Noise-Aware Encoder Training and Specialized Denoiser for Efficient JPEG AI Compression and Denoising

Journal of Physics: Conference Series **3128** (2025) 012021 doi:10.1088/1742-6596/3128/1/012021

Noise-Aware Encoder Training and Specialized Denoiser for Efficient JPEG AI Compression and Denoising

Abdellah El Mennaoui[1,2,3], Jean-Luc Dugelay[2] and Ghalia Hemrit[1]

[1]Huawei Technologies, Mougins, France
[2]EURECOM, Biot, France
[3]Sorbonne University, Paris, France

E-mail: abdellah.el.mennaoui1@huawei.com

Abstract. Joint image compression and denoising present a significant challenge for the emerging JPEG AI standard, which requires a single learned encoder to generate a compressed bitstream compatible with multiple decoders. In this work, we demonstrate an effective and adaptable approach to enhancing framework performance in the noise domain for different noise levels by employing distinct training strategies for the encoder and decoder branches of the multi-task system. We test our approach on real noisy data from the Smartphone Image Denoising Dataset (SIDD) across six rate–distortion trade-offs and two sensor gain ranges, representing low and high noise levels. Our approach reduces bitrate allocation in the noise domain by up to 30% at an equivalent peak signal-to-noise ratio (PSNR) while enabling accurate restoration of the original noise—all within a single decoding pass.

1 Introduction

In real-world imaging pipelines, denoising (D) and compression (C) are almost always applied in sequence—either D→C or C→D —yet each ordering incurs its own penalty. In the D→C pipeline, denoising removes sensor noise but also attenuates high-frequency details (e.g., fine textures and edges) that modern codecs leverage for efficient transform coding[1], resulting in either oversmoothed reconstructions or a compensatory bitrate increase. Conversely, the C→D order forces the compressor to allocate bits to irrelevant noise, which the subsequent denoiser then discards[15]; moreover, residual compression artifacts can amplify perceptual noise, yielding unnaturally speckled outputs. To reconcile these conflicting objectives, the forthcoming JPEG AI standard[16] adopts a unified, learned bit-stream: a single encoder produces a universal latent representation, and multiple task-specific decoders (e.g., for denoising, color enhancement, or high-level vision tasks like detection and segmentation) each extract exactly the information they require without re-encoding. Recent work has begun to validate this paradigm in the simplest two-task setting—joint compression and denoising [3, 14, 1]. This unified encoder–multiple decoder strategy eliminates the inefficient distribution of bits characteristic of serial C→D schemes by encoding all content in one latent space and letting each decoder pull only the information it needs. Moreover, JPEG AI further enables an adjustable trade-off between reconstruction fidelity and denoising capability, allowing practitioners to balance image detail preservation against noise suppression within a single learned framework, while still enabling access to the original noise residual for photography use cases [10, 11].

In conventional joint compression–denoising frameworks, the shared encoder is typically trained solely on clean images[1, 3, 5]. Consequently, when encoding real sensor noise, the encoder either discards valuable noise information or encodes it inefficiently, wasting bitrate on irrelevant artifacts. In contrast, JPEG AI advocates a unified encoder that preserves all image information, including noise, enabling each decoder to selectively extract the necessary details. To achieve this, we propose a two-stage, noise-aware

training approach. First, we fine-tune the shared encoder on real noisy images captured at different ISO levels, embedding realistic noise statistics directly into the encoder's latent representations to optimize bitrate allocation. Second, we specialize the denoising decoder for specific ISO ranges. Our implementation combines SegPIC [13], a segmentation-guided compressor, and CGNet [14], a lightweight global-context denoiser, into a unified forward pass producing both noisy and clean outputs. We refer to this architecture as SeggazNet.

To isolate the impact of the encoder's training domain, we compare the vanilla encoder, pretrained solely on noise-free images, with an identical encoder fine-tuned on real noisy captures from the SIDD benchmark. We then specialize the denoising decoder by fine-tuning its branch on latent codes from both low- and high-ISO noise distributions, ensuring it adapts precisely to different sensor noise levels. Across both scenarios, this training strategy (1) reallocates bit budget away from pure noise—saving up to 30 % of bits at equal PSNR; and (2) enriches the latent representation with denoising features—gaining up to 2.34 dB PSNR at matched bitrates, while still supporting on-demand noise recovery in a single forward pass. Our experiments confirm that fine-tuning the encoder on noisy domain data significantly enhances both rate–distortion performance and denoising fidelity within the JPEG AI framework.

Importantly, these improvements require no modifications to the SeggazNet architecture, demonstrating that simple encoder fine-tuning alone can enhance multi-task coding gains within the JPEG AI framework.

Our key contribution is a noise-domain encoder fine-tuning strategy that injects real sensor noise statistics into the shared backbone, paired with a dual-intensity denoising decoder trained to handle low- and high-level noise artifacts in latent space.

In what follows, we first survey prior work on joint image compression and denoising, parameter-sharing strategies in multi-task architectures, and latent-space alignment techniques. In Section 3, we introduce the end-to-end framework that we define for this study, detailing the segmentation-guided compression backbone, the lightweight neural adapter, and the global-context denoiser. In Section 4, we describes the training and evaluation protocol, including noisy-domain encoder fine-tuning and the associated loss formulations. Section 5 presents quantitative rate–distortion analyses and qualitative visual comparisons across low and high levels of noise.

2 Related Work

2.1 Lossy Image Compression

Lossy image compression is a well-studied problem. Many hand-crafted codecs like JPEG[6], JPEG2000[7] and VVC intra[8] have been proposed and widely adopted in the industry. In recent years, learning-based lossy image compression methods have attracted more interest as they generally provide a significant rate-distortion performance boost compared to JPEG/JPEG2000. Among these architectures, a specific type of generative model called Variational Autoencoder has achieved promising results [9] due to its deep connection with rate-distortion theory, i.e., the theoretical foundation of lossy image compression.

2.2 Joint Compression–Denoising

Methods that integrate denoising into the compression pipeline by sharing a single latent representation have shown dramatic bitrate savings and flexible noise handling. Alvar *et al.* [1] introduced a scalable latent-space partitioning approach, Joint Image Compression and Denoising (JICD), that splits the encoder's output into a "base" layer for the clean image and an "enhancement" layer for residual noise. When only a denoised output is needed, the enhancement layer is omitted, yielding up to 80 % bitrate savings; when the original noisy image is required, both layers are decoded. Cheng et al. [3] propose a noise-aware compressor with parallel clean-guidance branches, encouraging the shared encoder to suppress noise pre-quantization. Building on these, recent works leverage contrastive objectives for orthogonal signal–noise disentanglement [5] and multi-branch teacher-student guidance[20]. Departing from prior methods, our approach fine-tunes only the encoder on real noisy data to enrich the latent space with denoising features, while preserving compression performance and maintaining the original architecture. In parallel, we specialize the denoiser for different noise levels to enhance adaptation to real-world sensor characteristics.

In contrast to prior methods that modify the architecture [1, 3, 5], we preserve the original design and achieve strong performance through encoder fine-tuning and decoder specialization. Building on our earlier work in mobile image compression [17, 19], we address, in this work, compression and denoising under real sensor noise within a unified framework.

2.3 Parameter Sharing

The choice of how to share parameters across tasks critically affects multi-task performance[2, 25, 18]. Hard sharing enforces a single encoder for all tasks with task-specific decoders, as in the JPEG AI vision

[18]. Soft sharing augments a shared backbone with lightweight modules (e.g. cross-stitch units [21], adapter layers [22]) to balance shared vs. private capacity. We follow the same hard-and-soft sharing principles, but demonstrate that encoder pre-training on noisy data alone can unlock both improved compression efficiency and richer denoising features in a single shared backbone.

3 SeggazNet Architecture

We explore how an encoder's training, whether pretrained on clean images or fine-tuned in the noise domain—affects the joint compression and denoising performance of a single JPEG AI bitstream. Beyond fine-tuning on real noisy data, we introduce a specialized decoder trained to target either low-ISO or high-ISO latent patterns. To demonstrate this, we present SeggazNet, an end-to-end architecture that, from one compressed bitstream, simultaneously reconstructs a noise-retained image and produces a fully denoised output (see Figure 1).

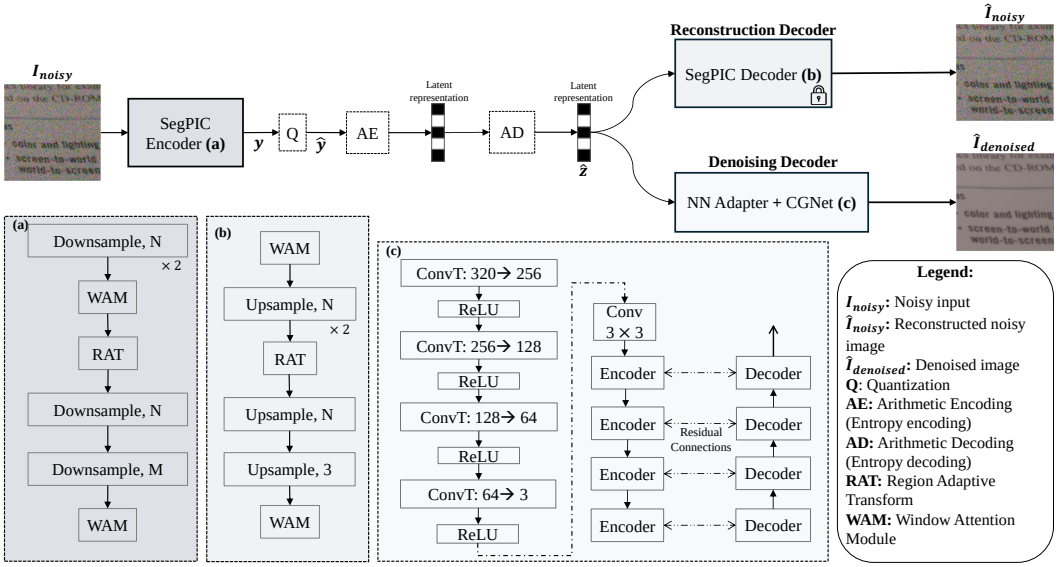


Figure 1: Overview of SeggazNet, showing the encoder SegPICEnc, quantizer Q, arithmetic coder (AC) and decoder (AD), and SegPIC decoder + neural adapter + CGNet denoiser. The shared SegPIC encoder **(a)** ingests the noisy input and produces a 320-channel latent vector code via segmentation-guided transforms and a hyperprior. That same latent is quantized (Q), entropy-coded (AC) to a bitstream and decoded (AD) back to $\hat{y}$, then decoded by SegPIC **(b)** to reconstruct the noisy image. In parallel, the latent vector passes through a lightweight neural adapter to reshape it for CGNet's global-context denoiser **(c)**, yielding the clean output. Both noisy and denoised outputs emerge from one bitstream in a single forward pass.

3.1 Compression Branch (SegPIC Encoder):

In our framework, the noisy input image $I_{\text{noisy}} \in \mathbb{R}^{H\times W\times 3}$ is first analyzed by the SegPIC encoder SegPICEnc to produce a continuous feature tensor $y = \text{SegPICEnc}(I_{\text{noisy}}) \in \mathbb{R}^{\frac{H}{16}\times\frac{W}{16}\times 320}$. This tensor is quantized via $\hat{y} = Q(y) \in \mathbb{Z}^{\frac{H}{16}\times\frac{W}{16}\times 320}$, entropy-encoded into a bitstream by $\text{AC}(\hat{y}|p_{\hat{y}}) := bit_{\hat{y}}$, and reconstructed by the arithmetic decoder $\hat{z} := \text{AD}(\text{bit}_{\hat{y}}|p_{\hat{y}})$. The single recovered latent $\hat{z}$ now drives both output branches in one forward pass: the SegPIC decoder SegPICDec reconstructs the noisy image as $\hat{I}_{\text{noisy}} = \text{SegPICDec}(\hat{z})$, while in parallel a lightweight neural adapter upsamples and reshapes $\hat{y}$ for the CGNet denoiser, yielding the clean result $\hat{I}_{\text{denoised}}$. By quantizing and coding only once, we avoid redundant transforms and jointly optimize rate–distortion performance for both noise-preserved reconstruction and denoising.

3.2 Reconstruction Branch (SegPIC Decoder):

The compressed latent $\hat{z}$ is fed into the SegPIC decoder, which applies learned transpose-convolutions and upsampling enhanced by a Window Attention Module (WAM) and a Region-Adaptive Transform (RAT)

to produce the noisy output $\hat{I}_{\text{noisy}}$. This branch preserves sensor noise in the reconstruction, enabling on-demand noise retention without additional passes through the network.

3.3 Denoising Branch (Neural Adapter & CGNet):
The same latent $\hat{\mathbf{z}}$ is routed through a neural adapter that first upsamples the spatial resolution and reduces the channel dimension from 320 to 3 via four transposed-convolution layers (doubling the height and width at each step).Only the adapter transforms the latent vector and produces the structural cues that feed into the CGNet denoiser. CGNet's encoder extracts multi-scale features, its Global Context Extractor (GCE) modules aggregate long-range dependencies through cascaded small-kernel convolutions, and its channel attention blocks. Finally, the CGNet decoder uses convolutional upsampling and skip connections to reconstruct the clean image $\hat{I}_{\text{denoised}} = \text{CGNet}\big(\text{Adapter}(\hat{z})\big)$.

4 Training and Inference

4.1 Dataset
We use in this work the Smartphone Image Denoising Dataset (SIDD) [15], a high-quality benchmark designed to reflect real-world smartphone image noise. SIDD contains approximately 30000 noisy images captured across 10 scenes using five smartphone cameras under varying lighting conditions. Each noisy image is paired with a noise-free ground truth image, estimated via a multi-frame averaging and alignment pipeline tailored to compensate for optical stabilization, radial distortion, and clipping artifacts. This procedure ensures accurate noise-free even in low-light settings.

For our experiments, we select two ISO-conditioned subsets from the SIDD dataset: a low-ISO (50–800) set, consisting of 64192 cropped images to simulate low noise levels; and a high-ISO (6400–12800) set, consisting of 2112 cropped images to capture the severe real noise typical of night photography.

4.2 Training
The training process consists of two stages: encoder adaptation and denoising decoder training.

Stage 1: Encoder Domain Adaptation. We initialize the encoder weights from the original SegPIC encoder pretrained on the COCO-Stuff dataset [23] for 400 epochs (batch size 32, 32 workers), using an initial learning rate of 1×10^{-4} decayed via ReduceLROnPlateau (16-epoch patience) to a minimum of 5×10^{-6}. To achieve domain adaptation to real noisy data and various rate–distortion trade-offs, we fine-tune the initialized encoder weights on real noisy captures from the SIDD dataset for 100 epochs, using identical hyperparameters and varying the rate–distortion weight $\lambda \in \{0.0018, 0.0035, 0.0067, 0.0130, 0.0250, 0.0483\}$ to balance bitrate and reconstruction quality. The optimization minimizes the reconstruction loss defined as:

$$\mathcal{L}_{\text{recon}} = r + \lambda \big\| \hat{I}_{\text{noisy}} - I_{\text{noisy}} \big\|_2^2, \tag{1}$$

where r is the estimated bitrate (in bpp), $\hat{I}_{\text{noisy}}$ is the SegPIC-decoded noisy image, I_{noisy} is the original noisy input, and λ is the selected rate–distortion weight.

Stage 2: Task-specific decoder training. With the encoder fixed—using either the original pretrained weights or the fine-tuned noisy domain weights—we train the neural adapter and the CGNet denoiser on latent representations extracted from noisy inputs. To specialize the decoder for different ISO levels, we use separate subsets of the SIDD data corresponding to low-ISO and high-ISO. Both the neural adapter and the CGNet denoiser are trained for 100 epochs on 128×128 patches (batch size 8, initial learning rate 5×10^{-5} with ReduceLROnPlateau scheduling). We optimize the denoising decoder by minimizing the loss

$$\mathcal{L}_{\text{denoise}} = \big\| \hat{I}_{\text{denoised}} - I_{\text{clean}} \big\|_2^2, \tag{2}$$

where $\hat{I}_{\text{denoised}}$ is the CGNet output and I_{clean} is the noise-free ground truth.
Note that for the reconstruction branch, we did not specialize training on either low or high ISO; the reconstruction decoder remained frozen throughout.

4.3 Inference
To analyze the effect of noise-domain encoder training, we evaluate four SeggazNet variants (see Table 1), each defined by a specific encoder configuration—either pretrained on clean images or fine-tuned on noisy data—and tested under distinct noise regimes (low or high ISO). For each model, we exploit six values of the rate–distortion trade-off parameter λ, giving bitrates ranging from 0.05 to 3.2 bpp. Performance is assessed using PSNR and BD-rate on both ISO-specific subsets and the complete SIDD test set.

Table 1: SeggazNet Encoder+Decoder combinations

Model	Encoder	Reconstruction Decoder	Denoising Decoder
SeggazNet_pl	Pretrained weights	Pretrained weights	Fine-tuned on low-ISO
SeggazNet_ph	Pretrained weights	Pretrained weights	Fine-tuned on high-ISO
SeggazNet_nl	Fine-tuned on full SIDD	Pretrained weights	Fine-tuned on low-ISO
SeggazNet_nh	Fine-tuned on full SIDD	Pretrained weights	Fine-tuned on high-ISO

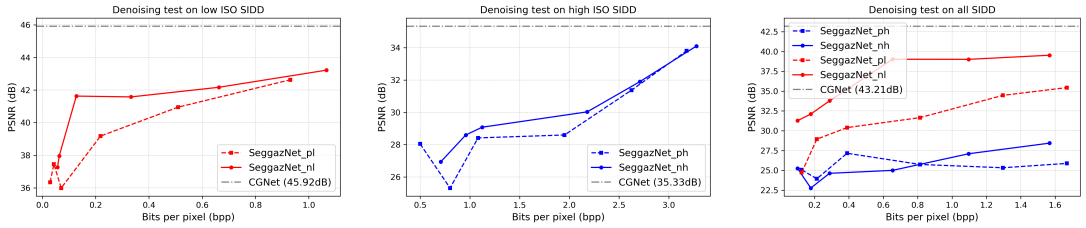
The reconstruction branch outputs from SeggazNet_pl and SeggazNet_ph are identical,the same applies to SeggazNet_nl versus SeggazNet_nh. Our method reuses the SegPIC encoder and fixed decoder with minimal overhead; the added CGNet branch runs at 52 ms with 62.1 G MACs per 256×256 image and <1 GB memory usage [14].

5 Results

In this section we quantify the rate–distortion performance of SeggazNet on SIDD and show visual reconstructions and denoising under varying noise levels.

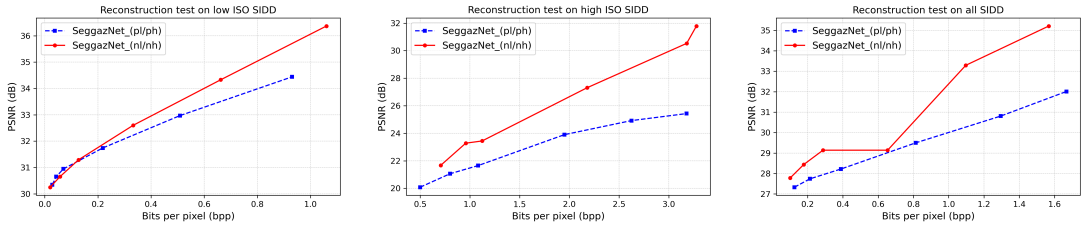
5.1 Quantitative Evaluation

Figure 2 shows PSNR versus bits-per-pixel (bpp) curves for the denoising branch on low-ISO, high-ISO, and the full SIDD dataset. SeggazNet_nl and SeggazNet_nh (solid) consistently outperform SeggazNet_pl and SeggazNet_ph (dashed), yielding BD-rate[24] savings of up to 30% (see Table 2) over the pretrained encoder SeggazNet_(pl/ph). As shown in the figure, CGNet applied to the uncompressed images (24.00 bpp) sits at the top of the PSNR curve—however, SeggazNet variants run just below it across all bitrates, illustrating their ability to perform denoising in tandem with compression from a single bitstream. As highlighted in Fig. 2(c), SeggazNet_nl achieves the highest PSNR across the entire bitrate spectrum, thanks to noisy-domain encoder fine-tuning and the high volume of low-ISO examples in SIDD. Meanwhile, the high-ISO variant SeggazNet_nh trained on fewer images examples—still exceeds its clean-only counterpart at higher bpp.



(a) Low-ISO (b) High-ISO (c) All-ISO

Figure 2: Rate–distortion (PSNR vs. bpp) for the denoising branch on SIDD under two noise levels.



(a) Low-ISO noisy reconstruction (b) High-ISO noisy reconstruction (c) All-ISO noisy reconstruction

Figure 3: Rate–distortion (PSNR vs. bpp) for the reconstruction branch under two noise levels.

Figure 3 plots PSNR versus bits-per-pixel (bpp) for the noisy-image reconstruction branch applied to low- and high-ISO images. In both ISO ranges, the encoder fine-tuned on noisy data, SeggazNet_nl and SeggazNet_nh (solid lines) consistently outperform their pretrained counterparts SeggazNet_pl and SeggazNet_ph (dashed lines), confirming that encoder adaptation to real noise enhances reconstruction fidelity. Overall, fine-tuning the encoder on noisy data significantly enhances noise robustness without compromising compression performance, fulfilling the dual-task objectives of JPEG AI.

Table 2: BD-rate reduction (%) and BD-PSNR gain (dB) when fine-tuning the encoder, for each ISO regime (Anchors: SeggazNet_(pl/ph) for reconstruction, SeggazNet_pl and SeggazNet_ph for denoising.)

Comparison	**Task**	**BD-Rate (%)**	**BD-PSNR (dB)**
SeggazNet_nl/SeggazNet_pl	Denoising (Low ISO)	–31.41	+2.19
SeggazNet_(nl/nh)/SeggazNet_(pl/ph)	Reconstruction (Low ISO)	–10.33	+0.20
SeggazNet_nh/SeggazNet_ph	Denoising (High ISO)	–0.53	+0.97
SeggazNet_(nl/nh)/SeggazNet_(pl/ph)	Reconstruction (High ISO)	–39.11	+2.34

5.2 Visual Results

Figure 4 presents qualitative comparisons on low- and high-ISO patches from the SIDD dataset. For each sample from the dataset, we visualize the noisy input, noise-free ground truth (GT), CGNet output, SeggazNet denoised and reconstructed outputs. Results are shown for both pretrained variants and their fine-tuned counterparts. As observed, encoder fine-tuning on noisy data significantly improves perceptual quality—preserving finer textures, and achieving sharper reconstructions. In the reconstruction branch, across the entire dataset SeggazNet_(nl/nh) variants yield BD-rate reductions of 10.33% at low ISO and 39.11% at high ISO (Table 2) while maintaining visually faithful outputs (see Figure 4).

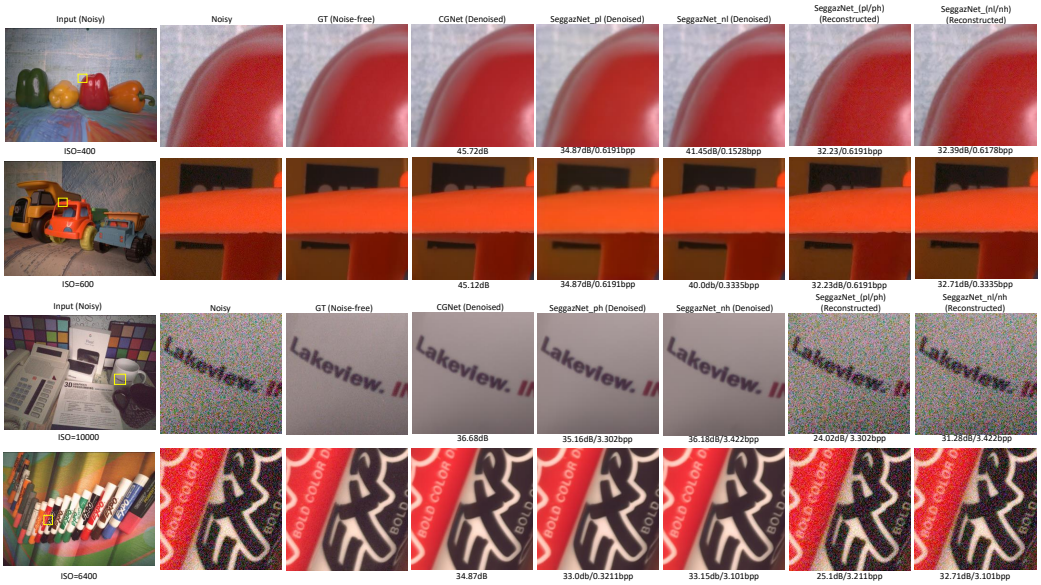


Figure 4: Visual results on the SIDD dataset. Top: low-ISO samples (ISO 400/600) comparing CGNet, SeggazNet_pl/nl. Bottom: high-ISO samples (ISO 6400/10000) comparing CGNet, SeggazNet_ph/nh. Each block shows: input, ground truth (GT), CGNet denoising, SeggazNet denoising, and noisy reconstruction by SeggazNet_p and SeggazNet_n . PSNR/bpp are shown under each image.

6 Conclusion

We have presented an effective training strategy on real noisy data that yields significant gains in a multi-task compression–denoising codec for JPEG AI. SeggazNet_n reduces bit allocation to pure noise by up to 30% and delivers up to 1.5 dB PSNR improvement at matched bpp. It also supports on-demand noise-preserved reconstructions in a single forward pass, all while lowering inference cost. A key advantage is that this noise-domain training operates independently of any architectural change, offering a lightweight strategy to boost both rate–distortion efficiency and denoising fidelity. Looking ahead, integrating learnable latent partitioning (e.g., dynamic gating or contrastive losses), perceptual/adversarial objectives, and extensions to video and other vision tasks promises to further close the gap toward the full JPEG AI vision of one bitstream powering every downstream applications.

References

[1] Ranjbar Alvar S, Ulhaq M, Choi H, Bajić IV 2022 Joint image compression and denoising via latent-space scalability *Front. Signal Process.* **2** https://doi.org/10.3389/frsip.2022.932873.

[2] Zhang Y, Yang Q 2022 A survey on multi-task learning *IEEE Trans. Knowl. Data Eng.* **34** 5586–5609 https://doi.org/10.1109/TKDE.2021.3070203.

[3] Cheng K L, Xie Y, Chen Q 2022 Optimizing image compression via joint learning with denoising *Proc. Eur. Conf. Comput. Vis. (ECCV)* (LNCS 13679) pp 56–73 https://doi.org/10.1007/978-3-031-19800-7_4.

[4] Lukas J, Fridrich J, Goljan M 2006 Digital camera identification from sensor pattern noise *IEEE Trans. Inf. Forensics Secur.* **1** 205–214 https://doi.org/10.1109/TIFS.2006.873602.

[5] Xie Y, Yu L, Pakdaman F, Gabbouj M 2024 Joint end-to-end image compression and denoising: leveraging contrastive learning and multi-scale Self-ONNs *Proc. IEEE Int. Conf. Image Process. (ICIP)* pp 3758–3764 https://doi.org/10.1109/ICIP51287.2024.10648189.

[6] Wallace G K 1992 The JPEG still picture compression standard *IEEE Trans. Consum. Electron.* **38** xviii–xxxiv https://doi.org/10.1109/30.125072.

[7] Skodras A, Christopoulos C, Ebrahimi T 2001 The JPEG 2000 still image compression standard *IEEE Signal Process. Mag.* **18** 36–58 https://doi.org/10.1109/79.952804.

[8] Pfaff J et al. 2021 Intra prediction and mode coding in VVC *IEEE Trans. Circuits Syst. Video Technol.* **31** 3834–3847 https://doi.org/10.1109/TCSVT.2021.3072430.

[9] Duan Z, Lu M, Ma Z, Zhu F 2023 Lossy image compression with quantized hierarchical VAEs *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)* pp 198–207 https://doi.org/10.1109/WACV56688.2023.00028.

[10] Tinio P P L, Leder H, Strasser M 2011 Image quality and the aesthetic judgment of photographs: Contrast, sharpness and grain teased apart and put together *Psychol. Aesthet. Creat. Arts* **5** 165–176 https://doi.org/10.1037/a0019542.

[11] Norkin A 2018 Film grain synthesis for video coding *Proc. Data Compression Conf. (DCC)* pp 425–434 https://doi.org/10.1109/DCC.2018.00008.

[12] Duan Z, Lu M, Ma J, Huang Y, Ma Z, Zhu F 2024 QARV: quantization-aware ResNet VAE for lossy image compression *IEEE Trans. Pattern Anal. Mach. Intell.* **46** 436–450 https://doi.org/10.1109/TPAMI.2023.3322904.

[13] Liu Y, Yang W, Bai H, Wei Y, Zhao Y 2025 SegPIC: region-adaptive transform with segmentation prior for image compression *Proc. Eur. Conf. Comput. Vis. (ECCV)* (LNCS 14203) pp 181–197 https://doi.org/10.1007/978-3-031-72952-2_11.

[14] Ghasemabadi A, Janjua M K, Salameh M, Zhou C, Sun F, Niu D 2024 CascadedGaze (CGNet): Efficiency in Global Context Extraction for Image Restoration *Trans. Mach. Learn. Res.* accepted https://openreview.net/forum?id=C3FXHxMVuq.

[15] Abdelhamed A, Lin S, Brown M S 2018 A high-quality denoising dataset for smartphone cameras (SIDD) *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* pp 1692–1700 https://doi.org/10.1109/CVPR.2018.00182.

[16] JPEG Committee 2025 *JPEG AI becomes an International Standard* (JPEG Press Release) https://jpeg.org/items/20250219_press.html.

[17] El Mennaoui A, Hemrit G, Dugelay J-L, 2025 Optimized image compression for mobile photography *Proc. Data Compression Conf. (DCC)* https://doi.org/10.1109/DCC62719.2025.00052.

[18] Ruder S 2017 An overview of multi-task learning in deep neural networks *arXiv* 1706.05098 https://doi.org/10.48550/arXiv.1706.05098.

[19] El Mennaoui A, Hemrit G, Dugelay J-L, 2025 Category-dependent learned image compression for smartphone photography with standard-compliant decoders *Proc. IEEE Int. Conf. Image Process. (ICIP)*. To appear.

[20] Li S, Azam M A, Gunalan A, Mattos L S 2022 One-step enhancer: deblurring and denoising of OCT images *Appl. Sci.* **12** 10092 https://doi.org/10.3390/app121910092.

[21] Misra I, Shrivastava A, Gupta A, Hebert M 2016 Cross-stitch networks for multi-task learning *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* pp 3994–4003 https://doi.org/10.1109/CVPR.2016.433.

[22] Rebuffi S A, Bilen H, Vedaldi A 2018 Efficient parametrization of multi-domain deep neural networks *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* pp 8119–8127 https://doi.org/10.1109/CVPR.2018.00847.

[23] Caesar H, Uijlings J, Ferrari V 2018 COCO-Stuff: thing and stuff classes in context *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* pp 1209–1218 https://doi.org/10.1109/CVPR.2018.00132.

[24] Bjøntegaard G 2001 Calculation of average PSNR differences between RD-curves ITU-T SG16/Q.6 Doc. VCEG-M33, Apr 2001.

[25] Caruana R 1997 Multitask learning *Mach. Learn.* **28** 41–75 https://doi.org/10.1023/A:1007379606734.