Title: Pathway Vision Transformer for Image Recognition

Introduction:

The Pathway Vision Transformer (PaViT), developed by Ajibola Emmanuel Oluwaseun, is a novel architecture for image recognition that draws inspiration from Google's PaLM (Pathways Language Model) and few-shot learning techniques used in natural language tasks. The goal of PaViT is to develop a model that can perform well in image recognition tasks, while also being easy to train on CPU and outperforming existing Vision Transformer models.

Methodology:

PaViT was trained on a dataset of 2000 Kaggle images belonging to 4 classes, using a 4GB RAM CPU. The model was designed with 4 self-attention heads, and achieved an accuracy of 74%. The model's accuracy was further improved to 87% by adding 10 self-attention heads and linear layers. The author experimented with different normalization techniques and found that the use of Batch normalization layer yielded better performance compared to custom layer normalization. To achieve the best accuracy, the model was trained with 12 self-attention heads and 18 linearly stacked Dense layers, resulting in 90% accuracy on the same dataset.

Results:

After training on a larger dataset of 15000 Kaggle plant images belonging to 15 classes, using Google Colab NVIDIA T4 Tensor Core GPUs, PaViT achieved an impressive 98% accuracy. These results demonstrate that PaViT has the potential to outperform existing Vision Transformer models, while also being easy to train on CPU.

Discussion:

The success of PaViT can be attributed to its unique architecture and its ability to incorporate insights from natural language processing techniques. The self-attention mechanism in PaViT allows the model to focus on relevant features and learn from correlations between different features. The use of Batch normalization layer also contributed to the improved performance of PaViT. Additionally, PaViT's ability to achieve high accuracy on a relatively small dataset is a promising feature that could be particularly useful in scenarios where acquiring a large dataset is challenging.

Conclusion:

In conclusion, the Pathway Vision Transformer (PaViT) developed by Ajibola Emmanuel Oluwaseun, has

demonstrated remarkable results in image recognition tasks, with the potential to outperform existing Vision Transformer models. The model's fast training speed on CPU and ability to achieve high accuracy on a relatively small dataset are particularly promising features that could make it a leading model in the field of image recognition. Continued development and contributions from the community will be key to further improving PaViT's performance and applicability in real-world scenarios.