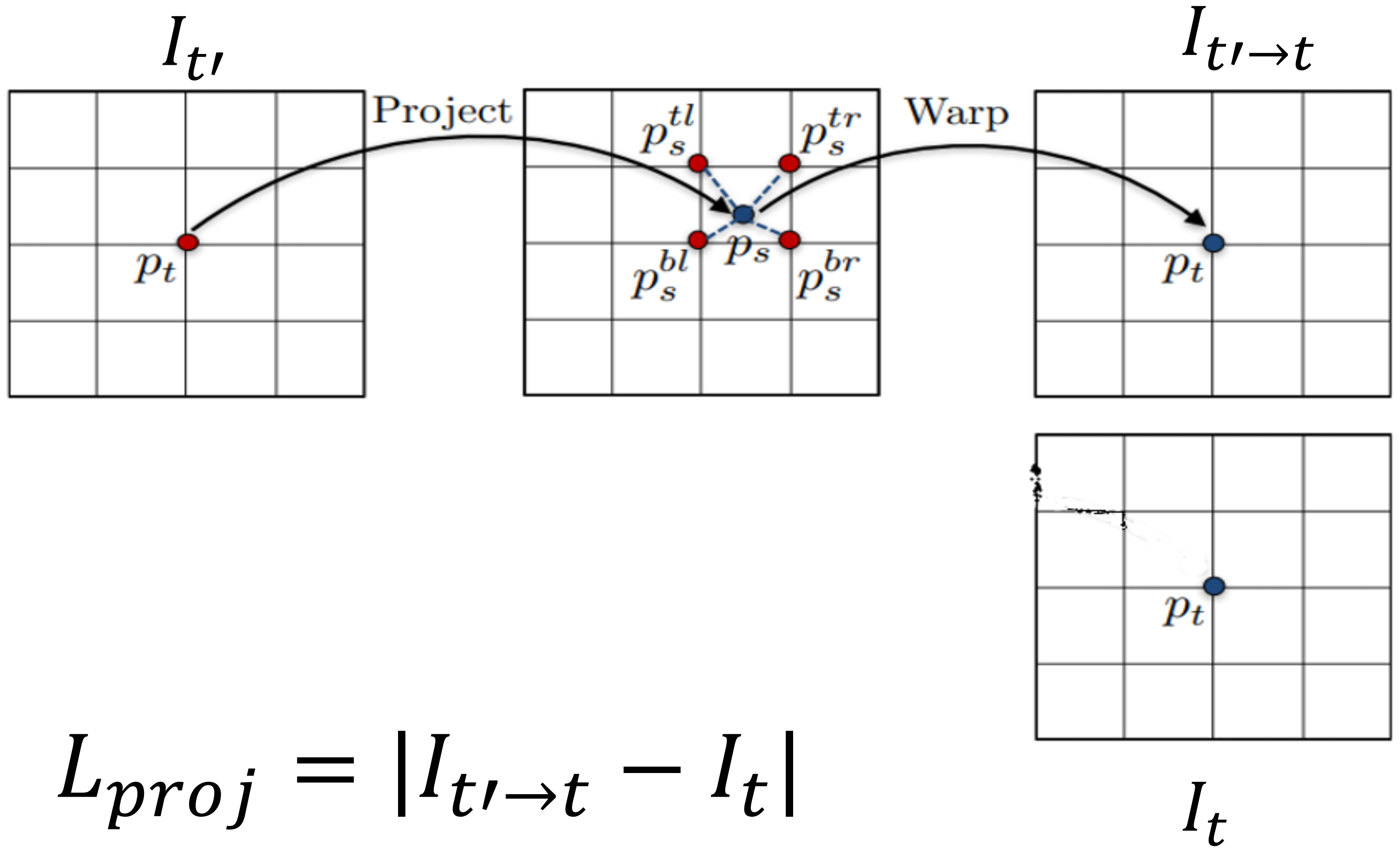
mono uncertainty(CVPR 2020)

- Recall uncertainty model in supervised learning

$$L = -\log p\left(d \middle| \tilde{d}, \sigma\right) = \frac{\left|d - \tilde{d}\right|}{\sigma} + \log(\sigma)$$

- D3VO, CVPR 2020
- Mono-uncertainty, CVPR 2020

$$L = \frac{|I_{t' \to t} - I_t|}{\sigma} + \log(\sigma)$$



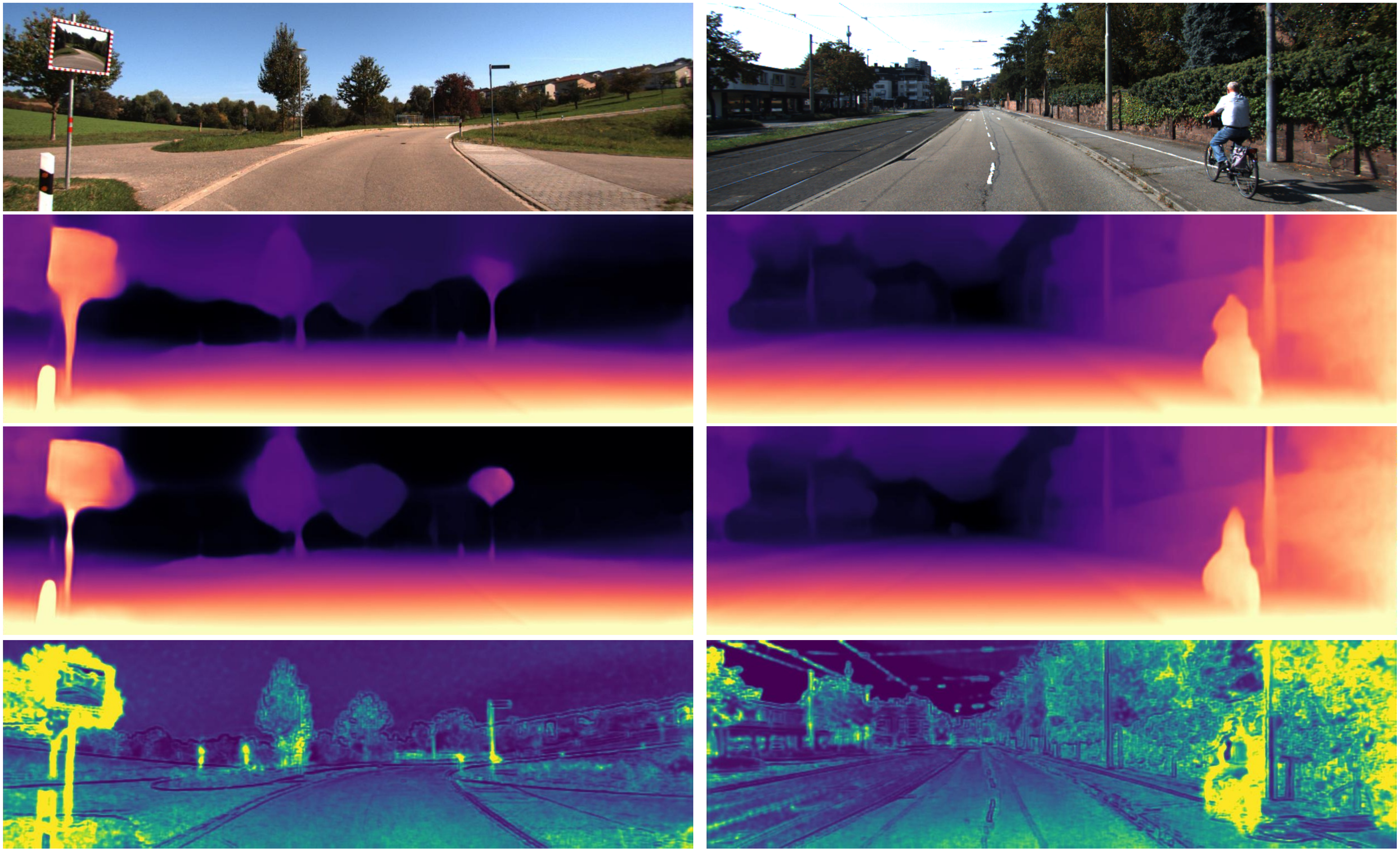
$$L_{proj} = |I_{t' \to t} - I_t|$$

I

Mono depth

Ours

Uncer map

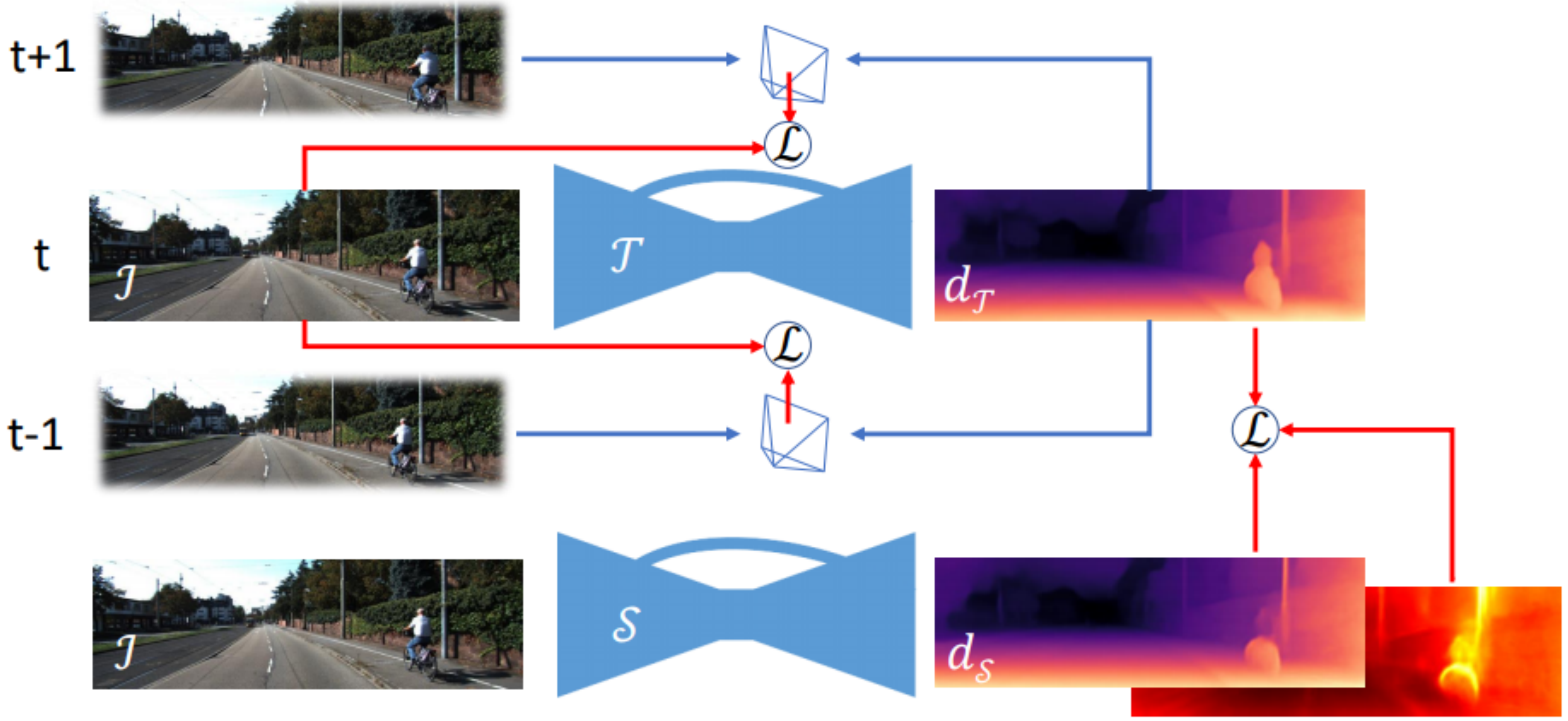


Further thoughts about uncertainty

- 解耦depth net和pose net分别带来的uncertainty? (CVPR 2020)
- High uncertainty区域对low uncertainty区域的影响降低?
- High uncertainty区域能否预测更准? (结合先验信息或mask? Hard...)
- Inference test预测阶段利用uncertainty信息? (CVPR 2021)

Further thoughts about uncertainty

- 解耦depth net和pose net分别带来的uncertainty?
  - Teacher-student framework (On the uncertainty of self-supervised monocular depth estimation, Matteo Poggi, CVPR 2020)



$$\mathcal{L}_{Self} = \frac{|\mu(d_{\mathcal{S}}) - d_{\mathcal{T}}|}{\sigma(d_{\mathcal{S}})} + \log \sigma(d_{\mathcal{S}})$$

Further thoughts about uncertainty

- High uncertainty区域对low uncertainty区域的影响降低?
  - 动态自适应调节σ和L1 loss的下降速率?

How σ works better? (Adaptive)

- Recall occlusion: what if we introduce $\boldsymbol{\sigma}$ here ?

I_1

200	200	0	20	20

I_2

200	200	255	20	20

$$L_{log} = \frac{|I_{1\to 2} - I_2|}{\sigma} + \log(\sigma)$$

$I_{1\to 2}$

Stage 1:

200	0	20	20	20

$$L_{proj} = 255 - 20 + 200 = 435$$

Stage 2:

200	200	0	20	20

$$L_{proj} = 255, \text{ gt!}$$

Stage 3:

200	200	200	0	20

$$L_{proj} = 255 - 200 + 20 = 75,$$

lower loss…

Thanks

- Q & A