3.3 Synthetic Gradients

Idea: Each member has their own subnetwork for predicting synthetic gradients w.r.t. to their set of parameters. Each epoch, update everyone’s subnetworks as well as parameters. I’m guessing there’s some property that can be exploited as training progresses like reducing the number of communications over time since the subnetworks should become better at predicting the gradients. Should also mention the similarity to neuroevolution and possible extensions of this with neuroevolutionary selection. Would be cool to combine hypernetworks with decoupled neural interfaces with more fine-grained synthetic gradients.

Further generalize and make more practical