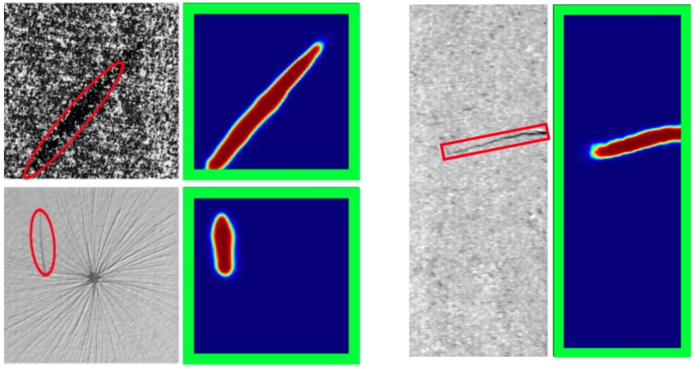
End-to-end training of a two-stage neural network for defect detection

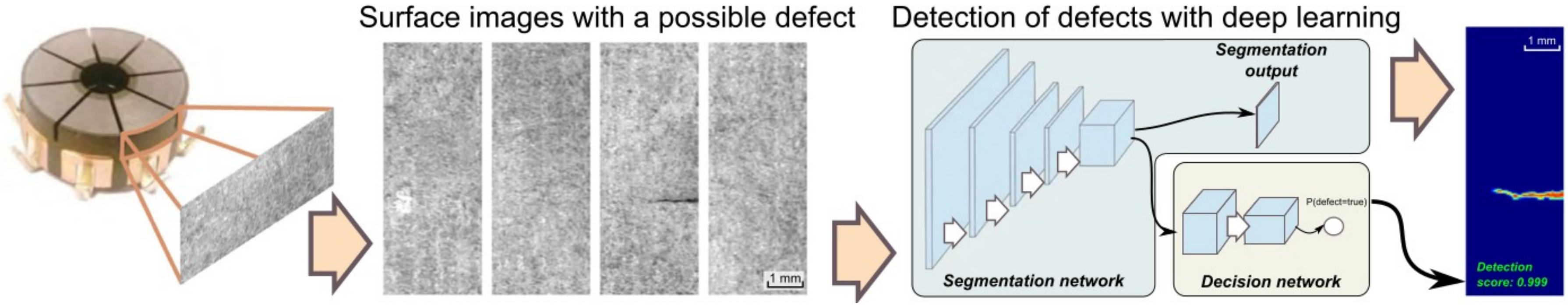
Jakob Božič, Domen Tabernik and Danijel Skočaj

University of Ljubljana
Faculty of Computer and Information Science, ViCoS

Industrial quality control

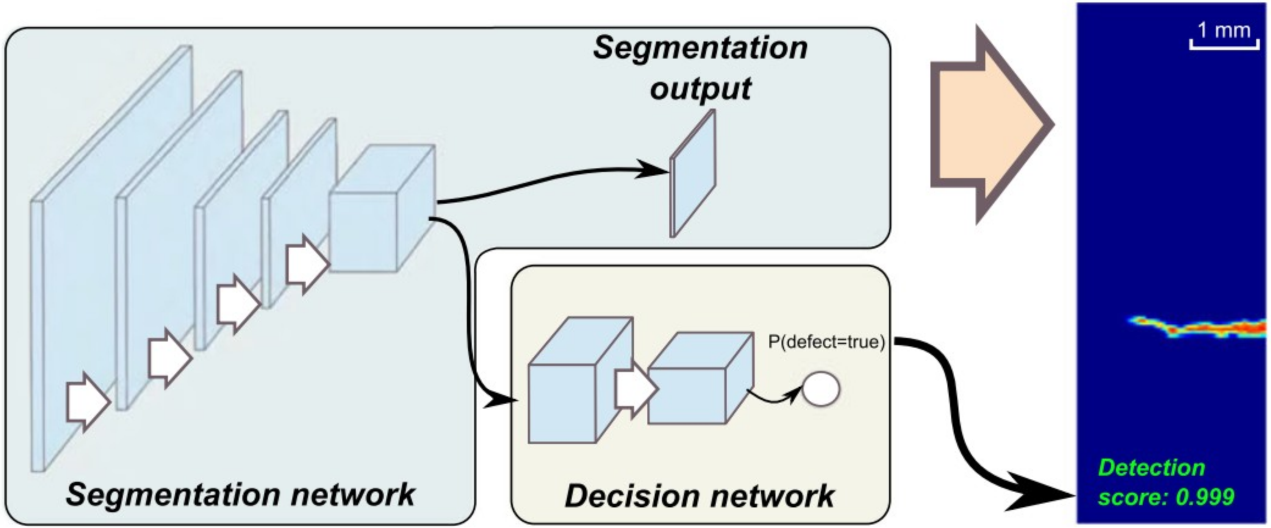
- One of the key processes in manufacturing
  - Classical machine vision methods
- Transition to deep learning
  - No feature hand-engineering





Previous work

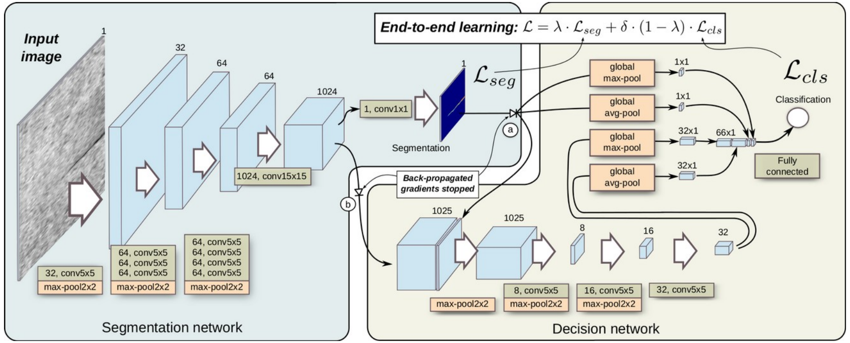
- Tabernik et al., 2019:
  - Two-stage architecture
  - Two learning phases
  - Cumbersome training process



D. Tabernik et al., Segmentation-based deep-learning approach for surface-defect detection, JIM 2019

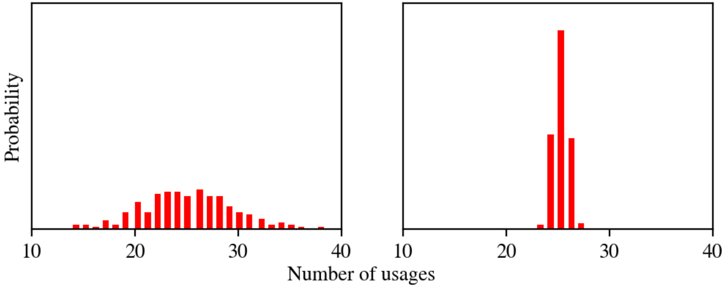
Our contributions

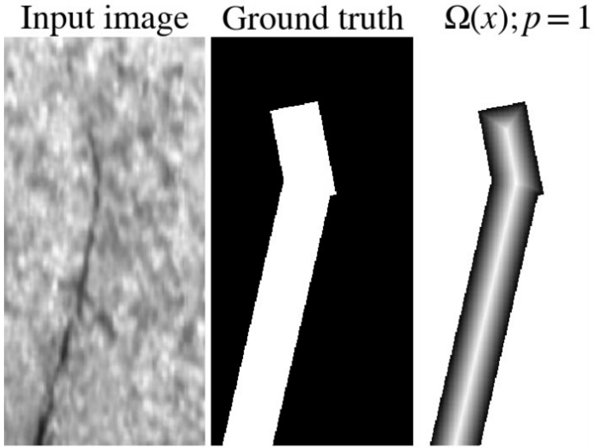
1. End-to-end learning
   - ⇒ Simultaneous training of both stages
   - ⇒ Gradient flow adjustments



2. Supervised learning with coarse annotations
   - ⇒ Coarse pixel-level (i.e., region-level) annotations
   - ⇒ Weighted segmentation loss

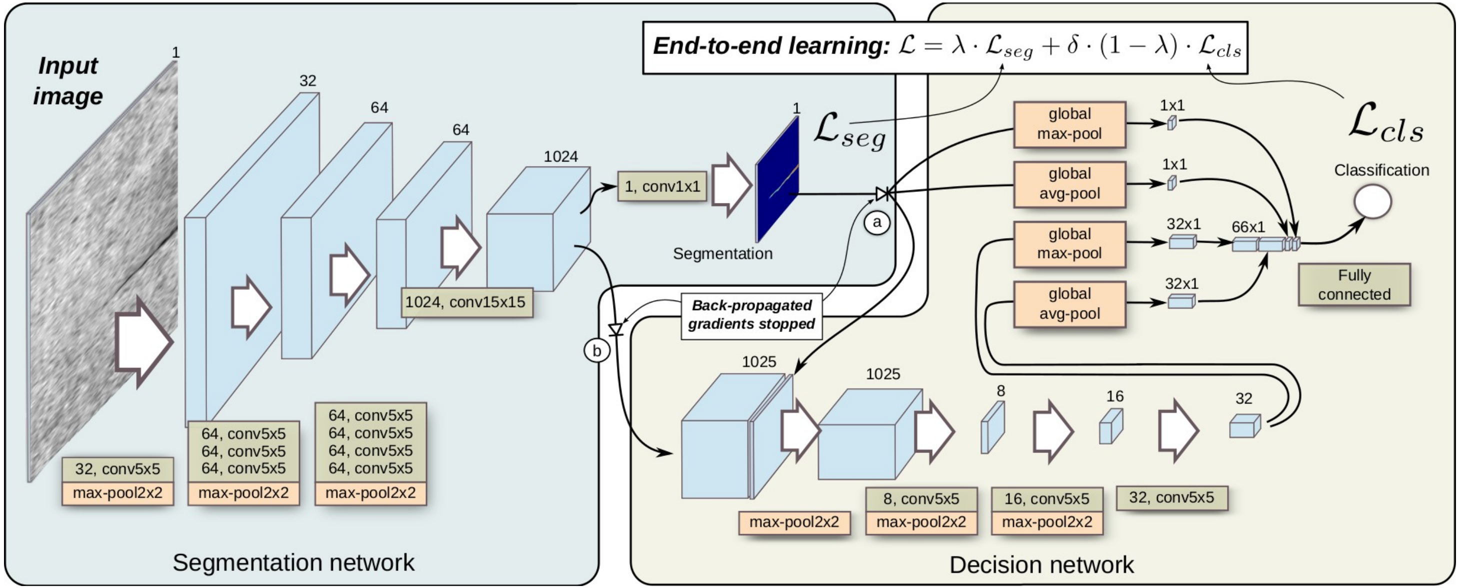
3. Frequency-of-use sampling
   - ⇒ Account for the unbalanced training data
   - ⇒ Undersampling with frequency-of-use based probabilities





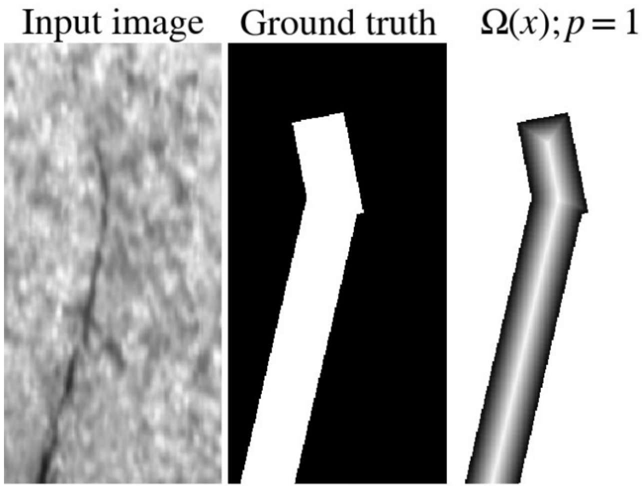
End-to-end training

- Joint loss: $\mathcal{L}_{total} = \lambda \cdot \mathcal{L}_{seg} + \delta \cdot (1-\lambda) \cdot \mathcal{L}_{cls}$ $\lambda = 1 - \frac{n}{total\_epoch}$
- Remove gradient flow from decision stage into segmentation stage



Weighted segmentation loss

- Varying uncertainty of presence of defect:
  - Greater probability of presence in center of region
  - Higher uncertainty near edges

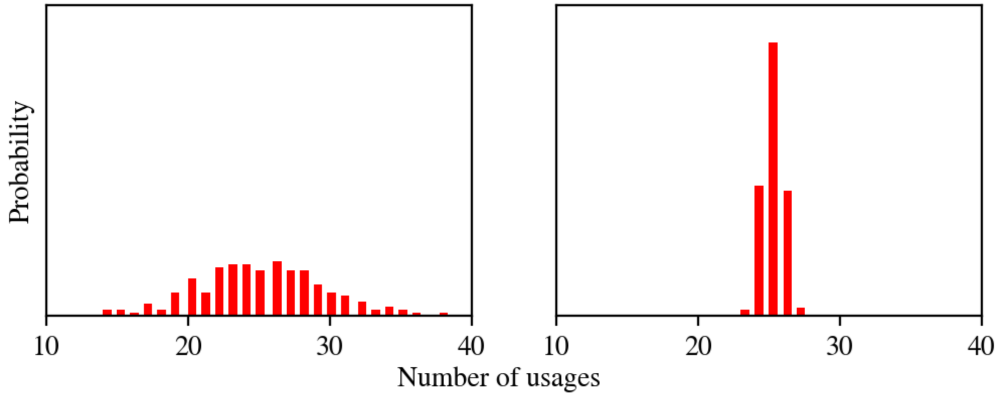


$$\mathcal{L}_{seg} = \Omega\left(\frac{\mathcal{D}(pix)}{\mathcal{D}(pix_{max})}\right) \cdot \hat{\mathcal{L}}(pix)$$

$$\Omega(x) = w_{pos} \cdot x^p$$

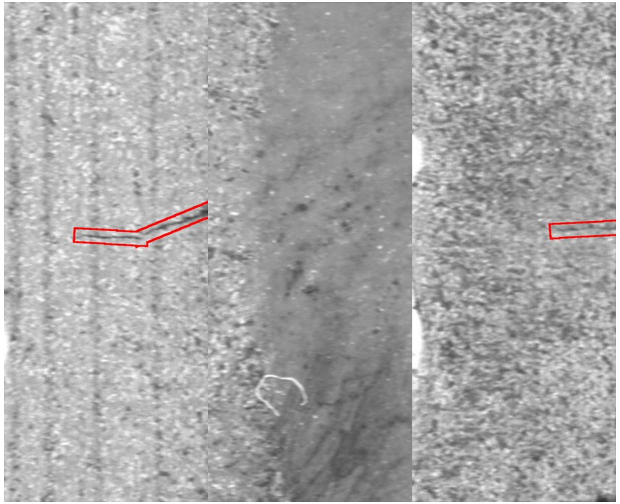
Frequency-based sampling

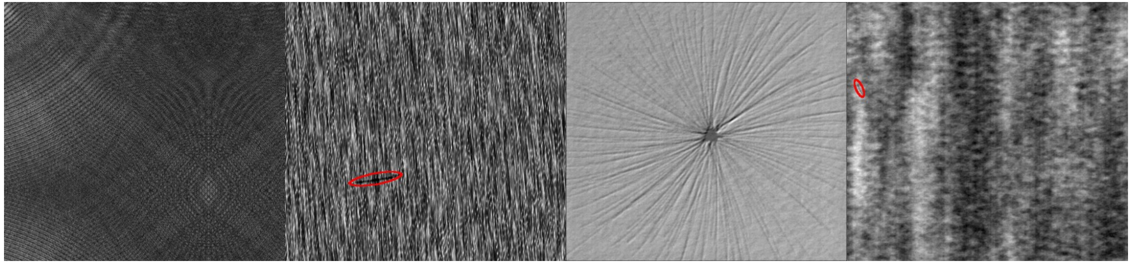
- Heavily unbalanced dataset
  - Undersampling of non-defective samples
- Sample negatives with probability inversely proportional to their frequencies of selection
  - Assures all samples are used approximately equal number of times

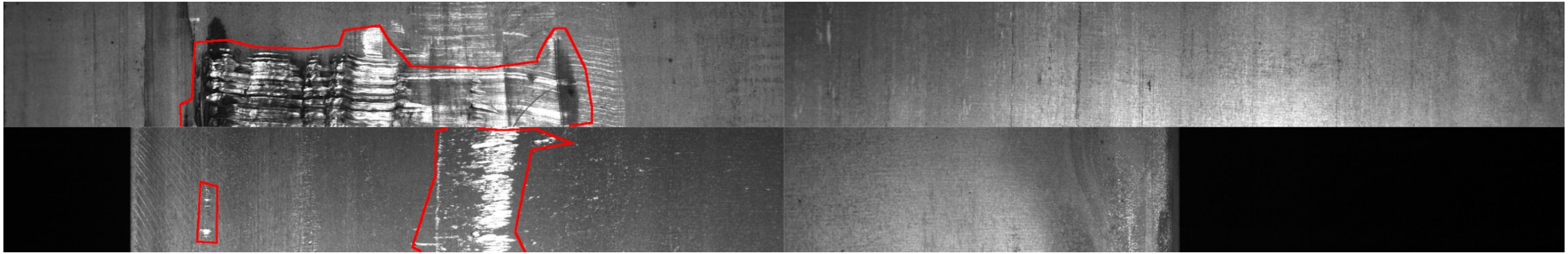


Evaluation

- KolektorSDD
- DAGM
- Severstal Steel Defect dataset

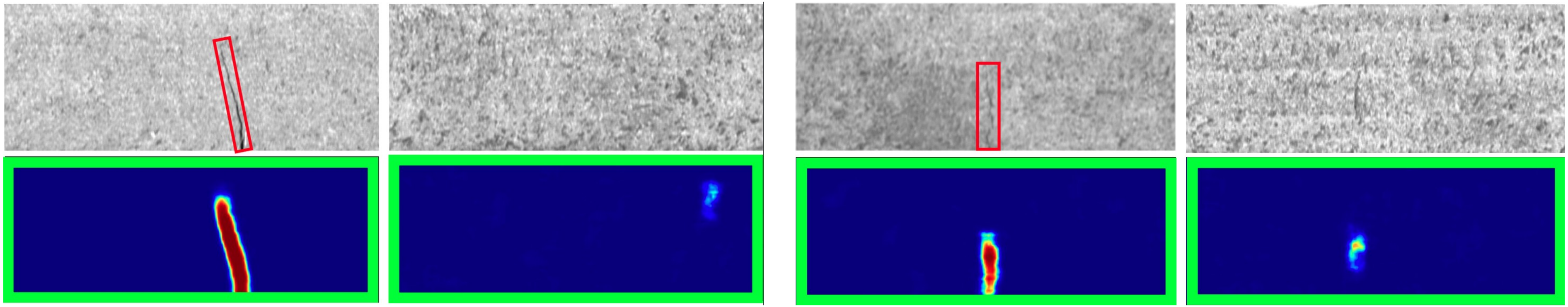






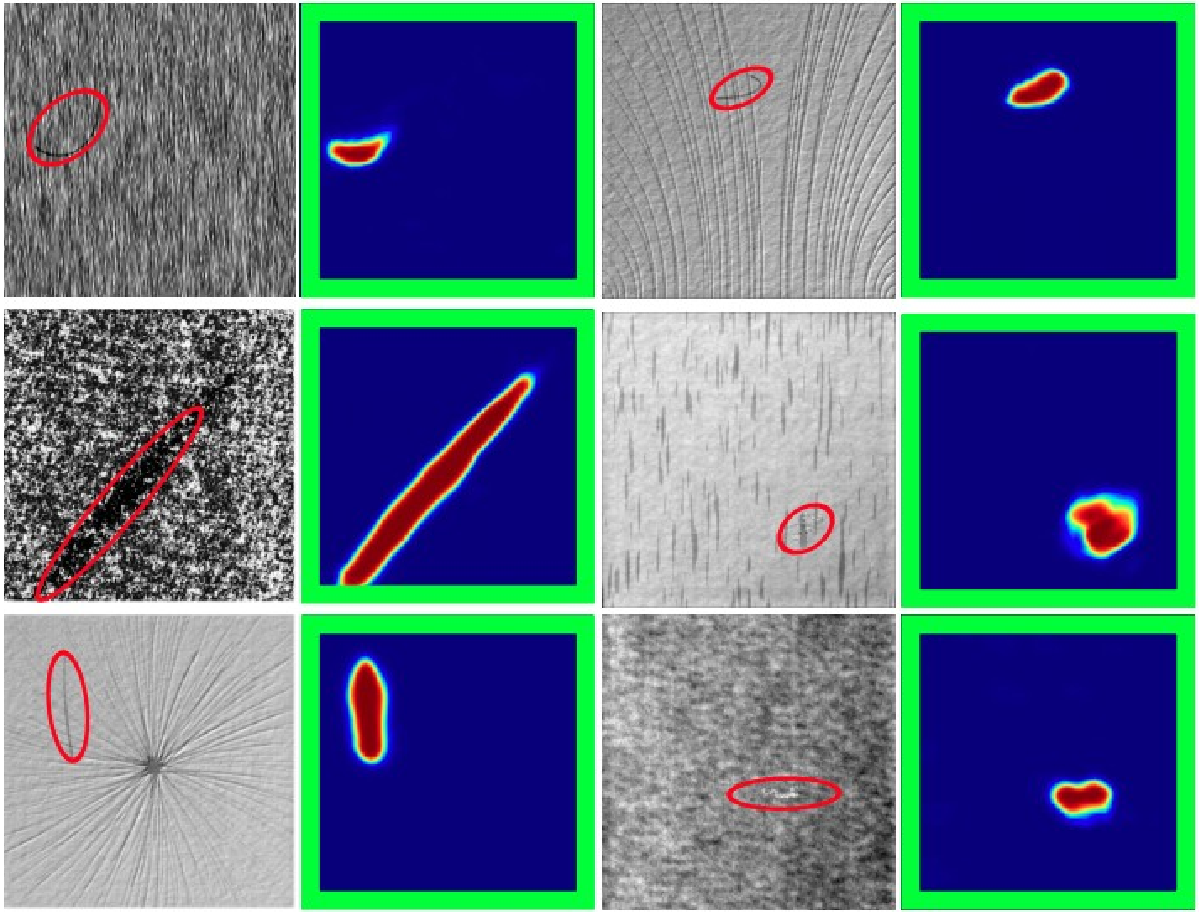
KolektorSDD

Architecture and approach	*Learning stages*	*Number of positive training samples*					
		33	*25*	*20*	*15*	*10*	*5*
Extended Segmentation+Decision Network (ours)	*end-to-end*	**100.00**	**99.78**	**100.00**	**99.88**	**99.31**	**96.71**
Segmentation+Decision Network [9]	*separate (two stages)*	99.0	97.5	99.5	97.4	98.8	95.8
Cognex ViDi (commercial software) [9]	-	99.0	97.4	95.7	97.1	95.6	89.2
Xu et al. [13] (image-level label only)	*separate (three stages)*	99.5	-	-	-	98.0	-
Pre-trained ResNet [13] (image-level label only)	-	97.8	-	-	-	-	-



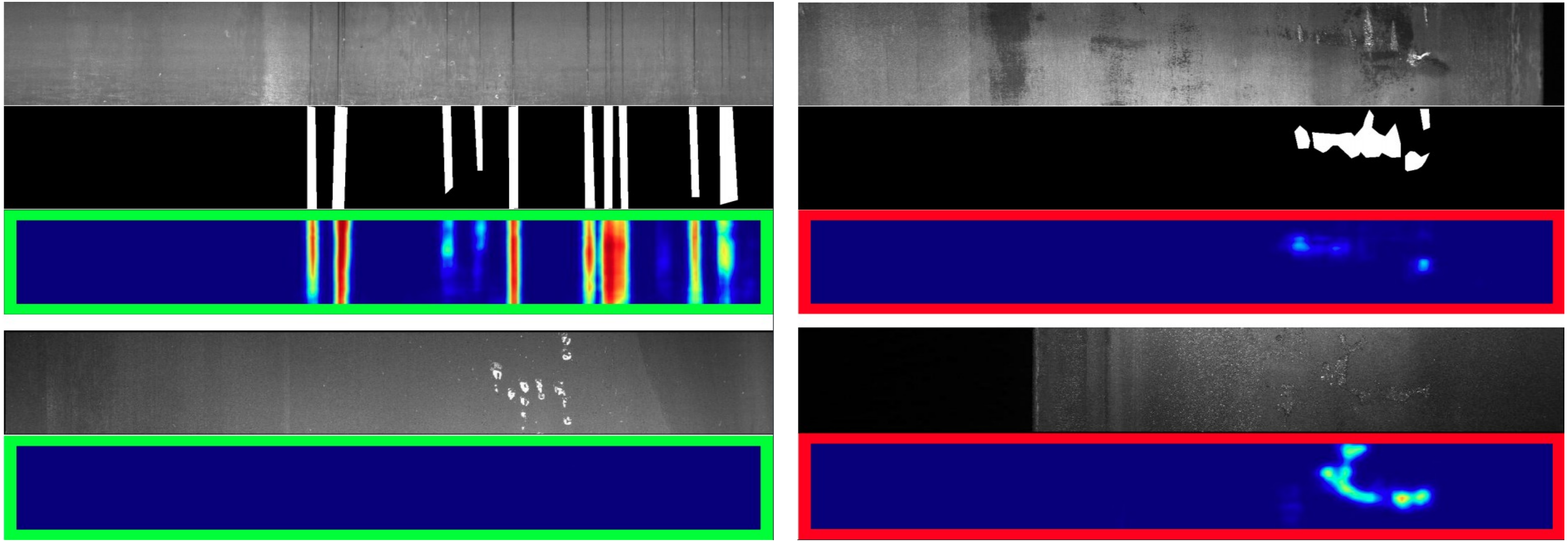
DAGM

Surface	Ours		Rački et al. [7]	
	TPR	TNR	TPR	TNR
1	100	100	100	98.8
2	100	100	100	99.8
3	100	100	100	96.3
4	100	100	98.5	99.8
5	100	100	100	100
6	100	100	100	100
7	100	100	100	100
8	100	100	100	100
9	100	100	100	99.9
10	100	100	100	100



Severstaal steel

Metric	*Number of positive training samples*			
	3000	*1500*	*750*	*300*
Average precision (AP)	99.04	99.00	98.91	97.78
False positives (FP)	34	41	52	95
False negatives (FN)	54	70	65	77



Ablation study

DAGM		*KolektorSDD*		*Severstal Steel*		*Dynamically balanced loss*	*Gradient-flow adjustment*	*Frequency-of-use sampling*	*Distance transform*
AP	FP+FN	AP	FP+FN	AP	FP+FN				
90.84	661+45	99.77	0+1	95.90	59+102				
97.60	26+24	99.88	1+0	97.43	76+72	✓			
99.998	1+0	99.90	1+0	97.59	65+61	✓	✓		
100.00	0+0	99.88	1+0	98.24	52+58	✓	✓	✓	
100.00	0+0	100.00	0+0	98.74	59+40	✓	✓	✓	✓

Conclusion

- Introduced end-to-end training
- Solved KolektorSDD and DAGM
- Extensive ablation study

End-to-end training of a two-stage neural network for defect detection

Jakob Božič, Domen Tabernik and Danijel Skočaj

University of Ljubljana, Faculty of Computer and Information Science
Večna pot 113, 1000 Ljubljana