Deep Learning for Automated Defect Detection in TEM Images: RetinaNet

Dongxia Wu[1], Varun Sreenivasan[1], Vanessa Nilsen[1], Mingren Shen[1], and Dane Morgan[1]
[1]University of Wisconsin-Madison

Introduction

Classifying defects in metallic microstructures is done widely in the metallographic industry to understand and assess the quality and performance of metal alloys. In nuclear applications this process is critical to understand the effects of radiation and design for safe operation of nuclear reactors.

Currently defects are frequently identified through electron microscopy, and the image analysis is done by hand, which is slow, error prone and inconsistent.
Our project focused on automating this process using computer vision techniques based on convolution neural networks. The Univ. of WI team is exploring four implementations of networks, and this project is focused on one known as RetinaNet [1].

Methods and Major Activities

"RetinaNet is a single, unified network composed of a backbone and two task-specific subnetworks". The FPN backbone on top of feedforward ResNet architecture helps generate a rich, multiscale convolutional feature pyramid (2). The model also utilizes the focal loss function that helps mitigate the accuracy gap between one and two stage detectors (2).

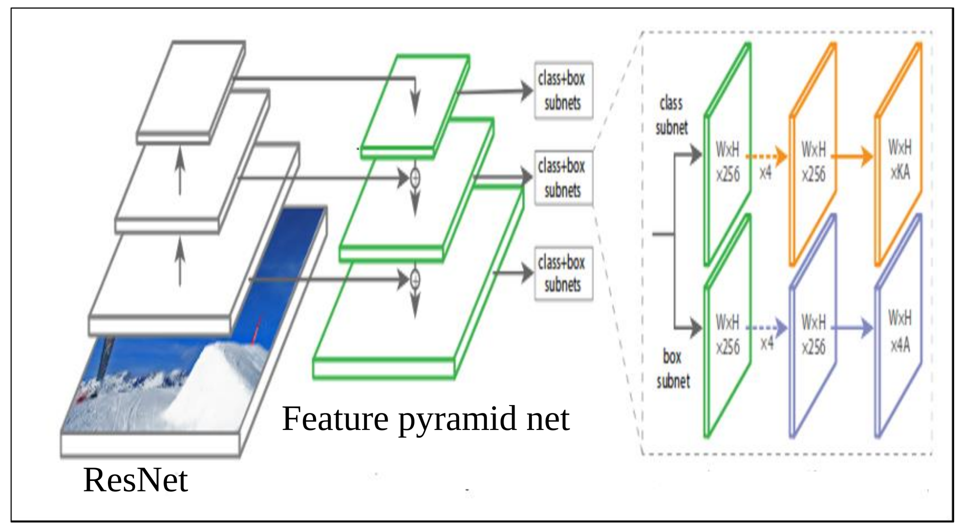

Fig 2: RetinaNet architecture (2)

Data Preparation

- A dataset of 298 micrograph images with 9566 total loop defects (3) was prepared by two researchers experienced in loop-labelling (set as "ground truth")
- The dataset was split into a training set and testing set at a fixed 9:1 ratio. After splitting, a training set of 270 images and test set of 20 images were obtained (3).

Training Model

- Multiple RetinaNet models were created by varying hyperparameters such as number of epochs, number of layers, batch size, and learning rate.
- Optimal parameters values were obtained by testing the various models.

Testing Model

- Recall and Precision metrics were used to quantify the performance of the model.
- The best model was selected by determining the model that maximizes recall and precision values and minimizes the difference between the two values.

Funding & Acknowledgements

Funding by Schmidt Futures grant number MSN214683 as part of Citrine Informatics' NextGen Fellowship

Results

Model Summary

Recall: 74% Precision: 89% F1 Score: 0.81

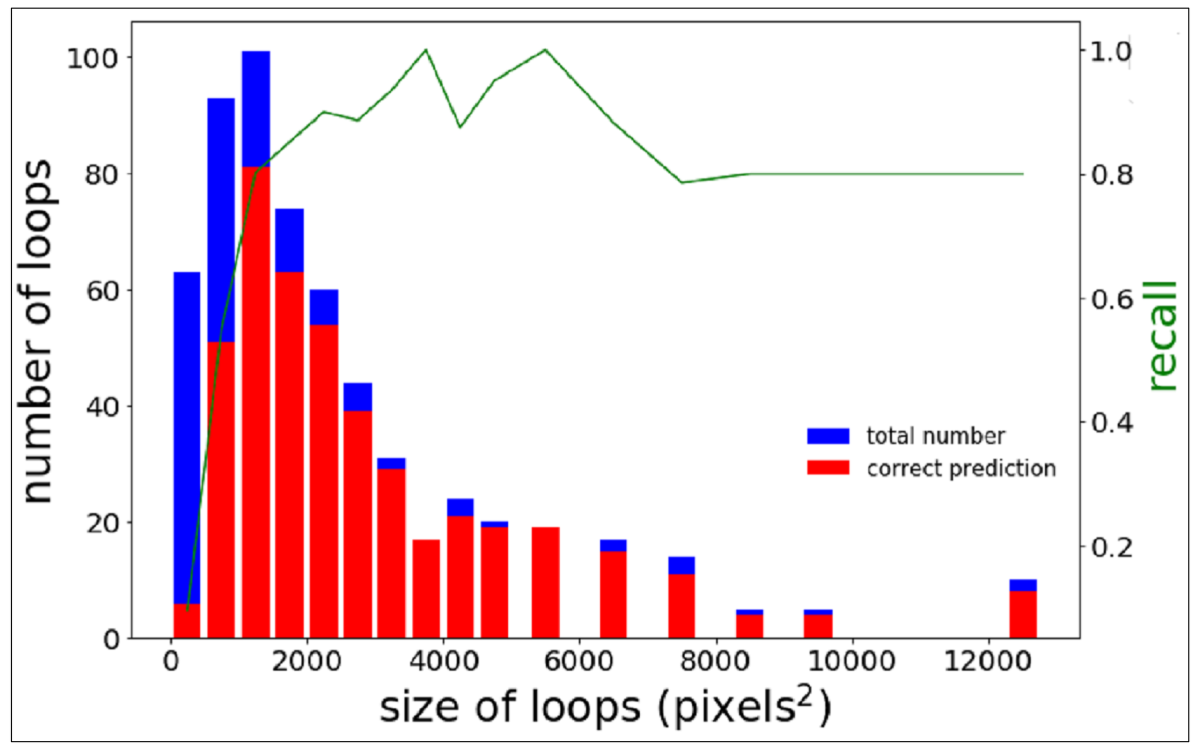

Fig 3: Recall vs Size of Loops

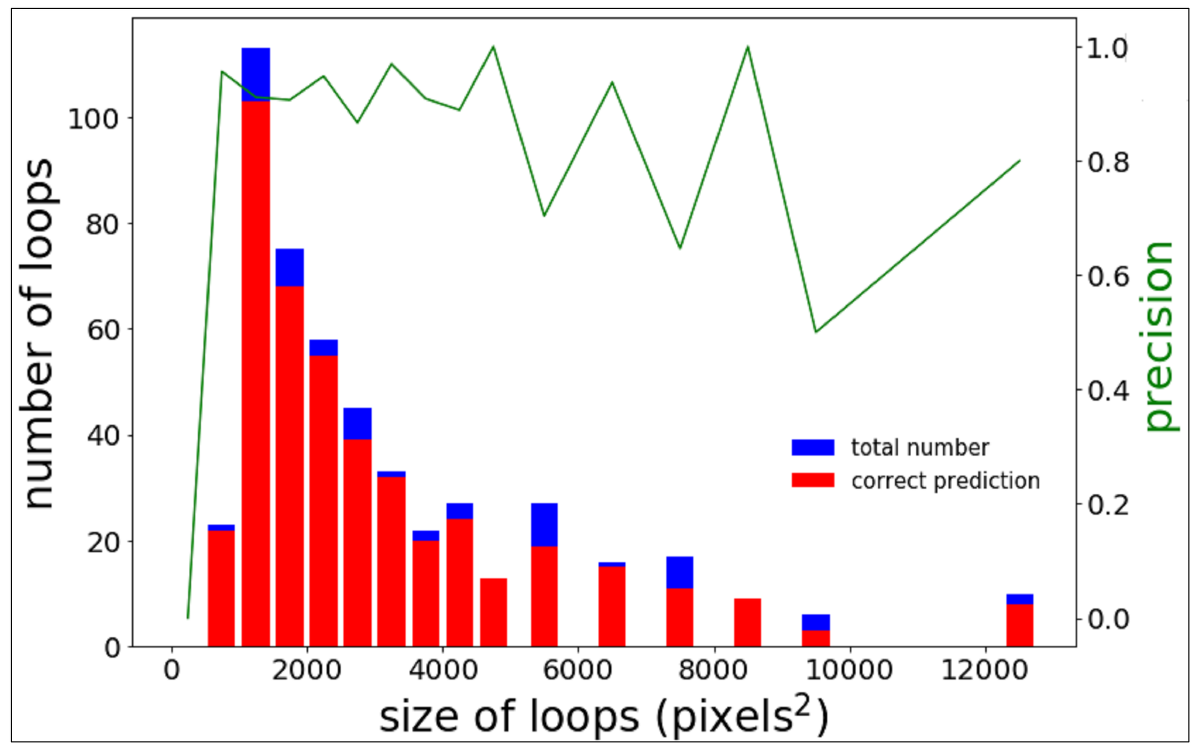

Fig 4: Precision vs Size of Loops

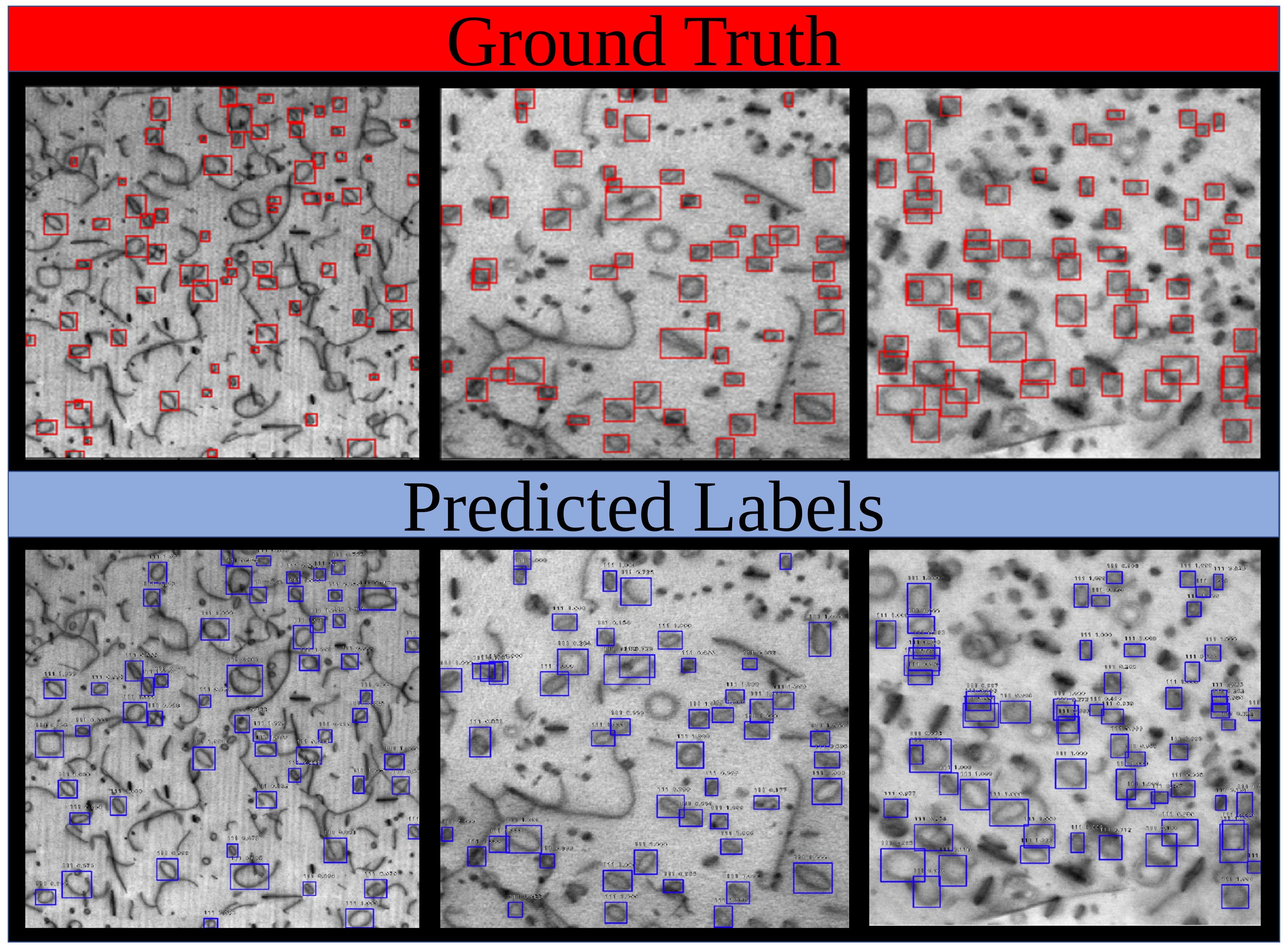

Fig 5: Ground Truth Labels vs Predicted Labels

RECALL AND PRECISION TRENDS

- Fig 3's 0-500 pixels^2 group suggests the model doesn't work well for small defects.
- Recall improves sharply in the 0-1000 pixels^2 region because the model is able to better discern defects.
- Fig 4 suggests that there is no coherent relationship between loop size and precision.
- Precision seems to be affected by the model's inability to effectively differentiate between 111, 100, and pre-existing defects

Conclusions

The model performs reasonably well in terms of recall and precision. However, there is still room for improvement:

- The model has an average recall of 68% and average precision of 85% on the test set.
- Relatively low recall stems from inability to detect small defects.

Future Work

The model still has further scope for improvement, especially recall. The following methods can be used:

- Preparing a larger dataset comprising of more micrographs of irradiated metals, focusing on smaller defects.
- Train the model for a longer duration.

References

1. Lin, et al. "Focal Loss for Dense Object Detection." *[1402.1128] Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition*, 7 Feb. 2018, arxiv.org/abs/1708.02002.
2. Fizyr. "Fizyr/Keras-Retinanet." *GitHub*, 18 July 2018, github.com/fizyr/keras-retinanet.
3. Li, Wei, Kevin G. Field, and Dane Morgan. "Automated defect analysis in electron microscopic images." npj Computational Materials 4.1 (2018): 36