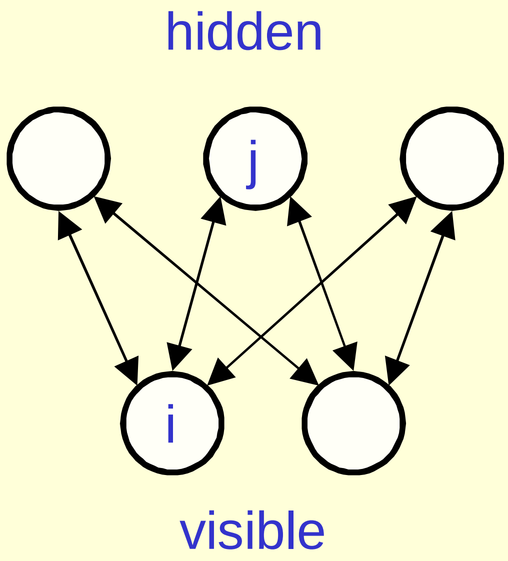
Neural Networks

Lecture 19
Learning Restricted Boltzmann Machines

A simple learning module: A Restricted Boltzmann Machine

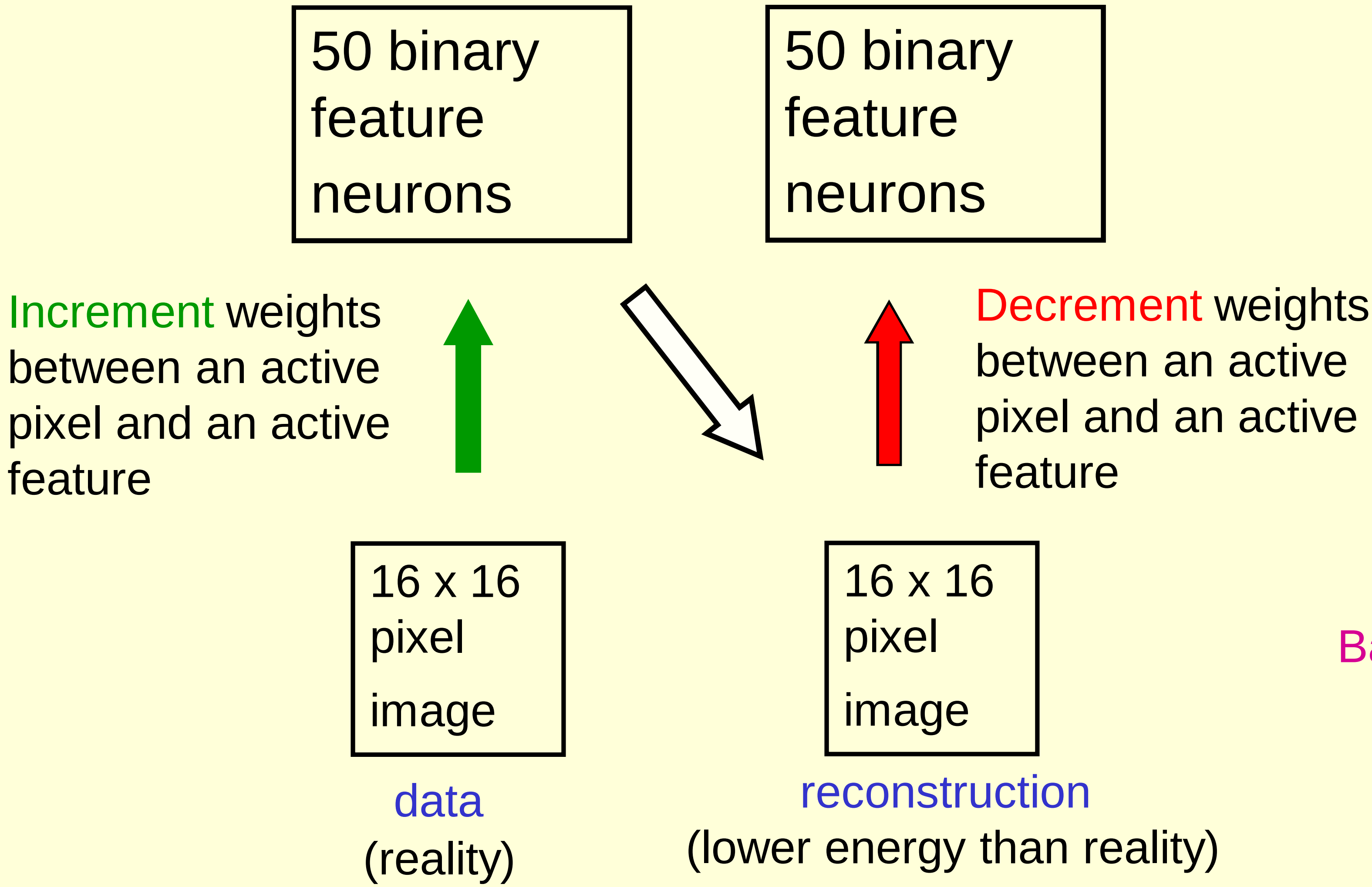
- We restrict the connectivity to make learning easier.
  - Only one layer of hidden units.
    - We will worry about multiple layers later
  - No connections between hidden units.
- In an RBM, the hidden units are conditionally independent given the visible states..
  - So we can quickly get an unbiased sample from the posterior distribution over hidden "causes" when given a data-vector
    .



Weights → Energies → Probabilities

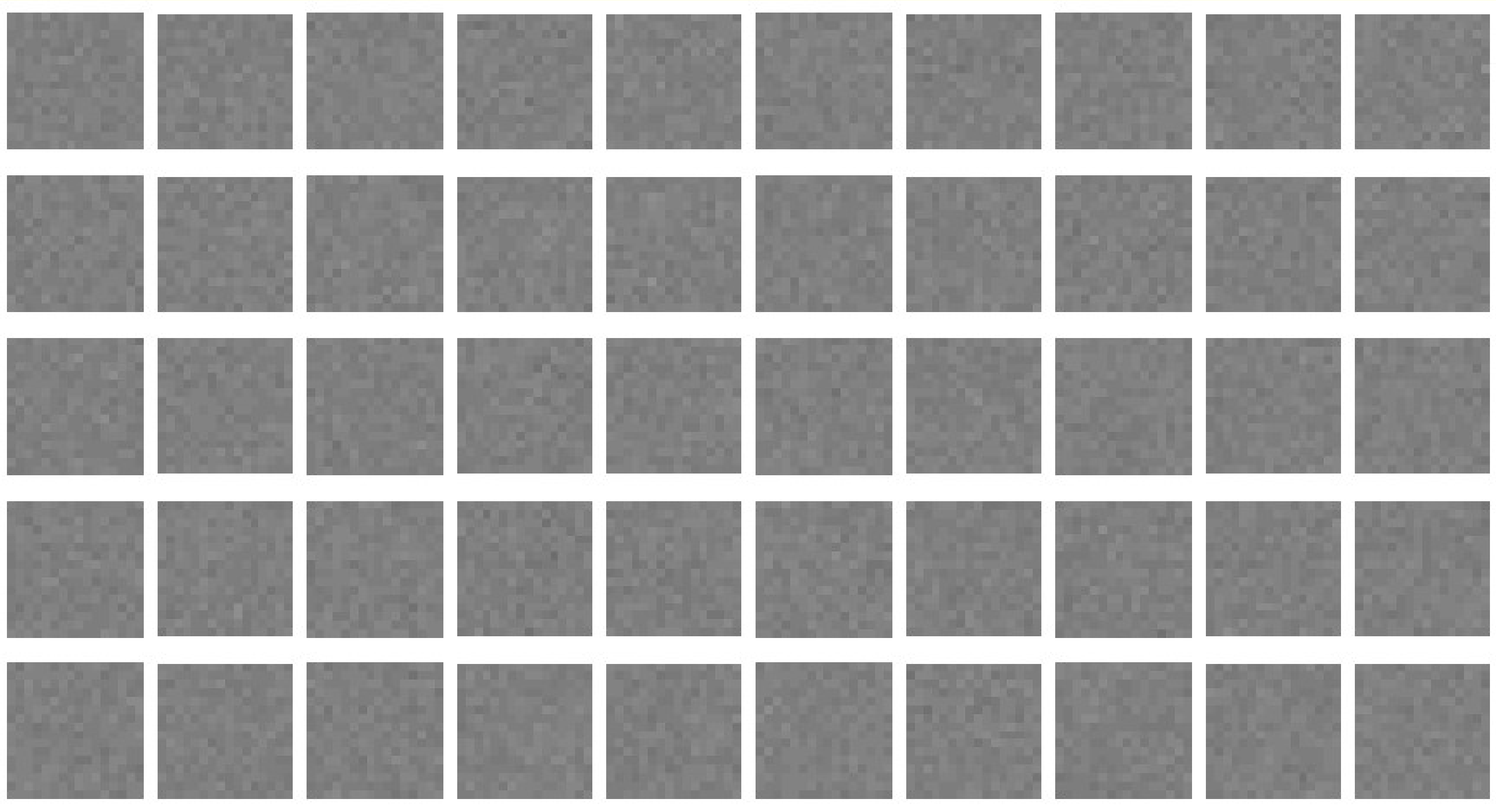
- Each possible joint configuration of the visible and hidden units has a Hopfield “energy”
  - The energy is determined by the weights and biases.
- The energy of a joint configuration of the visible and hidden units determines the probability that the network will choose that configuration.
- By manipulating the energies of joint configurations, we can manipulate the probabilities that the model assigns to visible vectors.
  - This gives a very simple and very effective learning algorithm.

How to learn a set of features that are good for reconstructing images of the digit 2

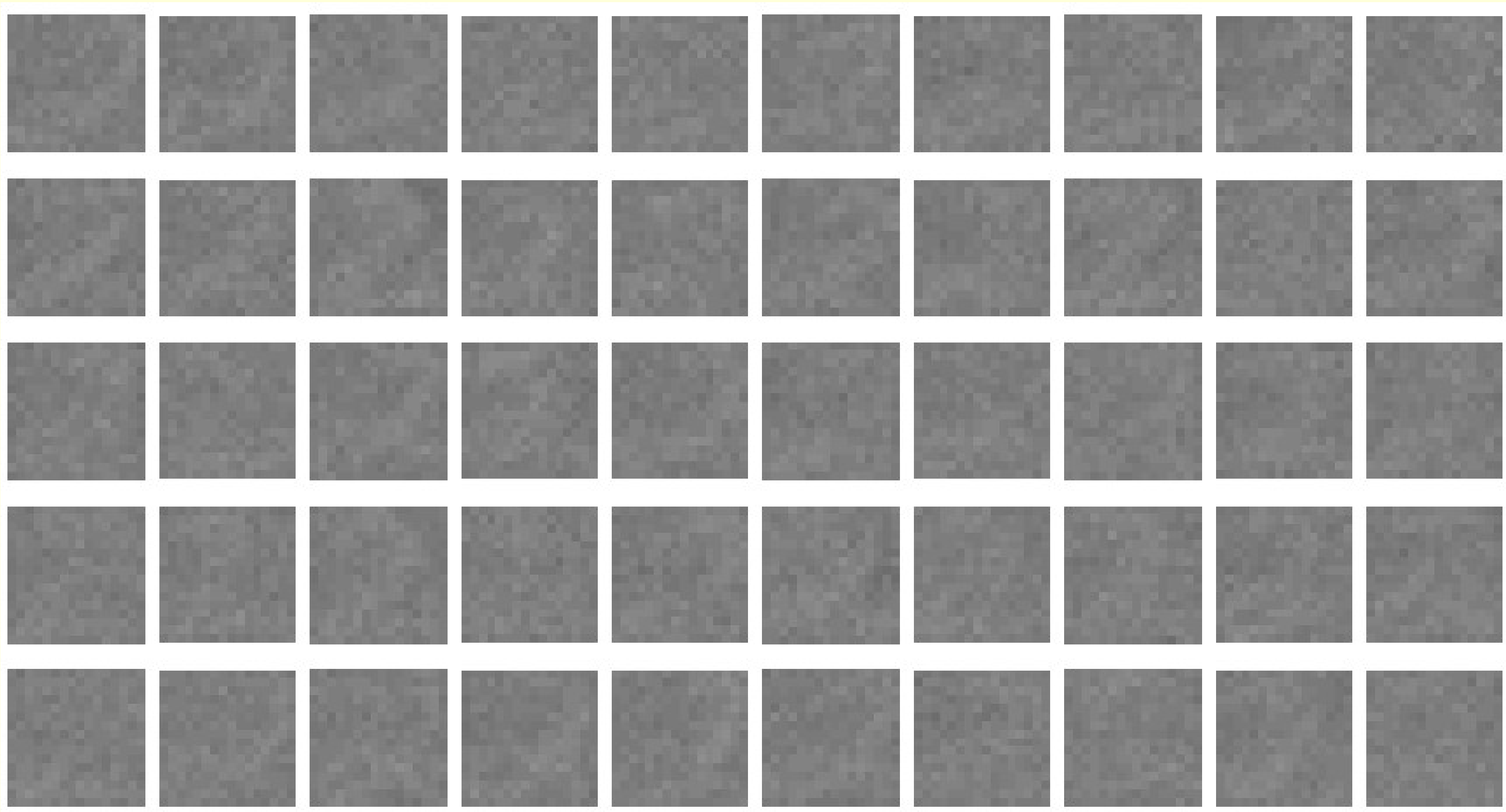


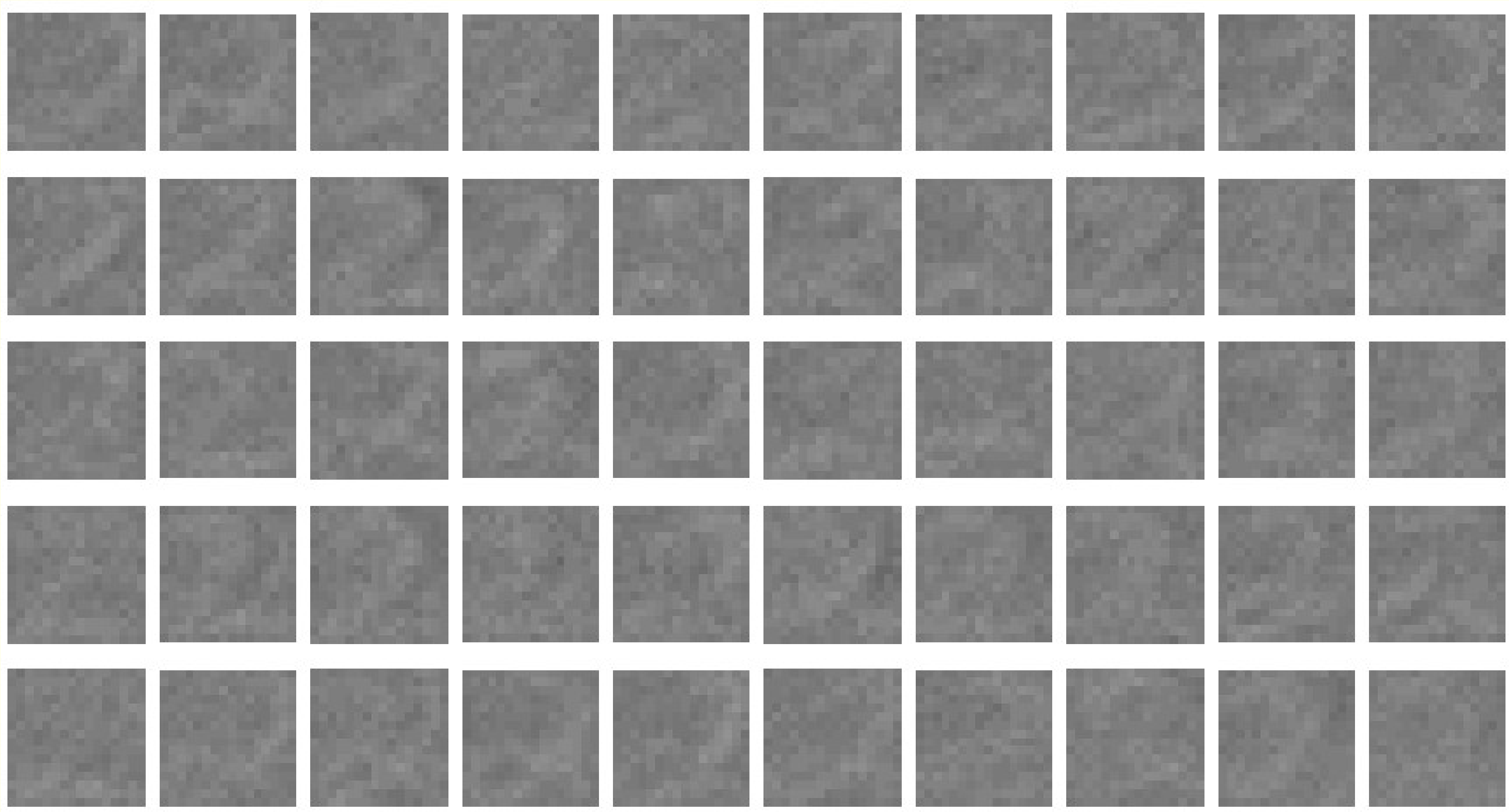
Bartlett

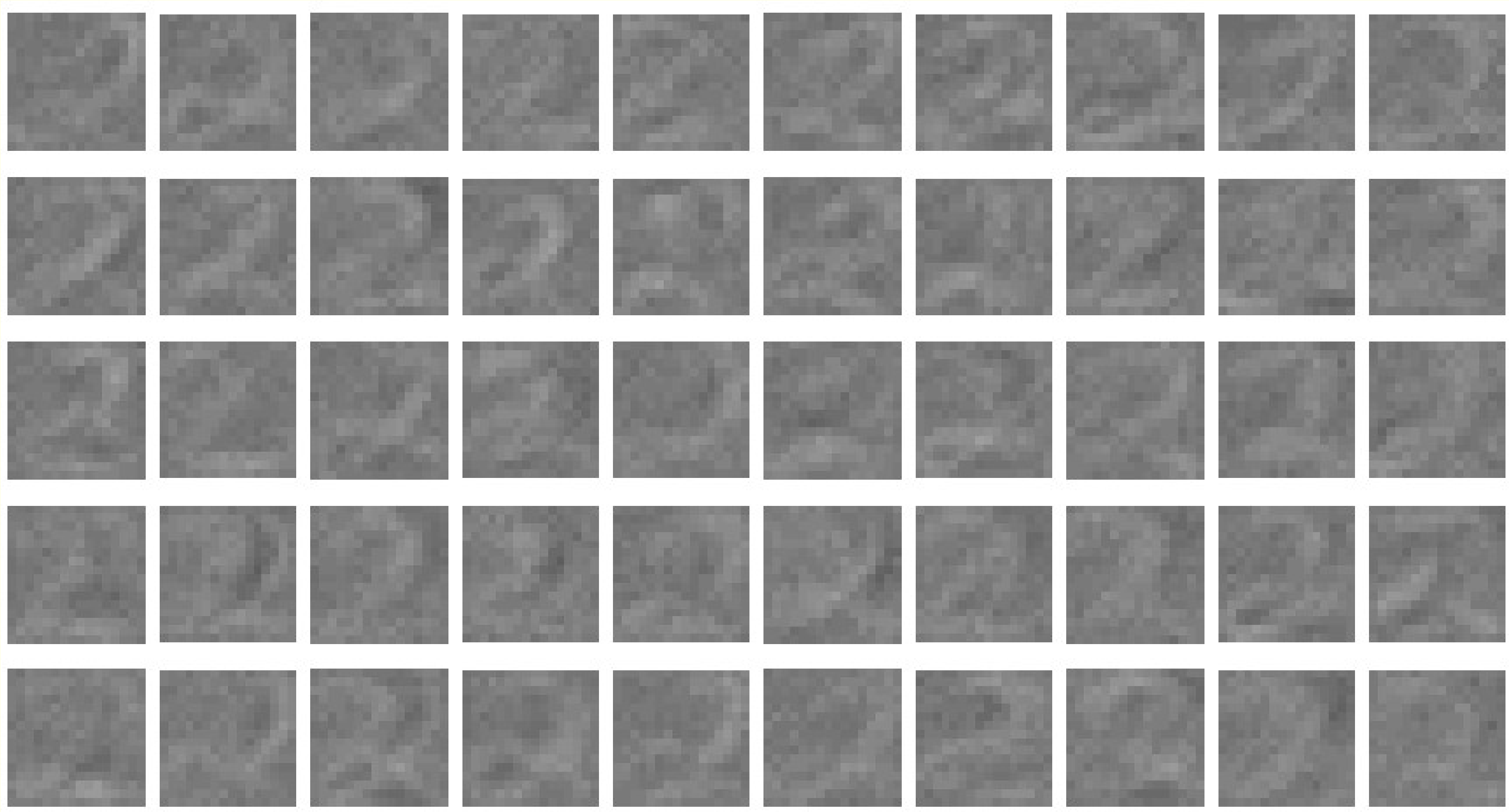
The weights of the 50 feature detectors

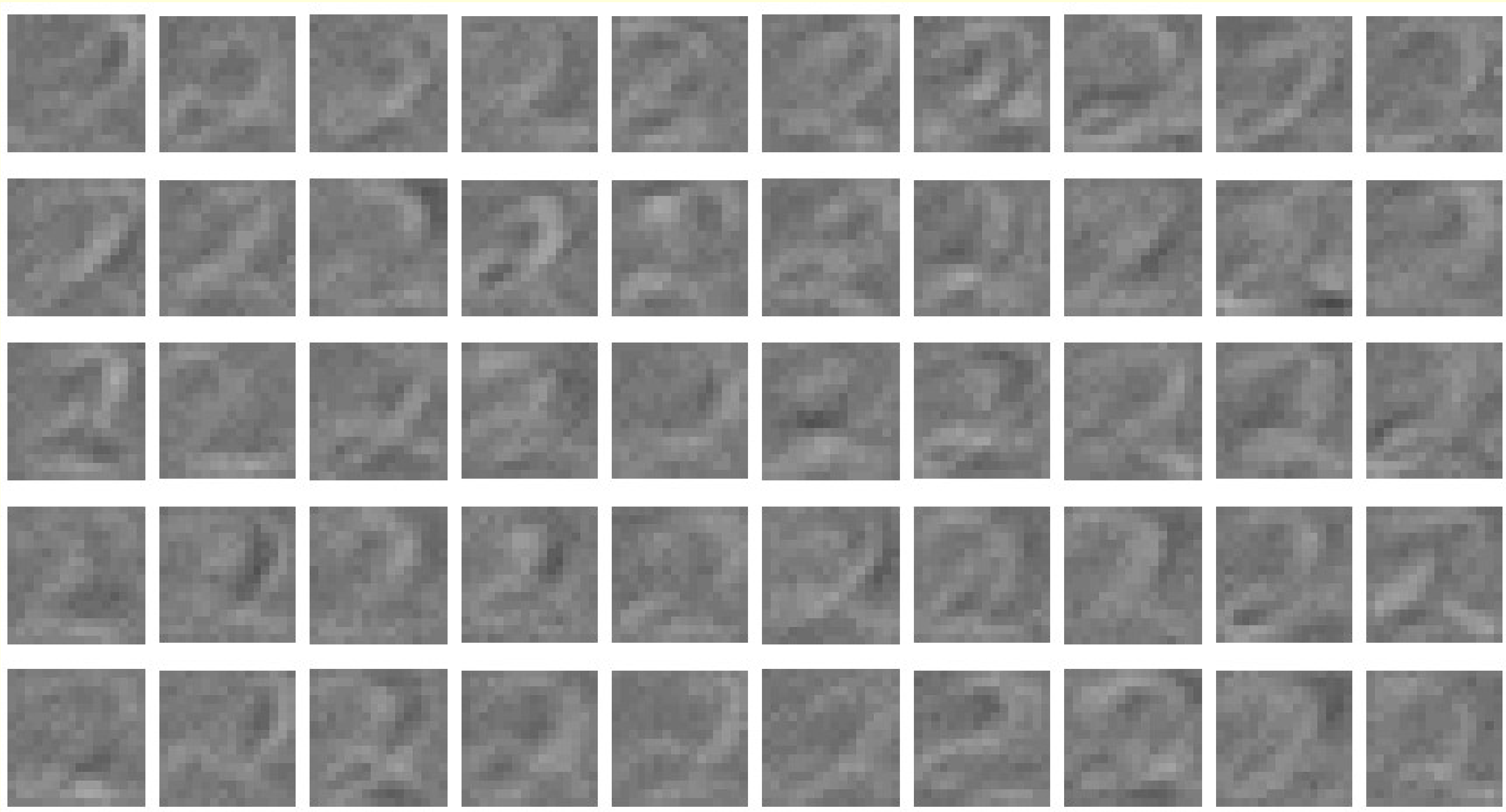


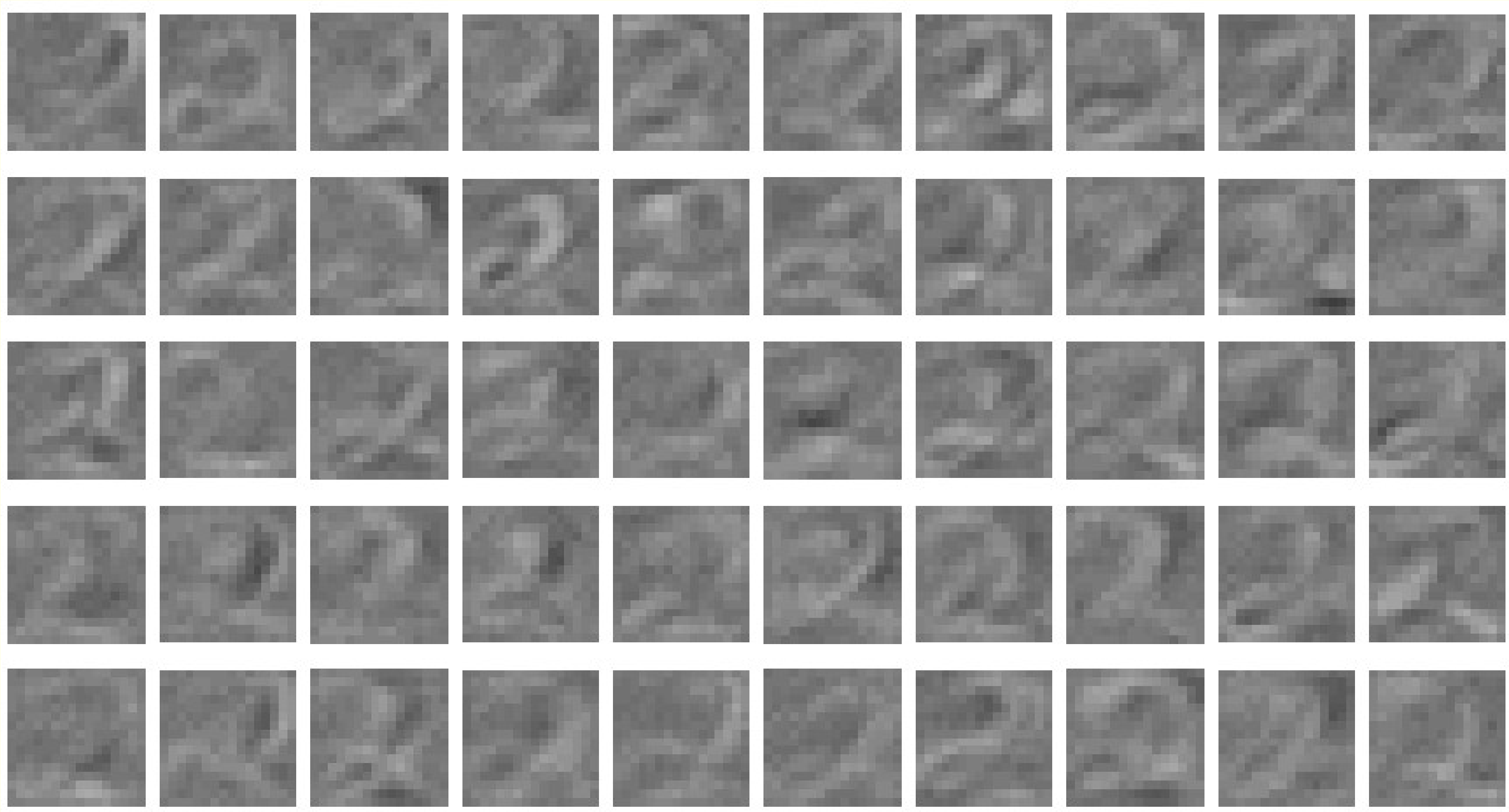
We start with small random weights to break symmetry

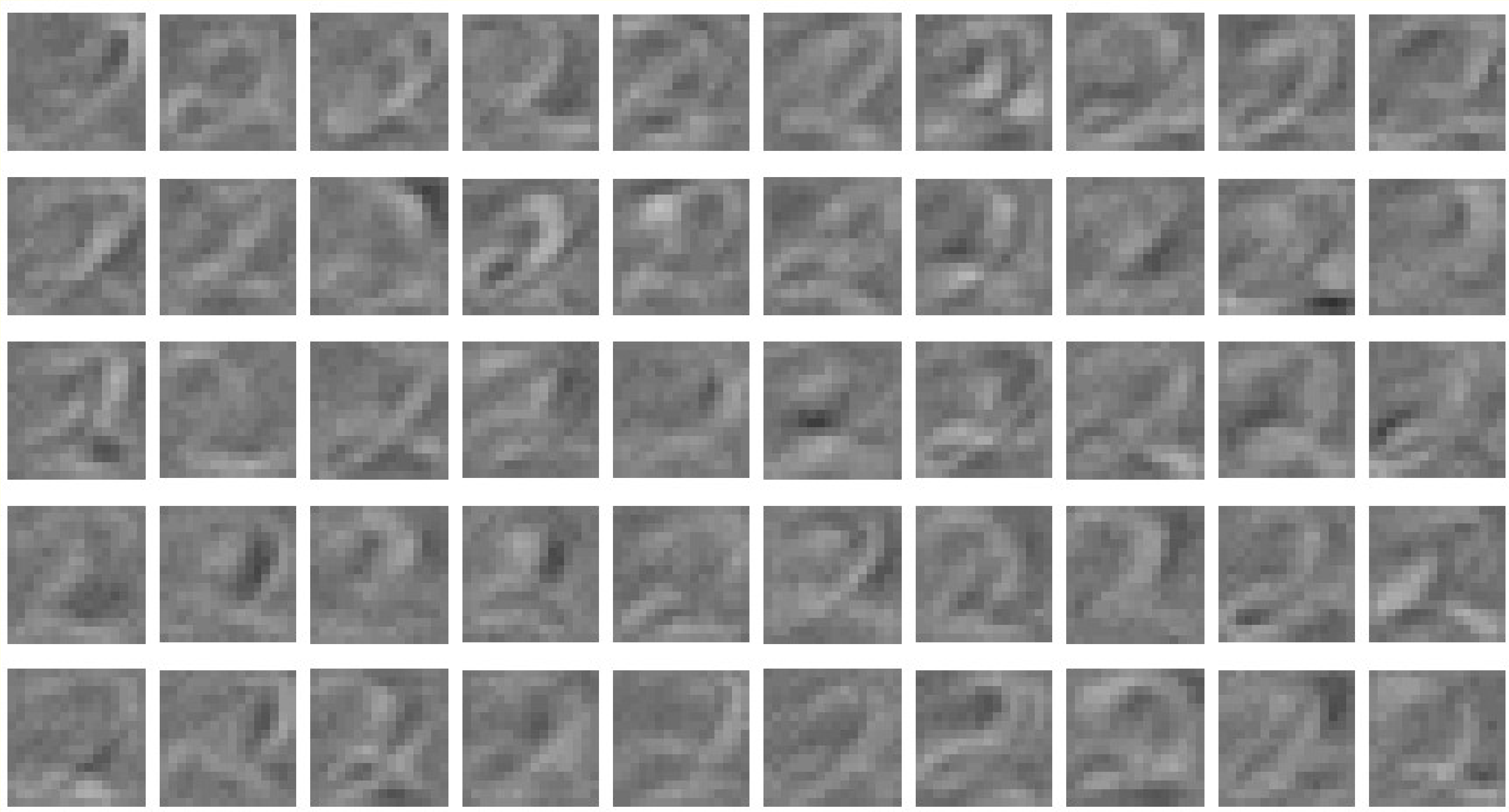


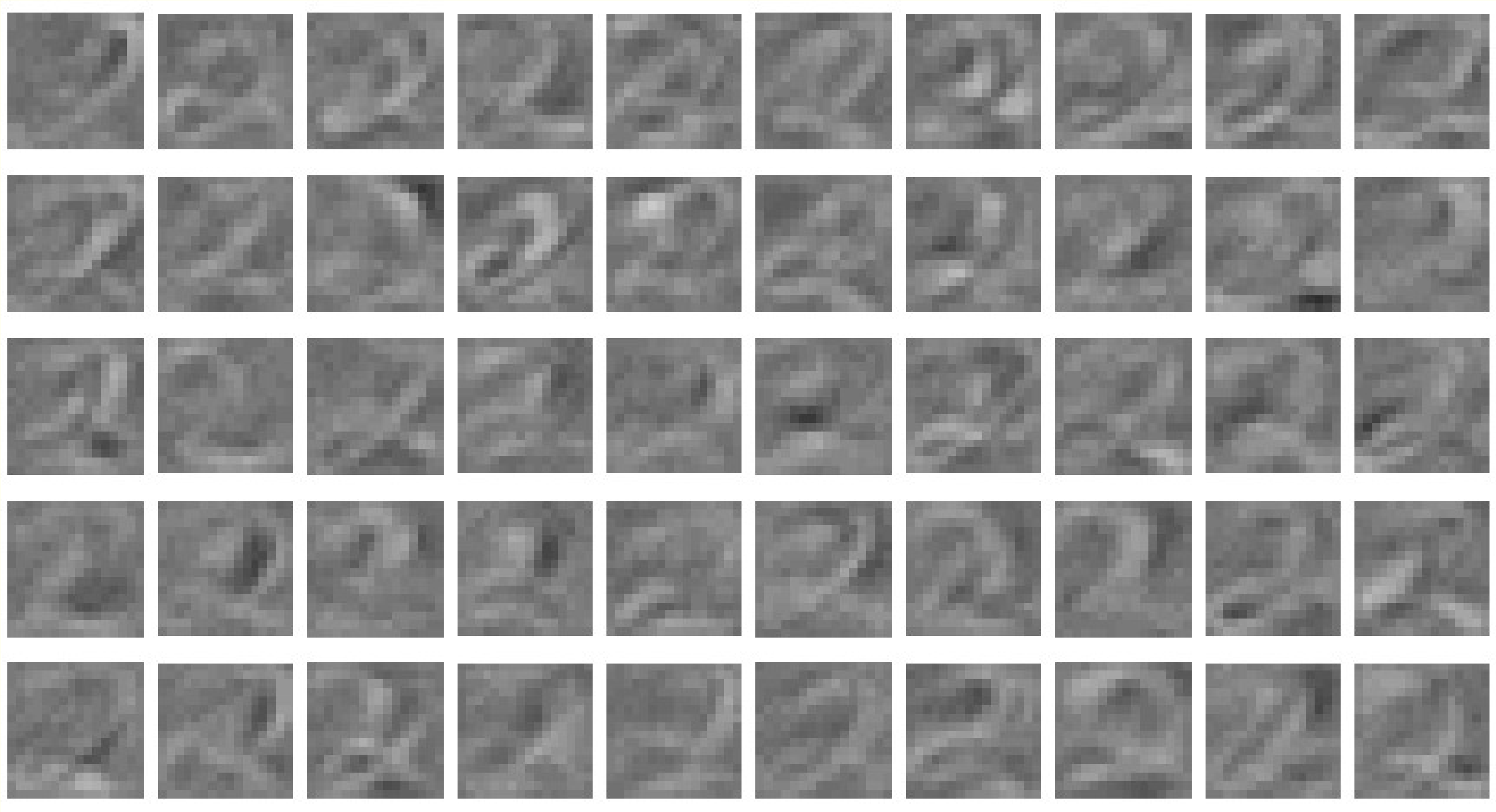


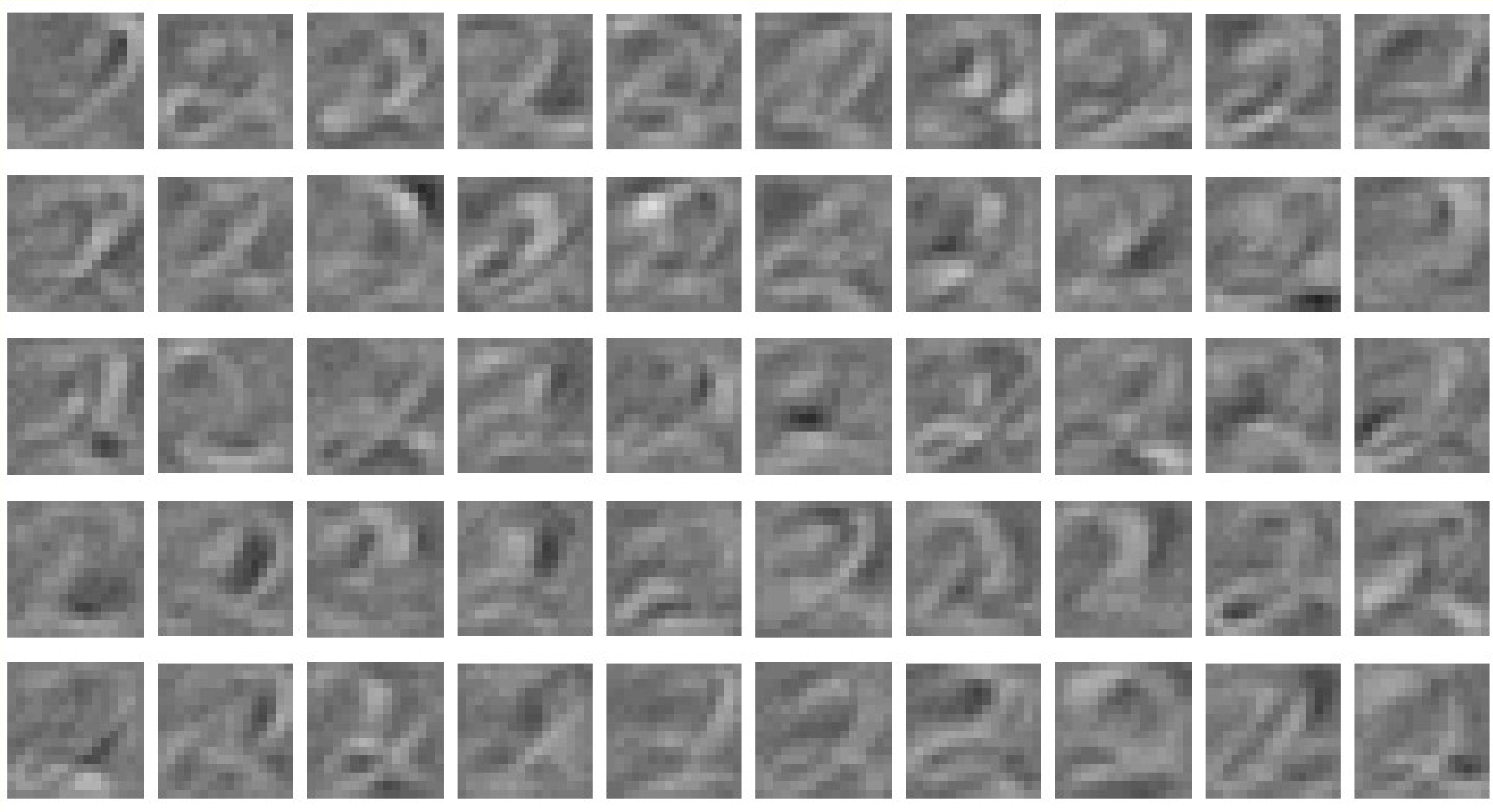


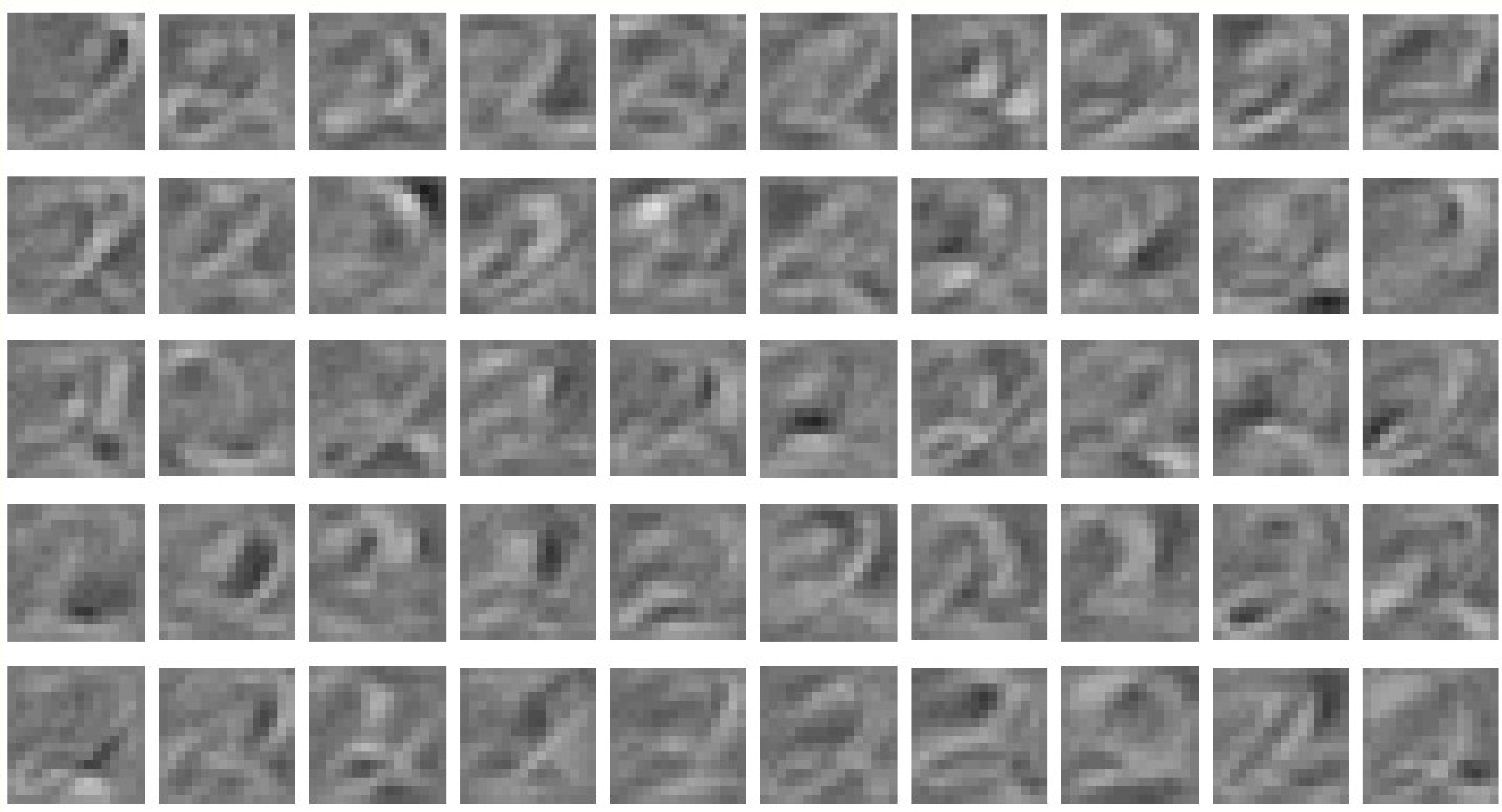


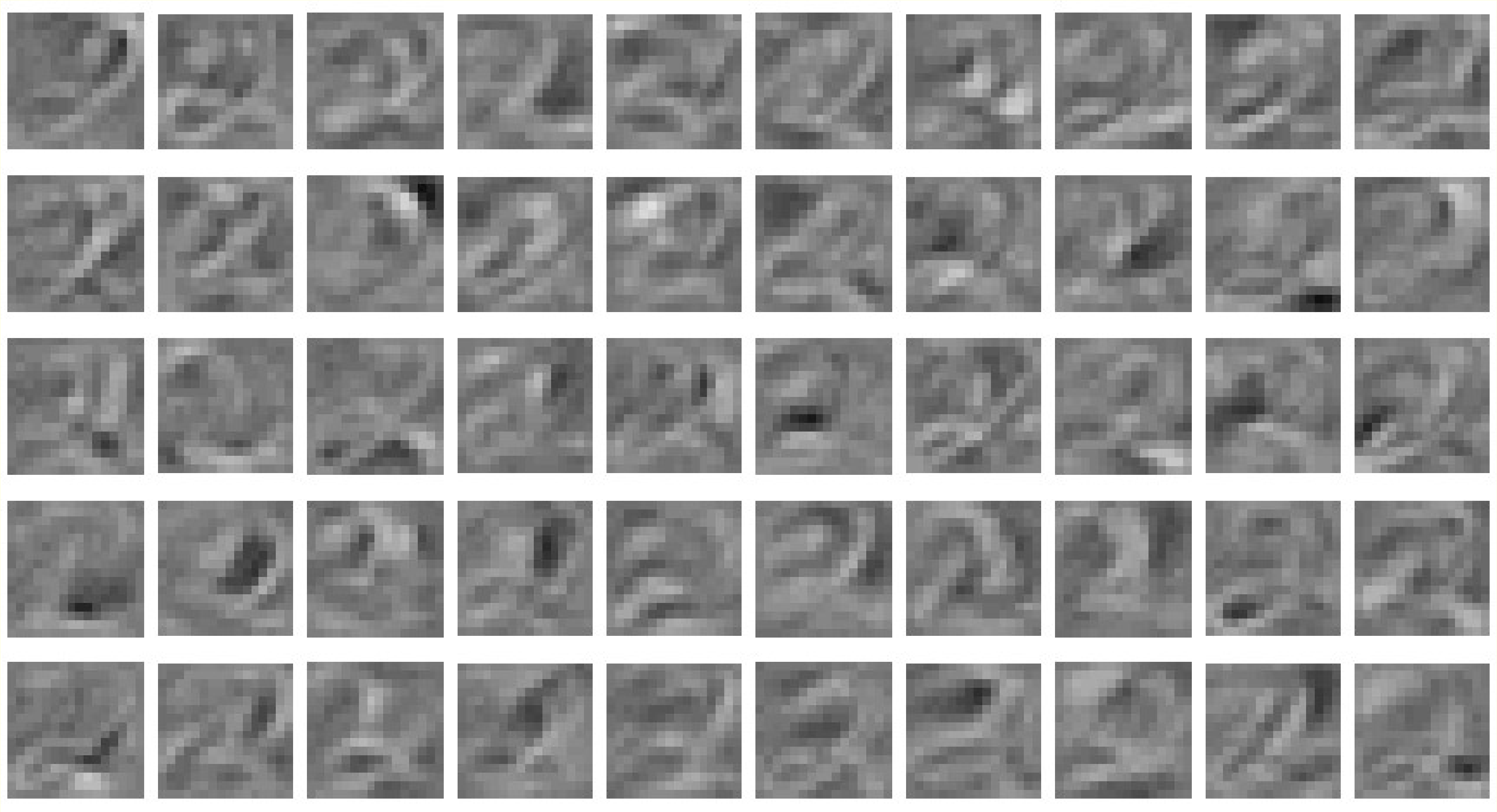


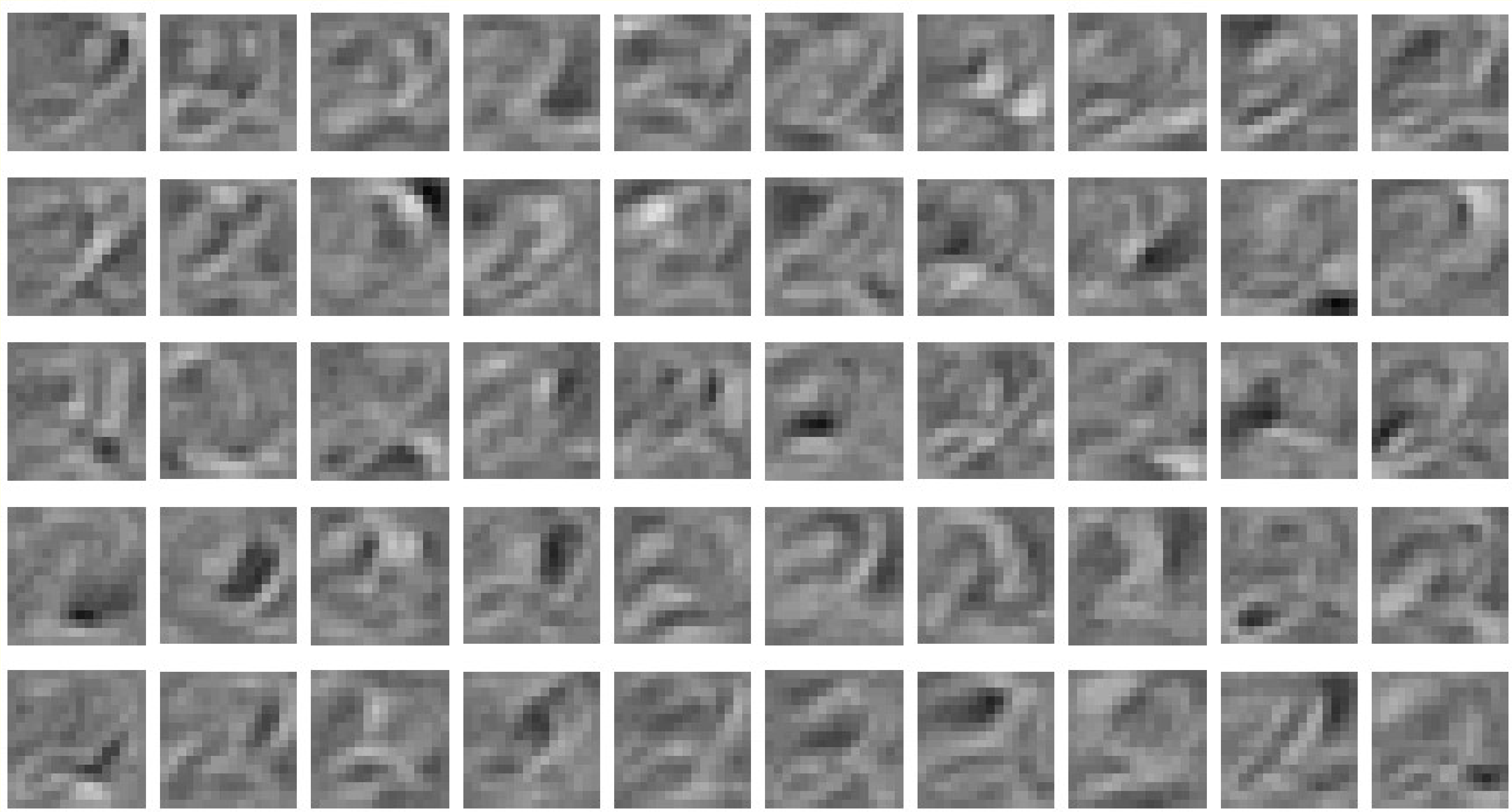


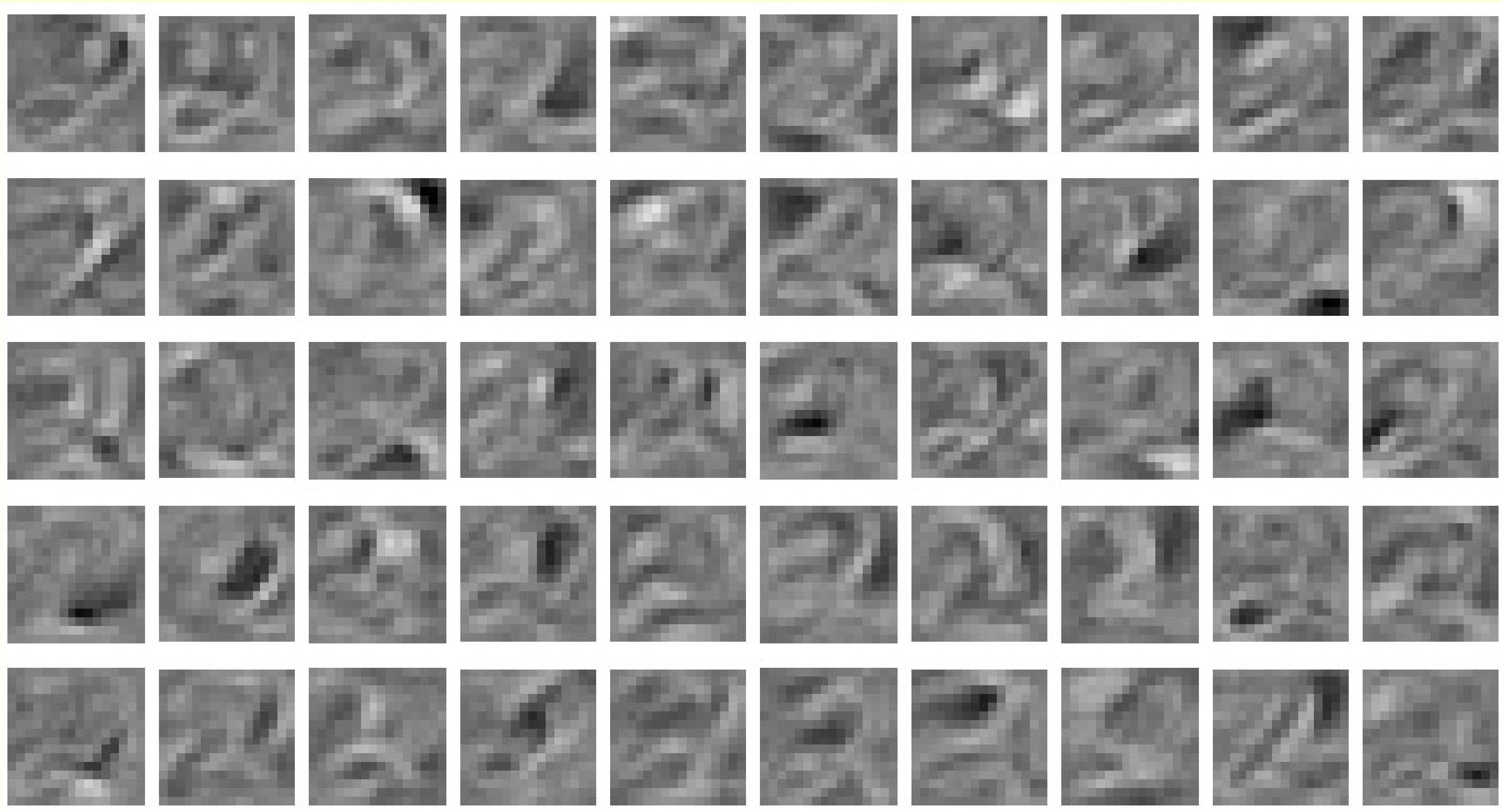


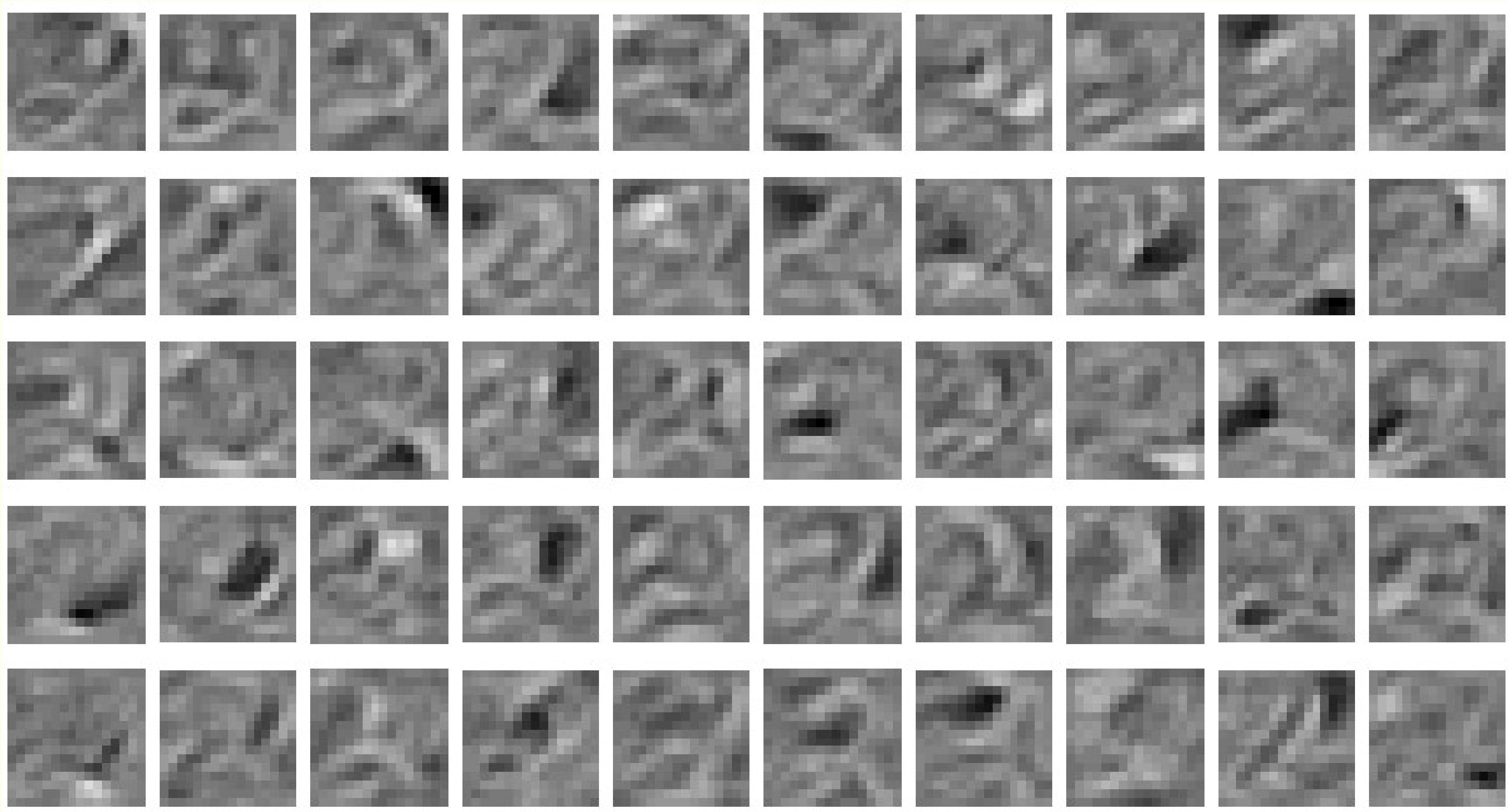


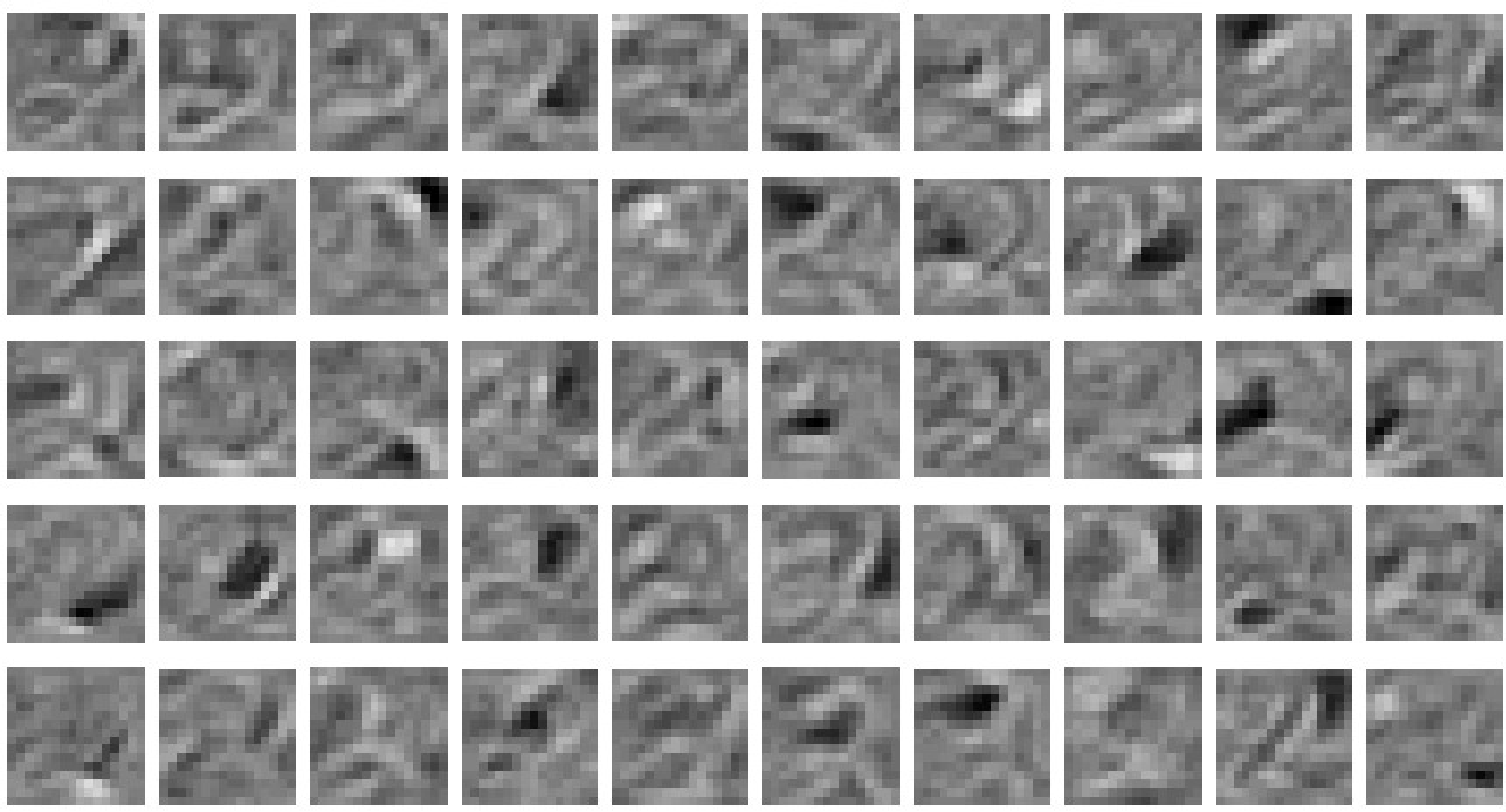


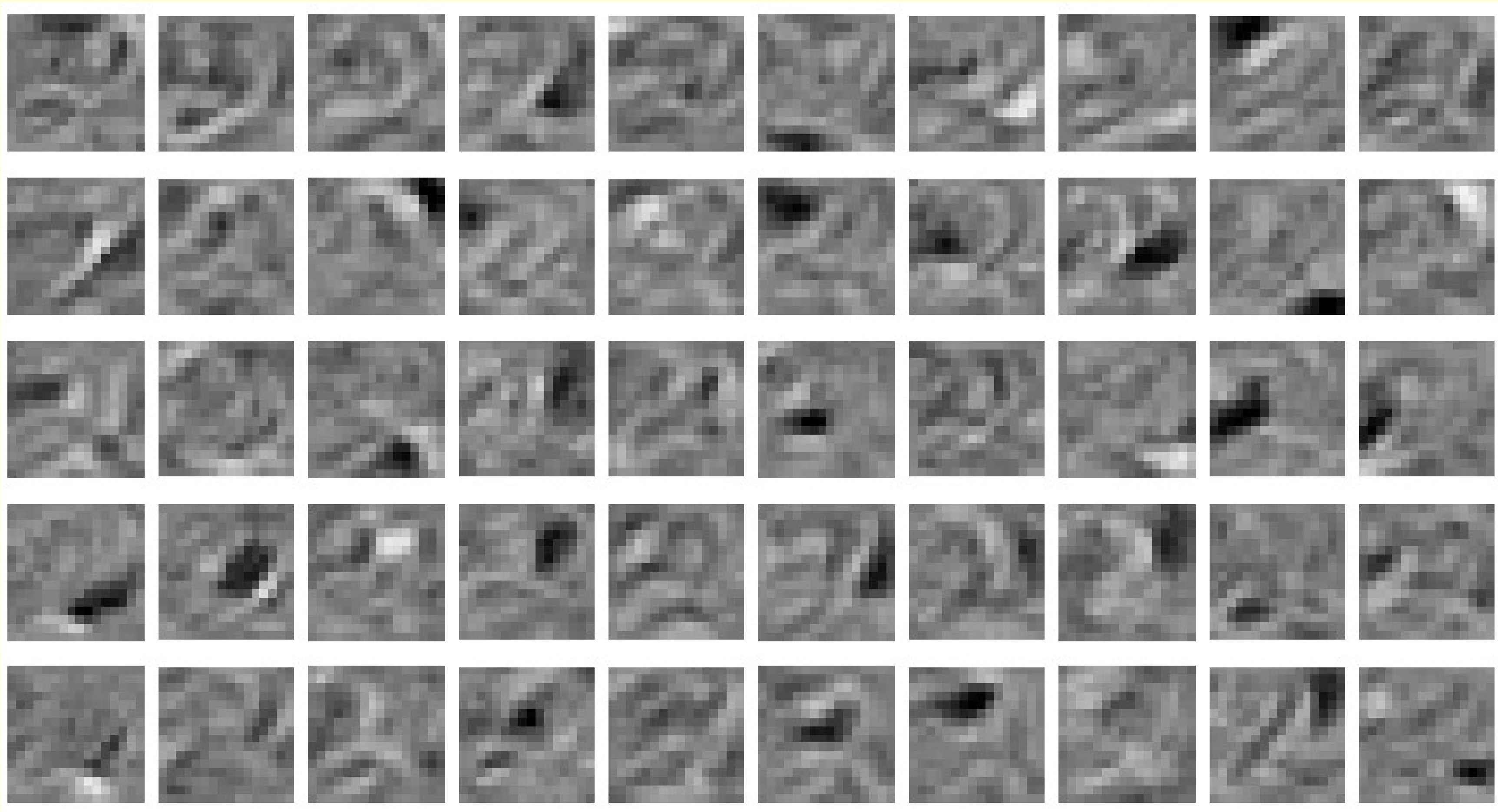


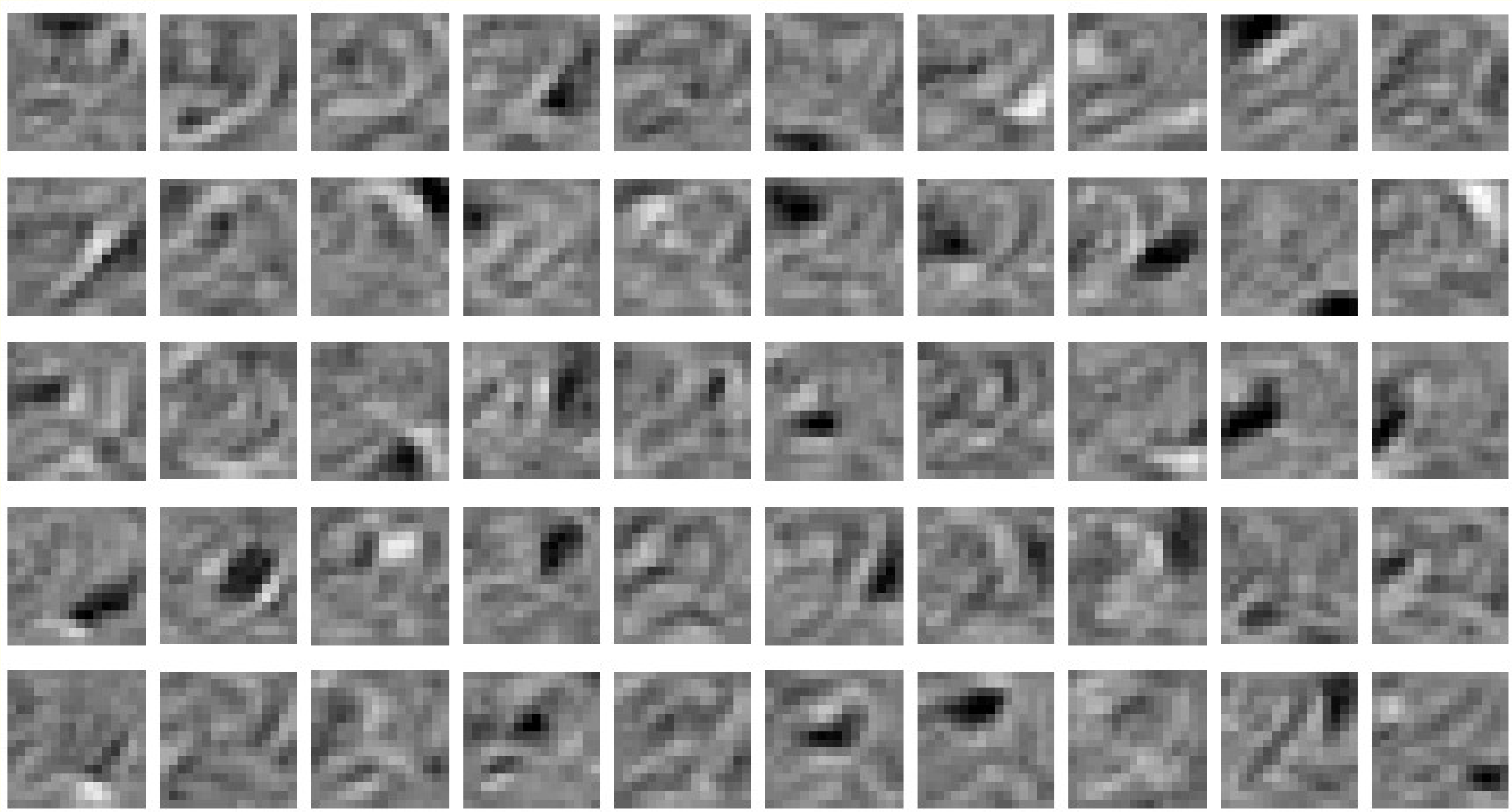




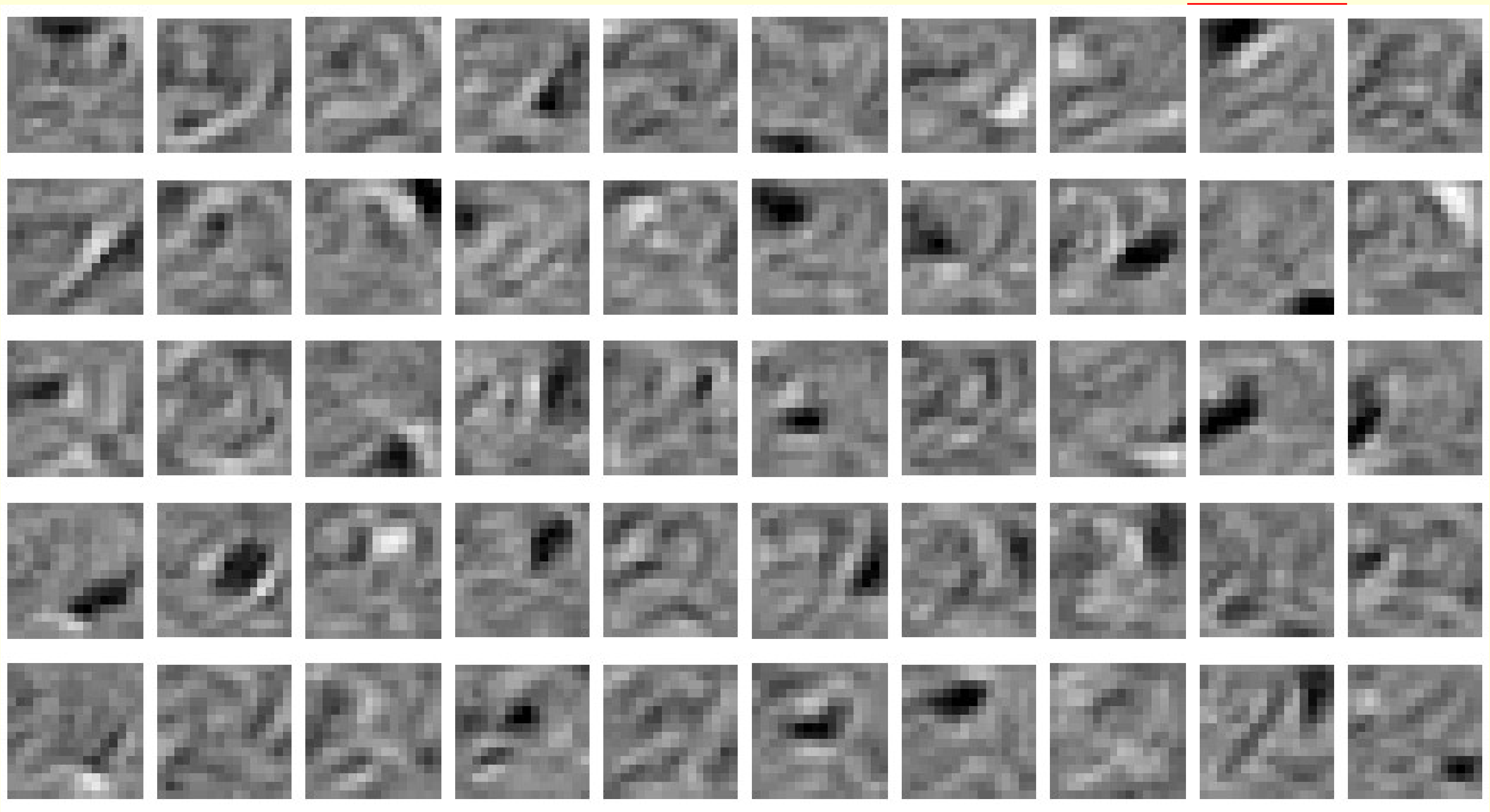




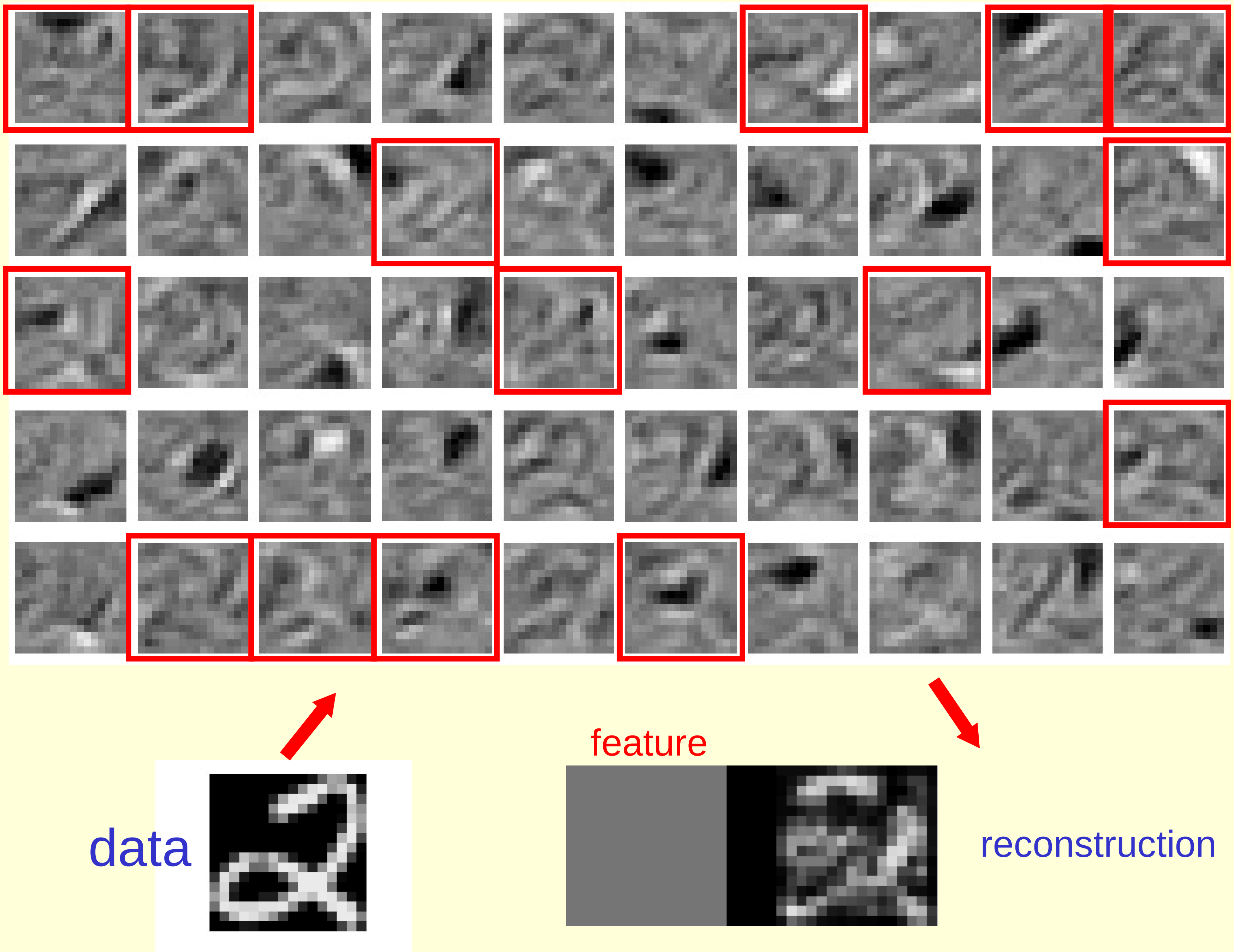




The final 50 x 256 weights



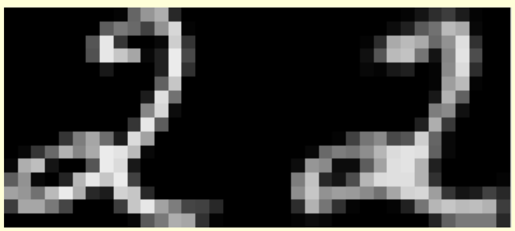
Each neuron grabs a different feature.


feature
data
reconstruction

How well can we reconstruct the digit images from the binary feature activations?

Data

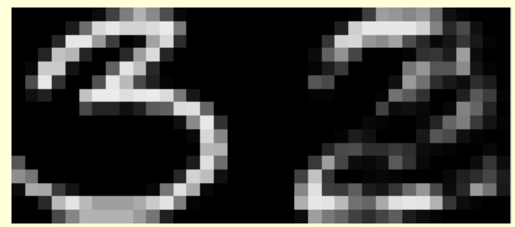
Reconstruction from activated binary features



New test images from the digit class that the model was trained on

Data

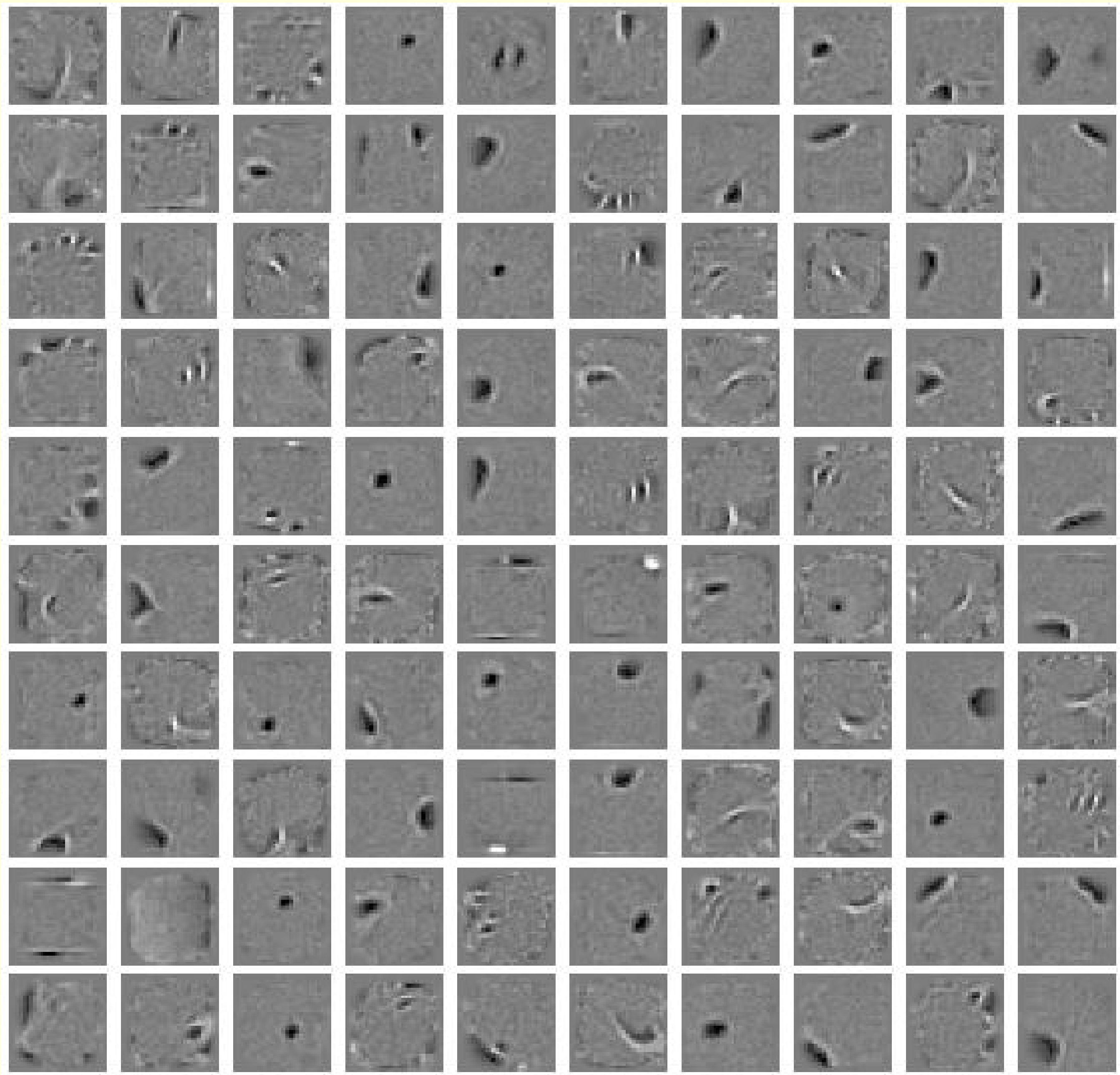
Reconstruction from activated binary features



Images from an unfamiliar digit class (the network tries to see every image as a 2)

Show the movies that windows 7 refuses to import even though they worked just fine in XP

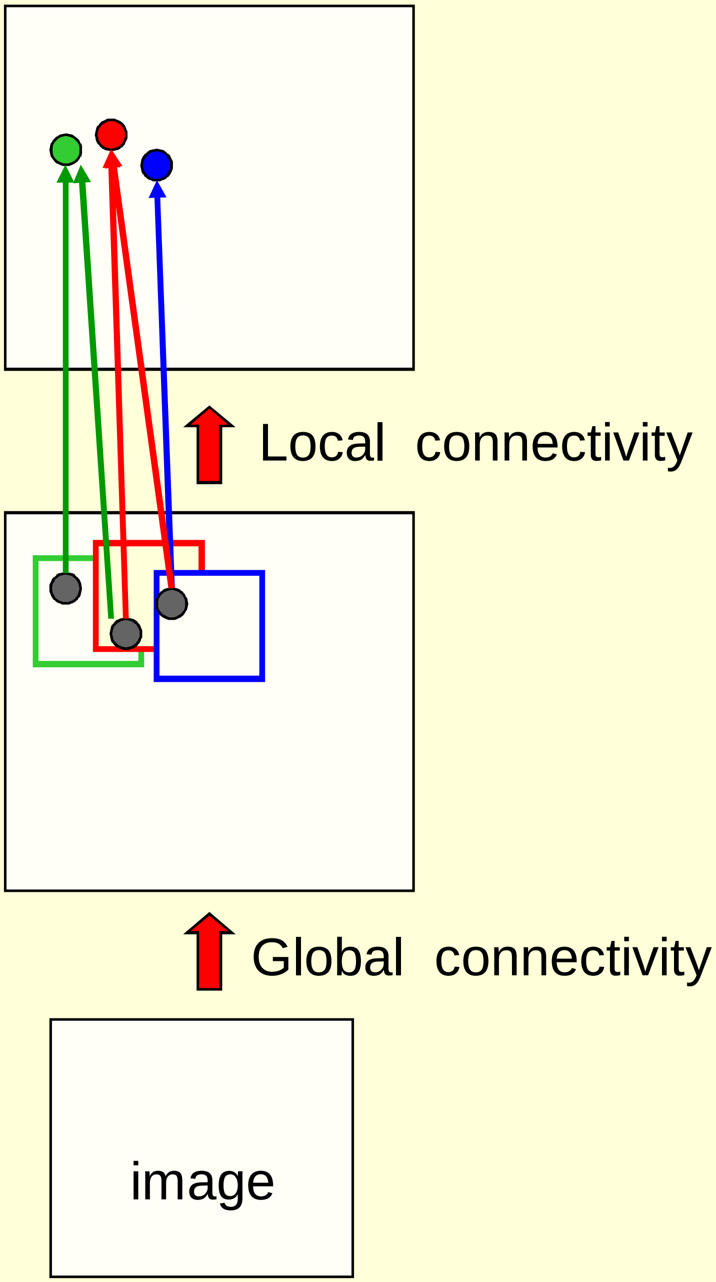
Some features learned in the first hidden layer for all digits



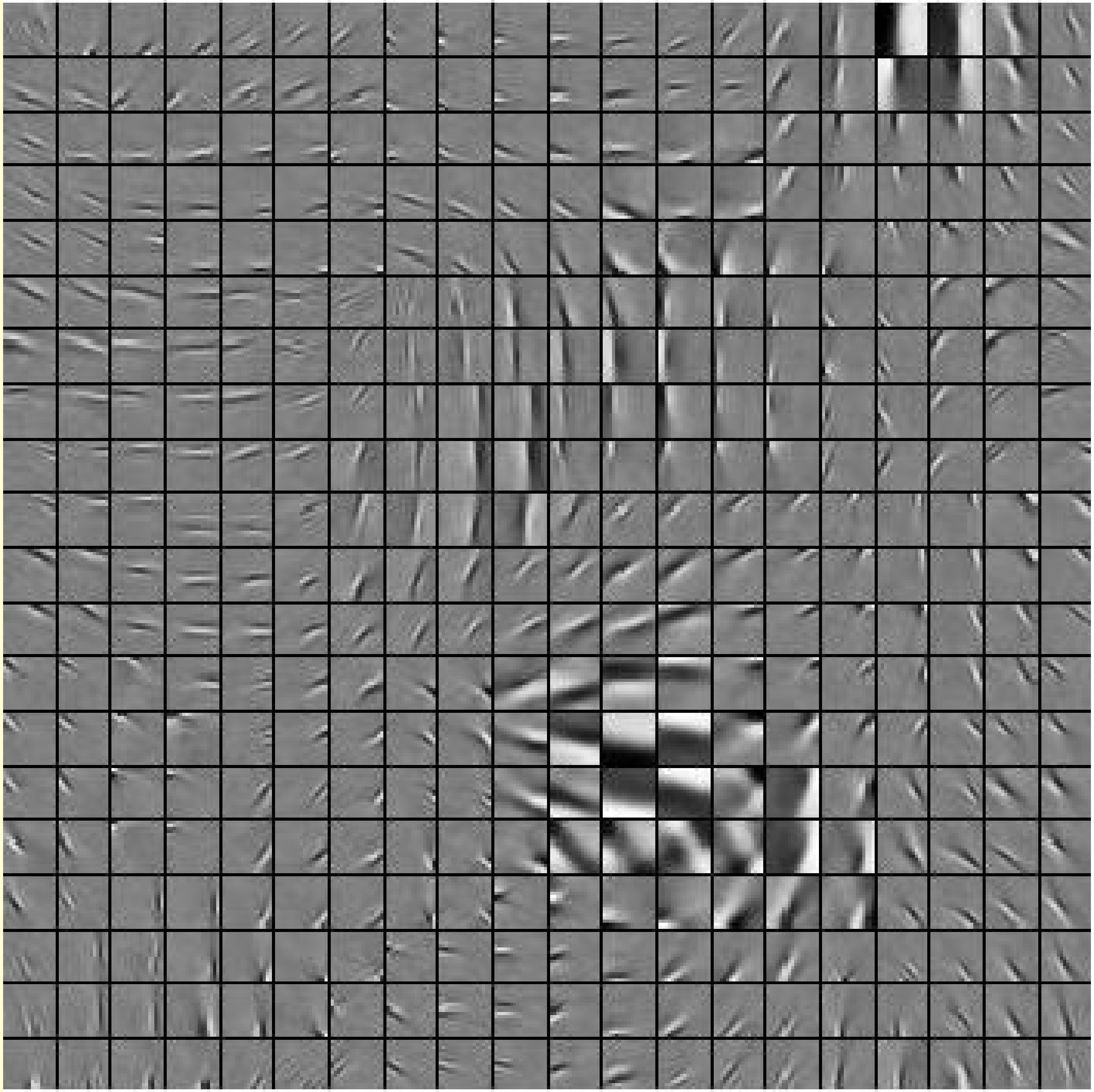
And now for something a bit more realistic

- Handwritten digits are convenient for research into shape recognition, but natural images of outdoor scenes are much more complicated.
  - If we train a network on patches from natural images, does it produce sets of features that look like the ones found in real brains?
  - The training algorithm is a version of contrastive divergence but it is quite a lot more complicated and is not explained here.

A network with local connectivity



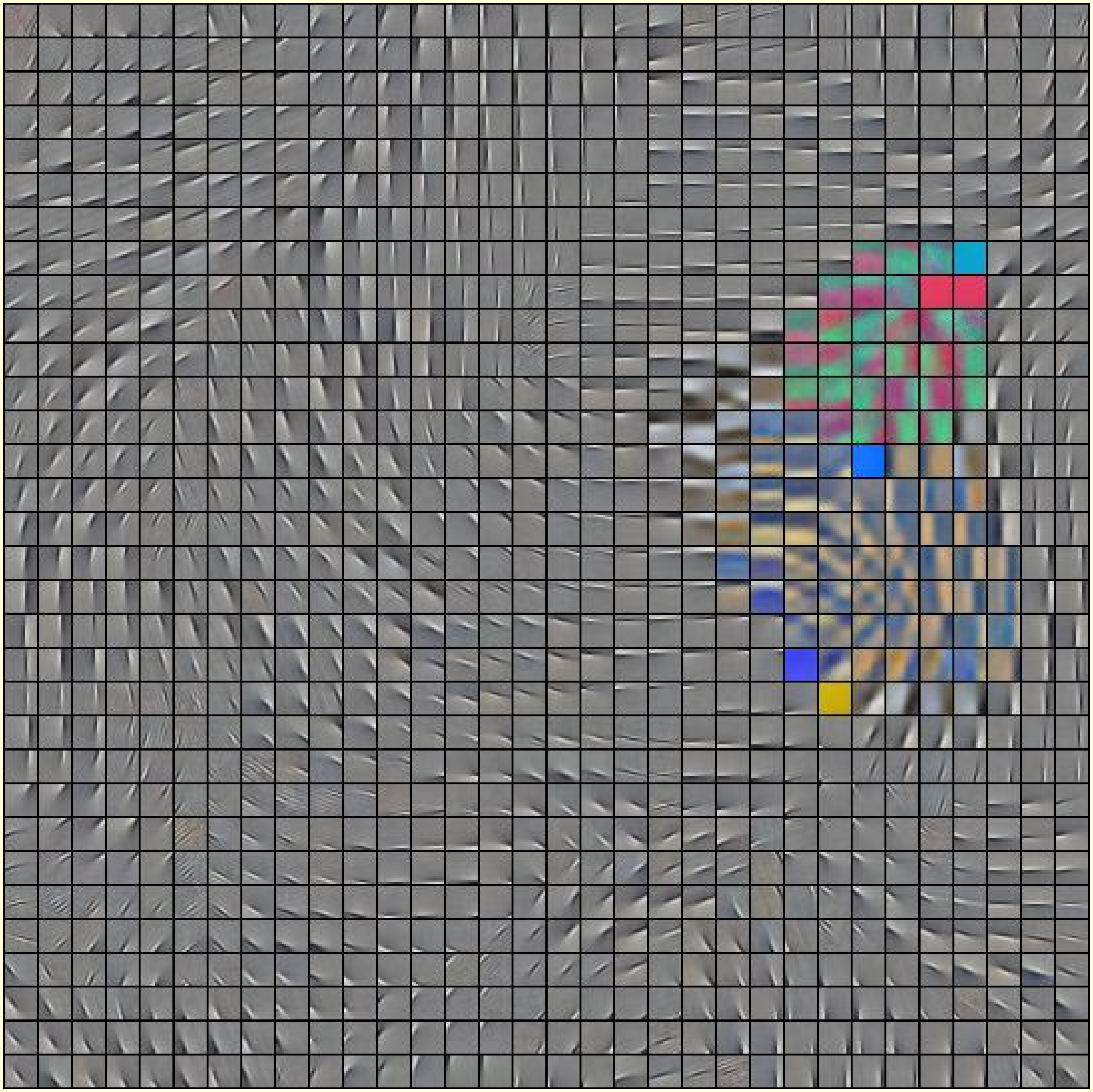
The local connectivity between the two hidden layers induces a topography on the hidden units.



Features learned by a net that sees 100,000 patches of natural images.

The feature neurons are locally connected to each other.

Osindero, Welling and Hinton (2006) Neural Computation



Filters learned for color image patches by an even more complicated version of contrastive divergence. Color “blobs” consisting of red-green and yellow-blue filters are found in monkey cortex.

Where do they come from?