Systematic Literature Review on technology of Hairstyle Transfer with Generative Adversarial Network

Glen Lordeo Tunjung
Computer Science Department
Student of Computer Science
Bina Nusantara University
Palangka Raya 73112, Indonesia
glen.tunjung@binus.ac.id

Joshua Chandra Lian
Computer Science Department
Student of Computer Science
Bina Nusantara University
Jakarta 14440, Indonesia
joshua.lian@binus.ac.id

John Wisely Laino
Computer Science Department
Student of Computer Science
Bina Nusantara University
Bekasi 17111, Indonesia
john.laino@binus.ac.id

Samuel Philip
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
samuel.philip@binus.ac.id

Hidayaturrahman
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
hidayaturrahman@binus.ac.id

***Abstract* — Deciding on a hairstyle can be a daunting task which could lead one to losing confidence instead of enhancing personality or boosting self-esteem. Recent studies see the emergence of means to simulate exchanging hairstyle using images via the technology of GAN. In this literature review, 13 studies of implementation of GAN to perform hairstyle transfer on images are sourced and screened following the PRISMA guidelines to be analyzed and discussed, as well as the potential of the technology for future applications. Older technologies suffer from the inability to transfer a hairstyle convincingly and accurately from target image to a source image which exhibits pose variance. Recent implementations tackle this issue by perfecting the algorithm for semantic segmentation mask which results in better alignment. Current state-of-the-art models exchange segmentation masks with alternative methods which outperform the former in terms of accuracy and preservation of hair. The issue of dataset diversity also comes into play in terms of bias and generalization exhibited by the models. Future developments towards efficiency and availability of the technology which would aid in transforming the hairstyling industry that hopefully leads to more creative and hassle-free hairstyling for the public.**

***Keywords* — Hairstyle transfer, generative adversarial network, StyleGAN, deep learning**

I. Introduction

Hair styling is an essential part of a person's appearance and has an important role in expressing individuality. Having a suitable hair style not only improves one's overall appearance but also boosts one's self-esteem. On the other hand, choosing the suitable hairstyle can be a challenging task, and even made more difficult by other factors such as face shape [1], hair texture [2], pose [3], and most importantly personal preferences of suitable hair styles. Therefore, it is necessary to build and improve a system that simulates hairstyles with high accuracy when an image of an individual's face and the image of the targeted hair style is inputted. Many implementations have demonstrated success in delivering convincing hairstyle transfer from an existing image to another. Most notable ones are through the usage of Generative Adversarial Network.

A GAN (Generative Adversarial Network) in a general definition is an architecture of deep learning made up of two neural networks that compete with one another in a framework similar to zero-sum game. This architecture is designed to produce new, synthetic data that closely mimics a preexisting data distribution. An instance of general adversarial network called StyleGAN2 is a modification of Style GAN.

In an early study of hair style image swapping and compositing, hair can be decomposed into hair structure and hair appearance. Those terms were previously used in the MichiGAN hair image generation, which introduced the concepts of identity, shape, structure, and appearance [4]. However, the research has not provided a clear description for such terms. Hair structure, which is closely related to hair shape, refers to the fundamental characteristics of hair, such as the form of hair's lock features, as opposed to hair appearance, which typically refers to the finer aspects of hair, such as hair color. Those two are the main attributes for the hair style image synthesizing process used in both prototypes.

To achieve high-quality hair style image synthesizing process, another implementation namely LOHO [5] implements the StyleGANv2 algorithm built on Style GAN and uses a novel approach to perform hair style transfer by optimizing pretrained StyleGANv2's extended latent space and noise space. This is followed by background inpainting, which is used to fill holes left over by misaligned hair masks during post-synthesizing process. Additionally, a two-stage optimization approach is provided in this prototype to enhance the photorealism of generated photos. Then, without interfering with one another, a technique known as gradient orthogonalization is used to maximize numerous qualities in the latent space.

The Barbershop prototype [6] also implements the StyleGANv2 algorithm similar to the LOHO prototype. The difference is, instead of using a network inpainting networks to fill holes left over by misaligned hair masks, this prototype uses only a single GAN-based semantic alignment step for the image synthesizing process. It produces high-quality photos that resemble the input images. This results in less noticeable artifacts caused by transparency, reflection, or interactions of the hair with the face when the semantic regions relevant to the task (which is the hair itself in this context) are aligned. As a result, this GAN-based semantic alignment step evolves to a new algorithm for GAN-based embedding used for aligned embedding that modifies image to follow a new segmentation mask.

Despite the improvements made in both Barbershop and LOHO from earlier predecessors, the implementations still suffer from imperfections and artifacts. A difference in the direction or angle between the two images results in faulty alignment as seen in LOHO. In the example of Barbershop, the segmentation masks could be misaligned which causes faulty reconstruction of the face resulting in unnatural images. Overall, the transfer of finer details is still lacking which made the resulting image less convincing upon closer inspection.

This literature review aims to outline the current state of development for hairstlye transfer techonology via the implementation of GAN. The improvements since the early days of the technology until the state-of-the-art method will be discussed in this paper. The problems solved as well as challenges faced by the proposed algorithms will be acknowledged and analyzed. Finding the problems that could be solved in the future for hairstyle transfer will contribute to the progress of the technology as demand for realistic means of hairstyle manipulation reaches.

II. Related Works

Current technologies for face detection or face recognition are used for many real-world applications. A comparative study on different methods of face recognition [7] showcases how face recognition can be achieved through differing approaches. Notably, image-based approaches, though more complicated than feature based approaches, reduce image feature corruption seen in Artificial Neural Network (ANN), Support Vector Machines (SVM), and Principal Component Analysis (PCA). Current face recognition technologies, however, encounter issues with presence of occlusion and non-uniform illumination.

Deep learning is a subfield of machine learning which involves training of artificial neural networks in facilitating the learning of data representations using multiple processing layers. A number of applications of deep learning had been demonstrated including computer vision, natural language processing, speech recognition, and genomics. In particular, the application in image understanding using deep convolutional networks has been shown to be successful with labelled data such as traffic sign, segmentation of biological images, and face recognition [8]. This technology plays a significant role in transferring hair in an image by distinguishing the facial part of the head with the hair target.

Generative adversarial network as proposed by the original paper is a neural network architecture composed of a generator and a discriminator [9]. The task of the generator as suggested is to generate realistic data samples while the discriminator then acts to identify which given data is fake or real. Feedback given to the generator improves its ability to generate more realistic data, while the discriminator becomes better at discriminating fake ones. The goal is for the generator to cause generated data to pass as real to the discriminator. An improvement to Style GAN, a high-quality image synthesis model, into StyleGANv2 provides better image quality through adaptive instance normalization, more stable training, and better control over the style or features of generated images [10].

Latent Optimization of Hairstyles via Orthogonalization (LOHO) is a proposed approach of hairstyle transfer in 2021 which attempts to solve the issue of hair misalignment artifacts by developing an algorithm specific to hairstyle transfer as opposed to general task [5]. A GAN inversion approach allows for better infilling of missing parts or artifacts caused by different poses exhibited between images [5]. LOHO attempts to produce a natural hairstyle transfer image via latent optimization to blend the face and the new hair, however the limitation of latent optimization itself leads to bad performance on cases with extreme poses discrepancy [3], [5]. Barbershop is another GAN based approach to hairstyle transfer from the same year which utilized a new latent space for image blending to produce a natural image instead of disjointed images put together [6]. This model solves the problem by aligning the two images to a target segmentation mask and copying an early style-block of the hair. Barbershop relies on the segmentation mask using naïve heuristic approach to determine the region where the composite image is created, which is also a limitation when handling cases with some geometric distortions [6].

III. Methodology

The literatures on the subject were sourced using Google Scholar, a search engine for the database of scientific literatures in many forms and scope. Terms used includes “hairstyle transfer”, “hairstyle replacement”, “hairstyle generation”, and “GAN”. These terms were matched to the title, abstract, and words found within the content of the literatures. Identification and screening for the papers are done by following the guidelines of The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) in the paper published in 2020 [11]. The PRISMA flowchart shows the process of filtering relevant papers for this literature review (Figure 1).

Initialy, 1,230 scientific papers were identified from the search terms. The first stage eliminates entries due to duplication, automated removal, and erronous search results. This includes papers of the same title published in different websites. Most common errors were literatures containing a partial term from the search keywords that were included in the results.

In the next stage, 200 papers were screened and a further 148 entries were excluded, some of which having no corellation between hairstyle and GAN. These were papers that only contained either an implementation of GAN or topics regarding hairstyle in general. The excluded entries also includes paper which validity cannot be confirmed. Out of 52 papers remaining, 12 was not able to be retrieved for further review.

The papers assessed for them to be eligible must include a strong corellation between the process of hairstyle transfer and the implementation of GAN. Entries which only mentions either hairstyle or GAN without attempting any GAN-based hairstyle transfer or generation are excluded. Literature reviews and exclusively case studies are excluded from elibible entries. The 13 literatures included in this review consist of novel implementation of GAN for hairstyle transfer ranging from late 2020 up to early 2023.

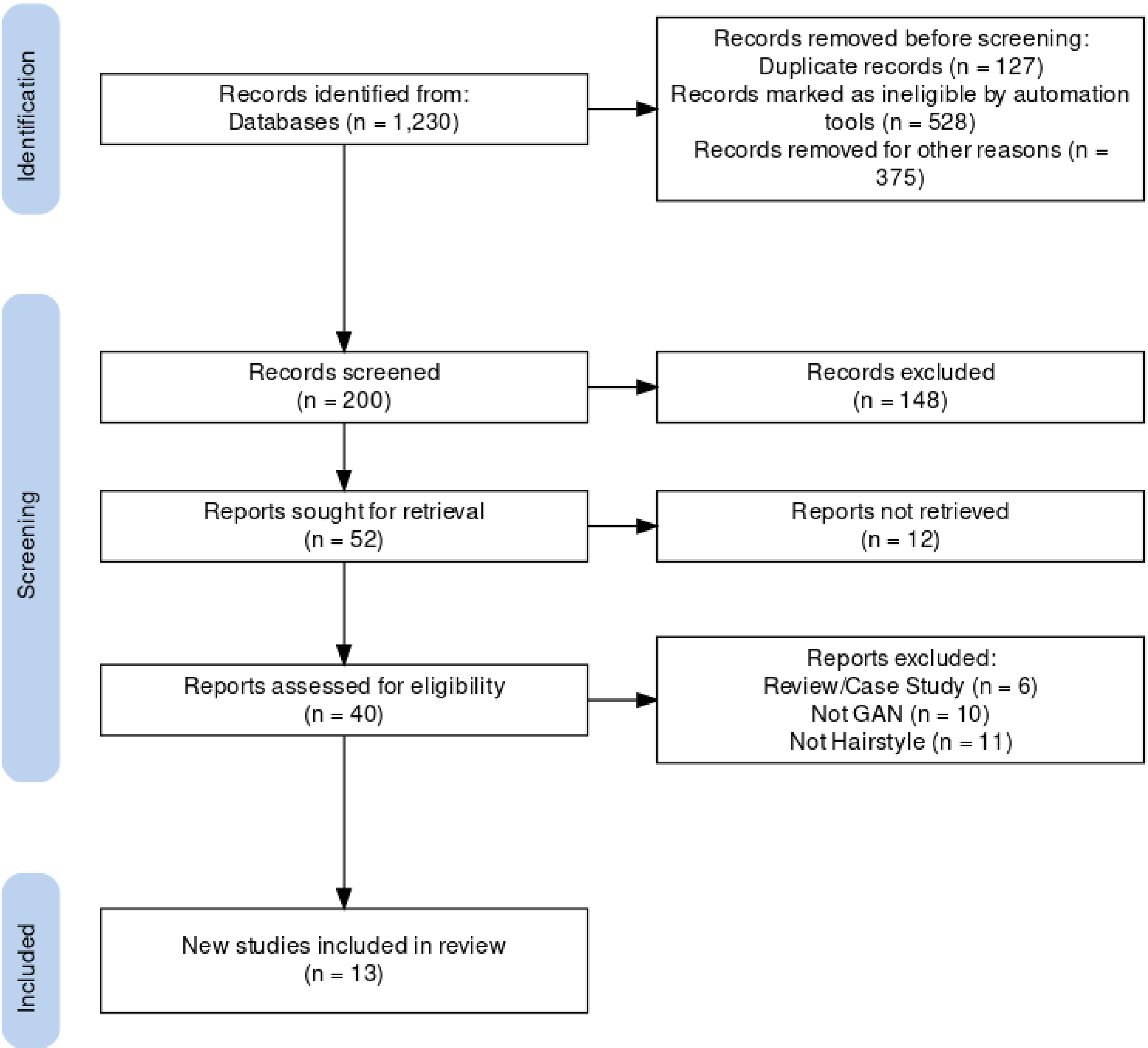

Figure 1. The PRISMA Flowchart

IV. RESULS AND DISCUSSION

The early development of GAN-based hairstyle transfer was a response to the success of facial image generation also implementing GAN. MichiGAN [4] achieves the feat of transferring details of hair in a manner which seems like they naturally belong on the new image, outperforming the state-of-the-art method of the time which is Semantic Image Generation that could not achieve simliar blending and preservation of the original hair. MichiGAN also introduced the ability for user tweaking which separated the outline, the texture, and the color of the hair via a combination of input images [4]. As pointed out by the authors however, the quality of the generated image is dependent on the accuracy and simplicity of the shape maks of the target hair (Figure 2). Drawing a mask that do not align to the face well or using a reference hair which shape differs dramatically would produce an unsatisfactory image.

Figure 2. MichiGAN mask painting [4].

The next study to improve on the implementation of MichiGAN is LOHO, a model which introduces a Two Stage Optimization strategy followed by Gradient Orthogonalization [5]. By separating optimization into reconstruction of the target image and then reconstructing all losses except for the reconstruction, LOHO is able to keep the shape of the target hair even when it does not match with the outline of the original hair where hair are or are not present by inpainting. The Gradient Orthogonalization step is to prevent interference between the gradient of the source and the target image. Visualy, this achieves a preservation of target hair which overlaps the source face without and gradient artifcats. In comparison to MichiGAN, this produces a more “blended-in” image as oppposed to “copy-pasted” [5]. Figure 3 shows a comparison where MichiGAN’s faulty mask without manual intervention, whereas LOHO adheres more to the face shape.

Figure 3. Source image, MichiGAN implementation, and LOHO implementation in order [5].

Other implementations attempting to improve on the quality from LOHO is Barbershop introducing Latent Space Manipulation [6] and Mask R-CNN Based Hairstyle Conversion [12]. Both of these models achieves segmentation of hair and face which results in stronger preservation of even finer details such as hair curls [6] and better color matching [12] (Figure 4). A problem with LOHO was that individuals’ faces were experiencing loss of detail or in extreme cases distortions due the the algorithm. This issue along with artifacts and misallignment are alleviated in the two later models due to the mapping of image segmentations

Barbershop

Mask R-CNN Based

Figure 4. Demonstration of shape and color mixing using Barbershop and Mask R-CNN Based Hairstyle Conversion [6, 12].

The most significant improvements made from previous studies is the ability to transfer hair from images where pose is significantly different. A problem occurs for example when the source image is facing one direction and the target is facing the other, tilted, or at some angle. Earlier algorithms could not adjust to the discrepancy resulting in an awkward image [4, 5, 6]. Later models improved on filling in the gap from missing information by perfecting the implementation of semantic segmentation masks [3, 13, 14]. Figure 5 showcases the ability to seamlessly paint hair into the source when target hair is facing a different direction.

Figure 5. A comparison between Barbershop (left) and Style Your Hair (right) when significant angle mismatch is present [3].

An issue of time cost performance is a limition that can be found even more recent models. For comparison, Barbershop takes in total 5 minutes from alignment, optimization, and generation using with top of the line graphics card at the time [6]. A more recent model of Style Your Hair achieves the feat in 8 minutes [15]. The earlier model LOHO can only finish in 15 minutes [15]. This presents a problem that hinders the use of the technology for broader market where high-end graphics card is not ubiquitous. From a consumer or business perspective, the time and computing power required is not feasible for practical applications.

One study attempting to remedy the computational cost is EHGAN (shortened from Efficient Hair Style Transfer with Generative Adversarial Network) [16]. The key to improving performance was to perform the hairstyle transfer on low-resolution images before being upscaled 4x using a super-resolution model. EHGAN also implements Hair Blending Block (HBB) which takes care of the process of blending backgrounds in a seamless manner. When compared with LOHO's 1448 seconds/image during its runtime, EHGAN outperforms by a vast margin with a time of 0.086 seconds/image [16]. This shows that a relatively real-time aplication of hairstyle transfer may be feasable in the future.

The majority of the models from these literatures are trained using high quality portrait images sourced from datasets such as Flickr-Faces-HQ (FFHQ) [4, 5, 6, 3, 14, 15, 16, 17, 14, 18], CelebA-HQ [12, 19, 3, 14, 18, 14], a more recent K-hairstyle [13], and video dataset VoxCeleb2 [13, 3]. Diverse datasets such as mentioned are required to ensure that the model recognizes as many types of hairstyles on faces ranging from all ages, genders, and ethnicities. Without diverse datasets of faces, the models might underperform on underrepresented groups of people. However, an outlier uses synthetic dataset generated by StyleGAN2 instead of real images [20]. Šubrtová et al. demonstrated that such algorithm is feasible to generate a comparatively convincing hairstyle transfer result against other methods. The algorithm proposed was simple and automatic. In contrast, other implementations require large database, tedious anotation and some user interactions [20].

A unique approach to presenting a hairstyle transfer to the public is by integrating a text-based application into the existing technology. HairCLIP introduces the Contrastive Language-Image Pre-Training (CLIP) model in a hairstyle transfer framework which allows user input of a combination of a reference image and text prompt (Figure 6) [19]. It allowed for a reduction of user interaction when adjusting the image into the desired outcome by simply providing a desciption of what the hair is supposed to look like. Combining text-based algorithm to image generation adds a

Figure 6. HairCLIP image output with respective textual input [19].

layer of human aspect which contributes to the ease-of-use in the user's prespective.

A challenge that is inevitable when dealing with two dimensional images is occlusion. The hairstyle transfer models are struggling when there are parts of the hair that is obstructed either due to the direction the head is facing or an object blocking the hair [6, 18, 13, 3, 14]. When a part of the hair is not visible, no information exists which would fill in the gap. If the faces do align well, it does not present much of an issue. However, if variable poses is introduced then the algorithm must be able to correctly predict the how the hair would appear without any data from the input [3]. The

Figure 7. Artifacts around areas outside of the hair [18].

opposite is also true when switching from longer to shorter hair [17], what will take place of the space that used to be hair?

What sets the state-of-the-art model apart from the previous ones is the extensive capability of transferring a hairstyle between source and target with complex pose variation. Both HairNET and StyleGAN Salon achieve the task by discarding the reliance of segmentation mask to predict the final output of the image [17, 15]. Although segmentation mask achieved seamless transfer of hair to another image it did have the tendency for overfitting in some cases. HairNET implements "baldification" which is the removal of hair earlier in the pipeline which simplifies the process of inpainting the areas outside the hair [15]. StyleGAN Salon achieves similar result by projecting 3D information into RGB space which preserves the geometric information of the hair and also generate a previously obstructed part of the face such as the forehead. The comparison of the two models can be seen in Figure 8.

Figure 8. Image generated using two state-of-the-art models of hairstyle transfer.

As the technology improves, it may become a commodity for the public consumers. Current accessible means of hairstyle transfer are only in the form of "filters" which do not represent an accurate image of replacing one's hairstyle, but rather a pasted image of hair on top of the existing one. Further development may see the possibility of even more creative uses such as eccentric hairstyles which is currently not viable even in the state-of-the-art model [17]. Expanding the support for more devices other than one manufacturer's graphics card would also improve usability, particularly the use of StyleGAN since the algorithm was originaly developed by NVIDIA.

Future implementations could focus on improving accessibility for the mass consumer. Optimizing the algorithm to generate output faster or downscalling the application would render it usable in lower powered devices such as mobile phones. From a technical standpoint, implementing a 3D image for the hair model would garner more interest from technology such as real-time live transfer and VR. Increased accessibility to the technology would improve productivity in the hairstyling industry. A better way to recommend and "try-out" a personalized hairstyle based on one's type of hair or shape of face would revolutionalize hair fashion and individuality.

V. Conclusion

The development of GAN-based hairstyle transfer has led to higher accuracy and performance of the technology. It has reached a point where preserving details and structure of the hair between differing poses is feasible. Introduction of new algorithms to combat pose variance shown in HairNET [15] and StyleGAN Salon [17] has broadened the scope of facial images which hairstyle transfer is possible. It has also been demonstrated with extensive optimization and upscaling low-resolution images it may become practical to run the algorithm at vastly more efficient times [16]. A future implementation of cloud-based service may substantially increase the usability of the technology across many devices, seeing as the current algorithm relies on one manufacturer's graphics processing unit.

A consensus regarding the datasets of faces for the training of the models is that it requires a diverse set ranging from all ages, ethnicities, and genders. Underrepresented groups of people may experience lower performance from the output. This can also be exhibited in types of hairstyles which are uncommon or eccentric [17]. The datasets used by the models in the order of most implemented are Flickr-Faces-HQ, CelebA-HQ, VoxCeleb2, and K-Hairstyle. More diverse datasets could decrease bias and decrease errors on more subjects of varying backgrounds and appearances.

Technical challenges yet to be solved are handling of occlusion in the images. Concealed or obstructed parts of the hair or face are missing information the algorithms need to synthesize. In the future, a better prediction algorithm may preserve or synthesize the geometric data of the hair in a 3D space. This increase in flexibility may render challenging poses of the head to be trivial. Isolating the hair may also improve upon the preservation of props present near the hair such as glasses, jewelry, or hands interacting with the hair.

The trajectory of the development of hairstyle transfer technology may see the revolution in the hairstyling industry. A virtual means of showcasing a hairstyle on an individual helps visualizing in an interactive manner before committing to a style. Further evolution may see the emergence of hairstyle recommendation algorithms which would improve efficiency of hairstyling businesses and overall satisfaction of hairstyling customers. However, this may also reduce the role of a hairstylist in deciding the most attractive or suitable hairstyles as the technology matures. On the other side, there may be a shift in trend where more creative hairstyles emerge which can only be attainable with a human touch.

VI. BIBLIOGRAPHY

[1] K. Pasupa, W. Sunhem and C. K. Loo, "A hybrid approach to building face shape classifier for hairstyle recommender system," *Expert Systems with Applications,* vol. 120, pp. 14-32, 2019.

[2] G. Daniels, S. Tamburic, S. Benini, J. Randall, T. Sanderson and M. Savardi, "Artificial Intelligence in hair research: A proof-of-concept study on evaluating hair assembly features," *International Journal of Cosmetic Science,* vol. 43, no. 4, pp. 405-418, 2021.

[3] T. Kim , C. Chung , Y. Kim , S. Park , K. Kim and J. Choo, "Style Your Hair: Latent Optimization for Pose-Invariant Hairstyle Transfer via Local-Style-Aware Hair Alignment," *Lecture Notes in Computer Science ,* p. 188–203). , 2022.

[4] T. Zhentao, C. Menglei, C. Dongdong, L. Jing, C. Qi, Y. Lu, T. Sergey and Y. Nenghai, "MichiGAN: multi-input-conditioned hair image generation for portrait editing," *ACM Transaction on Graphics,* vol. 39, no. 4, pp. 95:1-95:13, 2020.

[5] R. Saha, B. Saha, F. Shkurti, G. W. Taylor and P. Aarabi, "LOHO: Latent Optimization of Hairstyles via Orthogonalization," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2021.

[6] Z. Peihao, R. Peihao, J. Femiani and P. Wonka, "Barbershop: GAN-based Image Compositing using Segmentation Masks," arXiv, 2021 .

[7] A. Kumar, A. Kaur and M. Kumar, "Face detection techniques: a review," *Artificial Intelligence Review,* vol. 52, no. 2, pp. 927-948, 2019.

[8] L. Yann, B. Yoshua and H. Geoffrey, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[9] I. J. Goodfellow, M. Mirza, B. Xu, J. P. Abadie, D. W. Farley, S. Ozair and Y. Bengio, "Generative Adversarial Networks," arXiv, 2014.

[10] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2020.

[11] M. J. Page, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ,* p. p. n71, 2021.

[12] G. J. Jang, Q. Man and Y. I. Cho, "Development of a Hairstyle Conversion System Based on Mask R-CNN," *Electronics ,* vol. 11, no. 12, 2022.

[13] C. Chung, T. Kim, H. Nam, S. Choi, G. Gu, S. Park and J. Choo, "HairFIT: Pose-Invariant Hairstyle Transfer via Flow-based Hair Alignment and Semantic-Region-Aware Inpainting," *arXiv,* 2022.

[14] M. S. Abdallah and Y. I. Cho, "Virtual Hairstyle Service Using GANs & Segmentation Mask (Hairstyle Transfer System)," *Electronics ,* vol. 11, no. 20, 2022.

[15] P. Zhu, R. Abdal, J. Femiani and P. Wonka, "HairNet: Hairstyle Transfer with Pose Changes," *Computer Vision ,* vol. 13676, 2022.

[16] M. Pektas, B. Gecer and A. Ugur, "Efficient Hair Style Transfer with Generative Adversarial Networks," *arXiv,* 2022.

[17] K. Sasikarn, P. Pakkapon, S. Patsorn and S. Supasorn, "StyleGAN Salon: Multi-View Latent Optimization for Pose-Invariant Hairstyle Transfer," *arXiv,* 2023.

[18] Q. Man, Y. I. Cho, S. G. Jang and H. J. Lee, "Transformer-Based GAN for New Hairstyle Generative Networks," *Electronics,* vol. 11, no. 13, 2022.

[19] T. Wei, D. Chen, W. Zhou, J. Liao, Z. Tan, L. Yuan, W. Zhang and N. Yu, "HairCLIP: Design Your Hair by Text and Reference Image," *CoRR,* 2021.

[20] A. Šubrtová, J. Čech and V. Franc, "Hairstyle Transfer between Face Images," *16th IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 1-8, 2021 .