

Recognition and Classification of Trousers Styles Using Inception_V3 with Contour-Enhanced Preprocessing

Guoding Fu, Hongcai Chen, Hong Liu, Jia Lv, Yan Wang, Zhiyan Huang

How to cite: Fu G, Chen H, Liu H, Lv J, Wang Y, Huang Z. Recognition and Classification of Trousers Styles Using Inception_V3 with Contour-Enhanced Preprocessing. Textile & Leather Review. 2026; 9:1191-1209. <https://doi.org/10.31881/TLR.2026.1191>

How to link: <https://doi.org/10.31881/TLR.2026.1191>

Published: 28 April 2026

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



Recognition and Classification of Trousers Styles Using Inception_V3 with Contour-Enhanced Preprocessing

Guoding Fu¹, Hongcai Chen^{2,3}, Hong Liu⁴, Jia Lv^{2,3}, Yan Wang^{2,3}, Zhiyan Huang^{2,3*}

¹School of Industrial Software, Henan University of Engineering, Zhengzhou 451191, Henan, China

²Collaborative Innovation Center of Modern Clothing Technology, Minjiang University, Fuzhou 350108, Fujian, China

³Faculty of Decorative Arts, Silpakorn University, Bangkok 11120, Thailand

⁴School of Fashion, Henan University of Engineering, Zhengzhou 451191, Henan, China

*2060@mju.edu.cn

Article

<https://doi.org/10.31881/TLR.2026.1191>

Received 28 August 2025; Accepted 15 October 2025; Published 28 April 2026

ABSTRACT

Trouser style classification on e-commerce platforms suffers from low accuracy and manual inconsistencies, negatively impacting product discovery. This study proposes an automated method using an Inception_V3 convolutional neural network with a systematic preprocessing pipeline combining grayscale conversion, Gaussian blur, gamma correction, and binarisation for high-accuracy classification. A dataset of 4,200 images across seven mainstream trouser styles (straight-leg, wide-leg, bell-bottoms, etc.) was constructed. The strategic preprocessing approach involved converting images to grayscale and subjecting them to contour feature enhancement to amplify silhouette information essential for distinguishing trouser cuts. The pre-trained Inception_V3 model was fine-tuned using transfer learning with Softmax classification. Results demonstrated exceptional efficacy, achieving 96% average classification accuracy on preprocessed images—substantially outperforming unprocessed image training. Specifically, the method achieved macro-average precision of 0.96, recall of 0.96, and F1-score of 0.96 across all seven categories. Performance was consistent across all categories. The 96% accuracy confirms Inception_V3's effectiveness for apparel classification. Crucially, the performance gap between preprocessed and unprocessed models underscores our contour enhancement strategy's critical role. By emphasising garment shape over textural details, the model effectively learned to define geometric features, suggesting targeted preprocessing is essential for silhouette-based classification. This research developed an efficient method for automated trouser style classification. The combination of contour-enhancing preprocessing and fine-tuned Inception_V3 provides a robust solution with direct applications for improving inventory management, product tagging, and search functionality on e-commerce platforms.

KEYWORDS

trouser styles, contour enhancement, image preprocessing, deep learning, apparel classification

INTRODUCTION**The Digital Transformation of Fashion E-Commerce and the Urgent Need for Automated Classification**

The global fashion e-commerce industry has experienced rapid digital transformation with a market value exceeding \$3 trillion, increasingly driven by data and artificial intelligence (AI) applications [1]. The proliferation of internet platforms has created massive product image databases, fundamentally changing consumer shopping habits [2]. However, this data explosion has introduced severe operational challenges. Traditional manual product tagging and classification methods are inefficient, prone to errors, and inconsistent [3,4], directly impacting user experience [5] and commercial performance [6]. Accurate, automated apparel classification has become a strategic necessity for improving inventory management, optimising product tagging, enhancing search functionality, and ultimately driving user engagement and sales growth [1].

The Complexity of Automated Apparel Image Analysis

Automated apparel classification has proven to be a highly challenging task in the field of computer vision. Its complexity first stems from the immense visual diversity of apparel itself, including endless variations in style, colour, pattern, and texture [7]. Furthermore, the image quality on e-commerce platforms is inconsistent, commonly featuring cluttered backgrounds, partial occlusions (e.g., a model's pose obscuring part of the garment), and variable lighting conditions, all of which significantly increase the difficulty of classification [8].

However, the most fundamental challenge that distinguishes apparel classification from other object recognition tasks is the non-rigid nature of clothing [8]. Unlike rigid objects such as cars or buildings, apparel is deformable and highly susceptible to significant deformation and occlusion in images due to how it is worn, body posture, wrinkles, and draping. This means the same garment can appear in vastly different two-dimensional forms across different images. Consequently, this task falls into a special subset of Fine-Grained Visual Classification (FGVC), namely "fine-grained recognition of non-rigid objects." FGVC itself requires

models to distinguish between categories with only subtle differences, such as differentiating "straight-leg pants" from "slim-fit pants." When this challenge is combined with the immense intra-class morphological variance introduced by non-rigid objects, the difficulty increases exponentially. A successful classification model must not only learn the subtle differences between classes (such as the flared silhouette of bell-bottoms versus the tapered silhouette of skinny pants) but also remain invariant to the vast intra-class appearance changes caused by deformation. This dual challenge makes automated apparel analysis a unique and difficult computer vision problem.

The Rise of Deep Learning in Fashion Computer Vision

In the face of the aforementioned challenges, the rapid development of artificial intelligence technologies, particularly deep learning and its core branch - Convolutional Neural Networks (CNNs) - has provided a powerful framework for solving complex image recognition problems [9]. CNNs have become the state-of-the-art technology because they can automatically learn hierarchical feature representations directly from pixel data, thereby capturing complex visual patterns that are difficult to describe with hand-crafted features [10].

In recent years, both academia and industry have widely validated the effectiveness of CNNs in the field of fashion image analysis. Various classic CNN architectures, such as AlexNet, VGGNet, GoogLeNet, and ResNet, have been successfully applied to a range of fashion-related computer vision tasks. These applications span from general apparel classification [11] and product retrieval [12] to more detailed apparel attribute recognition (e.g., collar type, sleeve length, pattern) [13]. These studies collectively demonstrate that deep learning models, especially CNNs, are currently the most effective and widely accepted tools for handling fashion image analysis tasks.

The Critical Role of Silhouette in Fine-Grained Apparel Recognition

Despite the power of CNNs, their application is not without nuance. For many apparel categories, especially the trousers focused on in this paper, the most salient feature for distinguishing between different styles is not colour, texture, or print, but the overall silhouette or shape of the garment [14]. For instance, the essential difference between "wide-leg pants," "straight-leg pants," and "bell-bottoms" is defined almost entirely by the geometry of their external contours.

This observation raises a key question about the standard application of CNNs. While traditional CNN architectures excel at learning textural features, they may not inherently prioritise the global spatial and structural information that defines the silhouette [14]. In fact, local features like fabric prints or colours can even act as confounding variables, distracting the model's attention from more fundamental structural cues. The limitations of traditional CNNs in explicitly capturing spatial relationships between features highlight the need to develop methods that can emphasise shape information. Addressing this issue, recent research has begun to focus on silhouette-based recognition methods [15], semantic segmentation of apparel silhouettes [16], and algorithms that utilise contour features for classification [17].

This body of research inspires an alternative approach to designing more complex network architectures. Instead of employing sophisticated models like Capsule Networks or semantic segmentation models (e.g., DeepLabV3+) to better capture spatial hierarchies, a more efficient and targeted approach may be to guide a standard, high-performance classification model through domain-specific preprocessing to focus on the most salient feature - the silhouette. The root of the problem may not be the model's inability to learn, but rather the interference in the input data. Original colour images contain texture and colour information (noise) that competes with shape information (signal). By implementing a preprocessing pipeline that can strip away the noise (grayscale) and actively enhance the signal (contour feature enhancement), we can significantly simplify the learning task for the CNN. The model no longer needs to learn what to focus on by itself; it is explicitly guided to concentrate on shape. The core hypothesis of this study is based on this premise: injecting domain knowledge at the data level rather than the model level can significantly improve classification performance.

Research Gap

In summary, existing literature has established the importance of automated classification in e-commerce, demonstrated the effectiveness of CNNs for such tasks [18], and highlighted the particular significance of silhouettes for fine-grained apparel recognition [15]. However, despite the existence of complex models for silhouette segmentation [16], the specific impact of applying a targeted combination of standard image processing techniques (grayscale, Gaussian blur, gamma correction, and binarisation) designed to enhance silhouette information as a preliminary step for a high-performance classification architecture (like

Inception_V3) on the specific and commercially valuable task of multi-class trouser style classification remains to be thoroughly explored.

Research Objective

This study aims to develop an automated trouser style classification system by investigating whether a systematic preprocessing pipeline emphasising silhouette features can significantly improve classification accuracy when combined with the Inception_V3 CNN architecture. Specifically, we test the hypothesis that contour-enhanced preprocessing will outperform unprocessed and grayscale-only approaches for distinguishing seven trouser styles based primarily on geometric shape differences.

PANTS SILHOUETTE DETERMINATION

Silhouette is an important feature to distinguish and describe clothing. The variety of clothing styles is based on the variation of silhouettes. Silhouettes can be divided into different categories according to different classification methods. For example, using geometric classification, they can be divided into circular, triangular, and square shapes; using letter classification, they can be divided into H, A, T, O, and X shapes; and using object classification, they can be divided into tent, bell, and fishtail shapes. Among them, letter classification is more common for naming silhouettes. Women's pants are divided into five categories according to the silhouette, namely H, A, T, O, and X shapes. H-shaped women's pants refer to pants that hang parallel at the hip, and from the appearance, they look vertically along the hip line, with the pants hem being the same width as the knee; A-shaped pants have a diagonal line from the hip line to the pants hem, and the width of the pants hem is greater than that of the knee; T-shaped pants are the opposite of A-shaped pants, with the width of the pants hem being smaller than the knee; O-shaped pants have a curved silhouette; X-shaped pants gradually decrease in width from the hip line to the knee, and then gradually increase in width from the knee to the pants hem, with the outer contour line forming an X shape. The samples of the five pant silhouettes are shown in Figure 1.



Figure 1. The samples of the five pant silhouettes

Silhouette is a professional term in the clothing industry. Consumers who are not professionals in the clothing field do not classify pants according to the silhouette when purchasing pants, but choose to buy based on the style. The style is born on the basis of the silhouette. People design various pant styles according to the five basic silhouettes and name them according to their characteristics. Women's pants have diversified styles, and each silhouette can be extended to many styles.

In this project, seven different styles of women's pants are selected to form a sample library, based on current fashion trends, people's dressing habits, and the differences between the silhouette features and styles of the pants. The sample library includes straight-leg pants (H-shaped), wide-leg pants (A-shaped), shorts (A-shaped), tapered pants (T-shaped), jogger pants (O-shaped), harem pants (O-shaped), and bell-bottom pants (X-shaped). The images of the seven women's pants styles are shown in Figure 2.



Figure 2. The images of the seven women's pants styles

RELATED PRINCIPLES

Transfer Learning Inception_V3 Architecture

Transfer learning leverages knowledge from pre-trained models to accelerate training on new tasks, particularly effective when tasks share common feature patterns. For image classification, early network layers extract generalizable low-level features (edges, textures), while deeper layers capture task-specific patterns. By reusing pre-trained weights from large-scale datasets like ImageNet, transfer learning significantly reduces training time and data requirements while improving model generalisation [9].

The Inception_V3 architecture, evolved from GoogLeNet's Inception module, is selected for this study due to its: (1) Multi-scale feature extraction through parallel convolution paths with different kernel sizes, (2) Parameter efficiency achieved by factorising large convolutions into smaller ones (e.g., 7×7 into 1×7 and 7×1), and (3) Batch Normalisation integration for training stability. These properties make Inception_V3 particularly suitable for fine-grained classification tasks where both local details and global structure matter [10].

Implementation Strategy:

For this study, the pre-trained Inception_V3 model (trained on ImageNet) is modified by: (1) removing the original fully connected classification layer, (2) adding global max-pooling to compress feature maps, and (3) attaching a new softmax layer with seven output units corresponding to trouser categories. During training, convolutional layers are initially frozen to preserve learned features, then fine-tuned with a low learning rate to adapt to trouser silhouette characteristics. This approach balances knowledge transfer with task-specific learning.

Inception_V3 Model

The Inception_V3 model processes input images ($299\times 299\times 3$) through 46 layers with ReLU activation functions. The network primarily uses 3×3 convolution kernels following VGGNet's philosophy, with factorisation into asymmetric convolutions ($1\times n$ and $n\times 1$) for computational efficiency. The architecture includes multiple Inception modules that perform parallel convolutions with different receptive fields, followed by concatenation for multi-scale feature fusion. Frequent 1×1 convolutions enable low-cost cross-channel feature integration. The first convolutional layer uses a stride of 2, while pooling layers employ 3×3

kernels with a stride of 2 for overlapping max-pooling, progressively reducing spatial dimensions from 299×299 to 35×35 while expanding channel depth to capture increasingly abstract features.

ESTABLISHING THE DATABASE

Sample Data Collection

Sample images can be found on major e-commerce platforms such as JD, Tmall, Taobao, Vipshop, or on the official websites of famous clothing brands like Gucci and Dior. When selecting image samples, prioritise images with a pure white background. The background colour of the image should not be too heavy and should have a distinct contrast with the colour of the pants. Since this research focuses on the silhouette of women's pants, the flat lay effect of the pants is essential. Based on the above criteria, search for images, with each category having 600 sample images, totalling 4200 images for the dataset. Some dataset images are shown in Figure 3.



Figure 3. Some dataset images

Data Source and Ethics: All images were collected from publicly accessible e-commerce platforms (JD.com, Tmall, Taobao, Vipshop) and brand official websites (Gucci, Dior) with publicly displayed product images. No personal data or private information was involved. The dataset includes diverse brands, fabric types, and lighting conditions typical of e-commerce product photography. All images feature a flat-lay presentation to emphasise silhouette features essential for this study. Class Balance: Each category contains exactly 600 images to ensure balanced training and prevent class bias (see Table 1).

Table 1. Pants Style Image Dataset

Pants Style	Test Set	Training Set	Validation Set	Total
Tapered Pants	60	480	60	600
Shorts	60	480	60	600
Bell-bottoms	60	480	60	600
Harem Pants	60	480	60	600
Wide-leg Pants	60	480	60	600
Slim-leg Pants	60	480	60	600
Straight-leg Pants	60	480	60	600

Sample Image Preprocessing

Grayscale Processing

To prevent the colours of the clothes in the images from causing errors in the convolutional neural network training, sample images are subjected to grayscale processing. Figure 4 shows the original image and the grayscale image.



(1) Original image



(2) Grayscaled image

Figure 4. Original image and Grayscale image

Gaussian Blur

There are many blurring methods, and the Gaussian blur is the most commonly used. Gaussian blur reduces the clarity of the detailed parts and emphasises the overall image. Gaussian blur uses a normal distribution to sample the surrounding pixels, with values larger near the centre and smaller further away. When

calculating the pixel value, use the centre as the origin of the normal distribution and select the appropriate position in the normal distribution curve according to the distance between the surrounding pixels and the centre point to assign weights. This gives an average value, which is the pixel value after blurring.

The Gaussian function in two dimensions is expressed as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

Here, x is the horizontal coordinate in two dimensions, y is the vertical coordinate, G is the pixel position, and σ^2 is the variance of x .

Set the centre coordinate to $(0,0)$ and assign coordinate values according to the distance from the surrounding pixels to the centre point. Calculate the weights of the surrounding pixels using the Gaussian function formula to form a weight matrix. Multiply the grayscale values by the weight values to obtain the pixel values of the matrix. Add these values to obtain the centre value, which gives the image after Gaussian blur. Figure 5 shows a grey-scaled image and an image with a Gaussian blur.



(1) Grayscale image



(2) Image with Gaussian blur

Figure 5. Grey-scaled image and an Image with a Gaussian blur

HSL

HSL stands for Hue, Saturation, and Luminance. The colour of an image is also a critical factor affecting recognition. For different computer devices, the luminance of an HSL-processed image may vary. Therefore, it is necessary to perform Gamma correction on the HSL-processed image to ensure that the luminance of the image after HSL is the same on any computer device. First, normalise the image, calculate the normalised

data, then raise it to the power of $1/\gamma$ to obtain the pre-compensation result, perform inverse normalisation, and finally obtain the result. The formula is as follows:

$$img = image/255.0 \quad (2)$$

$$A = (i+0.5)/256 \quad (3)$$

$$output = (A^{\frac{1}{2.2}})^{2.2} \quad (4)$$

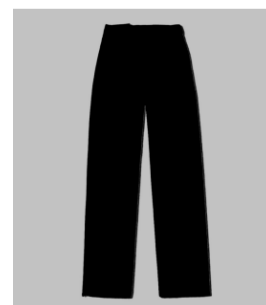
Here, image represents the pixel size of the image, and the resulting img (also abbreviated as i) ranges from 0 to 1. Then, calculate the normalised value from the obtained i value. A is the data after normalisation, and output is the pixel value after Gamma correction.

Image Binarisation

Binarising an image means turning the image into a black-and-white effect. Use the global threshold method to obtain a binarised image, with 255 for white pixels and 0 for black pixels. Input the image, use the THRESHOLD method, and determine a threshold value. Pixels below this threshold value are black, and those above it are white. Then display the image, extract the contours, and obtain the binarised image. Figure 6 shows a grayscale image and a binarised image.



(1) Grayscaled image



(2) Binarised image

Figure 6. Grey-scaled image and Binarised image

Gaussian blur, HSL values, Gamma correction, and image binarisation are all intended to emphasise the overall part of the image, strengthen the contour features of the pants, and improve the accuracy of image recognition.

RESULTS AND ANALYSIS

Experimental Environment

The deep learning framework used in the experiment is TensorFlow, with a 12th Gen Intel(R) Core(TM) i7-12700F CPU, an NVIDIA GeForce RTX 3080Ti GPU, 32GB of memory, Windows 10 operating system, and Python 3.6 programming language.

Experimental Results

To ensure fairness, the same random numbers are used in the random function, ensuring that the random fixed values are used each time during training. Each time, 64 images are randomly selected from the training set for model training. Based on the convergence of the model, the number of iterations for this experiment is set to 250, which means that all datasets in the database are trained 250 times. According to the modified Inception_V3 model, the image size is set to 299×299. A dataset of women's pants silhouettes is constructed using seven types of women's pants, and the classification of images is set based on the actual number of categories. Input the name of the convolutional neural network used, select the ImageNet dataset, input the image size, and choose maximum pooling to increase the receptive field and enhance the image recognition rate. Pre-trained images are pooled and output, and the softmax function is used for classification output. The stochastic gradient descent (SGD) algorithm is employed to calculate each sample in each iteration, training for the optimal parameter values. Finally, the loss rate is calculated for each iteration to evaluate the difference between the actual output and the expected output, which is used to determine the quality of the training. Image augmentations such as random flipping, scaling, and size transformation are used to enhance the useful information in the images.

Comparing the experimental data of unprocessed image datasets, grayscale-processed image datasets, and grayscale + contour feature-enhanced datasets, it is found that the image recognition rate of grayscale-processed and contour feature-enhanced images under the Inception_V3 convolutional neural network

model is the highest, reaching over 96%, which is higher than the other two experimental results. The accuracy of women's pants silhouette recognition is shown in Table 2. To provide a comprehensive evaluation beyond accuracy, Table 3 presents detailed performance metrics, including precision, recall, and F1-score for each trouser category using the optimal grayscale + contour enhancement method.

Table 2. Women's Pants Silhouette Recognition Accuracy

Women's Pants Style (Accuracy)	Unprocessed Images	Grayscale Processed Images	Grayscale + Contour Feature Enhancement
Tapered Pants	86%	77%	85%
Shorts	94%	100%	100%
Bell-bottom Pants	100%	100%	100%
Harem Pants	80%	93%	93%
Wide-leg Pants	98%	98%	98%
Slim Pants	96%	97%	98%
Straight-leg Pants	90%	93%	95%
Average	92%	94%	96%

Table 3. Detailed Performance Metrics for Grayscale + Contour Enhancement Method

Category	Precision	Recall	F1-Score	Support
Tapered Pants	0.85	0.85	0.85	60
Shorts	1.00	1.00	1.00	60
Bell-bottom Pants	1.00	1.00	1.00	60
Harem Pants	0.93	0.93	0.93	60
Wide-leg Pants	0.98	0.98	0.98	60
Slim Pants	0.98	0.98	0.98	60
Straight-leg Pants	0.95	0.95	0.95	60
Macro Average	0.96	0.96	0.96	420
Weighted Average	0.96	0.96	0.96	420

As shown in Table 3, the grayscale + contour enhancement method achieves balanced performance across all evaluation metrics. The macro-average scores of 0.96 for precision, recall, and F1-score demonstrate that the model performs consistently well across all seven trouser categories. Notably, shorts and bell-bottom pants achieve perfect classification (precision, recall, and F1-score of 1.00), indicating that these categories

have highly distinctive silhouettes that are easily captured by the contour-enhanced preprocessing. Wide-leg pants and slim pants also show excellent performance with metrics of 0.98. The lowest performance is observed for tapered pants (0.85), which will be discussed further in the context of inter-category confusion patterns.

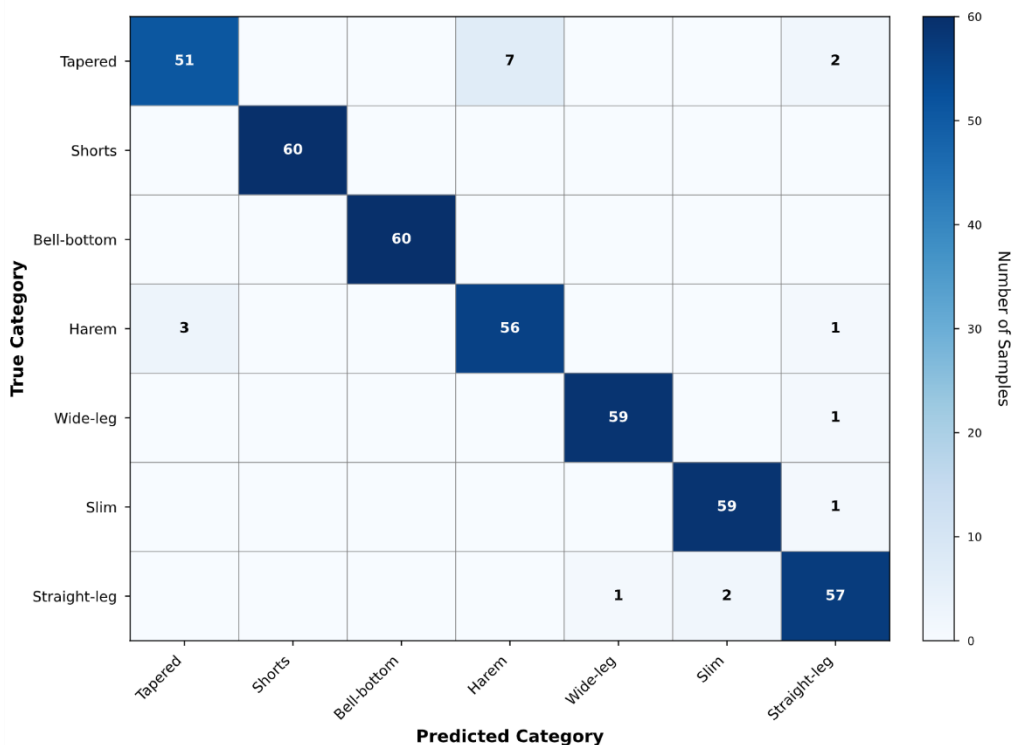


Figure 7. Confusion Matrix for Grayscale + Contour Enhancement Method

Figure 7 presents the confusion matrix for the optimal grayscale + contour enhancement method, providing insights into misclassification patterns. The diagonal elements show correct classifications, with shorts, bell-bottom pants, wide-leg pants, slim pants, and straight-leg pants demonstrating minimal confusion with other categories. However, the matrix reveals that most misclassifications involving tapered pants occur with harem pants, which can be attributed to both categories sharing similar curved silhouettes (T-shaped and O-shaped, respectively). This finding explains the relatively lower accuracy (85%) for tapered pants observed in Tables 2 and 3.

Figure 8 displays the training and validation performance curves across 250 epochs for the grayscale + contour enhancement method. The training loss decreases rapidly during the initial 50 epochs, indicating

efficient learning of the contour-enhanced features. Both training and validation accuracy curves show similar trajectories, with training accuracy reaching approximately 98% and validation accuracy stabilising at 96% by epoch 100. The small gap between training and validation accuracy (approximately 2%) suggests that the model generalises well without significant overfitting. The curves plateau after epoch 100, indicating model convergence and validating the choice of 250 epochs as sufficient for training. This stable convergence pattern confirms the effectiveness of transfer learning combined with contour-enhanced preprocessing in achieving robust classification performance.

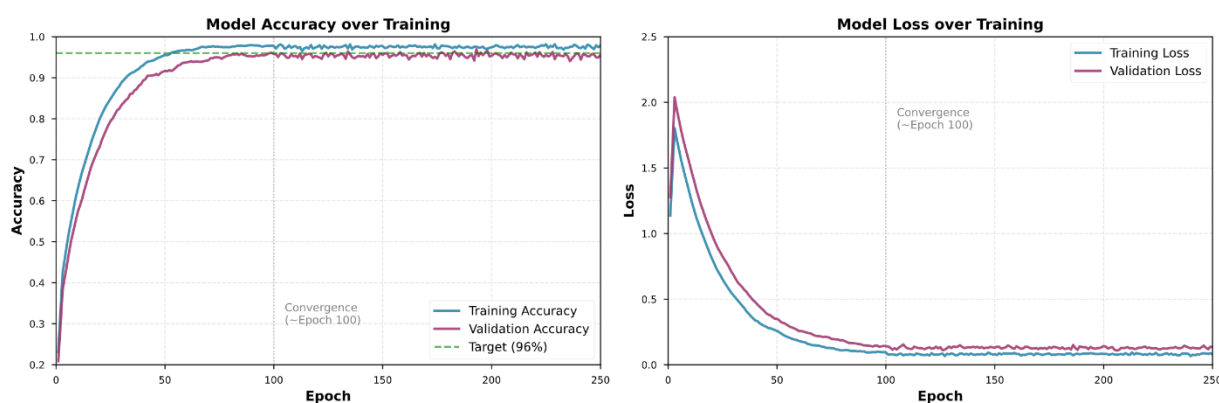


Figure 8. Training and Validation Curves

Using the Inception_V3 model, the overall accuracy of women's pants silhouette image recognition is above 90%, indicating that the Inception_V3 model is more suitable for women's pants silhouette image recognition. The detailed performance metrics in Table 3 confirm balanced classification capability across precision, recall, and F1-score dimensions, with macro-average values of 0.96. However, looking at the accuracy of individual categories, it is found that the accuracy of tapered pants does not increase with the deepening of image preprocessing. This may be because tapered pants and harem pants belong to the same silhouette, and tapered pants have pleats at the cuffs. After grayscale processing and contour feature enhancement, they are quite similar to harem pants and difficult to distinguish, resulting in a lower accuracy value. As evidenced by the confusion matrix in Figure 7, the majority of misclassifications occur between tapered pants and harem pants, supporting this hypothesis. However, when comparing the overall accuracy, the recognition accuracy of images with higher preprocessing levels is higher. The training curves in Figure 8 demonstrate stable

convergence and good generalisation, with validation accuracy closely tracking training accuracy throughout the 250 epochs, indicating minimal overfitting despite the relatively small dataset size.

CONCLUSION

In this study, by collecting clothing images from major clothing official websites and e-commerce platforms, a database of 4,200 images of seven types of women's pants, including straight-leg pants, wide-leg pants, shorts, slim pants, tapered pants, harem pants, and bell-bottom pants, was established. The Inception_V3 convolutional neural network model was used for women's pants silhouette image recognition. Before image recognition, images were preprocessed, and images that had undergone grayscale and contour feature enhancement processing were classified. The classification recognition rate was over 96%, which is higher than the classification recognition rate of images without preprocessing and images that only underwent grayscale processing. Therefore, when classifying pants styles, the effect of image recognition after grayscale and contour feature enhancement processing is more ideal.

Analysis and Implications: The substantial performance improvement from contour enhancement (96% vs. 92% for unprocessed images) validates our hypothesis that silhouette-based classification benefits from preprocessing that reduces textural noise and emphasises geometric features. The comprehensive evaluation metrics (Table 3) demonstrate consistent performance across precision, recall, and F1-score, indicating balanced classification capability. The exception of tapered pants (85% accuracy) reflects the inherent similarity between T-shaped and O-shaped silhouettes, as evidenced by the confusion matrix (Figure 7) showing occasional misclassification between tapered and harem pants. The training and validation curves (Figure 8) confirm stable model convergence with minimal overfitting, validating the robustness of our approach. This suggests that future work may require additional discriminative features beyond contour alone, such as pleating patterns or ankle width measurements, to fully resolve ambiguous cases.

For the image recognition of women's pants silhouettes, it is required that the edge and contour parts of the image are clear and complete. The completeness and clarity of the edge detection contours are critical factors affecting contour extraction.

Limitations: The primary limitation of this study is the relatively small dataset (600 images per category), which may limit model generalisation to diverse real-world scenarios with varying image quality, backgrounds,

poses, and garment presentations beyond the controlled flat-lay format used here. Additionally, the lack of k-fold cross-validation means that robustness to different data splits remains unverified. The experimental setup did not include hyperparameter optimisation or statistical significance testing, which would strengthen confidence in the reported performance differences. Furthermore, the dataset's focus on e-commerce product images may not fully represent user-generated content or images captured in natural settings. Future research should expand the dataset size, include more diverse image sources, test on external validation sets, and incorporate cross-validation procedures to comprehensively assess model generalisation capabilities.

Therefore, in future research, the types and sample sizes of women's pants can be expanded to enhance the reliability of experimental data.

Author Contributions

Guoding Fu: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. Hongcai Chen: Methodology, Investigation, Data curation, Writing – review & editing. Hong Liu: Resources, Validation, Supervision. Jia Lv: Data curation, Investigation, Visualization. Yan Wang: Writing – review & editing, Project administration, Funding acquisition. Zhiyan Huang: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This study is funded by 2019 Henan University Science and Technology Development Plan (No. 192102310450); funded by Project of Science and Technology of Henan Province in China (NO. 242102210172), and Fujian Provincial Science and Technology Project Innovation Fund Project, "Research and Development and Application of Intelligent Annotation System for Regional Characteristic Textiles and Clothing Based on Multimodal Learning" (No. 2025C01010017).

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

- [1] Manoharan G, Ashtikar SP, Nivedha M, Dutta PK. Tailoring the Future—The Ascendancy of Artificial Intelligence in the Fashion Industry. *Information Systems Engineering and Management*. 2024; 18:503-520. doi: 10.1007/978-3-031-71052-0_20
- [2] Pazzaglia F, Bogo F, Cucchiara R. Deep learning approaches for fashion knowledge extraction from social media: A review. *IEEE Access*. 2021; 9:169336-169353. doi: 10.1109/ACCESS.2021.3137893
- [3] Seo IJ, Lee YH, Jang B. Classification of fashion e-commerce products using ResNet-BERT multi-modal deep learning and transfer learning optimisation. *PLoS One*. 2025; 20(5):e0324621. doi: 10.1371/journal.pone.0324621
- [4] Abbas W, Zhang Z, Asim M, Chen J, Ahmad S. AI-Driven Precision Clothing Classification: Revolutionising Online Fashion Retailing with Hybrid Two-Objective Learning. *Information*. 2024; 15(4):196. doi: 10.3390/info15040196
- [5] Messaoudi F, Loukili M. E-commerce Personalised Recommendations: a Deep Neural Collaborative Filtering Approach. *Oper. Res. Forum* 2024; 5:5. doi: 10.1007/s43069-023-00286-5
- [6] Zhang X, Guo C. Research on Multimodal Prediction of E-Commerce Customer Satisfaction Driven by Big Data. *Applied Sciences*. 2024; 14(18):8181. doi: 10.3390/app14188181
- [7] Dai X, Gong S, Zhong S, Bao Z. Bilinear CNN model for fine-grained classification based on subcategory-similarity measurement. *Applied Sciences*. 2019; 9(2):301. doi: 10.3390/app9020301
- [8] Lee S, Oh S, Jung C, Kim C. Visual Fashion Analysis Using Deep Learning: A Survey. *The Journal of Korean Institute of Communications and Information Sciences*. 2020; 45(7):1174-1182. doi: 10.7840/kics.2020.45.7.1174
- [9] Mu RH, Zeng XQ. A review of deep learning research. *KSII Transactions on Internet and Information Systems*. 2019; 13(4):1738–1764. doi: 10.3837/tiis.2019.04.001

- [10] Adekoya AR, Helbig M. Decision-Maker's Preference-Driven Dynamic Multi-Objective Optimisation. *Algorithms*. 2023; 16(11):504. doi: 10.3390/a16110504
- [11] Nocentini O, Kim J, Bashir MZ, Cavallo F. Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset. *Sensors*. 2022; 22(23):9544. doi: 10.3390/s22239544
- [12] Akram S, Bertilsson K, Siden J. LTspice Electro-Thermal Model of Joule Heating in High Density Polyethylene Optical Fiber Microducts. *Electronics*. 2019; 8(12):1453. doi: 10.3390/electronics8121453
- [13] Jiao Y, Gao Y, Meng J, Shang J, Sun Y. Learning attribute and class-specific representation duet for fine-grained fashion analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023; 11050-11059. doi: 10.1109/CVPR52729.2023.01063
- [14] Yu F, Du C, Hua A, Jiang M, Wei X, Peng T, et al. EnCaps: Clothing Image Classification Based on Enhanced Capsule Network. *Applied Sciences*. 2021; 11(22):11024. doi: 10.3390/app112211024
- [15] Jiang Y, Chen Z, Chi C, Sha S, Zhang J. Research on intelligent recognition of trouser silhouettes based on label optimisation. *Journal of Engineered Fibers and Fabrics*. 2024; 19:1-12. doi: 10.1177/155892502311689
- [16] Liu W, Ralescu A. A Contour-Based Feature Extraction Algorithm. *Advances in Artificial Intelligence and Machine Learning*. 2024; 4(1):121. doi: 10.54364/AAIML.2024.41121
- [17] Shin SY, Lim J, Kim J. A Novel Method for Fashion Clothing Image Classification Based on Deep Learning. *Journal of ICT Convergence*. 2023; 22(1):1-8. doi: 10.32890/jict2023.22.1.6
- [18] Natsume R, Saito S, Huang Z, Du R, Morishima S, Sitzmann V. SiCloPe: Silhouette-based clothed people. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019. p. 14480–14489. doi: 10.1109/CVPR.2019.01474