Leo Milligan

In my exploration of machine learning models for image classification, I focused on two main approaches: Convolutional Neural Networks (CNNs) and Transformers. The CNN approach involves examining small portions of an image first, then gradually building up a holistic understanding, similar to assembling a puzzle starting from one corner. The Transformer, conversely, looks at the entire image at once, understanding how different parts relate on a global scale.

The project utilized a diverse image dataset, requiring standardization in terms of size and lighting conditions. To prevent the models from simply memorizing the dataset, I introduced data augmentation techniques like flipping and rotating the images.

My CNN architecture was relatively straightforward, comprising layers that detect patterns and interpret them to classify the images. This setup was efficient and less demanding in terms of computational resources. However, it had limitations in handling variations in the position or orientation of image subjects. To improve its adaptability, I trained it with images that varied in placement and orientation.

In contrast, the Transformer model required more nuanced management due to its propensity for overfitting and its higher computational demands. It's akin to having a detailed-oriented mindset — comprehensive but resource-intensive. To balance this, I employed regularization techniques and adjusted learning rates, ensuring effective learning without overburdening the system.

The Transformer ultimately achieved slightly higher accuracy than the CNN, but this came with the trade-off of increased complexity and resource requirements. The CNN, while marginally less accurate, was more efficient and practical for scenarios with limited computational resources.

I encountered specific challenges with each model. The CNN struggled with the vanishing gradient problem in its deeper layers, which I mitigated by incorporating batch normalization. This adjustment stabilized the training process, allowing for deeper and more effective learning. The Transformer's main challenge was achieving stable and consistent training convergence. I addressed this by carefully tuning the model's parameters and learning rate, guiding it towards more efficient learning paths.

In conclusion, this study provided valuable insights into the strengths and limitations of each architecture. While the Transformer offers a slight edge in accuracy, its resource-intensive nature makes the CNN a more feasible choice for situations with computational constraints. My project highlighted that in machine learning, the choice of model is highly dependent on the specific task requirements and available resources. Choosing the right model involves balancing these factors to achieve the best possible outcome.