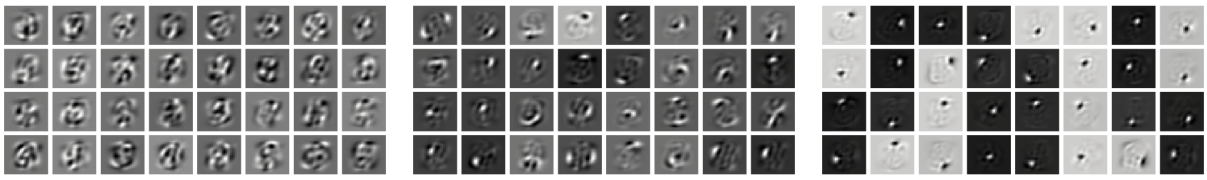
Lifted Regression/Reconstruction Networks
Supplementary material

Rasmus Kjær Høier and Christopher Zach
Chalmers University of Technology

1 Unsupervised learning: more results

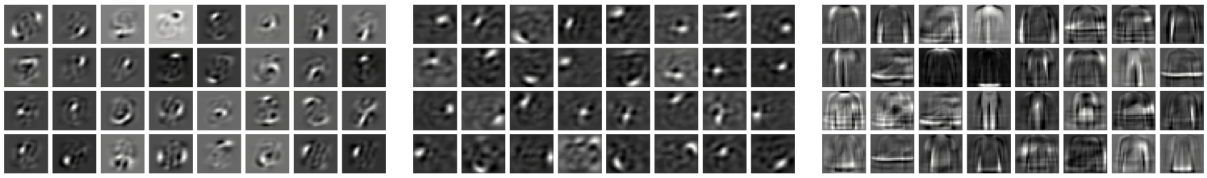
All networks were trained for 100 epochs with a learning rate $\eta = 0.005$. Fig. 1 shows filters obtained by unsupervised training of a 784-32-32 network with linear activation ($\mathcal{C}_k = \mathbb{R}^{d_k}$), ReLU-like activation ($\mathcal{C}_k = \mathbb{R}^{d_k}_{\geq 0}$) and hard-sigmoid activation function ($\mathcal{C}_k = [0,1]^{d_k}$). Fig. 2 illustrates the first layer filters for MNIST, KMNIST and FMNIST using the ReLU-type activation function.

The impact of the chosen activation function and dataset on the visual properties of the filters is evident. More constrained network activations leads to sparser filters, and the visual appearance of each dataset is also reflected in the filters.



(a) Linear activations (b) ReLU activations (c) Hard sigmoid activations.

Figure 1: Filters of a 784-32-32 unsupervised network trained on MNIST using different activation functions .



(a) MNIST (b) KMNIST (c) FMNIST

Figure 2: Filters of 784-32-32 ReLU networks trained unsupervised on MNIST, KMNIST and FMNIST.

2 Supervised Learning

2.1 Supervised learning from random initialization

A 784-64-64-10 ReLU-type LRRN network was trained for 100 epochs, using 20 BCD passes when inferring activations. A learning rate of 0.4 was used for MNIST and KMNIST and $\eta = 0.1$ was used for FMNIST. The reconstruction prefactors were chosen as $\beta = [1, 1, 0]1$ and the final layer was linear. The feedback parameter was chosen as $\gamma = 1/8$, and a batch size of 10 was used.

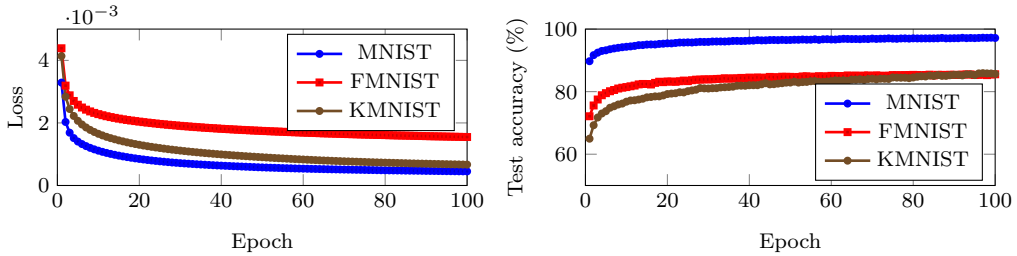
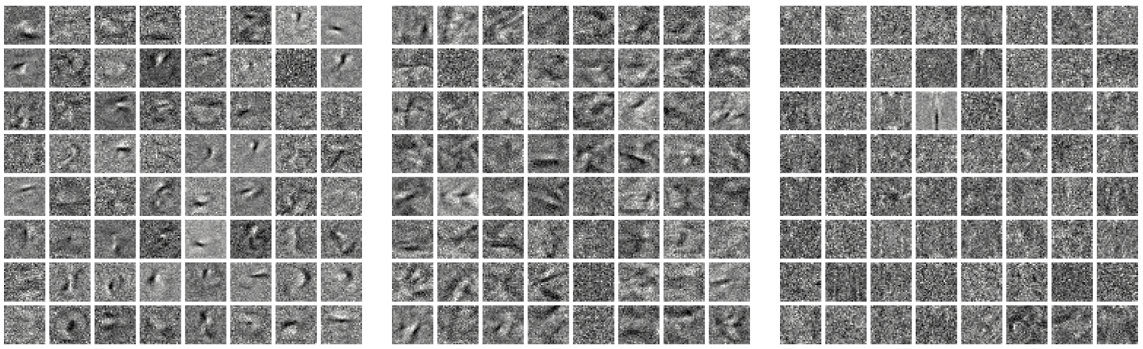


Figure 3: Loss and classification accuracies for a 784-64-64-10 ReLU-type LRRN. An accuracy of 97.2% is achieved for MNIST, 85.8% for KMNIST, and 86.4% for FMNIST.



(a) MNIST filters (b) KMNIST filters (c) FMNIST filters

Figure 4: 784-64-64-10 ReLU-type LRRN first layer filters.

2.2 Supervised learning with unsupervised pretraining

Three 784-64-64-10 LRRNs were trained on MNIST, KMNIST and FMNIST in an unsupervised manner by minimizing the free energy. A learning rate of $\eta = 0.005$ was used for all the networks. Furthermore $\beta = [1, 1, 0]$, $\gamma = 1/8$, and a mini-batchsize of 10. The resulting filters are shown in Fig. 6. The networks were then trained with supervision for additional 100 epochs. Fig. 7 depict the fine-tuned first layer filters, which retained most of their interpretable appearance. Fig. 5 shows the learning progress in terms of training loss and test accuracies. Pretraining leads to slightly better classification results, especially for the KMNIST dataset (85.8% vs. 87.91%).

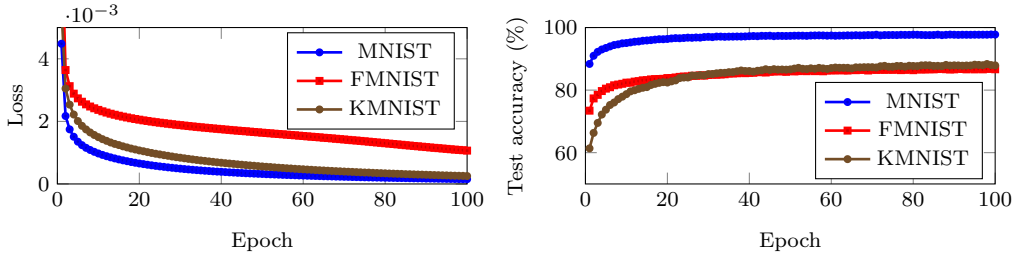
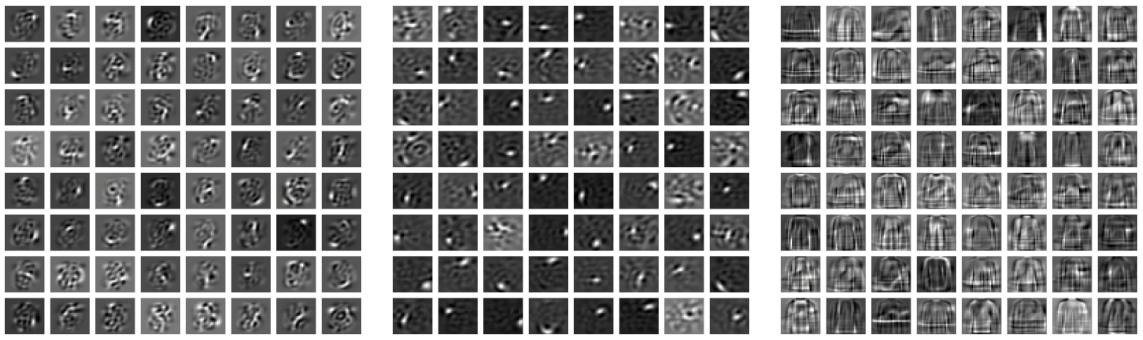


Figure 5: Loss and classification accuracies for the supervised training phase of a 784-64-64-10 ReLU-type LRRN. An accuracy of 97.76% is achieved for MNIST, 87.91% for KMNIST, and 86.66% for FMNIST.

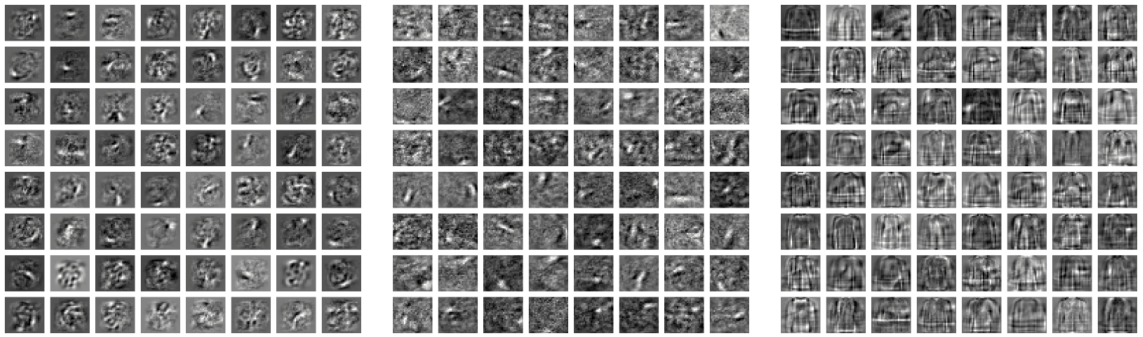
2.3 The impact of weight decay on the Lipschitz estimates

Tables 1 and 2 list the estimates for the Lipschitz constants ρ, the classification margin m and the allowed ℓ_2 norm δ for safe perturbations for two different weights on the weight decay term (10^{-5} and 10^{-4}). The upper bound on δ has been calculated via $\delta \leq \frac{m}{\sqrt{2}\rho}$. Higher weight decay regularization clearly induces a tradeoff between accuracy and perturbation robustness.



(a) MNIST filters (b) KMNIST filters (c) FMNIST filters

Figure 6: First layer filters of 784-64-64-10 ReLU-LRRNs after 20 epochs of unsupervised training.



(a) MNIST filters (b) KMNIST filters (c) FMNIST filters

Figure 7: First layer filters of the same 784-64-64-10 ReLU-LRRNs as shown in Fig. 6 after an additional 100 epochs of supervised training.

	ρ	$mean(m)$	$median(m)$	$std(m)$	Median ℓ_2 norm δ	test accuracy
MNIST	0.9387	0.8299	0.9270	0.2592	0.698	97.2%
KMNIST	0.9548	0.6195	0.6827	0.3442	0.506	85.6%
FMNIST	1.0716	0.6318	0.6913	0.3397	0.456	85.7%

Table 1: Lipschitz value of models trained on the three datasets (MNIST, KMNIST and FMNIST) and mean, median and standard deviation of classification margins. Weight decay factor 10^{-5}.

	ρ	$mean(m)$	$median(m)$	$std(m)$	Median ℓ_2 norm δ	test accuracy
MNIST	0.4604	0.6581	0.7038	0.2899	1.0810	95.5%
KMNIST	0.4174	0.4402	0.4101	0.3051	0.6948	80.5%
FMNIST	0.4142	0.5006	0.4776	0.3267	0.8153	82.4%

Table 2: Lipschitz value of models trained on the three datasets (MNIST, KMNIST and FMNIST) and mean, median and standard deviation of classification margins. Weight decay factor 10^{-4}.