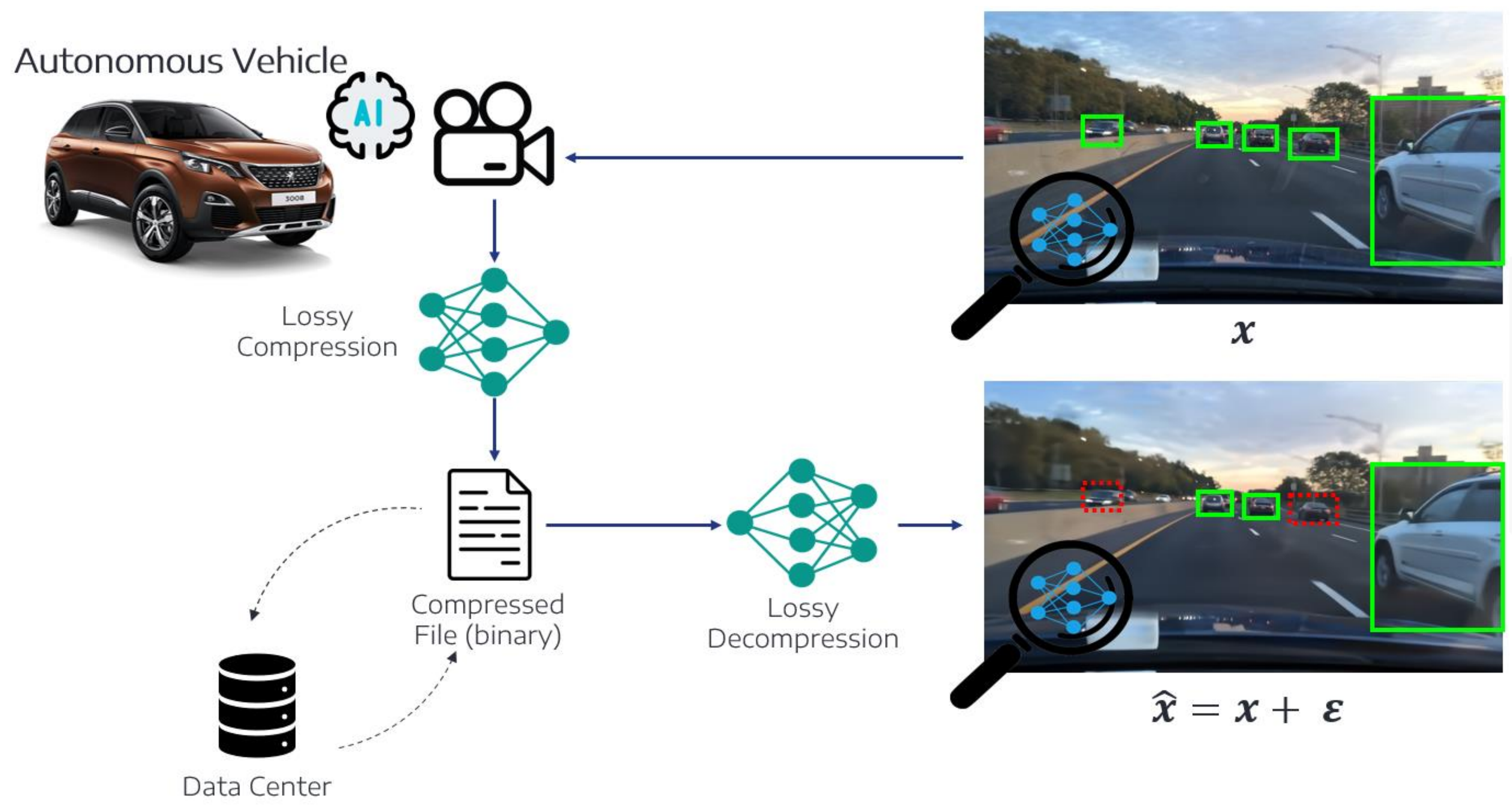
Obtaining AI system performance guarantees in the context of autonomous vehicles applied to image compression

To ensure that Autonomous Vehicles will make a safe decision in all life situations encountered, the Field Monitoring system was designed to enable vehicles to record (upon predefined triggers), transfer and store sequences of images (i.e. videos) in a Data Center, with the aim to reduce the amount of unknown, and especially unsafe, scenarios the vehicle might not have been trained for. In this context, with the objective of reducing data transfer and storage costs, we set out to supplement the Field Monitoring system with an image compression system, with the particularity of being optimized using deep neural networks. However, such compression must not subsequently too much impact the manual or automatic analyzes (according to a pre-defined acceptable level of error) performed on these images.

Field Monitoring System



1. Impact of Learned Image Compression on YOLOv7

Learned Image Compression : STF [1], TinyLIC [2] and Qres-VAe[3]
Datasets & classes : {Waymo,BDD100K} & {vehicle, pedestrian, cyclist}
Loss function :

$$L = R + \lambda . D(x, \hat{x})$$

where R is the bit rate (bpp : bits per pixel), D is the distortion between the raw image x and reconstructed image $\hat{x}$ (usually MSE or MS-SSIM), and λ is a hyperparameter that controls the trad-off between rate and distortion. For each λ we get a set of weights for a mean bpp and quality

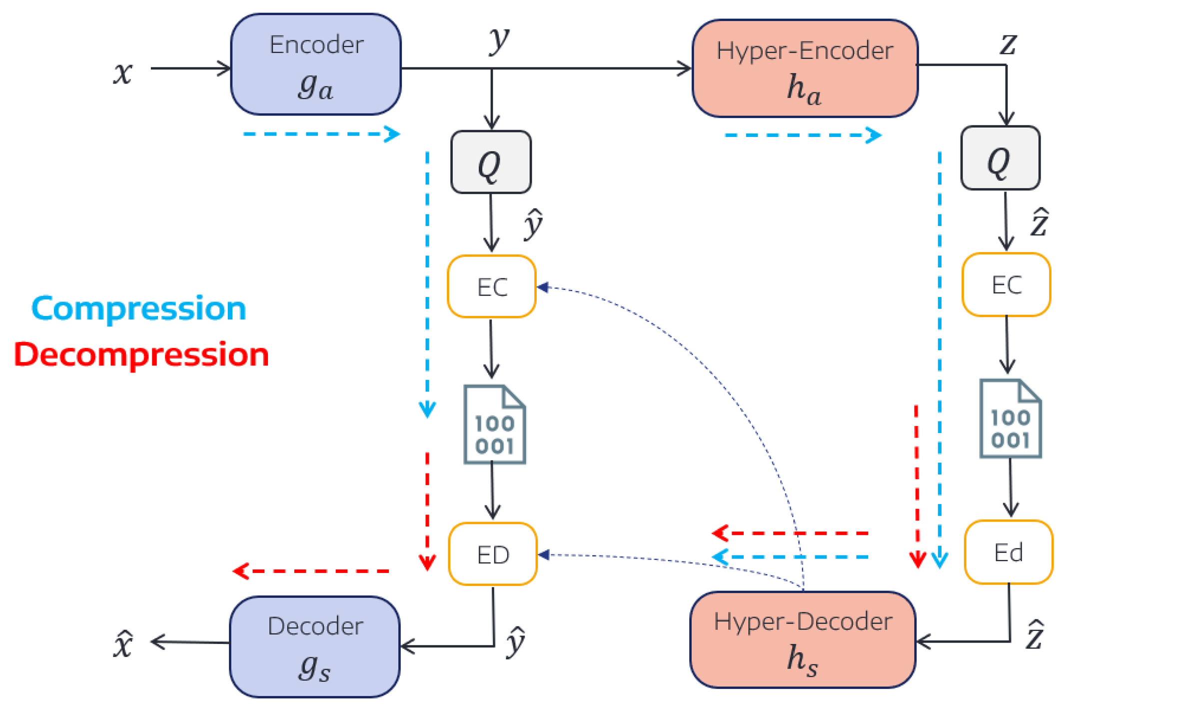


Fig 1. Common Learned Image Compression model architecture

Detection w/o compression | *Detection w/ compression at 0.313 bpp* | *Detection w/ compression at 0.045 bpp*

Fig 2: Impact of STF (mse) compression on YOLOv7

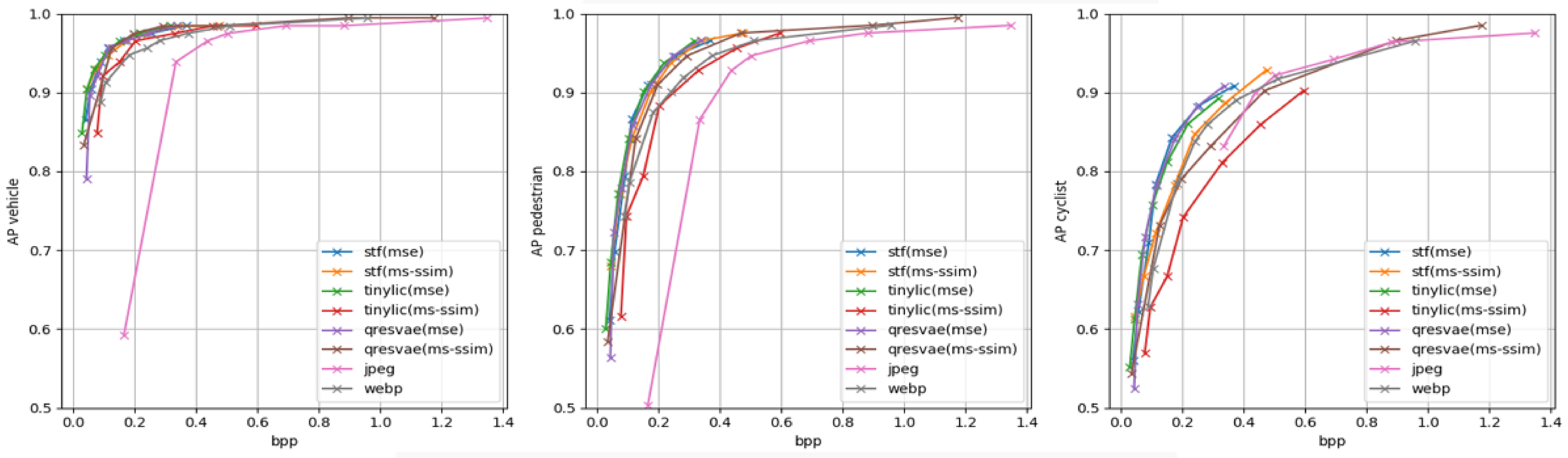


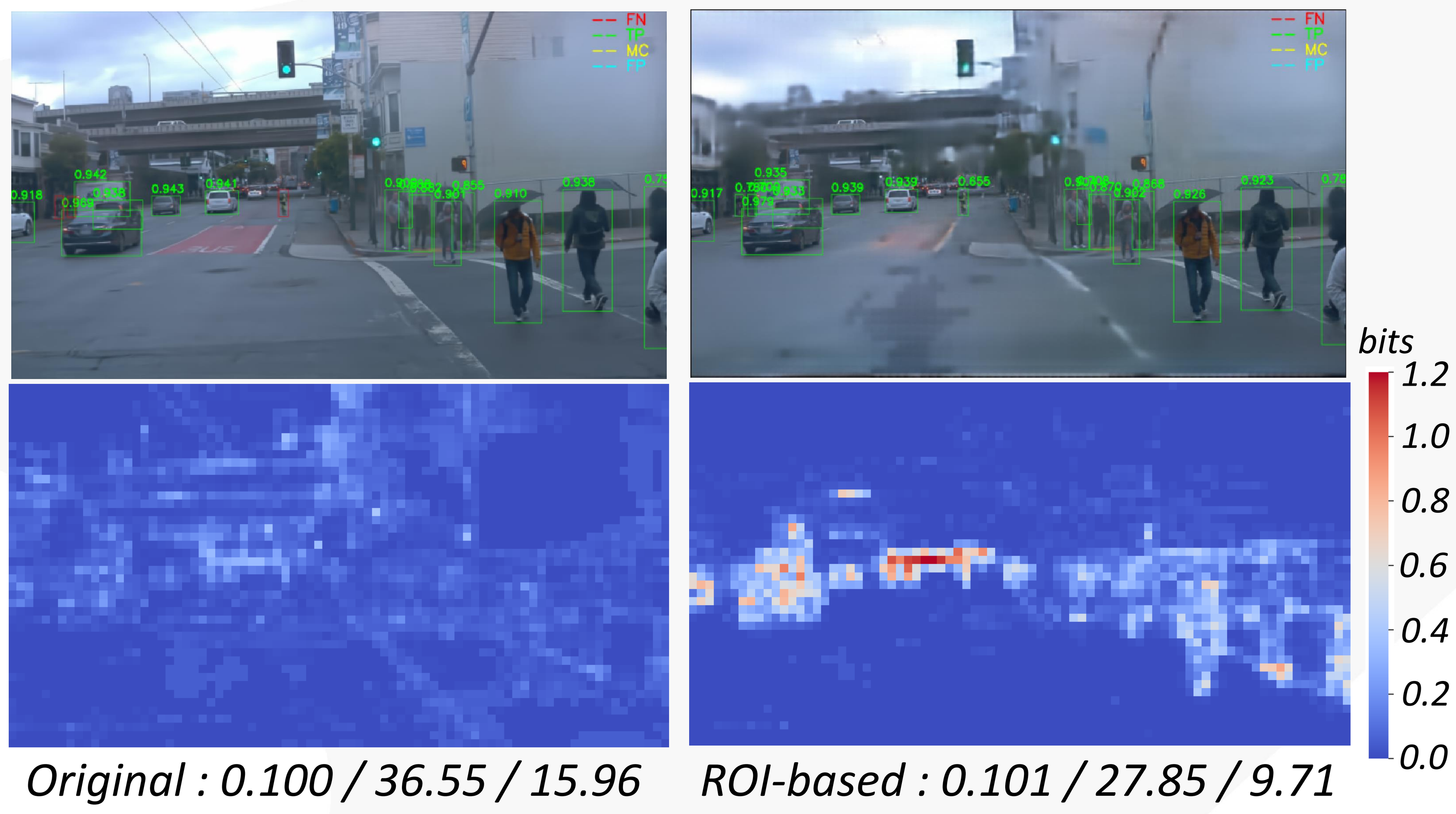
Fig 3 Average Precision per class between detection w/ and w/o compression depending on bpp (errors are mostly False Negatives and Mis-Classification)

2. ROI-based Learned Image Compression

Loss function (with MSE) :

$$L = R + \frac{1}{N} \sum_{i}^{N} \lambda_i (x_i - \hat{x}_i)^2$$

where λ_i is set high (λ^+) if the pixel belongs to an object and low (λ^-) otherwise, and N is the total number of pixels in the image.



Original : 0.100 / 36.55 / 15.96 *ROI-based : 0.101 / 27.85 / 9.71*

Fig 4: Qualitative & bit allocation comparison between Original and Roi-based STF compression (bpp / psnr / ms-ssim)

Impact on YOLOv7 (Average Precision) :

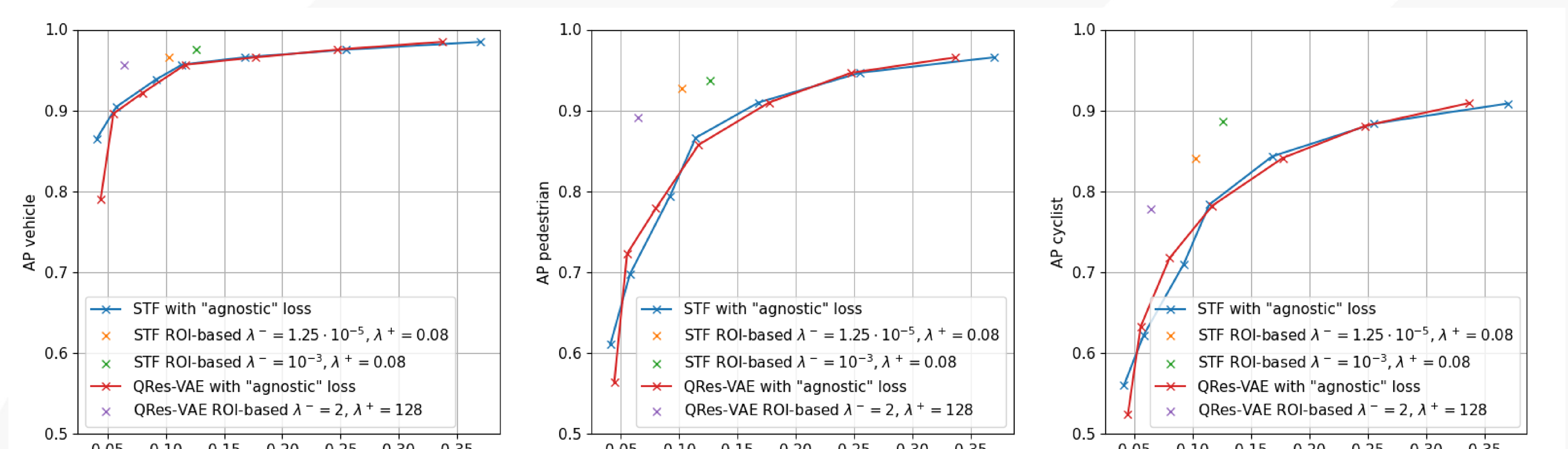


Fig 5: Average Precision per class between detection w/ and w/o compression depending on bpp (errors are mostly False Negatives and Mis-Classification)

3. Future work : Performance Guarantees

Looking for a methodology to estimate :

$$P\left(AP^c_{f(x)}\left[f\left(g_{\tau_i}(x)\right)\right] > 1 - \epsilon\right)$$

with AP^c the Average Precision for the object class c, f the object detector, g_{τ_i} the compressor for a certain bit rate τ_i and ϵ the pre-defined acceptable level of error.

References

[1] Zou and al., The Devil Is in the Details : Window-based Attention for Image Compression, CVPR 2022.
[2] Lu and al., High-Efficiency Lossy Image Coding Through Adaptive Neighborhood Information Aggregation, arXiv:2204.11448, 2022.
[3] Duan and al., Lossy Image Compression with Quantized Hierarchical VAEs, WACV 2023

PhD Student : JASPAR Jérémy
Supervisor : VIENNET Emmanuel (Paris 13 University, L2TI Laboratory)
Tutors : GUALANDRIS David & SAUVAGET Jean-Louis (Stellantis)

02/2022 – 02/2025