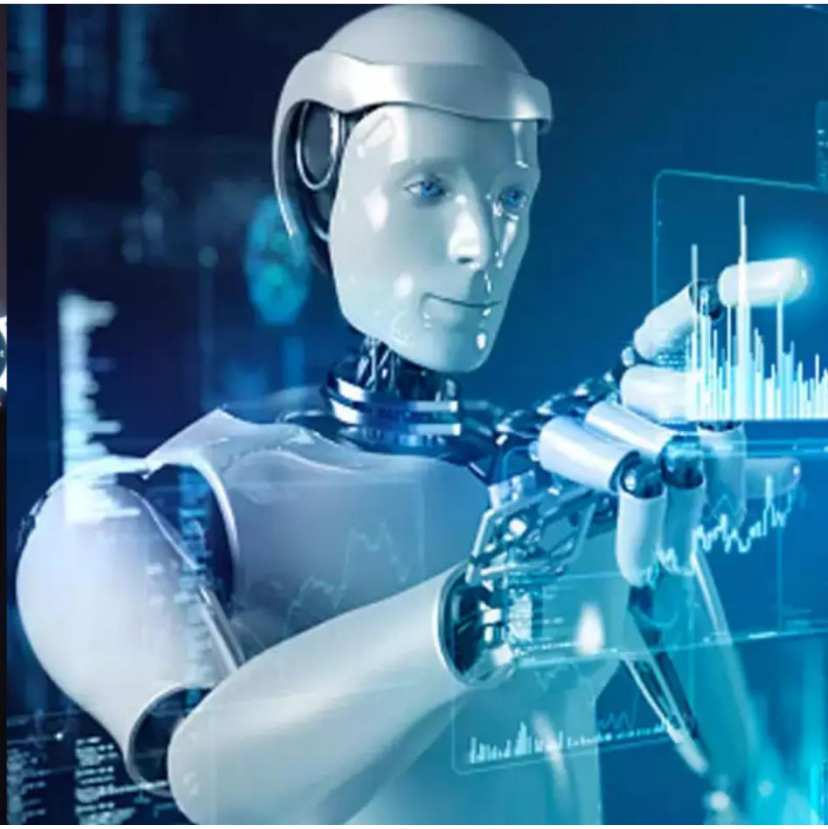
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

Vision-X: An Intelligent Object Removal in Text, Space and Human Interaction

Dr.P.Chiranjeevi[1], K. Sai Nikith[2], G. Chandrashekar[2], C. Valli Srujana[2], N. Hari Krishna[2]

Associate Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India[1]

IV B. Tech, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India[2]

ABSTRACT: VISION-X is an intelligent visual enhancement system designed to improve the clarity and quality of documents, images, and live video streams. It addresses common issues such as shadows, noise, uneven lighting, and background distractions using an adaptive computer vision approach instead of heavy deep learning models. By analyzing image characteristics, the system automatically selects the most suitable enhancement pipeline to preserve content while improving readability. With modules for Document Enhancement, Virtual Home Staging, and Live Video Background Refinement, VISION-X offers a scalable and practical solution for real-world applications.

I. INTRODUCTION

VISION-X addresses the need for improving the quality and professionalism of images and videos used in digital communication and documentation. Visual content captured through mobile devices often contains unwanted elements such as shadows, clutter, furniture, or sensitive items, which reduce clarity and privacy. Existing tools either blur entire backgrounds or require manual editing skills, making them unsuitable for many users. VISION-X provides an automated solution that detects and removes unwanted objects using deep learning and image inpainting techniques while preserving the original background. It supports document cleaning, virtual home staging, and real-time object removal during video calls through an easy-to-use web interface. The system focuses on object detection, removal, and quality enhancement but does not include full interior redesign or 3D modelling features.

II. EXISTING SYSTEM

Current object removal and visual enhancement solutions are fragmented and domain-specific. Document enhancement tools mainly focus on basic adjustments such as brightness and contrast but fail to effectively remove complex distractions like shadows or fingers. Virtual background features in video calls only blur or replace the entire background, offering limited privacy and no selective object removal capability. Furniture removal or interior editing applications are often computationally expensive and lack real-time performance, restricting their usability in dynamic environments such as real estate workflows. Overall, existing systems lack adaptability, selective object removal, and efficient real-time processing.

III. PROPOSED SYSTEM

VISION-X proposes a unified and intelligent visual enhancement system that integrates three distinct yet interconnected modules, leveraging Artificial Intelligence, Computer Vision, and Augmented Reality technologies. Unlike existing fragmented solutions, the proposed system provides selective object detection and removal while preserving the overall background and visual integrity.

The three core modules include:

- **Document Cleaner** – Automatically detects and removes shadows, fingers, and unwanted clutter from scanned or photographed documents, improving readability and clarity.
- **Virtual Home Staging (Furniture Remover)** – Identifies large furniture or distracting interior objects and removes them to visualize clean and spacious environments, especially useful in real estate applications.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

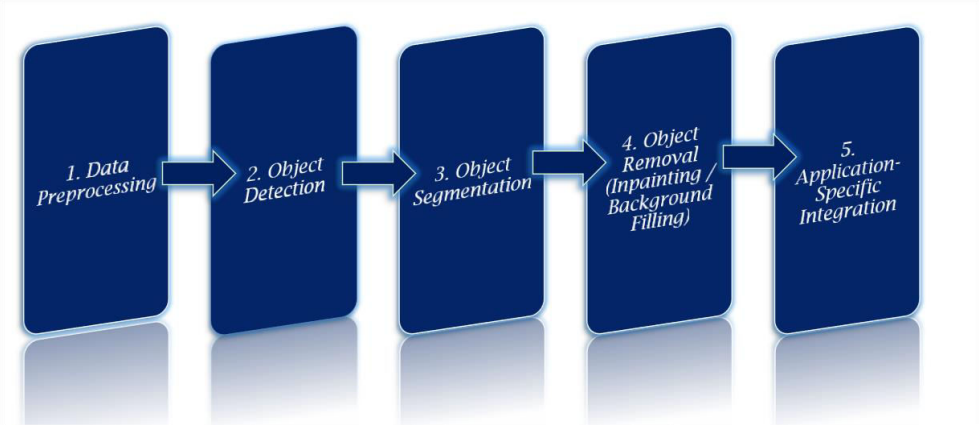
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

- **AR Object Remover for Video Calls** – Enables real-time detection and masking of distracting or sensitive objects during live video meetings without blurring the entire background.

IV. METHODOLOGY



i. Data Preprocessing

In this stage, input images and video frames are resized and converted into suitable formats to maintain uniformity across the system. Pixel values are normalized to ensure consistency, and noise reduction techniques are applied to remove unwanted distortions. Brightness and contrast adjustments are performed to correct uneven lighting conditions. Low-quality or corrupted samples are filtered out, ensuring that only clean and optimized data proceeds to the next stage for accurate processing.

ii. Object Detection

The system analyzes the preprocessed input to identify unwanted objects such as shadows, fingers, furniture, or background clutter. Deep learning–based detection models are used to locate these elements and generate bounding boxes around them. Each detected object is classified and evaluated based on confidence levels to ensure reliable identification. The selected regions are then forwarded for detailed segmentation.

iii. Object Segmentation

After detection, segmentation techniques are applied to generate precise pixel-level masks of the identified objects. This step separates the unwanted object from the background with high accuracy. Mask boundaries are refined to capture the exact shape and structure of the object while preserving surrounding details. The resulting mask clearly defines the region that will be removed or modified in the next stage.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

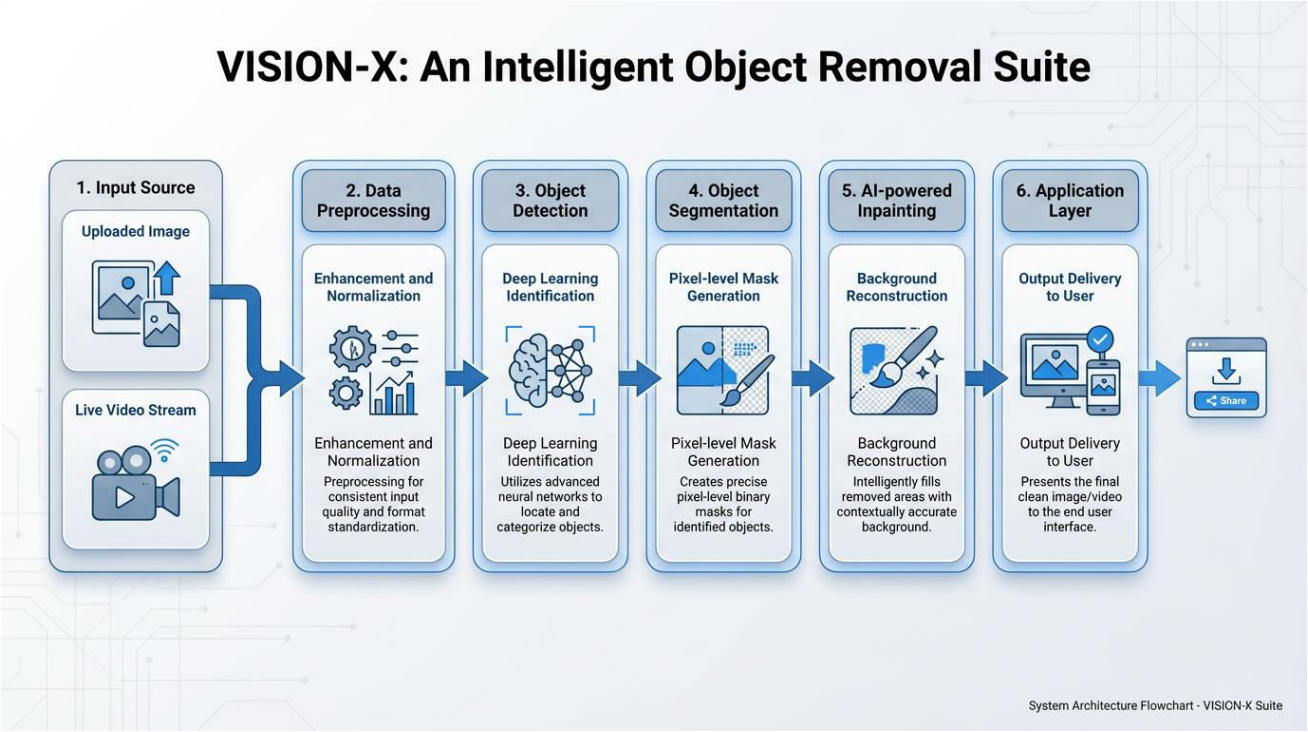
iv. Object Removal (Inpainting / Background Filling)
Once the object mask is created, the system removes the selected region and reconstructs the missing background using context-aware inpainting techniques. The surrounding texture, color patterns, and structural information are analyzed to fill the removed area naturally. Special care is taken to preserve edges, straight lines, and spatial consistency to ensure a realistic output.

v. Application-Specific Integration
In the final stage, the processed output is integrated into the respective application module, such as document cleaning, virtual home staging, or live video refinement. The system connects backend processing with the frontend interface, allowing users to upload media, preview results, and export enhanced outputs. Real-time optimization is applied where required.

V. SYSTEM ARCHITECTURE

The system architecture of VISIONX is based on a modular pipeline that sequentially processes the input. The system receives either an uploaded image or a live video stream. Once input is obtained, the data preprocessing unit performs enhancement and normalization. The refined input flows into the object detection component, which identifies unwanted objects using deep- learning models. The segmentation unit then extracts the precise object boundaries using pixel-level mask generation. Finally, the inpainting module reconstructs the removed region using advanced AI-based background fill techniques.

The output is rendered back to the user through the application layer. Since each module operates independently, the architecture ensures low coupling and high cohesion.



VI. RESULTS AND OUTPUT

STEP: 1 Open the URL in any of the web browsers. You can see UI as shown in the image.Create an account using your mail id and password. Or if you are an existing user, log in to the application using the credentials that you used while creating your account. You can also claim the premium access and use all the premium features (like no watermark and all).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

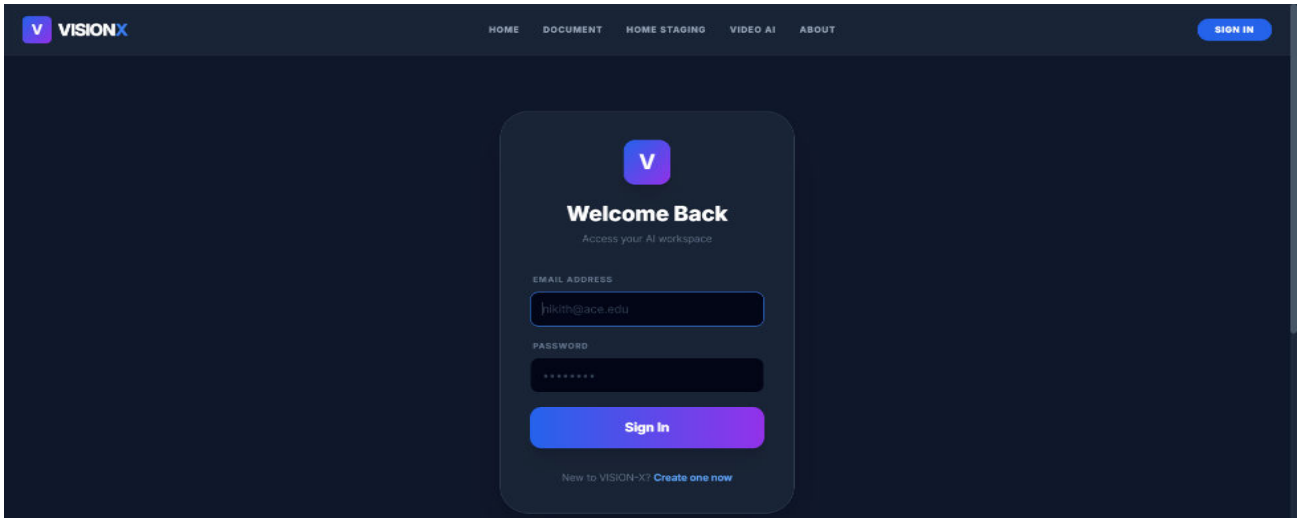


Fig : 6.1 Login page

STEP: 2 In the navbar, you can see various pages like document, home staging, and video AI. In that, select any of the required pages (say document).Upload an image of a document from your device. And click on the enhance button to remove unwanted objects, shadows and clutter from your documentation.

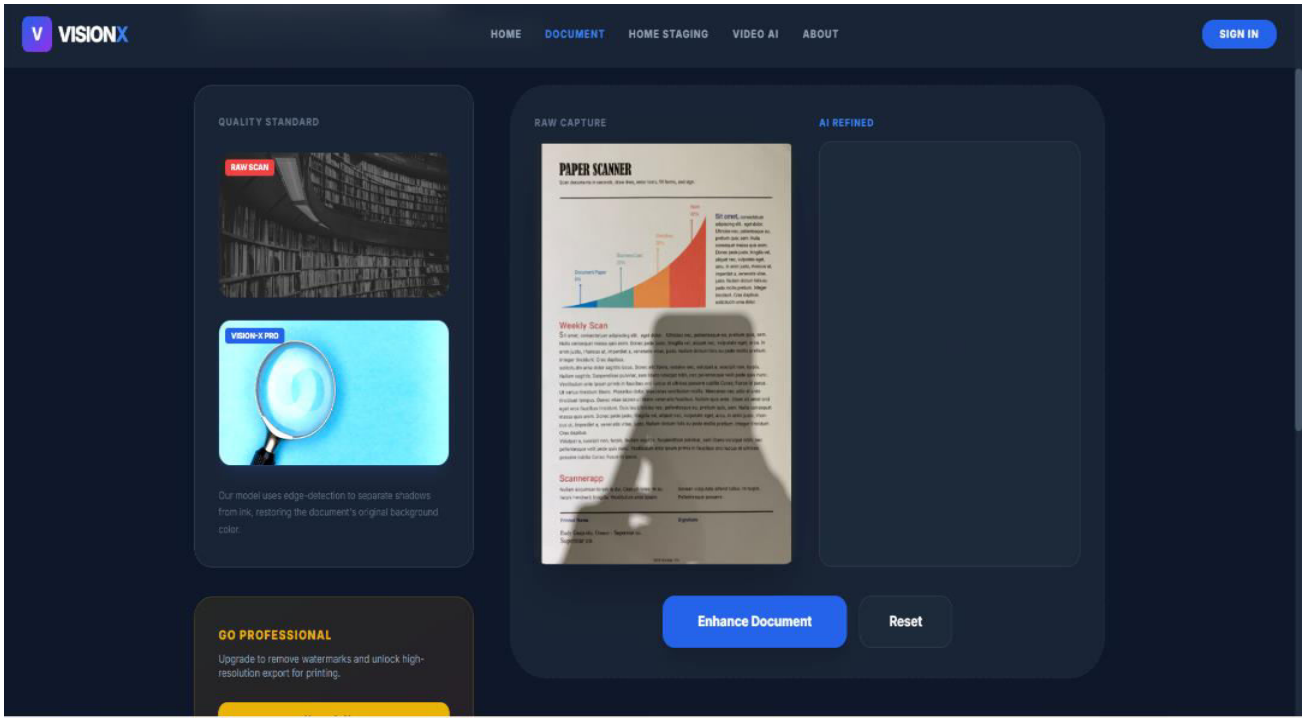


Fig : 6.2 input document

STEP: 3 The model processes the uploaded input and generates an output by removing the shadows or clutters, you can download this enhanced image by clicking on the DOWNLOAD RESULT button.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

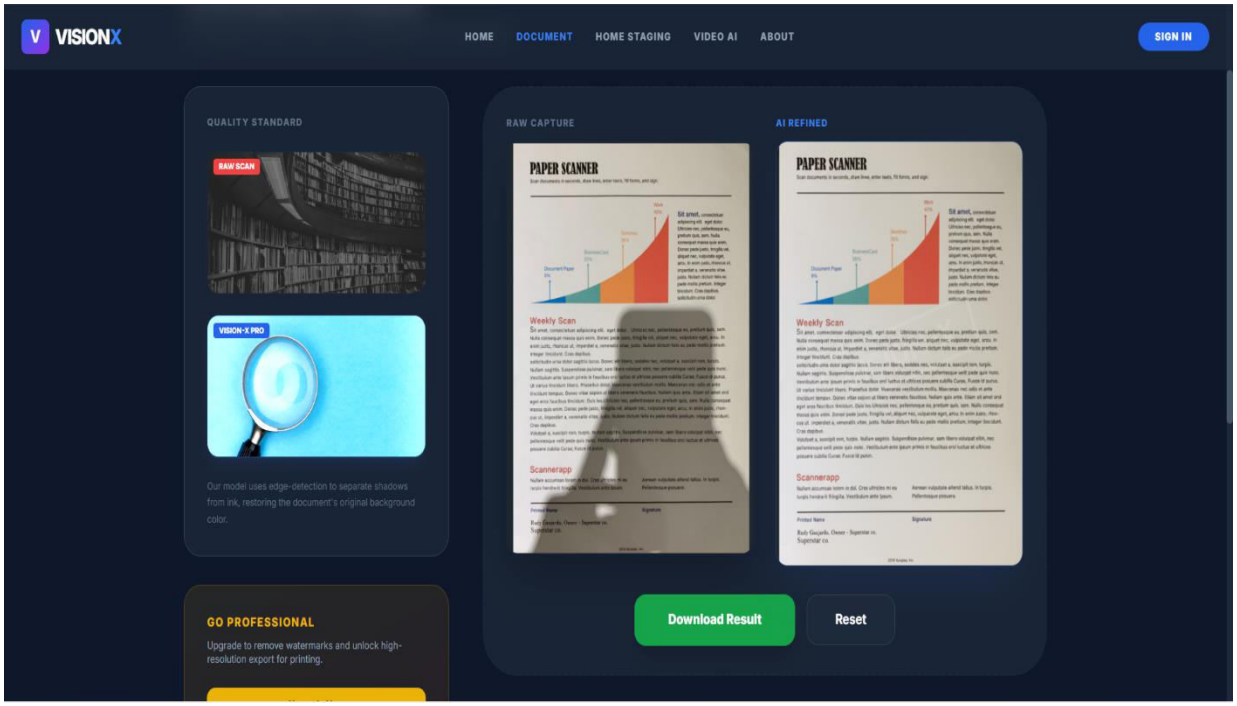
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

VII. CONCLUSION

- VISION-X successfully demonstrates the design and implementation of an AI-powered visual enhancement platform that improves the quality, clarity, and privacy of images and videos in real time. The system integrates modern web technologies with classical computer vision techniques to provide three major capabilities: document enhancement, virtual home staging, and live video refinement. By using lightweight, efficient algorithms such as contrast enhancement, shadow suppression, inpainting, and frame processing, the application delivers high-quality results without requiring deep learning models or GPU resources.
- The platform is designed to be user-friendly, allowing non-technical users to upload and process visual data through an intuitive interface easily. Its stateless request–response ar- chitecture ensures reliable performance, efficient resource utilization, and easy deployment across different environments. Additionally, the system emphasizes data privacy, as it avoids long-term storage and processes each request independently.
- The modular architecture ensures scalability, maintainability, and fast response times, making the system suitable for practical real-world applications such as education, real estate, document digitization, and video conferencing. Furthermore, the project provides a strong foundation for future enhancements, including AI-based automation, cloud integration, and support for additional visual processing modules.
- Overall, VISION-X proves that intelligent visual enhancement can be achieved through optimized engineering design, efficient processing pipelines, and thoughtful system architecture, making it a reliable and cost-effective solution for modern visual computing needs.



VII. FUTURE SCOPE

i. Deep Learning Integration

Integrating deep learning models can significantly improve the system's intelligence and automation. Convolutional Neural Networks (CNNs) can be used for document segmentation, enabling accurate separation of text, background, and noise. AI-based models can automatically detect furniture and objects in images, improving virtual home staging quality.

Additionally, semantic background removal can precisely distinguish foreground objects from complex

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503052|

backgrounds, resulting in more realistic visual outputs.

ii. OCR Integration

Optical Character Recognition (OCR) can be incorporated to extract text from enhanced documents such as scanned certificates, invoices, and forms. This allows users to convert images into searchable and editable digital documents, improving usability for offices, educational institutions, and digital archiving systems. OCR integration also enables automated data extraction and indexing.

iii. Smart Object Detection

Advanced object detection models can eliminate the need for manual masking by automatically identifying clutter, furniture, or unwanted elements in images. This enhancement can speed up the processing workflow and reduce user effort. Additionally, automatic face and sensitive object blurring can be implemented to enhance privacy, making the system suitable for surveillance, social media, and compliance-driven applications.

iv. Mobile Application

Developing an Android and iOS application will make VISION-X accessible on mobile devices. Users can perform on-the-go document scanning, image enhancement, and virtual staging using smartphone cameras. Mobile integration increases convenience, adoption, and real-world usability, especially for students, professionals, and real estate agents.

REFERENCES

[1] J. Redmon and A. Farhadi, "YOLOv3: An incremental improve- ment," *arXiv preprint arXiv:1804.02767*, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[2] G. Jocher et al., "YOLOv8: Real-time Object Detection and Segmentation," Ultralytics, 2024. [Online]. Available: https://docs.ultralytics.com/

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2961–2969. doi: 10.1109/ICCV.2017.322

[4] G. Liu et al., "Image Inpainting for Irregular Holes using Partial Con- volutions," *arXiv preprint arXiv:1804.07723*, 2018. [Online]. Available: https://arxiv.org/abs/1804.07723

[5] R. Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022. doi: 10.48550/arXiv.2112.10752

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real- time object detection with region proposal networks," *IEEE Trans. Pat- tern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. doi: 10.1109/TPAMI.2016.2577031

- OpenCV Official Documentation – https://opencv.org/
- FastAPI Official Documentation – https://fastapi.tiangolo.com/
- React Official Documentation – https://react.dev/

Tailwind CSS Documentation – https://tailwindcss.com/

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details