Open-Vocabulary Attribute Detection

TUE-PM-278


María A. Bravo


Sudhanshu Mittal


Simon Ging

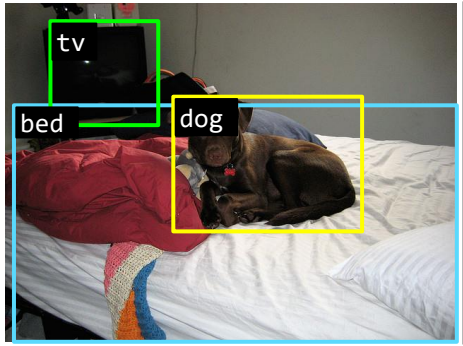

Thomas Brox

universität freiburg

[o] Vision
COMPUTER VISION University of Freiburg

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

OVAD: Open-vocabulary Attribute Detection



Object class: dog	Object class: bed	Object class: tv
Attributes	**Attributes**	**Attributes**
color quantity: single	color quantity: single	color quantity: single
color: brown	color: white	color: black
group: single	group: single	material: glass
maturity: adult	material: cloth	optical prop: reflective
position: lying	optical prop: opaque	pattern: plain
state: dry	size: large	state: off
texture: soft	texture: smooth	texture: smooth
tone: dark	tone: light	tone: dark

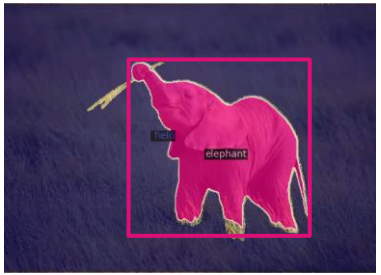
base
novel

Task Dataset Baseline Benchmark

Vision Language Models for Open Vocabulary Recognition

A gray baby elephant in a field

VLM



Classes: elephant, field

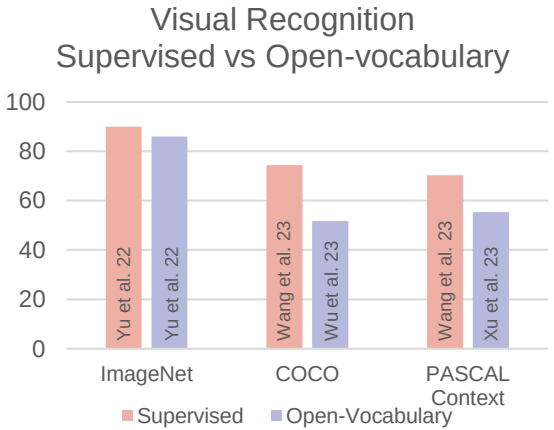
Beyond a closed-set of categories

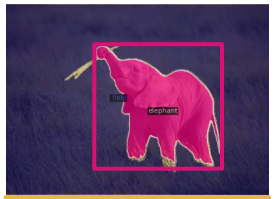
Use Natural Language to query classes

Use Image-Text pairs as supervision

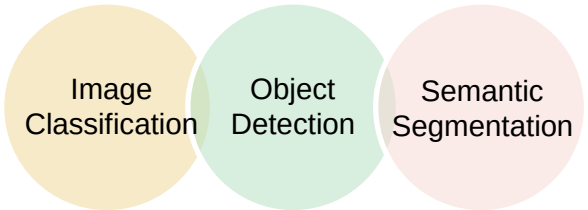
Vision Language Models for Open Vocabulary Recognition

Noun concepts



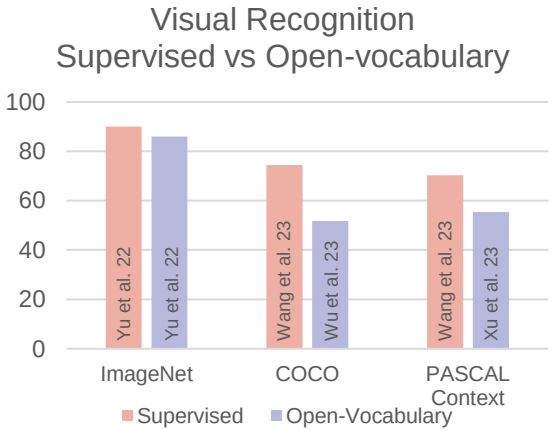


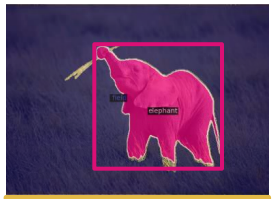
Classes: elephant, field



Vision Language Models for Open Vocabulary Recognition

Noun concepts





Classes: elephant, field

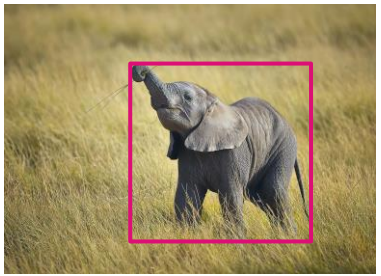
Image Classification

Object Detection

Semantic Segmentation

Attribute concepts

A gray baby elephant in a field



```
Object class: elephant
Attributes
color quantity: one
color: gray
group: single
maturity: baby
pattern: plain
position: vertical
size: small
state: dry
texture: rough
tone: light
```

Significance of attributes in an object's identity

A cup with **three** pairs of scissor, one is **blue**, and **two** are **metallic**.

A **white** and a **spotted** horse on a field of grass.

Red traffic signal in the middle of a **wide** street.

Images from COCO Captions. Chen et al. 2015

OVAD task: Open-vocabulary Attribute Detection

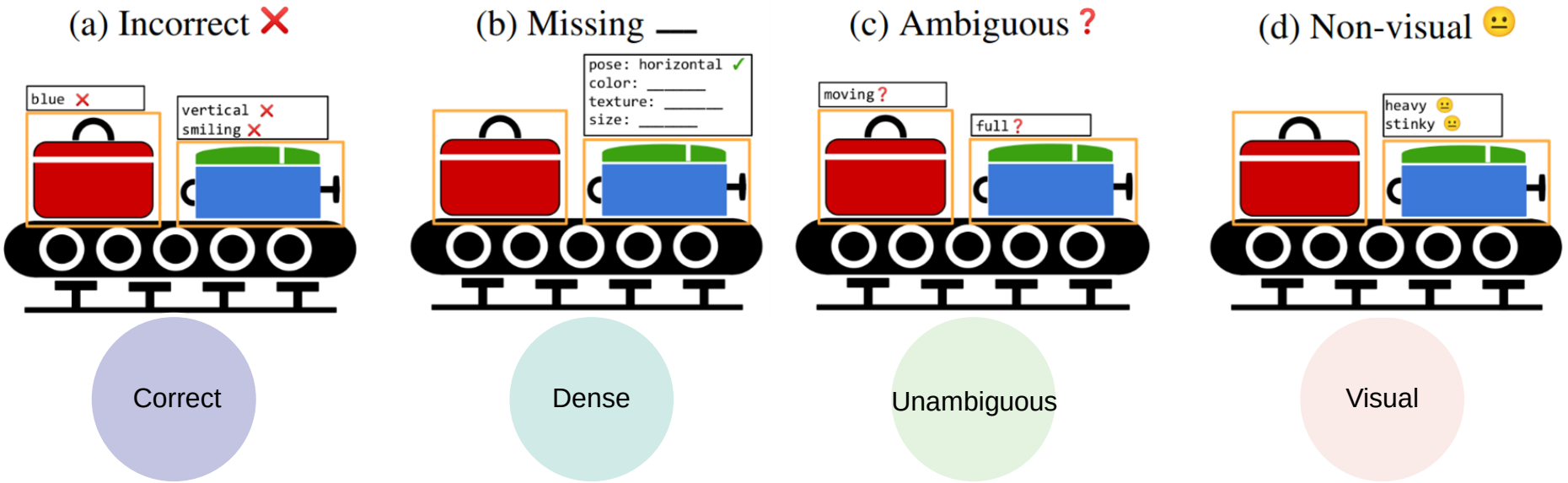
Object class: bear
Attributes
color quantity: two
color: black and brown
group: single
material: wood
maturity: adult
position: upright
size: big
state: dry
texture: smooth
tone: dark

Object class: car
Attributes
color quantity: one
color: red
group: single
material: wood
optical prop: opaque
patterns: lettered
state: piece / cut
texture: smooth
tone: dark

Object class: person
Attributes
face exp: surprise
group: single
hair color: black
hair length: short
hair tone: dark
hair type: straight
maturity: adult
position: upright
clothes color: white

*Solving the **OVAD** task:* → *(1) Detect all **object classes***
*(2) Detect their **attributes**.*

Attribute Detection Benchmarks
Type of errors:



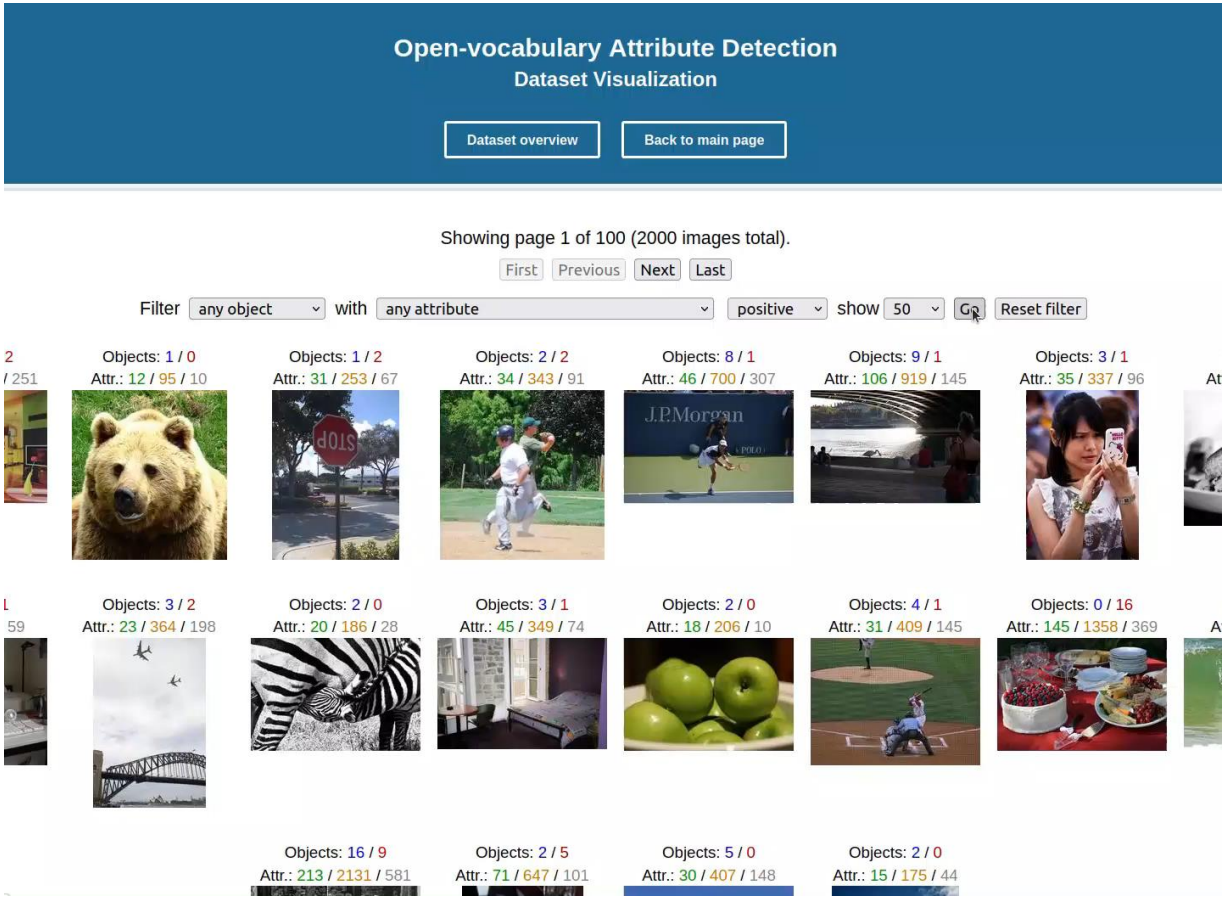
OVAD Benchmark

Dense

Correct

Unambiguous

Visual



OVAD Benchmark

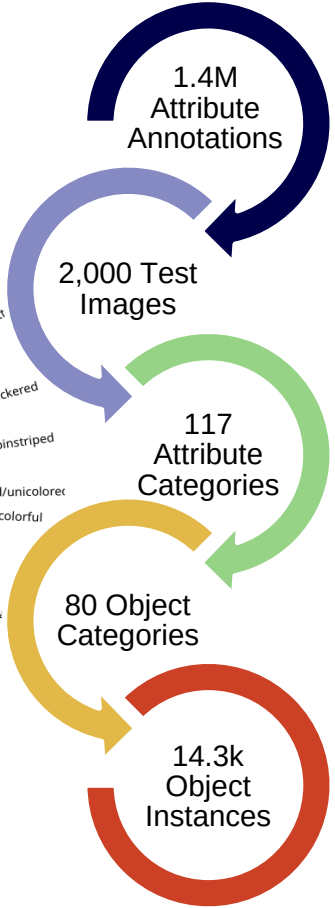
Human annotated

~ 96.8 attributes / instance

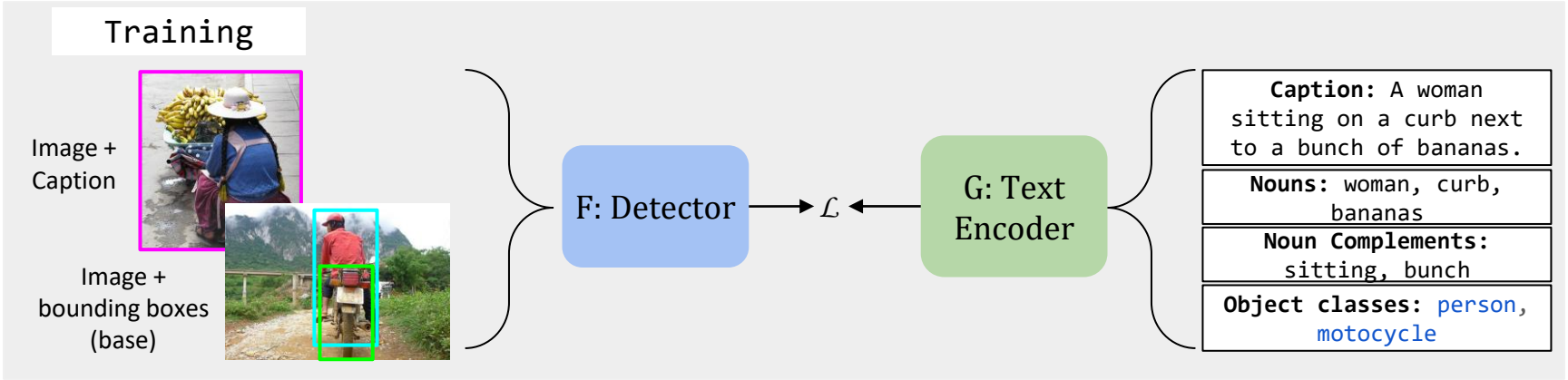
Evaluation modes:

1. Full benchmark:
   - Detect objects and their attributes

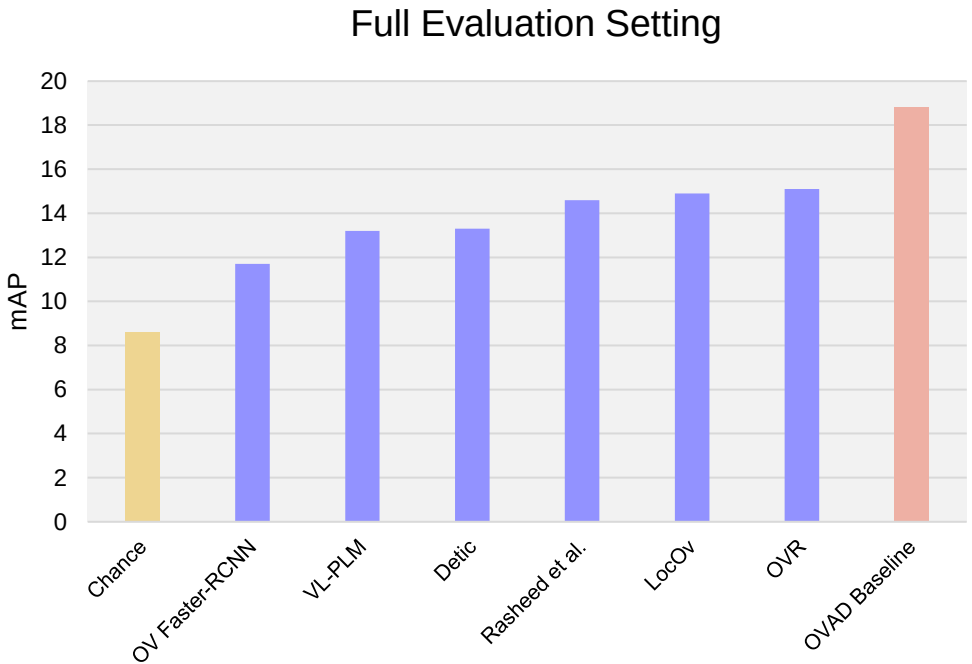
2. Box-oracle:
   - Detect attributes given box

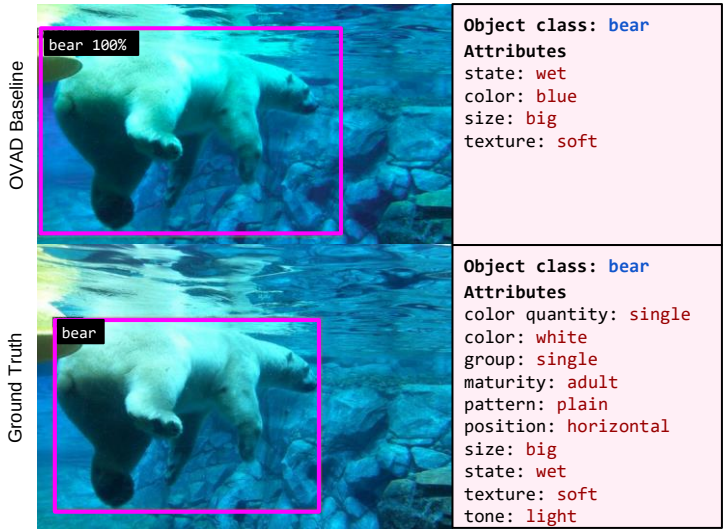


OVAD Baseline



Open-vocabulary Models on OVAD

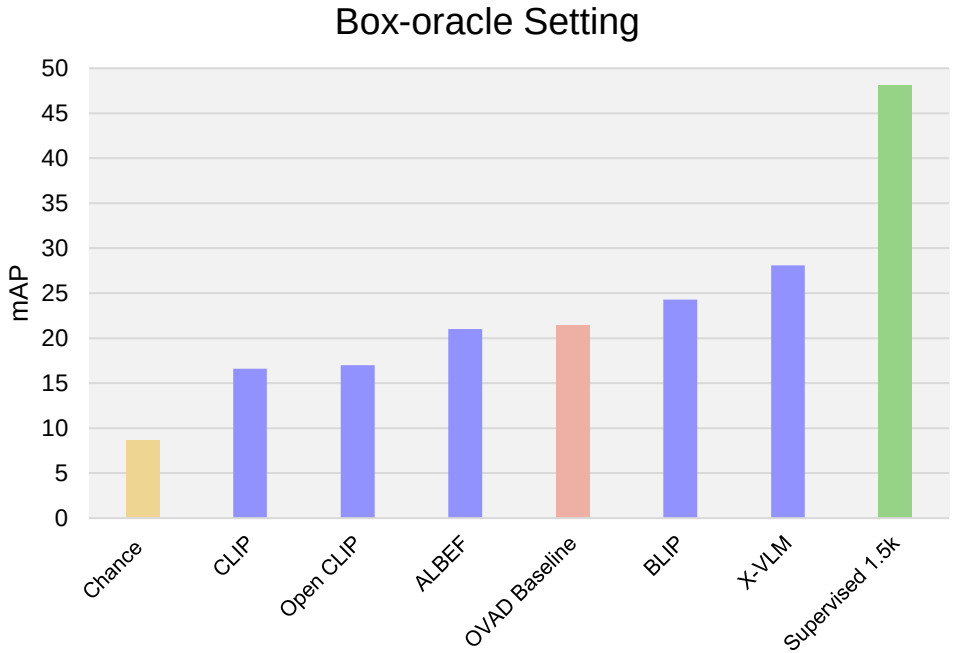




- Our OVAD Baseline outperforms latest OVD models on the OVAD task.

Performance of VLMs on OVAD

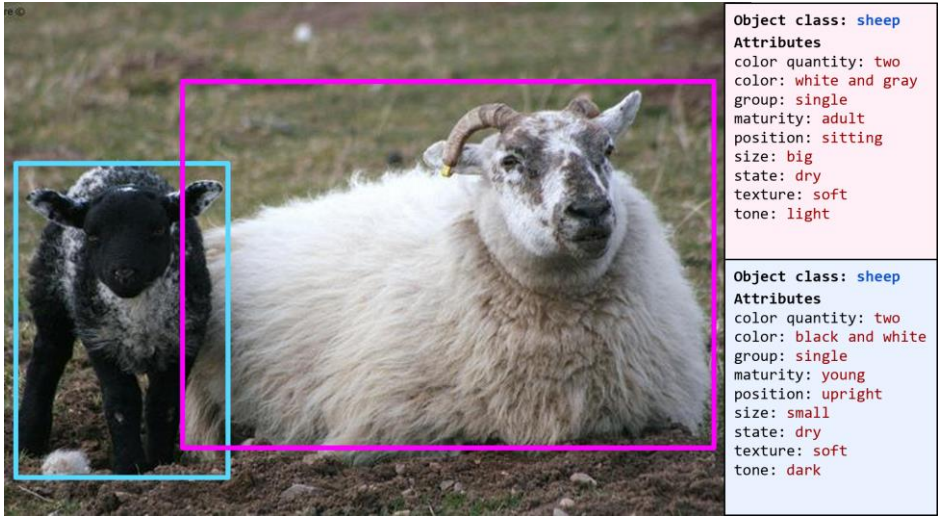
Method	mAP OVAD-Box oracle
Chance	8.6
CLIP VIT-B16 (400M)	16.6
Open CLIP VIT-B32 (2B)	17.0
ALBEF (14M+COCOft)	21.0
OVAD Baseline (110k+COCO)	21.4
BLIP (129M+COCOft)	24.3
X-VLM (16M+COCOft)	**28.1**
Supervised (OVAD-1.5k)	48.2



- X-VLM model performs the on the OVAD task in the box-oracle setting.
- Large pre-trained VLMs fail to capture fine-grained information.

Open-Vocabulary Attribute Detection

TUE-PM-278



[o] Vision
COMPUTER VISION University of Freiburg

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA