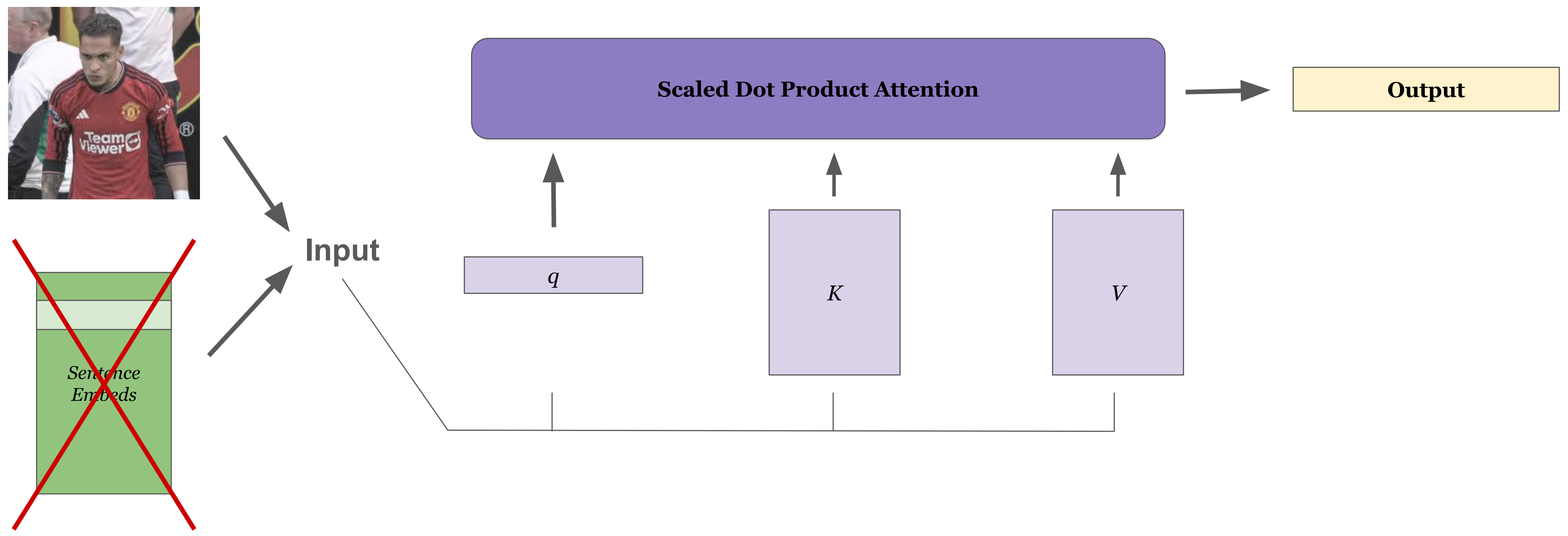
# An Image Is ~~Still~~ Worth 16×16 Words

Cody Torgovnik, Daniel Lines, Akaash Mahinth

## Motivation

Following the 2017 paper “Attention is All You Need”, the transformer architecture was at the forefront of the ML space. Researchers in CV wanted to answer the question: Can we apply a Transformer architecture to images for large-scale image recognition?



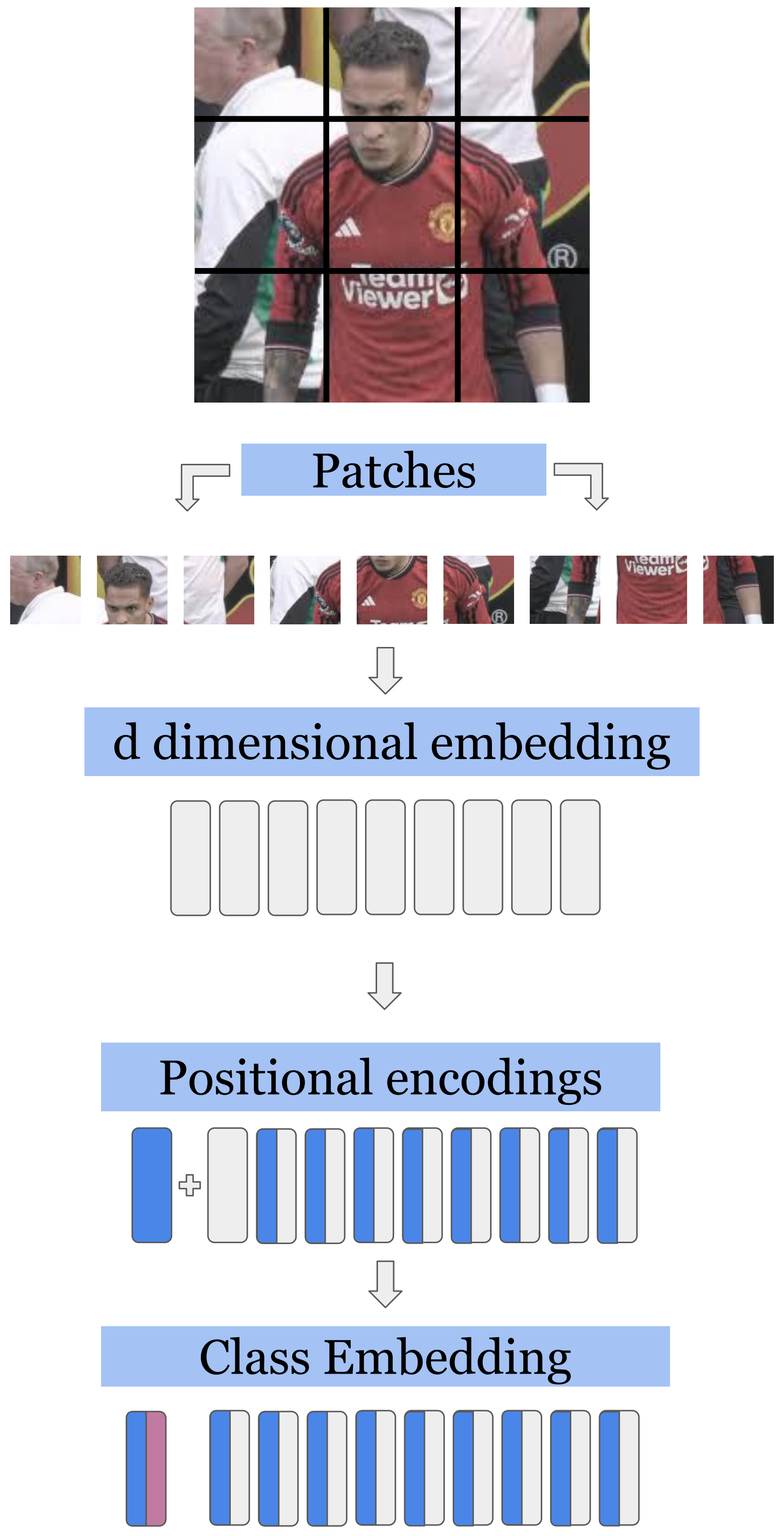
## Methodology/Goals

As pointed out in “An Image is Worth 16x16 Words”, training ViTs is very resource intensive. We used smaller scale models as well as pretrained starter models to test convolutional classifiers against attention based classifiers.

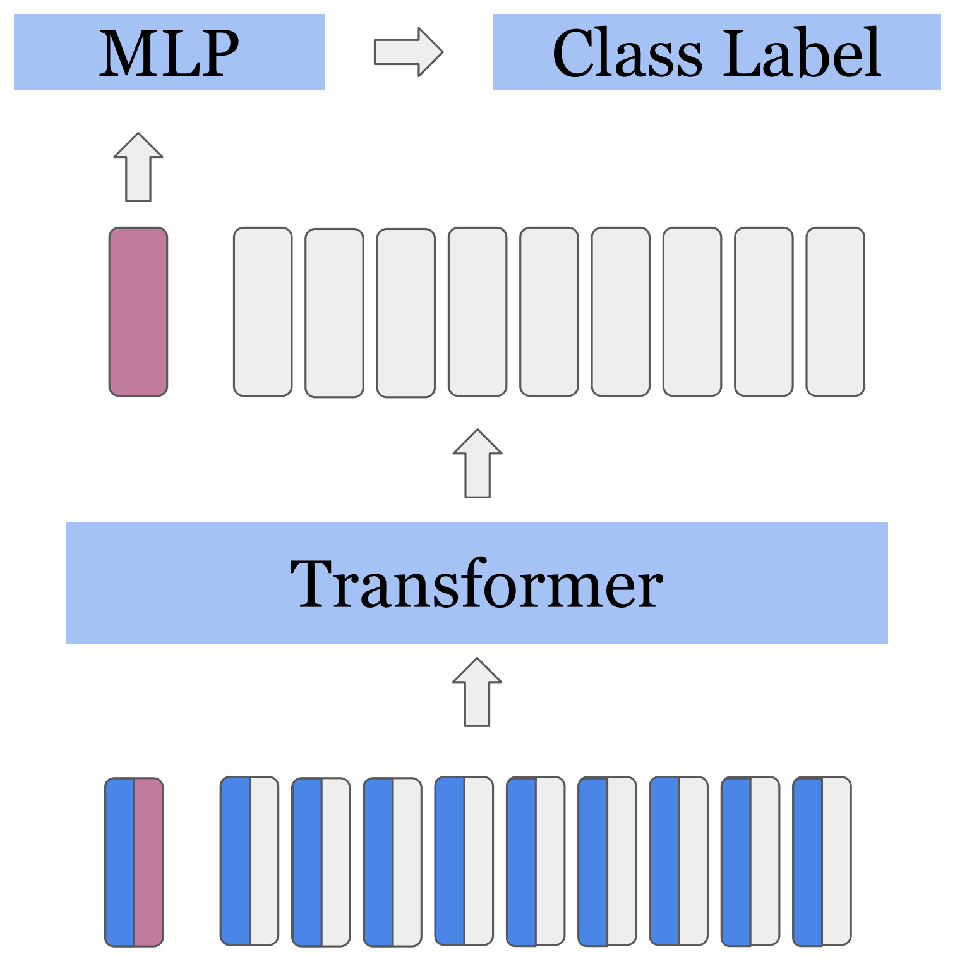
- **ViT (OC):** Our mini implementation of the ViT architecture. Pretrained on CIFAR100 and fine tuned for CIFAR10.
- **ViT_b_16:** The base model from the paper. Pretrained model from Pytorch and finetuned over CIFAR10.
- **DeiT-tiny:** A tiny transformer pulled from Pytorch. Pretrained model from Pytorch and finetuned over CIFAR10.
- **ResNet18:** A ResNet model pulled from PyTorch. We finetuned two versions of this model, one pretrained on CIFAR100, and one trained on Imagenet1k

| | Ours-JFT<br>(ViT-H/14) | Ours-JFT<br>(ViT-L/16) | Ours-I21k<br>(ViT-L/16) | BiT-L<br>(ResNet152x4) |
|---|---|---|---|---|
| CIFAR-10 | **99.50** ± 0.06 | 99.42 ± 0.03 | 99.15 ± 0.03 | 99.37 ± 0.06 |
| CIFAR-100 | **94.55** ± 0.04 | 93.90 ± 0.05 | 93.25 ± 0.05 | 93.51 ± 0.08 |

## Embeddings



## Model Architecture



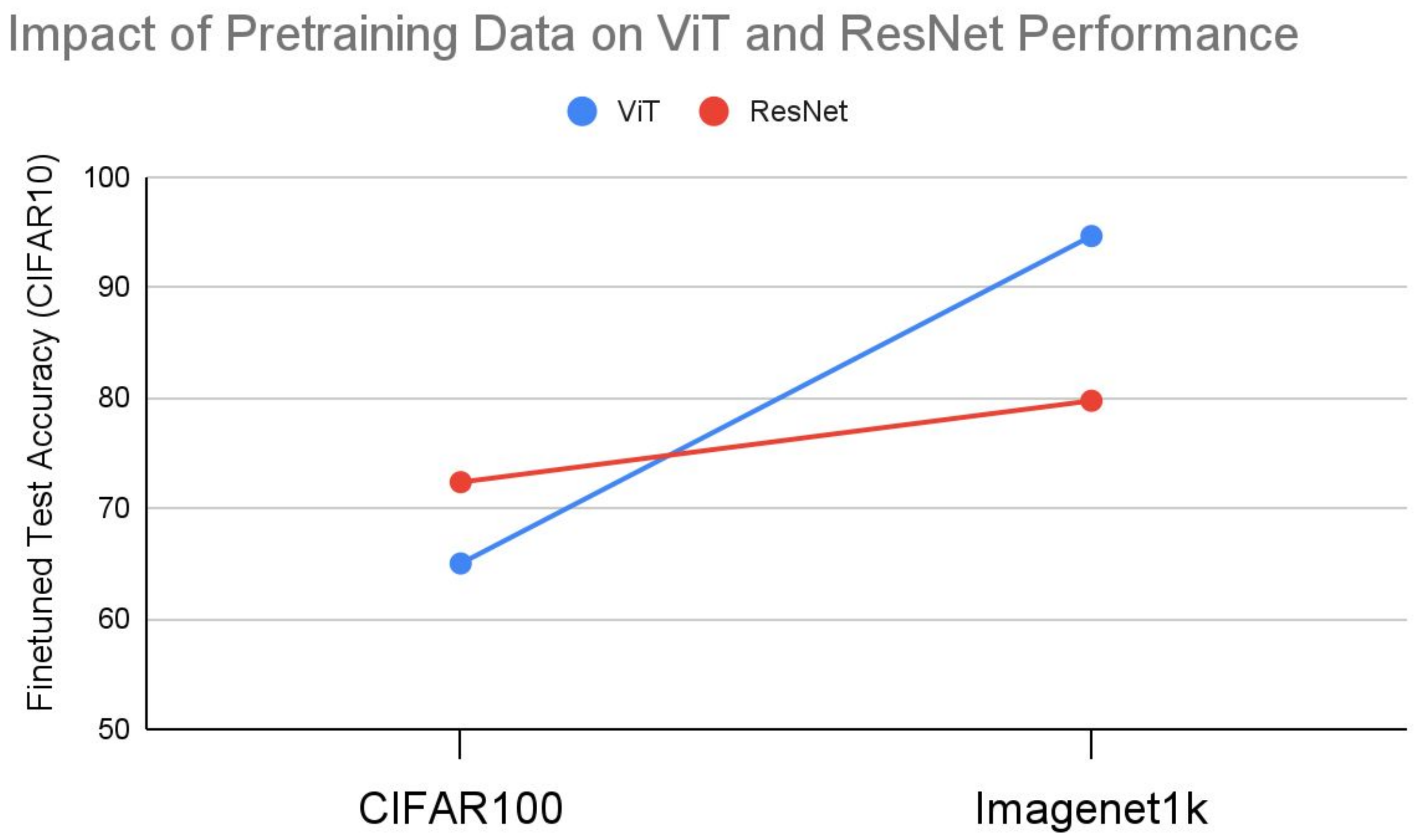
Specifications:

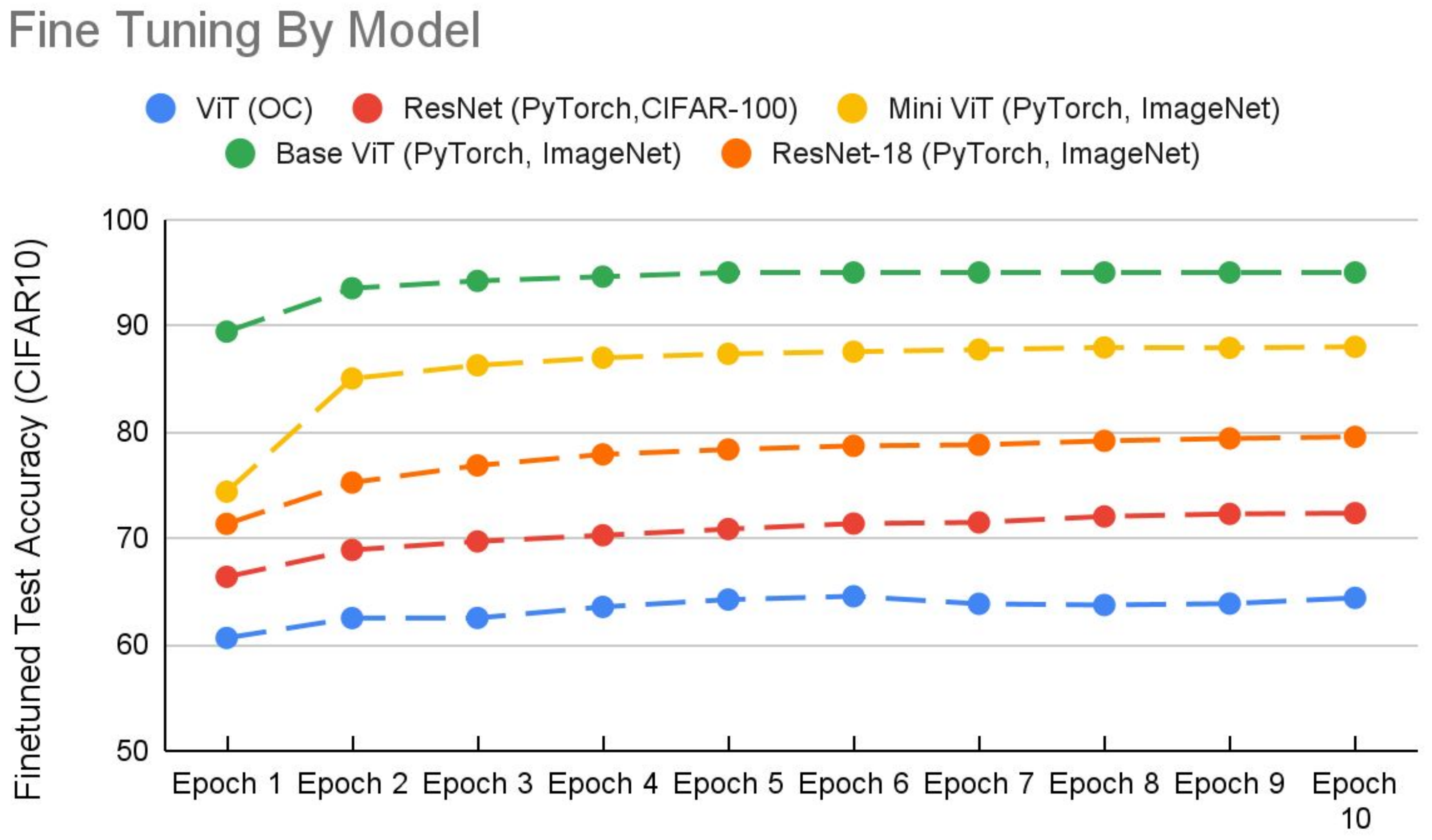
# of heads: 6

Embedding Dimension: 144

## Results

| Model | # Parameters | Pretraining Dataset | CIFAR-10 Accuracy |
|---|---|---|---|
| ViT (OC) | 1M | CIFAR-100 | 65.00% |
| ResNet-18 | 11.6M | CIFAR-100 | 72.36% |
| Mini ViT | 5M | ImageNet-1K | **87.33%** |
| Base ViT | 86.5M | ImageNet-1K | **94.64%** |
| ResNet-18 | 11.6M | ImageNet-1K | 79.73% |





## References

[1] https://doi.org/10.48550/arXiv.2010.11929
[2] https://x.com/FootballFunnnys/status/1789711042055975040
[3] https://pytorch.org/vision/main/models.html
[4] https://huggingface.co/facebook/deit-tiny-patch16-224