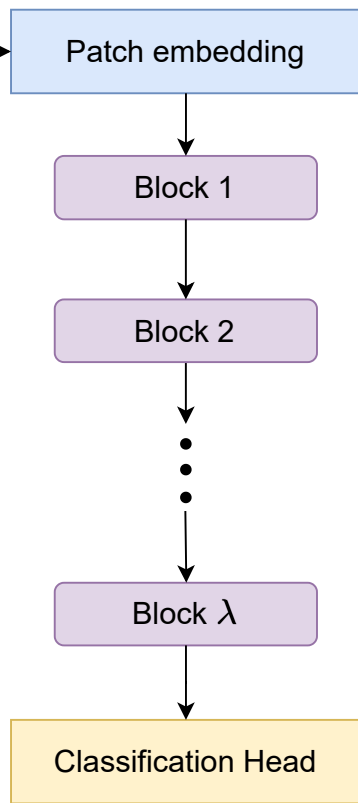
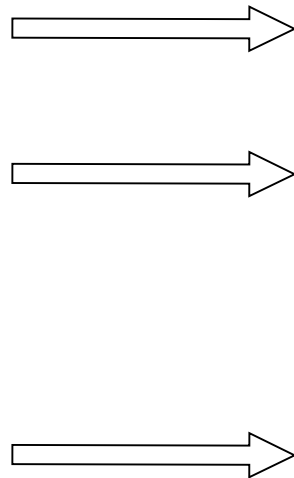


ViT



**Aggregate attention heads
with function ϕ**



Attention Matrix 1
(d, d)

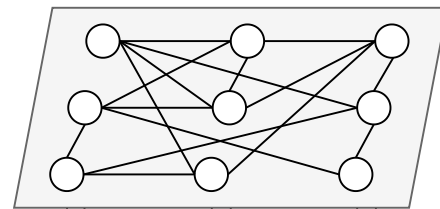
Attention Matrix 2
(d, d)

⋮

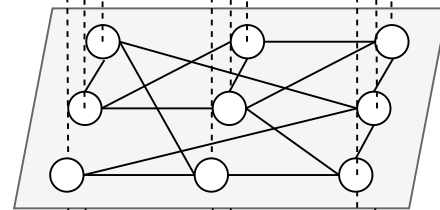
Attention Matrix λ
(d, d)

**Create the multiplex
network**

Layer 1



Layer 2



⋮

Layer λ

