Calibrating Deep Neural Networks by Pairwise Constraints
Supplementary Material

A. Empirical Results

In addition to Figure 7, more evaluation results for different combinations of datasets and networks are provided in Figures A.1 and A.2.
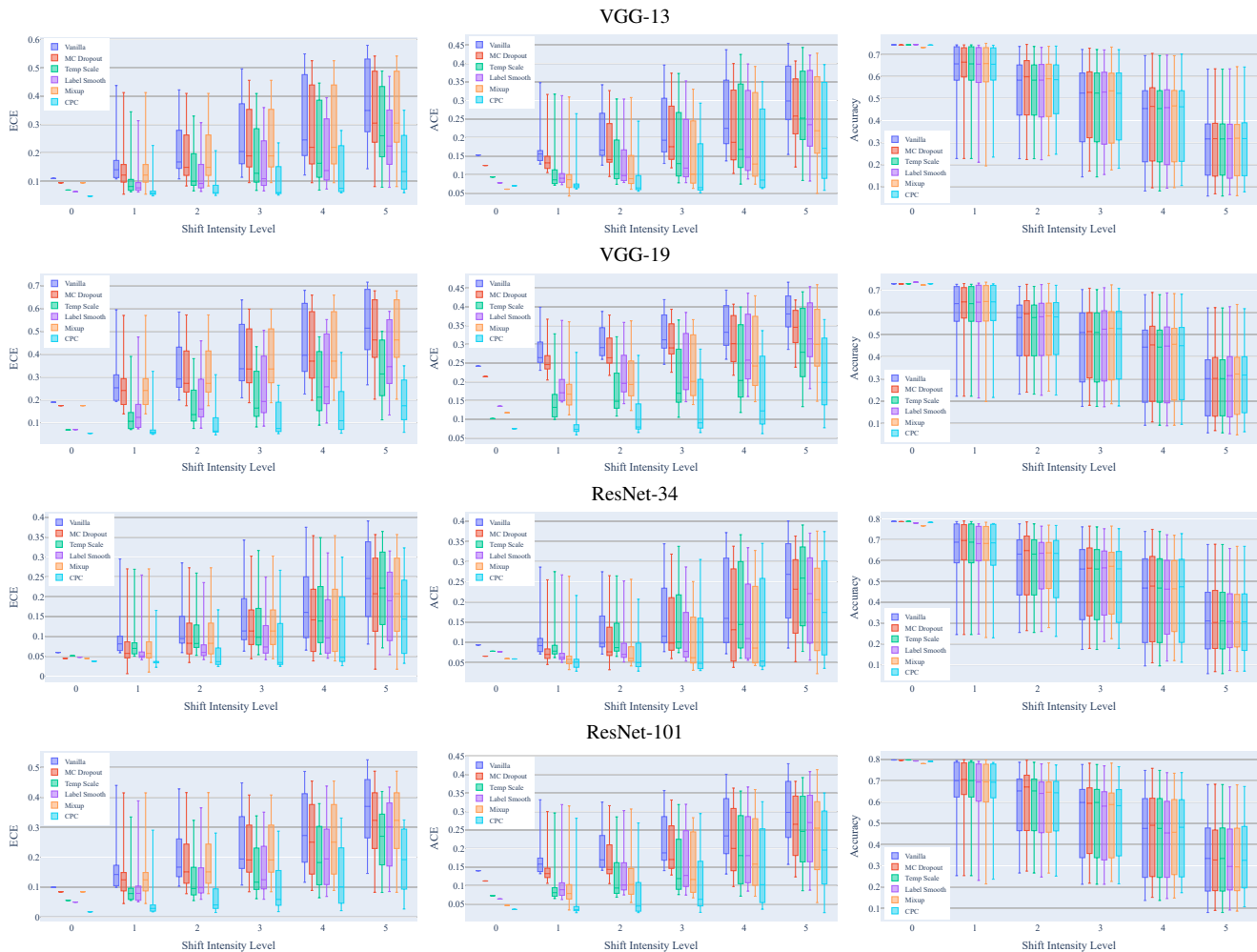

Figure A.1. Calibration and classification performance of different calibration methods on CIFAR-100 under different levels of data shift.
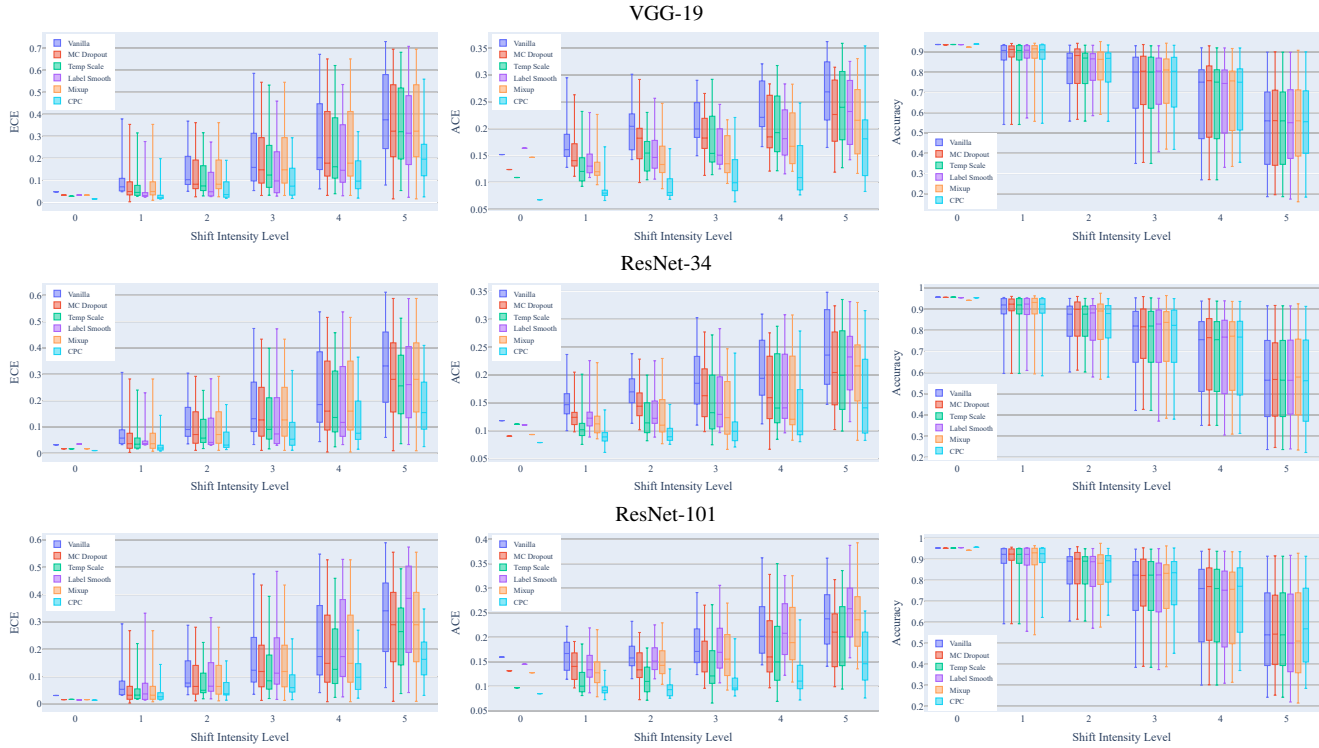
Figure A.2. Calibration and classification performance of different calibration methods on CIFAR-10 under different levels of data shift.

B. Implementation Details

For MC dropout, the number of stochastic inferences was set to be 10. For temperature scaling, the temperature hyperparameter was chosen from $\{0.1, 2, 4, 8, 16\}$. For label smoothing, the label smoothing parameter was chosen from $\{0.02, 0.04, 0.06, 0.08, 0.1\}$. These hyperparameters were determined by the same holdout validation procedure as that of CPC.