Large-capacity and Flexible Video Steganography via Invertible Neural Network

Chong Mou[1], Youmin Xu[1], Jiechong Song[1], Chen Zhao[2], Bernard Ghanem[2], Jian Zhang[1]*

[1] Peking University Shenzhen Graduate School, [2] King Abdullah University of Science and Technology (KAUST)

Reporter: Chong Mou

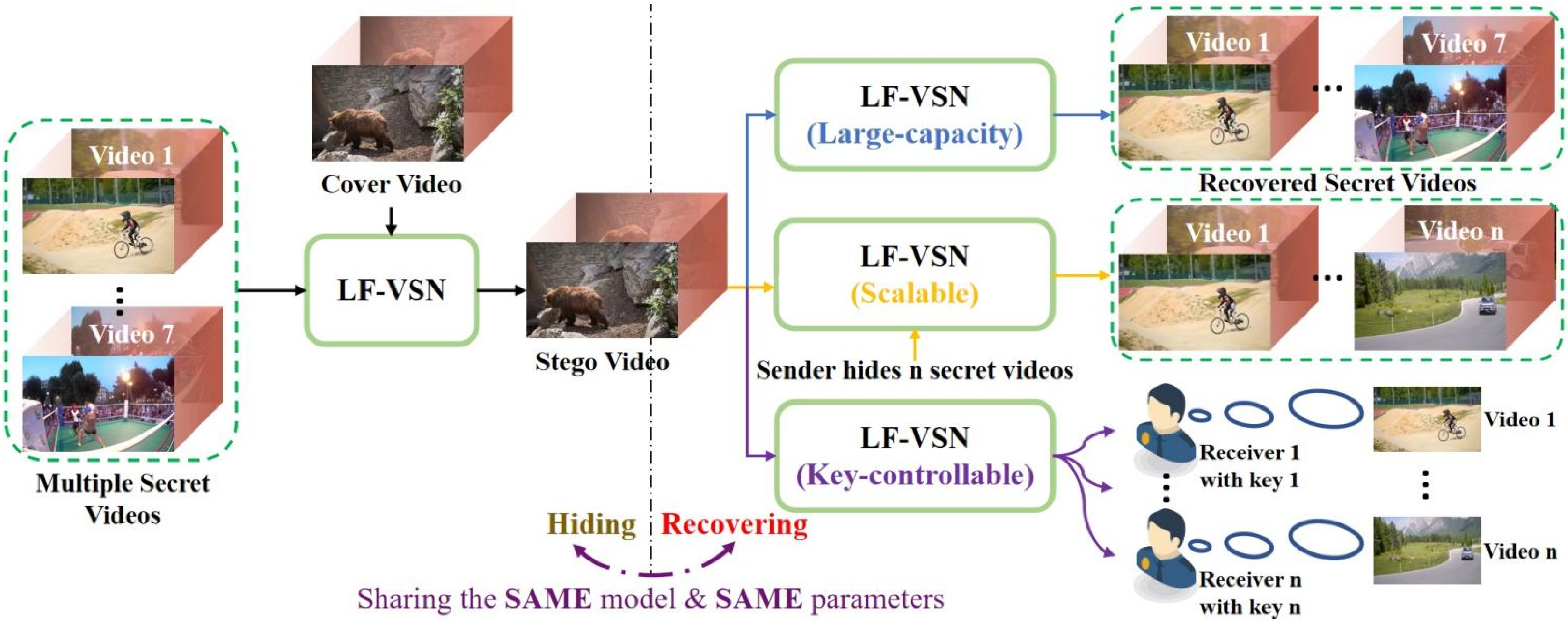
PEKING UNIVERSITY 北京大學 1898

VILLA Visual-Information Intelligent Learning LAB

King Abdullah University of Science and Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

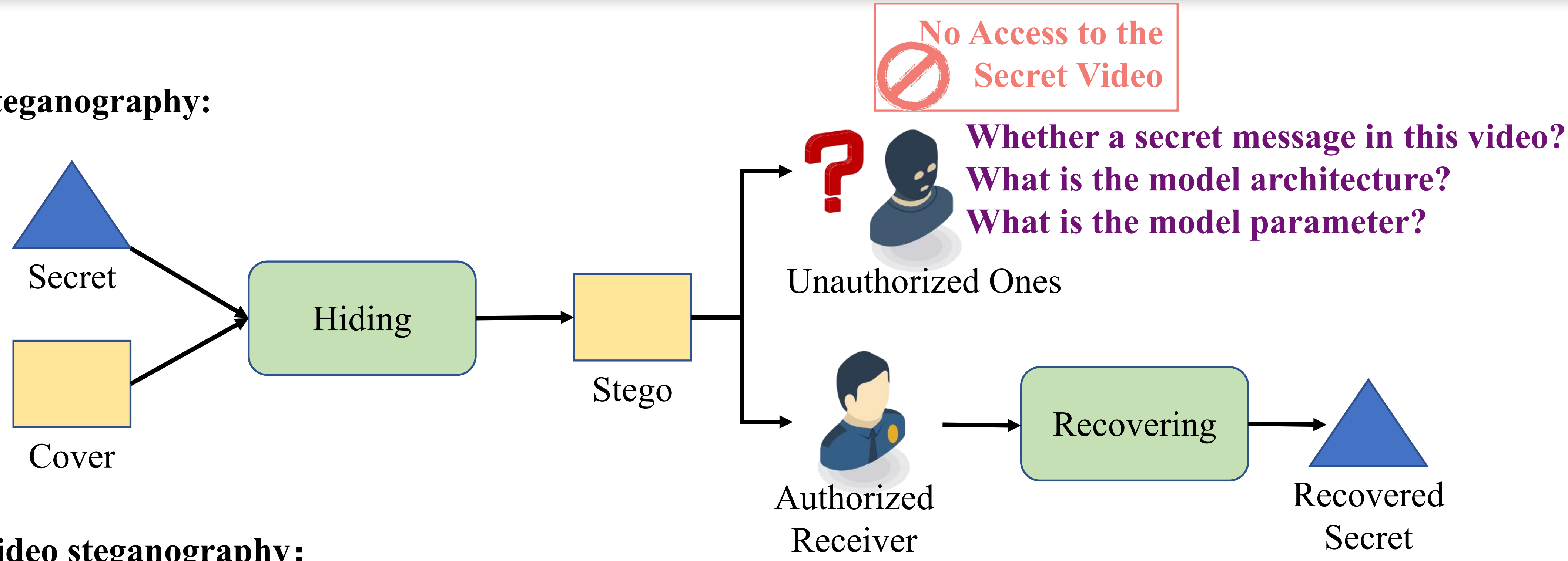
00 Abstract



Video 1
Video 7
Multiple Secret Videos
Cover Video
LF-VSN
Stego Video
Hiding
Recovering
Sharing the SAME model & SAME parameters
LF-VSN
(Large-capacity)
LF-VSN
(Scalable)
Sender hides n secret videos
LF-VSN
(Key-controllable)
Video 1
Video 7
Recovered Secret Videos
Video 1
Video n
Receiver 1
with key 1
Receiver n
with key n
Video 1
Video n

Background-The Task of Steganography

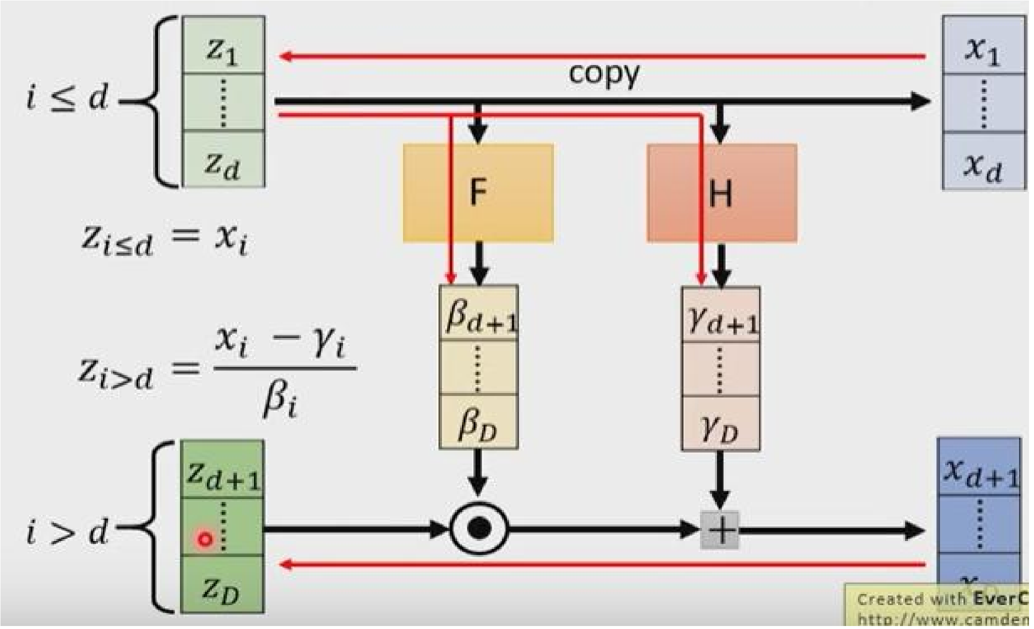
- **Steganography:**



- **Video steganography:**

Background-Invertible Neural Network

Invertible Block



$$Z \xrightarrow{\text{split}} Z_1,\ Z_2$$

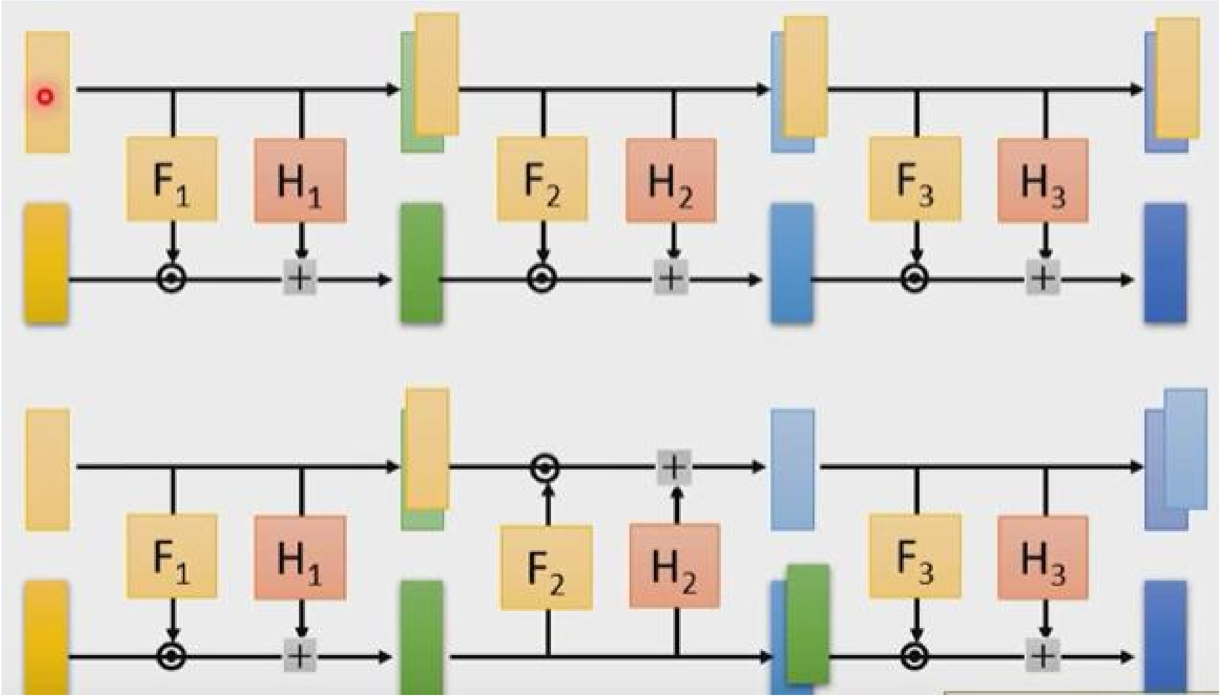
Forward:

$$X_1 = Z_1$$
$$X_2 = Z_2 \cdot F(Z_1) + H(Z_1)$$

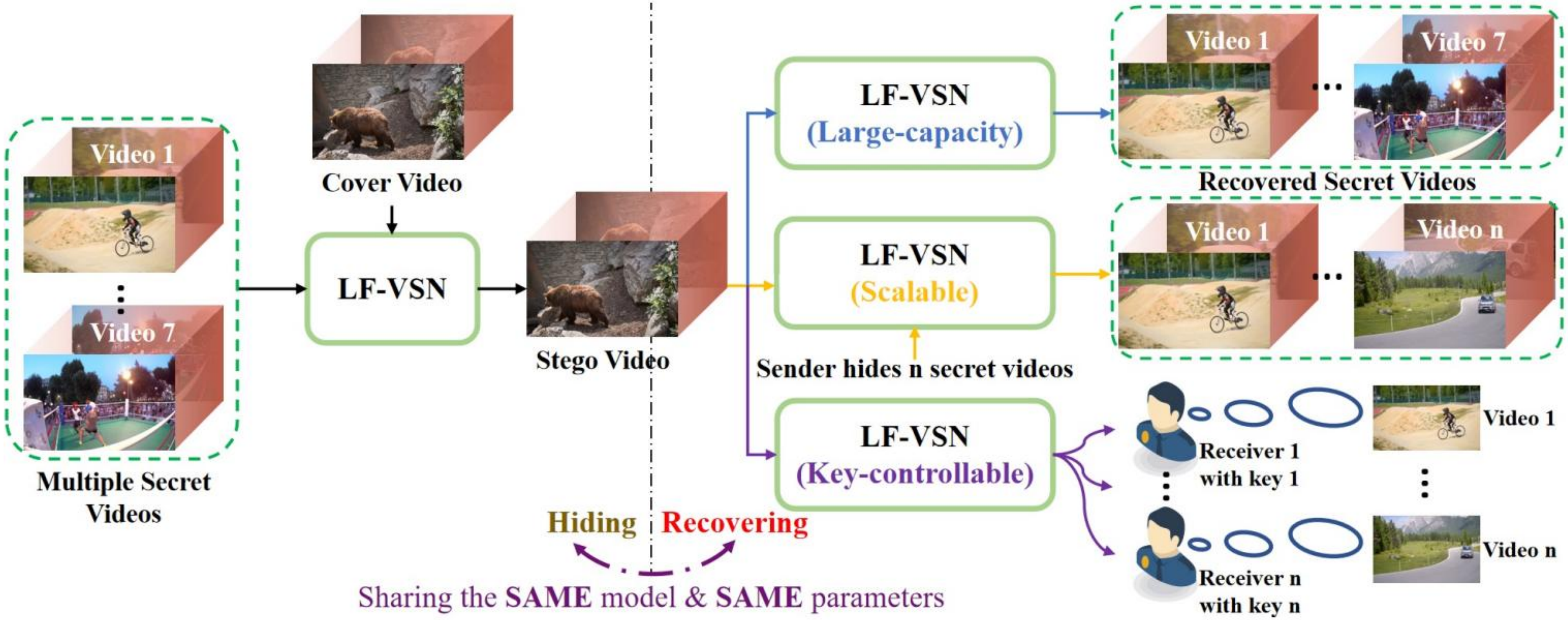
Backward:

$$Z_2 = \frac{X_2 - H(Z_1)}{F(Z_1)}$$
$$Z_1 = X_1$$

Invertible Neural Network



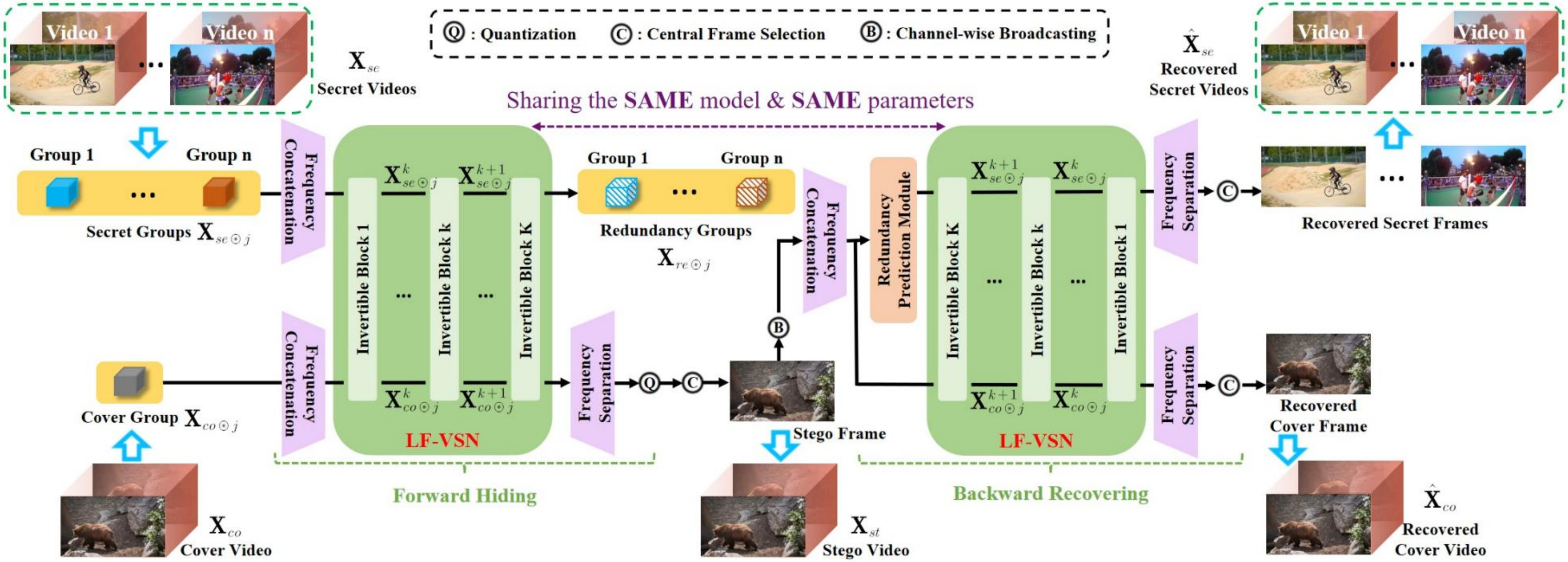
Overview



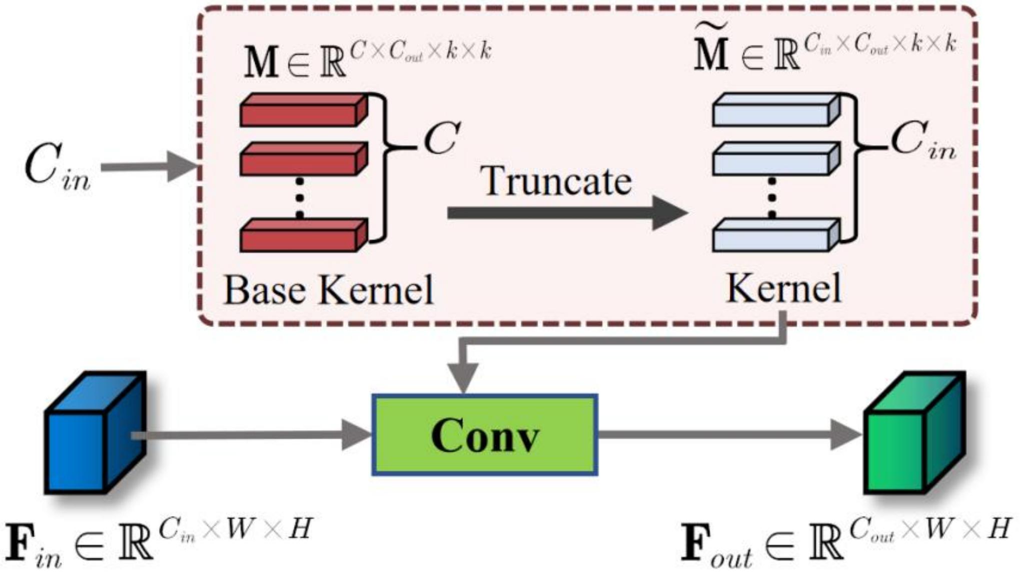
Contributions:

- ✓ Large hiding capacity
- ✓ Hiding a scalable number of secret videos into one video
- ✓ Different receivers can recover different secret videos through a specific key
- ✓ Invertible

Network Structure

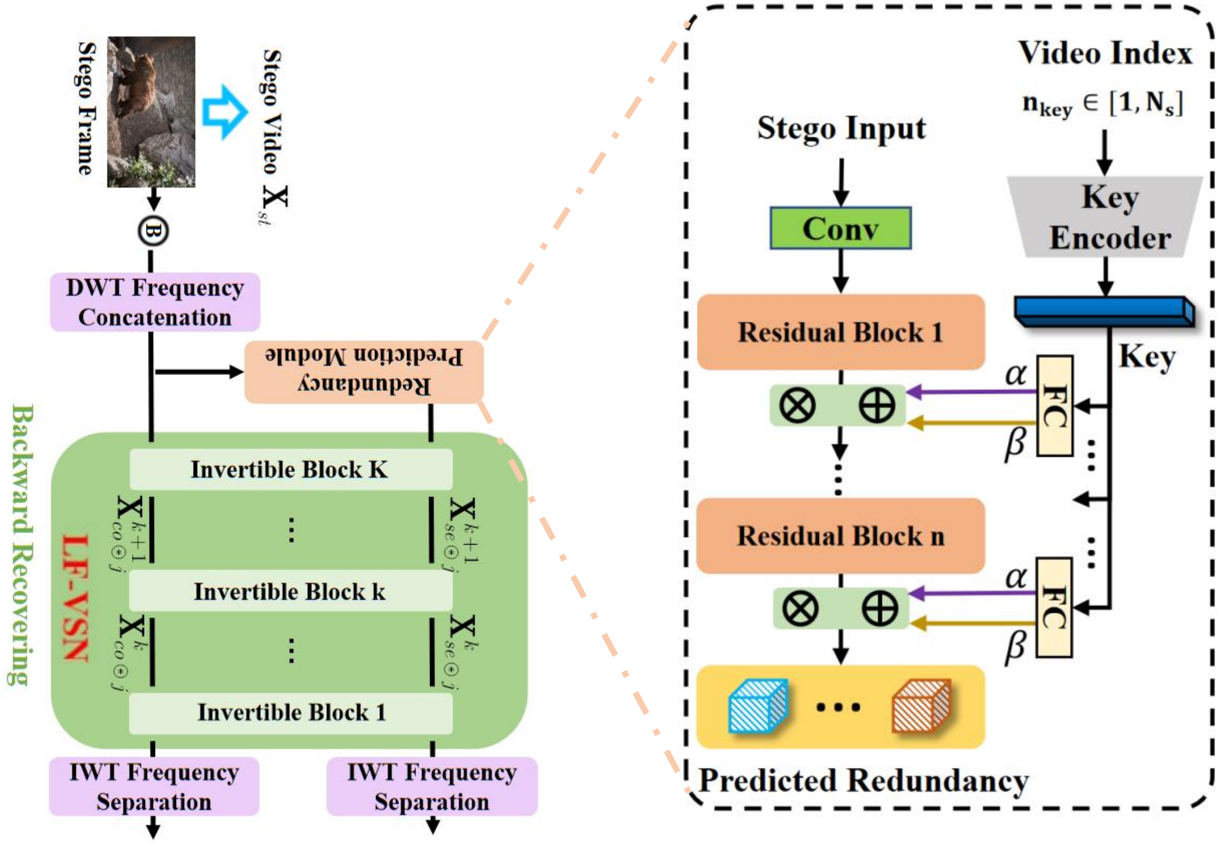


Scalable Design



- Customize the convolution kernel
- All convolution kernels share a parent matrix

Key-controllable Design



Loss Function:

- **Forward hiding:**

$$\mathcal{L}_f = ||\mathbf{X}_{st\circledast j}[I_c] - \mathbf{X}_{co\circledast j}[I_c]||_2^2,$$

- **Backward recovering (w/o key control):**

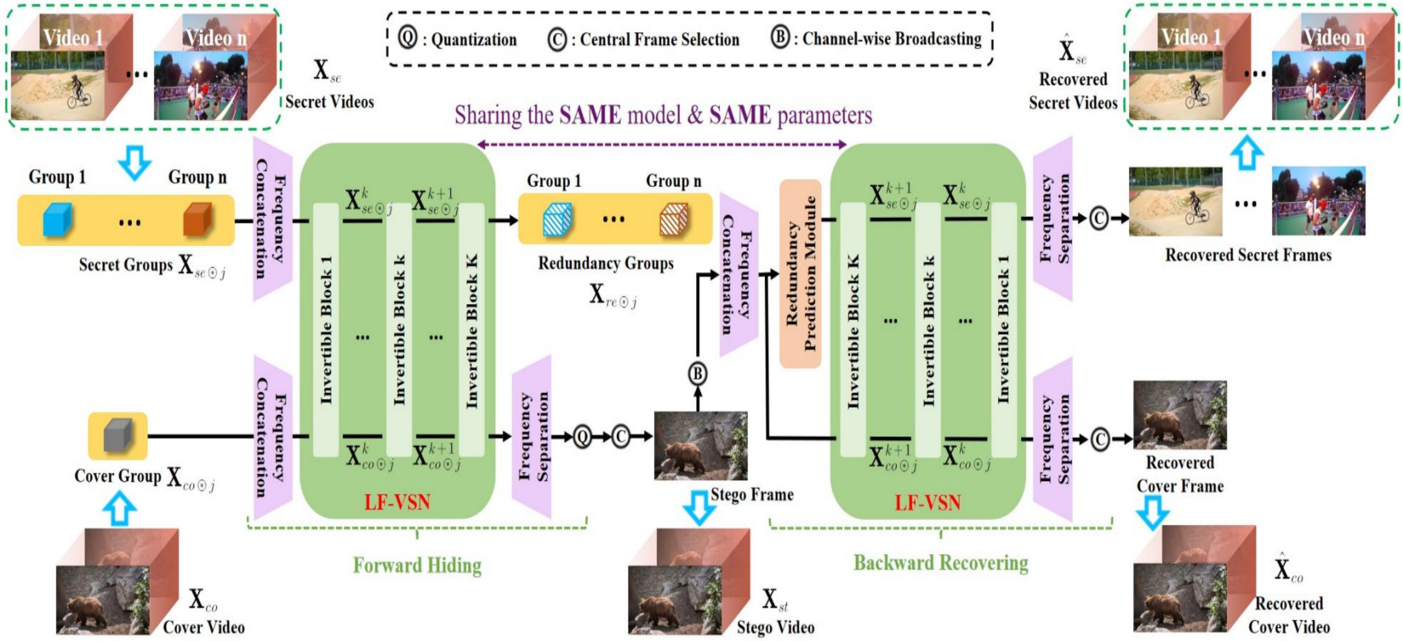
$$\mathcal{L}_b = \sum_{n=1}^{N_s} ||\hat{\mathbf{X}}_{se\circledast j}(n)[I_c] - \mathbf{X}_{se\circledast j}(n)[I_c]||_2^2 + ||\hat{\mathbf{X}}_{co\circledast j}[I_c] - \mathbf{X}_{co\circledast j}[I_c]||_2^2,$$

- **Backward recovering (w key control):**

$$\mathcal{L}_b = ||\frac{1}{N_s}\sum_{n=1}^{N_s} \hat{\mathbf{X}}_{se\circledast j}(n)[I_c] - \mathbf{X}_{se\circledast j}(n_{key})[I_c]||_2^2 + ||\hat{\mathbf{X}}_{co\circledast j}[I_c] - \mathbf{X}_{co\circledast j}[I_c]||_2^2.$$

- **Final loss:**

$$\mathcal{L} = \mathcal{L}_f + \lambda\mathcal{L}_b,$$



Experiment

Comparison on One-video Hiding:

Table 1. Quantitative comparison (PSNR/SSIM) on Vimeo-T200. The best and second-best results are **highlighted** and <u>underlined</u>. Our LF-VSN achieves the best performance in stego and secret quality with acceptable complexity.

	Weng *et al.* [43]	Baluja [4]	ISN [32]	HiNet [21]	RIIS [47]	PIH [11]	LF-VSN (Ours)
Stego	29.43/0.862	34.14/0.860	42.08/<u>0.965</u>	42.09/0.962	<u>43.50</u>/0.951	-	**45.17/0.980**
Secret	32.08/0.899	35.21/0.931	42.11/0.984	<u>44.44</u>/<u>0.991</u>	44.08/0.964	36.48/0.939	**48.39/0.996**
Params.	42.57M	<u>2.65M</u>	3.00M	4.05M	8.15M	**0.67M**	7.40M

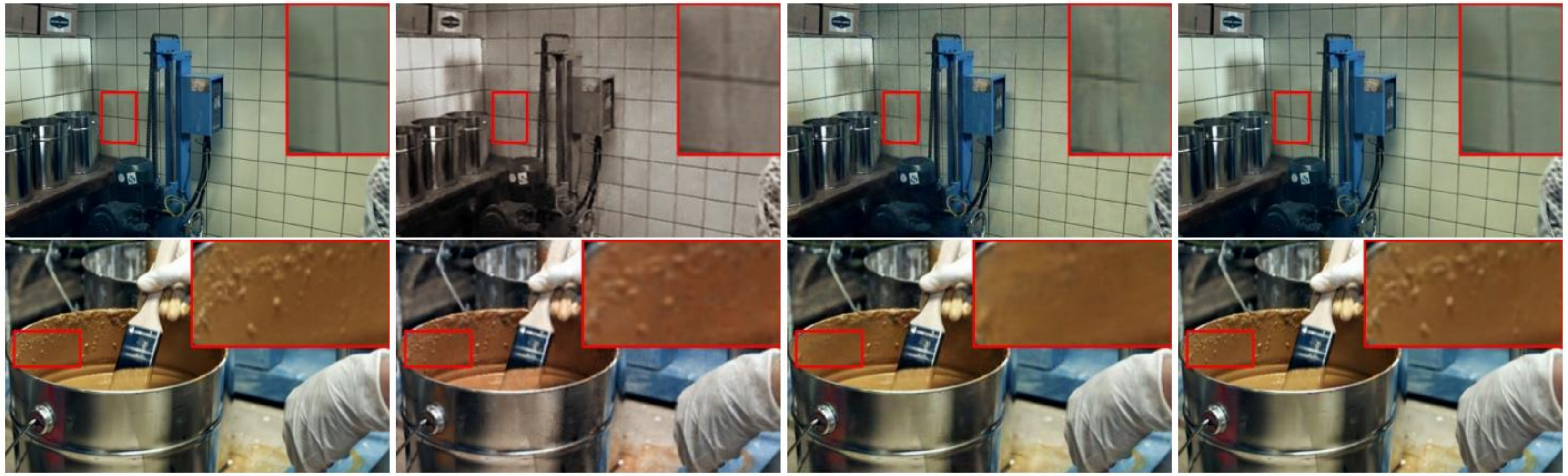


Figure 6. Visual comparison between our LF-VSN, ISN [32], and PIH [11] in 4 videos Steganography. We present the secret reconstruction results of video 2 and video 4. Our LF-VSN produces better result with intact color and details.

Comparison on Multiple-video Hiding:

Table 2. Multiple videos steganography comparison (PSNR) of our LF-VSN, ISN [32], and PIH [11] on Vimeo-T200 test set. Our LF-VSN can hide/recover 7 videos with promising performance.

	Videos	2	3	4	5	6	7
ISN	Stego	37.60	36.41	32.56	31.46	-	-
	Secret	41.47	38.76	33.42	33.39	-	-
PIH	Stego	-	-	-	-	-	-
	Secret	35.95	34.96	34.20	-	-	-
Ours	Stego	**40.97**	**38.55**	**37.55**	**36.57**	**35.68**	**35.01**
	Secret	**44.24**	**42.27**	**40.21**	**38.88**	**36.94**	**35.71**

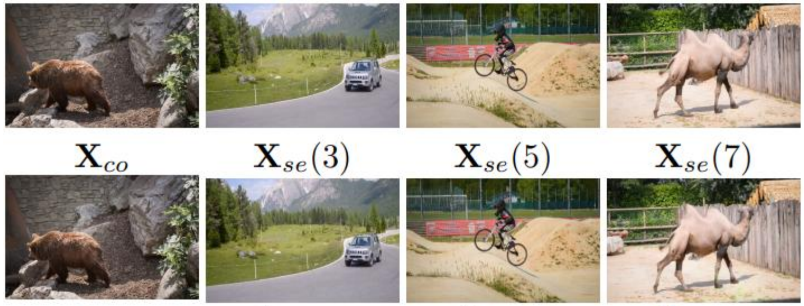


Figure 7. Visualization of our LF-VSN in 7 videos steganography, showing promising performance in such an extreme case.

Key-controllable Video Steganography:

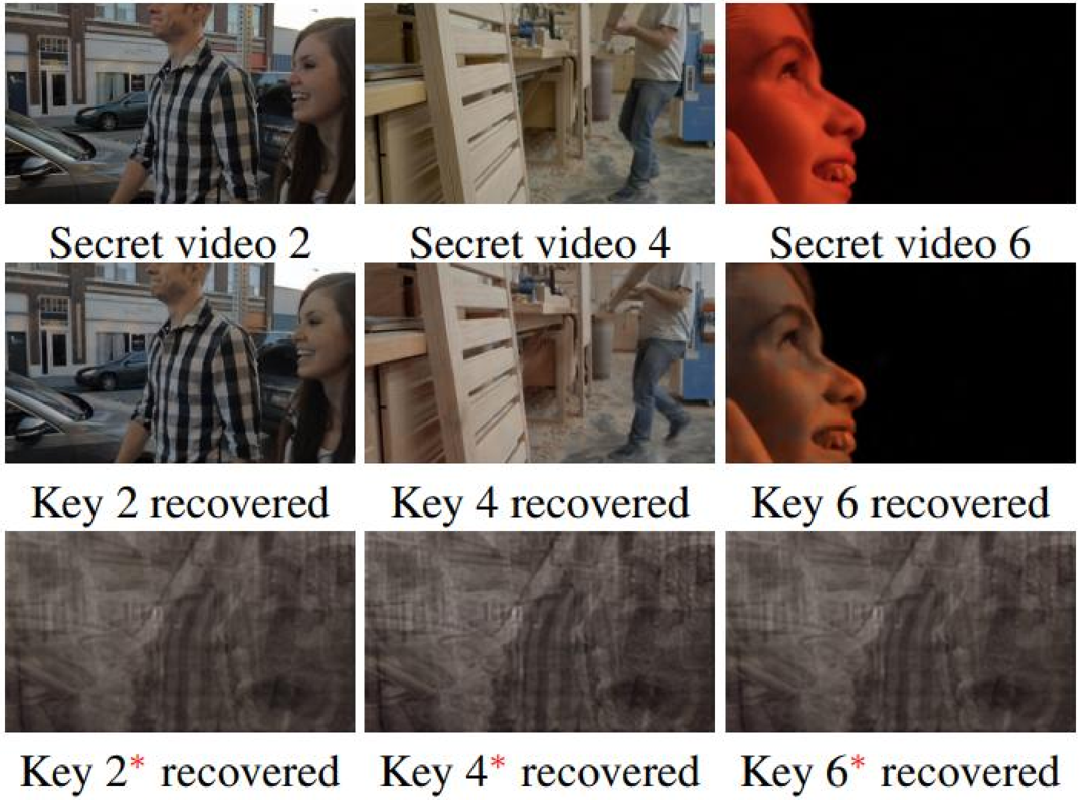


Figure 8. Visualization of our key-controllable scheme in 6 videos steganography. In the second and third rows, we use the correct and wrong (*) keys of 2, 4, 6 to recover secret videos, respectively.

Scalable Video Steganography:

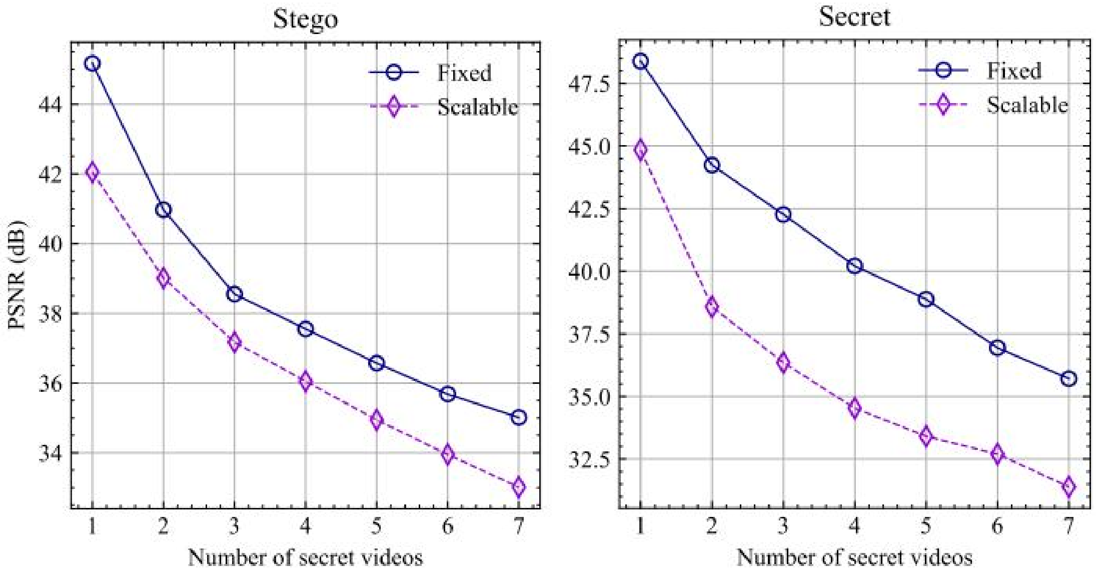


Figure 9. Performance comparison between our scalable and fixed design in multiple videos steganography.

Steganographic Analysis:

Statistics-based steganalysis:

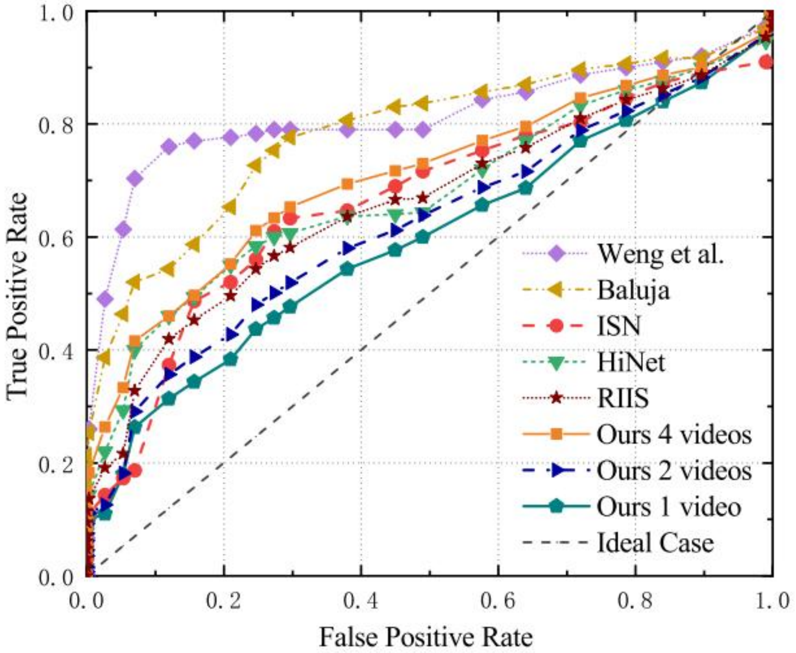


Figure 10. Statistics-based steganalysis by StegExpose [7]. The closer the detection accuracy is to 50%, the higher the security is.

Learning-based steganalysis:

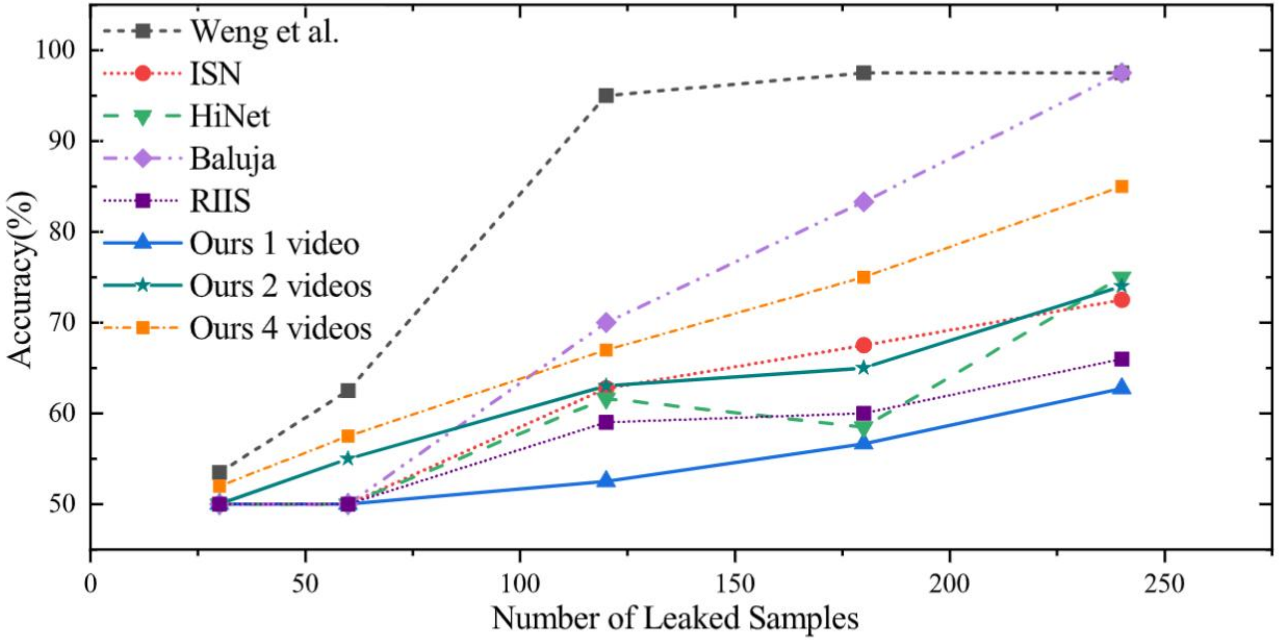


Figure 6. Deep-learning steganalysis results, which are produced by the latest Size-Independent-Detector (SID) steganalysis method [9]. The closer the detection accuracy is to 50%, the higher the steganography security is.

03

Experiment

Ablation Study:

Table 4. The ablation study of different components in our LF-VSN. It includes the sliding window size, number of invertible blocks (IB), frequency concatenation (FreqCat), and redundancy prediction module (RPM).

Num. videos		2			4			6			3		3	3
Window size	1	3	5	1	3	5	1	3	5	3	3 (ours)	3	3	3
Num. IB	16	16	16	16	16	16	16	16	16	12	16	20	16	16
FreqCat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
RPM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
Stego	39.64	40.97	41.08	36.41	37.55	37.86	34.47	35.46	35.96	38.03	38.55	38.91	38.28	36.85
Secret	42.97	44.24	44.43	37.67	40.21	40.42	35.11	36.83	39.97	41.99	42.27	42.40	41.69	40.36

Real-world Application:

Table 1. The impact of video compression. PSNR is presented.

qp	Video size	HiNet[21]	Ours 1video	Ours 3videos	Ours 5videos	Ours 7videos
0	100MB	32.04	44.15	39.55	37.11	32.05
5	80.1MB	23.05	35.23	34.32	32.47	29.87
10	35.7MB	18.61	31.25	29.49	28.30	26.94
15	21.6MB	13.45	23.55	21.78	20.97	19.80

Figure 1. Visualization of secret quality after video compression.

ffmpeg -s 1280x720 -i <input> -c:v libx264 -qp <qp> <output>

Conclusion

- ✓ **We propose a large-capacity video steganography method, which can hide/recover multiple (up to 7) secret videos in/from a cover video. Our hiding and recovering are fully reversible via a single INN.**

- ✓ **We propose a key-controllable scheme with which different receivers can recover particular secret videos from the same cover video via specific keys.**

- ✓ **We propose a scalable embedding module, utilizing a single model and a single training session to satisfy different requirements for the number of secret videos hidden in a cover video.**

- ✓ **Extensive experiments demonstrate that our proposed method achieves state-of-the-art performance with large hiding capacity and flexibility.**

感谢大家

Thanks for Your Attention!

eechongm@gmail.com

Code: https://github.com/MC-E/LF-VSN