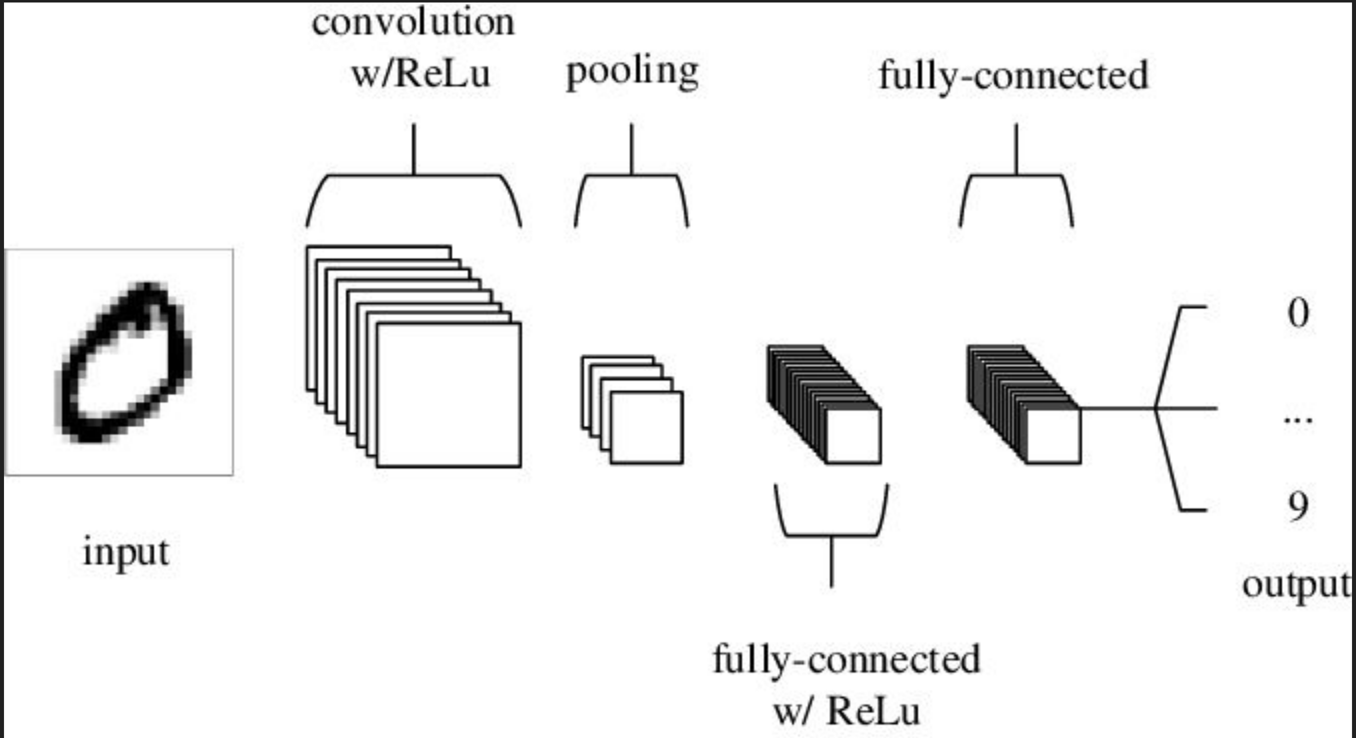
Deciphering Cognition: An Investigation into CNN Architectures

Justice Tomlinson and Shysta Sehgal

Introduction

- Convolutional Neural Networks are inspired by the biological **visual cortex**.
- CNNs consist of an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer.
- **Convolutional layers** capture spatial hierarchies and patterns in the images.
- **Pooling layers** help in developing a compressed representation of input data.
- **Fully connected layers** do the high-level reasoning after the feature extraction and down-sampling to classify the input.

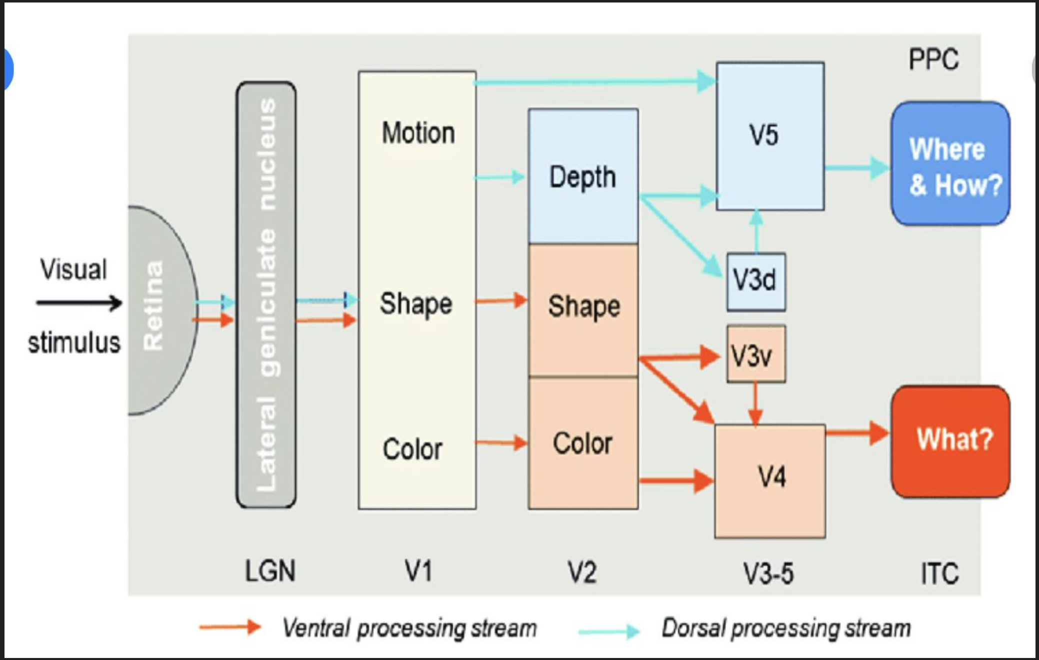
CNNs



convolution
w/ReLu
pooling
fully-connected
input
0
...
9
output
fully-connected
w/ ReLu

Similarities with Human Visual Cortex

- Hierarchical Processing: We start processing simple features like edges in V1 and complex objects in ITC.
- Spatial Invariance: We can detect objects regardless of orientation, or variation in size. Pooling layers mirror this property.



Motivation

- Given the similarities in CNNs and human visual cortex, we want to understand how varying CNN architecture can provide understanding of human visual processing.
- Specifically, how do network depth and width affect the performance and learning abilities of CNNs?

Hypothesis: Increasing depth will enhance learning more effectively than increasing width.

Related Work

- He et al (2015) demonstrated that networks with increased depth significantly improved performance on image recognition tasks.
- Deeper networks can achieve higher levels of feature abstraction, analogous to complex cognition in humans

- Tan and Le (2019) determined that by doubling the depth, the computational cost roughly squares, while doubling the width or resolution increases the computational cost linearly.

Methodology

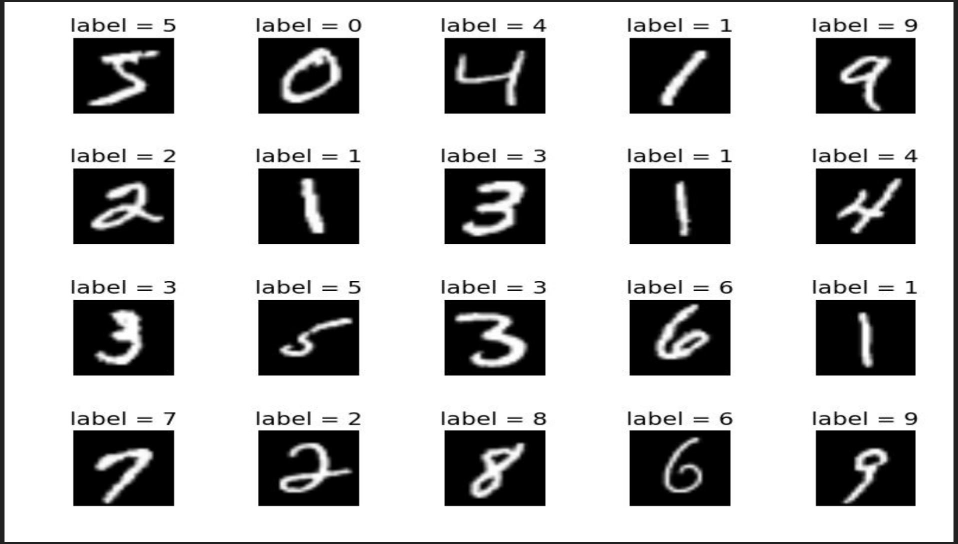
- **Models used**: One-layer, two-layer, three-layer CNNs (90 total). All layers are followed by pooling with the final layer in each model being a fully connected layer.

- **Training parameters**:
  - Epochs: 5
  - Batch size: 64
  - Learning rate: 0.001
  - Optimizer: Stochastic Gradient Descent (SGD) with momentum 0.9.

- **Training setup**:
  - Loss: Cross-entropy loss

Dataset

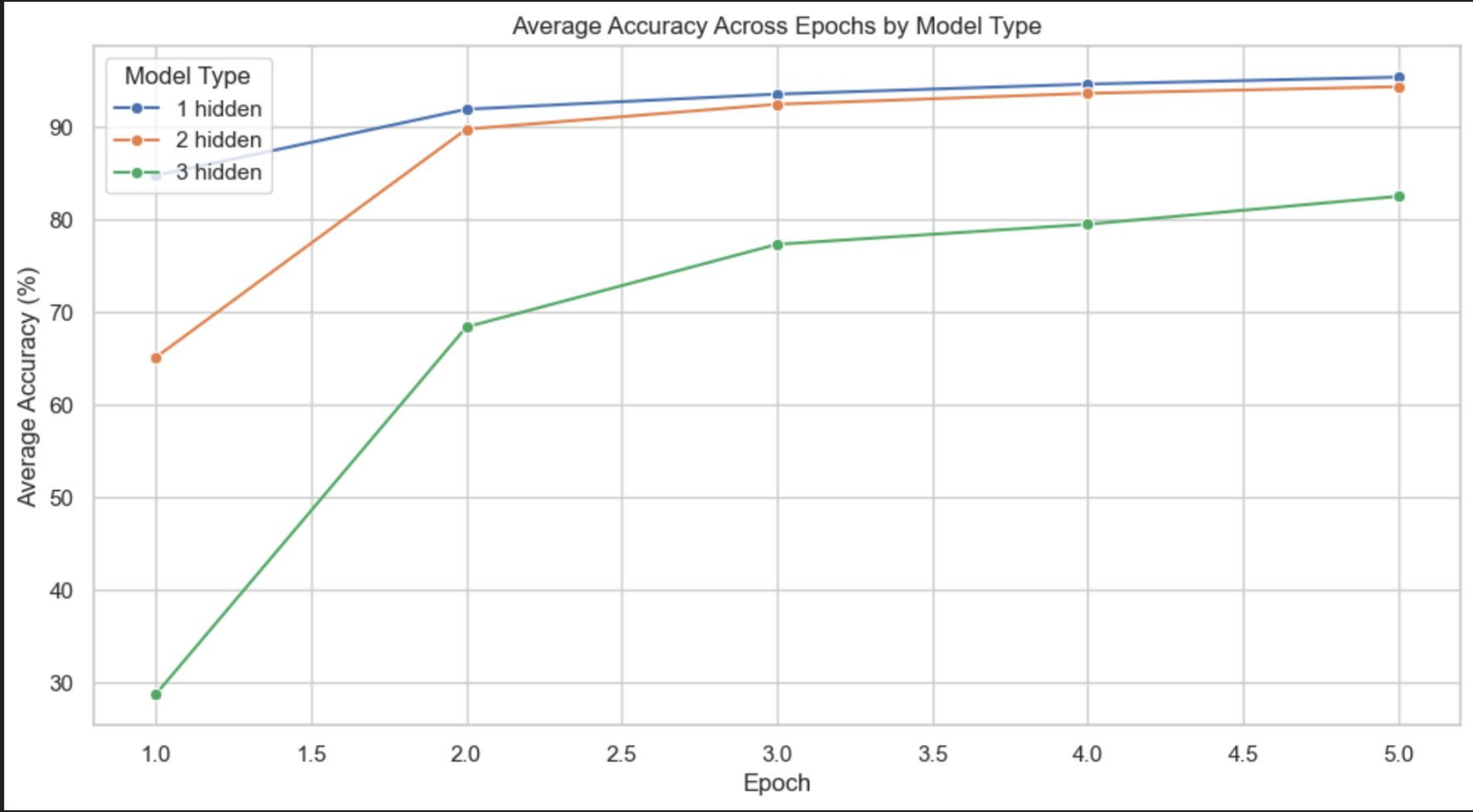
- **MNIST**: 60,000 training images and 10,000 testing images
- Black and white
- 10 classes



Results: Model-wise Summaries

Model Type	Mean Accuracy	Mean Loss
1 Hidden Layer	95 +- 1	0.14 +- 0.03
2 Hidden Layer	95 +- 3	0.16 +- 0.09
3 Hidden Layer	86 +- 22	0.39 +- 0.58

Accuracy Measures

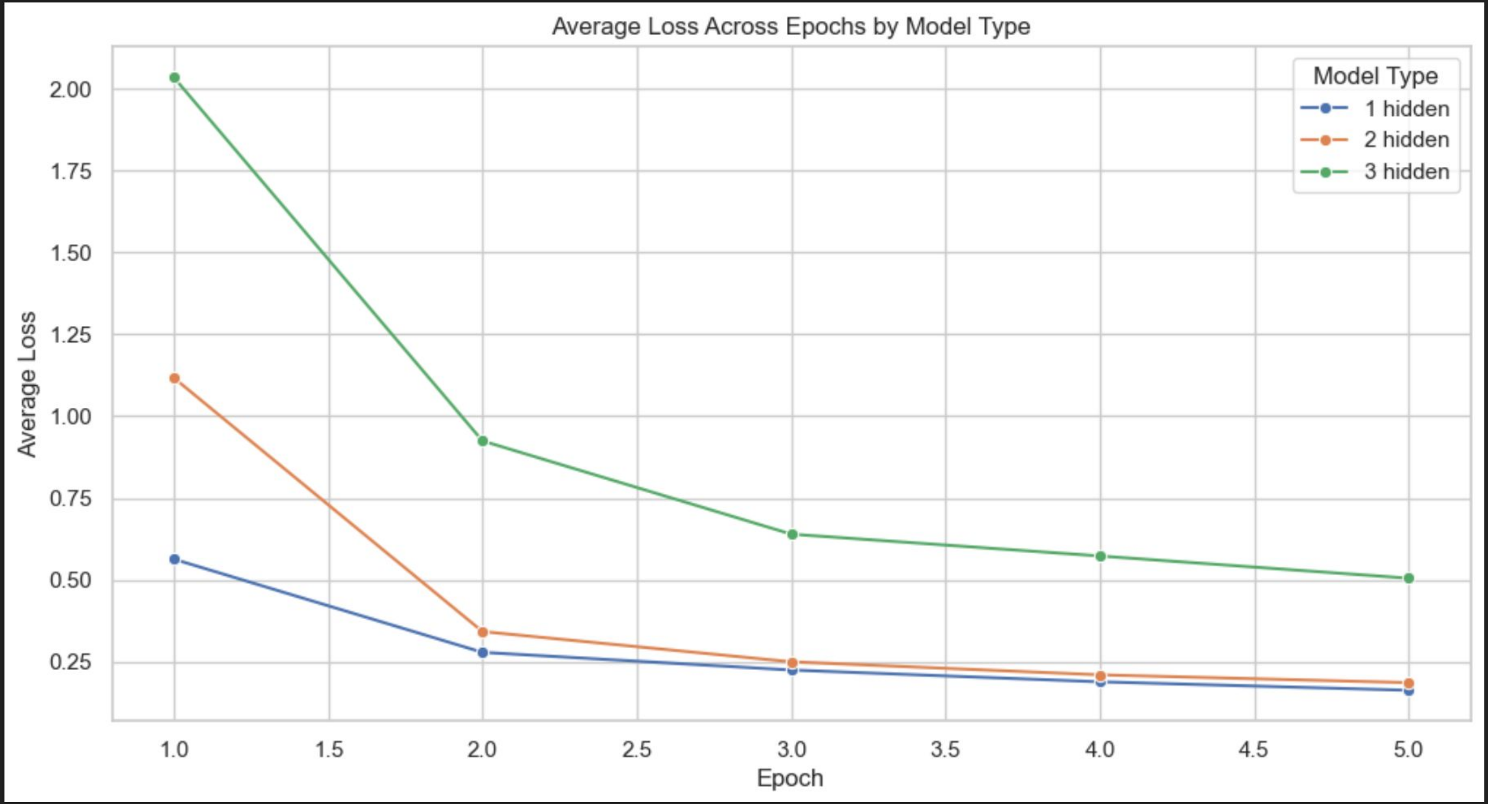


H1: 95%

H2: 94%

H3: 84%

Loss Measures



Interpretation

- Sometimes, higher-level abstraction offered by adding more layers might be unnecessary for the task.

- Adding more filters allows the model to learn a diverse set of features, providing richer representation for decision-making layers.

- More filters probably also help the model to generalize better to unseen data.

Conclusion

- The experimentation with CNN architectures on the MNIST dataset demonstrates the delicate balance between model complexity (depth and width) and performance.

- After adding a certain number of layers that mitigate any over-simplicity or over-complexity concerns, it is both computationally and performance-wise better to experiment with width to increase accuracy.

- **Future research**:
  a. Testing on datasets like Omniglot, because of the diversity of characters and languages
  b. Expanding testing on different cognitive tasks like auditory processing

References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science, 1*, 417-446. https://doi.org/10.1146/annurev-vision-082114-035447

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019-1025.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*. https://arxiv.org/abs/1905.11946

Thank you!

Do you have any questions?