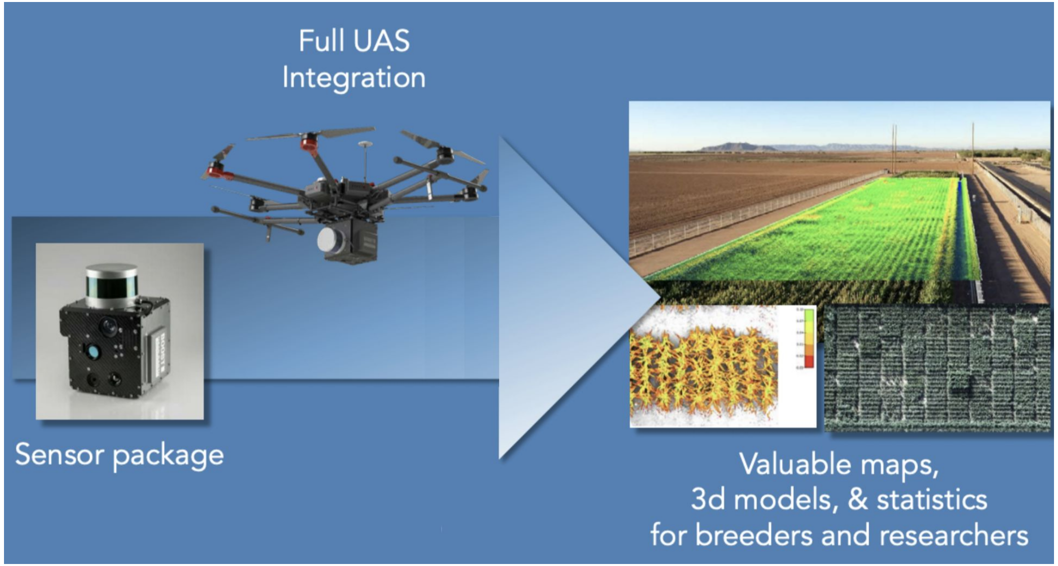
Carnegie Mellon University

Visual Inspection for Aircraft & Power Lines

Chang Gao & Anshuman Majumdar
Master of Science in Computer Vision, CMU

Summary



Motivation

Automate asset inspection with sensor data

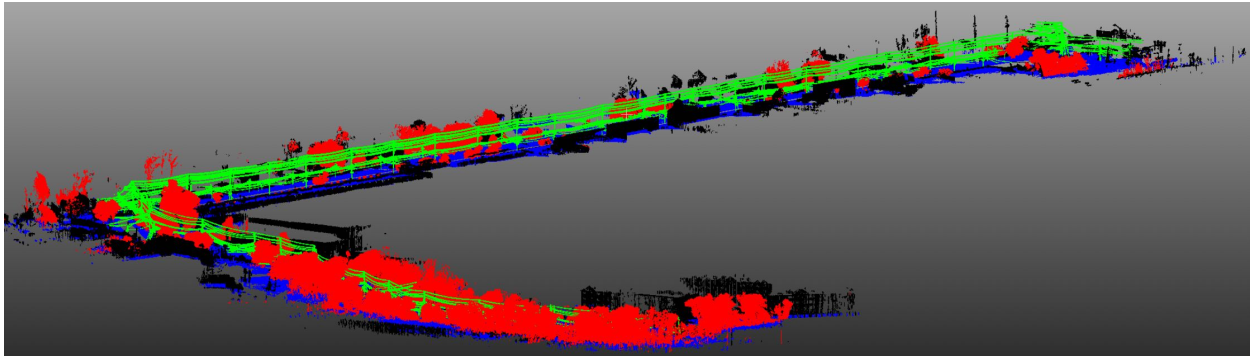
Problem

Powerline inspection & aircraft defect detection

Solution

Semantic segmentation & object detection

First Semester: Powerline Inspection



3D Semantic Segmentation Results

Green = Power lines

Red = Trees

Black = Houses

Blue & **Black** = Ground

New task: Aircraft Defect Detection

Sample defect images with bounding boxes

Carnegie Mellon University

Aircraft Defect Detection

1. Problem overview
2. Dataset study
3. Proposed approaches
4. Experimental results
5. Conclusion

Carnegie Mellon University

Problem Overview

For airplane companies and airport managers

- Planes need to be inspected before taking off
- It takes a long time and many workers

Proposed solution

- Just let drones fly around and take pictures
- Perform defect detection on these pictures

We now have the dataset collected from pictures taken around planes

Carnegie Mellon University

Dataset

Dataset

Key findings

Approach: Object Detection

Why object detection?

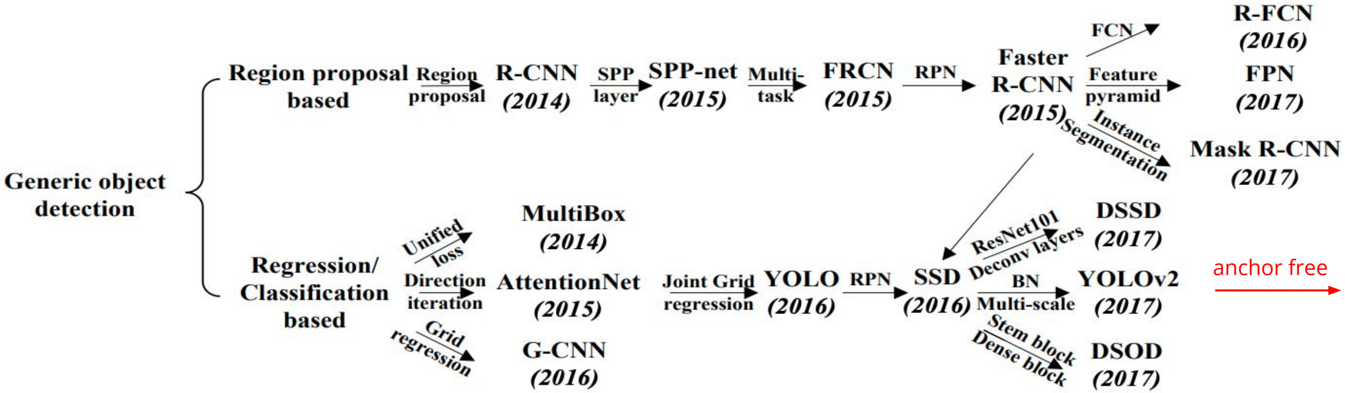
- Bounding boxes already provided as ground-truth
- Direct approach to solve the problem

What type of object detection?

- Speed & Feasibility to train and evaluate -> one stage over two stage
- Variant ratio of bounding box sizes -> anchor-free models

Literature Review

Object detection roadgraph

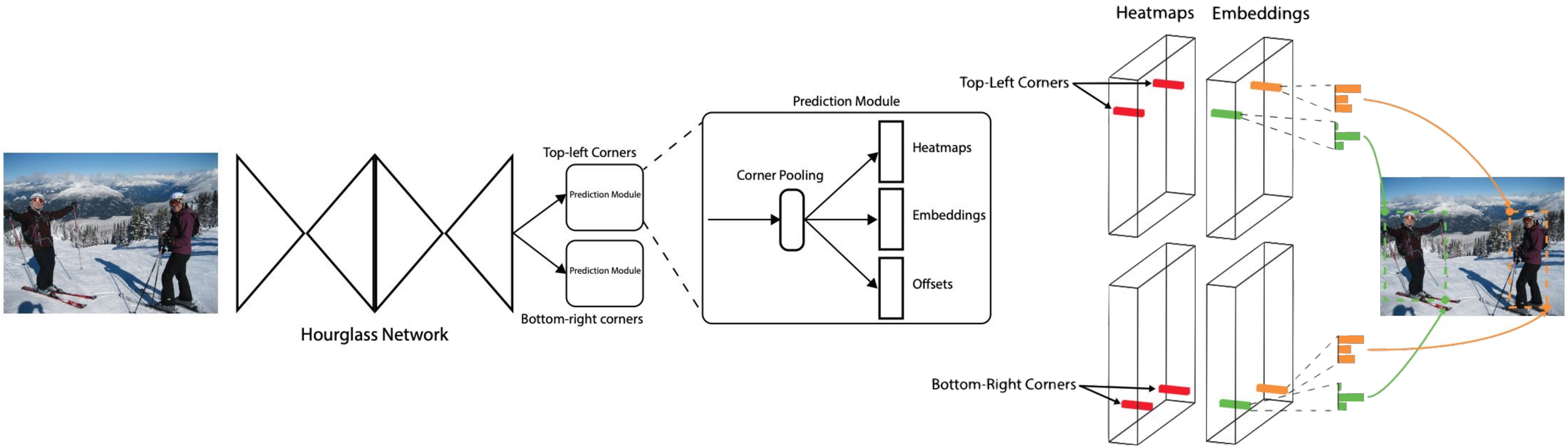


anchor free →

CornerNet (2018)
CenterNet (2019)
FCOS (2019)

Literature Review: CornerNet

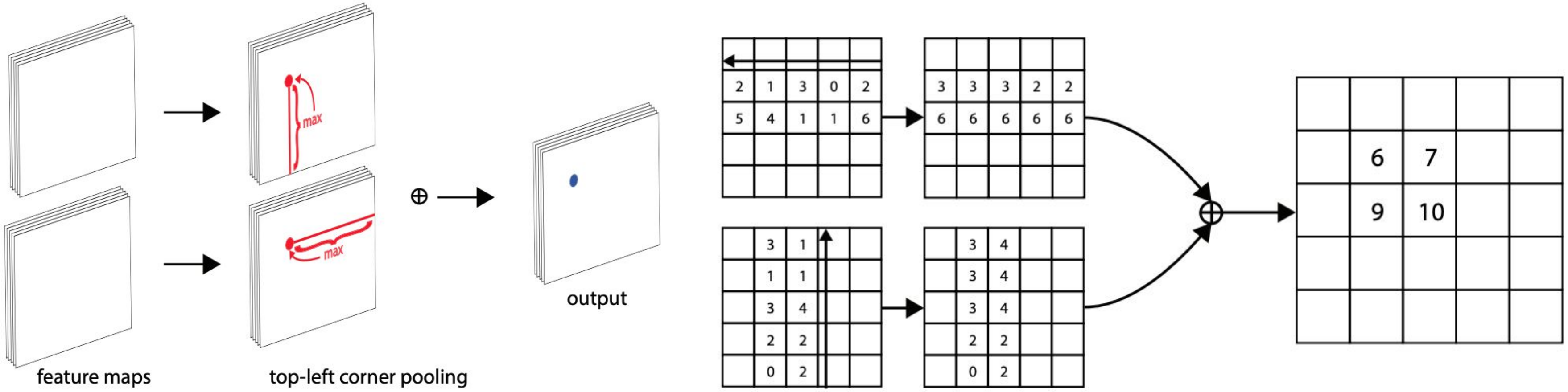
CornerNet Architecture



Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Literature Review: CornerNet

Corner Pooling Module

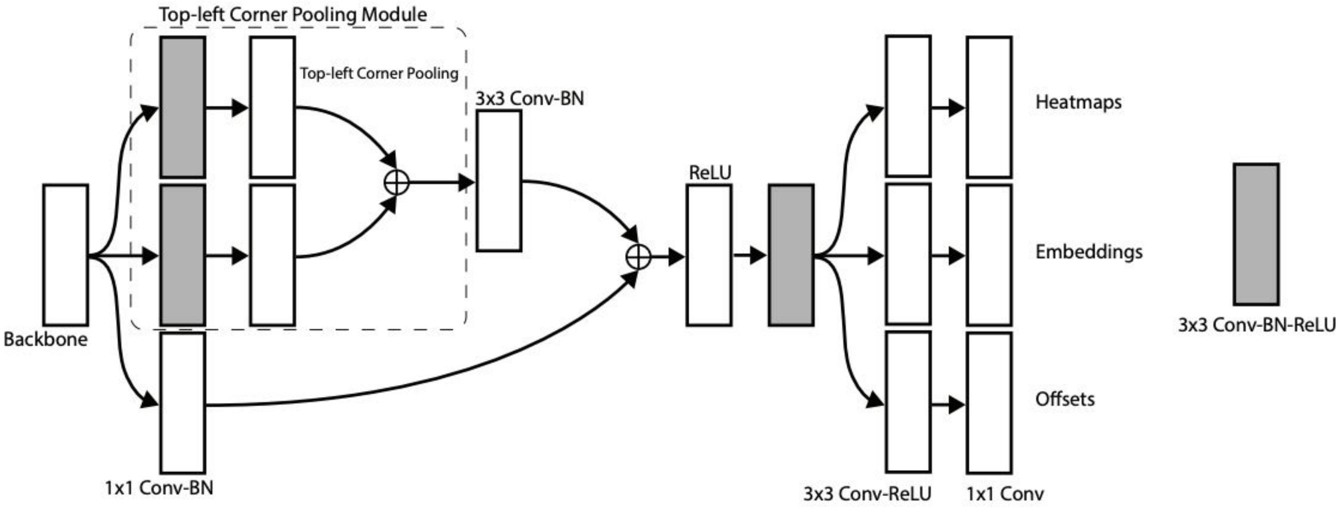


Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Carnegie Mellon University

Literature Review: CornerNet

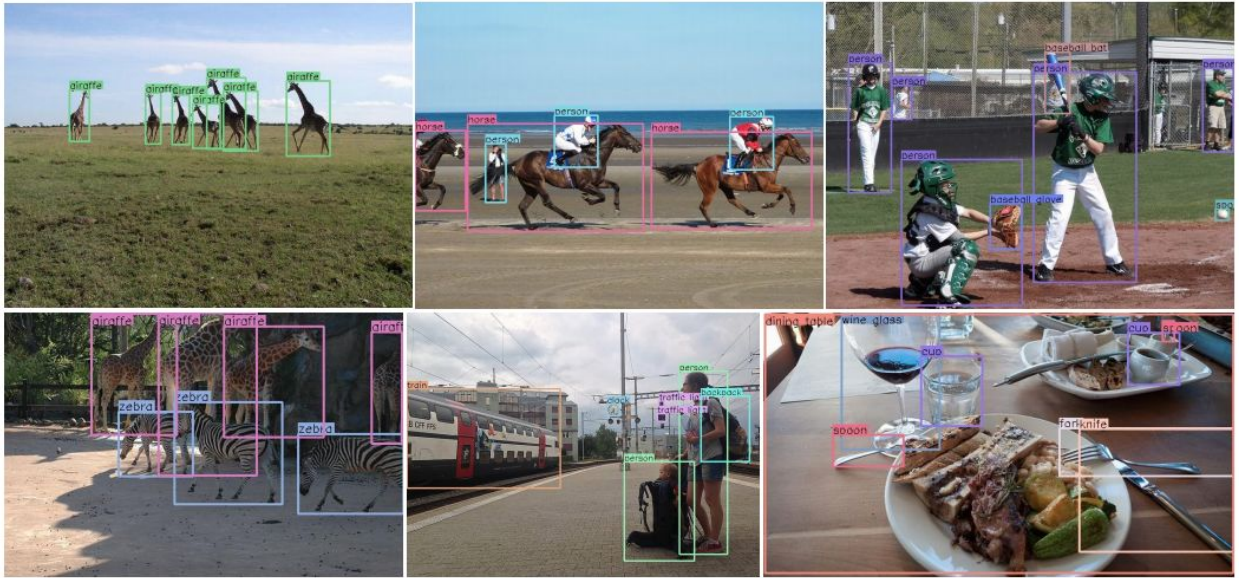
Bounding Box Prediction Module (Top-left branch)



Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Literature Review: CornerNet

Qualitative Examples on MSCOCO



Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Literature Review: CenterNet

Improvements from CornerNet:

- Center pooling module: inherits the functionality of RoI pooling
- Cascade corner pooling: perceives internal information

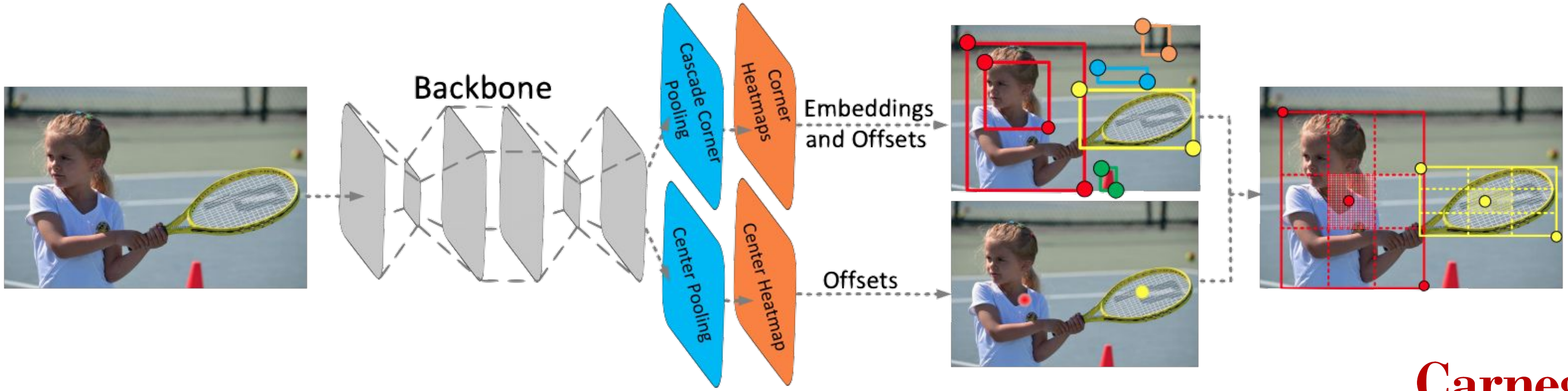
Duan, Kaiwen, et al. "CenterNet: Keypoint Triplets for Object Detection." ICCV 2019.

Carnegie Mellon University

Literature Review: CenterNet

CenterNet Architecture

- Similar to CornerNet, a pair of detected corners and the similar embeddings are used to detect a potential bounding box. Then the detected center keypoints are used to determine the final bounding boxes.

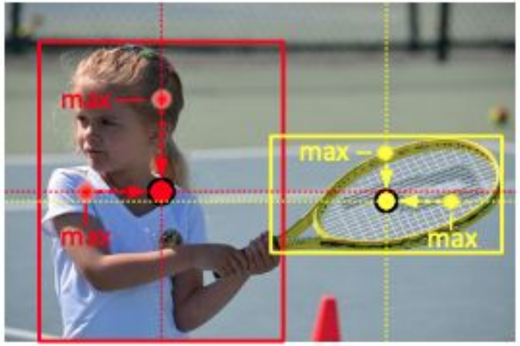


Duan, Kaiwen, et al. "CenterNet: Keypoint Triplets for Object Detection." ICCV 2019.

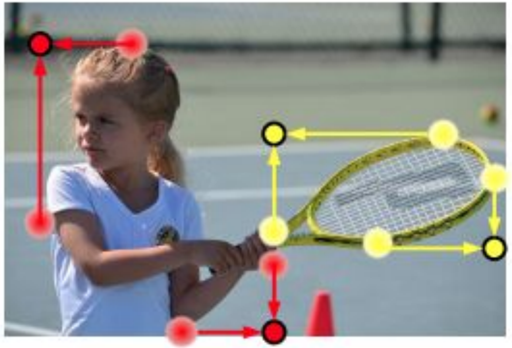
Carnegie Mellon University

Literature Review: CenterNet

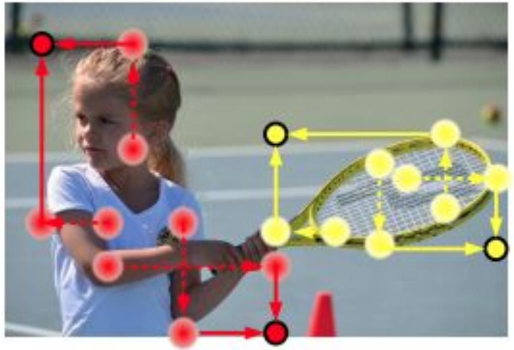
Center pooling & cascade corner pooling module



Center Pooling



Corner Pooling



Cascade Corner Pooling

Duan, Kaiwen, et al. "CenterNet: Keypoint Triplets for Object Detection." ICCV 2019.

Carnegie Mellon University

Literature Review: CenterNet

Quantitative Results on MSCOCO (Apr 2019)

		Average Precision
Two-stage Models	Mask R-CNN	39.8
	PANet (SOTA)	47.4
One-stage Models	RetinaNet800	39.1
	CornerNet	42.1
	CornerNet-Saccade	43.2
	CenterNet-104	47.0

Duan, Kaiwen, et al. "CenterNet: Keypoint Triplets for Object Detection." ICCV 2019.

Network Training & Evaluation

- We use CenterNet-52 as our network structure (52-layer Hourglass Network)
- Multi-scale training
- batch size of 4 on each of 2 Nvidia 1080 Ti GPUs
- 4K as training set and 1K as evaluation set

Experimental Results: Quantitative

Training set (NMS threshold = 50%, proposal confidence = 30%):

- mAP = 92.2301%
- Foreground overall recall [1] = 97.7200%
- Foreground overall precision [2] = 71.7203%
- CenterNet-52 successfully converged on the training set

[1] A ground-truth bounding box is considered recalled if it is predicted as any foreground class.
[2] A predicted bounding box is considered correct if any foreground ground-truth bounding box overlaps it larger than a threshold (default is 50%).
* We also define image level recall and precision in following slides, which denotes an image as positive if there is at least a bounding box provided or predicted.

Carnegie Mellon University

Experimental Results: Quantitative

Validation set (NMS threshold = 50%, proposal confidence = 30%):

- mAP = 18.1666%
- Foreground overall recall = 49.3571%
- Foreground overall precision = 49.3860%
- Severe overfitting effect observed!

Experimental Results: Quantitative

Validation set (NMS threshold = 50%, proposal confidence = 30%):

- Class-wise confusion matrix
- Findings: P/R among each class is good, P/R against background is terrible

Experimental Results: Qualitative

Why overfitting is so severe?

1. Ground-truth labels are actually ill-posed

Ground-truth Predicted

Experimental Results: Qualitative

More ill-posed or bad bounding boxes and labels

Experimental Results: Qualitative

We sometimes, not too rarely, performs better than the ground-truth!

Ground-truth

Predicted

Experimental Results: Qualitative

Why overfitting is so severe?

1. Ground-truth labels are actually ill-posed
    - Possible solution: Lower overlapping threshold
    - With overlapping threshold changed from 50% to 30%
    - mAP = 18.1666% -> 25.6855%
    - Foreground overall recall = 49.3571% -> 55.8153%
    - Foreground overall precision = 49.3860% -> 55.7909%

Carnegie Mellon University

Experimental Results: Qualitative

Why overfitting is so severe?

2. Too few training data (only ~4000 images)
   - It is really hard to train a supervised detection model given limited data
   - We apply anti-overfitting techniques, including
     1. random rescaling, random cropping, color jittering
     2. gaussian bump of corner/center ground-truth
     3. class-balanced weights for losses [1]
     4. download similar images online

[1] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." CVPR 2019.

Carnegie Mellon University

Experimental Results: Qualitative

Download similar images online

- Search for images with captions “airplane close up”, “aircraft zoom in”, etc
- We tried our best and downloaded ~1500 images, yet not many of them are in the same domain as the provided dataset, and some of them look completely different

Carnegie Mellon University

Experimental Results: Qualitative

Download similar images online

- Add around online 700 images to the training set, with no ground truth bounding boxes
- Add around online 700 images to the test set, with no ground truth bounding boxes

Carnegie Mellon University

Experimental Results: Qualitative

Why overfitting is so severe?

2. Apply more data augmentation techniques:
   - mAP: 25.6855% -> 26.4279%
     - Adding online data is the main contributor
   - To consider the online data, we also evaluated image-level recall and precision
   - On the test set with both original and online images
     - Image-level recall = 96.5368%
     - Image-level precision = 95.0554%
   - Although the domain is likely different, we perform fairly good on detecting whether a region has defects or not

Carnegie Mellon University

Experimental Results: Qualitative

Why overfitting is so severe?

3. Model complexity is too high
   - Reduce channel sizes and layers
   - Apply dropouts
   - Result: None of these methods work, probably because given too few images, we do not have enough features to learn

More Qualitative Examples

Conclusion

1. On a new image domain with limited data, our detection model performs fare on detecting bounding boxes and performs well on detecting image-level defects
2. The provided data is ill-posed and is hard to learn itself, and we sometimes perform better than the labeler
3. In future, one could possibly use weakly-supervised methods to perform segmentation instead of bounding box detections to reduce the ambiguity in the labels

Carnegie Mellon University