

Computer Vision for Embedded Systems

Yung-Hsiang Lu
Purdue University
yunglu@purdue.edu

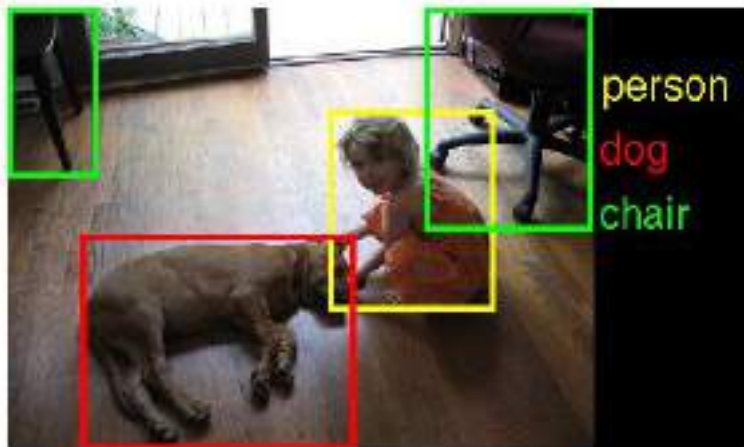


Yung-Hsiang Lu, Purdue University



Bottleneck in Supervised Learning

- Labeling data is time-consuming and expensive
- Computer vision (and many other machine learning tasks) is not perfect and needs "teachers" to provide correct answers.
- Labeling requires human effort, slow and expensive
- Acquiring "rare" events is difficult (or impossible)



ImageNet



DAVIS: Densely Annotated Video Segmentation

DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort (CVPR 2021)

"Labeling a complex scene with 50 objects can take anywhere between 30 to 90 minutes"
(for semantic segmentation)

30 frames / second x 60 second/minute x 90 minutes = 162,000

Top Data labeling Companies

Top ranked companies for keyword search: Data labeling

Export

Scale

Private Company

Founded 2016

USA

Our API provides access to human-powered data for hundreds of use cases. After sending us your data via API call, our platform through a combination of human work and review, smart tools, statistical confidence checks and machine learning checks...

<http://scale.com/>

CrowdAI

Private Company

Founded 2016

USA

At CrowdAI, we provide scalable, high-quality image annotation. We combine machine learning, computer vision and human intelligence to maximize value for self-driving car, automated drone and satellite image companies. Leaders in data insights, our...

<http://crowdai.com>

Falkonry

Private Company

Founded 2012

USA

Falkonry separates the data into unique patterns in your data and presents them as clear bands. It looks for time trends, multi-variate correlations across signals, and more. Falkonry automatically presents patterns for labeling, just like photos.

<http://falkonry.com/>

Snorkel AI

Private Company

Founded 2019

USA

The only AI platform that lets you label data programmatically, train models efficiently, improve performance iteratively, and deploy applications rapidly. Instead of hand-labeling millions of data points by hand, automatically label vast amounts of...

<https://www.snorkel.ai/>

Docugami

Private Company

Founded 2017

USA

Founded in 2018, Docugami creates SaaS solutions that harness a wide range of artificial intelligence techniques, including natural language processing, image recognition, declarative markup, and other approaches, to enable businesses of all sizes...

<http://www.docugami.com/>

icoMetrix

Private Company

Founded 2011

Belgium

icometrix provides clinicians with standardized measurements on their patients' brain MRI scans to improve personalized care of people with a neurological disorder. icometrix was founded in 2011 by Dirk Loebck & Wim Van Hecke as a spin-off of the...

<https://icometrix.com/>

Heex Technologies

Private Company

Founded 2013

France

Heex provides a data management solution that enables relevant data to move faster and more reliably from vehicles to those who need it the most via the cloud. Thanks to the pre-set triggers, the data generated is directly classified and sorted on...

<https://heex.io/>

SuperAnnotate

Private Company

Founded 2018

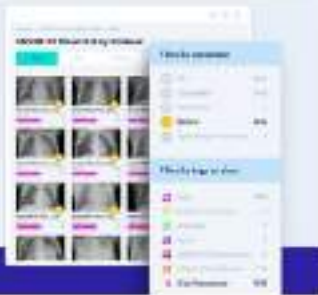
USA

The fastest annotation platform and services for training AI. A complete set of solutions for image and video annotation and an annotation service with integrated tooling, on-demand narrow expertise in various fields, and a custom neural network...

<https://www.superannotate.com/>

— USE CASE

Medical Diagnostics



Medical Diagnostics

Artificial intelligence is transforming healthcare by allowing practitioners to use big data to identify and treat diseases.

— USE CASE

Retail Automation

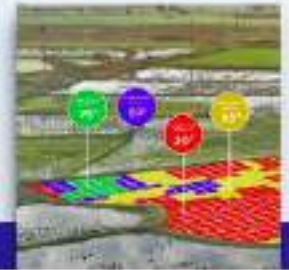


Retail Automation

As customer expectations evolve, companies are turning to AI to make the retail experience more convenient and customized.

— USE CASE

Precision Agriculture



Precision Agriculture

Technology holds great promise for solving the many challenges and inefficiencies in the production and distribution of food.

<https://info.cloudfactory.com/>



Pricing

<https://www.v7labs.com/pricing>

| Icon | Icon | Icon | Icon |
|-------------------|-------------------|-------------------|----------------------|
| | | | |
| Coastal | Team | Business | Enterprise |
| \$150 | \$450 | | |
| | | | |
| 300 hours of work | 300 hours of work | 700 hours of work | 1,000+ hours of work |
| 10,000 characters | 50,000 characters | 70,000 characters | 100k+ characters |

Crowdsourcing Annotations for Visual Object Detection (Conference on Artificial Intelligence 2012)

Desired outcomes: objects' bounding boxes in images

- Quality: tight bounding boxes
- Coverage: every object is labeled (for positive and negative examples)

Challenge: how to know the labels are correct and high quality?

- How to obtain trustable results from crowds?
- What is the right incentive to the crowds?
- "chicken-egg" problem:
 - no labels and no truth
 - no truth and cannot verify
 - cannot verify and no trusted label



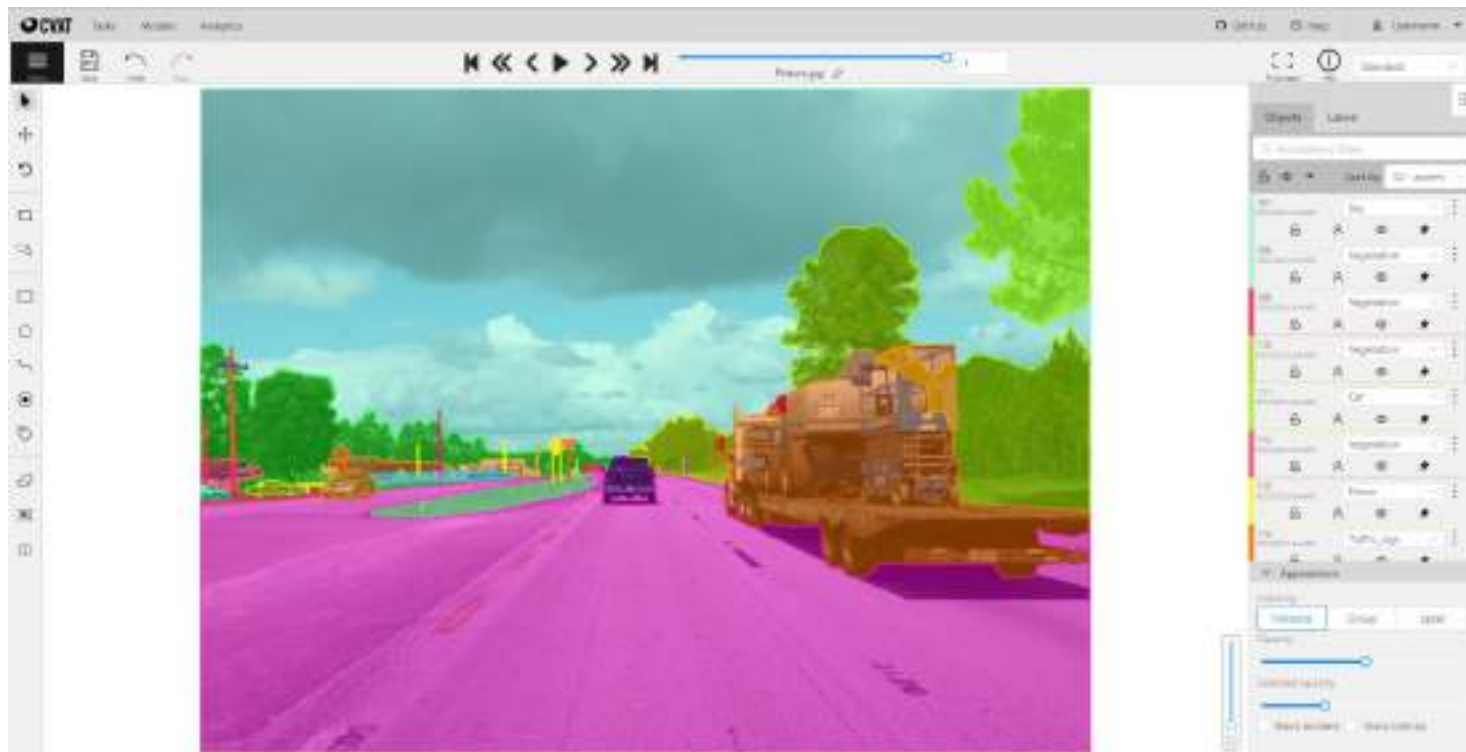
"bottle" category

Crowdsourcing for Labeling



Labeling Tools

Computer Vision Annotation Tool (CVAT)



Yung-Hsiang Lu, Purdue University

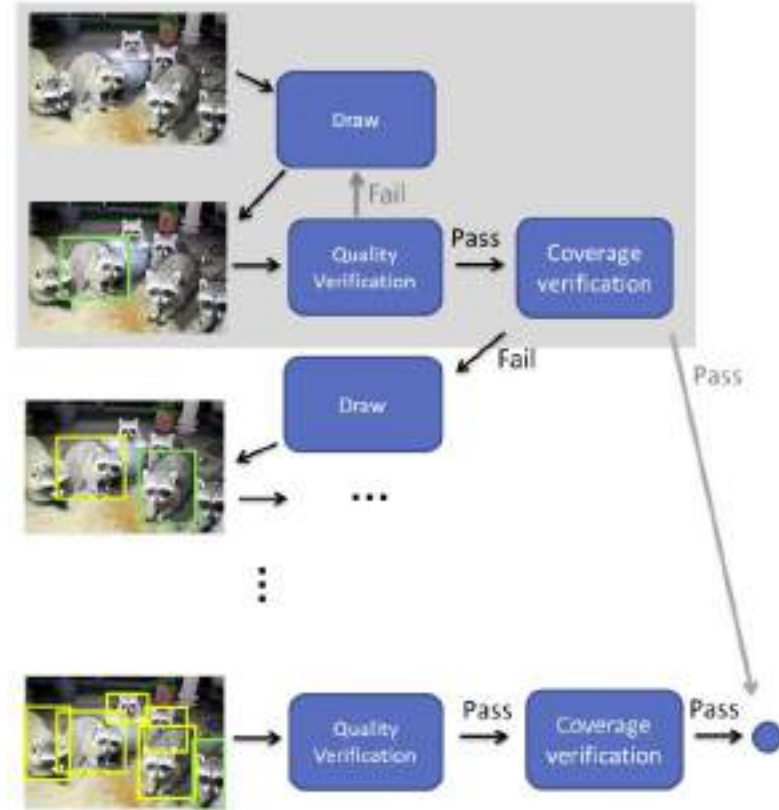


Draw-Verification Procedure

30 second/image x 1M image x 3
people
= 25,000 hours
= 1,042 days

If person x 2 hours/day
⇒ 520 days

Each person draws only one
bounding box



Procedure for Crowdsourced Labeling

Drawing Task:

1. Include all visible parts and draw as tightly as possible.
2. If there are multiple instances, include only ONE (any one).
3. Draw on a new instance if an instance has a bounding box.
4. If every instance already has a bounding box, check the check box.

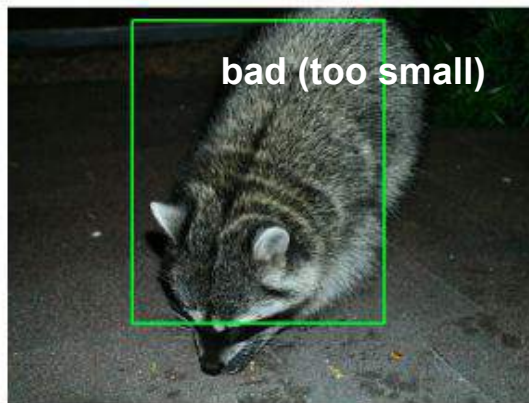
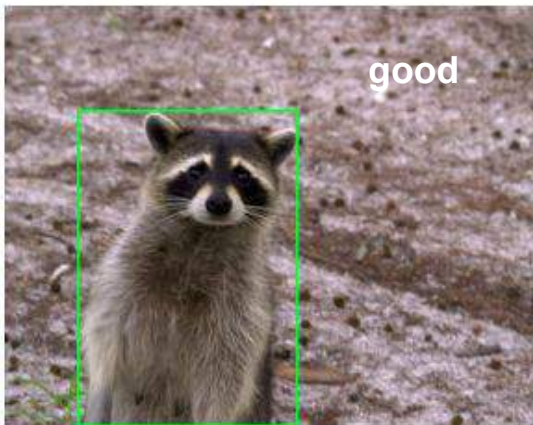
Verification Task:

1. A bounding box must include an instance of the required object.
2. A bounding box must include all visible parts and be as tight as possible.
3. If there are multiple instances, a good bounding box must include only ONE (any one)

Before given a real labeling task, a participant must pass a test.

Experiments

- 20,000 categories
- 14 million images
- 97.9% correct bounding boxes
- common errors: bounding boxes too small
- 88 second per bounding box





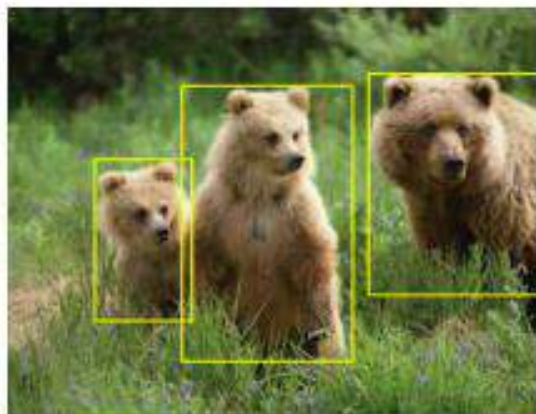
bottle



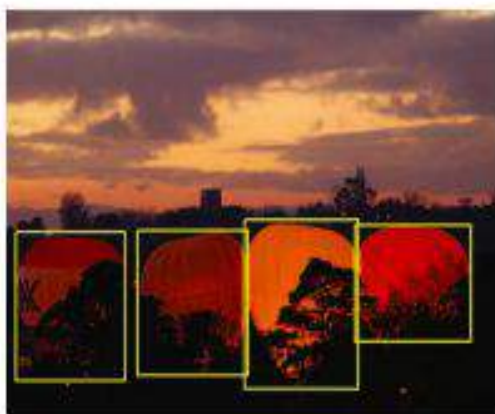
bed



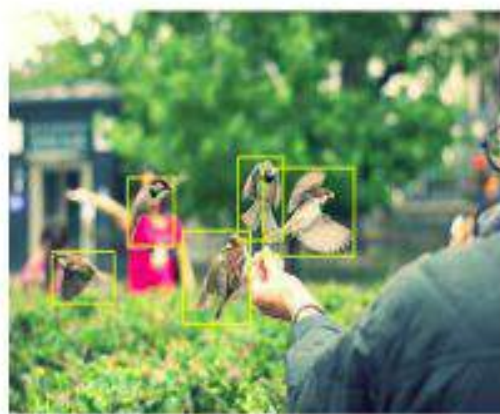
bear



bear



balloon



bird

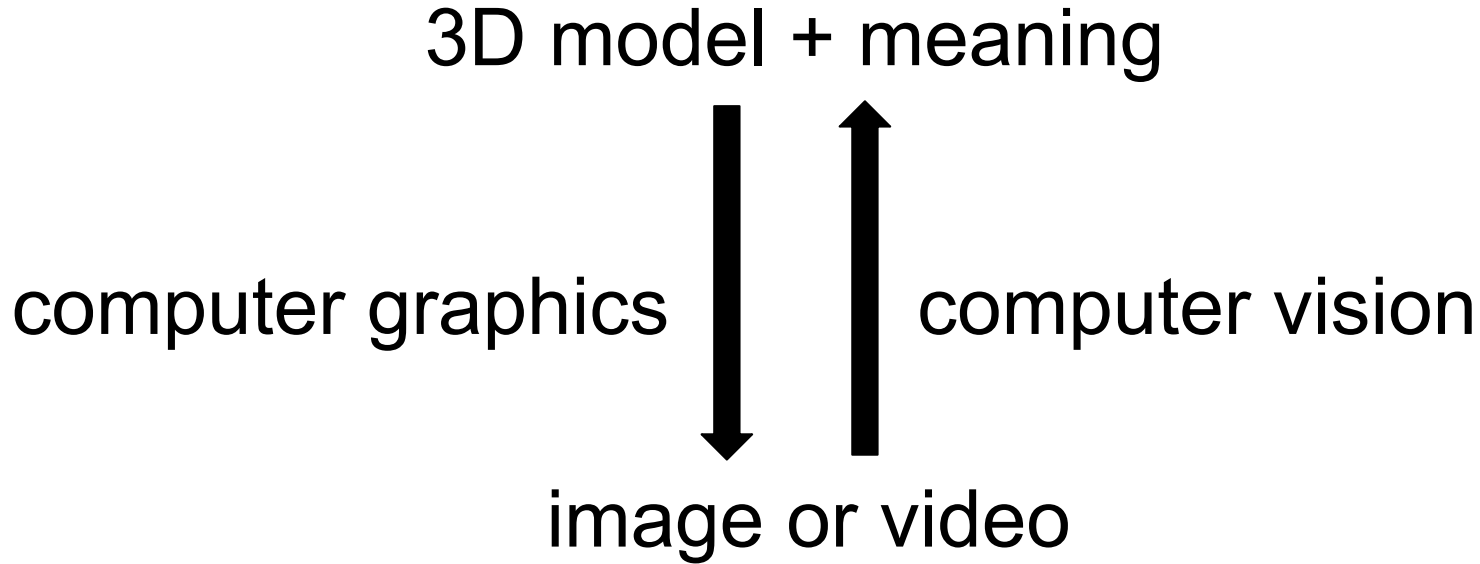


Research Environments for Machine Learning

- social networks + photo hosting services \Rightarrow easy to collect data
- web-based human interaction \Rightarrow crowdsourcing
- crowdsourcing framework \Rightarrow distribute work
- "micro payment" \Rightarrow monetary incentive
- select the right tasks \Rightarrow "ordinary" people can contribute
- well-defined problems \Rightarrow crowd can participate

Synthesized Data + Labels




Vision and Graphics are Reverse Problems



Computer Graphics: also called CG, animation, special effects ... in movies

Computer Graphics and Vision

Graphics

| Meaning | → | Visual Data |
|--|---------------|---|
| Objects: Airplane, Human, Vehicle, Tree | Vision | |
| Season, Time, environment: Autumn, football game | |  |
| Actions: Walking, Running, Flying | |  |
| Relationship of objects: Above, Fighting | |  |

<https://brianmmurray.wordpress.com/2013/02/28/feeling-anxious-spend-time-with-nature/>

<https://www.jconline.com/picture-gallery/sports/2020/10/31/zander-horvath-look-purdue-football-running-back/6103553002/>

<https://www.rogerebert.com/reviews/ip-man-4-the-finale-movie-review-2019>

Why to synthesize data?

- rare events (disasters, accidents, endangered species)
- dangerous environments (fire, pedestrians jumping into traffic)
- seasonal delays (evaluate vision's responses to winter)
- augmented reality
- flexible viewing angles
- scale up to interactions of multiple objects
- repeatable evaluation
- additional information: depth, speed, weight and volume

Sim4CV: A Photo-Realistic Simulator for Computer Vision Applications *International Journal of Computer Vision (2018)*

| | | | |
|--|---|--|---|
| <p>Object Tracking</p>  <p>● ● ● / ●</p> | <p>Pose Estimation</p>  <p>● ● / ● ● ●</p> | <p>Object Detection</p>  <p>● ● / ● ●</p> | <p>Action Recognition</p>  <p>● ● ● / ● ● ●</p> |
| <p>Autonomous Navigation</p>  <p>● ● ● ● / ● ● ●</p> | <p>3D Reconstruction</p>  <p>● ● ● ● / ● ● ●</p> | <p>Crowd Understanding</p>  <p>● ● ● / ● ● ● ●</p> | <p>Urban Scene Understanding</p>  <p>● ● ● / ● ● ●</p> |
| <p>Indoor Scene Understanding</p>  <p>● ● ● ● / ● ●</p> | <p>Multi-agent Collaboration</p>  <p>● ● ● / ● ● ● ● ●</p> | <p>Human Training</p>  <p>● ● ● ● ●</p> | <p>Aerial Surveying</p>  <p>● ● ● ● / ● ● ● ●</p> |

- Image
- Depth/Multi-View
- Video
- Segmentation/Bounding Box
- Image Label
- User Input
- Physics
- Camera Localization

Factors to consider in synthesized data

- Photo-realistic or not
- Physics (inertia, gravity, turbulence, mass, size, elasticity)
- Weather (wind, rain, fog, sun, shadow)
- Time
- Power and energy
- Surface properties (e.g., reflection)
- Human behavior
- Interaction with physical components

Applications



Photorealistic Image Synthesis for Object Instance Detection

IEEE International Conference on Image Processing 2019

Physics-based modeling:

- scattering
- refraction and reflection
- diffuse
- usually very slow



Photorealistic images synthesized by the proposed approach



Real images from LineMod [4] and Rutgers APC [2] datasets

Scene 1



Scene 2



Scene 3



Scene 4



Scene 5



Scene 6



Virtual Worlds as Proxy for Multi-Object Tracking Analysis

Adrien Gaidon, Qiao Wang, Yohann Cabon, Eleonora Vig

CVPR 2016

Is this real or virtual?



Is this real or virtual?



Segmentation and Depth



Challenge in Data Labeling (for Video)

- volume of data (30 frames per second, multiple objects)
- movement
- occlusion
- diversity

Meanwhile, synthesizing data also has many challenges:

- photorealism
- time-consuming to generate high-quality data
- Playing video games need human players

Generate Virtual World

1. acquire real-world data as a starting point for calibration
2. clone this real-world data into a virtual world
3. generate synthetic sequences with different weather conditions
4. create ground truth annotations
5. evaluate the “usefulness” of the synthetic data



Figure 2: Frames from 5 real KITTI videos (left, sequences 1, 2, 6, 18, 20 from top to bottom) and rendered virtual clones (right).



Figure 3: Simulated conditions. From top left to bottom right: clone, camera rotated to the right by 15° , to the left by 15° , “morning” and “sunset” times of day, overcast weather, fog, and rain.



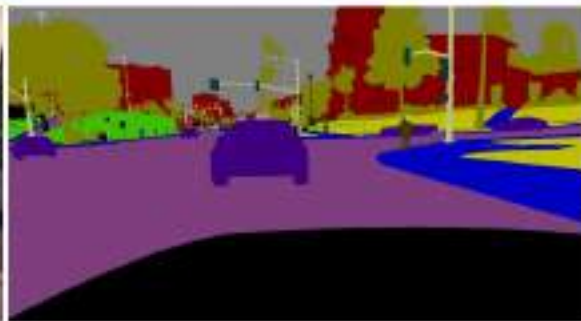
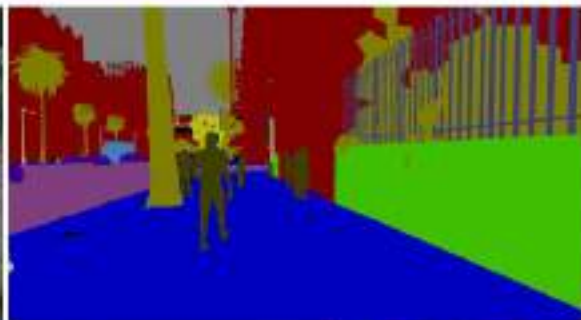
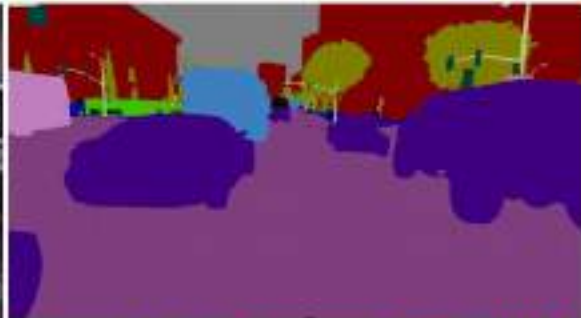
Figure 5: Predicted tracks on matching frames of two original videos (top) and their synthetic clones (bottom) for both DP-MCF (left) and MDP (right). Note the visual similarity of both the scenes and the tracks. Most differences are on occluded, small, or truncated objects.

Playing for Data: Ground Truth from Computer Games

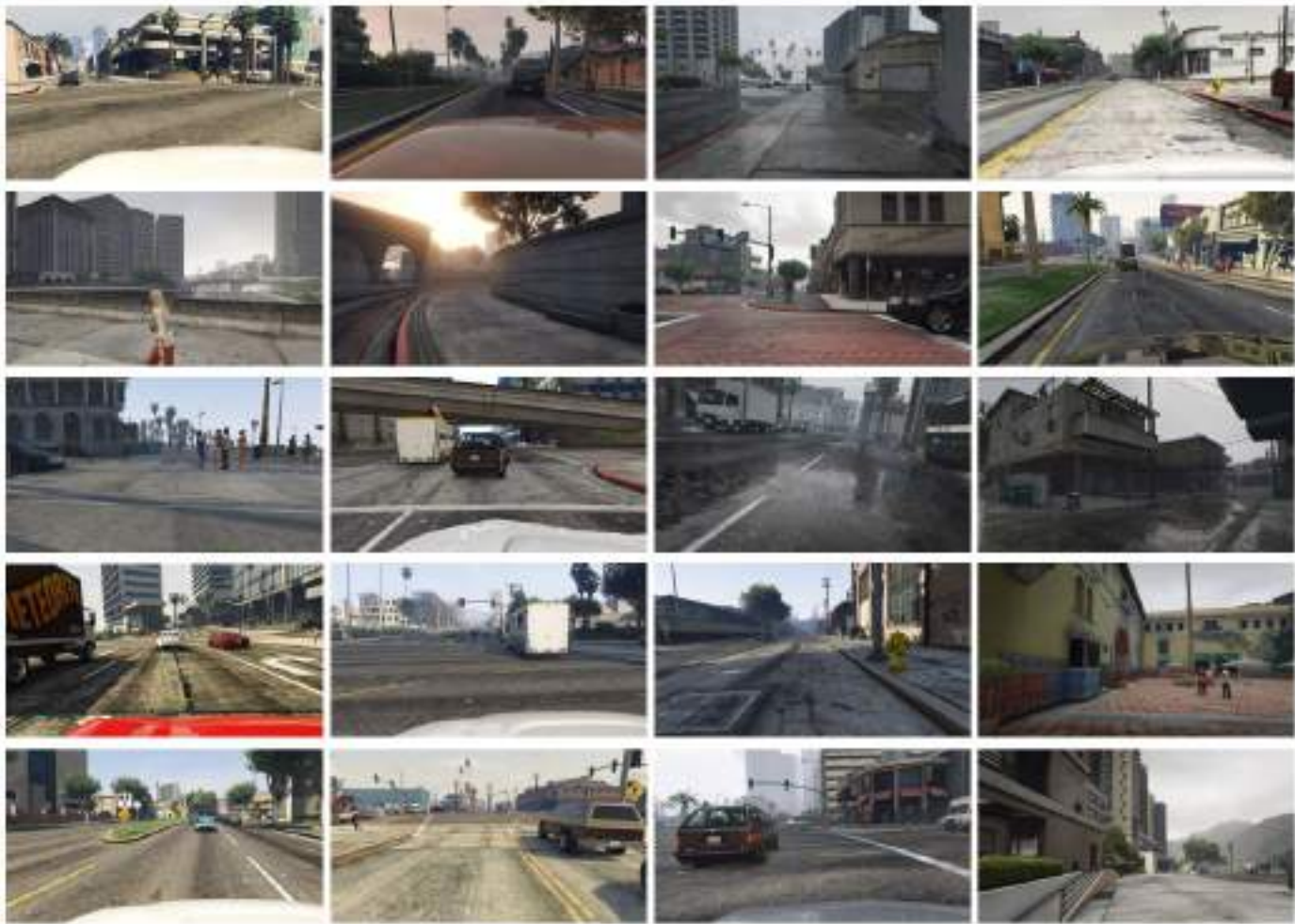
Stephan R. Richter, Vibhav Vineet, Stefan Roth, Vladlen Koltun
ECCV 2016

Capture Graphics Commands in Game

- open-source games lack details like commercial games
- source code of commercial game engine unavailable
- create pixel-wise semantic labels
- record and reproduce rendering commands at operating systems
- method: intercept communication between software and hardware
 - identify relevant function calls
 - identify hardware resources
 - format for annotation
- 25,000 images, 49 hours for labeling, about 7 seconds / image







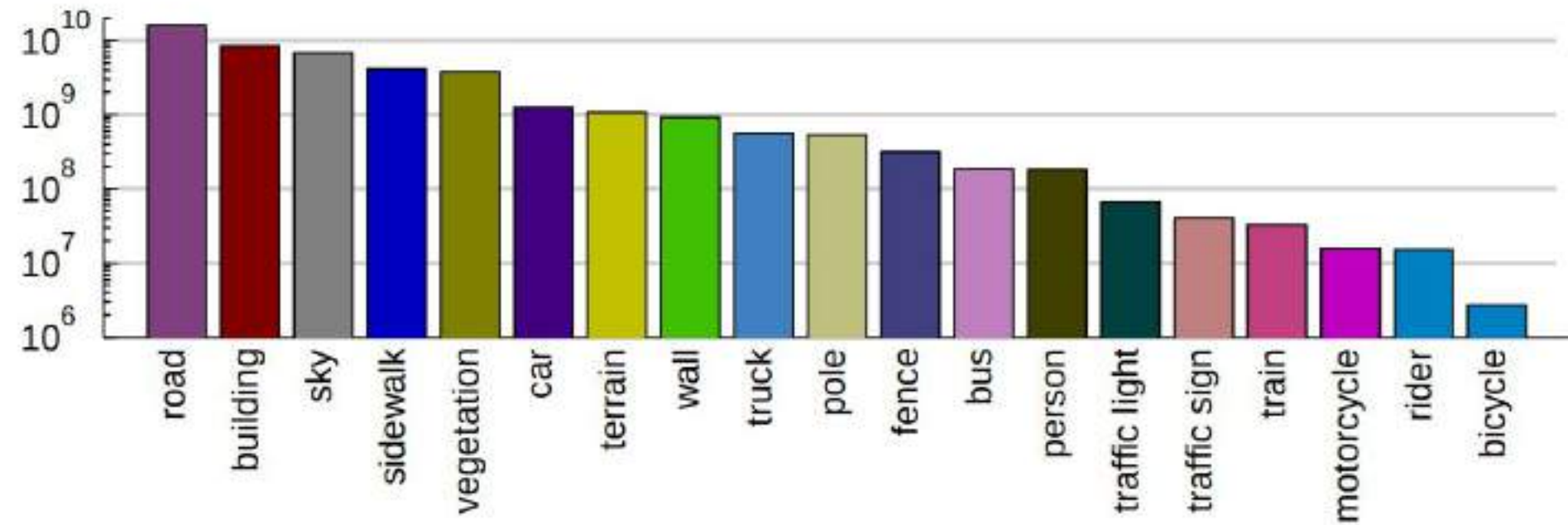


Fig. 4. Number of annotated pixels per class in our dataset. Note the logarithmic scale.

| | #pixels [10^9] | annotation density [%] | annotation time [sec/image] | annotation speed [pixels/sec] |
|--------------------------|-----------------------|---------------------------|--------------------------------|----------------------------------|
| GTA5 | 50.15 | 98.3 | 7 | 279,540 |
| Cityscapes (fine) [11] | 9.43 | 97.1 | 5400 | 349 |
| Cityscapes (coarse) [11] | 26.0 | 67.5 | 420 | 3095 |
| CamVid [8] | 0.62 | 96.2 | 3,600 | 246 |
| KITTI [39] | 0.07 | 98.4 | N/A | N/A |