SWIN TRANSFORMER (V2).

Buildupon the original Swin Transformer by improving scalability, training stability and performance on high resolution images.

It included/introduces key enhancements like log spaced attention, post normalization and MoE layers; making it more efficient and capable of handling ultra large datasets.

PROBLEMS IN SWIN T:-

① The original swin T used fixed -size local windows which limited its ability to model very large/longe range dependancies.

Solution: In swin V2 ; instead of evenly spaced windows, log spaced attention increases window sizes exponentially at deep layers.

② The layer Norm (LN) was placed before self attention and MLP layers causing instability when training deep models

Solution: They were simply moved afterwards (after residual connections).

③ The model had instability/difficulty in handling large feature variances indeeper layers.

Solution: V2 T of Swin introduced a new weight initialization technique to stabalize training of very deep networks.

What is MOE:?

Instead of using a single MLP layer; SWin V2 introdufed MoE (mixture of experts) where multiple MLP exist but only a few are activated per input.

*** This enchanced performance & scalability on large dataset.**

Log-spaced Attention.

Instead of keeping all attention windows uniform in size, Swin V2 progressively increase the window size at deeper level.

* The increase in the window size followed a logrithmic scale rather than a linear scale.

Note:

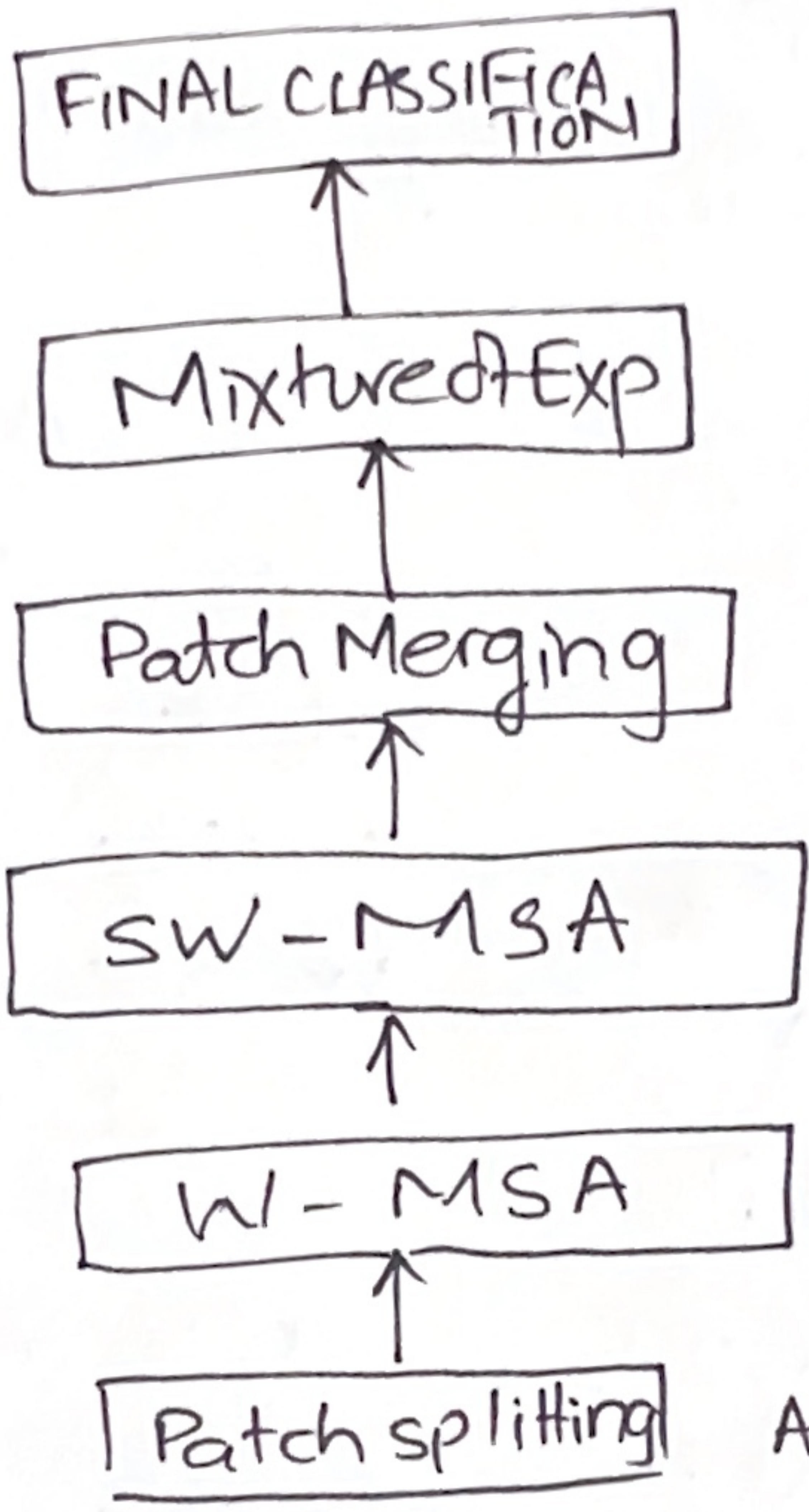
In the earlier layers, the attention windows are small, focusing on local detals.

In deeper layer they are exponentially large, i.e logspaced capturing global dependancies efficiently.

⇒ The approach maintains computational efficiency while capturing long range dependancies.

Basic Archietecture.

here is asimple SWIN V2 Transformer Model.



Replaced single MLP.

AkA Patch embedding.