Flexible image decoding in learned image compression

Hossein Motamednia
Faculty of Computer Science and Engineering
Shahid Beheshti University
Tehran, Iran
h_motamednia@sbu.ac.ir
h.motamednia@ipm.ir

Azadeh Mansouri
Department of Electrical and Computer Engineering
Faculty of Engineering
Kharazmi University
Tehran, Iran
a_mansouri@khu.ac.ir

Fariba Saadati-Monem
Department of Electrical and Computer Engineering
Faculty of Engineering
Kharazmi University
Tehran, Iran
std_fariba.saadati@khu.ac.ir

Ahmad Mahmoudi-Aznaveh
Cyber Research Institute
Shahid Beheshti University
Tehran, Iran
a_mahmoudi@sbu.ac.ir

Abstract

Digital images in real-world applications mainly suffer from several quality degradations. Most learning-based codecs rely on a predefined compression process with perfect or high-quality images as the input. Nevertheless, compared to high-quality photos, images in the wild present very different features, making learning-based image coding sub-optimal. This paper offers a framework for compressing distorted images while estimating distortion-free ones. The reconstructed signals are represented using two categories of latent space features: the features that illustrate Visual Signal (VS) and the latent vector that represents Distortion Signal (DS). The distorted input can be reconstructed using both latent space features when merged employing different weight factors. In our method, various types of distortions have been explored. Based on our experiments, the presented method has competitive results with the state-of-the-art with augmented capability. The implementation of our method is available at https://github.com/MotamedNia/flexible_compression

1 Introduction

Image compression has become an important area of study in signal processing, enabling effective image transmission and storage. Learned compression techniques recently have shown a rapid development trend with promising results. Due to the rapid spread of mobile devices, visual data has recently been converted from professionally acquired photographs and videos to user-generated content (UGC) Li et al. [2021].

The acquired images inevitably include distortions, added after several processing stages. Most techniques treat denoising and compression separately, disregarding the shared concept Liu et al. [2020], Xing and Egiazarian [2021]. Additionally, most image compression models do not fully consider the application scenarios when assuming the input as pristine images Ballé et al. [2016, 2018], Minnen et al. [2018b], Cheng et al. [2020]. This leads to constraints on the applicability of these algorithms when employing UGC with varying quality as input images. Moreover, by using the noisy or corrupted input, the rate-distortion optimization may result in sub-optimal coding Pavez et al. [2022]. Many approaches suggest using an optimally denoised signal as both a source and a reference in the theoretically ideal User-Generated Content (UGC) compression system. However,

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

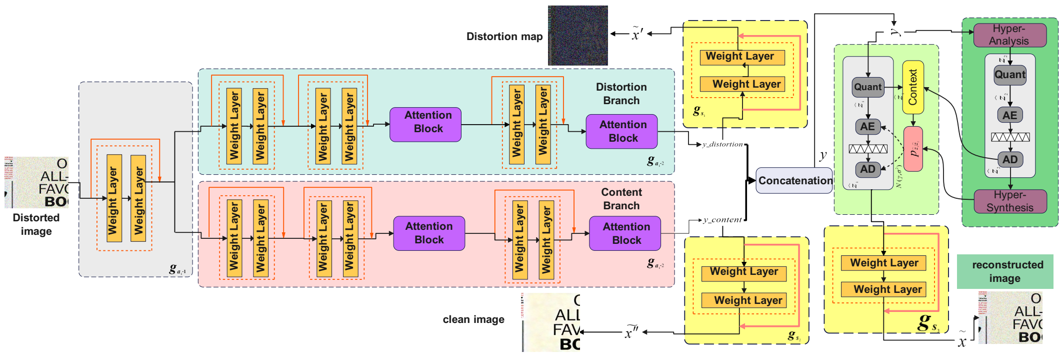

Figure 1: The proposed flexible decoding framework.

such a system is not applicable due to the lack of an ideal denoiser, and the removed noise during pre-processing cannot be restored, leading to a loss of the original noise's information. While most applications prefer noise-free stored images, noise can carry crucial information for applications like medical image analysis, multimedia forensics, and synthetic and artistic purposes. These types of disturbances should also be present in the compressed image Alvar et al. [2022].

Furthermore, a recent JPEG-AI call for proposals states that the coding system to be standardized will utilize an end-to-end learning-based architecture. In this scenario, compressed-domain representation and analysis can be employed in image processing and computer vision tasks cfp [2022]. The use of compressed-domain denoising, along with noise preservation methods, is recommended in Alvar et al. [2022]. A multi-task, scalable image compression system is proposed in which the base features and the enhancement features are utilized to reconstruct the noise-free and noisy input reconstruction respectively Alvar et al. [2022].

In this work, we present a learning-based image compression for distorted images. Our study exploits a wide range of distortion types and levels. To establish a set of acceptable distortion types and levels, we utilized a publicly available image quality database, tid2013 Ponomarenko et al. [2013] which includes 24 distinct types of corruption in 5 levels. we separate the Distortion Signal(DS) and the Visual Signal (VS) in the latent space. Here, the DS features reflect the imposed corruption and the VS stands for the pristine content. The overall structure of the proposed model is depicted in Figure 1. The proposed method's contribution is that it simultaneously enables the retrieval of an enhanced, noise-free image and the original noisy image. Additionally, it allows for adjusting the level of noise reduction during image decoding. The conducted experiments and analysis confirm the power of the presented latent space illustration.

2 Related works

DNN-based compression. Learning based image compression methods Minnen and Singh [2020], Lee et al. [2019], Minnen et al. [2018a], Klopp et al. [2018], Ballé et al. [2018], Minnen et al. [2018b], Ballé et al. [2016], Theis et al. [2017], Minnen et al. [2017] have surpassed traditional methods. These techniques mostly transform the input image into a latent representation using a convolutional auto-encoder. One of the primary works is Ballé et al. [2016]; they also have presented a framework to optimize their end-to-end rate-distortion performance. The later work Ballé et al. [2018], introduces a hyperprior network to extract the side information from the latent representation. In Minnen and Singh [2020], the authors improve performance over context-adaptive models by introducing channel conditioning and a latent residual prediction. In Cheng et al. [2020], the authors found that accurate entropy models for rate estimation significantly influence the performance of the method. They propose parameterizing the distributions of latent codes with discretized Gaussian mixture likelihoods. They incorporate attention modules as well.

Multi-Task compression. In compression, the primary objective is to reconstruct the input image. However, recent studies have demonstrated that in addition to the reconstruction, it is feasible to accomplish other tasks such as denoising, classification, and object detection. Testolina et al. [2021] introduces a method for learning-based denoising and compression. The method uses Poisson-Gaussian noise on the input training data and expands the decoder to have twice as many layers. It

employs a universal encoder and specialized decoders to delegate the denoising task to the decoder. This architecture does not reduce the bit rate because the task of eliminating noise is delegated to the decoder, whereas it is the encoder that can prevent incompressible noise from reaching the bitstream. Alvar et al. [2022] discusses the significance of preserving noise in applications like court evidence, image forensics, and artistic intent. It points out that once noise is removed in the pre-processing stage, it cannot be restored. In this case, the authors propose a learning-based image compression that combines compression and denoising. The network maps the input image to a scalable latent space divided into a base layer (containing clean image information) and an enhancement layer (carrying noise information). Depending on the need for a clean or noisy image, only the base layer or both layers are decoded. The framework includes two decoders for image construction.
Cheng et al. [2022] discusses how traditional compressions, with fixed parameters, can lead to artifacts. To solve this, the authors propose a noise-aware image compression algorithm that transforms the original noisy image into noise-free bits during compression. The method uses a two-branch, weight-sharing architecture. A guidance branch is pre-trained on clean images and then fine-tuned on noisy-clean image pairs, while the denoising branch takes noisy images and generates denoised features.
Li et al. [2021] investigates the peculiarities of the rate-distortion behavior when the input images are noisy. They provide a novel data-driven method for the noisy input without knowing the noise level beforehand. In Zhang et al. [2022], another noisy image compression framework is presented which operates under the assumption that a specific noise type and level always exists. The encoder divides the representation into two features which represent intrinsic content (FIC) and consider additive degradation (FAD). The presented approach in this paper provides flexibility to adjust the degree of noise reduction during the image decoding process. In the following sections we will show how leveraging the contribution levels of features extracted in the latent space can provide the desired flexibility in the output.

3 Proposed Method

Within the context of signal compression, the variable x represents an image that can be either in its pristine form or in a distorted state, such as user-generated content (UGC). The distortion exhibits a distinct nature of the contextual image signal. The proposed compression model utilizes a transform coding technique that effectively controls both context and distortion. In transformation coding, the encoder representation is illustrated as:

$$b = en(x) \tag{1}$$

where b shows the bitstream that can provide a bit-rate $R(b)$. The output of the decoder is depicted as:

$$\hat{x} = de(en(x)) \tag{2}$$

where de is the decoder. Then the learned image coding problems can be illustrated as:

$$J = \min_{en \backslash de} D(x, \hat{x}) + \lambda.R(b) \tag{3}$$

where J shows the rate-distortion optimization equation. λ is a rate-distortion trade-off to regulate bit rates. $D(x, \hat{x})$ is the distortion term. D represents the distortion term, indicating the difference between the input image and the reconstructed image. R stands for the bit rate needed to store or transmit the bit stream.

When the input signal is distorted, or UGC, this presents a challenging issue. In some scenarios Pavez et al. [2022], Xiong et al. [2023], as the input is assumed UGC or noisy, it is suggested to reformulate the RDO so that distortion is calculated concerning the original signal, $x_{pristine}$:

$$J = \min_{en \backslash de} D(x_{pristine}, \hat{x}) + \lambda.R(b) \tag{4}$$

To satisfy the optimal requirement in (4), the decoded signal $\hat{x}$ must be similar to the distorted free original content x. However, the pre-processed eliminated distortion cannot be brought back. As a result, the compressed image will lose information about the original distortion, which may include crucial information for some purposes, such as medical image analysis, photos as legal proof, and forensics.

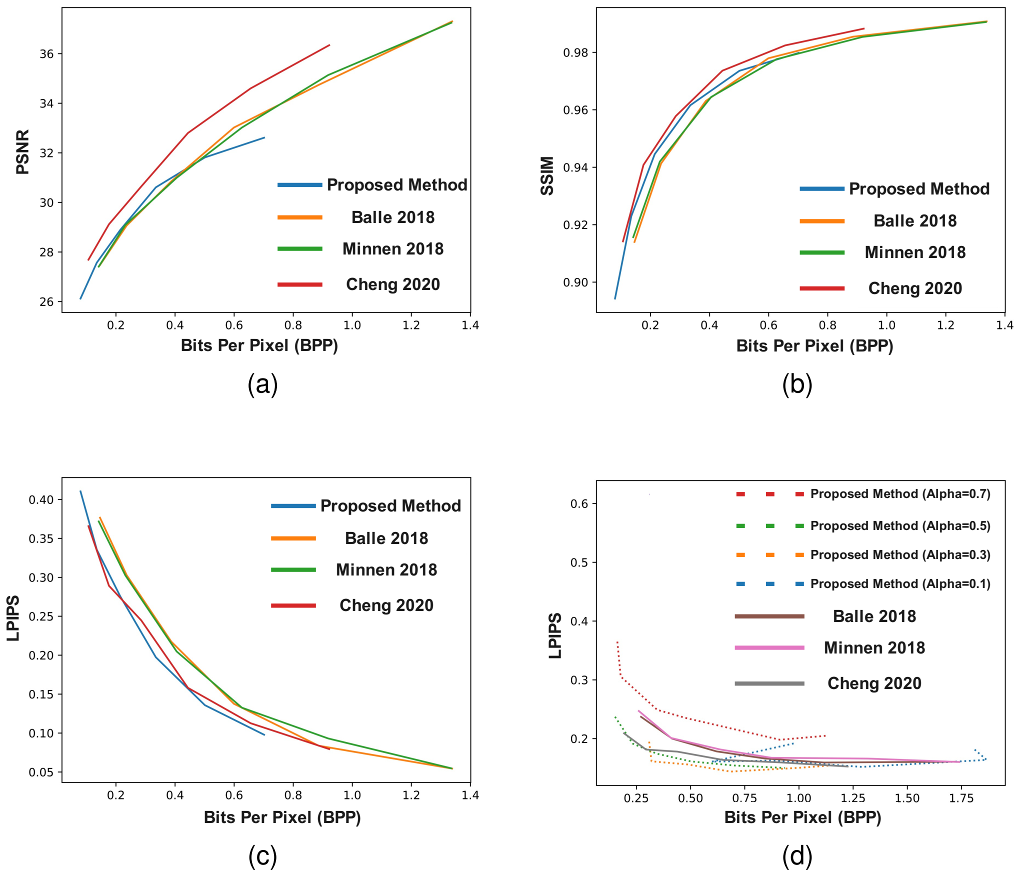

Figure 2: Rate-distortion (RD) curves collected employing the Kodak and SIQAD datasets. (a): The RD curves of image compression methods, using the PSNR quality metric. (b): The RD curves applying the SSIM quality evaluator. (c): The RD curves using LPIPs. (d): The RD curves when input images are distorted and selected from the SIQAD dataset; output images are evaluated against pristine images using LPIPS employing different αs.

We adopted the end-to-end model Cheng et al. [2020] and proposed a multi-task image compression that provides distortion-free images and distortion maps. The proposed two branches framework can be explored for controlling the output signal's quality, employing the latent space information. As a result, we present a joint framework that includes two branches: Visual Signal (VS) and Distortion Signal (DS) illustrations, and a sub-network that merges and reconstructs the distorted input signal using the VS and DS features adaptively.

As it is illustrated in Figure 1, the Encoder divides into two parts to extract distortion features and visual content. The first branch learns to extract distortion features which is fed into the g_{s1} as the decoder to construct the Distortion Map, while the second is trained to represent image Visual Content features which are decoded using g_{s2} to reconstruct the clean image.

The features are combined. A weight factor is considered for each portion which indicates their respective contributions to the output:

$$y = \alpha D_f + (1 - \alpha) C_f. \tag{5}$$

where D_f indicates distortion features and C_f shows the content features. The resulting feature vector ,y, is obtained representing the latent space feature corresponding to the input data. The achieved feature vector is then quantized and prepared for entropy coding. The main decoder ,g_{s3}, receives y and provides the reconstructed image. Figure 1 demonstrates our architecture in detail.

Figure 3: Reconstructing an image using a sample dataset given by Kodak. (a): An input image that has been distorted by Gaussian noise. (b): The reconstructed image with equal contributions of VS and DS components. (c): The model's output when VS influences it more than DS. (d): The resulting image when the effect of DS is more than that of VS.

A range of outputs with flexible distortion levels can be achieved in the presented multi-task framework. we conducted experiments to estimate the quality of the image using latent space representation.

Quality-aware latent space features can provide a flexible mechanism for compromising the balance between fidelity (the distorted/UGC input) and the reconstructed output's quality (the pristine image).

Using RDO, the encoder-decoder parameters in conventional compression methods are acquired by resolving the optimization problem as follows:

$$min_{\theta} D(\theta) + \lambda R(\theta) \tag{6}$$

in which θ illustrates the sets of coding parameters, D shows distortion between the reconstructed signal and the reference, and R refers to the dedicated bit-rate regarding the reconstructed image. As mentioned before, the reference image can be considered distorted or in some scenarios the pristine image. λ controls the trade-off between the rate and quality of the reconstructed signal. Small λ may result in larger bit-rates; hence when encoding UGC or distorted image, a small λ encourages quality saturation. In Xiong et al. [2023], a saturation detection method is presented, and the effect of different λ values is analyzed. In a traditional video compression system, it is presented that the quantization parameter (QP) can be easily converted into λ Richardson [2004].

In this case, we introduce a framework that estimates clean image features and generates a distortion map simultaneously. Considering this information, the flexible signal reconstruction mechanism is presented. Based on the application, the rate of the noise combination can be regulated. The effect of different levels of noise combination is illustrated in Figure 3 which can be modeled as λ in RDO.

4 Experiments

We assessed the suggested framework for distortion-aware learning image compression by utilizing latent space vectors. The network is assumed to get a distorted image, and the proposed model aims to provide flexible decoding by utilizing extracted features from the encoder. This allows the decoder to reconstruct the image from a distorted state to a nearly perfect image.

To train the proposed model, we combined the TID2013 and SCID datasets to construct the training and validation sets. The dataset comprises 3366 images, of which 2106 images originated from TID2013, and 1260 were obtained from SCI datasets. To introduce variability, we incorporate various types and intensities of distortion during training. The proposed model was trained on a workstation with an NVIDIA RTX 3090 GPU and an AMD Ryzen 5 5600X processor.

To assess the effectiveness of the proposed flexible learned image compression model, we performed various experiments using a well-known image compression dataset, Kodak Kodak [1993]. Moreover, we selected an image quality assessment dataset, SIQAD Yang et al. [2015], and a distorted Kodak dataset for flexible image quality decoding evaluation. The Kodak dataset collection has 24 pristine photos. The SIQAD dataset contains distorted screen content images and their corresponding clean images. The distorted Kodak collection comprises Kodak images corrupted by Gaussian noise. In addition, three cutting-edge models were chosen to assess the proposed model's performance by comparing rate-distortion curves. The Ballé et al. [2018], Minnen et al. [2018a], and Cheng et al. [2020] are excellent models chosen as the most advanced in the learned image compression.

To evaluate the proposed method, we selected three state-of-the-art approaches Balle, Minnen, and Cheng, with six pre-trained models for each method. Each pre-trained model encodes images at a particular bitrate. We additionally trained the proposed model in six distinct configurations for each compression level measured in bits-per-pixel. The reconstructed image by each specific model is evaluated using three metrics PSNR, SSIM, and LPIPS Zhang et al. [2018]. The LPIPS compares images using features extracted with deep neural networks and provides better performance than PSNR and SSIM. For each plot, the rate-distortion score is determined indicating compression level and reconstructed quality score.

Figure 2 provides a rate-distortion evaluation for the Kodak dataset. Regarding basic quality metrics, specifically PSNR(Fig. 2a) and SSIM(Fing. 2b), the corresponding RD curves demonstrate that the proposed method outperforms Ballé and Minnen works. However, Cheng et al. indicate better reconstruction quality in terms of PSNR and SSIM. As it is illustrated in Figure 2, results indicate that our proposed model effectively reconstructs the structural information. To illustrate this result quantitatively we evaluate the quality of the results using LPIPS as the SOTA method as well. The LPIPS criterion shows that our technique exhibits superior or in some cases similar performance. Figure 2c provides a detailed comparison of various setups of the models. We conducted another experiment to analyze the model's flexibility in decoding various types of image quality. Quality-aware latent space features can provide flexibility for balancing fidelity and reconstructed quality. To decode images in different ranges of quality, the latent spaces of Visual Signal(VS) and Distortion Signal(DS) features are explored.

The proposed model controls the involvement of VS and DS using the $alpha$ coefficient, as seen in Equation 5. Based on this coefficient, we designed five distinct configurations in which $alpha$ is set as 0.1,0.3,0.5,0.7 and 0.9. During the evaluation process, the model gets distorted images and reconstructs the input based on the given configuration. Subsequently, the resulting image is compared with a pristine image from which the distorted image was constructed. VS and DS have equal weight in image reconstruction when $alpha$ is set to 0.5. In the two alternative setups, the level of VS involvement exceeds that of DS when $alpha$ is set to 0.1 and 0.3, resulting in VS being weighted by 0.9 and 0.7, respectively. In the two remaining configurations, $alpha$ is set as 0.9 and 0.7 respectively in which the effect of noise is clearly shown in Figure 2. In this experiment, the assessment is conducted utilizing the SIQAD dataset. The results in Figure 2 show that augmenting the impact of VS leads to the model generating more refined images while reducing the effects of VS has the opposite effect.

Moreover, the effect of DS and VS contributions is illustrated visually. As shown in Figure 3 when the $alpha$ is decreased, the VS effect increases and the quality of the reconstructed image improves. Figure 3 b,c, and d illustrate the results when $alpha$ is set to 0.5,0.3 and 0.7 respectively. It is clearly shown that $alpha$ =0.3 provides the image with sharper text and edges than images reconstructed in regular configurations or models with more DS contribution.

5 Conclusion

Learning-based image compression models are mostly optimized for reconstructing high-quality images. However, wild images present very different characteristics, which causes the learning-based image coding to be sub-optimal. Some distortions are reduced/eliminated during compression, which may be preferred to be kept in some applications. This paper offers a framework for compressing UGC images, providing the capability of providing a range of distortion-free and level of distortions in reconstructed images using two latent space features: Visual Signal (VS) and Distortion Signal (DS). The reconstruction of the wild or distorted input involves merging the whitening VS and DS features. The presented multi-task flexible framework can generate an adaptable decoded signal. Various distortions have been utilized in our experiment, providing competitive results with the state-of-the-art baseline while enhancing the capacity to obtain a flexible range of distortion. The experiments illustrate the effective results of the presented flexible multi-task framework based on various quality metrics such as LPIPS.

References

Jpeg - final call for proposals on jpeg ai. *94th JPEG Meeting, ISO/IEC JTC 1/SC29/WG1 N100095*, 2022.

Saeed Ranjbar Alvar, Mateen Ulhaq, Hyomin Choi, and Ivan V Bajić. Joint image compression and denoising via latent-space scalability. *arXiv preprint arXiv:2205.01874*, 2022.

Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018.

Ka Leong Cheng, Yueqi Xie, and Qifeng Chen. Optimizing image compression via joint learning with denoising. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pages 56–73. Springer, 2022.

Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7948, 2020.

Jan Klopp, Yu-Chiang Frank Wang, Shao-Yi Chien, and Liang-Gee Chen. Learning a code-space predictor by exploiting intra-image-dependencies. In *BMVC*, page 124, 2018.

Eastman Kodak. Kodak lossless true color image suite (photocd pcd0992). *URL http://r0k.us/graphics/kodak*, 6, 1993.

Jooyoung Lee, Seunghyun Cho, and Seung-Kwon Beack. Context-adaptive entropy model for end-to-end optimized image compression. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=HyxKIiAqYQ`.

Binzhe Li, Shurun Wang, and Shiqi Wang. Defending against noise by characterizing the rate-distortion functions in end-to-end noisy image compression. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3727–3731. IEEE, 2021.

Lin Liu, Xu Jia, Jianzhuang Liu, and Qi Tian. Joint demosaicing and denoising with self guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2240–2249, 2020.

David Minnen and Saurabh Singh. Channel-wise autoregressive entropy models for learned image compression. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3339–3343. IEEE, 2020.

David Minnen, George Toderici, Michele Covell, Troy Chinen, Nick Johnston, Joel Shor, Sung Jin Hwang, Damien Vincent, and Saurabh Singh. Spatially adaptive image compression using a tiled deep network. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2796–2800. IEEE, 2017.

David Minnen, Johannes Ballé, and George D Toderici. Joint autoregressive and hierarchical priors for learned image compression. *Advances in neural information processing systems*, 31, 2018a.

David Minnen, George Toderici, Saurabh Singh, Sung Jin Hwang, and Michele Covell. Image-dependent local entropy models for learned image compression. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 430–434. IEEE, 2018b.

Eduardo Pavez, Enrique Perez, Xin Xiong, Antonio Ortega, and Balu Adsumilli. Compression of user generated content using denoised references. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4188–4192. IEEE, 2022.

Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database tid2013: Peculiarities and preliminary results. In *European workshop on visual information processing (EUVIP)*, pages 106–111. IEEE, 2013.

Iain E Richardson. *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

Michela Testolina, Evgeniy Upenik, and Touradj Ebrahimi. Towards image denoising in the latent space of learning-based compression. In *Applications of Digital Image Processing XLIV*, volume 11842, pages 412–422. SPIE, 2021.

Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017.

Wenzhu Xing and Karen Egiazarian. End-to-end learning for joint image demosaicing, denoising and super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3507–3516, 2021.

Xin Xiong, Eduardo Pavez, Antonio Ortega, and Balu Adsumilli. Rate-distortion optimization with alternative references for ugc video compression. *arXiv preprint arXiv:2303.06254*, 2023.

Huan Yang, Yuming Fang, and Weisi Lin. Perceptual quality assessment of screen content images. *IEEE Transactions on Image Processing*, 24(11):4408–4421, 2015.

Pingping Zhang, Meng Wang, Baoliang Chen, Rongqun Lin, Xu Wang, Shiqi Wang, and Sam Kwong. Learning-based compression for noisy images in the wild. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: It is clearly expressed that we present a framework for compressing distorted images while estimating distortion-free ones.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: As usual in learning-based image compression, security issues such as backdoor attacks were not investigated.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The implementation of our method is available at `https://anonymous.4open.science/r/flexible_compression-9E80`

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The implementation of our method is available at `https://anonymous.4open.science/r/flexible_compression-9E80`

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The experiment details are provided with the code at `https://anonymous.4open.science/r/flexible_compression-9E80`

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: We followed the usual procedure in the state-of-the-art methods.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provide sufficient information on the computer resources

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This paper conforms to the NeurIPS Code.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We presented a learning-based image compression algorithm that seems to have no negative societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [No]

    Justification: This paper does not describe any safeguards.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We cited the employed code and datasets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: A trained model has been provided.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: We have not used crowdsourcing. In addition, human subjects were not involved.

    Guidelines:

    - The answer NA means that the paper does not involve crowd sourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: In our research, we have not employed human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.