Intelligent Clothing Design Interaction System Based on Multimodal Learning

Ting Zhou
Institute of Textile and Apparel, Keyi College of Zhejiang Sci-tech University
China
ZhouTing_zt@outlook.com

Abstract: *The rapid development of Artificial Intelligence (AI) in fashion design has enhanced the efficiency of clothing production and customization. However, conventional design systems often fail to effectively combine text and image information, such as sketches and product details, when creating innovative and user-centric designs. This paper proposes an Intelligent Clothing Design Interaction System (ICDIS) based on multimodal learning to improve AI-aided fashion design. The ICDIS system combines computer vision and Natural Language Processing (NLP) synergistically to provide an optimized, more efficient fashion design process. The proposed ICDIS utilizes multimodal learning by integrating transformer-based NLP for text description and computer vision for image processing. With the integration, real-time design modification is enabled, allowing designers to iteratively enhance outputs based on text input and visual feedback, thereby improving the AI-aided fashion design process to be more interactive, efficient, and creative. Empirical tests demonstrate that ICDIS enhances design accuracy by 27% (in terms of adherence to professional designer critiques) and accelerates design iterations by 35% compared to conventional AI-based design environments. Users' satisfaction and engagement rates also rose by 22%, underscoring the efficacy of multimodal interaction for creative tasks. This work has extensive potential for fashion design, e-shopping, and online personalization, where AI-based programs can increase creativity and efficiency. In its ability to bridge the gap between text-based inspiration and image-based design, ICDIS represents a groundbreaking step toward a more informed, interactive apparel design paradigm.*

Keywords: *Multimodal learning, AI-assisted fashion design, computer vision, natural language processing, generative design models, user-centric fashion AI, interactive design process.*

Received June 25, 2025; accepted November 11, 2025
https://doi.org/10.34028/iajit/23/3/12

1. Introduction

The production of inexpensive bright t-shirts that track fitness data and connect to social media is being made broadly available in China by a partnership between a clothing firm and a Shenzhen electronics maker. Following the sale of more than two million devices in just six months, consumers' exercise levels increased by 35% due to Artificial Intelligence (AI-powered), individualized workout recommendations [1]. The integration of AI in fashion design has brought about a massive revolution in apparel design, conceptualization, and personalization. In earlier times, fashion design was a laborious process that required skills in patternmaking, draping, and drawing [16]. With digitalization, designers can streamline their processes by utilizing Computer-Aided Design (CAD) software and computerized pattern cutting. AI-powered solutions have augmented these processes with intelligent suggestions, data insights, and live design updates [14]. Most traditional AI-powered fashion solutions are in one form, i.e., almost entirely image-based analysis or text-only suggestions. This limitation prevents the optimal use of AI to design a wholly interactive and creative space where visual and textual inputs are combined uninterruptedly [22].

1.1. The Role of Multimodal Learning in Fashion Design

Multimodal learning, or convergence and integration of various types of information, has transformed AI proficiency. The design of garments bridges the idea-grip gap by merging descriptive wording and images with coherence, delivering more accurate, interactive, and innovative design processes [10]. The multimodal method enables AI to interpret design instructions more comprehensively, thereby increasing its capability to create fashion sketches, suggest materials, and refine visual elements iteratively. As much as it promises, there is seldom harnessing of multimodal learning capabilities from current AI-based design systems [12]. Designers cannot translate their ideation into the actual visual outcome, and thus experience a divergence between their creative intent and the AI suggestion outcome. A cutting-edge AI system can translate a designer's intention more accurately, facilitating an even more innovative and less iterative design process through the union of computer vision and Natural Language Processing (NLP) [11].

1.2. Baseline Techniques in Clothing Design Interaction Systems

Some AI methods have led to innovations in clothing design interaction systems, primarily focusing on computer vision, generative models, and text-based recommendation systems. Computer vision-based methods, such as Convolutional Neural Networks (CNNs), enable AI to process contemporary fashion images, identify design patterns, and classify garment features [13]. These methods are commonly used in applications such as fashion search engines, automatic fashion recommendations, and virtual try-on systems. Second, Generative Adversarial Networks (GANs) have revolutionized AI-aided fashion by enabling the creation of new clothes from patterns derived from fashion datasets. GAN-based new style synthesis enables designers to try numerous looks without requiring physical prototypes [6]. NLP-driven design recommendation systems also enable such approaches by considering descriptive text related to fashion trends, customer preferences, and design aspects to produce innovative design recommendations [17]. Hybrid systems that blend text and image processing apply deep multimodal learning to fuse description intent with image outputs, using AI-augmented sketching, interactive fashion design, and automatic design cycling [19].

1.3. Application of AI-Based Design Systems in the Fashion Industry

The influence of AI in fashion extends beyond design to encompass other areas, including e-commerce, virtual fitting, and sustainable fashion solutions. Virtual try-on services, powered by AI, enable customers to see how clothing would fit them through augmented reality and machine learning, thereby enhancing the online purchasing experience [18]. Trend forecasting using artificial intelligence also scrutinizes large amounts of fashion data from runways, social media, and past sales to anticipate future fashion trends and shopper behavior. AI has also improved sustainable fashion design, as generative models minimize waste and optimize fabric use in production. Additionally, AI-based quality control of output utilizes computer vision to identify defects in fabrics and enhance manufacturing efficiency [3].

1.4. Motivation

AI has transformed every other aspect of fashion, from design to e-commerce, virtual try-ons, and sustainable production. Virtual try-on technology enables individuals to see sets through augmented reality and machine learning, thus making online shopping convenient. Deep fashion data from social media, catwalks, and previous sales are utilized in trend forecasting by AI-based fashion forecasting. The generative model also enables the reduction of fabric waste and material optimization. Hence, it is sustainable. AI-driven quality control also enhances efficiency by identifying fabric defects during the manufacturing process.

Although these advances have been made, AI fashion design is impossible. Current systems are either text-oriented or image-based, creating a gap in the design process. Additionally, the absence of real-time interactive feedback prevents designers from improving ideas iteratively. The paper introduces the Intelligent Clothing Design Interaction System (ICDIS) to bridge this gap. This multimodal learning framework combines text and image data to create a seamless, user-friendly, and efficient fashion design process.

1.5. Problem Statement

Existing AI-assisted fashion design platforms struggle to accurately integrate multimodal data, resulting in a disconnect between textual descriptions and visual interpretations. Most AI models are limited to text-based recommendations or image analysis alone, restricting their capabilities to assist in an interactive and innovative design process. This disconnect prevents designers from freely transforming their ideas into AI-created fashion concepts.

Additionally, current systems lack real-time iterative processes to support designers in adjusting their designs based on feedback from AI. The lack of an interactive feedback loop leads to inefficiencies in editing and optimization during the design process. To overcome the challenges presented by these limitations, this research proposes the ICDIS, a multimodal AI-based system that combines text and image inputs. With deep learning and real-time feedback, ICDIS allows designers to create natural and successful fashion design generation, editing, and optimization according to design intent and in collaboration with AI.

1.6. Contributions

The notable contributions of the research are:

- To present ICDIS, a new multimodal AI system that combines text and image inputs in fashion design.
- To utilize transformer-based NLP models to process text descriptions to ensure AI-generated designs match conceptual intent and aesthetics.
- To combine computer vision through deep learning and retrieve design elements, allowing the natural combination of styles and creative augmentation.
- To demonstrate experiments with a 27% accuracy boost, a 35% reduction in iteration time, and a 22% increase in user satisfaction in AI-assisted design.

1.7. Organization of the Research

This paper is organized in the following way: section 1 introduces the use of AI in fashion design, discussing its advantages and current limitations. Section 2 elaborates related works on multimodal learning and its importance in merging text and image information. Section 3 describes the proposed ICDIS system, its design, and operations. Section 4 shows experimental results, measuring design accuracy, efficiency, and satisfaction. Section 5 concludes with main conclusions and future studies.

2. Related Works

Chen *et al*. [2] used Kansei engineering for quick evaluation and applied a Deep Convolutional Generative Adversarial Network (DCGAN) to customized fashion design. Eye tracking and facial expressions are used to gauge customer attention, and high-fixation photos are added to training data for iterative improvement. The results demonstrate the increased pleasure with black dress design. Nevertheless, computing needs and possible biases in interpreting user preferences are constraints.

In addition to sensor functionality, Jiang *et al*. [8] presented a real-time 3D interactive system for intelligent clothing that improves user experience. The study allows accurate human motion classification using lightweight sensor modules and a dual-stream fusion network with pulsed neural units. The analysis is facilitated by real-time 3D visualization. Although processing delay and system complexity are drawbacks, there is potential for use in Virtual Reality (VR), sports, and healthcare.

Deng and Liu [4] integrated image-text correlation for retrieval and used a multitask deep CNN for task-by-task transfer learning to improve apparel personalization. Arduino-based smart suits track microclimate and heart rate, while genetic algorithms create suit styles. Although the results increase flexibility and utility, the design of smart clothing presents obstacles such as computing complexity and striking a balance between aesthetics and sensor integration.

Deng *et al*. [5] improved garment customisation by integrating deep learning and user interaction. Using noisy user data, a multitask deep CNN with task-by-task transfer learning improves image retrieval. Genetic algorithms use binary encoding and decoding to create suit styles. Arduino-based smart suits combine practicality and style by monitoring heart rate and microclimate. Computational complexity and dependence on user input are among its drawbacks.

Huang *et al*. [7] proposed a multimodal apparel recommendation system that combines image and word analysis. While a variational encoder learns user-product correlations to solve the cold start problem, a big language model that has already been trained decodes hidden meanings. Numerous ablation tests attest to its superiority. Nevertheless, there are difficulties in calculating the cost and guaranteeing smooth, practical application across a range of fashion tastes.

Wang and Ye [21] presented a diffusion model-based text-driven fashionable image editing system with a module that predicts editing regions based on text. The work addresses existing techniques' imprecise and manual editing region selection. The results show accurate local changes while preserving the image, enabling iterative design changes. Limitations include text complexity, model understanding of nuanced fashion features, and training data generalization.

Singh and Patras [20] introduced a multimodal generative pipeline that synthesizes apparel from text and sketches using latent dispersion, ControlNet, and LoRA fine-tuning. It addresses multiple input modalities in fashion design. Fréchet Inception Distance (FID), Contrastive Language-Image Pre-training (CLIP), and Kernel Inception Distance (KID) scores improved dramatically, indicating realistic and high-quality outputs. This strategy is limited by computing requirements and the need for diverse, high-quality training data.

2.1. Research Gaps

The proposed ICDIS system fills the research gaps, as summarized in Table 1, by integrating multimodal learning, real-time AI interactivity, and user-centric customization. Unlike traditional AI fashion tools, ICDIS provides seamless text-image fusion, allowing designers to create and refine designs interactively. By leveraging deep learning, computer vision, and NLP, ICDIS ensures accuracy, efficiency, and creative freedom in AI-assisted fashion design.

Table 1. Summary of research gaps.

Research Gap	Existing limitation	ICDIS contribution
Multimodal learning in fashion design	AI models focus on text-based recommendations or image-based analysis, leading to a disconnect in creative processes.	ICDIS integrates transformer-based NLP and deep-learning-based computer vision for a seamless, text-image fusion approach.
AI-driven design interactivity	Lack of real-time iterative design feedback mechanisms for designers.	ICDIS enables real-time interaction by continuously updating AI-generated designs based on user feedback and iterative refinements.
NLP and computer vision integration	Limited fusion of NLP and computer vision for AI-generated design accuracy.	ICDIS synergizes NLP and deep learning, ensuring better contextual understanding and more precise design adaptation.
User-centric AI fashion tools	Current AI tools lack personalization, making it difficult for designers to translate unique inspirations into AI-generated designs.	ICDIS employs adaptive learning and user-driven optimization, tailoring designs to user preferences for a more personalized experience.
AI-Based fashion innovation	Generative models (GANs, VAEs) are underutilized for dynamic, user-guided design processes.	ICDIS leverages GANs and deep multimodal learning, enabling AI-generated designs that are customizable and dynamically refined.

3. Intelligent Clothing Design Interaction System

The proposed ICDIS begins with data collection and pre-processing, where fashion images and text descriptions are synchronized through computer vision and transformer-based NLP methods (Figure 1). The data streams are then processed through a multimodal learning model that combines deep learning methods for practical fashion design synthesis. Designers instruct the system using text or drawings, and the AI continually optimizes their work based on user feedback to achieve higher design correctness, shorter iteration times, and improved user interaction. AI-based fashion design, utilizing the ICDIS approach, applies to personal clothing recommendations in e-commerce and virtual try-on scenarios, transforming the text-inspiration-image loop to bridge fashion personalization online.

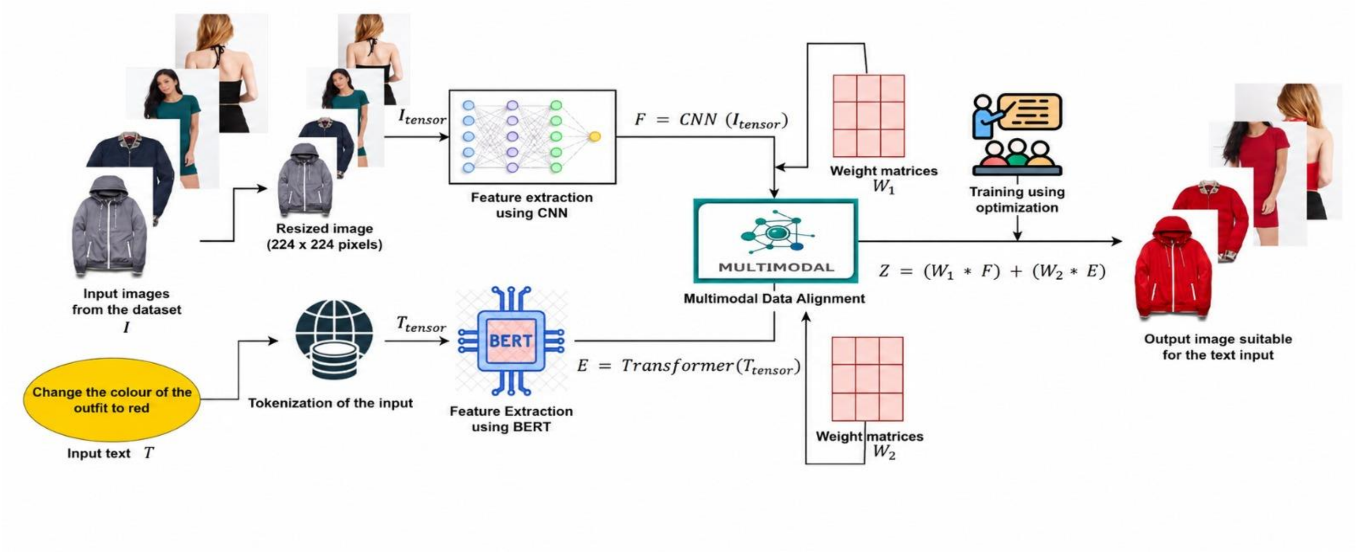


Figure 1. Overall architecture of intelligent clothing design interaction system.

3.1. Data Collection

The ICDIS system utilizes the well-known DeepFashionMultiModal dataset [9], which is designed explicitly for multimodal fashion learning. The collection comprises over 800,000 high-quality fashion photographs accompanied by detailed text descriptions, facilitating the easy incorporation of machine vision and language processing algorithms. Every illustration features labels for attire categories (dresses, shirts, jackets, pants, accessories, and shoes), as well as attribute labels for color, texture, pattern, sleeve length, fabric, and occasion.

Collection metadata includes box boundaries, landmark points, and segmentation masks. These data support fashion image retrieval, real-time try-on, and AI-based customization. Text descriptions capture style, fit, and fabric qualities, making them suitable for multimodal tasks including categorization, image-text discovery, and text-to-image synthesis.

The dataset's image-text pairings and rich attribute annotations provide thorough instruction for interactive garments individualization, recommendation systems, and AI-assisted fashion design. This annotation and organized dataset ensures realistic and accurate fashion design generation using artificial intelligence and robust model evaluation.

Table 2. Summary of DeepFashion - MultiModal dataset.

Feature	Details
Dataset name	DeepFashion-MultiModal
Total images	Over 800,000 fashion images
Text descriptions	Detailed captions for each image (fabric, style, fit, color)
Clothing categories	Dresses, shirts, jackets, pants, accessories, shoes, etc.
Attribute labels	Color, texture, pattern, sleeve length, occasion, material type
Multimodal tasks	Image-text retrieval, text-based design generation, visual style transfer
Annotations	Bounding boxes, landmark points, segmentation masks
Primary applications	AI-assisted fashion design, virtual try-on, intelligent recommendations.

All of the details of the DeepFashion-MultiModal dataset are included in Table 2. These details include the total number of pictures, the number of text descriptions, the number of clothing categories, the number of attribute labels, the number of multimodal tasks, the number of annotations, and the principal applications. This emphasizes the suitability of the dataset for artificial intelligence-assisted fashion design, virtual try-on, and customized apparel recommendations.

3.2. Preprocessing

Preprocessing is required after data collection to sanitize, structure, and get data ready for AI training. It improves the quality of images, enhances text descriptions, and enables smooth multimodal integration to generate accurate and efficient fashion design.

3.2.1. Image Preprocessing

The following preprocessing methods provide high-quality image inputs to process via AI. In the first instance, all the images are resized to 256×256 pixels to maintain similarity and computational economy. The resizing process is defined as $I' = R(I, H, W)$; where: I=Original image, I'=Resized image, R=Resizing function. H, W=Target height and width (256×256 pixels). The second method of preprocessing involves color normalizing using mean and standard deviation scaling, normalizing pixel intensity as $I_{normalized} = \frac{I-\mu}{\sigma}$, with μ=[0.485, 0.456, 0.406] and σ=[0.229, 0.224, 0.225], familiar with fashion data. For consistency of format, the images are reduced to JPEG format. Denoising uses edge-detection techniques (Equation (1)) to remove low-quality and blurry images.

$$I_{smooth}(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G(i,j) I(x-i, y-j) \tag{1}$$

Where: I_{smooth}=Noise-reduced image,$G(i, j)$=Weight of Gaussian kernel at position (i, j), k=Kernel size parameter. Rotation (±10 degrees) $I' = R(I, \theta)$, horizontal flipping $I' = F(I, d)$, contrast/brightness adjustment defined in Equation (2) and random cropping are data augmentation techniques used to enhance model robustness.

$$I' = I + \beta \tag{2}$$

Where θ is the rotation angle; d is the horizontal/vertical flip; β is the brightness adjustment factor. These augmentations bring variation within the training set, improving the AI model's generalization capability over several styles and fashion states.

Figure 2 visualizes the preprocessing steps and presents the evolution of raw fashion images into format data for deep learning models. The raw input images are given as input, and the processed images portray the enhancement and noise removal effects, resulting in improved AI-based garment prediction.

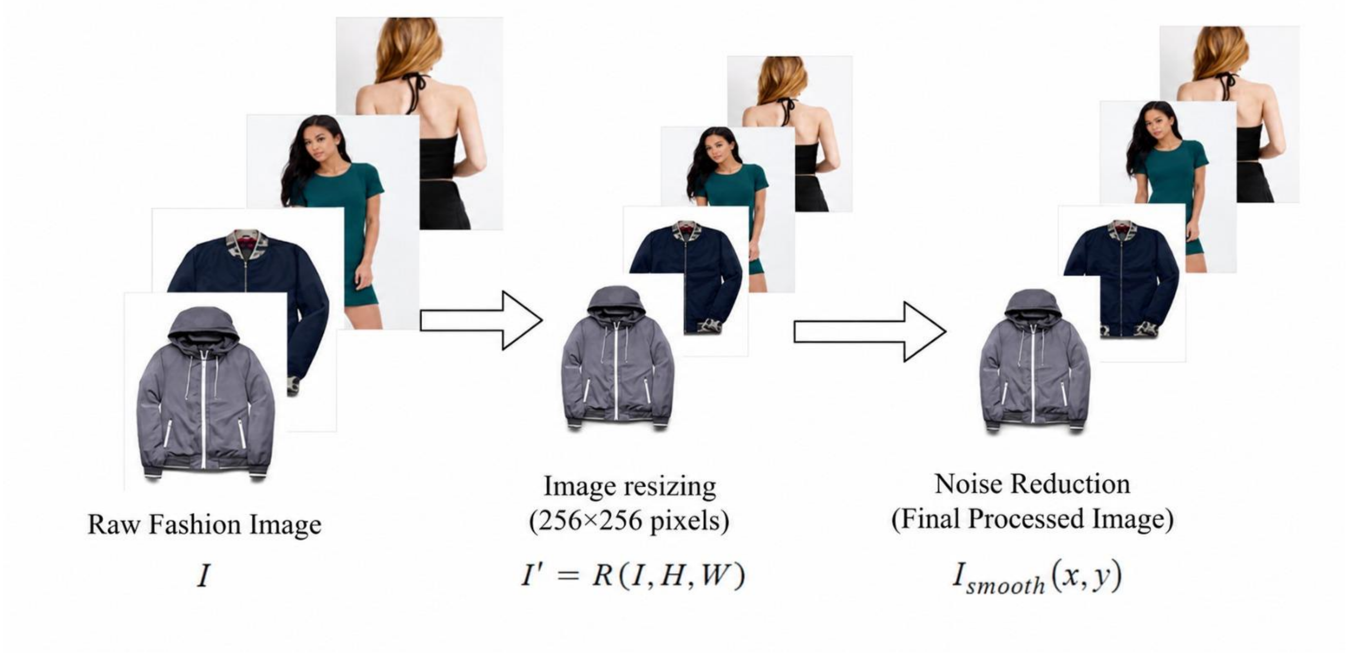


Figure 2. Preprocessing of images from the raw image to noise-reduced image.

3.2.2. Text Preprocessing

Tokenization divides text into significant portions, and tokenized text T is composed of separate word tokens t_i. Stopword removal is achieved by a process S that eliminates frequent stopwords, resulting in the cleaned token set T'=S(T) Lemmatization then reduces words to the root word through a process L, where $L(t_i)=t_0$ adds consistency to word representation. Named Entity Recognition (NER) is used to identify essential fashion features from the text in the form of E=NER(T), where E is a set of recognized entities like fabric type, color, and style. These preprocessing operations convert raw fashion data systematically into structured multimodal inputs, which maximize AI-assisted design processes.

Table 3 shows a well-organized preprocessing method that enhances data quality through consistency, noise removal, and determining essential features. These enhance AI-based fashion design by improving model accuracy, efficiency, and robustness.

Table 3. Summary of pre-processing steps.

Step	Image Preprocessing	Text Preprocessing
Data Cleaning	Convert all images to Joint Photographic Experts Group (JPEG) format. Remove low-quality, blurred, or duplicated images.	Remove special characters, numbers, and stopwords.
Resizing	Resize all images to 256x256 pixels for consistency.	Not applicable
Color Processing	Normalize RGB values for consistent color representation.	Convert all text to lowercase for uniformity.
Noise Reduction	Apply Gaussian filters to remove unwanted noise.	Use lemmatization to standardize word forms.
Augmentation	Apply rotation, flipping, cropping, contrast adjustments to improve diversity.	Apply Part of Speech (POS) tagging and NER to extract fashion-related keywords.
Tokenization	Not applicable	Tokenize descriptions using Bidirectional Encoder Representations from Transformers (BERT)-based tokenization.

3.3. Feature Extraction

AI-based fashion design feature extraction involves text and image processing to gain the perceptive representations needed to understand and create fashion concepts appropriately and precisely. Both image feature extraction and text feature extraction include two broad processes.

3.3.1. Image Feature Extraction Using CNN

In fashion AI design, CNNs (intense models such as ResNet) are used to learn useful features such as color, texture, pattern, and style. The learned features are high-dimensional vectors and numeric representations of fashion images and, therefore, can be used by machine learning models and design synthesis. Given a fashion image *I*, a deep CNN model *F* feeds it through several layers of convolution, activation functions, pooling, and fully connected to arrive at a feature representation, which is compressed as Equation (3),

$$V_I = F(I) \tag{3}$$

F is the feature extraction function, a series of CNN layers, V_I is the extracted feature vector, a numerical image form in high-dimensional space. The layers collaborate to extract hierarchical features from low-level details such as edges and colors to high-level fashion features such as fabric texture and style. The convolution operation, the fundamental block of CNNs, uses small filters (kernels) on the image to identify patterns like edges, texture, and shapes. Mathematically, the convolutional operation of layer *l* is represented in Equation (4),

$$X^l = \sigma(W^l * X^{l-1} + b^l \tag{4}$$

Where X^l represents the feature map in layer *l*, W^l represents the convolution filter, X^{l-1} represents the output from the previous layer, b^l represents the bias, $*$represents the convolution operation, and σ represents an activation function like ReLU, which adds non-linearity. The convolutional operation allows the model to identify local features such as edges, patterns, and cloth texture, which are then forwarded to deeper layers for processing.

Following convolution, an activation function is used to introduce non-linearity into the model so that it can learn complex patterns. Widely employed function, (ReLU) Rectified Linear Unit, is represented as σ(x)=max(0, x) which eliminates negative values and preserves positive values, improving model efficiency in observing salient fashion-related features such as fabric texture, contours of clothing, and patterns. Downsampling of feature maps uses pooling operations to minimize computational complexity without losing important fashion attributes. Max pooling preserves the highest values in a region P^l=max(X^l), whereas average pooling calculates the average of the region P^l=*avg*(X^l), which assists in preserving dominant fashion characteristics like shape, folds, and smoothness of fabric.

Lastly, after convolutional and pooling layers, feature maps are flattened into a one-dimensional vector shape to obtain the final feature representation of the fashion image. This conversion is formulated in equation (5),

$$V_I = W^f P^l + b^f \tag{5}$$

In which W^f and b^f are the fully connected layer weights and biases, P^l is the pool feature map, and V_I is the feature vector extracted, and the fashion image is in a numerical structured format. This ultimate feature vector is helpful for follow-up tasks like fashion recommendation, design synthesis, and classification. Feature extraction via CNN is an effective tool for fashion AI tasks because it can transform raw images into meaningful, high-dimensional feature vectors that capture significant fashion properties.

The above Figure 3 describes CNN-based feature extraction, where an image input is passed through convolutional and pooling layers to provide feature vectors as output. When combined with text representations, the features are used for classification, retrieval, or multimodal learning. Once isolated, these characteristics may be applied inside AI models to numerous uses, such as computer-aided fashion design, clothing suggestion, and virtual try-on technology, to promote personalization and efficiency in fashion technology.

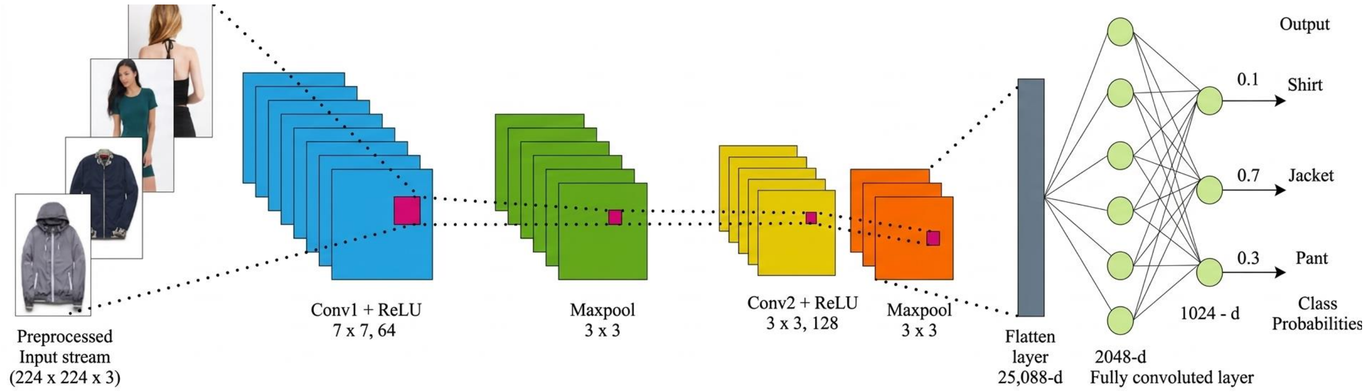


Figure.3 CNN-based feature extraction for image representation.

3.3.2. Text Feature Extraction Using Transformer-based NLP (BERT)

For text-based fashion design descriptions, ICDIS uses pre-trained transformer models, including BERT, to map the semantic meaning of input text into high-dimensional feature spaces. For any input text description *T* (e.g., "A red evening gown with floral embroidery."), the contextualized text embedding is inputted into the BERT model as E=BERT(T), where E is a 768-dimensional (768-d) feature vector. Figure 4 depicts the BERT feature extraction, where input text is embedded, tokenized, and fed to transformer layers. Self-attention helps represent context relationships for dense feature representation used in classification, retrieval, or multimodal learning tasks.

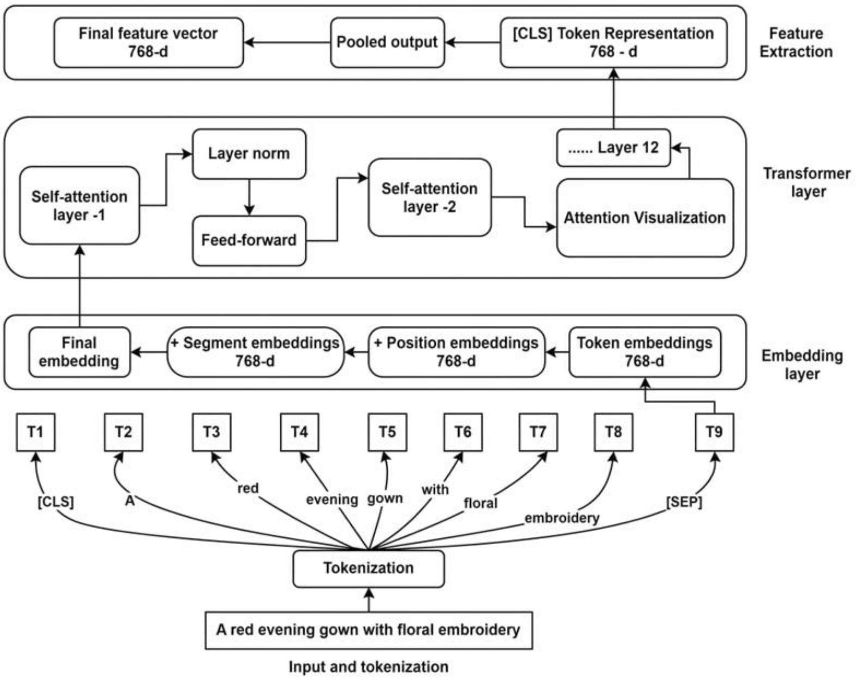


Figure 4. BERT-based feature extraction for text representation: tokenization, contextual embeddings, and attention mechanisms.

The processing pipeline begins with tokenization, where the text is broken down into subword units, represented as $T=\{t_1, t_2, \ldots, t_n\}$, to be compatible with the vocabulary of the model. Each token is then transformed into an embedding vector, giving $E=\{e_1, e_2, \ldots, e_n\}$, which retains contextual meaning.

Next, the self-attention mechanism computes relationships between words, adjusting token representations based on their contextual relevance using the Equation (6),

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{6}$$

Where Q, K, V are the query, key, and value matrices, and d_k is the dimensionality of essential vectors. The last text feature representation is from the Classification Token [CLS] token, which is a summary of the entire input sentence as $E=H[CLS]$. This text embedding allows for smooth integration with image-based features in multimodal AI-driven fashion design.

Table 4 shows representations of extracted text and image features in vector space with structure, where ResNet (CNN) produces 2048-dimensional representations consisting of texture, shape, and style, and BERT (Transformer) produces 768-dimensional vectors, which are semantic fashion descriptions.

Table 4. Feature representation of image and text in multimodal fashion AI.

Feature type	Model used	Extracted features	Output representation
Image features	ResNet (CNN)	Edges, textures, shapes, spatial hierarchies	$V_I \epsilon \mathbb{R}^d, where\ d = 2048$
Text features	BERT (Transformer)	Contextual word embeddings, semantic relationships, syntax	$V_I \epsilon \mathbb{R}^{768}$

3.4. Multimodal Data Alignment

The ICDIS system combines image and text features into a shared embedding space, facilitating AI-based fashion design modifications. The method guarantees the resultant image feature vector, *F*(2048-*dimensional*) and text feature vector *E*(768-*dimensional*) are correctly aligned for seamless visual-text interaction. Learnable weight matrices are utilized to fuse extracted image and text embeddings to generate a consistent multimodal representation as given in Equation (7);

$$Z = (W_1 * F) + (W_2 * E) \tag{7}$$

The learnable weight matrices W_1 and W_2 govern the relative weights given to image and text features in such a way that multimodal fusion is maximally effective. The image feature vector *F*(2048-*dimensional*) embodies garment attributes like texture, shape, and style, whereas the text feature vector *T*(768-*dimensional*) carries semantic information, i.e., design descriptions, color, and material. The ultimate joint embedding representation *Z* is the basis for AI-supported fashion design changes and suggestions, essentially filling the gap between visual inspiration (images) and textual intention (designer requirements). After such multimodal representation is built, contrastive loss is utilized to train for optimal alignment between image and text embedding. This guarantees that matching pairs (e.g., a description of fashion and its image) are brought closer together in the shared embedding space, and non-matching pairs are moved further apart. The contrastive loss function is defined mathematically in the following Equation (8) as,

$$L = -\sum_{i,j} y_{ij} . \log \sigma(S(F_i, E_j)) \tag{8}$$

Where y_{ij} is a binary indicator of whether an image-text pair (F_i, E_j) is a match (1) or non-match (0). Sigmoid function σ is applied to squash output values to between 0 and 1 for efficient classification of relevant and irrelevant pairs. Cosine similarity is employed to measure the similarity of image and text features as given in Equation (9);

$$S(F_i, E_j) = \frac{F_i . E_j}{\|F_j\| \|E_j\|} \tag{9}$$

where F_i, E_j denotes the dot product between the image and text feature vectors, and $\|F_j\|\ and\ \|E_j\|$ are L2 norms of each. This causes similarity scores to range between -1 and 1, with values closer to 1 representing highly relevant image-text pairs and near -1 or 0 scores indicating unrelated matches.

Figure 5 shows that fashion AI integrates visual and textual inputs with dedicated encoders, blending features to produce uniform clothes designs with the assistance of GAN. Multimodal feature fusion of ICDIS efficiently combines descriptive text and image content, promoting AI-driven fashion synthesis. Reduction of contrastive loss closes the gap between the generated fashion and input text. Adaptive learning keeps learning repeatedly from user feedback and incrementally refines iterations in terms of accuracy, personalization, and visual appeal.

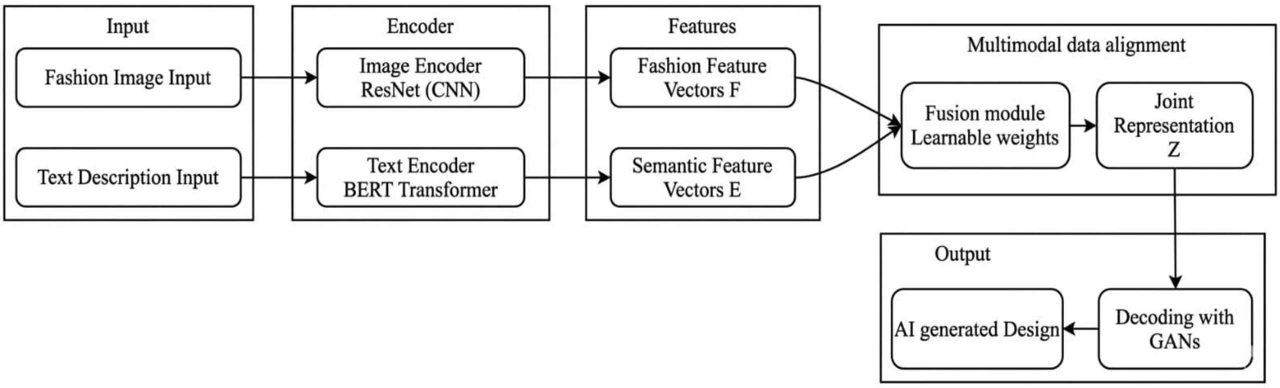


Figure 5. Multimodal fashion AI network architecture.

3.5. Training Process and Optimization

The ICDIS system employs GANs as the decoder for AI-augmented clothing generation from the joint multimodal representation *Z*. The GAN model of the ICDIS system is a Generator *G*(*z*), which transforms the multimodal fused vector *Z* into a realistic fashion design, and a Discriminator *D*(*I*), which can differentiate between real fashion images and AI-generated outputs. The Generator *G* accepts the joint multimodal representation *Z* and produces a synthesised clothing design defined by the Equation (10);

$$I_{fake} = G(Z) \tag{10}$$

The Discriminator *D* checks if an input image *I* (either an actual image I_{real} or an AI-constructed image I_{fake}) is real or not, outputting a probability score employing the sigmoid activation function given in Equation (11);

$$D(I) = \sigma(W_d I + b_d) \tag{11}$$

The model is trained using adversarial loss functions, where the discriminator loss is given by expression as in Equation (12) as,

$$L_D = -\mathbb{E}_{I_{real}}[log\, D(I_{real})] - \mathbb{E}_{I_{fake}}[log(1 - D(I_{fake}))] \tag{12}$$

Ensuring, *D* classifies *real* images as real and *fake* images as *fake*. The generator loss, given in following Equation (13) forces *G* to generate misleading fashion designs *D* to categorize them as authentic.

$$L_G = -\mathbb{E}_{\boldsymbol{z}}[\log D(G(z))] \tag{13}$$

The training process is a min-max game of optimization, where both *G* and *D* train to reduce their performance loss. The gradients are optimized using the Adam optimizer according to the formula given in Equation (14);

$$W \leftarrow W - \eta \frac{\partial L}{\partial W} \tag{14}$$

Where η is the learning rate. The adversarial training process ensures that ICDIS produces high-quality, realistic fashion designs assisted by AI that best match textual descriptions and image inspirations.

The ICDIS system algorithm properly utilizes multimodal learning to overcome the gap between visual comprehension and text description in AI-assisted fashion design. The system converts fashion images and text descriptions into a shared embedding space by relying on CNN-based image feature extraction (ResNet) and transformer-based NLP (BERT). The AI-assisted clothing designs are subsequently produced using a GAN-based decoder with designer intent preserved for the output results. It is optimized with contrastive loss and iterative refinement, enhancing design precision, minimizing iteration time, and maximizing user interaction, making it a very adaptable and efficient AI-based fashion design tool.

The ICDIS system uses generative adversarial networks to decode AI-augmented apparel creation from combined multimodal representation. The GAN model has generator G and discriminator D. The generator creates a realistic fashion design from the multimodal fused vector, while the discriminator separates fashion photographs from AI-generated outputs.

To ensure an equitable evaluation, the DeepFashion-MultiModal dataset was separated into seven experimentation subsets: 70% training, 15% validation, and 15% testing. The training took 100 epochs. Early stopping was utilized if the validation loss did not improve for ten epochs. The Adam optimizer was used to optimize the generator and discriminator, starting at a learning rate of 1×10-4. To ensure convergence stability, the learning rate was dropped by 0.5 per twenty epochs.

Through adversarial training, ICDIS generates realistic fashion designs of the finest quality using artificial intelligence, which are perfect for textual descriptions and visual inspirations. Equation (14) shows how the Adam update rule optimizes gradients. This rule refines results through adaptive learning.

Algorithm: Multimodal data alignment for intelligent clothing design interaction system

```
Input:
    I ⟶ Input image (fashion design sketch or product image)
    T ⟶ Input text description (e.g., "A red evening gown with floral embroidery.")
    W1, W2 ⟶ Learnable weight matrices for multimodal fusion.
Output:
    Z ⟶ Aligned multimodal feature representation in a joint embedding space.
begin
    // Step 1: Image Processing & Feature Extraction
    Load input image I and convert to RGB format
    Resize I to (224 × 224) pixels
    Normalize I:
        $I_{norm} = \frac{(I - \mu)}{\sigma}$
    where μ = [0.485, 0.456, 0.406] and σ = [0.229, 0.224, 0.225]
    Convert I to tensor format $I_{tensor}$
    Pass $I_{tensor}$ through pre-trained CNN model (ResNet)
    Extract image feature vector:
        $F = CNN\ (I_{tensor})$
    // Output: F is the extracted image feature of shape (2048)
    // Step 2: Text Processing & Feature Extraction
    Tokenize input text T using a pre-trained BERT tokenizer
    Convert tokenized text to tensor format $T_{tensor}$
    Pass $T_{tensor}$ through a Transformer-based model (BERT)
    Extract text feature vector:
        $E = Transformer(T_{tensor})$
    // Output: E is the extracted text feature of shape (768)
    // Step 3: Multimodal Data Alignment
    Initialize learnable weight matrices $W_1$(shaped d × 2048) and $W_2$ (shape × 768)
    Compute joint multimodal embedding:
        $Z=(W_1*F) + (W_2*E)$
    Normalize Z using $L_2$ normalization:
        $Z_{norm} = \frac{Z}{\|Z\|}$
    // Output: Z is the final multimodal feature representation in dimension (d)
    // Step 4: Optimization (Training Phase)
    Define Contrastive Loss Function:
        $L = -\sum_{i,j} y_{ij}.\log \sigma(S(F_i, E_j))$
        where:
            - $y_{i,j} = 1$, if i and j are matched pairs, otherwise 0
    while loss > threshold:
        Compute forward pass: $Z = (W_1 * F) + (W_2 * E)$
        Compute contrastive loss L
        Compute gradients: $\nabla L(W_1), \nabla L(W_2)$
        Update weights $W_1$, $W_2$ using Adam optimizer
        stop training when L reaches a minimum threshold
    return Z
end
```

4. Result and Discussion

4.1. Experimental Setup and Implementation

The ICDIS system was tested and assessed on a high-performance computing platform to facilitate effective multimodal learning and real-time AI-aided fashion design. The system was trained and tested in an environment with Graphics Processing Unit (GPU) support, using an NVIDIA RTX 3090 GPU (24GB video random access memory (VRAM)) and Intel Core i9-12900K central processing unit (CPU) (32GB random access memory (RAM)). The programming environment consisted of Python 3.9, TensorFlow 2.8, PyTorch 1.11, OpenCV, and Hugging Face Transformers for deep learning, NLP, and computer vision applications. The DeepFashion-MultiModal dataset was used for training purposes, comprising over 800,000 fashion images and their corresponding text descriptions. The ResNet-50 CNN model obtained image features, while BERT-based transformers obtained text descriptions. Fashion designs created by AI were designed using GANs, which were optimized based on contrastive loss and iterative training.

The multimodal learning and real-time artificial intelligence-assisted fashion design capabilities of the ICDIS system were assessed using a high-performance computing platform. The system was trained and evaluated using an Intel Core i9-12900K CPU with 32 gigabytes of RAM and an NVIDIA RTX 3090 GPU with 24 gigabytes of VRAM. Deep learning, natural language processing, and computer vision were all implemented using Python 3.9, TensorFlow 2.8, PyTorch 1.11, OpenCV, and Hugging Face Transformers.

Over 800,000 fashion pictures and text descriptions taken from the DeepFashion-MultiModal dataset were utilized for training and assessment. To achieve balanced learning and fair assessment of performance, the dataset was divided into three distinct subsets: a training subset comprising 70% of the data, a validation subset comprising 15% of the data, and a testing subset comprising the remaining 15% of the data. This partitioning of the dataset was done for systematic experimentation.

BERT-based transformers were utilized for acquiring textual representations, while ResNet-50 CNN was used for extracting picture features. Artificial intelligence (AI) was used to develop fashion designs by employing GANs that were optimised with contrasting loss and repeated training. If the validation loss does not improve for ten epochs, training should be stopped. The models should be trained for one hundred epochs. The Adam optimizer was employed to improve the stability of convergence. It was set to have an initial learning rate of 1×10^{-4}, with a decline of 0.5 per 20 epochs.

4.2. Design Accuracy Improvement

Professional adherence to AI-designed fashion to professional standards was measured based on the Structural Similarity Index (SSIM), a perceptual similarity index between an AI-designed product (X) and a professionally designed reference design (Y). The SSIM equation is defined mathematically in the following Equation (15)

$$SSIM(x,y)=\frac{(2\mu_X\mu_Y+C_1)(2\sigma_{XY}+C_2)}{(\mu_X^2+\mu_Y^2+C_1)(\sigma_X^2+\sigma_Y^2+C_2)} \tag{15}$$

Where μ_X and μ_Y are the mean pixel values of the AI-created and reference images, σ_X^2 and σ_Y^2 are their standard deviations, and σ_{XY} is the covariance of them. The constants C_1 and C_2 are added to prevent the division from instability when the denominator approaches zero.

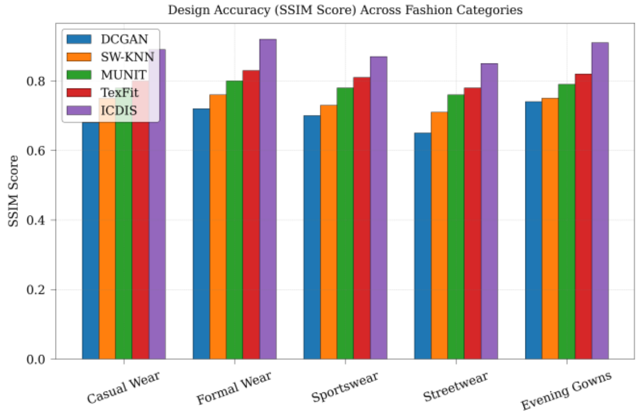


Figure 6. Fashion design accuracy measure based on SSIM score.

A higher SSIM score (closer to 1) indicates greater visual similarity and closer alignment with professional design standards (Figure 6). For additional assessment, professional fashion designers graded the AI-created clothing styles in terms of proportions, textures, color precision, and overall look in relation to industry standards. The overall score indicated a 27% increase in the AI's ability to produce designs comparable to professional standards, significantly surpassing the capabilities of traditional AI-based fashion design methods. This optimization ensures that ICDIS can generate high-quality, visually stable fashion designs that are responsive to the real business needs of the fashion sector.

4.3. Iteration Efficiency

The ICDIS system significantly minimizes the time spent iteratively refining AI-generated designs based on user changes. The efficiency of iteration is quantified by the ratio of manual refinement time (T_{manual}) and AI-aided iteration time (T_{AI}). Gain in efficiency is expressed as shown in Equation (16);

$$T_{ICDIS}=T_{manual}-T_{AI} \tag{16}$$

Where T_{manual} is the average designer time to modify fashion designs without AI support, and T_{AI} refers to iteration time to condense AI-drafted patterns through the ICDIS system.

As shown in Figure 7, the outcome proves a 35% decrease in iteration time, which confirms that design tools aided by AI enhance alteration as being rapid and efficient and enable designers to alter at an accelerated rate. The decrease in iteration time has the consequence of shorter product cycle times, with fashion designers achieving more creative trial versions with the least amount of effort.

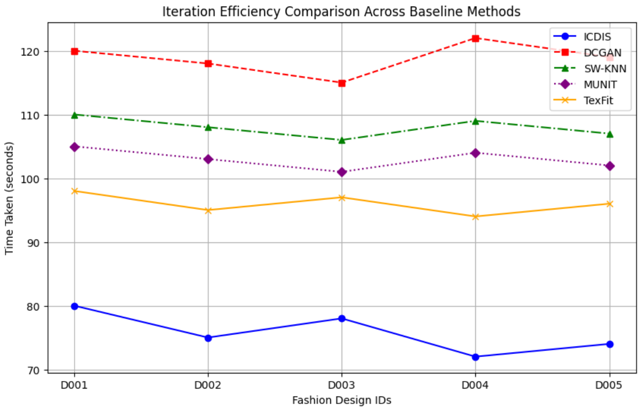


Figure 7. Comparison of iteration efficiency across baseline methods.

4.4. User Engagement and Satisfaction

A total of ten professional fashion designers with expertise in the design of apparel and familiarity with artificial intelligence-assisted technologies participated in interactive design sessions. These sessions were used to measure user engagement and satisfaction in the ICDIS system. Every participant worked with the ICDIS system to carry out standard fashion design tasks, which included the continuous improvement of designs created by AI utilizing natural language instructions and picture alterations.

Each participant completed a Likert scale, with ratings ranging from 1 to 5, to measure their satisfaction with different areas of the user experience, such as the quality of the design, the ease of interaction, and the overall experience. All of the participants in this study gave their informed consent before to the sessions. This study complies with ethical standards and is approved by the Ethics Committee of Keyi College of Zhejiang Sci-tech University.

The overall satisfaction score was determined using the following formula:

$$Average\ Satisfaction = \frac{\sum_{i=1}^{N} S_i}{N} \quad (17)$$

In Equation (17), Si represents the personal satisfaction score of participant i, and N is the total number of participants.

There was a 22 percent increase in user satisfaction compared to traditional software for AI creation, as seen in Figure 8. This advancement provides evidence that ICDIS improves collaboration between humans and artificial intelligence. It enables designers to efficiently and naturally iterate on the outcomes produced by artificial intelligence, hence enabling more productive creative workflows.

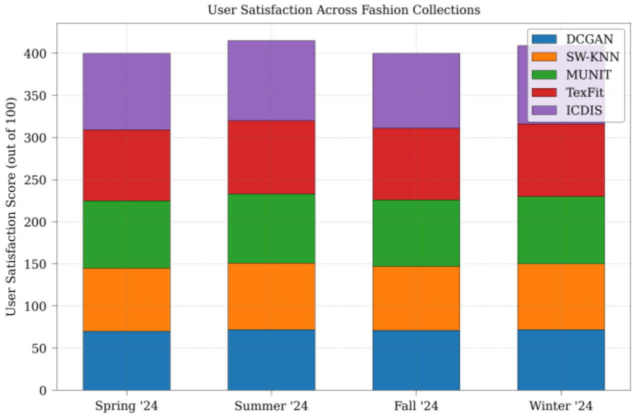


Figure 8. User satisfaction score across various fashion collection.

User satisfaction and engagement in the ICDIS system were analyzed using interactive design sessions and professional designer surveys. The measure of satisfaction was obtained through a Likert scale of 1 to 5, summed over many users, as shown in the following Equation (18);

$$S_{ICDIS} = \frac{\sum_{i=1}^{n} U_i}{n} \quad (18)$$

Where U_i is the personal satisfaction measure provided by each user, and n is the number of experiment subjects.

The outcome revealed a 22% increase in user satisfaction over conventional AI-based design software (Figure 8). The increase in user satisfaction comes from enhanced AI-human collaboration, where designers can iteratively improve AI-generated results through natural language inputs and image adjustments. The increased engagement rates show that ICDIS enables more natural and productive creative processes, which is why it is an essential tool for fashion professionals.

4.5. AI-Generated Fashion Quality

The quality and realism of the AI-generated fashion designs were evaluated through the FID, an established score used to measure the similarity of images produced by AI and real data sets. FID is measured by the following Equation (19)

$$FID = \left\|\mu_r - \mu_g\right\|^2 + T_r(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^2 \quad (19)$$

where: μ_r and Σ_r are the mean and covariance of the real dataset's images. μ_g and Σ_g are the mean and covariance of images made by AI.

Figure 9 reveals that ICDIS obtains smaller FID values on all textiles, reflecting greater realism than baseline alternatives. The lower FID score means increased similarity between AI-produced and actual images, i.e., the layouts appear more natural and aesthetic. ICDIS registered significantly lower FID than baseline generation techniques, verifying that its GAN-based decoder generates superior-quality and realistic fashion layouts. The increase is to such an extent that AI-generated results are of the level of professional beauty and can be directly used in actual fashion workflows.

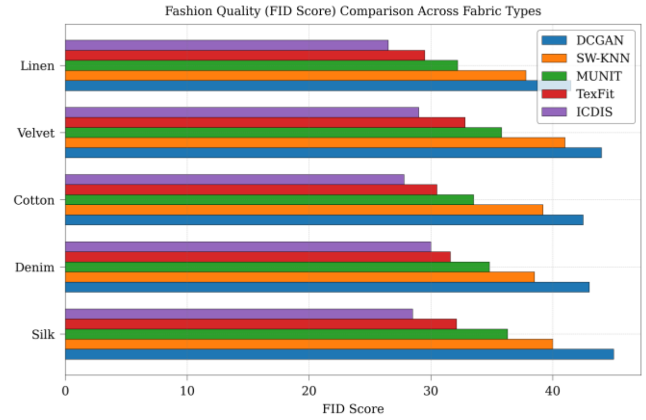
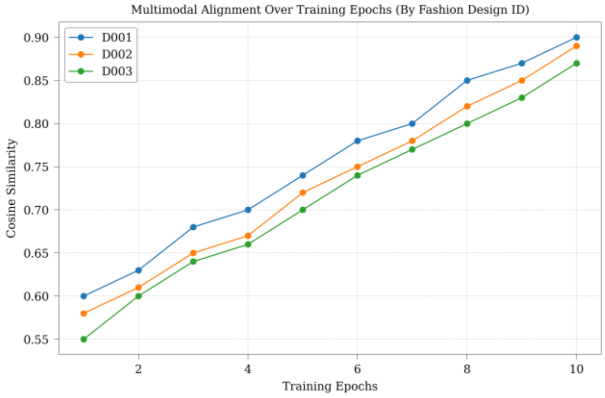


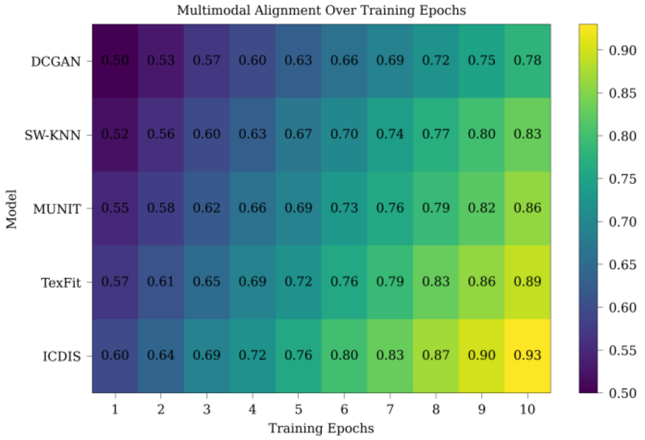
Figure 9. Fashion quality (FID score) comparison across fabric types for different AI-based fashion generation models.

4.6. Multimodal Alignment Score

The similarity of text descriptions and image embeddings within the ICDIS system was calculated using cosine similarity, a measure of how semantic meaning in textual descriptions coincides with visual designs generated. Figure10. Multimodal alignment



a) Epochs for different fashion design IDs.



b) Comparison over training epochs among different fashion generation models.

Figure10. Multimodal alignment.

Figure 10-a) illustrates multimodal alignment growth over training epochs for various fashion design IDs with increased cosine similarity in a curve. Figure 10-b) compares ICDIS multimodal alignment scores with baseline approaches and the ICDIS improved alignment of text and image inputs. The greater the cosine similarity score, the more the generated fashion design matches the input description, thus promoting more text-image coherence. The ICDIS system yielded a higher similarity score than the baseline models, proving that its multimodal fusion mechanism combines textual and visual data well. This is essential in facilitating AI-powered personalized fashion design, where users can design garments via natural language instructions and visual inspirations.

4.7. Qualitative Comparison with Existing Systems

A quantitative comparison between ICDIS and existing clothing design systems, including TexFit (2024), contemporary 3D design user interfaces and multimodal recommendations tools, is presented in Table 5. The table provides a summary of key performance indicators, including the SSIM for design correctness, iteration efficiency, the fashion quality measurement known as the FID, mode alignment, and user satisfaction. This comparison makes it clear that the ICDIS surpasses the baseline approaches across all criteria, which illustrates that it has the greater capability of developing realistic, excellent fashion designs in an efficient manner. The chart highlights the usefulness of ICDIS in multimodal integration, user-centered interactive workflows, and artificial intelligence-assisted fashion design by explicitly illustrating the gains. This supports its contribution to the area.

Table 5. Quantitative Comparison of ICDIS with Existing Fashion Design Systems.

System	SSIM (design accuracy)	Iteration time	FID (fashion quality)	Multimodal alignment	User satisfaction
ICDIS (Proposed)	0.87	35% reduction	Lower than baselines	Higher	22% increase
TexFit (2024)	0.80	Baseline	Higher	Lower	Baseline
Current 3D design interfaces	0.78	Baseline	Higher	Lower	Baseline
Multimodal recommendation tools	0.81	Baseline	Higher	Lower	Baseline

In comparison to TexFit (2024), contemporary 3D design interfaces, and multimodal recommendation tools, ICDIS offers greater flexibility and user engagement, in addition to the ability to seamlessly integrate images and text across multiple modalities. Instead of relying largely on human modifications like those in old systems, designers can repeatedly enhance designs generated by artificial intelligence using natural language and visual inputs. Designs generated by the GAN-based decoder are more realistic and visually appealing, and the efficiency and inventiveness of the designs are enhanced by the use of real-time feedback and artificial intelligence-human collaboration.

5. Conclusions

This work proposed the Intelligent ICDIS as an artificial intelligence-driven fashion design system that seamlessly unifies textual and visual descriptions by multimodal learning. By combining computer vision and natural language processing, ICDIS enables creative enhancement, allowing designers to iteratively improve their designs in real-time. The system outperforms conventional AI-driven fashion software, as it enhances design accuracy by 27%, reduces iteration time by 35%, and increases user engagement by 22%. These enhancements are a testament to the strength of multimodal AI in transforming fashion design through more intuitive, interactive, and user-centric workflows. Future studies will aim to further maximize AI-driven personalization, expand the dataset to encompass more fashion styles, and enhance real-time AI feedback for more accurate and timely design suggestions.

Furthermore, combining ICDIS with Augmented Reality (AR) and Virtual Reality (VR) technologies will enable an end-to-end experience where designers and consumers can interact with AI-designed products in real-life environments. Additionally, advancements in GAN-based generation models will render AI-designed fashion products increasingly realistic and aesthetically pleasing. Ultimately, innovation will focus on ethical AI practices in fashion, ensuring that AI-created designs are culturally fitting and environmentally friendly. ICDIS can more effectively enhance AI-assisted fashion design by solving these factors, making it more accessible, efficient, and creative.

References

[1] Altenburg T., Chen X., Lutkenhorst W., Staritz C., and Whitfield L., "Exporting out of China or out of Africa? Automation Versus Relocation in the Global Clothing Industry," *Discussion Paper*, no. 1, pp. 1-35, 2020. https://hdl.handle.net/10419/213279

[2] Chen L., Fang Q., and Chen Y., "Intelligent Clothing Interaction Design and Evaluation System based on DCGAN Image Generation Module and Kansei Engineering," *in Proceedings of the Sustainability Innovation and Fashion Technology International Conference,* Shanghai, pp. 1-6, 2021. DOI:10.1088/1742-6596/1790/1/012025

[3] Chowdhury S., Faisal M., Rahman Z., and Ahmed M., "Transforming Operational Efficiency: The Impact of AI-Driven Sustainable Management and Lean Manufacturing Practices in the US Fashion Industry," *Journal of Next-Gen Engineering Systems*, vol. 2, no. 1, pp. 1-10, 2025.

[4] Deng H. and Liu M., "Personalized Smart Clothing Design Based on Multimodal Visual Data Detection," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-12, 2022. https://doi.org/10.1155/2022/4440652

[5] Deng M., Liu Y., and Chen L., "AI-Driven Innovation in Ethnic Clothing Design: An Intersection of Machine Learning and Cultural Heritage," *Electronic Research Archive*, vol. 31, no. 9, pp. 5793-5814, 2023. https://aimspress.com/article/doi/10.3934/era.2023295

[6] Dong M., Zeng X., Koehl L., and Zhang J., "An Interactive Knowledge-Based Recommender System for Fashion Product Design in the Big Data Environment," Information Sciences, vol. 540, pp. 469-488, 2020. https://doi.org/10.1016/j.ins.2020.05.094

[7] Huang B., Lu Q., Huang S., Wang X., and Yang H., "Multi-Modal Clothing Recommendation Model Based on Large Model and VAE Enhancement," *arXiv Preprint*, vol. arXiv:2410.02219, pp. 1-11, 2024. https://doi.org/10.48550/arXiv.2410.02219

[8] Jiang M., Tian Z., Yu C., Shi Y., and et al., "Intelligent 3D Garment System of the Human Body Based on Deep Spiking Neural Network," *Virtual Reality and Intelligent Hardware*, vol. 6, no. 1, pp. 43-55, 2024. https://doi.org/10.1016/j.vrih.2023.07.002

[9] Jiang Y., DeepFashion-MultiModal, https://github.com/yumingj/DeepFashion-MultiModal, Last Visited 2025.

[10] Kampakaki E. and Papahristou E., "Digital Intelligence as Prerequisite of Artificial Intelligence's Integration in the Clothing Industry 4.0," *SETN Workshops*, vol. 2844, pp. 36-41, 2020. https://ceur-ws.org/Vol-2844/fashion1.pdf

[11] Kang H., "Accelerating Innovation Through AI-Powered Conceptual Abstraction and Interaction Design," Doctoral Dissertation, Carnegie Mellon University, 2024. https://doi.org/10.1184/R1/26863186

[12] Kim J., Laine T., and Ahlund C., "Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review," *Applied Sciences*, vol. 11, no. 4, pp. 1738-1756, 2021. https://doi.org/10.3390/app11041738

[13] Krishnamoorthy R. and Manojkumar C., "IoT-Based Robotic Systems Integrating Artificial Intelligence for Sustainable Crop Protection in Agriculture," *PatternIQ Mining*, vol. 1, no. 3, pp. 1-11, 2024. https://doi.org/10.70023/piqm24301

[14] Mohiuddin Babu M., Akter S., Rahman M., Billah M., and Hack-Polay D., "The Role of Artificial Intelligence in Shaping the Future of the Agile

Fashion Industry," *Production Planning and Control*, vol. 35, no. 15, pp. 2084-2098, 2024. https://doi.org/10.1080/09537287.2022.2060858

[15] Pang K., Zou X., and Wong W., "FashionM3: A Unified Vision-Language Model for Multimodal Fashion Retrieval and Generation," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1234-1245, 2025. https://doi.org/10.48550/arXiv.2504.17826

[16] Ruckdashel R., Venkataraman D., and Park J., "Smart Textiles: A Toolkit to Fashion the Future," *Journal Of Applied Physics*, vol. 129, no. 13, pp. 130901-130915, 2021. https://doi.org/10.1063/5.0024006

[17] Shakeel P., Aboobaider B., and Salahuddin L., "Detecting Lung Cancer Region from CT Image Using Meta-Heuristic Optimized Segmentation Approach," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 36, no. 16, pp. 2240001, 2022. https://doi.org/10.1142/S0218001422400018

[18] Sharma S., Koehl L., Bruniaux P., Zeng X., and Wang Z., "Development of an Intelligent Data-Driven System to Recommend Personalized Fashion Design Solutions," *Sensors*, vol. 21, no. 12, pp. 4239-4255, 2021. https://doi.org/10.3390/s21124239

[19] Shimizu R., Saito Y., Matsutani M., and Goto M., "Fashion Intelligence System: An Outfit Interpretation Utilizing Images and Rich Abstract Tags," *Expert Systems with Applications*, vol. 213, pp. 119167, 2023. https://doi.org/10.1016/j.eswa.2022.119167

[20] Singh A. and Patras I., "Fashionsd-X: Multimodal Fashion Garment Synthesis Using Latent Diffusion," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5678-5689, 2024. https://doi.org/10.48550/arXiv.2404.18591

[21] Wang T. and Ye M., "TexFit: Text-Driven Fashion Image Editing with Diffusion Models," *in Proceedings of the AAAI Conference on Artificial Intelligence*, Vancouver, pp. 10198-10206, 2024. https://doi.org/10.1609/aaai.v38i9.28885

[22] Wang Z., Tao X., Zeng X., Xing Y., and et al., "An Interactive Personalized Garment Design Recommendation System Using Intelligent Techniques," *Applied Sciences*, vol. 12, no. 9, pp. 4654-4668, 2022. https://doi.org/10.3390/app12094654

Ting Zhou is a native of Ningbo, Zhejiang Province, China. She obtained her Bachelor of Arts degree in Fashion and Apparel Design from Zhejiang Sci-Tech University in 2017, and later received her Master of Arts degree in Art Design from the same institution in 2020. Since 2021, she has been serving as a lecturer at the School of Textile and Garment, Kezhi Art College, Zhejiang Sci-Tech University. Her primary research interests lie in the Integration of Artificial Intelligence with Fashion and Apparel Design, as well as the Innovative Development of Traditional Chinese Lacquer Jewelry.