



"The image features a cityscape with a large, old, and colorful building situated on a hill. The building has a distinct red roof and a tall tower, which adds to its architectural charm. The city is surrounded by a forest, creating a picturesque and serene atmosphere.

The style of the image is black and white, which gives it a timeless and classic feel. The contrast between the black and white tones and the vibrant colors of the building and the surrounding landscape creates a visually striking composition."

### Reference Caption

Caption-01  
Average noun count: 4

Caption-02  
Average noun count: 8

Caption-03  
Average noun count: 21

Caption-04  
Average noun count: 45

Caption-05  
Average noun count: 118

Caption-06  
Average noun count: 234

### Refined Caption



### Human Instruction

"Describe the image based on the provided caption. If the caption is simple (less than 25 words), make the output as similar to the input as possible. Otherwise, if the caption is not simple (more than 25 words), slightly refine it to better align with the style in which the user writes prompts."

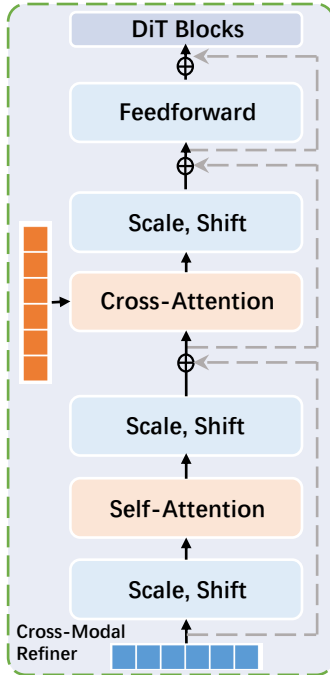


Text encoder

Large Language Model



Language Representation



Train



Inference



x N

