SSA-GAN: End-to-End Time-Lapse Generation with Spatial Self-Attention
Daichi Horita, Keiji Yanai
The University of Electro Communications, Tokyo, Japan

1. Introduction
- We usually predict how objects will move in the near future in our daily lives. However, how do we predict? To address this problem, we propose a GAN-based network to predict the near future for fluid object domains.
- Our model takes one frame and is able to predict future frames.
- We propose introducing the spatial self-attention mechanism into the model.

2. Related Works
Video prediction
- Since VGAN[1] generates the background and foreground of the image separately, the background is fixed.
- MDGAN[2] generate rough movements in the first stage and add detailed appearances and motions in the second stage.

3. Method - Spatial Self Attention Module
- Propose a spatial self-attention module to learn the long-range dependence within a frame.
- The network assigns more weight to areas outside the neighborhood.
- γ play the important role to avoid the over-weighting(See 5.2).

4. Method - Spatial Self-Attention GAN
- (Blue) 3D convolutional/transposed convolutional layers.
- (Orange) Spatial Self Attention Modules.

5. Experiments
1. Quantitative evaluation

6. Future Works
- Use more temporal features to generate more realistic frames.
- Experiment with mode frames.