Ensemble loss (w. hyperparameters as in Wenzel et al. (2020)) Mean loss 0.7 $\pm 1\sigma$ Best individual model Wenzel et al. (2020) 0.6 0.5 0.3 0.2 10 14

Ensemble size

Categorical cross-entropy