Appendices

A. Training and Test Errors

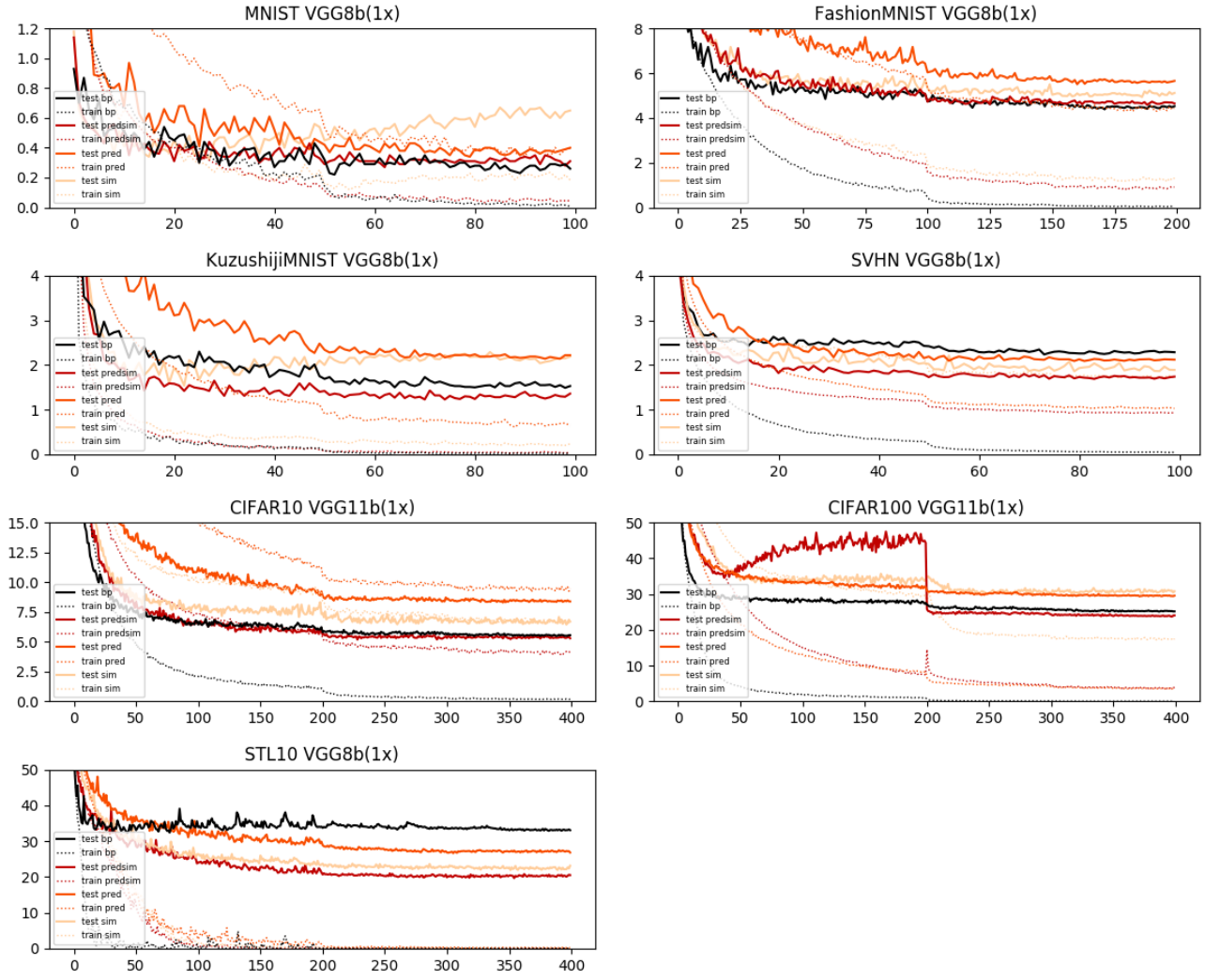

Figure 1. Training and test classification errors on all datasets with different loss functions. Note that the CIFAR100 runs are less comparable to each other, because the sim and predsim runs had batches sampled to have only 20 classes per batch during training, which we found to cause a higher training error, but lower test error.

B. Similarity Matching as a Complementary Objective

Although somewhat out of scope of this work, we performed some experiments with the **sim** loss as a local complementary loss. We trained the networks with global back-propagation combined with a local **sim** loss. In this way, hidden layers were trained based on a global cross-entropy loss, and back-propagated similarity matching losses from the layers above. Hyper-parameters and training details were identical to the experiments with the local **sim** loss, except that we did not detach the computational graph.

The results are summarized in Table **??**. We can see that similarity matching is a powerful auxillary objective for classification, also in a global loss context. For all datasets we can see an improvement in test error compared to global back-propagation alone.

Table 1. Similarity matching as a complementary objective. Test error in percent.

Dataset	Model	#par	glob	predsim	glob+sim
MNIST	VGG8B	7.3M	0.26	0.31	**0.24**
Fashion-MNIST	VGG8B	7.3M	4.53	4.65	**4.16**
Kuzushiji-MNIST	VGG8B	7.3M	1.53	1.36	**1.13**
CIFAR-10	VGG11B	12M	5.56	5.30	**4.33**
CIFAR-100	VGG11B	12M	25.2	24.1	**22.2**
SVHN	VGG8B	8.9M	2.29	**1.74**	1.95
STL-10	VGG8B	12M	33.1	**20.5**	25.6