STEM Summer Camp

06/21/2024

Hand Mesh Recovery

1) 3D hand reconstruction

2) Backbone: Large Vision Transformer(ViT) model with large training datasets

3) Other Related work: 3D hand pose estimation; MANO parametric hand model.

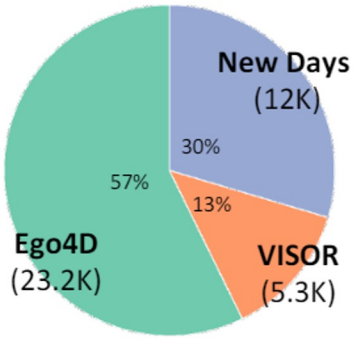
https://arxiv.org/abs/2312.05251

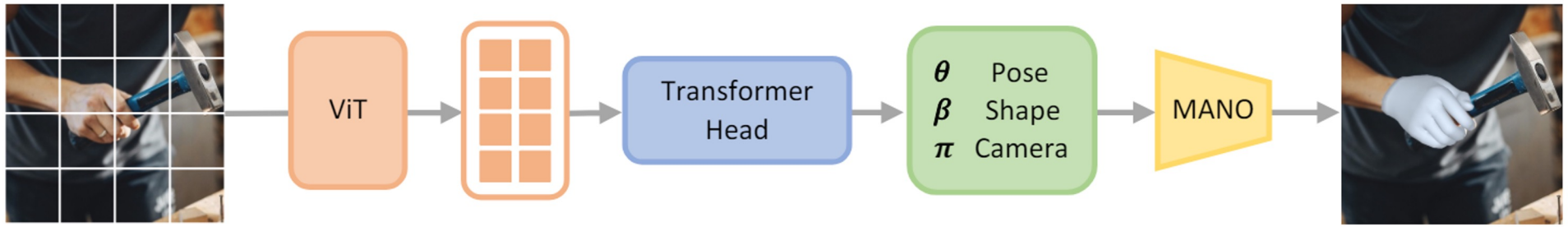
3D hand reconstruction

https://geopavlakos.github.io/hamer/

Transformer-based Architecture

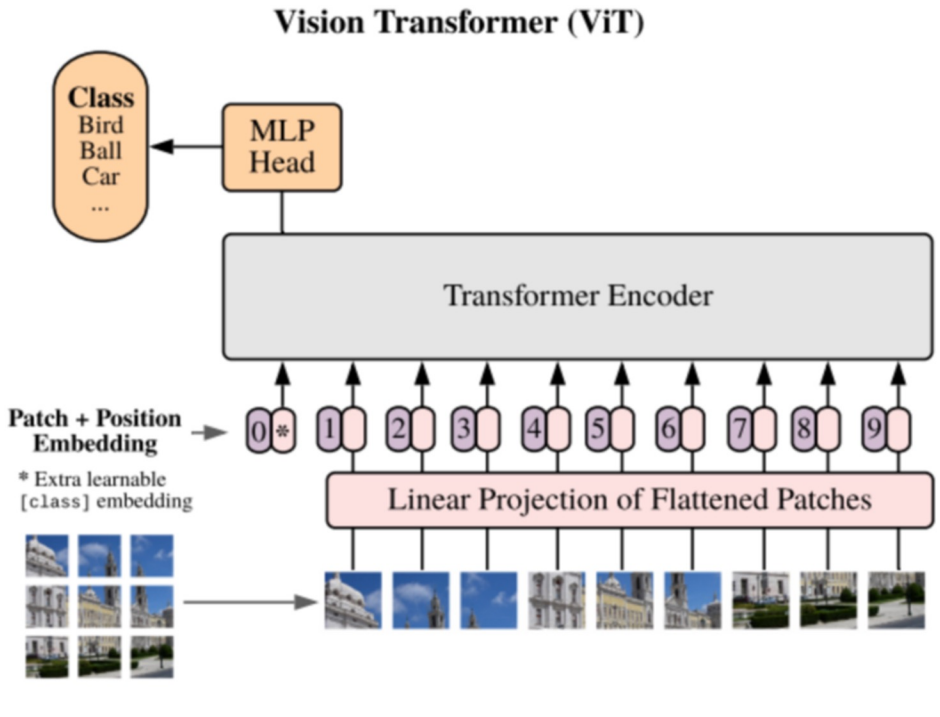
Backbone: Vision Transformer(ViT) model

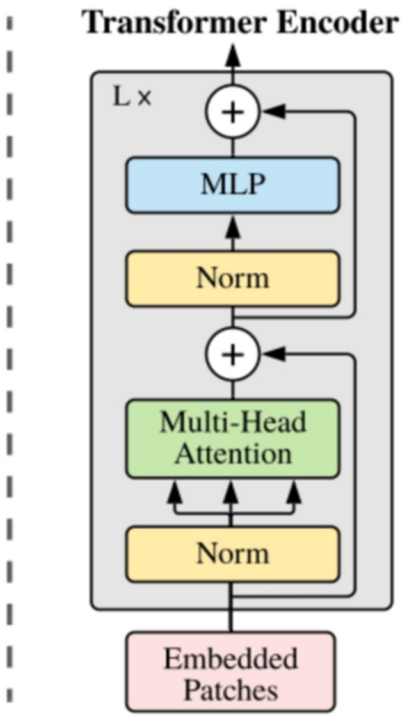




Vision Transformer

Vision Transformer

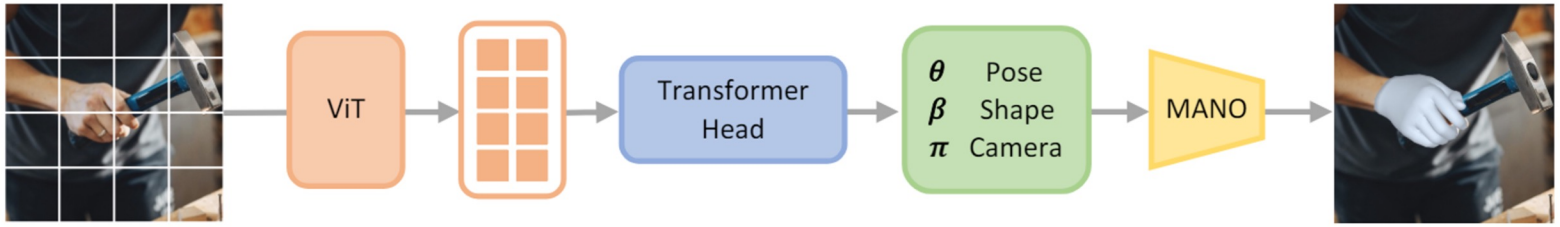




- Development of large language models
  Same loss function to update the model
- Better performance than CNN
- Pre-trained on millions of images(Pre-trained model)
- Easy to find internal patterns between data

https://arxiv.org/abs/2010.11929

Transformer-based Architecture



1) learn the mapping f from image pixels to MANO parameters and camera parameters

1) MANO takes as input the pose parameters θ and shape parameters β and defines a function M(θ,β) that returns the mesh of the hand, MANO additionally returns the joints X of the hand, for a total of K = 21 joints.