Object Recognition on Long Range Thermal Image Using State of the Art DNN

Veta Ghenescu[1], Eduard Barnoviciu[2], Serban-Vasile Carata[1], Marian Ghenescu[1], Roxana Mihaescu[2], Mihai Chindea[2]

[1]Institute of Space Science, Bucharest, ROMANIA
[2]UTI Grup, Bucharest, Romania

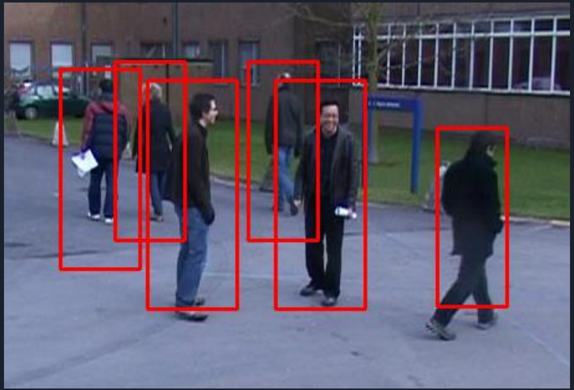
RO-LCG workshop, 17 – 19 October, Cluj-Napoca, ROMANIA

Content

- Introduction
- Proposed Method
  - YOLO model based on CNN
  - Retraining YOLO
  - The Database
- Results
- Conclusions

Introduction

Object detection: a solved problem ?



VS

In our case, the difficulty comes mainly from the distance at which the images were taken, but that’s not all.

Proposed method

Object detection was approached in many different ways.

HOG based classification, feature based classification (SURF, SIFT), etc.

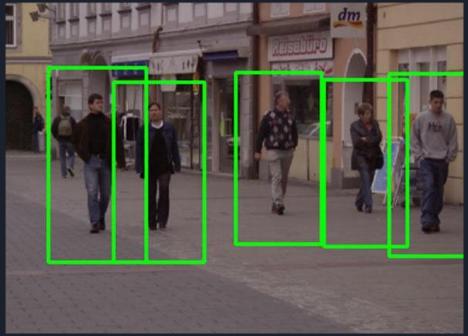
Neural Networks:

Why are neural networks so popular today?

- GPU computing
- Data availability

Feature extraction

Histogram of Oriented Gradients



Neural network based detection

Proposed method

Why use YOLO?

- It’s well documented, fast and computationally-cheap.
- Their architecture proved to be very versatile.
- There is also TinyYOLO. It can run on GPU-equipped SoCs !

Proposed method

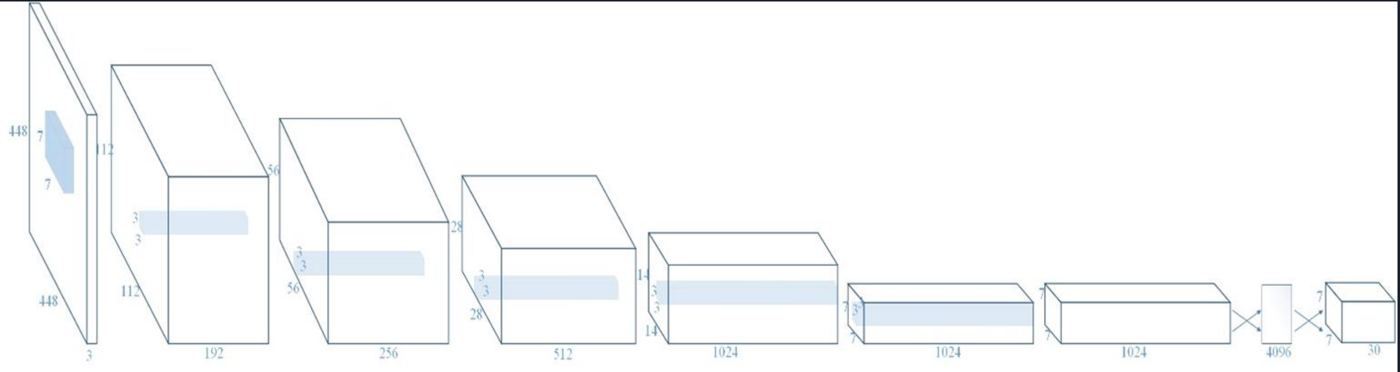
YOLO: Segmentation and classification in one run.

YOLO features:

- ❑ treats segmentation and prediction as one big regression problem.
- ❑ different from sliding window and region proposal classification networks.
- ❑ encodes contextual information about the classes, not just their appearance.

Proposed method

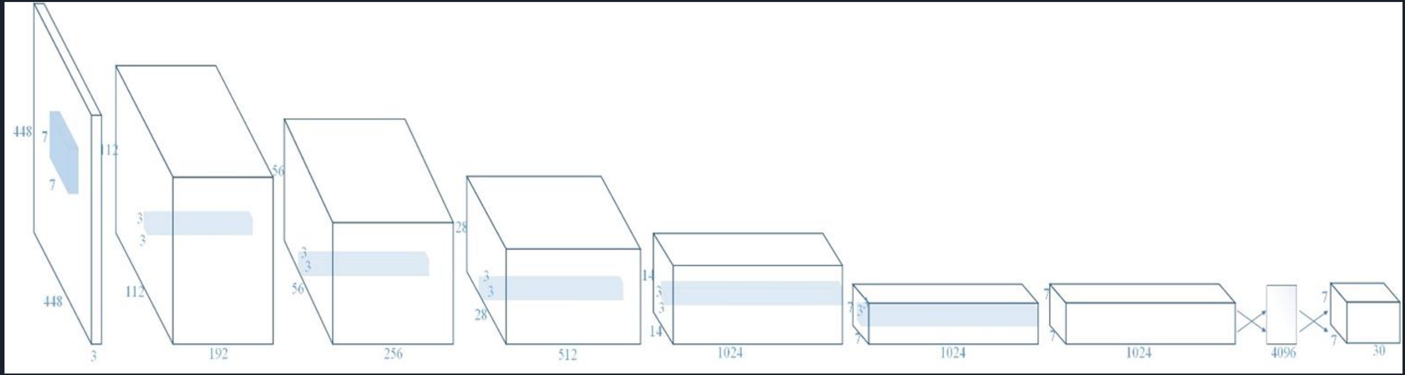
YOLO: Re-training



- The last convolutional layer is removed and replaced with another convolutional layer, specific to this application.
- The final layer is added, with the appropriate number of outputs:
- For each class, the network produces the bounding box and the probability.

Proposed method

YOLO: Re-training



- We considered that because the number of features is different from most detection cases, we should try different ranges of filters:

- The smallest configuration (filter-wise), had 2 filters at the first convolutional layer and 64 at the last one, while the biggest configuration we have tried had 128 filters at the first convolutional layer and 4096 at the last one.

Proposed method

YOLO: Re-training

- To enable YOLO to work on our database, we trained it on 34 different configurations.
- The most important difference between these networks is the size of the convolutions and the number of filters.
- All the configurations were trained on 800.000 generations.
- Each of them was tested at every 10.000 generation.
- In the end, we obtained 2720 results.

The Database

The complete database comprised 500.000 images that were manually annotated, some images containing up to 50 distingusable subjects.

Black is hot image
Tiny humans

White is hot Image
Big Humans

The Database

Eight classes were used: Human, Tiny Human, Animal, Tiny Animal, Boat, Tiny Boat, Vehicle, Tiny Vehicle

Black is hot image
Tiny humans

White is hot Image
Big humans

The Database

Another huge difference between them is the representation of the thermal imaging. It can switch at any time between “black is hot” and “white is hot” representations

Black is hot image
Tiny humans

White is hot Image
Big humans

Experimental Results

- Because this use-case of person detection is very different, comparing our results with results of person detection systems by other methods in a normal conditions, seems to be unsuitable.

- As a compromise, we compared our network with the standard YOLO network, which already offered some of the best human detection results at that time.

Experimental Results

- As mentioned earlier, 2720 total results were obtained.
- Out of all of those, one particular configuration yielded the best accuracy, at 68.75%.
- This result is a significant improvement on the standard model, which had a 45.23% accuracy.

Experimental Results

- Our best configuration used more filters and a smaller resolution than the YOLO standard model.

- This is a good understanding of how these particular types of images should be approach.

Experimental Results

Detection examples:

Example of differences in "Human" and "Tiny Human" classes

Conclusion

- Taking into account the difficult conditions that our database presented (small image resolution, noisy images and difficult to detect targets even for the human eye), we consider the obtained accuracy to be more than adequate. Our results are in a good agreement with previously published results.
- We proved that this algorithm can be trained for detection and classification in extreme conditions.
- This work has immediate real world applications, such as border security scenarios, which can reduce the cost and labor force of those who secure our countries.

Conclusion

Future Work

- In the future we intend to extend the number of classes that this model recognizes.
- Furthermore we intend to port this entire work to the new YOLO V3 network that has been launched since we have finished work on this particular stage of the problem.
- Besides this, in the next iteration we intend to change the number and the types of layers in the network, in order to obtain better results.

Acknowledgements

This work was supported by the Romanian Space Agency (ROSA), contract STAR-C3-160/2017 and Romanian UEFISCDI Agency under projects PN3-483/2017, PN3-P2-2.1-PED-2016-1065 and PN3-PED-147/2017.

Thank you for listening