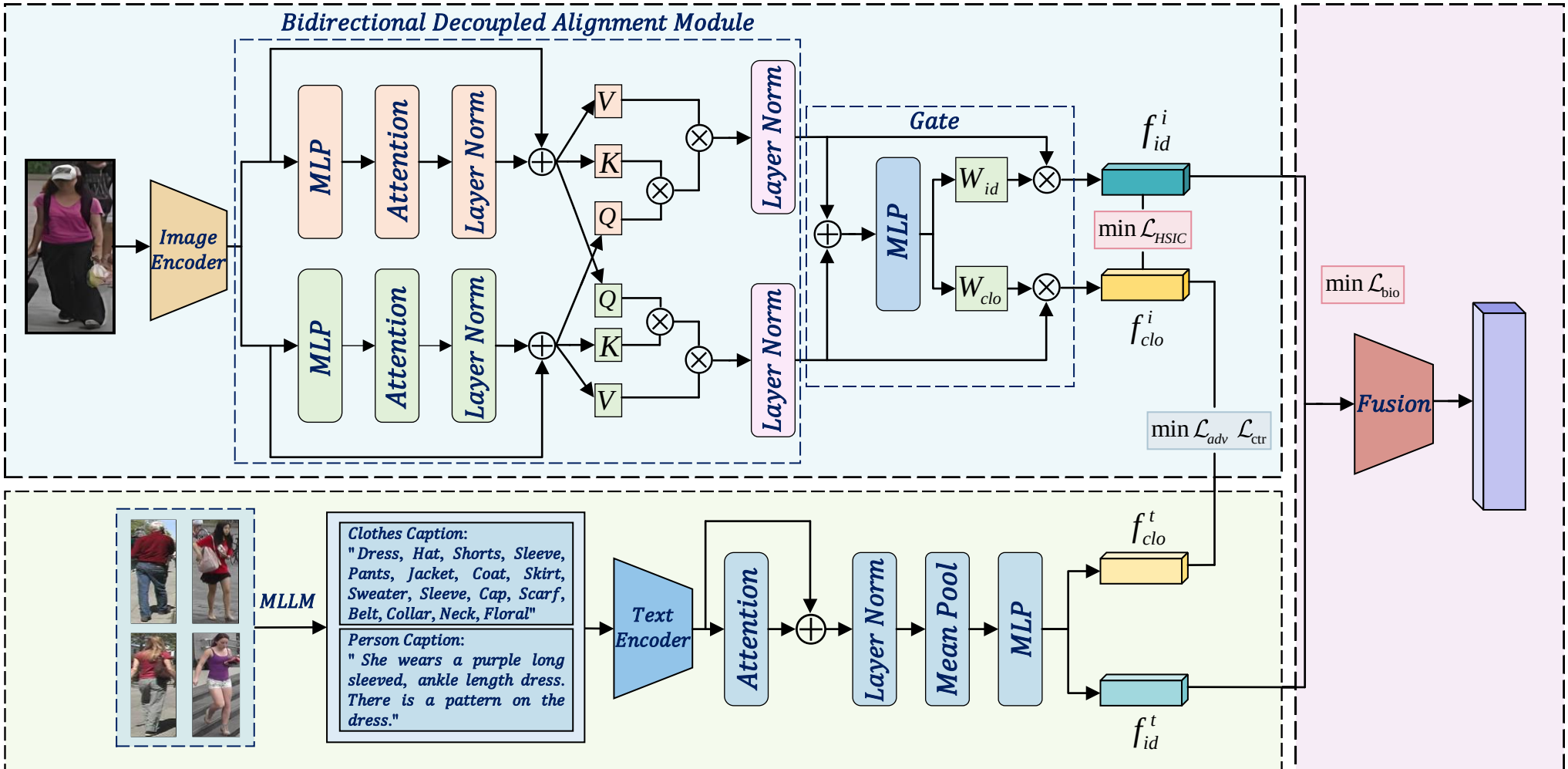
Bidirectional Decoupled Alignment Module
Image Encoder
MLP
Attention
Layer Norm
MLP
Attention
Layer Norm
V
K
Q
Q
K
V
Layer Norm
Layer Norm
Gate
MLP
W_{id}
W_{clo}
f^i_{id}
f^i_{clo}
$\min \mathcal{L}_{HSIC}$
$\min \mathcal{L}_{adv}$ $\mathcal{L}_{ctr}$
$\min \mathcal{L}_{bio}$
Fusion
MLLM
Clothes Caption:
" Dress, Hat, Shorts, Sleeve, Pants, Jacket, Coat, Skirt, Sweater, Sleeve, Cap, Scarf, Belt, Collar, Neck, Floral"
Person Caption:
" She wears a purple long sleeved, ankle length dress. There is a pattern on the dress."
Text Encoder
Attention
Layer Norm
Mean Pool
MLP
f^t_{clo}
f^t_{id}