Improving RetinaNet for CT Lesion Detection with Dense Masks from Weak RECIST Labels

Martin Zlocha, Qi Dou, Ben Glocker
Biomedical Image Analysis Group, Imperial College London, UK

BioMedIA
project mira ERC StG 2017 grant No 757173
erc European Research Council

A simple RetinaNet with **optimized configuration** beats previous methods!

Adding **dense supervision** from weak RECIST labels plus **attention**, improves sensitivity for **small lesions** <10 mm by over **8%**.

Beat our AI on **spot-the-lesion.com**

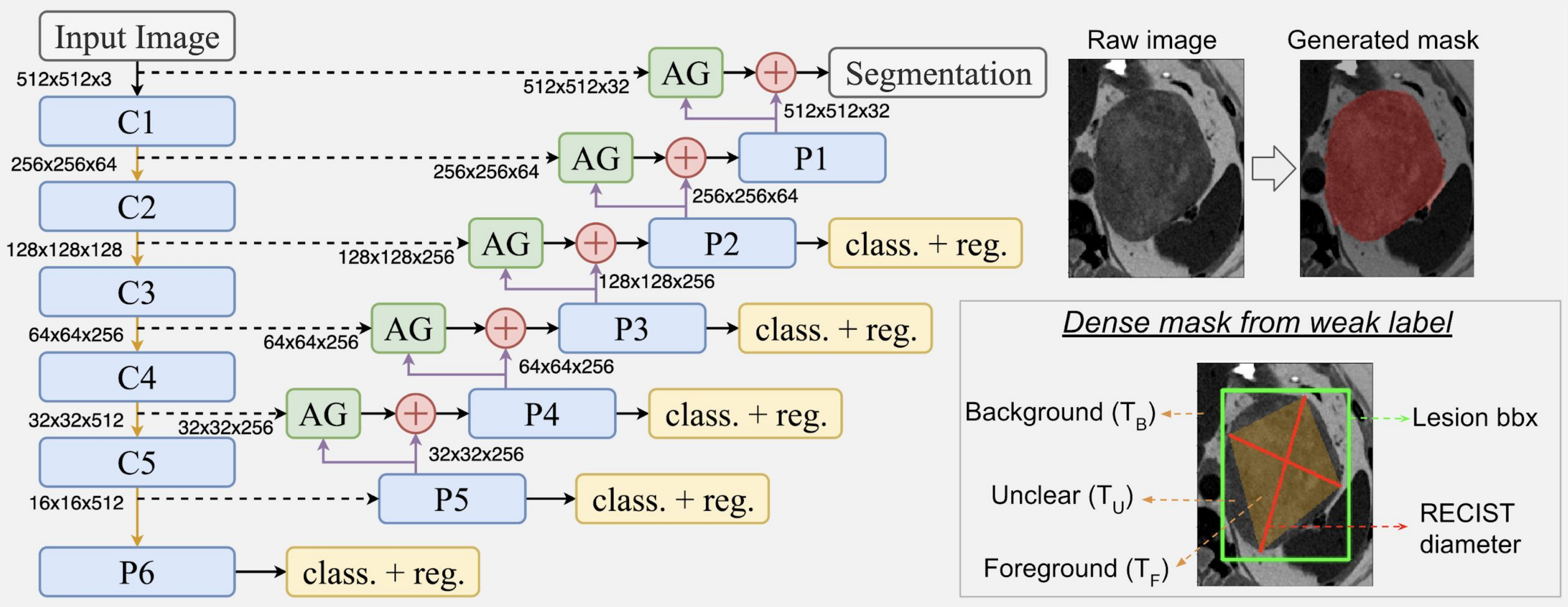
MOTIVATION

Problem: Accurate, automated lesion detection in Computed Tomography (CT) is an important yet challenging task due to the large variation of lesion types, sizes, locations and appearances. Most previous work focuses on a specific type of lesion within a relatively constrained (anatomical) context.

Solution: We propose a highly accurate and efficient one-stage lesion detector, by re-designing a RetinaNet. We evaluate our method on the public DeepLesion benchmark achieving a sensitivity of **90.77%** at 4 false positives per image, **significantly outperforming** the best reported methods by over **5%**.

METHOD

- **RetinaNet** is used as it can detect objects of various sizes well. **VGG-19** is used as the backbone.
- Focal loss addresses the problem of class imbalance.
- To output a segmentation mask the network is modified similar to modifications made in **Retina U-Net**.
- **Attention gates** are used to filter feature responses propagated through skip connections.



Anchor Optimization

- Instead of using trial and error, frame the anchor optimisation problem as a maximisation of the overlap between the lesion bounding-box and the best anchor
- The ratios and scales of bounding boxes are optimised using differential evolution search algorithm.
- **Code:** bit.ly/anchor_optimization

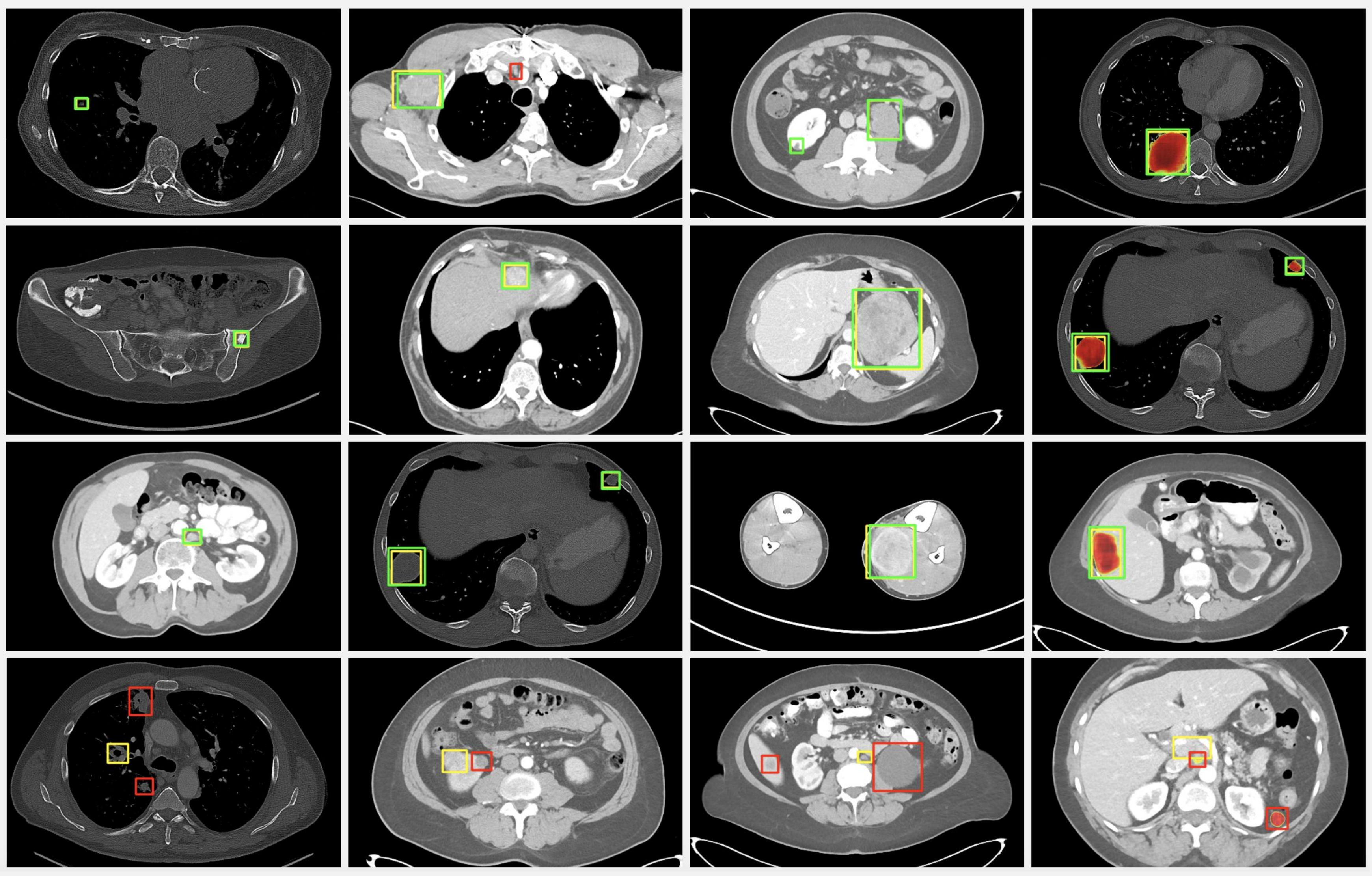
Dense Supervision

- Build a quadrilateral by consecutively connecting the four vertices of the RECIST diameters.
  - **Foreground:** Every pixel within the quadrilateral
  - **Background:** Every pixel outside of the bounding box
- Use GrabCut to generate a mask

RESULTS

Methods	0.5	1	2	4	8	16	runtime
Faster R-CNN [7]	56.90	67.26	75.57	81.62	85.83	88.74	32 ms
Mask R-CNN [3]	39.82	52.66	65.58	77.73	85.54	91.80	-
ULDor (Tang et al. [12])	52.86	64.80	74.84	84.38	87.17	91.80	-
3DCE (Yan et al. [13])	62.48	73.37	80.70	85.65	89.09	91.06	114 ms
original RetinaNet [6]	45.80	54.17	62.50	69.80	75.34	79.48	28 ms
+ anchor optimization	64.82	74.98	82.29	87.87	92.20	94.90	31 ms
+ dense supervision	70.24	78.28	85.10	90.39	93.81	96.01	39 ms
+ attention gate	**72.15**	**80.07**	**86.40**	**90.77**	**94.09**	**96.32**	41 ms

- RetinaNet with default anchor configuration is performing poorly
- **Automatic** anchor optimisation **outperforms previous state-of-the-art** by **2.34%** at 0.5 FP.
- Adding dense supervision with segmentation masks **automatically** generated from RECIST diameters significantly boosts detection sensitivity, with **5.42%** improvement at 0.5 FP.



REFERENCES

[1] Rother, C., Kolmogorov, V., Blake, A.: GrabCut: Interactive foreground extraction using iterated graph cuts. In: ACM Transactions on Graphics. vol. 23 (2004)
[2] Jaeger, P.F., Kohl, S.A., Bickelhaupt, S., Isensee, F., Kuder, T.A., Schlemmer, H.P., Maier-Hein, K.H.: Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. arXiv preprint arXiv:1811.08661 (2018)
[3] Yan, K., Bagheri, M., Summers, R.M.: 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. In: MICCAI (2018)
[4] Yan, K., Wang, X., Lu, L., Summers, R.M.: DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of Medical Imaging 5(3) (2018)

Contact: martinzlocha@google.com

Imperial College London