July 2022

PATCH-BASED END-TO-END IMAGE LEARNED CODECS USING OVERLAPPING

Marwa TARCHOULI

Thomas Guionnet
Sebastien Pelurson

Olivier Deforges
Wassim Hamidouche
Meriem Outtas

ATEME
Captivate your audience

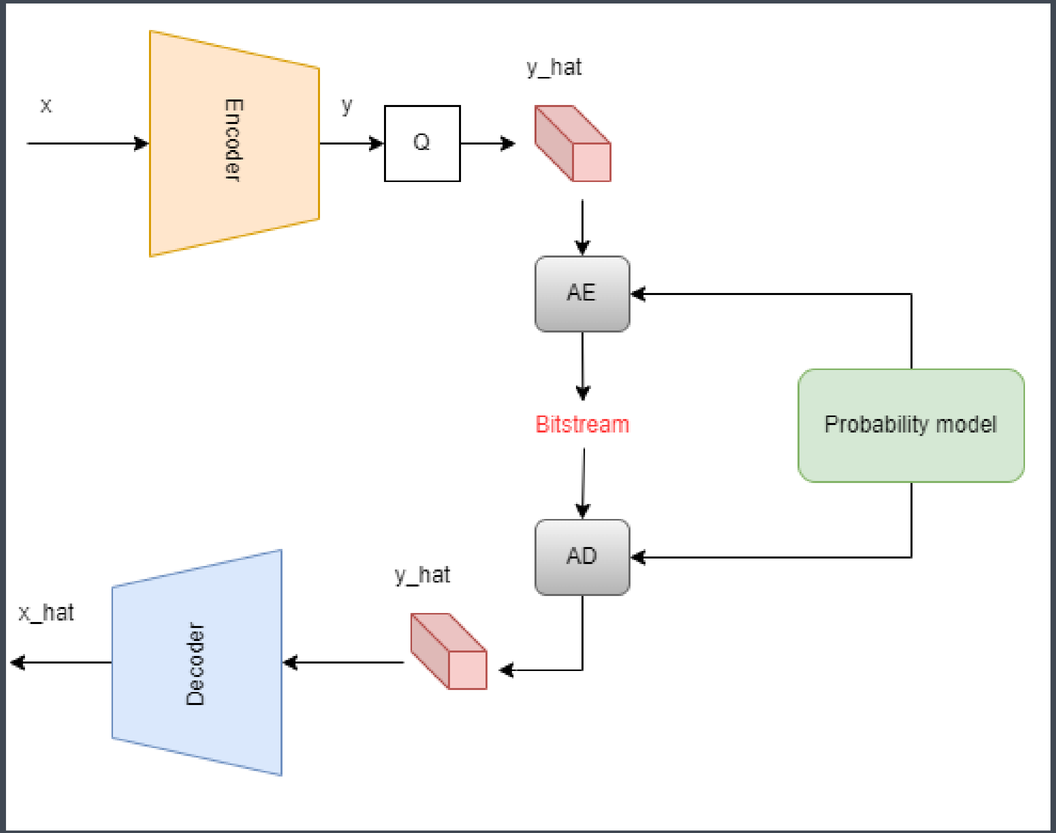
INSA
INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

OUTLINE

1. End-to-end Learned image codecs
2. Patch-based End-to-end image learned codecs using Overlapping
3. Results
4. Conclusion

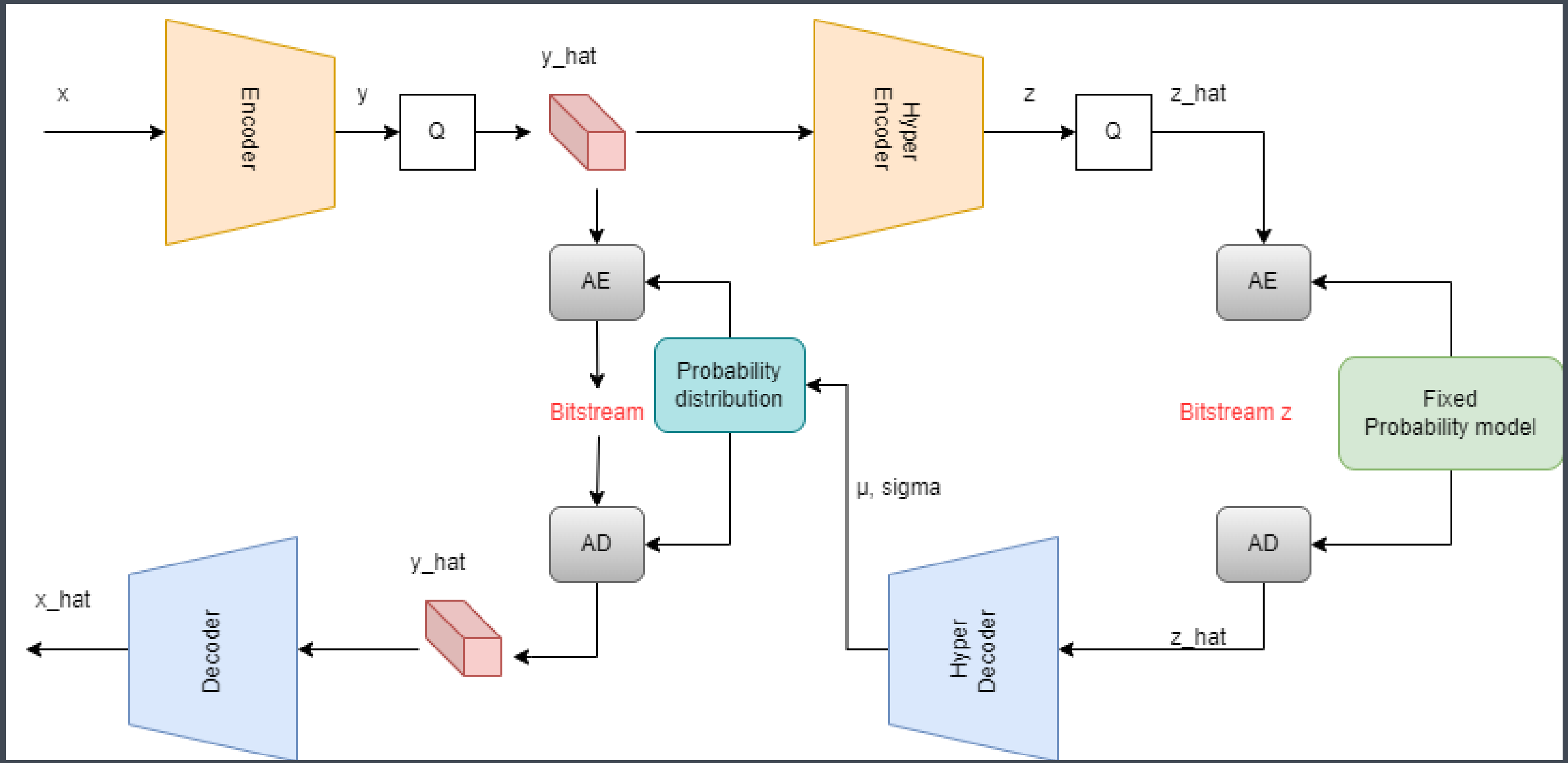
END-TO-END IMAGE LEARNED CODECS

- End-to-end image learned codecs based on auto-encoder architecture
- Encoder : Transform the input image x to a latent representation y
- Probability model : Estimation the probability distribution of the latent representation
- Decoder : Recontruct the input image from the latent representation



ATEME

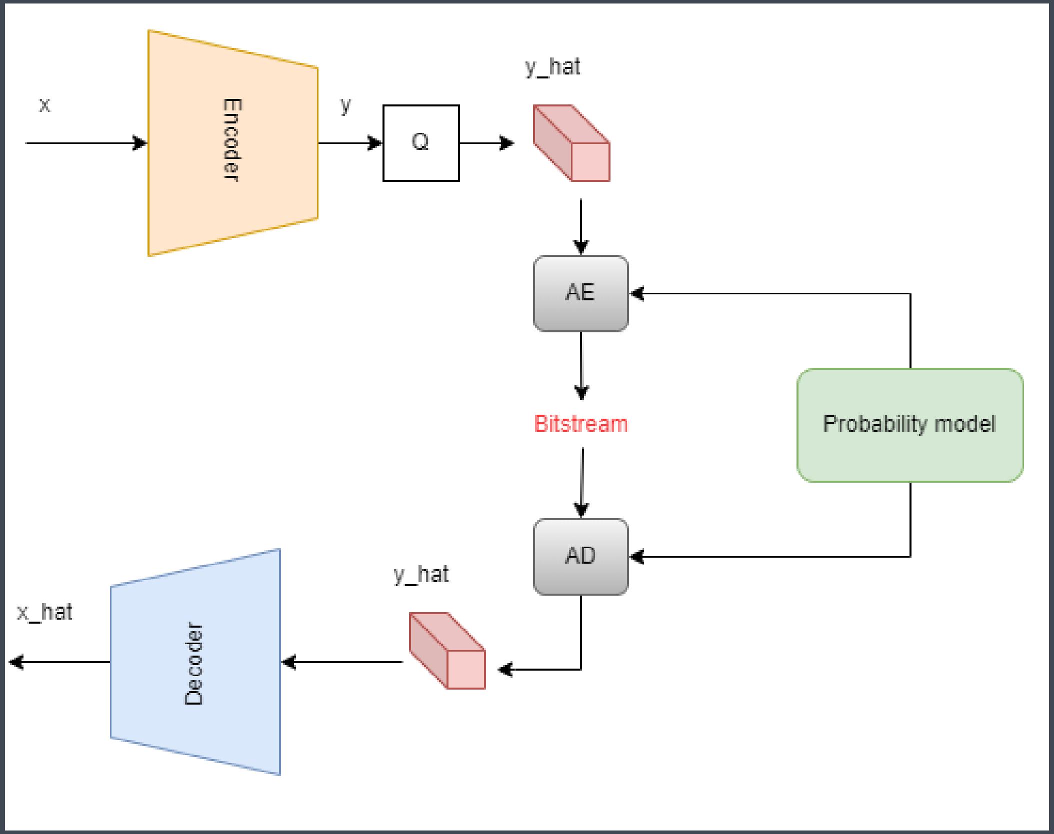
END-TO-END IMAGE LEARNED CODECS



[1] J. Balle, D. Minnen, S. Singh, and N.Johnston S.J Hwan,
“Variational image compression with a scale hyper-prior,” ICLR 2018 - Conference Track Proceedings, 2018.

ATEME

END-TO-END IMAGE LEARNED CODECS : LIMITATIONS

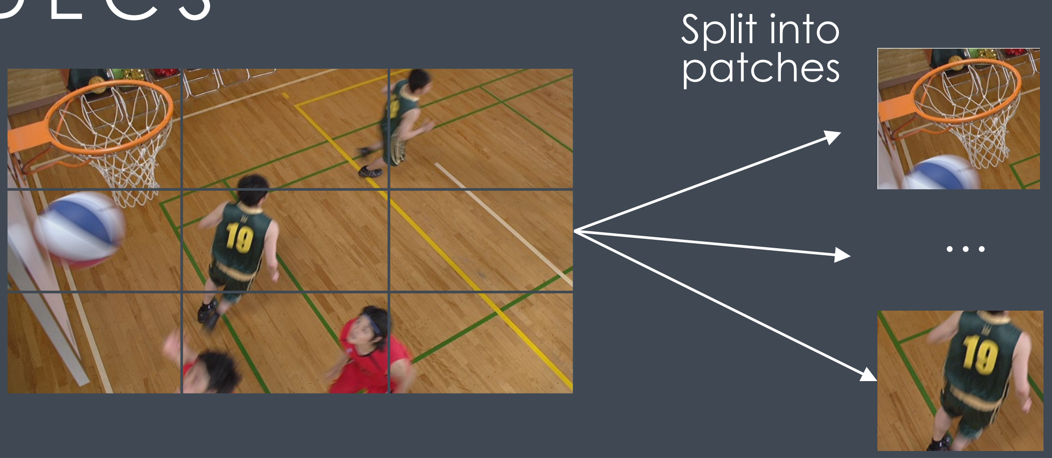


Architecture allows multi-resolution image coding.

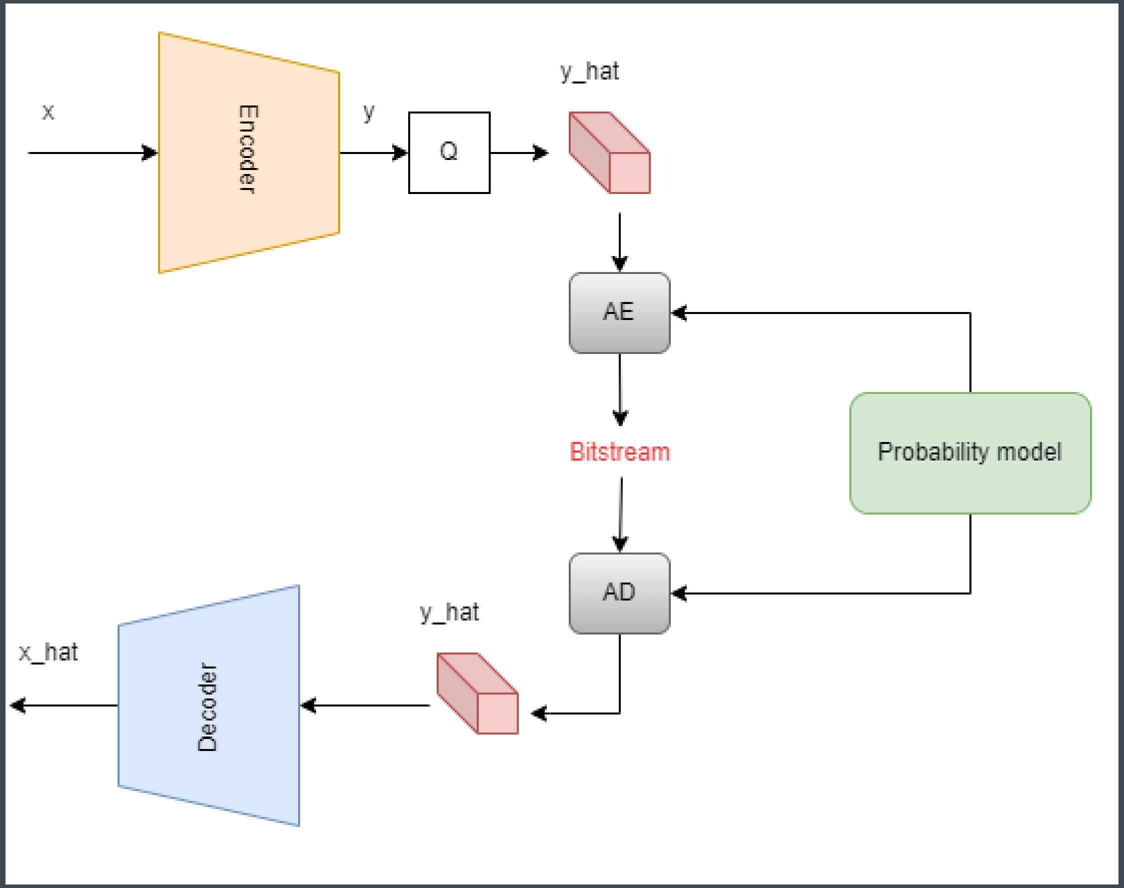
Hardware limitation (memory saturation) for big resolutions

Ex : OOM error for coding HD Image on GPU 2080ti with 11Go

PATCH-BASED END-TO-END IMAGE LEARNED CODECS



Border artifacts



PATCH-BASED END-TO-END IMAGE LEARNED CODECS USING OVERLAPPING

Goal

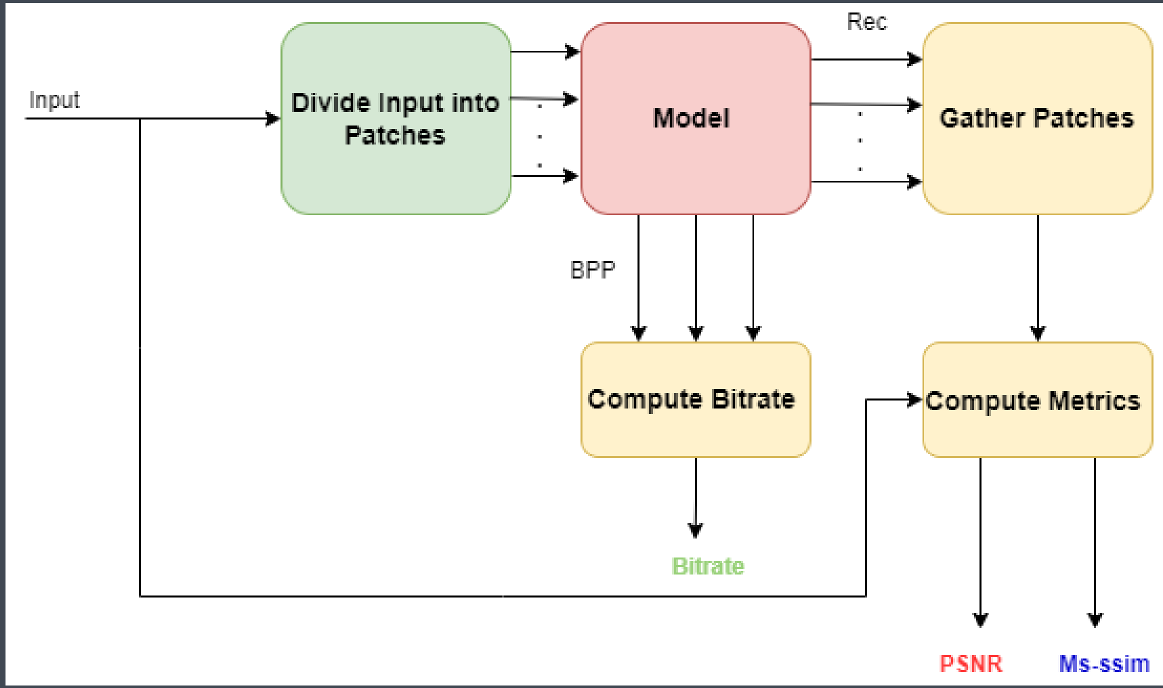
- Benefit from the advantages of patch-based solutions to address the hardware limitation
- Eliminate Border artifacts from the decoded images

Contribution

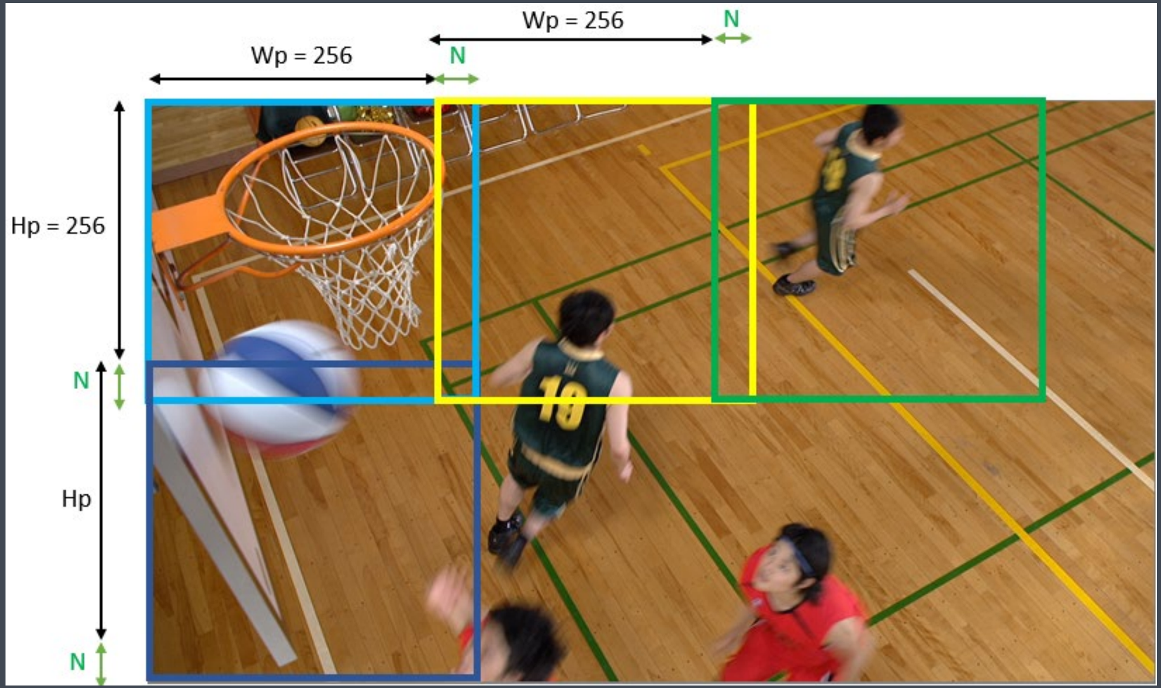
- Patch-based end-to-end learned image codec using overlapping method
- The proposed method is compatible with any learned codec based on an auto-encoder architecture

ATEME

PATCH-BASED END-TO-END IMAGE LEARNED CODECS USING OVERLAPPING



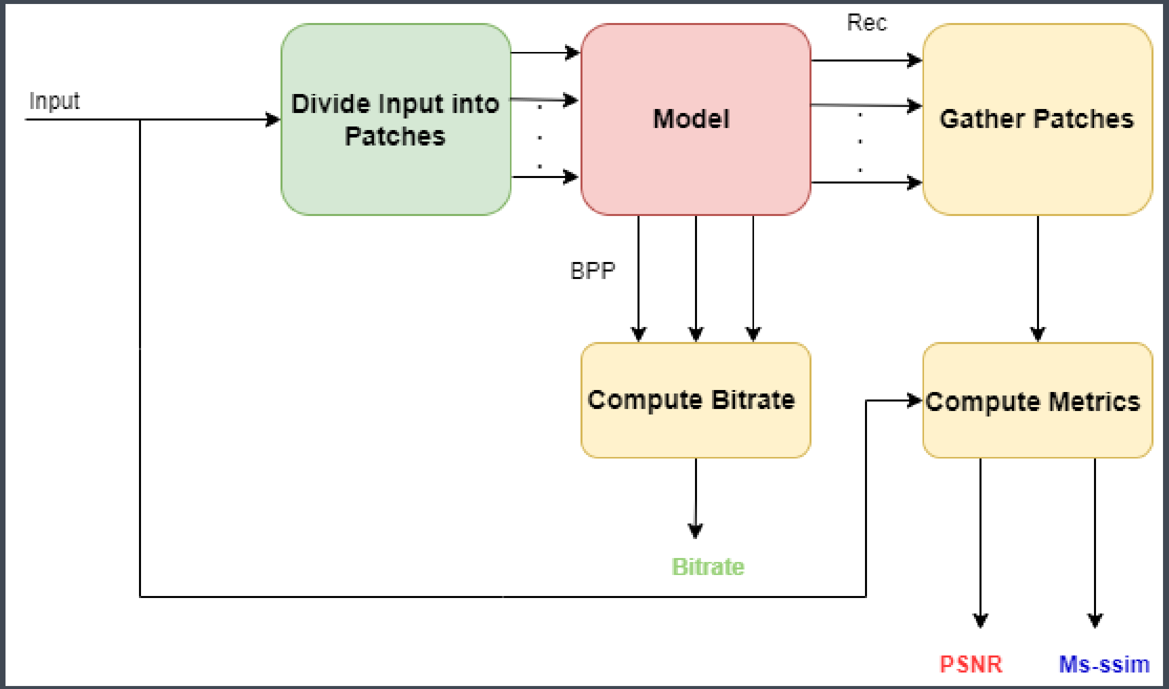
Steps to patch-based image coding



Divide into patches

ATEME

PATCH-BASED END-TO-END IMAGE LEARNED CODECS USING OVERLAPPING



Steps to patch-based image coding

- End-to-end model is an implementation of cheng 2020 [2].
- Training process :
  - Dataset : Clic 2020
  - Training resolution : 256x256
  - Total number of steps 500 000
- Loss function :

$$J = D + \lambda R$$

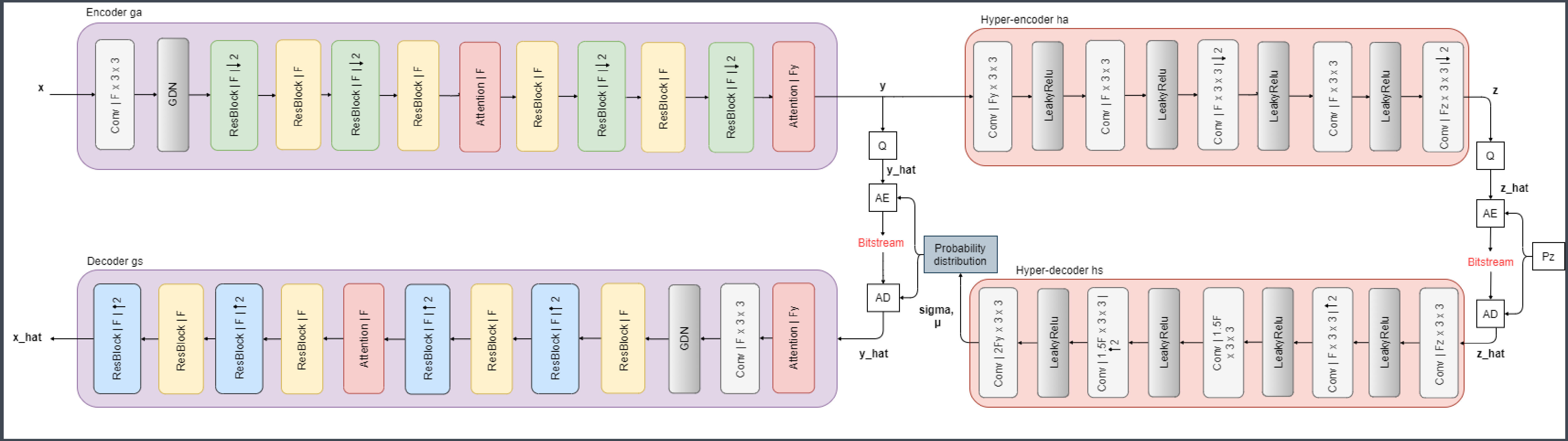
  - D : distorsion measured By MSE or MS-SSIM
  - R : rate used to transmit the bitstream, estimated using the shannon entropy
  - λ Lagrangian multipliers

Model

[2] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Learned image compression with discretized gaussian mixture likelihoods and attention modules," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7936–7945, 2020.

ATEME

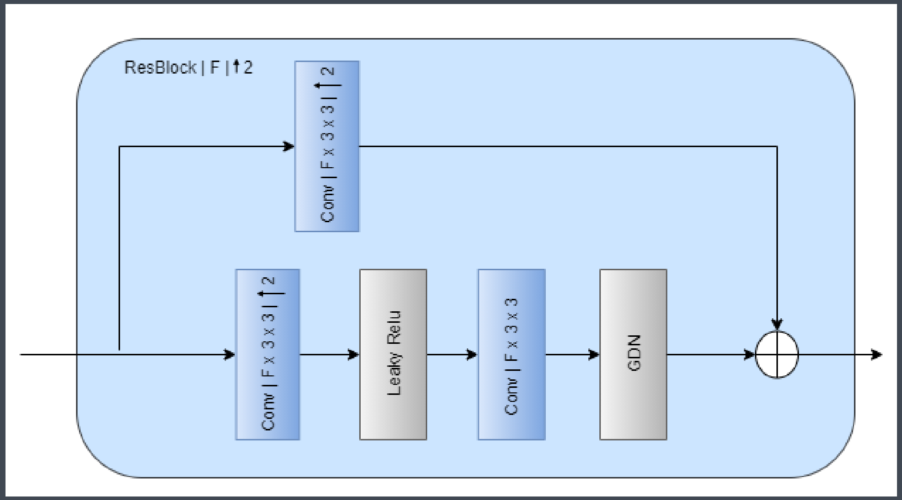
END-TO-END MODEL ARCHITECTURE

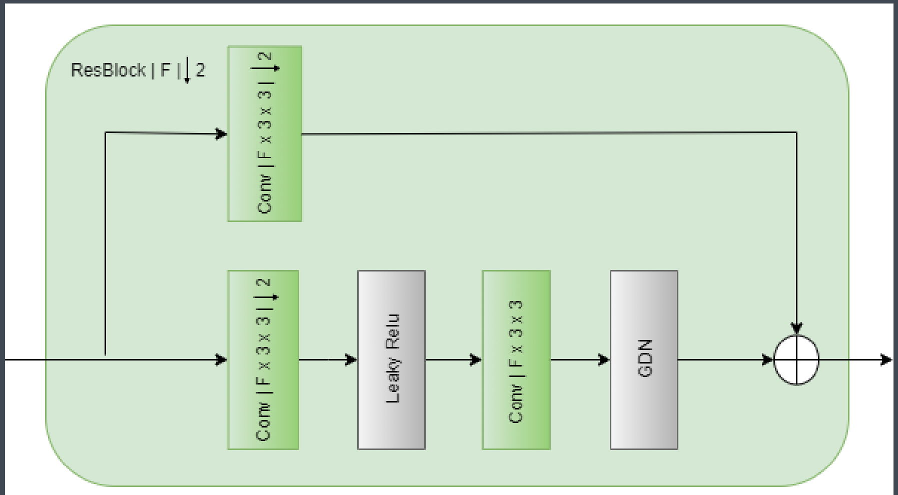


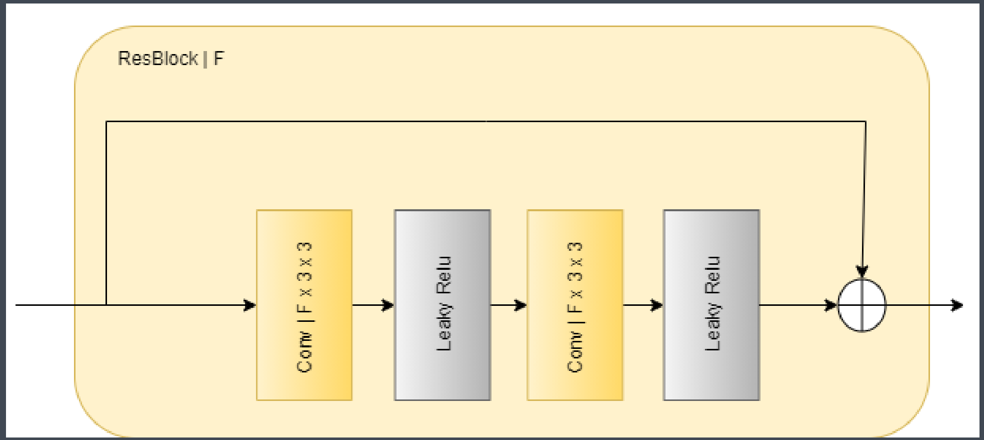
[2] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with discretized gaussian mixture likelihoods and attention modules,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7936–7945, 2020.

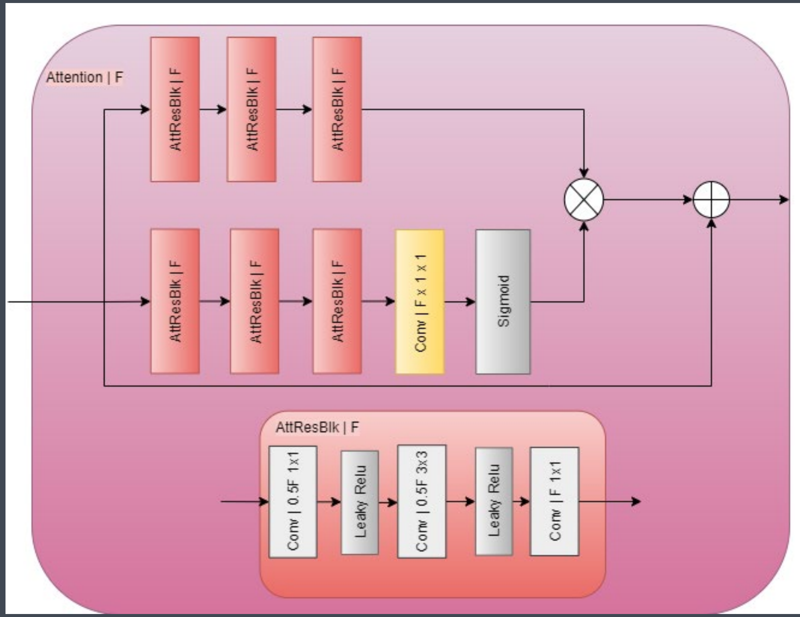
ATEME

END-TO-END MODEL ARCHITECTURE

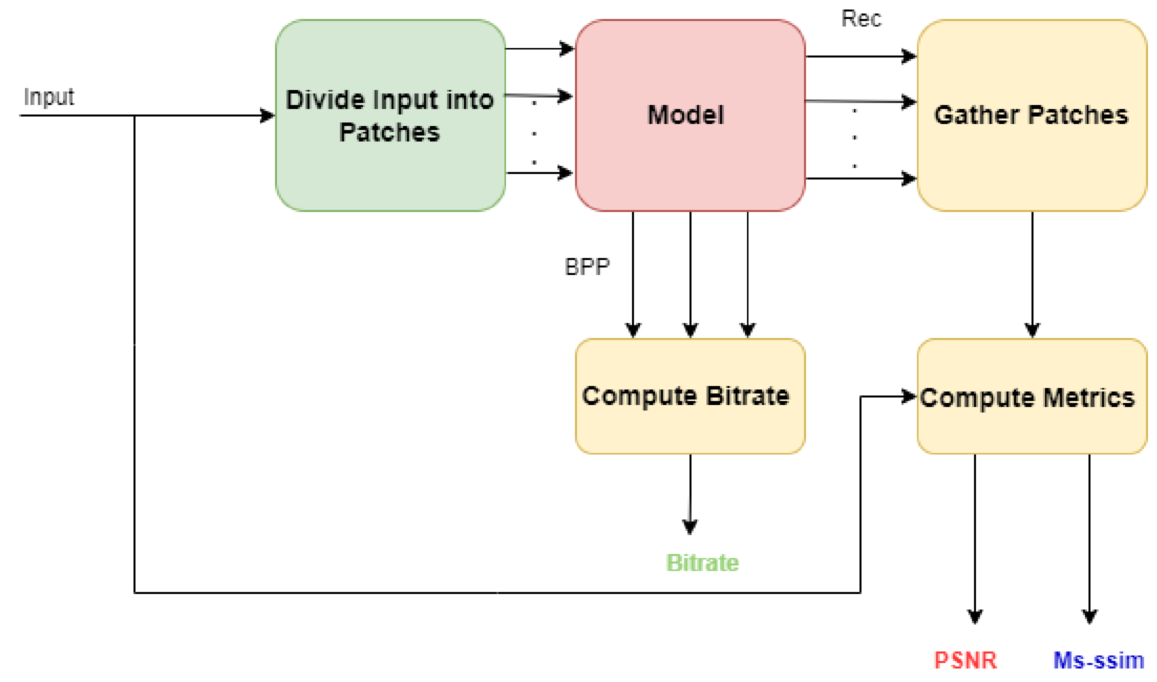




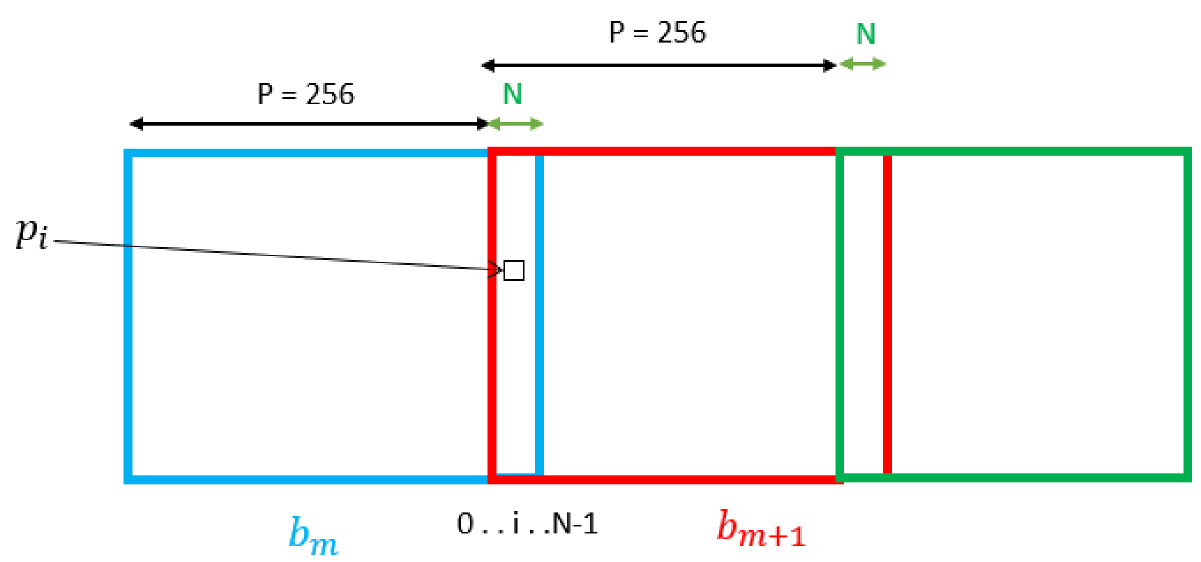




PATCH-BASED END-TO-END IMAGE LEARNED CODECS USING OVERLAPPING



Steps to patch-based image coding



$$p_i = (1 - \frac{i}{N-1})p_{b_m}(P+i) + (\frac{i}{N-1})p_{b_{m+1}}(i)$$

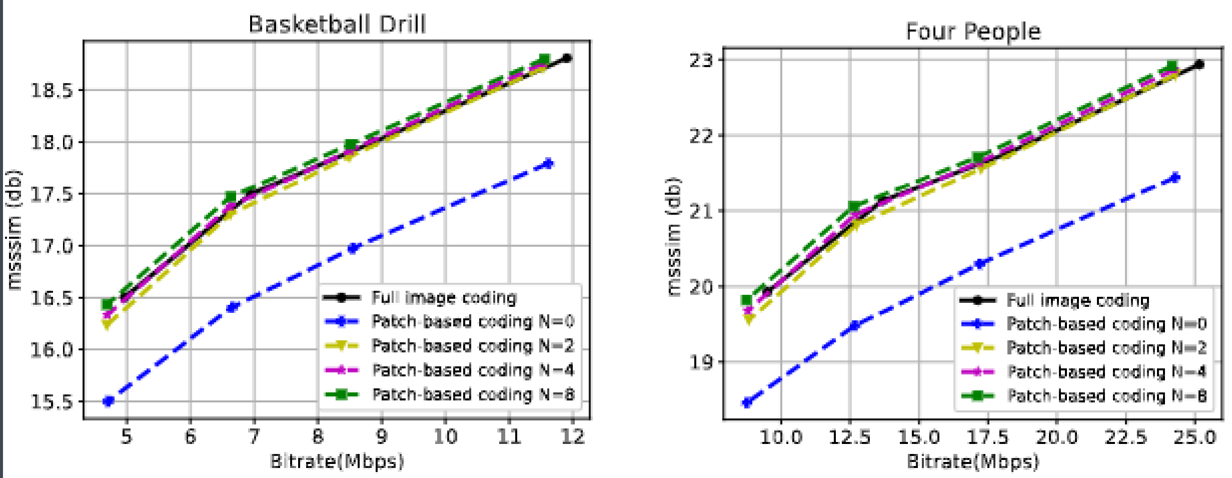
Gathering patches

ATEME

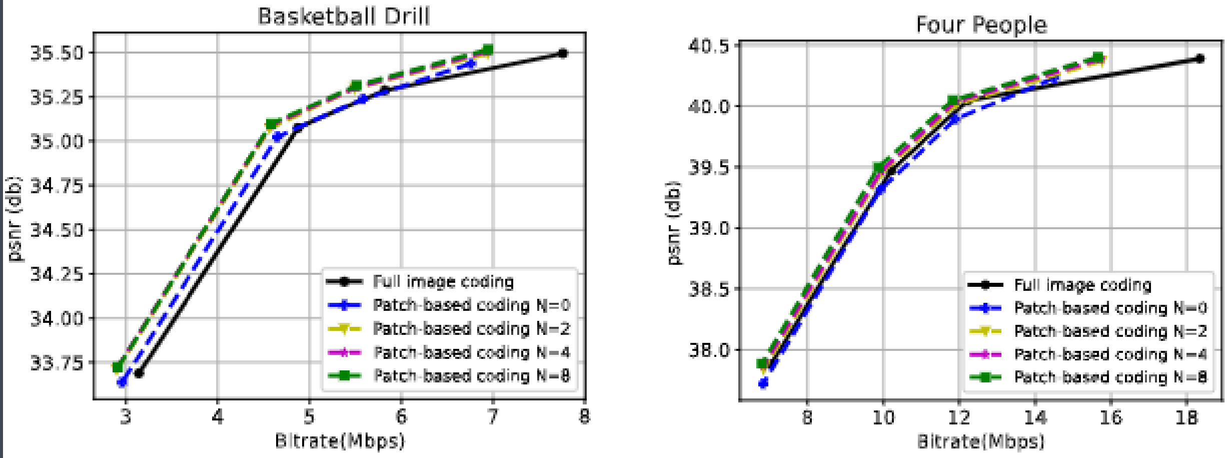
EVALUATION PROCESS

- Models : Eight models were trained. 4 for each quality metric (MSE and MS-SSIM) .
  - Metric MSE : $\lambda = \{4096, 3140, 2048, 1024\}$
  - Metric MS-SSIM : $\lambda = \{420, 220, 120, 64\}$
- Evaluation Sequences : Frame extracted from JVET Common Test Condition (CTC).
- Frame is compressed in :
  - Full resolution
  - Per patch without overlapping
  - Per patch with overlapping $N \in \{2, 4, 8, 16, 32\}$

RESULTS



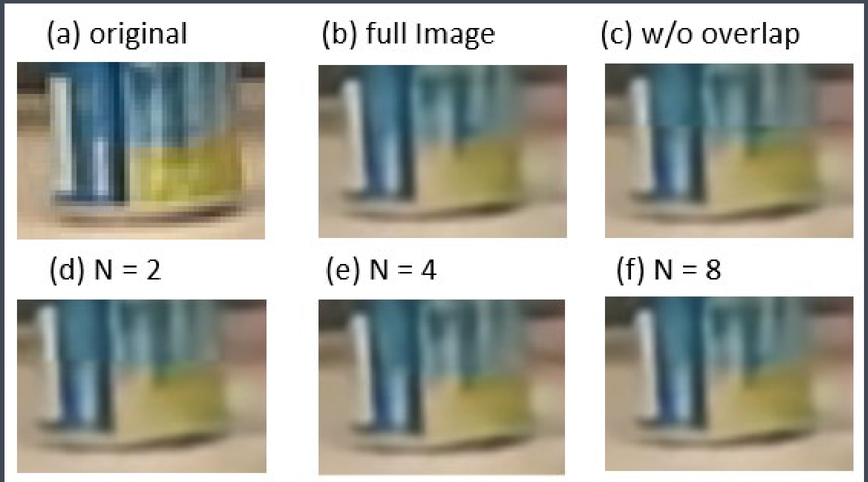
(a) MS-SSIM Rate-Distortion results



(b) PSNR Rate-Distortion results

Rate-distorsion Results

Visual results for MS-SSIM models



Visual results for MSE models

ATEME

RESULTS

BD-rate (MS-SSIM) gains of patch-based coding schemes compared to full resolution coding system for CTC sequences.

	Patch w/o Overlapping	Patch with Overlapping				
		N = 2	N = 4	N = 8	N = 16	N = 32
Class B	0.649	0.029	-0.009	-0.046	-0.066	-0.066
Class C	0.536	0.030	0.0002	-0.029	-0.044	-0.041
Class D	0.231	0.022	0.008	-0.009	-0.016	-0.013
Class E	0.958	0.057	0.016	-0.021	-0.045	-0.047
Class F	0.685	0.045	0.010	-0.018	-0.033	-0.032

BD-rate (PSNR) gains of patch-based coding schemes compared to full resolution coding system for CTC sequences.

	Patch w/o Overlapping	Patch with Overlapping				
		N = 2	N = 4	N = 8	N = 16	N = 32
Class B	-0,006	-0.061	-0.067	-0.072	-0.080	-0.074
Class C	0,0065	-0.030	-0.035	-0.038	-0.040	-0.035
Class D	0,0042	-0.010	-0.011	-0.014	-0.014	-0.007
Class E	0.035	-0.002	-0.014	-0.027	-0.050	-0.036
Class F	0.030	-0.022	-0.025	-0.027	-0.030	-0.023

BD-rate Results

CONCLUSION

- Hardware problem addressed.
- Block artifacts removed.
- Slight gains are observed comparing to Full resolution coding
- Other applications of this method such as denoising.

REFERENCES

- [1] J. Balle, D. Minnen, S. Singh, and N.Johnston S.J Hwan, “Variational image compression with a scale hyper-prior,” ICLR 2018 - Conference Track Proceedings, 2018.

- [2] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with discretized gaussian mixture likelihoods and attention modules,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7936–7945, 2020.

- [3] “Workshop and c. on learned image compression,” Https://www.compression.cc/, 2020.

ATEME

THANK YOU FOR YOUR ATTENTION

RESULTS

Resolution	Method	Coding Time GPU 2080 11Go	Coding time GPU 3090 24Go	Total Number of patches	Batch size	N
HD	Full Resolution	OOM	OOM	-	-	-
	Patch in parallel w/o overlaping	3.40s	1.967s	40	8	-
	Patch in parallel with overlaping	3.82s	2.05s		8	16
	Patch sequentially with overlaping	6.15s	2.861s		-	16
1280x720	Full Resolution	OOM	0.93s	-	-	-
	Patch in parallel w/o overlaping	1.75s	0.95s	15	5	-
	Patch in parallel with overlaping	1.91s	1.012s		5	16
	Patch sequentially with overlaping	2.73s	1.25s		-	16
832x480	Full Resolution	1.06s	0.52	-	-	-
	Patch in parallel w/o overlaping	1.05s	0.54s	8	8	-
	Patch in parallel with overlaping	1.109s	0.55s		8	16
	Patch sequentially with overlaping	1.586s	0.70s		-	16

Coding time Results