MOTION-COMPENSATED FRAME INTERPOLATION

SUPER SLOMO

M S Nishanth
CSMTECH11005

ABSTRACT

We propose a ML algorithm for video processing in which intermediate animation frames are generated between existing ones by means of interpolation, to observe the the video/ action in slow motion to understand the content/ infer more data.

[Slow Motion]

[1]: Source WiKi

SUPER SLOMO: HIGH QUALITY ESTIMATION OF MULTIPLE INTERMEDIATE FRAMES FOR VIDEO INTERPOLATION

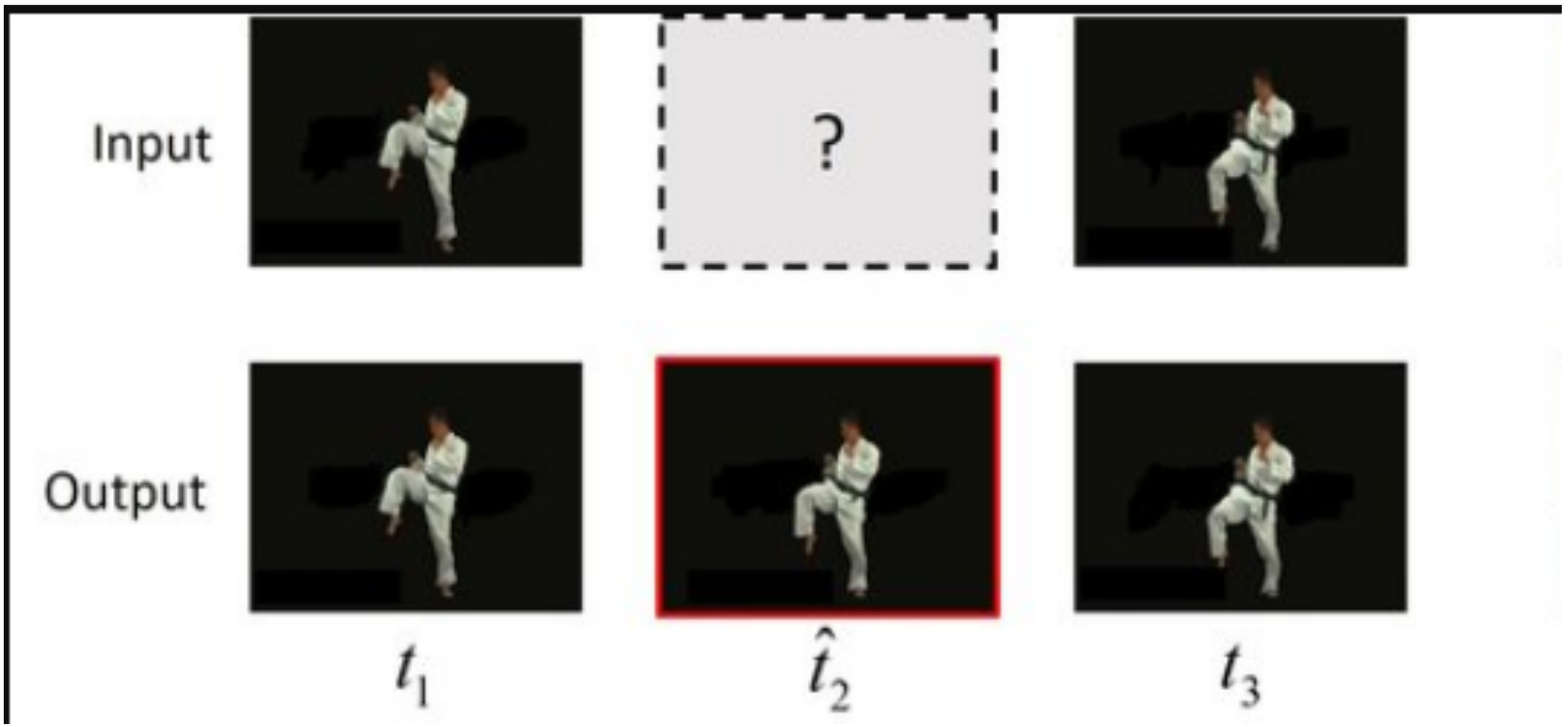
EXISTING MODEL:

Single frame interpolation [1] [2]

- Can't generate higher frame rate videos
- Cannot fully parallelize
- Errors are accumulative
- Can only generate frames in powers(2)

METHODS TO LEARN THE INTERMEDIATE FRAME

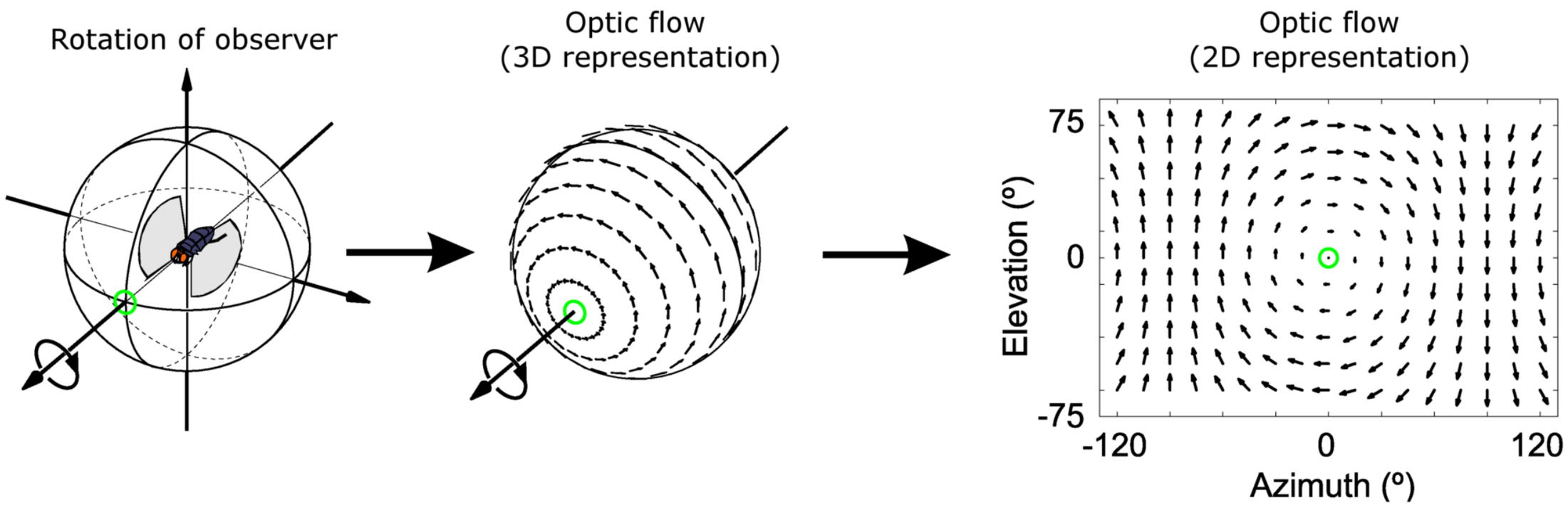
1. VIDEO INTERPOLATION



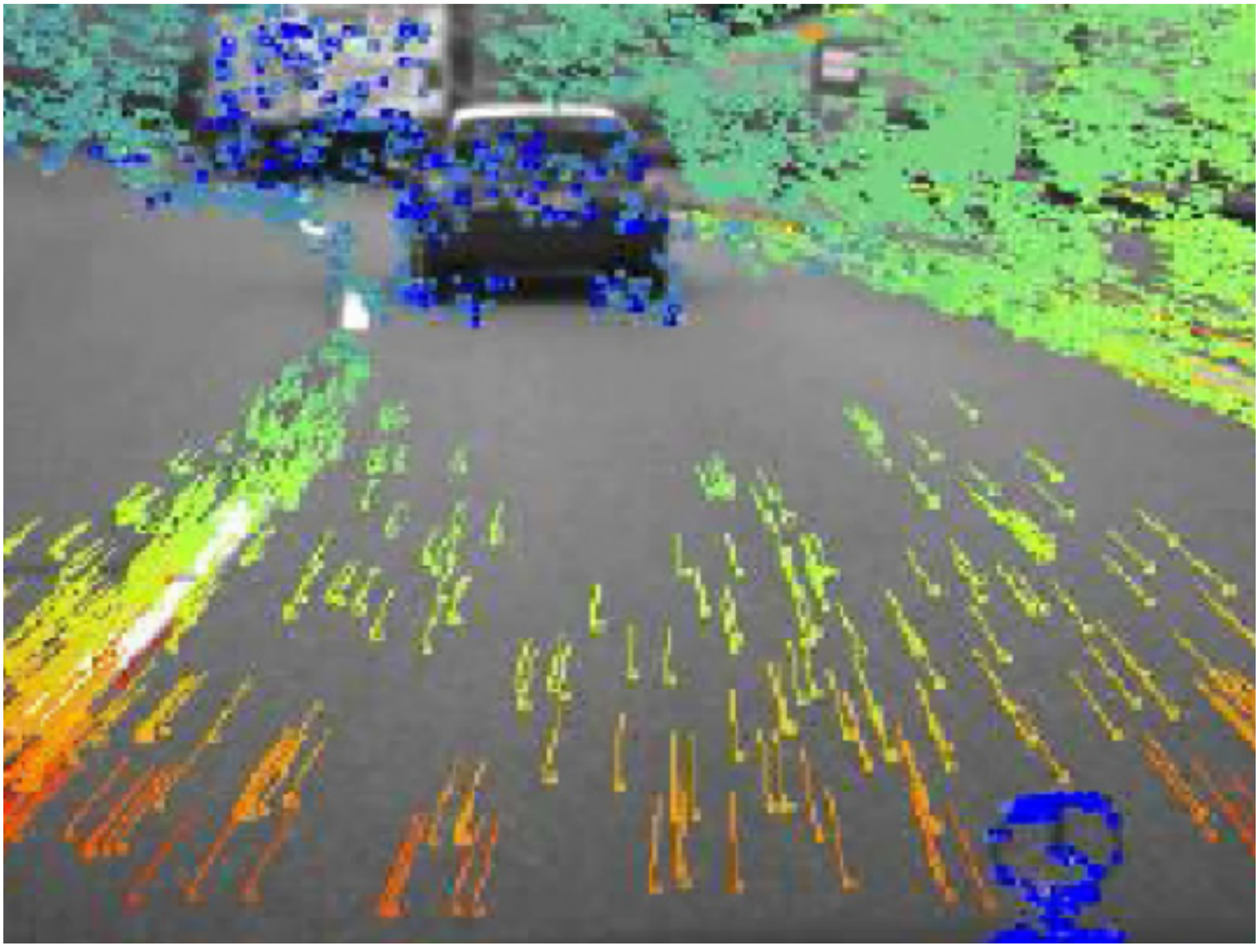
VIDEO INTERPOLATION (WORKS)

Methods	Drawbacks
Move the image gradients to a given time step and solve a Poisson equation to reconstruct the	Computationally expensive because of the complex optimization problems.
propagating phase information across oriented multi-scale pyramid levels for video interpolation.	Fails for high-frequency contents with large motions.
using frame interpolation as a supervision signal to learn CNN models for optical flow.	Memory intensive to predict a kernel for every pixel
improving the efficiency by predicting separable kernels.	But the motion that can be handled is limited by the kernel size (up to 51 pixels).
a CNN model for frame interpolation that has an explicit sub-network for motion estimation.	nNt well-suited for multi-frame interpolation.

2. OPTICAL FLOW



SUPER SLOMO: HIGH QUALITY ESTIMATION OF MULTIPLE INTERMEDIATE FRAMES FOR VIDEO INTERPOLATION



2. OPTICAL FLOW (WORKS)

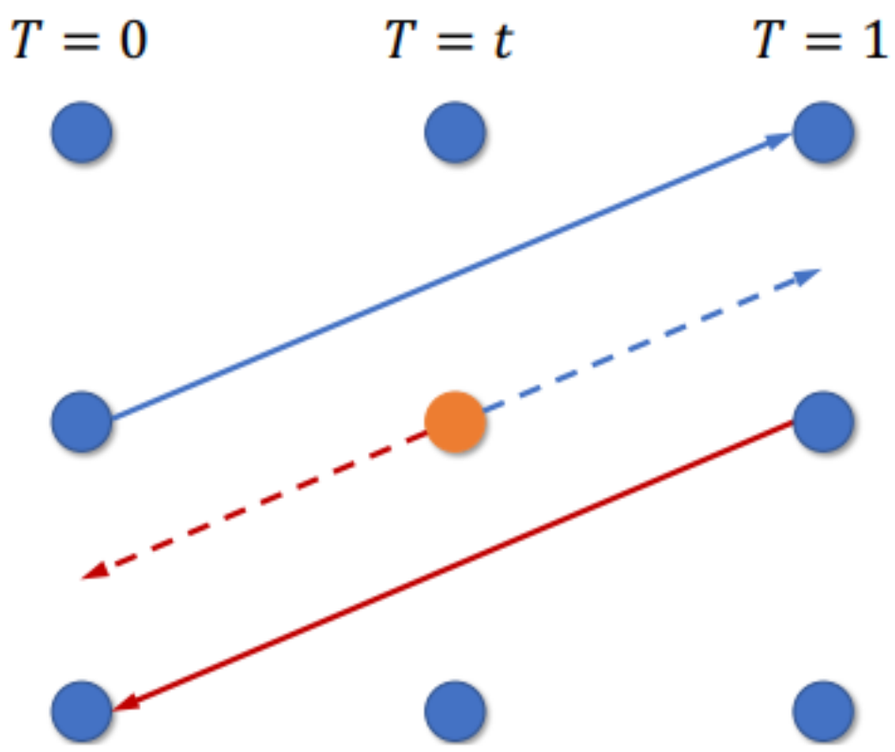
Methods	Drawbacks
Variational approach	Complex optimization function to solve.
Feature Matching	Learning limited to few parameters.
FLowNetC FlowNetS	Computation Cost

PROPOSED APPROACH OF USING THESE TWO FEATURES

1. INTERMEDIATE FRAME SYNTHESIS

Given two time frames i(0) and i(1), we want to predict an intermediate image i(p). where 0 < p < 1

for this we first try to find the warped images at time t = p.



Let $F_{t\to 0}$ and $F_{t\to 1}$ denote the optical flow from I_t to I_0 and I_t to I_1, respectively. If these two flow fields are known, we can synthesize the intermediate image $\hat{I}_t$ as follows:

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

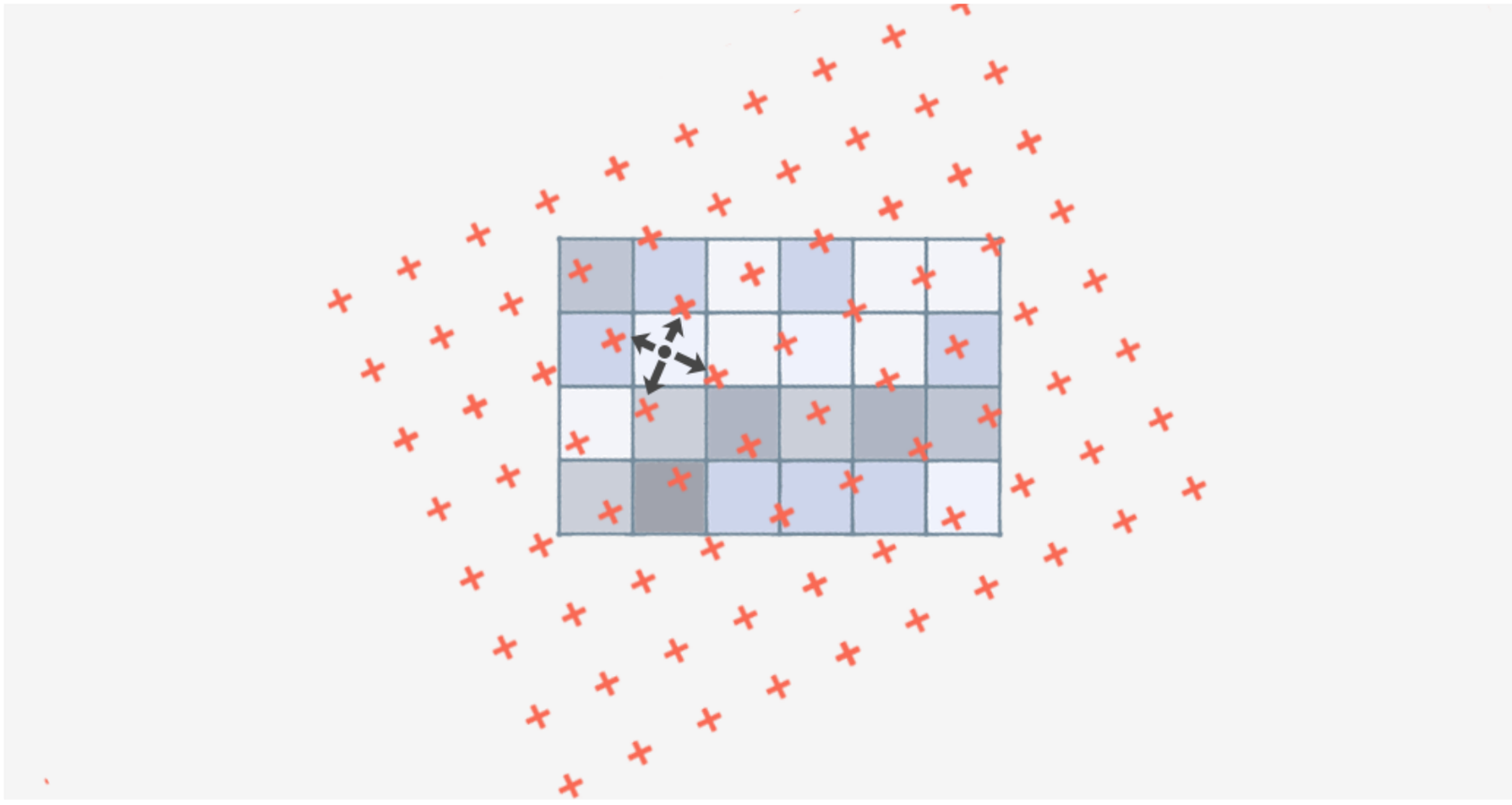
where,

g(.) : backward warping function

α_0 : temporal consistency and occlusion reasoning

$\odot$: Element wise multiplication

BILINEAR INTERPOLATION:



We introduce :

: *visibility maps* $V_{t\leftarrow 0}$ and $V_{t\leftarrow 1}$. $V_{t\leftarrow 0}(p) \in [0,1]$

Whether pixel P remains visible when moving from t= 0 to t= 1

Now combining occlusion reasoning and temporal consistency, we get,

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

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

2. TIME- FLOW INTERPRETATION

Since it is hard to compute the flow fields, we try to compute the intermediate optical flow optical flow between the two images.

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

or

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

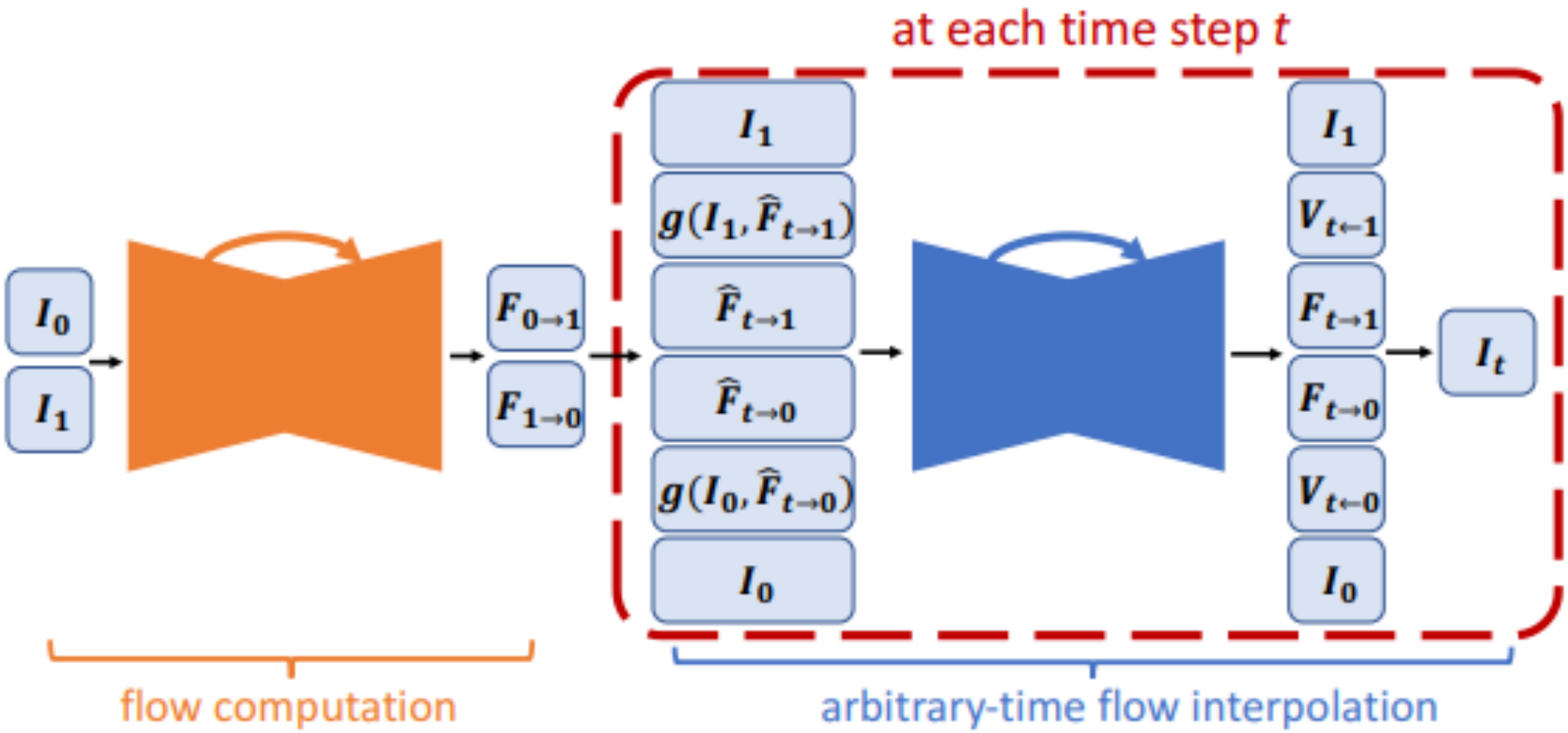
Similar to temporal consistency for image synthesis we can combine bi-directional input optical flow as follows:

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

Since occlusions are better handled by visibility maps, We have to predict the visibility map at V(t) using the flow interpolation CNN, and also enforce the following condition on it. (matting effect)

$$V_{t\leftarrow 0} = 1 - V_{t\leftarrow 1}.$$

ARCHITECTURE



For flow interpolation and flow computation we use U-Net architecture. It is a fully connected CNN, consisting of an encoder and a decoder.

LOSS FUNCTION

We take loss to be a linear combination like:

$$l = \lambda_r l_r + \lambda_p l_p + \lambda_w l_w + \lambda_s l_s.$$

Lr Reconstruction Loss
Lp Perceptual loss
Lw warping loss
Ls Smoothing loss

REFERENCES

1. Z. Liu, R. Yeh, X. Tang, Y. Liu, and A. Agarwala. Video frame synthesis using deep voxel flow. In ICCV, 2017.
2. G. Long, L. Kneip, J. M. Alvarez, H. Li, X. Zhang, and Q. Yu. Learning image matching by simply watching video. In ECCV, 2016.
3. B. Horn and B. Schunck. Determining optical flow. Artificial Intelligence
4. U-Net: Convolutional Networks for Biomedical Image Segmentation