SCALED: Surrogate-gradient for Codec-Aware Learning of Downsampling in ABR Streaming

Esteban Pesnel [1,2] Julien Le Tanou [1] Michael Ropert [1] Thomas Maugey [2] Aline Roumy [2]

[1]MediaKind, Rennes, France [2]INRIA, Rennes, France

ABR streaming

- **ABR (Adaptive Bit Rate):** Match video bitrate and resolution to client bandwidth and device.
- **Pipeline:** Server (downscale/encode) $\rightarrow$ Client (decode/upscale).

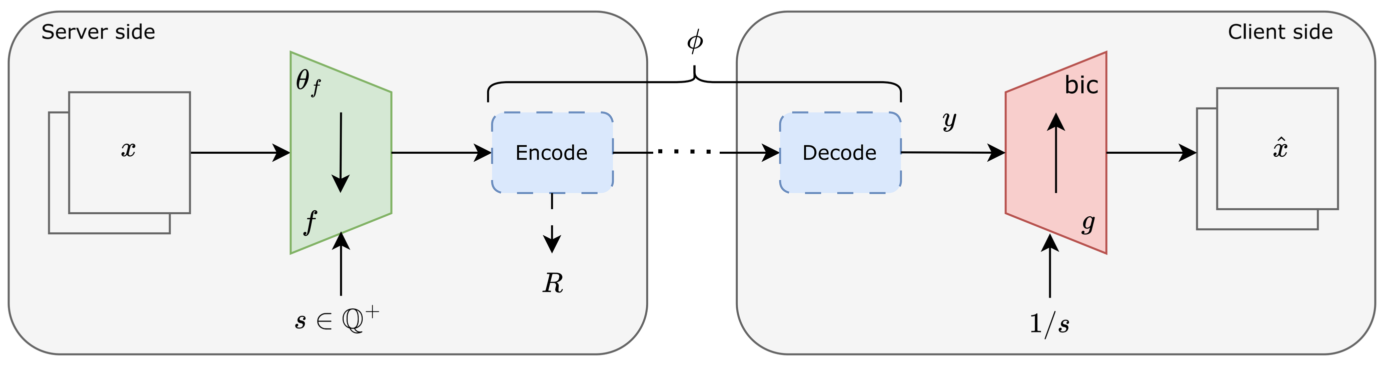

Figure 1. ABR streaming pipeline

Problem formulation

- **Constraint :** Fixed bicubic upscaler (standard set-top box)
- **Problem:** Given fixed upscaling, fixed codec, learn donwscaler f optimizing **end-to-end rate-distortion**

$$\boldsymbol{\theta}_f^* = \arg\min_{\boldsymbol{\theta}_f} \left\| \boldsymbol{x} - g\big(\phi\big(f(\boldsymbol{x};\boldsymbol{\theta}_f)\big)\big) \right\|_2^2 + \lambda R\big(\phi\big(f(\boldsymbol{x};\boldsymbol{\theta}_f)\big)\big) \tag{1}$$

Differentiability issue

Challenge: Learned downscaler f while codec ϕ being non differentiable.
Existing solutions: Substitute ϕ with a differentiable proxy $\hat{\phi}$:

- **Bypass**: $\phi = 1$.
- **Surrogate model:** NN, codec proxies.
- **Straight-Through-Estimator (STE):** Uses true codec ϕ at the forward, and assumes identity gradient at the backward ($\frac{\delta \boldsymbol{y}}{\delta f} = \mathbf{1}^{N\times M}$):

$$y = f(\boldsymbol{x};\boldsymbol{\theta}_f) + sg(\underbrace{\phi(f(\boldsymbol{x};\boldsymbol{\theta}_f)) - f(\boldsymbol{x};\boldsymbol{\theta}_f)}_{\boldsymbol{\epsilon}}) \tag{2}$$

	Design Complexity	Fidelity	
		Forward	Backward (Grad.)
Bypass	✓	×	⊗
Proxy model	×	∼	∼
STE	✓	✓	⊗

Table 1. Comparison of Bypass, Proxy and STE across complexity and fidelity criteria.

Question: How to improve backward fidelity while keeping low complexity?

A novel surrogate gradient for codec aware downscaling : SCALED

Based on [?], introduce function of $\boldsymbol{\epsilon}$ that is easier to differentiate (here, $\sigma(\boldsymbol{\epsilon})$):

$$\boldsymbol{y} = f(\boldsymbol{x};\boldsymbol{\theta_f}) + sg(\boldsymbol{\epsilon}) \cdot \frac{\sigma(\boldsymbol{\epsilon})}{sg(\sigma(\boldsymbol{\epsilon}))} \tag{3}$$

Consequently, the gradient w.r.t the input f becomes **dependent on the quantization error** ϵ:

$$\frac{\partial \boldsymbol{y}}{\partial f} = \mathbf{1}^{N\times M} + \frac{\boldsymbol{\epsilon}}{\sigma(\boldsymbol{\epsilon})} \odot \frac{\partial \sigma(\boldsymbol{\epsilon})}{\partial f} \tag{4}$$

Benefits:

- **Accurate backpropagation:** gradient explicitly depends on the true $\boldsymbol{\epsilon}$.
- **General:** The formulation is compatible with a wide range of learning objectives.

Application 1: Optimizing MSE only (SCALED$_D$)

Leveraging our surrogate gradient, we minimize the **end-to-end reconstruction error**:

$$\boldsymbol{\theta_f^*} = \arg\min_{\boldsymbol{\theta_f}} \left\| \boldsymbol{x} - g\left(f(\boldsymbol{x};\boldsymbol{\theta_f}) + sg(\boldsymbol{\epsilon}) \cdot \frac{\sigma(\boldsymbol{\epsilon})}{sg(\sigma(\boldsymbol{\epsilon}))} \right) \right\|_2^2 \tag{5}$$

Application 2: Optimizing Rate-Distortion (SCALED$_{RD}$)

We extend SCALED$_D$ by adding a **differentiable rate regularization** $\hat{R}$ [?]:

$$\boldsymbol{\theta_f^*} = \arg\min_{\boldsymbol{\theta_f}} \left\| \boldsymbol{x} - g\left(f(\boldsymbol{x};\boldsymbol{\theta_f}) + sg(\boldsymbol{\epsilon}) \cdot \frac{\sigma(\boldsymbol{\epsilon})}{sg(\sigma(\boldsymbol{\epsilon}))} \right) \right\|_2^2 + \lambda \hat{R} \tag{6}$$

Performance assessments

- **Evaluation:** Comparison based on the **convex hull** of R-D curves for various scale ratios.
- **Encoding settings:** Sequences are encoded using x264 with the `medium` preset.
- **Training strategy:** A separate model is trained for each individual target QP.

Filter	Training dataset	Training	Codec aware	XIPH				UVG			
				PSNR	SSIM	VMAF	VMAF$_{NEG}$	PSNR	SSIM	VMAF	VMAF$_{NEG}$
ProgDownLite$_{RGB}$	DIV2K	D-only [?]	No	-0.66	-1.24	-5.67	-3.58	1.22	-1.78	-7.72	-5.24
ProgDownLite$_{YUV}$	Google [?]	D-only [?]	No	-0.85	-1.19	-4.71	-3.60	1.91	-2.31	-6.11	-5.02
ProgDownLite$_{YUV}$	Google [?]	STE (2)	Yes	68.57	48.34	11.22	20.69	49.39	45.33	4.46	5.06
ProgDownLite$_{YUV}$	Google [?]	Proxy [?]	Yes	-1.47	-1.87	-2.58	-1.67	-0.08	-2.99	-5.88	-4.79
ProgDownLite$_{YUV}$	Google [?]	SCALED$_D$ (6)	Yes	-4.67	**-5.21**	**-9.80**	**-7.78**	-2.84	**-5.79**	**-10.95**	**-9.07**
ProgDownLite$_{YUV}$	Google [?]	SCALED$_{RD}$ (7)	Yes	**-6.11**	-4.29	-8.80	-7.59	**-5.07**	-5.21	-8.88	-8.02

Table 2. Comparison of filtering methods on the XIPH and UVG datasets. Metrics are BD-BR relative to the convex-hull Lanczos baseline.

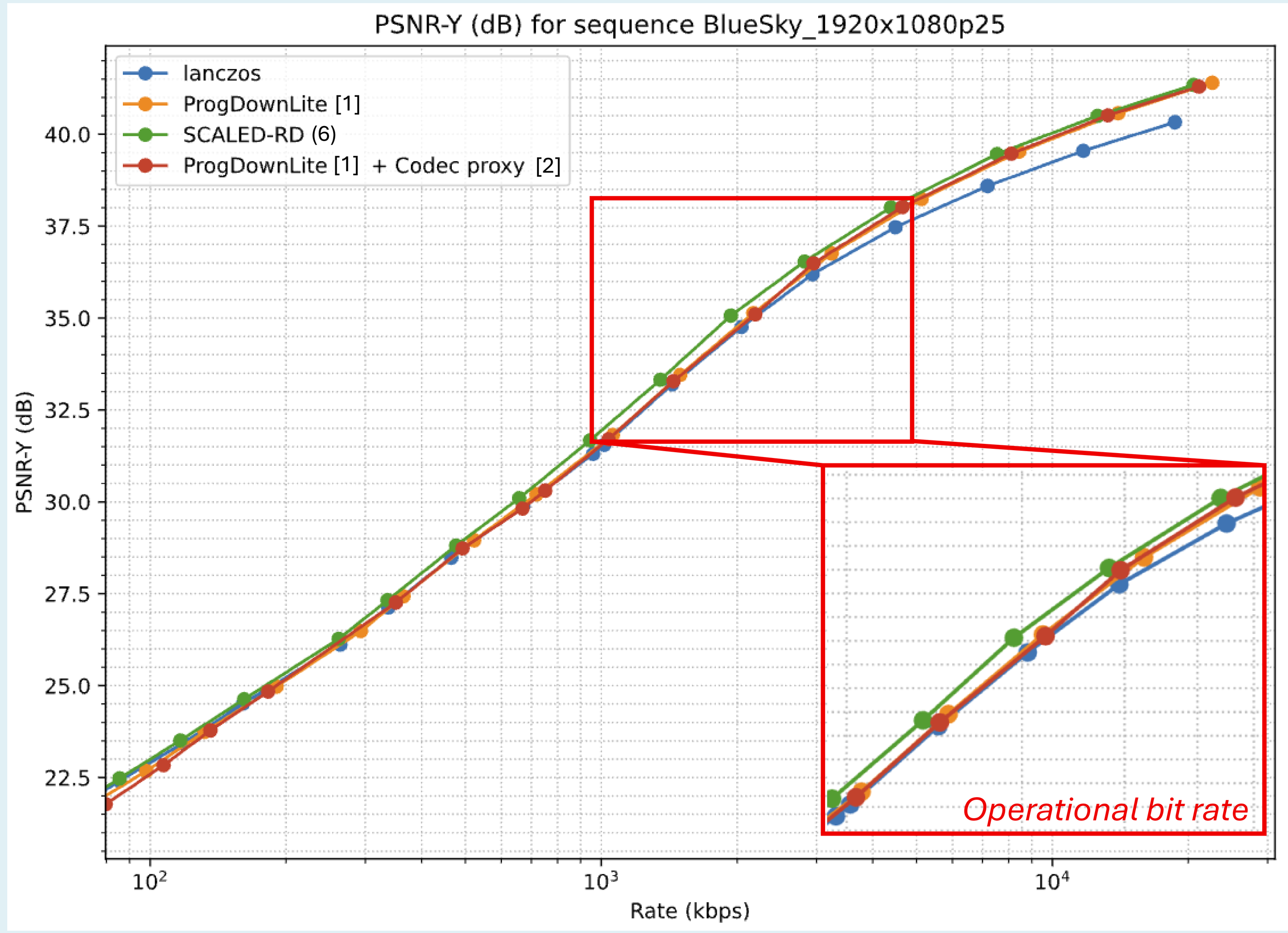

Figure 2. Convex hull comparison for the content BlueSky, from XIPH dataset

Figure 3. Pre-encoding frame Lanczos downscaler.

Figure 4. Pre-encoding frame from ProgDownLite [?].

Figure 5. Pre-encoding frame from SCALED$_{RD}$ downscaler (QP 17 model).

Figure 6. Pre-encoding frame from SCALED$_{RD}$ downscaler (QP 47 model).

Conclusion

We introduced **SCALED**, a novel surrogate gradient enabling **end-to-end optimization** with **true non-differentiable codecs**.

- **Design complexity:** Eliminates the need for complex, trainable proxies.
- **Fidelity:** Accurate forward/backward based on true codec.
- **Performance:** Outperforms SOTA methods (STE, Proxies), achieving up to **6.11% BD-Rate savings** (PSNR) on XIPH.
- **Adaptability:** The model learns to smooth content based on the target QP to mitigate compression artifacts.

References

[1] Li-Heng Chen, Christos G Bampis, Zhi Li, Joel Sole, Chao Chen, and Alan C Bovik. "Learned fractional downsampling network for adaptive video streaming." *Signal Processing: Image Communication*, 128:117172, 2024.

[2] Onur G Guleryuz, Philip A Chou, Berivan Isik, Hugues Hoppe, Danhang Tang, Ruofei Du, Jonathan Taylor, Philip Davidson, and Sean Fanello. "Sandwiched compression: Repurposing standard codecs with neural network wrappers." *arXiv preprint arXiv:2402.05887*, 2024.

[3] Wolfgang Mack, Ahmed Mustafa, Rafał Łaganowski, and Samer Hijazy. "Efficient evaluation of quantization-effects in neural codecs." *arXiv preprint arXiv:2502.04770*, 2025.

esteban.pesnel@mediakind.com