Computer Vision (CV25) — Midterm Project Proposal
Instructor: Dr. I. Atadjanov, Dr. B. Kiani
Team Name: VisionAR
Project Title: Real-Time Artifact Removal for Passthrough AR (Temporal + Spatial Denoising)

Team Members:

- Azizbek — Project Lead
- Azim — Model & Evaluation Lead
- Ikromjon — Data & Integration Lead

GitHub Repository: https://github.com/azizbekb/CV25_Proposal_VisionAR

---

## 1. Abstract

Augmented Reality (AR) devices rely on real-time video passthrough to merge the physical world with digital content. However, passthrough video streams often suffer from spatial and temporal artifacts such as motion blur, sensor noise, and compression distortion, which degrade both visual quality and tracking accuracy.

Our project, VisionAR, proposes a real-time artifact removal system that enhances AR video streams using a combination of spatial denoising via a lightweight Convolutional Neural Network (CNN) and temporal smoothing to ensure visual consistency across frames. The system is designed to maintain low-latency processing (<30ms per frame), enabling seamless integration in real-world AR applications.

We will evaluate performance using publicly available AR datasets and custom webcam recordings. Key metrics include PSNR, SSIM, and latency (ms). The final deliverables will include an optimized denoising model, a quantitative evaluation report, and a demonstration video comparing raw and cleaned AR video streams.

---

## 2. Problem & Motivation

Modern AR and MR (Mixed Reality) systems depend on camera passthrough feeds for rendering the real-world environment. However, these video feeds often contain visible artifacts such as noise, compression

blocks, flickering, and inconsistent lighting. These imperfections can lead to eye strain, unstable virtual object placement, and reduced immersion for the user.

Traditional denoising algorithms such as Gaussian or Bilateral filtering fail to preserve edges while maintaining real-time performance. Therefore, there is a growing demand for fast, lightweight AI-based denoisers that can improve perceptual quality without introducing lag.

The motivation behind VisionAR is to design a low-compute, low-latency video denoising system optimized for AR devices. Instead of maximizing image quality alone, our focus is achieving the optimal trade-off between clarity and speed. This solution directly benefits the AR/VR industry, robotics, and real-time telepresence systems where visual consistency and responsiveness are crucial.

---

## 3. Related Work

Several deep learning–based denoising approaches have demonstrated strong results in computer vision tasks.

- DnCNN (Zhang et al., 2017) introduced a CNN architecture for single-image denoising that generalizes across multiple noise levels.
- FastDVDnet (Tassano et al., 2020) extended this concept to videos by grouping consecutive frames to learn temporal relationships without explicit optical flow estimation, achieving near real-time results.
- EDVR (Wang et al., 2019) further improved restoration using deformable convolutions but requires high computational resources, making it unsuitable for AR hardware.

While these models achieve excellent denoising quality, few are optimized for on-device inference where latency and memory are constrained.

Our approach builds upon FastDVDnet for its computational efficiency and adds temporal post-filtering and OpenCV-based smoothing to further stabilize frame-to-frame consistency. We also take inspiration from Li et al. (2023) and Xu et al. (2024), who emphasized low-latency rendering pipelines for AR and MR systems.

---

4. Data & Resources

Datasets:

- OpenAR Video Dataset — Synthetic AR passthrough dataset suitable for denoising evaluation.
- Custom Webcam Dataset — Short recordings under varying lighting and motion conditions to simulate real-world AR use cases.

Resources:

- Hardware: NVIDIA RTX 2050 GPU (local) or Google Colab Pro GPU.
- Software: Python 3.10, PyTorch, TorchVision, OpenCV, NumPy.
- Storage: 10–20 GB total video data.

Ethics & Privacy:
All data will be self-recorded or publicly available. No personal faces, private environments, or sensitive materials will be captured. No IRB approval is required as human subjects are not involved.

---

5. Method

5.1 Overview

The VisionAR pipeline integrates both spatial and temporal denoising modules to enhance AR passthrough feeds in real time. It consists of the following stages:

1. Frame Acquisition — Capture live video from a webcam or AR passthrough input.
2. Preprocessing — Normalize frames, detect artifacts, and stabilize camera motion.
3. Denoising Network — Apply a lightweight CNN based on *FastDVDnet* for temporal and spatial noise removal.
4. Rendering Output — Stream cleaned frames back to the AR pipeline at 30–60 FPS.

### 5.2 Model Architecture

The CNN receives three consecutive frames ((I_{t-1}, I_t, I_{t+1}))\ as input. Shared convolutional layers extract features, followed by temporal fusion layers that learn motion consistency.

Loss functions used include:

- L1 Loss — Minimizes pixel-wise error.
- Edge Consistency Loss — Preserves fine spatial structures.
- Temporal Smoothness Loss — Ensures consistent brightness and color between frames.

The final output is re-injected into the AR rendering loop with optimized buffering to maintain <30ms latency.

---

## 6. Experiments & Metrics

We will benchmark VisionAR using both synthetic and real datasets. The following metrics will be used:

| Metric | Description |
|---|---|
| PSNR (Peak Signal-to-Noise Ratio) | Measures quantitative image quality after denoising. |
| SSIM (Structural Similarity Index) | Evaluates perceptual similarity to ground truth. |
| Latency (ms) | Measures per-frame processing time. |
| MOS (Mean Opinion Score) | Subjective user evaluation of visual stability. |

Success Criteria:

- Maintain ≥30 FPS (≤30ms per frame).
- Achieve +2dB PSNR and +0.03 SSIM improvement over baseline OpenCV filters.
- Ensure visually stable, flicker-free AR feed in demo videos.

7. Risks & Mitigations

| Risk | Mitigation Strategy |
| --- | --- |
| Limited GPU access | Use Colab Pro or CPU-optimized inference models. |
| Model too slow for real-time | Prune model layers or lower input resolution. |
| Dataset too small | Augment data with synthetic noise and generated frames. |
| Unstable performance under lighting changes | Use adaptive normalization and temporal averaging. |

---

8. Timeline & Roles

See full roadmap: ROADMAP.md

Summary:

- Week 1–2: Setup, topic approval, dataset collection.
- Week 3: Baseline OpenCV filters + metrics.
- Week 4: CNN model integration.
- Week 5: Optimization and evaluation.
- Week 6: Final report and video demo.

Team Roles:

- Azizbek: Project coordination, optimization, final documentation.
- Azim: Model development and evaluation.
- Ikromjon: Data preparation, integration, and preprocessing.

---

9. Expected Outcomes

- Working prototype of a real-time AR denoising pipeline.
- Comparative report: baseline vs. AI-based performance.
- Demonstration video showing before/after improvement.

- Optional: lightweight deployment version (ONNX/TensorRT).

The project's broader impact lies in improving real-time visual quality for future AR/VR applications, reducing motion artifacts, and increasing user comfort.

---

10. Ethics & Compliance

This project adheres to all academic integrity and data privacy standards.
All code and datasets used will be properly cited.
No personal data, biometric, or medical information will be processed.
All materials will be released under permissible open-source or research-use licenses.

---

11. References


1. Zhang, K., et al. "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising." *IEEE Transactions on Image Processing*, 2017.
2. Tassano, M., et al. "FastDVDnet: Towards Real-Time Deep Video Denoising Without Flow Estimation." *CVPR*, 2020.
3. Wang, X., et al. "EDVR: Video Restoration with Enhanced Deformable Convolutions." *CVPR Workshops*, 2019.
4. Li, C., et al. "Latency-Aware AR Rendering for Head-Mounted Displays." *IEEE VR*, 2023.
5. Xu, Y., et al. "Adaptive Temporal Smoothing for Real-Time Mixed Reality." *ACM SIGGRAPH*, 2024.