Introduction to Computer Vision

Instructor - Simon Lucey

16-720B - Computer Vision

Carnegie Mellon
THE ROBOTICS INSTITUTE

Today

- 16-720B Logistics
- Computer Vision - The Science
- Computer Vision - The Applications
- Computer Vision - The Trends

16-720B Waitlist

- We are at 143 enrolled with 79 students on wait list.
- I’m getting numerous requests of the form:-
  - “how likely is it that I’ll get registered?”
  - “can I switch from A to B (or vice-versa)?”
- **Answer**: “unlikely”
- If you are considering dropping, please do so quickly.

16-720 A versus B

- 16-720B is an accelerated version of 16720-A.
- Homeworks are longer and harder.
- All homeworks will be in python not MATLAB.
- Assumed previous exposure to core topics such as:

i) solving linear systems (e.g. $\mathbf{y} = \mathbf{A}\mathbf{x}$)

ii) gradient descent optimization,

iii) convolution and the Fourier transform.

16-720 A versus B

- Slides between A & B will be different, but cover similar topics (with B slightly accelerated).
- Assessment will be similar to that offered in A, with additional "advanced" questions per homework.
- **Be patient**: even though B will be moving quickly it may cover concepts that some of you have previously covered. This is unavoidable in such a large class.

16-720 B - Assessment

- All course logistics, schedule and assessment details can be found on the website.

  www.cs.cmu.edu/~16720b

- The first 5 homeworks (with considerable python implementation) worth 18% each, and the last homework (HW6) worth 9%. 1% of grade will be attributed to Piazza use.
- There is **no** final project.
- Homeworks and due dates can be found on the lecture webpage.
- Homeworks must be submitted on Canvas by 11:59 pm on the given due date.

16-720 B - Late hand in policy

- Each student has a total of 5 late-day points for the course. You can extend an assignment deadline by one day using one point.
- Rules for the late-day points are as follows:
  - Late-day points CANNOT be used for HW6.
  - If all late-day points are used up, late homework submissions will be graded as 0%.
  - A maximum of 3 late-day points are allowed for the same assignment. If the submission is late by more than 3 days, it will be graded as 0%.

16-720B - TAs

- TA contact hours will be Mon-Fri (6:00pm - 7:30pm) in EDSH 200, from 3rd of September until the end of semester.

- TAs in this class are (please stand up):-
  - Adam Villaflor (lead TA)
  - Xian Zhou
  - Kushal Vyas
  - Abhay Gupta
  - Anuj Pahuja
  - Nitin Singh
  - George Joseph

16-720B - Piazza & Canvas

- All course assessment will be submitted through Canvas.
- All course discussions will be through Piazza (see link in Canvas).
- 1% of your grade will be linked to how actively you use Piazza.
- If appropriate try to answer each others queries.
- Please use Piazza properly.
  - **Good** place for discussing concepts and ideas.
  - **Good** place for finding mistakes in homeworks.
  - **Bad** place for dissecting coding mistakes and asking answers for homeworks.
  - We will be watching!!!

Python

- 16720B is now being offered using only Python.
- You must have a recent version of python to complete the homeworks.
- The recommended approach is to use the Anaconda Python 3 distribution.
- Ensure you install:- pytorch and opencv (you can use conda).
- I would encourage anyone who wants to brush up on their Python skills to check out:-
  - Cyrille Rossant, IPython Interactive Computing and Visualization Cookbook, 2nd Edition, Packt publishing.

Recommended Texts

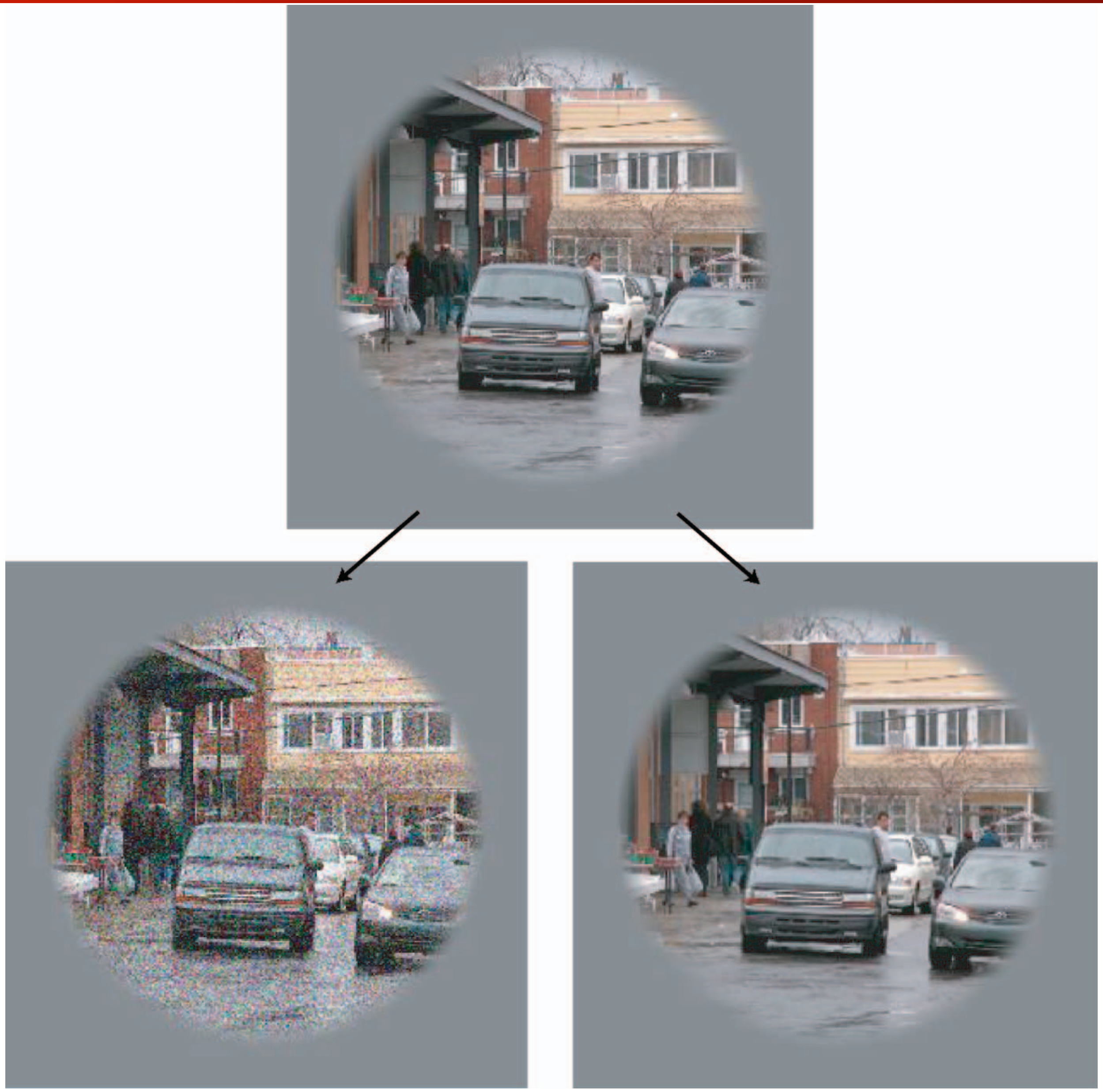
- There is no set text for this course as we will be drawing from numerous sources.

- Other books we will use (but are not essential) include:-
  - Simon J. D. Prince, Computer Vision – Models, Learning and Inference
  - Peter Corke, Robotics, Vision and Control: Fundamental Algorithms in MATLAB
  - Richard Hartley & Andrew Zisserman, Multiple View Geometry in Computer Vision
  - Ian Goodfellow, Yoshua Bengio & Aaron Courville, Deep Learning

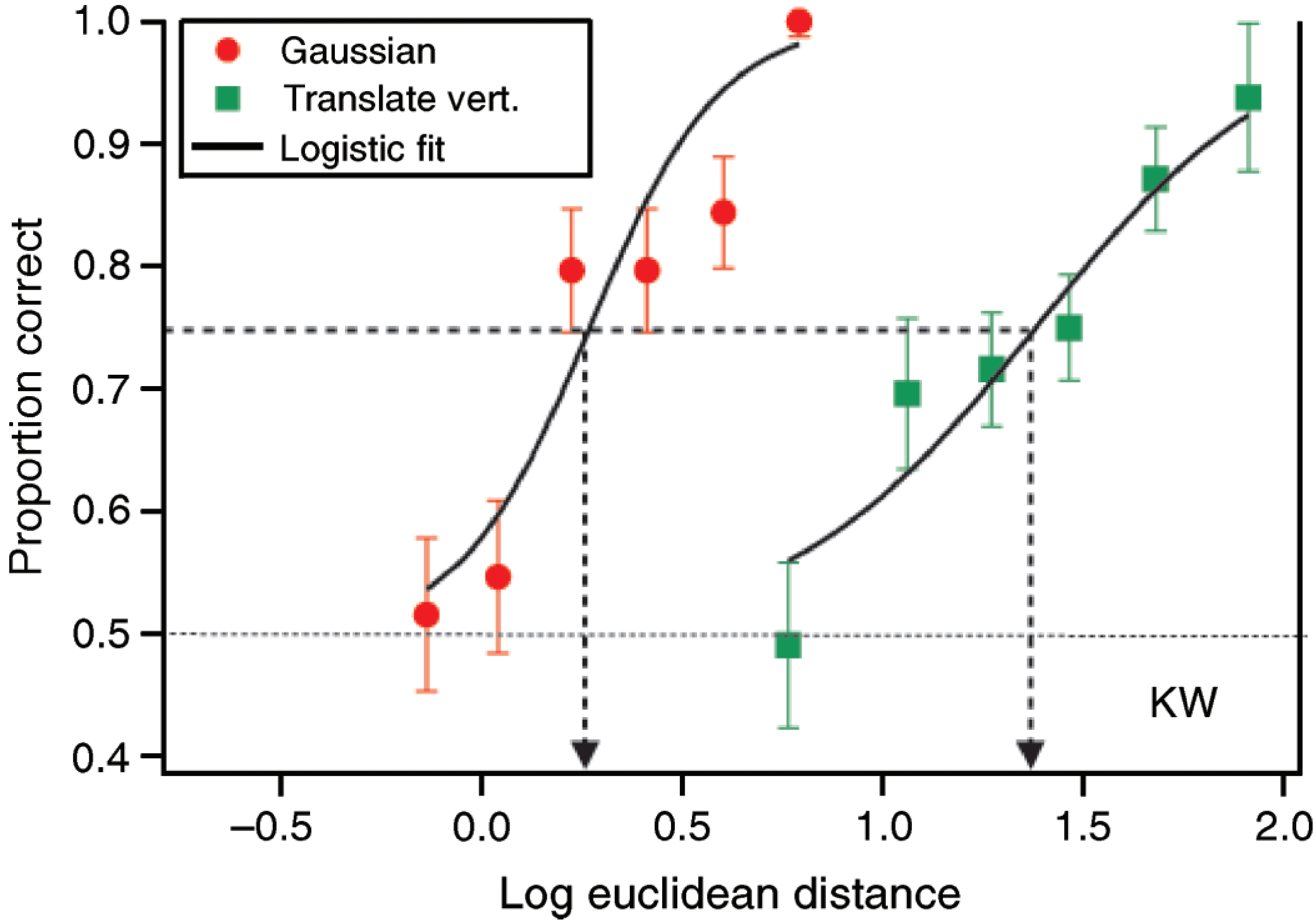
Today

- 16-720B Logistics
- Computer Vision - The Science
- Computer Vision - The Applications
- Computer Vision - The Trends

Question?



Answer



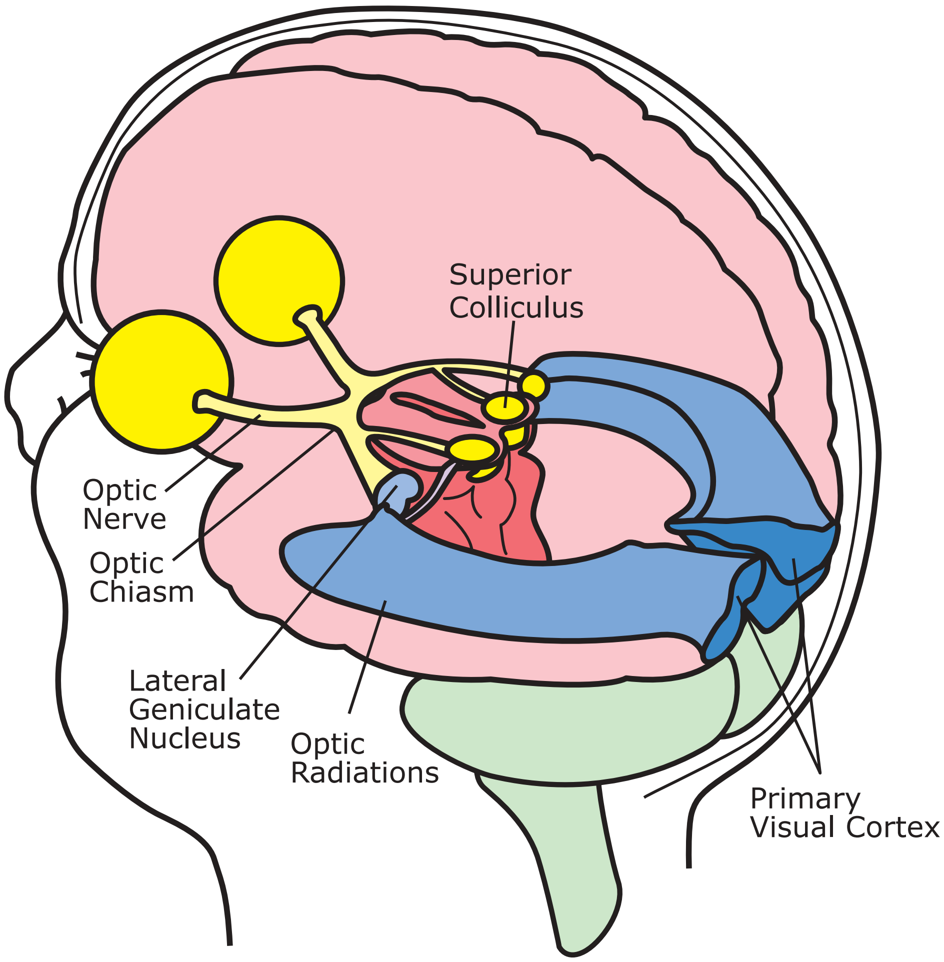
Kingdom, Field, Olmos, 2007

Why?

Why?

207	245	77	21	247	211	240	1
219	41	58	179	161	154	184	98
215	145	187	71	251	249	65	100
192	2	189	247	166	63	232	213
105	94	66	190	156	61	89	145
159	154	87	184	101	105	72	71
192	111	6	94	60	70	65	226
175	120	210	226	80	183	168	184
134	56	36	240	159	178	76	135
239	244	199	9	132	104	188	185
245	210	78	199	0	92	9	246
5	121	187	122	107	47	12	119
230	171	135	36	82	54	65	37
61	140	79	19	161	96	127	187
56	223	46	6	180	186	142	244
28	20	61	2	178	187	98	220

Primary Visual Cortex



What are the fundamental principles at play?

Challenge

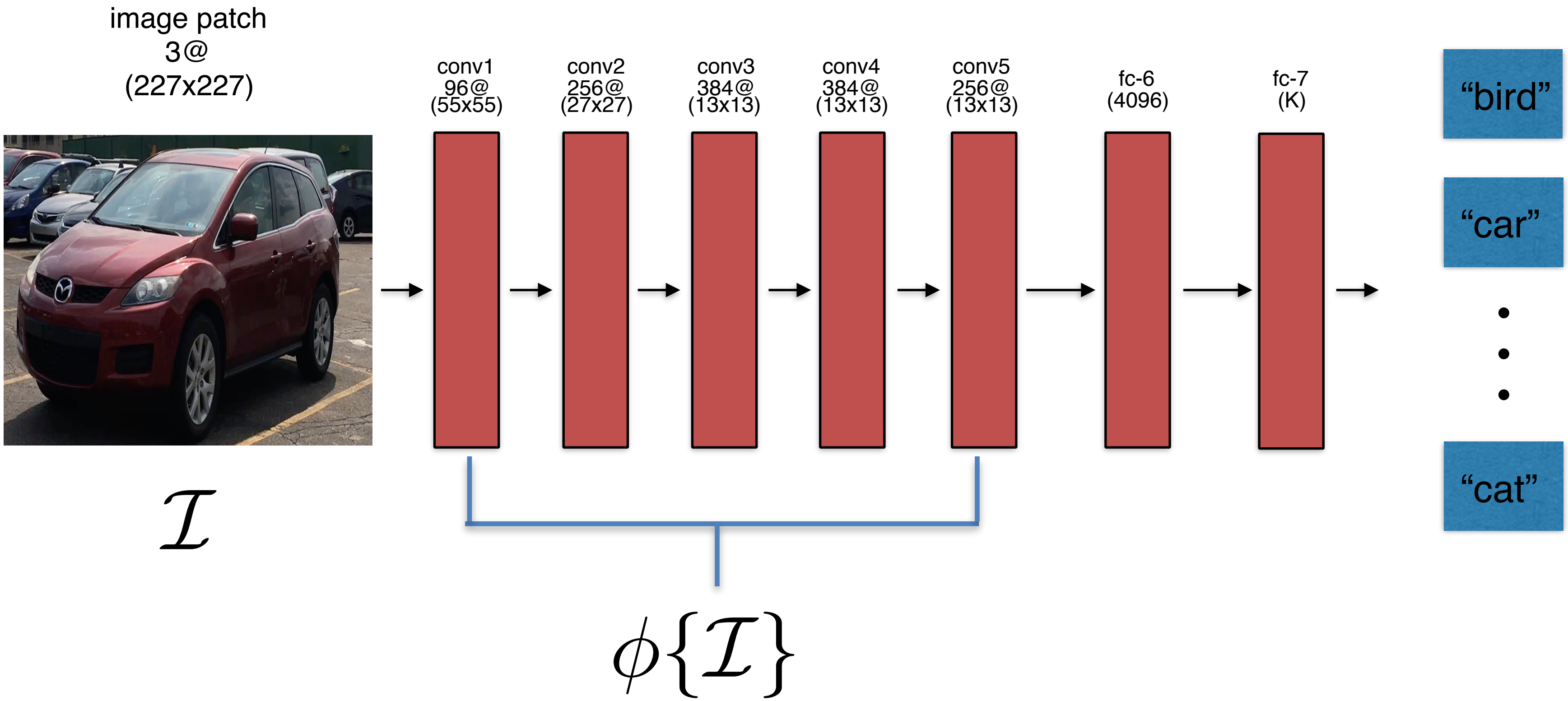
207	245	77	21	247	211	240	1
219	41	58	179	161	154	184	98
215	145	187	71	251	249	65	100
192	2	189	247	166	63	232	213
105	94	66	190	156	61	89	145
159	154	87	184	101	105	72	71
192	111	6	94	60	70	65	226
175	120	210	226	80	183	168	184
134	56	36	240	159	178	76	135
239	244	199	9	132	104	188	185
245	210	78	199	0	92	9	246
5	121	187	122	107	47	12	119
230	171	135	36	82	54	65	37
61	140	79	19	161	96	127	187
56	223	46	6	180	186	142	244
28	20	61	2	178	187	98	220

$$\mathcal{I}$$

Challenge

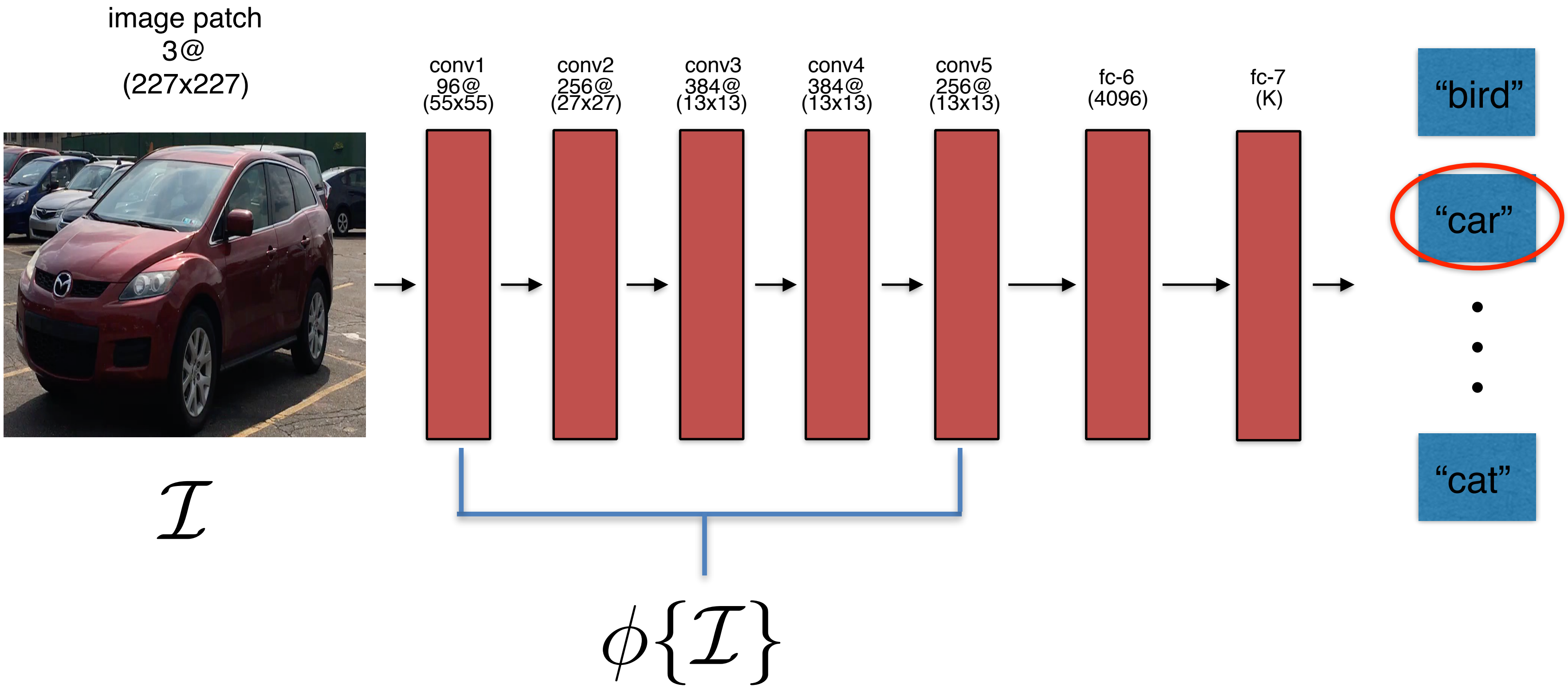
$$\phi\{\mathcal{I}\}$$

Deep Learning



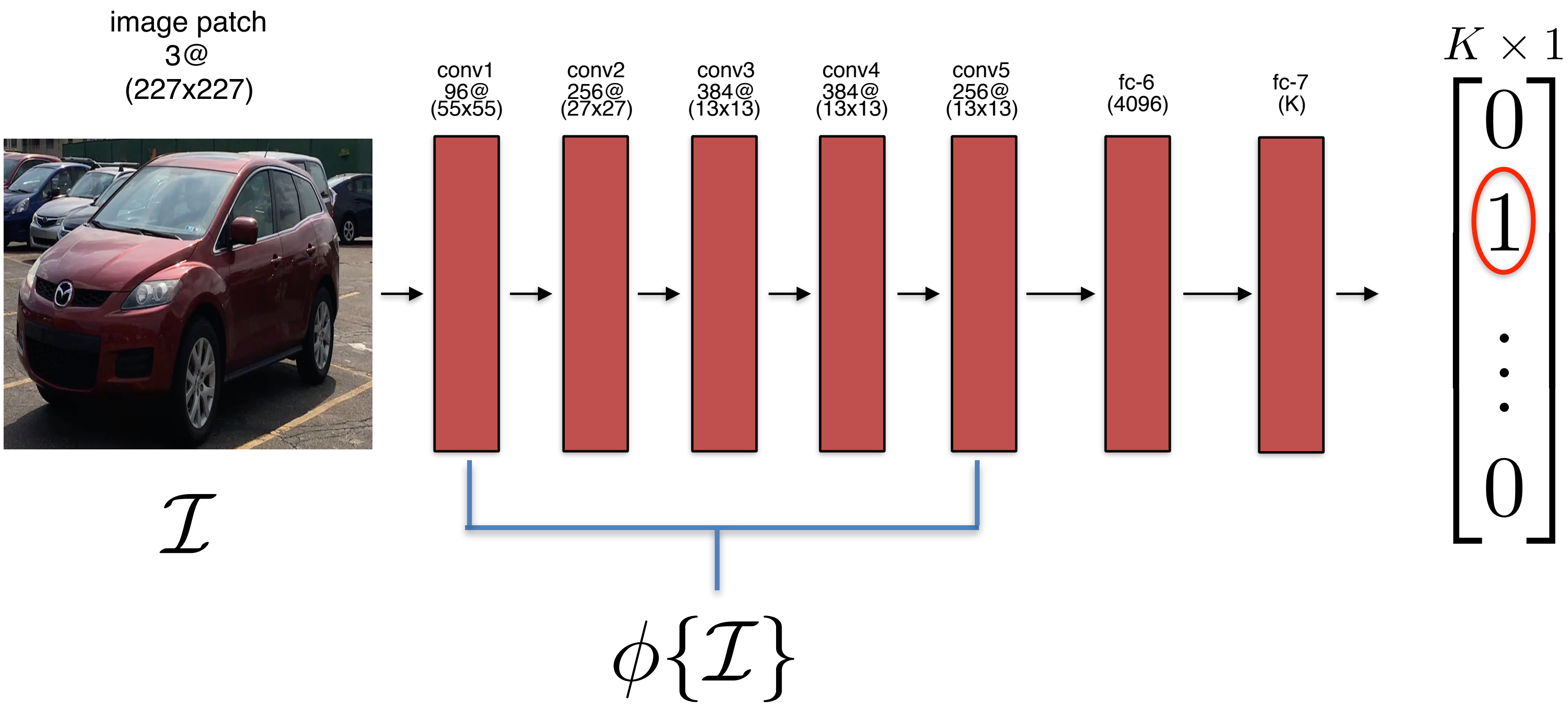
A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

Deep Learning



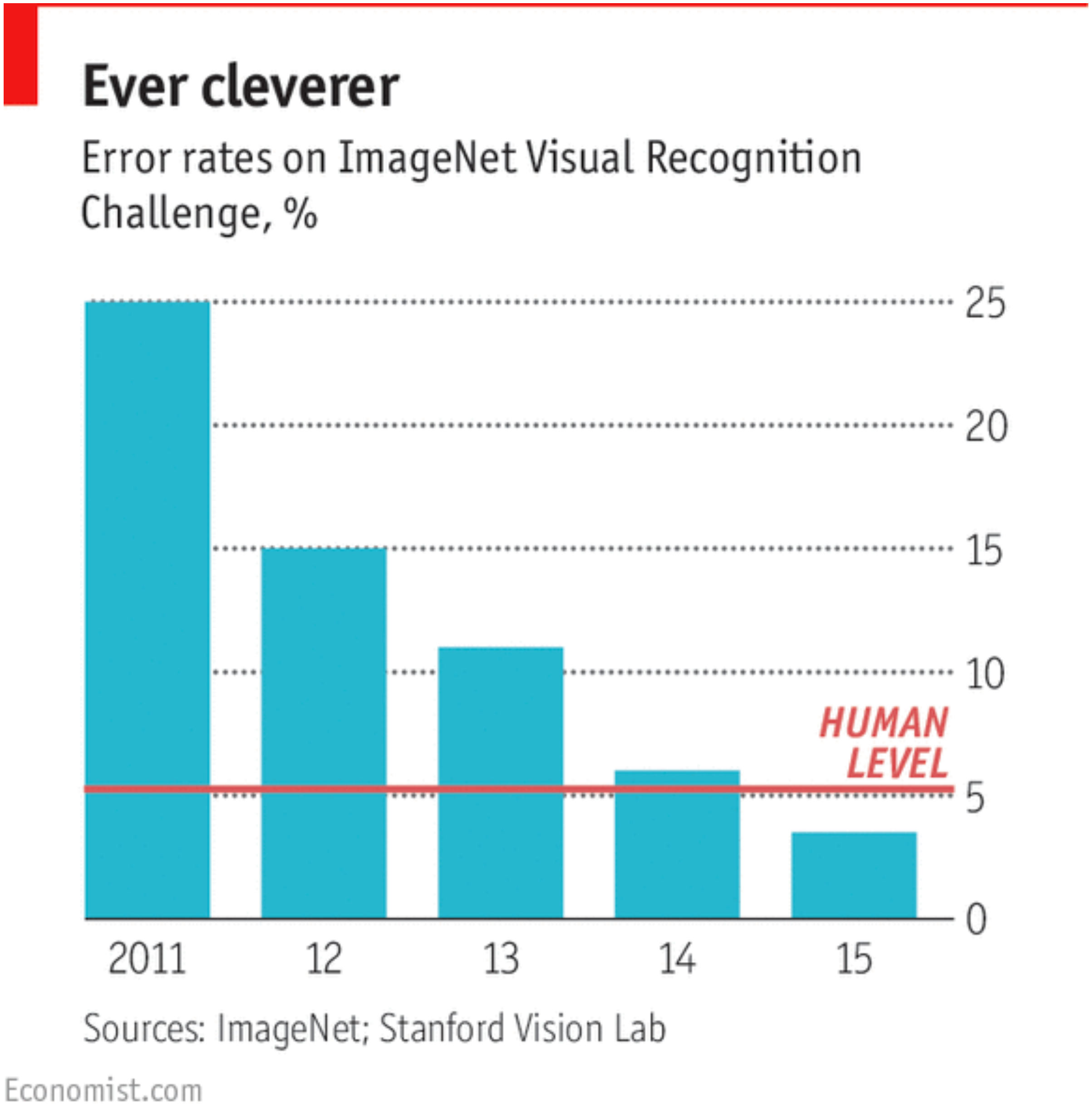
A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

Deep Learning

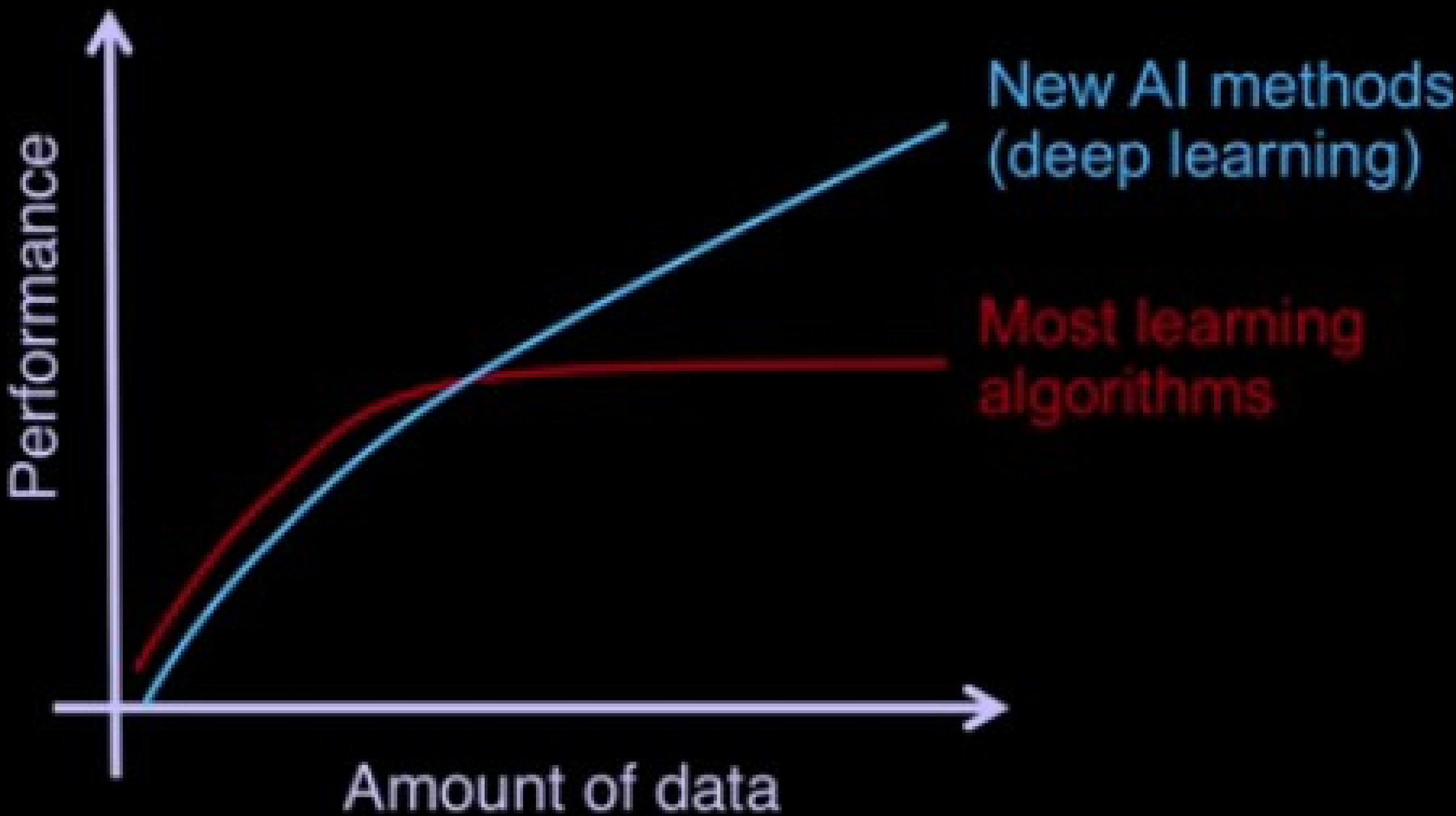


A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

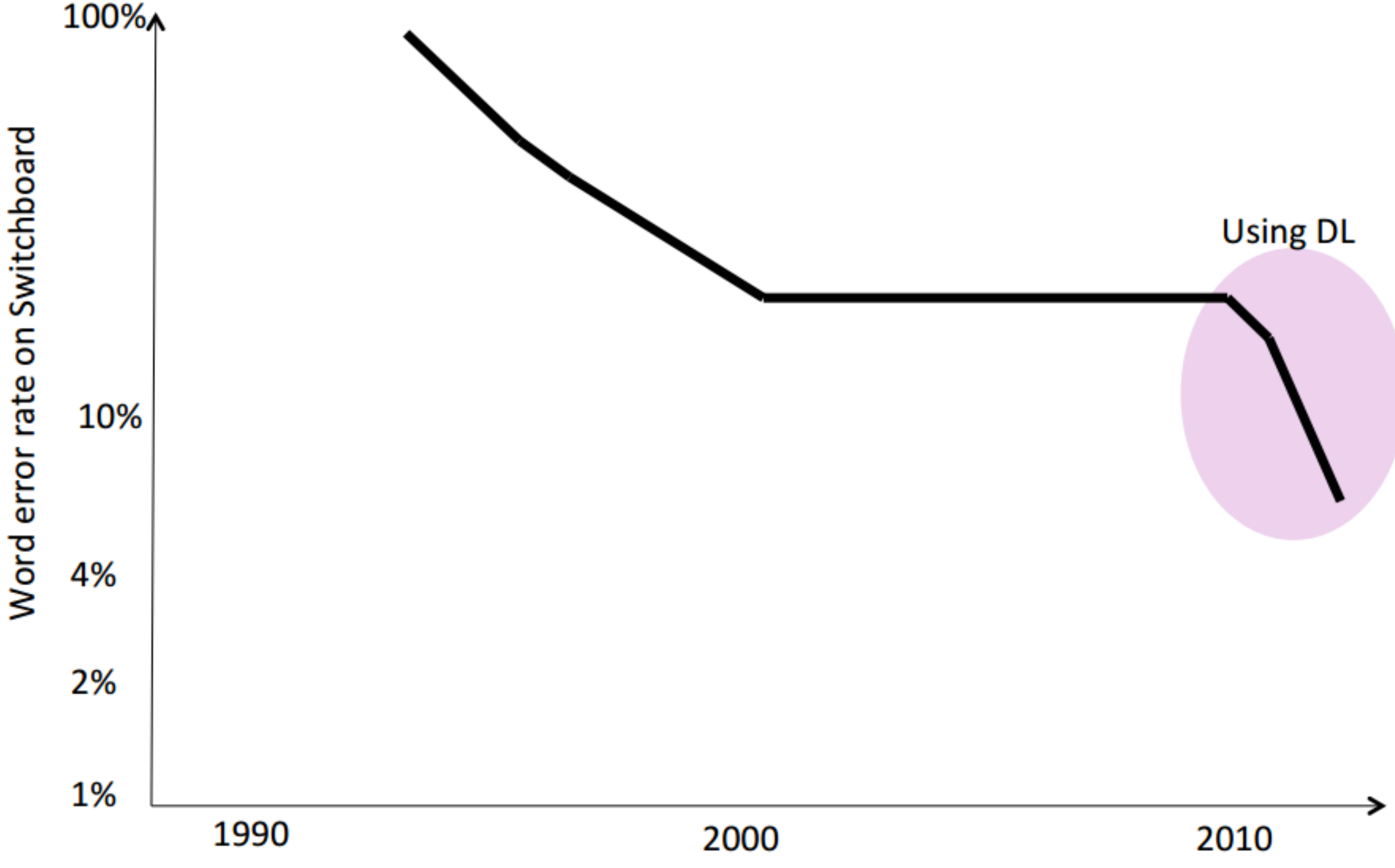
Deep Learning



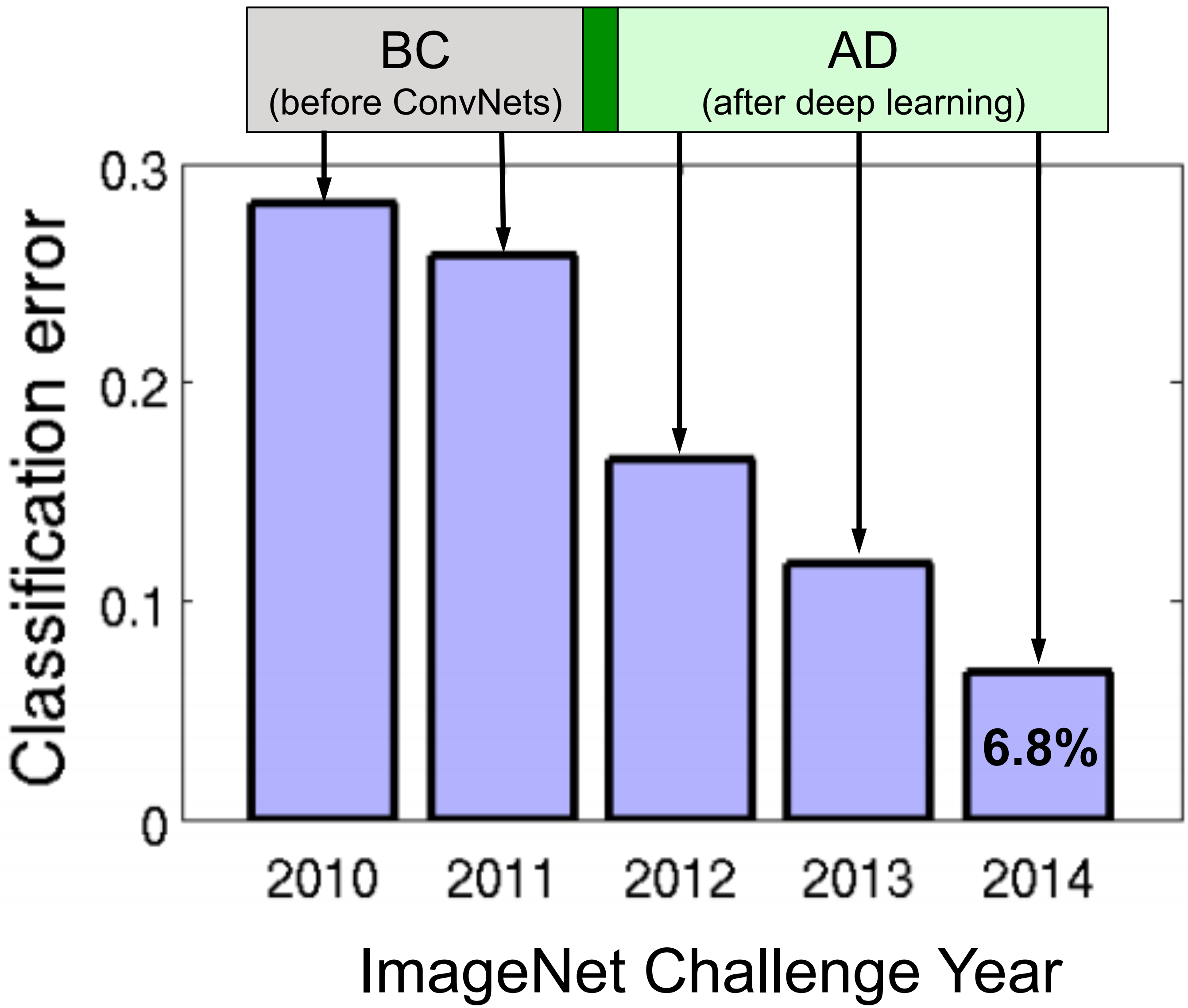
Data and machine learning



Impact on Speech Recognition



Impact on Object Recognition



Audio

TIMIT Phone classification	Accuracy
Prior art (Clarkson et al.,1999)	79.6%
Feature learning	**80.3%**

TIMIT Speaker identification	Accuracy
Prior art (Reynolds, 1995)	99.7%
Feature learning	**100.0%**

Images

CIFAR Object classification	Accuracy
Prior art (Ciresan et al., 2011)	80.5%
Feature learning	**82.0%**

NORB Object classification	Accuracy
Prior art (Scherer et al., 2010)	94.4%
Feature learning	**95.0%**

Video

Hollywood2 Classification	Accuracy
Prior art (Laptev et al., 2004)	48%
Feature learning	**53%**

YouTube	Accuracy
Prior art (Liu et al., 2009)	71.2%
Feature learning	**75.8%**

KTH	Accuracy
Prior art (Wang et al., 2010)	92.1%
Feature learning	**93.9%**

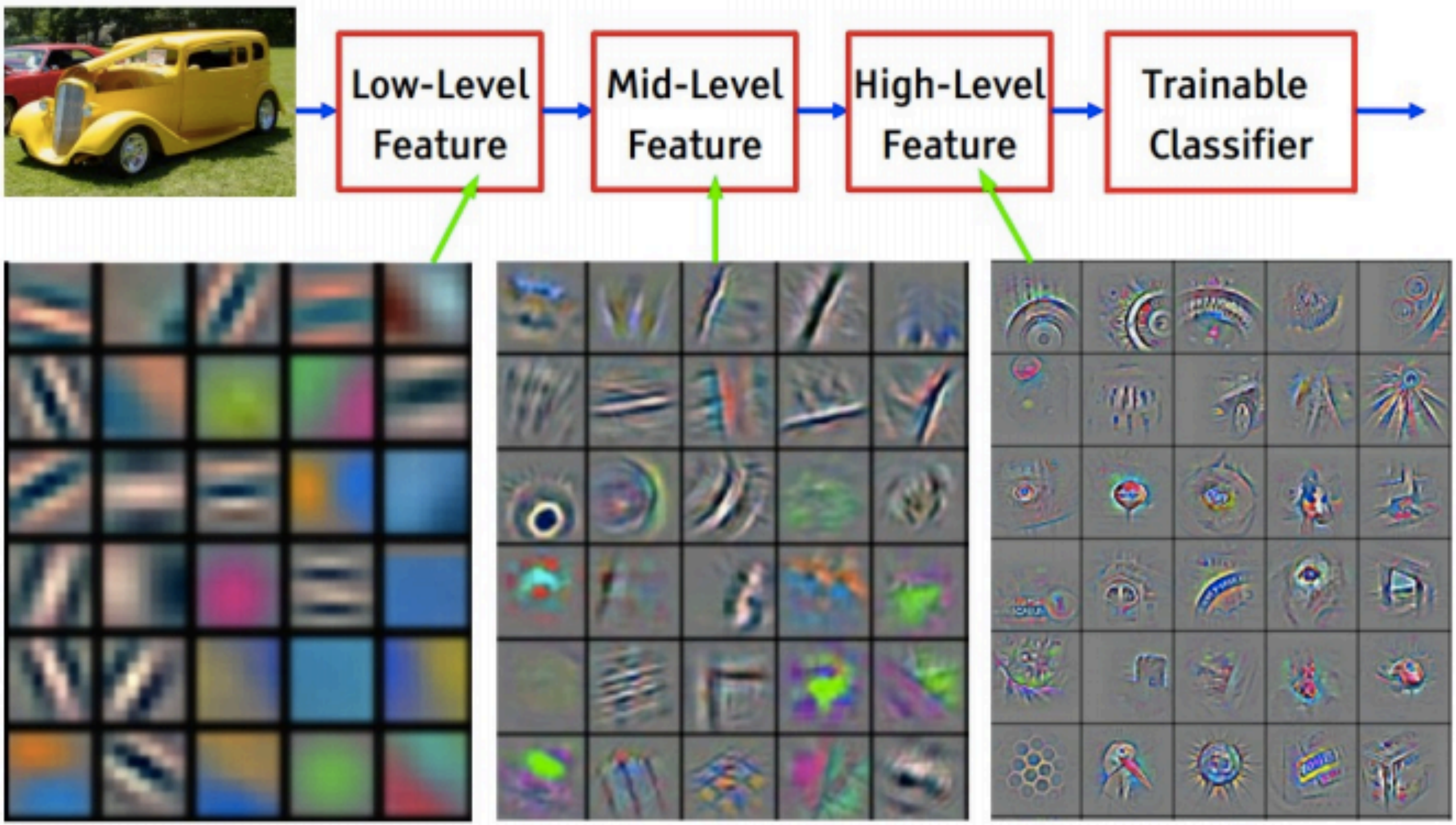
UCF	Accuracy
Prior art (Wang et al., 2010)	85.6%
Feature learning	**86.5%**

Text/NLP

Paraphrase detection	Accuracy
Prior art (Das & Smith, 2009)	76.1%
Feature learning	**76.4%**

Sentiment (MR/MPQA data)	Accuracy
Prior art (Nakagawa et al., 2010)	77.3%
Feature learning	**77.7%**

Visualizing CNNs



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

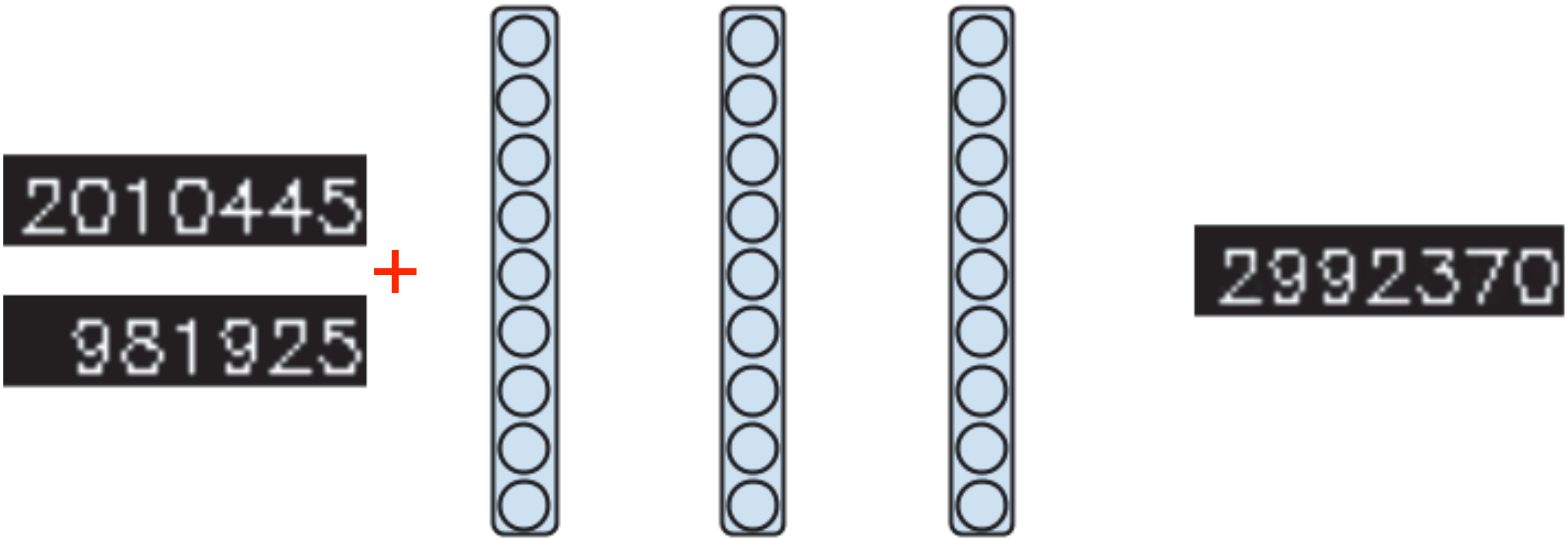
Mimicking $\neq$ Understanding

Deep Learning & Visual Arithmetic

+

Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic



Input Layer
Nodes: 60X15
Depth: 2

HL1
256 nodes
Fully Con.
ReLU

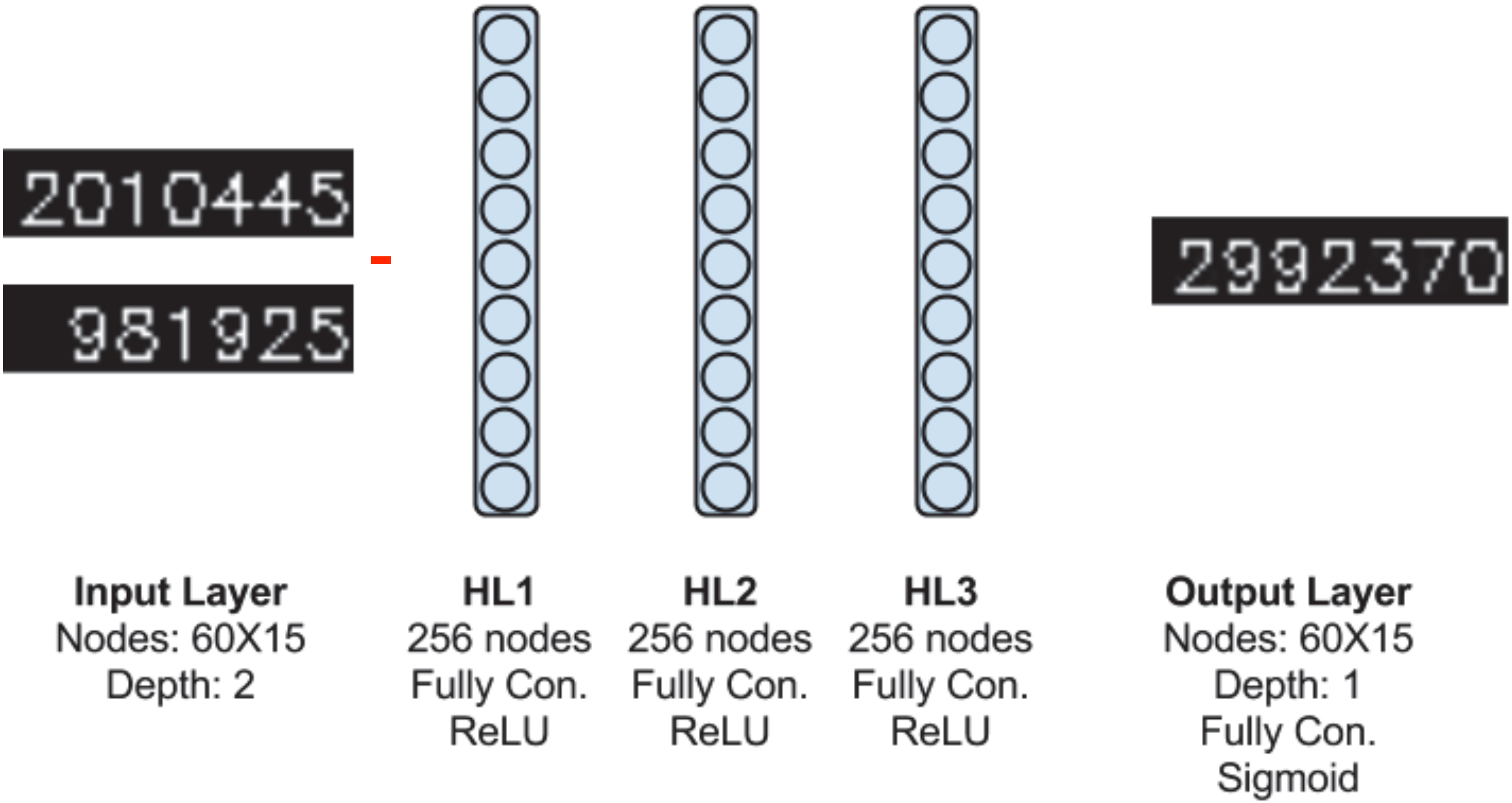
HL2
256 nodes
Fully Con.
ReLU

HL3
256 nodes
Fully Con.
ReLU

Output Layer
Nodes: 60X15
Depth: 1
Fully Con.
Sigmoid

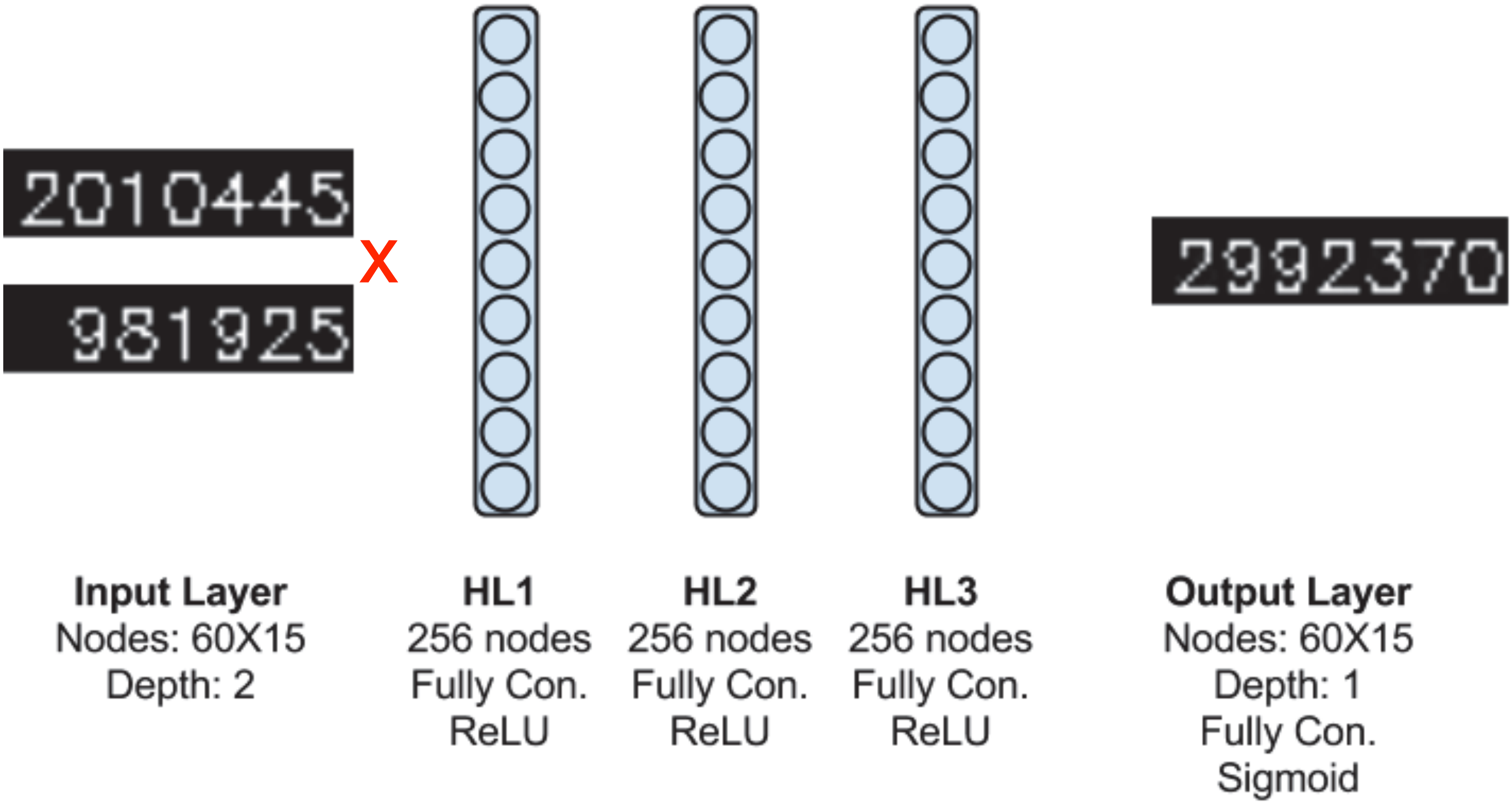
Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic



Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic



Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic

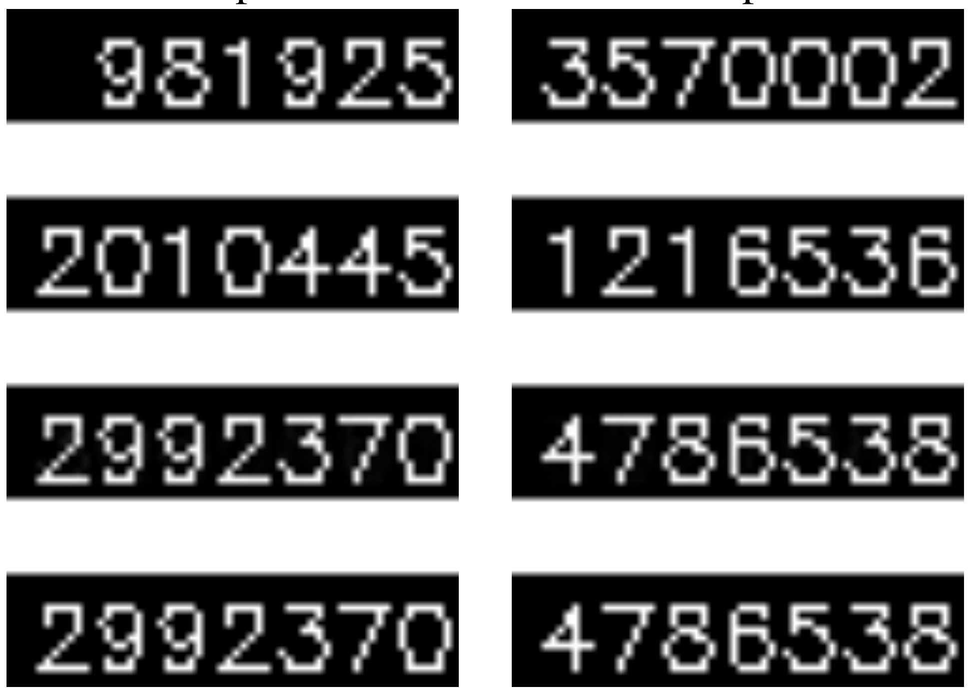
Example A Example B

Input Picture 1

Input Picture 2

Network Output Picture

Ground Truth Picture



Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic

	Example A	Example B	Failure Example
Input Picture 1	981925	3570002	3750668
Input Picture 2	2010445	1216536	3643197
Network Output Picture	2992370	4786538	7393855
Ground Truth Picture	2992370	4786538	7393865

Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Deep Learning & Visual Arithmetic

Operation	Pictures		1-hot Vectors	
	No. Layers	% Error	No. Layers	% Error
Add	3	1.9%	1	1.7%
Subtract	3	3.2%	1	2.1%
Multiply	5	71.5%	3	37.6%
Roman Addition	5	74.3 %	3	0.7 %

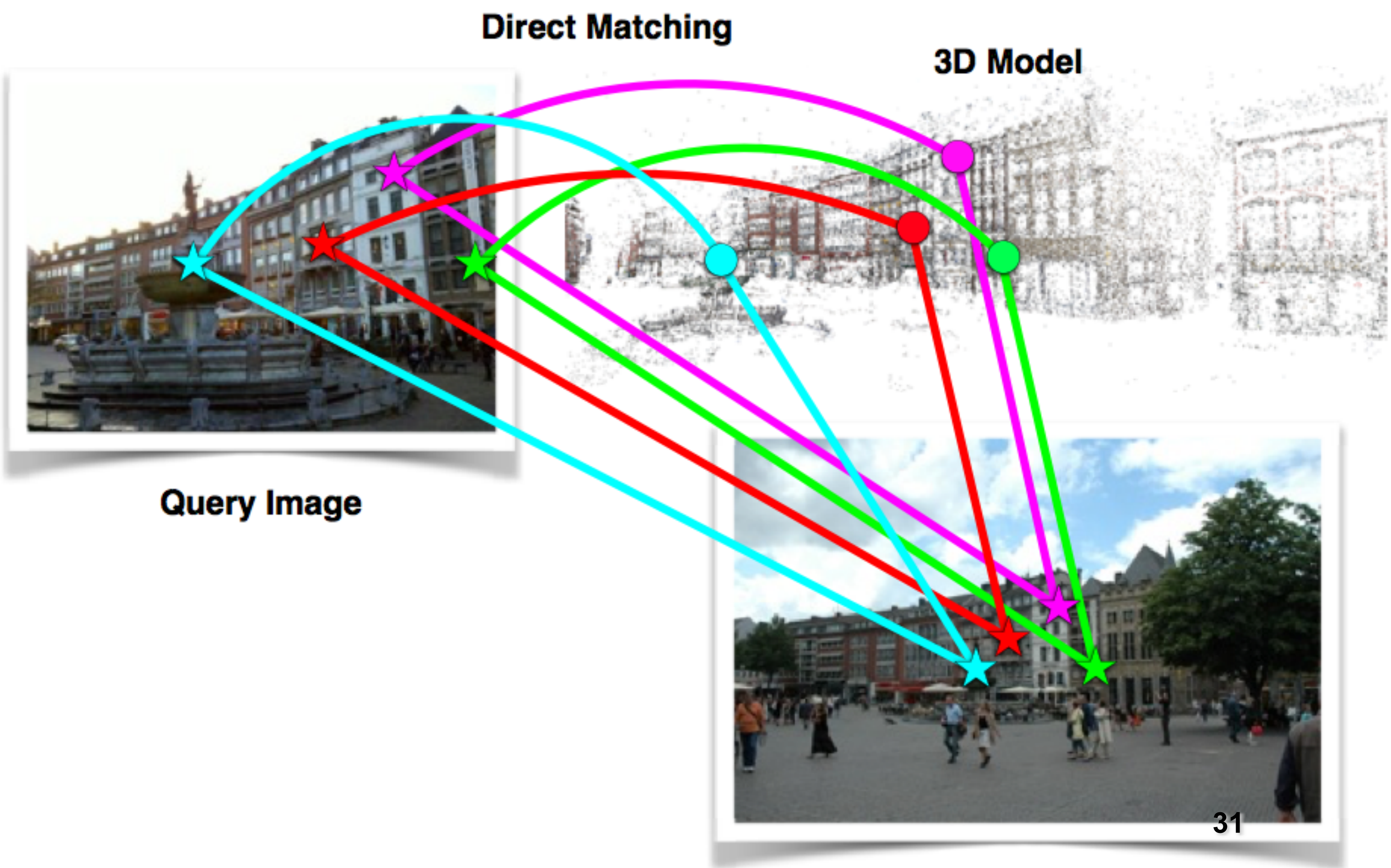
Y. Hoshen and S. Peleg, Visual Learning of Arithmetic Operations, AAAI 2016.

Today

- 16-720B Logistics
- Computer Vision - The Science
- Computer Vision - Applications & Trends

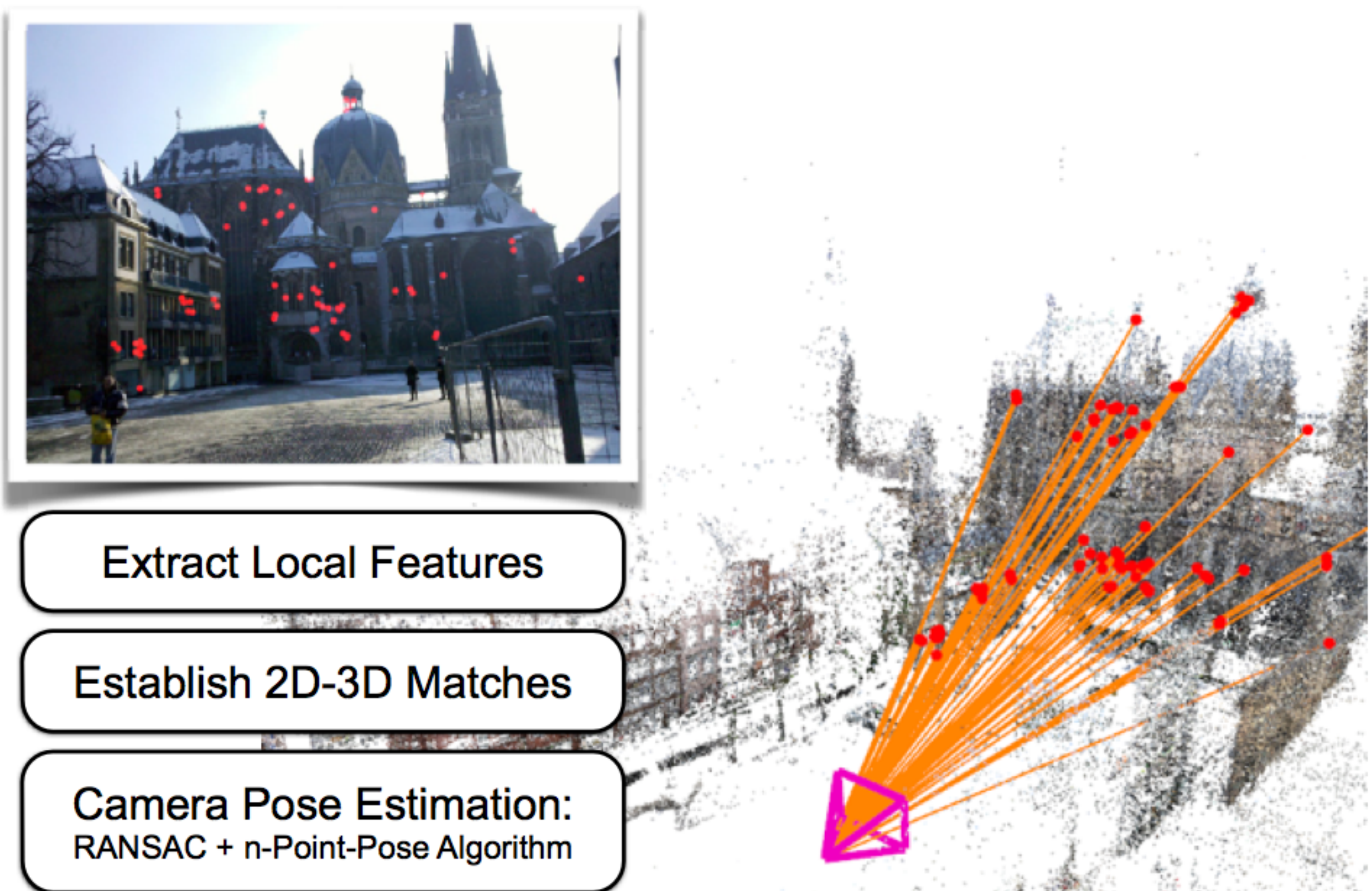
Trend - Visual SLAM

Simultaneous Localization and Mapping

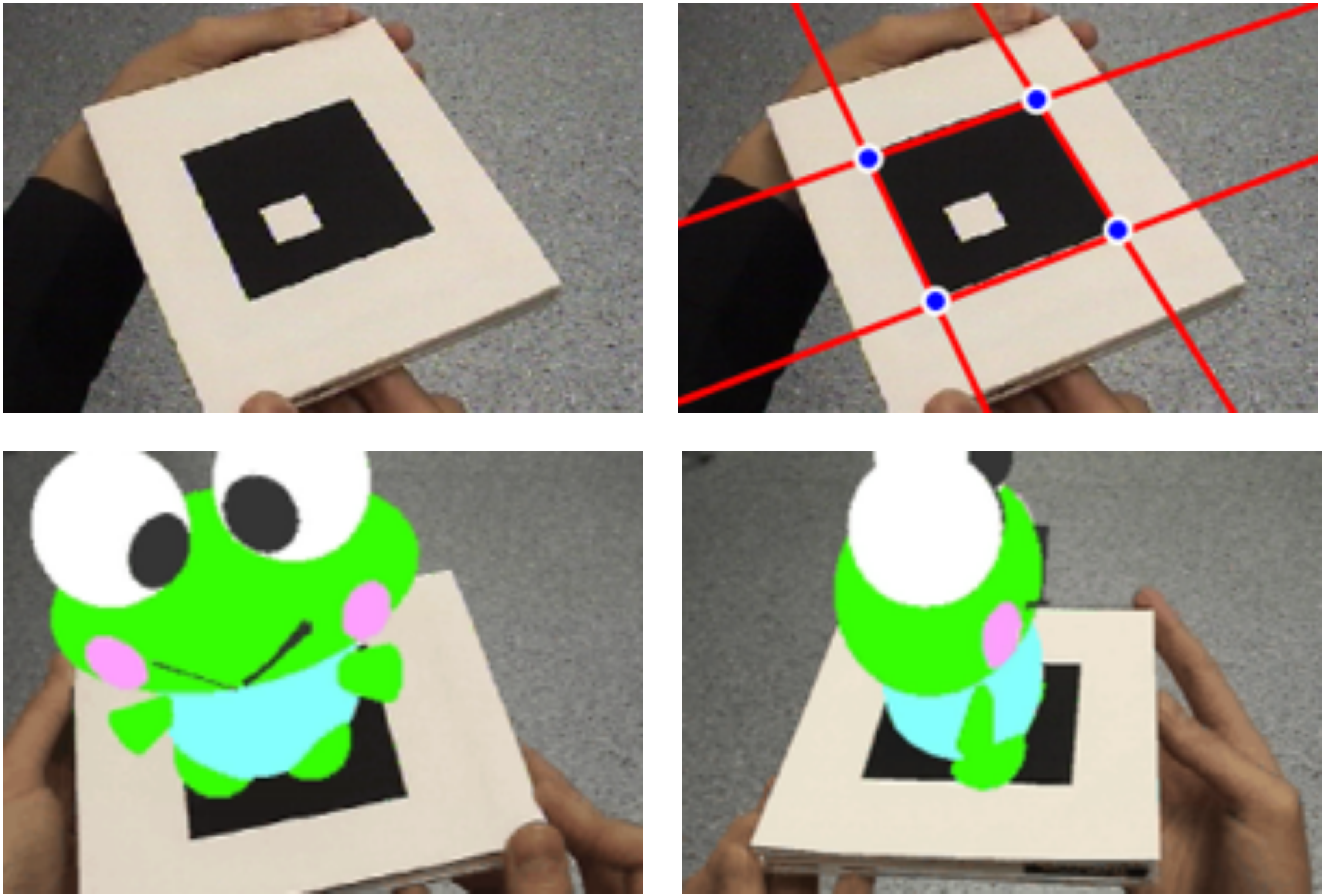


Visual SLAM

Attempts to reconstruct the 3D world & camera position in real-time.

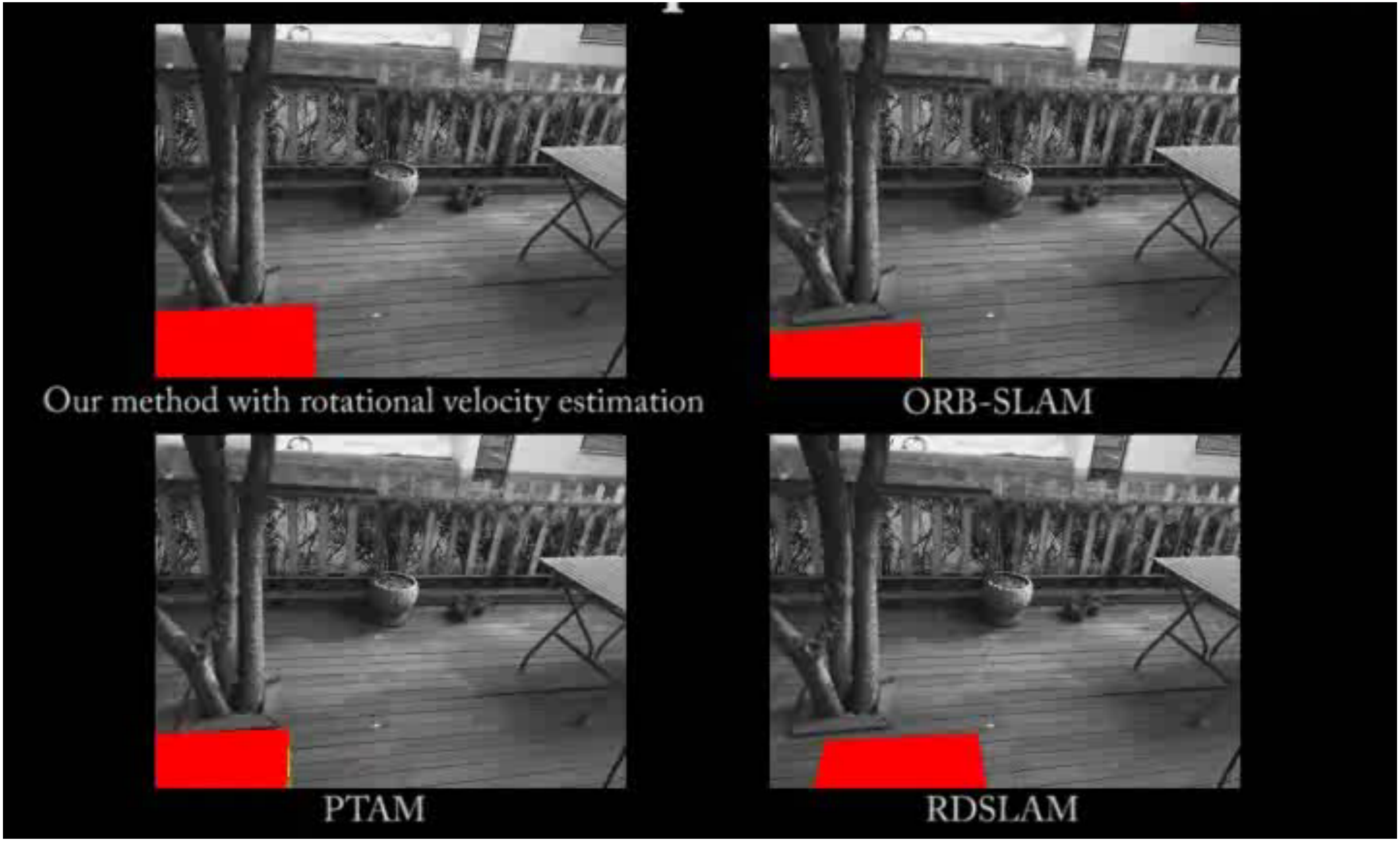


Motivation - Augmented Reality



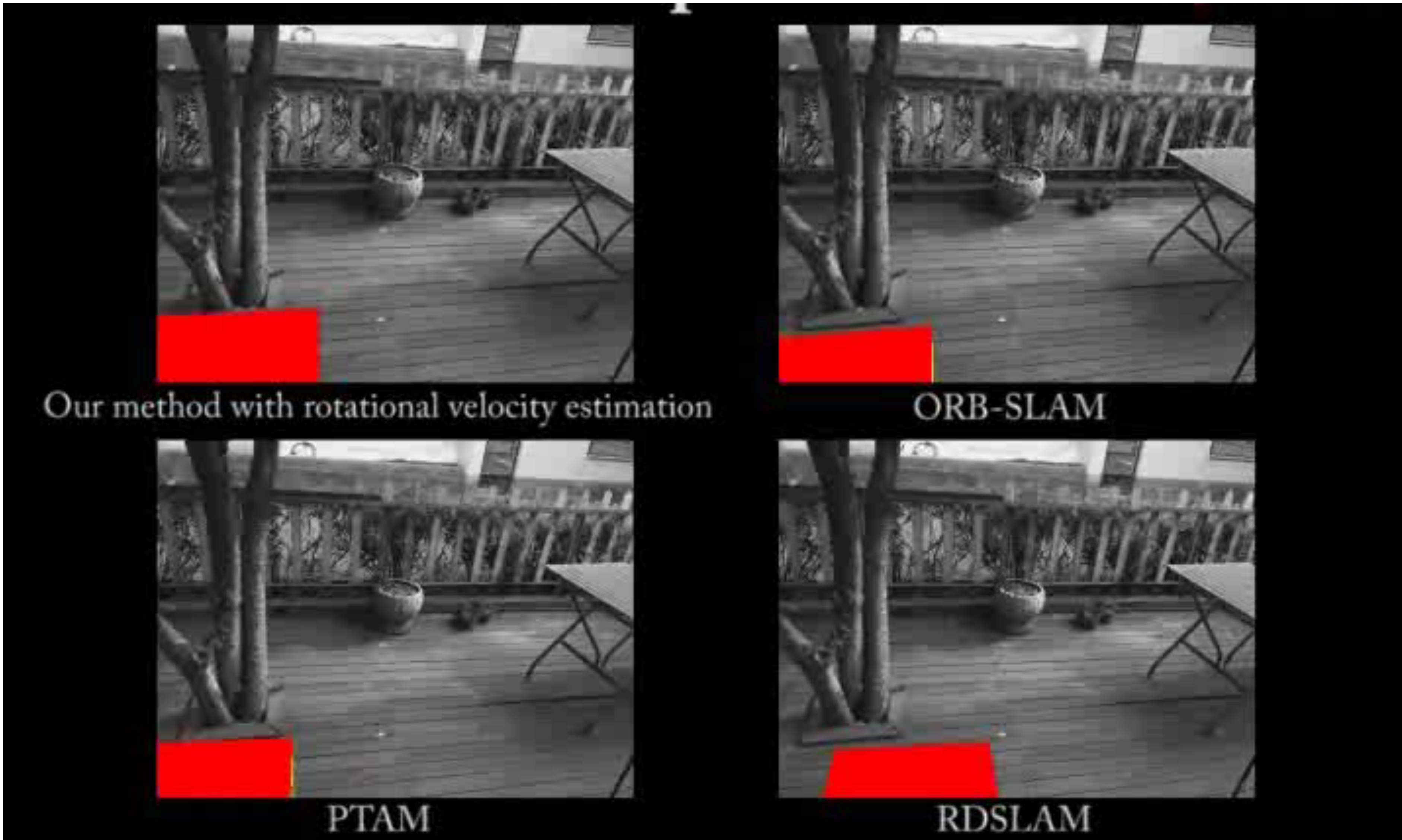
Adapted from: Computer vision: models, learning and inference. Simon J.D. Prince

Examples of SLAM for AR



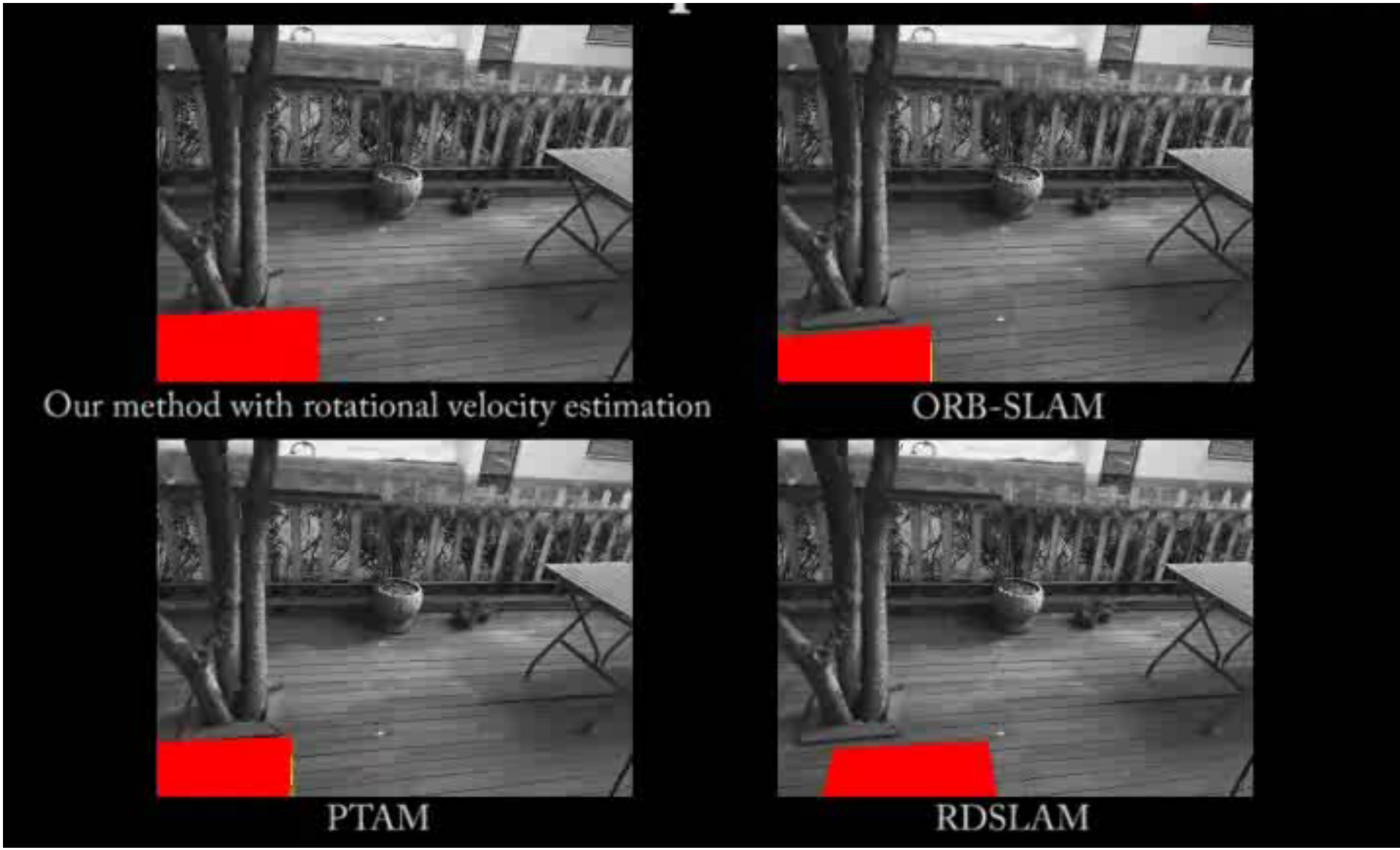
Taken from: H. Liu et al. “Robust Keyframe-based Monocular SLAM for Augmented Reality”, ISMAR 2016.

Examples of SLAM for AR



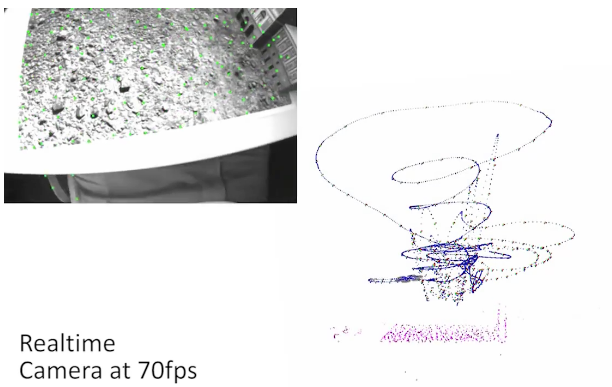
Taken from: H. Liu et al. “Robust Keyframe-based Monocular SLAM for Augmented Reality”, ISMAR 2016.

Examples of SLAM for AR

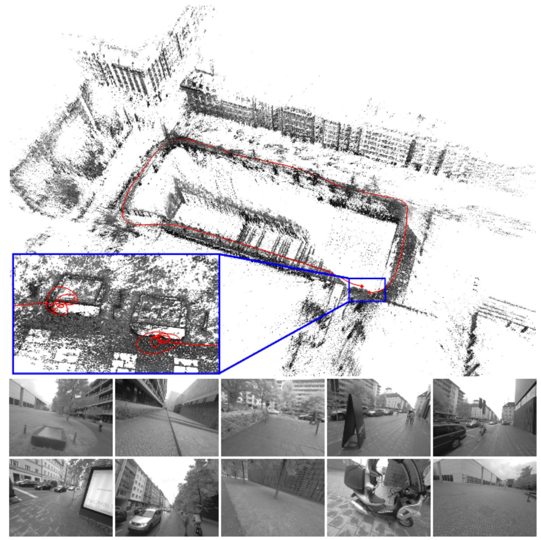


Taken from: H. Liu et al. “Robust Keyframe-based Monocular SLAM for Augmented Reality”, ISMAR 2016.

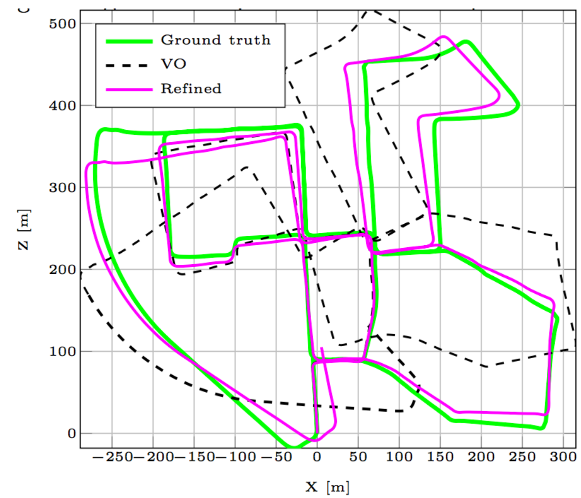
Direct SLAM



SVO (2014) [1]
Tracking at 300FPS.

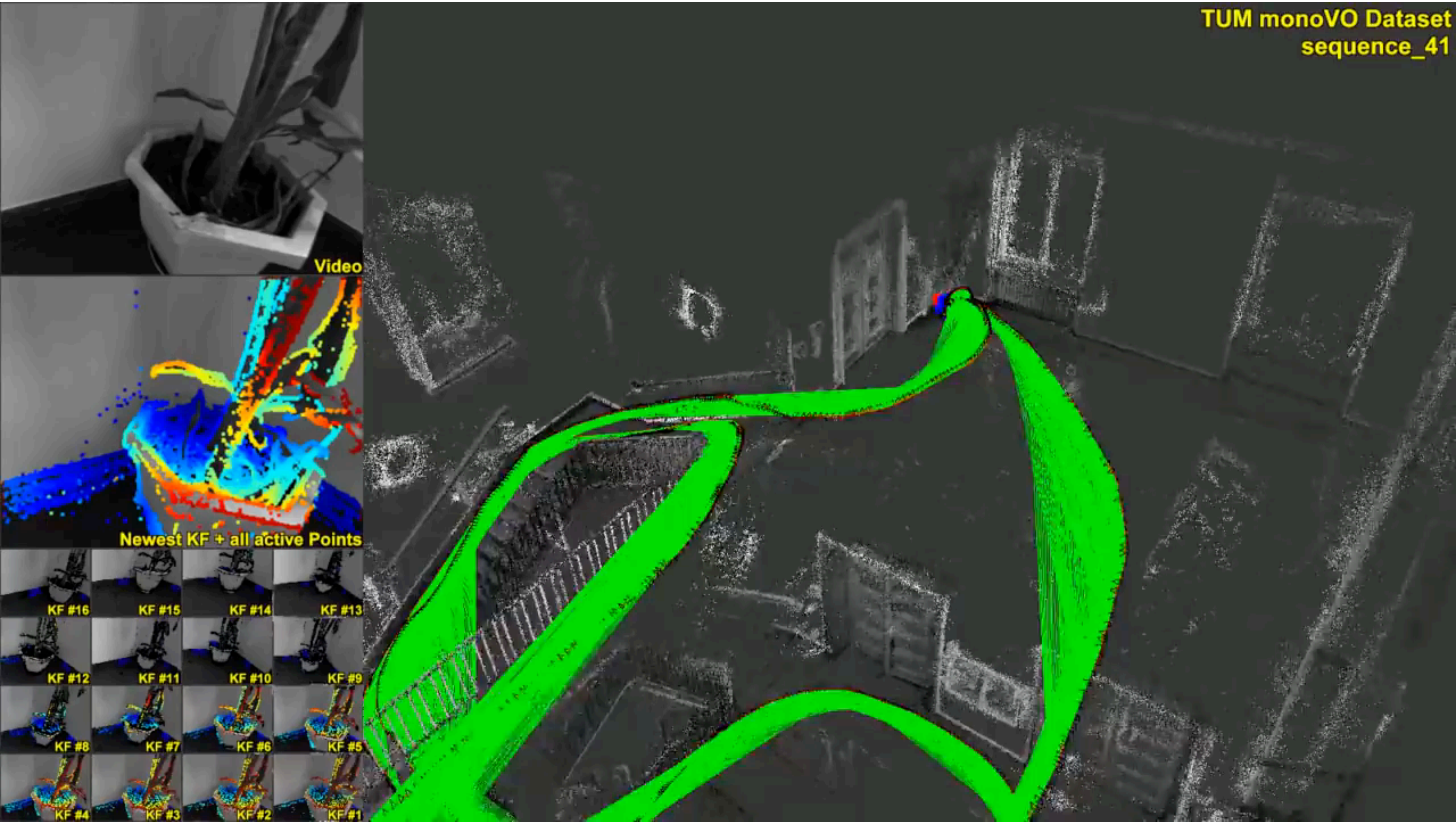


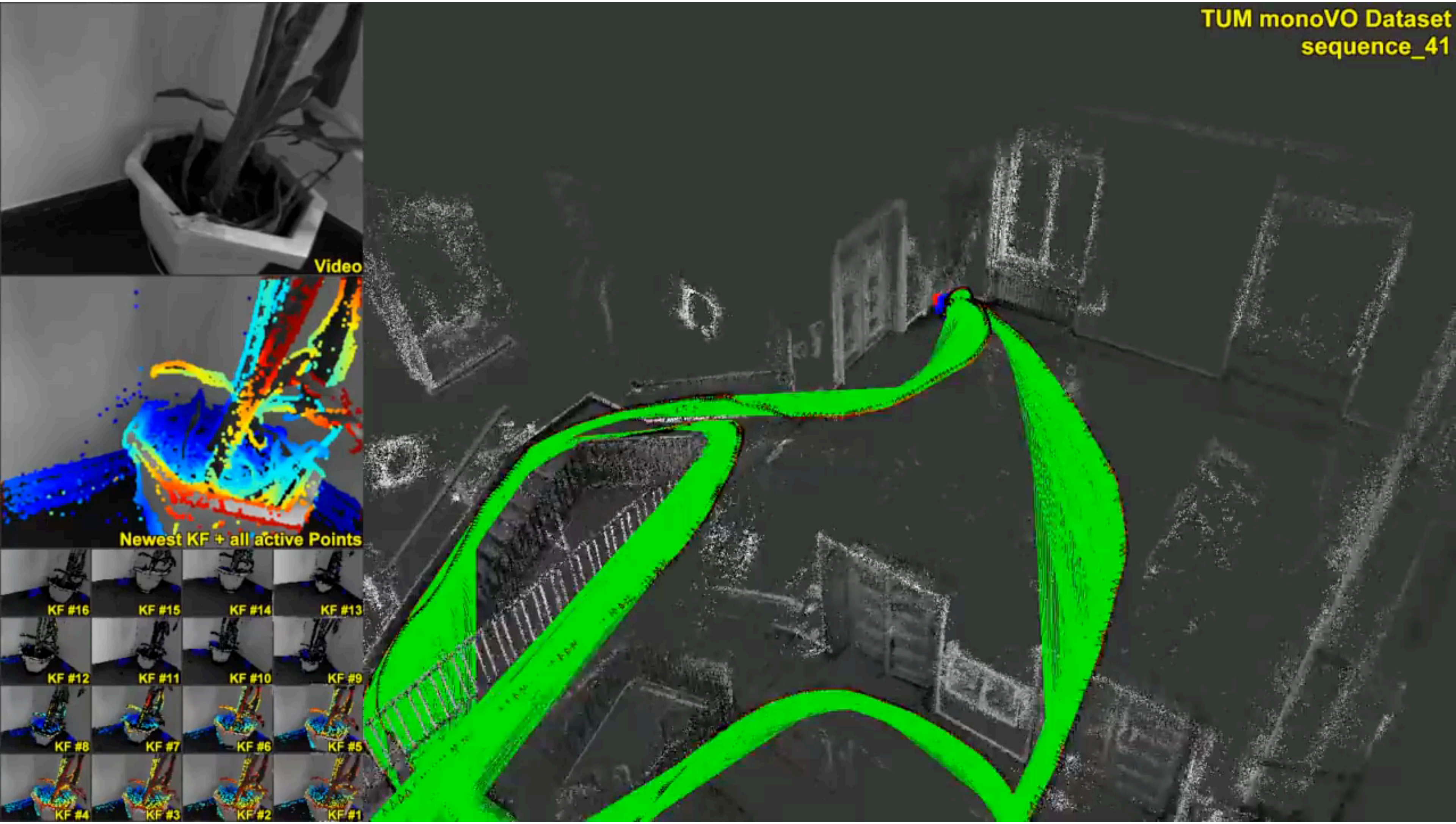
DSO (2017) [2]
Large exposure changes.
Direct bundle adjustment.



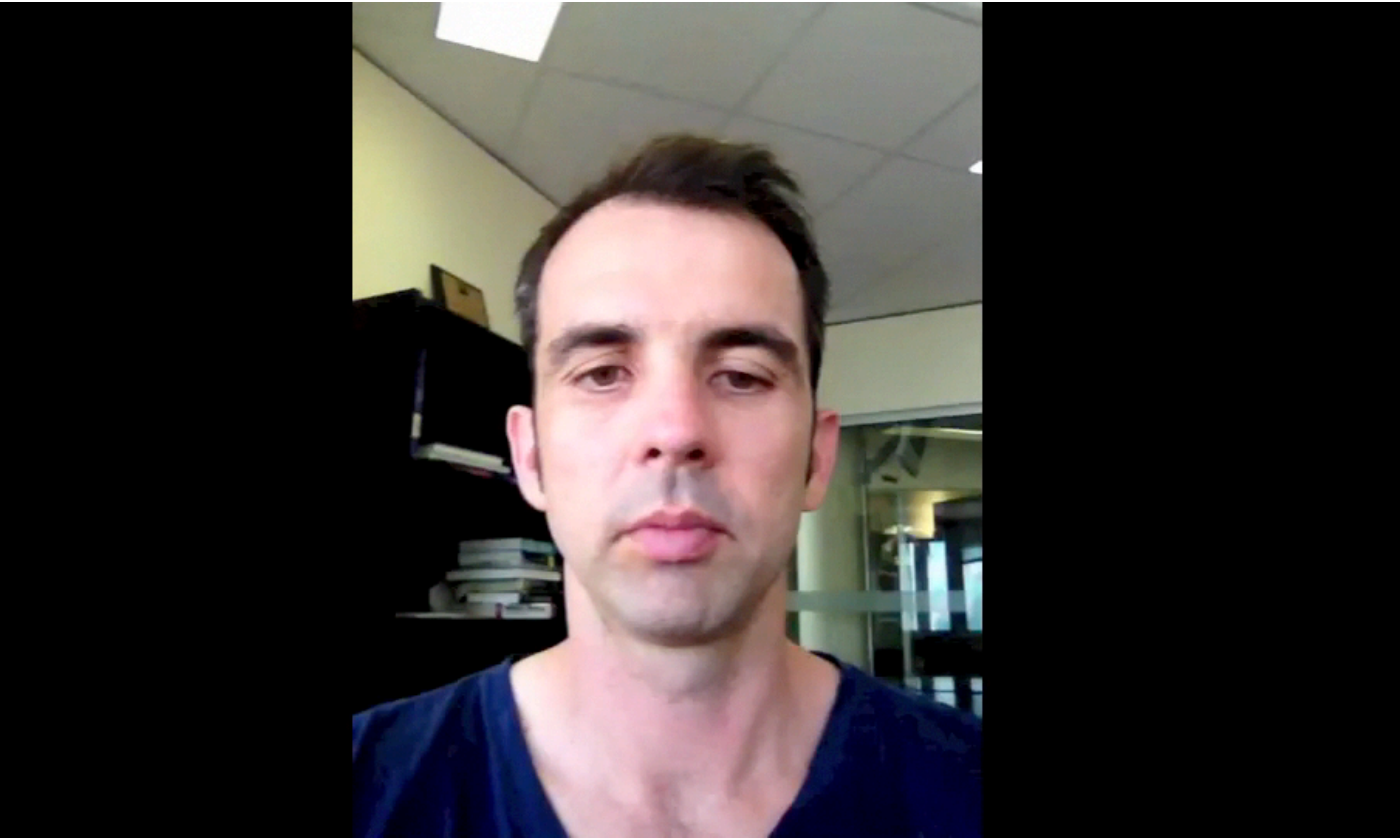
Alismail et al. (2016) [3]
Direct bundle adjustment

[1] C. Forster, M. Pizzoli, D. Scaramuzza, SVO: Fast Semi-Direct Monocular Visual Odometry, ICRA, Hong Kong, 2014.
[2] Engel, J., Koltun, V. and Cremers, D., 2017. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
[3] Alismail, H., Browning, B. and Lucey, S., 2016. Photometric Bundle Adjustment for Vision-Based SLAM. *ACCV 2017*.


TUM monoVO Dataset
sequence_41
Video
Newest KF + all active Points
KF #16
KF #15
KF #14
KF #13
KF #12
KF #11
KF #10
KF #9
KF #8
KF #7
KF #6
KF #5
KF #4
KF #3
KF #2
KF #1



TUM monoVO Dataset
sequence_41
Video
Newest KF + all active Points
KF #16
KF #15
KF #14
KF #13
KF #12
KF #11
KF #10
KF #9
KF #8
KF #7
KF #6
KF #5
KF #4
KF #3
KF #2
KF #1



"Without question, the glasses.com app represents the state of the art in virtual style shopping. It's superb technology that really works and could save a lot of people a lot o embarrassment."

- DAVID POGUE, THE NEW YORK TIMES

"The Best: Try on as much as you want without actually having to try it on."

- ASHLEY FEINBERG, GIZMODO

"You don't need to have 20/20 vision to see the appeal of the new and free glasses.com app for iPad."

- MARC SALTZMAN, USA TODAY

"Hate Trying On Glasses? A New App Ends The Misery."

- BRUCE UPBIN, FORBES

"Thanks to a new iPad app by Glasses.com, dragging a friend to the store may be a thing of the past."

- BRIGITT HAUCK, REALSIMPLE.COM


Demo
cnet

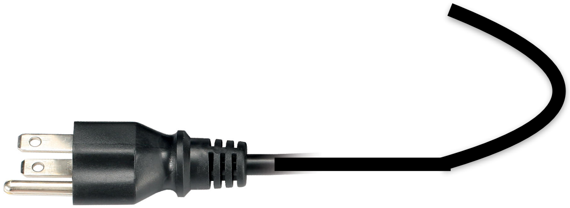

Demo

Trend - Mobile Computer Vision

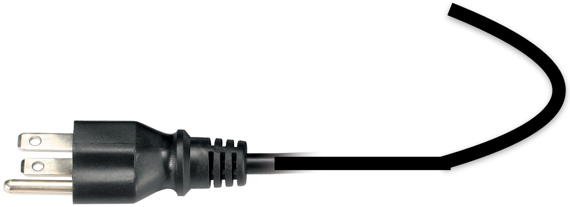
Trend - Mobile Computer Vision

Why is Mobile CV Different?

Why is Mobile CV Different?

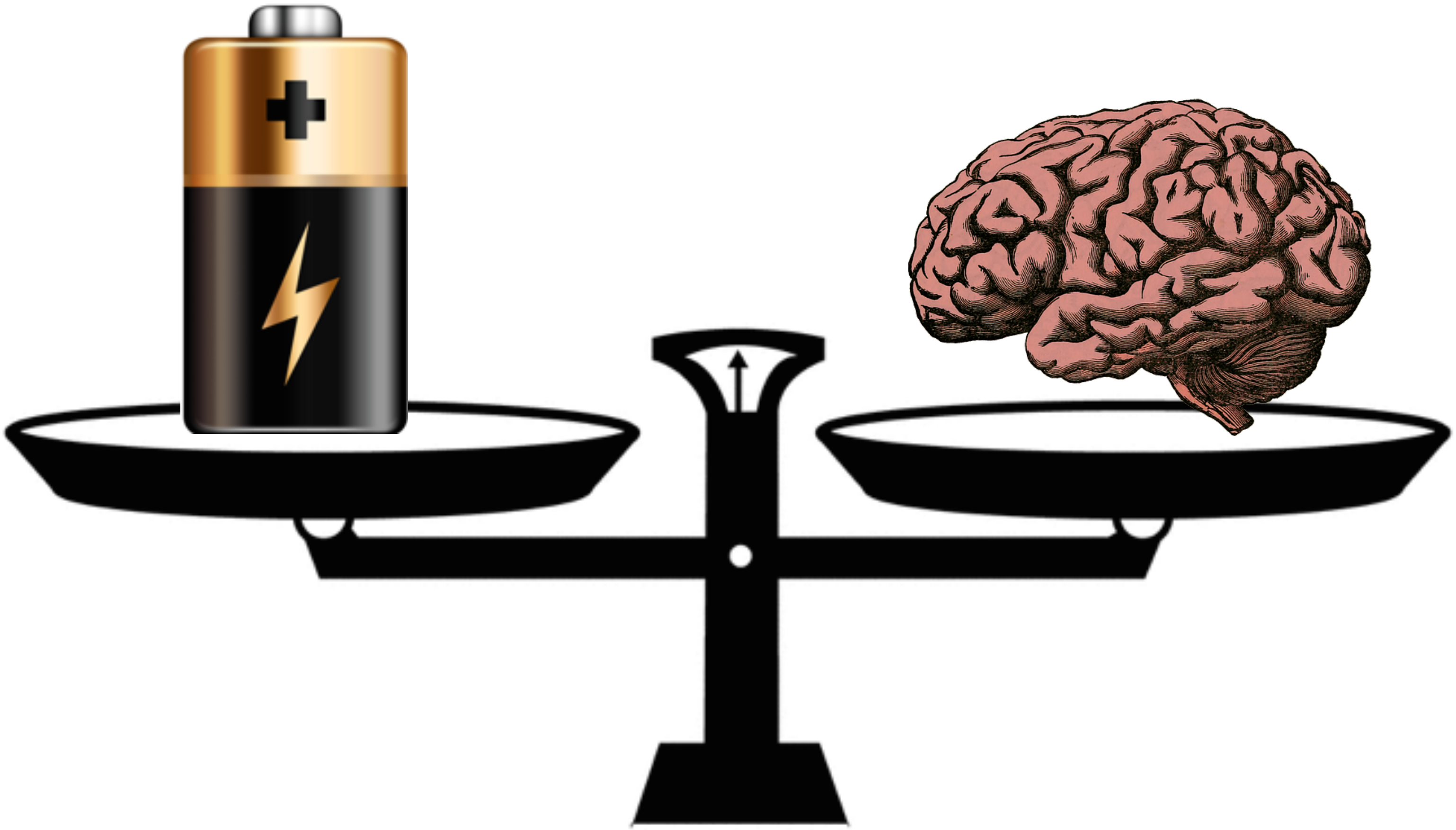


Why is Mobile CV Different?

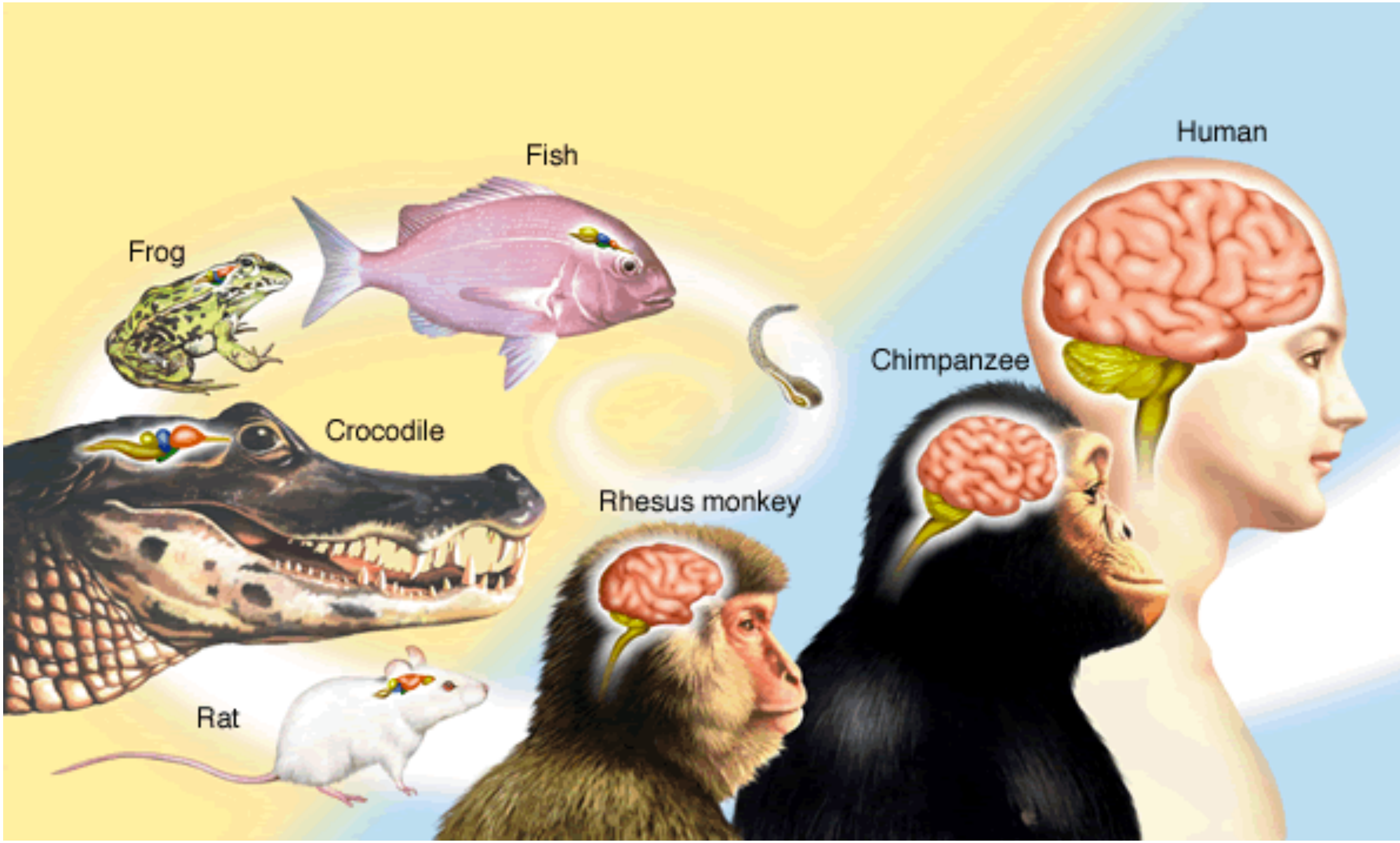


Connection to Robotics…..

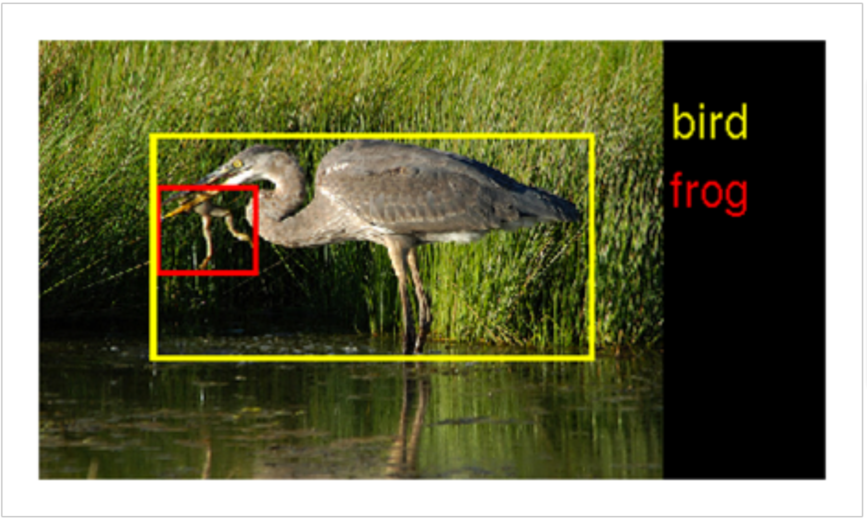
Balancing Power versus Perception

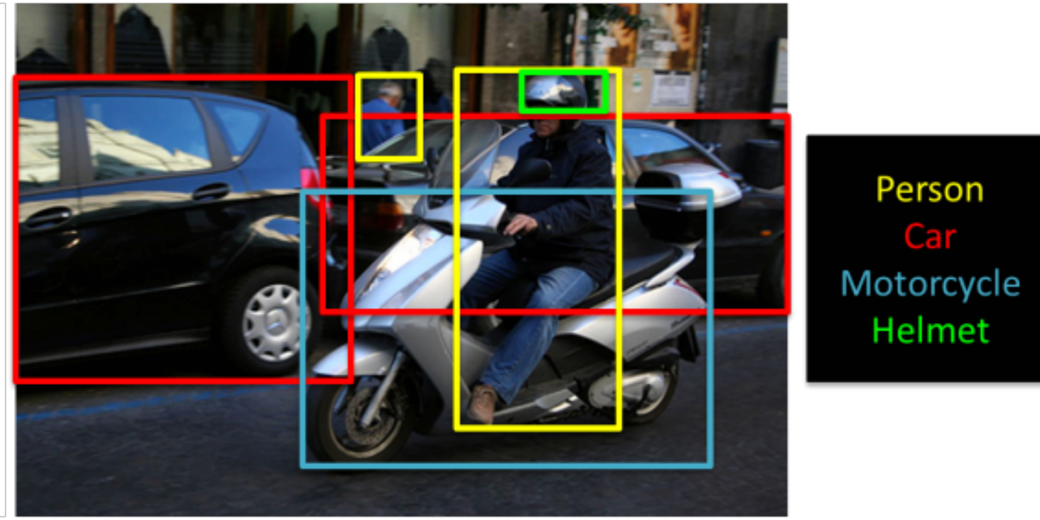


Balancing Power versus Perception



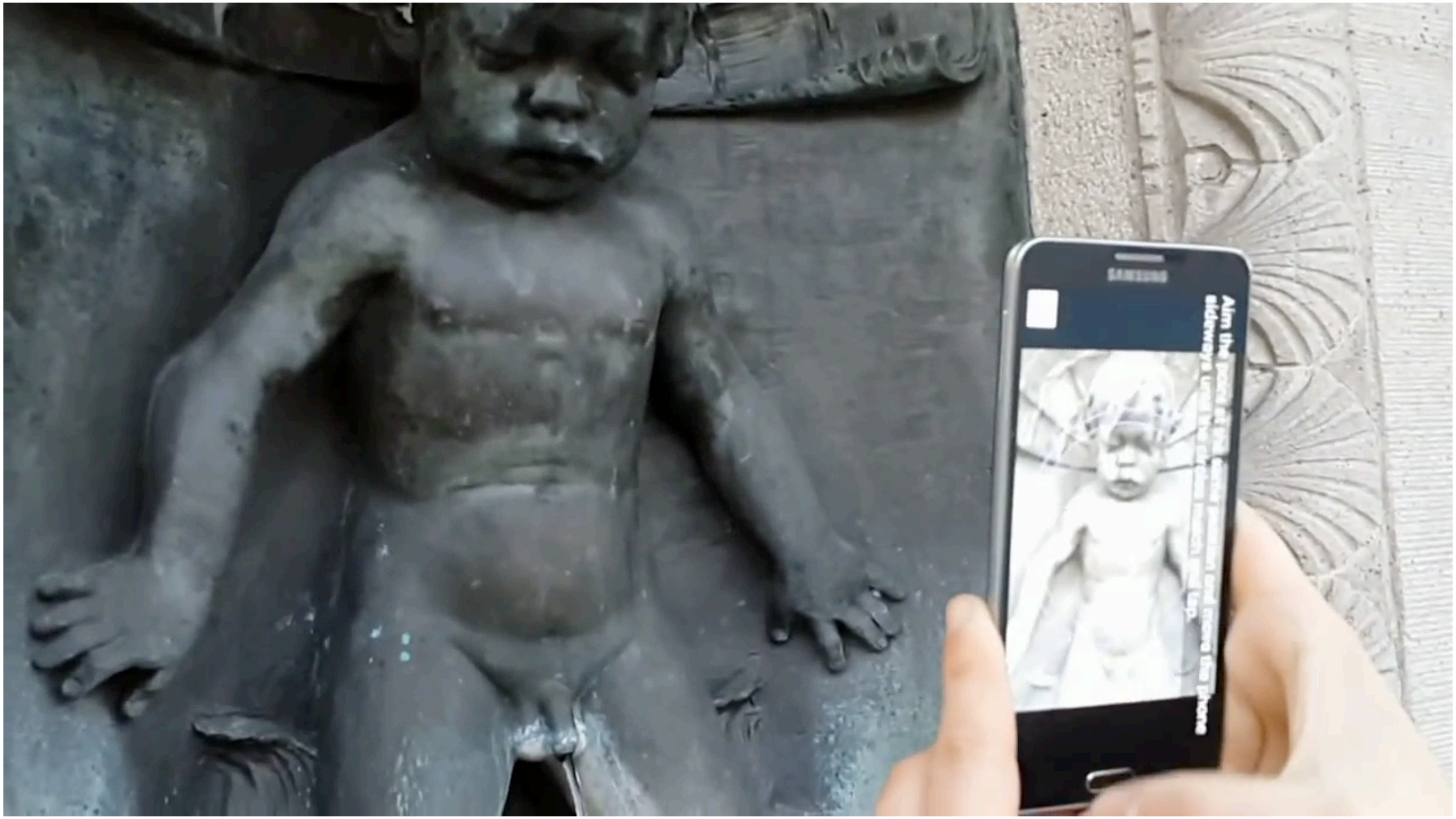
Low Power Image Recognition Challenge





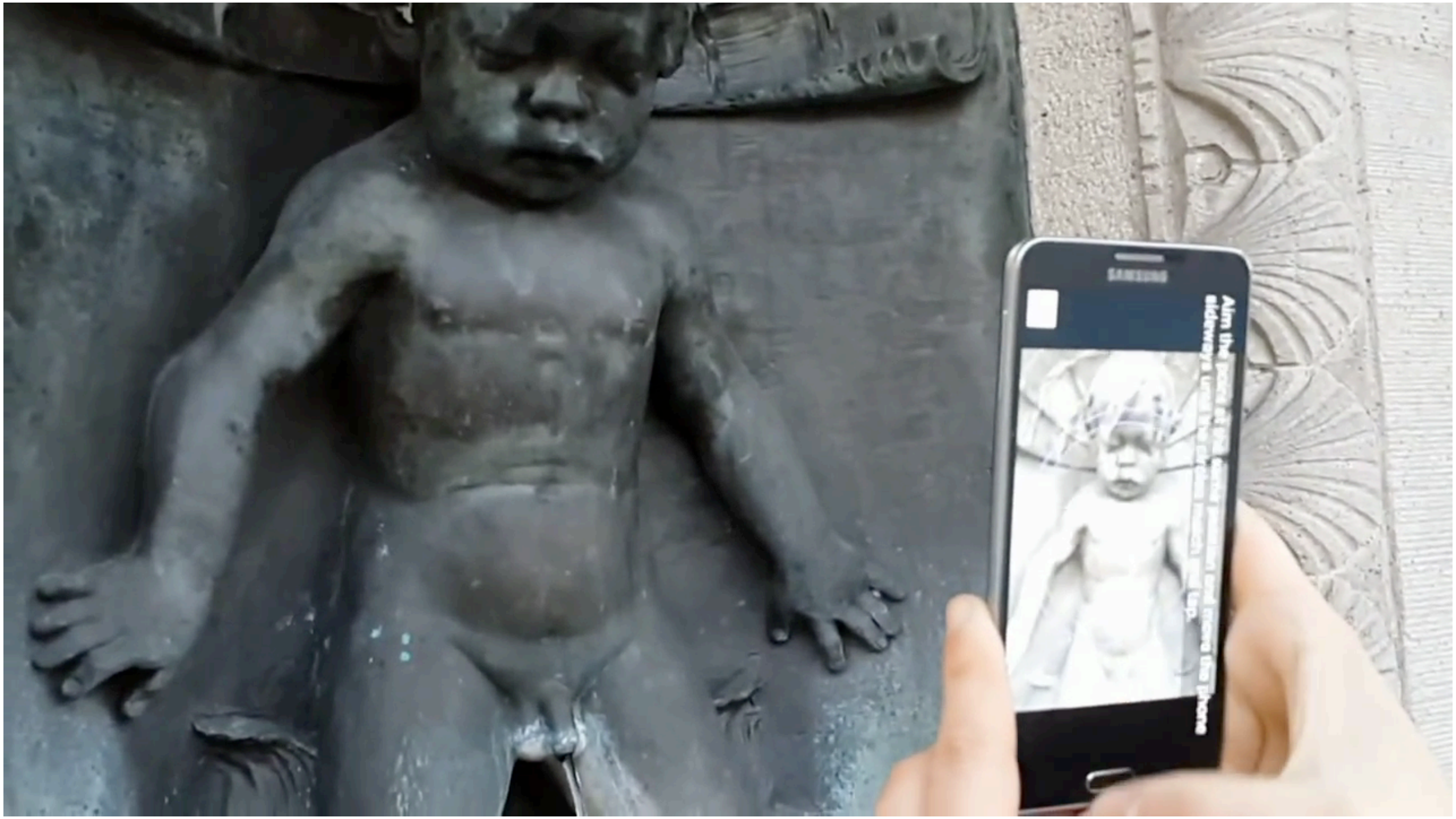
Taken from: http://lpirc.net/

Mobile Visual SLAM + IMU



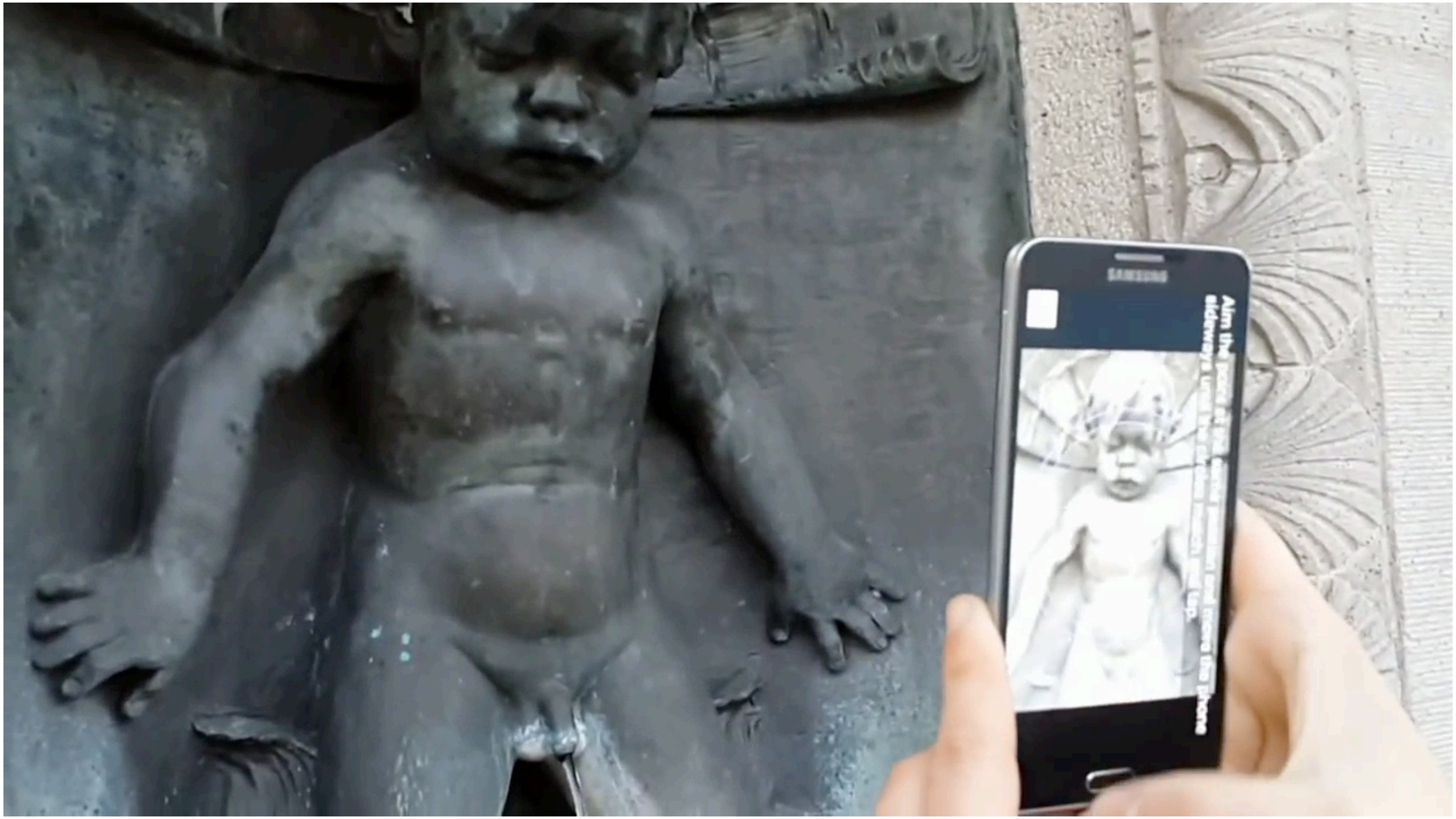
P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, M. Pollefeys : Live metric 3d reconstruction on mobile phones. (ICCV 2013)

Mobile Visual SLAM + IMU



P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, M. Pollefeys : Live metric 3d reconstruction on mobile phones. (ICCV 2013)

Mobile Visual SLAM + IMU



P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, M. Pollefeys : Live metric 3d reconstruction on mobile phones. (ICCV 2013)

Trend - High Speed Cameras

Trend - High Speed Cameras

Trend - High Speed Cameras

Trend - High Speed Cameras

Sony Xperia Phone - 960 FPS - https://www.facebook.com/bbcnews/videos/662989307239621/

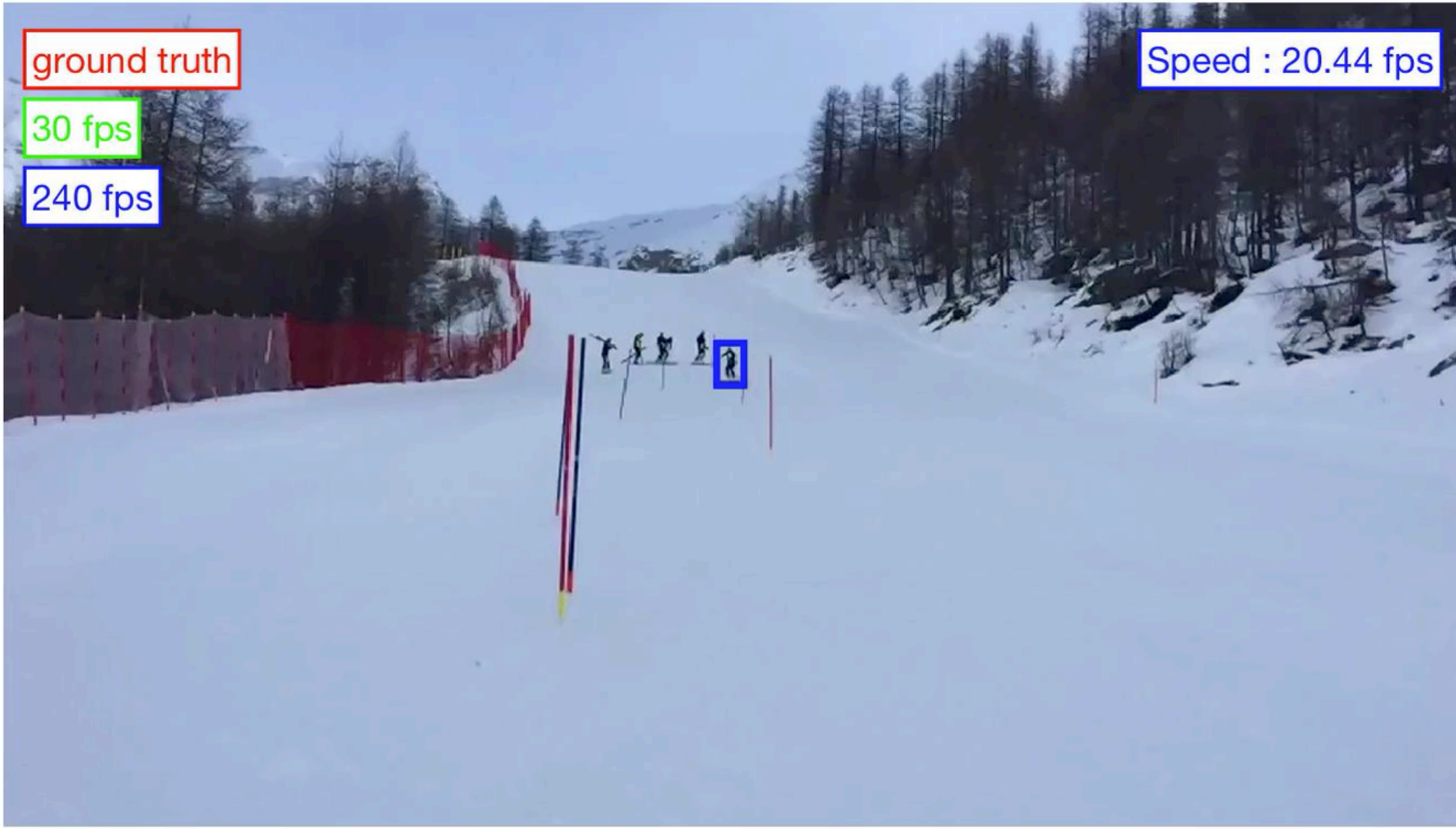
Sony Xperia Phone - 960 FPS - https://www.facebook.com/bbcnews/videos/662989307239621/

Release - “Need for Speed: Benchmark for High Frame Rate Object Tracking”

H. Kiani Galoogahi, et al. “Need for Speed: A Benchmark for High Frame Rate Object Tracking”, ICCV 2017

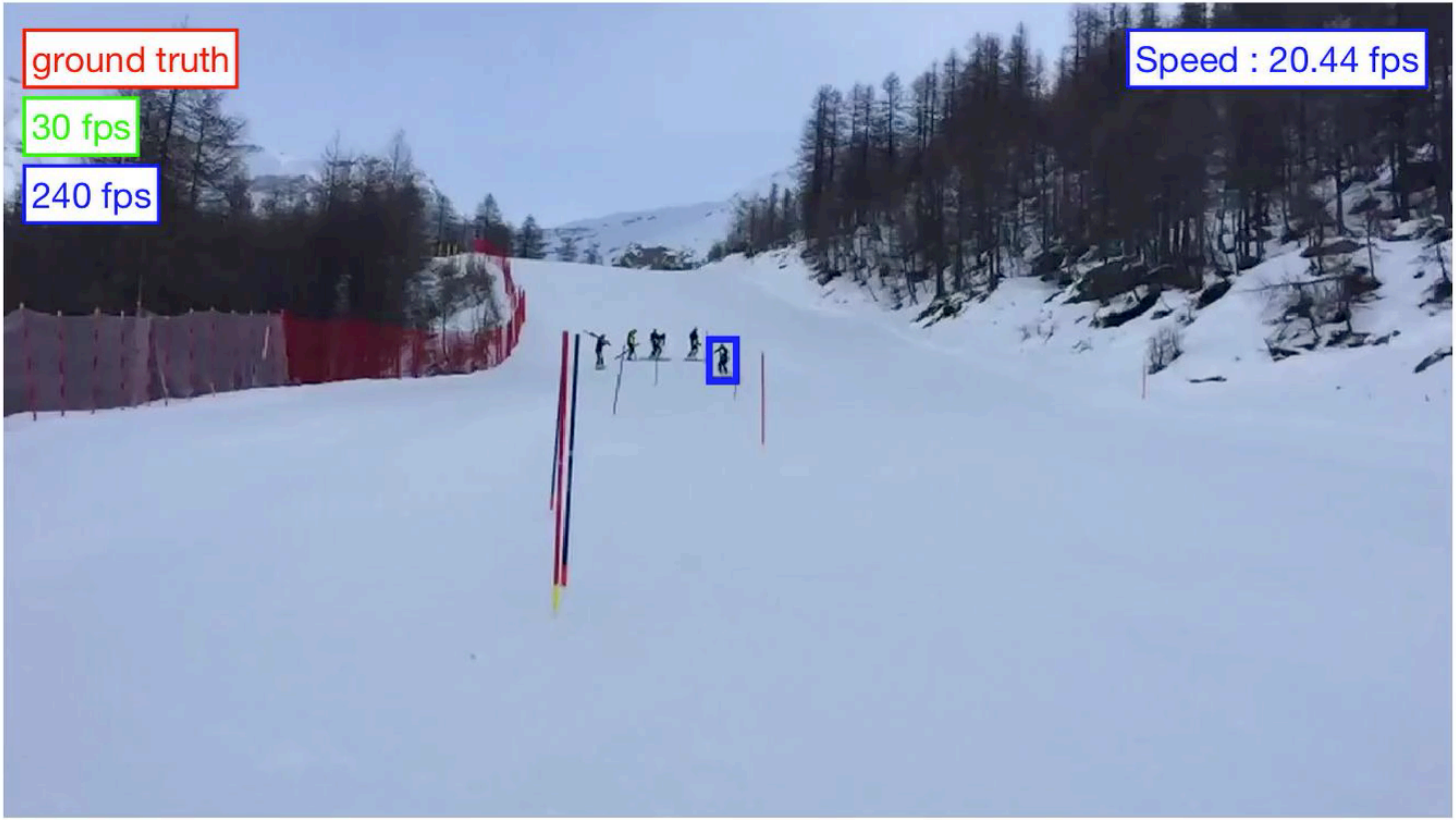
Release - “Need for Speed: Benchmark for High Frame Rate Object Tracking”

H. Kiani Galoogahi, et al. “Need for Speed: A Benchmark for High Frame Rate Object Tracking”, ICCV 2017



Release - "Need for Speed: Benchmark for High Frame Rate Object Tracking"

H. Kiani Galoogahi, et al. "Need for Speed: A Benchmark for High Frame Rate Object Tracking", ICCV 2017



Release - “Need for Speed: Benchmark for High Frame Rate Object Tracking”

H. Kiani Galoogahi, et al. “Need for Speed: A Benchmark for High Frame Rate Object Tracking”, ICCV 2017


X-STAR
4K

Drone-Based Object Tracking

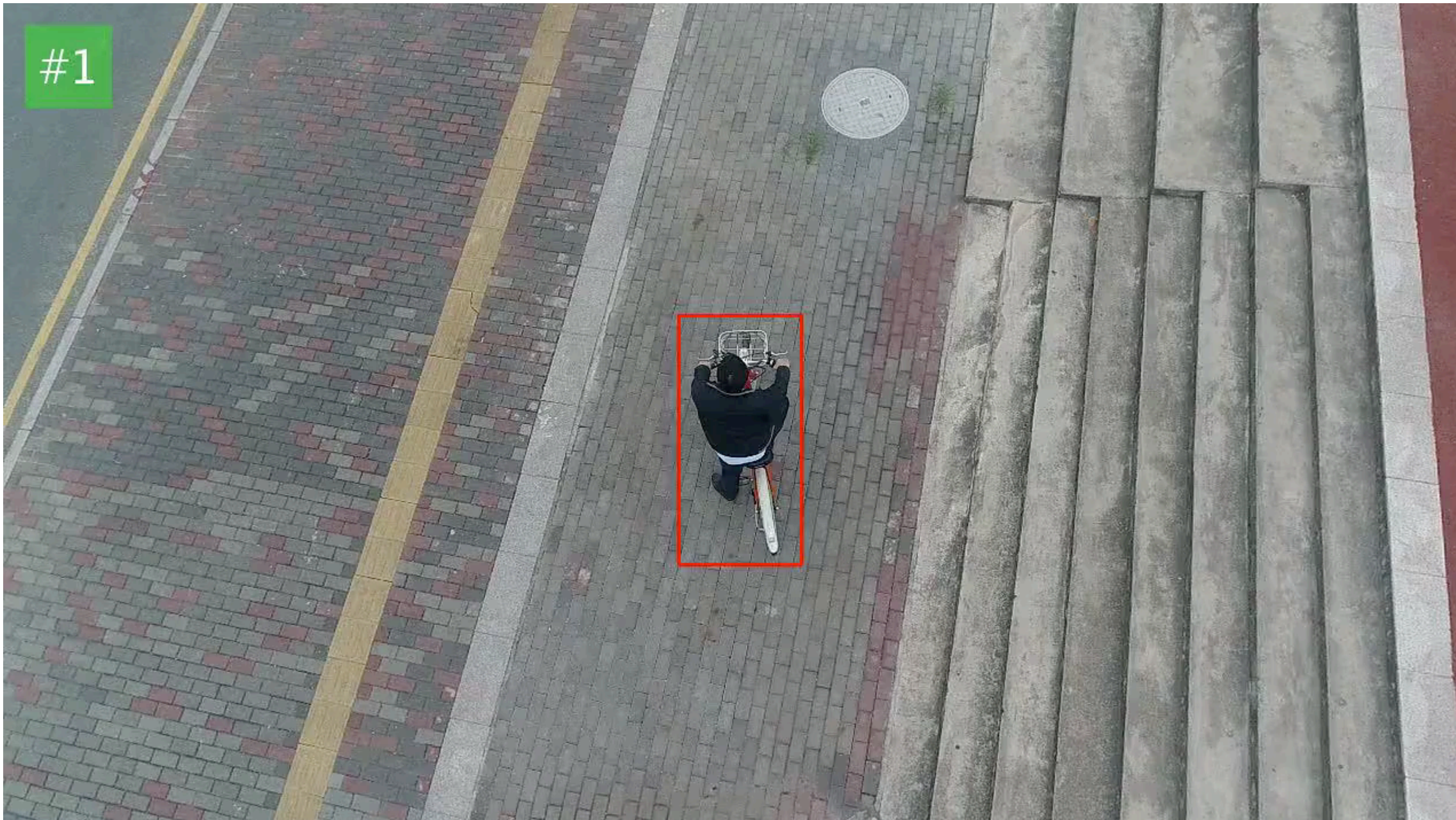
Ground truth | Deep-LK original | Deep-LK improved

Drone-Based Object Tracking

Ground truth | Deep-LK original | Deep-LK improved

Drone-Based Object Tracking

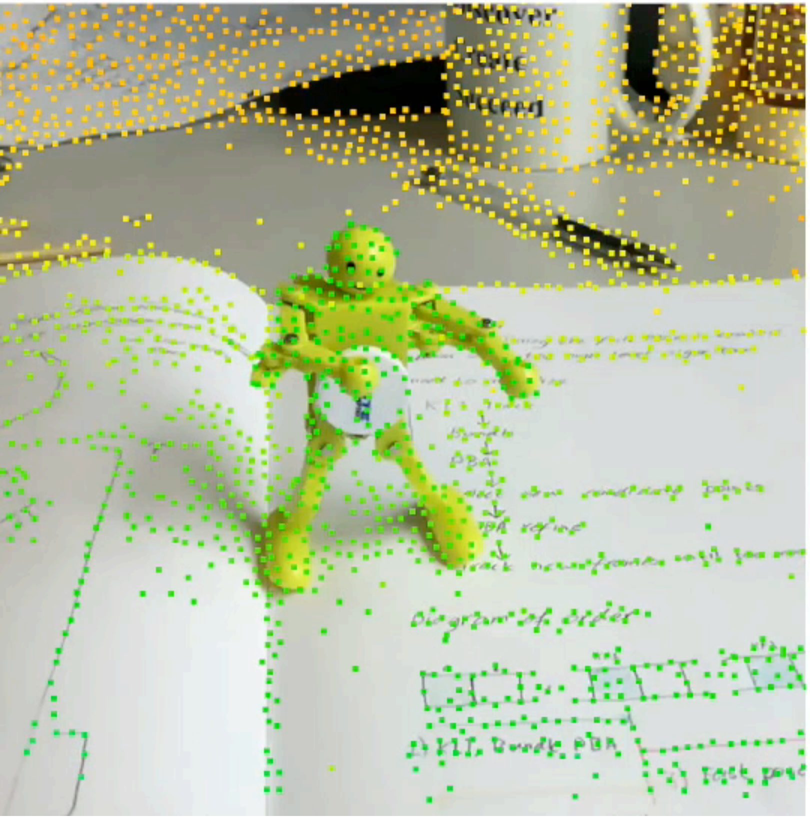
Ground truth
Deep-LK original
Deep-LK improved



High frame rate: 240 FPS

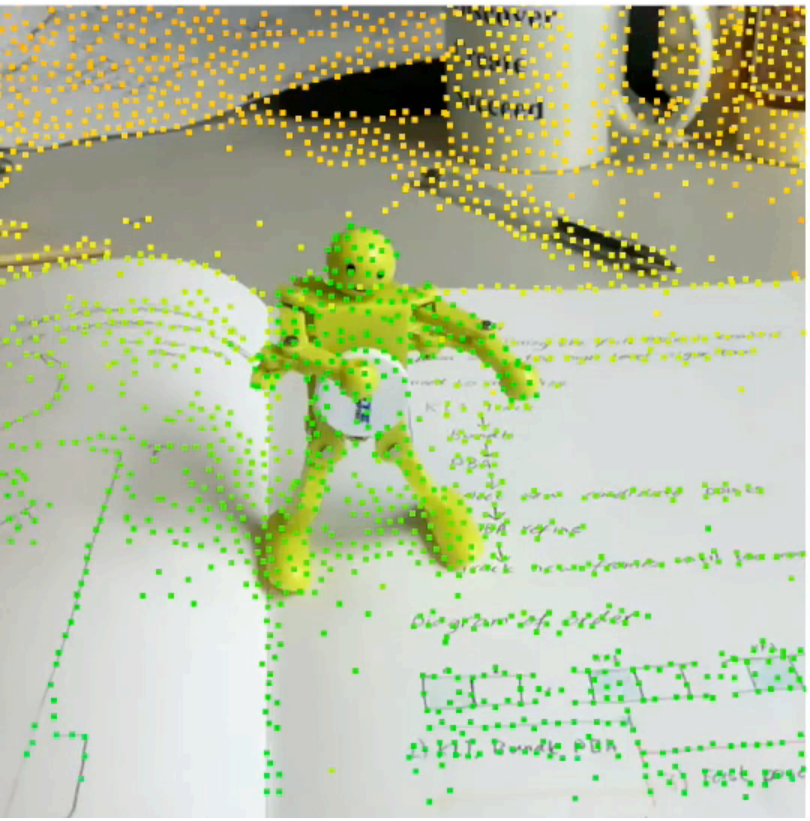
Normal: 30 FPS

Sparse Photometric Track



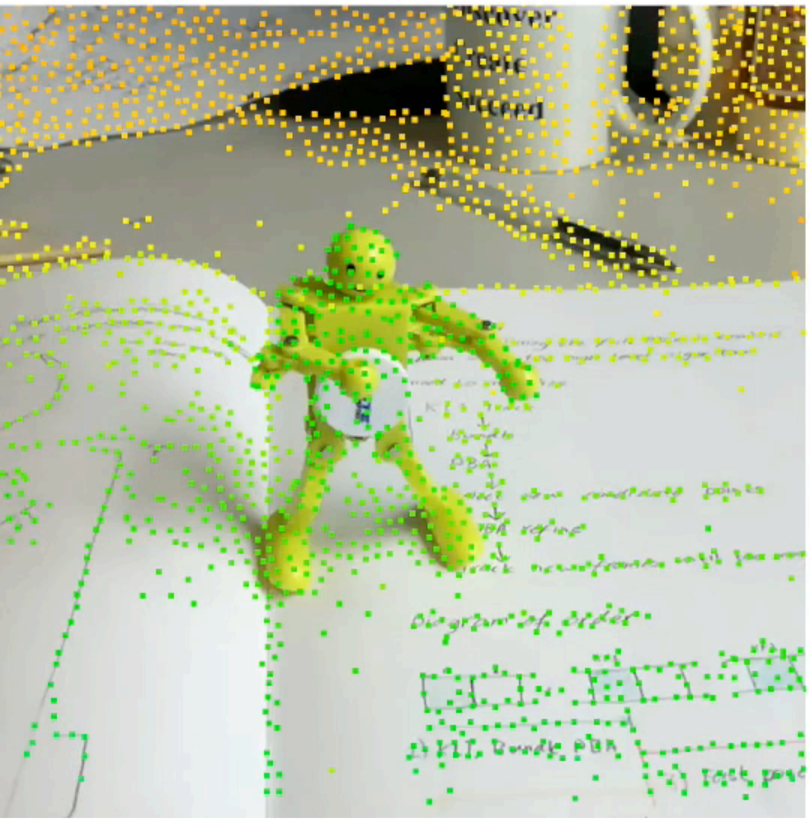
Dense Cloud Reconstruction

Sparse Photometric Track



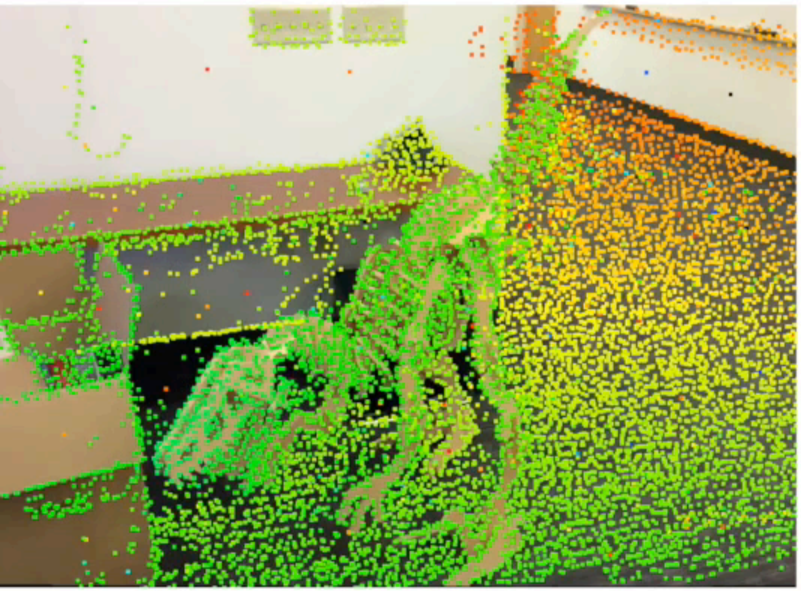
Dense Cloud Reconstruction

Sparse Photometric Track



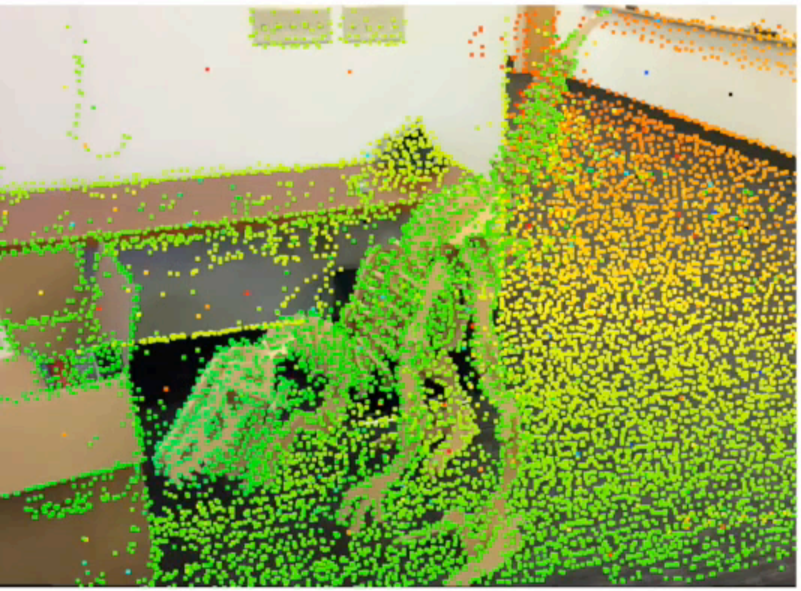
Dense Cloud Reconstruction

Sparse Photometric Track



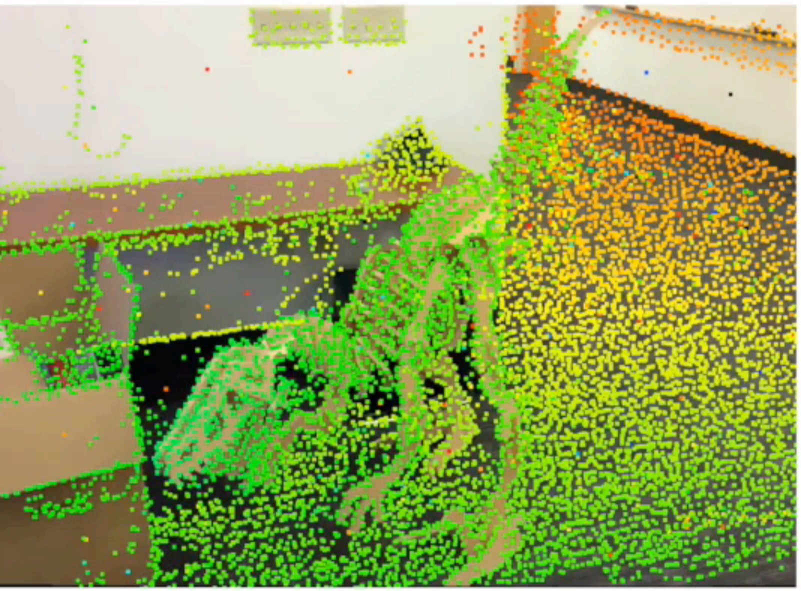
Dense Cloud Reconstruction

Sparse Photometric Track



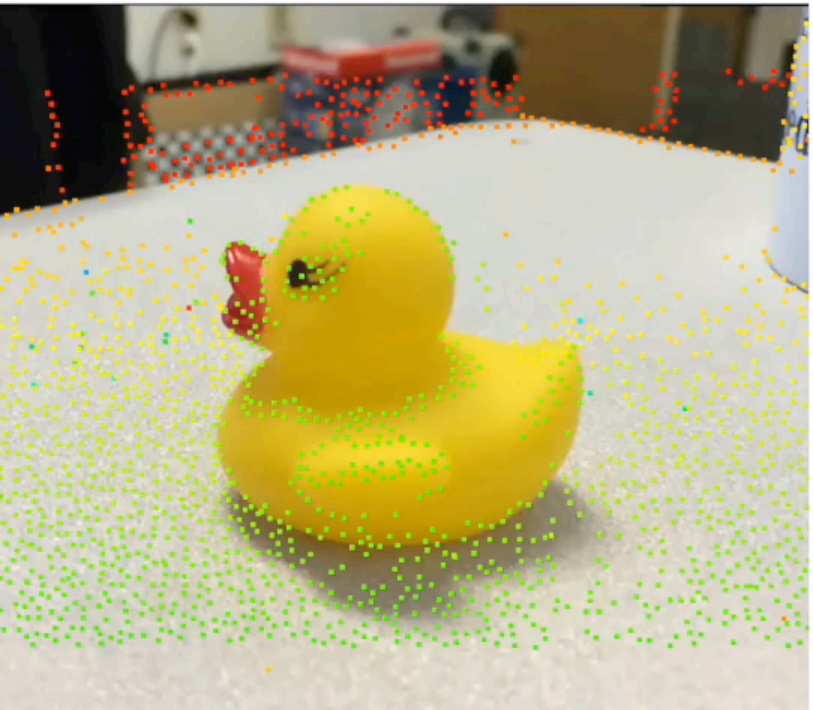
Dense Cloud Reconstruction

Sparse Photometric Track



Dense Cloud Reconstruction

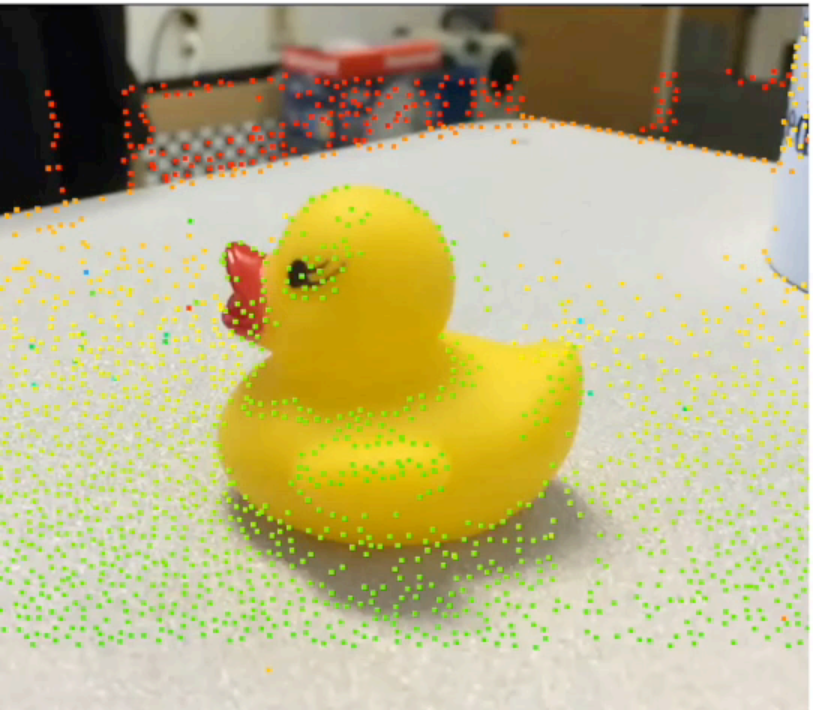
Sparse Photometric Track



Dense Cloud Reconstruction

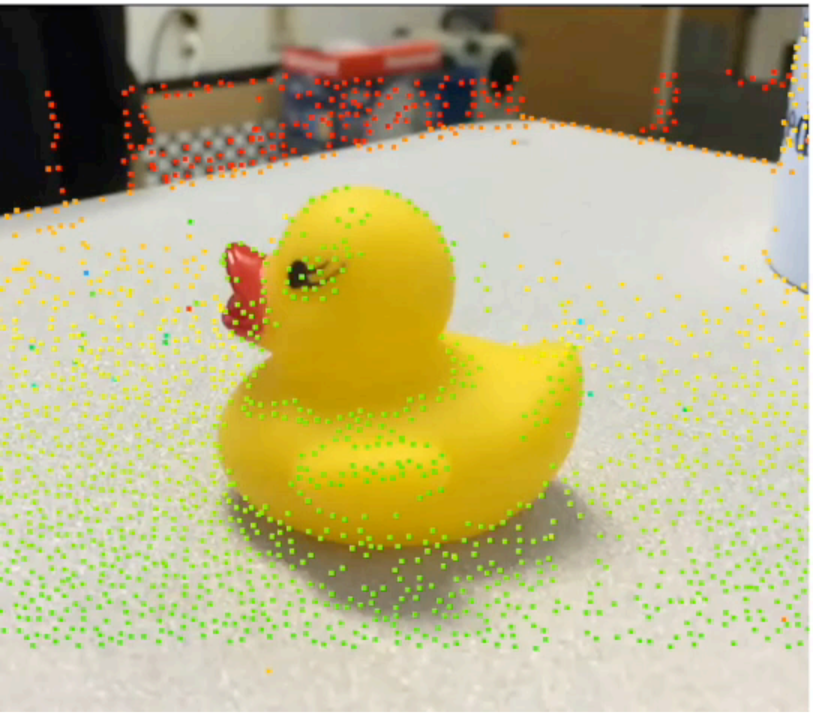
Sparse Photometric Track

Dense Cloud Reconstruction



Sparse Photometric Track

Dense Cloud Reconstruction



Sparse Photometric Track

Dense Cloud Reconstruction

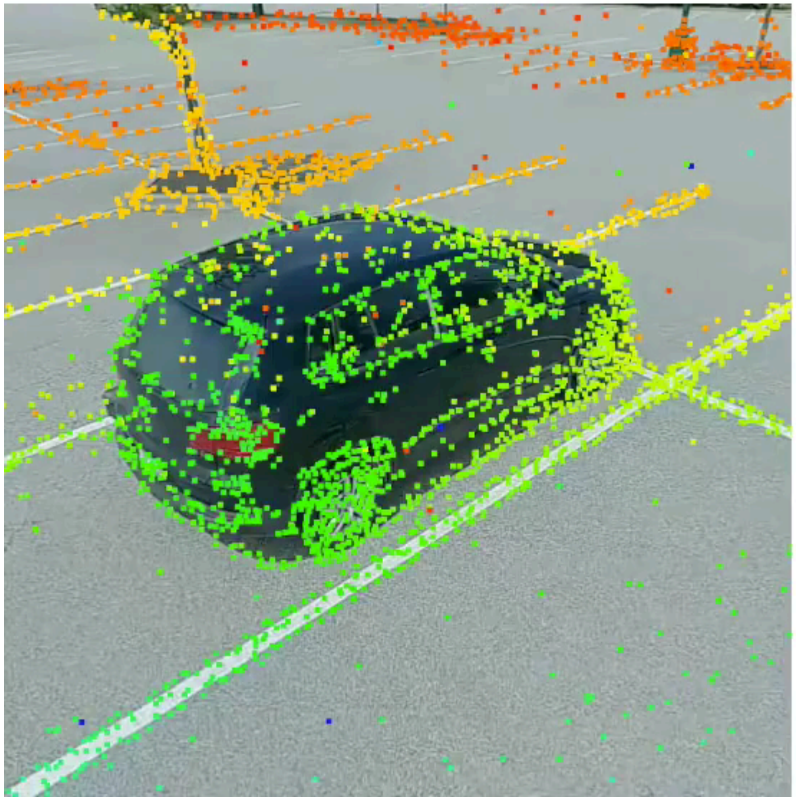
Sparse Photometric Track

Dense Cloud Reconstruction

Sparse Photometric Track

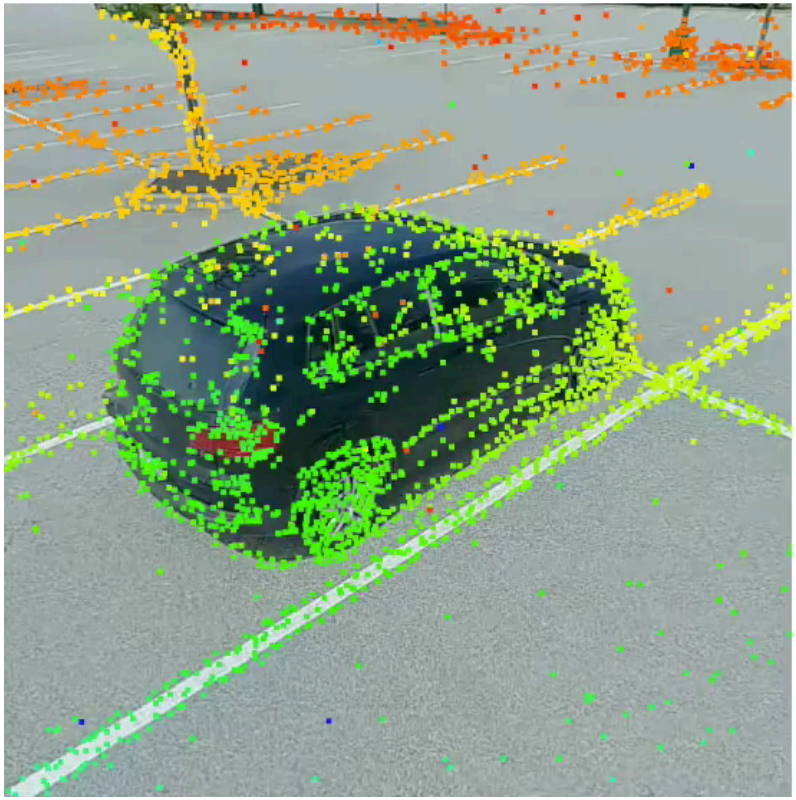
Dense Cloud Reconstruction

Sparse Photometric Track



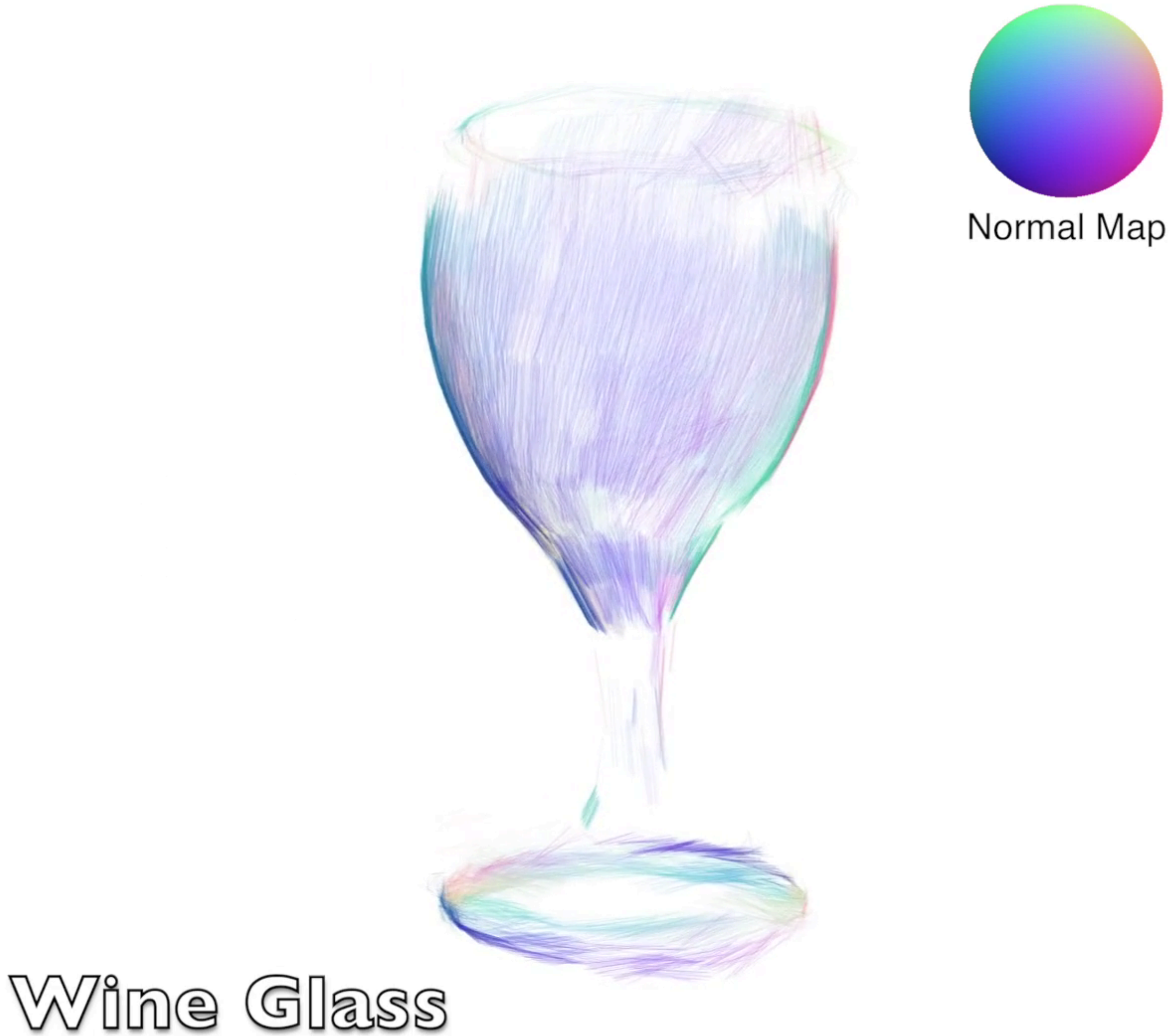
Semi Dense Cloud Reconstruction

Sparse Photometric Track



Semi Dense Cloud Reconstruction

Extension to Transparent Objects



Ham, Singh and Lucey “Occlusions are Fleeting - Texture is Forever: Moving Past Brightness Constancy” WACV 2017.

Extension to Transparent Objects

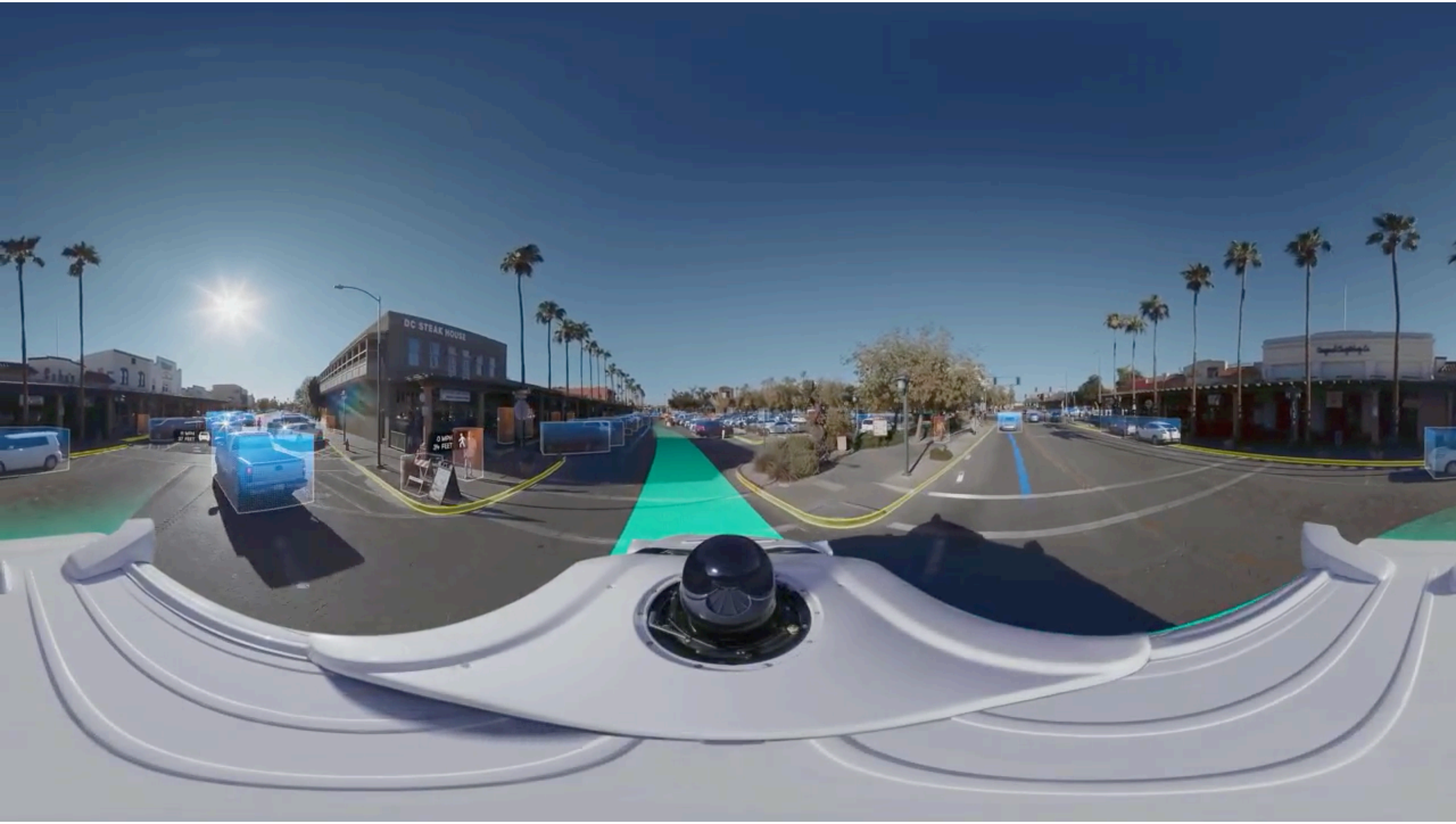
Ham, Singh and Lucey "Occlusions are Fleeting - Texture is Forever: Moving Past Brightness Constancy" WACV 2017.

Autonomous Vehicles!!!

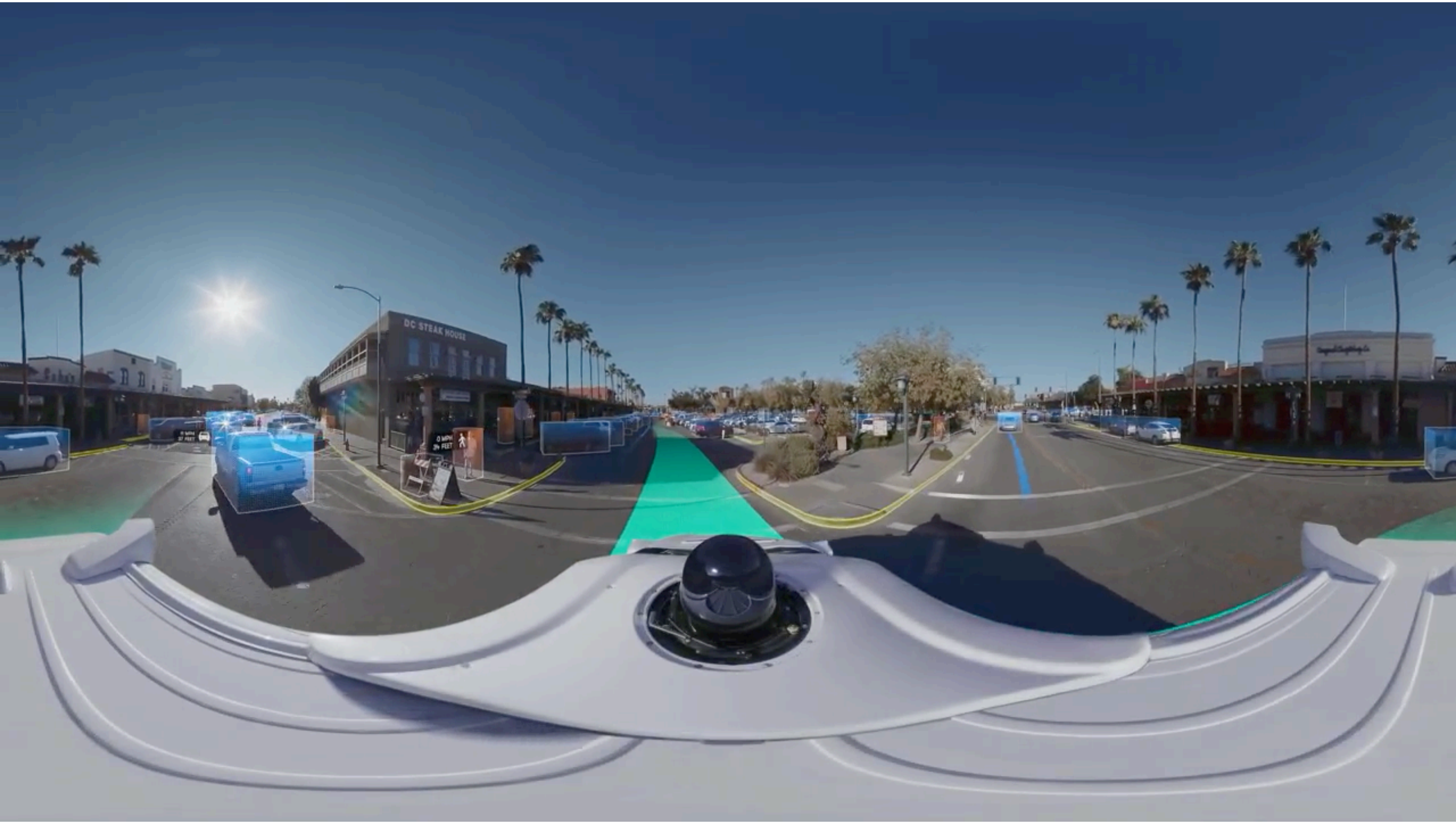
2007

2017

Autonomous Vehicles!!!



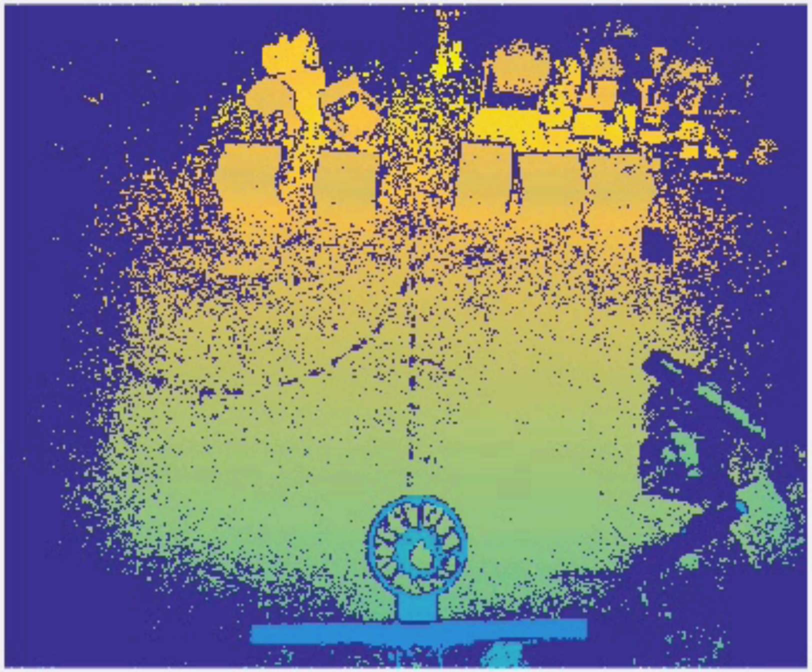
Autonomous Vehicles!!!



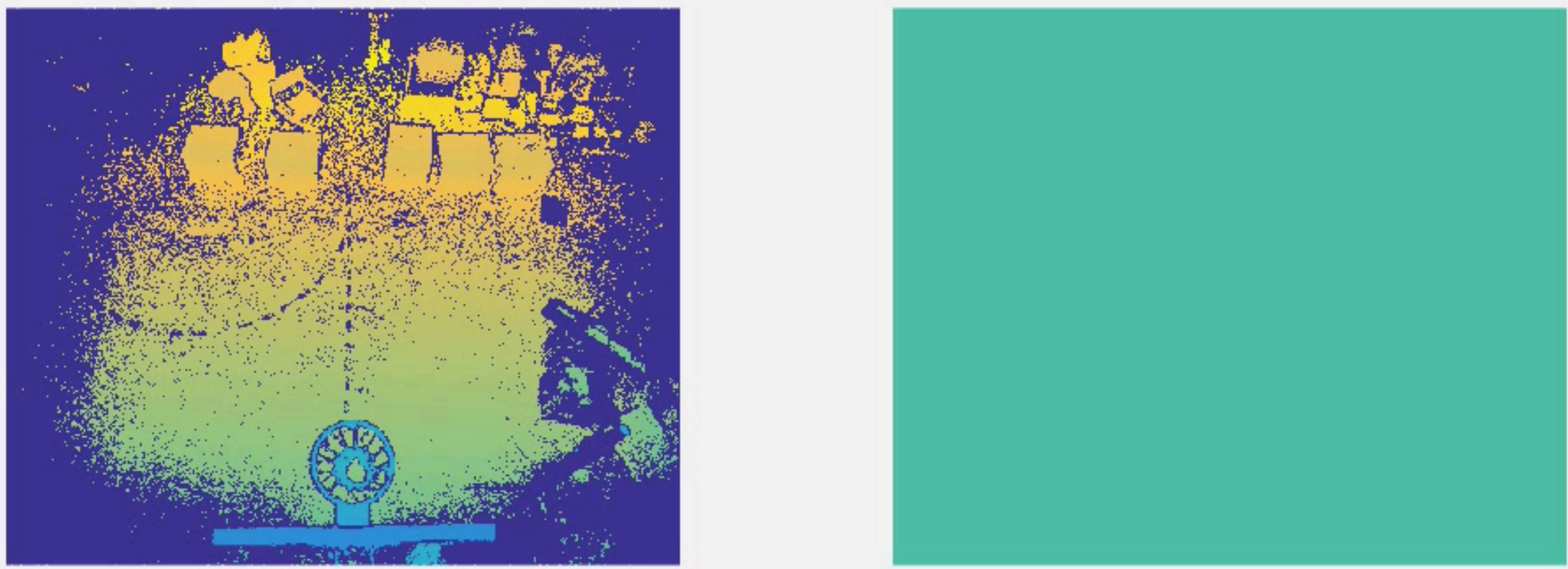
Trend - Depth Cameras

Kinect 2 Sensor
KINECT2
MOCK IMAGE - THISISXBOX.COM
Standard Basketball Hoop

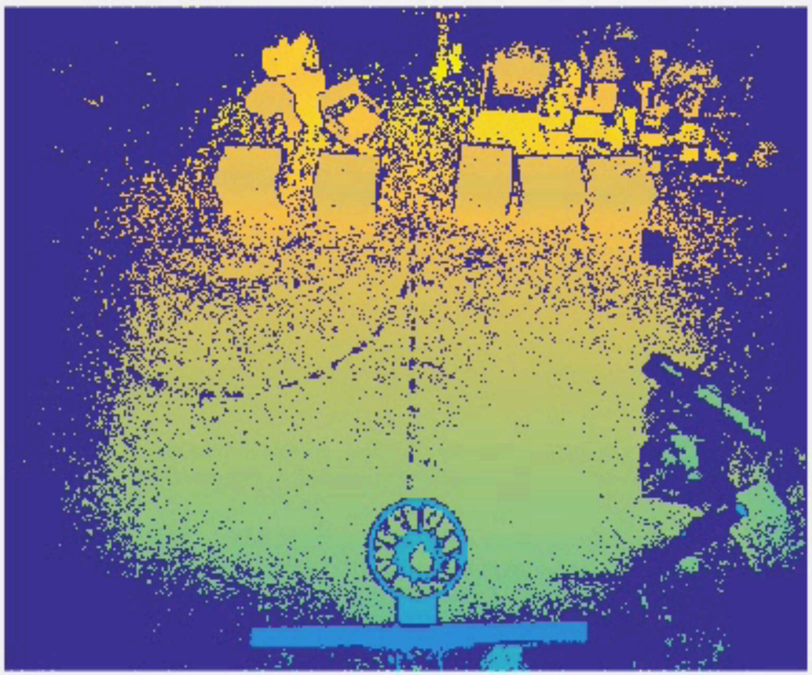
Ball Tracking

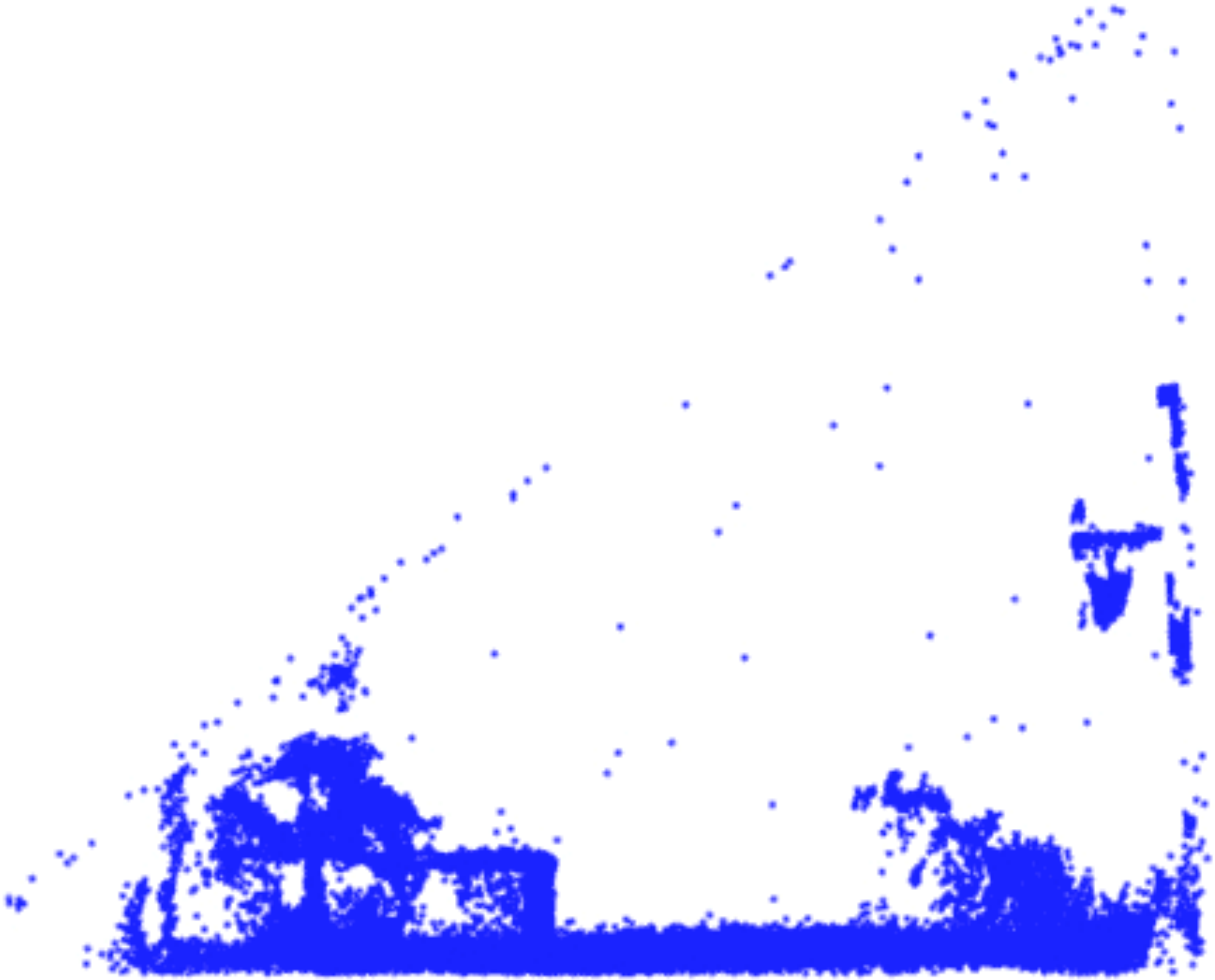


Ball Tracking

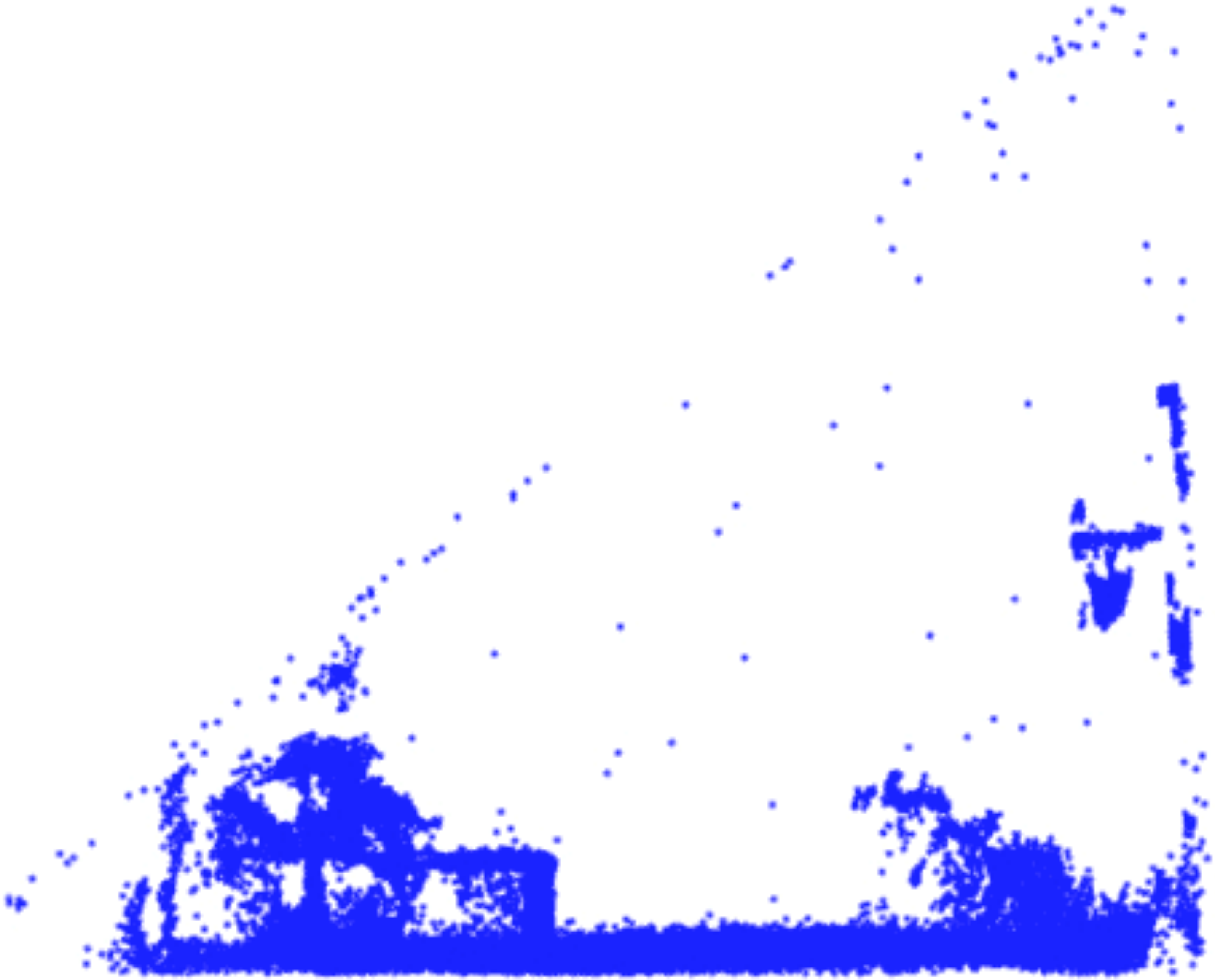


Ball Tracking



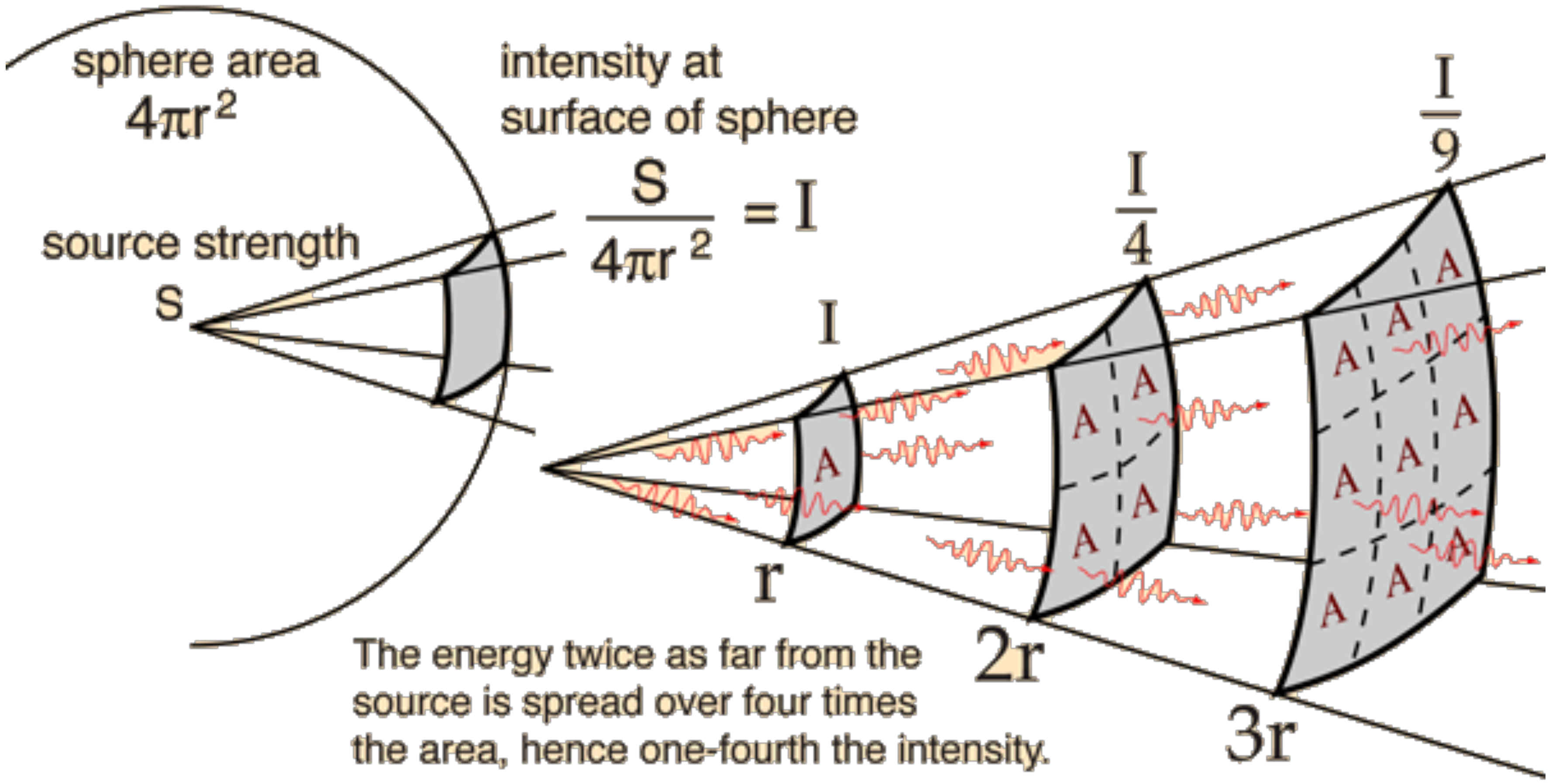


Current Time: 0.00 (secs), Hoop Impact Time: 2.35 (secs)
Angle of Entry: 53.69 (degs)



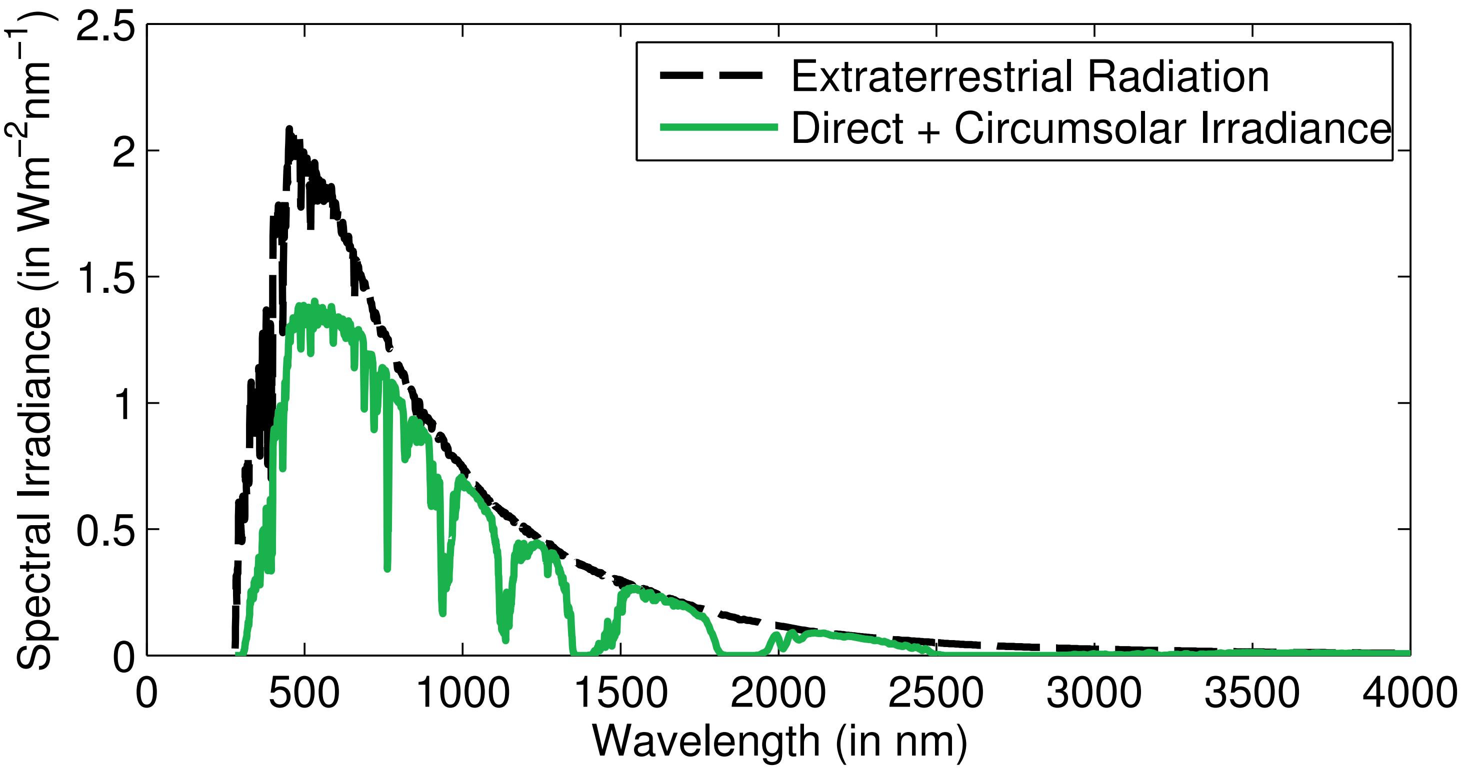
Current Time: 0.00 (secs), Hoop Impact Time: 2.35 (secs)
Angle of Entry: 53.69 (degs)

Limitations - Range



Limitations - Ambient Light

- A sunny day on Earth can reach up to $1120 Wm^{-2}$
- Tabletop projector releases on average 10W of light.



The Future

Homogenous Codes for Energy-Efficient Illumination and Imaging

Matthew O'Toole, Supreeth Achar,
Srinivasa G. Narasimhan, Kiriakos N. Kutulakos

ACM SIGGRAPH 2015

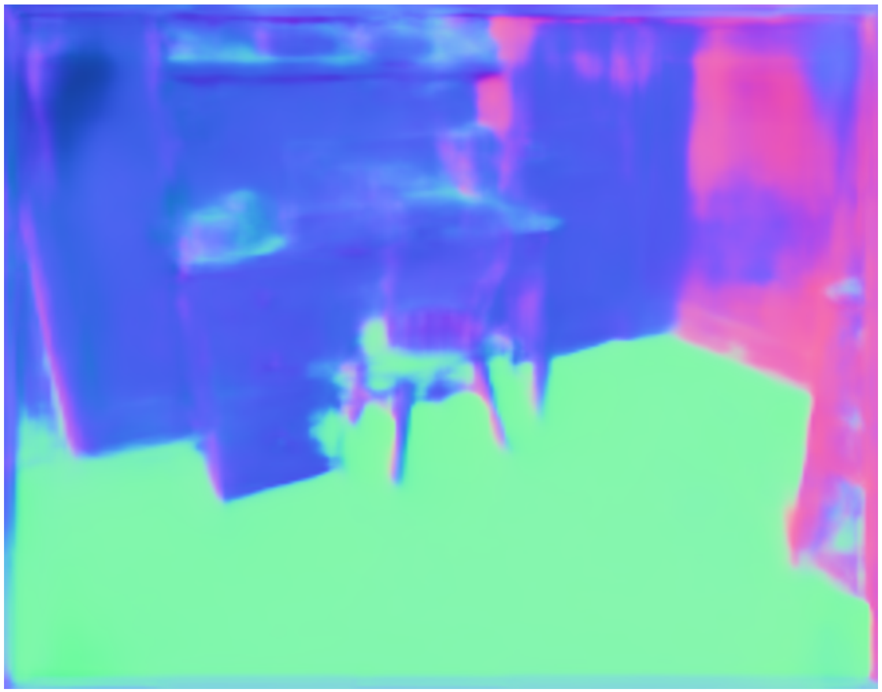
The Future

Homogenous Codes for Energy-Efficient Illumination and Imaging

Matthew O’Toole, Supreeth Achar,
Srinivasa G. Narasimhan, Kiriakos N. Kutulakos

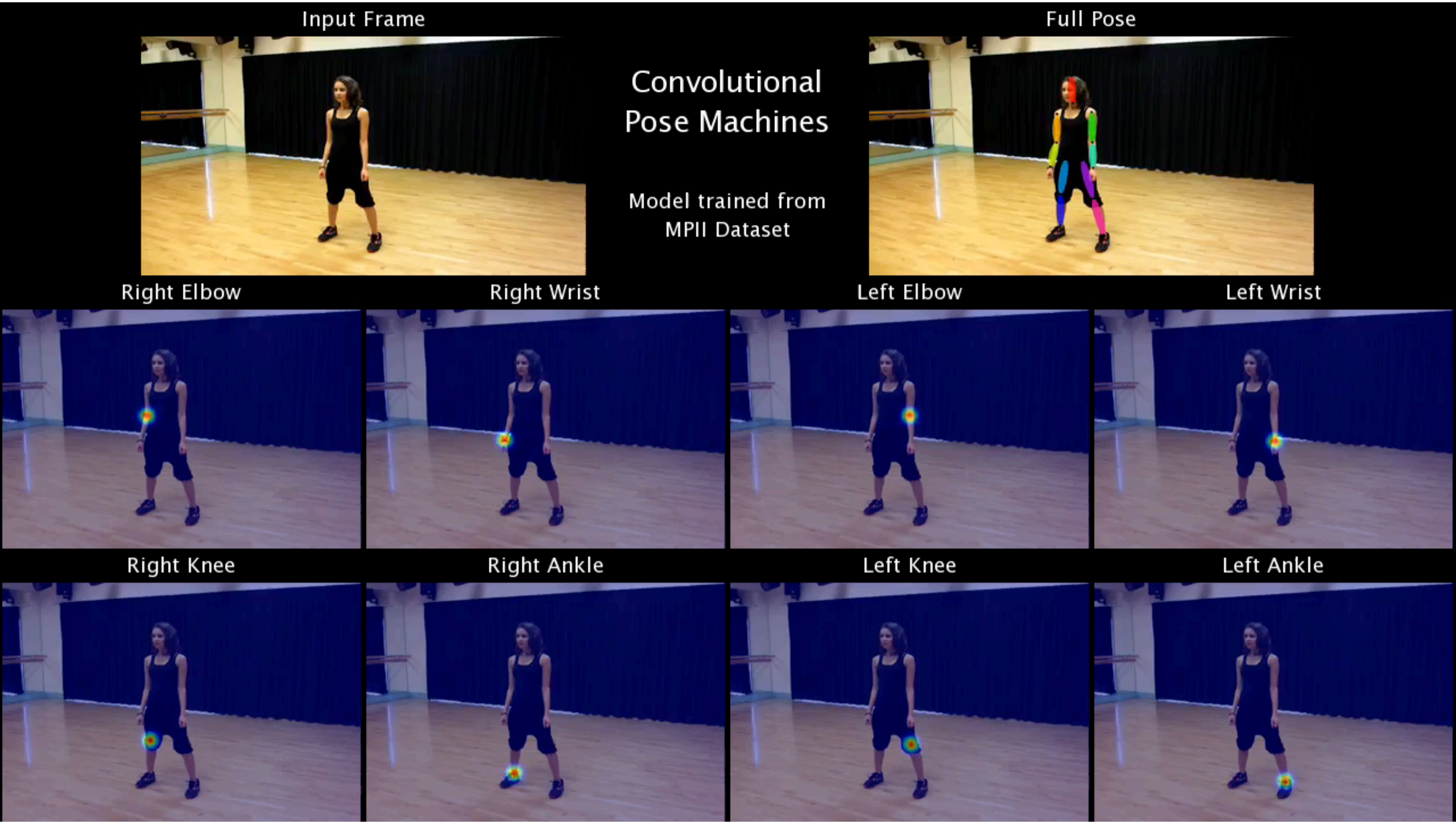
ACM SIGGRAPH 2015

Trend - High Fidelity Deep Learning



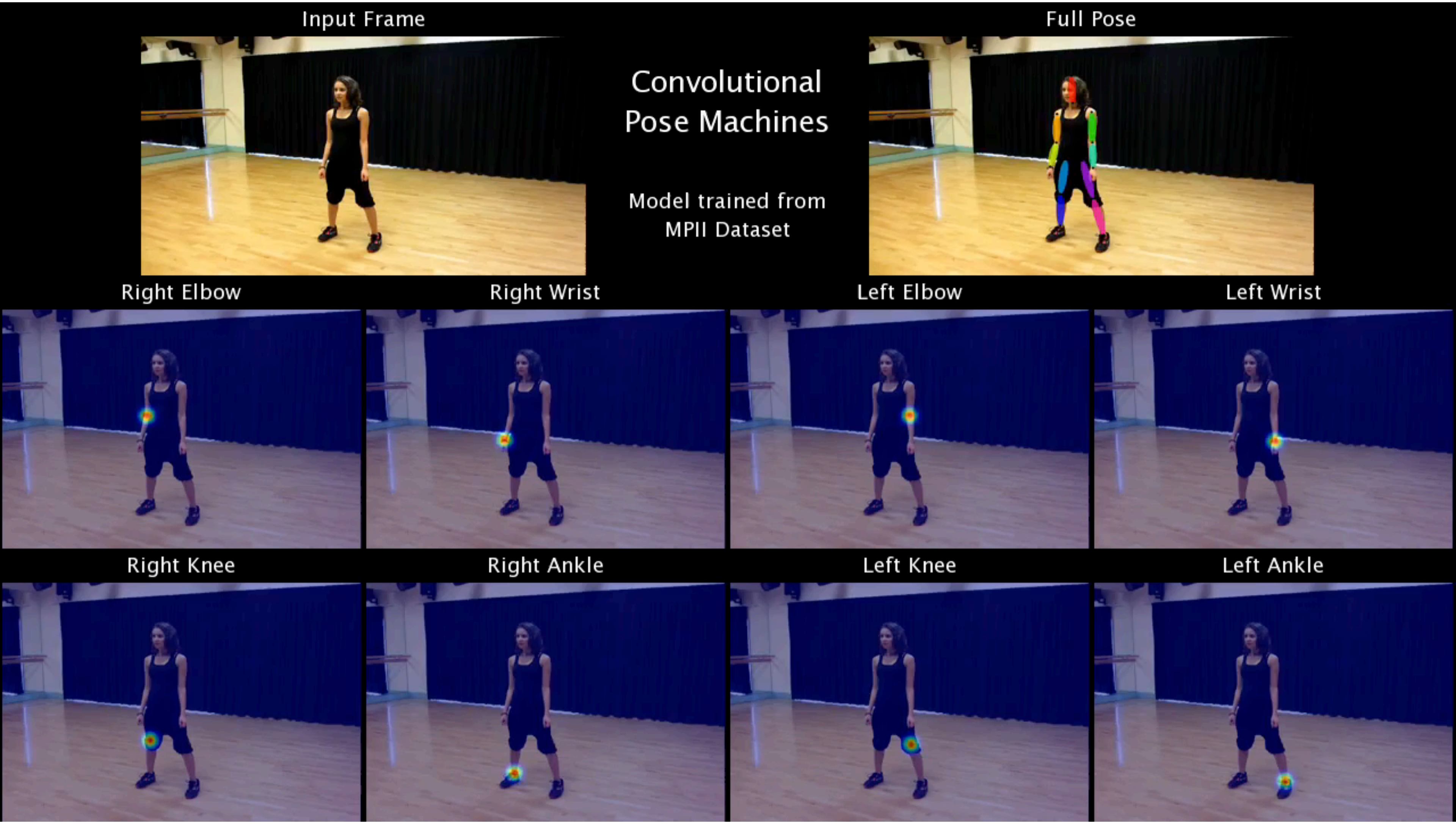
A. Bansal, et al. “Marr Revisited: 2D-3D Alignment via Surface Normal Prediction”. In CVPR 2016.

High-Fidelity Deep Learning



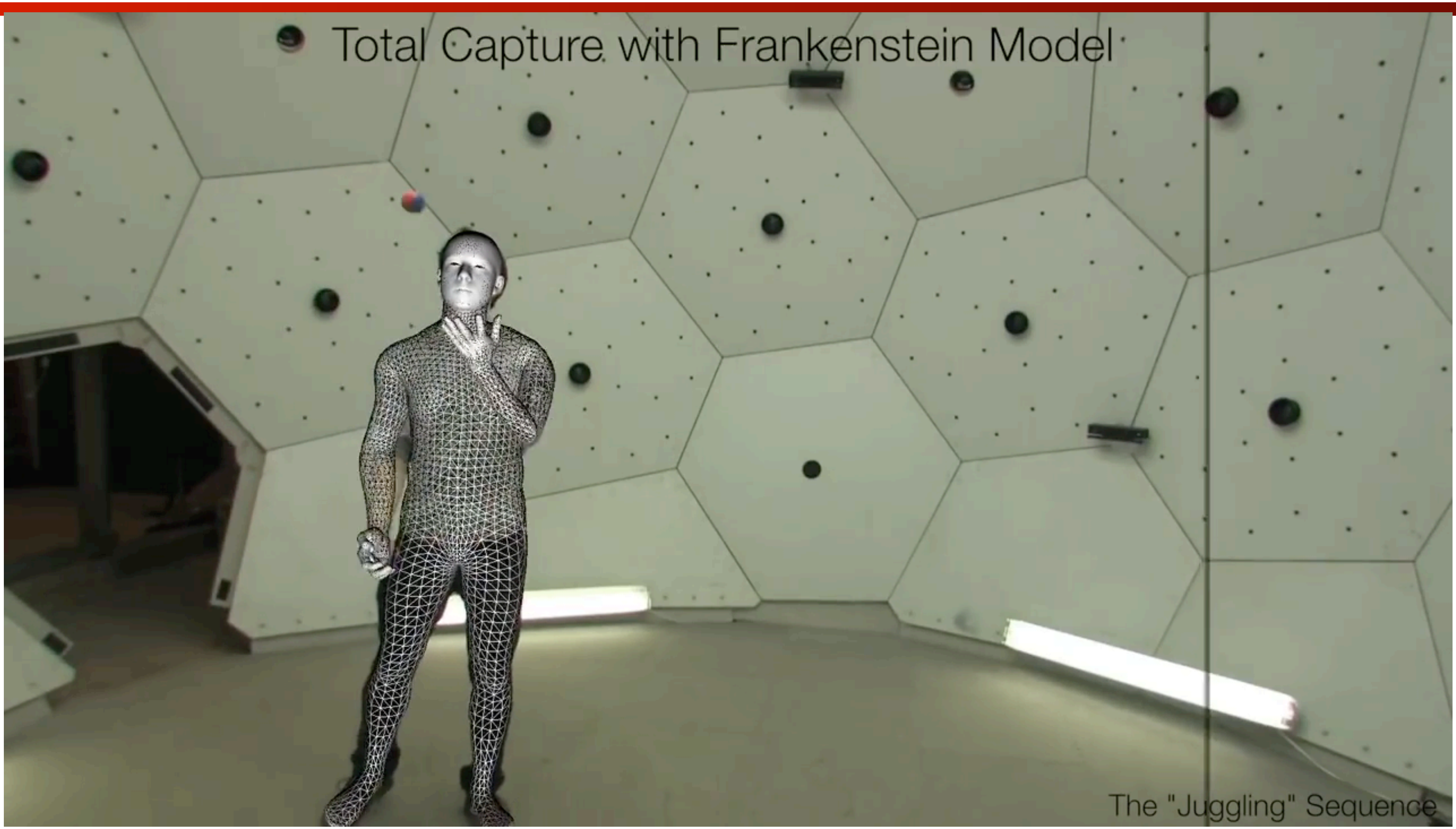
S. Wei, et al. “Convolutional Pose Machines”. In CVPR 2016.

High-Fidelity Deep Learning



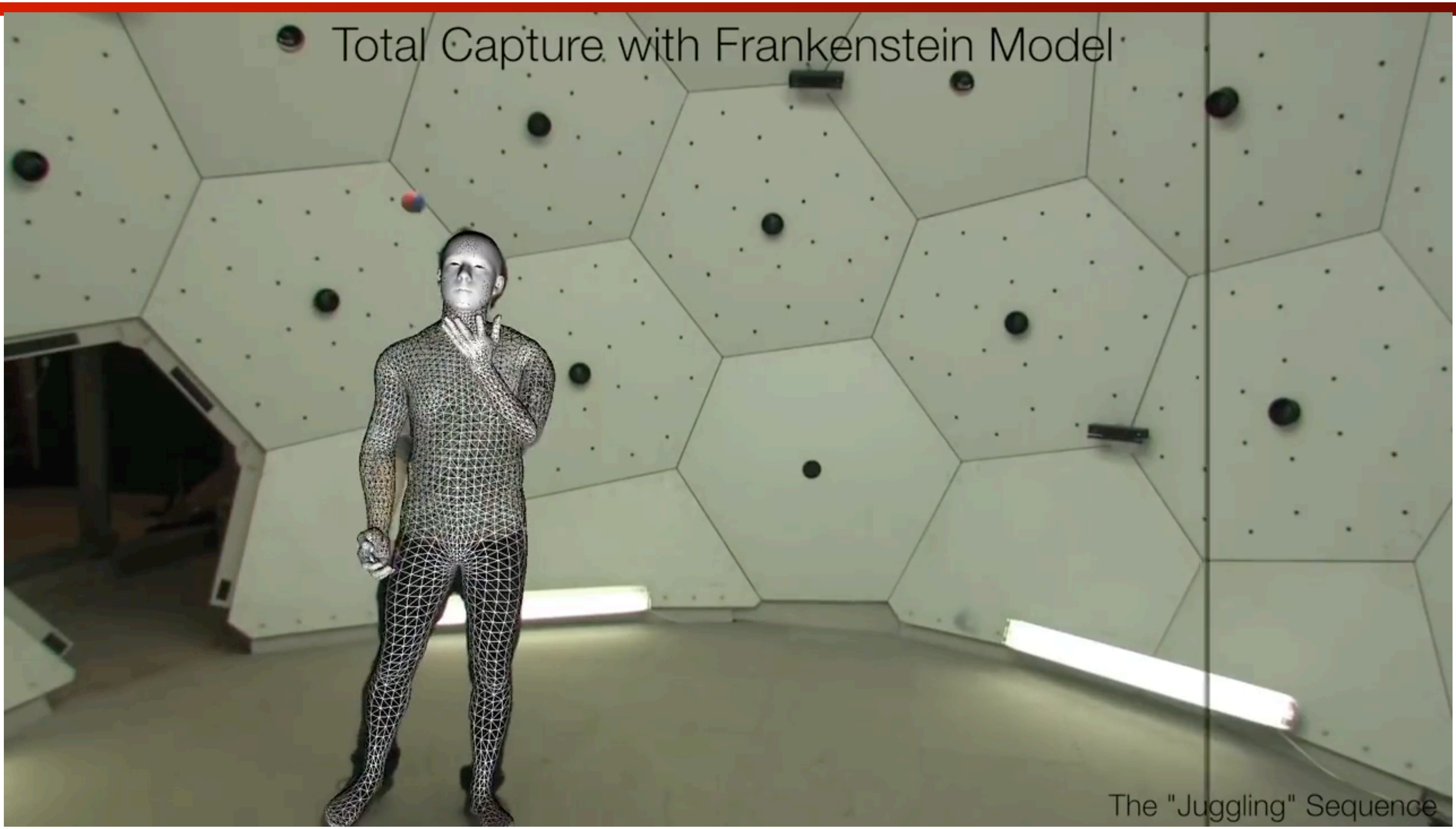
S. Wei, et al. “Convolutional Pose Machines”. In CVPR 2016.

Total Capture

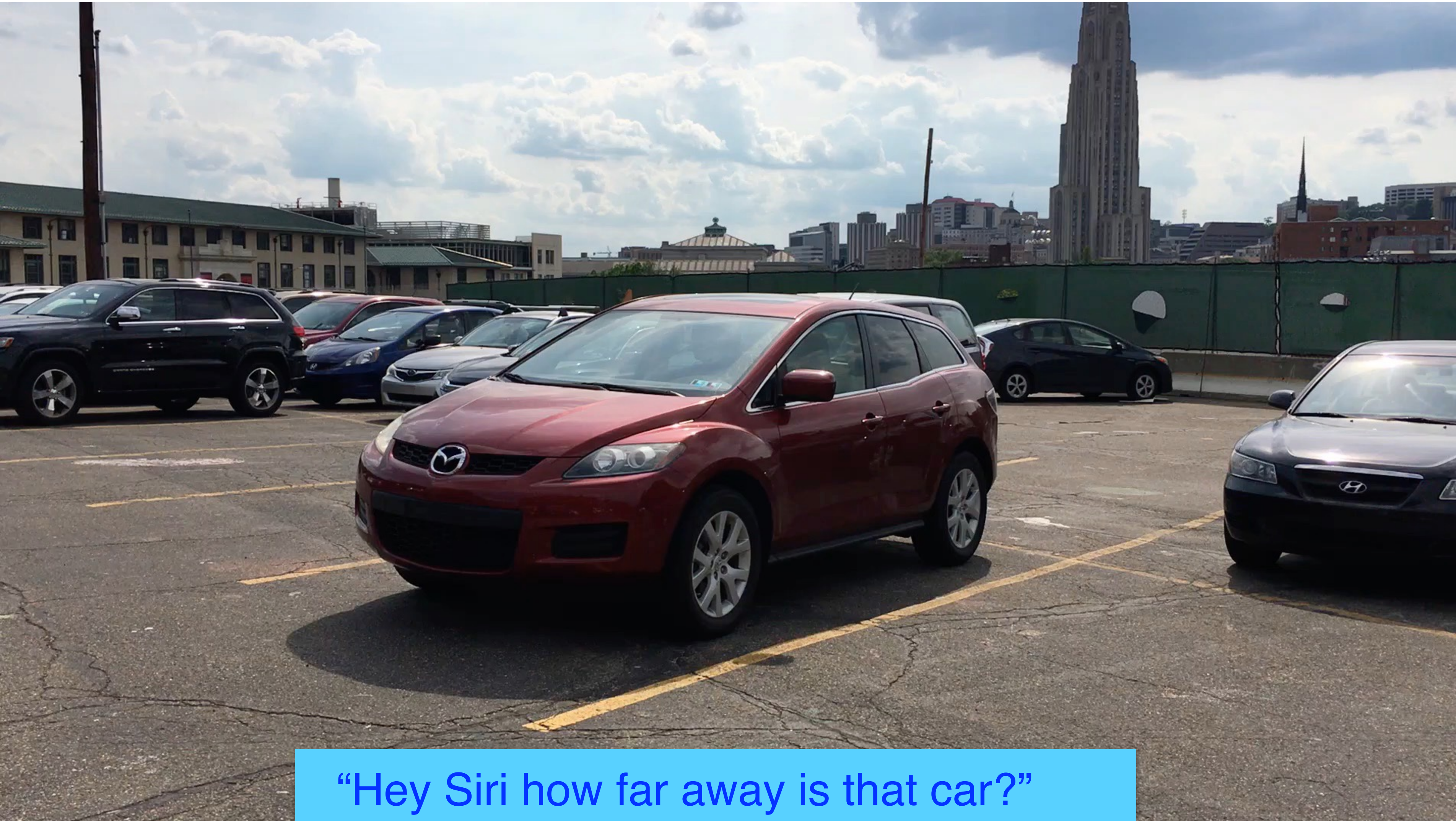


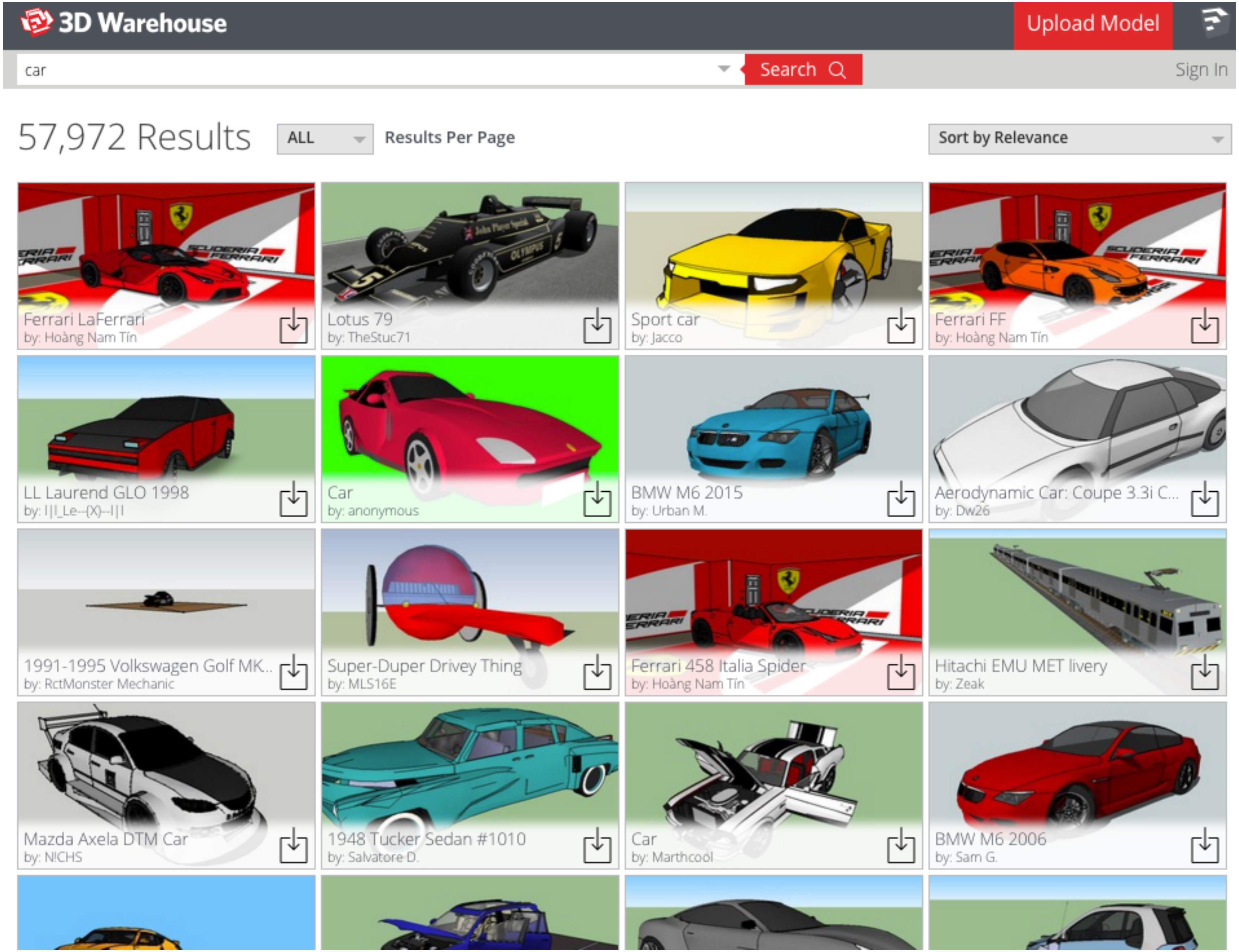
H. Joo, et al. "Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies ". In CVPR 2018.

Total Capture



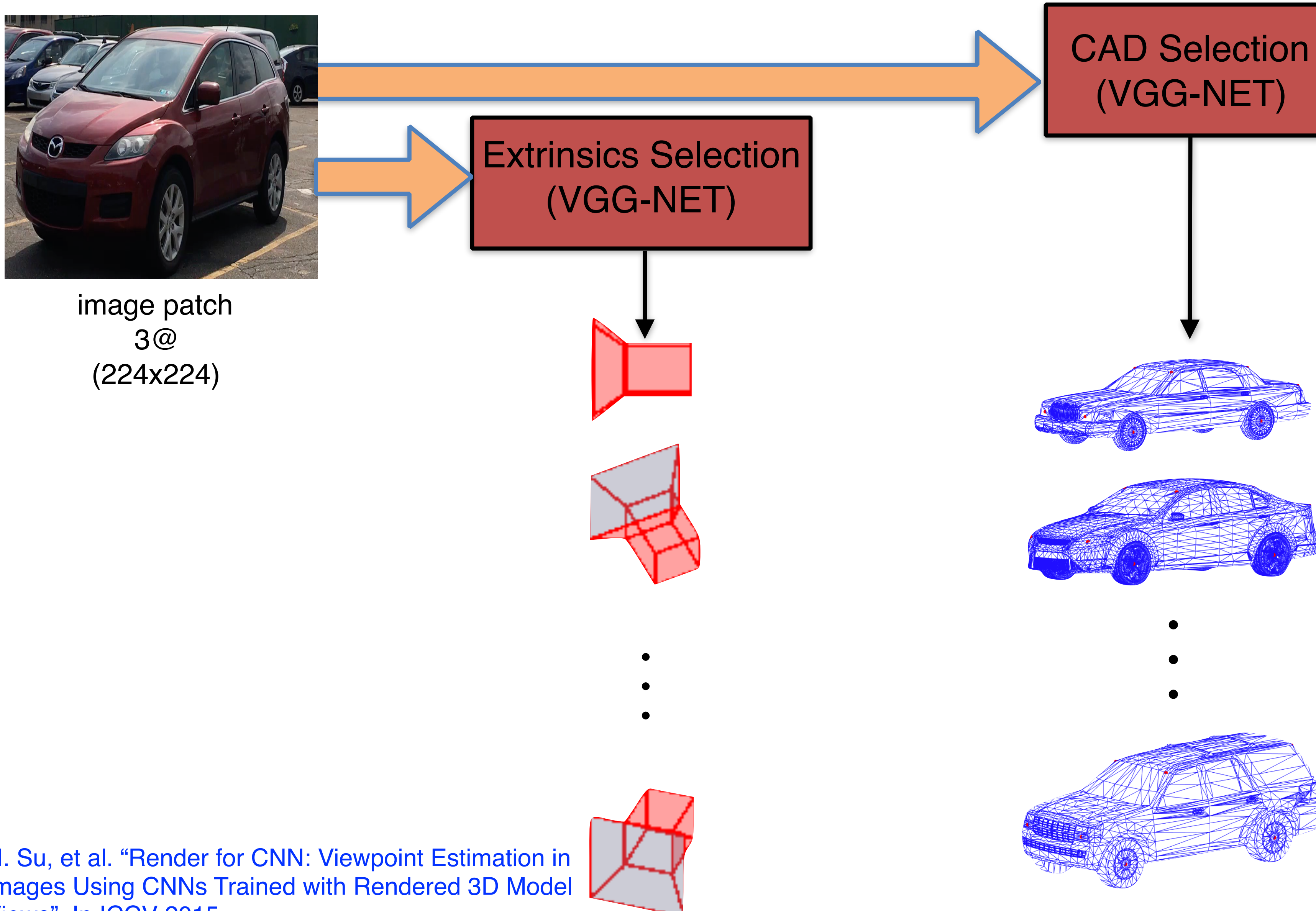
H. Joo, et al. “Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies ”. In CVPR 2018.


“Hey Siri how far away is that car?”



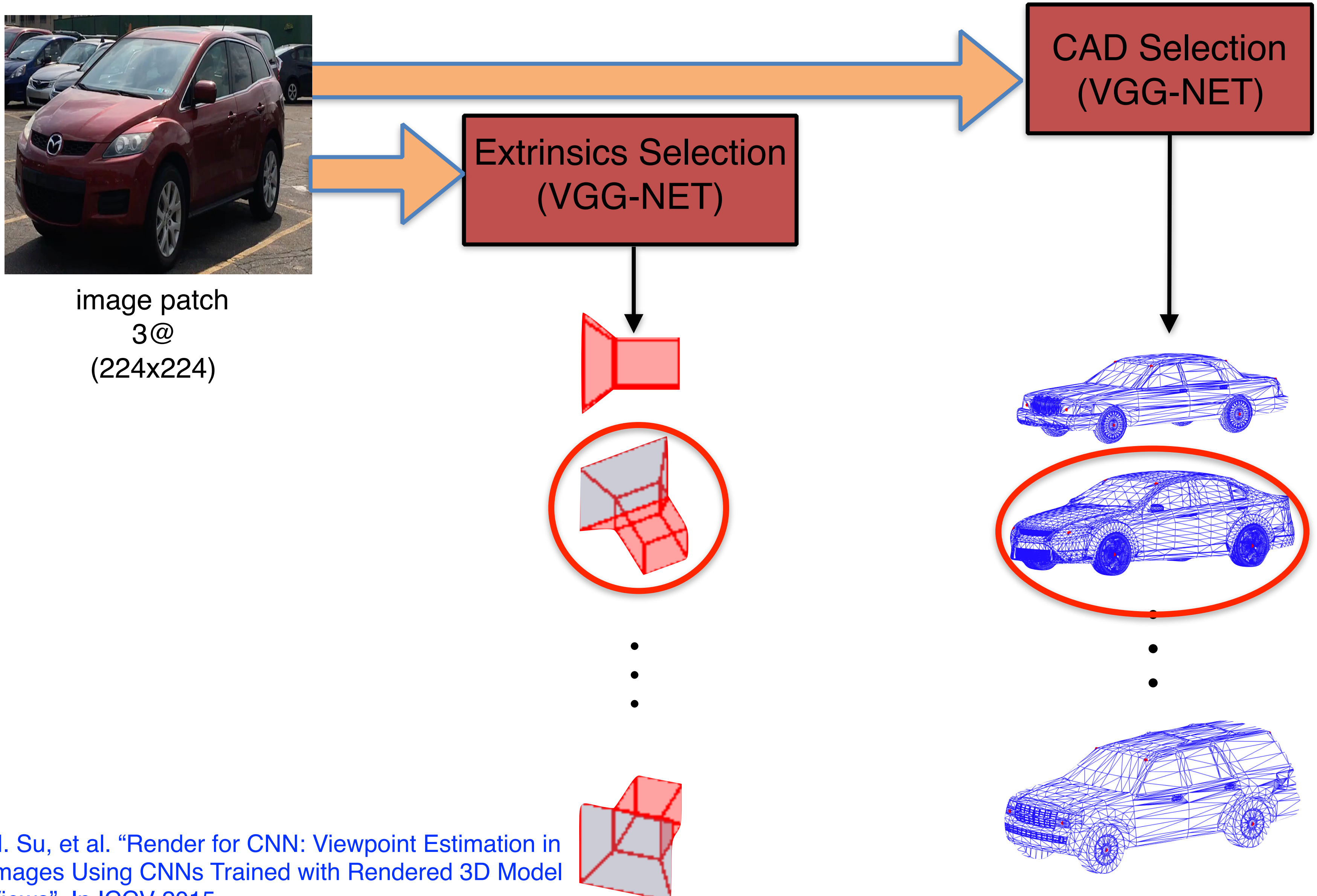
A. Chang et al., “ShapeNet: An Information-Rich 3D Model Repository”, arXiv 1512.03012, 2015

Current State of the Art



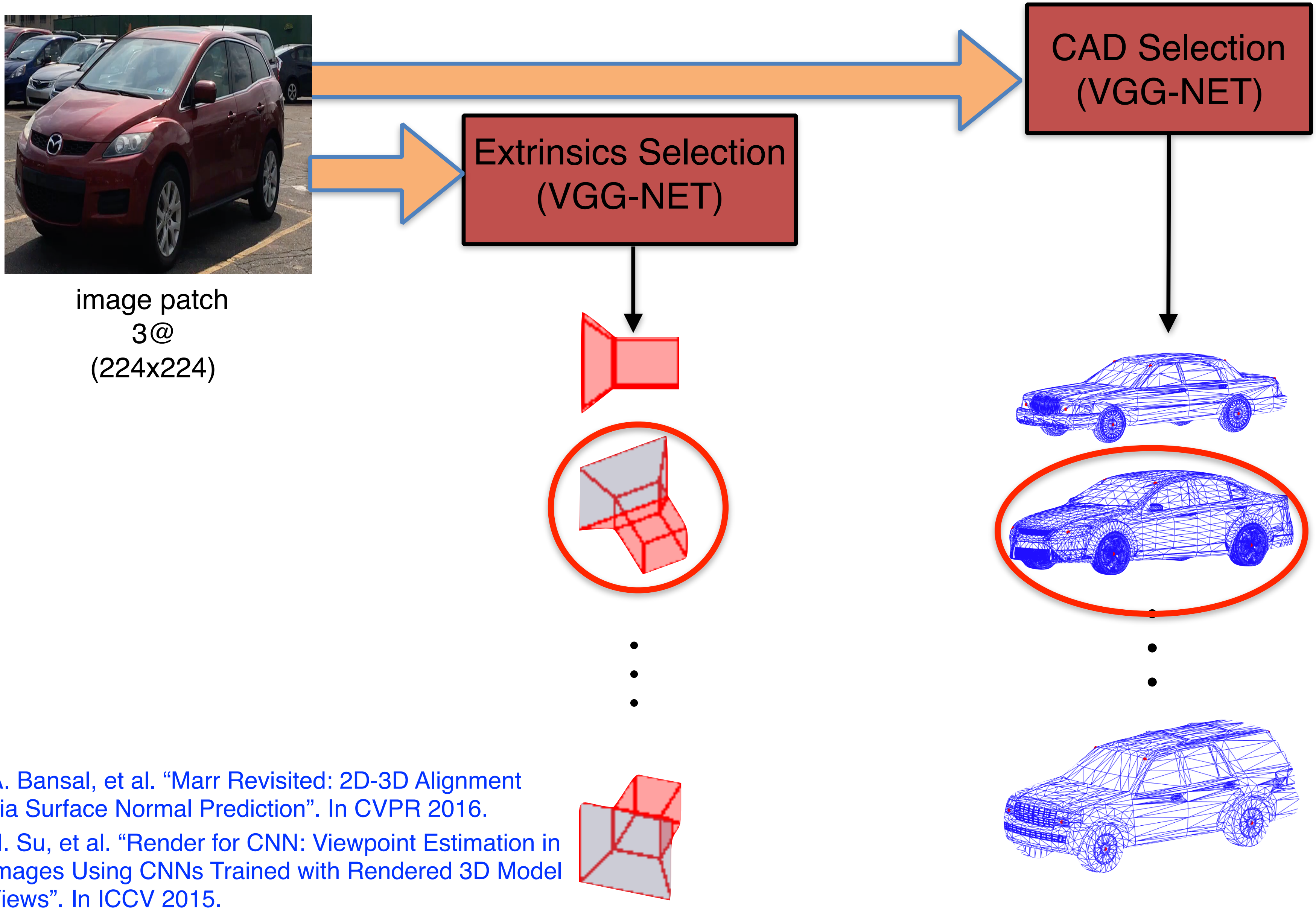
H. Su, et al. "Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views". In ICCV 2015.

Current State of the Art



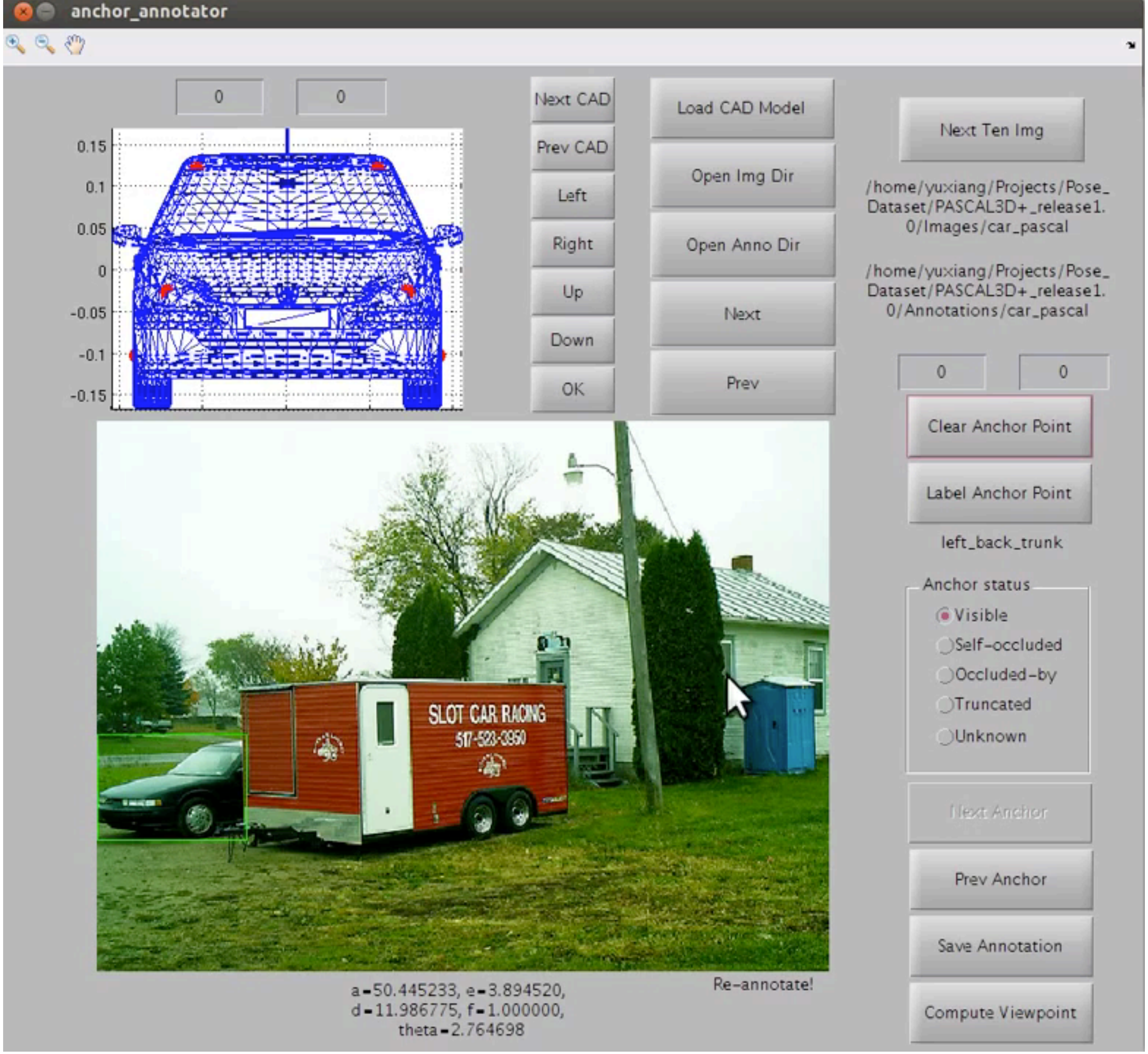
H. Su, et al. “Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views”. In ICCV 2015.

Current State of the Art

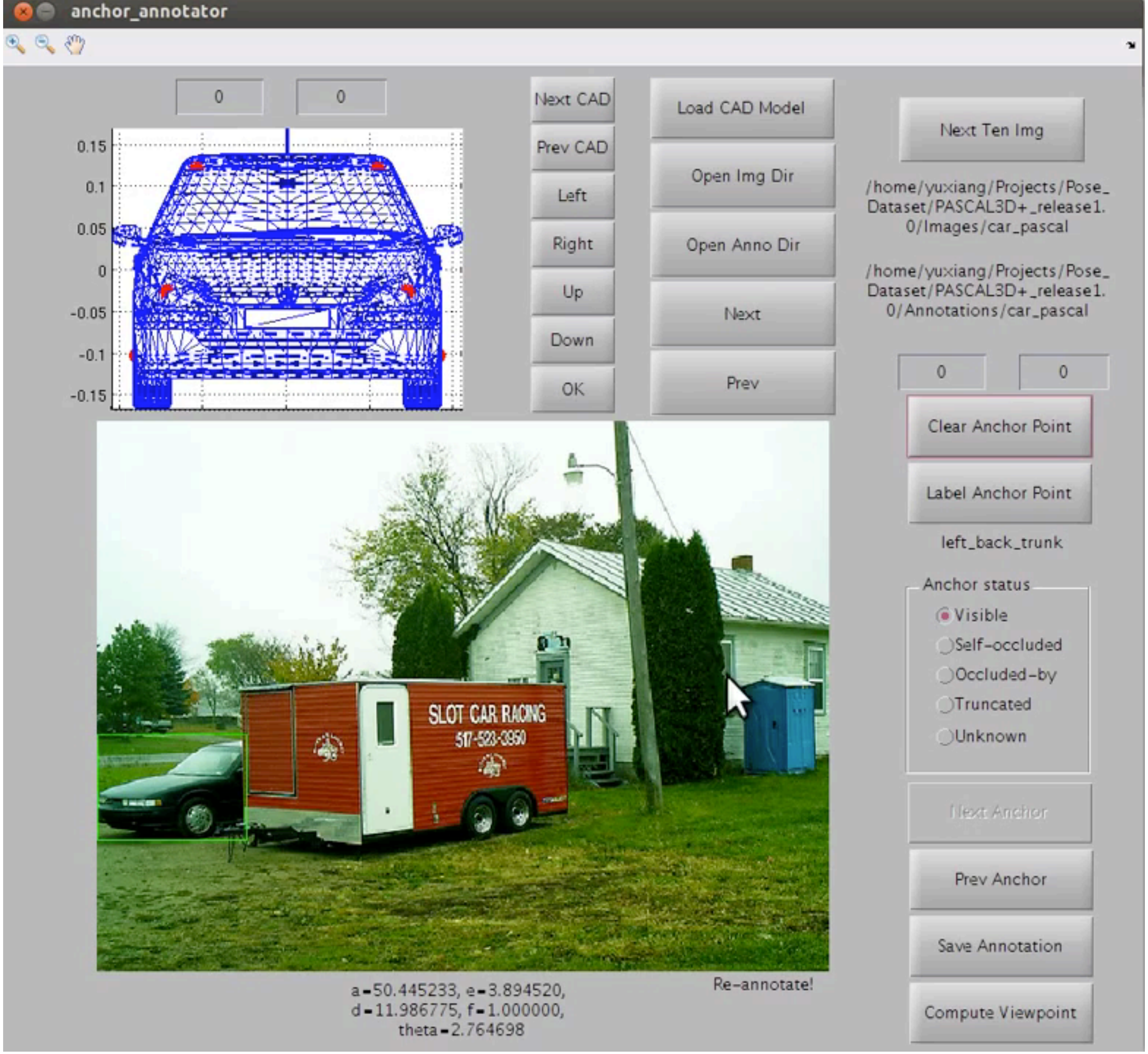


A. Bansal, et al. “Marr Revisited: 2D-3D Alignment via Surface Normal Prediction”. In CVPR 2016.

H. Su, et al. “Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views”. In ICCV 2015.



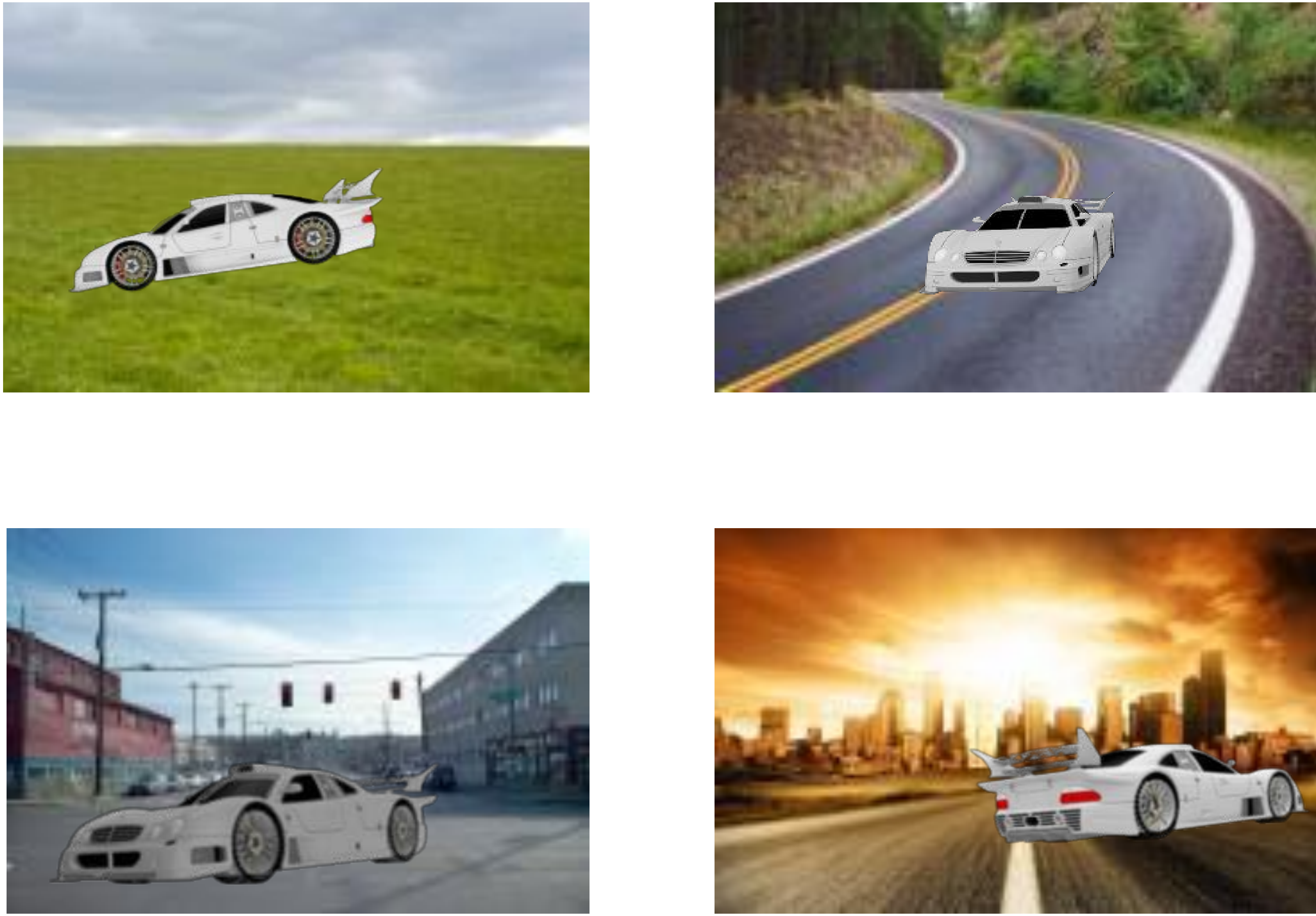
Y.Xiang, et al. " Beyond PASCAL: A benchmark for 3D object detection in the wild". In WACV 2014.



Y.Xiang, et al. " Beyond PASCAL: A benchmark for 3D object detection in the wild". In WACV 2014.


13


Volunteers of America



H. Su, et al. “Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views”. In ICCV 2015.