Automated Structural Component Recognition Using Scribble Annotations: A Cost-effective Method for Creating the Digital Twin for Transportation Asset Management

Chenyu Zhang [a], Ke Li [a], Zhaozheng Yin [b], Ruwen Qin [a,*]

[a] Department of Civil Engineering, Stony Brook University
[b] Department of Computer Science, Department of Biomedical Informatics, and AI Institute, Stony Brook University

Introduction

Digital Twins for Transportation Asset Mgt.

- Sensor and robotic technologies have enabled acquisition of large amount of data
- Digital twins can enhance safety and efficiency

A Challenge in Creating Digital Twins

- Rely on fully labeled mask annotations
- Creating masks is time-consuming and labor-intensive

Girder
Bearing
Substructure

Original image   Mask annotation   Scribble annotation

Research Question

How can scribble annotations enable efficient and accurate segmentation of structural components?

Experimental Settings

SBCIV-ScribComp Dataset

- 440 high-resolution images from bridge inspections
- 306 images for training, 34 for validation, and 100 for testing
- Six structural components: bearing, bracing, deck, floor beam, girder, and substructure

Illustrative Examples

Mask   Scribble   Mask   Scribble

Results

Comparison with State-of-the-Art

Type	Method	mIoU ↑	mAcc ↑	pAcc ↑
	U-Net	82.60	88.95	93.98
	PSPNet	85.83	91.44	95.01
F	DeepLabv3+	86.57	92.31	95.36
	SegFormer	89.41	94.22	96.20
	HRNet	**91.00**	**95.58**	**96.74**
	ScribbleSup	67.80	80.28	89.41
	M Loss	70.27	82.13	90.20
S	EM	74.52	85.04	90.25
	WSL4MIS	75.97	86.71	90.45
	R Loss	83.23	91.56	93.87
	ScribCompNet (Ours)	**90.19**	**94.54**	**96.27**

- ScribCompNet surpasses other scribble-supervised methods
- It achieves comparable performance to the best HRNet

Qualitative Evaluation

Input Image   Ground Truth   HRNet (F)   ScribCompNet
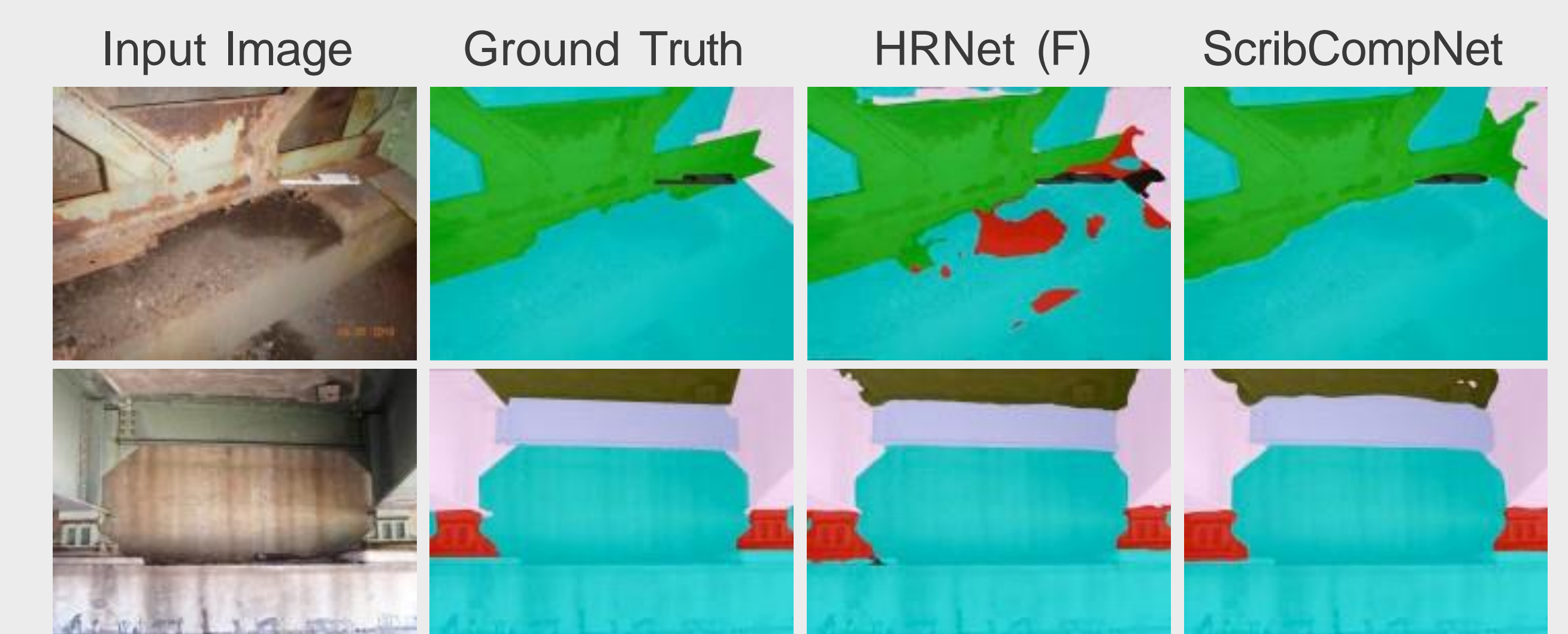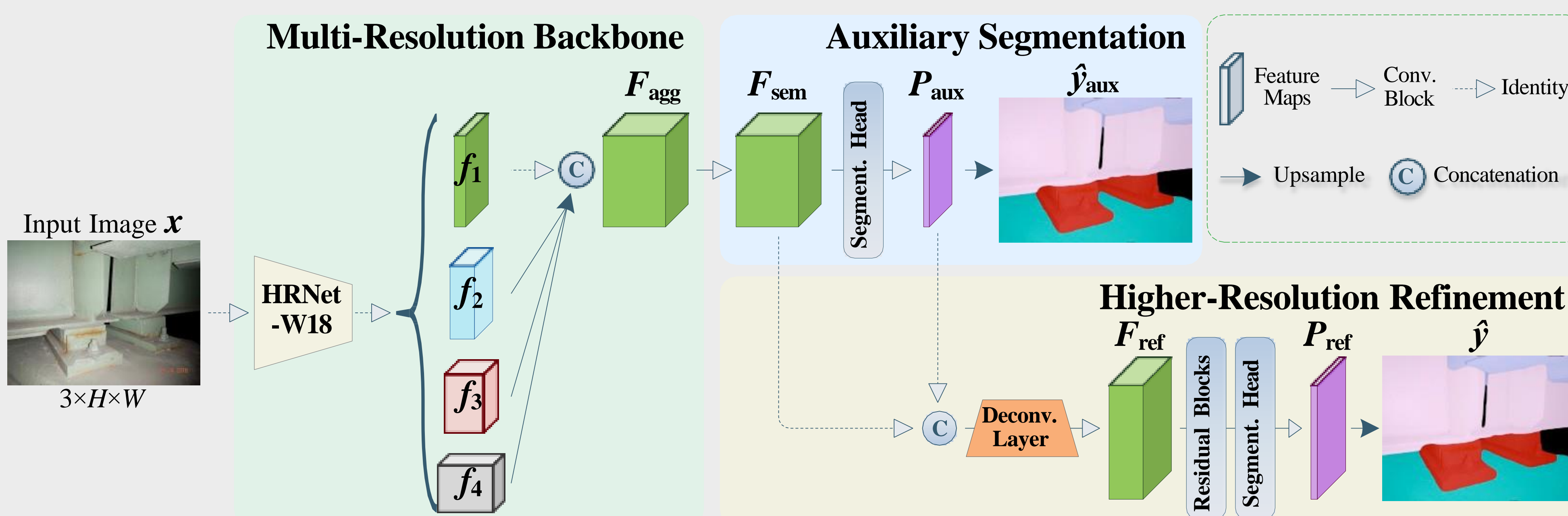
Labeling Time Cost
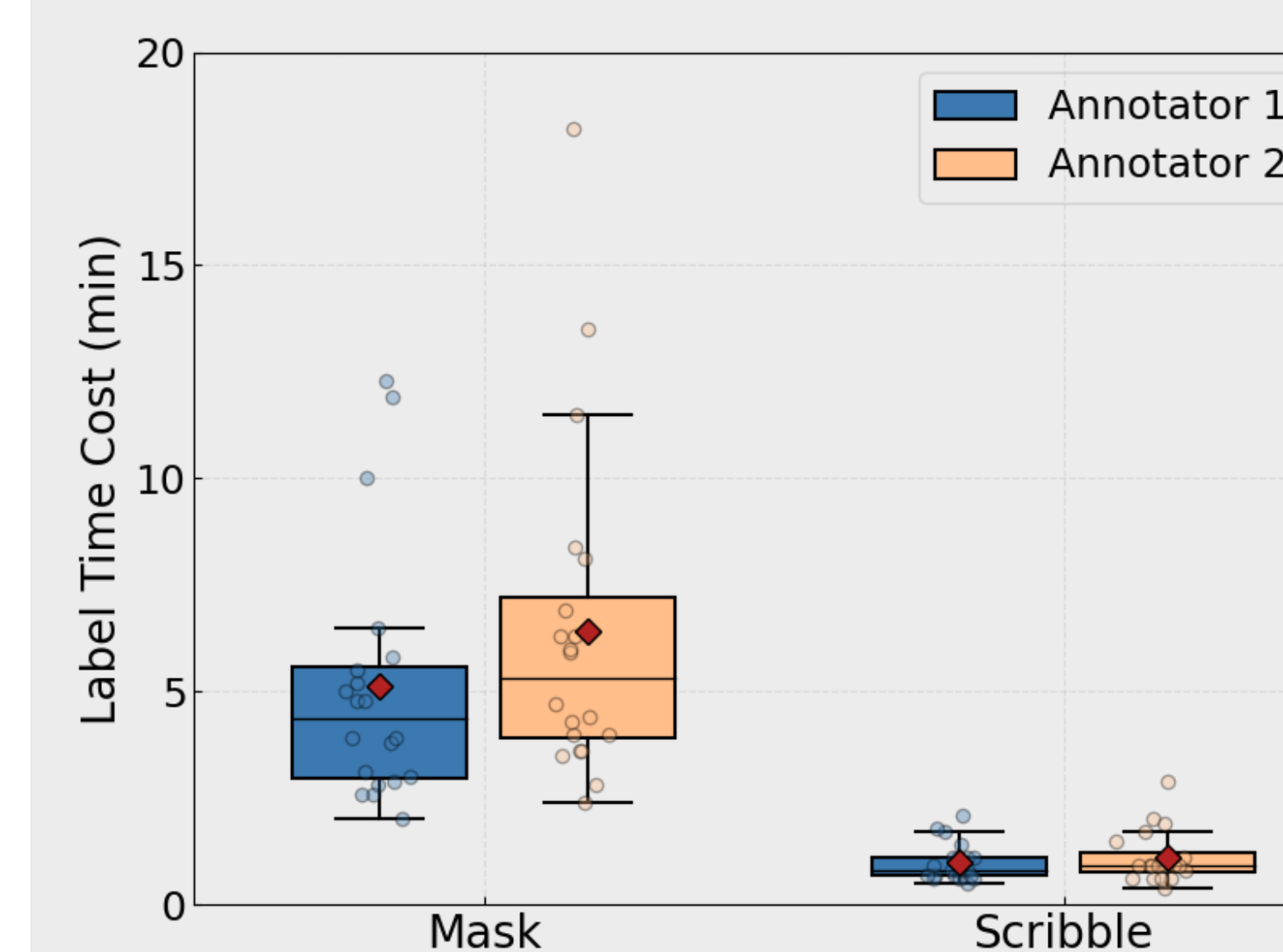
- Scribbles take about 1 minute per image
- Scribble labeling is 80% faster than mask annotation
- This efficiency offers faster deployment of inspection models

Methodology

ScribCompNet Architecture

Multi-Resolution Backbone

F_{agg} F_{sem} P_{aux} $\hat{y}_{aux}$

Input Image x HRNet-W18 f_1 f_2 f_3 f_4

$3 \times H \times W$

Auxiliary Segmentation

Feature Maps → Conv. Block → Identity
Upsample   C Concatenation

Higher-Resolution Refinement

F_{ref} P_{ref} $\hat{y}$

Deconv. Layer   Residual Blocks   Segment. Head

Joint Objective Function

- Scribble annotation loss provides labeled-pixel information
- Dynamic pseudo label loss integrates outputs of both branches
- Semantic context enhancement loss leverages pixel semantic context
- Scale-adaptive harmony loss ensure consistency across scales

Contact Information

ruwen.qin@stonybrook.edu

Code   Paper

Acknowledgement