

SAVGAN: SELF-ATTENTION BASED GENERATION OF TUMOUR ON CHIP VIDEOS

ISBI 2022, Kolkata, India

Sandeep Manandhar, Irina Veith,
Maria Carla Parrini and Auguste Genovesio



Introduction

Generation of videomicroscopy sequences will become increasingly important in order to train and evaluate dynamic image analysis methods.

Generative models have allowed us to create low dimensional embedding (latent space) of a high dimensional data.

Exploring and manipulating the so called latent space have also lead to powerful image processing techniques like unpaired image-image translation, text to image translation.

However, video generation is a bigger and denser problem than the popular methods like image or text generation.

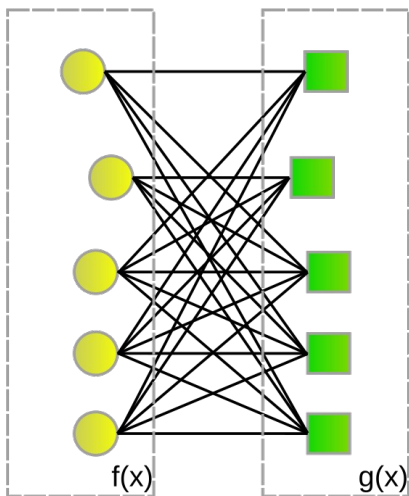
Spatio-temporal Self-Attention

Understanding of long range interactions between spatio-temporal points in a feature map is vital in order to generate realistic looking videos. Self-attention takes those relationships into account while computing intermediate features.

Nonetheless, computation of self-attention is computationally heavy. Therefore, we rely on approximation of the self-attention via Nystrom's method.

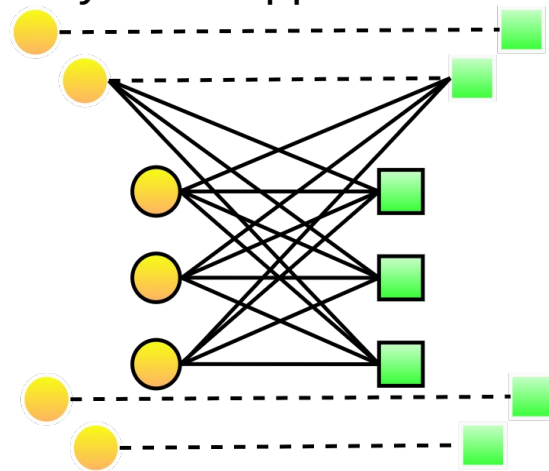
Saptio-temporal Self-Attention

Full Self- Attention



$$\mathbf{S} = \sigma\left(\frac{f(x)^T(g(x))}{\sqrt{C}}\right)$$

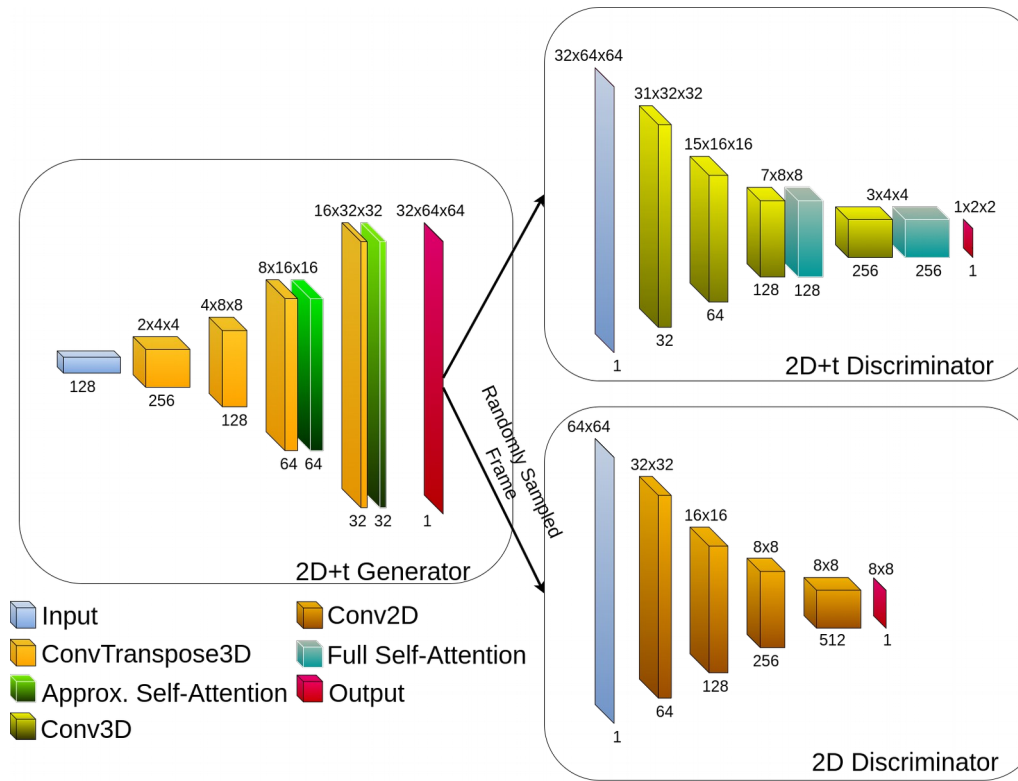
Nystrom Approximation



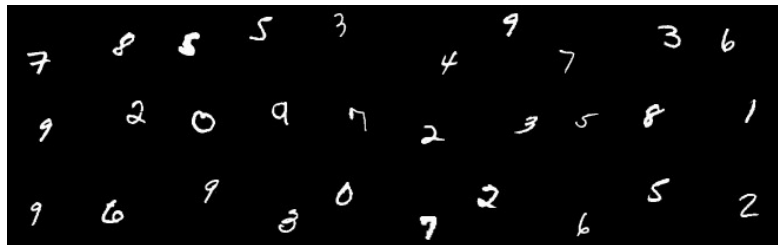
$$\tilde{\mathbf{S}} = \sigma\left(\frac{f(x)^T(g(\tilde{x}))}{\sqrt{C}}\right) \sigma\left(\frac{f(\tilde{x})^T(g(\tilde{x}))}{\sqrt{C}}\right)^+ \sigma\left(\frac{f(\tilde{x})^T(g(x))}{\sqrt{C}}\right)$$

Xiong et al, AAAI, 2021

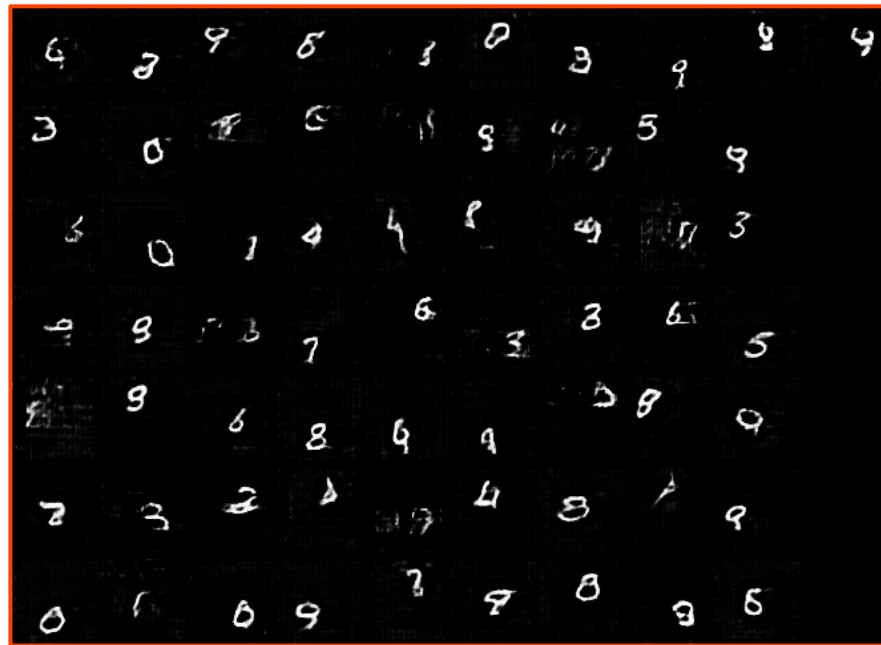
SAVGAN: Architecture



Results

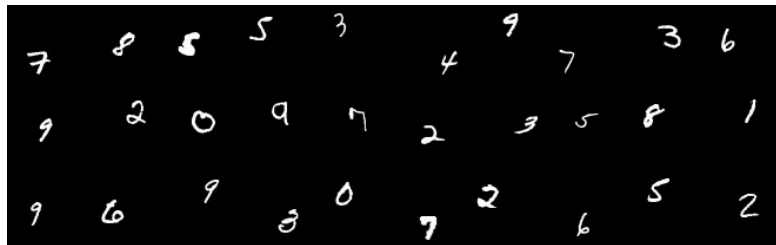


Moving MNIST (single digit)

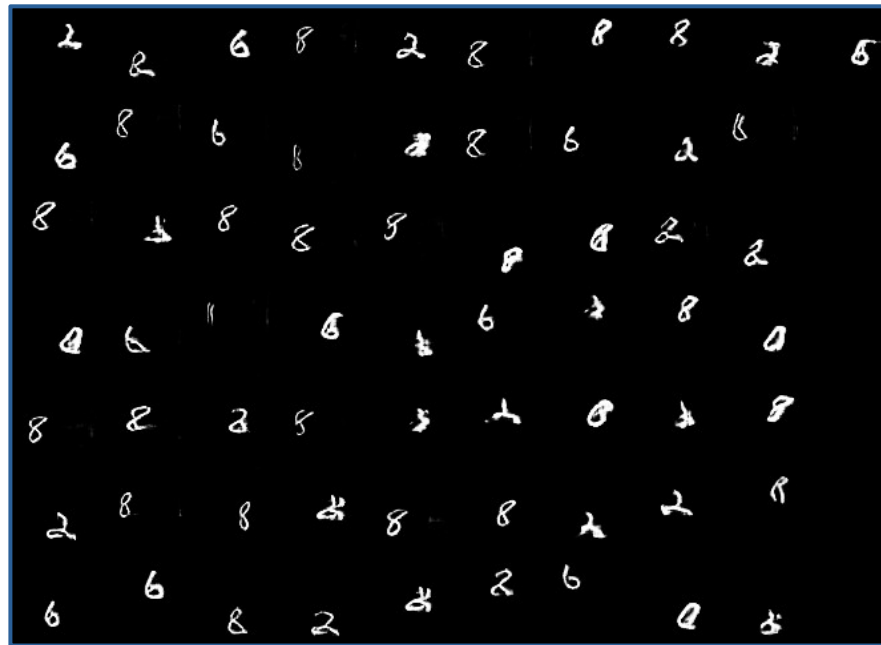


Generated with 3DGAN (no self-attention)

Results

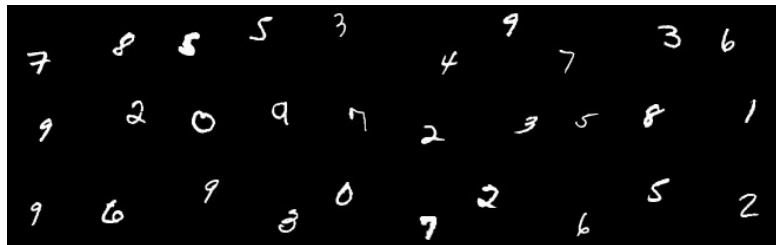


Moving MNIST (single digit)

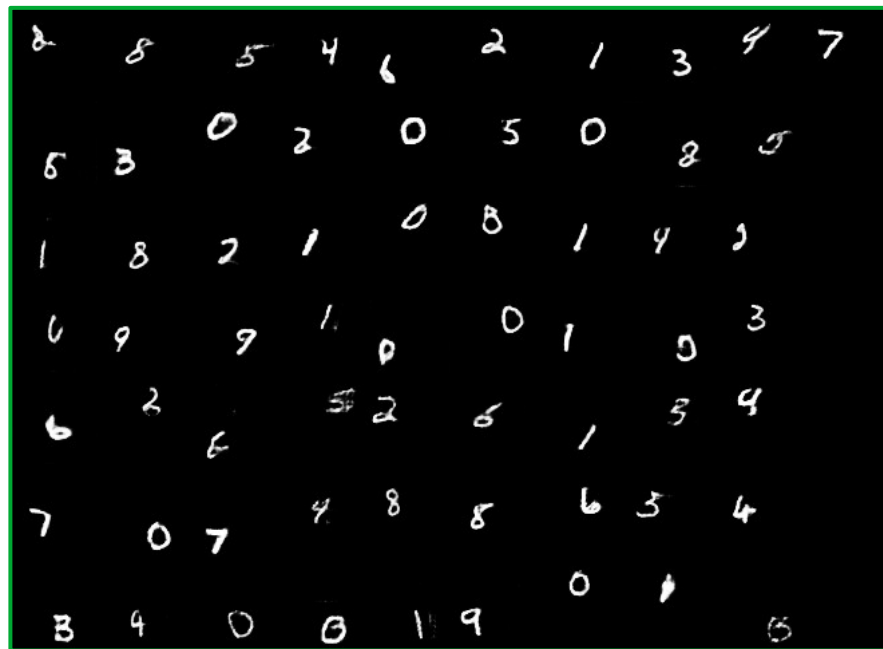


Generated with TGANv2 [Saito et al., IJCV 2020](#)

Results

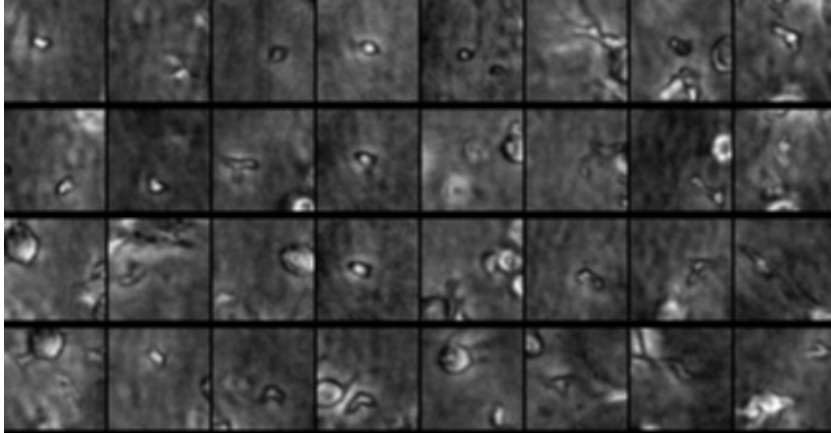


Moving MNIST (single digit)



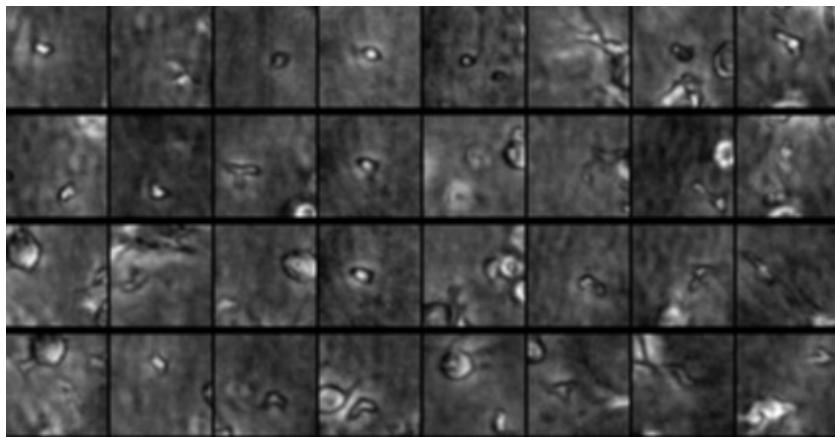
Generated with SAVGAN (ours)

Results

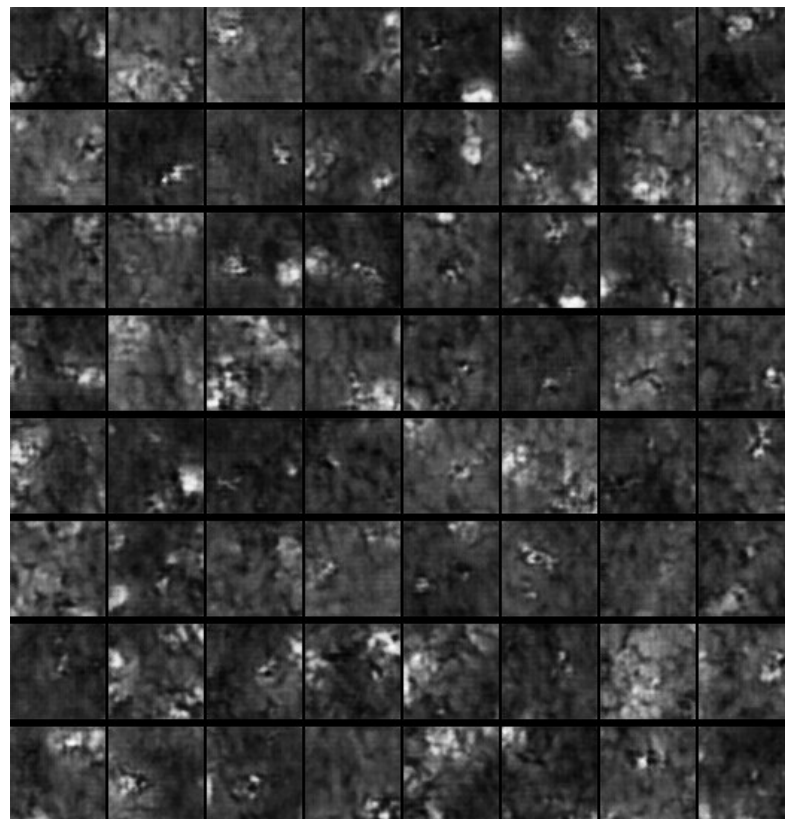


Tumour-on-chip videos

Results

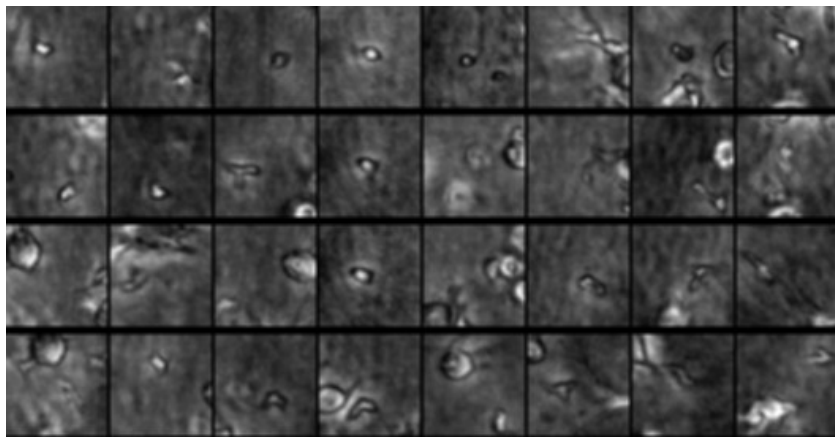


Tumour-on-chip videos

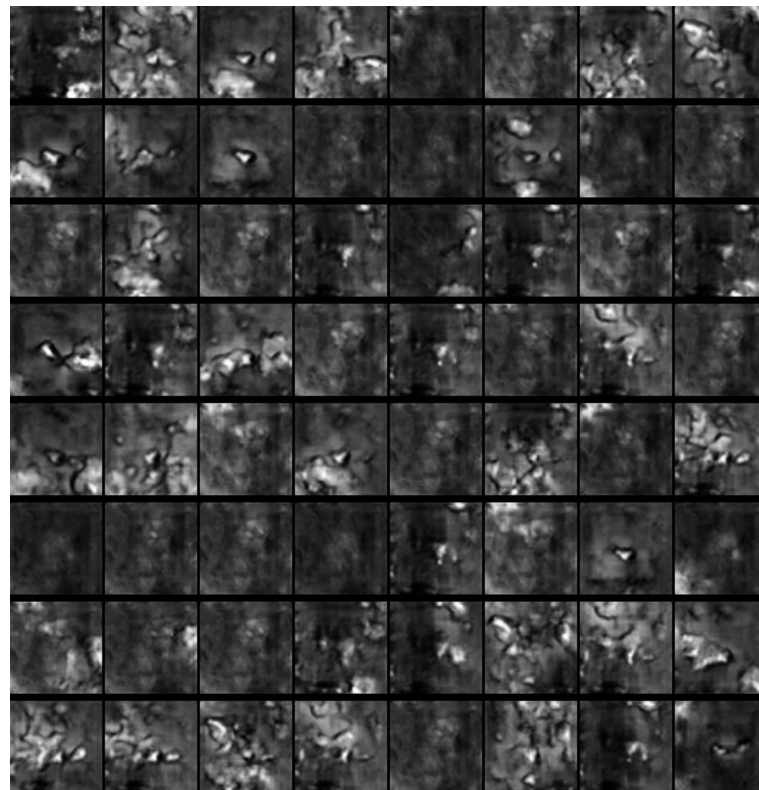


Generated with 3DGAN (no self-attention)

Results

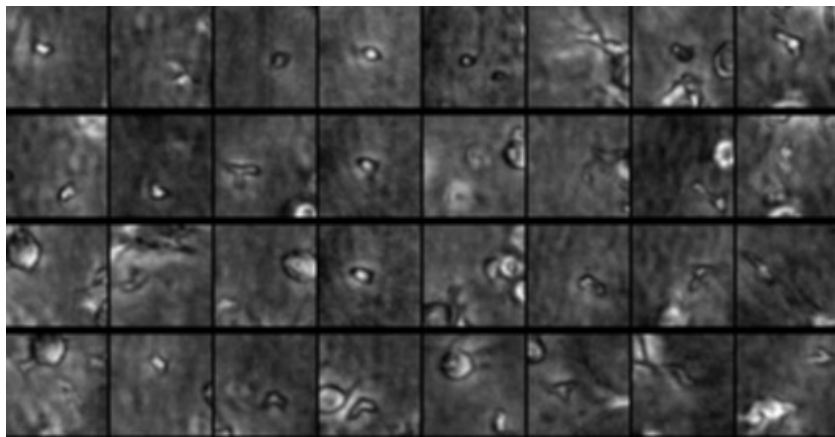


Tumour-on-chip videos

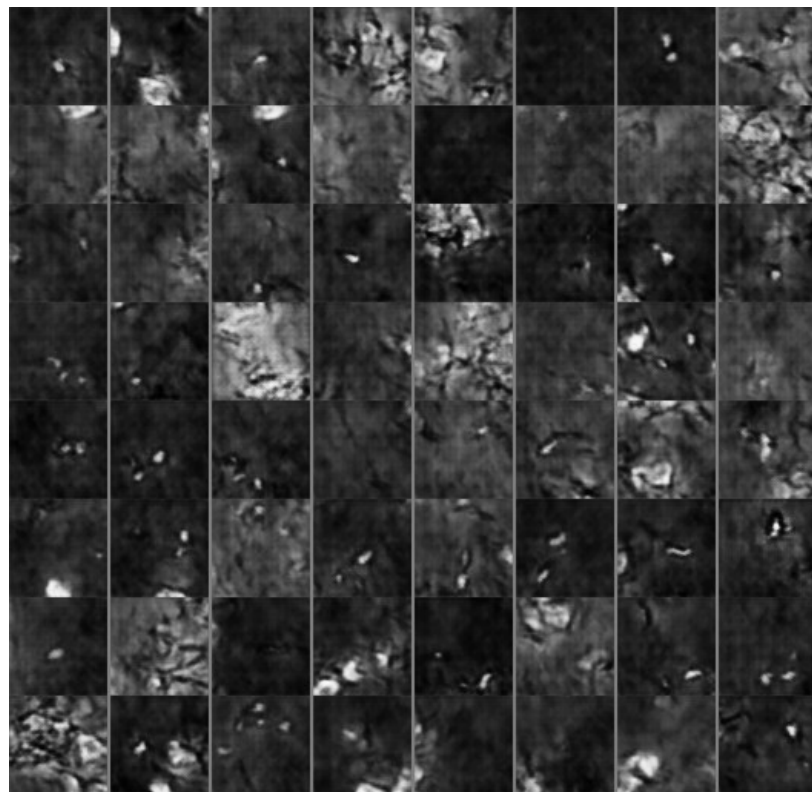


Generated with TGANv2 [Saito et al., IJCV 2020](#)

Results



Tumour-on-chip videos



Generated with SAVGAN (ours)

Results: Quantitative Comparison

Dataset	TOC			MMNIST		
Method	TGANv2	3DGAN	SAVGAN	TGANv2	3DGAN	SAVGAN
FVD	738.75	729.25	389.3	84.97	80.4	81.29
AACD	0.739	2.26	0.477	1.218	1.040	0.469
Flips/seq	-	-	-	3.42	3.1	1.76
Res18-cosTS	-	-	-	0.63	0.69	0.76

References

- M. Saito et al. “Train Sparsely, Generate Densely: Memory-efficient Unsupervised Training of High-resolution Temporal GAN”. In: IJCV. May 2020.
- Y. Xiong et al. “Nyströmformer: A Nyström-based Algorithm for Approximating Self-Attention”. In: AAAI. 2021

The End