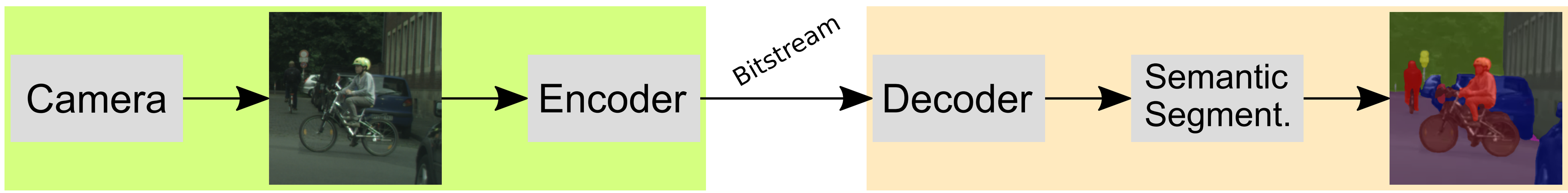
Video Coding for Machines: Large-Scale Evaluation of DNNs Robustness to Compression Artifacts for Semantic Segmentation

Alban Marie[1] Karol Desnos[1] Luce Morin[1] Lu Zhang[1]
[1] Univ Rennes, INSA Rennes, CNRS, IETR - UMR6164; Rennes, France
{Alban.Marie, Karol.Desnos, Luce.Morin, Lu.Ge}@insa-rennes.fr

Example of a Video Coding for Machines Pipeline

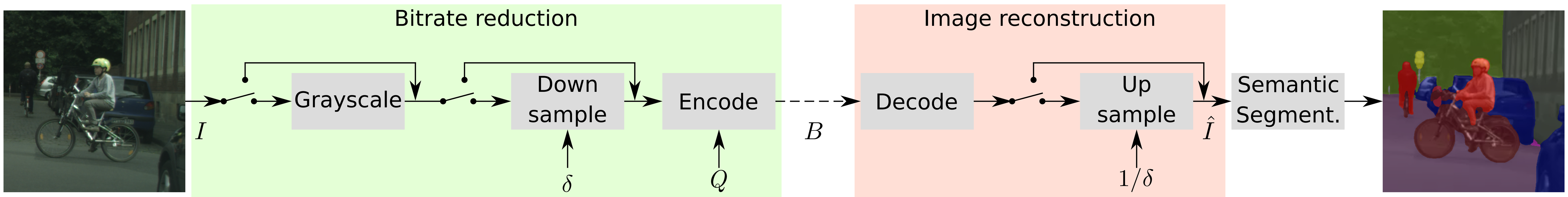


✔ Low computation on camera side.
✔ Low data transmission to the vision task side.
✘ Degradation introduced by lossy compression.

How resilient a semantic segmentation algorithm is to various compression artifacts?

Considered Coding Configurations

Which encoding strategy should be followed to achieve optimal bitrate accuracy trade-off?



A total of 1486 coding configurations are considered, including:

✔ Images with or without colors ✔ Wide range of image resolutions ✔ 5 codecs from JPEG to VVenC ✔ Wide range of bitrates

Progressive Training Procedure

✘ A vision task trained on pristine images performs poorly on distorted content.

✔ Mitigated by re-training with distorted images using the proposed progressive training:

$$f(e) = p_\infty + \Delta p \lfloor \frac{1}{\Delta p}(p_0 - p_\infty)\exp(-se) \rceil$$

✔ Progressive training allows to re-train one DNN on a large amount of distortion at once, ranging from undistorted to highly distorted.

[1] Kristian Fischer, Christian Blum, Christian Herglotz, and André Kaup. Robust Deep Neural Object Detection and Segmentation for Automotive Driving Scenario with Compressed Image Data. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2021.

Comparison of the proposed progressive training with other training strategies.

	Complexity↓	BDR↓
Baseline	—	486.57%
Separate train	100.00%	0.00%
Data augm. [1]	46.99%	21.92%
ours, $s = 0.085$	**26.51%**	16.44%
ours, $s = 0.045$	48.19%	9.51%
ours, $s = 0.025$	86.75%	**−2.62%**

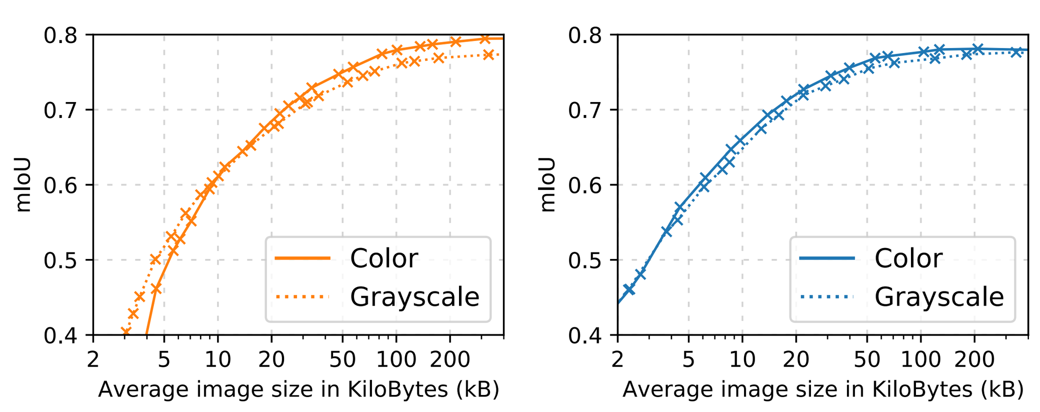
Experimental Results

BDR gains with* or without optimal resolution.

	JPG	JPG*	JM	JM*	x265	x265*	VVenC	VVenC*
JPG	0.0	139.6	53.3	125.7	123.3	178.6	193.8	224.1
JPG*	−58.3	0.0	−43.1	27.6	−2.7	68.6	26.8	73.0
JM	−34.8	75.9	0.0	99.1	71.8	157.5	115.3	179.6
JM*	−55.7	−21.6	−49.8	0.0	−14.2	34.6	4.3	37.1
x265	−55.2	2.7	−41.8	16.5	0.0	50.3	24.7	63.0
x265*	−64.1	−40.7	−61.2	−25.7	−33.5	0.0	−19.9	2.0
VVenC	−66.0	−21.1	−53.5	−4.2	−19.8	24.9	0.0	32.1
VVenC*	−69.1	−42.2	−64.2	−27.1	−38.7	−2.0	−24.3	0.0

✔ 58.3%, 49.8%, 33.5% and 24.3% bitrate savings with optimal image resolution for JPG, JM, x265 and VVenC, respectively.

Rate-mIoU trade-off with or without chroma channels for JPG and VVenC.



✘ BDR increase when removing chrominance channels, except at very low rates for JPG/J2K

JPG BDR gains over VTM using CTC. (i) Re-trained (ii) Optimal resolution (iii) Color and grayscale

(i)	(ii)	(iii)	BDR
			644.68%
✓			−4.06%
✓	✓		−73.41%
✓		✓	−25.65%
✓	✓	✓	−76.13%

✔ JPG can significantly outperform VTM with re-training and optimal image resolution