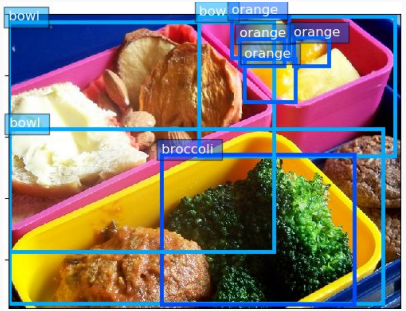
Localizing Grouped Instances for Efficient Detection in Low-Resource Scenarios

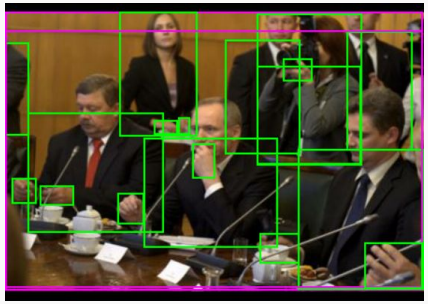
Amélie Royer, Christoph H. Lampert
IST Austria
WACV 2020

Motivation

Heterogeneous distributions and computational efficiency problems
for specific applications, e.g. aerial imagery.
Trade-off between exhaustiveness / accuracy and computational efficiency

Natural Scenes Datasets





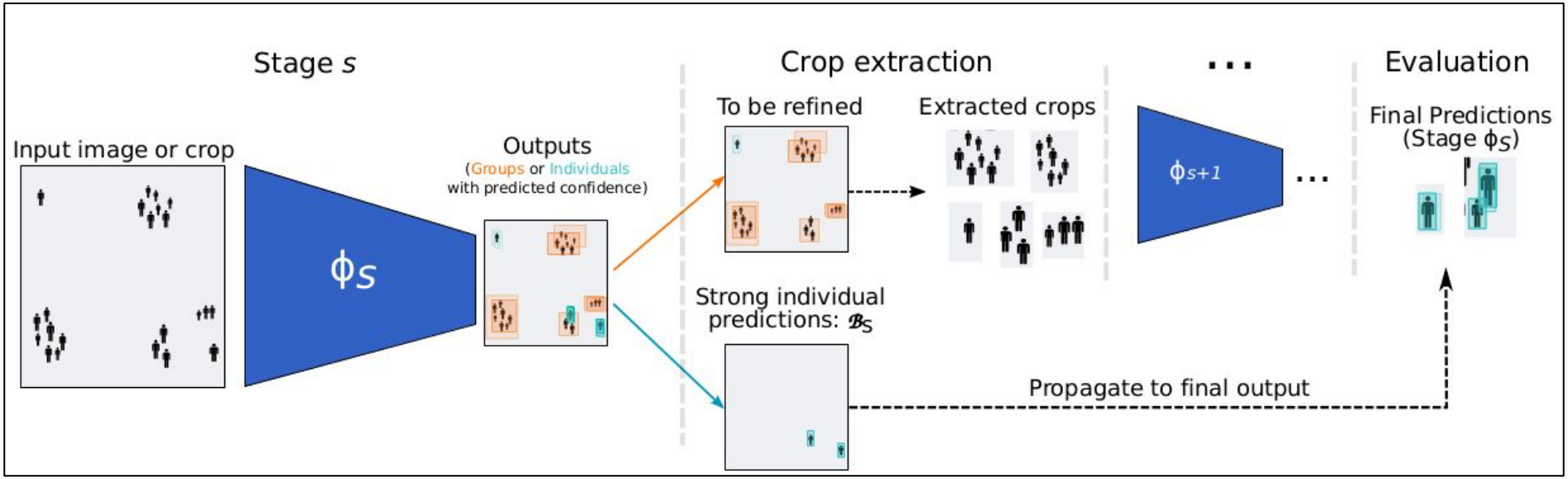
Densely distributed objects
salient objects
Varied semantic categories

Aerial Views Datasets

Small objects
Only a few relevant regions
Often specific application (e.g., pedestrians)

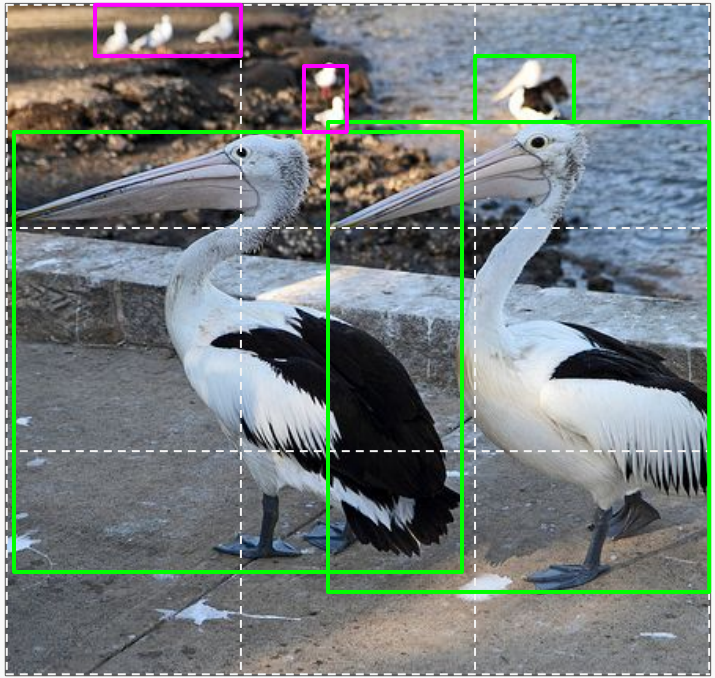
Model Overview

ODGI (**O**bject **D**etection with **G**rouped **I**nstances)
A Refinement cascade for object detection.



Defining groups of objects

$$\mathcal{L} = \sum_S \mathcal{L}^s_{\text{loc}} + \mathcal{L}^s_{\text{groups}} + \mathcal{L}^s_{\text{offsets}}$$



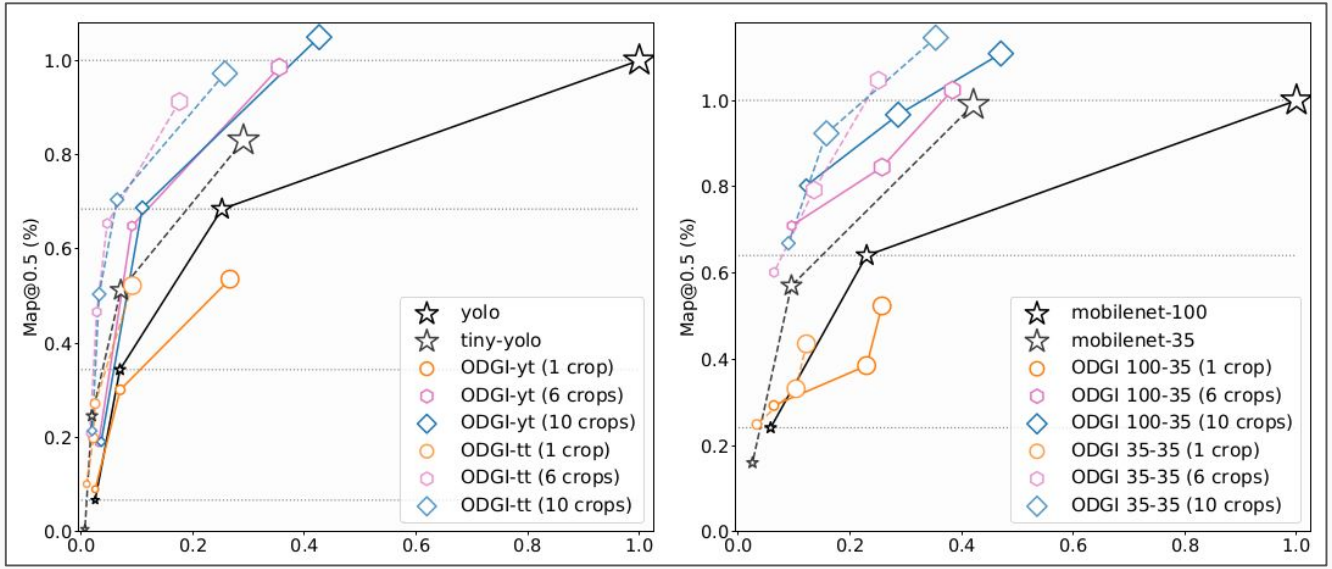
Intuitively,

- If one object in a cell: map to the unique predictor (standard detection)
- If multiple objects in a cell: group them and define the union of bounding boxes as the new ground-truth

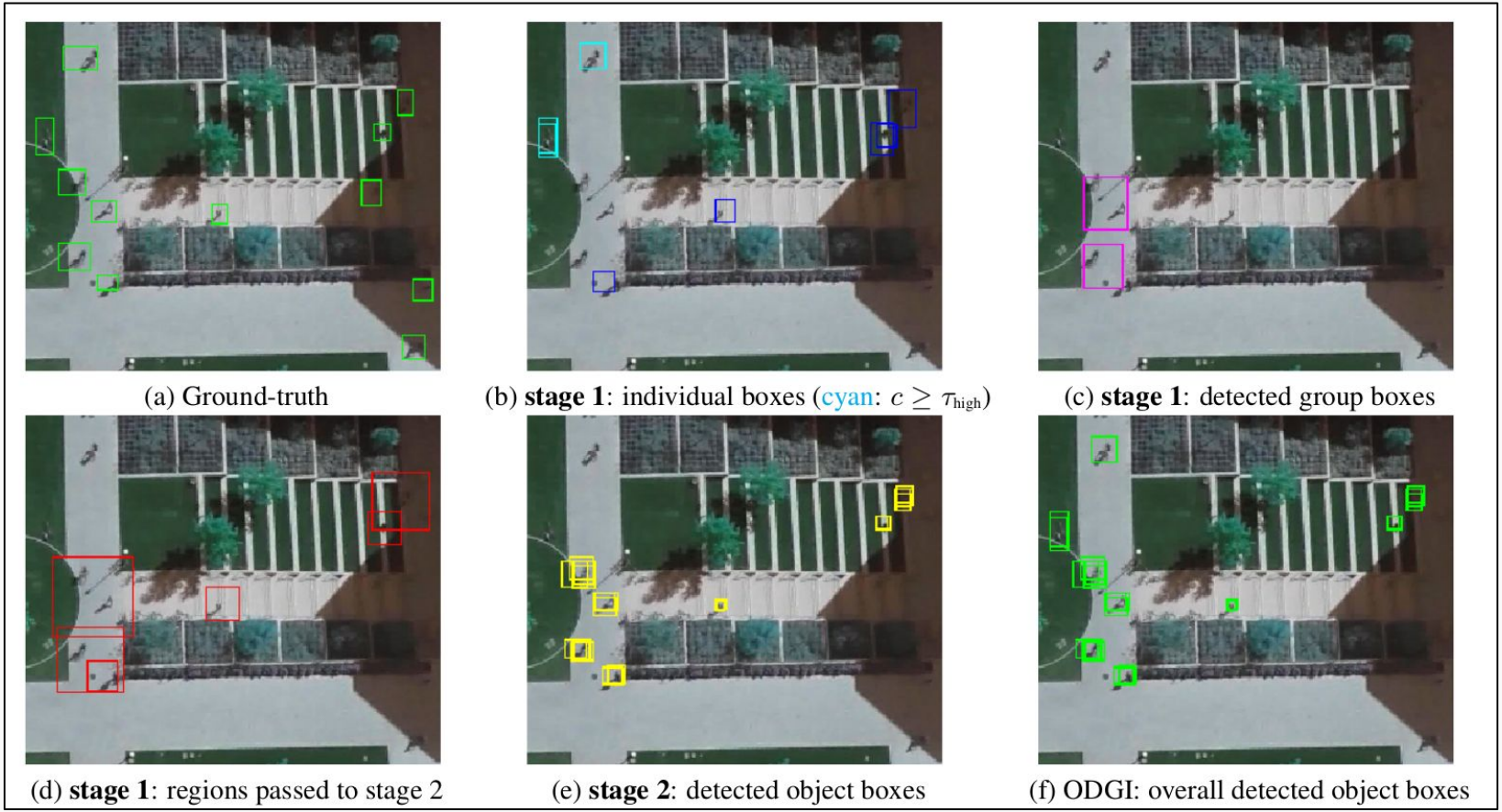
Experiments

Two aerial views datasets

- VEDAI[6] (1268 images, ~3 objects per image)
- SDD[7] (~13k images, ~12 objects per image)



Experiments



(a) Ground-truth

(b) **stage 1**: individual boxes (cyan: $c \geq \tau_{\text{high}}$)

(c) **stage 1**: detected group boxes

(d) **stage 1**: regions passed to stage 2

(e) **stage 2**: detected object boxes

(f) ODGI: overall detected object boxes