Deep Learning application for style control of virtual characters motions

Ali GHAMMAZ
Master 1/INFO4
Polytech Grenoble

Supervisors :
Katja ZIBREK , Ludovic HOYET ,
Yuliya PATOTSKAYA

Referent teacher:
Fabien RINGEVAL

INTRODUCTION

Exploring the animation qualities of photorealistic virtual humans.

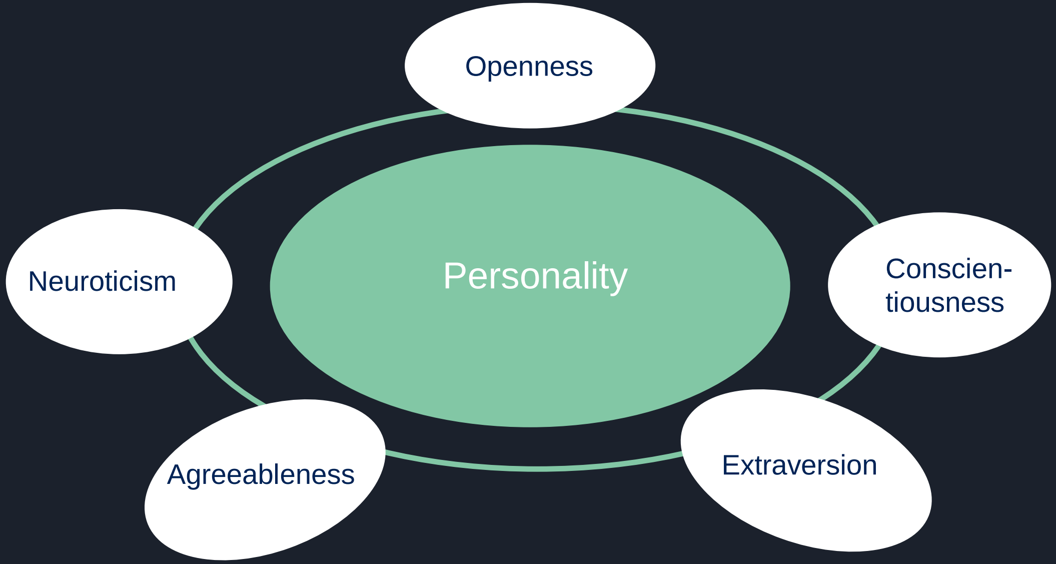
Identifying which aspects of the character animation affect the user’s perception and interaction with the character.

Appealing Character $\xrightarrow{\text{Impact ?}}$ User

What are our objectives?

1. Improve out Dataset
2. Generation of stylized motions (Motion synthesis)
3. Style tuning and Style transfer

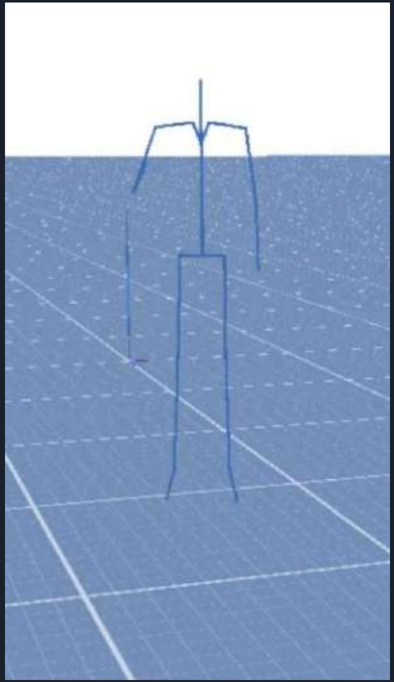
Styles



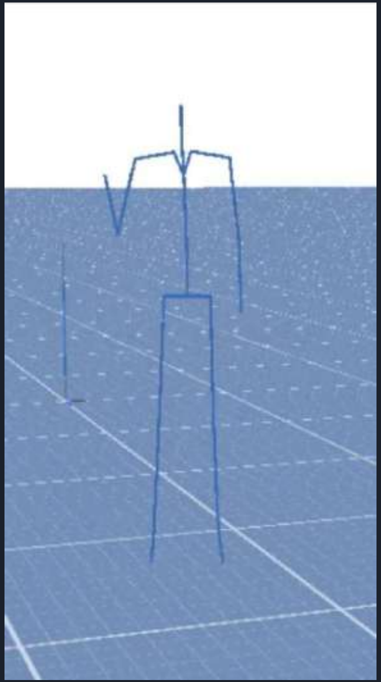
The big five personality traits: OCEAN Model

Our Styles

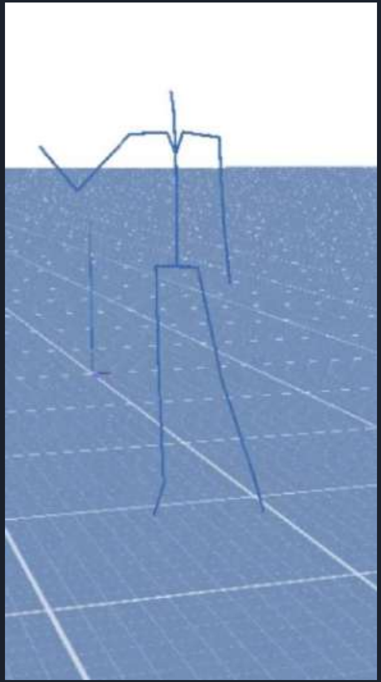
ES



Neutral

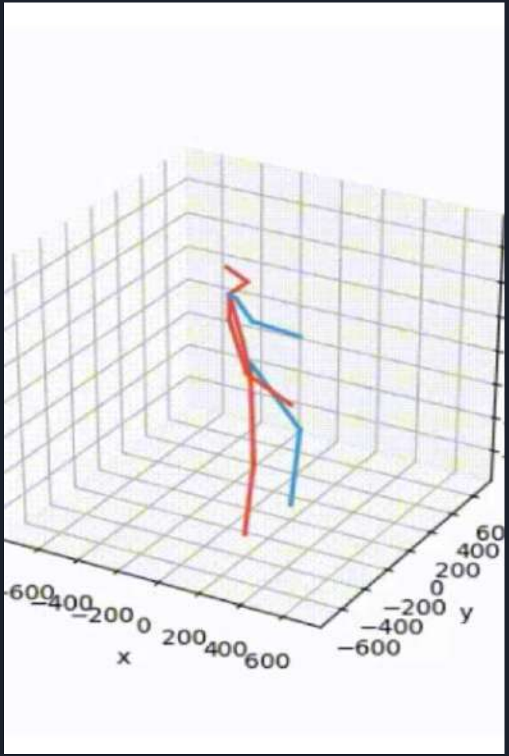


Neurotic

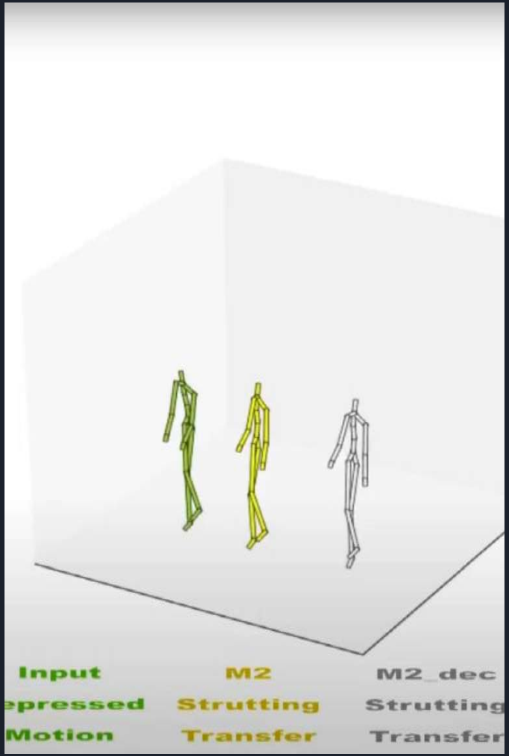


Motion Synthesis VS Style Transfer

- Motion Synthesis

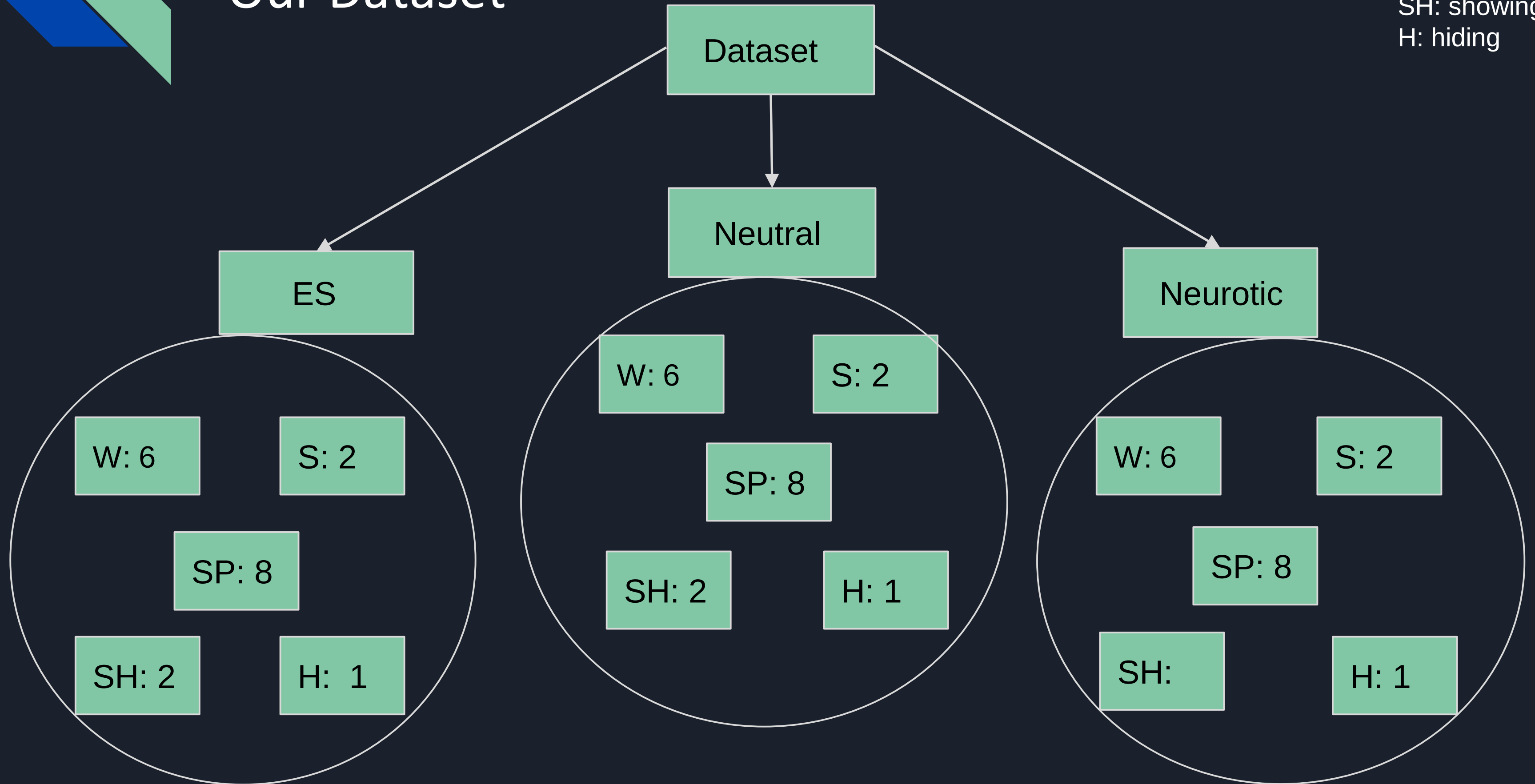


- Style Transfer



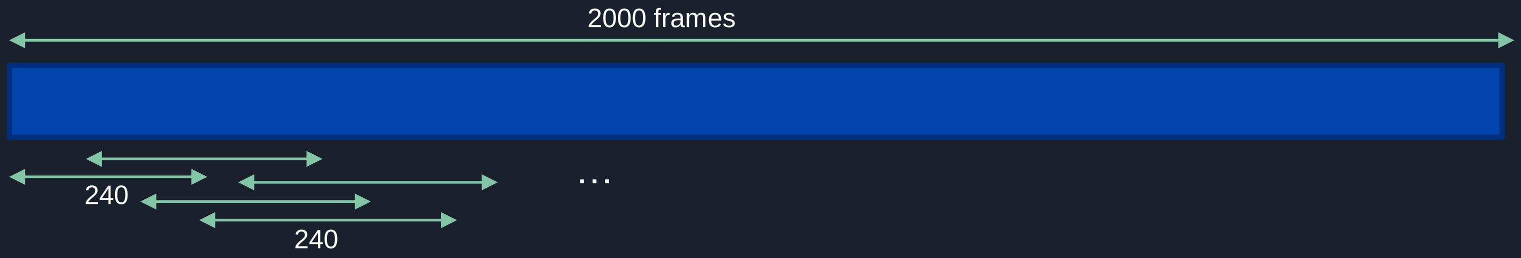
Our Dataset

W: waving
S: stopping
SP: showing phone
SH: showing
H: hiding



Expanding the dataset

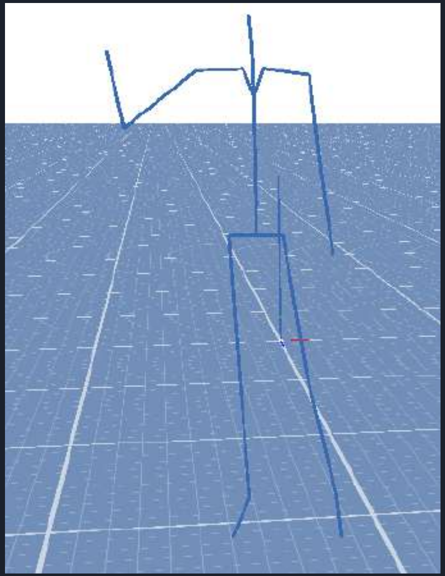
1-**Overlapping chunks**(for 2000 frame motion, cutting it in 240-frame sequences)

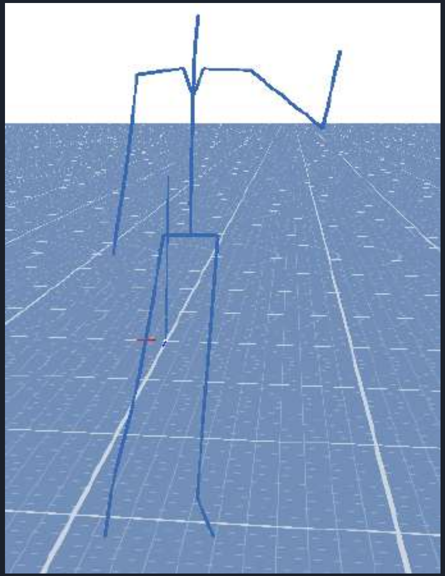


1 animation with 2000 frames → 15 animations with 240 frames

Expanding the dataset

2- Mirroring Dataset (Duplication of informations)





Motion Synthesis and Style Transfer in neural network related work

- On human motion prediction using recurrent neural networks, J.Martinez et al. 2017

RNN architecture for both pose labeling and generating future motion based on the past frames

- A Deep Learning Framework for Character Motion Synthesis and Editing ,Holden et al. ,2016

Realistic generation of motion for long periods of time based only on a high level control signal (such as a trajectory drawn on a plane

- A recurrent variational autoencoder for human motion synthesis, Habibe et al., 2017

A similar system for modelling periodic motion that additionally takes into account the stochastic nature of human motion by using the sampling behaviour of variational autoencoders

1st Approach: Human motion prediction using recurrent neural networks, Martinez et al. , 2017

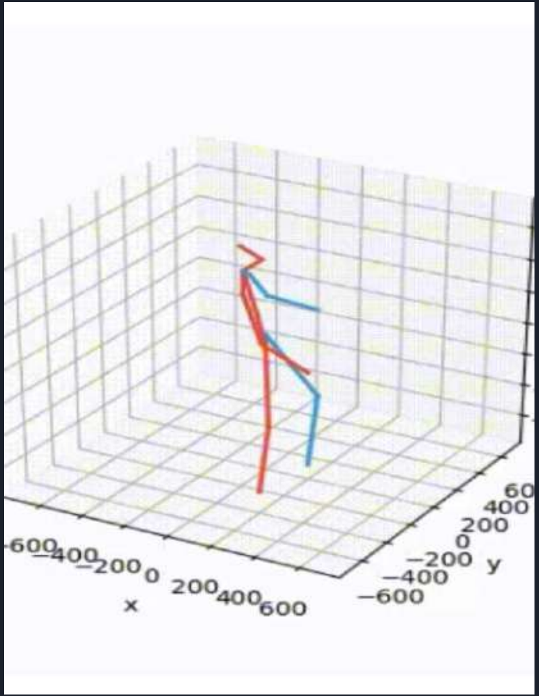
Reason :

- Easy approach
- Good results for many actions (walking , eating , smoking..)
- RNN architecture for both pose labeling and generating future motion based on the past frames

Constraints:

- Different Data format
- Not including learning from style
- Require a large dataset

https://arxiv.org/pdf/1705.02445.pdf



1-Converting our dataset to Exp Map format

Fbx Format

For each frame of an action and for each joint of the skeleton :

Two 4x4 matrix (Global and local)

```
[[ 0.93582509,  0.07633924, -0.34409841, -0.00790402],
 [-0.03896056,  0.99268532,  0.11427129,  0.90152412],
 [ 0.35030482, -0.09353167,  0.93195405,  3.42280487],
 [ 0.        ,  0.        ,  0.        ,  1.        ]]
```

Rotations

Positions

Exp Map Format

For each joint:

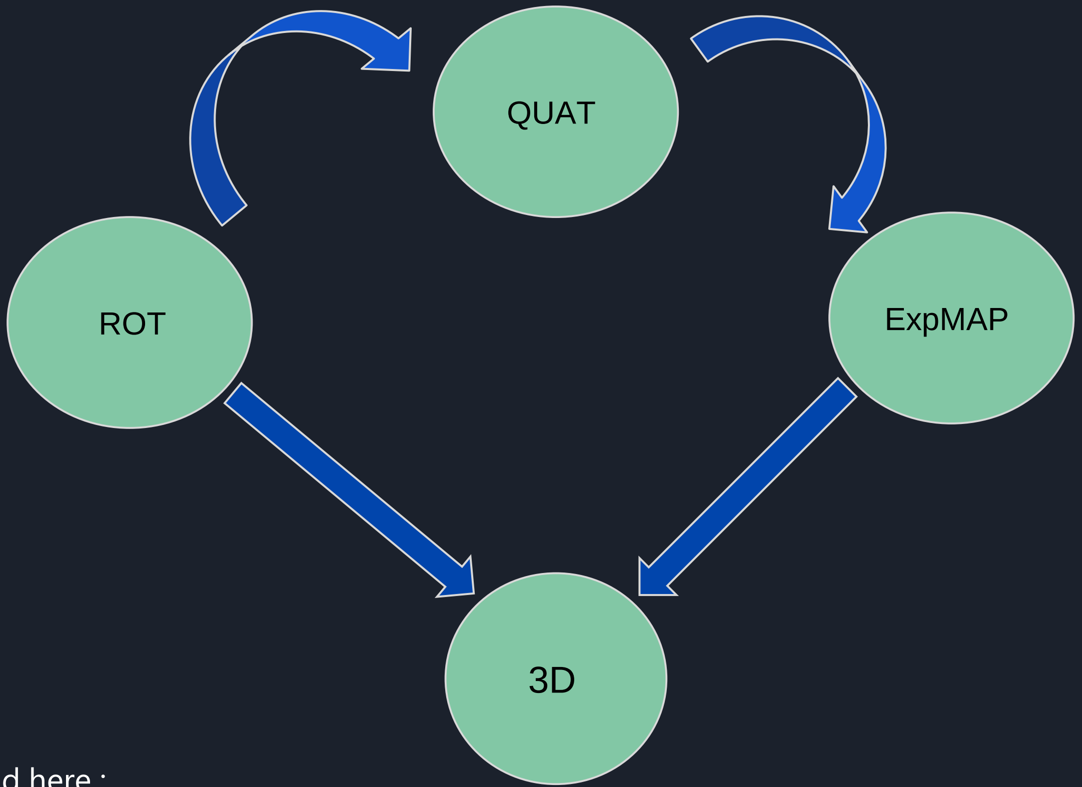
3 local rotations [x,y,z]

Joint 1

Joint n

```
[[-0.007904   0.9015241  3.4228048 ... -0.2135291  0.5530091 -0.0235806]
 [-0.0081086  0.9015495  3.4229658 ... -0.2123501  0.5523702 -0.0224363]
 [-0.0082679  0.9015689  3.423084  ... -0.2114942  0.5518609 -0.021585 ]
 ...
 [ 0.0110366  0.9013949  3.4029183 ... -0.3584471 -0.1209618 -0.0424051]
 [ 0.0102839  0.9017921  3.401889  ... -0.358592  -0.1213635 -0.0423724]
 [ 0.0096719  0.9021094  3.4010482 ... -0.3587122 -0.1217161 -0.0423553]]
```

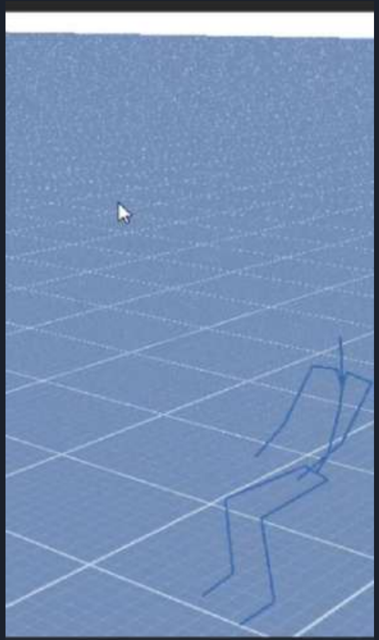
1-Converting our dataset to Exp Map format



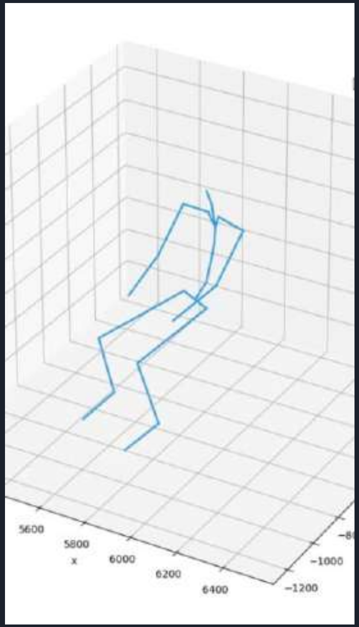
Code could be found here :
https://github.com/ichiro10/DeepL_VirtualCharactersMotions/blob/main/Pynimation/fbx2exp.py

1-Converting our dataset to Exp Map format

3D visualization from fbx format:

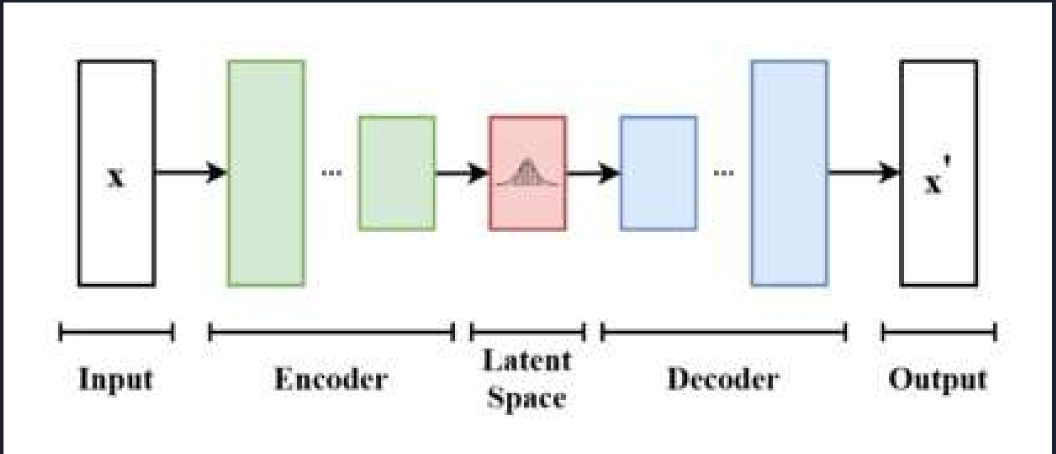


3D visualization from Exp Map format:



2- Adding a VAE to the model

VAE is a powerful tool that can generate a stylized motion



By manipulating the values in the latent space, motion sequences with different styles can be generated

Style labeling(One hot encoding)

0 1 2

```
styles = ['ES', 'Neurotic', 'Normal']
motions = ['waving', 'stopping', 'showingphone', 'showing', 'hiding']
```

0 1 2 3 4

```
[0, 0] [0, 0] [0, 0] [0, 0] [0, 0] [0, 0] [1, 0] [1, 0] [1, 0] [1, 0] [1, 0]
[1, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [3, 0] [3, 0]
[4, 0] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [1, 1] [1, 1]
[1, 1] [1, 1] [1, 1] [1, 1] [2, 1] [2, 1] [2, 1] [2, 1] [2, 1] [3, 1] [3, 1]
[4, 1] [0, 2] [0, 2] [0, 2] [0, 2] [0, 2] [0, 2] [1, 2] [1, 2] [1, 2] [1, 2]
[1, 2] [1, 2] [2, 2] [2, 2] [2, 2] [2, 2] [3, 2] [3, 2] [4, 2]
```

Testing for later ..

Style Tuning & Style transfer

Ian Mason et al. Works

2022

DeepPhase: Periodic Autoencoders for Learning Motion Phase Manifolds, SIG/ACM Transactions on Graphics, 2022.

-Most interesting
-Missing data

Real-Time Style Modelling of Human Locomotion via Feature-Wise Transformations and Local Motion Phases, i3D/ACM 2022

2020

…

2018

Few-shot Learning of Homogeneous Human Locomotion Styles,2018

2017

Thesis : Motion Generation and Style Transfer in Variational Autoencoder Systems, 2017

Solution

Motion Generation code

Style labeling ideas

Dataset

Same dataset

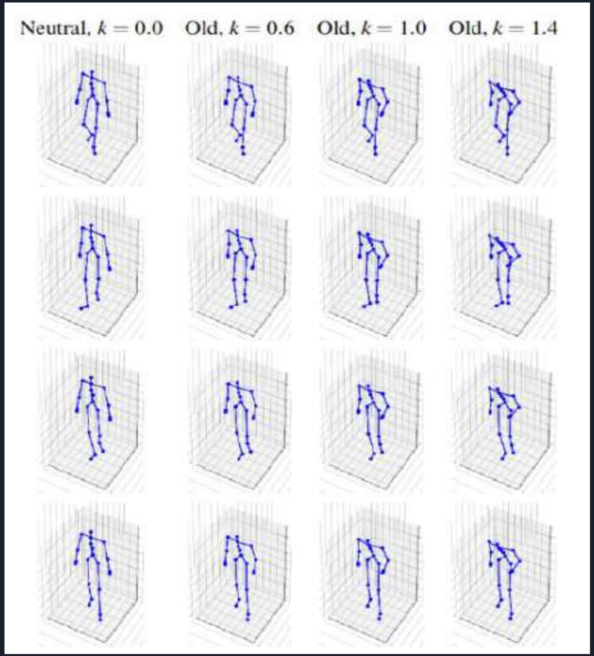
Habibie, I., Holden, D., Shwarz, J., Yearsley, J., and Komura, T. (2017). A recurrent variational autoencoder for human motion synthesis. In BMVC

Holden, D., Saito, J., and Komura, T. (2016). A deep learning framework for character motion synthesis and editing. ACM Trans. Graph., 35

Motion Generation and Style Transfer

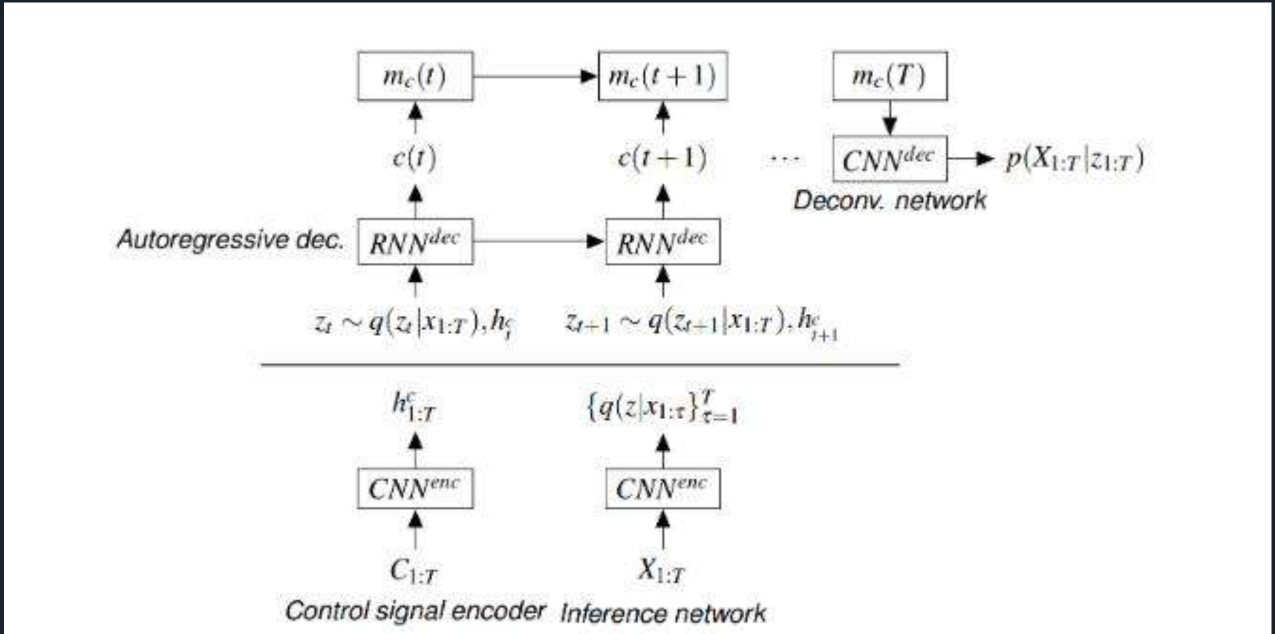
- Easy to implement
- Capable of generating stylised human motion from a high level control signal input

- By using the VAE-LSTM architecture recently proposed by Habibie et al. (2017) along with a one hot label representing a style of motion, our system is able to learn to switch between styles in a manner that is both efficient and consistently generates realistic and natural human motion.



https://project-archive.inf.ed.ac.uk/msc/20172562/msc_proj.pdf

Style Tuning : Motion Generation and Style Transfer in Variational Autoencoder Systems for Human Motion Synthesis (Ian MASON) 2017, Thesis



The VAE-LSTM architecture diagram from Habibie et al. (2017). C1:T represents the control signal over T frames, and X1:T the joint positions

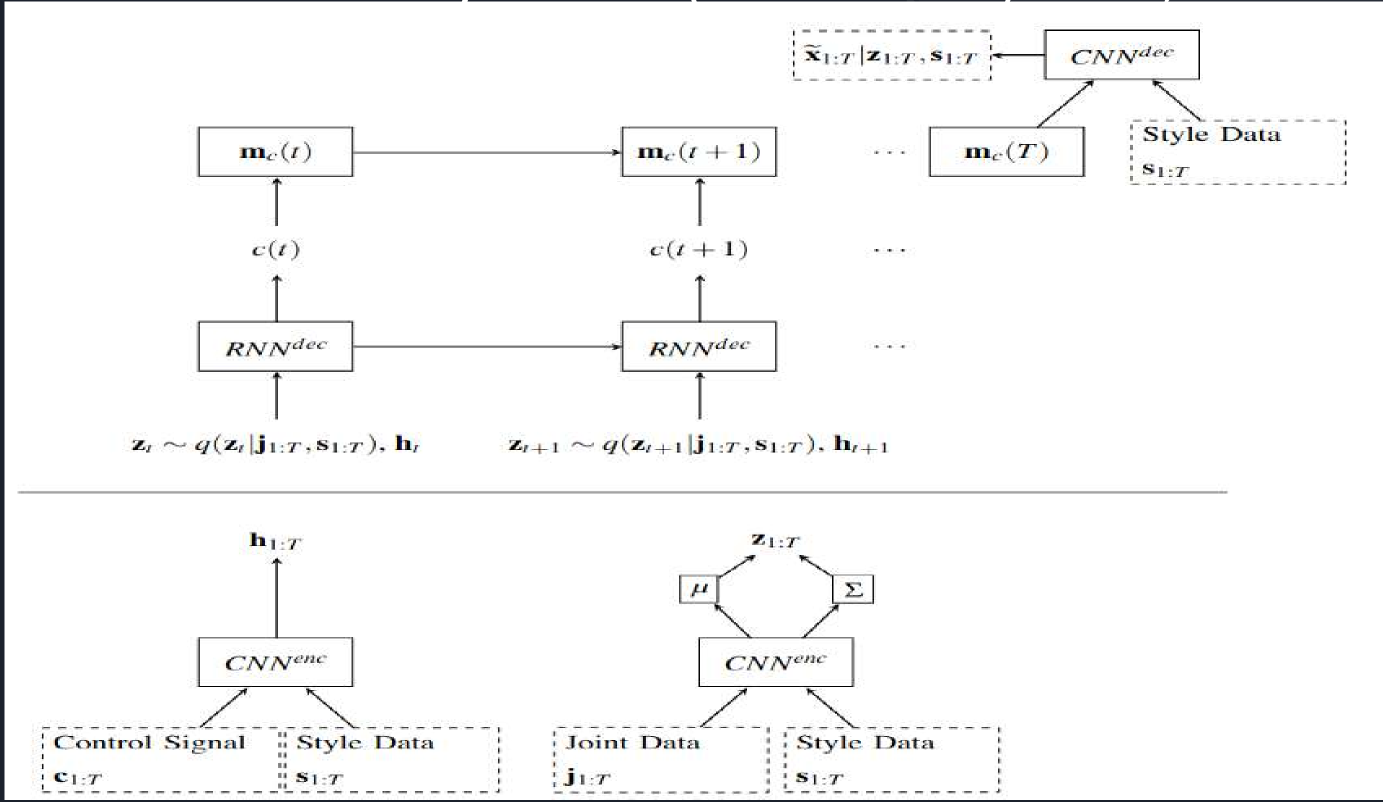
Style Tuning : Motion Generation and Style Transfer in Variational Autoencoder Systems for Human Motion Synthesis (Ian MASON) 2017, Thesis

- Addition of a style label to VAE-LSTM architecture of Habibee to alter the style.

- Where adding a style label ?

1. M2 dec - Style label only applied to convolutional decoder.
2. M2 enc2 - Style label only applied to encoder for joint positions.
3. M2 enc2 dec - Style label applied to encoder for joint positions and convolutional decoder.
4. M2 enc1 enc2 - Style label applied to encoder for joint positions and encoder for control signal.

Style Tuning : Motion Generation and Style Transfer in Variational Autoencoder Systems for Human Motion Synthesis (Ian MASON) 2017, Thesis

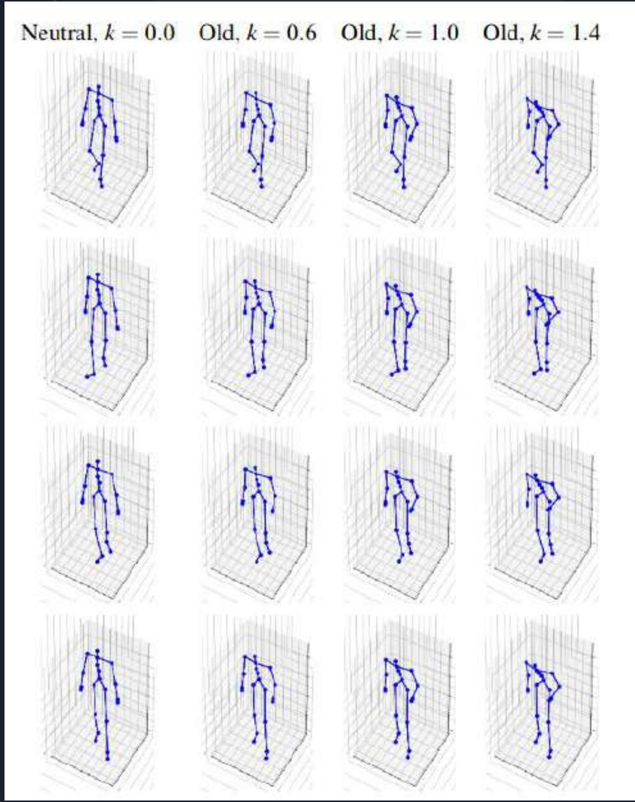


Style Tuning : Motion Generation and Style Transfer in Variational Autoencoder Systems for Human Motion Synthesis (Ian MASON) 2017, Thesis

Model	Mean Reconstruction Error
M2_enc2	0.0235804
M2_enc1_enc2	0.016751
M2_enc2_dec	0.0143762
M2_enc1_enc2_dec	0.0155206
M2_dec	0.0159735
M1+M2_enc	0.0161324
M1+M2_dec	0.0697627

Performing a simple linear interpolation between output motions to create a form of continuous style transfer

$$\widetilde{\mathbf{x}}_{1:T} * (1-k) + \mathbf{a}_{1:T} * k.$$

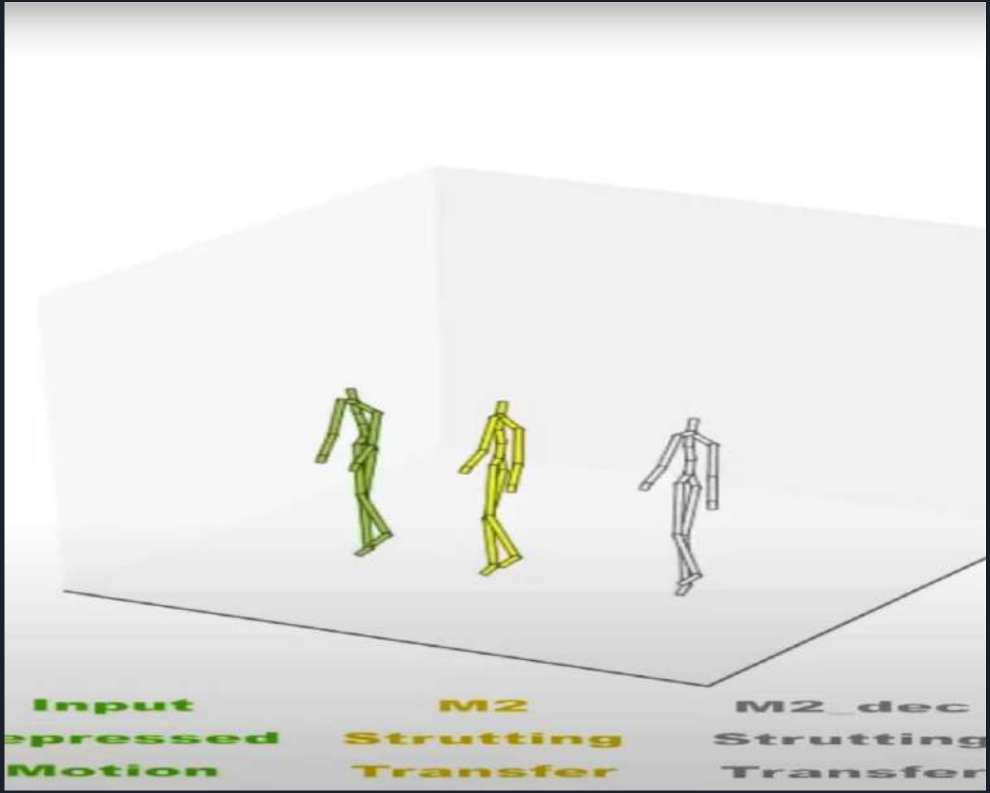


Style Tuning : Motion Generation and Style Transfer in Variational Autoencoder Systems for Human Motion Synthesis (Ian MASON) 2017, Thesis

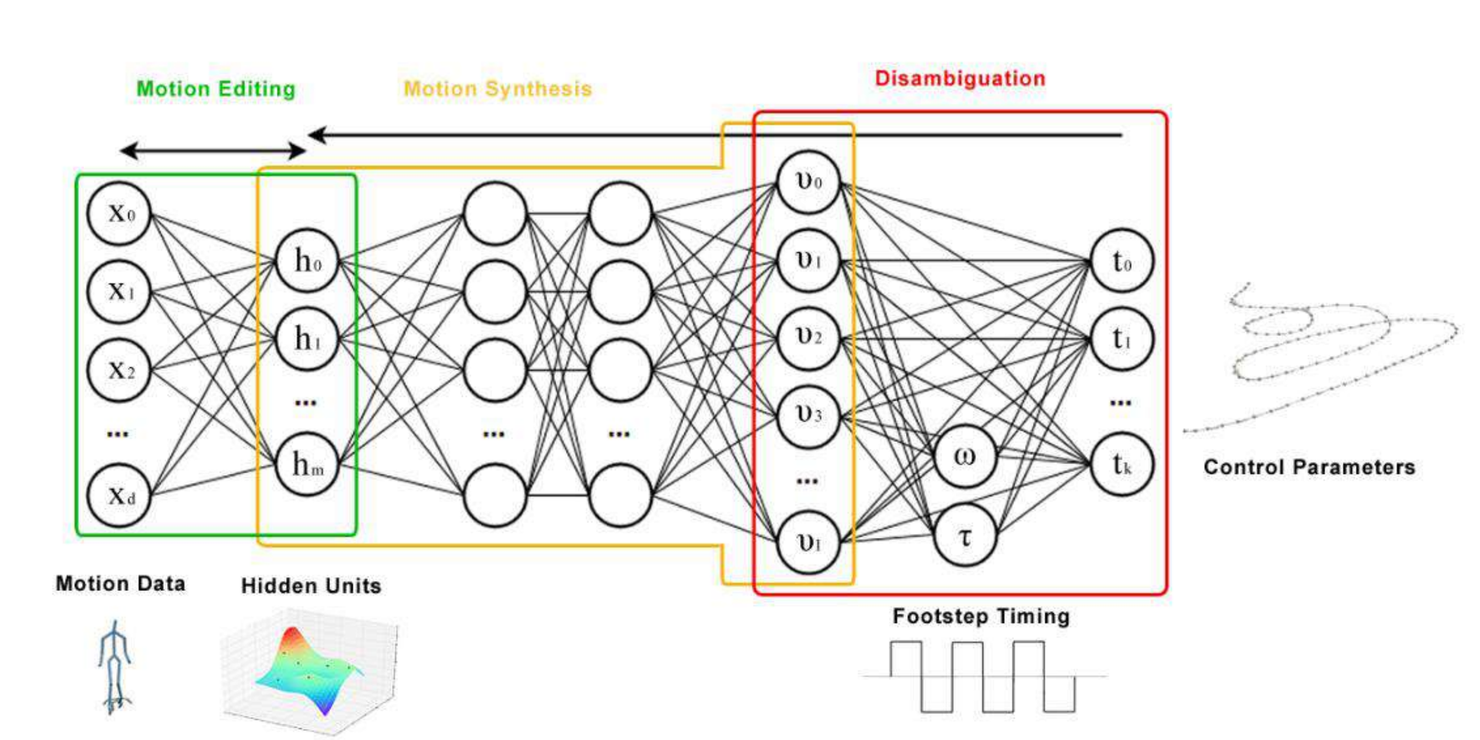
- I received code of the Motion Generation part from Habibee , but the style transfer code is still missing .

- It Would be a good idea to implement the style transfer missing part to test this approach on our dataset

Style Transfer: Deep learning framework for character motion synthesis and editing Daniel HOLDEN et al., 2016

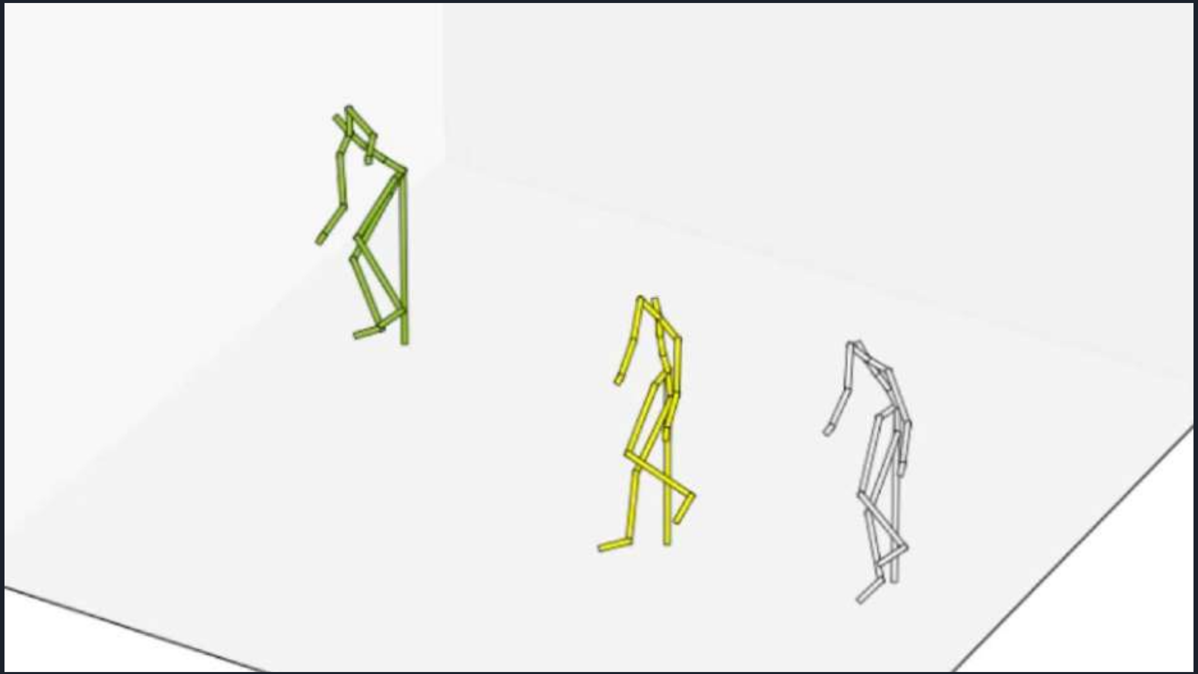


Style Transfer: Deep learning framework for character motion synthesis and editing,Daniel HOLDEN(2016)



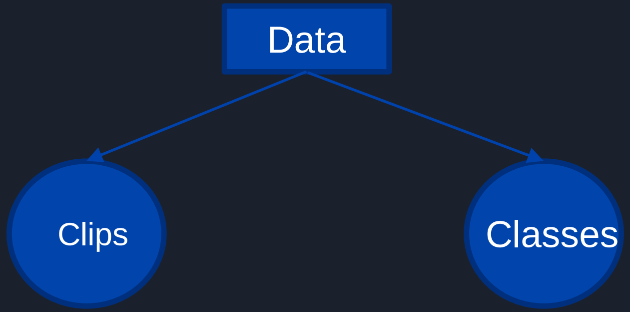
Style Transfer : Deep learning framework for character motion synthesis and editing,Daniel HOLDEN, 2016

Testing the model with Holden dataset:



Style Transfer : Deep learning framework for character motion synthesis and editing,Daniel HOLDEN (2016)

- Converting data from txt to npz



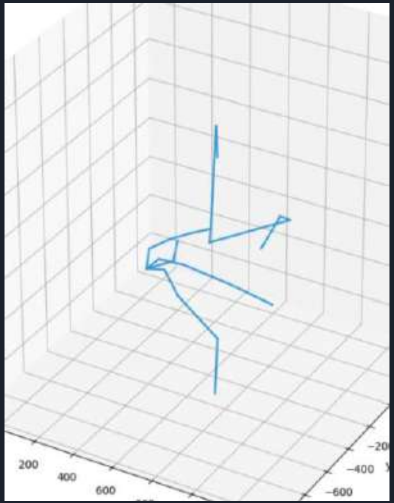
- BVH format
- Retargeting the dataset to a unifed BVH structure

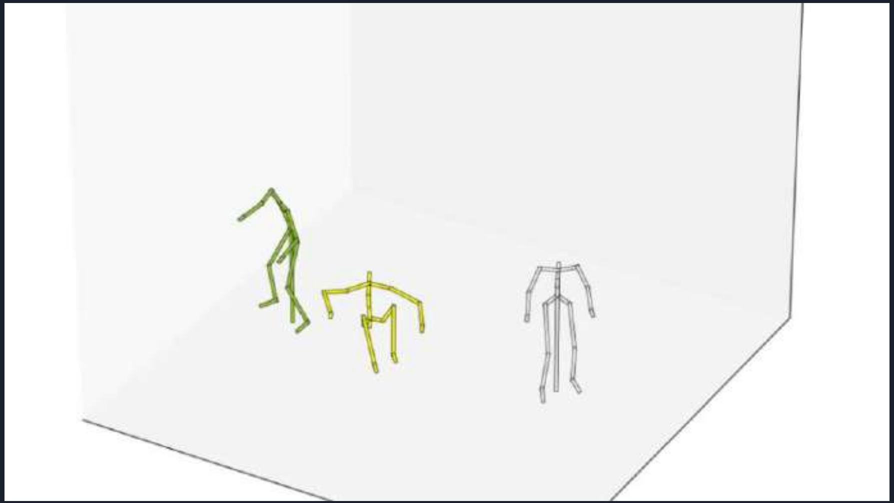
```
             0           1            2
styles = ['ES', 'Neurotic', 'Normal']
motions = ['waving', 'stopping', 'showingphone', 'showing', 'hiding']
              0           1              2              3          4
```

```
[0, 0] [0, 0] [0, 0] [0, 0] [0, 0] [0, 0] [1, 0] [1, 0] [1, 0] [1, 0] [1, 0]
[1, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [2, 0] [3, 0] [3, 0]
[4, 0] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [0, 1] [1, 1] [1, 1]
[1, 1] [1, 1] [1, 1] [1, 1] [2, 1] [2, 1] [2, 1] [2, 1] [2, 1] [3, 1] [3, 1]
[4, 1] [0, 2] [0, 2] [0, 2] [0, 2] [0, 2] [0, 2] [1, 2] [1, 2] [1, 2] [1, 2]
[1, 2] [1, 2] [2, 2] [2, 2] [2, 2] [2, 2] [3, 2] [3, 2] [4, 2]
```

Style Transfer : Deep learning framework for character motion synthesis and editing, Daniel HOLDEN (2016)

Results:





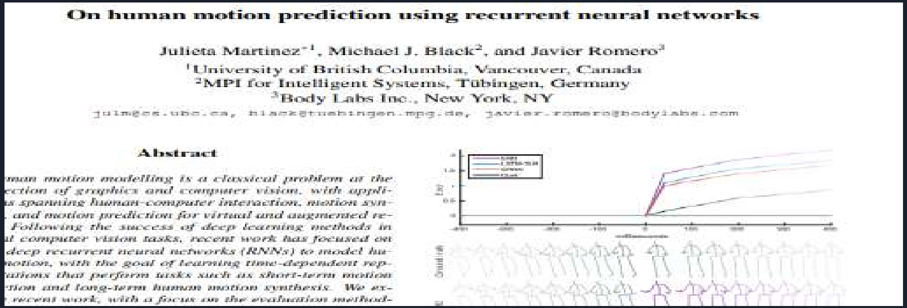
- Most of the recent work used a periodic and locomotion dataset which is not what we are working with .

Conclusion:

- Improving dataset :
  - Overlapping chunks
  - Data mirroring
  - Converting from and to different MOCAP format
  - Retargeting data to an unified BVH structure
- Motion Synthesis:
  - Adding a VAE to make to the model learn from style
  - Style labeling
- Style transfer & Style tuning
  - Investigating and testing two different approaches

Further Work

- Back to the 1st approach : Human motion prediction using recurrent neural networks , Martinez et al. [cs.CV] (2017)

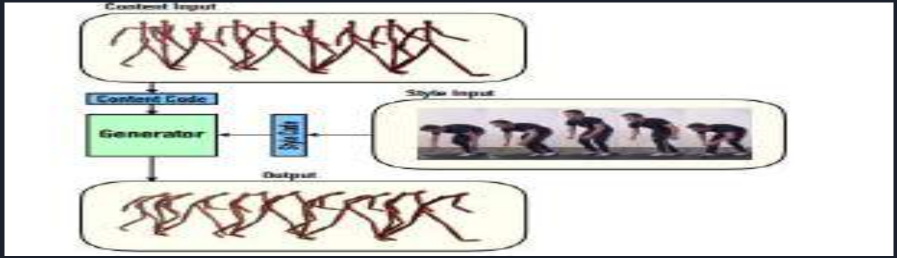

On human motion prediction using recurrent neural networks

Julieta Martinez*[1], Michael J. Black[2], and Javier Romero[3]
[1]University of British Columbia, Vancouver, Canada
[2]MPI for Intelligent Systems, Tübingen, Germany
[3]Body Labs Inc., New York, NY
julm@cs.ubc.ca, black@tuebingen.mpg.de, javier.romero@bodylabs.com

Abstract

man motion modelling is a classical problem at the
ection of graphics and computer vision, with appli-
s spanning human-computer interaction, motion syn-
, and motion prediction for virtual and augmented re-
. Following the success of deep learning methods in
al computer vision tasks, recent work has focused on
deep recurrent neural networks (RNNs) to model hu-
motion, with the goal of learning time-dependent rep-
tations that perform tasks such as short-term motion
ction and long-term human motion synthesis. We ex-
recent work, with a focus on the evaluation method-

- Unpaired motion style transfer from video to animation, Aberman et al. ACM Transactions on Graphics (2020)



- Representing motion as a sequence of latent primitives, a flexible approach for human motion modelling, Mathieu Marsot et al. [cs.CV] 1 (2022)
- A Structured Latent Space for Human Body Motion Generation, Mathieu Marsot et al. [cs.CV] (2022)

References:

- Habibie, I., Holden, D., Shwarz, J., Yearsley, J., and Komura, T. (2017). A recurrent variational autoencoder for human motion synthesis. In BMVC.
- Holden, D., Saito, J., and Komura, T. (2016). A deep learning framework for character motion synthesis and editing. ACM Trans. Graph., 35(4)
- Ian Mason, Sebastian Starke, and Taku Komura. 2022. Real-Time Style Modelling of Human Locomotion via Feature-Wise Transformations and Local Motion Phases. arXiv preprint arXiv:2201.04439 (2022).
- On human motion prediction using recurrent neural networks Julieta Martinez∗1 , Michael J. Black2 , and Javier Romero3. arXiv:1705.02445v1 [cs.CV] 6 May (2017)
- Unpaired motion style transfer from video to animation, Aberman et al. ACM Transactions on Graphics (2020)