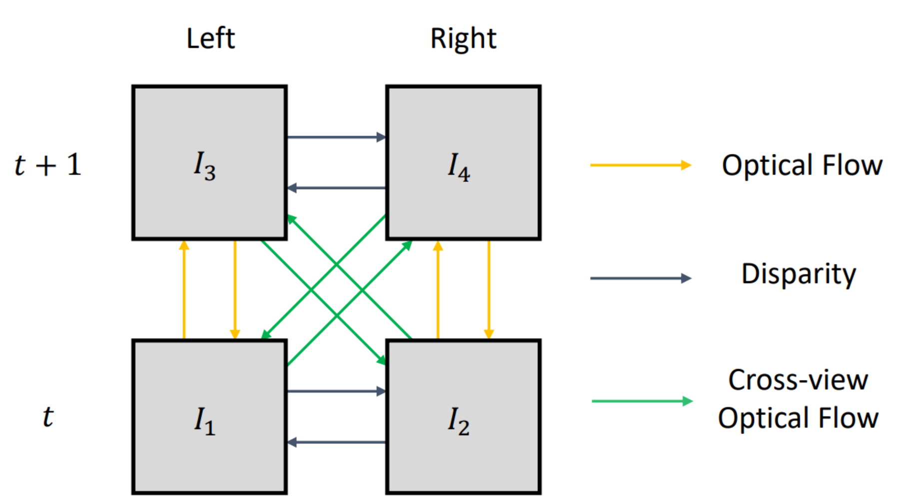
StereoFlowGAN: Co-training for Stereo and Flow with Unsupervised Domain Adaptation

BMVC 2023

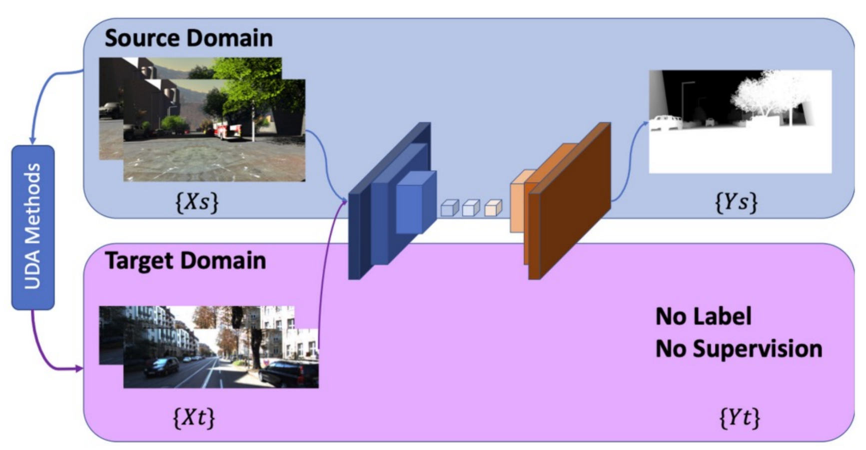
Zhexiao Xiong[1], Feng Qiao[2], Yu Zhang[3], Nathan Jacobs[1]
[1]Washington University in St. Louis, [2]RWTH Aachen University, [3]University of Kentucky

Washington University in St. Louis
James McKelvey School of Engineering

Background



- Stereo matching and optical flow estimation are closely related computer vision tasks



- Unsupervised Domain Adaptation: Relieve the reliance on real data

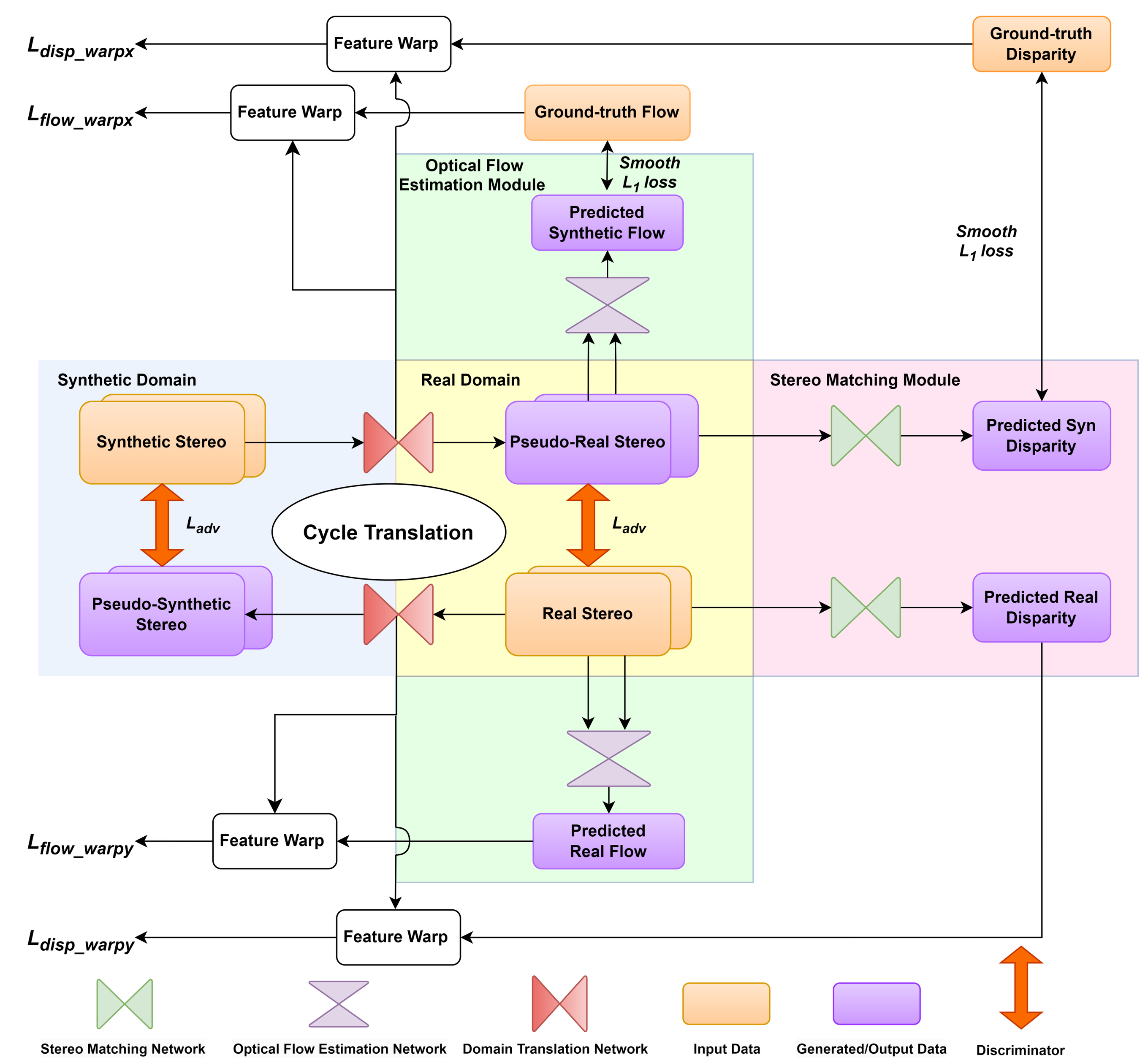
How to utilize the relationships between stereo matching and optical flow estimation to solve unsupervised domain adaptation problem?

Contribution

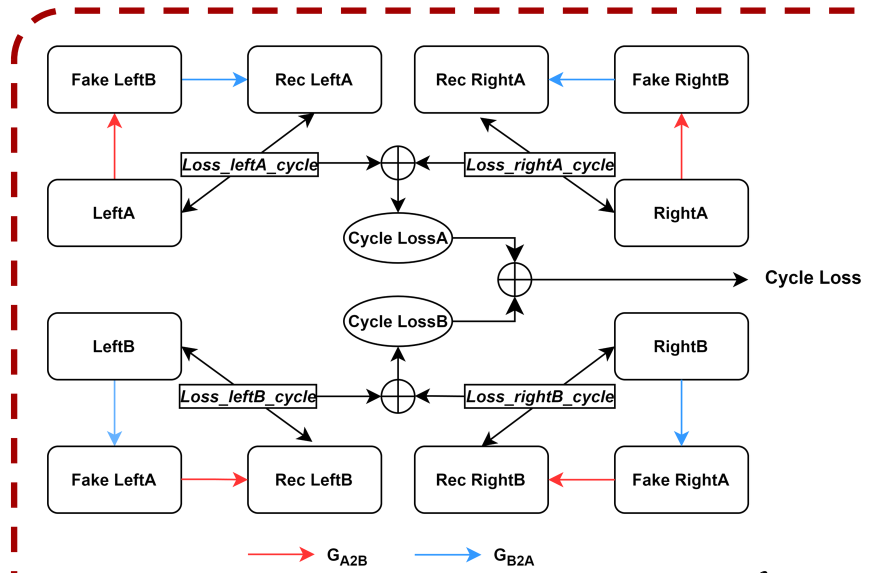
- We build an end-to-end joint learning framework to combine unsupervised domain translation with optical flow estimation and stereo matching in the absence of real ground truth optical flow and disparity, which facilitates the co-optimization of models, yielding superior performance compared with executing each task in isolation.
- We apply novel constraints on the cycle domain translation process to achieve cross-domain translation with global and local consistency, which significantly reduces the pixel distortion during the domain translation stage.
- We employ task-specific multi-scale feature warping loss and iterative feature warping loss during the training phase to regulate the training process of the shared domain translation module and task-specific module in both spatial and temporal dimensions.

Reference:
1. Liu et al. StereoGAN: Bridging synthetic-to-real domain gap by joint optimization of domain translation and stereo matching. In CVPR, 2020.
2. Liu P, King I, Lyu M R, et al. Flow2stereo: Effective self-supervised learning of optical flow and stereo matching, CVPR, 2020.

Framework



Domain Translation



- Cycle Loss:

$$\mathcal{L}_{cyc}(G_{A2B},G_{B2A}) = \mathbb{E}_{y\sim\{\mathcal{Y}_L,\mathcal{Y}_R\}}[\|G_{A2B}(G_{B2A}(y))-y\|_1 + (1-SSIM(G_{A2B}(G_{B2A}(y)-y)))] + \mathbb{E}_{x\sim\{\mathcal{X}_L,\mathcal{X}_R\}}[\|G_{B2A}(G_{A2B}(x))-x\|_1 + (1-SSIM(G_{A2B}(G_{B2A}(y)-y)))],$$

- Perceptual Loss:

$$\mathcal{L}_p(G_{A2B},G_{B2A}) = \mathcal{L}_{perceptual}(G_{A2B}(G_{B2A}(y)),y) + \mathcal{L}_{perceptual}(G_{B2A}(G_{A2B}(x)),x)$$

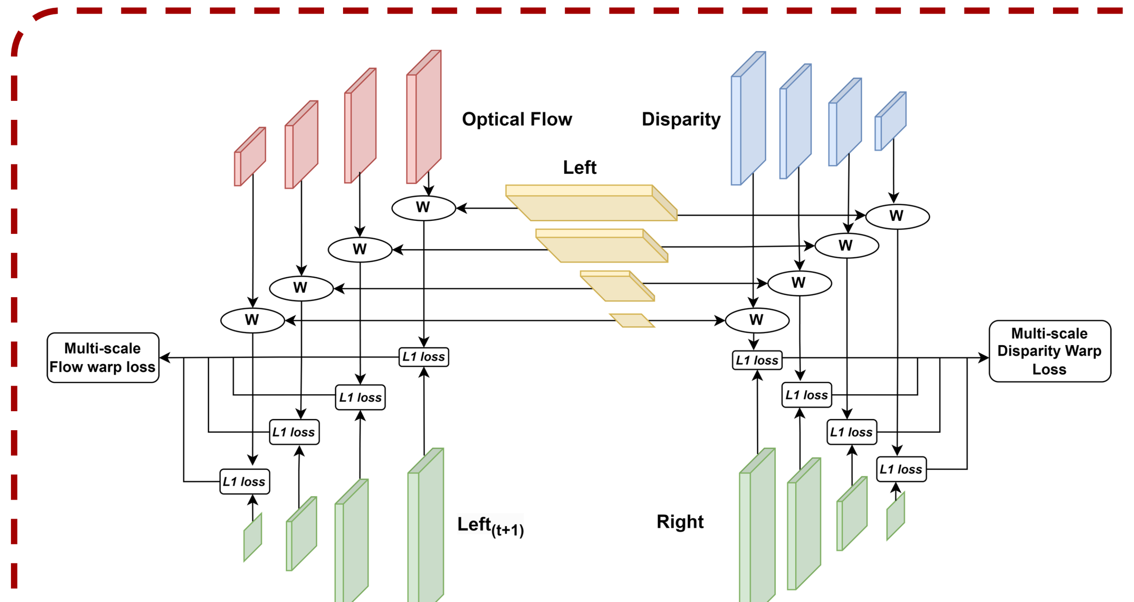
- Cosine Similarity Loss:

$$\mathcal{L}_{cos}(G_{A2B},G_{B2A}) = [1-\cos(G_{A2B}(G_{B2A}(y)),y)] + [1-\cos(G_{B2A}(G_{A2B}(x)),x)]$$

- Domain Translation Loss:

$$\mathcal{L}_{translation}(G_{A2B},G_{B2A},D_A,D_B) = \mathcal{L}_{adv}(G_{A2B},D_B,\mathcal{X},\mathcal{Y}) + \mathcal{L}_{adv}(G_{B2A},D_A,\mathcal{Y},\mathcal{X}) + \lambda_{cyc}\mathcal{L}_{cyc}(G_{A2B},G_{B2A}) + \mathcal{L}_p(G_{A2B},G_{B2A}) + \mathcal{L}_{cos}(G_{A2B},G_{B2A})$$

Feature Warping



- Left-right feature warping
- Forward-backward feature warping
- Multi-Scales
- Both in the real domain and synthetic domain

Co-training

- Domain Translation:

$$\mathcal{L}_T(G_{A2B},G_{B2A},D_A,D_B) = \mathcal{L}_{translation}(G_{A2B},G_{B2A},D_A,D_B) + \lambda_{f_{disp\_warpx}}\mathcal{L}_{f_{disp\_warpx}}(G_{A2B},G_{B2A}) + \lambda_{f_{flow\_warpx}}\mathcal{L}_{f_{flow\_warpx}}(G_{A2B},G_{B2A}) + \lambda_{corr}\mathcal{L}_{corr}(G_{A2B},G_{B2A}) + \lambda_{ms}\mathcal{L}_{ms}(G_{A2B})$$

- Stereo Matching:

$$\mathcal{L}_d(F_{disp},G_{B2A}) = \lambda_{disp}\mathcal{L}_{disp}(F_{disp}) + \lambda_{f_{disp\_warpy}}\mathcal{L}_{f_{disp\_warpy}}(G_{B2A})$$

- Optical Flow Estimation:

$$\mathcal{L}_f(F_{flow},G_{B2A}) = \lambda_{flow}\mathcal{L}_{flow}(F_{flow}) + \lambda_{f_{flow\_warpy}}\mathcal{L}_{f_{flow\_warpy}}(G_{B2A})$$

Feature warping: a link between domain translation module and task-specific modules

Experiments

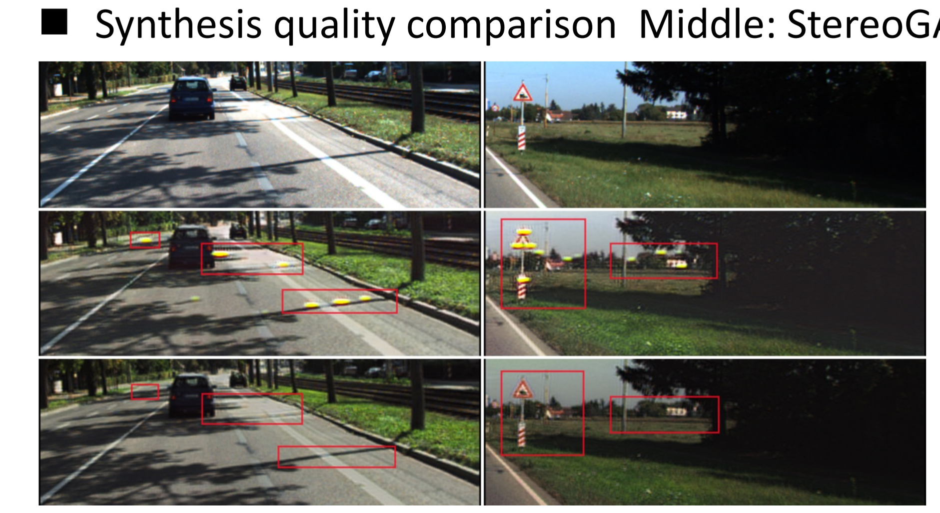
Virtual-KITTI2 to KITTI2015:

Method	EPE	D1-all(%)	>2px(%)	>4px(%)	>5px(%)	EPE(flow)	F1-all(%)
IGEV-Stereo source only	1.01	3.80	7.91	2.78	2.23	–	–
Stereo GAN [22]	0.98	3.59	7.52	2.67	2.13	–	–
Unimatch-flow source only	–	–	–	–	–	5.79	21.81
proposed	**0.93**	**3.18**	**7.04**	**2.37**	**1.90**	**5.19**	**18.32**

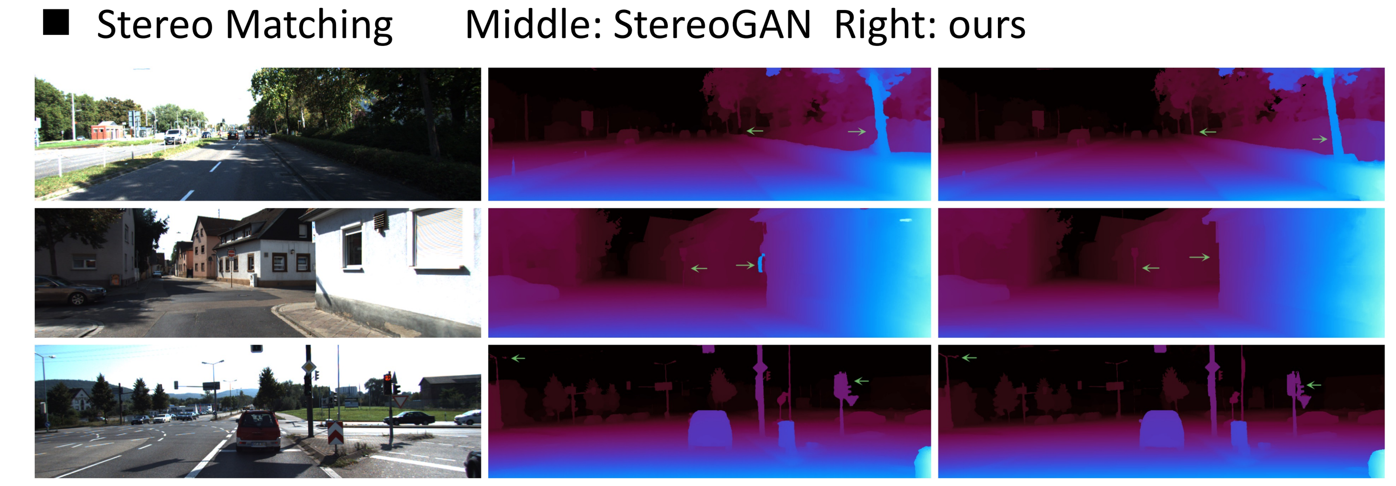
Better performance than source-only results on both stereo matching and optical flow estimation tasks!

Visualization:

- Synthesis quality comparison Middle: StereoGAN Bottom: ours



- Stereo Matching Middle: StereoGAN Right: ours



- Optical Flow Estimation: Middle: Unimatch-flow Source-only Right: ours