Object Detection Model on Thermal Images

Enes Özipek
Advisor: Tambet Matiisen

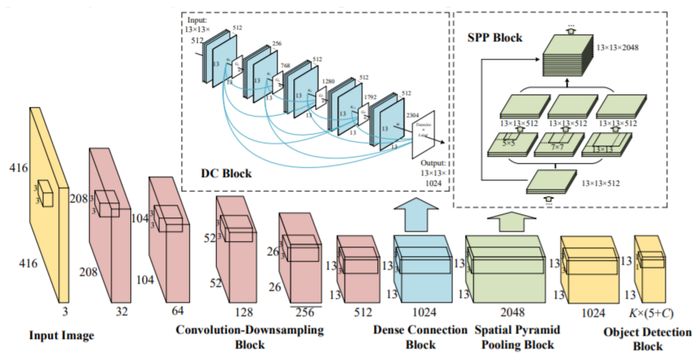
Motivation

For UGVs(Unmanned Ground Vechicles), it's vital to have capability to navigate around at any time of the day and night. Furthermore, in some cases, it's required that navigation for military UGVs should take place without being detected by enemies. Therefore, for UGVs, passive only terrain perception is a must to possess. Moreover, it's significant to have a reliable object detection model to predict the objects on thermal images captured by FLIR(Forward Looking Infrared) cameras.

Objectives

- My main objective in this task is to train an object detection model on FLIR dataset[1]. This model is aimed to achieve the same mAP(mean average precision) as baseline or preferably better than it.
- Another goal is to get familiar with training a model based on using known repositories, that is appyling theoritical knowledge on a practical problem.

Backbone Network(Yolov3-SPP)[5]



Experiments

First of all, I tried training a RefineDet512 model based on the repository[2]. Obtained bad results and getting the repository up&ready for training was tedious. Second option was to try out Yolov3[3]. Following steps are recorded to obtain the best weight in the current configuration:

- **Imbalanced dataset**
  - Only three categories(People,Bicycles and Cars) can be observable enough with annotations. -> *Oversampling !*
- **Custom Anchors**
  - Recalculating anchors based on the size distribution in the dataset helped the network have a fast convergence and better mAP score.
- **Pretrained Convolution Weights on Imagenet**
  - It takes the network less time to converge.
  - More better mAP score in the first calculation.
- **Channel Number**
  - Original images are 3-channels gray-scale images and pretrained network has been trained on 3-channel images.
- **Adam Optimizer**
  - Do not use Adam
  - Darknet has *momentum*, **i.e accumalation of movement.**
  - Darknet has *decay*, **a weaker updating of the weights for typical features.**

Ablation Experiments(Darknet mAP)[4]

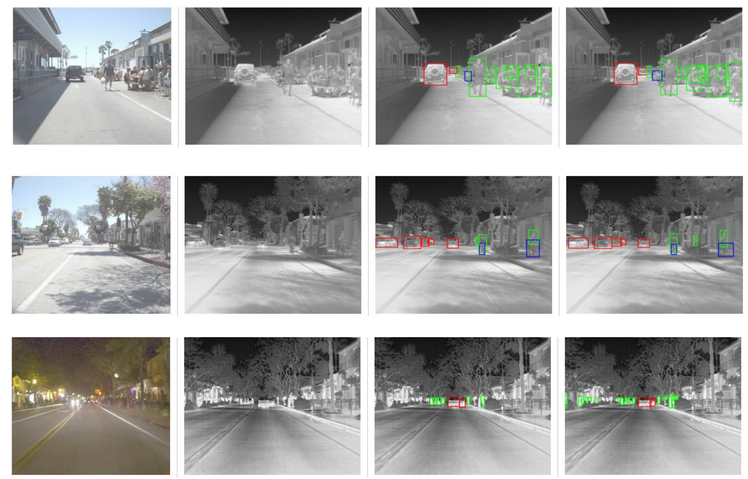
Model	People	Bicycles	Cars	All*	Epoch Number**
Yolov3-spp-thermal (without any ablations)	0.75	0.71	0.87	0.78	4
Yolov3-spp-thermal (with imbalanced dataset)	0.61	0.14	0.87	0.47	8
Yolov3-spp-thermal (with original anchors)	0.63	0.71	0.85	0.73	4
Yolov3-spp-thermal (without pretrained weight)	0.37	0.45	0.77	0.53	4

** Number of epoch at which corresponding results were obtained.

PyCOCO Results[6]

Model	People	Bicycles	Cars	All*
Yolov3-spp-thermal(Small area excluded)	0.749	**0.879**	**0.865**	**0.831**
Yolov3-spp-thermal	0.590	0.756	0.795	0.714
FLIR	**0.794**	0.580	0.856	0.743

* mAP scores when all of three categories are included.



Future Work

For the future work, configuration of the network might be modified in a way that it can better generalize on different datasets. Please note that current weight is obtained from early stopping point.

Conclusion

In this study, a thermal object detection model is trained using Yolov3-SPP as backbone. Numerical results are quite satisfying considering the baseline score. It's very likely to have better results if annotation errors are fixed.

Contact

Enes Özipek
University of Tartu
enesozipek05@gmail.com
+37258925631

References

1. https://www.flir.com/oem/adas/adas-dataset-form/
2. https://github.com/sfzhang15/RefineDet.
3. https://github.com/AlexeyAB/darknet
4. https://github.com/AlexeyAB/darknet/blob/master/src/detector.c#L649-L1088
5. https://arxiv.org/abs/1903.08589
6. http://cocodataset.org/#detection-eval