Ensemble loss (w. hyperparameters as in Wenzel et al. (2020)) Mean loss $\pm 1\sigma$ 0.7 Best individual model Wenzel et al. (2020) 0.6 0.3 0.2 12 14 16 18

Ensemble size

Categorical cross-entropy