Human-Machine Vision and Scalable Image Coding: A Survey

Abstract

In the rapidly evolving field of image processing, the integration of human visual cognition with machine-based image processing has emerged as a transformative approach. This survey paper explores key advancements in human-machine vision, focusing on scalable image coding, frequency decoupling, frequency invariance, collaborative vision systems, image compression, and visual perception. Traditional image compression methods often prioritize human perception over machine analysis, leading to a trade-off between image quality and computational efficiency. Recent advancements in deep learning methodologies have revolutionized image compression, optimizing both visual fidelity and recognition performance at reduced bitrates. The integration of Just Noticeable Difference (JND) principles into learned compression strategies has further enhanced perceptual quality while maintaining efficient compression rates. Techniques such as frequency decoupling and invariance have been pivotal in maintaining image quality across varying resolutions, addressing the challenges of scalable image coding. Hybrid and adaptive compression models, which combine traditional transformations with learned components, have also emerged as significant advancements in the field, offering potential for substantial improvements in image representation and compression efficiency. By aligning technological advancements with human perceptual standards and machine vision requirements, these methodologies are poised to play a crucial role in the future of image processing, ensuring high-quality visual data at reduced bitrates for diverse applications, including healthcare, smart cities, and autonomous vehicles. This survey paper provides a comprehensive overview of the current state of research in human-machine vision, scalable image coding, frequency decoupling, frequency invariance, collaborative vision systems, image compression, and visual perception, highlighting the transformative potential of these advancements in enhancing image processing systems.

1 Introduction

1.1 Integration of Human Visual Cognition and Machine-based Image Processing

The integration of human visual cognition with machine-based image processing represents a significant advancement in computer vision, striving to align human perceptual insights with machine computational capabilities. This synergy is essential for developing systems that efficiently process and interpret visual data, particularly in light of the increasing volume of visual information, as noted by Zhang [1].

Optimizing image compression for both human and machine vision tasks poses a challenge, as traditional methods often favor human perception over machine understanding, as discussed by Liu [2]. However, merging human visual perception with machine analysis can yield methods that satisfy both domains, as emphasized by Mao [3]. This is particularly relevant in image compression, where methods must accommodate diverse perceptual needs while ensuring efficiency.

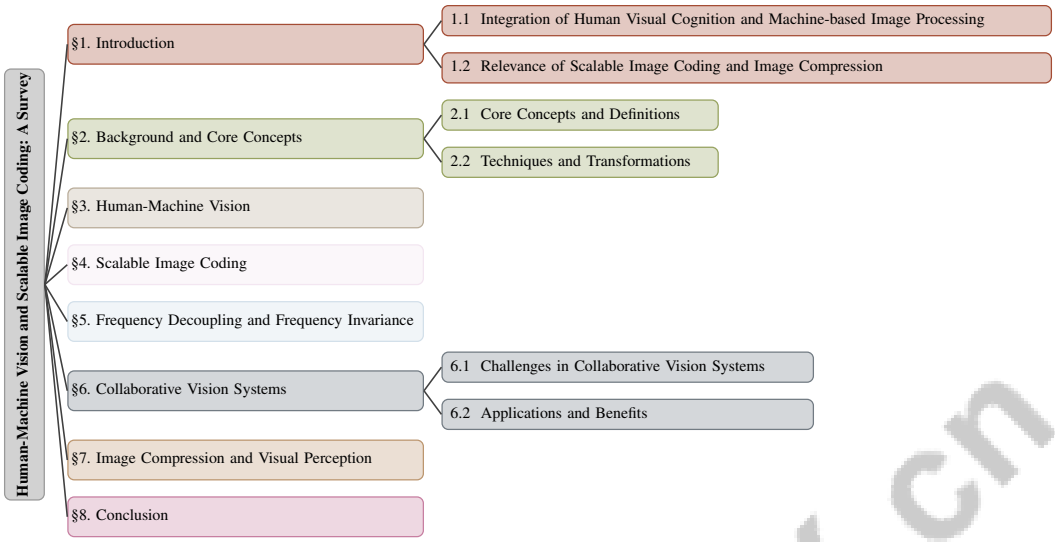


Figure 1: chapter structure

Bhown [4] highlights the necessity of prioritizing semantic content over pixel-level accuracy, advocating for the integration of human visual cognition into conventional lossy image compression techniques. Weber [5] further explores the relationship between human visual perception and machine classification accuracy, revealing enhancements in large-scale machine learning applications.

In deep neural network-based image compression, Torfason [6] discusses the potential for executing image understanding tasks directly on compressed representations. This aligns with the need for a codec that serves both human and machine purposes, as proposed by Chen [7], which combines high-level semantic information with low-level features.

The integration of human visual cognition with machine-based image processing is crucial for improving predictive accuracy in healthcare, as noted by Hu [8]. This integration facilitates scalable image coding methods that support both human verification and machine recognition, as highlighted by Shindo [9]. Kao [10] addresses the adaptation of compressed image latents for downstream vision tasks utilizing Multimodal Large Language Models (MLLMs), exemplifying improved efficiency in dataset annotation through innovative frameworks like that proposed by Zhang [11].

This integration enhances the efficiency and accuracy of image analysis by leveraging advanced algorithms and preprocessing techniques, ensuring that technological advancements align with human perceptual standards. This alignment is vital across various applications, including object detection, image compression, and visual artifact detection [12, 13, 14, 15]. The significance of this integration is particularly pronounced in fields such as medical imaging, surveillance, and multimedia applications, where a balance between automated processing and human interpretation is essential.

1.2 Relevance of Scalable Image Coding and Image Compression

Scalable image coding and image compression are fundamental to contemporary image processing, driven by the exponential growth of visual data and the need for efficient storage and transmission. Traditional compression techniques, including established standards like JPEG, often struggle to meet the dual demands of high fidelity and adaptability, particularly as high-resolution imagery becomes more prevalent [16]. The shift towards learning-based techniques marks a significant paradigm change, optimizing for visual distortion and recognition performance, which is increasingly critical as compressed images serve as inputs for deep learning tasks [17].

End-to-end optimized image compression methods represent a transformative approach, integrating high-level semantics into codecs to enhance perceptual quality, especially at low bitrates [18]. This approach is vital in applications such as smart cities and visual sensor networks, where efficient visual data compression is paramount [19]. The necessity to tailor image codecs for machine consumption, rather than traditional codecs designed for human viewers, highlights the differing requirements between human and machine vision [20].

Scalable image coding techniques are essential for adapting to varying channel bandwidths in modern networks, addressing existing methods' limitations regarding robustness to scale and illumination variations [21]. Optimizing JPEG image compression for deep learning tasks is particularly crucial in distributed learning systems, where bandwidth and storage constraints are significant considerations [22]. Additionally, the exploration of ultra-low bitrate compression, as discussed by Li [23], presents an underexplored area with potential for significant advancements.

The integration of attention mechanisms in transformer-based architectures, as explored by Luka [24], exemplifies the transition from traditional handcrafted codecs to learned image compression methods, offering promising avenues for enhancing compression efficiency. Furthermore, performing image classification and segmentation directly from compressed representations, as proposed by Torfason [6], presents opportunities for substantial computational savings by eliminating the need to decode images into RGB format.

The continuous advancement of scalable image coding and image compression techniques is crucial for supporting the growing demand for efficient image data processing across various applications. Meeting the requirements of both human perceptual standards and machine vision remains a critical goal, underscoring the need for methods that optimize both visual quality and recognition performance [7].The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Core Concepts and Definitions

The interdisciplinary field of human-machine vision and scalable image coding is grounded in key concepts that merge human visual cognition with machine processing. These include human-machine vision integration, scalable coding methodologies tailored to both human and machine visual needs, frequency decoupling for enhanced image processing, collaborative vision systems for optimized analytics, and sophisticated image compression strategies balancing fidelity and efficiency [25, 26, 8, 27].

Human-machine vision enhances applications like autonomous driving and medical imaging by combining human perceptual insights with machine capabilities. This integration is crucial for precision in complex tasks, exemplified by Object-Based Image Coding (OBIC) which efficiently represents arbitrarily shaped objects [28]. Scalable image coding, employing techniques that encode data at multiple resolutions, addresses diverse network conditions and application needs, prompting advancements like learned wavelet-like transforms [23]. Joint compression and classification frameworks integrate image compression with deep learning, enhancing both image reconstruction and machine vision performance [29].

Frequency decoupling maintains image quality across resolutions through frequency domain transformations, ensuring consistent performance despite frequency changes [30]. Optimizing rate-distortion performance in learned image compression highlights balancing entropy and quantization error [24]. Collaborative vision systems leverage human intuition and machine learning for data interpretation, facing challenges in resource allocation and feedback integration [6]. Integrating color component correlations in neural wavelet compression enhances coding efficiency.

Image compression aims to reduce bandwidth while preserving visual quality. Modern approaches balance compression rates and quality, particularly in high-frequency regions [16]. Learning-based compression methods address limitations in interpretability and scalability [31, 32, 33, 34, 35]. Visual perception prioritizes perceptual quality over pixel accuracy, optimizing for human perception to enhance compressed representations in tasks like segmentation and classification. This is crucial in applications such as human pose estimation and object detection, where lossy compression impacts performance. Frameworks for distributed image-to-image translation emphasize efficient image transformation across domains.

These core concepts are vital for advancing image processing systems that meet modern application demands while ensuring high visual fidelity and computational efficiency. Research into integrating local and non-local modeling in CNNs and transformers significantly enhances the rate-distortion performance of learned image compression methods. The Transformer-CNN Mixture (TCM) block, fusing local and non-local modeling, achieves state-of-the-art results across datasets [36, 37, 38, 17,

39]. Challenges remain in achieving high-quality compression at extremely low bitrates without losing semantic information.

2.2 Techniques and Transformations

Scalable image coding employs diverse techniques and transformations to enhance compression efficiency and quality, addressing human perceptual standards and machine vision needs. The integration of deep learning with traditional frameworks has been pivotal in this advancement. The LearntOBIC method exemplifies enhanced coding efficiency by optimizing the representation of arbitrarily shaped objects through segmentation and compression networks in an end-to-end learning framework [28].

Traditional transformation theory, including Block Discrete Cosine Transformation (BDCT) and Haar transformation, continues to inspire contemporary methods by modeling feature sparsity and improving reconstruction quality [23]. These transformations are foundational in developing modern coding paradigms balancing compression efficiency with high-fidelity reconstruction.

UniCompress utilizes wavelet transforms and knowledge distillation to efficiently compress multiple volumetric medical images using a single Implicit Neural Representation (INR) network [30]. This approach underscores frequency decoupling's significance in maintaining image quality across resolutions, a critical aspect of scalable coding.

In deep learning, the Direct Inference from Compressed Representations (DICR) method leverages compressed representations for image understanding tasks, eliminating the need for RGB decoding [6]. This highlights potential computational savings and the importance of aligning compression strategies with machine learning objectives.

The integration of CNNs and transformers through TCM blocks optimizes model complexity and performance in learned image compression, showcasing the synergy between local and non-local modeling capabilities [39]. The Cross-Component Context Model (CCM) enhances neural wavelet image coding by incorporating cross-component dependencies into the entropy modeling process, improving coding efficiency [40].

Adaptive resource management and parallel execution capabilities, as outlined in perceptually tuned enhanced datasets, are crucial for optimizing compression performance in dynamic environments [41]. These advancements in techniques and transformations are essential for meeting modern application demands, ensuring both human perceptual standards and machine vision requirements are effectively addressed.

The advancements in human-machine vision have been significantly influenced by the integration of machine learning and artificial intelligence, which have enhanced image processing capabilities. As illustrated in Figure 2, this figure depicts the hierarchical structure of these advancements, emphasizing the critical role of domain-specific knowledge in the evolution of human-machine vision systems. Furthermore, it highlights the synergy between human and machine vision, showcasing how efficient visual data processing can lead to improved outcomes in various applications, including autonomous vehicles and medical imaging. This integrated approach not only optimizes image processing tasks but also paves the way for innovative solutions in fields reliant on visual data interpretation.

3 Human-Machine Vision

3.1 Role of Machine Learning and AI

Machine learning (ML) and artificial intelligence (AI) significantly enhance human-machine vision systems by employing computational techniques that mimic human visual cognition, thereby improving image processing capabilities. Zhang's end-to-end optimized framework leverages human visual system properties to enhance image compression efficiency, especially in lower frequency domains where human sensitivity decreases [42]. Le's machine-specific image codecs, utilizing a neural network-based approach, prioritize task performance over visual fidelity, demonstrating ML's potential to tailor image processing strategies to machine vision needs [20]. Qi's JCC method illustrates ML's ability to boost classification performance while maintaining bandwidth efficiency, emphasizing the alignment of compression strategies with machine learning tasks [22].

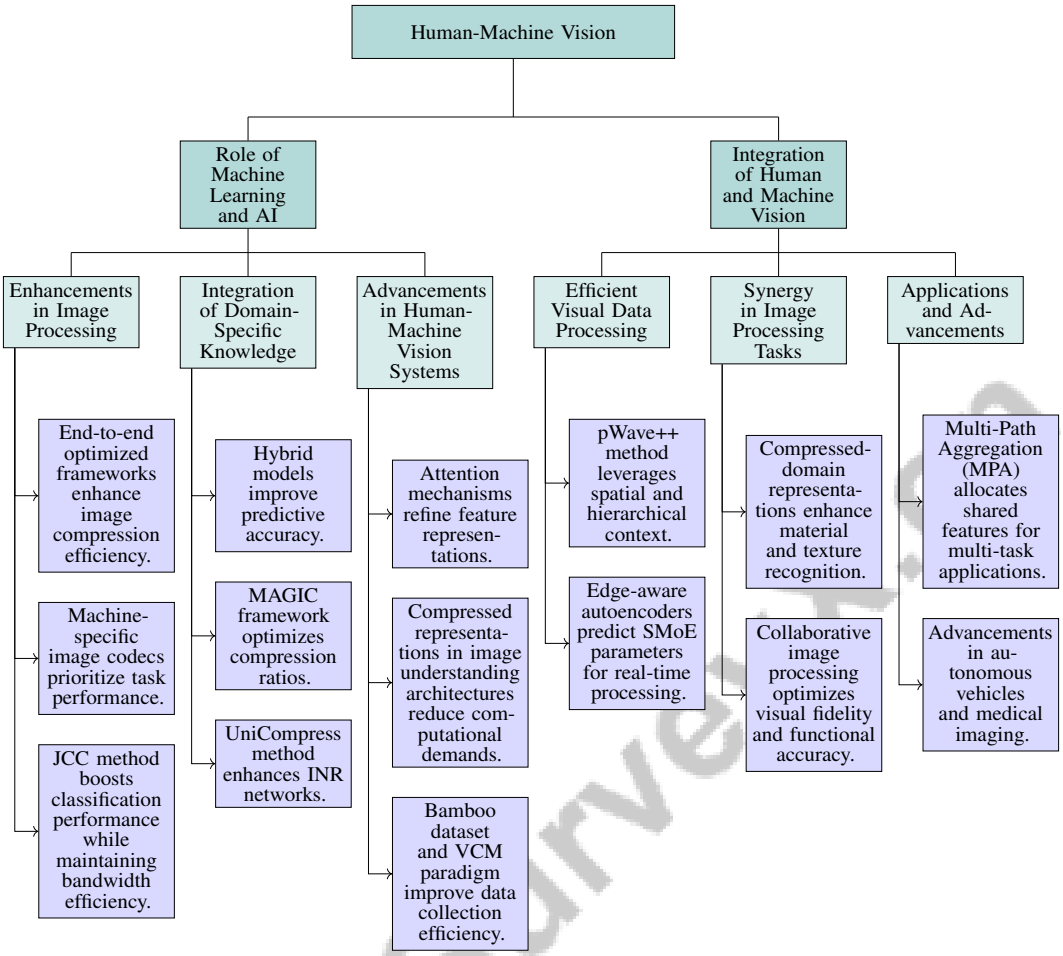


Figure 2: This figure illustrates the hierarchical structure of advancements in human-machine vision, highlighting the role of machine learning and AI in enhancing image processing, integrating domain-specific knowledge, and advancing human-machine vision systems. It also showcases the integration of human and machine vision, emphasizing efficient visual data processing, synergy in image processing tasks, and applications in autonomous vehicles and medical imaging.

Integrating domain-specific knowledge enhances ML technique performance, particularly in hybrid models aimed at improving predictive accuracy in sectors like healthcare. Advanced image compression methods, such as the MAGIC framework, leverage application-specific insights to optimize compression ratios while preserving essential features for machine-to-machine communication [11, 43, 13, 44]. The UniCompress method enhances Implicit Neural Representation (INR) networks by integrating discrete codebooks with frequency domain information, thereby improving system efficiency.

Attention mechanisms, as demonstrated by Luka, refine feature representations, reducing computational complexity and enhancing image analysis accuracy in human-machine vision systems [24]. Torfason's integration of compressed representations into image understanding architectures achieves performance comparable to traditional methods while reducing computational demands [6].

Recent ML and AI advancements in human-machine vision systems are exemplified by frameworks like the Bamboo dataset and the Video Coding for Machines (VCM) paradigm, which improve data collection efficiency and scalability while enhancing performance in downstream tasks. These developments facilitate effective integration of machine perception tasks with visual data compression, leading to more efficient and accurate image processing aligned with human perceptual standards and machine vision requirements [13, 8, 11, 27].

3.2 Integration of Human and Machine Vision

Integrating human and machine vision advances image processing by combining human perceptual insights with machine computational strengths, enabling efficient visual data processing and interpretation. The pWave++ method exemplifies this integration by leveraging spatial and hierarchical context from coded subbands, facilitating faster processing and reduced decoding times [45]. Fleig's edge-aware autoencoders further illustrate this integration by directly predicting SMoE parameters, eliminating time-consuming iterative optimization and enhancing real-time image processing [46].

Deng's discussion on learning-based image codecs highlights the integration of compressed-domain representations to enhance material and texture recognition without full decoding [47]. This synergy optimizes image processing tasks, allowing efficient analysis of complex visual data while maintaining high perceptual fidelity.

Advanced image coding frameworks utilizing compressive and generative models enhance collaborative image processing by optimizing visual fidelity for human perception and functional accuracy for machine analysis. This dual approach improves efficiency and facilitates seamless transitions between human-oriented and machine-oriented tasks, as demonstrated by methodologies like Multi-Path Aggregation (MPA), which intelligently allocates shared features for multi-task applications while minimizing parameter overhead [8, 48]. Merging human visual cognition with machine learning algorithms paves the way for advancements in applications such as autonomous vehicles and medical imaging, where precise and reliable image analysis is critical.

4 Scalable Image Coding

Category	Feature	Method
Learned Image Compression Techniques	Loss Optimization Strategies	JND-LC[49], HiFiC[50], JIQ[29]
	Model and Representation Utilization	DGML[51], SMoE-AE[46], DICR[6]
	Segmentation and Saliency Methods	LOBIC[28], NICE[16]
	Frequency and Scale Techniques	MSFDPM[52], E2E-FOT[42]
Integration of JND Principles in Learned Compression	JND Optimization	TSM[53], FGS[21]

Table 1: This table presents a comprehensive overview of recent advancements in learned image compression techniques and the integration of Just Noticeable Difference (JND) principles. It categorizes various methods based on their focus areas, such as loss optimization strategies, model utilization, segmentation techniques, and frequency methods, highlighting their contributions to enhancing compression efficiency and perceptual quality.

The burgeoning demand for efficient image coding has spotlighted scalable image coding as a pivotal area of research, emphasizing compression efficiency and perceptual quality. Table 1 provides a detailed summary of the cutting-edge methods in learned image compression and the application of JND principles, underscoring their role in advancing scalable image coding. Additionally, Table 2 provides a comprehensive comparison of learned image compression techniques and the integration of JND principles, emphasizing their role in advancing compression efficiency and perceptual quality in scalable image coding. This section explores innovative learned image compression techniques, leveraging deep learning to optimize visual data representation. These advancements not only redefine traditional paradigms but also align closely with human visual perception, deepening our understanding of their implications for modern image processing.

4.1 Learned Image Compression Techniques

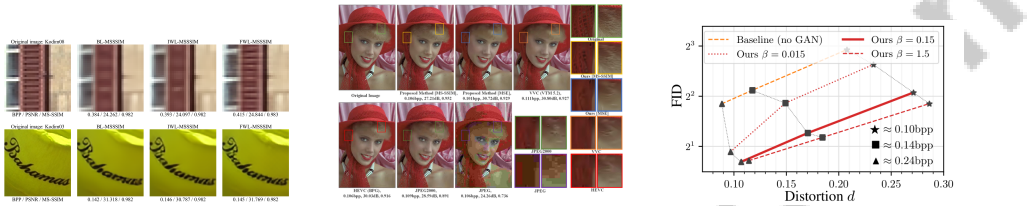
Recent strides in learned image compression, integrating deep learning methodologies, have significantly enhanced both efficiency and perceptual quality. Zhang's end-to-end optimized framework exemplifies this by leveraging human visual system properties, particularly in lower frequency domains, to optimize compression [42]. The LearntOBIC method further refines this by integrating segmentation and compression networks in an end-to-end framework, enhancing the representation of arbitrarily shaped objects [28].

Lee's use of a Gaussian Mixture Model (GMM) for entropy minimization improves probability estimation accuracy for latent representations, optimizing the balance between compression efficiency and image quality at lower bitrates [29]. Li's Neural Image Compression Explanation (NICE)

generates sparse masks to highlight salient pixels, creating mixed-resolution images for efficient storage [16].

Huang's MSFDPM method aligns and fuses multi-scale features from side information images, significantly improving image representation and compression efficiency [52]. The pWave++ method exploits spatial and hierarchical context from previously coded subbands, facilitating faster processing and reduced decoding times [46]. Torfason's integration of compressed representations into image understanding architectures achieves comparable performance to traditional methods while reducing computational complexity [6].

Attention mechanisms in learned image compression, as demonstrated by methods employing only attention layers, enhance compression efficiency and the representational power of compressed images [16].



(a) Comparison of Image Restoration Techniques Using MSSSIM Metrics[49]

(b) Comparison of Image Quality Metrics for Different Image Compression Techniques[51]

(c) Comparison of FID values for different values and distortion levels[50]

Figure 3: Examples of Learned Image Compression Techniques

Figure 3 illustrates the comparative analysis of various image compression methods. The first figure compares image restoration techniques using MSSSIM metrics, evaluating efficiency with metrics like Bit Per Pixel (BPP) and Peak Signal-to-Noise Ratio (PSNR). The second figure contrasts image quality metrics across compression techniques, emphasizing the effectiveness of proposed MS-SSIM and MSE techniques against traditional methods. The third figure plots the relationship between FID values and varying values across distortion levels, highlighting their impact on compression quality [49, 51, 50].

4.2 Integration of JND Principles in Learned Compression

Incorporating Just Noticeable Difference (JND) principles into image compression enhances the trade-off between efficiency and perceptual quality. By identifying JND-critical regions and utilizing a novel framework for assessing these differences, researchers have developed the KonJND++ dataset to inform compression algorithms. JND-based perceptual quality loss in learned models allows effective distortion allocation, leading to superior perceptual quality at equivalent bit rates compared to traditional methods [54, 55].

The Adaptive Importance Coding based on Transformations (AICT) framework by Bao utilizes a scale adaptation module to enhance compression efficiency, particularly at low bitrates, aligning with human visual system characteristics [53]. Zhai's Deep Fine-Grained Scalable Feature (DeepFGS) method exemplifies forward dependence between channels, ensuring smooth reconstruction quality increase as more features are decoded [21].

Bao's Transformation-based Scalable Model (TSM) interprets quantized representations in the frequency domain, enhancing compression by enabling nuanced handling of image details [53]. These methods underscore the potential of integrating JND principles in learned compression strategies, meeting demands for high-quality visual data at reduced bitrates. Recent innovations, such as text-guided encoding techniques, enhance pixel-level and perceptual quality by utilizing semantic information from text, addressing visual artifacts through targeted detection methods. Experimental results indicate these frameworks can outperform traditional codecs, ensuring a reliable and effective compression process [35, 56, 15].

AI-generated, for reference only.

Feature	Learned Image Compression Techniques	Integration of JND Principles in Learned Compression
Compression Efficiency	Enhanced Efficiency	High Efficiency
Perceptual Quality	Improved Perceptual Quality	Superior Perceptual Quality
Optimization Technique	Deep Learning Integration	Jnd Principles

Table 2: Comparison of learned image compression techniques and the integration of Just Noticeable Difference (JND) principles. The table highlights the differences in compression efficiency, perceptual quality, and optimization techniques, illustrating the advancements in scalable image coding. This comparison underscores the significance of deep learning and JND principles in enhancing image compression methodologies.

5 Frequency Decoupling and Frequency Invariance

5.1 Frequency Decoupling and Invariance in Image Processing

Frequency decoupling and invariance are crucial for scalable image coding, enhancing compression and transmission by separating image signals into distinct frequency bands. Recent innovations, such as frequency-oriented transforms and frequency-aware transformer models, have significantly improved rate-distortion performance and semantic fidelity [42, 57, 58]. These methods facilitate selective frequency component transmission and capture multiscale directional details, improving interpretability and overcoming traditional codec limitations. The strategic handling of frequency components preserves image quality across varying resolutions, essential for efficient storage and high perceptual fidelity.

The LearntOBIC method demonstrates effective frequency decoupling through local spatial correlations, optimizing arbitrarily shaped object representation while maintaining distinct structures at lower bitrates [28]. This approach aligns image processing systems with human perception and machine learning objectives.

Frequency invariance ensures consistent performance despite frequency changes, as exemplified by the UniCompress method, which compresses multiple volumetric medical images using a single Implicit Neural Representation (INR) network, leveraging wavelet transforms and knowledge distillation for enhanced efficiency [30]. The JCC method's use of Discrete Cosine Transform (DCT) coefficients highlights the importance of frequency domain transformations in maintaining image quality across resolutions [22]. These advancements are critical for high-precision applications, emphasizing the integration of compression strategies with machine learning goals, as shown by Torfason's work on embedding compressed representations into image understanding architectures to reduce computational complexity [6].

5.2 Hybrid and Adaptive Compression Models

Hybrid and adaptive compression models represent a significant advancement in image processing, particularly through frequency-oriented techniques that enhance compression efficiency and visual fidelity. Recent developments, such as end-to-end optimized models utilizing frequency-oriented transforms, outperform traditional codecs by effectively separating image signals into frequency bands, optimizing compression performance, and enabling selective frequency component transmission [42, 25, 58, 59, 60]. These models balance compression rate and image quality, ensuring high perceptual fidelity while minimizing storage needs.

Incorporating perceptual quality metrics into compression algorithms is a notable innovation. Pakdaman emphasizes perceptual learned image compression, prioritizing visual quality at lower bitrates [49]. By integrating perceptual metrics, these models preserve visually significant features crucial for human perception.

Jiang's introduction of a Just Noticeable Difference (JND)-based perceptual quality loss, replacing traditional MSE loss, allows dynamic distortion assignment based on perceptual thresholds, shifting focus from pixel-level accuracy to perceptual quality [61, 49].

Hybrid models, combining traditional and learned components, optimize image compression by integrating neural networks with classical transformation techniques. These models adapt to varying

image content and quality requirements, ensuring consistent performance across different resolutions and frequency bands.

Advancements in hybrid and adaptive models are essential for applications demanding high-quality visual data at reduced bitrates. By incorporating frequency-based methodologies and perceptual metrics, these models enhance compression efficiency and effectiveness. This alignment improves compressed image fidelity according to human perceptual standards and meets machine vision application needs. Frequency-oriented transforms facilitate scalable coding and preserve salient features, while perceptual similarity metrics tailored to human perception enhance model performance in visual tasks, surpassing traditional codecs and established metrics like MS-SSIM and PSNR [42, 31, 35].

6 Collaborative Vision Systems

6.1 Challenges in Collaborative Vision Systems

Collaborative vision systems, which merge human visual cognition with machine-based image processing, encounter several challenges that impede optimal performance. A primary concern is preserving image quality across multiple resolutions, essential for high fidelity and adaptability, but complicated by the integration of variable resolution methods into existing frameworks [62]. Saliency detection accuracy is critical; errors in saliency maps can result in suboptimal compression, potentially missing crucial features in non-salient regions [63, 64]. This issue is exacerbated by the demand for high-quality real-time data in applications requiring immediate processing [65].

Emerging learned compression techniques introduce additional challenges related to computational complexity, which varies with deployment environments and hardware capabilities, affecting real-time processing [66]. Robust complexity assessment benchmarks are necessary to ensure effective implementation [67]. The NLAIC method's complexity further complicates real-time applications, particularly in resource-constrained environments [38].

Incorporating multiple quality assessment models can improve prediction accuracy and resilience to image distortions [68]. However, achieving consistent performance across various visual tasks, especially in dynamic settings, remains challenging. The lack of a theoretical framework for the Rate-Distortion-Complexity (RDC) trade-off limits the development of efficient compression strategies [69].

Maintaining exemplar diversity while optimizing memory use is vital, particularly in real-world applications where efficient memory management is crucial [70]. Traditional methods struggle with image quality at varying resolutions, necessitating new strategies leveraging frequency decoupling and invariance [71].

Collaborative vision systems may also falter in scenarios with minimal view disparities [72], highlighting the need for compression methods adept at managing these disparities. The reliance on high-quality text descriptions, which might be inaccurate or unavailable, further complicates system deployment [73].

Addressing these challenges is crucial for the advancement of collaborative vision systems. Innovative frameworks like the Bamboo dataset, with its extensive annotations, can significantly enhance system efficiency and accuracy across diverse applications. Techniques such as Multi-Path Aggregation (MPA) optimize feature representation, facilitating seamless task transitions without performance compromise. These developments are key to creating high-quality datasets and effective models for real-world scenarios [11, 48].

6.2 Applications and Benefits

Collaborative vision systems, integrating human visual cognition with sophisticated machine-based image processing, offer substantial applications and benefits across diverse sectors. These systems enhance visual data interpretation and accuracy through methodologies like Multi-Path Aggregation (MPA) and Hybrid-Analyze-Then-Compress (HATC). MPA optimizes feature representation for varied tasks with minimal parameter overhead, while HATC balances visual content transmission with local feature representation, maximizing resource use and task efficiency. The development of large-scale datasets like Bamboo, utilizing active learning frameworks, enriches these systems by

providing high-quality annotations that bolster model performance in various tasks. This synergy between human and machine vision fosters a robust analytical framework for complex visual data interpretation [25, 26, 11, 48].

In image recognition, Shindo's adaptable approach [9] is applicable across models without specific optimizations, proving valuable for autonomous vehicles and security systems. The ability to classify directly on compressed representations, as shown by Deng [47], enhances efficiency and reduces computational needs, particularly advantageous in resource-limited environments.

Mital's approach [74] demonstrates potential improvements in image quality through effective side information leveraging via feature alignment, crucial for applications like remote sensing and environmental monitoring requiring high-quality visual data. This aligns with Yan's findings [75] on low computational complexity and high accuracy, making the method suitable for diverse applications without specific encoder dependencies.

In visual data compression, Li [76] showcases superior rate-accuracy performance compared to existing standards, validating its potential for real-world applications where efficient storage and transmission of high-quality visual data are critical. Kawawabeaudan [17] supports this by highlighting RALIC's advantages in achieving higher recognition accuracy at lower bitrates, making it efficient for recognition-prioritized tasks.

Chamain's joint optimization of codec and task models [77] demonstrates improved task performance through collaborative vision systems, allowing better adaptation to specific machine task needs, benefiting applications like autonomous navigation and smart city infrastructure.

Le's ICM method [20] illustrates efficient compression designed for machine consumption, achieving lower bitrates while maintaining high accuracy, crucial for applications requiring rapid and precise image processing, such as real-time decision-making in autonomous systems.

7 Image Compression and Visual Perception

7.1 Perceptual Quality and Image Compression Techniques

Modern image compression techniques increasingly focus on balancing compression rates with perceptual quality, a vital aspect in image processing. Traditional metrics like Peak Signal-to-Noise Ratio (PSNR) often fail to capture human perception nuances, necessitating advanced metrics for better alignment with perceived image quality [78]. Li's study on learned lifting-based transform methods underscores the need for metrics evaluating both reconstructed image fidelity and perceptual quality [79]. This dual emphasis ensures visual consistency alongside optimized rate-distortion performance.

The integration of Just Noticeable Difference (JND) principles in compression methods marks significant progress in enhancing visual quality and optimizing rate-distortion performance. Ding's JND-based optimization method achieves superior perceptual quality by effectively guiding distortion assignment [54]. The HFLIC method maintains high perceptual quality while achieving over 25% bitrate savings compared to prior methods [80]. Cai's approach allows for precise control over compression rates, ensuring image fidelity across multiple re-encodings without additional semantic data [81].

Kirmemis highlights the optimal perception-distortion trade-off for fixed bitrates, emphasizing the need to optimize both rate-distortion performance and perceptual quality [82]. Luo's work on context modeling in learned compression directly impacts perceptual quality, necessitating advanced methodologies for improved visual results [83]. Rafi's survey supports incorporating perceptual metrics, comparing Pulse-Coupled Neural Network (PCNN) applications across tasks [84].

Wu's PRFNet method captures multi-scale features, enhancing accuracy and visual outcomes over traditional methods [14]. Ball's approach achieves high visual quality at low bitrates with minimal computational complexity, crucial for applications like healthcare where balancing accuracy and complexity is essential [8].

Advancements in image compression prioritizing perceptual quality and efficiency are crucial for contemporary applications like streaming media and machine vision. Innovations such as the "Kuchen" framework enhance codecs with personalized preprocessing, improving performance.

Models integrating saliency and perceptual similarity metrics demonstrate traditional measures like MS-SSIM and PSNR inadequately reflect human perception. These developments result in visually superior images and enhance tasks like object detection and segmentation, leading to more efficient data transmission and storage solutions for both human and machine analysis [12, 31, 85]. By integrating perceptual metrics and advanced methodologies, these techniques ensure compressed images maintain high visual fidelity, benefiting both human and machine vision tasks.

7.2 Balancing Compression Rate and Perceptual Fidelity

Balancing compression rate and perceptual fidelity is essential in image compression, especially where storage efficiency and visual quality are critical. Recent advances in learned image compression focus on optimizing perceptual quality, achieving better visual outcomes at lower bitrates than traditional methods [49]. This optimization involves integrating perceptual metrics into the compression process, ensuring high fidelity to original content.

Mohammadi's results show perceptually optimized codecs produce images perceived as more similar to references, especially at lower bitrates, surpassing traditional methods that often compromise visual quality for compression efficiency [86]. By prioritizing perceptual quality, these codecs deliver images meeting human visual standards even at minimal sizes.

Ball's C3-WD method achieves a quality-rate trade-off comparable to generative methods with fewer computational resources, highlighting the importance of optimizing compression for perceptual fidelity and reduced computational demands, suitable for real-time and resource-constrained environments [87].

Luo's ELIC approach balances compression rate with perceptual quality by optimizing the context model and decoder within the learned compression framework, refining encoding and decoding processes to enhance compressed image quality, ensuring visual indistinguishability from originals [83].

8 Conclusion

8.1 Future Directions and Research Opportunities

The integration of human-machine vision and scalable image coding continues to evolve, offering vast potential for enhancing image processing systems. Key areas for future exploration include refining segmentation robustness and optimizing bit allocation strategies, which are vital for advancing both human-machine vision and scalable image coding. Enhancements in reconstruction quality and artifact minimization, alongside adaptations for diverse image data types, remain critical areas of focus.

Innovations in Implicit Neural Representation (INR) architectures and knowledge distillation processes are essential for boosting model generalizability across various medical image datasets. These developments promise versatile systems capable of handling a wide range of inputs effectively.

Research into improving the quality enhancement sub-network and assessing JointIQ-Net's performance on varied datasets presents promising directions. Such endeavors could foster the integration of machine learning techniques into image compression frameworks, yielding more resilient systems.

Expanding Torfason's method to encompass additional computer vision tasks and deepening the understanding of features learned by compression networks are crucial for broadening the applicability of learned image compression techniques. This could lead to significant breakthroughs in image-to-image translation and related fields.

Further investigation should concentrate on enhancing the prior model employed in QPressFormer and incorporating additional attention-based techniques to optimize image compression. These advancements will ensure that technological progress aligns with human perceptual standards.

Moreover, extending NICE's capabilities to areas such as text and bioinformatics, where interpretability is crucial, could amplify the applicability of learned image compression techniques. This expansion would ensure high recognition accuracy across various compression levels and broaden the methods' reach.

Addressing these research opportunities will propel the field of human-machine vision and scalable image coding towards continued innovation, leading to more effective and efficient image processing systems.

8.2 Applications and Implications for Image Compression

Progress in image compression techniques significantly impacts various sectors by improving the efficiency and quality of visual data processing. The Enhanced Deep Image Compression (EDIC) framework exemplifies these advancements, offering superior performance with faster decoding speeds and enhanced compression efficiency compared to traditional methods. This is particularly beneficial for real-time applications such as video streaming and telemedicine, where latency and bandwidth are critical constraints.

The Neural Image Compression Explanation (NICE) framework provides substantial advantages, operating significantly faster than conventional methods while delivering high-quality explanations. This efficiency makes NICE ideal for smart surveillance systems and autonomous vehicles, where rapid decision-making based on compressed image data is crucial.

In smart cities and Internet of Things (IoT) networks, the efficient compression and transmission of high-resolution imagery are essential for applications like traffic monitoring and environmental surveillance. By integrating learned image compression techniques with IoT devices, bandwidth utilization is optimized, ensuring high-quality visual data is available for real-time analysis and decision-making.

These advancements have profound implications in medical imaging, where high-quality image compression is crucial for accurate diagnosis and treatment planning. By optimizing compression techniques to maintain perceptual fidelity, medical professionals can rely on compressed images for detailed analysis without compromising diagnostic accuracy, particularly in telemedicine applications where efficient and reliable image transmission is essential.

Furthermore, advancements in image compression have the potential to reduce storage costs and energy consumption in data centers, as more efficient techniques decrease data size without sacrificing quality. This aligns with global sustainability goals by minimizing the environmental impact of data processing and storage.

The ongoing development of image compression technologies is set to play a pivotal role across various industries, ensuring that both human perceptual standards and machine vision requirements are met effectively. By addressing the challenges of balancing compression rates and perceptual fidelity, these advancements pave the way for more efficient and effective image processing systems, with extensive implications across multiple applications.

References

[1] Yuefeng Zhang, Chuanmin Jia, Jiannhui Chang, and Siwei Ma. Machine perception-driven image compression: A layered generative approach, 2023.

[2] Jinming Liu, Yuntao Wei, Junyan Lin, Shengyang Zhao, Heming Sun, Zhibo Chen, Wenjun Zeng, and Xin Jin. Tell codec what worth compressing: Semantically disentangled image coding for machine with lmms, 2024.

[3] Qi Mao, Chongyu Wang, Meng Wang, Shiqi Wang, Ruijie Chen, Libiao Jin, and Siwei Ma. Scalable face image coding via stylegan prior: Towards compression for human-machine collaborative vision, 2023.

[4] Ashutosh Bhown, Soham Mukherjee, Sean Yang, Shubham Chandak, Irena Fischer-Hwang, Kedar Tatwawadi, Judith Fan, and Tsachy Weissman. Towards improved lossy image compression: Human image reconstruction with public-domain images, 2019.

[5] Maurice Weber, Cedric Renggli, Helmut Grabner, and Ce Zhang. Observer dependent lossy image compression, 2020.

[6] Robert Torfason, Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, and Luc Van Gool. Towards image understanding from deep compression without decoding, 2018.

[7] Sien Chen, Jian Jin, Lili Meng, Weisi Lin, Zhuo Chen, Tsui-Shan Chang, Zhengguang Li, and Huaxiang Zhang. A new image codec paradigm for human and machine uses, 2021.

[8] Yueyu Hu, Shuai Yang, Wenhan Yang, Ling-Yu Duan, and Jiaying Liu. Towards coding for human and machine vision: A scalable image coding approach, 2020.

[9] Takahiro Shindo, Taiju Watanabe, Yui Tatsumi, and Hiroshi Watanabe. Scalable image coding for humans and machines using feature fusion network, 2024.

[10] Chia-Hao Kao, Cheng Chien, Yu-Jen Tseng, Yi-Hsin Chen, Alessandro Gnutti, Shao-Yuan Lo, Wen-Hsiao Peng, and Riccardo Leonardi. Bridging compressed image latents and multimodal large language models, 2025.

[11] Yuanhan Zhang, Qinghong Sun, Yichun Zhou, Zexin He, Zhenfei Yin, Kun Wang, Lu Sheng, Yu Qiao, Jing Shao, and Ziwei Liu. Bamboo: Building mega-scale vision dataset continually with human-machine synergy. *arXiv preprint arXiv:2203.07845*, 2022.

[12] Preprocessing enhanced image compression for machine vision.

[13] Vassilios S. Vassiliadis. The perceptron algorithm: Image and signal decomposition, compression, and analysis by iterative gaussian blurring, 2006.

[14] Kaiqun Wu, Xiaoling Jiang, Rui Yu, Yonggang Luo, Tian Jiang, Xi Wu, and Peng Wei. Progressive feature fusion network for enhancing image quality assessment, 2024.

[15] Daria Tsereh, Mark Mirgaleev, Ivan Molodetskikh, Roman Kazantsev, and Dmitriy Vatolin. Jpeg ai image compression visual artifacts: Detection methods and dataset, 2024.

[16] Xiang Li and Shihao Ji. Neural image compression and explanation, 2020.

[17] Maxime Kawawa-Beaudan, Ryan Roggenkemper, and Avideh Zakhor. Recognition-aware learned image compression, 2022.

[18] Yueyu Hu, Wenhan Yang, Zhan Ma, and Jiaying Liu. Learning end-to-end lossy image compression: A benchmark, 2021.

[19] Shurun Wang, Shiqi Wang, Wenhan Yang, Xinfeng Zhang, Shanshe Wang, Siwei Ma, and Wen Gao. Towards analysis-friendly face representation with scalable feature and texture compression, 2021.

[20] Nam Le, Honglei Zhang, Francesco Cricri, Ramin Ghaznavi-Youvalari, and Esa Rahtu. Image coding for machines: an end-to-end learned approach, 2021.

[21] Yongqi Zhai, Yi Ma, Luyang Tang, Wei Jiang, and Ronggang Wang. Deepfgs: Fine-grained scalable coding for learned image compression, 2024.

[22] Siyu Qi, Lahiru D. Chamain, and Zhi Ding. End-to-end optimization of jpeg-based deep learning process for image classification, 2023.

[23] Zhiyuan Li, Chenyang Ge, and Shun Li. Traditional transformation theory guided model for learned image compression, 2024.

[24] Natacha Luka, Romain Negrel, and David Picard. Image compression using only attention based neural networks, 2023.

[25] Luca Baroffio, Matteo Cesana, Alessandro Redondi, Marco Tagliasacchi, and Stefano Tubaro. Hybrid coding of visual content and local image features, 2015.

[26] Wenhan Yang, Haofeng Huang, Yueyu Hu, Ling-Yu Duan, and Jiaying Liu. Video coding for machine: Compact visual representation compression for intelligent collaborative analytics, 2021.

[27] Lingyu Duan, Jiaying Liu, Wenhan Yang, Tiejun Huang, and Wen Gao. Video coding for machines: A paradigm of collaborative compression and intelligent analytics. *IEEE Transactions on Image Processing*, 29:8680–8695, 2020.

[28] Qi Xia, Haojie Liu, and Zhan Ma. Object-based image coding: A learning-driven revisit, 2020.

[29] Jooyoung Lee, Seunghyun Cho, and Munchurl Kim. An end-to-end joint learning scheme of image compression and quality enhancement with improved entropy minimization, 2020.

[30] Runzhao Yang, Yinda Chen, Zhihong Zhang, Xiaoyu Liu, Zongren Li, Kunlun He, Zhiwei Xiong, Jinli Suo, and Qionghai Dai. Unicompress: Enhancing multi-data medical image compression with knowledge distillation, 2024.

[31] Yash Patel, Srikar Appalaraju, and R. Manmatha. Saliency driven perceptual image compression, 2020.

[32] Felipe Codevilla, Jean Gabriel Simard, Ross Goroshin, and Chris Pal. Learned image compression for machine perception, 2021.

[33] Shiyu Qin, Bin Chen, Yujun Huang, Baoyi An, Tao Dai, and Shu-Tao Xia. Perceptual image compression with cooperative cross-modal side information, 2023.

[34] Shiyu Duan, Huaijin Chen, and Jinwei Gu. Jpd-se: High-level semantics for joint perception-distortion enhancement in image compression, 2022.

[35] Chen-Hsiu Huang and Ja-Ling Wu. Exploring compressed image representation as a perceptual proxy: A study, 2024.

[36] Yuanchao Bai, Xu Yang, Xianming Liu, Junjun Jiang, Yaowei Wang, Xiangyang Ji, and Wen Gao. Towards end-to-end image compression and analysis with transformers, 2021.

[37] Renjie Zou, Chunfeng Song, and Zhaoxiang Zhang. The devil is in the details: Window-based attention for image compression, 2022.

[38] Tong Chen, Haojie Liu, Zhan Ma, Qiu Shen, Xun Cao, and Yao Wang. Neural image compression via non-local attention optimization and improved context modeling, 2019.

[39] Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-cnn architectures, 2023.

[40] Anna Meyer and André Kaup. A novel cross-component context model for end-to-end wavelet image coding, 2023.

[41] Shilv Cai, Xiaoguo Liang, Shuning Cao, Luxin Yan, Sheng Zhong, Liqun Chen, and Xu Zou. Powerful lossy compression for noisy images, 2024.

[42] Yuefeng Zhang and Kai Lin. End-to-end optimized image compression with the frequency-oriented transform, 2024.

[43] Kartik Gupta, Kimberley Faria, and Vikas Mehta. Learning-based image compression for machines, 2024.

[44] Prabuddha Chakraborty, Jonathan Cruz, and Swarup Bhunia. Leveraging domain knowledge using machine learning for image compression in internet-of-things, 2020.

[45] Anna Meyer, Srivatsa Prativadibhayankaram, and André Kaup. Efficient learned wavelet image and video coding, 2024.

[46] Elvira Fleig, Jonas Geistert, Erik Bochinski, Rolf Jongebloed, and Thomas Sikora. Edge-aware autoencoder design for real-time mixture-of-experts image compression, 2022.

[47] Yingpeng Deng and Lina J. Karam. Learning-based compression for material and texture recognition, 2021.

[48] Xu Zhang, Peiyao Guo, Ming Lu, and Zhan Ma. All-in-one image coding for joint human-machine vision with multi-path aggregation, 2024.

[49] Farhad Pakdaman, Sanaz Nami, and Moncef Gabbouj. Perceptual learned image compression via end-to-end jnd-based optimization, 2024.

[50] Fabian Mentzer, George Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression, 2020.

[51] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7939–7948, 2020.

[52] Yujun Huang, Bin Chen, Shiyu Qin, Jiawei Li, Yaowei Wang, Tao Dai, and Shu-Tao Xia. Learned distributed image compression with multi-scale patch matching in feature domain, 2022.

[53] Youneng Bao, Fangyang Meng, Wen Tan, Chao Li, Yonghong Tian, and Yongsheng Liang. Transformations in learned image compression from a modulation perspective, 2024.

[54] Feng Ding, Jian Jin, Lili Meng, and Weisi Lin. Jnd-based perceptual optimization for learned image compression, 2023.

[55] Guangan Chen, Hanhe Lin, Oliver Wiedemann, and Dietmar Saupe. Localization of just noticeable difference for image compression, 2023.

[56] Hagyeong Lee, Minkyu Kim, Jun-Hyuk Kim, Seungeon Kim, Dokwan Oh, and Jaeho Lee. Neural image compression with text-guided encoding for both pixel-level and perceptual fidelity, 2024.

[57] Ali Zafari, Atefeh Khoshkhahtinat, Piyush Mehta, Mohammad Saeed Ebrahimi Saadabadi, Mohammad Akyash, and Nasser M. Nasrabadi. Frequency disentangled features in neural image compression, 2023.

[58] Han Li, Shaohui Li, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Frequency-aware transformer for learned image compression, 2024.

[59] Oren Rippel and Lubomir Bourdev. Real-time adaptive image compression. In *International Conference on Machine Learning*, pages 2922–2930. PMLR, 2017.

[60] Rehna. V. J and Jeyakumar. M. K. Hybrid approaches to image coding: A review, 2012.

[61] Wei Jiang and Wei Wang. Image and video compression using generative sparse representation with fidelity controls, 2024.

[62] Shahrokh Paravarzar and Javaneh Alavi. A decade of research for image compression in multimedia laboratory, 2021.

[63] Xi Zhang and Xiaolin Wu. Attention-guided image compression by deep reconstruction of compressive sensed saliency skeleton, 2021.

[64] Kristian Fischer, Fabian Brand, Christian Blum, and André Kaup. Saliency-driven hierarchical learned image coding for machines, 2023.

[65] Siddharth Reddy, Anca D. Dragan, and Sergey Levine. Pragmatic image compression for human-in-the-loop decision-making, 2021.

[66] Xihua Sheng, Li Li, Dong Liu, and Houqiang Li. Vnvc: A versatile neural video coding framework for efficient human-machine vision, 2023.

[67] Farhad Pakdaman and Moncef Gabbouj. Comprehensive complexity assessment of emerging learned image compression on cpu and gpu, 2022.

[68] S. Farhad Hosseini-Benvidi, Hossein Motamednia, Azadeh Mansouri, Mohammadreza Raei, and Ahmad Mahmoudi-Aznaveh. Compressed image quality assessment using stacking, 2024.

[69] Yichi Zhang, Zhihao Duan, and Fengqing Zhu. On efficient neural network architectures for image compression, 2024.

[70] Justin Yang, Zhihao Duan, Andrew Peng, Yuning Huang, Jiangpeng He, and Fengqing Zhu. Probing image compression for class-incremental learning, 2024.

[71] Vijayaraghavan Thirumalai and Pascal Frossard. Correlation estimation from compressed images, 2011.

[72] Kedeng Tong, Xin Jin, Yuqing Yang, Chen Wang, Jinshi Kang, and Fan Jiang. Learned focused plenoptic image compression with microimage preprocessing and global attention, 2023.

[73] Xuhao Jiang, Weimin Tan, Tian Tan, Bo Yan, and Liquan Shen. Multi-modality deep network for extreme learned image compression, 2023.

[74] Nitish Mital, Ezgi Ozyilkan, Ali Garjani, and Deniz Gunduz. Neural distributed image compression with cross-attention feature alignment, 2023.

[75] Xiao Yan, Zhangxin Gong, Wenqiang Wang, Xiaoyang Zeng, and Yibo Fan. A high accuracy and low complexity quality control method for image compression, 2022.

[76] Binzhe Li, Shurun Wang, Shiqi Wang, and Yan Ye. High efficiency image compression for large visual-language models, 2024.

[77] Lahiru D. Chamain, Fabien Racapé, Jean Bégaint, Akshay Pushparaja, and Simon Feltman. End-to-end optimized image compression for machines, a study, 2020.

[78] Zhengxue Cheng, Pinar Akyazi, Heming Sun, Jiro Katto, and Touradj Ebrahimi. Perceptual quality study on deep learning based image compression, 2019.

[79] Xinyue Li, Aous Naman, and David Taubman. Exploration of learned lifting-based transform structures for fully scalable and accessible wavelet-like image compression, 2024.

[80] Peirong Ning, Wei Jiang, and Ronggang Wang. Hflic: Human friendly perceptual learned image compression with reinforced transform, 2023.

[81] Shilv Cai, Zhijun Zhang, Liqun Chen, Luxin Yan, Sheng Zhong, and Xu Zou. High-fidelity variable-rate image compression via invertible activation transformation, 2022.

[82] Ogun Kirmemis and A. Murat Tekalp. A practical approach for rate-distortion-perception analysis in learned image compression, 2021.

[83] Jixiang Luo. Rethinking learned image compression: Context is all you need, 2024.

[84] Nurul Rafi and Pablo Rivas. A review of pulse-coupled neural network applications in computer vision and image processing, 2024.

[85] Moqi Zhang, Weihui Deng, and Xiaocheng Li. A unified image preprocessing framework for image compression, 2022.

[86] Shima Mohammadi, Yaojun Wu, and João Ascenso. Fidelity-preserving learning-based image compression: Loss function and subjective evaluation methodology, 2024.

[87] Jona Ballé, Luca Versari, Emilien Dupont, Hyunjik Kim, and Matthias Bauer. Good, cheap, and fast: Overfitted image compression with wasserstein distortion, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.