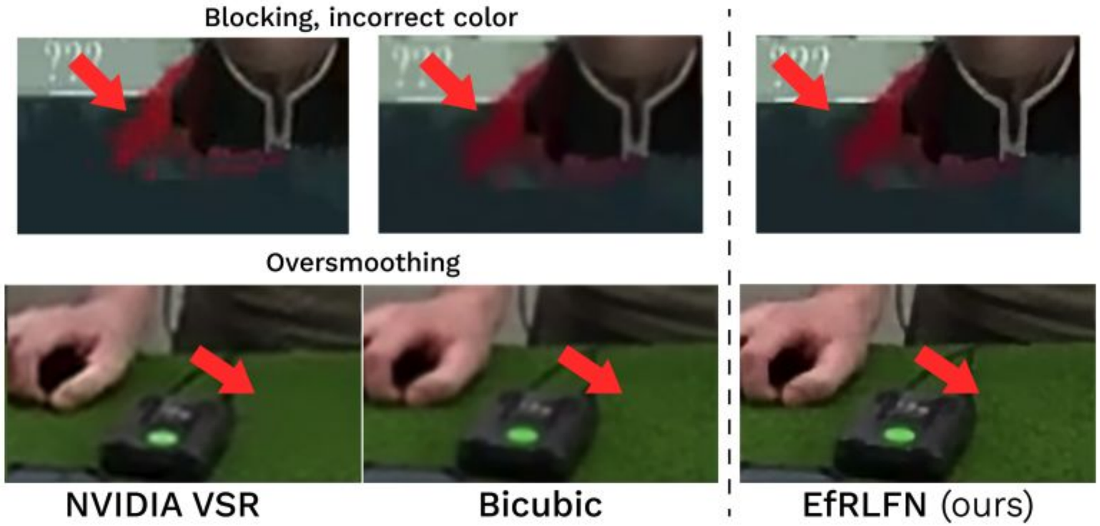
Exploring Real-Time Super-Resolution: Benchmarking and Fine-Tuning for Streaming Content

Evgeney Bogatyrev, Khaled Abud,
Ivan Molodetskikh, Nikita Alutis, Dmitriy Vatolin

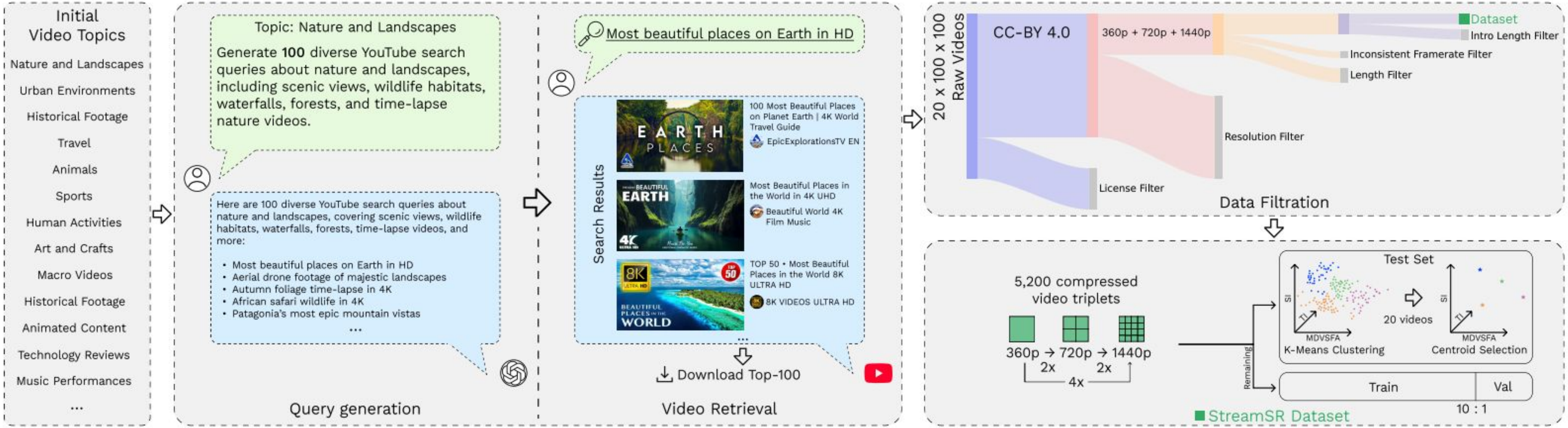
Challenges of existing SR

- Real-Time SR models are typically trained on synthetic data
- Fail to generalize to real-world compression artifacts
- Streaming video is heavily compressed
- Training data does not reflect these distortions => degraded SR performance



StreamSR Dataset

- 5,200 YouTube videos
- Real compression artifacts
- Multi-scale streams (360p–1440p)
- Large user study (3,800+ participants)



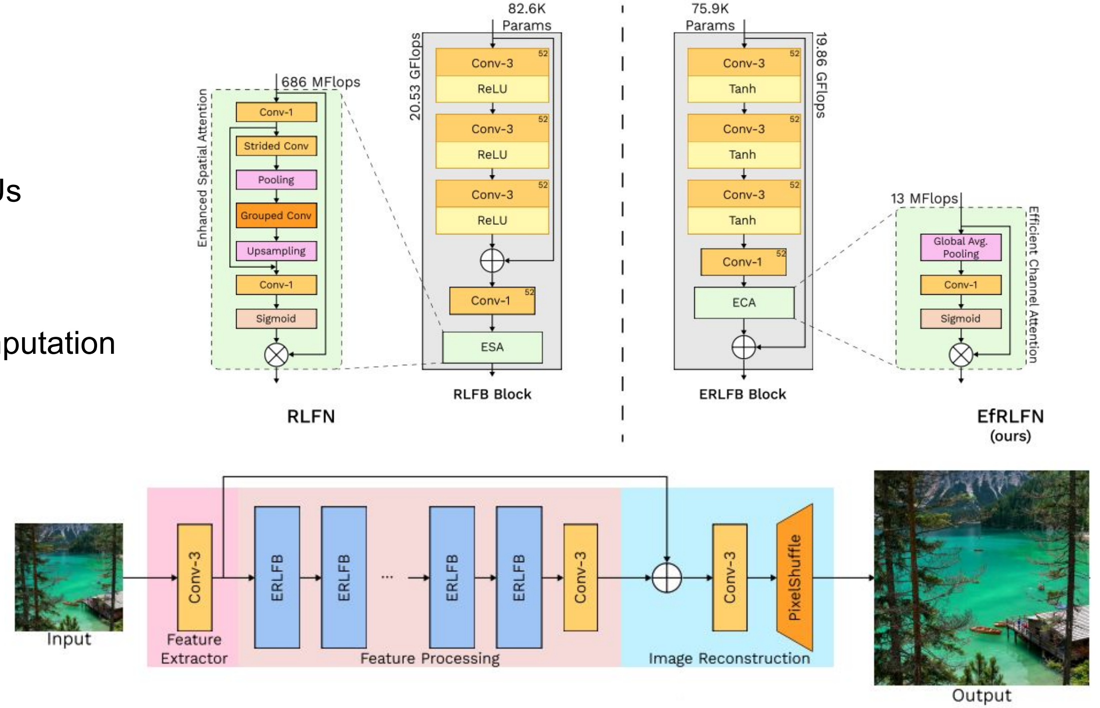
EfRLFN

EfRLFN: Efficient Real-Time SR Model

- Based on RLFN architecture
- ~0.37M parameters
- Real-time performance on consumer GPUs

Efficient Design

- Improved residual blocks for reduced computation
- Lightweight channel attention mechanism
- *tanh* activation for stable reconstruction
- Better efficiency–quality trade-off



Loss function

$$\mathcal{L} = \underbrace{\lambda_{Charb}\mathcal{L}_{Charb}}_{\text{Reconstruction}} + \underbrace{\lambda_{VGG}\mathcal{L}_{VGG}}_{\text{Perception}} + \underbrace{\lambda_{Sobel}\mathcal{L}_{Sobel}}_{\text{Edge Sharpness}}$$

$$\mathcal{L}_{Charb} = \sqrt{\|\mathbf{I}_{HR} - \mathbf{I}_{SR}\|^2 + \epsilon^2}, \quad \epsilon > 0$$

$$\mathcal{L}_{VGG} = \|\phi_{\text{VGG}}(\mathbf{I}_{HR}) - \phi_{\text{VGG}}(\mathbf{I}_{SR})\|_1,$$

$$\mathcal{L}_{Sobel} = \|S(\mathbf{I}_{HR}) - S(\mathbf{I}_{SR})\|_2^2,$$

Benchmark results

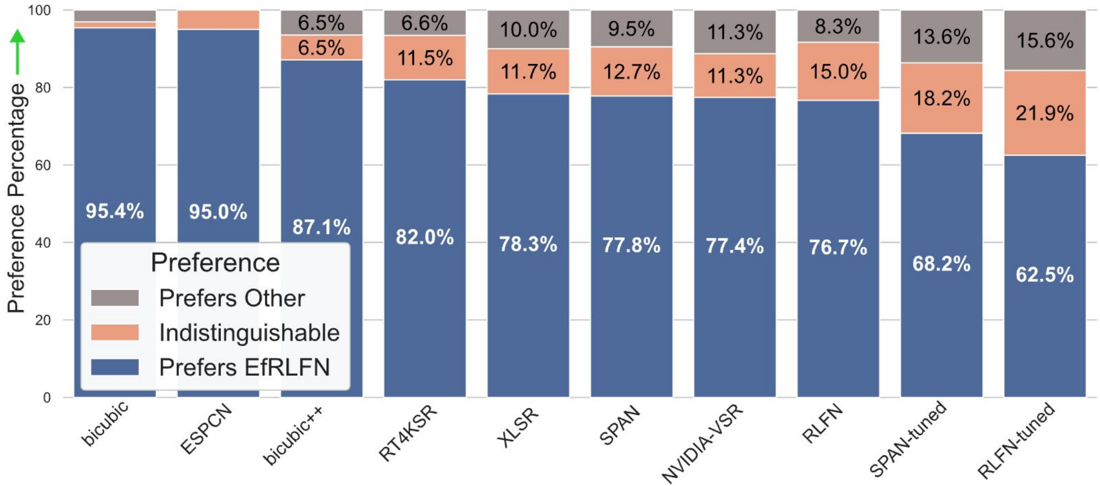
- Best perceptual quality among real-time SR models
- Competitive performance on objective metrics
- Real-time inference: 30+ FPS

Method	BSD100 SSIM↑ / LPIPS↓	Urban100 SSIM↑ / LPIPS↓	DIV2K SSIM↑ / LPIPS↓	FPS↑
AsConvSR^T	0.832 / 0.270	0.824 / 0.208	0.883 / 0.206	213±8
RT4KSR	0.805 / 0.296	0.769 / 0.244	0.860 / 0.221	102±5
RT4KSR^T	0.812 / 0.254	0.766 / 0.223	0.864 / 0.186	102±5
bicubic	0.752 / 0.386	0.711 / 0.324	0.820 / 0.292	**1829±50**
ESPCN	0.774 / 0.534	0.505 / 0.437	0.514 / 0.515	201±7
ESPCN^T	0.814 / 0.236	0.840 / 0.150	0.862 / 0.180	201±7
Bicubic++^T	0.768 / 0.360	0.725 / 0.298	0.832 / 0.270	*1629±45*
NVIDIA VSR	0.788 / 0.243	0.786 / 0.152	0.858 / 0.180	52±3
XLSR^T	0.817 / 0.244	0.828 / 0.158	0.864 / 0.183	429±15
SMFANet^T	0.803 / 0.239	0.798 / 0.153	0.865 / 0.177	327±12
SAFMN^T	0.813 / 0.236	0.808 / 0.150	0.871 / 0.175	273±10
RLFN	0.805 / 0.238	0.803 / 0.147	0.876 / 0.175	225±8
RLFN^T	0.834 / 0.239	0.845 / 0.153	0.881 / 0.178	225±8
SPAN	0.836 / *0.222*	0.841 / 0.139	0.887 / *0.168*	60±3
SPAN^T	*0.837* / 0.239	*0.847* / *0.118*	*0.890* / 0.175	60±3
EfRLFN^T	**0.847** / **0.190**	**0.856** / **0.116**	**0.892** / **0.145**	271±10

Pairwise subjective results

We conducted a user study on subjectify.us platform:

- 37K+ pairwise comparisons conducted
- 3,800+ participants
- Strong user preference for EfRLFN outputs



Conclusion

Contributions:

- **StreamSR**: realistic dataset for streaming video SR
- **EfRLFN**: efficient, real-time SR model
- Improved perceptual quality on compressed streaming video

Code and Dataset:
https://github.com/EvgeneyBogatyrev/EfRLFN