Master seminar: Solving localization problem in first person computer games with deep learning

Yauheni Selivonchyk[1],
Prof. Dr. Christian Bauckhage[2],
Prof. Dr. Stefan Wrobel[2],
Mr. Sc. Rafet Sifa[2]

[1]Institute of Computer Science, University of Bonn

[2]Fraunhofer IAIS

April 26, 2017

Outline

Recent progress in AI

Some of the recent advances in machine learning:

- Image classification
- Machine translation

Recent progress in AI

Some of the recent advances in machine learning:

- Image classification
- Machine translation

Important attributes of it:

- Low cost and high processing abilities of modern computer chips
- Advances in machine learning algorithms
- Access to large labeled datasets

Importance of unsupervised learning for AI

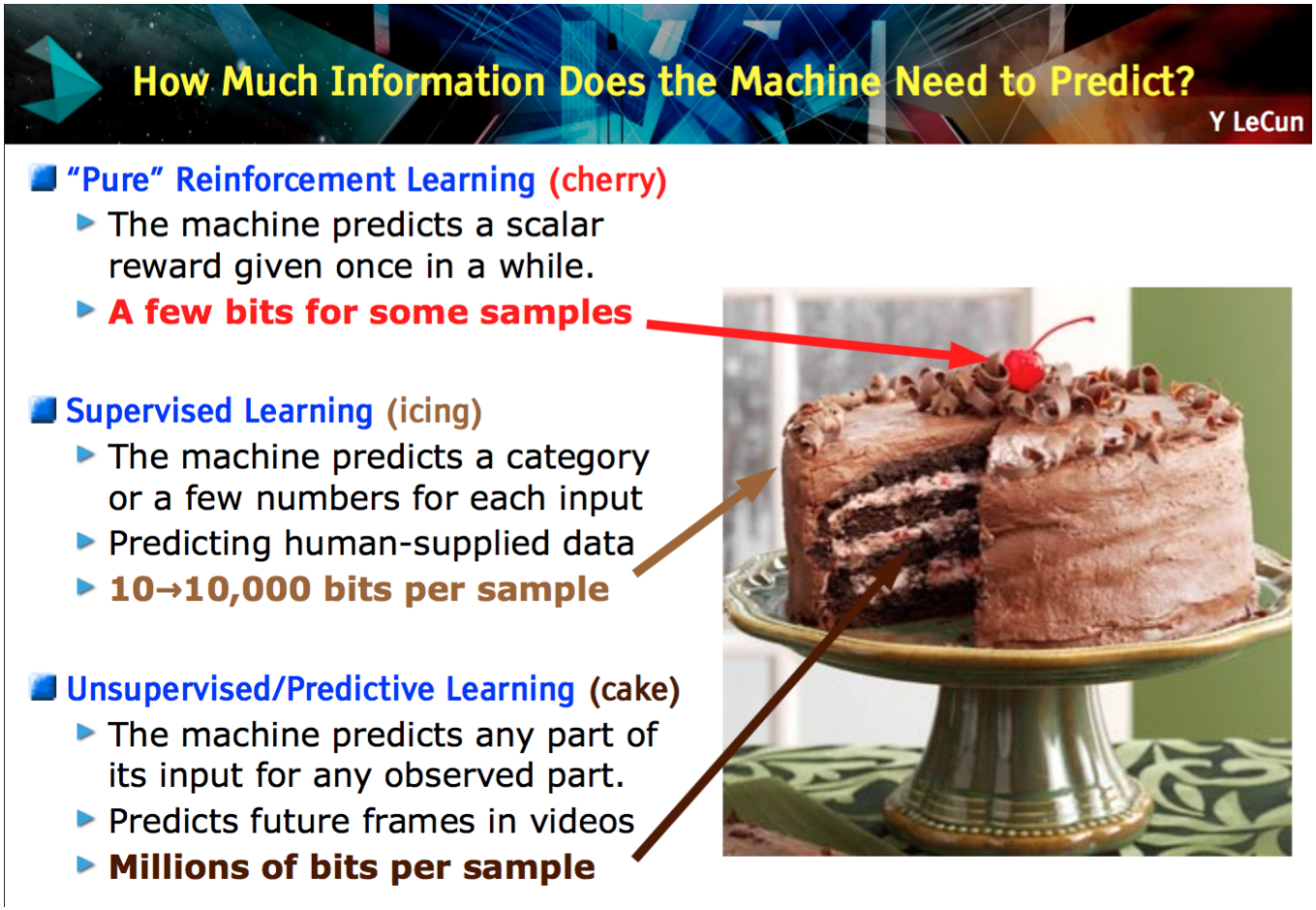


Figure: Slide from "Predictive learning" opening address given by Yann Lecun at NIPS2016.

Recent advances in Unsupervised learning

Some of the recent influential models:

- Word embedding (T. Mikolow, 2013)
- Variational autoencoders (D. Kingma, 2013)
- Generative Adversarial Networks (I. Goodfellow, 2014)

Localization

Localization as a task of extracting, tracking or predicting object's position in some environment from available sensory data.

Localization

Localization as a task of extracting, tracking or predicting object's position in some environment from available sensory data.

Types of data:

- Visual data: images or video sequences
- Depth map
- Information about position/direction of the sensors
- etc.

Localization example: Tracking

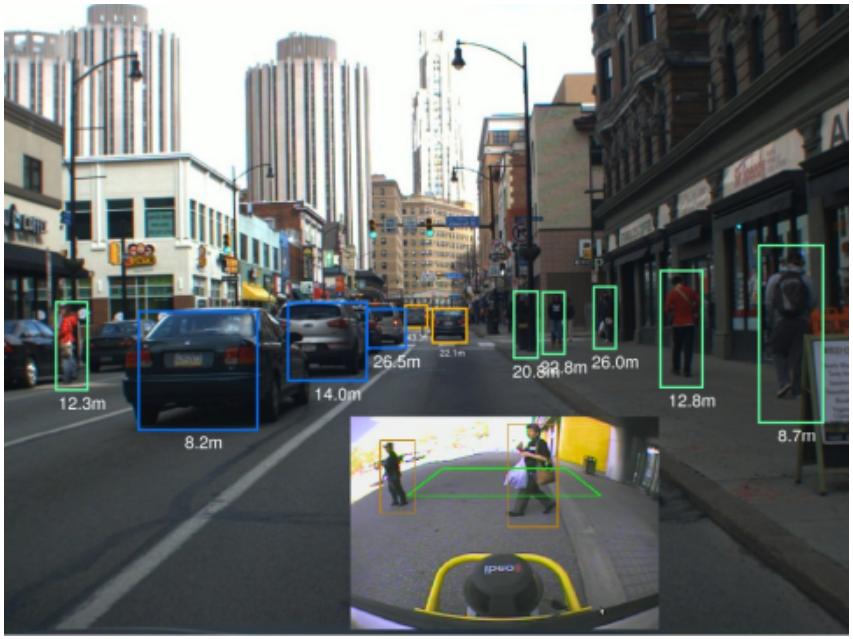


Figure: Pedestrian tracking visualization [1].

[1] H. Cho et. al. "Real-Time Pedestrian and Vehicle Detection for Automotive Active Safety Systems"

Localization example: SLAM

Figure: Example solution of SLAM problem on PC3 dataset (courtesy of University of Michigan).

Localization example: surgery

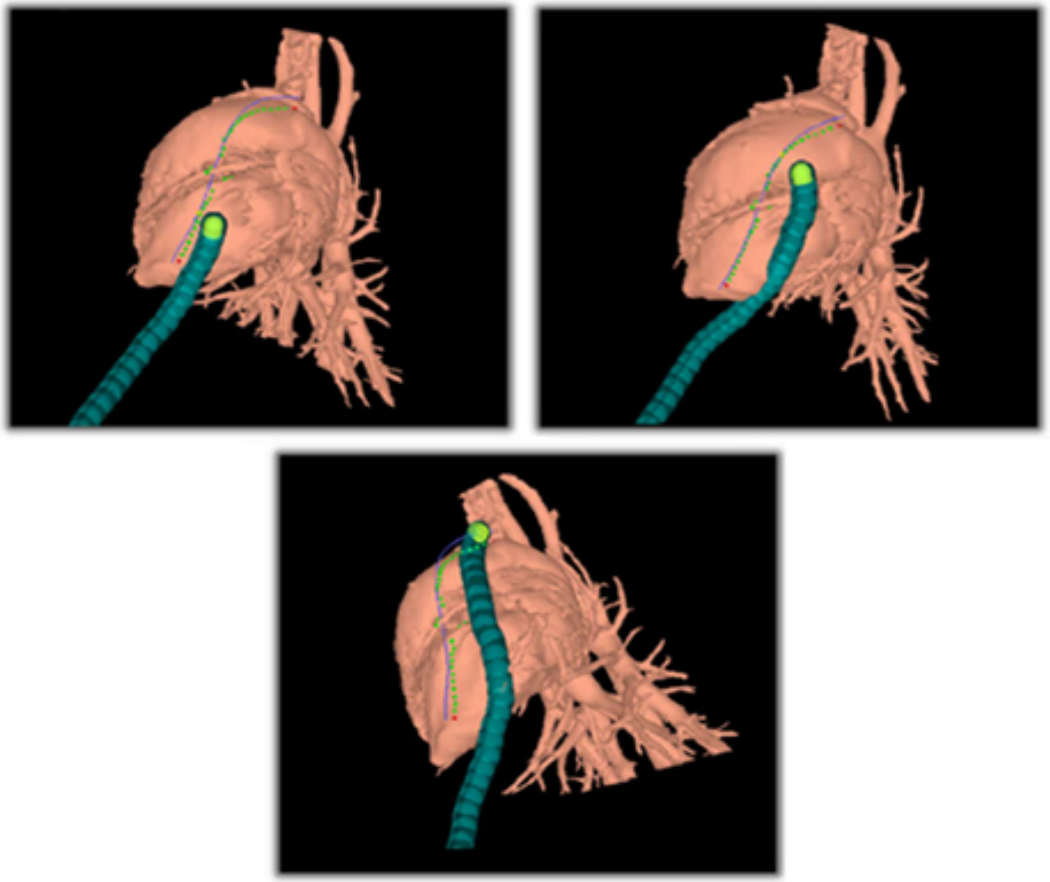


Figure: Mapping the position of a tool in minimally invasive surgery [http://biorobotics.ri.cmu.edu/research/medicalSLAM.html].

Motivation. Continuied

Goal of this work: extracting interpretable low-dimensional representation of players movements in first-person shooter (games) from visual (video) data.

Motivation. Continuied

Goal of this work: extracting interpretable low-dimensional representation of players movements in first-person shooter (games) from visual (video) data.

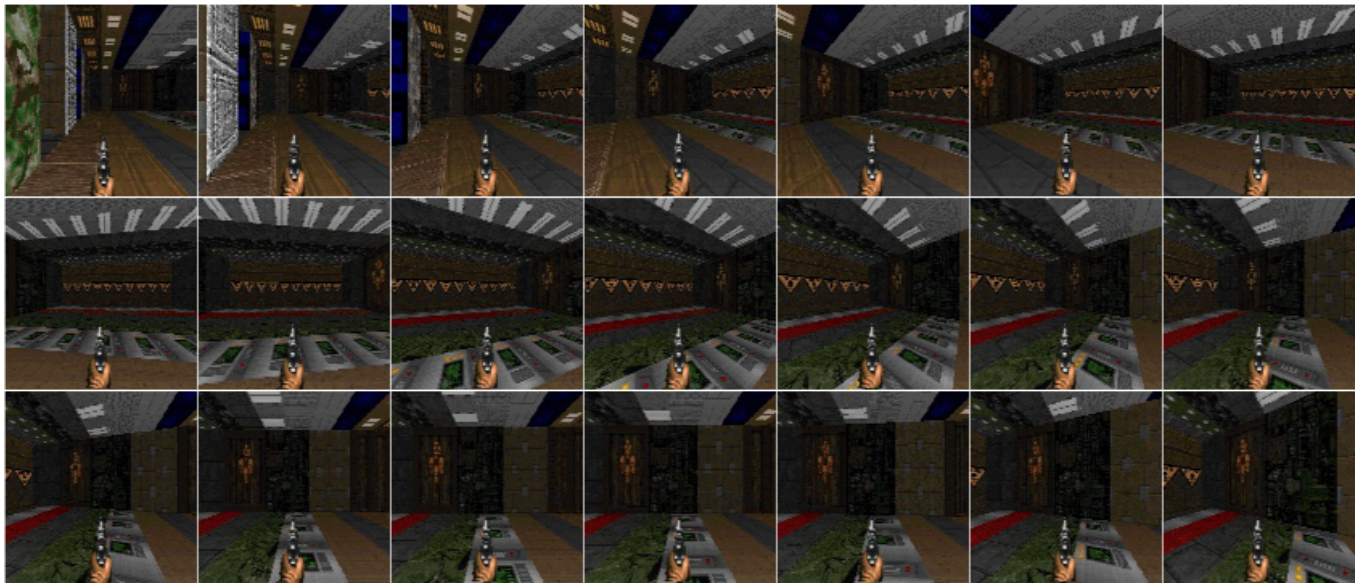
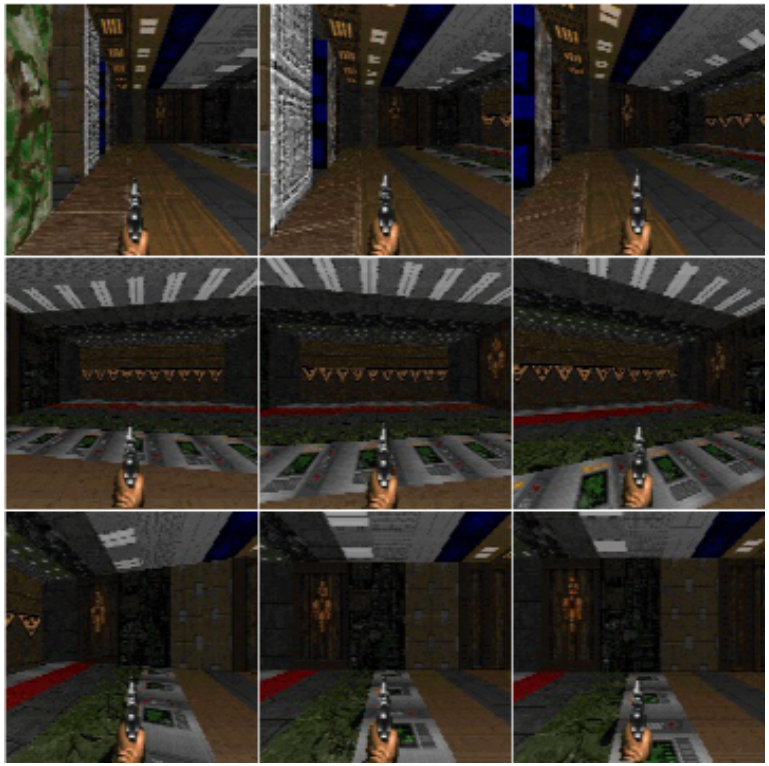


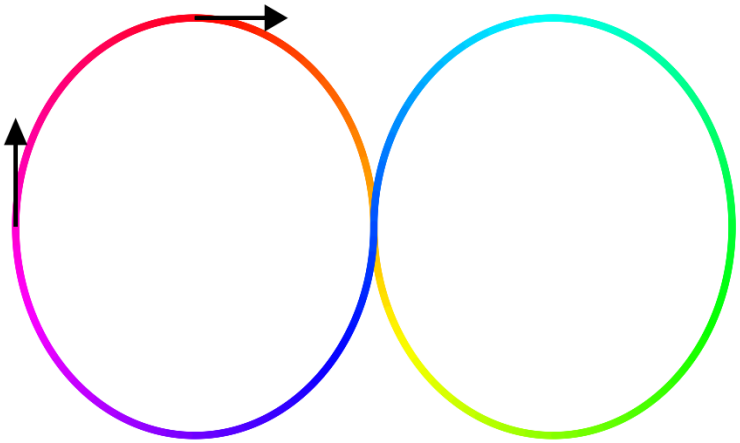
Figure: Example visual data.

Autoencoder model

Input video information:

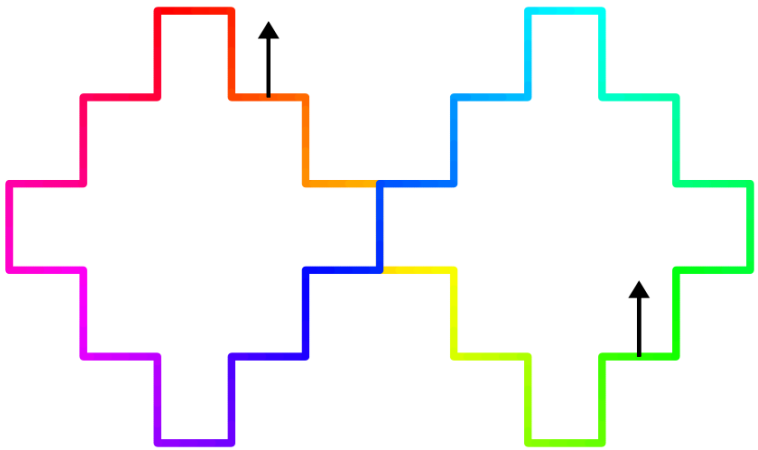


Corresponding player's path:

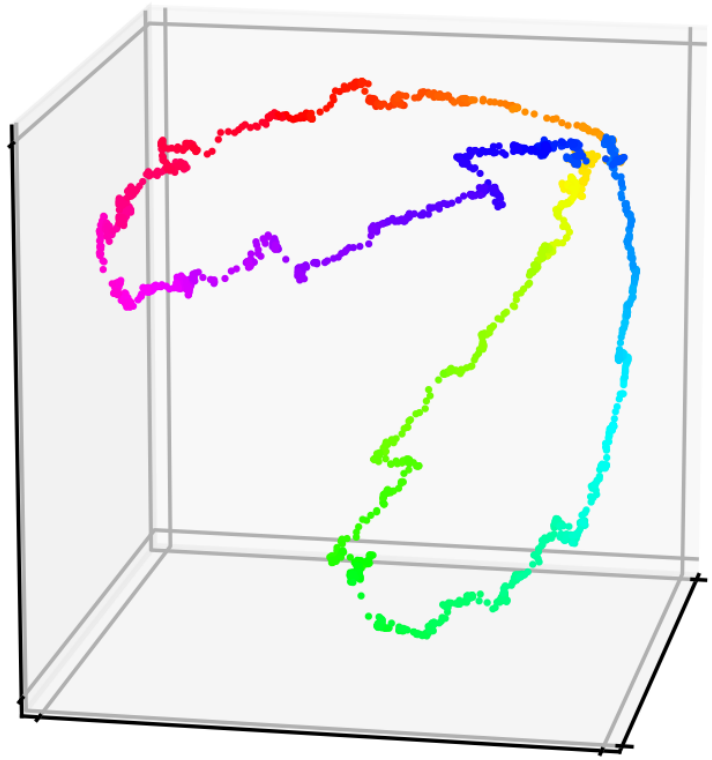


Example reconstructions

Original trajectory on the video:



Our embedding in a 3D space:

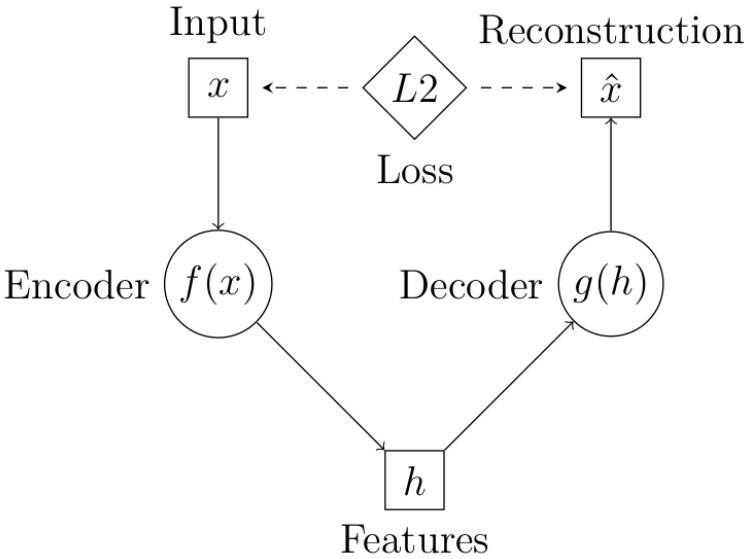


Autoencoder model

Autoencoders learn to project the input x into some embedding space $h \in H$ and simultaneously reconstruct the original information $\hat{x}$.

Autoencoder model

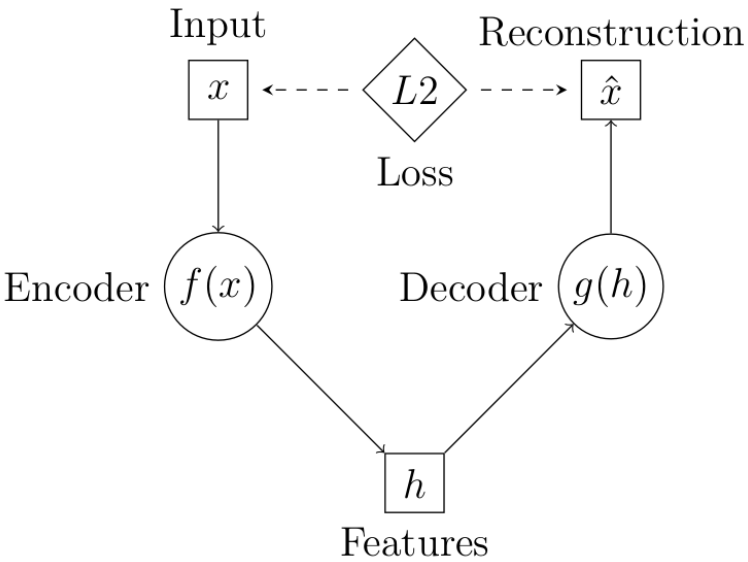
Autoencoders learn to project the input x into some embedding space $h \in H$ and simultaneously reconstruct the original information $\hat{x}$.



- $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$
- $g : \mathbb{R}^M \rightarrow \mathbb{R}^N$

Autoencoder. Primary goal

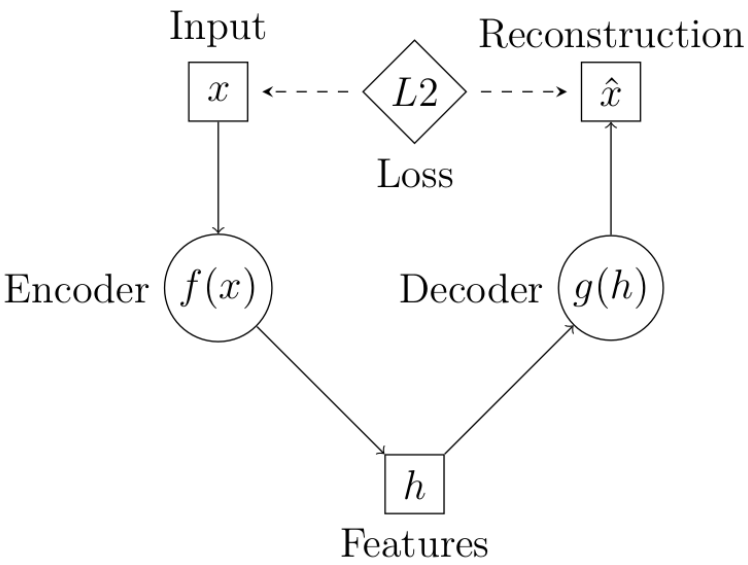
While good quality reconstruction is desirable, we will focus on producing high quality representation in the embedding space H.

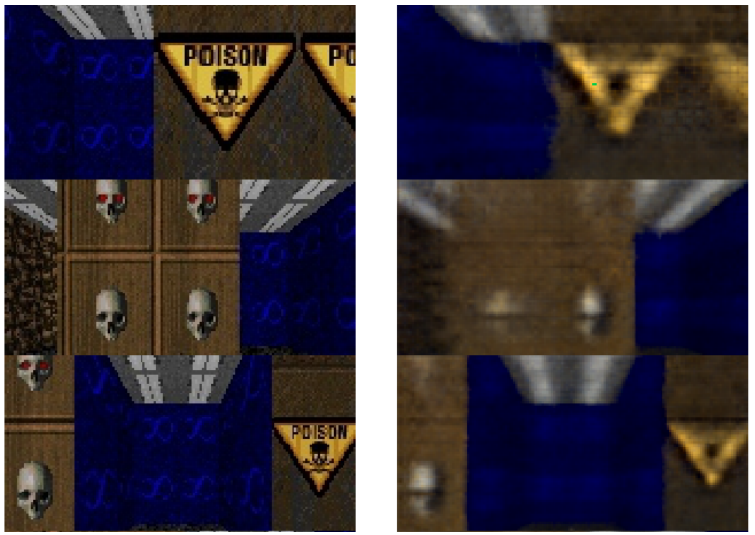


Original images and their reconstruction

Autoencoder. Primary goal

While good quality reconstruction is desirable, we will focus on producing high quality representation in the embedding space H.





Our contributions. Training process

- We propose a robust training technique that allows unsupervised visual data embedding into extremely low dimensional space.
  Our model requires compression ratio of 10000:1, but allows information loss.
- We describe regularization techniques that allow to preserve spatial relations in the embedding space

Data collection

We use VizDoom scientific research platform, which is based on a 3D engine of FPS game DoomII.

Figure: Visual data in DoomII computer game.

Data collection. Continued

We collect several trajectories for our evaluation:

- Trivial trajectories: showing simple transitions as walking along the axis with limited degress of freedom.
- Trajectories of natural, yet, predictable movement. As running in an *eight*.
- Random movement on the map.

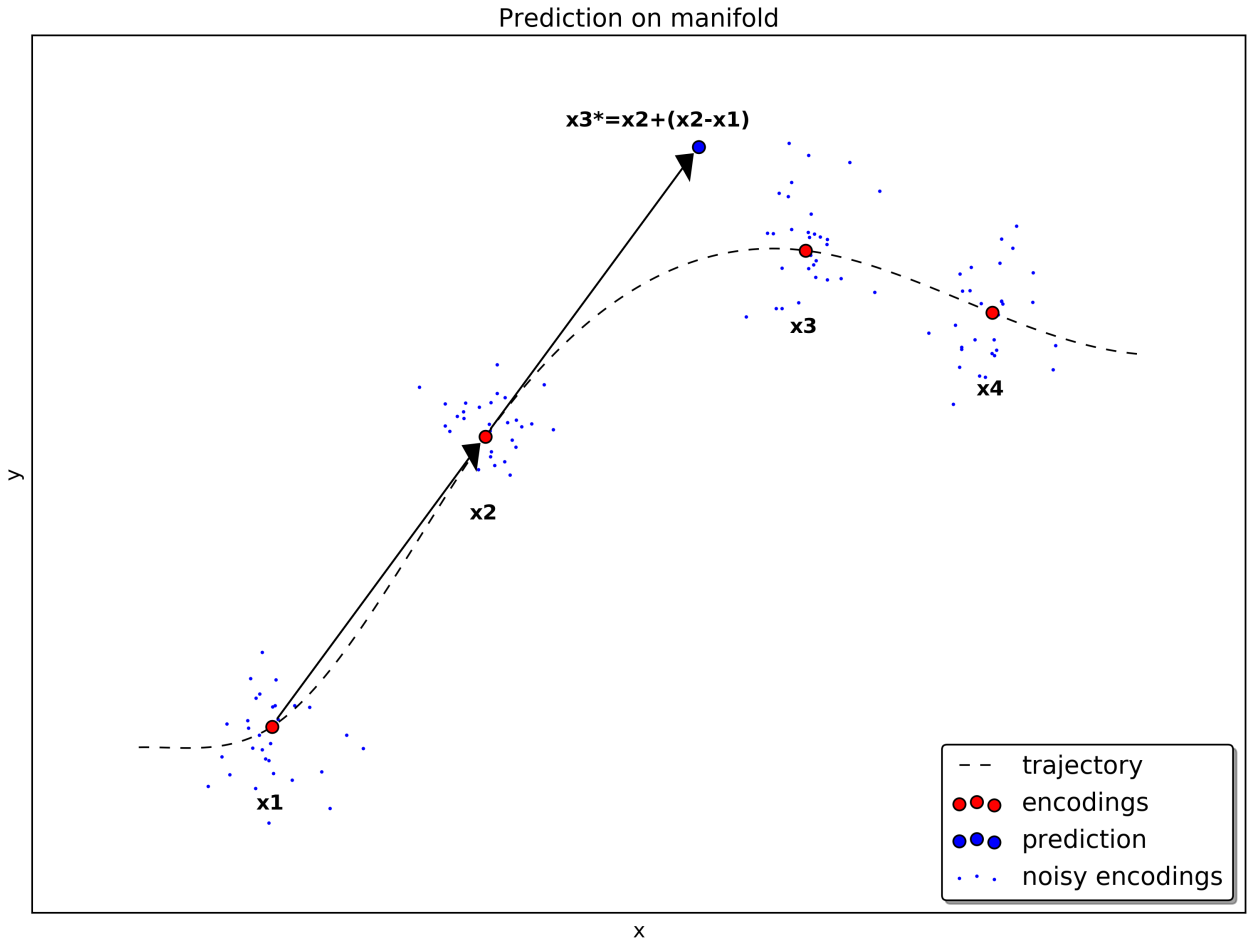
Our goals

Qualities of a good spatial embedding:

- It is visually interpretable;
- It preserves spatial structure of the trajectories in Euclidean space:
  - Key trajectory elements: turns, intersections
  - Point-to-point relations
- It allows prediction of the future frames.

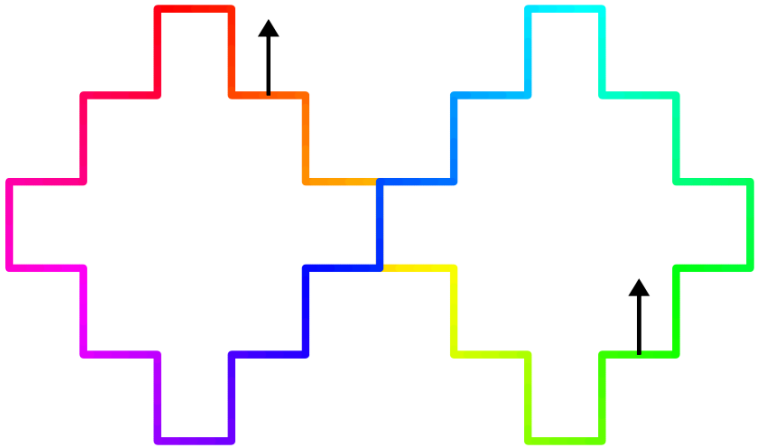
Prediction in the embedding space

Try to estimate positional encoding of the next frame using last two frames of the video.

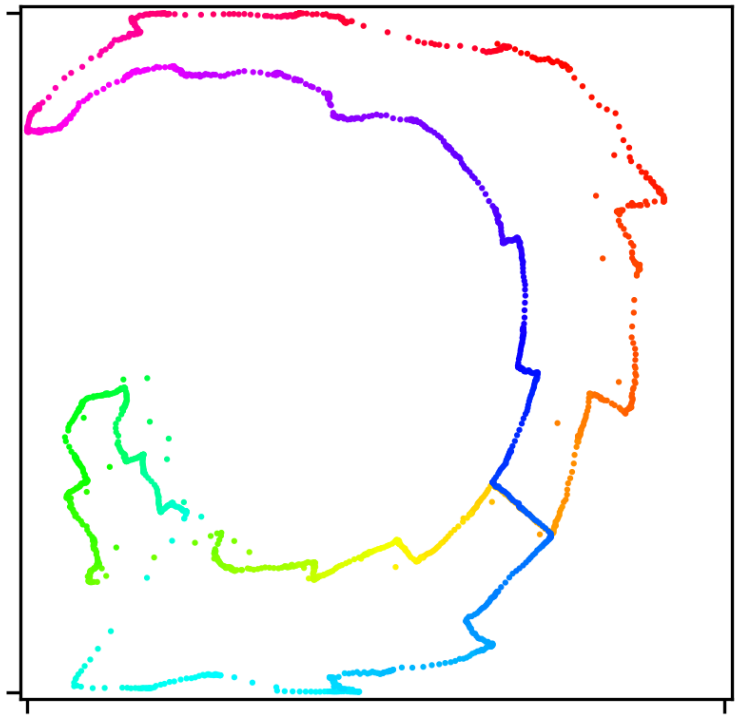


Trivial trajectory

Original trajectory on the video:

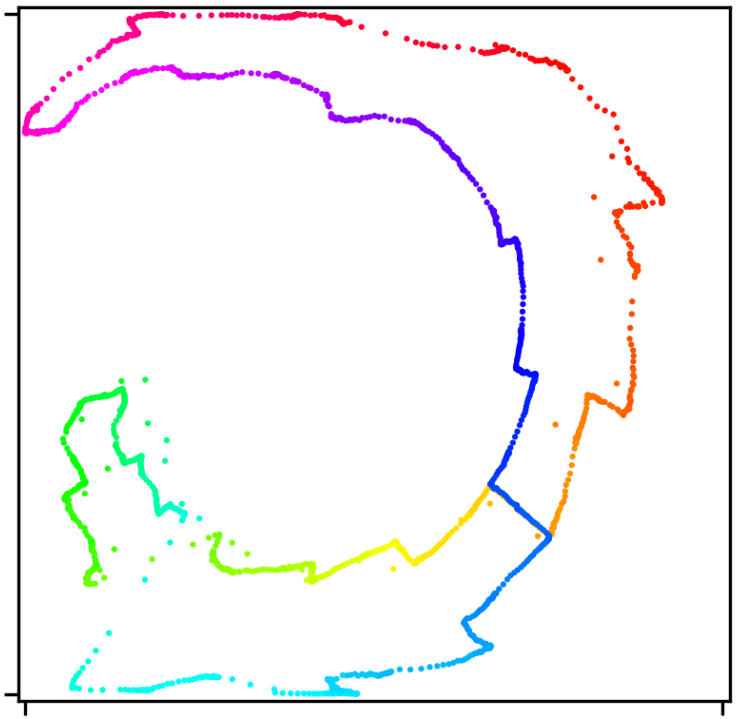


Our embedding in a 3D space:



Trivial trajectory. Continued

Trajectory:

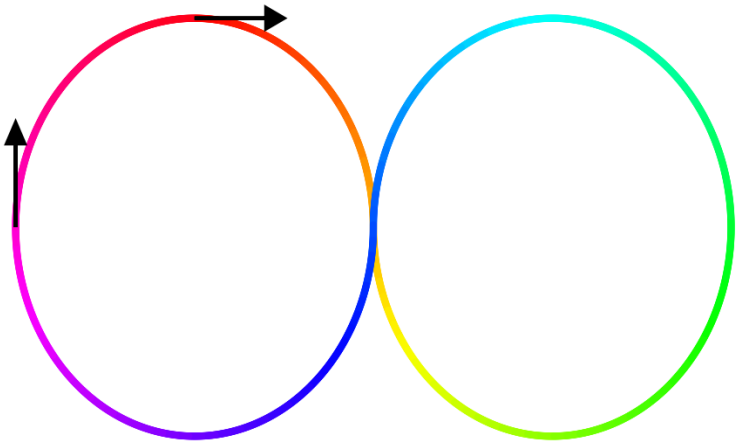


Key characteristics

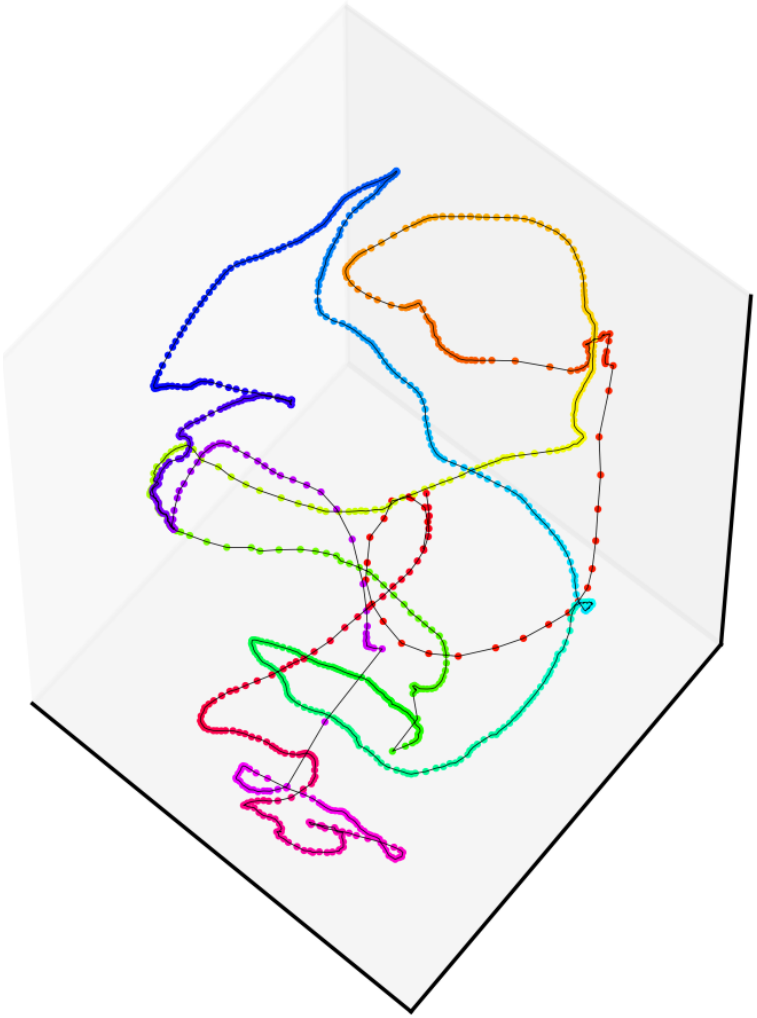
- Trajectory is visually recognizable
- Subsequent frames remain nearest neighbors 78% of the time
- Predicted frames are on average 37% closer to actual embedding comparing to current frame

Natural trajectory

Original trajectory on the video:

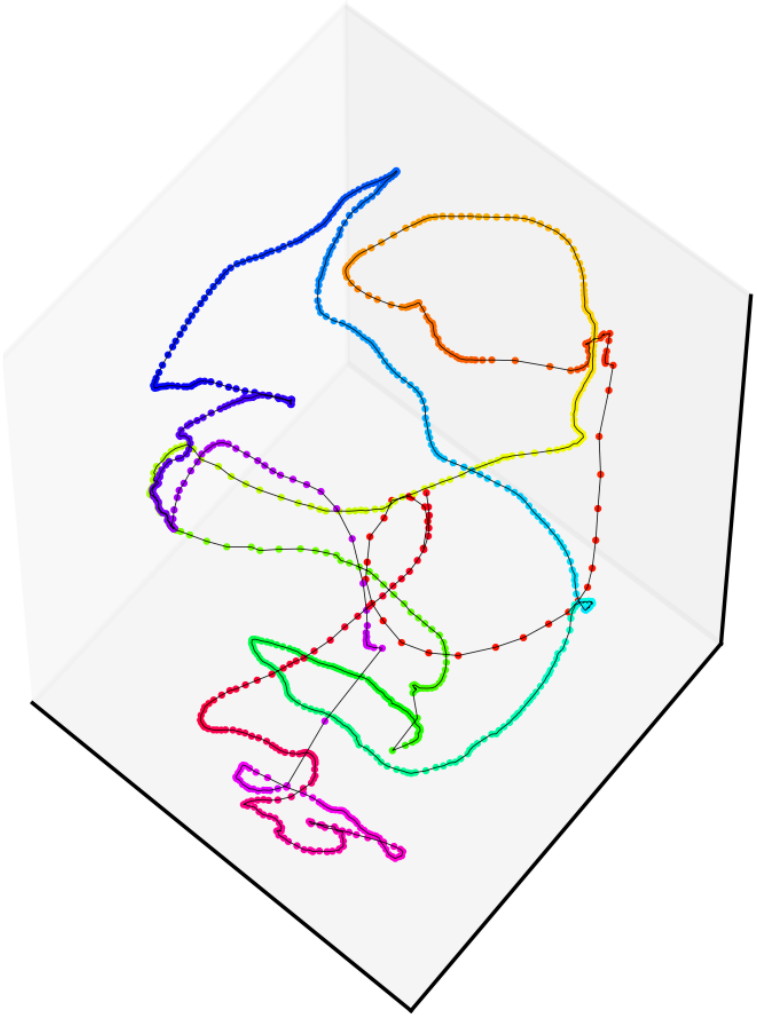


Our embedding in a 3D space:



Natural trajectory. Continued

Trajectory embedding:

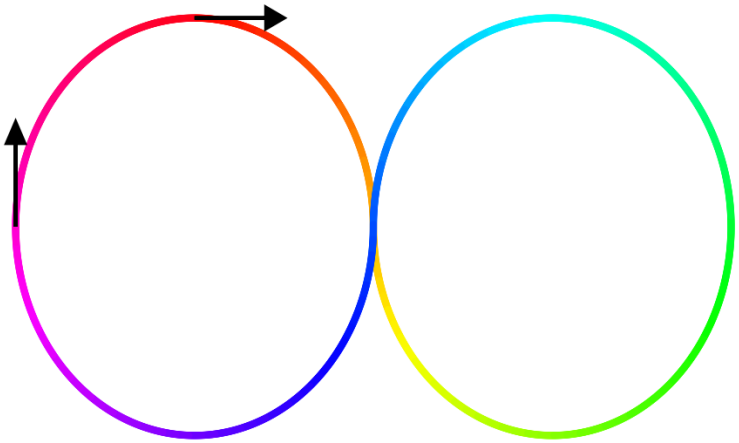


Key characteristics

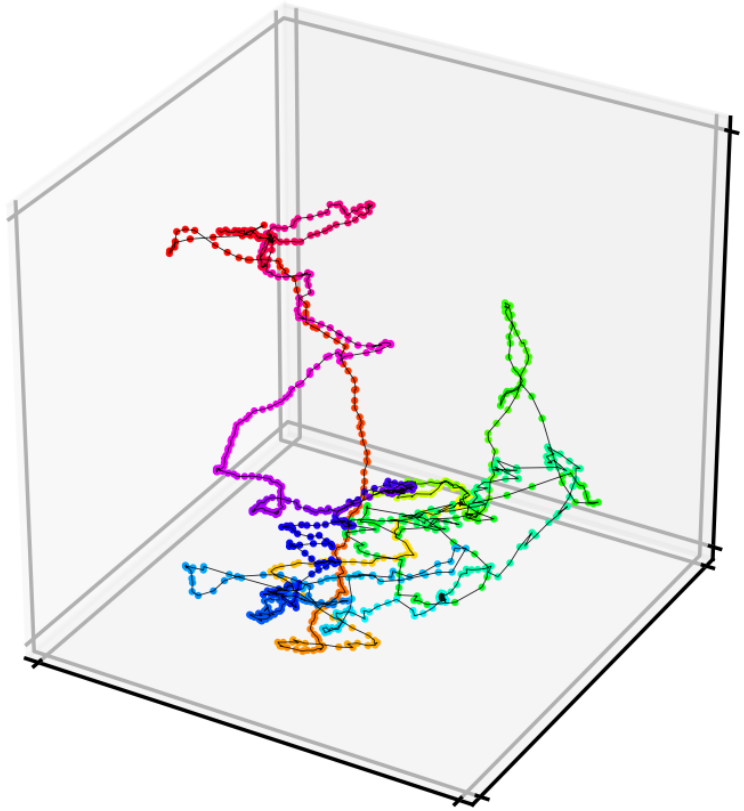
- Trajectory is hard to recognize
- Subsequent frames remain nearest neighbors 91% of the time
- Predicted frames are on average 58% closer to actual embedding comparing to the current frame

General trajectory

Original trajectory on the video:

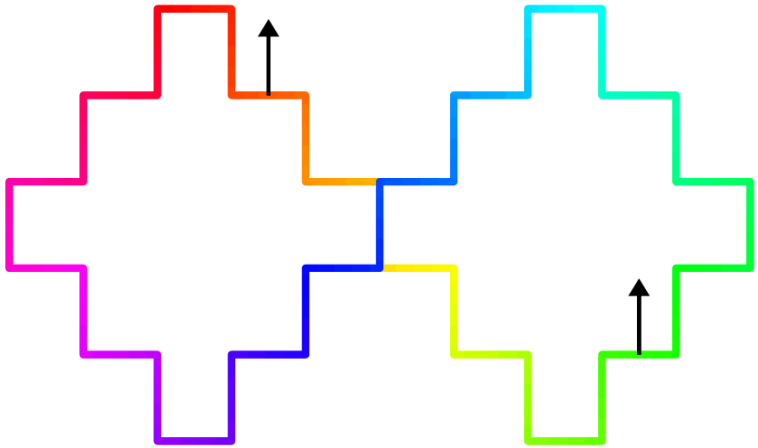


Our embedding in a 3D space:

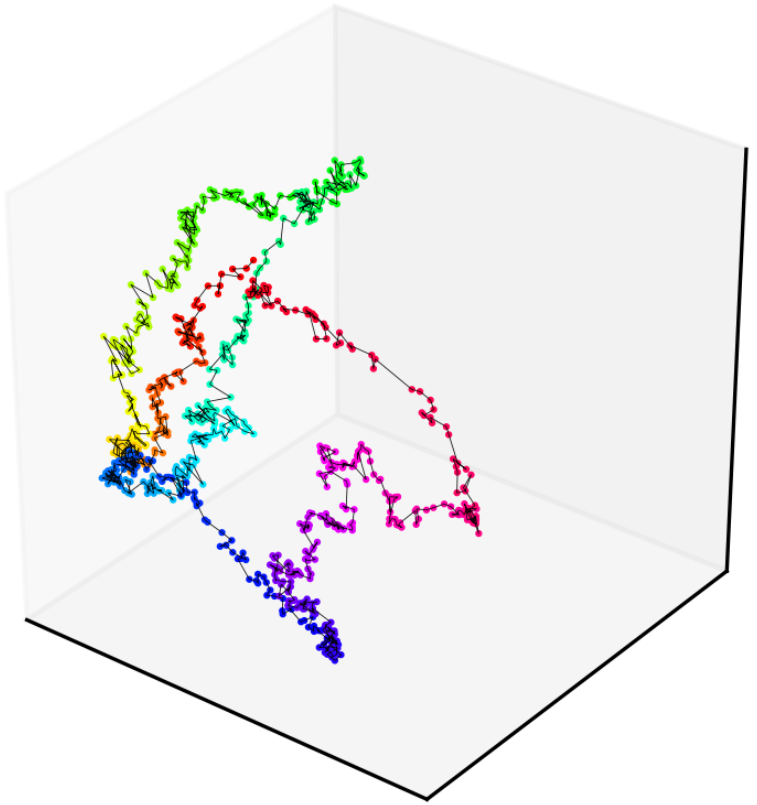


General trajectory

Original trajectory on the video:

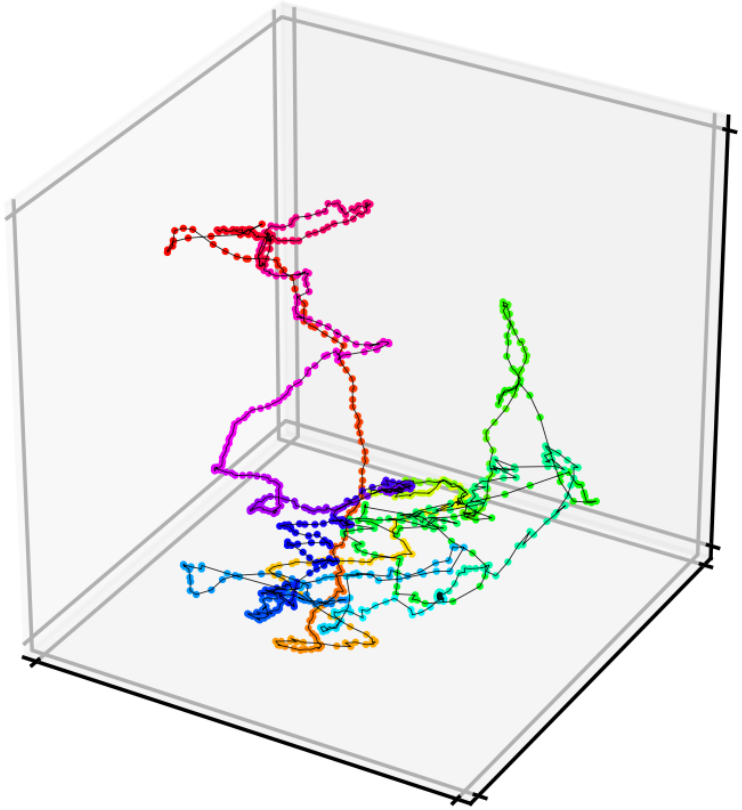


Our embedding in a 3D space:



General trajectory. Continued

Trajectory embedding:



Key characteristics

- Trajectory is not recognizable
- Subsequent frames remain nearest neighbors 58% of the time
- Only 19% of predicted frames were at least somewhat closer to the actual embedding than current frame

Summary

- We proposed a model for unsupervised learning of topological trajectory from first-person video
- We successfully trained the model on trajectories of various complexities
- We identified the limits of our model depending on the complexity of training concepts

- Shortcomings:
  - The embedding is not robust against some common transformations.
  - Current concept do not allow to separate positional information of the player from direction of players view.

Future work

- Manifold
- Outlook
  - Something you haven't solved.
  - Something else you haven't solved.