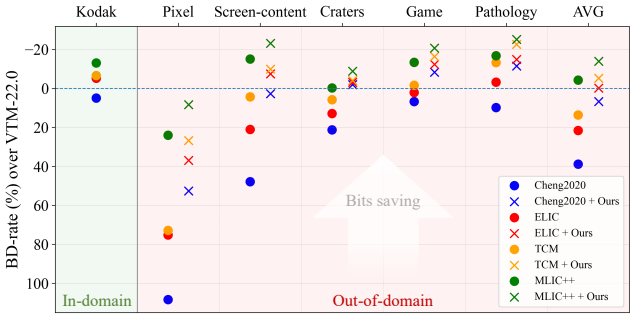
Few-Shot Domain Adaptation for Learned Image Compression

Tianyu Zhang, Haotian Zhang, Yuqi Li, Li Li, Dong Liu*

University of Science and Technology of China
{zhangtianyu,zhanghaotian,lyq010303}@mail.ustc.edu.cn, {lil1,dongeliu}@ustc.edu.cn

Abstract

Learned image compression (LIC) has achieved state-of-the-art rate-distortion performance, deemed promising for next-generation image compression techniques. However, pre-trained LIC models usually suffer from significant performance degradation when applied to out-of-training-domain images, implying their poor generalization capabilities. To tackle this problem, we propose a few-shot domain adaptation method for LIC by integrating plug-and-play adapters into pre-trained models. Drawing inspiration from the analogy between latent channels and frequency components, we examine domain gaps in LIC and observe that out-of-training-domain images disrupt pre-trained channel-wise decomposition. Consequently, we introduce a method for channel-wise re-allocation using convolution-based adapters and low-rank adapters, which are lightweight and compatible to mainstream LIC schemes. Extensive experiments across multiple domains and multiple representative LIC schemes demonstrate that our method significantly enhances pre-trained models, achieving comparable performance to H.266/VVC intra coding with merely 25 target-domain samples. Additionally, our method matches the performance of full-model finetune while transmitting fewer than 2% of the parameters.



(a) Performance of LIC models and our adaptation on different domains, compared with VTM

(b) Natural Images (c) Pixel (d) Screen-content (e) Craters

Figure 1: (a) BD-rate (↓) of four advanced LIC models with or without our method on different domains. (b)-(e) Images from different domains have visible different characteristics.

Introduction

With the rapid growth of images in multimedia, image compression has become essential for efficient storage and transmission. Over the past decade, conventional methods including JPEG (Wallace 1991), H.265/HEVC (Sullivan et al. 2012) and H.266/VVC (Bross et al. 2021), have significantly contributed to the real-world applications of image compression. Recently, learned image compression (LIC) has achieved impressive progress in rate-distortion (RD) performance using non-linear transforms and end-to-end optimization. Several studies (He et al. 2022; Jiang and Wang 2023) have even surpassed VTM (the reference software for H.266/VVC) on both PSNR and MS-SSIM, indicating LIC's potential for future image compression techniques.

Despite the encouraging progress, some studies (Shen, Yue, and Yang 2023; Lv et al. 2023) have pointed out the performance degradation of pre-trained LIC models on out-of-domain images (we simplify in/out-of-training-domain as in/out-of-domain for convenience). It is common practice to train and evaluate LIC models on natural image benchmarks (Franzen 1993; Agustsson and Timofte 2017; Asuni and Giachetti 2014; CLIC 2020; Li et al. 2024b). However, practical codecs may well be challenged by images from various domains, many of which have significant domain gaps compared to natural images. For demonstration purposes, we test advanced LIC models including Cheng2020 (Cheng et al. 2020), ELIC (He et al. 2022), TCM (Liu, Sun, and Katto 2023), and MLIC++ (Jiang and Wang 2023) in Fig. 1. While all these models perform comparably or better than the conventional codec VTM on natural images, they exhibit varying degrees of performance drop and generally fall behind VTM on out-of-domain images such as pixel-style art and screen content, highlighting the necessity for efficient adaptation methods for pre-trained LIC models.

In this regard, most existing methods (Shen, Yue, and Yang 2023; Lv et al. 2023; Tsubota, Akutsu, and Aizawa 2023) suggest instance adaptation (IA), which aims to adapt LIC models to a single image at inference. Despite its effectiveness, IA does not perform actual domain adaptation

*Corresponding Author

(DA) for pre-trained models, and is time-consuming and computationally expensive due to its per-image online training. On the other hand, Katakol et al. (2021) proposed DA for LIC using limited target samples. They suggested selective finetune, and adapted pre-trained models to a target domain with one-time training and transmission, demonstrating great adaptation efficiency for deployed models. Nevertheless, the proposed method is restricted to GDN structures (Ballé, Laparra, and Simoncelli 2015), and suffers from large additional parameters on recent schemes (He et al. 2022; Liu, Sun, and Katto 2023; Jiang et al. 2023).

Despite the promising applications like codec calibration for batch transmission, DA for LIC remains largely unexplored. Recently, Presta et al. (2024) pre-trained a gate network on multiple domains, enhancing the generalization capacity of LIC models. However, they did not address the problem of adapting LIC models to a specific domain. Therefore, in this paper, we focus on **adapting pre-trained LIC models to a target domain**. For practicality, we follow Katakol et al. (2021) and study DA for LIC with limited samples. We also highlight the versatility of the DA method to different domains and mainstream LIC schemes.

Building on these insights, we propose a universal few-shot domain adaptation method for LIC using compact adapters, achieving superior performance across various domains and mainstream LIC schemes. Inspired by (Li et al. 2024a), we explore domain gaps in LIC through disturbed channel-wise decomposition. We demonstrate that pre-trained LIC models suffer from scattered channel-wise energy allocation on out-of-domain images. To address this problem, we introduce convolution-based adapters (Conv-Adapters) and low-rank adapters (LoRA-Adapters) to the pre-trained models for channel-wise re-allocation. These adapters are trained with a few target samples, and are plug-and-play for processing specific domains.

Our main contributions can be summarized as follows:

- We propose a universal few-shot domain adaptation method for learned image compression by incorporating compact adapters. Our method is applicable to different domains and mainstream LIC schemes.
- We explore the domain gaps in LIC through disturbed channel-wise decomposition, which results in scattered energy allocation on out-of-domain images. The proposed channel-wise re-allocation with adapters strengthens energy compaction.
- Extensive experiments verify the effectiveness of our method across mainstream domains and LIC schemes. We generally enhance pre-trained models to VTM on every domain, achieving comparable improvements to full-model finetune with tiny additional parameters.

Related Work

Learned Image Compression

Learned image compression has demonstrated competitive potential against traditional codecs. Ballé, Laparra, and Simoncelli (2016) proposed the first end-to-end LIC model with non-linear transforms and uniform quantizer,

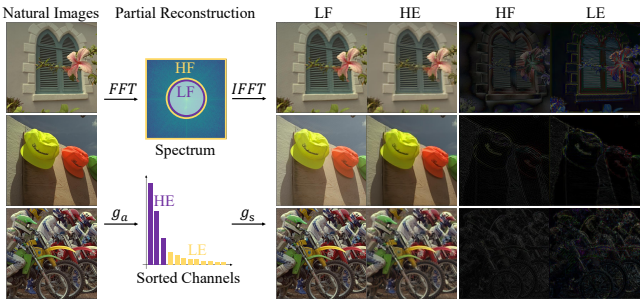


Figure 2: Analogy between frequency components and channels of LIC latents. For one image, we perform Fast Fourier Transform (FFT) and then reconstruct from low-frequency (LF) or high-frequency (HF) components respectively. Similarly, we perform a learned analysis transform (g_a) and then reconstruct from high-energy (HE) or low-energy (LE) channels respectively. Please view on screen and zoom in to observe the reconstructions from HF/LE.

which was further strengthened by hyperprior (Ballé et al. 2018) and context model (Minnen, Ballé, and Toderici 2018). Cheng et al. (2020) first achieved comparable performance with VTM using Gaussian mixture likelihoods. He et al. (2022) proposed uneven channel autoregressive model, demonstrating pleasant performance and complexity. Recently, Liu, Sun, and Katto (2023) designed Transformer-CNN mixture block to incorporate the ability of both structures for LIC. Jiang et al. (2023) introduced multi-reference context to capture different correlations, achieving the state-of-the-art performance (Jiang and Wang 2023).

In the regard of interpretability, early work (Ballé et al. 2018; Cheng et al. 2020) visualized the latent in LIC to evaluate the preciseness of the entropy model. He et al. (2022) sorted latent channels by energy, suggesting low-frequency components concentrated on a few channels while the rest was extremely sparse. Duan et al. (2022) interpreted latent channels as orthogonal transform coefficients, while Li et al. (2024a) suggested frequency decomposition is an intrinsic property for LIC. Nevertheless, these results on latent representation did not fully consider out-of-domain images.

Adaptation for Learned Image Compression

Two kinds of adaptation namely instance adaptation (IA) and domain adaptation (DA) have been studied for LIC. IA (Campos et al. 2019; Van Rozendaal et al. 2021; Tsubota, Akutsu, and Aizawa 2023; Shen, Yue, and Yang 2023; Lv et al. 2023) adapts pre-trained models to a single image rather than a domain using per-image online training. Despite its effectiveness, IA is computationally expensive and time-consuming as shown in **Appendix A**. Contrarily, Katakol et al. (2021) introduced the problem of DA for LIC, adapting pre-trained LIC models to a target domain with one-time training. They suggested finetuning GDN layers with channel-wise parameters and the entire entropy model, which was restricted by specific structures, and additional parameters due to the complex entropy models. We alleviate this issue by designing compact and universal adapters.

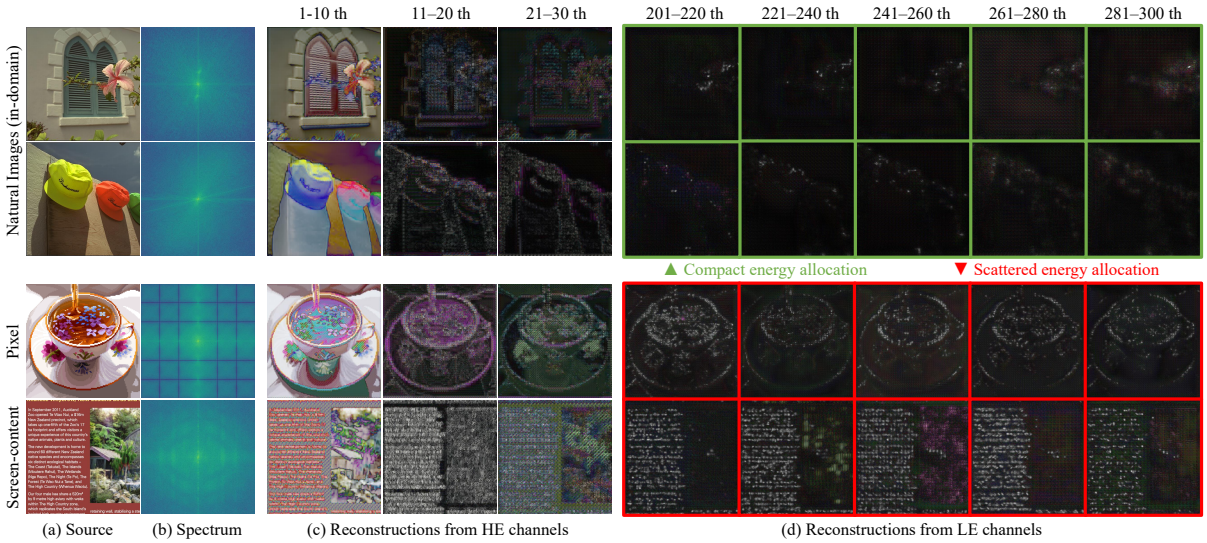


Figure 3: Following Fig. 2, we explore the domain gaps by observing the channel-wise decomposition. The top two rows display in-domain natural images, while the bottom two rows show out-of-domain images. (a) Source image. (b) Spectrum (using FFT). (c, d) Reconstructions from different HE channels and LE channels respectively. Here the total number of channels is 320. *Out-of-domain images have more HF components as shown in (b), as well as more information embedded into LE channels as shown in (d). Thus, our key idea is to re-allocate the information from LE channels to HE channels for out-of-domain images.*

Method

Motivation

Recent studies (Duan et al. 2022; Li et al. 2024a) suggest that the transform in LIC can be interpreted as channel-wise frequency decomposition in a non-linear fashion. We compare latent channels to frequency components of a source image using linear transforms (e.g. FFT), as reconstructions from partial channels or partial frequency components demonstrate obvious similarities in Fig. 2. Motivated by this, we further study the channel-wise decomposition on different domains by reconstructions from specific channels in Fig. 3 (c) and (d). Consistent with previous findings (He et al. 2022; Duan et al. 2022; Li et al. 2024a), decomposed information concentrates on high-energy channels (respectively low-frequency). For in-domain images, the energy allocation on channels is more compact, with little information of the source images found in low-energy channels. However, for out-of-domain images, more high-frequency components are presented in the spectrum (Fig. 3 (b)), and obvious contours of source images can be observed in reconstructions from low-energy channels, as shown in Fig. 3 (d). This indicates a scattered energy allocation of the pre-trained LIC model on out-of-domain images, which differs from the compact energy allocation on in-domain images.

Given these channel-wise differences among latents of in-domain and out-of-domain images, we assume that the following entropy estimation and reconstruction modules can not operate effectively in the pre-trained manner due to improper latents, leading to degraded rate-distortion performance. Therefore, our key idea is to perform channel-wise re-allocation, which concentrates the information from LE channels to HE channels for out-of-domain images. Specifically, we propose gently re-allocating the channels of intermediate latents by inserting compact adapters into the pre-trained models. The implementation and performance of these adapters are detailed in the following sections.

Channel-wise Re-allocation with Adapters

We present the deployment of our method on ELIC (He et al. 2022) in Fig. 4. Two types of adapters, namely Conv-Adapters and LoRA-Adapters, are introduced. To refine the channel-wise decomposition in the transform, we insert Conv-Adapters after non-linear blocks. For entropy estimation, LoRA-Adapters are applied to the entropy parameters network. We highlight the compatibility of the proposed method to mainstream LIC schemes.

Conv-Adapters We introduce Conv-Adapters to the transform to adjust the channels of intermediate latents. Denoting the output latent of a specific block as $L \in \mathbb{R}^{c \times h \times w}$ and its ith channel $L_i \in \mathbb{R}^{h \times w}$, we aim to re-allocate the information on channels before passed to the next block:

$$L' = (L'_0, L'_1, ..., L'_{c-1}) = f(L_0, L_1, ..., L_{c-1}) \quad (1)$$

where L' denotes the output latent with refined information allocation on channels, and f denotes the re-allocation mechanism introduced by Conv-Adapters. Considering the restriction on computational complexity and limited samples, we simplify this mechanism into a channel-wise linear transform, and apply Conv1×1 as the Conv-Adapter. To maintain the training stability, we initialize the weight matrix $W \in \mathbb{R}^{c \times c \times 1 \times 1}$ of Conv1×1 to identity matrix $I \in \mathbb{R}^{c \times c \times 1 \times 1}$, and the bias $b \in \mathbb{R}^c$ to zeros. Therefore, we refine intermediate latents in the transform by the following linear re-allocation:

$$L' = W \cdot (L_0, L_1, ..., L_{c-1})^T + b \quad (2)$$

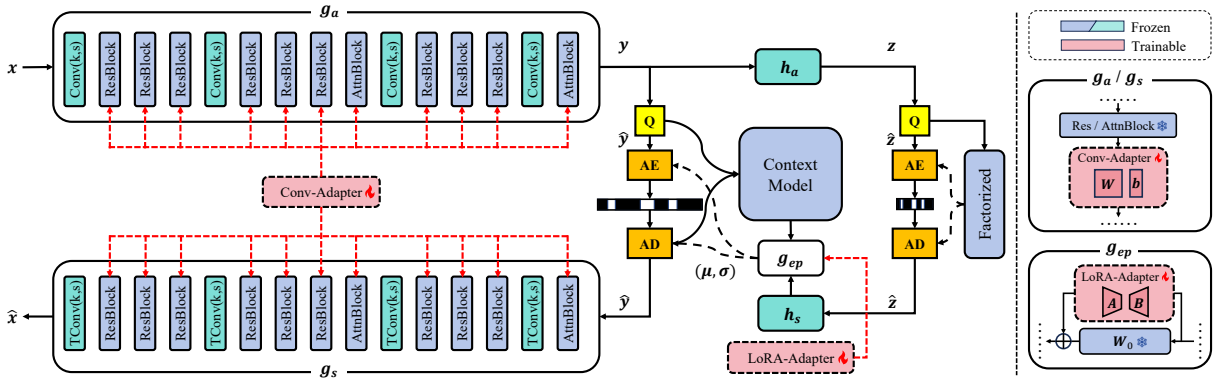


Figure 4: Deployment of our method on ELIC (He et al. 2022). We denote g_{ep} as the entropy parameters network, while the other notations follow the same explanations in (He et al. 2022). As detailed on the right, Conv-Adapters are inserted serially after non-linear blocks in the transform, while LoRA-Adapters are added to the pre-trained weight matrix W_0 in g_{ep}. W and b are weight and bias of Conv-Adapter, respectively, while A and B are low-rank matrices. Only adapters are trainable.

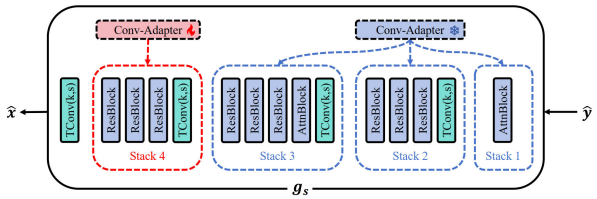


Figure 5: Division of g_s into four stacks. In the second stage, we only finetune adapters in Stack 4 for reconstruction.

LoRA-Adapters To accommodate the channel-wise reallocation in the transform, we adopt Low-Rank Adaptation (LoRA) (Hu et al. 2021; Lv et al. 2023) in the entropy model. Mainstream LIC schemes have large variances on the entropy model, resulting in difficulty for a universal adaptation method on entropy estimation. Besides, hidden dimensions in the entropy model are empirically large for better performances, indicating convolution-based adapters will boost the amount of additional parameters. Therefore, we introduce low-rank matrices (LoRA-Adapters) to the common entropy parameters network. Namely, we modify the weight of a pre-trained layer $W_0 \in \mathbb{R}^{c_{out} \times c_{in} \times h \times w}$ by:

$$\begin{aligned} W_0' &= W_0 + \Delta W \\ \Delta W &= BA \end{aligned} \tag{3}$$

where $A \in \mathbb{R}^{r \times c_{in}}$ is initialized with Gaussian, $B \in \mathbb{R}^{c_{out} \times r}$ is set to zeros, ensuring $\Delta W = 0$ thus the training stability. The rank $r \ll \min\{c_{in}, c_{out}\}$, and is set to 10 empirically. LoRA-Adapters serve as bias on channel dimension for pre-trained weights, and can be merged once trained without introducing additional latency at inference.

Two-Stage Training Strategy

The training of adapters consists of two stages, as detailed in Fig. 4 and Fig. 5 respectively. We denote $\theta = \{W, b, A, B\}$ as the total parameters of adapters. Given the analysis transform g_a, synthesis transform g_s, hyper analysis transform h_a, uniform scalar quantization Q, and λ to balance bit-rate $\mathcal{R}$ and distortion $\mathcal{D}$, we first freeze all pre-trained parameters ϕ and train adapters θ jointly with:

$$\mathcal{L}(\theta) = \mathcal{R}(\hat{y}) + \mathcal{R}(\hat{z}) + \lambda \cdot \mathcal{D}(x, \hat{x}) \tag{4}$$

where:

$$\begin{aligned} \hat{y} &= Q(g_a(x; \phi, \theta)) \\ \hat{x} &= g_s(\hat{y}; \phi, \theta) \\ \hat{z} &= Q(h_a(y; \phi)) \end{aligned} \tag{5}$$

We suggest adapters optimized by a joint objective function suffer from insufficient training under limited samples. Inspired by the two-stage training strategy in (Guo et al. 2021), we further conduct a second training stage. Concretely, we divide g_s into four stacks by upsample operations, as shown in Fig. 5. We freeze all the adapters except the ones in Stack 4, and finetune this small set (denoted as $\Delta\theta$) with hard quantization and distortion only:

$$\mathcal{L}(\Delta\theta) = \mathcal{D}(x, \hat{x}), \text{switch } Q \text{ to } rounding. \tag{6}$$

We demonstrate in Section 4.3 that this stage is simple but effective without introducing more parameters. Once trained, adapters in g_a are stored at the encoder, while the rest should be transmitted to the decoder. In this way, a full set of adapters corresponds to a certain domain, and is plug-and-play for flexibly processing specific domains.

Experiments

Settings

Domains and Datasets. We follow Lv et al. (2023); Shen, Yue, and Yang (2023) and sort five domains with 150 images each, including (i) pixel-style images (Lv et al. 2023); (ii) screen content images from SCID (Ni et al. 2017), SCI1K (Yang et al. 2021), SIQAD (Yang, Fang, and Lin 2015); (iii) craters images cropped from Lunar Reconnaissance Orbiter Camera (LROC) (Robinson et al. 2010); (iv) game images cropped from GamingVideoSET (Barman et al. 2018); (v) pathology images cropped from BRACS (Brancati et al. 2022). All crops have the same resolution of 600×800. We

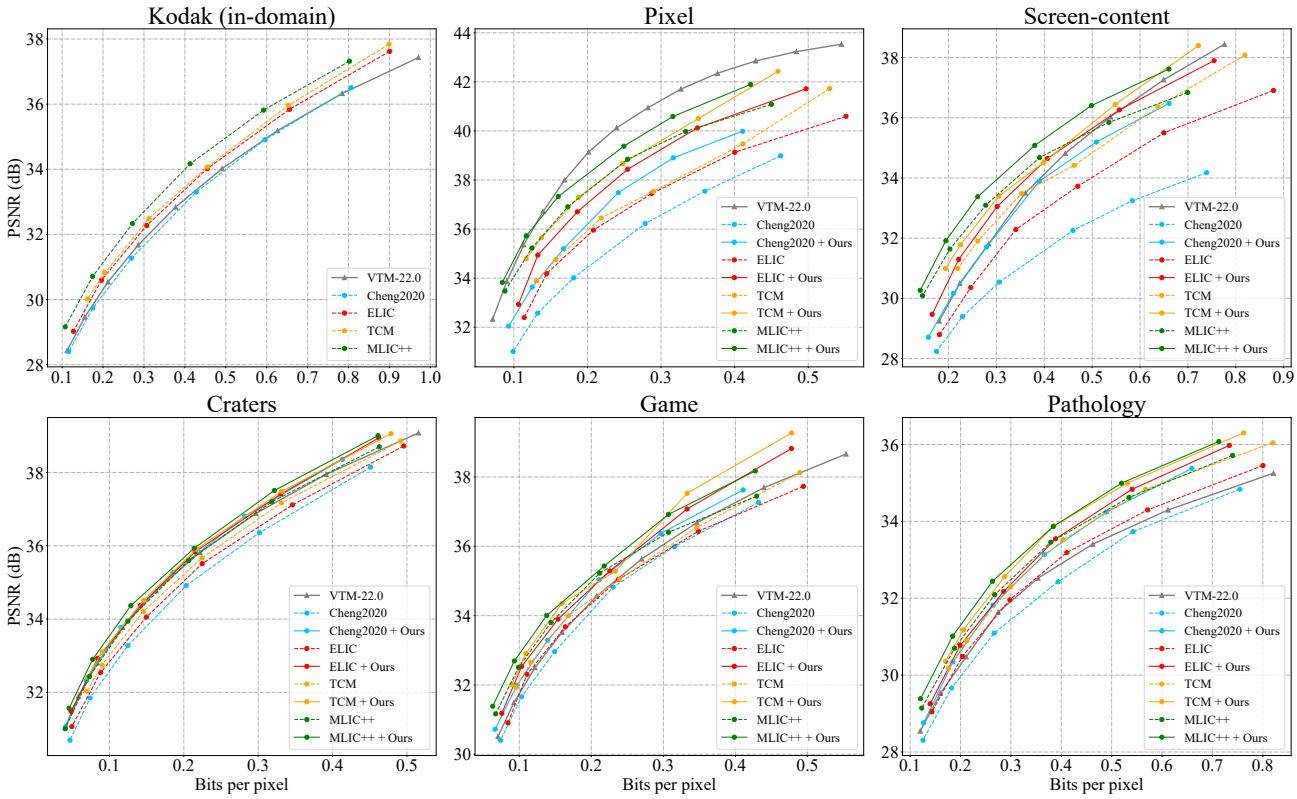


Figure 6: RD curves including VTM-22.0 intra coding, pre-trained LIC models, and models adapted with 25 samples in our method. We generally enhance pre-trained models to the level of VTM on the other five domains.

select a fixed test set with 100 images for each domain. We use Kodak (Franzen 1993) for the natural image domain.

Models. We test our method on advanced LIC models, namely Cheng2020 (Cheng et al. 2020), ELIC (He et al. 2022), MLIC++ (Jiang and Wang 2023) and TCM (Liu, Sun, and Katto 2023). We use the official pre-trained weights except ELIC, for which we train ourselves for a lack of official weights. More details can be found in **Appendix F**.

Implementation. We train adapters with $N = \{5, 10, 25\}$ samples, respectively. We split the samples into training and validation set in a proportion of 4:1. We use a patch size of 256 and a batch size of 4. In the first stage, we set learning rate stages $\{50, 10, 7.5, 5, 2.5, 1\} \times 10^{-5}$. For different N, we further set max epoch $= \{250, 500, 750\}$. In the second stage, we train selected adapters for another max epoch with fixed learning rate 5×10^{-4}.

Adaptation Performance

Comparisons with Traditional Codec We compare the proposed method with the conventional standard VVC by VTM-22.0 intra coding. We plot the rate-distortion (RD) curves on Kodak (in-domain) and the other five domains in Fig. 6. All pre-trained models perform comparably or better than VTM on Kodak, but most of them fall behind VTM on the other five domains. Our method enhances these models to the level of VTM except on the Pixel domain, where the domain gap is too wide to fully compensate (Lv et al. 2023).

Baseline	Params	Method	Transmit	Prop.
Cheng2020	31.48M	Finetune	24.12M	76.62%
		DANICE	10.86M	34.50%
		Ours	**0.37M**	**1.17%**
ELIC	40.72M	Finetune	30.98M	76.08%
		DANICE	/	/
		Ours	**0.71M**	**1.74%**
TCM	45.18M	Finetune	41.93M	92.81%
		DANICE	35.17M	77.84%
		Ours	**0.88M**	**1.95%**
MLIC++	116.72M	Finetune	110.45M	94.71%
		DANICE	98.41M	84.31%
		Ours	**0.73M**	**0.63%**

Table 1: The number of additional parameters to transmit compared with full-model finetune and DANICE (Katakol et al. 2021), Note that DANICE can not be applied to ELIC.

Comparisons with Domain Adaptation Method To evaluate the efficiency, we also compare with existing DA method DANICE (Katakol et al. 2021) and full-model finetune by RD performance and complexity. We demonstrate detailed adaptation results under different N by computing BD-rate (Bjontegaard 2001) in Table 2. It is noted that DANICE is restricted to GDN structures (Ballé, Laparra, and Simoncelli 2015), making it inapplicable to baselines like

N	Baseline	Method	Pixel	Screen-content	Craters	Game	Pathology	AVG
5	Cheng2020	Finetune	**-23.09%**	**-25.51%**	4.35%	**-10.69%**	**-16.49%**	-14.29%
		DANICE	-4.92%	-4.48%	-7.31%	-2.65%	-5.62%	-5.00%
		Ours	-21.95%	-23.99%	**-8.55%**	-8.93%	-13.58%	**-15.40%**
	ELIC	Finetune	-15.82%	-15.82%	-3.26%	-2.73%	-9.57%	-9.44%
		DANICE	/	/	/	/	/	/
		Ours	**-17.36%**	**-16.27%**	**-6.07%**	**-4.78%**	**-10.70%**	**-11.04%**
10	Cheng2020	Finetune	**-27.76%**	**-28.42%**	**-16.79%**	**-13.48%**	-17.48%	**-20.79%**
		DANICE	-8.19%	-7.78%	-10.88%	-4.50%	-7.14%	-7.70%
		Ours	-27.44%	-28.00%	-15.25%	-11.34%	**-17.78%**	-19.96%
	ELIC	Finetune	-20.84%	-21.92%	**-13.06%**	**-9.25%**	-10.41%	**-15.10%**
		DANICE	/	/	/	/	/	/
		Ours	**-21.38%**	**-22.35%**	-10.64%	-6.32%	**-11.87%**	-14.51%
25	Cheng2020	Finetune	**-30.37%**	**-34.99%**	**-22.83%**	**-19.66%**	-17.72%	**-25.11%**
		DANICE	-11.30%	-11.59%	-13.58%	-6.88%	-9.83%	-10.64%
		Ours	-27.85%	-31.10%	-19.12%	-14.20%	**-18.52%**	-22.16%
	ELIC	Finetune	**-24.63%**	**-25.62%**	**-17.15%**	**-15.95%**	-10.76%	**-18.82%**
		DANICE	/	/	/	/	/	/
		Ours	-23.35%	-24.95%	-13.77%	-13.83%	**-12.29%**	-17.64%
	TCM	Finetune	**-28.76%**	**-15.53%**	**-13.16%**	-12.03%	-10.21%	**-15.94%**
		DANICE	-4.12%	-4.81%	-8.52%	2.30%	-4.34%	-3.90%
		Ours	-27.82%	-13.81%	-8.96%	**-14.86%**	**-10.71%**	-15.24%
	MLIC++	Finetune	-12.97%	-9.78%	**-11.82%**	**-10.61%**	-9.35%	-10.91%
		DANICE	-0.61%	-0.79%	-7.15%	-0.01%	-4.90%	-2.69%
		Ours	**-14.63%**	**-12.09%**	-8.50%	-9.70%	**-10.27%**	**-11.04%**

Table 2: Detailed BD-rate (↓) using different numbers of target samples (N) for domain adaptation. We compare our method with full-model finetune and DANICE. The anchors are corresponding pre-trained models. Note that DANICE can not be applied to ELIC as GDN is not used. Please refer to the **Appendix B** for more results when $N = \{5, 10\}$.

Stage 1		Stage 2		Pixel	Screen-content	Craters	Game	Pathology	AVG
Conv	LoRA	Stack 1234	Stack 4						
✓				-19.38%	-23.73%	-10.84%	-12.73%	-10.30%	-15.40%
✓	✓			-20.40%	-24.28%	-12.57%	-13.40%	-11.50%	-16.43%
✓	✓	✓		-19.92%	-22.46%	-11.95%	-10.68%	-10.95%	-15.19%
✓	✓		✓	**-23.35%**	**-24.95%**	**-13.77%**	**-13.83%**	**-12.29%**	**-17.64%**

Table 3: Ablation studies conducted on ELIC adapted with 25 samples. The anchor of BD-rate (↓) is pre-trained ELIC models.

ELIC. Additionally, DANICE fails to achieve domain adaptation in some conditions, such as TCM and MLIC++ on the Game domain, indicating a limited versatility. Our method, however, achieves excellent adaptation performance in all conditions, performing comparably with full-model finetune and significantly outperforming DANICE.

In the regard of complexity, we highlight the compactness of adapters in Table 1. The proposed method only needs to transmit fewer than 2% of the total parameters, while much more parameters need considering in DANICE and finetune due to the complex entropy model. Besides, the proposed method costs only a tiny increase on the FLOPs and the decoding time, while enjoys a faster training speed than finetune as suggested in **Appendix C**.

Ablation Studies

We conduct ablation studies to evaluate the contributions of Conv-Adapters, LoRA-Adapters and the two-stage training strategy. Using pre-trained ELIC as the baseline, we perform the proposed adaptation with 25 samples on the five domains. The results are shown in Table 3. Conv-Adapters achieve an average reduction of 15.40% in BD-rate compared to pre-trained models, while LoRA-Adapters and the two-stage training strategy each contributes an additional 1% BD-rate reduction. We also compare our second stage strategy with finetuning all Conv-Adapters in the decoder and display the superiority of our method.

Given that Conv-Adapter, which involves linear re-allocation on channels, has the most significant impact on adaptation performance, we carry out further explorations on its structure and deployment in **Appendix D**, and demonstrate the efficiency of the proposed method on both adaptation performance and additional parameters.

Analyses on Channel-wise Re-allocation

We propose channel-wise re-allocation with adapters in the pre-trained LIC model for domain adaptation. We get motivated by the observation that the pre-trained channel-wise

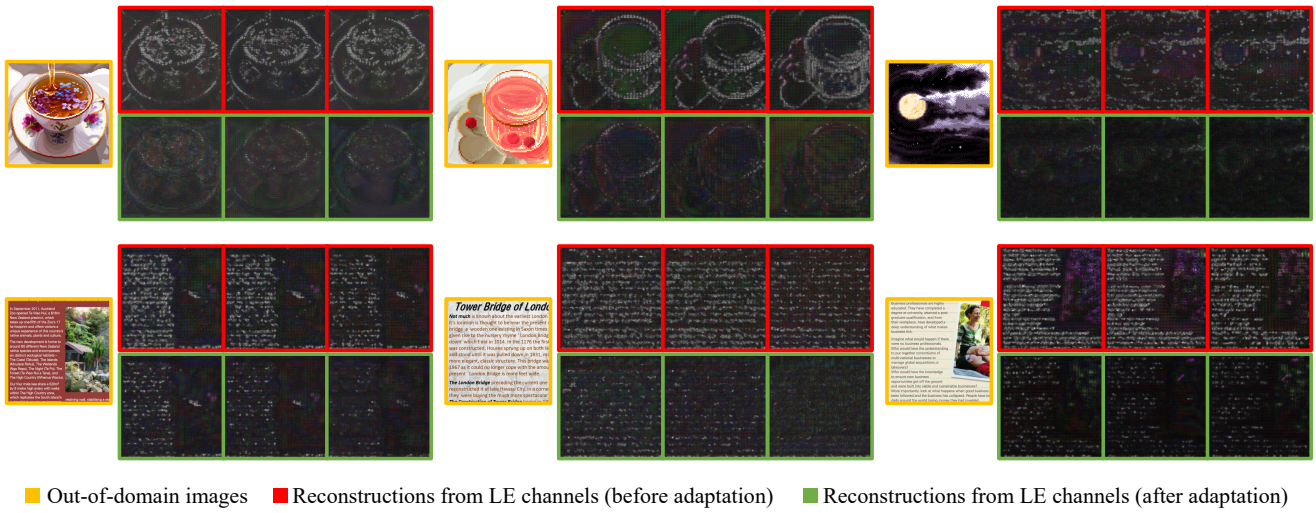


Figure 7: Following Fig. 3, we display reconstructions from low-energy (LE) channels on out-of-domain images. The top row is Pixel, and the bottom row is Screen-content. As observed, our method efficiently reduces the information from LE channels.

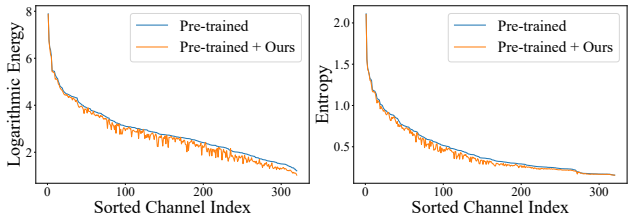


Figure 8: Distributions of energy and entropy on channels. The channels are sorted by the distributions of pre-trained models. The results are averaged on the Pixel domain.

decomposition is disturbed by out-of-domain images, resulting in scattered energy allocation across channels. To highlight this issue and demonstrate the effectiveness of our method, we perform detailed analyses focusing on the channels of latent representation. We conduct experiments on pre-trained ELIC models. We perform the proposed method on the Pixel domain using 25 target samples.

Variation on Channels We use the output latent of the analysis transform, and study the distribution of logarithmic energy and entropy on channels following (He et al. 2022) in Fig. 8. The results are averaged on the Pixel domain. We find that both energy and entropy on specific channels drop after adapting with adapters. Especially, the variation is more apparent on low-energy channels, while high-energy channels remain almost unchanged. This suggests that the energy becomes more concentrated to high-energy channels on out-of-domain images.

Reconstructions from Low-Energy Channels To further investigate the variation on channel-wise decomposition, we compare the reconstructions from low-energy channels before and after adaptation. Concretely, we keep specific channels unchanged and mask the rest into zeros, reconstruct the modified latent, then subtract the reconstruction of all-zero one. We sort the channels by energy, and display the reconstructions from specific ranges of channels after $200th$ in Fig. 7, which is similar to Fig. 3. We find weaker details of inputs can be observed in reconstructions from the model adapted with our method, suggesting high-frequency components have been better moved out of the low-energy channels, which is more in line with the compact energy allocation in processing in-domain images. Thus, we suggest the pre-trained frequency decomposition on high-frequency components is strengthened by our method. More reconstructions of images from different domains are shown in **Appendix E**.

Conclusion

In this paper, we propose a universal few-shot domain adaptation method for learned image compression. With pleasant RD improvement and limited complexity, we expand the proposed method to multiple domains and multiple representative LIC schemes, demonstrating its versatility. We explore the domain gaps in LIC through disturbed channel-wise decomposition, and introduce compact adapters for channel-wise re-allocation. With our method the domain adaptation of LIC models can be simplified into lightweight training, transmission, and storage of compact adapters, achieving flexibility and efficiency. We conduct sufficient experiments showing the proposed method enhances pre-trained models to the level of VTM generally, and significantly outperforms existing domain adaptation method. We hope our findings provide inspiration for future research on domain adaptation for LIC, thus contributing to the real-world deployment of versatile learned image compression.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant 62036005. We acknowledge the

support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

Agustsson, E.; and Timofte, R. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 126–135.

Asuni, N.; and Giachetti, A. 2014. TESTIMAGES: a Large-scale Archive for Testing Visual Devices and Basic Image Processing Algorithms. In *STAG: Smart Tools & Apps for Graphics*, 63–70.

Ballé, J.; Laparra, V.; and Simoncelli, E. P. 2015. Density modeling of images using a generalized normalization transformation. *arXiv preprint arXiv:1511.06281*.

Ballé, J.; Laparra, V.; and Simoncelli, E. P. 2016. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*.

Ballé, J.; Minnen, D.; Singh, S.; Hwang, S. J.; and Johnston, N. 2018. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*.

Barman, N.; Zadtootaghaj, S.; Schmidt, S.; Martini, M. G.; and Möller, S. 2018. GamingVideoSET: a dataset for gaming video streaming applications. In *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 1–6. IEEE.

Bjontegaard, G. 2001. Calculation of average PSNR differences between RD-curves. *ITU SG16 Doc. VCEG-M33*.

Brancati, N.; Anniciello, A. M.; Pati, P.; Riccio, D.; Scognamiglio, G.; Jaume, G.; De Pietro, G.; Di Bonito, M.; Foncubierta, A.; Botti, G.; et al. 2022. BRACS: A dataset for breast carcinoma subtyping in h&e histology images. *Database*, 2022: baac093.

Bross, B.; Wang, Y.-K.; Ye, Y.; Liu, S.; Chen, J.; Sullivan, G. J.; and Ohm, J.-R. 2021. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10): 3736–3764.

Campos, J.; Meierhans, S.; Djelouah, A.; and Schroers, C. 2019. Content adaptive optimization for neural image compression. *arXiv preprint arXiv:1906.01223*.

Cheng, Z.; Sun, H.; Takeuchi, M.; and Katto, J. 2020. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7939–7948.

CLIC. 2020. Workshop and challenge on learned image compression. https://www.compression.cc/.

Duan, Z.; Lu, M.; Ma, Z.; and Zhu, F. 2022. Opening the black box of learned image coders. In *2022 Picture Coding Symposium (PCS)*, 73–77. IEEE.

Franzen, R. 1993. Kodak lossless true color image suite (photocd pcd0992). http://r0k.us/graphics/kodak/.

Guo, Z.; Zhang, Z.; Feng, R.; and Chen, Z. 2021. Soft then hard: Rethinking the quantization in neural image compression. In *International Conference on Machine Learning*, 3920–3929. PMLR.

He, D.; Yang, Z.; Peng, W.; Ma, R.; Qin, H.; and Wang, Y. 2022. ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5718–5727.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Jiang, W.; and Wang, R. 2023. MLIC++: Linear Complexity Multi-Reference Entropy Modeling for Learned Image Compression. In *ICML 2023 Workshop Neural Compression: From Information Theory to Applications*.

Jiang, W.; Yang, J.; Zhai, Y.; Ning, P.; Gao, F.; and Wang, R. 2023. MLIC: Multi-reference entropy model for learned image compression. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7618–7627.

Katakol, S.; Herranz, L.; Yang, F.; and Mrak, M. 2021. DANICE: Domain adaptation without forgetting in neural image compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1921–1925.

Li, S.; Dai, W.; Fang, Y.; Zheng, Z.; Fei, W.; Xiong, H.; and Zhang, W. 2024a. Revisiting Learned Image Compression With Statistical Measurement of Latent Representations. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4): 2891–2907.

Li, Z.; Liao, J.; Tang, C.; Zhang, H.; Li, Y.; Bian, Y.; Sheng, X.; Feng, X.; Li, Y.; Gao, C.; et al. 2024b. USTC-TD: A Test Dataset and Benchmark for Image and Video Coding in 2020s. *arXiv preprint arXiv:2409.08481*.

Liu, J.; Sun, H.; and Katto, J. 2023. Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14388–14397.

Lv, Y.; Xiang, J.; Zhang, J.; Yang, W.; Han, X.; and Yang, W. 2023. Dynamic Low-Rank Instance Adaptation for Universal Neural Image Compression. In *Proceedings of the 31st ACM International Conference on Multimedia*, 632–642.

Minnen, D.; Ballé, J.; and Toderici, G. D. 2018. Joint autoregressive and hierarchical priors for learned image compression. *Advances in neural information processing systems*, 31: 10794–10803.

Ni, Z.; Ma, L.; Zeng, H.; Chen, J.; Cai, C.; and Ma, K.-K. 2017. ESIM: Edge similarity for screen content image quality assessment. *IEEE Transactions on Image Processing*, 26(10): 4818–4831.

Presta, A.; Spadaro, G.; Tartaglione, E.; Fiandrotti, A.; and Grangetto, M. 2024. Domain Adaptation for Learned Image Compression with Supervised Adapters. In *2024 Data Compression Conference (DCC)*, 33–42. IEEE.

Robinson, M. S.; Brylow, S.; Tschimmel, M.; Humm, D.; Lawrence, S.; Thomas, P.; Denevi, B. W.; Bowman-Cisneros, E.; Zerr, J.; Ravine, M.; et al. 2010. Lunar reconnaissance orbiter camera (LROC) instrument overview. *Space science reviews*, 150: 81–124.

Shen, S.; Yue, H.; and Yang, J. 2023. Dec-adapter: Exploring efficient decoder-side adapter for bridging screen content and natural image compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12887–12896.

Sullivan, G. J.; Ohm, J.-R.; Han, W.-J.; and Wiegand, T. 2012. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12): 1649–1668.

Tsubota, K.; Akutsu, H.; and Aizawa, K. 2023. Universal deep image compression via content-adaptive optimization with adapters. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2529–2538.

Van Rozendaal, T.; Brehmer, J.; Zhang, Y.; Pourreza, R.; Wiggers, A.; and Cohen, T. S. 2021. Instance-adaptive video compression: Improving neural codecs by training on the test set. *arXiv preprint arXiv:2111.10302*.

Wallace, G. K. 1991. The JPEG still picture compression standard. *Communications of the ACM*, 34(4): 30–44.

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

Yang, J.; Shen, S.; Yue, H.; and Li, K. 2021. Implicit transformer network for screen content image continuous super-resolution. *Advances in Neural Information Processing Systems*, 34: 13304–13315.