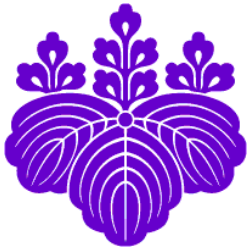
A Study on Image in-betweening using Generative Model

Paulino Cristovao[1], Hidemoto Nakada[1,2]
Yusuke Tanimura[1,2], Hideki Asoh[2]



1- University of Tsukuba
2- National Advanced Institute of Science and Technology

AIST

Objective

- This work focus on generating an image in-betweening using latent model.

First Frame

In-between Frame

Second Frame

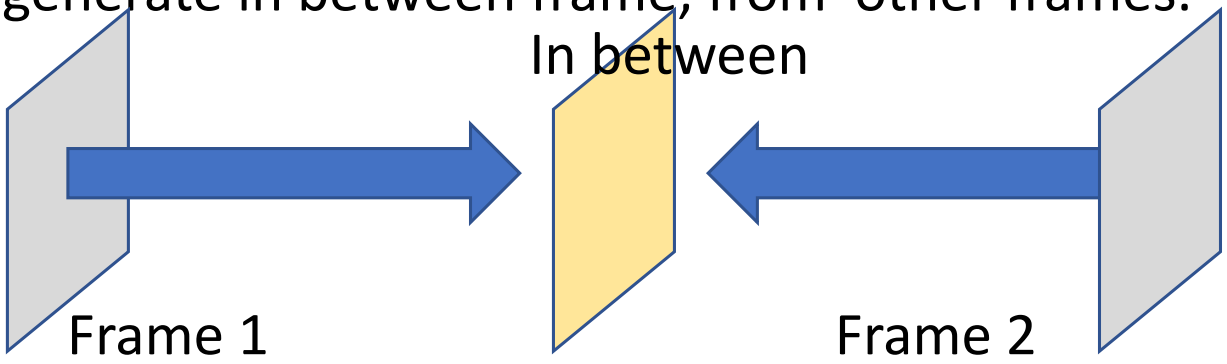
Introduction

- Existing Methods:
  - Optical flow
  - Pixel to pixel-based approach
- Issue:
  - They can not capture what is not present in the frame.
  - When frames are non-consecutive, they do not perform well.

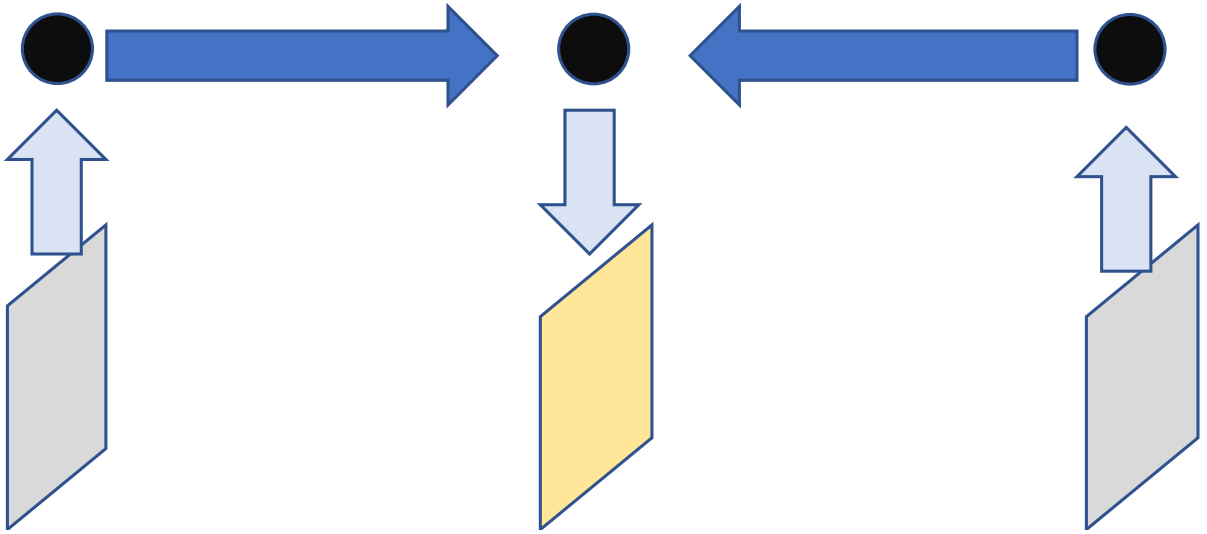
- In this work:
  - We design a new network architecture (based on VAE) which generates the image inbetween
  - This model can make assumption about the objects in the images

Proposed Method – Overview Latent-variable based in-betweening

- Existing methods: ‘pixel-based’
  - Directly generate in between frame, from other frames.



- ‘Latent-variable’ based method
  - Interpolate in the latent space and generate in-between frame.

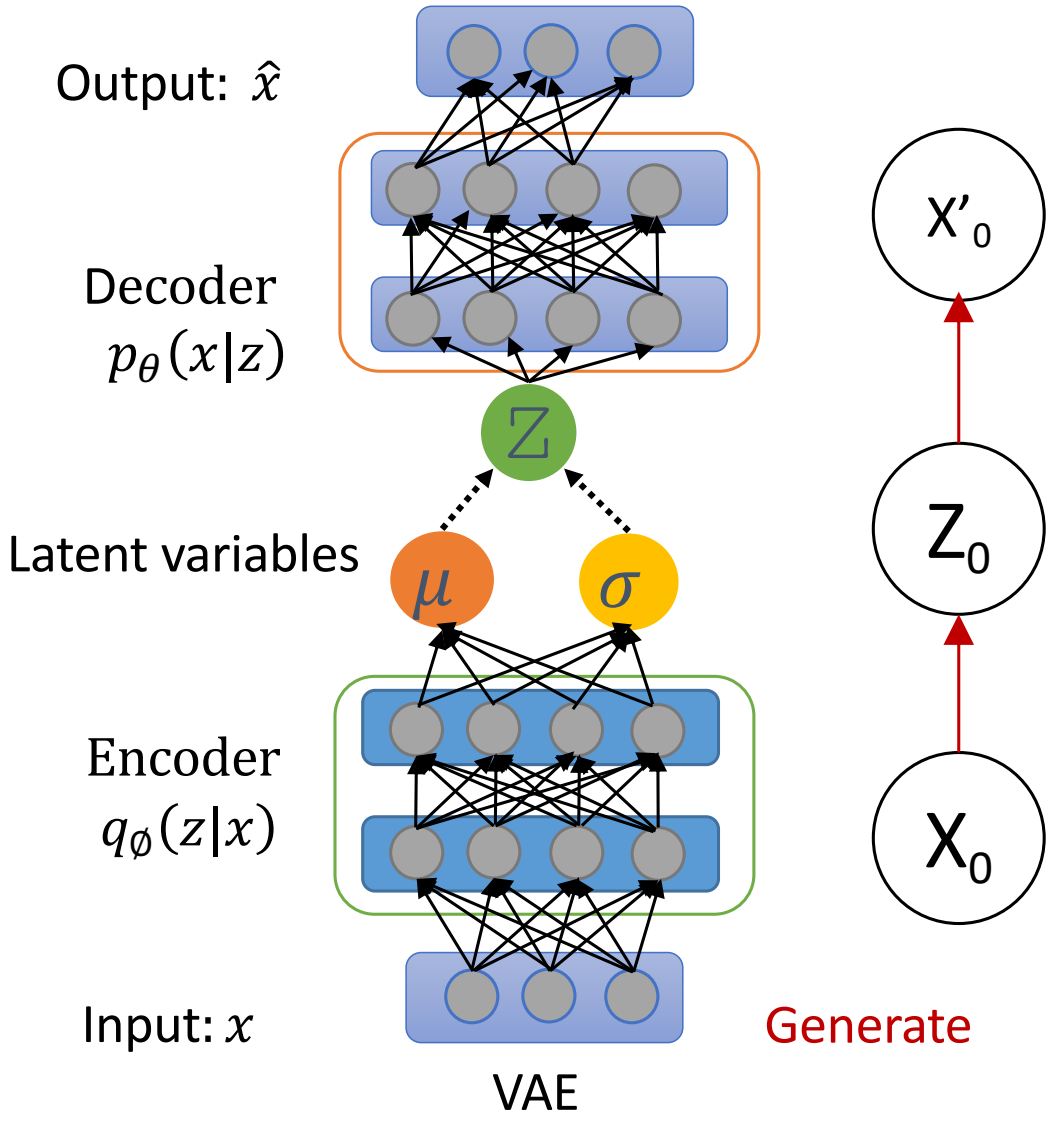


Content list

- Background
- Proposed Method
- Evaluation
- Conclusion
- Future Work

Variational autoencoder (VAE)

- VAE is a modern version of autoencoder
- Learns a latent representation of the hidden structures of its input data.
- Issue of VAE
  - It is hard to define exactly what information the latent variables care to capture latent information.

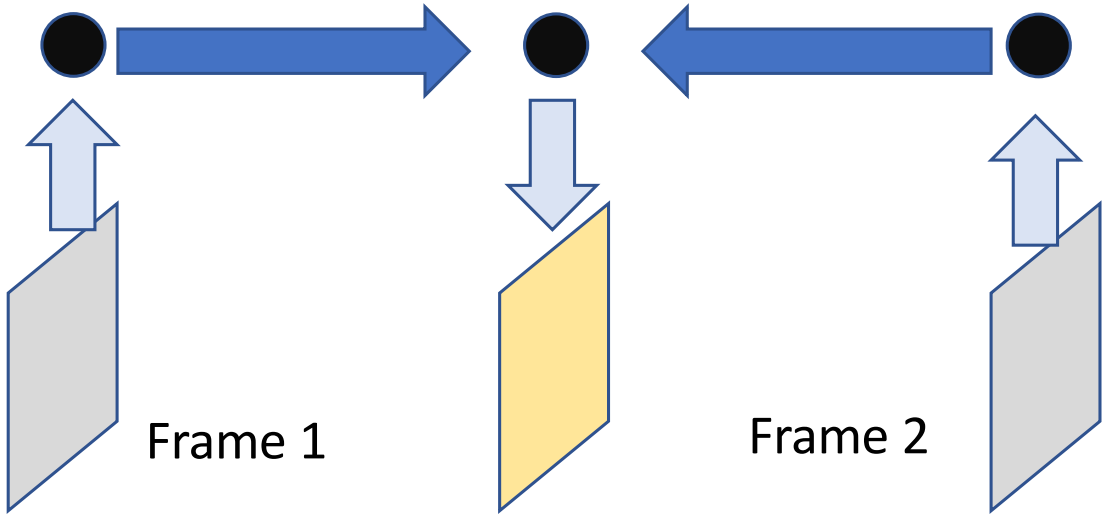


Content list

- Background
- Proposed Method
- Evaluation
- Conclusion
- Future Work

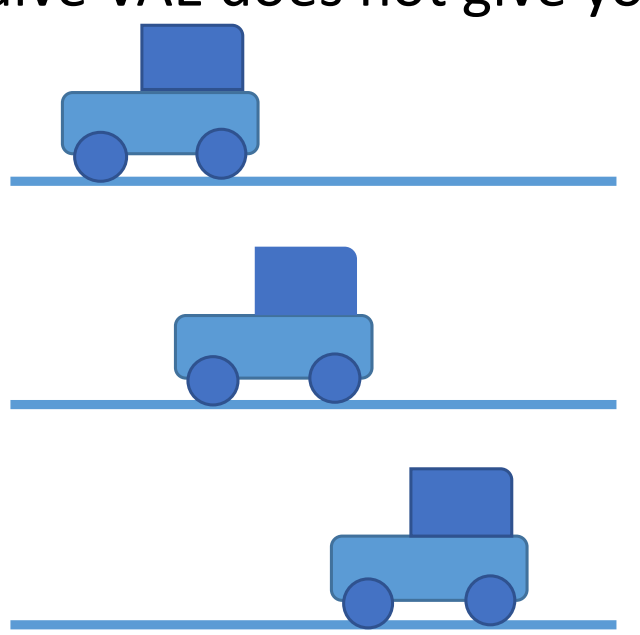
Proposed Method Latent-variable based in-betweening

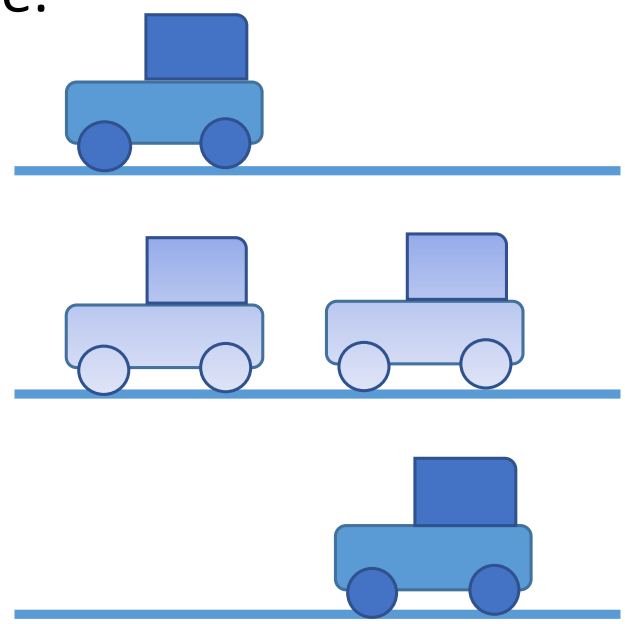
- ‘Latent-variable’ based method
  - Interpolate in the latent space and generate in-between frame.



How to have ‘such’ latent space?

- We need to have latent space such that,
  - Interpolation in the latent space
  - In-betweening in the image space
- Naïve VAE does not give you such latent space.



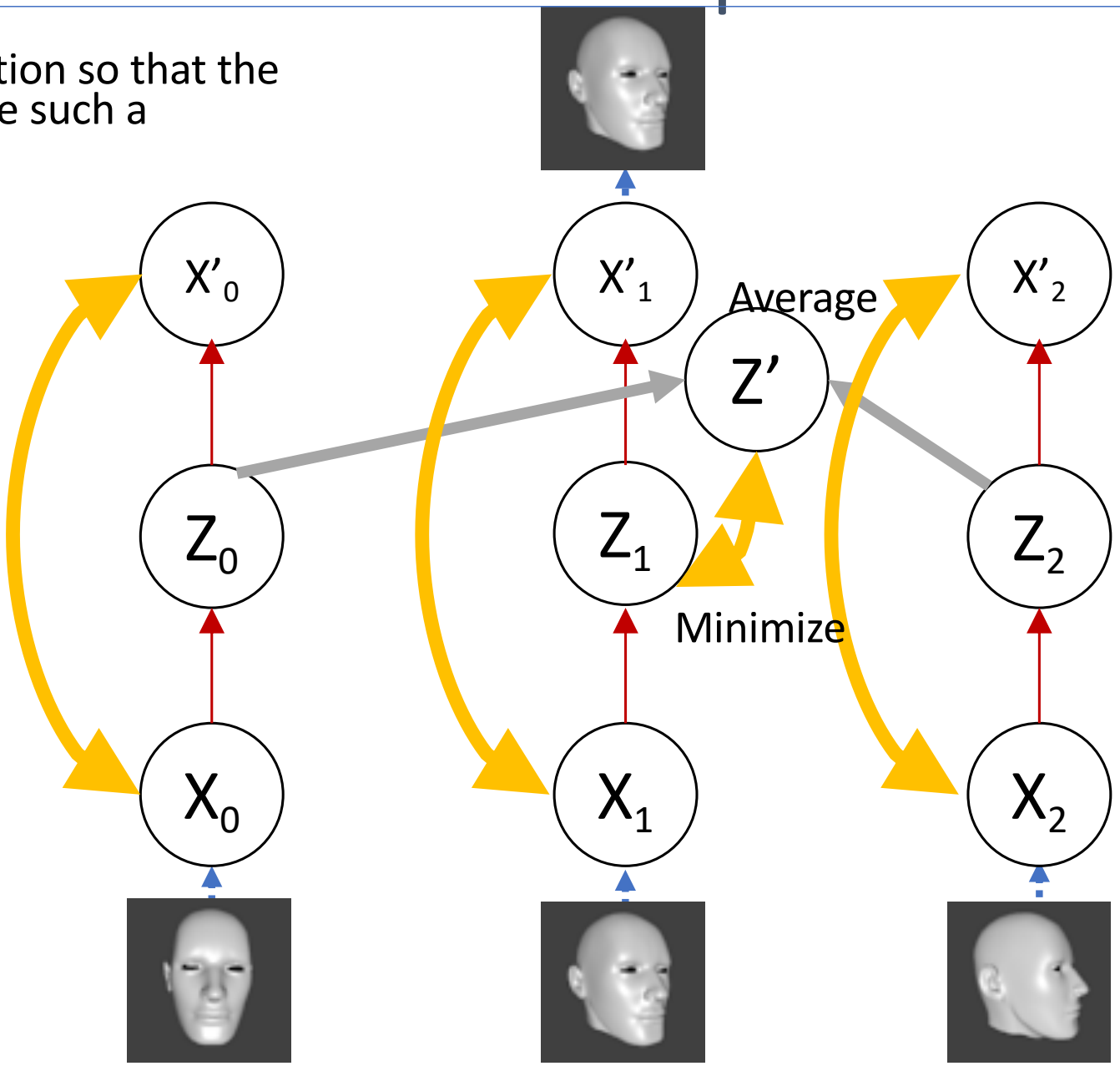


We might actually get..

➔ Need to control the latent space structure

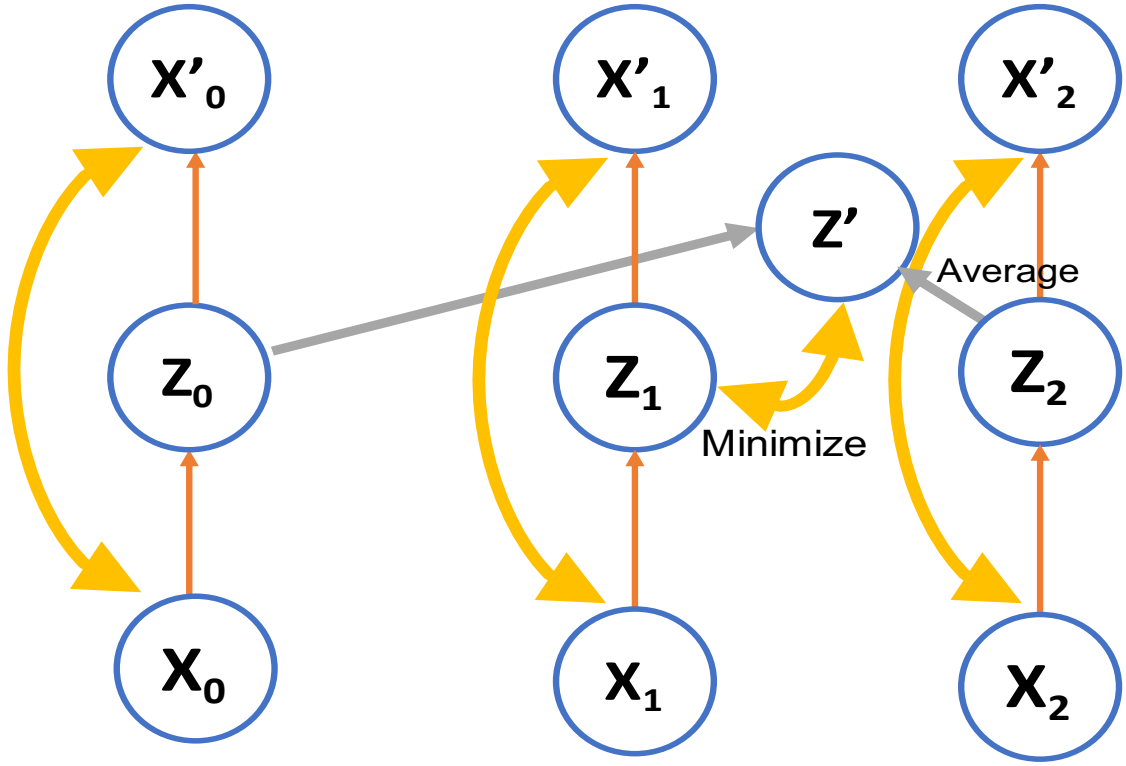
How to have ‘such’ latent space?

- Design loss function so that the latent space have such a structure.



Proposed - Loss Function

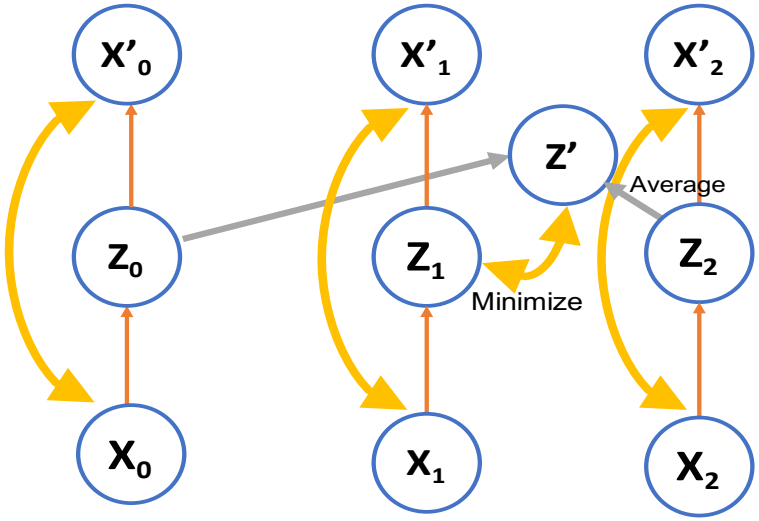
- $l_{(x_0,x_1,x_2)} = l_{VAE}(x_0) + l_{VAE}(x_1) + l_{VAE}(x_2) + \alpha\left(D_{KL}\left(q_{(x_1)} || \frac{q_{(x_0)} + q_{(x_2)}}{2}\right)\right)$
- Minimize: difference of latent variables(Z_1 and Z')



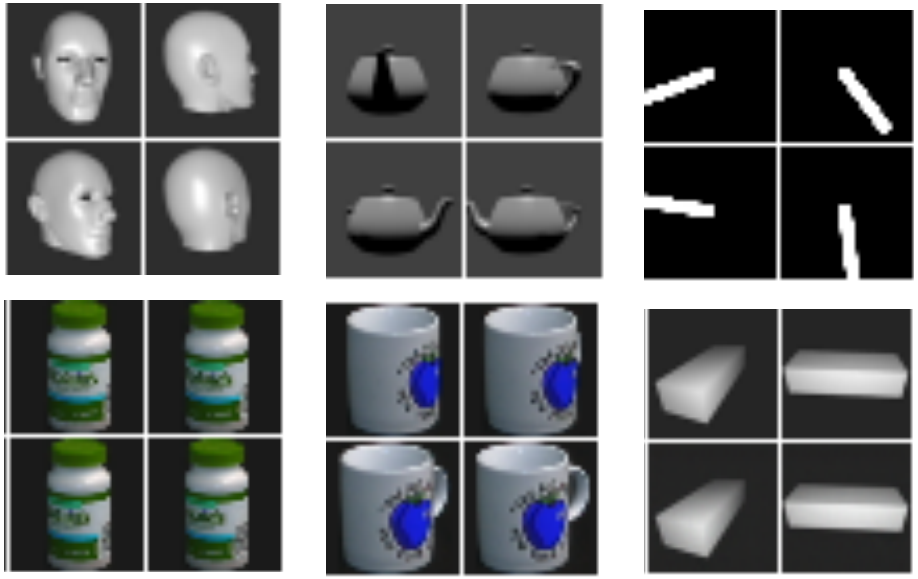
Content list

- Background
- Proposed Method
- Evaluation
- Conclusion
- Future Work

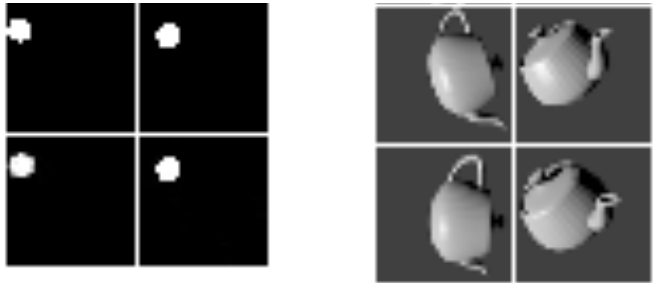
Dataset



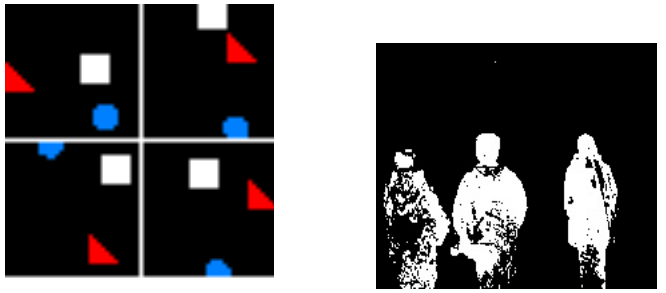
First phase – One degree of freedom



Second phase – Two degrees of freedom

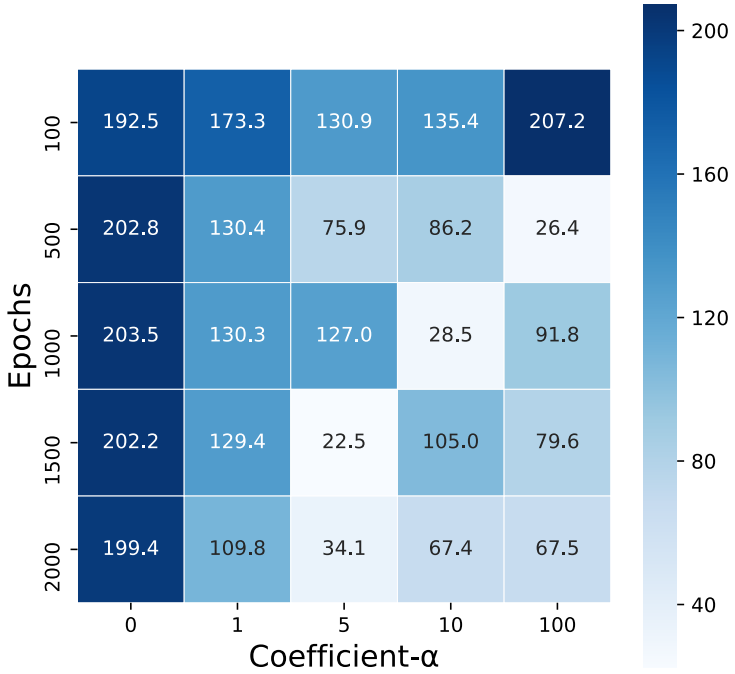


6 and above degrees of freedom



Evaluation

- Claim: This loss function generates an image inbetween.
- Adding adjustable hyper-parameter α is very important to generate a plausible image inbetween.



For evaluation:

- Image reconstruction
- Image in-between
- Evaluation Metrics
  - MSE

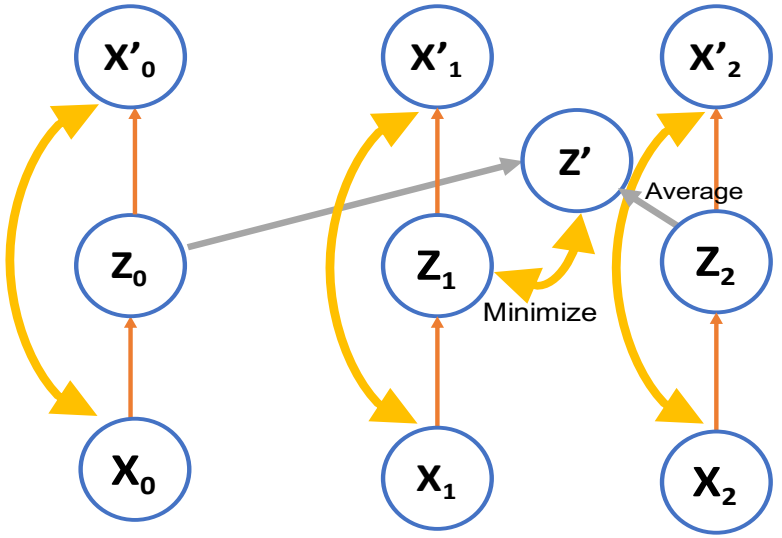


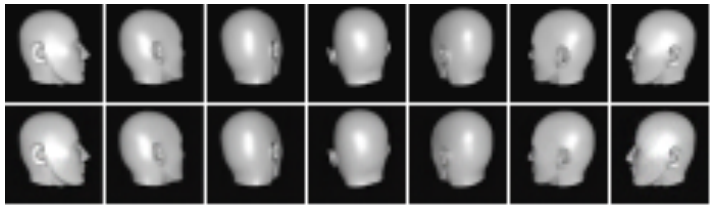
Image Reconstruction

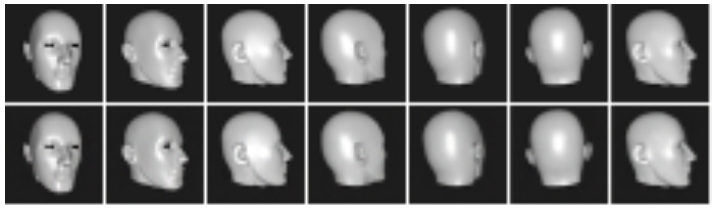
- Goal - Test the reconstruction location accuracy

Alpha: 0, epochs:1500

Original Image

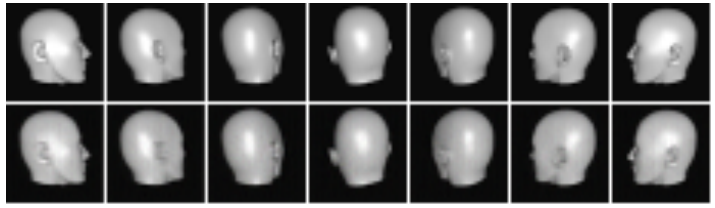
Reconstructed



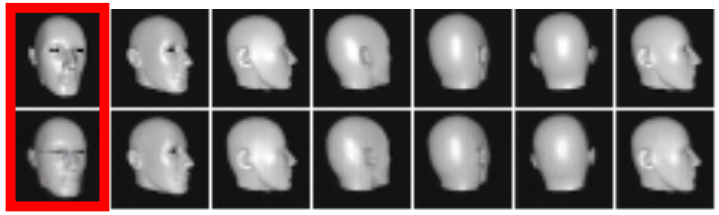


Our model - Alpha: 5, epochs:1500

Training



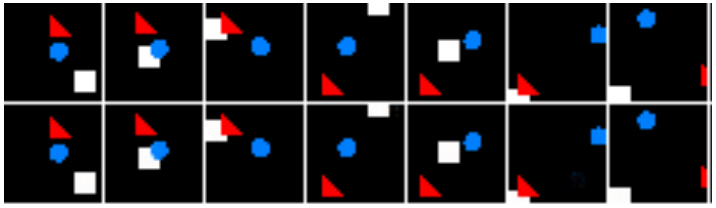
Testing

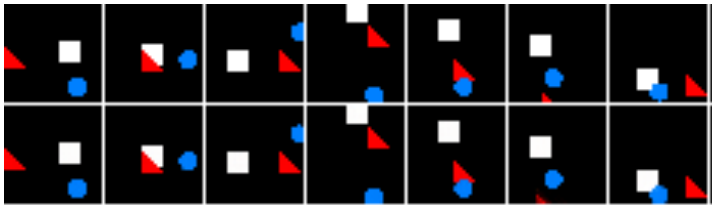


Alpha: 0, epochs:1500

Original Image

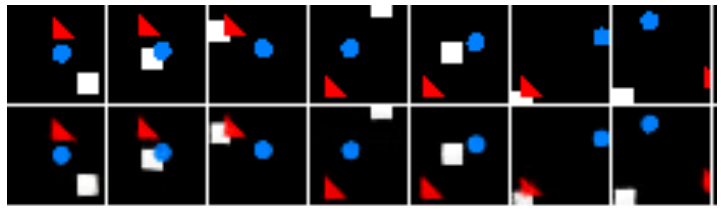
Reconstructed





Our model - Alpha: 100, epochs:1500

Training



Testing

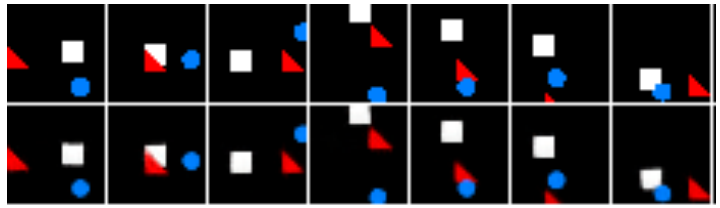


Image In-between

- Goal –Test Image in-between

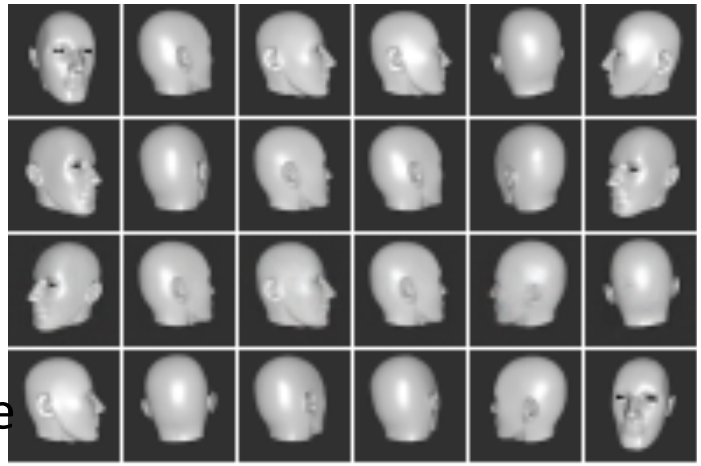
Alpha: 0, epochs:1500

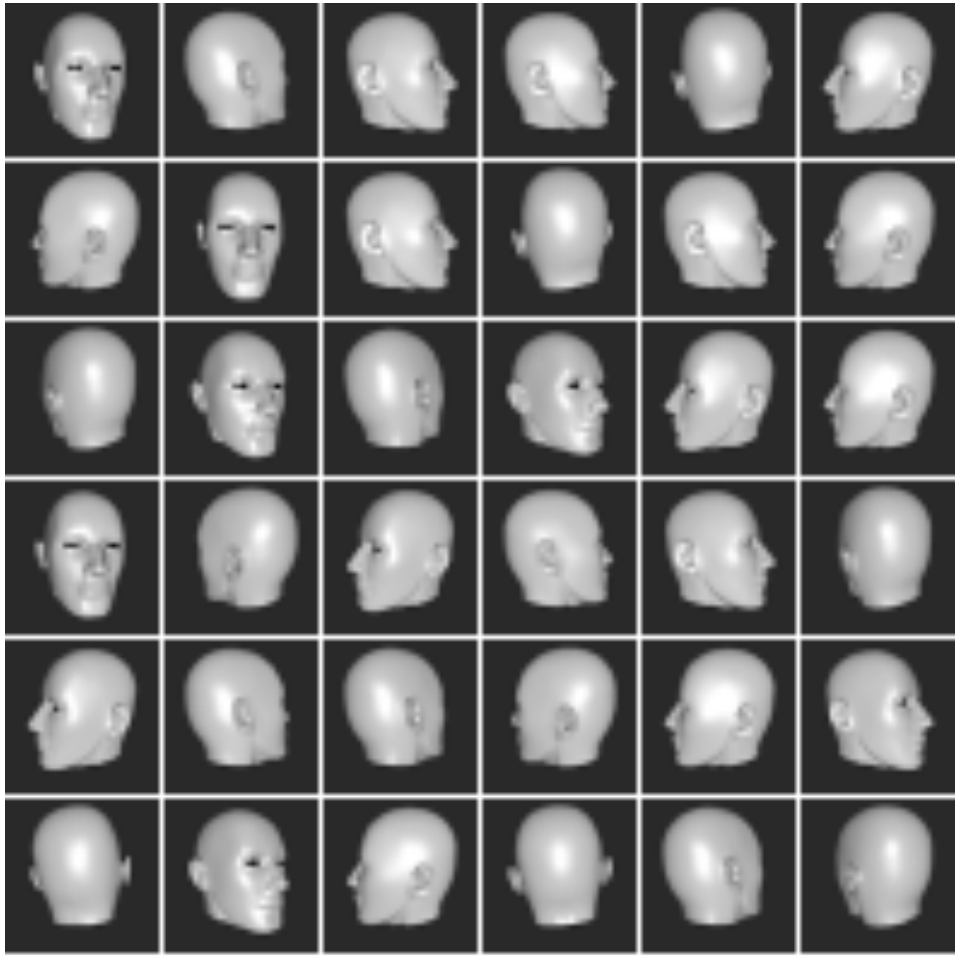
First frame

Ground Truth

Inbetween frame

Second frame





Our model - Alpha: 5, epochs:1500

First frame

Ground Truth

Inbetween frame

Second frame

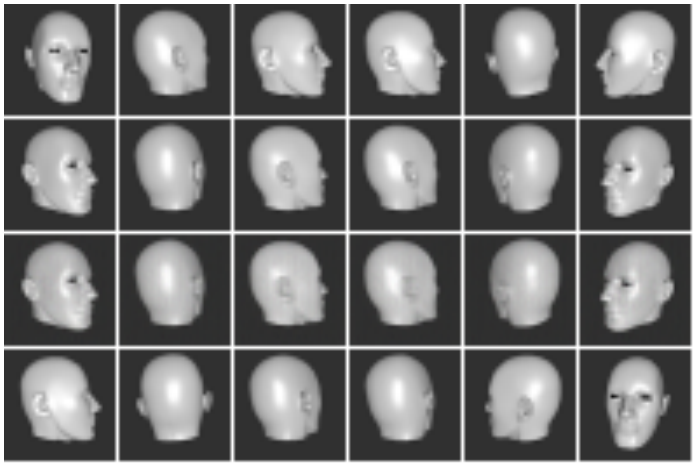


Image In-between - Two degrees of freedom

- Goal – test the image in-between

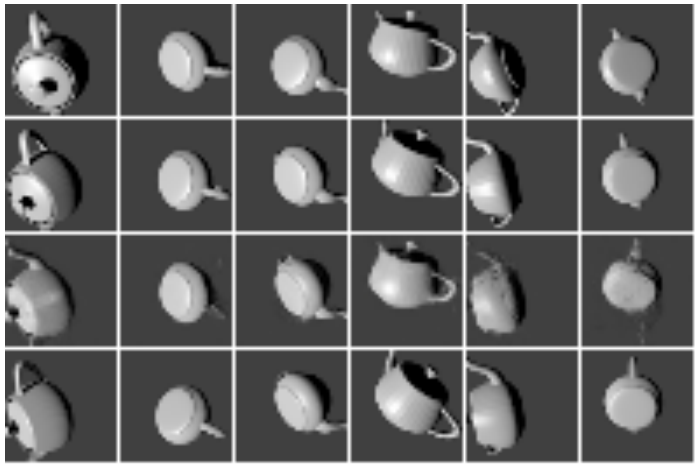
Alpha: 0, epochs:2000

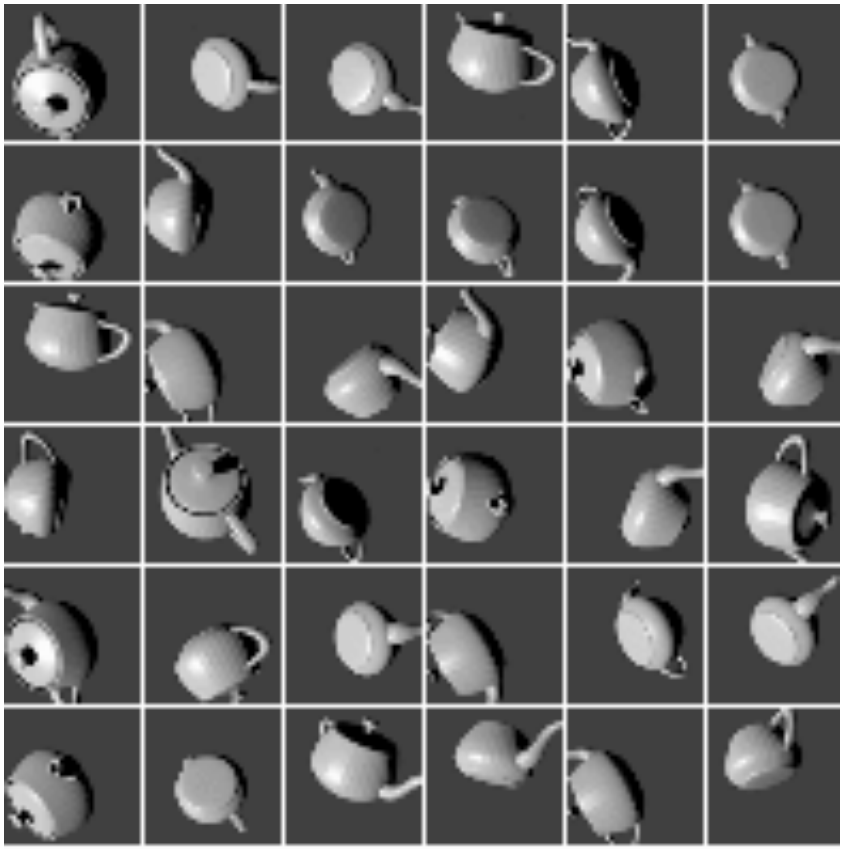
First frame

Ground Truth

Inbetween frame

Second frame





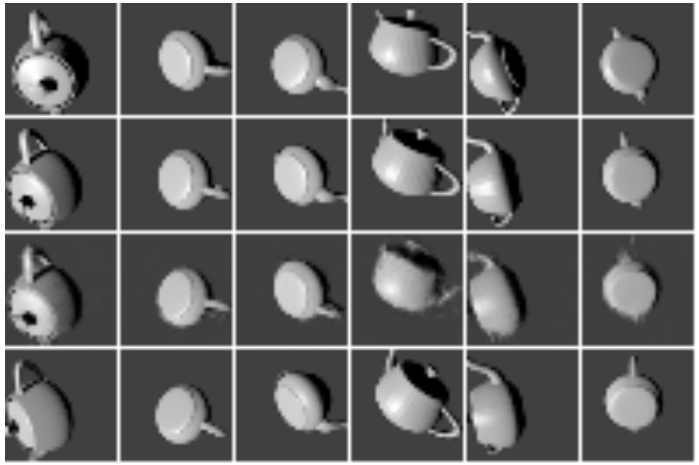
Our model - Alpha: 10, epochs:2000

First frame

Ground Truth

Inbetween frame

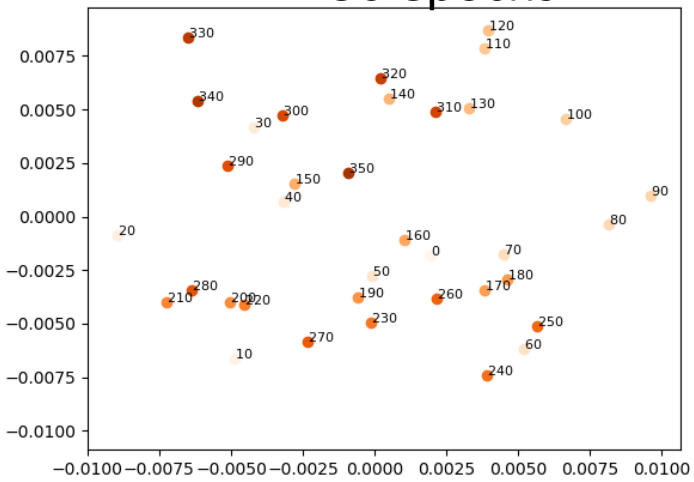
Second frame



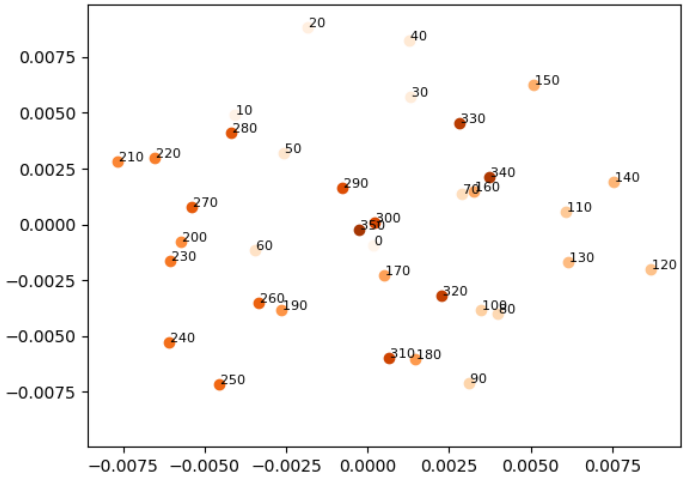
Latent Space Visualization - PCA

coeff.=0, Z=5

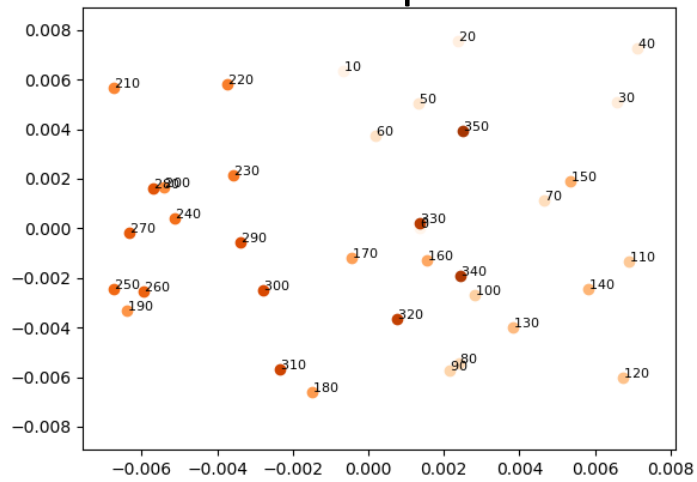
50 epochs



100 epochs

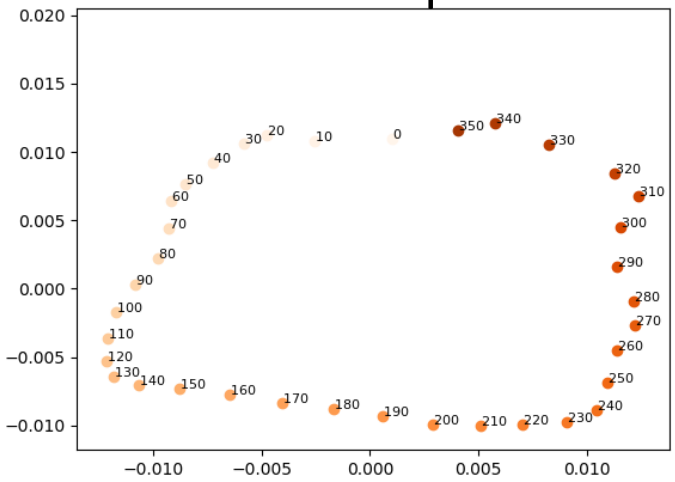


300 epochs

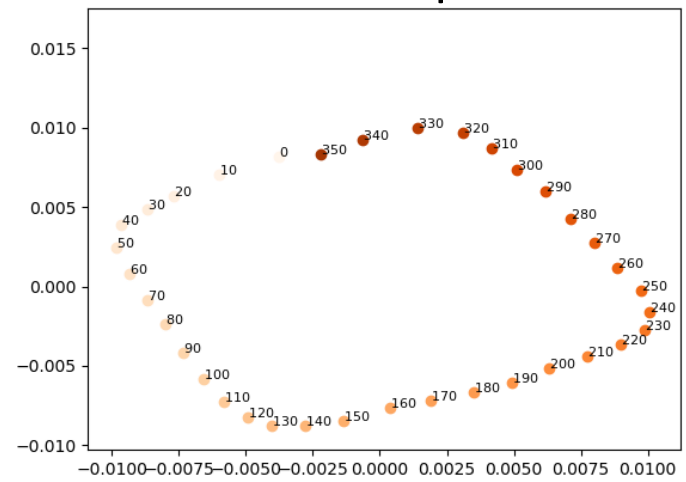


coeff.=10, Z=5

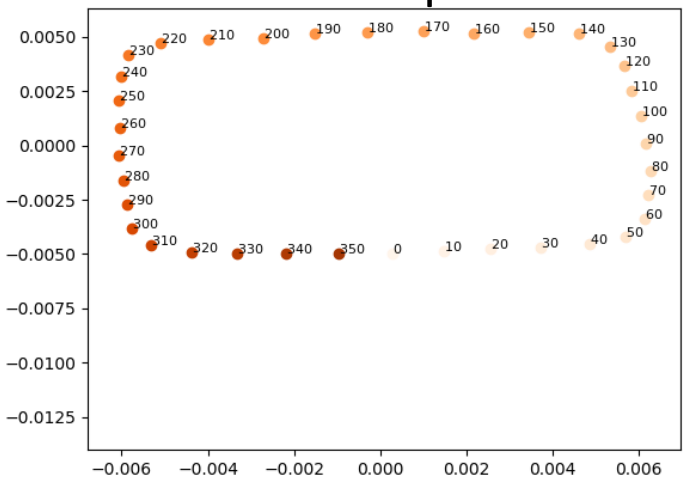
50 epochs



100 epochs

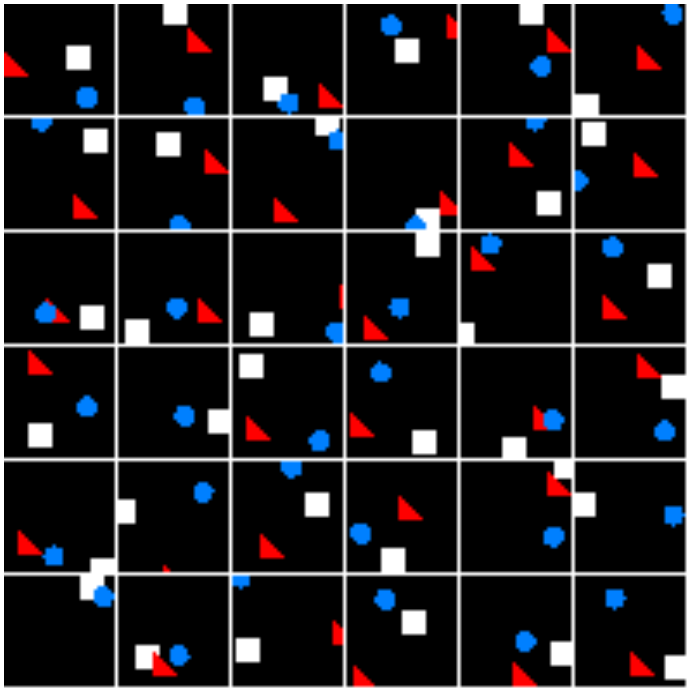


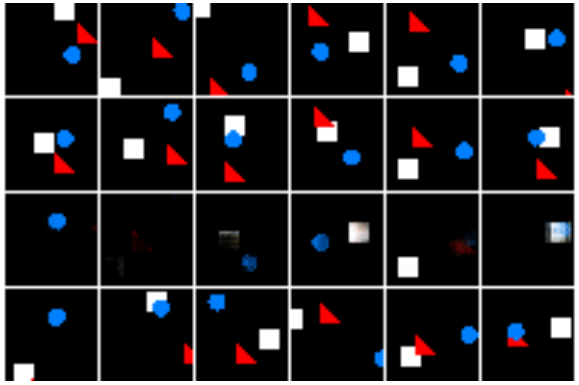
300 epochs



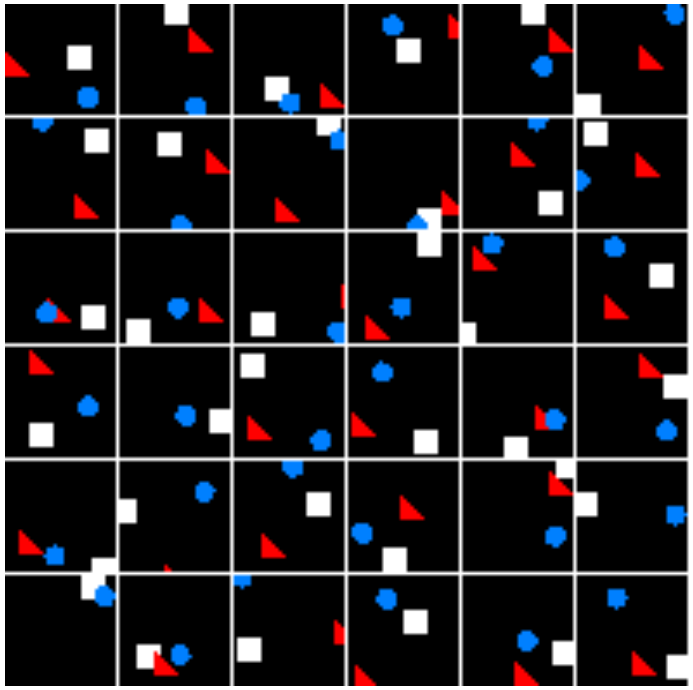
Multiple objects image inbetween

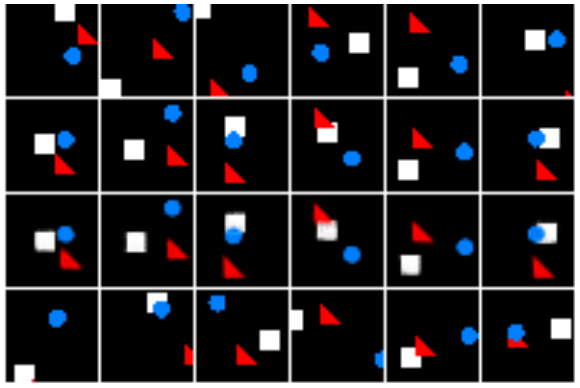
Coeff.0



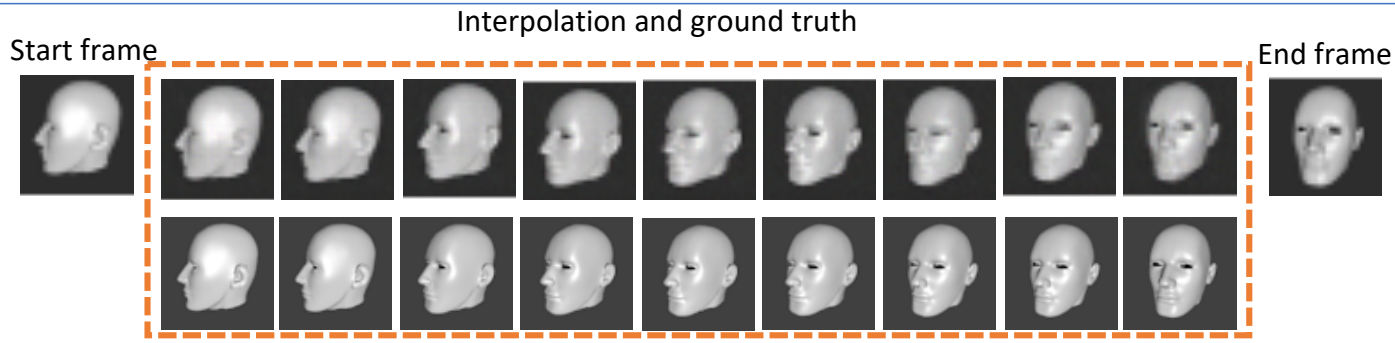


Coeff.100

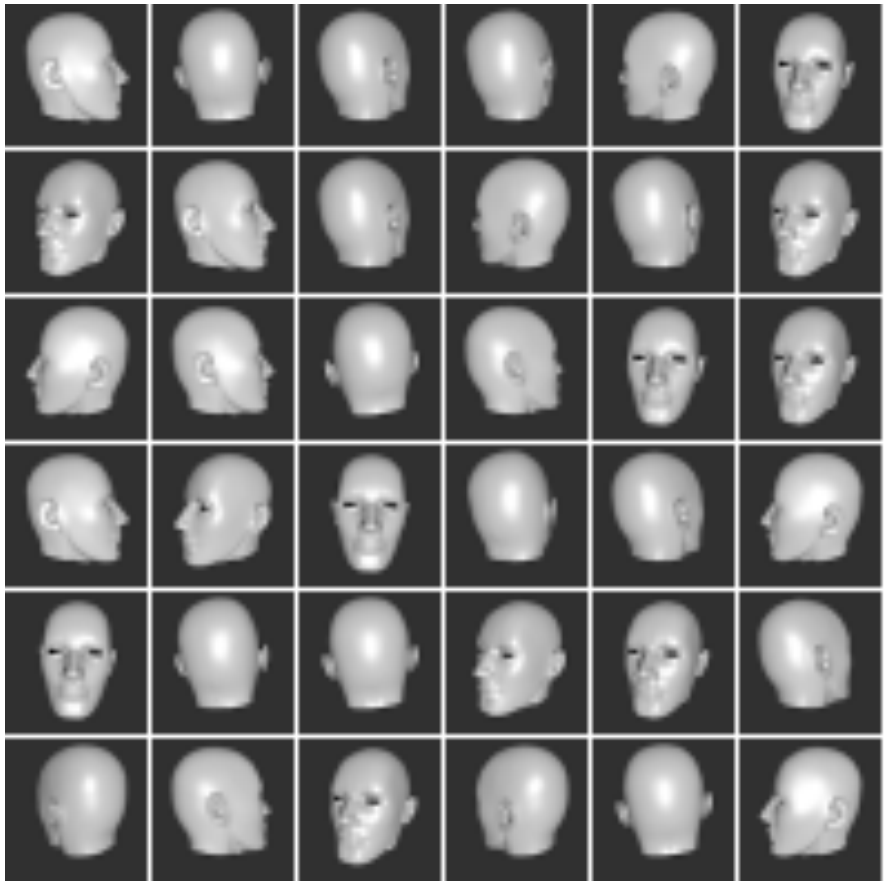




Linear Long-term Latent Interpolation

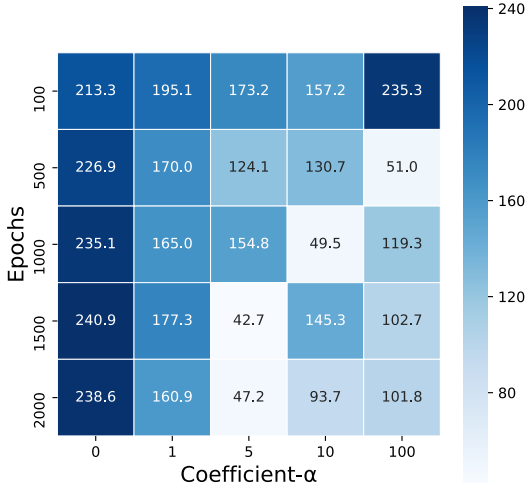


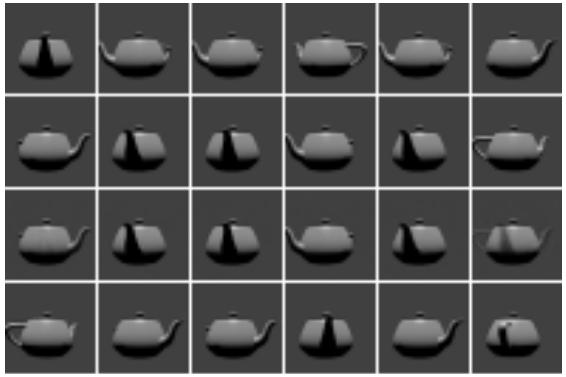
Testing



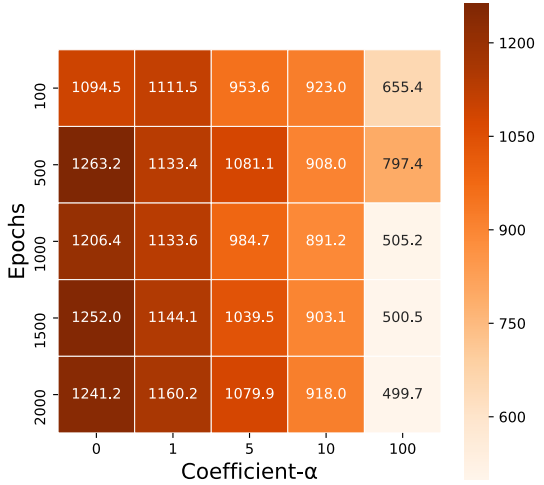
Evaluation based on metrics - MSE

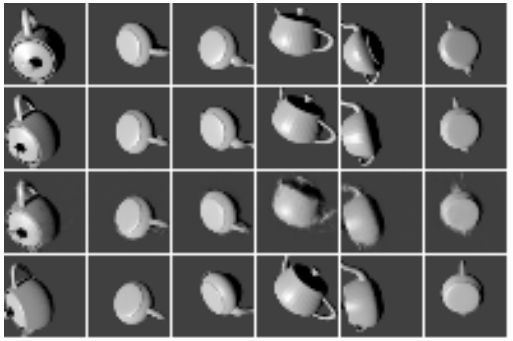
- "Teapot" dataset – One degree of freedom





- "Teapot" dataset – Two degrees of freedom



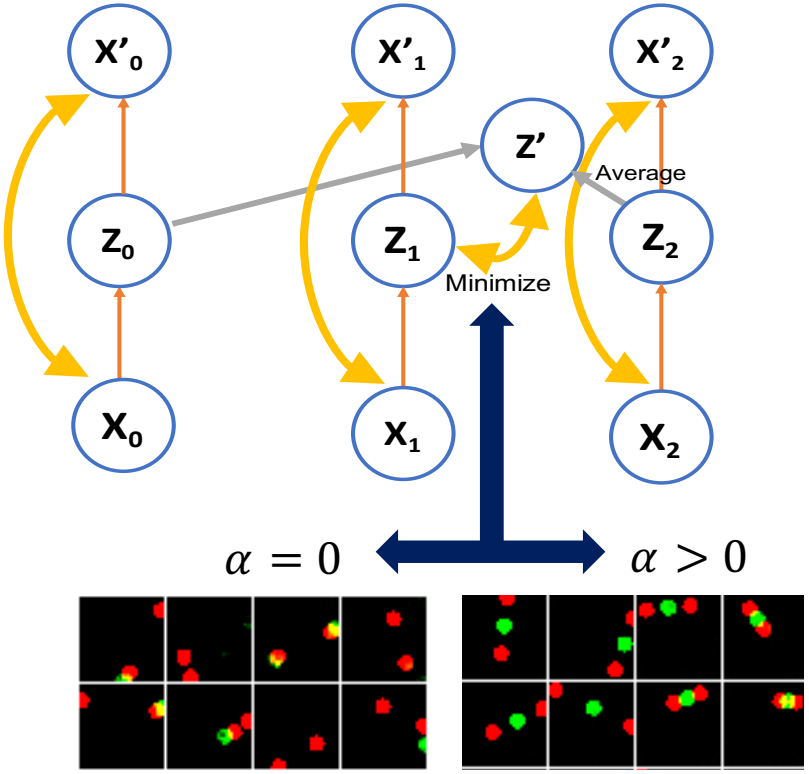


Content list

- Introduction
- Background
- Proposed Method
- Evaluation
- Conclusion
- Future Work

Conclusion

- We presented an alternative approach for generating an image interpolation by giving nearby frames which are non-consecutive images using a latent model.
- This model excels at predicting the interpolated image spatial location of the object in the image in addition the model generalizes for different datasets.
- Being able to twist the latent model gave us a desired outcome.



Content list

- Introduction
- Background
- Proposed Method
- Evaluation
- Conclusion
- Future Work

Future Work

- Try it with:
  - More complex physical models, such as linked arms.
  - Non-image data, for instance: text and audio data
  - Complex video i.e. video with fast motions and more moving objects.

Acknowledgement

- This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).
- This work was supported by JSPS KAKENHI Grant Number JP16K00116.