Does using images in the frequency domain improve the performance of an object counter?

Introduction

Traditional CNNs: bounding box to point at object locations.
CNN used for this paper: single-point to denote object center.

$$d_{AH}(X,Y) = \frac{1}{|X|}\sum_{x\in X}\min_{y\in Y} d(x,y) + \frac{1}{|Y|}\sum_{y\in Y}\min_{x\in X} d(x,y)$$

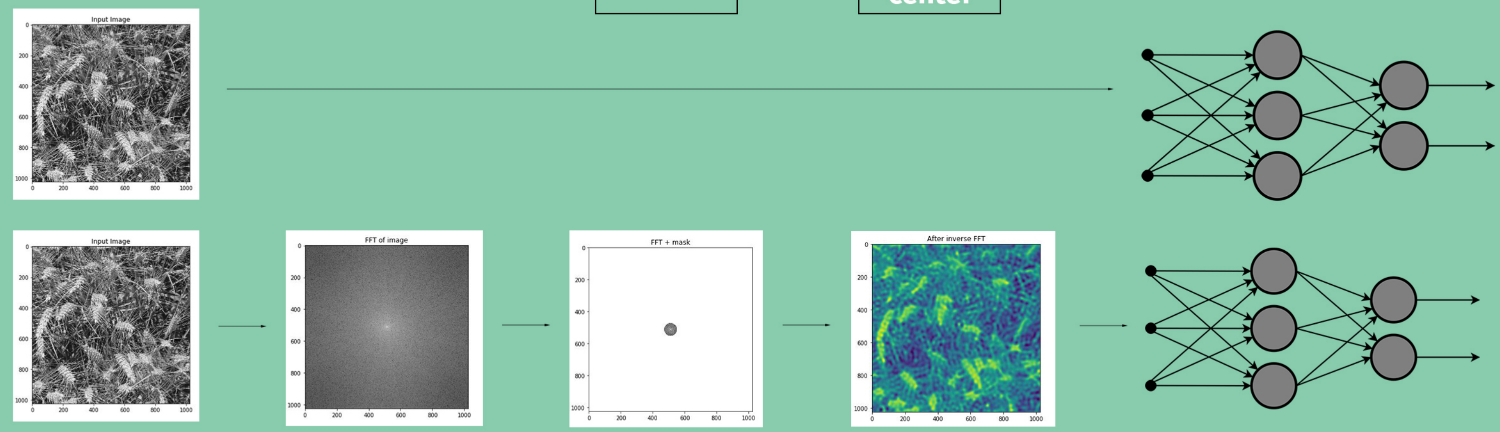
Kaggle dataset: Problematic dataset
- Crowded wheat heads
- Overlapping elements
- Wheat heads taken at different regions of the world.

Goal of the research: does filtering frequency information in images as a pre-processing step improve the accuracy of our single-point object locator when trained on the Kaggle Global Wheat dataset?

Methodology

Image labels (wheat head locations)

Bounding box → Single pixel center

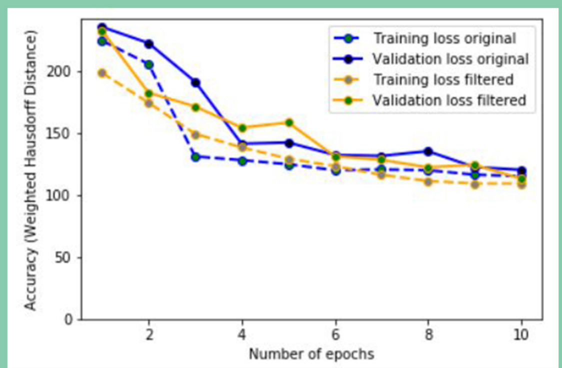


Evaluation metrics for counting and locating objects; does the neural network perform better on filtered images?

Precision, recall, RMSE

Exclude 120, threshold 0.2 * 10e-7, high pass filter: analysis of 115 images

Results



	Precision (%)	Recall (%)	RMSE
Original images	16.57	18.26	9.11
Filtered images	**18.11**	**18.30**	**9.02**

Relative performance: FFT did not bring significant improvements in performance.
Absolute performance: poor in both scenarios.

Discussion

- Wheat heads are better represented by bounding boxes than by single-pixel centers.

Average height: 76px
Average width: 84px

Solution: modified average Hausdorff distance.

$$d_{AH}(X,Y) = \frac{1}{|X|}\sum_{x\in X}\min_{y\in Y} \frac{d(x,y)}{A_x} + \frac{1}{|Y|}\sum_{y\in Y}\min_{x\in X} \frac{d(x,y)}{A_y}$$

- Other limitations due to FFT parameters

Conclusion

- No significant improvement through FFT.

- Most important novely of the neural network was its biggest hinderence.

- A loss function that is less distance such as cross-entropy or the modified average Hausdorff distance proposed should be used.

Supervised by Yancong Lin & Silvia Pintea | d.z.rogmans@student.tudelft.nl