Fast Convergence for Object Detection by Learning how to Combine Error Functions

Benjamin Schnieders[1], Karl Tuyls[1], Katie Atkinson[1], Andy Cooper[2]

[1]Department of Computer Science [2]Department of Chemistry

University of Liverpool, L69 3BX Liverpool, United Kingdom

{bsc, ktuyls, katie, aicooper}@liverpool.ac.uk

UNIVERSITY OF LIVERPOOL

Abstract

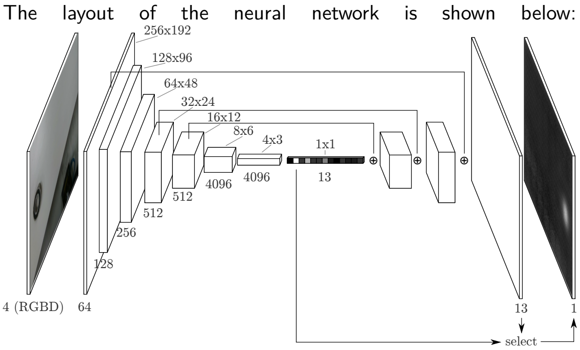
We introduce an innovative method to improve the **convergence speed** and **accuracy** of object detection neural networks [1]. Our approach, Converge-Fast-AuxNet, employs multiple, dependent loss metrics and weights them optimally using an auxiliary network.

Experiments are performed in the RoboCup@Work challenge environment.

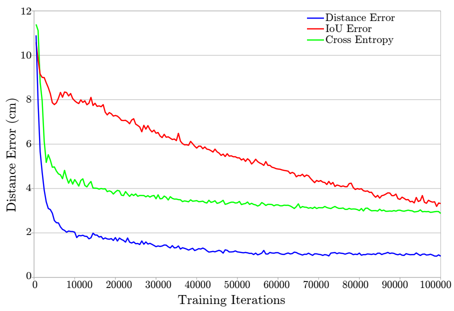
Our experiments show that adding an optimally weighted Euclidean distance loss, compared to a network trained on IoU alone, reduces the convergence time by **42.48%**. The estimated pickup rate is improved by **39.90%**. Compared to state-of-the-art methods [2], the improvement is **24.5%** in convergence, and **15.8%** on the estimated pickup rate.

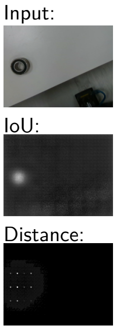
Approach

A Deep Convolutional Neural Network is trained to detect center points of RoboCup@Work objects. The layout of the neural network is shown below:



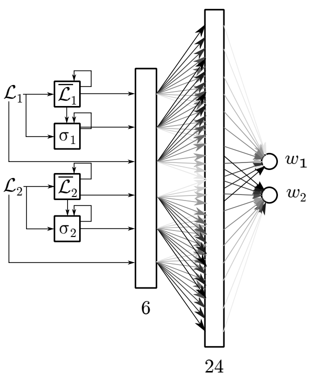
Training on the distance error metric alone converges faster than training on IoU, however, activations are sparse. Multiple, close objects can not be separated automatically with too sparse and distant activations. We hypothesize that a combination can produce superior results.





The network is thus tasked to reduce a weighted sum of the available error metrics:

$$\mathcal{L}_{Total} = \sum_i \frac{1}{w_i}\mathcal{L}_i + \log w_i$$



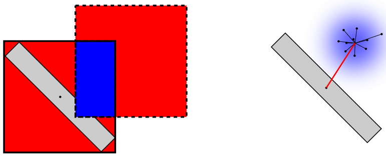
An auxiliary network is trained alongside the main network. It observes the current error, its average and its variance for all metrics. It is tasked to learn weights that produce the sharpest decline in overall error. The term to be minimized reads as follows:

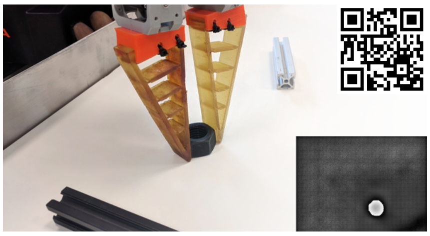
$$\mathcal{L}_{\text{AuxNet}} = \frac{\mathcal{L}_{Total} - \overline{\mathcal{L}_{Total}}}{\mathcal{L}_{Total}}$$

Introduction

Object detection: Output a list of trained objects that are present in a given image, along with their positions. A robot can use object detection to find and pick up objects.

Typically, object detection networks are trained on an IoU (Intersection over Union) metric (below left). Minimizing Euclidean distance is an alternative:





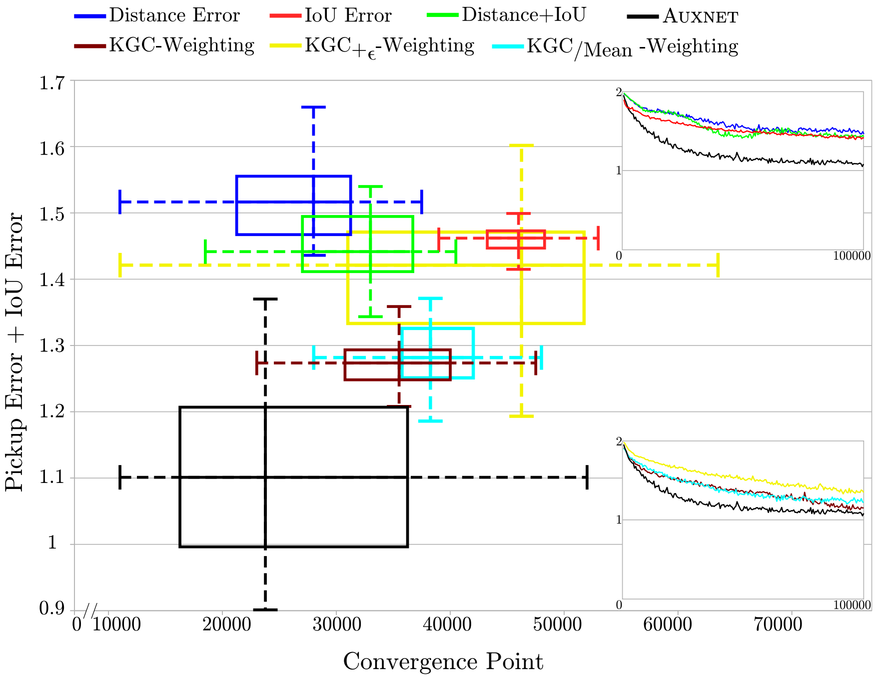
We model object detection as a multi-objective learning problem [3], learning both error metrics simultaneously. This leads to lower training times and lower error. AuxNet helps achieving an optimal balance of our error metrics during training.

Experiments

- Train 20 networks for all combinations of error metrics and weighting methods.
- Learning rate and other hyperparameters are chosen so all networks converge, then frozen
- Trained on 35000 images strong, self-recorded RoboCup@Work dataset
  Available from `https://airesearch.de/ObjectDetection@Work/`

Results

A double box plot showing the median and quartiles of each tested method in both X (convergence time) and Y (combined distance and IoU error). AuxNet outperforms state-of-the-art weighting methods (KGC-variants).



Conclusion

- Combining dependent error functions can reduce error and convergence time
- AuxNet reduces the average error by 15.85%; compared to the state of the art, convergence time reduced by 25%
- Results are statistically significant, as determined with a 2-Sample Kolmogorov-Smirnov test

References

[1] Benjamin Schnieders and Karl Tuyls. Fast Convergence for Object Detection by Learning how to Combine Error Functions. *Submitted to 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[2] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. *arXiv preprint arXiv:1705.07115*, 2017.

[3] Rich Caruana. Multitask Learning. In *Learning to Learn*, pages 95–133. Springer, 1998.