FPT UNIVERSITY

CAPSTONE PROJECT PRESENTATION

FPT Education
FPT UNIVERSITY

CAPSTONE PROJECT

Topic
BareFace Restoration: Using ResNet-34 for Makeup Removal

Group: AIP490_G1
List of members:

- HE150829 Ngo Sach Trung
- HE150321 Nguyen Hai Dang
- HE150601 Uong Khanh Duy

Instructor: Dr. Phan Duy Hung

FPT Education
FPT UNIVERSITY

Table of Content

FPT Education
FPT UNIVERSITY

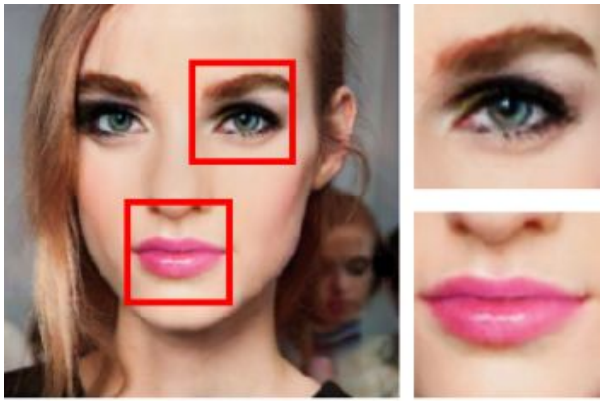
Introduction

1. Problem

FPT Education
FPT UNIVERSITY

Introduction

2. Related Works



BeautyGAN

- Makeup transfer using SVR[1]
- BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network[2]

1. 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.
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

FPT Education
FPT UNIVERSITY

Introduction

2. Related Works

- Makeup Removal via Bidirectional Tunable De-makeup Network [3]

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.

FPT Education
FPT UNIVERSITY

Introduction

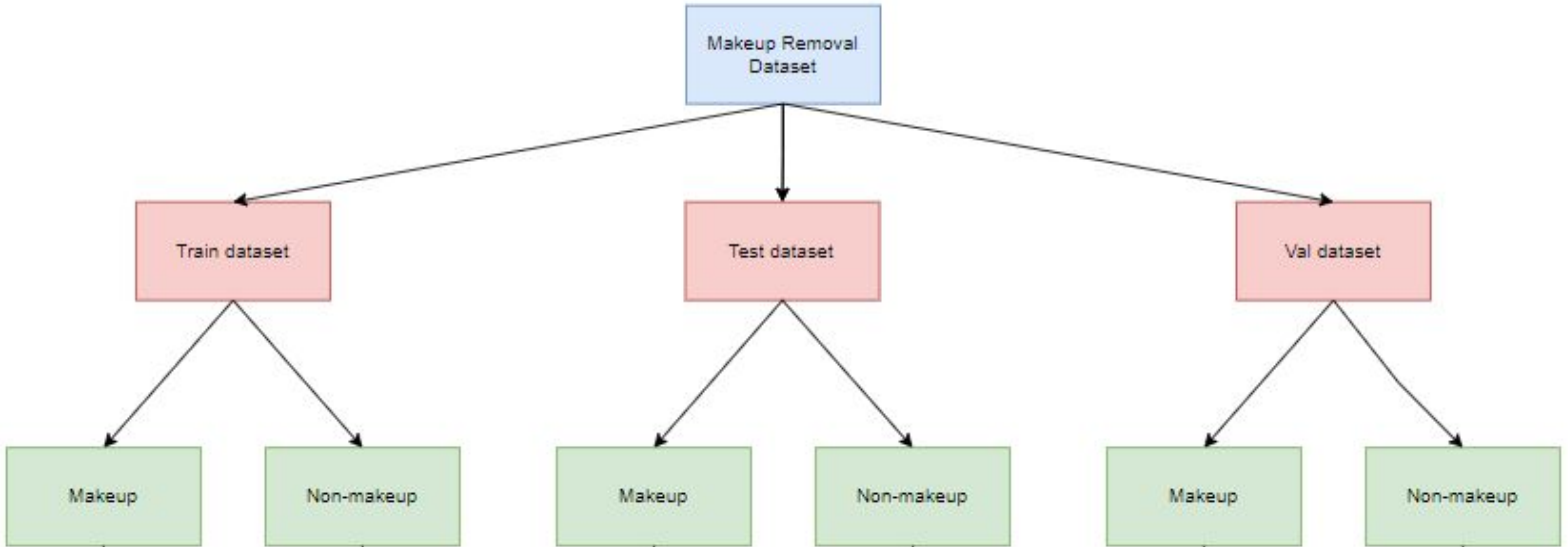
3. Objectives

The difficulties is that training deep convolutional neural networks is increasingly complex, and vanishing gradients must be avoided.

=>The utilization of a combined ResNet and U-Net network has been proposed to enhance the efficiency of training the makeup removal model, aiming to improve its accuracy and augment its learning capability.

FPT Education
FPT UNIVERSITY

Data Collection



Train : 5000 pairs

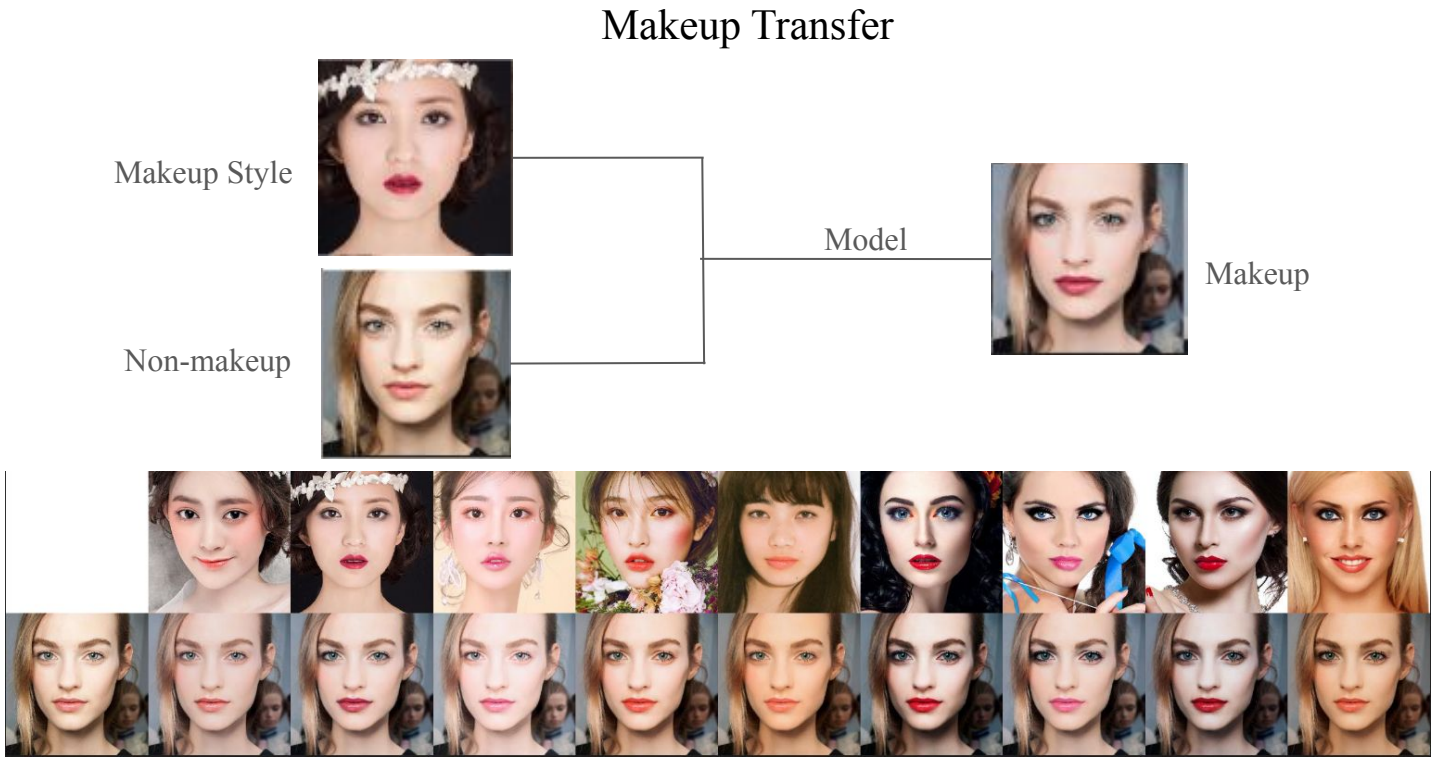
Test : 500 pairs

Validation : 500 pairs

Image size : 224x224x3

FPT Education
FPT UNIVERSITY

Data Collection



FPT Education
FPT UNIVERSITY

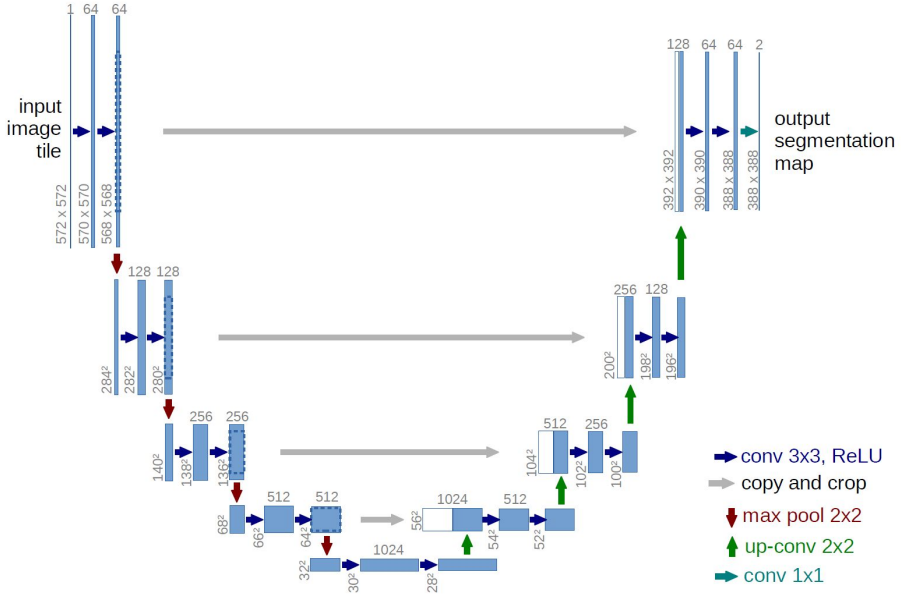
Data Collection

The first row is the makeup image after it was generated and the second row is the non-makeup image taken from kaggle

FPT Education
FPT UNIVERSITY

Methodology

1. U-Net

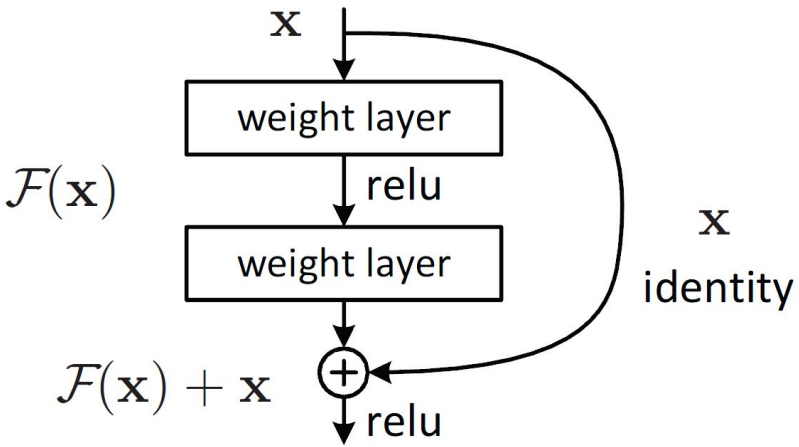


U-Net architecture

- U-Net was introduced in 2015.
- Designed for semantic segmentation and image analysis.
- Higher performance than previous methods.
- The U-Net architecture consists paths:
  - Encoder: extract features from the input images.
  - Decoder: reconstruct spatial dimensions gradually.
- Skip connections: retain fine-grained details and help mitigate the vanishing gradient problem.
- U-Net architecture is capable of working with inputs of different sizes.

Methodology

2. ResNet and ResNet-34



Residual block

- ResNet or Residual Network was introduced in 2015.
- Deeper neural networks are more difficult to train.
- ResNet primarily uses residual blocks.
- Residual block contain two paths:
  - Identity path: transmitting the input of the block directly to the output.
  - Residual path: contains a series of transformations and learned during the training process.
- ResNet is deeper and much better than previous models. Also make the training step better.
- Provides very easy optimization and gains accuracy from considerably increased depth.

Methodology

2. ResNet and ResNet-34

layer name	output size	34-layer
conv1	112×112	
conv2_x	56×56	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$

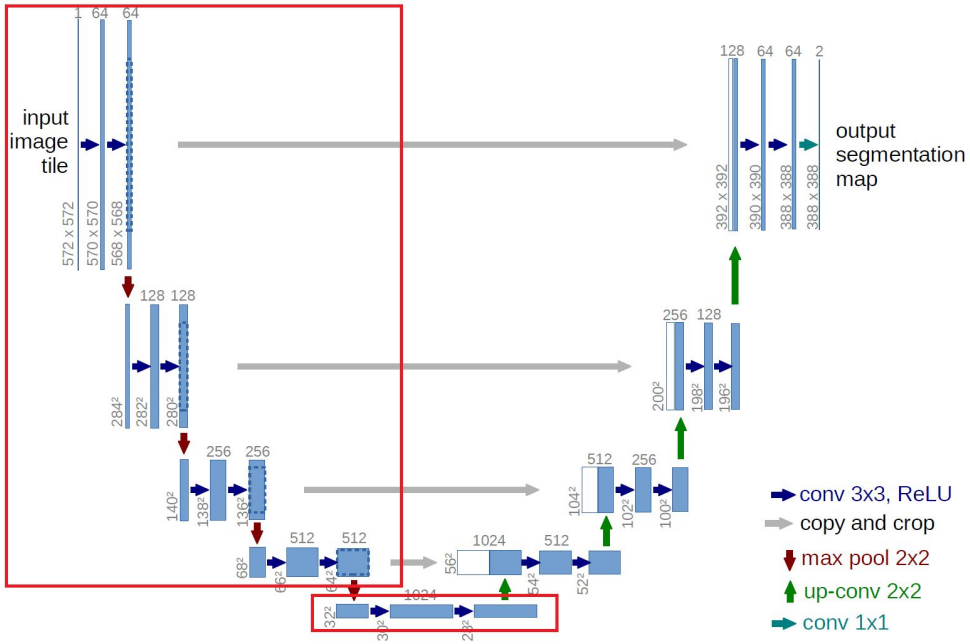
ResNet-34

- ResNet-34 is a specific variant of the Residual ResNet architecture. It is a part of the ResNet family.
- Deep architecture and the use of residual blocks to the training of very deep neural networks.
- ResNet-34 consists of 34 layers.
- Uses basic building blocks called residual blocks. Also using skip connections reduce vanishing gradient.
- ResNet-34 is a computationally efficient variant of the ResNet architecture while still providing good performance.

FPT Education
FPT UNIVERSITY

Methodology

3. Res34Unet



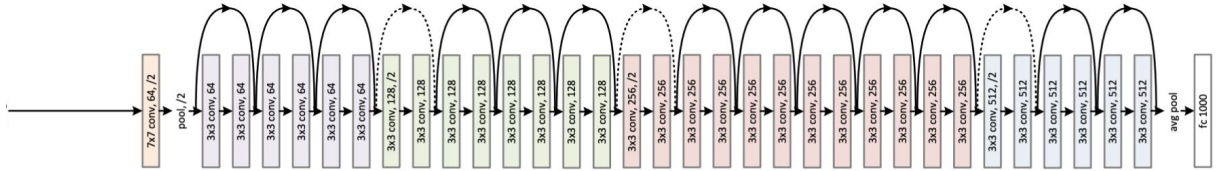
- U-Net architecture using ResNet-34.
- Combination of ResNet-34 and U-Net.
- Encoder using ResNet-34.
- The encoder and bridge components have undergone substitution with the ResNet-34 architecture
- In order to maintain these advantages while integrating ResNet-34 with U-Net.
- In U-Net architecture have skip connections directly from encoder to decoder. Therefore, ResNet-34 need add skip connections to keep the advantages of U-Net.

Methodology

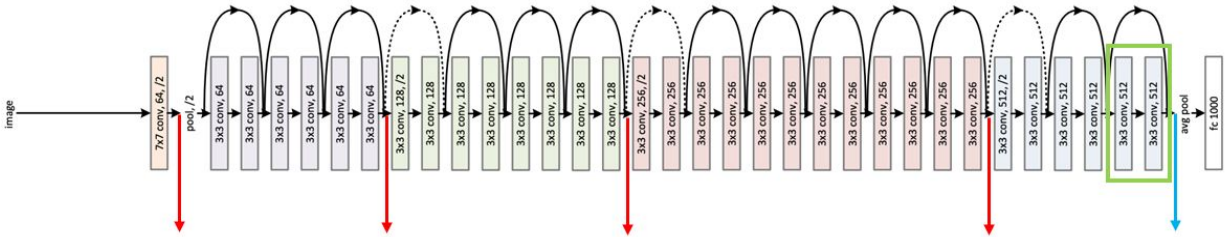
3. Res34Unet

layer name	output size	34-layer
conv1	112×112	
conv2_x	56×56	[3×3, 64; 3×3, 64] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×4
conv4_x	14×14	[3×3, 256; 3×3, 256] ×6
conv5_x	7×7	[3×3, 512; 3×3, 512] ×3

ResNet-34



ResNet-34



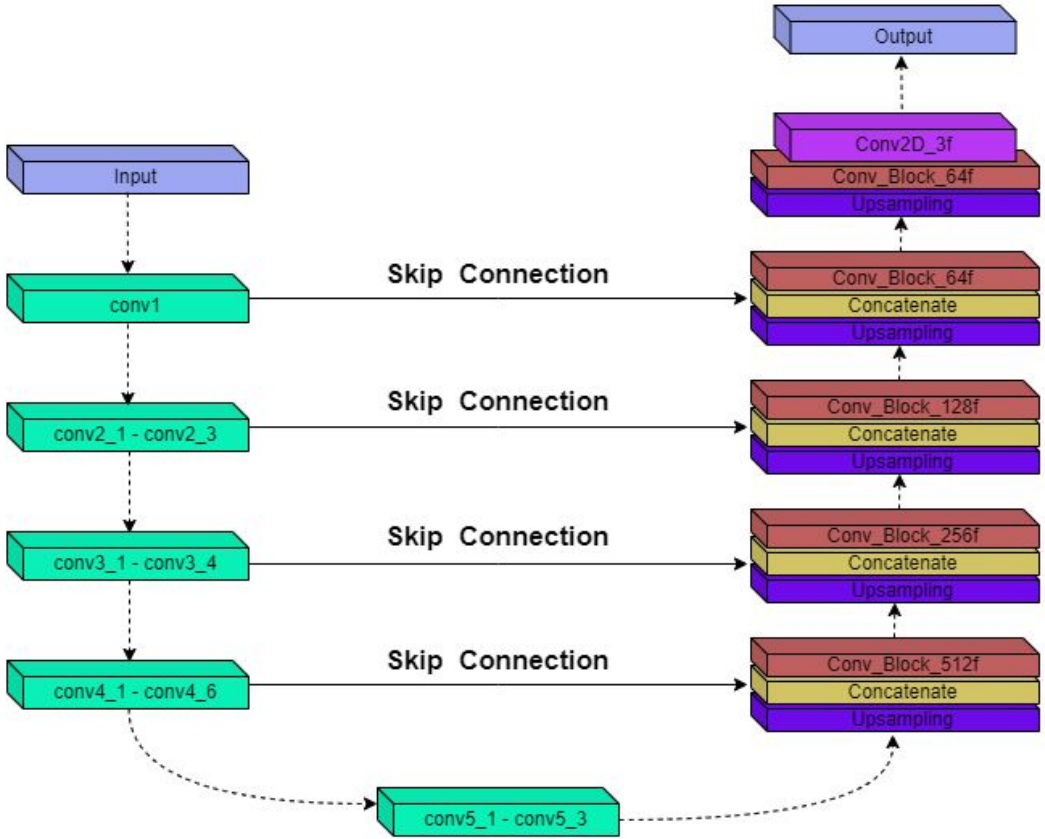
ResNet-34 added skip connections

Methodology

3. Res34Unet

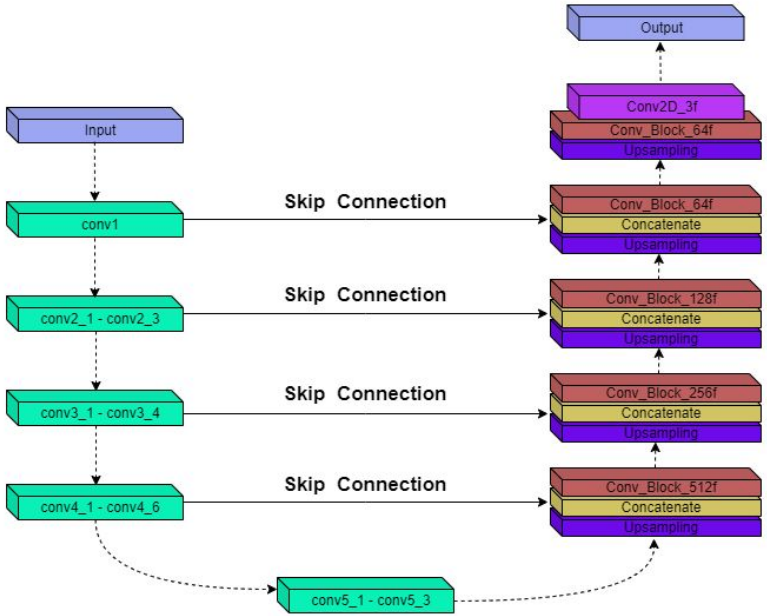
layer name	output size	34-layer
conv1	112×112	
conv2_x	56×56	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$

ResNet-34



Methodology

3. Res34Unet



- 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.
- U-Net excels in semantic segmentation tasks, owing to its U-shaped architecture that facilitates the seamless transfer of high-resolution information.
- Moreover, U-Net's ability to accommodate input images of diverse sizes enhances its adaptability across different applications.

FPT Education
FPT UNIVERSITY

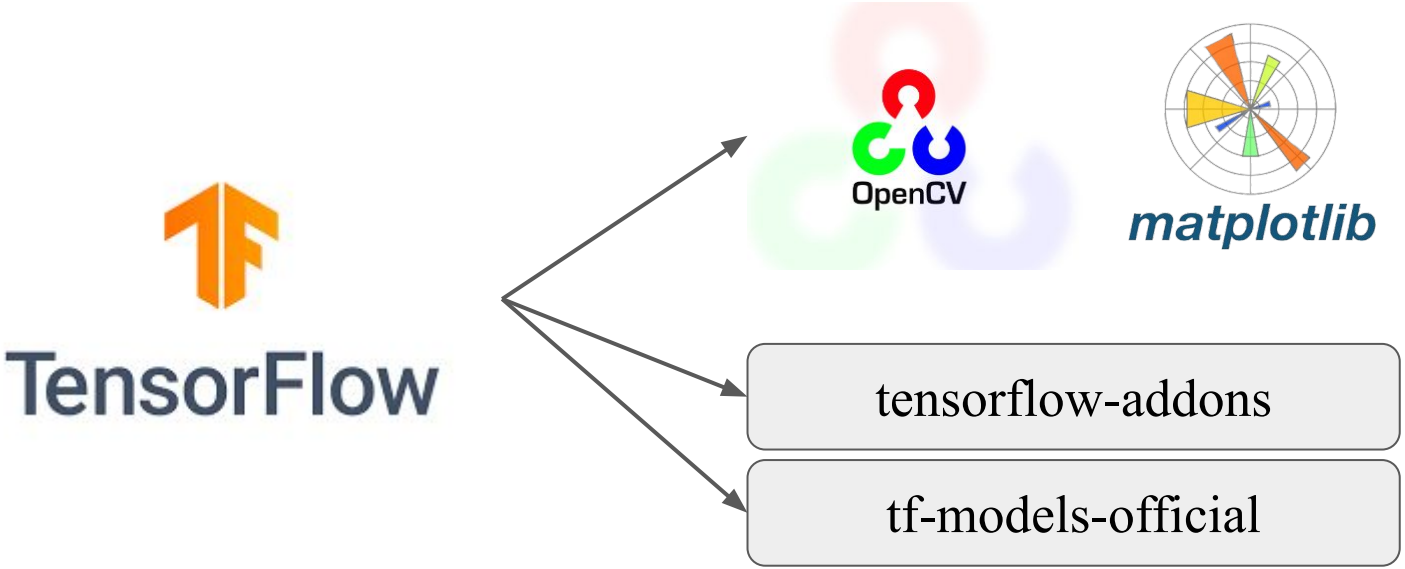
Methodology

3. Res34Unet

- Rectified Linear Unit ⟹ Gaussian Error Linear Units
- Instance Normalization, padding = “same"
- Custom the output to size 224x224 and 3 channels

FPT Education
FPT UNIVERSITY

Frameworks and libraries



FPT Education
FPT UNIVERSITY

Environment

- AMD Ryzen 5 (2.1GHz)
- Ram 16GB
- Radeon RX 560X 4GB VRAM

Google colab

- Intel Xeon (2.5GHz)
- Ram 12.7GB
- NVIDIA Tesla T4 16GB VRAM

FPT Education
FPT UNIVERSITY

Hyperparameters

Hyperparameters	
Initial learning rate	0,01
Batch size	32
Num epochs	330
Optimizer	AdamW

- Loss function: Mean squared error

- Evaluation index:
  + Peak signal-to-noise ratio (PSNR)
  + Structural similarity index measure (SSIM)

FPT Education
FPT UNIVERSITY

Evaluation

Peak signal-to-noise ratio (PSNR)

$$\mathrm{PSNR} = 20 \cdot log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

With:

$$\mathrm{MSE} = \frac{1}{m * n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

Structural similarity index measure (SSIM)

$$\mathrm{SSIM(x,y)} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

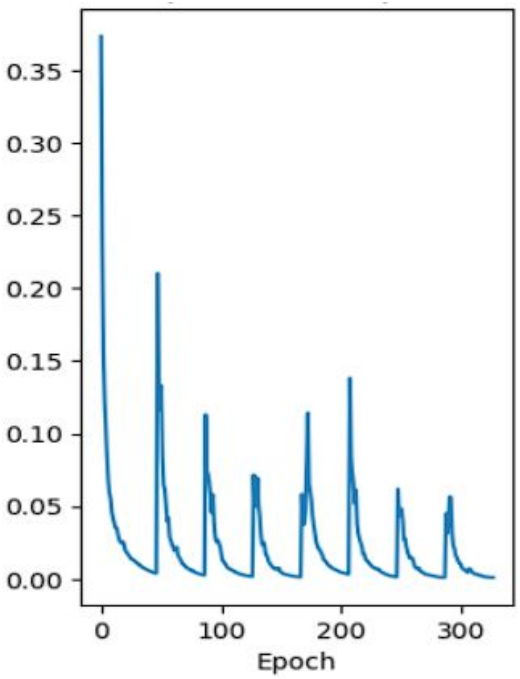
$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$

$k_1 = 0.01$ and $k_2 = 0.03$ by default.

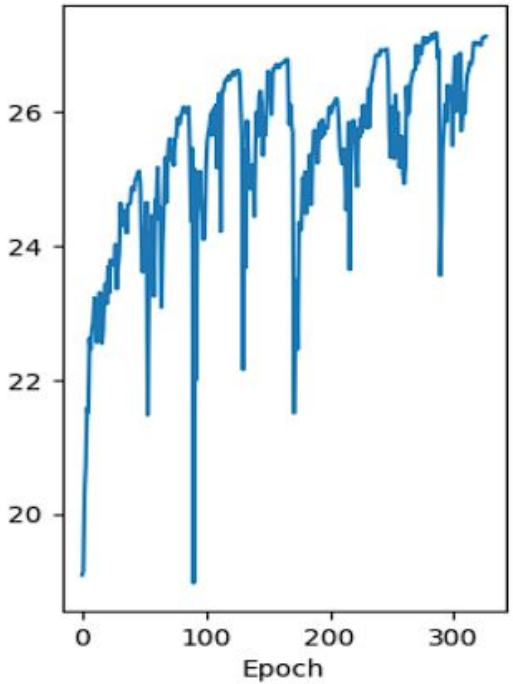
L: dynamic range of pixels (255 for 8-bit images)

FPT Education
FPT UNIVERSITY

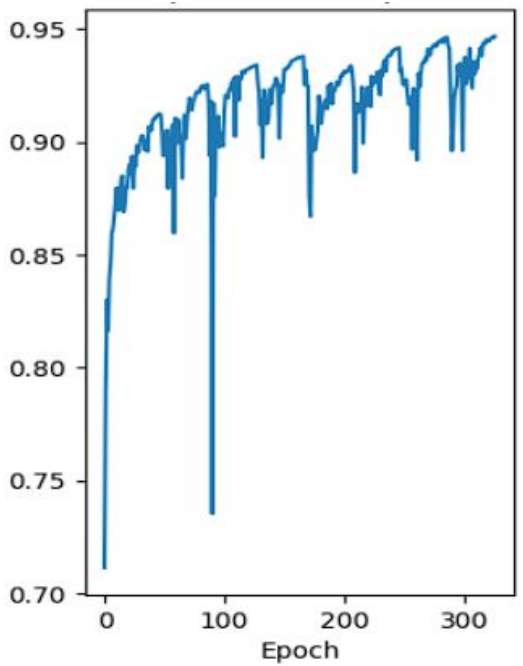
Result



Loss



PSNR



SSIM

FPT Education
FPT UNIVERSITY

Result

Values after 330 training epochs:

+ loss: 0,00105 + psnr: 27,133 + ssim: 0,947

APPLY DE-MAKEUP ON NON-MAKEUP FACES AND VIRTUAL MAKEUP FACES AND COMPARE THE RESULTS TO CYCLEGAN [10] AND RESIDUAL LEARNING [32] IN TERMS OF PSNR AND SSIM.

	Non-makeup		Makeup	
	PSNR	SSIM	PSNR	SSIM
CycGAN	23.0	0.873	22.4	0.855
Residual	22.2	0.860	22.0	0.850
BTD-Net	**25.1**	**0.929**	**24.3**	**0.907**

FPT Education
FPT UNIVERSITY

Result

The predicted output image has PSNR = 28.26 and SSIM = 0.956

Input

Non-makeup

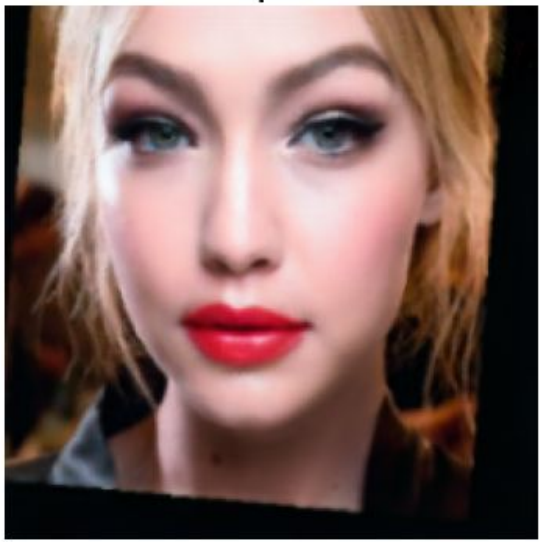
Predicted

FPT Education
FPT UNIVERSITY

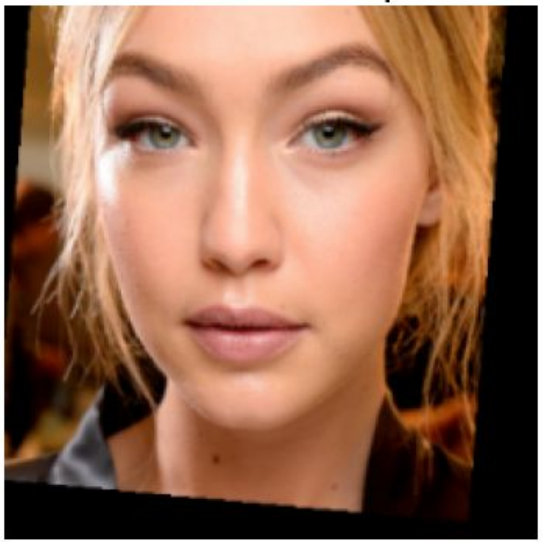
Result

The predicted output image has PSNR = 19.99 and SSIM = 0.801

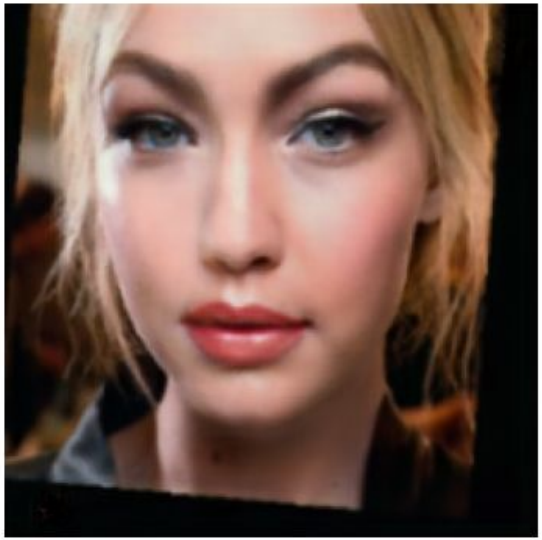
Input



Non-makeup



Predicted



FPT Education
FPT UNIVERSITY

Conclusion

Developed a model with high makeup removal accuracy ≈ 94.7%

Advantage :

+ High performance
+ Applicability to many different types of makeup photos
+ Easily customizable

Disadvantages:

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

FPT Education
FPT UNIVERSITY

Future Works

The results of this research can be applied in many fields such as beauty and identification through machines.

Application to real life

+ Model support for makeup application
+ Face recognition application, distinguishing between real and fake
+ Expand the model to remove makeup from different parts

Enhance performance

+ Use different optimization techniques.
+ Combine with other models
+ Search for more data sets to cover more types of makeup


thankyou