Supplementary Material of Bidirectional Stereo Image Compression with Cross-Dimensional Entropy Model

Zhening Liu, Xinjie Zhang, Jiawei Shao, Zehong Lin*, and Jun Zhang

Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR
{zhening.liu,xzhangga,jiawei.shao}@connect.ust.hk,
{eezhlin,eejzhang}@ust.hk

The supplementary material provides additional visualization results in Sec. 6, extra ablation studies in Sec. 7, more experimental details in Sec. 8, and discussion about potential extensions in Sec. 9.

6 Additional Visualization Results

6.1 Qualitative Results

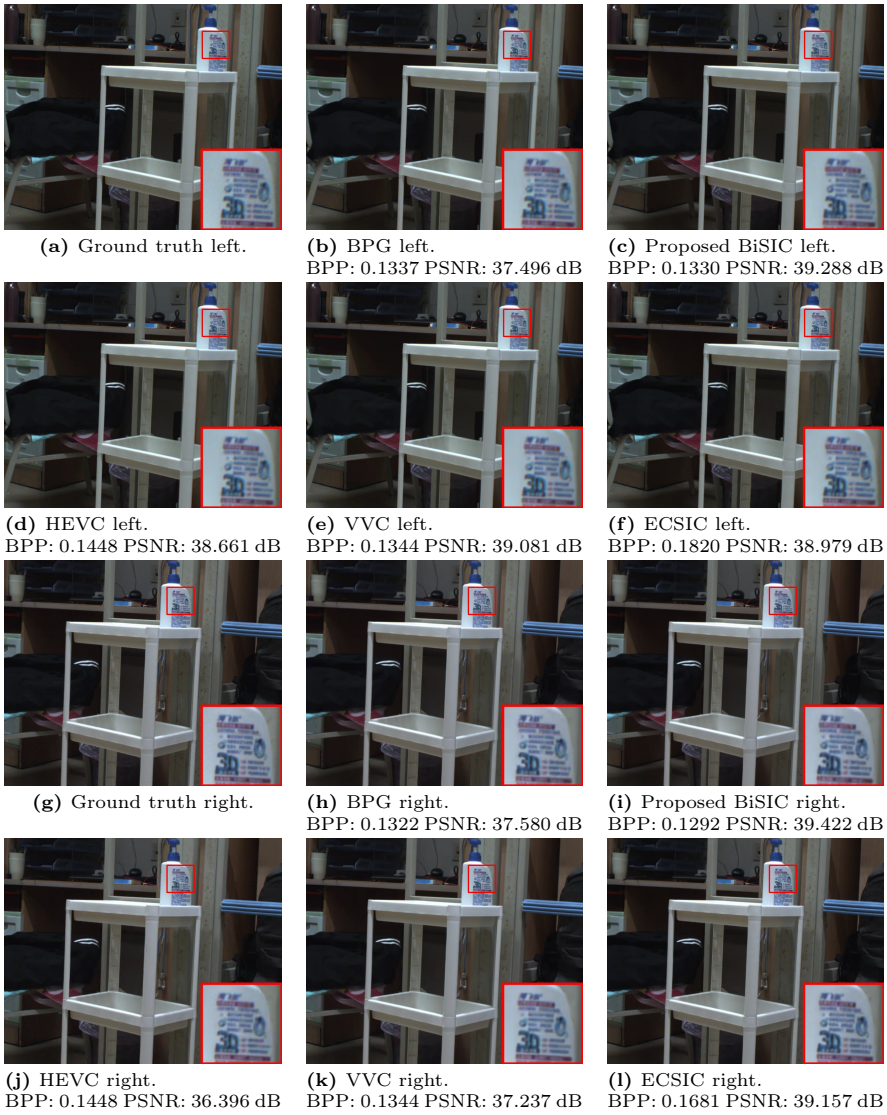
We visualize the qualitative results in Fig. 9, Fig. 16, and Fig. 17 to show the effectiveness of the proposed method, compared with baselines BPG [1], HEVC [12], VVC [3], SASIC [14], and ECSIC [13].

As shown in Fig. 9, our proposed BiSIC achieves higher PSNR quality with a lower BPP for both left and right views, compared with other methods. Besides, the reconstruction details and texture of BiSIC are closer to the ground truth. Moreover, the image qualities of the left and right views in our bidirectional design remain close, mitigating the imbalance issue in unidirectional methods. In contrast, HEVC and VVC adopt a predictive compression pipeline where one view is compressed normally, and the other view is generated through the disparity between the prediction and the real view. The unidirectional compression results in a 2.265 dB PSNR gap for HEVC and a 1.844 dB PSNR gap for VVC between stereo views, as seen in Fig. 9. ECSIC utilizes the spatial context from the left image to compress the right one, resulting in a higher compression quality of the right image. In Fig. 16, we illustrate another example on InStereo2K, where we can visually observe that the same area appears differently in the left and right views between Fig. 16d and Fig. 16j, as well as Fig. 16e and Fig. 16k, due to the unidirectional compression. Another group of visualization comparisons on Cityscapes is shown in Fig. 17.

6.2 Downstream Task Verification

The previous subsection provides a visual comparison between our method and the baselines. Note that imbalanced stereo quality is unfavorable for machine

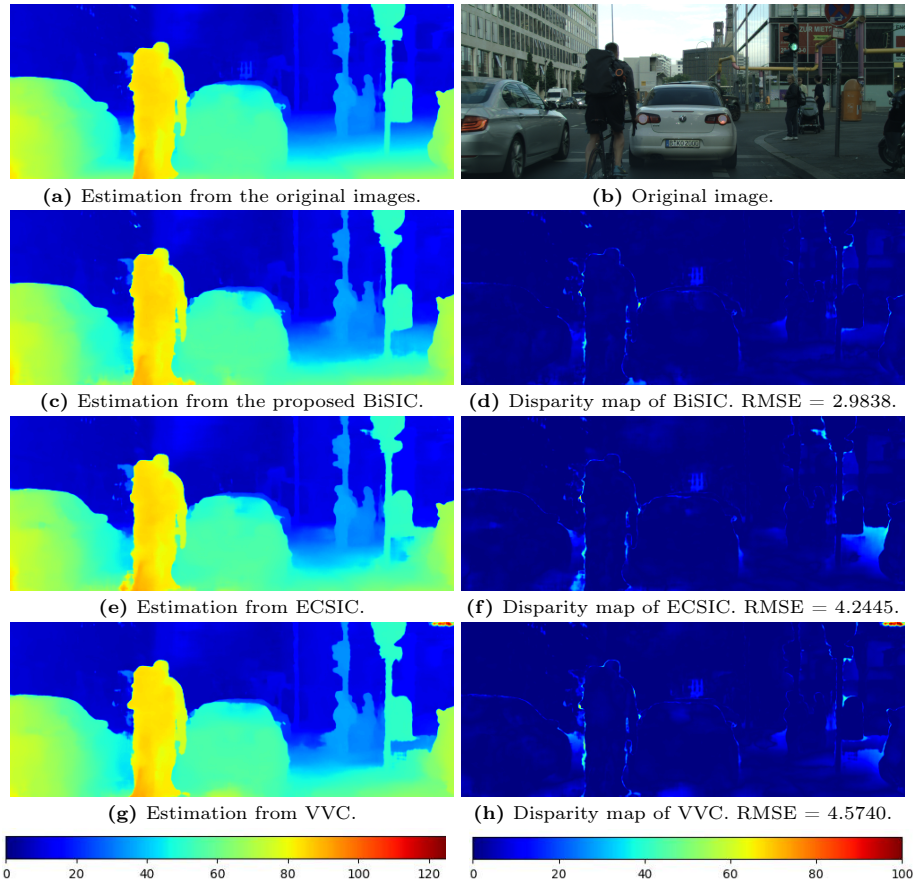
* Corresponding author



(a) Ground truth left.
(b) BPG left. BPP: 0.1337 PSNR: 37.496 dB
(c) Proposed BiSIC left. BPP: 0.1330 PSNR: 39.288 dB
(d) HEVC left. BPP: 0.1448 PSNR: 38.661 dB
(e) VVC left. BPP: 0.1344 PSNR: 39.081 dB
(f) ECSIC left. BPP: 0.1820 PSNR: 38.979 dB
(g) Ground truth right.
(h) BPG right. BPP: 0.1322 PSNR: 37.580 dB
(i) Proposed BiSIC right. BPP: 0.1292 PSNR: 39.422 dB
(j) HEVC right. BPP: 0.1448 PSNR: 36.396 dB
(k) VVC right. BPP: 0.1344 PSNR: 37.237 dB
(l) ECSIC right. BPP: 0.1681 PSNR: 39.157 dB

Fig. 9: Visualization of the reconstructed images. For classical video coding methods, such as HEVC and VVC, BPP is calculated as an average across two views.

vision and downstream tasks [9]. Therefore, it is interesting to investigate the degradation caused by different compression methods. In this subsection, we compare their performance on the stereo matching task. We employ a benchmark stereo matching method [4] on both ground truth stereo image pairs and reconstructed stereo image pairs from various compression methods to illustrate the degradation effect brought by compression.



(a) Estimation from the original images. **(b)** Original image.

(c) Estimation from the proposed BiSIC. **(d)** Disparity map of BiSIC. RMSE = 2.9838.

(e) Estimation from ECSIC. **(f)** Disparity map of ECSIC. RMSE = 4.2445.

(g) Estimation from VVC. **(h)** Disparity map of VVC. RMSE = 4.5740.

Fig. 10: Stereo matching results of different compression methods. The original image, the estimation from the original images, the estimations from reconstruction results, and their disparity maps are provided. All stereo matching results are estimated using the method in [4], and RMSE is calculated to reflect the accuracy [11]. The corresponding BPPs for BiSIC, ECSIC, and VVC are 0.103, 0.116, and 0.131, respectively.

The stereo matching results are visualized in Fig. 10. We calculate the root-mean-square-error (RMSE) [11] to quantify the disparity between the estimations from original images and reconstructed images. As illustrated in Fig. 10a and Fig. 10c, the estimation from the reconstructed results of our proposed BiSIC achieves a nearly identical estimation to the one from the ground truth. Moreover, it achieves the lowest RMSE among others, while requiring the lowest BPP. This finding demonstrates that BiSIC preserves most of the features and information in the stereo images after compression. In contrast, the decompression result of ECSIC (see Fig. 10e) fails to accurately estimate the right part of the image, where objects at different depths are confused. Although VVC (38.04 dB) achieves approximately the same average PSNR level as our BiSIC

(38.44 dB), there exists a discrepancy of 2.5769 dB in VVC between the left view (39.3268 dB) and the right view (36.7499 dB). Moreover, the estimation from VVC (see Fig. 10g) produces severe artifacts on the top right region and incorrect estimations in the right part. These observations suggest that balanced stereo image quality, which can be achieved through bidirectional compression, is beneficial to both visual perception and downstream tasks.

6.3 Bit Allocation Visualization

In this subsection, we examine the effectiveness of the proposed mutual attention blocks on compression performance. Fig. 11 visualizes the bit allocation of latents $\hat{\boldsymbol{y}}_l$ and $\hat{\boldsymbol{y}}_r$ in BiSIC with and without the mutual attention blocks. In particular, darker-colored regions indicate a greater number of allocated bits for image encoding (i.e., higher BPP), while regions with lighter colors are encoded with fewer bits (i.e., lower BPP). By incorporating the mutual attention blocks, our BiSIC method effectively identifies shared features between stereo views for redundancy reduction, thereby increasing the compression ratio.

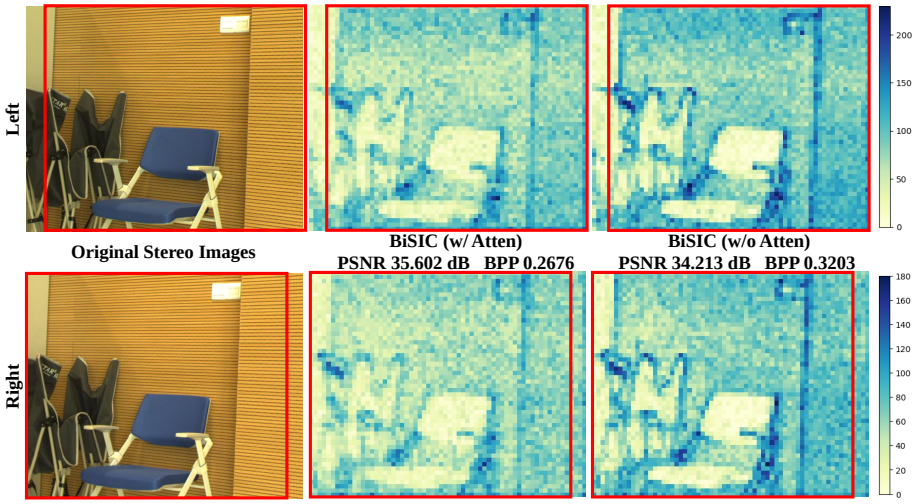


Fig. 11: Visualization of the bit allocation. The regions within red boundaries present the common areas of stereo images. (Left) Original images. (Middle) Bit allocation of BiSIC. (Right) Bit allocation of BiSIC without mutual attention blocks. BiSIC with mutual attention blocks achieves a higher average PSNR of 35.602 dB with a lower average BPP, compared to the baseline without attention (34.213 dB).

7 Extra Ablation Studies

7.1 Ablation Study on BiSIC-Fast

The ablation studies for the proposed BiSIC method are shown in Sec. 4, which illustrates the impact of each proposed component, including the 3D convolution backbone, cross-dimensional entropy model, and mutual attention block. In this subsection, we provide ablation studies for our fast variant, BiSIC-Fast. Specifically, we investigate the effect of our designed stereo-checkerboard structure and evaluate the significance of channel context. The RD performance is illustrated in Fig. 12, and we also calculate the Bjøntegaard Delta PSNR (BD-PSNR) [2] for comparison.

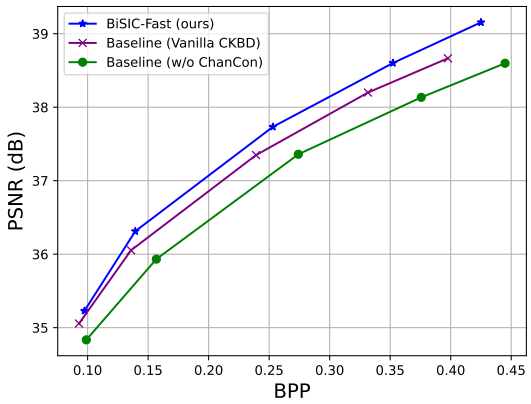


Fig. 12: Ablation study of BiSIC-Fast on InStereo2K dataset. The Baseline (Vanilla CKBD) replaces the stereo-checkerboard structure in BiSIC-Fast with the vanilla checkerboard, while the Baseline (w/o ChanCon) removes the utilization of channel context in BiSIC-Fast.

Effectiveness of Stereo-Checkerboard. The stereo-checkerboard structure enables joint learning from both views, aided by 3D convolution. To illustrate the effect of the stereo-checkerboard structure, we replace it with the vanilla checkerboard in [7] and present the ablation results in Fig. 12. We observe an RD performance degradation in this baseline compared to our BiSIC-Fast, specifically, with a BD-PSNR of -0.213 dB. This is because the cooperation of the stereo-checkerboard and 3D convolution enables the utilization of references from both self-view and the other view, and thus extracts more information compared with the vanilla checkerboard method, as shown in Fig. 13.

Effectiveness of Channel Context. To evaluate the contribution of channel context, we remove the slicing process on channel axis and the channel context

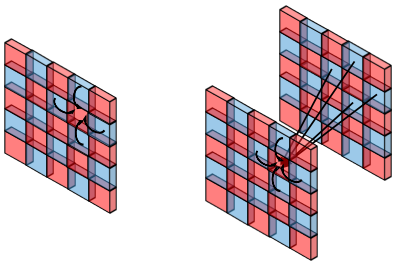


Fig. 13: (Left) Illustration of dependencies learned in a vanilla checkerboard structure. (Right) Illustration of the dependency utilized by a stereo-checkerboard structure. For simplicity, in both examples, we only show the conditional effect of the four neighboring entries around one target entry.

model. As shown in Fig. 12, channel context contributes to a significant improvement in performance, quantified as an improvement of 0.6173 dB in BD-PSNR. Therefore, without channel context, the stereo anchor part is conditioned only on hyperprior, which is relatively insufficient. Consequently, an off-the-optimal stereo anchor progressively provides an inadequate condition for the stereo non-anchor part, resulting in unsatisfactory performance.

7.2 Ablation Study on Number of Slices

In the channel-wise auto-regressive entropy model, the previously decoded part serves as a condition for the later part. Thus, it is interesting to investigate the relationship between the number of slices, compression performance, and model efficiency. Note that a higher precision in slicing generates abundant conditions, but more slices directly increase the time consumption of compression. This forms a trade-off between compression performance and speed. In this subsection, we conduct an ablation study for our proposed BiSIC on the number of slices K and investigate its effect on the trade-off between performance and efficiency. The RD performance on InStereo2K is shown in Fig. 14, along with several baselines for comparison. The Bjøntegaard Delta Bitrate (BDBR) [2] results relative to BPG are shown in Tab. 4. As demonstrated, reducing the number of slices leads to a slight decrease in the RD performance and accelerates the encoding/decoding process. Specifically, when $K = 6$, the time consumption is reduced by 42%, while the RD performance experiences a degradation of 3.19%. Nonetheless, it still outperforms other baselines, as shown in Fig. 14.

7.3 Ablation Study on ELIC Backbone

In our work, we employ 3D convolutional layers as the backbone of the codec. In Section 4.4 of the main body of the paper, we have provided ablation study compared with plain 2D convolution backbone baseline. Here, we provide the comparison results with previous SOTA codec backbone of ELIC [6], to further

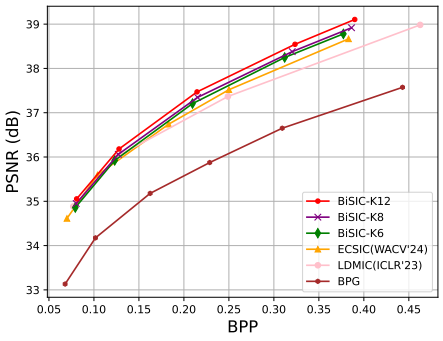


Fig. 14: RD performance of our BiSIC model with various numbers of slices K.

Method	Time	BDBR
BiSIC-K12	167.25s	-48.07%
BiSIC-K8	116.41s	-45.95%
BiSIC-K6	96.84s	-44.88%

Table 4: Runtime and BDBR results with various numbers of slices K.

validate the effectiveness of our 3D convolution based backbone in the codec for stereo image compression. Specifically, we maintain the other part of our model and replace our backbone with the residual block and attention based paradigm as in ELIC. We refer to this variant as Baseline (ELIC). Note that this baseline (43.1M parameters, 4353G FLOPs) has similar model size but higher computation cost compared with ours (49.3M parameters, 2978G FLOPs). Fig. 15 shows that our method achieves a BD-PSNR gain of 0.237 dB over this baseline.

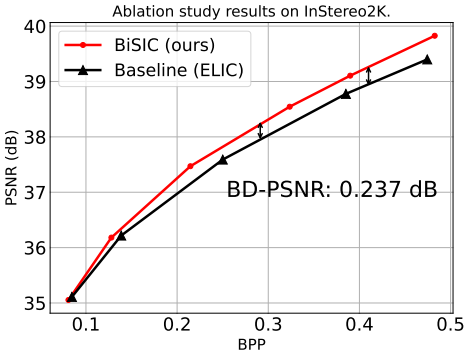


Fig. 15: RD performance of our BiSIC model compared with Baseline (ELIC).

8 Experimental Details

In this section, we provide more details about the neural network architectures and the process of model training.

Details of Entropy Model. The proposed cross-dimensional entropy model aggregates the hyperprior, spatial context, channel context, and stereo dependency to estimate the probability distributions of the compact latents. Let $\mathbf{G}$ denote the network and θ denote its learnable parameters. The hyperprior, channel context model, and spatial context model are formulated as follows:

$$\tilde{\boldsymbol{z}}_l, \tilde{\boldsymbol{z}}_r = h_s(\hat{\boldsymbol{z}}_l, \hat{\boldsymbol{z}}_r), \tag{11}$$

$$\boldsymbol{\Theta}_l, \boldsymbol{\Theta}_r = \mathbf{G}_{\text{ch}}(\hat{\boldsymbol{y}}_l^{<k}, \hat{\boldsymbol{y}}_r^{<k}; \theta_{\text{ch}}), \tag{12}$$

$$\boldsymbol{\Upsilon}_l, \boldsymbol{\Upsilon}_r = \mathbf{G}_{\text{sp}}(\hat{\boldsymbol{y}}_{l,<i}^{k}, \hat{\boldsymbol{y}}_{r,<i}^{k}; \theta_{\text{sp}}). \tag{13}$$

The hyperprior dependency is obtained through a hyper decoder, which is also depicted in Fig. 1 and Eq. (4) in the main text. The channel context model $\mathbf{G}_{\text{ch}}$ comprises four convolutional layers with 1×1 kernel size, and it produces the channel dependency feature with 128 channels. The spatial context model $\mathbf{G}_{\text{sp}}$ is obtained through one layer of masked 3D convolution, which is illustrated in Fig. 4(a) in the main body of the paper. The estimated mean and variance are produced with the aggregation model as follows:

$$\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2 = \mathbf{G}_{\text{ag}}(\tilde{\boldsymbol{z}}_l, \boldsymbol{\Theta}_l, \boldsymbol{\Upsilon}_l; \theta_{\text{ag}}), \tag{14}$$

$$\boldsymbol{\mu}_r, \boldsymbol{\sigma}_r^2 = \mathbf{G}_{\text{ag}}(\tilde{\boldsymbol{z}}_r, \boldsymbol{\Theta}_r, \boldsymbol{\Upsilon}_r; \theta_{\text{ag}}), \tag{15}$$

where $\mathbf{G}_{\text{ag}}$ represents the network that aggregates multiple references and provides estimations. This network consists of four convolutional layers with 1×1 kernels. The conditional estimations of the two views are as follows:

$$p_{\hat{\boldsymbol{y}}_l}(\hat{\boldsymbol{y}}_l | \hat{\boldsymbol{z}}_l; \theta_{\text{ag}}, \theta_{h_s}, \theta_{\text{ch}}, \theta_{\text{sp}}) = \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2), \tag{16}$$

$$p_{\hat{\boldsymbol{y}}_r}(\hat{\boldsymbol{y}}_r | \hat{\boldsymbol{z}}_r; \theta_{\text{ag}}, \theta_{h_s}, \theta_{\text{ch}}, \theta_{\text{sp}}) = \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\sigma}_r^2). \tag{17}$$

Details of Stereo-Checkerboard. The proposed fast variant relies on the stereo-checkerboard structure, which transforms the entry-by-entry auto-regressive process into a two-fold operation. Specifically, the stereo views are split into two parts: stereo anchor part $\hat{\boldsymbol{y}}_{\text{ach}}$ and stereo non-anchor part $\hat{\boldsymbol{y}}_{\text{nac}}$, as shown in Fig. 5 in the main body. The anchor part is encoded/decoded with a hyperprior and the channel-wise condition, where the estimations of mean $\boldsymbol{\mu}_{\text{ach}}$ and variance $\boldsymbol{\sigma}_{\text{ach}}^2$ for stereo anchor part are formulated as:

$$\boldsymbol{\mu}_{l,\text{ach}}, \boldsymbol{\sigma}_{l,\text{ach}}^2 = \mathbf{G}_{\text{ag-ach}}(\tilde{\boldsymbol{z}}_l, \boldsymbol{\Theta}_l; \theta_{\text{ag-ach}}), \tag{18}$$

$$\boldsymbol{\mu}_{r,\text{anc}}, \boldsymbol{\sigma}_{r,\text{anc}}^2 = \mathbf{G}_{\text{ag-ach}}(\tilde{\boldsymbol{z}}_r, \boldsymbol{\Theta}_r; \theta_{\text{ag-ach}}). \tag{19}$$

Then, with the existing anchor part, we obtain the anchor context feature $\boldsymbol{\Upsilon}_{\text{ach}}$ using 3D convolution as:

$$\boldsymbol{\Upsilon}_{\text{ach}} = \mathbf{G}_{\text{ach}}(\hat{\boldsymbol{y}}_{\text{ach}}; \theta_{\text{ach}}). \tag{20}$$

Notably, $\mathbf{G}_{\text{ach}}$ is an ordinary 3D convolutional layer with a kernel size of $(3, 5, 5)$, as the whole stereo anchor part has been obtained and the non-anchor entries

are set to zero, eliminating the need for a mask. The anchor context feature $\boldsymbol{\Upsilon}_{\text{ach}}$ serves as a reference for the stereo non-anchor part. Thus, the mean $\boldsymbol{\mu}_{\text{nac}}$ and variance $\boldsymbol{\sigma}^2_{\text{nac}}$ of the stereo non-anchor part are estimated by:

$$\boldsymbol{\mu}_{l,\text{nac}}, \boldsymbol{\sigma}^2_{l,\text{nac}} = \mathbf{G}_{\text{ag-nac}}(\tilde{\boldsymbol{z}}_l, \boldsymbol{\Theta}_l, \boldsymbol{\Upsilon}_{l,\text{ach}}; \theta_{\text{ag-nac}}), \tag{21}$$

$$\boldsymbol{\mu}_{r,\text{nac}}, \boldsymbol{\sigma}^2_{r,\text{nac}} = \mathbf{G}_{\text{ag-nac}}(\tilde{\boldsymbol{z}}_r, \boldsymbol{\Theta}_r, \boldsymbol{\Upsilon}_{r,\text{ach}}; \theta_{\text{ag-nac}}), \tag{22}$$

where the aggregation networks $\mathbf{G}_{\text{ag-ach}}$ and $\mathbf{G}_{\text{ag-nac}}$ consist of four convolutional layers with 1×1 kernels.

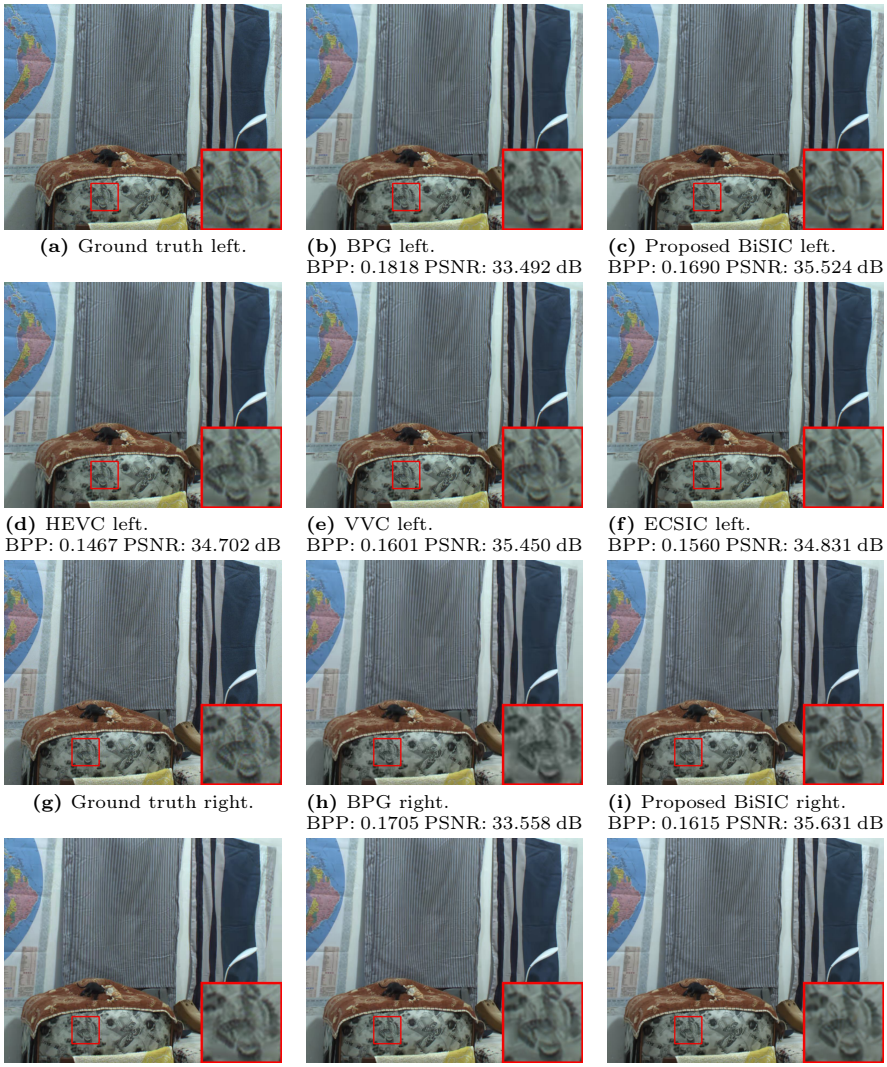
Implementation Details. All training and testing settings on datasets follow previous works [13–15] to ensure a fair comparison. Specifically, each image in the InStereo2K dataset is pre-processed to ensure that its size is divisible by 64. For the Cityscapes dataset, rectification artifacts and the self-vehicle are removed, with 64, 256, and 128 pixels being cut off from the top, bottom, and sides, respectively, of every image. In the test phase, we evaluate the performance using images of size $1,024 \times 832$ from the InStereo2K dataset and images of size $1,792 \times 704$ from the Cityscapes dataset.

For the traditional codec baselines, BPG [1] is implemented with YUV 4:4:4 to maintain its good performance. HEVC and VVC are implemented based on JVET[1], where we first convert the stereo image pair into a YUV 4:4:4 video using ffmpeg[2], followed by video compression. The left view is regarded as an I frame and the other one is regarded as a P frame during video compression. Notably, MV-HEVC only supports the 4:2:0 chroma mode, which results in suboptimal PSNR scores at higher bitrates [13]. In addition, we reproduce BCSIC [8] and test it under the same test image settings as in [13–15]. This is because the original RD curves reported in the paper [8] are tested on 512×512 images, which yields much lower values compared to other shown baselines. Therefore, we present the results under the same testing setup for a fair comparison.

9 Extensions

Based on our proposed bidirectional stereo image compression model, BiSIC, and its fast variant, several interesting follow-up directions are worth investigating. Firstly, the Vision Transformers [5, 10] has proven effective in single image compression [16, 17] due to its ability in feature learning and latent representation. Thus, it has the potential to further optimize the RD performance when used as the backbone of our model. Secondly, it is interesting to extend this work to multi-view video compression or immersive video compression pipelines, which further cater to the current boom in AR/VR technology.

[1] `https://vcgit.hhi.fraunhofer.de/jvet`
[2] `https://ffmpeg.org/`



(a) Ground truth left.

(b) BPG left.
BPP: 0.1818 PSNR: 33.492 dB

(c) Proposed BiSIC left.
BPP: 0.1690 PSNR: 35.524 dB

(d) HEVC left.
BPP: 0.1467 PSNR: 34.702 dB

(e) VVC left.
BPP: 0.1601 PSNR: 35.450 dB

(f) ECSIC left.
BPP: 0.1560 PSNR: 34.831 dB

(g) Ground truth right.

(h) BPG right.
BPP: 0.1705 PSNR: 33.558 dB

(i) Proposed BiSIC right.
BPP: 0.1615 PSNR: 35.631 dB

(j) HEVC right.
BPP: 0.1467 PSNR: 32.558 dB

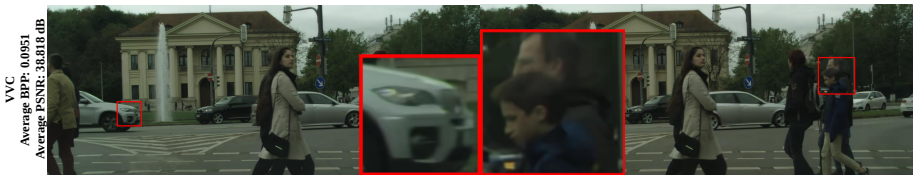
(k) VVC right.
BPP: 0.1601 PSNR: 33.477 dB

(l) ECSIC right.
BPP: 0.1431 PSNR: 35.010 dB

Fig. 16: Visualization of the reconstructed images. For classical video coding methods, such as HEVC and VVC, BPP is calculated as an average across two views.

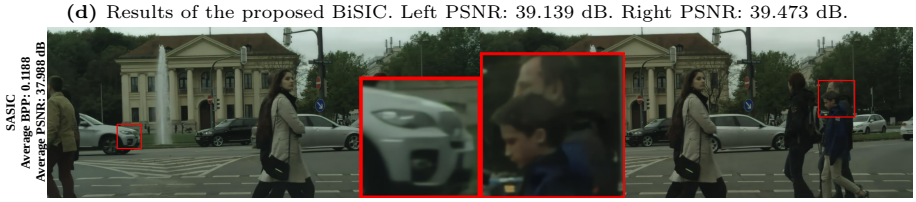
(a) Ground truth images.

(b) Results of HEVC. Left PSNR: 38.619 dB. Right PSNR: 37.871 dB.

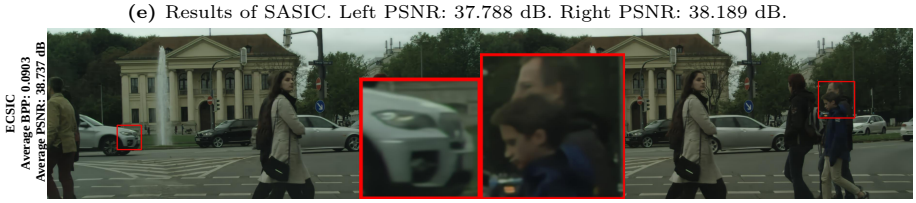


(c) Results of VVC. Left PSNR: 39.665 dB. Right PSNR: 37.971 dB.

(d) Results of the proposed BiSIC. Left PSNR: 39.139 dB. Right PSNR: 39.473 dB.



(e) Results of SASIC. Left PSNR: 37.788 dB. Right PSNR: 38.189 dB.



(f) Results of ECSIC. Left PSNR: 38.402 dB. Right PSNR: 39.072 dB.

Fig. 17: Visualization on the Cityscapes dataset. We compare our BiSIC with HEVC, VVC, SASIC, and ECSIC. BiSIC achieves the best PSNR performance with a relatively low BPP. Moreover, BiSIC maintains balanced qualities between stereo views.

References

1. Bellard, F.: BPG image format. Website (2014), https://bellard.org/bpg/
2. Bjontegaard, G.: Calculation of average psnr differences between rd-curves. ITU SG16 Doc. VCEG-M33 (2001)
3. Bross, B., Wang, Y.K., Ye, Y., Liu, S., Chen, J., Sullivan, G.J., Ohm, J.R.: Overview of the versatile video coding (VVC) standard and its applications. IEEE Transactions on Circuits and Systems for Video Technology **31**(10), 3736–3764 (2021)
4. Chang, J.R., Chen, Y.S.: Pyramid stereo matching network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5410–5418 (2018)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. He, D., Yang, Z., Peng, W., Ma, R., Qin, H., Wang, Y.: 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. pp. 5718–5727 (2022)
7. He, D., Zheng, Y., Sun, B., Wang, Y., Qin, H.: Checkerboard context model for efficient learned image compression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14771–14780 (2021)
8. Lei, J., Liu, X., Peng, B., Jin, D., Li, W., Gu, J.: Deep stereo image compression via bi-directional coding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19669–19678 (June 2022)
9. Liu, Y., Ren, J., Zhang, J., Liu, J., Lin, M.: Visually imbalanced stereo matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2029–2038 (2020)
10. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
11. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision **47**, 7–42 (2002)
12. Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on circuits and systems for video technology **22**(12), 1649–1668 (2012)
13. Wödlinger, M., Kotera, J., Keglevic, M., Xu, J., Sablatnig, R.: ECSIC: Epipolar cross attention for stereo image compression. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3436–3445 (2024)
14. Wödlinger, M., Kotera, J., Xu, J., Sablatnig, R.: SASIC: Stereo image compression with latent shifts and stereo attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 661–670 (June 2022)
15. Zhang, X., Shao, J., Zhang, J.: LDMIC: Learning-based distributed multi-view image coding. arXiv preprint arXiv:2301.09799 (2023)
16. Zhu, Y., Yang, Y., Cohen, T.: Transformer-based transform coding. In: International Conference on Learning Representations (2021)

17. Zou, R., Song, C., Zhang, Z.: The devil is in the details: Window-based attention for image compression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17492–17501 (2022)