Deep Restoration of Archival Videos: Developments, Challenges, and Opportunities

Shiv Gehlot, Sri Harsha Musunuri, Sutanu Bera, Aupendu Kar, Guan-Ming Su
Dolby Laboratories, Inc.
firstname.secondname@dolby.com, guanmingsu@ieee.org

***Abstract*—The visual quality of archival videos is shaped by degradations introduced across the stages of their lifecycle and corresponding enhancement methods. This survey consolidates research on restoration and enhancement techniques spanning the capture, content creation, content delivery, and content consumption stages. Beyond stage-specific approaches, the survey also reviews emerging all-in-one frameworks designed to address multiple degradations jointly. By unifying these perspectives, it offers insights for developing pipelines that enhance visual fidelity while preserving historical authenticity.**

***Index Terms*—Archival Content, Degradations, Enhancement**

I. Introduction

Archival videos, preserved across analog and early digital formats, represent invaluable historical and cultural records but often exhibit degradations accumulated throughout their lifecycle. These degradations primarily originate at three stages of the content pipeline: 1) capture, 2) content creation, and 3) content delivery, while the content consumption stage focuses on enhancing visual quality for modern viewing. At capture stage, limitations of early imaging devices and recording media introduce film grain, sensor noise, and color fading. During content creation, operations such as editing, duplication, or poor exposure contribute to blur, low-light degradation, and color inconsistencies. In delivery stage, repeated compression and format migration further compromise spatial and temporal fidelity. Finally, at consumption stage, contemporary enhancement techniques, such as frame interpolation, and super-resolution are applied to improve the visual experience and adapt archival content to current display standards.

Although degradations can be taxonomized by their origin within these stages, archival video rarely presents artifacts in such isolation. Instead, distortions frequently manifest as compound, mutually reinforcing degradations, for example, noise intertwined with blur, limiting the applicability of stage-specific solutions. These observations have motivated the emergence of all-in-one restoration frameworks, which aim to model heterogeneous, ambiguously sourced degradations within a unified formulation and deliver more reliable restoration performance under the uncertain and overlapping conditions. Consequently, this survey organizes the literature around stages-specific and all-in-one frameworks, reviewing representative algorithms along with their evolution. By unifying degradations and their restoration tasks into a single taxonomy, this work guides restoration pipelines that can improve perceptual quality while preserving archival authenticity.

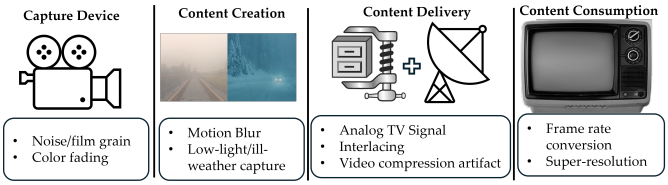

Fig. 1: Sample degradations at different stages.

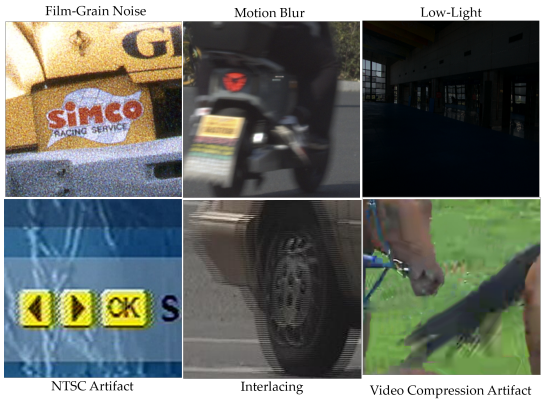

Fig. 2: Examples of representative degradations.

II. Video Content Degradations

Archival video quality is affected by a range of degradations that originate at different points in the content pipeline, each shaped by distinct physical and operational factors. To structure this variability, the following subsections outline the characteristic artifacts associated with each stage.

A. Capture Device

Degradations at the capture stage arise from the physical and technological limitations of early imaging systems and media, including film grain, sensor noise, and color fading.

1) Film Grain: Film grain is a stochastic, high-frequency artifact arising from the random distribution of silver halide crystals in analog film stock [1]. When digitized, it appears as temporally varying, signal-dependent noise with non-Gaussian statistics [2], making it significantly harder to suppress compared to stationary additive noise. It degrades perceived sharpness, impairs motion estimation, and reduces compression efficiency, while excessive removal risks oversmoothing details and diminishing intentionally preserved cinematic texture

979-8-3315-7111-5/26/$31.00 ©2026 IEEE

[3]. Classical solutions include adaptive spatial filtering and wavelet-domain shrinkage [4], which suppress noise while attempting to preserve edges [5]. Temporal filtering strategies exploit frame correlation but often blur dynamic content. Non-local means and block-matching approaches improve performance by leveraging self-similarity across frames [6]. Recent learning-based methods better capture the signal-dependent distribution of grain [7], employing encoder–decoder architectures, conditional GANs, and perceptual losses to balance grain removal with texture preservation [8], [9]. Additionally, neural grain-analysis models enable parameter estimation compatible with coding standards, improving compression while maintaining artistic intent [10].

2) Denoising: Denoising is a fundamental problem in video and image restoration, aiming to suppress noise from sensors, compression, or transmission while preserving structural and temporal fidelity. Early statistical methods relied on self-similarity: Buades *et al.* introduced Non-Local Means (NLM), averaging pixels over neighborhoods of similar patches to reduce noise without excessive blurring [11]. Dabov *et al.* extended this concept with BM3D, grouping similar 2D patches into 3D arrays and applying collaborative transform-domain filtering, which became the benchmark in image denoising [12]. For video, Maggioni *et al.* generalized BM3D into VBM4D, exploiting spatio-temporal redundancy with 4D transforms for significant improvements in temporal coherence [13]. With the rise of deep learning, Zhang *et al.* proposed DnCNN, a residual convolutional neural network trained to predict noise, establishing CNNs as flexible denoisers capable of handling diverse noise distributions [14]. For video, Tassano *et al.* introduced FastDVDnet, a flow-free, real-time architecture that processes consecutive frames jointly, achieving state-of-the-art trade-offs between quality and speed [15]. These approaches highlight the progression from handcrafted self-similarity priors to end-to-end learned mappings, while balancing computational complexity with restoration quality.

3) Colorization: Colorization, also known as gray-to-color conversion, is a transformative technique for restoring old archival footage that was originally captured in black and white. Direct gray-to-color transfer approaches develop model architectures to learn the mapping from grayscale input to color outputs (L channel to a, b chrominance channels in CIELab space). The integration of generative models [16] and modern architectures [17] addressed the need for diversity and vivid, photorealistic coloring. Transformer-based models [18] leverage the architecture's capacity for capturing long-range dependencies (global context) to overcome issues such as incorrect semantic colors and undersaturation. User-guided methods require the user to provide external assistance or conditions, such as strokes [19], hint points [20], to achieve customized results aligned with user preferences. Multimodal colorization approaches are often driven by reference color images or text prompts to describe the color of the objects present in the input grayscale video [21]. Reference-driven approaches learns how to select, propagate, and predict colors from large-scale data [22].

B. Content Creation

During content creation, attributes such as poor exposure introduce blur, low-light artifacts, and color inconsistencies.

1) Video Deblurring: Video deblurring aims to restore temporally sharp and spatially detailed frames degraded by motion or camera shake. The blur arises from temporal integration of scene motion during exposure, causing loss of fine details and motion discontinuities. Early approaches modeled this degradation through motion compensation and optical flow, followed by deconvolution or energy minimization, but these methods were fragile under severe or spatially varying blur [23], [24]. With deep learning, networks jointly modeled spatial detail and temporal coherence, while attention and feature-level correspondence improved robustness to complex motion [25], [26]. Transformer and state-space models further enhanced long-range temporal reasoning with real-time efficiency [27], [28]. Recent works explore multi-modal fusion with event-based sensors to recover high-frequency motion cues under fast motion and low light [29], [30]. Generative and diffusion-based frameworks have reframed deblurring as a conditional synthesis task, producing perceptually realistic and temporally stable outputs [31], [32]. Finally, domain generalization, test-time adaptation, and Mixture-of-Experts designs promote robustness and adaptability to diverse blur distributions in real-world deployment [33].

2) Exposure Correction: The quality of old archival footage suffers due to poor lighting, aging film stock, or limitations of analog recording. Enhancement algorithms can intelligently boost brightness while preserving details, avoiding overexposure. The Retinex theory [34] provides a fundamental framework for Low-Light Image Enhancement (LLIE) by modeling how human vision perceives color, serving as a basis for algorithms that aim to isolate and correct inadequate lighting. Early approaches utilize different priors, e.g., illumination prior [35] and reflectance prior [36]. To handle noise in low-light regions, a robust retinex model is developed to include an explicit noise term [36], or an additional denoising procedure applied to the reflectance map before final reconstruction [37] are generally used. Deep learning integrates the Retinex decomposition paradigm into neural networks, e.g., Retinex-Net [37], Retinexformer [38]. End-to-End deep learning methods utilize deep networks [39], [40], particularly CNNs or Transformers, to directly learn a mapping function from the low-light input image to the desired enhanced image. Zero-DCE [41] reformulates LLIE as an image-specific curve estimation problem. AnlightenDiff [42] explores the capability of diffusion models.

3) Color Transfer: Colors in early film stocks often feel unattractive compared to the era of motion pictures' color. This is due to limited dynamic range and narrow color gamuts, biased color reproduction, inconsistent color calibration across cameras, and poor color grading resulting from the limited availability of tools, lighting setups, and film development processes. In early works, researchers relied on classical

methods, matching global statistics such as means, variances, or full histograms of pixel colors. These techniques excelled at preserving content detail or good photorealism, but they could only transfer simple styles like tone curves and global color shifts, failing to capture regional complexity. The advent of deep learning brought powerful tools, but also a central conflict. Initial artistic style transfer algorithms leveraged the deep features of networks, particularly the VGG architecture, to encode and transfer style based on feature correlations (Gram matrices). However, when applied to photographs, this potent magic resulted in severe spatial distortions and unrealistic, painterly artifacts. DPST [43] addressed this by introducing a photorealism constraint. To overcome the efficiency limitations, subsequent works adopted a feed-forward and post-processing framework. These methods formulate stylization by directly matching feature correlations between content and style features using Whitening and Coloring Transform (WCT) [44]. The family of approaches that utilize WCT employs various techniques to enhance photorealism [45]. Recent work [46] has shown that the first layer of VGG is sufficient for color representation and has also demonstrated that semantic guidance helps produce photo-realistic, controlled output. A powerful parallel narrative emerged around deterministic color transformations, viewing style transfer not as a complex feature matching problem, but as a solvable mapping problem using Look-Up Tables (LUTs) [47].

C. Content Delivery

The delivery stage introduces degradations due to transmission format and video compression leading to format-specific and compression artifacts.

1) Analog TV Signal: Composite analog systems such as NTSC, PAL, and SECAM encode luminance and chrominance into a single signal for compatibility with monochrome displays. Because chroma is modulated onto a subcarrier and spectrally interleaved with luma, limited bandwidth and imperfect filtering inevitably introduce interference [48], manifested as dot crawl, cross-color, and cross-luminance artifacts. In NTSC, QAM-encoded chroma at 3.579545 MHz produces checkerboard edge patterns (dot crawl) and rainbowing in fine textures [49]. PAL reduces hue errors via line-wise phase alternation but still exhibits cross-luminance and color bleeding, especially under motion or saturation. SECAM's FM chroma avoids some crosstalk but introduces line-by-line color delay and reduced vertical chroma resolution. Suppression of such artifacts historically relied on comb filters and later 2D/3D adaptive filters exploiting spatial-temporal correlations [49]. However, these often blur detail or create motion-dependent artifacts, motivating modern learning-based methods that more effectively disentangle luma-chroma interference across frames.

2) Interlacing: Interlacing was introduced in broadcast television standards (e.g., NTSC, PAL) to reduce flicker and conserve bandwidth by transmitting two interleaved fields per frame, each capturing odd or even scanlines at successive time instants [50], [51]. This increases temporal resolution while halving spatial resolution, but introduces artifacts when motion is present. Characteristic degradations include *combing*, where motion between fields yields jagged edges, vertical resolution loss from naive line duplication, and motion aliasing due to field misalignment [50]. These artifacts are especially visible in dynamic content and may propagate to chroma channels due to subsampling. Classical deinterlacing strategies include *weaving* (field combination) and *bob* (line interpolation). Weaving preserves detail in static regions but fails under motion; bob maintains temporal smoothness at the cost of vertical detail. Motion-adaptive and edge-directed interpolation techniques attempt to balance these trade-offs by selectively applying interpolation depending on motion cues [52], [53]. Motion-compensated methods further estimate and compensate inter-field motion before merging, achieving higher fidelity at increased complexity [54]. Recent learning-based approaches leverage spatio-temporal priors. Zhu et al. pioneered deep CNNs for deinterlacing, reconstructing only missing scanlines while preserving known pixel values and exploiting temporal information from both fields [55]. Bernasconi et al. extended this by using multi-field fusion with residual dense blocks, improving reconstruction quality and reducing flickering [56]. Zhao et al. proposed a deinterlacing network employing cooperative vertical interpolation followed by motion-aware field merging with ghost suppression [57]. Multi-frame joint enhancement frameworks exploit deformable convolutions and recurrent modules to align and fuse multiple fields, improving temporal consistency [58].

3) Compression Artifact Reduction: To mitigate bandwidth constraints, lossy compression techniques (such as H.264/AVC [59] and H.265/HEVC [60]) are extensively employed to reduce bitrates while preserving acceptable visual quality. However, such schemes inevitably introduce visual artifacts, including blocking, blurring, and ringing, particularly under high compression ratios. Early strategies sought to mitigate these effects via deterministic post-filters embedded in codec loops, notably the in-loop deblocking and deringing filters. These approaches are computationally inexpensive and analytically interpretable but inherently limited. Later, the advent of deep learning has been increasingly applied to the restoration of compressed visual content [61], [62]. Although these methods aim to suppress compression artifacts, they typically optimize objective metrics such as PSNR or SSIM (in contrast to perceptual quality), which measure pixel-level accuracy and may not align with human perception. In this direction, GAN-based frameworks [63], [64] incorporate temporal information while focusing on perceptual quality enhancement. While GAN-based methods can recover plausible textures, diffusion models offer superior generative capabilities due to their iterative denoising process [65]. Hence, they have been successfully applied to various image and video restoration tasks [66], [67]. Collectively, these advancements reflect a paradigm shift toward perceptual fidelity, temporal coherence, and adaptive restoration strategies, paving the way for more visually convincing video enhancement systems. In this di-

rection, frameworks such as [68], [69] introduce a latent-diffusion based mechanism for video compression artifact reduction. The evolution of artifact reduction reveals a persistent tension between interpretability, perceptual realism, and computational efficiency. Classical filters offered analytical transparency, yet limited adaptivity; CNNs provided precision, but suppressed texture; generative models achieved realism at the expense of determinism and speed.

D. Content Consumption

At the consumption stage, restoration shifts towards enhancement and adaptation for modern viewing.

1) Video Frame Rate Conversion: Video Frame Rate Conversion (VFR) refers to the process of synthesizing new frames between existing ones in a video sequence to enhance temporal resolution, achieve smooth motion, enable slow-motion playback, or recover missing frames. This process is inherently ill-posed because temporal degradation arises when frames are sparsely sampled in time, leading to motion ambiguity, occlusions, large displacements, and appearance variations that cannot be directly inferred from neighboring frames. Early approaches primarily relied on optical flow–based warping techniques that estimated pixel-wise motion between input frames and warped them to generate intermediate frames [70]. While conceptually elegant, these methods were highly sensitive to flow estimation errors and struggled in the presence of occlusions, motion blur, or large displacements. Subsequent deep learning methods improved robustness by learning bidirectional flow and context-aware blending [71], [72], yet residual artifacts and temporal inconsistencies persisted due to imperfect motion modeling. To address these limitations, recent research has shifted from explicit flow estimation to dense correlation modeling and transformer-based motion reasoning, enabling networks to capture long-range dependencies and non-rigid motion without relying on explicit flow representations [73], [74]. In parallel, generative and diffusion-based frameworks have redefined VFR as a conditional synthesis problem, achieving high perceptual fidelity through motion-aware or patch-based diffusion processes [75], [76].

2) Video Super-Resolution: Video super-resolution leverages temporal correlations across frames to recover high-quality sequences from low-resolution inputs, facing fundamental challenges in balancing temporal consistency, detail generation, and computational efficiency. Early approaches relied on multi-frame fusion and motion-compensated filtering, where adjacent frames were registered using optical flow or block matching, followed by iterative reconstruction to enhance spatial resolution [77], [78]. These classical methods emphasized accurate motion estimation and regularization to suppress noise and aliasing, but struggled with complex motion, occlusions, and fine texture recovery. BasicVSR++ [79], employed bidirectional feature propagation with second-order grid propagation and flow-guided deformable alignment for flexible information aggregation across frames. RVRT [80] is a recurrent video restoration transformer that processes local frames in parallel within a globally recurrent framework, incorporating guided deformable attention for accurate clip-to-clip alignment while balancing model size and memory consumption. Zhou *et al.* advanced transformer-based approaches with MIA-VSR, introducing masked inter- and intra-frame attention mechanisms that exploit temporal continuity to reduce redundant computations while maintaining state-of-the-art accuracy [81]. Recent generative models have pushed the boundaries in detail synthesis [67], [82].

III. All-In-One Video Restoration

All-in-one frameworks seek to address the heterogeneous and uncertain degradation profiles characteristic of archival video by learning a unified representation that can accommodate multiple artifact types within a single model. These approaches rely on scalable generative priors, flexible conditioning mechanisms, and adaptive inference strategies to generalize across mixed or poorly specified degradation conditions. Within this paradigm, research has progressed along several complementary fronts: degradation synthesis for constructing broad training distributions, text-guided restoration for controllable model conditioning, zero-shot methods for adaptation without retraining, GAN-based architectures for holistic appearance modeling, and diffusion backbones for high-capacity video enhancement. Collectively, these developments reflect a shift toward integrated restoration systems capable of robust operation across diverse archival scenarios.

1) Degradation Synthesis: Realistic training data is crucial for generalization to real-world scenarios. Wang *et al.* introduced high-order degradation modeling in Real-ESRGAN, applying classical degradation processes iteratively-combining blur, resize, noise, and JPEG compression-to simulate complex real-world corruptions [83]. They further incorporated sinc filters to synthesize ringing and overshoot artifacts, enabling models trained purely on synthetic data to handle authentic degradations effectively. Building on this, Li *et al.* proposed AirNet, which uses contrastive learning to implicitly encode unknown degradation types without requiring explicit priors, demonstrating robust all-in-one restoration across denoising, deraining, and dehazing Overall, degradation synthesis provides the foundation for training unified models that can generalize across heterogeneous and compositionally complex artifact distributions. [84].

2) Text-Guided Restoration: Vision-language models have introduced flexible, human-centric control mechanisms for restoration. Conde *et al.* presented InstructIR, the first method to use natural language instructions for restoration, leveraging GPT-4 generated prompts and CLIP encoders to guide task-agnostic networks toward user-specified improvements [85]. Potlapalli *et al.* proposed PromptIR, which employs learnable prompt tokens to encode degradation-specific information, enabling all-in-one blind image restoration without requiring explicit degradation labels [86]. Qi *et al.* proposed SPIRE, enabling both semantic content prompts and quantitative degradation specifications within a diffusion framework for fine-grained restoration control [87]. Luo *et al.* introduced DA-CLIP, adapting pretrained CLIP models via a trainable

controller that predicts degradation features and pilots cross-attention modules for unified restoration [88]. In summary, text-guided restoration enables controllable, semantically informed adjustments that extend all-in-one frameworks beyond fixed degradation assumptions.

3) Zero-Shot Frameworks: Although supervised learning methods have demonstrated efficacy in addressing image/video restoration, their applicability is often constrained by the need to retrain to handle novel or unseen degradation patterns. This limitation has motivated the exploration of unsupervised and zero-shot paradigms that leverage the inherent structural regularities of natural images. Consequently, some frameworks employ GANs as implicit priors to model the distribution of natural images [89]. Alternatively, restoration frameworks have been developed based on denoiser-driven regularization, where pre-trained denoisers serve as image priors within iterative optimization schemes [90], [91]. Recently, diffusion models have emerged as powerful priors for image restoration, and recent frameworks [92], [93] leverage pretrained diffusion models for targeting blurring, missing pixels, compression artifacts, or composite degradations. For video data, [94], [95] introduce different fusion or temporal-alignment strategies to obtain temporally-consistent enhanced video sequences in zero-shot setting. While these methods demonstrate flexibility and generalization, they remain computationally demanding due to iterative sampling or explicit degradation model requirements. Nevertheless, this paradigm establishes a promising bridge between powerful pretrained diffusion priors and real-world video restoration, which may have potential for developing a universally adaptive framework.

4) GANs for Video Restoration: Adversarial learning has been a principal instrument for promoting perceptual realism in image and video restoration. The seminal SRGAN work demonstrated that a perceptual loss combined with a generative adversarial objective yields significantly improved single-image super-resolution, motivating subsequent research to focus on texture fidelity rather than pixel fidelity [83], [96]. Transferring adversarial objectives to video necessitates explicit temporal regularization. TecoGAN [97] introduced temporally coherent adversarial training by employing spatio–temporal discriminators and recurrent generators, and by proposing the Ping-Pong loss to mitigate error accumulation in recurrent propagation. Recently, [98] proposed VideoGigaGAN, which adapts GigaGAN [99], an image upsampler for video data through a temporal module and flow-guided feature propagation. Despite their perceptual strengths, GAN-based frameworks face persistent challenges such as training stability. Further, as diffusion models have shown better generative capabilities as compared to GANs, the recent frameworks utilize former for robust enhancement frameworks.

5) Pretrained Diffusion Backbones for Video Enhancement: A parallel research line investigates adapting large text-to-image (T2I) and text-to-video (T2V) diffusion models for video restoration. Instead of training diffusion networks from scratch on restoration datasets, these methods leverage the expressive priors learned by generative backbones, such as Stable Diffusion or video diffusion transformers, and repurpose them through conditioning or lightweight adaptation. This paradigm seeks to combine the semantic richness and texture realism of large generative models with restoration-specific conditioning, thereby reducing data requirements and improving perceptual quality and generalization. Upscale-A-Video [67] pioneers this direction by adapting an image-trained latent diffusion model for real-world video super-resolution. Similarly, STAR [100] extends pretrained text-to-video backbones (I2V and CogVideoX) for video enhancement. It complements the global attention block with a local information enhancement module and utilizes dynamic frequency loss during training. Further, DOVE [101] builds upon CogVideoX to achieve remarkable performance on the real-world degraded videos. Effectively, it utilizes a two-stage (latent-pixel) training strategy to adapt the video generation model for the video enhancement task. Overall, leveraging pretrained diffusion backbones enables restoration frameworks to inherit rich generative structure and robust generalization without extensive retraining.

IV. Challenges and Opportunities

Despite rapid progress in restoration and enhancement, archival video remains uniquely challenging due to its diverse, compound degradations and the need to preserve both perceptual quality and historical authenticity. We highlight key challenges and corresponding opportunities for future research.

1. Heterogeneous and interacting degradations. Archival footage often contains multiple, interdependent artifacts spanning capture, creation, and delivery stages, making joint modeling difficult. Opportunities lie in degradation-aware architectures and causal models that disentangle and reason over mixed degradations.

2. Scarcity of representative datasets. Annotated archival datasets remain limited due to rights, privacy, and curation costs. Building ethically accessible archives and hybrid real-synthetic datasets can support scalable benchmarking and improve generalization.

3. Faithfulness vs. perceptual enhancement. Restoration must enhance visual quality without altering historical content. Developing fidelity-preserving objectives can help maintain authenticity while improving perceptual appeal.

4. Evaluation and benchmarking. Existing metrics fail to capture perceptual and historical fidelity. Domain-specific evaluation protocols, expert-in-the-loop assessment, and long-range temporal metrics are needed.

5. Human-centered and multimodal guidance. Integrating textual cues, archival notes, and curator input can promote semantically consistent and historically aware restoration, with multimodal conditioning offering promising avenues.

These challenges underscore the need for algorithmically robust, data-driven, and heritage-aware restoration frameworks. Advancing physically grounded degradation models alongside adaptive learning architectures will support scalable and reproducible pipelines that preserve archival authenticity while

enabling high-quality, analyzable, and interoperable video content for research, preservation, and public dissemination.

V. Conclusion

This survey systematically reviewed degradations and restoration methodologies across different stages of archival video, along with emerging frameworks for joint degradation modeling. By consolidating restoration tasks, and learning paradigms, it traced the evolution from physically inspired filtering to modern foundational models. The findings highlight a shift toward data-driven, physically grounded, and scalable restoration frameworks. Future progress will depend on integrating explicit degradation modeling, multimodal priors, and adaptive learning mechanisms to achieve restoration pipelines that are both computationally reliable and historically faithful.

References

[1] A. C. Kokaram, "Detection and removal of film dirt and scratches in archived film sequences," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1993, pp. 473–476.

[2] X. Li, "Noise modeling and estimation for video processing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 574–579, 2004.

[3] A. C. Kokaram, *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*. Springer, 1998.

[4] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[5] T.-W. Huang *et al.*, "Film grain removal using metadata," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 1520–1524.

[6] A. Buades *et al.*, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. 60–65.

[7] S. H. Musunuri *et al.*, "Spatio-temporal film grain removal in video using attention-guided neural networks," in *IEEE International Conference on Multimedia Information Processing and Retrieval*, 2025.

[8] Z. Ameur *et al.*, "Deep-based film grain removal and synthesis," *IEEE Trans. Image Process.*, vol. 32, pp. 5046–5059, 2023.

[9] J. Liang *et al.*, "Flow-based video denoising with consistent texture preservation," in *Proc. IEEE Int. Conf. Computer Vision*, 2021, pp. 5436–5445.

[10] Z. Ameur *et al.*, "FGA-NN: Film grain analysis neural network," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2025, pp. 1–5.

[11] A. Buades *et al.*, "A non-local algorithm for image denoising," in *CVPR*, 2005, pp. 60–65.

[12] K. Dabov *et al.*, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[13] M. Maggioni *et al.*, "Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3952–3966, 2012.

[14] K. Zhang *et al.*, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[15] M. Tassano *et al.*, "Fastdvdnet: Towards real-time deep video denoising without flow estimation," in *CVPR*, 2020, pp. 1354–1363.

[16] Y. Wang *et al.*, "Palgan: Image colorization with palette generative adversarial networks," in *European Conference on Computer Vision*. Springer, 2022, pp. 271–288.

[17] R. Zhang *et al.*, "Colorful image colorization," in *European conference on computer vision*. Springer, 2016, pp. 649–666.

[18] S. Weng *et al.*, "Ct 2: Colorization transformer via color tokens," in *European Conference on Computer Vision*. Springer, 2022, pp. 1–16.

[19] H. Lyu *et al.*, "Lga-net: Learning local and global affinities for sparse scribble based image colorization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 8144–8153.

[20] J. Yun *et al.*, "icolorit: Towards propagating local hints to the right region in interactive colorization by leveraging vision transformer," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 1787–1796.

[21] Z. Huang *et al.*, "Unicolor: A unified framework for multi-modal colorization with transformer," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 1–16, 2022.

[22] M. He *et al.*, "Deep exemplar-based colorization," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–16, 2018.

[23] J. Pan *et al.*, "Blind image deblurring using dark channel prior," in *CVPR*, 2016.

[24] ——, "Cascaded deep video deblurring using temporal sharpness prior," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3043–3051.

[25] D. Li *et al.*, "Arvo: Learning all-range volumetric correspondence for video deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7721–7731.

[26] S. Zhou *et al.*, "Spatio-temporal filter adaptive network for video deblurring," in *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[27] M. Suin and A. Rajagopalan, "Gated spatio-temporal attention-guided video deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7802–7811.

[28] K. Zhang *et al.*, "Stripformer: Strip transformer for fast image deblurring," in *European Conference on Computer Vision (ECCV)*, 2022.

[29] Z. Jiang *et al.*, "Learning event-based motion deblurring," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3320–3329.

[30] Y. Chen *et al.*, "Event-based video reconstruction via exploiting complementary information for motion deblurring," in *IEEE Transactions on Image Processing*, 2022.

[31] H. Long *et al.*, "Divd: Deblurring with improved video diffusion model," *arXiv preprint arXiv:2412.00773*, 2024.

[32] J. Ho *et al.*, "Video deblurring with conditional diffusion models," in *Neural Information Processing Systems (NeurIPS)*, 2022.

[33] D. Feijoo *et al.*, "Towards unified image deblurring using a mixture-of-experts decoder," *arXiv preprint arXiv:2508.06228*, 2025.

[34] B. Cai *et al.*, "A joint intrinsic-extrinsic prior model for retinex," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4000–4009.

[35] X. Ren *et al.*, "Lr3m: Robust low-light enhancement via low-rank regularized retinex model," *IEEE Transactions on Image Processing*, vol. 29, pp. 5862–5876, 2020.

[36] M. Li *et al.*, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE transactions on image processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[37] C. Wei *et al.*, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[38] Y. Cai *et al.*, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12 504–12 513.

[39] C. Li *et al.*, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern recognition letters*, vol. 104, pp. 15–22, 2018.

[40] T. Wang *et al.*, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 3, 2023, pp. 2654–2662.

[41] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *CVPR*, 2020, pp. 1780–1789.

[42] C.-Y. Chan *et al.*, "Anlightendiff: Anchoring diffusion probabilistic model on low light image enhancement," *IEEE Transactions on Image Processing*, 2024.

[43] F. Luan *et al.*, "Deep photo style transfer," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4990–4998.

[44] Y. Li *et al.*, "Universal style transfer via feature transforms," *Advances in neural information processing systems*, vol. 30, 2017.

[45] T.-Y. Chiu and D. Gurari, "Photowct2: Compact autoencoder for photorealistic style transfer resulting from blockwise training and skip connections of high-frequency residuals," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2868–2877.

[46] A. Kar and G.-M. Su, "Temporal consistent semantic video color transfer from multiple references," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 6207–6215.

[47] Y. Chen *et al.*, "Nlut: Neural-based 3d lookup tables for video photo-realistic style transfer," *arXiv preprint arXiv:2303.09170*, 2023.

[48] E. Dubois, "Sampling and reconstruction of NTSC video signals at twice the color subcarrier frequency," *IEEE Trans. Commun.*, vol. 29, no. 12, pp. 1823–1828, 1979.

[49] J. W. Lee *et al.*, "Reduction of dot crawl and rainbow artifacts in the NTSC video," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 740–748, 2007.

[50] C. Poynton, *Digital Video and HDTV: Algorithms and Interfaces*. Morgan Kaufmann, 2003.

[51] International Telecommunication Union, "ITU-R BT.601: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios," 2011, recommendation ITU-R BT.601-7.

[52] X. Li and M. T. Orchard, "Edge-directed interpolation for deinterlacing," in *Proc. Int. Conf. Image Process.*, vol. 3, 2001, pp. 690–693.

[53] S. Traverso *et al.*, "Motion adaptive deinterlacing based on edge pattern analysis," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1423–1431, 2009.

[54] Z. Zhang *et al.*, "Motion-compensated deinterlacing using adaptive edge-oriented interpolation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1037–1049, 2005.

[55] H. Zhu *et al.*, "Real-time deep video deinterlacing," *arXiv preprint arXiv:1708.00187*, 2017.

[56] M. Bernasconi *et al.*, "Deep deinterlacing," in *SMPTE Annual Tech. Conf. Exhib.*, 2020.

[57] Y. Zhao *et al.*, "Rethinking deinterlacing for early interlaced videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4872–4878, 2022.

[58] Y. Chen *et al.*, "MFDIN: Multi-frame joint enhancement for video deinterlacing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021.

[59] T. Wiegand *et al.*, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[60] G. J. Sullivan *et al.*, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[61] Y. Dai *et al.*, "A convolutional neural network approach for post-processing in hevc intra coding," in *MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part I 23*. Springer, 2017, pp. 28–39.

[62] Y. Xu *et al.*, "Boosting the performance of video compression artifact reduction with reference frame proposals and frequency domain information," in *CVPR*, 2021, pp. 213–222.

[63] S. Zhang *et al.*, "Dcngan: A deformable convolution-based gan with qp adaptation for perceptual quality enhancement of compressed video," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 2035–2039.

[64] J. Wang *et al.*, "Mw-gan+ for perceptual quality enhancement on compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4224–4237, 2021.

[65] J. Ho *et al.*, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[66] Y. Li *et al.*, "Tdm: Temporally-consistent diffusion model for all-in-one real-world video restoration," in *International Conference on Multimedia Modeling*. Springer, 2025, pp. 155–169.

[67] S. Zhou *et al.*, "Upscale-A-Video: Temporal-consistent diffusion model for real-world video super-resolution," in *CVPR*, 2024, pp. 14 766–14 776.

[68] Y. Wang *et al.*, "Pixrevive: Latent feature diffusion model for compressed video quality enhancement," in *CVPR*, 2024, pp. 12 345–12 356.

[69] S. Gehlot and G.-M. Su, "Lvicar: Diffusion models for perceptual quality enhancement in video compression artifact reduction," in *Proceedings of the ACM Multimedia Workshop on Deep Multimodal Generation and Retrieval (MMGR) / ACM Multimedia Workshops*, 2025, pp. 1–8.

[70] S. Niklaus *et al.*, "Video frame interpolation via adaptive separable convolution," in *ICCV*, 2017.

[71] S. Niklaus and F. Liu, "Context-aware synthesis for video frame interpolation," in *CVPR*, 2018.

[72] W. Bao *et al.*, "Depth-aware video frame interpolation," in *CVPR*, 2019.

[73] Z. Li *et al.*, "Amt: All-pairs multi-field transforms for efficient frame interpolation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9801–9810.

[74] H. Sim *et al.*, "Xvfi: Cross-space video frame interpolation transformer," in *CVPR*, 2023.

[75] Z. Huang *et al.*, "Motion-aware latent diffusion models for video frame interpolation," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 1043–1052.

[76] J.-H. Lee *et al.*, "Videodiff: Diffusion-based video frame interpolation with temporal consistency," in *European Conference on Computer Vision (ECCV)*, 2024.

[77] M. Irani and S. Peleg, "Improving resolution by image registration," in *CVGIP: Graphical Models and Image Processing*, 1991, pp. 231–239.

[78] S. C. Park *et al.*, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[79] K. C. K. Chan *et al.*, "BasicVSR++: Improving video super-resolution with enhanced propagation and alignment," in *CVPR*, 2022.

[80] J. Liang *et al.*, "Recurrent video restoration transformer with guided deformable attention," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022, pp. 378–393.

[81] X. Zhou *et al.*, "Video super-resolution transformer with masked inter&intra-frame attention," in *CVPR*, 2024, pp. 13 398–13 407.

[82] Y. Xu *et al.*, "VideoGigaGAN: Towards detail-rich video super-resolution," *arXiv preprint arXiv:2404.12388*, 2024.

[83] X. Wang *et al.*, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 1905–1914.

[84] B. Li *et al.*, "All-in-one image restoration for unknown corruption," in *CVPR*, 2022, pp. 17 452–17 462.

[85] M. V. Conde *et al.*, "InstructIR: High-quality image restoration following human instructions," in *ECCV*, 2024, pp. 15–31.

[86] V. Potlapalli *et al.*, "PromptIR: Prompting for all-in-one blind image restoration," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[87] C. Qi *et al.*, "SPIRE: Semantic prompt-driven image restoration," *arXiv preprint arXiv:2312.11595*, 2023.

[88] Z. Luo *et al.*, "Controlling vision-language models for universal image restoration," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024.

[89] J. Gu *et al.*, "Image processing using multi-code gan prior," in *CVPR*, 2020, pp. 3012–3021.

[90] W. Dong *et al.*, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, 2018.

[91] Y. Romano *et al.*, "The little engine that could: Regularization by denoising (red)," *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[92] B. Fei *et al.*, "Generative diffusion prior for unified image restoration and enhancement," in *CVPR*, 2023, pp. 9935–9946.

[93] D. Y. Lee *et al.*, "Sd-urm: Stable-diffusion based zero-shot universal restoration model," in *Proceedings of the IEEE International Conference on Image Processing (ICIP) Workshops*, 2025.

[94] C.-H. Yeh *et al.*, "Diffir2vr-zero: Zero-shot video restoration with diffusion-based image restoration models," *arXiv preprint arXiv:2407.01519*, 2024.

[95] S. Gao and et al., "Ditvr: Diffusion transformer for zero-shot video restoration," *arXiv preprint arXiv:2508.07811*, 2025.

[96] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 4681–4690.

[97] M. Chu *et al.*, "Tecogan: Temporally coherent gans for video super-resolution," *arXiv preprint arXiv:1811.09393*, 2018.

[98] Y. Xu *et al.*, "Videogigagan: Towards detail-rich video super-resolution," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 2139–2149.

[99] M. Kang *et al.*, "Scaling up gans for text-to-image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 10 124–10 134.

[100] R. Xie *et al.*, "Star: Spatial-temporal augmentation with text-to-video models for real-world video super-resolution," *arXiv preprint arXiv:2501.02976*, 2025.

[101] Z. Chen *et al.*, "Dove: Efficient one-step diffusion model for real-world video super-resolution," *arXiv preprint arXiv:2505.16239*, 2025.