TOWARD LEARNING LATENT-VARIABLE REPRESENTATIONS OF MICROSTRUCTURES BY OPTIMIZING IN SPATIAL STATISTICS SPACE

Sayed Sajad Hashemi
Dept. of Mechanical & Industrial Engineering
University of Toronto
sajjads.hashemi@gmail.com

Michael Guerzhoy
Division of Engineering Science and
Dept. of Mechanical & Industrial Engineering
University of Toronto
guerzhoy@cs.toronto.edu

Noah Paulson
Argonne National Laboratory
npaulson@anl.gov

ABSTRACT

In Materials Science, material development involves evaluating and optimizing the internal structures of the material, generically referred to as microstructures. Microstructures structure is stochastic, analogously to image textures. A particular microstructure can be well characterized by its *spatial statistics* Paulson et al. (2017), analogously to image texture being characterized by the response to a Fourier-like filter bank Varma & Zisserman (2002). Material design would benefit from low-dimensional representation of microstructures Paulson et al. (2017).

In this work, we train a Variational Autoencoders (VAE) to produce reconstructions of textures that preserve the spatial statistics of the original texture, while not necessarily reconstructing the same image in data space. We accomplish this by adding a differentiable term to the cost function in order to minimize the distance between the original and the reconstruction in spatial statistics space.

Our experiments indicate that it is possible to train a VAE that minimizes the distance in spatial statistics space between the original and the reconstruction of synthetic images. In future work, we will apply the same techniques to microstructures, with the goal of obtaining low-dimensional representations of material microstructures.

1 INTRODUCTION

Computer-aided materials design holds the promise of designing materials with favorable properties by enabling researchers to computationally optimize potential materials.

Material microstructure is often represented as a large 3D image. 2D images are often also used. In research, both real images (e.g., X-ray) and simulated microstructures are used.

In "data space," the dimensionality of the microstructure representation is the number of pixels/voxels, which can be quite large. This project is motivated by the need for working in low-dimensional latent space when optimizing material properties. Previous work includes representing the microstructure using the first principal components of the spatial statistics representation Paulson et al. (2017).

In this paper, we modify the Variational Autoencoder (VAE) Kingma & Welling (2014) to learn a low-dimensional representation that preserves the statistical properties of the material by making the original and the reconstruction be similar in spatial statistics space. We work with 2D images.

2 BACKGROUND

In Materials Science, the microstructure m is often characterized using spatial statistics Cecen et al. (2016), $f\left(h, h' \mid \boldsymbol{r}\right) = \frac{1}{\mathrm{Vol}(\Omega_{\boldsymbol{r}})} \int_{\Omega_{\boldsymbol{r}}} m(h, \boldsymbol{x}) m\left(h', \boldsymbol{x} + \boldsymbol{r}\right) d\boldsymbol{x}$.

Here, $m(h, \boldsymbol{x})$ is the local state h (e.g., the crystal lattice orientation) at location $\boldsymbol{x}$ in the microstructure. The spatial statistics (or correlations) can be interpreted as the set of correlations between all locations at distance r, in states h and h' respectively. In practice, the spatial statistics can be computed as the inverse Fourier transform of the magnitudes of the Fourier transform of the microstructure Paulson et al. (2017) Einstein (1914).

Generation of images with a specific texture that is specified by the response to a Fourier-like filter bank (or indeed the Fourier transform) goes back decades Matsuyama et al. (1983). Recently, the literature on neural style transfer Gatys et al. (2016) constrained the image generator with two cost functions: one that controlled the content and one that controlled the style. The style cost function constrained the response to a ConvNet in the lower layers, some of which are known to be Fourier-response-like Zeiler & Fergus (2014).

Prior work includes incorporating the loss from the neural style transfer literature to train a VAE that reconstructs microstructure Sardeshmukh et al. (2021) and using a Generative Adversarial Network (GAN) when training a VAE to produce reconstructions of microstructure that preserve the texture Zhang et al. (2024). Unlike previous work, we directly optimize in spatial statistics space. We aim to use limited embedding space to only store texture information.

Embedding microstructure in a low-dimensional space while preserving microstructure information is useful since it can enable researchers to rapidly explore different microstructures in the process of materials design Kalidindi (2015) McDowell et al. (2009) Sundararaghavan & Zabaras (2009).

3 METHODS

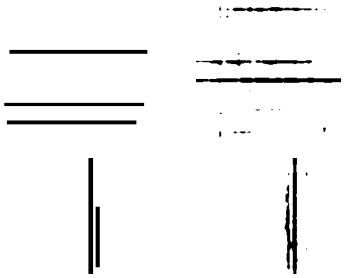
We train a VAE to reconstruct images of texture such that, in spatial statistics space, the original and the reconstruction are close. That is, the VAE loss we use is

$\mathcal{L} = \alpha \cdot ||f(\boldsymbol{x}) - f(\boldsymbol{x}_{recon})||_2 + \beta \cdot \mathrm{Loss}_{KL}$.

That is, instead of minimizing the distance between the input x and the reconstruction $\boldsymbol{x}_{recon}$, we minimize the distance between the spatial statistics f of $\boldsymbol{x}$ and $\boldsymbol{x}_{recon}$. The spatial statistics function can be efficiently computed using FFT and is differentiable.

4 RESULTS

On a dataset of images of 100,000 randomly-placed vertical and horizontal lines of various lengths (details in Appendix A), we show that our network successfully produces reconstructions that are close to the original in spatial statistics space, but not necessarily in data space, and does so better than the baseline VAE. Qualitative and quantitative comparisons are in Appendix B. More samples are in Appendix C.

Data		**Spatial statistics**	
Original	**Reconstructed**	**Original**	**Reconstructed**
			

REFERENCES

Ahmet Cecen, Tony Fast, and Surya R Kalidindi. Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure. *Integrating Materials and Manufacturing Innovation*, 5:1–15, 2016.

Albert Einstein. Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Archives des Sciences*, 37:254–256, 1914.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423, 2016.

Surya R Kalidindi. *Hierarchical materials informatics: novel analytics for materials data*. Elsevier, 2015.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *Proc. ICLR*, 2014.

Takashi Matsuyama, Shu-Ichi Miura, and Makoto Nagao. Structural analysis of natural textures by fourier transformation. *Computer vision, graphics, and image processing*, 24(3):347–362, 1983.

David L McDowell, Jitesh Panchal, Hae-Jin Choi, Carolyn Seepersad, Janet Allen, and Farrokh Mistree. *Integrated design of multiscale, multifunctional materials and products*. Butterworth-Heinemann, 2009.

Noah H Paulson, Matthew W Priddy, David L McDowell, and Surya R Kalidindi. Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics. *Acta Materialia*, 129:428–438, 2017.

Avadhut Sardeshmukh, Sreedhar Reddy, BP Gautham, and Pushpak Bhattacharyya. Texturevae: Learning interpretable representations of material microstructures using variational autoencoders. In *AAAI Spring Symposium: MLPS*, 2021.

Veera Sundararaghavan and Nicholas Zabaras. A statistical learning approach for the design of polycrystalline materials. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 1(5):306–321, 2009.

Manik Varma and Andrew Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III 7*, pp. 255–271. Springer, 2002.

Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pp. 818–833. Springer, 2014.

Yichi Zhang, Paul Seibert, Alexandra Otto, Alexander Raßloff, Marreddy Ambati, and Markus Kästner. Da-vegan: Differentiably augmenting vae-gan for microstructure reconstruction from extremely small data sets. *Computational Materials Science*, 232:112661, 2024.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

A EXPERIMENTAL SETUP

In our experiment, we trained a VAE on a dataset consisting of 100,000 images, each sized 224x224 pixels. The VAE utilizes a pretrained ResNet-152-based encoder with a bottleneck of size 9 and an untrained decoder with a spatial statistics loss. The training hyper-parameters were set to the following: a total of 1500 epochs, a learning rate of 0.001, a batch size of 32, and a KLD beta value of 1. For detailed insights into the experimental configuration, including access to the code, models, and dataset, interested readers are directed to the README.md file of the GitHub repository: `https://github.com/spatial-stats-vae/spatial-stats-vae.git`. This repository provides comprehensive resources for replicating this experiment.

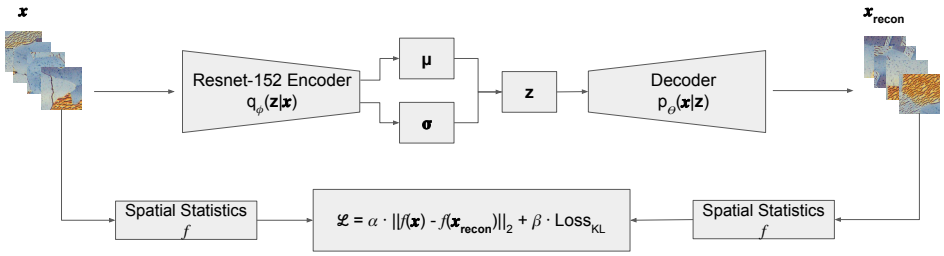


Figure 1: The VAE with spatial statistics loss.

B EVALUATION

We assessed the performance of the VAE with 100 image pairs from the test set. From the 100 reconstructions, 98 had the correct line orientation, and 63 had the correct number of lines as the original image

The reconstructions are characterized by randomly arranged lines, except one line remains constant across all images, varying only in size and orientation in response to the source image. The black pixels in the reconstructions, which represent lines, exhibited an average difference in volume fraction of 17%.

Our model demonstrated a consistent ability to accurately reconstruct images that originally contained a single line. However, its performance varied with images having more lines. Notably, in instances where the original images included two or three lines, our model correctly identified the number of lines in most cases. It is also worth mentioning that when comparing the original images to those with spatial statistics most similar to the reconstructed images, there was a noticeable increase in accuracy regarding the correct identification of the number of lines (see A3).

To determine if our model was overfitting, we analyzed 1000 reconstructed samples from both the training and validation datasets, comparing them against their respective complete sets using Mean Squared Error (MSE) as the metric. Our focus was on the distribution of average MSE values for each reconstruction, where each average denotes the mean MSE of a single reconstruction compared against all samples in the respective dataset. Our findings indicate that the distribution of the average MSEs for training reconstructions relative to the training set closely mirrors the distribution of the average MSEs for validation reconstructions relative to the validation set (refer to Figure 4). This similarity suggests that our model did not overfit to the training data. We also analyzed the MSE of spatial statistics, comparing the spatial statistics of 1000 training reconstruction samples with the spatial statistics of the entire training set, and similarly for the validation set. The findings revealed similar results for both training and validation sets (refer to Figure 5).

Lastly, we have tabulated reconstruction examples in Appendix B. In the post-processing stage of the VAE reconstructions, a threshold value of 0.05 was utilized on the reconstructed tensors. This thresholding transforms grayscale images into binary ones, a step taken to improve the clarity of the output images. Crucially, this thresholding process preserves the integrity of the reconstructions, ensuring that the resulting binary images are appropriate for further visual analysis and comparison. It should be noted that the reported MSE scores below do not depend on this thresholding function, and that the thresholding was solely done for visual purposes. For additional details, refer to the Appendix A2.

B.1 CALCULATION OF VOLUME FRACTION DIFFERENCE

The volume fraction difference percentage is a metric in evaluating the accuracy of the reconstructed images in terms of the black pixel content. This metric quantifies the difference in the number of black pixels (representing lines in images) between the original and the thresholded (or reconstructed) images. The calculation of this percentage is vital for understanding the degree to which the reconstructed images deviate from the original in terms of pixel composition.

The formula for calculating the volume fraction difference percentage is as follows:

$$\left| \frac{\text{original_black_pixel_count} - \text{thresholded_reconstruction_black_pixel_count}}{\text{original_black_pixel_count}} \right| \times 100 \tag{1}$$

After calculating the percentage difference, the value is rounded to four decimal places for precision and ease of interpretation. This rounding process helps in standardizing the results for comparison and further analysis.

Note: The calculated volume fraction percentage is rounded to four decimal places.

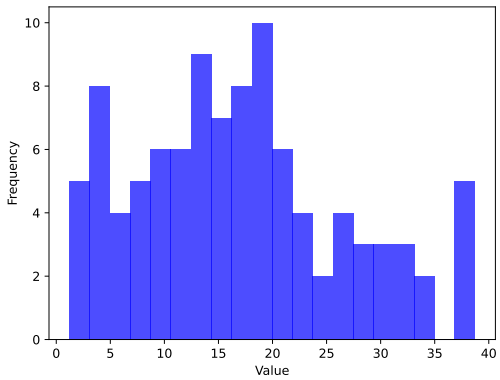


Figure 2: The difference between percentage of black pixels in the input image and the percentage of black pixels in the reconstruction.

B.2 IMAGE THRESHOLDING FUNCTION

The threshold function is implemented to modify the image based on a predefined threshold value, setting pixel values below the threshold to zero and those equal to or above the threshold to one. This process is crucial for enhancing image clarity in the VAE's reconstructed outputs.

The mathematical representation of the threshold function is as follows:

$$\text{threshold_pixel}(pixel, threshold) = \begin{cases} 0 & \text{if } pixel < threshold, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

B.3 NUMBER OF LINES IN ORIGINAL IMAGES VS. RECONSTRUCTIONS

We randomly select 100 images in the test set, compute their reconstructions, and also find the closest image in spatial statistics space to the original. We then compare the number of lines in the test image to the number of lines in the reconstruction (left) and the number of lines in the closest image in spatial statistics (right).

We show that images that are close together in spatial statistics space have similar numbers of lines, and that our reconstructions produce images with numbers of lines related to the number of lines in the original.

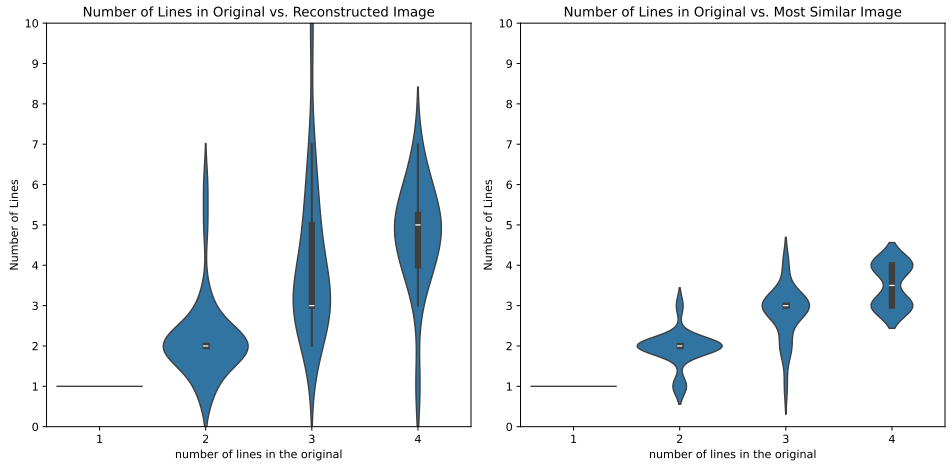


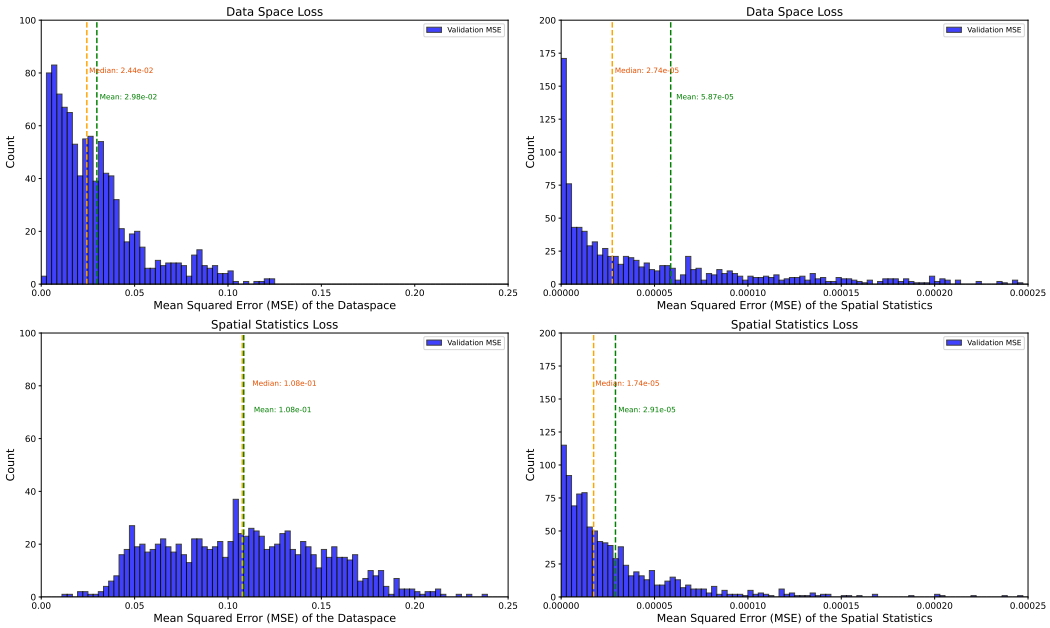
Figure 3: Line count in the reconstructed image vs original image (left), line count in image closes in spatial statistics space vs line in count in original image (right)

B.4 Quantitative results for original VAE and our VAE

We train a "vanilla" VAE that minimizes the distance in data space between the original and the reconstruction ("data space loss"), and our proposed model, which minimizes the distance in spatial statistics space between the original and the reconstruction ("spatial statistics loss"). Our method produces reconstructions that are closer together in spatial statistics space, but further apart in data space.

This is to be expected: we directly try to make the reconstructions be as close as possible as the original in spatial statistics space, and do not try to match the original in data space at all.

Note the exact reconstructions that are close to the original in data space would also be close to the original in spatial statistics space. For that reason, the spatial statistics error when training with data space loss is larger, but not much larger, than when training with data space loss.



C MORE ILLUSTRATIVE EXAMPLES

C.1 RECONSTRUCTIONS AND THEIR SPATIAL STATISTICS

The following table contains illustrative instances of reconstructions from the 100 samples. The table shows the original image from the dataset, and the reconstructed image by our model in the two columns on the left. The two columns on the right show the auto-correlations (aka spatial statistics) of the original and the reconstructed images, respectively. We have also included the 2 reconstruction samples (out of 100) that did have the correct line orientation at the end of the table.

Table 1: Original, reconstructed, and spatial statistics difference

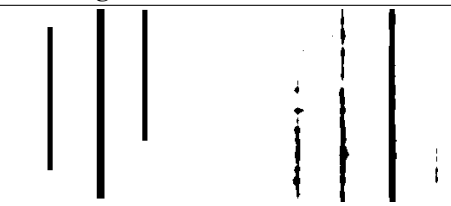
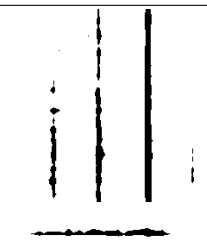
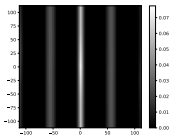
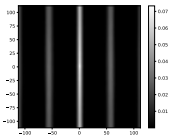
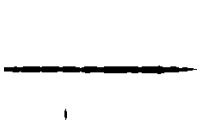
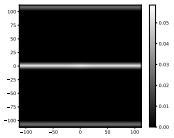
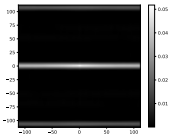
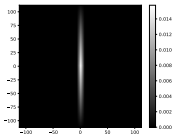
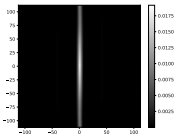
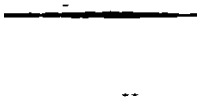
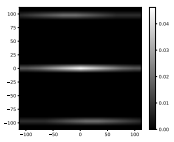
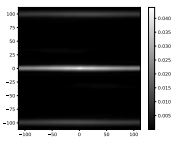
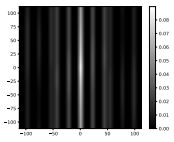
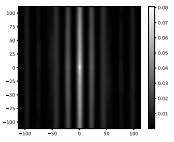
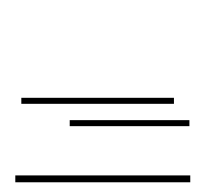
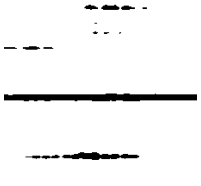
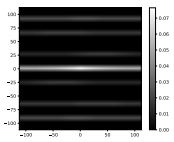
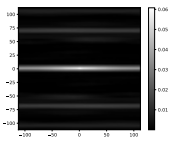
Data		**Auto-correlations**	
Original	**Reconstructed**	**Original**	**Reconstructed**
			
			
			
			
			
			

Table 2: Original, reconstructed, and spatial statistics difference

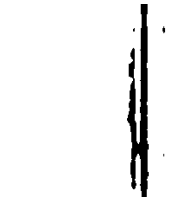
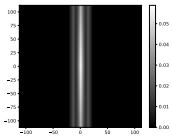
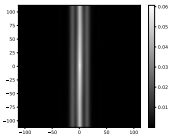
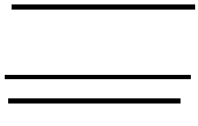
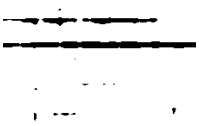
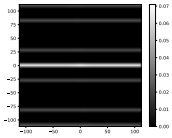
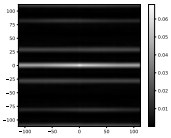
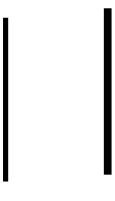
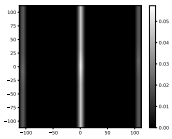
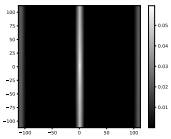
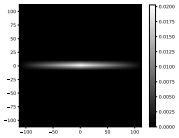
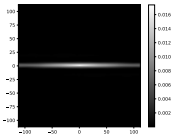
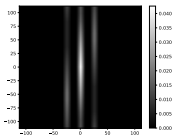
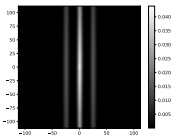
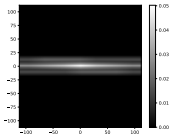
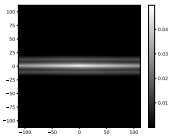
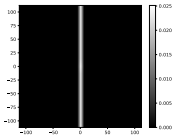
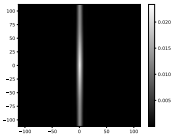
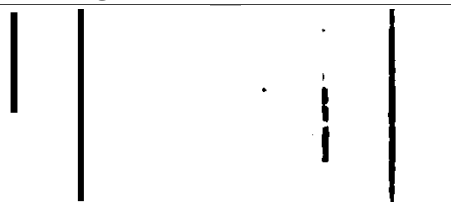
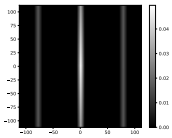
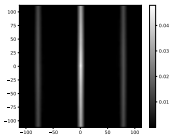
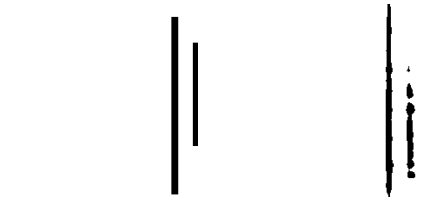
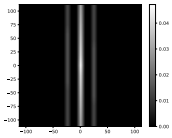
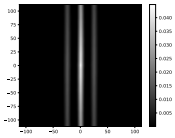
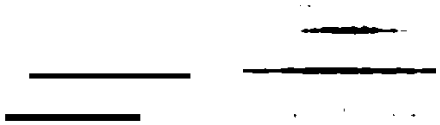
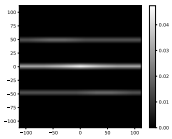
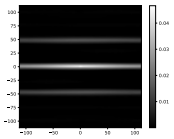
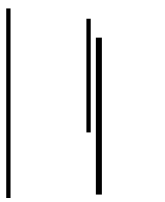
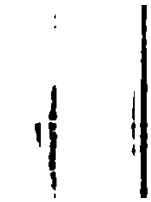
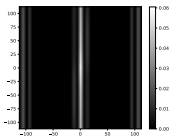
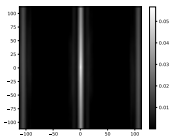
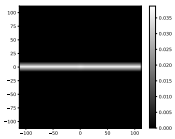
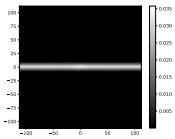
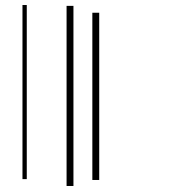
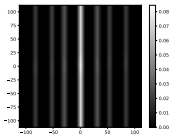
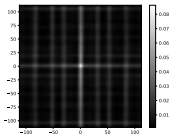
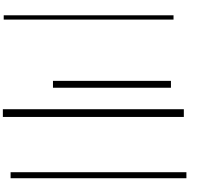
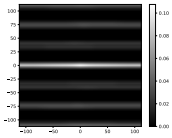
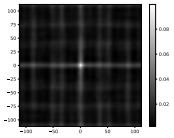
Data		Auto-correlations	
Original	**Reconstructed**	**Original**	**Reconstructed**
			
			
			
			
			
			
			

Table 3: Original, reconstructed, and spatial statistics difference

Data		Auto-correlations	
Original	**Reconstructed**	**Original**	**Reconstructed**
			
			
			
			
			
			
			

C.2 RECONSTRUCTIONS ALONG WITH MOST SIMILAR IMAGES FROM THE TRAINING SET

The following table contains illustrative instances of reconstructions, and the most similar images to them. The table shows the original image from the dataset, and the reconstructed image by our model in the two columns on the left. The third columns shows the most similar image to the reconstructed image from the training set based on MSE of the reconstruction and the sample from the training set. The fourth column shows the most similar image to the reconstructed image from the training set based on MSE of the spatial statistics of the reconstruction and the sample from the training set.

Table 4: Original, reconstructed, and spatial statistics difference

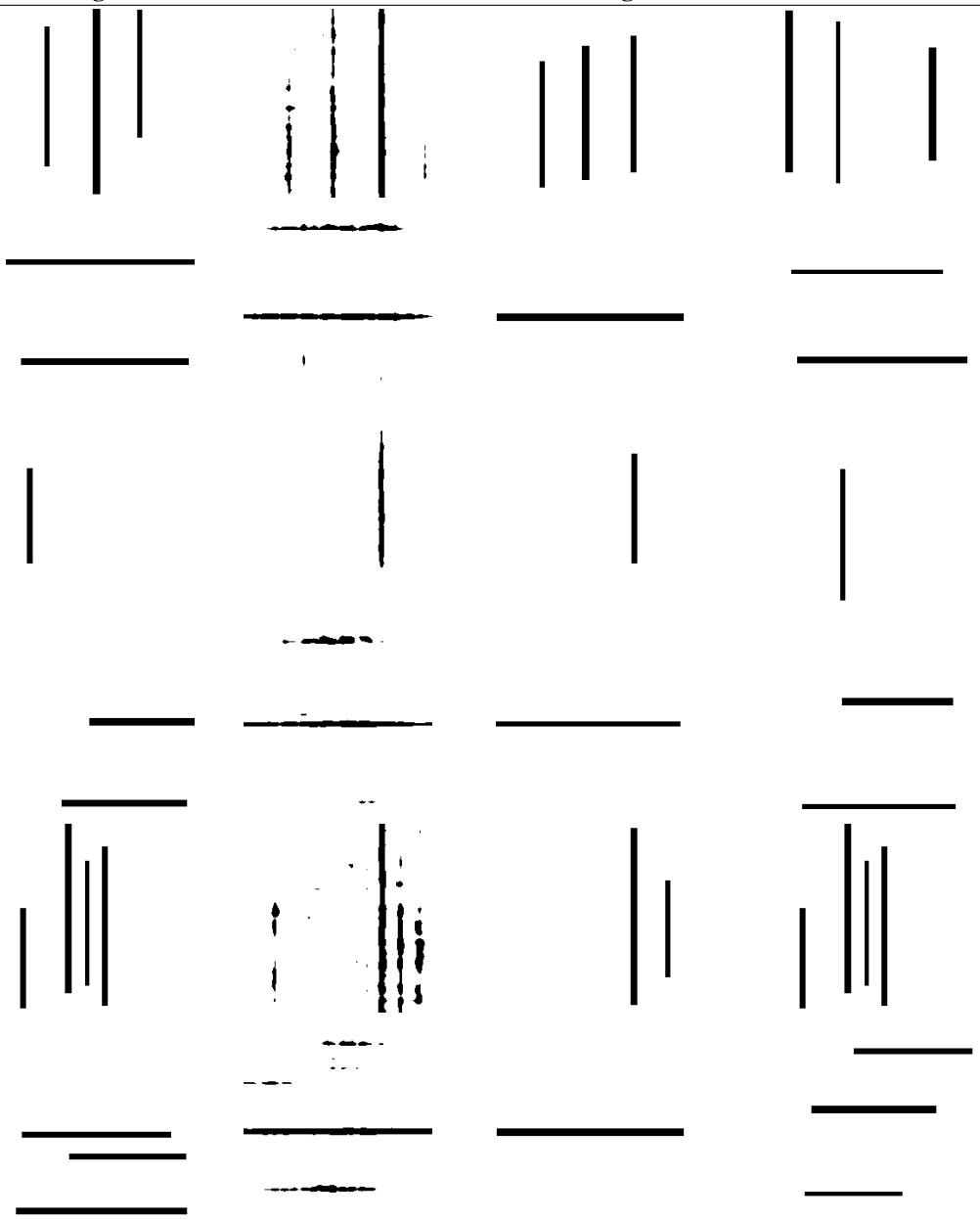
Data		**Most similar images**	
Original	**Reconstructed**	**Based on image MSE**	**Based on auto-corr. MSE**
			

Table 5: Original, reconstructed, and spatial statistics difference

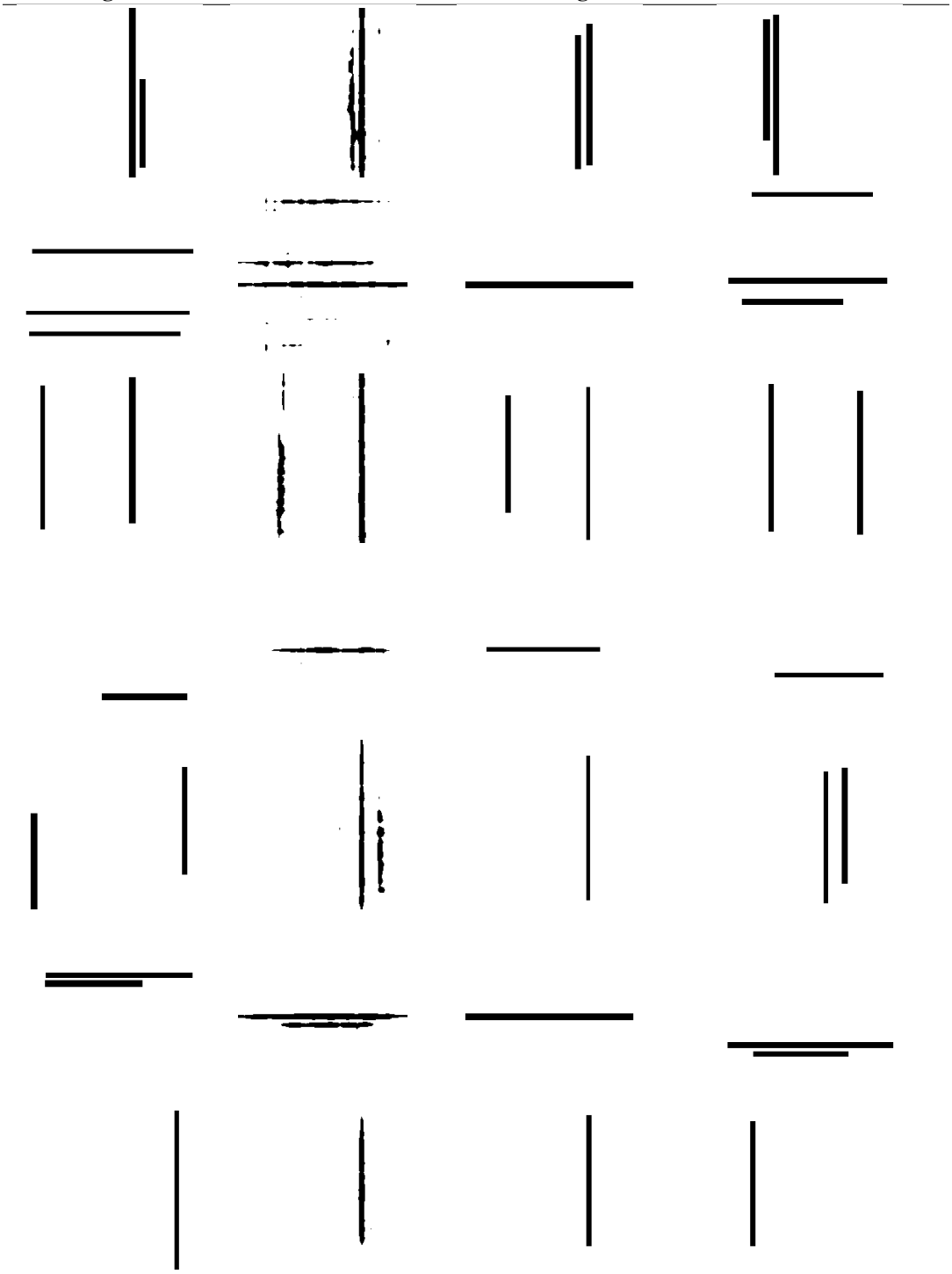
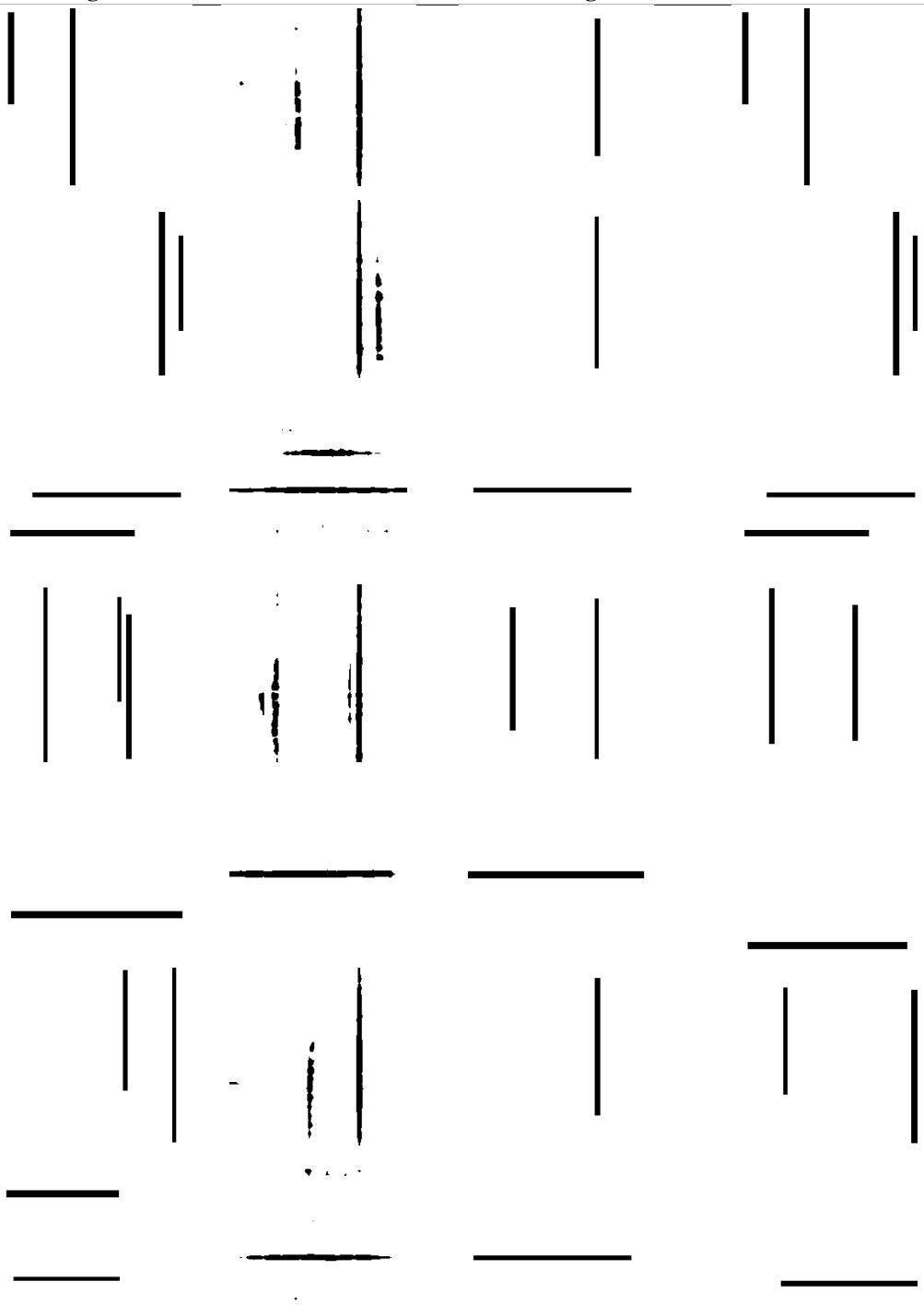
Data		Most similar images	
Original	**Reconstructed**	**Based on image MSE**	**Based on auto-corr. MSE**
			

Table 6: Original, reconstructed, and spatial statistics difference

Data		Most similar images	
Original	**Reconstructed**	**Based on image MSE**	**Based on auto-corr. MSE**