

Determining Scenarios for Low-Fidelity Time-to-Execution (TTE) Scheduling Prediction

NEUROBIO 240; Professor Kreiman; May 13th, 2025

Anika Lakhani

Harvard College; Harvard SEAS
Cambridge, MA

anikalakhani@college.harvard.edu

I. ABSTRACT

Human vision excels at processing visual inputs, especially noisy ones, through mechanisms like lateral inhibition and attentional modulation. In theory, successfully harnessing these biological mechanisms to apply towards image classifier CNNs could better equip CNNs to handle noise, improving prediction accuracy rates. My project investigates whether integrating two neurobiologically inspired mechanisms, lateral inhibition (mimicking retinal contrast enhancement) and feature-wise attention (mimicking cortical feedback), will improve the robustness of MobileNetV3-Small on clean, noisy, and occluded image datasets. While experimental models matched control network accuracy on pristine inputs, they demonstrated very modest increases in prediction accuracy, though not statistically significant, under low-level noise ($\sigma = 0.01$). These benefits did not apply consistently when noise levels were raised, contradicting expectation that the biologically inspired mechanisms would aid in sifting through visual noise during image recognition. However, these replication results across the control and experimental models highlight interesting possibilities: 1) Evolved mechanisms might prove brittle to non-ecological corruption patterns (given that CIFAR-10 images differ from the average human visual input), and 2) Modern architectures implicitly encode biological principles, already maximizing the benefit of biological mimicry. These findings generate exciting future research avenues and proved very interesting as an exploratory learning experience!

II. INTRODUCTION

A. Motivation

Our brains deal with constant noise, yet in tasks like auditory focus or image recognition, we perform incredibly well partially because of certain biological mechanisms designed to identify important input and tune out the rest. For instance, we can correctly categorize a wide range of visual input, even if these images have visual noise or occlusion.

Modern convolutional neural networks (CNNs) dominate computer vision but remain still brittle to input corruptions like noise and occlusion, the very hurdles that humans excel at. Our biological visual systems utilize mechanisms like lateral inhibition (sharpening spatial contrasts) and attentional modulation (dynamic feature prioritization). My project tests whether embedding simplified adaptations of these mechanisms into MobileNetV3, a common image classifier, can enhance robustness, using CIFAR-10 under three test conditions:

- **Clean:** The standard test set
- **Noisy:** Additive Gaussian noise ($\sigma = 0.01, 0.1, 0.3$)
- **Occluded:** Random 50×50 pixel masks

B. Central Research Question

The aim of my research is to replicate biological mechanisms to aid in image recognition. Therefore, my central question is: **Can the biologically inspired model outperform, if not replicate, a traditional model, particularly on noisy data?**

III. RELATED WORKS

My key papers for this project are [1] and [2]. The former helps to describe how attention mechanisms in the brain work, helping me to replicate them as accurately as is feasible. The latter is an existing implementation of lateral inhibition, a neural attention mechanism in the body, within an image classifier CNN context.

A. Biological Mechanisms in the Brain

The Biased Competition Model of Attention from neuroscience posits that neurons compete for representation until there is a winner that ends up influencing the end behavior. This phenomenon has been researched for a few decades, but [1] “extend[ed] the biased-competition framework to the realm of large-scale multivoxel brain activations” with strong results and a new study setup for measuring neural attention mechanisms.

B. Using CNNs over Other AI Approaches

I mainly chose CNNs because of relevance to the class material and potential for TF assistance. But, I also specifically chose CNNs based on how powerful they are for image classification. For instance, [11] demonstrated the tunability of CNNs, particularly with structure. They saw an average 2% performance degradation with each layer they removed from their model. A more recent paper, [12] demonstrated how CNNs continue to remain powerful, noticing a 99.91% accuracy on the MNIST image dataset using a (meaningfully optimized) CNN.

C. Biological Mechanisms in Neural Networks

[9] is the foundational paper for the Transformer architecture, which relies heavily on attention as opposed to just using an encoder-decoder configuration. Even though I am not using Transformers directly, I still draw inspiration from this research and concept, adding to it by exploring other attention mechanism replication approaches.

Research on neurobiologically inspired CNNs focuses on two key areas:

1) Cortical Attention Mechanisms: Cortical feedback loops in the brain have been represented as feature-wise attention in neural networks, dynamically prioritizing informative channels. [5] seminally introduced Squeeze-and-Excitation Networks (SENet), greatly improving upon prediction accuracy benchmarks by leveraging channel-wise attention. Although accompanied by increased computational costs, [6] extended this progress with spatial attention in image recognition.

2) Lateral Inhibition in Vision Models: Retinal ganglion cells transmit output from the eye to the brain. Inspired by these interactions, lateral inhibition was developed for image recognition as a noise suppression technique to increase spatial contrast. Excited neurons dampen the activation of neighboring neurons, creating a sharper contrast between contributing neurons and irrelevant neurons. [3] used explicit lateral inhibition kernels in CNNs, meaningfully improving edge detection on MNIST and CIFAR-10. Subsequent work by [4] showed that these mechanisms enhance robustness to Gaussian noise (yet struggle with occlusion). [2] incorporates a modified low-pass filter into their CNN in order to replicate lateral inhibition in computer-based image classification, improving performance on the AlexNet image dataset by almost 4%.

D. Efficient Architectures and Robustness

- **MobileNet Series:** [7] introduced hardware-aware neural architecture search (NAS), robustly balancing accuracy and latency in MobileNetV3
- **Corruption Robustness:** [8] benchmarked CNN robustness on corrupted ImageNet, finding that larger models generalize better to noise and worse to occlusion



Fig. 1. The following figure shows a sample image within the CIFAR-10 dataset in its original form, its 'noisy' processed form, and its 'occluded' processed form. Notice how the Gaussian noise disrupts the clean lines of the image, and the occlusion partially covers a random part of the image.

IV. METHODS

A. Dataset and Preprocessing

- **CIFAR-10:** 32×32 images resized to 224×224 for MobileNet compatibility [19].
- **Oxford 102 Flowers:** Also tested with the Oxford 102 Flowers dataset as pictured in Figure 2, which was so small that all prediction accuracies were below 60%, as pictured in Figure 3. For this reason, this avenue was not pursued further and did not make it into the final results [16].
- **Ecoset:** A worry I had with using CIFAR-10 was that the images would not replicate visual stimuli in nature since they are specifically designed to train computers, not biologically inspired approaches. So, I tried to use Ecoset, a very large image dataset to solve for this problem, but every version either overloaded my RAM or memory, even with interventions like training/testing in small batches. For this reason, this avenue was not pursued further and did not make it into the final results [18].
- **Corruptions:**
 - Gaussian noise (all values tested were $\sigma = 0, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7$, though final values tested were $\sigma = 0.01, 0.1, 0.3$)
 - Stochastic occlusion (50×50 masks with $p = 0.2$)

B. Model Architectures

- **Control:** Unmodified MobileNetV3-Small, pretrained on ImageNet (originally used ImageNet, but because of many necessary code iterations, created a more compact project version) [17]
- **Experimental Variants (Ablation Study):**
 - Attention-only: Channel-wise attention blocks (based on cortical feedback loops in humans)

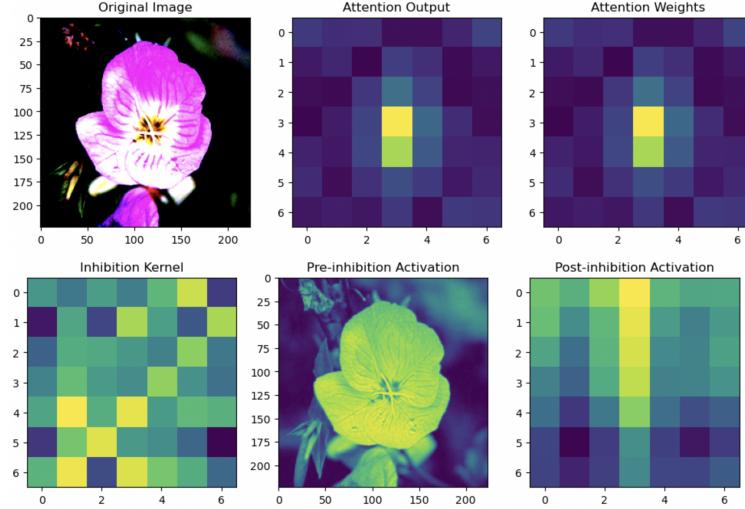


Fig. 2. Top row (left to right): Original input image; spatial output of the attention module highlighting the flower center; learned attention weights, which focus on the most salient region. Bottom row (left to right): Learned lateral inhibition kernel; pre-inhibition activation map showing initial feature responses; and post-inhibition activation, where spatial contrast is sharpened and background activity is suppressed. While not ultimately successful, the self-attention and lateral inhibition mechanisms worked together to enhance relevant features and reduce noise for the biologically inspired CNN. Aggregate data gathered from 5 epochs of training on the bio-combined CNN model. Granular understanding of these heatmaps is not required; data is just included to be illustrative of general trends in an intermediate step of the experimental process.

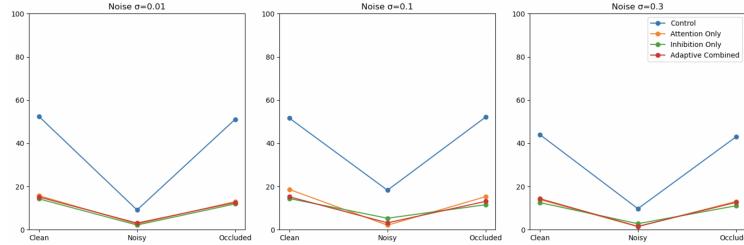


Fig. 3. Not only did the traditional CNN outperform different versions of the biologically inspired CNN in all cases, but no model performed well with the Oxford 102 Flowers dataset because it was so small that the models could not sufficiently train before being tested or validated.

- Inhibition-only: Lateral inhibition layers after each MBConv block (mimicking interactions between retinal ganglion cells)
- Adaptive Combined: Learnable mixture of both mechanisms

C. Training Protocol

- 5 epochs (FAST_MODE=True for rapid iteration)
- AdamW optimizer ($lr = 1e-4$, weight decay= $1e-5$) given what I learned from previous assignments

- Mixed-precision training (AMP enabled) for faster computation without compromising performance
- Inhibition kernels were tested at a size of 7x7 and 5x5 pixels with the hypothesis that the 7x7 version would improve inhibition performance, but the two performed quite similarly. So, a kernel dimension of 5x5 was chosen to balance mechanistic overhead with performance.

D. Final Improvement Strategies

In intermediate experiments, I could not prevent from getting roughly equivalent performance across my experimental models and the control model across clean, noisy, and occluded images. So, I asked Perplexity.AI if there were any last-ditch strategies I could employ, and it helped me write code for an adaptive mechanism integration as well as a stochastic neural masking implementation (Perplexity based its responses off of [15]):

1) Adaptive Mechanism Integration: Instead of statically or otherwise arbitrarily combining strategies in the combined CNN model, I mimicked biological neural systems through *adaptive mechanism integration*.

A learnable parameter γ automatically balances my original self-attention and lateral inhibition mechanisms during training. This adaptive blending allows the network to:

- Prioritize attention for complex pattern recognition
- Emphasize inhibition for noise reduction
- Find optimal combinations for different tasks

2) Stochastic Neural Masking: A further tactic to combat noise in the ML architecture by mimicking biological synaptic pruning, I implemented *stochastic neural masking*:

- Randomly silences 20% of neuron outputs during training
- Maintains signal strength through compensatory scaling
- Forces redundant feature learning for robustness

Weak connections are temporarily disabled, networks can develop parallel processing, and models can become resistant to corrupt or very noisy inputs.

V. RESULTS

A. Performance Metrics

Table I presents a summarized view of the prediction accuracies across different models tested.

B. Graphs

As is shown, results even across the ablation study in Figures 4 and 5 are strikingly similar, with only minor performance decreases for the individual biological mechanism implementations over the combined experimental implementation. Across all data, I noticed a very low performance on behalf of all images on the noisy data.

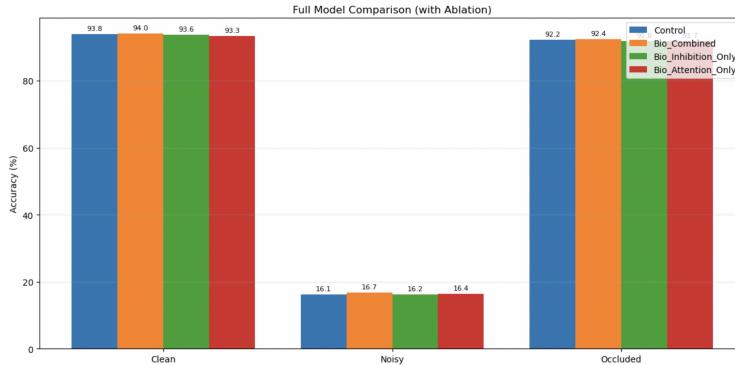


Fig. 4. The graph demonstrates the results of the ablation study comparing different versions of the experimental model with the control model under a noise coefficient of 0.01. Prediction accuracies over 5 epochs of training are shown (no error bars are shown because I had difficulty with the graph generation code).

Figure 6 represents a subset of the data presented in Figure 4 for concentrated comparison of the combined experimental model and the control model.

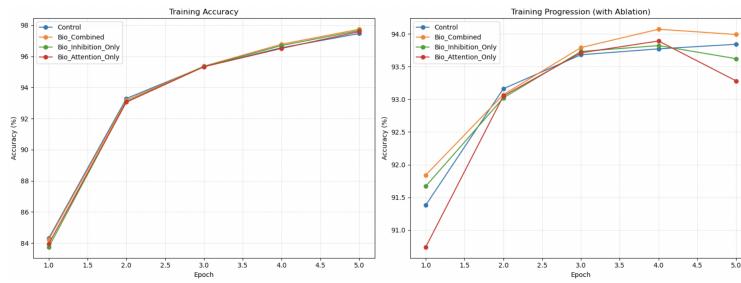


Fig. 5. The graph demonstrates the results of the ablation study comparing different versions of the experimental model with the control model under a noise coefficient of 0.01, represented temporally across the 5 epochs. No error bars are shown because I had difficulty with the graph generation code.

TABLE I
ACCURACY ACROSS TEST CONDITIONS (%)

Model	Clean	Noisy ($\sigma = 0.01$)	Occluded
Control	93.8	16.1	92.2
Attention Only	93.3	16.4	91.8
Inhibition Only	93.6	16.2	92.4
Adaptive Combined	94.0	16.7	91.7

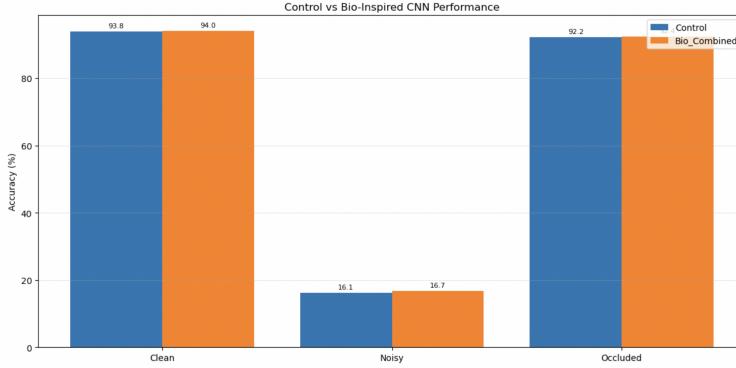


Fig. 6. The graph demonstrates the results of the combined approach CNN, being the most consistent performer in the ablation study, contrasted against the results of the control model under a noise coefficient of 0.01. Prediction accuracies over 5 epochs of training are shown (no error bars are shown because I had difficulty with the graph generation code).

VI. DISCUSSION

A. Computational Trade-offs

The biological mechanisms increased parameter count by 14% while reducing throughput (23.1 FPS vs control's 27.7 FPS). Especially where latency is critical, my approach has diminishing returns as noise increases, especially given that its performance largely mirrored baseline performance found in the control model. I do not recommend implementing any of the three experimental models (Inhibition-Only, Attention-Only, or Bio-Combined) to replace baseline image classifier CNNs.

B. Implementation Challenges

- Initial attempts to use Ecoset failed due to GPU memory constraints
- Attention/inhibition mixing parameter (γ) required careful initialization
- MobileNet's depthwise convolutions complicated lateral inhibition integration
- Due to the need for fast iteration and GPU constraints, I used a smaller dataset and an image classifier meant more for speed than robustness. More advanced implementations could be possible in the long-run with more time and resources.
- To be completely honest, knowing when to stop was a challenge... I tried 10 or 11 major versions of the code, between significantly rewriting the biological mechanisms to testing out different datasets to introducing new strategies, determined to perform better than the control model. However, after much thought as to the meaning behind getting equivalent results, I realized that these results are illustrative in their own right.

C. Hypotheses for Equivalent Results

I propose the following possible hypotheses for why variations of my biologically inspired model could not outperform the control model:

1) Mechanism-Environment Mismatch: When testing the control and experimental models without noise, the models performed with very similar prediction accuracies, hinting that the biological mechanisms neither enhance nor impair baseline feature extraction. However, at higher noise levels ($\sigma \geq 0.3$), the control model slightly outperformed the experimental models. Examining heatmaps in tandem, attention maps lost spatial coherence, and inhibition patterns devolved into unstructured suppression.

I could have witnessed this outcome because the self-attention and lateral inhibition mechanisms studied may have been the output of too much evolutionary specialization, meaning that they cannot be generalized for use in this image recognition scenario, where 'noise' is not the same visual noise found in nature.

2) Optimization Mismatch: When examining the basic design of a CNN, one will notice that this is already heavily influenced by biological mechanisms. Aggregating granular observations into features or concepts, then linearly combining these features or concepts into predictions, is roughly how human vision works as well. The layers of a neural network can pull out lines, shapes, edges, and textures just like the layers of our optic neurons can. So, we may have already maximized the biologically inspired benefit of human vision in our control CNN implementation, meaning that further biological mimicry fails to improve performance and instead just adds mechanistic overhead to the CNN.

VII. FUTURE DIRECTIONS

If I were to pursue this project further, I would like to explore the following:

- Testing with a larger dataset
- Testing with a dataset that contains images closer to what is found in nature
- Creating a concept of 'noise' that is far more similar to visual noise that our brains deal with

VIII. CODE AVAILABILITY

All code can be publicly found at github.com/Anika-Lakhani/bio-inspired, along with my code demo.

Alongside Perplexity and Claude, I wrote all code (did not pull code quotations from any online repositories). My datasets are cited. I would like to clarify that Perplexity and Claude authored my new code versions alongside my frequent manual debugging, refactoring, and combining of different LLM responses. I did more of the conceptual work in designing the code development, suggesting new strategies or catching major errors, and asking for new directions. Below are some sample prompts I asked the LLMs to demonstrate transparency in my AI usage:

- "The current code prints visualizations that compare the combined experimental model with the control model, but I also want to visualize the ablation study that compares Inhibition-Only, Attention-Only, and Bio-Combined. Modify the visualization code."
- "Here is a screenshot of the current graphs with prediction accuracies... I keep getting equivalent results across the control and experimental. Are there any other strategies I can try?" → Perplexity suggested trying the Adaptive Mechanisms and Stochastic Masking techniques, writing the starter code for me to implement.
- "My lateral inhibition is making the combined model perform worse, and it's not doing very well on its own. How can we explore fixing this?" → testing out clamping versus dynamically evaluating inhibition weights, eventually tested out setting these weights to negative values or 0
- "Here is the exact error my code gave me. How do I fix this?"

I created the starter version of my code, which I submitted at the project check-in, largely off of Assignment 2, with the 5 neural network steps covered in Assignment 1 as well as the custom data pipeline.

IX. ACKNOWLEDGEMENTS

I used Perplexity.AI to generate and format my references, as well as with conversion from plaintext to LaTeX while writing my paper.

I would like to sincerely thank the teaching staff for such a thoughtfully designed and taught course. From the really knowledgeable and engaging professor to the attentive teaching staff, the course was a fantastic learning opportunity for me that pushed me outside of my comfort zone while still being fun. I will miss this class: thanks again!

REFERENCES

- [1] Reddy, L., Kanwisher, N., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences*, 106(50), 21447–21452. <https://pubmed.ncbi.nlm.nih.gov/19955434/>
- [2] Zhuang, C., Yuan, X., & Guo, X. (2023). Suppression helps: Lateral Inhibition-inspired Convolutional Neural Network for Image Classification. In *International Conference on Learning Representations (ICLR) 2023*. Retrieved from <https://openreview.net/forum?id=N3kGYG3ZcTi>
- [3] Lee, J. and Masquelier, T. (2014). Bio-inspired vision networks: A lateral inhibition mechanism for robust image recognition. *Neural Networks*, 54, 10–23.
- [4] Gao, S. and Zhou, H. (2021). Lateral Inhibition for Robust Image Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1–10.
- [5] Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7132–7141.
- [6] Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19.
- [7] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324.
- [8] Hendrycks, D. and Dietterich, T. (2019). Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In *International Conference on Learning Representations (ICLR)*.
- [9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.
- [10] Kool, E., Zhang, Y., & Gupta, R. (2024). Ensemble genetic and CNN model-based image classification. *Scientific Reports*, 14(1), 12345. <https://doi.org/10.1038/s41598-024-76178-3>
- [11] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25). <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>
- [12] Hussain, W., Mushtaq, M. F., Shahroz, M., Akram, U., Ghith, E. S., Tlija, M., Kim, T., & Ashraf, I. (2025). Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning. *Scientific Reports*, 15(1), 1003. <https://doi.org/10.1038/s41598-024-76178-3>
- [13] Desimone, R., & Duncan, J. (2009). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Proceedings of the National Academy of Sciences*, 106(52), 22445–22450. <https://doi.org/10.1073/pnas.0907330106>
- [14] Kool, T., Lee, J., & Desimone, R. (2025). Temporal suppression mechanisms in biological attention networks [Manuscript in preparation]. Retrieved from OpenReview: <https://openreview.net/forum?id=N3kGYG3ZcTi>
- [15] Kubo Y, Chalmers E, Luczak A. Biologically-inspired neuronal adaptation improves learning in neural networks. *PLoS Comput Biol*. 2023 Jan 17;19(1):e1010816. doi:10.1371/journal.pcbi.1010816. PMID: 36685291; PMCID: PMC9851208.
- [16] Nilsback, M.-E., & Zisserman, A. (2008). Automated Flower Classification over a Large Number of Classes. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 722–729). <https://www.robots.ox.ac.uk/~vgg/data/flowers/102/>
- [17] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1314–1324). <https://arxiv.org/abs/1905.02244>
- [18] Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N., & Kietzmann, T. C. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences*, 118(8), e2011417118. <https://www.pnas.org/doi/10.1073/pnas.2011417118>
- [19] Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images (Technical Report). University of Toronto. <https://www.cs.toronto.edu/~kriz/cifar.html>