RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2667

Comparative Evaluation of Benchmarking SuperPoint against Classical Feature Extractors for Fiducial Markers with Explainable AI Insights

P.K. Anjana [a,*], V. Ashok Immanuel [b]

[a] Department of Computer Science, Christ Deemed to be University, Bangalore, 560029, India
* Corresponding Author Email: anjana.pk@res.christuniversity.in
DOI: https://doi.org/10.54392/irjmt2667
Received: 17-06-2026; Revised: 23-10-2026; Accepted: 10-11-2026; Published: 24-11-2026

Abstract: This study describes the feature extraction methods for fiducial marker detection in collaborative augmented reality systems and a comparative experimental analysis across multiple fiducial marker families under controlled degradation and evaluation conditions. Feature extraction plays a crucial role in visual localization, SLAM (Simultaneous Localization and Mapping), and autonomous navigation by improving the extraction and tracking of distinctive visual features for both mapping and localization tasks. Classic feature extraction methods applied and failed to detect a greater number of features from the fiducial marker images. Here, the dataset comprises images of fiducial markers, including AprilTag, ChArUco, STag, and TopoTag. SuperPoint: Self-Supervised Interest Point Detection and Description presented for detecting and describing key points in the images. Our study analyzes SuperPoint, which is able to constantly detect a major set of interest points than any other traditional corner detector methods like ORB, SURF, and SIFT. The extracted features are evaluated using repeatability, confidence, matched feature count, matching score, and matching accuracy, followed by degradation experiments to assess their robustness under varying conditions. Experimental results show that among the evaluated markers, SuperPoint with STag (8 cm) performed the highest repeatability score, that is, 99.92 %, and matching accuracy of 99.68% while SuperPoint with ChArUco (4 cm) achieved the highest matching score, 83.77%, with the highest confidence score of 0.031652. AprilTag (1 cm) and TopoTag (4 cm) also demonstrated robust matching performance with high numbers of matched features, which are 1954 and 1879, and reliable matching accuracy of 99.33% and 99.52% respectively. For real-time performance, SuperPoint with STag (1 cm) achieved the highest frame rate (147.16 FPS) and throughput (73.5353 MB/s), with low frame latency (0.0067 s/frame). Additionally, this work integrated with Explainable Artificial Intelligence (XAI) techniques such as occlusion sensitivity tests and heatmap visualization to clarify the characteristics of the SuperPoint model as regards the performance evaluation.

Keywords: Augmented Reality (AR), Collaborative, Explainable Artificial Intelligence (XAI), Feature Extraction, Fiducial Marker, Heatmap Visualization, Occlusion Sensitivity, Superpoint.

1. Introduction

Fiducial markers are essential for accurate alignment and placement in augmented reality applications and provide a cost-effective solution for labeling, tracking, and localization issues. They are specially designed visual patterns and reliable reference points for detecting, tracking, and estimating the position and orientation of objects in a real-world environment. AprilTag, ChArUco, STag, and TopoTag are the established planar fiducial systems, and they employ significantly different visual structures with detection principles. AprilTags are black and white mosaic patterns optimized for robust detection and precise pose estimation. They are lightweight and computationally efficient and can run on embedded devices at real-time frame rates [1]. ChArUco is a hybrid fiducial marker system that combines the strengths of ArUco markers (binary square fiducials) with a chessboard grid [2]. Now ChArUco2 has developed an enhanced calibration board system and C++ library that builds on the original ChArUco design to overcome some key limitations of the standard ChArUco boards, mainly double marker density with wider spatial spread, which is more informative for calibration. STags are a hybrid design of an outer square for coarse detection and homography estimation, paired with an inner circle for precise refinement. Recent study about STag reveals that it provides a robust foundation for robot localization, especially in environments where stability and resilience to occlusion are critical [3]. TopoTags encode

information using topological structures, and unlike square markers like AprilTag, it uses all vertices of its structure for pose estimation, improving accuracy [4]. Hence, these differences amid marker designs connect fiducial markers as a suitable test bench for discovering how feature-extraction methods respond to different planar visual structures.

Feature extraction is selecting the best essential corner features for model training, and eliminating redundant features makes the process very defined. Especially corners of an image are basically considered the area in which there is heterogeneity in a large amount of gradient in all possible dimensions and directions. Moreover, [5] specified corners as the points where such changes occur, unlike edges, which have changes in intensity in only one direction, and flat regions have very little change in any direction. It is noted that generally handcrafted feature approaches accurately demonstrated a good performance for identified corners in the multiview camera but were slow due to noncornered images. On the other hand, [6] explained deep learning approaches automatically learn complex and more discriminative features from the data, adapt to variations in scale, rotation, and noise, and achieve more accuracy, robustness, and generalization performance than handcrafted classical feature extraction methods.

The SuperPoint approach was introduced [7] by Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich in 2018 to conquer all these restrictions. SuperPoint is an inclusive convolutional neural network that simultaneously discovers the identification and descriptor extraction of interest points inside a well-integrated, end-to-end architecture. SuperPoint can now mutually learn keypoint detection and descriptors in a single forward pass, which is more outstanding than traditional techniques like SIFT, SURF, and ORB on homography estimation challenges and reaching real-time speeds. After SuperPoint, collectively other studies followed that basic methodology and enhanced algorithms, such as SuperGlue [8], which introduced a graph neural network with attention-based context aggregation to establish correspondences between two sets of local features while rejecting non-matchable points, and then LightGlue [9] proposed an adaptive learned matcher that improved the computational efficiency of sparse feature matching while retaining strong matching performance. Its adaptive inference mechanism allows computational effort to vary according to the difficulty of an image pair, making the approach particularly relevant to latency-sensitive vision applications. Next, recently OmniGlue [10] was introduced as a learnable image feature matcher that focuses on generalization across diverse visual domains. SuperGlue and LightGlue [8, 9] performed well mainly in their training domains, whereas OmniGlue [10] is explicitly designed to handle out-of-distribution data like aerial, object-centric, and synthetic images.

Detecting fiducial markers with a greater number of reliable features is very prominent in augmented reality (AR) applications because these rich feature details help to improve marker recognition accuracy, tracking stability, and pose estimation performance. Thus, strong feature extraction techniques are needed to encounter numerous issues that can affect detection accuracy, including occlusions, motion blur, camera noise, and varying lighting conditions. Traditional feature extraction techniques have many limitations, and they often have limited detection accuracy [11]. DeepArUco++ proposed in [12] a deep-learning-based framework for square fiducial marker detection, corner refinement, and decoding, particularly targeting challenging illumination conditions and image degradation such as blur and sensor noise. Such approaches demonstrate that neural networks can be incorporated directly into the markerdetection pipeline to improve robustness when conventional image-processing assumptions become unreliable.

Hence, there is a possibility to conduct a comparative evaluation of SuperPoint and classical local feature extractors like SIFT, SURF, and ORB for planar fiducial markers, considering AprilTag, ChArUco, STag, and TopoTag as representative marker structures. Instead of introducing a new marker design, this study isolates the feature-extraction stage and investigates how the same feature-extraction framework responds to different marker topologies, corner configurations, internal patterns, and geometric structures. In recent works, the effects of marker structure and varying imaging conditions on keypoint detection, feature matching, computational efficiency, and robustness are not sufficiently clear. Consequently, a standardized experimental framework is required to evaluate these methods and identify their strengths and limitations for accurate fiducial-marker detection in real-time augmented reality applications.

2. Related Work

This section briefly reviews the prior research in fiducial marker systems, including AprilTag, ChArUco, STag, and TopoTag; examines classical local feature extraction methods, namely SURF, SIFT, and ORB; discusses SuperPoint and its subsequent developments; and finally reviews fiducial marker detection approaches for collaborative augmented reality applications.

The literature review in Table 1, which we have done here, mainly focuses on improving computer vision tasks through local feature extraction and matching processes of different fiducial markers and by comparing classical methods with deep learning methods, which help to find out how many improvements have been developed for the detection process of fiducial markers.

Table 1. Comparison of Research Works: Key Improvements from Base Research to Recent Research

Paper	Key Improvements in Research	Limitations
Olson, 2011 [13]	Robust 2D fiducial detection, fast line detection, 6-DOF localization, and tolerance to occlusion and distortion.	Partial border occlusion can affect detection; pose estimation mainly relies on outer marker corners.
Wang *et al.*, 2016 [14]	Redesigned detector with higher detection rate, fewer false positives, lower computational cost, and improved small-tag detection.	Improved speed and false-positive performance is obtained at the cost of reduced partial-border occlusion detection.
Krogius *et al.*, 2019 [15]	Faster processing, higher recall, improved small/distant tag detection, flexible layouts, and increased tag capacity.	Smaller borders may reduce pose accuracy and detection distance; detector speed and recall require a trade-off, especially for small tags.
Garrido-Jurado *et al.*, 2014 [16]	Reliable coded-marker detection and improved robustness to occlusion.	Stochastic dictionary generation may produce suboptimal solutions; occlusion handling depends on multiple markers and color segmentation.
Hu *et al.*, 2019 [17]	CNN-based marker detection and corner refinement for low-light and motion-blur conditions.	Additional corner-refinement network increases computation and requires camera calibration for pose estimation.
Munoz-Salinas˜ *al.*, 2026 [18]	Approximately doubled marker density, improved detection range and robustness, and increased diamond reference points from 4 to 9 pose reference points.	Further evaluation of calibration accuracy and repeatability under diverse conditions is still required.
Benligiray *et* 2019 [19]	Stable pose estimation using square and circular borders with homography refinement.	High computational cost, with EDPF contributing more than 80% of the reported detection runtime.
Heselden *et* 2025 [3]	> 50% improved low-contrast detection, up to ~ 11 trillion IDs, > 75% tested occlusion detection, and 6-DoF robot localization.	Chromatic extensions increase processing time, while long-term material robustness and atmospheric-condition performance require further investigation.
Yu *et al.*, 2020 [4]	Topology-based detection with improved detection/pose accuracy, occlusion robustness, flexible shapes, and million-scale ID scalability.	Occlusion handling depends on preserving the marker topology; severe occlusion may require multiple markers.
Shewell *et al.*, 2016 [20]	Combines ORB's low-light and motion-blur resilience with ArUco's high-accuracy marker recognition.	Missed detections reduce recall; noise and multiple objects may cause false positives.
DeTone *et al.*, 2018 [7]	Self-supervised keypoint detection and description with Homographic Adaptation for improved repeatability. 70 FPS at 480×640; HPatches Rep.=0.581 and MScore=0.470.	Generic rather than marker-specific; large keypoint sets can increase subsequent matching cost.
Sarlin *et al.*, 2020 [8]	GNN-based self/cross-attention and optimal transport for context-aware matching and outlier rejection. AUC@5°: 9.80 → 34.18; precision: 22.5% → 84.9%.	Computational cost increases with the number of keypoints, reducing suitability for high-keypoint real-time applications.
Lindenberger *et al.*, 2023 [9]	Adaptive-depth matching and early stopping for efficient, low-latency matching. Runtime: 70.0 → 44.2 ms; AUC@5°: 93.0 → 94.5.	Inference speed depends on image-pair difficulty and feature set.
Jiang *et al.*, 2024 [10]	DINOv2-guided attention improves generalization, with 20.9% and 9.5% gains over SuperGlue and LightGlue on unseen domains.	Primarily developed for general image matching rather than fiducial-marker detection.
Zhang *et al.*, 2023 [11]	Supports multiple marker families with dense keypoint detection and subpixel refinement; achieved 100% recall and precision on Yu's dataset.	Deep-learning processing introduces greater computational requirements than lightweight classical approaches.
Berral-Soler *et al.*, 2024 [12]	CNN-based detection, refinement, and decoding for challenging conditions; outperformed ArUco and DeepTag with competitive throughput.	Multi-stage neural pipeline; occasional false detections still require thresholding.

The recent advancements have improved detection speed, small-tag performance, flexibility, and pose estimation of AprilTag while referring to [13–15]; there is an increased marker density, larger markers, additional reference points, and improved pose estimation from ChArUco to ChArUco2 [17,18]; STag to enhanced STag [3,19] have increased customizability, marker capacity, occlusion resistance, and practical robot localization; and then [4] the introduction of topology-based, scalable, and customizable marker design with improved occlusion robustness has been developed for TopoTag. Traditional feature extraction techniques have progressively highlighted the balanced relationship between robustness and computational efficacy. Considering the studies, SIFT offers robust matching that is changeless to scale and rotation but requires considerable computing expense, whereas SURF was created as a more advisable option with similar resilience. Recent fiducial marker research has mostly shifted toward dedicated geometric, template-based, and deep learning detectors in place of using SIFT or SURF as the primary detection mechanism. ORB [20] improved computing performance through binary feature representation, sketching it as appropriate for real-time applications. To be precise, under large-scale and illumination variations, ORB generates less robust and unique features, which can result in lower matching accuracy in complicated situations.

Even though SIFT has dependable qualities, mainly its large computational complexity makes it slower for real-time applications. In a similar way, SURF outperforms SIFT in speed, but it still has issues with resilience and processing efficiency under extreme perspective and occlusion fluctuations. On the other hand, advanced techniques like SuperPoint [7] have delivered strong feature representations with impressive matching efficacy. This base work is the main link across all sources, connecting as the base front-end architecture for specialized contributions of works, such as [8–10]. Then, [11,12] introduced deep learning–based frameworks for fiducial marker design and detection, allowing support for both existing and newly designed marker families while improving detection robustness and pose accuracy compared to traditional handcrafted methods. However, their comparative efficiency is heavily impacted by the dataset, hardware, and evaluation parameters. Hence, a direct experimental assessment of AprilTag, ChArUco, STag, and TopoTag markers is very needed to identify the optimal approach.

All these works considered, previous studies have progressed from handcrafted local feature extraction to learned feature detection and matching, while recent fiducial-marker research has progressively addressed difficult scenarios such as blurriness, changes in lighting, and occlusion. Even so, the majority of current methodologies focus on a specific marker design or an entire marker-detection system. A methodical evaluation of a widely utilized feature extractor, especially SuperPoint, among structurally diverse recognized fiducial markers is still comparatively limited. This inspires the current research, which evaluates SuperPoint in comparison to SIFT, SURF, and ORB with AprilTag, ChArUco, STag, and TopoTag.

3. Materials and Methods

This paper presents a systematic cross-marker evaluation framework that compares classical and deep-learning-based feature extraction methods across multiple fiducial-marker families. The study further analyzes the relationship between marker structure and feature extraction performance using quantitative evaluation and explainability analysis. Initially the dataset, which includes images of AprilTags, ChArUco, STag, and TopoTag, is compiled. Then, this dataset undergoes preprocessing, which mainly includes resizing images, standardizing lighting conditions, and augmentation such as scaling and rotating the images. Followed by an effective feature extraction technique by SuperPoint, it uses a neural network trained in a self-supervised way to automatically learn where good key points are presented. Along with that, a comprehensive analysis is conducted with some important metrics such as the number of features detected, processing time for getting those features, FPS (frames per second), frame latency, computation time, throughput, repeatability score, matched features, matching score, matching accuracy, and confidence score with this set of fiducial markers, AprilTags, ChArUco, STag, and TopoTag. In addition, Explainable Artificial Intelligence (XAI) techniques are integrated for the spatial characteristics of SuperPoint feature responses through occlusion sensitivity analysis and heatmap visualization.

Here, Figure 1 shows the overall workflow of the proposed fiducial marker feature extraction and evaluation framework.

3.1 Pseudocode of the Proposed Model

Algorithm 1. Integrated Image Processing Pipeline and Deep Feature Extraction Framework

```
Algorithm
Phase 1: Matlab Pipeline Initialization and Pre-processing
Clear MATLAB workspace
Define projectpath and imagefolder
A ← Read input images from dataset
    for i ← 1 to users do
        A1{i} ← standardizeLighting(A{i})
        A2{i} ← imresize(A1{i},[256 × 256])
Import Python module spmod
result ← spmod.runpipeline(imagefolder,modelpath)
numFeatures ← result{1}
```

pyLatency ← result{3}
all keypoints ← result{4}
all heatmaps ← result{5}
Phase 2: Python Code - Deep Feature Extraction
Load pretrained SuperPointNet model
Set conf thresh ← 0.0015
Set nms radius ← 4
Set max keypoints ← 2500
for i ← 1 to 10 do
    image ← Read RGB splitting image
    image ← Resize image to 256 × 256
    tensor ← Convert image into tensor
    Start inference timer
    semi, desc ← model(tensor)
    heatmap ← decode superpoint(semi)
    pts ← Extract points above conf thresh
    scores ← Sort confidence scores
    keypoints ← Apply NMS using nms radius
    descriptors ← Extract 256D descriptors
    Store heatmaps, keypoints and descriptors

Phase 3: Matlab Post-processing and Performance Metrics
Start total super timer
for i ← 1 to users do
    img ← imread(current_image)
    kp ← double(all keypoints {i})
    Store feature count and frame statistics
Compute FPS super, lat super and thr super

3.2 Dataset and Marker Categories

The feature-extracting model [21] utilizes a Fiducial Marker Analysis dataset for fiducial marker detection. The dataset integrates multiple fiducial marker types, including 100 samples of AprilTag (1 cm), 100 samples of ChArUco (4 cm), and STag markers at three scales (1 cm, 8 cm, and 16 cm), with 100 samples each, respectively. Then TopoTag markers are collected at three different sizes (2 cm, 4 cm, and 5 cm), with each size category comprising 100 samples. The complete dataset includes variations in scale, viewpoint, illumination, rotation, blur, and partial occlusion to simulate realistic augmented reality environments. Table 2 provides the information of the dataset, and Table 3 gives the additional dataset and model-data information.

3.3 Dataset Preprocessing

Some of the existing AR systems depend on basic grayscale conversions, global thresholding, or non-unified resizing techniques. Because of that, they commonly struggle with adaptive management of varying lighting conditions and offer little in terms of augmentation. Here in preprocessing, standardization ensures that uniform data scaling across lighting variations really makes a difference in generalizability. Resizing aligns all images to a consistent input dimension, 256×256 pixels, avoiding feature map misalignment during the learning stage. Augmentation such as rotation, flipping, and scaling increases dataset diversity, improving model robustness against real-world tag distortions. Since the dataset includes fiducial marker images with different original dimensions, a uniform spatial resolution of 256×256 pixels is adopted for quantitative evaluation.

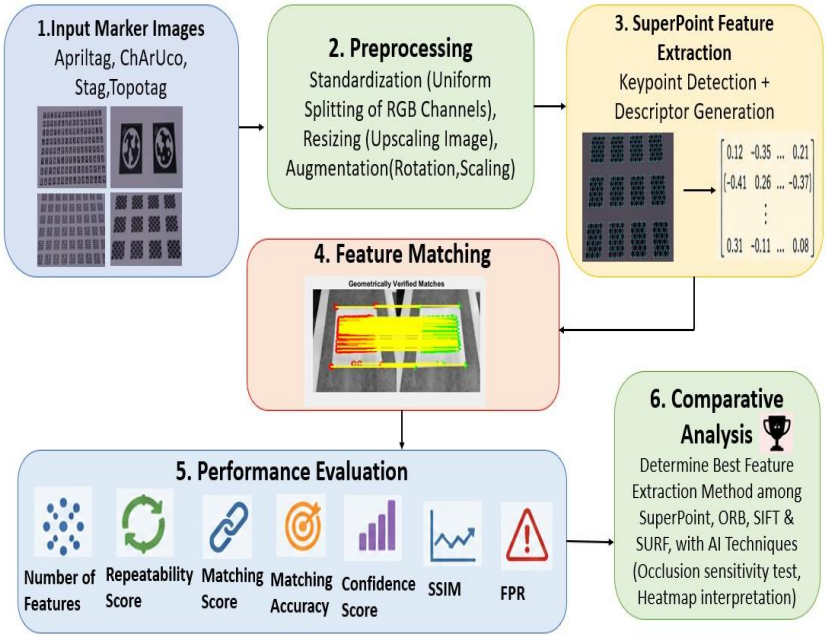


Figure 1. Overview of the proposed model methodology

Table 2. Repository Information

Repository Owner	Y.Z. Guleray¨
Repository Title	Fiducial Markers Analysis Dataset
Year	2025
Version	1.0
Access Date	17 Aug 2026
DOI	https://github.com/yusufguleray/ Fiducial-Markers-Analysis?tab= readme-ov-file

Table 3. Additional Dataset and Model-Data Information

Parameter	Information
Camera Type	Not reported in the source dataset
Capture Distance	Not reported in the source dataset
Viewing Angle	Not reported in the source dataset
Illumination Range	Not reported in the source dataset
Blur-Generation Procedure	Not specified in the source dataset
Occlusion Percentage	Not specified in the source dataset
Annotation Procedure	Not specified in the source dataset
SuperPoint Training Dataset	No additional training was performed using the study dataset
SuperPoint Fine-Tuning Dataset	No fine-tuning was performed using the study dataset
SuperPoint General Evaluation Dataset	AprilTag (1 cm), ChArUco (4 cm), STag (1 cm, 8 cm, and 16 cm), and TopoTag (2 cm, 4 cm, and 5 cm) marker images
Degradation AnalysisDataset	ChArUco (4 cm) images evaluated using SuperPoint and SURF under the defined degradation conditions

The corresponding original image is subjected to the preprocessing operation and thereafter resized to 256×256 pixels to obtain the test image. SSIM and PSNR are then calculated between the reference and test images at identical spatial dimensions for all fiducial marker types. The AprilTag (1 cm) image used in the experiments has a resolution of 1024 × 1024 pixels with a 24-bit color depth and a display resolution of 96 dpi. The ChArUco (4 cm) image, the STag images of sizes 1 cm, 8 cm, and 16 cm, and the TopoTag images of sizes 2 cm, 4 cm, and 5 cm used for evaluation have a resolution of 1920 × 1080 pixels with a 24-bit color depth and a display resolution of 96 dpi.

As shown in Equation 1, the Structural Similarity Index Measure (SSIM) measures the structural similarity between the reference image *x* and a test image *y*.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (1)$$

Where μ_x, μ_y : mean intensity values of images *x* and *y*,

σ_x^2, σ_y^2 variances of images *x* and *y*, σ_{xy} : covariance between images *x* and *y*,

C_1, C_2 : stabilization constants to avoid division by zero.

The Peak Signal-to-Noise Ratio (PSNR) is a common measure that quantifies the quality of a processed image by comparing it with a reference image.

The PSNR and MSE are calculated using Equations 2 and 3, respectively.

$$PSNR = 10log^{10}(\frac{MAX_I^2}{MSE}) \quad (2)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (3)$$

where

MAX_I : maximum possible pixel value

MSE : Mean Squared Error

3.4 Feature Extraction Methods

Conventional methods like SIFT, SURF, and ORB are based on handcrafted methodologies to find out keypoints and create descriptions. Regardless of indicating robust performance in manually controlled environments, these approaches repeatedly confront difficulties in generalizing under major differences in illumination, perspective, and scene intricacy. SuperPoint by DeTone *et al.* [7] has its self-supervised training approach, and the model is primarily trained on synthetic data with a forerunner network called MagicPoint and then afterward transformed by a process referred to as homographic adaptation. This method authorizes the network to attain consistent and repeatable keypoints from real-world images without the requirement for explicit ground-truth annotations.

The core of SuperPoint is the concept of covariance under homographic transformations, which states that identified interest spots change consistently as the image is warped. As shown in Equation 4 and Equation 5,

$$H_x = f_\theta(H(I)) \quad (4)$$

Since real detectors are imperfect, SuperPoint enforces this property empirically through homographic adaptation:

$$F^{(I;f_\theta)} = \frac{1}{N_h} \sum_{i=1}^{Nh} H_i^{-1} f_\theta\left(H_i(I)\right) \quad (5)$$

Where:

I : Input image

H_i : Randomly sampled homography

f_θ : SuperPoint detector

N_h : Number of homographic transformations

Homographic adaptation enhances keypoint detection robustness by applying multiple random homographic transformations to an image and combining the detected keypoints.

Next, Equation 6 ensures that both the detector and descriptor are optimized together, sharing features that improve geometric consistency.

$$L = L_p(X,Y) + L_p(X',Y') + \lambda L_d(D,D',S) \quad (6)$$

- L_p : Cross-entropy loss for interest point detection
- L_d : Hinge loss for descriptor similarity
- λ : Weighting factor

Then in Equation 7 , the extracted descriptors are normalized to unit length to improve matching consistency and robustness.

$$d^{\wedge} = \frac{d}{\|d\|2} \quad (7)$$

d : Descriptor vector

$d^{\wedge}$: Normalized descriptor

$\|d\|_2$: L2 norm of the descriptor vector

3.5 Experimental Parameter Settings for Feature Extraction and Matching

In this section, Table 4 provides a common matching and geometric verification protocol that is applied to all four feature extraction methods such as SuperPoint, ORB, SURF, and SIFT, on ChArUco marker of size 4 cm, while method-specific feature extraction parameters are configured independently. All these parameters are very important because they directly influence keypoint detection sensitivity, scale-space analysis, descriptor generation, keypoint selection, and computational cost. Hence, evaluating them and reporting ensures a transparent comparison among the four feature extraction methods. To maintain a consistent computational budget and provide a controlled comparison of processing performance, a maximum of 25 features are selected for the ensuing feature matching procedure for all four feature extraction methods. In Table 4, twenty-five represents the maximum number of correspondences used for matching and does not represent the maximum number of keypoints detected by the feature extraction methods. The actual number of detected keypoints is image dependent and is recorded separately as the feature count. Therefore, the detected feature count could exceed 25, whereas a maximum of 25 selected features is used for the subsequent matching procedure. Here also marking non-applicable settings as ”N/A” for each method.

Next, Table 5 shows the selected features and their corresponding descriptors are subsequently matched between the image pairs. A maximum ratio of 0.80 is used to filter unclear matches, and unique matching is enabled to retain one-to-one feature correspondences. The resulting matches are then subjected to geometric verification using RANSAC with a similarity transformation model. RANSAC is used to identify geometrically consistent inliers and reject outlier correspondences. The verified correspondences are subsequently used for the calculation of the matching evaluation metrics.

Table 6 displays the degradation analysis for SuperPoint data configuration.

Therefore, for each degradation condition is Equation 8,

Degradation Levels × Observations per Level = 10 × 5 = 50 (8)

Thus, each of the four performance measures contains 50 observations distributed across 10 degradation levels.

Table 4. Feature Extraction Parameter Configuration for ORB, SIFT, SURF, and SuperPoint on ChArUco marker (4 cm)

Parameter	ORB	SIFT	SURF	SuperPoint
Maximum Correspondences Used for Matching	25	25	25	25
Scale Factor	1.3000	N/A	N/A	N/A
Pyramid/Octaves	3	3	3	N/A
Contrast Threshold	N/A	0.0133	N/A	N/A
Edge Threshold	1.2	10	N/A	N/A
Sigma	N/A	1.60	N/A	N/A
Hessian Threshold	N/A	N/A	1000	N/A
Scale Levels	N/A	N/A	4	N/A
Descriptor Type	ORB Binary	SIFT-128D	SURF-64D	256-D
FAST Threshold	0.08	N/A	N/A	N/A
Score Type	HARRISSCORE	N/A	N/A	N/A
Confidence Threshold	N/A	N/A	N/A	0.0015
NMS Radius	N/A	N/A	N/A	4 pixels
Pretrained Weights	N/A	N/A	N/A	superpointv1.pth

Table 5. Common Matching and Geometric Verification Parameters

Parameter	Configuration
Maximum Ratio	0.80
Unique Matching	True
Geometric Model	Similarity Transform
Geometric Verification	RANSAC

Table 6. SuperPoint Data Configuration for Degradation Analysis

Parameter	Value
Repeatability observations	50
Matching accuracy observations	50
Processing time observations	50
Feature count observations	50
Degradation levels	10
Observations per degradation level	5

3.6 Evaluation Metrics

The performance of the evaluated feature extraction and matching methods is analyzed using a set of quantitative metrics such as feature count, repeatability, confidence score, matching score, matching accuracy, processing time, computational performance, and False Positive Rate (FPR). The selected metrics are used to provide a comprehensive comparison of SuperPoint, ORB, SIFT, and SURF across the different fiducial-marker configurations.

3.6.1 Feature Count

Feature count represents the total number of keypoints detected by a feature extraction method in an input image. It provides a direct measure of the amount of local visual information identified by each method. For an image *I*,as shown in Equation 9,

$$Nf = |\mathrm{K}| \quad (9)$$

where N_f denotes the number of detected features and K represents the set of detected keypoints.

3.6.2 Repeatability

It is the ability of a feature extraction method to detect corresponding or spatially consistent keypoints across related images of the same marker. Therefore, it indicates the stability of feature detection when the imaging condition changes.

For Repeatability, Equation 10 measures the ability of the detector to consistently identify the same keypoints under image transformations.

$$Repeatability = \frac{N_{repeated}}{N_{reference}} \quad (10)$$

$N_{repeated}$: Number of repeated keypoints

$N_{reference}$: Number of reference keypoints

3.6.3 Mean Confidence Score

The Mean confidence score calculated by Equation 11, represents the average confidence assigned by SuperPoint to the detected keypoints and reflects the overall reliability of feature detection.

$$C_{mean} = \frac{1}{N}\sum_{i=1}^{N} C_i \quad (11)$$

C_{mean} : Mean confidence score

c_i : Confidence score of the i^{th} keypoint

N : Total number of detected keypoints

3.6.4 Matched Features

The matched features are defined as the total number of accepted feature correspondences obtained between the two images by Equation (12),

$$M = N_{matched} \quad (12)$$

where M represents the number of matched feature pairs.

3.6.5 Matching Score

The matching score is calculated as the percentage of matched features relative to the total number of reference keypoints using Equation (13),

$$MS(\%) = \frac{M}{K} * 100 \quad (13)$$

where M is the number of matched feature pairs and K is the total number of reference keypoints.

3.6.6 Matching Accuracy

The matching accuracy was calculated as the ratio of correct matches to the total number of matched features using Equation (14),

$$MA = \frac{C}{M} \quad (14)$$

where C represents the number of correct or geometrically verified matches and M represents the total number of accepted matched features.

3.6.7 Processing Time and Frame Rate

Processing time is the amount of time required by the algorithm to process an image or frame. Here, the processing time stands for the total elapsed time for processing 20 images in each fiducial marker category.

Frame rate, expressed as frames per second (FPS), indicates how many images/frames the system process in one second. The frame rate is calculated using Equation (15),

$$FPS = \frac{N}{T_p} \quad (15)$$

where N is the total number of processed images and T_p is the total processing time in seconds.

3.6.8 Frame Latency

Frame processing latency represents the average time required to process one input image frame and generate the corresponding feature detection and matching result. It is calculated from the total processing time and the number of processed frames using Equation 16:

$$L = \frac{T_p}{N} = \frac{1}{FPS} \quad (16)$$

where L represents the average processing latency per frame in seconds, T_p denotes the total processing time for N processed frames, and FPS represents the corresponding frame rate in frames per second. Lower latency indicates faster processing and greater suitability for realtime fiducial-marker detection.

3.6.9 Throughput

Throughput represents the amount of input data processed by the feature extraction and matching system per unit of time. It is expressed in megabytes per second (MB/s) and is calculated as the ratio of the total input data size to the total elapsed processing time using Equation (17),

$$Thoughput = \frac{D}{T_p} \quad (17)$$

where D is the total input data size in megabytes (MB) and T_p is the total processing time in seconds. Here, total input data size in megabytes is 10MB. Computation time is reported per image, whereas processing time represents the total elapsed time for the 20-image evaluation.

3.6.10 False Positive Rate

False Positive Rate (FPR) is a metric that is used in tasks like object detection to quantify the number of actual negative regions, such as irrelevant backgrounds or objects that are incorrectly classified as positive. It is calculated using Equation 18.

$$FPR = \frac{FP}{FP+TN} \quad (18)$$

where

FP : number of False Positives,

TN : number of True Negatives.

3.7 Experimental Protocol

3.7.1 Evaluation Metrics Performance for Feature Extraction and Matching

A controlled comparative experimental protocol is established to evaluate the performance of SuperPoint and the classical feature extraction methods SIFT, SURF, and ORB for fiducial marker image analysis. The evaluation included eight marker configurations: AprilTag (1 cm), ChArUco (4 cm), STag (1 cm), STag (8 cm), STag (16 cm), TopoTag (2 cm), TopoTag (4 cm), and TopoTag (5 cm). The same experimental image set and evaluation procedure are used for the comparative assessment to ensure consistency across the evaluated feature extraction methods. First, we conducted a comparative study that analyzes some significant approaches on the basis of metrics for 20 images: the processing time for getting features; FPS (frames per second), which shows how many frames the system can process in one second; frame latency, which indicates how long it takes for one frame to travel through the entire system; computation time; and lastly, throughput, which means how much data the system can process over time, and GPU/CPU speed. Then for the baseline marker evaluation, the extracted keypoints and descriptors are used for feature matching, and the resulting correspondences are assessed using repeatability, confidence score, matched-feature count, matching score, and matching accuracy.

3.7.2 Matching Performance of SuperPoint, ORB, SIFT, and SURF

Next, for the pair-based matching evaluation, 50 image pairs are considered, as specified $N = 50$ in the comparative evaluation. The same matching protocol is applied to SuperPoint, ORB, SIFT, and SURF, allowing their matching performance to be compared under equivalent experimental conditions. The resulting measurements are summarized using the mean, standard deviation (SD), and 95% confidence interval (CI), where applicable. The CI is used to quantify the uncertainty associated with the estimated mean matching accuracy.

Then, to determine whether the observed differences among the four evaluated methods (ORB, SIFT, SURF, and SuperPoint) are statistically significant, a Friedman repeated-measures test is conducted for repeatability, matching accuracy, feature count, and processing time. It is calculated by Equation 19

$$X_F^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3N(k+1), \quad (19)$$

where N is the number of observations, k is the number of methods, and R_j is the sum of ranks for the j-th method. The resulting χ^2, degrees of freedom, and p-value are reported together with Kendall's coefficient of concordance W as a measure of effect size using Equation 20:

$$W = \frac{X_F^2}{N(k-1)} \quad (20)$$

A significance level of $p < 0.05$ is used.

Following the overall Friedman test, pairwise Wilcoxon signed-rank tests are conducted to identify which method pairs differed significantly. To account for multiple pairwise comparisons, the resulting p-values are adjusted using the Holm correction. For each pair, the Wilcoxon signed-rank statistic W, raw p-value, Holm-adjusted p-value, and significance decision are reported.

For the Wilcoxon signed-rank test, the paired differences are ranked according to their absolute magnitude, and the test statistic is based on the smaller of the positive- and negative-rank sums using Equation 21:

$$W = \min(W^+, W^-) \quad (21)$$

where W^+ and W^- denote the sums of positive and negative ranks, respectively.s

For m pairwise comparisons, the Holm-adjusted significance level for the ordered p-values is using Equation 22:

$$\alpha_i = \frac{\alpha}{m-i+1} \quad (22)$$

where i denotes the rank of the ordered p-value and $\alpha = 0.05$.

3.7.3 Robustness Performance under Controlled Image Degradations

Degradation experiments are performed for ORB, SIFT, SURF, and SuperPoint on the marker category, ChArUco marker of size 4 cm, based on increasing disturbance severity to evaluate their robustness in terms of processing time, matching accuracy, repeatability, and feature count. Since SuperPoint demonstrated the most favorable overall performance in the preceding comparative evaluation, its degradation results are presented in detail to further assess its robustness under progressively degraded conditions. Moreover, the degradation results of SURF are evaluated and presented here. The corresponding degradation results for ORB and SIFT, are also evaluated but are omitted from the main manuscript for brevity. Each degradation process is evaluated at controlled severity levels from 0% to 90% in increments of 10% , and the resulting changes in repeatability, matching accuracy, feature count, and processing time help to analyze and determine the sensitivity and robustness of the evaluated method. For the degradation experiments, the processing time, feature count, repeatability, and matching accuracy are evaluated separately for each degradation condition.

Feature counts are recorded as the number of detected keypoints/features produced by the corresponding feature-extraction method, while processing time is recorded for the processing of the degraded images. Repeatability measures how many of the same or matching feature points detected in the original image are detected again in the degraded image, and matching accuracy helps to find out how many of the matched feature pairs are actually correct or geometrically valid out of all accepted matches. The same severity levels and degradation categories are used for the evaluated methods to provide a consistent basis for comparison.

Then, a detailed evaluation is conducted under five degradation conditions: illumination variation, blur, rotation, viewpoint change, and partial occlusion on ChArUco markers of size 4 cm to assess the robustness of the deep learning feature extraction methods SuperPoint and SURF based on various image disturbances. Each condition is evaluated independently at ten controlled severity levels ranging from 0% to 90%. The mean of the measured performance metrics is calculated for each severity level.

3.7.4 Keypoint Detection and Matching Validation

Next, a complete processing procedure is systematically defined from keypoint detection to geometric verification to ensure the reproducibility of the SuperPoint feature detection and matching experiment. Evaluation of the image pairs of ChArUco (4 cm) is completed by detecting keypoints in both images, identifying repeated keypoints within the specified pixel tolerance, and calculating repeatability. Then detected descriptors are matched using a fixed ratio threshold, followed by mutual consistency filtering and RANSAC-based geometric verification. Ultimately, the verified matches are classified as true or false matches, and the corresponding true-match and false-match rates were calculated.

3.8 Explainability of SuperPoint

3.8.1 Occlusion Sensitivity Test

SuperPoint is a black box, which refers to a system whose internal workings and algorithms are always hidden from the user and interacts with only inputs and observed outputs. So, the user arrives at some conclusions without knowing how the system reached them. SuperPoint gives feature points or keypoints and descriptors.

Occlusion sensitivity test is a good explainable AI (XAI) technique applied to identify which parts of the image are significant for a model's conclusion. An important drop confidence against occlusion specifies that the region is vital for the model's prophesy and also it provides perception into the spacial dependencies informed by the network. Explainable AI (XAI) examines that the Convolutional Neural Networks (CNNs) have learned to impose high confidence to such corner-like patterns during training, as proved by strong activations in the confidence heatmap. Descriptor matching occurs as the network receives the ability to associate visually and geometrically analogous patches with the closest points in the feature space while differentiating dissimilar ones via contrastive learning. During training, positive pairs are aggregated while negative pairs are separated, leading to the formation of discriminative and invariant representations. Explainable AI analysis explains that equivalent descriptors focus on similar prominent regions, signifying common structural information. Additionally, occlusion sensitivity testing discloses that concealing essential areas diminishes descriptor similarity, so we can infer that the correspondence is influenced by significant and informative local patterns.

3.8.2 Heatmap Interpretation

A visual representation that contains colors those are encode numerical values. The probability of each pixel represents a keypoint or interest point. Generally, SuperPoint detects particular structured locations, not all the points. The bright spots are white/yellow in color, indicating high-confidence keypoints. And these are corners which shows sharp intensity changes and strong feature location on the specific image. Dark red spots have a low confidence score, and orange has a medium confidence score. The region that is black indicates very low confidence, near to zero. In this study we have analyzed SuperPoint heatmap activation of all the markers: AprilTag, ChArUco, STag, and TopoTag. The created heatmap exposes a sparse activation pattern, with only a few pixels confirming high confidence values. This sparsity indicates that the SuperPoint model superiorly locates remarkable and consistent keypoints while diminishing responses in non-informative areas. Moreover, getting a sparse heatmap is good because it reduces noise, improves matching accuracy being used by strong points, which results in fewer false matches, helps to reduce computational efficiency due to fewer points being selected, and has better geometrical stability by focusing on true structures like corners. SuperPoint outshines other methods using fiducial markers like AprilTags, ChArUco, STag, and TopoTag.

SuperPoint does not detects markers, it detects repeatable, high-information keypoints based on learned visual patterns. The number of detected points depends on local gradient structure. Each image's local gradient structure commands the detection process, like areas with momentous intensity variations in several directions, such as corners or intersections, deferring high-confidence replies. But in the case of edges, which show gradient variation in only one direction, and flat regions with insignificant gradients present minimally to the detection process. Moreover, these are the locations

where the image has strong structural information, and each point can easily and reliably be detected again if any rotation, change in illumination, or change in viewpoint happens.

3.9 Instruments and Software

Table 7 and Table 8 provide the information about the hardware configuration used for SuperPoint evaluation and the software configuration used for experiments, respectively.

4. Results and Discussion

The experimental results confirm a clear contrasting view among ORB, SIFT, SURF, and SuperPoint in terms of feature robustness, matching performance, and computational efficiency for the fiducial marker extraction process. The evaluation considers complementary metrics, including repeatability, matching score, matching accuracy, confidence score, number of matched features, processing time, frame rate, frame latency, computation time, and throughput. The experimental results exhibit that no single feature-extraction method provides the best performance across all evaluation criteria and marker types. The relative performance of SuperPoint, ORB, SIFT, and SURF alters according to the structural characteristics of the fiducial marker, the number and distribution of detectable visual features, and the computational conditions of the feature-extraction and matching stages. Accordingly, feature count alone cannot be reasoned as an indicator of detection quality. The combined analysis of repeatability, matching score, matching accuracy, confidence, and processing time provides a more meaningful assessment of feature-extraction performance. SuperPoint provides strong feature robustness and matching performance; at the same time, the classical methods achieve competitive computational efficiency with particular marker configurations.

4.1 Performance Evaluation of the Preprocessing Pipeline

The experimental results of Table 9 confirm the potency of the analysis. The structural similarity index (SSIM) in the range of near 1 reveals that the reconstructed images keep most of the structural and intuitive information from the reference images. Moreover, a peak signal-tonoise ratio (PSNR) between 20 and 40 dB confirms a good balance between noise reduction and reconstruction quality. Then, for all the markers, the peak signal-to-noise ratios (PSNRs) are between the ChArUco of size 4 cm, which shows a minimum value of 24.2064 dB, and the TopoTag of size 2 cm, which shows a maximum value of 27.0313 dB, clearly showing stability. The structural similarity index (SSIM) for all the markers signifies good results of a range near 1. Again, the STag of size 16 cm shows the highest value, that is, 0.9361, and the AprilTag of size 1 cm shows the lowest, 0.9236.

4.2 Feature Point Detection Comparison of SuperPoint, ORB, SIFT, and SURF

In this comparative analysis, each graph shows a comparison of four different feature extraction techniques (SuperPoint, ORB, SIFT, and SURF) evaluated over an experimental threshold ranging from 0 to 3000 units, and for 10 users we calculated all the unique features. The experimental outcomes announce a clearly noticeable tier of difference between traditional handcrafted feature detectors and contemporary deep learning techniques. SuperPoint delivers a greater number of feature points than the traditional opponents, ORB, SIFT, and SURF.

Table 7. Hardware Configuration Used for SuperPoint Evaluation

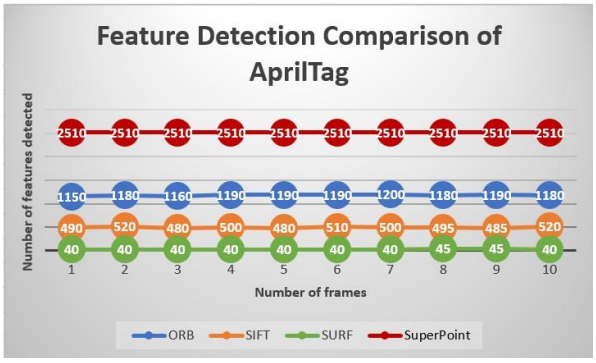
Component	Specification
Processor	Intel Core i3-1115G4 (11th Gen, 3.00 GHz)
Graphics Processing Unit (GPU)	Intel UHD Graphics
Shared GPU Memory	3.8 GB
RAM	8 GB

Table 8. Software Configuration Used for Experiments

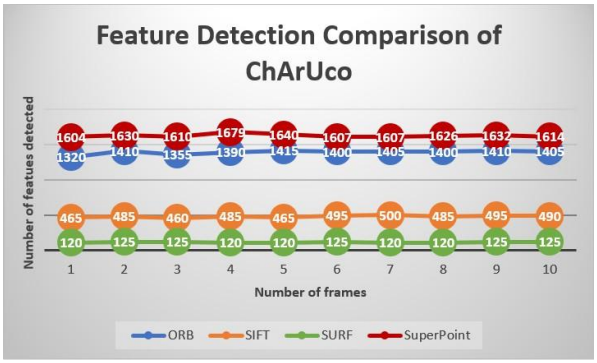
Component	Specification
Operating System	Windows 11, Version 25H2
Programming Language	Python 3.10.0
Computer Vision Library	OpenCV
Feature Extraction Model	SuperPoint
Development Environment	Matlab IDE build in with python integrated compiler

Table 9. Evaluation of Preprocessing Performance

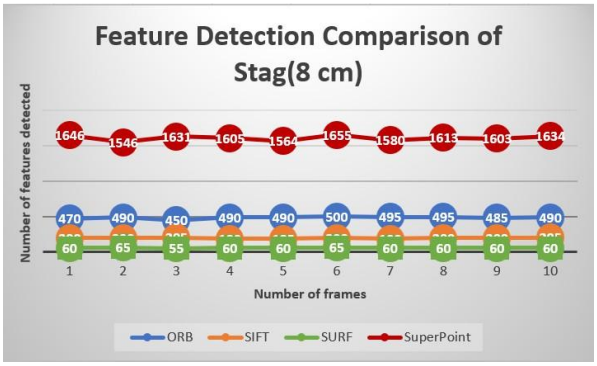
Fiducial Markers	Preprocessing Metrics and Values	
	SSIM	PSNR
AprilTag (1 cm)	0.9236	26.4914
ChArUco (4 cm)	0.9359	24.2064
STag (1 cm)	0.9327	26.6190
STag (8 cm)	0.9358	26.5520
STag (16 cm)	0.9361	26.7946
TopoTag (2 cm)	0.9341	27.0313
TopoTag (4 cm)	0.9343	26.2443
TopoTag (5 cm)	0.9348	26.6268



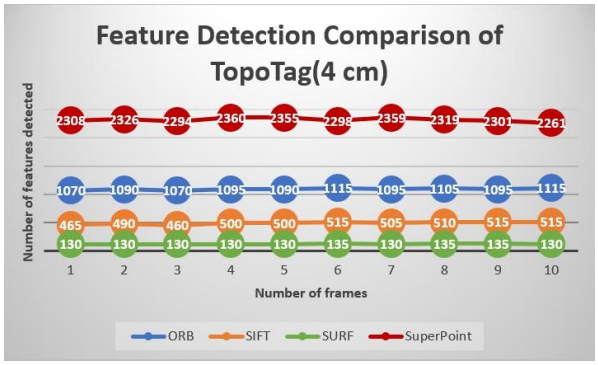
(a) Feature Detection Comparison of AprilTag (1 cm)



(b) Feature Detection Comparison of ChArUco (4 cm)



(c) Feature Detection Comparison of STag (8 cm)



(d) Feature Detection Comparison of ChArUco (4 cm)

Figure 2. Feature Detection Comparison of Fiducial Markers.

Figure 2(a) exhibits SuperPoint has the highest overall capacity, persistently maintaining an elite baseline value of approximately 2,510 points all over the entire testing range. ORB performs as the top-scoring classical, hand-crafted detector. It secures within a tight variance band between 1,150 and 1,200 points. SIFT holds the next performance bracket, indicating slight fluctuations around a mean value of around 500 points. SURF shows the lowest quantitative outline among the tested algorithms, yielding a flat baseline changing marginally between 40 and 45 points.

Next, in Figure 2(b), SuperPoint performs the highest baseline result, which shows a consistent track near the top of the evaluation scale with values varying within a stable band between 1,604 and 1,680 points. Next, ORB shows values between 1,320 and 1,415 points, tracing nearly below the deep learning baseline. SIFT performed as the mid-tier, exhibiting very small variations around a mean value of approximately 480 points. SURF is showing a flat baseline between 120 and 125 points.

Figure 2(c) shows SuperPoint only indicates the highest value range of feature points in between 1546 and 1655. Followed by ORB, the line shows values in between 450 and 495. Conversely, SIFT and SURF exhibit considerably lower performances. SIFT produces less than half the capacity of ORB, while SURF constantly remains flat-lined at the base of the metric chart. Finally in Figure 2(d), SuperPoint shows a premium range of feature points between 2261 and 2360. ORB shows, as always, the secondary position of getting feature points in the range of 1070 and 1115. This time comparatively SIFT attained almost half of the range of ORB, which is 460 to 515, whereas SURF shows a bottom constant line of approximate 130 to 135.

4.3 Computational and Real-Time Performance Comparison of Feature Extraction Methods

Here, Table 10 results show a clear trade-off between computational efficiency and overall processing performance among the four feature extraction methods. SuperPoint consistently demands the lowest computation time per image (approximately 0.0062–0.0075 s/image) throughout the evaluated fiducial markers, pointing to efficient feature extraction. However, except for STag (1 cm), SuperPoint shows considerably higher total processing times, lower FPS, higher frame latency, and lower throughput than the classical methods. Considering TopoTag (4 cm), SuperPoint requires 6.3339 s, achieving only 3.1576 FPS with a latency of 0.3166 s/frame, whereas SURF requires only 0.3420 s, achieving 58.49 FPS and 0.0171 s/frame. Between the classical methods, SURF regularly provides overall real-time performance, achieving the highest FPS and throughput and generally the lowest processing time and frame latency across the marker categories. SIFT then performs second best in terms of processing speed, while ORB provides medium performance. The STag (1 cm) outcome is an exception, where SuperPoint attains a 0.1359 s processing time, 147.16 FPS, 0.0067 s/frame latency, and 73.5353 MB/s throughput, outperforming all three classical methods.

Overall, the results show that SuperPoint has low feature-extraction computation time but does not constantly translate this gain into low end-to-end processing time, whereas SURF indicates the most consistent balance between computational demand and real-time processing performance across the evaluated fiducial markers. Every marker category is evaluated using 20 images, and it includes image loading, preprocessing, and other processing stages.

Next, in Table 11, the confusion matrix and classification performance of SuperPoint for 100 images are calculated. Therefore, positive pairs corresponded to image pairs containing the same marker instance or valid corresponding marker content, whereas negative pairs contained non-corresponding marker pairs.

Here in Equation(23),

$$FPR = \frac{25}{25+0} * 100 = 100\% \quad (23)$$

A negative pair incorrectly accepted as a valid correspondence is counted as a false positive, while a correctly rejected negative pair is counted as a true negative.

Table 11: SuperPoint Confusion Matrix and Classification Performance for 100 Images

4.4 Evaluation of Repeatability and Matching Performance of SuperPoint across Fiducial Markers

Next, an evaluation is conducted for repeatability score and matching score for each marker type in order to evaluate the robustness and consistency of the SuperPoint feature extraction network across fiducial marker designs. The ability of SuperPoint to recognize the same keypoints under different imaging settings, including viewpoint, scale and lightings, indicating feature stability is reflected by the repeatability score. Whereas the matching score measures feature matching reliability by assessing the identified keypoints and descriptors' ability to match marker pictures. High repeatability and matching scores indicate the marker's unique and consistent features, which are necessary for successful marker recognition, homography estimation, and pose tracking in augmented reality.

The results in Table 12 demonstrate that SuperPoint consistently achieved high repeatability, matching performance and detection accuracy across all fiducial marker types and sizes. 50 image pairs are matched for each marker category, and two images in each pair are selected as the first, third, fifth, seventh, and so on, matched with the second, fourth, sixth, eighth, and so on. Also, matching is like one-to-one. Among the evaluated markers, Stag (8 cm) illustrated the best overall performance with the highest repeatability score (99.92%), a good matching score (83.27%) and excellent matching accuracy (99.68%) indicating robust and reliable feature extraction. STag (8 cm) has an outer square boundary with strong corner features, while the internal circular structure produces weaker edge-based features. But strong corner features, which give good matching accuracy. While ChArUco (4 cm) attained the highest matching score (83.77%) and matching accuracy (99.61%) reflecting superior feature correspondence quality. ChArUco has a checkerboard structure with many well-defined corner intersections. So these corners produce strong gradients in multiple directions, which result in a high matching score. AprilTag (1 cm) produced the highest number of matched features (1954) with matching accuracy (99.33%).

Table 10. Comparison of Feature Extraction Methods Across Different Fiducial Markers

Evaluation Metrics	Fiducial Marker	Feature Extraction Method			
		ORB	SIFT	SURF	SuperPoint
Processing Time(s)	AprilTag (1 cm)	0.7477	0.6288	0.3252	3.2185
	ChArUco (4 cm)	0.7050	0.7257	0.4367	6.3214
	STag (1 cm)	1.8970	0.7877	0.7208	0.1359
	STag (8 cm)	0.8361	0.6674	0.3644	6.3834
	STag (16 cm)	0.6108	0.5635	0.4120	6.2360
	TopoTag (2 cm)	0.9956	0.6949	0.4346	6.0006
	TopoTag (4 cm)	0.6723	0.5792	0.3420	6.3339
	TopoTag (5 cm)	0.6506	0.6374	0.3229	5.8977
Frame Rate (FPS)	AprilTag (1 cm)	26.75	31.80	61.51	6.2140
	ChArUco (4 cm)	28.37	27.56	45.80	3.1638
	STag (1 cm)	10.54	25.39	27.75	147.16
	STag (8 cm)	23.92	29.97	54.89	3.1331
	STag (16 cm)	32.74	35.49	48.54	3.2072
	TopoTag (2 cm)	20.09	28.78	46.01	3.3330
	TopoTag (4 cm)	29.75	34.53	58.49	3.1576
	TopoTag (5 cm)	30.74	31.38	61.93	3.3911
Frame Latency (s/frame)	AprilTag (1 cm)	0.0373	0.0314	0.0162	0.1609
	ChArUco (4 cm)	0.0352	0.0362	0.0218	0.3160
	STag (1 cm)	0.0948	0.0392	0.0360	0.0067
	STag (8 cm)	0.0418	0.0333	0.0182	0.3191
	STag (16 cm)	0.0305	0.0282	0.0206	0.3117
	TopoTag (2 cm)	0.0497	0.0347	0.0217	0.3000
	TopoTag (4 cm)	0.0336	0.0289	0.0171	0.3166
	TopoTag (5 cm)	0.0325	0.0318	0.0161	0.2948
Computation Time(s/image)	AprilTag (1 cm)	0.0175	0.0303	0.0149	0.0073
	ChArUco (4 cm)	0.0165	0.0349	0.0198	0.0068
	STag (1 cm)	0.0459	0.0379	0.0339	0.0068
	STag (8 cm)	0.0175	0.0322	0.0167	0.0069
	STag (16 cm)	0.0147	0.0269	0.0188	0.0070
	TopoTag (2 cm)	0.0214	0.0335	0.0199	0.0067
	TopoTag (4 cm)	0.0153	0.0278	0.0156	0.0075
	TopoTag (5 cm)	0.0156	0.0306	0.0148	0.0062

Throughput (MB/s)	AprilTag (1 cm)	13.3737	15.9022	30.7527	3.1070
	ChArUco (4 cm)	14.1840	13.7806	22.8997	1.5819
	STag (1 cm)	5.2715	12.6969	13.8738	73.5353
	STag (8 cm)	11.9597	14.9840	27.4432	1.5666
	STag (16 cm)	16.3714	17.7468	24.2703	1.6036
	TopoTag (2 cm)	10.0445	14.3896	23.0074	1.6665
	TopoTag (4 cm)	14.8747	17.2641	29.2427	1.5788
	TopoTag (5 cm)	15.3716	15.6893	30.9666	1.6956

Table 11. SuperPoint Confusion Matrix and Classification Performance for 100 Images

Parameter	**Value**
Total Images	100
Total Image Pairs	50
Positive Pairs	25
Negative Pairs	25
Confusion Matrix	
True Positive (TP)	25
True Negative (TN)	0
False Positive (FP)	25
False Negative (FN)	0
Performance Measures	
Sensitivity / TPR	100.00%
Specificity / TNR	0.00%
Precision	50.00%
Accuracy	50.00%
False Positive Rate (FPR)	100.00%

Table 12. Performance Evaluation of SuperPoint on Different Fiducial Markers

Marker	Repeatability Score (%)	Mean Confidence Score	Matched Features	Matching Score (%)	Matching Accuracy (%)
AprilTag (1 cm)	99.64	0.023886	1954	78.25	99.33
ChArUco (4 cm)	96.01	0.031652	1285	83.77	99.61
STag (1 cm)	99.65	0.028891	748	66.49	97.33
STag (8 cm)	99.92	0.017209	1244	83.27	99.68
STag (16 cm)	98.77	0.017593	1095	73.59	98.54
TopoTag (2 cm)	99.02	0.030119	1877	73.87	96.06
TopoTag (4 cm)	97.53	0.025986	1879	83.29	99.52
TopoTag (5 cm)	95.08	0.019155	281	61.22	96.80

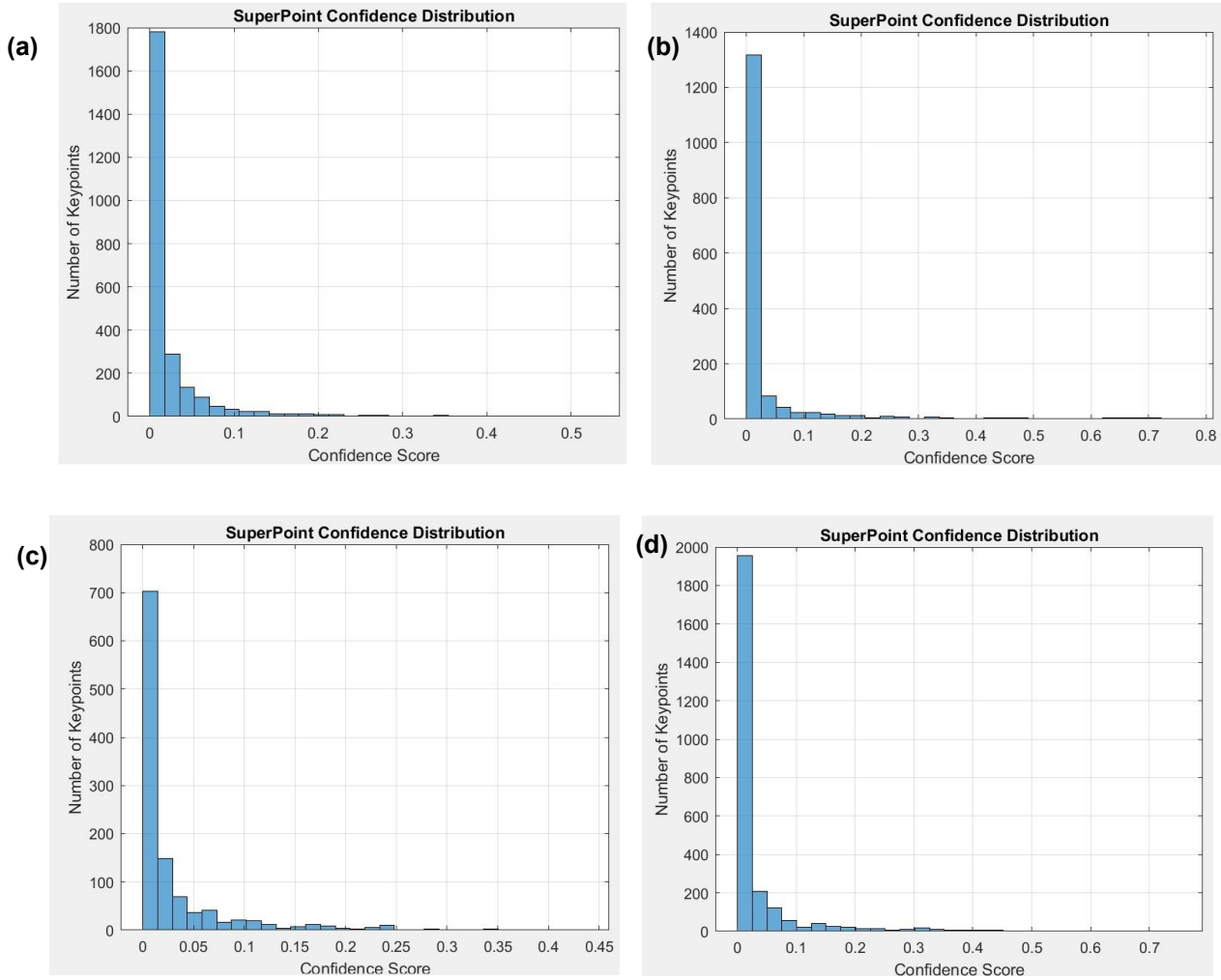


Figure 3. Confidence-score Distributions of Fiducial Markers, (a) Confidence distribution on AprilTag (1 cm), (b) Confidence distribution on ChArUco (4 cm), (c) Confidence distribution on STag (1 cm), (d) Confidence distribution on TopoTag (2 cm).

TopoTag (4 cm) also showed good performance with a matching score of (83.29%) and matching accuracy of (99.52%). In general, the results confirm that SuperPoint can constructively detect and describe stable keypoints on different fiducial markers, providing reliable feature matching and high detecting accuracy for augmented reality and marker-based vision applications.

ChArUco (4 cm) achieved the highest mean confidence score, that is, 0.031652, indicating that its corner-rich checkerboard structure generates the most distinctive and reliable SuperPoint keypoints. TopoTag (2 cm) and STag (1 cm) also exhibit good confidence values, 0.030119 and 0.028891, respectively, due to their unique geometric patterns. Although STag (8 cm) and STag (16 cm) produced the lowest confidence scores, such as 0.017209 and 0.017593, respectively, suggesting that repeatability and confidence measure different aspects of feature quality. Overall, markers with dense, high-contrast, and geometrically distinctive corner structures tend to produce higher SuperPoint confidence scores.

Here, Figure 3 shows a right-skewed distribution of SuperPoint confidence distribution with few keypoints scoring high confidence scores and majority of detected keypoints with low confidence values concentrated near zero. SuperPoint creates a wide range of candidate features but only holds a small fraction of very unique and reliable keypoints that connect to high-gradient areas, marker boundaries, and strong corners. As per the result shown by the steady decrease in keypoint frequency with rising confidence, it illustrated that more discriminative features are given greater confidence values in effective feature prioritization. A steady and robust feature detection is offered by SuperPoint by combining a high density of detected keypoints with a smaller set of highly confident keypoints that greatly support precise feature matching and marker localization.

4.5 Statistical Significance Evaluation of Feature Matching Accuracy across Methods

Here, Table 13 analogizes the matching accuracy performance of SuperPoint, ORB, SIFT, and

SURF over 50 image pairs. In this evaluation, ORB attained the highest mean accuracy (100.00%) with zero standard deviation, then followed by SuperPoint (99.83%), SIFT (99.62%), and SURF (98.02%). The 95% confidence intervals mean that ORB displayed completely consistent performance, while SURF exhibited the greatest variability. Overall, all methods achieved high matching accuracy, with ORB providing the highest and most consistent performance among the four methods.

Interpretation relative to prior work: SuperPoint was originally proposed as a joint learned detector and descriptor and evaluated on general image-matching benchmarks [7]. In the present ChArUco comparison, ORB has the highest mean matching accuracy (100.00%), while SuperPoint is 99.83%; hence these data do not show SuperPoint superiority in this individual metric. A different benchmark, selected-correspondence cap, and geometric verification procedure also preclude direct numerical comparison with the original SuperPoint paper [7] or learned matchers such as SuperGlue [8] and LightGlue [9].

Now, the resulting Friedman test statistics, *p*-values, and Kendall's *W* values are presented in Table 14. Repeatability showed a significant difference ($\chi^2(3) = 570.34$, $p < 0.05$, $W = 0.19011$), while matching accuracy also differed significantly ($\chi^2(3) = 1089.20$, $p < 0.05$, $W = 0.36308$). Significant differences are further observed for feature count ($\chi^2(3) = 1659.90$, $p < 0.05$, $W = 0.55330$) and processing time ($\chi^2(3) = 2659.80$, $p < 0.05$, $W = 0.88660$). Among the evaluated metrics, processing time exhibited the largest Kendall's *W*, indicating the strongest method-related effect. The Friedman test produces an overall comparison and does not determine which specific pairs of methods differ. The analysis is performed using 1,000 matched experimental blocks, with statistical significance assessed at $\alpha = 0.05$.

Table 14: Friedman repeated-measures statistical significance analysis of ORB, SIFT, SURF, and SuperPoint across the evaluated metrics.

Next in Table 15, pairwise comparisons are performed between ORB, SIFT, SURF, and SuperPoint using the Wilcoxon signed-rank test on 1,000 matched experimental blocks. The Holm method is applied to adjust the *p*-values for multiple comparisons. Statistical significance is assessed at $\alpha = 0.05$. "Yes" indicates a statistically significant difference after Holm adjustment, whereas "No" indicates that the difference is not statistically significant.

4.6 Robustness Performance Evaluation under Controlled Image Degradations

SuperPoint and SURF are presented here and evaluated under independently controlled illumination, blur, rotation, viewpoint, and partial-occlusion conditions, with performance quantified at multiple severity levels using repeatability, matching accuracy, feature count, and processing time by ChArUco (4 cm), which is selected as one of the markers from the dataset category. Here, Tables 16 and 17 present the overall degradation analysis, whereas Tables 18–25 present the subsequent condition-specific analysis in which illumination, blur, rotation, viewpoint, and occlusion are independently evaluated at severity levels from 0 to 90. Since these analyses use different experimental settings and aggregation procedures, their numerical values are not expected to be identical.

Table 16 shows the performance of SuperPoint under increasing degradation severity levels from 0% to 90%. Here, repeatability ranges from 96.824% to 99.992%, while matching accuracy remains between 99.745% and 99.900%. The mean processing time varies from 0.00089276 s to 0.010692 s, indicating consistently low computational time. The mean feature count remains relatively stable, ranging from 1609.6 to 1642.8 features, indicating that the number of detected features is not substantially affected by increasing degradation. Thus, overall the results show that SuperPoint maintains stable feature detection and matching performance under increasing degradation severity.

Table 17 presents the overall SURF performance across degradation severity levels from 0% to 90%. The mean repeatability ranges from 95.2% to 99.9%, while the mean matching accuracy varies from 93.979% to 100.000%. The processing time increases substantially from 6.8426 s at 0% severity to 98.047 s at 90% severity, indicating increased computational time under higher degradation levels. The matching stage consistently retains the maximum 25 correspondences at all severity levels.

Table 13. Matching Accuracy Performance of ORB, SIFT, SURF, and SuperPoint

Method	N	Mean (%)	SD (%)	95% CI Lower	95% CI Upper
SuperPoint	50	99.83	0.12	99.80	99.86
ORB	50	100.00	0.00	100.00	100.00
SIFT	50	99.62	1.14	99.30	99.95
SURF	50	98.02	3.66	96.99	99.06

Table 14. Friedman repeated-measures statistical significance analysis of ORB, SIFT, SURF, and SuperPoint across the evaluated metrics.

Metric	N	χ2	df	p-value	Kendall's W
Repeatability	1000	570.34	3	2.3554 × 10−167	0.19011
Matching Accuracy	1000	1089.20	3	4.6144 × 10−317	0.36308
Feature Count	1000	1659.90	3	< 2.2 × 10−16	0.55330
Processing Time	1000	2659.80	3	< 2.2 × 10−16	0.88660

Table 15. Pairwise statistical significance analysis using the Wilcoxon signed-rank test with Holm adjustment.

Metric	Method A	Method B	N	Wilcoxon W	Raw p-value	Holm-adjusted p-value	Significant
Repeatability	ORB	SIFT	1000	18413	1.547 × 10−87	9.2822 × 10−87	Yes
Repeatability	ORB	SURF	1000	20130	6.1245 × 10−85	3.0622 × 10−84	Yes
Repeatability	ORB	SuperPoint	1000	44334	1.4087 × 10−53	4.2262 × 10−53	Yes
Repeatability	SIFT	SURF	1000	57750	9.3185 × 10−36	1.8637 × 10−35	Yes
Repeatability	SIFT	SuperPoint	1000	1.37 × 105	0.10659	0.10659	No
Repeatability	SURF	SuperPoint	1000	2.3384 × 105	8.6391 × 10−85	3.4554 × 10−84	Yes
Matching Accuracy	ORB	SIFT	1000	0	7.7678 × 10−143	4.6607 × 10−142	Yes
Matching Accuracy	ORB	SURF	1000	0	6.0664 × 10−134	3.0332 × 10−133	Yes
Matching Accuracy	ORB	SuperPoint	1000	0	4.5119 × 10−129	1.8048 × 10−128	Yes
Matching Accuracy	SIFT	SURF	1000	67983	1.8412 × 10−15	5.5236 × 10−15	Yes
Matching Accuracy	SIFT	SuperPoint	1000	1.2883 × 105	9.7039 × 10−11	1.9408 × 10−10	Yes
Matching Accuracy	SURF	SuperPoint	1000	1.1604 × 105	0.054395	0.054395	No
Feature Count	ORB	SIFT	1000	4.9295 × 105	1.6494 × 10−155	6.5978 × 10−155	Yes
Feature Count	ORB	SURF	1000	4.9314 × 105	9.4011 × 10−156	4.7006 × 10−155	Yes
Feature Count	ORB	SuperPoint	1000	4.886 × 105	4.7497 × 10−150	1.4249 × 10−149	Yes
Feature Count	SIFT	SURF	1000	4.9651 × 105	1.4928 × 10−164	8.9576 × 10−164	Yes
Feature Count	SIFT	SuperPoint	1000	4.7529 × 105	5.5173 × 10−134	1.1035 × 10−133	Yes
Feature Count	SURF	SuperPoint	1000	2.8968 × 105	1.5861 × 10−105	1.5861 × 10−105	Yes

Processing Time	ORB	SIFT	1000	0	3.3259 × 10−165	1.9955 × 10−164	Yes
Processing Time	ORB	SURF	1000	0	3.3259 × 10−165	1.6629 × 10−164	Yes
Processing Time	ORB	SuperPoint	1000	0	3.3259 × 10−165	9.9776 × 10−165	Yes
Processing Time	SIFT	SURF	1000	5.005 × 105	3.3259 × 10−165	1.3303 × 10−164	Yes
Processing Time	SIFT	SuperPoint	1000	1.8924 × 105	2.4227 × 10−11	2.4227 × 10−11	Yes
Processing Time	SURF	SuperPoint	1000	1130	9.8189 × 10−164	1.9638 × 10−163	Yes

Table 16. SuperPoint performance under different degradation severity levels

Severity (%)	Rep. Mean (%)	Match. Acc. Mean (%)	Proc. Time Mean (s)	Feature Count Mean
0	99.992	99.889	0.010692	1632.6
10	99.727	99.775	0.0010098	1617.2
20	98.391	99.843	0.0010496	1630.8
30	96.824	99.837	0.00089556	1620.0
40	97.626	99.745	0.00089402	1609.6
50	97.680	99.900	0.00092382	1631.2
60	98.195	99.816	0.00091640	1614.8
70	97.943	99.890	0.00089442	1631.4
80	98.486	99.856	0.00089276	1642.8
90	98.256	99.744	0.0010034	1628.6

Table 17. SURF performance under different degradation severity levels

Severity (%)	Rep. Mean (%)	Match. Acc. Mean (%)	Proc. Time Mean (s)	Retained Correspondences (Max. 25)
0	99.9	97.5	6.8426	25
10	95.2	93.979	17.608	25
20	97.6	97.889	27.856	25
30	99.2	97.778	37.858	25
40	98.4	97.025	48.083	25
50	96.0	99.091	58.114	25
60	98.4	100.000	68.002	25
70	96.0	98.889	78.102	25
80	97.6	99.091	88.037	25
90	96.8	99.000	98.047	25

Table 18. SuperPoint condition-specific repeatability under different degradation and transformation severity levels.

Severity Level	Illumination Mean (%)	Blur Mean (%)	Rotation Mean (%)	Viewpoint Mean (%)	Occlusion Mean (%)
0	100.00	100.00	100.00	100.00	100.00
10	96.60	97.80	96.12	89.24	81.56
20	94.32	93.08	94.25	85.08	74.08
30	93.28	92.00	93.00	81.68	71.92
40	91.96	90.88	91.56	78.12	68.80
50	91.28	90.64	91.17	71.96	64.44
60	90.56	90.20	90.01	70.60	61.96
70	88.72	87.12	87.98	67.64	59.96
80	85.68	84.08	84.33	65.48	57.96
90	82.32	80.76	81.45	64.40	55.96

Table 19. SURF condition-specific repeatability under different degradation and transformation severity levels.

Severity Level	Illumination Mean (%)	Blur Mean (%)	Rotation Mean (%)	Viewpoint Mean (%)	Occlusion Mean (%)
0	100.00	99.28	99.97	100.00	100.00
10	99.80	91.76	99.12	93.88	92.84
20	99.80	89.60	99.10	93.88	72.96
30	99.68	88.52	99.00	75.16	72.96
40	99.44	85.84	99.34	73.84	57.64
50	99.52	85.60	99.46	74.44	53.64
60	99.52	85.40	99.43	69.92	36.36
70	99.52	83.04	99.41	82.64	32.36
80	99.28	81.00	99.22	82.88	24.40
90	99.11	80.00	89.55	65.52	24.40

Table 20. SuperPoint condition-specific matching accuracy under different degradation and transformation severity levels.

Severity Level	Illumination Mean (%)	Blur Mean (%)	Rotation Mean (%)	Viewpoint Mean (%)	Occlusion Mean (%)
0	100.00	99.875	99.98	100.00	100.00
10	99.715	97.539	99.65	94.263	98.513
20	99.254	93.443	99.20	88.887	98.300
30	99.197	91.752	99.14	84.888	87.839
40	99.184	76.665	99.11	82.624	94.996
50	99.097	72.896	99.00	77.967	93.490
60	99.219	72.533	99.20	76.424	93.922
70	99.506	77.869	99.46	72.617	96.167
80	99.031	73.231	99.00	69.00	95.825
90	98.805	77.361	98.76	63.45	95.611

Table 21. SURF condition-specific matching accuracy under different degradation and transformation severity levels.

Severity Level	Illumination Mean (%)	Blur Mean (%)	Rotation Mean (%)	Viewpoint Mean (%)	Occlusion Mean (%)
0	100.00	99.045	100.00	100.00	100.00
10	99.958	81.033	83.374	47.246	99.667
20	100.00	89.279	84.905	37.914	99.526
30	100.00	53.867	81.630	39.433	99.354
40	100.00	42.283	74.788	31.627	99.804
50	100.00	4.500	81.181	30.596	95.110
60	99.957	2.000	77.938	4.7333	98.236
70	99.828	0.500	74.646	3.400	89.938
80	99.578	1.230	68.860	7.000	89.831
90	97.00	1.200	46.800	4.600	89.831

Table 22. SuperPoint condition-specific feature count under different degradation and transformation severity levels.

Severity Level	Illumination Mean	Blur Mean	Rotation Mean	Viewpoint Mean	Occlusion Mean
0	1641.1	1580.7	1637.78	1641.1	1641.1
10	1644.7	1298.0	1641.61	1497.4	1608.5
20	1645.8	1092.9	1639.57	1840.8	1506.7
30	1642.6	1007.1	1636.64	1819.9	1521.9
40	1645.8	1020.4	1641.11	1647.1	1320.4
50	1657.9	1063.4	1652.42	1600.6	1320.8
60	1650.7	1155.3	1647.98	1565.3	1202.8
70	1628.5	1296.4	1619.88	1485.2	1101.7
80	1598.0	1373.4	1590.33	1443.8	1068.5
90	1577.9	1430.7	1569.48	1474.7	1036.4

Hence, Table 16 and Table 17 show the comparison of ChArUco markers of size 4cm between SuperPoint and SURF, based on increasing disturbance severity, to evaluate their robustness in terms of processing time, matching accuracy, repeatability, and feature count. SuperPoint shows consistently higher repeatability and matching accuracy with substantially lower processing time than SURF across the degradation severity levels.

Here, Table 18 presents the SuperPoint repeatability under different degradation and transformation severity levels. At 0% severity, repeatability is 100.00% for all conditions. At 90% severity, it decreases to 82.32% for illumination, 80.76% for blur, 81.45% for rotation, 64.40% for viewpoint, and 55.96% for occlusion. Overall, repeatability decreases with increasing severity, with occlusion and viewpoint changes producing the largest reductions, while illumination, blur, and rotation retain comparatively higher repeatability.

And Table 19 presents the SURF repeatability under different degradation and transformation severity levels. At 0% severity, repeatability is 100.00% for illumination, viewpoint, and occlusion, with 99.28% for blur and 99.97% for rotation. At 90% severity, it decreases to 99.11% for illumination, 80.00% for blur, 89.55% for rotation, 65.52% for viewpoint, and 24.40% for occlusion. Overall, SURF maintains high repeatability under illumination changes but shows substantial degradation under occlusion, viewpoint changes, and blur as severity increases.

Next, Table 20 presents the SuperPoint matching accuracy under different degradation and transformation severity levels. At 0% severity, the matching accuracy is 100.00% for illumination, viewpoint, and occlusion, 99.875% for blur, and 99.98%

for rotation. At 90% severity, it decreases to 98.805% for illumination, 77.361% for blur, 98.76% for rotation, 63.45% for viewpoint, and 95.611% for occlusion. Overall, SuperPoint maintains high matching accuracy under illumination and rotation, while viewpoint and blur produce the largest reductions at high severity levels.

Comparison with the literature: The marked declines under severe blur and viewpoint changes should qualify generalized robustness claims. The fiducial-detection studies DeepTag [11] and DeepArUco++ [12] likewise motivate evaluation under difficult imaging conditions, although their reported detection or pose metrics are not numerically interchangeable with the feature-matching accuracy used here. A fair cross-paper comparison would require identical marker images, degradation definitions, ground-truth correspondences, and matching protocols.

Then, Table 21 presents the SURF matching accuracy under different degradation and transformation severity levels. At 0% severity, the accuracy is 100.00% for illumination, rotation, viewpoint, and occlusion, while blur gives 99.045%. At 90% severity, the accuracy decreases to 97.00% for illumination, 1.200% for blur, 46.800% for rotation, 4.600% for viewpoint, and 89.831% for occlusion. Overall, SURF remains highly accurate under illumination changes but shows substantial degradation under blur, viewpoint, and rotation, while occlusion retains relatively higher accuracy despite a decline from the baseline.

Table 22 presents the SuperPoint feature count under different degradation and transformation severity levels. At 0% severity, the detected feature counts range from 1580.7 (blur) to 1641.1 (illumination, viewpoint, and occlusion). At 90% severity, the counts are 1577.9 for illumination, 1430.7 for blur, 1569.48 for rotation, 1474.7 for viewpoint, and 1036.4 for occlusion. Overall, SuperPoint maintains a relatively stable number of detected features under illumination and rotation, while occlusion produces the largest reduction, followed by blur and viewpoint changes.

Next, Table 23 presents the mean number of SURF features detected under different degradation and transformation severity levels. At 0% severity, the feature count ranges from 1987.0 for blur to 2133.2 for illumination, rotation, viewpoint, and occlusion. At 90% severity, the counts decrease to 46.72 for illumination, 325.04 for blur, 2476.2 for rotation, 1069.7 for viewpoint, and 859.1 for occlusion. Overall, SURF shows a substantial reduction in detected features under illumination and blur, while the feature count under rotation remains comparatively high and even increases at higher severity levels. Table 23 values represent the mean number of features detected by SURF before correspondence selection. Up to 25 correspondences are subsequently retained for matching evaluation where sufficient valid features are available. That is, SURF detected more than 25 candidate features at this severity level, allowing up to 25 correspondences to be retained for the standardized matching evaluation. The detected feature count and retained correspondence count represent different stages of the SURF pipeline. Hence, Table 23 reports the mean number of SURF features detected under each degradation condition, whereas Table 17 reports the number of correspondences retained for subsequent matching evaluation. A maximum of 25 correspondences is used for all methods to maintain a consistent computational budget.

Table 24 presents the SuperPoint condition-specific processing time under different degradation and transformation severity levels. At 0% severity, the processing time is 0.43072 s for illumination, 0.38205 s for blur, 0.089606 s for rotation, 0.40690 s for viewpoint, and 0.40876 s for occlusion. At 90% severity, these values are 0.38033 s, 0.32992 s, 0.090791 s, 0.35225 s, and 0.21382 s, respectively. Overall, rotation remains the most computationally stable condition, while illumination shows the largest processing-time variation across severity levels.

Table 23. Mean number of SURF features detected under different degradation and transformation severity levels.

Severity Level	Illumination Mean	Blur Mean	Rotation Mean	Viewpoint Mean	Occlusion Mean
0	2133.2	1987.0	2133.2	2133.2	2133.2
10	2042.9	1830.2	2058.0	1922.8	2019.7
20	1923.3	1317.8	2101.3	1963.8	1818.4
30	1728.5	631.92	2113.4	1798.6	1650.1
40	1443.5	473.98	2319.1	1584.5	1485.4
50	1081.6	427.21	2563.7	1459.3	1357.0
60	962.09	403.57	2625.7	1396.2	1128.2
70	249.95	377.29	2584.9	1280.0	965.03
80	49.34	353.95	2517.9	1129.5	860.67
90	46.72	325.04	2476.2	1069.7	859.1

Table 24. SuperPoint condition-specific processing time under different degradation and transformation severity levels.

Severity Level	Illumination Mean (s)	Blur Mean (s)	Rotation Mean (s)	Viewpoint Mean (s)	Occlusion Mean (s)
0	0.43072	0.38205	0.08966	0.40690	0.40876
10	0.85974	0.28573	0.089357	0.35210	0.42524
20	1.44090	0.23108	0.089555	0.48627	0.36062
30	1.45440	0.20739	0.093988	0.47713	0.35912
40	1.43690	0.21497	0.090422	0.40614	0.29603
50	1.30520	0.22103	0.09129	0.38850	0.29291
60	0.51957	0.24800	0.091174	0.37882	0.25927
70	0.45200	0.28689	0.091239	0.36411	0.22971
80	0.42411	0.31039	0.092578	0.33120	0.22234
90	0.38033	0.32992	0.090791	0.35225	0.21382

Table 25. SURF condition-specific processing time under different degradation and transformation severity levels.

Severity Level	Illumination Mean (s)	Blur Mean (s)	Rotation Mean (s)	Viewpoint Mean (s)	Occlusion Mean (s)
0	0.12046	0.11473	0.11528	0.11680	0.11439
10	0.11698	0.11063	0.11426	0.11262	0.11484
20	0.11494	0.10687	0.11524	0.11167	0.11389
30	0.11286	0.10277	0.11522	0.11007	0.11285
40	0.11294	0.10123	0.11621	0.11018	0.11113
50	0.10941	0.099486	0.11749	0.11071	0.10824
60	0.10504	0.10012	0.12014	0.11143	0.10513
70	0.10130	0.098737	0.11837	0.10762	0.10370
80	0.10068	0.098997	0.11685	0.10614	0.10254
90	0.097815	0.098789	0.11688	0.10603	0.10409

Finally, Table 25 presents the SURF condition-specific processing time under different degradation and transformation severity levels. At 0% severity, the processing times are 0.12046 s for illumination, 0.11473 s for blur, 0.11528 s for rotation, 0.11680 s for viewpoint, and 0.11439 s for occlusion. At 90% severity, these values become 0.097815 s, 0.098789 s, 0.11688 s, 0.10603 s, and 0.10409 s, respectively. Overall, SURF processing time remains relatively stable across the severity levels, with rotation showing the least variation, while illumination and blur show modest decreases as severity increases.

4.7 Evaluation of SuperPoint Feature Detection, Matching, and Verification

Feature descriptors extracted from the corresponding marker images are matched to create feature correspondences between image pairs. The initial matches are further refined according to the defined matching criteria to exclude unreliable correspondences. Here in Figure 4, we discuss geometrically verified feature correspondences of the ChArUco (4 cm) marker after feature matching and geometric consistency verification. Here, the red circles on the left image indicate the detected keypoints, and the green markers on the right image show their respective matched keypoints. The yellow lines connect those matching feature pairs together. Within the ChArUco region, it observes a dense concentration of verified matches, which indicates that the marker contains numerous distinctive and repeatable feature points. The major part of matching lines are mostly parallel and well-aligned, which suggests a strong geometric consistency between the two views.

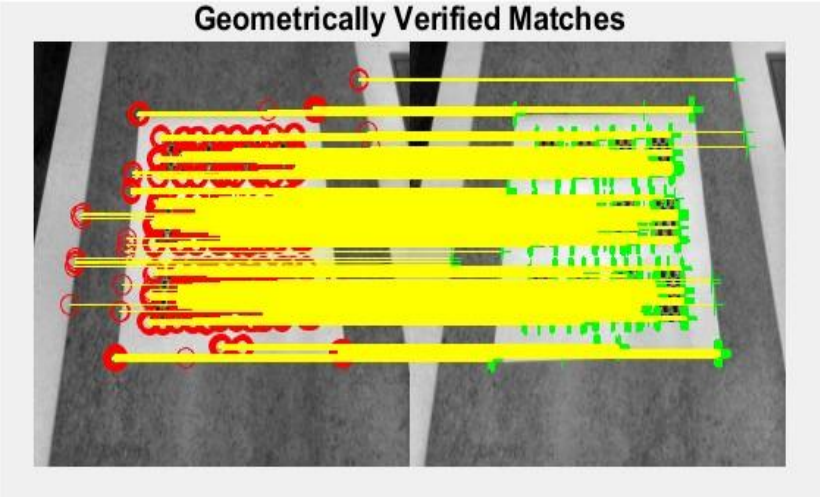


Figure 4. Geometrically verified matches of ChArUco (4 cm)

Table 26. SuperPoint Keypoint Detection and Initial Matching Analysis

Image Pair	KP1	KP2	Tol.	Repeated KP	Repeatability (%)	Initial Matches	Ratio Threshold
1	1604	1630	3	1542	96.13	1281	0.8
3	1610	1679	3	1579	98.07	1307	0.8
5	1640	1607	3	1592	99.06	1344	0.8
7	1607	1626	3	1545	96.14	1307	0.8
9	1632	1614	3	1569	97.21	1337	0.8
11	1684	1590	3	1630	96.79	1287	0.8
13	1622	1620	3	1566	96.66	1255	0.8
15	1638	1589	3	1615	98.59	1297	0.8
17	1611	1607	3	1569	97.63	1324	0.8
19	1687	1606	3	1641	97.27	1282	0.8

Table 27. SuperPoint Match Verification and Accuracy Analysis

Image Pair	Mutual Matches	RANSAC Tolerance	True Matches	False Matches	True Match Rate (%)	False Match Rate (%)
1	1211	3	1208	3	99.75	0.24773
3	1246	3	1246	0	100	0
5	1279	3	1279	0	100	0
7	1257	3	1257	0	100	0
9	1276	3	1276	0	100	0
11	1212	3	1211	1	99.91	0.082508
13	1178	3	1176	2	99.83	0.16978
15	1213	3	1213	0	100	0
17	1255	3	1255	0	100	0
19	1208	3	1205	3	99.75	0.24834

Although a few matches are located outside the marker boundary, the correspondences are correctly associated with the marker structure. Hence, the result demonstrates the high-contrast structure and abundant, reliable corner feature matching of ChArUco, making it highly suitable for robust marker detection and pose estimation.

Now, Table 26 outlines the keypoint detection and initial matching results for the evaluated image pairs. SuperPoint detected approximately 1,590–1,687 keypoints per image, with 1,542–1,641 repeated keypoints identified using a 3-pixel tolerance. Using a fixed 0.80 ratio threshold, the number of initial descriptor matches ranged from 1,255 to 1,344, while the resulting repeatability ranged from 96.14% to 99.06%. All these outcomes produce the key parameters and intermediate values required to reproduce the keypoint detection and initial matching process.

Then,Table 27 sketches the mutual matching and RANSAC-based geometric verification outcomes. With a fixed 3-pixel RANSAC tolerance, the number of mutual matches ranges from 1,178 to 1,279. True matches range from 1,176 to 1,279, although false matches are limited to 0–3. As a result, the true-match rate remains very high (99.75–100%), with false-match rates below 0.25%, confirming reliable descriptor matching and geometric verification.

4.8 Explainable Evaluation of SuperPoint Keypoint Detection across Fiducial Markers

The reason for dissimilarity in the number of keypoints detected by SuperPoint across various fiducial markers is a result of the difference in their elementary visual structures especially in terms of corner density, texture richness, gradient strength, and scale- contrast characteristics. It can be clearly interpreted using Explainable AI techniques like heatmap visualization. Sharp black-to-white transitions result in increased detection confidence; markers with a higher density of corners and fuller internal patterns offer more unique and repeatable features. When markers are well- scaled and have distinct contrast then the extraction of features is best. SuperPoint's capacity to identify high-information, stable local features are impacted by this combination of factors, resulting in fluctuations in the number and distribution of detected keypoints. Also, SuperPoint is fundamentally biased toward corners and high-gradient regions.

In AprilTag, Figure 5 shows SuperPoint detects strong keypoints at square boundaries and internal binary grid intersections which shows a high heatmap activation at corners and sharp black and white transitions. Additionally, a major drop in some detection points indicates the occlusion of corner points. When it comes to ChArUco, SuperPoint responds strongly to its checkboard junctions, and a dense heatmap peak indicates a repetitive corner intersection. Occluding a few corner points does not significantly affect detection due to redundancy.

Next, Figure 6 shows that in the case of TopoTag, SuperPoint struggles to detect feature points due to fewer sharp gradients, which results in spars and inconsistent heatmap activations. Because of these activations, when a small region is occluded, the effect on SuperPoint is normally unpredictable. TopoTag is constructed as a collection of various distributed nodes and connections that create various local patterns. Because of this fact, SuperPoint works effectively on topotag. It has excess of internal keypoint locations, each of which has robust local intensity variations, and SuperPoint can reliably detect them instead of relying exclusively on the outer corners. This results in a dense and evenly distributed set of keypoints all over the entire marker, which helps resilience to occlusion and viewpoint changes. But due to this dense and repetitive node structure of TopoTag, it can lead to overlapping activations and less discriminative features.

It is very clear in Figure 7 that Stag carries an outer square boundary that offers robust corner features for SuperPoint detection, whereas the inner circular region shows weakened edgebased features. So, a moderate heatmap response gets from fewer true corners, and a gradual performance drop highlights occlusion of circular edges. Even though the circular encoding of STag carries an outer square boundary that offers robust corner features for SuperPoint detection, the inner circular region provides weakened edge-based features. Because of this contrast, Stag shows a mix of strong feature points (from the outer square) and weak edge points (from the inner circle). Hence, the number of keypoints are moderate, not as many as grid-based markers, but still stable due to this kind of contrast in nature.

Then in Figure 8, ChArUco has a regular chessboard structure with countless well-defined corner intersections that produce significant intensity changes in multiple directions. In each grid intersection, the SuperPoint's preference for high-information and repeatable keypoints are precisely aligned, which results in dense and stable detections that are concentrated at these junctions. Moreover, the stark contrast between black and white regions and the gradient responses are further strengthened, and the uniform geometric layout promises that the detected keypoints are highly reliable for pose estimation and matching across various perspectives.

An AprilTag contains many internal bit junctions and outer square corners. Hence, it produces high-confidence keypoints. Next, AprilTag has high-contrast, black and white patterns. They show strong gradients and high-intensity changes, which leads to clear heatmap activations. Then, it has a structured grid layout with regular and repeatable geometry.

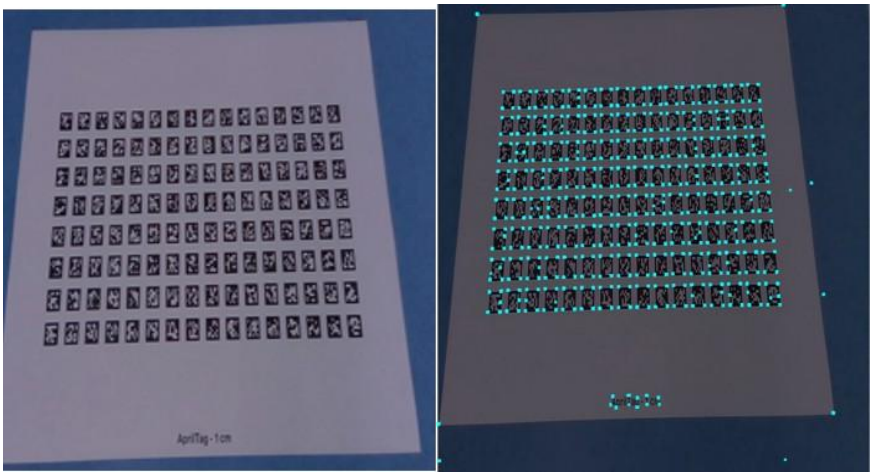


Figure 5. SuperPoint Heatmap with Keypoints of AprilTag (1cm), **(a)** Before Applying SuperPoint **(b)** After Applying SuperPoint.

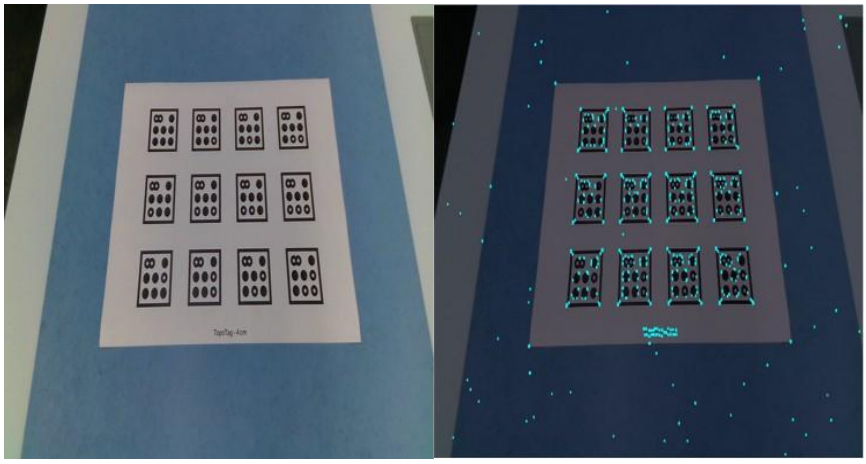


Figure 6: SuperPoint Heatmap with Keypoints of TopoTag (4cm), **(a)** Before Applying SuperPoint **(b)** After Applying SuperPoint

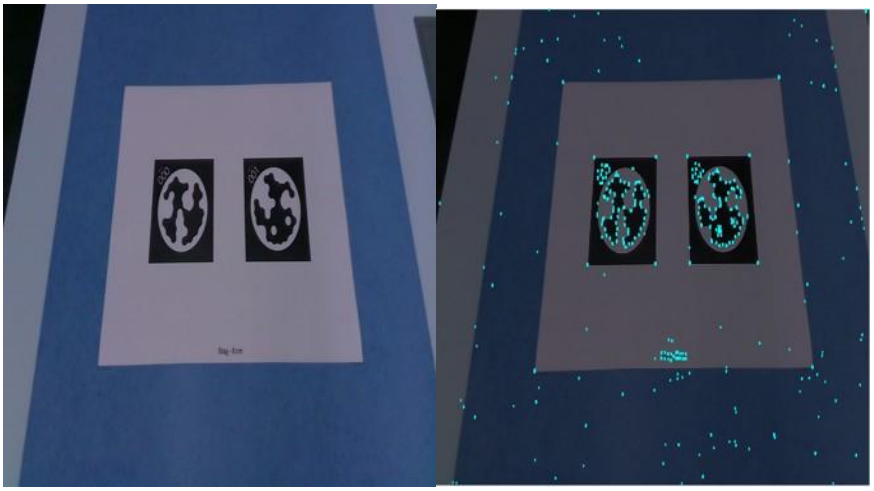


Figure 7. SuperPoint Heatmap with Keypoints of STag (8cm), **(a)** Before Applying SuperPoint **(b)** After Applying SuperPoint

So, it is very consistent and very easy to match across frames. Next, if we consider each small region of AprilTag patterns, it shows distinct patterns, and it improves descriptor matching quality. Also, it is good for robustness to occlusion and warping. Even with the fact that ChArUco offers a high density of corner features. Conversely, the repetitive grid pattern in ChArUco results in similar descriptors and potential mismatches, whereas the unique arrangement of intensity patterns in AprilTag generates highly discriminative descriptors, reducing ambiguity during feature matching.

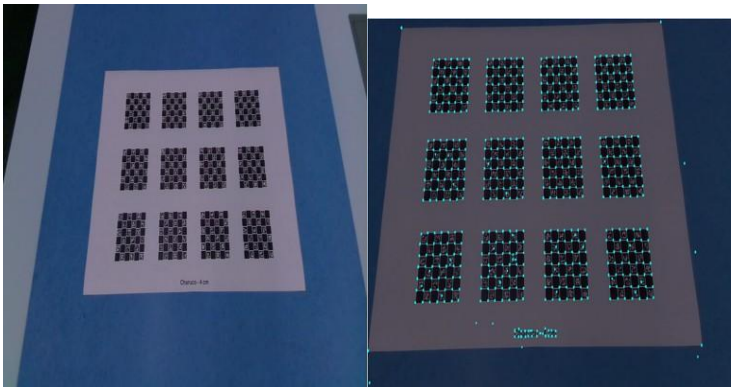


Figure 8. SuperPoint Heatmap with Keypoints of ChArUco (4cm), **(a)** Before Applying SuperPoint **(b)** After Applying SuperPoint

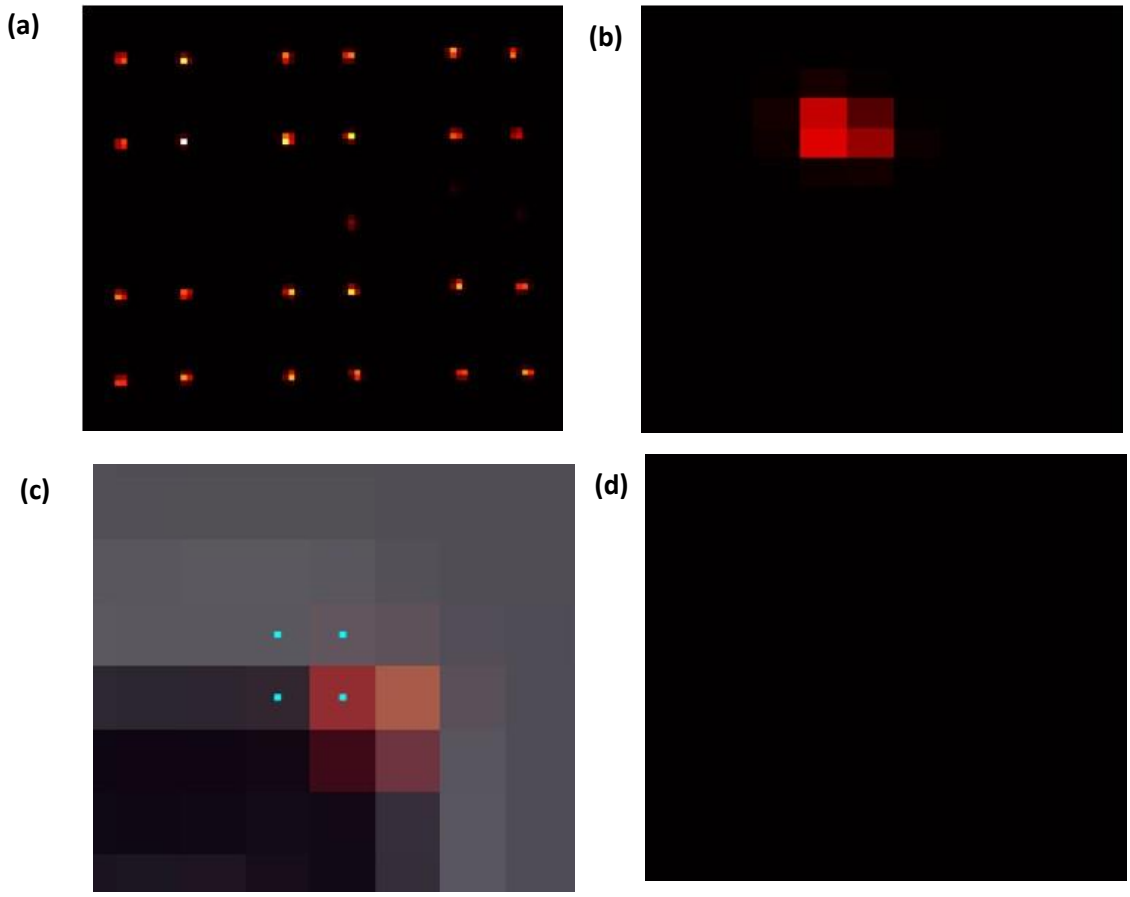


Figure 9. Experimental results of Occlusion sensitivity test, **(a)** Maximized Corner Response, **(b)** Low Confidence Regions Detected, **(c)** Removal impact test (Score drops on Occlusion), **(d)** Zero response corner region (black) detected

Consider balanced geographic distribution of keypoints of AprilTag and remaining robustness in the case of occlusion, also with the analysis report that indicates more stable nature and reliable feature detection. AprilTag is very compatible with the SuperPoint method. Hence, AprilTag works well with SuperPoint's learned features, and we got the maximum number of true corners compared to all other tags such as ChArUco, STag, and TopoTag.

Now, Figure 9 shows experimental results of an occlusion sensitivity test. It shows that these points correspond to a high response region in the detector output, which indicates those points are highly informative for matching. Here, an occlusion patch size of 32x32 pixels is moved across the image with a stride of 8 pixels.

Also, a keypoint is more dependable when the gradients exist in multiple directions. Also, that particular location shows strong gradients in both horizontal and vertical directions. Hence Occlusion sensitivity test validates that each direction contributes substantially to the decision. The fact is a corner point makes it more robust for matching other than edge and flat regions. Next, occlusion sensitivity analyses when a specific

region is masked or removed, the confidence score drops notably. That shows this region is very significant for the model's decision. Finally, the SuperPoint confidence heatmap shows a clear peak at the location, that helps to confirm the network allocates a high probability to being a unique and reliable feature point. The structure tensor, or second-moment matrix, is a computer vision mathematical tool which examines how image intensity changes in a small region or patch. Simply it helps to identify whether a region is flat, an edge, or a corner. Regions with high eigenvalues in both directions identify the corners and are considered highly informative for feature detection. Next point is that the local image patch about the corner has very high intensity fluctuations. On the part of geometric perspective, the corner is a stable and repeatable feature beneath various changeovers.

Here in Figure 10, a high-confidence region is detected on the AprilTag (1 cm) displayed.

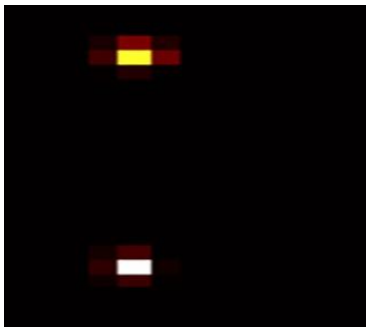


Figure 10. High-confidence regions detected—(white/yellow) on AprilTag(1 cm)

4.8.1 Quantitative Validation of Heatmap Interpretations

For the visual interpretation of the SuperPoint heatmaps, a quantitative validation is performed by comparing the marker and background regions. Moreover, it evaluates the average activation, activation ratio, detected keypoint count, keypoint density, and additional activation statistics, including maximum and median activation values. This is the numerical evidence of the spatial distribution and concentration of SuperPoint responses within the fiducial marker region compared with the surrounding background. Here, heatmap interpretation analysis is executed with ChArUco marker of size 4 cm.

According to Table 28, the quantitative heatmap analysis of the ChArUco marker, SuperPoint shows a stronger overall response within the marker region than in the background region. The mean activation in the marker region is 0.009813, compared with 0.001104 in the background, resulting in a reported activation ratio of 8.8886. Similarly, the marker region included 67 detected keypoints with a density of 0.32367150, while the background comprised 1537 keypoints with a density of 0.02352707. Hence, the marker region keypoint density is approximately 13.76 times higher than that of the background. These results hold that SuperPoint responses are more densely concentrated within the ChArUco marker region, although isolated high activation values occur in the background, as indicated by the maximum background activation of 0.836845.

Table 28. Automatic Heatmap Numerical Analysis for ChArUco (4cm)

Parameter	Value
Marker Type	ChArUco
Frame Number	1
Activation Analysis	
Marker-region activation	0.009813
Background activation	0.001104
Activation difference	0.008709
Activation ratio	8.8886
Keypoint Analysis	
Marker-region keypoints	67
Background keypoints	1537
Marker keypoint density	0.32367150
Background keypoint density	0.02352707
Keypoint density ratio	13.7574
Additional Activation Statistics	
Maximum marker activation	0.443862
Maximum background activation	0.836845
Median marker activation	0.000308
Median background activation	0.000000

Discussion in relation to published research: The greater concentration of SuperPoint keypoint density within the ChArUco marker than in the background is compatible with SuperPoint's learned interest-point response to repeatable local image structures [7]. Nonetheless, this experiment measures the spatial distribution of responses, not the causal internal reasoning of the network. The marker-specific heatmaps complement research on engineered fiducial geometry in TopoTag [4] and STag [19], but they do not directly validate or refute those marker systems' published pose-estimation claims.

Here, Figure 11 shows a color scale ranging from 0 (low activation) to 1 (high activation), indicating bright-colored points where SuperPoint detects stronger keypoint region responses, mainly around the marker structure and its edges/corners. The other view of the dark-blue background represents low activation.

4.8.2 Analysis of Occlusion Sensitivity

In Table 29, occlusion analysis indicates that masking has a limited effect on SuperPoint keypoint detection and confidence, while matching performance is more sensitive to localized information loss. The consistent reduction in matching score and the larger decreases in matching accuracy for some marker configurations demonstrate that occlusion primarily affects the reliability of feature correspondences rather than the initial feature-detection capability.

The percentage change in each metric under occlusion is calculated relative to the original unoccluded image using Equation (24),

$$\Delta X(\%) = \frac{X_{occ} - X_{base}}{X_{base}} * 100$$

where X_{base} represents the value obtained from the original unoccluded image and X_{occ} represents the value obtained after applying the occlusion mask. Positive values indicate an increase, whereas negative values indicate a decrease relative to the baseline.

4.9 Interpretation of Feature Extraction and Matching Performance

The cross-marker results in Table 12 indicate that marker geometry affects the number and reliability of local correspondences. In this study, STag (8 cm) achieved 99.92% repeatability and 99.68% matching accuracy, whereas ChArUco (4 cm) achieved the highest reported matching score (83.77%). AprilTag (1 cm) yielded 1,954 matched features. These outcomes are consistent with the broader premise that fiducial-marker layout influences detection and localization: TopoTag uses topological and geometric structure [4], while STag combines square and circular features to stabilize marker localization [19]. However, those publications evaluate their own marker-detection or pose-estimation systems, not SuperPoint-based matching on the present dataset. Accordingly, the present percentages should be interpreted as new observations from this study, not as demonstrated numerical improvements over [4] or [19]. The differing feature counts and matching scores also show that more detected or matched features do not automatically imply a higher fraction of successful correspondences.

Comparison with the foundational SuperPoint study [7] provides a more direct methodological context. DeTone et al. introduced learned joint detection and description and assessed repeatability and geometric matching on generic-image benchmarks. The current manuscript instead uses pretrained SuperPoint weights and evaluates eight configurations from four fiducial-marker families, alongside ORB, SIFT, and SURF. Thus, the contribution is a marker-specific benchmarking application rather than a new SuperPoint network or training method. Importantly, Table 13 reports mean matching accuracy of 100.00% for ORB, 99.83% for SuperPoint, 99.62% for SIFT, and 98.02% for SURF under the study's matching protocol. SuperPoint therefore cannot be described as uniformly best on matching accuracy. Its merits should be judged jointly using repeatability, feature count, runtime, and matching outcomes. Because the evaluation retains a maximum of 25 correspondences for the common comparison, the reported accuracy should also be distinguished from the overall detection rate or end-to-end pose-estimation performance.

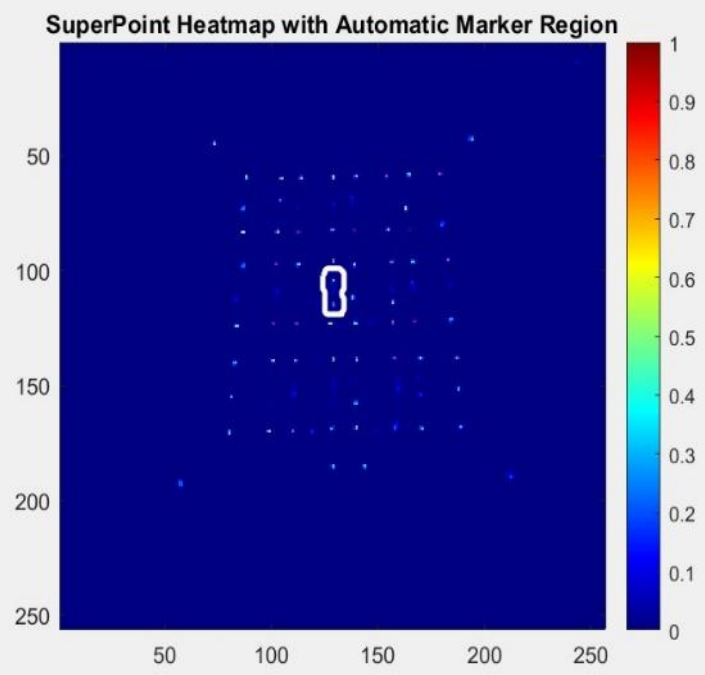


Figure 11. SuperPoint Heatmap with Automatic Marker Region.

Table 29. SuperPoint occlusion sensitivity analysis under the controlled experimental settings

Marker	Images/Pairs	Occlusion Setting	SP/Matching Setting	Keypoints	Confidence	Matching Score	Matching Accuracy (%)
AprilTag (1 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	340.70 / 340.24 −0.136%	0.077726 / 0.080026 +2.958%	0.23617 / 0.22382 −5.226%	70.000 / 52.610 −24.843%
ChArUco (4 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	195.80 / 199.49 +1.882%	0.169060 / 0.169260 +0.119%	0.65313 / 0.61675 −5.570%	99.333 / 97.873 −1.471%
STag (1 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	358.40 / 355.48 −0.814%	0.060731 / 0.063723 +4.927%	0.28129 / 0.26638 −5.303%	86.667 / 52.334 −39.615%
STag (8 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	154.00 / 156.77 +1.799%	0.074287 / 0.081910 +10.262%	0.50986 / 0.48388 −5.095%	100.000 / 85.219 −14.781%
STag (16 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	166.10 / 168.23 +1.284%	0.071619 / 0.079415 +10.885%	0.58111 / 0.54604 −6.035%	90.000 / 60.283 −33.019%
TopoTag (2 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	537.50 / 529.66 −1.458%	0.099402 / 0.100900 +1.511%	0.25539 / 0.24577 −3.767%	90.000 / 85.467 −5.037%
TopoTag (4 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	263.70 / 265.18 +0.560%	0.151300 / 0.152460 +0.770%	0.52312 / 0.49507 −5.362%	99.667 / 97.820 −1.853%
TopoTag (5 cm)	20 / 10	Input: 256×256 Mask: 32×32 Stride: 8; Value: 0 Positions: 841	Conf.: 0.0015 NMS: 1; Limit: 2500 Ratio: 0.80 RANSAC: 3 px	76.000 / 81.105 +6.7169%	0.053368 / 0.067696 +26.848%	0.71223 / 0.64652 −9.227%	99.167 / 31.163 −68.575%

The degradation analysis gives additional insight into applicability. At 90% degradation severity, SuperPoint repeatability on ChArUco (4 cm) fell to 55.96% for occlusion and 64.40% for viewpoint variation (Table 18), while its matching accuracy fell to 77.361% under blur and 63.45% under viewpoint changes (Table 20). The results therefore support condition-dependent robustness rather than uninterrupted near-perfect

performance. Existing studies such as DeepTag [11] and DeepArUco++ [12] address challenging fiducial-marker detection, including degraded imaging, but examine different detector pipelines and outcome measures. They supply relevant context, not directly comparable accuracy baselines. Likewise, SuperGlue [8], LightGlue [9], and OmniGlue [10] develop learned feature matchers, whereas the present experiments use SuperPoint descriptors followed by ratio filtering and RANSAC verification; their published performance figures cannot establish superiority or inferiority for the current setup.

The distinctive contribution of the present work is the controlled comparison of established feature extractors across marker structures and selected degradations, supplemented by heatmap and occlusion analyses. Its limitations include reliance on a dataset with unspecified acquisition conditions (Table 3), detailed degradation reporting focused on ChArUco (4 cm), and the absence of direct tests against these more recent matching and detection pipelines. Broader datasets and standardized end-to-end detection and pose benchmarks would be needed to establish generalizability.

5. Conclusion

This particular study illustrated a thorough comparative evaluation of classical feature extraction methods (such as ORB, SIFT, and SURF) and the deep learning method. SuperPoint perspective for different categories of fiducial markers (such as AprilTag, ChArUco, STag, and TopoTag). As per the experimental evaluation, the SuperPoint provides strong feature robustness and matching reliability for fiducial marker detection in comparison with evaluated classical feature extraction methods. The highest repeatability of 99.92% and matching accuracy of 99.68% are attained by Stag (8 cm), whereas ChArUco (4 cm) achieved the highest matching score of 83.77%. AprilTag (1 cm) generated the largest number of matched features, with 1954 correspondences. For real-time operation, SuperPoint with Stag (1 cm) achieved 147.16 FPS with a frame latency of 0.0067s/frame, showing its suitability for computationally constrained AR applications. We have clarified that no single marker demonstrates consistently superior performance across all evaluated criteria. Moreover, an additional observation showed by degradation analysis that SuperPoint sustains stable feature detection and matching performance under increasing disturbance severity. Altogether, the findings indicate that marker scale and structure influence the balance between matching reliability and computational efficiency, with SuperPoint ensuring a promising solution for robust and real-time fiducial marker detection in collaborative augmented reality applications. Future work can guide the insights of exploring hybrid models combining geometric priors with deep learning and integrate the outcomes with advanced Explainable Artificial Intelligence (XAI) techniques to improve interpretability in real-time AR systems.

References

[1] A.D. Fletcher, E. Cheadle, J. Davidson, D. Conniffe, F. Podd, A.J. Peyton, Evaluation of a low power computer vision-based positioning system for a handheld landmine detector using AprilTag markers. Instruments, 9(4), (2025) 27. https://doi.org/10.3390/instruments9040027

[2] M.C. Roshan, M. Isaksson, A. Pranata, A geometric calibration method for thermal cameras using a ChArUco board. Infrared Physics & Technology, 138, (2024) 105219. https://doi.org/10.1016/j.infrared.2024.105219

[3] J. R. Heselden, D. Paparas, R. L. Stevenson, G. P. Das, Enhanced STag marker system: Materials and methods for flexible robot localisation. Machines (Basel), 13(1), (2025) 2. https://doi.org/10.3390/machines13010002

[4] G. Yu, Y. Hu, J. Dai, TopoTag: A robust and scalable topological fiducial marker system. IEEE Transactions on Visualization and Computer Graphics, IEEE, 27(9), (2021) 3769–3780. https://doi.org/10.1109/TVCG.2020.2988466

[5] G. Petrakis, A SuperPoint neural network implementation for accurate feature extraction in unstructured environments. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-1/W2-2023, (2023) 1215–1222. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1215-2023

[6] S. Wang, Z. Zhu, Y. Jiang, Superpoint network-based video stabilization technology for mine rescue robots. Applied Sciences (Basel), 15(22), (2025) 12322. https://doi.org/10.3390/app152212322

[7] D. DeTone, T. Malisiewicz, A. Rabinovich, (2018) SuperPoint: Self-supervised interest point detection and description. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, USA. https://doi.org/10.1109/CVPRW.2018.00060

[8] P.E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich, (2020) SuperGlue: Learning feature matching with graph neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, USA. https://doi.org/10.1109/CVPR42600.2020.00499

[9] P. Lindenberger, P.-E. Sarlin, M. Pollefeys, (2023) LightGlue: Local feature matching at light speed. Proceedings of the IEEE/CVF International Conference on Computer Vision

(ICCV), IEEE, France. https://doi.org/10.1109/ICCV51070.2023.01616

[10] H. Jiang, A. Karpur, B. Cao, Q. Huang, A. Araujo, (2024) OmniGlue: Generalizable feature matching with foundation model guidance. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, USA. https://doi.org/10.1109/CVPR52733.2024.01878

[11] Z. Zhang, Y. Hu, G. Yu, J. Dai, DeepTag: A general framework for fiducial marker design and detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3), (2023) 2931–2944. https://doi.org/10.1109/TPAMI.2022.3174603

[12] R. Berral-Soler, R. Munoz-Salinas, R. Medina-Carnicer, M. J. Marín-Jimenez, DeepArUco++: Improved detection of square fiducial markers in challenging lighting conditions. Image and Vision Computing, 152, (2024) 105313. https://doi.org/10.1016/j.imavis.2024.105313

[13] E. Olson, AprilTag: A robust and flexible visual fiducial system. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), (2011) 3400–3407. https://doi.org/10.1109/ICRA.2011.5979561

[14] J. Wang, E. Olson, (2016) AprilTag 2: Efficient and robust fiducial detection. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Korea (South). https://doi.org/10.1109/IROS.2016.7759617

[15] M. Krogius, A. Haggenmiller, E. Olson, (2019) Flexible layouts for fiducial tags. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, China. https://doi.org/10.1109/IROS40897.2019.8967787

[16] S. Garrido-Jurado, R. Munoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jimenez, Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47(6), (2014) 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005

[17] D. Hu, D. DeTone, T. Malisiewicz, Deep ChArUco: Dark ChArUco marker pose estimation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (2019) 8428–8436. https://doi.org/10.1109/CVPR.2019.00863

[18] R. Munoz-Salinas, F.J. Romero-Ramirez, S. Garrido-Jurado, ChArUco2: Enhanced calibration boards with dual black-and-white marker detection. SoftwareX, 35, (2026) 102832. https://doi.org/10.1016/j.softx.2026.102832

[19] B. Benligiray, C. Topal, C. Akinlar, STag: A stable fiducial marker system. Image and Vision Computing, 89, (2019) 158–169. https://doi.org/10.1016/j.imavis.2019.06.007

[20] C. Shewell, J. Medina-Quero, M. Espinilla, C. Nugent, M. Donnelly, H.Y. Wang, Comparison of fiducial marker detection and object interaction in activities of daily living utilising a wearable vision sensor. International Journal of Communication Systems, 30(5), (2017) e3223. https://doi.org/10.1002/dac.3223

[21] Y. Z. Guleray, (2025) Fiducial Markers Analysis Dataset. GitHub repository, version 1.0.

Authors Contribution Statement

P.K. Anjana: Conceptualization, Methodology, Software, Validation, Formal analysis, investigation, Resources, Data curation, Original draft preparation, Writing Review and Editing, and Visualization. V. Ashok Immanuel: Supervision and Project Administration. Both the authors read and approved the final version of this manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.