

Task 2:

Many Deep Learning based models are "CUDA-based research". One of the more recent CUDA-based research that I find interesting is the "Attention Is All You Need[1]" paper.

Why I find it interesting:

- This paper introduced the "Transformer" architecture, which is now one of the primary model architecture of choice.
- This architecture allows massive parallelisation really putting to use the numerous CUDA cores of GPUs.
- This architecture makes it really possible to do NLP, Image and multiple domains in a unified architecture.

[1] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

Name: Ramadoss, Adhithyan; Matriculation Number: 22322010

IdM: um41adoh