Bored by Classification ConvNets?

End-to-end Learning of other Computer Vision Tasks

Thomas Brox

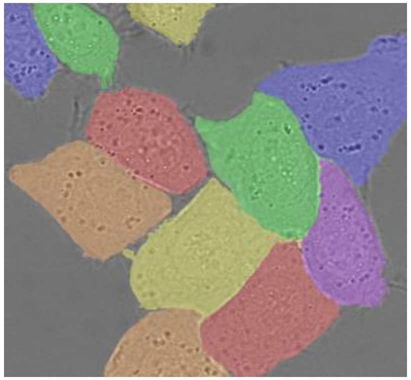
University of Freiburg

Germany

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

Outline

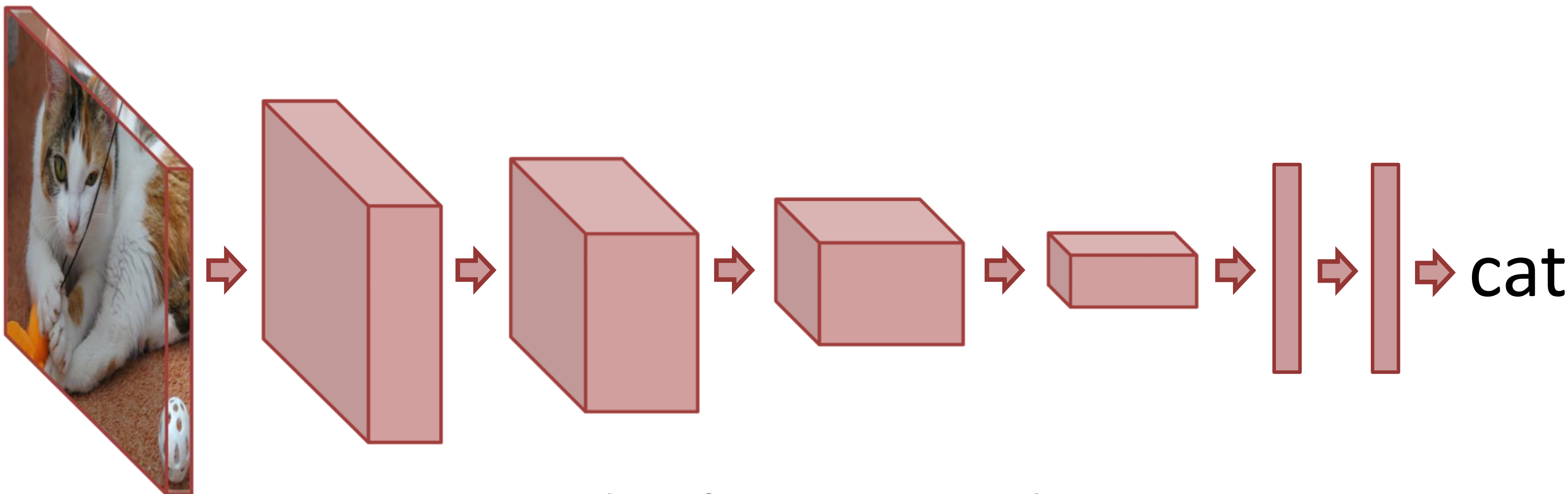
Generative networks



U-Net: Multi-instance segmentation

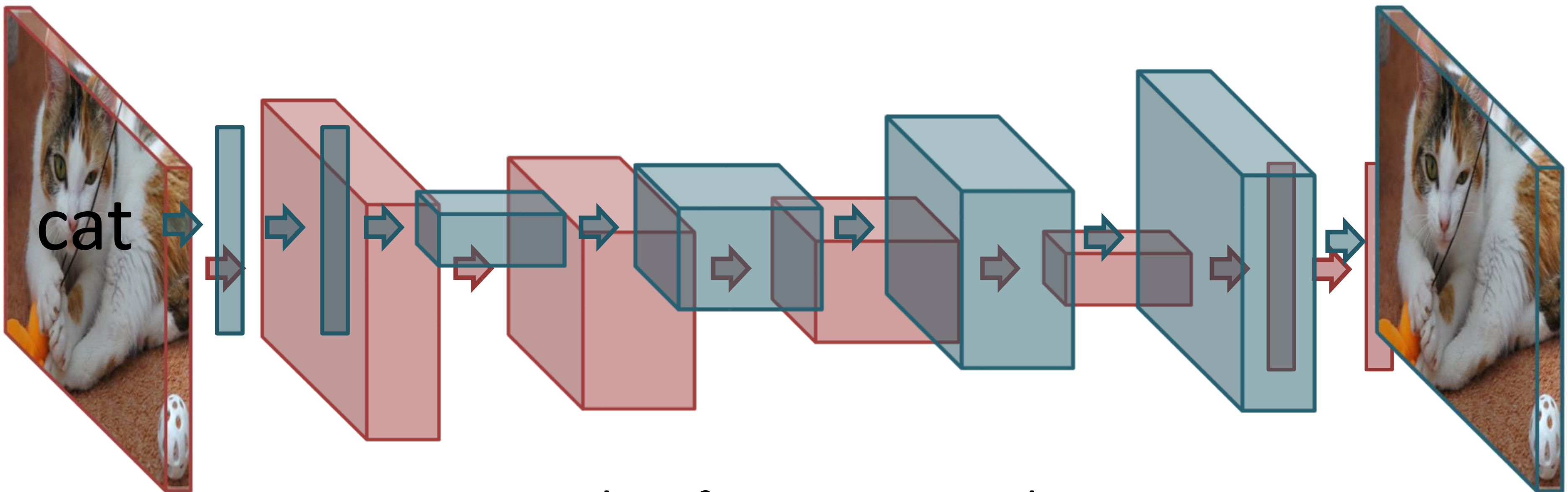
FlowNet: Estimating optical flow

Typical ConvNet architecture



Classification network

Typical ConvNet architecture



Classification network

Up-convolutional network

Alexey Dosovitskiy
CVPR 2015

New: Expanding network architecture

Small gray office chair, side view

cat

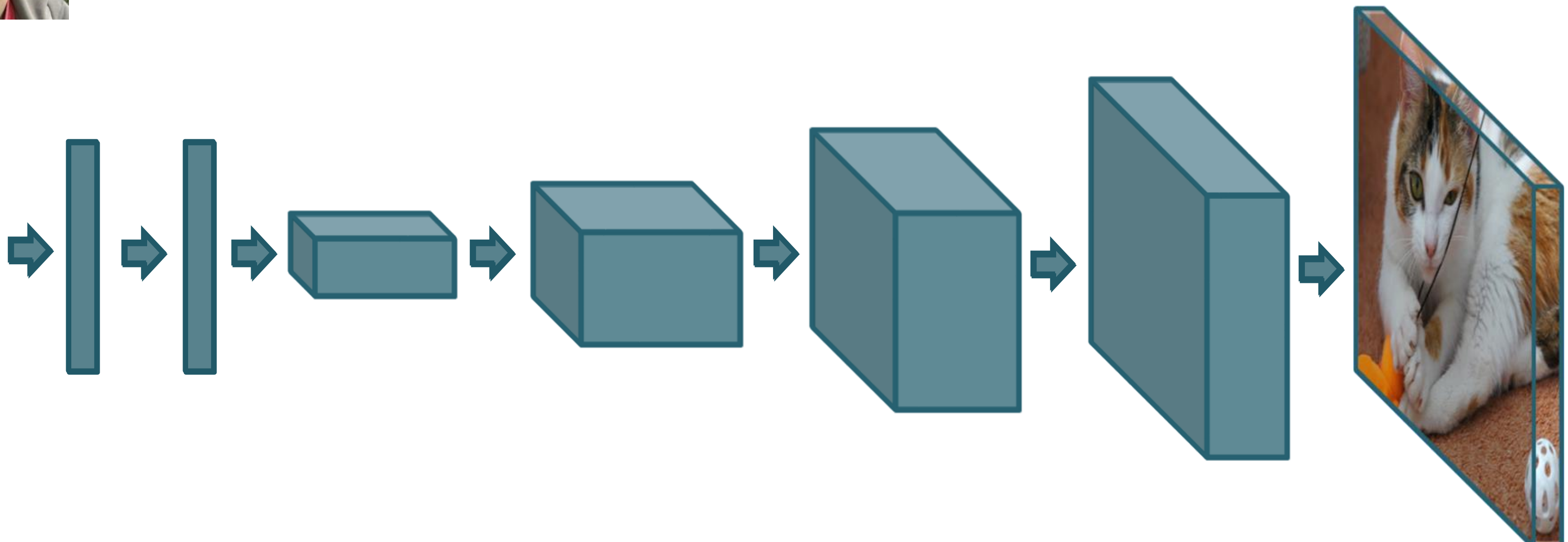
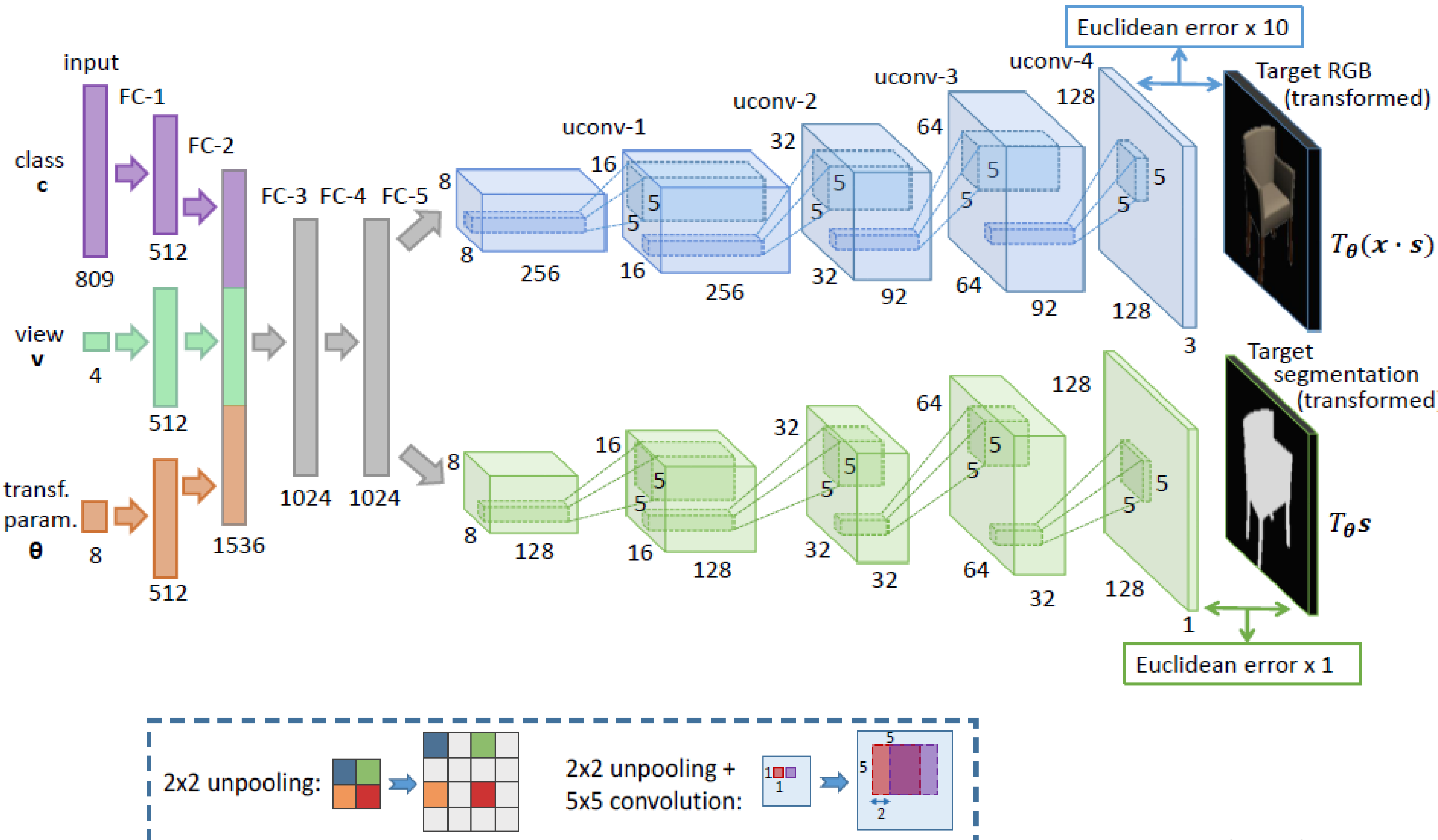


Image generation

Related work:

- Eigen et al. NIPS 2014: Network for depth map prediction
- Long et al. CVPR 2015: Network for semantic segmentation

Generating chair images with a network



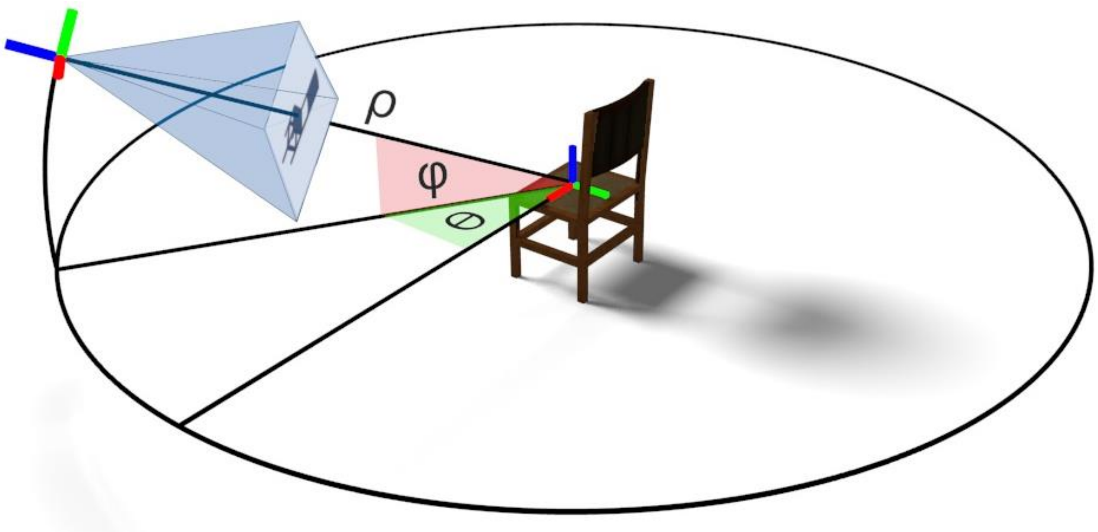
Dosovitskiy et al., CVPR 2015

Training set

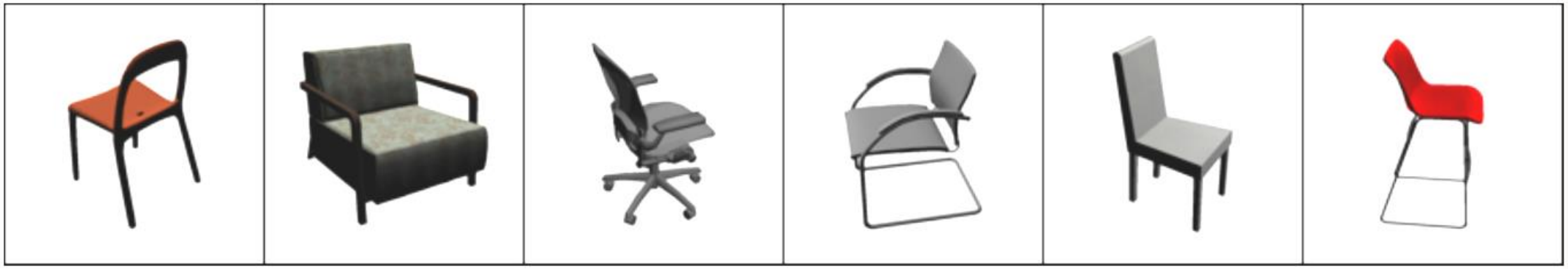
3D chair dataset

Aubry et al. CVPR 2014

Rendering 809 chair styles
From 62 viewpoints



Source: https://github.com/dimatura/seeing3d



Some of the rendered chairs

Generating images of unseen views

<u>Training set split into two subsets:</u>

Source set: 62 viewpoints available (90% of all chair models)

Target set: fewer viewpoints available (10% of all models)

Generating images of unseen views

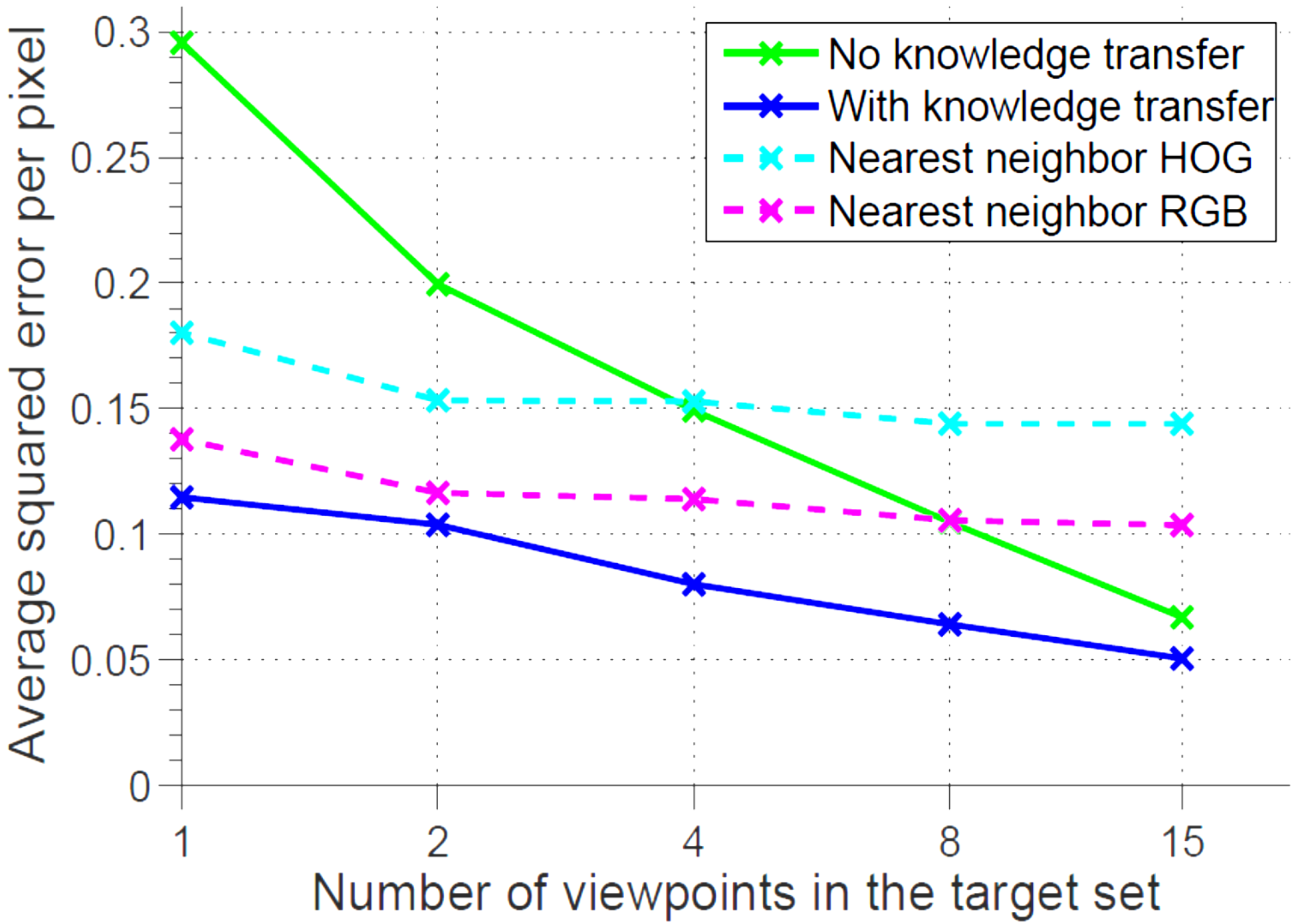
8 azimuths available

4 azimuths available

2 azimuths available

1 azimuth available

Comparison to baselines

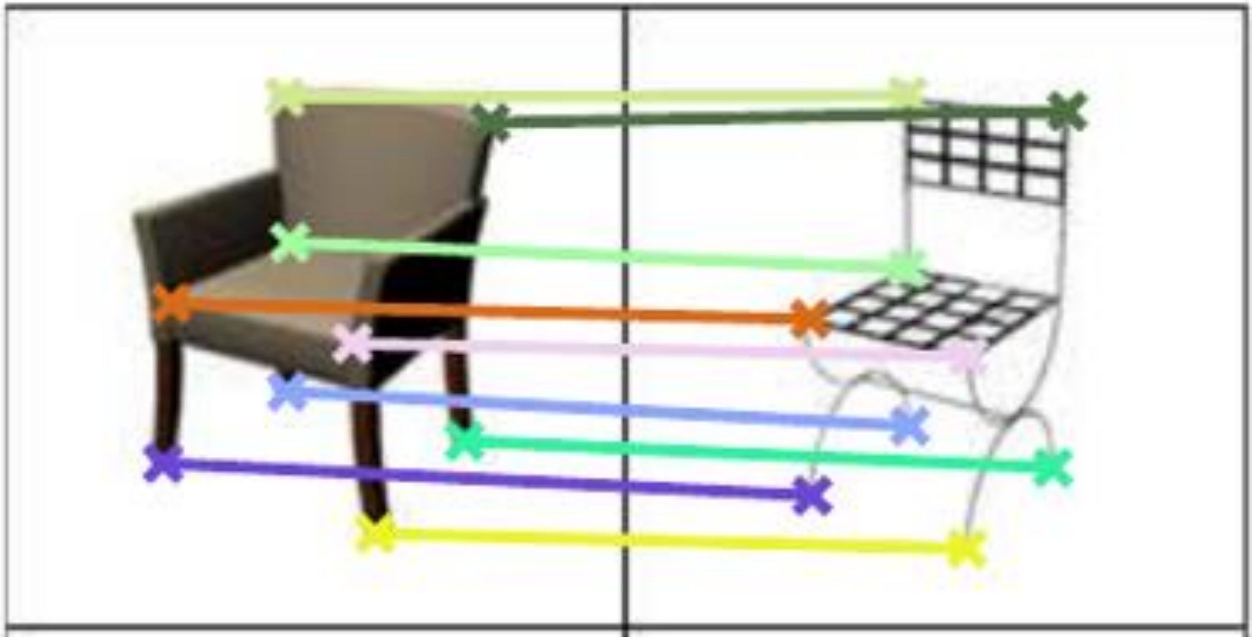


Alexey Dosovitskiy
CVPR 2015

Interpolation of chair styles

Alexey Dosovitskiy
CVPR 2015

Correspondences between chair instances



Alexey Dosovitskiy
CVPR 2015

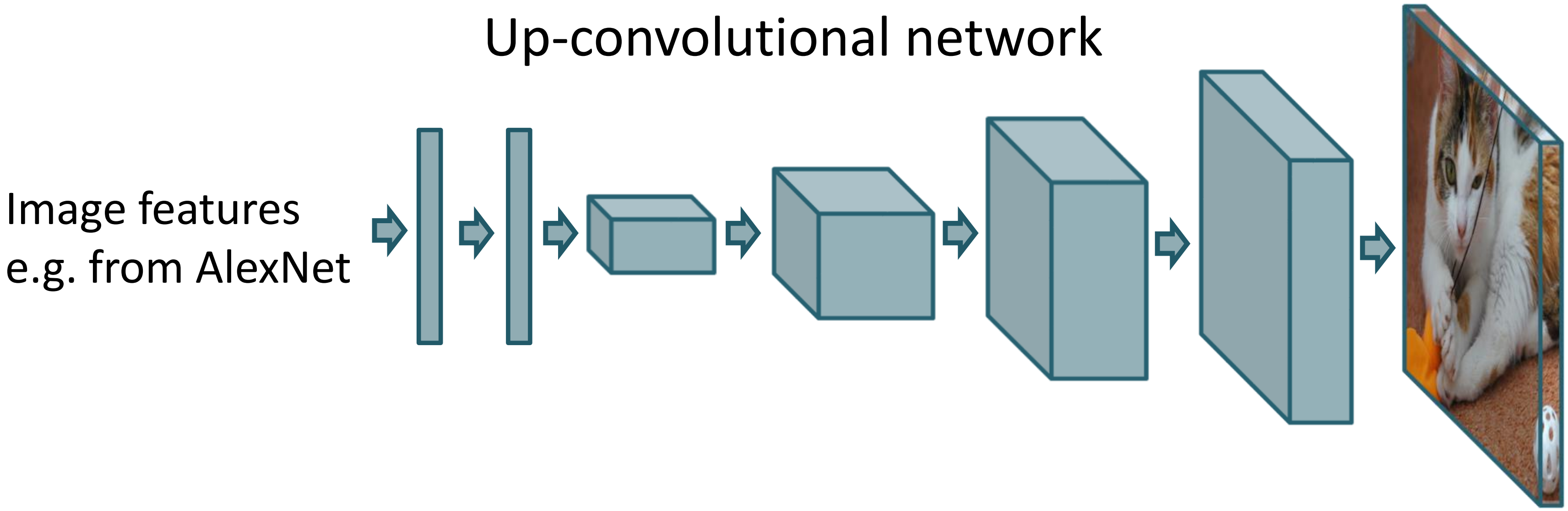
Correspondences between chair instances

- Generate intermediate images with the network

- Track points with optical flow (LDOF) along the sequence

	all	easy	difficult
Deformable Spatial Pyramid Matching (Kim et al. 2013)	5.2	3.3	6.3
SIFT Flow (Liu et al. 2008)	4.0	**2.8**	4.8
Ours	**3.9**	3.9	**3.9**
Human performance	1.1	1.1	1.1

Preview: Inverting ConvNets with ConvNets



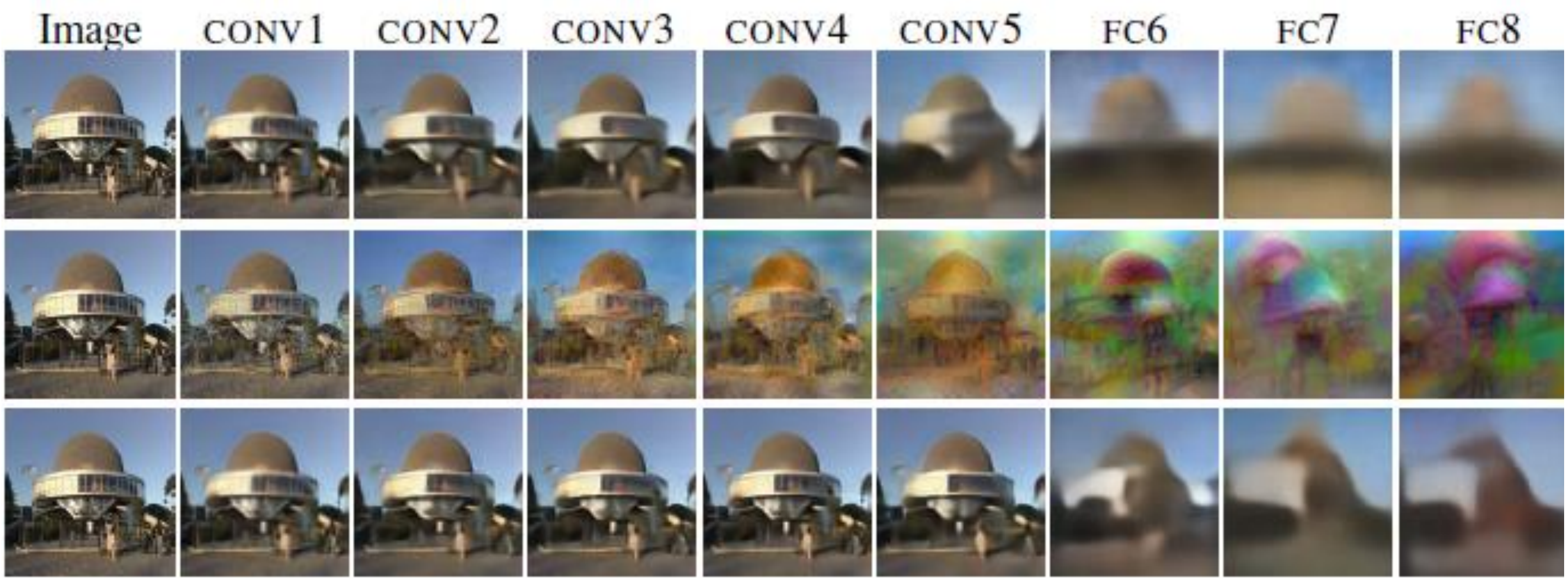
Learn to re-generate the input image from its feature representation

Related work:

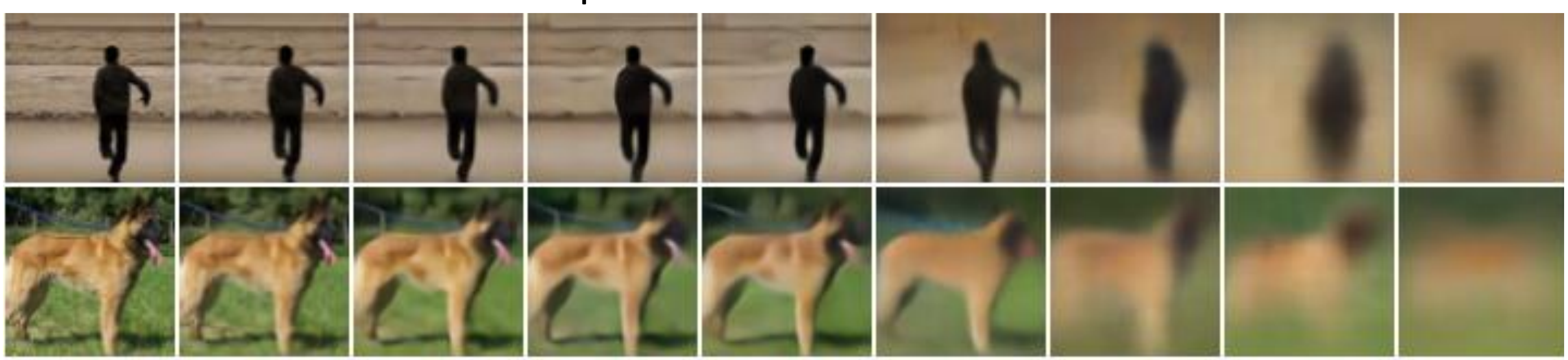
- Mahendran & Vedaldi CVPR 2015
- Zeiler & Fergus ECCV 2014

Alexey Dosovitskiy
arXiv 2015

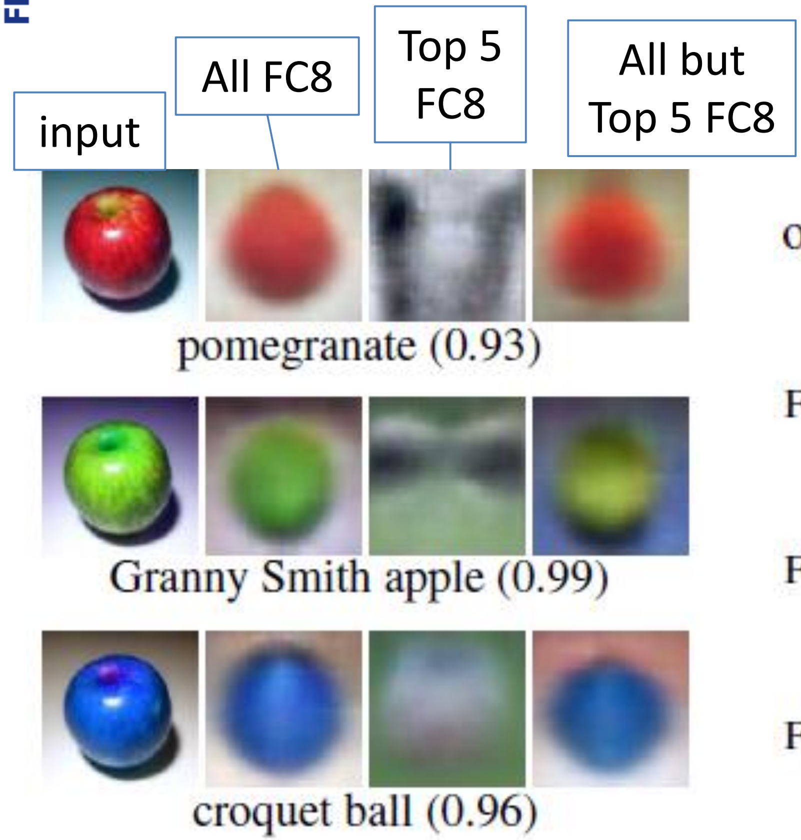
Reconstruction results



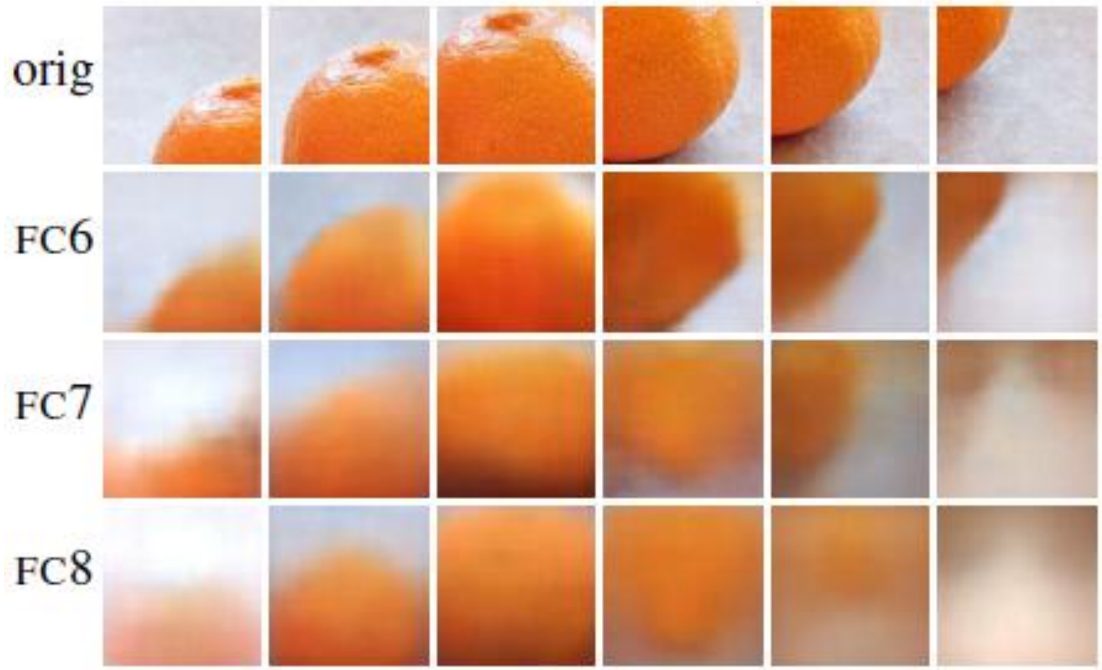
More reconstructions with up-convolutional network:



Color and position are preserved in high layers



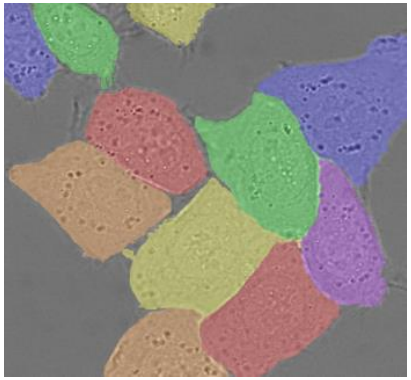
Color experiment



Position experiment

Outline

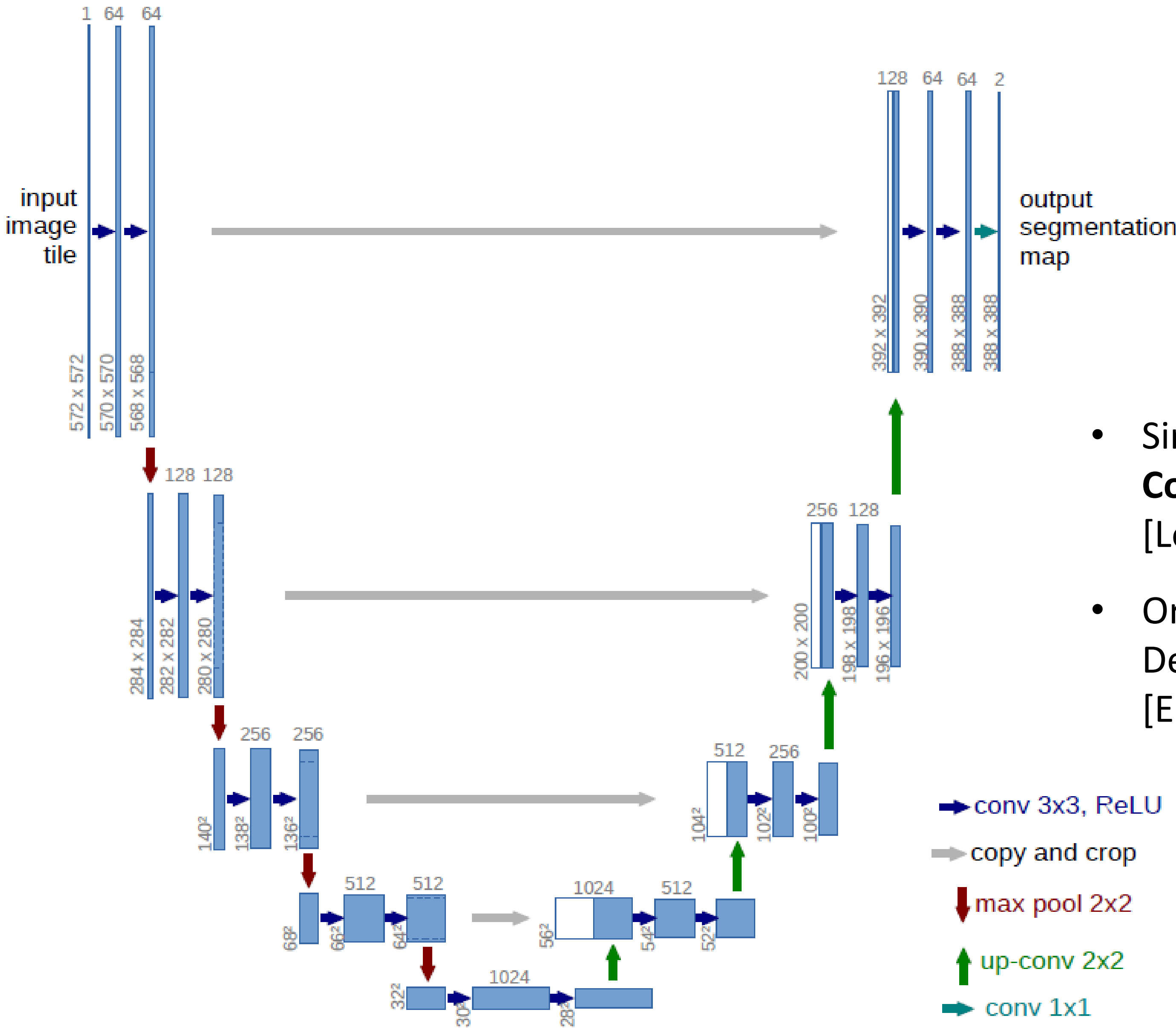
A generative network



U-Net: Multi-instance segmentation

FlowNet: Estimating optical flow

U-Net: Image segmentation with a ConvNet



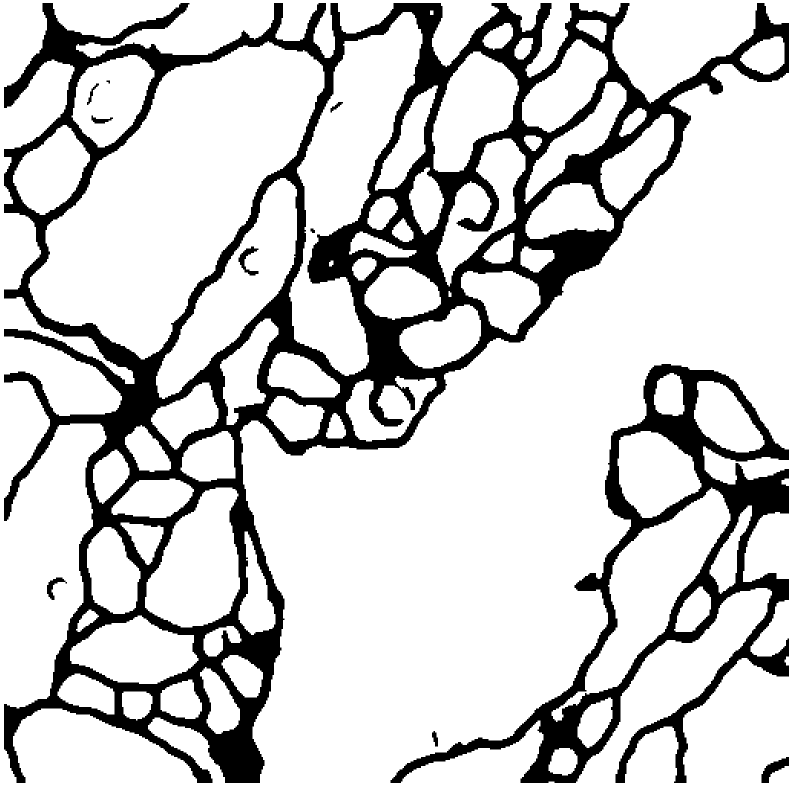
Olaf Ronneberger

Philipp Fischer

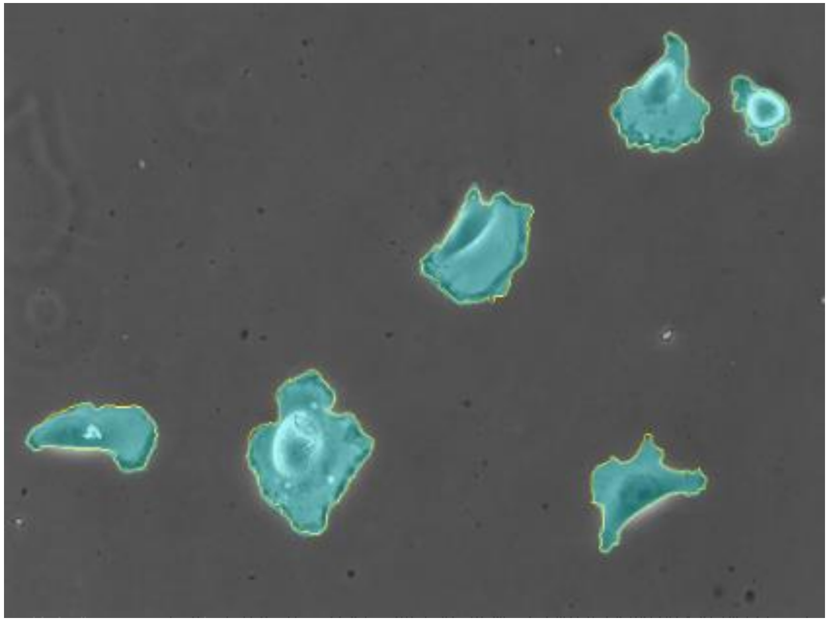
MICCAI 2015

- Similar to **Fully Convolutional Network** [Long et al., CVPR 2015]
- Original inspiration: Depth map prediction [Eigen et al., NIPS 2014]

Binary segmentation

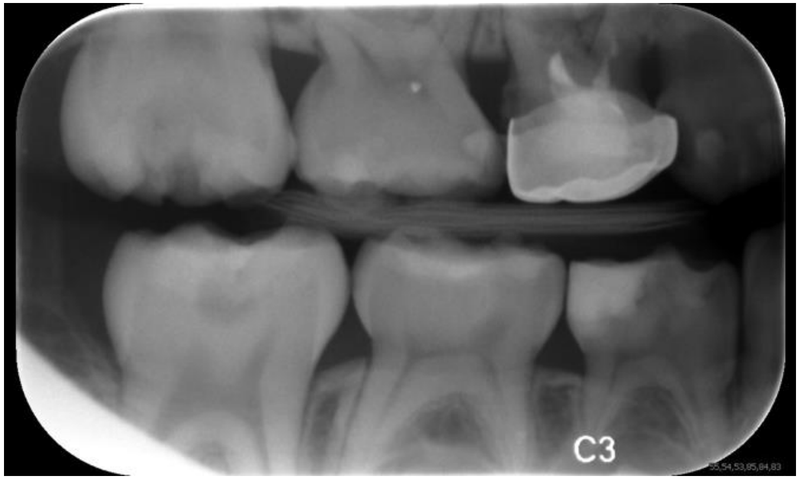


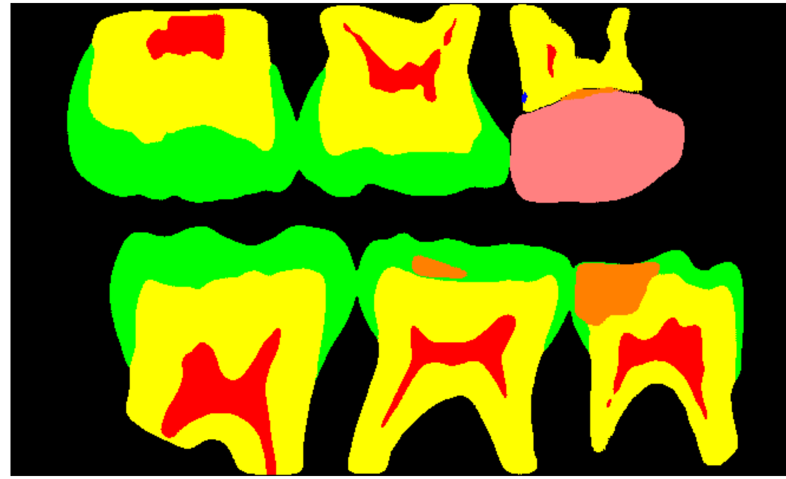
Electron Microscopy
ISBI 2012 Challenge
Rank 1



Light microscopy cell tracking
ISBI 2015 Challenge
Rank 1

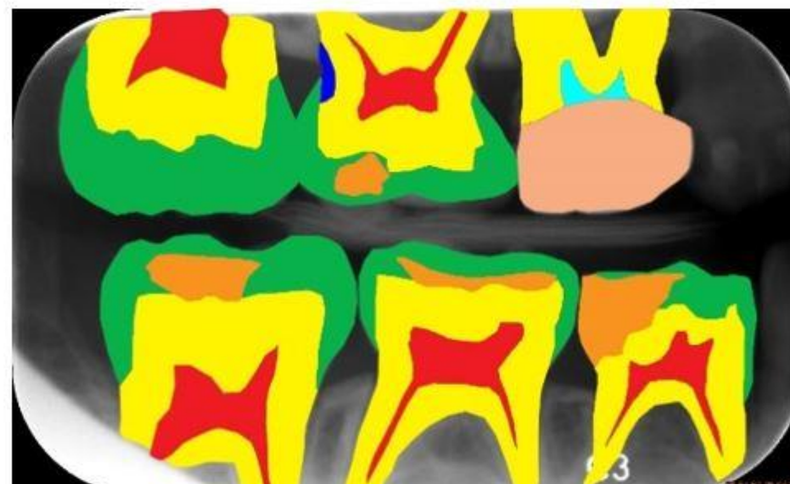
Multi-class semantic segmentation



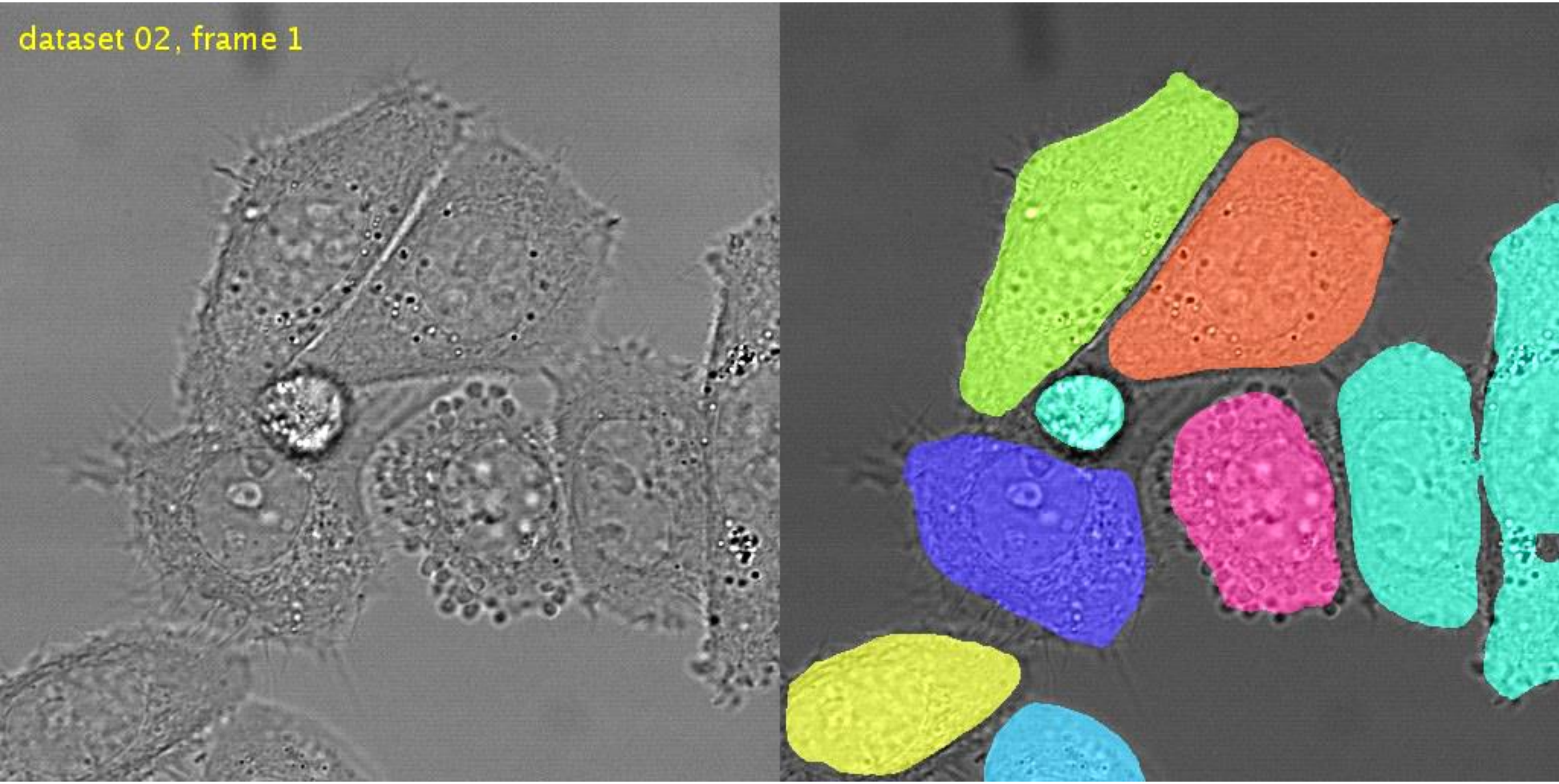


X-ray dental segmentation, ISBI 2015 Challenge, Rank 1

No.	Important Properties Parts
1	caries (blue color)
2	enamel (green color)
3	dentin (yellow color)
4	pulp (red color)
5	crown (skin color)
6	restoration (orange color)
7	root canal treatment (cyan color)



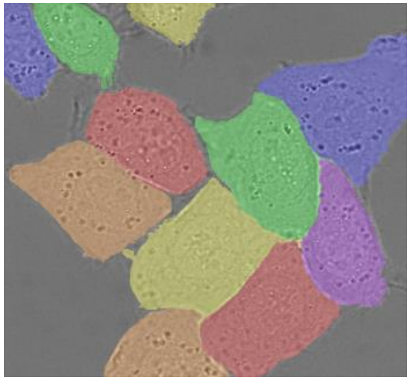
Multi-instance segmentation



Light microscopy, DIC-HeLa cell tracking
ISBI 2015 Challenge: Rank 1

Outline

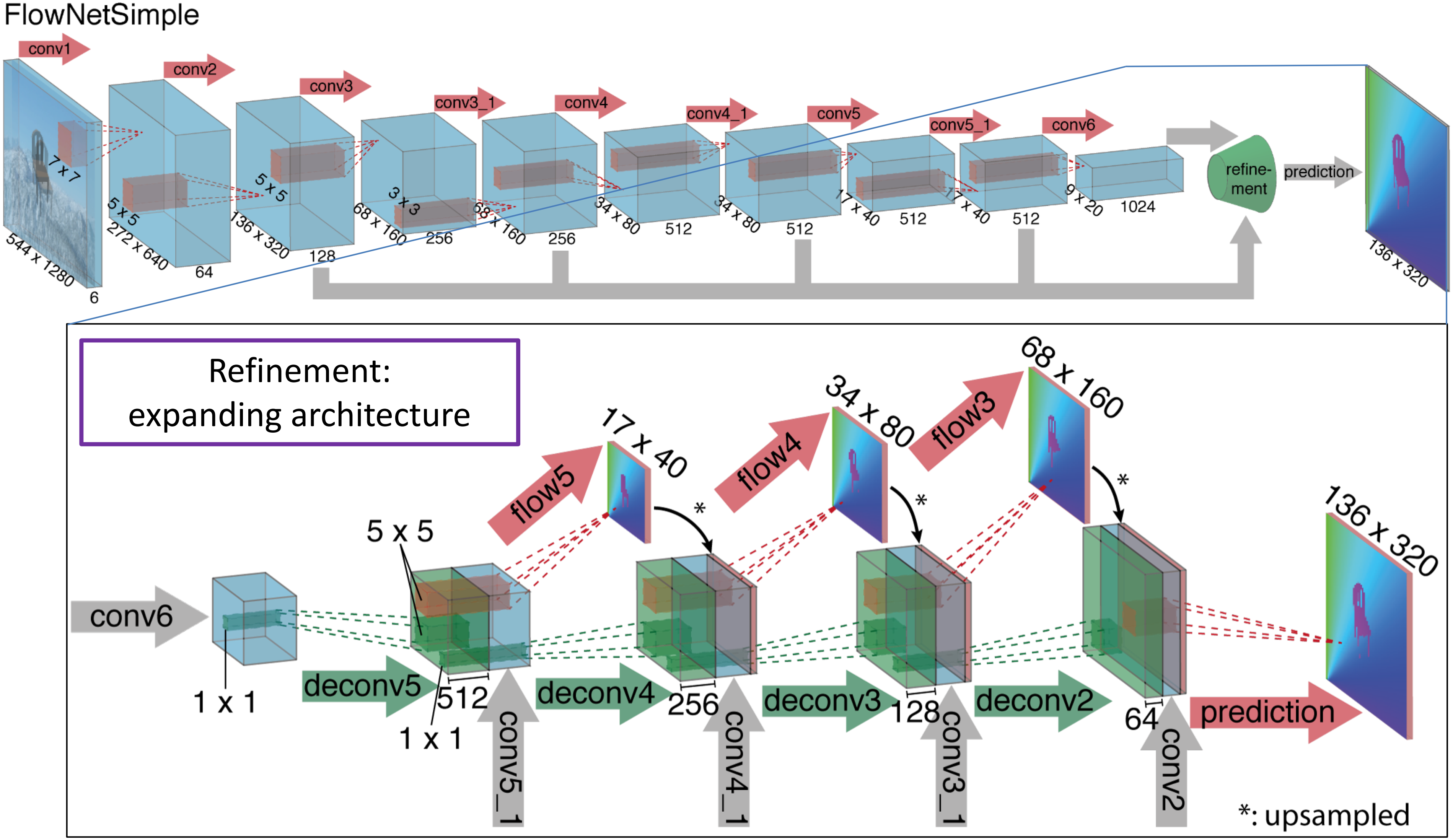
A generative network



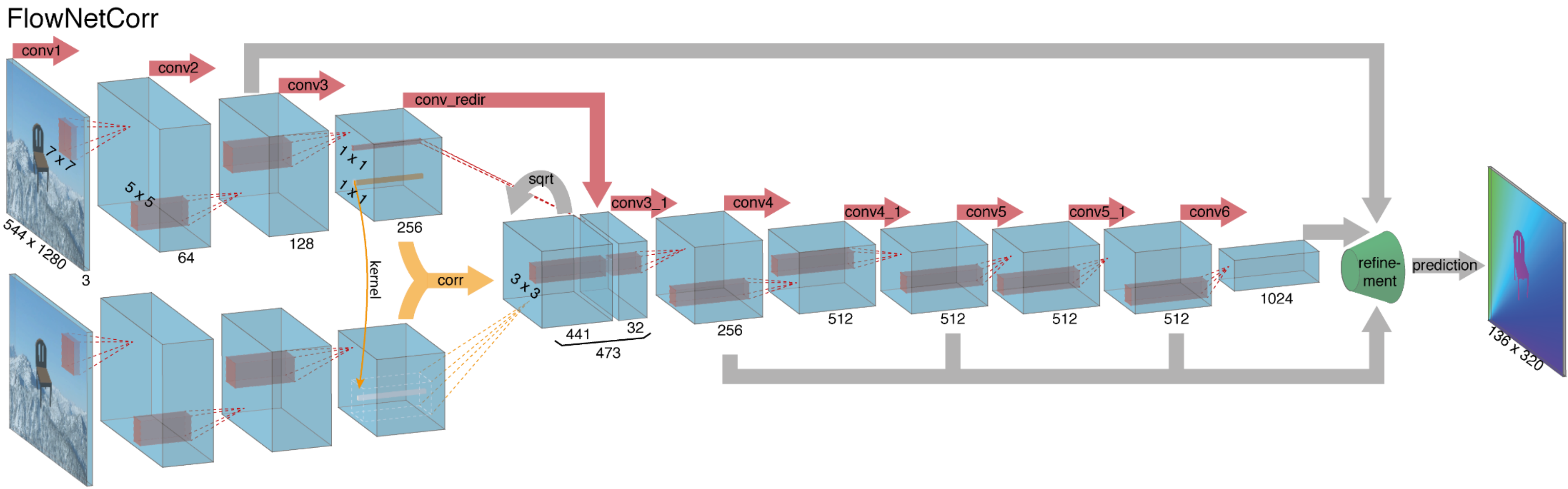
U-Net: Multi-instance segmentation

FlowNet: Estimating optical flow

FlowNet: Estimating optical flow with a ConvNet



Helping the network with a correlation layer



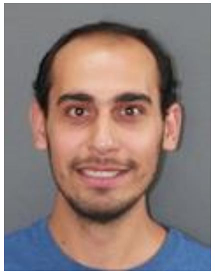
Joint work with the group of Daniel Cremers

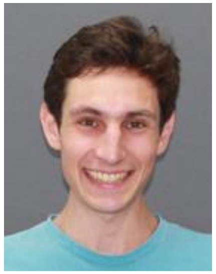

Philipp Fischer


Alexey Dosovitskiy


Eddy Ilg


Philip Häusser


Caner Hazirbas


Vladimir Golkov

Enough data to train such a network?

- Getting ground truth optical flow for realistic videos is hard

- Existing datasets are small:

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

Realism is overrated: the “flying chair” dataset

Rendered image Optical flow

It works!

Input images

Ground truth

FlowNetSimple

FlowNetCorr

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

Results on various datasets

	Middlebury	KITTI	Sintel Clean	Sintel Final	Flying Chairs
EpicFlow	0.39	3.8	4.1	6.3	2.9
DeepFlow	0.42	5.8	5.4	7.2	3.5
LDOF	0.56	12.4	7.6	9.1	3.5
FlowNetS	-	-	7.4	8.4	2.7
FlowNetS+v	-	-	6.5	7.7	2.9
FlowNetS+ft	-	9.1	7.0	7.8	3.0
FlowNetS+ft+v	0.47	7.6	6.2	7.2	3.0
FlowNetC	-	-	7.3	8.8	2.2
FlowNetC+v	-	-	6.3	8.0	2.6
FlowNetC+ft	-	-	6.9	8.5	2.3
FlowNetC+ft+v	0.5	-	6.1	7.9	2.7

Networks can compete with state-of-the-art conventional optical flow estimation methods

Can handle large displacements

Input images

Ground truth

FlowNetSimple

FlowNetCorr

DeepFlow (Weinzaepfel et al. ICCV 2013)

EpicFlow (Revaud et al. CVPR 2015)

Sometimes wrong direction

Input images

Ground truth

FlowNetSimple

FlowNetCorr

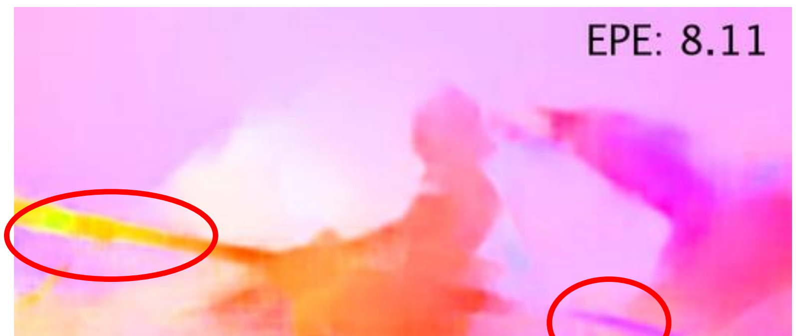
DeepFlow (Weinzaepfel et al. ICCV 2013)

EpicFlow (Revaud et al. CVPR 2015)

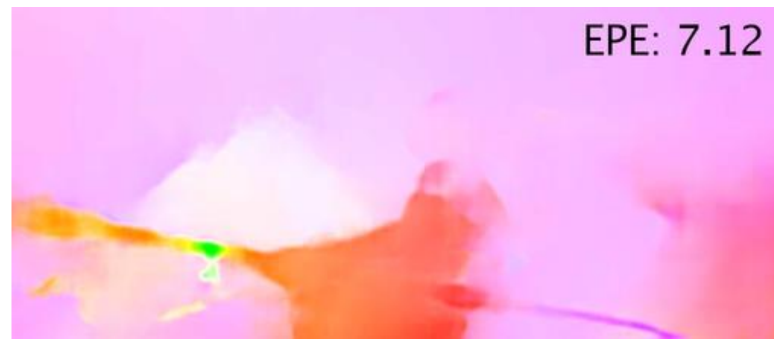
Often captures fine details

Input images

Ground truth



FlowNetSimple



FlowNetCorr

DeepFlow (Weinzaepfel et al. ICCV 2013)

EpicFlow (Revaud et al. CVPR 2015)

Results on “Flying chairs” test set

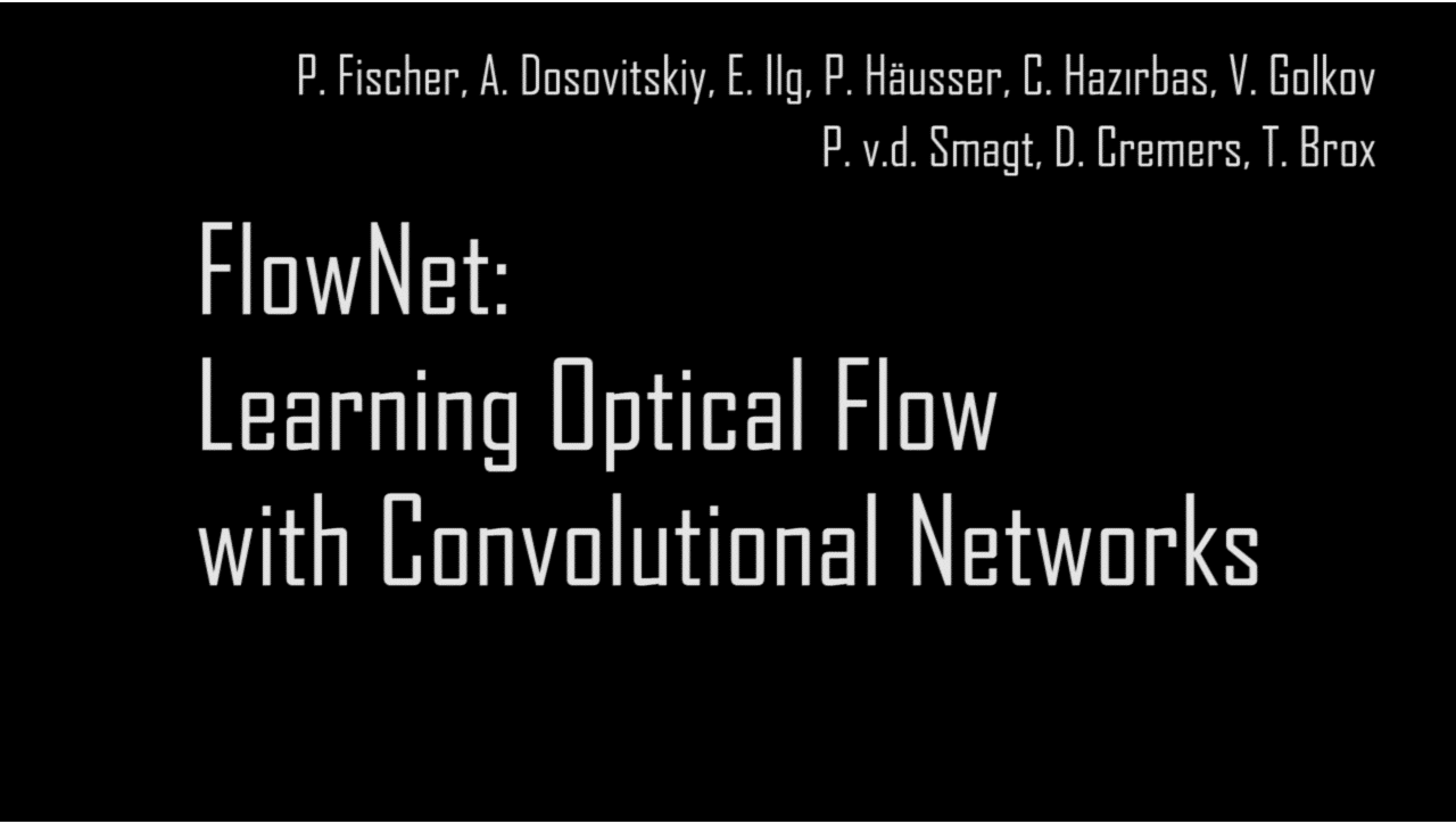
Input images

Ground truth

EpicFlow (Revaud et al. CVPR 2015)

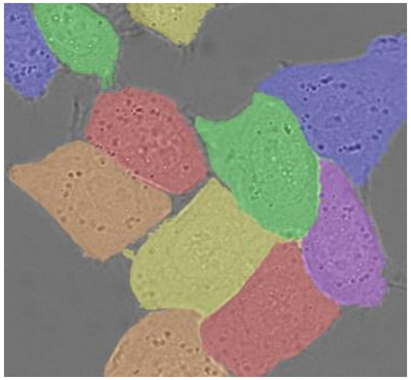
FlowNetCorr

Runs with 10fps on the GPU



Summary

A generative network



U-Net: Multi-instance segmentation

FlowNet: Estimating optical flow

Tip of the day

Replace student descent by gradient descent