Super SloMo

(https://arxiv.org/abs/1712.00080)

High Quality Estimation of Multiple Intermediate Frames for Video Interpolation

Huaizu Jiang[1] Deqing Sun[2] Varun Jampani[2]
Ming-Hsuan Yang[3,2] Erik Learned-Miller[1] Jan Kautz[2]
[1]UMass Amherst [2]NVIDIA [3]UC Merced
{hzjiang,elm}@cs.umass.edu,{deqings,vjampani,jkautz}@nvidia.com, mhyang@ucmerced.edu

목차

Abstract

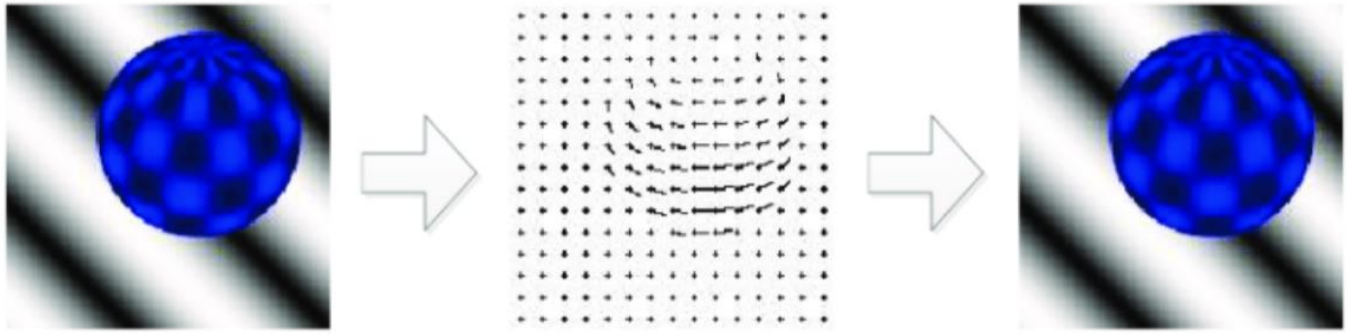
- 연구 배경
  - Video interpolation은 두 frame 사이의 spatially and temporally coherent한 intermediate frame(s)을 만들어 내는 작업을 뜻한다.
  - 대부분의 video interpolation 연구는 single-frame interpolation에 집중되어 있다.
    Single-frame interpolation은 intermediate frame을 $2^i - 1$단위의 생성만 할 수 있다는 한계가 있다.

- 목표
  - End-to-End convolutional neural network(U-Net)를 이용하여 variable-length multi-frame video interpolation을 한다.
  - 두 input images에 대한 bi-directional optical flow를 이용하여 intermediate frame을 계산하기 위한, intermediate flows와 visuality maps을 예측한다.

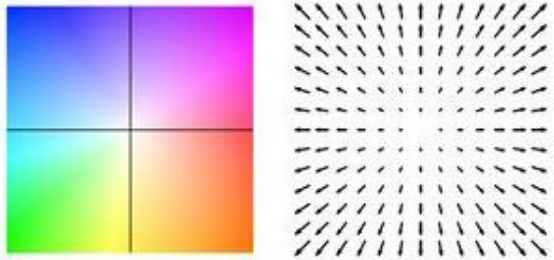
Optical Flow

What is optical flow?

- Optical Flow
  - 어떤 물체의 표면(surface)혹은 모서리(edge)의 움직임(Motion)을 나타내는 Vector Map을 말한다.
  - 연속되는 frame 사이에서 물체의 외관상의 움직임을 보여주는 값이다. (각 pixel을 각각 x, y 방향에 대한 값을 가진다.)



- Optical Flow visualization
  - 화살표로 모든 픽셀을 표현하기 힘들기 때문에, 주로 color map을 이용해서 optical flow를 표현한다.



Ex) MPI sintel dataset : https://youtu.be/ZmiBI4tPk_o?t=59

Optical Flow를 기본적인 컨셉은 다음과 같다.

- t시간일 때, x, y 위치에 있는 image의 값을 $I(x, y, t)$라고 할 때, Δt 동안 Δx, Δy만큼 움직였을 때, 밝기는 변하지 않고 위치만 움직였다고 가정하면 다음과 같이 표현 할 수 있다.

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

- 여기서 $I(x, y, t)$를 Taylor series로 표현하면,

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \text{higher-order terms}$$

- Higher order term을 무시하고, $I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$ 이므로.

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$$

- Δt로 나누어주면.

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0$$

- $V(x) = \frac{\Delta x}{\Delta t}$, $V(y) = \frac{\Delta y}{\Delta x}$ 로 바꾸면

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0$$

- 결국 각각 Image의 x,,y,t 방향으로의 편미분을 I_x, I_y, I_t 로 바꾸면.

$$I_x V_x + I_y V_y = -I_t$$

- 이를 행렬식으로 바꾸면 다음과 같이 바꿀수 있다.

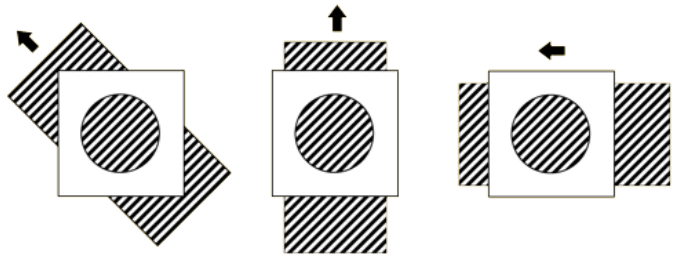
$$\nabla I \cdot \vec{V} = -I_t$$

Optical Flow

What is optical flow?

- Hard to estimate optical flow
  - Aperture(조리개) problem

    작은 부분을 통해 본 움직임의 방향을 확정 지을 수 없다.



  - Large displacement에 취약

    큰 움직임(Large displacement)에 대해서 estimate error가 크게 발생

- 응용분야
  - 물체의 이동 분석
  - 비디오 압축
  - 비디오 안정화

Approach 1: Intermediate Frame Synthesis

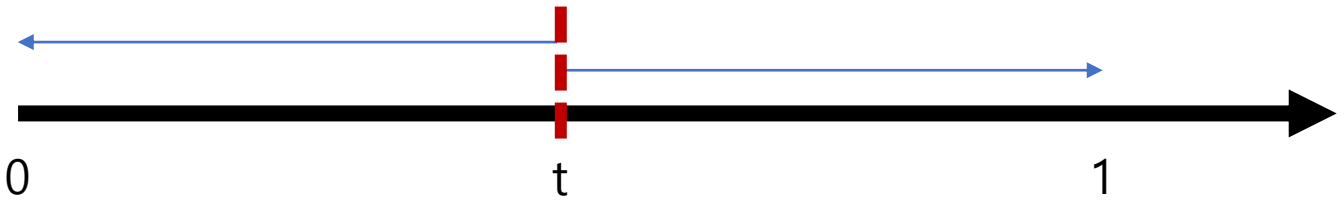
Super SloMo Approach 1.1

- Problem : 주어진 두 입력 이미지 I_0, I_1와 시간 $t \in (0,1)$에 대해 intermediate image $\hat{I}_t$를 예측하는 문제
  - > 이를 예측하기 위해서, 모델은 motion pattens 와 두 이미지의 appearance를 학습하여야 한다.
  - > $\hat{I}_t$를 directly predict 하기에는 rich RGB color space로 인해, 높은 성능의 이미지를 생성하기 힘들다.

- 위의 이유로, optical flow와 이미지를 통해 다음 frame을 복원하는 함수 warping function을 통해 $\hat{I}_t$를 계산한다.
  따라서 시간 t에서의 두 optical flow($F_{t\to0}$, $F_{t\to1}$) 가 있을 때, $\hat{I}_t$는 다음과 같이 계산 할 수 있다.

$$\hat{I}_t = \alpha_0 \odot g(I_0, F_{t\to0}) + (1-\alpha_0) \odot g(I_1, F_{t\to1}),$$

$F_{t\to0}$: t에서 0 으로의 optical flow
$F_{t\to1}$: t에서 1 으로의 optical flow
g : backward warping function
α_0: contribution parameter



Approach 1: Intermediate Frame Synthesis

Super SloMo Approach 1.2

$$\hat{I}_t = \alpha_0 \odot g(I_0, F_{t\to 0}) + (1-\alpha_0) \odot g(I_1, F_{t\to 1}),$$

$F_{t\to 0}$: t에서 0 으로의 optical flow
$F_{t\to 1}$: t에서 1 으로의 optical flow
g : backward warping function
α_0: contribution parameter

- 여기서 α_0를 결정하는 요소로 두가지를 들 수 있다.
  1. Temporal consistency

     만약 t가 0에 가깝다면, $\hat{I}_t$는 I_0에 영향을 많이 받을 것이고, I_1에 대해서 영향을 적게 받을 것이다.
  2. Occlusion reasoning
     - If a pixel p is visible at T=t, it is most likely at least visible in one of the input images.
     - Introduce *visibility maps* $V_{t\leftarrow 0}$, $V_{t\leftarrow 1}$ $(V(p) \in [0,1])$

- 위 두 요소를 고려하여 (1) 식을 수정하면.

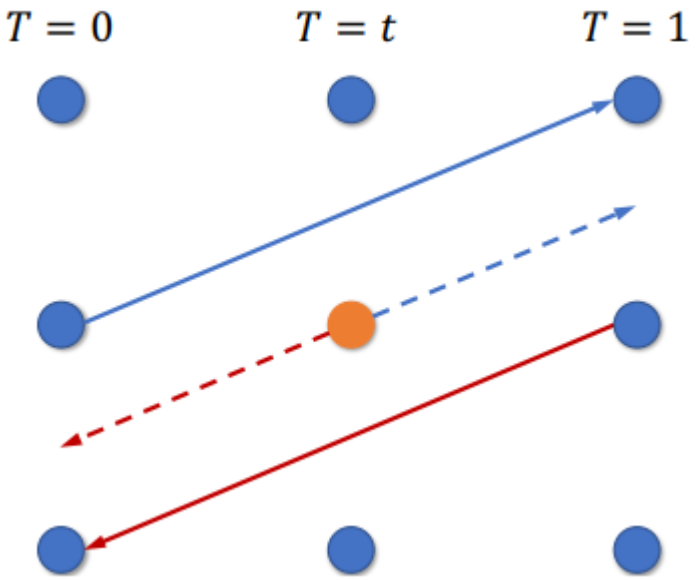
$$\hat{I}_t = \frac{1}{Z} \odot \big((1-t)V_{t\leftarrow 0} \odot g(I_0, F_{t\to 0}) + tV_{t\leftarrow 1} \odot g(I_1, F_{t\to 1})\big),$$

where $Z = (1-t)V_{t\to 0} + tV_{t\to 1}$ is a normalization factor.

Approach 2: Arbitrary-time Flow Interpolation

Super SloMo Approach 2.1

- Intermediate optical flow ($F_{t\to 0}$, $F_{t\to 1}$) approximation synthesis
  - t에서의 optical flow는, 두 input images의 optical flow ($F_{0\to 1}$, $F_{1\to 0}$)을 통해 근사 할 수 있다.



- 근사의 가장 쉬운 방법은 같은 pixel의 optical flow를 가져오는 것이다.
- 예를 들어, p위치의 t->1 로의 optical flow는 0->1로의 optical flow에 $(t-1)$배 만큼의 값일 것이다.

$$\hat{F}_{t\to 1}(p) = (1-t)F_{0\to 1}(p)$$
$$\text{or}$$
$$\hat{F}_{t\to 1}(p) = -(1-t)F_{1\to 0}(p),$$

- 이를 Temporal consistency를 고려하여 bi-directional하게 표현하면.

$$\hat{F}_{t\to 0} = -(1-t)tF_{0\to 1} + t^2F_{1\to 0}$$
$$\hat{F}_{t\to 1} = (1-t)^2F_{0\to 1} - t(1-t)F_{1\to 0}.$$

Approach 2: Arbitrary-time Flow Interpolation

Super SloMo Approach 2.2

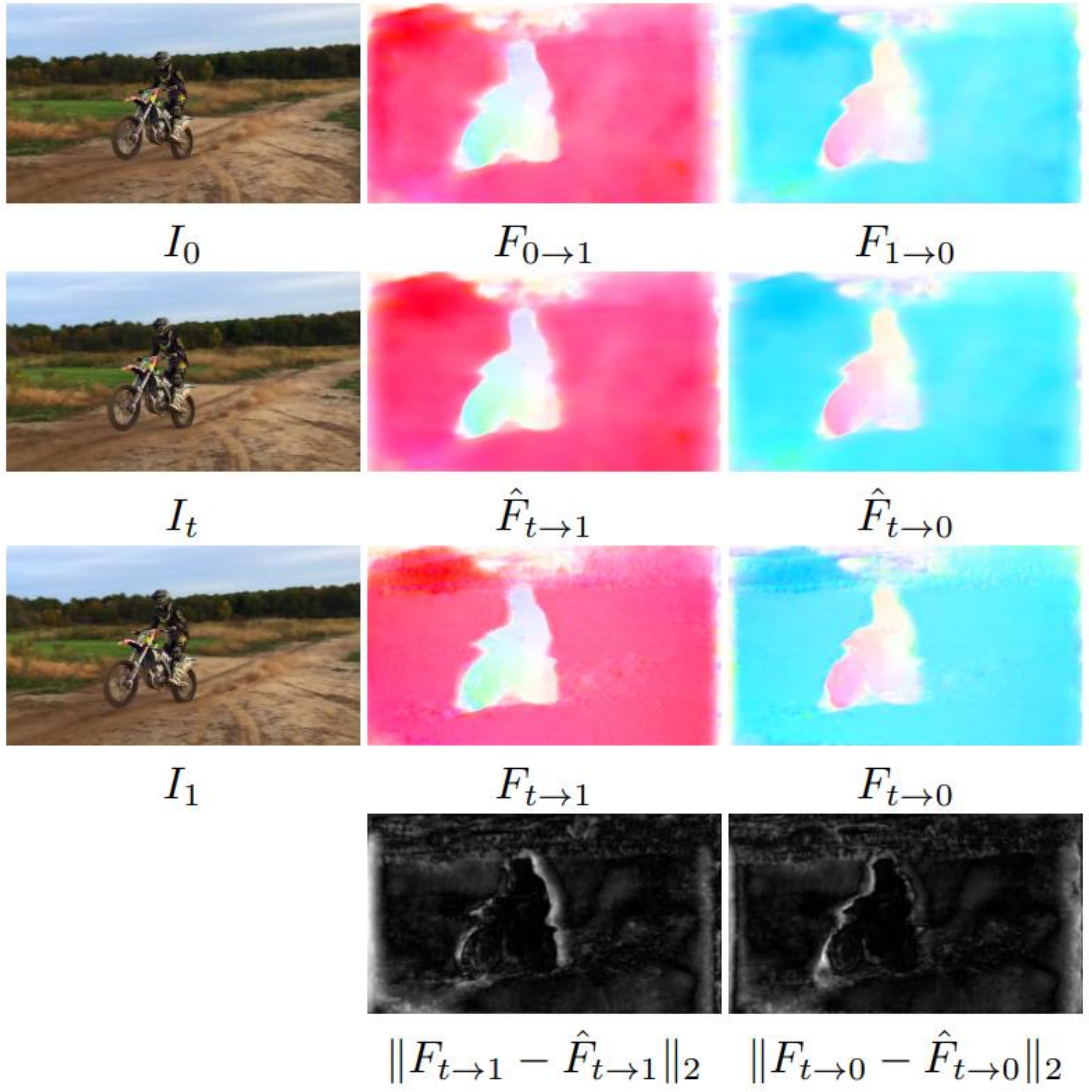
$$\hat{F}_{t\to 0} = -(1-t)tF_{0\to 1} + t^2 F_{1\to 0}$$
$$\hat{F}_{t\to 1} = (1-t)^2 F_{0\to 1} - t(1-t)F_{1\to 0}.$$

- Refine optical flow
  - 위의 approximation은 smooth regions에는 좋지만, motion boundaries에는 좋지 않다. 따라서 이를 refine할 필요가 있다.
  - Network를 통해 refine된 intermediate optical flow를 예측해준다.

- Visibility maps
  - visibility maps 역시 필요하므로, $V_{t\leftarrow 0}$, $V_{t\leftarrow 1}$를 위의 Network를 통해 같이 예측해 준다.
  - 두 visibility maps은 [0,1]의 범위 이고, 서로 반대 값을 가지기 때문에, $V_{t\leftarrow 0} = 1 - V_{t\leftarrow 1}$ constrain를 강제해 준다.

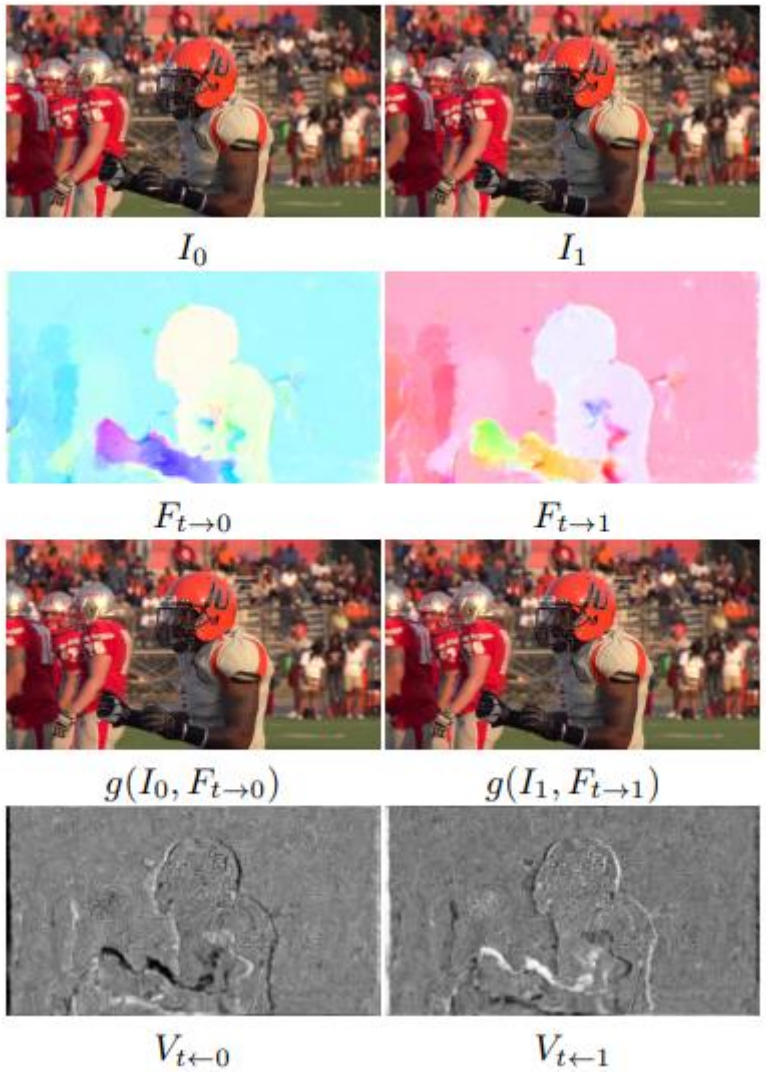
Approach 2: Arbitrary-time Flow Interpolation

Super SloMo Approach 2.3



Approach 2: Arbitrary-time Flow Interpolation

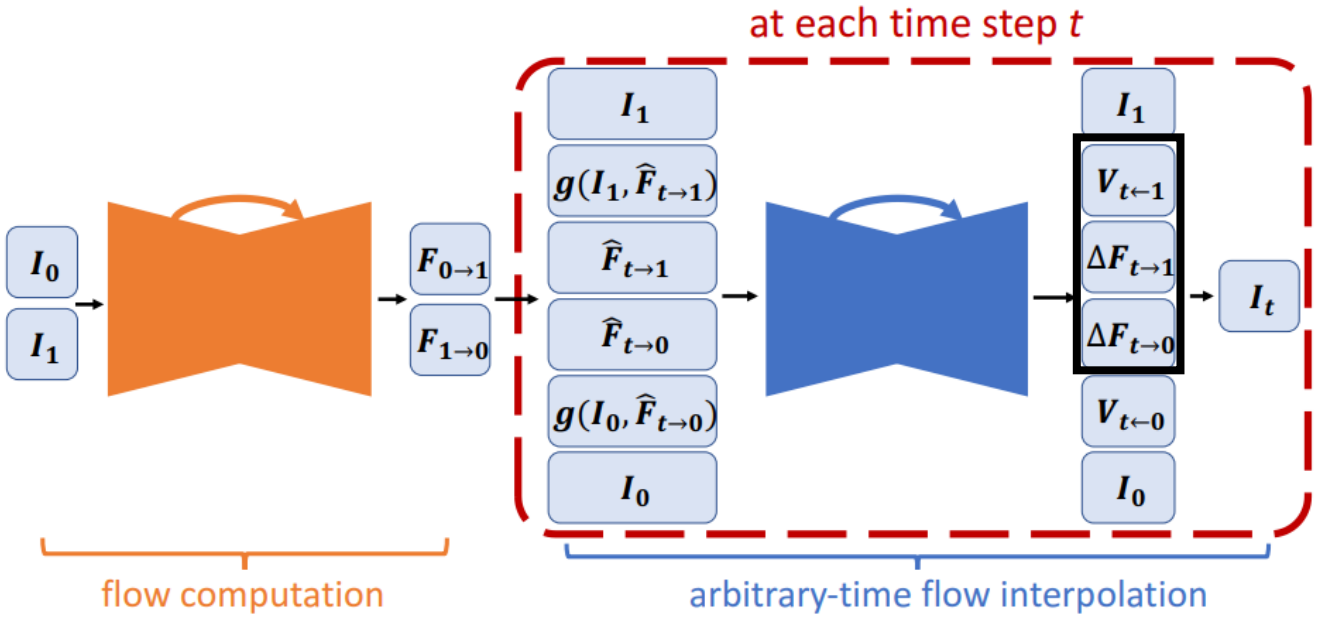
Super SloMo Approach 2.3

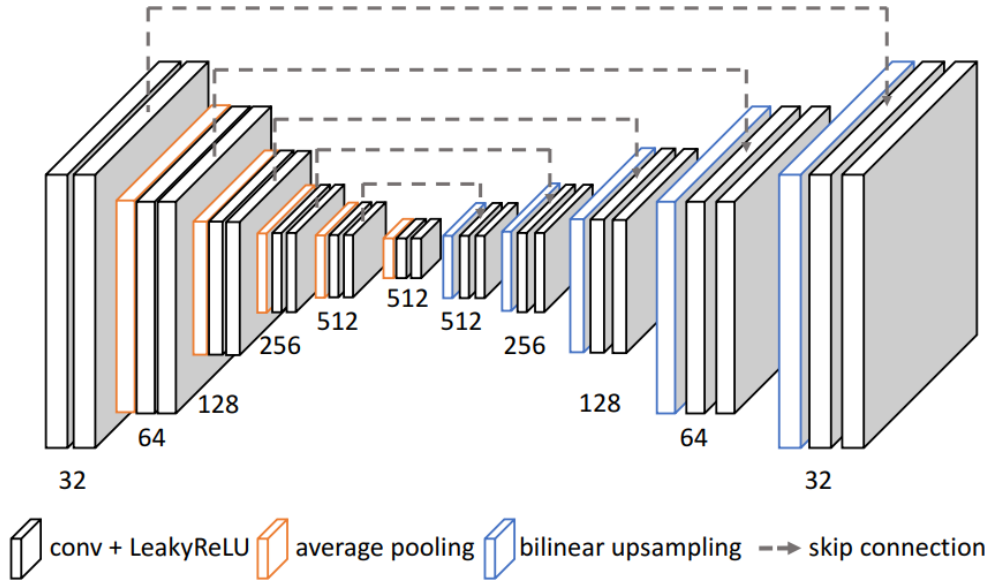


$\hat{I}_t$
PSNR=30.23

$\hat{I}_t$ w/o visibility maps
PSNR=30.06

Super SloMo architecture





Super SloMo Loss

1. Reconstruction loss : $$l_r = \frac{1}{N}\sum_{i=1}^{N} \|\hat{I}_{t_i} - I_{t_i}\|_1.$$

2. Perceptual loss : $$l_p = \frac{1}{N}\sum_{i=1}^{N} \|\phi(\hat{I}_t) - \phi(I_t)\|_2,$$

3. Warping loss : $$l_w = \|I_0 - g(I_1, F_{0\to1})\|_1 + \|I_1 - g(I_0, F_{1\to0})\|_1 + \frac{1}{N}\sum_{i=1}^{N} \|I_{t_i} - g(I_0, \hat{F}_{t_i\to0})\|_1 + \frac{1}{N}\sum_{i=1}^{N} \|I_{t_i} - g(I_1, \hat{F}_{t_i\to1})\|_1$$

4. Smoothness loss : $$l_s = \|\nabla F_{0\to1}\|_1 + \|\nabla F_{1\to0}\|_1.$$

Super SloMo Result

https://youtu.be/LBezOcnNJ68?t=61