

HA-NC v/s APE: varying text encoder  
( $N = 12$ ,  $M = 3$ ), best model shown: DeiT3-L/16

