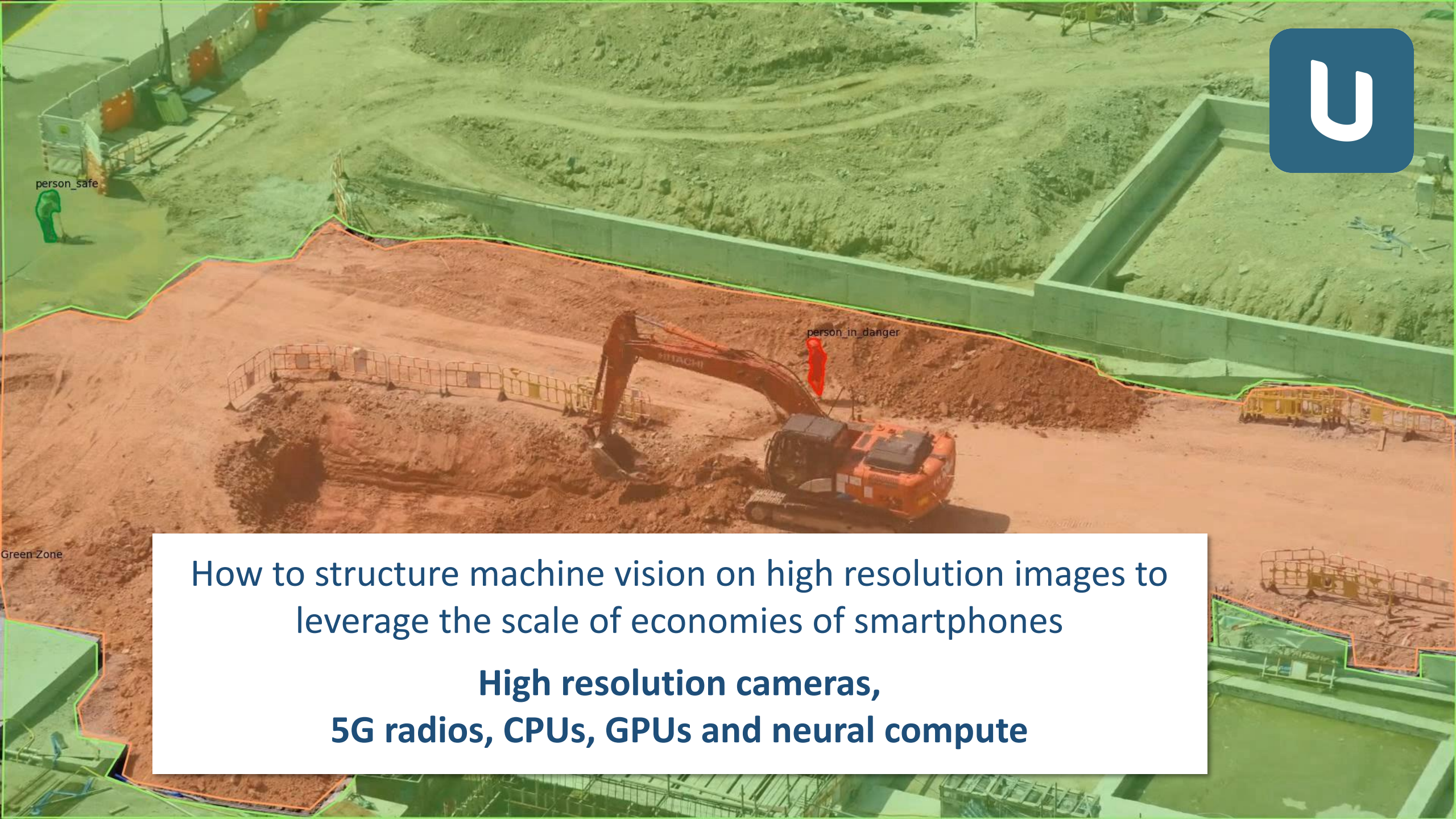


How to structure machine vision on high resolution images to leverage the scale of economies of smartphones

High resolution cameras,
5G radios, CPUs, GPUs and neural compute

Industrial customers

- Frontline Applications
- Embedded Applications
- Cloud Applications

The Umajin Platform

umajin

- Powers Applications
- Supports Customisation
- Supports Integration

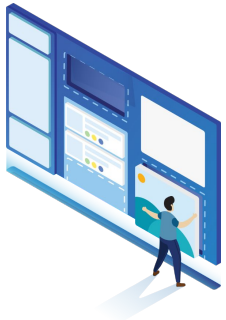
Enterprise devices

Enterprise systems

Umajin: Platform for Machine Vision & AI

All parts of the Umajin Platform contribute to delivering a flexible pipeline which allow new innovations in AI to be quickly applied in applications using your valuable datasets

- Umajin Server is a turnkey database and API layer making it easy to build applications and to collect raw data securely, to store, clean, classify and label data, to queue up training models based on your data using the latest machine learning approaches and finally distribute them using Umajin Cloud
- Umajin Editor and the Umajin Native Runtime on devices (computers, smartphones and embedded devices) allows high performance applications using cameras, lighting, CPU, GPU and TPU to be updated and distributed to the edge
- Umajin Cloud is a content management system that manages versions of assets. The content can be distributed from global content distribution networks right through to point to point encrypted connections



Umajin Server

Umajin Editor

Umajin Cloud

Transmit data for analysis and storage

Design, Develop & Deploy applications

Deploy application updates and models

Umajin on Devices

Umajin: Machine Vision Cameras

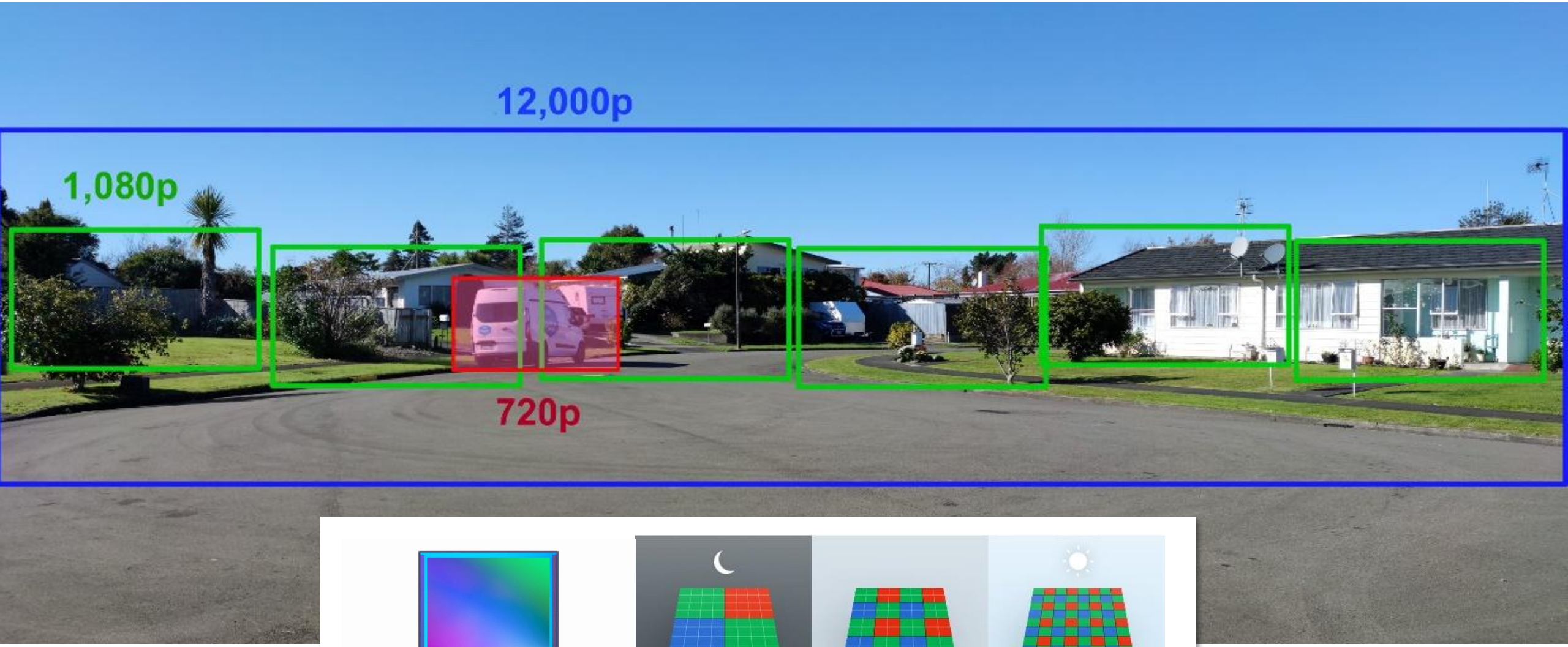
Leverage Continual Improvement in Smartphone price/performance

- 1.43 Billion smartphones sold in 2021 – leverage the scale of economies, 5G radio, compute and high quality cameras
- Over 6.8 Billion active smartphones

Hybrid architecture with on device event processing for broad phase detection of people, vehicles and objects

Edge severs to apply classification and specialised pattern matching like OCR. This enables 10,000 scale camera deployments with cost effective bandwidth and compute.

Resolution: 108MP now 200MP!!



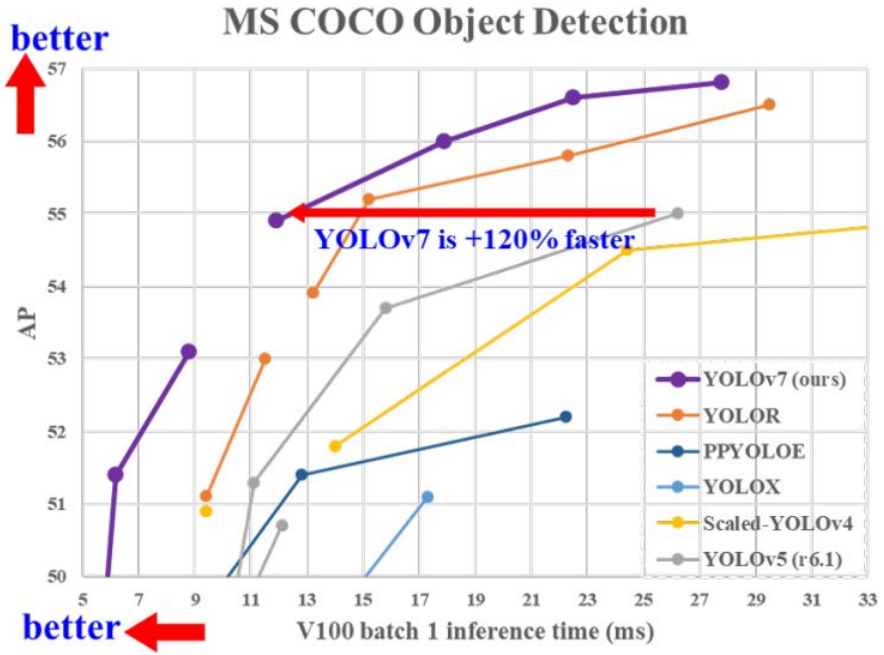
Algorithms: Object detector breakthroughs

YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

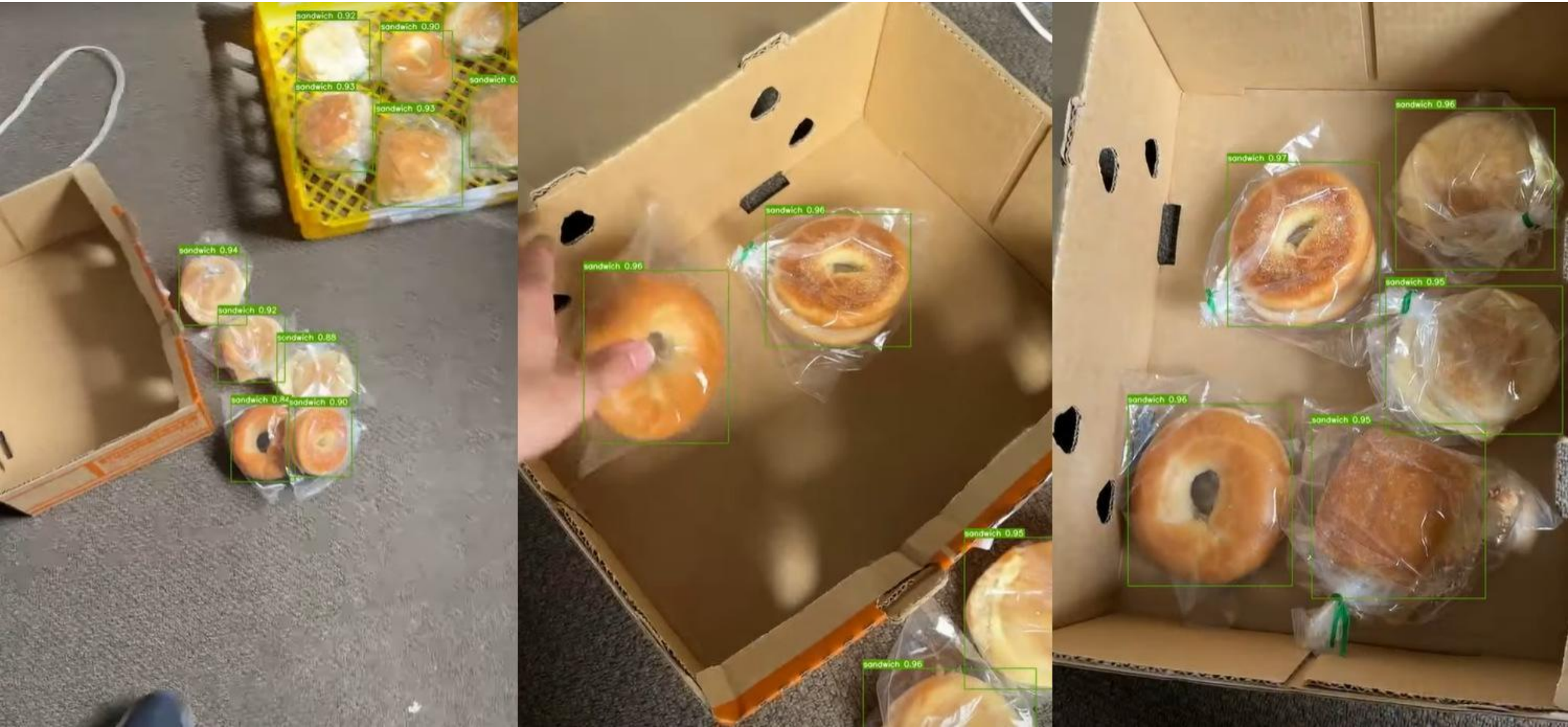
Chien-Yao Wang[1], Alexey Bochkovskiy, and Hong-Yuan Mark Liao[1]
[1]Institute of Information Science, Academia Sinica, Taiwan
kinyiu@iis.sinica.edu.tw, alexeyab84@gmail.com, and liao@iis.sinica.edu.tw

Abstract

YOLOv7 surpasses all known object detectors in both speed and accuracy in the range from 5 FPS to 160 FPS and has the highest accuracy 56.8% AP among all known real-time object detectors with 30 FPS or higher on GPU V100. YOLOv7-E6 object detector (56 FPS V100, 55.9% AP) outperforms both transformer-based detector SWIN-L Cascade-Mask R-CNN (9.2 FPS A100, 53.9% AP) by 509% in speed and 2% in accuracy, and convolutional-based detector ConvNeXt-XL Cascade-Mask R-CNN (8.6 FPS A100, 55.2% AP) by 551% in speed and 0.7% AP in accuracy, as well as YOLOv7 outperforms: YOLOR, YOLOX, Scaled-YOLOv4, YOLOv5, DETR, Deformable DETR, DINO-5scale-R50, ViT-Adapter-B and many other



Object detector: Fast, Reliable, Small training set



Resilience: Segmentation & OCR

Algorithms: Transformers breakthrough

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

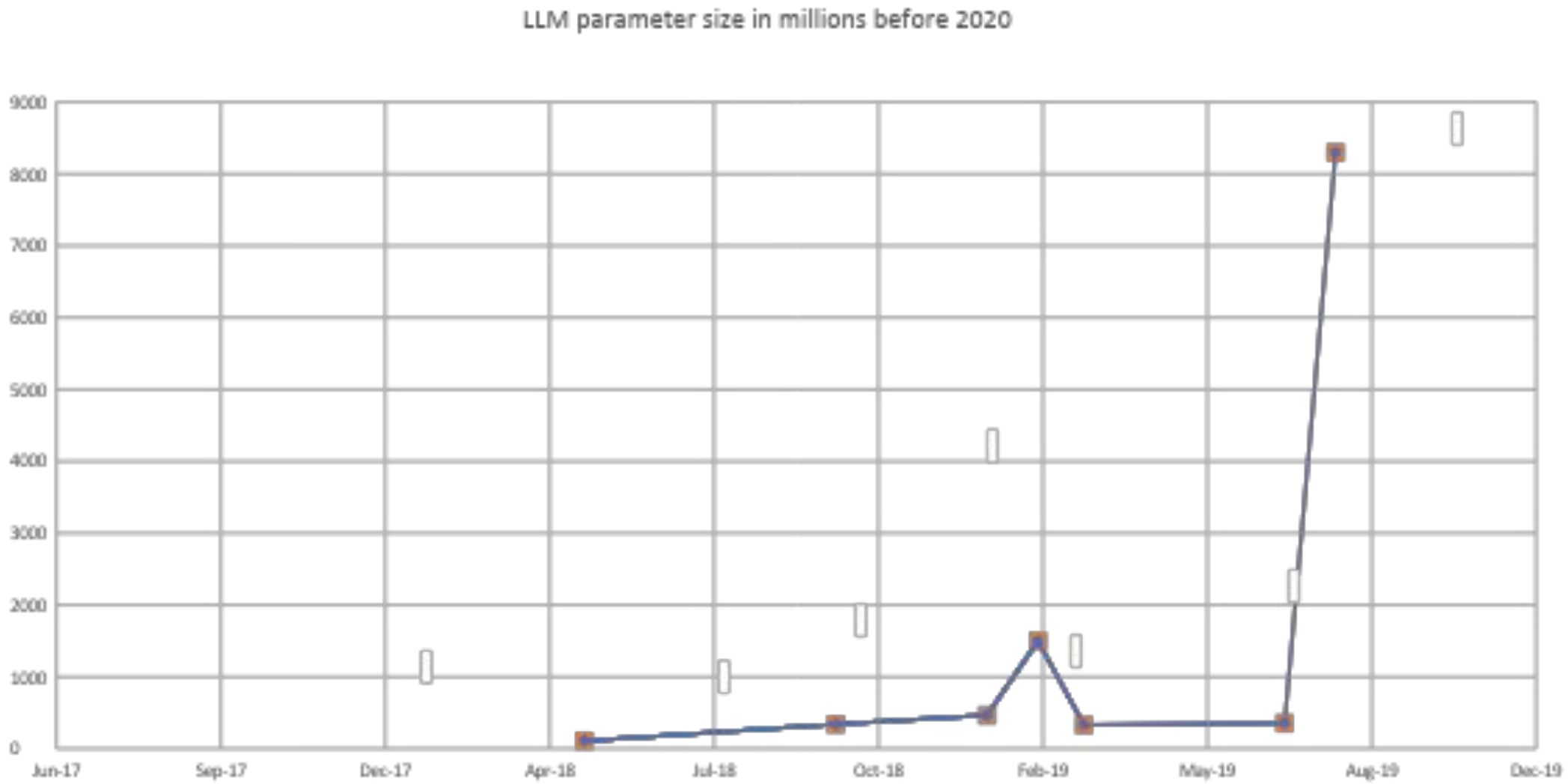
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

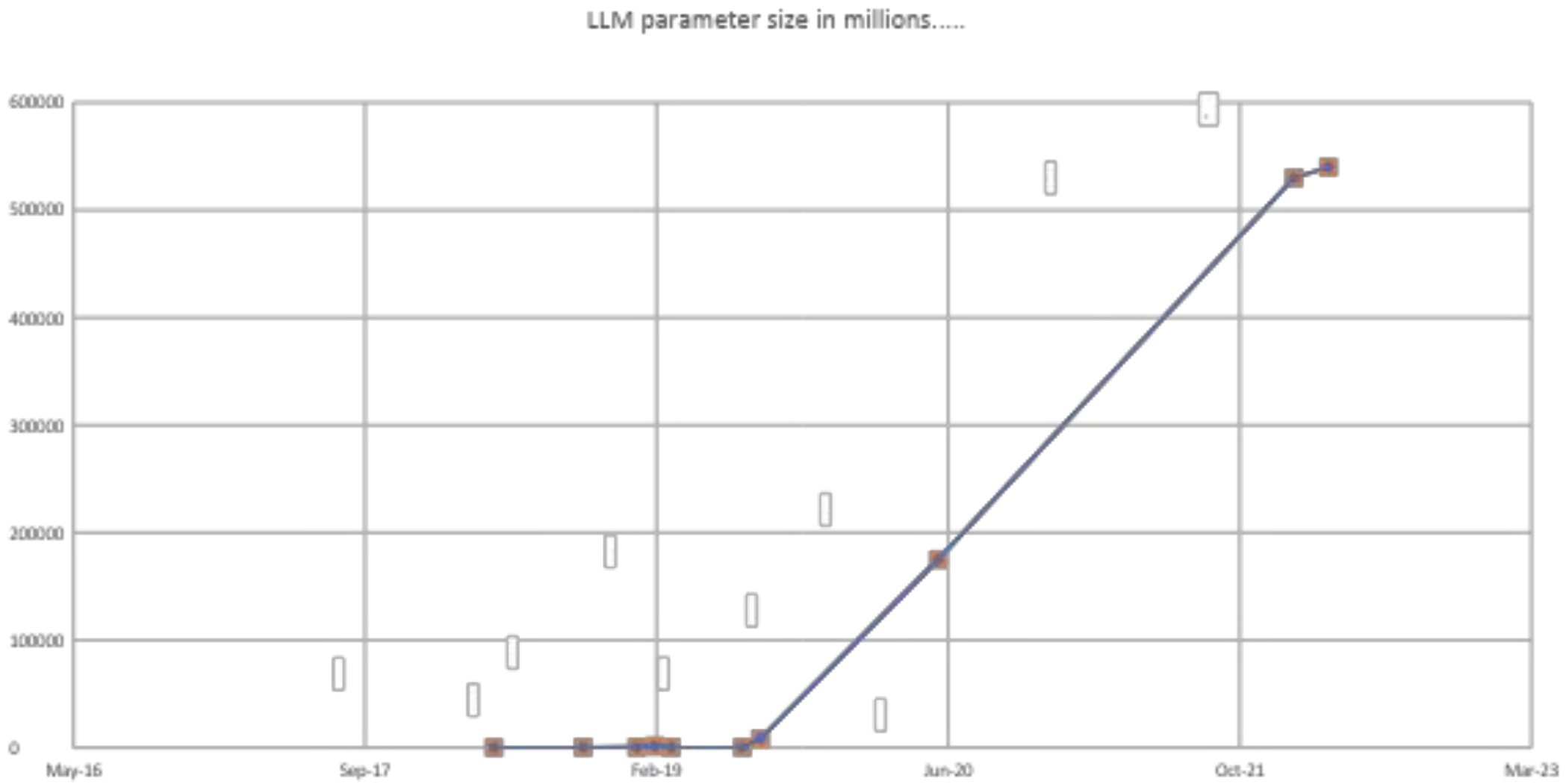
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

We propose a new simple network architecture, **the Transformer**, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely.

Algorithms: Throwing scale at models

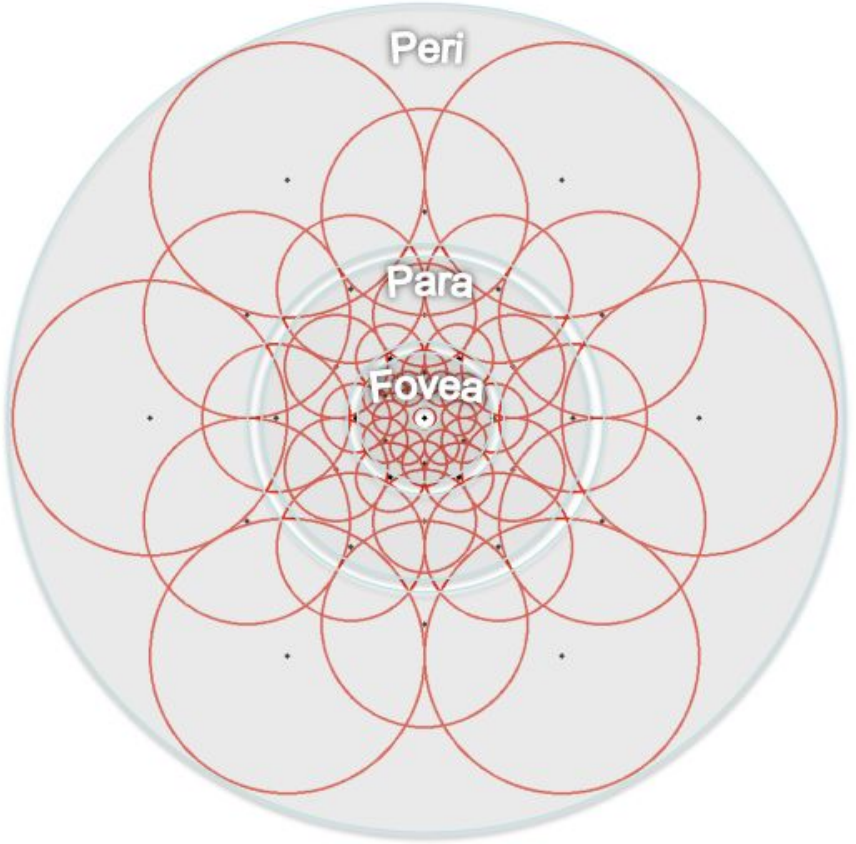


Algorithms: Throwing scale at models



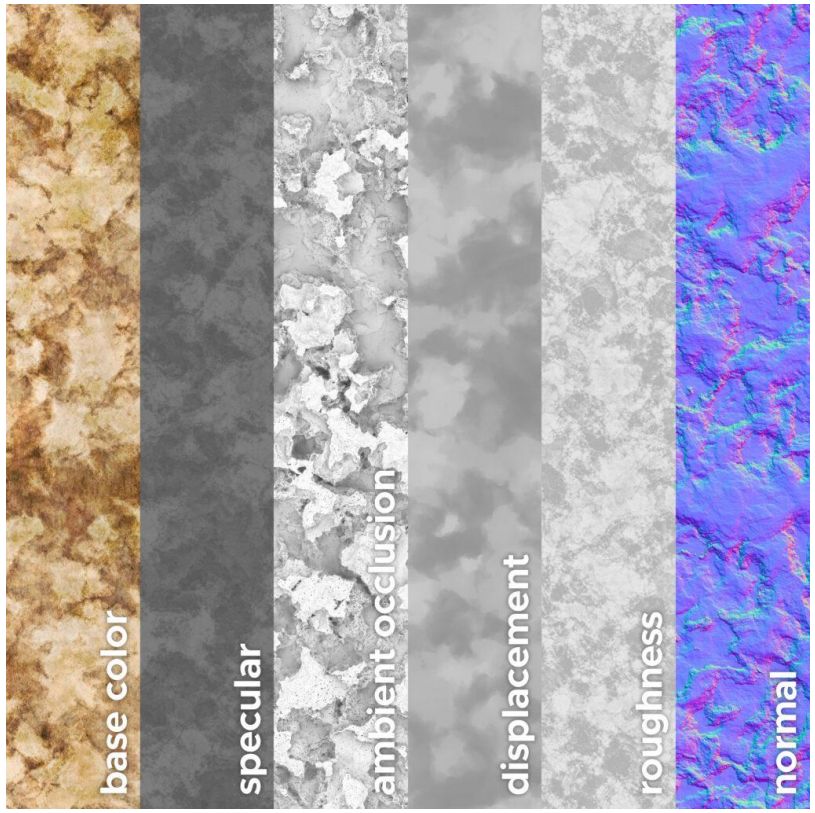
Transformers on images?

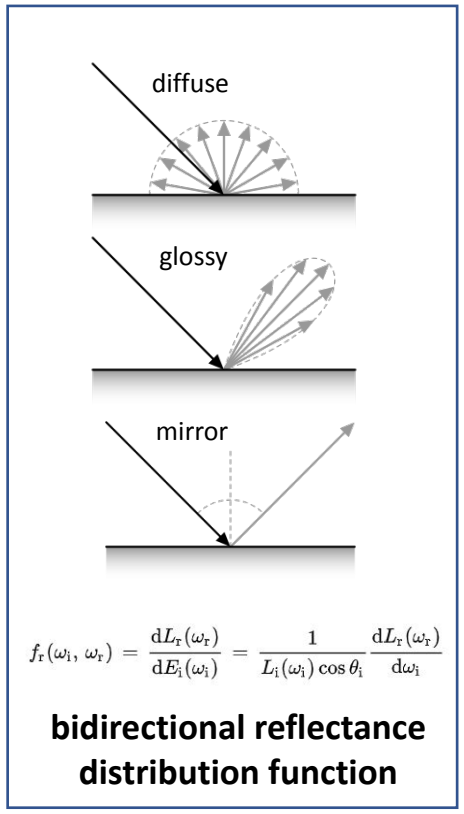
- Transformers for linear bidirectional content makes a huge amount of sense (audio and text)
- ViT Vision Transformers – linear patches are working.. But is this the best model?
- What do our visual systems do?
  e.g. FREAK: Fast Retina Keypoint



AI Solvers: Complex function “lookup”

- high-dimensional nonlinear functions (even ones we don’t have formulations for)
- recast as a stochastic problem
- approximate the gradient of the solution using deep neural networks





rendering

photometry

AI Solvers: Complex function “lookup”