POLITECNICO DI TORINO
Repository ISTITUZIONALE

A vision-driven framework for robotic garment stitching: edge-aware perception and trajectory planning

Original
A vision-driven framework for robotic garment stitching: edge-aware perception and trajectory planning / Bao, N., Wang, K., Simeone, A., Fan, Y., Xiang, Q., Liang, Y., Bao3, R.. - In: JOURNAL OF INTELLIGENT MANUFACTURING. - ISSN 0956-5515. - (2026). [10.1007/s10845-026-02941-0]

Availability:
This version is available at: 11583/3014747 since: 2026-08-22T09:15:27Z

Publisher:
Springer

Published
DOI:10.1007/s10845-026-02941-0

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

26 November 2026

Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-026-02941-0

A vision-driven framework for robotic garment stitching: edge-aware perception and trajectory planning

Nengsheng Bao[1] · Kewei Wang[1] · Alessandro Simeone[2] · Yuchen Fan[2] · Qingjiang Xiang[1] · Yali Liang[1] · Runxuan Bao[3]

Received: 29 April 2026 / Accepted: 20 July 2026
© The Author(s) 2026

Abstract

The garment sewing industry faces persistent operator shortages and inconsistent seam quality. This work presents a vision-based robotic system for automating the perimeter-stitch, i.e. the Yun operation in shirt manufacturing, covering cuff, collar, and pocket-flap assembly. An improved Holistically-nested Edge Detection (HED) network is proposed, incorporating a directionally-aware Frequency-Spatial (FS) attention module that preserves spatial coordinate information and enhances boundary localisation in texture-rich garment images. A domain-specific dataset was constructed using garment-panel images captured under front-lit and backlit illumination conditions. The improved model achieves an Optimal Dataset Scale (ODS) F-measure of 0.8517, improving over the baseline HED model by 4.02 percentage points in ODS and 4.21 percentage points in Optimal Image Scale (OIS). Comparisons with representative edge-detection models, including PiDiNet, DexiNed, and EDTER, further demonstrate that the proposed model achieves a favourable balance between accuracy and efficiency. Stitch coordinates are generated in approximately 10 s via contour extraction and polygon approximation. The MS6MT six-degree-of-freedom (6-DOF) manipulator executes Cartesian-space linear and circular-arc trajectories derived from these coordinates through camera and eye-to-hand calibration. Physical experiments demonstrate stitch placement accuracy within ± 1 mm, validating the feasibility of the system under controlled experimental conditions.

Keywords Stitching robot · Edge detection · HED · Attention mechanism · Trajectory planning · Garment manufacturing · Intelligent manufacturing

Introduction

The textile and apparel sector is one of the most significant traditional manufacturing industries in China, contributing substantially to national export revenue and employment. Following China's accession to the World Trade Organisation in 2001, the industry underwent sustained expansion in production scale and export volume, establishing China as the world's largest exporter of textiles and clothing, with its global market share rising from approximately 13.5% in 2000 to over 36% by 2021 (Guan et al., 2019; Yang & Kumarasinghe, 2024). Nevertheless, the sector faces mounting structural pressures. Labour costs have risen sharply, driven by an ageing workforce, declining recruitment of younger workers, and a persistent shortage of qualified operators, while international buyers and regulatory bodies are simultaneously demanding higher product quality and reduced environmental impact (CICC Research, CICC Global Institute 2024). China's dual-carbon commitment, announced in 2020 and targeting peak carbon emissions by 2030 and neutrality by 2060, has placed the textile industry under particular decarbonisation pressure given its status as one of the country's most energy-intensive manufacturing sectors (Chang & Zhu, 2022).

Shirt manufacturing is representative of these challenges. A single shirt comprises five major subassemblies, front

✉ Alessandro Simeone
alessandro.simeone@polito.it

1 College of Engineering, Shantou University, Shantou, Guangdong, China

2 Department of Management and Production Engineering, Politecnico Di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy

3 College of Engineering, Guangdong Technion-Israel Institute of Technology, Shantou, Guangdong, China

Published online: 29 July 2026

panel, back panel, sleeves, collar, and cuffs, each requiring a multi-step stitching sequence. The most prevalent of these steps is the Yun (transport-stitch) operation, a perimeter sewing process commonly used in shirt manufacturing, in which a cut garment panel is guided through a lockstitch sewing machine to form a seam at a prescribed offset from the panel edge. This operation is functionally equivalent to edge-following seam stitching and is performed repeatedly across multiple garment components such as cuffs, collars, and pocket flaps. It is currently performed either by skilled operators relying on tactile and visual feedback or by semi-automatic fixtures that require manual measurement and CAD entry of panel dimensions. Both approaches are slow to changeover when garment styles change, produce variable stitch quality, and consume significant operator attention, rendering them incompatible with the digital and intelligent manufacturing paradigms described in the 14th Five-Year Plan for the textile industry (China National Textile and Apparel Council (CNTAC) 2021).

Robotic sewing systems offer a compelling alternative. By coupling machine vision with robotic manipulation and a sewing head, it becomes possible to detect a panel's perimeter autonomously, plan a stitch trajectory, and execute the seam with limited operator intervention when required for ambiguous cases. This reflects a design choice in which automated perception handles the majority of cases, while operator supervision is retained to ensure reliability in the presence of geometric ambiguity. Human-in-the-loop learning has been identified as a key paradigm for advancing robot autonomy in smart manufacturing environments (Chen et al., 2025b; Fan et al., 2025). This is closely aligned with the broader emphasis on safe and effective human–robot collaboration in industrial settings, where balancing operator oversight with robot autonomy remains a central research theme (W. Li et al., 2024b; Patalas-Maliszewska et al., 2025). Recent advances in deep learning, particularly in pixel-level edge detection, enable improved boundary localisation under varying lighting and fabric conditions. Simultaneously, mature industrial robot kinematics and Cartesian-space trajectory planning methodologies provide the motion precision and repeatability required by garment quality standards.

Existing learning-based perception-planning (Chen et al. 2025c) frameworks for deformable object manipulation, such as those reviewed by Yin et al. (Yin et al., 2021) and demonstrated by Shen et al. (Shen et al., 2025), address general manipulation tasks but do not provide the boundary localisation precision required for stitch-path generation in garment manufacturing. The present work differs from these frameworks in three respects: it targets a specific industrial sewing operation (the Yun perimeter stitch) with defined geometric accuracy requirements ($\pm$ 1 mm); it introduces a direction-aware attention mechanism (Frequency-Spatial, FS module) designed specifically for texture-rich garment-panel boundary detection; and it integrates perception, planning, and operator supervision into a deployable workflow rather than addressing each component in isolation.

State-of-the-art

Sewing robot systems

Robotic approaches to garment sewing have evolved from early rule-based systems toward vision-integrated and learning-enabled platforms. Choi et al. (Choi et al., 2025) demonstrated vision-guided fabric manipulation using edge detection and geometric feature extraction for robotic cloth alignment, establishing the perception-motion coupling paradigm that subsequent systems have refined. The fundamental challenge of manipulating deformable textile objects, arising from their infinite-dimensional configuration space, self-occlusion, and complex dynamics, has since been comprehensively surveyed by Yin et al. (Yin et al., 2021), who identified garment sewing as a priority open problem at the intersection of perception, planning, and physical control.

More recent systems have demonstrated end-to-end automated sewing for specific seam geometries. Ku et al. (Ku et al., 2023) developed a machine-vision-enabled sewing system that deploys YOLOv5-based instance segmentation to localise an opening between acrylic guide plates, subsequently generating a constant-offset top-stitch trajectory with $\pm$ 68 μm/pixel spatial resolution; the work validates the viability of deep learning for stitch-path generation but is constrained to a single fabric template geometry. Song et al. (F. Li et al., 2024a) developed a robot sewing methodology based on process modelling of the collaborative feeding dynamics between a robot manipulator and a sewing machine head, addressing trajectory synchronisation but relying on manually specified seam boundaries. Zhu et al. (Zhu et al., 2022) provide a recent outlook on challenges in robotic manipulation of deformable objects, noting that the absence of reliable boundary perception under varying lighting and material conditions remains the primary bottleneck for production deployment. Recent work has also explored integrated robotic sewing systems combining machine vision and compliant manipulation. Shen et al. proposed an autonomous sewing system that integrates soft robotic end-effectors with visual perception to improve adaptability when handling deformable fabrics (Shen et al., 2025). However, such systems primarily focus on manipulation and hardware compliance, with limited emphasis on precise perception-driven stitch-path generation. Similarly, vision-guided robotic systems have been developed for garment handling and manipulation. Choi et al. proposed an adaptive fabric gripping system that integrates

visual perception with machine learning to determine optimal manipulation strategies for deformable materials (Choi et al., 2025).

Existing perception methods for deformable objects focus primarily on garment-level tasks such as classification, pose estimation, and unfolding, and do not address the sub-millimetre boundary localisation required for stitch-path generation. A consistent limitation of prior sewing systems is their reliance on template-matching or threshold-based vision, which requires manual parameter adjustment and fails to generalise across the material and lighting variation of production environments.

Edge detection

Classical gradient-based operators, Sobel, Canny (Canny, 1986), Prewitt, and Roberts, exploit local intensity discontinuities but are sensitive to texture and noise. Dollár and Zitnick (Dollar & Zitnick, 2015) achieved an ODS of 0.75 on Berkeley Segmentation Dataset 500 (BSDS500) using structured forests with random forests and structured output regression. The gPb framework (Arbeláez et al., 2011) introduced spectral clustering of brightness, colour, and texture cues, reaching ODS 0.71. A comprehensive review of developments in both classical and deep learning-based edge detection is provided by Jing et al. (Jing et al., 2022).

Deep learning transformed edge detection from the mid-2010s. The Holistically-nested Edge Detection (HED) network (Xie and Tu 2015) replaced the fully connected layers of Visual Geometry Group 16-layer network (VGG-16) (Simonyan & Zisserman, 2014) with side-output branches at each convolutional stage, fusing multi-scale feature maps through deep supervision and achieving ODS 0.788 on BSDS500. Subsequent models improved upon HED: Richer Convolutional Features (RCF) (Liu et al., 2017) introduced richer multi-level feature fusion (ODS 0.806); Bi-Directional Cascade Network (BDCN) (He et al., 2019) proposed bidirectional cascade supervision (ODS 0.828); DexiNed (Soria et al., 2020) combined HED with Xception; PiDiNet (Su et al. 2021) embedded gradient-operator priors into convolutional kernels (ODS 0.807, < 1 M parameters); and EDTER (Pu et al., 2022) leveraged transformer global–local attention to reach ODS 0.848. More recent transformer-based detectors have continued to push performance on BSDS500; EdgeNAT (Jie et al., 2024), which uses a Dilated Neighbourhood Attention encoder, achieves ODS 0.860, outperforming EDTER on standard benchmarks. Attention mechanisms such as Convolutional Block Attention Module (CBAM) (Woo et al., 2018), Efficient Channel Attention (ECA) (Wang et al., 2020), and squeeze-and-excitation (SE) (Hu et al., 2018) have also been incorporated into edge detection backbones with varying success. Attention mechanisms coupled with texture-contextual feature extraction have similarly improved boundary and defect localisation in other industrial surface-inspection tasks (Zhang et al., 2024), underscoring the broader relevance of directional and texture-aware attention design beyond general-purpose benchmarks.

Recent work has also explored task-specific edge detection in textile applications. Gao et al. proposed a real-time cloth edge extraction method based on ENet semantic segmentation, achieving robust contour detection for fabric materials under practical conditions (Gao et al., 2022). Nevertheless, segmentation-based approaches primarily focus on region-level accuracy and do not explicitly optimise for precise boundary localisation, which is critical for downstream geometric tasks such as stitch-path generation.

Despite these advances, existing edge detection methods are primarily optimised for general-purpose benchmarks and do not explicitly address the requirement for precise boundary localisation in texture-rich, low-contrast industrial fabrics. In particular, attention mechanisms integrated into these models often rely on global pooling operations that reduce spatial specificity, limiting their effectiveness for tasks where pixel-level geometric accuracy is critical. This limitation motivates the development of a directionally-aware attention mechanism tailored to garment-panel edge detection.

Robot trajectory planning

Industrial robot trajectory planning in Cartesian space has been studied since the foundational resolved-rate control framework of Whitney (Whitney, 1969). Achieving smooth continuous motion requires the jerk to be bounded; Macfarlane and Croft (Macfarlane & Croft, 2003) demonstrated that quintic-polynomial-based jerk-bounded trajectory generation reduces mechanical vibration and improves path-tracking accuracy in industrial applications. Cubic spline interpolation provides velocity continuity but cannot guarantee acceleration continuity, while Bézier curve methods address smooth line-arc transitions in Cartesian space (Xu et al., 2014). In vision-integrated manufacturing scenarios, trajectory planning is increasingly coupled with real-time perception. Coverage path planning methods have also been extended to surface inspection tasks, where optimised trajectory generation is critical for reliable defect detection (Chen et al., 2025a). For instance, Zhong et al. proposed a visual-feedback-based trajectory planning method in which contour features extracted from images are directly used to generate smooth robot paths (Zhong et al., 2025). More broadly, recent reviews have framed such vision-coupled motion generation as an integrated perception–planning–execution architecture for industrial manipulators, reflecting a growing convergence between visual perception and Cartesian-space trajectory generation in intelligent manufacturing (Liu et al., 2026).

For spatial arc generation, three-waypoint circular arc interpolation has been employed in electronic assembly robots (Douglas & Peucker, 1973).

In the context of robotic sewing, Li et al. proposed a process-model-based approach in which stitch trajectories are decomposed into linear and curved segments, enabling coordinated motion between the manipulator and the sewing process (F. Li et al., 2024a). However, the method relies on predefined seam geometry and does not consider perception-driven trajectory generation from visual input. More recently, Ma et al. (Ma et al., 2021) proposed a time-optimal trajectory planning approach with explicit torque and jerk limits, demonstrating applicability to six-DOF serial manipulators performing constrained Cartesian-space tasks. The present work adapts Cartesian-space linear and arc interpolation to the specific geometry of garment-panel stitch trajectories, incorporating jerk-bounded constraints to ensure consistent stitch spacing under the synchronised motion requirements of robotic sewing. Recent studies in intelligent manufacturing have explored vision-driven trajectory generation frameworks, in which task paths are directly derived from perceptual inputs rather than predefined geometry (Tafuro et al., 2025). Parameter optimisation for trajectory planning has also been approached through data-efficient methods, including LLM-augmented Bayesian optimisation (Xia et al., 2025). Such approaches demonstrate the potential of integrating perception and planning to improve flexibility in complex manufacturing tasks.

Contributions

The primary contribution of this work is the development and experimental validation of a vision-based sewing robot system structured as a human-supervised workflow for the Yun operation. The specific technical contributions are as follows.

(1) A directionally-aware attention mechanism (FS module) for edge detection, which preserves spatial coordinate information through axis-wise feature aggregation. Unlike conventional attention modules, the proposed design enhances boundary localisation accuracy in low-contrast, texture-rich garment images, enabling reliable extraction of stitch-relevant contours.
(2) A systematic stitch generation pipeline that converts edge-detection output into ordered stitch-point coordinates through contour extraction, polygon approximation, interactive corner selection, and enabling operator-guided correction of ambiguous geometric features.
(3) An integrated Cartesian-space trajectory planning framework for the MS6MT six-DOF manipulator, employing linear interpolation for rectilinear seams and Bézier-arc interpolation for curved seams, with coordinate transformation via camera calibration and eye-to-hand calibration.
(4) A three-module software system integrating the above components into an operator-facing application supporting a human-supervised workflow, including a human-in-the-loop correction step for ambiguous corner detection, tested under laboratory conditions and validated to place stitches within $\pm$ 1 mm of the target seam line.

This work is positioned within the paradigm of human-centric intelligent manufacturing, addressing the integration of perception, planning, and operator supervision for deformable material processing tasks.

Overall system architecture

The proposed vision-based sewing robot system comprises two primary subsystems: a vision subsystem responsible for panel perception and stitch-coordinate generation, and a robot subsystem responsible for panel manipulation and sewing execution. The overall system concept is depicted in Fig. 1.

The operational sequence is shown in Fig. 2. A Hikrobot industrial camera mounted above the work surface captures an image of the garment panel. A deep-learning model processes the image on a host computer to yield an edge-probability map, from which contour extraction and polygon approximation algorithms derive the panel outline and sewing stitch coordinates. Camera calibration and eye-to-hand calibration transform these image-plane coordinates to the robot base frame. Kinematic analysis and Cartesian-space trajectory planning then produce a configuration file that is loaded by the robot controller, which drives the MS6MT manipulator to transport the panel through the sewing machine head along the planned stitch trajectory. When contour approximation produces ambiguous corner configurations, the operator can refine the detected corner set through the graphical interface prior to trajectory generation.

Target operations

Field investigations were conducted at a garment factory in Guangdong Province to identify the operations most suitable for robotic automation. A shirt comprises multiple structural components, including the front panel, back panel, sleeves, collar, and cuffs. The Yun operations targeted in this study are (*i*) cuff stitching (Fig. 3a), in which a cuff interlining is aligned and sewn to a shell fabric along three sides; (*ii*) collar stitching (Fig. 3b), in which collar interlining and shell panels

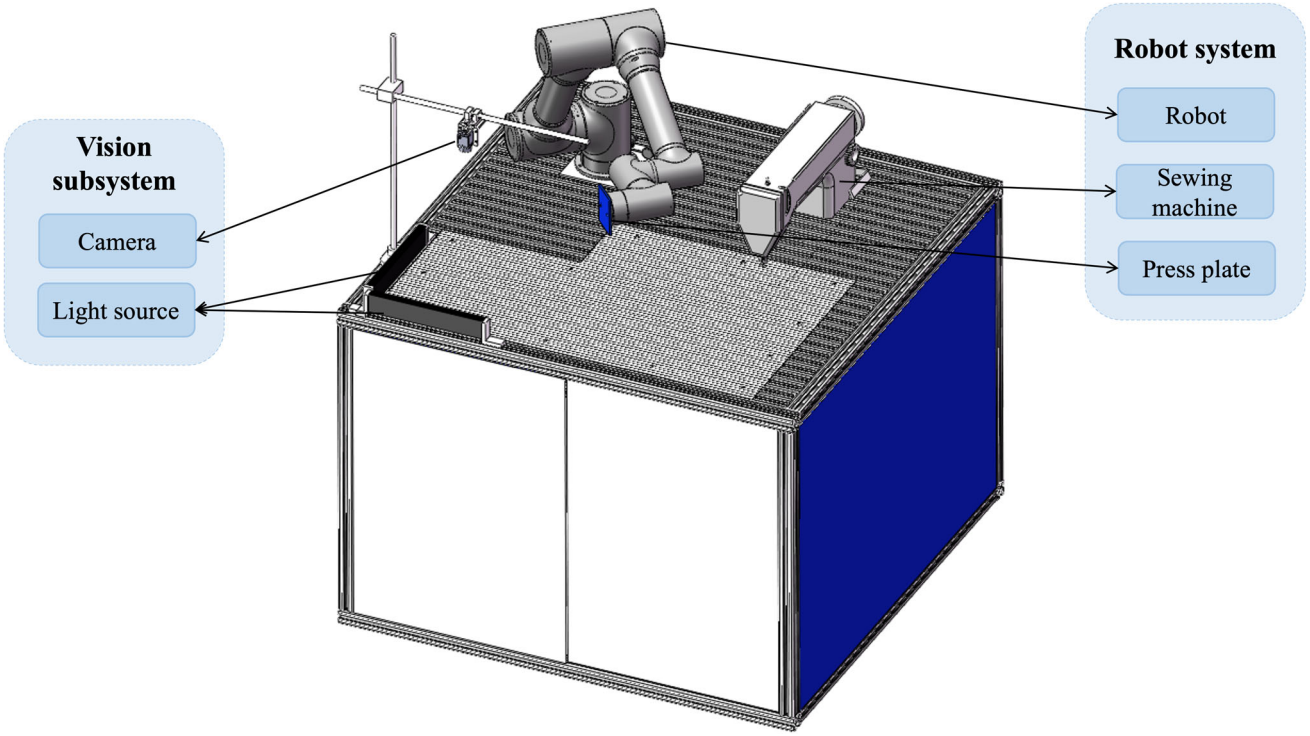


Fig. 1 Overall architecture of the vision-based sewing robot system

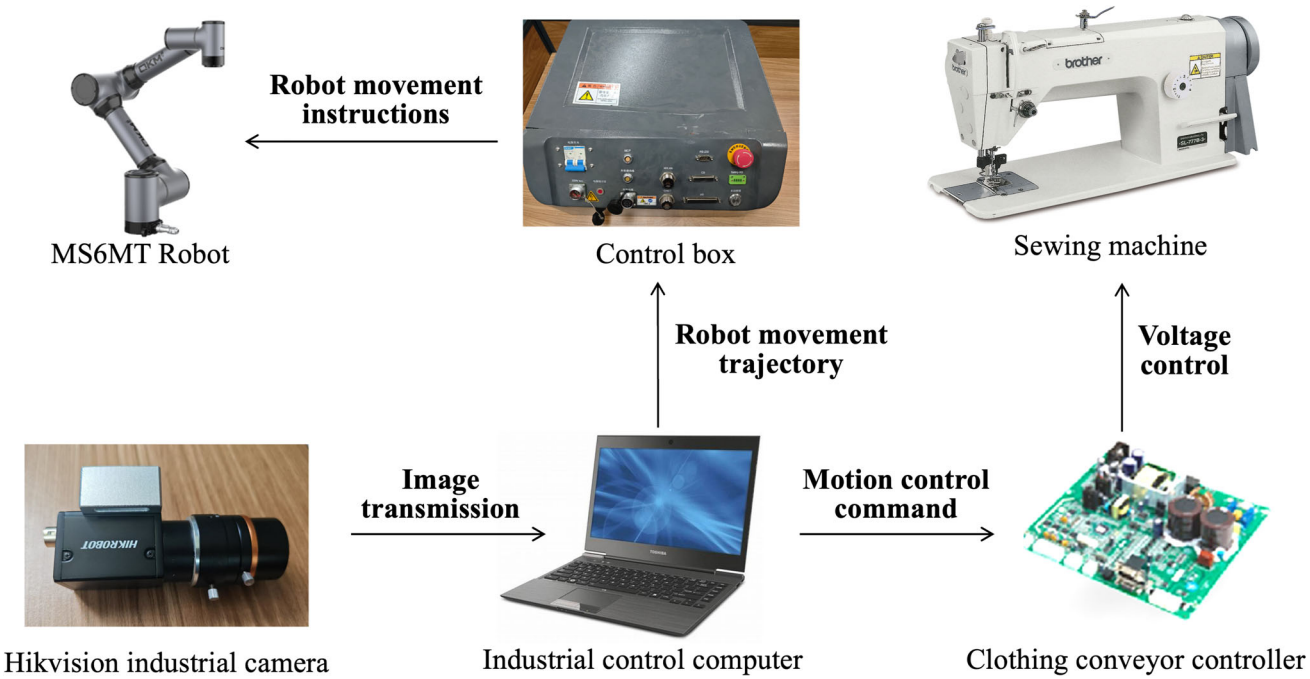


Fig. 2 Functional decomposition of the system from perception to robot execution

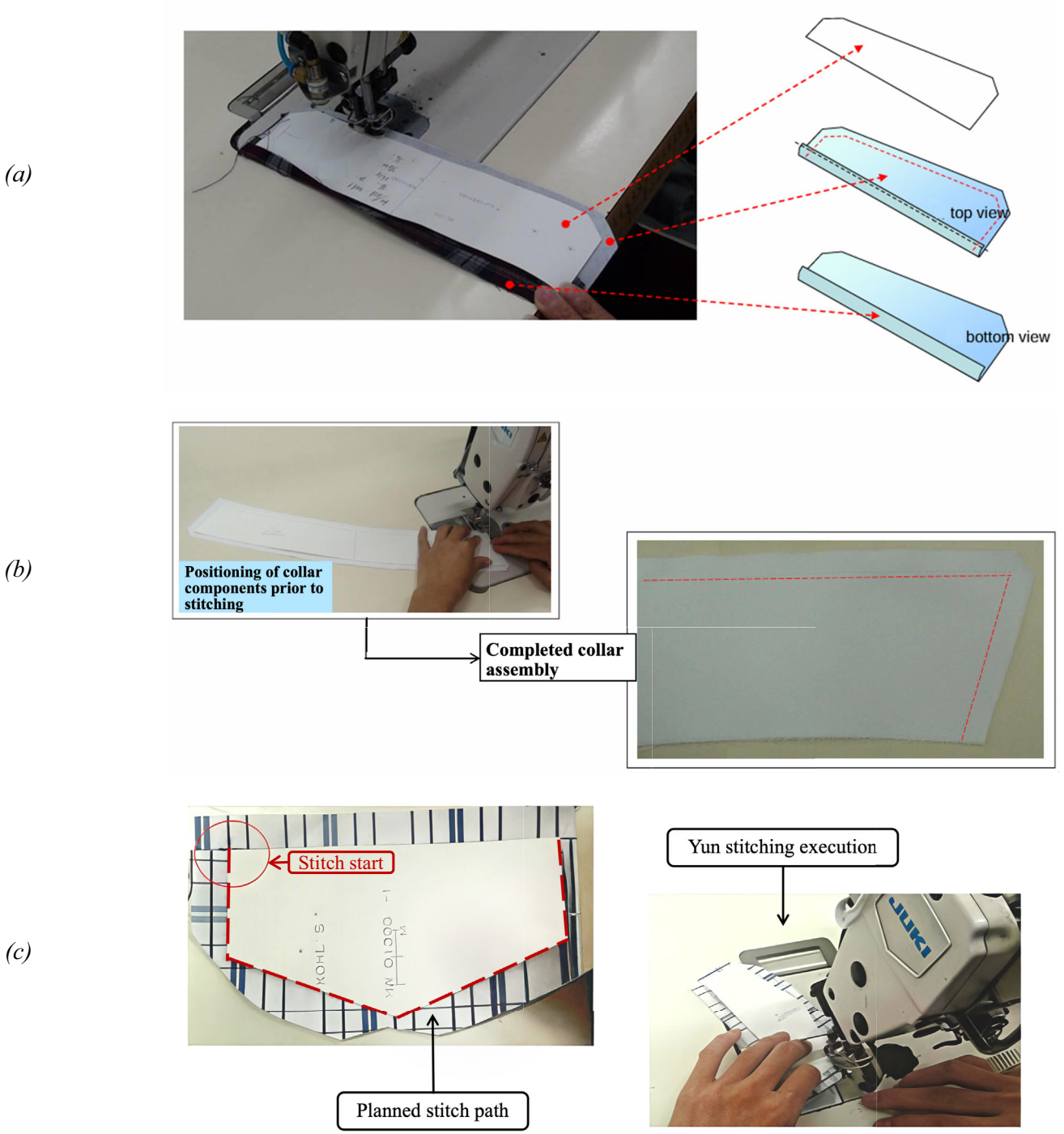


Fig. 3 Representative Yun operations considered in this study: **a** cuff stitching; **b** collar stitching; **c** pocket-flap stitching. Insets illustrate the corresponding panel geometry and target stitch path

are sewn together along the outer perimeter; and (*iii*) pocket-flap stitching (Fig. 3c), in which the flap interlining is sewn to the flap shell along the curved lower edge.

Hardware configuration and experimental platform

Based on task analysis and laboratory resources, the following hardware components were selected. The MS6MT six-degree-of-freedom collaborative arm (Liqun Automation) was selected as the robot platform. The manipulator has a reach of 962 mm, a rated payload of 5 kg, and a repeatability of ± 0.1 mm, which is consistent with the target seam placement accuracy of the proposed system and sufficient to meet the workspace requirements of garment panel handling and stitching operations. The robot provides an open software development kit (SDK) that supports trajectory configuration file input, enabling integration with the vision-based stitch generation pipeline. A two-dimensional industrial camera was selected in preference to a depth camera, as the panel geometry is planar and depth information is not required. The target camera resolution requirement of ± 0.2 mm per pixel demands a minimum pixel resolution of 350/0.2 = 1750 px per axis over the 350 mm × 350 mm field of view. Accordingly, the Hikrobot MV-CE050-30GC camera (2592 × 1944 px, 5 MP resolution, 2.2 μm pixel pitch, 24 fps frame rate) was selected, providing sufficient spatial resolution for accurate boundary localisation and stitch-path generation. Lens focal length was computed as f = WD × H_s/V_f = 380 × 7.1/350 ≈ 7.7 mm, leading to the selection of the LOMOSEN ZX-SF08 8 mm industrial lens (5 MP, f/2.0, 1/1.8-inch format, ± 0.1% TV distortion). The ± 0.2 mm figure specifies the camera resolution requirement, not the end-to-end system accuracy. The ± 1 mm trajectory execution accuracy reported in Sect. "Full workflow experimental testing and results" reflects combined calibration, robot, and mechanical errors, and is consistent with the factory tolerance for the Yun operation. An LED ring light source with adjustable intensity was employed to ensure uniform illumination across fabric textures.

The experimental platform integrates the camera and LED illumination on a fixed overhead frame, the MS6MT arm mounted on an adjacent support, and a crank-slider mechanism that replicates the needle-bar motion of a lockstitch sewing machine. This surrogate sewing head is used for trajectory validation experiments, with a pen tip substituted for the needle to trace stitch trajectories on paper overlaid on the panel. This setup evaluates trajectory generation and execution accuracy only. Fabric deformation, thread tension, and needle–material interaction are outside the scope of the current validation and represent the primary gap between laboratory testing and production deployment. The modular design of the platform allows independent evaluation of perception, calibration, and trajectory execution components.

Camera calibration and hand-eye calibration

Accurate coordinate transformation from image pixels to robot Cartesian space requires both intrinsic camera calibration and hand-eye calibration. The pinhole imaging model relates a world point $P = (X_w, Y_w, Z_w)^T$ to its pixel coordinate $(u, v)^T$ through the composite transformation, as shown in Eq. (1):

$$Z_c[u, v, 1]^T = M_1 M_2[X_w, Y_w, Z_w, 1]^T \quad (1)$$

where Z_c denotes the depth of the point in the camera coordinate frame, M_1 is the 3 × 4 intrinsic matrix parameterised by focal lengths (f_x, f_y) and principal point (u_0, v_0), and M_2 is the 3 × 4 extrinsic matrix comprising rotation R and translation t. The model is extended by a five-parameter distortion model (k_1, k_2, k_3 for radial distortion and p_1, p_2 for tangential distortion). The coordinate transformations are based on the standard pinhole camera model relating the world, camera, image, and pixel coordinate systems. The intrinsic and extrinsic parameters are estimated using Zhang's planar calibration method (Zhang, 2000). A 9 × 6 checkerboard target was captured from 15 to 25 viewpoints using MATLAB Camera Calibration Toolbox, which solves for the intrinsic and distortion parameters by minimising reprojection error via Levenberg–Marquardt optimisation.

The camera is mounted at a fixed position above the worktable (eye-to-hand configuration), which avoids potential collisions with the sewing head and eliminates the need for recalibration during robot motion. The calibration is formulated using the standard hand–eye relationship: $AX = XB$ (Tsai & Lenz, 1989), where $X = {}^{base}H_{cam}$ is the unknown homogeneous transformation from the camera frame to the robot base frame. The matrices A and B are constructed from pairs of robot and camera observations. Specifically, the robot motion between two poses i and j is defined as: $A_{ij} = ({}^{base}H_{tool,i})^{-1}\,{}^{base}H_{tool,j}$ and the corresponding motion of the calibration target observed in the camera frame is: $B_{ij} = {}^{cam}H_{cal,i}\,({}^{cam}H_{cal,j})^{-1}$ In this setup, the calibration board is rigidly attached to the robot flange, so its motion is induced by the robot end-effector. A total of 15–25 image—pose pairs are collected at different robot configurations to ensure sufficient motion diversity. The transformation X is then estimated using OpenCV's calibrateHandEye function. The resulting transformation X, together with the camera intrinsic parameters, enables mapping of detected stitch features from image space to the robot base coordinate system. Quantitative calibration results, including reprojection error and coordinate transformation accuracy, are reported in Sect. "Experimental Validation".

Vision-based perception and stitch generation

This section presents the perception-driven approach for extracting garment boundaries and generating stitch coordinates from visual input. The overall processing workflow is first outlined before detailing its individual components.

Stitch generation pipeline

The complete stitch generation pipeline is illustrated in Fig. 4. A raw camera image passes through the improved HED edge detection model to yield an edge-probability map. Standard OpenCV algorithms then binarise and smooth the map, extract the dominant contour, approximate it as a polygon, and identify the corner vertices that define stitch-path waypoints. To ensure robust stitch-path generation under geometric ambiguity, the pipeline incorporates a human-supervised refinement step in which the detected corner set can be inspected and interactively adjusted prior to trajectory generation. This design enables selective operator intervention only when required, maintaining automation efficiency while preserving reliability in challenging cases. Straight lines between adjacent corners are offset inward by the target seam allowance (typically 5 mm from the panel edge, or along the interlining boundary when present) using a parallel-offset operation. Arc segments between corners are represented with Bézier curves (Mineur et al., 1998). The resulting stitch coordinate set is exported as a text file for the robot controller. The full pipeline completes in approximately 10 s. A breakdown of the processing time by stage is provided in Table 5.

The advantage of deep learning over traditional operators for this task is illustrated through a side-by-side comparison on representative panels (Fig. 5). The Canny operator, which relies on a fixed threshold, produces fragmented or incomplete boundaries when applied to textured or low-contrast fabrics. In contrast, the HED-based approach yields continuous and well-localised contours that can be directly processed by downstream stitch-path generation algorithms.

Edge detection model

Holistically-nested Edge Detection (Xie and Tu 2015) adapts VGG-16 by (*i*) discarding the fully connected layers, (*ii*) attaching side-output branches with deep supervision to the final convolutional layer of each VGG stage, and (*iii*) fusing the resulting multi-scale side outputs by a learned linear combination. During training, a class-balanced cross-entropy loss is applied at each side output and at the fused output. The multi-scale architecture allows the network to capture both fine boundary detail from shallow layers and semantic context from deep layers, producing full-resolution edge probability maps.

Baseline HED architecture

Analysis of HED performance on garment-panel images reveals two primary limitations: (*i*) imprecise corner localisation, which leads to inaccurate stitch-path waypoints, and (*ii*) sensitivity to fabric texture gradients with similar magnitude to true boundary gradients. These limitations arise from insufficient positional discrimination in the learned feature representations, particularly in regions where edge contrast is weak or locally ambiguous. Existing attention mechanisms, such as SE, CBAM, and efficient channel attention ECA, improve feature weighting but rely on global pooling operations that partially discard spatial coordinate information. While effective for high-level recognition tasks, this loss of positional specificity is detrimental for edge detection problems requiring pixel-level localisation accuracy. This limitation is especially critical in garment-panel images, where boundaries must be precisely identified for downstream geometric processing and stitch-path generation.

To address this limitation, we propose a directionally aware attention mechanism that explicitly preserves spatial coordinate information during feature aggregation. The FS module shares the axis-wise pooling principle with Coordinate Attention (Hou et al., 2021) and axial attention (Wang et al., 2020), but differs in two respects. First, those designs were developed for object recognition and segmentation tasks, where global spatial context is the primary objective. The FS module is instead optimised for pixel-level boundary localisation in texture-rich garment images, where the critical requirement is to distinguish structured edge gradients from fabric texture noise along specific orientations. Second, strip pooling (Hou et al., 2020) aggregates long-range context along strips but does not decompose the resulting attention map into directional components applied separately to the feature map. In contrast, the proposed FS module is designed specifically for HED side-output features in edge detection. It decomposes boundary responses into horizontal and vertical directional components and applies them to multi-scale side-output branches, thereby enhancing structured garment boundaries without introducing heavy transformer-style computation.

Inspired by the directional gradient decomposition of the Sobel operator, the proposed FS attention module enhances the network's sensitivity to structured edge directions. By maintaining axis-wise spatial dependencies, the module improves discrimination between true panel boundaries and texture-induced gradients, enabling more accurate and stable edge localisation for stitch trajectory generation.

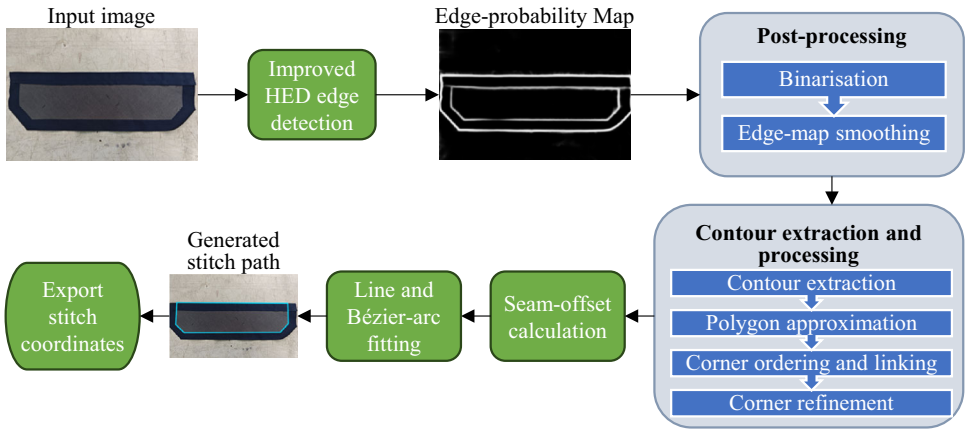


Fig. 4 Vision-based stitch generation pipeline from raw camera image to stitch-coordinate export

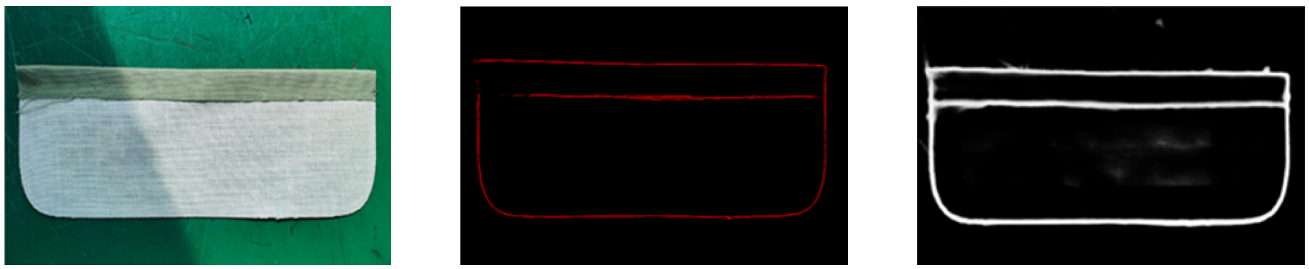


Fig. 5 Comparison of edge detection methods on representative garment panels: **a** input image; **b** Canny edge detection showing fragmented boundaries; **c** HED-based edge detection producing continuous contours suitable for stitch-path computation

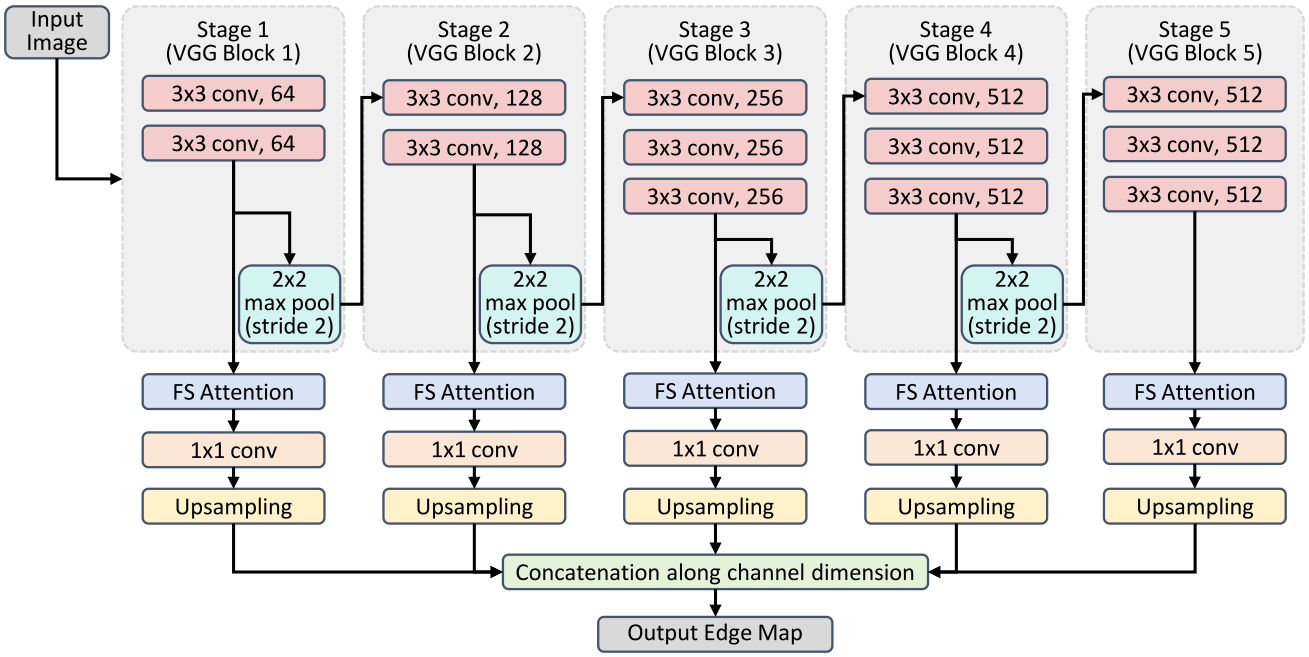


Fig. 6 HED architecture with FS attention modules applied to the five side-output branches

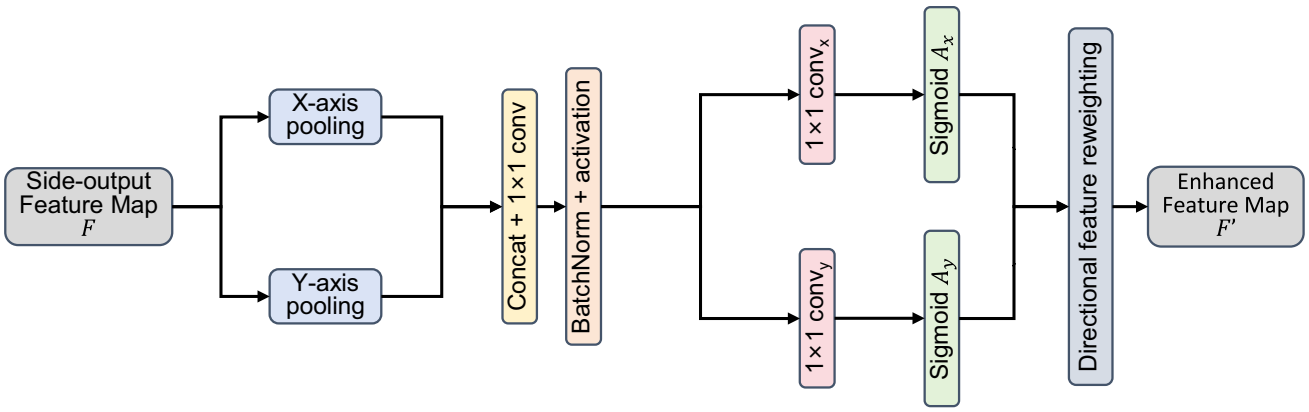


Fig. 7 Structure of the Frequency-Spatial attention module

FS attention module

Let $F \in \mathbb{R}^{C\times H\times W}$ denote a side-output feature map. The FS module applies directional global average pooling along the horizontal axis to produce a vertical feature descriptor $f_v \in \mathbb{R}^{C\times H\times 1}$, and along the vertical axis to produce a horizontal feature descriptor $f_h \in \mathbb{R}^{C\times 1\times W}$. Unlike conventional attention modules that compress spatial information through two-dimensional global pooling, the proposed FS module performs axis-wise feature aggregation, preserving directional spatial dependencies along horizontal and vertical axes. This design maintains explicit positional cues that are critical for accurate boundary localisation, particularly at corners and curved regions, which are essential for downstream stitch-path generation. The two descriptors are concatenated and passed through a shared 1×1 convolution, batch normalisation, and nonlinear activation to obtain a shared directional representation. This representation is projected by two separate 1×1 convolution branches followed by sigmoid activations to generate A_x and A_y, corresponding to the horizontal and vertical axes. These directional attention components are applied to the input feature map F through element-wise multiplication. The resulting feature maps are combined and normalised to produce the enhanced feature representation. This directional attention mechanism preserves spatial coordinate information that is often lost in pooling-based attention modules such as CBAM (Woo et al., 2018) and SE (Hu et al., 2018). The FS module is inserted after each of the five HED side-output layers. Figure 6 illustrates the overall architecture of the improved HED model, and Fig. 7 provides a detailed view of the FS module. This directional attention mechanism enables the network to better distinguish structured garment boundaries from texture-induced gradients, which often exhibit less consistent directional patterns. In summary, the FS module improves elongated textile boundary extraction by preserving directional contour information and enhancing multi-scale edge features, thereby improving garment edge localisation accuracy and contour continuity.

Dataset construction and training

No publicly available dataset provides garment-panel boundary annotations under controlled industrial lighting conditions. To address this gap, a purpose-built dataset was constructed from two sources: 400 images captured using the experimental vision platform in a factory environment, and 932 images obtained via infrared backlighting, resulting in a total of 1332 images. The dataset comprised 757 cuff images, 384 collar images, and 191 pocket-flap images, with each category represented in the training, validation, and test subsets. The dataset captures variability in fabric texture, colour, shape, and lighting conditions. The backlit images highlight garment-panel contours and folded-edge structures, whereas the front-lit images preserve texture, shadow, and fabric appearance variations under real workstation conditions. The combination of these two illumination settings helps improve the robustness of the model for garment boundary detection under different lighting conditions. In addition, the dataset includes multiple garment panel types, ensuring coverage of both straight and curved boundary geometries relevant to sewing trajectory generation.

Representative samples from both subsets are shown in Fig. 8, illustrating the variability in fabric appearance and imaging conditions. All images were annotated using Labelme by tracing the garment-panel boundaries. The boundaries were independently labelled by four annotators following the BSDS500 multi-annotator protocol to reduce individual labelling bias. This multi-annotator strategy improves the consistency and reliability of the ground-truth edge maps used for training and evaluation. The dataset was partitioned into 80% training, 10% validation, and 10% test sets. Experiments were conducted on a system equipped with an NVIDIA GeForce RTX 4090 D GPU. The

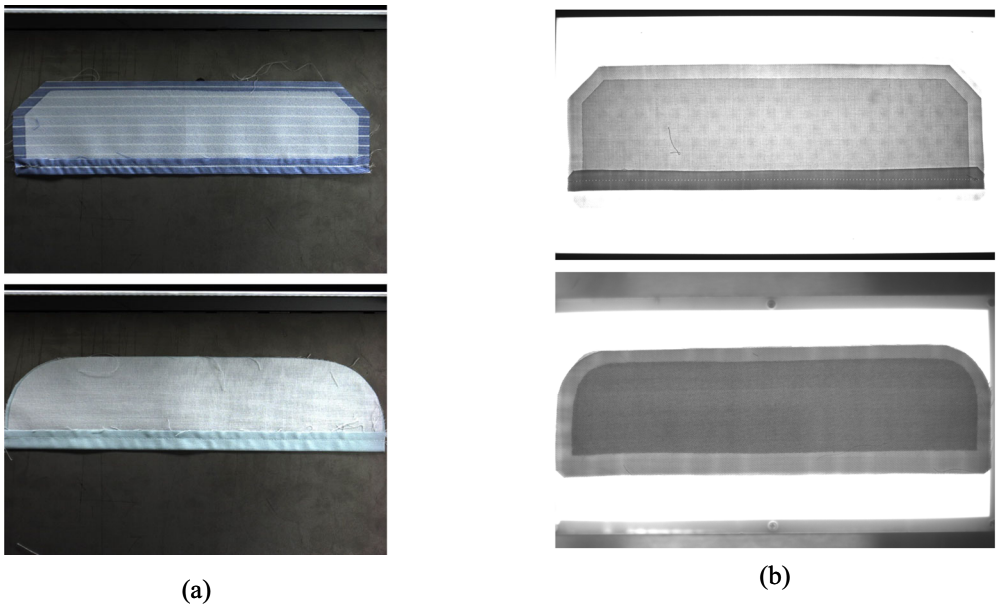


Fig. 8 Representative samples from the garment-panel dataset: **a** front-lit garment panel images; **b** backlit garment panel images

improved HED model was initialised from VGG-16 weights pre-trained on ImageNet and fine-tuned using the Adam optimiser with an initial learning rate of 1×10^{-6} and a weight decay of 2×10^{-4} for 40 epochs with a batch size of 1. These hyperparameters were selected based on standard practice for fine-tuning HED-based models. The learning rate was determined by a grid search over $\{1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$, and 40 epochs were found sufficient for convergence on the training set, with the model achieving the highest validation ODS F1 score selected as the final model for testing. During training, the edge annotations were dilated with a radius of 1 pixel to alleviate the severe class imbalance between positive and negative samples caused by thin edge annotations. During evaluation, a matching tolerance of 1 pixel was adopted to reduce evaluation errors caused by single-pixel annotation shifts. A class-balanced cross-entropy loss was applied at each side output and at the fused output. This dataset design is particularly suited to evaluating edge detection models in conditions representative of industrial garment manufacturing.

Evaluation metrics and results

Edge detection performance is quantified using the standard precision-recall framework employed by BSDS500. Precision (P) and Recall (R) are computed at varying thresholds, and the F-measure $F = 2PR/(P + R)$ is reported at two operating points: the Optimal Dataset Scale threshold (ODS, one threshold for all images) and the Optimal Image Scale threshold (OIS, per-image best threshold). To evaluate the computational complexity and deployment efficiency of different models, the parameter count (Params), floating-point operations (FLOPs), and inference latency per image (Time) were introduced as additional evaluation metrics. This section presents the FS module ablation study, comparative evaluations against HED variants and representative edge-detection models, PR-curve analysis, cross-domain experiments under different illumination conditions, and robustness analysis on challenging garment-panel samples.

To verify the effectiveness of each component of the FS module, an ablation study was conducted within the HED framework, as shown in Table 1. The original HED model achieved an ODS of 0.8115 and an OIS of 0.8421. After introducing 2D global pooling attention, the ODS and OIS decreased by 0.48% and 0.56%, respectively, indicating that excessive compression of spatial information weakens pixel-level boundary localisation. In contrast, introducing directional pooling improved the ODS and OIS by 2.79% and 2.92%, respectively, suggesting that preserving axis-wise spatial dependencies along the horizontal and vertical directions helps distinguish true garment-panel boundaries from fabric texture interference. When the FS module was applied only to the final side-output branch, the performance was further improved. Finally, when FS modules were inserted into all HED side-output branches, the model achieved the highest ODS of 0.8517 and OIS of 0.8842. These results demonstrate that directional enhancement of multi-scale side-output features is more effective than applying attention only at the final fusion stage. Meanwhile, the complete FS structure increases the number of parameters and FLOPs by only 0.029 M and 0.024 G over the baseline HED framework, respectively, indicating that the proposed module significantly improves edge

Table 1 Ablation study of the FS module components

Variant	ODS ↑	OIS ↑	Params(M) ↓	FLOPs(G) ↓
HED	0.8115	0.8421	**14.716**	**76.018**
HED + 2D global pooling attention	0.8067	0.8365	14.795	76.049
HED + directional pooling	0.8346	0.8657	14.736	76.022
HED + FS on final side-output only	0.8429	0.8741	14.724	76.019
HED + FS full	**0.8517**	**0.8842**	14.745	76.042

Bold values indicate the best performance

detection performance with negligible computational overhead.

Table 2 aims at evaluating the effectiveness of the proposed FS module in comparison with other attention mechanisms and recent advanced edge detection methods. The proposed HED + FS model achieves an ODS of 0.8517 and an OIS of 0.8842, outperforming the baseline HED model by 4.02 and 4.21 percentage points, respectively. Among the attention-based variants, CBAM shows inferior performance, while ECA and SE provide only moderate improvements. Strip Pooling achieves better performance through strip-wise context aggregation; however, its ODS and OIS are still lower than those of the proposed FS module. This indicates that the directional decomposition strategy adopted by FS provides a larger performance gain with lower parameter overhead.

Table 2 also compares the proposed method with other representative edge detection models. PiDiNet, a representative lightweight edge detection network, has the fewest parameters and FLOPs and achieves the fastest inference speed, but its ODS and OIS are clearly lower than those of the HED-based models. DexiNed, which has stronger contour representation capability, achieves moderate detection accuracy but requires more parameters and computational cost. EDTER obtains strong edge detection performance, with ODS and OIS only 0.28 and 0.39 percentage points lower than those of HED + FS, respectively. Based on three repeated runs, HED + FS exhibits lower standard deviations, and its improvement over EDTER is statistically significant under a paired t-test ($p < 0.05$). This emphasis on statistically validated comparison, rather than point-estimate metrics alone, echoes recent calls for reproducible, statistically rigorous benchmarking practices in industrial vision model evaluation (Lema et al., 2026). Moreover, EDTER has substantially higher computational complexity and inference latency. In contrast, the proposed HED + FS model achieves high detection accuracy while introducing only a marginal increase in parameters and FLOPs compared with the original HED model, and its inference time remains at the millisecond level. These results demonstrate that the proposed FS module preserves the multi-scale edge detection capability of HED while further enhancing direction-sensitive boundary localisation, achieving a better balance between accuracy and efficiency for deployment-oriented garment-panel edge detection tasks.

Figure 9 shows the precision–recall (PR) curves of the compared models on the garment-panel boundary dataset. The proposed HED + FS method maintains the highest precision across most recall ranges, and its overall PR curve consistently outperforms those of other HED variants and representative edge-detection models. These results further validate the effectiveness of the proposed FS module for garment-panel boundary localisation.

To evaluate the generalisation ability of the proposed HED + FS model under different industrial imaging conditions, two illumination subdomains, namely front illumination and back illumination, were constructed for cross-domain evaluation, as shown in Table 3. In this experiment, the illumination condition was regarded as the main domain variable: the model was trained on the source domain and evaluated on the test set of the target domain. The domain gap is defined as: $Domaingap(X \rightarrow Y) = ODS(Y \rightarrow Y) - ODS(X \rightarrow Y)$. The results show that the model achieves higher performance in the in-domain settings, with the best performance obtained when both training and testing are conducted on the back-illumination domain. In contrast, the performance decreases noticeably in the cross-domain settings, indicating the existence of domain shift between different illumination conditions. Nevertheless, HED + FS still maintains relatively stable contour detection performance under both cross-domain settings, demonstrating a certain degree of generalisation ability. These results further suggest that mixed training with both front-illumination and back-illumination datasets can improve the robustness and practical applicability of the model under complex illumination conditions.

To analyse the robustness of the model, quantitative error analysis was conducted under three typical challenging scenarios, including flying threads, jagged edges, and blurred boundaries. The ODS and OIS scores for each category are reported in Table 4. Compared with HED, the proposed HED + FS achieves higher ODS and OIS scores on all three challenging subsets. The improvement is particularly evident in the flying-thread and blurred-boundary scenarios,

Table 2 Comparative evaluation of HED variants and representative edge detectors

Method	ODS ↑	OIS ↑	Params(M) ↓	FLOPs(G) ↓	Time(ms/image) ↓
HED [21]	0.8115	0.8421	14.716	76.018	4.701
HED + CBAM [28]	0.7969	0.8303	14.751	76.035	6.518
HED + ECA [29]	0.8219	0.8527	14.716	76.028	4.727
HED + SE [30]	0.8446	0.8723	14.751	76.028	4.746
HED + Strip Pooling	0.8452	0.8756	14.758	76.275	5.021
PiDiNet	0.7771	0.8044	**0.710**	**18.596**	**3.588**
DexiNed	0.8364	0.8645	35.215	140.444	9.372
EDTER	0.8489 ± 0.0014	0.8803 ± 0.0018	71.310	332.035	145.487
HED + FS (proposed)	**0.8517 ± 0.0012**	**0.8842 ± 0.0015**	14.745	76.042	5.223

Bold values indicate the best performance

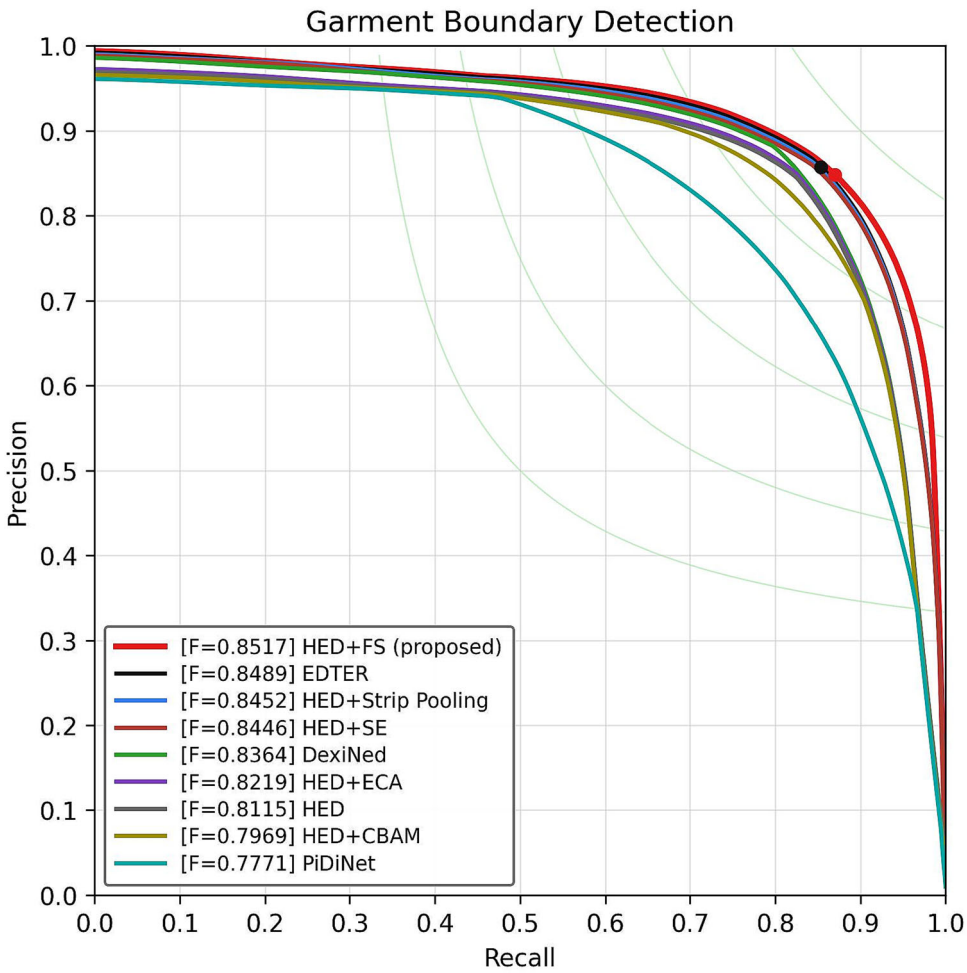


Fig. 9 Precision–recall curves on the garment-panel test set comparing HED variants, representative edge-detection models, and the proposed HED + FS model

Table 3 Cross-domain evaluation under illumination-domain shift

Training domain	Testing domain	Setting	ODS ↑	OIS ↑	Domain gap ↓
Front illumination	Front illumination	In-domain	0.829	0.862	–
Front illumination	Back illumination	Cross-domain	0.671	0.708	**0.188**
Back illumination	Back illumination	In-domain	**0.859**	**0.894**	–
Back illumination	Front illumination	Cross-domain	0.632	0.674	0.197

Bold values indicate the best performance

Table 4 Robustness evaluation under challenging garment-panel conditions

Category	HED ODS	HED OIS	HED + FS ODS	HED + FS OIS
Flying threads	0.6688	0.6798	0.7429	0.7529
Jagged edges	0.7709	0.7812	0.7817	0.8113
Blurred Boundaries	0.6338	0.6442	0.7011	0.7066

indicating that the FS module helps preserve contour continuity and reduce the influence of thread-like interference and weak edges. These results demonstrate that HED + FS has better robustness under practical manufacturing variations.

Qualitative results on representative garment panels are shown in Fig. 10. The improved model accurately delineates both straight and curved boundaries while suppressing texture-induced false positives, demonstrating improved localisation compared to the baseline HED.

Stitch coordinate generation

Following edge detection, the predicted edge-probability map is binarised using Sauvola adaptive thresholding (Sauvola & Pietikäinen, 2000) and smoothed by bilateral filtering (Tomasi & Manduchi, 1998). OpenCV's findContours is applied with RETR_EXTERNAL to extract the dominant outer contour, which is approximated by a polygon using the Douglas-Peucker algorithm (Douglas & Peucker, 2011). The epsilon parameter of the Douglas-Peucker algorithm was set to 1.5% of the contour arc length, a value determined empirically to retain corner vertices while suppressing minor boundary irregularities. An interactive graphical user interface (GUI) supports operator supervision of the stitch generation process by allowing the operator to select and reorder corner vertices, correct spurious corners introduced by fabric fraying, and specify the seam-allowance offset. This human-supervised step is designed to ensure reliable stitch-path generation in cases where purely automatic corner detection is uncertain. Straight edge segments are offset using the standard parallel-offset formula; curved segments are fitted by quadratic Bézier curves (Mineur et al., 1998). The resulting stitch trajectory is exported as an ordered list of (x, y) pixel coordinates, converted to robot Cartesian coordinates via the calibration matrices, and written to a text file for the robot controller.

The improvements in ODS and OIS achieved by HED + FS translate directly to more complete and accurately localised contours, which in turn reduce polygon approximation errors and the frequency of ambiguous corner configurations requiring operator intervention. The reduced intervention rate observed in Sect. "Full workflow experimental testing and results" (approximately 15% of panels) thus serves as a system-level indicator of the downstream benefit of improved edge detection quality.

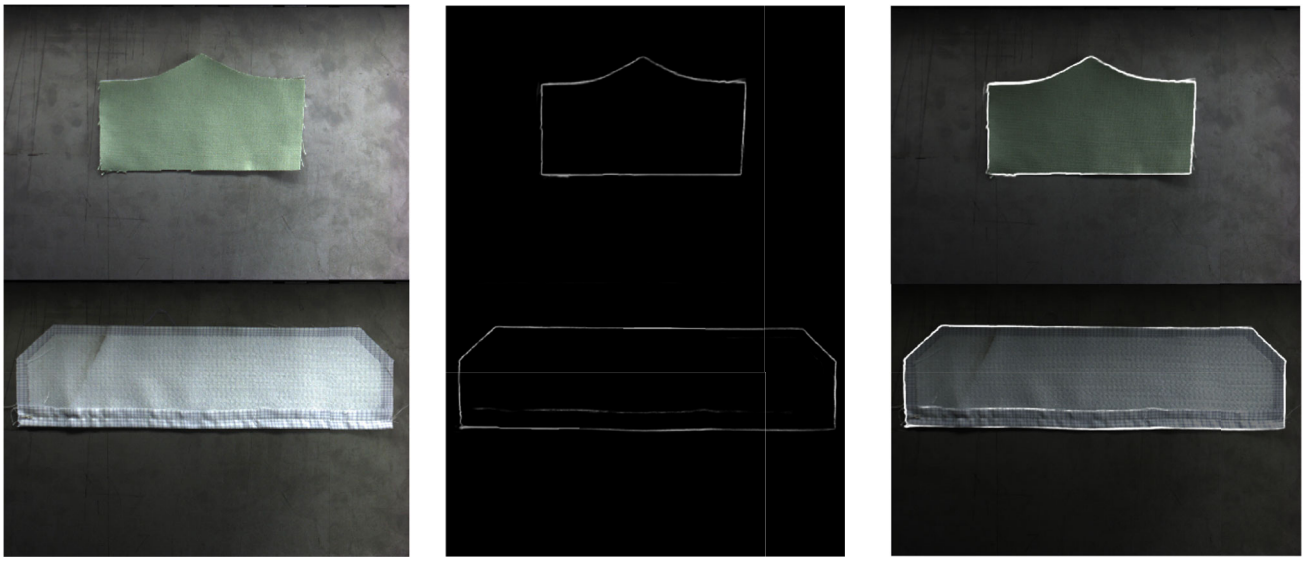
Representative stitch generation results are shown in Fig. 11 (contour outlines and the final stitch coordinates shown in blue). As shown in Table 5, the total processing time from raw image to exported stitch file is approximately 10 s: 5 s for the edge map generation module, including image preprocessing, HED-based edge prediction, and edge-map post-processing, 2 s for contour extraction and corner detection, 2 s for edge fitting, and 1 s for offset computation and export.

Robot kinematics and trajectory planning

This section addresses the transformation of generated stitch coordinates into executable robot motion through kinematic modelling and trajectory planning. The task requirements and corresponding planning strategy are introduced as the foundation for subsequent formulation.

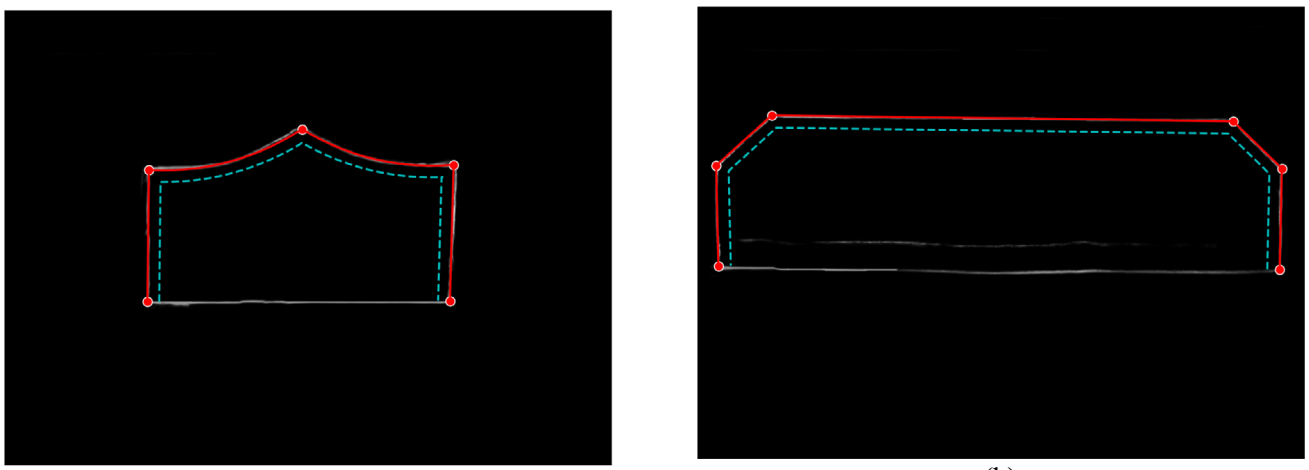
Task analysis and planning strategy

The robot is required to grasp the garment panel and transport it through the sewing head while following a predefined stitch trajectory derived from vision-based perception. This task



(a) *(b)* *(c)*

Fig. 10 Visualisation of improved HED detection on diverse panel types. **a** input image; **b** predicted edge map; **c** edge map overlaid on input



(a) (b)

Fig. 11 Computed sewing stitch outlines (blue lines) for representative garment panels: **a** cuff panel; **b** pocket-flap panel

Table 5 Processing time breakdown for stitch coordinate generation

Processing stage	Time (s)
Edge-map generation module	5
Contour extraction and corner detection	2
Edge fitting (line + Bézier arc)	2
Offset computation and coordinate export	1
Total	10

can be formulated as a Cartesian trajectory tracking problem with geometric constraints imposed by the panel boundary. Two trajectory planning strategies are adopted depending on the seam geometry. For panels composed of piecewise linear segments, the trajectory is defined by a sequence of corner waypoints, and the robot motion is generated using linear interpolation in Cartesian space. For panels containing both linear and curved segments, a hybrid representation is employed, in which straight segments are interpolated linearly, while curved segments are approximated using arc-based waypoint sets obtained from Bézier curve sampling. This formulation enables consistent trajectory generation across different panel geometries while maintaining compatibility with the kinematic and control constraints of the manipulator. Accurate execution of the stitch trajectory requires tight coordination between perception and motion planning, particularly in the presence of geometric variability

and boundary uncertainty. The trajectory planning contribution lies in the application-specific coupling of vision-derived stitch coordinates with jerk-bounded Cartesian interpolation, rather than in novel planning algorithms. The design choices are driven by the geometric constraints of the Yun operation: seam-allowance offsets, corner handling, and consistent stitch spacing.

Kinematic modelling

The Denavit–Hartenberg (D–H) convention [43] is used to model the six-joint MS6MT manipulator. The D–H parameters of the MS6MT manipulator, derived from the manufacturer's geometric data, are summarised in Table 6.

These parameters are used for forward and inverse kinematic computations required for trajectory execution, but no modification to the standard kinematic model is introduced. The overall homogeneous transformation from the robot base frame to the end-effector frame is given by Eq. (2):

$$T = \left[{}_1{}^0T_2 {}^1T_3 {}^2T_4 {}^3T_5 {}^4T_6 {}^5T \right] \tag{2}$$

where each transformation matrix ${}_{i+1}{}^iT$ is constructed according to the standard D–H formulation. The forward kinematics model was verified numerically by substituting known joint angle configurations and comparing the resulting end-effector pose with the manipulator's internal forward kinematics solver, yielding consistent results.

The inverse kinematic solution was derived analytically using the Pieper criterion (Donald L. Peiper, 1968), which is satisfied because the last three joint axes of the MS6MT intersect at a common wrist point. The solution decomposes into a position sub-problem for joints 1–3 and an orientation sub-problem for joints 4–6. Multiple solutions exist; the appropriate one is selected by checking joint-limit feasibility and minimising the total joint displacement from the current configuration. The computed joint angles are subsequently used by the robot controller to execute the Cartesian-space stitch trajectories defined in Sect. "Cartesian-space trajectory planning". Verification was performed by back-substituting the computed joint angles into the forward kinematics model described in Sect. "Kinematic modelling", confirming that the resulting end-effector pose matches the target to within 0.01 mm and 0.001 rad.

Cartesian-space trajectory planning

Between two consecutive waypoints $P_i = (x_i, y_i, z_i)$ and P_{i+1}, the end-effector position is parameterised by a scalar $s \in [0,1]$, as given in Eq. (3):

$$P(s) = P_i + s \cdot (P_{i+1} - P_i) \tag{3}$$

Orientation is interpolated using spherical linear interpolation (SLERP) applied to the quaternion representations of the start and end poses. The velocity profile follows a trapezoidal scheme with configurable acceleration and deceleration phases to limit jerk and ensure smooth motion. Velocity continuity across segment transitions is maintained by matching exit and entry velocities, ensuring C1 continuity in the Cartesian velocity profile. End-effector feed speed is bounded by the operational constraints of the sewing machine head. In the current setup, a fixed end-effector feed speed is set within the operational range of the surrogate mechanism. Full dynamic synchronisation with a production sewing machine needle and feed dog motion is not implemented and represents a critical requirement for real deployment. A full quantitative dynamic analysis of jerk, acceleration, and tracking smoothness is outside the scope of the current validation and is identified as a direction for future work.

For curved segments, a three-point arc formulation (Xu et al., 2014) is used to determine the arc centre C, radius r, and angular span $\Delta\phi$ from three points (start, mid, end). The arc trajectory is parameterised by the angle ϕ, as expressed in Eq. (4):

$$P(\phi) = C + r \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{bmatrix} \tag{4}$$

for in-plane arcs defined in the robot sewing plane. For out-of-plane arcs, the centre C and the arc plane are expressed in the robot base frame. A Bézier curve representation is used to smooth transitions between adjacent linear and arc segments (Xu et al., 2014).

Experimental validation

The experimental validation focuses on evaluating the accuracy of coordinate transformation and trajectory execution under controlled laboratory conditions. The objective is to assess trajectory generation and coordinate transformation accuracy under controlled conditions. The reported ± 1 mm accuracy reflects trajectory execution performance and does not account for fabric deformation, thread tension, or needle–material interaction.

The checkerboard calibration procedure was performed using the MATLAB Camera Calibration Toolbox on 20 captured images. The mean reprojection error was 0.34 pixels, indicating that the calibration accuracy is sufficient to meet the target seam placement requirement of ± 0.2 mm.

Fifteen robot pose–camera image pairs were collected with the checkerboard rigidly attached to the robot flange. The eye-to-hand transformation ${}^{base}H_{cam}$ was computed

Table 6 D–H parameter table for the MS6MT six-DOF manipulator

Joint i	a_i (mm)	d_i (mm)	α_i (rad)	θ_i
1	0	159.5	π/2	θ_1
2	450	0	0	θ_2
3	397.5	0	0	θ_3
4	0	138.5	− π/2	θ_4
5	0	114.5	π/2	θ_5
6	0	103.5	0	θ_6

using the Tsai–Lenz algorithm (Tsai & Lenz, 1989), as implemented in OpenCV. The resulting transformation matrix is given in Eq. (5):

$$ {}^{base}H_{cam} = \begin{bmatrix} -0.917 & -0.395 & 0.046 & 17.776 \\ 0.391 & -0.916 & -0.078 & -11.574 \\ 0.074 & -0.053 & 0.995 & 111.254 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5) $$

Coordinate transformation accuracy was verified by mapping image points to the robot base frame and comparing the resulting end-effector positions with ground-truth measurements. The positioning error was within ± 0.5 mm across the evaluated panel locations.

A representative pocket-flap panel was selected as the primary validation specimen, as it combines both straight and curved seam segments. Additional trials were conducted on cuff and collar panels to verify the generality of the approach across different seam geometries. Stitch coordinates were generated by the vision pipeline, transformed to robot Cartesian space, and loaded into the MS6MT controller via the configuration text file.

Figure 12 shows the robot executing the planned trajectory at four representative waypoints. The pen-tip surrogate traces the stitch trajectory on a paper overlay, enabling direct measurement of seam-placement accuracy. Trajectory accuracy was evaluated by measuring the point-wise deviation between each planned stitch waypoint and the corresponding pen-traced position on the paper overlay, across multiple panels and repeated trials. In the evaluated demonstrations, the observed trajectory deviations remained within ± 1 mm, with most measurements falling between approximately 0.5 and 0.7 mm. No significant outliers or systematic drift were observed along either linear or curved segments.

Integrated system implementation

This section describes the integration of the perception and motion modules into a unified system for practical deployment. The overall software architecture is first introduced to illustrate the organisation of system components.

System implementation flow

The three hardware-level modules described in Sects. "Vision-based perception and stitch generation" and "Robot kinematics and trajectory planning" are integrated into a unified operator-facing application that supports a human-supervised sewing workflow developed in Python 3.9 using the PyQt5 GUI framework and OpenCV. The software architecture is illustrated in Fig. 13 and comprises three principal modules:

(1) Stitch generation module: manages camera acquisition, model inference, contour extraction, operator-guided corner editing, offset computation, and coordinate export.
(2) Calibration module: handles checkerboard image import, camera calibration parameter loading (.mat file), and hand-eye calibration execution.
(3) Robot motion module: manages TCP/IP connection to the MS6MT controller, uploads trajectory configuration files, and provides real-time joint-state monitoring and emergency stop.

Stitch generation module interface

The stitch generation module GUI (Fig. 14) provides controls for camera activation, model selection, contour generation, and export. When the edge-probability map is generated, the system automatically proceeds to contour extraction and enters the interactive corner-selection sub-window, in which the operator can adjust the corner count, seam-allowance distance, and individual corner positions. This interactive refinement is a key component of the proposed human-supervised workflow, allowing the operator to resolve ambiguous geometric features that would otherwise degrade stitch-path reliability. This interactive capability is not intended as a fallback mechanism, but as an integral component of the system design. In the context of deformable object manipulation, where perception uncertainty cannot be fully eliminated, selective human intervention provides a practical means to

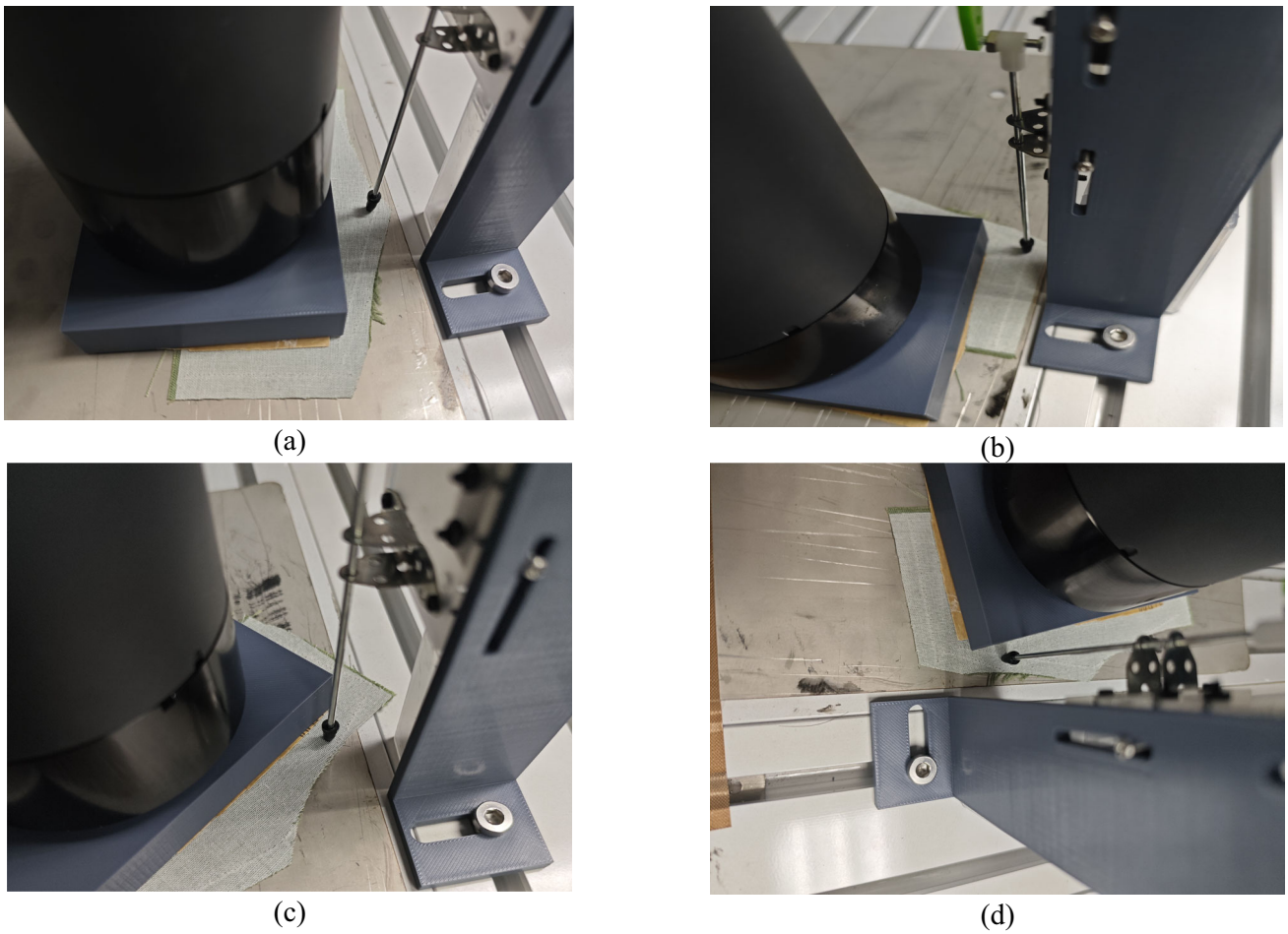


Fig. 12 Robot executing stitch trajectory; **a** waypoint 1 (corner position); **b** waypoint 2 (corner position); **c** mid-arc between waypoints 2 and 3; **d** waypoint 3 (corner position)

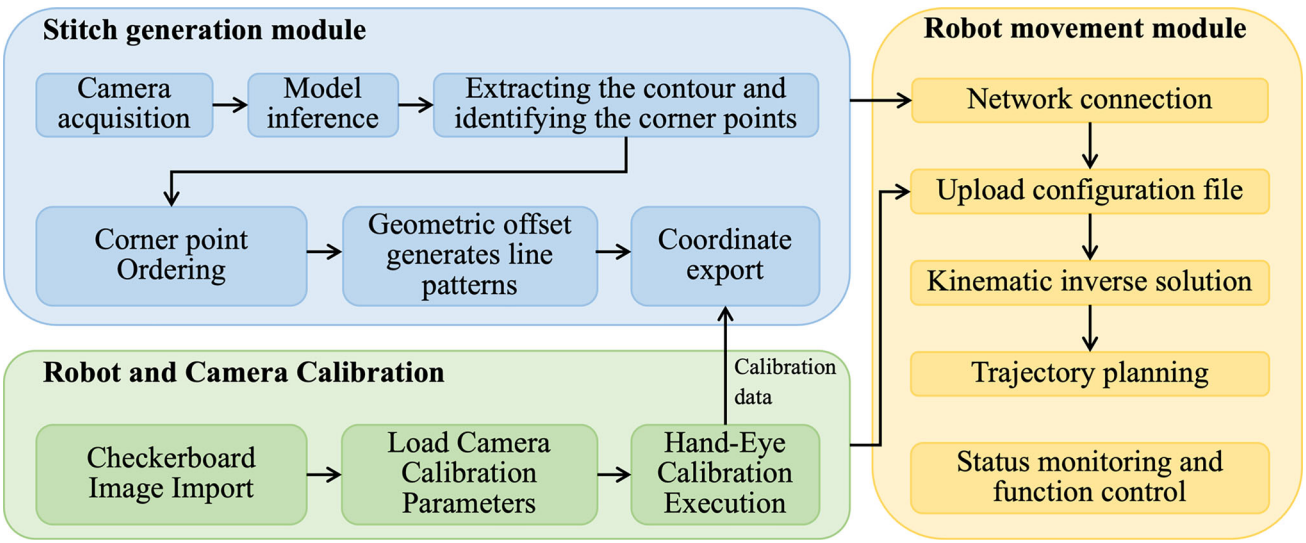


Fig. 13 Three-module software architecture

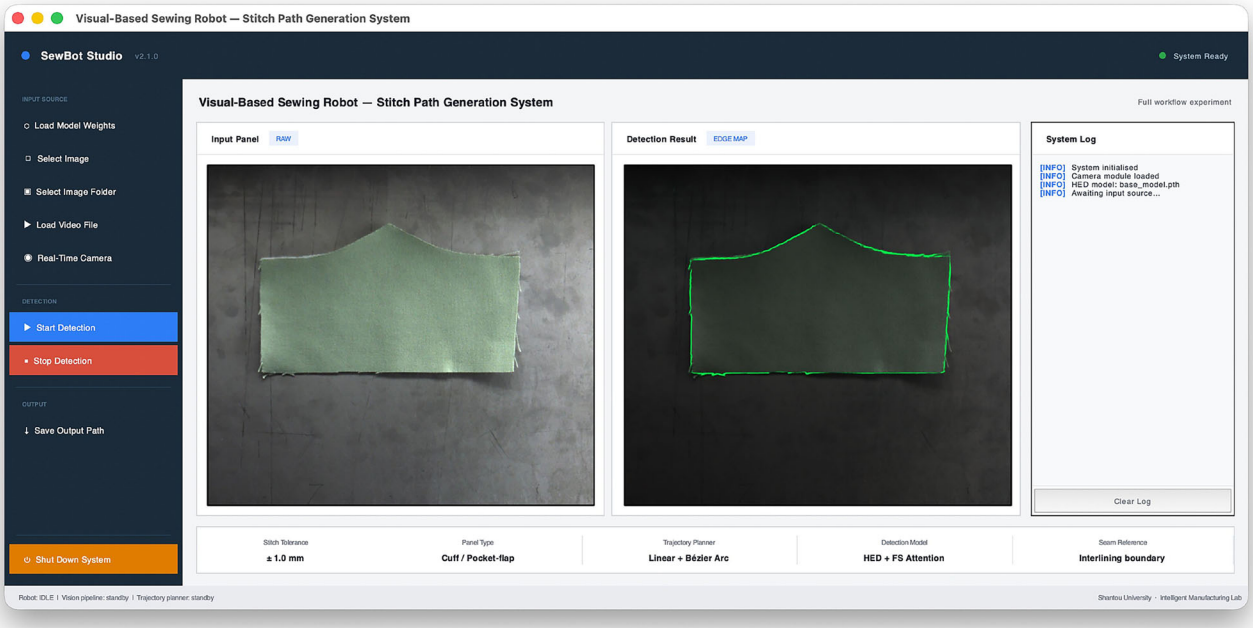


Fig. 14 Stitch generation module graphical user interface, showing the edge map, contour overlay, and interactive corner selection controls

ensure geometric accuracy without compromising overall system efficiency.

Full workflow experimental testing and results

The complete system workflow was validated by executing the end-to-end sewing cycle on two panel types: (i) a cuff panel with an interlining bond line used as the seam reference, and (ii) a pocket-flap panel with a 5 mm seam allowance from the raw edge. Where required, the operator used the graphical interface to refine detected stitch-path corners prior to trajectory generation, consistent with the intended human-supervised operation of the system. Manual corner refinement was required in approximately 15% of tested panels, arising in cases where fabric fraying or low-contrast boundaries produced spurious or missing corner detections in the polygon approximation step. These cases fell into three categories corresponding to the challenging conditions characterised in Table 4: fabric fraying at cut edges producing spurious contour segments (flying-thread condition), low-contrast boundaries under non-uniform illumination (blurred-boundary condition), and geometrically complex corners where polygon approximation produced ambiguous vertex configurations (jagged-edge condition). The distribution across these categories was qualitatively consistent with the relative difficulty levels reflected in the ODS scores reported in Table 4.

For the cuff panel, the maximum deviation from the interlining boundary was $\pm$ 1.0 mm; for the pocket-flap panel, deviations at individual waypoints were within $\pm$ 0.5 mm, with the slightly larger cuff deviation attributable to the lower repeatability of the crank-slider surrogate compared with a production sewing head. These results are consistent with the quality tolerances applied at the partner factory, where $\pm$ 1 mm is considered acceptable for the Yun operation. It should be noted that the experiments were conducted using a pen-tracing surrogate rather than a production sewing head, and therefore do not account for fabric deformation, thread tension, or needle–material interaction effects. All system modules operated within the required response time and functional constraints during testing, as shown in Table 7.

Table 7 Functional validation of system modules

Function under test	Pass criterion	Result
Camera image acquisition	Response time $\leq$ 10 ms	5 ms—Pass
Deep learning contour detection	Contour generated within 10 s	5 s—Pass
Corner selection and stitch generation	Stitch generation rate 100%	100%—Pass
Hand-eye calibration matrix computation	Matrix generated correctly	Pass
Robot connection and motion command	Response time $\leq$ 100 ms	< 50 ms—Pass

From an industrial perspective, the proposed system reduces setup time from several tens of minutes required by

manual measurement and CAD-based workflows to approximately 10 s, representing a substantial improvement in production flexibility for high-mix manufacturing scenarios.

Discussion

The proposed system incorporates several design choices that distinguish it from prior work in vision-based robotic sewing. First, the combination of a domain-specific dataset and the proposed FS attention module address the domain gap between general-purpose edge detection benchmarks and the industrial garment-panel setting. The observed performance differences between HED + FS (ODS 0.8517) and the compared models suggest that the proposed model can learn task-relevant garment-boundary representations with relatively low computational cost, which is beneficial for downstream stitch-path generation. Second, the integration of a human-supervised workflow represents a key aspect of the system design. Rather than being treated as a fallback mechanism, the interactive corner selection interface enables selective operator intervention to resolve geometric ambiguities arising from fabric fraying, cutting irregularities, or weak boundary contrast. In practice, operator input was required for approximately 15% of test panels, indicating that the majority of cases can be processed fully automatically. This selective intervention strategy reduces operator workload compared to fully manual workflows while maintaining high reliability and reflects a collaborative human–robot paradigm suitable for deformable object manipulation tasks. Third, the eye-to-hand calibration configuration simplifies system integration by avoiding camera–robot interference and eliminating the need for recalibration during robot motion. Although this setup requires recalibration when the camera or workspace is repositioned, the relatively low frequency of style changeovers in garment production (approximately weekly in the partner factory) makes this trade-off acceptable in practice.

The primary sources of residual error in the system arise from the combined effects of vision, robot kinematics, and experimental setup limitations. Specifically, the coordinate measurement accuracy of the vision system ($\pm$ 0.2 mm theoretical, $\pm$ 0.5 mm demonstrated after calibration), the repeatability of the MS6MT manipulator ($\pm$ 0.1 mm), and the limited fidelity of the crank-slider surrogate collectively contribute to the observed stitch placement error of approximately $\pm$ 1.0 mm. Replacing the surrogate with a production sewing head and improving illumination conditions are expected to reduce this error toward $\pm$ 0.5 mm, which aligns with the tolerance requirements for fine-fabric garments. From a production perspective, the proposed system offers a significant reduction in setup time compared to conventional workflows based on manual measurement, CAD modelling, and machine code generation. The end-to-end pipeline generates stitch trajectories in approximately 10 s with minimal operator interaction, representing a substantial productivity advantage in high-mix, low-volume manufacturing environments.

The current system has several practical limitations. The perception pipeline is sensitive to severe fabric fraying and low-contrast boundaries, which are the primary causes of corner detection failure and operator intervention. Although the robustness experiments considered flying threads, jagged edges, blurred boundaries, and illumination-domain shift, the current validation does not yet systematically evaluate severe wrinkles, partial occlusions, or a wide range of fabric material properties; these conditions will require further testing before production deployment. The calibration-based coordinate transformation assumes a planar panel on a fixed work surface; out-of-plane deformation or panel shifting would degrade trajectory accuracy. The 10-s processing time and operator supervision requirement currently limit throughput in high-volume production scenarios. In the current system, this processing time is acceptable for high-mix, low-volume production and rapid style changeover, where the main bottleneck is manual measurement and CAD-based setup rather than per-piece cycle time. However, for high-throughput continuous sewing lines, the current pipeline remains slower than the cycle-time requirements of production sewing operations. Future deployment will therefore require acceleration of the edge-map generation module through lightweight model compression, quantisation, GPU/edge-device optimisation, and automated corner classification to reduce both computation time and operator intervention. The reliance on operator intervention for approximately 15% of panels reflects a deliberate design trade-off: full automation of corner detection would reduce throughput latency but increase the risk of undetected geometric errors propagating to trajectory execution. The human-supervised workflow prioritises stitch-path reliability over full automation, which is appropriate for the current system maturity level and consistent with human-centric manufacturing paradigms.

Three concrete directions are identified for extending the current system. First, multi-layer fabric handling requires modelling of inter-layer slip and thickness variation, which could be addressed by integrating force-torque sensing at the end-effector to detect and compensate for layer misalignment during feeding. Second, dynamic fabric deformation during sewing execution could be handled by incorporating a real-time vision feedback loop that updates stitch waypoints based on observed boundary displacement, building on the existing perception pipeline. In addition, integration with an industrial

lockstitch head will require closed-loop synchronisation between robot feed speed, needle phase, and feed-dog motion to avoid fabric wrinkling, stitch-length inconsistency, and needle damage. Third, processing throughput could be improved through model acceleration and a learned corner classification module to reduce operator intervention frequency.

Conclusions

A human-supervised, vision-based robotic sewing system has been developed and experimentally validated for the Yun (perimeter stitch) operation in shirt manufacturing. The system integrates perception, calibration, and robot motion into a coherent pipeline capable of converting raw garment images into executable stitch trajectories with minimal operator intervention. The proposed HED + FS model achieves an ODS of 0.8517 and an OIS of 0.8842 on a purpose-built garment-panel dataset, outperforming the baseline HED and slightly surpassing EDTER while maintaining computational efficiency. The directional attention mechanism enhances boundary localisation under texture-rich and low-contrast conditions, enabling reliable downstream geometric processing. Building on this, the stitch generation pipeline generates ordered seam coordinates in approximately 10 s and incorporates an operator-guided refinement step to address geometric ambiguity, ensuring a balance between automation efficiency and robustness. The combination of Cartesian-space linear and arc trajectory planning with camera and hand–eye calibration enables the MS6MT manipulator to execute stitch trajectories with a placement accuracy within $\pm$ 1 mm under laboratory conditions, consistent with industrial tolerance requirements for the Yun operation.

The results demonstrate a substantial reduction in setup time compared with conventional manual and CAD-based workflows while maintaining acceptable accuracy. The main limitations of the current work are the use of a pen-tip surrogate rather than a real sewing head for experimental validation, the absence of fabric deformation and thread tension modelling, and the restriction of testing to controlled laboratory conditions. Future work will focus on validation with a production lockstitch head and full fabric-on-fabric sewing, extension of the stitch generation pipeline to multi-seam garment assemblies, and further optimisation of edge detection models to improve both accuracy and computational efficiency.

Acknowledgements The authors would like to acknowledge the support provided by the Intelligent Manufacturing Modern Industry College of Shantou University (Project No. 140-181253002).

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Arbeláez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(5), 898–916. https://doi.org/10.1109/TPAMI.2010.161

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 679–698. https://doi.org/10.1109/TPAMI.1986.4767851. PAMI-8.

Chang, R., & Zhu, S. (2022). Probe into the low-carbon economic environment challenges and opportunities of textile and 7 edited by W. Zeng. *Journal of Environmental and Public Health, 2022*(1), 6043903. https://doi.org/10.1155/2022/6043903

Chen, H., Huo, S., Muddassir, M., Lee, H.-Y., Liu, Y., Li, J., Duan, A., Zheng, P., & Navarro-Alarcon, D. (2025a). PSO-based optimal coverage path planning for surface defect inspection of 3C components with a robotic line scanner. *IEEE Transactions on Instrumentation and Measurement, 74*, 1–12. https://doi.org/10.1109/TIM.2025.3552466

Chen, H., Li, S., Fan, J., Duan, A., Yang, C., Navarro-Alarcon, D., & Zheng, P. (2025b). Human-in-the-loop robot learning for smart manufacturing: A human-centric perspective. *IEEE Transactions on Automation Science and Engineering, 22*, 11062–11086. https://doi.org/10.1109/TASE.2025.3528051

Chen, Hongpeng, Liu., Shimin, Li., Zhiyuan, David Navarro-Alarcon., & Zheng, Pai. (2025c). Multimodal autonomous robotic long-horizon task planning via embodied language model and behavior trees. In: *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hangzhou, China: IEEE. Pp. 20283–90

China National Textile and Apparel Council (CNTAC). (2021). *14th Five-Year Plan for the Development of the Textile Industry.* http://www.cntac.org.cn

Choi, Y. W., Lee, J., Lee, Y., Lee, S., Jeong, W., Lim, D. Y., & Lee, S. W. (2025). A vision-guided adaptive and optimized robotic fabric gripping system for garment manufacturing automation. *Robotics and Computer-Integrated Manufacturing, 92*, Article 102874. https://doi.org/10.1016/j.rcim.2024.102874

CICC Research, CICC Global Institute. (2024). Textiles and apparel—-expanding production bases overseas while moving up the industry chain. *The reshaping of China's industry chains* (pp. 331–350). Springer Nature Singapore.

Dollar, P., & Zitnick, C. L. (2015). Fast edge detection using structured forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(8), 1558–1570. https://doi.org/10.1109/TPAMI.2014.2377715

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica, 10*(2), 112–122. https://doi.org/10.3138/FM57-6770-U75U-7727

Douglas, D. H., & Peucker, T. K. (2011). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Classics in cartography* (pp. 15–28). John Wiley & Sons, Ltd.

Fan, Yuchen., Simeone, Alessandro., & Antonelli, Dario. (2025). Reciprocal human–machine learning flow modelling for assisted assembly systems. In: *Materials Research Proceedings*. Vol. 57. Pp. 457–65

Gao, X., Lizhe, Q., Chuangjia, M., & Sun, Y. (2022). Research on real-time cloth edge extraction method based on ENet semantic segmentation. *Journal of Engineered Fibers and Fabrics, 17*, 15589250221131890. https://doi.org/10.1177/15589250221131890

Guan, Z., Xu, Y., Jiang, H., & Jiang, G. (2019). International competitiveness of Chinese textile and clothing industry—A diamond model approach. *Journal of Chinese Economic and Foreign Trade Studies, 12*(1), 2–19. https://doi.org/10.1108/JCEFTS-01-2018-0003

He, Jianzhong, Shiliang, Zhang., Yang, Ming., Shan, Yanhu., & Huang, Tiejun. (2019). Bi-directional cascade network for perceptual edge detection. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE. Pp. 3823–32

Hou, Qibin., Li, Zhang., Cheng, Ming-Ming., & Feng, Jiashi. (2020). Strip pooling: Rethinking spatial pooling for scene parsing. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE. Pp. 4002–11

Hou, Qibin, Daquan, Zhou., & Jiashi, Feng. (2021). Coordinate attention for efficient mobile network design. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE. Pp. 13708–17

Hu, Jie, Li, Shen., & Gang Sun. (2018). Squeeze-and-excitation networks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE. Pp. 7132–41

Jie, Jinghuai., Guo, Yan., Wu, Guixing., Wu, Junmin., Hua, Baojian. (2024). EdgeNAT: Transformer for efficient edge detection

Jing, J., Liu, S., Wang, G., Zhang, W., & Sun, C. (2022). Recent advances on image edge detection: A comprehensive review. *Neurocomputing, 503*, 259–271. https://doi.org/10.1016/j.neucom.2022.06.083

Ku, S., Choi, HyunWoong, Kim, H.-Y., & Park, Y.-L. (2023). Automated sewing system enabled by machine vision for smart garment manufacturing. *IEEE Robotics and Automation Letters, 8*(9), 5680–5687. https://doi.org/10.1109/LRA.2023.3300284

Lema, D. G., Sánchez-González, L., Usamentiaga, R., & de la Calle, F. J. (2026). Benchmarking deep learning models for surface defect detection: A reproducible and statistically-rigorous approach. *Journal of Intelligent Manufacturing, 37*(7), 3001–3018. https://doi.org/10.1007/s10845-025-02672-8

Li, F., Hou, D., Fu, T., Song, J., He, W., & Song, R. (2024a). Research on robot sewing method based on process modeling. *International Journal of Intelligent Robotics and Applications, 8*(2), 401–421. https://doi.org/10.1007/s41315-024-00326-1

Li, W., Hu, Y., Zhou, Y., & Pham, D. T. (2024b). Safe human–robot collaboration for industrial settings: A survey. *Journal of Intelligent Manufacturing, 35*(5), 2235–2261. https://doi.org/10.1007/s10845-023-02159-4

Liu, Y., Yi, J., Chi, P., Liao, H., Zhang, Q., & Wang, Z. (2026). Review of trajectory planning for humanoid welding robot manipulators: From fundamentals to industrial applications in intelligent manufacturing. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-025-02765-4

Liu, Yun., Hu, Xiaowei., Wang, Kai., & Bai, Xiang. (2017). Richer convolutional features for edge detection. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Pp. 5872–81

Ma, J.-W., Gao, S., Yan, H.-T., Lv, Q., & Hu, G.-Q. (2021). A new approach to time-optimal trajectory planning with torque and jerk limits for robot. *Robotics and Autonomous Systems, 140*, Article 103744. https://doi.org/10.1016/j.robot.2021.103744

Macfarlane, S., & Croft, E. A. (2003). Jerk-bounded manipulator trajectory planning: design for real-time applications. *IEEE Transactions on Robotics and Automation, 19*(1), 42–52. https://doi.org/10.1109/TRA.2002.807548

Mineur, Y., Lichah, T., Castelain, J. M., & Giaume, H. (1998). A shape controled fitting method for Bézier curves. *Computer Aided Geometric Design, 15*(9), 879–891. https://doi.org/10.1016/S0167-8396(98)00025-9

Patalas-Maliszewska, J., Łosyk, H., & Dudek, A. (2025). Improving safety in human–robot collaboration towards sustainable production in industry 5.0. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-025-02676-4

Peiper, Donald L. (1968). The kinematics of manipulators under computer control. PhD dissertation, Stanford University

Pu, Mengyang., Huang, Yaping., Liu, Yuming., Guan, Qingji., & Ling, Haibin. (2022). EDTER: Edge detection with transformer

Saining, Xie., & Tu, Zhuowen. (2015). Holistically-Nested Edge Detection. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE Pp. 1395–1403

Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition, 33*(2), 225–236. https://doi.org/10.1016/S0031-3203(99)00055-2

Shen, J., Ramírez-Gómez, Á., Wang, J., & Zhang, F. (2025). Autonomous sewing technology and system: A new strategy by integrating soft fingers and machine vision technology. *Textiles, 5*(4), Article 45. https://doi.org/10.3390/textiles5040045

Simonyan, Karen., & Zisserman, Andrew. (2014). Very deep convolutional networks for large-scale image recognition

Su, Zhuo., Liu, Wenzhe., Yu, Zitong., Hu, Dewen., Liao, Qing., Tian, Qi., Pietikainen, Matti., & Liu, Li. (2021). Pixel difference networks for efficient edge detection. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE. Pp. 5097–5107

Tafuro, A., Molinaro, M., Zanchettin, A. M., & Rocco, P. (2025). Self-supervised vision-driven trajectory planning for intelligent robotic deburring. *Machine Intelligence Research, 22*(4), 655–676. https://doi.org/10.1007/s11633-025-1560-6

Tomasi, C., and Manduchi. R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Bombay, India: Narosa Publishing House. Pp. 839–46

Tsai, R. Y., & Lenz, R. K. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation, 5*(3), 345–358. https://doi.org/10.1109/70.34770

Wang, Qilong., Wu, Banggu., Pengfei, Zhu., Peihua, Li., Wangmeng, Zuo., and Qinghua, Hu. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE. Pp. 11531–39

Whitney, D. E. (1969). Resolved motion rate control of manipulators and human prostheses. *IEEE Transactions on Man-Machine Systems, 10*(2), 47–53. https://doi.org/10.1109/TMMS.1969.299896

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In V. Ferrari, M. Hebert, C.

Sminchisescu, & Y. Weiss (Eds.), *Computer vision—ECCV 2018* (Vol. Vol. 11211, pp. 3–19). Springer International Publishing.

Xavier, Soria., Riba, Edgar., & Angel, Sappa. (2020). Dense extreme inception network: Towards a robust CNN model for edge detection. In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Snowmass Village, CO, USA: IEEE. Pp. 1912–21

Xia, Liqiao., Hongpeng Chen, Jiazhen Pang., Liu, Shimin., Pai, Zheng., and Fazel Ansari. (2025). LLM-Augmented multi-fidelity bayesian optimization for parameter optimization in human-robot collaborative assembly. In: *2025 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. Hangzhou, China: IEEE Pp. 1–6

Xu, Z., Wei, S., Wang, N., & Zhang, X. (2014). Trajectory planning with Bezier curve in Cartesian space for industrial gluing robot. In X. Zhang, H. Liu, Z. Chen, & N. Wang (Eds.), *Intelligent robotics and applications* (Vol. Vol. 8918, pp. 146–154). Springer International Publishing.

Yang, L., & Kumarasinghe, P. (2024). Analysis of the impact of RCEP on the industrial and innovation Chains of China's textile and clothing industry edited by M Hashim. *PLoS ONE, 19*(8), Article e0309708. https://doi.org/10.1371/journal.pone.0309708

Yin, H., Varava, A., & Kragic, D. (2021). Modeling, learning, perception, and control methods for deformable object manipulation. *Science Robotics, 6*(54), Article eabd8803. https://doi.org/10.1126/scirobotics.abd8803

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330–1334. https://doi.org/10.1109/34.888718

Zhang, C., Cui, J., Wu, J., & Zhang, X. (2024). Attention mechanism and texture contextual information for steel plate defects detection. *Journal of Intelligent Manufacturing, 35*(5), 2193–2214. https://doi.org/10.1007/s10845-023-02149-6

Zhong, M., Liang, Y., Xie, W., Cui, W., Lv, H., & Zhong, D. (2025). Real-time dynamic look-ahead trajectory planning for industrial robots based on visual feedback. *International Journal of Advanced Manufacturing Technology, 140*(11–12), 6045–6063. https://doi.org/10.1007/s00170-025-16604-6

Zhu, J., Cherubini, A., Dune, C., Navarro-Alarcon, D., Alambeigi, F., Berenson, D., Ficuciello, F., Harada, K., Kober, J., Li, X., Pan, J., Yuan, W., & Gienger, M. (2022). Challenges and outlook in robotic manipulation of deformable objects. *IEEE Robotics & Automation Magazine, 29*(3), 67–77. https://doi.org/10.1109/MRA.2022.3147415

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.