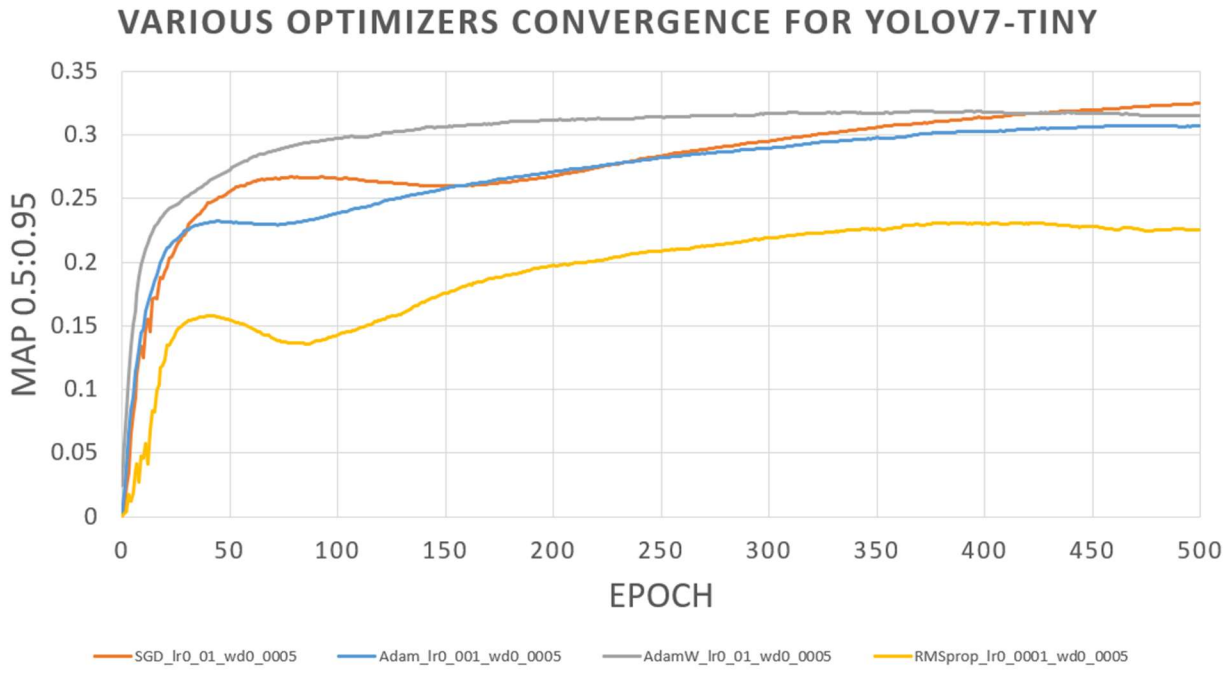
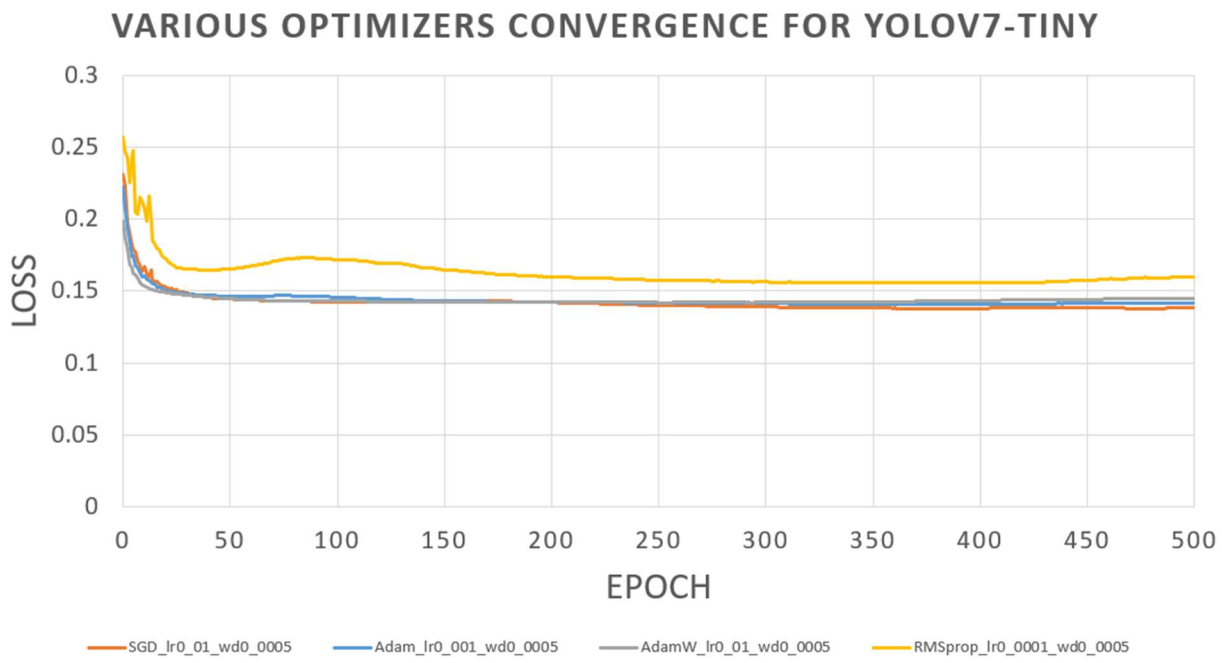
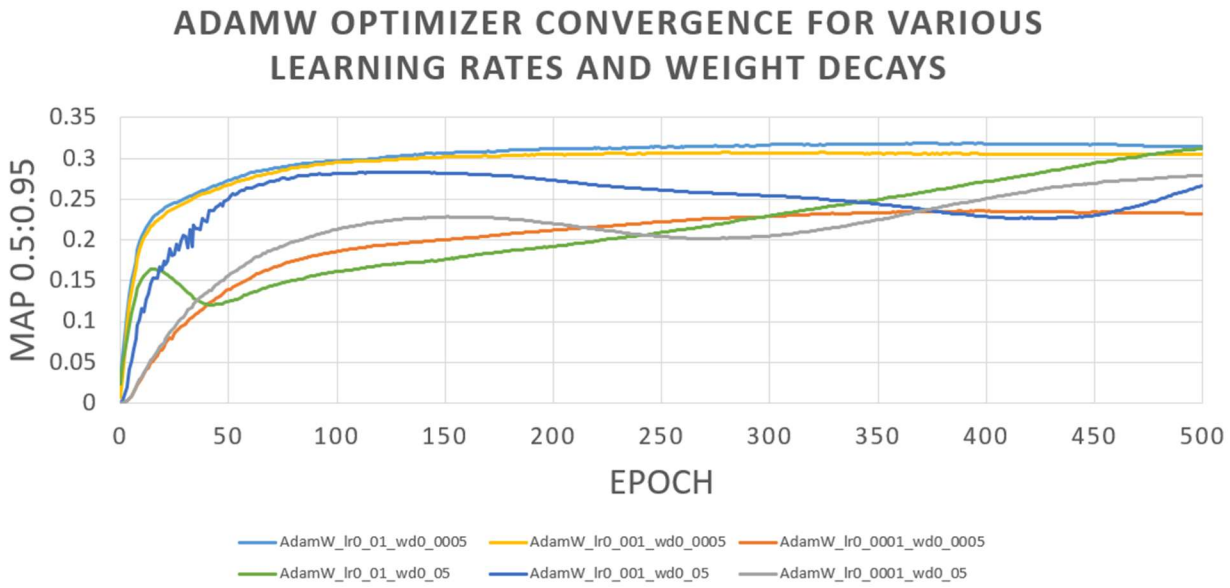
Various optimizers convergence on MS COCO for YOLOv7-tiny

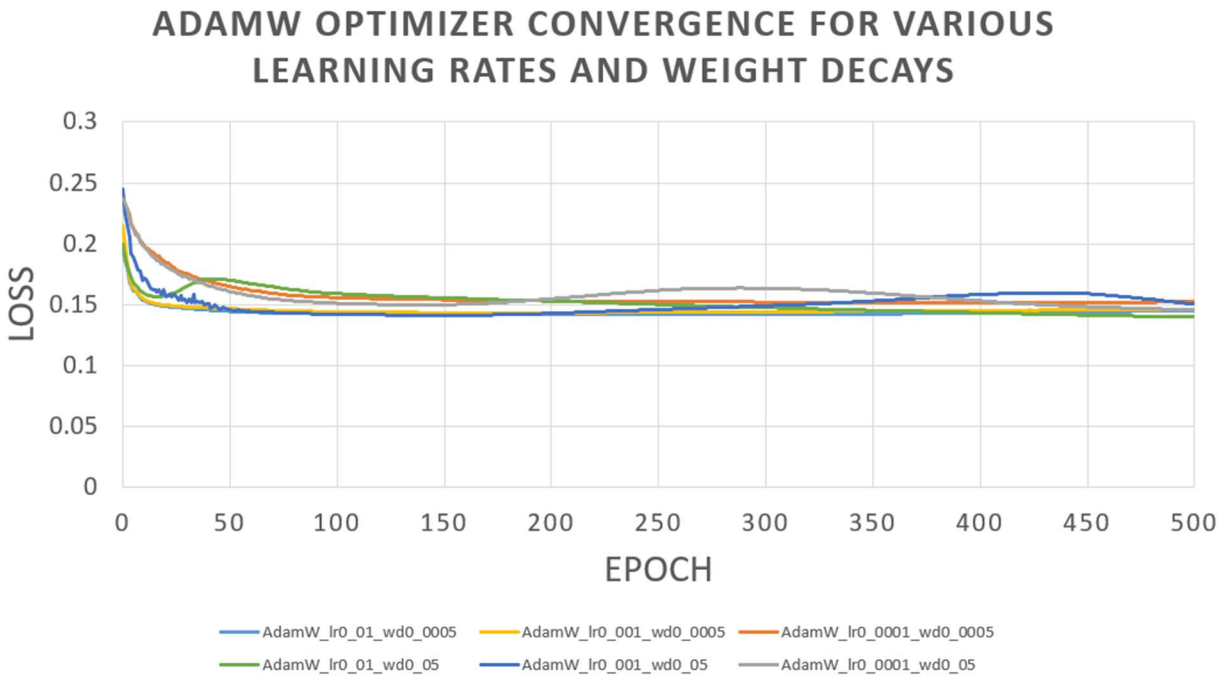
1. Best for each of SGD, Adam, AdamW, RMSProp



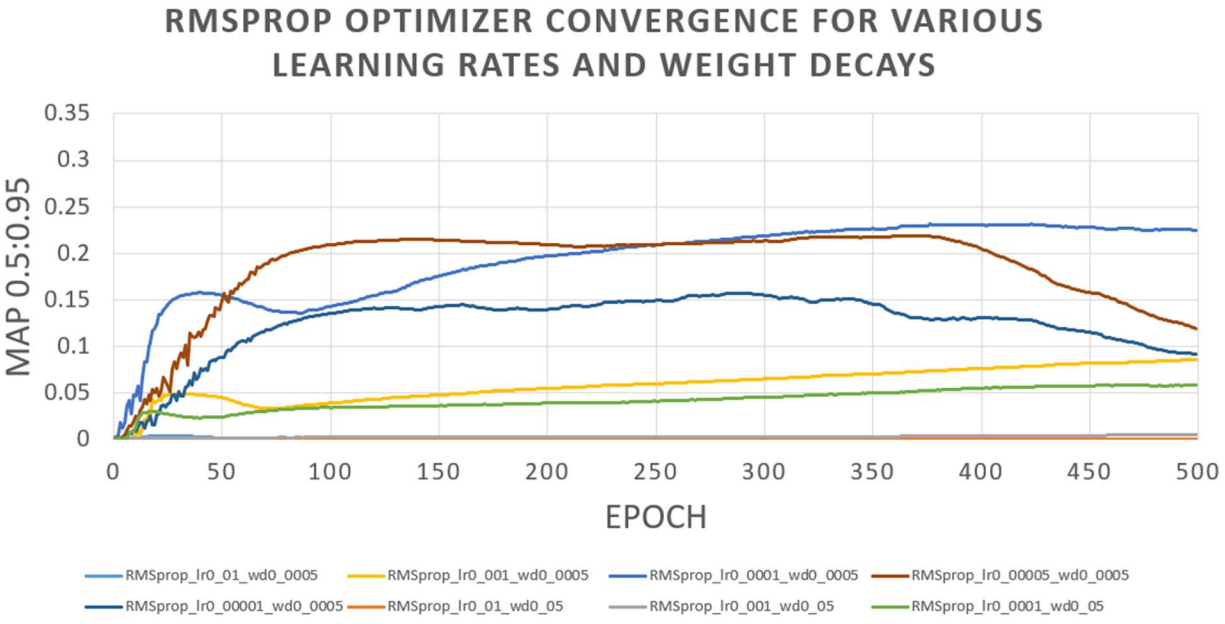


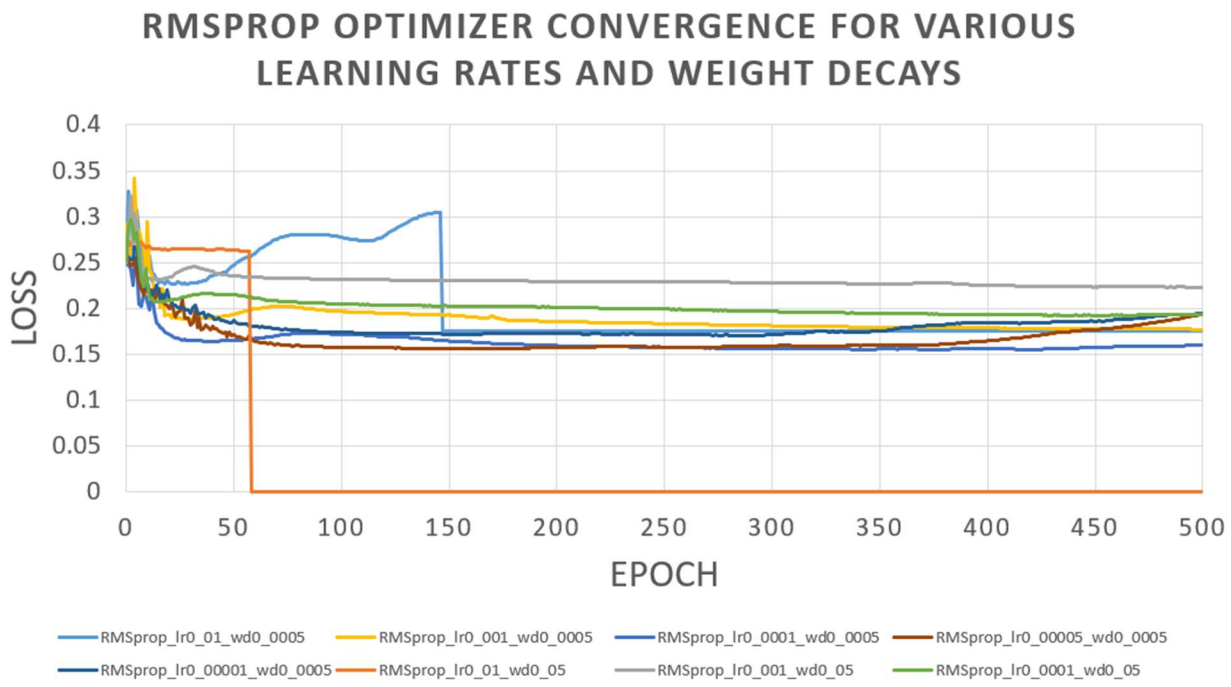
2. AdamW various learning rates and weight decays





3. RMSprop various learning rates and weight decays





RMSProp has an increased sensitivity to the learning rate setting, learning rates like 0.01 and 0.001 that work well for the other optimizers (SGD, Adam and AdamW) do not converge (see the RMSProp chart below). The best performing learning rate that we found was 0.0001 with weight decay of 0.0005. We also tried to lower learning rate setting to 0.00001, but it does not perform better. It could be that the combination of MS COCO data and the RMSProp updates do not lead to a good convergence for the lightweight YOLO7-tiny.