Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Bc. Jan Waltl
Název práce	Autoregressive action-conditioned 3D human motion synthesis using latent discrete codes
Rok odevzdání	2022
Studijní program	Informatika **Studijní obor** Umělá inteligence

Autor posudku	Mgr. Martin Mirbauer	**Role**	Oponent
Pracoviště	KSVI		

Text posudku:

This work focuses on the problem of 3D motion generation: generating a realistic sequence of human poses, performing a specified activity, either starting from a single pose, or continuing a given motion. The author uses ideas coming from state-of-the-art research on text-to-image generation networks and applies them to the motion encoding/generation domain, namely discretization of the latent space.

The author splits the problem into two sub-tasks: encoding and motion prediction, conditioned by motion type label. Both parts share the same latent space – sequences of (discrete) tokens. An auto-encoder is trained to convert frame(s) of the input motion to the token sequence, and then reconstruct it. The motion predictor uses a Transformer conditioned on the desired motion category label, processing previous motion tokens including those just generated (autoregression). Finally the predicted tokens are decoded into animation frames using the auto-encoder's decoder part.

Design decisions are sufficiently described with sound reasoning given. In cases where multiple approaches are likely to work well for a specific task, the author conducts experiments or ablation studies to compare their performance, and similarly for architecture hyper-parameters optimization. Extensive evaluation is done w.r.t. fundamental design choices like auto-encoder architecture block type (*residual 1D-convolutional* or *attention*) and discretization type (*nearest neighbor* or *Gumbel-softmax*).

Formal quality of the text is very high with good text structure, level of detail, and only a few, negligible typographical errors. There are just minor aspects reducing readability:
- Some figures and tables in the evaluation are several pages before/after their reference in the text, likely caused by their large size and count.
- Parameter sweep values written in a table (e.g. Table 4.4) may be better interpretable if shown in a plot.

Overall the thesis is well-written, as well as the provided source code.

The presented method reaches state-of-the-art results outperforming previous method (ACTOR) based on the FID metric. In addition, it is able to generate sequences of arbitrary length (previous method's limitation).
During evaluation, when showing “smoothing” in the supplementary videos, it would be useful to show samples from the A-UESTC dataset in addition to showing results of your model trained on that dataset.

All discovered artifacts of the generated motions are already mentioned in the thesis.
I would appreciate if the thesis text contained a note about the total model size (parameter count and/or memory usage) and inference speed (frames per second).

Additional questions:
- Table 4.4: Changing weight of a loss function may impact learning speed. Did the training converge in all training runs in this table?
- Is the 6D orientation representation unambiguous? (Quaternions can represent rotation in just 4 values.) If this representation is ambiguous, using a loss ($\mathcal{L}_{6D}$) limits the produced values more than necessary.

Future work ideas:
- If this method was released as a plug-in for 3D animation software, it could simplify animators' workflows. Please consider this as a possible future work.
- It would be useful if, in addition to continuation of a motion, the motion predictor could take one or more keyframes in the future, which it should smoothly reach in n frames.
- It may be worth to try using adversarial loss, which may detect the observed global rotation issue, assuming the dataset size is sufficient.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 30. 8. 2022 **Podpis**