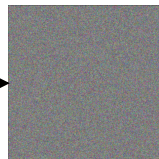
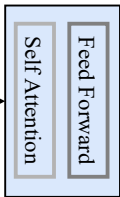


DiT Forward Pass



DiT Block with Caching

if not using cache

Feed Forward

x

+

Cached FF

if using cache

if not using cache

Self Attention

x

+

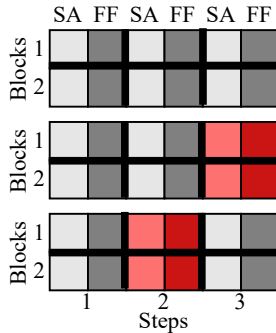
Cached SA

if using cache

x

Evolutionary Step

Generation 1

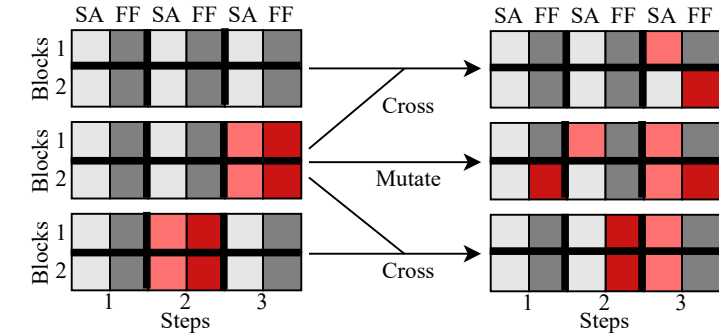


Cross

Mutate

Cross

Generation 2



Recompute Self Attention (SA)

Use Cached Self Attention (SA)

Recompute Feed Forward (FF)

Use Cached Feed Forward (FF)