Reg R-CNN: Lesion Detection and Grading under Noisy Labels

Using ordinal context in the training signal to improve lesion detection and grading by swapping a CE classifier for a regressor.

G. N. Ramien, P. F. Jaeger, Simon A. A. Kohl, K. H. Maier-Hein

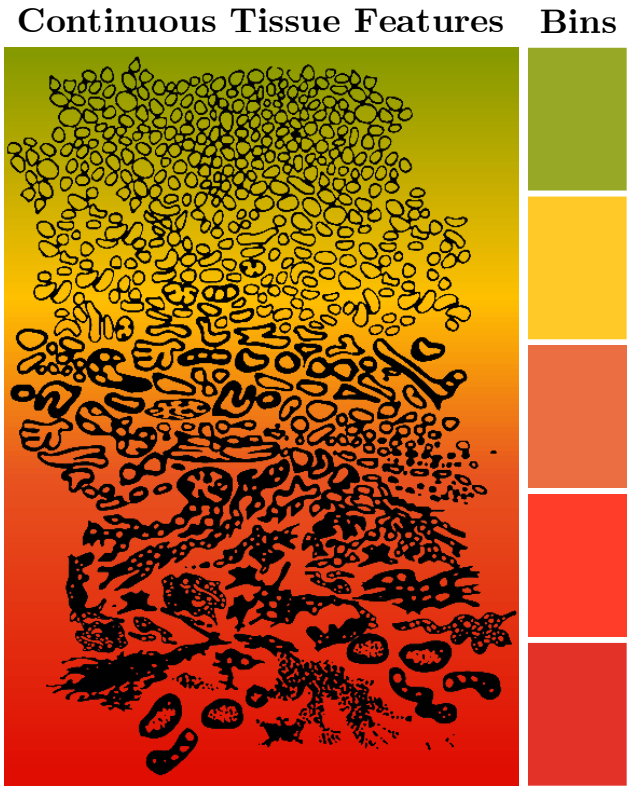
*Presented by **David Zimmerer***

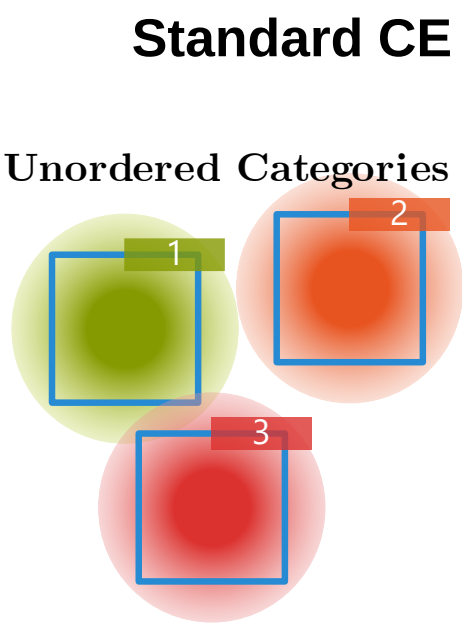
dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

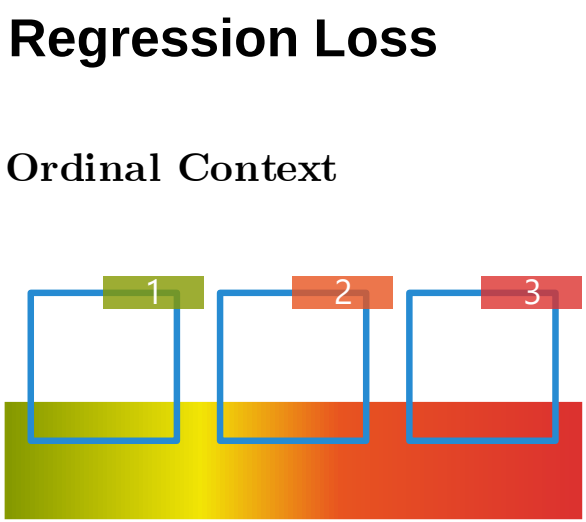
Research for a Life without Cancer

Motivation

Discarding of ordinal context: cross-entropy loss penalizes predictions irrespective of distance to target.



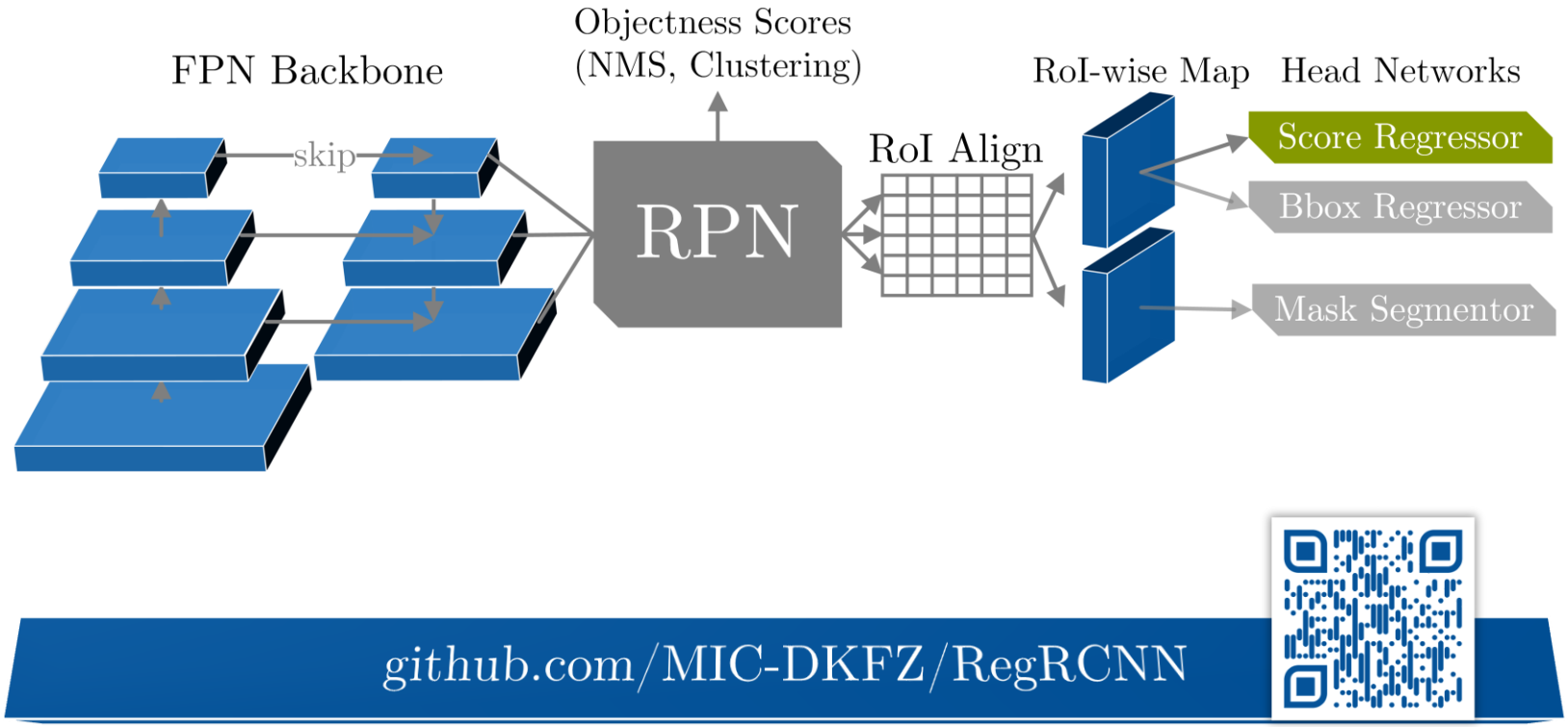




(left) NIH. Morphology & Grade: ICD-O-3 Morphology Codes, 2019

Methods

Reg R-CNN is Mask R-CNN with regressed instead of categorized object classed.

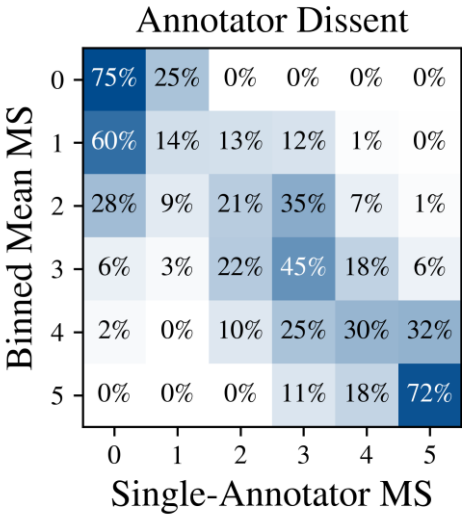


Experiments

Reg vs. Mask R-CNN on public LIDC and specific toy data set.

LIDC

- 1026 patient CT scans
- 4 annotations per volume
- Raters often disagree



Toy

- 3D cylinders
- 1.5k training, 1k test samples
- Task: detection and radius determination

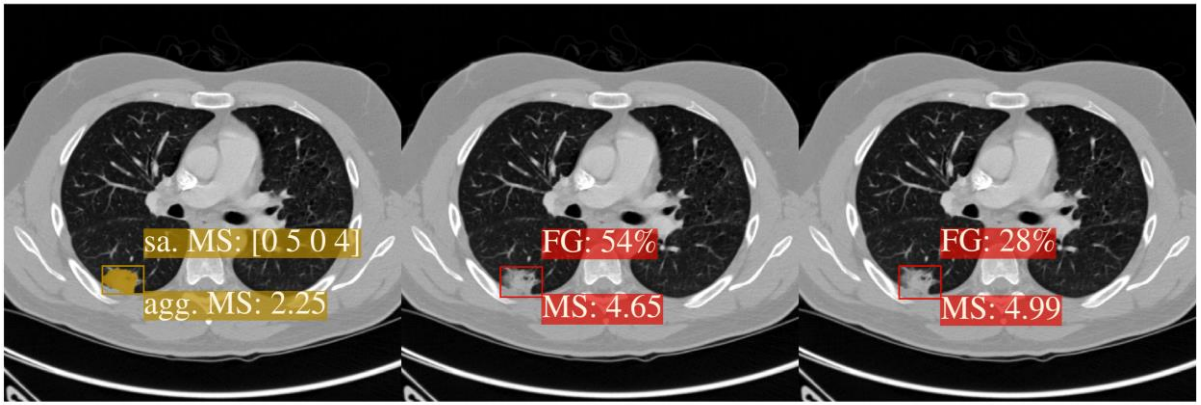
[1] Jaeger et al., Retina U-Net, 2018

Results

Reg R-CNN outperforms Mask R-CNN in all cases except 3D toy AP.

Data & Dim	Network	AVP_{10}	AP_{10}	Bin Accuracy
LIDC 3D	Reg R-CNN	0.259±0.035	0.628±0.038	0.477±0.035
	Mask R-CNN	0.235±0.027	0.622±0.029	0.411±0.026
	Δ (Reg – Mask)	**0.024±0.044**	**0.006±0.048**	**0.066±0.044**
Toy 3D	Reg R-CNN	0.881±0.014	0.998±0.004	0.887±0.014
	Mask R-CNN	0.822±0.070	1.000±0.000	0.826±0.069
	Δ(Reg – Mask)	**0.059±0.071**	**-0.002±0.004**	**0.061±0.070**

LIDC example result

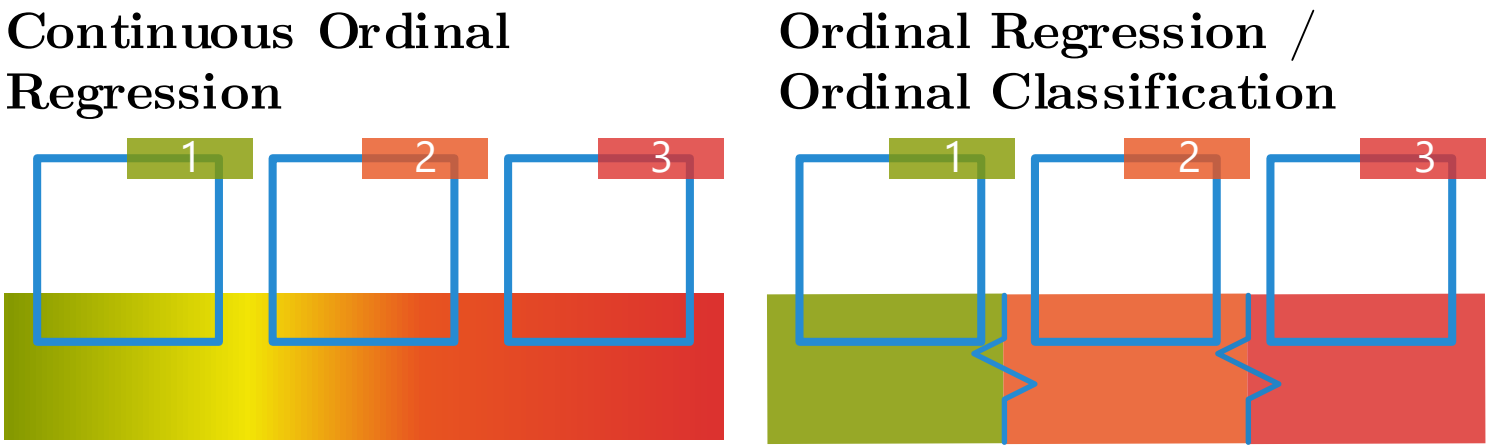


Ground Truth | Reg R-CNN | Mask R-CNN

Discussion & Outlook

Continuous regressor implies metric scale.

- We imply metric grading scale but is not guaranteed.[1]
- Future work: replace linear regressor by categorical ordinal regressor[2].



[1]Liddell, Kruschke, Analyzing ordinal data, 2018; [2]Feindt, A neural Bayesian estimator, 2004

Thank you for attending!

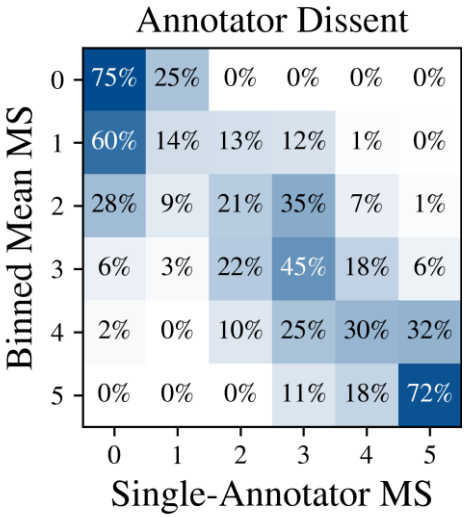
Full paper:

Experiments

Reg vs. Mask R-CNN on public LIDC and specific toy data set.

LIDC

- 1026 patient CT scans
- 4 annotations per volume
- Raters often disagree



Toy

- 3D cylinders
- 1.5k training, 1k test samples
- Task: detection and radius determination

Setup:

- 2D (slice sampling) or 3D input
- 5-fold cross-validation
- Alternating single-rater (LIDC) or noisy (toy) labels during
- Averaged multi-rater (LIDC) or exact (toy) labels during testing
- Weighted ensemble and view aggregation (WBC[1])

[1] Jaeger et al., Retina U-Net, 2018

All Results

Reg R-CNN outperforms Mask R-CNN in all cases except 3D toy AP.