Improvement of Basic CNN Via Biological Mechanism Imitation For Classification of Fashion MNIST Dataset

Eknoor Singh and Gurman Brar
Department of Systems Design Engineering
University of Waterloo
Waterloo, ON N2L 3G1
eknoor.singh@uwaterloo.ca
g9brar@uwaterloo.ca

April 22, 2024

Abstract

This paper investigated the functional capabilities of CNNs in image classification, while varying network architectures that mimic certain biological mechanisms (sparse connectivity and lateral inhibition). It was hypothesized that a simple CNN model would outperform these biologically altered models in measures of training loss and test accuracy. Varying train and test set sizes were used to assess each model's ability to generalize the image classification task with smaller amounts of data. It was found that the model that implemented lateral inhibition showed the greatest test accuracies. Analysis was provided on how these mechanisms can affect the learning process for the CNN models and future improvements to the investigation were discussed.

1 Introduction

1.1 Motivation

Convolutional neural networks (CNN) are a type of deep-learning model where hidden layers are placed in between the input and output layers, with a specified number of nodes in each layer [1]. CNNs are largely used in image classification projects, due to their ability to mimic certain aspects of the human visual cortex [2]. This will be reviewed further in Section 1.3. Two of the characteristics of the visual cortex that can be recreated within CNNs are sparse connectivity and lateral inhibition. Sparse connectivity is the mechanism where neuronal connections are limited [3]. This function has a direct example, through the output from the thalamus before serving as the input to the visual cortex [4].

Sparse connectivity can be implemented into CNNs as forms of SparseLayers, where neurons don't connect to all neurons contained in the surrounding layers. Lateral inhibition is a biological mechanism, where neighboring neurons prevent the excitation of a neuron if excited first [5]. This mechanism is largely used by photoreceptors within the retina, with the purpose of light contrast [5]. This can be recreated within a CNN's architecture by applying inhibition to neurons on the layer level if neighboring neurons have been stimulated. With this in mind, a suitable dataset for this experimentation was chosen to be the MNIST Fashion dataset. This will be discussed further in Section 2.2.

1.2 Objectives & Hypothesis

With the topic of the article stated, the objective of this experiment is that of comparative standards. The following are the objectives this investigation seeks to research:

1. To create three CNN models, one being with standard network architecture, the other two having sparse connectivity and lateral inhibition implemented

2. To evaluate the loss curves of each model and the accuracy of the models, using 25% and 100% of a train and test set that belongs to the MNIST Fashion dataset

3. To discuss the difference in results from the models and deliberate about the effects of the imported biological mechanisms

With the objectives of the experiment outlined, a general hypothesis was formed regarding the comparison of a simple CNN's performance against the "biological-tuned" CNN models. The hypothesis that will be tested is,

The simple CNN model will outperform the altered CNN models that implement biological-like mechanisms, in measures of test accuracy and loss curves.

The rationale behind this hypothesis is that by implementing these mechanisms into the CNN models, models will likely be inhibited by their respective effects. The reduction of neuronal connections through sparse connectivity and the inhibition of neurons through lateral inhibition are the factors that can hinder the models' performance.

1.3 Background Literature

The field of image classification has seen great leaps of progress in the 21st century. Recent literature has explored the Fashion MNIST dataset in image classification. From the review of these articles, it can be said with great emphasis that CNNs are one of the most common and well-performing techniques employed when doing image classification. In [6], multiple CNNs, of varying architecture, were tested in comparison to classify the dataset. In contrast to the other CNN models, the most simple architecture CNN performed best in the category of model loss, while only slightly being outperformed by the developed CNN-dropout-3 model, which peaked at 99.1% accuracy. A separate study varied the architecture of two CNN models to test on the same dataset, the MNIST Fashion Dataset [7]. The study investigated the use of a linear SVM being the output layer of the model, instead of the softmax function. The respective accuracies from the two separate models were found to be 90.72% and 91.86%. In [8], sparse connectivity was tested on multiple artificial neural networks, proposing that the implementation of sparsity can allow ANNs to outperform current-day standards. This article was dated in 2018, with many years available for future improvements. Lastly, a separate paper explored the capabilities of a Spiking Neural Network with lateral interactions against other spiking models with the Fashion MNIST, MNIST, and N-MNIST datasets [9]. The model with integrated lateral interactions performed the best on Fashion MNIST, with an accuracy of 92.07%, showing true potential in the use of lateral interactivity of neurons within layers.

2 Methodology

2.1 Sparsity, Lateral Inhibition, and CNN architecture

As described in Section 1.2, the first objective is to create three distinct CNNs, one being a simple CNN and two others implementing the biological mechanisms previously described. The first layer is the Conv2D layer, with the 'relu' activation function. It's followed by a MaxPooling Layer, which is then flattened and fed into two dense layers. The CNN with sparse connectivity is of similar architecture as the simple CNN but has the first dense layer replaced with a sparse layer. This causes the outputs of the sparse layer to randomly feed inputs to neurons of the dense layer. Lastly, the CNN model with lateral inhibition has an implementation where outputs from the Conv2D layer are inhibited, but similar architecture to the simple CNN. All implementations can be viewed in the Python notebook associated with this paper.

2.2 Fashion MNIST Dataset [10]

The Fashion MNIST dataset comprises 70,000 images, sized 28x28 pixels, of clothing items of 10 different categories. The 10 categories and examples of images within these classifications are shown in Figure 1. The train and test sets of data are split into 60,000 and 10,000 images respectively. To better evaluate the models, the accuracy of each model will be tested with 25% and 100% of the test dataset, as mentioned as the second objective in Section 1.2. The loss curves will also be tested at 25% and 100% of the training dataset. This will be done to see how well each model generalizes with relatively smaller datasets, while also checking their capabilities with the entire train/test sets.

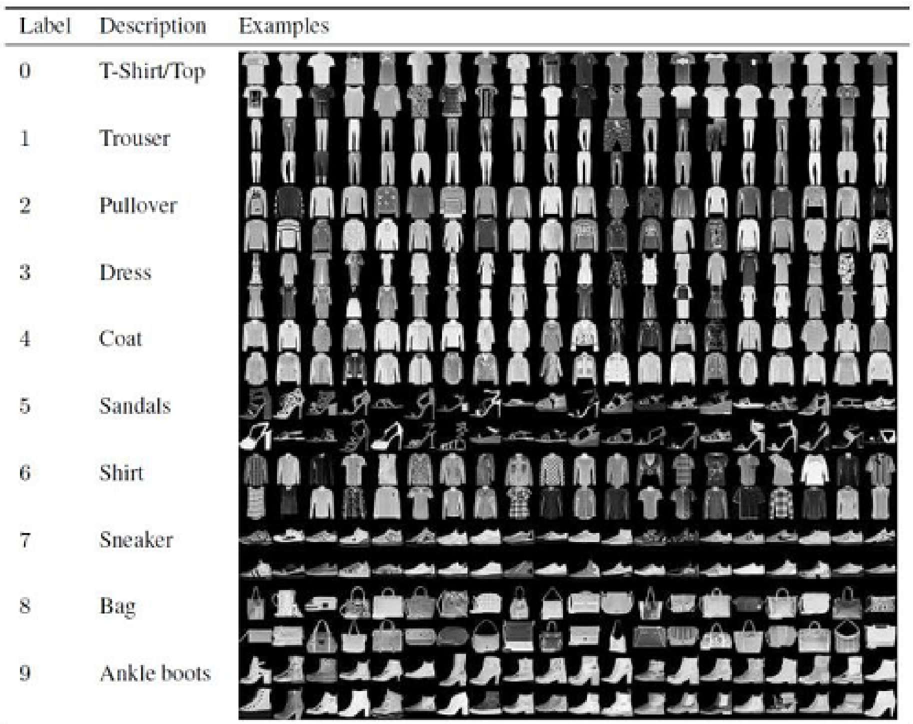

Figure 1: Categories of Fashion MNIST Dataset [11]

2.3 Analysis Methods

As mentioned in Section 1.2, the hypothesis stated that the models will be compared in measures of accuracy against the test set (both 25% and 100% splits) and in measures of model loss. Each of the models described will be trained over 50 epochs, and the training and validation losses will be compared for the designed models. Additionally, the final accuracies with the two test dataset compositions will also be compared for all 3 models. From these results, discussions can be conducted about the performance of the 3 models, and how the network architecture can contribute to the varying metrics of evaluation.

3 Results

Before exploring the results of implementing the two human neurophysiology mimicking mechanisms, as described in the methodology, we must formulate our baseline using the basic CNN architecture. After training the basic CNN model on the fashion MNIST training dataset, the resulting test set accuracy was 100% (99.99% to be exact, Refer to Table 1).

Table 1: Model performance on test data with optimized hyper-parameters

Basic CNN (Full Test Set)	99.99%
Sparse Connectivity CNN (Full Test Set)	98.73%
Basic CNN (25% Test Set)	92.65%
Sparse Connectivity CNN (25% Test Set)	91.33%
Lateral Inhibition CNN (Full Test Set)	100%
Lateral Inhibition CNN (25% Test Set)	92.71%

In addition, following a review of the loss curves for both the training and validation sets, it was evident that there was minimal over-fitting with the basic CNN model (refer to Figure 2). The observed results were expected as the fashion MNIST dataset provides sufficient data for effective model training and allows the model to generalize well to both the validation and test sets. Furthermore, as previously mentioned, CNNs are the "baseline" and "standard" for solving basic image classification tasks such as that being investigated with the fashion MNIST dataset.

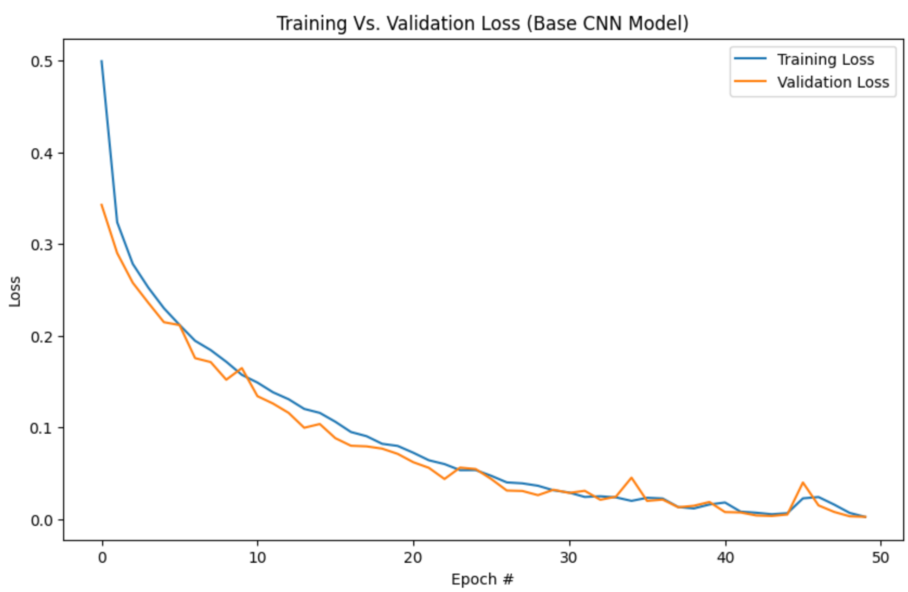

Figure 2: Loss Curves for Basic CNN (Full Training Set)

In theory, the accuracy of the following two models cannot exceed that of the CNN as it already almost achieves a 100% accuracy on the test set. However, other mechanisms of change such as changes to the loss curves, as well as changes in over-fitting patterns, in addition to potential decreases in accuracy can still be investigated.

3.1 Sparse Connectivity

The first experiment for the new CNN model with an additional sparse-connectivity layer was to train it on the same data as the basic CNN model and produce a training vs. validation loss plot. In addition, we want to test the accuracy of this model compared to the basic CNN model (testing on the same set of testing data).

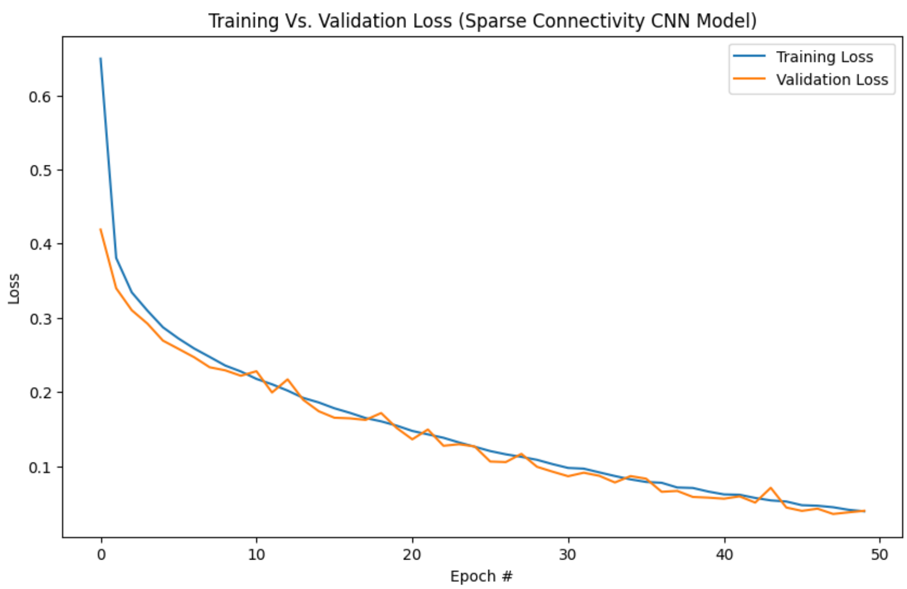

Figure 3: Loss Curves for Sparse Connectivity CNN (Full Training Set)

Referring to Table 1, it can be noticed that this model (with the sparse-connectivity layer) had a sub-optimal accuracy score when compared to that of the basic CNN. However, following an analysis of the loss curves plot, it could be noticed that the model did not end up converging to a loss of almost 0 (like was observed with the basic CNN). In addition, the loss curves themselves seemed more stable and had less frequent spiking and even when there were spikes, they had a smaller amplitude (please refer to Figure 3.

Our final experiment was to reduce the size of the training set to 25% of its original size and compare how this affects the generalizability between the basic CNN and the CNN with a sparse connectivity layer. Referring to Figure 4, it can be seen that the adapted CNN's training loss is still higher than that of the basic CNN as the training EPOCH approaches the limit (70). However, it seems that the validation loss of the adapted model stabilizes before that of the basic CNN. Although both models seem to be showing over-fitting, it seems that the overfitting is greater in the basic CNN compared to the adapted CNN.

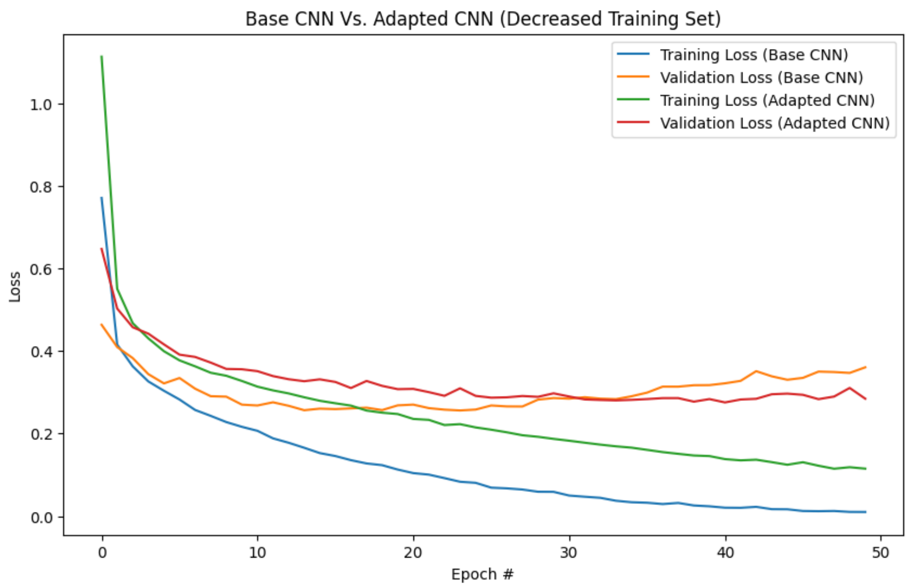

Figure 4: Loss Curves for Sparse Connectivity CNN (Reduced Training Set)

Comparing the accuracy of the two models on the complete testing set (when trained with 25% of the training set), both models performed practically the same with no large improvements or deficits in either model (please refer to Table 1)

3.2 Lateral Inhibition

Similar to that of the sparse connectivity model, for the lateral inhibition model, the model will be trained on the same full set of training data to produce the loss curves for training and validation data, and finally, get a testing set accuracy.

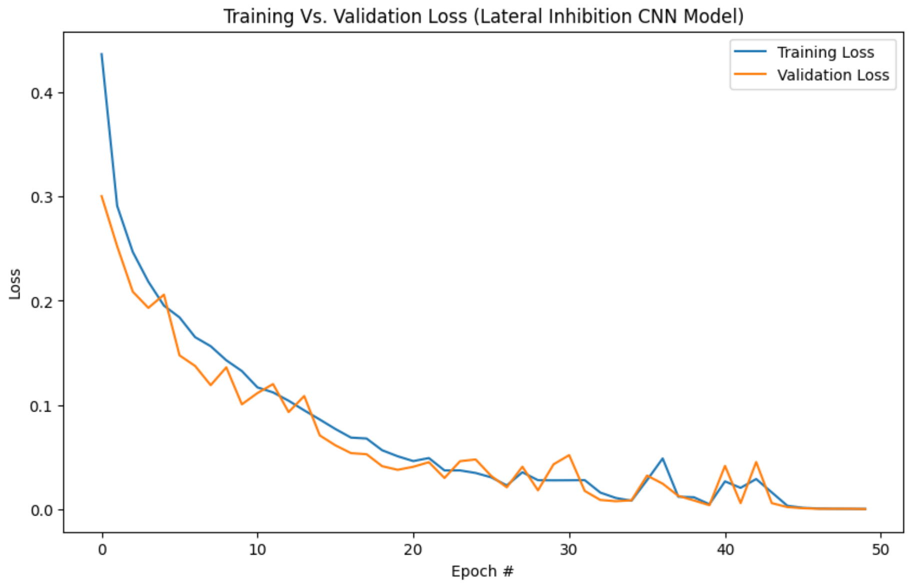

Figure 5: Loss Curves for Lateral Inhibition CNN (Full Training Set)

Comparing the accuracies of the basic CNN model to that of this adapted CNN, it can be noted that the adapted CNN actually performed better but by a negligible amount (0.01% better). After an analysis of the loss curves (please refer to Figure 5) it could also be noted that the loss curves themselves were more choppy and less smooth. This means that there were more frequent and larger amplitude variations/spikes in the loss curves.

The final experiment for this model was to re-train it on only 25% of the training set it was originally trained on and compare that to the basic CNN model. Referring to Figure **??**, it can be noted that the adapted model's training loss was able to converge quicker than that of the basic model. However, the loss curve for the validation set reveals that the adapted CNN is more over-fit to the training data than the basic CNN model. Furthermore, when comparing the resulting test accuracies (on the full test set), it could be noted that the adapted CNN performed better than the basic CNN model (refer to Table 1) and for that matter, all the other models that were also trained on 25% of the training set (sparse connectivity model).

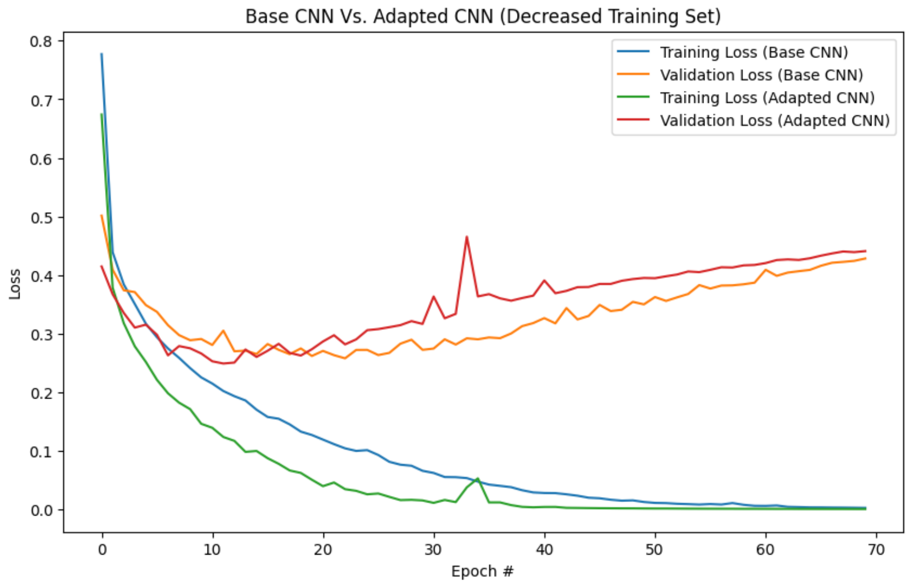

Figure 6: Loss Curves for Lateral Inhibition CNN (Reduced Training Set)

4 Discussion

4.1 Interpretation of Results

Observing the preliminary results in Table 1, models behaved similarly when challenged with different-sized test datasets. Models when tested with the complete test sets, showed great results, with the simple CNN producing a 99.99% accuracy. This accuracy shows little need for improving mechanisms like Sparse Connectivity and Lateral Inhibition, however, these mechanisms also contribute to the model's reduction in overfitting and its ability to generalize at a quicker rate. The basic CNN performed at a slightly lower accuracy when it came to the use of 25% of the test set, in comparison to the Lateral Inhibition CNN, 92.65% vs 92.71% respectively.

When judging the performance of the sparse connectivity CNN, the model underperformed in the accuracy of both test set compositions, pointing towards the possibility of underfitting. This can be attributed to the regularization effects that are produced through using sparse connections. However, while evaluating the loss curves for both trials of the sparse connected CNN, the loss curves contain less volatility and spiking moments, pointing towards a more smoothened learning process. This would aid in scenarios where biological data is highly variable and can allow for efficient training of the network towards its task, although opening it up to possible limitations, like lower accuracy.

When viewing the performance of the laterally inhibited CNN, it performs the best through measures of accuracies, with both percentages being highest in comparison to the other models. When tying in the loss curves of the lateral inhibition model, the curve appears to be volatile, but looking at Figure 6, the model seems to generalize better with the lower training set, as compared to the simple CNN model. This can be attributed to the ability of the lateral inhibition model to create a more whole feature representation, where it's able to adapt quickly and extract more meaningful features. As mentioned previously, lateral inhibition is seen in the visual cortex, with a prime example being the photoreceptors of the retina. The model can be drawn likeness to this example as these receptors adapt their responses to stimuli, becoming more well-adjusted when excited.

4.2 Future Work

Although this paper conducted direct comparisons of models of simple architecture vs varying additive mechanisms, the true difference in performance isn't to be expected since CNNs are renowned for their performance in the prescribed task. However, this paper allowed for astute insights to be drawn on the effect of these biological mechanisms and how they affect the learning dynamics of a CNN model. In large part, the literature presented in Section 1.3 implemented a wide range of CNN models, but all showed relatively similar results. The main extrapolations from those papers were that variation will allow for future growth of CNNs and change how traditional models are trained. As for future works, it would be interesting to view the resiliency of models that are implemented with biological mechanisms, like sparse connectivity and lateral inhibition, but others as well, on new/unseen data. This could lead to an investigation of how these biologically tuned models would fare in standards of adaptability and transfer learning capabilities. Since accuracies of the models were reportedly near/at 100%, the task involved in image classification can/should be made more difficult, allowing for a larger lens to evaluate the models/mechanisms in question. Lastly, more metrics, aside from test accuracy and loss curves, can be studied to get a better grasp on the learning process of these biologically altered models, and they can be better tested in metrics of versatility and robustness as well.

5 Conclusion

In all, we found the model that implemented lateral inhibition, to perform the best in measures of test accuracy, by a slight margin. This goes against the hypothesis that states, the simple CNN model will outperform the other models in measures of test accuracy and loss curves. Although this difference is small, the mechanisms of sparse connectivity and lateral inhibition allow for the training dynamics of the models to be altered, providing their own benefits, alongside shortcomings. Future work has been discussed on changing the metrics used in the evaluation, tasks assigned for the models, as well as other testing methods for versatility and robustness.

6 References

[1] IBM, "What are convolutional neural networks?," IBM, https://www.ibm.com/topics/convolutional-neural-networks (accessed Apr. 20, 2024).

[2] "Image classification using CNN: Introduction and tutorial," Datagen, https://datagen.tech/guides/image-classification/image-classification-using-cnn/ (accessed Apr. 20, 2024).

[3] M. Thom and G. Palm, "Sparse activity and sparse connectivity in supervised ...," Journal of Machine Learning Research, https://www.jmlr.org/papers/volume14/thom13a/thom13a.pdf (accessed Apr. 21, 2024).

[4] David Orenstein, The Picower Institute for Learning and Memory, "Sparse, small, but diverse neural connections help make perception reliable, efficient," MIT News | Massachusetts Institute of Technology, https://news.mit.edu/2023/sparse-small-diverse-neural-connections-help-make-perception-reliable-efficient-0202#:~:text=perception%20reliable%2C
,Sparse%2C%20small%2C%20but%20diverse%20neural%20connections %20help%20make%20perception%20relia
%20process%20sensory%20information. (accessed Apr. 21, 2024).

[5] R. A. Cohen, "Lateral inhibition," Encyclopedia of Clinical Neuropsychology, pp. 1436–1437, 2011. doi:10.1007/978-0-387-79948-3_1379

[6] A. S. Henrique et al., "Classifying garments from fashion-mnist dataset through cnns," Advances in Science, Technology and Engineering Systems Journal, vol. 6, no. 1, pp. 989–994, Feb. 2021. doi:10.25046/aj0601109

[7] A. F. Agarap, "An architecture combining convolutional neural network (CNN) and Support Vector Machine (SVM) for Image Classification," arXiv.org, https://arxiv.org/abs/1712.03541 (accessed Apr. 21, 2024).

[8] D. C. Mocanu et al., "Scalable training of artificial neural networks with adaptive sparse connectivity inspired by Network Science," Nature News, https://www.nature.com/articles/s41467-018-04316-3 (accessed Apr. 21, 2024).

[9] X. Cheng, Y. Hao, J. Xu, and B. Xu, "Lisnn: Improving spiking neural networks with lateral ...," LISNN: Improving Spiking Neural Networks with Lateral Interactions for Robust Object Recognition, https://www.ijcai.org/proceedings/2020/0211.pdf (accessed Apr. 22, 2024).

[10] K. Team, "Keras Documentation: Fashion Mnist Dataset, an alternative to mnist," Keras, https://keras.io/api/datasets/fashion_mnist/ (accessed Apr. 21, 2024).

[11] Greeshma K V, "Fashion-MNIST dataset images with labels and description II. literature... | download scientific diagram," Fashion-MNIST Dataset Im-

ages with Labels and Description, https://www.researchgate.net/figure/Fashion-MNIST-Dataset-Images-with-Labels-and-Description-II-LITERATURE-REVIEW-In-image_fig1_340299295 (accessed Apr. 22, 2024).