- 2023 Fall EECS598-007 -

Domain transfer of sketched facial image into realistic facial image

Wonseok Oh

Umich ECE master's student

Contents

- Introduction
- Related works
- Method
- Results
- Future work

Introduction

- Prevent the criminal
    1. CCTV videos for Dataset

Introduction

- Prevent the criminal

  II. Sketch – Montage of a criminal

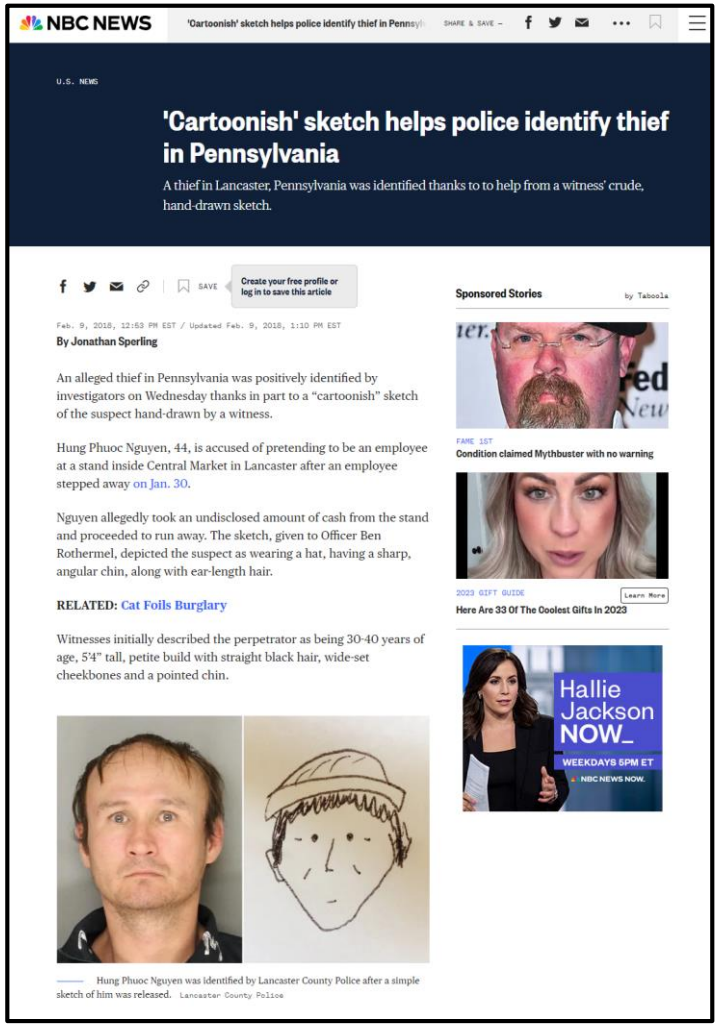
Introduction

- Prevent the criminal

  II. Sketch – Montage of a criminal

Introduction

- Prevent the criminal



NBC NEWS

'Cartoonish' sketch helps police identify thief in Pennsyl

SHARE & SAVE

U.S. NEWS

'Cartoonish' sketch helps police identify thief in Pennsylvania

A thief in Lancaster, Pennsylvania was identified thanks to to help from a witness' crude, hand-drawn sketch.

SAVE

Create your free profile or log in to save this article

Feb. 9, 2018, 12:53 PM EST / Updated Feb. 9, 2018, 1:10 PM EST

By Jonathan Sperling

An alleged thief in Pennsylvania was positively identified by investigators on Wednesday thanks in part to a "cartoonish" sketch of the suspect hand-drawn by a witness.

Hung Phuoc Nguyen, 44, is accused of pretending to be an employee at a stand inside Central Market in Lancaster after an employee stepped away on Jan. 30.

Nguyen allegedly took an undisclosed amount of cash from the stand and proceeded to run away. The sketch, given to Officer Ben Rothermel, depicted the suspect as wearing a hat, having a sharp, angular chin, along with ear-length hair.

RELATED: Cat Foils Burglary

Witnesses initially described the perpetrator as being 30-40 years of age, 5'4" tall, petite build with straight black hair, wide-set cheekbones and a pointed chin.

Hung Phuoc Nguyen was identified by Lancaster County Police after a simple sketch of him was released. Lancaster County Police

Sponsored Stories by Taboola

FAME 1ST

Condition claimed Mythbuster with no warning

2023 GIFT GUIDE

Learn More

Here Are 33 Of The Coolest Gifts In 2023

Hallie Jackson NOW_

WEEKDAYS 5PM ET

NBC NEWS NOW.

Related Works

Pix2Pix : Requires paired data

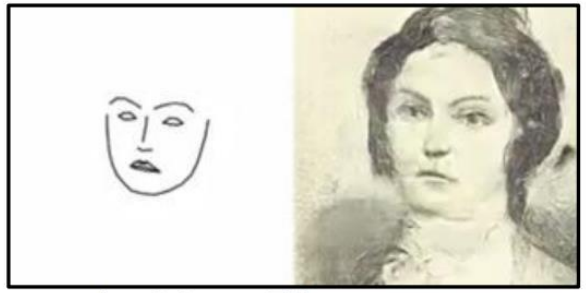
Background removal

by Kaihu Chen

Palette generation

by Jack Qiao

Sketch→ Portrait



by Mario Klingemann

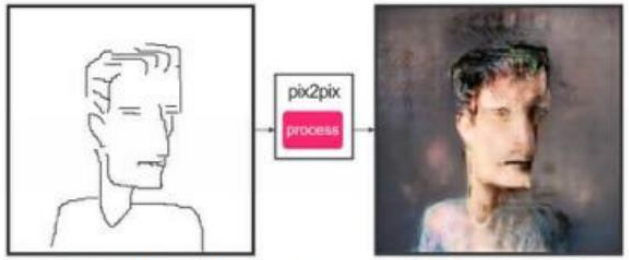
Sketch → Pokemon

by Bertrand Gondouin

"Do as I do"

by Brannon Dorsey

#fotogenerator

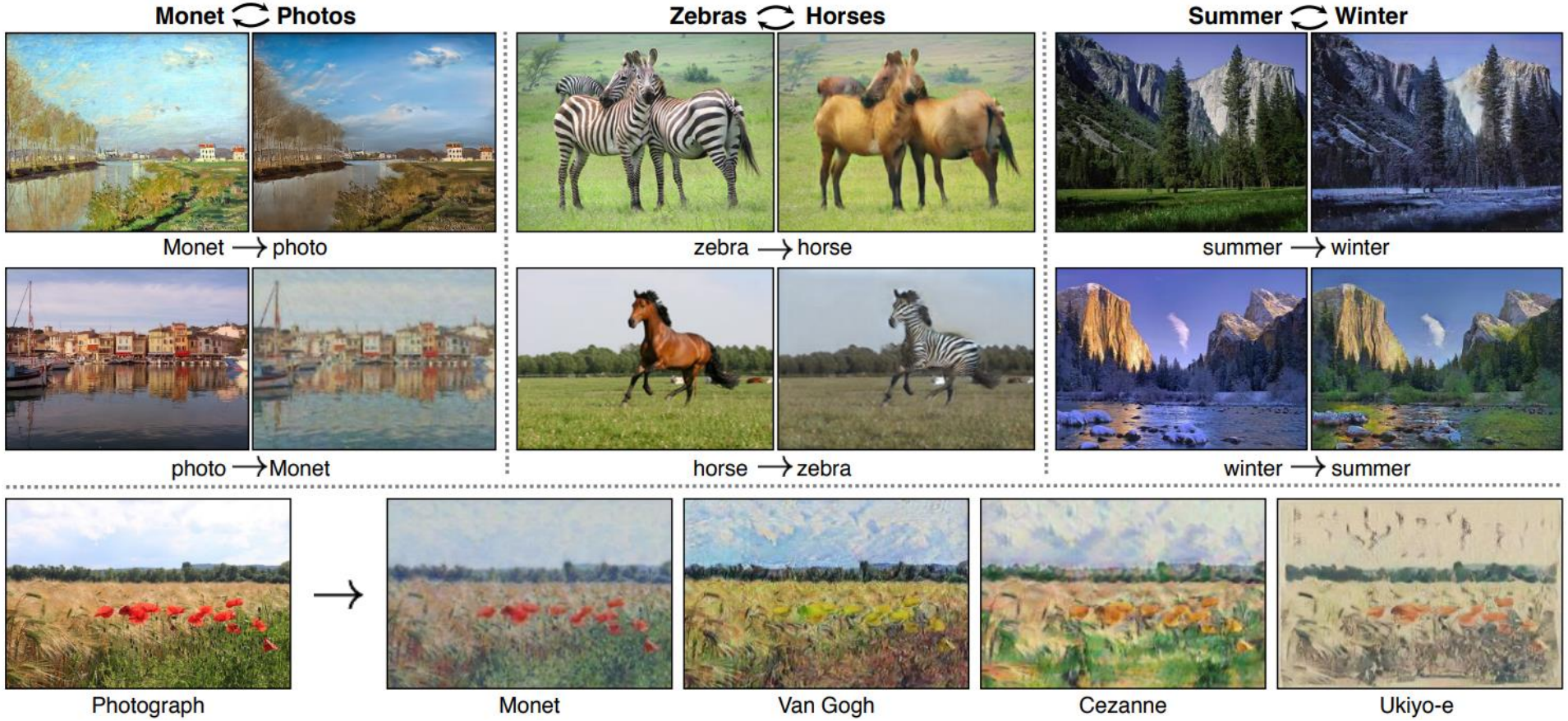


sketch by Yann LeCun

Image-to-image translation with conditional adversarial networks, Isola, Phillip, et al. CVPR 2017

Related Works

CycleGAN : unpaired data



Unpaired image-to-image translation using cycle-consistent adversarial networks, Zhu, Jun-Yan, et al. ICCV 2017

Method

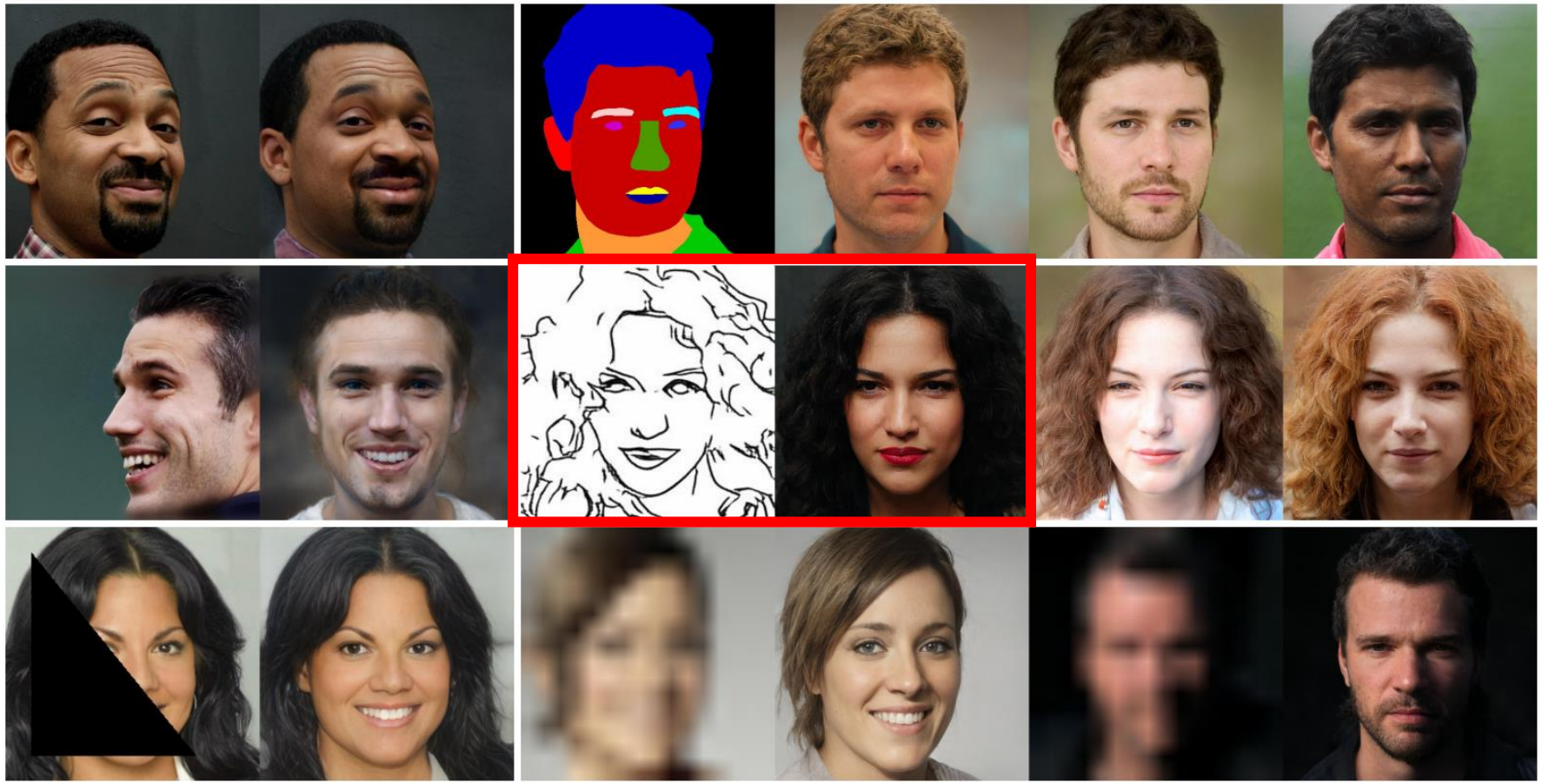
1. Image2Image Translation

2. Image editing

Encoding in style: a stylegan encoder for image-to-image translation., Richardson, Elad, et al, CVPR 2021

Method

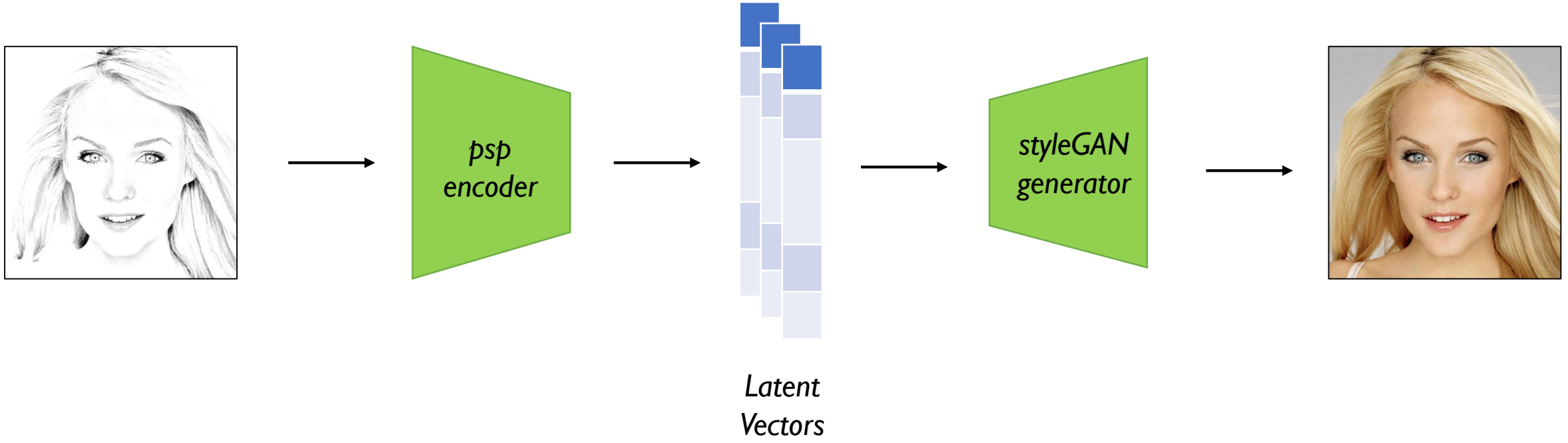
Image2Image Translation



Encoding in style: a stylegan encoder for image-to-image translation., Richardson, Elad, et al, CVPR 2021

Method

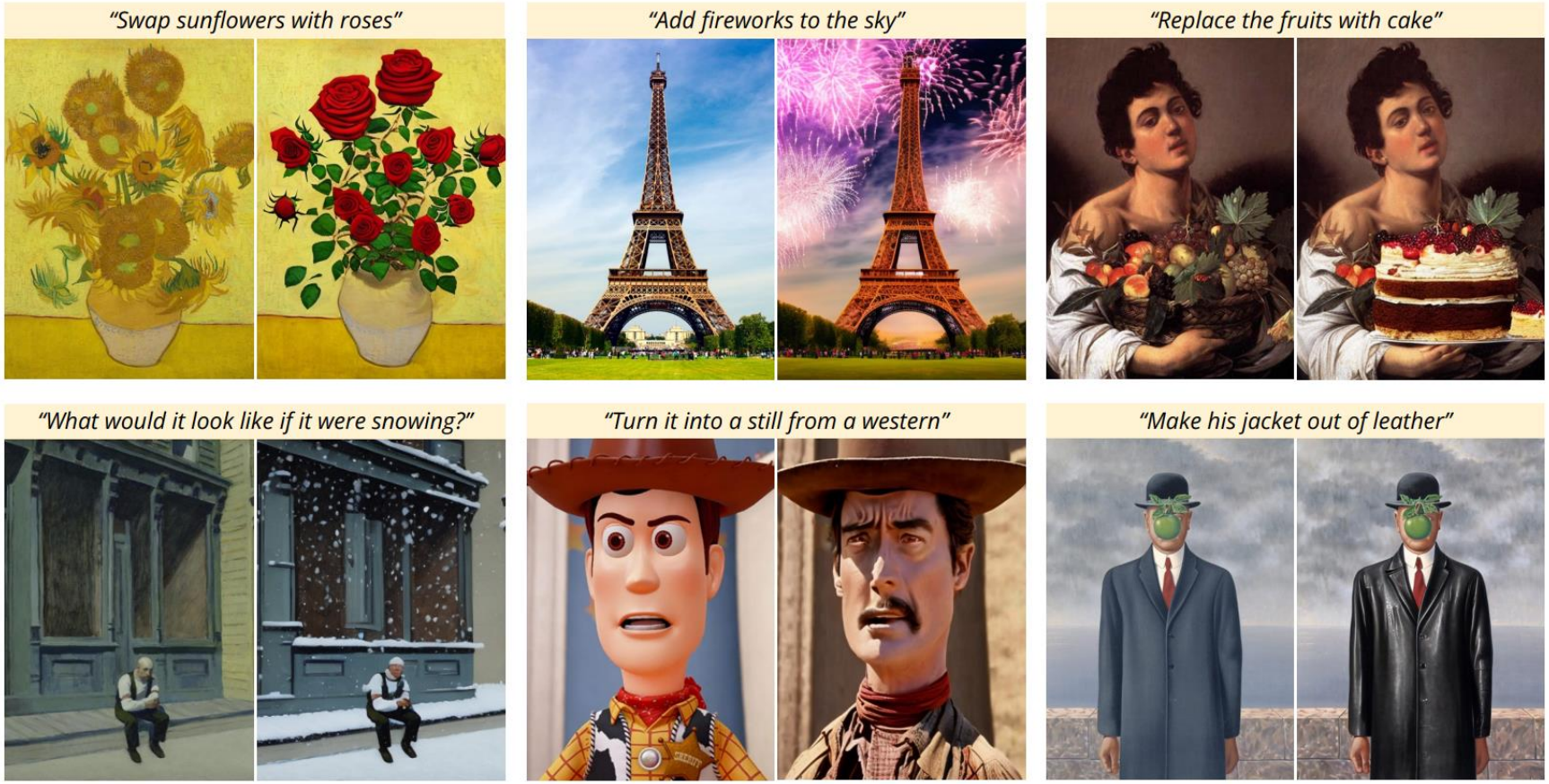
Image2Image Translation



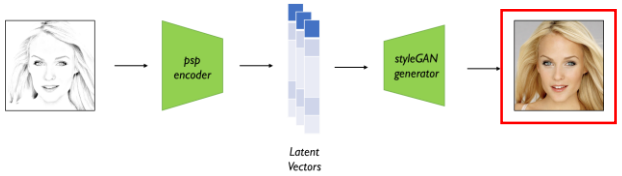
Encoding in style: a stylegan encoder for image-to-image translation., Richardson, Elad, et al, CVPR 2021

Method

Image editing

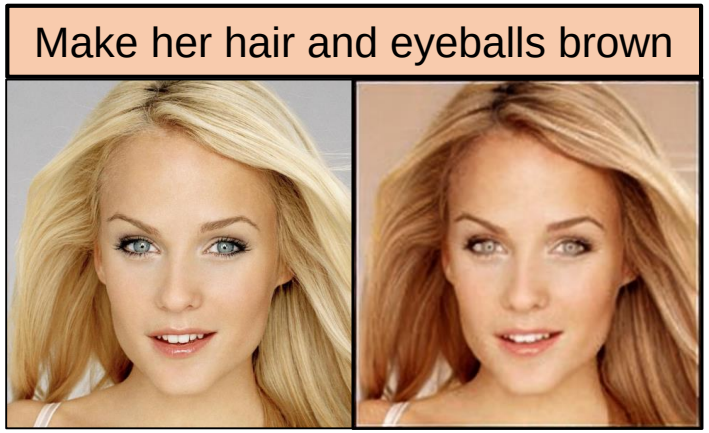


Instructpix2pix: Learning to follow image editing instructions., Brooks, Tim, et al, CVPR 2023

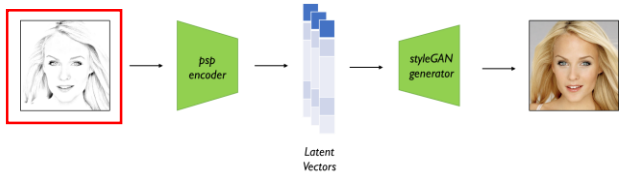


Method

Image editing



Instructpix2pix: Learning to follow image editing instructions., Brooks, Tim, et al, CVPR 2023

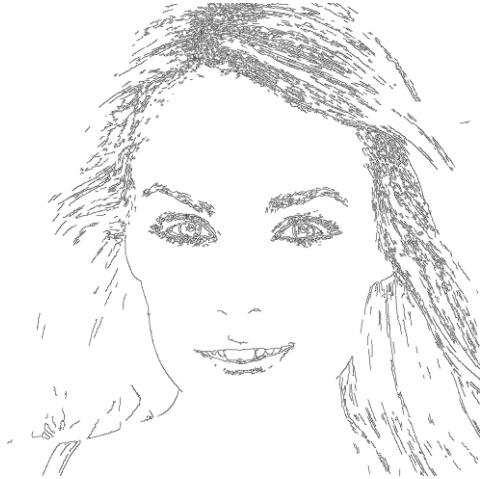


Method

Dataset – Preprocessing Sketches

Original Image

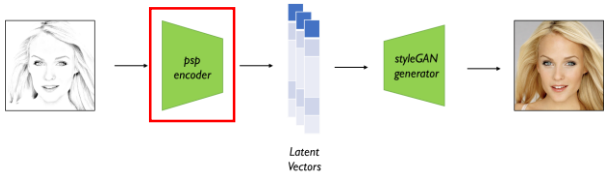
Black and White Image



Canny edge Image

Sketched Image

Instructpix2pix: Learning to follow image editing instructions., Brooks, Tim, et al, CVPR 2023



Method

Encoder – loss function

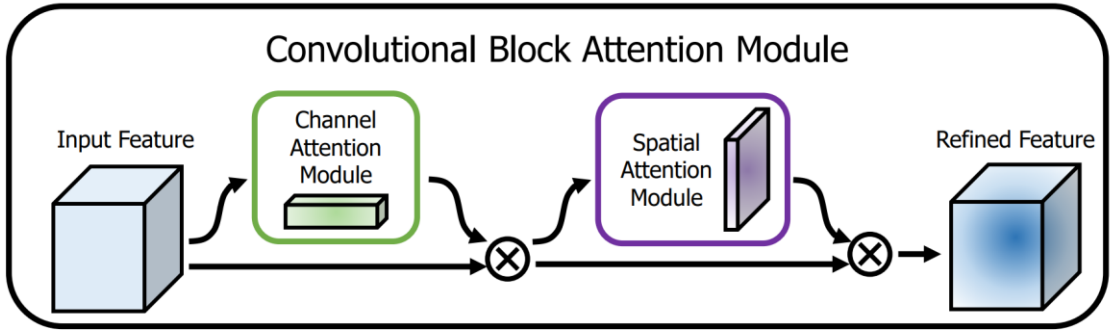


Fig. 1: **The overview of CBAM**. The module has two sequential sub-modules: *channel* and *spatial*. The intermediate feature map is adaptively refined through our module (CBAM) at every convolutional block of deep networks.

$$\mathbf{F}''(\mathbf{x}) = \mathbf{M_s}(\mathbf{F}') \otimes \mathbf{F}$$

$$\mathcal{L}_{attention}(\mathbf{x}) = \lambda||\mathbf{F}''(\mathbf{x}) - \mathbf{F}''(G(E(\mathbf{x})))||_2$$

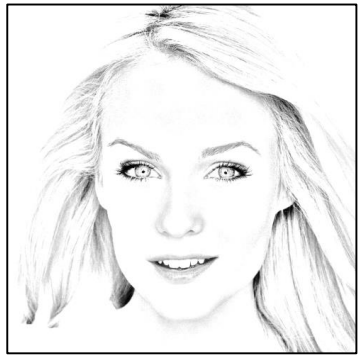
$$\mathcal{L}_{total}(\mathbf{x}) = \lambda\mathcal{L}_E + \lambda_4\,\mathcal{L}_{attention}(\mathbf{x})$$

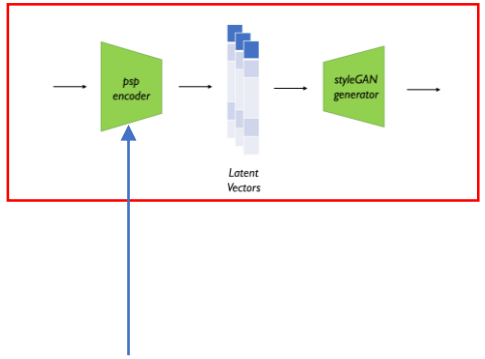
CBAM: Convolutional Block Attention Module, Woo et al, ECCV 2018

Method

Overview

psp with attention loss





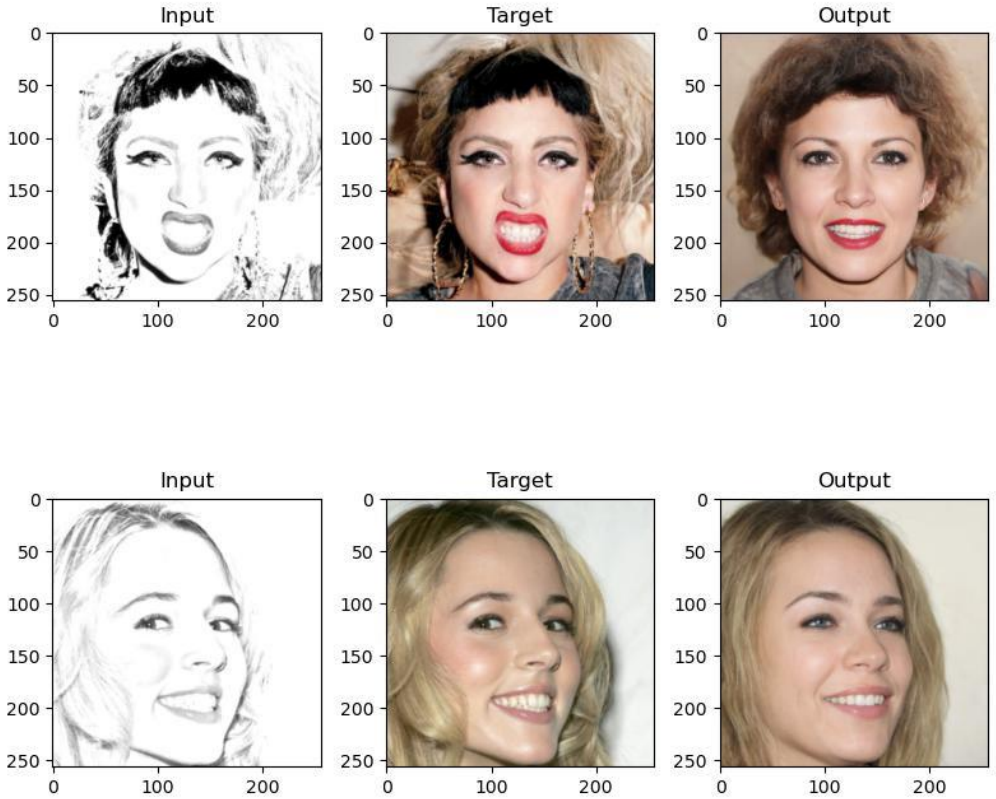
$$\mathcal{L}_{total}(\mathbf{x}) = \lambda \mathcal{L}_E + \lambda_4 \mathcal{L}_{attention}(\mathbf{x})$$

Instruct pix2pix with editing

Make her hair and eye brown

Results

Qualitative comparisons



Results

Image editing

“ Make her eyeballs brown ”

Input sketch

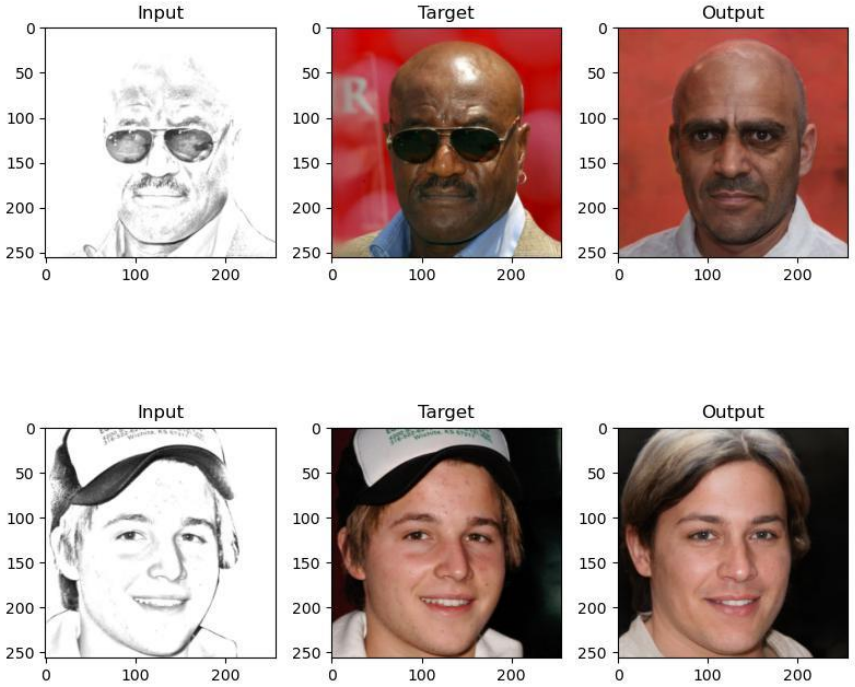
Output 1 | Final Output | Ground Truth

Results

Multimodal + Real Case

Future work

Limitations



Future work

Hyperparameter tuning

Model	Runtime	MSE ↓	LPIPS ↓	Similarity ↑
pSp	0.0398 ± 0.1031	0.0780	0.291	**0.340**
Ours ($\lambda = 1000$)	0.0395 ± 0.1110	**0.0773**	**0.288**	0.341
Ours ($\lambda =?$)				

Thank you

And thank you for your effort!