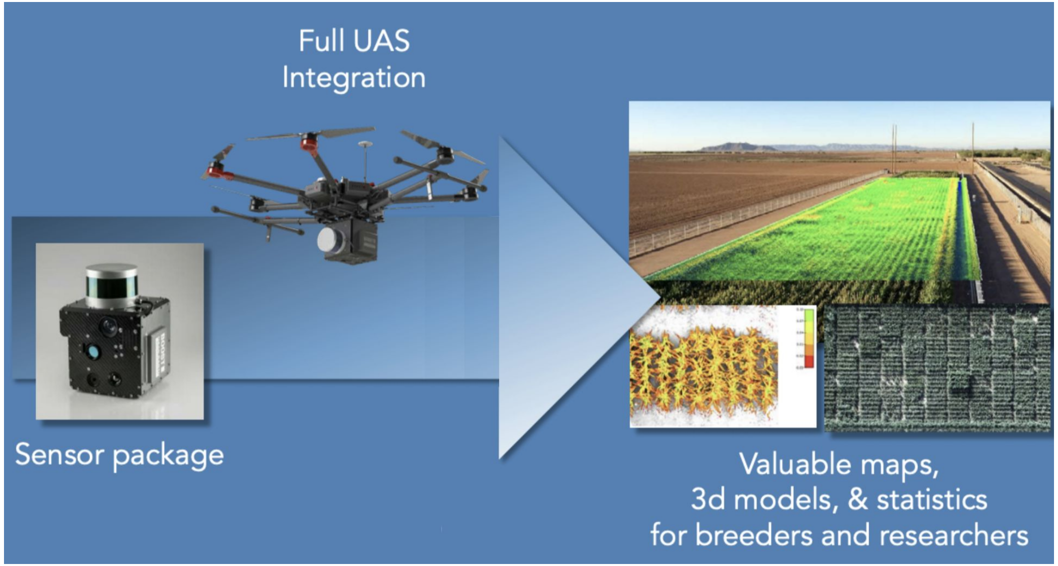
Carnegie Mellon University

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

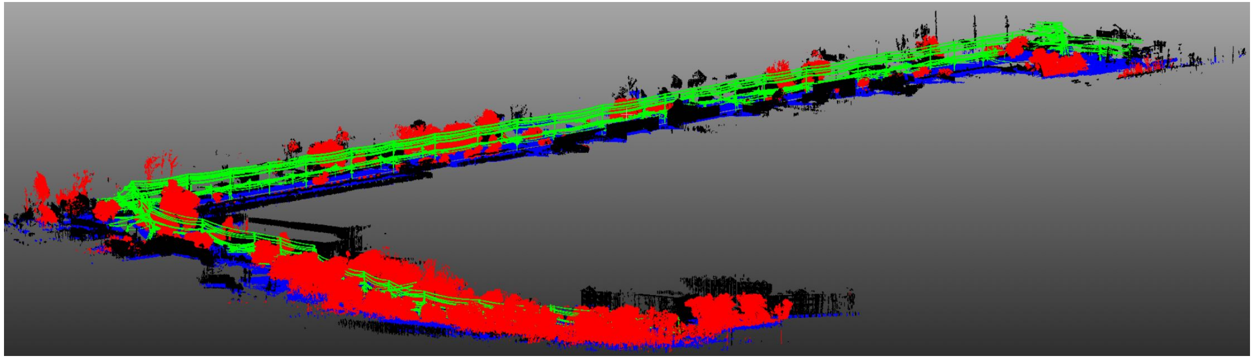
Problem

Aircraft defect detection

Solution

Object detection and/or semantic segmentation

Past: 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

Aircraft Defect Detection

1. Problem Overview
2. Dataset Study
3. Proposed Approaches
4. Preliminary Results
5. Timeline

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

Proposed Approaches

- Object detection (Chang)
- Semantic Segmentation (Anshuman)

Approach 1: Object Detection

Why object detection?

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

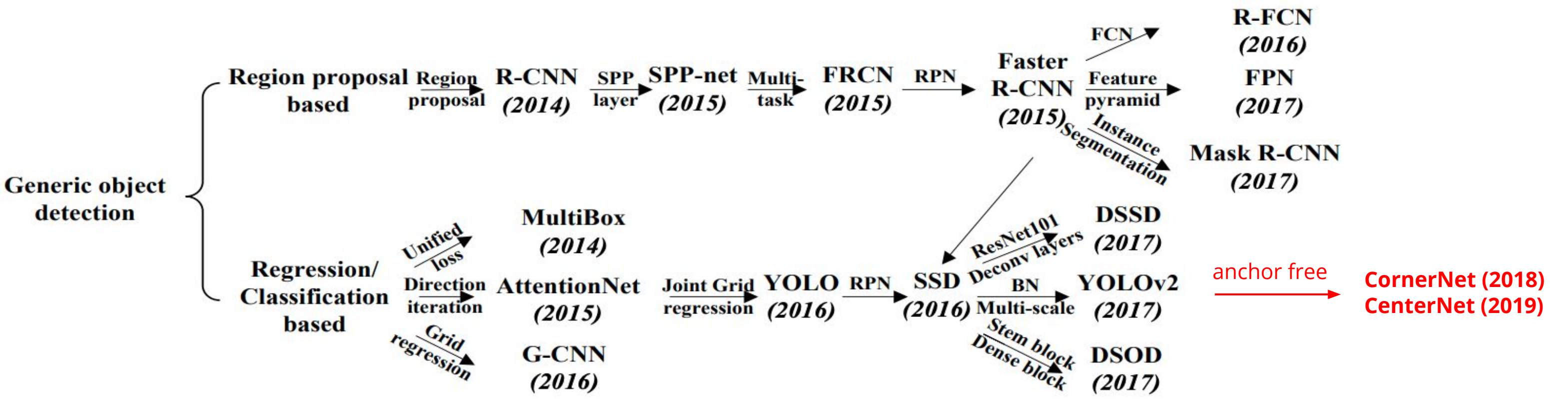
What type of object detection?

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

Carnegie Mellon University

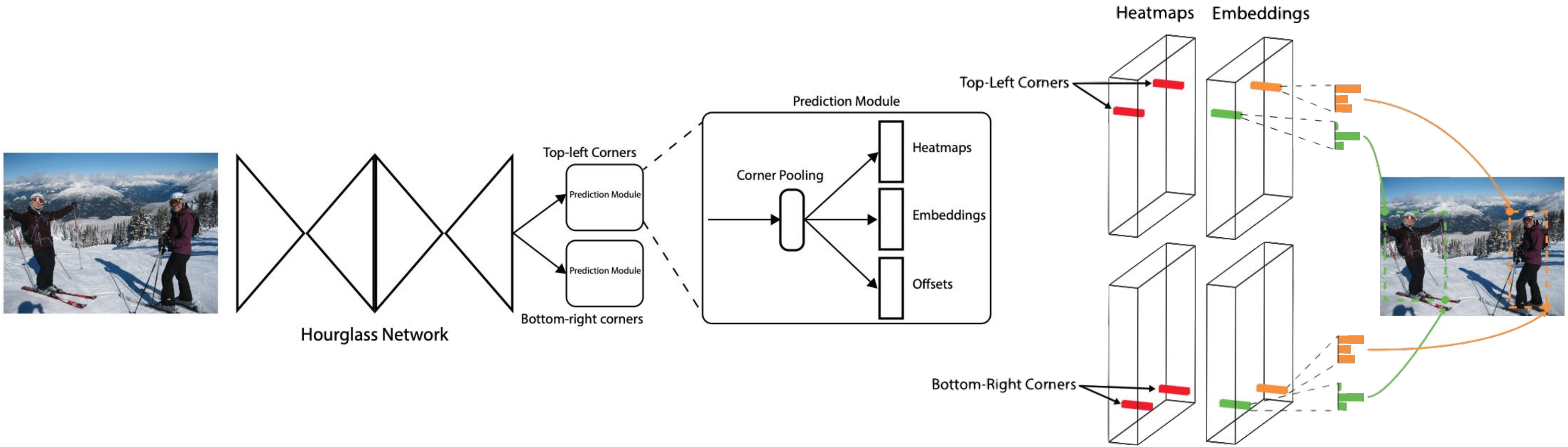
Literature Review

Object detection roadgraph



Literature Review: CornerNet

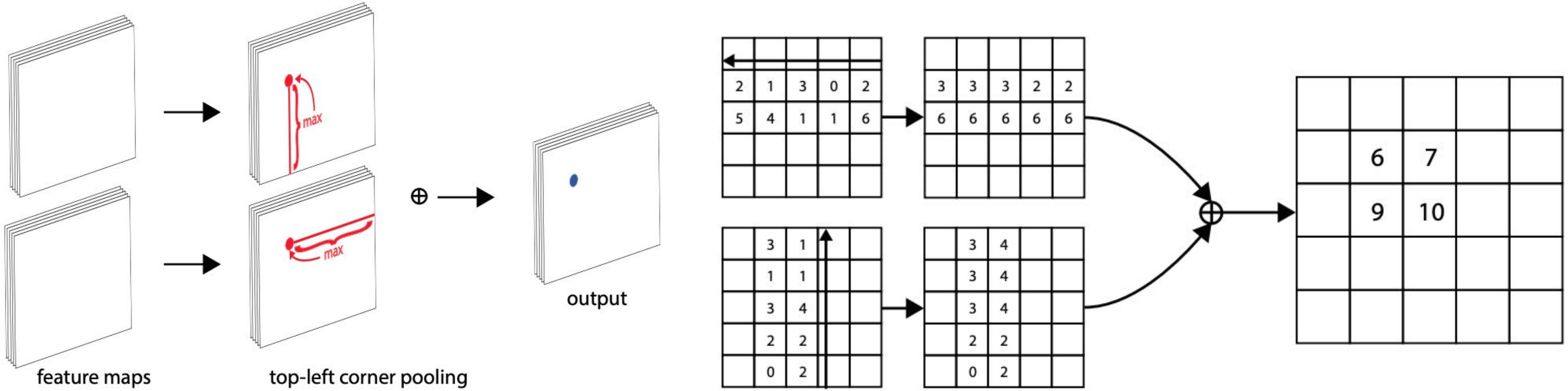
CornerNet Architecture



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

Literature Review: CornerNet

Corner Pooling Module

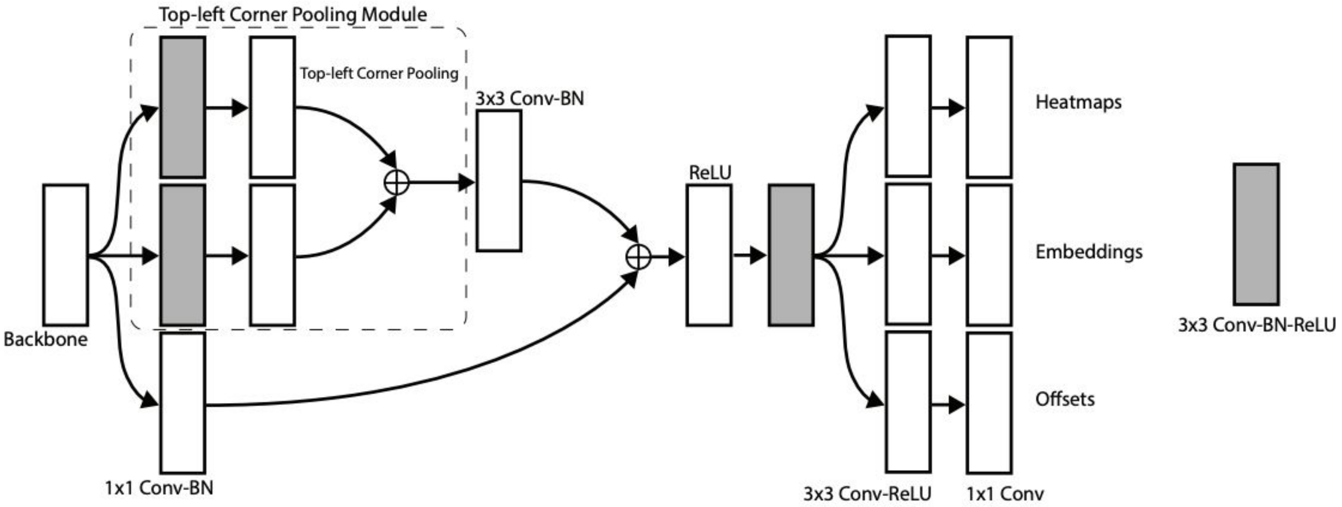


Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Carnegie Mellon University

Literature Review: CornerNet

Bounding Box Prediction Module (Top-left branch)

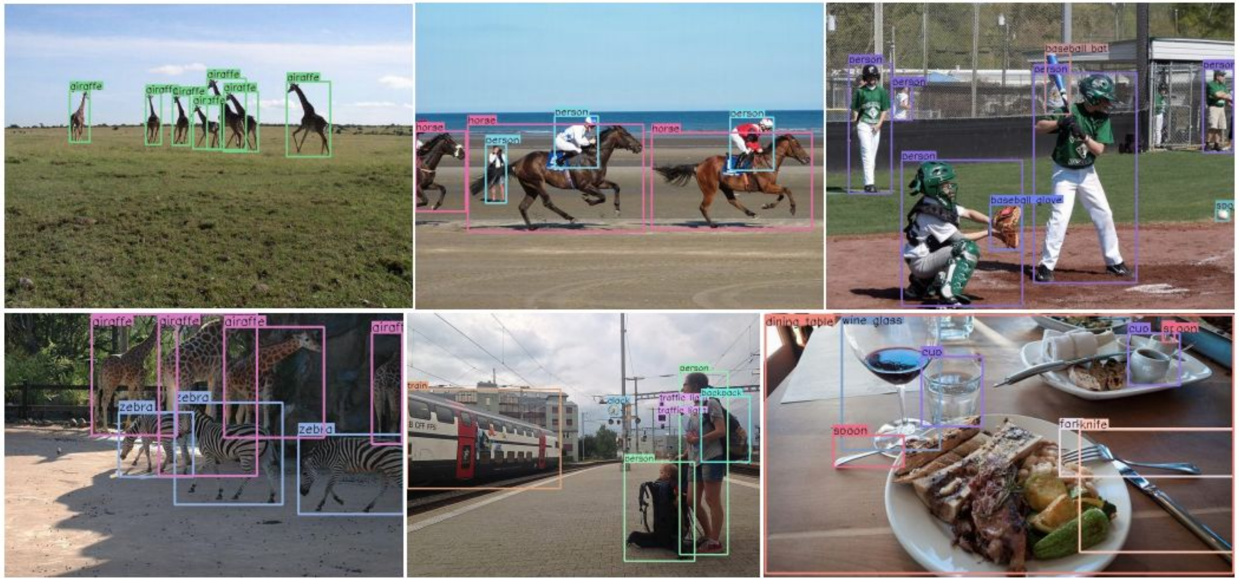


Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." ECCV 2018.

Carnegie Mellon University

Literature Review: CornerNet

Qualitative Examples on MSCOCO



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

Carnegie Mellon University

Literature Review: CenterNet

Improvements from CornerNet:

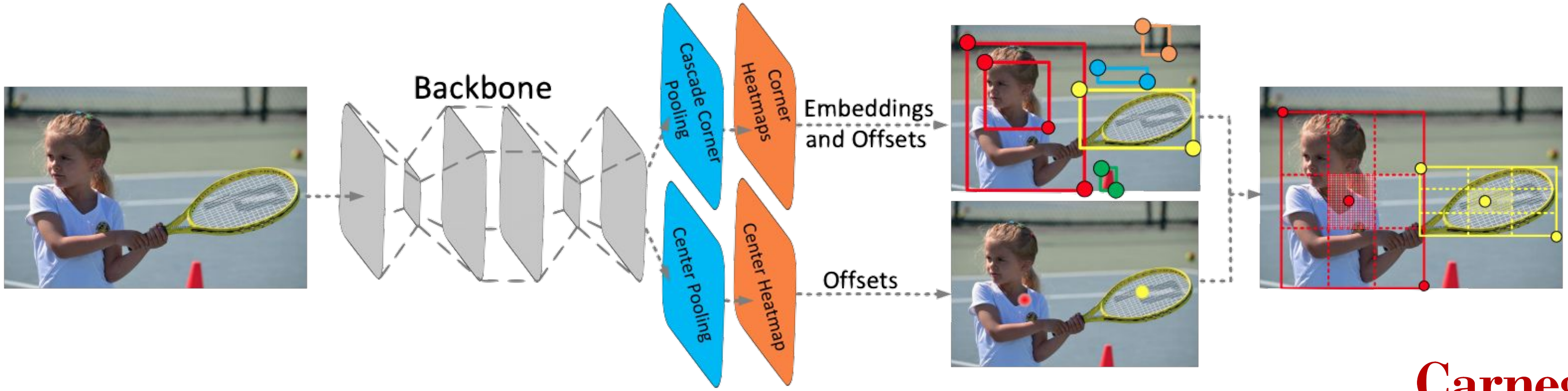
- Center pooling module: inherits the functionality of RoI pooling
- Cascade corner pooling: perceives internal information

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

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.

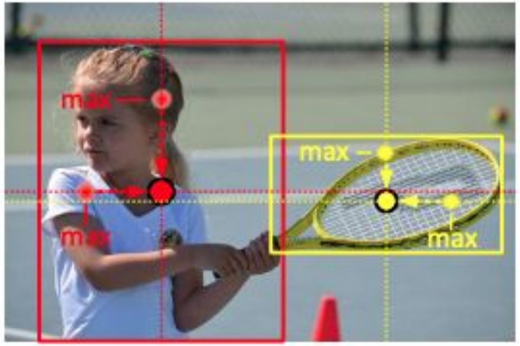


Duan, Kaiwen, et al. "CenterNet: Keypoint Triplets for Object Detection." ICCV 2019.

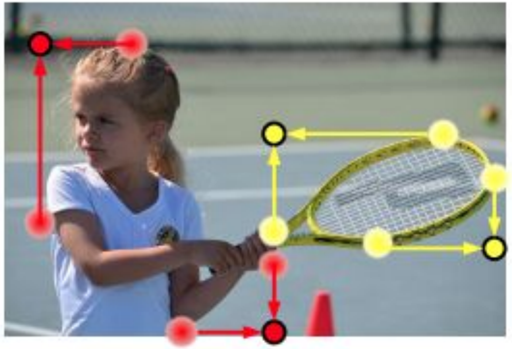
Carnegie Mellon University

Literature Review: CenterNet

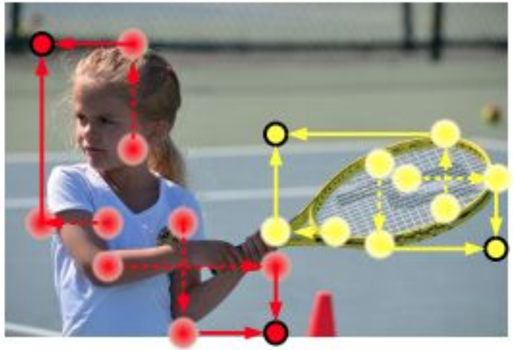
Center pooling & cascade corner pooling module



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)
- Single-scale training
- batch size of 4 on each of 2 Nvidia 1080 Ti GPUs
- 4K as training set and 1K as evaluation set

Preliminary 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%).

Preliminary 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!

Preliminary 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

Preliminary Results: Qualitative

Why overfitting is so severe?

1. Ground-truth labels are actually ill-posed

Ground-truth

Predicted

Carnegie Mellon University

Preliminary Results: Qualitative

More ill-posed or bad bounding boxes and labels

Preliminary 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

Preliminary Results: Qualitative

Why overfitting is so severe?

2. Current data augmentation techniques:
    - random rescaling, random cropping, color jittering
    - gaussian bump of corner/center ground-truth
    - class-balanced weights for losses [1]

Seems not enough for this small dataset. More to explore:

- random flipping and rotation
- mixup [2]

[1] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." CVPR 2019.
[2] Zhang, Hongyi, et al. "mixup: Beyond empirical risk minimization." ICLR 2018.

Preliminary Results: Qualitative

Why overfitting is so severe?

3. Model complexity is too high
   - Reduce channel sizes and layers
   - Results: With 1/13 parameters, network still converges on training set, yet still low mAP on evaluation set
   - Possible solutions:
     i. *Dropout/Dropblock*
     ii. *Even simpler models*

Preliminary Results: Qualitative

Examples: Low recall case

Ground-truth

Predicted

Preliminary Results: Qualitative

Examples: complex scene

Ground-truth

Predicted

Preliminary Results: Qualitative

Examples: various sizes

Ground-truth

Predicted

Future Work

1. Reduce the overfitting problem
2. Collect more data from the company (probably the easiest solution :p)

Approach 2: Semantic Segmentation

Why semantic segmentation?

- Higher recall for defects
- More fine-grained classification
- Solve ill-posed bounding box cases like

Literature Review

Classical Approaches

- UNet: Convolutional Networks for Biomedical Image Segmentation
- SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

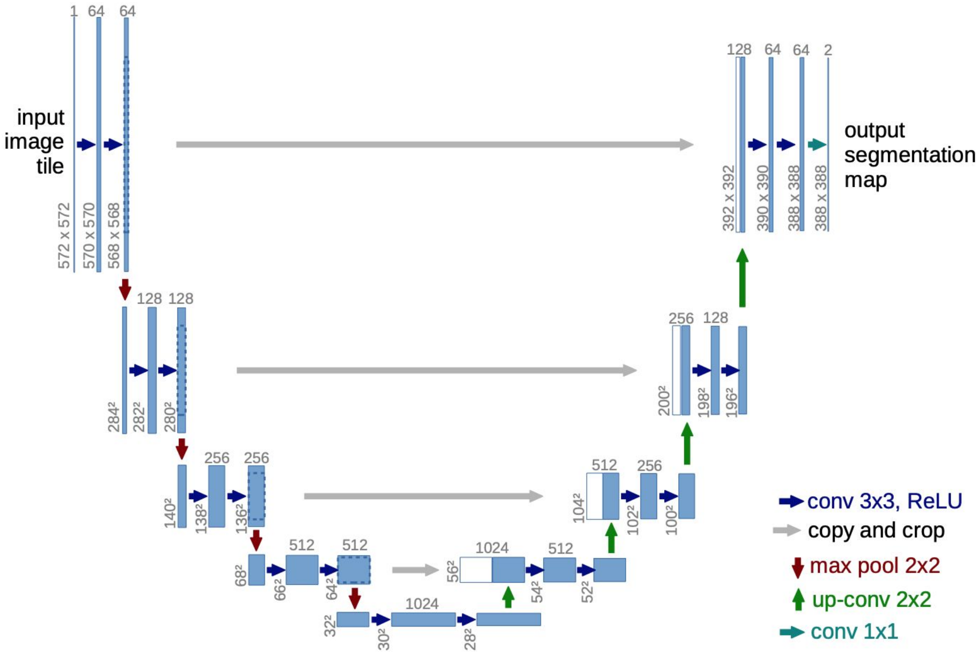
Why?

Ronneberger et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation"

Kendall et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation"

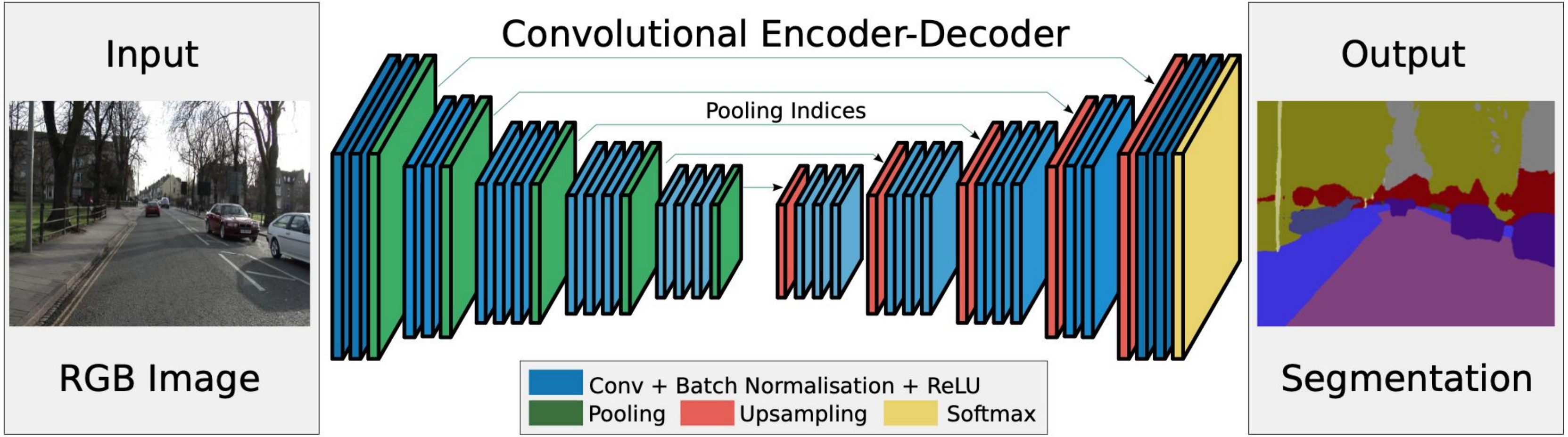
Carnegie Mellon University

Literature Review: UNet



Ronneberger et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation"

Literature Review: SegNet



Kendall et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation"

Timeline

Date	Task
Oct 31	Examination of the overfitting problems in CenterNet Preliminary results on semantic segmentation
Nov 15	Combining object detection and segmentation (E.g. weakly-supervised segmentation)
Nov 31	Fix final model and finishing training

Carnegie Mellon University