Tree Instance Segmentation with Temporal Contour Graph

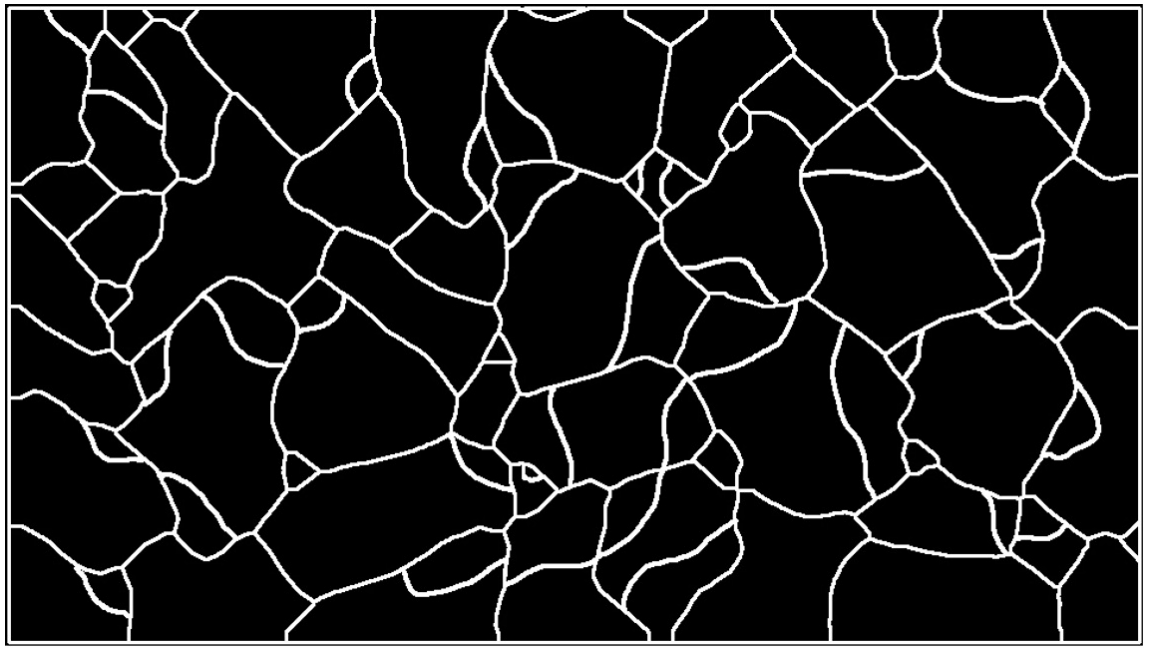
Adnan Firoze, Cameron Wingren,
Raymond A. Yeh, Bedrich Benes, Daniel Aliaga

Purdue University, West Lafayette, Indiana

[Poster Number: 210, Paper Tag: TUE-AM-210]

We ask how many trees are in a densely packed scene like this and can we segment them?

Our work solves this with high accuracy



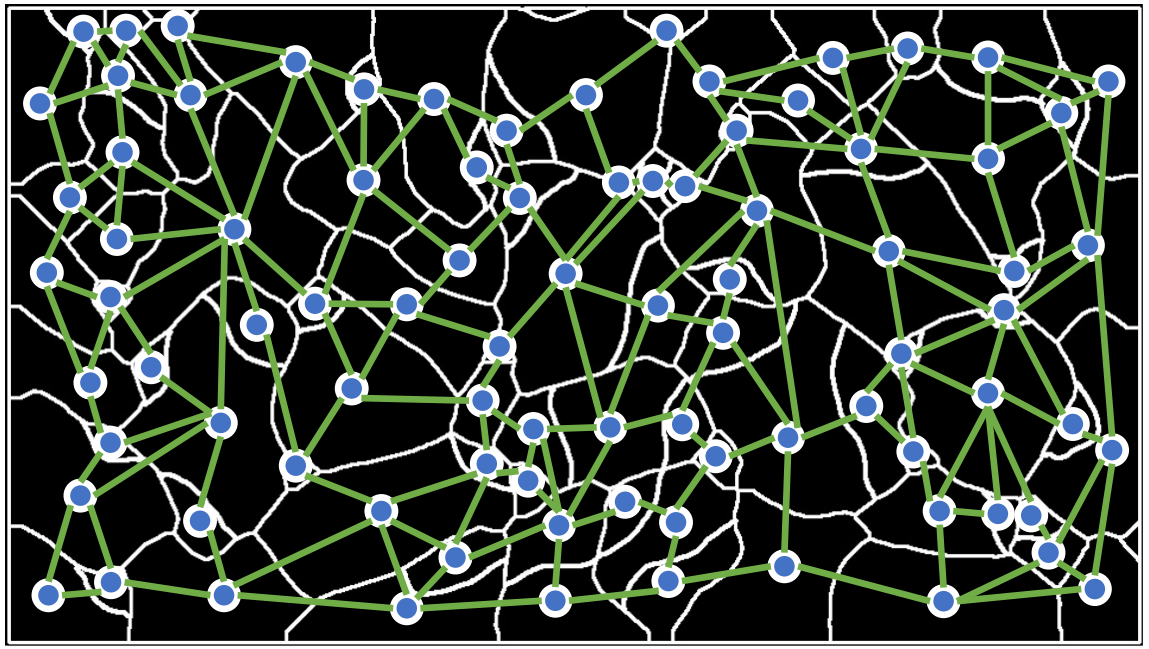
Input frames

Deep edge detection using PIDINet[1]

[1] Su, Z. et al. “Pixel Difference Networks for Efficient Edge Detection.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 5097-5107.

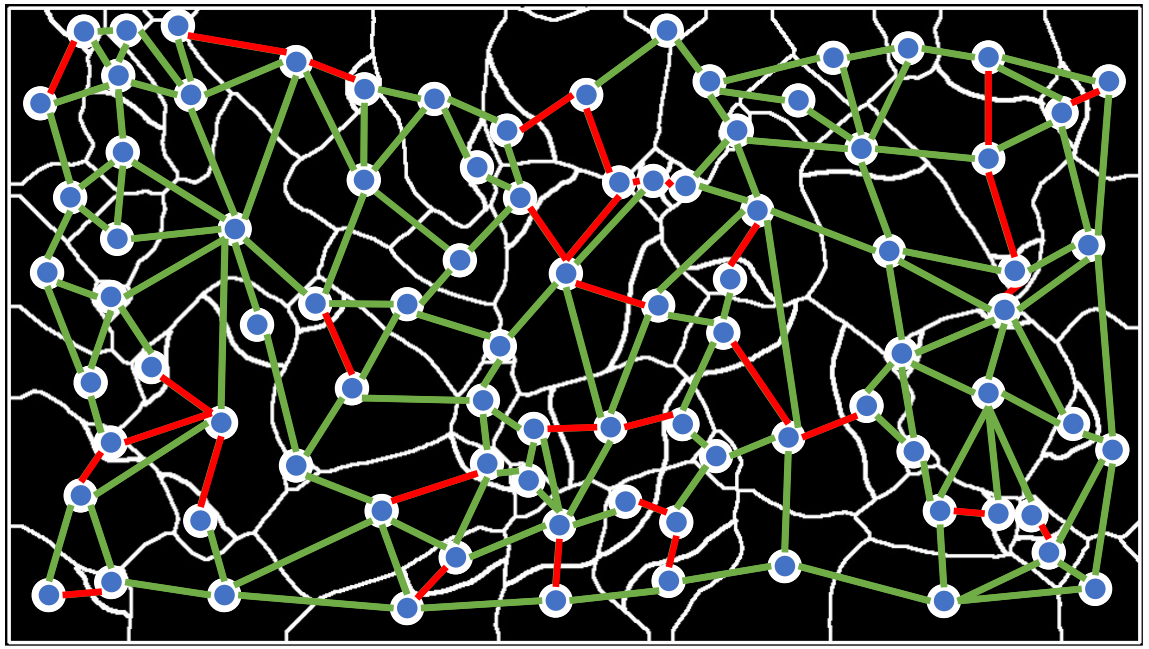
Our work solves this with high accuracy

Input frames



Converting edge map to a graph

Our work solves this with high accuracy

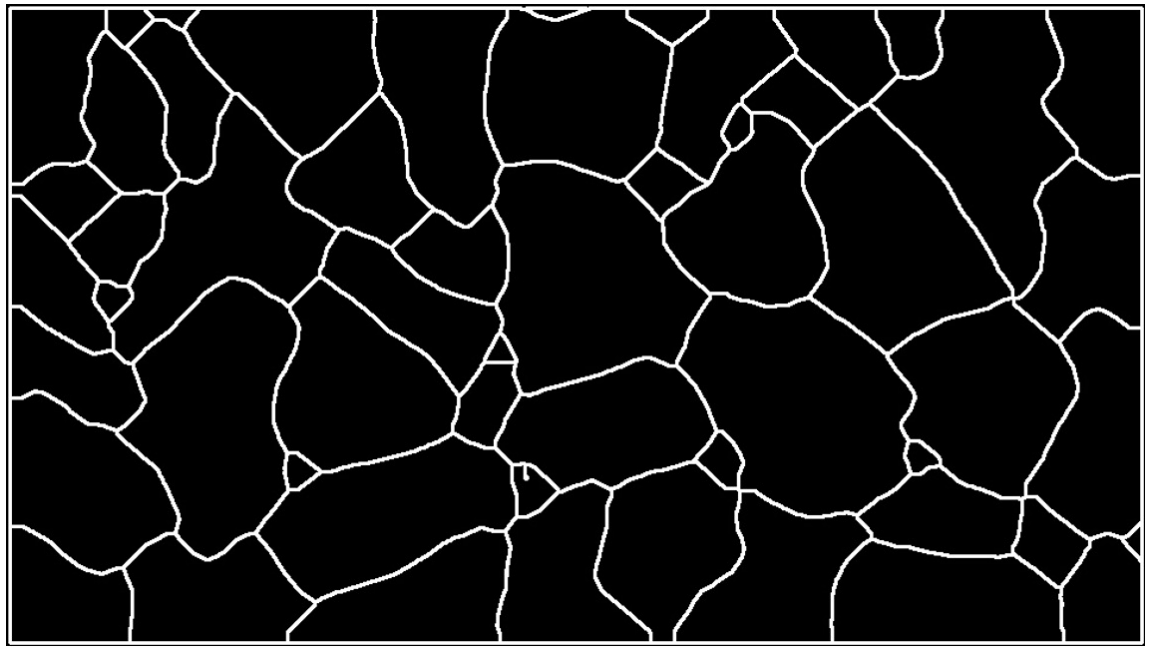


Input frames

Classify edges to merge nodes using motion (optical flow)

Our work solves this with high accuracy

Input frames



Refined mask generated

Our work solves this with high accuracy

Input frames

Final output of tree instances

We validated using three real-world forests

Olympic National Forest, WA

We validated using three real-world forests

Olympic National Forest, WA

AP = 70.1

We validated using three real-world forests

Martell Forest, IN

We validated using three real-world forests

Martell Forest, IN

AP = 74.5

We validated using three real-world forests

Ridge State Forest, KY

We validated using three real-world forests

Ridge State Forest, KY

AP = 69.8

Contributions

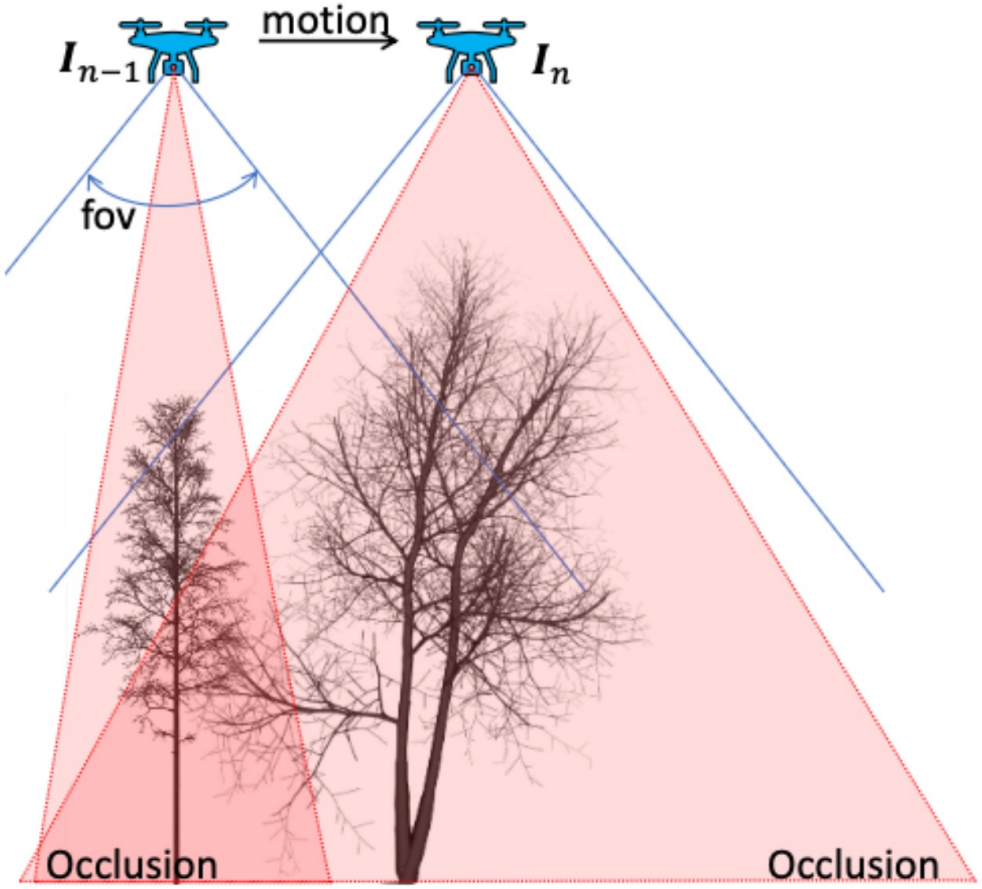
- **Instance Segmentation**
  - of dense and tightly packed trees in forests from overhead UAV frames
- **Counting**
  - tree crowns that are highly similar, prone to occlusion, self-similar, and overlapping
- **Dataset Creation**
  - the largest UAV video dataset of forests to date (to our knowledge)
  - linked for all in the paper

Roadmap

- Overview
- Observations
- Pipeline
- Quantitative Results
- Qualitative Results
- Future Works

Observations

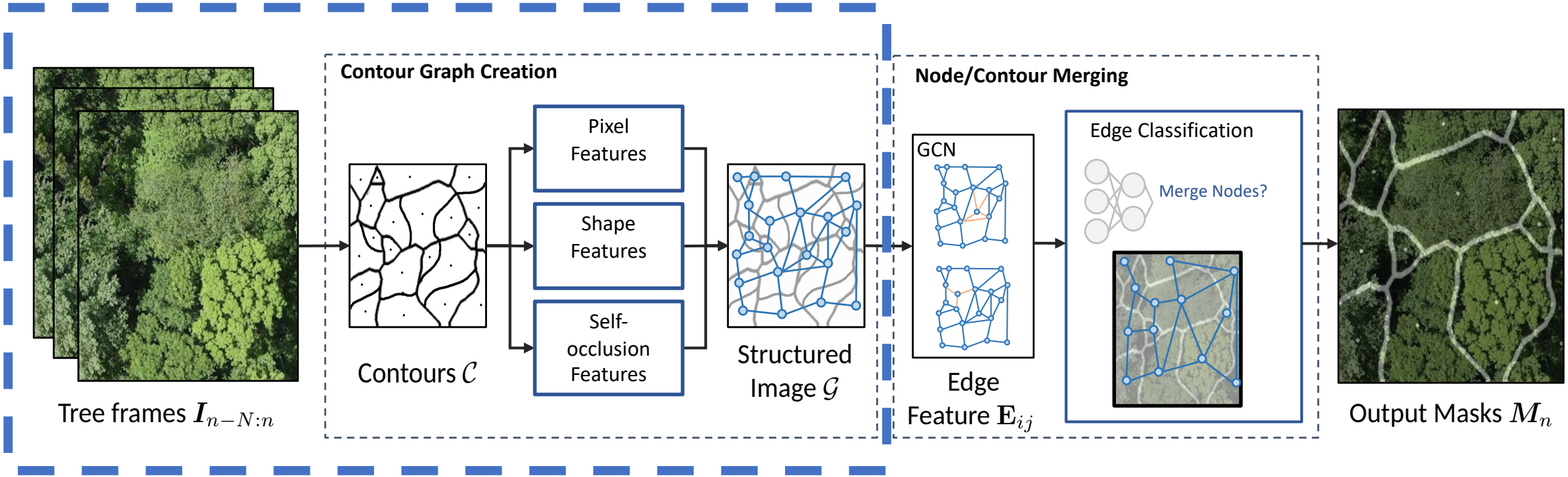
- In dense forests, two adjacent tree crowns are:
  - tightly packed.
  - highly similar, and
  - often occluding.
- Making it a hard problem
- Using temporal features capturing motion leads to higher instance segmentation accuracy.



Roadmap

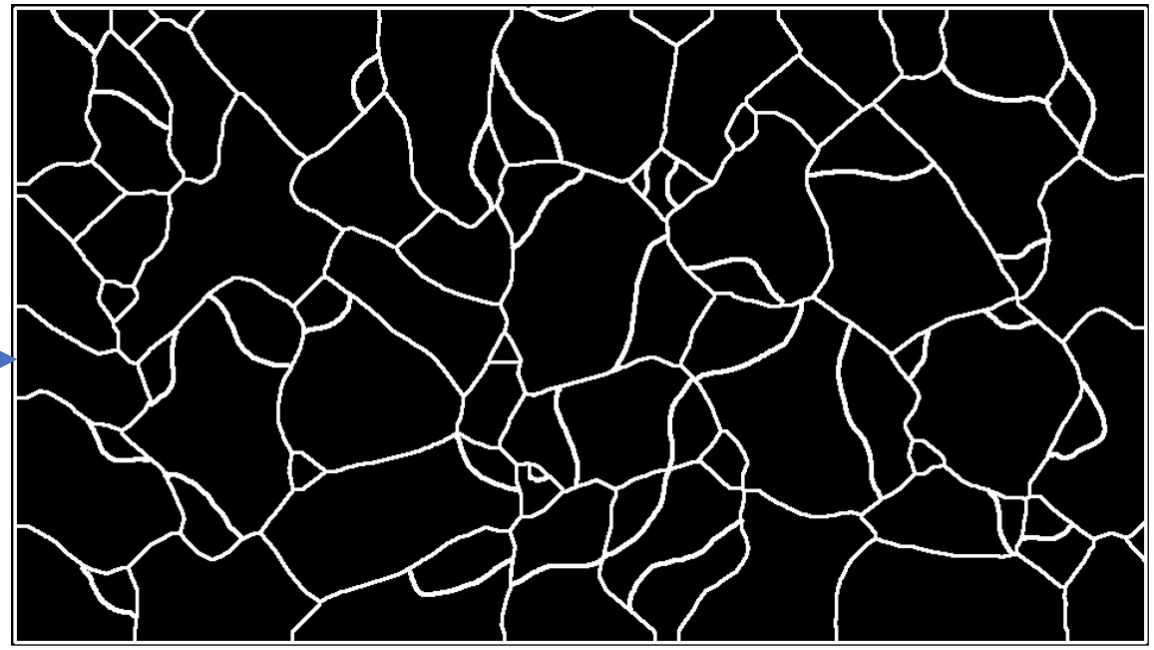
- Overview
- Observations
- Pipeline
- Quantitative Results
- Qualitative Results
- Future Works

Pipeline


Contour Graph Creation
Node/Contour Merging
Tree frames $I_{n-N:n}$
Contours $\mathcal{C}$
Pixel Features
Shape Features
Self-occlusion Features
Structured Image $\mathcal{G}$
GCN
Edge Feature $\mathbf{E}_{ij}$
Edge Classification
Merge Nodes?
Output Masks M_n

Contour Graph/Structured Image Creation

Tree Frames

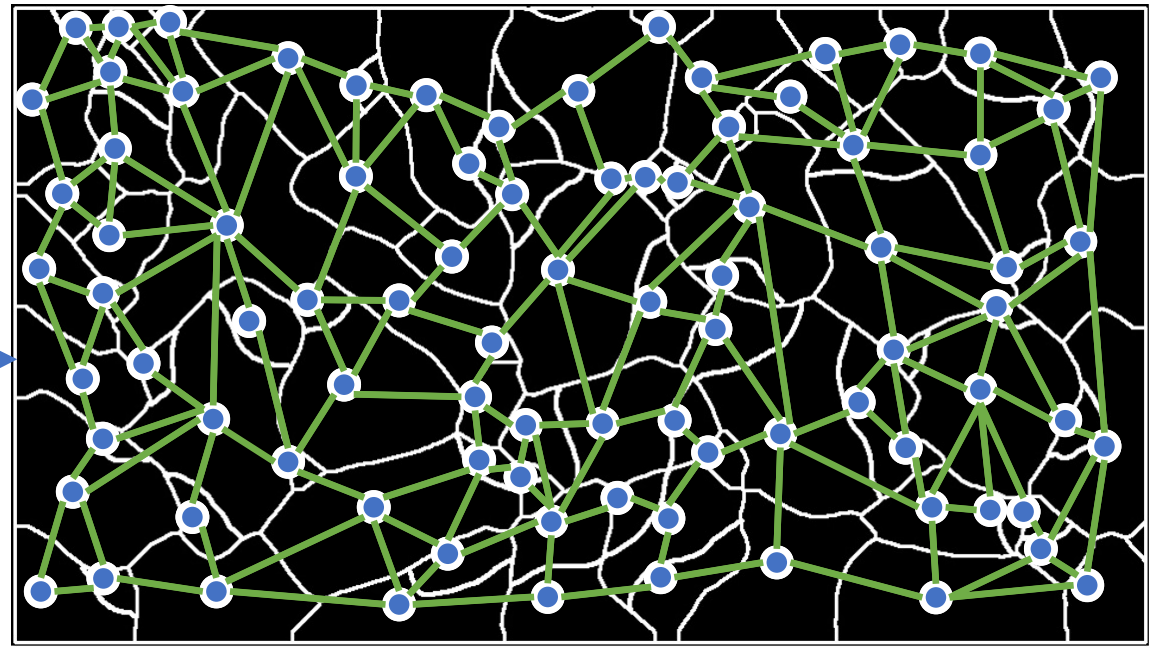


Deep edge map with overestimation

For edge detection:
Su, Z. et al. “Pixel Difference Networks for Efficient Edge Detection.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 5097-5107.

Contour Graph/Structured Image Creation

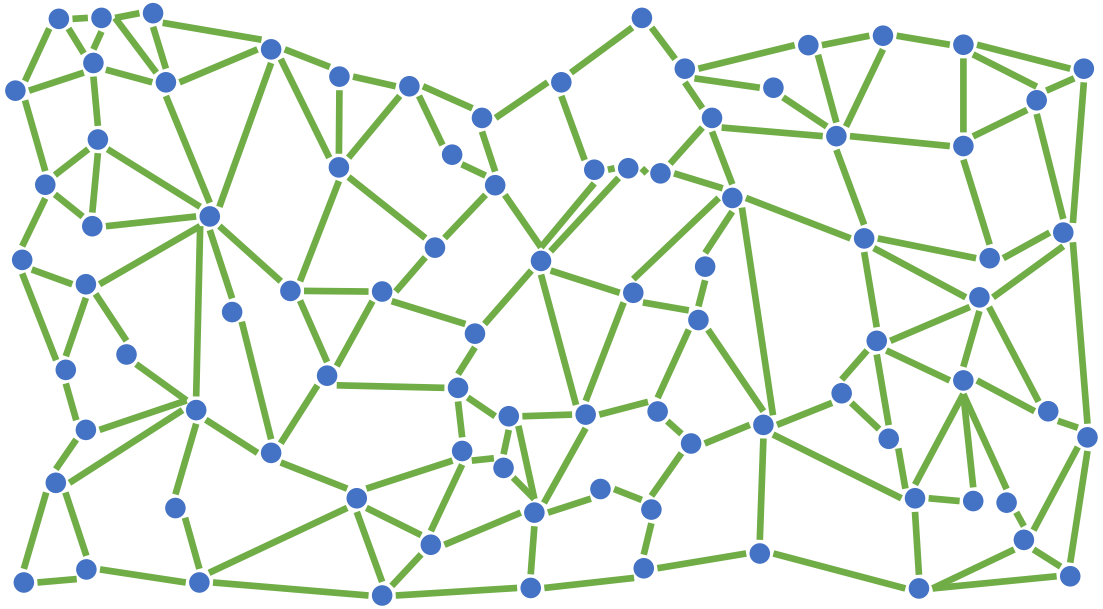
Tree Frames



Conversion to graph

Contour Graph/Structured Image Creation

Tree Frames

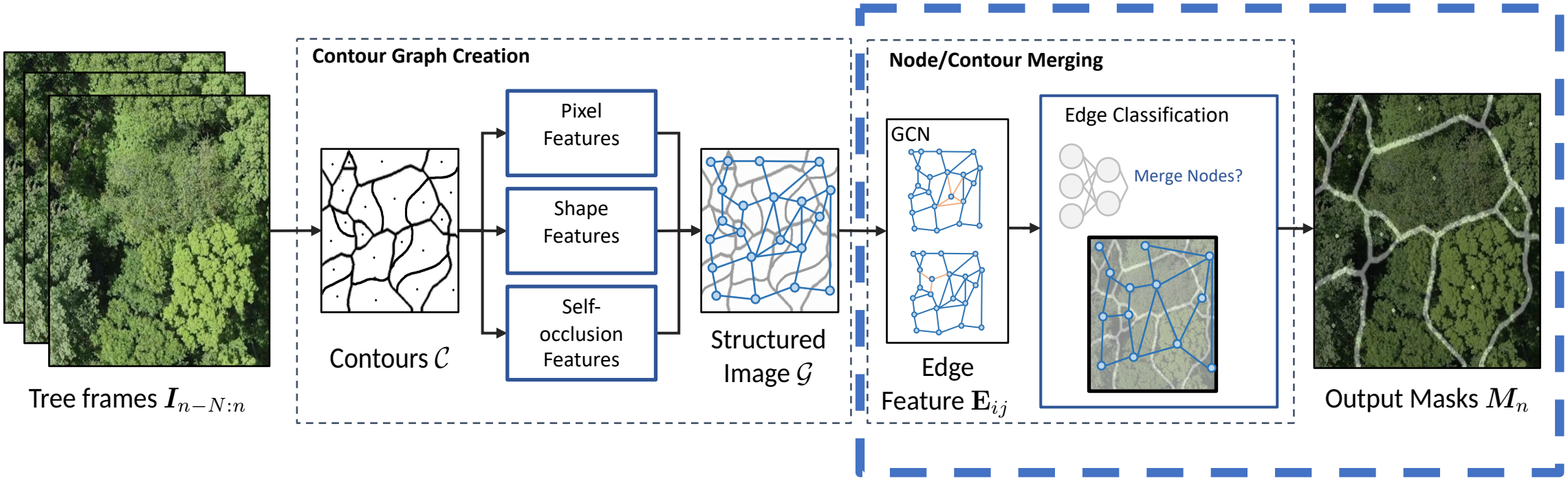


(Temporal) Contour Graph

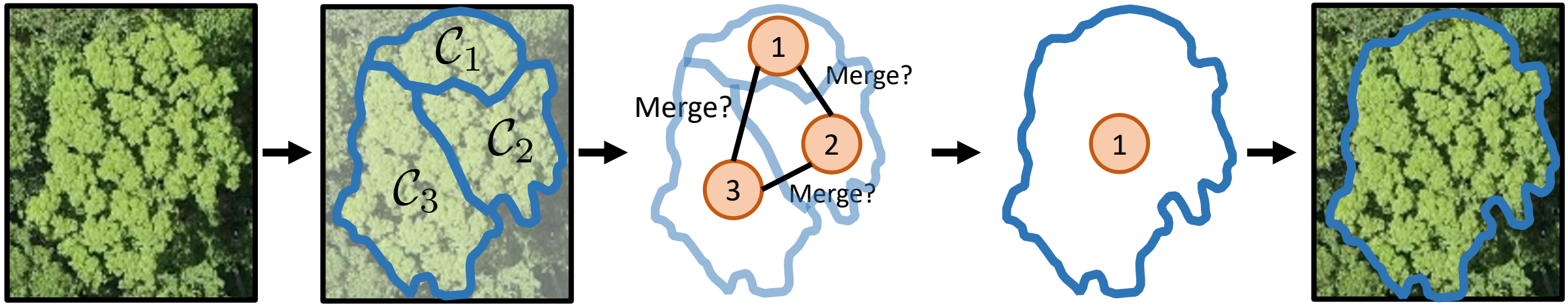
Node Features

Pixel Features: RGB patch features, LPIPS similarity to neighboring nodes
Shape Features: Area, Extent, Aspect Ratio, Solidity
Self-occlusion Features: Temporal features (Optical Flows)

Pipeline

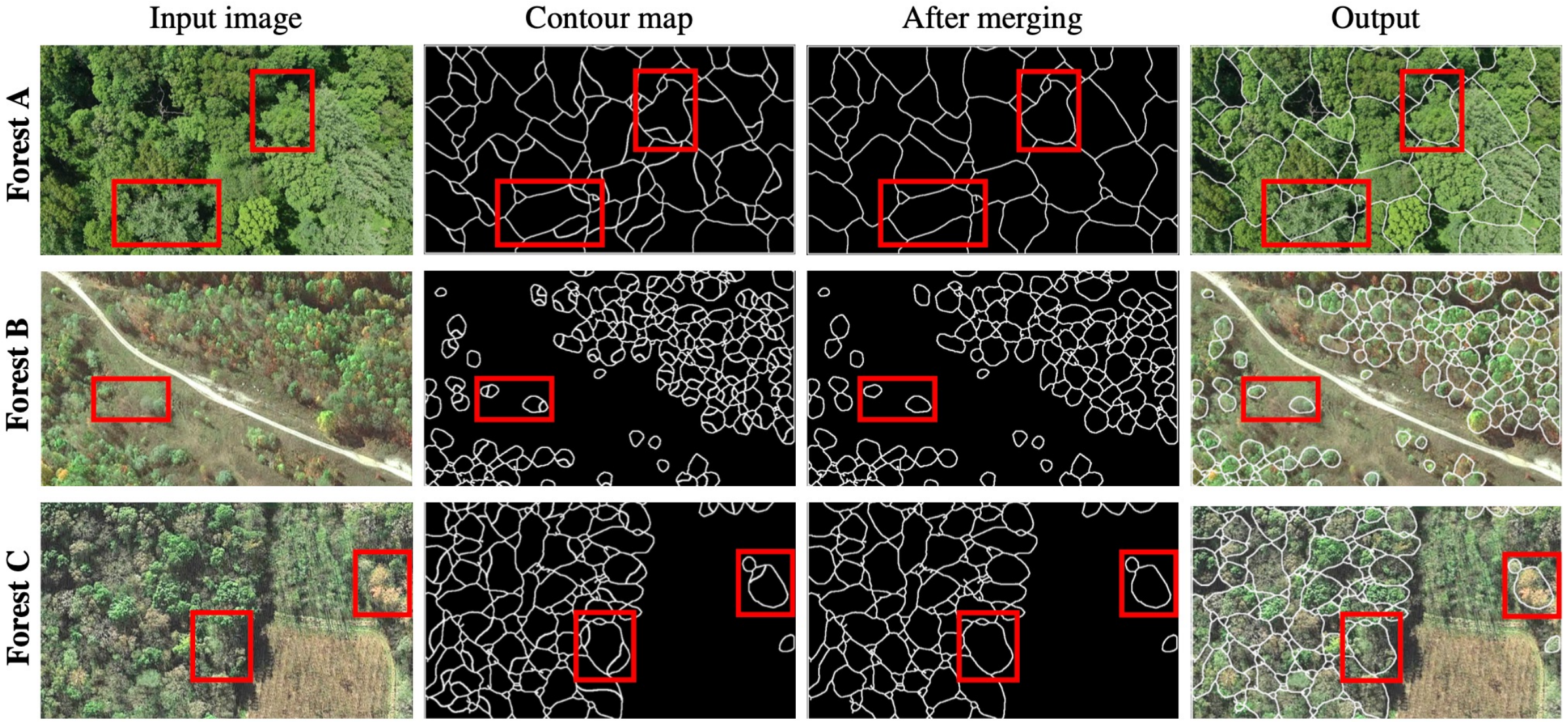

Contour Graph Creation
Pixel Features
Shape Features
Self-occlusion Features
Contours $\mathcal{C}$
Structured Image $\mathcal{G}$
Node/Contour Merging
GCN
Edge Feature $\mathbf{E}_{ij}$
Edge Classification
Merge Nodes?
Tree frames $\boldsymbol{I}_{n-N:n}$
Output Masks $\boldsymbol{M}_n$

Node/Contour Merging



Input Image $\boldsymbol{I}_n$ | Contours $\mathcal{C}$ | Edge Classification | Merged Node | Output Mask $\boldsymbol{M}_t$

Pipeline using Visual Examples (Data Flow)



For edge detection:
Su, Z. et al. “Pixel Difference Networks for Efficient Edge Detection.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 5097-5107.

Roadmap

- Overview
- Observations
- Pipeline
- Quantitative Results
- Qualitative Results
- Future Works

Quantitative Results

Instance Segmentation

Methods	Synthetic Forest			Forest A			Forest B			Forest C		
	AP ↑	AP_{50} ↑	AP_{70} ↑	AP ↑	AP_{50} ↑	AP_{70} ↑	AP ↑	AP_{50} ↑	AP_{70} ↑	AP ↑	AP_{50} ↑	AP_{70} ↑
Mask-RCNN (ResNet)	27.1	53.6	50.1	33.4	57.3	54.1	39.2	58.8	56.1	35.1	57.9	58.4
Mask-RCNN (Swin-T)	59.8	70.2	68.3	64.6	74.1	70.5	69.5	**77.3**	**72.4**	63.2	72.6	70.3
TraDeS	61.1	55.2	64.4	58.1	71.3	66.8	63.7	73.9	70.5	59.6	70..4	64.1
BoundaryFormer	56.3	65.1	57.5	60.9	72.9	66.2	64.1	73.4	69.2	58.9	71.2	61.8
Ours	**74.6**	**73.1**	**69.5**	**74.5**	**81.6**	**72.8**	**69.8**	76.2	71.5	**70.1**	**75.4**	**72.5**

Tree Count

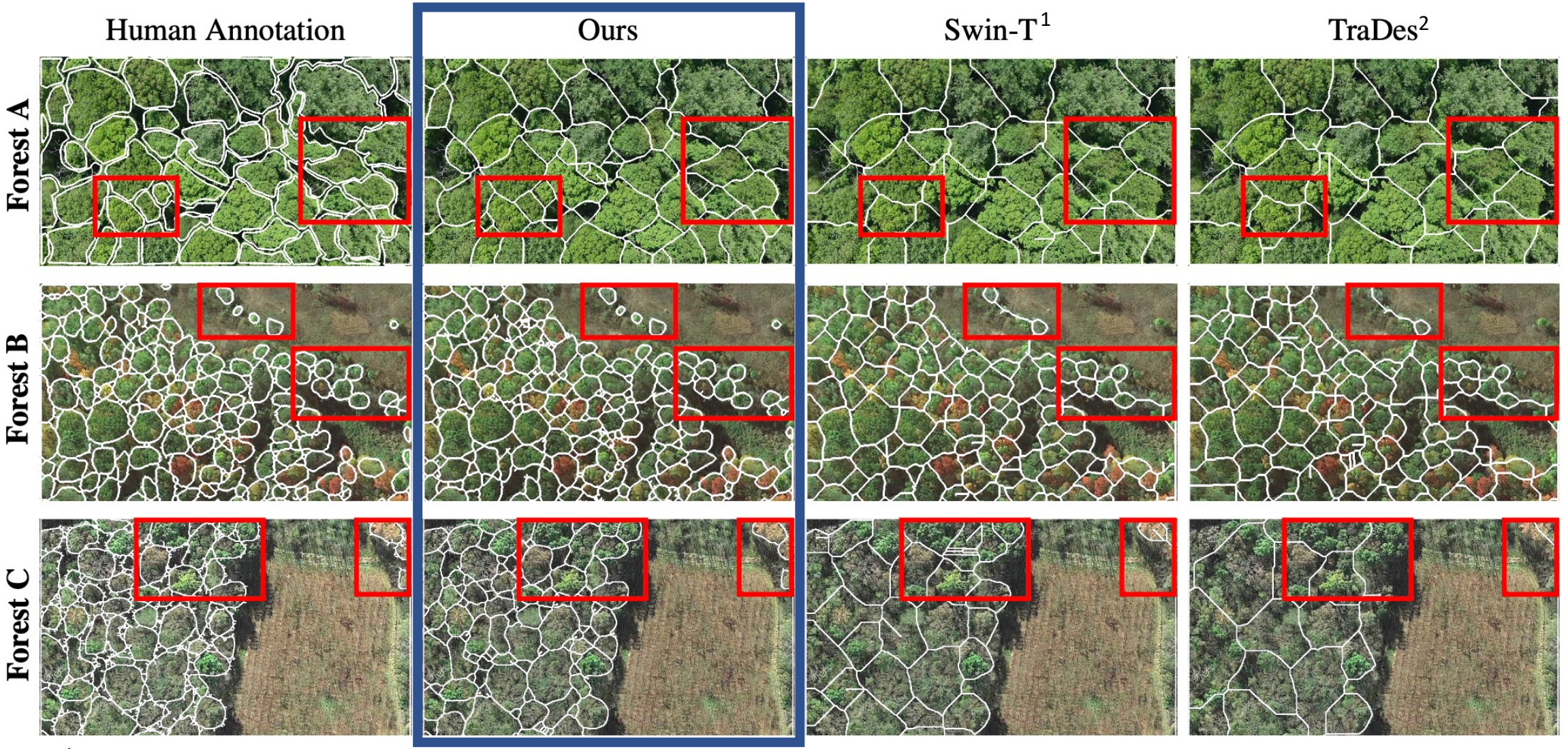
Methods	Synthetic Forest			Forest A			Forest B			Forest C		
	GT	Pred.	Acc. ↑	GT	Pred.	Acc. ↑	GT	Pred.	Acc. ↑	GT	Pred. ↑	Acc. ↑
Mask-RCNN (ResNet)	5157	3291	63.8	2314	1691	73.1	2041	1281	62.8	2172	1403	64.6
Mask-RCNN (Swin-T)	5157	4275	82.9	2314	1816	78.5	2041	1655	81.1	2172	1791	82.5
TraDeS	5157	4131	80.1	2314	1987	85.9	2041	1596	78.2	2172	1611	74.2
BoundaryFormer	5157	3981	77.2	2314	1950	84.3	2041	1549	75.9	2172	1713	78.9
Ours	5157	5636	**90.7**	2314	2510	**91.5**	2041	1771	**86.8**	2172	2384	**90.2**

Forest A = Martell Forest, IN
Forest B = Kentucky Ridge State Forest, KY
Forest C = Olympic National Forest, WA
[More baseline comparisons are given in the paper.]

Roadmap

- Overview
- Observations
- Pipeline
- Quantitative Results
- Qualitative Results
- Future Works

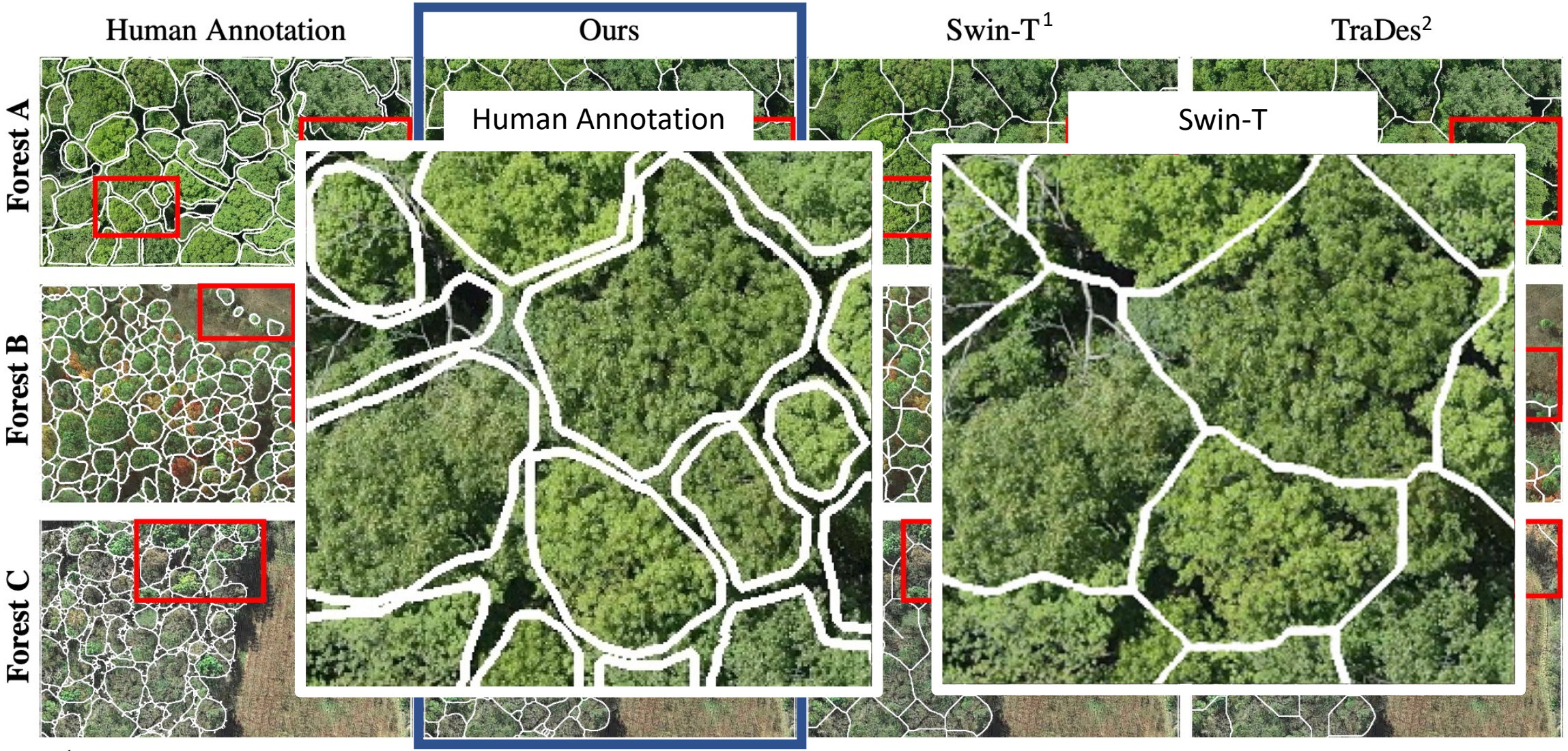
Qualitative Results



[1] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Int. Conf. Comput. Vis.*, 2021

[2] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Int. Conf. Comput. Vis.*, 2021

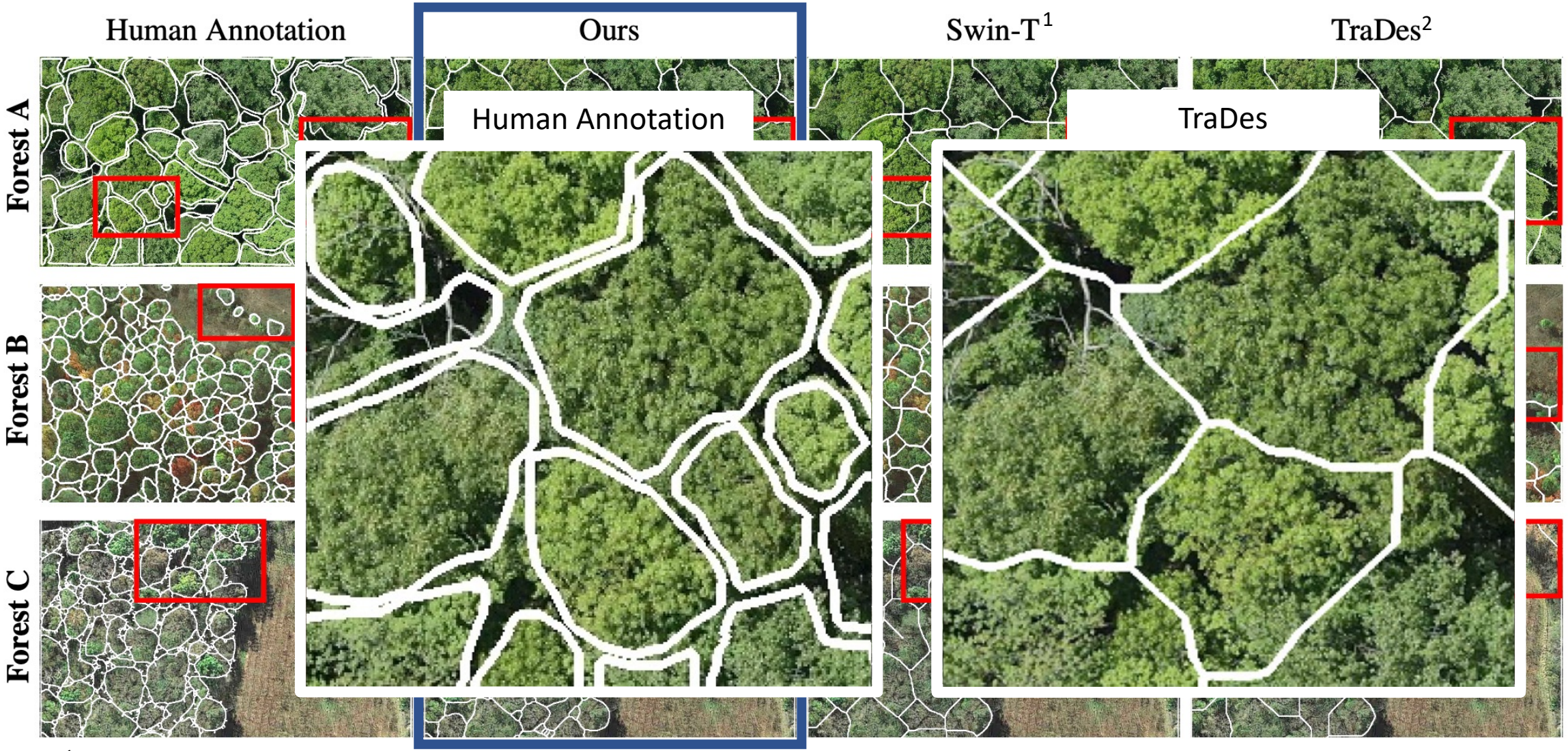
Qualitative Results



[1] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Int. Conf. Comput. Vis.*, 2021

[2] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Int. Conf. Comput. Vis.*, 2021

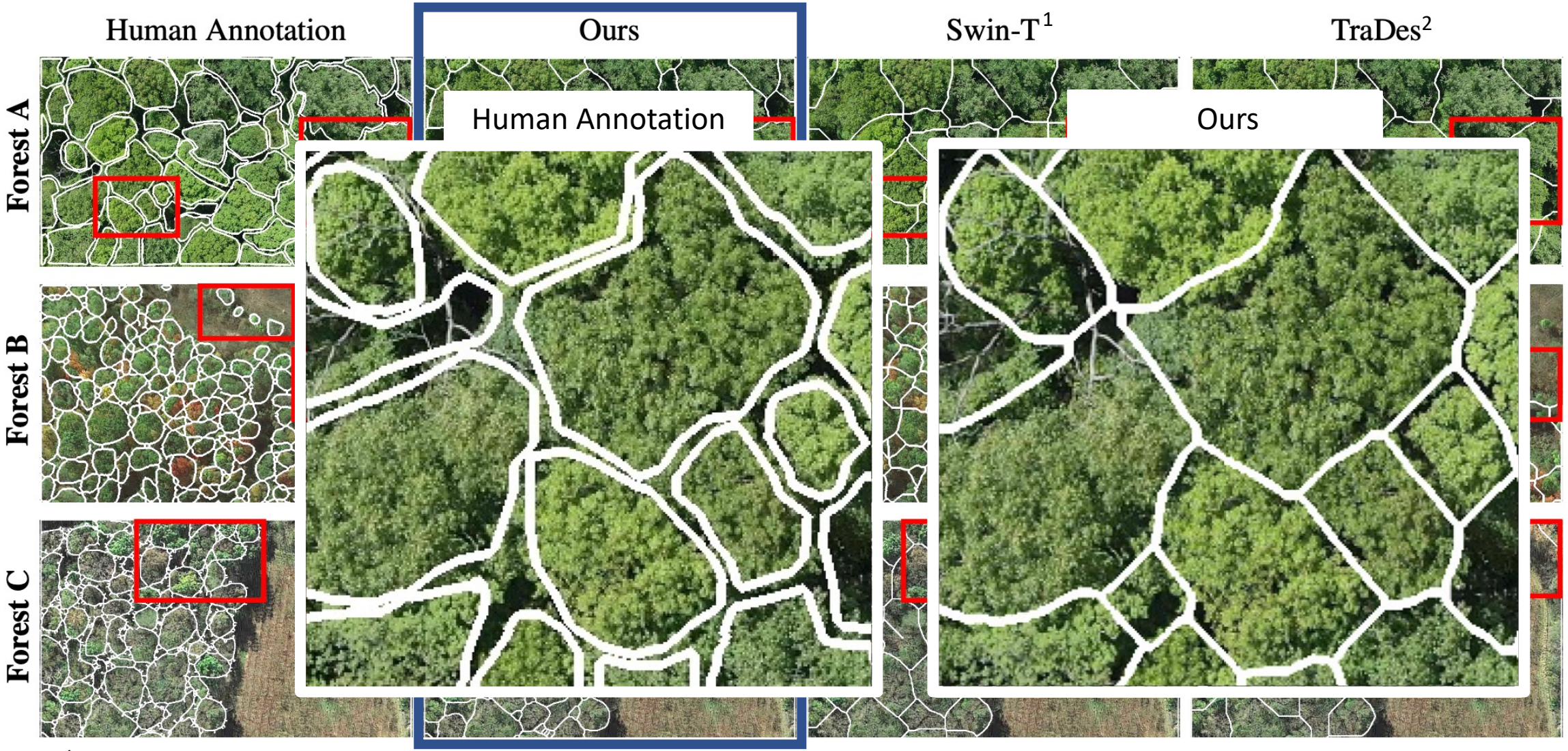
Qualitative Results



[1] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Int. Conf. Comput. Vis.*, 2021

[2] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Int. Conf. Comput. Vis.*, 2021

Qualitative Results



[1] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Int. Conf. Comput. Vis.*, 2021

[2] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Int. Conf. Comput. Vis.*, 2021

Roadmap

- Overview
- Observations
- Pipeline
- Quantitative Results
- Qualitative Results
- Future Works

Future Works

- Extend the approach to other tightly packed objects such as:
  - Dense Crowds
  - Animals
  - Bees
  - Traffic
- Explore other continents
- Implement an online solution to the approach.

Thank you very much for your time and attention.