Average Peak Signal to Noise Ratio (APSNR), text Recover Rate (RR), and label Recover Accuracy (RA) are adopted to evaluate the reconstruction of images, texts, and labels respectively:

APSNR: It is used to evaluate the restoration quality of pictures. The larger the value, the higher the image quality.

text Recover Rate (RR): The proportion of recovered text to the total number of texts in the datasets.

label Recover Accuracy (RA): The proportion of recovered label from multi-model data (image, text) to the total number of datasets.