Bilateral Cyclic Constraint and Adaptive Regularization for Unsupervised Monocular Depth Prediction

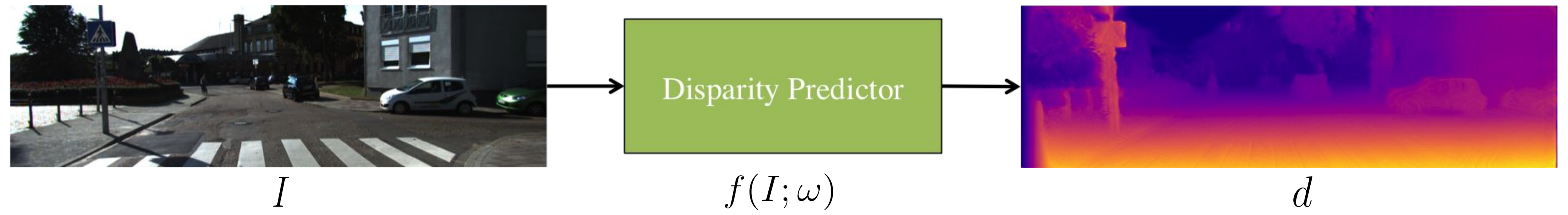
Code available at:
https://tinyurl.com/yx93mylh

Alex Wong[1] alexw@cs.ucla.edu
Byung-Woo Hong[2] hong@cau.ac.kr
Stefano Soatto[1] soatto@cs.ucla.edu
University of California, Los Angeles[1] Chung-Ang University, Korea[2]
UCLA VISION LAB

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

Goal: Learn a function $d = f(I; \omega)$ to recover the scene from a single image I

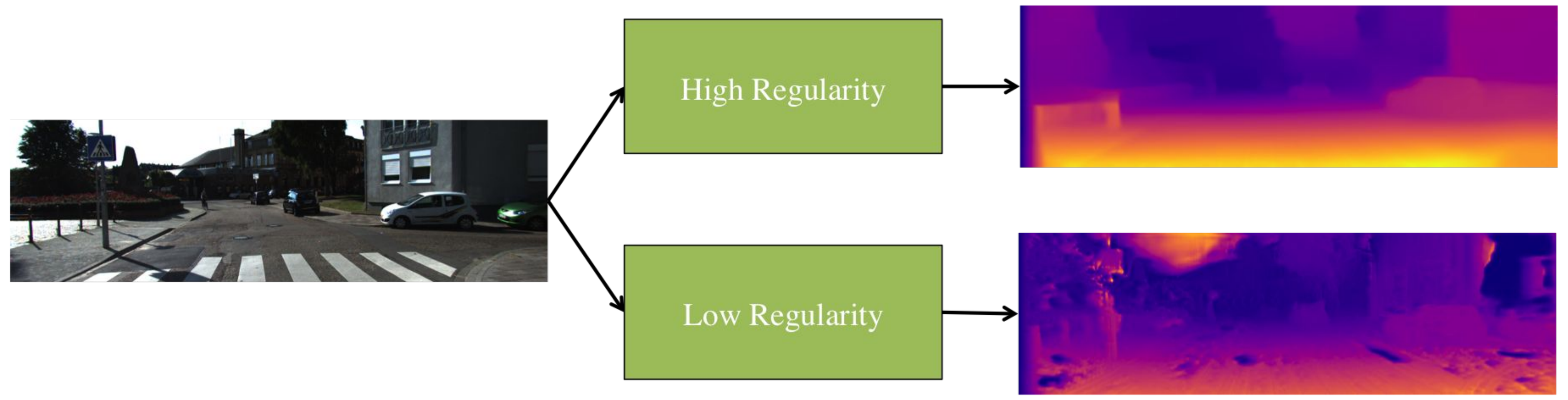


Recovering 3D geometry from a single image is an **ill-posed** problem

We must rely on a prior e.g. piecewise smoothness

We formulate this as an energy minimization problem: $\mathcal{D}(d) + \alpha\mathcal{R}(d)$

where $\mathcal{D}$ denotes data fidelity, $\mathcal{R}$ regularization, and α a static scalar

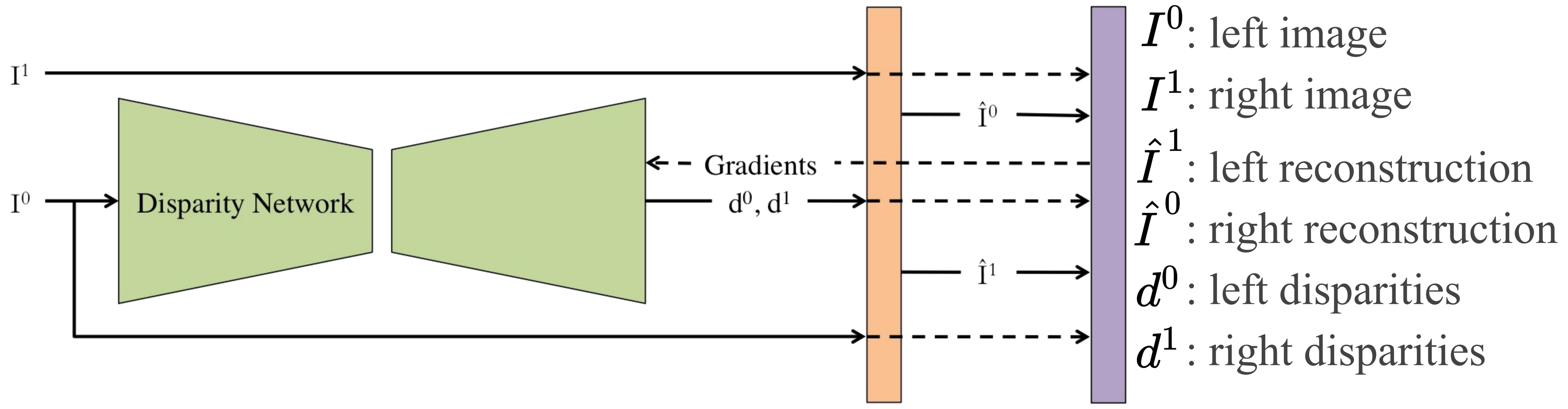


To better modulate the amount of regularity imposed: α should be adaptive

System Diagram

Exploiting stereo pairs and view synthesis loss for training

Given a single image we predict its left and right disparities



where the block in **orange** denotes a bilinear sampler and **purple** the loss layer

Loss Function

Data-fidelity terms: L_1 photometric ($\mathcal{L}_{ph}$) and illumination invariant SSIM ($\mathcal{L}_{st}$)

Regularization: local smoothness ($\mathcal{L}_{sm}$) and bilateral cyclic consistency ($\mathcal{L}_{bc}$)

$$\mathcal{L} = \underbrace{w_{ph}\mathcal{L}_{ph} + w_{st}\mathcal{L}_{st}}_{\text{data-fidelity}} + \underbrace{w_{sm}\mathcal{L}_{sm} + w_{bc}\mathcal{L}_{bc}}_{\text{regularization}}$$

$\mathcal{L}_{sm}$ and $\mathcal{L}_{bc}$ are adaptively weighted by a spatially and training time varying α

Adaptive Regularization

As $\mathcal{R}(d)$ biases the solution based on an assumption about d,
we should only impose regularity if $\mathcal{D}(d)$ is met; hence, we want α to vary for:

- each position $x \in \Omega$ of d depending on the residual of $\mathcal{D}(d)$
- each time step of training

$$\alpha(x) = \exp\left(-\frac{c\rho(x)}{\sigma}\right)$$

for local residual $\rho(x) = |I(x) - \hat{I}(x)|$ and global residual $\sigma = \frac{1}{\frac{1}{|\Omega|}\sum_{x\in\Omega}|I(x) - \hat{I}(x)|}$

$\alpha(x)$ begins low, and $\alpha(x) \rightarrow 1$ as global residual decreases

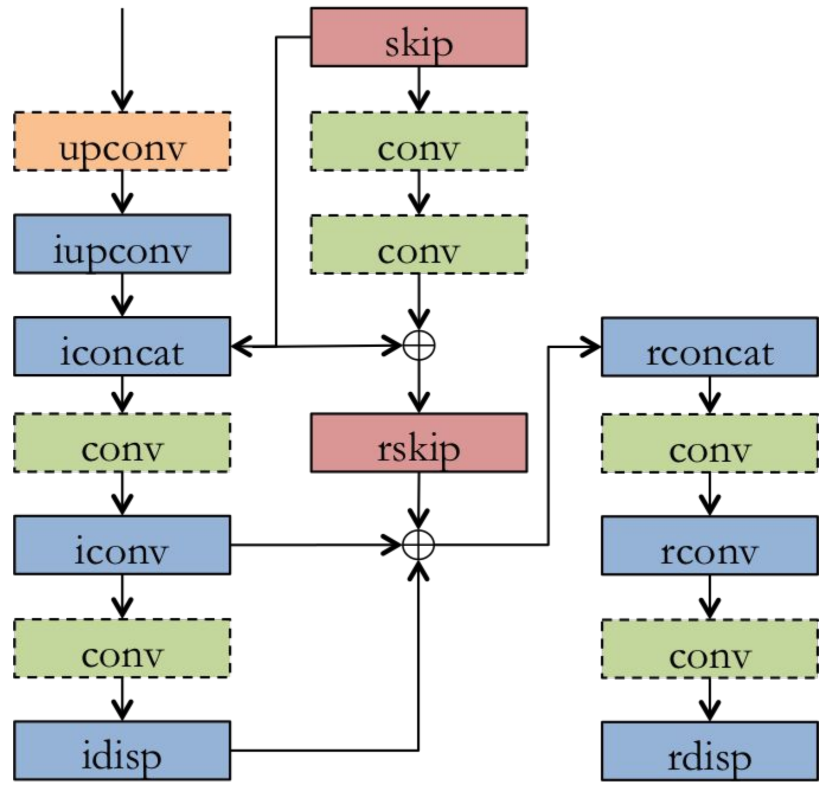
$$\mathcal{L}_{sm} = \sum_{s\in\{0,1\}}\sum_{x\in\Omega}\alpha^s(x)\left(\lambda^s(x)|\partial_X d^s(x)| + \lambda^s(x)|\partial_Y d^s(x)|\right)$$

where $\lambda^s(x) = \exp(-|\nabla^2 I^s(x)|)$

$$\mathcal{L}_{bc} = \sum_{s\in\{0,1\}}\sum_{x\in\Omega}\alpha^s(x)|d^s(x) - \hat{d}^s(x)|$$

where $d^{1p}(x) = d^0(x + d^1(x))$ and $\hat{d}^0(x) = d^{1p}(x - d^0(x))$

Two-Branch Decoder



α relies on data-fidelity residual. To ensure we have the necessary features to satisfy $\mathcal{D}(d)$:

Dedicate one branch to minimizing just the data-fidelity term

Give its features and coarse prediction to a second branch to minimize entire loss

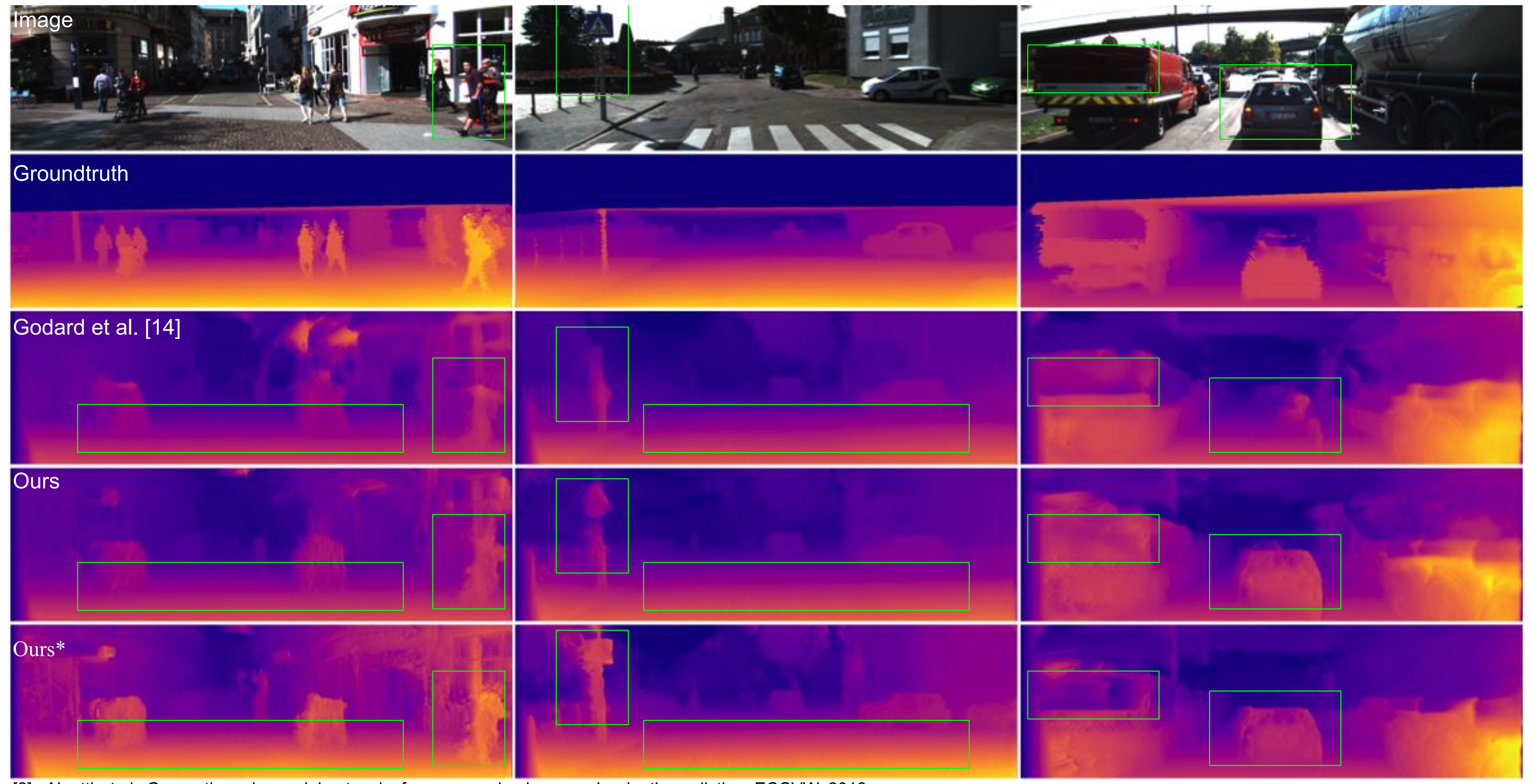
idisp for minimizing : $\mathcal{L}_0 = w_{ph}\mathcal{L}_{ph} + w_{st}\mathcal{L}_{st}$

rdisp for minimizing : $\mathcal{L}$

We use rdisp as our final output

Results

Method	Error Metrics: Abs Rel	Sq Rel	RMS	logRMS	D1-all	Accuracy Metrics: $\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
$ph + st + \lambda^G sm + lr$ [14]	0.124	1.388	6.125	0.217	30.272	0.841	0.936	0.975
$ph + st + \alpha\lambda^G sm + \alpha lr$ ([14] w/ Our Adaptive Regularization)	0.120	1.367	6.013	0.211	30.132	0.849	0.942	0.975
Aleotti et al. [2]	0.119	1.239	5.998	0.212	29.864	0.846	0.940	0.976
$ph + st + \lambda^L sm + bc$ (Ours w/o Adaptive Regularization)	0.117	1.264	5.874	0.207	29.793	0.851	0.944	0.977
$ph + st + \alpha\lambda^L sm + \alpha lr$ (Ours w/o Bilateral Cyclic Consistency)	0.117	1.251	5.876	0.206	29.536	0.851	0.944	0.977
$ph + st + \alpha\lambda^G sm + \alpha bc$ (Ours w/o Bidirectional Edge-Awareness)	0.115	1.211	5.743	0.203	28.942	0.852	0.945	0.977
$ph + st + \alpha\lambda^L sm + \alpha bc$ (Ours Full Model)	0.114	1.172	5.651	0.202	28.142	0.855	**0.947**	0.979
$ph + st + \alpha\lambda^L sm + \alpha bc$ * (Ours Full Model w/ 2 Branch Decoder)	**0.110**	**1.119**	**5.576**	**0.200**	**27.149**	**0.856**	**0.947**	**0.980**



[2] Aleotti *et al.*, Generative adversarial networks for unsupervised monocular depth prediction. ECCVW, 2018.
[14] Godard *et al.*, Unsupervised Monocular Depth Estimation with Left-Right Consistency. CVPR, 2017.

This work is supported by NRF-2017R1A2B4006023, NRF-2018R1A4A1059731, ONR N00014-17-1-2072, ARO W911NF-17-1-0304