Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Jan Waltl
Název práce	Autoregressive action-conditioned 3D human motion synthesis using latent discrete codes
Rok odevzdání	2022
Studijní program	Computer Science
Studijní obor	Artificial Intelligence

Autor posudku	Josef Šivic
Role	Vedoucí práce
Pracoviště	CIIRC ČVUT

Posudek byl připraven ve spolupráci s Tomášem Součkem (CIIRC ČVUT), který práci spoluškolil.

Text posudku:

Thesis overview. The objective of the thesis is to develop a new method for synthesis of realistic and diverse 3D human motions conditioned on categorical action labels such as "Running" or "Toe touching". This is an important problem with applications in, for example, special effects in movies or computer games, where such technology could significantly simplify and reduce the cost of the motion authoring process. The problem is also hard. Specific motions can be performed in very different ways by different people and it is difficult to capture and realistically reproduce this variability. The current state-of-the-art solutions to this problem use transformer neural architectures to capture this variability. However, one of the key limitations of those solutions is that they produce only motions of a fixed and given length. The objective of this thesis is to address this issue and develop a new approach that can produce human 3D motion of potentially arbitrary length. This objective is achieved by developing a new autoregressive neural architecture with latent discrete codes that enables generation of motions in an autoregressive manner with potentially arbitrary length. The thesis presents results on the UESTC dataset, which is one of the standard datasets for this problem, and demonstrates reaching and, in some cases outperforming, the state-of-the-art for this problem while enabling the generation of motions of variable length.

Content of the thesis and its main contributions. The thesis is divided in five chapters. Chapter 1 describes the motivation, challenges and objectives. Chapter 2 gives the overview of the main related work for this problem. Chapter 3 describes the problem formulation, the developed neural architecture and details of its individual modules and their tested variants. It also describes the details of the training procedure. Chapter 4 then provides the experimental results including the details of the dataset and performance metrics, comparison with the state-of-the-art and extensive ablations of the main components of the model. Finally, the main limitations and future work are discussed in Chapter 5. The thesis has the following three main contributions. First, the thesis develops a new neural architecture for 3D human motion synthesis that is based on latent discrete codes. Similar type of discretization has been used for 2D image synthesis tasks. The thesis develops a new latent discrete code architecture for synthesizing 3D human motion, which is a major innovation in this area. Second, discretizing the 3D human pose space than allows for developing an autoregressive transformer-based architecture for synthesis of potentially arbitrary length human motions. Finally, the experiments demonstrate reaching and, in some cases outperforming, the state-of-the-art for

this problem.

Evaluation. Overall, the student has demonstrated independent work on a challenging problem. This was a an open-ended research-oriented project. The student has mastered the state-of-the-art for 3D human motion synthesis and designed, implemented and tested several new neural architectures for this problem. From this perspective, this work was more akin to the level of difficulty of the first year of Phd research. While the writing process required quite substantial feedback from the supervisor/consultant and was done somewhat in a rush, the final version is comprehensible and provides a solid basis for future research. The obtained results reach and in some cases outperform the current state-of-the-art in this area. With some additional effort including results on a second dataset, we would like to submit the work to a major computer vision conference such as CVPR. Such level of innovation and results is not very common for a MSc thesis.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Datum **Podpis**

30.8.2022