DISENTANGLING IMAGES WITH LIE GROUP TRANSFORMATIONS AND SPARSE CODING

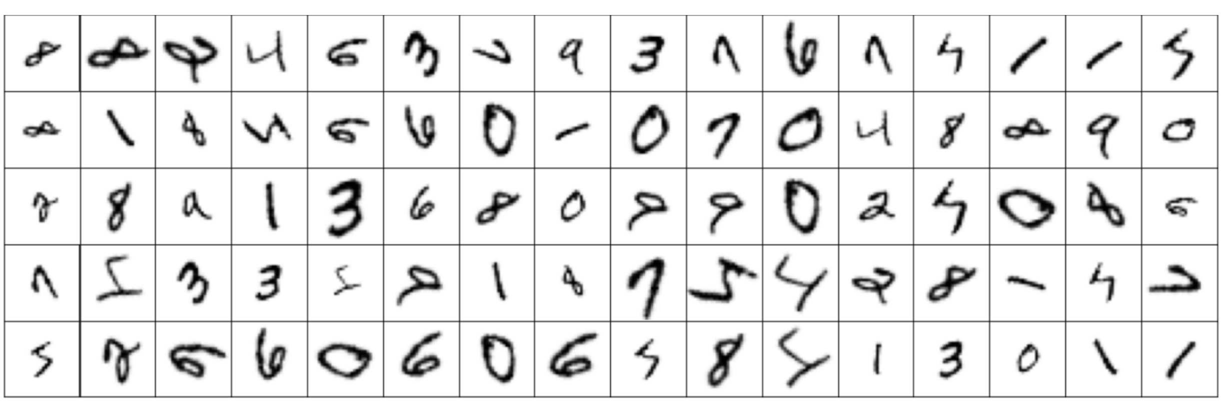
Report

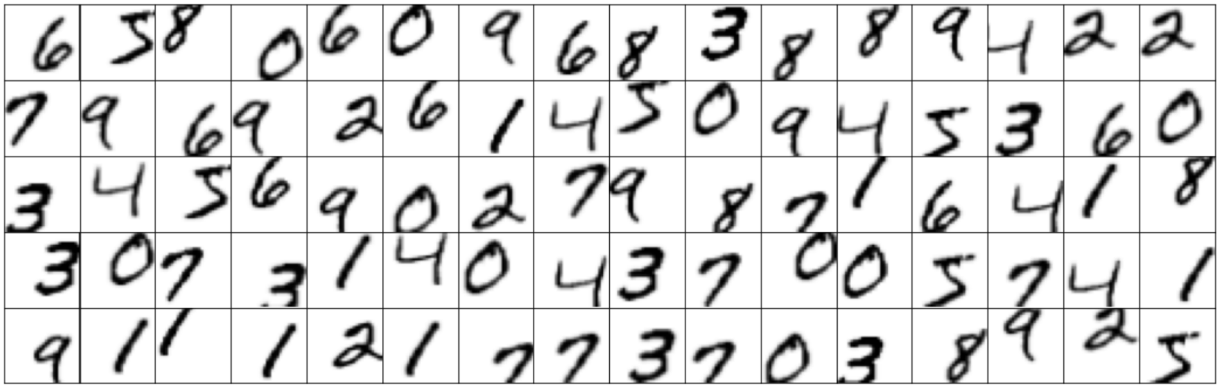
Aim and Experimentation

A synthetic dataset is created from MNIST with transformations and translations such that it is learned by the model. There is initial supervision provided for the perturbed dataset. The aim is to prove that model learns different transformations and translations in an unsupervised manner when provided with limited rotations for training. In other words, we want that our model to automatically learn different transformations such that the model can recognize and classify any form of rotational transformation applied to digit images.

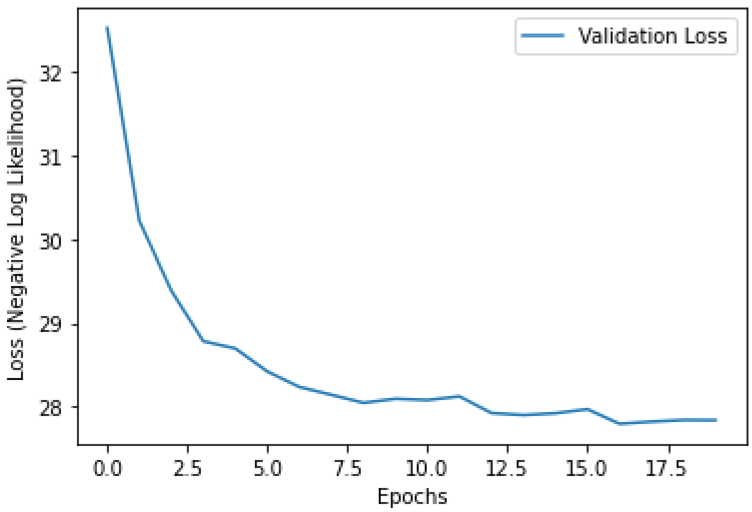
Methodology and Results

Parameters n and K are set before such that n is the number of datasets that need to be created and K is the number of MNIST classes that need to be considered. For example, for n=2 and K=10, we pick 2 images from each of the 10 classes. In other words, we will have 1 image for each MNIST class totaling 10 images for n=1 and then apply transformations and translations such that you get 6000 images for each MNIST class. Hence totaling 60000 images in the first dataset. Again, you will have 60000 images for the second dataset which was different transformations and translations of the second picked 10 images from the MNIST dataset. Now again for the sake of populating the dataset, the author creates the dataset from each of the original images from the MNIST dataset such that for each image you have 6000 different sized, scaled, rotated, perturbed images. Rotation is drawn uniformly between -75° and 75° , while scaling is drawn uniformly between 0.5 and 1.0

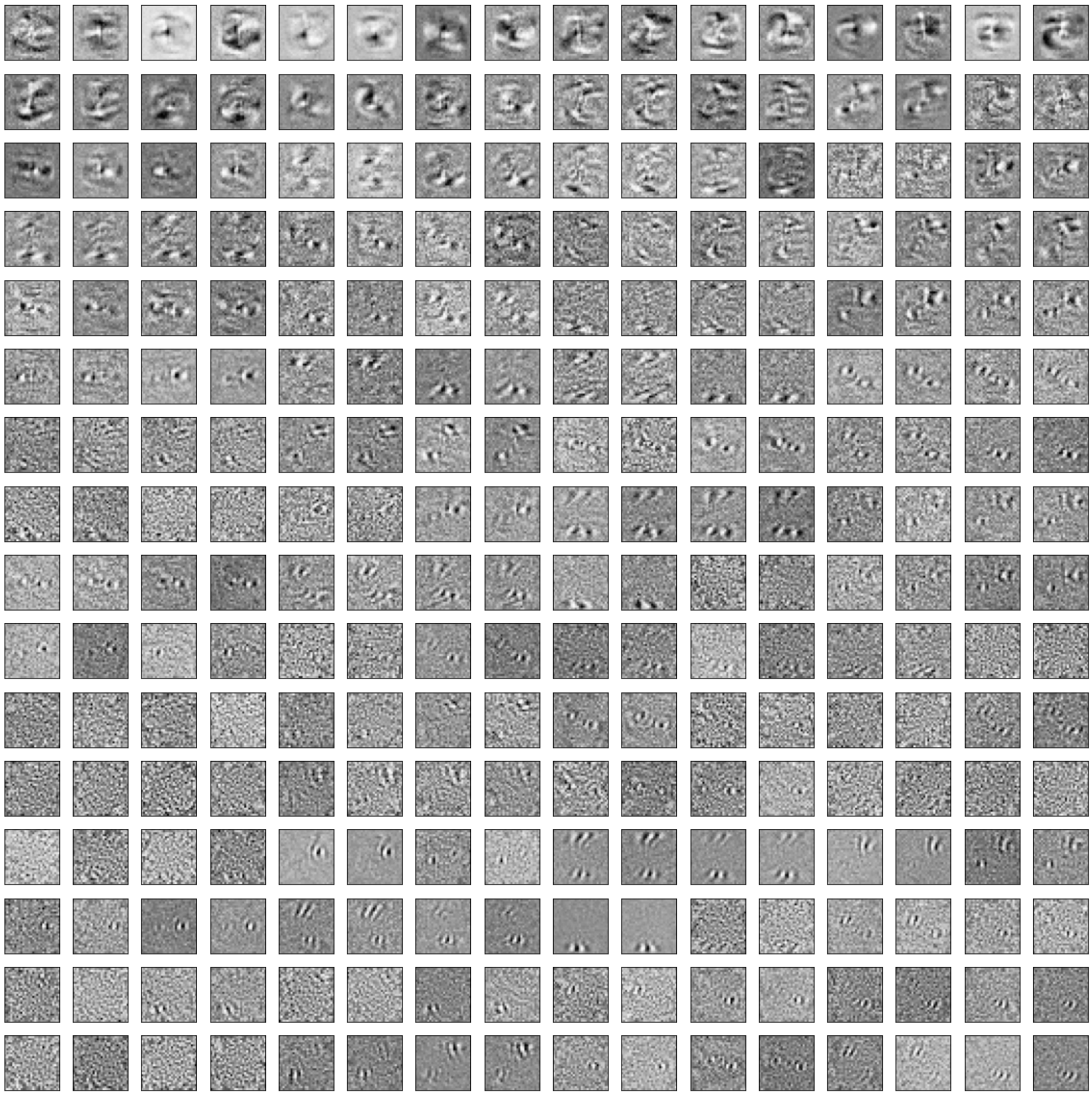




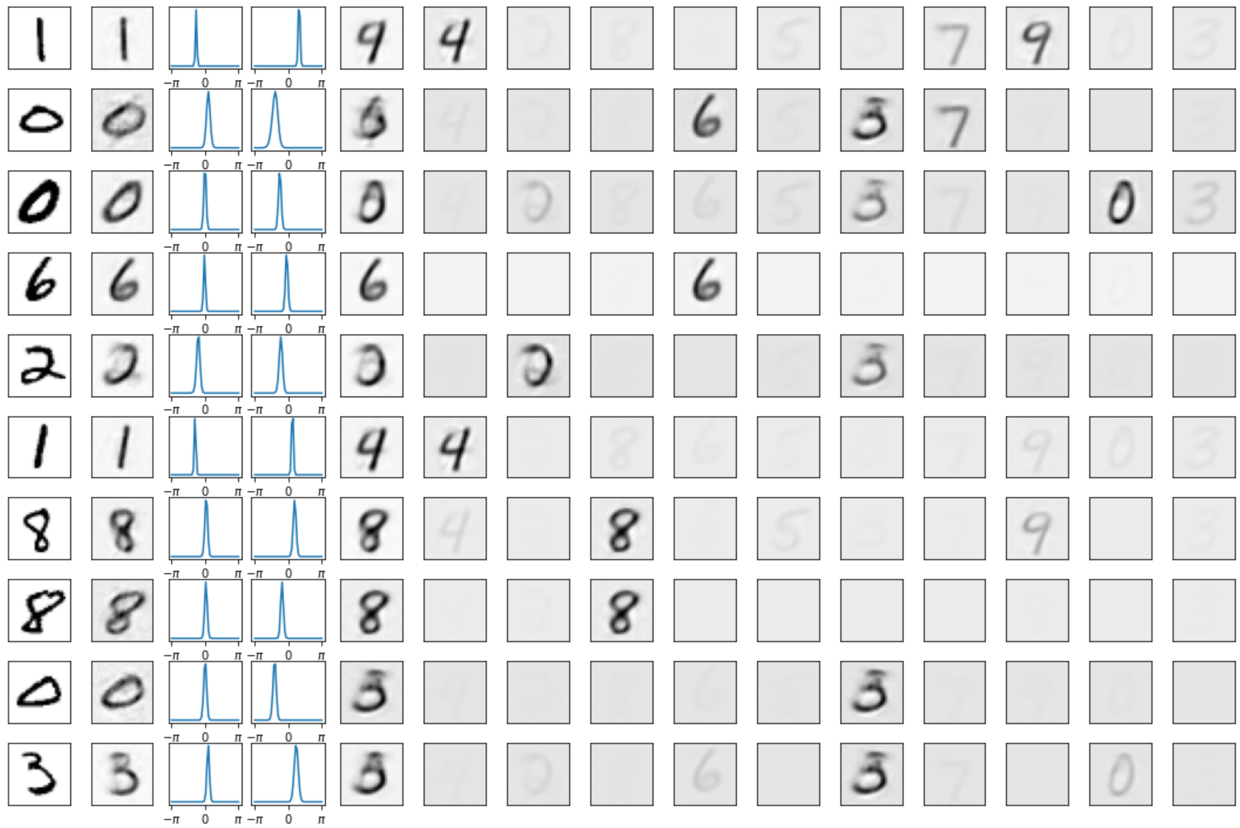
In the images above, the first set of 80 images on the previous page is the result of ROTATION + SCALING while the second set of 80 images is the result of only TRANSLATION of the dataset. Now train using the datasets and plot the curve for loss.



The Lie group sparse coding classifier is able to learn 10 digits of the MNIST classes. In addition, W matrices and PSI dictionary are visualized to check the learned representations. Even though the rotation + scaling dataset contains only rotations between -75° and 75°, the model learns the full 360° rotation. This ability to generalize and extrapolate correctly the transformation present in the dataset is a feature of the Lie group structure that is built into LSC.



The image above shows learned W matrices for each of the digits. With the help of these visualizations, we want to see that how MNIST and perturbed MNIST datasets look and model learns what kind of weights for the learning algorithm. In other words, each of the deep visualizations depicts the inner representations for each of the digits. For only 2D translations, the weights visualization looks like horizontal and vertical lines however, for rotations and transformed datasets the deep visualization looks like spiral deep dreams. (for rotation and scaling both) In the image above, initial images are for 0 transformations MNIST original images while the later images are of increasing transformations and translations along with scaling.



In this image first column shows original MNIST images of the digits, the second column shows reconstructed images of the sparse encoding. The peaks in the graphs of the third and fourth columns show the transformations parameters from -pi to +pi such that the values on the x-axis denote the transformation value or the angle. An image I is given to the network, which then performs the inference procedure given in the earlier section to yield the MAP estimate of the sparse coefficients αˆ and the posterior distribution of the transformation parameter P. In addition, the rest of the columns shows different reconstructions of that particular MNIST class for the row. For example, the third row has zero class. Then the most highlighted digit is expected to be 0 and the reconstruction of the digit is also coming out to be 0 as the optimal one (the one which is in sharp colors row 3 column 14). In the same figure, we can also observe that the transformation value T(s) is being changed in each row which is chosen randomly.

Another model is trained for sparse coding to compare with Lie group Sparse coding (LSC). It was empirically tested that LSC performs well than sparse coding for the MNIST dataset and permuted MNIST dataset. Additionally, LSC can classify and reconstruct MNIST digits even for the unseen transformations on which it was not trained.