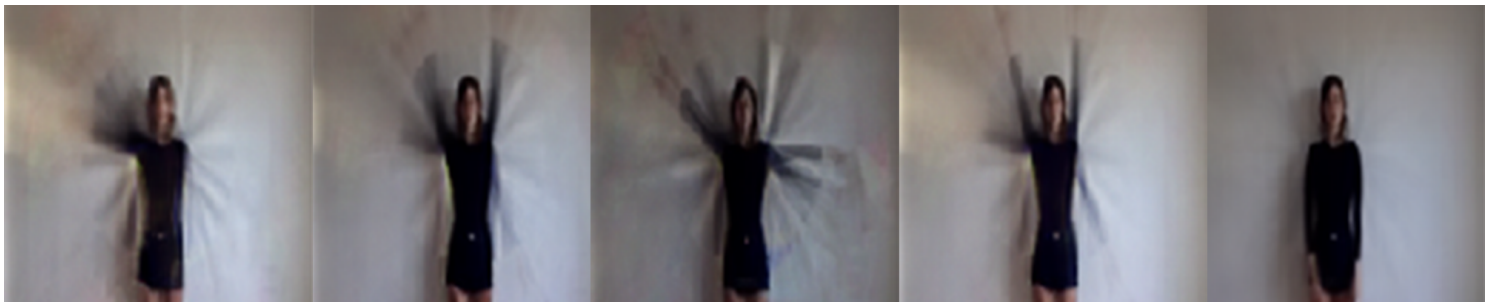
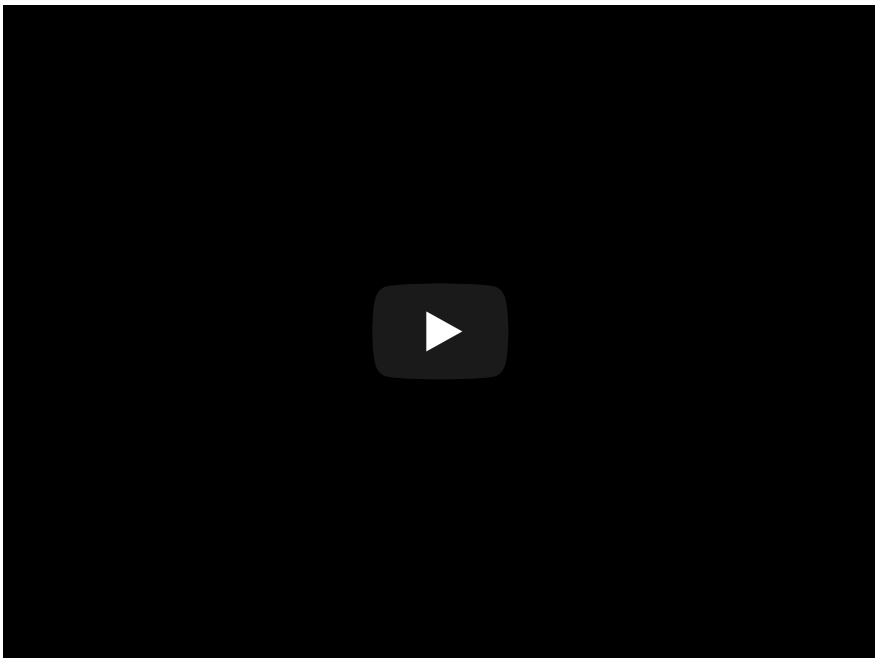
Arthur Juliani Follow
Deep Learning @Unity3D & Cognitive Neuroscience PhD student.
Jul 19, 2016 · 3 min read

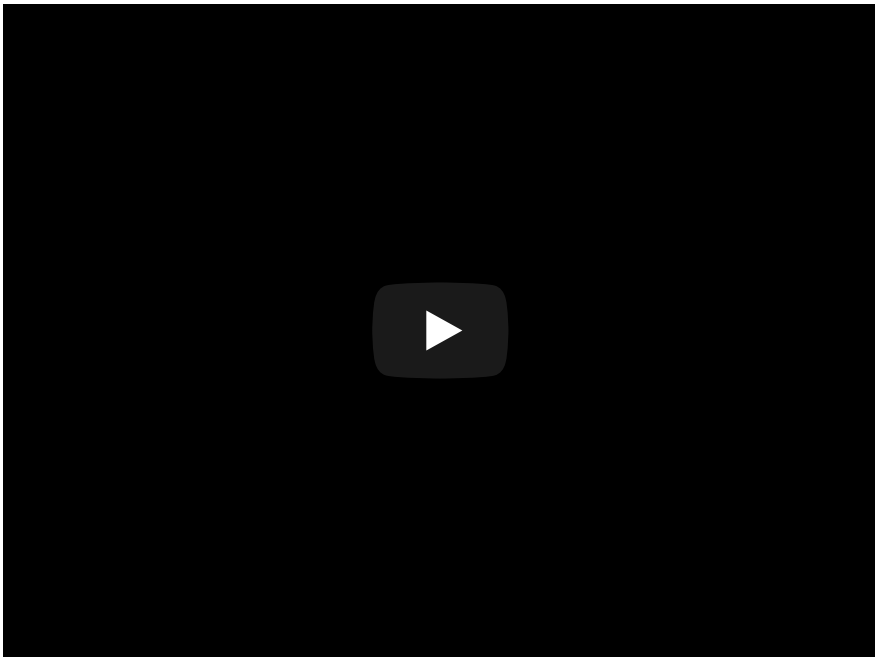
Introducing Neural Dream Videos



Over the past few weeks I have been working on a way to combine two deep learning architectures, a variational auto-encoder and a recurrent neural network, to allow for the generation of what I am calling *Neural Dream Videos*. These videos capture both the spatial and temporal properties of a given source video, and produce potentially endless new variations on the substance of the video itself in a hallucinogenic way. Below are a couple examples of Neural Dream Videos I generated from classic video games (see further down for live-action videos).

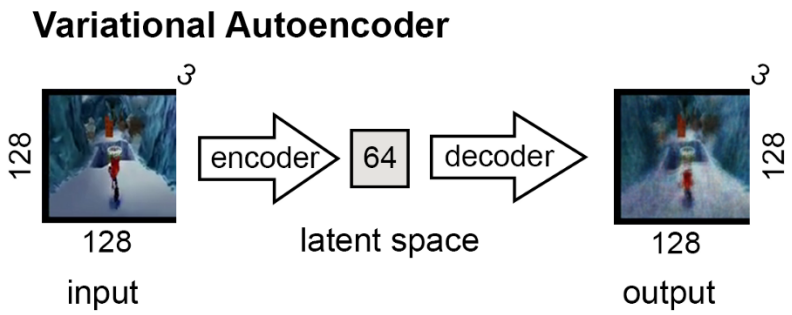


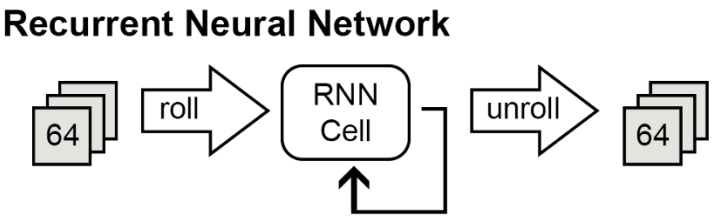
Mario Kart 64 Rainbow Road level recreated as neural dream video. (Original)



Crash Bandicoot 2 N. Gin Boss Fight recreated as neural dream video. (Original)

The videos above were generated completely from the two neural network working together, and once trained don't rely on the source videos at all. Let me explain the process below.

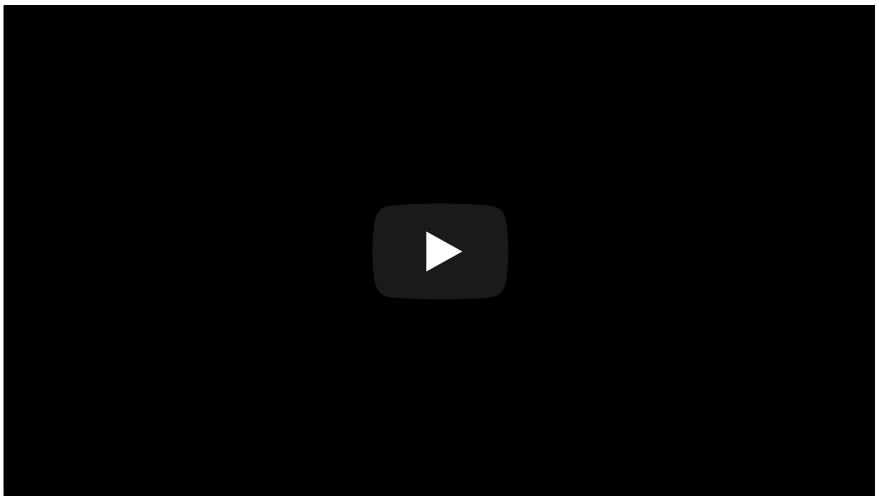




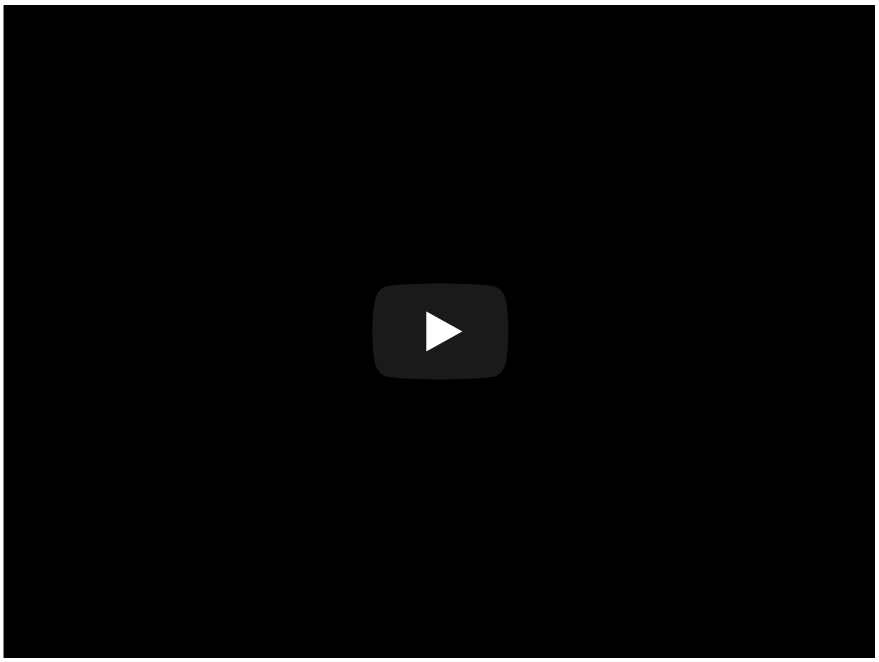
Variational Auto-encoders (VAE) have been used as a way to allow neural networks to create new examples of images. They do this by learning an efficient representation of the spatial characteristics of thousands of training images. For example, give a VAE thousands of horse photos, and it will learn a representation that allows it to produce novel horse photos from it's own representation. Like humans, this can be thought of as a kind of imagination at work. While they work well for still images, videos are far too complex to capture using a VAE however, and that is where the current network comes into play.

Recurrent Neural Networks (RNNs) are used to learn temporal patterns in data, and create new patterns from old ones. A common use case for RNNs that has gained attention lately is language modeling. For example an RNN learns to produce a piece of text, such as a new Shakespearian play, when trained on a corpus of Shakespeare's works. The problem with RNNs by themselves is that they are too simple to understand the relationships between video frames, which contain thousands of numbers and complex spatial relationships. By combining a VAE and RNN, we can train the VAE to learn a compact and semantically meaningful representation of the frames of a video, and then train an RNN to learn how to model the temporal patterns of those latent representations. Once we have a newly generated sequence of latent representations, we can run them through the trained VAE to generate a new set of full video frames. Stitch them back into a video, and thats it!

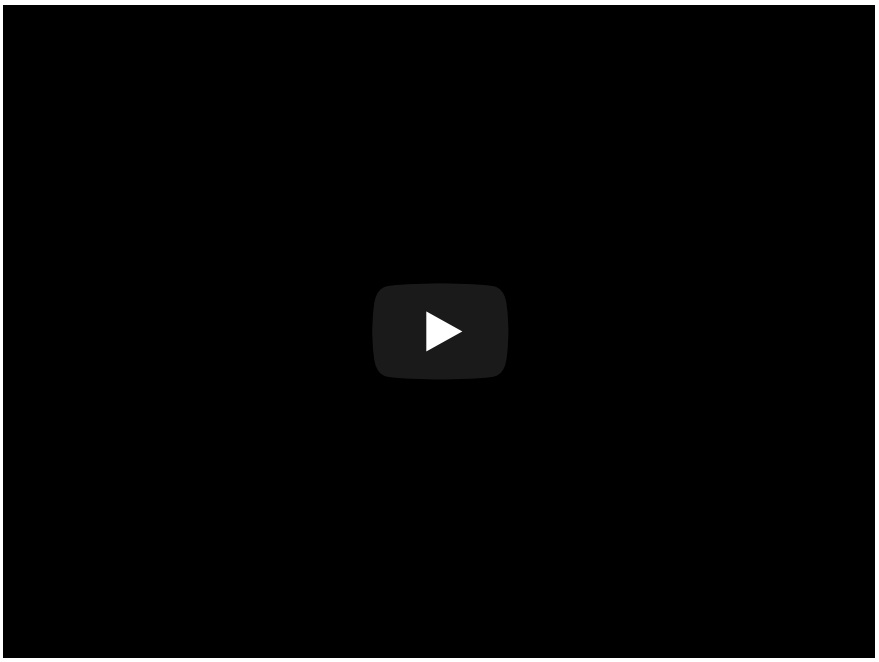
Below are a few videos created from live-action footage.



A Neural Dream Video made in collaboration with artist Laura Figa.



A recreation of a recent Donald Trump interview. Sometimes the method can produce pretty frightening results...



Dance of the Sugar Plum Fairy. (Original)

I am happy to be releasing this code for creating these videos on Github here, and I hope others are able to have fun making these kinds of videos as well. There are also a lot of ways in which the quality of both the video and the logic of the temporal flow of the videos could be improved by using more sophisticated kinds of auto-encoders and recurrent networks. If you work on those kinds of things, please feel free to contact me or contribute additions or changes to the repository.

We are still in the early times of creative applications for generative networks. Google is currently developing models that produce basic melodies, and the videos here don't stray too far from their source material. Still, such networks are the first steps to neural networks that will be able to accomplish the exciting tasks of write songs, making art, and create films on their own.

. . .

I am a Phd student in Cognitive Neuroscience, currently looking for internship and work opportunities in the San Francisco Bay area this coming fall. If you are part of a company working on Deep Learning and AI problems, I would love to chat about joining your team!