CS229: Machine Learning

Steel defect detection with high-frequency camera images

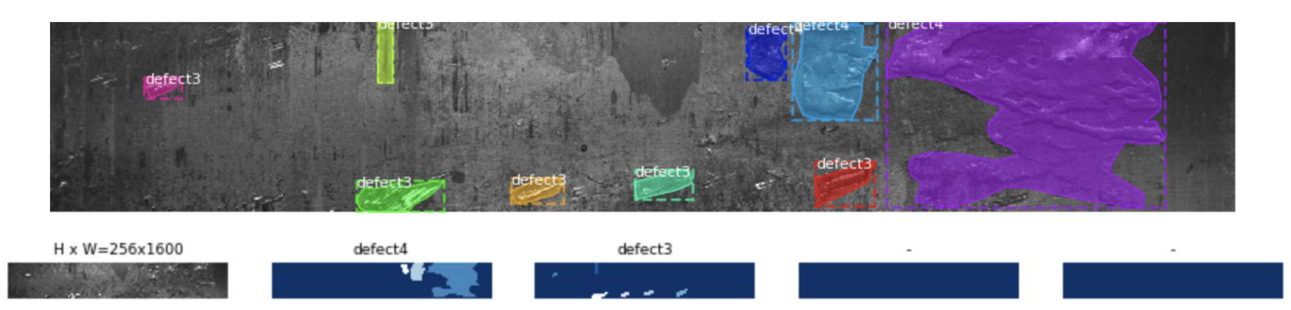
Zhuoran Ma, Xinrui Wang, Mengxi Li

Background

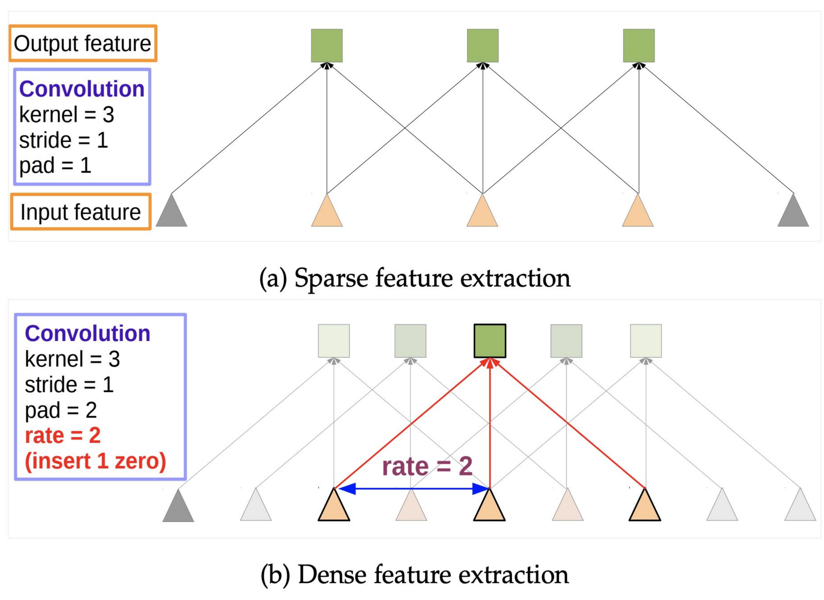
Steel is one of the most important building materials of modern times. And the production process of flat sheet steel is especially delicate. From heating and rolling to drying and cutting, several machines touch flat steel by the time it's ready to ship. Before delivering the product, steel sheets need to undergo careful inspection to avoid defects and thus localizing and classifying surface defects on a steel sheet is crucial. Hence, automating the inspection process would accelerate the steel sheet production.

Dataset Explanation

The images provided are of size 1600 × 256 × 1 and totals 12568 counts. Out of all the images, 5902 images are are with defects and 6666 images are without defects. There are four labels of defects 1, 2, 3, and 4. A Sample image with instance-level defects is shown below.



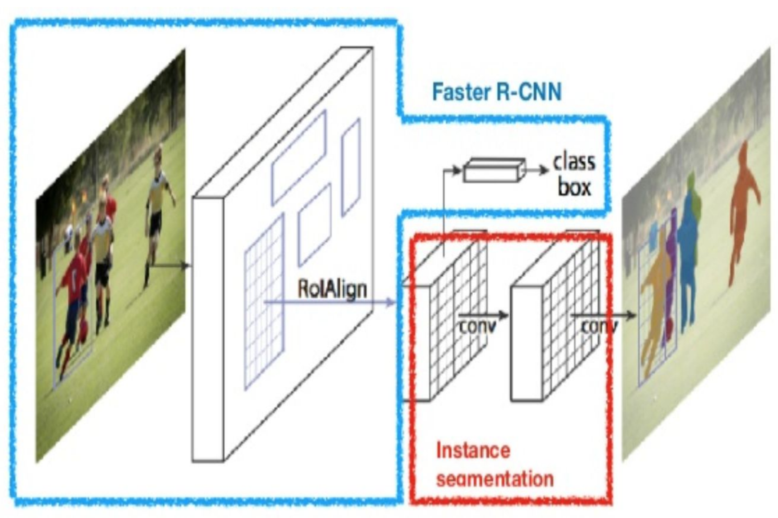
Method



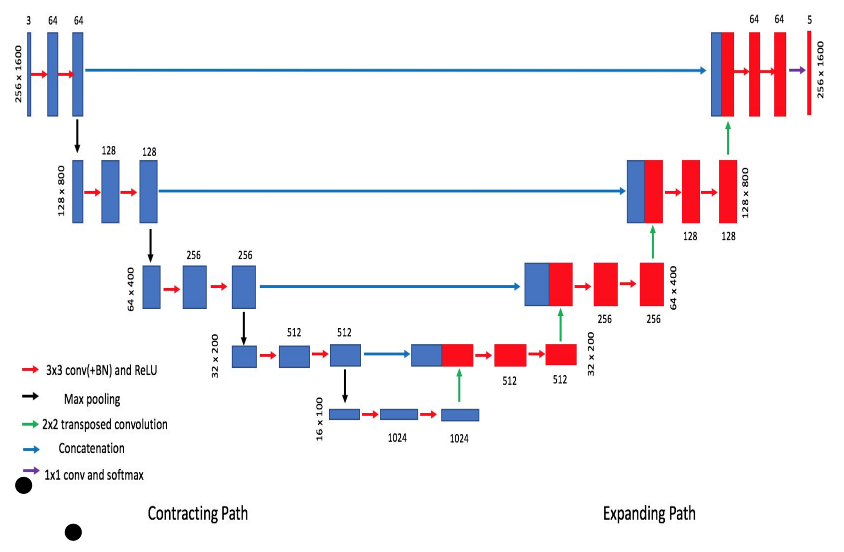
(a) Sparse feature extraction

(b) Dense feature extraction

Xception
We removed the last fully-connected layers of the original Xception model[1] and followed the idea of "atrous convolution" as used in deeplab[2], which enlarges the feature map by linear interpolation.



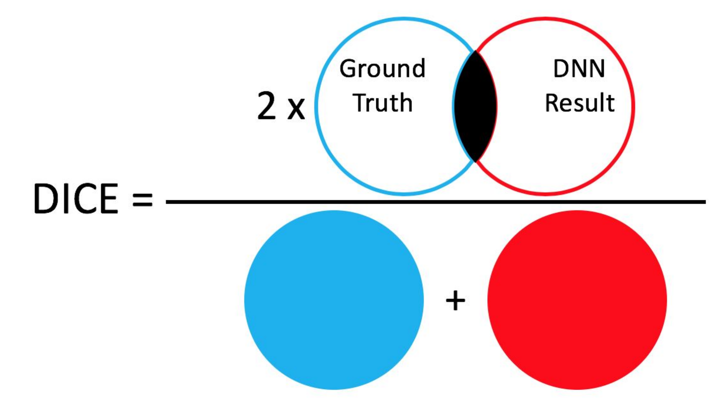
Mask R-CNN
Mask R-CNN algorithm is an instance segmentation algorithm[3], which identifies object outlines at the pixel level. It consists of two stages: the first stage scans the image and generates proposals and the second stage classifies the proposals and generates bounding boxes and masks



U- Net
U-Net is a semantic segmentation algorithm[4]. It contains a contracting path, which captures context of an image, and a symmetric expanding path, which enables precise localization.

Benchmark: Dice coefficient

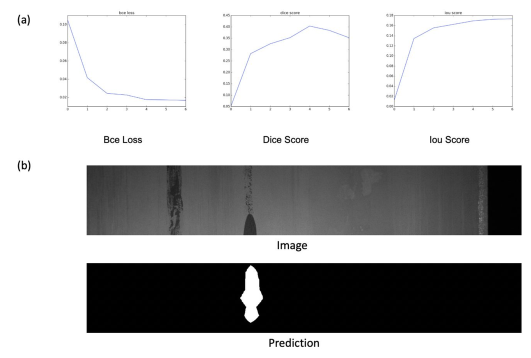
The Dice coefficient can be used to compare the pixel-wise agreement between a predicted segmentation and its corresponding ground truth.where X is the predicted set of pixels and Y is the ground truth. The Dice coefficient is defined to be 1 when both X and Y are empty. The formula is given by:



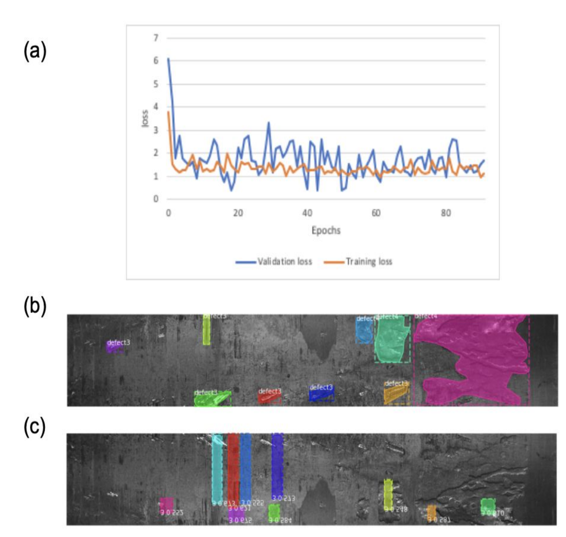
$$Dice(X, Y) = \frac{2 \cdot |X \cap Y|}{|X| + |Y|}$$

Experiments & Result

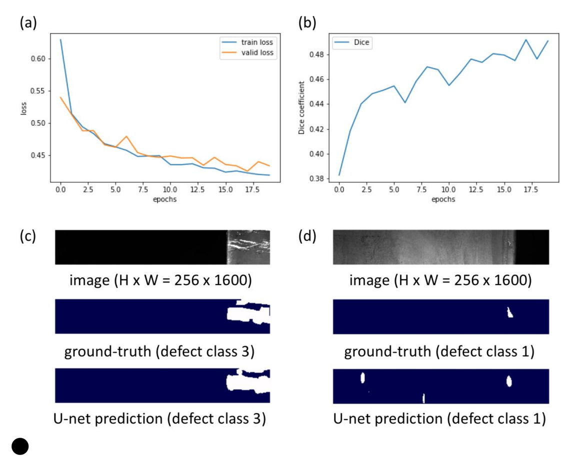
	Xception[1]	Mask R-CNN[3]	U-Net[4]
positive dice	0.369	0.083	0.492
negative dice	0.978	0.898	0.977



The results of Xception model[1]. (a) Binary cross-entropy loss, dice loss and dice score during training. (b) Original image along with the prediction output by Xception model.



The figure on the left shows the results of the Mask R-CNN[3]. (a) Loss of the training and the validation sets. (b) A sample labeled image containing instances of defect. (c) The defect instances prediction result generated by Mask R-CNN. The training loss quickly drops to around 1.1 and remain steady throughout the training process. And the loss for Region Proposal Network localizing objects dominates the loss.



The results of the U-Net[4]. (a) Loss of the training and the validation sets. (b) Average Dice score of the validation set. (c) Prediction result for an image containing defects of class 3. (d) Prediction result for an image containing defects of class 1.

Discussion/Future Work

All of the three methods suffer from imbalanced dataset to some degrees. For example, the training samples are insufficient for the Mask R-CNN[3] so that it overfits severely, the predictions of U-Net[4] and Xception[1] could mainly capture defect type 3 because the majority of defect data belongs to type 3. An interesting future direction would be designing an efficient algorithm to train a neural network with imbalanced/insufficient data.

References

[1] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[2] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence 40.4 (2017): 834-848.

[3] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[4] Ronneberger O., Fischer P., and Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015