Simple Video Generation using Neural ODEs

Vikram Voleti, David Kanaa, Samira Kahou, Chris Pal

NeurIPS 2019 Workshop

Vikram Voleti

PhD student, Mila
University of Montreal

voletiv.github.io

VIKRAM VOLETI

2019 Dec
Visiting Researcher
Prof. Graham Taylor, University of Guelph

2019 Sep-Dec
Research Intern
Google AI Perception

2018 Sep
PhD student
Mila, University of Montreal - Supervisor: Prof. Chris Pal

2017 May
Research Fellow
IIIT Hyderabad - Supervisor: Prof. C. V. Jawahar

2016 Feb
GreyOrange Robotics - Image Processing Engineer

2014 Jun
Airbus - Associate Engineer

2009 - 2014
B.Tech. + M.Tech.
IIT Kharagpur

Neural ODEs

can be trained to model

latent dynamics in video.

Video Generation

- Conditioned on *m* frames, generate next *n* frames

Video Generation

- Conditioned on *m* frames, generate next *n* frames

Stochastic Adversarial Video Prediction:
https://arxiv.org/abs/1804.01523

High Fidelity Video Prediction with Large Stochastic Recurrent Neural Networks:
https://arxiv.org/abs/1911.01655

Improved Conditional VRNNs for Video Prediction - https://arxiv.org/abs/1904.12165

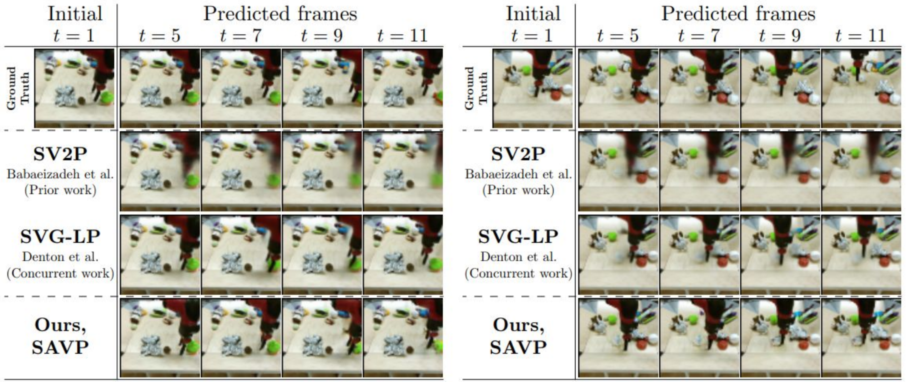


Fig. 1: **Example results.** We show example predictions for two video sequences comparing our method to prior and concurrent work. All methods are conditioned on two initial frames and predict 10 future frames (only some frames are shown). The predictions are stochastic, and we show the closest sample to the ground truth (out of 100 samples). While the prior SV2P method [1] produces blurry, unrealistic images, our method maintains sharpness and realism through time. We also compare to the concurrent SVG-LP method [2], which produces sharper predictions, but still blurs out objects in the background (left) or objects that interact with the robot arm such as the baseball (right).

Idea

What if we model the latent dynamics of a video using Neural ODEs?

Neural ODEs

3 Neural Ordinary Differential Equations

Neural Ordinary Differential Equations (Chen et al., 2018) (Neural ODEs) represent a family of parameterised algorithms designed to model the evolution across time of any system, of state $\boldsymbol{\xi}(t)$ at an arbitrary time t, governed by continuous-time dynamics satisfying a Cauchy (or initial value) problem

$$\begin{cases} \boldsymbol{\xi}(t_0) & = \boldsymbol{\xi}_0 \\ \dfrac{\partial \boldsymbol{\xi}}{\partial t}(t) & = f(\boldsymbol{\xi}(t), t) \end{cases}$$

By approximating the differential with an estimator $f_\theta \simeq f$ parameterized by θ, such as a neural network, these methods allow to learn such dynamics (or, trajectories) from relevant data. Thus

https://arxiv.org/abs/1806.07366

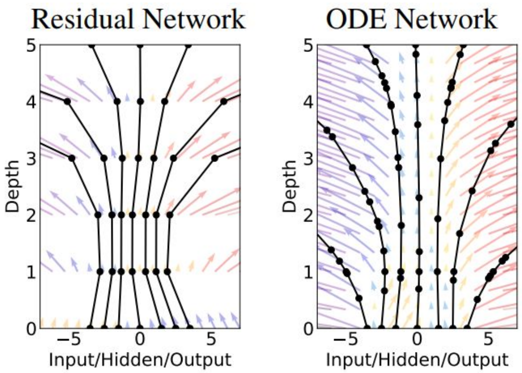
Neural ODEs

3 Neural Ordinary Differential Equations

Neural Ordinary Differential Equations (Chen et al., 2018) (Neural ODEs) represent a family of parameterised algorithms designed to model the evolution across time of any system, of state $\boldsymbol{\xi}(t)$ at an arbitrary time t, governed by continuous-time dynamics satisfying a Cauchy (or initial value) problem

$$\begin{cases} \boldsymbol{\xi}(t_0) & = \boldsymbol{\xi}_0 \\ \dfrac{\partial \boldsymbol{\xi}}{\partial t}(t) & = f(\boldsymbol{\xi}(t), t) \end{cases}$$

By approximating the differential with an estimator $f_\theta \simeq f$ parameterized by θ, such as a neural network, these methods allow to learn such dynamics (or, trajectories) from relevant data. Thus



https://arxiv.org/abs/1806.07366

Ours

Prev. video generation papers:

- train on generation

Neural ODE paper:

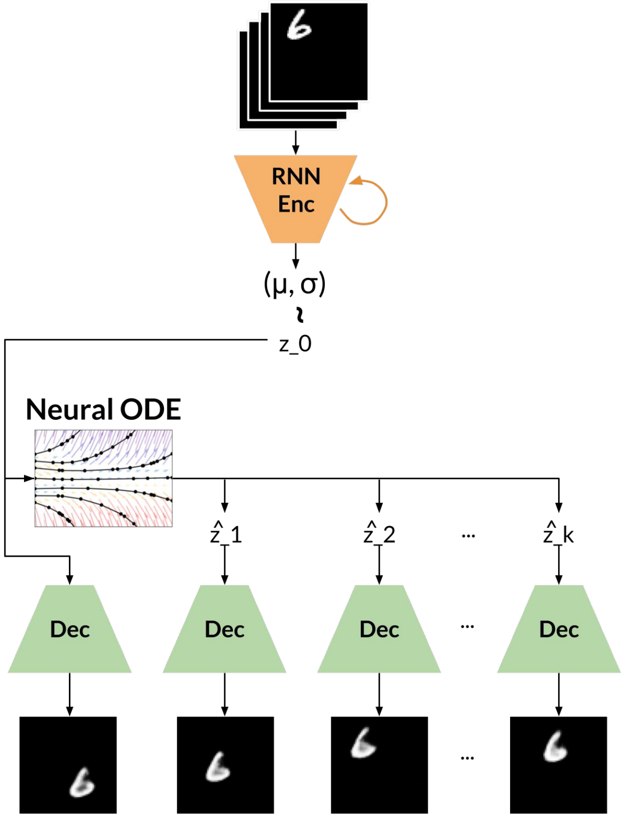
- replaces Residual layers with Neural ODE layer.

Ours:

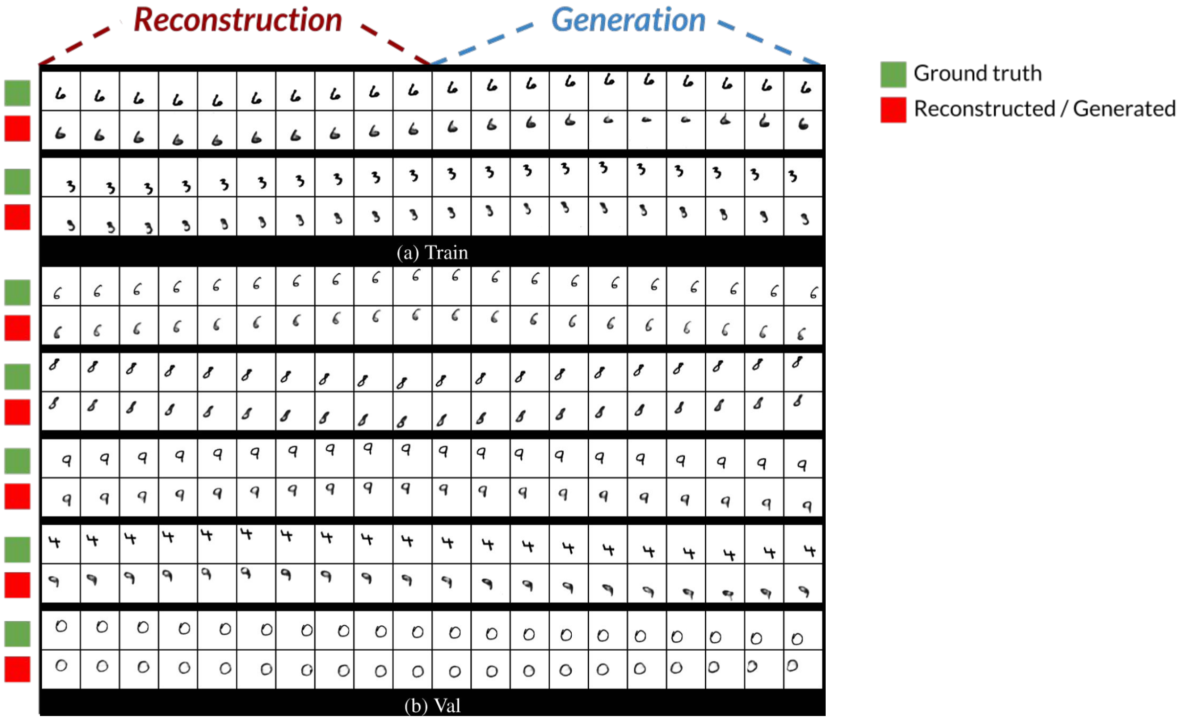
- **trains on only reconstruction.**
- **uses Neural ODE in latent space.**

https://arxiv.org/abs/1806.07366

Architecture

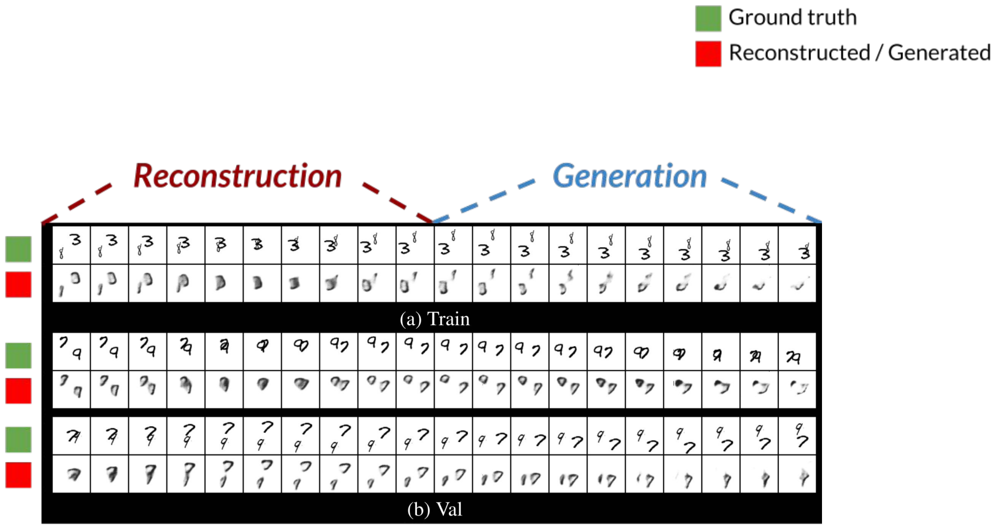


Results



Moving MNIST 1-digit

Results



Moving MNIST 2-digit

Simple Video Generation using Neural ODEs

- We study the use of Neural ODEs in video generation.
- We train an encoder-decoder architecture with a Neural ODE in latent space to generate video frames in future time steps.
- We show that Neural ODEs can model the latent dynamics in video for the 1-digit and 2-digit Moving MNIST dataset.
- We list the potential use cases and future steps in this direction.

Future (current) work

- Scaling up to natural datasets
- Visualization of latent space
- Evaluation using new video quality metrics (FVD, ...)

Links

Vikram Voleti - https://voletiv.github.io

Paper

Mila

Thank you.

voletiv.github.io