



# **Bài 7:**

# **Kiến trúc Transformer**

# **Semantic Segmentation**

AI Research Team - Sun\* AI Research

# Nội dung chính

01

Lý thuyết  
Transformer

02

Thực hành  
với PyTorch

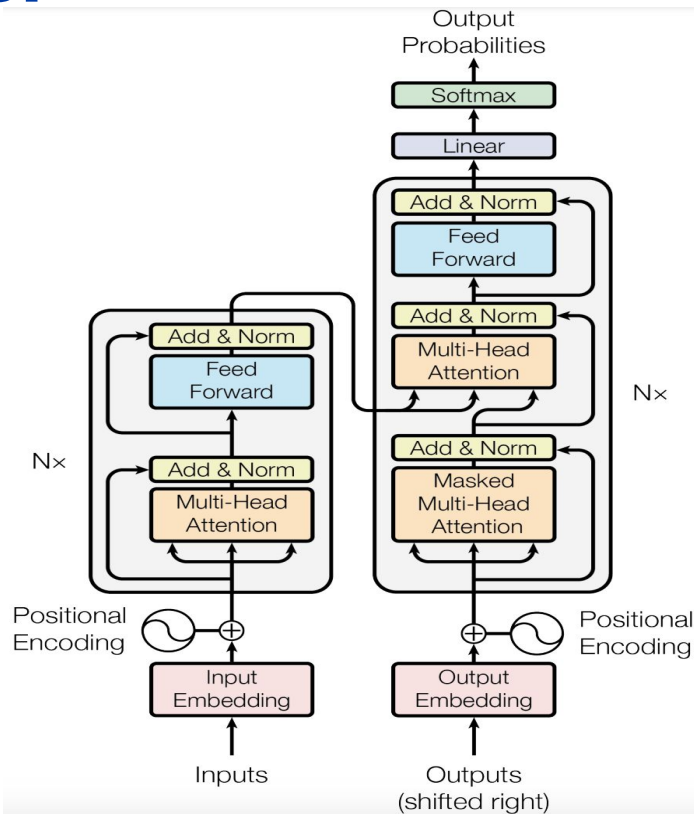


01

# Kiến trúc Transformer

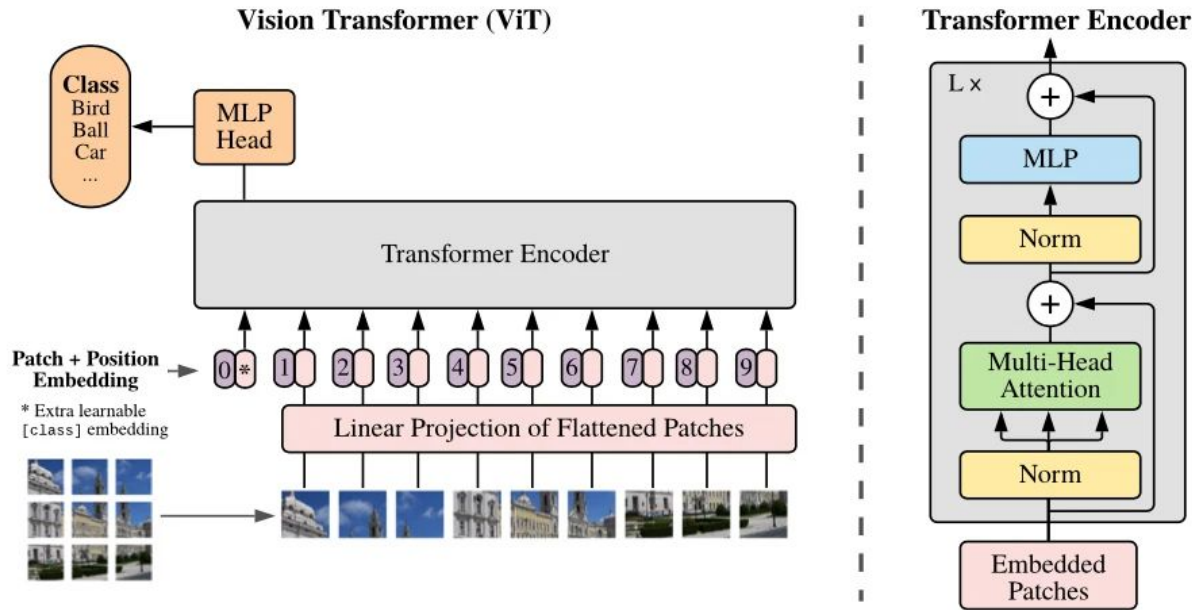
# Transformer

- Kiến trúc Transformer xuất hiện cho bài toán dịch máy
- Đạt hiệu năng cao trong các bài toán NLP
- Dựa trên cơ chế Multi Head Self Attention
- Xử lý dữ liệu dưới dạng chuỗi
- Input: (B, L, dim)
- Output: (B, L, dim')



Nguồn - Attention Is All You Need

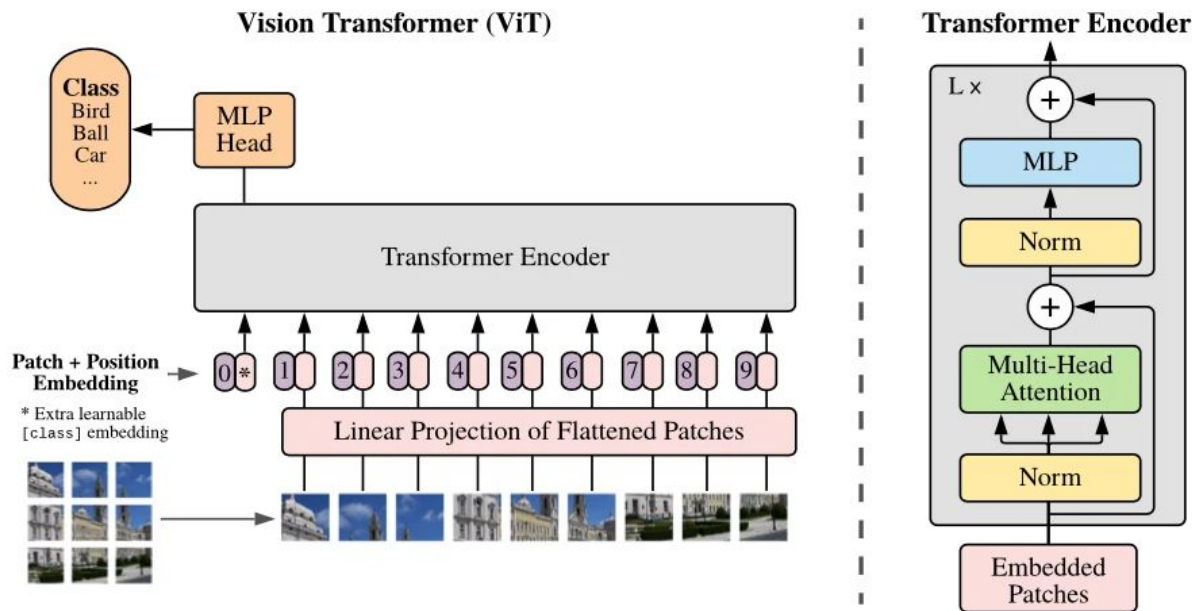
# Vision Transformer



Nguồn - Vision Transformer

- Input: Image  $(B, 3, H, W)$
- Embedded Patch:  $(B, C, H, W) \rightarrow (B, H/16*W/16, C, 16, 16)$
- Đặt  $L = H/16*W/16$ ,  $\text{dim} = C*16*16 \rightarrow (B, L, \text{dim}) \rightarrow \text{Transformer}$

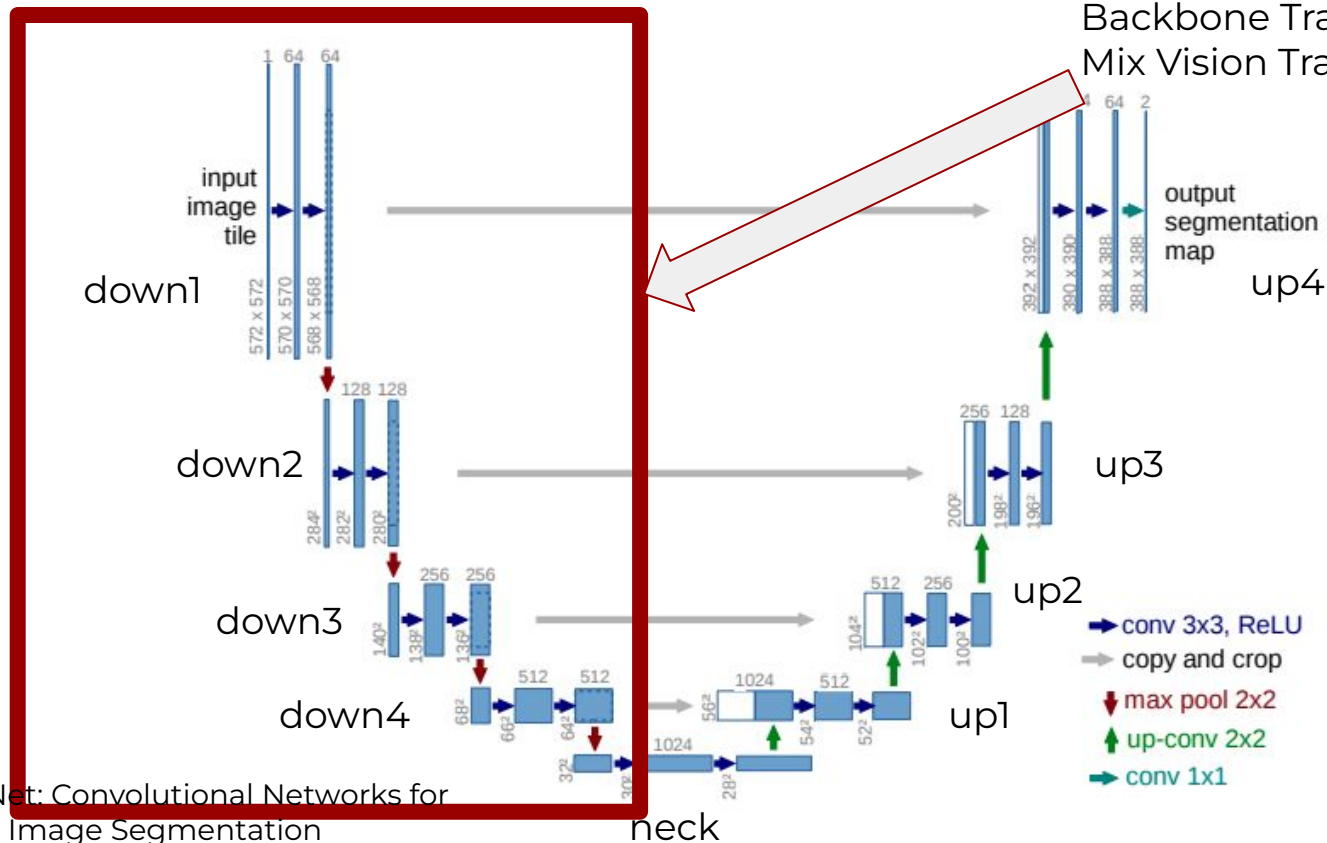
# Vision Transformer



- Transformer thường được train sẵn với nhiều dữ liệu trên ImageNet
- Transformer thường nắm global feature rất tốt so với mạng CNN -> hiệu năng cao
- Nhiều biến thể của ViT như: Pyramid Vision Transformer, Swin Transformer, ...

Nguồn - Vision Transformer

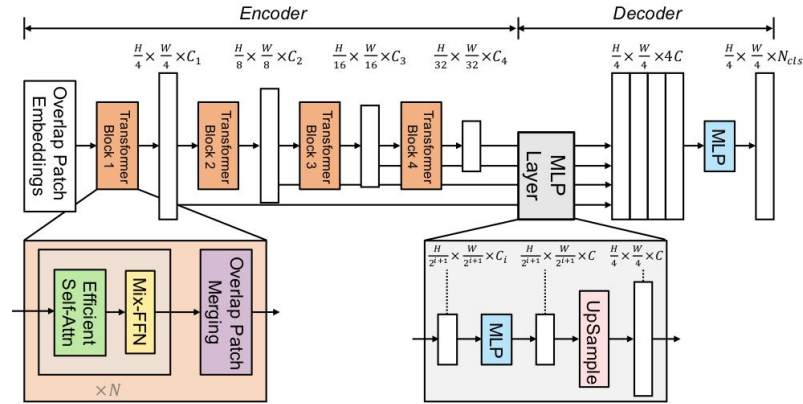
# Sử dụng Transformer làm backbone



Nguồn: U-Net: Convolutional Networks for Biomedical Image Segmentation

neck

# SegFormer



- Là mô hình Transformer chuyên dụng cho bài toán Semantic Segmentation
- Đạt hiệu năng cao
- Sức mạnh đến chủ yếu từ backbone Mix Vision Transformer (MiT)

<https://viblo.asia/p/paper-explain-segformer-mo-hinh-transformer-don-gian-hieu-qua-cho-bai-toan-segmentation-1VgZv2XmZAw>

Nguồn: Paper SegFormer





02

**Thực hành với PyTorch**