

Vision Transformers are NOT flexible!

ViTs offer different variants

- High hardware demand due to large attention matrices in MHA.
- ViTs provide multiple variants with different hardware demands.

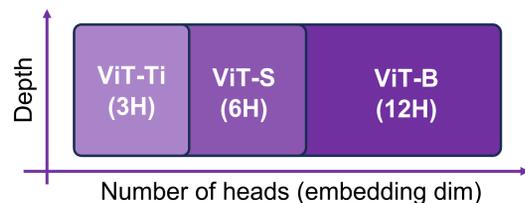
	ViT-Ti	ViT-S	ViT-B
# Layers	12	12	12
Dim	192	384	768
# Heads	3	6	12
Dim per Head	64	64	64
# Params	5.7 M	22 M	86 M

Limitations of ViT's variants

- Each must be individually trained, tuned, and stored.
- The number of configurations is limited, offering only a few options.

Observation

- The variants share the same architecture, differing only in the number of attention heads and embedding sizes, i.e., $ViT-Ti \subseteq ViT-S \subseteq ViT-B$.



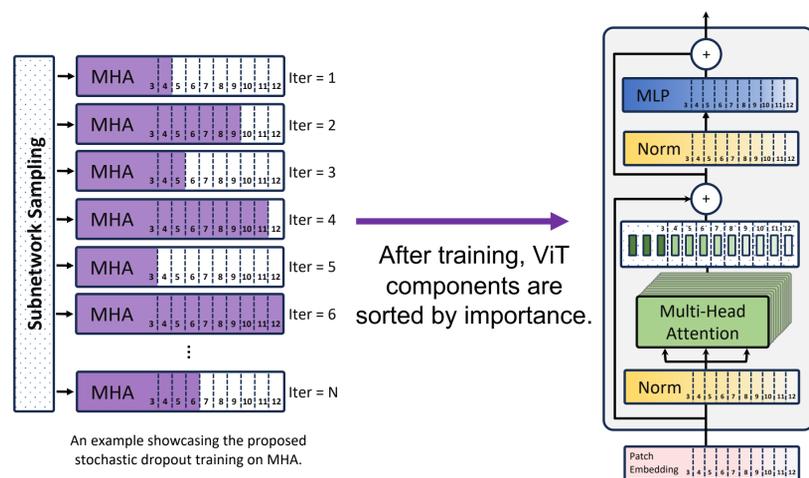
HydraViT: One model, many configurations

Train a **single universal ViT model** adaptable to diverse hardware:

- In each iteration, we train a **randomly sampled subnetwork** corresponding to the first k heads, including the associated weights in MLP, Norm, and Patch Embedding layers.

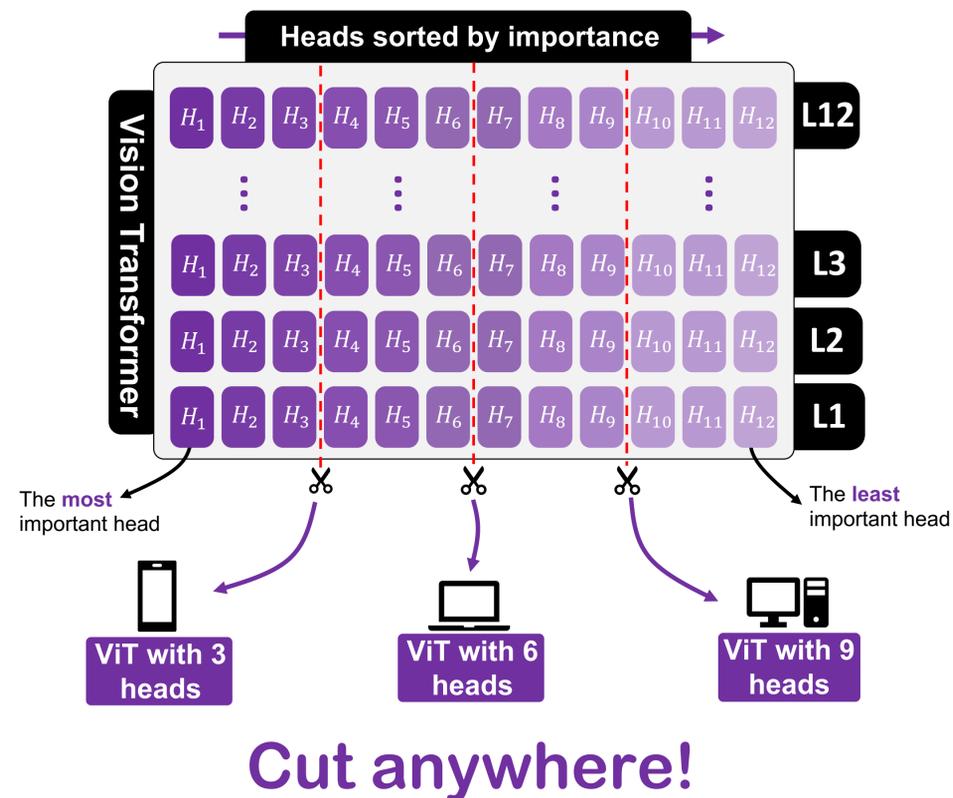
Result:

- **Initial heads** are involved more in training \rightarrow more important.
- **Later heads** are involved less in training \rightarrow less important.

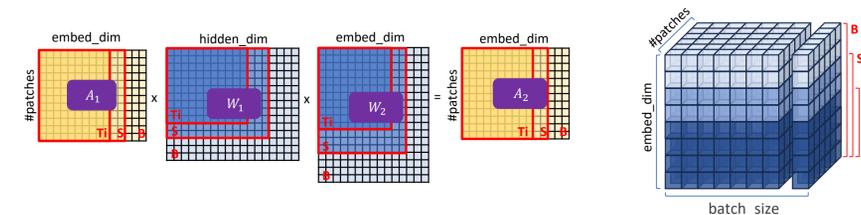


Sort Heads, Scale Anywhere!

TLDR: By sorting attention heads during training, we enable flexible inference that adapts to diverse hardware constraints by dropping the least important heads.

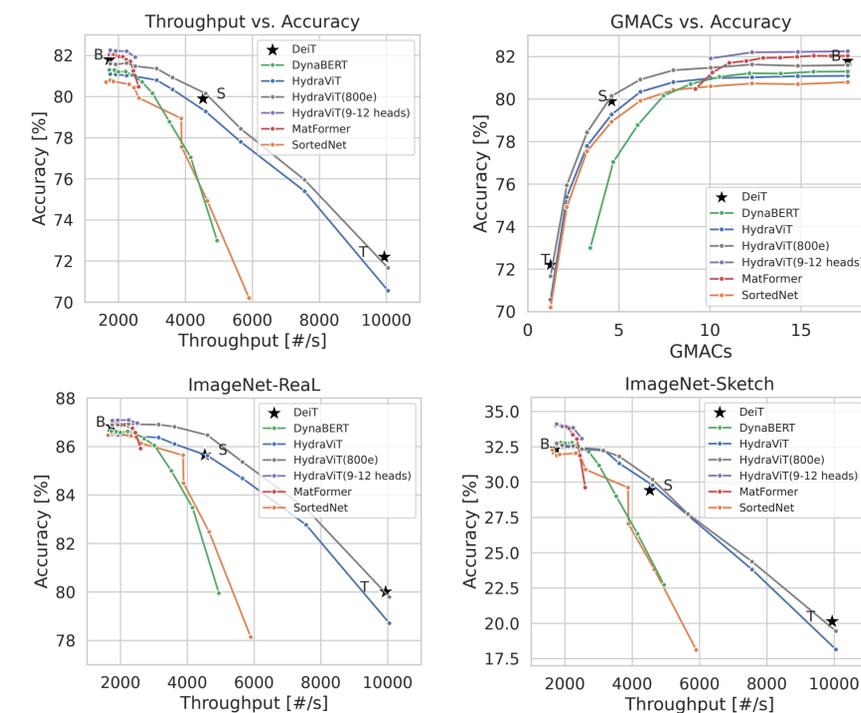


How do we extract subnetworks?



Results

- HydraViT outperforms baselines by up to 7 p.p. at the same throughput.



- The accuracy of HydraViT with our different design choices.

Weighted Sampling?	Separate Classifiers?	Epochs	Acc [%] 3 Heads	Acc [%] 6 Heads	Acc [%] 12 Heads
X	X	300	72.56	79.35	80.63
X	X	400	73.16	79.63	80.90
X	X	500	73.54	80.09	81.30
✓	X	300	72.02	79.35	80.98
✓	X	400	72.45	79.85	81.49
✓	X	500	72.50	79.89	81.63
X	✓	300	72.78	79.44	80.52
X	✓	400	73.24	79.88	81.13
X	✓	500	73.42	80.12	81.13
✓	✓	300	72.13	79.45	81.18
✓	✓	400	72.46	79.93	81.58
✓	✓	500	72.65	80.08	81.77
DeiT-tiny/small/base			72.2	79.9	81.8



GitHub



arXiv