

Key Questions: 18.11.2022

Q: How could positional information of detected text / text regions be retrieved?

A: The current text region detector was built on the basis of YOLO v3

(<https://arxiv.org/pdf/1804.02767.pdf>), which is a general-purpose object detection approach. On an abstract level, it predicts boxes. In reality, our network outputs a set of

- center points (x,y) and per center point
 - a box height
 - a box width

This value set uniquely identifies each text region.

Q: How can high accuracy and scalability be achieved on low-resource devices?

A: There exists an approach called Neural Network Pruning. Basically, a large-scale network is trained and evaluated. Then within a monitored environment, certain parts of the network are 'turned off'. The impact of that in turn gets propagated to surrounding network parts. Surprisingly, this approach can maintain the original network's accuracy. In some cases, up to 90% of the original network could be removed without any loss in accuracy.

For more on pruning, please refer to this summary:

[Translue - Optimal Brain Damage - Pruning Neural Networks.pdf](#)