Chapter 5
EVALUATION AND COMPARISON

As shown in Table1, there is no existing work shares the completely same task settings as ours. To make fair comparison to other works and baseline, we perform both qualitative inspection and novel quantitative evaluation.

Frame quality assessment. Quantitative evaluation of generative models is a difficult, unsolved problem. The video imagination task is a multi-modality problem. But traditional full reference image quality assessment methods (FIQA) requires a precise ground truth image as reference hence they are no longer appropriate. We employ popular Blind Image Quality Assessment (BIQA) method BRISQUE as our non-reference quantitative evaluation metric. Since BRISQUE is based on natural scene statistic, it is not applicable in synthetic image. We implement it on those methods that can synthesize natural scene images in UCF101 dataset. A key problem of employing this metric is that the scenes and resolutions of the synthesized videos may be varied, so it is unfair to make comparison among those samples directly. Fortunately, the quality of the input image can be a solid quality reference. We calculate the decreasing proportion of quality score between inputs and outputs, and take it as our assessment metric: Relative image quality assessment (RIQA).

$$\text{RIQA} = \text{BRISQUE}(\text{Input}) - \text{BRISQUE}(\text{Output}) \,/\, \text{BRISQUE}(\text{Input}) \quad (6)$$

It is fair and reasonable because RIQA eliminates the natural quality differences between scenes and resolutions while have the ability of reflecting the crucial reconstruction quality well.