Interactive Multi-Class Tiny-Object Detection

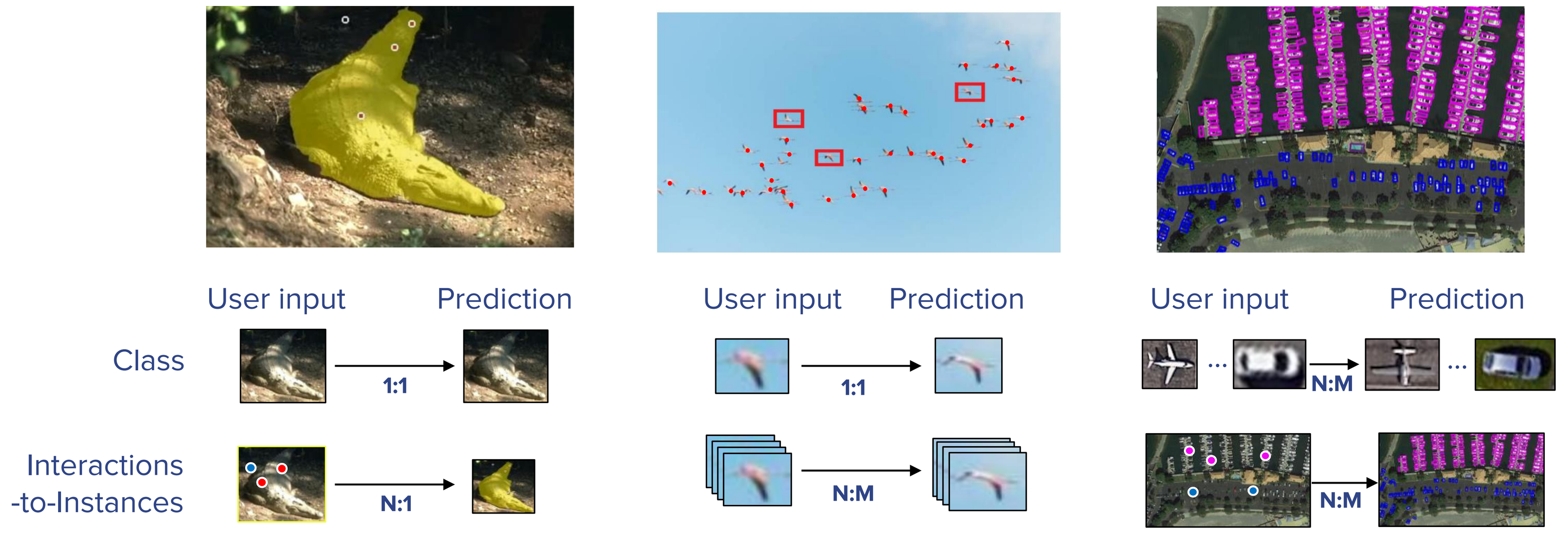
Chunggi Lee, Seonwook Park, Heon Song, Jeongun Ryu, Sanghoon Kim, Haejoon Kim, Sérgio Pereira, and Donggeun Yoo

PAPER GITHUB

Lunit

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation



Annotating tiny objects is an important Computer Vision task, but annotating these many objects is very expensive

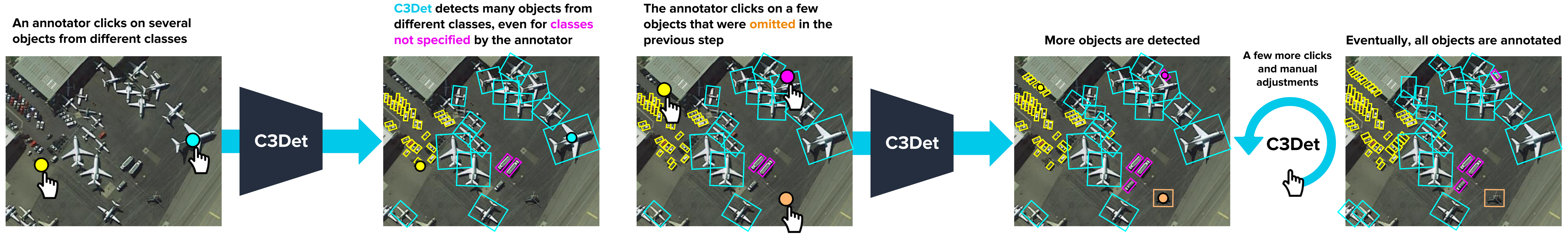
- Tedious + Laborious
- Time-consuming
- therefore, Expensive

C3Det

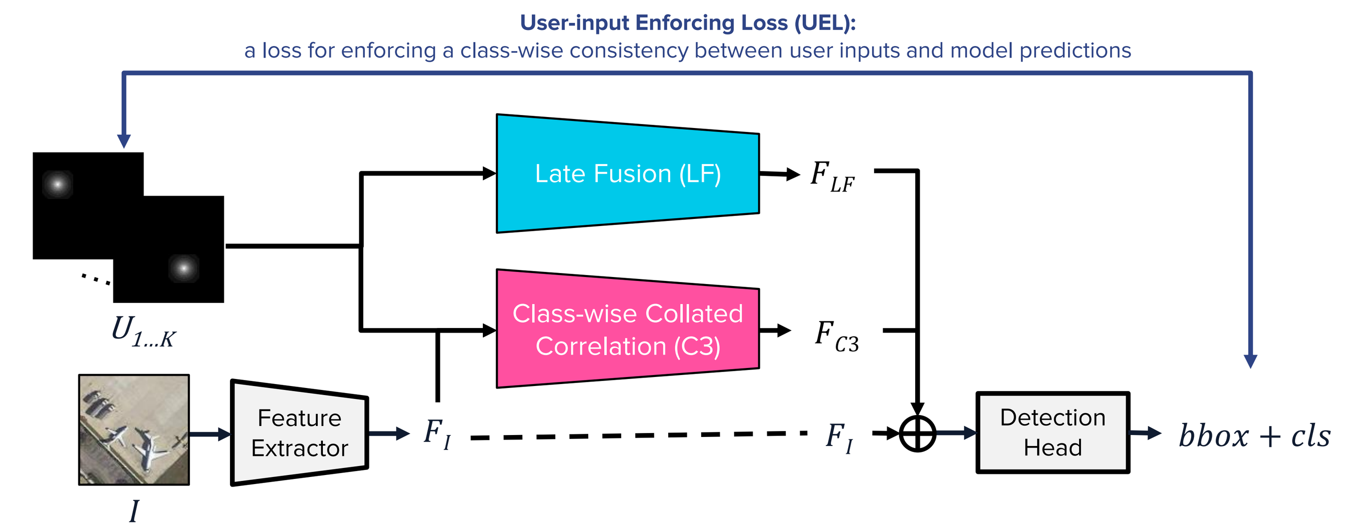
C3Det is an effective interactive annotation framework for tiny object detection that understands the effect of user inputs in a local and global manner via late fusion and feature correlation.

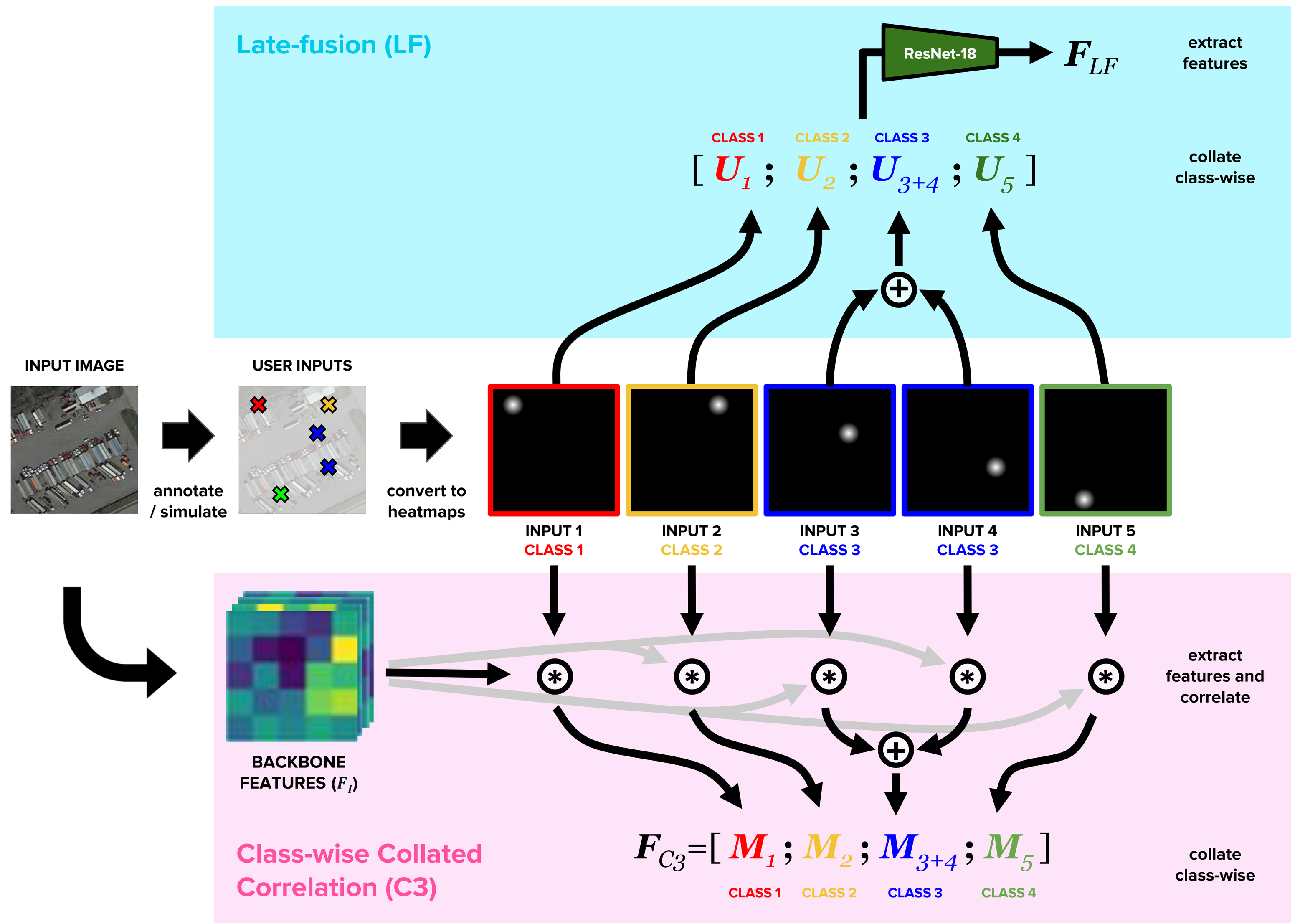
- Works at Interactive Rates
- Reduces workload (0.36x lower)
- Reduces annotation time (2.85x faster)

Overview



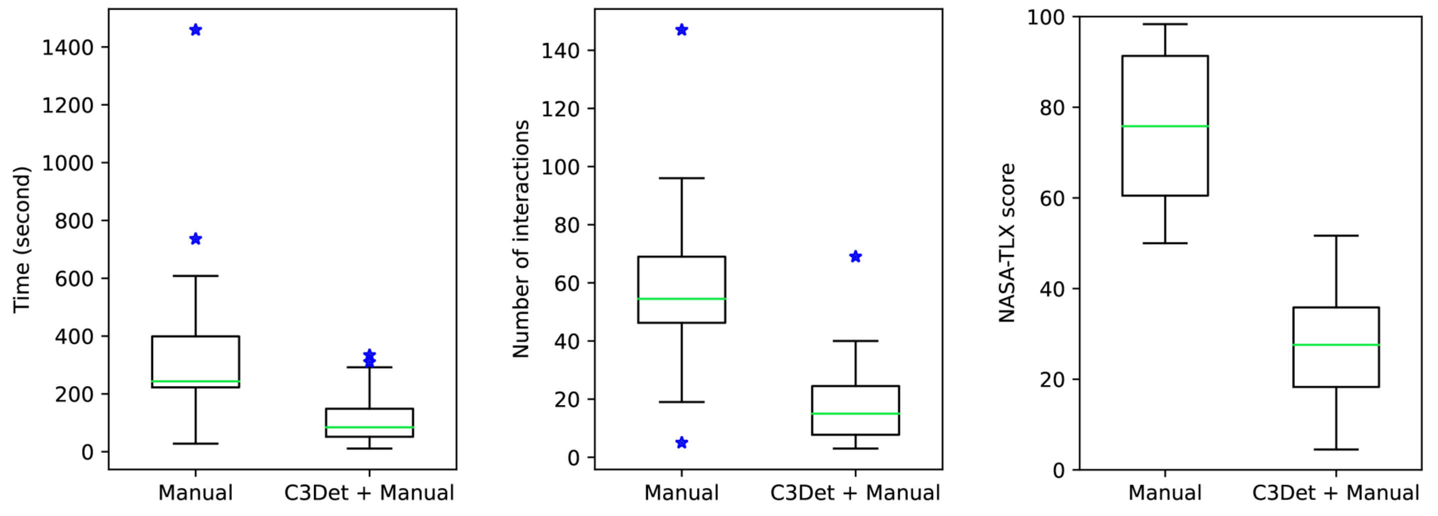
C3Det Architecture





User Study

Our study is a within-subject study, in which 10 participants perform their tasks with (a) fully-manual annotation or (b) interactive annotation using C3Det.

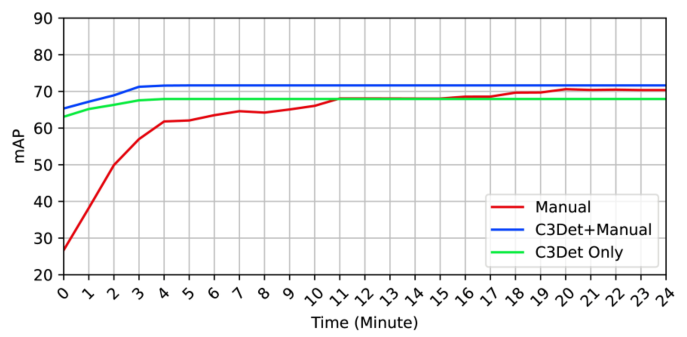


The **average annotating time spent** for C3Det + Manual (114.7s) is **2.85 times lower** than the Manual condition (327.73s).

The **number of interactions** required for C3Det + Manual (17.93) is **3.25 times fewer** than the Manual condition (58.33).

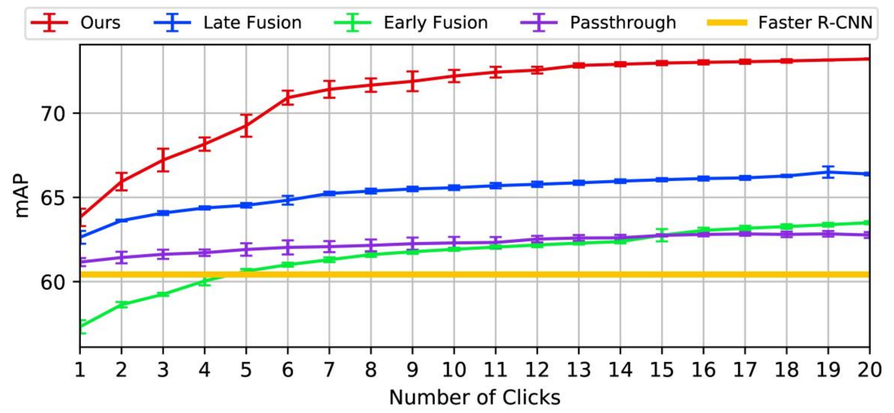
The **median NASA-TLX score** with the Manual approach is 75.83, and the score with C3Det + Manual is 27.58.

To achieve 67.9 mAP, the Manual condition takes 714.3s, while the "C3Det Only" and "C3Det + Manual" conditions take 294.2s and 144.2s respectively. Allowing manual edits after C3Det results in more complete annotations - over 5 times faster than fully manual annotation of Tiny-Dota.

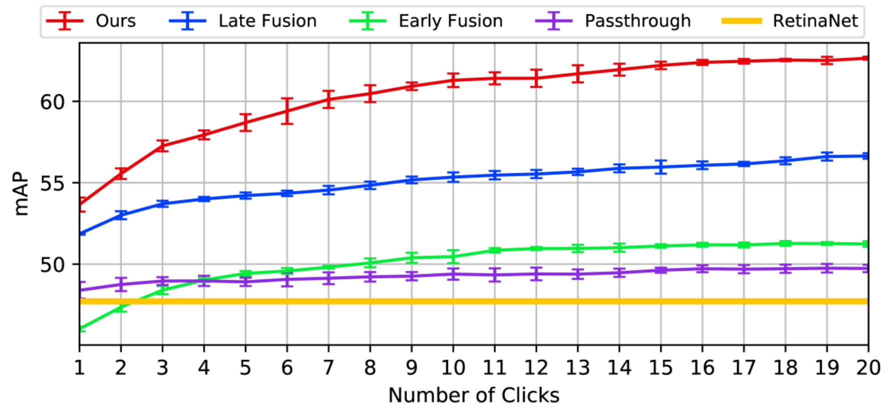


Results

Two-Stage Model (Faster R-CNN R50-FPN) on the Tiny-Dota Dataset

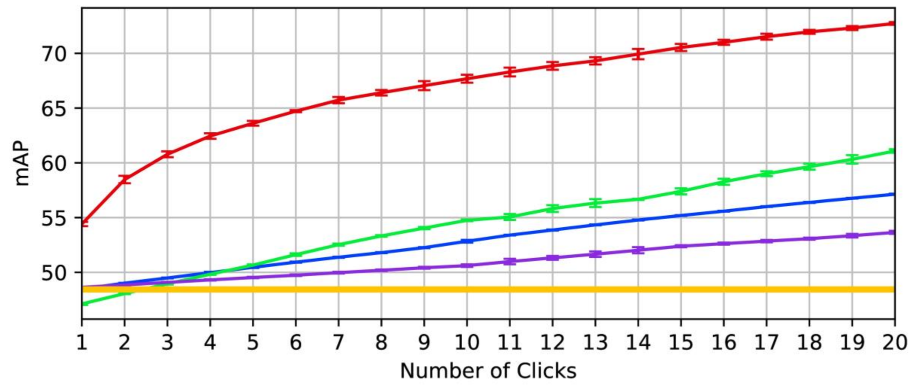


One-Stage Model (RetinaNet R50) on the Tiny-Dota Dataset



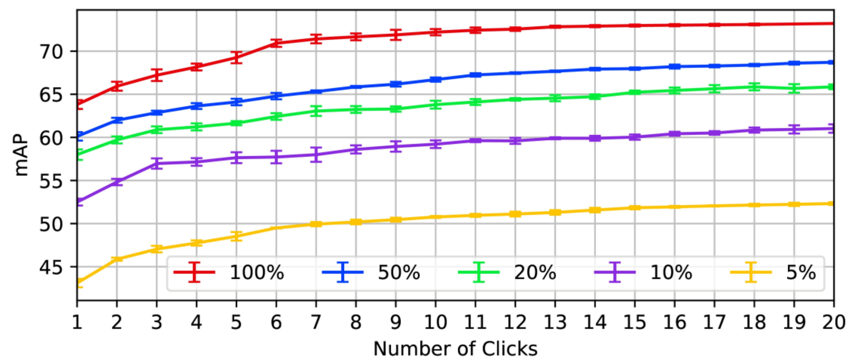
Our method outperforms all baselines in both one- and two-stage models,
1. quickly increasing in mAP with a few number of clicks
2. reaching higher mAP when the maximum number of clicks are provided

Two-Stage Model (Faster R-CNN R50-FPN) on the LCell Dataset

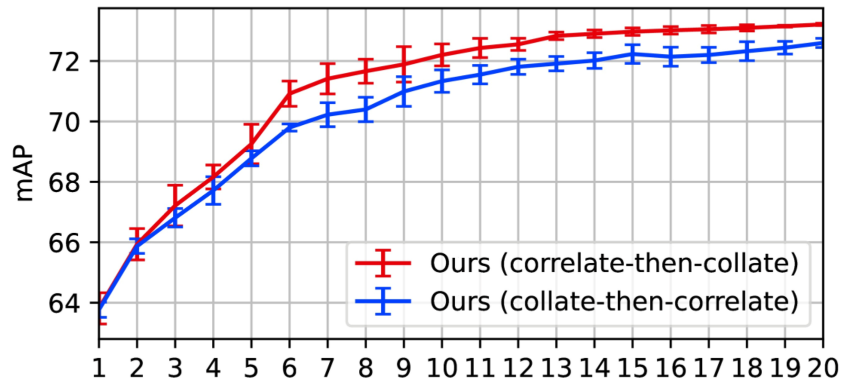


We find that similar trends can be seen on LCell dataset.

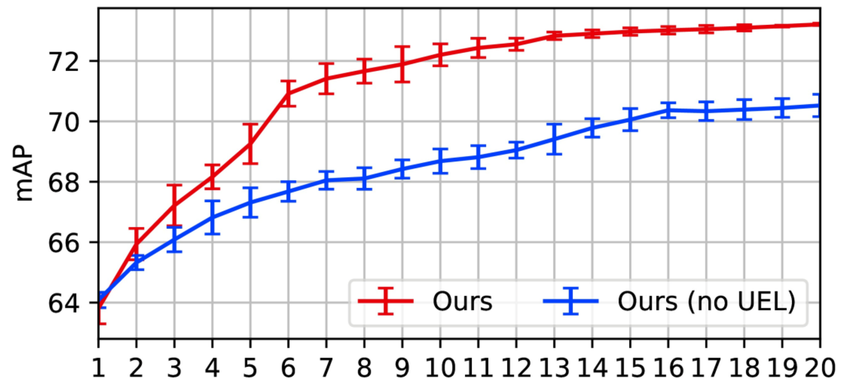
Decreasing the amount of training data (%: percentage of full Tiny-Dota training subset). Our approach predicts bounding boxes with increasing mAP with increasing clicks, even with as little training data as 5%. We show that C3Det is applicable in the real-world environment.



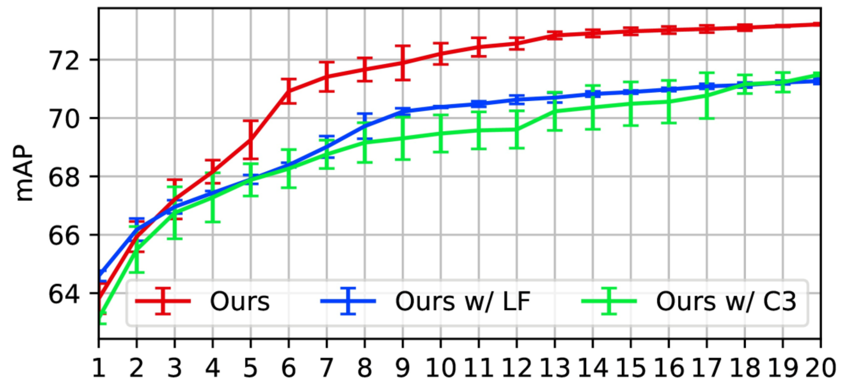
Correlation and Collation order in C3 module. First correlating then collating allows multiple user inputs from the same class to be captured better, showing consistent improvements especially at high number of clicks.



Effect of User-input Enforcing Loss (UEL). UEL ensures the consistency between user inputs and model predictions both in terms of instances as well as class.



When used together, C3 and LF modules help. Late-fusion (LF) and class-wise collated correlation (C3) when used together capture both local and global contexts of user inputs, yielding large performance improvements.



Take Home Message

- We introduce a training data synthesis and an **evaluation procedure** for the problem of interactive multi-class tiny-object detection.
- Our proposed C3Det architecture considers **local**-context (LF module) and **global**-context (C3 module) holistically.
- Our **real-world user-study** (10 annotators) shows that C3Det is **2.85x faster** and yields **0.36x lower task-load** compared to manual annotation.