CS221 Fall 2017 Project Proposal

Members:
Cary Huang (carykh)
Cindy Jiang (cindyj)
Connie Xiao (coxiao)

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Subject

Description: Our goal is to generate dance moves from a dataset consisting of videos with people dancing. More concretely, our input would be dancing videos and we would output a generated dancing video. Our dataset would be an image sequence of videos of real human dancing from either online or taken ourselves. To make training easier, we could use low-resolution images (maybe 30 by 30) in black-and-white and with a low frame rate (about 10). In order to attempt this task, we will break it down into two stages: training and generating.

Stage 1: Train an auto-encoder on still images of people dancing. If this works, now we have a way to quickly convert between 30x30 pixel images (large) and 50-dimensional vectors (small) (These numbers are approximations.) Now, every dance is represented as a sequence of 50-dimensional vectors across time.

Stage 2: In order to produce a good-looking dance, you merely need to produce 50D vectors well. We could use Markov models to train the computer to do that by using our dataset and treating the videos as discrete images. We are also considering using an RNN because each action is a 50D-to-50D vector, which would not be discrete (the 50D space of all dancing poses is continuous).

Evaluation: For evaluating the success of our generated dance moves, we would ask people to observe our generated dance and administer a survey that would ask them to measure how feasible the dance looks along with other metrics.

Approach

For both baseline and oracle, well only focus on Stage 2 (the actual dance generating), because Stage 1 (auto-encoding) has already proven to work reliably. (There are open-source image auto-encoders on GitHub that we can snatch)

Baseline: Given that Stage 1 works correctly, our baseline algorithm will generate a sequence of 50-D vectors that corresponds to some dance of some sort, and hope that it looks remotely natural. To do this, perhaps well use a basic feedforward neural network. Well feed in only the current frame vector as input (50 neurons), have one (or two) hidden layers of neurons, and output the next frame vector as output (50 neurons). To train, we just take an arbitrary frame from the actual dancing video as the input and the following frame as the output, and repeat thousands of times, performing stochastic gradient descent.

When its done training, we can now feed a randomly generated vector into this NN, feed its output back as input, and repeat a few hundred times to create a sequence of 50-D vectors. After using the auto-encoder to decode, we have a video of dance poses. If this looks anything remotely like a dance, weve succeeded!

If we want to be *fancy* with our baseline (oxymoron, I know) we could feed in the last TWO dance frames as input (100 neurons), which would encode useful information such as the speed and direction each body part is moving. But it definitely isnt necessary for the barebones baseline.

Oracle: For our oracle, we can have a person dance because ideally, our generated dance moves would look like a person dancing. [change later]

Addressing the gap and potential challenges: Our challenges include being able to use data that is continuous to its full potential. In order to address the gap between our baseline and our oracle, we plan to use topics like Markov models or an RNN as mentioned in Stage 2.

Similar Work

There are some related works regarding using models to generate dance moves or activity. One such project called Chor-rnn uses neural networks to generate choreography represented in the training set. This project was evaluated by a choreographer, who also learned and performed the choreography to evaluate its feasibility. [1] Another project uses neural networks on top of an autoencoder, trained on an extensive dataset of motions, to generate realistic-looking movements. [2]

\(ˆ▽ ˆ)/ \(ˆ▽ ˆ)/ ← (YAY! WE <3 DANCING!)

[1] Crnkovic-Friis, L., & Crnkovic-Friis, L. (2016). Generative Choreography using Deep Learning. arXiv preprint arXiv:1605.06921.

[2] Holden, D., Saito, J., & Komura, T. (2016). A deep learning framework for character motion synthesis and editing. ACM Transactions on Graphics (TOG), 35(4), 138.