Emre Gülsoylu[1], André Kelm[1], Lennart Bengtson[2], Matthias Hirsch[1], Christian Wilms[1], Tim Rolff[3], Janick Edinger[2] and Simone Frintrop[1]

[1] University of Hamburg, CV Group [2] University of Hamburg, DOS Group
[3] University of Hamburg, HCI Group

TRUDI and TITUS: A Multi-Perspective Dataset and A Three-Stage Recognition System for Transportation Unit Identification

Motivation

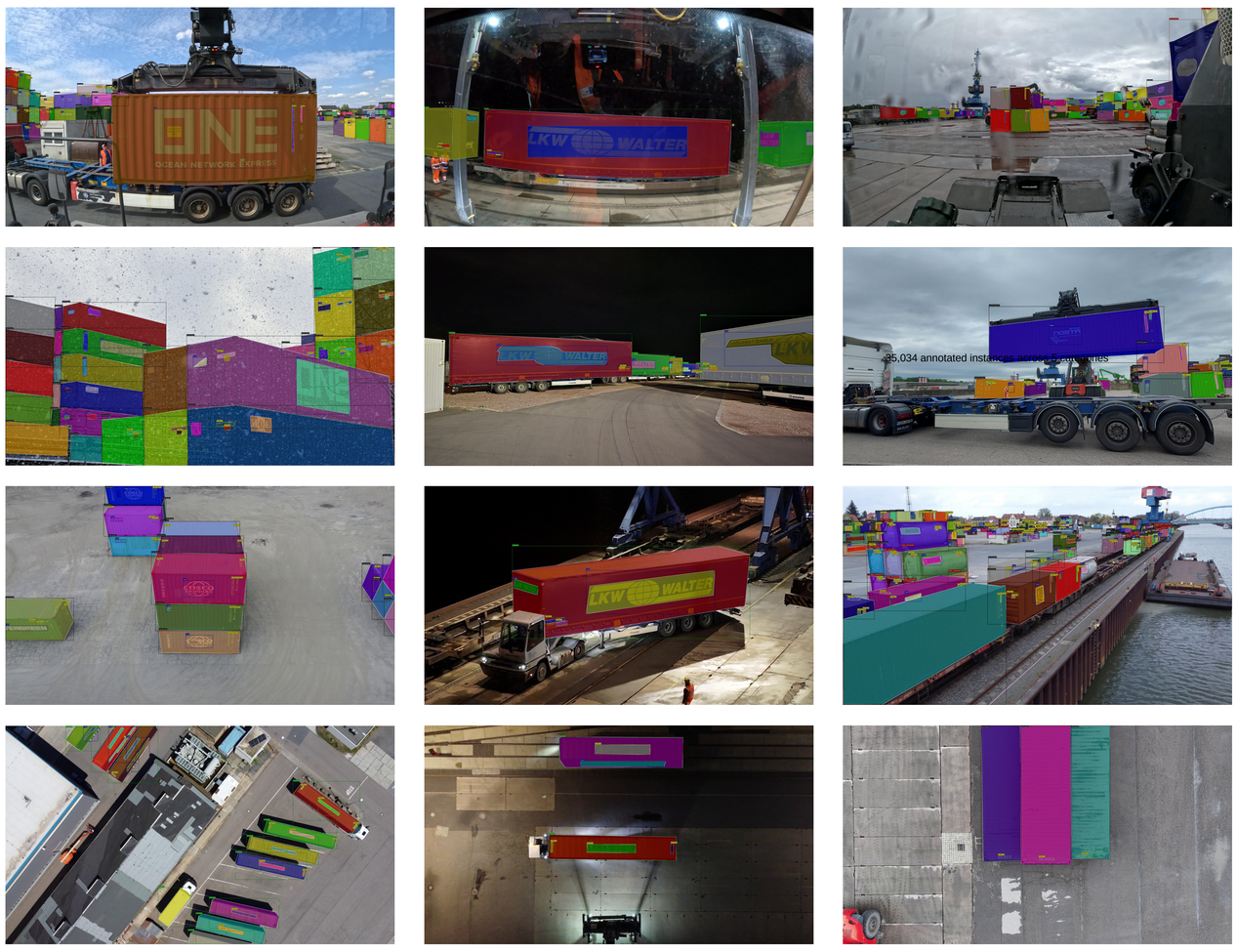
Task: Detect and recognise ID codes on transportation units (TU), such as containers and trailers.

Problem:
- Lack of publicly available datasets for TU identification and benchmarking.
- Reliance on fixed camera settings and dynamic, real-world conditions.

Solution:
- **TR**ansportation **U**nit **D**etection and **I**dentification (TRUDI) dataset contains aerial and ground imagery from multiple terminals, taken under various weather conditions.
- **T**hree-stage **I**dentification of **T**ransportation **U**nit**S** (TITUS) pipeline was developed for robust, multi-perspective TU identification.

TRUDI Dataset

35,034 annotated instances (containers, trailers, tank containers, ID text, logos) from ground and aerial perspectives.



- Suitable for various CV tasks such as object detection, instance segmentation, logo detection, text detection, text recognition and text spotting.
- Varying lighting, weather, perspectives (ground/aerial), and image quality.
- 18-month collection across multiple countries, times of day, and seasons.

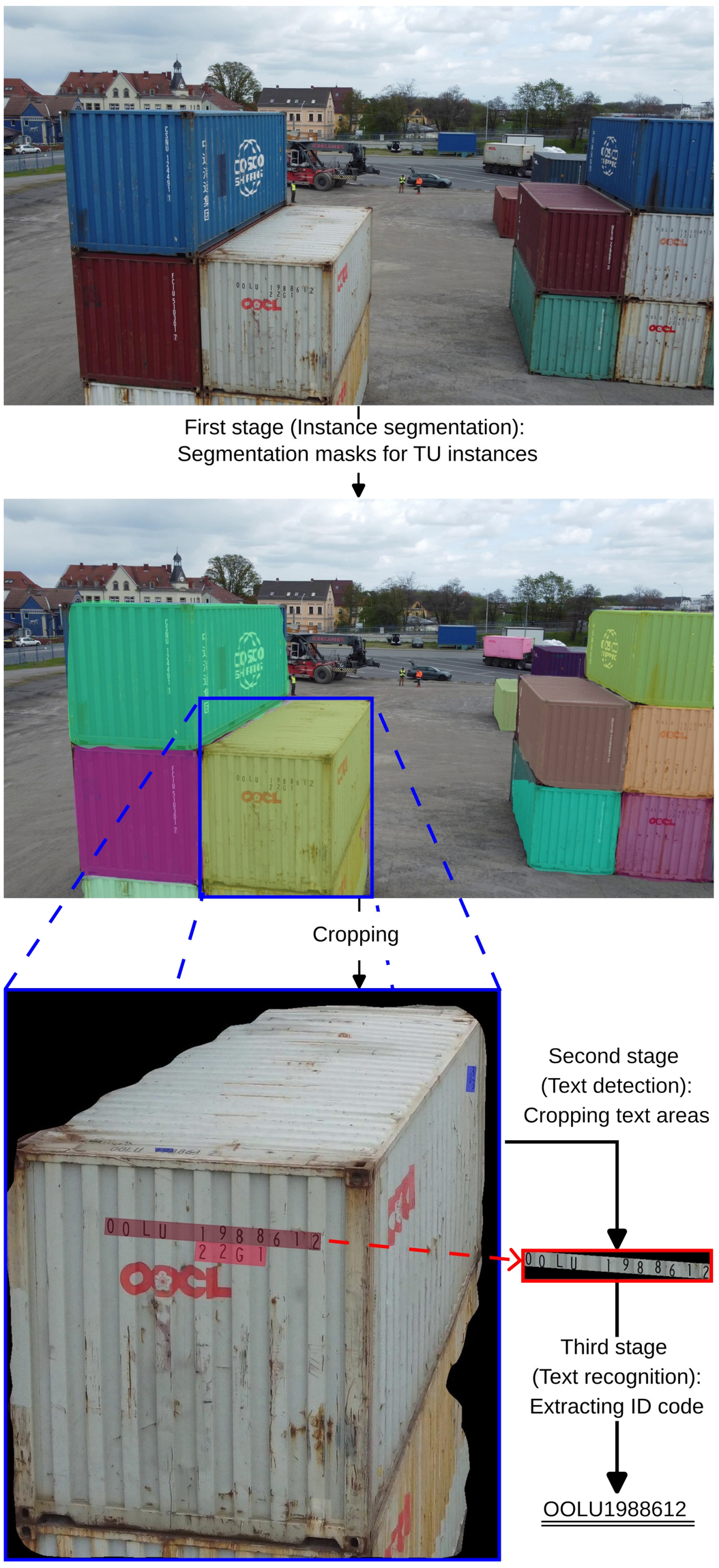
Category	# Instances
Container	11109
Tank Container	808
Trailer	2780
ID Text	14009
Logo	6328

Evaluation

Stage	Perspective	Model	AP@0.50:0.95	AP@0.50	AR@0.50:0.95
Instance Seg.	Aerial	Mask R-CNN-Swin(1C)[1,2]	0.43	0.62	0.48
	Ground	Mask R-CNN-Swin(3C)-C[1,2]	**0.44**	**0.67**	**0.50**
	Combined	Mask R-CNN-Swin(3C)[1,2]	**0.44**	0.65	**0.50**
Stage	Perspective	Model	Recall	Precision	Hmean
Text Det.	Aerial	DBNet++-C[3]	**0.79**	0.68	0.73
	Ground	DBNet++-C[3]	**0.79**	**0.69**	**0.74**
	Combined	DBNet++[3]	**0.79**	**0.69**	**0.74**
Stage	Perspective	Model	Word acc.	Char. recall	Char. precision
Text Rec.	Aerial	RobustScanner-C[4]	**0.68**	**0.88**	**0.88**
	Ground	RobustScanner-C[4]	0.54	0.73	0.73
	Combined	RobustScanner[4]	0.63	0.83	0.84
Stage	Perspective	Precision	Recall	F1 Score	Accuracy
End-to-end	Aerial	**0.45**	**0.30**	**0.36**	**0.22**
	Ground	0.25	0.19	0.22	0.12
	Combined	0.39	0.27	0.32	0.19

Acknowledgment: The project is supported by the German Federal Ministry for Digital and Transport in the funding program Innovative Hafentechnologien II.

TITUS Pipeline



Paper + Dataset + Weights

References

[1] Kaiming He, et al. Mask R-CNN. In Proc. of the IEEE International Conference on Computer Vision, 2961–2969, 2017.

[2] Ze Liu, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In Proc. of the IEEE/CVF International Conference on Computer Vision, 10012–10022, 2021.

[3] Minghui Liao, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion. IEEE Trans. on Pattern Analysis and Machine Intelligence, 45(1):919–931, 2022.

[4] Xiaoyu Yue, et al. RobustScanner: Dynamically enhancing positional clues for robust text recognition. In Proc. European Conference on Computer Vision, 135–151. Springer, 2020.