Cassy Cormier

Lab 07 Reflective Journal: Chihuahua or Muffin with CNN

15698 Computer Vision 1378

Professor Anna Devarakonda

April 26th, 2025

Distinguishing Chihuahuas from Muffins: A Reflective Journey on Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have become essential tools in modern computer vision, enabling machines to perform complex image classification tasks with remarkable accuracy (What are convolutional neural networks?, n.d.). These networks process data by passing it through an input layer, then through hidden layers with activation functions, and finally, through an output layer. In traditional neural networks (NN), using a Multi-Layer Perceptron (MLP) is beneficial when the order of the data is not crucial (HCC). However, in image classification, preserving the position of pixels is vital for understanding the context of the image and cannot be changed as it can be in other datasets. CNNs use matrix multiplication to identify patterns within an image without the traditional need to flatten the image first (What are convolutional neural networks?, n.d.). A CNN contains an input layer, convolutional layers for feature extraction, fully connected layers for classification, and an output prediction layer. The transition to this deep-learning pipeline has automated the feature learning and extraction process (HCC).

Distinguishing between visually similar objects remains a significant challenge, as illustrated by the "Chihuahua vs. Muffin" classification problem. This lab explored the ability of a CNN to accurately differentiate between images of Chihuahuas and muffins, two classes that can appear deceptively alike in certain contexts. The purpose of the experiment was to deepen understanding of CNN architecture, feature extraction, and model evaluation in cases where traditional visual cues are insufficient. This reflective journal examines the development and performance of a CNN for image classification in the "Chihuahua vs. Muffin" task, analyzing its architectural advantages over traditional NNs, evaluating model accuracy and misclassification trends, reflecting on technical challenges encountered and resolved, and considering both the real-world applications and ethical implications of deploying deep learning models for visual recognition.

During the early stages of training, the model exhibited unstable learning with very high loss values and fluctuating validation accuracy. However, after adjusting the learning rate from 0.005 to 0.0005, there was a significant improvement in both training and validation performance: losses steadily decreased and accuracies rose consistently, with minimal signs of overfitting. Further lowering the learning rate to 0.0001 resulted in slower but more stable training, although progress plateaued around 80% validation accuracy. The best results were achieved with a learning rate of 0.0005, leading to strong generalization and validation accuracy reaching 86.7%. In later experiments without further changes, training accuracy consistently surpassed 91%, and validation accuracy peaked at 100%, showing that the model had reached robust and reliable performance.

Phase	Learning Rate	Epochs	Training Behavior	Validation Behavior	Notes
Initial	0.005	0–9	Unstable, high loss	Poor generalization	Learning rate too high
Test 2	0.0005	10–19	Smooth convergence	Good generalization	Best performance observed
Test 3	0.0001	20–29	Very slow learning	Plateaued performance	Learning rate too low
Final	0.0005	30–44	Excellent accuracy	Outstanding validation	Confirmed 0.0005 as optimal

(ChatGPT, n.d.)

An interesting pattern in the misclassifications was that, depending on the test instance, either some muffins were incorrectly classified as chihuahuas while all chihuahuas were correctly classified, or some chihuahuas were incorrectly classified as muffins while all muffins were correctly classified. At no point were both classes simultaneously misclassified. This suggests that the model was picking up on certain features that, in some cases, made it confuse one class for the other without causing mistakes in both direction at once.

The basic NN (MySkynet) achieved high training accuracy quickly but showed signs of overfitting, as it could not fully capture the spatial features of the images. In contrast, the ChihuahuaMuffinCNN learned more gradually but ultimately achieved better validation accuracy, demonstrating stronger generalization to unseen data. This highlights the advantages of CNNs for image classification tasks where preserving pixel structure is important.

One major challenge was the initial overfitting, where the model achieved high training accuracy but failed to generalize to validation data, indicating a patter of memorizing rather than learning. Lowering the learning rate (LR) from 0.005 to 0.0005 significantly improved performance, boosting training accuracy from 62% to 86% and greatly improving validation results. Especially with the Adam optimizer, too high a learning rate caused unstable convergence, while different optimizers have varying sensitivity to learning rate choices

(Brownlee, 2021). Setting the learning rate too low, as in the third test, caused the model to improve extremely slowly and eventually plateau without reaching optimal performance. Overall, this lab provided valuable insights into the strengths and limitations of CNNs in image classification tasks, setting the foundation for broader reflections on the role of deep learning in both technical and ethical contexts.

This type of image classification model has the potential for valuable real-world applications such as in property management. For example, CNNs can be used by property managers and owners by enabling smart security systems that detect intruders, identify license plates, and monitor suspicious activity, reducing manual labor costs. They can also streamline asset inventory by automatically reporting property contents from images, helping landlords save time and reduce insurance-related expenses. These technologies can address key challenges in property management while improving efficiency and cost control.

While CNN-based models offer efficiency and enhanced security, they also raise ethical concerns related to privacy, surveillance, and potential biases in image recognition. In property management, constant monitoring through smart security systems could infringe on tenants' privacy rights if not properly regulated. Additionally, inaccuracies or biases in classification could lead to wrongful accusations or discrimination, highlighting the need for transparent model training practices and ethical deployment guidelines. Overall, balancing technological advancement with ethical responsibility is critical to ensuring that CNN-based solutions provide real value without compromising individual rights or fairness.

References

Brownlee, J. (2021, January 13). *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*. Retrieved from Machine Learning Mystery: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/

ChatGPT. (n.d.). Retrieved from ChatGPT: https://chatgpt.com/

HCC. (n.d.). Convolutional Neural Networks. *Class Notes*. (HCC, Ed.) Houston, Texas, USA: HCC.

What are convolutional neural networks? (n.d.). Retrieved April 26, 2025, from IBM: https://www.ibm.com/think/topics/convolutional-neural-networks