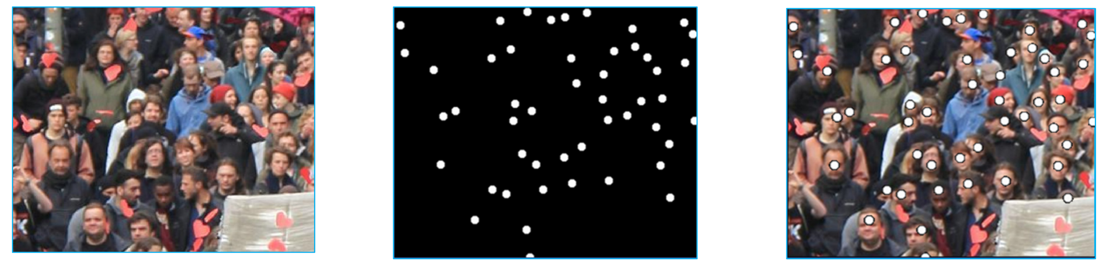
Localization in the Crowd with Topological Constraints

Shahira Abousamra, Minh Hoai, Dimitris Samaras, Chao Chen
Stony Brook University, USA

AAAI 2021

Introduction

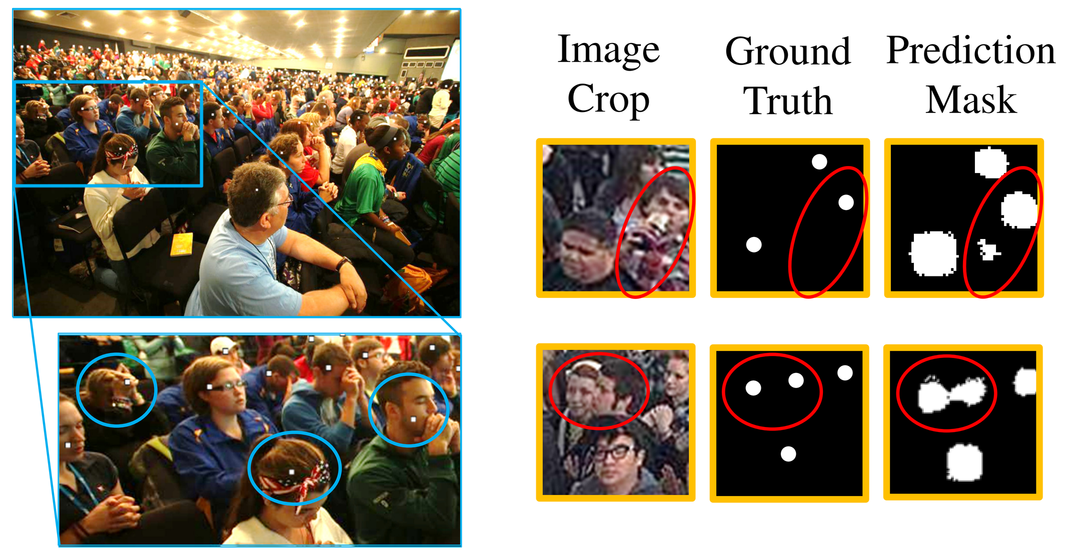
Crowd Localization Problem: Finding the location of each person in a crowded scene.
Ground truth: a single dot on each head.



Input Image · Ground Truth Dots · Input Image w/ Ground Truth Dots Overlay

Challenges in Crowd Localization

1. Perspective, occlusion, and cluttering.
2. The features of dots are not specific.
3. Difficult to prevent spatial semantic errors:

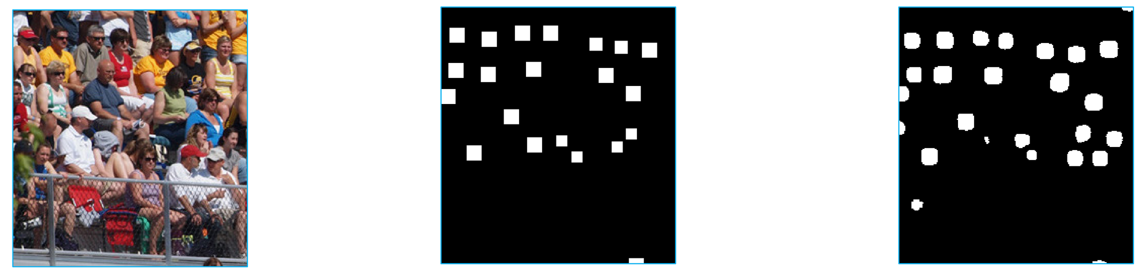


Contributions:

1. Overcome these challenges by introducing topological constraints in the training phase.
2. Propose persistence loss to enforce topological constraints.
3. Achieve high quality localization that is useful for crowd counting and spatial analysis.

Method: TopoCount

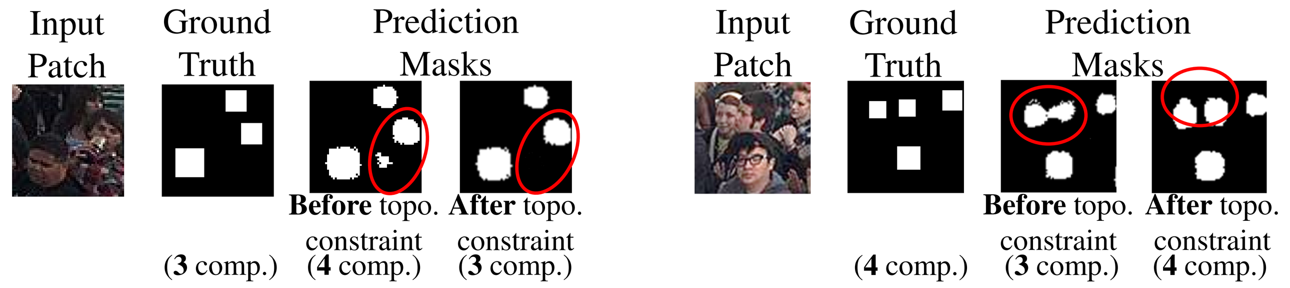
- Formulate crowd localization as a structured prediction problem.
- Each component in the binary prediction represents one dot.



Input Image · Ground Truth Dot Map · Prediction Mask

Topological Constraint for Crowd Localization

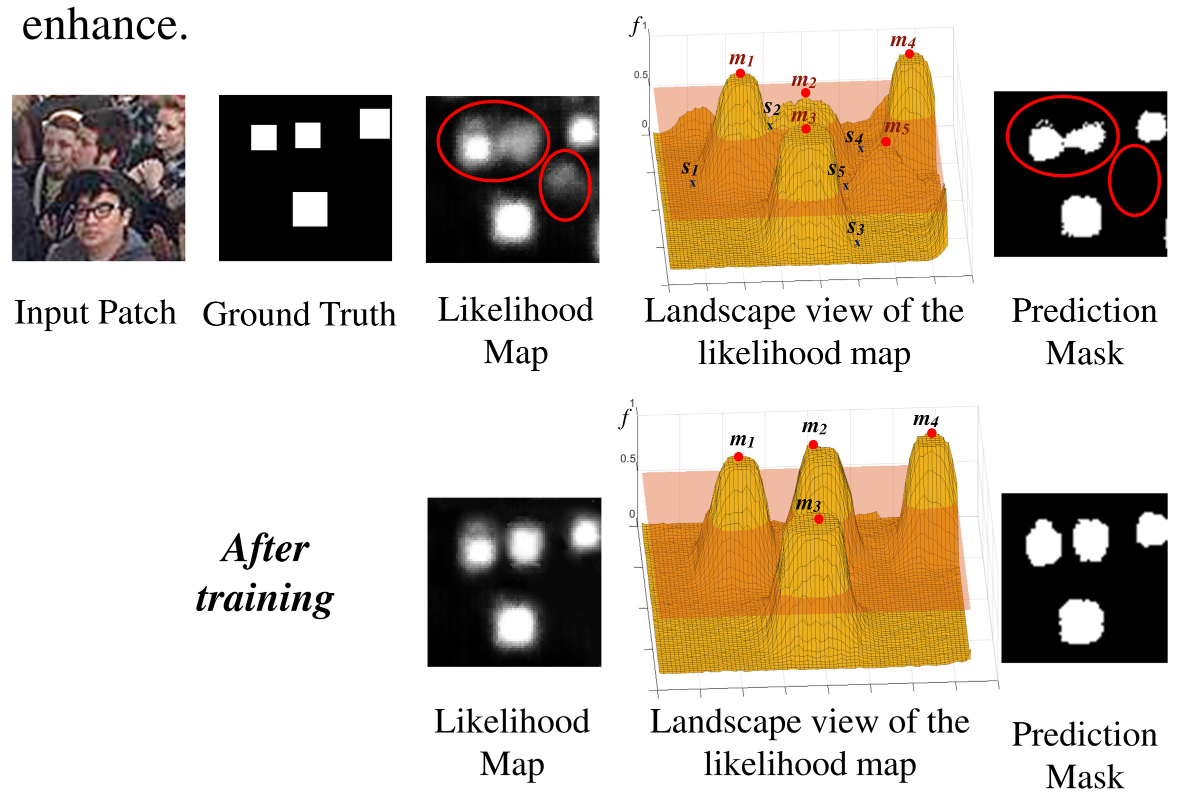
Within any local patch, the number of connected components in the prediction equals to the number of ground truth dots.



Topological errors ↔ Semantic errors

Persistence Loss $\mathcal{L}_{Pers}$

- To enforce topological constraints.
- Consider likelihood map as a terrain function *f*.
- Each mode of *f* corresponds to a possible dot prediction.
- Persistence Loss captures all modes and chooses to suppress or enhance.



Given a patch δ, with c ground truth dots:

Persistence loss reinforces the total ***saliency*** of the top ***c*** modes of *f* and suppresses the saliency of the rest.

Saliency/Persistence of a mode m_i = $f(m_i) - f(s_i)$

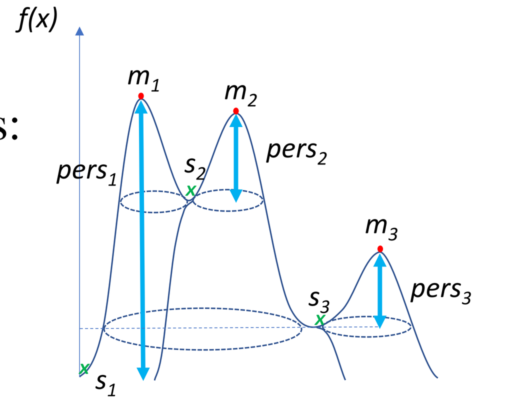
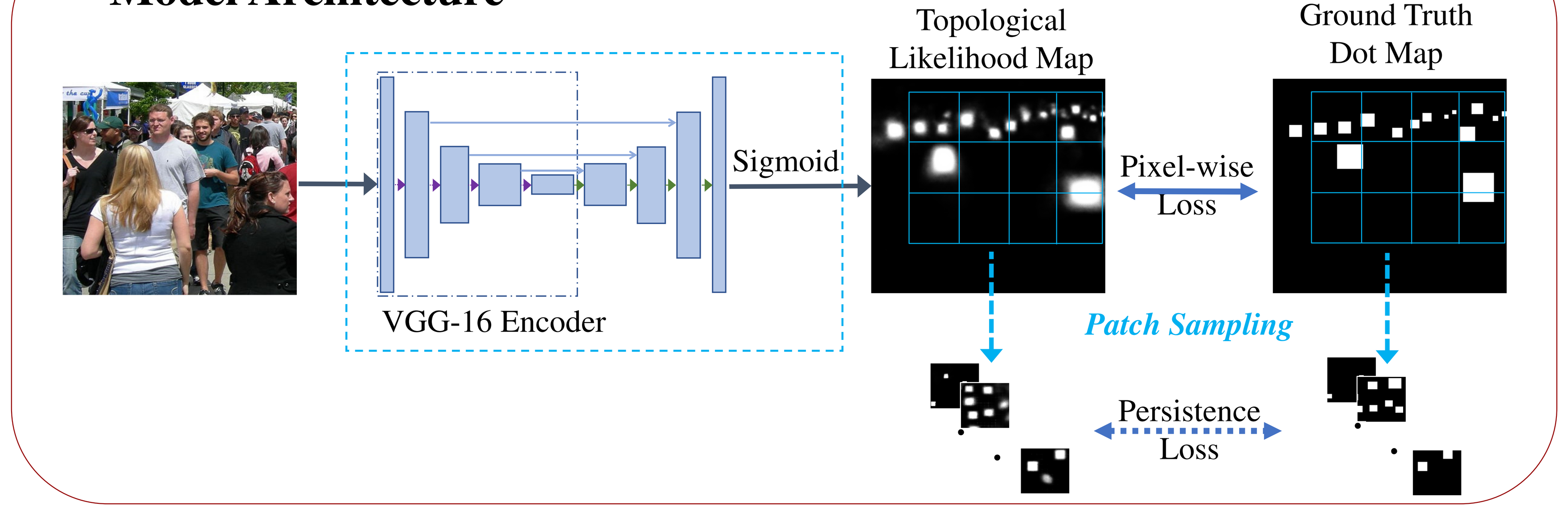


Illustration of Persistent Homology

$$\mathcal{L}_{Pers}(f,\delta) = -\sum_{m\in\mathcal{M}_c}\mathrm{Pers}(m) + \sum_{m\in\bar{\mathcal{M}}_c}\mathrm{Pers}(m)$$

Training Loss $\mathcal{L} = \mathcal{L}_{DICE} + \lambda_{pers}\mathcal{L}_{Pers}$

Model Architecture



Evaluation

1. Localized Counting

Grid Average Mean Absolute Error (**G(L)**): divide the image into 4^L non-overlapping cells and computes the mean MAE over all grid cells.

Method	ShanghaiTech A			ShanghaiTech B			UCF QNRF		
	G(1)	G(2)	G(3)	G(1)	G(2)	G(3)	G(1)	G(2)	G(3)
CSRNet (Li et al. 2018)	76	113	149	13	21	29	157	187	219
Bayesian (Ma et al 2019)	75	90	130	**10**	**14**	23	**100**	**117**	150
LSC-CNN (Babu Sam et al. 2019)	70	95	137	**10**	17	27	126	160	206
TopoCount	**69**	**81**	**104**	**10**	**14**	**20**	102	119	**148**

2. Ablation Study: Loss function

Compare training with and without persistence loss:

- Dice loss only better than binary cross entropy (BCE) loss only.
- Dice loss + Persistence loss give lowest error.

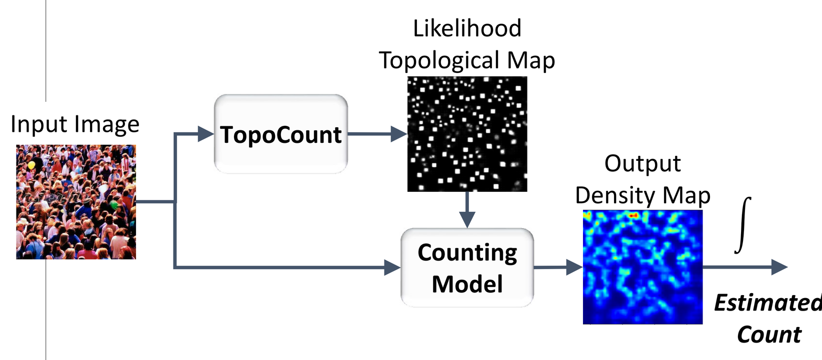
Loss	G(3)
BCE Loss	122
DICE Loss	114
DICE Loss + Pers. Loss	**104**

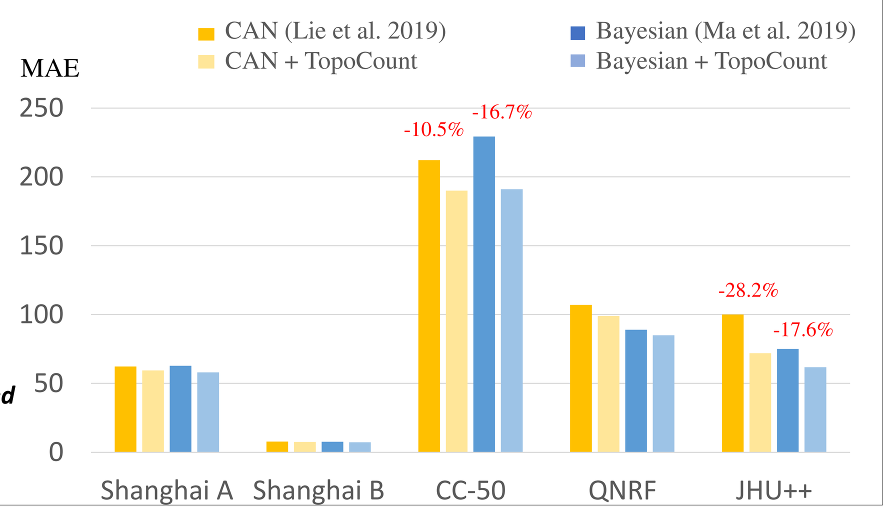
3. Dot Matching Accuracy

Compare precision, recall, and F-score on NWPU localization challenge.

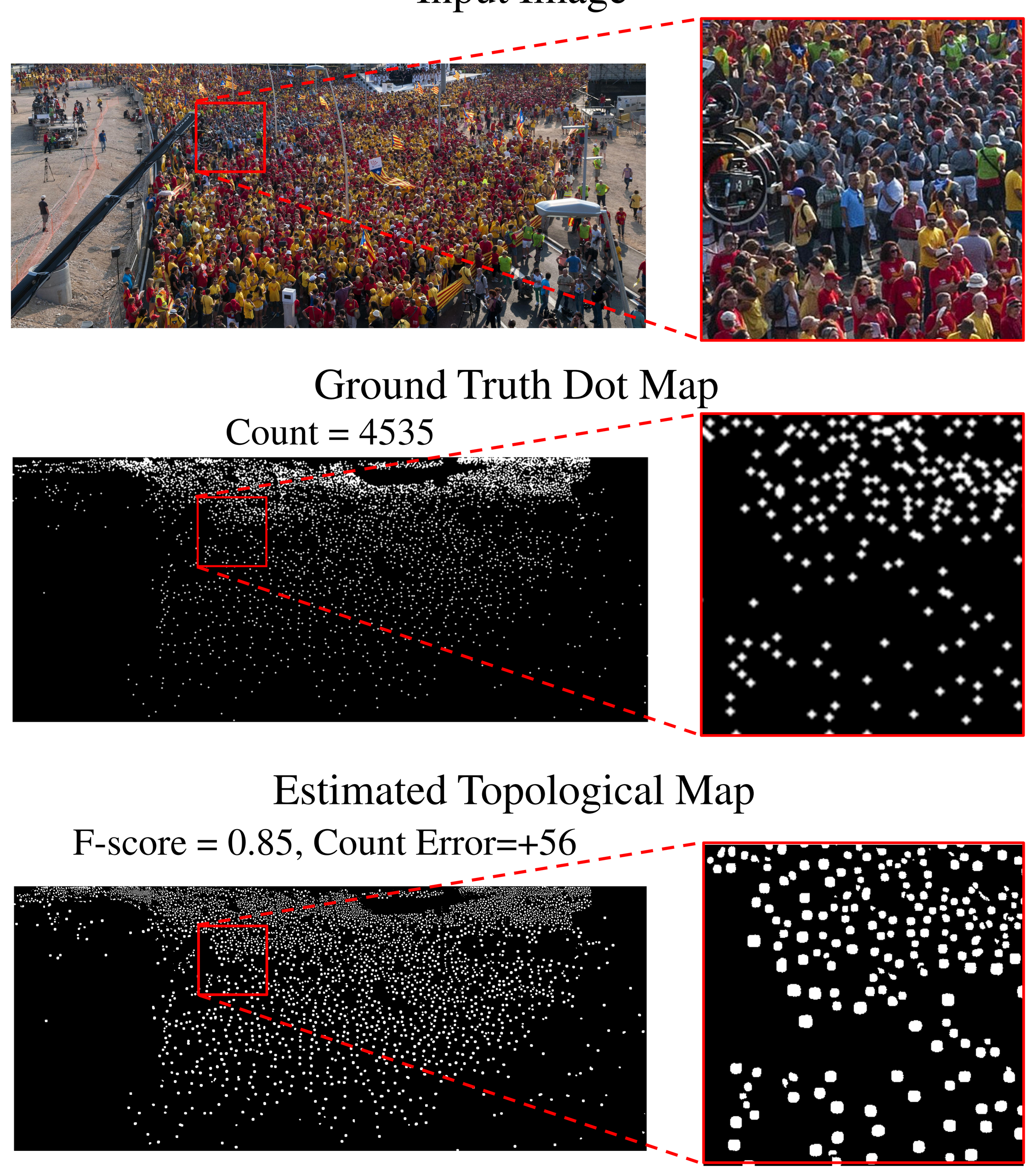
Method	F1 / Pre. / Rec. (%)
Faster RCNN (Ren et al. 2015)	6.7 / **95.8** / 3.5
TinyFaces (Hu et al. 2017)	56.7 / 52.9 / 61.1
VGG+GPR (Gao et al. 2019)	52.5 / 55.8 / 49.6
RAZ Loc (Liu et al. 2019)	59.8 / 66.6 / 54.3
TopoCount	**69.1** / 69.5 / **68.7**

4. Integration with density-map counting:

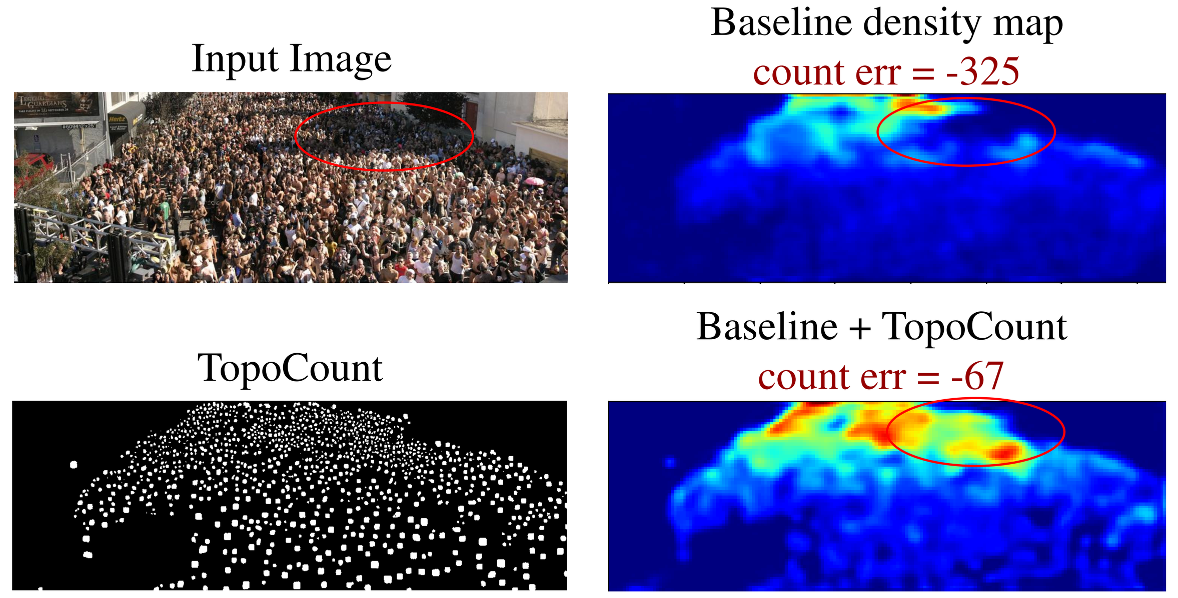




Qualitative Results



Integration with Density-map Counting



code: https://github.com/TopoXLab/TopoCount
email: shahira.abousamra@stonybrook.edu