MosaicOS: A Simple and Effective Use of Object-Centric Images for Long-Tailed Object Detection

Cheng Zhang[1], Tai-Yu Pan[1*], Yandong Li[2], Hexiang Hu[3], Dong Xuan[1], Soravit Changpinyo[2], Boqing Gong[2], Wei-Lun Chao[1]*

[1]The Ohio State University, [2]Google Research, [3]University of Southern California

2021 ICCV OCTOBER 11-17 VIRTUAL

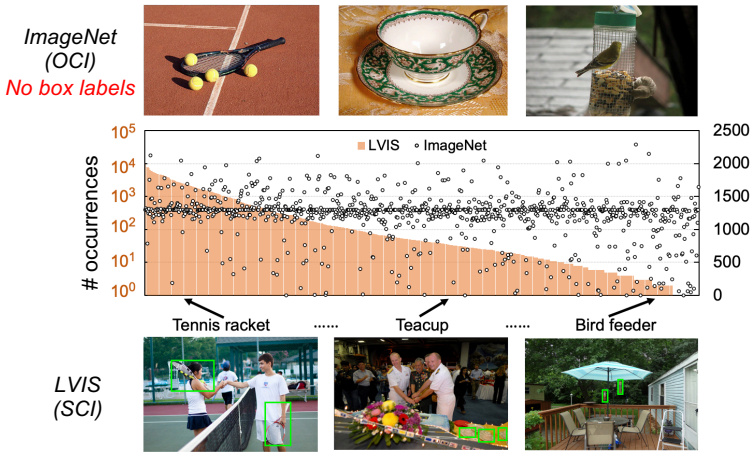
Highlights

- *Investigate the use of **object-centric images (OCI)** to facilitate long-tailed object detection on **scene-centric images (SCI)***
- *Propose **MosaicOS** framework to leverage OCI, consisting of **pseudo SCI generation**, **multi-stage training**, etc.*
- *Achieve significant accuracy gains on LVIS, e.g., boost object detection mAP of **rare** classes from 13% to **20%***

Introduction

- Long-tailed object detection
  - ➢ Many objects do not appear frequently enough in SCIs
  - ➢ Most methods address this issue in training objectives
- **Our key insights** to resolve "long-tailed" distributions
  - ➢ Rare objects in SCIs appear **more frequently** in OCIs
  - ➢ Good OCI sources exist: ImageNet & Google Images

Object-centric images *vs.* scene-centric images: different *object frequencies*, *focuses*, and *object sizes*

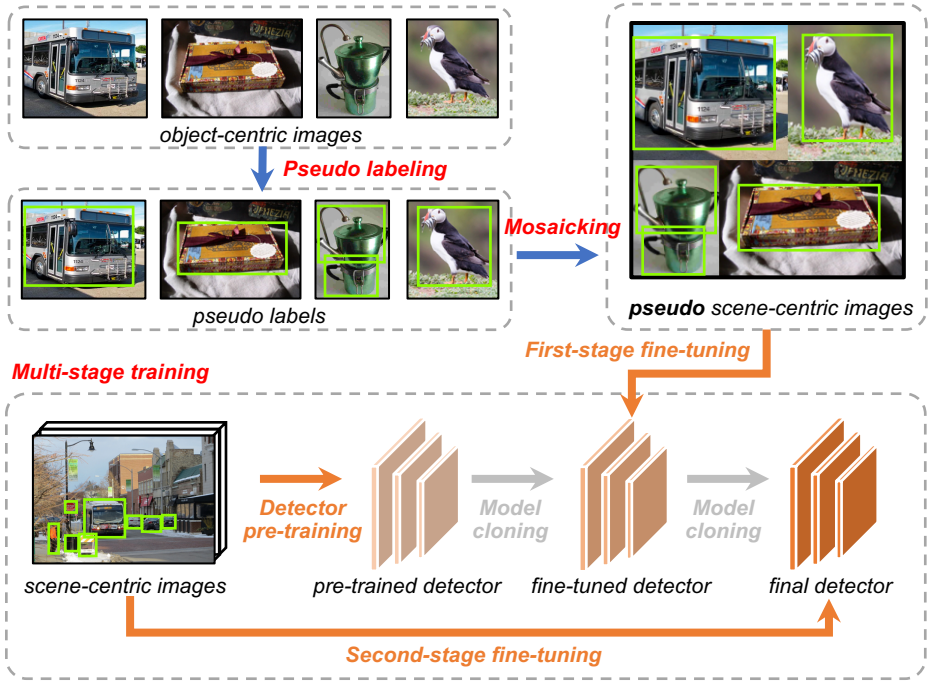


Can we leverage abundant OCIs?

- Missing box locations in OCI
  - ➢ Only image-level label is available in OCIs
- Domain gap between OCI and SCI
  - ➢ Visual discrepancy in object sizes and contexts
- Learning strategy
  - ➢ How to train with images from different domains

Proposed MosaicOS Framework

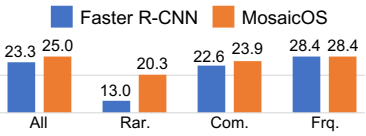
Main idea: "Mosaicking" Object-centric images to create pseudo Scene-centric images for long-tailed object detection



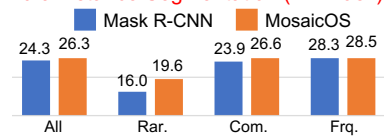
Experiments

- **Dataset**: LVIS [1], categories are divided into frequent (>100 images), common (11-100), and rare (<10) groups
- **Metric**: mAP on all, frequent, common, and rare classes
- **Object-centric images**: ImageNet-21K [2] & Google Images
- **Class matching**: using WordNet [3] synset names

v0.5 Object Detection (AP box)

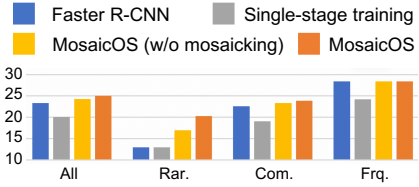


v0.5 Instance Segmentation (AP mask)

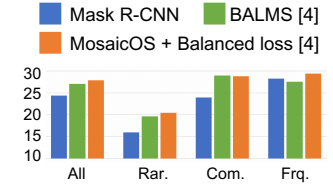


For both tasks, MosaicOS notably improves rare object categories without sacrificing frequent classes

v0.5 Ablations (AP box)



v0.5 SOTA Comparison (AP mask)

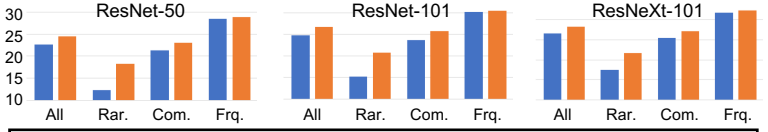


Mosaicking OCIs is effective

Multi-stage training matters

MosaicOS can be compatibly used with SOTA methods

v1 Instance Segmentation (AP mask)



MosaicOS is strong upon different model architectures

See more ablations on mosaicking, pseudo-label generation, multi-stage training in the paper:

[1] LVIS: A dataset for large vocabulary instance segmentation. In CVPR, 2019.
[2] ImageNet: A large-scale hierarchical image database. In CVPR, 2009.
[3] WordNet: A lexical database for English. Communications of the ACM, 1995.
[4] Balanced meta-softmax for long-tailed visual recognition. In NeurIPS, 2020.