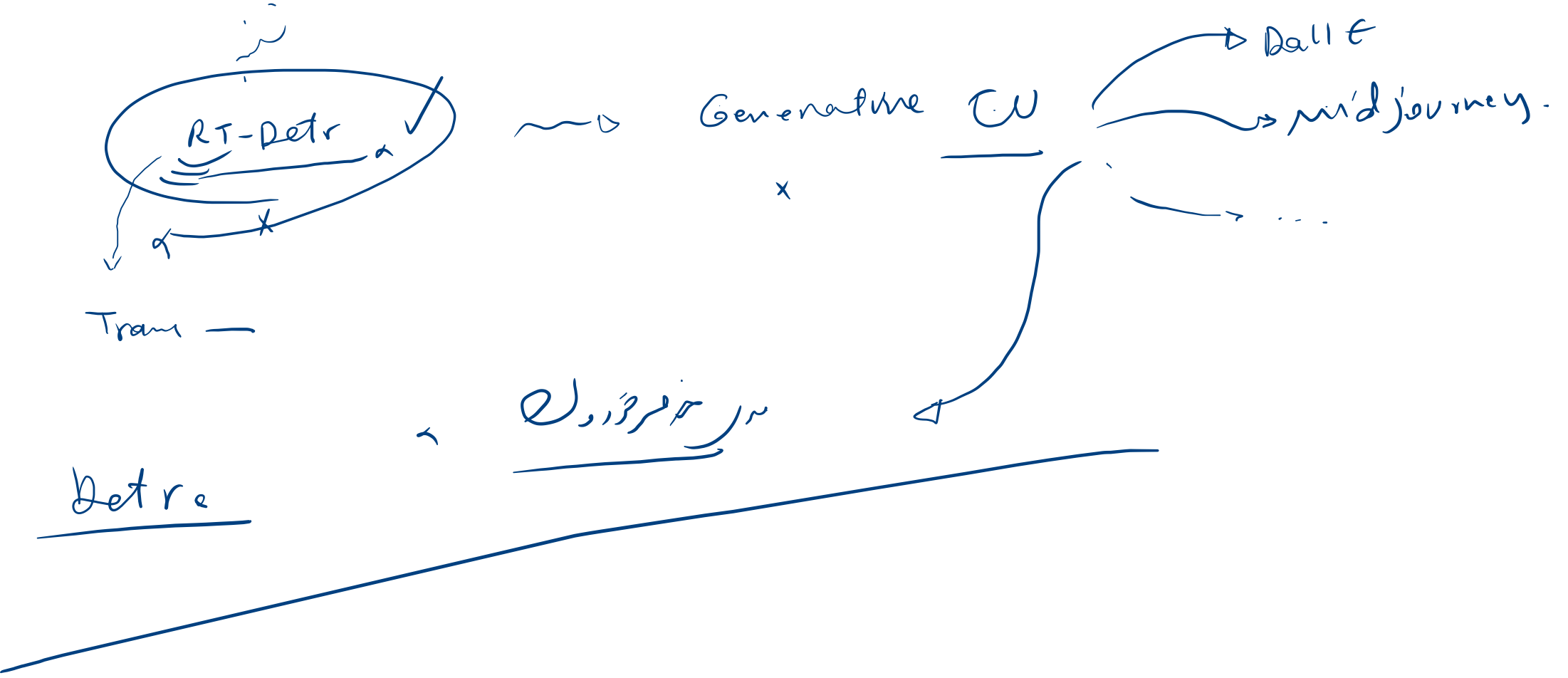
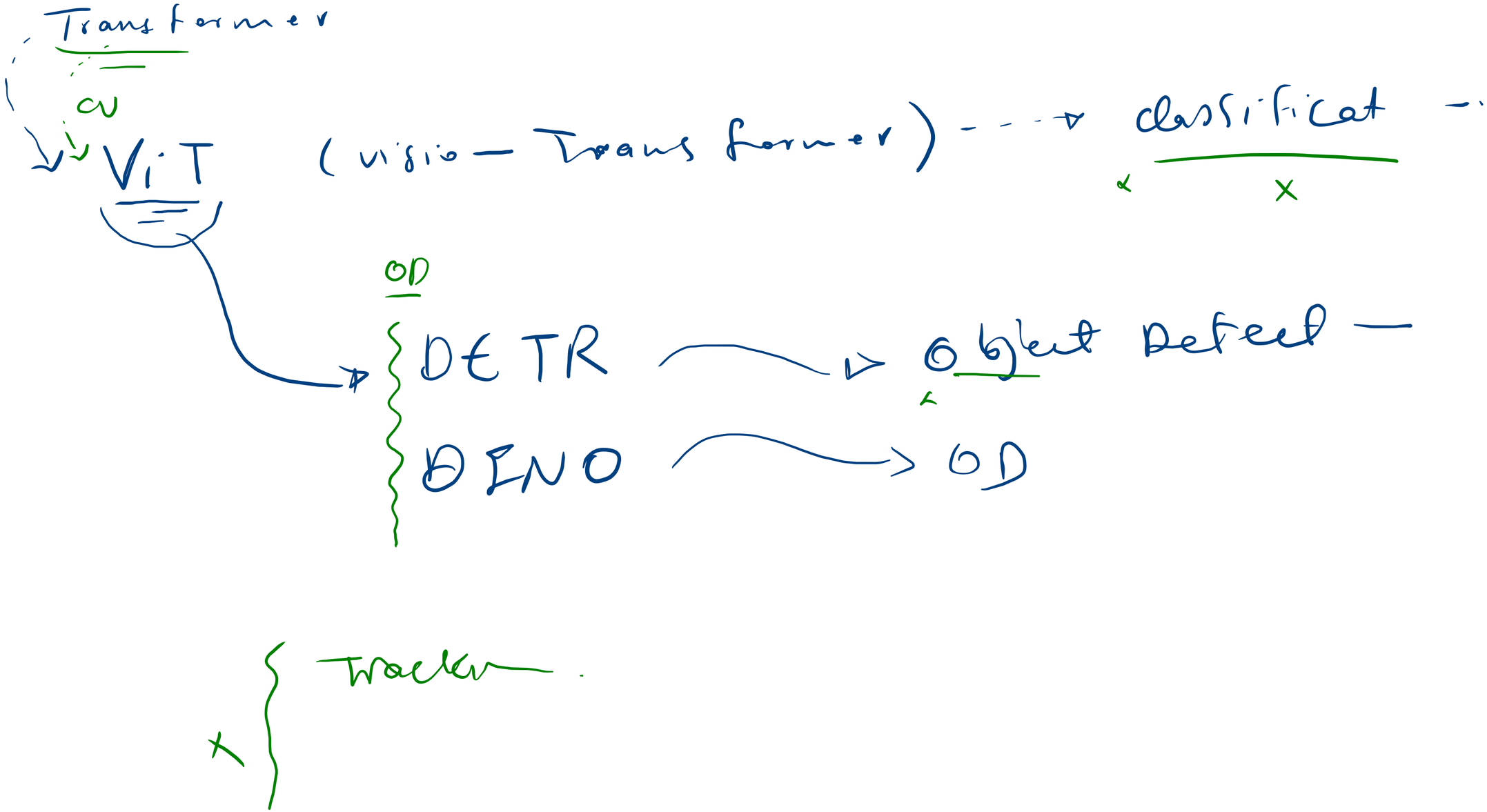

RT-Detr
Generative AI
Dall E
midjourney.
Trans
Detre

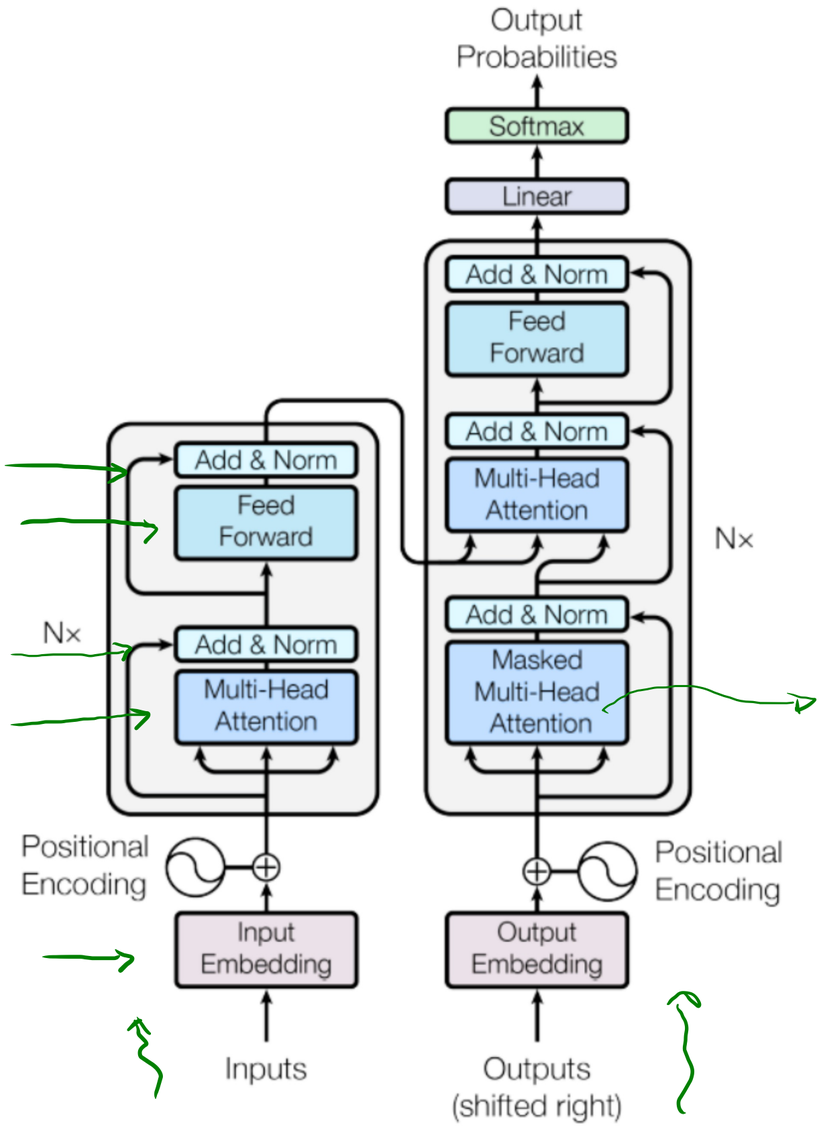

Transformer
CV
ViT
(vision — Transformer)
classificat
OD
DETR
Object Detect
DINO
OD
Tracker

شبکه ی ویژن ترنسفورمر

Vision Transformer

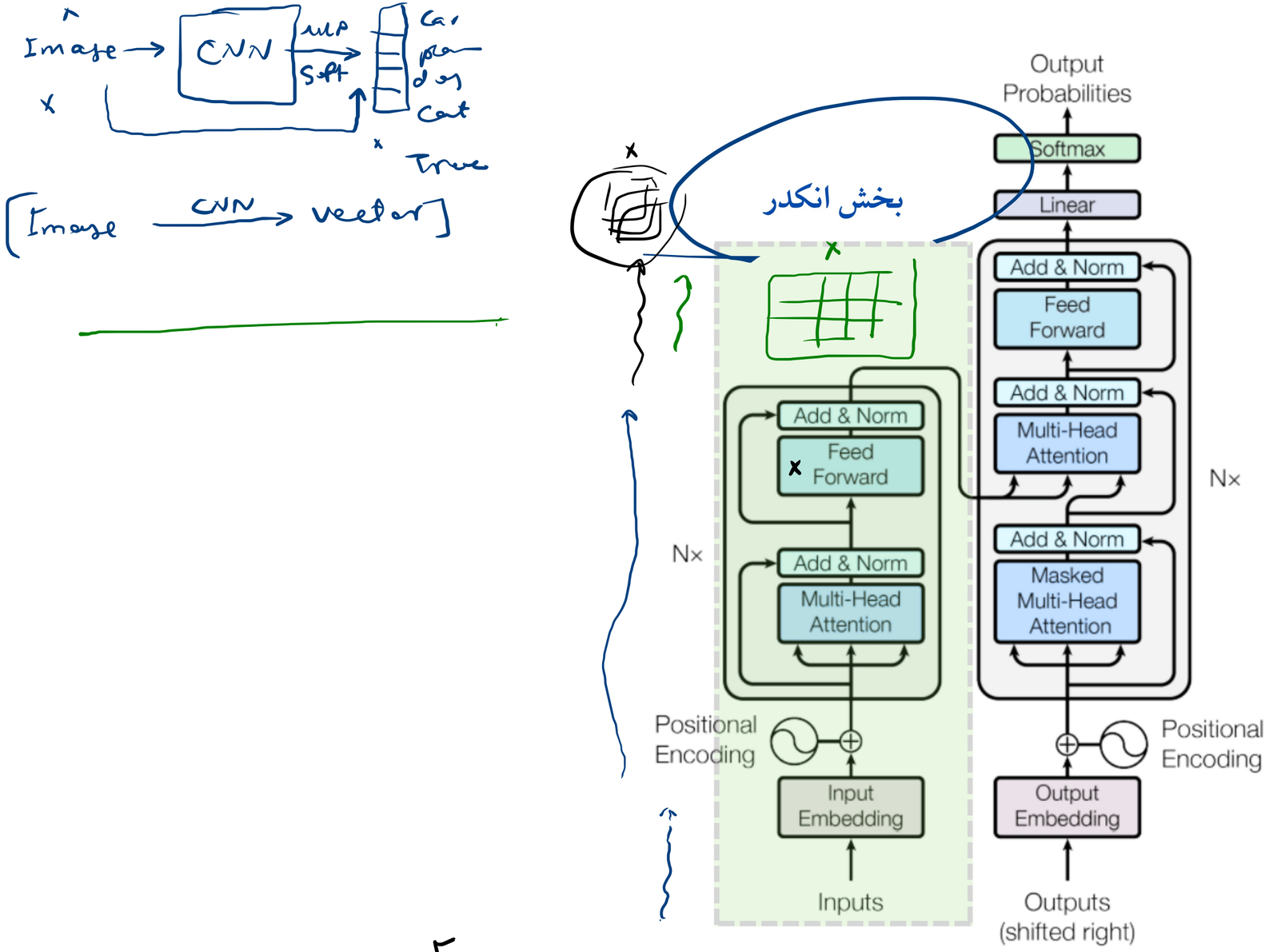
(ViT)

شبکه ترنسفورمر



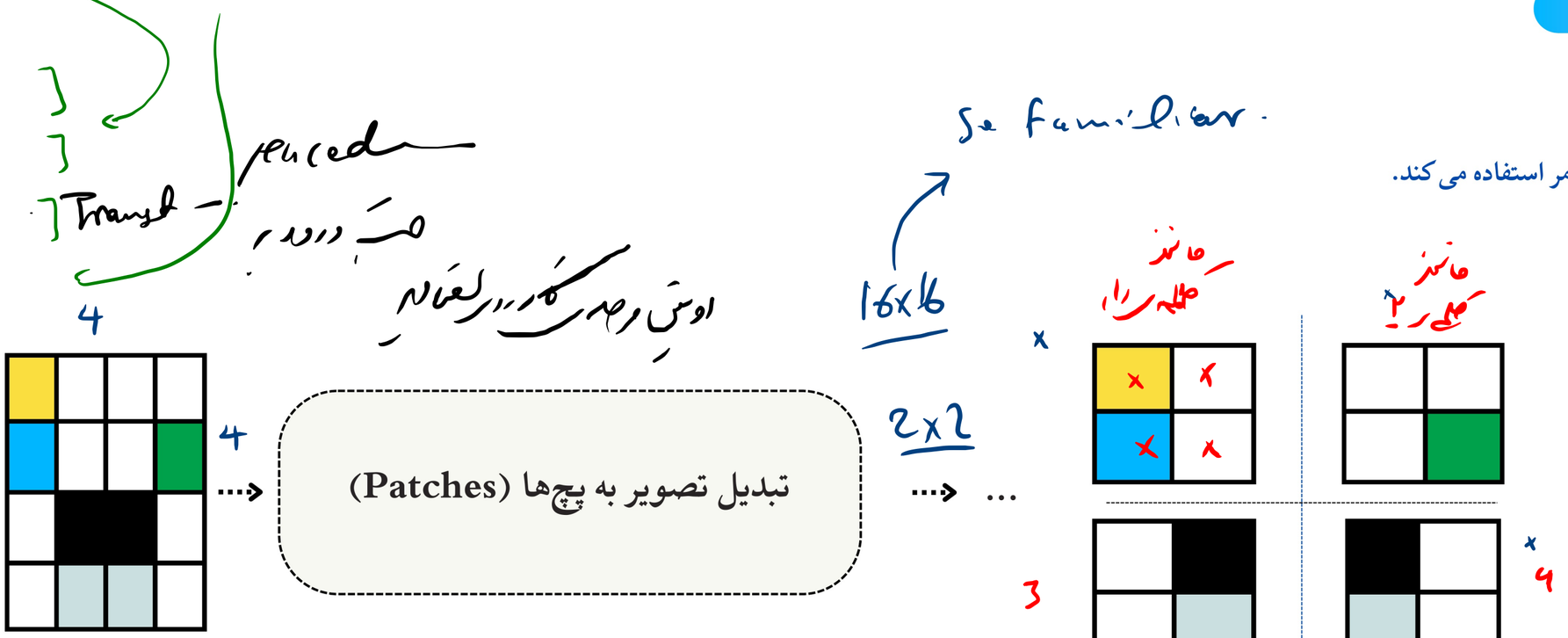
شبکه ویژن ترنسفورمر

شبکهViT از بخش انکدر ترنسفورمر استفاده می کند.



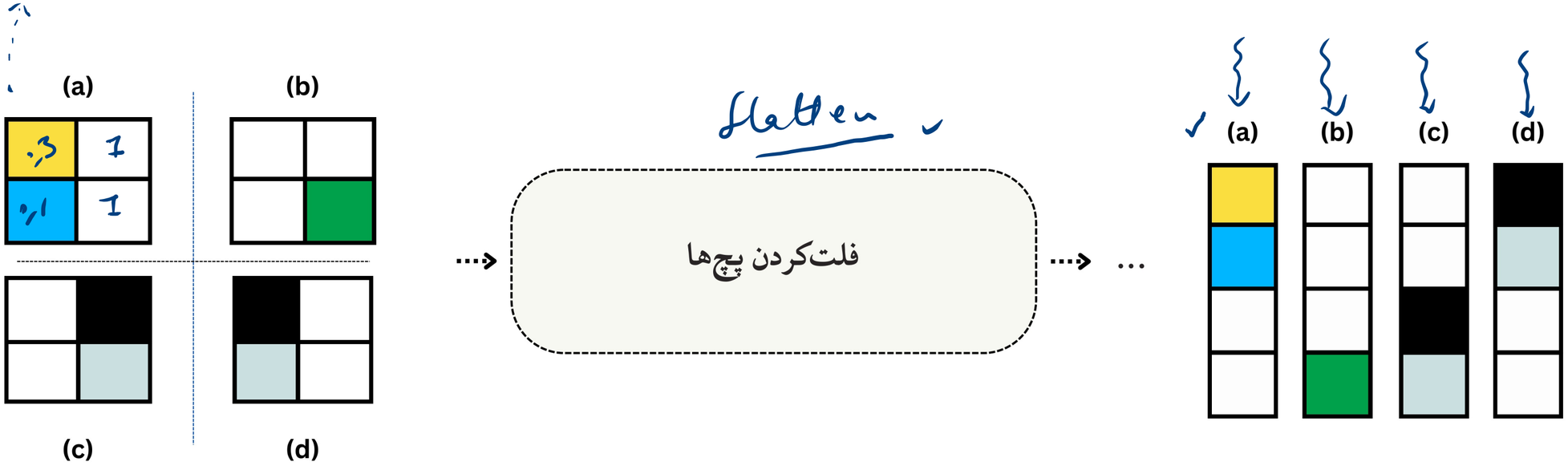
شبکه ویژن ترنسفورمر

شبکهViT از بخش انکدر ترنسفورمر استفاده می کند.



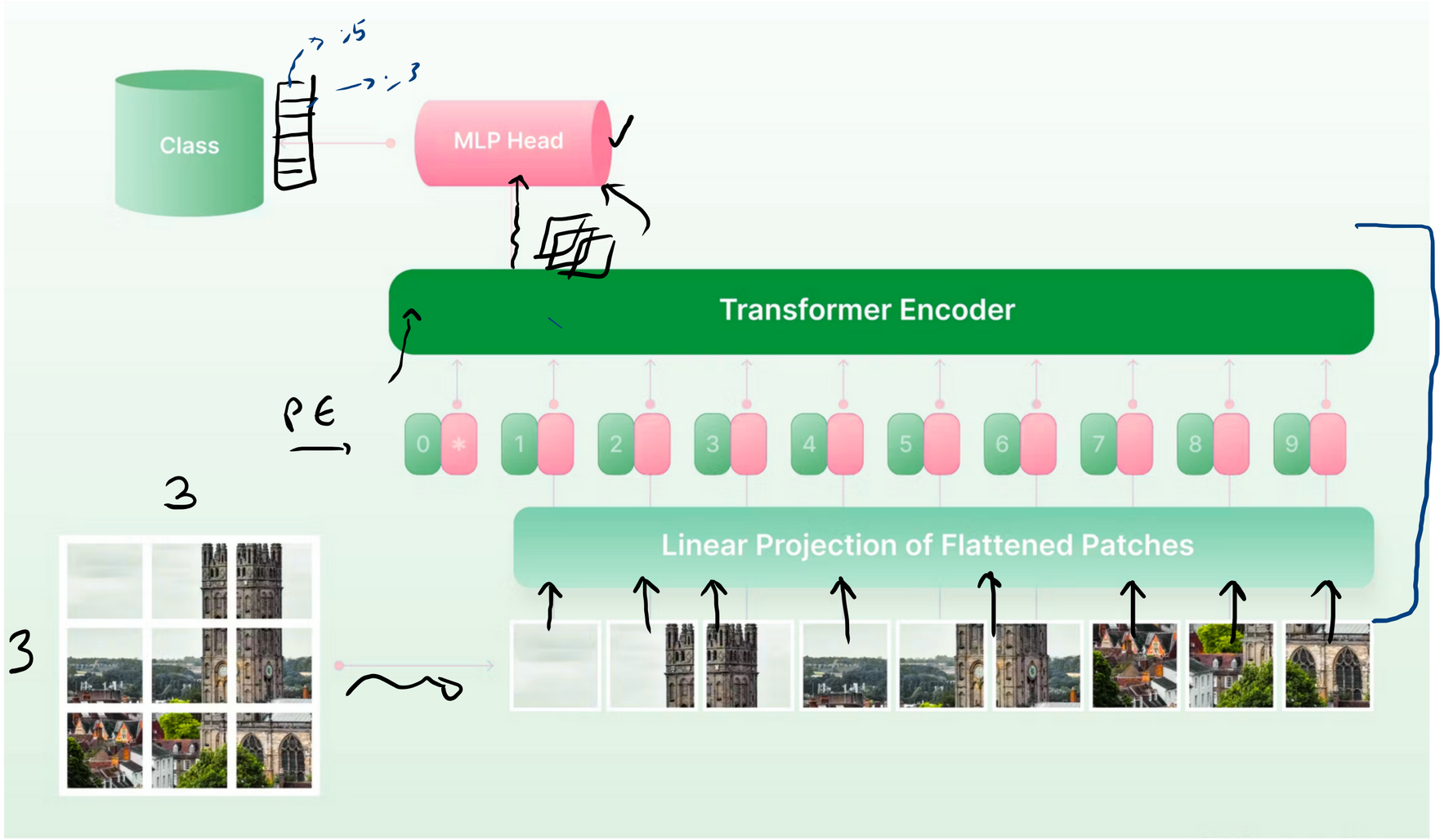
شبکه ویژن ترنسفورمر

بخش دوم : فلت کردن پچ ها



شبکه ویژن ترنسفورمر (برای کلاسیفای تصویر)

شبکهViT از بخش انکدر ترنسفورمر استفاده می کند.



شبکه ویژن ترنسفورمر برای تشخیص اشیا

OD

شبکه ViT از بخش رمزگذار ترنسفورمر استفاده می‌کند.

- DETR
- ViTDet
- DINO

RT-DETR V3

Train

CLIP

Image → Image encoder

Text → Text encoder

Image encoder

Text encoder

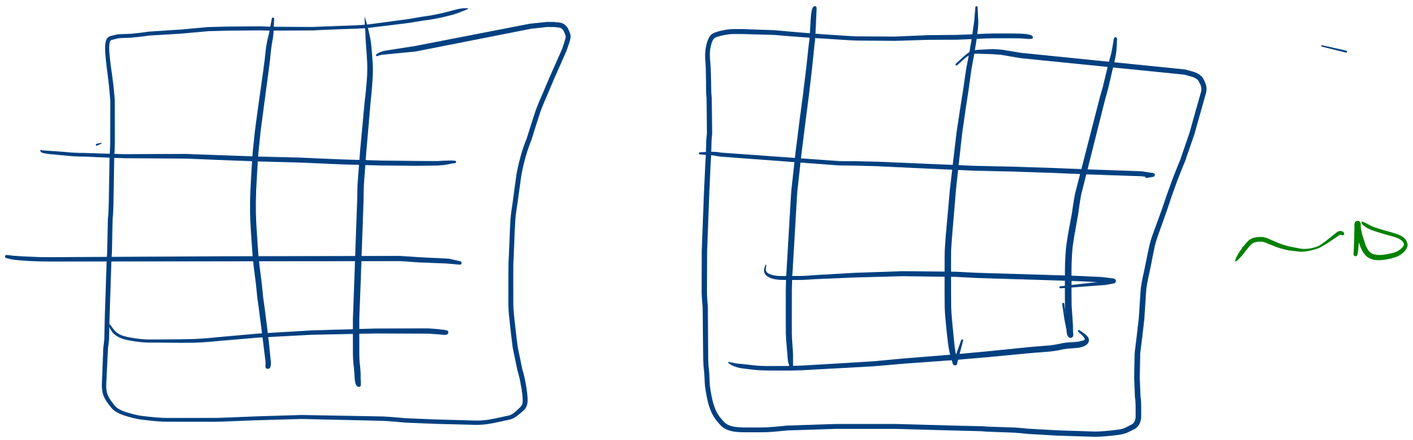


Image vector 1

Text vector 2

loss →

Similarity ✓

loss

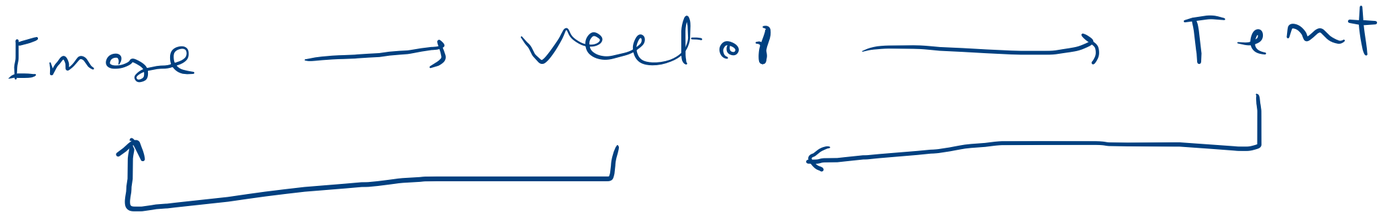


Image
vector
Text

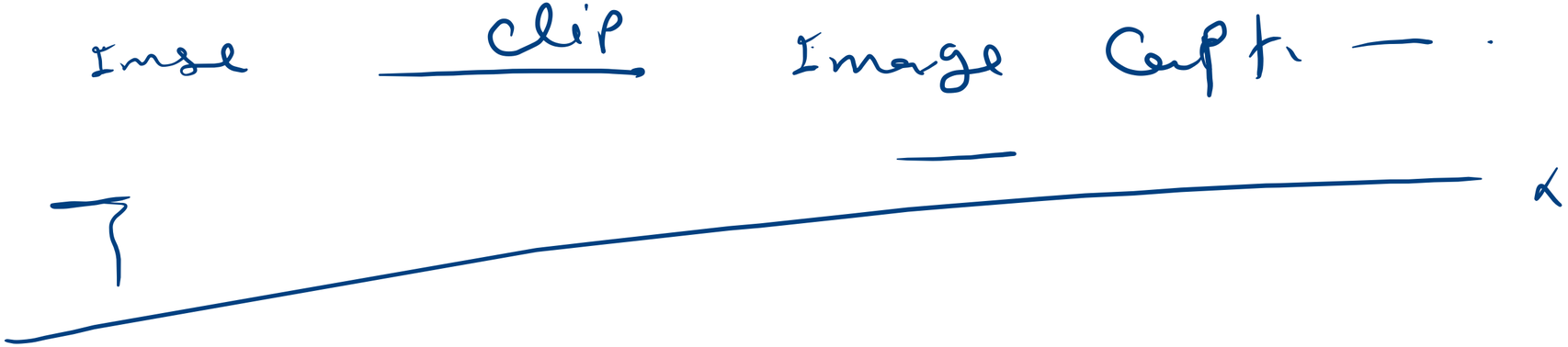
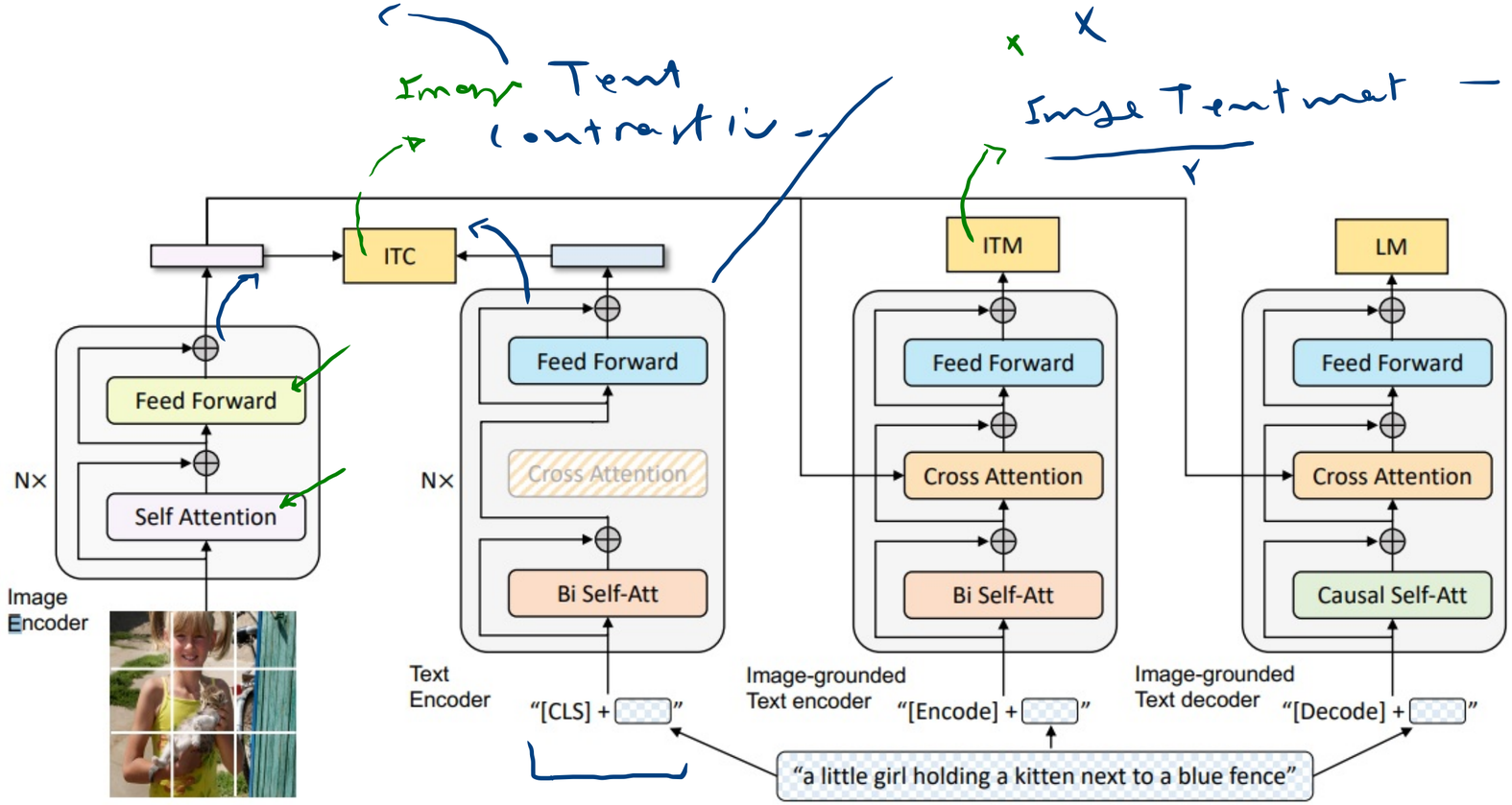


Image
clip
Image Capt.
x


ITC
ITM
LM
Feed Forward
Self Attention
N×
Cross Attention
Bi Self-Att
Causal Self-Att
Image Encoder
Text Encoder
Image-grounded Text encoder
Image-grounded Text decoder
"[CLS] +
"[Encode] +
"[Decode] +
"a little girl holding a kitten next to a blue fence"