How to satisfy the Thirst for Data?

Andreas Geiger

Autonomous Vision Group
University of Tübingen / MPI for Intelligent Systems

September 9, 2018

University of Tübingen
MPI for Intelligent Systems
Autonomous Vision Group

The Deep Learning Revolution

What is the problem?

Data

Data Data Data Data

Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data
Data Data Data Data Data Data Data Data Data Data

Labels

Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels
Labels Labels Labels Labels Labels Labels Labels

Money

Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money
Money Money Money Money Money Money Money

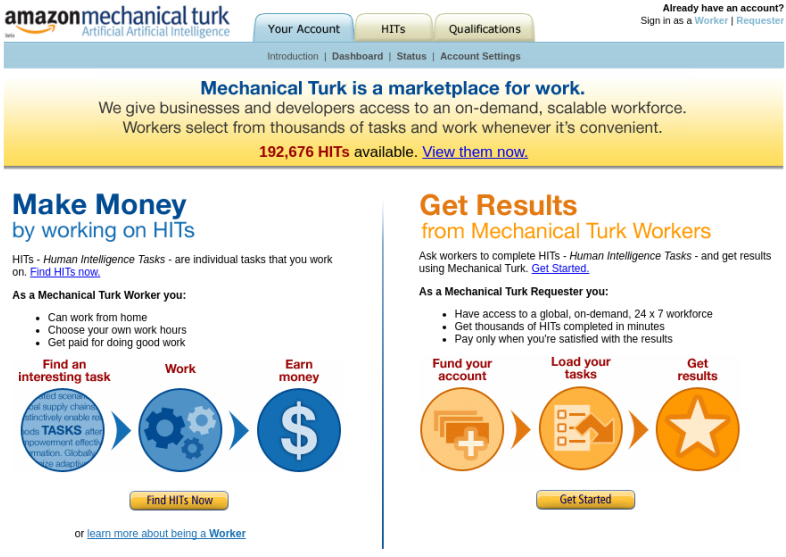
Image Classification

[Deng *et al*., CVPR 2009]

Semantic Segmentation

[Cordts *et al.*, CVPR 2016]

Amazon Mechannical Turk



Data Annotation Industry

BOSCH
Technik fürs Leben

Baidu 百度

MOBILEYE®

Mapillary

Mighty AI

DATATANG

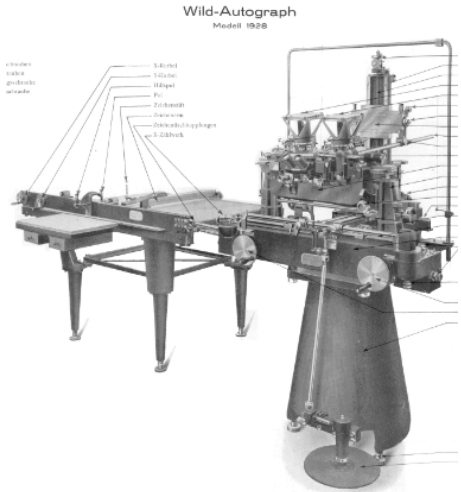
PALLASLUDENS
VISION LAB

CrowdFlower

SURFING TECH

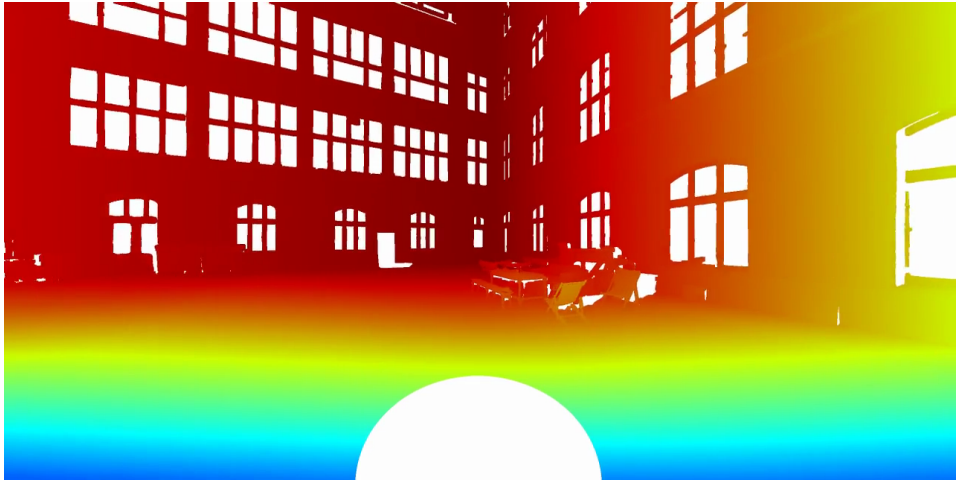
What about other tasks?

3D Reconstruction



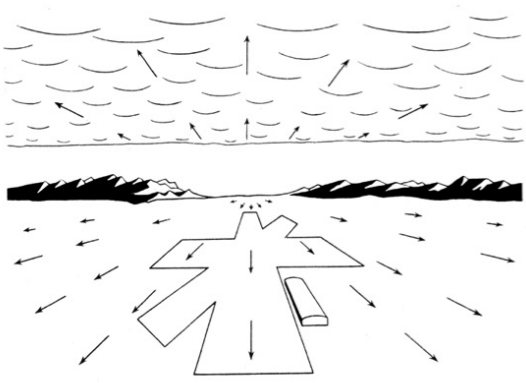
www.wild-heerbrugg.com/photogrammetry1.htm

3D Reconstruction

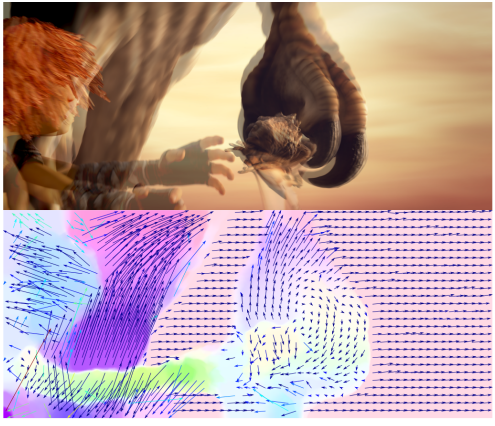


www.eth3d.net

Optical Flow



J. Gibson, 1950



MPI Sintel, 2012

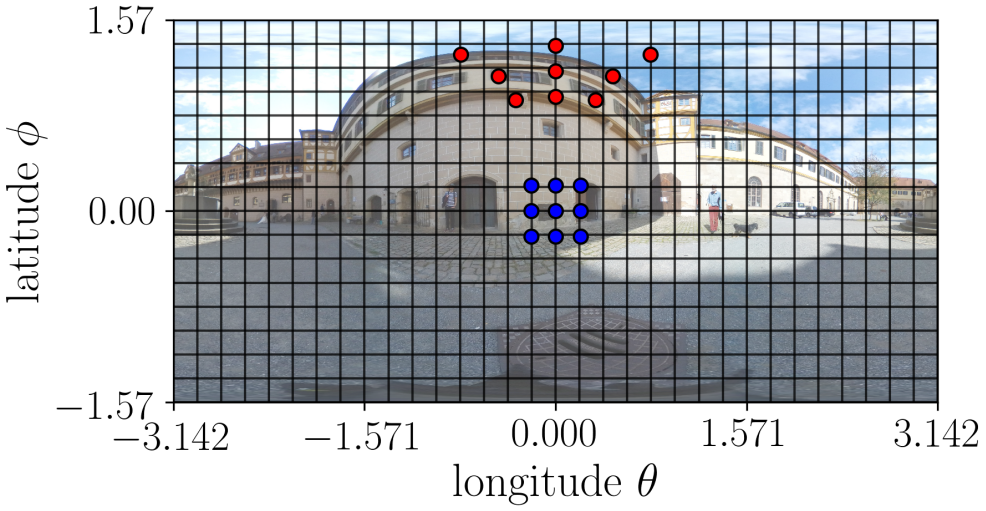
What can we do?

6 Strategies

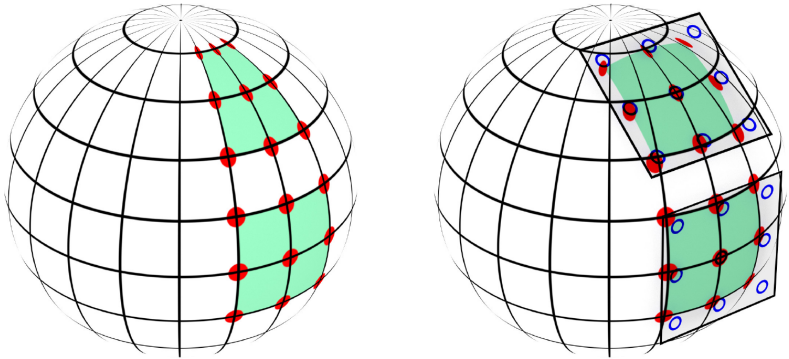
Strategy 1
Specialized Architectures

SphereNet: Learning Spherical Representations for Detection

SphereNet: Learning Spherical Representations for Detection

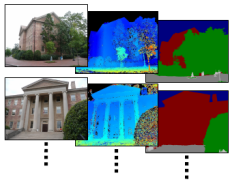


SphereNet: Learning Spherical Representations for Detection



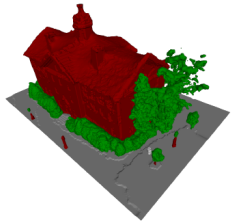
- Allows for easy transfer learning between camera models

Learning Priors for Semantic 3D Reconstruction



Input Images &
Depth & Semantics

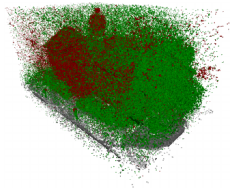
Häne *et al.* [1]
(50 iters.)



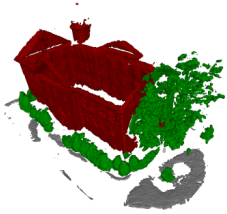
Häne *et al.* [1]
(2750 iters.)

ground [1]

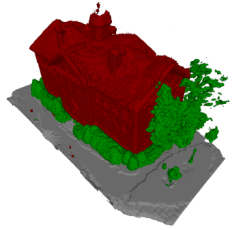
building [1]
Shape priors [1]



Input Data Cost



TV-L1 (50 iters.)

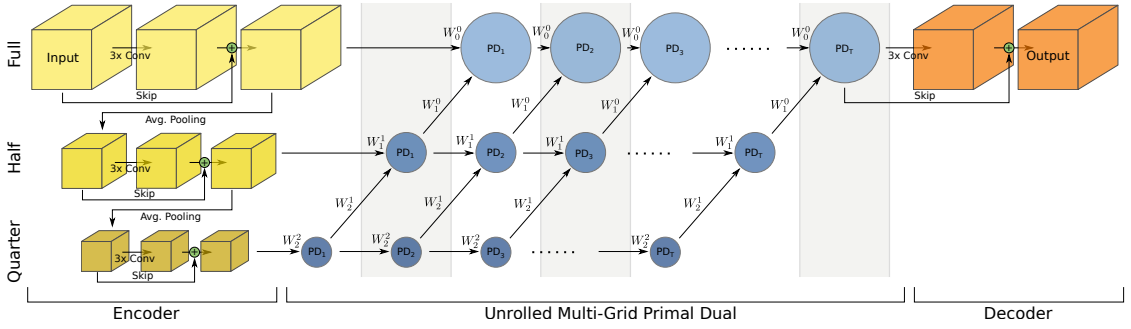


Ours (50 iters.)

ground (ours)

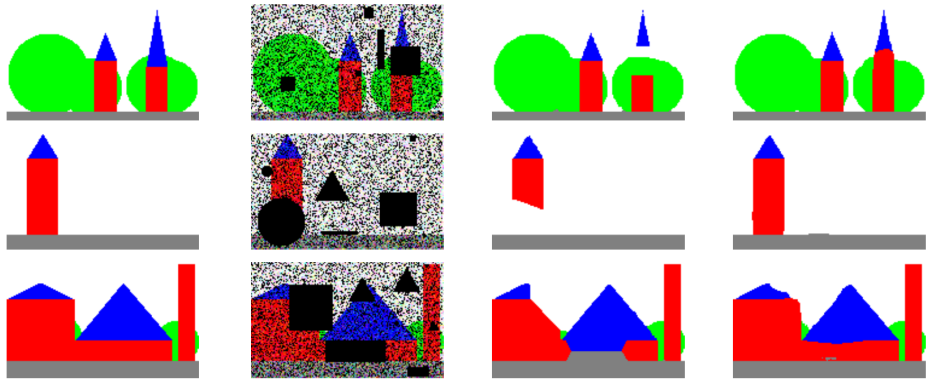
building (ours)
Our shape priors

Learning Priors for Semantic 3D Reconstruction



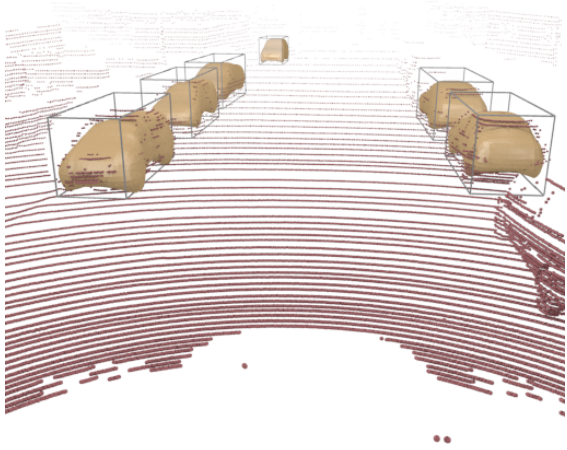
$$\underset{u}{\text{minimize}} \quad \int_{\Omega} \left(\|Wu\|_2 + fu \right) d\mathbf{x} \quad \text{subject to} \quad \forall \mathbf{x} \in \Omega : \sum_{\ell} u_{\ell}(\mathbf{x}) = 1$$

Learning Priors for Semantic 3D Reconstruction

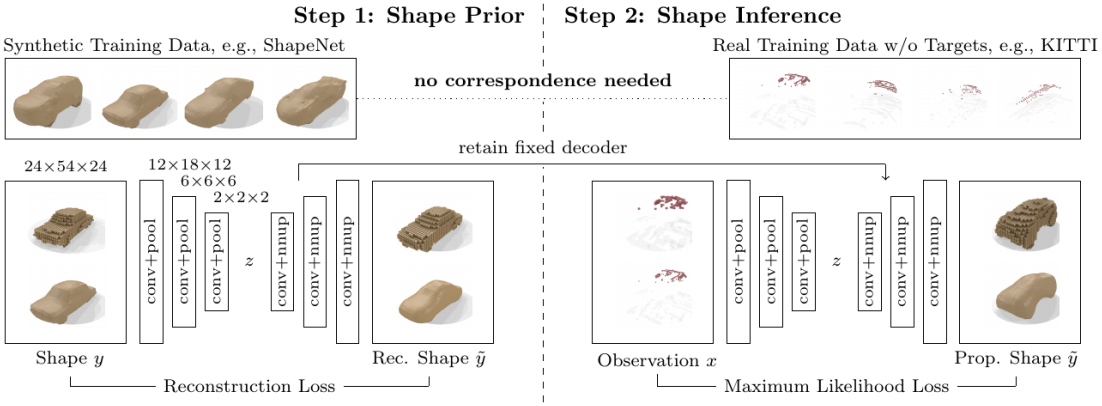


Strategy 2
Strong Priors

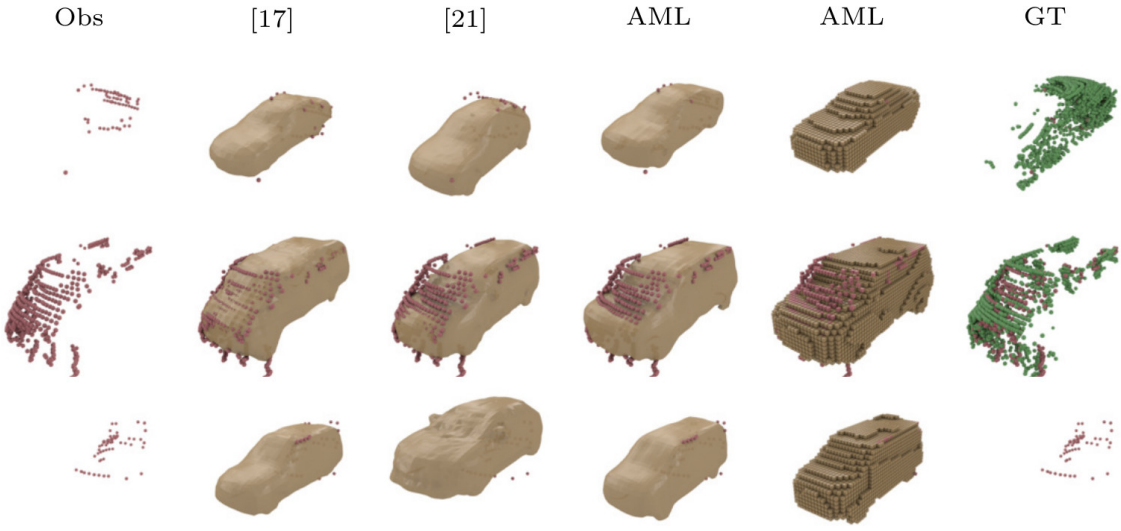
Learning 3D Shape Completion with Weak Supervision



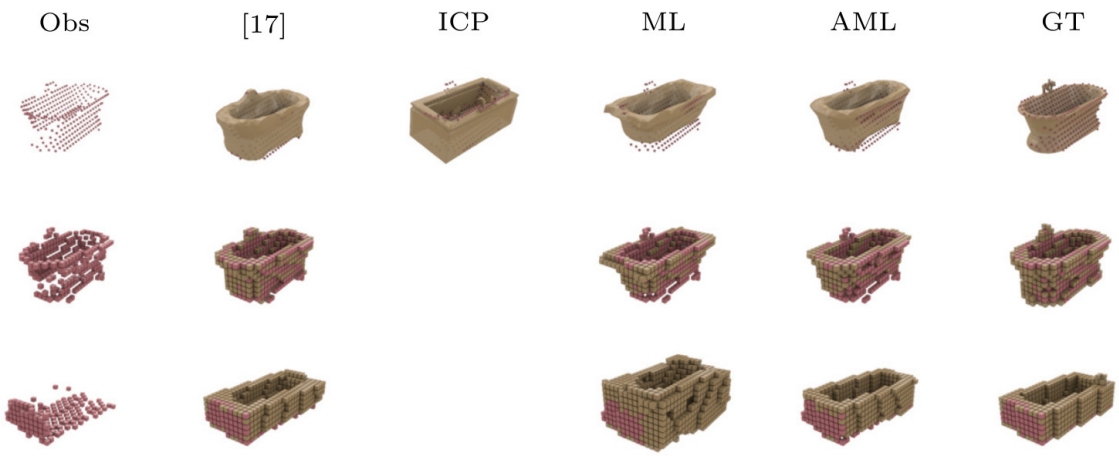
Learning 3D Shape Completion with Weak Supervision



Learning 3D Shape Completion with Weak Supervision

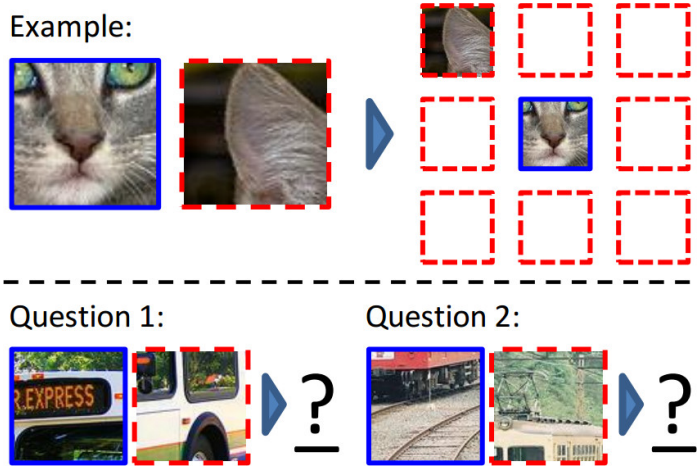


Learning 3D Shape Completion with Weak Supervision

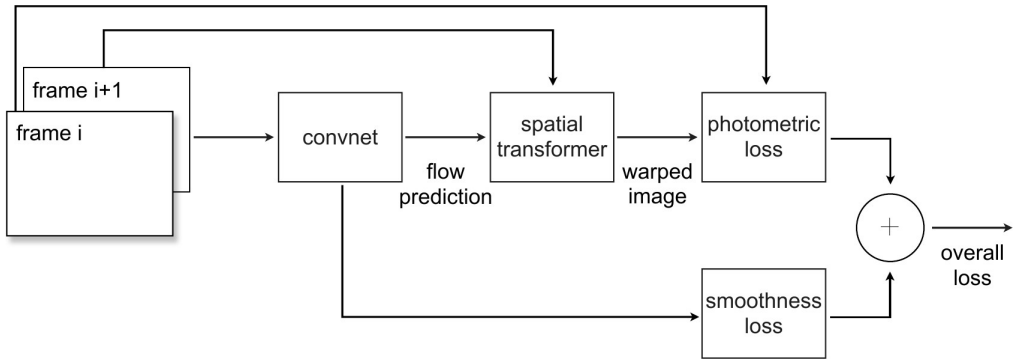


Strategy 3
Self Supervision

Unsupervised Visual Representation Learning by Context Prediction

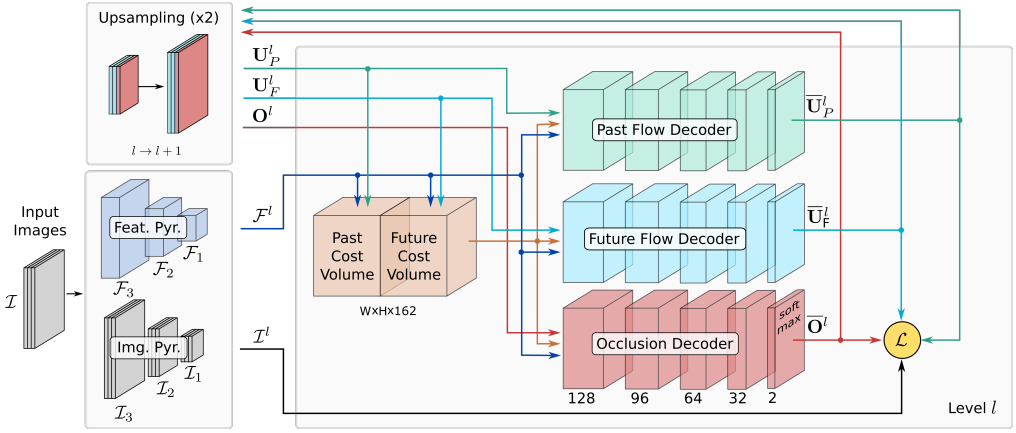


Unsupervised Learning of Multi-Frame Optical Flow with Occlusions



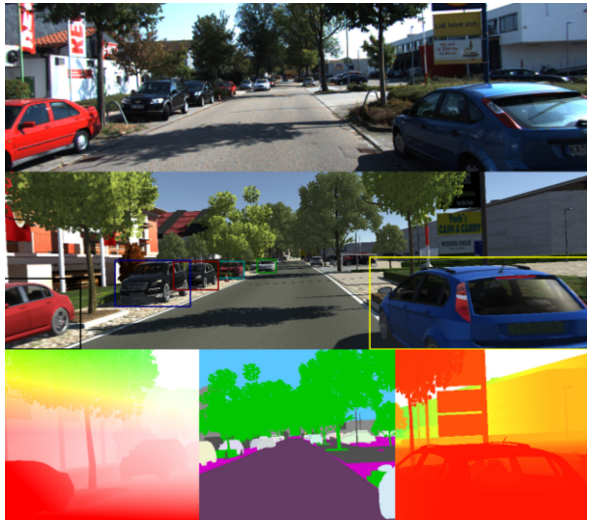
[Courtesy: Jason J. Yu]

Unsupervised Learning of Multi-Frame Optical Flow with Occlusions



Strategy 4
Simulation

Virtual Worlds as Proxy for Multi-Object Tracking Analysis



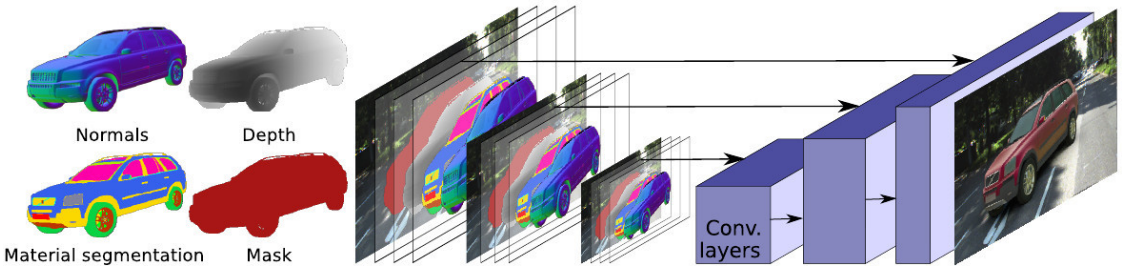
Augmented Reality Meets Deep Learning

Real scene (KITTI)

Synthetic scene (Virtual KITTI)

Real scene augmented with synthetic cars (Ours)

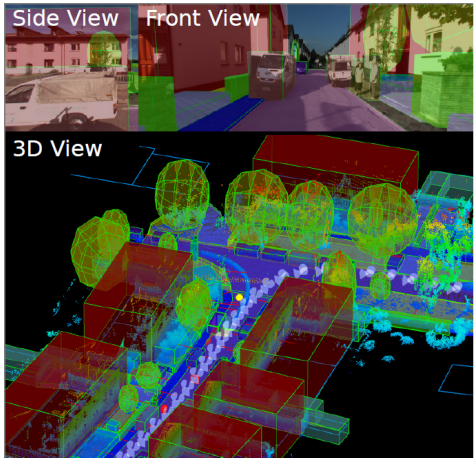
Geometric Image Synthesis

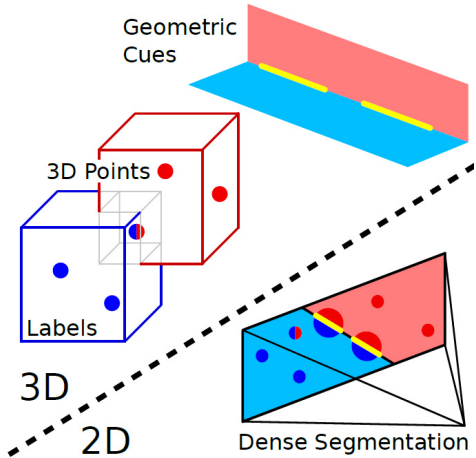


Alhaija, Mustikovela, Geiger & Rother, 2018

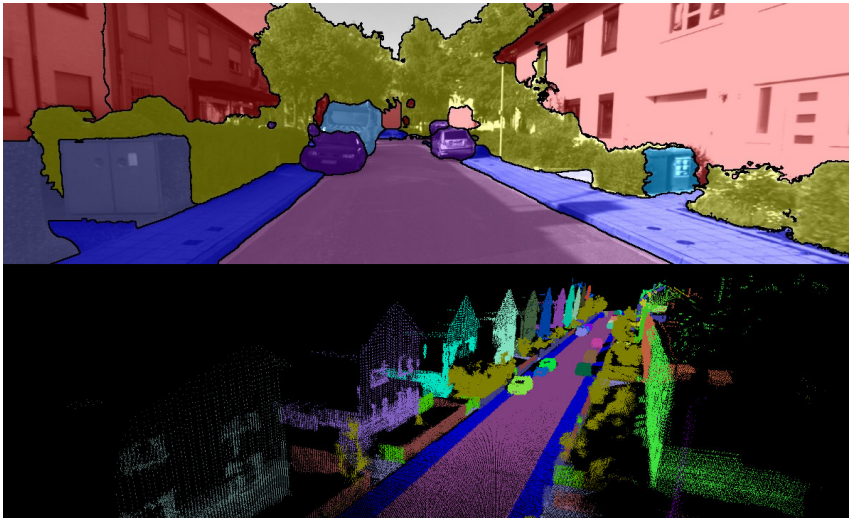
Strategy 5
Label Transfer

3D to 2D Semantic Label Transfer



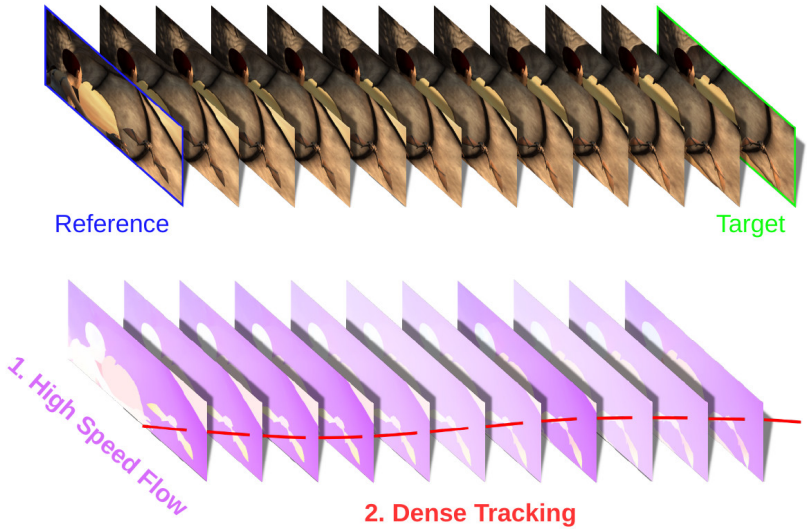


3D to 2D Semantic Label Transfer



Xie, Kiefel, Sun & Geiger, CVPR 2016

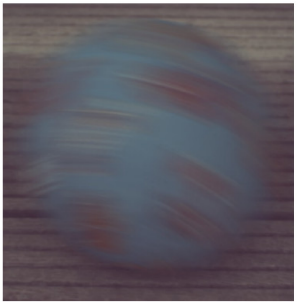
Slow Flow



Slow Flow

Slow Flow

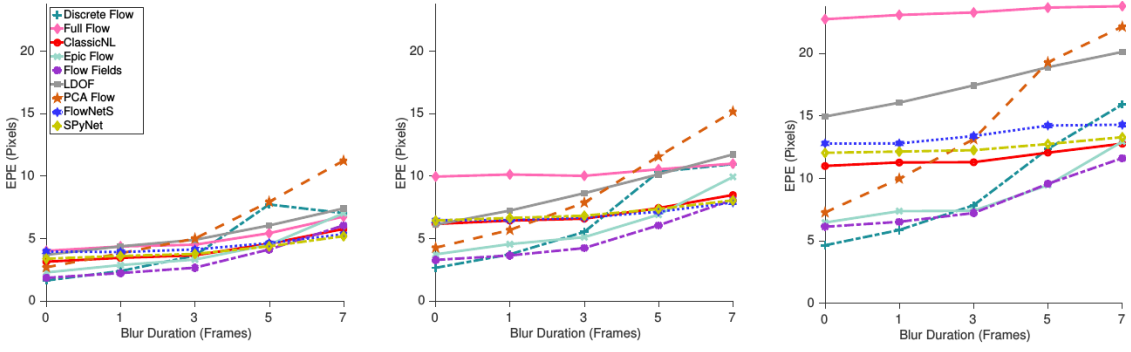
Input Image



High frame rate

Low frame rate

Slow Flow

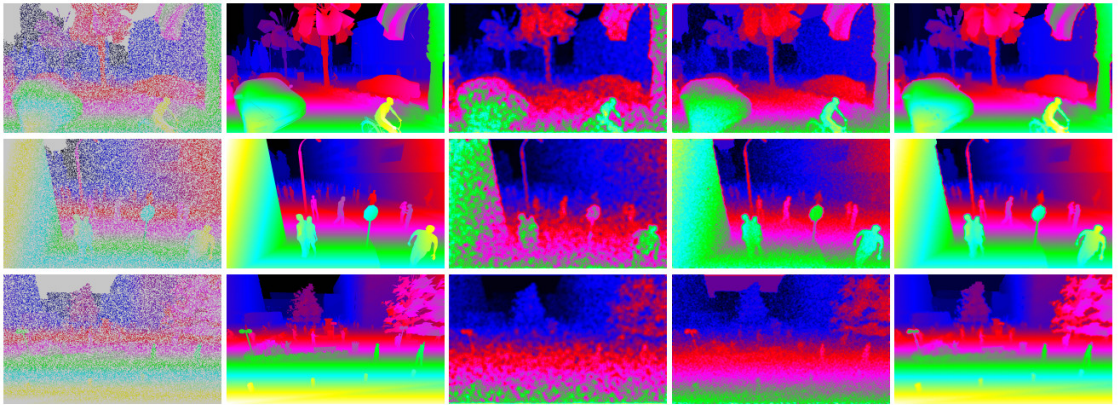


(a) 100px Flow Magnitude (b) 200px Flow Magnitude (c) 300px Flow Magnitude

Figure 4: State-of-the-art comparison on the generated reference data wrt. motion magnitude and blur.

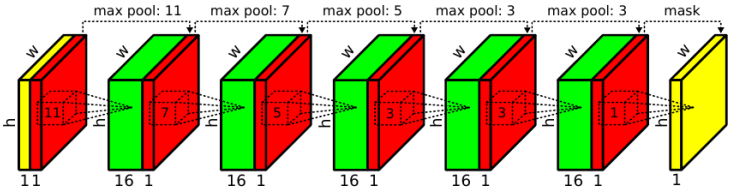
Strategy 6
Model Generalization

Sparsity Invariant CNNs

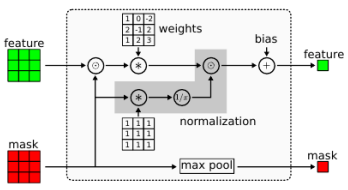


(a) Input (visually enhanced) (b) Groundtruth (c) ConvNet (d) ConvNet + mask (e) SparseConvNet (ours)

Sparsity Invariant CNNs



(a) **Network Architecture**



(b) **Sparse Convolution**

Thank you!