UNIVERSIDAD NACIONAL DE COLOMBIA

Visual Transformers

Sebastian López ·, Santiago Pineda, · Rafael Mejia , · Andrés Álvarez

Digital Signal Processing and Control Group - (GCPDS)
Universidad Nacional de Colombia
Manizales, Colombia
July 2023

Contenido

Split the image in Patches

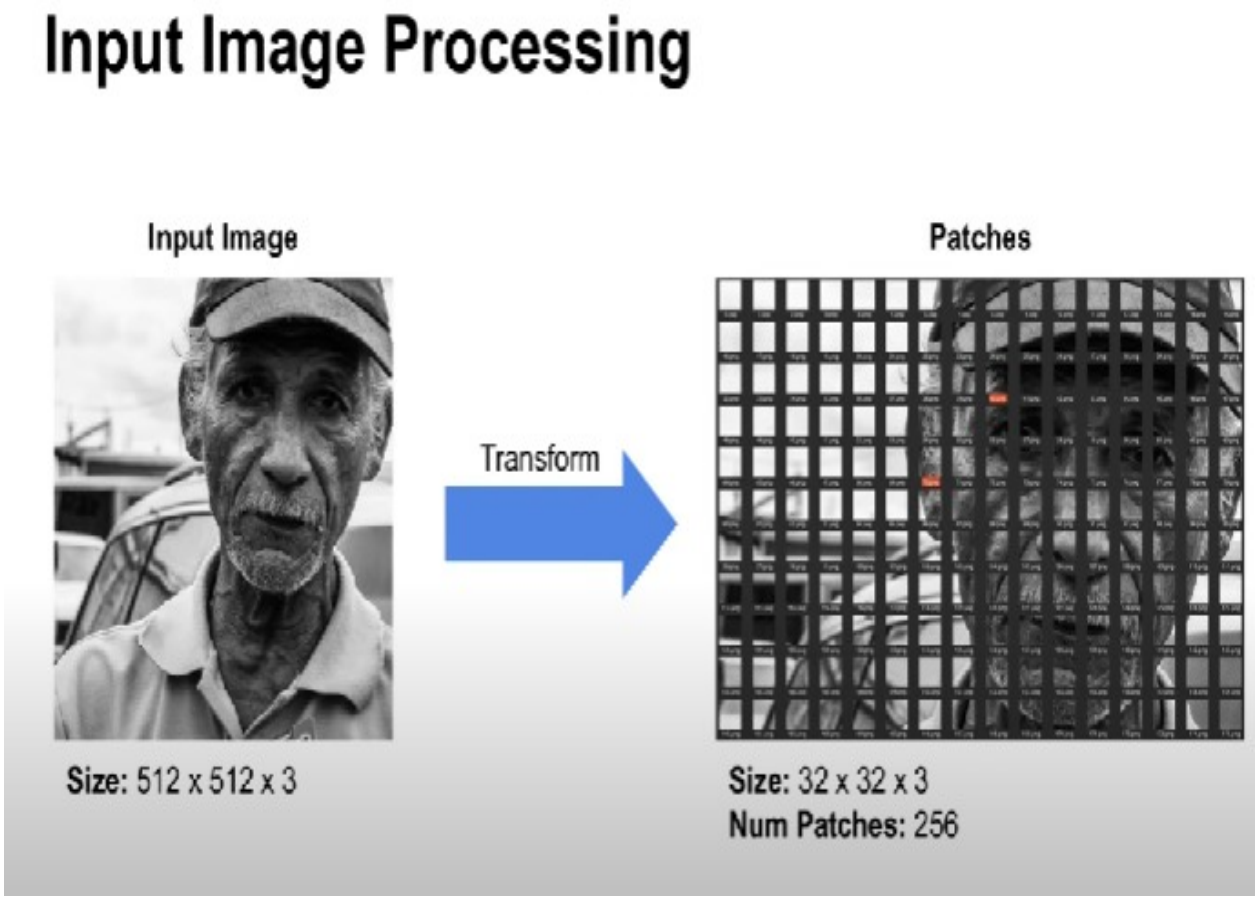


Figure: Split the image into patches

Split the image in Patches

UNIVERSIDAD NACIONAL DE COLOMBIA

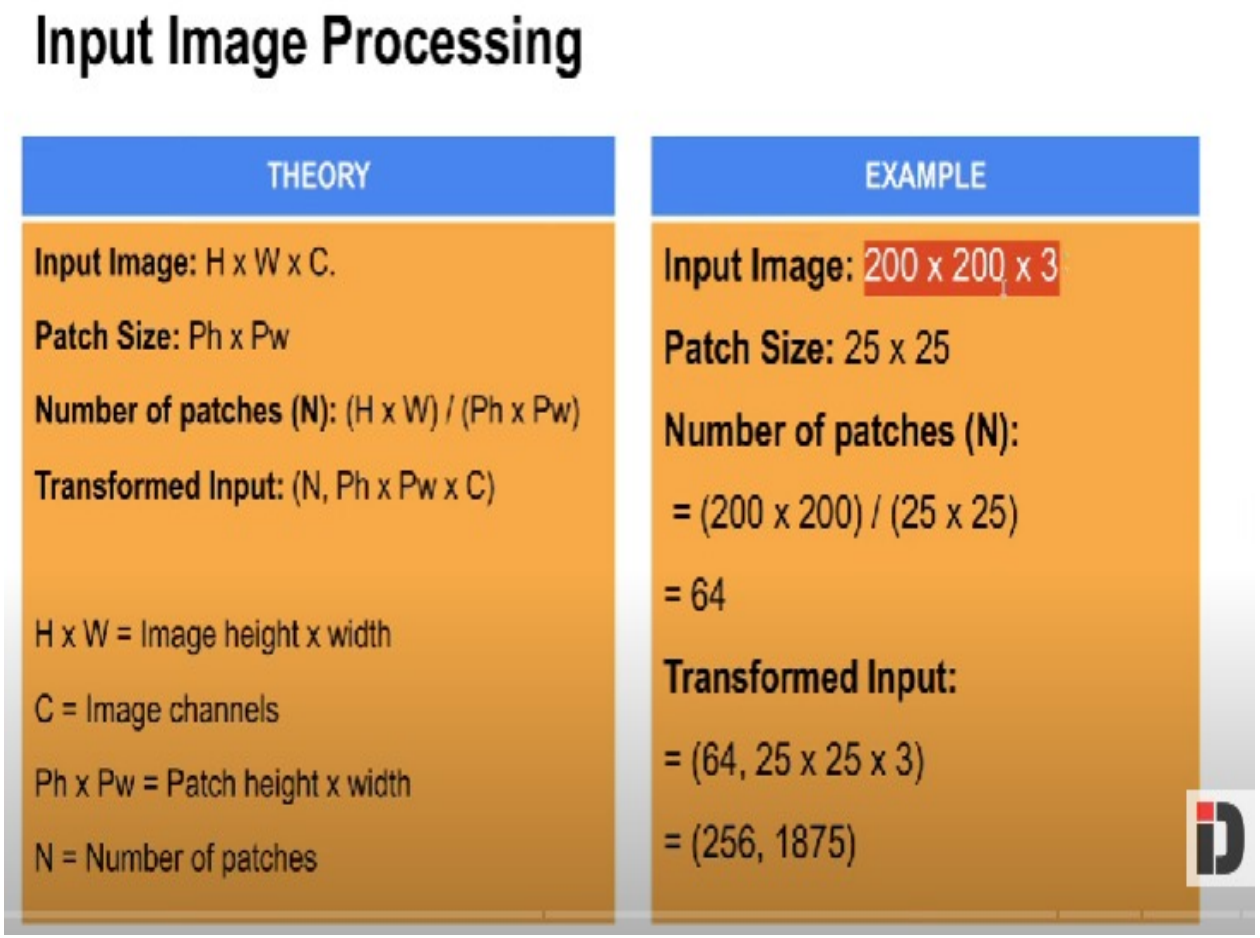
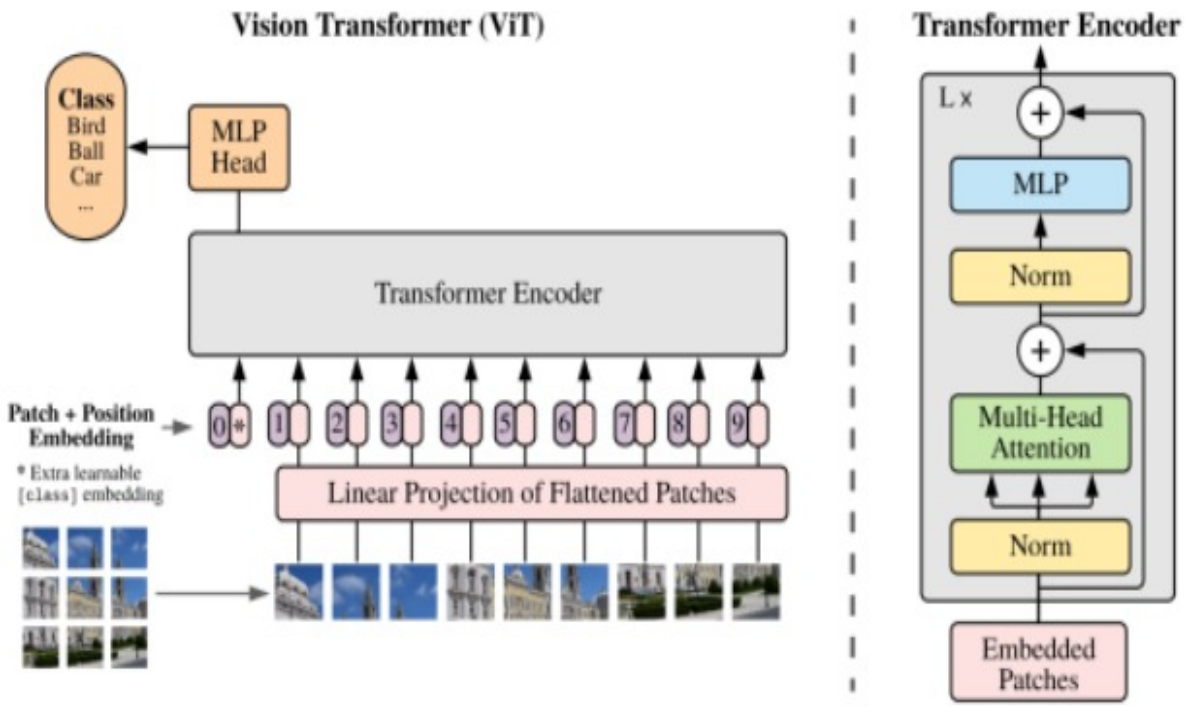


Figure: Split the image into patches

Visual Transformer Architecture



The architecture of the proposed Vision Transformer (ViT)

Figure: Visual Transformer Architecture

Positional Embedding

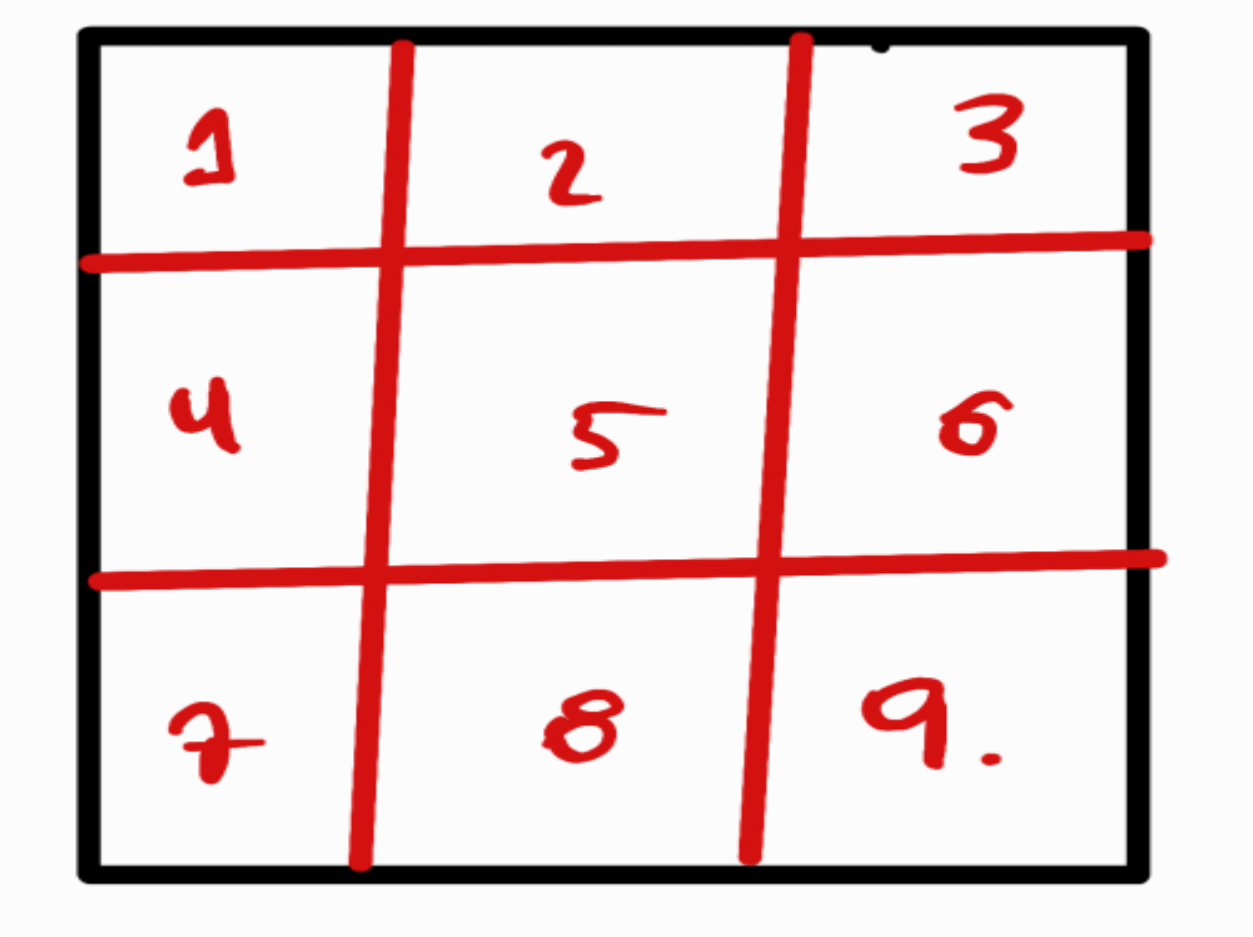


Figure: Positional Embedding

Class Token

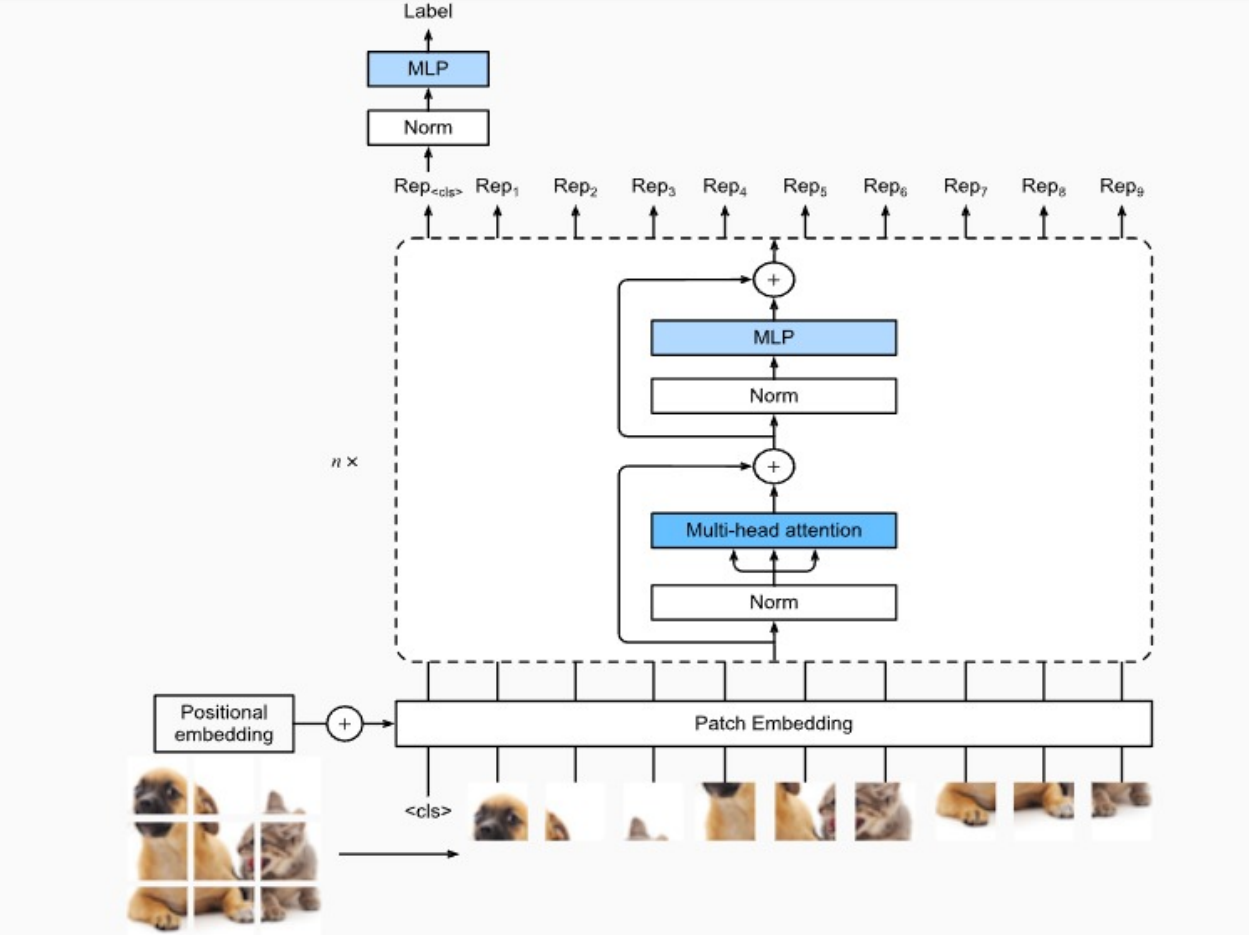


Figure: Class Token

Transformer Encoder

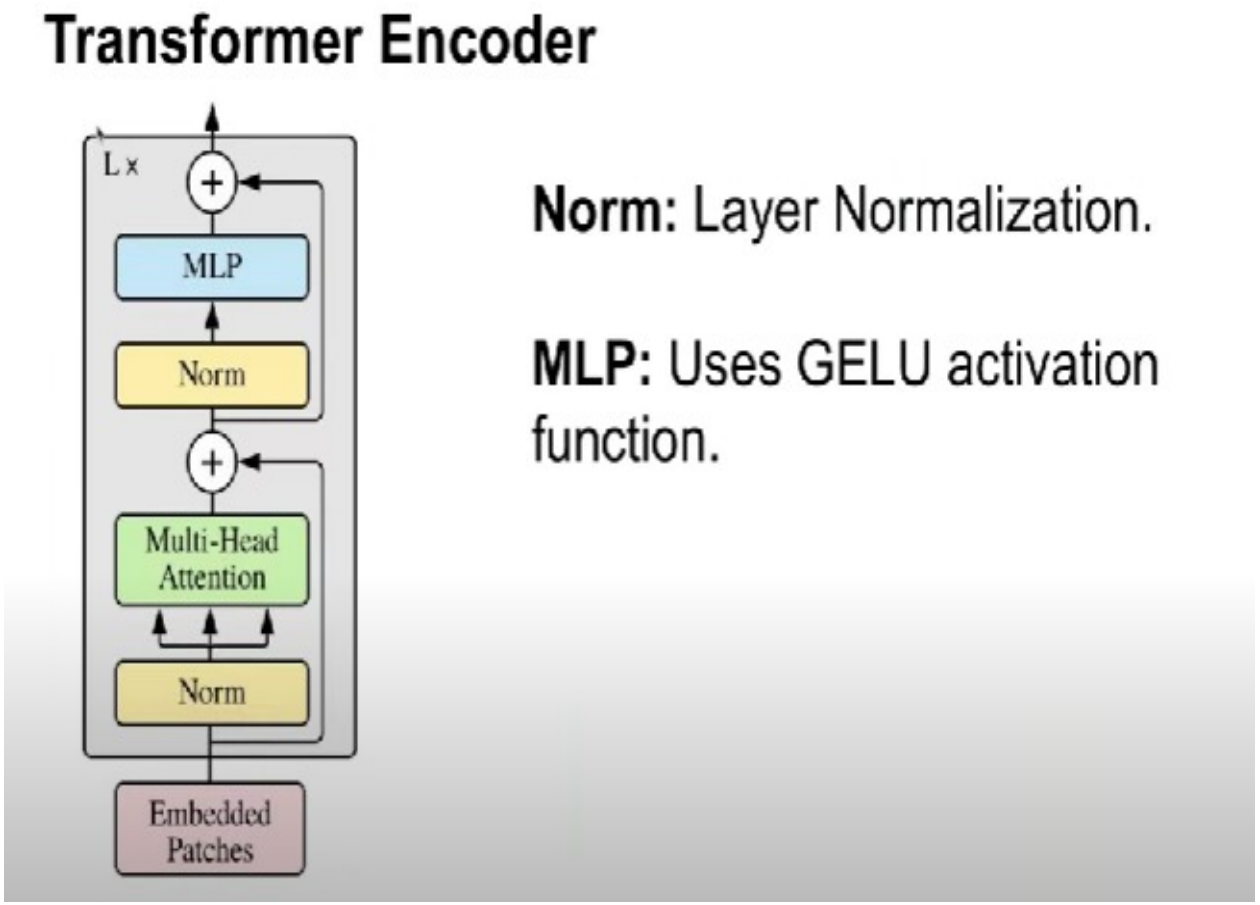


Figure: Transformer Encoder

Clasics ViTs

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

Figure: Clasic ViTs

UNIVERSIDAD NACIONAL DE COLOMBIA

Thanks!