Novelty Assessment Report

Paper: LiveMoments: Reselected Key Photo Restoration in Live Photos via Reference-guided Diffusion
PDF URL: https://openreview.net/pdf?id=02mgFnnfqG
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-05

Abstract

Live Photo captures both a high-quality key photo and a short video clip to preserve the precious dynamics around the captured moment. While users may choose alternative frames as the key photo to capture better expressions or timing, these frames often exhibit noticeable quality degradation, as the photo capture ISP pipeline delivers significantly higher image quality than the video pipeline. This quality gap highlights the need for dedicated restoration techniques to enhance the reselected key photo. To this end, we propose LiveMoments, a reference-guided image restoration framework tailored for the reselected key photo in Live Photos. Our method employs a two-branch neural network: a reference branch that extracts structural and textural information from the original high-quality key photo, and a main branch that restores the reselected frame using the guidance provided by the reference branch. Furthermore, we introduce a unified Motion Alignment module that incorporates motion guidance for spatial alignment at both the latent and image levels. Experiments on real and synthetic Live Photos demonstrate that LiveMoments significantly improves perceptual quality and fidelity over existing solutions, especially in scenes with fast motion or complex structures.

Core Task Landscape

This paper addresses: **Reselected Key Photo Restoration in Live Photos**
A total of **6 papers** were analyzed and organized into a taxonomy with **7 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Reference-Guided Image Restoration**
- **Video Quality Enhancement**
- **Microscopy Imaging Enhancement**
- **Real-Time Visual Processing**

Complete Taxonomy Tree

- Reselected Key Photo Restoration in Live Photos Survey Taxonomy
- Reference-Guided Image Restoration
  - Multi-Frame Photo Restoration with Motion Alignment ★ (1 papers)
  - [0] LiveMoments: Reselected Key Photo Restoration in Live Photos via Reference-guided Diffusion (Anon et al., 2026) View paper
  - Real-Time Facial Reenactment and Manipulation (1 papers)
  - [2] Face2face: Real-time face capture and reenactment of rgb videos (Thies, 2016) View paper
- Video Quality Enhancement
  - Super-Resolution for Live Streaming (1 papers)
  - [1] A super-resolution flexible video coding solution for improving live streaming quality (Qing Li, 2022) View paper
  - Archival Photo and Video Restoration (1 papers)
  - [6] VIDEO ENHANCEMENT AND RESTORATION FOR ARCHIVAL PHOTOGRAPHS (M HARITHA, n.d.) View paper
- Microscopy Imaging Enhancement
  - Trajectory-Aware Space-Time Super-Resolution (1 papers)
  - [4] Efficient Trajectory Space-Time Super-Resolution for Fast Live-cell Imaging (Ruian He, 2025) View paper
  - Implicit Neural Representation for Artifact Removal (1 papers)
  - [5] CellINR: Implicitly Overcoming Photo-induced Artifacts in 4D Live Fluorescence Microscopy (Zhao Cun-min, 2025) View paper
- Real-Time Visual Processing
  - Keyframe-Based Camera Tracking (1 papers)
  - [3] Keyframe-based real-time camera tracking (Zilong Dong, 2009) View paper

Narrative

Core task: Reselected key photo restoration in live photos. This field addresses the challenge of enhancing a user-selected frame from a live photo sequence, leveraging temporal information from neighboring frames to improve quality. The taxonomy organizes work into four main branches. Reference-Guided Image Restoration focuses on methods that exploit multiple frames or reference images to restore a target photo, often requiring careful motion alignment and feature aggregation. Video Quality Enhancement encompasses techniques for improving temporal sequences, including super-resolution and artifact removal across frames. Microscopy Imaging Enhancement targets specialized scientific imaging where temporal data can reveal finer structural details. Real-Time Visual Processing emphasizes low-latency methods suitable for interactive applications. Representative works such as SuperResolution Streaming[1] and Trajectory SuperResolution[4] illustrate how temporal context can be harnessed for quality improvement, while CellINR[5] demonstrates domain-specific enhancements in microscopy.

Several active lines of work explore trade-offs between computational efficiency and restoration quality, particularly when aligning frames with complex motion or handling degraded archival content. Within Reference-Guided Image Restoration, a small cluster of

methods tackles multi-frame photo restoration with motion alignment, where the central challenge is to register and fuse information from temporally adjacent frames without introducing artifacts. LiveMoments[0] sits squarely in this cluster, emphasizing the restoration of a reselected keyframe by aligning and aggregating features from the live photo burst. Compared to approaches like Archival Enhancement[6], which may prioritize static degradation repair, LiveMoments[0] leverages the temporal redundancy inherent in live photo sequences. This positions it closer to video-inspired techniques such as Trajectory SuperResolution[4], yet with a focus on single-frame output rather than continuous playback, reflecting the unique user interaction model of live photos.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

Both subtopics address restoration or manipulation of visual content using reference frames, but target fundamentally different domains and temporal relationships. The original leaf focuses on enhancing still photo quality by leveraging temporal information from associated video frames (live photos), while the sibling focuses on real-time facial animation transfer between video streams. The original leaf emphasizes photo restoration quality through multi-frame aggregation, whereas the sibling prioritizes real-time performance for facial reenactment.

Similarities: - Both leverage reference frames to guide visual content generation or enhancement - Both involve motion-aware processing to handle temporal relationships between frames - Both use spatial alignment techniques to establish correspondence between source and reference content - Both aim for photo-realistic output quality in their respective domains

Differences: - Application domain: photo quality restoration from live photos vs. real-time facial expression transfer - Temporal direction: retrospective enhancement of a selected frame vs. real-time forward synthesis - Content scope: general photo content (landscapes, objects, people) vs. facial regions only - Performance constraints: quality-focused offline processing vs. real-time rendering requirements - Reference relationship: using multiple frames from same capture session vs. continuous video-to-video mapping - Primary goal: restoration/denoising of degraded frames vs. animation/manipulation of facial expressions

Suggested Search Directions: - Hybrid approaches combining photo restoration with facial-specific enhancement in live photos - Real-time multi-frame photo restoration techniques that could bridge both domains - Reference-guided diffusion models for facial restoration in photo-video pairs

Sibling Subtopics

- **Real-Time Facial Reenactment and Manipulation** (leaves: 1, papers: 1)
- Scope: Methods for real-time facial expression transfer and photo-realistic re-rendering using monocular video as reference.
- Exclude: Excludes non-facial or offline methods; see Video Quality Enhancement or Real-Time Visual Processing.

Contributions Analysis

Overall novelty summary. The paper introduces LiveMoments, a reference-guided restoration framework for reselected key photos in Live Photos, addressing quality degradation when users choose alternative frames from the video clip. According to the taxonomy, this work occupies the 'Multi-Frame Photo Restoration with Motion Alignment' leaf under Reference-Guided Image Restoration, where it appears as the sole paper. This positioning suggests the paper targets a relatively sparse and specialized research direction within the broader image restoration landscape, focusing specifically on the photo-video quality gap in live photo capture systems.

The taxonomy reveals that neighboring research directions include Real-Time Facial Reenactment (focusing on expression transfer) and broader Video Quality Enhancement branches (super-resolution for streaming, archival restoration). LiveMoments diverges from these by exploiting the unique structure of live photos: a high-quality reference frame paired with lower-quality video frames. Unlike general video enhancement methods that lack reference guidance, or facial reenactment techniques targeting expression manipulation, this work specifically addresses the ISP pipeline quality disparity between photo and video capture modes, carving out a distinct problem space at the intersection of multi-frame fusion and reference-based restoration.

Among the 30 candidates examined through semantic search, none clearly refute the three main contributions. The reselected key photo restoration task (10 candidates examined, 0 refutable) appears novel within this limited scope, as does the reference-guided diffusion framework (10 candidates, 0 refutable) and the LiveMoments benchmark dataset (10 candidates, 0 refutable). The absence of refutable prior work across all contributions suggests either genuine novelty or limitations in the search scope. The specialized nature of the live photo restoration problem may explain why existing multi-frame restoration or video enhancement methods do not directly overlap with these specific contributions.

Based on the limited literature search covering 30 semantically similar papers, the work appears to address an underexplored problem space with no direct prior solutions identified. However, the small search scope and the paper's isolation within its taxonomy leaf warrant caution: a broader survey of reference-guided restoration, burst photography, or computational photography venues might reveal closer related work. The analysis captures top semantic matches but may not reflect the full landscape of multi-frame image enhancement research.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Reselected Key Photo Restoration task for Live Photos

Description: The authors define a new problem of restoring a blurry frame that users select as their preferred key photo in Live Photos by leveraging adjacent sharp frames from the same capture sequence as reference guidance.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Cascaded deep video deblurring using temporal sharpness prior

URL: View paper

Brief Assessment

Cascaded Deblurring[8] focuses on video deblurring using temporal sharpness priors from adjacent frames, not on restoring user-selected key photos from Live Photo sequences with ISP pipeline quality gaps.

---

2. Cascaded deep video deblurring using temporal sharpness prior and non-local spatial-temporal similarity

URL: View paper

Brief Assessment

Cascaded Spatial Temporal[9] focuses on video deblurring by exploiting sharp pixels from adjacent frames in video sequences, not on the specific task of restoring user-reselected key photos in Live Photos with ISP pipeline quality gaps.

---

3. Bringing events into video deblurring with non-consecutively blurry frames

URL: View paper

Brief Assessment

Events Deblurring[16] focuses on video deblurring using event cameras to handle non-consecutively blurry frames in videos, not on restoring user-selected key photos from Live Photo sequences using adjacent sharp frames as references.

---

4. Real-time large-motion Deblurring for Gimbal-based imaging systems

URL: View paper

Brief Assessment

Gimbal Deblurring[10] focuses on real-time motion deblurring for gimbal-based surveillance systems using gyroscopic priors and contrastive learning, not on restoring user-selected frames from Live Photo sequences using temporally adjacent frames as references.

---

5. Adversarial spatio-temporal learning for video deblurring

URL: View paper

Brief Assessment

Adversarial Deblurring[14] focuses on video deblurring using spatio-temporal learning with 3D convolutions to remove motion blur from video frames. It does not address the specific task of restoring reselected key photos in Live Photos using adjacent sharp frames as reference guidance from the same capture sequence.

---

6. Reference-based motion blur removal: Learning to utilize sharpness in the reference image

URL: View paper

Brief Assessment

Reference Deblur[15] focuses on motion blur removal using reference images from video sequences or alternative scenes, not the specific task of restoring reselected key photos in Live Photos with ISP pipeline quality gaps.

---

7. Frequency-aware event-based video deblurring for real-world motion blur

URL: View paper

Brief Assessment

Event Deblurring[11] focuses on video deblurring using event cameras to restore sharp frames from motion-blurred video sequences, not on restoring reselected key photos from Live Photo sequences using adjacent sharp frames as references.

---

8. Spatio-temporal filter adaptive network for video deblurring

URL: View paper

Brief Assessment

Filter Adaptive[13] focuses on video deblurring by aligning consecutive blurry frames and removing spatially variant blur, not on restoring a user-selected key photo from Live Photos using adjacent sharp frames as reference guidance.

---

9. Video extrapolation using neighboring frames

URL: View paper

Brief Assessment

Video Extrapolation[12] focuses on extrapolating video frames using neighboring frames for temporal prediction tasks, not on restoring blurry frames selected from Live Photo sequences using adjacent sharp frames as reference guidance.

---

10. Deep video deblurring for hand-held cameras

URL: View paper

Brief Assessment

Deep Video Deblurring[7] addresses general video deblurring from camera shake using temporal information from neighboring frames, but does not define or address the specific task of restoring reselected key photos in Live Photos sequences where users choose alternative frames with quality degradation.

Contribution 2: Reference-guided diffusion framework for key photo restoration

Description: The authors develop a diffusion-based restoration method that incorporates temporal information from neighboring frames in the Live Photo sequence to enhance the quality of the user-selected blurry key frame.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. LViCAR: Diffusion Models for Perceptual Quality Enhancement in Video Compression Artifact Reduction

URL: View paper

Brief Assessment

LViCAR[25] focuses on video compression artifact reduction using diffusion models with spatiotemporal architectures for temporal consistency. The original paper addresses Live Photo key frame restoration using neighboring frames as reference guidance, which is a fundamentally different task and application domain.

---

2. SVFR: A Unified Framework for Generalized Video Face Restoration

URL: View paper

Brief Assessment

SVFR[26] addresses generalized video face restoration tasks (blind restoration, inpainting, colorization) using Stable Video Diffusion priors, not reference-guided restoration of reselected key frames from Live Photo sequences with temporal context from neighboring frames.

---

3. Ditvr: Zero-shot diffusion transformer for video restoration

URL: View paper

Brief Assessment

DitVR[17] focuses on zero-shot video restoration using optical flow trajectories across video frames, not on reference-guided restoration of single frames from Live Photo sequences with temporal context from neighboring frames.

---

4. Long-term talkingface generation via motion-prior conditional diffusion model

URL: View paper

Brief Assessment

Motion Prior[24] focuses on talkingface video generation using motion priors from archived and present video clips for temporal consistency in facial animation. The original paper addresses still image restoration in Live Photos using temporal information from neighboring frames in a photo sequence. These are fundamentally different tasks with different architectures and objectives.

---

5. SeedVR: Seeding Infinity in Diffusion Transformer Towards Generic Video Restoration

URL: View paper

Brief Assessment

SeedVR[22] focuses on generic video restoration with arbitrary length/resolution using shifted window attention, not reference-guided single-frame restoration from Live Photo sequences with temporal context.

---

6. Video diffusion posterior sampling for seeing beyond dynamic scattering layers

URL: View paper

Brief Assessment

Scattering Layers[23] focuses on video diffusion models for imaging through dynamic scattering layers (e.g., fog, turbulence), not on restoring blurry frames in Live Photo sequences using neighboring frames as references.

---

7. Dual-Conditioned Temporal Diffusion Modeling for Driving Scene Generation

URL: View paper

Brief Assessment

Dual Conditioned[20] focuses on generating long driving videos with temporal consistency for autonomous driving scenarios, not on restoring degraded frames using reference images from the same sequence. The technical approaches and application domains are fundamentally different.

---

8. Learning temporally consistent video depth from video diffusion priors

URL: View paper

Brief Assessment

Consistent Depth[18] focuses on video depth estimation from arbitrary-length videos using temporal consistency across frames, not on restoring blurry key frames in Live Photos using neighboring frames as references. The tasks and applications are fundamentally different.

---

9. One-step diffusion for detail-rich and temporally consistent video super-resolution

URL: View paper

Brief Assessment

OneStep Diffusion[19] focuses on video super-resolution using temporal information across video frames for consistency, not on restoring a single reselected key photo from Live Photo sequences using a reference-guided approach with dual-branch architecture.

---

10. TDM: Temporally-Consistent Diffusion Model for All-in-One Real-World Video Restoration

URL: View paper

Brief Assessment

TDM[21] focuses on all-in-one video restoration across multiple degradation types using task prompt guidance, not on reference-guided restoration of specific frames using temporal information from neighboring frames in a Live Photo sequence.

Contribution 3: LiveMoments benchmark dataset

Description: The authors create a dedicated benchmark dataset called LiveMoments to facilitate evaluation and research on the task of restoring reselected key photos in Live Photo sequences.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. SIDAR: Synthetic Image Dataset for Alignment & Restoration

URL: View paper

Brief Assessment

SIDAR[28] focuses on synthetic data generation for image alignment and restoration tasks using 3D rendering techniques, not on benchmark datasets for photo restoration in Live Photo sequences. The candidate addresses different technical challenges (homography estimation, dense matching) compared to the original paper's focus on reselected key photo restoration in temporal video sequences.

---

2. A real-time interactive restoration system for intraoral digital videos using segment anything model

URL: View paper

Brief Assessment

Intraoral Restoration[31] focuses on intraoral video restoration using the vident-lab dataset for dental applications, not photo restoration in Live Photo sequences. The domains and tasks are fundamentally different.

---

3. VRT: A Video Restoration Transformer

URL: View paper

Brief Assessment

VRT[27] focuses on general video restoration tasks (super-resolution, deblurring, denoising) using standard benchmark datasets like REDS, Vimeo-90K, DVD, and GoPro. It does not address photo restoration in Live Photo sequences or create datasets for reselected key photo restoration.

---

4. Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency

URL: View paper

Brief Assessment

PPR10K[36] focuses on portrait photo retouching with human-region masks and group-level consistency, not on restoring reselected key photos in Live Photo sequences. The datasets address fundamentally different tasks and application domains.

---

5. Introduction to the issue on deep learning for image/video restoration and compression

URL: View paper

Brief Assessment

Deep Learning Issue[35] mentions 'several benchmark datasets for image restoration' only in passing without providing specific details about datasets for photo restoration in Live Photo sequences or temporal image sequences with reference-guided restoration.

---

6. Quanta Video Restoration

URL: View paper

Brief Assessment

Quanta Restoration[32] introduces I2-2000fps, a high-speed video dataset for quanta image restoration, not photo restoration in Live Photo sequences. The datasets serve fundamentally different purposes and imaging modalities.

---

7. Video restoration based on deep learning: a comprehensive survey

URL: View paper

Brief Assessment

Deep Learning Survey[33] is a comprehensive review of video restoration methods and datasets, but does not introduce any new benchmark datasets. It surveys existing datasets like GOPRO, DVD, REDS, and Vimeo90K for various video restoration tasks, but LiveMoments (a dataset specifically for reselected key photo restoration in Live Photos) is not mentioned and represents a distinct contribution addressing a different problem domain.

---

8. Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset

URL: View paper

Brief Assessment

Underwater Recovery[30] focuses on underwater depth recovery and image restoration with the DRUVA dataset containing underwater videos of artifacts. This is fundamentally different from LiveMoments' focus on photo restoration in Live Photo sequences captured by smartphones.

---

9. : A Large-Scale Multi-Modality Multi-View Synthesized Benchmark Dataset for Image Restoration and Content Creation

URL: View paper

Brief Assessment

MultiModality Benchmark[34] focuses on super-resolution, novel view synthesis, and controllable video generation for gaming content, not on restoring reselected key photos in Live Photo sequences. The datasets serve fundamentally different restoration tasks.

---

10. Towards real-world video face restoration: A new benchmark

URL: View paper

Brief Assessment

RealWorld Benchmark[29] focuses on video face restoration with diverse face poses and degradations from video frames, while LiveMoments addresses reselected key photo restoration in Live Photo sequences with temporal context between reference and target frames from the same capture session.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] LiveMoments: Reselected Key Photo Restoration in Live Photos via Reference-guided Diffusion View paper
- [1] A super-resolution flexible video coding solution for improving live streaming quality View paper
- [2] Face2face: Real-time face capture and reenactment of rgb videos View paper
- [3] Keyframe-based real-time camera tracking View paper
- [4] Efficient Trajectory Space-Time Super-Resolution for Fast Live-cell Imaging View paper
- [5] CellINR: Implicitly Overcoming Photo-induced Artifacts in 4D Live Fluorescence Microscopy View paper
- [6] VIDEO ENHANCEMENT AND RESTORATION FOR ARCHIVAL PHOTOGRAPHS View paper
- [7] Deep video deblurring for hand-held cameras View paper
- [8] Cascaded deep video deblurring using temporal sharpness prior View paper
- [9] Cascaded deep video deblurring using temporal sharpness prior and non-local spatial-temporal similarity View paper
- [10] Real-time large-motion Deblurring for Gimbal-based imaging systems View paper
- [11] Frequency-aware event-based video deblurring for real-world motion blur View paper
- [12] Video extrapolation using neighboring frames View paper
- [13] Spatio-temporal filter adaptive network for video deblurring View paper
- [14] Adversarial spatio-temporal learning for video deblurring View paper
- [15] Reference-based motion blur removal: Learning to utilize sharpness in the reference image View paper
- [16] Bringing events into video deblurring with non-consecutively blurry frames View paper
- [17] Ditvr: Zero-shot diffusion transformer for video restoration View paper
- [18] Learning temporally consistent video depth from video diffusion priors View paper
- [19] One-step diffusion for detail-rich and temporally consistent video super-resolution View paper
- [20] Dual-Conditioned Temporal Diffusion Modeling for Driving Scene Generation View paper
- [21] TDM: Temporally-Consistent Diffusion Model for All-in-One Real-World Video Restoration View paper
- [22] SeedVR: Seeding Infinity in Diffusion Transformer Towards Generic Video Restoration View paper
- [23] Video diffusion posterior sampling for seeing beyond dynamic scattering layers View paper
- [24] Long-term talkingface generation via motion-prior conditional diffusion model View paper
- [25] LViCAR: Diffusion Models for Perceptual Quality Enhancement in Video Compression Artifact Reduction View paper
- [26] SVFR: A Unified Framework for Generalized Video Face Restoration View paper
- [27] VRT: A Video Restoration Transformer View paper
- [28] SIDAR: Synthetic Image Dataset for Alignment & Restoration View paper
- [29] Towards real-world video face restoration: A new benchmark View paper
- [30] Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset View paper
- [31] A real-time interactive restoration system for intraoral digital videos using segment anything model View paper

- [32] Quanta Video Restoration View paper
- [33] Video restoration based on deep learning: a comprehensive survey View paper
- [34] : A Large-Scale Multi-Modality Multi-View Synthesized Benchmark Dataset for Image Restoration and Content Creation View paper
- [35] Introduction to the issue on deep learning for image/video restoration and compression View paper
- [36] Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency View paper