IFSCC 2025 full paper (IFSCC2025-465)

"Development of the Generative AI based Pipeline Architecture for Advanced Virtual Makeup"

Myeongjin Goh [1,*], Seongmin Jeong [1], Heechan Jeon [2], Geonyeong Park [3], Serin Yang [3], Inhwa Han [3], Jin Nam [1] and Jongchul Ye [2]

[1] R&I Center, Amorepacific, Yongin, South Korea; [2] Amorepacific, Seoul, South Korea; [3] Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology, Seoul, South Korea

* Myeongjin Goh, R&I Center, Amorepacific, 1920 Yonggu-daero, Giheung-gu, Yongin, Gyeonggi-do, South Korea, +82-31-280-5862, stcanopus@amorepacific.com

1. Introduction

The pursuit of personalized and aesthetically optimized makeup remains a central desire among cosmetic consumers, reflecting their diverse facial features, skin tones, and style preferences. Traditionally, users have relied on physical product trials or in-person consultations to discover suitable makeup products and application techniques. However, these approaches are often time-intensive, costly, and subjective, leading to inconsistent outcomes and reduced consumer satisfaction.

In response, the global beauty industry has increasingly adopted virtual makeup technologies, driven by advances in artificial intelligence (AI) and augmented reality (AR). These systems offer real-time simulations, AI-assisted consultations, and personalized product recommendations, aligning with the broader trend of digital transformation and the growing demand for immersive, data-driven beauty experiences [1-4].

Despite recent progress, current virtual makeup systems face significant limitations in realism, personalization, and adaptability. Many rely on static overlays or rule-based algorithms, which inadequately capture the variability of human facial anatomy and dynamic makeup styles. Consequently, there is a pressing need for more sophisticated solutions that offer photorealistic rendering, individualized adaptation, and semantically driven customization [5, 6].

To address these challenges, we propose a novel generative AI pipeline for virtual makeup simulation, structured into three distinct phases to enable end-to-end user-specific beauty experiences (Figure 1).

The pipeline employs a three-phase process. First, facial morphology analysis is conducted using advanced computer vision to extract structural and anatomical features. Then, a large language model (LLM)—trained on beauty-related datasets—interprets user preferences through a multimodal interface that integrates text and image inputs. Finally, a stable diffusion model generates high-resolution, personalized virtual makeup simulations that maintain facial identity while allowing flexible style transformation.

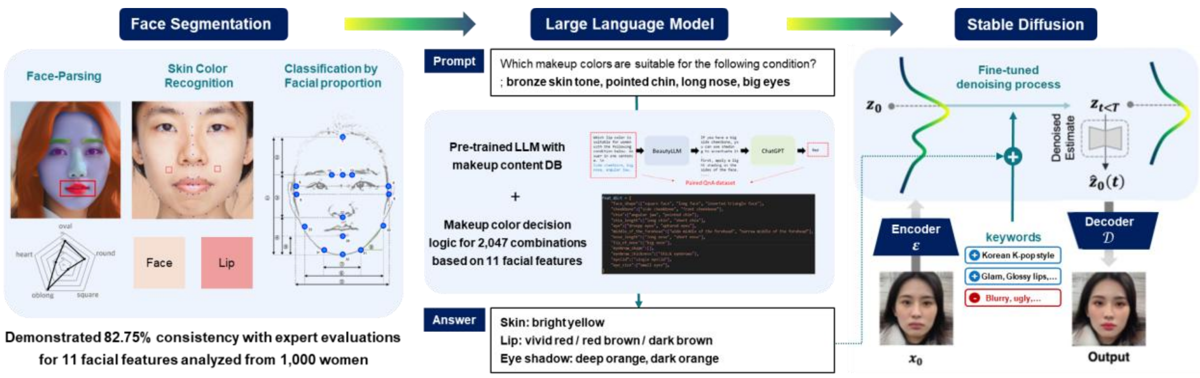


Figure 1. Overview of generative AI based pipeline architecture for virtual makeup. The pipeline integrates seamlessly integrates computer vision, large language models, and stable diffusion to enable a comprehensive process from facial diagnosis to personalized and realistic virtual application.

In particular, the pipeline supports a wide array of conditioning inputs—ranging from reference images and RGB values to descriptive text—enabling nuanced and diverse makeup effects. In contrast to existing AR techniques that apply beauty filters using presets, which require users to specify their comprehensive — makeup preferences manually, this method enables users to swiftly experience virtual makeup without the necessity of complicated information input or search processes.

In this study, we present an advanced virtual makeup pipeline that integrates generative diffusion modeling, semantic control, and identity preservation within a unified architecture. To the best of our knowledge, this is the first pipeline to cohesively combine LLM-based intent modeling with generative image synthesis for personalized digital beauty applications [7, 8]. The following sections detail the pipeline design, implementation methodology, and future directions for advancing intelligent makeup simulation technologies.

2. Materials and Methods

2.1. Construction of the Full-Face Image Database

A total of 4,003 full-face images of 519 Korean women were collected to define facial characteristics and support diagnostic system development. Publicly available datasets from the Open AI Dataset Project (AI-Hub, Republic of Korea) were utilized. All data are accessible via AI-Hub (www.aihub.or.kr).

2.2. Development of the Frontal Face Parsing System

A frontal face parsing system was developed using a pre-trained ResNet backbone within the BiSeNet v1 architecture [9]. The CelebAMask-HQ dataset served for training, validation, and functional evaluation [10]. The system enabled real-time extraction of facial structures for downstream analysis.

2.3. Facial Skin Color Analysis

Representative facial skin tones were determined by extracting pixel color values adjacent to the nasolabial folds, following the anatomical guidelines established by Jeong. Facial landmark detection employed the dlib library's 68-point shape predictor. Skin tone and lip color were defined by sampling the most frequent RGB values within regions of interest (ROIs) centered on key anatomical points, based on preliminary research [11].

2.4. Definition of Facial Classification Logic

Facial morphology was characterized using 64 landmark points based on anthropometric standards [12]. From these landmarks, 39 diagnostic indicators were computed, including distance ratios and angular relationships. Classification criteria were established with reference to a curated dataset of 1,000 Korean female faces, refined typology of facial types, and validated for diagnostic accuracy [13].

2.5. Language Model Setup

The dolly-v2-3b model, an instruction-tuned LLM based on the GPT-J architecture and developed by Databricks, was employed [14]. The model, comprising approximately 3 billion parameters, was accessed through Hugging Face and deployed on a desktop PC equipped with two NVIDIA RTX 4090 GPUs (24 GB VRAM each). All operations were conducted in a controlled research environment.

2.6. Diffusion-Based Makeup Customization

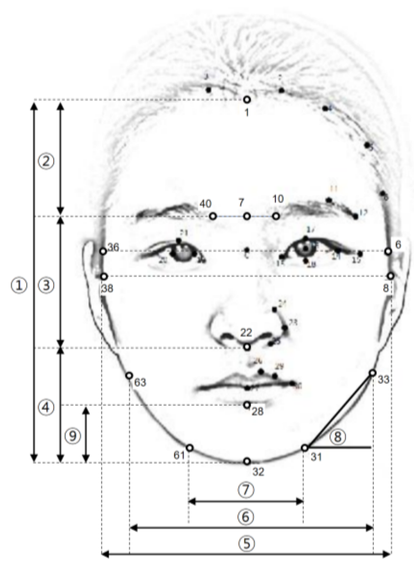
Stable Diffusion v1.5 [15] and SDXL [16] served as the base models for image generation. Makeup Transfer (MT) dataset images [17] and synthetically generated non-makeup images of Asian women were used as inputs. The pipeline was further enhanced by integrating Low-Rank Adaptation (LoRA) [18] weights and pre-trained models from the open-source platform CivitAI [19], facilitating improved fidelity and diversity in virtual makeup synthesis.

3. Results

3.1. Verification of the Frontal Face Parsing System

The development of a robust facial recognition and automated landmark detection system plays a crucial role in accurately analyzing facial features for virtual makeup customization. This system, based on the BiSeNet v1 architecture, was trained using 2,500 previously unseen frontal face images to enhance its accuracy. The system was optimized over 80,000 training iterations, resulting in a high accuracy of 80.78%. The system's ability to process frontal face images in real-time was further demonstrated by a mean Intersection over Union (IoU) score of 80.78% and a processing speed of 67.03 frames per second (FPS). These results show the system’s efficiency in detecting and analyzing facial features, which are key to providing precise makeup applications.

3.2. Development of a Model for Face Segmentation



Measurement indicators		Definition
Face shape	① Face length	distance between 1 and 32
	② Upper face	distance between 1 and 7
	③ Middle face	distance between 7 and 22
	④ Lower face	distance between 22 and 32
	⑤ Face width	distance between both sides 8
	⑥ Lower jaw width	distance between both sides 33
Chin	⑦ Chin width	distance between both sides 31
	⑧ Chin angle	angle between the line connecting 31–33 and the horizontal line
Chin length	⑨ Lower face length	distance between 28 and 32
	⑩ Upper face length	distance between 22 and 28
Cheek	⑪ Cheek prominence	(distance between 8 and 38) - (distance between 6 and 36)

Figure 2. Measurement indicators for female facial features and quantification methods. Landmark points and diagnostic measurements are defined to systematically classify facial morphology.

Accurate face segmentation is essential for the virtual application of makeup to ensure that the correct cosmetic products are applied to the right facial regions. To achieve this, 64 standardized facial measurement points were utilized, representing key facial landmarks, which serve as a basis for quantifying structural characteristics. The classification model, using the BiSeNet v1 architecture, enabled real-time semantic segmentation. As shown Figure 2, the model's ability to segment the face into individual components (e.g., facial shape, chin length, and eye size) within approximately one second highlights its efficiency. The facial attributes were then analyzed by calculating specific length, ratio, and angle values.

These measurements were categorized into five facial shapes based on diagnostic logic, as illustrated in Figure 3. Additionally, the model employed reinforcement learning using 940 annotated images, evaluated by makeup artists, to refine the classification process. This ensures that diagnostic logic of the model is both accurate and user-adaptable.

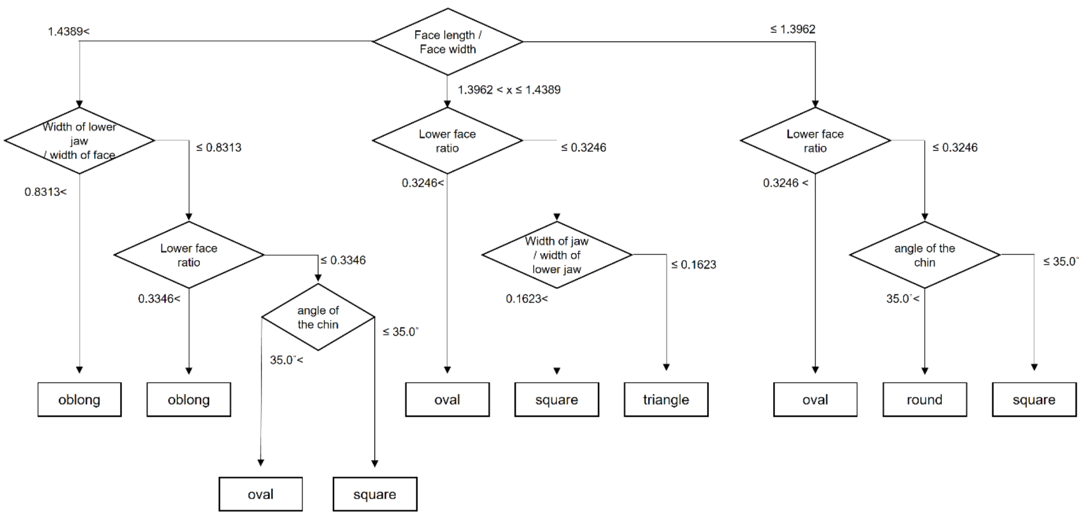


Figure 3. Schematic of classification criteria for female facial shapes. Facial shapes are categorized based on ratios and angles derived from landmark-based diagnostic indicators.

The model achieved an 82.59% concordance with expert assessments based on a validation study of 1,000 facial datasets, confirming reliability and precision of the model (Table 1).

Indicators	Definition	Number of items	Classification					Accuracy
Face shape	Overall shape of the face	5	Oval	Round	Square	oblong	triangle	78.33%
Cheekbones	Protrusion and direction of cheekbones	2	Regular Cheekbones	Side Cheekbones				88.33%
Chin	Chin tip shape	3	Square chin	Regular chin	Pointed chin			86.66%
Chin length	Distance between lower lip and chin	3	Long chin	Regular chin	Short chin			86.66%
Eye slant	Degree of upward or downward slant of the outer corners of the eyes	3	Drooping eyes	Regular eyes	Raised eyes			75%
Brow	Distance between the inner corners of the eyes	3	Wide brow	Regular brow	Narrow brow			78.33%
Nose	Length between the brow and the bottom of the nose	3	Long nose	Regular nose	Short nose			78.33%
Eyelid	Presence and shape of eyelids	2	Single eyelid	Double eyelid				86.66%
Eye size	Total width of the eye	3	Large eyes	Regular eyes	Small eyes			85%

Table 1. Face segmentation classification metrics and accuracy of the frontal face diagnosis

The face segmentation analysis was thus refined to achieve a high level of personalization and accuracy for makeup recommendations.

3.3. Evaluation of Facial Color Analysis

An essential aspect of makeup customization is the accurate assessment of skin tone and lip color, as these are critical to creating a harmonious look. To evaluate skin tone classification, a collaboration with professional makeup artists and R&D researchers led to iterative refinement of the color classification algorithm. The skin tone and lip color were determined based on pixel colors extracted from regions of interest (ROIs) near the nasolabial folds. As shown in Figure 4, the dominant RGB color values from the ROIs were used to define skin tones and lip colors.

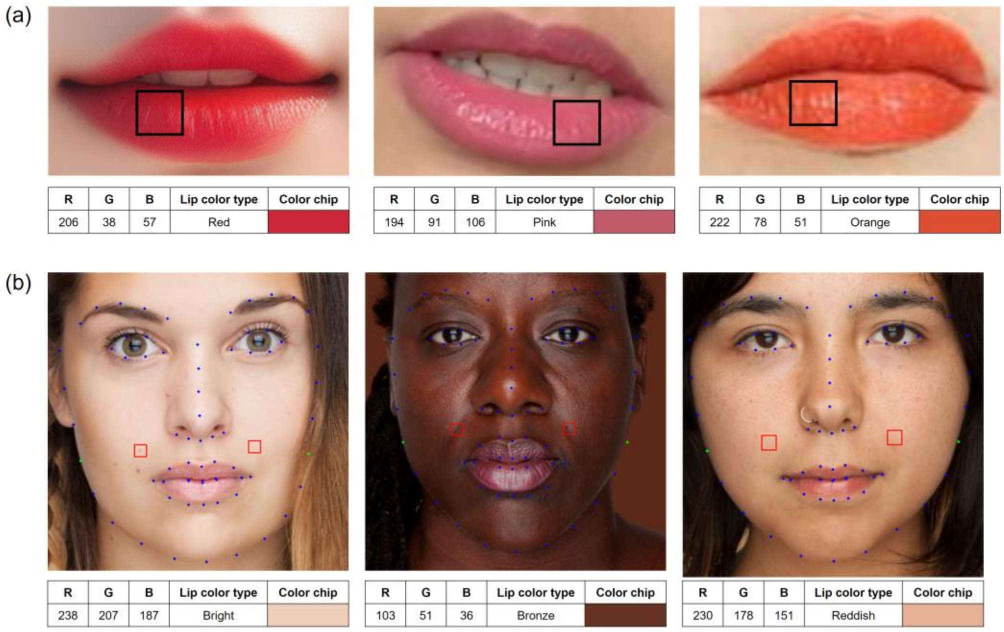


Figure 4. Results of facial color analysis. Representative RGB values for lip color (a) an

d skin tone (b) are extracted from defined regions of interest (ROIs).

3.4. Development and Fine-Tuning of a Beauty-Focused Language Model

To enhance the pipeline's ability to generate personalized makeup recommendations, an LLM was fine-tuned with makeup-specific datasets. The LLM, built on the dolly-v2-3b architecture, was optimized with 50,000 Q&A pairs sourced from the ChatGPT-3.5 API [20]. This dataset provided a broad range of beauty-related topics, ensuring that the LLM could understand and respond to a variety of user inputs. As depicted in Figure 5, this fine-tuning enabled the LLM to generate personalized makeup suggestions based on facial features and skin tone. The recommendation logic of the LLM prioritizes makeup combinations for the base, lips, eyeshadow, and blush, starting with lip color, followed by eyeshadow and blush suggestions, all tailored to the user's unique facial traits and preferences. The trained LLM allows for seamless integration with the facial analysis part of the pipeline, ensuring that makeup recommendations are grounded in scientific understanding and expert guidance.

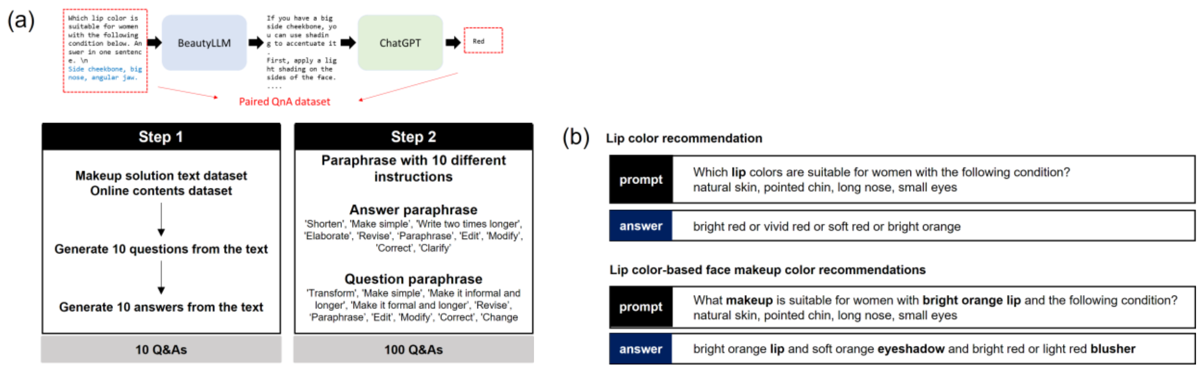


Figure 5. Makeup recommendation learning process based on facial features. A two-step approach generates Q&A datasets (a) and fine-tunes the LLM to recommend personalized makeup colors (b).

3.5. Development of a Generative Model for Makeup Application

The central innovation of this study is the development of a generative AI-based diffusion model that applies makeup to virtual facial images while preserving facial identity and minimizing distortion. The model employs a diffusion-based technique, specifically the Denoising Diffusion Implicit Models (DDIM) inversion method [21], which samples pixel-by-pixel from the original image to achieve the desired makeup style. This approach enables high-fidelity customization, maintaining the structural integrity of facial features while accurately applying makeup colors.

Figure 6 presents the core principle of the diffusion model, which demonstrated its ability to apply RGB color values with high accuracy, without causing distortion or loss of facial features. The integration of diffusion models with color control techniques enables the precise application of makeup shades, overcoming the challenges previously faced by diffusion-based image generation methods.

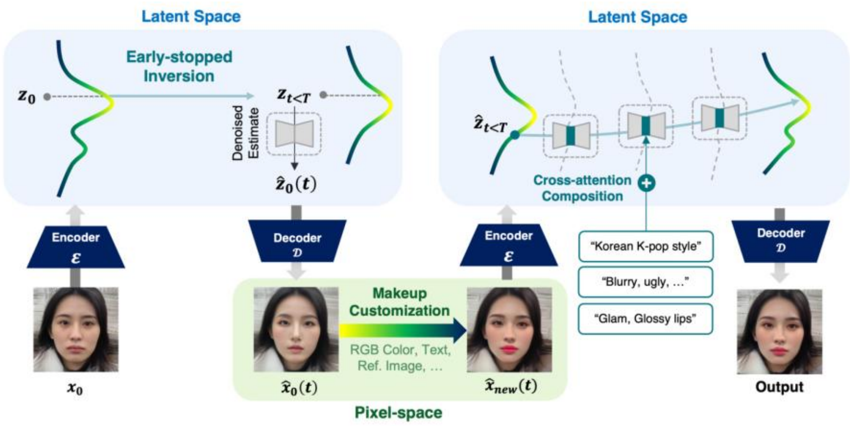


Figure 6. Overview of the generative AI-based makeup customization diffusion model. Precise guidance is applied in pixel space during reverse sampling, and text prompts are used to harmonize local variations into a consistent global style.

In Figure 7, the makeup transformation process is illustrated. The upper three rows show the application of eye shadow, skin, and lip colors as indicated by the RGB values in the bottom right corner. The bottom row presents the combined results of these makeup transformations, demonstrating how the model successfully applies specific RGB colors to various facial regions.

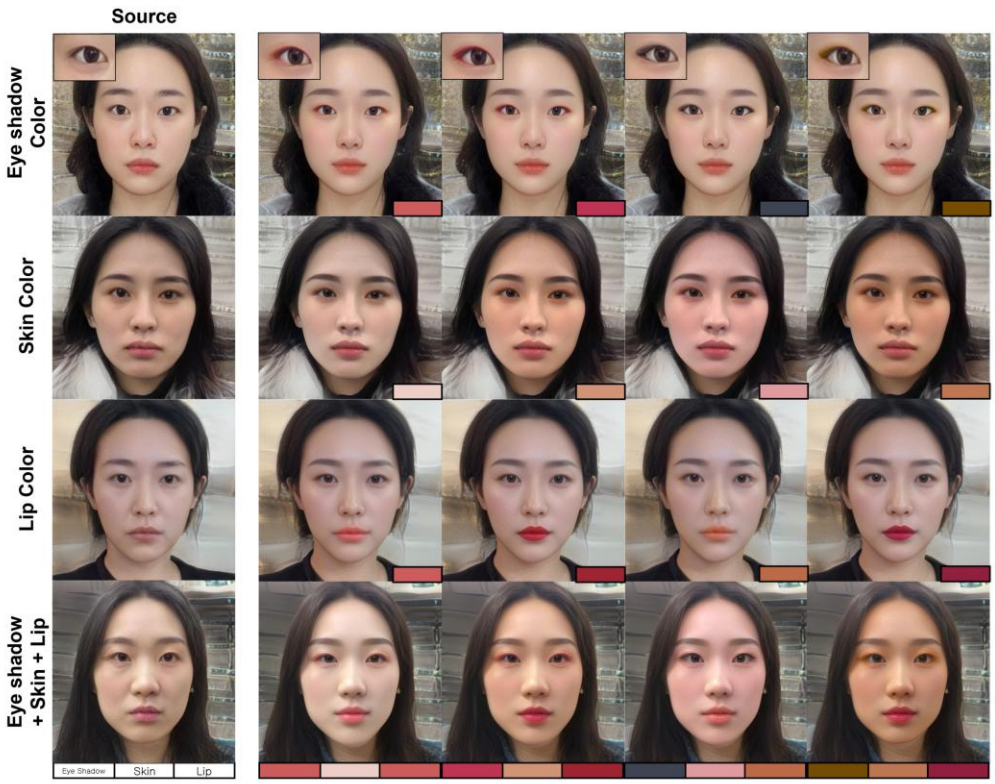


Figure 7. RGB color-based virtual makeup application results. Specified RGB values are applied to skin, lip, and eye shadow regions while maintaining facial identity, with individual and combined results presented across the image grid.

In Figure 8, the integration of the generative AI pipeline with the existing makeup transfer model is depicted. This model not only enhances structural consistency, a challenge in previous diffusion models, but also minimizes the occurrence of abnormal artifacts, showcasing its potential as an ideal solution for personalized, high-quality virtual makeup transfer.

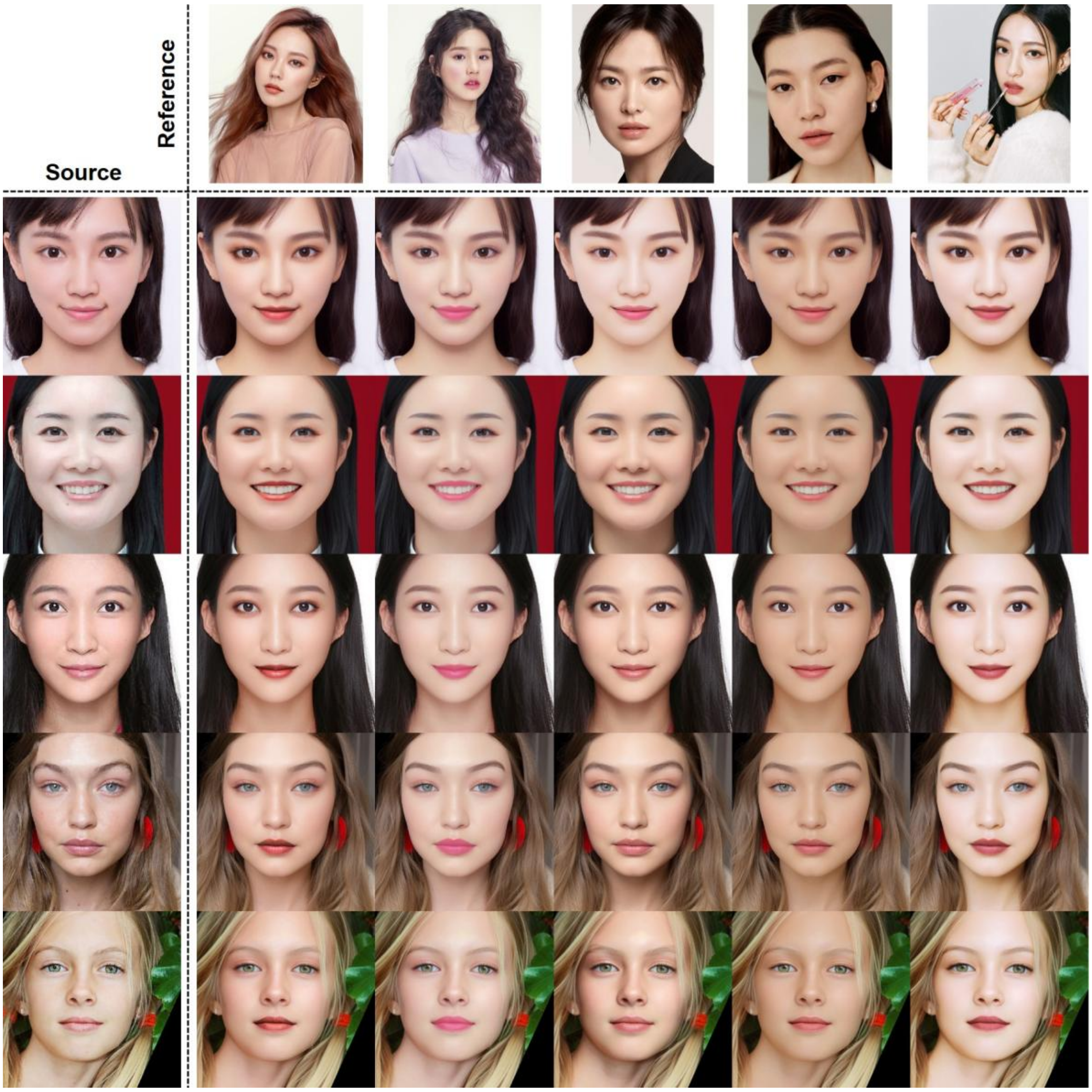


Figure 8. Makeup transfer results using the generative AI-based diffusion model. The model generates diverse makeup styles and colors while preserving facial structure and minimizing artifacts.

4. Discussion

This study presents a novel generative AI pipeline that technically advances virtual makeup applications by integrating face segmentation, a large language model, and stable diffusion within a unified architecture. Unlike previous approaches that treated face analysis and recommendation as separate or manual steps, this pipeline enables automated, end-to-end virtual makeup experiences. The integration of customer face analysis with LLM, allowing the pipeline to understand user preferences and generate personalized makeup recommendations-eliminating the need for users to manually search or select desired styles.

Compared to prior virtual try-on tools, which often relied on static presets or limited image synthesis capabilities, this pipeline supports a far greater diversity of makeup styles and color variations. These features allow for precise, realistic application while preserving the user's unique facial identity. These advances position the pipeline as a significant leap forward in both user experience and technical capability.

In the broader context, this research demonstrates how integrating advanced computer vision, natural language understanding, and generative image synthesis can overcome long-standing limitations in virtual beauty technology. Future research may focus on expanding the pipeline's makeup repertoire, enhancing real-time performance, and incorporating augmented reality for even more immersive experiences.

5. Conclusion

This work establishes that the generative AI pipeline enables a level of virtual makeup experience previously unattainable-allowing customers to receive expert, personalized makeup consultations and try-ons anytime and anywhere. The pipeline's integration of face analysis, conversational AI, and advanced image synthesis eliminates the need for manual selection and professional intervention, making high-quality beauty advice universally accessible.

From a business perspective, this technology offers a powerful digital tool for enhancing customer engagement, efficiently linking makeup products to individual needs, and collecting valuable insights on user preferences. The pipeline's scalable, cloud-based design ensures smooth deployment across devices and platforms, setting a new benchmark for digital transformation in the beauty industry. These innovations not only improve user satisfaction but also provide a strategic foundation for future advancements in personalized cosmetics and customer-centric beauty solutions.

6. References

[1] Scholz, J.; Duffy, K. We ARe at Home: How Augmented Reality Reshapes Mobile Marketing and Consumer-Brand Relationships. *J. Retail. Consum. Serv.* 2018, 44, 11–23.

[2] Wedel, M.; Bigné, E.; Zhang, J. Virtual and Augmented Reality: Advancing Research in Consumer Marketing. *Int. J. Res. Mark.* 2020, 37(3), 443–465.

[3] Smink, A.R.; van Reijmersdal, E.A.; van Noort, G.; Neijens, P.C. Shopping in Augmented Reality: The Effects of Spatial Presence, Personalization and Intrusiveness on App and Brand Responses. *J. Bus. Res.* 2020, 118, 474–485.

[4] Rathgeb, C.; Dantcheva, A.; Busch, C. Impact and Detection of Facial Beautification in Face Recognition: An Overview. *IEEE Access* 2019, 7, 152667–152678.

[5] Alashkar, S.; Jiang, S.; Wang, S.; Fu, Y. Examples-Rules Guided Deep Neural Network for Makeup Recommendation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.

[6] Gulati, K.; Verma, G.; Mohania, M.; Kundu, A. BeautifAI: A Personalised Occasion-Oriented Makeup Recommendation System. In Proceedings of the 2021 International Conference on Big Data Analytics (BDA), Guwahati, India, 15–18 December 2021.

[7] He, F.; et al. Makeup Transfer: A Review. IET Comput. Vis. 2023, 17(5), 513–526.

[8] Goel, P.; Sharma, B.; Kumari, A.; Gupta, A.K.; Chhajed, P. A Review on Virtual Try-On. In Proceedings of the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024; pp. 1–5.

[9] Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.

[10] Yan, Y.; Choi, H.; Suk, H.-J. Exploring the Facial Color Representative Regions Using the Humanae Images. J. Imaging Sci. Technol. 2020, 64(6), 060406-1–060406-7.

[11] Lee, Y.; Liu, Z.; Hu, Z.; Luo, P.; Kautz, J. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; pp. 5549–5558.

[12] Martin, R.; Saller, K. Lehrbuch der Anthropologie in Systematischer Darstellung; Gustav Fischer Verlag: Stuttgart, Germany, 1957; Vol. 3.

[13] Baek, K.-J.; Kim, Y.-I. Formative Elements of the Facial Image of Korean Women and the Effects of Makeup Design. J. Korean Soc. Cost. 2014, 64(4), 1–20.

[14] Databricks. Dolly v2-3B: An Instruction-Tuned Open LLM. 2023. Available online: https://www.databricks.com/blog/2023/04/12/hello-dolly-democratizing-magic-open-models.html/

[15] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 10684–10695.

[16] Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; Rombach, R. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. arXiv 2023, arXiv:2307.01952.

[17] Li, L.; Zhang, S.; Chen, T.; Zha, Z.-J. BeautyGAN: Instance-Level Facial Makeup Transfer with Deep Generative Adversarial Network. In Proceedings of the 26th ACM International Conference on Multimedia (ACM MM 2018), Seoul, Korea, 22–26 October 2018; pp. 645–653.

[18] Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. arXiv 2021, arXiv:2106.09685.

[19] CivitAI. An Open Platform for Sharing and Discovering AI Models. Available online: https://civitai.com/

[20] OpenAI. ChatGPT-3.5; OpenAI: San Francisco, CA, USA. Available online: https://platform.openai.com/

[21] Song, J.; Meng, C.; Ermon, S. Denoising Diffusion Implicit Models. arXiv 2020, arXiv:2010.02502.