Title:

Deep learning for multiple prediction of head motion

Type of contract:

6-month research internship for a Master 2 student

Supervisor:

Lucile Sassatelli (Université Côte d'Azur, CNRS, I3S and IUF)

Contact:

lucile.sassatelli@univ-cotedazur.fr

Location :

Laboratoire I3S (Université Côte d'Azur and CNRS)
2000 route des Lucioles
06900 Sophia Antipolis - France

Salary:

French internship gratification (~600eur/month)

To apply:

Send CV and master transcripts to above email address

Keywords:

machine learning, deep learning, computer vision, time series, variational autoencoders

Description:

Immersive environments such as Virtual Reality (VR) are developping fast. For many problems related to these environments (video compression, transmission, editing in films or guidance in games), it is crucial to understand and predict where the user is going to turn to or look at. This corresponds to a trajectory prediction problem, common to other domains such as autonomous driving (requiring to predict car and pedestrian trajectories).

This prediction problem involves heterogeneous high-dimensional input data with the video content watched by the user, and their past motion. A number of works have tackled this problem [1,2,3]. Yet, their performance is far from satisfactory because they do not consider a key charactaristic of the human data: for the same input (past trajectory and video content), there can be different output future trajectories. This is known as the diversity of the problem, and confuses the training which can quickly lead to overfitting.

Objective: Investigate (design, implementation and evaluation) deep variational methods able to provide diverse predictions of future trajectories of users watching 360° videos.

To do so, successive steps will be taken:

1. Step 1: The main methods available in the literature for Dynamical Variational Auto-Encoders (DVAE) will be reviewed [4], focusing on the key concepts of Variational Recurrent Autoencoders (VRAE) [5] and Stochastic Recurrent Neural Networks (SRNN) [6].
2. Step 2: Mastering the available code of single future prediction based on TRACK introduced in [3].
3. Step 3: Investigation of multiple future prediction considering DVAE framework to extend the architecture of TRACK, analyzing the meaning of the latent space. Approaches of Factorial Hierarchical VAE (FHVAE) [7] will be considered to disentangle user-related and video-related features impacting the motion.

Requirements:

Strong background in Machine Learning, Deep Learning and Statistics

Language/tools: Python, Keras, Tensorflow, PyTorch, OpenCV

References:

[1] Y. Xu, Y. Dong, J. Wu, Z. Sun, Z. Shi, J. Yu, and S. Gao.Gaze prediction in dynamic 360° immersive videos. in IEEE CVPR, 2018, pp. 5333-5342.
[2] M. Xu, Y. Song, J. Wang, M. Qiao, L. Huo, and Z. Wang. Predicting head movement in panoramic video: A deep reinforcement learning approach. IEEE Trans. on PAMI, 2018.
[3] M. Romero, L. Sassatelli, R. Aparicio-Pardo, and F. Precioso. TRACK: A New Method from a Re-examination of Deep Architectures for Head Motion Prediction in 360◦ Videos. Preprint arXiv: 1911.11702, Feb. 2020.
[4] L. Girin, S. Leglaive, X. Bie, J. Diard, T. Hueber, and X. Alameda- Pineda, “Dynamical Variational Autoencoders: A Comprehensive Review,” arXiv:2008.12595 [cs, stat], Dec. 2020, arXiv: 2008.12595. [Online]. Available: http://arxiv.org/abs/2008.12595
[5] O. Fabius and J. R. van Amersfoort, “Variational Recurrent Auto- Encoders,” arXiv:1412.6581 [cs, stat], Jun. 2015, arXiv: 1412.6581. [Online]. Available: http://arxiv.org/abs/1412.6581
[6] M. Fraccaro, S. K. Sønderby, U. Paquet, and O. Winther. Sequential neural models with stochastic layers. In Advances in Neural Information Processing Systems, 2016.
[7] W.-N. Hsu, Y. Zhang, and J. Glass. Unsupervised learning of disentangled and interpretable representations from sequential data. In Advances in Neural Information Processing Systems, pages 1878–1889, 2017.