Contrastive Learning for Low-Light Raw Denoising (Student Abstract)

Taoyong Cui [1], **Yuhan Dong** [1]

[1]Shenzhen International Graduate School, Tsinghua University,Shenzhen, China
cty21@mails.tsinghua.edu.cn, dongyuhan@sz.tsinghua.edu.cn

Abstract

Image/video denoising in low-light scenes is an extremely challenging problem due to limited photon count and high noise. In this paper, we propose a novel approach with contrastive learning to address this issue. Inspired by the success of contrastive learning used in some high-level computer vision tasks, we bring in this idea to the low-level denoising task. In order to achieve this goal, we introduce a new denoising contrastive regularization (DCR) to exploit the information of noisy images and clean images. In the feature space, DCR makes the denoised image closer to the clean image and far away from the noisy image. In addition, we build a new feature embedding network called Wnet, which is more effective to extract high-frequency information. We conduct the experiments on a real low-light dataset that captures still images taken on a moonless clear night in 0.6 millilux and videos under starlight (no moon present). The results show that our method can achieve a higher PSNR and better visual quality compared with existing methods.

Introduction

Due to the very limited photons count and the presence of escapeable noise, it is a great challenge for the quality of low-light photography and videography. In general, we can increase the aperture setting to use high ISO, and extend the exposure time to collect more light. However, high ISOs, while effectively making each pixel more sensitive to light and allowing shorter exposures, can also amplify the noise in each frame. Therefore, more efficient algorithms are needed to obtain high-quality images and videos under low-luminance light.

Over the years, several denoising algorithms have been developed to improve the image/video quality, from classic methods (e.g., spatial domain methods, and transform domain filtering methods) to deep learning based approaches. Each of these methods always attempts to extract the signal from the noise based on some assumptions about the statistical distributions of the image and noise (Monakhova et al. 2022). While these methods have achieved reasonably good performance in some denoising tasks, most are built upon simplistic noise models (Gaussian or Poisson-Gaussian noise), which do not well reflect the severe quantization, bias, and clipping that arise in extreme low-light conditions. Without a good understanding of the structure of the noise in the images/videos, the denoising effect will be poor in the actual low-light denoising task.

The challenge of denoising in low light is well-known in the computational photography community but remains open. Rather than assuming a certain noise model, recently, some learning-based methods automatically account for the low-light noise through a deep neural network and massive training image pairs. However, they always require robust alignment techniques to account for any motion in the scene, which is difficult in the presence of extreme noise. So they always fail to remove such noise in extremely low-light conditions and do not effectively address the color bias.

In order to tackle this challenge, we propose a new approach with three novel contributions: **(1)** We design a new denoising contrastive regularization (DCR) to exploit the information of noisy images and clean images. The DCR makes that the denoised images (anchors) are pulled closer to the clean images (positive samples) and pushed far away from the noisy images (negative samples) in the feature space. **(2)** In DCR, different from the existing methods which adopt a prior model (e.g., pre-trained VGG model) to obtain the feature embedding, we build a novel and task-related one called Wnet, which is more effective to extract high-frequency texture information. **(3)** Experiments show that our method outperforms several existing methods in terms of quantitative and qualitative results. In addition, we achieve state-of-the-art performance on the starlight dataset.

Our Method

Contrastive learning is one of the most powerful approaches for representation learning. It aims at pulling the anchor sample close to the positive samples and pushing it far away from negative samples in latent space (Wu et al. 2021). In this work, we propose a new contrastive regularization (DCR) to generate higher quality denoised images. Most importantly, in DCR, we need to consider three aspects: first, constructing "positive" and "negative" pairs; second, finding the latent feature space of these pairs for comparison; and third, specific loss function. Specifically, in our DCR, the positive pair and negative pair are generated by the group of a clear image $\boldsymbol{P}$ and its denoised image $\boldsymbol{F}^*$, and the group of $\boldsymbol{F}^*$ and a noisy image $\boldsymbol{N}$, respectively. For the latent feature

space, we select the task-related latent space from the proposed Wnet for practical feature embedding. And we propose the Closs function to fully utilize the network.

We develop a simple but efficient network called Wnet to achieve feature embedding. Specifically, in order to learn more noise information, we use the Haar wavelet transform to extract the informative high-frequency components (HL, LH, and HH) and then use CNN to learn these features. For pre-training this feature network, we use the prior noise model (Monakhova et al. 2022) to synthesize additional noisy samples and then train with all clean training samples for binary classification, the criterion we choose is Cross-entropy loss. With this simple classification task, Wnet is able to utilize prior knowledge and focus on more noise information in the latent feature space of these "positive" pairs and "negative" pairs for contrast. Later, Wnet will be used in the denoising stage to feature embedding in DCR. To enhance the contrastive ability, we extract hidden features from different convolution layers of Wnet.

To fully utilize Wnet, we use multi-intermediate features from Wnet in our contrastive loss and propose a new loss function based on Eq. (1). For a target denoised image $\boldsymbol{F_l^*}$, its positive and negative counterparts are noted as $\boldsymbol{P_l}$ and $\boldsymbol{N_l}$ respectively. The feature representations for the denoised image, positive and negative counterparts are noted as $\boldsymbol{f}$, $\boldsymbol{p}$, and $\boldsymbol{n}$, respectively. Our contrastive loss for the $\boldsymbol{l}$-th sample on the $\boldsymbol{i}$-th convolution layer is defined as follows:

$$\mathcal{L}_{\boldsymbol{Closs}} = \sum_{l=1}^{L}\sum_{i=1}^{N} \omega_i \cdot \frac{s\left(G_i(p), G_i(f)\right)}{s\left(G_i(n), G_i(f)\right)}, \tag{1}$$

where $\boldsymbol{G_i}$ = 1, 2, $\cdots$, N extracts the $\boldsymbol{i}$-th hidden features from the Wnet, and w_i is a weight coefficient. Let's note the shape of the feature map as $\boldsymbol{C}(channels) \times \boldsymbol{H}(height) \times \boldsymbol{W}(width)$. Inspired by (Wu et al. 2021; Wu et al. 2021), we adopt the mean value of pixel-wise cosine similarity with L1 loss as the similarity between feature maps. The function $s(f^x, f^y)$ is defined as follows:

$$s\left(f^x, f^y\right) = \frac{1}{2HW}\Big(\sum_{h=1}^{H}\sum_{w=1}^{W} \frac{f_{hw}^x f_{hw}^y}{\|f_{hw}^x\|\,\|f_{hw}^y\|} + 2|f_{hw}^x - f_{hw}^y|\Big). \tag{2}$$

Experiments

To test the effectiveness of the DCR, we conducted extensive experiments on a real low-light image/video dataset. For datasets, we use all datasets from (Monakhova et al. 2022). All images/videos are captured in RAW format. Before the denoising process, we used the noise generator (Monakhova et al. 2022) to generate training data due to the difficulty in obtaining clean/noisy image video pairs. And actual noisy data was used to verify its effectiveness during testing. In the denoising process, the inputs to Wnet are noisy frames/images, processed denoised frames/images, and clean frames/images. The output feature representations of Wnet is used to generate the Closs. Other experimental settings are the same as (Li et al. 2022). We quantitatively and qualitatively compare our performance on the dataset, and the results are shown in Table 1 and Figure 1,respectively. Extensive experiments have shown that our proposed approach can help existing benchmarks achieve better performance.

	PSNR	SSIM	LPIPS
Baseline	31.44	0.852	0.0696
Baseline+(VGG+L1)	30.23	0.842	0.0635
Baseline+DCR (Ours)	**32.23**	**0.853**	**0.0630**

Table 1: Performance on still images from the test set by training FastDVDnet with still images and videos."Baseline+(VGG+L1)" is from (Wu et al. 2021)

	PSNR	SSIM	LPIPS
FastDVDnet*	23.8	0.618	0.282
Baseline (retrained)	27.3	0.837	0.091
Baseline (**SOTA**)*	27.7	**0.931**	0.078
Baseline+(VGG+L1)	28.1	0.831	0.103
Baseline+DCR (Ours)	**30.5**	0.889	**0.060**

Table 2: Performance on still images from the test set by training Modified FastDVDnet (Monakhova et al. 2022) with still images and videos. "retrained" represents that we retrain the model according to the method of the paper (Monakhova et al. 2022), "*" data is from (Monakhova et al. 2022).

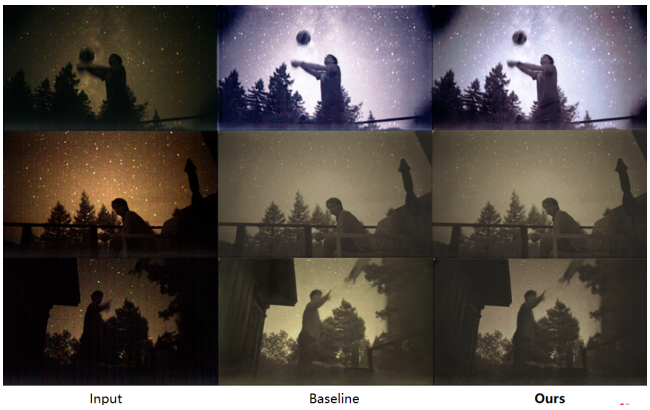

Figure 1: Visual comparison of the Modified FastDVDnet performance on unlabeled submillilux videos (after the same post-processing).

Acknowledgements

This research is supported in part by Science, Technology and Innovation Commission of Shenzhen Municipality under grant KCXFZ20211020163813019 and WDZC20200818121348001.

References

Li, D.; Zhang, Y.; Law, K. L.; Wang, X.; Qin, H.; and Li, H. 2022. Efficient Burst Raw Denoising with Variance Stabilization and Multi-frequency Denoising Network. *arXiv e-prints*, arXiv:2205.04721.

Monakhova, K.; Richter, S. R.; Waller, L.; and Koltun, V. 2022. Dancing Under the Stars: Video Denoising in Starlight. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16241–16251.

Wu, G.; Jiang, J.; Liu, X.; and Ma, J. 2021. A Practical Contrastive Learning Framework for Single Image Super-Resolution. *arXiv e-prints*, arXiv:2111.13924.

Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive Learning for Compact Single Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10551–10560.