Kalpa Publications in Computing

Volume 22, 2025, Pages 473–485

Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

Towards Higher Information Density in Image Transmission: Learned Image Compression for Construction Site Monitoring

Jiucai Liu[1], Chengzhang Chai[1], Linghan Ouyang[1], and Haijiang Li[1*]

1 School of Engineering, Cardiff University, Cardiff, UK.

LiuJ151@cardiff.ac.uk, lih@cardiff.ac.uk

Abstract

The digital transformation in the Architecture, Engineering, and Construction (AEC) sector underscores the growing need for efficient data transmission, especially in computer vision tasks that depend on the transfer of large volumes of images. In this work, a novel method is introduced to enhance data transmission efficiency in an edge-cloud coordinated architecture using Learned Image Compression (LIC). By integrating the LIC model with multiple downstream task models (Mask R-CNN and Faster R-CNN), the proposed framework aligns their respective latent features, resulting in a task-oriented LIC model that optimises compression for specific tasks. The approach increases the proportion of task-relevant information—referred to as information density—in the transmitted bitstream. Experimental results demonstrate that this method significantly reduces data transmission load while concentrating the transmitted bits on regions essential for downstream tasks, all without a notable decrease in task accuracy.

1 Introduction

1.1 Background

The Architecture, Engineering, and Construction (AEC) industry is actively pursuing productivity improvements through advanced technologies such as robotics (Zhang et al., 2023), digital twins (Tuhaise et al., 2023), and extended reality (XR) (J. C. P. Cheng et al., 2020). These technologies provide opportunities for automating human-robot interaction, robotic-based construction, and infrastructure inspection. These technologies rely on fundamental processes like data collection, transmission, and processing. With advancements in Deep Learning (DL) and camera sensors, Computer Vision (CV) has gained significant attention in the AEC sector, raising the demand for extensive image data to support tasks such as image classification, object detection, and instance segmentation (Xu et al., 2021). However, most current research focuses on enhancing task-specific accuracy through increasingly complex methods, with limited attention to evaluating the usability and real-time performance of CV algorithms from the perspectives of system architecture and data transmission.

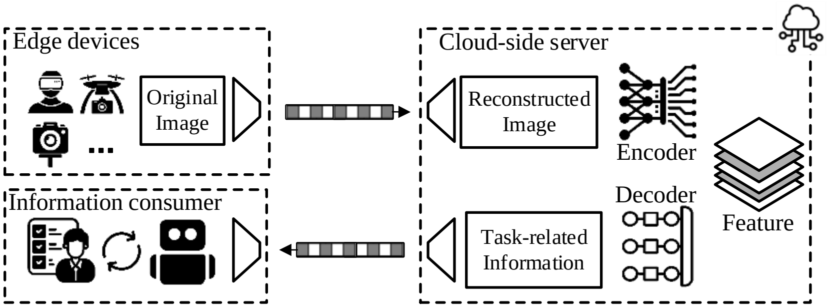


Figure 1: The edge-cloud system in AEC (from the perspective of information flow)

Due to the high computational demands of deep learning, equipping each edge device with powerful GPUs is inefficient and costly. Consequently, many studies propose an edge-cloud architecture as shown in Figure 1, where computationally intensive tasks are handled in the cloud-side server, while edge devices focus on data collection, data transmission, and information visualization (Alizadehsalehi et al., 2020; Cheng et al., 2020). For CV-related applications, on the edge side, captured images need to be compressed and encoded as bitstream for data transmission. On the cloud side, latent feature will be extracted from the reconstructed image by the encoder of pre-trained DL models and the task-related semantic information will be generated by the decoder of pre-trained DL models. This setup addresses the limited processing capacity of edge devices but highlights the requirements for data transmission and information extraction.

Although there is research on data transmission in the AEC sector, it mainly focuses on hardware and communication layer protocols (Tuhaise et al., 2023), without addressing how to handle transmitted data and improve transmission efficiency. On the other hand, for information extraction in CV applications, more dense DL models are increasingly used in AEC, yet they are often deployed on powerful GPUs, neglecting the interaction between edge devices and the cloud-side server. Additionally, inspired by the research of Huang and Wu, high-level features extracted by encoders can be reused across multiple downstream tasks (Huang and Wu, 2024). However, current AEC research mainly trains independent DL models for each task, leading to inefficient resource utilization and reduced real-time performance.

To address these limitations, this research focuses on two primary aspects: First, it aims to improve data transmission efficiency by enhancing the information density of transmitted image data within AEC domain. Specifically, Learned Image Compression (LIC) is introduced to prioritize task-relevant information rather than transmitting the entire image content. Second, a multi-task processing solution is proposed to integrate multiple downstream tasks within a unified framework, thereby maximizing computational reusability and improving efficiency.

1.2 Image Compression and Learned Imgae Compression

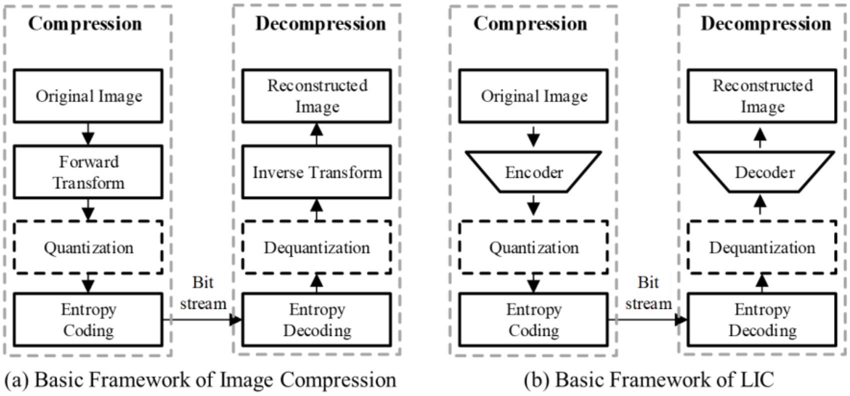


Figure 2: Basic framework of image compression and learned image compression

The basic format of data transmission is bitstream. For computer vision tasks, digital images are encoded into bitstreams and reconstructed by codecs, a process known as image compression. As shown in Figure 2, image compression generally consists of two main phases: encoding and decoding. The encoding process involves forward transform, quantization, and encoding, while the corresponding decoding process includes inverse encoding, inverse quantization, and image reconstruction. Traditional image compression algorithms, such as JPEG (Wallace, 1992), use handcrafted operators to remove redundant information (e.g., large uniform color areas), effectively reducing storage and transmission load with minimal quality loss of image according to human perception.

Recently, with the rise of deep learning, LIC has emerged, transforming images into feature representations and removing redundancy through deep learning-based feature extractors, thereby supporting efficient downstream image reconstruction. Ballé et al. introduced the first end-to-end optimised model for image compression (Ballé et al., 2017). Later, Ballé et al. extended this work by incorporating a hyper-prior to better capture spatial dependencies in the latent representation (Ballé et al., 2018). Building on the success of auto-regressive priors in probabilistic generative models, Minnen et al. further improved the entropy model by adding an auto-regressive component (Minnen et al., 2018). Z. Cheng et al, enhanced the network architecture using residual blocks and integrated a simplified attention module, replacing the commonly used Single Gaussian Model (SGM) with a Gaussian Mixture Model (GMM) (Z. Cheng et al., 2020). To reduce the need for serial processing in autoregressive context models, Minnen and Singh proposed a channel-wise auto-regressive entropy model (Minnen & Singh, 2020).

1.3 LIC-enhanced Machine Vision

With advancements in computer vision, images captured by cameras are increasingly processed by Artificial Intelligence (AI) algorithms rather than being solely consumed by humans. Consequently, LIC has evolved in two directions: human-perception-oriented LIC and machine-vision-oriented LIC. Human-perception-oriented LIC focuses on transmitting data for reconstructing high-fidelity digital images while machine-vision-oriented LIC aims to convey task-relevant information. Figure 3 illustrates the high-level JPEG AI framework referenced in the JPEG AI white paper, featuring three distinct pipelines. In this learning-based image coding framework, digital images serve as input, and the output bitstream can be processed in two ways: it can either be reconstructed through a standard pipeline for human visualization or it can be optimised for machine-vision-oriented LIC. This dual functionality demonstrates the framework's capability to meet both human and machine processing requirements.

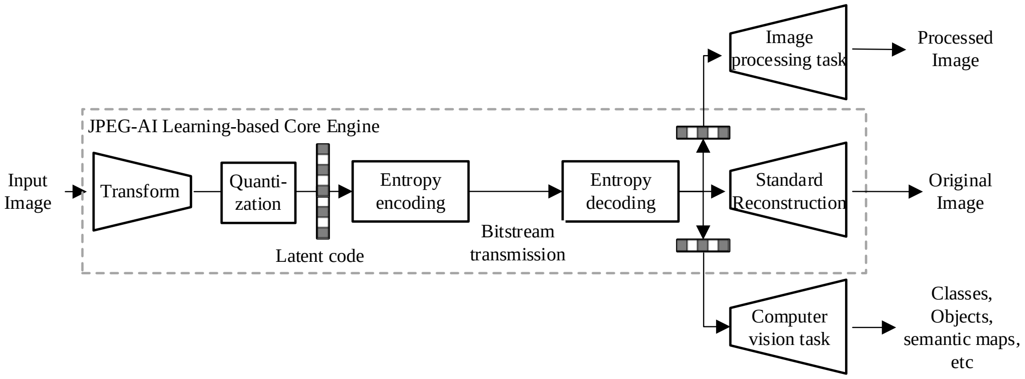


Figure 3: JPEG AI learning-based image coding framework

The concept of machine-vision-oriented LIC aligns well with the core principles of machine learning, where the objective is to extract essential features from raw data to create a latent representation for downstream decision-making or restoration tasks. This parallels the Minimum Description Length (MDL) principle in information theory, which seeks to find the most efficient model by minimizing the code length required to describe both the model and the data (Grünwald and Roos, 2019). Thus, machine-vision-oriented LIC not only enhances compression efficiency by focusing on task-specific information but also reflects the broader objective of machine learning to extract and compress the most relevant features, creating synergies between visual analysis and data compression.

The research of machine-vision-oriented LIC starts from Zhang et al.'s research, which they propose a research question of whether the bitstream of images and image features could be unified to serve both compression and retrieval simultaneously (Zhang et al., 2017). To address that, they proposed a content-based image retrieval system in which images are encoded once, and the encoded bitstream can be used for both image reconstruction and direct comparison for similar image retrieval. Several studies have focused on designing bitrate-efficient quantization for image compression while minimizing classification accuracy loss (Chamain et al., 2019; Liu et al., 2018). These joint rate–distortion–accuracy approaches aim to optimise quantization steps for JPEG and JPEG 2000 encoders to reduce both classification loss and bitrate.

With the success of LIC, interest in visual compression for machine vision has surged, focusing on maintaining machine task performance on compressed data. Le et al. introduced the first end-to-end learned system that optimises the rate-distortion trade-off, where the distortion term includes the training loss of a pre-trained neural network task (Le et al., 2021a). Le et al. introduced a content-adaptive fine-tuning method applied during inference, which aims to optimise the latent representation to enhance compression efficiency for machine-based tasks (Le et al., 2021b). Wang et al. developed an inverted bottleneck structure for the encoder and explored ways to refine the network architecture, specifically to improve compression performance for machine vision applications (Wang et al., 2021).

2 Method

This article presents a unified machine vision framework enhanced by LIC and applies it to the real-world scenario of construction site monitoring in the AEC domain. This chapter will introduce the proposed method from three aspects: overall structure, feature alignment, and training details.

2.1 Overall Structure

As shown in Figure 4, the proposed architecture comprises a single LIC model along with two downstream task models tailored for object detection and instance segmentation. Serving as the bottleneck of the overall framework, the LIC model is divided into an Encoder and a Decoder. The Encoder, deployed on an edge device, is primarily responsible for mapping raw images into a high-dimensional latent feature while effectively filtering out redundant information. It dynamically selects high-level features pertinent to the downstream tasks and encodes these features into a bitstream. In contrast, the Decoder operates in the cloud side alongside the downstream task models, focusing on reconstructing high-level features from the bitstream and subsequently regenerating the digital images. These reconstructed digital images are then input into the pre-trained models of the downstream tasks, enabling the extraction of essential high-level semantic features. For object detection tasks, the extracted semantic feature includes class labels and bounding boxes; for instance segmentation tasks, it encompasses class labels and segmentation masks.

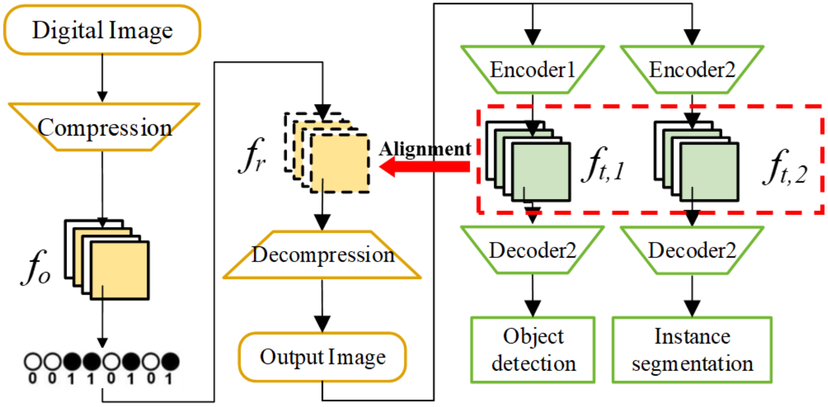
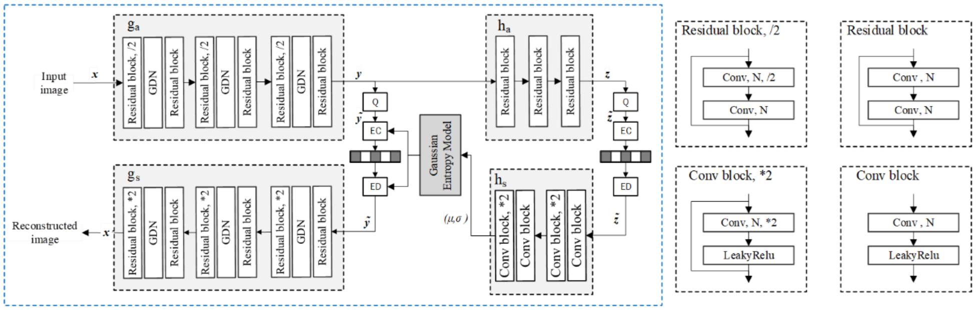


Figure 4: Overall framework of LIC-enhanced multi-task machine vision

2.2 Detail of LIC Model



(a) overall structure of LIC model (b) detail of components

Figure 5: Detail of LIC model

Figure 5 illustrates the fundamental architecture of the proposed LIC model, which primarily draws from the hyper-prior model presented by Ballé et al. (Ballé et al., 2018). This model is a dual-branch Variational AutoEncoder (VAE), designed to efficiently encode and compress image data while leveraging a hyperprior to enhance the modelling of latent representations. Q, EC and ED denote the quantization, entropy coding and entropy decoding, respectively.

The g_a–g_s branch is the main branch of the model, focusing on transforming the input image into its latent representation through encoding and then reconstructing it during decoding. The main branch is responsible for the actual compression and reconstruction of the image, ensuring that the salient details necessary for downstream tasks are preserved.

The h_a–h_s is the hyperprior of the model. It acts as an auxiliary model that captures the statistical dependencies among the latent representations generated by the encoder. By modelling these dependencies, the hyperprior enhances the efficiency of entropy coding, allowing for more compact representations of the image data. It essentially predicts the probability distribution of the latent variables, which is crucial for effective lossless compression. This hierarchical approach improves reconstruction quality by reducing the overall bitrate while maintaining or enhancing image fidelity.

The hyperprior and the main branch optimise the image compression process by enabling efficient representation and reconstruction of images. The hyperprior enhances the overall model's capacity to predict and encode data, while the main branch ensures that the essential features of the image are effectively captured and transmitted.

Notably, the model depicted in Figure 5 is deployed in a distributed manner, with the encoder implemented on the edge device and the decoder on the cloud side. The encoder comprises components such as Q, EC, the Gaussian Entropy Model, along with the h_a and g_a modules, while the decoder includes Q, EC, the Gaussian Entropy Model, and the h_s and g_s modules, executing on the cloud side.

For downstream tasks, this study focuses on two classic CV applications: object detection and instance segmentation. For the object detection task, the Faster R-CNN model is selected, while the Mask R-CNN model is chosen for instance segmentation. Both models utilise the Feature Pyramid Network (FPN) with a ResNet-50 backbone.

Table 1 presents the parameter counts for the various models within this architecture. Notably, the LIC model exhibits a relatively smaller model size compared to the downstream task models. Furthermore, from the perspective of model deployment, the distribution of the computational load is facilitated by the separation of the LIC encoder and decoder across different devices, allowing for an efficient allocation of processing resources.

Margin	LIC Encoder	LIC Decoder	Mask R-CNN	Faster R-CNN
Total params	6,604,337	9,237,565	43,982,622	41,356,561
FLOPs (G)	9.13	9.26	14.180	12.058

Table 1: Comparison of model sizes between LIC encoder, LIC decoder, and downstream task models

2.3 Latent Feature Alignment

In deep learning models, the carrier of information is the latent feature. The primary task of human-perception-oriented LIC is to ensure that the latent feature contains more information relevant to human perception, such as colour and texture. However, machine-vision-oriented LIC differs in that the latent feature it reconstructs should be more relevant to downstream tasks, often forming a subset of the information related to human perception. In light of this, this paper proposes a Latent Feature Alignment method, allowing the LIC model to selectively retain information pertinent to downstream tasks in an end-to-end manner, thereby improving compression efficiency without significantly impacting task accuracy.

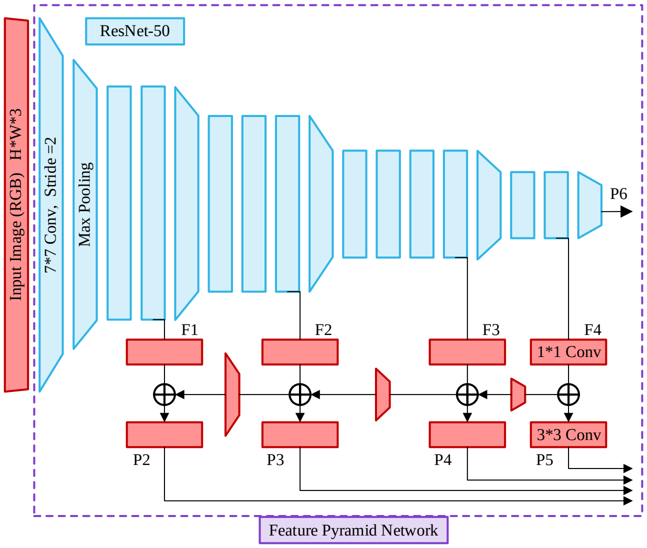


Figure 6: Feature pyramid network used for Faster R-CNN and Mask R-CNN

To adapt LIC model for generating a decoded image suitable for machine perception, three steps are employed.

- First, feature extraction is performed on the original image using the FPN of a pre-trained task model, selecting several latent features—P2, P3, P4, P5, and P6—as $\hat{f}$, as illustrated in the figure.
- Next, the LIC model is combined with the FPN, and the LIC model is trained while keeping the FPN frozen. In the same way, several latent features extracted by the FPN are denoted as f.
- Finally, the perceptual loss, defined as the difference between f and $\hat{f}$, is used as an additional term in the loss function, and the LIC model is trained through backpropagation.

Through this end-to-end training setup, the information transmitted by the LIC model is aligned with the information extracted by the FPN relevant to downstream tasks, allowing the LIC model to prioritize the transmission of task-relevant information instead of the complete set of information present in the image.

2.4 Training Detail

The training data utilised in this study is derived from the publicly available dataset, which is divided into a training set (19,404 images), a validation set (4,000 images), and a testing set (18,264 images) (Xuehui et al., 2021). This dataset encompasses a total of 13 object classes, amounting to 116,380 labels. For the purposes of this research, a subset of the overall dataset was randomly sampled, with the distribution of the data illustrated in Table 2.

Notably, the original dataset exhibits a significant long-tail distribution issue, which can adversely affect the recognition accuracy of classes with fewer instances. Consequently, this study focuses exclusively on four object classes, each having a label count greater than 500, to ensure sufficient representation and improve recognition performance.

Category	Worker	Excavator	Static crane	Truck
Train	12103	1431	1124	740
Val	586	116	75	55

Table 2: Distribution of instances across different types

The neural network constructed in this study has two primary optimization objectives: to reduce the bit rate (the amount of data transmitted) and to minimize the distortion (the loss of quality or accuracy). However, there generally exists a trade-off between these two objectives. To address this, the loss function was developed following the standard format of rate-distortion cost commonly used in the LIC domain, as illustrated in Equation (1).

$$L = R + \lambda D \tag{1}$$

where λ is a Lagrange multiplier that controls the trade-off between rate R and distortion D. By adjusting λ, one can prioritize minimizing either the rate or the distortion depending on the specific requirements of the application.

$$R = -\log p(\tilde{z}) - \log p(\tilde{y}|\tilde{z}) \tag{2}$$

R represents the number of bits required to represent the compressed data, which is typically expressed in bits per symbol or bits per pixel in image compression contexts. It is calculated based on information theory principles, particularly using concepts from Shannon's entropy.

$$D = d\left(f, \hat{f}\right) = \frac{1}{5}\sum_{i=2}^{6} \mathrm{MSE}(P_i(x), P_i(x')) \tag{3}$$

In LIC, D typically reflects the difference between the original and the reconstructed digital image. In this paper, the perceptual loss D measures the Mean Squared Error (MSE) between latent features extracted from the pretrained FPN as shown in Figure 6.

3 Results

To evaluate the effectiveness of the proposed method, this paper discusses the results from both qualitative and quantitative perspectives.

3.1 Qualitative Result

Figure 7 contrasts the original image and the image reconstructed by the traditional LIC model, Faster-R-CNN-enhanced LIC model and Mask-R-CNN-enhanced LIC model, while Figure 8 contrasts the task output based on the reconstructed image.

The traditional LIC model effectively reconstructs some details of the original image but also introduces some artifacts. However, these artifacts have little impact on the accuracy of downstream tasks.

On the contrary, the proposed machine-vision-oriented LIC model focuses specifically on regions relevant to downstream tasks during the image compression and reconstruction process, achieving high data reconstruction quality in those regions, which in turn results in certain accuracy for downstream tasks. For irrelevant regions, fewer bits are allocated, leading to lower image reconstruction quality.

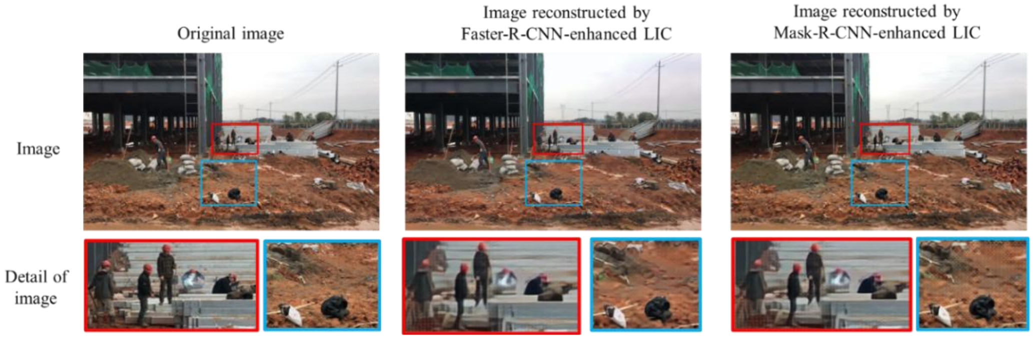

Original image
Image reconstructed by Faster-R-CNN-enhanced LIC
Image reconstructed by Mask-R-CNN-enhanced LIC
Image
Detail of image

Figure 7: Example of reconstructed images

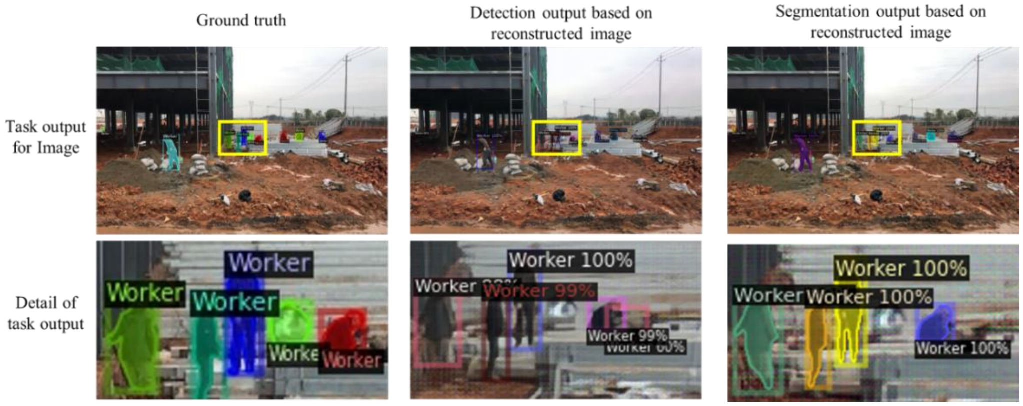


Figure 8: Example of task output based on reconstructed images

Figure 9 illustrates the comparison of bit allocation across various regions in images reconstructed by the traditional LIC model and the Faster-R-CNN-enhanced LIC model. The results demonstrate a clear trend where the Faster-R-CNN-enhanced model concentrates a higher proportion of bits in regions crucial for downstream tasks, while allocating fewer bits to less relevant areas. This suggests that the proposed model effectively boosts the density of task-relevant information in the bitstream, thereby enhancing information efficiency.

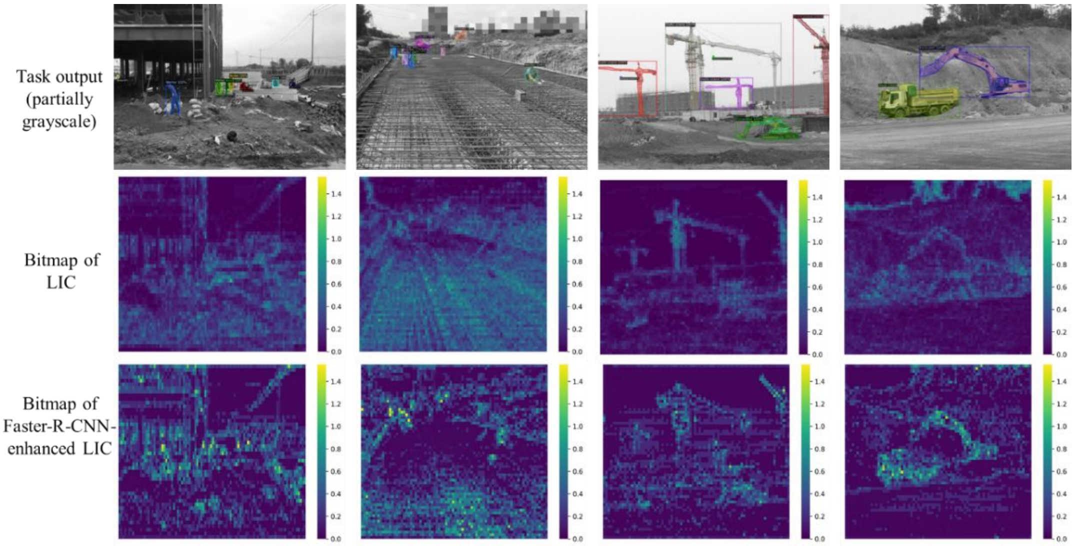

Task output (partially grayscale)
Bitmap of LIC
Bitmap of Faster-R-CNN-enhanced LIC
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

Figure 9: Example of reconstructed images and corresponding bitmaps during image compression

3.2 Quantitative Result

The quantitative assessment of the proposed approach primarily centers on the trade-off between compression efficiency and the accuracy of downstream tasks. The selected evaluation metrics include the following:

For image compression, the key performance indicator is bits per pixel (bpp), which indicates the average transmission load of the image. Higher bpp indicates lower compression ratio.

$$\text{bpp} = \frac{\text{File Size(in bits)}}{\text{Number of Pixels}} \tag{4}$$

For object detection, the primary metric for evaluation is:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 P_i(R)dR = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

where N is the number of object classes; $P_i(R)$ is the precision as a function of recall for class i; $\int_0^1 P_i(R)dR$ represents the area under the precision-recall curve for class i , which gives the AP for that class.

For instance segmentation, the main metrics for evaluation are:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{|T|}\sum_{t\in T}\int_0^1 P_i^t(R)dR\right) = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{6}$$

where N is the Number of object classes; T is set of IoU thresholds, typically ranging from 0.5 to 0.95 with a step of 0.05; $P_i^t(R)dR$ represents the precision at recall R for class i at threshold t.

Category	JPEG	LIC	Faster-R-CNN-enhanced LIC	Mask-R-CNN-enhanced LIC
bpp	3.0281	0.133	0.078	0.108
MSE	---	0.00116	0.00394	0.00291
mAP (object detection)	51.354	48.512	44.817	48.179
mAP (instance segmentation)	36.495	37.026	35.910	32.954

Table 3: Comparison of bpp for LIC and mAPs for downstream task

Task	Instance type	JPEG	LIC	Faster-R-CNN-enhanced LIC	Mask-R-CNN-enhanced LIC
object detection	Worker	52.157	50.995	47.071	45.481
	Excavator	65.875	65.724	61.961	58.231
	Static crane	48.995	46.78	43.608	37.046
	Truck	56.813	54.924	49.452	47.352
instance segmentation	Worker	44.528	43.481	39.398	36.509
	Excavator	41.557	41.207	39.773	37.643
	Static crane	26.148	26.079	22.283	20.803
	Truck	43.388	42.757	40.765	38.086

Table 4: Comparison of $\boldsymbol{AP}$ for different instance types

Table 3 presents a comparison of four methods—JPEG, LIC, Faster-R-CNN-enhanced LIC, and Mask-R-CNN-enhanced LIC—in terms of image compression efficiency and downstream task accuracy. The results indicate that the proposed method significantly improves data transmission efficiency compared to the original image, albeit with a minor reduction in task performance. When compared to traditional LIC models, the proposed approach substantially reduces the transmission load while maintaining similar levels of accuracy for downstream tasks, highlighting its effectiveness in increasing the density of task-relevant information in the bitstream. Table 4 further examines the detection accuracy of the four label types outlined in Table 2, reinforcing the findings and supporting the same conclusions.

When combined with the conclusions drawn from Figure 7, these results suggest that the proposed method reduces the overall data transmission load while increasing the proportion of relevant information (information density) in the transmitted data, with minimal impact on the accuracy of downstream tasks.

Moreover, the table compares the accuracy of semantic segmentation using Mask-R-CNN on images reconstructed by the Faster-R-CNN-enhanced LIC and the accuracy of object detection using Faster-R-CNN on images reconstructed by the Mask-R-CNN-enhanced LIC. The accuracy difference between these tasks is relatively small, indicating that the latent features extracted by the FPNs in different task models exhibit a notable degree of generalizability across different downstream tasks. This finding suggests the potential to further improve the reusability of modules within the proposed multi-task processing framework. Future work could focus on exploring strategies to further enhance this reusability to optimise calculation efficiency across multiple tasks.

4 Colclusion

The AEC sector is experiencing a transformation towards automation and intelligence. With rapid advancements in computer vision-based intelligent sensing, issues such as efficient data transmission and large-scale data storage have become increasingly important. However, current research on data transmission tends to focus either on optimizing existing hardware or on developing frameworks, with limited attention given to improving the information density of transmitted data.

To address this gap, this paper introduces two perspectives: LIC and machine vision. LIC offers a research pathway for effective data compression using end-to-end methods, while the machine vision perspective integrates data transmission and multiple downstream computer vision tasks within a unified framework.

Building upon these perspectives, this paper proposes a machine-vision-oriented LIC model that unifies LIC and machine-vision models through latent feature alignment. The proposed approach was validated using a construction site monitoring dataset, and the results demonstrate that the model focuses transmission on task-relevant areas, allocating fewer resources to irrelevant image regions. Consequently, the method achieves a substantial improvement in data transmission efficiency without significantly compromising downstream task accuracy.

ACKNOWLEDGMENTS

This research is funded by the Cardiff University – China Scholarship Council joint programme.

References

Alizadehsalehi, S., Hadavi, A., Huang, J.C., 2020. From BIM to extended reality in AEC industry. Automation in Construction 116, 103254. https://doi.org/10.1016/j.autcon.2020.103254

Ballé, J., Laparra, V., Simoncelli, E.P., 2017. End-to-end Optimized Image Compression.

Ballé, J., Minnen, D., Singh, S., Hwang, S.J., Johnston, N., 2018. Variational image compression with a scale hyperprior.

Chamain, L.D., Cheung, S.S., Ding, Z., 2019. Quannet: Joint Image Compression and Classification Over Channels with Limited Bandwidth, in: 2019 IEEE International Conference on Multimedia and Expo (ICME). Presented at the 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, Shanghai, China, pp. 338–343. https://doi.org/10.1109/ICME.2019.00066

Cheng, J.C.P., Chen, K., Chen, W., 2020. State-of-the-Art Review on Mixed Reality Applications in the AECO Industry. J. Constr. Eng. Manage. 146, 03119009. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001749

Cheng, Z., Sun, H., Takeuchi, M., Katto, J., 2020. Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules.

Grünwald, P., Roos, T., 2019. Minimum Description Length Revisited. Int. J. Math. Ind. 11, 1930001. https://doi.org/10.1142/S2661335219300018

Huang, C.-H., Wu, J.-L., 2024. Unveiling the Future of Human and Machine Coding: A Survey of End-to-End Learned Image Compression. Entropy 26, 357. https://doi.org/10.3390/e26050357

Le, N., Zhang, H., Cricri, F., Ghaznavi-Youvalari, R., Rahtu, E., 2021a. Image coding for machines: an end-to-end learned approach, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1590–1594. https://doi.org/10.1109/ICASSP39728.2021.9414465

Le, N., Zhang, H., Cricri, F., Ghaznavi-Youvalari, R., Tavakoli, H.R., Rahtu, E., 2021b. Learned Image Coding for Machines: A Content-Adaptive Approach, in: 2021 IEEE International Conference on Multimedia and Expo (ICME). Presented at the 2021 IEEE International Conference on Multimedia and Expo (ICME), IEEE, Shenzhen, China, pp. 1–6. https://doi.org/10.1109/ICME51207.2021.9428224

Liu, Z., Liu, T., Wen, W., Jiang, L., Xu, J., Wang, Y., Quan, G., 2018. DeepN-JPEG: A Deep Neural Network Favorable JPEG-based Image Compression Framework, in: 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). Presented at the 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC), IEEE, San Francisco, CA, pp. 1–6. https://doi.org/10.1109/DAC.2018.8465809

Minnen, D., Ballé, J., Toderici, G., 2018. Joint Autoregressive and Hierarchical Priors for Learned Image Compression.

Minnen, D., Singh, S., 2020. Channel-wise Autoregressive Entropy Models for Learned Image Compression.

Tuhaise, V.V., Tah, J.H.M., Abanda, F.H., 2023. Technologies for digital twin applications in construction. Automation in Construction 152, 104931. https://doi.org/10.1016/j.autcon.2023.104931

Wallace, G.K., 1992. The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics 38, xviii–xxxiv. https://doi.org/10.1109/30.125072

Wang, Shurun, Wang, Z., Wang, Shiqi, Ye, Y., 2021. End-to-end Compression Towards Machine Vision: Network Architecture Design and Optimization.

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., Wang, X., 2021. Computer Vision Techniques in Construction: A Critical Review. Arch Computat Methods Eng 28, 3383–3397. https://doi.org/10.1007/s11831-020-09504-3

Zhang, M., Xu, R., Wu, H., Pan, J., Luo, X., 2023. Human–robot collaboration for on-site construction. Automation in Construction 150, 104812. https://doi.org/10.1016/j.autcon.2023.104812

Zhang, Q., Liu, D., Li, H., 2017. Deep network-based image coding for simultaneous compression and retrieval, in: 2017 IEEE International Conference on Image Processing (ICIP). Presented at the 2017 IEEE International Conference on Image Processing (ICIP), pp. 405–409. https://doi.org/10.1109/ICIP.2017.8296312