**Video Swin Transformers**

**Outline:**

* Introduction to Transformers [1]
* Limitations of Standard Transformers
* Visual Transformers
* What is SWIN Transformers
* Architecture
* Advantages over transformers
* Limitations on Temporal Modelling
* Video Swin Transformers
* Architecture
* Step by step representation of forward pass
* Advantages on performance and accuracy
* Usage

**Introduction to Transformers**

**Transformers**

**References:**

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. Retrieved from https://arxiv.org/abs/1706.03762

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Retrieved from https://arxiv.org/abs/2010.11929

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. Retrieved from https://arxiv.org/abs/2103.14030