TCG DIGITAL

Custom Object Detection and Object Tracking on Video Data

May 19, 2023

TCG DIGITAL

CONTENTS

Executive Summary

The detection and tracking of objects within a video are fundamental elements in contemporary surveillance systems, as well as in anomaly detection video systems, traffic monitoring, crowd control, and industrial automation. However, pre-trained object detection models are often not suitable for specific use cases, such as tracking unique objects or individuals. Custom object detection and tracking models can be built to address this challenge. The development of custom object detection models involves data collection and annotation, model selection, and training. This white paper provides a comprehensive guide on custom object detection and tracking in video, covering the fundamental concepts, the process of building a custom model, a case study, and evaluation metrics.

This white paper provides a comprehensive guide on custom object detection and tracking in video, covering the fundamental concepts, the process of building a custom model, a case study, and evaluation metrics.

Problem Statement

The main goal is to identify and locate the target object in a video, which could come from a variety of sources, such as drone footage, CCTV recordings, or other video feeds. The suitability of pre-trained models such as R-CNN, YOLO, or SSD versus the necessity of training a custom object detection model using transfer learning will depend on the characteristics of the object being tracked. Once the object has been detected, object tracking must be employed to determine the object or person's movement within and outside the video frame.

Building a Custom Object Detection and Tracking Model

Building a custom object detection and tracking model involves the following steps: data collection and annotation, model selection, and training.

Phase 1

Detecting: In the detection phase, we employ our computationally intensive object tracker to perform two tasks: (1) identify newly appeared objects in our field of view and (2) locate any objects that were lost during the tracking phase. We update or establish object trackers for each detected object with the new bounding box coordinates. Due to the high computational cost of the object detector, we only execute this phase once every N frames.

Algorithms to be employed for Detecting – These include HOG + Linear SVM, Haar cascades, and deep learning-based object detectors such as YOLO, Single Shot Detectors (SSDs), and Faster R-CNNs.

R – CNN: R-CNN makes use of a region proposal method to create about 2000 ROIs (regions of interest). The regions are warped into fixed-size images and fed into a CNN network individually. Subsequently, it employs fully connected layers to classify the object and further improve the precision of the bounding box.

This method is slow as it requires the regions of interest to be generated. Although fast and faster R-CNN uses a feature extractor to improve upon the time, it still falls behind the single shot detectors.

YOLO:

It is an object detection algorithm, classed under the single shot detection category, in fact, the name YOLO is an acronym for You Only Look Once. It can run the convolutions and bounding boxes in the same run without the need for any recursive process. This makes it significantly faster than region–based algorithms such as the variations of R-CNNs that take several steps to detect the object.

The way it works is by dividing the input image into a G x G grid. Then for each grid, it runs a CNN and calculates the probabilities for the B number of bounding boxes of the object being detected in the cell. Finally, it accounts for the excess bounding boxes in a process called Non–Maximal Suppression by suppressing all the bounding boxes with lower probabilities and then returns the bounding box with the best Intersection over Union (IoU) from the remaining boxes.

Phase 2

Tracking:

The “detecting” phase is followed by the “tracking” phase. We create an object tracker for each of the objects that we detect to monitor their movement within the frame. Our objective is to make our object tracker faster and more efficient than the object detector. We will maintain the tracking process until we reach the N-th frame, after which we will re-run the object detector. The entire process is then redone.

Algorithms to be used for Tracking – Deepsort/Sort, MedianFlow, MOSSE, GOTURN, kernalized correlation filters, and discriminative correlation filters.

SORT:

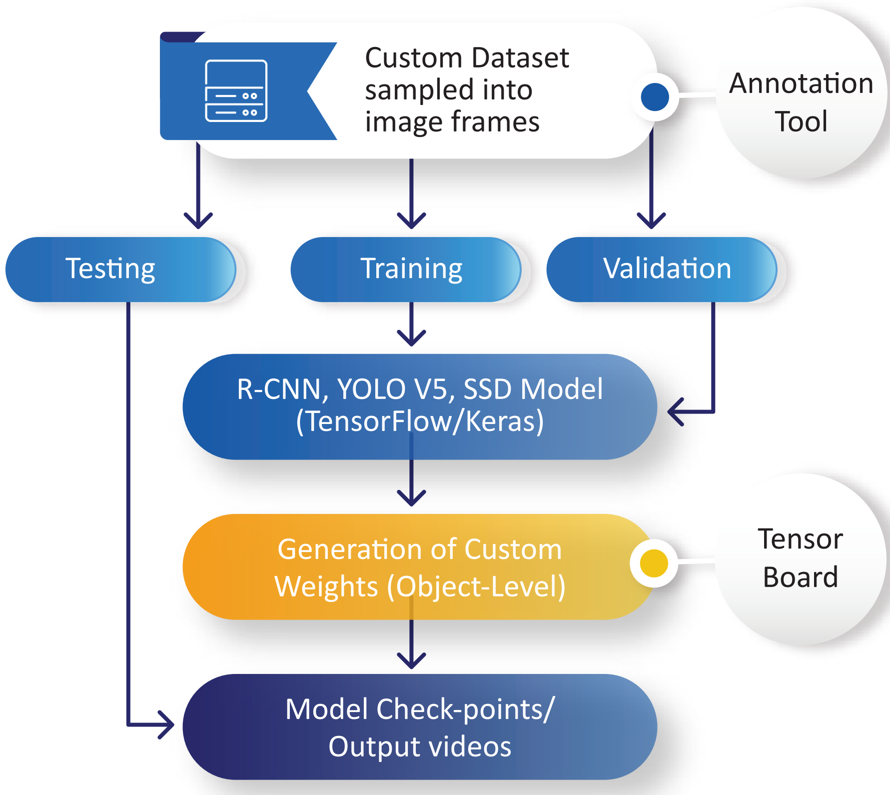
The abbreviation SORT stands for Simple Online Realtime Tracking. SORT is a real-time multiple object tracking system that employs straightforward heuristics to associate the objects in each frame with those detected in earlier frames.

It mainly relies on overlapping (Intersection over Union) in order to make an association, with objects across different frames. As an example, let us assume a car going down the road. Say, the car is detected at frame t. Now the filter will be used to run an estimate on where it can possibly be during the next frame (t+1), and if the car is detected within that estimated region or with a good overlapping, it will be automatically assumed that the car is the same car that was detected in the previous frame. Now the catch is, that any car detected in that position will be associated with the car in the previous frame, which will make the whole system inconsistent and unreliable.

DeepSORT:

DeepSort improves upon this by adding a pre-trained neural net to generate features for objects, so the association can be made based on feature similarities and not just on the overlap. It also tries to make use of cascaded matches to associate objects in multiple previous frames. This can often solve the problem of occlusion and track already tracked objects once they come back into visibility.

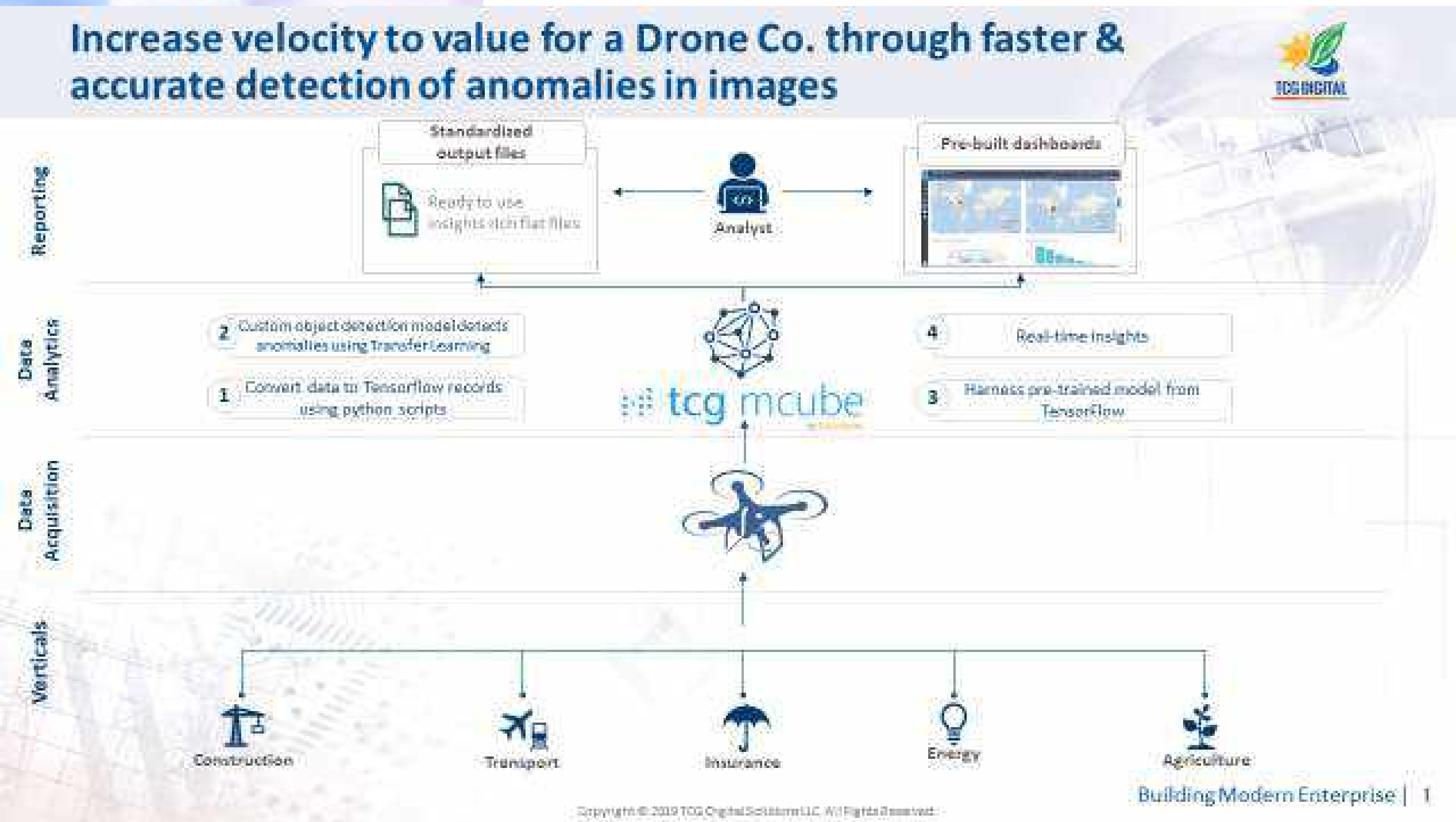
Custom Object Detection workflow in tcg mcube



Evaluation Metrics:

The performance of a custom object detection and tracking model can be evaluated using metrics such as precision, recall, and the F1 score. Precision measures the percentage of correctly detected objects among all detected objects. Recall measures the percentage of correctly detected objects among all ground-truth objects.
The F1 score is the harmonic mean of precision and recall.

Case Study:
A custom object detection and tracking model is used for anomaly detection in videos for a drone services company. It can be trained to track damage and maintain solar farms automatically using drone videos. It is also used for agricultural field monitoring and crop maintenance.



Conclusion

Custom object detection and tracking models are essential for addressing the limitations of pre-trained models in specific use cases. This white paper throws light on custom object detection and tracking in video, covering the fundamental concepts, the process of building a custom model, a case study, and evaluation metrics. Building custom models involves data collection and annotation, model selection, and training. Transfer learning and data augmentation techniques can be used to improve the accuracy and robustness of the model. The case study demonstrates the effectiveness of custom object detection and tracking in anomaly detection on drone video.

TCG DIGITAL

Author

Ashutosh Ranjan
Manager – Data Science Competency,
TCG Digital

TCG Digital is the flagship data science and technology solutions company of 'The Chatterjee Group' (TCG), a multi-billion dollar conglomerate. We leverage hyper-contemporary technologies and deep domain expertise to engage enterprises with full-spectrum digital transformation initiatives in operational support systems, enterprise mobility, app development and testing, cloud and microservices, automation, security, big data, AI/ML, and advanced analytics.

In addition to our digital transformation practices, by using our end-to-end AI and advanced analytics platform, **tcgmcube**, enterprises are extracting highly actionable insights from their invaluable data assets, and achieving Velocity to Value.
tcgmcube democratizes data science with scalability, performance, and flexibility. For more information, please visit our website at www.tcgdigital.com