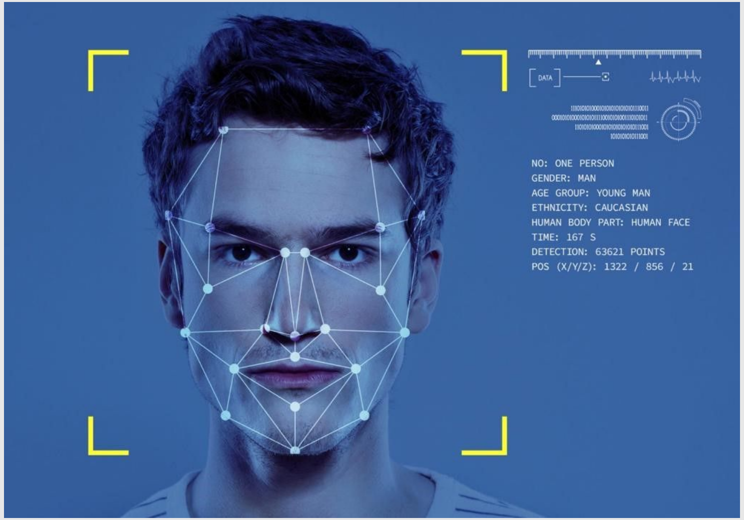
CELEBVISION: COMPREHENSIVE CELEBRITY IMAGE PROCESSING

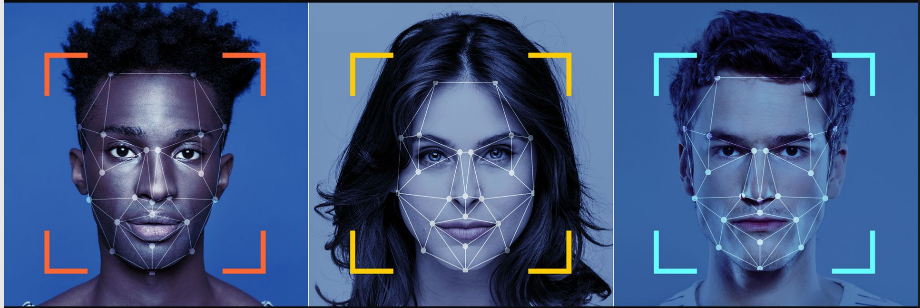


2024SP_MLDS_432-1_SEC20 **Deep Learning Final Project**

Siyan Li, Stella Wang, Xinran Wang, Yumin Zhang

ARTIFICIAL INTELLIGENCE (AI)

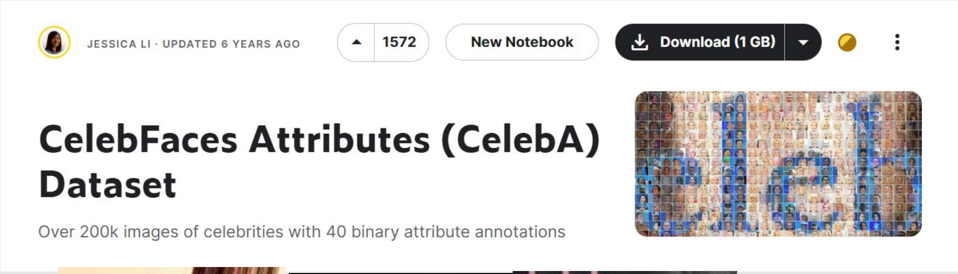
TABLE OF CONTENTS



DATA OVERVIEW

Tabular Data:

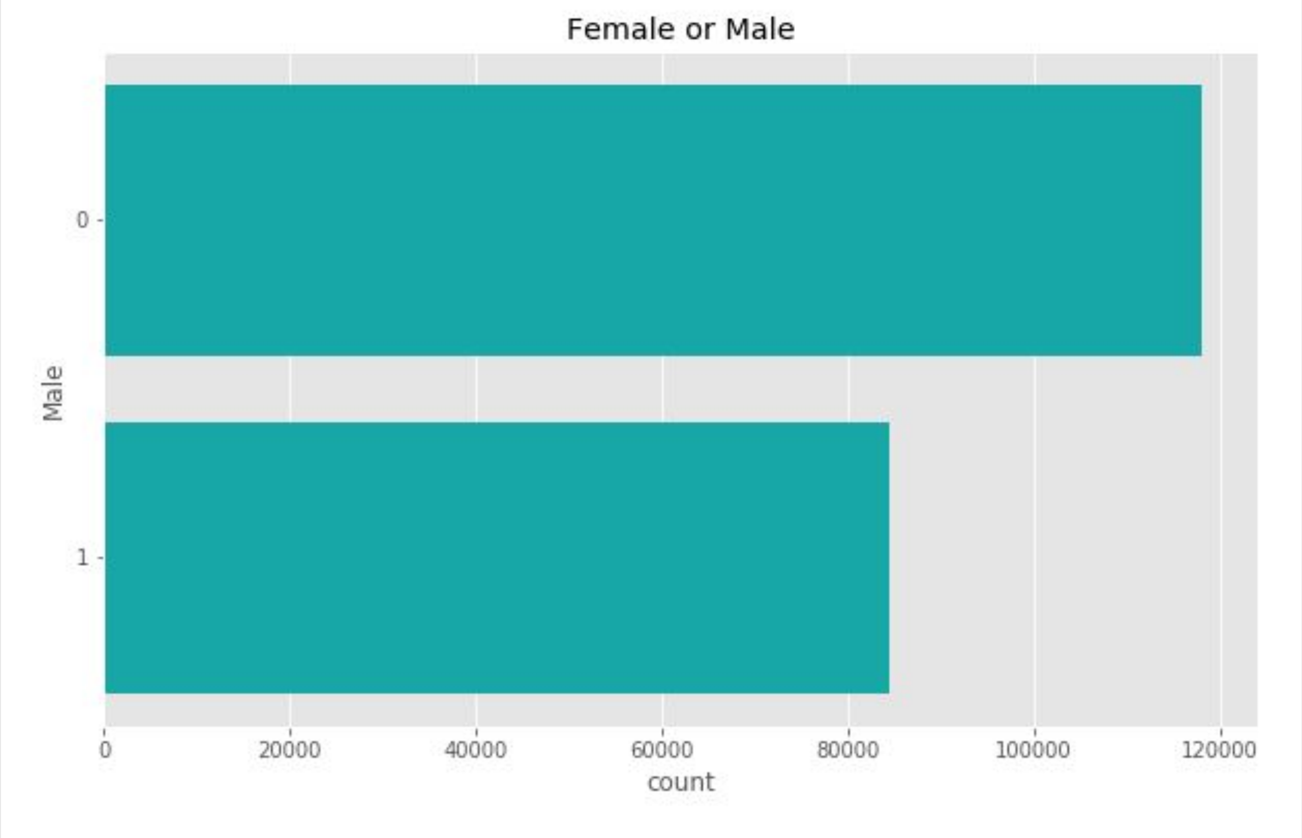
- Total Rows: 202,599
- Total Columns: 40 (Attributes)
- No Missing Values



Data Characteristics:

Type: Dataset Type: Image dataset with corresponding attributes.
Images: 202,599 images of celebrities, each annotated with various facial attributes.

VISUALIZATION



TASK 01.

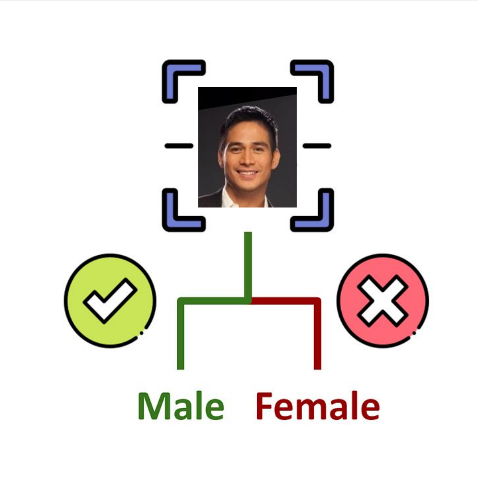
IMAGE CLASSIFICATION

Classify the gender of the faces

OBJECTIVE

GOAL

To develop and evaluate deep learning models capable of accurately **classifying the gender** of individuals in the CelebA dataset based on facial attributes.



MODELS

Customized CNN & InceptionV3

> > > >

CUSTOMIZED MODEL: CNN

ARCHITECTURE

Customized CNN is a deep convolutional neural network (CNN) designed specifically for the CelebFaces data to optimize performance.

PROS

Tailored Architecture:
Custom design for better performance.

Control Over the Model:
Optimize the architecture and adapt to specific data characteristics

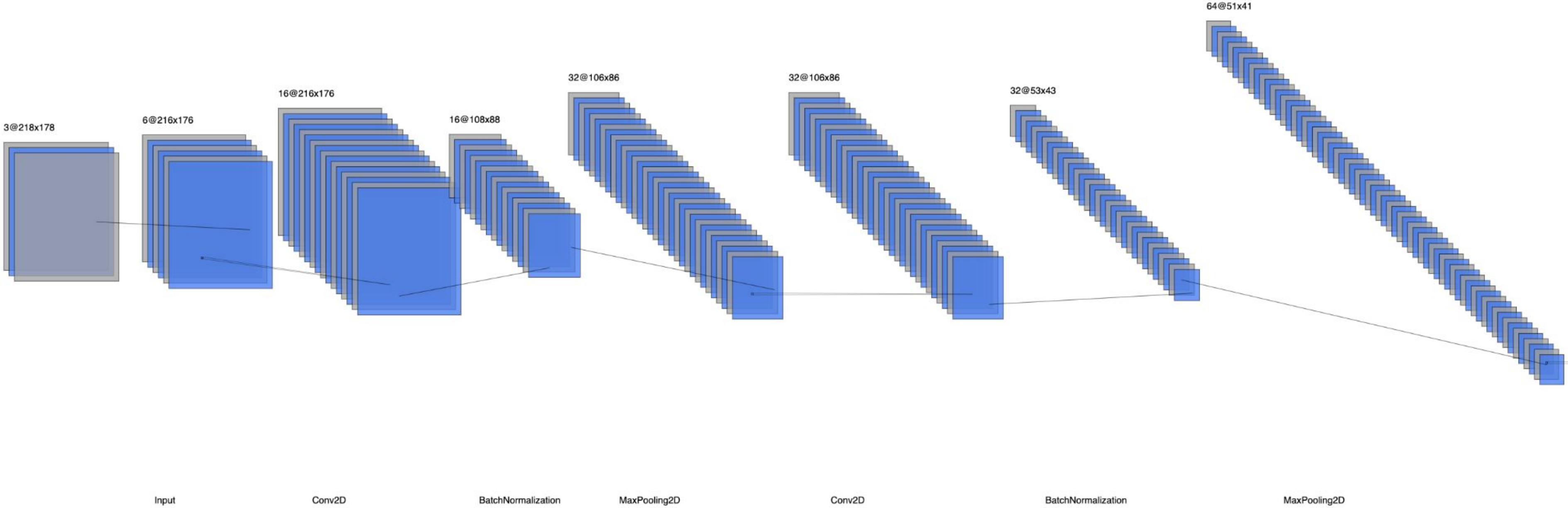
CONS

Resource and Expertise Demands:
Time-consuming and computationally intensive

Risk of Overfitting:
Higher risk of overfitting for small or unbalanced datasets

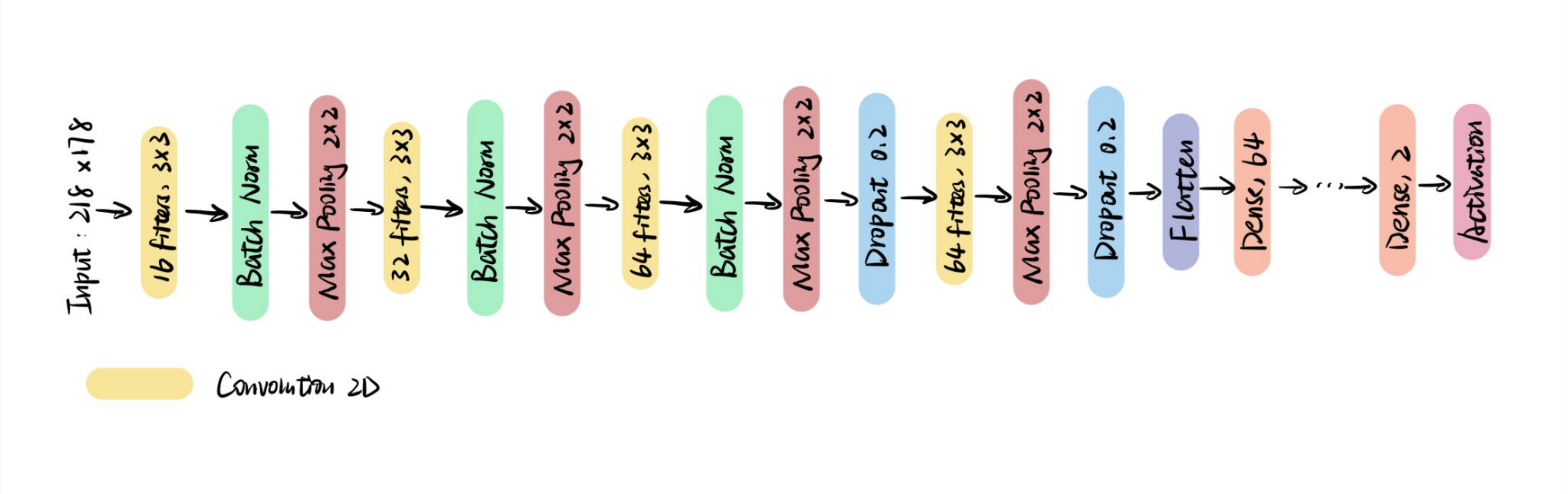
CUSTOMIZED MODEL: CNN

Model Architecture (first few layers)



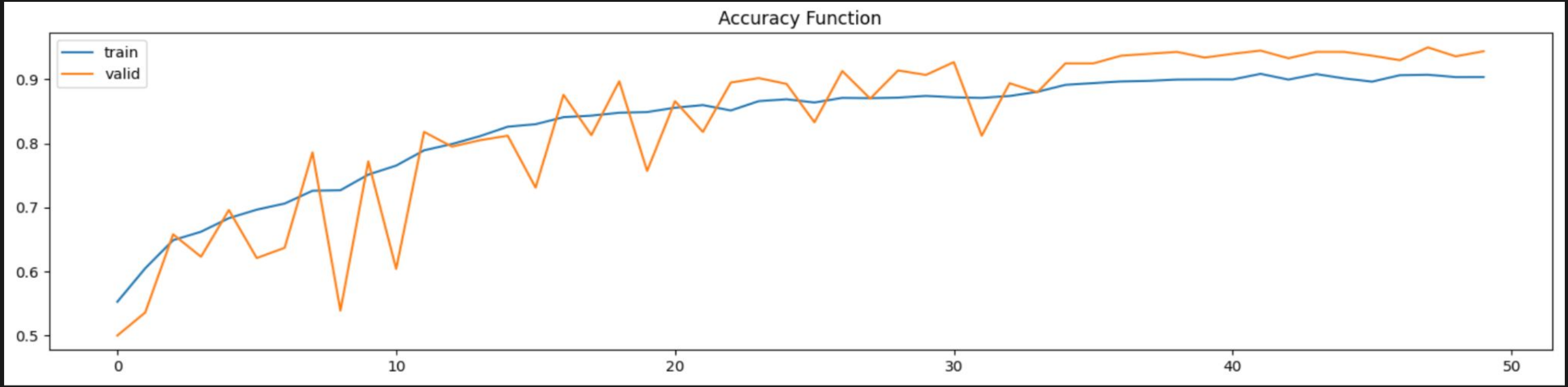
CUSTOMIZED MODEL: CNN

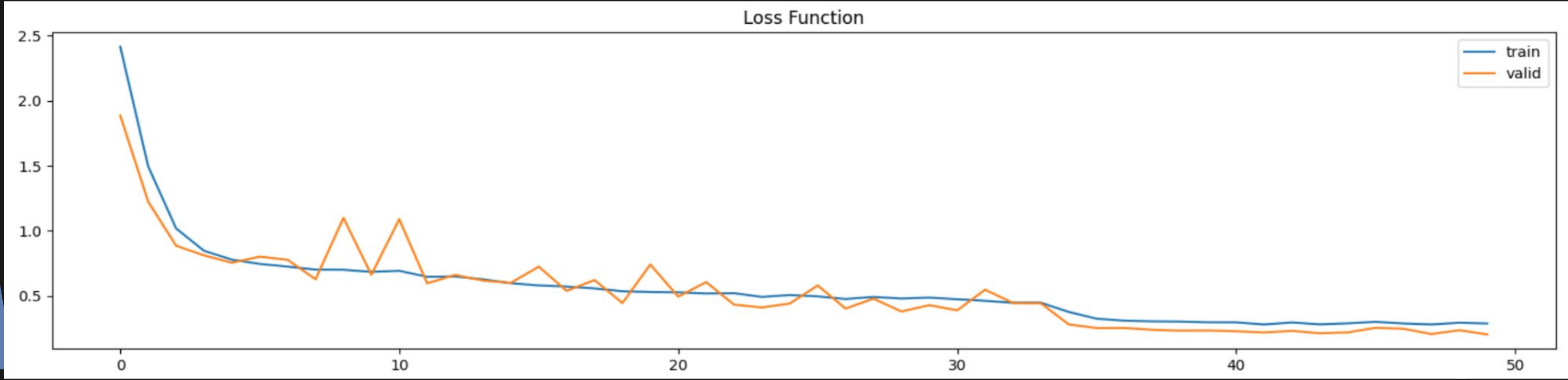
Model Architecture (Continued)



full model architecture.

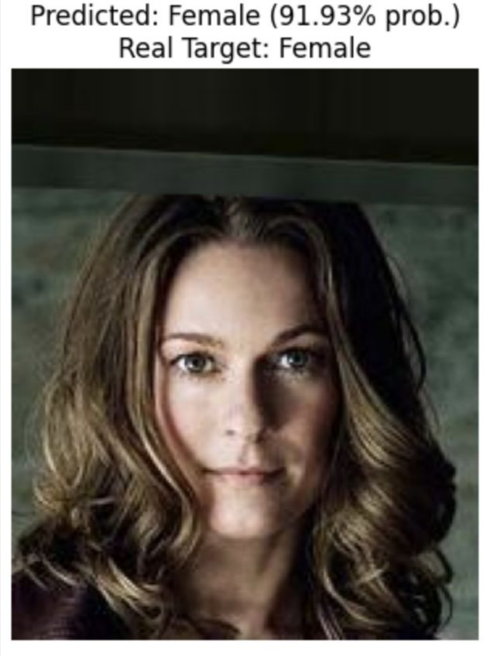
CUSTOMIZED MODEL: CNN





CUSTOMIZED MODEL: CNN

Demo:



PRETRAINED MODEL: INCEPTIONV3

ARCHITECTURE

InceptionV3 is a deep convolutional neural network (CNN) designed by Google.

PROS

<u>High Accuracy:</u> InceptionV3 has demonstrated excellent performance on a variety of image classification tasks, making it well-suited for gender classification on the CelebA dataset.
<u>Scalability:</u> InceptionV3 can handle large-scale datasets like CelebA with many images and attributes, making it versatile for extensive facial attribute analysis.

CONS

<u>Complexity:</u> The architecture is complex, with multiple types of convolutions and pooling operations within each module, making it harder to implement and tune compared to simpler models.

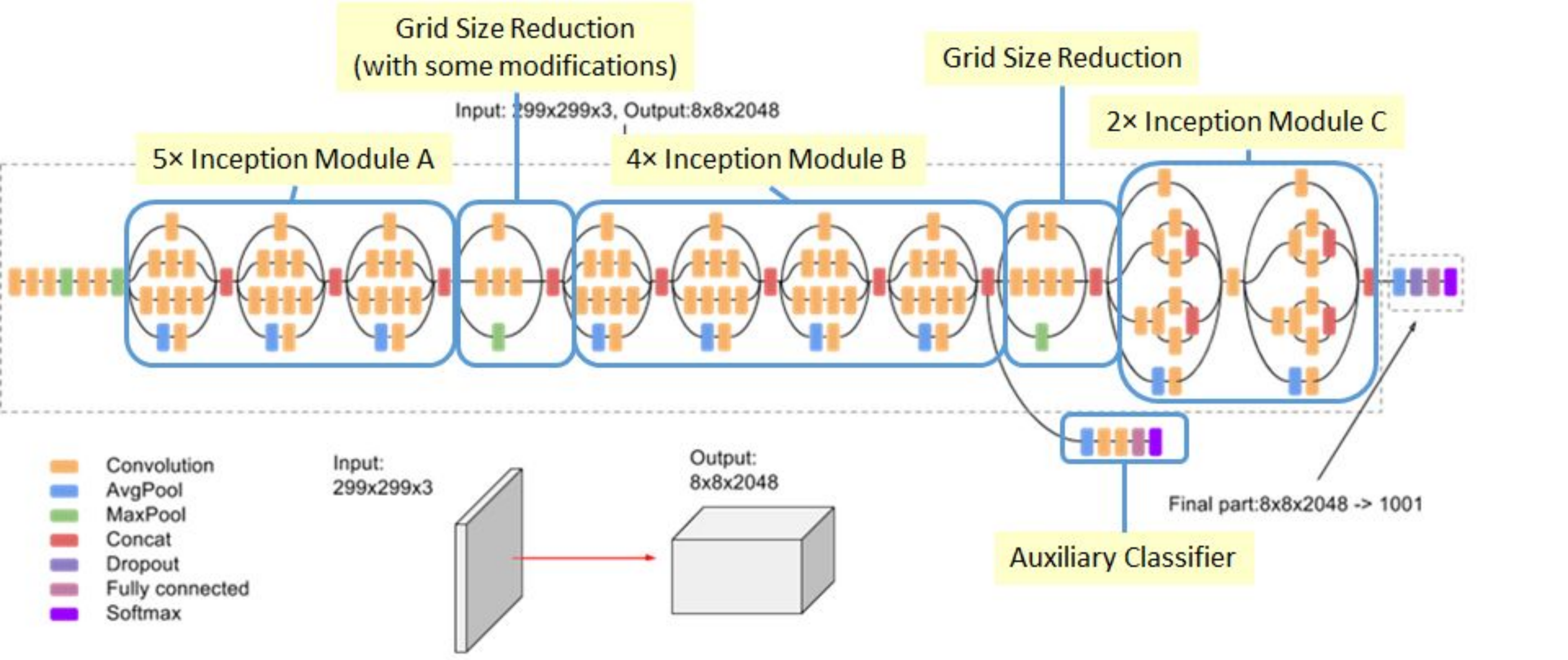
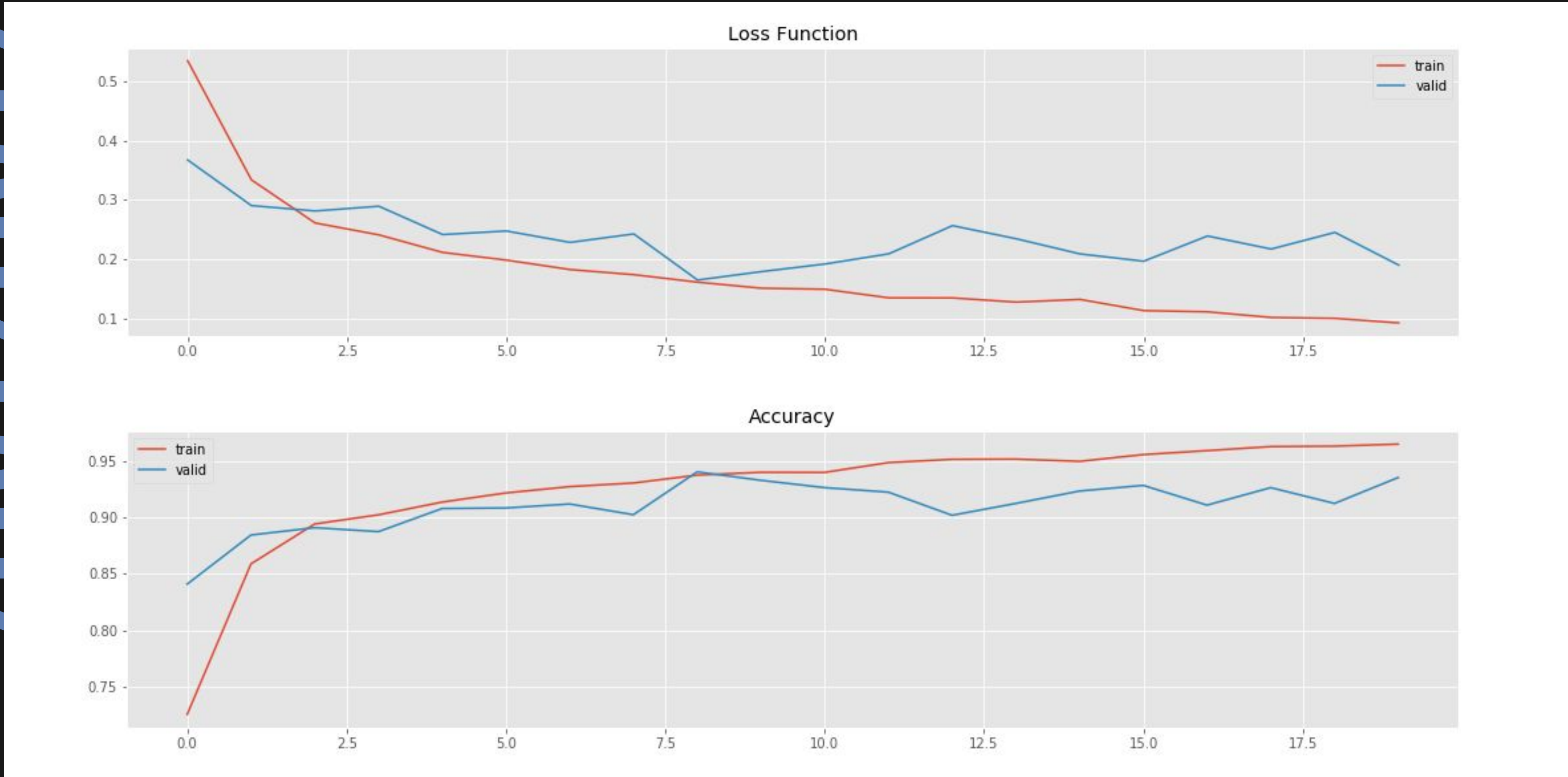


Figure from: Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, 2016

PRETRAINED MODEL: INCEPTIONV3



PRETRAINED MODEL: INCEPTIONV3

Demo:

Male

98.07% prob.
Real Target: Male
Filename: 186397.jpg

Female

99.96% prob.
Real Target: Female
Filename: 192481.jpg

MODEL EVALUATION:

CUSTOMIZED CNN VS INCEPTIONV3

	ACCURACY	F1 SCORE	PRECISION	RECALL	ROC-AUC
CUSTOMIZED CNN	94.4%	0.944	0.945	0,944	0,980
INCEPTION V3	93.2%	0.931	0.934	0.932	0.987

TASK 02.

STYLE TRANSFER

Transfer the style of the given face image

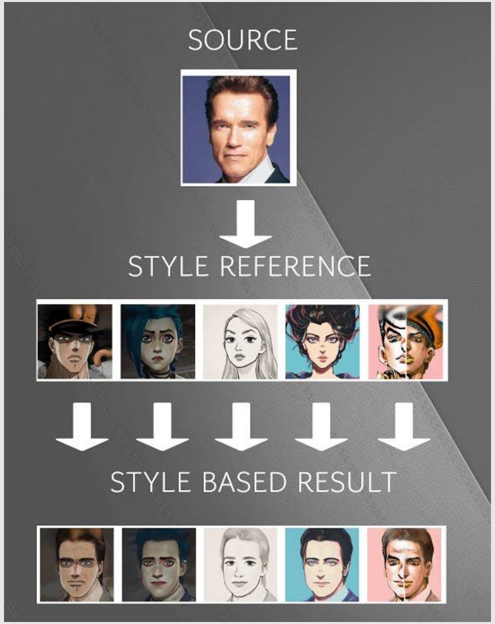
OBJECTIVE

GOAL

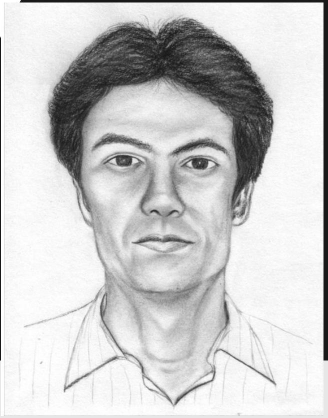
Train and evaluate a model to **apply a specific artistic style** to the **content** of a image.

MODELS

CycleGAN & VGG19



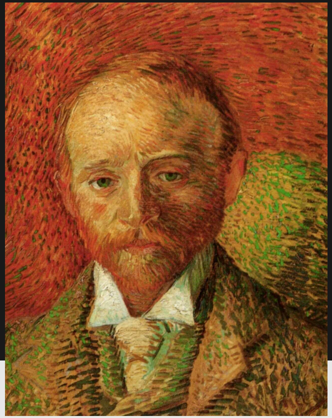
ADDITIONAL kaggle DATASETS



SKETCH

CUHK Face Sketch Database
Sketch Drawn Based on photos

<<<<



VANGOGH

Van Gogh Paintings
Art-Work Sample of Vincent

<<<<

CARTOON

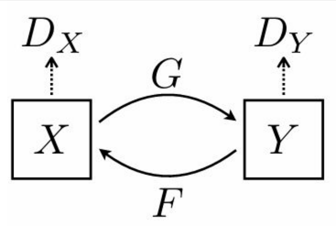
cartoonset10k
Cartoon Images by Google

<<<<

CUSTOMIZED MODEL: CYCLEGAN

ARCHITECTURE

GAN vs CycleGAN



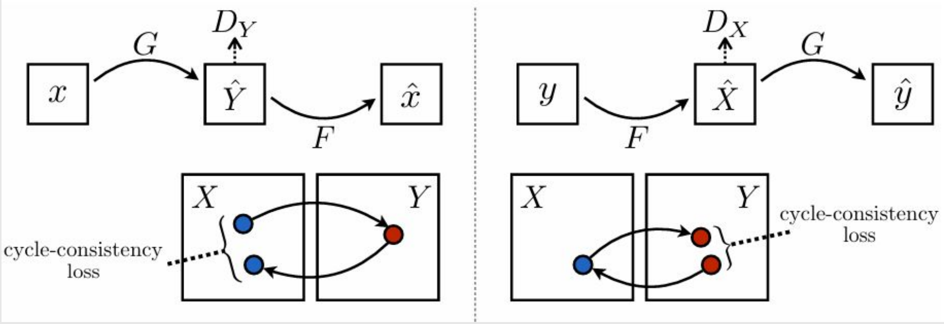


Figure from: Zhu, Jun-Yan, 2017.

GENERATOR LOSS

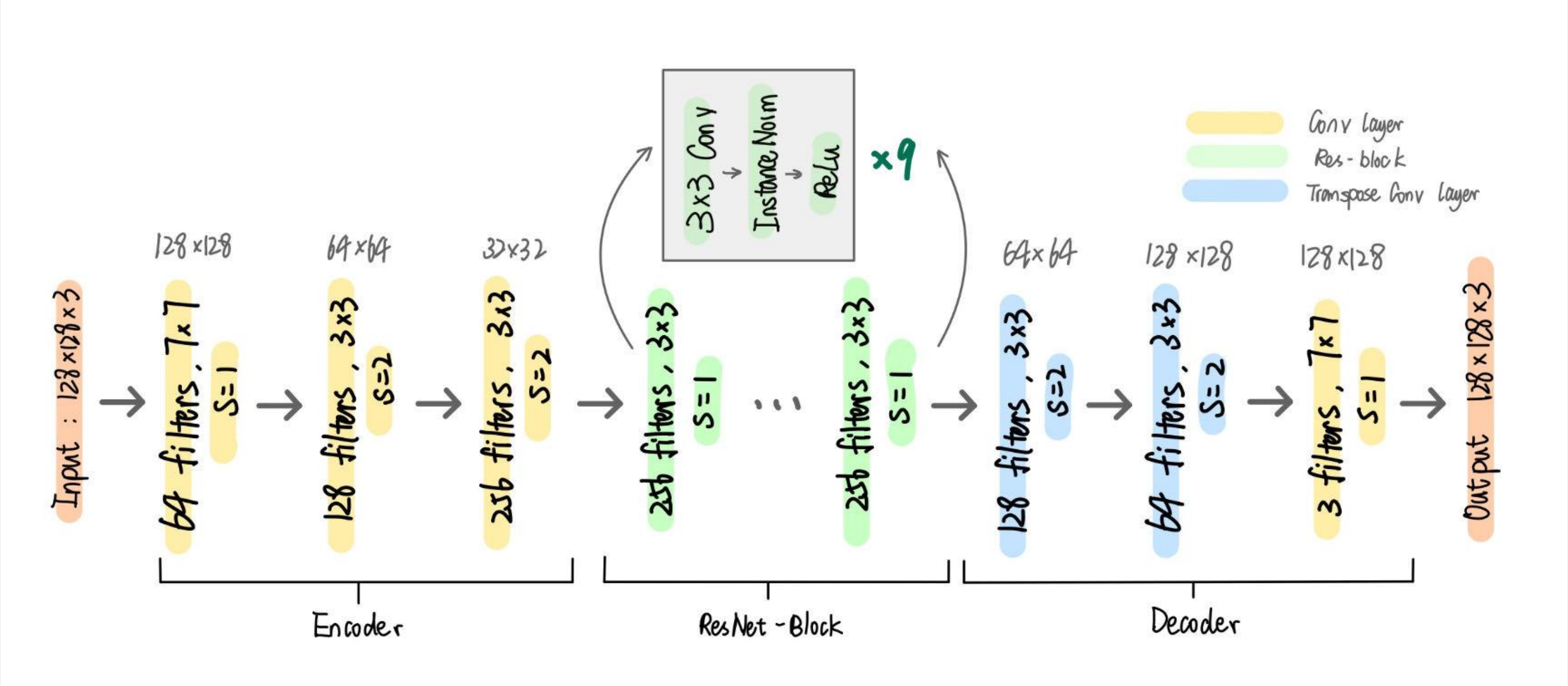
- Adversarial loss (L2/MSE).
- Identity loss (L1/MAE).
- Forward cycle loss (L1/MAE).
- Backward cycle loss (L1/MAE).

L2/MSE: for **label**

L1/MAE: for **image**

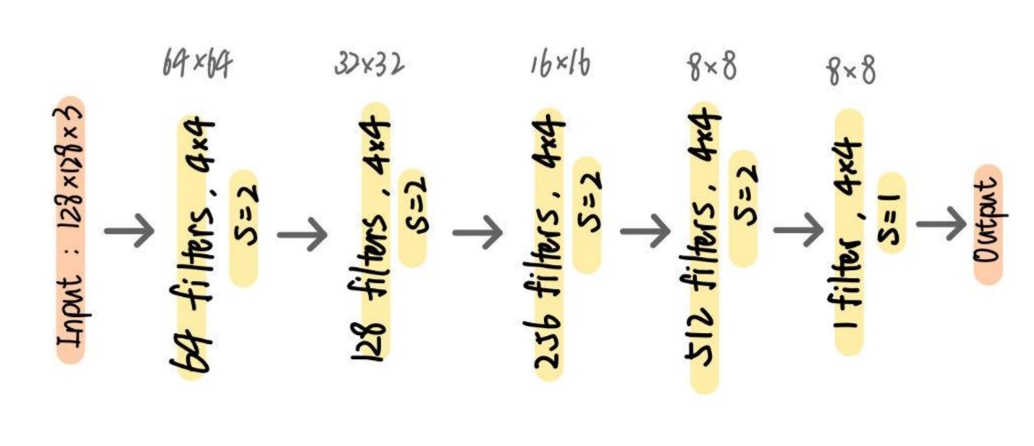
CUSTOMIZED MODEL: CYCLEGAN

GENERATOR



CUSTOMIZED MODEL: CYCLEGAN

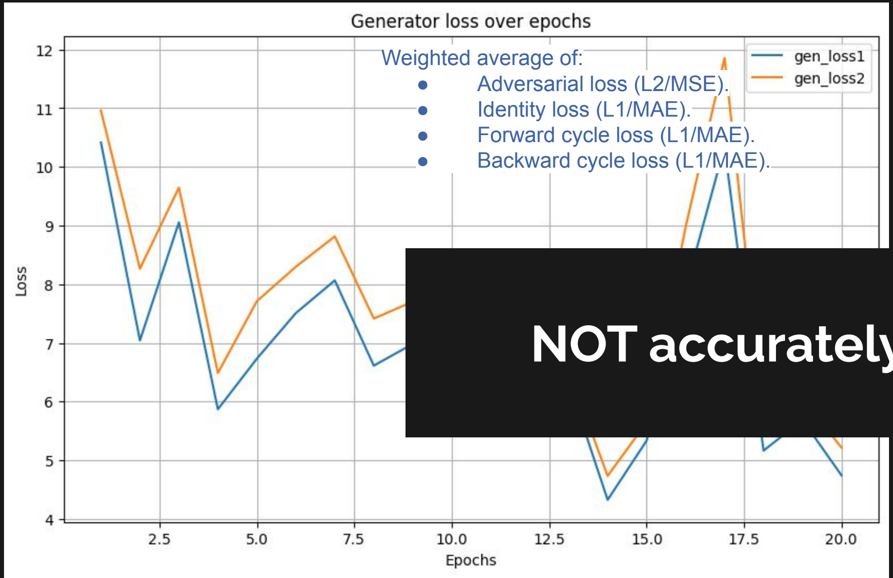
DISCRIMINATOR

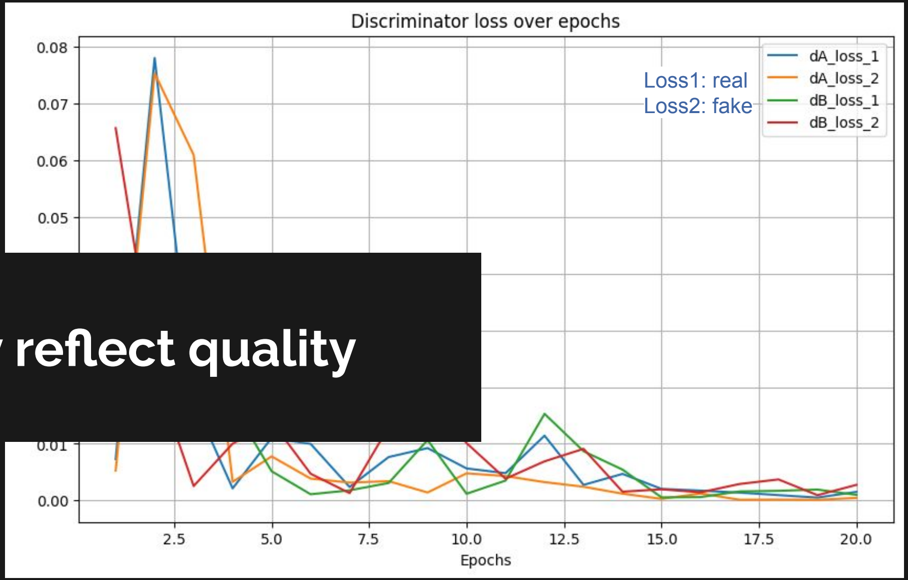


NOTES

1. batchsize = 1
2. Instance normalization instead of batch normalization

CUSTOMIZED MODEL: CYCLEGAN



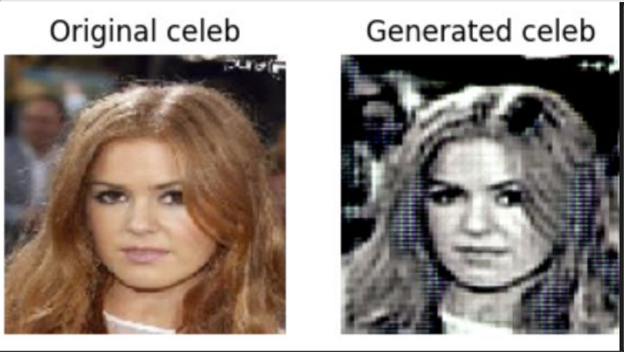


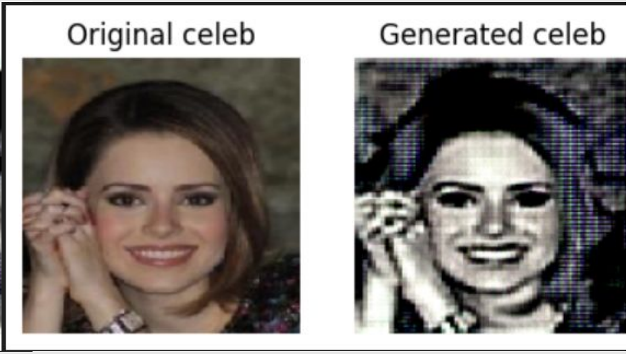
NOT accurately reflect quality

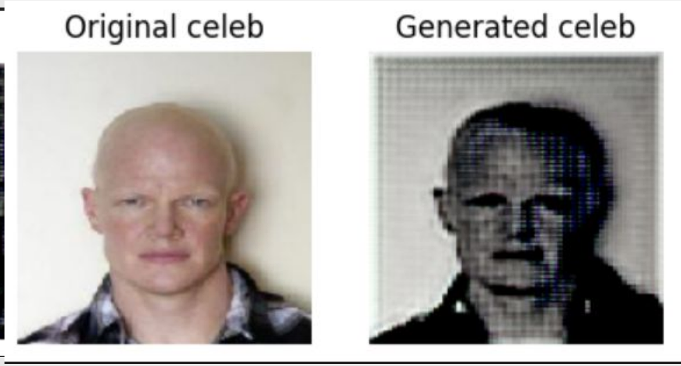
ARTIFICIAL INTELLIGENCE (AI)

CUSTOMIZED MODEL: CYCLEGAN

SKETCH STYLE







ARTIFICIAL INTELLIGENCE (AI)

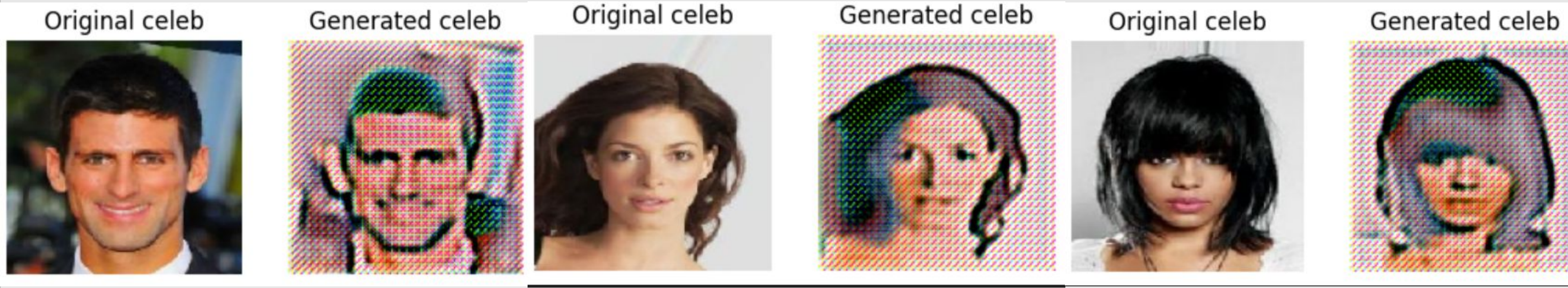
CUSTOMIZED MODEL: CYCLEGAN

VAN GOGH STYLE

ARTIFICIAL INTELLIGENCE (AI)

CUSTOMIZED MODEL: CYCLEGAN

CARTOON STYLE



ARTIFICIAL INTELLIGENCE (AI)

ARTIFICIAL
INTE
(AI)

APPLICATION EXAMPLE:
BY STREAMLIT

(AI)

CycleGAN Style Transfer

Choose a celebrity image...

Drag and drop file here

Limit 200MB per file • JPG, JPEG

Browse files

test_siyan.jpg 10.2KB

Choose a style transfer model

Cartoon Style

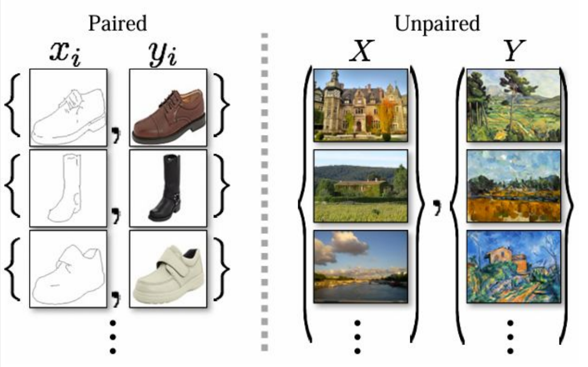
Uploaded Image

Generate Style Transferred Image

CUSTOMIZED MODEL: CYCLEGAN

PROS

1. <u>Unpaired data</u>, unlike *Pix2Pix*



CONS

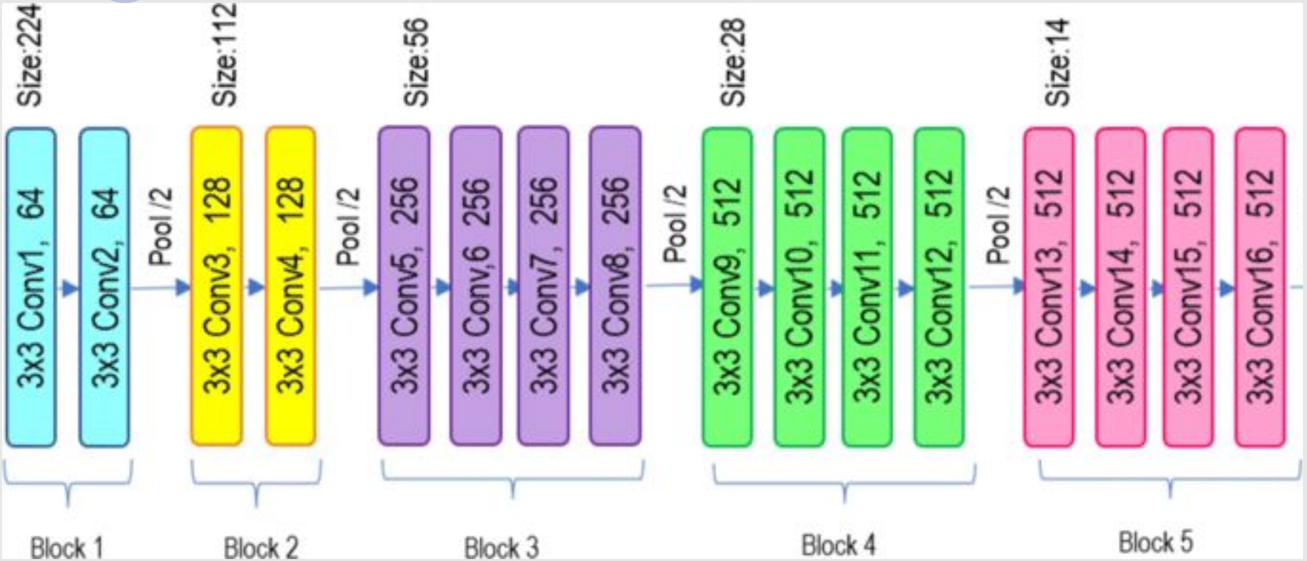
1. <u>Tuning Difficulty</u>
   The performance can be highly sensitive to hyperparameters such as weights of loss components.

2. <u>Blurriness</u>
   The generated images may sometimes contain artifacts or appear blurry.

Figure from: Zhu, Jun-Yan, 2017.

PRETRAINED MODEL: VGG-19

ARCHITECTURE

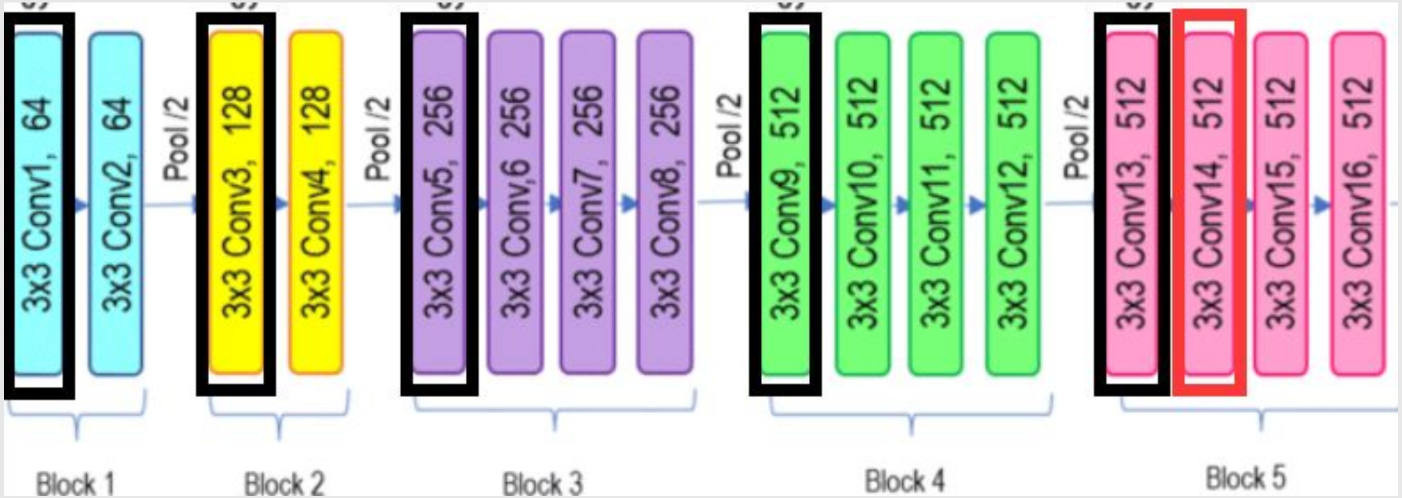


OVERVIEW

- 19 layers CNN
- Pretrained model on *ImageNet* classification
- Employ it for feature extraction

PRETRAINED MODEL: VGG-19

EXTRACTED LAYERS



Black Rectangle:
Style extraction

Red Rectangle:
Content extraction

PRETRAINED MODEL: VGG-19

STYLE FORMULA

- For each style layer extracted, we can calculate style matrix by the means and correlations across the different feature map channels

$$G^l_{cd} = \frac{\sum_{ij} F^l_{ijc}(x) F^l_{ijd}(x)}{IJ}$$

LOSS FUNCTION=STYLE LOSS + CONTENT LOSS

$$\sum_{Style\ Layers\ x} mse(G_x(target), G_x(generated))$$

$$\sum_{Content\ Layers\ x} mse(C_x(target), C_x(generated))$$

PRETRAINED MODEL: VGG-19

TRAINING? GENERATION!

- Loss function : L(target content, target style, generated image)
- We have a loss function, so we can do gradient descent, but on which parameters?
- The trainable parameters are exactly the generated image
- The gradient descent “training” process is actually the “generation” process

$$\underset{y_{generated}}{\mathrm{argmin}}\,(L(y_{generated}, y_{style\_target}, y_{content\_target}))$$

PRETRAINED MODEL: VGG-19

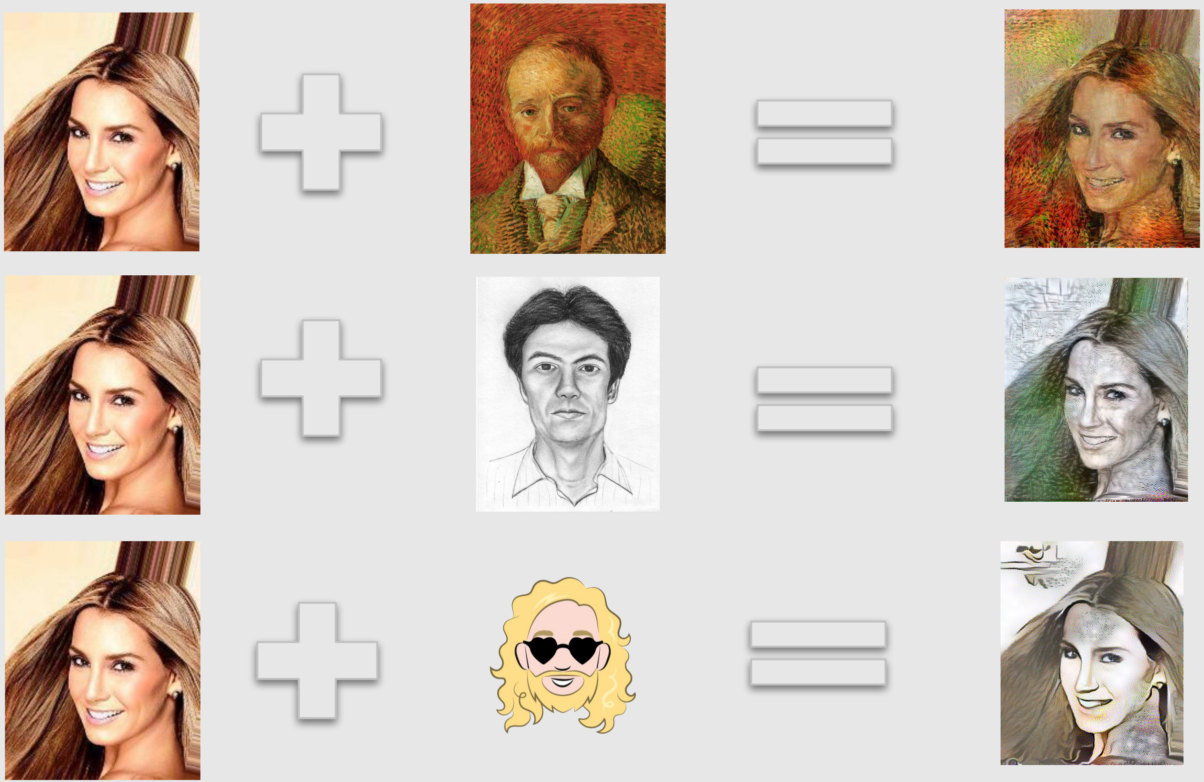
PROS

- Easy to implement
- Computation efficiency
- Only need one graph as style graph instead of large training dataset

CONS

- No model saved
- Need to generate for every new input style/content images
- Hard to compare the loss across different generation tasks

PRETRAINED MODEL: VGG-19

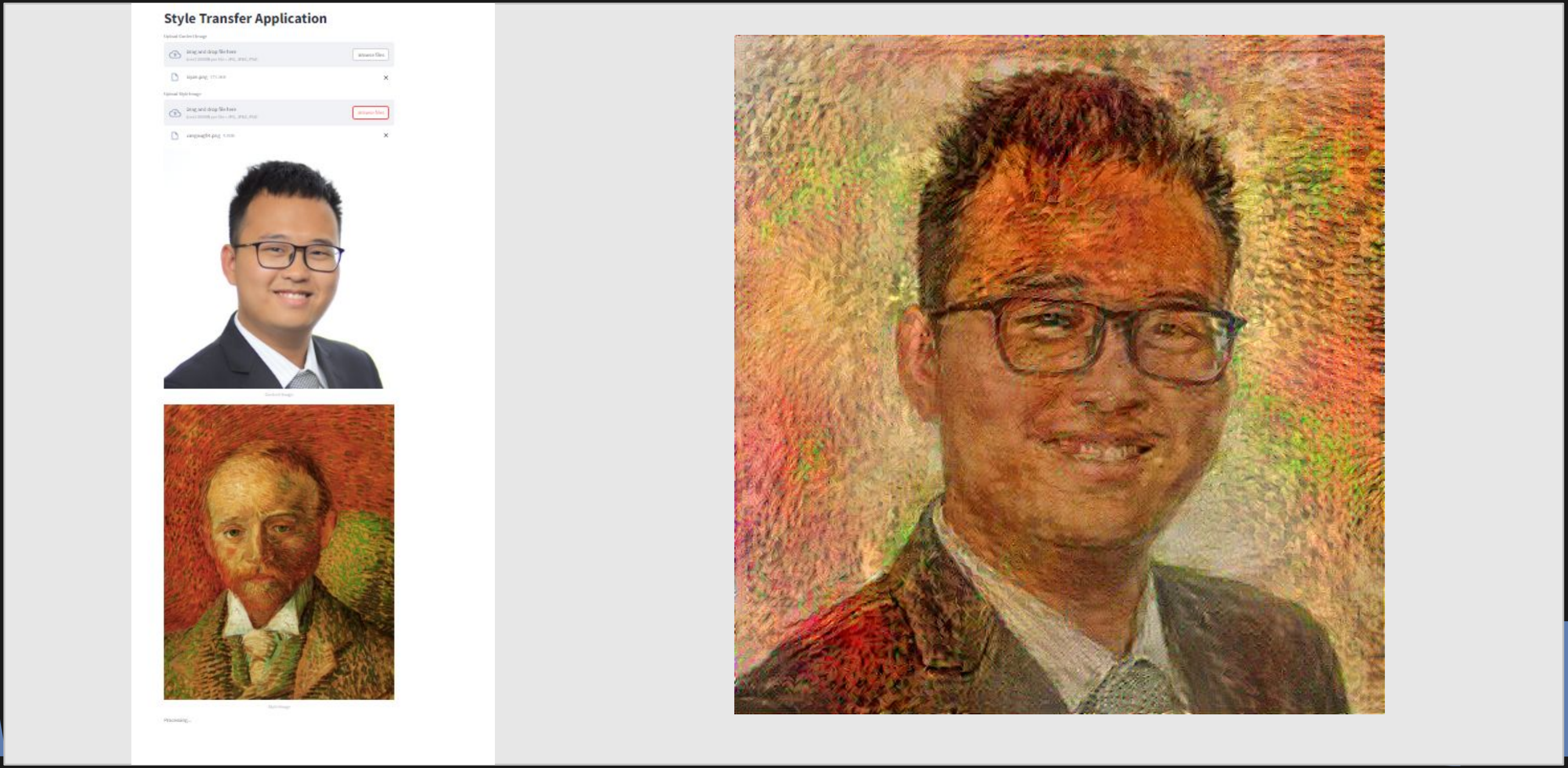


ARTIFICIAL
INTE
(AI)

APPLICATION EXAMPLE:
BY STREAMLIT

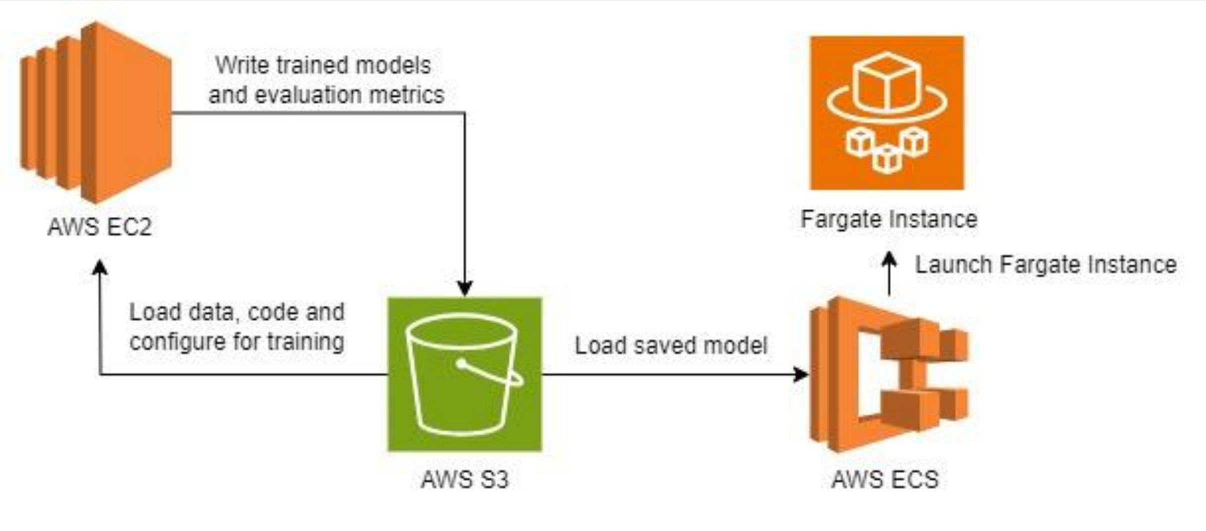
(AI)

STREAMLIT DEPLOYMENT



MODEL OPERATIONS

OPERATION ARCHITECTURE



MODEL MAINTENANCE AND PARAMETER UPDATE

MAINTENANCE

- Infrastructure: AWS
- Fetch new data from AWS S3 and retrain the model
- Use CloudWatch to keep track of the application health

PARAMETER UPDATE

- Model can read config file to customize the neural network structure and hyperparameters
- Can set periodical retraining for parameter updates

CONCLUSION

Image classification task:
Both high accuracy, customized CNN slightly higher.

Style transfer task:
Depends on style:
VGG-19 models better transfer styles with texture features, such as the *Van Gogh style*.
For simpler styles like *sketches*, both the CycleGAN and VGG-19 models performed well.
However, both models struggled with the *cartoon style*.

Our findings emphasize the importance of model selection and customization based on the task requirements and dataset properties. While pre-trained models like InceptionV3 and VGG-19 offer robust performance and ease of implementation, customized models can provide superior results with sufficient tuning and resource investment.

ARTIFICIAL INTE (AI)

REFERENCES

[1] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In *Proceedings of the IEEE international conference on computer vision*, pp. 2223-2232. 2017.

[2] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. “Rethinking the Inception Architecture for Computer Vision.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016. https://doi.org/10.1109/cvpr.2016.308.

[3] VGG-19 architecture [39]. VGG-19 has 16 convolution layers grouped into... | download scientific diagram. Accessed May 31, 2024. https://www.researchgate.net/figure/VGG-19-Architecture-39-VGG-19-has-16-convolution-layers-grouped-into-5-blocks-After_fig5_359771670.

[4] “Neural Style Transfer : Tensorflow Core.” TensorFlow. Accessed May 30, 2024. https://www.tensorflow.org/tutorials/generative/style_transfer.

(AI)

ARTIFICIAL INTE (AI)

DATASET LINKS

[1] https://www.kaggle.com/datasets/jessicali9530/celeba-dataset
[2] https://www.kaggle.com/datasets/arbazkhan971/cuhk-face-sketch-database-cufs
[3] https://www.kaggle.com/datasets/ipythonx/van-gogh-paintings
[4] https://www.kaggle.com/datasets/imreallyjohn/cartoonset10k

(AI)

APPENDIX: DATA PREPROCESSING

DATA LOADING

Combine gender and partition data into a single DataFrame.

DATA SAMPLING

Define a function to load and reshape images. Sample an equal number of male and female images for training, validation, and testing.

DATA SPLITTING

Split data into training, validation, and test sets using the sampling function

DATA AUGMENTATION

Use ImageDataGenerator for data augmentation. Apply transformations such as rotation, shift, shear, zoom, and flip.

ARTIFICIAL
INTE
(AI)

THANK YOU !

(AI)