BareFace Restoration:

Using ResNet-34 for Makeup Removal

Final Year Project Final Report

A 4th Year Student Name

Uong Khanh Duy

Ngo Sach Trung

Nguyen Hai Dang

Associate Professor

Phan Duy Hung

Bachelor of Artificial Intelligence

Hoa Lac campus - FPT University

18 December 2023

ACKNOWLEDGEMENT

We would like to thank our instructor, Phan Duy Hung for his patience and time, and for instructing and advising us enthusiastically.

We would like to thank all at my University, FPT, for giving us the best environment to study and grow over the years.

We would like to thank our classmates in AI1601, for letting us meet amazing people and learn a lot from them.

We always remember our family's encouragement and support. Thanks to them, we have the will, the energy and the confidence to pursue our goals.

DECLARATION

We declare that the work in this dissertation titled "BareFace Restoration: Using ResNet-34 for Makeup Removal" has been carried out by our research with FPT University Artificial Intelligence Department. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signed:___ ___ ___ HDang ______________________

Date:___18/12/2023____

Uong Khanh Duy, Ngo Sach Trung, Nguyen Hai Dang and full qualifications

FPT University

Table of Contents

List of Figures

List of Table

List of Abbreviations and Acronyms

Abbreviations	**Meaning**
AI	Artificial Intelligence
U-Net	U-shaped neural network
ResNet	Residual Networks
ResNet-34	Residual Network with 34 layers
LSTM	Long Short-Term Memory
ReLU	Rectified Linear Unit
Res34UNet	Residual Networks with U-Net architecture
Conv2D	Convolutional 2D
GeLU	Gaussian Error Linear Unit
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure

ABSTRACT

We currently investigate the use of ResNet-34 as an encoder in the U-Net architecture, in the field of decoupling. This problem poses a major challenge because cosmetics obscure basic facial features, which is important in applications in many fields of security, entertainment and social networks. By effectively exploiting deep learning techniques to automatically remove makeup from facial images. The algorithm enhances the performance and feature extraction capabilities inherent in ResNet-34. Through testing, the results have shown that this new makeup removal model is very effective.

It helps to remove makeup that not only simplifies the process but also contributes significantly to advancing computer vision applications. This investigation represents an important step forward in the development of smart solutions for image enhancement and potential applications. The architectural framework of U-Net with ResNet-34 as the encoder contributes significantly to achieving these advances.

Keywords : Makeup Removal; U-Net; ResNet-34; Res34Unet

1. INTRODUCTION

With the development of the cosmetic industry worldwide, the need for women to use daily cosmetics is increasing. In Vietnam, the proportion of women using beauty cosmetic products has increased from 76% to 86% between 2018 and 2022. Especially nearly 70% of individuals aged 25-30, not only women but also men often use it [1]. Using cosmetics for yourself daily and weekly is a habit to make yourself more beautiful. Makeup has become an indispensable part of most women's daily lives, it helps them be confident and maintain a beautiful appearance when going out to work or out every day.

Since the advent of cosmetics, a significant and widespread challenge has emerged, covering the natural contours of the face. This phenomenon has created formidable obstacles for various security and social media sectors, hindering the ability to recognize and differentiate faces. Imagine a situation where individuals are navigating social networks or law enforcement agencies are looking for a specific person armed with only a photo as reference. The layers of make-up, used to such an extent, made it difficult for anyone, even seasoned police officers, to accurately identify individuals. One more example, there is a female office worker at the company who regularly uses the attendance system using facial recognition. However, every day she changes a different makeup style, from light makeup to heavy makeup. dark. So facial recognition may have some difficulty in identifying her every day. This has created problems for the facial recognition system, making the attendance process difficult and inaccurate. body. To solve this problem, we have experimented with a task that can support the face recognition process, which is the makeup removal technique.

In addition to solving specific problems in the corporate environment, makeup removal techniques also bring many benefits to the fields of security and social networks. With increasing emphasis on privacy, it is extremely important to ensure that facial recognition systems can operate accurately and securely. The makeup removal technique not only improves the accuracy of the identification process but also enhances security by ensuring that information about personal appearance is not confused through daily makeup changes. This is an important step forward for the development and application of facial recognition technology in areas such as security and human resource management.

Makeup removal models are much more complex than makeup models, partly due to the diverse nature of cosmetics and partly due to makeup styles. Previous models used pairs of images with and without makeup, which yielded commendable results but still had many limitations. Previous models were built complex and deep, these models were trained on large data sets and large systems. If these models are applied on small data sets and resource-limited systems, they are difficult to implement and can easily lead to overfitting on small data sets, and excessive complexity in their implementation must be avoided. Training deep convolutional makeup removal models.

To comprehensively address these challenges, we experiment with using Res34Unet which is a combination of U-Net and ResNet-34 to remove makeup from facial images. Therefore, let's take advantage of the advantages gained from the two models U-Net and ResNet-34. A slight innovation in our approach involves re-characterization of part of the available dataset with the aim of developing a deeper understanding of makeup styles across makeup styles daily of the European Union using BeautyGAN: Individual-level facial makeup transfer using Deep Generative Adversarial Network method [2].

It is hoped that combining the U-Net model with ResNet-34 and the extended dataset will yield more positive results in makeup removal and on small systems. It contributes to promoting the development of applications in the fields of security, entertainment, social networks and especially the field of computer vision.

2. RELATED WORKS

2.1. Makeup

In recent years, the application of recognizing and executing images have been widely researched. Especially, the rise of new research on human faces contributes an important part for recognizing human faces. In addition to facial recognition applications, researchers also explore problems related to creating an overlay on facial images also known as the makeup problem. A variety of applications were found, which use models from research articles to create simulated makeup on the human face.

In the field of researching and evaluating makeup models, some studies can be mentioned such as: [3] Makeup Transfer Using Support Vector Regression published from 2018 using SVR and PCA to generate a virtual makeup facial image considering personal facial features. From researching this paper successfully to estimate the texture after makeup considering the facial features. [2] BeautyGan: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network is also used in makeup problems by using Generative Adversarial Network. Facial makeup transfer aims to makeup human facial images from specific makeup styles. The results achieved when using BeautyGan are positive. And in this study, we use BeautyGan to generate additional data from the non-makeup and makeup images.

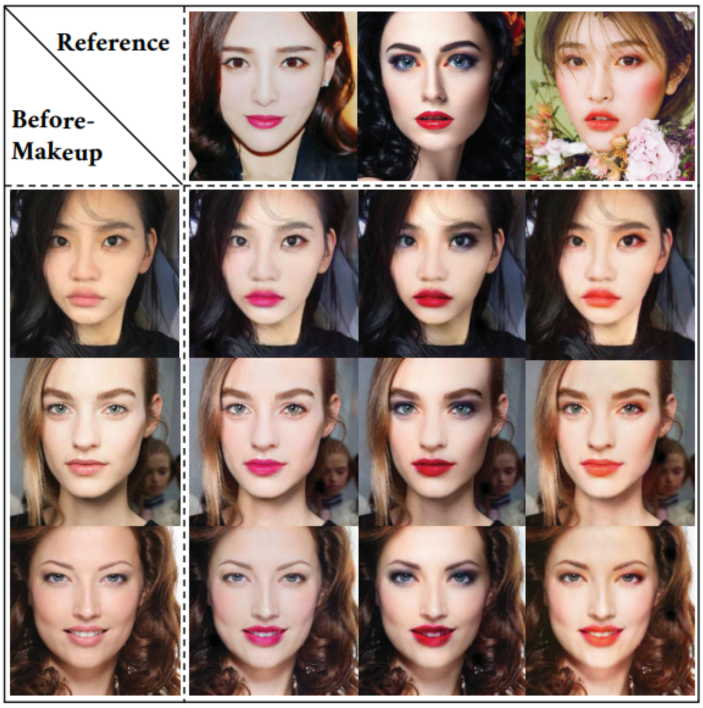


Figure 2.1: Example results of BeautyGan model for makeup transfer. Three makeup styles on reference images (top row) are translated to three before-makeup images (left column). Nine generated images are shown in the middle [2].

2.2. Makeup removal

On the contrary, compared to makeup synthesis [2], [3], [4], [5] makeup removal is a more challenging inverse task that has received less attention. But there are also articles that are very valuable in terms of research. However, the efficacy attained by those models is somewhat limited. Makeup Removal via Bidirectional Tunable De-makeup Network [6] published in 2018 and using a deep learning based method for removing makeup effects in facial images. To make makeup removal tasks less complicated, previous techniques [6], [7], [8] require training on image pairs of a face with makeup and a bare face. By implementing previous techniques and the model BeautyGan can generate makeup images, therefore this study will build the dataset consisting of a large number of image pairs of makeup and non-makeup. One of the troubles is that the training of deeper convolutional neural networks poses escalating difficulties, especially with makeup removal needed to build a deep neural network. Therefore, this model needs a solution to make the training part easier. Deep Residual Learning for Image Recognition [9] or ResNet was undertaken with the objective of addressing the aforementioned challenge. ResNet was present to ease the training of networks that are substantially deeper than those used previously. In [9], an experiment shows that more importantly, the 34-layer ResNet exhibits considerably lower training error and is generalizable to the validation data. U-Net also known as Convolutional Networks for Biomedical Image Segmentation [10] is also a recommendation to enhance the training phase. This approach entails the deployment of a network and training strategy that intricately leverages data augmentation to optimize the efficiency of the available annotated samples.

3. CONTRIBUTION

This study focuses on testing image reconstruction methods, with the input data set selected as a set of facial images of a woman wearing makeup. Our goal is to try to experimentally build a model that can remove women's facial makeup and achieve good results with a small architecture. We also use a method to evaluate the similarity between the output image and the target image, specifically Structural Similarity Index Measurement (SSIM) [11] and Maximum Signal-to-Noise Ratio (PSNR) [12] as two indexes of evaluation of the BTD-Net model [6]. These two indices are commonly used in image reconstruction problems because they can evaluate the similarity between two images by comparing brightness and contrast and calculating the difference between pixels. For the makeup remover image dataset, the study U-Net architecture with ResNet-34 encoder and skipping connection layers yielded satisfactory results. By gradient descent disappearance and enhancing the model's learning ability, these skip connections enhance the model's resistance to deformation. In particular, the skip connection layers help the model extract image features while still retaining information about the position in the image, avoiding the case of the output image being distorted in the wrong position. Includes many popular makeup looks for women, such as light makeup, heavy makeup, and special effects makeup, all of which are considered when selecting input images.

The effectiveness of the makeup remover image reconstruction method and its potential for use in mobile applications including image editing and face recognition are enormous. It can help clearly identify the real faces of both men and women. Furthermore, removing makeup can help people be more confident in their true beauty.

4. METHODOLOGY

4.1. U-Net

U-Net [10] constitutes a convolutional neural network architecture explicitly crafted for segmentation within the domains of image analysis and especially biomedical image segmentation. The U-Net architecture differs from traditional convolutional neural networks. In the International Symposium on Biomedical Imaging (ISBI) challenge [10] on structural segmentation Neurons in electron microscopy stacks, U-Net is considered to have higher performance than previous models. On the paperswithcode page, U-Net is the most used semantic segmentation model as of the time of the report with nearly 1800 papers used. The U-Net architecture consists of 2 main part of an encoder and decoder.

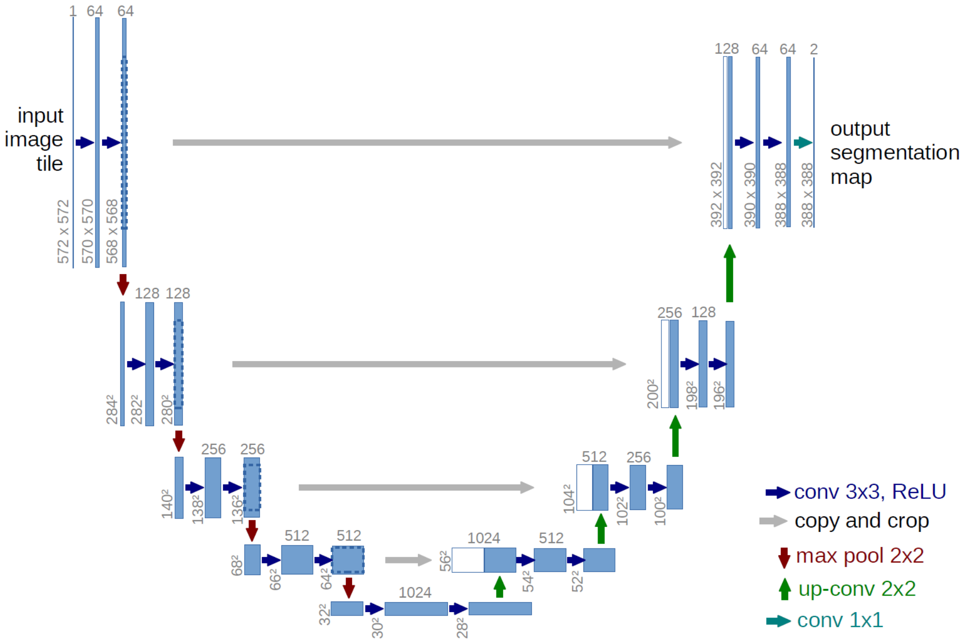


Figure 4.1: U-Net architecture [10].

Encoder:

The encoder or contracting path is the upper part of the U-Net architecture. It is combined using multiple convolutional layers and max pooling and spatial dimensionality reduction operations. This process helps to extract features from the input image, which is necessary for segmentation tasks.

Decoder:

The decoder or expansive path is the lower part of the U-Net. It involves upsampling the feature maps and applying convolutional layers to reconstruct spatial dimensions gradually. Skip connections play a crucial role, concatenating feature maps from the contracting path to corresponding layers in the expansive path. These connections preserve fine details during upsampling, addressing information loss challenges.

One of the distinctive features of U-Net is the use of skip connection, that is a key innovation of U-Net. These connections directly link corresponding layers in the encoder and decoder paths. Skip connections enable the network to retain fine-grained details and help mitigate the vanishing gradient problem.

4.2. ResNet-34

ResNet [9], short for Residual Network, is a type of deep convolutional neural network architecture. It was introduced in 2016 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Was created to solve the problem of deeper neural networks being more and more difficult to train. With 2016 papers using ResNet, this was the most used convolutional neural networks reported on paperswithcode. The architecture of ResNet is characterized by its use of residual blocks.

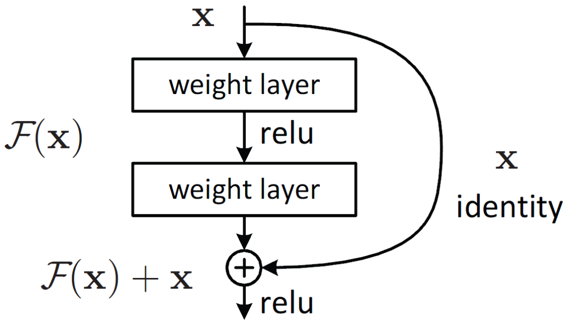
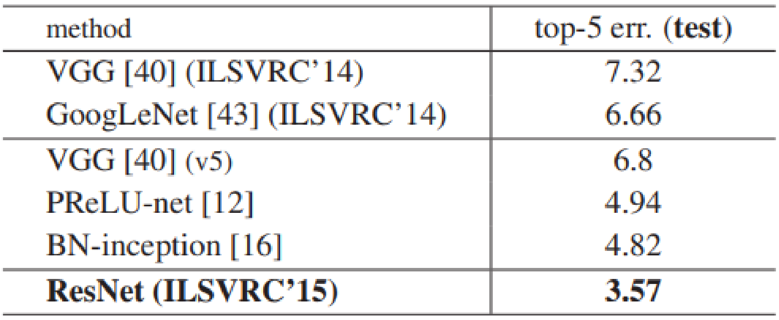

x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 4.2: Residual block [9]

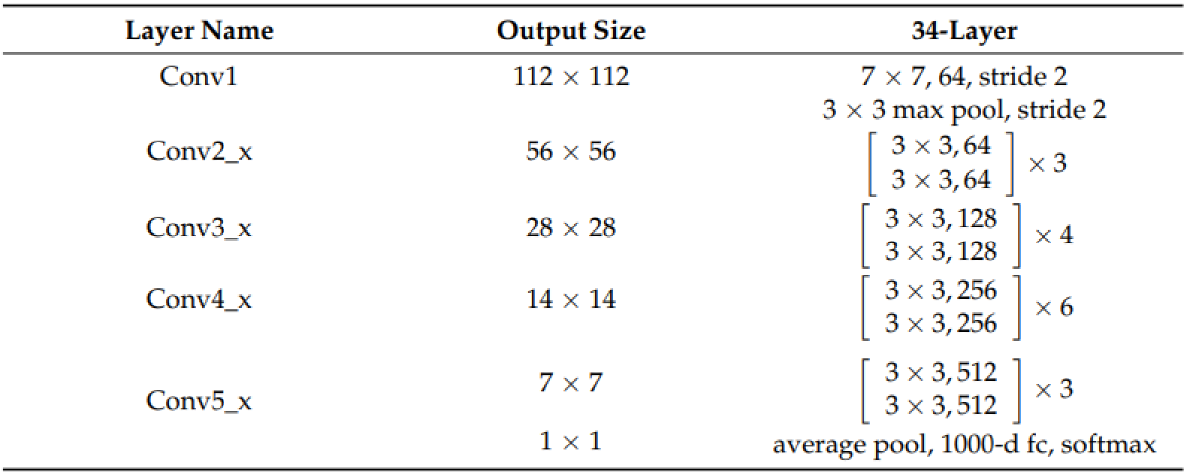
Residual block consists of two main paths: identity path and residual path. The identity path constitutes a direct shortcut connection, transmitting the input of the block directly to the output. This design facilitates the acquisition of an identity mapping by the network when deemed as the optimal transformation. In essence, if the identity path proves to be the most suitable means of representing the input, the network can dynamically adapt by adjusting the weights of the residual path to zero, effectively rendering it an identity function. The residual path contains a series of transformations applied to the input, which are learned during the training process. And the result is the ResNet substantially deeper than those used previously and make the training step better. The experimental results in Figure 4.3 show that the ResNet model is much better than previous models. Additionally, ResNet provides very easy optimization and can gain accuracy from considerably increased depth.



method	top-5 err. (**test**)
VGG [40] (ILSVRC'14)	7.32
GoogLeNet [43] (ILSVRC'14)	6.66
VGG [40] (v5)	6.8
PReLU-net [12]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

Figure 4.3: Error rates (%) of ensembles. The top-5 error is on the test set of ImageNet and reported by the test server [9].

ResNet-34 [13] is a specific variant of the Residual Network (ResNet) architecture. It was introduced by Kaiming He, et al., in the paper "Deep Residual Learning for Image Recognition"[9] in 2015. ResNet-34 [10] is part of the ResNet family, which is known for its deep architecture and the use of residual blocks to facilitate the training of very deep neural networks. ResNet-34 [13] consists of 34 layers, including convolutional layers, batch normalization, ReLU activation functions, max pooling layers, and fully connected layers.



Layer Name	Output Size	34-Layer
Conv1	112×112	$7 \times 7, 64$, stride 2
Conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax

Figure 4.4: The structure of ResNet-34 [13].

ResNet-34 [13] primarily uses basic building blocks called residual blocks. Each residual block consists of two convolutional layers with batch normalization and ReLU activation functions. The key innovation in Residual Networks is the use of skip connections (or shortcut connections) that allow the network to learn residual mappings, making it easier to train very deep networks. Skip connections in ResNet-34 enable the gradient to flow more directly during backpropagation, addressing the vanishing gradient problem and helping with the training of deep neural networks.

ResNet-34 is considered a computationally efficient variant of the ResNet architecture [9] while still providing good performance. It is often used in computer vision tasks such as image classification, object detection, and segmentation. The choice of ResNet-34 versus other ResNet variants depends on factors such as the available computational resources, the size of the dataset, and the specific requirements of the task at hand.

4.3. Res34UNet

Res34UNet, short for U-Net architecture using ResNet-34 and skip connection. That is the combination of 2 models: ResNet-34 and U-Net. Res34UNet will execute through 2 steps encoder using ResNet-34, decoder. In accordance with the U-Net [10] architectural framework, the encoder and bridge components have undergone substitution, with the ResNet-34 architecture being integrated in lieu of the original structures.

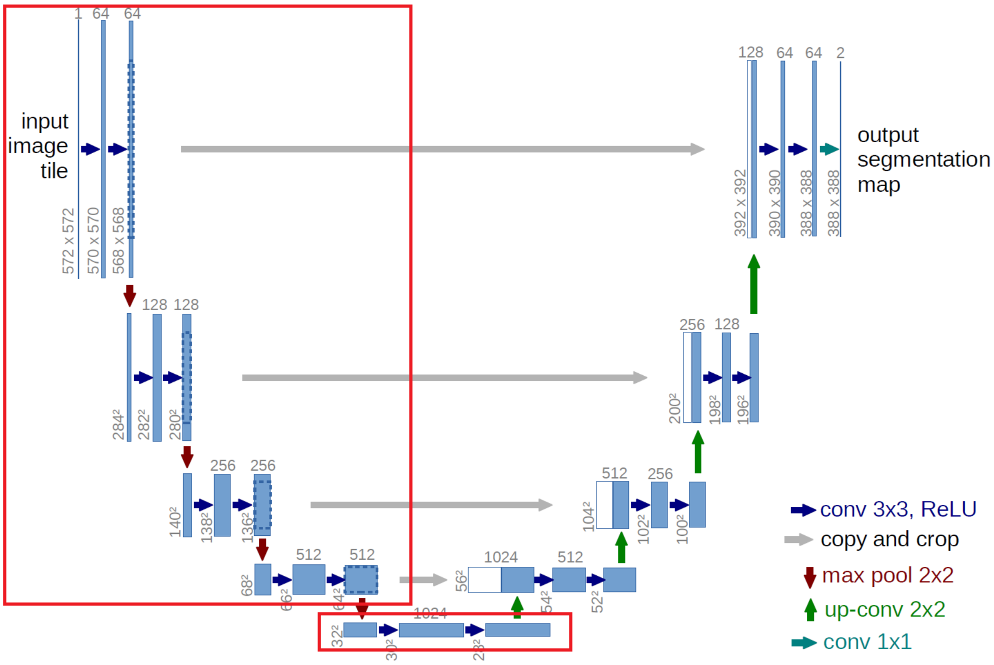


Figure 4.5: Substitute the encoder and bridge (red square) components within the U-Net architecture with the ResNet-34 architecture. [10]

In figure 4.5, the decoder part (outside the red square) added layers, custom to suit the problem and output size is equal to the input size. The U-Net architecture leverages skip connections to preserve fine-grained details and address issues related to vanishing gradients. In order to maintain these advantages while integrating

ResNet-34 with U-Net, we tried to using skip connections to each block of ResNet-34, as illustrated in Figure 4.6.

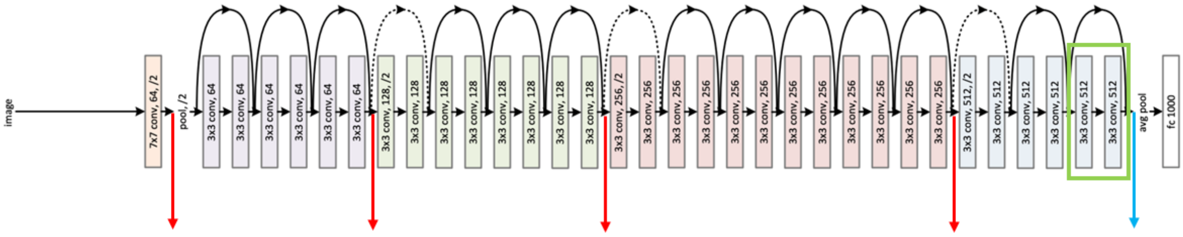


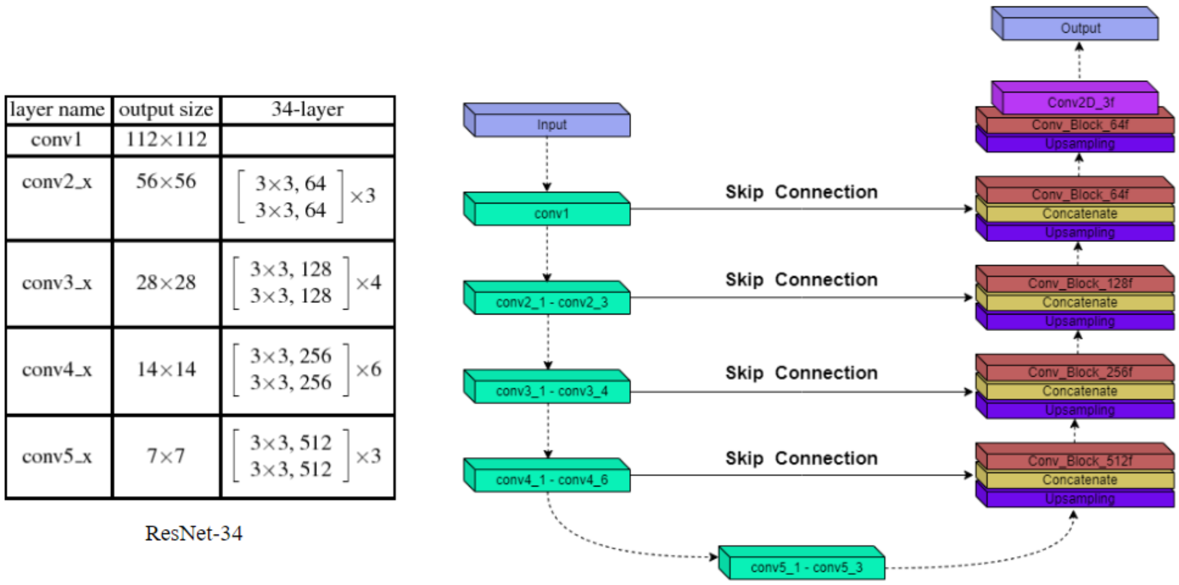
Figure 4.6: The ResNet-34 model along with the incorporation of skip connections, has been implemented as a replacement for the encoder segment within the U-Net architecture.

The ResNet-34 architecture needs 4 skip connections and a part to replace the bridge. And in the ResNet-34 architecture, we added 4 skip connections after conv1, conv2_x, conv3_x and conv4_x. There are four layer names of ResNet-34. And conv5_x from conv5_1 to conv5_3 replaced the bridge in U-Net architecture. In the layer conv5_3 which is the last layer of conv5_x is the output of the encoder path and transmitted directly to the decoder path to reproduce images.

The structure of U-Net and ResNet-34 is refined with regard to the construction of identity blocks consisting of two Conv2D layers using the activation function GeLU [14], replacing the usual ReLu [15] because GeLU [14] can helps the model learn more complex relationships than ReLU [15] and adds Instance Normalization [16] to increase the learning speed of the model.

The GeLU [14] (Gaussian Error Linear Unit) activation function is a variant of the ReLU [15] (Rectified Linear Unit) that aims to capture more complex dependencies and potentially improve the representational power of neural networks. It was introduced by Dan Hendrycks and Kevin Gimpel in their paper "Gaussian Error Linear Units (GELUs)" [16] in 2016.

In addition, in medicine, it is necessary to learn accurately, so padding is 0. In de-makeup, you only need to learn enough to remove makeup, so padding is equal to "same" to avoid complicating the problem.



layer name	output size	34-layer
conv1	112×112	
conv2_x	56×56	$\begin{bmatrix}3\times3, 64\\ 3\times3, 64\end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\ 3\times3, 128\end{bmatrix}\times4$
conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\ 3\times3, 256\end{bmatrix}\times6$
conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\ 3\times3, 512\end{bmatrix}\times3$

ResNet-34

Figure 4.7: U-Net architecture using ResNet-34 and skip connections.

In Figure 4.7, the layers in ResNet-34 replace each block of encoder and bridge. U-Net is capable of paying attention to details during the decoding process, aiding the model in generating non-makeup images while preserving essential facial features. The ResNet-34 model facilitates the recognition of both large and small features during makeup removal. It improves feature encoding and extraction from makeup images compared to the default encoder of U-Net. Increasing the number of channels in the encoder section enhances feature extraction, and reducing the number of feature maps lost positional information due to multiple pooling layers. To generate an image, the size of the feature map must increase, and simultaneously, to reduce computational load, we decrease the channels, requiring up-sampling. Up-sampling we use learning based (transposed convolution) instead of interpolation based. Although interpolation based is simpler and less complicated, it is not effective when it is necessary to recreate small and complex details of the face after removing makeup while Transposed convolution can learn to reproduce fine and complex details of a face based on complex relationships in the data. However, this process still loses significant image information, hence the need to skip connections to incorporate some feature information from the corresponding level layer in the encoder to the decoder, ensuring reduced information loss. The accuracy of predictions relies on evaluating the differences between the predicted non-makeup image and the actual non-makeup image. Although mathematical comparisons are commonly used, it is essential to acknowledge that achieving complete accuracy is not always possible. Human perception is necessary to determine the similarity between the predicted image and reality.

The combination of the ResNet-34 and U-Net models leverages the distinctive advantages inherent in each architecture. ResNet-34 incorporates residual blocks and skip connections, thereby mitigating the vanishing gradient problem and facilitating the successful training of deep networks. The residual blocks, in particular, play a pivotal role in information transfer throughout the network, thereby enhancing generalization across diverse tasks. Conversely, U-Net excels in semantic segmentation tasks, owing to its U-shaped architecture that facilitates the seamless transfer of high-resolution information. This attribute proves particularly valuable in scenarios involving the transmission of human facial image data characterized by varying resolutions. Moreover, U-Net's ability to accommodate input images of diverse sizes enhances its adaptability across different applications. The synergistic integration of these two models not only yields discernible improvements but also engenders heightened efficiency, ensuring that constraints on computational resources minimally impact the efficacy of the training process.

5. EXPERIMENTS RESULT AND CONCLUSION

5.1. Data Collection

The dataset used in this study, referred to as the Makeup Removal Dataset, consists of two types of images: non-makeup and makeup-applied. It is divided into three subsets for training, validation, and testing, with 5000 image pairs in the training set, 500 image pairs in the validation set, and another 500 image pairs in the testing set. Each image has a size of 224x224x3.

Initially, the dataset was sourced from Kaggle, where we obtained everyday images of Europeans with and without makeup. However, it was observed that this dataset had many makeup images that closely resembled non-makeup images. Therefore, we chose to employ a makeup transfer approach using the pre-existing BeautifulGAN model [2] to generate additional and regenerate makeup images data from the non-makeup from Kaggle.

Makeup transfer involves applying makeup to non-makeup images using makeup-styled images as references, creating makeup-applied output images. Subsequently, we use these output images as inputs and non-makeup images as targets for the Makeup Removal Dataset.

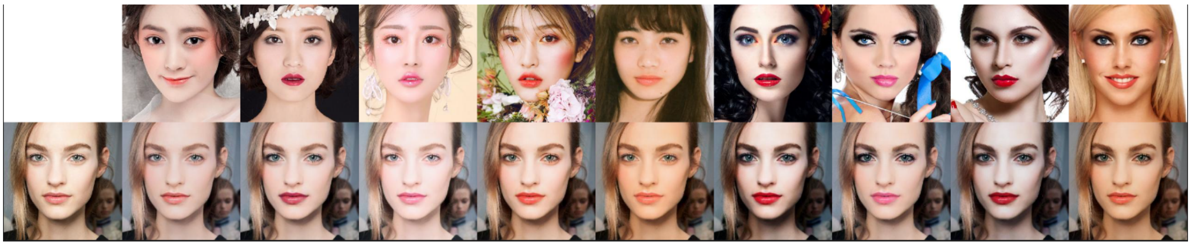


Figure 5.1: Makeup Transfer Example

In figure 5.1, the top row shows makeup style photos, while the leftmost photo in the bottom row is the original photo, the remaining photos are photos after makeup transformation.

Below is an example of the dataset after generating. The top row is the generated makeup from the non-makeup and the bottom row is the non-makeup image preserved.

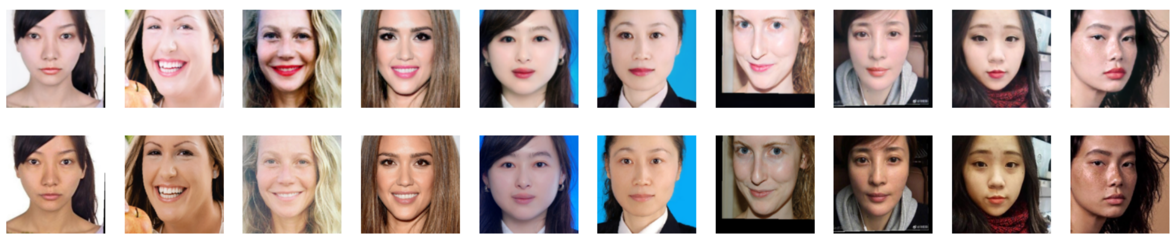


Figure 5.2: Dataset example after generating

5.2. Experiments

We have put in significant effort and thoroughness in generating data, implementing the Res34UNet model, training, testing, and analyzing and visualizing results. In this section, we will provide detailed information about the implementation process.

Frameworks and libraries: Our Res34UNet model is coded using Python programming language and the TensorFlow framework. With TensorFlow, we can seamlessly read, preprocess and feed the training data and easily implement the Res34UNet model during training and inference. TensorFlow also includes a Tensorboard tool, which enables users to conveniently track training progress and view training loss, testing loss, and evaluation scores during training. Apart from TensorFlow, we also use a few other support libraries such as (Table 5.1):

Table 5.1: List of libraries support

List of libraries support	Describe
OpenCV	read, write, and process images
Matplotlib	draw graphs and analyze results
Tensorflow addons	contains functions not yet available in Tensorflow core
Tf-models-official	contains pretrain models and optimization algorithms

Environment: use a laptop during the implementation process, run code tests, process data synthesis, debug and create additional images with many different styles. During model training, we run on Google Colab to shorten training time because it has a more specialized configuration for machine learning tasks.

Code: We reused a pre-existing codebase from the U-Net model and ResNet-34 to build our own model. In addition, we developed all of the source code ourselves, including data generation, data loading, training-testing-inference procedures, and visualization.

Hyperparameters: Table 5.2 provides a detailed breakdown of the hyperparameters used during our training process. While most of the parameters remain the same as in the baseline, there are a few changes that we made:

Each dataset contains a predetermined number of training and testing data. Because the data allocation for each dataset is quite large, as almost all makeup images need to focus only on the face, we set the input size for the model to (224 × 224). Then train the model with a learning rate of 0.01, batch size 32, and num epochs 330. We apply the AdamW optimizer with the same hyperparameters and use the first 10% of steps to gradually increase the learning rate until the initial learning value then the learning rate will be adjusted according to AdamW.

Table 5.2: Hyperparameters of Res34UNet model

Hyperparameters	
Input shape	224 x 224
Initial learning rate	0.01
Batch size	32
Num epochs	330
Optimizer	AdamW

Evaluation methods: In this study, two indices: SSIM [11] is used to compare the structural similarity between two images and PSNR [12] calculates the error rate between pixels in the image, expressed in terms of decibels.

5.3. Result and Analysis

In this research, we developed a deep learning model aimed at effectively removing makeup from images. The model is constructed using the Res34UNet architecture and is trained on a dataset comprising 5,000 image pairs, each consisting of a makeup-applied and a non-makeup image. Aligning with established practices in image restoration research, we utilized two widely employed image similarity metrics, PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index), to evaluate the model's performance. The mean values of PSNR and SSIM were calculated based on 500 validation images from the dataset. PSNR assesses image compression quality, with values exceeding 30 generally considered visually identical. SSIM measures illumination, contrast, and structural similarity, with higher values indicating closer resemblance to the original reality (ranging from -1 to 1, where 1 signifies identical images).

Following 330 epochs of training, the model achieved the following evaluation metrics.

Note: The three graphs below (Figure 5.4, 5.5, 5.6) including the image on the left are the values for all 330 epochs. The image on the right ignores the large values of the first 3 epochs to make the graph easier to see.

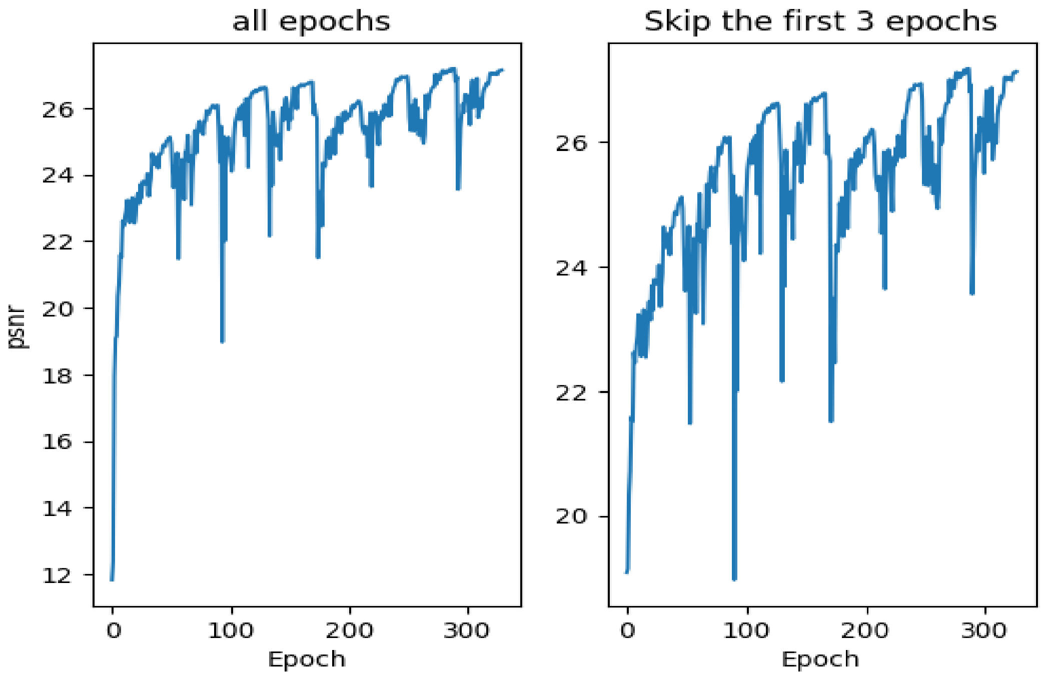


Figure 5.3: PSNR curve from Res34UNet model on validation dataset.

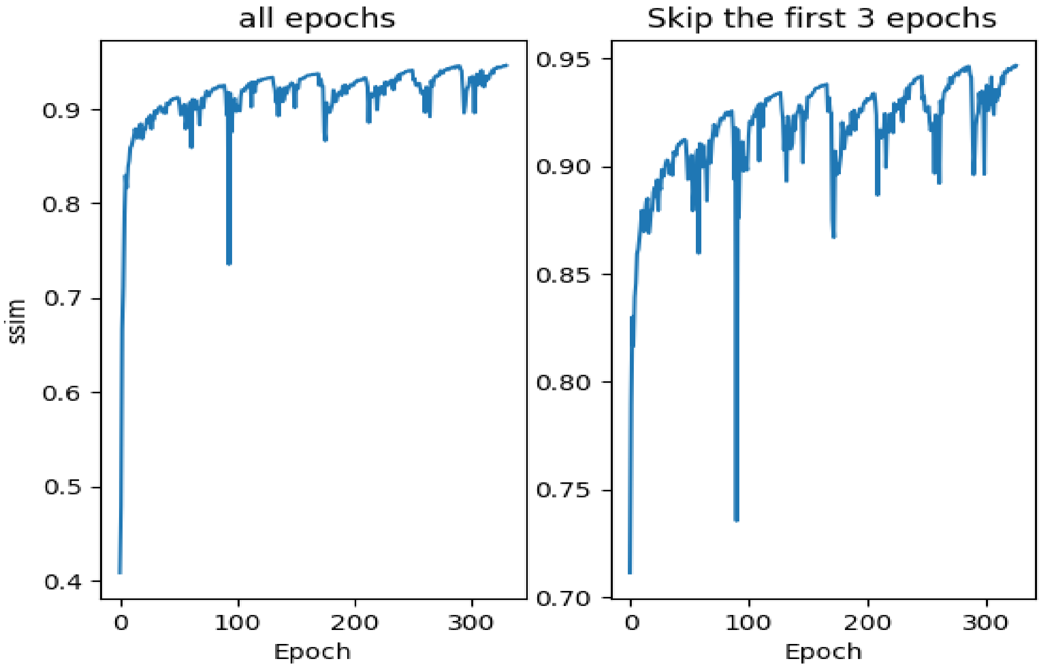


Figure 5.4: SSIM curve from Res34UNet model on validation dataset.

During the initial stages of model training, we used a personal laptop to estimate the training time for each epoch. However, due to the large dataset and the complexity of the Res34UNet architecture, the training duration reached up to 3 hours per epoch. Recognizing this as impractical, we transitioned to using Google Colab for training.

Colab imposes a daily training time limit of no more than 12 hours per account. Consequently, instead of training for 330 epochs continuously, we divided the training into multiple sessions, each consisting of a smaller number of epochs (e.g., 50, 40, 40, ..., 40). This led to irregular fluctuations in the results graph at certain epochs, as the model needed initial epochs to adapt to previously saved weights. Fortunately, all models exhibited improvement, as depicted in the loss graph.

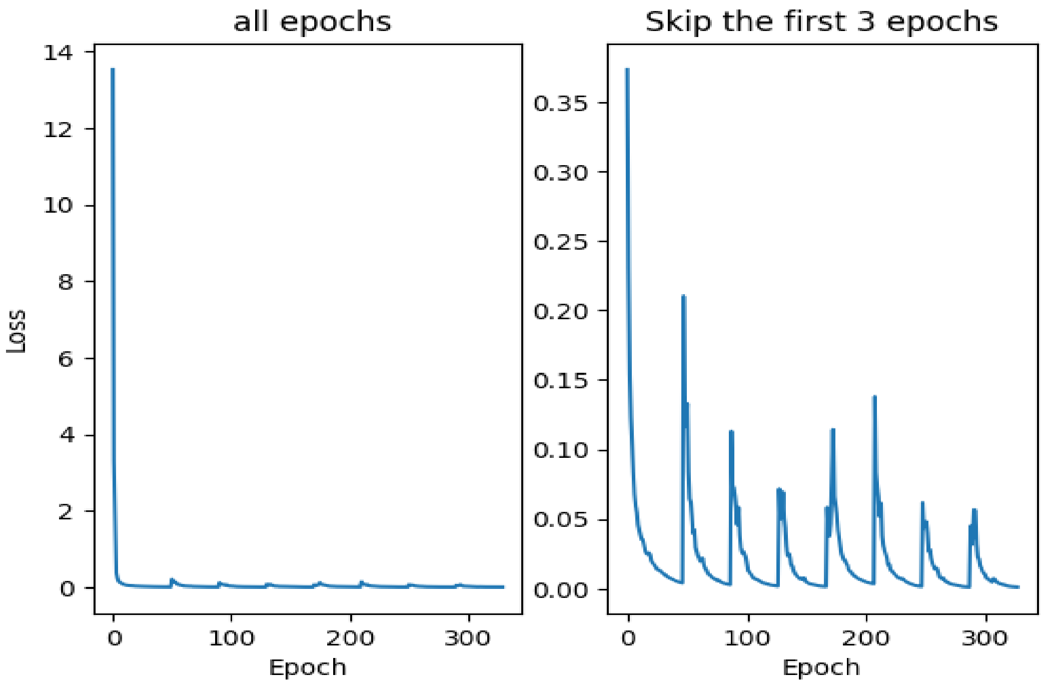


Figure 5.6: Loss curve from Res34UNet model on training dataset.

On the makeup face, the average makeup removal result of Res34UNet with PSNR is 27.133 and SSIM is 0.947, showing close similarity to the original face, the accuracy is considered quite good.

The experimental results show that Res34UNet can accurately adjust the makeup components on the face and keep the original shape of the non-makeup face. But the Res34UNet model also has both advantages and disadvantages:

Advantage :

+ High performance

+ Applicability to many different types of makeup photos

Disadvantages:

- A large amount of training data is needed
- High requirements for hardware systems used for training

Makeup removal images in the test dataset were processed by the training Res34UNet model. In Figure 5.8, in the first image, with the input image created from the no-makeup image in the middle through the makeup transfer model, the output predicted image has almost all the makeup removed, leaving only a light makeup layer compared to the image input but clearly see the difference between the two images. The predicted output image has PSNR = 28.26 and SSIM = 0.956, which is close to the original image, so considered a good predicted image.

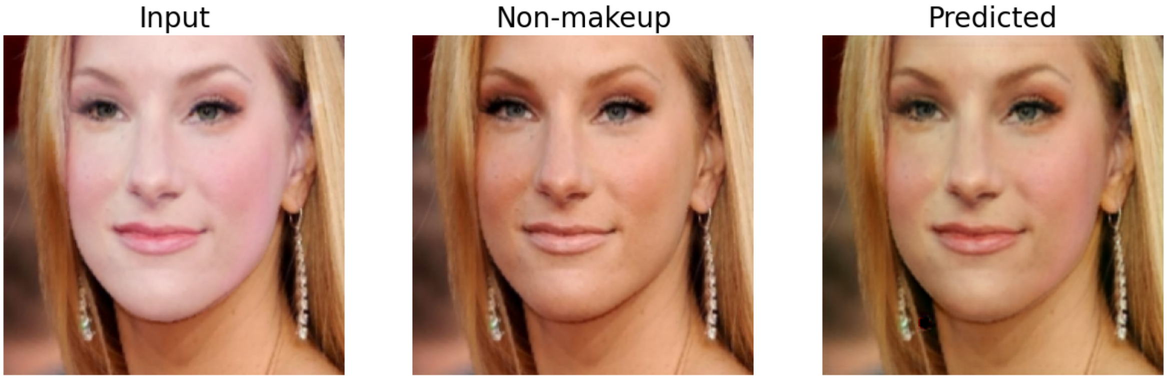


Figure 5.8: Image predicted by Res34Uet model. Image order: input makeup image, target non-makeup image, predicted makeup removal image.

5.4. Conclusion and Future Works

We present a contribution to the field of artificial intelligence and image processing using U-net that uses ResNet-34 as an encoder to perform the facial makeup removal task. It creates a model with higher makeup removal precision than previous models. We conducted many tests and obtained impressive results to prove the effectiveness of U-net with the ResNet-34 model in removing makeup. The results of this research can be applied in many fields such as beauty and identification through machines.

This presentation still has many potential directions to develop, enhance and expand the scope of future applications in optimizing performance by researching and using optimization techniques that help the system operate effectively. Starting from a complex foundation, distinguishing between real people and fake people, supporting makeup applications, researching deep learning structures to be able to apply makeup on specific parts, to improve accuracy.

So we can see a lot of potential in improving and applying the U-Net makeup removal method in the future not only with ResNet-34 but also with the possibility of combining with other models for each specific application and efficiency.

References

1. “Tong quan thi truong my pham tai Viet Nam – Xu huong thi truong my pham 2024.” GMPc Viet Nam

2. Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. 2018. BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network. In Proceedings of the 26th ACM international conference on Multimedia (MM '18). Association for Computing Machinery, New York, NY, USA, 645–653. https://doi.org/10.1145/3240508.3240618

3. A. Tsuji, M. Seo, Y. Muto and Y. -W. Chen, "Makeup Transfer Using Support Vector Regression," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 289-293, doi: 10.1109/FSKD.2018.8687239.

4. Q. Gu, G. Wang, M. T. Chiu, Y. -W. Tai and C. -K. Tang, "LADN: Local Adversarial Disentangling Network for Facial Makeup and De-Makeup," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 10480-10489, doi: 10.1109/ICCV.2019.01058.

5. W. Jiang et al., "PSGAN: Pose and Expression Robust Spatial-Aware GAN for Customizable Makeup Transfer," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 5193-5201, doi: 10.1109/CVPR42600.2020.00524.

6. C. Cao, F. Lu, C. Li, S. Lin and X. Shen, "Makeup Removal via Bidirectional Tunable De-Makeup Network," in IEEE Transactions on Multimedia, vol. 21, no. 11, pp. 2750-2761, Nov. 2019, doi: 10.1109/TMM.2019.2911457.

7. S. Wang and Y. Fu, “Face behind makeup,” in The Thirtieth AAAI Conference on Artificial Intelligence, 2016.

8. Y.-C. Chen, X. Shen, and J. Jia, “Makeup-go: Blind reversion of portrait edit,” in Proc. of Int’l Conf. on Computer Vision (ICCV), 2017.

9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.90

10. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

11. Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004, doi: 10.1109/TIP.2003.819861.

12. K. Joshi, R. Yadav and S. Allwadhi, "PSNR and MSE based investigation of LSB," 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), New Delhi, India, 2016, pp. 280-285, doi: 10.1109/ICCTICT.2016.7514593.

13. Mingyu Gao;Dawei Qi;Hongbo Mu;Jianfeng Chen; (2021). A Transfer Residual Neural Network Based on ResNet-34 for Detection of Wood Knot Defects . Forests, (), –. doi:10.3390/f12020212

14. Hendrycks, Dan and Kevin Gimpel. “Gaussian Error Linear Units (GELUs).” arXiv: Learning (2016): n. Pag.

15. Agarap, Abien Fred. “Deep Learning using Rectified Linear Units (ReLU).” ArXiv abs/1803.08375 (2018): n. pag.

16. Ulyanov, Dmitry et al. “Instance Normalization: The Missing Ingredient for Fast Stylization.” ArXiv abs/1607.08022 (2016): n. pag.