

HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.





ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

AUTOTOON

Automatic Geometric Warping for Face
Cartoon Generation

Nguyen Thi Hoang Linh - 20224324
Le Ngoc Huong - 20224315
Class: 152446

ONE LOVE. ONE FUTURE.

The logo graphic features a red background with a dense pattern of small white dots. A large, stylized white 'H' shape is formed by a path of dots, starting from the top left and curving down towards the bottom right. The word 'HUST' is written in a bold, white, sans-serif font, positioned to the left of the 'H' shape.

HUST

Outline

Introduction

Dataset

Model Architecture

Progress

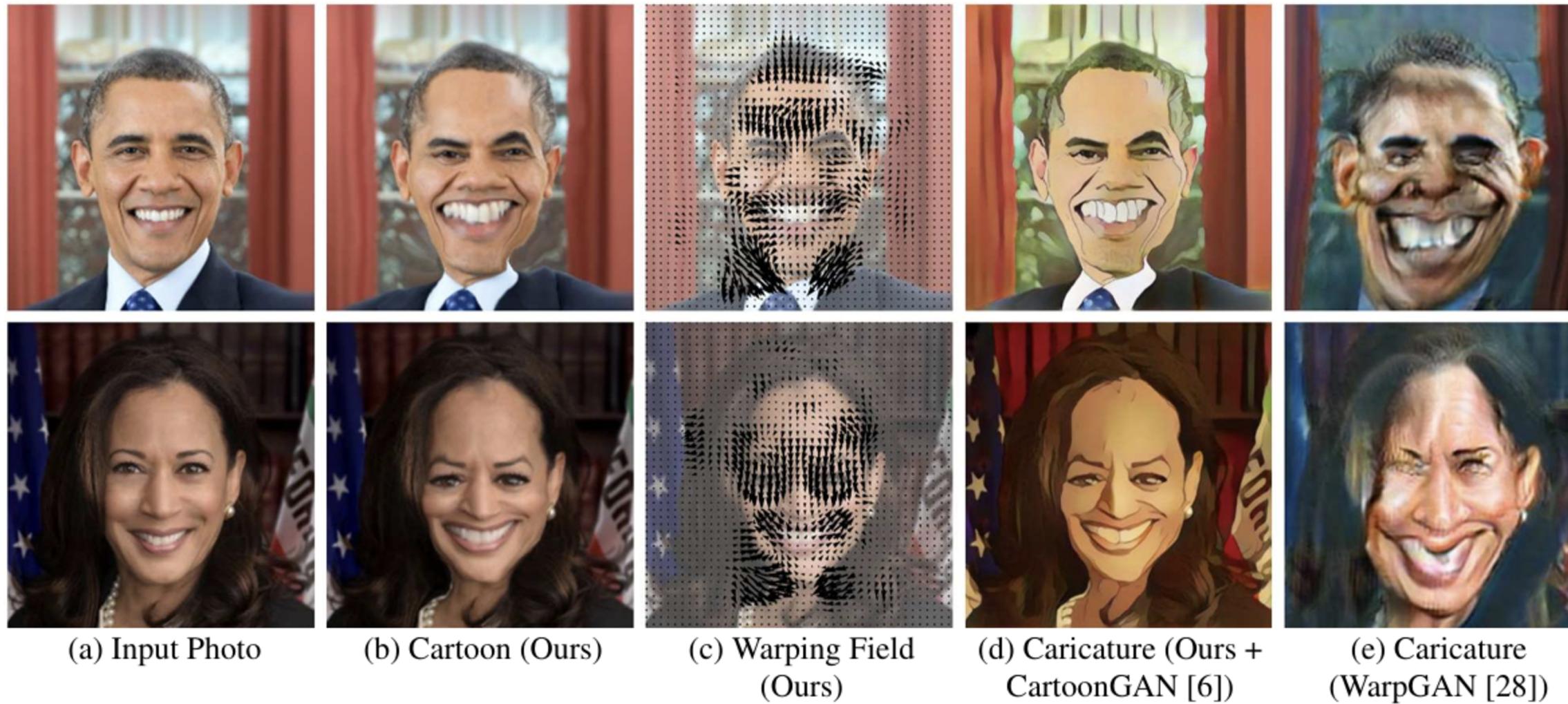
Evaluation

New Idea

Conclusion



Introduction



About this project:

Caricature: artists use creative exaggeration to accentuate unique facial traits

Input: a standard portrait photo

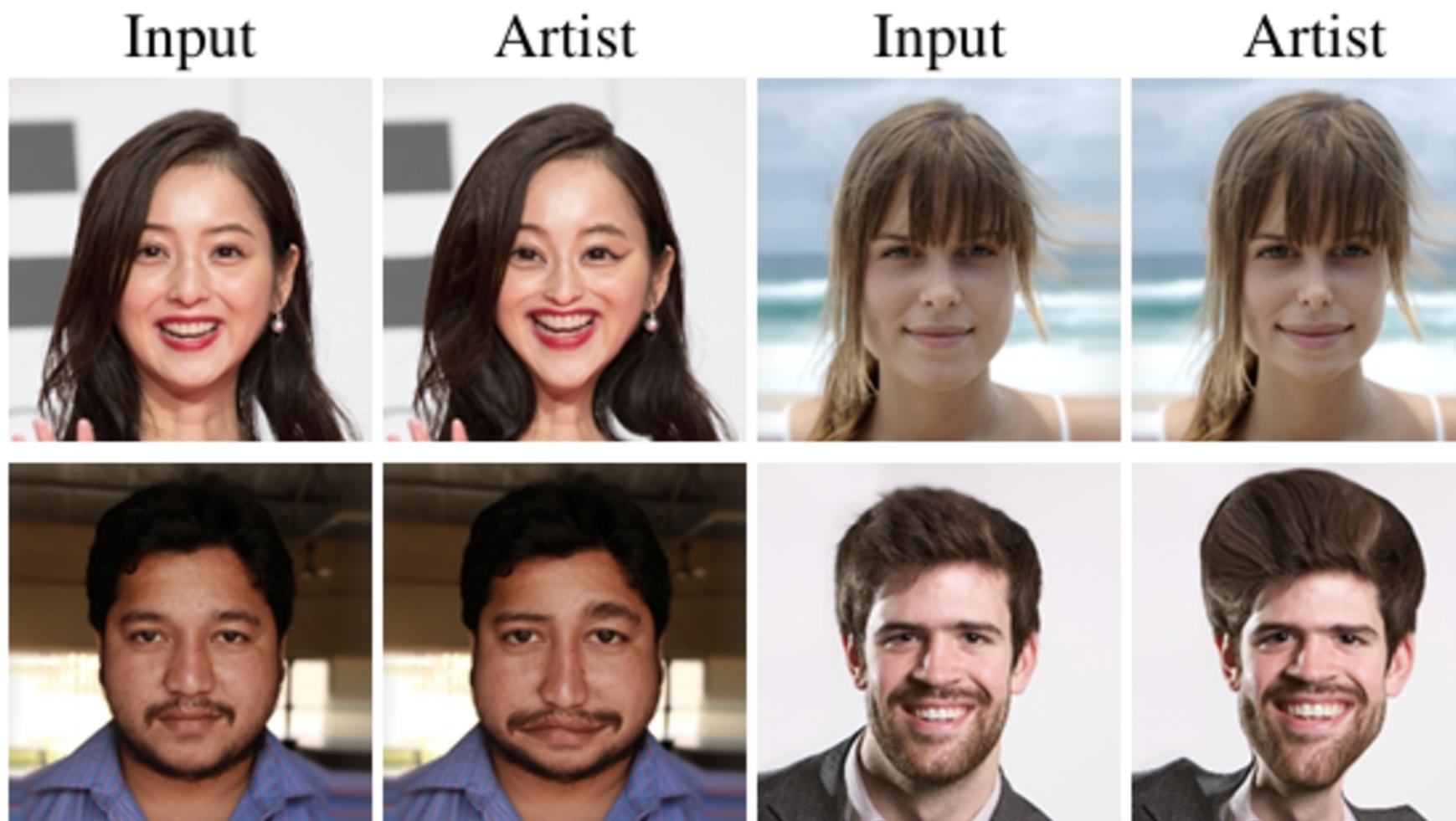
Output: an exaggerated cartoon version

Target: an automated, end-to-end pipeline

Step 1: Adding a geometric warp to the face to exaggerate features

Step 2: Stylized the twisted image for an artistic impact (style transfer).

Dataset



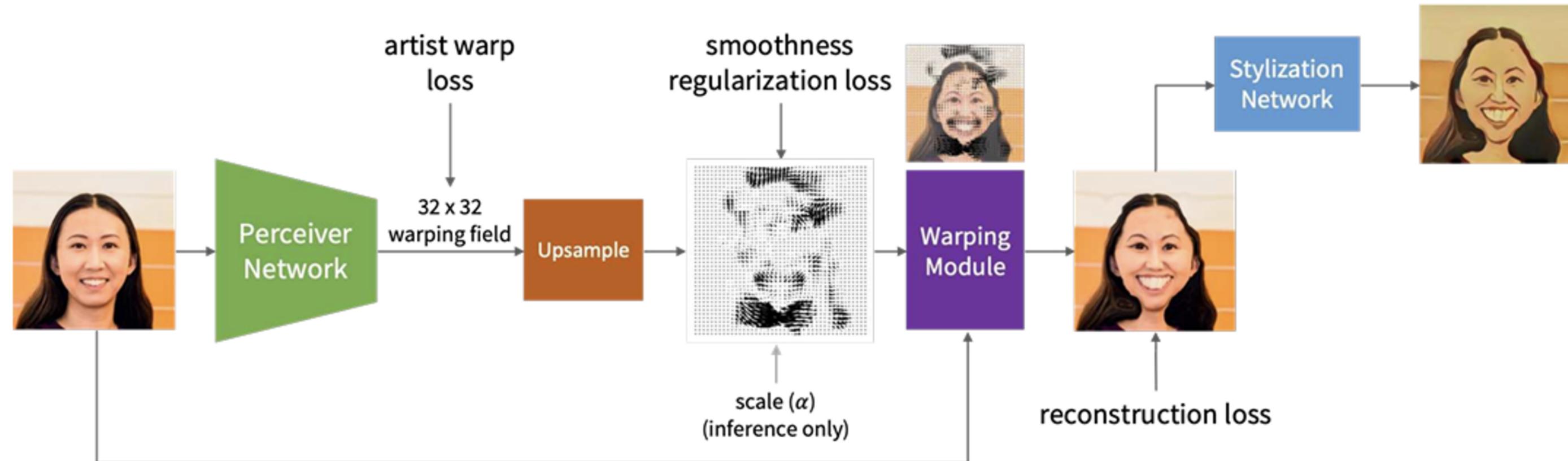
101 frontal-facing portrait images will be collected from Flickr.

Ground truth: were warped via Adobe Photoshop by two caricature artists with similar styles.

The dataset includes 101 image pairs, with 90 for training and 11 for validation.

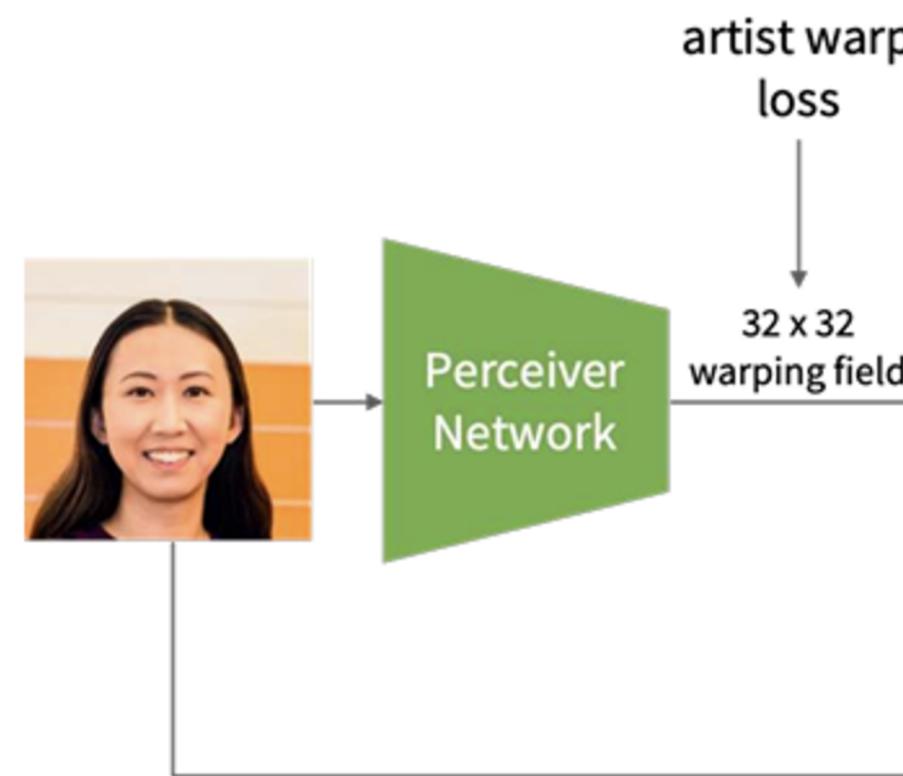
The test set, collected separately from public platforms, contains images without ground-truth labels.

Model Architecture



THE PRECEIVER NETWORK

A shortened version of the Squeeze-and-Excitation Network (SENet50) that has been pretrained using the VGGFace2 Dataset.



Retaining only the initial layers up to and including the second bottleneck block, followed by an adaptive average pooling layer with an output size of $32 \times 32 \times 2$.

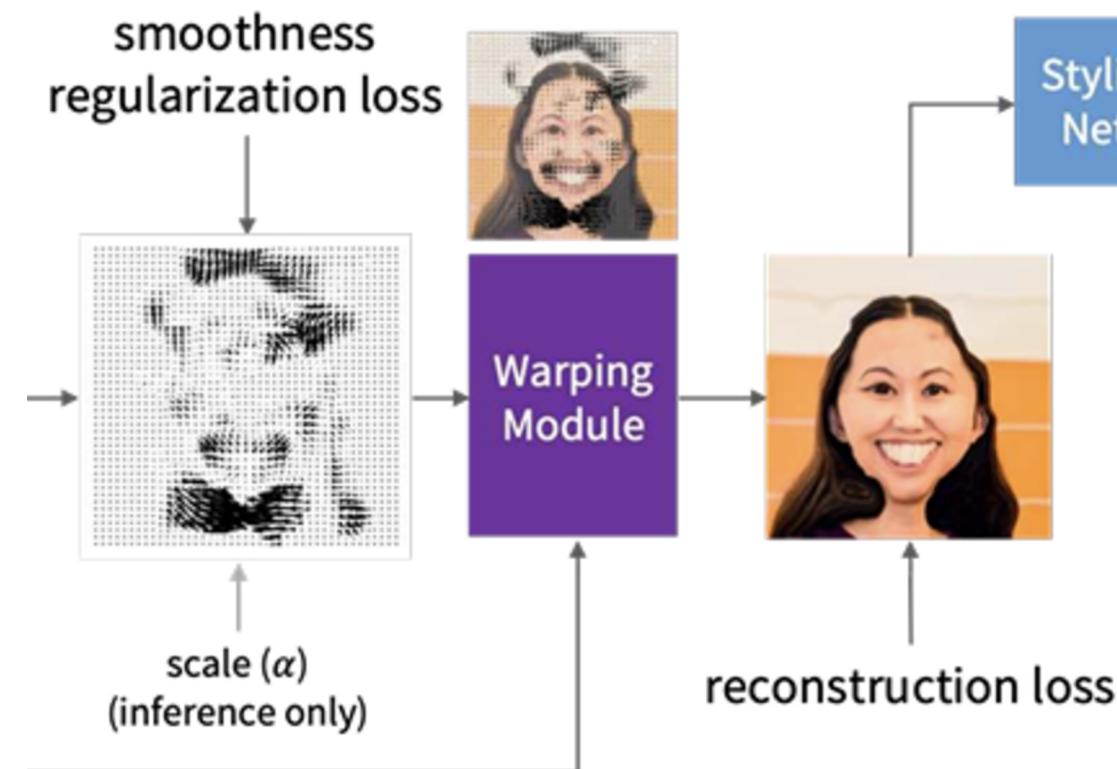
Truncating the network aims to decrease its capacity and prevent overfitting to the limited dataset.

Input: original image.

Output: the warping field, a flow field size $H \times W \times 2$

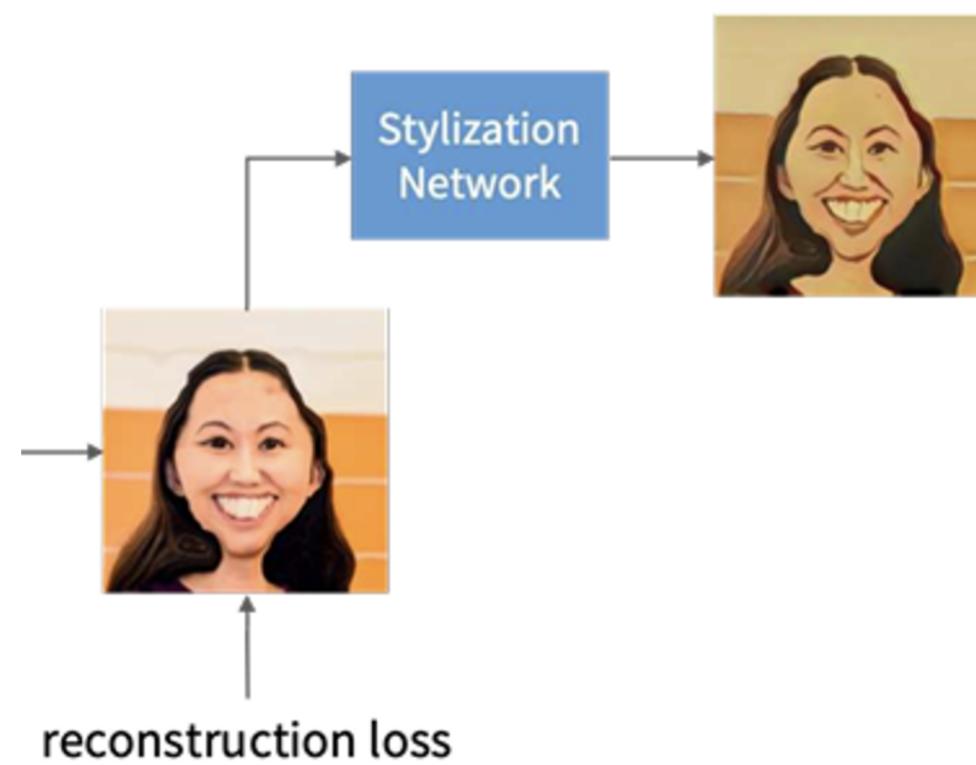
WARPING MODULE

Applies the warping field to the image.



Input: smoothed warping field and original image
Output: warped image

STYLIZATION NETWORK



Applying style transfer to the warped image for an artistic effect.

Using a pretrained model: CartoonGAN

Input: warped image
Output: final image

Progress

PREPROCESS

Handle gray scale image.

Augmented data: random crop, flip, resize.

MODEL DEVELOPMENT

Implement pretrained stylize model CartoonGAN.

Investigating suggested models, which include SENet50

TRAINING AND OPTIMIZATION

Training: Train the model using the preprocessed data with appropriate hyperparameters.

Validation: Evaluate the model's performance on a validation dataset.

Optimization: Fine-tune model architecture or hyperparameters for better results.



Evaluation

HUMAN EVALUATION

Good



Bad



METRICS

MSE (Mean Squared Error): 4.73

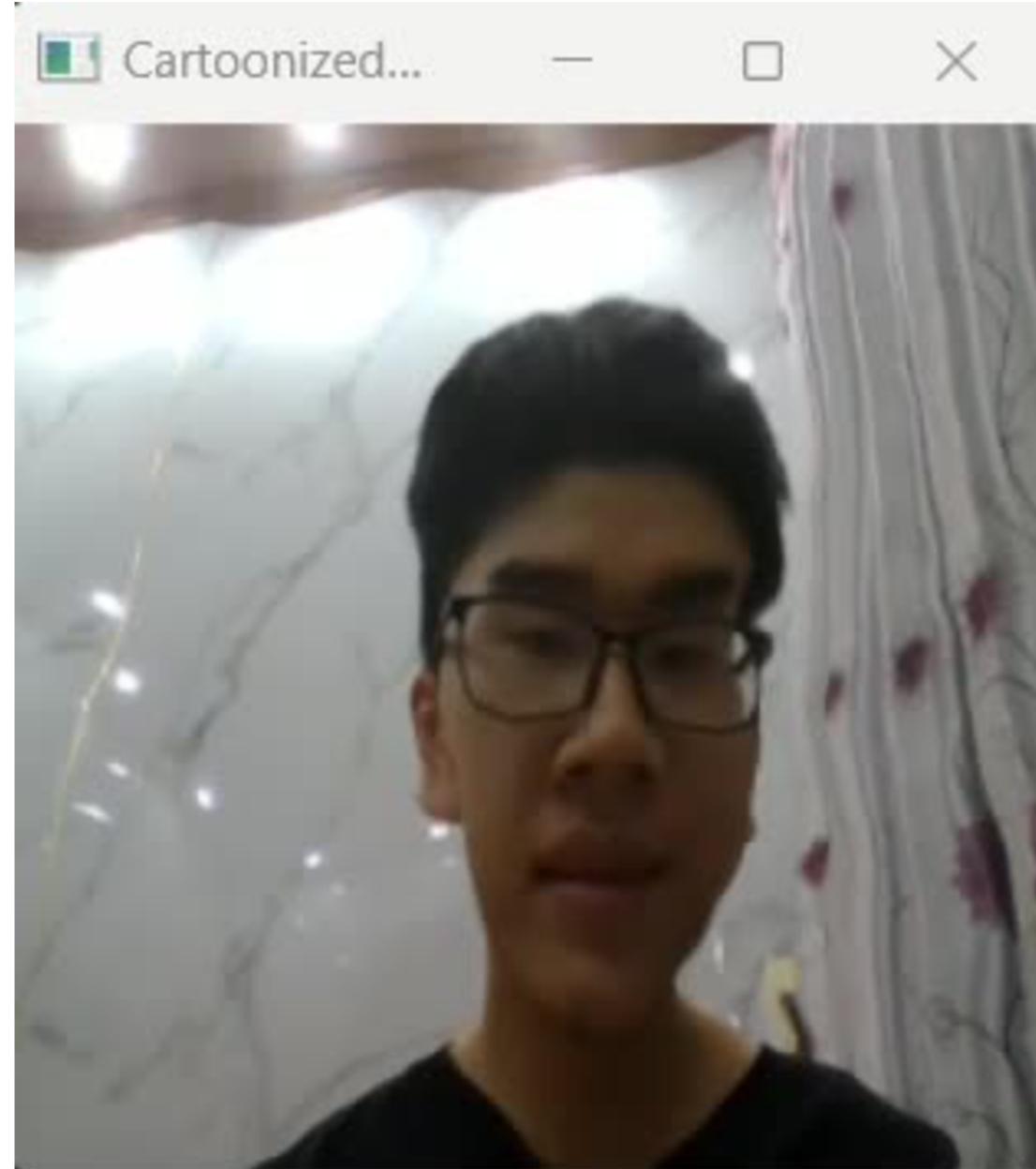
-> Quite small relative to the range of pixel values, meaning there are no extreme discrepancies between the ground truth and the output images.

SSIM (Structural Similarity Index): 0.982

-> Excellent, the model is able to preserve the structural features of the images very well, with minimal perceptual differences.

PSNR (Peak Signal-to-Noise Ratio): 40.58 dB

-> Very good, the output image is of high quality and very similar to the ground truth, with minimal distortion or noise.



CARTOON WEBCAM FEED

The Cartoon Webcam Feed is a real-time application that processes live video feed from a webcam and applies a cartoon-style transformation to each frame.

Use OpenCV for webcam feed handling.

Features: Real-Time Processing; Deep Learning-Based Cartoonization and User Interaction.

Conclusion

AutoToon offers flexibility and detail, outperforming state-of-the-art methods in exaggerating facial features. AutoToon is highlighted to be a potential project in addressing real-world applications.

However, while this system shows good results in caricature exaggeration, it still faces some challenges.

The future work can explore refining warping smoothness, improving identity preservation, and adapting to diverse artist styles through few-shot learning.





ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

A large, semi-transparent watermark of the HUST logo is positioned on the left side of the slide. The logo consists of the letters "HUST" in a bold, white, sans-serif font, with a stylized orange and red dotted pattern forming a background shape.

HUST

THANK YOU !