Learned Video Compression with Feature-level Residuals

Runsen Feng   Yaojun Wu   Zongyu Guo
Zhizheng Zhang   Zhibo Chen

University of Science and Technology of China

Motivation

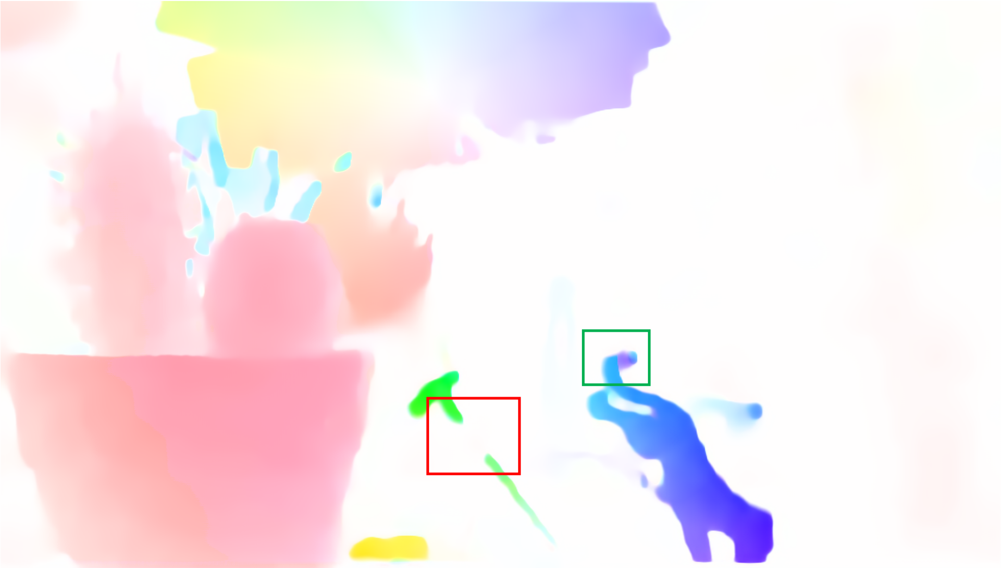
Prediction errors of optical flow

⬇

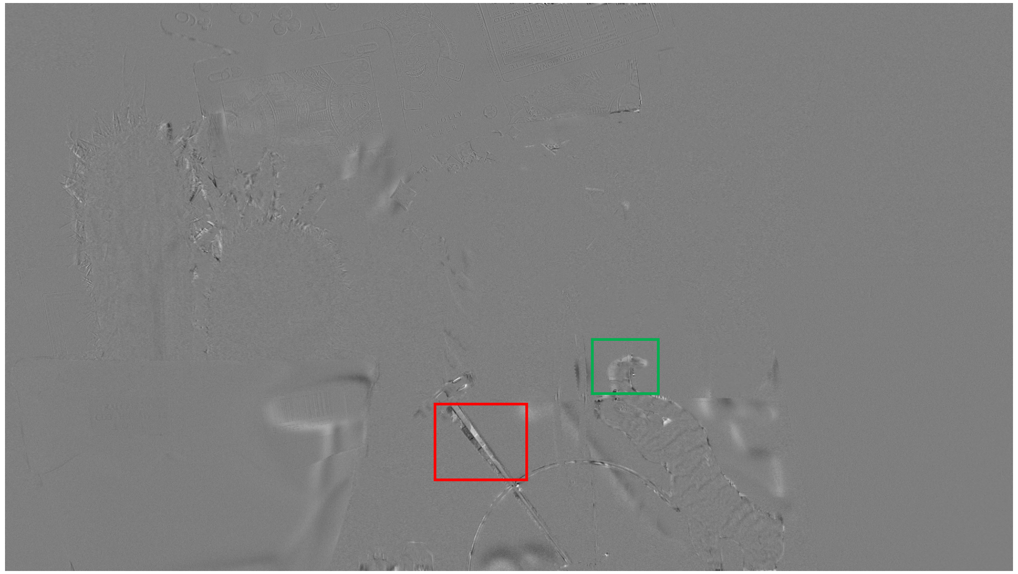
Large pixel space residual

Compensation Frame

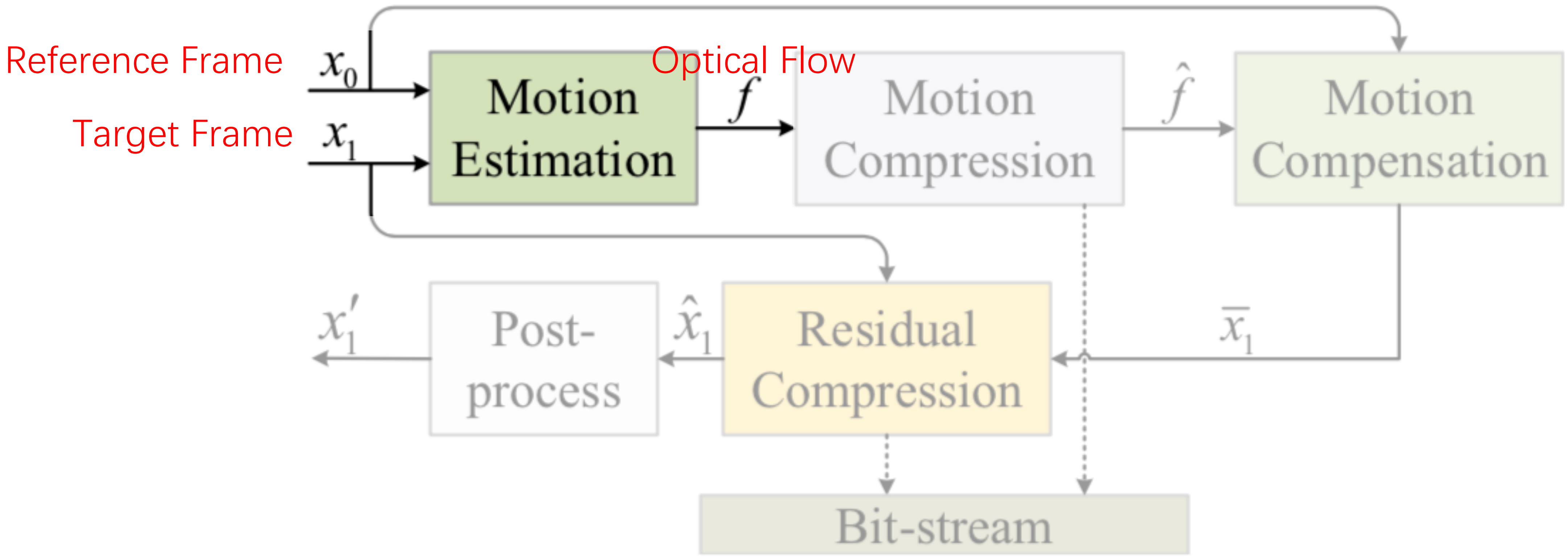
Decoded Optical Flow



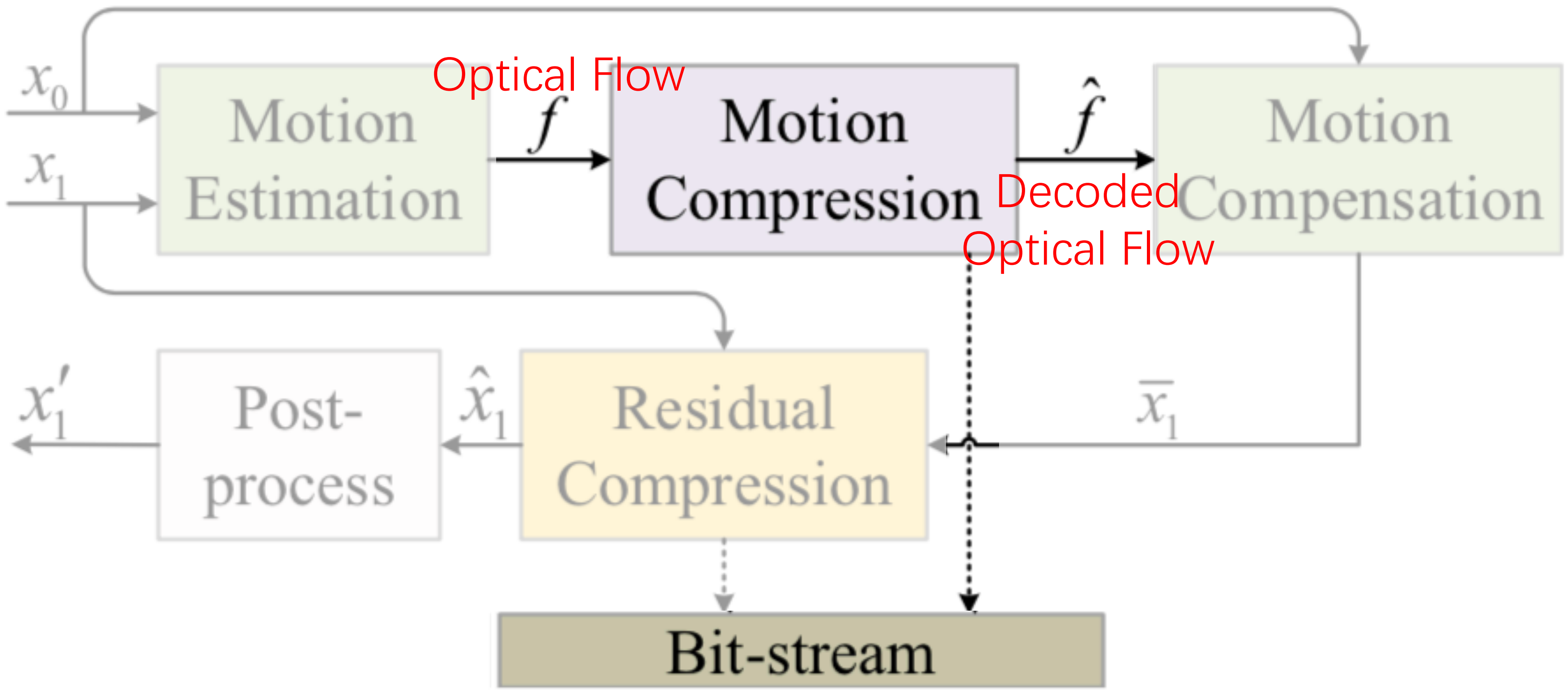
Pixel Space Residual



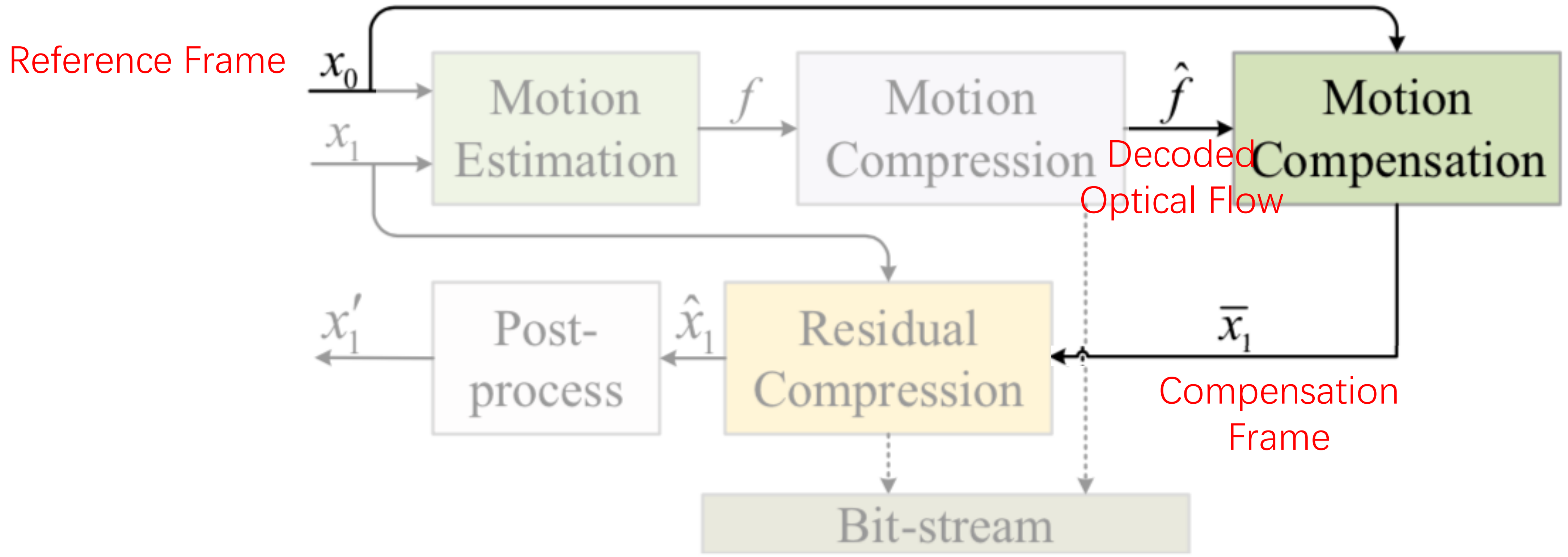
The overall coding pipeline of our P-frame compression framework.



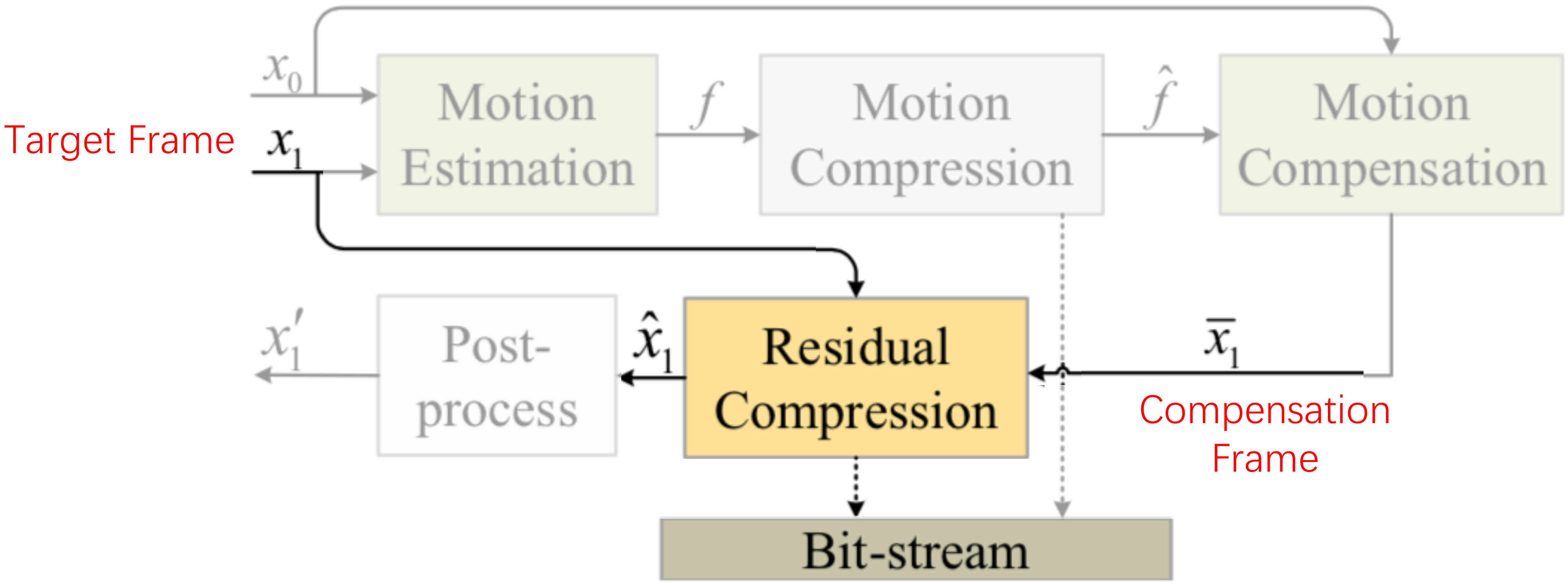
The overall coding pipeline of our P-frame compression framework.



The overall coding pipeline of our P-frame compression framework.

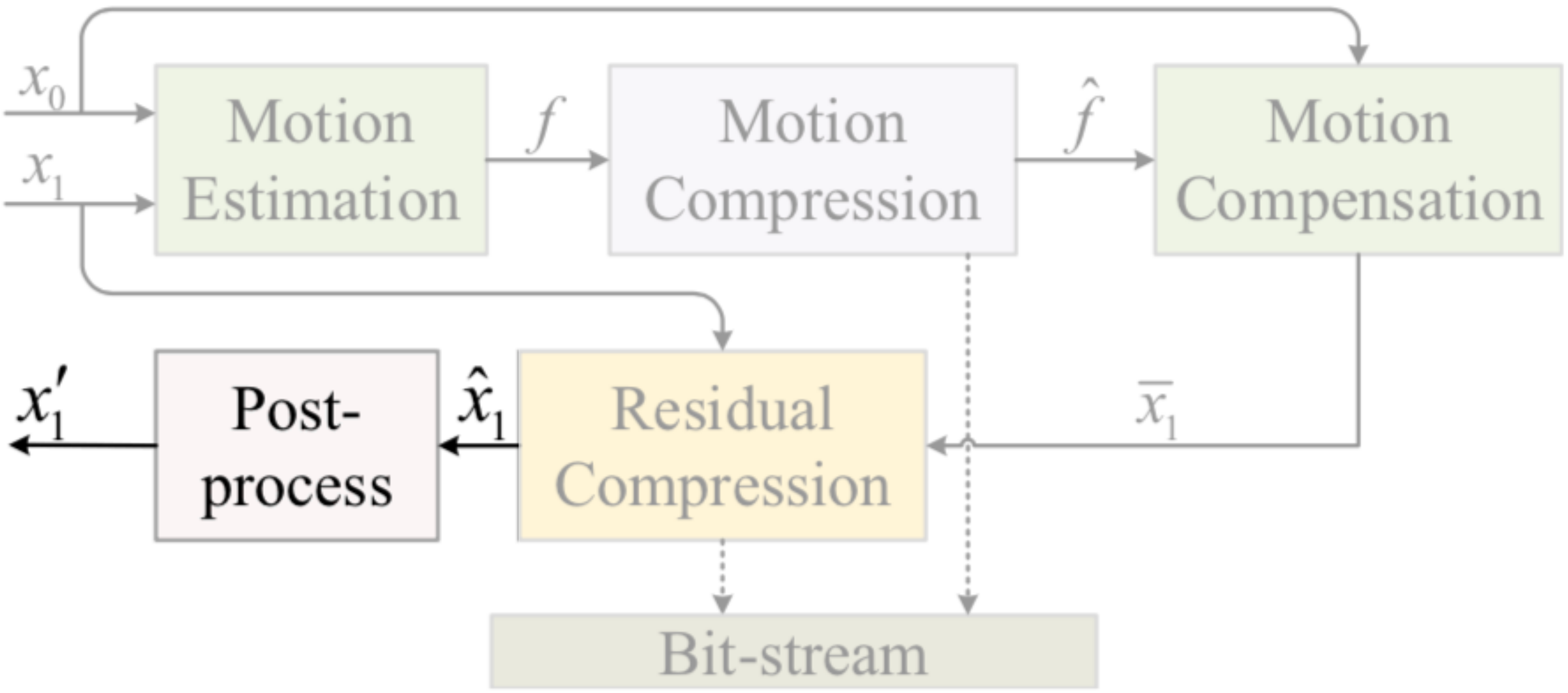


The overall coding pipeline of our P-frame compression framework.

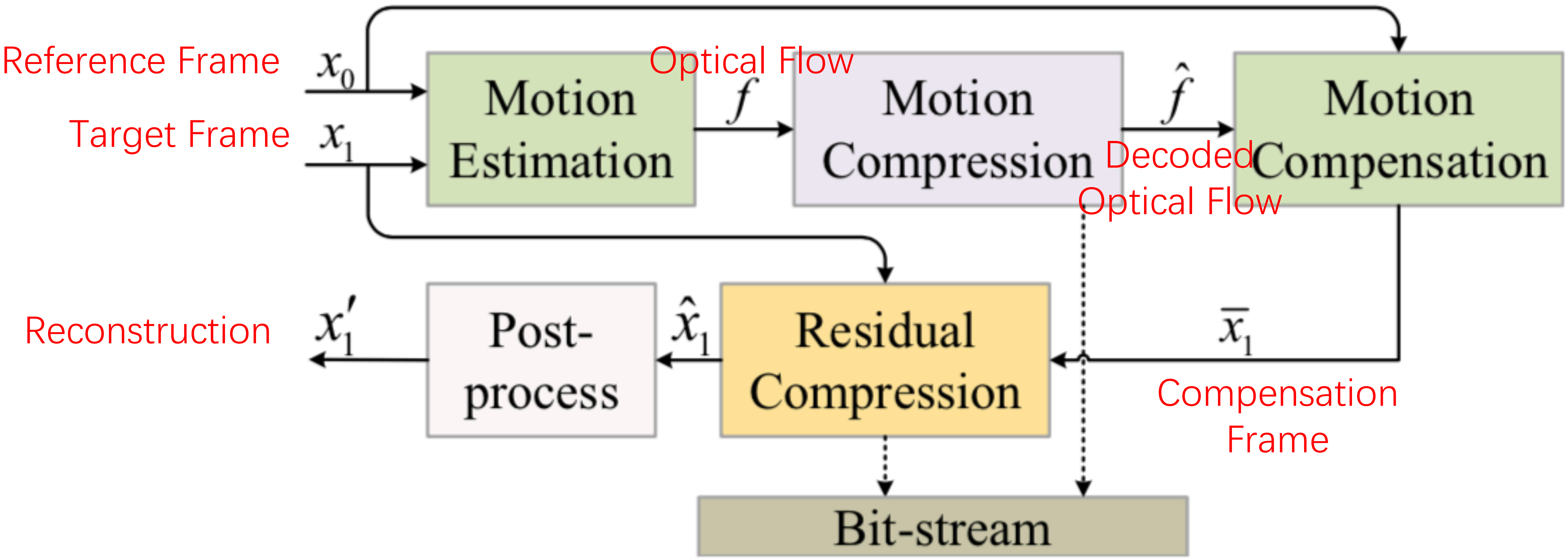


The overall coding pipeline of our P-frame compression framework.

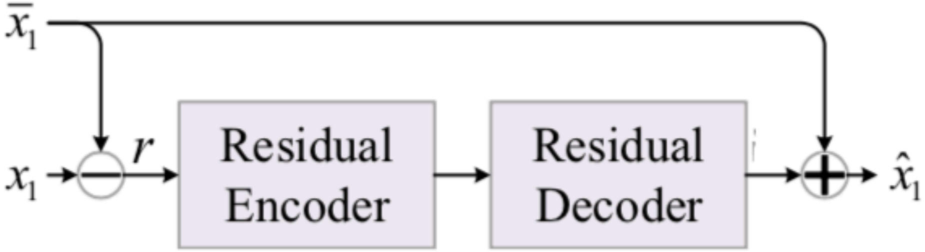
Reconstruction



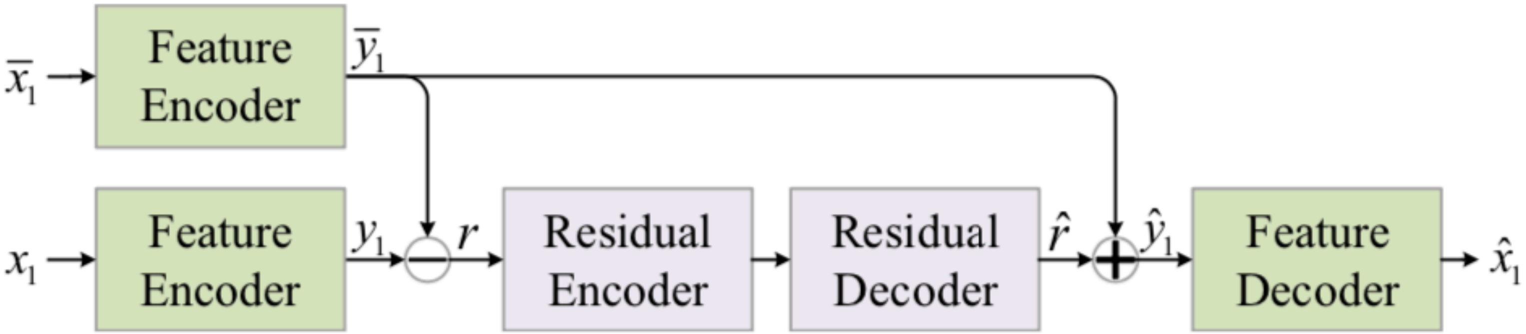
The overall coding pipeline of our P-frame compression framework.



Computing residual in feature space.

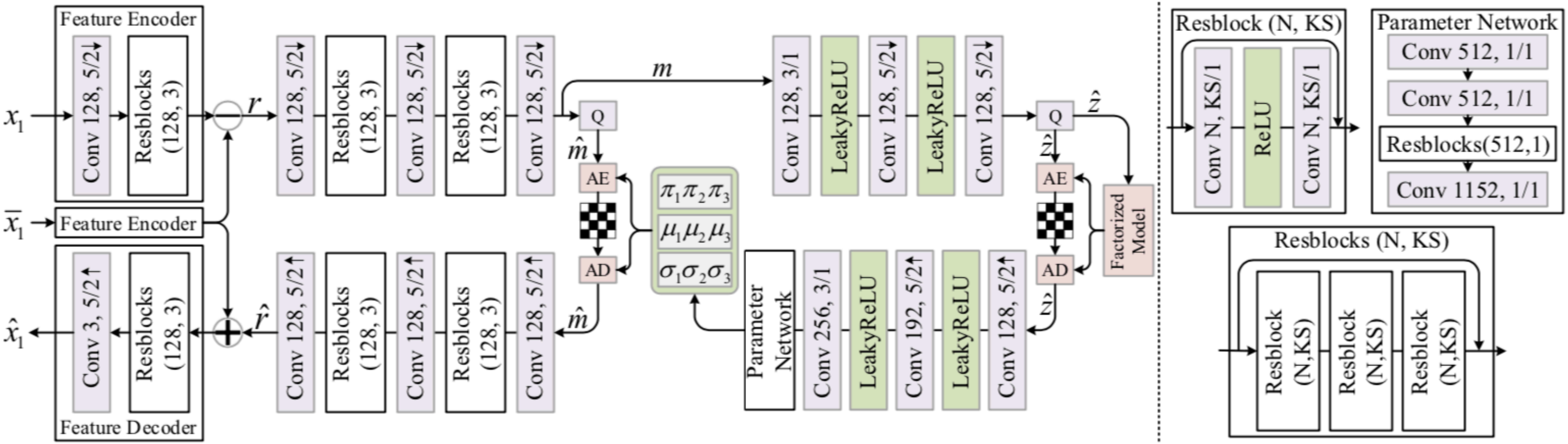


(a) Residual compression on pixel level.



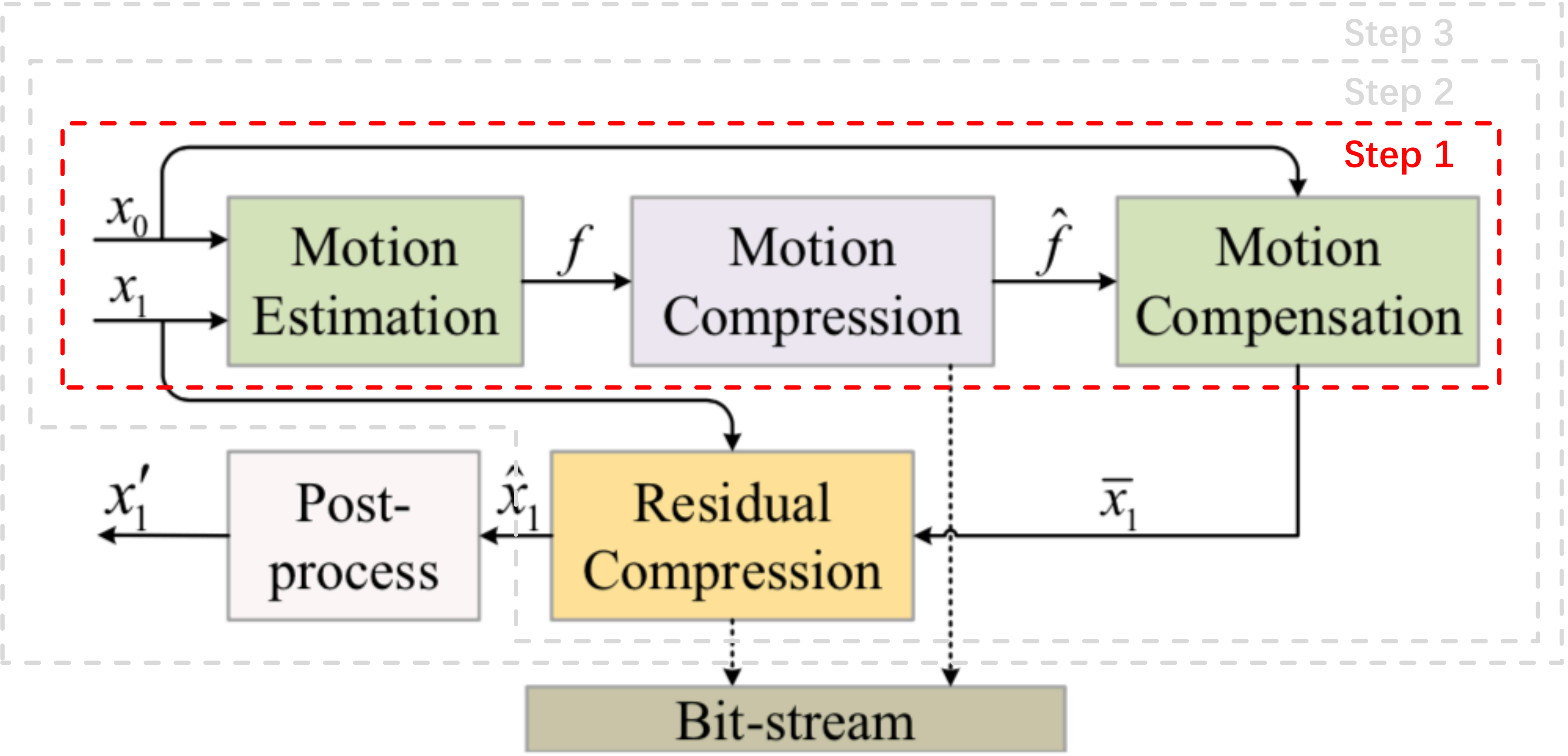
(b) Residual compression on feature level.

Detailed structure of our feature residual compression method.

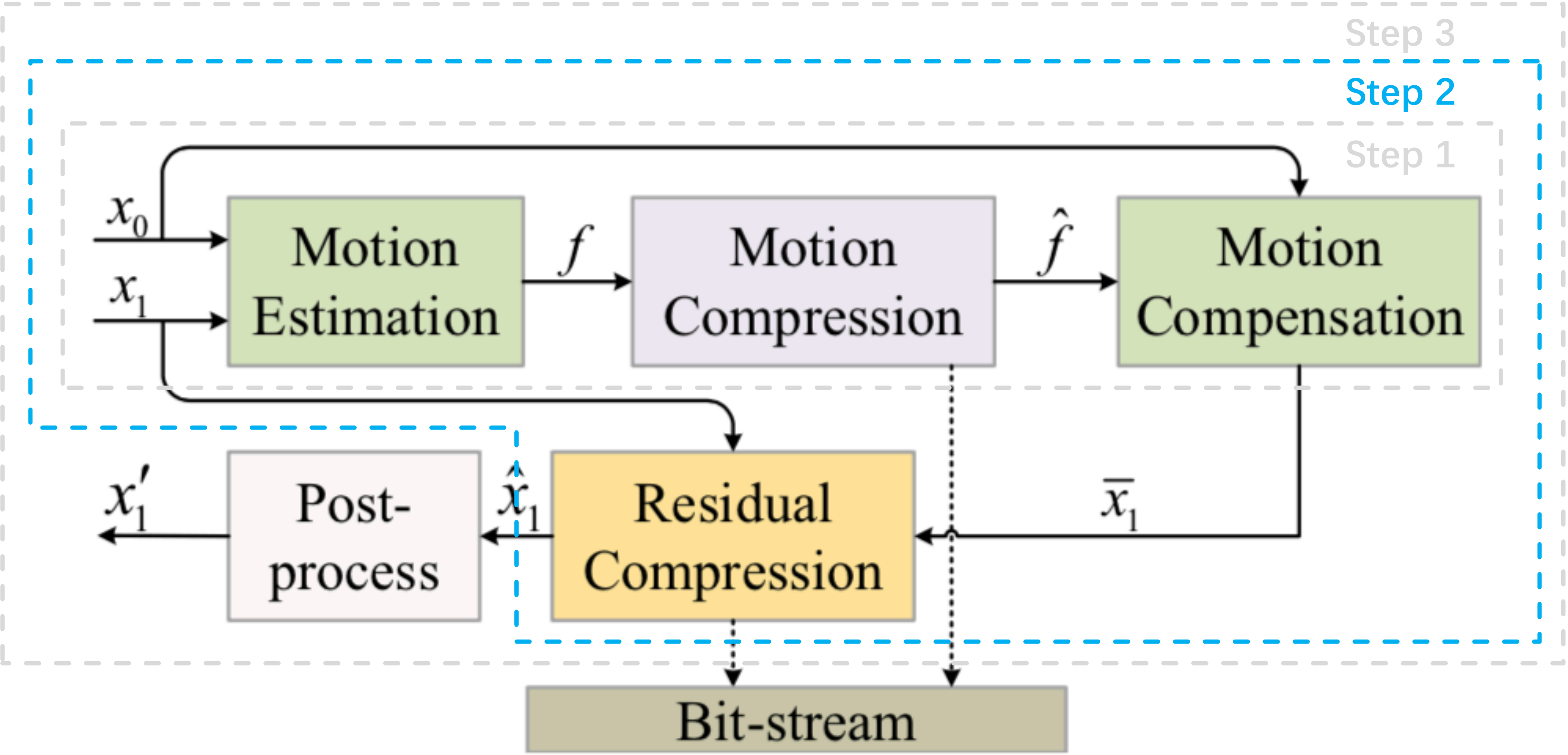


$$p_{\hat{\boldsymbol{m}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{m}} \mid \hat{\boldsymbol{z}}) = \prod_i \left(\sum_{k=1}^{K} \pi_{i,k}\mathcal{N}(\mu_{i,k}, \sigma_{i,k}^2) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2})\right)(\hat{m}_i)$$

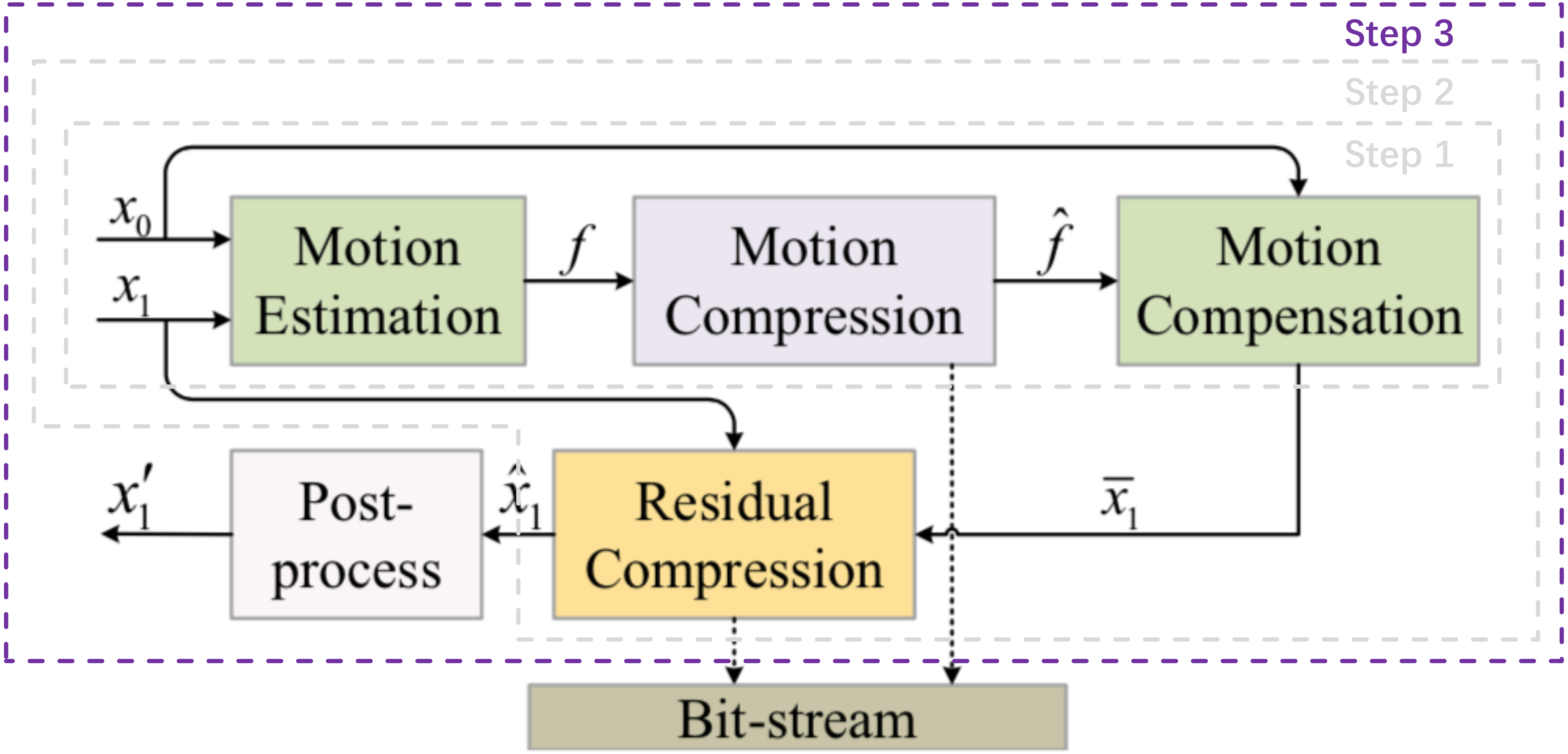
Step-by-step training strategy.



Step-by-step training strategy.



Step-by-step training strategy.



The overall rate-distortion (R-D) loss:

$$\mathcal{L} = R_f + R_r + \lambda d(\boldsymbol{x}_1, \boldsymbol{x}_1')$$

Evaluation results

Table 1: Evaluation results on P-frame validation dataset.

Model #	Pixel-level residuals	Feature-level residuals	GRDN	Ensemble	Data size	Model size	MS-SSIM
# 1	✓				38205911	79114205	0.996302
# 2	✓		✓		37788576	120847836	0.996619
# 3		✓			38133059	86399558	0.996645
# 4		✓	✓		37716003	128105152	0.996700
# 5	✓	✓	✓	# 1, # 4	37960950	103610323	0.996792
# 6	✓	✓	✓	# 2, # 4	37735411	126164272	0.996866

* The unit of data/model size is Byte. For P-frame challenge in CLIC, we limit the total size to 3,900,000,000 bytes, which is calculated as follows: model size + 100 $\times$ data size.

Evaluation results

Table 1: Evaluation results on P-frame validation dataset.

Model #	Pixel-level residuals	Feature-level residuals	GRDN	Ensemble	Data size	Model size	MS-SSIM
# 1	✓				38205911	79114205	0.996302
# 2	✓		✓		37788576	120847836	0.996619
# 3		✓			38133059	86399558	0.996645
# 4		✓	✓		37716003	128105152	0.996700
# 5	✓	✓	✓	# 1, # 4	37960950	103610323	0.996792
# 6	✓	✓	✓	# 2, # 4	37735411	126164272	0.996866

* The unit of data/model size is Byte. For P-frame challenge in CLIC, we limit the total size to 3,900,000,000 bytes, which is calculated as follows: model size + 100 $\times$ data size.

Evaluation results

Table 1: Evaluation results on P-frame validation dataset.

Model #	Pixel-level residuals	Feature-level residuals	GRDN	Ensemble	Data size	Model size	MS-SSIM
# 1	✓				38205911	79114205	0.996302
# 2	✓		✓		37788576	120847836	0.996619
# 3		✓			38133059	86399558	0.996645
# 4		✓	✓		37716003	128105152	0.996700
# 5	✓	✓	✓	# 1, # 4	37960950	103610323	0.996792
# 6	✓	✓	✓	# 2, # 4	37735411	126164272	0.996866

* The unit of data/model size is Byte. For P-frame challenge in CLIC, we limit the total size to 3,900,000,000 bytes, which is calculated as follows: model size + 100 $\times$ data size.

Evaluation results

Table 1: Evaluation results on P-frame validation dataset.

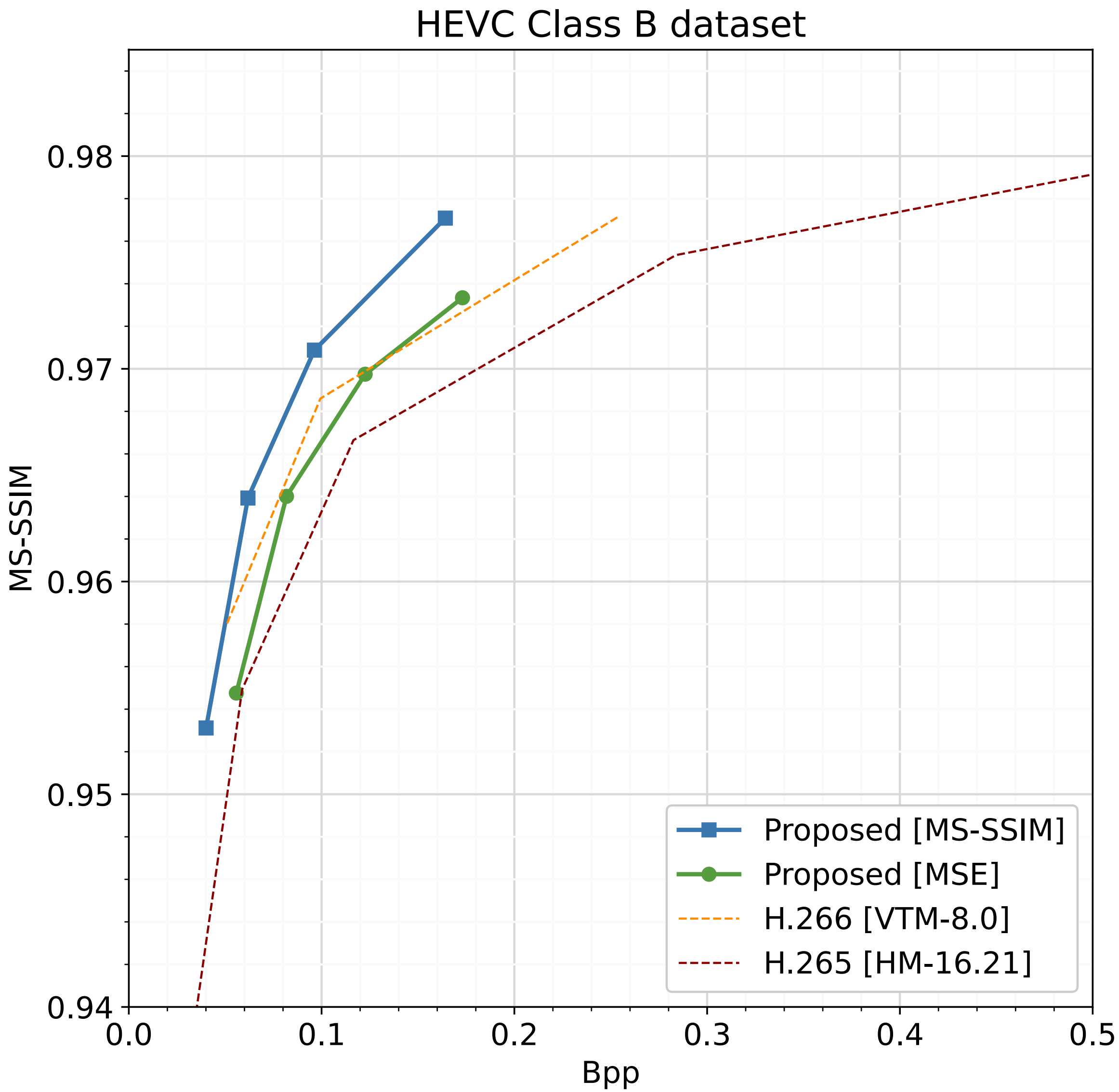
Model #	Pixel-level residuals	Feature-level residuals	GRDN	Ensemble	Data size	Model size	MS-SSIM
# 1	✓				38205911	79114205	0.996302
# 2	✓		✓		37788576	120847836	0.996619
# 3		✓			38133059	86399558	0.996645
# 4		✓	✓		37716003	128105152	0.996700
# 5	✓	✓	✓	# 1, # 4	37960950	103610323	0.996792
# 6	✓	✓	✓	# 2, # 4	37735411	126164272	0.996866

* The unit of data/model size is Byte. For P-frame challenge in CLIC, we limit the total size to 3,900,000,000 bytes, which is calculated as follows: model size + 100 × data size.

Team name: IMCL_MSSSIM

Workshop poster ID: 5

Our recent work of learned video compression for low-latency scenarios.



Thank you

E-mail: chenzhibo@ustc.edu.cn

IMCL-Intelligent Media Computing Lab
University of Science and Technology of China

This work was supported in part by NSFC under Grant U1908209, 61632001 and the National Key Research and Development Program of China 2018AAA0101400.