Application of RetinaNet in localizing cancer in digital breast tomosynthesis

Rina Ding

Motivation and goal

Motivation:

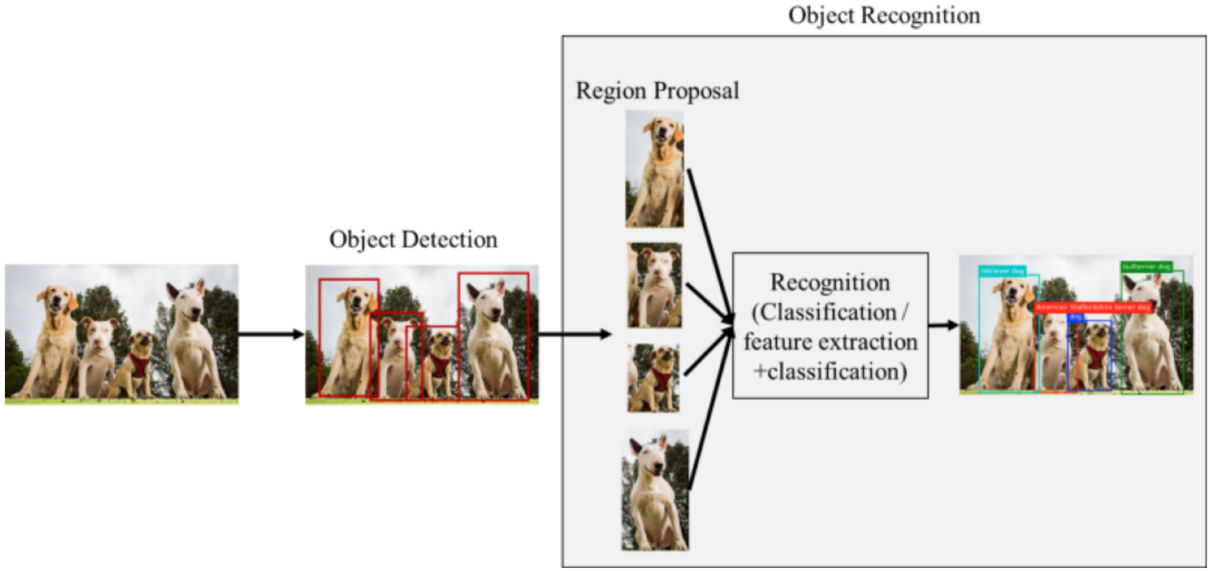
- Breast cancer screening to catch cancer early and treat it
- Traditionally mammography, but now have a new modality digital breast tomosynthesis (DBT)
- Use deep learning to improve detection and save time

Goal: to detect biopsy-proven breast lesions in DBT which include mass and architectural distortion.

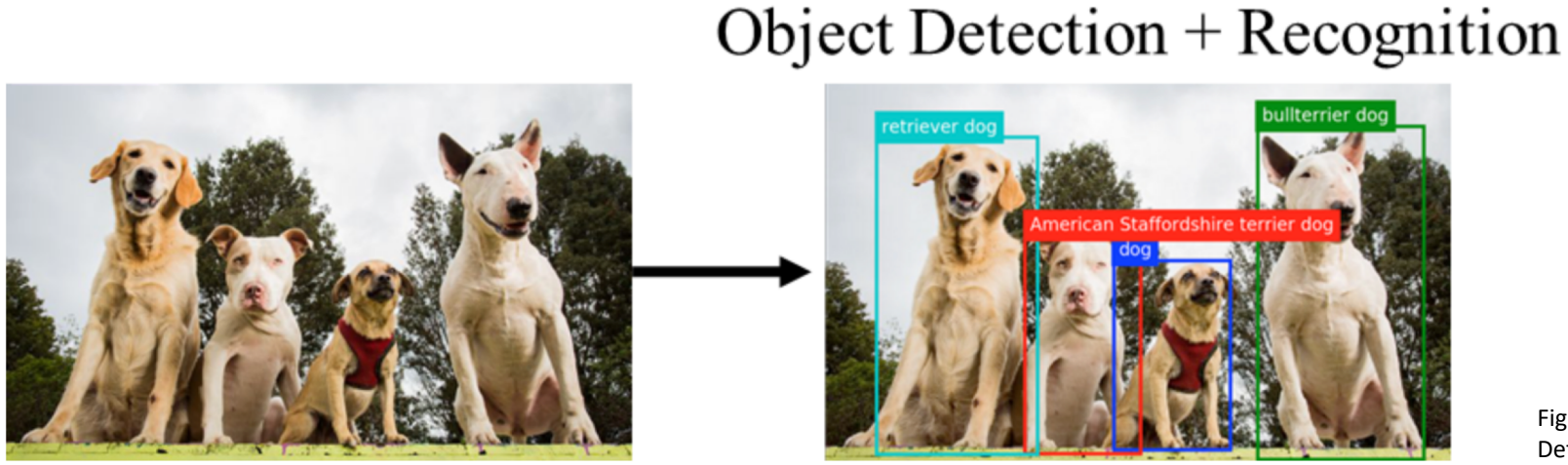
- Provide bounding box (our team)
- Classify the detected box (classification team)

One vs two stage detectors

Two-stage:
More accurate



One-stage:
Faster



Figures: https://github.com/yehengchen/Object-Detection-and-Tracking/blob/master/Two-stage%20vs%20One-stage%20Detectors.md

Focal loss

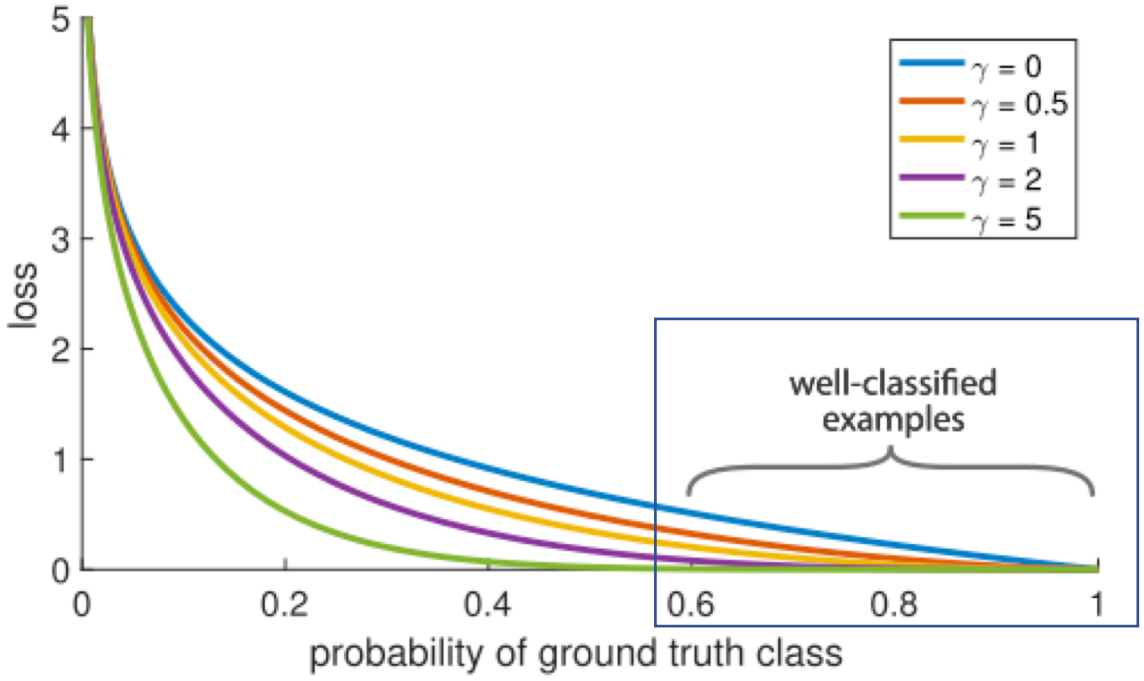
- Recall binary cross entropy loss (blue curve)

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases}$$

- Focal loss: down-weight the loss of easy background examples

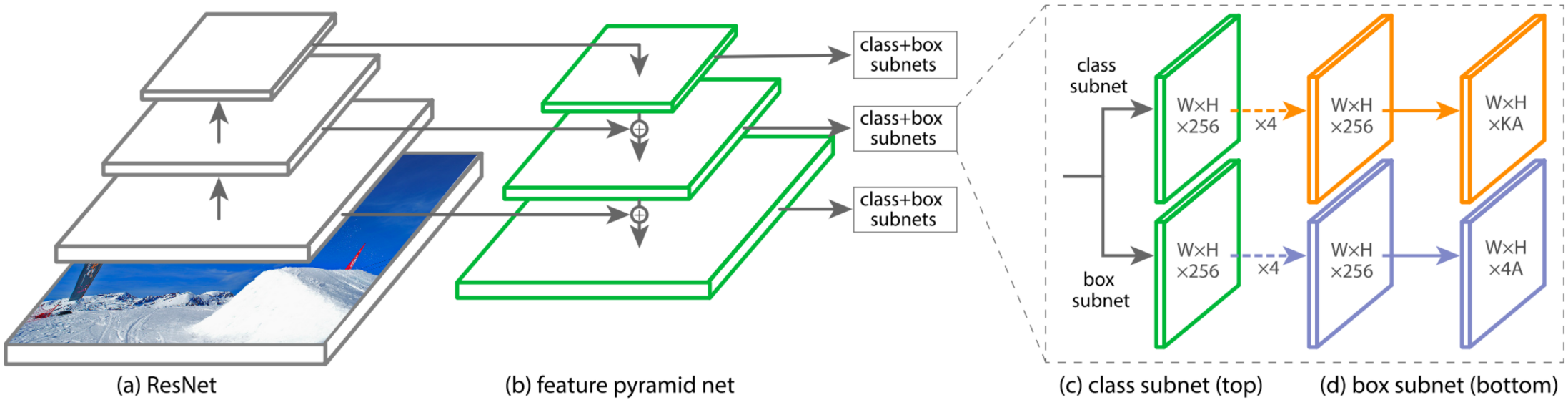
$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma} \log(p_t)$$

P→ 1, loss is down-weighted
Gamma: adjusts the rate of down-weighting easy examples



Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).

Architecture



(a) ResNet (b) feature pyramid net (c) class subnet (top) (d) box subnet (bottom)

Produces feature maps at multiple scales (deeper, smaller scales)

- 5 different levels (showing 3)
- Each level detects features at a different scale (2^{Level} difference in resolution)
- Anchor box areas: 32^2, 64^2, 128^2, 256^2, 512^2. Three aspect ratio at each level.

Box regression:
For each of the A anchors per pyramid level

- Compute offset between each anchor and assigned nearby GT box
- IoU >= 0.5: assigned

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).

Study design

General principles is to start simple:

- Use the default RetinaNet model parameters
- Run for less epochs, use less data

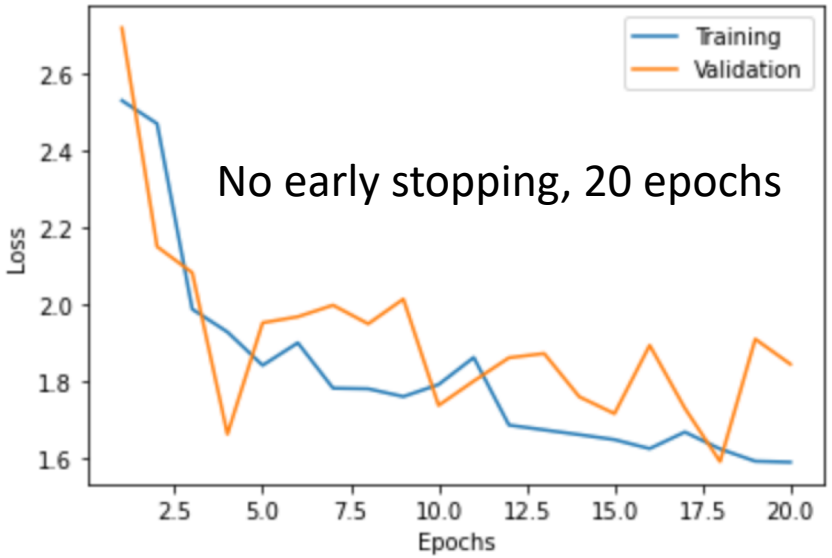
Training parameters

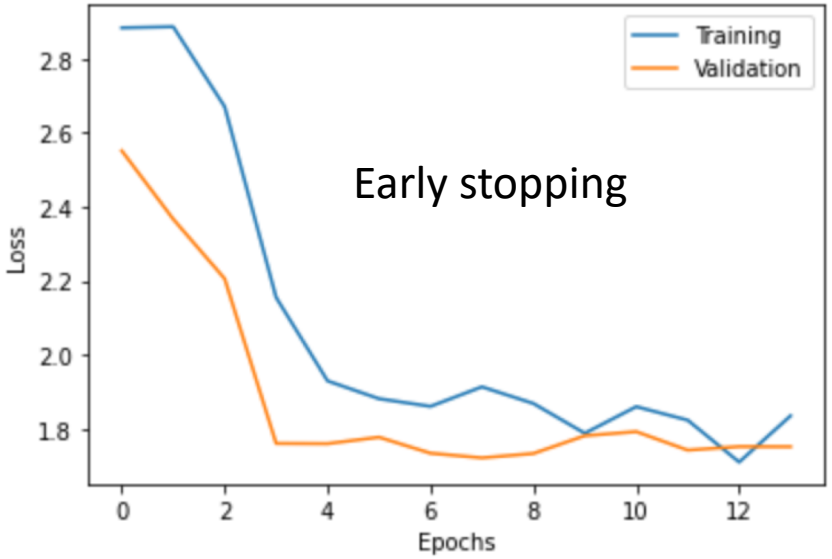
- Optimization: SGD lr = 0.001 with momentum 0.9
- Batch size = 6
- 50 epochs with early stopping
- Anchor sizes and aspect ratios: default
- Detections per image: 300

Training split

- 80%/10%/10% train validation test stratified split.
- 180 train (70 cancer), 21 validation (8 cancer), 23 test (9 cancer)

Initial results





Evaluation metric:

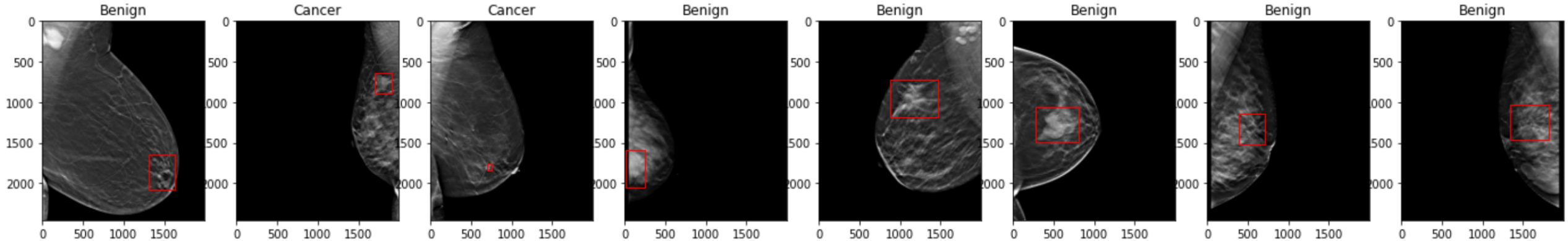
- A box is TP if
  - Distance between centers < half of annotation box diagonal length
  - Or distance between centers < 100 pixels

Result on test set (n = 23):

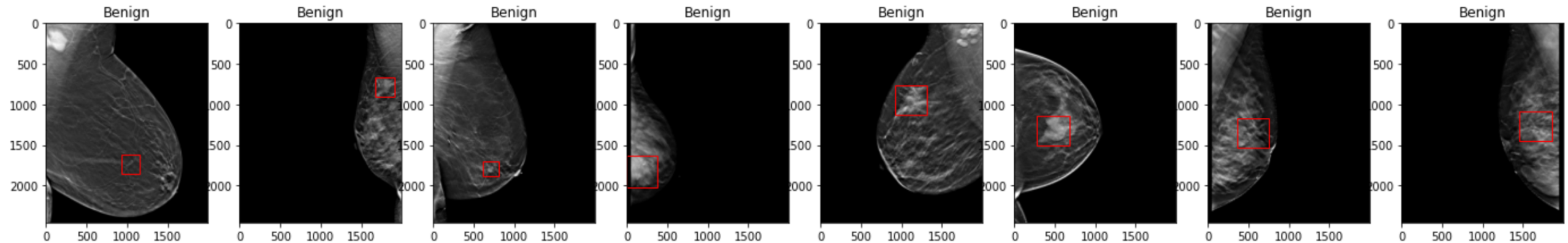
Mean sensitivity at 2 FP per slice was **0.601**

On test set (n = 23 with 9 cancer) trained with 20 epochs

Annotations



Model predictions



IoU	0, not TP	0.765, TP	0.153, TP	0.584, TP	0.526, TP	0.610, TP	0.716, TP	0.609, TP

Limitations and remaining steps

Limitations:

- Undertrained
- Not using patient level splitting
- Did not account for normal cases
- Visualization of predicted boxes only limited to one top-confidence box

Remaining steps:

- Compute the free response ROC curves at 1, 2, 3, 4 FP per slice
- Discuss with Joy and Al about the 3D bounding box inference
- Integrate with classification and segmentation team
- Heatmap vis and user interface

Main references

- Buda, Mateusz, et al. "Detection of masses and architectural distortions in digital breast tomosynthesis: a publicly available dataset of 5,060 patients and a deep learning model." *arXiv preprint arXiv:2011.07995* (2020).
- Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.
- Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Integration with classification team

1860 normal, benign, cancer cases

In the future: validation and test cases provided by the challenge.

RetinaNet

- Image ID
- Predicted boxes
- Slice index
- Class label

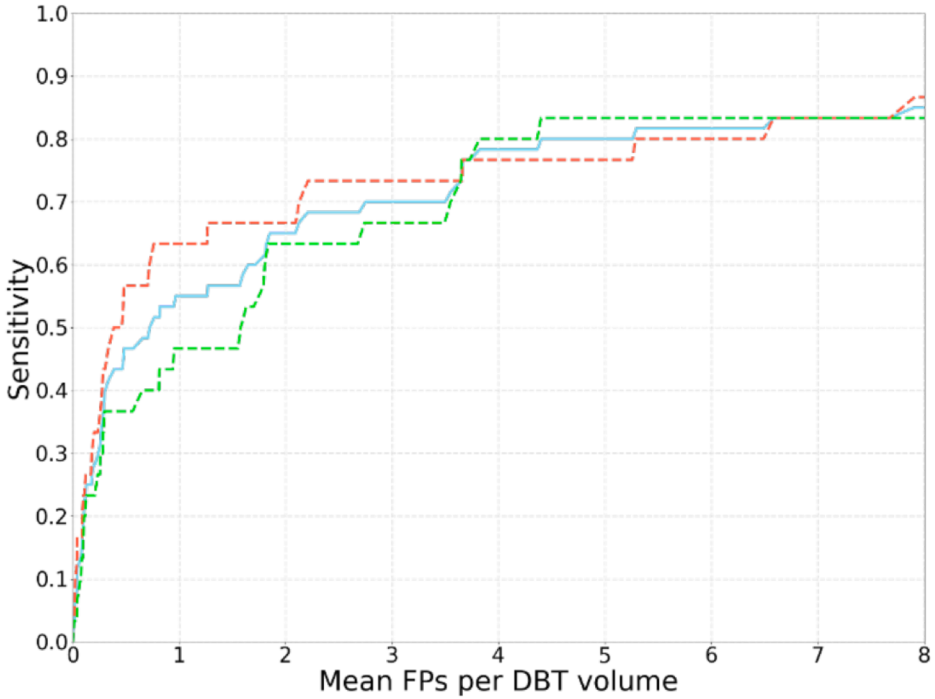
VGG or ResNet

Predicted class

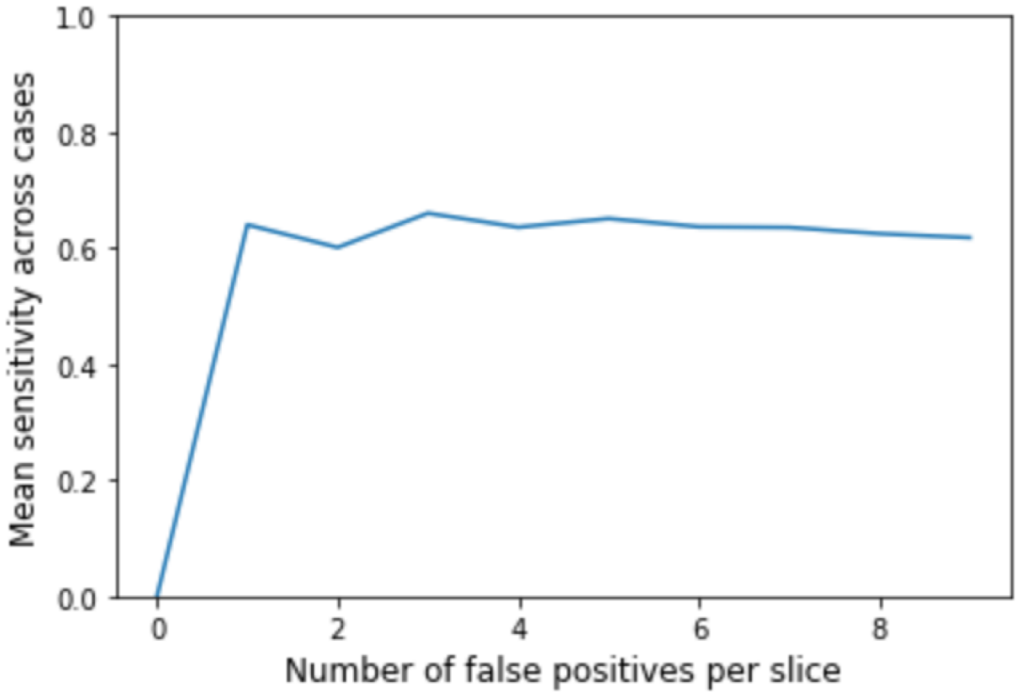
Previous version:
For each slice:

- If the predicted box has confidence > 0.3: save the box coordinates
- Else, save the box as an empty box

Current version:
3 top boxes per slice



From the challenge paper



Our RetinaNet