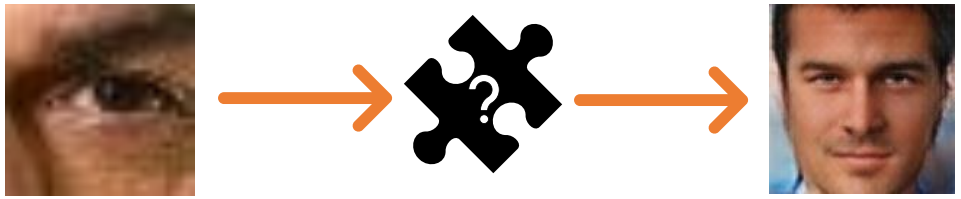
A New Approach to Eye-to-Face Synthesis

Rex Tse, Peter Ng, Ronnie Jok

Background and Innovation

Project Background

- The reference project aims to generate a face from a pair of eyes
- Main differences
  - Single eye vs two-eye input
  - Standard deviation loss utilized in our model

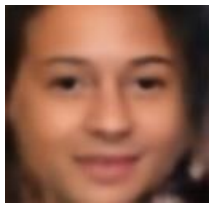
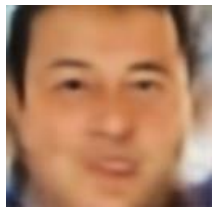
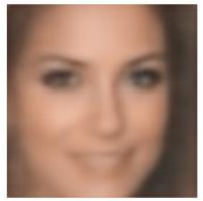
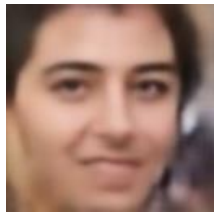


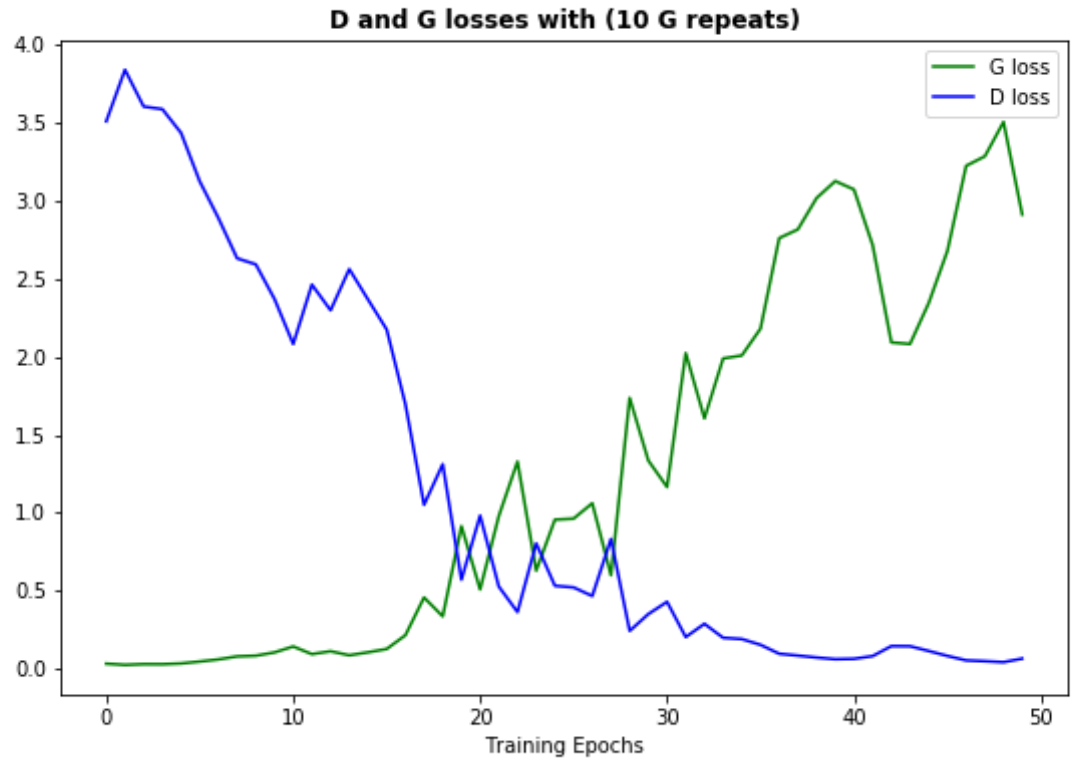
From Eyes to Face Synthesis:

a New Approach for Human-Centered Smart Surveillance

Problems with Current Models

L1	• Blurry • Mode collapse (insufficient variation)
GAN	• Unstable • Mode collapse
Conditional GAN	• Unstable

REAL	L1	REAL	L1
			
			
			



Inspiration of Standard Deviation Loss

Standard Deviation

Metric to measure variation of an image batch

New Loss?

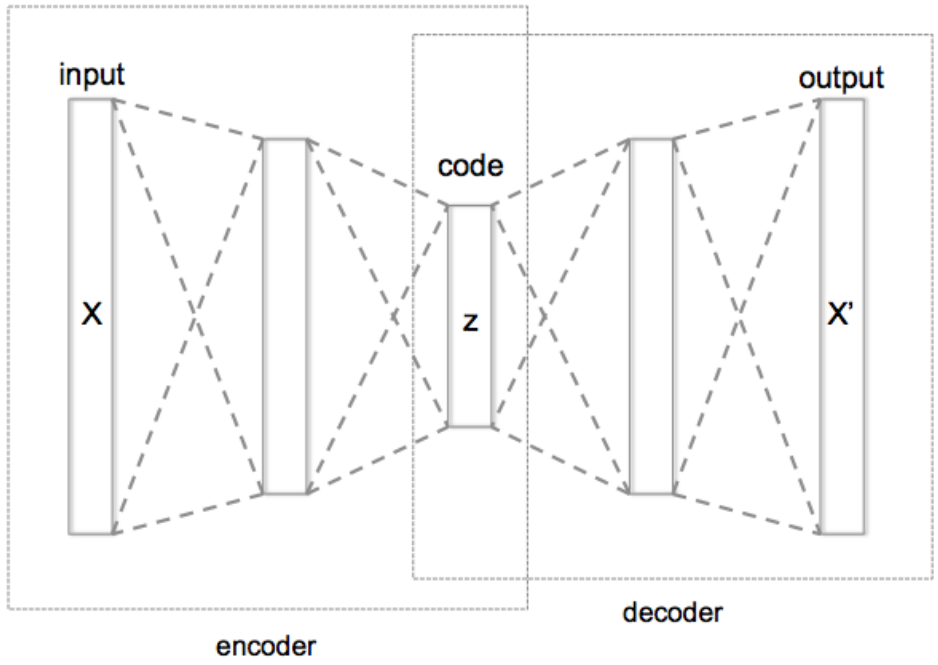
Compare standard deviation of generated batch and target batch

Against Mode Collapse?

As the loss becomes minimized, variation of output becomes similar to the target batch → No mode collapse!

Methodology

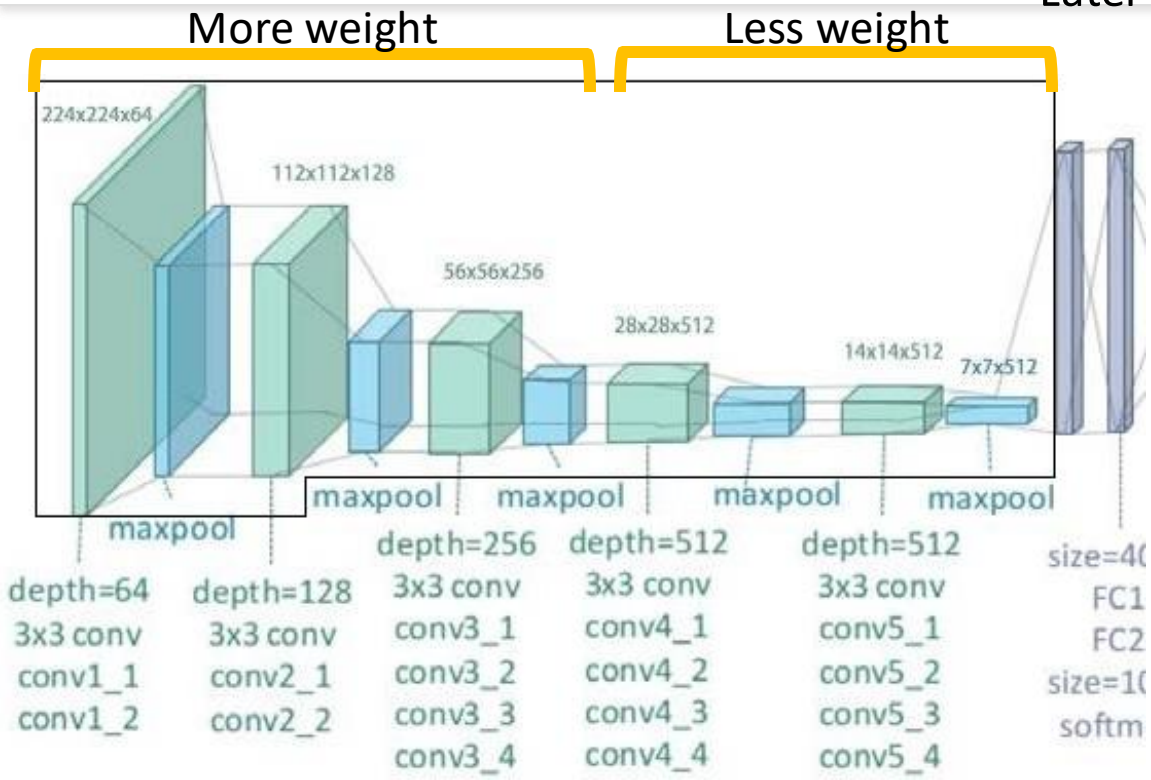
Autoencoder

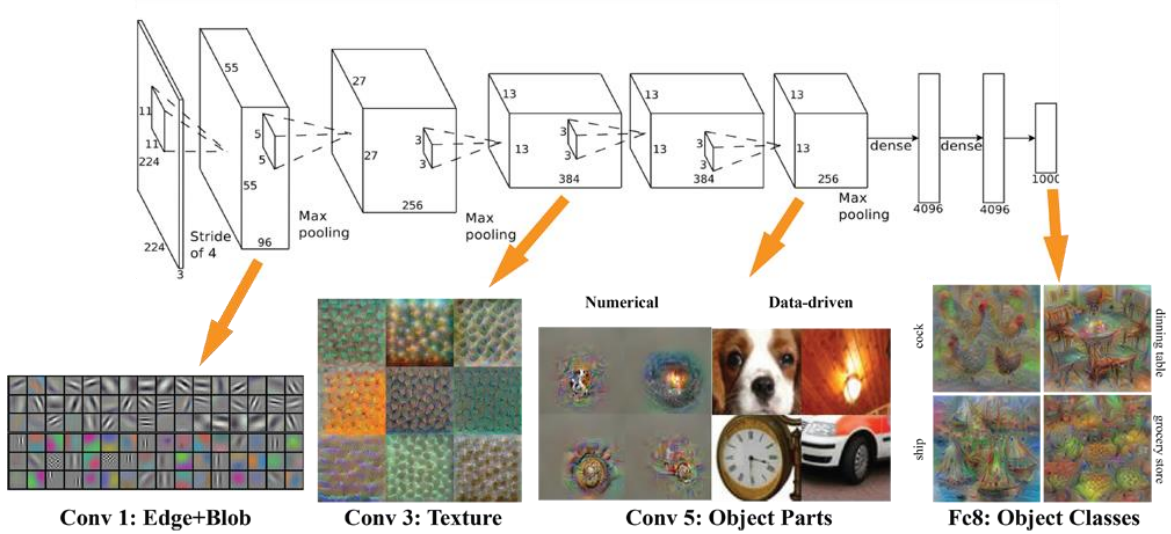


- Encoder consists of blocks of convolutional layers and maxpooling layers, which is then flattened and condensed into a vector
- Decoder consists of blocks of convolutional layers and upsampling layers, which expands an encoded-eye vector to a generated face
- Batch normalization is used

VGG19 and Feature Loss

- The highlighted part shows the layers we utilize in the training process
- L1 loss is used to compare the features grasped from the generated and target batch
- Early layers grasp local features
- Later layers grasp global features





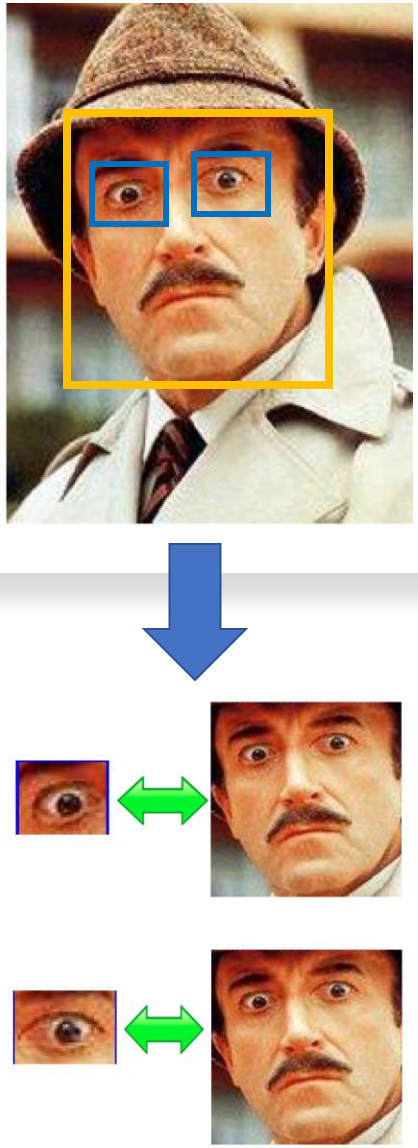
Standard Deviation Loss

- $\sigma_{loss} = |\sigma(y_{true}) - \sigma(y_{pred})|$,where:
- $\sigma(B) = \frac{\sum_{k=1}^{n} \sigma(k)}{n}$,where B represents a batch of images, k represents a pixel, n represents the total number of pixels in one image

Data

Data Input and Processing

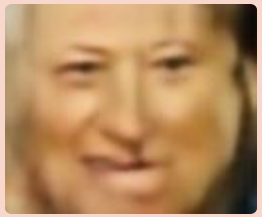
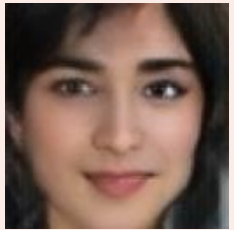
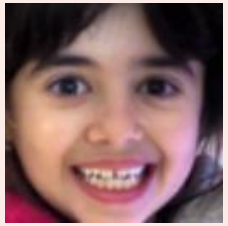
- 3100 faces from the CelebA dataset
- OpenCV Library to crop faces and eyes using the Haar-cascade algorithm
- Both eyes are cropped separately → grasping the correlation between the eyes to the face
- The pixel value are then normalized to the range of 0 to 1 to ensure better training



Result and Summary

Result Analysis (Clearness)

The faces generated by our model is much clearer than the L1 counterpart

Our model	L1 loss model	Target image
		
		
		

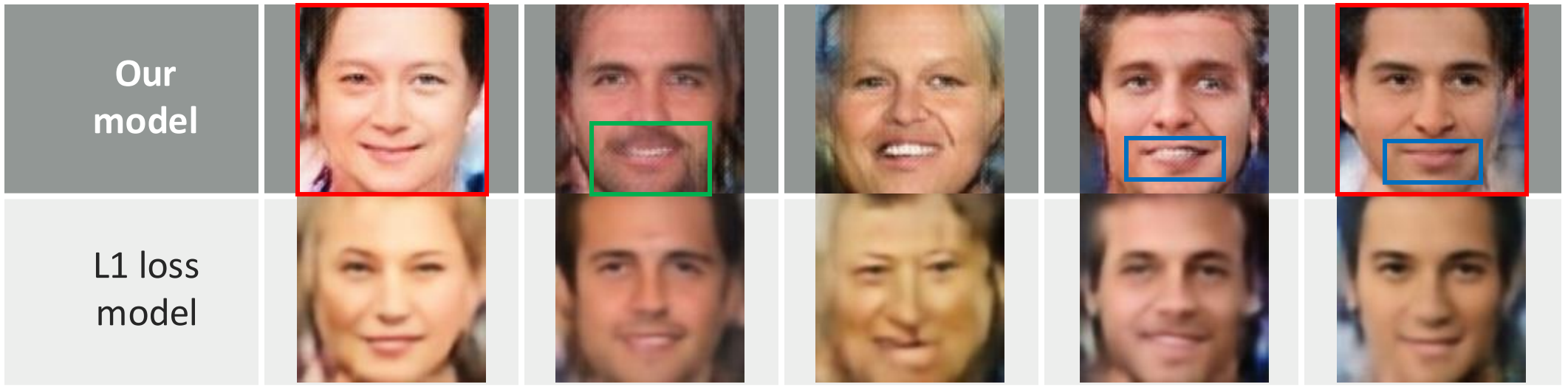
Result Analysis (Face Features)

Our model is accurate in predicting features such as:

- Moustache
- Face shapes
- Eye shapes

Our model	Target image
	
	
	

Result Analysis (Variety)



- Opening of mouth
- Angle of lighting
- Moustache and beard

Quantitative Comparison

	L2 Loss	Feature Loss	Standard Deviation Loss of Features
Our model	0.0570	161.7686	41.7407
L1	0.0532	158.1259	75.5827

- Susceptibility of pixel-wise losses and pure feature loss to blurriness and mode collapse
- Effectiveness of standard deviation loss

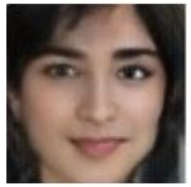
Flaws in Our Model

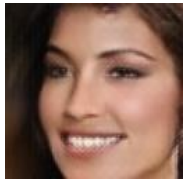
- Fails to generate some special facial features, such as very fat or skinny faces
- Performance is unsatisfactory if the eye is too tilted, too small or the lighting is not optimal
- Cannot generate fully photorealistic photos as the VGG model ignores background

Summary and next steps

- provided a possible alternative for replacing GAN networks in image generation.
- Explore more possibility to expand the usage of our Standard Deviation Loss on different models
- In the future, we may even create models that can generate a human face from other partial facial features, e.g. mouth and nose.

Result Demonstration

our model	real
	
	
	
	
	

Our model	real
	
	
	
	
	

Acknowledgment

- University of Hong Kong Professor Wong Ngai
- Steven Luo (Captain of AI Team)

Attribution

SLIDE	CONTENT	SOURCE
4	Image of double-eye-to-face synthesis	From Eyes to Face Synthesis: a New Approach for Human-Centered Smart Surveillance
5	Image of generator loss and discriminator loss during GAN training	https://stackoverflow.com/questions/44313306/dcgans-discriminator-getting-too-strong-too-quickly-to-allow-generator-to-learn
8	Image of structure of an autoencoder	https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png
9	Image of structure of the VGG network	https://www.researchgate.net/figure/Illustration-of-the-network-architecture-of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356

Attribution

SLIDE	CONTENT	SOURCE
9	Image of features grasped by the VGG network	https://www.researchgate.net/figure/Illustration-of-the-network-architecture-of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356
12,13	Image of a person	Large-scale CelebFaces Attributes (CelebA) Dataset

Attribution (Test Batch)

- Boris Yeltsin: By Kremlin.ru, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=67546416
- Lee Hsien Loong: By Prime Minister's Office (GODL-India), GODL-India, https://commons.wikimedia.org/w/index.php?curid=87653410
- FFHQ (Flickr-Faces-HQ) dataset