

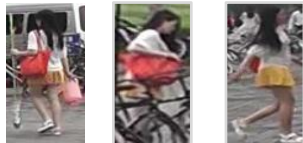
## Progressive Multi-view Generation

She **adopt it one a to** and a green broom ...with thick white shoes

⋮

She carries ... a **label the score dash will** wears white sandals with thick white shoes

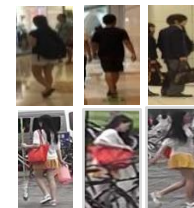
Selected reliable candidates



Randomly Word Replacing  
or  
LLM-based Paraphrasing



She carries a red tote bag and a green broom in her left hand, wearing a shirtsleeve white top over a pleated mustard-colored miniskirt. She wears white sandals with thick white shoes



Unlabeled Set



Labeled Set

TIReID Model

## Multi-view Contrastive Learning

Text Encoder

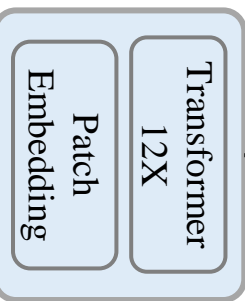
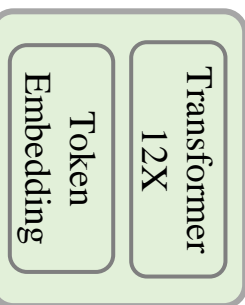


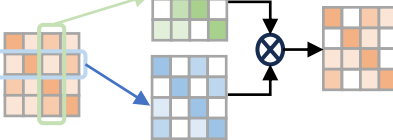
Image Encoder

$I_1^m$   
 $\vdots$   
 $I_N^m$

$T_1^1 \dots T_N^1$

$I_1^T T_1^1$	$\dots$	$I_1^T T_B^1$
$\vdots$	$\ddots$	$\vdots$
$I_N^T T_1^1$	$\dots$	$I_N^T T_N^1$

softmax



Compact Cross-modal Matching

Noisy Image-Text Pair

She **more can a any of** and...shoes.

Cross-modal Attention

She **more can any of** and...shoes.

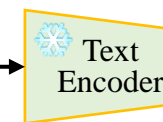
Labeled Image-Text Pair

She **carries a red tote bag** and... shoes.

Cross-modal Attention



Cross-view Discrepancy



Triplet Loss



Cross-view Discrepancy Learning



Multi-view Data