Kakao Brain's Approach for Task 2 in DOAI2019

Jihoon Lee*, Taegoo Kim*, Ildoo Kim*, Jonghyuck Park, Woonhyuk Baek, Sungbin Lim
Kakao Brain
{jihoon.lee, taegoo.kim, ildoo.kim, jonghyuck.park, wbaek, sungbin.lim}@kakaobrain.com

Introduction

Despite recent remarkable advancements in computer vision era, detecting objects in aerial imagery has been challenging in both industry and academia. Objects displayed on aerial images have similar characteristics which include crowded small objects, scale variations, arbitrary orientations, and various aspect ratio of objects. In this competition, it was mainly focused on scale variation to improve the performance of the detector.

In general, sizes of anchor boxes are crucial to improve the performance of an object detector. Due to the extreme extent of instance size in images, the result score was not promising by only adjusting sizes of anchor boxes. In the DOTA-v1.5 dataset, the sizes of the smallest and largest instances are 8 px^2 and 2,884,902 px^2 respectively. Detecting both small and large objects in aerial images with a set of anchor box strategy requires a large number of anchor boxes, which requires more computation power and deteriorates model performance.

Due to the variation of GSD (Ground Sampling Distance) metric for each image, the different size of instances exist regardless of their same class type, it was essential to normalize the image size based on the given GSD value. However, GSD was not annotated for the test dataset unlike the train and the validation dataset; thus, a simple regression model was developed to estimate GSD.

An ensemble model with a combination of Cascaded R-CNN (C-RCNN) and Single Shot multi-box Detector (SSD) was utilized for the final submission. Additionally, due to the lack of the number of instances for container cranes in the dataset, a fine-tuned C-RCNN detector was developed and trained with a manually annotated external dataset, which only includes annotations for container cranes.

Methods

GSD Normalization & Estimator

In order to reduce the variance of object size, GSD Normalization was conducted. GSD Normalization is an image augmentation technique to change the size of images so that the newly modified image gets the target GSD. However, it is required to estimate GSD value during test time since DOTA does not provide GSD for the test dataset. Thus, a GSD regression model has been invented using Resnet-34 backbone network.

Detection Models

C-RCNN (w/ Resnet-50) and SSD (w/ Resnet-101) were mainly employed to detect objects in the DOTA dataset. C-RCNN acted as the main detector that was responsible for detecting the majority of objects in scenes and SSD as a supplementary detector improved the recall of several classes.

---
* Equal Contribution

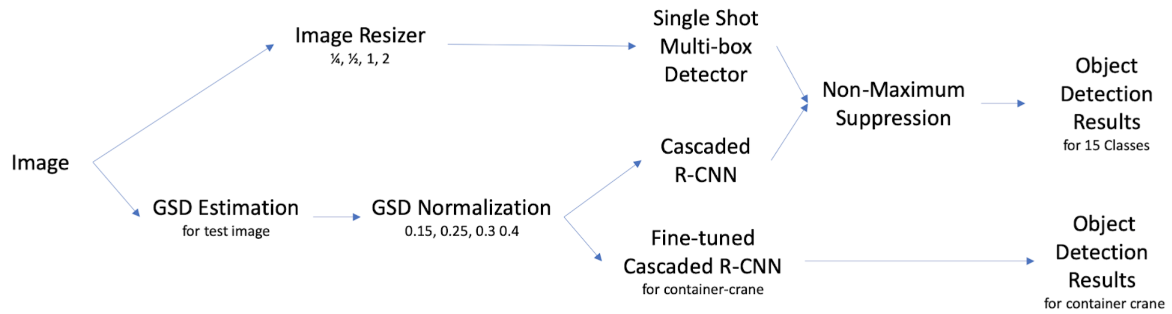

Figure 1. A diagram of the overall pipeline for the proposed object detector

The C-RCNN was trained with images that converted its GSD to 0.15, 0.25, 0.3 and 0.4 via GSD Normalization as described earlier. The GSD values were empirically chosen based on the anchor box size calculation, which was capable of covering the size of both the smallest to the largest instances. Additionally, deformable convolution layer was used. As a result, the C-RCNN performed as a major detector with mAP 73.0% in the validation set as a single model.

The SSD was trained with image patches in multiple scales splited using DOTA-devkit. Unlike the C-RCNN approach, GSDs of images were not converted for both training and inferencing. The single model mAP was 51.7%.

Weighted ensembling was performed to take benefits of two imbalanced models. The confidence scores of the SSD results were downscaled; then, the result was concatenated with the C-RCNN results. Lastly, non-maximum suppression was conducted. As a result, the ensemble technique slightly improved recall of classes that have a massive number of instances such as large-vehicle and small-vehicle and obtained mAP 73.3% in the validation set.

External Dataset for Fine Tuning

Container-cranes, the newly introduced label in the DOTA-v1.5, were hard to detect; only 142 container-cranes were annotated in the train dataset. The insufficient number of container-cranes was considered to be one of the main factors causing poor performance on the detection. In order to increase the number of container-cranes, the images of several container ports from Google Earth were collected. The container-cranes in each image were then annotated in a manner similar to that shown in DOTA dataset. 1,344 container-cranes were newly annotated in a total of 90 images. The dataset was only used to fine-tune the C-RCNN to learn the container-crane class.

Conclusion

GSD is significantly meaningful information that object detectors in natural scenes do not have for aerial imagery. During the competition, it has been proved that GSD Normalization can actually enhance the performance of object detection in aerial images by reducing the variance of object sizes. In addition, the proposed GSD estimator is expected to be a breakthrough tool that obtains more accurate GSD without manual installation of GCP (Ground Control Point) in the industry.