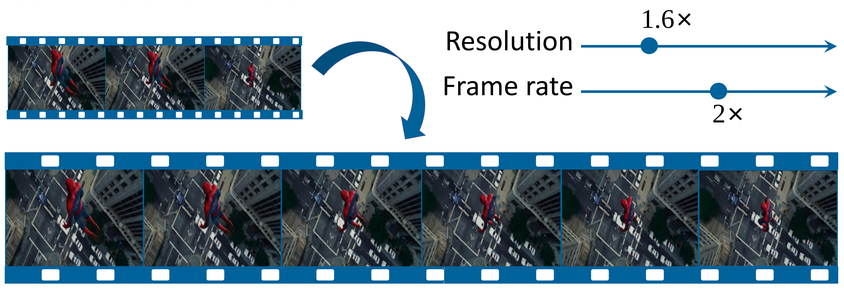
VideoINR: Learning Video Implicit Neural Representation for Continuous Space-Time Super-Resolution

Zeyuan Chen[1], Yinbo Chen[1], Jingwen Liu[1], Xingqian Xu[3], Vidit Goel[2], Zhangyang Wang[4], Humphrey Shi[2,4,5], Xiaolong Wang[1]

[1]UCSD, [2]Picsart AI Research (PAIR), [3]UIUC, [4]UT Austin, [5]U of Oregon

Code and more results at http://zeyuan-chen.com/VideoINR

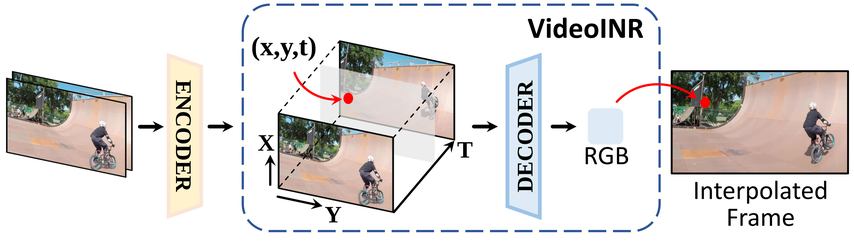
Continuous Video



◆ Videos are recorded with limited resolutions and frame rates, introducing challenges in down-stream vision tasks.

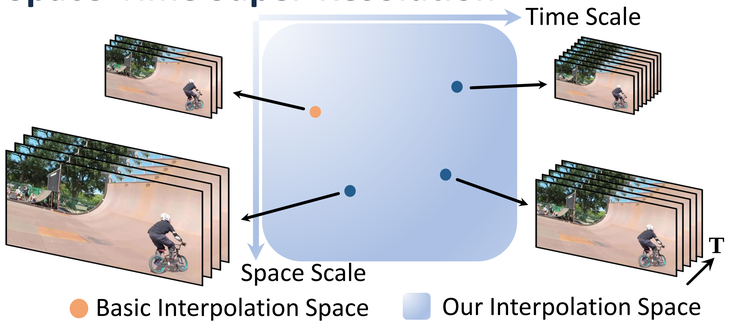
◆ We propose VideoINR for continuous video representation, which is capable of generating new videos at arbitrary spatial resolutions and frame rates from the original input video.

How to represent a video continuously?



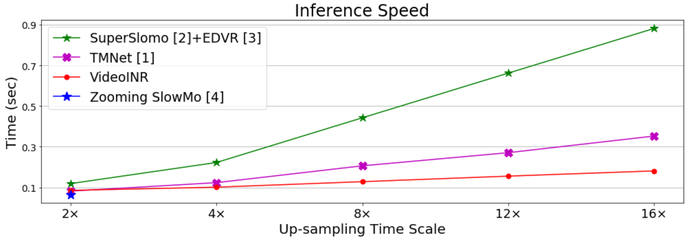
1. Define a canonical 3D space (2D coordinate + Time)
2. Predict RGB signals for any 3D coordinate in the space
3. Query coordinates of all pixels in the target frame for prediction

Space-Time Super-Resolution

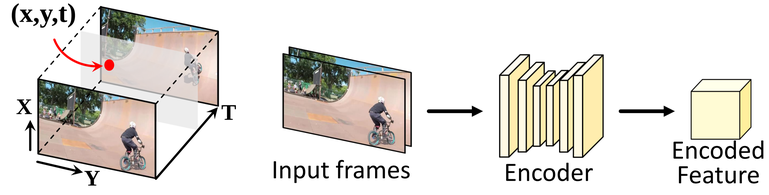


◆ VideoINR expands interpolation space, and supports extrapolation on out-of-distribution up-sampling scales.

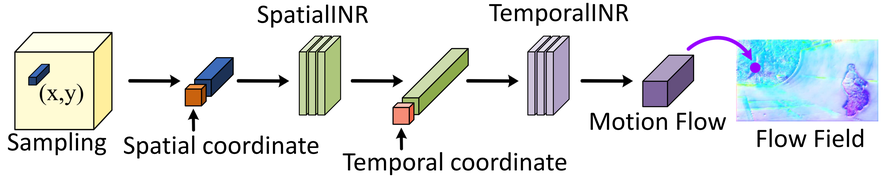
Results



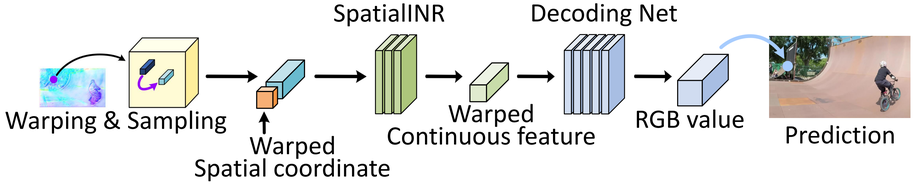
Approach



■ Extract discrete feature from input using an encoder



- Spatial INR defines a continuous spatial feature space
- Temporal INR defines a continuous motion flow field



◆ Perform spatial warping based on motion flow

◆ Decode for the final prediction