Learning Optimized Low-Light Image Enhancement for Edge Vision Tasks

S M A Sharif, Azamat Myrzabekov, Nodirkhuja Khujaev, Roman Tsoy, Seongwan Kim, Jaeho Lee [*]

Opt-AI Inc.

LG Sciencepark, Seoul, South Korea

{sharif, azamat, nodir, roma, swan.kim, jaeho.lee }@opt-ai.kr

Abstract

Low-light image enhancement (LLIE) has a significant role in edge vision applications (EVA). Despite its widespread practicability, the existing LLIE methods are impractical due to their high computational costs. This study proposed a framework to learn optimized low-light image enhancement to tackle the limitations of existing enhancement methods for accelerating EVA. The proposed framework incorporates a lightweight and mobile-friendly deep network. We optimized our proposed model with INT8 precision with a post-training quantization strategy and deployed it on an edge device. The LLIE model has achieved over 199 frames per second (FPS) on a low-power edge board. Additionally, we evaluated the practicability of an optimized model for accelerating the vision application of an edge environment. The experimental results illustrate that our optimized method can significantly accelerate the performance of SOTA vision algorithms in challenging low-light conditions for numerous everyday vision tasks, including object detection and image registration.

1. Introduction

LLIE refers to enhancing the quality of images captured under challenging low-light conditions. Enhancing poorly lightened images is typically considered among the most complex image restoration tasks due to the lack of visible details and aggressive sensor noise [1, 38]. Therefore, enhancing low-light images requires comprehensive reconstruction techniques such as contrast enhancement, noise reduction, and detail restoration. Apart from its aesthetic importance, LLIE has numerous real-world applications, including accelerating generic vision tasks on low-power edge devices.

LLIE on-edge devices incorporate widespread applications, ranging from surveillance and security systems to autonomous vehicles and industrial monitoring, often op-

erating in real time and relying heavily on visual information for decision-making. Regrettably, the performance of edge applications such as object detection (OD), segmentation, stereo vision, etc., when camera sensors suffer from quantum inefficiency and produce darker images [38]. Improving the quality of such low-quality images can enhance their ability to detect and recognize objects accurately, even in challenging lighting conditions. Overall, strengthening the low-quality images can improve safety and efficiency in real-world scenarios.

Recent literature has seen a surge in efforts to tackle LLIE challenges through innovative learning-based approaches [18]. These approaches primarily fall into two categories. The first category intends to learn LLIE as an image-to-image translation task [3, 8, 9, 17, 23, 28, 40, 44, 44, 52]. Typically, these methods leverage convolutional neural networks (CNN) for end-to-end learning. Also, a few recent methods from the same category [6, 51] combined their CNN architectures with transformers [42], similar to the recent image restoration methods [48, 55, 56]. The second category of LLIE incorporates the Retinex theory [21] into CNN architecture for enhancing low-light images[4, 41, 47, 49, 50, 54, 57, 59, 60]. However, most existing LLIE methods are developed without considering the feasibility of deploying their novel networks on edge devices. Retinex-based and attention-based methods' complex nature incorporates high computational cost edge unfriendly convolution maneuvers. Consequently, these SOTA LLIE solutions could be more practical for deploying on-edge devices.

To address the limitation and take LLIE research in a realistic direction, we propose a novel edge-friendly LLIE learning and deployment framework. Our proposed framework incorporates a lightweight U-shape deep model [35] and optimization strategy. We optimized our model by leveraging Post-training quantization (PTQ)[25] and downgraded precision from the floating weight into INT8 for faster inference in an edge environment. Later, we deployed our method to real-edge devices and perceived an FPS of 199. The practicability of such a low-precision LLIE model
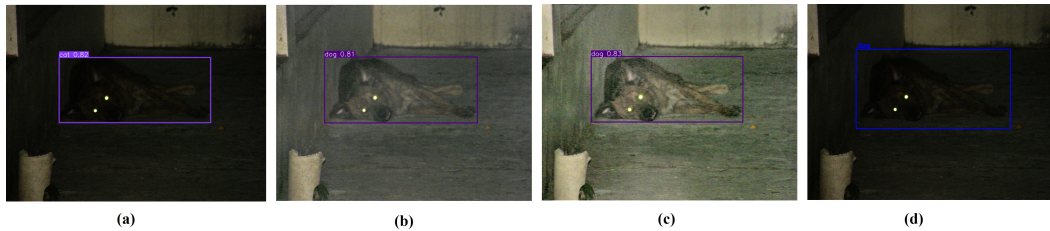
---

[*]Corresponding author

Figure 1. Real-world OD enhancement leveraging proposed optimization framework. (a) Low-light image with Yolo-V3. (b) Our model (INT8) with Yolo-V3. (c) Our model (full precision) with Yolo-V3. (d) Ground-truth (for detection).

is extensively evaluated by incorporating it into generic vision tasks (e.g., OD and image registration) [50]. The experimental results illustrate that our optimized model can significantly improve the performance of such vision applications even in stochastic lighting conditions. To the best of concern, this is the first open literature exploring quantized models' practicability in edge vision tasks. Our framework can help with everyday vision tasks like OD, even in very dark scenarios, as shown in Fig. 1. Our contributions are as follows:

- We propose a LLIE learning framework for edge devices. Our proposed framework incorporates a lightweight deep model and deployment strategy in an edge environment.
- We optimized our model and achieved an FPS of 299 on real-edge hardware by sacrificing minor fidelity scores.
- We incorporate our optimized model for accelerating the performance of generic vision tasks.

2. Related Work

This section briefly describes the related works, including methods for LLIE and Quantization methods on image enhancement.

2.1. Low-light Image Enhancement

Researchers have developed various algorithms for low-light image enhancement to improve the quality of images captured in low, insufficient-light environments. The most straightforward approach used for this task is a Histogram equalization-based approach[15, 27]. These methods are computationally effective and only sometimes produce well-defined results. Meanwhile, Retinex-based [21] enhancement algorithms work much better due to leveraging the human visual system's ability to perceive colors and brightness relative to surrounding regions, resulting in visually pleasing enhancements. With the development of deep learning methods, current deep learning methods were inspired by Retinex theory, such as Retinex-Net [49], which shows promising results for image enhancement. Besides CNN models, transformer architecture was used to learn or adjust the parameters of Retinex and produce promising results. One such model is RetinexFormer[4], where a one-

stage transformer was used to learn illumination information from low-light images. Moreover, several diffusion-based approaches, such as Diff-Retinex[54], were made, where generative diffusion models were used to learn low-light image features. Although all these methods show excellent results on low-light image enhancement tasks, they need more interpretability. Notably, the complex nature of this SOTA method makes them edge unfriendly and impractical for edge deployment.

2.2. Quantization on Enhancement

Model quantization is a crucial step for deploying networks on mobile devices. Unfortunately, only a few recent methods from the image enhancement domain leverage the capability of quantization for edge optimization. Among them is a recent method, Fully Quantized Image Super-Resolution Framework (FQSR)[46], which aims at obtaining end-to-end quantized models for all layers, mainly including skip connections. PAMS [22] applied layer-wise symmetric quantization to extract high-level features. A channel-wise distribution-aware quantization scheme for image enhancement was introduced by DAQ[12]. In contrast, DDTB[61] introduced the design of an asymmetric activation quantizer, which compresses to 2-4 bit by introducing dynamic dual trainable clip values. Apart from that, a recent method applied PTQ for Image Super Resolution[45], authors introduce the density-based dual clipping to cut off the outliers by analyzing the asymmetric bounds of activations.

It is worth noting that the existing quantitative enhancement methods have yet to explore the challenges and practicability of quantization on LLIE. However, a few novel quantized methods [2, 7, 29] are introduced for image restoration, like super-resolution. In contrast, such restoration tasks have different characteristics from those of the LLIE. Image enhancement techniques like super-resolution focus on improving the details, whereas LLIE incorporates more challenging tasks due to low visibility and missing details in the darker regions. In addition to that, these methods mainly conclude their study without exploiting real-world edge hardware. Thus, the practicability of quantization on image enhancement, particularly for LLIE, is still undis-

closed. This study explored the practicability of quantized LLIE on EVA utilizing actual edge hardware.

3. Method

This section describes the proposed LLIE edge optimization framework and its learning strategies. Fig. 2 illustrates the overview of the proposed framework. We develop a lightweight deep model with hardware-friendly vanilla convolution operation. We quantized our trained network into INT8 precision for faster inference. Later, we deployed our network on a real-edge device to evaluate its practicability on real-world edge vision applications (EVA).

3.1. Network

We consider LLIE for EVA to be an image-to-image translation task. Thus, it aims to map a low-light image (I_L) to a perceptually enhanced image (I_H) using a mapping function M : $I_L \rightarrow I_H$. Here, I_L is constrained to the range $[0,1]^{H \times W \times 3}$, where H and W denote the height and width of both input and output images, respectively. Fig. 2 provides an overview of the proposed deep model architecture.

3.1.1 Architecture

The proposed deep network is structured as a fully convolutional encoder-decoder architecture [5, 10], featuring convolutional skip connections. The initial layer of the generator transforms the input image (I_L) into a 64-depth feature map. This input convolutional layer employs a kernel size of 3×3, padding of 1, and stride of 1. Following this, an encoder consists of four consecutive feature levels, with alternating feature depths of $d = 64, 96, 128, 196$. Each feature level integrates a residual block, which processes an input feature X using the equation:

$$F' = \mathrm{C}(X) + X \qquad (1)$$

Here, $\mathrm{C}(\cdot)$ comprises two convolution layers, each with a kernel size of 3×3, padding of 1, and stride of 1, with the first layer activated using PReLU activation. The output feature dimension of a residual block remains unchanged from its input (i.e., 64 in the initial feature level). Notably, the residual blocks are well-known for achieving better denoising performance. Such architectural choices help us in handling sensor noise in low-light conditions.

Following each residual block in the encoder, a convolutional downsampling layer is applied as follows:

$$F_\downarrow = C_\downarrow(X_0) \qquad (2)$$

Here, $C_\downarrow$ represents a 3×3 convolution operation with a stride of 2.

The decoder section of the network mirrors the encoder in terms of the number of feature levels, with an upsampling layer following each residual block. The upsampling operation is implemented as follows:

$$F_\uparrow = C_\uparrow(X_0) \qquad (3)$$

Here, $F_\uparrow$ involves a transpose convolution operation [38]. The transpose convolution helps us make our model fully convolutional with effective restoration performance.

In addition to the basic indoor-decoder design, we incorporated a gated skip connection to propagate refined low-level features. It is worth noting that the U-shaped encoder encodes the noise along with salient information of low-light images. We found that propagating such sensor noises in the decoder pushes the deep model to achieve desired enhanced images [39]. Therefore, we leverage a convolution residual-gating mechanism as follows:

$$F_G = C_{1 \times 1}(X_R) \qquad (4)$$

Here, $C_{1 \times 1}$ represents point-wise convolution that refines the spatial information of a given tensor X_R.

Finally, the decoder portion culminates in a final convolutional layer, which produces the three-channel enhanced image using a convolutional kernel size of 3×3, padding of 1, and a stride of 1. This final output layer is activated with a tanh function to yield the final images within the range of $[0, 1]$.

3.1.2 Learning Objective

We leverage a pixel-wise reconstruction loss to guide our deep model as a coarse-to-refine reconstruction process. An L1 or L2 distance typically serves as the pixel-wise loss function. While both options are standard, the L2-loss is directly related to the Peak Signal-to-Noise Ratio (PSNR) metric and tends to yield smoother images [20, 38]. Notably, low-light images comprise a substantial sensor noise. Thus, we utilized an L1 objective function as the reconstruction loss. Our reconstruction loss can be represented as follows:

$$\mathcal{L}_R = \| I_G - I_H \|_1 \qquad (5)$$

Here, I_H represents the output obtained through M, while I_G denotes the reference high-light image. This loss function quantifies the absolute differences between corresponding pixels in the output and reference images, thereby driving the network to minimize these differences during training.

3.2. Post-training quantization (PTQ)

PTQ is a technique used to reduce the memory footprint and improve the inference speed of deep neural
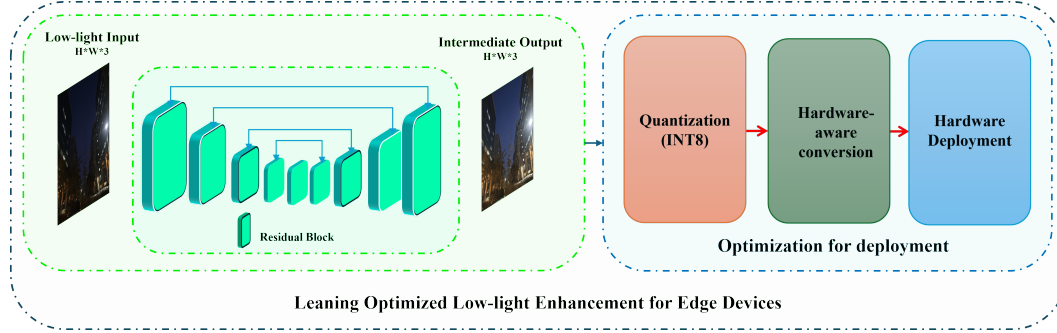
Figure 2. Overview of the proposed LLIE edge optimization framework. Our framework incorporates a lightweight deep model, a PTQ, and a real-hardware deployment strategy for faster inference speed in an edge environment. We deployed our model on actual low-power hardware to find its practicability in the later phase.

networks after training [25, 45]. It involves converting the network's weights and activations from high-precision floating-point representations (e.g., 32-bit floating point) to lower-precision fixed-point representations (e.g., 8-bit integers). This process significantly reduces the network's memory requirements and computational complexity, making it more efficient to deploy on resource-constrained devices such as edge devices. It is worth noting that PTQ can achieve even lower bit rates for faster optimization. However, we found that most edge boards available for real-world use (e.g., autonomous driving and security) support 8-bit inference [62]. Therefore, considering the actual usage, we quantized our model into INT8 precision for real-edge deployment.

3.2.1 Calibration.

Model calibration is crucial for determining appropriate scaling factors for lower bit precision when performing PTQ on deep networks. The calibration process is typically perceived using numerous techniques (e.g., histogram and max calibration) [14]. We found the histogram calibration process fits our application well. Notably, the low-light images often have a compressed histogram towards darker values. Thus, such a calibration process helps us expand the distribution, revealing details obscured in the shadows by brightening them proportionally.

$$\mathbf{Q_h} = \frac{M_h}{M_q} \qquad (6)$$

Here, we obtained histogram calibration Q_h with maximum histogram values (M_h) and maximum quantization value (Q_h). We tailor the quantization process to the specific distribution of activation values encountered during inference. This meticulous calibration technique ensures that the quantized model maintains accuracy and performance levels commensurate with the original floating-point

model, thus facilitating seamless deployment on resource-constrained devices.

3.2.2 INT8 Quantization

INT8 quantization for neural networks involves scaling and rounding the original floating-point values (like weights) to a lower precision integer format (typically 8 bits for INT8). This is achieved by considering the minimum and maximum values of the original data and then rescaling the data to a range suitable for the 8-bit integer representation. We perceive INT8 quantization as follows:

$$\mathbf{INT8_Q} = \mathbf{R}\left(\frac{\mathbf{P}}{\mathbf{S}}\right) \qquad (7)$$

Here, $\mathbf{R}$ presents The round function that ensures the rounded to the nearest integer. $\mathbf{P}$ is presents the original value to be quantized, and $\mathbf{S}$ is the scaling factor determined during calibration.

3.3. Learning Details

We leverage three low-light enhancement datasets for learning real-world LLIE, including a dataset from the New Trends in Image Restoration and Enhancement'24 (NTIRE) low-light enhancement challenge [24], LOL-v1 [49], and LOL-v2 (real data) [53]. NTIRE challenge provides 230 large-dimension images for training purposes. We extracted 4,500 non-overlapping patches from the dataset and combined them with LOL-v1 (485 training samples) and LOL-v2 (689 training samples). In the testing phase, we leverage LOL-V1 and LOL-v2 for quantatitive evaluation and the NTIRE dataset of qualitative assessment.

The proposed u-shape lightweight model has been implemented with the PyTorch framework [33] and optimized with the Pytorch Quantization library [31]. We optimized our method in the learning phase with an Adam optimizer [19]. We tuned its hyperparameters as $\beta_1 = 0.9$, $\beta_2 = 0.99$. Initially, we set the learning rate for the model as

$lr = 1e - 4$. We trained our LLIE method for 50,000 steps with a batch size 64. For the training phase, we utilized image patches of dimension $128 \times 128 \times 3$. The machine learning phase took around 6 hours, comprising an NVIDIA A6000 GPU with 32GB RAM and a 16-core CPU. Later, we deployed our model on the NVIDIA Jetson Orin edge board for real-world evaluation.

4. Experiments

This section compares the proposed deep models with SOTA lightweight deep models, real-world quantized performance, and performance on the acceleration of edge vision tasks.

4.1. Comparison with SOTA method

We compared the performance of lightweight mobile-friendly deep image enhancement networks, including Unet [35], Mobile Unet [13], and DnCNN [16]. In addition to image enhancement methods, we also compared our model with the LLIE methods like [11] and Retinexnet [49]. We trained and tested all comparing methods under the same data samples for a fair comparison. We summarised the performance of deep methods with a peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [37] and perceptual image patch similarity (LPIPS) [58]

4.1.1 Quantitative evaluation

Table. 1 illustrates the performance of the deep models on LOL-v1 and LOL-v2 testing sets. The proposed method can outperform the existing methods despite its dedicated nature for the edge device. Particularly in a diverse dataset like LOL-v2, the proposed method illustrates notable domination over its counterparts. In contrast, Retinexnet [49] can achieve a higher fidelity score in LOL-v1 despite illustrating significant perceptual degradation. It is worth noting that Retinexnet has been developed to perform well, particularly on the LOL-v1 dataset [49].

4.1.2 Qualitative Evaluation

Fig. 3 further confirms the practicability of the proposed method in real-world LLIE. The proposed method can handle complex real-world low-light scenes compared to its counterparts. The typical mobile-friendly deep architecture like mobile Unet illustrates significant deficiencies in addressing real-world low-light scenes. Such architectures' limitations confirm the practicability of an efficient and effective lightweight LLIE method for real-world applications.

4.2. Performance on Edge Device

We selected our best pre-trained weight for quantization and deployment. Infer efficiently on edge devices. Apart from our quantization strategy, we converted our quantized weight with TensorRT optimization library [30] to directly deploy in the edge board. We leverage an NVIDIA Jetson Orin board to deploy and test our proposed LLIE method. We leveraged Docker to enable the utilization of CUDA cores for GPU acceleration of the target edge board. We deployed the optimized weight on our Docker environment and operated the hardware in its efficient mode (30 watts) [32] to ensure the practicability of our proposed method on edge platforms.

Table. 2 illustrates the performance of the proposed optimization framework on different hardware platforms. The proposed framework allowed us to infer speed 8x compared to the full precision model. It achieved over 199 FPS on Jetson Orin with an image dimension of $256 \times 256 \times 3$. Notably, image dimension on inferencing edge devices can affect the perceptual quality. Overall, it makes a trade-off between FPS and fidelity score for LLIE on edge devices.

Apart from the quantatitive evaluation, we also perform a subjective comparison between the full-precision and quantized models. In our assessment, the PTQ can produce similar visual results in most cases compared to the full-precision variants, as shown in Fig. 4.

4.3. Real-world Edge Applications

EVA typically refers to computer vision tasks performed directly on edge devices, such as smartphones, IoT devices, or embedded systems, without relying heavily on cloud computing or external servers. Such vision computing approaches are becoming popular due to their ability to perform real-time visual data processing directly on edge devices, mitigating privacy concerns and reducing latency. Despite countless possibilities, developing efficient and accurate vision solutions for low-power devices is among the most complicated tasks. In this study, we aim to improve the performance of such vision tasks by incorporating our proposed LLIE.

4.3.1 Object Detection

Among existing EVA, OD is among the most widely used applications of edge vision. It allows real-time identification and localization of objects within images or video streams directly on the device. This technology has countless real-world uses, such as autonomous vehicles for identifying pedestrians and obstacles and intelligent retail systems for tracking inventory and customer behavior. Unfortunately, this widely used application also suffers in low-light conditions due to missing visible information.

Method	LOL-v1			LOLv-2			Average		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Unet [35]	15.37	0.6326	0.4007	17.22	0.6516	0.3718	16.30	0.6421	0.3863
Retinexnet [49]	**17.63**	**0.7185**	0.4051	17.71	0.6049	0.4403	17.67	0.6617	0.4227
LIME [11]	16.28	0.6278	0.4315	17.52	0.5674	0.4115	16.90	0.5976	0.4215
MobileUnet [13]	15.80	0.6508	0.4047	18.54	0.6994	0.3668	17.17	0.6751	0.3857
DnCNN [16]	16.18	0.6839	0.4209	18.91	0.7063	0.3909	17.54	0.6951	0.4059
Ours	16.44	0.7006	**0.3499**	**19.04**	**0.7558**	**0.3190**	**17.74**	**0.7282**	**0.3344**

Table 1. Qunatatitive comparison between deep methods for LLIE. Overall, the proposed network outperforms its counterpart in all evaluation metrics.
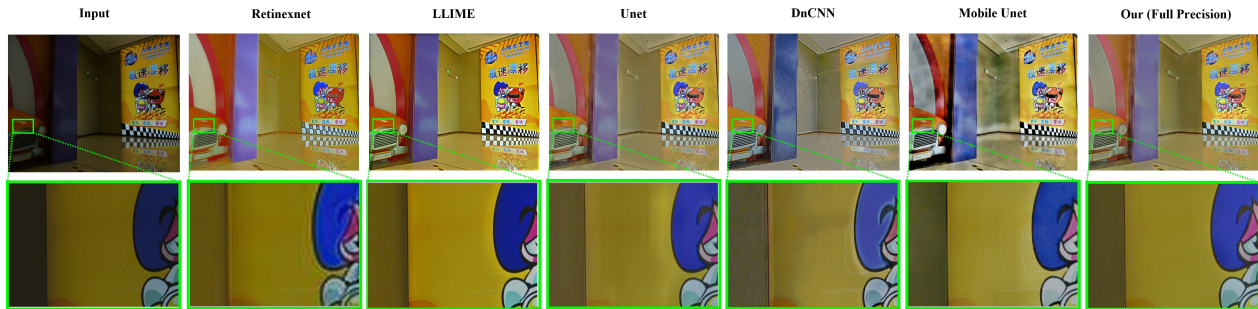

Figure 3. Real-world complex low-light scene enhancement with deep enhancement methods. The proposed method outperforms the existing method in visual comparison. Despite being lightweight and mobile-friendly, the proposed method can generate plausible highlight images without producing artifacts. From left to right: input image, Retinexnet [49], LIME [11], Unet [35], DnCNN [16], Mobile Unet[13], Ours (full-precision)
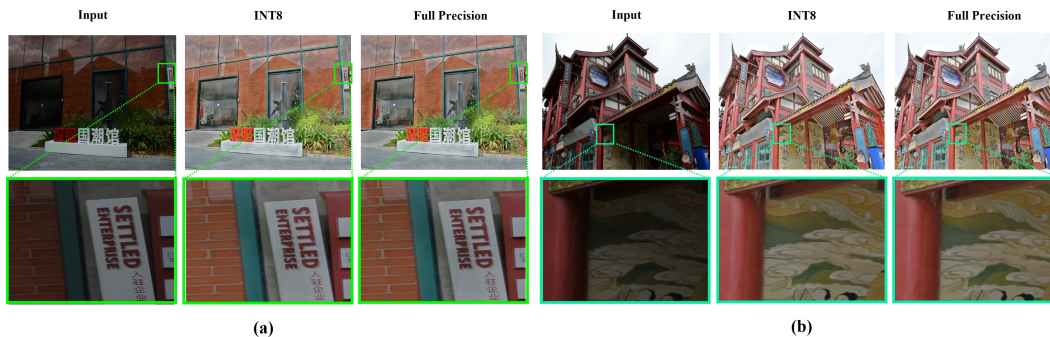

Figure 4. Visual comparison between full-precision LLIE model and its quantized version. In most cases, the quantized version can produce perceptually plausible images similar to its counterparts with a significant acceleration in inference speed.

To make OD efficient on edge devices, we incorporate our optimized LLIE method and SOTA OD to improve its performance on edge devices. We chose YOLOv3 [34] for its simplicity and compatibility with edge devices, enabling efficient object detection without compromising accuracy. Its streamlined architecture and optimization options make it ideal for real-time applications where resource constraints are a concern. Table. 3 illustrates the detection improvement achieved on the SOTA detector in the ExDark [26] dataset. Full-precision and quantized versions can improve the detector's performance by adding a small overhead computation.

Fig. 5 further confirms the effectiveness of our proposed method on OD acceleration in edge devices. Our models, including the INT8 version, substantially enhance the visibility of the images captured in extreme and tricky scenarios. Such perceptual improvement helps the detector extract salient details from the given scenes and improves performance. Such visibility enhancement can accelerate the safety of security and autonomous applications in real-world scenarios.

Figure 5. Accelerating OD performance on edge devices by leveraging proposed optimization framework. (a) Low-light image with YOLOv3 [34] . (b) Our model (INT8) with YOLOv3 [34] . (c) Our model (full precision) with YOLOv3 [34]. (d) Ground-truth (for detection)

Device	GPU	ARM64	ARM64	GPU	ARM64	ARM64
Dimension	$512 \times 512 \times 3$			$256 \times 256 \times 3$		
Precesion	Full		INT8	Full		INT8
PSNR	17.738		16.34	17.31		15.92
SSIM	0.7282		0.6325	0.6719		0.5457
LPIPS	0.3344		0.3506	0.2169		0.2403
Param. (M)	4.46					
Time (ms)	9.33	140.87	18.37	4.4	39.39	5.01
FPS	164.50	7.09	54.42	227.14	25.39	**199.67**

Table 2. Performance of the proposed LLIE method numerous hardware platforms. The proposed method achieved 199 FPS on edge devices by utilizing PTQ.

4.3.2 Image registration

Image registration is a fundamental process in computer vision that involves aligning multiple images of the same scene or object from different viewpoints, angles, or times. This alignment facilitates various applications across diverse domains, including autonomous vehicles, augmented reality, industrial inspection, healthcare, and environmental monitoring. However, image registration can face challenges in low-light conditions, where images may suffer from reduced visibility, noise, or distortion. In this section, we explore the concept of image registration, its wide-ranging applications, and the specific challenges encountered when performing registration in low-light environments, focusing on its feasibility in autonomous driving scenarios.

We incorporated our proposed method for improving image registration in an edge environment. Therefore, we used the SOTA Superglue network [36] on real-world low-light stereo image datasets to perform registration and matching in the edge platform. Due to the lack of visibility, Superglue illustrates a significant deficiency in finding perfect matches on the image pairs. However, our proposed method can improve its performance by significantly enhancing the darker images. Table. 4 illustrates the performance of Superglue on matching-in-the-dark (MID) dataset [43]. Our proposed method significantly improves the performance of the Superglue network. Without LLIE, the Superglue network struggles to identify the key features and perfect match on the stereo images. Our proposed method helps it by improving visibility to achieve more accurate findings while mitigating false matches. On average, our model improves the key point detection on both left and right images. The proposed method helps the Superglue improve its performance by over 90 matches per pair.

Fig. 6 visually confirms the impact of the proposed LLIE on image matching. Our models substantially improve the performance of the SUperglue network. These help the image-matching algorithm minimize false matching and find perfect matches among stereo images.

It is important to note that most algorithms used for matching images are designed to work with grayscale images to simplify the computational complexity of the matching process. Our INT8 model has been optimized for image matching with quantized weights and can perform with the same efficiency level as its full-precision counterpart. The high-performance levels achieved by our model in image matching with quantized weights indicate that it is well-suited for EVA. Using a quantized model for image matching in EVA applications can significantly reduce computational and memory requirements while maintaining the matching process's accuracy and reliability. Hence, our INT8 model emerges as an optimal (EVA) solution, offering efficient and precise image-matching capabilities.

4.4. Discussion

This study reveals a new dimension of LLIE on EVA. Our optimization framework can significantly improve the performance of EVA by adding minor computational overhead. The low computational cost of our proposed network with post-training quantization helps our model achieve an efficient inference speed on edge devices with a decent fertility score. We tested our model on the NTIRE'24 low-light challenge. Despite being significantly lightweight, our model can be on par with the SOTA image restoration models like Uformer, MIRNet, etc. However, the SOTA LLIE method can achieve a better fidelity score than the proposed method. However, complex blocks like transformers make that model impractical for edge devices due to their high

Class Name	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Bike	People	Table	All
ExDark	0.3930	0.1890	0.2570	0.6520	0.4210	0.4140	0.3150	0.2720	0.3960	0.3730	0.3960	0.0152	0.3410
INT8	0.4230	0.2050	0.2300	0.6520	0.4300	0.4180	0.3180	0.2790	0.3960	0.4210	0.4120	0.0150	**0.3500**
Full Precision	0.4240	0.2030	0.2350	0.6590	0.4330	0.4280	0.3270	0.2800	0.4050	0.4320	0.4150	0.0149	**0.3550**

Table 3. Our proposed method accelerates the SOTA object detector in low-light conditions. Despite being significantly faster, our quantized LLIE can also improve the performance of OD in challenging lighting conditions.

Method	Pair	1	2	3	4	5	6	7	8	9	Average
MID [43]	Keypoints	536 : 321	459 : 371	742 : 664	519 : 433	335 : 311	609 : 494	531 : 477	574 : 461	646 : 653	550:465
	Matches	123	204	297	164	191	212	242	195	333	217.89
	Matches / Total	0.1435	0.2458	0.2112	0.1723	0.2957	0.1922	0.2401	0.1884	0.2564	0.2162
INT8	Keypoints	655 : 488	599 : 666	910 : 966	764 : 823	434 : 397	705 : 741	744 : 643	587 : 563	985 : 1024	709:701
	Matches	166	297	407	317	231	302	305	213	506	**304.89**
	Matches / Total	0.1452	0.2348	0.2170	0.1997	0.2780	0.2089	0.2199	0.1852	0.2519	0.2156
Full	Keypoints	655 : 488	599 : 666	910 : 966	764 : 823	434 : 397	705 : 741	744 : 643	587 : 568	985 : 1024	709:702
	Matches	166	297	407	317	231	302	305	213	506	**304.89**
	Matches / Total	0.1452	0.2348	0.2170	0.1997	0.2780	0.2089	0.2199	0.1844	0.2519	0.2155

Table 4. Image registration with Superglue on MID dataset. Our proposed LLIE method significantly improves its performance by enhancing visibility.
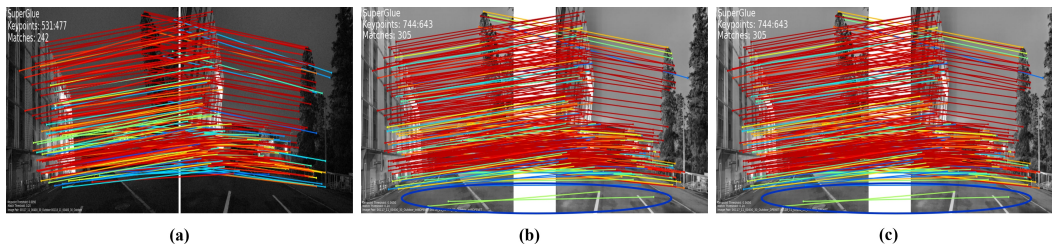

Figure 6. Real-world image registration with Superglue network on MID dataset. Our model(s) significantly improve the performance of the image-matching algorithms. (a) Superglue + Low-light. (b) Superglue + Ours (INT8). (c) Superglue + Ours (full-precision)

complexity and unsupported learning mechanisms.

Our model can show visual artifacts caused by aggressive sensor noise in extremely low-light situations. However, it's important to note that the number of real-world training samples we could access was limited. Despite combining three different single-shot low-light image datasets, our total number of training samples was insufficient for handling real-world extreme low-light scenarios. Additionally, the images used in our study were taken with professional cameras, which have imaging characteristics different from those of compact camera sensors used in edge environments like autonomous driving. Thus, it would be beneficial to create a new LLIE dataset captured with edge camera sensors to improve the performance of future edge LLIE methods.

Besides the data limitation, our quantized model tends to clip color and contrast due to INT8 conversion. Consequently, the model can illustrate on-par performance (e.g., for image matching) with a full-precision model in grayscale images. However, due to the range clipping in RGB space, the quantized model can illustrate performance downgrade (e.g., object detection) in such instances. Notably, the clipping range issues are a very classical problem of PTQ. Therefore, we planned to study the effectiveness of perceptual quantization-aware training for enhancing LLIE in a future study.

5. Conclusion

This study proposes a novel framework for learning optimized LLIE for EVA. Our method incorporates a lightweight deep model, an optimization strategy (e.g., PTQ), and deployment on edge hardware. We compared our proposed method with SOTA LLIE and outperformed them in quantatitive and qualitative comparison. Later, we deployed our method edge device to achieve an FPS of 199. We also illustrated the practicability of such optimized on generic vision tasks. The experimental results demonstrate that the proposed method can significantly improve the performance of SOTA vision algorithms like YOLOv3 and Superglue networks. We planned to collect a real-world LLIE dataset for edge environment and study the quantization-aware training in the foreseeable future.

References

[1] SM A Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 233–242, 2021. 1

[2] Mustafa Ayazoglu. Extremely lightweight quantization robust real-time single-image super resolution for mobile devices. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2472–2479, 2021. 2

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 1

[4] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 1, 2

[5] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014. 3

[6] Z Cui, K Li, L Gu, S Su, P Gao, Z Jiang, Y Qiao, and T Harada. You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. arxiv 2022. *arXiv preprint arXiv:2205.14871*, 238, 2022. 1

[7] Zongcai Du, Jie Liu, Jie Tang, and Gangshan Wu. Anchor-based plain net for mobile image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2494–2502, 2021. 2

[8] Saijie Fan, Wei Liang, Derui Ding, and Hui Yu. Lacn: A lightweight attention-guided convnext network for low-light image enhancement. *Engineering Applications of Artificial Intelligence*, 117:105632, 2023. 1

[9] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22252–22261, 2023. 1

[10] Fei Gao, Teresa Wu, Xianghua Chu, Hyunsoo Yoon, Yanzhe Xu, and Bhavika Patel. Deep residual inception encoder–decoder network for medical imaging synthesis. *IEEE journal of biomedical and health informatics*, 24(1):39–49, 2019. 3

[11] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 5, 6

[12] Cheeun Hong, Heewon Kim, Sungyong Baik, Junghun Oh, and Kyoung Mu Lee. Daq: Channel-wise distribution-aware quantization for deep image super-resolution networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2675–2684, 2022. 2

[13] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 5, 6

[14] Itay Hubara, Yury Nahshan, Yair Hanani, Ron Banner, and Daniel Soudry. Accurate post training quantization with small calibration sets. In *International Conference on Machine Learning*, pages 4466–4475. PMLR, 2021. 4

[15] Rahul Jaiswal, AG Rao, and HP Shukla. Image enhancement techniques based on histogram equalization. *International Journal of Electrical and Electronics Engineering*, 1:69–78, 2010. 2

[16] Dongsheng Jiang, Weiqiang Dou, Luc Vosters, Xiayu Xu, Yue Sun, and Tao Tan. Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network. *Japanese journal of radiology*, 36:566–574, 2018. 5, 6

[17] Xin Jin, Ling-Hao Han, Zhen Li, Chun-Le Guo, Zhi Chai, and Chongyi Li. Dnf: Decouple and feedback network for seeing in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18135–18144, 2023. 1

[18] Wonjun Kim. Low-light image enhancement: A comparative review and prospects. *IEEE Access*, 2022. 1

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[20] Furkan Kınlı, Sami Menteş, Barış Özcan, Furkan Kıraç, Radu Timofte, Yi Zuo, Zitao Wang, Xiaowen Zhang, Yu Zhu, Chenghua Li, et al. Aim 2022 challenge on instagram filter removal: methods and results. In *European Conference on Computer Vision*, pages 27–43. Springer, 2022. 3

[21] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 1, 2

[22] Huixia Li, Chenqian Yan, Shaohui Lin, Xiawu Zheng, Baochang Zhang, Fan Yang, and Rongrong Ji. Pams: Quantized super-resolution via parameterized max scale. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 564–580. Springer, 2020. 2

[23] Xiaokai Liu, Weihao Ma, Xiaorui Ma, and Jie Wang. Laenet: a locally-adaptive embedding network for low-light image enhancement. *Pattern Recognition*, 133:109039, 2023. 1

[24] Xiaoning Liu, Zongwei Wu, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, and Radu Timofte. NTIRE 2024 challenge on low light enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024. 4

[25] Zhenhua Liu, Yunhe Wang, Kai Han, Wei Zhang, Siwei Ma, and Wen Gao. Post-training quantization for vision transformer. *Advances in Neural Information Processing Systems*, 34:28092–28103, 2021. 1, 4

[26] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 6

[27] Nungsanginla Longkumer, Mukesh Kumar, and Rohini Saxena. Contrast enhancement techniques using histogram equalization: a survey. *International Journal of Current Engineering and Technology*, 4(3):1561–1565, 2014. 2

[28] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 1

[29] Ziwei Luo, Youwei Li, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, and Shuaicheng Liu. Fast nearest convolution for real-time efficient image super-resolution. In *European conference on computer vision*, pages 561–572. Springer, 2022. 2

[30] NVIDIA Corporation. TensorRT, Accessed: 2024. 5

[31] NVIDIA Corporation. TensorRT PyTorch Quantization Toolkit Documentation. `https://docs.nvidia.com/deeplearning/tensorrt/pytorch-quantization-toolkit/docs/`, Accessed: 2024. Accessed: April 6, 2024. 4

[32] NVIDIA Corporation. Jetson Orin NX Series and Jetson AGX Orin Series, Accessed: 2024. 5

[33] Pytorch. PyTorch Framework code. `https://pytorch.org/`, 2016. Accessed: 2020-08-24. 4

[34] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 6, 7

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 1, 5, 6

[36] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020. 7

[37] De Rosal Igantius Moses Setiadi. Psnr vs ssim: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6):8423–8444, 2021. 5

[38] SMA Sharif, Rizwan Ali Naqvi, Farman Ali, and Mithun Biswas. Darkdeblur: Learning single-shot image deblurring in low-light condition. *Expert Systems with Applications*, 222:119739, 2023. 1, 3

[39] SMA Sharif, Rizwan Ali Naqvi, and Woong-Kee Loh. Two-stage deep denoising with self-guided noise attention for multimodal medical images. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2024. 3

[40] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 1

[41] Kavinder Singh and Anil Singh Parihar. Illumination estimation for nature preserving low-light image enhancement. *The Visual Computer*, 40(1):121–136, 2024. 1

[42] Sushant Singh and Ausif Mahmood. The nlp cookbook: modern recipes for transformer based deep learning architectures. *IEEE Access*, 9:68675–68702, 2021. 1

[43] Wenzheng Song, Masanori Suganuma, Xing Liu, Noriyuki Shimobayashi, Daisuke Maruta, and Takayuki Okatani. Matching in the dark: A dataset for matching image pairs of low-light scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6029–6038, 2021. 7, 8

[44] Li Tao, Chuang Zhu, Guoqing Xiang, Yuan Li, Huizhu Jia, and Xiaodong Xie. Llcnn: A convolutional neural network for low-light image enhancement. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2017. 1

[45] Zhijun Tu, Jie Hu, Hanting Chen, and Yunhe Wang. Toward accurate post-training quantization for image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5856–5865, 2023. 2, 4

[46] Hu Wang, Peng Chen, Bohan Zhuang, and Chunhua Shen. Fully quantized image super-resolution networks. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 639–647, 2021. 2

[47] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 1

[48] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 1

[49] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2, 4, 5, 6

[50] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2022. 1, 2

[51] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 1

[52] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9893–9903, 2023. 1

[53] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 4

[54] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023. 1, 2

[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling

Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 1

[56] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 1

[57] Qing Zhang, Yongwei Nie, and Wei-Shi Zheng. Dual illumination estimation for robust exposure correction. In *Computer graphics forum*, pages 243–252. Wiley Online Library, 2019. 1

[58] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 5

[59] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. 1

[60] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1076–1088, 2021. 1

[61] Yunshan Zhong, Mingbao Lin, Xunchao Li, Ke Li, Yunhang Shen, Fei Chao, Yongjian Wu, and Rongrong Ji. Dynamic dual trainable bounds for ultra-low precision super-resolution networks. In *European Conference on Computer Vision*, pages 1–18. Springer, 2022. 2

[62] Qihua Zhou, Song Guo, Zhihao Qu, Jingcai Guo, Zhenda Xu, Jiewei Zhang, Tao Guo, Boyuan Luo, and Jingren Zhou. Octo:{INT8} training with loss-aware compensation and backward quantization for tiny on-device learning. In *2021 USENIX Annual Technical Conference (USENIX ATC 21)*, pages 177–191, 2021. 4