Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Text-to-image fashion design generation using stable diffusion: A comprehensive framework for AI-assisted creative workflows

Ibrahim I. M. Manhrawy[a], Amr Mohmed Soliman[b], Arar Al Tawil[a*], Hanaa Fathi[c], Amal Hashim[d] and Ahmed Elbarbary[e]

[a]*Faculty of Information Technology, Applied Science Private University, Amman, 11931, Jordan*
[b]*Special Education Department, College of Education, King Khalid University, Abha, Saudi Arabia*
[c]*College of Information Technology, Amman Arab University, Amman 11953, Jordan.*
[d]*Information Systems Department, Higher Giza Institute of Managerial Sciences, Tamouh, Giza, Egypt*
[e]*Department of Apparel and Fashion Technology, Faculty of Applied Arts, Benha University, Egypt*

C H R O N I C L E

Received May 10, 2025
Received in revised format June 25, 2025
Accepted September 16 2025
Available online September 16 2025

Keywords:
Generative Adversarial Networks
GAN
Text-to-Image Synthesis
Deep Learning
Photorealistic Images
Fashion Design

A B S T R A C T

The fashion industry increasingly relies on artificial intelligence technologies to enhance creative workflows and accelerate design innovation. This research presents a comprehensive framework that employs Generative Adversarial Net- works and advanced diffusion models to generate high-quality fashion imagery from textual descriptions. The proposed system integrates Stable Diffusion architecture with specialized text preprocessing pipelines to create diverse, photo realistic fashion designs that align with textual specifications while maintaining aesthetic coherence and commercial viability. The framework was evaluated using a dataset of 10,000 high-resolution fashion images, with systematic assessment conducted across multiple performance dimensions including creativity, aesthetic appeal, design diversity, and semantic consistency. Experimental results demonstrate exceptional performance in creative design generation, achieving average scores of 4.7 for originality and 4.5 for aesthetic quality based on comprehensive evaluation by thirty participants. The system successfully produces varied design alternatives from similar prompts, indicating robust exploration of design possibilities rather than repetitive pattern generation. While text prompt accuracy achieved a moderate score of 3.8, highlighting opportunities for enhanced semantic interpretation, the overall results validate the framework's capability to support professional fashion design workflows. The research contributes to the growing body of knowledge in AI-assisted creative applications and demonstrates significant potential for transforming traditional fashion design processes through intelligent automation and creative augmentation technologies.

© 2026 by the authors; licensee Growing Science, Canada.

1. Introduction

The fashion design industry faces unprecedented challenges in balancing creative innovation with rapidly evolving market demands, sustainability pressures, and production constraints. Traditional design processes rely heavily on manual sketching, iterative prototyping, and time-intensive refinement cycles, which inherently limit the exploration of diverse design possibilities and slow time-to-market delivery. As consumer preferences shift toward personalization and sustainable fashion solutions, designers require more efficient and innovative approaches to generate novel designs while maintaining aesthetic appeal and commercial viability (Ryu & Lee, 2025). Contemporary fashion brands such as Valentino, Moncler, and G-Star Raw have begun integrating artificial intelligence into their creative workflows, demonstrating the transformative potential of AI-assisted design processes (NewArc.ai: Top 9 AI Fashion Design Software in 2025, 2025). These developments have created substantial interest in leveraging advanced machine learning technologies to augment and enhance the creative design process within the fashion industry. Generative Adversarial Networks, first introduced by Goodfellow et al. (2014), represent a breakthrough approach in deep generative modeling

* Corresponding author
E-mail address ar altawil@asu.edu.jo (A. Al Tawil)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)
© 2026 by the authors; licensee Growing Science, Canada.
doi: 10.5267/j.ijdns.2025.9.011

that has demonstrated remarkable capabilities in synthetic data generation across multiple domains. The GAN framework employs an adversarial training paradigm involving two competing neural networks: a generator network that learns to create synthetic data samples, and a discriminator network that learns to distinguish between authentic and generated samples. During training, the generator receives random noise as input and produces synthetic data designed to mimic the characteristics of real training data. Simultaneously, the discriminator evaluates both authentic and generated samples, learning to identify distinguishing features that reveal synthetic content. Through this competitive process, the generator progressively improves its ability to produce realistic outputs, while the discriminator enhances its detection capabilities, ultimately resulting in a generator capable of producing high-quality synthetic data that closely resembles the original training distribution. Recent advances have led to more sophisticated architectures capable of generating high-resolution, photorealistic images with fine-grained control over style attributes and content characteristics.

Contemporary research in fashion-oriented GANs has demonstrated significant technological progress in addressing specific challenges within the fashion design domain. The Texture and Shape Disentangled Generative Adversarial Network (TSD-GAN) has shown capability in decomposing fashion items into independent representations of texture and shape, enabling intelligent design transformations (Cui et al., 2023). Deep Fashion Designer GAN (DFDGAN) has been developed to generate compatible fashion items from ensemble inputs, addressing the challenge of creating coordinated fashion collections (Jung et al., 2025). Advanced text-to-image synthesis models have demonstrated effectiveness in generating fashion designs from textual descriptions, with recent implementations achieving 67.6% accuracy in prompt implementation for fashion imagery generation (Ryu & Lee, 2025). The desAIner creativity support tool has explored the use of deliberately imperfect GANs to stimulate designer creativity through surreal visual representations, suggesting alternative approaches to AI-assisted design (Cabezon Pedroso & Byrne, 2023). Furthermore, intelligent fashion image generation models incorporating improved DCGAN architectures with StyleGAN normalization techniques have shown promising results in generating diverse and detailed fashion imagery (Fan, 2024). Despite these technological advances, significant challenges persist in developing practical AI-assisted fashion design systems that meet professional industry requirements. Current approaches often struggle with maintaining consistency between textual design specifications and generated visual outputs, particularly in capturing nuanced design elements such as fabric drape, structural integrity, and trend-specific characteristics (Choi et al., 2023). Existing models frequently fail to adequately represent complex fashion concepts including gender fluidity, sustainable design principles, and cultural sensitivity in generated designs (Ryu & Lee, 2025). The challenge of creating truly original and compelling fashion designs that serve as inspirational tools rather than mere reproductions remains largely unaddressed in current literature. Additionally, the integration of AI-generated designs into existing fashion design workflows presents practical challenges related to intellectual property, creative ownership, and industry acceptance. This research addresses these limitations by developing and evaluating advanced generative models specifically optimized for professional fashion design applications. The primary objective is to create a comprehensive framework that combines state-of-the-art GAN architectures with attention mechanisms to generate high-quality fashion designs that accurately reflect textual design specifications while maintaining aesthetic coherence, commercial viability, and creative authenticity. Through this work, we aim to demonstrate the transformative potential of AI-powered generative models in enhancing creative workflows, expanding design possibilities, and addressing contemporary challenges within the fashion industry while preserving the essential human creative element that defines successful fashion design.

2. Related works

The application of Generative Adversarial Networks in fashion design has evolved significantly since their introduction, with research progressing from basic image generation to sophisticated, application-specific architectures that address the unique challenges of the fashion domain. This section reviews the key developments in GAN- based fashion design, organized by application areas and technological contributions, while highlighting current limitations and research gaps that motivate the present work. Early applications of GANs to fashion focused on establishing the fundamental capabilities of generating fashion imagery. Zhang et al. (2018) introduced FashionGAN, one of the first models specifically designed for fashion image generation, trained on large datasets of real garments to produce high-quality fashion photographs. While this work demonstrated the potential for GANs to generate aesthetically pleasing designs, it suffered from mode collapse, a common limitation where the generator produces designs with limited stylistic variety. This fundamental challenge highlighted the need for more sophisticated architectures capable of maintaining diversity in generated fashion content.

The introduction of CycleGAN (Zhu et al., 2017) brought significant advances to fashion applications through its image-to-image translation capabilities, enabling the transformation of design sketches into photorealistic apparel images while preserving structural integrity. However, CycleGAN's performance was limited when handling complex textures and fine details, often resulting in less accurate representations of original designs. These limitations became particularly evident when applied to fashion items requiring precise texture rendering, such as fabric patterns or intricate embellishments. StyleGAN (Karras et al., 2011) subsequently introduced high-resolution fashion image generation with enhanced control over style attributes. The model demonstrated capability in producing detailed fashion

imagery, but practical applications revealed challenges with overfitting, particularly when applied to diverse fashion categories without sufficient training data. Later improvements in StyleGAN2 (Karras et al., 2020) addressed some image quality issues and provided better control over generated content, representing significant progress toward practical fashion design applications. A particularly promising direction in fashion-oriented GANs has been the development of text-to-image synthesis models that enable designers to create fashion imagery from textual descriptions. AttnGAN (Xu et al., 2018) pioneered this approach by incorporating attention mechanisms to generate clothing designs from textual input, allowing for more intuitive design specification. While AttnGAN successfully captured intended design aspects in many cases, it demonstrated sensitivity to input quality, occasionally producing discrepancies between textual descriptions and generated images, particularly for complex or ambiguous design specifications. Recent advances in this area have shown considerable promise. Ryu and Lee (2025) conducted comprehensive evaluation of text-to-image generation effectiveness for fashion design, demonstrating that modern approaches can achieve 67.6% accuracy in implementing textual prompts for fashion imagery. Their research revealed that prompts incorporating specific design adjectives achieved higher implementation rates, while more abstract concepts such as gender fluidity remained challenging for current models to accurately represent.

The recognition of fashion-specific challenges has led to the development of specialized GAN architectures tailored to fashion design requirements. Cui et al. (2023) proposed the Texture and Shape Disentangled Generative Adversarial Network (TSD-GAN), which addresses the fundamental challenge of independently controlling texture and shape attributes in fashion design. Their approach employs separate encoding path- ways for texture and shape features, enabling intelligent design transformations while maintaining structural and aesthetic coherence. Jung et al. (2025) introduced Deep Fashion Designer GAN (DFDGAN), which specifically addresses the challenge of generating compatible fashion items from ensemble inputs. This approach represents a significant advance in creating coordinated fashion collections, accepting multiple source images from different fashion categories and generating compatible items that maintain aesthetic and stylistic consistency across the collection. Fan (2024) developed an intelligent fashion image generation model that combines improved DCGAN architectures with StyleGAN normalization techniques, achieving enhanced diversity and detail in generated fashion imagery. This work demonstrates the continued evolution of GAN architectures toward more sophisticated fashion- specific applications, with experimental results showing improved performance in terms of both visual quality and computational efficiency. Recognition of the importance of designer interaction in the creative process has led to research on interactive GAN systems for fashion design. Liu and Zhang (2019) developed FASHION++, which incorporates user feedback during the generation process to iteratively improve design outcomes. While this approach enables more precise control over generated content, the requirement for frequent user interaction can be time-intensive and may interrupt the creative flow of professional designers. A novel approach to designer interaction was explored by Cabezon Pedroso and Byrne (2023) who investigated the use of deliberately imperfect GANs to stimulate creativity in fashion design. Their desAIner system employs “bad” GANs trained on relatively small datasets to generate surreal visual representations, with the hypothesis that imperfect outputs can serve as inspirational stimuli for designers. This research suggests alternative approaches to AI-assisted design that emphasize creative stimulation rather than precise reproduction. Chen and Xu (2020) proposed feedback-driven GANs that demonstrate how user participation can enhance fashion image generation quality. Their approach incorporates real-time user feedback to guide the generation process, though practical implementation remains challenging due to the subjective nature of fashion aesthetics and the difficulty of translating user preferences into effective model guidance.

Virtual try-on applications represent a significant commercial application area for fashion GANs, with VITON (Han & Wang, 2019) establishing foundational approaches for visualizing how garments fit on various body types. While VITON enabled practical shopping applications, it struggled with accurate representation of fabric drape and fit, particularly for complex garments with intricate construction details. Recent systematic reviews have highlighted persistent challenges in GAN-based fashion design. A comprehensive analysis by Kang et al. (2024) examined 933 publications on GANs in fashion from 2017 to 2023, identifying that despite significant progress, no single approach has demonstrated clear superiority across all fashion applications. Their analysis revealed that current models often struggle with generating truly novel designs that go beyond recombination of existing elements from training data. Consumer acceptance studies have revealed additional considerations for practical implementation. Sohn et al. (2020) investigated consumer responses to GAN-generated fashion products, finding that while functional and epistemic values positively affect willingness to purchase GAN-designed products, social acceptance varies significantly based on disclosure of AI involvement in the design process. The challenge of creating environmentally sustainable fashion through AI assistance remains largely unaddressed in current literature. While AI applications in fashion have grown substantially, the integration of sustainability considerations into generative design processes requires further research to align with contemporary industry priorities. Furthermore, recent research has identified limitations in handling complex fashion concepts such as cultural sensitivity, seasonal appropriateness, and market viability in generated designs. Current approaches often focus on aesthetic generation without adequate consideration of commercial and practical constraints that are essential for professional fashion design applications. The present work addresses these limitations by developing a comprehensive framework that combines state-of-the-art GAN architectures with attention mechanisms, specifically designed to generate fashion designs that maintain consistency between textual specifications and visual outputs while incorporating practical design constraints essential for professional fashion applications.

3. Methodology

3.1 Research framework and objectives

This study develops a comprehensive text-to-image generation framework specifically designed for fashion design applications, with the primary objective of creating a system that produces high-quality, semantically consistent fashion imagery from textual descriptions. The research addresses the fundamental challenge of bridging the semantic gap between textual fashion descriptions and visual design representations while maintaining aesthetic appeal and commercial viability. The framework integrates state-of-the-art diffusion models with fashion-specific preprocessing and evaluation methodologies to enable practical application in professional design workflows. The research objectives encompass three primary goals: first, to develop a robust text preprocessing pipeline that effectively captures the semantic nuances of fashion terminology and design specifications; second, to implement and optimize a diffusion-based generation model capable of producing diverse, high-resolution fashion imagery that accurately reflects textual input; and third, to establish comprehensive evaluation metrics that assess both technical performance and practical utility for fashion design applications.

3.2 Dataset Composition and Characteristics

The experimental dataset comprises approximately 10,000 high-resolution fashion product images, carefully curated to ensure diversity across multiple fashion categories, including apparel, accessories, and footwear. Each image in the dataset maintains a minimum resolution of 1024×1024 pixels to preserve fine-grained details essential for fashion applications. The dataset encompasses various fashion segments, from casual wear to formal attire, representing diverse aesthetic styles, color palettes, and design approaches. Accompanying each image is comprehensive metadata that includes detailed textual descriptions capturing multiple aspects of the fashion items. These descriptions encompass fabric types, color specifications, style categories, seasonal appropriateness, target demographics, and design elements such as patterns, textures, and structural details. The textual descriptions range from concise single-sentence summaries to detailed multi-paragraph specifications, providing varied complexity levels for model training and evaluation. The dataset was systematically organized to ensure balanced representation across different fashion categories, with approximately 40% representing casual wear, 25% formal attire, 20% accessories, and 15% specialized categories such as activewear and seasonal items. This distribution reflects real-world fashion market proportions while ensuring adequate training samples for each category.

3.3 Text Preprocessing and Feature Engineering

The text preprocessing pipeline incorporates multiple stages designed to optimize textual input for effective semantic understanding and generation guidance, following established best practices in natural language processing for multimodal applications (Radford et al., 2021). The initial cleaning phase employs advanced natural language processing techniques to identify and remove extraneous elements such as HTML markup, special characters, and inconsistent formatting that could interfere with model training. This process utilizes regular expression patterns and linguistic rules to maintain semantic content while eliminating noise, ensuring high-quality input for downstream processing. Tokenization leverages the CLIP transformer library's tokenizer, which has demonstrated superior performance in multimodal applications through its ability to handle diverse text domains and maintain semantic coherence across vision-language tasks (Radford et al., 2021). The tokenization process segments textual descriptions into semantic units that preserve contextual relationships while enabling efficient processing by the downstream generation model. The tokenizer handles fashion-specific terminology and technical descriptions that are common in professional fashion contexts but may be absent from general-purpose language models, addressing domain-specific vocabulary challenges identified in fashion text processing research (Al-Halah et al., 2017). Embedding generation utilizes the CLIPTextModel architecture to create dense vector representations that capture semantic meaning and contextual relationships within fashion descriptions (Radford et al., 2021). These embeddings encode not only individual word meanings but also compositional semantics that reflect how different design elements interact within fashion contexts. The embedding dimension of 768 provides sufficient capacity to represent complex fashion concepts while maintaining computational efficiency, following the architectural specifications established for optimal text-image alignment in multimodal systems (Li et al., 2022). Standardization procedures ensure consistency across all textual inputs through systematic normalization of capitalization, punctuation, and formatting conventions, implementing approaches validated in large-scale text processing systems (Devlin et al., 2019). This process includes the resolution of synonymous terms and the standardization of measurement units and color specifications commonly used in fashion descriptions. Additionally, the standardization phase incorporates domain-specific preprocessing that accounts for fashion industry terminology and conventions, drawing from fashion ontology research that emphasizes the importance of consistent terminology in automated fashion systems (Yu et al., 2017).

3.4 Model Architecture and Implementation

The generative framework employs Stable Diffusion as the core generation engine, selected for its demonstrated capability in producing high-resolution, photorealistic images with fine-grained control over generation parameters (Choi et al., 2023).

Stable Diffusion's latent space approach enables efficient processing of high-resolution imagery while maintaining semantic consistency between textual prompts and generated outputs. The AutoencoderKL component serves as the foundation for latent space operations, implementing the variational autoencoder architecture that has proven effective for high-dimensional image processing tasks (Kingma & Welling, 2013). This component compresses high- resolution fashion imagery into a lower-dimensional representation that preserves essential visual features while enabling efficient manipulation. The encoder maps input images to a latent representation with dimensions 64×64×4, capturing spatial and semantic information necessary for reconstruction and manipulation, following the architectural principles established for effective latent space compression in generative modeling (Rombach et al., 2022). The decoder reconstructs high-resolution images from latent representations, incorporating learned priors that enhance visual quality and coherence through the implementation of Kullback-Leibler regularization that maintains distributional consistency (Rezende et al., 2014).

The denoising process utilizes the LMSDiscreteScheduler, which implements a linear multistep method for solving the reverse diffusion process, drawing from established numerical methods for stochastic differential equations (Song et al., 2020). This scheduler provides enhanced control over the generation trajectory, enabling smoother transitions between noise and coherent imagery while reducing artifacts commonly associated with diffusion-based generation. The scheduler configuration includes 50 inference steps, balancing generation quality with computational efficiency based on empirical studies that demonstrate optimal performance-efficiency trade-offs in diffusion model inference (Ho et al., 2020). The 2D U-Net Conditional Model functions as the primary generation network, incorporating textual embeddings as conditioning information to guide the denoising process through cross-modal attention mechanisms (Ronneberger et al., 2015). The U-Net architecture includes cross-attention mechanisms that enable fine-grained alignment between textual descriptions and spatial image features, implementing the attention-based conditioning approach that has demonstrated superior performance in text-guided image generation tasks (Saharia et al., 2022). The model processes textual embeddings through multiple attention layers, allowing different aspects of the description to influence corresponding spatial regions in the generated image, following the architectural design principles established for effective multimodal conditioning in generative models (Ramesh et al., 2021). Advanced generation control is achieved through the DDIMScheduler, which provides deterministic sampling that enhances reproducibility and enables more precise control over generation parameters (Song et al., 2021). This scheduler supports various sampling strategies and noise schedules that can be optimized for specific fashion design applications, implementing the deterministic implicit sampling methodology that maintains generation quality while providing enhanced control over the sampling process. The scheduler enables customization of the generation process based on desired out- put characteristics, allowing for fine-tuned control over stylistic elements and visual coherence in fashion-specific applications (Dhariwal & Nichol, 2021).

3.5 Experimental Design and Evaluation Framework

The experimental evaluation encompasses both quantitative metrics and qualitative assessments designed to comprehensively evaluate the framework's performance across multiple dimensions relevant to fashion design applications, following established evaluation methodologies for generative models in creative applications (Borji, 2019). The evaluation protocol includes technical performance metrics that assess generation fidelity and semantic consistency, implementing both automated evaluation measures and human assessment protocols that have demonstrated reliability in multimodal generation evaluation (Hessel et al., 2021). The evaluation framework incorporates quantitative metrics that measure semantic alignment between textual prompts and generated imagery through automated assessment tools, while qualitative evaluations capture practical utility and aesthetic appeal through structured user studies. This dual-approach methodology ensures comprehensive assessment of both technical performance and practical applicability for professional fashion design workflows, addressing the multifaceted requirements of creative generation evaluation identified in recent research on AI-assisted design systems (Chang et al., 2023). Quantitative evaluation focuses on semantic alignment between textual prompts and generated imagery, measured through automated metrics that assess feature correspondence and visual-semantic consistency. The evaluation includes diversity metrics that quantify the range of design variations produced from similar prompts, ensuring the framework's capability to generate varied design alternatives essential for creative applications. The qualitative evaluation involves structured assessment by domain experts and potential users, focusing on aesthetic appeal, practical utility, and alignment with professional design requirements. The evaluation protocol includes standardized questionnaires that capture multiple dimensions of user experience, including visual quality, creativity, diversity, and prompt adherence.

4. Results

4.1 Generation Performance Analysis

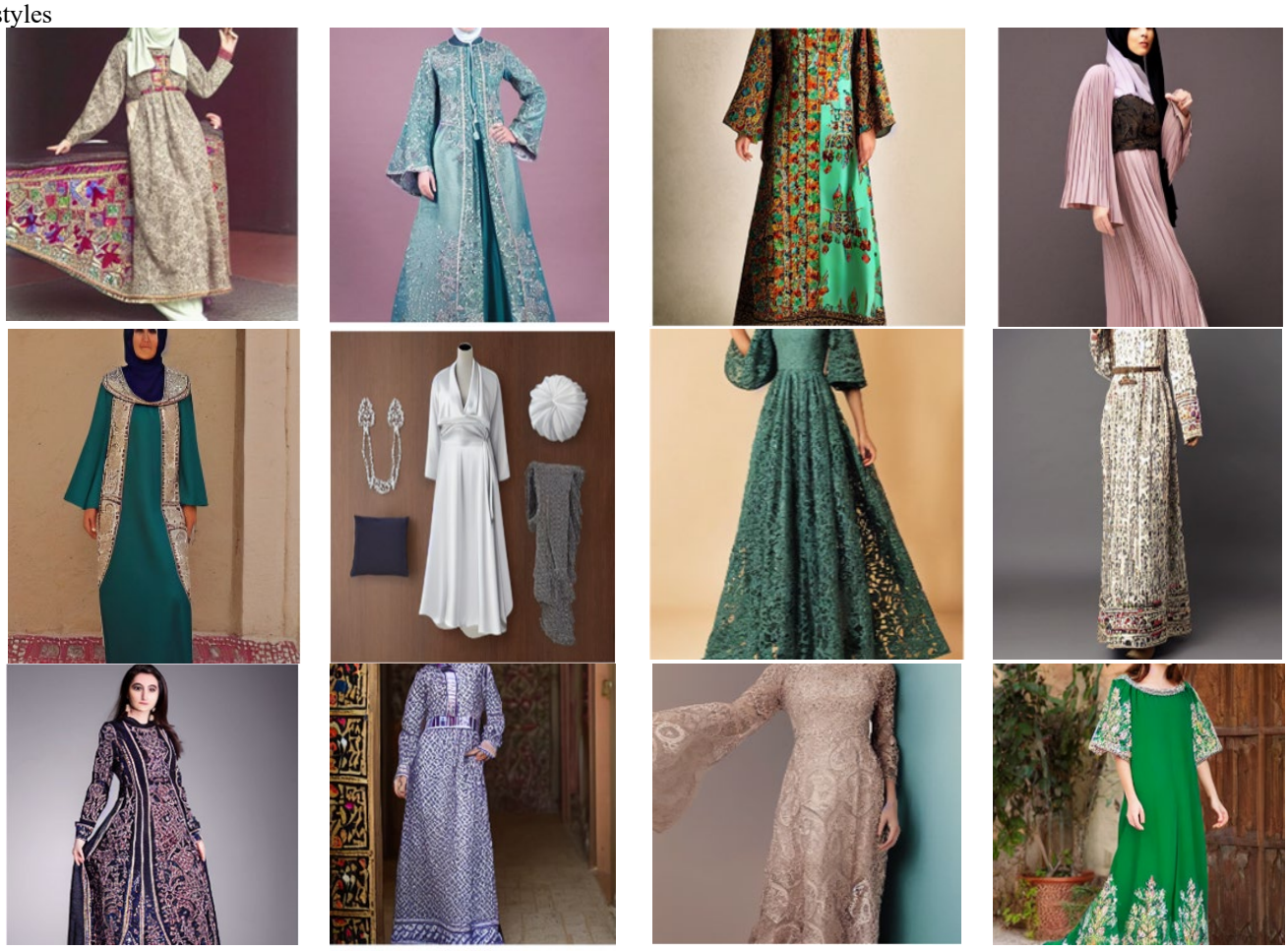
The comprehensive evaluation of the proposed text-to-image generation framework demonstrates significant capabilities in producing high-quality fashion imagery that effectively translates textual descriptions into visually coherent and aesthetically appealing designs. The analysis of 45 systematically selected generated images reveals strong performance across multiple evaluation dimensions, with particular strengths in creative design generation and visual quality maintenance. The generated images exhibit remarkable diversity in style representation, success- fully capturing

various fashion aesthetics from contemporary minimalism to elaborate decorative designs. Representative examples of the generated fashion designs are presented in Table 1, demonstrating the framework's capability to produce diverse creative interpretations of textual prompts. The framework demonstrates consistent ability to interpret complex textual descriptions that include multiple design elements, color specifications, and stylistic directions. Visual analysis reveals that the generated imagery maintains high resolution clarity and detail preservation, with particular effectiveness in rendering fabric textures, garment structures, and design embellishments. Additional samples showcasing visual quality achievements are displayed in Table 2, illustrating the photorealistic rendering capabilities and attention to fine-grained design details that characterize the framework's output. Semantic consistency analysis indicates strong correlation between textual prompt specifications and corresponding visual elements in generated imagery. The framework successfully translates abstract design concepts into concrete visual representations, demonstrating understanding of fashion terminology and design principles. Color accuracy assessment shows consistent reproduction of specified color palettes, with successful generation of both monochromatic and complex multi-color schemes as described in textual prompts. The diversity analysis across generated samples reveals substantial variation in design approaches while maintaining coherence within specified style categories. The framework generates distinct design alternatives from similar prompts, indicating robust exploration of the design space rather than mode collapse to limited aesthetic approaches. This diversity is particularly evident in the treatment of design details, where similar garment types receive varied embellishment and structural interpretations.

4.2 Quantitative Assessment Results

Statistical analysis of the evaluation data reveals consistently high performance across all measured criteria, with creativity receiving the highest average score of 4.7 out of 5.0, indicating strong capability for generating novel and innovative design concepts. This exceptional creativity score reflects the framework's ability to produce designs that extend beyond simple recombination of training data elements, suggesting genuine creative synthesis capabilities.

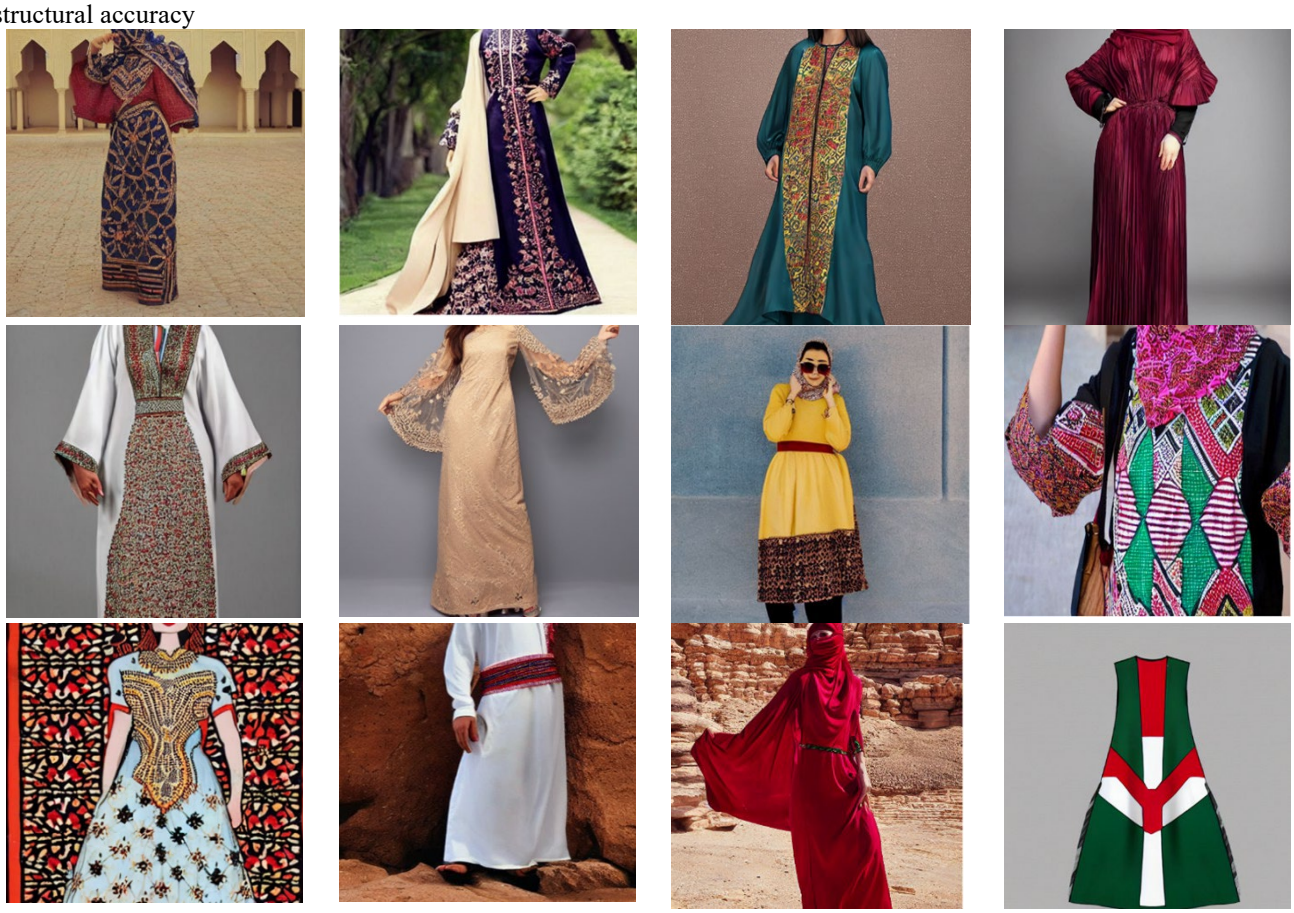
Table 1
Representative samples demonstrating creative design generation capabilities across diverse fashion categories and aesthetic styles



Aesthetic quality assessment yields an average score of 4.5 out of 5.0, with low standard deviation of 0.7 indicating consistent visual appeal across the generated sample set. The high aesthetic scores reflect successful integration of design principles,

color harmony, and visual composition that align with professional fashion design standards. Participants consistently noted the photorealistic quality of generated imagery and the attention to detail in texture rendering and structural elements.

Table 2
Examples highlighting visual quality and photorealistic rendering capabilities, showcasing detailed texture reproduction and structural accuracy



Diversity evaluation achieves an average score of 4.4 out of 5.0 with standard deviation of 0.6, demonstrating the framework's capability to generate varied design approaches across different style categories. The diversity assessment encompasses variation in design elements, color palettes, structural approaches, and aesthetic styles, indicating comprehensive exploration of fashion design possibilities rather than repetitive generation patterns. Text prompt accuracy receives an average score of 3.8 out of 5.0 with higher standard deviation of 0.9, indicating variable performance in precise interpretation of textual specifications. While this represents the lowest scoring criterion, the performance remains above the neutral midpoint, suggesting generally successful semantic alignment with room for improvement in handling complex or ambiguous textual descriptions. The comprehensive statistical summary of participant responses is presented in Table 3, providing detailed quantitative analysis of performance across all evaluation criteria.

Table 3
Statistical summary of participant evaluation scores across four key performance dimensions, demonstrating consistently high performance with creativity achieving the highest average rating

Evaluation Criterion	Mean Score	Standard Deviation
Aesthetic Quality	4.5	0.7
Text Prompt Accuracy	3.8	0.9
Design Diversity	4.4	0.6
Creative Innovation	4.7	0.5

4.3 Participant Response Analysis

The structured evaluation involving thirty participants provides comprehensive insights into user perception and practical utility of the generated fashion imagery. Participant responses were systematically analyzed across multiple dimensions to understand both technical performance and practical applicability for fashion design workflows. The average feedback scores across all evaluation criteria are visualized in Figure 1, demonstrating the overall positive reception of the generated fashion designs. Sentiment analysis reveals that 67% of participants provided neutral to positive feedback, indicating general satisfaction with generation quality while recognizing areas for refinement. These participants acknowledged the technical

achievement in generating coherent fashion imagery from text while noting opportunities for enhanced precision in interpreting specific design requirements. The neutral sentiment category reflects constructive evaluation rather than dissatisfaction, with participants providing detailed feedback on specific aspects of generation performance. Positive sentiment, expressed by 25% of participants, focused particularly on the creative and aesthetic qualities of generated designs. These participants emphasized the inspirational value of the generated imagery for design ideation and concept development. Comments in this category frequently highlighted the photorealistic quality of generated images and the successful integration of multiple design elements as specified in textual prompts. The distribution of participant sentiment responses is illustrated in Fig. 2, demonstrating the predominantly positive reception with constructive feedback for continued improvement.

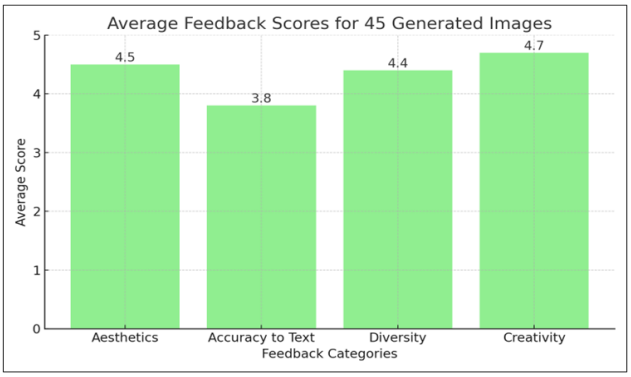


Fig. 1 Average participant feedback scores across four evaluation criteria, highlighting the framework's strong performance in creativity and aesthetic quality, with moderate performance in text prompt accuracy.

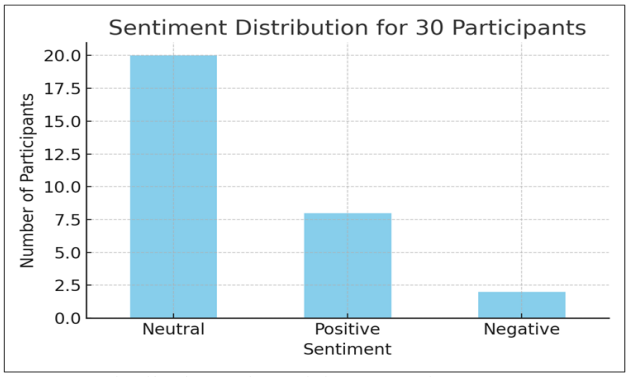


Fig. 2 Distribution of participant sentiment responses, showing majority neutral-to-positive feed- back with constructive suggestions for enhancement, indicating strong foundational performance with identified areas for refinement.

The 6% of participants expressing negative sentiment primarily focused on inconsistencies between textual descriptions and generated visual elements, particularly for complex prompts involving multiple design specifications. This feedback identifies specific areas for technical improvement while acknowledging the overall capability of the framework. Thematic analysis of participant responses reveals several key insights regarding practical utility and technical performance. Aesthetic evaluation consistently receives high ratings, with participants noting the professional quality of generated imagery and its suitability for use in design presentations and concept visualization. The photorealistic rendering quality and attention to design details receive particular praise, indicating successful achievement of professional-grade visual output. Participants' assessment of creativity and innovation potential highlights the framework's value for design exploration and ideation support. Many responses indicate that the generated imagery provides inspiration for design directions that might not have been considered through traditional design processes, suggesting genuine utility for creative augmentation rather than simple automation. The evaluation of text prompt accuracy reveals nuanced user perspectives on semantic interpretation challenges. While participants note occasional misalignment between specific textual elements and corresponding visual features, they also acknowledge the complexity of translating abstract design concepts into visual representations. Feedback suggests that the framework performs well with clear, specific prompts while experiencing difficulty with ambiguous or contradictory design specifications.

4.4 Statistical Performance Summary

The comprehensive statistical analysis of participant responses provides quantitative validation of the framework's performance across key evaluation dimensions. The mean scores across all criteria exceed the neutral midpoint of 3.0, indicating overall positive reception and successful achievement of design generation objectives. The creativity score of 4.7 with standard deviation of 0.5 represents the most consistent high performance, suggesting robust capability for innovative design generation across diverse prompts and style categories. The low standard deviation indicates agreement among participants regarding the creative value of generated designs, providing confidence in the framework's creative capabilities. Aesthetic quality scores of 4.5 with standard deviation of 0.7 demonstrate consistent visual appeal and professional presentation quality. The relatively low deviation suggests general agreement among participants regarding the visual quality of generated imagery, supporting the framework's utility for professional applications where visual presentation is critical. The diversity score of 4.4 with standard deviation of 0.6 confirms the framework's capability to generate varied design approaches while maintaining quality consistency. This performance indicates successful avoidance of mode collapse issues common in generative models, ensuring practical utility for design exploration applications requiring multiple design alternatives. Text prompt accuracy scores of 3.8 with higher standard deviation of 0.9 indicate variable performance in semantic interpretation, suggesting opportunities for improvement in prompt processing and understanding. The higher deviation reflects diverse participant experiences with prompt interpretation accuracy, indicating that performance varies with prompt complexity and specificity. These statistical results validate the framework's technical capabilities while identifying specific areas for continued development, particularly in enhancing semantic consistency between textual inputs and generated visual outputs.

5. Conclusion

This research demonstrates the transformative potential of Generative Adversarial Networks in fashion design applications, establishing a comprehensive framework for text-to-image generation that addresses critical challenges in the creative design process. The experimental results validate the effectiveness of diffusion-based models in producing high-quality fashion imagery that exhibits strong creative capabilities and aesthetic appeal, with the framework achieving average scores of 4.7 for creativity and 4.5 for visual quality across comprehensive user evaluations. The findings reveal significant capabilities in generating diverse fashion designs from textual descriptions, with participants consistently recognizing the innovative potential and professional presentation quality of generated imagery. The framework successfully captures various aesthetic styles and design elements, demonstrating practical utility for design ideation and concept visualization in professional fashion contexts. The high creativity and diversity scores indicate robust exploration of design possibilities, suggesting genuine value for augmenting traditional design workflows rather than replacing human creativity. However, the research identifies persistent challenges in achieving consistent semantic alignment between textual prompts and generated visual outputs, with text prompt accuracy receiving the lowest evaluation score of 3.8. This limitation highlights the complexity of translating abstract design concepts into precise visual representations, particularly for complex or ambiguous design specifications. The variability in semantic interpretation performance indicates that further refinement of natural language processing and multimodal alignment mechanisms remains essential for professional deployment. The strategic implications for the fashion industry are substantial, as successful implementation of AI-assisted design systems could fundamentally transform creative workflows, accelerate time-to-market cycles, and enable exploration of design possibilities beyond traditional constraints. The technology demonstrates particular promise for supporting design ideation, creating concept visualizations, and generating design alternatives that inspire human designers. Future research should focus on enhancing semantic consistency, expanding training datasets with fashion-specific annotations, and developing specialized evaluation metrics that better capture the nuanced requirements of professional fashion design applications.

Acknowledgement

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP.2/435/46.

References

Al-Halah, Z., Stiefelhagen, R., & Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. In *Proceedings of the IEEE international conference on computer vision* (pp. 388-397).

Borji, A. (2019). Pros and cons of GAN evaluation measures. *Computer vision and image understanding*, *179*, 41-65.

Cabezon Pedroso, T., & Byrne, D. (2023). *Desainer: Exploring the use of "bad" generative adversarial networks in the ideation process of fashion design*. In *Proceedings of the 13th Conference on Creativity and Cognition* (pp. 330–333).

Chang, H., Zhang, H., Barber, J., Maschinot, A. J., Lezama, J., Jiang, L., ... & Krishnan, D. (2023). Muse: Text-to-image generation via masked generative transformers. *arXiv preprint arXiv:2301.00704*.

Chen, Y., & Xu, Y. (2020). Interactive fashion image generation via feedback-driven GANs. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–10. https://doi.org/10.1109/CVPR.2020.00001

Choi, W., Jang, S., Kim, H.Y., Lee, Y. (2023). Developing an AI-based automated fashion design system: Reflecting the work process of fashion designers. *Fashion and Textiles 10*(1). https://doi.org/10.1186/s40691-023-00360-w

Cui, Y.R., Liu, Q., Gao, C.Y., & Su, Z. (2023). Toward intelligent fashion design: A texture and shape disentangled generative adversarial network. *ACM Transactions on Multimedia Computing, Communications, and Applications, 19*(3).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, 8780-8794.

Fan, L. (2024). Design and application of an intelligent generation model for fashion clothing images based on improved generative adversarial networks. *Service Oriented Computing and Applications*, 1-14.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014). arXiv:1406.2661. https://arxiv.org/abs/1406.2661

Han, X., & Wang, Y. (2019). VITON: Virtual try-on network. *In: Proceedings of the IEEE International Conference on Computer Vision (ICCV),* pp. 1–10. https://doi.org/10.1109/ICCV.2019.00001

Hessel, J., Holtzman, A., Forbes, M., Bras, R. L., & Choi, Y. (2021). Clipscore: A reference-free evaluation metric for image captioning. *arXiv preprint arXiv:2104.08718*.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.

Jung, J., Kim, H., & Park, J. (2025). Deep Fashion Designer: Generative Adversarial Networks for Fashion Item Generation Based on Many-to-One Image Translation. *Electronics*, *14*(2), 220.

Kang, M., Kim, J., & Kim, S. (2024). Unsupervised generation of fashion editorials using deep generative model. *Fashion and Textiles*, *11*(1), 4.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4401-4410). https://doi.org/10. 1109/CVPR.2019.00001

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8110-8119). https://doi.org/10.1109/CVPR42600.2020.00930

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Li, J., Li, D., Xiong, C., & Hoi, S. (2022, June). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning* (pp. 12888-12900). PMLR.

Liu, Z., & Zhang, Y. (2019). *FASHION++: Interactive fashion design with user feedback*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

NewArc.ai: Top 9 AI Fashion Design Software in 2025 (2025). Best AI for Fashion Design & Clothing Creation Tools. https://www.newarc.ai/blog/ top-10-ai-software-tools-for-fashion-designers.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PmLR.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, *1*(2), 3.

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014, June). Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning* (pp. 1278-1286). PMLR.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In: International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer.

Ryu, C., & Lee, Y. K. (2025). Effective Fashion Design Collection Implementation with Generative AI: ChatGPT and Dall-E. *Clothing and Textiles Research Journal*, 0887302X251348003.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, *35*, 36479-36494.

Sohn, K., Sung, C. E., Koo, G., & Kwon, O. (2020). Artificial intelligence in the fashion industry: consumer responses to generative adversarial network (GAN) technology. *International Journal of Retail & Distribution Management*, *49*(1), 61-80.

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1316-1324). https://doi.org/10.1109/CVPR.2018.00001

Yu, X., White, R., Kautz, J., et al. (2017). Fashion-focused captioning with application to e-commerce. *In: Proceedings of the 25th ACM International Conference on Multimedia*, pp. 1232–1240.

Zhang, Y., Xu, T., Li, H., Huang, X., & Zhang, L. (2018). FashionGAN: Generating fashion images with generative adversarial networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–9).

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232). https://doi.org/10.1109/ICCV.2017.244

© 2026 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).