



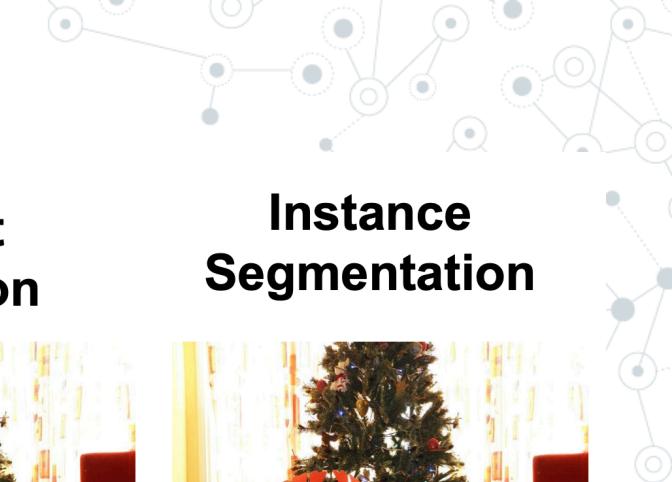
Lab 2: Computer Vision Anomaly Detection in Industrial Applications

11220IEEM513600

Deep Learning for Industrial Applications

2024/03/21 Taco





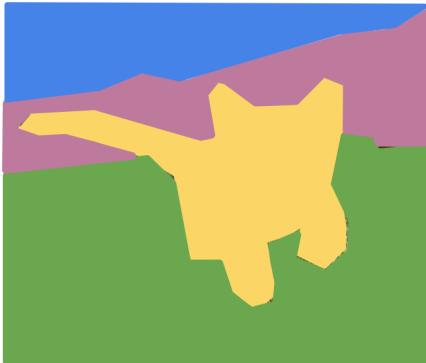
Classification



CAT

No spatial extent

Semantic Segmentation



GRASS, CAT,
TREE, SKY

No objects, just pixels

Object Detection



DOG, DOG, CAT

Multiple Object



DOG, DOG, CAT

Instance Segmentation

[This image](#) is CC0 public domain

CVPR 2023 by the Numbers

Selecting a category below changes the paper list on the right.

SELECT ↓ Top 10 overall by number of authors

	AUTHORS	PAPERS
1	3D from multi-view and sensors	1,090 246
2	Image and video synthesis and generation	889 185
3	Humans: Face, body, pose, gesture, movement	813 166
4	Transfer, meta, low-shot, continual, or long-tail learning	688 153
5	Recognition: Categorization, detection, retrieval	673 139
6	Vision, language, and reasoning	631 118
7	Low-level vision	553 126
8	Segmentation, grouping and shape analysis	524 113
9	Deep learning architectures and techniques	485 92
10	Multi-modal learning	450 89
11	3D from single images	431 91
12	Medical and biological vision, cell microscopy	420 53
13	Video: Action and event understanding	373 83
14	Autonomous driving	359 69
15	Self-supervised or unsupervised representation learning	349 71
16	Datasets and evaluation	344 54
17	Scene analysis and understanding	276 54
18	Adversarial attack and defense	274 61
19	Efficient and scalable vision	252 48
20	Computational imaging	226 53
21	Video: Low-level analysis, motion, and tracking	215 46
22	Vision applications and systems	171 35
23	Vision + graphics	155 32
24	Robotics	141 23
25	Transparency, fairness, accountability, privacy, ethics in vision	129 30
26	Explainable computer vision	107 24
27	Embodied vision: Active agents, simulation	80 14
28	Document analysis and understanding	72 12
29	Machine learning (other than deep learning)	65 14

Select
 All
 Award Candidate
 Highlight
 Paper

PICK INSTITUTIONS (All)

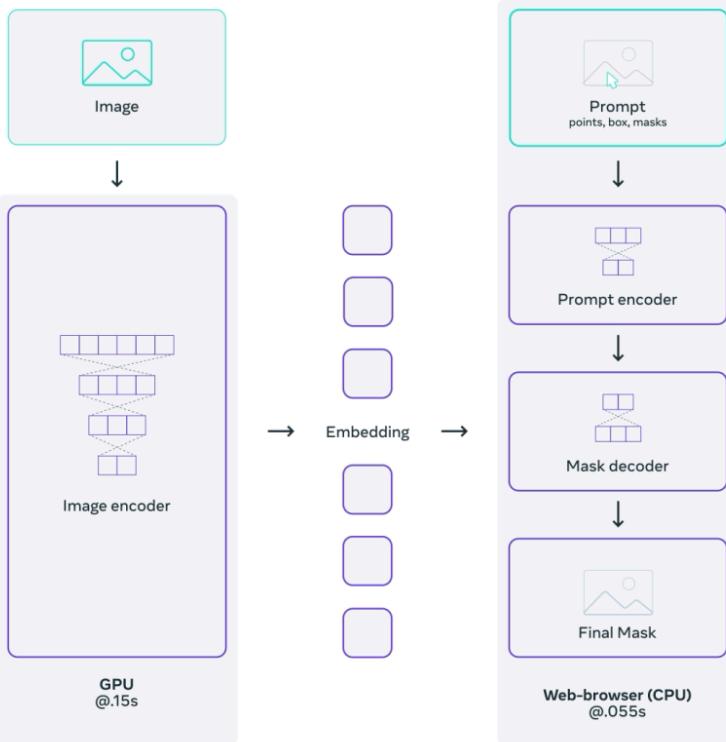
AUTHORS **# PAPERS**

3D from multi-view and sensors

- 33 NeuMap: Neural Coordinate Mapping by Auto-Transdecoder for Camera Localization
- 76 Object Pose Estimation with Statistical Guarantees: Conformal Keypoint Detection and Geometric Uncertainty Propagation
- 120 NeuralUDF: Learning Unsigned Distance Fields for Multi-view Reconstruction of Surfaces with Arbitrary Topologies
- 143 NEF: Neural Edge Fields for 3D Parametric Curve Reconstruction from Multi-view Images
- 330 Looking Through the Glass: Neural Surface Reconstruction Against High Specular Reflections
- 357 Multi-View Azimuth Stereo via Tangent Space Consistency

3

Segment Anything



Efficient & flexible model design

SAM is designed to be efficient enough to power its data engine. We decoupled the model into 1) a one-time image encoder and 2) a lightweight mask decoder that can run in a web-browser in just a few milliseconds per prompt.

<https://segment-anything.com>

YOLO-World

<https://www.yoloworld.cc/>

The screenshot shows a web browser window displaying the YOLO-World website. The page features a large green circular logo composed of dots, followed by the text "YOLO-World". Below the logo, the title "Real-Time Open-Vocabulary Object Detection" is displayed in a large, bold, dark font. Underneath the title, the authors' names are listed: Tianheng Cheng^{2,3*}, Lin Song^{1,*}, Yixiao Ge^{1,2}, Wenyu Liu³, Xinggang Wang³, Ying Shan^{1,2}. A note below indicates that ¹Tencent AI Lab, ²ARC Lab, Tencent PCG, ³Huazhong University of Science and Technology. Contribution roles are defined: *Equal Contribution, ☝ Corresponding Author, ⭐ Project Lead. The CVPR 2024 acceptance is mentioned. At the bottom, there are links to arXiv, Code, HuggingFace, and YOLO-World-EfficientSAM. A "What's New" section at the bottom left announces acceptance into CVPR 2024.

Real-Time Open-Vocabulary Object Detection

Tianheng Cheng^{2,3*}, Lin Song^{1,*}, Yixiao Ge^{1,2}, Wenyu Liu³, Xinggang Wang³, Ying Shan^{1,2}

¹Tencent AI Lab, ²ARC Lab, Tencent PCG, ³Huazhong University of Science and Technology

*Equal Contribution ☝ Corresponding Author ⭐ Project Lead

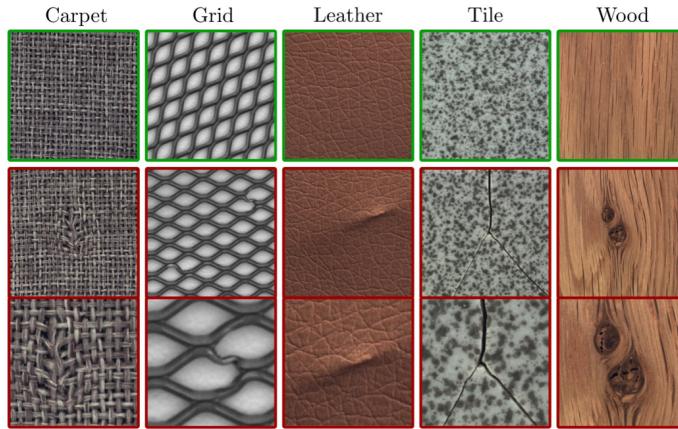
CVPR 2024

arXiv Code HuggingFace YOLO-World-EfficientSAM

🔥 What's New

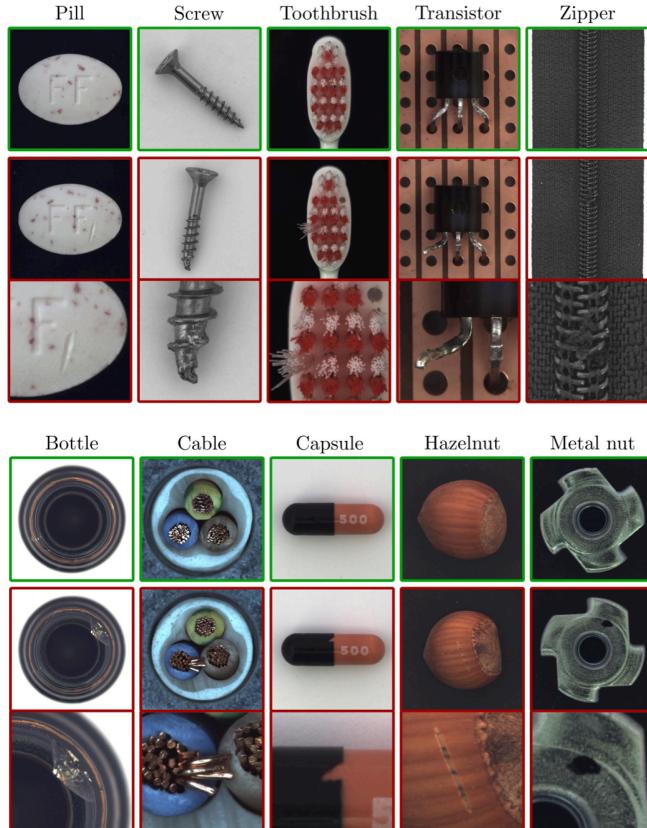
[2024-3-18] We are excited to announce that YOLO-World has been accepted by [CVPR 2024](#), hope to see you in Seattle! Now, YOLO-World supports prompt tuning, image prompts, high-resolution images (1280x1280), and ONNX export.

MVTec AD Dataset

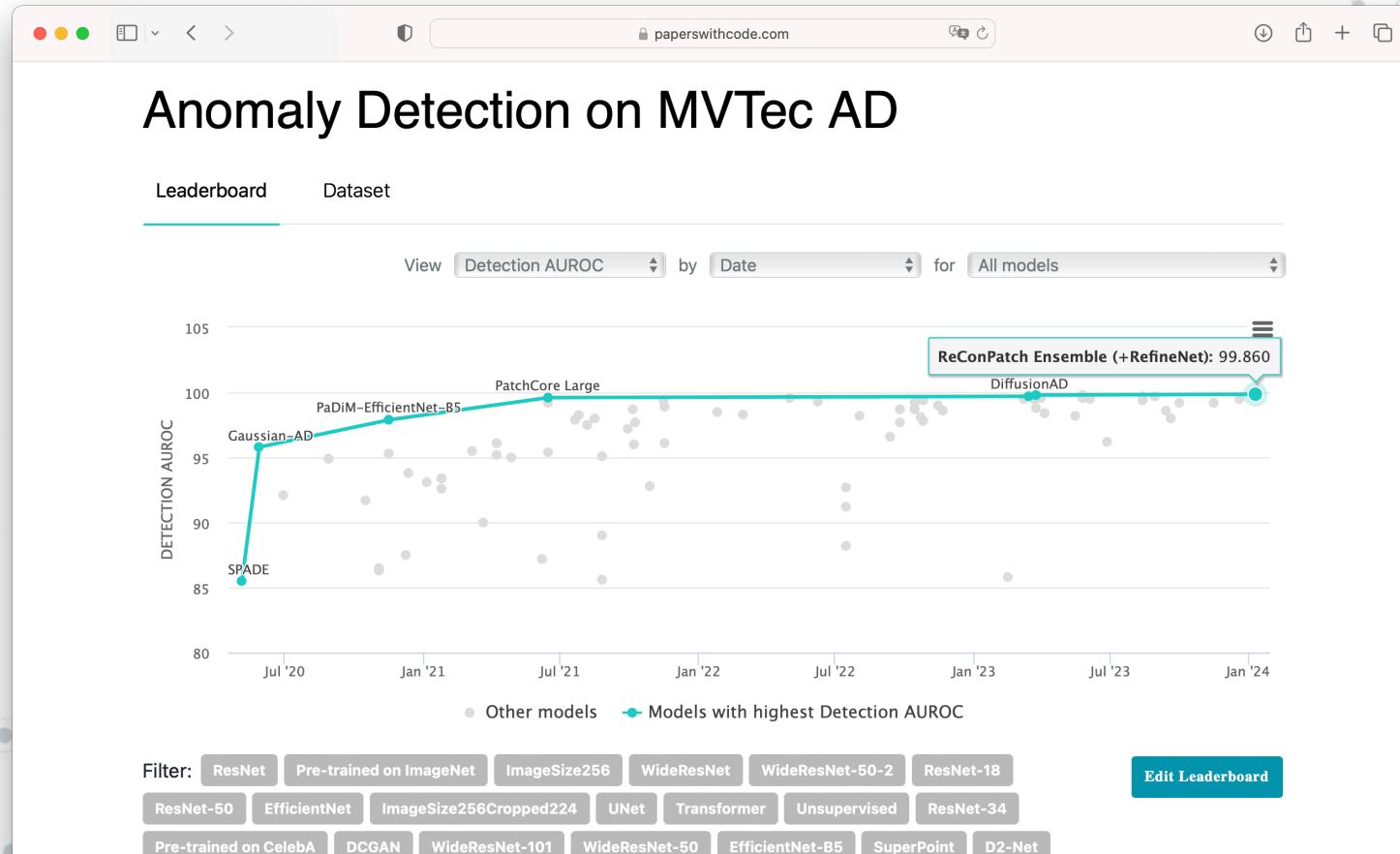


MVTec Anomaly Detection (MVTec AD) dataset containing 5354 high-resolution color images of different object and texture categories. It contains normal, i.e., defect-free, images intended for training and images with anomalies intended for testing.

https://openaccess.thecvf.com/content_CVPR_2019/papers/Bergmann_MVTec_AD -- A Comprehensive Real-World_Dataset_for_Unsupervised_Anomaly_CVPR_2019_paper.pdf



<https://paperswithcode.com/sota/anomaly-detection-on-mvtac-ad>



Homework 3

- ▶ **Deadline:** 2024.04.11 23:59
- ▶ **Github:** Create a "hw3" folder in your repository, "NTHU_2024_DLIA_HW", containing "hw3.ipynb" and "hw3.pdf". Ensure that you run your code and all outputs are saved within the .ipynb files.
- ▶ **EEclass:** You are required to submit only the GitHub link of your Homework 3. Do not upload files directly to EEclass.
- ▶  **Important:** Make sure your commit is timestamped before the deadline. Late submissions might not be graded or could incur a penalty. Only the GitHub link is required on NTHU EEclass.

CODING TIME!!

People: *fearing* AI takeover

AI:

