



Dual-Prompt Templates

Multimodal Large Language Model



"A person with X  
body shape..."



"Wears a red shirt,  
blue pants..."

Identity Description

Clothing Description

Original Text  
Description

CLIP  
Text Encoder



DBSCAN

