ACM multimedia

ALANET: Adaptive Latent Attention for Joint Video Deblurring and Interpolation



Akash Gupta Abhishek Aich Amit Roy-Chowdhury

University of California, Riverside

UC RIVERSIDE Bourns College of Engineering

VCG Video Computing Group Center for Robotics and Intelligent Systems

Problem Overview

- Generate **high frame-rate sharp** video from **low frame-rate poor** quality video.
- Learn video deblurring and interpolation jointly.

UCR Bourns College of Engineering
Video Computing Group

Problem Overview

- Generate **high frame-rate sharp** video from **low frame-rate poor** quality video.
- Learn video deblurring and interpolation jointly.

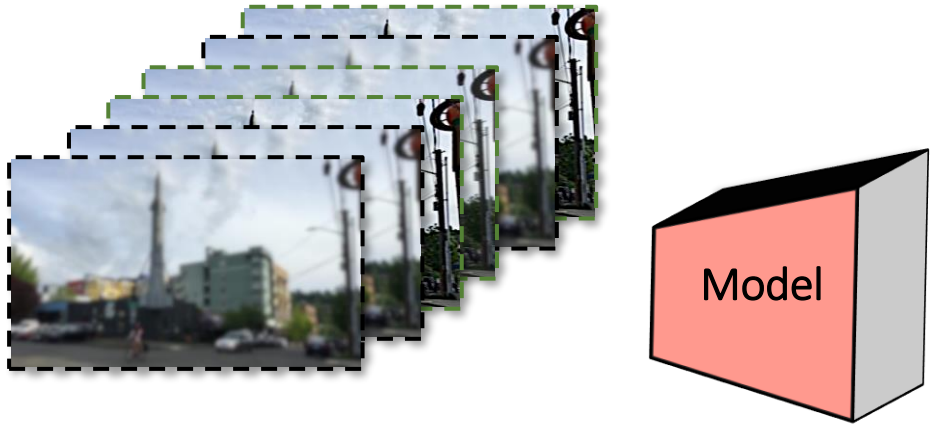
input frame | sharp frame | blur frame

output frame | sharp frame | interpolated frame

UCR Bourns College of Engineering
Video Computing Group

Problem Overview

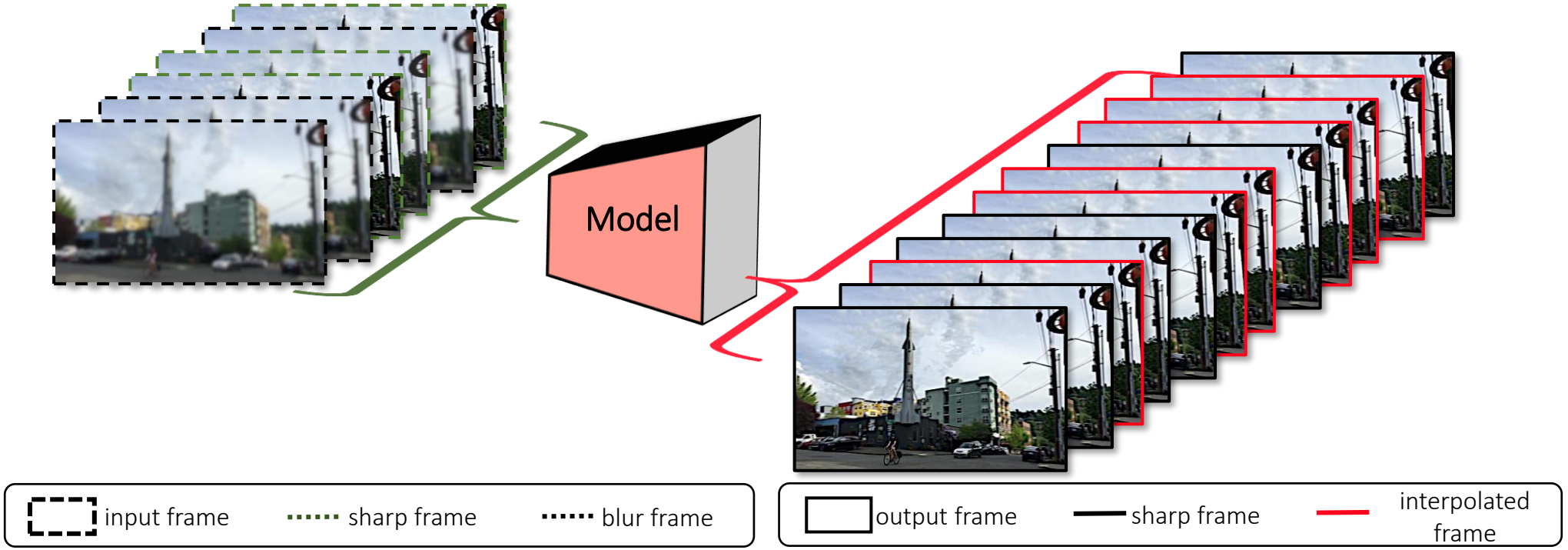
- Generate **high frame-rate sharp** video from **low frame-rate poor** quality video.
- Learn video deblurring and interpolation jointly.



UCR Bourns College of Engineering
Video Computing Group

Problem Overview

- Generate **high frame-rate sharp** video from **low frame-rate poor** quality video.
- Learn video deblurring and interpolation jointly.



UCR Bourns College of Engineering
Video Computing Group

Motivation

iPhone 13 Camera Specifications:

- **4K video** recording at maximum **60 fps**
- Slo-mo video **1080p** at **120 fps** or **240 fps**

24 fps

60 fps

GoPro Hero 7 Black (240 fps)

Pic Credits: Alisa Cassiel

Prior Works

- Unrealistic assumption that input video contains all blurry frame.
- The proposed approach uses latent attention for video synthesis.

Table 1: **Categorization of prior works**. ALANET demonstrates adaptive attention in latent space to perform joint deblurring and interpolation.

Methods	Settings			
	Interpolate?	Deblur?	Joint Deblur & Interpolate?	Latent Attention?
DAIN[1]	✓	✗	✗	✗
Jin[2]	✓	✓	✗	✗
BIN[3]	✓	✓	✓	✗
ALANET	✓	✓	✓	✓

[1] Wenbo Bao et al. "Depth-aware video frame interpolation". *CVPR*. 2019.

[2] Meiguang Jin et al. "Learning to extract a video sequence from a single motion-blurred image". *CVPR*. 2018.

[3] Wang Shen et al. "Blurry Video Frame Interpolation". 2020. arXiv: 2002.12259.

Conceptual Overview

How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?

Conceptual Overview

How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?

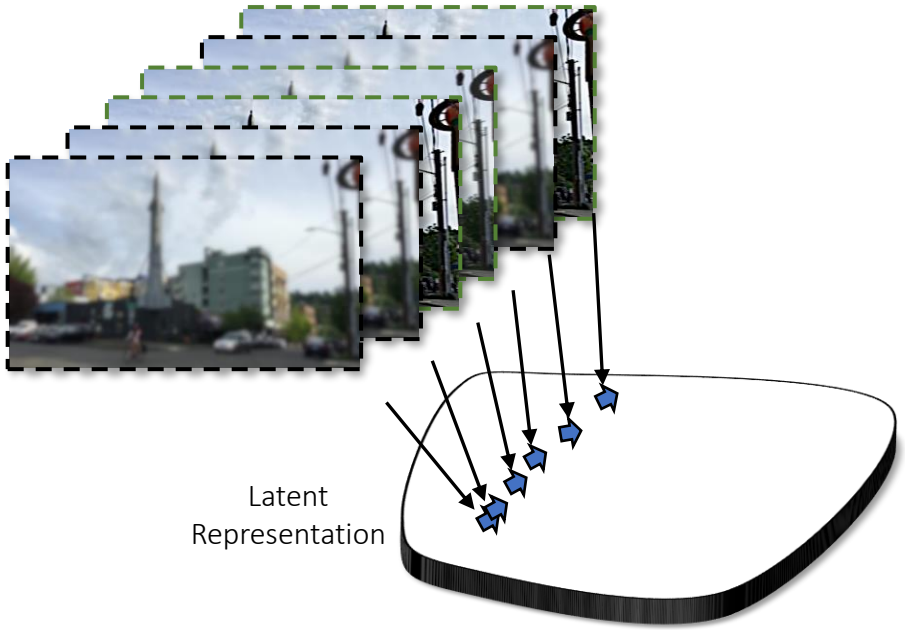
input frame | sharp frame | blur frame

output frame | sharp frame | interpolated frame

Conceptual Overview

How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?



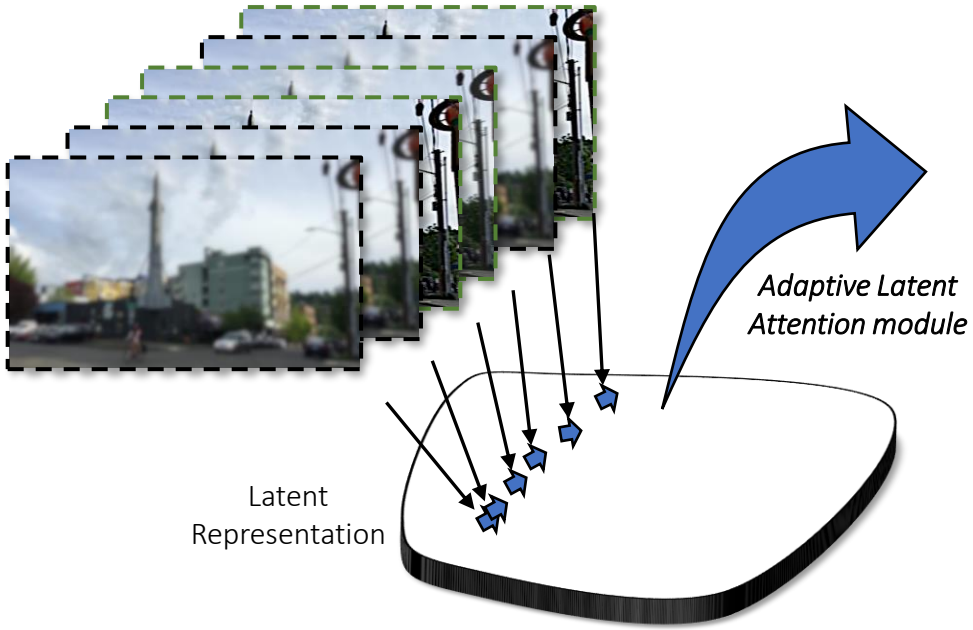
input frame | sharp frame | blur frame

output frame | sharp frame | interpolated frame

Conceptual Overview

How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?



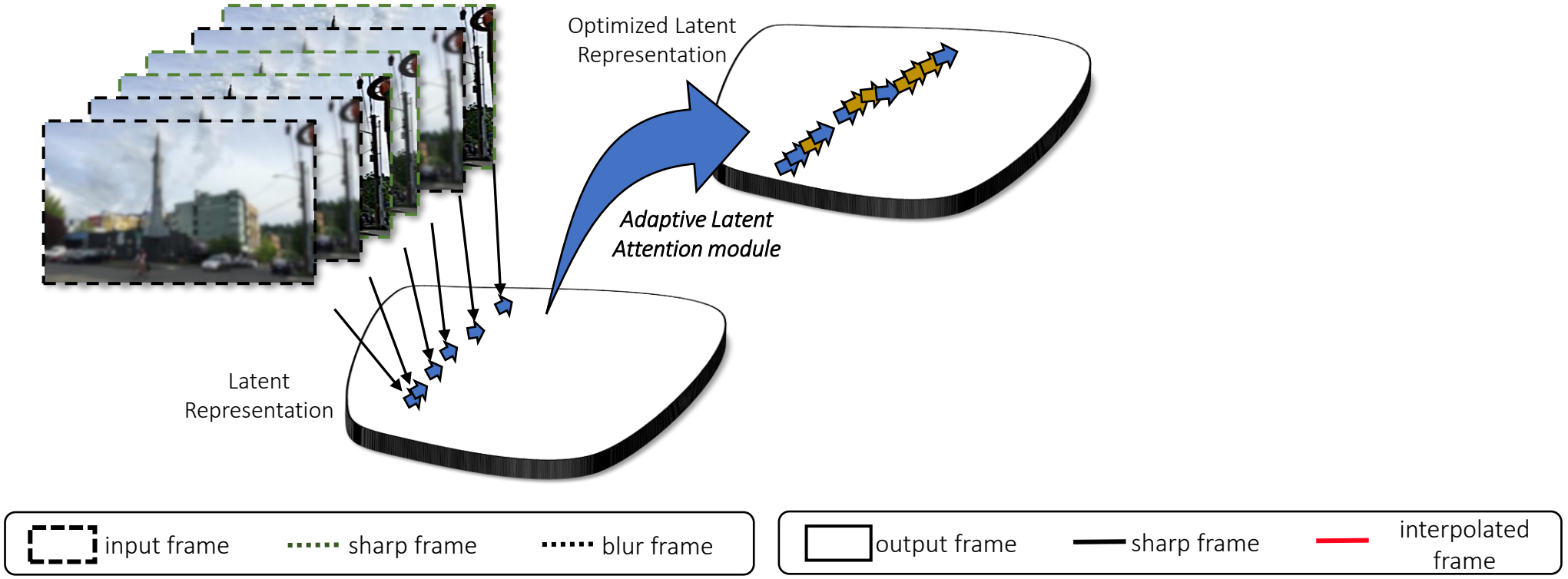
input frame | sharp frame | blur frame

output frame | sharp frame | interpolated frame

Conceptual Overview

How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?

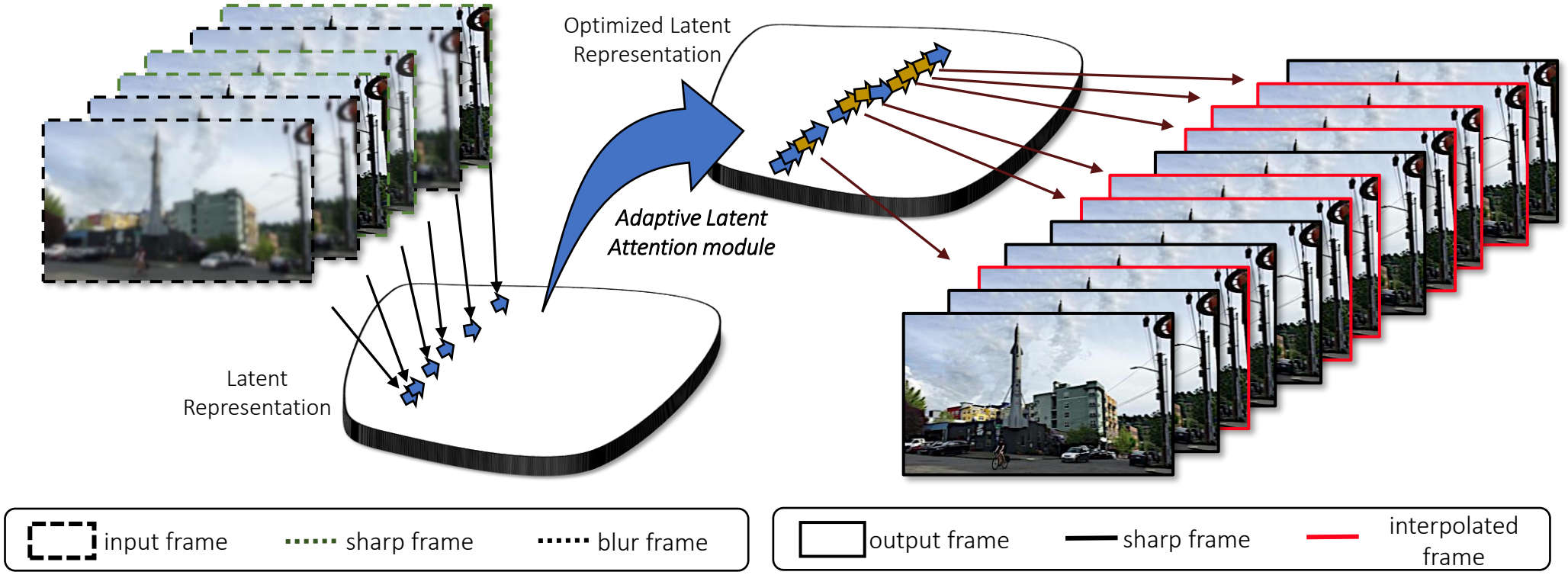


UCR Bourns College of Engineering
Video Computing Group

Conceptual Overview

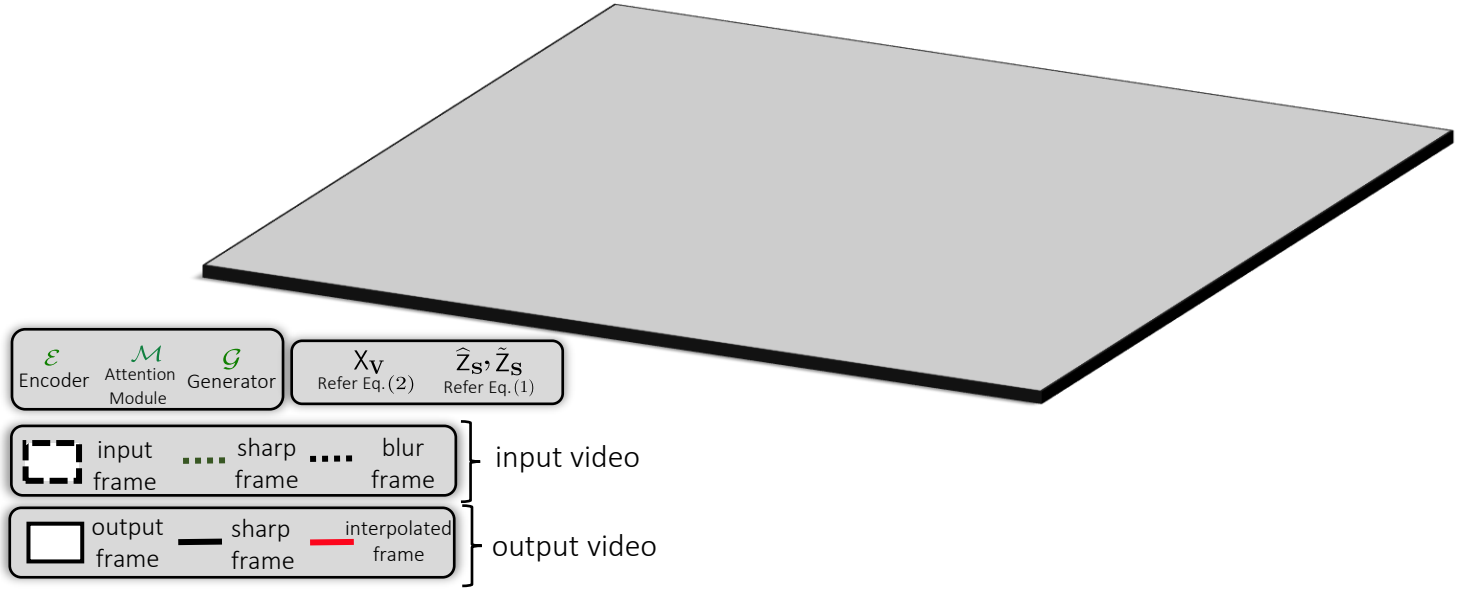
How to jointly deblur and increase frame-rate?

Can you utilize information within the poor-quality video to increase frame rate and deblur?



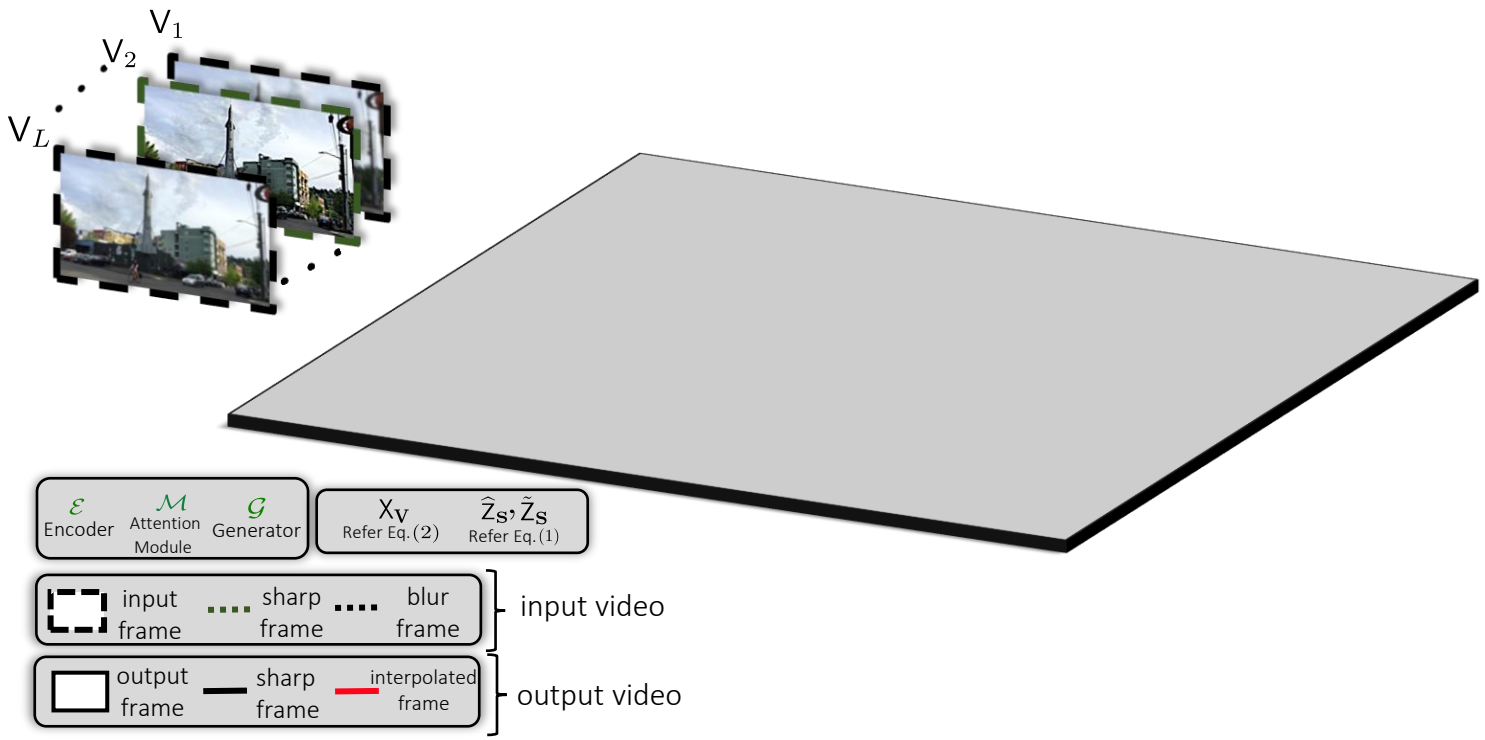
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



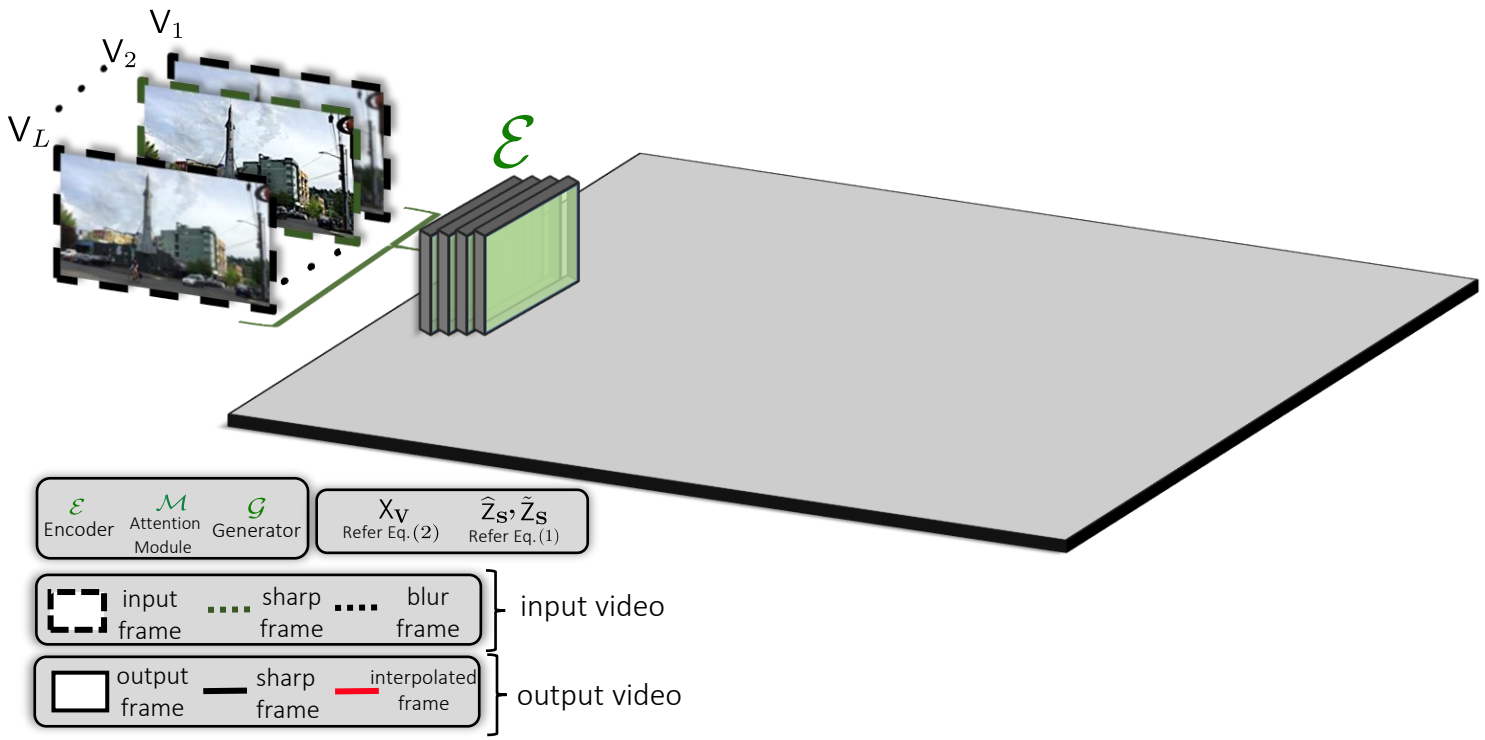
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach

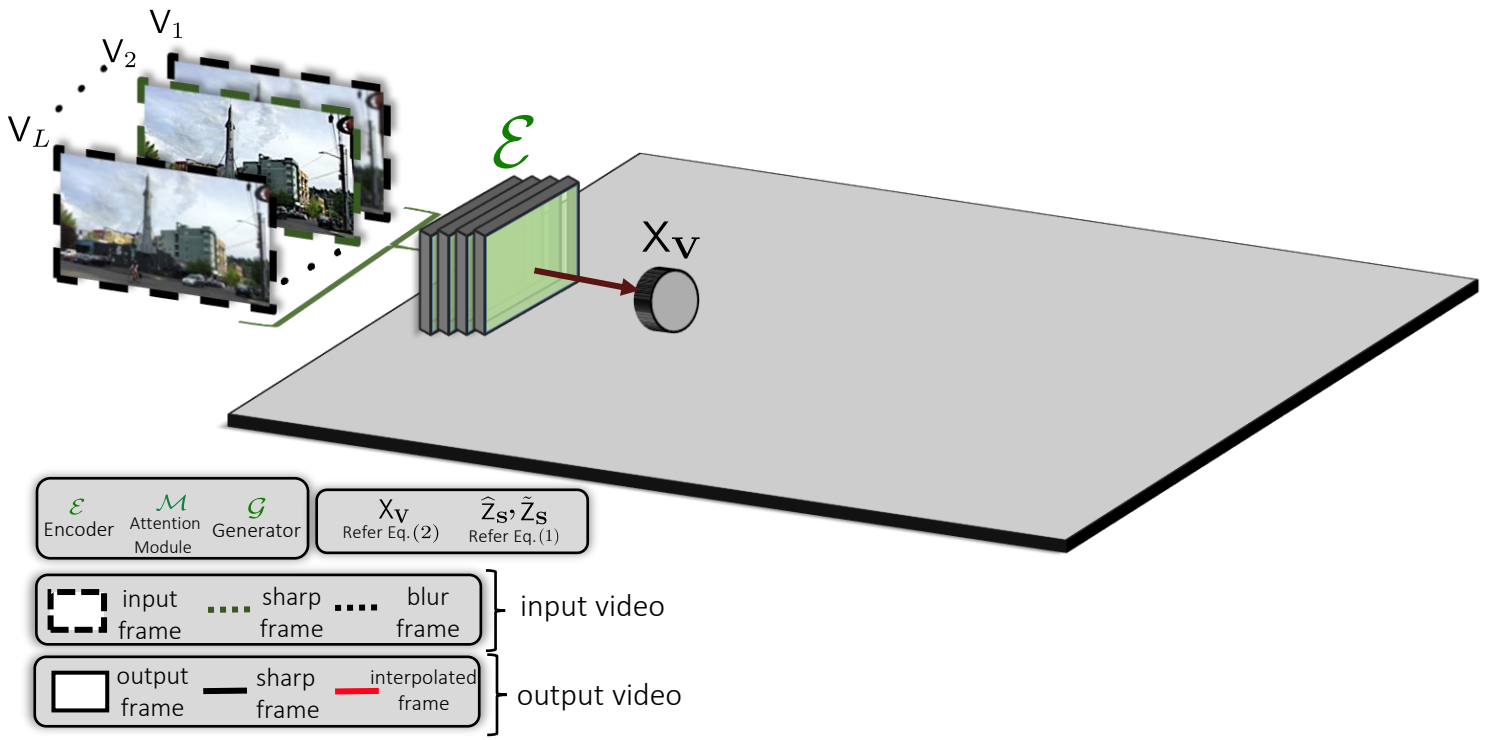


UCR Bourns College of Engineering
Video Computing Group

Proposed Approach

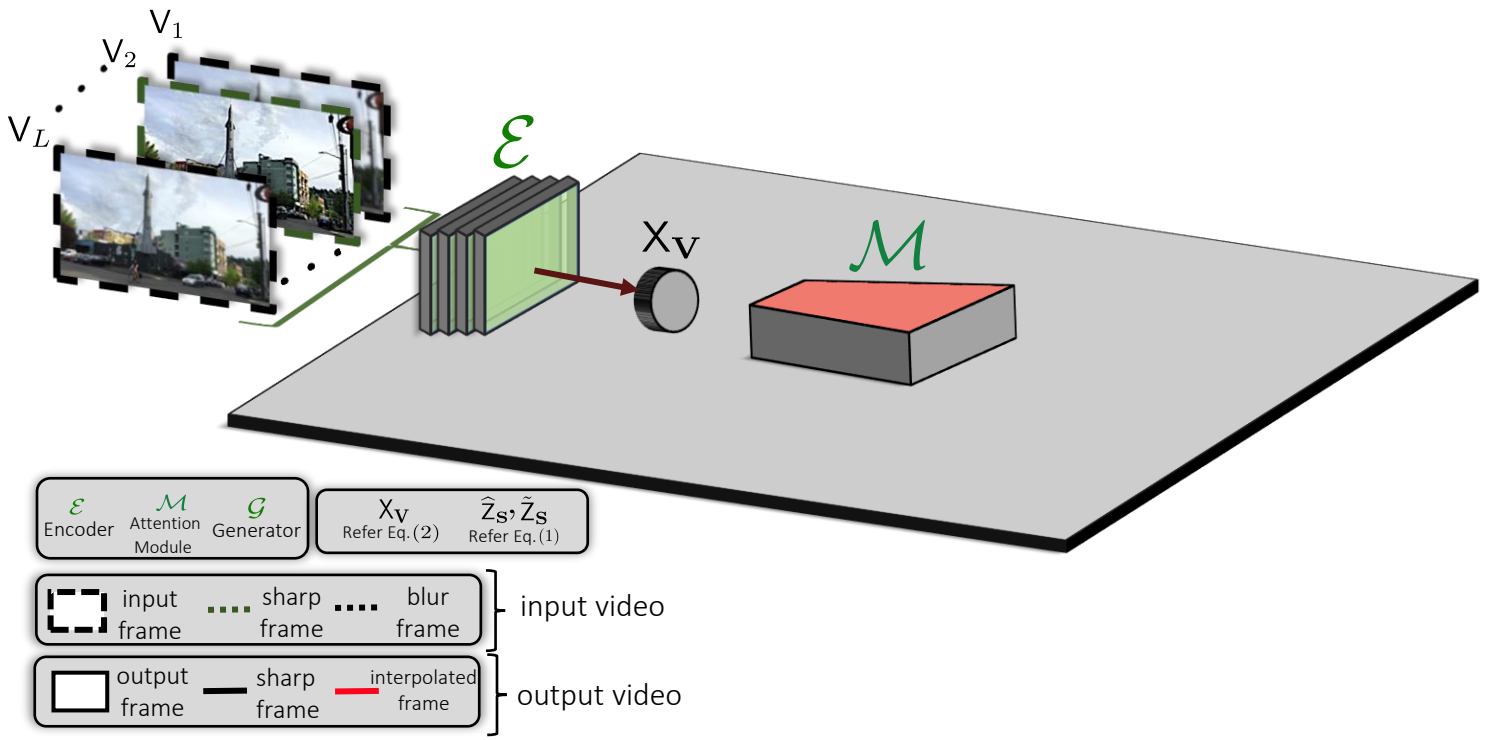


Proposed Approach



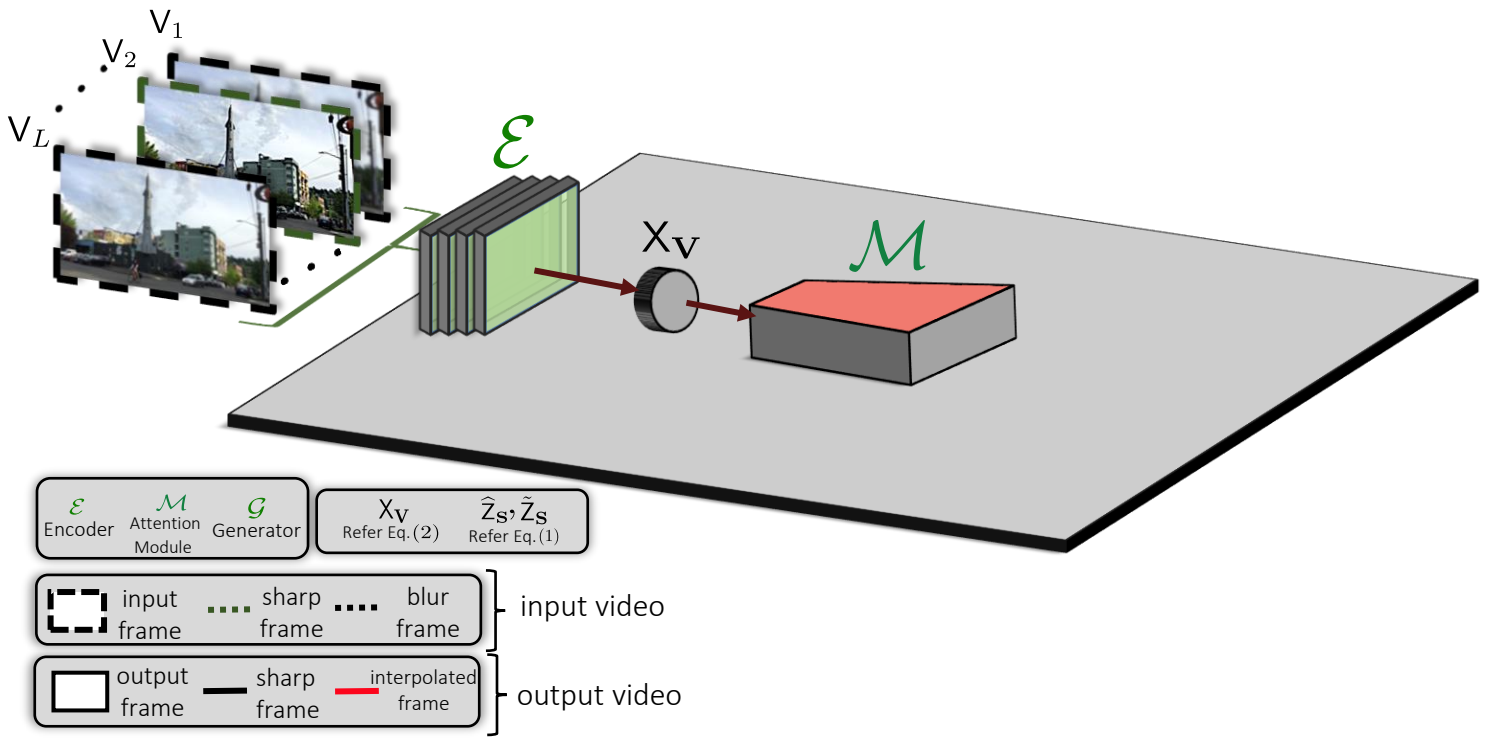
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



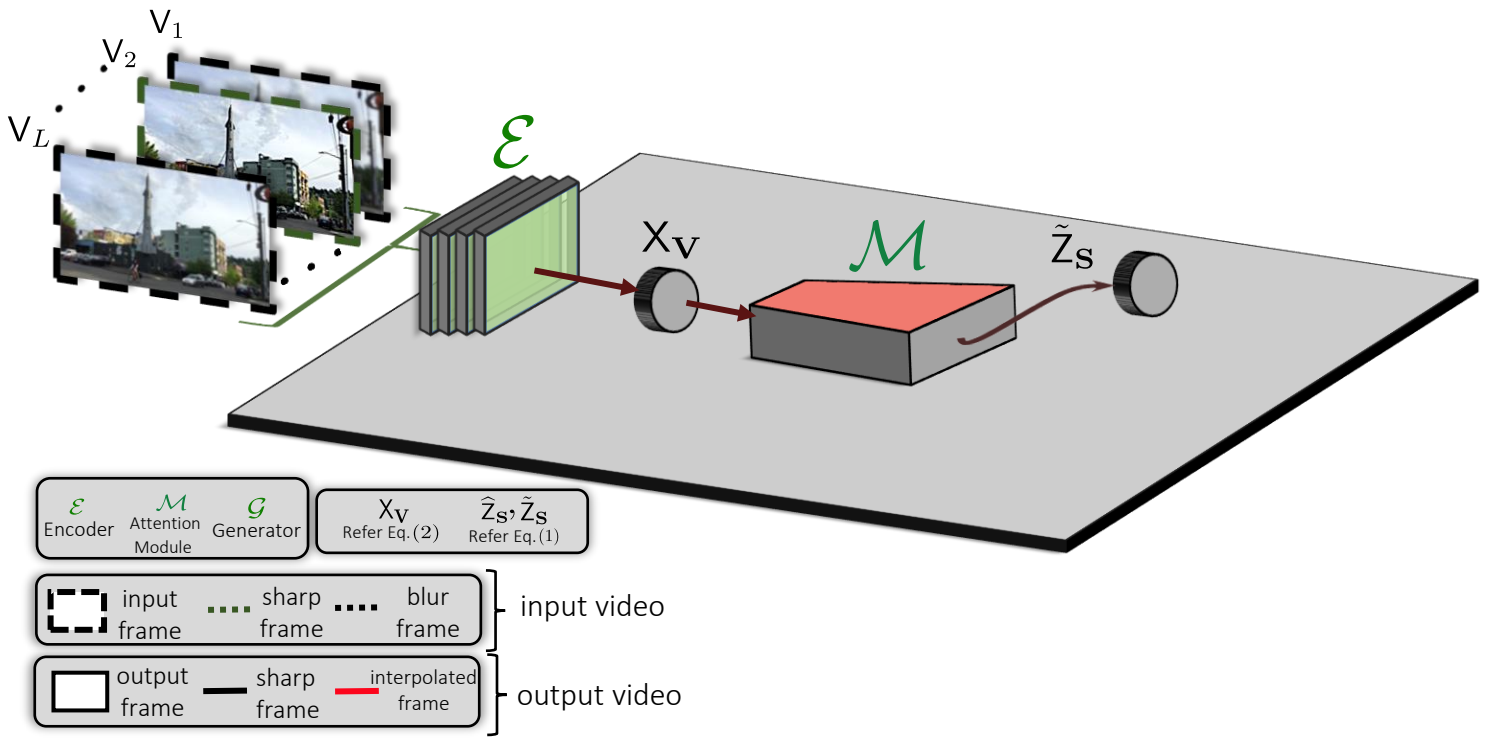
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



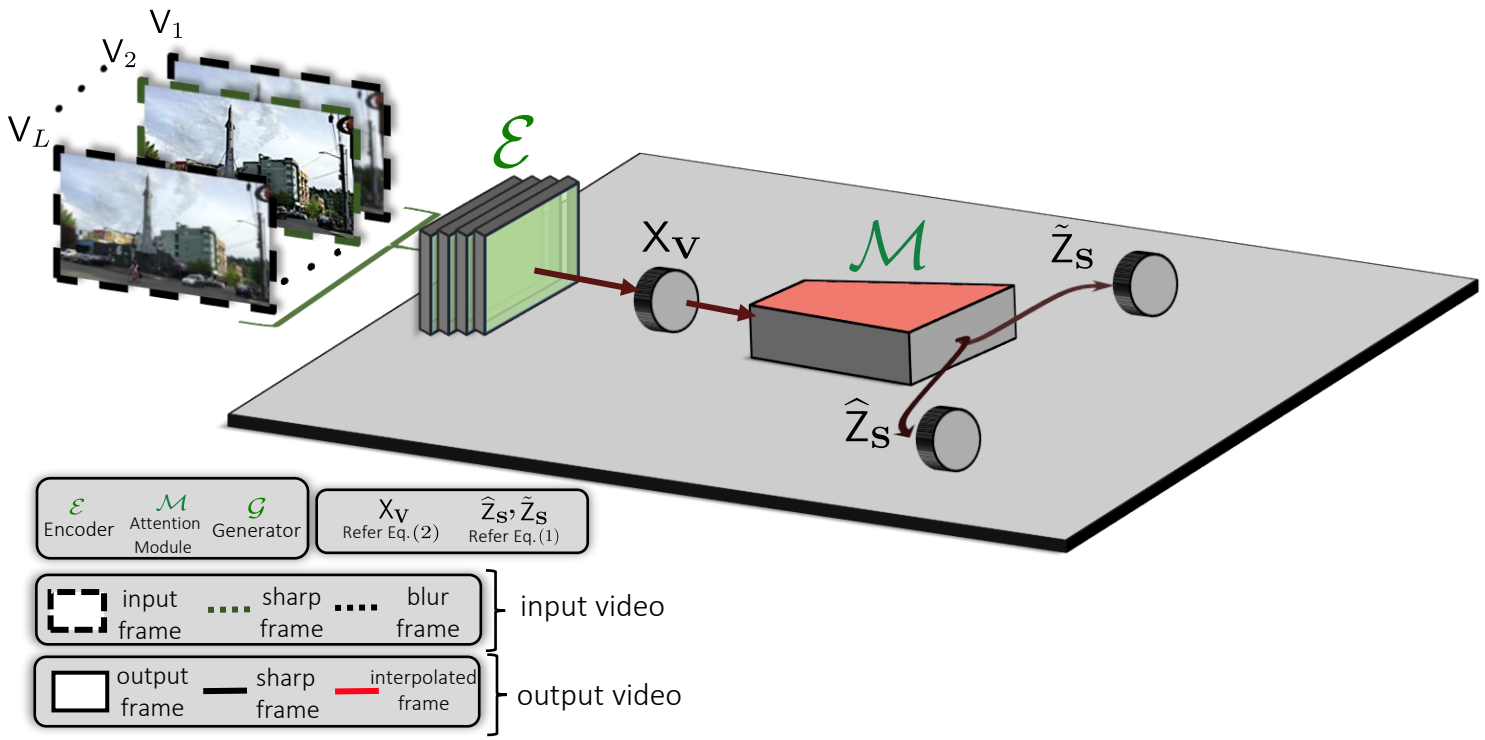
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



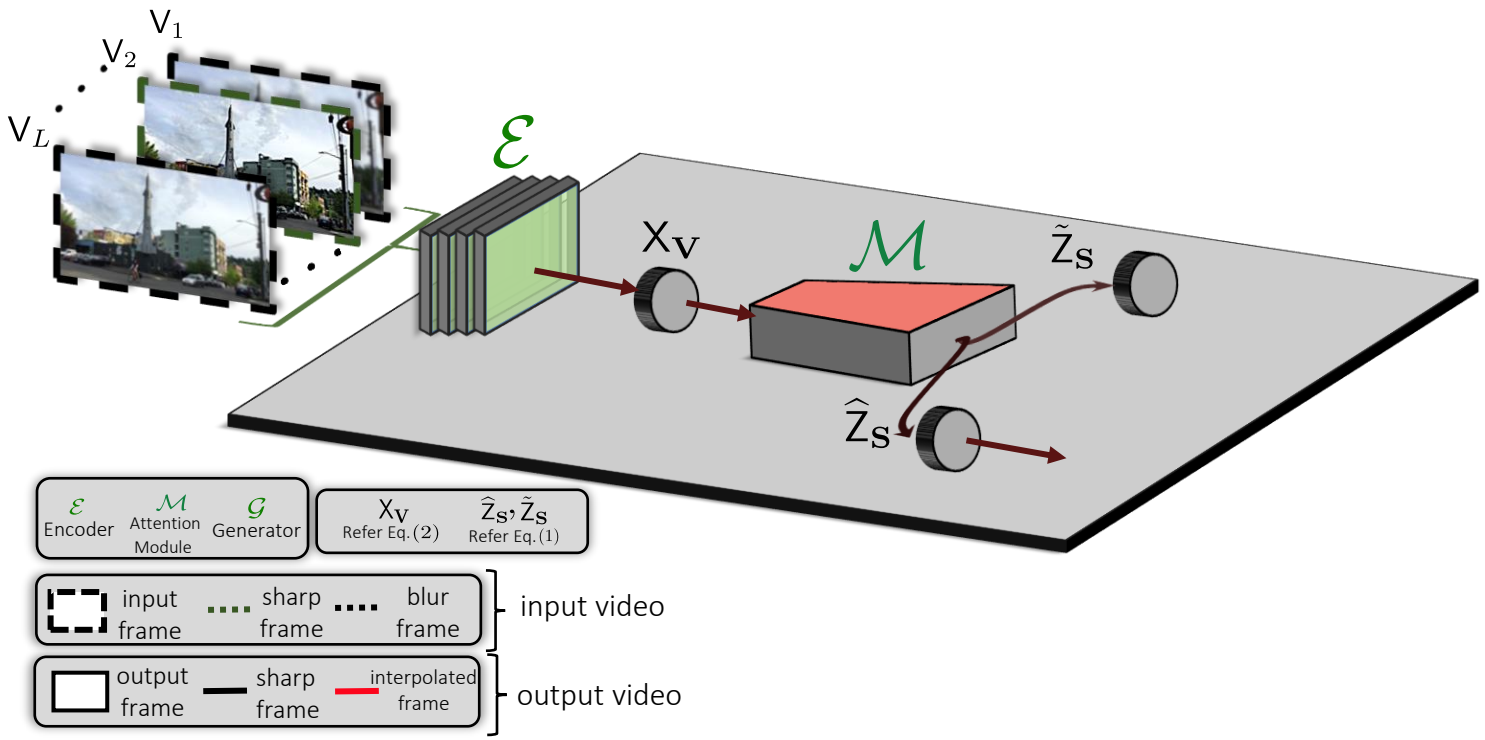
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



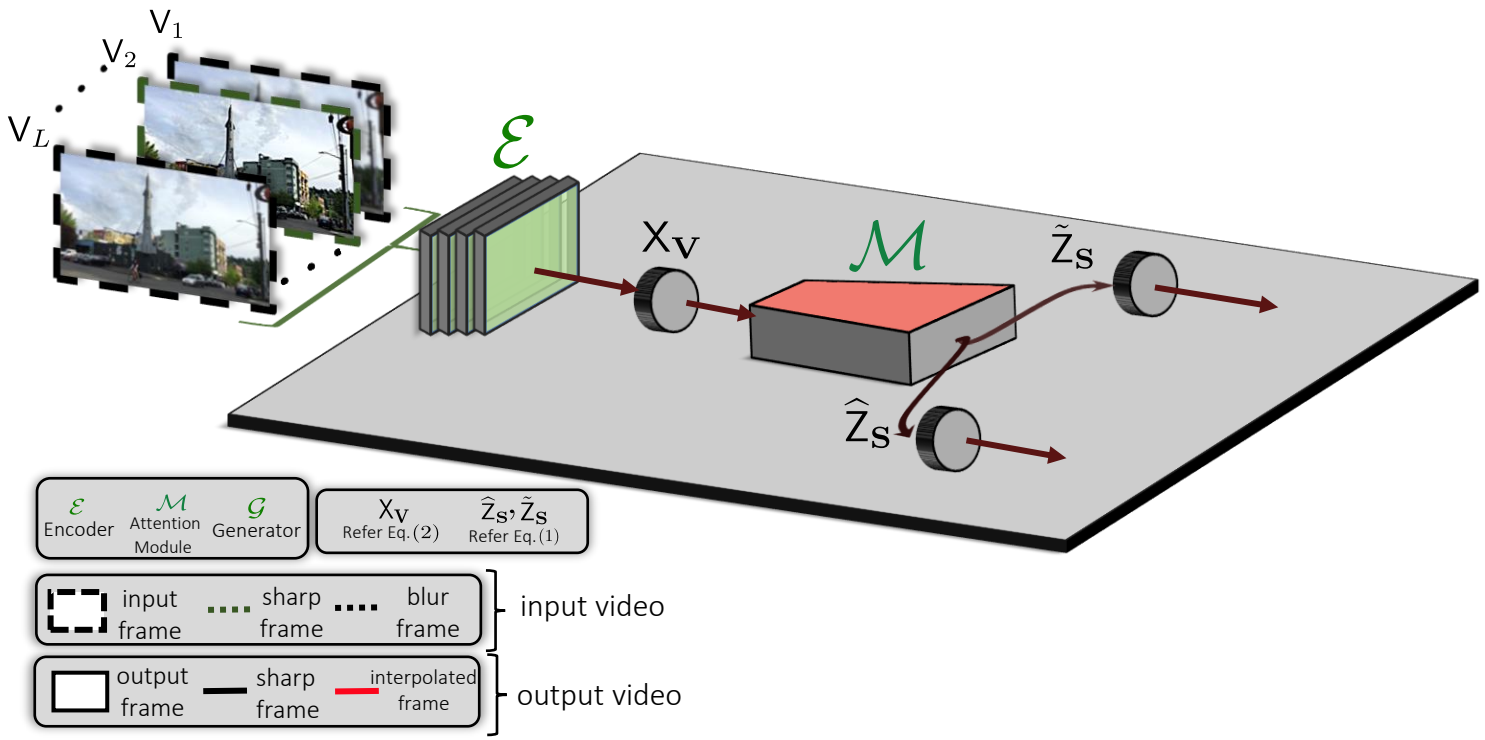
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



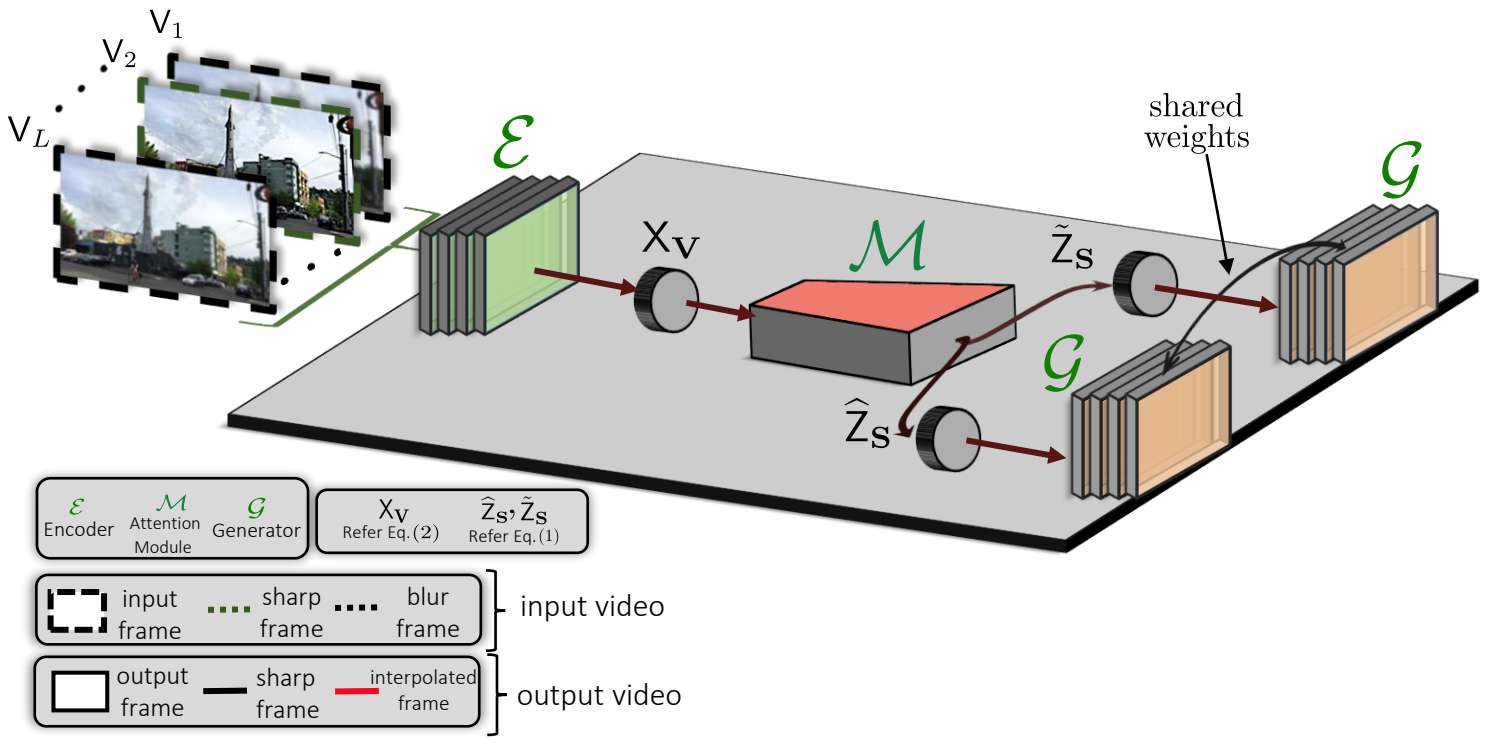
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



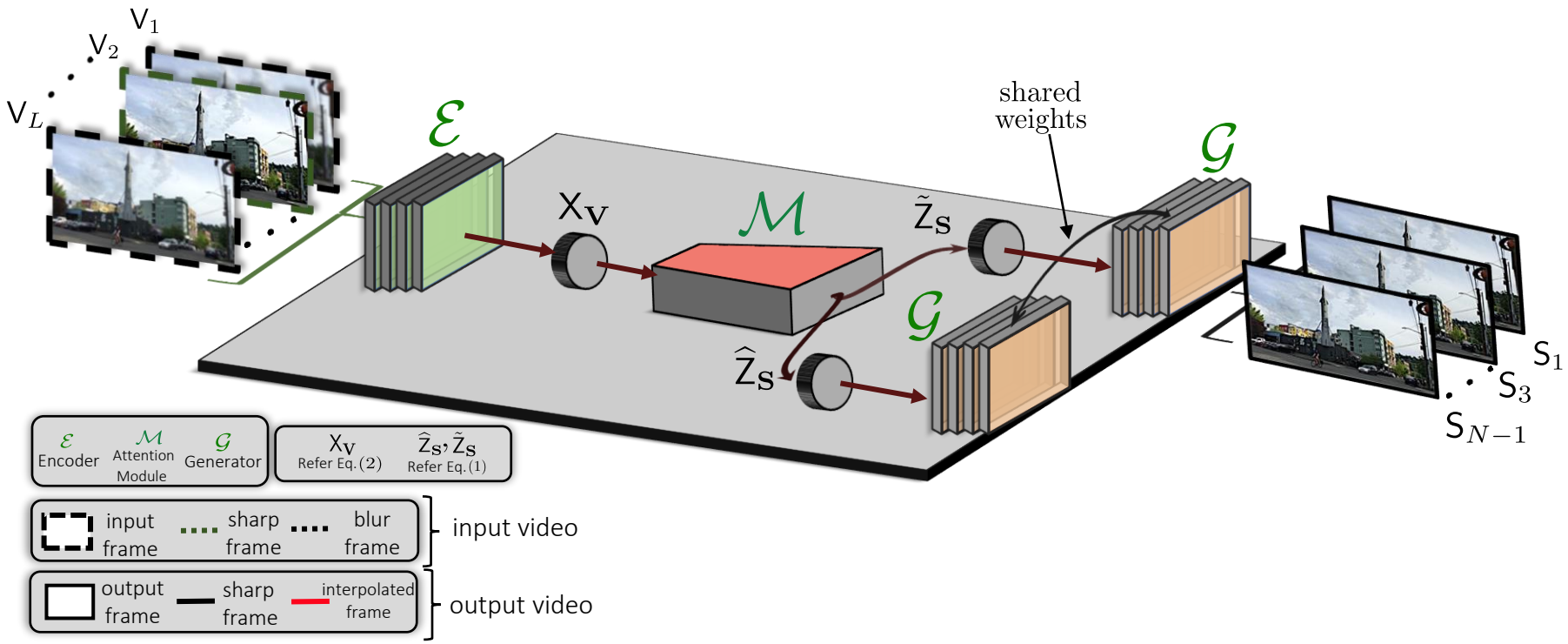
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



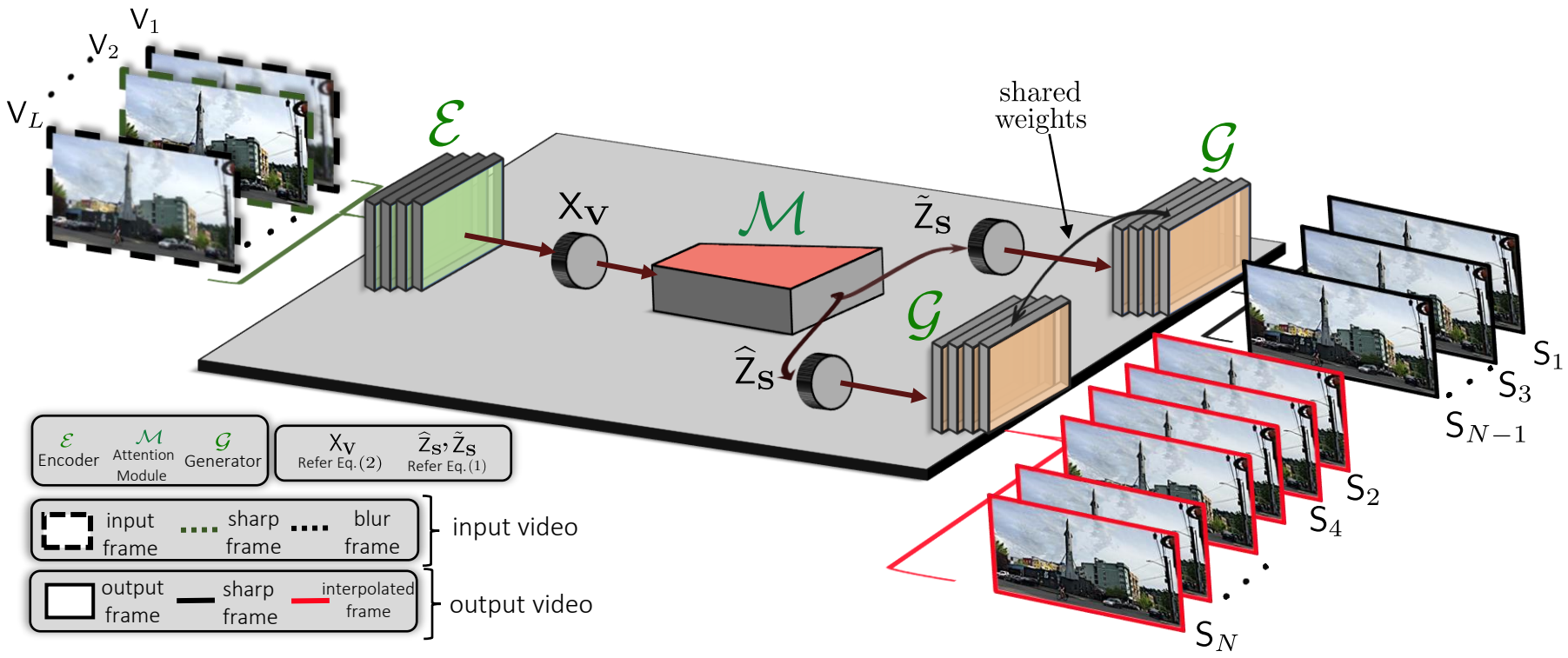
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach



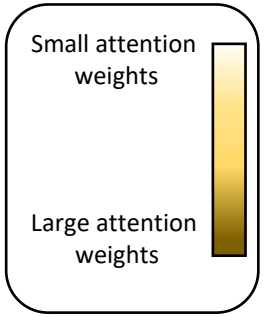
UCR Bourns College of Engineering
Video Computing Group

Proposed Approach

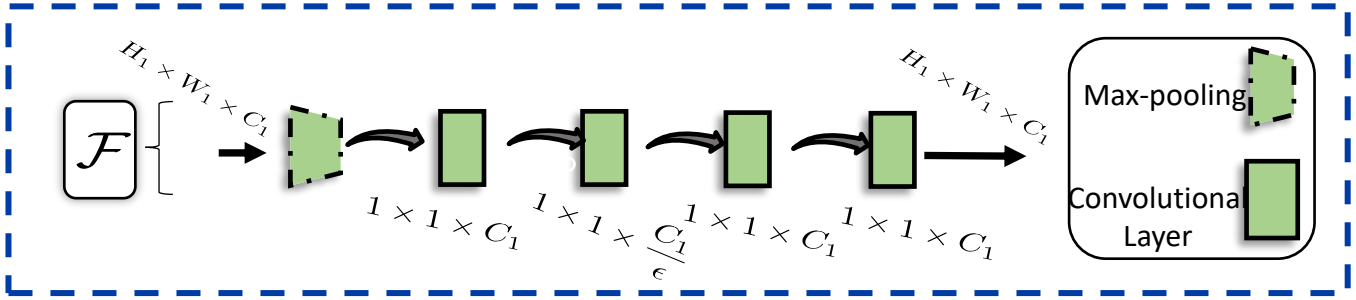


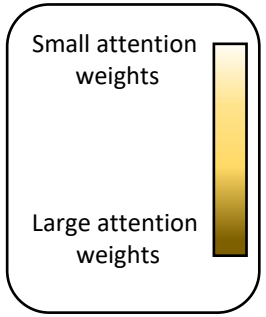
UCR Bourns College of Engineering
Video Computing Group

Attention Mechanisms

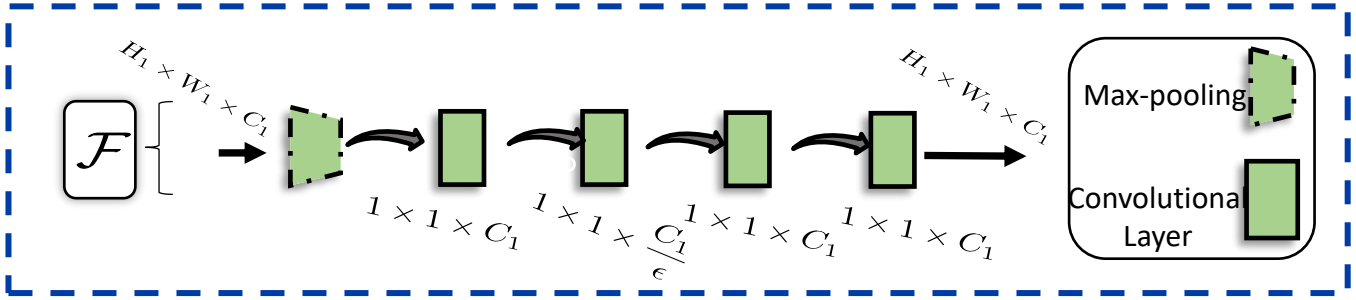


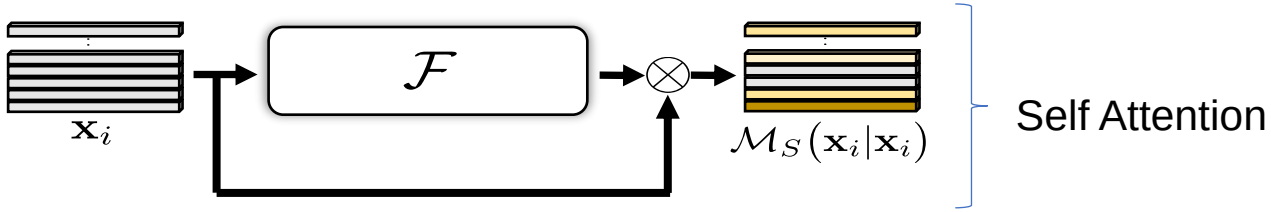
Attention Mechanisms

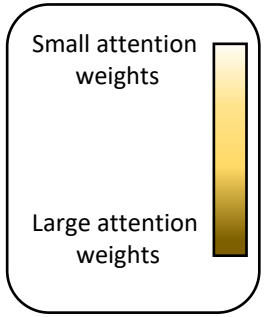




Attention Mechanisms

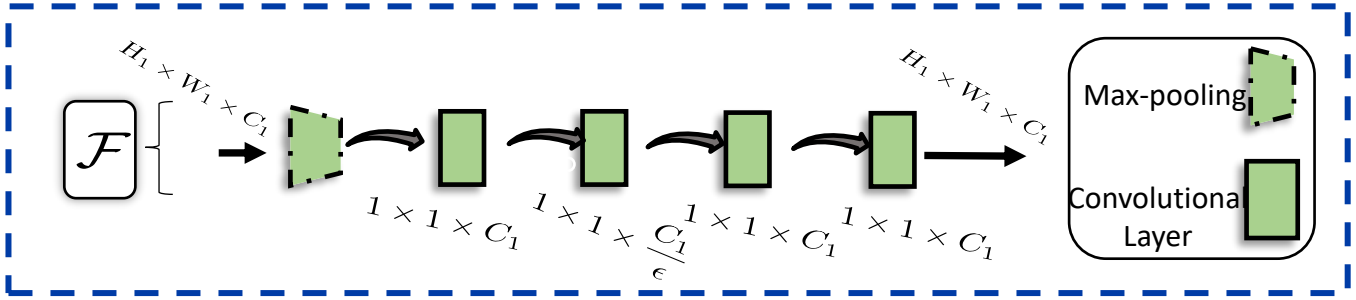


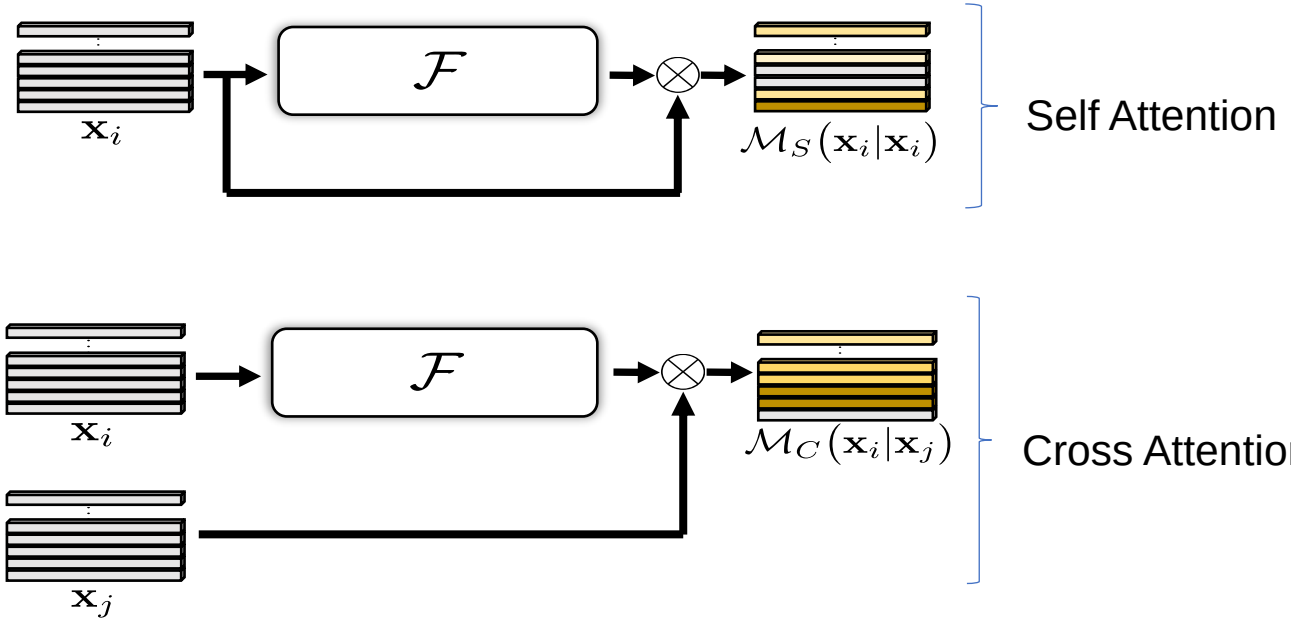


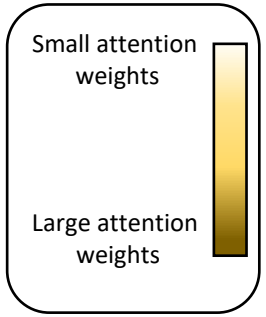


UCR Bourns College of Engineering
Video Computing Group

Attention Mechanisms







UCR Bourns College of Engineering
Video Computing Group

Adaptive Latent Attention

Combination of self-attention and cross-attention is used for deblurring and interpolation

Self-Attention: $\mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) = \mathbf{x}_i \otimes \mathcal{F}(\mathbf{x}_i)$

Cross-Attention: $\mathcal{M}_C\left(\mathbf{x}_j|\mathbf{x}_i\right) = \mathbf{x}_j \otimes \mathcal{F}(\mathbf{x}_i)$

UCR Bourns College of Engineering
Video Computing Group

Adaptive Latent Attention

Combination of self-attention and cross-attention is used for deblurring and interpolation

For deblurring:

$$\mathbf{z}_{2i} = \mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) + \sum_{j\in\mathbb{Q}} \mathcal{M}_C\left(\mathbf{x}_j|\mathbf{x}_i\right)$$

Self-Attention: $\mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) = \mathbf{x}_i \otimes \mathcal{F}(\mathbf{x}_i)$

Cross-Attention: $\mathcal{M}_C\left(\mathbf{x}_j|\mathbf{x}_i\right) = \mathbf{x}_j \otimes \mathcal{F}(\mathbf{x}_i)$

UCR Bourns College of Engineering
Video Computing Group

Adaptive Latent Attention

Combination of self-attention and cross-attention is used for deblurring and interpolation

Self-Attention: $\mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) = \mathbf{x}_i \otimes \mathcal{F}(\mathbf{x}_i)$

Cross-Attention: $\mathcal{M}_C\left(\mathbf{x}_j|\mathbf{x}_i\right) = \mathbf{x}_j \otimes \mathcal{F}(\mathbf{x}_i)$

For deblurring:

$$\mathbf{z}_{2i} = \mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) + \sum_{j \in \mathbb{Q}} \mathcal{M}_C\left(\mathbf{x}_j|\mathbf{x}_i\right)$$

For interpolation:

$$\widehat{\mathbf{z}}_{2i+1} = \mathcal{M}_S\left(\mathbf{x}_i|\mathbf{x}_i\right) + \mathcal{M}_C\left(\mathbf{x}_i|\mathbf{x}_{i+1}\right) + \mathcal{M}_S\left(\mathbf{x}_{i+1}|\mathbf{x}_{i+1}\right) + \mathcal{M}_C\left(\mathbf{x}_{i+1}|\mathbf{x}_i\right)$$

Training Loss

Our objective function consists of

- ℓ_1 pixel reconstruction loss : $\mathcal{L}_r = \sum_i |\mathsf{G}_i - \mathsf{S}_i|_1$
  - $\mathsf{G}_i \rightarrow$ Ground truth, $\mathsf{S}_i \rightarrow$ generated frame
- perceptual loss[4] $\mathcal{L}_p$: computed using a pre-trained VGG16 network

$$\mathcal{L} = \mathcal{L}_r + \lambda \mathcal{L}_p$$

We use $\lambda = 0.2$ for all our experiments.

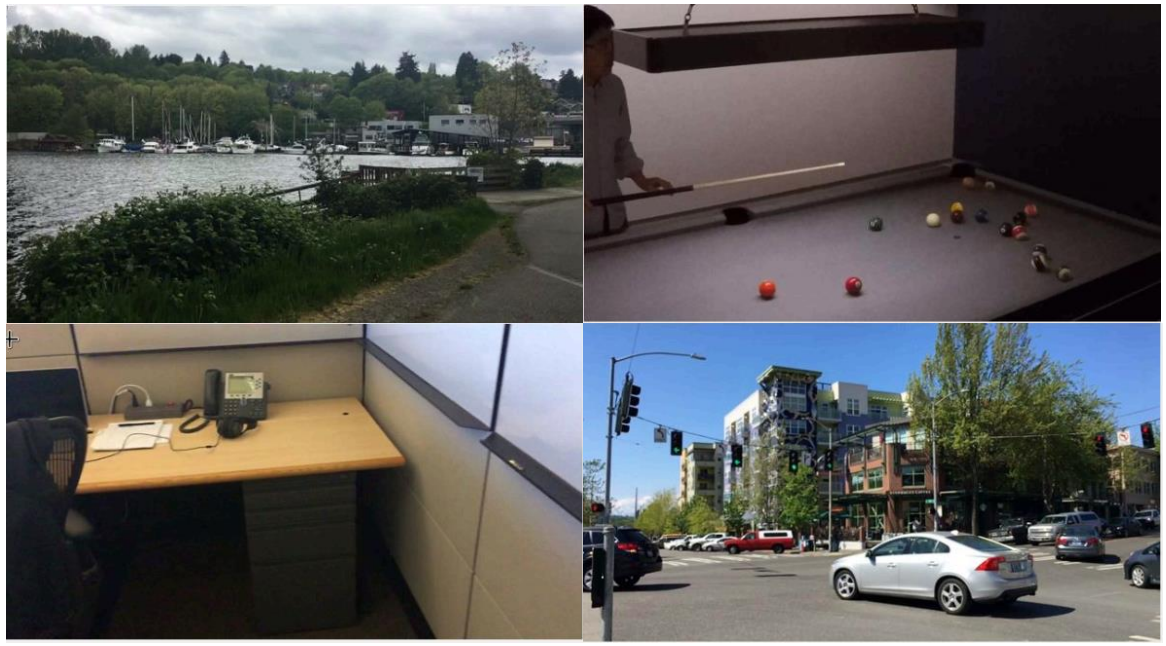
Datasets

Adobe240 Dataset

of videos: 118 @ 240 fps (100 training, 8 testing)

Resolution: 1280 × 720

Training video is reduced to 30fps



YouTube240 Dataset

Random 50 videos from YouTube @ 240 fps (for testing)

Resolution: 1280 × 720

Training video is reduced to 30fps

Quantitative Results

Adobe240: Improvement of **+2.3%** in terms of PSNR

Table 2: **Quantitative results comparison.** Best scores have been highlighted in bold. † indicates results reported from[3].

Method	Deblurring				Interpolation				Joint Deblurring and Interpolation			
	Adobe240		YouTube240		Adobe240		YouTube240		Adobe240		YouTube240	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Blurry Inputs†	28.68	0.8584	31.96	0.9119	-	-	-	-	-	-	-	-
Super SloMo† [5]	-	-	-	-	27.52	0.8593	30.84	0.9107	-	-	-	-
MEMC-Net† [6]	-	-	-	-	30.83	0.9128	34.91	0.9596	-	-	-	-
DAIN† [1]	-	-	-	-	31.03	0.9172	35.09	0.9615	-	-	-	-
Jin† [2]	29.40	0.8734	32.06	0.9119	29.24	0.8754	32.24	0.9140	29.32	0.8744	32.15	0.9130
BIN_4† [3]	32.67	0.9236	35.10	0.9417	32.51	0.9280	35.10	0.9468	32.59	0.9258	35.10	0.9443
ALANET (Ours)	**33.71**	**0.9429**	35.94	0.9496	**32.98**	**0.9362**	35.85	0.9513	**33.34**	**0.9355**	35.89	0.9504

[3] Wang Shen et al. "Blurry Video Frame Interpolation". 2020. arXiv: 2002.12259.

[5] Huaizu Jiang et al. "Super slomo: High quality estimation of multiple intermediate frames for video interpolation". *CVPR*. 2018.

[6] Wenbo Bao et al. "MEMC-Net: Motion estimation and motion compensation driven ···". *TPAMI* (2019).

[1] Wenbo Bao et al. "Depth-aware video frame interpolation". *CVPR*. 2019.

[2] Meiguang Jin et al. "Learning to extract a video sequence from a single motion-blurred image". *CVPR*. 2018.

Additional Results

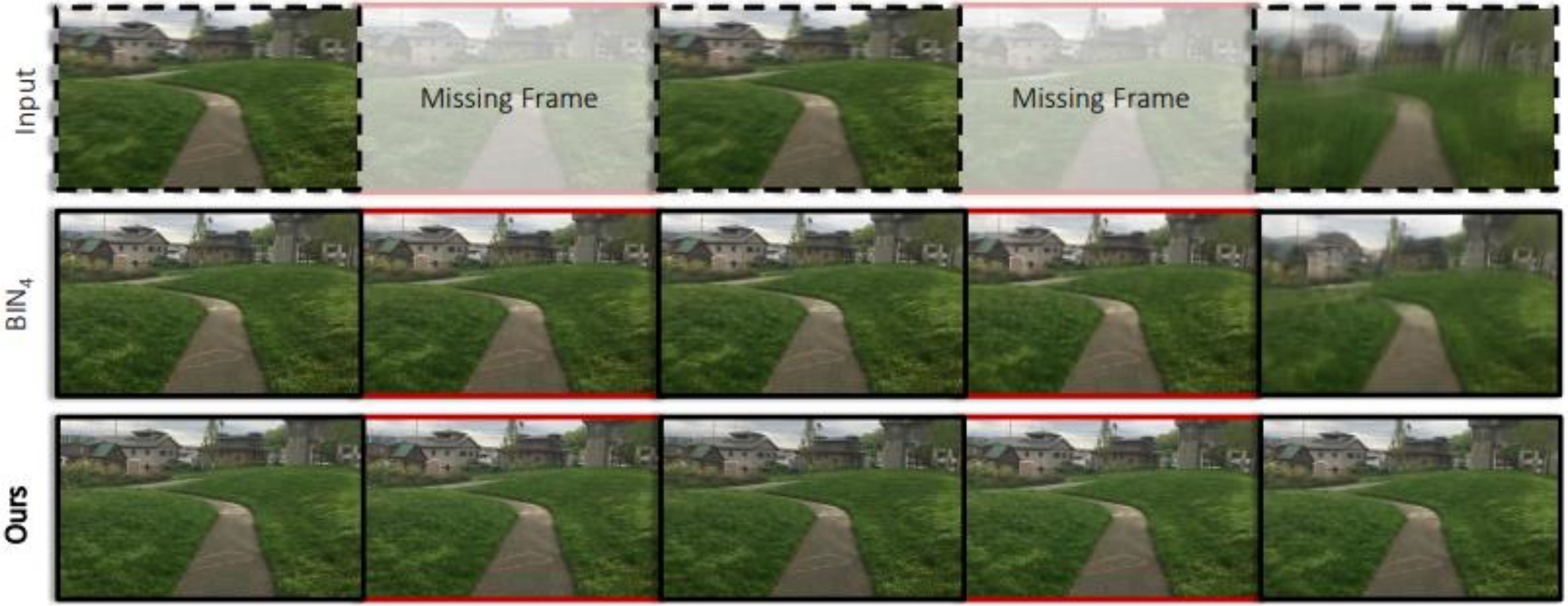
Qualitative Results

UCR Bourns College of Engineering
Video Computing Group

Qualitative Results

UCR Bourns College of Engineering
Video Computing Group

Qualitative Results



More Results

UCR Bourns College of Engineering
Video Computing Group

Conclusion

- **Neighbors can help!:** We introduce a novel framework ALANET, Adaptive Latent Attention Network, to jointly deblur and interpolate for high frame-rate sharp video generation.
- **One model does both:** This is the first work to generate high frame-rate sharp video from low frame-rate poor quality video by applying attention in the latent space.
- **Effective results:** Experiments demonstrate consistently effective results on two datasets, the benchmark Adobe240 and crawled YouTube240 in both deblurring and interpolation.