Learning from Synthetic Data

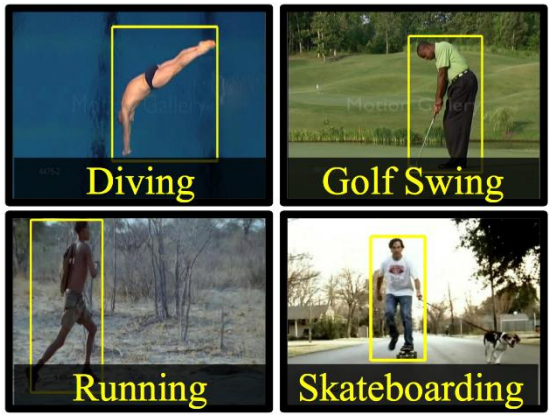
Thomas Brox

University of Freiburg, Germany

Research funded by the ERC Starting Grant VideoLearn, the German Research Foundation, and the Deutsche Telekom Stiftung

Deep learning wants data… input and output

IMAGENET



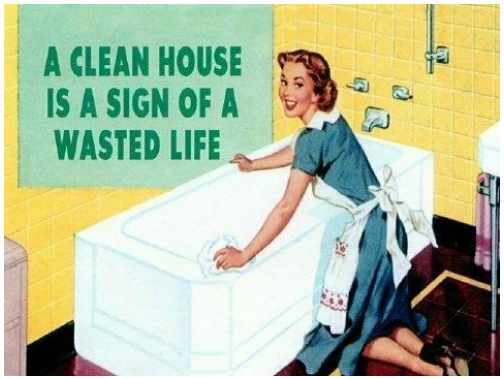
UCF 101

Typical strategy to set up a data set

1. Collect lots of images/videos
   - usually from the web

2. Manually annotate the desired output

Lock students in a cellar

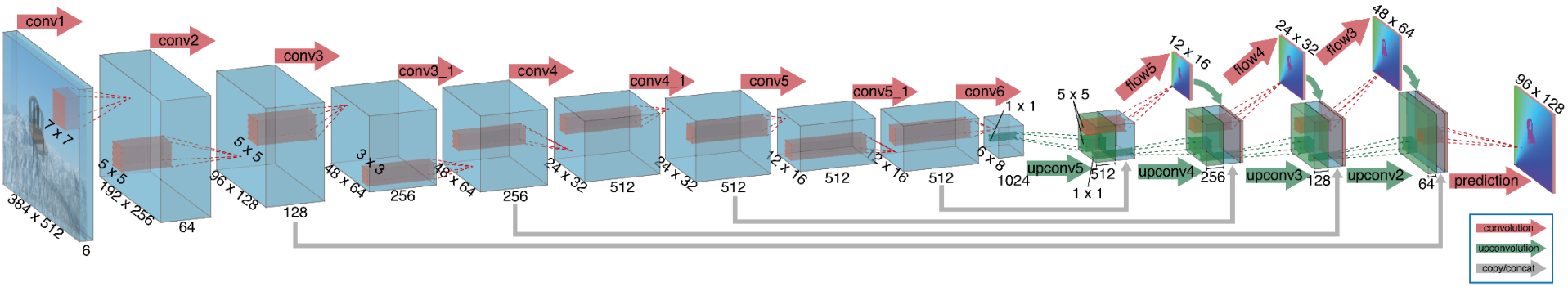


Crowd sourcing

Workers at companies

Works well for recognition

Learning tasks besides recognition: FlowNet



- Can networks learn to find correspondences?
- New learning task!

  (very different from classification, etc.)

Dosovitskiy et al.
ICCV 2015

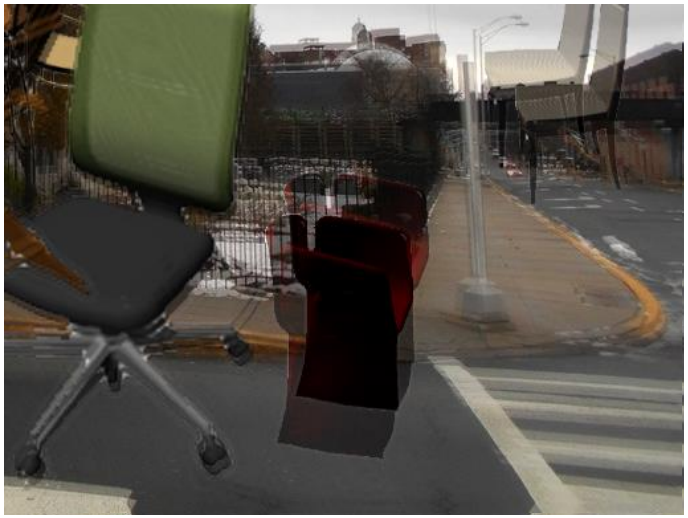
Enough data to train such a network?

- Getting ground truth optical flow for real videos is hard
- Existing datasets are small:

	Frames with ground truth
Middlebury	8
KITTI	194
Sintel	1041
Needed	>10000

Solution: large scale synthetic data

Realism is overrated: the “flying chairs” dataset



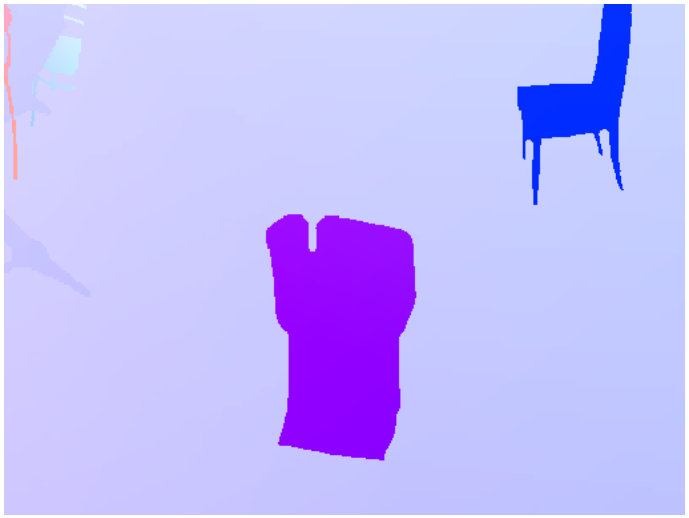


Image pair

Optical flow

What about generalization?

Input images

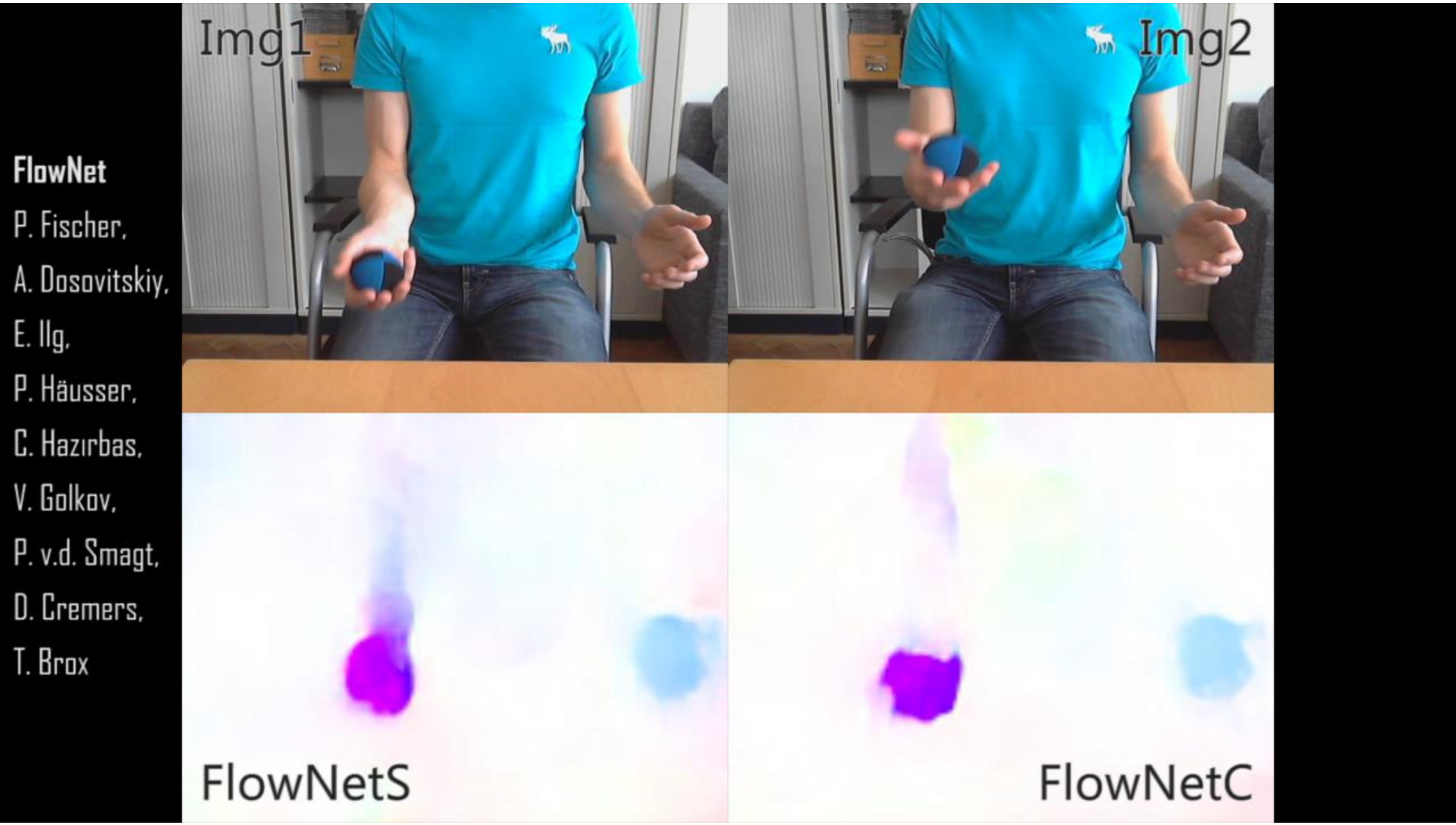
Ground truth

FlowNetSimple

FlowNetCorr

Although the network has only seen flying chairs for training, it predicts good optical flow on Sintel

Generalization to real data



Can handle large displacements

Input images

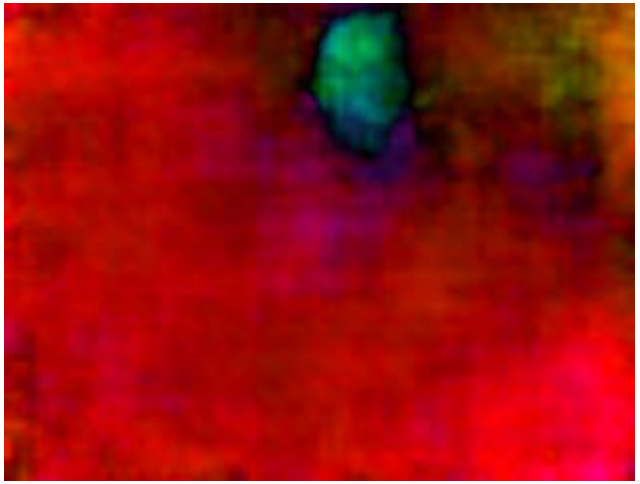
Ground truth

FlowNet

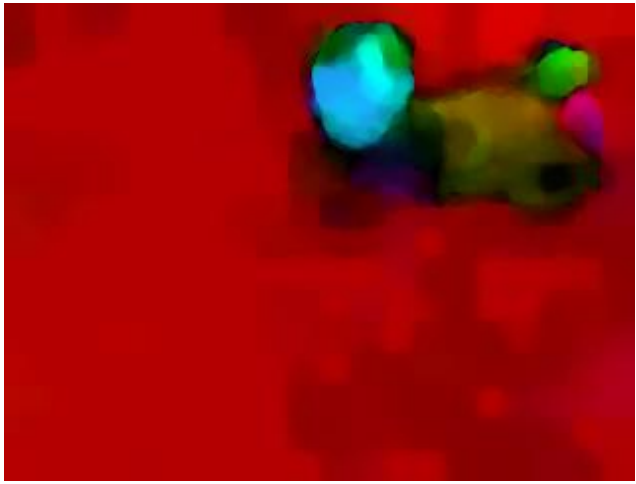
EpicFlow (Revaud et al. CVPR 2015)

Weak spot (old FlowNet): small displacements

Example from UCF 101

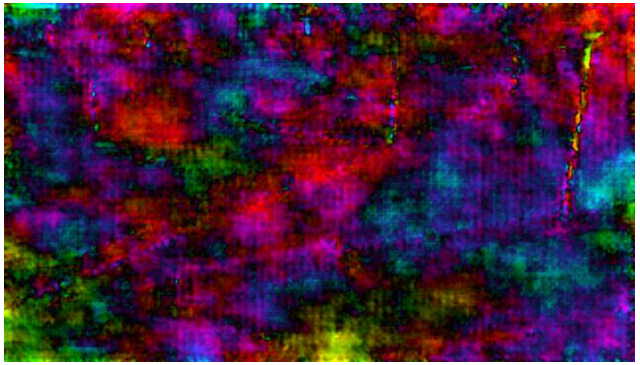


FlowNet



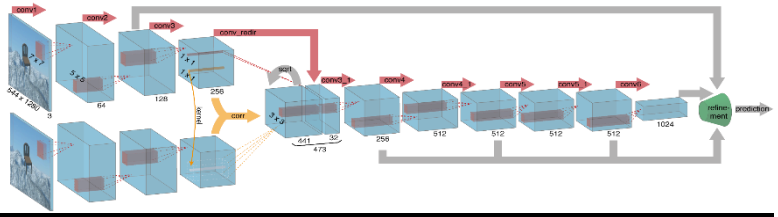
EpicFlow

Identical images

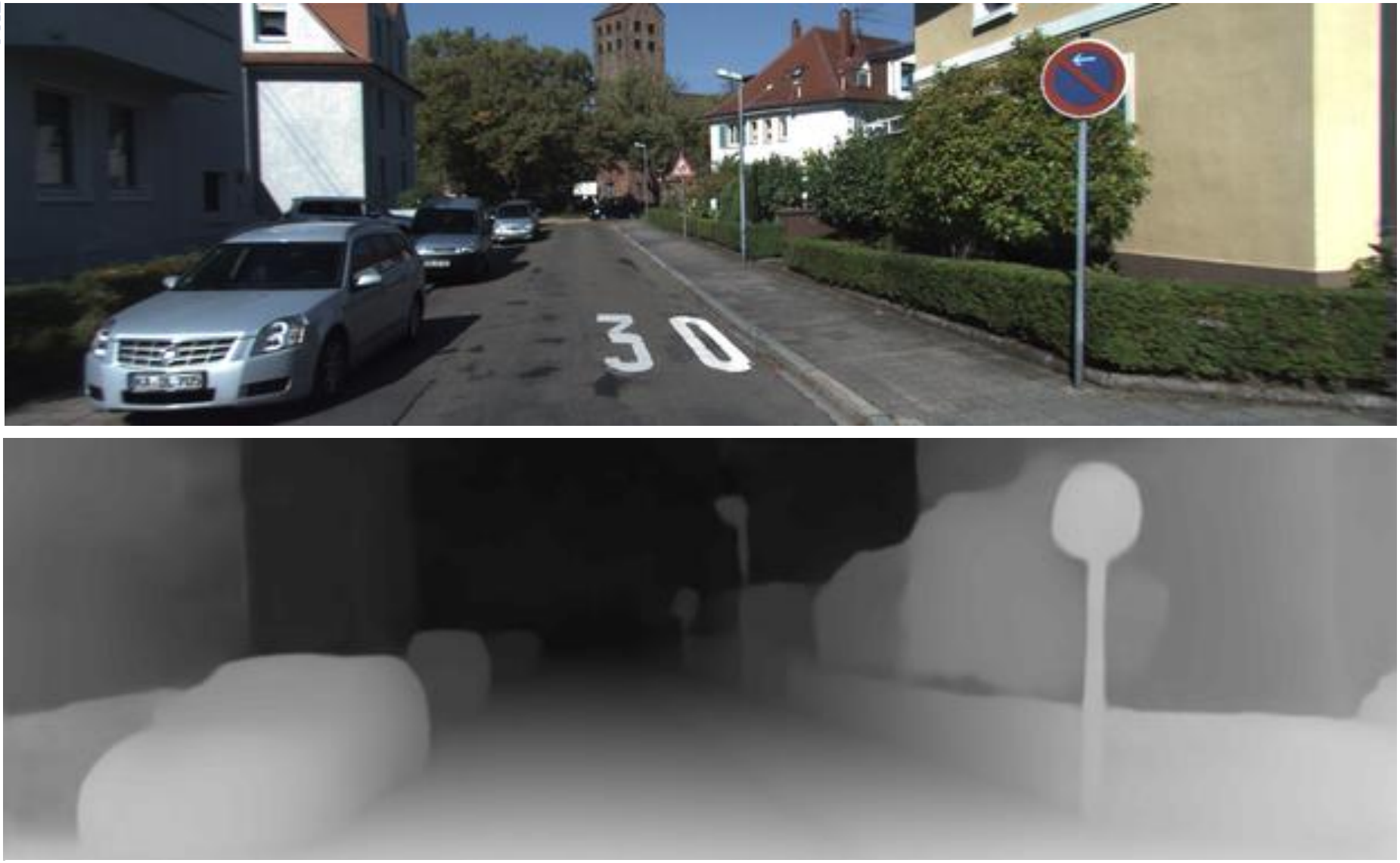


DispNet: disparity estimation in 60ms

Mayer et al.
CVPR 2016



DispNet: disparity estimation in 60ms



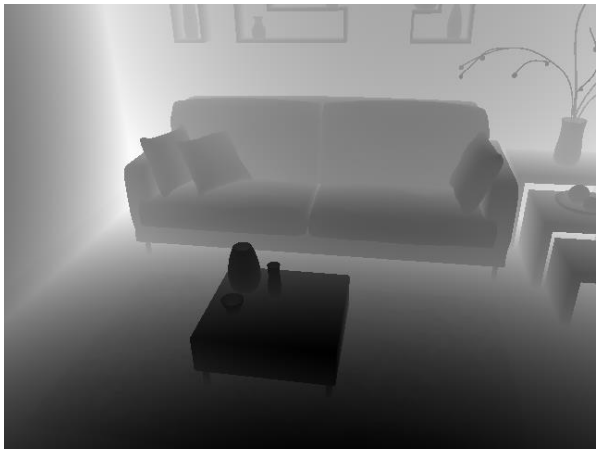
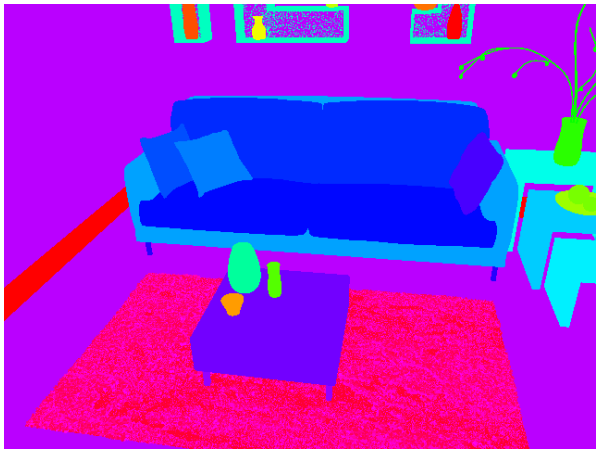
DispNet result on KITTI 2015

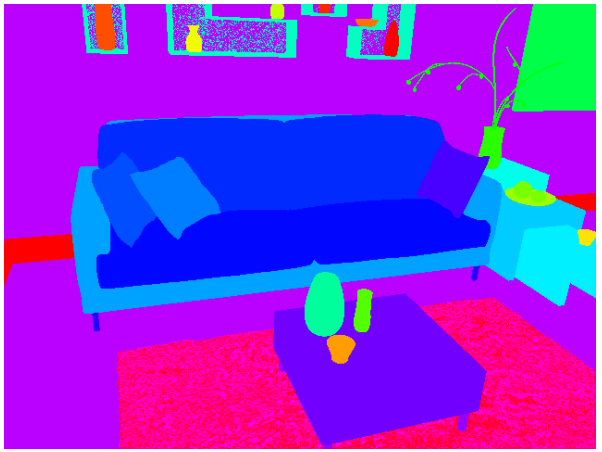
Synthetic 3D datasets

Mayer et al. CVPR 2016

Driving, Monkaa, FlyingThings3D datasets publicly available

Powerful concept: render input and any output

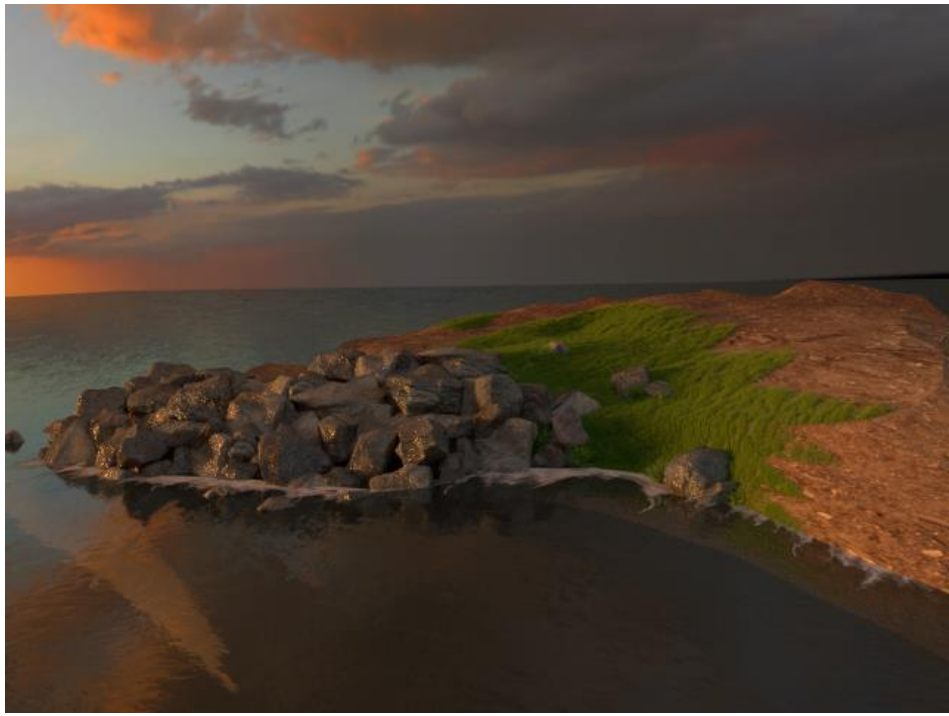



Rendered input

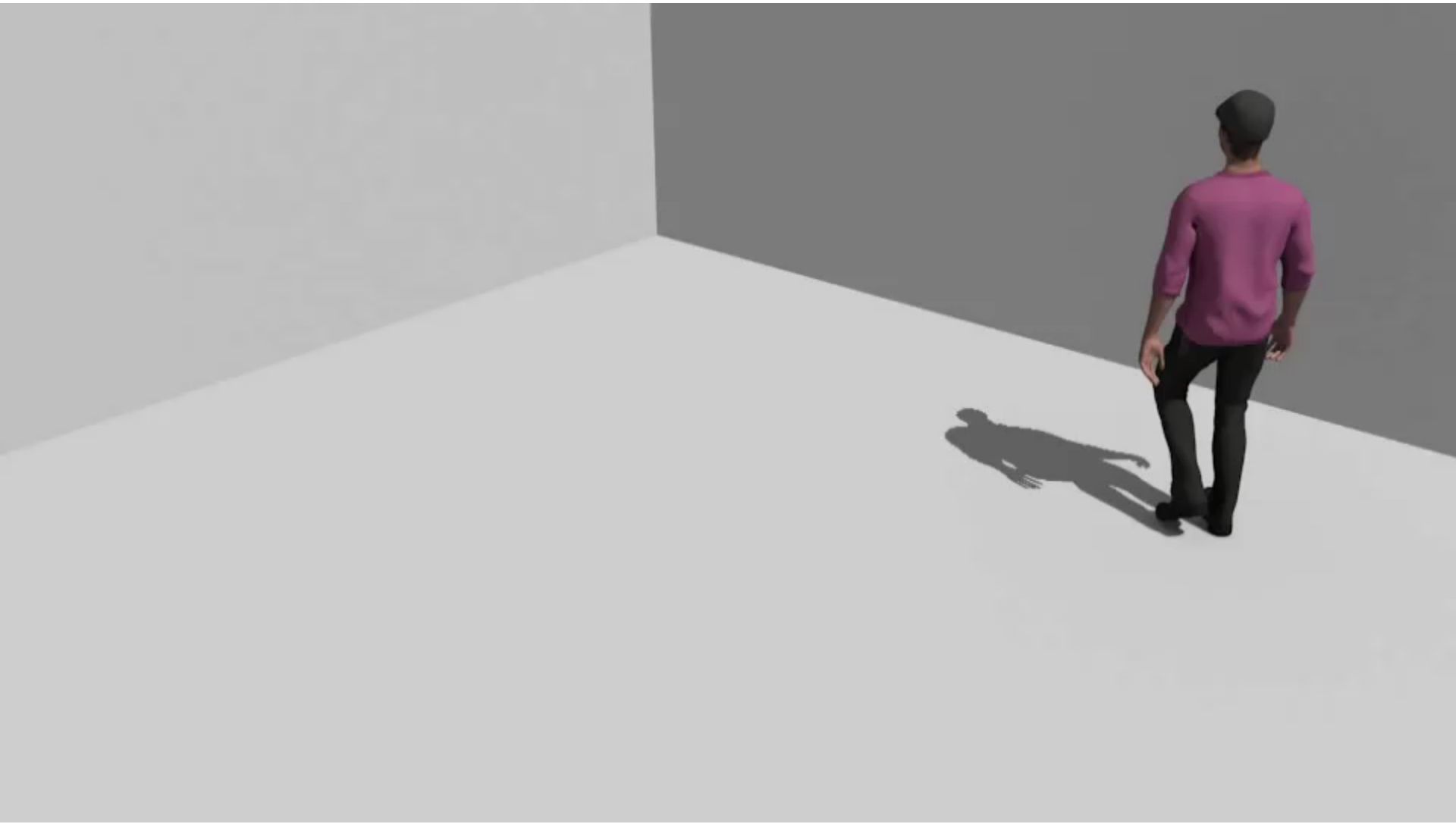
Rendered depth image

Rendered segmentation

Indoor and outdoor scenes possible



Adding animated persons

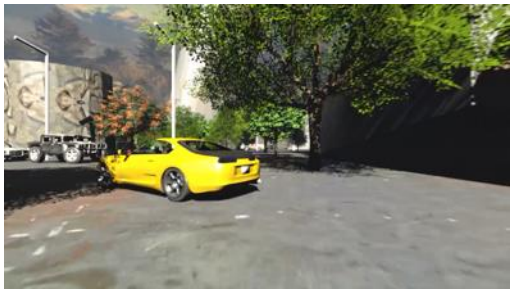


Modeling plants swinging in the wind

Existing scenes vs. new scenes

- Open source movies (Sintel, Monkaa), 3D Games
  - Effort: hacking the rendering engine to produce ground truth
  - Limited to the scenes that exist

- Manually created scenes
  - Full control about the scene
  - Enormeous work



- Procedurally created scenes
  - Scales well, allows for diversity
  - Harder to create realistic scenes

Some dataset studies on FlowNet

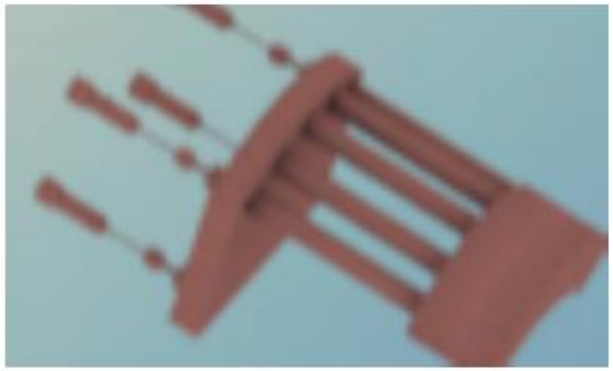
	Sintel Clean Validation
FlowNet, 600k on FlyingChairs	4.45
FlowNet, 1.2M on FlyingChairs	4.24
FlowNet, 1.7M on FlyingChairs	4.21
FlowNet, 1.2M on FlyingThings3D	5.07
FlowNet, 1.7M on FlyingThings3D	4.51
FlowNet, 1.2M on FlyingChairs, 500k on FlyingThings3D	3.79

- Lessons:
  - Longer training helps a little
  - FlyingChairs works better than FlyingThings3D
  - Fine-tuning on FlyingThings3D improves results

How important are realistic scenes?





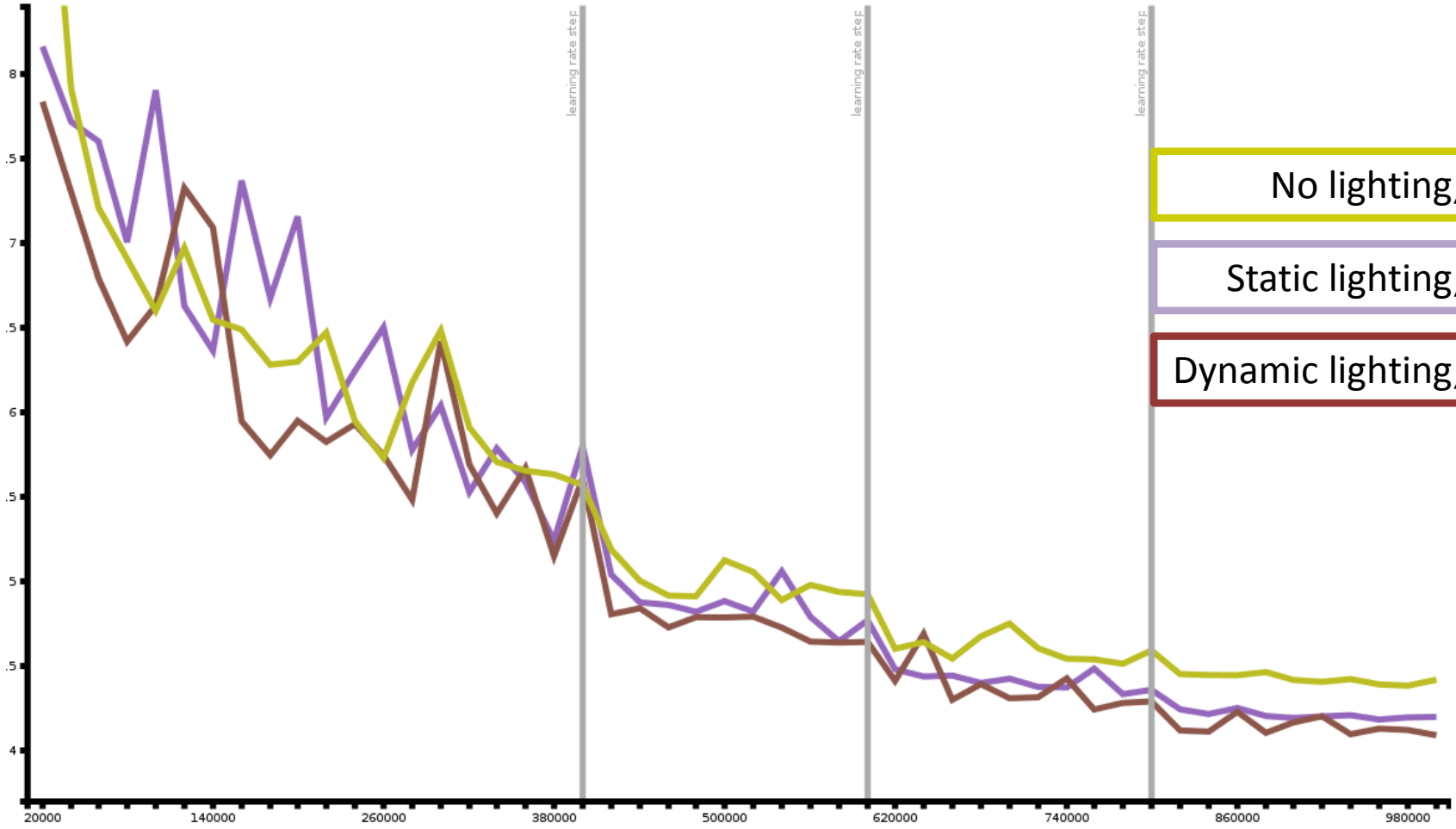



Dynamic lighting

Static lighting

No lighting

How important is realistic lighting?



No lighting, 4.41

Static lighting, 4.19

Dynamic lighting, 4.08

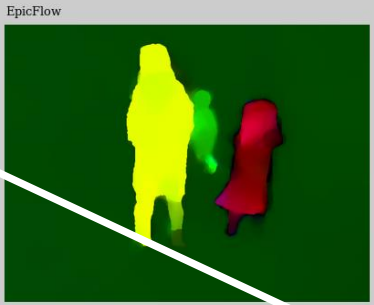
Breaking news: FlowNet 2.0

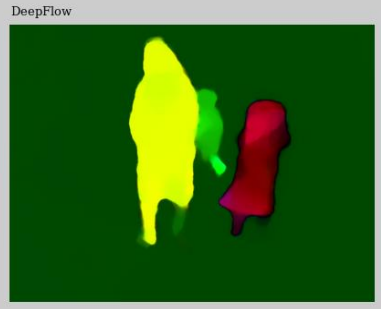
EpicFlow DeepFlow LDOF

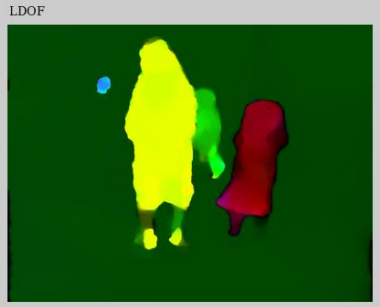
FlowNet 1.0
12fps

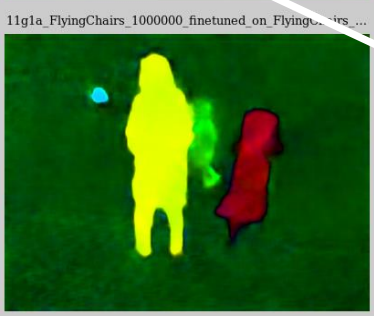
FlowNet 2.0
@20fps

FlowNet 2.0
@10fps





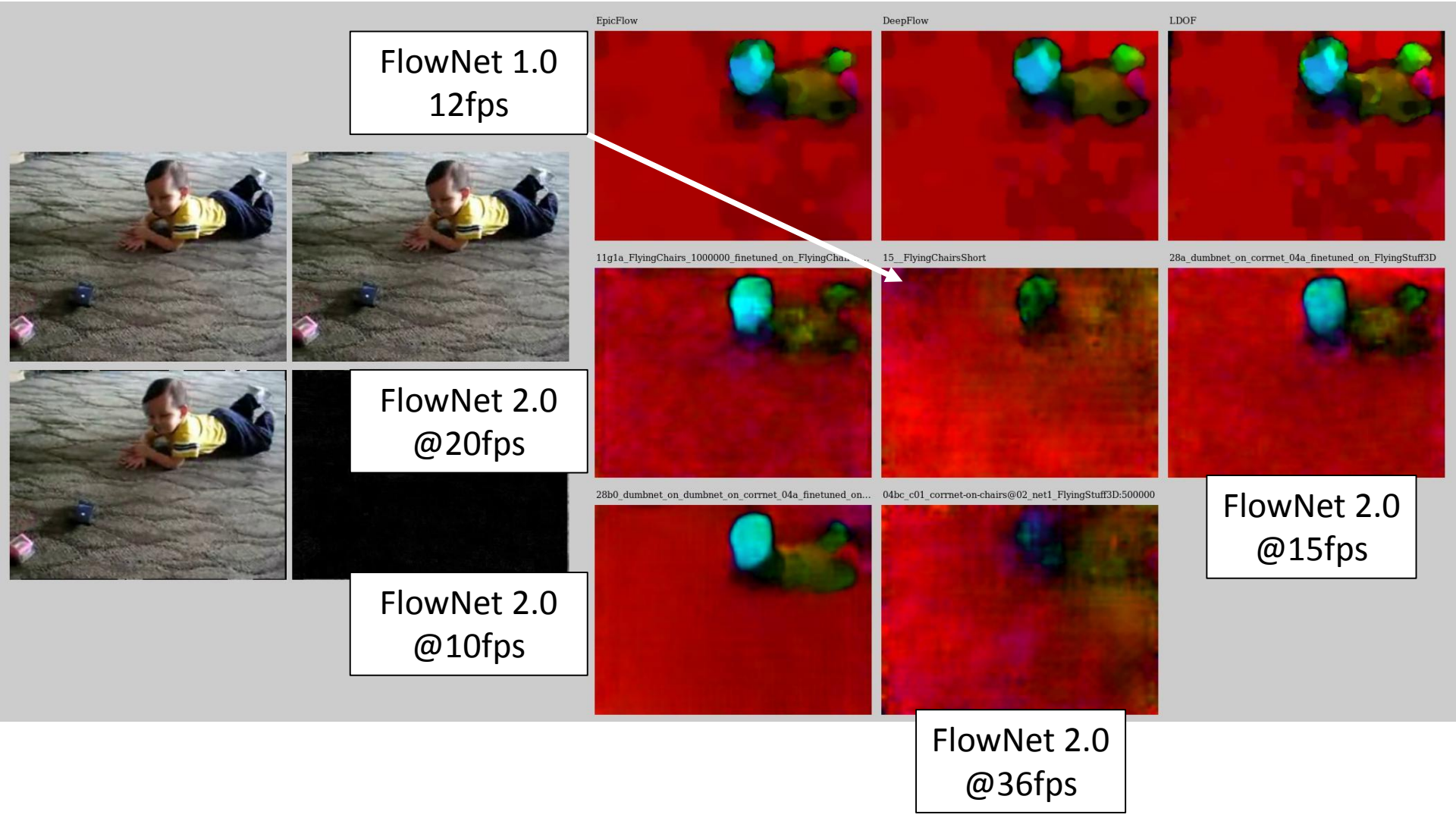




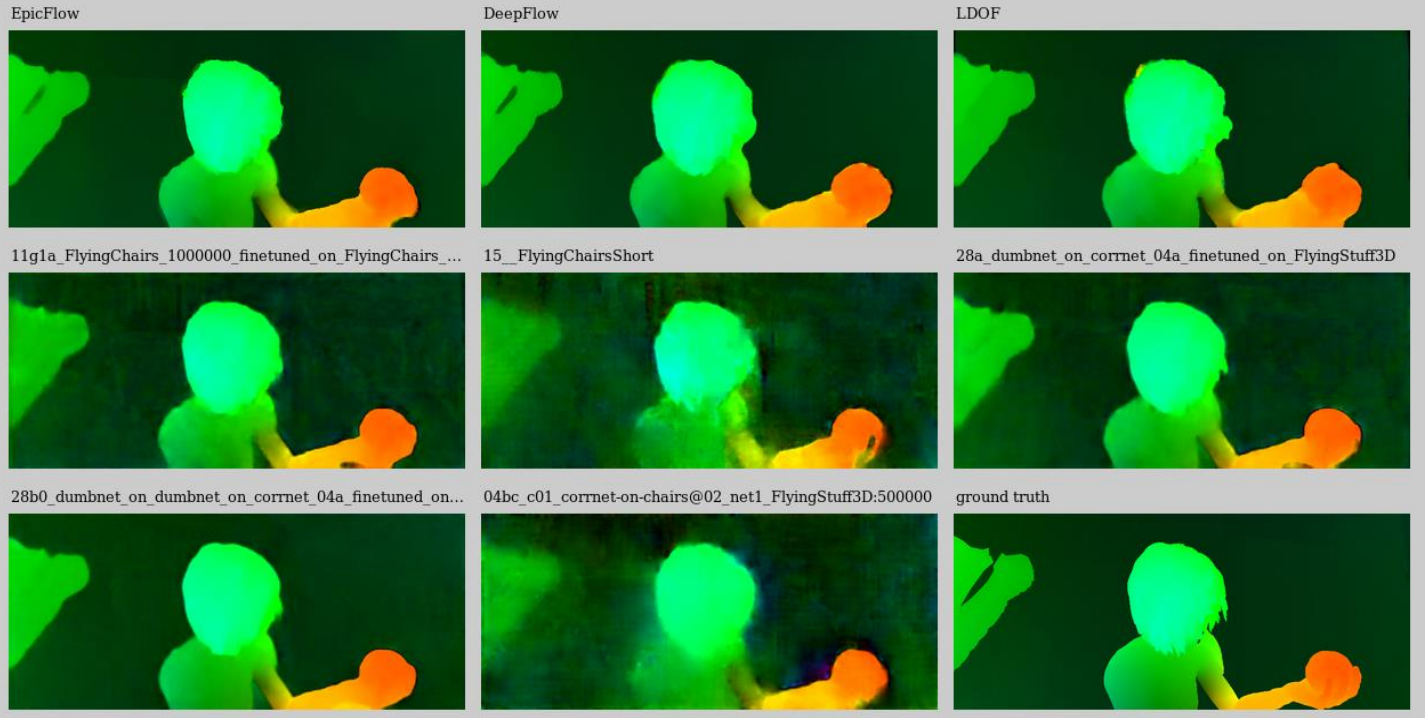
FlowNet 2.0
@15fps

FlowNet 2.0
@36fps

Good on small motion, too

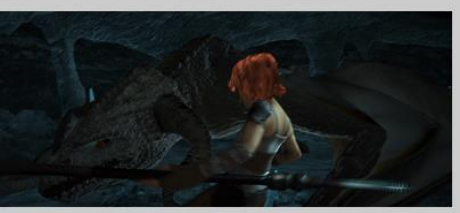


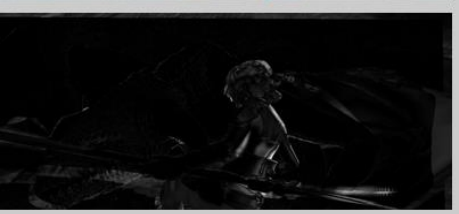
Good on small motion, too

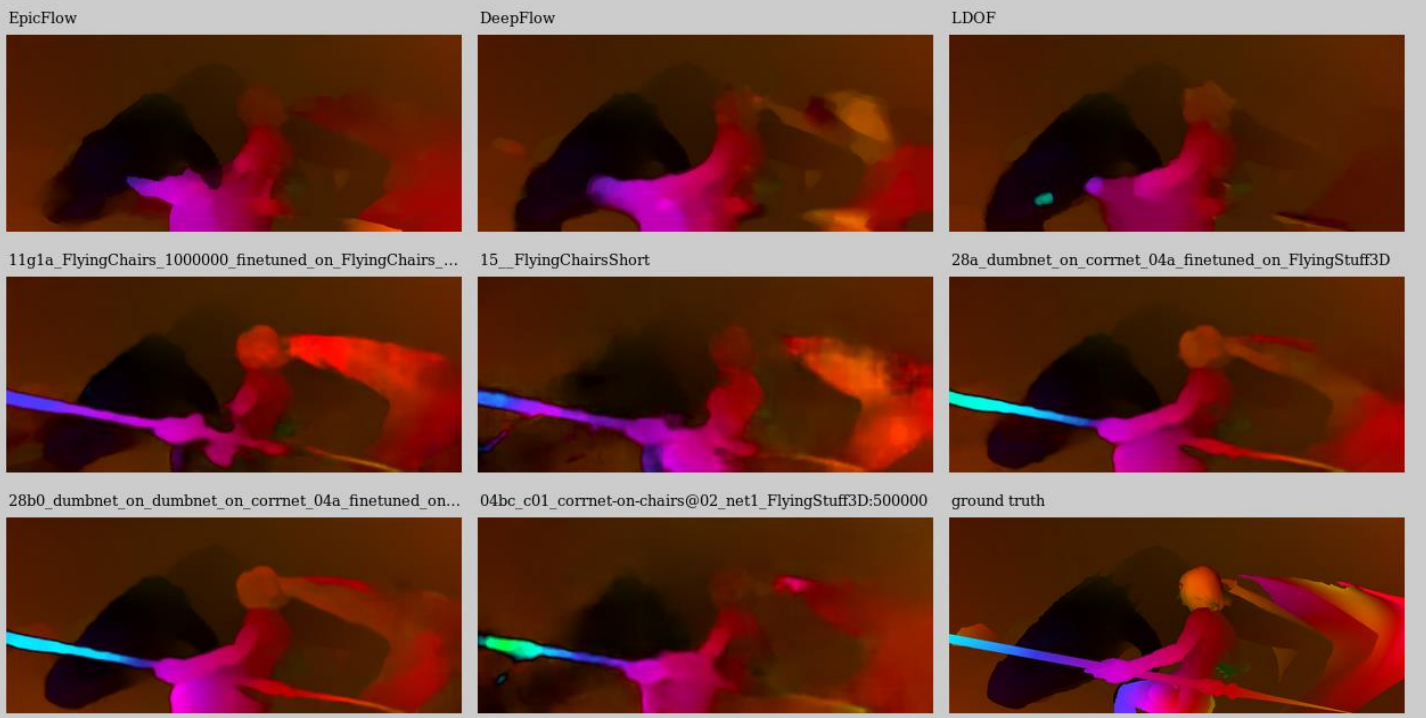


EpicFlow	DeepFlow	LDOF
@20fps	Old @12fps	@15fps
@10fps	@36fps	Ground truth

Super precise at small details

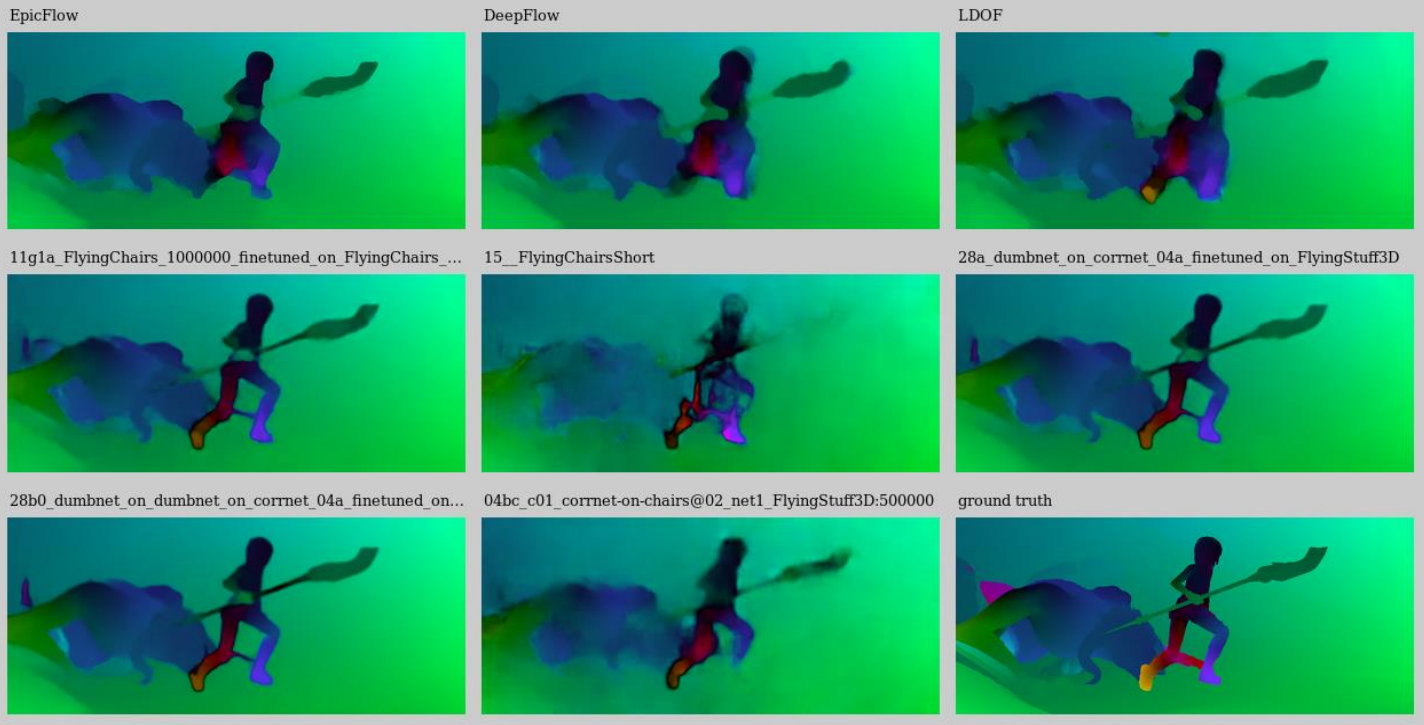






EpicFlow	DeepFlow	LDOF
@20fps	Old @12fps	@15fps
@10fps	@36fps	Ground truth

Super precise at small details



EpicFlow	DeepFlow	LDOF
@20fps	Old @12fps	@15fps
@10fps	@36fps	Ground truth

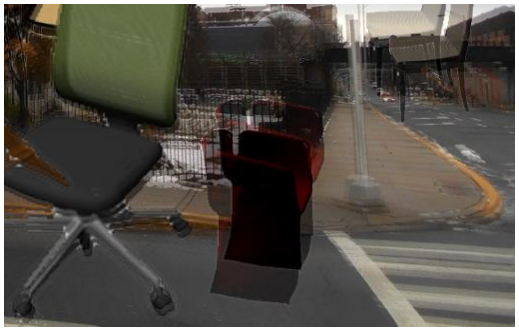
Preliminary numbers on Sintel.train.clean

	EPE	runtime
EpicFlow (Revaud et al. 2015)	2.4	15-20s
FlowNet 1.0 (Dosovitskiy et al. 2015)	4.5	80ms (12 fps)
FlowNet 2.0	**2.1**	90ms (11 fps)
FlowNet 2.0	2.6	28ms (36 fps)
FlowNet 2.0	4.5	**9ms (110 fps)**

Sintel was not used for training

(Training on FlyingChairs and FlyingThings3D)

Summary



Realism is “only” important for learning good priors

There are many ways to create synthetic training data

Procedurally generated datasets are important for diversity

FlowNet 2.0 with up to 100fps and very precise flow to come