An Introduction to Computer Vision (yeah, right!)

D.A. Forsyth, University of Illinois at Urbana Champaign

SOME APPLICATION SLIDES MISSING

Caveats

- Compression
  - Like a 3hr “Introduction to Mathematics” or “Introduction to Physics”
    - there are missing bits!
  - There are likely missing references in the slides: look at chapters
- Scope
  - entire disciplines are hardly touched
- On Web Page, see
  - Slides
  - Expanded versions of slides on topics
  - Corresponding movies on topics
  - Chapters from TNCV late draft

Structure

- What Computer Vision does
- What Images are Like
- Regression and Classification
- Geometry and Physics
  - or, more details on what images are like
- The uses of correspondence
  - or, elementary 3D, tracking, and so on
- Where it’s all going

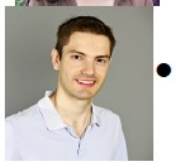
Vision group at Illinois

- David Forsyth
  - Fulton Watson Copp chair; Marr prize, 1993; 2 ex students with Marr prizes; IEEE Tech. Achievement, Fellow; ACM Fellow; EIC IEEE TPAMI; 4 patents

- Jim Rehg
  - HCESC director; multiple famous ex-students, best paper awards; 26 patents

- Derek Hoiem
  - best paper, CVPR 2006; ACM Doctoral Dissertation honorable mention; Sloan Fellow; PAMI-TC Young Researcher; major startup

- Lana Lazebnik
  - Microsoft Faculty Fellow; Sloan Fellow; Koenderink Prize (2016); EIC IJCV



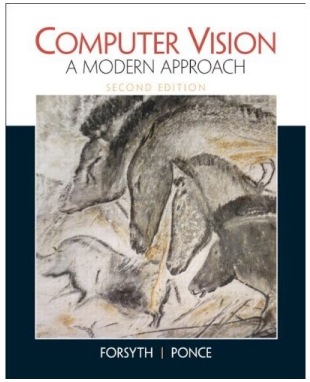
- Alex Schwing
  - Visual learning, segmentation and GAN models

Saurabh Gupta
- Linking visual sensing to motion

Liangyan Gui
- Understanding human movement

Shenlong Wang
- Simulation and sensing for autonomous vehicles

Yuxiong Wang
- Learning to detect and classify with very little data



Vision group

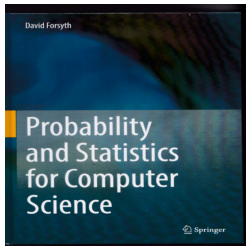
The New Computer Vision

D.A. Forsyth

Likely about 2024 ~~2026~~ 2027

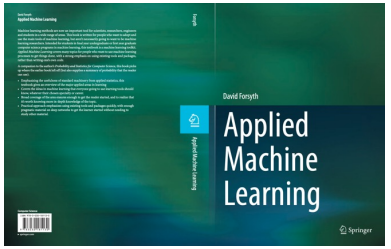
Cover design opportunity!

Well-known ex-students:
Lana Lazebnik (UIUC)
Tamara Berg (UNC)
Pinar Duygulu (Hacettepe U.)
Ian Endres
Ali Farhadi (UW)
Varsha Hedau
Nazli Ikizler (Hacettepe U.)
Brett Jones
Kevin Karsch
Zicheng Liao
Deva Ramanan (CMU)
Raj Sodhi
Gang Wang (now Alibaba)
Amin Sadeghi
Zicheng Liao (Zhejiang U.)
Anand Bhattad (JHU)



Startups:

Lightform
Revery.ai
Reconstruct
Depix



Top publications

Categories English

	Publication	h5-index	h5-median
1.	Nature	488	745
2.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	440	689
3.	The New England Journal of Medicine	434	897
4.	Science	409	633
5.	Nature Communications	375	492
6.	The Lancet	368	678
7.	Neural Information Processing Systems	337	614
8.	Advanced Materials	327	420
9.	Cell	320	482
10.	International Conference on Learning Representations	304	584
11.	JAMA	298	498
12.	Science of The Total Environment	297	436
13.	IEEE/CVF International Conference on Computer Vision	291	484
14.	Angewandte Chemie International Edition	281	361
15.	Nature Medicine	274	474
16.	Journal of Cleaner Production	272	359
17.	International Conference on Machine Learning	268	424
18.	Chemical Reviews	267	461
19.	Proceedings of the National Academy of Sciences	267	405
20.	IEEE Access	266	364

.com

Home / Best Conferences - Computer Science

Best Computer Science Conferences

The ranking of leading conferences for Computer Science was prepared by Research.com, one of the primary websites for Computer Science research offering accurate data on scientific output since 2014.

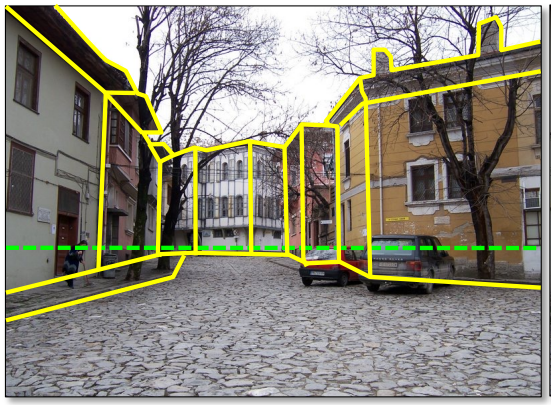
The spot on the list is based on Impact Score data collected on 21-11-2023. It was based on a careful analysis of as much as 3,040 conference profiles and websites. Show more

All resear | All publisl | All countr | Paper submission open

Search by name

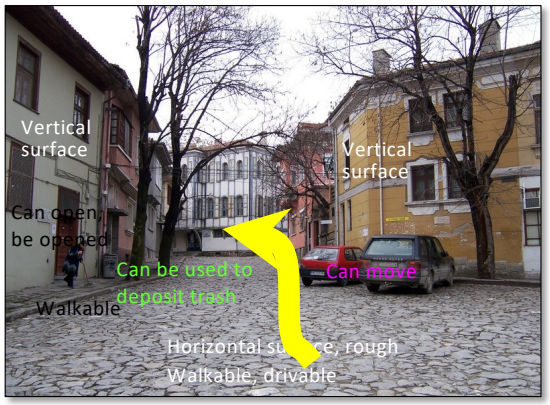
Rank	Conference Details	Impact Score
1 Rank	**Computer Vision and Pattern Recognition** 18-06-2023 - 22-06-2023 - **Vancouver**	62.60
2 Rank	**International Conference on Computer Vision** 11-10-2021 - 11-10-2021 - **Montreal**	43.80
3 Rank	**Neural Information Processing Systems** 12-12-2023 - 14-12-2023 - **New Orleans**	40.60
4 Rank	**International Conference on Learning Representations** 01-05-2023 - 05-05-2023 - **Kigali**	38.10
5 Rank	**AAAI Conference on Artificial Intelligence** 07-02-2023 - 14-02-2023 - **Washington DC**	35.00
6 Rank	**International Conference on Machine Learning** 17-07-2022 - 23-07-2022 - **Baltimore**	30.10
7 Rank	**European Conference on Computer Vision** 29-09-2024 - 04-10-2024 - **Milan**	29.60
8 Rank	**Meeting of the Association for Computational Linguistics** 22-05-2022 - 27-05-2022 - **Dublin**	28.90
9 Rank	**Empirical Methods in Natural Language Processing** 06-12-2023 - 10-12-2023 - **Singapore City**	21.90
10 Rank	**International Joint Conference on Artificial Intelligence** 23-07-2022 - 29-07-2022 - **Vienna**	21.10

What you can extract from an image



Geometric information

Geometric information
Semantic information



Geometric information
Semantic (?) information – *affordances*

Vision for action

Trad. View:

Images are fundamentally ambiguous!

...still, vision is hard even for humans

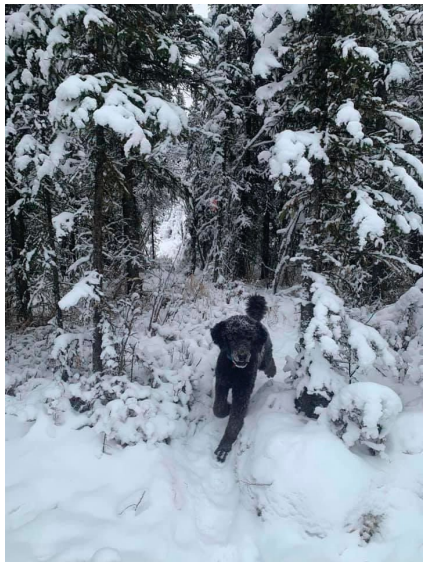


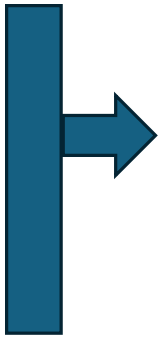
Image source

...still

Image source

The cornerstones

- Geometry and Physics
  - Cameras, multiple cameras, etc
  - Illumination, shading, color, etc
- For low compute regimes:
  - Fast and distinctive local image representations
  - Correspondence
- For high compute regimes:
  - Learned image/video representations
  - Learned decoding procedures
- Data Science
  - Acquiring and cleaning datasets
  - Clustering and matching
  - Very large scale optimization



The border between these two is surprisingly impermeable.

Current computing demands are absurd.

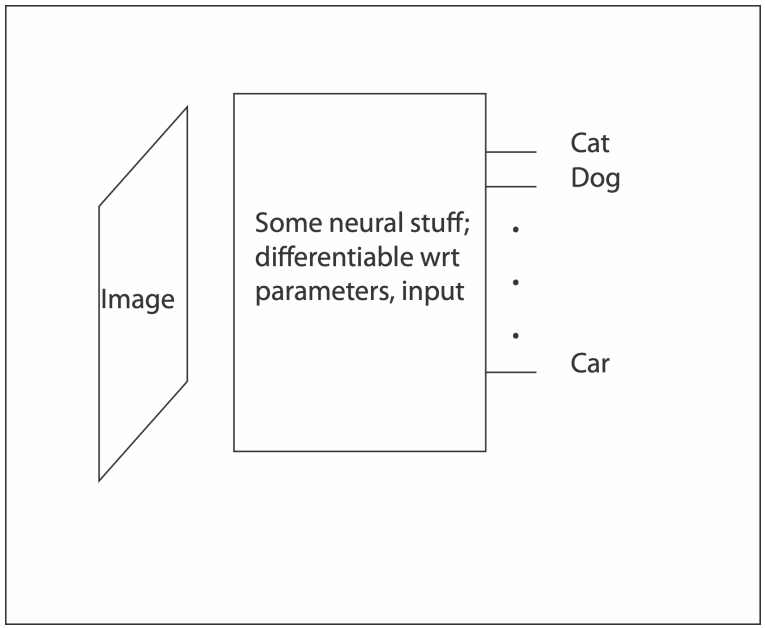
They are justified only by our ignorance.

What we are good at

- Classification and regression
- Examples:
  - image classification
  - object detection
  - depth/normal/X from single image
- Bolt together:
  - learned representation
  - learned decoding
  - data engineering

Mostly, big compute, but you
can do a moderate job of classification
and detection in small compute

Classification



Detection

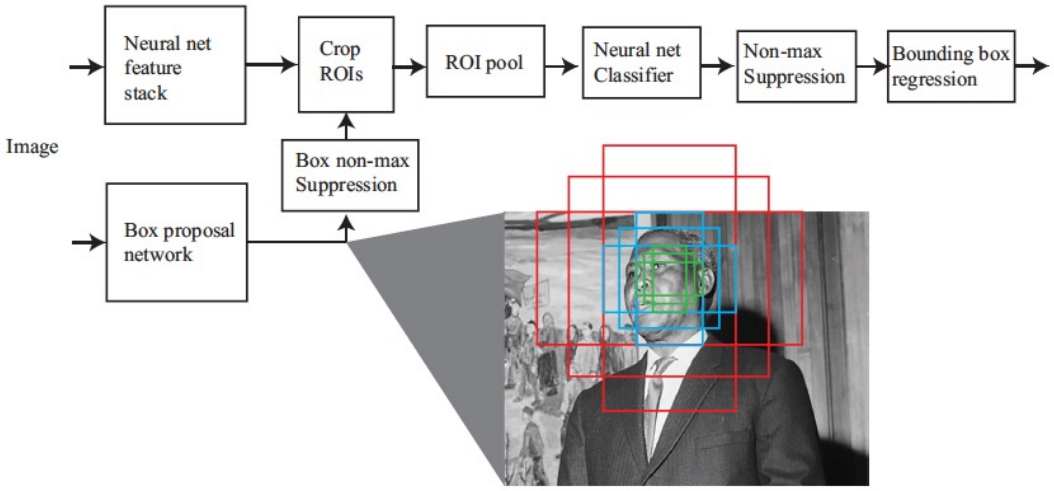
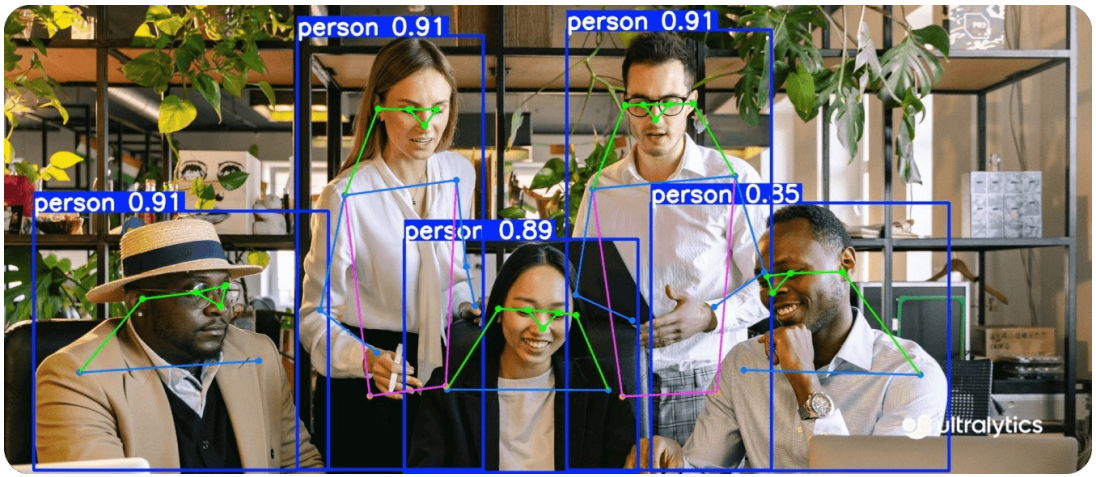


FIGURE 18.8: *Faster RCNN uses two networks. One uses the image to compute "objectness" scores for a sampling of possible image boxes. The samples (called "anchor boxes") are each centered at a grid point. At each grid point, there are nine boxes (three scales, three aspect ratios). The second is a feature stack that computes a representation of the image suitable for classification. The boxes with highest objectness score are then cut from the feature map, standardized with ROI pooling, then passed to a classifier. Bounding box regression means that the relatively coarse sampling of locations, scales and aspect ratios does not weaken accuracy.*

Detection and localization in 2D



YOLOv11 documentation by Ultralytics

Depth predictors work really well

What we are good at

- Segmentation
- Examples:
  - Which pixels are the cat?
  - break image into objects, known or unknown
  - depth/normal/X from single image
- Bolt together:
  - learned representation
  - learned decoding
  - clustering
  - data engineering

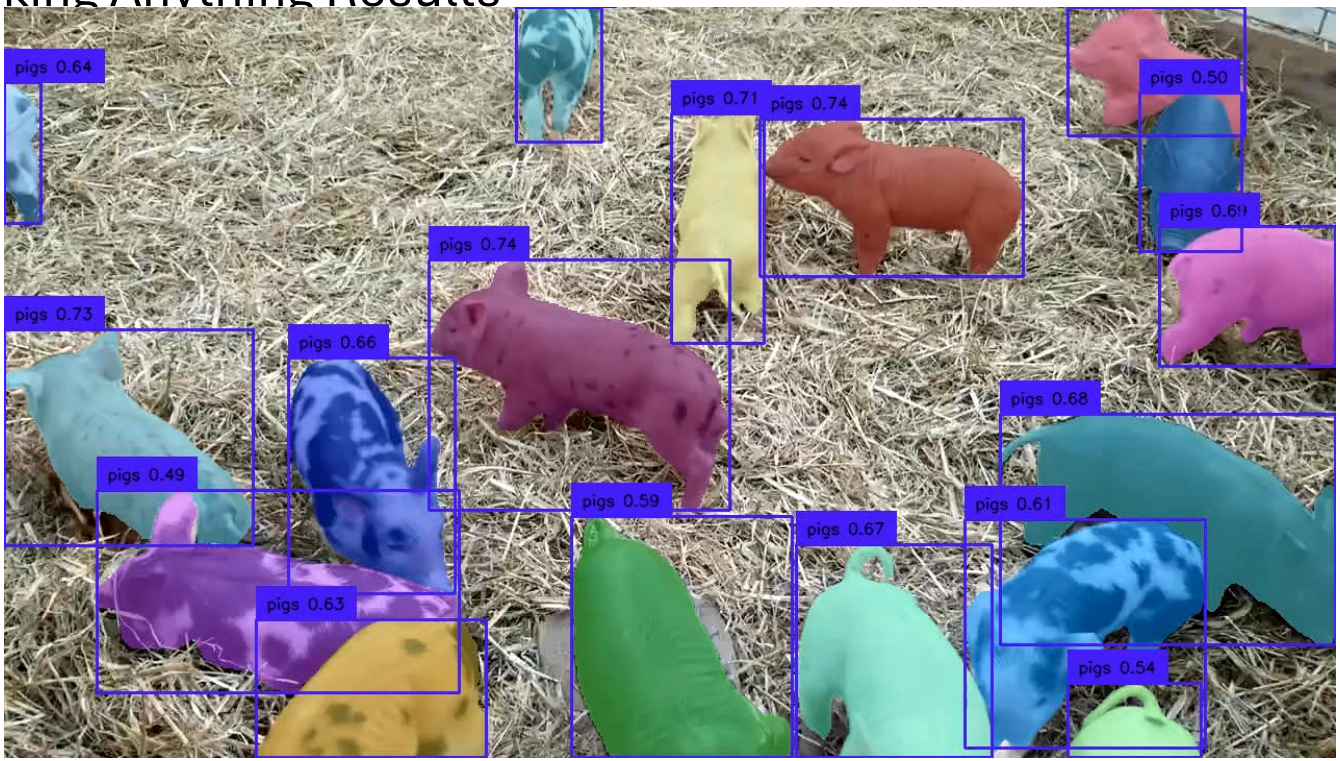
Mostly, big compute

What we are good at

- Tracking and correspondence
- Examples:
  - tracking people
  - tracking by/and detection
  - point tracking
- Bolt together:
  - learned representation
  - learned decoding
  - data engineering

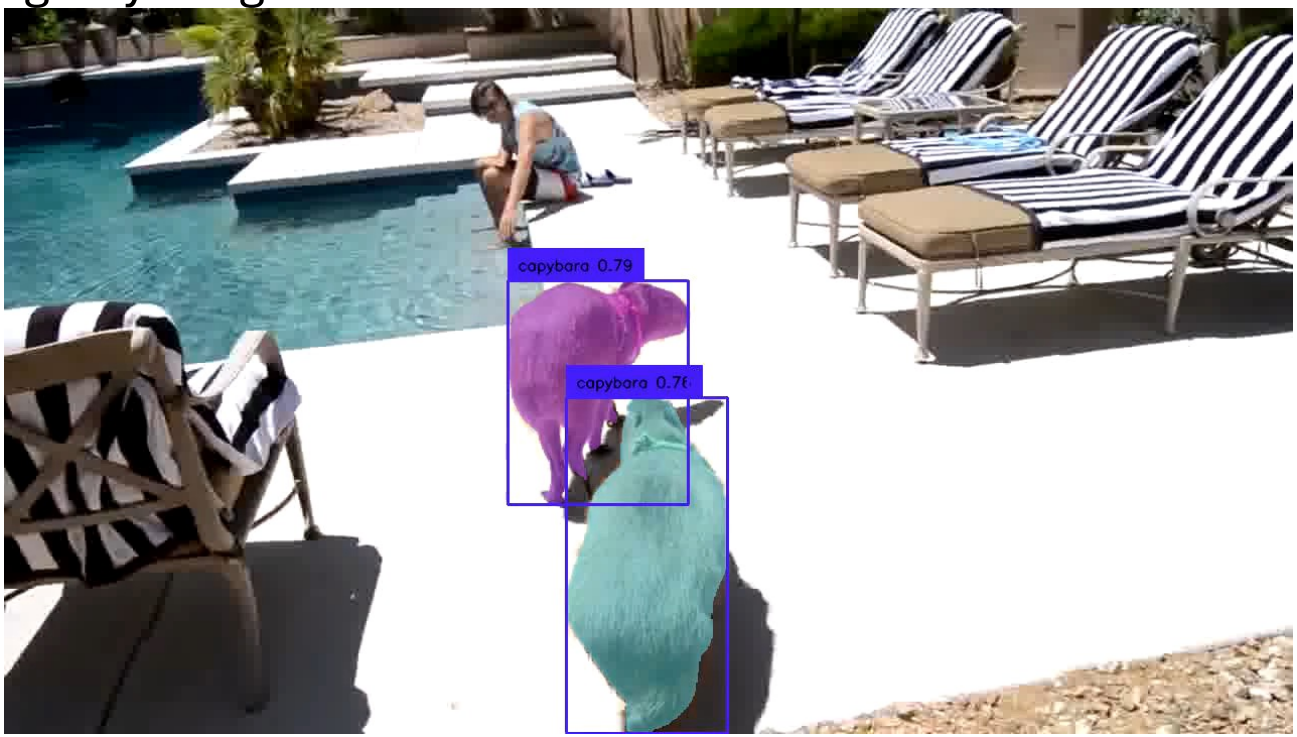
Mostly, big compute, but you
can do a moderate job of some kinds of
tracking in very small compute

Tracking Anything Results



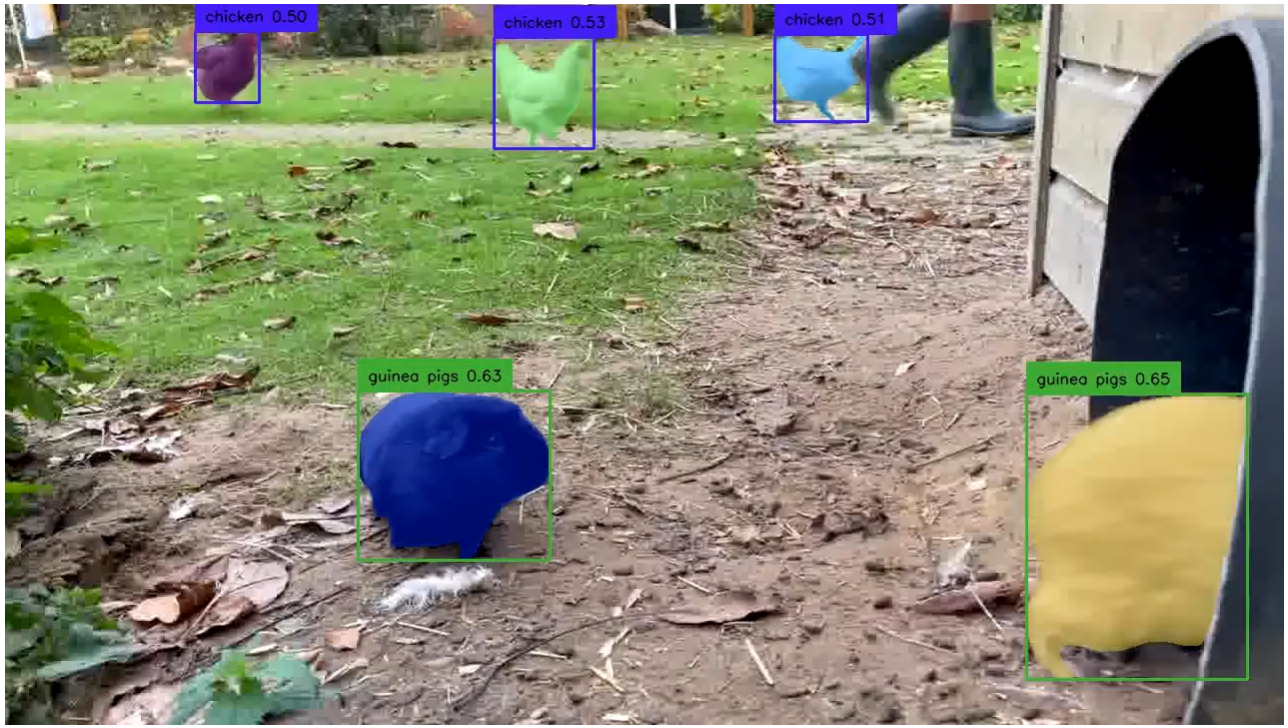
Cheng et al.; DEVA: Tracking anything with decoupled video segmentation; 2023

Tracking Anything Results



Cheng et al.; DEVA: Tracking anything with decoupled video segmentation; 2023

Tracking Anything Results



Cheng et al.; DEVA: Tracking anything with decoupled video segmentation; 2023

What we are good at

- Linking the visual and the textual
- Examples:
  - Vision Language Models
- Bolt together:
  - learned representation of images
  - learned representation of text
  - data engineering

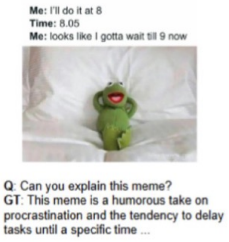
No small compute here!

Evaluation strategies - I

1) *Answer Matching*

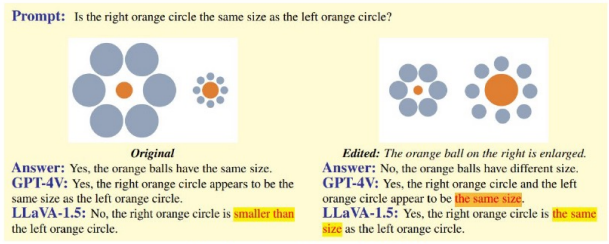
ST-VQA [20]

Evaluation: Accuracy
Metric: Exact Match
Format: Specific short-form answers, such as objects…



MM-Vet [245]

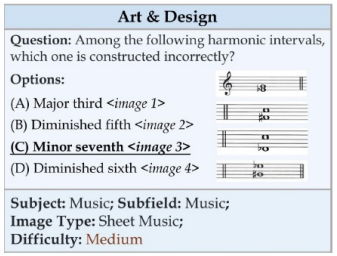
Evaluation: Average Score
Metric: ROUGE, LLM Eval
Format: Long-form open-ended answers



HallusionBench [67]

Evaluation: Accuracy / Precision / Recall
Metric: Exact Match
Format: Yes/No question

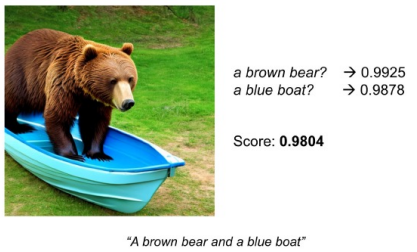
2) *Multiple Choice*



MMMU-Pro [248]

Evaluation: Accuracy
Metric: Exact Choice Match
Format: Multiple choice questions

3) *Image-Caption Similarity*



T2I-CompBench [63]

Evaluation: Average Similarity
Metric: CLIPScore, GenEval
Format: text to image generation

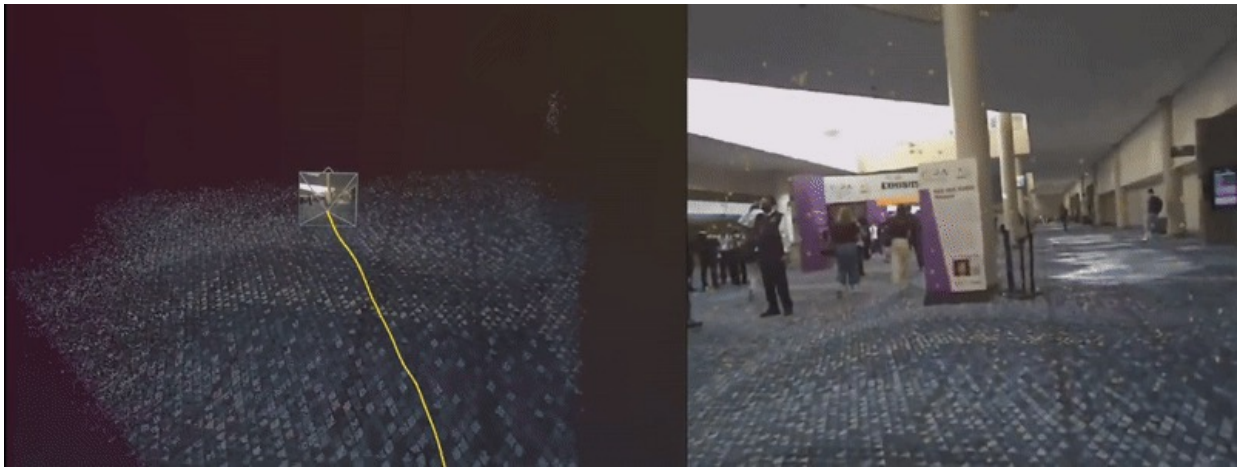
A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges, Li et al, 2025

What we are good at

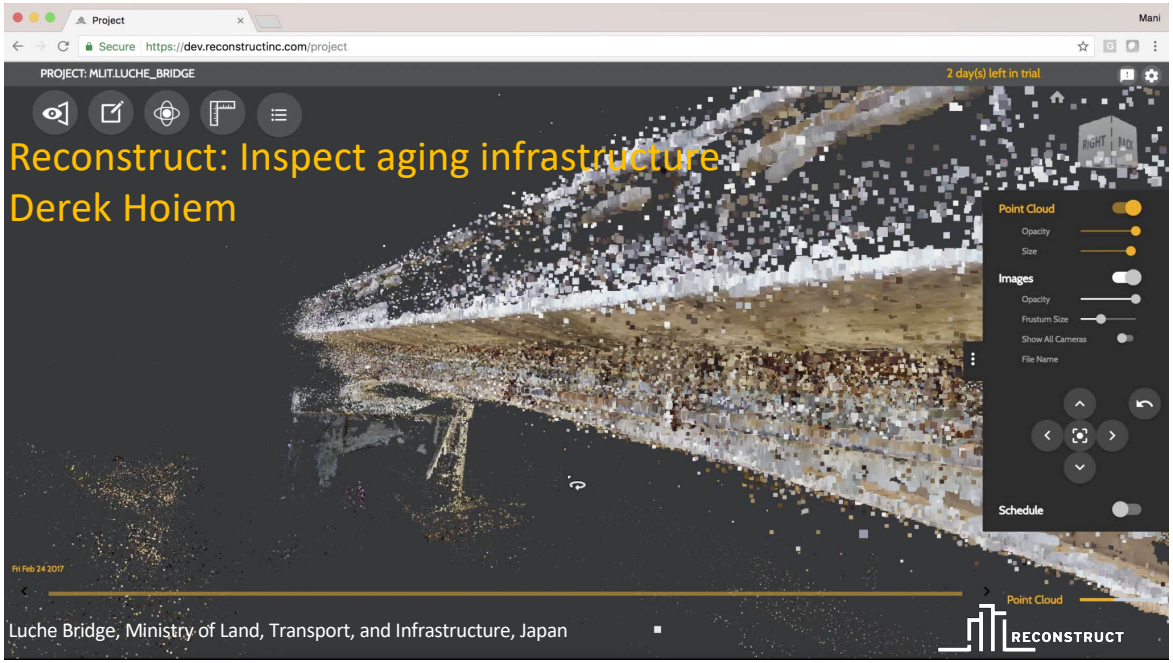
- 3D Reasoning
- Examples:
  - localization
  - reconstruction
  - odometry
- Bolt together:
  - in the low compute world
    - Fast distinctive image representations; correspondence; geometry
  - in the high compute world
    - learned image representations; data engineering; learned decoding

Very different kinds of activity

Visual odometry



https://github.com/MAC-VO/MAC-VO/blob/main/asset/ICRAvideo.gif


Project
Mani
Secure https://dev.reconstructinc.com/project
PROJECT: MLIT.LUCHE_BRIDGE
2 day(s) left in trial
Reconstruct: Inspect aging infrastructure
Derek Hoiem
Point Cloud
Opacity
Size
Images
Opacity
Frustum Size
Show All Cameras
File Name
Schedule
Fri Feb 24 2017
Point Cloud
Luche Bridge, Ministry of Land, Transport, and Infrastructure, Japan
RECONSTRUCT

Reconstruct: Align reality to plans for construction management

Derek Hoiem

New technologies

- Generating images/video
- Unsupervised learning
  - (not really new; but now actually becoming very useful)
- Learned 3D reconstructions