Leveraging a weakly adversarial paradigm for joint learning of disparity and confidence estimation

Matteo Poggi, Fabio Tosi, Filippo Aleotti, Stefano Mattoccia

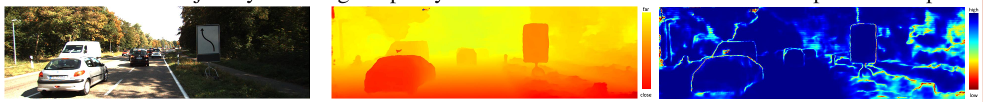
University of Bologna - Department of Computer Science and Engineering

{m.poggi, fabio.tosi5, filippo.aleotti2, stefano.mattoccia}@unibo.it

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Problem Definition and Contributions

Goal: We aim at jointly learning disparity and confidence estimation from input stereo pairs

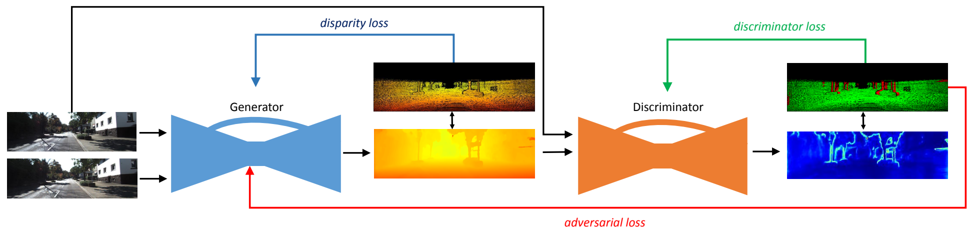


Reference image   Disparity map   Confidence map

Key Contributions:

- We train a confidence **discriminator** to detect wrong disparities estimated by a disparity **generator** and forcing the latter to correct them by means of an adversarial loss
- As training progresses, the wrong disparities reduce, making the adversarial term **weaker**
- Experimental results show improved results on both disparity and confidence estimation

Architecture & Method



Our **generator** estimates a disparity map d from a stereo pair (i^L, i^R) as $d = G(i^L, i^R)$. It is trained to minimize the error wrt ground truth $\hat{d}$ by means of loss $\mathcal{L}_G$

$$\mathcal{L}_G = \mathbb{E}_{\substack{i^L, i^R \sim p_{data}(i^L, i^R) \\ \hat{d} \sim p_{data}(\hat{d})}} [\mathcal{L}_1(G(i^L, i^R), \hat{d})]$$

According to such an error, pixels are classified as correct, $G_1(i^L, i^R)$, or wrong, $G_0(i^L, i^R)$. The **discriminator** infers a confidence map and is trained on labels $G_1(i^L, i^R)$ and $G_0(i^L, i^R)$ to solve a binary classification problem, with loss $\mathcal{L}_D$

$$\mathcal{L}_D = \mathbb{E}_{i^L, i^R \sim p_{data}(i^L, i^R)} [\log D(G(i^L, i^R))]$$

An adversarial term $\mathcal{L}_{\text{WAN}}(G, D)$ is added to $\mathcal{L}_G$ in order to force G to correct pixels $G_0(i^L, i^R)$

$$\mathcal{L}_{WAN}(G, D) = \mathbb{E}_{i^L, i^R \sim p_{data}(i^L, i^R)} [\log(1 - D(G_0(i^L, i^R)))]$$

Experiments & Results

Training on KITTI 2012, testing on KITTI 2015

Disparity evaluation

	>2(%)		>3(%)		>4(%)		>5(%)		MAE	
Model	Noc	All	Noc	All	Noc	All	Noc	All	Noc	All
PSMNet [1]	5.850	6.490	2.736	3.131	1.911	2.186	1.561	1.765	1.163	1.203
Heteroscedastic-PSMNet [2]	5.871	6.562	2.903	3.439	2.047	2.487	1.675	2.052	1.087	1.164
Reflective-PSMNet [3]	**5.670**	**6.209**	2.736	3.108	1.936	2.216	1.585	1.804	1.325	1.369
WAN-PSMNet (ours)	5.687	6.246	**2.681**	**3.062**	**1.885**	**2.176**	**1.528**	**1.762**	**0.972**	**1.025**

Confidence estimation

Estimator	AUC_{opt}	AUC	AUCM
CCNN	0.398	1.265	0.867
ConfNet	0.398	2.282	1.884
LGC-Net	0.398	1.059	0.661
Heteroscedastic	0.395	0.955	0.560
Reflective	0.450	1.250	0.800
WAN	0.358	0.908	**0.550**

Training on Middlebury *trainingQ*, testing on *additionalQ*

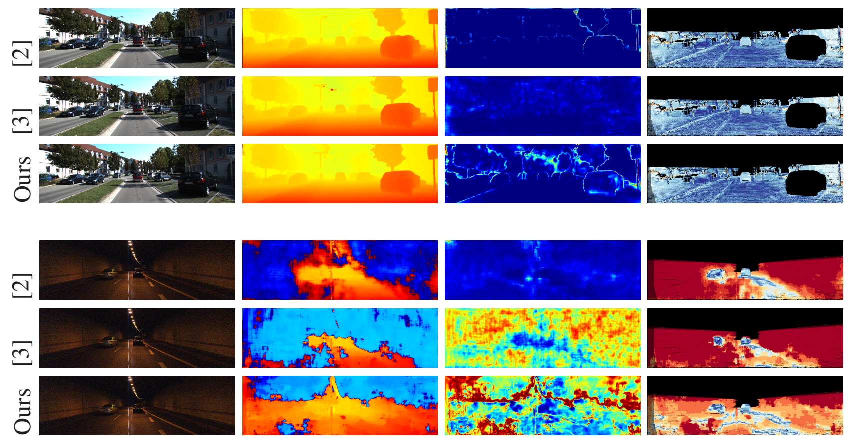
Disparity estimation

Model	>1(%)	>2(%)	>4(%)	MAE
PSMNet [1]	26.121	14.547	8.536	1.920
Heteroscedastic-PSMNet [2]	33.458	18.887	11.722	2.874
Reflective-PSMNet [3]	26.002	14.689	7.159	1.911
WAN-PSMNet (ours)	**25.496**	**14.476**	**7.132**	**1.906**

Confidence estimation

	AUC_{opt}	AUC	AUCM
CCNN	0.046	0.217	0.176
ConfNet	0.046	0.248	0.207
LGC-Net	0.046	0.194	**0.148**
Heteroscedastic	0.090	0.363	0.273
Reflective	0.045	0.166	0.191
WAN	0.041	0.194	0.153

Qualitative results



References

[1] Chang and Chen: Pyramid Stereo Matching Network (CVPR 2018)

[2] Kendal and Gal: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? (NIPS 2017)

[3] Shaked and Wolf: Improved stereo matching with constant highway networks and reflective confidence learning (CVPR 2017)