Exploring 3D Modeling and Counting as Tools for Precision Agriculture

Dr. Minglun Gong

School of Computer Science, University of Guelph

(Based on collaboration with Shenzhen Univ. and the PhD. dissertation of Dr. Mingjie Wang)

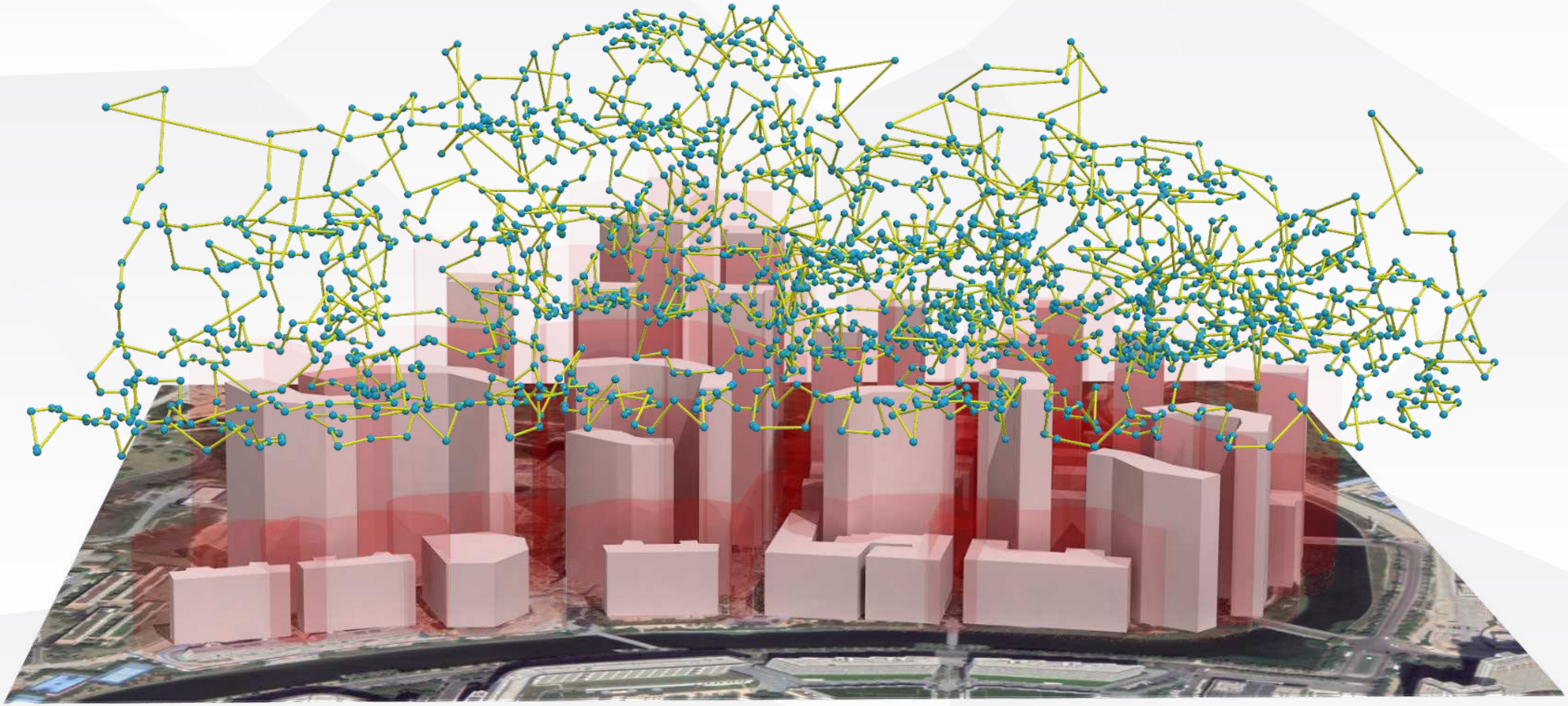
UNIVERSITY of GUELPH

UNIVERSITY of GUELPH

Large-scale 3D Modeling Problem

- Overview:
  - Involves creating detailed and complex three-dimensional digital representations of real-world objects, environments, and systems on a large scale.
- Applications in Digital Agriculture:
  - Crop Monitoring: Detailed modeling of crop fields for better monitoring and management.
  - Soil Analysis: 3D representations of soil layers and compositions for optimized farming practices.
  - Precision Farming: Enhanced decision-making through accurate models of farmland and crop growth.
  - Irrigation Management: Designing and simulating efficient irrigation systems.
  - Landscape Planning: Optimizing land use and layout for agricultural purposes.

Example in Urban Modeling



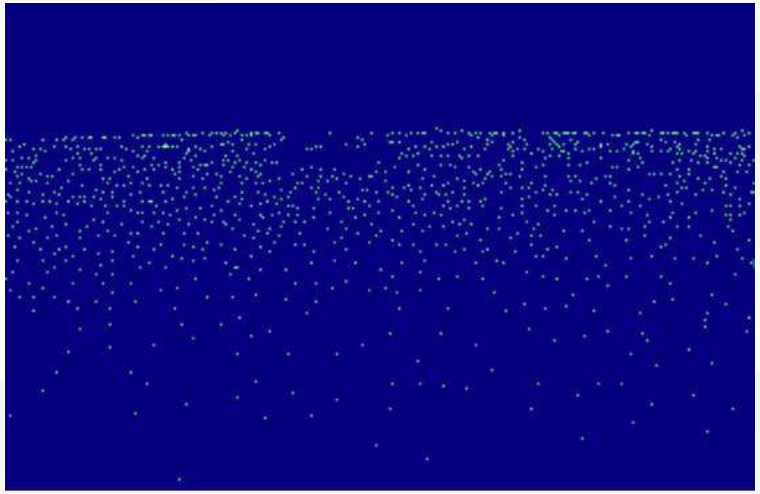
Object Counting Problem

- Overview:
  - Involves the automated detection and enumeration of objects within a given area using advanced imaging and processing techniques.
  - Essential for precision agriculture, enabling farmers to make data-driven decisions to optimize productivity and resource use.
- Applications in Digital Agriculture:
  - Yield Estimation: Accurate predictions of crop yield by counting plants, fruits, or grains.
  - Growth Monitoring: Tracking plant growth stages and identifying anomalies early.
  - Animal Tracking: Monitoring livestock numbers and movement patterns for health and management.
  - Resource Allocation: Efficient distribution of feed and other resources based on accurate animal counts.
  - Insect Counting: Detecting and counting pests to assess infestation levels and plan interventions.
  - Disease Spread Monitoring: Tracking disease symptoms across fields to control outbreaks.

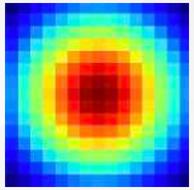
Ground Truth Annotation & Density Map

Input Crowd Image

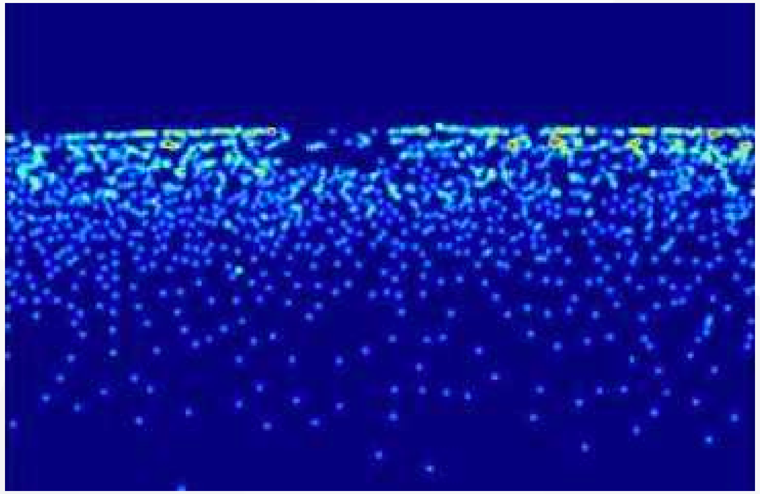
$$F(x) = \sum_{i=1}^{N} \delta(x - x_i) \times G_{\sigma_i}(x)$$



Dot Map



Gaussian Kernel, G



Density Map

Evaluation Metrics

- Mean Absolute Error (MAE):
  - $MAE = \frac{1}{N}\sum_{i=1}^{N}|C_i - C_i^{GT}|$
- Mean Square Error (MSE):
  - $MSE = \frac{1}{N}\sum_{i=1}^{N}\left(C_i - C_i^{GT}\right)^2$
- Root Mean Square Error (RMSE):
  - $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(C_i - C_i^{GT}\right)^2}$
- Notations:
  - N is the number of test images;
  - C_i denotes the predicted count;
  - C_i^{GT} is the ground truth of count number annotated for the i_{th} input image.

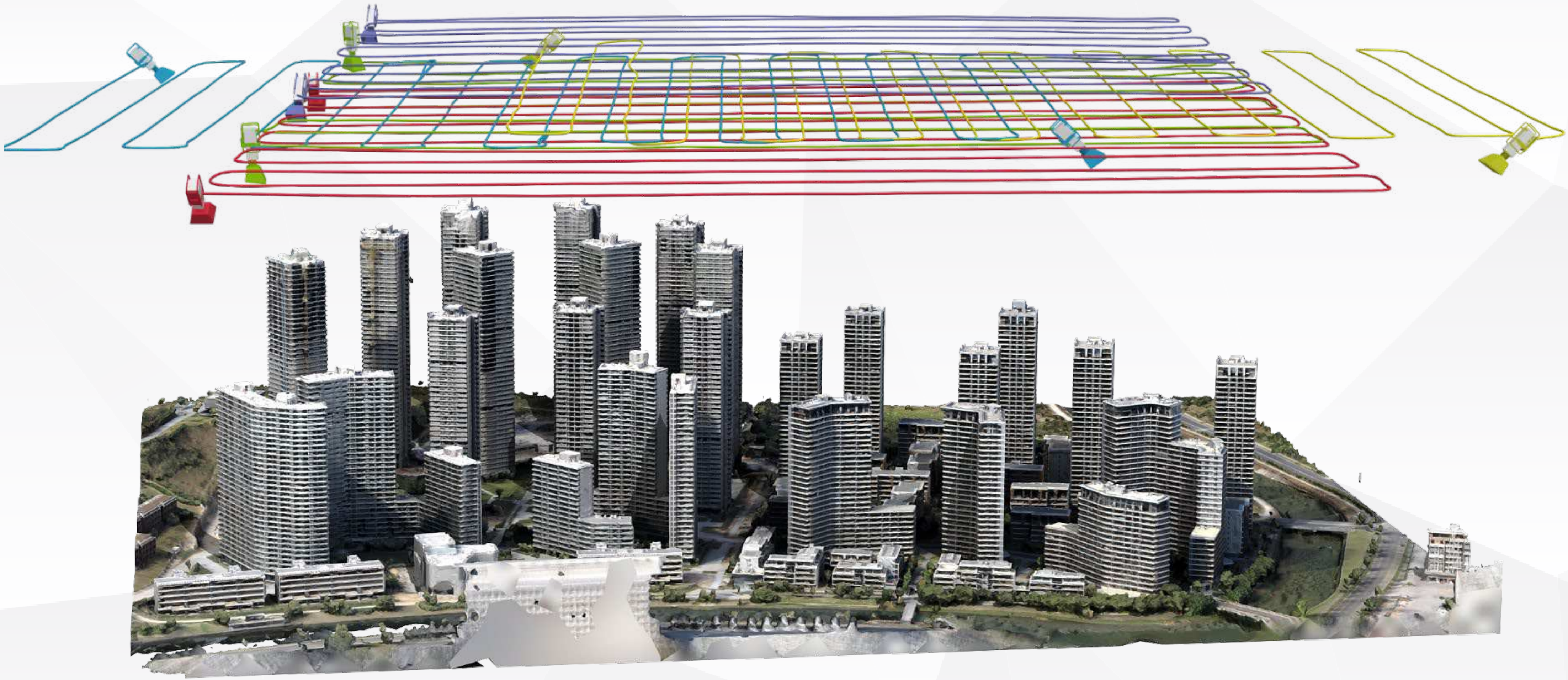
UNIVERSITY of GUELPH

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- General Object Counter
- Conclusion

Aerial Path Planning for Offsite 3D Scene Modeling

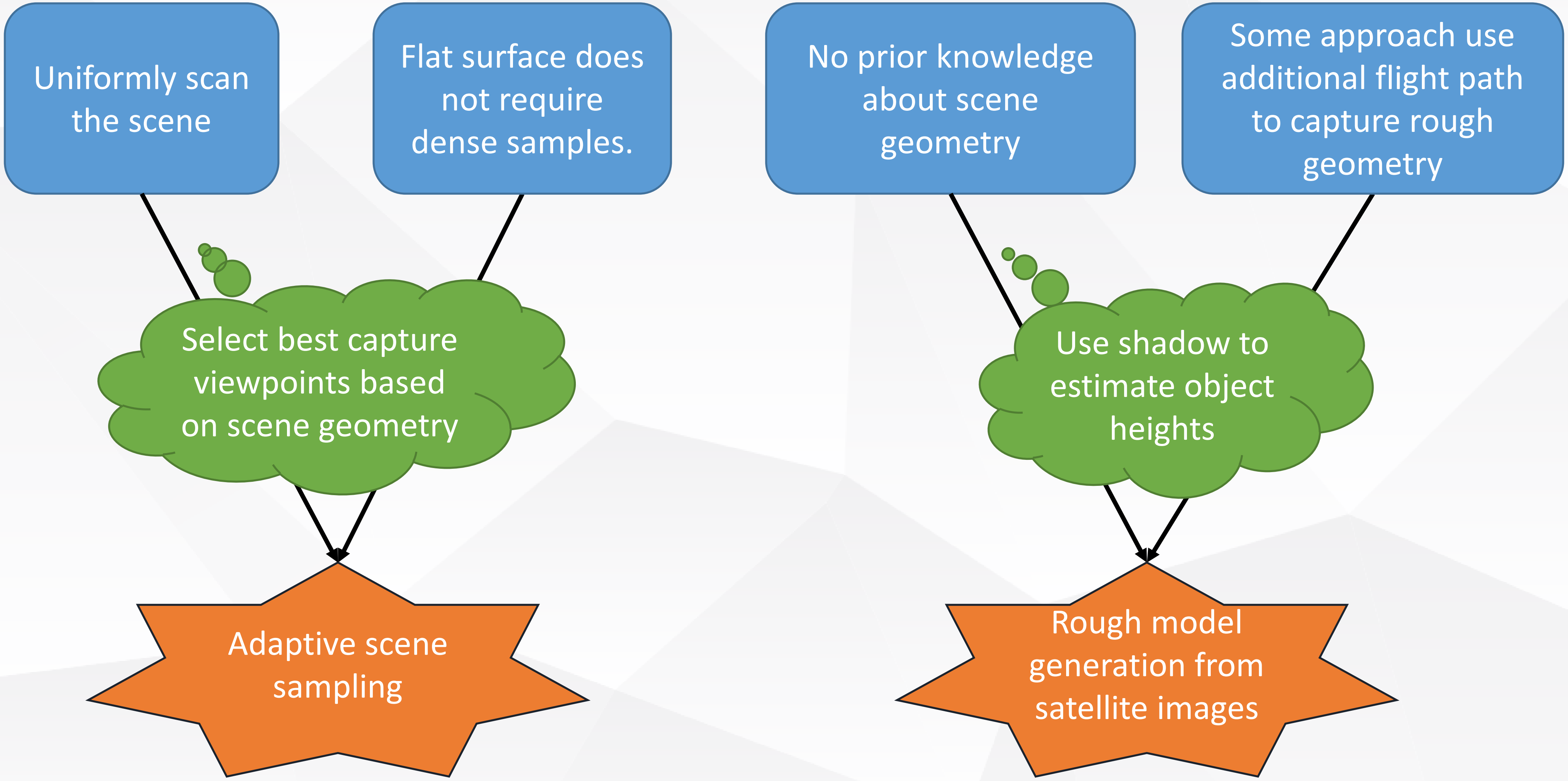
SIGGRAPH Asia, 2020

Existing Oblique Photography Approach



UNIVERSITY of GUELPH

Motivations



UNIVERSITY of GUELPH

Footprint & Shadow Detection

Footprint Extraction

Input Satellite Image

Shadow Detection

UNIVERSITY of GUELPH

2.5D Proxy Construction

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- General Object Counter
- Conclusion

UNIVERSITY of GUELPH

Samples on Proxy Surface

Introduction

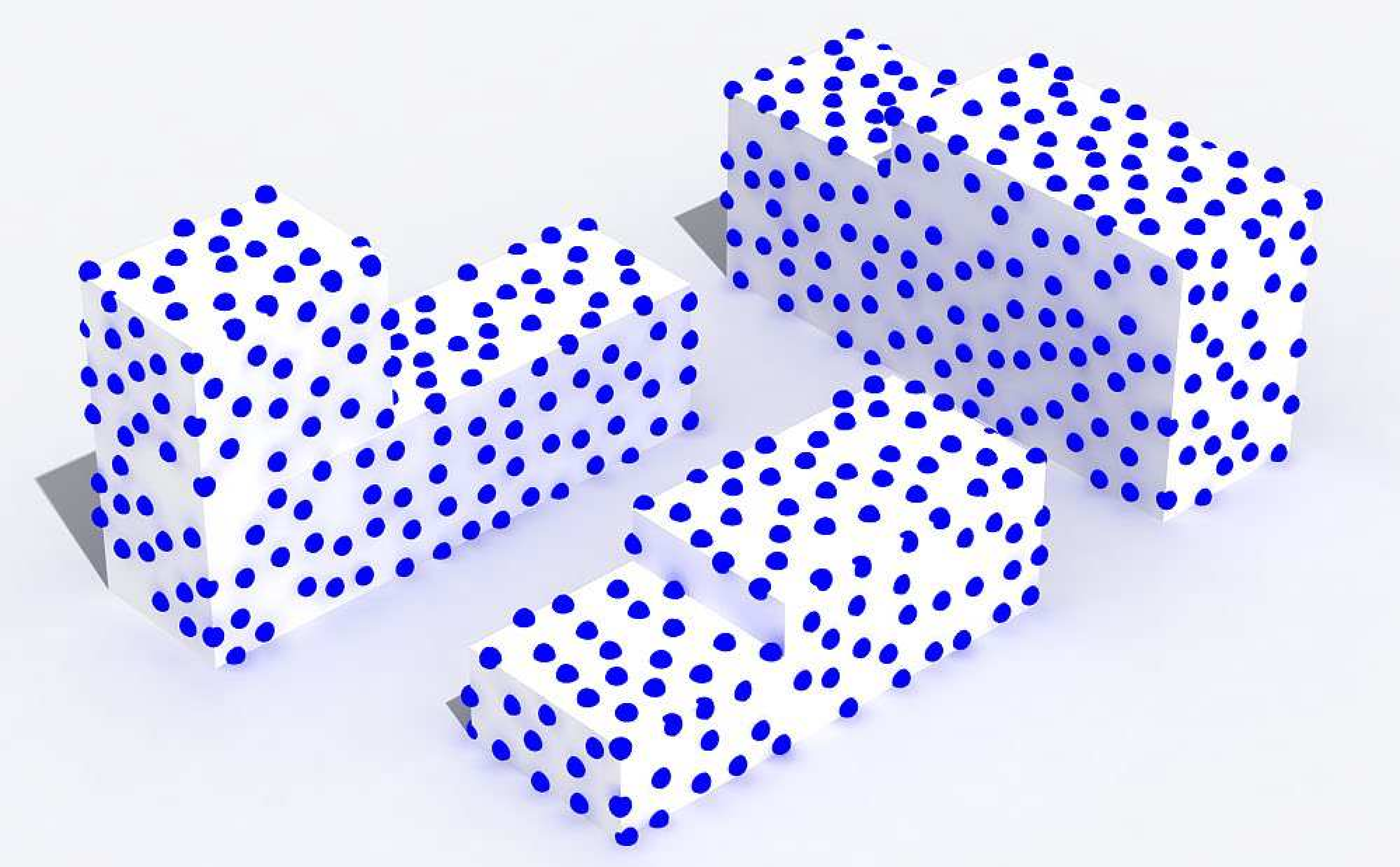
Path Planning for Offsite Modeling

Path Planning for Online Exploration

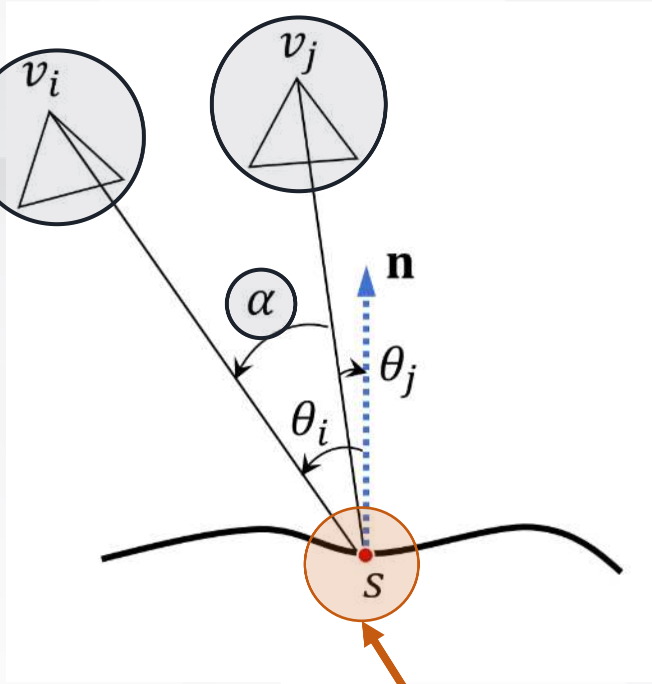
Weakly-Supervised Counter

General Object Counter

Conclusion



Reconstructability Heuristics



$$q(s,v_i,v_j)=w_1(\alpha)w_2(d_m)w_3(\alpha)cos(\theta_m)$$

[Smith et al. 2018]

$$w_1(\alpha)=(1\quad exp(\quad k_1*(\alpha\quad \alpha_1)))^{-1}$$

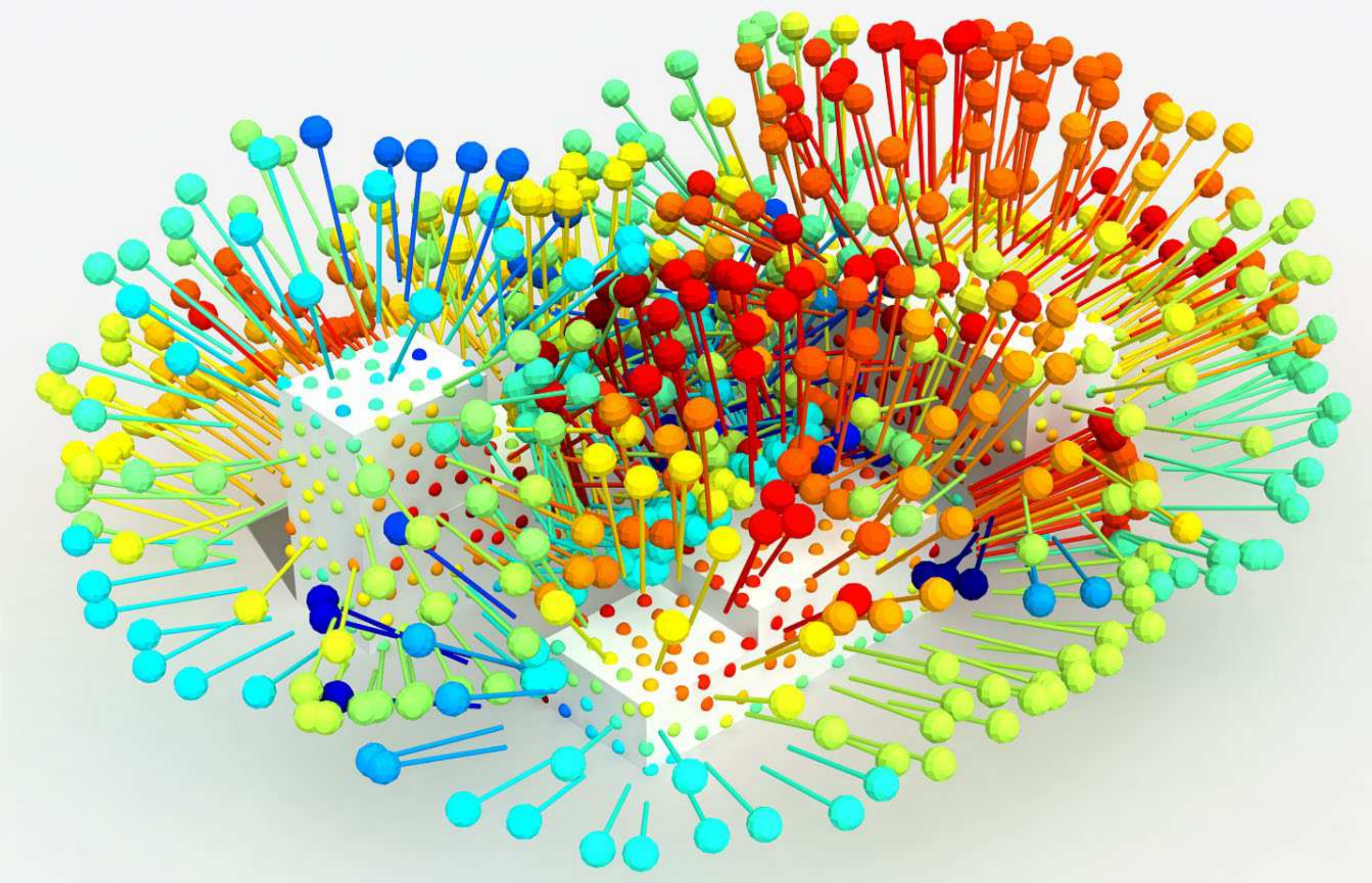
$$w_2(d_m)=1-min(d_m/d_{max},1) \qquad d_m=max(\|\overrightarrow{sv_i}\|,\|\overrightarrow{sv_j}\|)$$

$$w_3(\alpha)=1-(1+exp(-k_3*(\alpha-\alpha_3)))^{-1}$$

where $\delta(s,v)$ is a binary function, which equals to 1 if sample s is visible on view v, and 0 if not.

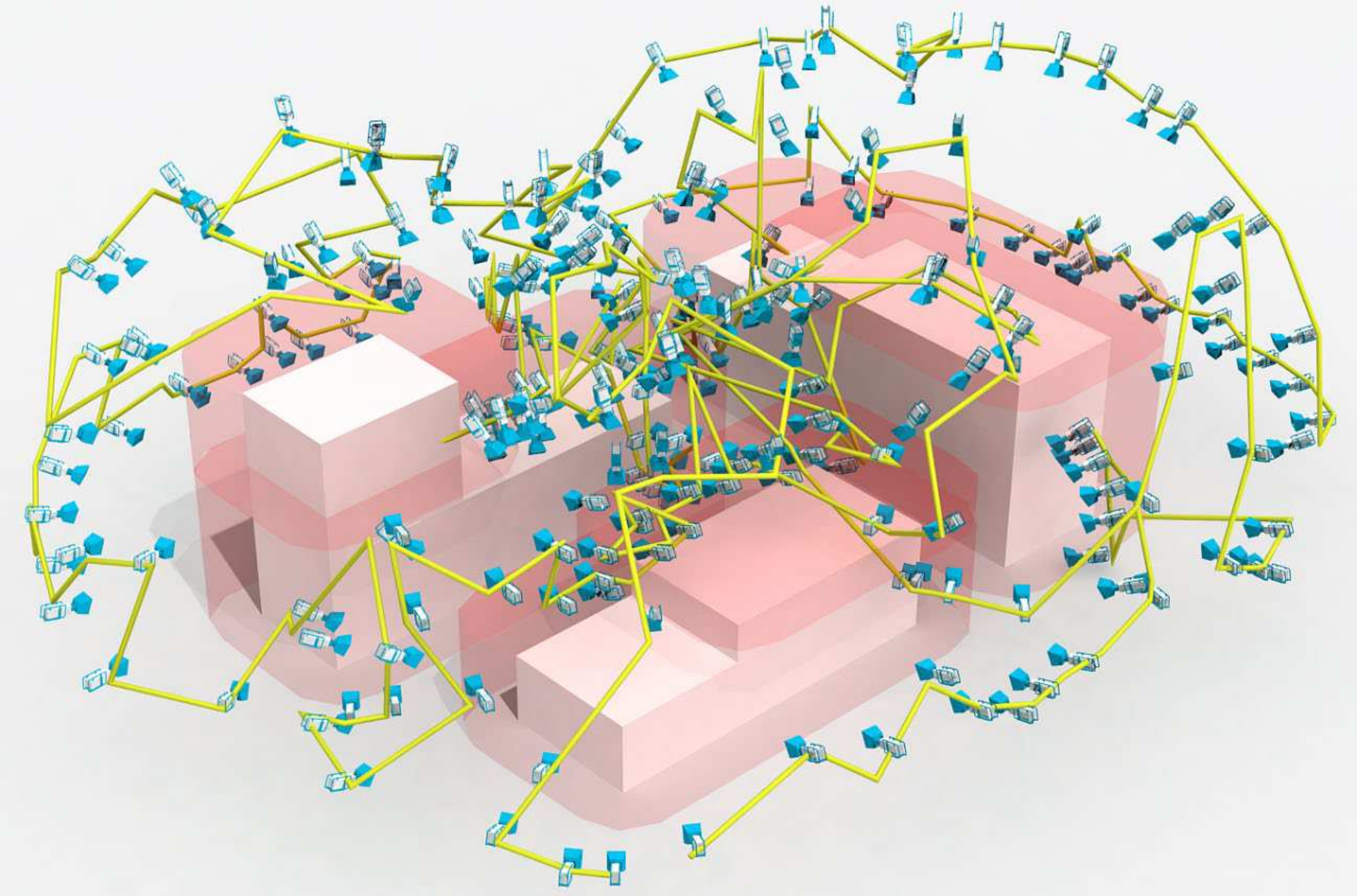
$$h(s,U)=\sum_{\substack{i=1,\ldots,|U| \\ j=i+1,\ldots,|U|}}\delta(s,v_i)\delta(s,v_j)q(s,v_i,v_j)$$

Reconstructability of Samples & Redundancy of Views



UNIVERSITY of GUELPH

TSP Path Planning



UNIVERSITY of GUELPH

Multi-View Reconstruction Result

UNIVERSITY of GUELPH

Comparison on Reconstruction Quality

Scene	Geometry Proxy	Method	#Images	Path Len(m)	Error 90% (m)	Error 95% (m)	Comp 0.02m(%)	Comp 0.05m(%)	Comp 0.075m(%)	Average Distance (m)
NY-1	Recon	[Smith et al. 2018]	433	2808	0.053	0.792	36.01	44.74	49.47	0.029
		Ours	262	**2197**	**0.030**	**0.342**	**38.19**	**45.22**	**49.78**	**0.020**
	2.5D	[Smith et al. 2018]	433	2876	0.037	0.675	39.59	46.97	51.46	0.022
		Ours	248	**1816**	**0.028**	**0.315**	**40.24**	**47.57**	**52.04**	**0.020**
UK-1	Recon	[Smith et al. 2018]	923	7819	**0.028**	**0.051**	**32.04**	**37.74**	**40.62**	**0.030**
		Ours	418	**4696**	0.030	0.054	30.75	35.96	38.77	0.032
	2.5D	[Smith et al. 2018]	918	7448	**0.027**	0.052	**31.83**	**37.29**	40.11	**0.030**
		Ours	418	**4648**	0.028	**0.049**	30.75	35.96	**40.29**	0.031
Bridge-1	Recon	[Smith et al. 2018]	565	4548	**0.015**	0.024	**52.97**	**59.29**	**62.53**	**0.020**
		Ours	389	**4241**	**0.015**	**0.023**	52.26	59.05	62.28	**0.020**
	2.5D	[Smith et al. 2018]	565	4569	**0.014**	**0.022**	49.47	55.73	59.07	0.022
		Ours	372	**4217**	0.015	0.023	**52.78**	**59.81**	**63.06**	**0.020**

Visual Comparison

Introduction

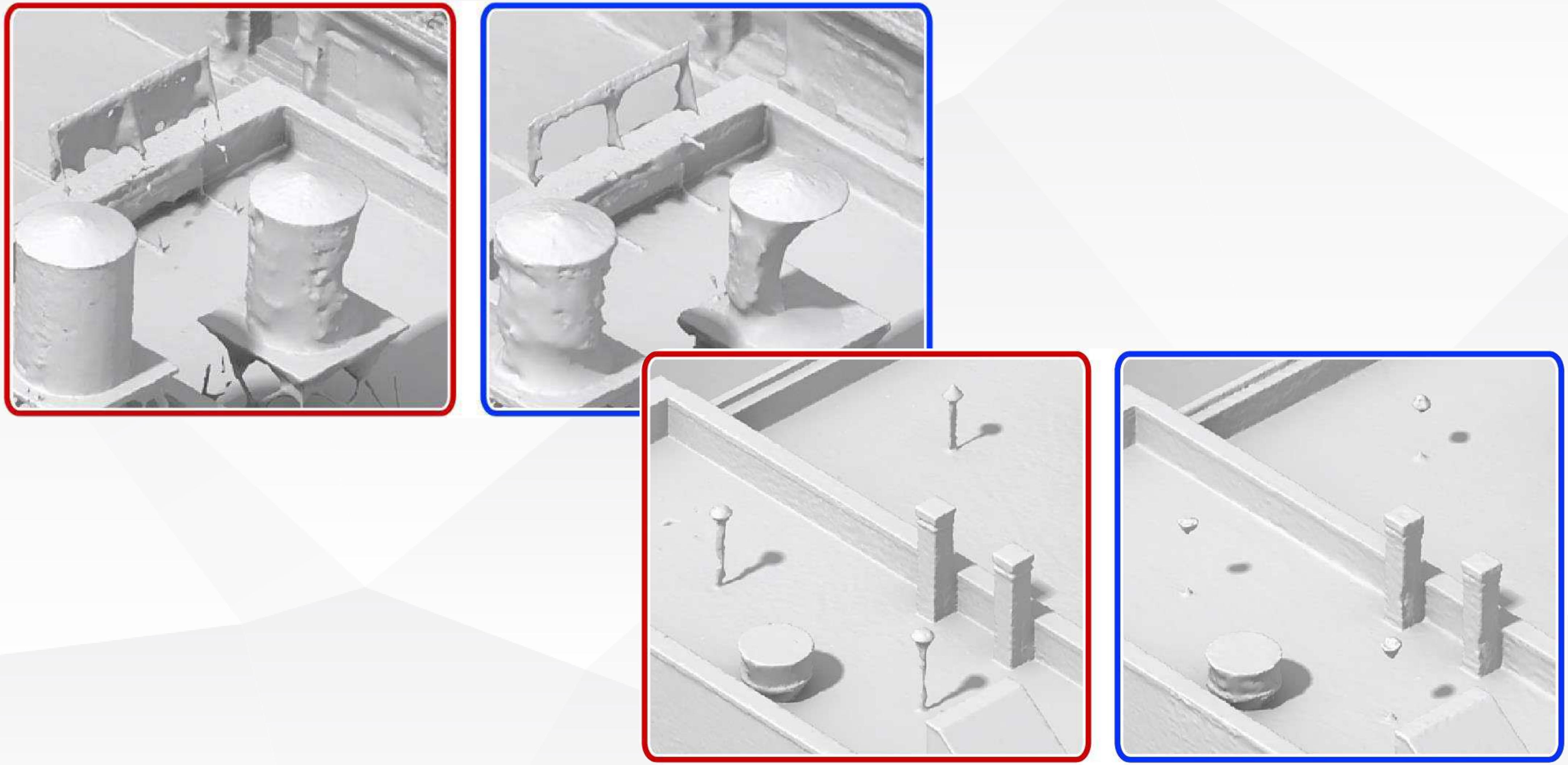
Path Planning for Offsite Modeling

Path Planning for Online Exploration

Weakly-Supervised Counter

General Object Counter

Conclusion



UNIVERSITY of GUELPH

Introduction

Path Planning for Offsite Modeling

Path Planning for Online Exploration

Weakly-Supervised Counter

General Object Counter

Conclusion

Aerial Path Planning for Online Exploration and Offline Modeling

SIGGRAPH Asia 2021

Priori Requirement in Existing Approaches

- Coarse scene geometry is needed for designing flight paths:
  - To avoid collision between the UAV and the objects in the scene.
  - To optimize the viewpoint and view direction for image capture.
- Existing approaches for generating coarse models:
  - Manual generation: Users manually create rough models.
  - Initial Scene Capture: Conduct an initial flight to capture the scene, followed by process the data offline to generate a coarse model.
  - Our Siggraph Asia 2020 Approach: Requires satellite images captured on sunny days to utilize shadows for better depth estimation.

UNIVERSITY of GUELPH

Motivations

All existing approaches require priori knowledge for the scene

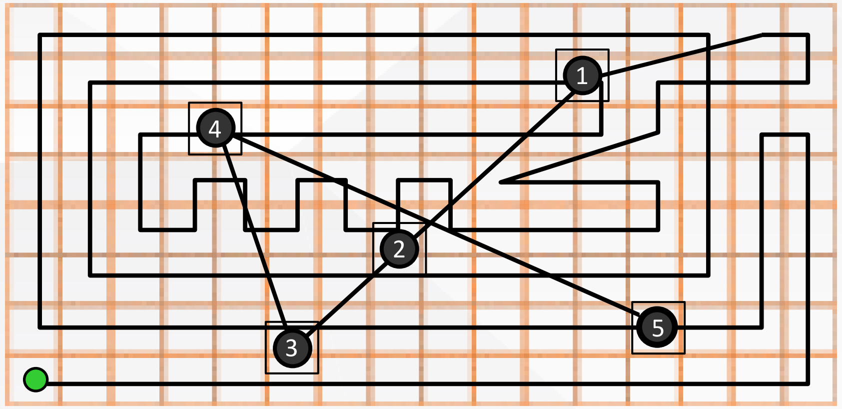
Long preparation time needed for capturing unknown scenes

How to explore and reconstruct a region without any prior?

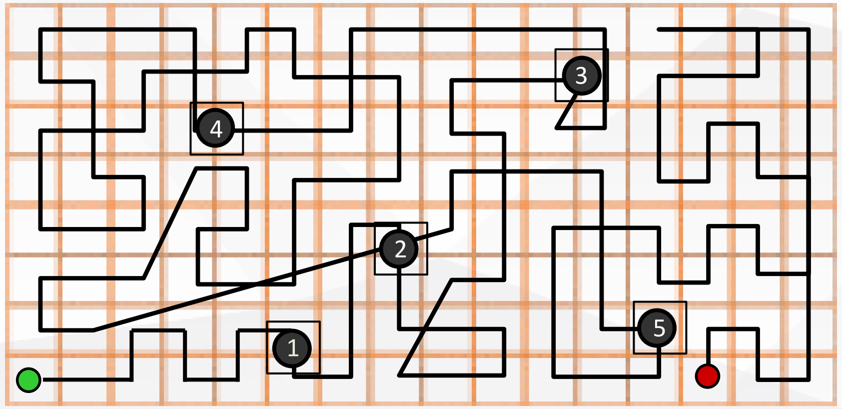
Generate proxy models online and use them to guide the capture

UNIVERSITY of GUELPH

Proposed Region-Division Method

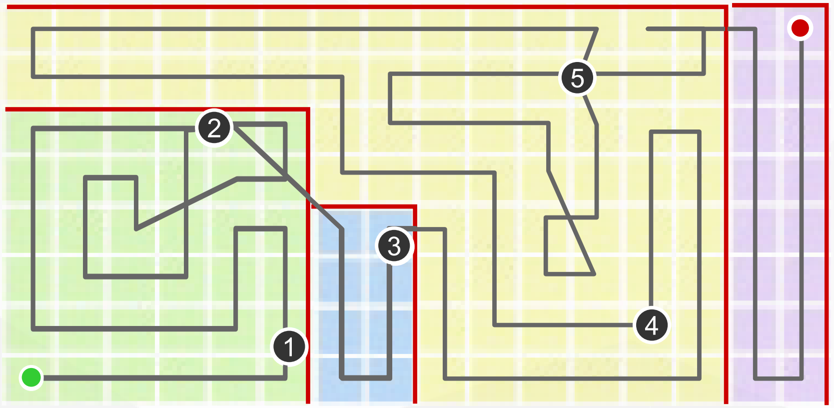


Two-Step 25170m



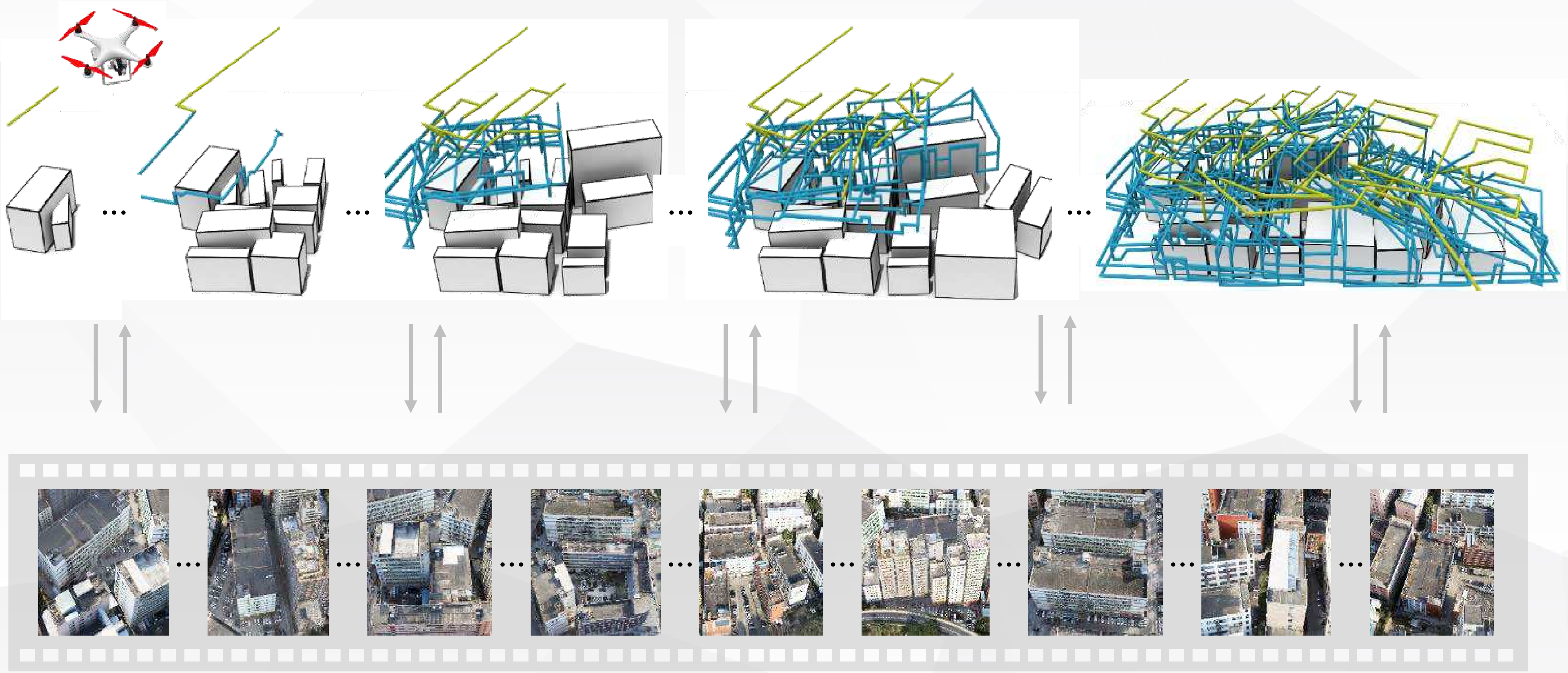
Greedy-Nearest 23990m

Region-Division

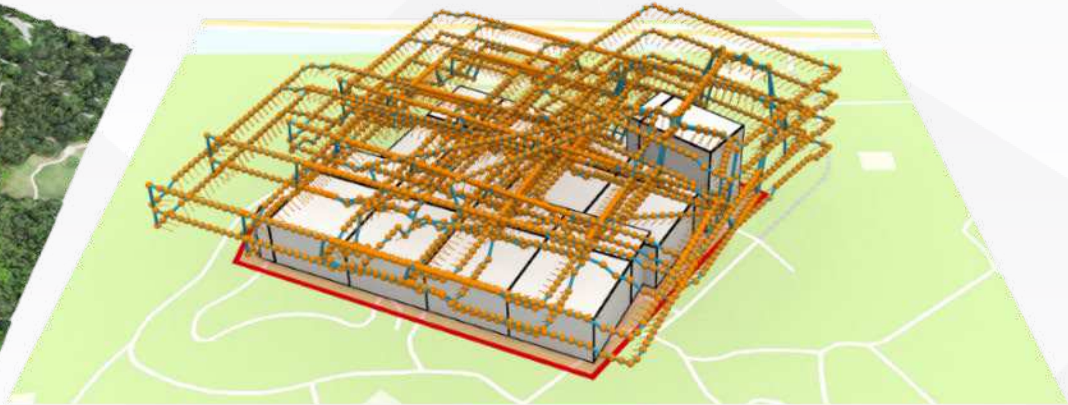


22150m

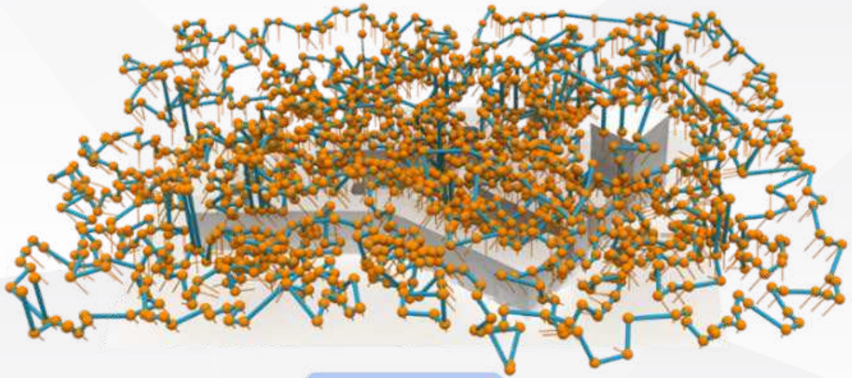
Example on Scene Exploration



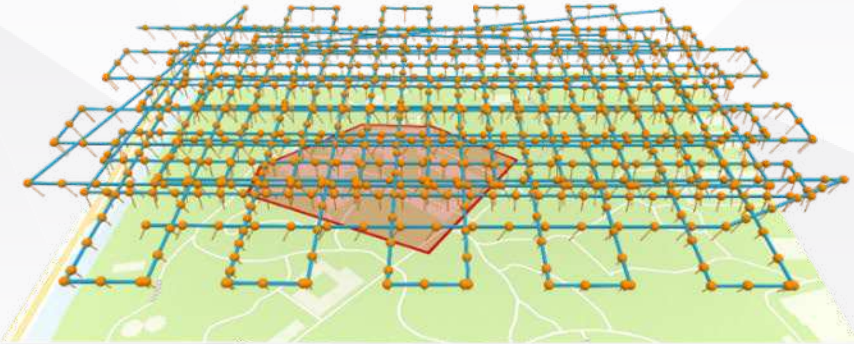
Experiment on Real Scene



Ours



SIGA20

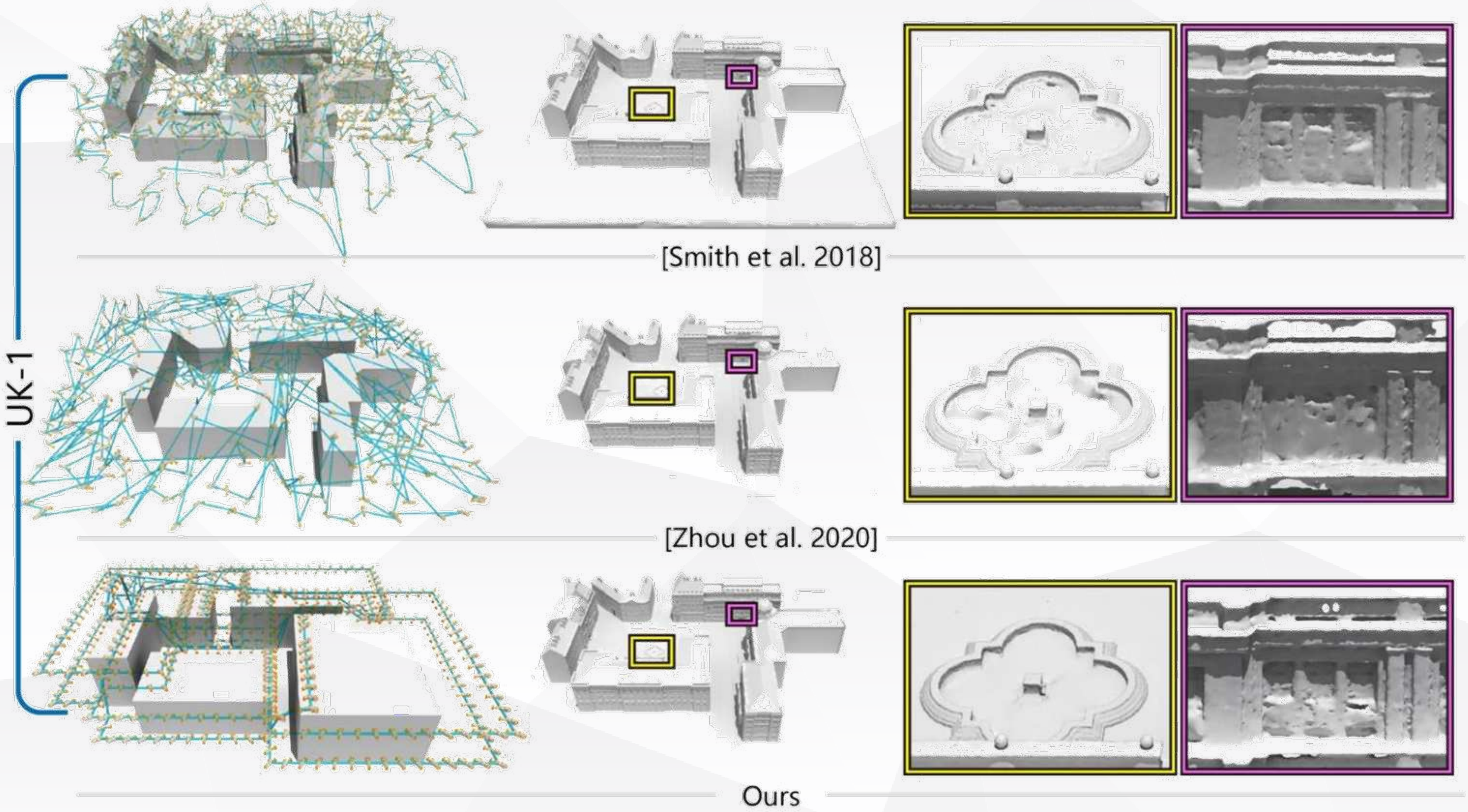


OP

UNIVERSITY of GUELPH

Performance Comparison

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- General Object Counter
- Conclusion



UNIVERSITY of GUELPH

Comparison on Reconstruction Quality

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- General Object Counter
- Conclusion

Scene	Method	#Images	Process Time (sec)	Path Len (m)	Flight Time (sec)	Error 90% (m)	Error 95% (m)	Comp. 0.05m (%)	Comp. 0.075m (%)
NY-1	Smith et al.	433	150	2876	1053	0.053	0.342	44.74	49.47
	Zhou et al.	248	120	1816	679	**0.028**	0.315	**47.57**	**52.04**
	Ours	**244**	**0.002**	**1605**	**338**	0.029	**0.107**	44.72	49.33
UK-1	Smith et al.	918	306	7448	2599	0.028	0.051	**37.74**	**40.62**
	Zhou et al.	**418**	246	4648	1451	0.028	**0.049**	35.96	40.29
	Ours	486	**0.003**	**3940**	**817**	0.028	0.052	36.71	39.58
Bridge-1	Smith et al.	565	222	4569	1482	0.015	0.024	59.29	62.53
	Zhou et al.	**372**	168	4217	1116	0.015	**0.023**	**59.81**	**63.06**
	Ours	395	**0.002**	**3397**	**761**	**0.013**	**0.023**	57.27	60.42

UNIVERSITY of GUELPH

MLP-based Weakly-supervised Counter & Self-supervised Proxy Task

Pattern Recognition, 2023

Introduction

Path Planning for Offsite Modeling

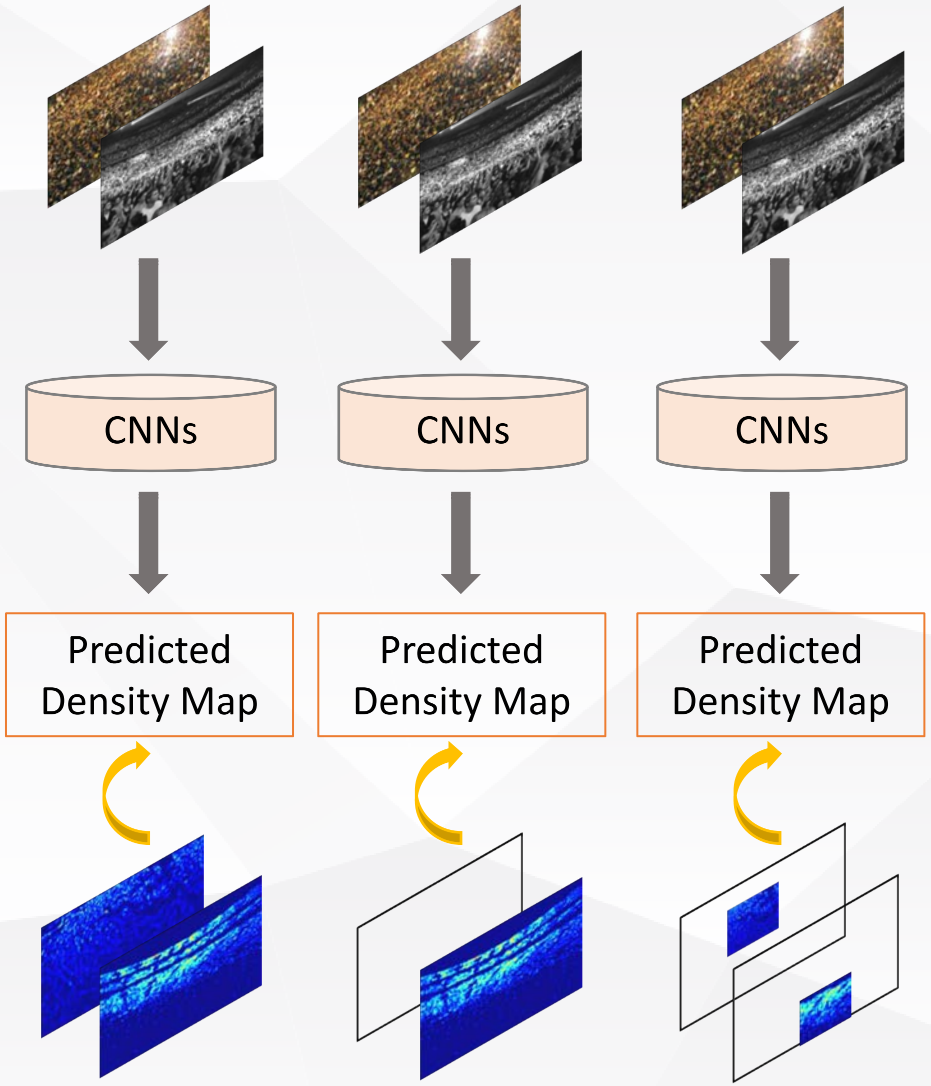
Path Planning for Online Exploration

Weakly Supervised Counter

General Object Counter

Conclusion

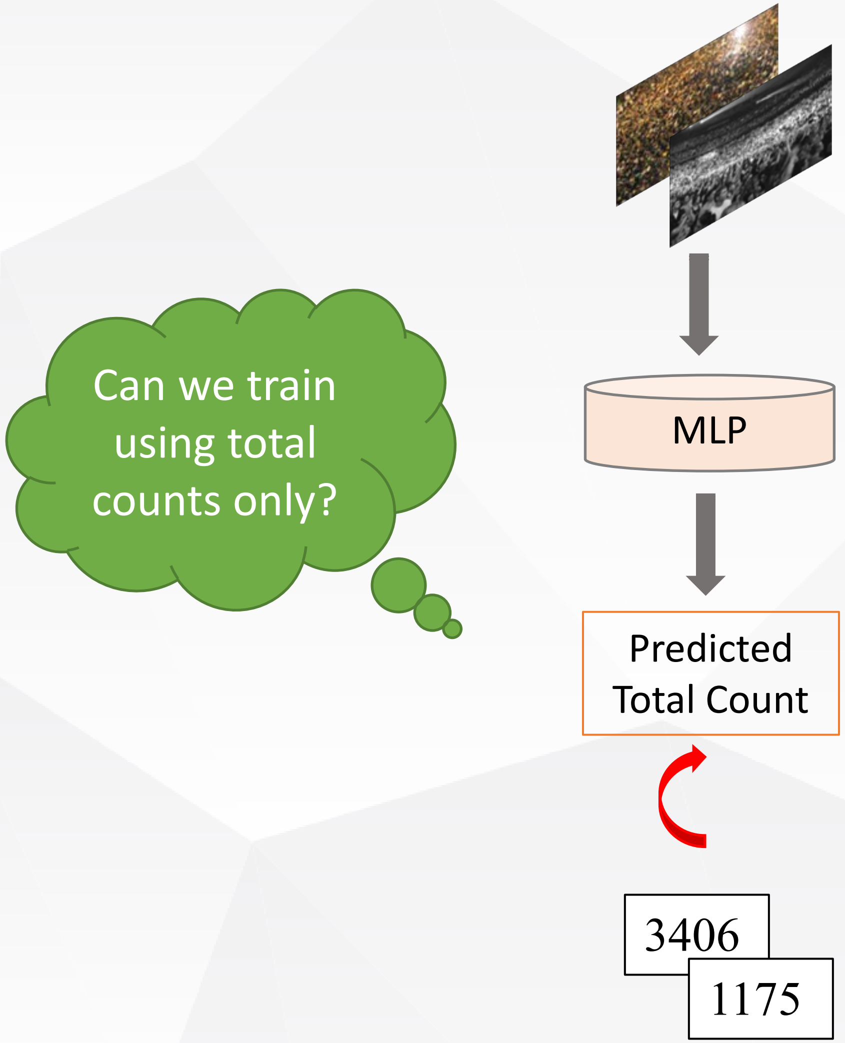
Existing Location-level Supervision Methods



- Collecting location-level annotation is time-consuming & labour-intensive.
  - Efforts have been made to reduce annotation needs.
- Strongly supervised:
  - Learn from an entire set of location-level density maps.
- Weakly supervised:
  - Learn from density maps of partially selected images.
  - Learn from density maps of partial regions in crowd scenes.

Limitations of Location-level Supervision

- Introduces a domain gap between training & inference phases
  - Models are trained to predict accurate2D density maps but evaluated solely on total counts.
- 2D density maps are derived from ground-truth dot maps through a heuristic procedure
  - May inductive bias that confuses the models on what to learn
- Collecting single counts can be much easier in many real-world scenarios.
  - E.g., scenes with controlled access



Motivations

Using single count values to train model is ambiguous

Lack of location cues causes performance degeneration

Use proxy task to guide model on what to learn

Self-supervised proxy task

Locally-focused CNN models cannot accurately & directly predict global count

Transformers require complex self-attention mechanisms & large amount training data

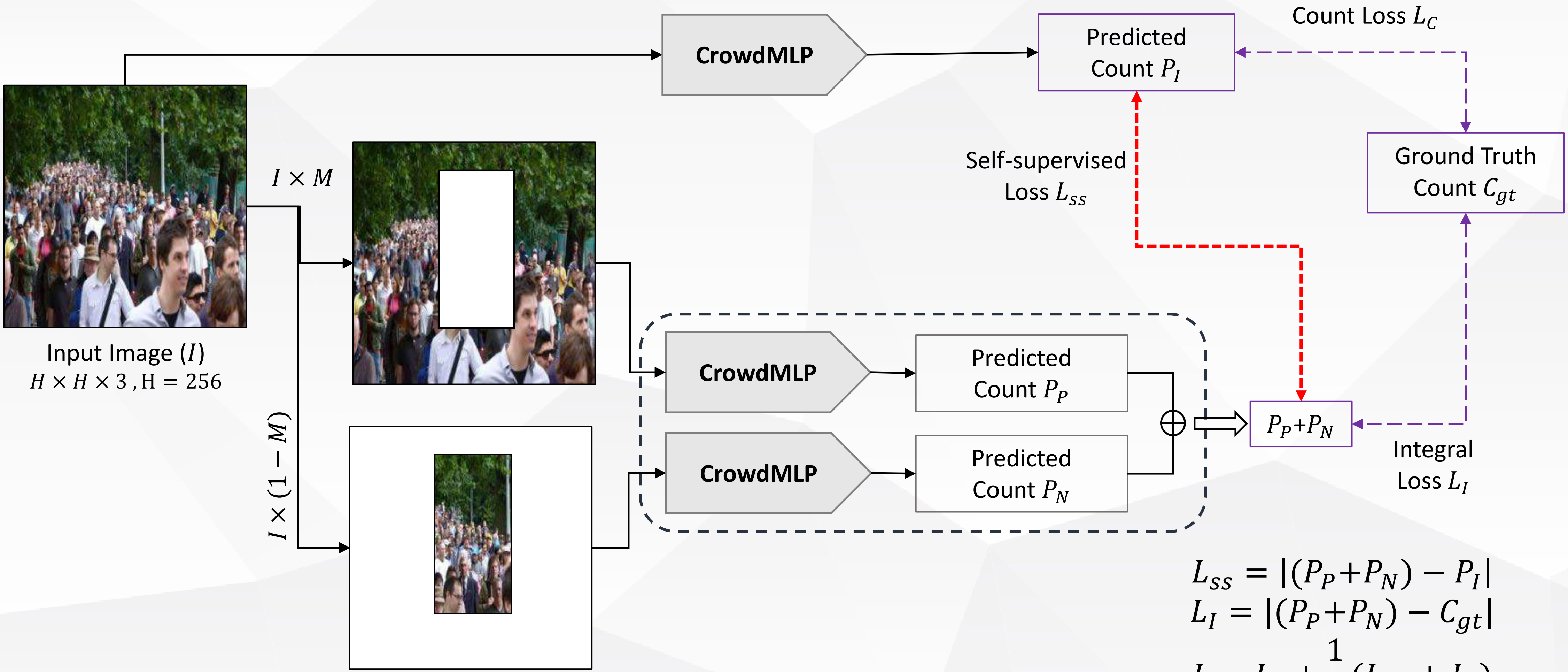
Add multi-layer perceptrons (MLP) for regression

Use pretrained CNN to distillate image features

Pretrained CNN frontend + MLP regressor

Overall Architecture w/ Split-Counting

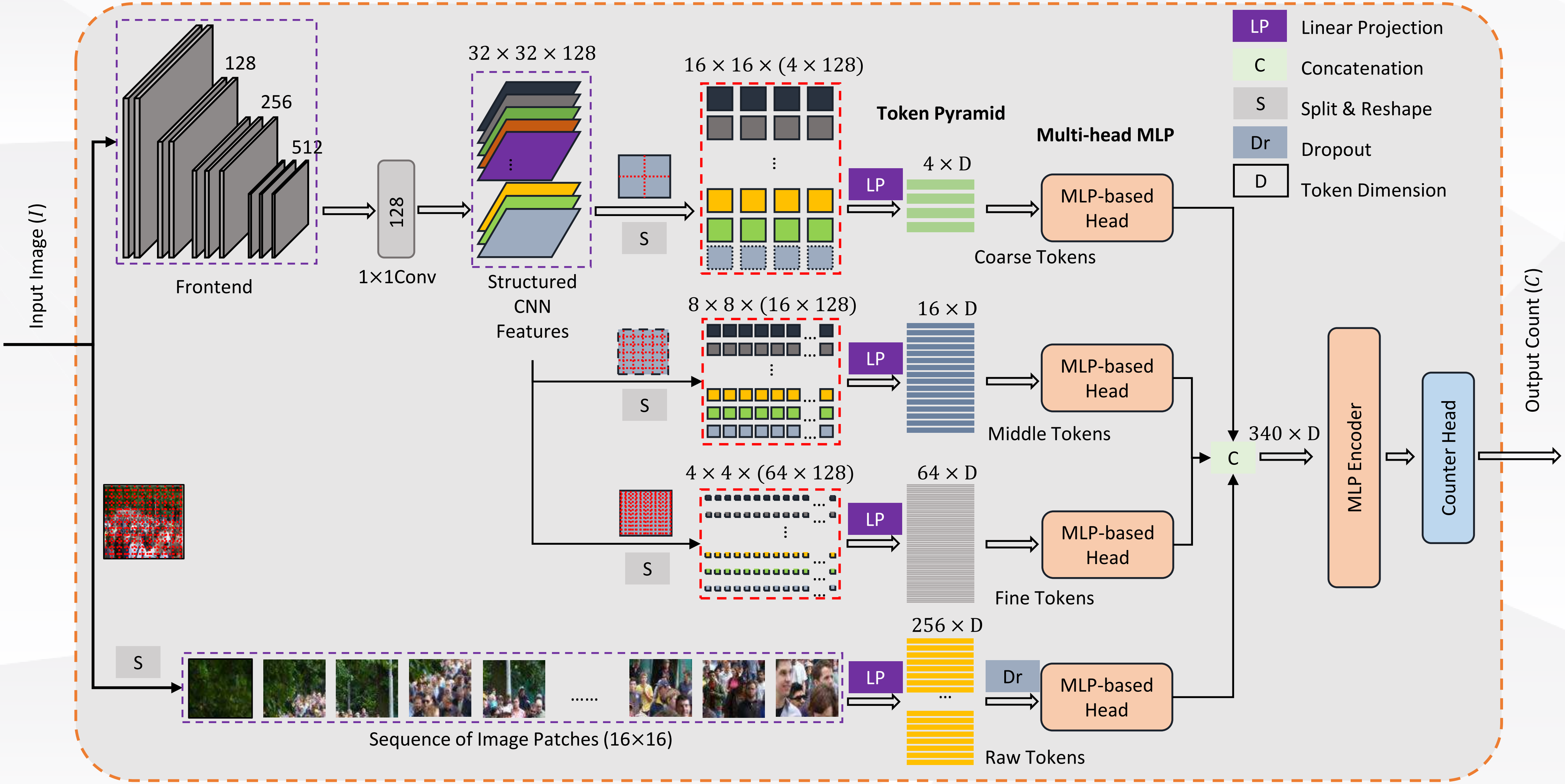
- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- **Weakly Supervised Counter**
- General Object Counter
- Conclusion



$$L_{ss} = |(P_P + P_N) - P_I|$$
$$L_I = |(P_P + P_N) - C_{gt}|$$
$$L = L_c + \frac{1}{2}(L_{SS} + L_I)$$

Details of the CrowdMLP Counter

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- **Weakly Supervised Counter**
- General Object Counter
- Conclusion



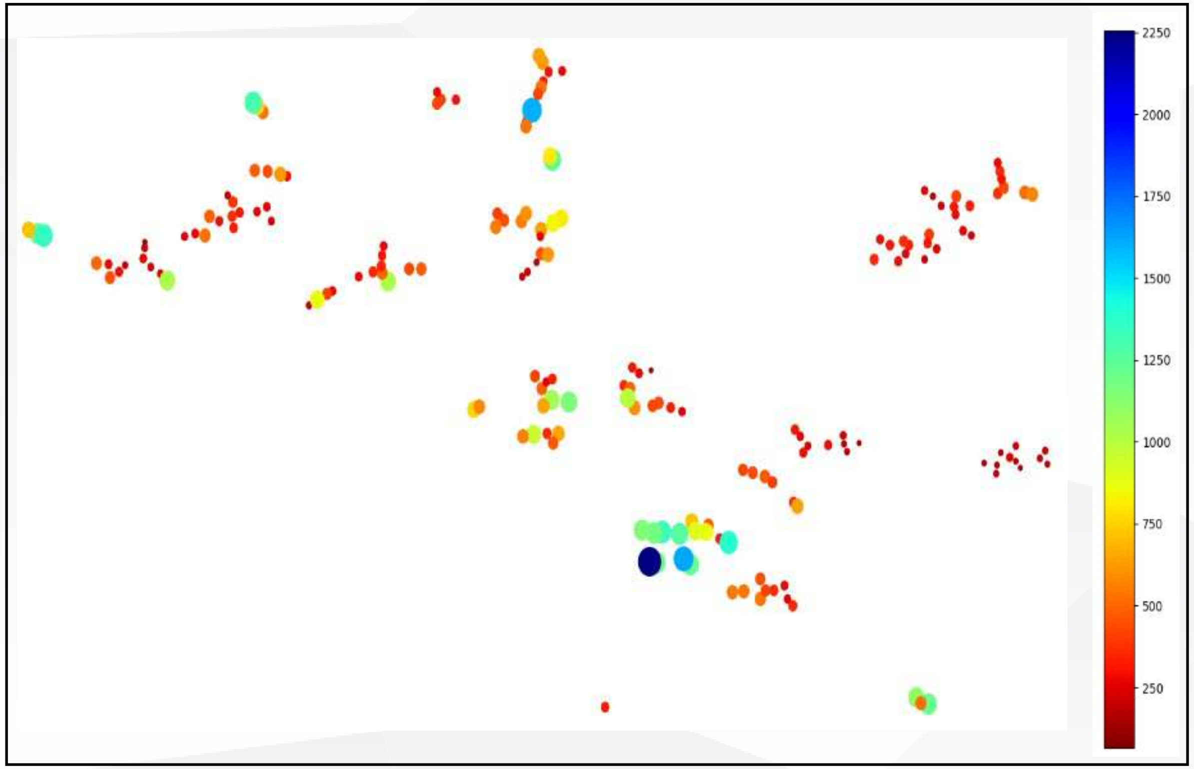
Quantitative Evaluation

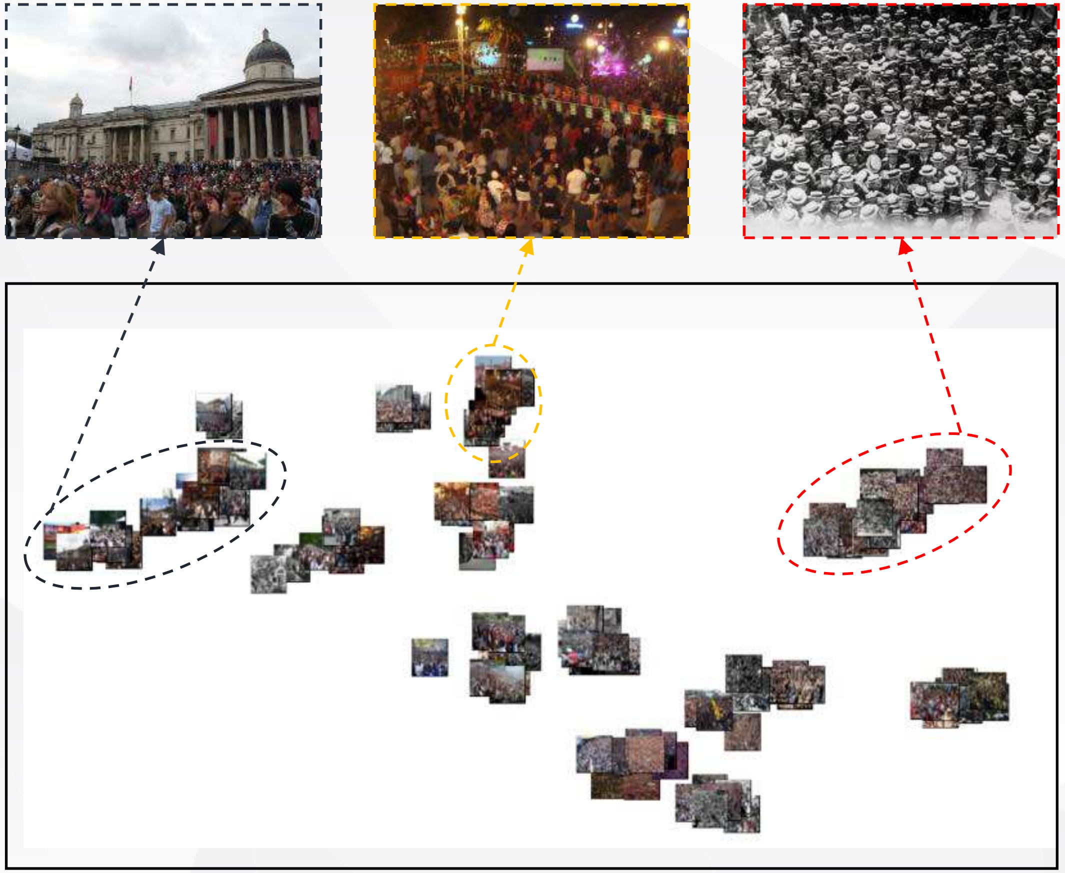
Methods	Location Label	Part A		Part B		UCF-QNRF		JHU++	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ADCrowdNet [77]	✓	63.2	98.9	7.6	13.9	-	-	-	-
PACNN [114]	✓	62.4	102.0	7.6	11.8	-	-	-	-
CAN [79]	✓	62.3	100.0	7.8	12.2	107	183	100.1	314.0
MBTTBF [123]	✓	60.2	94.1	8.0	15.5	97.5	165.2	81.8	**299.1**
DSSINet [76]	✓	60.63	96.04	6.85	10.34	99.1	159.2	133.5	416.5
S-DCNet [155]	✓	58.3	95.0	6.7	10.7	104.4	176.1	-	-
ASNet [54]	✓	57.78	90.13	-	-	91.59	159.71	-	-
AMRNet [81]	✓	61.59	98.36	7.02	11.00	86.6	152.2	-	-
DM-Count [146]	✓	59.7	95.7	7.4	11.8	85.6	**148.3**	-	-
BL [91]	✓	62.8	101.8	7.7	12.7	88.7	154.8	**75.0**	299.9
P2PNet [126]	✓	**52.7**	**85.0**	**6.2**	**9.9**	**85.3**	154.5	-	-
MATT [64]	×	80.1	129.4	11.7	17.5	-	-	-	-
Sorting [161]	×	104.6	145.2	12.3	21.2	-	-	-	-
TransCrowd-T [73]	×	69.0	116.5	10.6	19.7	98.9	176.1	76.4	319.8
TransCrowd-G [73]	×	66.1	105.1	9.3	16.1	97.2	**168.5**	74.9	295.6
CrowdMLP	×	**57.8**	**84.4**	**7.6**	**12.1**	**94.1**	170.3	**67.5**	**256.1**

UNIVERSITY of GUELPH

T-SNE Visualization on Feature Clusters

- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- **Weakly Supervised Counter**
- General Object Counter
- Conclusion





Ablation Studies

Methods	Parameters	Part A	
		MAE	MSE
w.o. Raw Token	23.70M	60.411	88.613
w.o. 16×16 Token	17.65M	61.896	93.894
w.o. 8×8 Token	23.74M	61.697	87.717
w.o. 4×4 Token	24.56M	59.294	87.483
Baseline	26.63M	57.828	84.412

The Effects of Coarse-to-fine Token Streams

Using Different Learning Paradigms

Methods	Parameters	Part A	
		MAE	MSE
Crowd-CNN	26.63M	159.925	251.649
Crowd-Transformer	36.43M	61.121	92.922
Baseline	26.63M	57.828	84.412

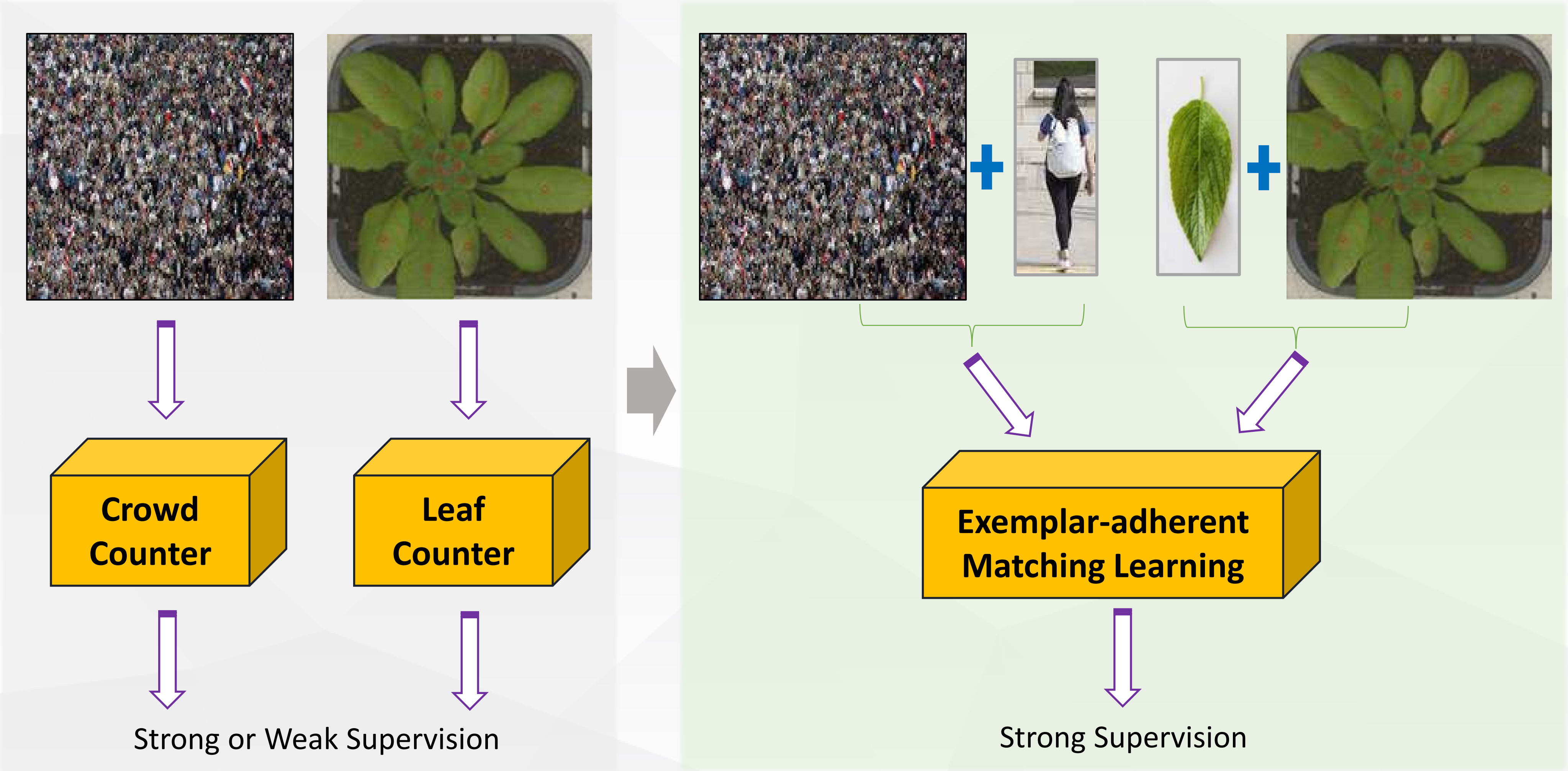
The effect of Proxy Split-Counting: 0.55% & 0.32% improvements on MAE & MSE

General Counting through Zero-shot Self-similarity Learning

Pattern Recognition, 2024

Existing Class-specific & Class-agnostic Counting

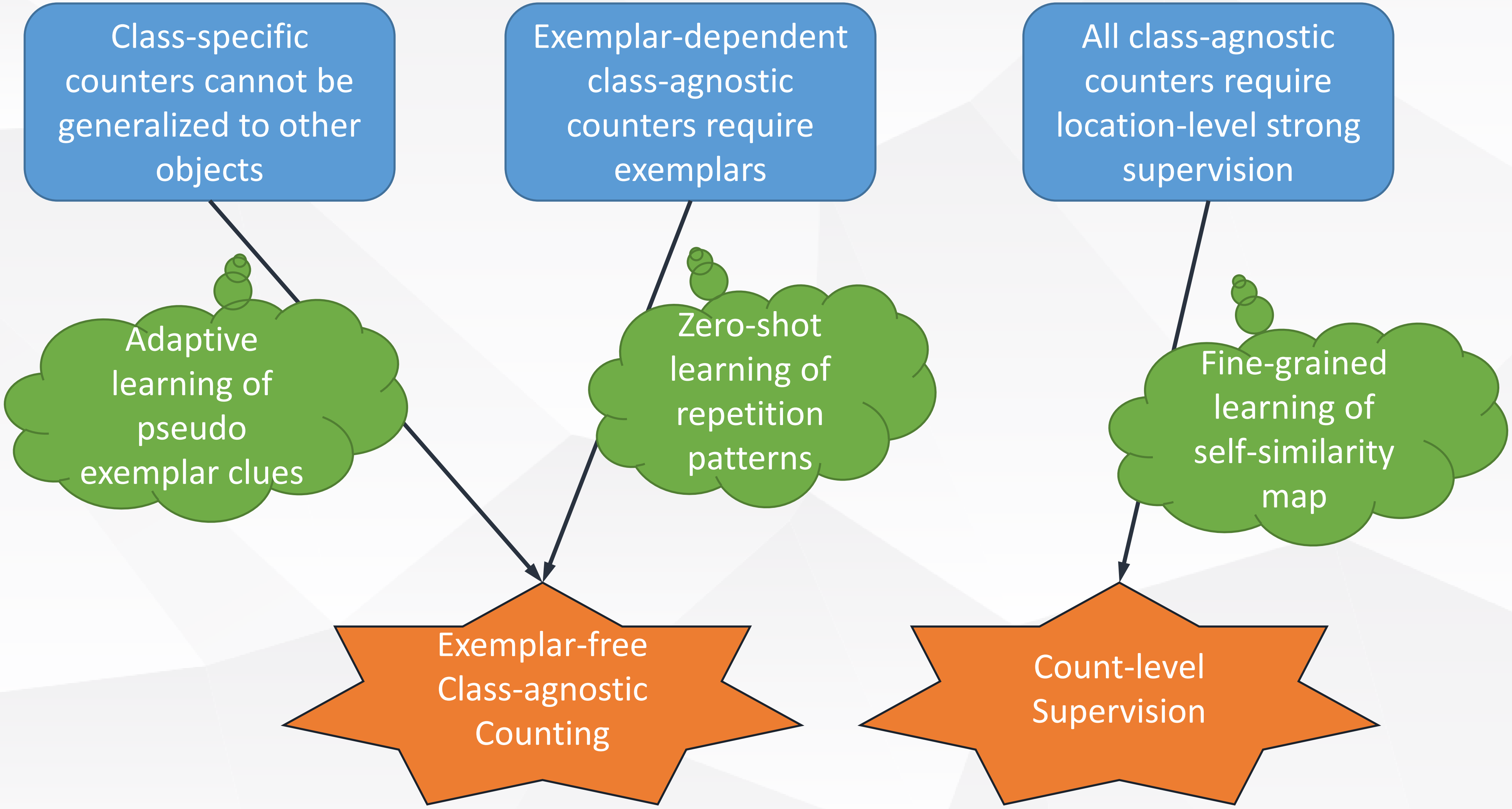
- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- General Object Counter
- Conclusion



(a) Class-specific Counting

(b) Exemplar-dependent Class-agnostic Counting

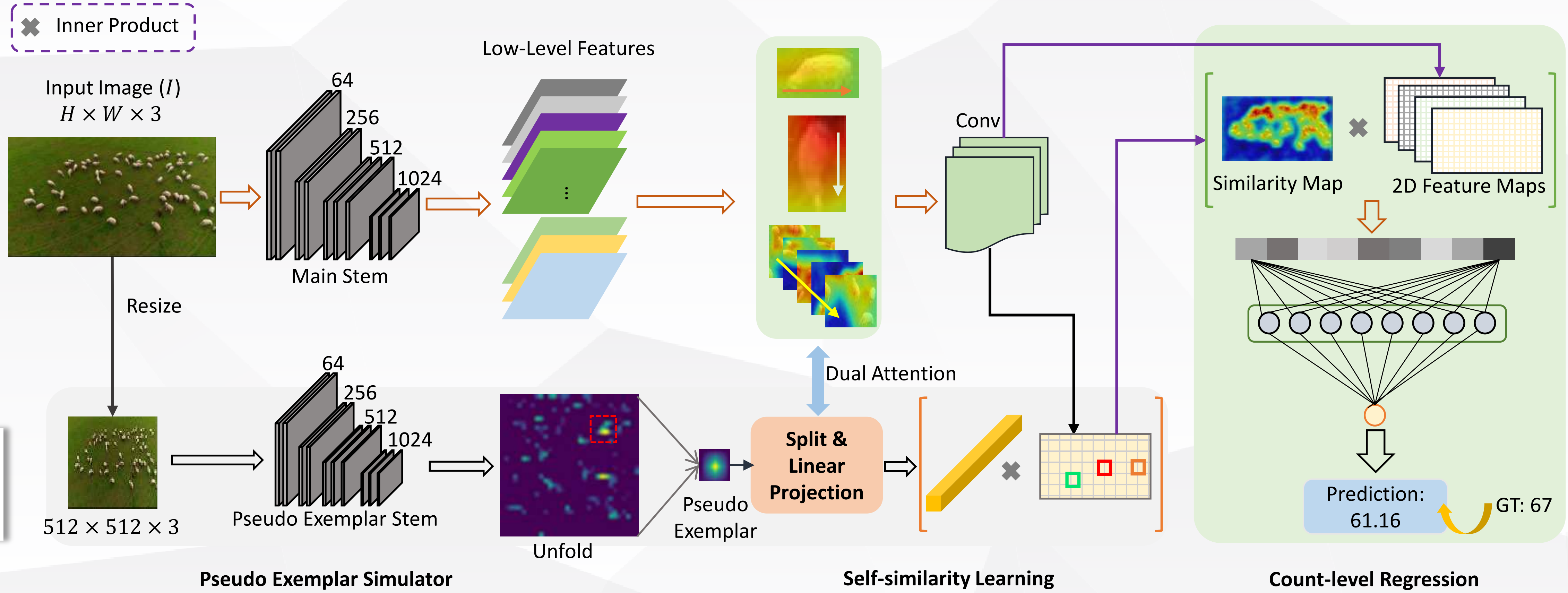
Motivations



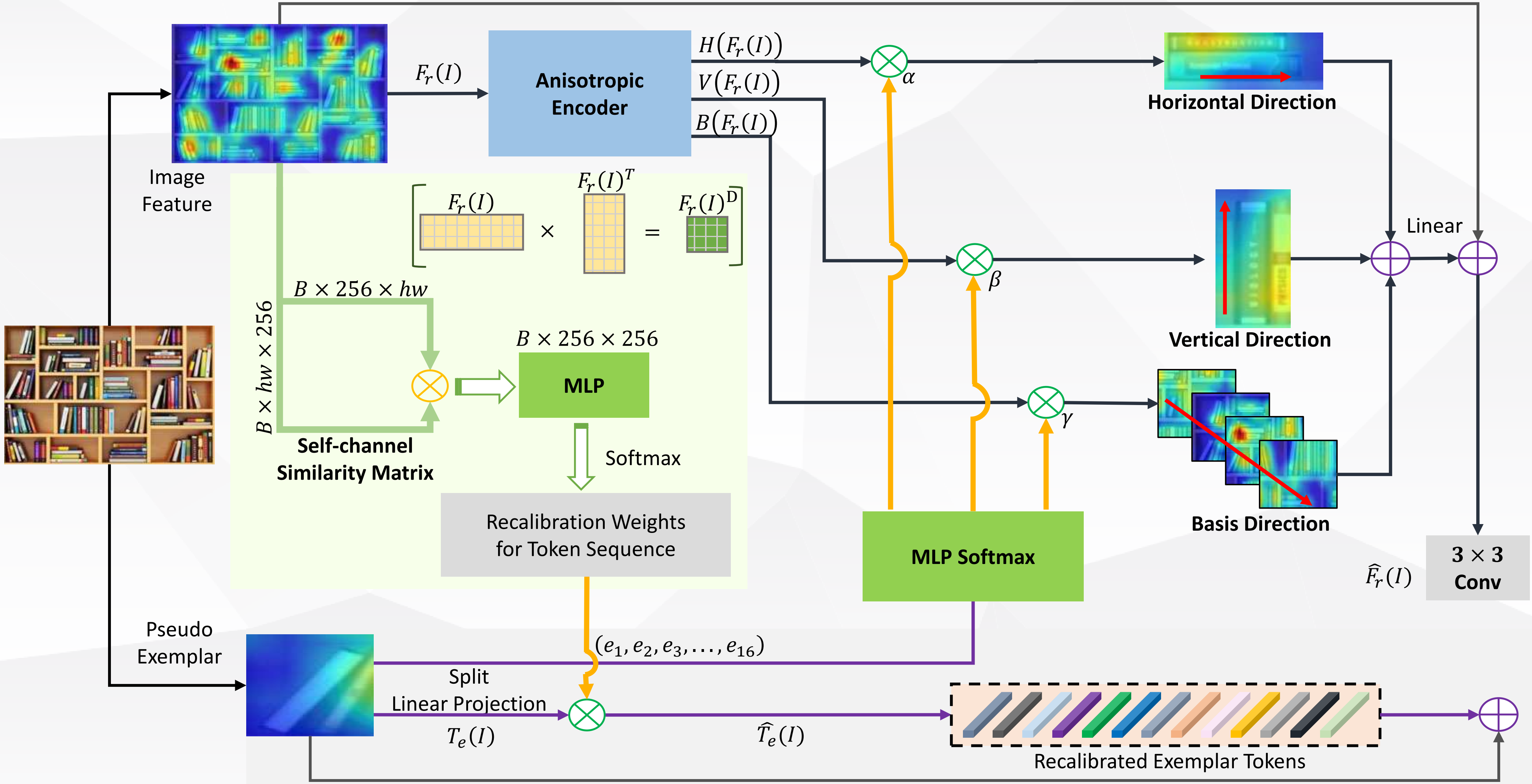
UNIVERSITY of GUELPH

Overall Architecture for General Counter

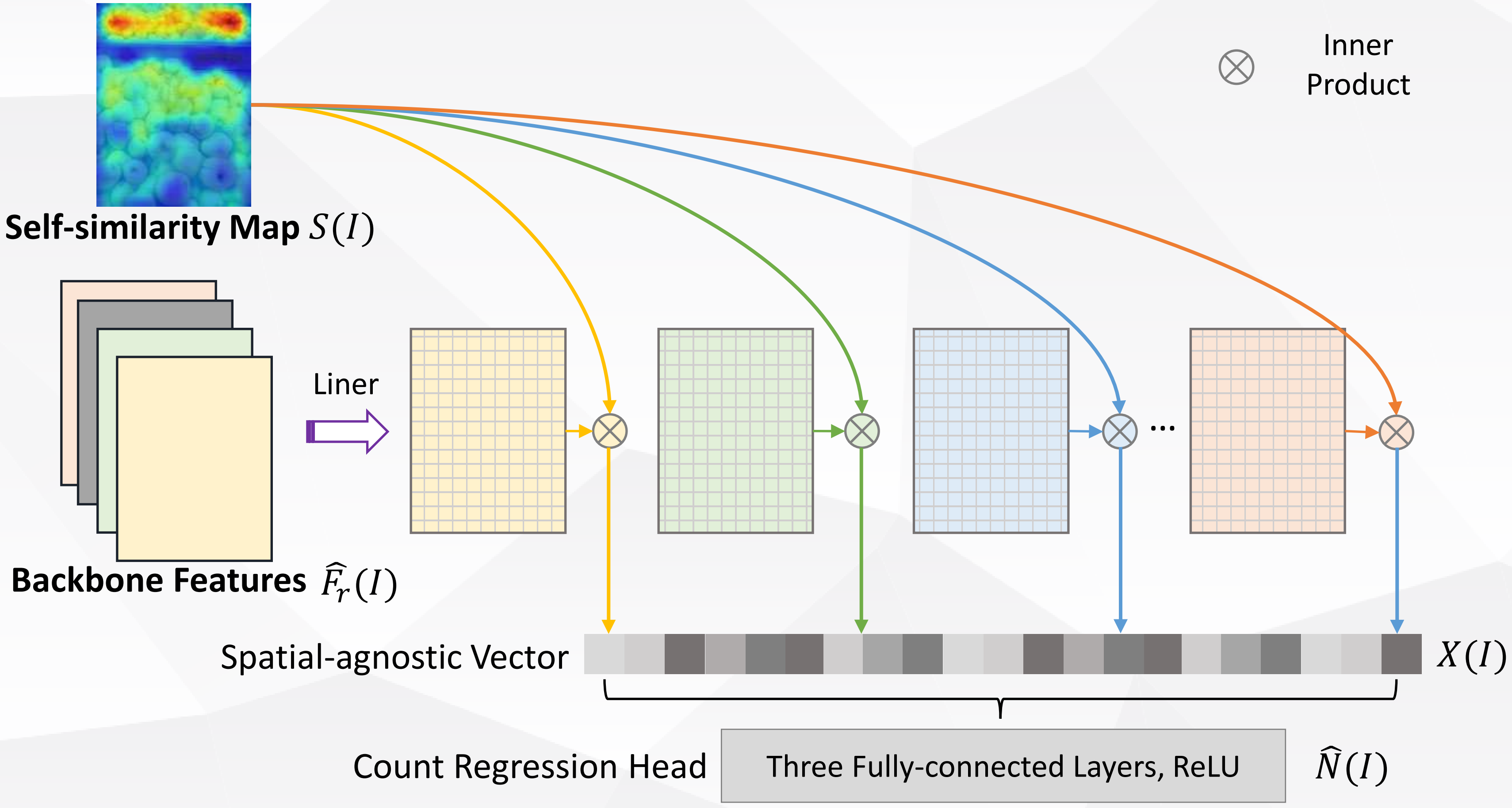
- Introduction
- Path Planning for Offsite Modeling
- Path Planning for Online Exploration
- Weakly-Supervised Counter
- **General Object Counter**
- Conclusion



Details of Self-similarity Learning



Weakly-supervised Location-aware Counter



UNIVERSITY of GUELPH

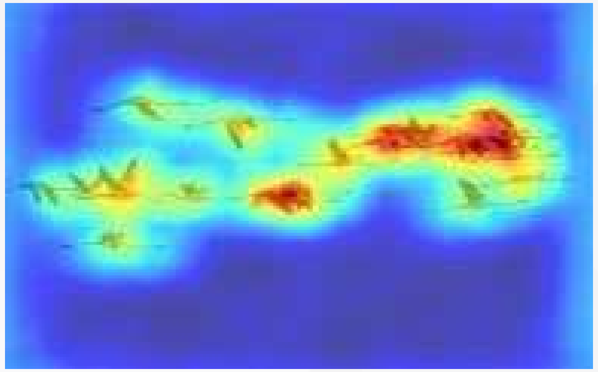
Quantitative Evaluation

Frameworks	Exemplar	Location	**MAE** for *Val*	**MSE** for *Val*	**MAE** for *Test*	**MSE** for *Test*
GMN [4]	✓	✓	29.66	89.81	26.52	124.57
FamNet [4]	✓	✓	24.32	70.94	22.56	101.54
FamNet+ [4]	✓	✓	23.75	69.07	22.08	**99.54**
CFOCNet [5]	✓	✓	21.19	**61.41**	22.10	112.71
BMNet [7]	✓	✓	**19.06**	67.95	**16.71**	103.31
RepRPN-Counter [54]	×	×	29.24	98.11	26.66	129.11
Baseline	×	×	23.14	77.30	21.88	112.37
GCNet+Exemplar (ours)	✓	×	19.61	66.22	17.86	106.98
GCNet (ours)	×	×	**19.50**	**63.13**	**17.83**	**102.89**

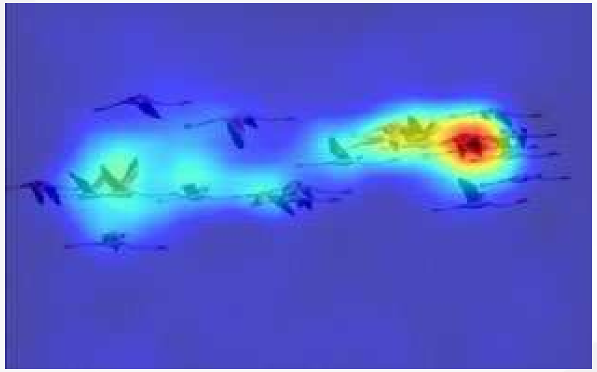
Results on General Objects

GT: 22

Ours: 21.45

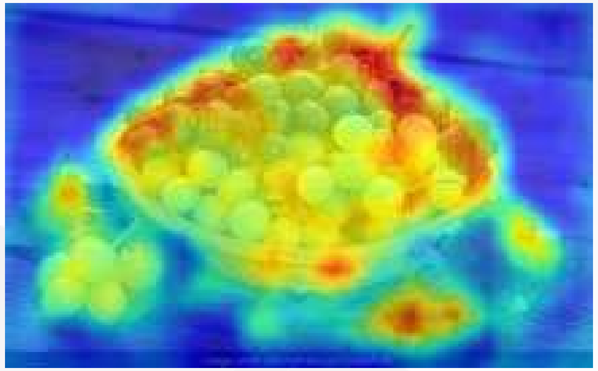


Baseline: 28.76



GT: 80

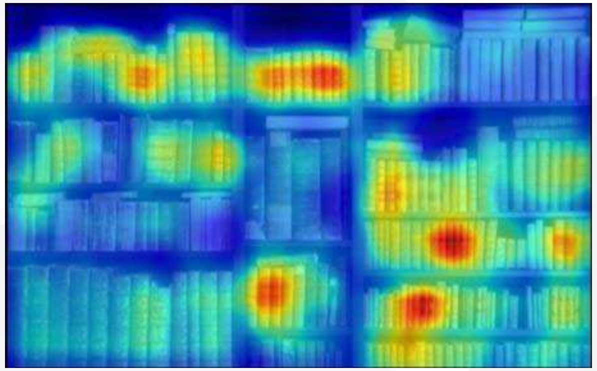
Ours: 79.49



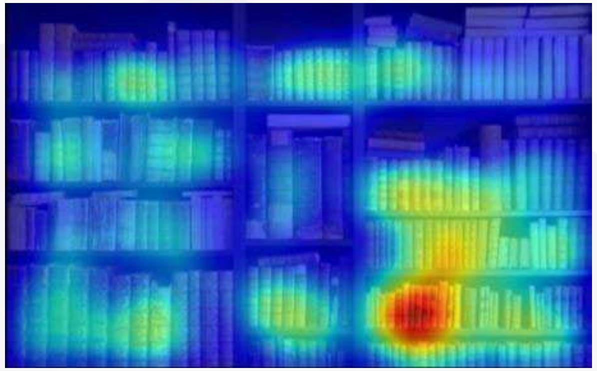
Baseline: 58.37

GT: 230

Ours: 232.35



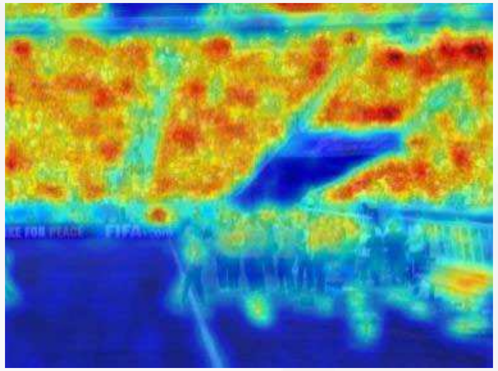
Baseline: 170.32



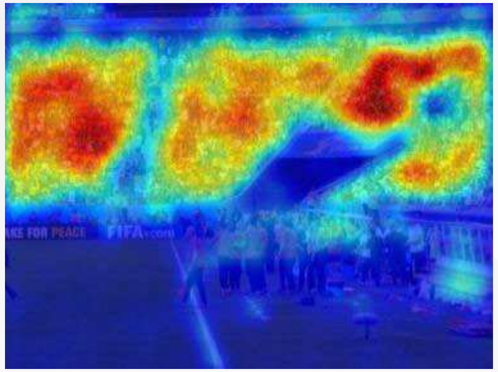
Results on Crowd Counting

GT: 1357

Ours: 1233.55

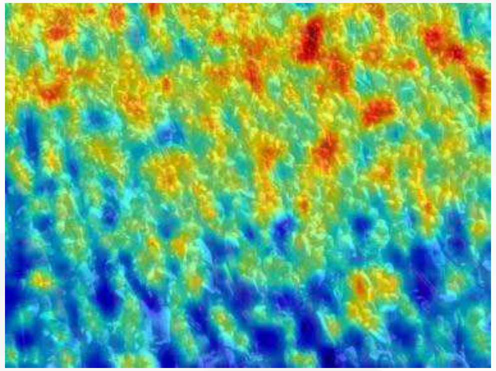


Baseline: 357.00

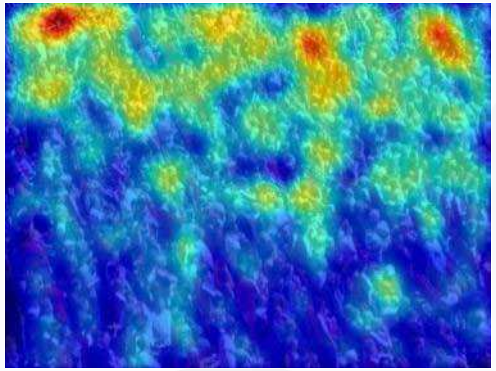


GT: 817

Ours: 818.97

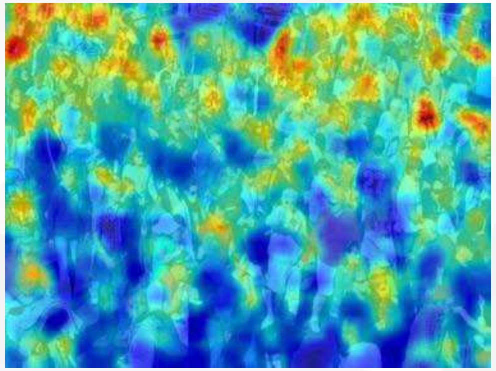


Baseline: 412.92



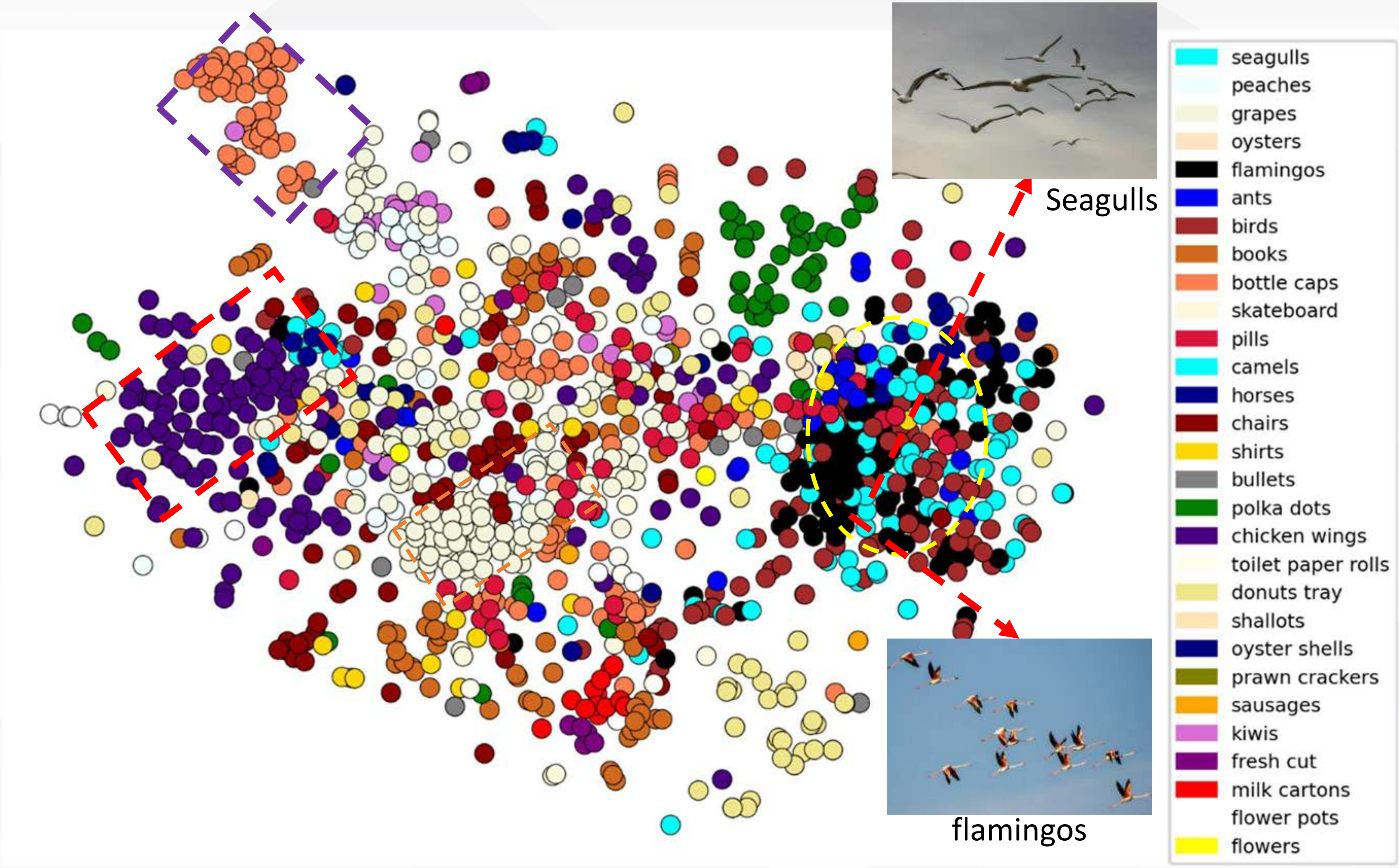
GT: 170

Ours: 168.65



Baseline: 116.10

Visualization of Intermediate Features

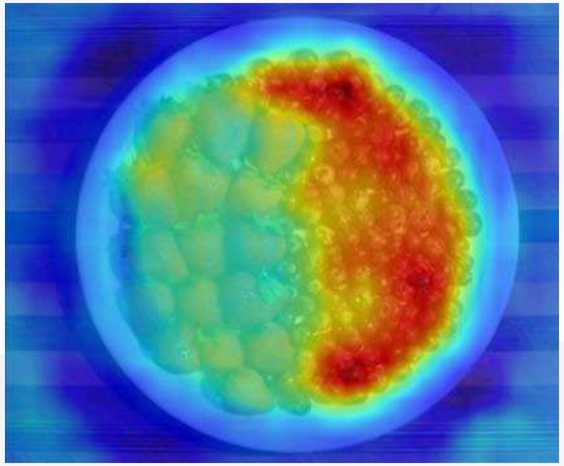


Failure Cases

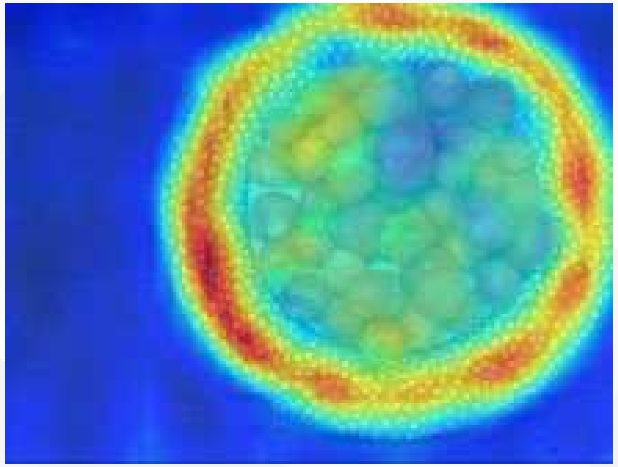
GT: 14

GT: 16

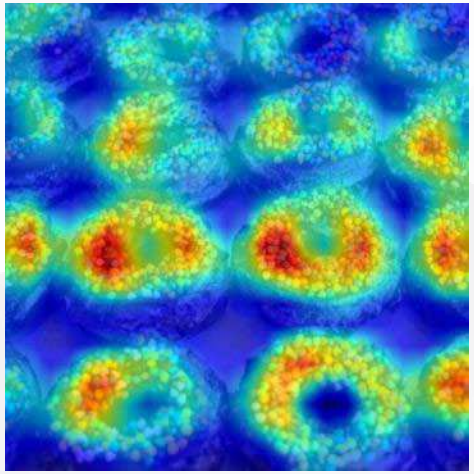
GT: 43



Ours: 63.88



Ours: 85.56



Ours: 137.43

Conclusions

- Two approaches are proposed for capturing large-scale scenes efficiently:
  - Optimize flight path to minimize the cost on scene capture and maximize the reconstruction quality.
    - *Use satellite images to generate proxy model*
  - Use explore-then-capture approach to design flight path without priori.
- Several strategies are explored to generalize object counting networks:
  - Generalization on location-agnostic counting
    - *With global MLP-based regressor, it is possible to train counters using only global count information.*
  - Generalization on class-agnostic counting
    - *Exploring self-similarity learning enables the models to adaptively* learn pseudo exemplar clues from inherent repetition patterns.

Future Work in Precision Agriculture

- Advanced Crop Health Monitoring:
  - Automated Growth Analysis: Utilize AI-driven 3D models to continuously monitor crop growth stages and detect anomalies in real-time.
- Enhanced Soil and Water Management:
  - 3D Soil Mapping: Develop detailed 3D models of soil layers to optimize planting strategies and design efficient irrigation systems.
- Precision Pest and Disease Control:
  - Real-time Pest Detection: Implement AI-driven object counting systems for the real-time detection and tracking of pests, enabling prompt interventions.
  - Disease Spread Modeling: Use 3D models to predict and visualize the spread of diseases, facilitating effective containment strategies.
- Optimized Resource Allocation:
  - Dynamic Resource Management: Create systems that utilize real-time data from 3D models and object counts to adjust the allocation of resources such as fertilizers and pesticides.
  - Livestock Tracking and Health Monitoring: Use 3D modeling and object counting to monitor livestock health and movement, improving overall farm management.

- This presentation explores cutting-edge research in 3D scene modeling and object counting, techniques that hold significant potential for precision agriculture.
- For large-scale scene reconstruction, we developed aerial path planning methods to efficient scene capture using UAVs. Our initial method features an adaptive path planning algorithm that utilizes just a 2D map and a satellite image for pre-planning. We generate an initial 2.5D model by analyzing the relationship between buildings and their shadows, then implement a Max-Min optimization strategy to reduce the number of viewpoints needed. Additionally, we introduce a real-time explore-and-reconstruct algorithm that requires no prior knowledge of the scene, using real-time 2D images to estimate 3D boundaries of buildings. This method facilitates efficient exploration and detailed observation of the scene, yielding high-quality 3D models more quickly and without prior data collection.
- On the object counting front, we designed advanced neural network for accurately counting objects. We first introduced a novel counter and a multi-granularity MLP regressor, which does not depend on location annotations but instead train solely with total counts. To address challenges related to limited samples and the lack of spatial cues, we developed a self-supervised technique called Split-Counting. Additionally, for counting various object categories without specific exemplars, we further proposed a zero-shot generalized counter that derives pseudo exemplar clues from repetitive patterns and captures spatial location hints through a self-similarity learning approach.

- 本次演讲探讨了3D场景建模和对象计数的前沿研究，这些技术对精确农业具有重要的潜力。
- 在大规模场景重建领域，我们优化了使用无人机(UAVs)进行场景捕捉的航迹规划方法。我们的初步方法采用了一种自适应航迹规划算法，该算法仅利用2D地图和卫星图像进行预规划。我们通过分析建筑物及其阴影的关系生成初步的2.5D模型，然后实施Max-Min优化策略以减少所需观察点的数量。此外，我们还引入了一种实时探索和重建算法，该算法不需要事先了解场景，使用实时2D图像来估计建筑物的3D边界。这种方法提高了场景探索和观察的效率，使得可以更快速地生成高质量的3D模型，且无需初步数据收集。
- 在对象计数方面，我们设计了用于精确计数对象的先进神经网络模型。我们首先引入了一个新颖的计数器和一个多粒度MLP回归器，这些模型仅依靠总计数进行训练，而无需位置注释。为了应对样本有限和空间线索缺失的挑战，我们开发了一种自监督技术，名为Split-Counting。此外，为了计数没有具体示例的各种对象类别，我们进一步提出了一种零样本泛化计数器，该计数器从重复模式中提取伪示例线索，并通过自相似学习策略捕获空间位置线索。