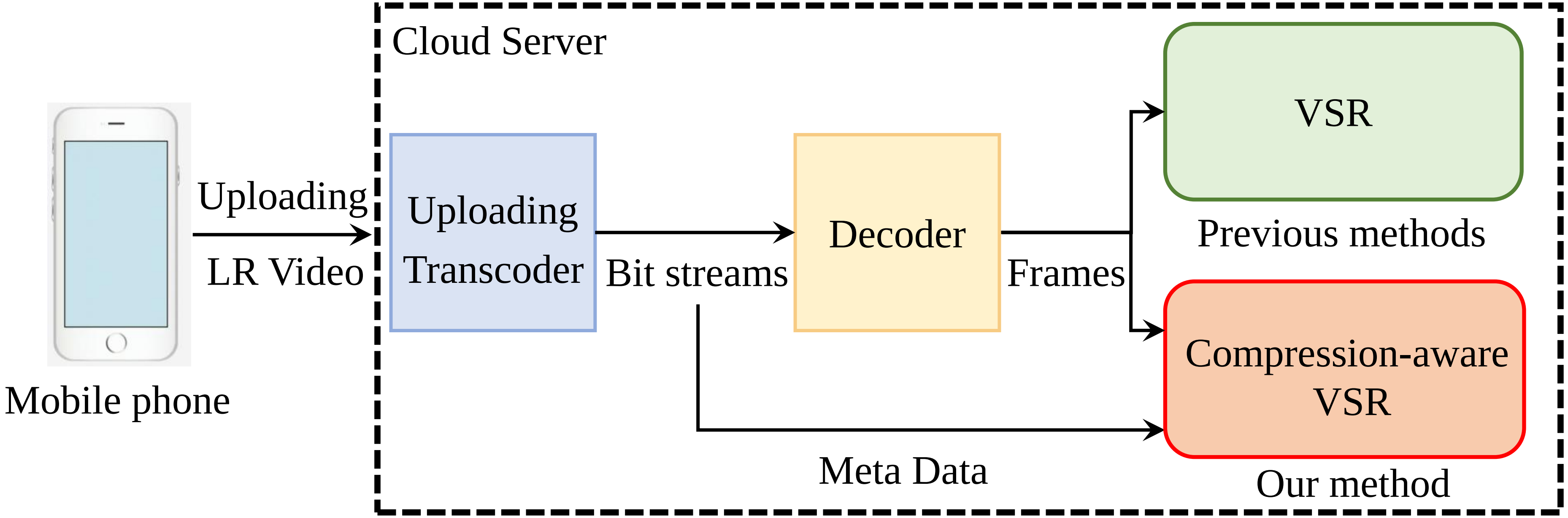
Compression-Aware Video Super-Resolution

Yingwei Wang[1]*, Takashi Isobe[2]*, Xu Jia[1], Xin Tao[3], Huchuan Lu[1,4], Yu-Wing Tai[5]

[1]Dalian University of Technology,
[2]Xiaohongshu Inc.,
[3]Kuaishou Technology,
[4]Peng Cheng Laboratory,
[5]The Hong Kong University of Science and Technology

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Compressed Video Super-Resolution

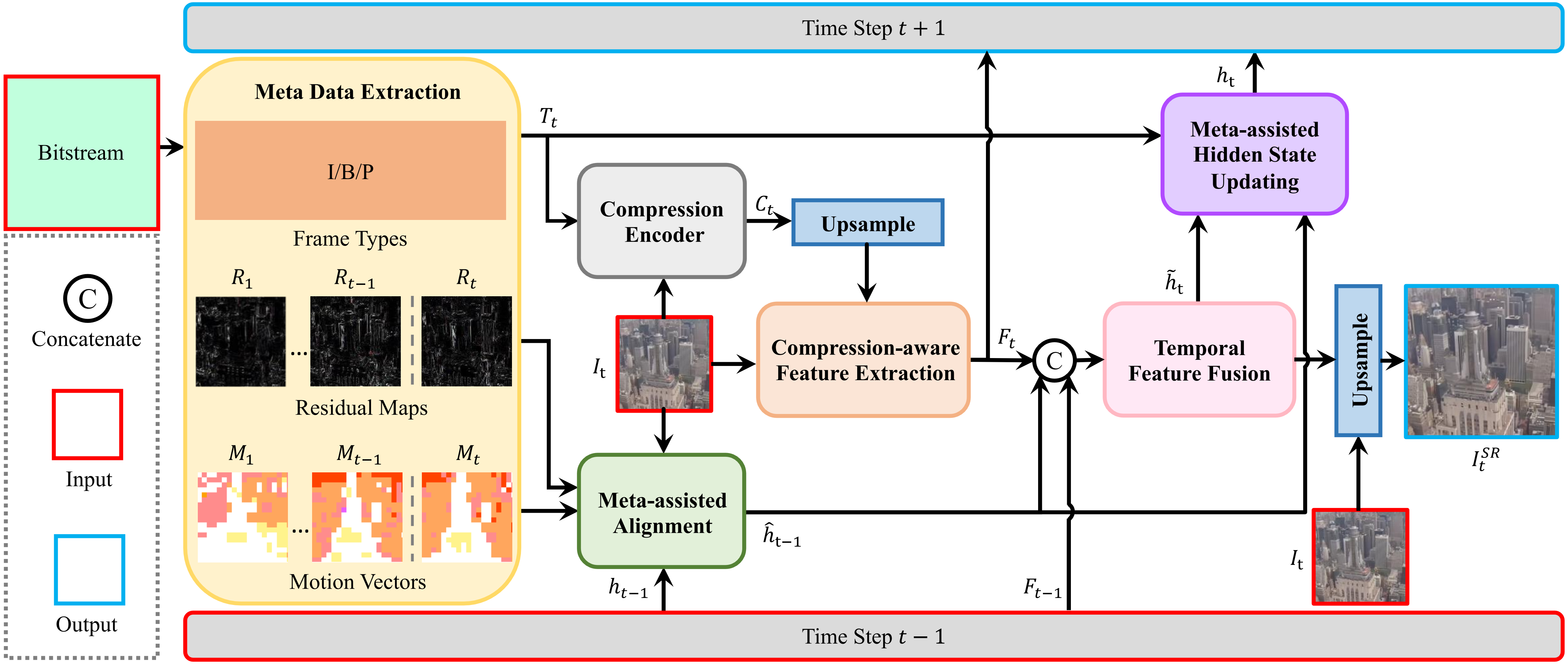


EDVR

RealBasic VSR

COMISR

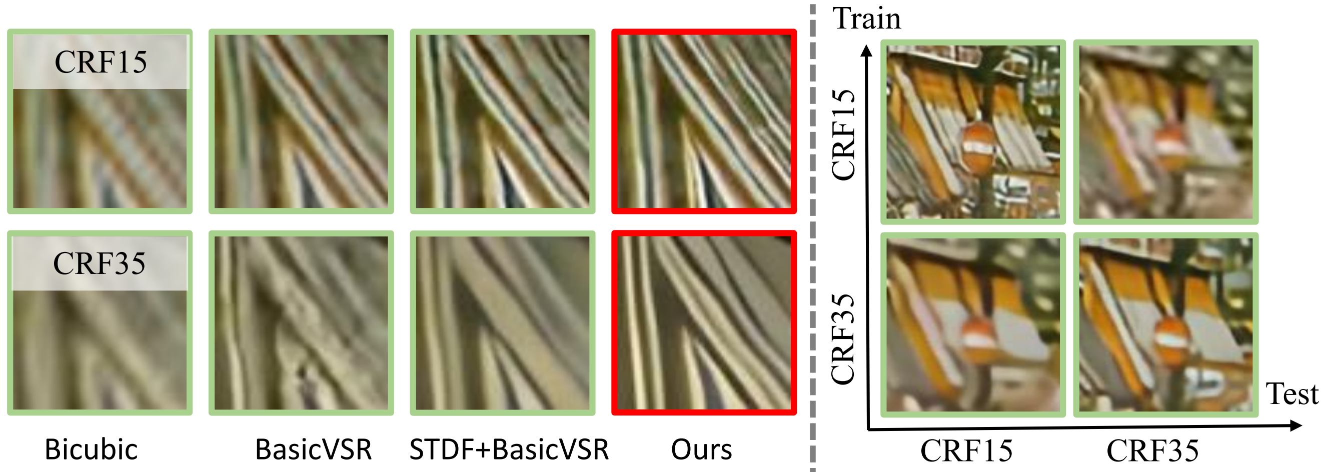
Network



Time Step $t+1$
Bitstream
Meta Data Extraction
I/B/P
Frame Types
R_1
R_{t-1}
R_t
Residual Maps
M_1
M_{t-1}
M_t
Motion Vectors
Concatenate
Input
Output
T_t
Compression Encoder
C_t
Upsample
I_t
Compression-aware Feature Extraction
F_t
Meta-assisted Alignment
$\hat{h}_{t-1}$
h_{t-1}
F_{t-1}
Temporal Feature Fusion
$\tilde{h}_t$
Meta-assisted Hidden State Updating
h_t
Upsample
I_t^{SR}
Time Step $t-1$

Motivation

- Problems:
  - ❑ unaware of compression level
    - ▪ artifacts
    - ▪ detail loss
  - ❑ disregard of meta data

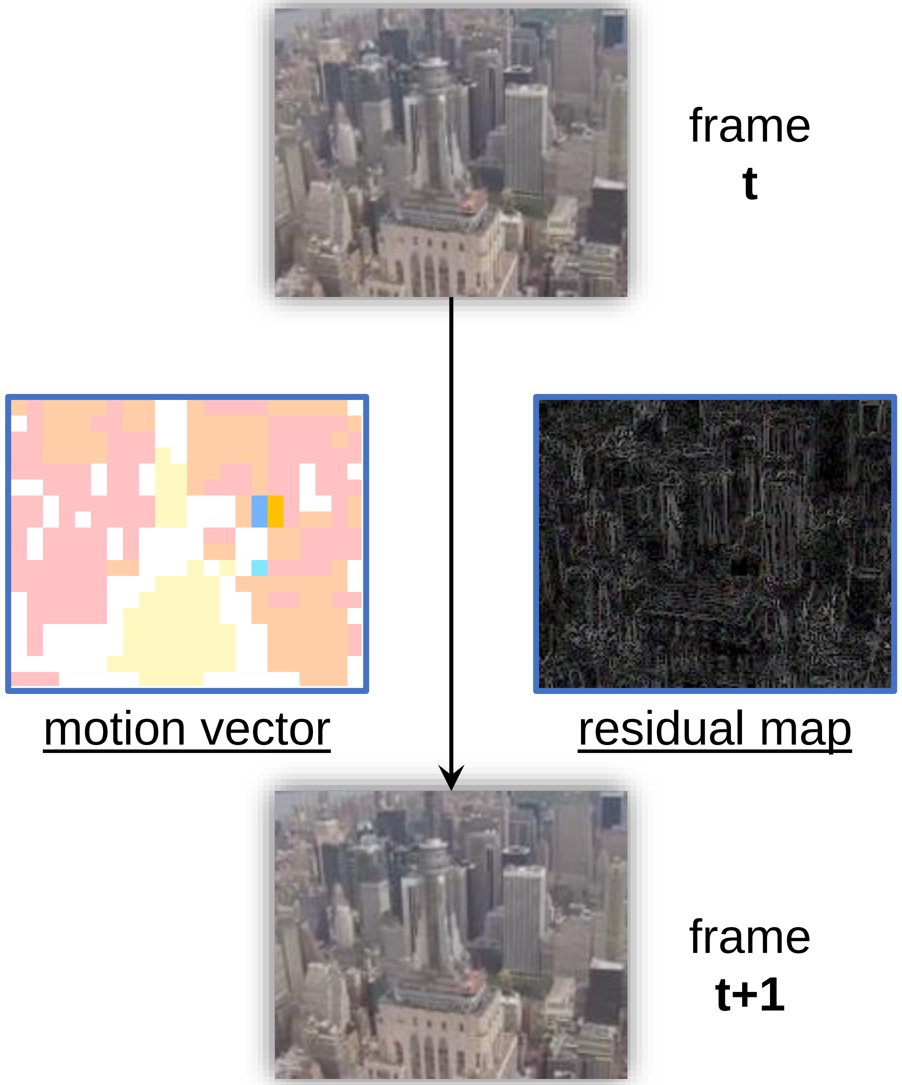


- Solution:
  - ❖ take advantage of meta data to facilitate the base VSR
  - ❖ be aware of compression with input videos and exert power adaptively.

Contribution

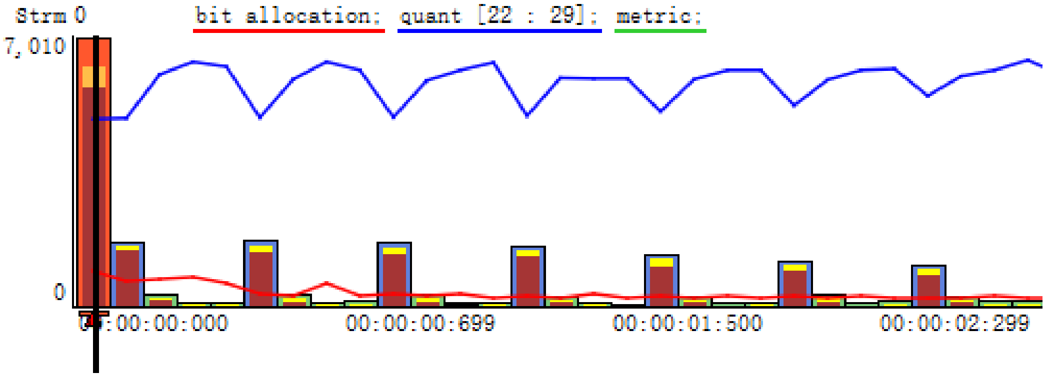
- A compression encoder to perceive compression levels of input frames.
- A compression-aware modulation module to encourage the base model to perform adaptively under various compression.
- Alignment and propagation process assisted by metadata.

Meta Data of Compressed Videos

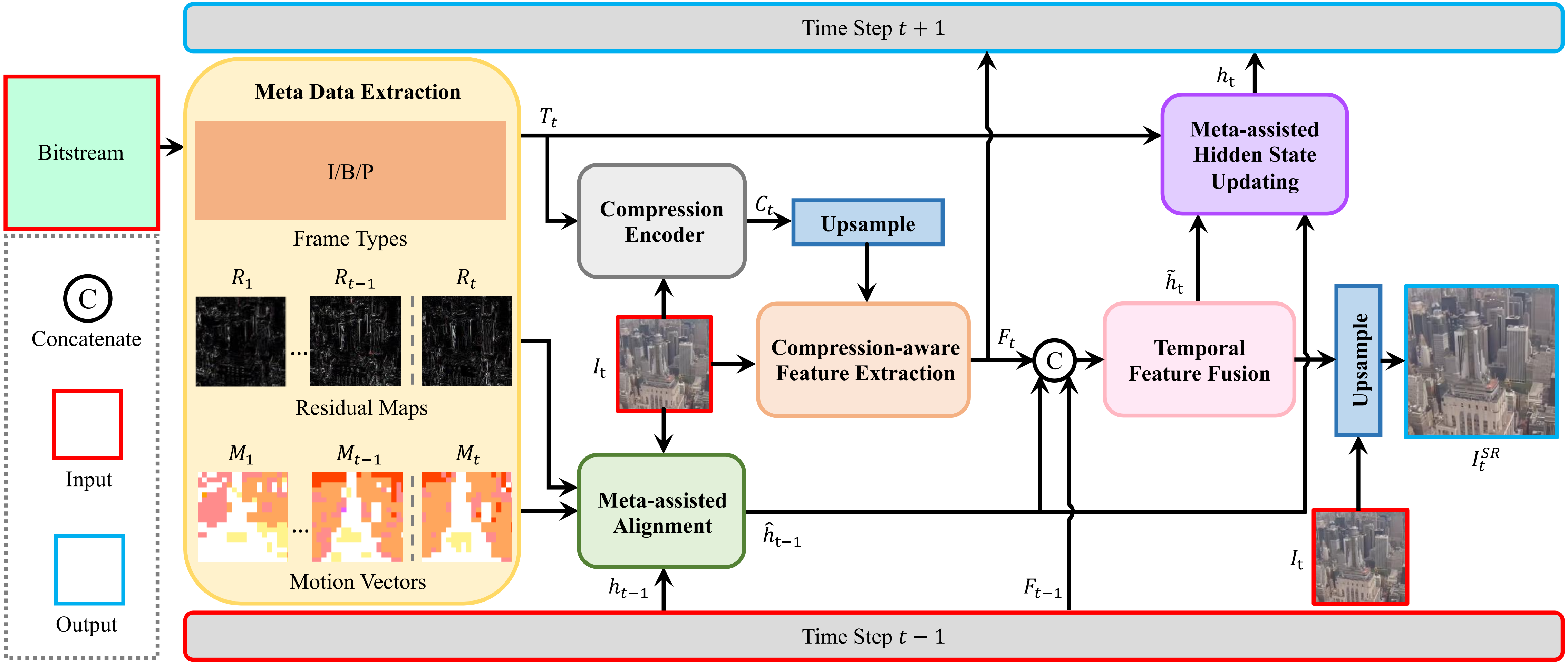


CRF (Constant Rate Factor)

Frame Type

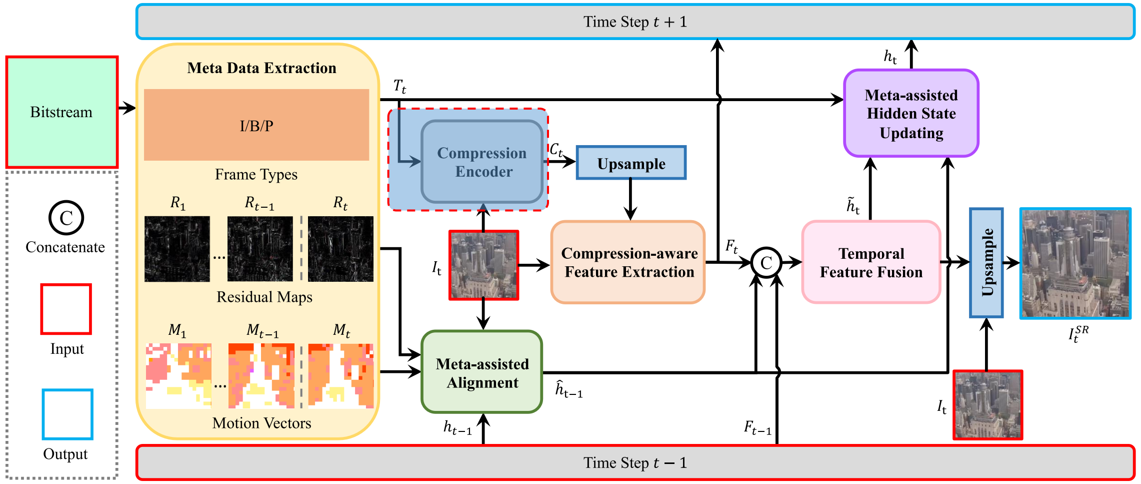


Approach

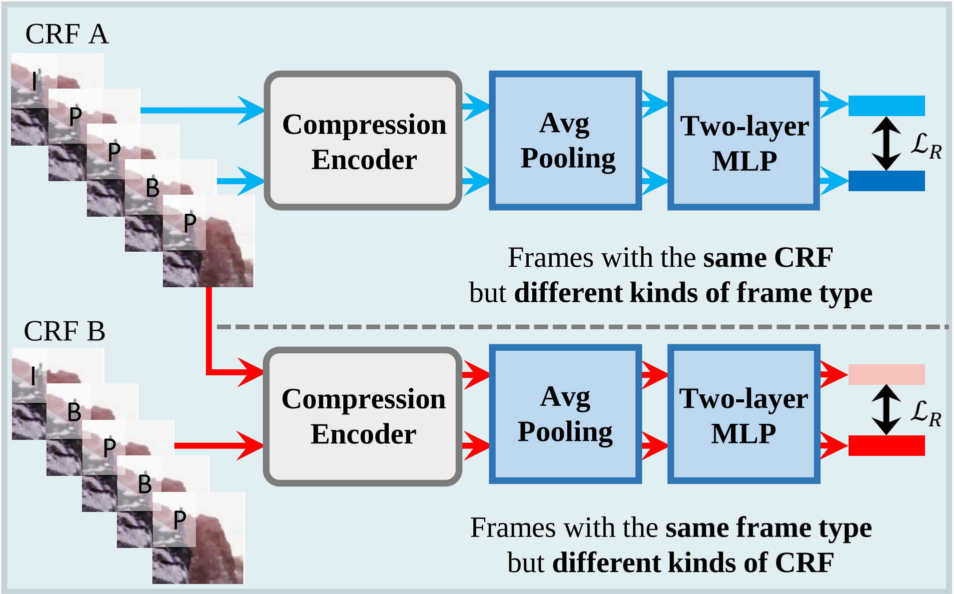


Overall

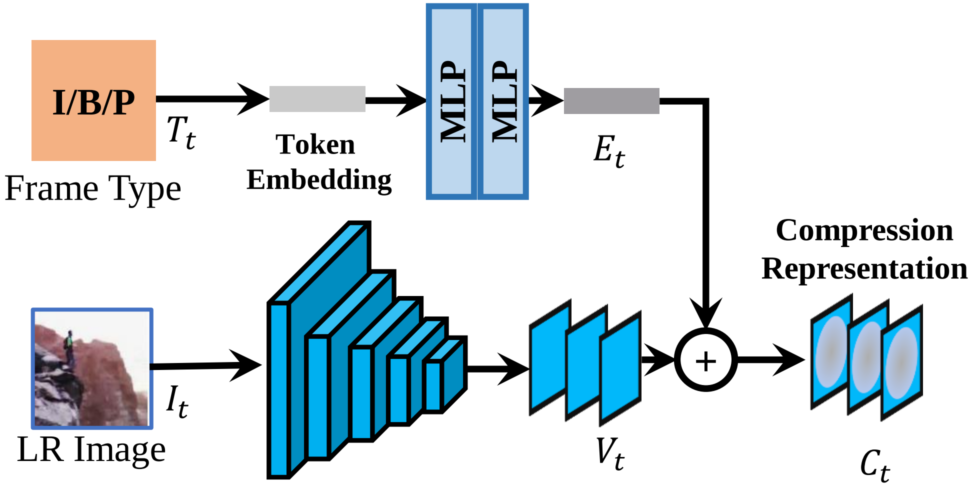
Approach



- Pretraining
  - ➢ Learning to **rank**



- Compression Encoder

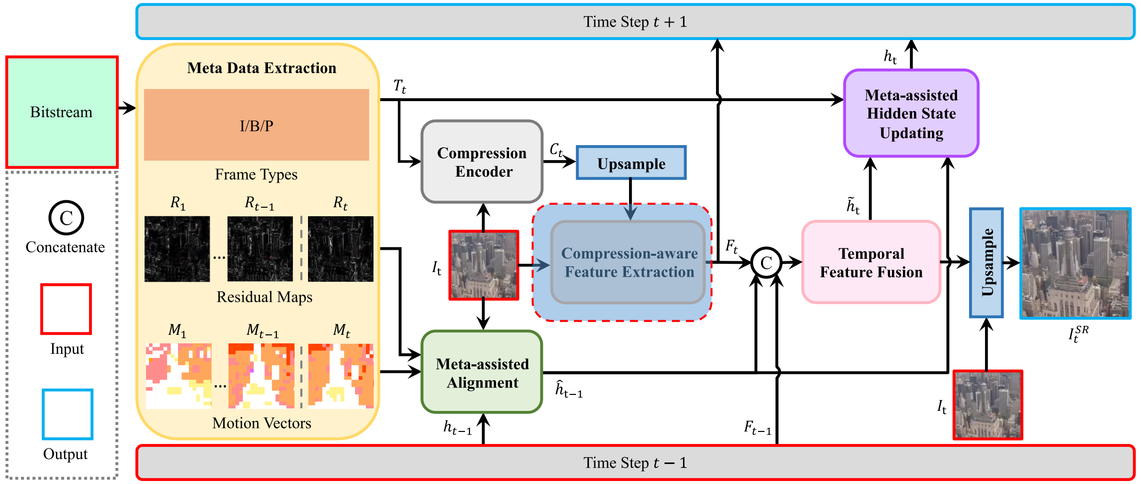


Ranking Loss Function:

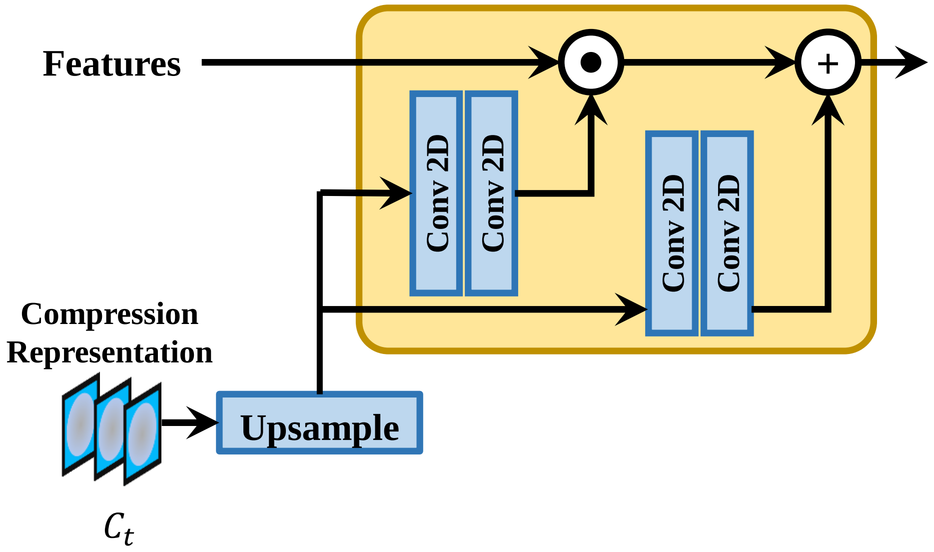
$$\mathcal{L}_R = max\left(0, (s_i - s_j) * \kappa + \xi\right)$$

$$where \begin{cases} \kappa = 1 & if\ \ Q_{f/c}(i) < Q_{f/c}(j) \\ \kappa = -1 & if\ \ Q_{f/c}(i) > Q_{f/c}(j) \end{cases}$$

Approach

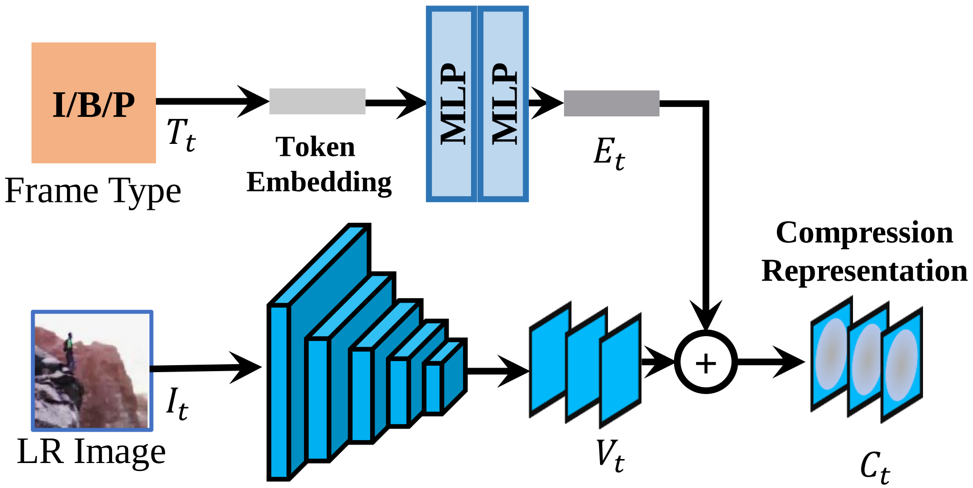


- Compression-Aware Modulation

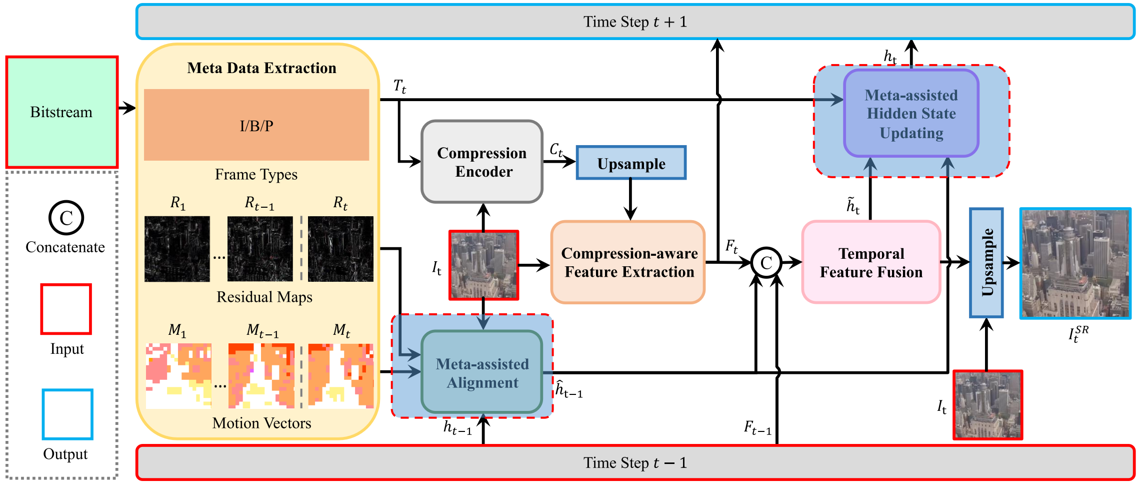


$$CAM(\mathbf{F}|\gamma_i, \beta_i) = \gamma_i \odot \mathbf{F} + \beta_i$$

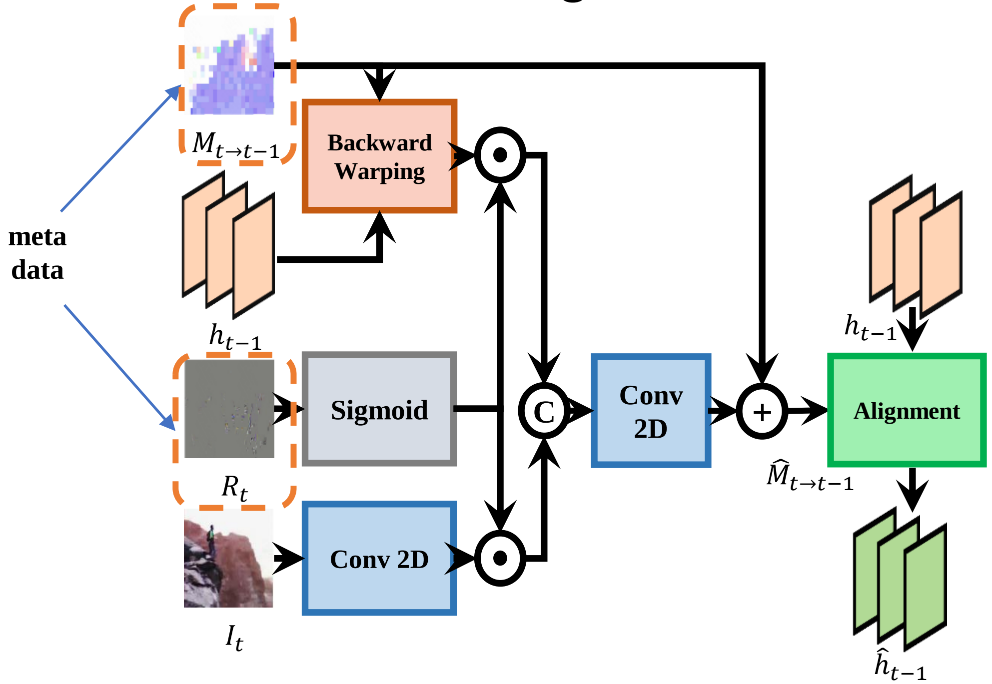
- Compression Encoder



Approach



- Meta-assisted alignment.



- Meta-assisted Propagation.

$$\begin{cases} h_t = \alpha * \tilde{h}_t + (1-\alpha) * \hat{h}_{t-1} & if\ T_t = B \\ h_t = \tilde{h}_t & otherwise, \end{cases}$$

where,

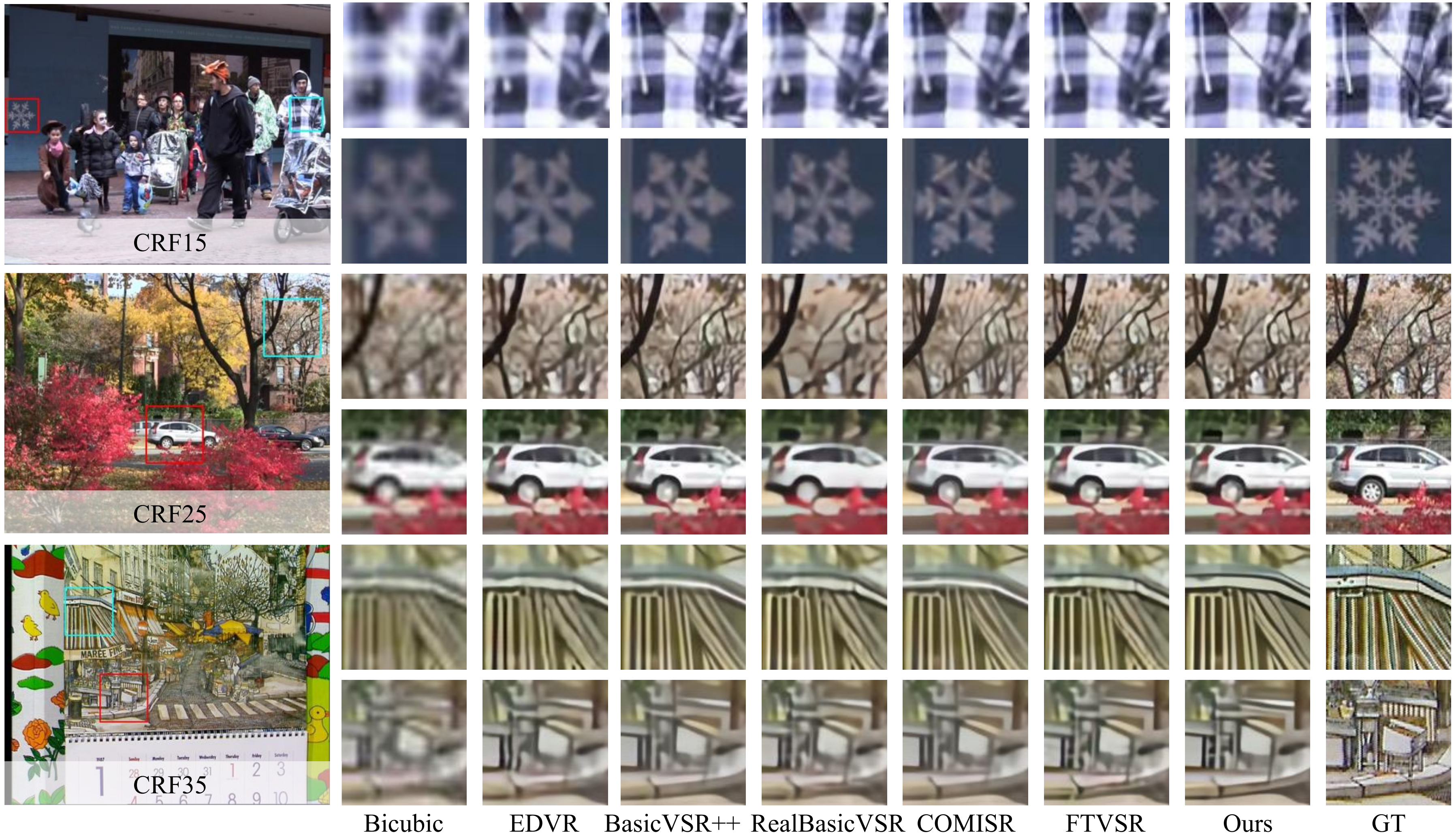
$\tilde{h}_t$ current hidden state,
$\hat{h}_{t-1}$ aligned previous hidden state
α momentum coefficient

Evaluation

	Params (M)	Runtime (ms)	Vid4 (Y)			REDS4 (RGB)		
			CRF 15	CRF 25	CRF 35	CRF 15	CRF 25	CRF 35
EDVR	20.6	378	26.53/0.794	24.76/0.694	22.39/0.544	29.31/0.836	26.27/0.742	23.78/0.625
RealBasic VSR	6.3	63	26.94/0.813	24.87/0.701	2239/0.531	29.76/0.849	26.49/0.746	23.63/0.626
COMISR	6.2	73	26.66/0.801	25.14/0.713	22.62/0.546	29.76/0.832	26.96/0.749	23.87/0.629
FTVSR	10.8	850	27.50/0.826	25.51/0.732	22.79/0.561	30.89/0.864	28.10/0.786	24.83/0.674
Ours	8.9	93	27.42/0.833	25.65/0.742	22.84/0.574	30.76/0.873	28.15/0.798	24.93/0.682

- Our method outperforms most of the previous VSR methods on the three compression levels both in PSNR and SSIM.
- Compressed to the latest FTVSR model, our method obtains comparable performance with lighter computation and GPU memory usage

Comparison to Existing Methods



- smoothing out noise,
- preserving details,
- maintaining temporal consistency

Ablation Studies

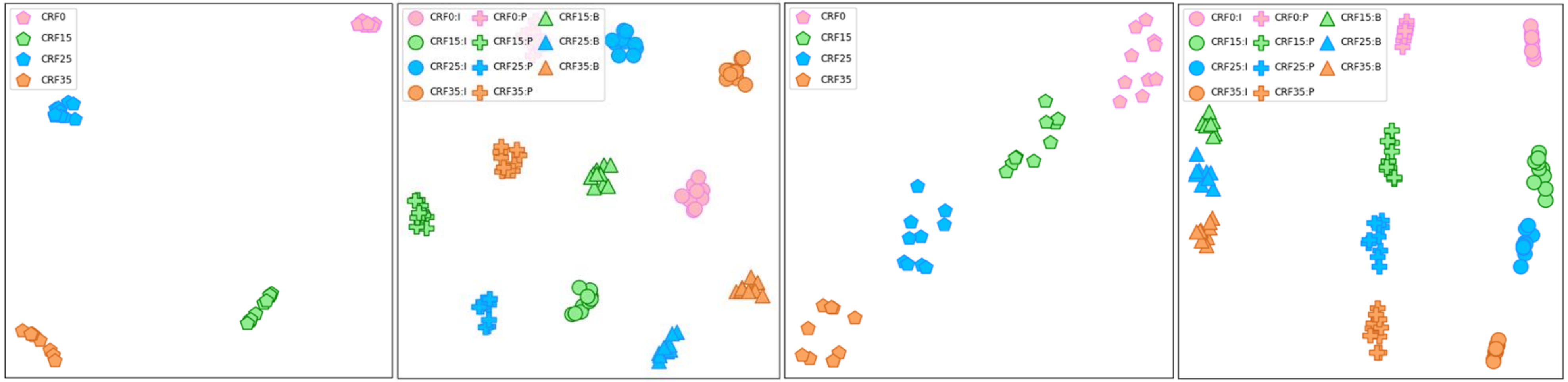
CAM	OA	MA	MH	CRF 15	CRF 25	CRF 35
	✓			26.76	24.54	22.06
✓	✓			27.25	25.41	22.74
✓		✓		27.40	25.60	22.80
✓		✓	✓	27.42	25.65	22.84

- With the compression-aware modulation (CAM)
  - Being awarded of compression level performance improved significantly
- Replacing optical flow alignment (OA) with meta-assisted alignment (MA)
  - More accurate motion estimation and improved temporal consistency
- With meta-assisted propagation (MP)
  - The propagation process is more stable, resulting in fewer artifacts

Performance of Compression Encoder

	loss		data		CRF 15	CRF 25	CRF 35
	CL	RL	CRF	IBP			
(a)	✓		✓		26.76	24.54	22.06
(b)	✓		✓	✓	27.25	25.41	22.74
(c)		✓	✓		27.40	25.60	22.80
(d)		✓	✓	✓	27.42	25.65	22.84

- Pretraining with ranking learning is more effective than contrastive learning and training
- Introduction of frame type information can improve the performance



(a) Contrastive-based Learning (CRF) (b) Contrastive-based Learning (CRF+IBP) (c) Ranking-based Learning (CRF) (d) Ranking-based Learning (CRF+IBP)

Conclusion

- A compression encoder and a compression-aware modulation
  - Perceiving compression level using rank-based pretrained encoder
  - Modulating feature extraction stage based on compression representation

- A meta-assisted alignment and propagation process
  - Leveraging the information from bitstream to enhance motion and temporal consistency modeling

- A meta-assisted propagation strategy
  - The propagation process is more stable, resulting in fewer artifacts
  - Reducing the computational cost and parameters of the optical flow network