**[Video Generation paper review]**

1. make-a-video: text-to-video generation without text-video data

[abstract]

* learn what the world looks like and how it is described from paired text-image data
* learn how the world moves from unsupervised video footage

[why is it hard?]

* lack of large scale datasets with high-quality text-video pairs
* complexity of modeling higher-dimensional video data
* natural to consider leveraging image priors for videos to simplifying the learning process

[Why Make-A-Video]

* breaks the dependency on text-video pairs
* fine tune the T2I model for video generation
* pseudo-3D convolution & temporal attention layers

[Method]