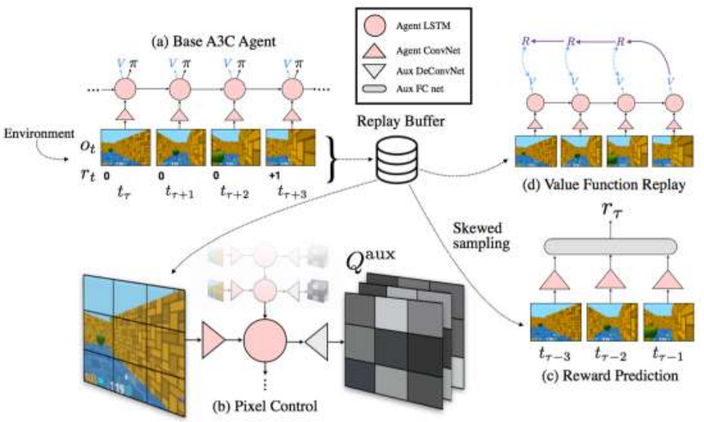
Occlusion Aware Unsupervised Learning of Optical Flow

Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang, Wei Xu
Baidu Research, Institute of Deep Learning (IDL)

Background – Self-supervised Learning

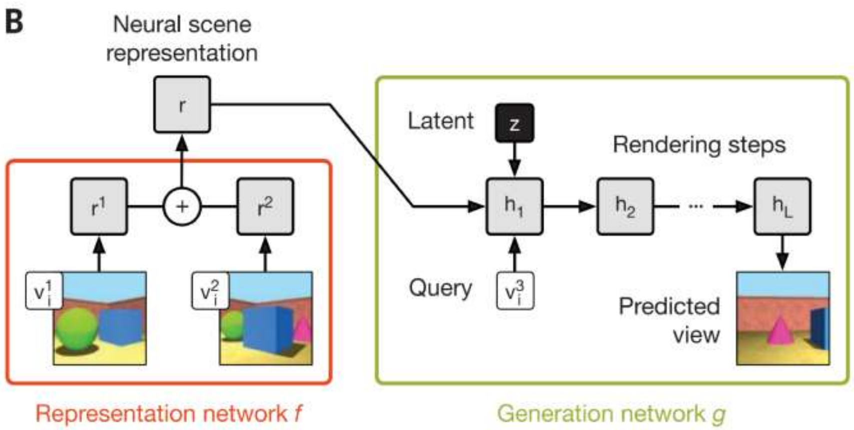
- Learning without human annotation



[1]

(a) learn to explore on Level-1

[2]



[3]

[1]Jaderberg et al, Reinforcement learning with unsupervised auxiliary tasks, ICLR 2017

[2] Pathak et al, Curiosity-driven exploration by self-supervised prediction, ICML 2017

[3] Eslami et al, Natural scene representation and rendering, Science 2018

Cherry On the Cake

Reinforcement Learning

Supervised Learning

Self-supervised Learning

Yann LeCun
March 14, 2016 ·

Statement from a Slashdot post about the AlphaGo victory: "We know now that we don't need any big new breakthroughs to get to true AI"

That is completely, utterly, ridiculously wrong.

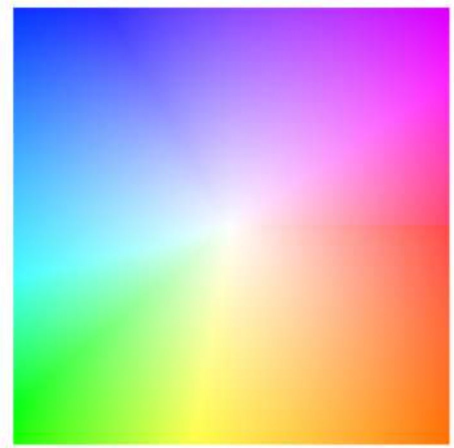
As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.

We need to solve the unsupervised learning problem before we can even think of getting to true AI. And that's just an obstacle we know about. What about all the ones we don't know about?

#deeplearning #AI #AlphaGo

But We Focus on Self-supervised Optical Flow

- Optical flow: $[\Delta x_i, \Delta y_i]$ encodes 2D motion for every pixel i



Flow field color coding. The central pixel does not move, isplacement of every other pixel is the vector from the this pixel.

Brox and Malik, Large displacement optical flow: descriptor matching in variational motion estimation, PAMI 2011

Why Optical Flow?

- Optical flow is very useful but ground truth are hard to obtain
- Optical flow technique can also apply to stereo depth estimation.
- Can be evaluated with standard benchmark dataset.

- Depth and flow extend “state” representation from 2D to 4D in RL.
- RGB, depth and flow are complimentary to each other.
- Stereo Video -> Optical flow + Depth -> Attention / Recognition.

Why Optical Flow?

- A 4 month baby can
  - Follow an object (Tracking, Flow)
  - Reach and grasp an object (Depth, Shape)
  - Pay attention to small objects (Attention)


CHAPTER SEVEN

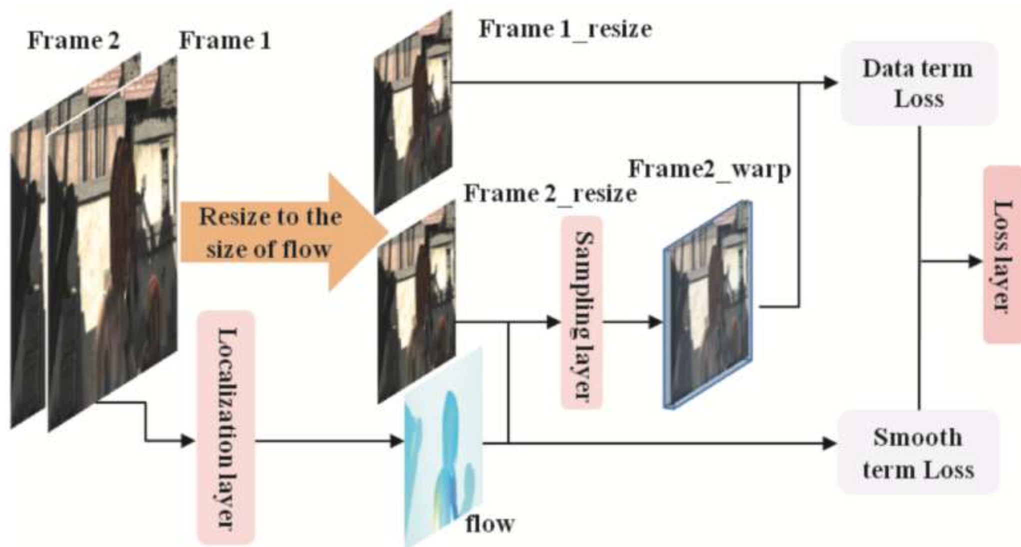
The Fourth Month

WHAT YOUR BABY MAY BE DOING

WHAT YOU CAN EXPECT AT THIS MONTH'S CHECKUP

Previous Work

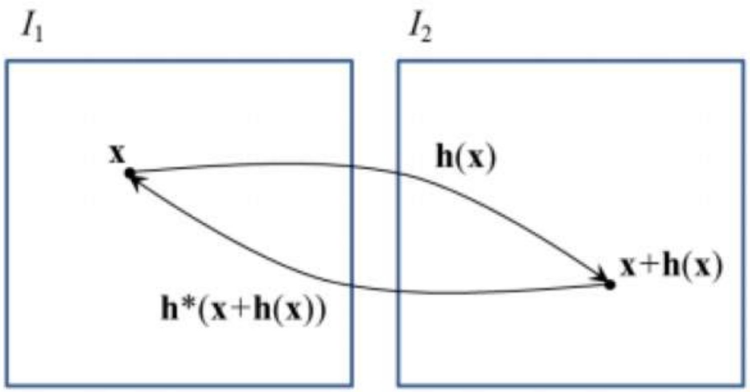
- Idea: Learning to predict optical flow to minimize photometric loss

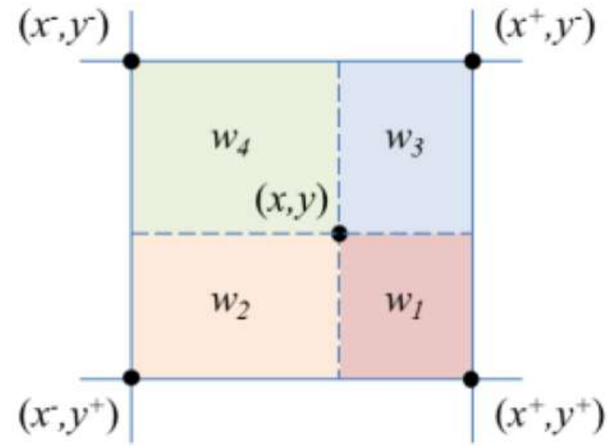


[1] Ahmadi et al, Unsupervised convolutional neural networks for motion estimation, ICIP 2016
[2] Jason et al, Backtobasics: Unsupervised learning of optical flow via brightness constancy and motion, ECCV 2016W
[3] Zhe et al, Unsupervised Deep Learning for Optical Flow Estimation, AAAI 2017

Warping Using Optical Flow

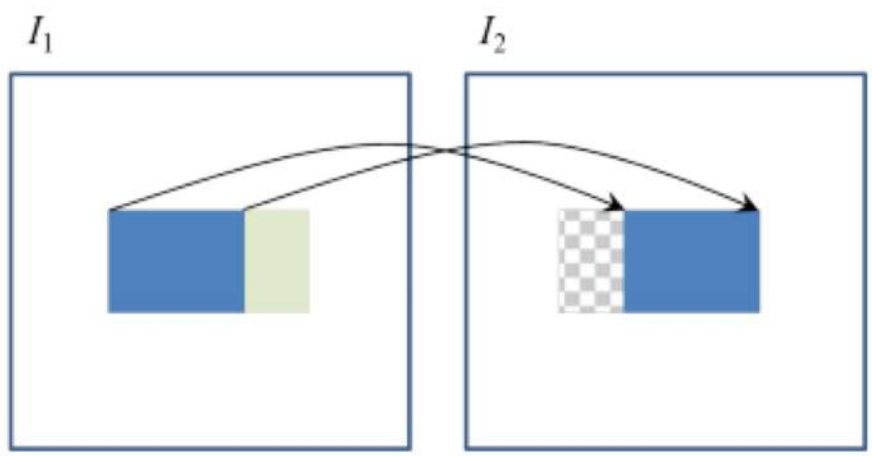
Use image2 and optical flow to re-construct image1 with bilinear interpolation.
Can be implemented through Spatial Transformer Networks.



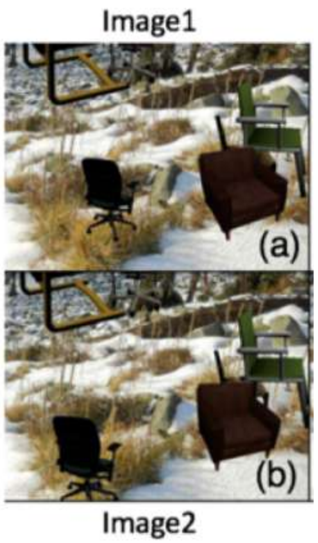


$$\frac{\partial V_i^c}{\partial x_i^s} = \sum_n^H \sum_m^W U_{nm}^c \max(0, 1 - |y_i^s - n|) \begin{cases} 0 & \text{if } |m - x_i^s| \geq 1 \\ 1 & \text{if } m \geq x_i^s \\ -1 & \text{if } m < x_i^s \end{cases}$$

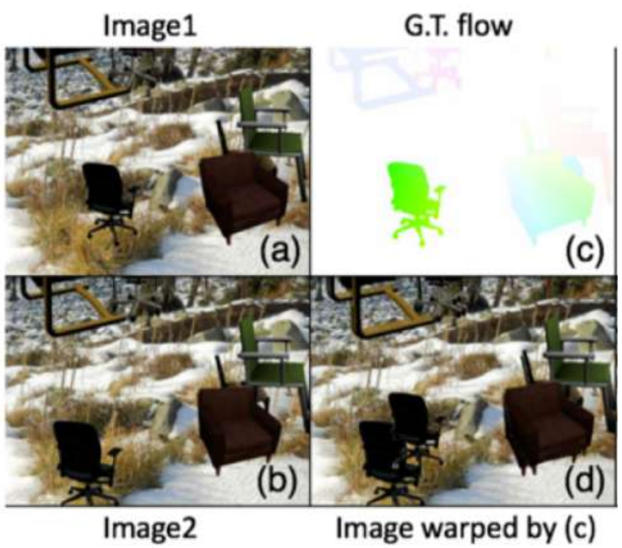
Occlusion Problem in Warping: correct flow but wrong reconstruction



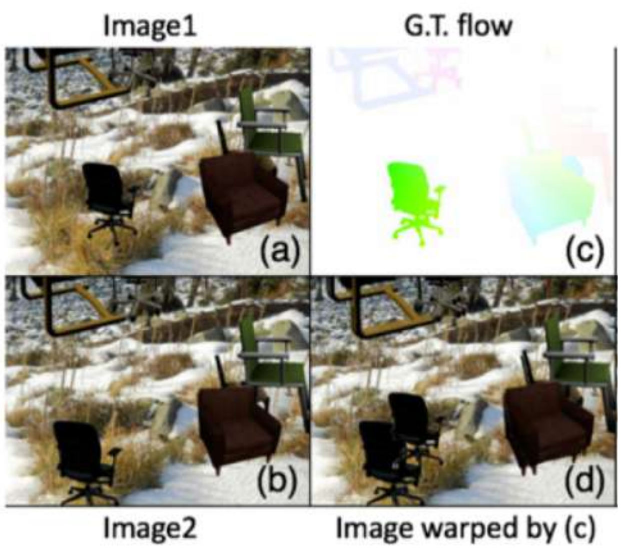
Problem – Occlusion affects Performance

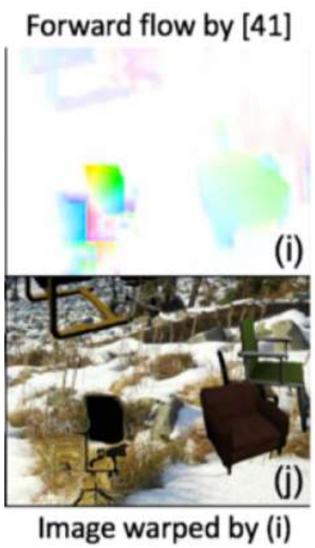


Problem – Occlusion affects Performance

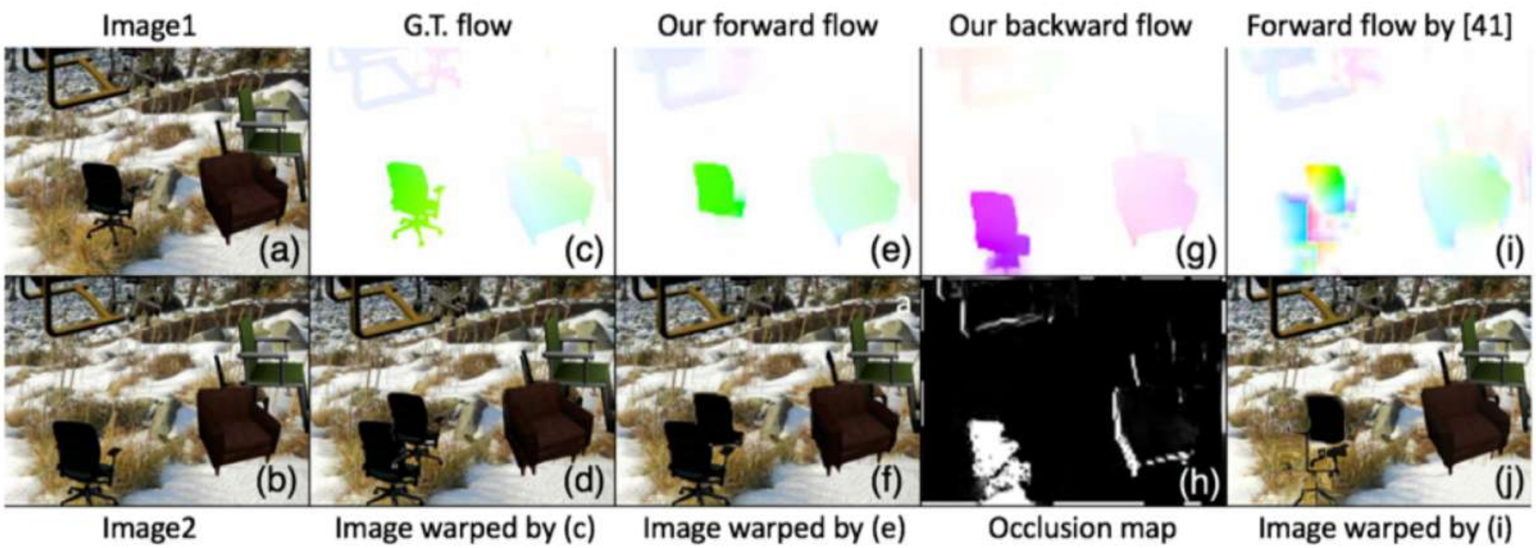


Problem – Occlusion affects Performance

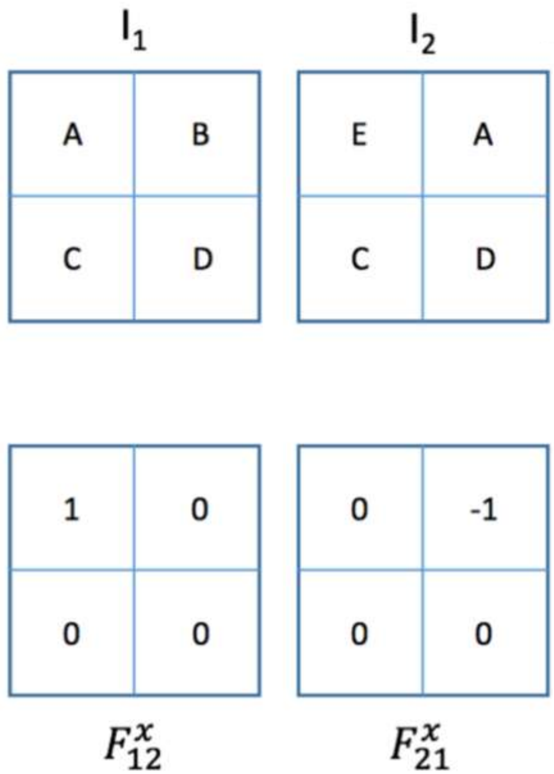




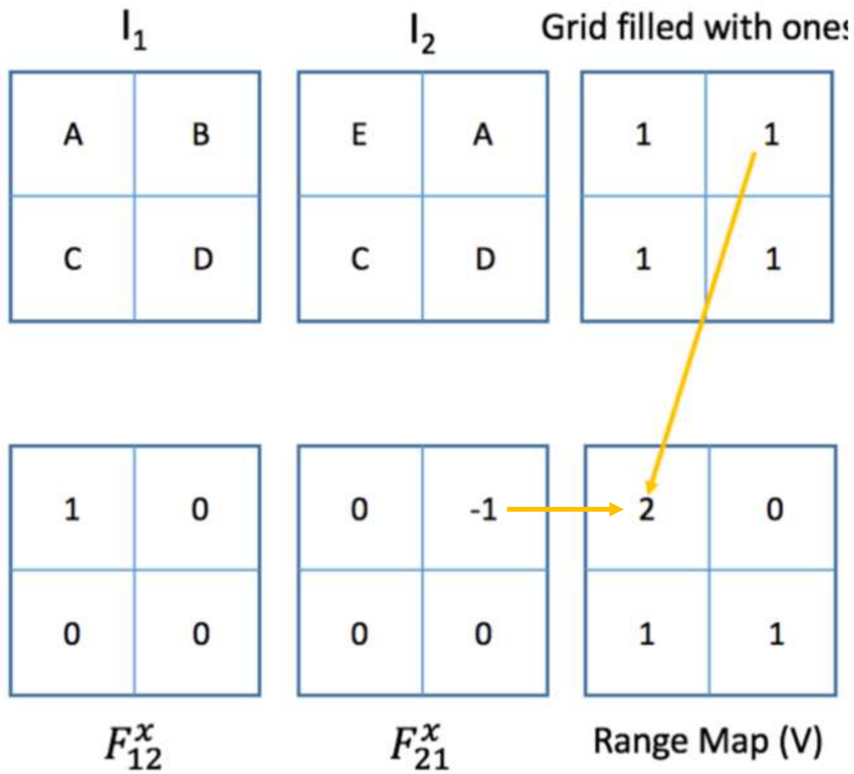
Our Solution – Model Occlusion Explicitly



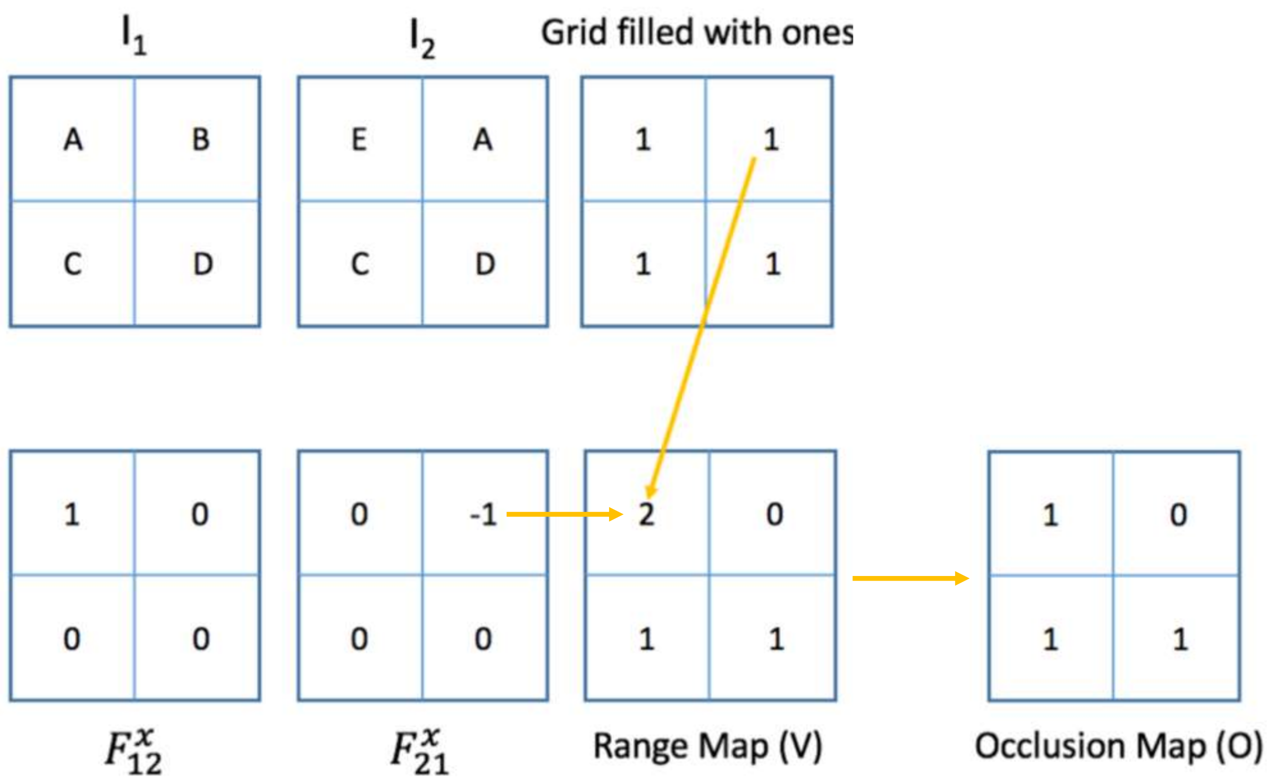
Occlusion Map



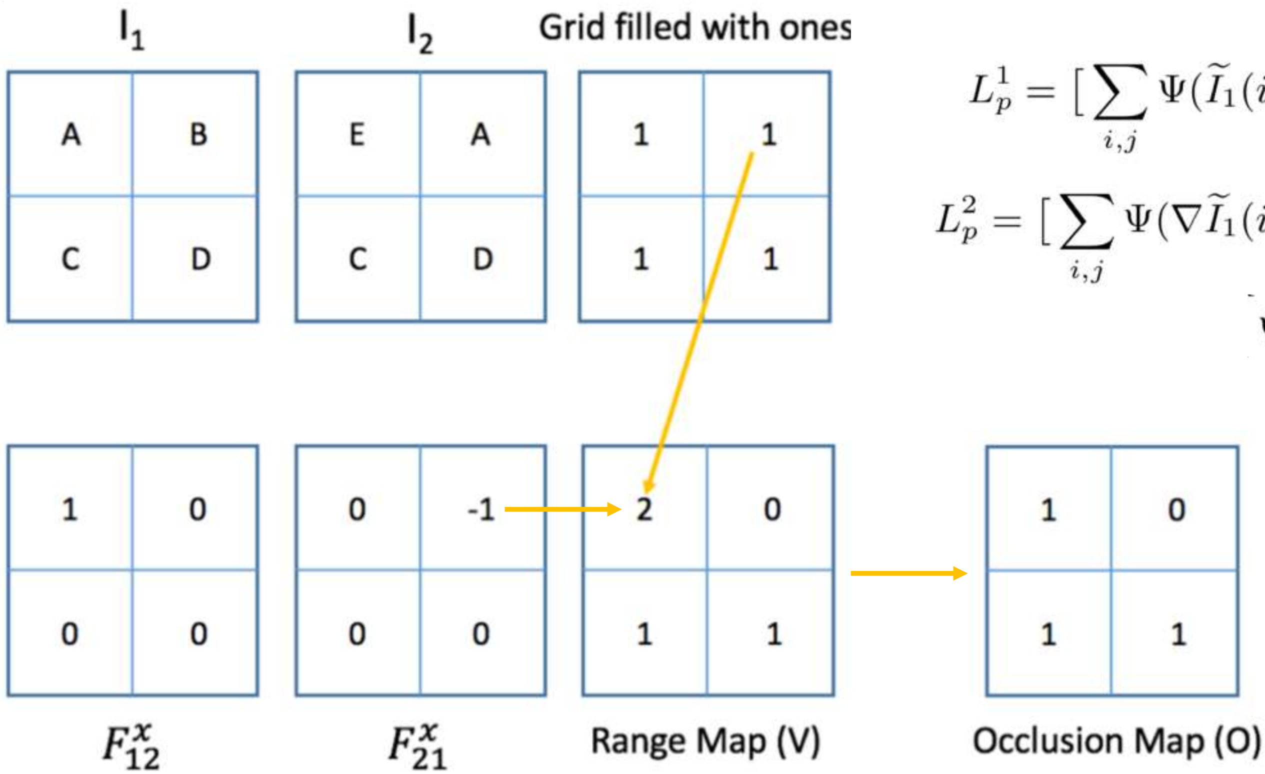
Occlusion Map



Occlusion Map



Occlusion-Aware Photometric Loss

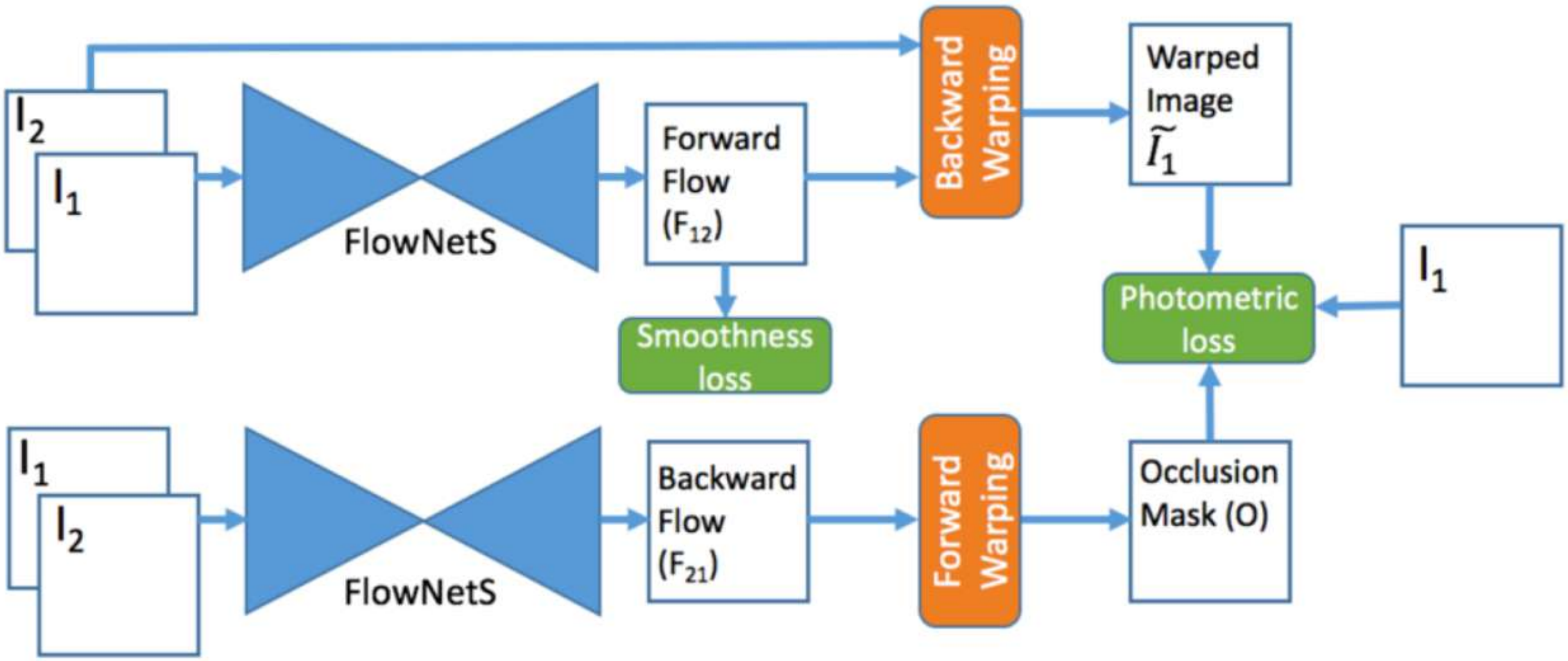


$$L_p^1 = \Big[\sum_{i,j} \Psi(\widetilde{I}_1(i,j) - I_1(i,j)) \cdot O(i,j)\Big] / \Big[\sum_{i,j} O(i,j)\Big]$$

$$L_p^2 = \Big[\sum_{i,j} \Psi(\nabla\widetilde{I}_1(i,j) - \nabla I_1(i,j)) \cdot O(i,j)\Big] / \Big[\sum_{i,j} O(i,j)\Big]$$

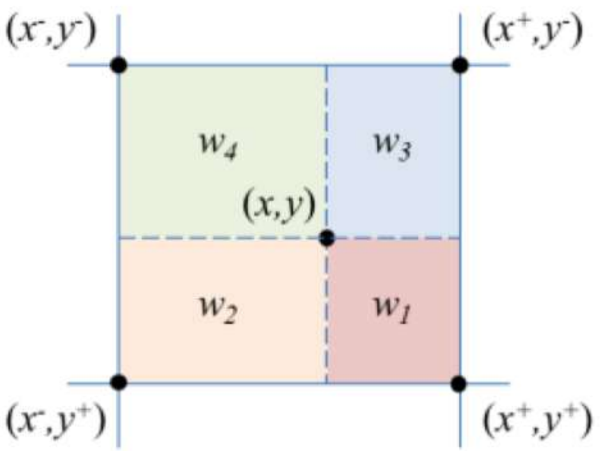
$$\Psi(s) = \sqrt{s^2 + 0.001^2}$$

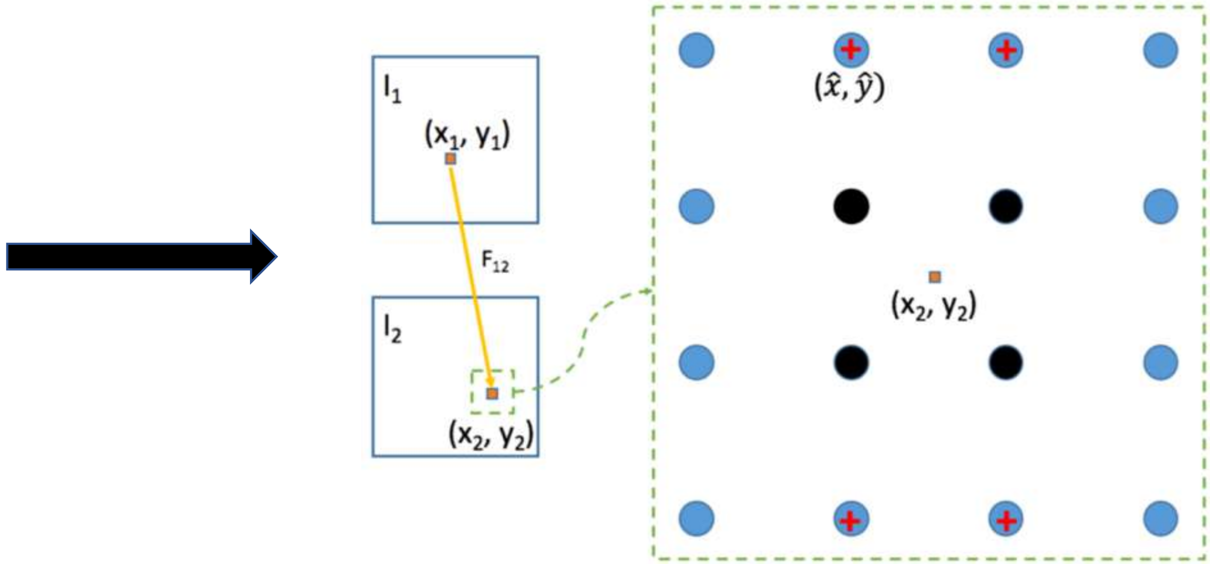
Model Structure



Improving Backpropagation - Enlarged Search

- The warped pixel only depends on its four nearest neighbors, so if the target position is far away from the proposed position, the network will not get meaningful gradient signals during backpropagation.
- We search for the best bilinear interpolation in different scales.



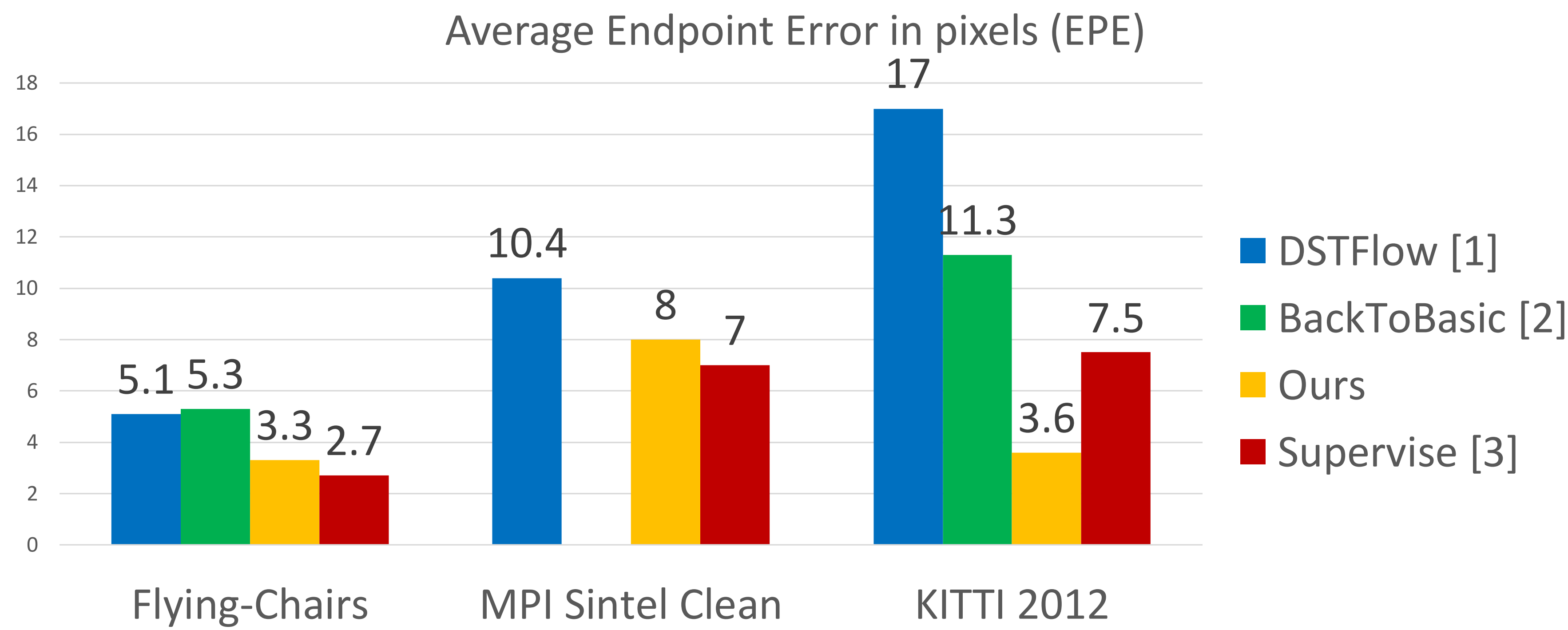


Experiments: Benchmark Datasets

- Dataset Statistics

Dataset	#Unsupervised Train Paris	#Validation Paris	Train Has GT Flow
Flying Chairs	22232	640	
MPI Sintel	908	133	
KITTI 2012	13372	194	✗

Quantitative Results – The Lower the Better

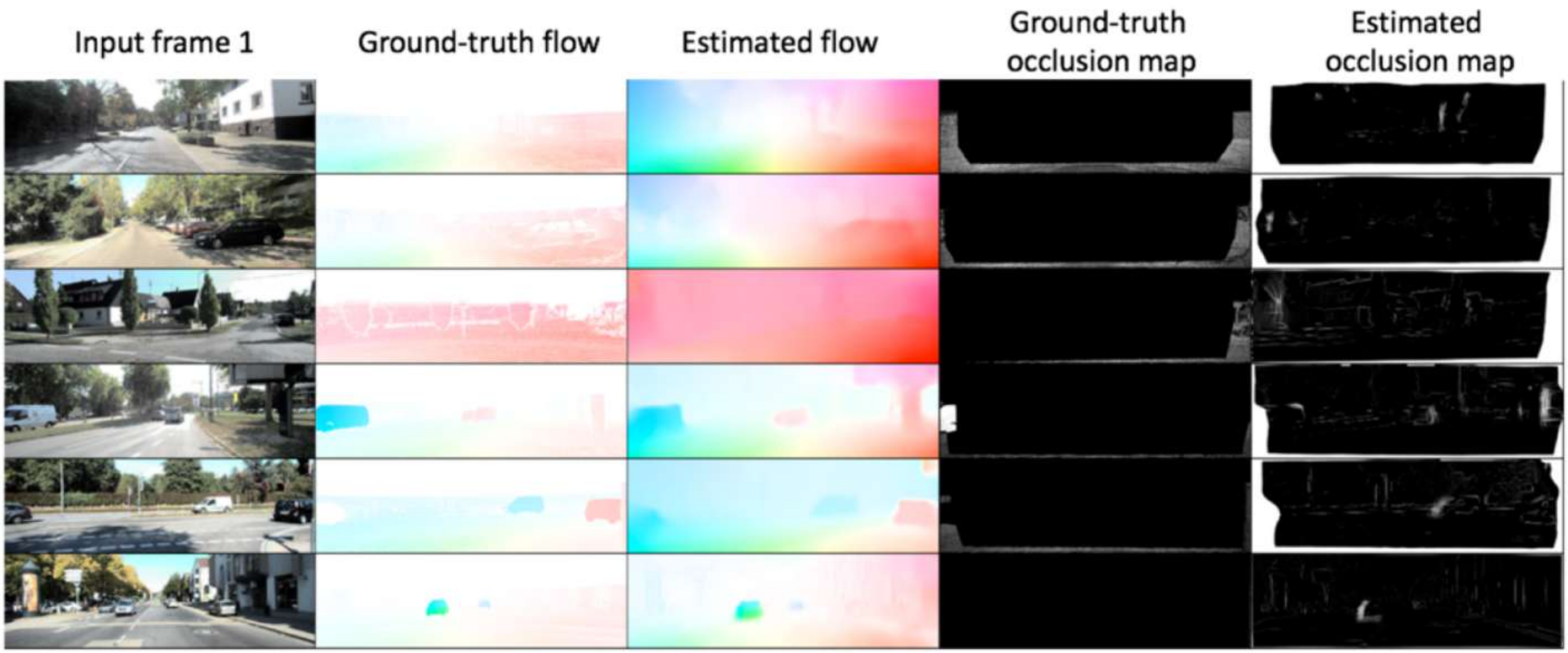


[1] Zhe et al, Unsupervised Deep Learning for Optical Flow Estimation, AAAI 2017

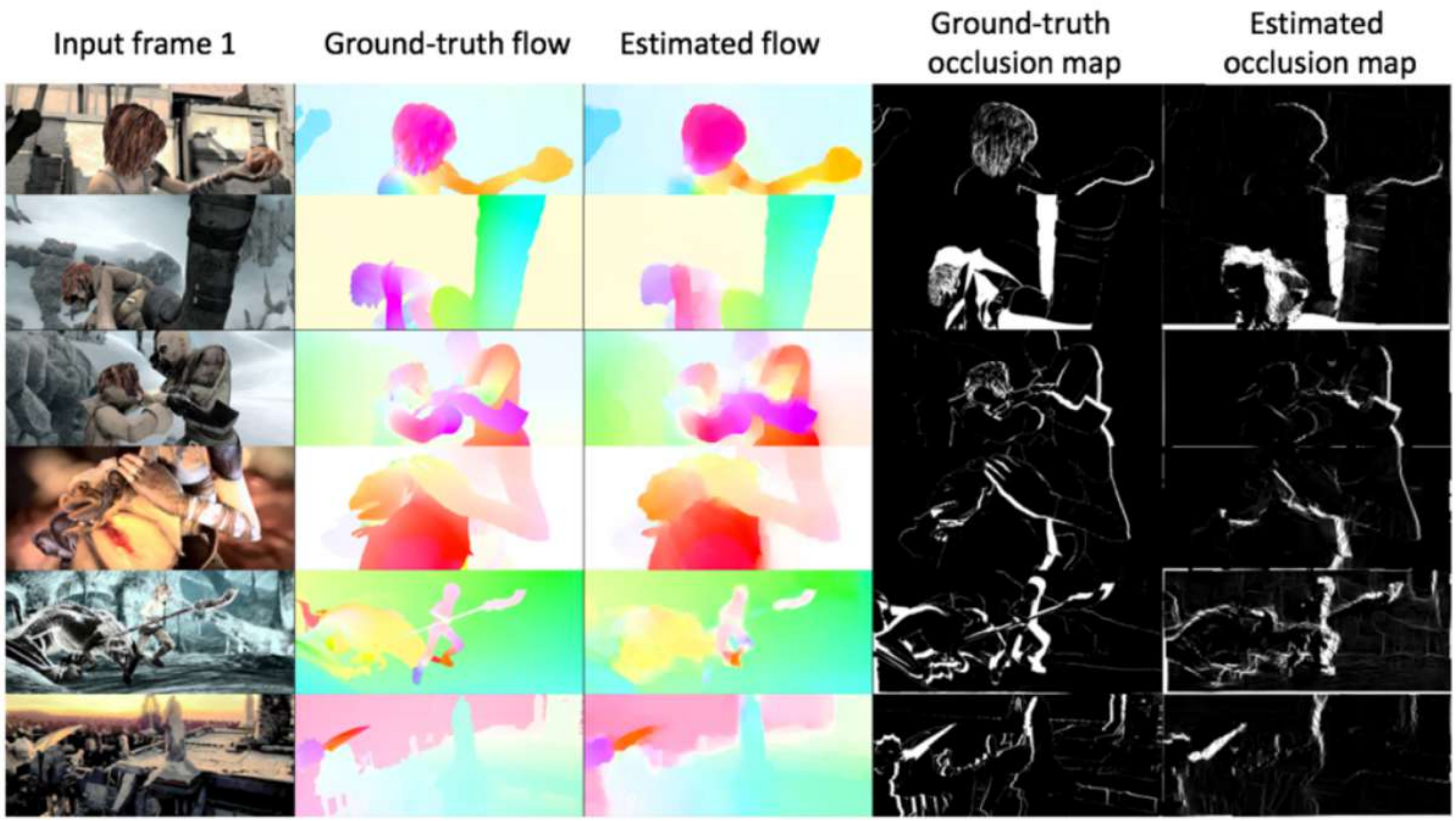
[2] Jason et al, Backtobasics: Unsupervised learning of optical flow via brightness constancy and motion, ECCV 2016W

[3] Flownet: Learning optical flow with convolutional networks, ICCV 2015

Cherry Pick Examples – KITTI



Cherry Pick Examples – MPI Sintel

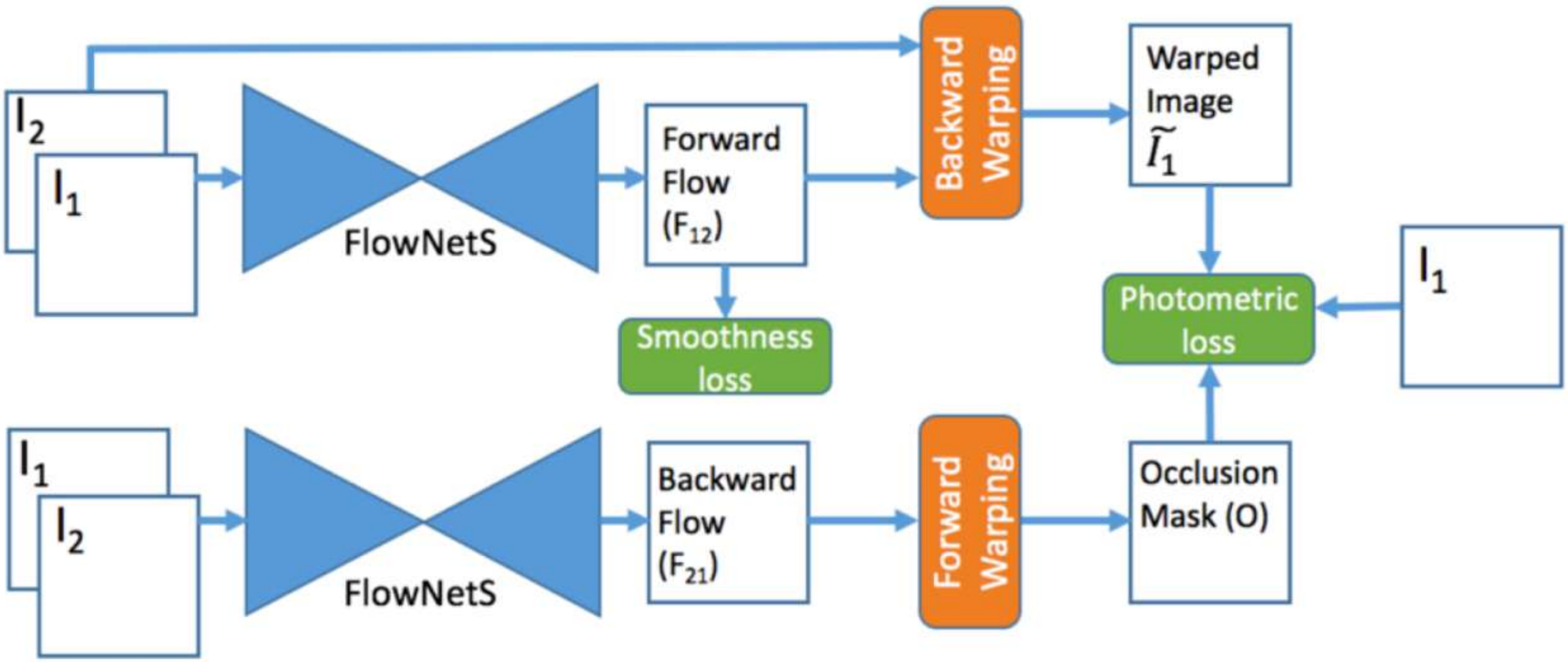


Ablation Study

occlusion handling	enlarged search	modified FlowNet	contrast enhancement	Chairs test	Sintel Clean train	Sintel Final train
				5.11	6.93	7.82
✓				4.51	6.80	7.32
✓	✓			4.27	6.49	7.11
✓	✓	✓		4.14	6.38	7.08
		✓		4.62	6.60	7.33
		✓	✓	4.04	6.09	7.04
✓		✓	✓	3.76	5.70	6.54
✓	✓	✓	✓	**3.30**	**5.23**	**6.34**

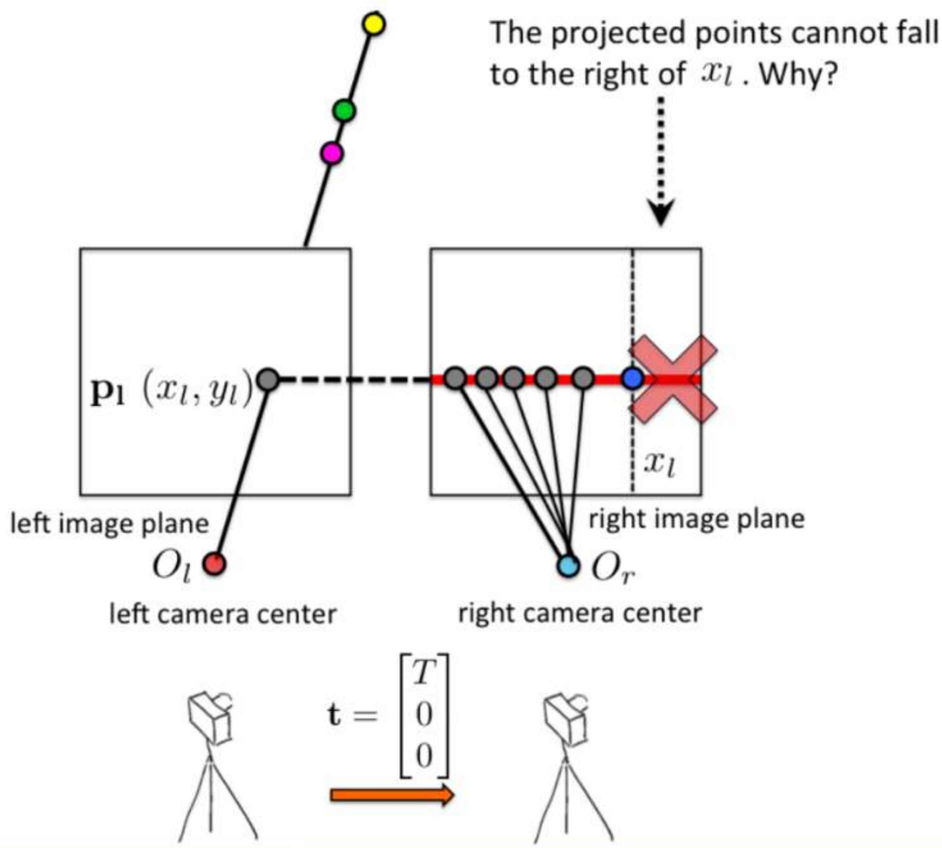
Extension: Occlusion Aware Unsupervised Learning of Stereo Depth (Disparity)

- I_1 and I_2 are left and right stereo image pairs.
- For calibrated cameras, the output disparity is only x channel with ReLU.

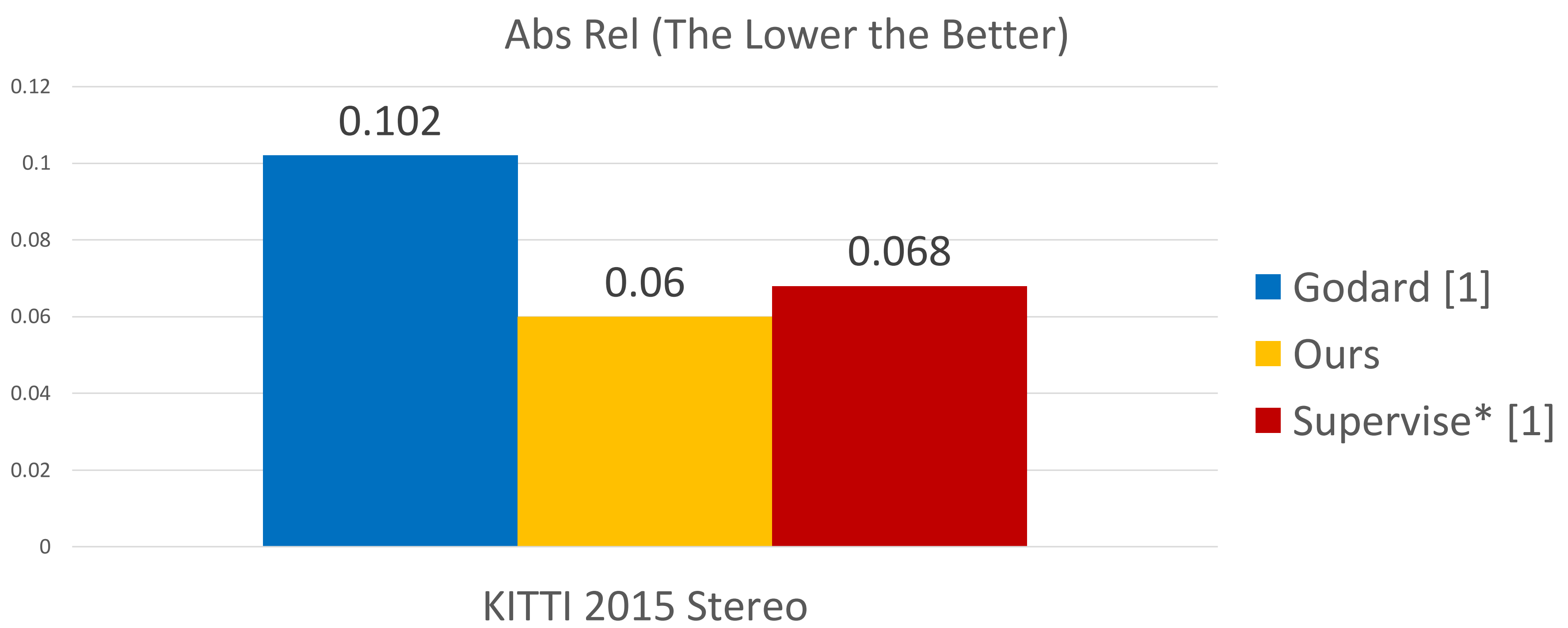


Stereo: Parallel Calibrated Cameras

- Another observation: No point from $\mathbf{O_l p_l}$ can project to the right of x_l in the right image. **Why**?

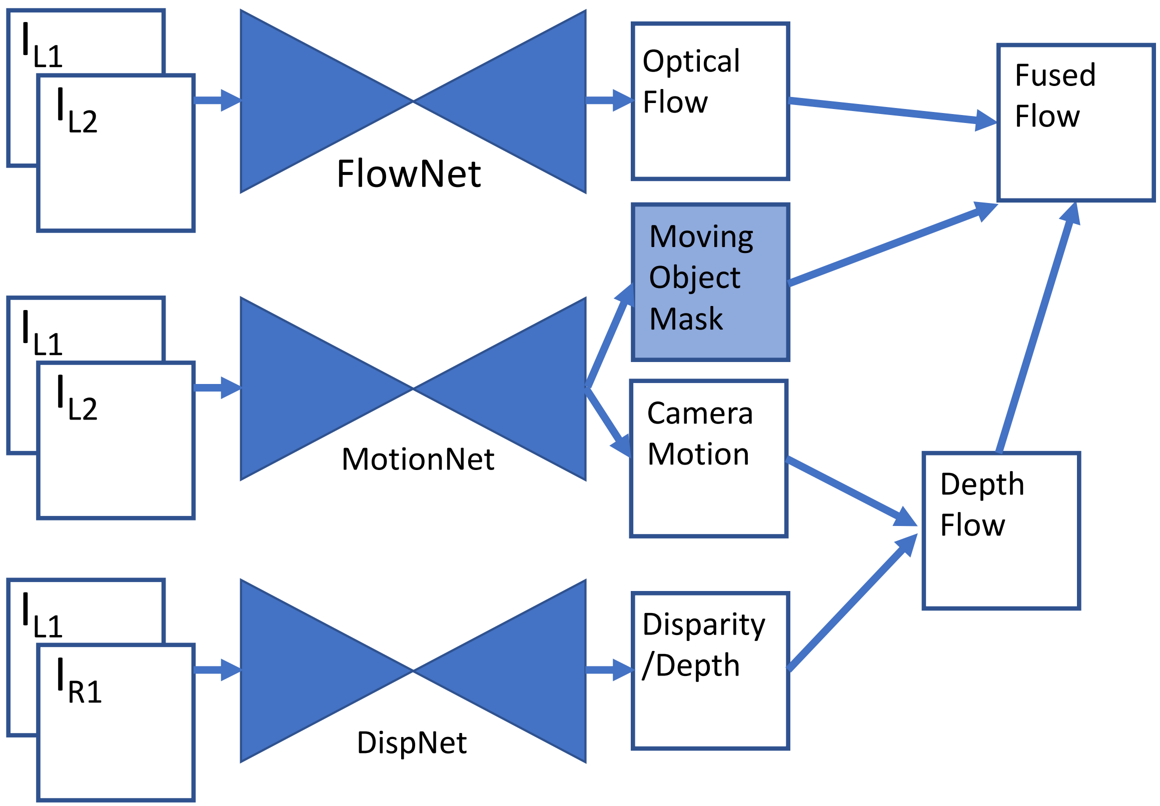


Results on KITTI Stereo Disparity Estimation



[1] Godard et al, Unsupervised monocular depth estimation with left-right consistency, CVPR 2017

Extension 2: Unsupervised Learning of Scene Flow (Flow + Disparity)



Cherry-Pick Examples of Moving Objects Mask