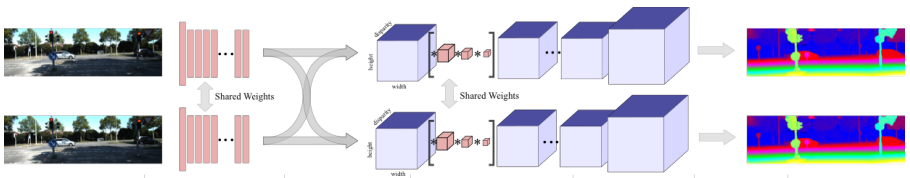
End-to-End Learning of Geometry and Context for Deep Stereo Regression

Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry,
Ryan Kennedy, Abraham Bachrach, Adam Bry

Skydio

ICCV17
International Conference on Computer Vision 2017



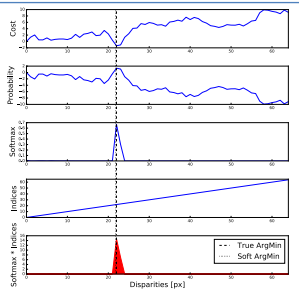
Geometry and Context Network

- End-to-end learning from rectified stereo pair to estimate sub-pixel disparity
- Form a differentiable cost volume which leverages the problem's geometry
- Learn context with multi-scale 3-D convolutions across cost volume
- Differentiable soft ArgMax to evaluate disparity curve

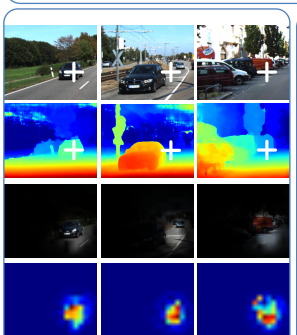
Results & Conclusions

- Pretrain on synthetic data [SceneFlow]
- Regression obtains better results than previous methods which classify disparities
- Learning end-to-end outperforms learned unary features + semi-global matching [MC-CNN, Content-CNN]
- Using the problem's geometry to learn cost volume explicitly outperforms other end-to-end approaches [DispNetC]
- Our model is relatively fast

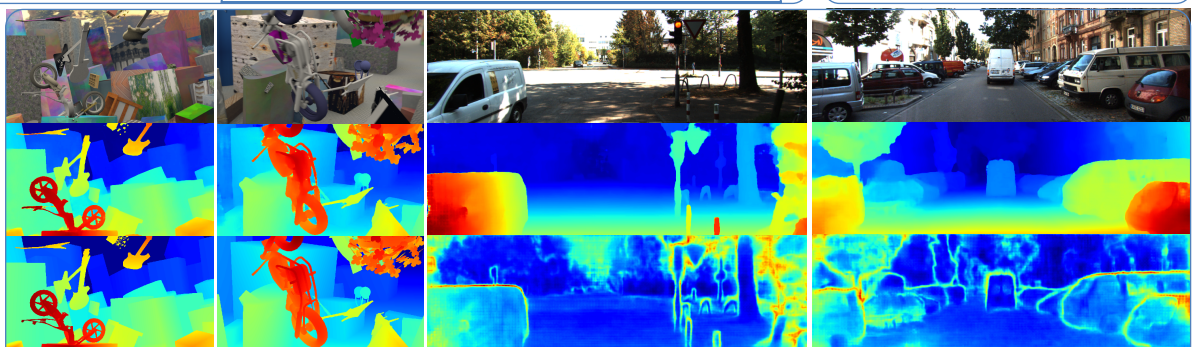
KITTI 2015 Benchmark	Error All Pixels	Runtime
GC-Net (this work)	2.87 %	0.9 s
Displets v2, Guney & Geiger	3.43 %	265 s
MC-CNN, Zbontar & LeCun	3.89 %	67 s
DispNetC, Mayer et al.	4.34 %	0.06 s
Content-CNN, Luo et al.	4.54 %	1.0 s
ELAS, Geiger et al.	9.72 %	0.3 s



Soft ArgMax Function



Saliency Maps



Scene Flow: left input, prediction, label

KITTI 2015 Results: left image, prediction, uncertainty (future work)