Geometry-Aware Symmetric Domain Adaptation for Monocular Depth Estimation

Shanshan Zhao, Huan Fu, Mingming Gong, Dacheng Tao

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

1. Introduction

Challenge

1. The depth labels are expensive to acquire in supervised learning.
   Solvable: Exploring unsupervised cues.
2. It tends to be vulnerable to illumination change, occlusion, etc.
   Solvable: Exploiting synthetic data with ground truth depth.
3. The model fails to perform well on real data due to the domain shift.
   Solvable: Utilizing domain adaptation techniques, like GAN based image-to-image translation.

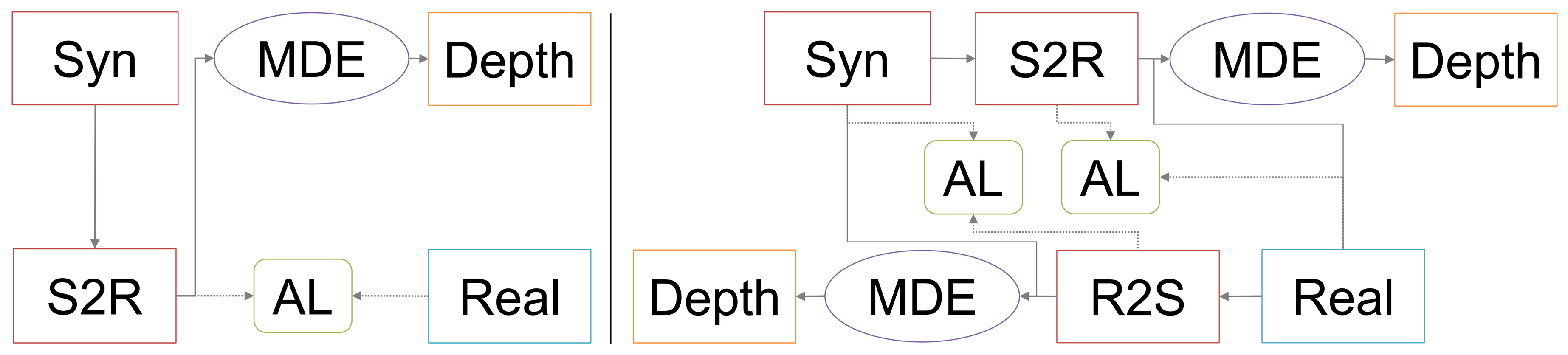
Existing Methods

- Overlooking the *specific geometric structure* in real domain.
- *Undesirable distortions* introduced by the I2I translation process.

Our Solution

Exploring *the labels in the synthetic data* and *epipolar geometry in the real data* jointly (***GASDA***).

Comparison



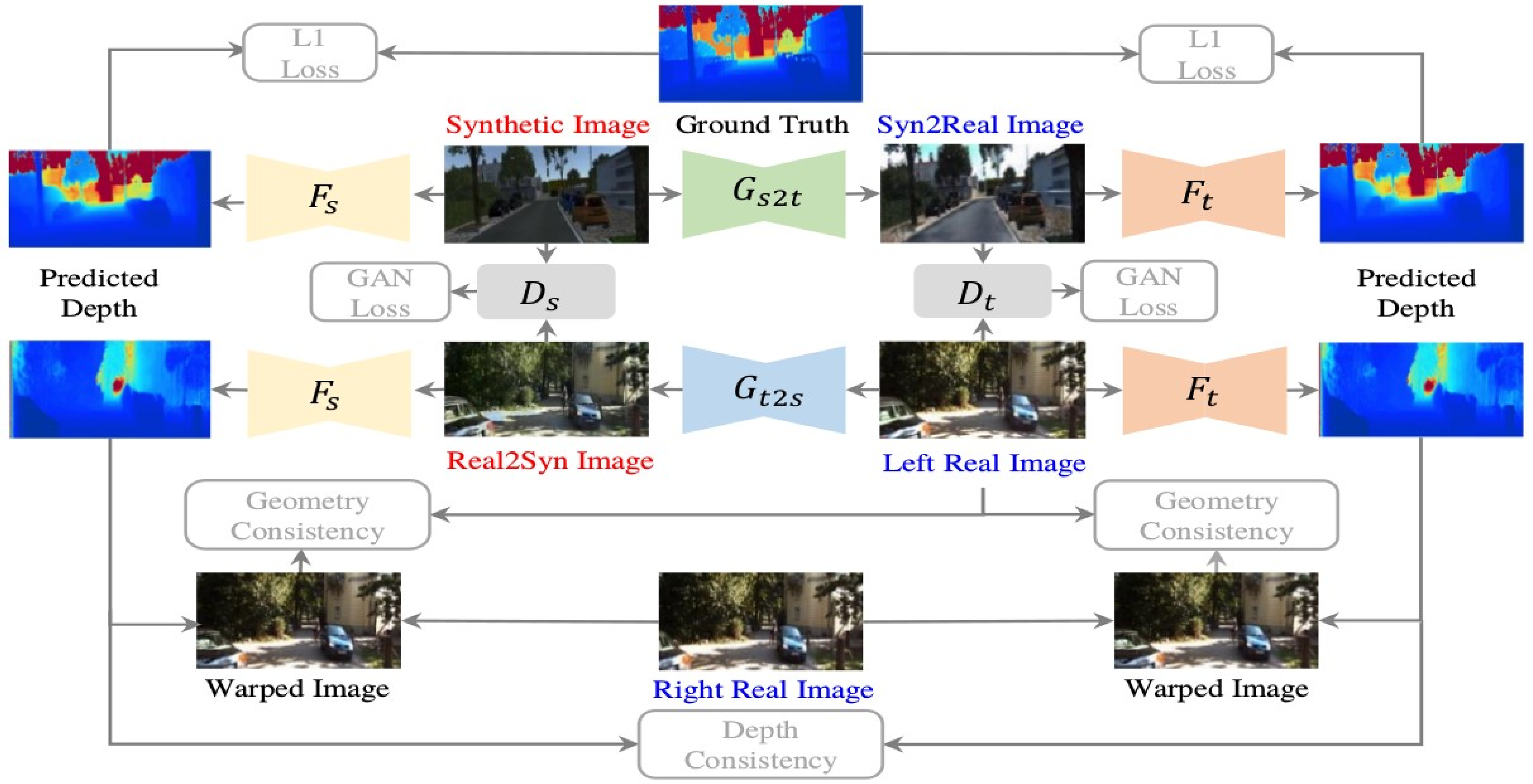
Basic *GASDA*

Contributions

- We propose a novel geometry-aware symmetric domain adaptation network *by exploiting the epipolar geometry of the stereo images*.
- The proposed model can generate high-quality results *for both image style translation and depth estimation*.

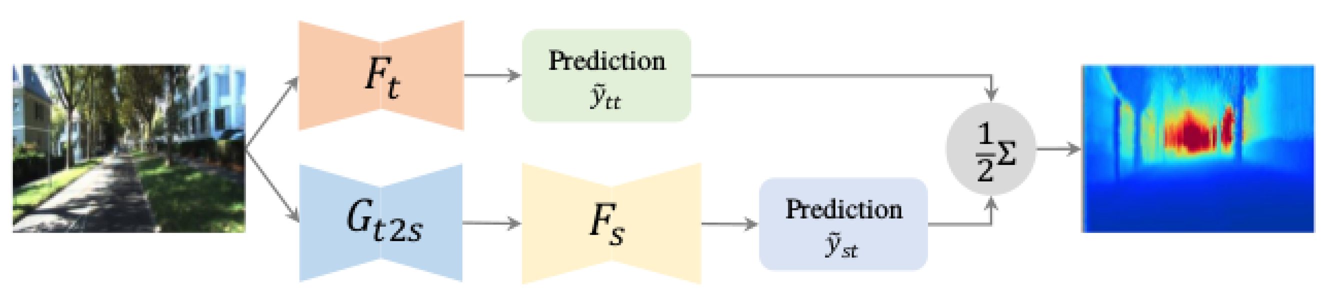
2. Network Architecture

Training



- Simultaneously utilizing both synthetic (with labels) and real (stereo) images.
- Learning domain adaptation and depth estimation in an end-to-end framework.
- Bidirectional style transfer and symmetric structure.

Inference



Project Page

Paper
Code
Poster

Modules

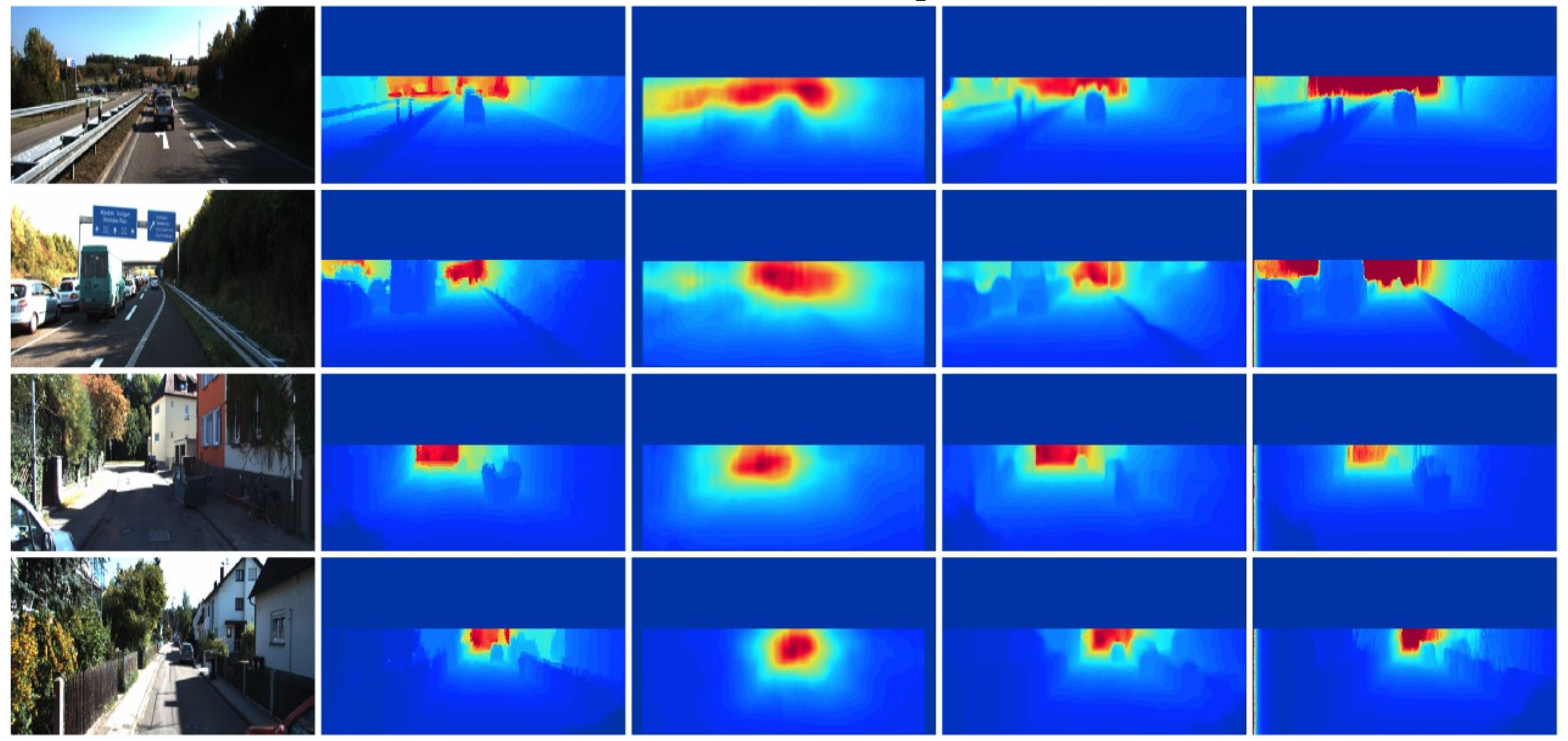
- G_{s2t}/G_{t2s}: transfer the synthetic/real data into the real/synthetic domain.
- D_s/D_t: discriminate the synthetic/real data from the translated data.
- F_s/F_t: depth estimator trained on {Syn, Real2Syn}/{Real, Syn2Real}.

3. Experimental Results

Quantitative Results on KITTI

Method	Supervised	Dataset	Cap	Error Metrics (lower, better)				Accuracy Metrics (higher, better)		
				Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Eigen *et al.* [9]	Yes	K	$80m$	0.203	1.548	6.307	0.282	0.702	0.890	0.958
Liu *et al.* [32]	Yes	K	$80m$	0.202	1.614	6.523	0.275	0.678	0.895	0.965
Zhou *et al.* [56]	No	K	$80m$	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Zhou *et al.* [56]	No	K+CS	$80m$	0.198	1.836	6.565	0.275	0.718	0.901	0.960
Kuznietsov *et al.* [25]	Semi	K	$80m$	0.113	0.741	4.621	0.189	0.862	0.960	0.986
Godard *et al.* [16]	No	K	$80m$	0.148	1.344	5.927	0.247	0.803	0.922	0.964
All synthetic(baseline1)	No	S	$80m$	0.253	2.303	6.953	0.328	0.635	0.856	0.937
All real(baseline2)	No	K	$80m$	0.158	1.151	5.285	0.238	0.811	0.934	0.970
Kundu *et al.* [24]	No	K+S(DA)	$80m$	0.214	1.932	7.157	0.295	0.665	0.882	0.950
Kundu *et al.* [24]	Semi	K+S(DA)	$80m$	0.167	1.257	5.578	0.237	0.771	0.922	0.971
GASDA	No	K+S(DA)	$80m$	**0.152**	**1.044**	**5.065**	**0.230**	**0.820**	**0.940**	**0.972**
Kuznietsov *et al.* [25]	Yes	K	$50m$	0.117	0.597	3.531	0.183	0.861	0.964	0.989
Garg *et al.* [14]	No	K	$50m$	0.169	1.080	5.104	0.273	0.740	0.904	0.962
Godard *et al.* [16]	No	K	$50m$	0.140	0.976	4.471	0.232	0.818	0.931	0.969
All synthetic(baseline1)	No	S	$50m$	0.244	1.771	5.354	0.313	0.647	0.866	0.943
All real(baseline2)	No	K	$50m$	0.151	0.856	4.043	0.227	0.824	0.940	0.973
Kundu *et al.* [24]	No	K+S(DA)	$50m$	0.203	1.734	6.251	0.284	0.687	0.899	0.958
Kundu *et al.* [24]	Semi	K+S(DA)	$50m$	0.162	1.041	4.344	0.225	0.784	0.930	0.974
Zheng *et al.* [55]	No	K+S(DA)	$50m$	0.168	1.199	4.674	0.243	0.772	0.912	0.966
GASDA	No	K+S(DA)	$50m$	**0.146**	**0.784**	**3.895**	**0.219**	**0.832**	**0.945**	**0.975**

Some Examples



Input Image | Ground Truth | Eigen *et.al.*[9] | Zheng *et.al.*[55] | GASDA

Real Image | CycleGAN[57] | GASDA | Synthetic Image | CycleGAN[57] | GASDA