Deep 360° Optical Flow Estimation Based on Multi-Projection Fusion

Yiheng Li[1], Connelly Barnes[2], Kun Huang[1], and Fang-Lue Zhang[1,*]

[1]Victoria University of Wellington, New Zealand [2]Adobe Research, Seattle US

ECCV TEL AVIV 2022

Motivation

Optical flow learning for 360° videos remains an interesting and open question in this field. As shown below, the existing methods formulated on 2D regular grids with convolutions that do not inherently deform according to distortion or area changes in the equirectangular projection.

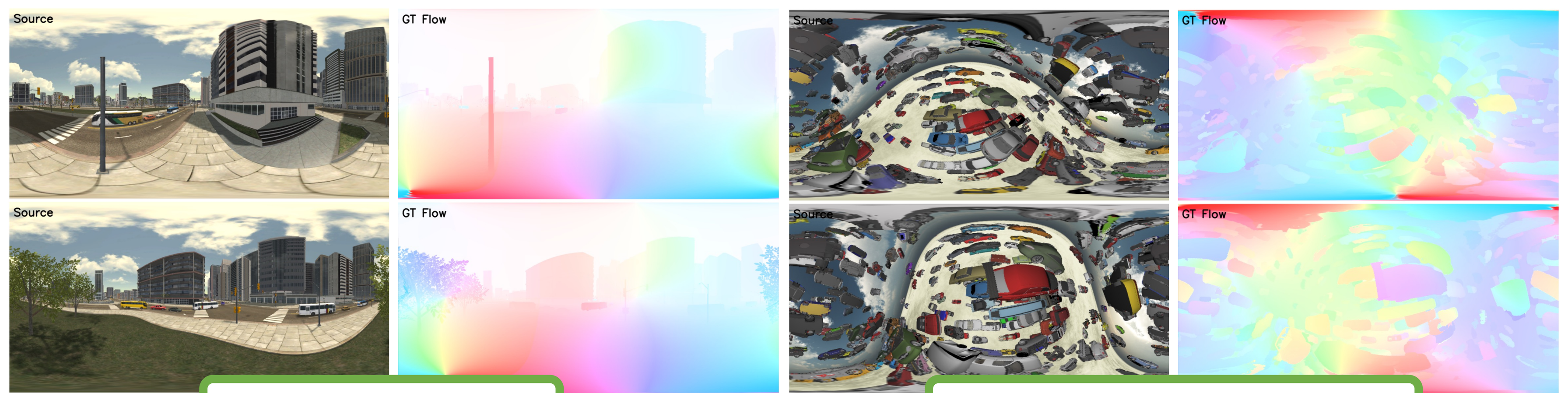
Estimation using narrow-FoV optical flow model

In this research, we were initially motivated to discover good projections for optical flow learning for 360° videos. We investigated three different projections from the sphere to the plane: cylindrical and cube-map, which are conformal within each chart, and equirectangular, which is neither area- nor angle-preserving. To our surprise, our experiments did not seem to indicate that any of our projections is always better than the others, but rather that these projections make complex tradeoffs in optical flow accuracy that appear to depend jointly on the input image and the mathematical properties of each projection.

Panoramic Optical Flow Dataset

We build the first large-scale panoramic optical flow dataset to support the training of neural networks and the evaluation of panoramic optical flow estimation methods. We generate ground truth panoramic optical flow maps by rendering synthetic 360° videos of dynamic 3D scenes. We built our panoramic RGB and optical flow frames using two types of scenes:

City Scene We designed the City scene to resemble the data with similar properties as KITTI and Driving. We generated City2000, City200, and City100 datasets, containing 2000, 217, and 138 frames, respectively. We also generated City100UR (Upright frames) dataset to conduct preliminary experiments.



City Scene **EFT Scene**

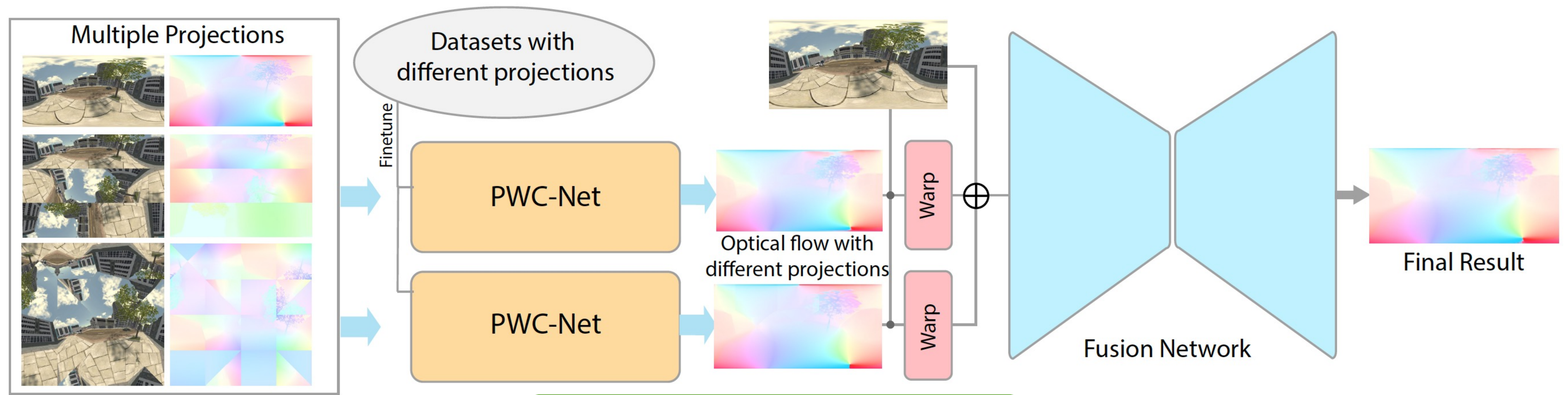
Equirectangular FlyingThings (EFT) Scene We designed the FlyingThings scene to resemble FlyingChairs and FlyingThings3D. We built EFT2000, EFT200, and EFT100 datasets, and each of them contains 2211, 199, and 99 image pairs.

Model	FlowNet2	PWC-Net	PWC-Sph	RAFT-12	RAFT-24	RAFT-48
EPE	4.15	**3.17**	3.80	5.12	4.93	4.72

Table 1. End point errors (EPE) on City100UR using the finetuned models. The models are finetuned using Equirectangular projection.

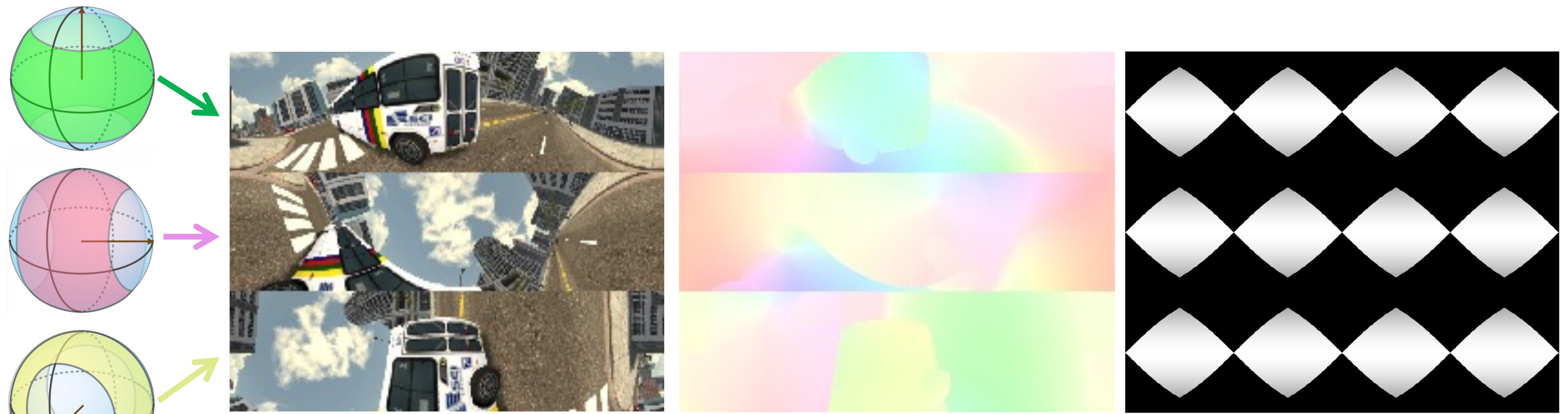
Multi-Projection Fusion

To maximally leverage existing narrow FOV pre-trained models while minimizing the impact of various distortions that occur in spherical projections, we propose to use different methods to project the spherical images to planar images and feed them to finetune the models. A fusion layer is finally built to combine the transformed optical flow results for the final 360° optical flow.



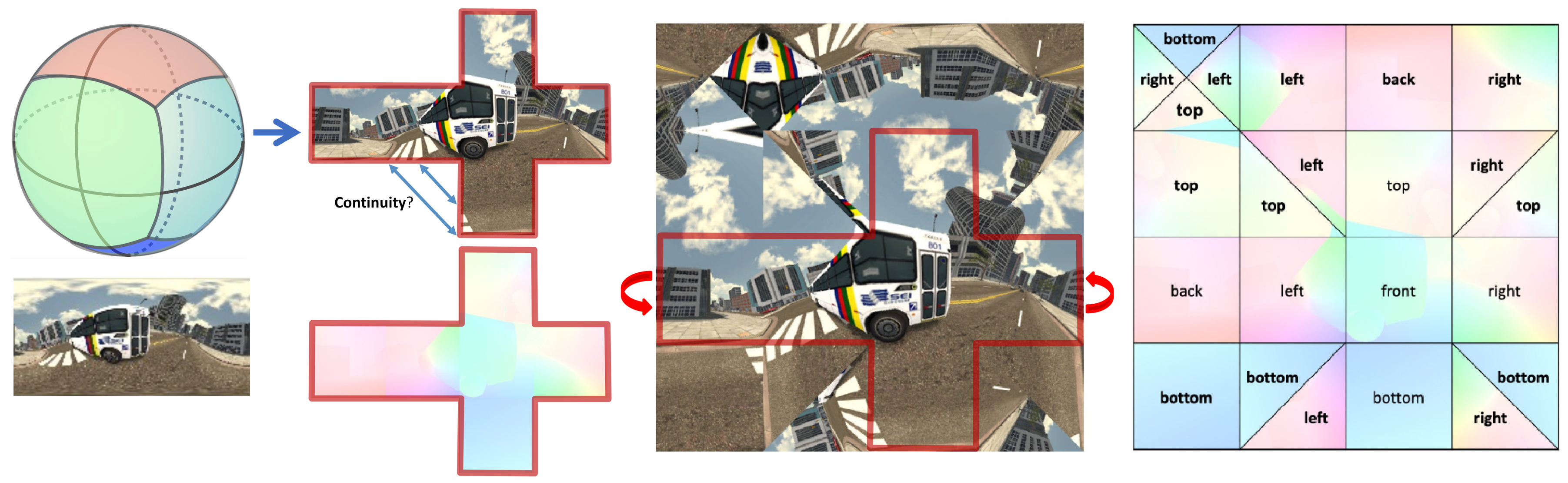
Our Framework

Equirectangular Projection It can keep both C_0 and C_1 continuity in the optical flow map except for the boundary areas. However, it is not conformal and introduces large distortions due to excessively sampling



Tri-Cylindrical Projection

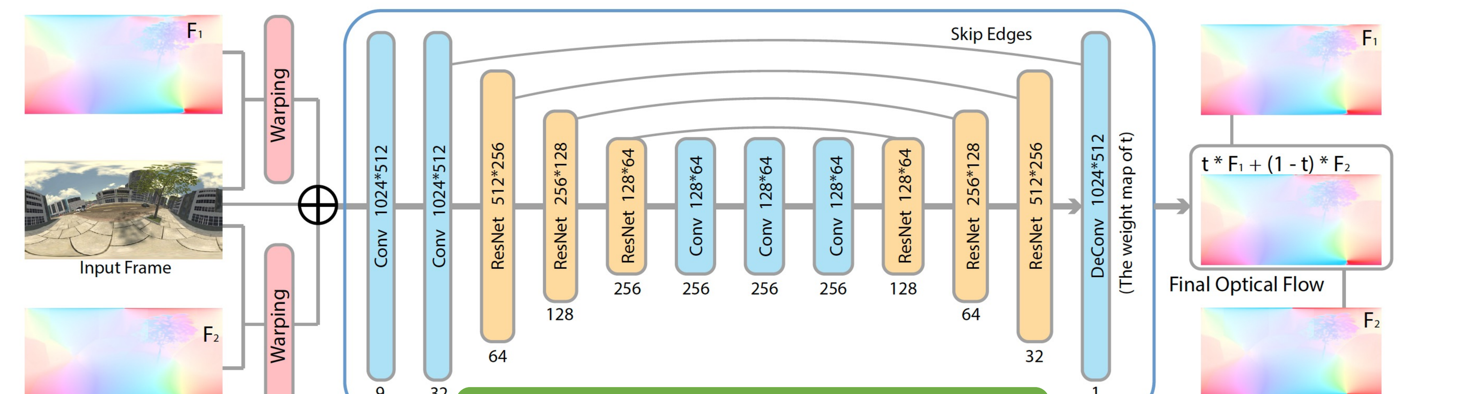
Tri-Cylindrical Projection Unlike equirectangular projection, cylindrical projections preserve angle but are not equal-area. To ensure that every part of the spherical surface has an equal contribution, we stack three cylindrical projection images together, where each cylinder to project is aligned with one of the X, Y, and Z-axis, and use solid angle weights for all the pixels.



Cube-Padding Projection

Cube-Padding Projection Within each face of the cubemap, the projection is angle-preserving but not area preserving, and is suitable for the pre-trained model to be further finetuned. But cube map loses the original adjacency information along the boundary of the cross. We propose to repeatedly pad the faces to stitch the six faces into one picture, keeping the spatial continuity in the projected image.

Fusion Network

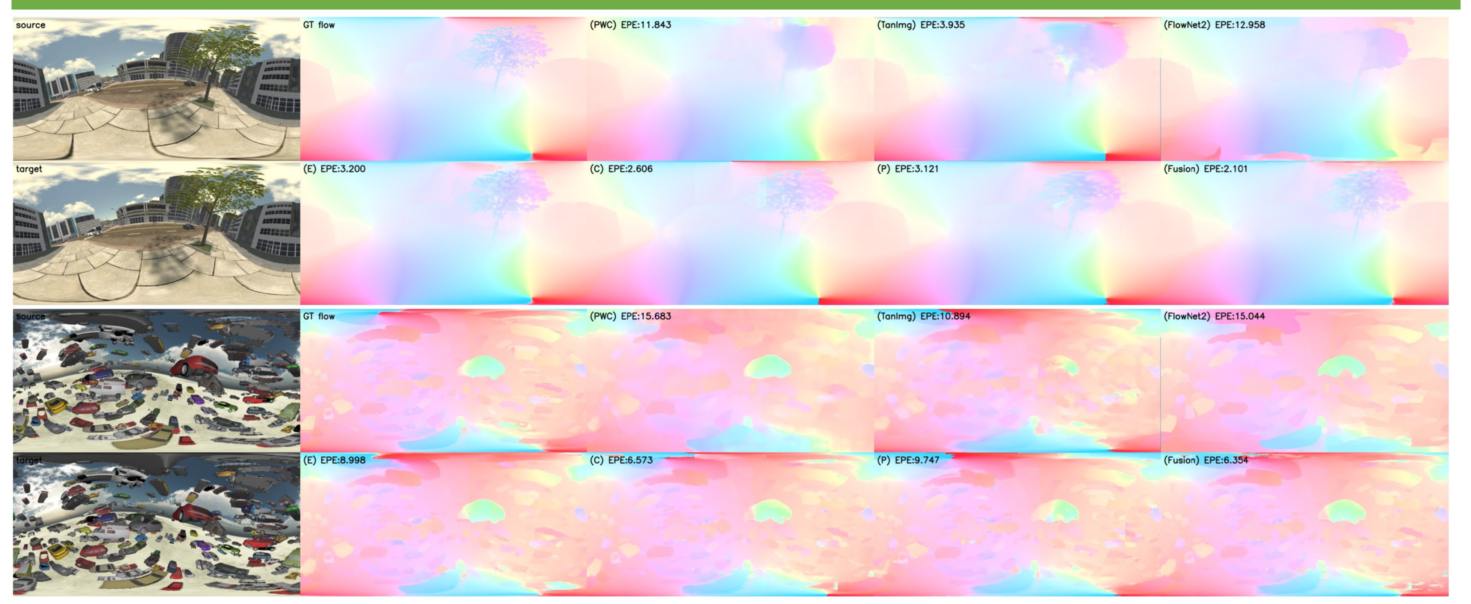


Network Architecture

SD/EPE	Equi (E)	Pad (P)	Cyl (C)	C+P	E+P	E+C
City100	0.90/2.41	0.92/3.32	1.34/2.64	0.82/2.15	**0.75**/1.95	0.83/**1.79**
EFT100	2.98/6.65	2.98/6.25	2.82/5.2	2.77/5.06	2.95/5.84	**2.64/5.01**
Average	1.94/4.91	1.95/5.68	2.08/5.06	1.80/3.60	1.85/3.89	**1.74/3.40**

Table 2. Spherical Distance (SD) and EPE using equirectangular (Equi), cube-padding (Pad), cylindrical (Cyl) projections and their fusion models.

Results



SD/EPE	PWC-Net[1]	RAFT[2]	TanImg[3]	FlowNet2[4]	Ours
City100	3.86/9.84	12.10/21.5	1.27/3.69	2.72/10.85	**0.82/1.79**
EFT100	6.26/15.64	20.21/29.6	4.61/8.06	4.91/14.88	**2.64/5.01**
Average	5.06/12.74	16.15/25.6	2.94/5.88	3.81/12.87	**1.73/3.40**

Table 3. Comparison with other methods using Spherical Distance (SD) / EPE. The SD/EPE of our method is 65.8%/73.3%, 89.3%/86.7%, and 54.6%/73.5% lower than PWC, RAFT, FlowNet2 respectively.

Application

We develop an application that utilizes the predicted optical flow to propagate edits across frames to evaluate our method visually.

Acknowledgement

This work is supported by Marsden Fund Council managed by Royal Society of New Zealand (No. MFP-20-VUW-180).

References

[1] Sun, D. et al.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume, CVPR 2018
[2] Teed, Z. and Deng, J.: RAFT: recurrent all-pairs field transforms for optical flow. ECCV 2020
[3] Yuan, M. and Richardt, C.: 360-degree optical flow using tangent images. BMVC 2021
[4] Ilg, E. et al.: Flownet 2.0: Evolution of optical flow estimation with deep networks. CVPR 2017

Project page: https://github.com/HenryLee0314/ECCV2022-MPF-net