# Dance Machine

DSR project (batch 27)
Oct.2021

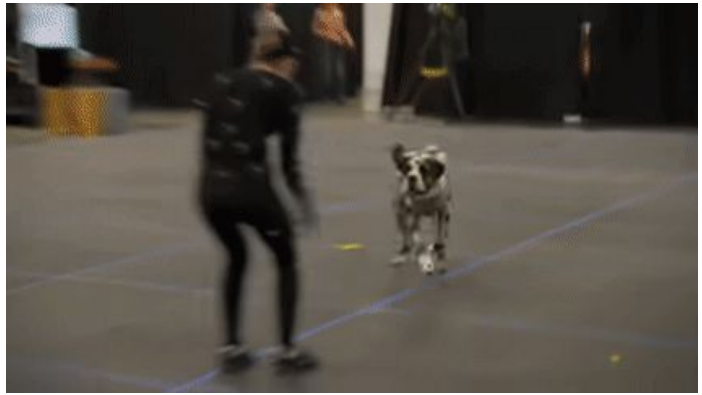
Mohammad Farahani, and Mohammad Saleh

# Outline:

- About dance learning
- Challenges
- Approach
  - Dataset
  - Modelling
  - Results
- Conclusion

# Why?

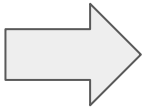
- General movement generation
  - Game industry
  - Animation industry
  - Customized choreography

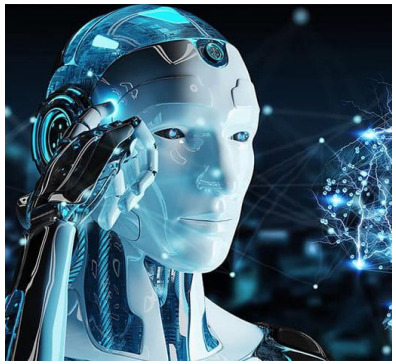


# How humans learn dancing?

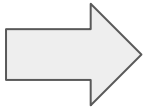
Can machine learn dancing as well?

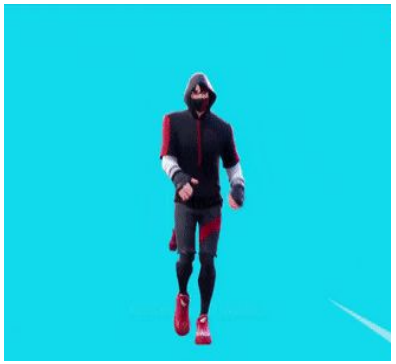
Real dances





Machine learning





Generated dances

# Generative methods

- Text generation: e.g. GPT-2

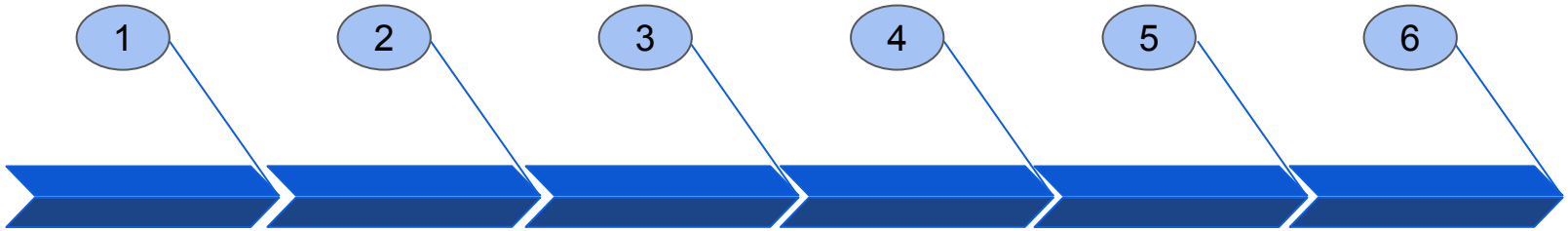

- Image generation: e.g. GANS

- What about movements ?

# Challenges

1. Human body movement detection.
2. Training a model with detected data.
3. Generate new movements.

# The approach




| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| **Gathering data** | **Pose detection** | **Loading data** | **Training Model** | **Future moves generation** | **Dance visualisation** |
| Gathering solo dancing about 10 hours. | Breaking down videos into poses and frames per second. | Loading X,Y,Z points for every joint in every frame into Tabular data. | Choosing model and training it on the gathered data | Generation of future coordinates in a specified period of time. | Loading these coordinates into a moving animation to evaluate. |

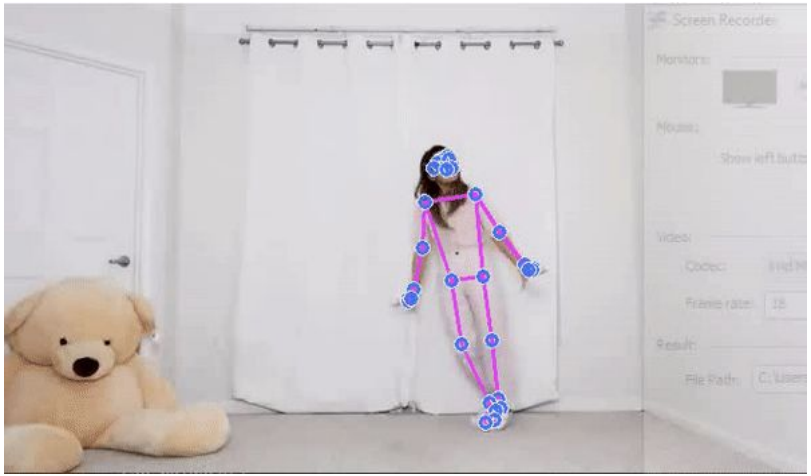
# Gathering data

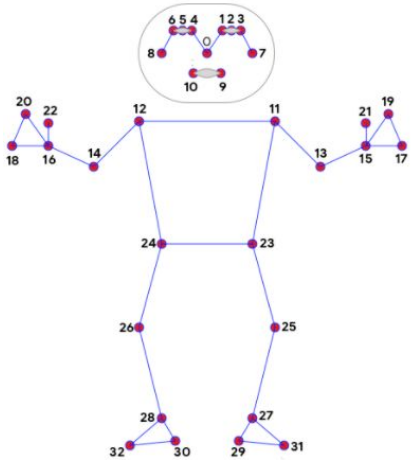


- Collection about 10 hours of hip hop dance.

Videos link: https://www.youtube.com/c/imlisarhee

# Pose detection: Mediapipe






0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

- Detect human pose landmarks
- 33 points, each point has X, Y, Z coordinates

# Raw data

| | x0 | y0 | z0 | x1 | y1 | z1 | x2 | y2 | z2 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.421101 | 0.550777 | -0.022540 | 0.410861 | 0.550942 | -0.018178 | 0.411158 | 0.549630 | -0.018285 |
| 1 | 0.420971 | 0.549533 | -0.039764 | 0.410825 | 0.549494 | -0.032597 | 0.411138 | 0.546786 | -0.032689 |
| 2 | 0.427683 | 0.534817 | -0.023420 | 0.419265 | 0.540632 | -0.016715 | 0.419211 | 0.539021 | -0.016817 |
| 3 | 0.426893 | 0.536704 | -0.023477 | 0.418657 | 0.540746 | -0.016965 | 0.418610 | 0.539225 | -0.017064 |
| 4 | 0.426909 | 0.498420 | -0.025411 | 0.418803 | 0.505658 | -0.018127 | 0.418611 | 0.506281 | -0.018222 |

# Input data

Past frames

| | x0 | y0 | z0 | x1 | y1 | z1 | x2 | y2 | z2 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.421101 | 0.550777 | -0.022540 | 0.410861 | 0.550942 | -0.018178 | 0.411158 | 0.549630 | -0.018285 |
| 1 | 0.420971 | 0.549533 | -0.039764 | 0.410825 | 0.549494 | -0.032597 | 0.411138 | 0.546786 | -0.032689 |
| 2 | 0.427683 | 0.534817 | -0.023420 | 0.419265 | 0.540632 | -0.016715 | 0.419211 | 0.539021 | -0.016817 |
| 3 | 0.426893 | 0.536704 | -0.023477 | 0.418657 | 0.540746 | -0.016965 | 0.418610 | 0.539225 | -0.017064 |

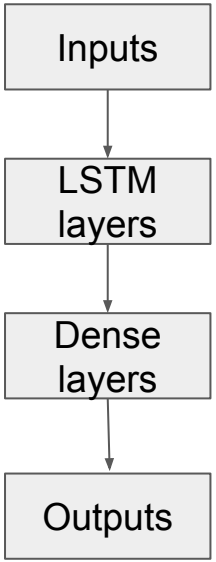
# Output data

Future frames

| | x0 | y0 | z0 | x1 | y1 | z1 | x2 | y2 | z2 |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 0.426909 | 0.498420 | -0.025411 | 0.418803 | 0.505658 | -0.018127 | 0.418611 | 0.506281 | -0.018222 |

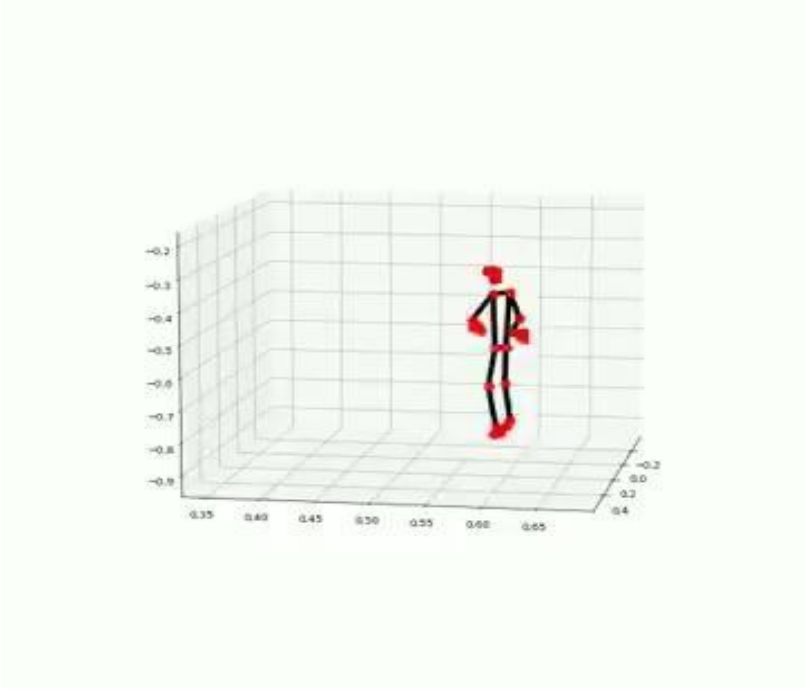
# Modeling

## Multivariate Time Series Using LSTM



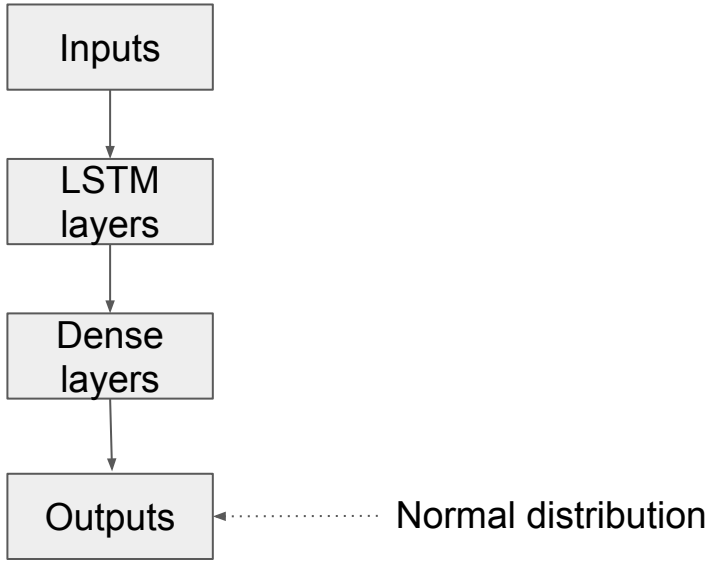
# Results

initial




Generation

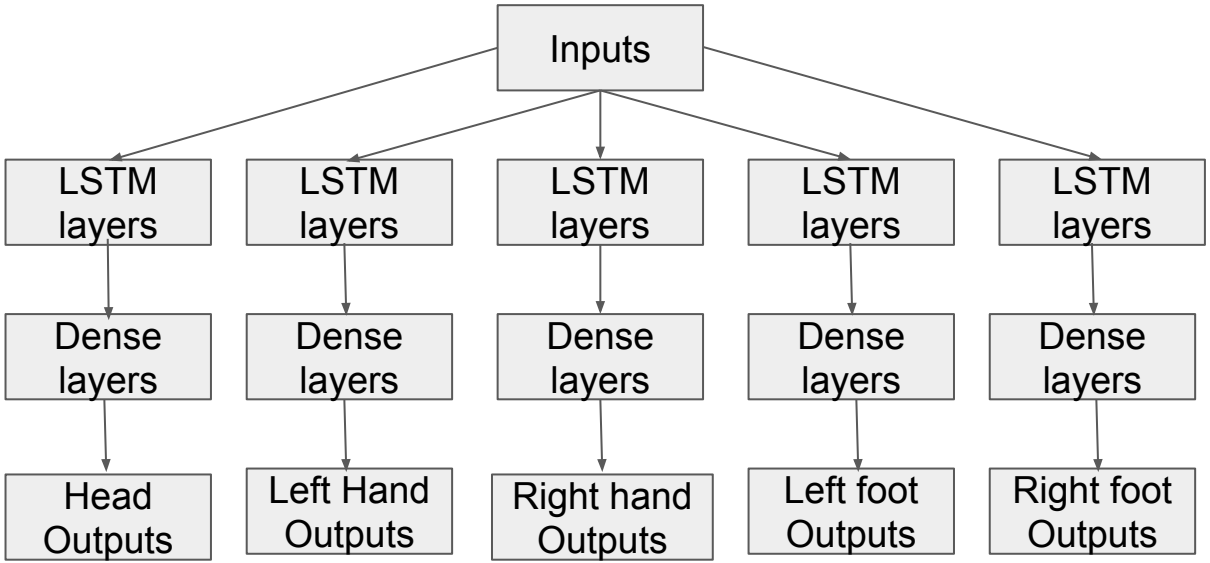
# Modeling




| | x0 | y0 | z0 | x1 | y1 | z1 | x2 | y2 | z2 |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 0.426909 | 0.498420 | -0.025411 | 0.418803 | 0.505658 | -0.018127 | 0.418611 | 0.506281 | -0.018222 |

# Results

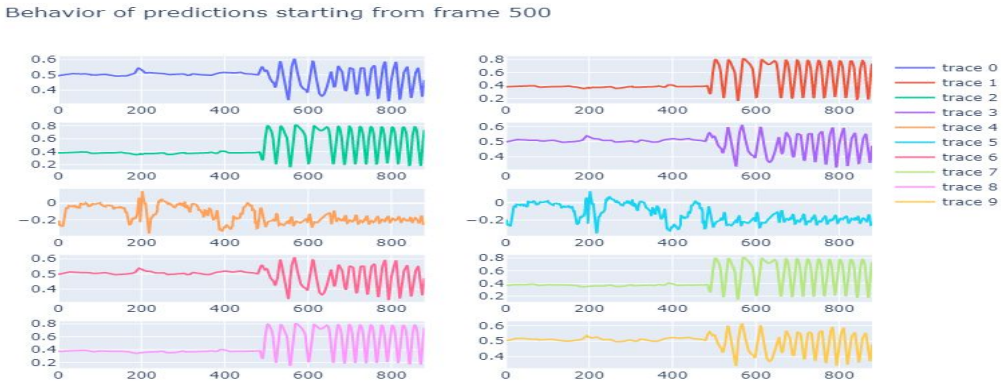
# Modeling


Inputs
LSTM layers
LSTM layers
LSTM layers
LSTM layers
LSTM layers
Dense layers
Dense layers
Dense layers
Dense layers
Dense layers
Head Outputs
Left Hand Outputs
Right hand Outputs
Left foot Outputs
Right foot Outputs

# Results

# Conclusion and Outlook

- The multipart model seems promising.
  - Use different type of layers for each part
- Different dance styles data
- Define key frames
- Use other model e.g. TCN



# Acknowledgement


- Dr. Tristan Behrens (our mentor)
- DSR Teachers
- DSR Batch 27

# Thank you for your attention!

Mohammad Farahani
linkedin.com/in/**mohfarahani**/

Mohammad Saleh
linkedin.com/in/**mohammadSalehSaudi**