ELEVATER: A Benchmark and Toolkit for Evaluating Language-Augmented Visual Models

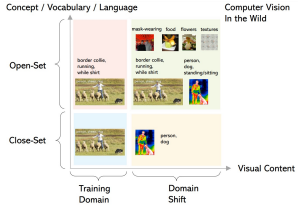
Chunyuan Li*, Haotian Liu*, Liunian Harold Li, Pengchuan Zhang, Jyoti Aneja,
Jianwei Yang, Ping Jin, Houdong Hu, Zicheng Liu, Yong Jae Lee, Jianfeng Gao
(Equal Contribution)

Microsoft WISCONSIN UCLA

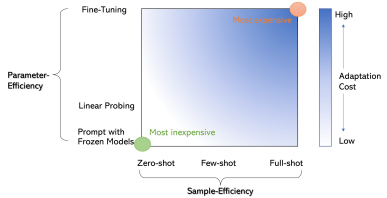
What is Computer Vision in the Wild (CVinW)?

Developing a transferable system that can adapt to *a large range of visual tasks* in the wild *effortlessly*.

1 The task transfer scenarios are broad

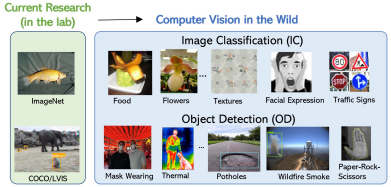


2 The task transfer cost is low.



Why ELEVATER?

A starting point of CVinW,
Evaluation of task-level visual transfer



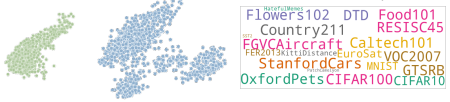
Fairness: Customized evaluation sets may favor individual model

Transparency: Detailed model adaptation process is inaccessible

Benchmark

ImageNet → Image Classification in the Wild (ICinW): **20** datasets

1151 classes: train 638K / eval 193K



LVIS → Object Detection in the Wild (ODinW): **35** datasets

314 classes: train 132K / eval 20K

External Knowledge

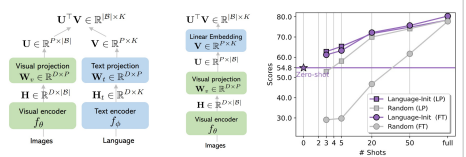
Concept name: risotto

- Def_wiki: An Italian savoury dish made with rice and other ingredients
- Def_wn: rice cooked with broth and sprinkled with grated cheese
- Path_wn: [risotto, dish, nutriment, food, substance, matter, physical_entity, entity]
- GPT3: A rice dish made with arborio rice and typically served with meat or fish

Toolkits

Automatic Hyper-parameter Tuning: learning rate and weight decay

Adaptation Pipeline: Language-augmented fine-tuning and linear probing

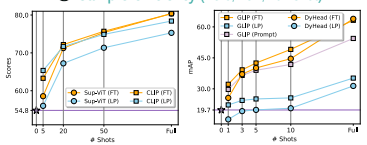


CLIP zero-shot

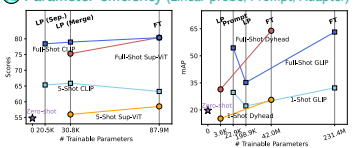
Language-Init Adaptation

Playgrounds

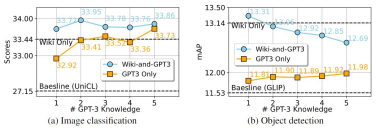
1 Sample-efficiency (Zero, Few, Full-shot)



2 Parameter-efficiency (Linear probe, Prompt, Adapter)



3 External Knowledge



(a) Image classification (b) Object detection

CVinW

Computer Vision in the Wild

https://computer-vision-in-the-wild.github.io

Challenge & Leaderboards

ICinW — Image Classification in the Wild

ODinW — Object Detection in the Wild

Benchmark Website
https://computer-vision-in-the-wild.github.io/ELEVATER/

ECCV Workshop
https://computer-vision-in-the-wild.github.io/eccv-2022/

Reading List for Beginners
https://github.com/Computer-Vision-in-the-Wild/CVinW_Readings

Call for Collaboration

- Benchmarking the transfer ability of SoTA vision models
- Contributors with valid submissions are encouraged for co-authorship