Video Frame Synthesis using Deep Voxel Flow

Ziwei Liu1  Raymond A. Yeh2  Xiaou Tang1  Yiming Liu3  Aseem Agarwala4

Motivation
Video Frame Synthesis

* Complex motion (camera motion & scene motion)
* High-res video generation (720p - 1080p)

Previous Attempts

Hallucination-based Methods
[Ranjan et al.], [Srivastava et al.], [Mathieu et al.]

Flow-based Methods

Our 3D Voxel Flow

* Symmetric bi-directional flows
* Selection mask between frames
* Differentiable bilinear sampling

Unsupervised Training
Any video can be used as training data by dropping, and then learning to predict, existing frames.

Approach
Deep Voxel Flow

Conv. encoder-decoder with voxel flow layer

Voxel Flow Layer

\[
W^{00} = (1 - (L_0^0 - L_0^1))(1 - (L_0^0 - L_0^1))(1 - \Delta t)
\]
\[
W^{01} = (L_0^0 - L_0^1)(1 - (L_0^0 - L_0^1))(1 - \Delta t)
\]
\[
\vdots
\]
\[
W^{11} = (L_0^0 - L_0^1)(L_0^0 - L_0^1)\Delta t ,
\]

Multi-scale Voxel Flow

Input Video

Convolutional Encoder-Decoder

Synthesized Frame

Experiments
Video/View Interpolation & Extrapolation

(UCF-101, KITTI)

Self-supervised Learning

Slow-motion Effects from 15fps to 30fps

User Study on 720p Videos

The result of voxel flow is indistinguishable from ground truth.

Project Page:
https://liuziwei7.github.io/projects/VoxelFlow