

EfficientDet: Scalable and Efficient Object Detection

Summary

1. follow the one-stage detector paradigm
2. introduce bidirectional feature pyramid network to fuse features from multiple input resolution.
argue that the multiple resolution contributes unequally to the final features, therefore let the network learn weights to determine the importance of each input resolution.
3. To scale up the base model, jointly scale up the resolution/depth/width for all backbone, feature network, box/class prediction network.
4. Use EfficientNet as the backbone.

Questions

1. It seems odd that using multiple resolution of the input helps. Shouldn't the network be able to learn to downsample?