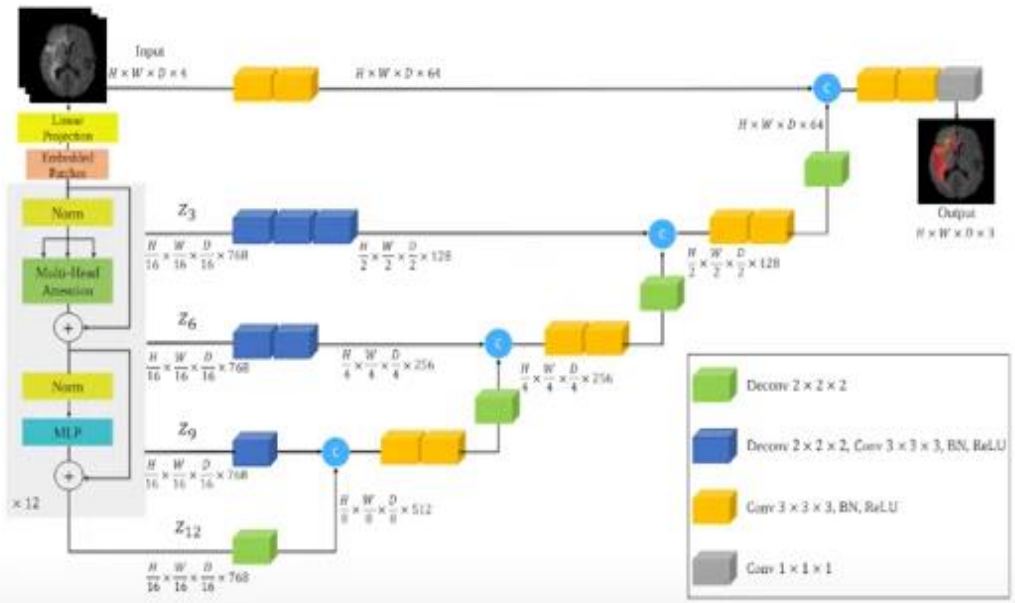
UNETR

UNEt TRansformers

What is UNETR?

UNETR is a 3D medical image segmentation model that uses a transformer encoder and a CNN-based decoder to predict the segmentation mask.

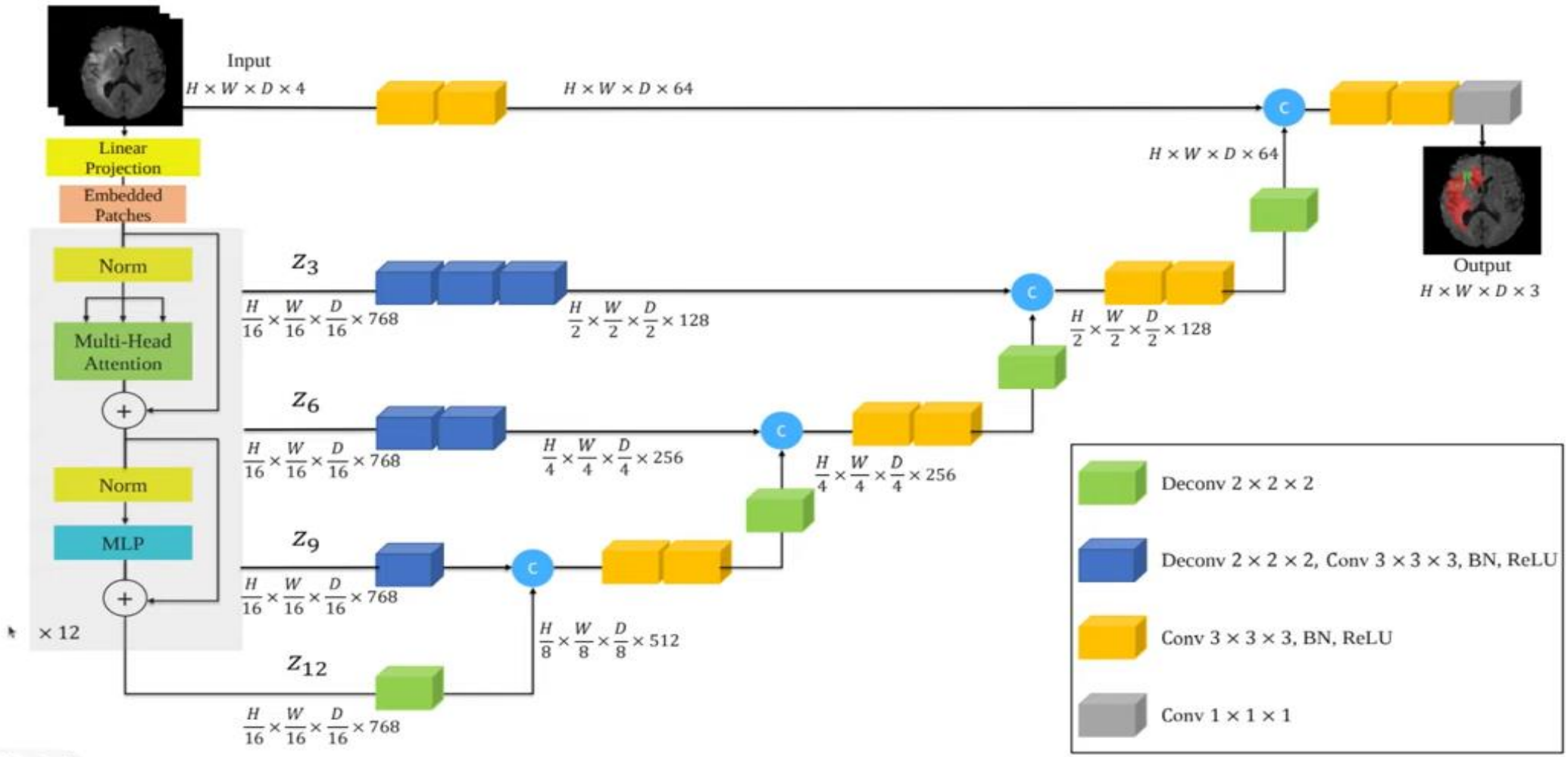


Key Features of UNETR?

1. It uses Vision Transformer as encoder to learn global contextual representations.

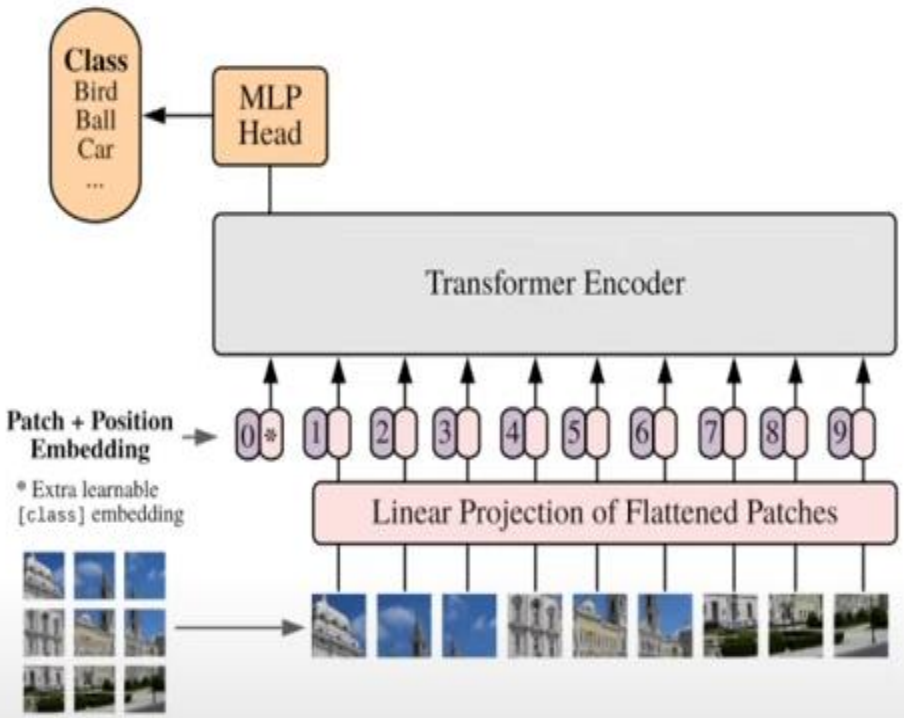
2. It uses a CNN Decoder to upsample the global representations and generate the final segmentation mask.

UNETR Architecture



What is Vision Transformer?

Vision Transformer (ViT) is an architecture that is used for image recognition. It is based on the Transformer architecture, which was originally developed for natural language processing. ViTs have been shown to achieve state-of-the-art results on a variety of image recognition tasks, including ImageNet classification.



ViT Variants

Model	Layers	Hidden size D	MLP size	Heads	Params
ViT-Base	12	768	3072	12	86M
ViT-Large	24	1024	4096	16	307M
ViT-Huge	32	1280	5120	16	632M