Open-world Text-specified Object Counting

Niki Amini-Naieni, Kiana Amini-Naieni, Tengda Han, Andrew Zisserman
Visual Geometry Group, University of Oxford

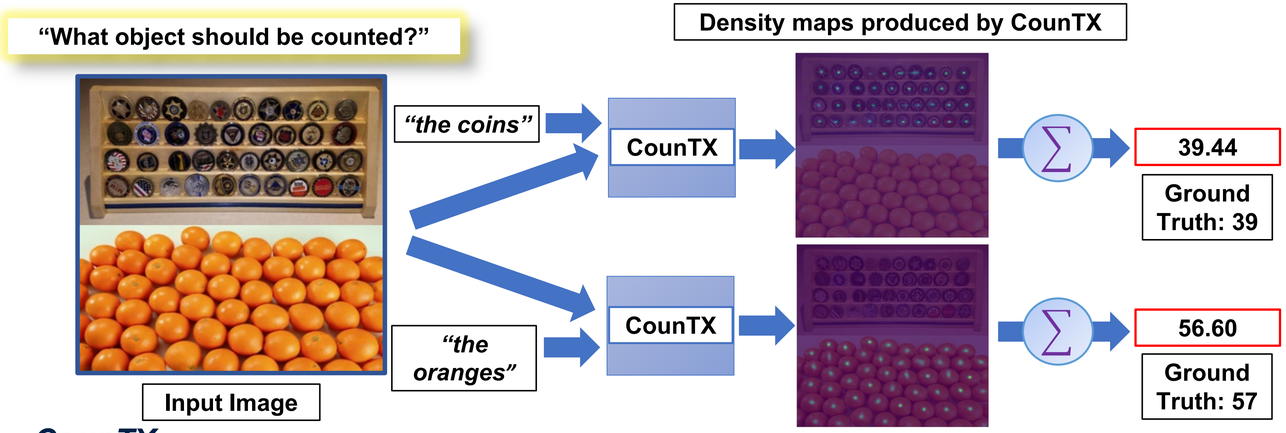
UNIVERSITY OF OXFORD

BMVC 2023

Objective and Contributions

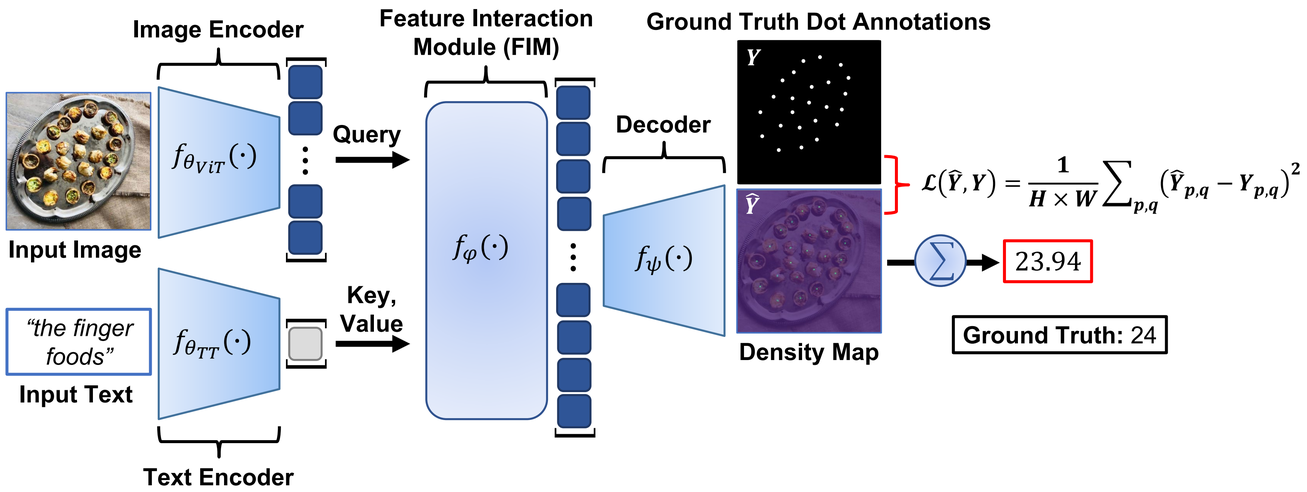
The objective of this paper is open-world object counting in images, where the target object class is specified by a text description. Classes unseen during training may be encountered during inference. To this end:

- We present ***CounTX***, a novel open-world object counting model that accepts an image and an ***arbitrary*** class description and directly uses these inputs to predict the object count.
- We augment the open-world object counting dataset, FSC-147, with class descriptions and release the enhanced dataset, FSC-147-D, for future research on open-world text-specified object counting.
- We demonstrate that CounTX achieves SOTA on FSC-147 for methods that use text, and CounTX performs competitively on the CARPK car-counting dataset. We also include qualitative results on the CountBench dataset.



CounTX

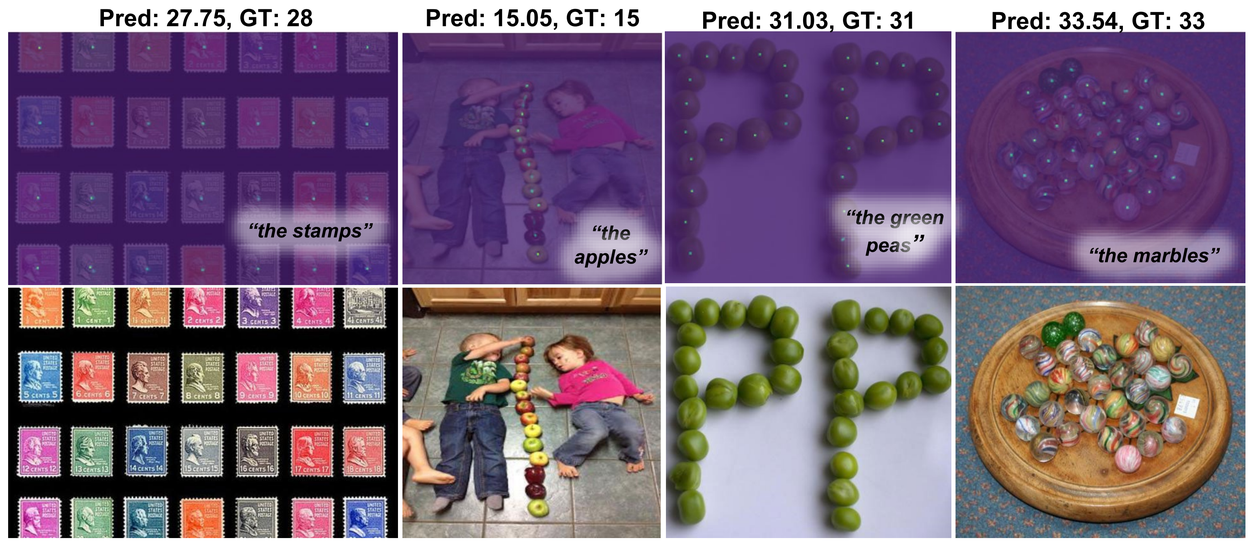
- Uses pretrained CLIP vision and text transformers to map image and text to joint image-text embedding space.
- Transformer decoder layers in FIM relate image and text features as query, key, and value vectors respectively.
- FIM attention map is upsampled in decoder, generating a density map matching the dots and summing to object count.



FSC-147

Method	Year	Published	How to Specify the Class	Validation MAE	Validation RMSE	Test MAE	Test RMSE
RepRPN-Counter	2022	✓	None	31.69	100.31	28.32	128.76
RCC	2022	✗	None	20.39	64.62	21.64	103.47
CounTR (0-shot)	2022	✓	None	17.40	70.33	14.12	108.01
LOCA (0-shot)	2022	✗	None	17.43	54.96	16.22	103.96
Patch-selection	2023	✓	Text (class name)	26.93	88.63	22.09	115.17
CounTX (FSC-147-D)	**2023**	-	**Text (class name)**	17.70	**63.61**	**15.73**	106.88
CounTX (FSC-147-D)	**2023**	-	**Text (FSC-147-D)**	**17.10**	65.61	15.88	**106.29**
CounTR (3-shot)	2022	✓	3 Visual Exemplars	13.13	49.83	11.95	91.23
LOCA (3-shot)	2022	✗	3 Visual Exemplars	10.24	32.56	10.79	56.97

CounTX ***achieves SOTA on FSC-147*** across all measures ***for methods that use text to specify the task.***



CountBench

CARPK

Project page: https://www.robots.ox.ac.uk/~vgg/research/countx/