# ReFIn: A Refinement Approach for Video Frame Interpolation

Saikat Dutta and Anurag Mittal

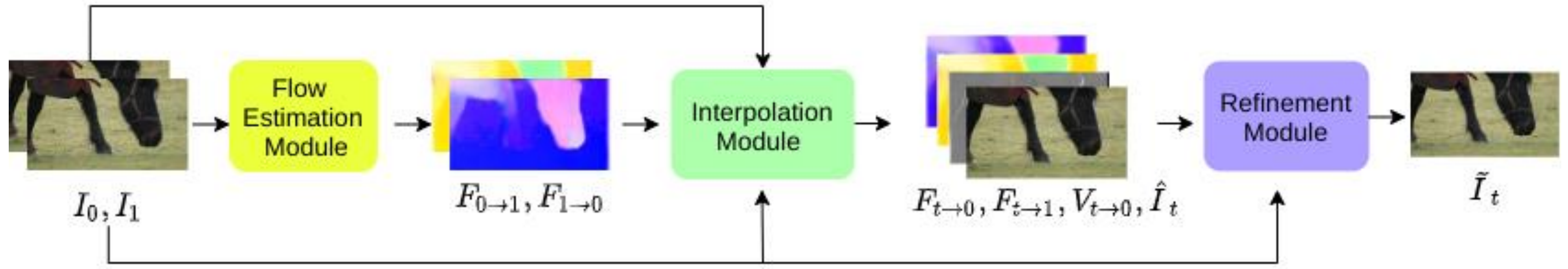
*IIT Madras, India*



## Introduction:

- Video frame interpolation algorithms generate one or multiple frames between two consecutive frames in a video.
- Optical flow-based frame interpolation approaches estimate intermediate optical flow from interpolated frame to input frames and warped frames are fused to generate interpolated frame.
- However, intermediate flow estimates can itself be erroneous leading to inaccurate interpolation results.
- In this work, we improve an optical flow-based interpolation algorithm, Super-SloMo by residual refinement.
- Specifically, we feed intermediate flowmaps, visibility map, warped input frames and intermediate interpolation estimate to a refinement network to predict a frame residual.
- We have also experimented with different architecture choices to be used in different modules to further improve the results.
- We found out that GridNet with four pyramid levels achieves the best results whereas UNet++ performs moderately well with significantly less number of parameters.

## Methodology:




- **Flow Estimation Module:**
  - First step of our algorithm is to estimate bidirectional optical flowmaps between input frames.
  - Instead of using UNet for flow estimation as in SuperSloMo, we use pretrained state-of-the-art flow estimator PWCNet.
- **Interpolation Module:**
  - After computing bidirectional optical flows using Flow Estimation module, we compute approximation of flow from intermediate image to input images.

$$\hat{F}_{t\to0} = -(1-t)tF_{0\to1} + t^2F_{1\to0}$$
$$\hat{F}_{t\to1} = (1-t)^2F_{0\to1} - t(1-t)F_{1\to0}$$

  - However, this approximation is not true in case of motion boundaries. Interpolation module acts as a correction step for computing final intermediate flowmaps.
  - This module generates residual flowmaps to refine the intermediate flowmaps.

$$F_{t\to0} = \hat{F}_{t\to0} + \triangle\hat{F}_{t\to0}$$
$$F_{t\to1} = \hat{F}_{t\to1} + \triangle\hat{F}_{t\to1}$$

  - Additionally, it computes soft visibility maps required for fusing the warped input frames.
  - Intermediate frame is synthesized as,

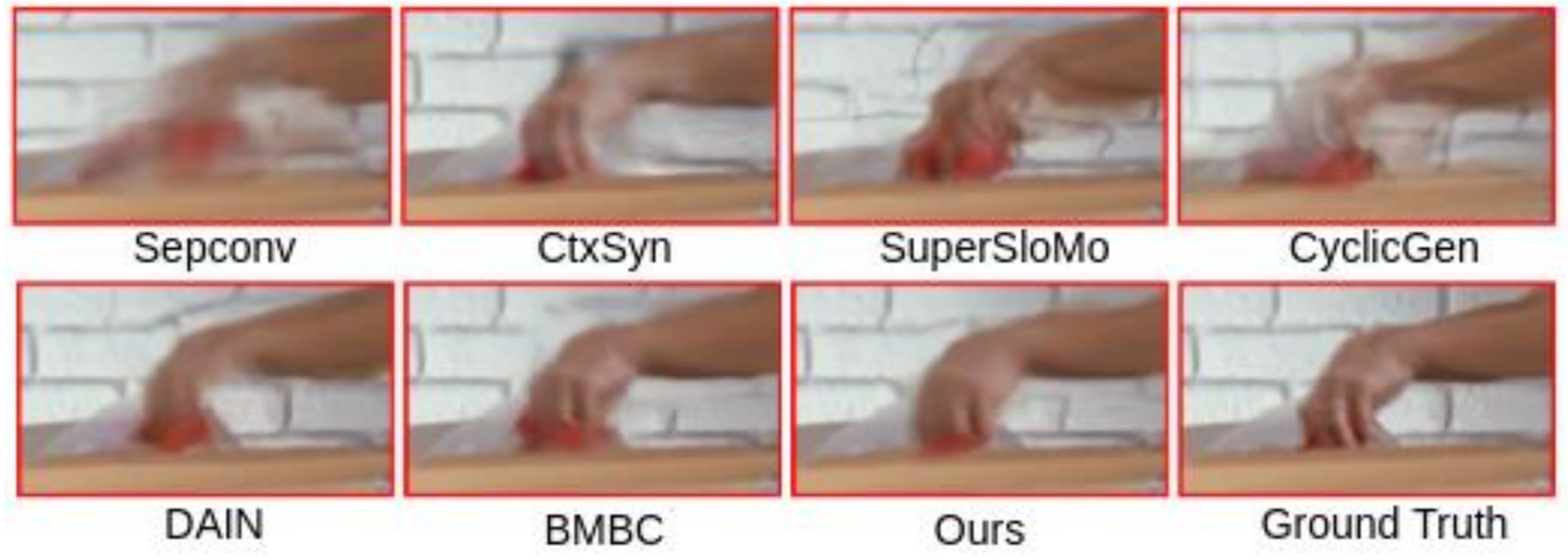
$$\hat{I}_t = \frac{(1-t)V_{t\to0} \odot bw\left(I_0, F_{t\to0}\right) + tV_{t\to1} \odot bw\left(I_1, F_{t\to1}\right)}{(1-t)V_{t\to0} + tV_{t\to1}}$$

- **Refinement Module:**
  - Although the formulation of intermediate frame is compact, the synthesized frame can still have erroneous predictions.
  - Hur et al. has shown that iterative residual refinement of optical flow predictions can improve optical flow accuracy.
  - Inspired by this, we introduce one step of refinement. We compute residual of the predicted frame using a Refinement module.
  - Final frame is synthesized as,

$$\tilde{I}_t = \hat{I}_t + \triangle\hat{I}_t$$

## Results:

| Method | Vimeo-Triplet | | UCF101 | |
|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM |
| TOFlow | 33.73 | 0.9682 | 34.58 | 0.9667 |
| SepConv | 33.79 | 0.9702 | 34.78 | 0.9669 |
| SuperSloMo | 33.44 | 0.9673 | 34.09 | 0.9655 |
| CtxSyn | 34.39 | 0.9610 | 34.62 | 0.9490 |
| CyclicGen | 32.10 | 0.9490 | 35.11 | 0.9680 |
| MEMC-Net | 34.29 | 0.9739 | 34.96 | 0.9682 |
| DAIN | 34.71 | 0.9756 | 34.99 | 0.9683 |
| CAIN | 34.65 | 0.9730 | 34.91 | 0.9690 |
| BMBC | 35.01 | 0.9764 | 35.15 | 0.9689 |
| AdaCoF | 34.35 | 0.9714 | 35.16 | 0.9689 |
| Ours (GridNet-4) | 35.15 | 0.9773 | 34.96 | 0.9690 |




## Conclusion:

- In this work, we adopt SuperSloMo and improve interpolation accuracy with the help of a refinement network.
- We show this refinement module can help in case of incorrect intermediate flow estimation.
- It can also help in enhancing edges and textures in the generated frame.
- We have explored a variety of network architectures to be used in Interpolation and Refinement modules.
- Our experiments show that GridNet-4 performs better than the rest.
- On the other hand, UNet++ achieves good performance while being parameter-efficient.