

(a) Generation pipeline

Generate detailed descriptions for each images involved in the following text. You can refer to the general captions and additional information for the targets

The general captions are as follows:







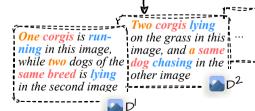
 C^2



Additional information for the targets:

Question: What are the breeds of dog in the first two images? Answer: They are corgis.

Question: For the image #3, how many cats are there and what are they doing? Answer: 3 cats are sleeping.



(b) An example