faculty of
computers and artificial
intelligence

cairo university

PICTALK GENERATOR

Graduation project

SUPERVISED BY:
DR/ Doaa Saleh
DR/Ghada Dahy

Academic year
2023-2024

Our Team

Alaa

ALMoatasim

Arwa Sallam

Esraa

Mariam

Yara

Agenda

1. Problem definition
2. Introduction
3. Literature review
4. Proposed System
5. Experimental result
6. Conclusion

01

PROBLEM DEFINITION

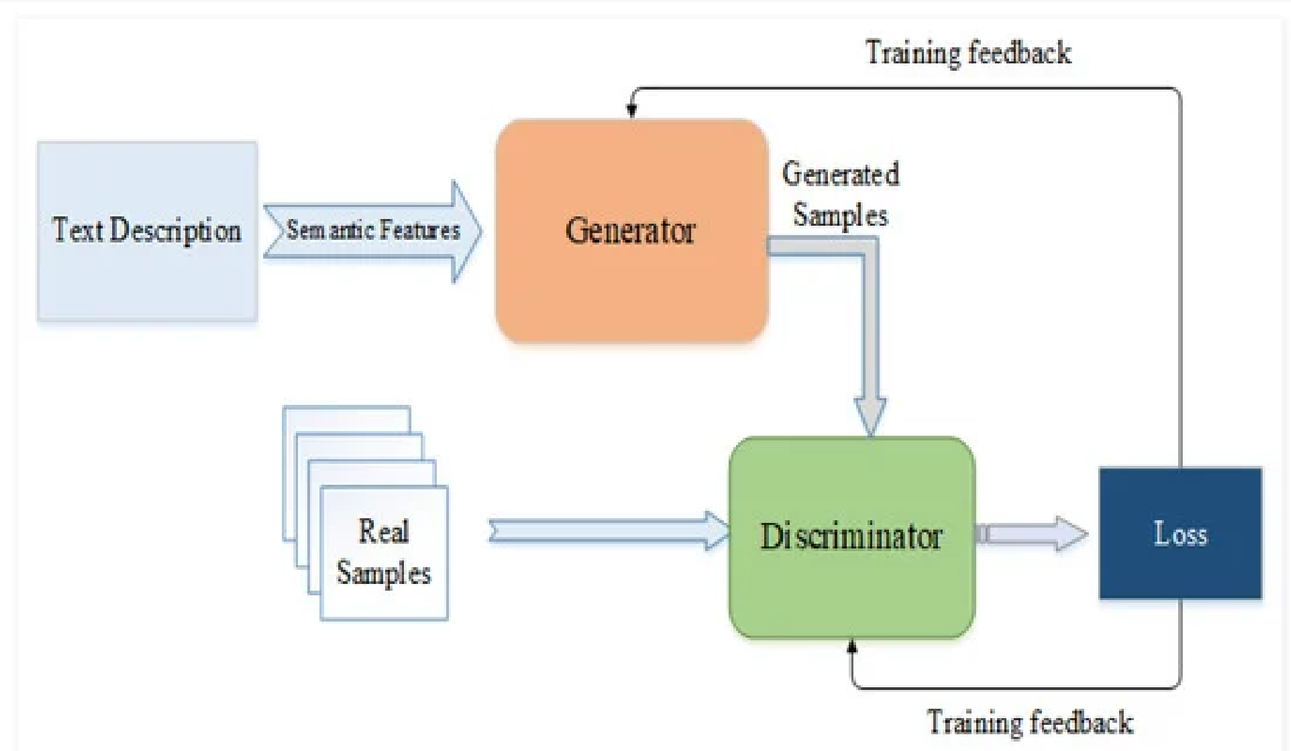
Problem Definition

In criminal justice, two key issues are inefficient suspect identification from subjective witness descriptions and the exclusion of blind individuals from visual evidence in courtrooms. This leads to inaccurate investigations and an information gap affecting fair outcomes.

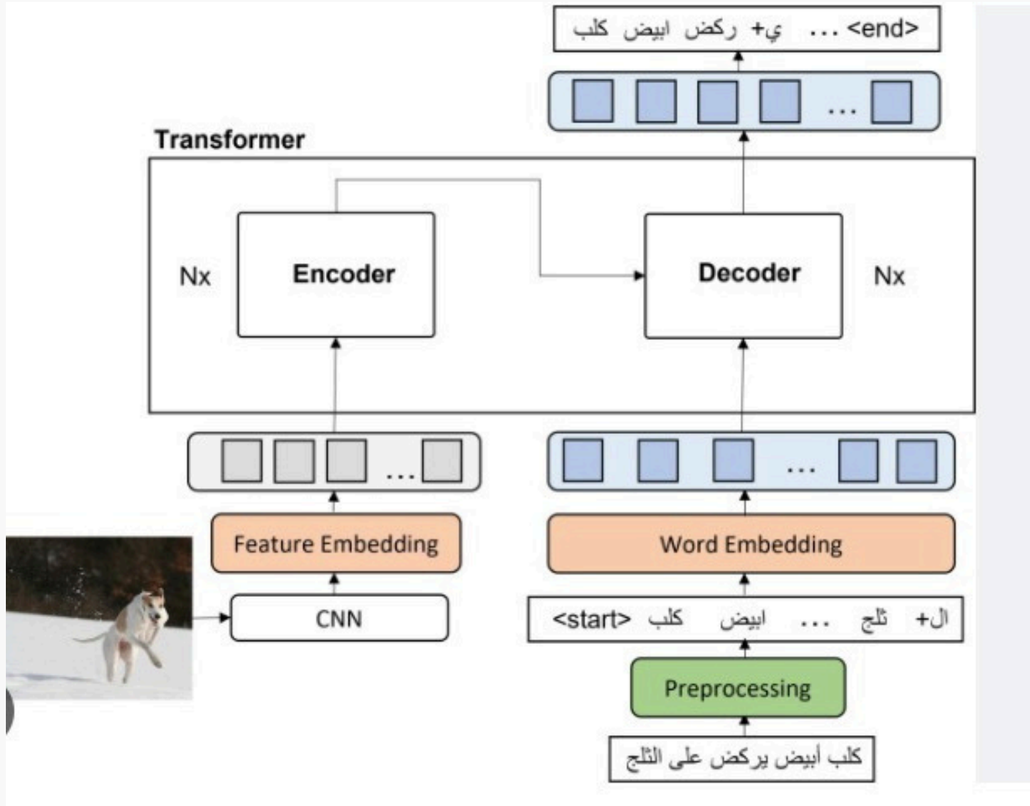
02

INTRODUCTION

1. The DF GAN model generates realistic faces from Arabic and English textual descriptions, aiding criminal investigations by providing visual representations of suspects for quicker identifications. It encodes text descriptions into numerical data, which is used by a generator to create images, while a discriminator distinguishes between real and generated images. This model effectively produces facial images that closely match the given descriptions.



2. Our AI project features an image captioning system that aids visually impaired individuals by converting visual data into descriptive sentences, enhancing situational awareness and safety. We use two transformer-based models, one for Arabic captions and another for English. Initial results are promising, laying a strong foundation for future enhancements.



03

LITERATURE REVIEW

model1

approach 1

Generative adversarial networks (GANs) are models designed to produce realistic samples. Conditional GANs generate data based on specific conditions. Control GANs enhance this by incorporating a classifier alongside the generator and discriminator.

approach 2

Self-Attention Generative Adversarial Networks (SAGANs) introduces a self-attention mechanism. The self-attention module helps with modeling long range, multi-level dependencies across image regions.

approach 3

StackGAN generates high-resolution images. AttnGAN introduces the cross-modal attention mechanism MirrorGAN regenerates text descriptions from generated images for text-image semantic consistency. SD-GAN employs the Siamese structure. DM-GAN introduces the Memory Network to refine fuzzy images.

approach 4

Deep Fusion GAN (DF-GAN) differs significantly from previous methods by generating high-resolution images directly with a single-stage backbone in one generator, avoiding the complications of using multiple generators. This makes DF-GAN simpler yet more effective at synthesizing realistic and text-matching images.

model2

approach 1

Template-based approaches have fixed templates with a number of blank slots to generate captions. In these approaches, different objects, attributes, and actions are detected first and then the blank spaces in the templates are filled.

approach 2

Retrieval-based image captioning involves retrieving captions from a set of existing captions. Similar images and their captions are selected as candidate captions, from which captions for the query image are chosen.

Back to Agenda

approach 3

Novel captions can be generated from both visual space and multimodal space. A general approach of this category is to analyze the visual content of the image first and then generate image captions from the visual content using a language model.

approach 4

Recent advancements in Transformer-based approaches have significantly improved image captioning, particularly in capturing semantic relationships and spatial dependencies.

Back to Agenda

Research Gap

PickTalk Offers a two-in-one solution for image understanding, the project contains two aggregate models which are text-to-image and image-to-text, its image captioning transformer forges detailed descriptions, while the DF-GAN conjures suspect sketches, empowering investigations like never before

we did not stop with the promising results, we're building a dedicated image captioning dataset and pushing training boundaries with increasing number of epochs in training to grasp the scene's soul in word and image

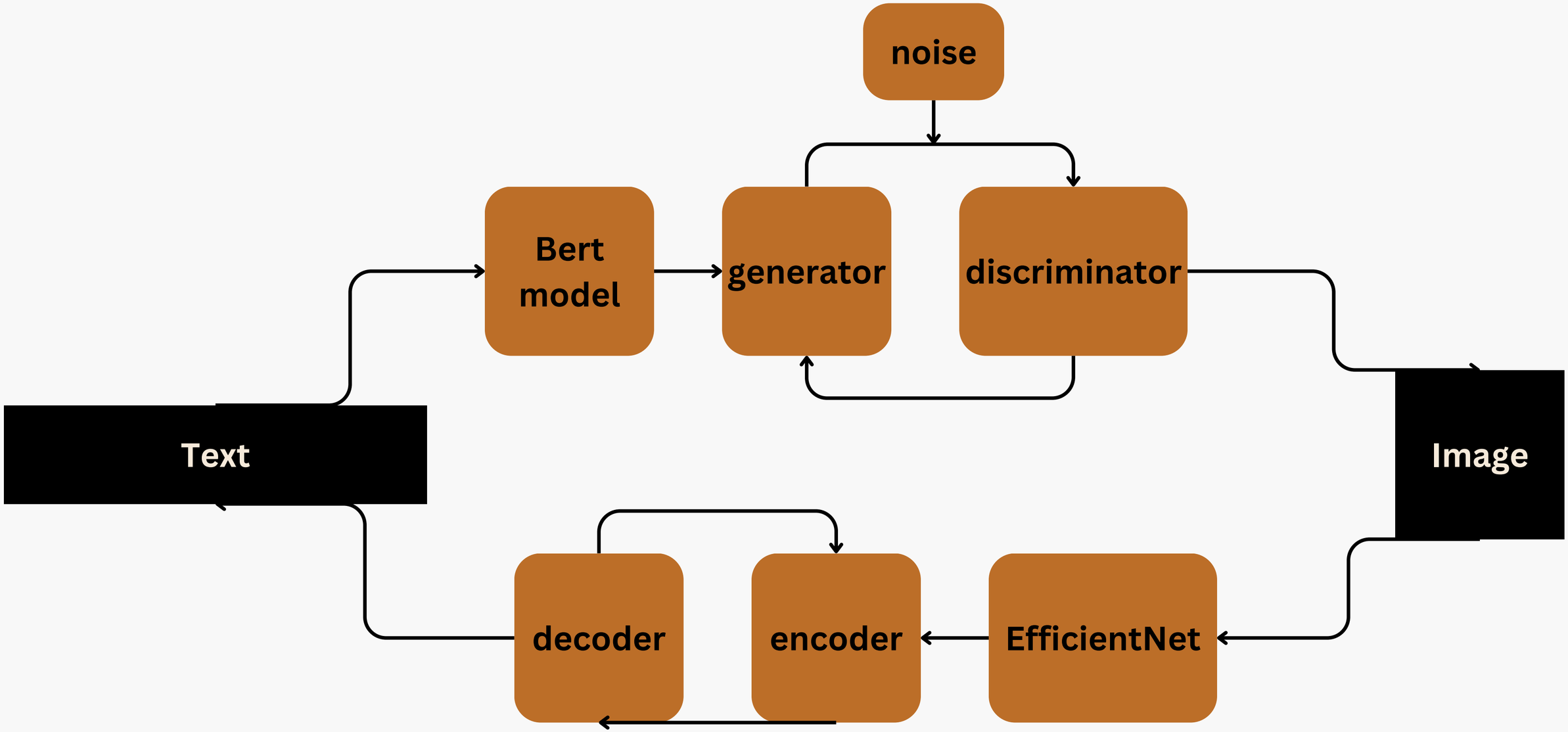
transformer-based approach for generative captioning offers several advantages: Higher creativity and flexibility, Adaptability to unseen scenarios, Capturing complex relationships, and State-of-the-art results.

our DF-GAN has better object shapes and realistic fine-grained details. Comparing the text-image semantic consistency, we find that our DF-GAN can also capture more fine-grained details in text descriptions which helps avoid the "fuzzy shape". it generates high-resolution images directly by a one-stage backbone in one generator, it avoids the entanglements between different generators.

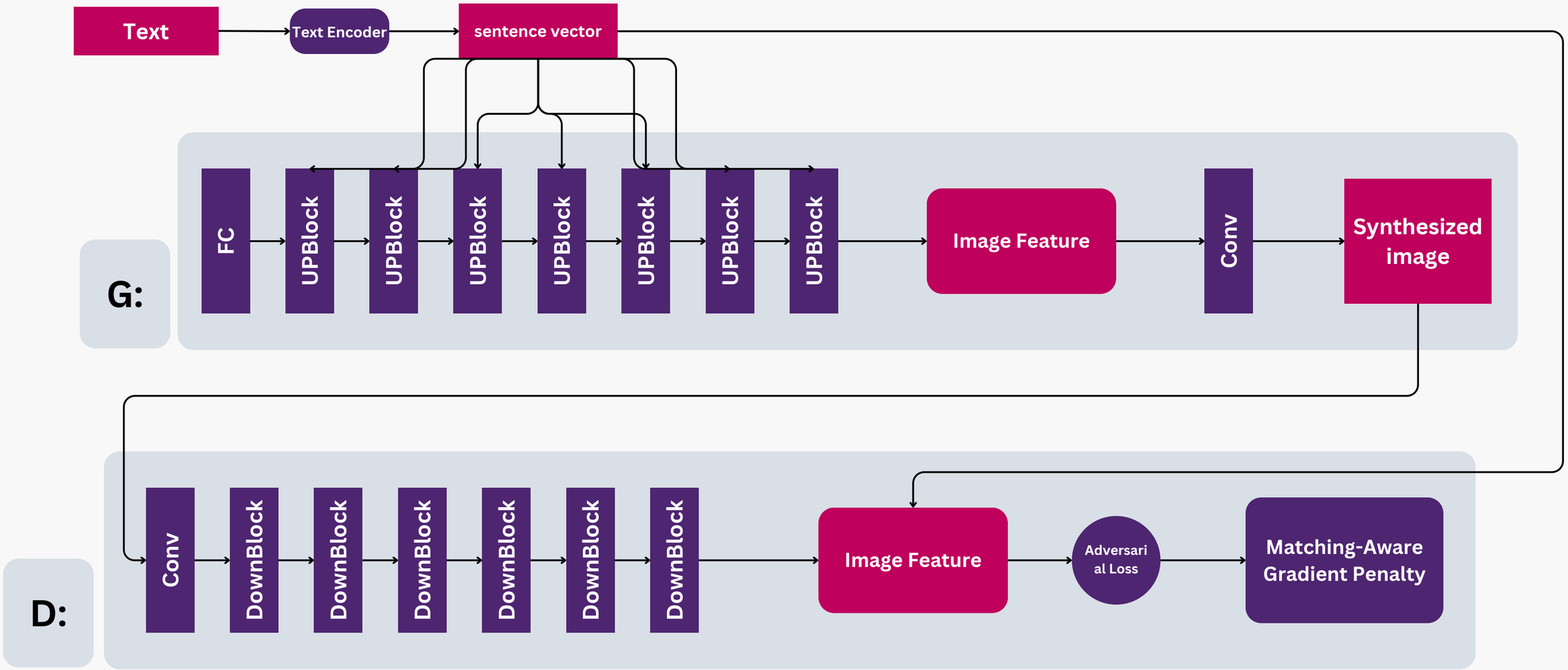
04

PROPOSED MODEL

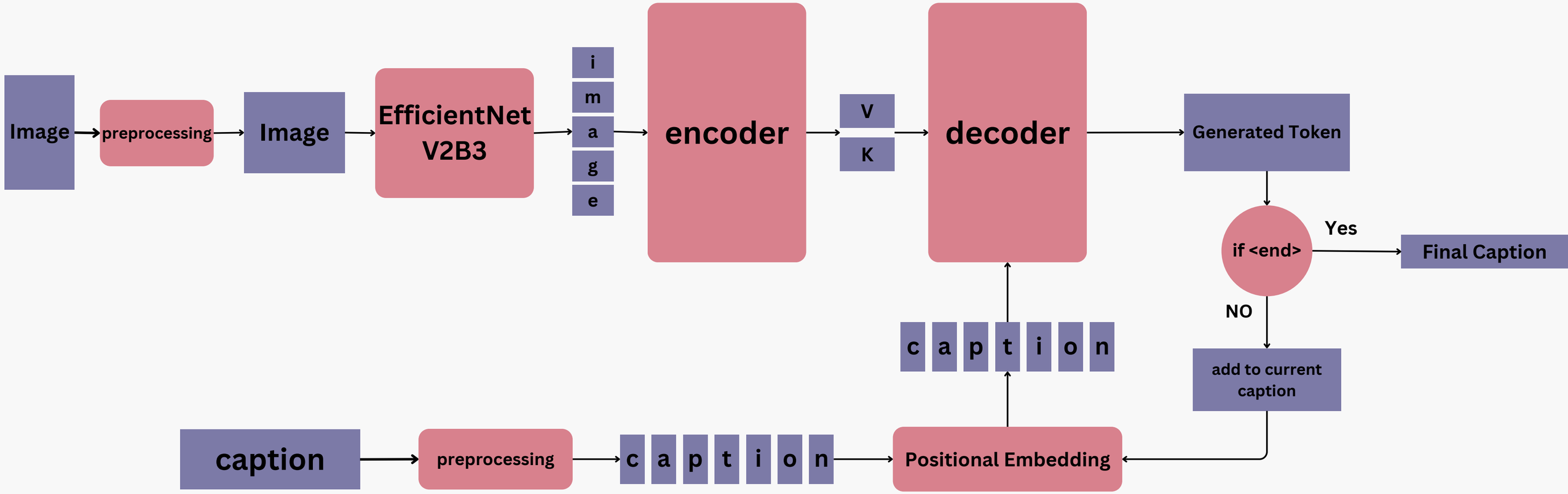
SUMMARY OF THE MODELS

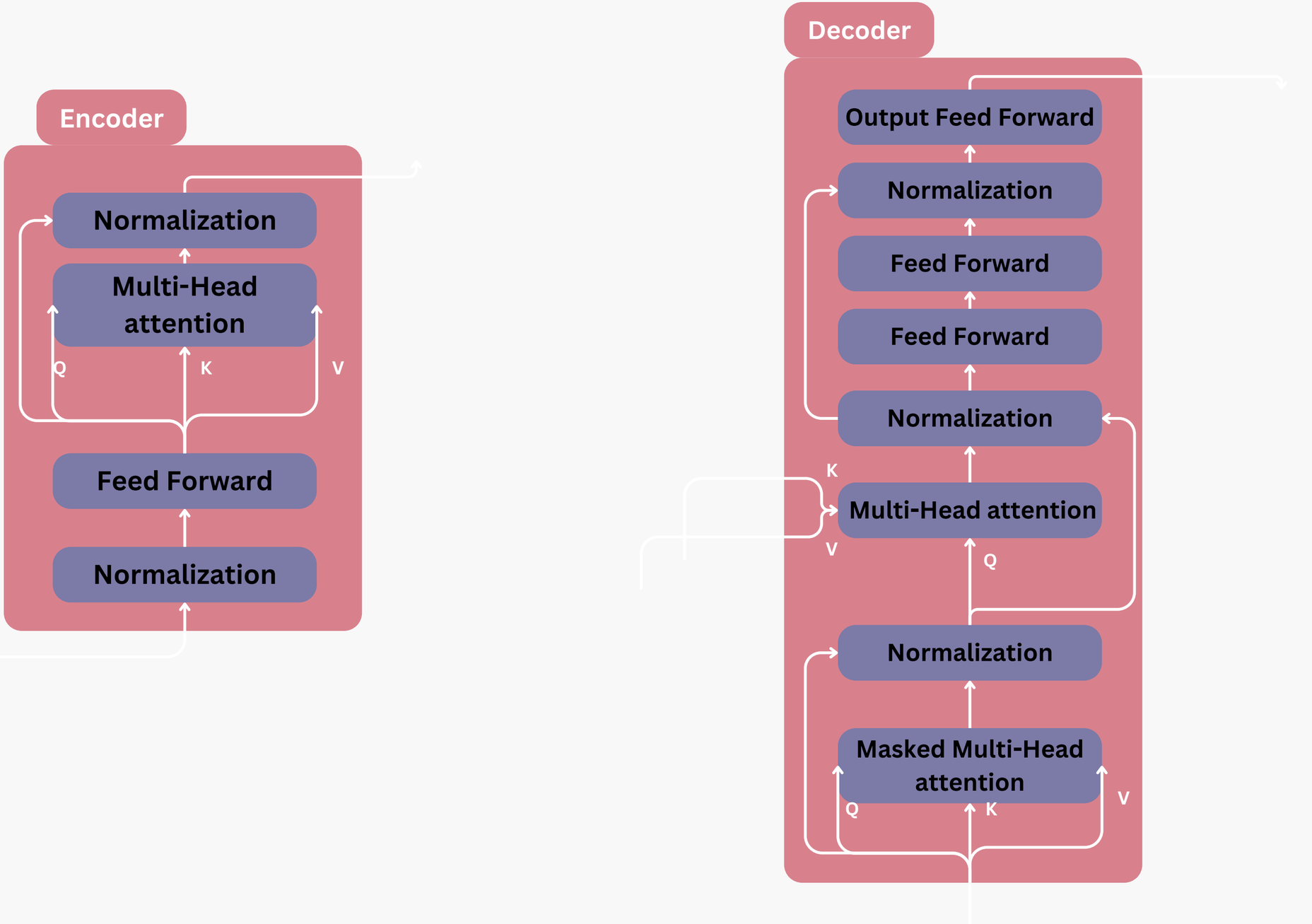


Proposed MODEL 1



Proposed MODEL 2




Encoder
Normalization
Multi-Head attention
Q
K
V
Feed Forward
Normalization
Decoder
Output Feed Forward
Normalization
Feed Forward
Feed Forward
Normalization
K
Multi-Head attention
V
Q
Normalization
Masked Multi-Head attention
Q
K
V

05

EXPERIMENTAL RESULT

Datasets

I

CelebA: CelebFaces Attributes Dataset, it great for face detection, particularly for recognizing facial attributes, it annotated with 40 different attribute labels per image, This data was originally collected by researchers at MMLAB, The Chinese University of Hong Kong.

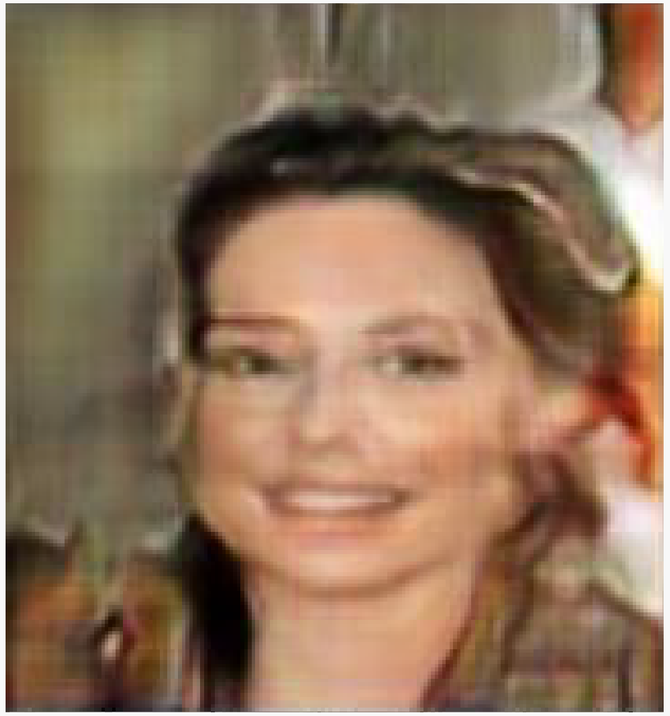
In this draft we were able to train on 25,000 image for 22 epoch which is a significant improvement from the first draft where we was only able to train on 20,000 image for 4 epoch

II

Our GAN model achieved a high Inception Score of 9.127, indicating diverse and quality-generated images. The Fréchet Inception Distance (FID) of 38.689 shows good image quality with room for improvement. These results highlight strong performance with opportunities for refinement.

Score	value
Inception Score (IS)	9.127
Fréchet inception distance (FID)	38.689

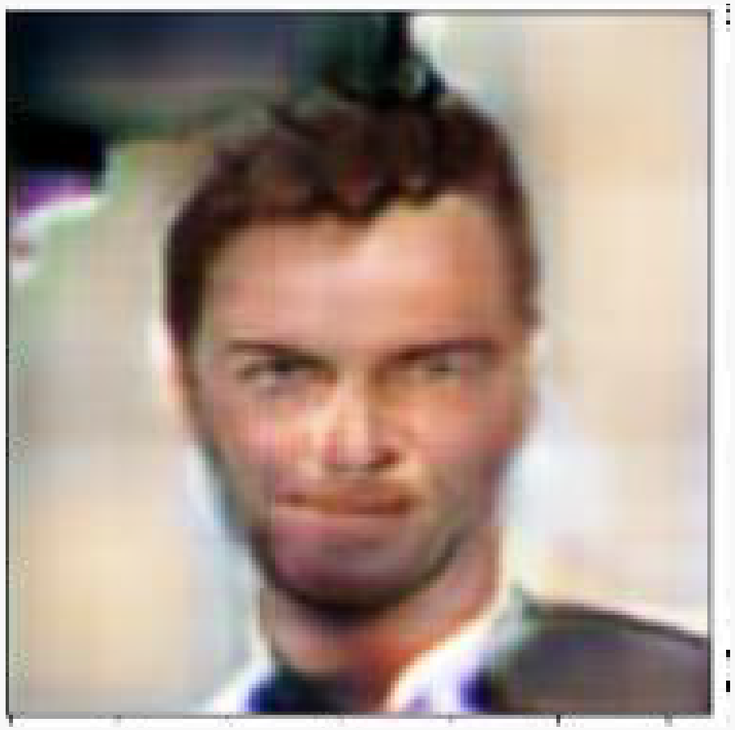
old version



" The woman has high cheekbones. She has straight hair which is brown in colour. She has arched eyebrows and a slightly open mouth. The smiling, young attractive woman has heavy makeup. She is wearing lipstick. "

new version

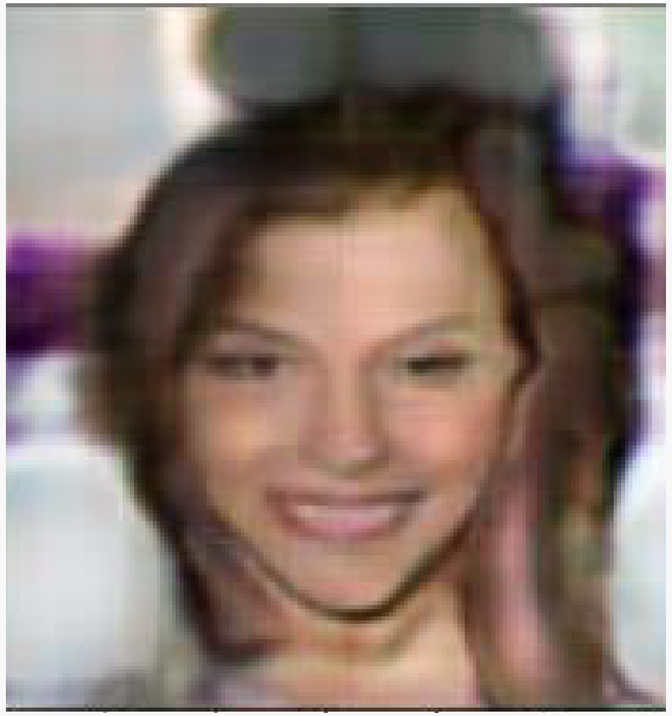
old version



" The criminal is a male. He has an oval face, short brown hair, narrow eyes, and a beard. "

new version

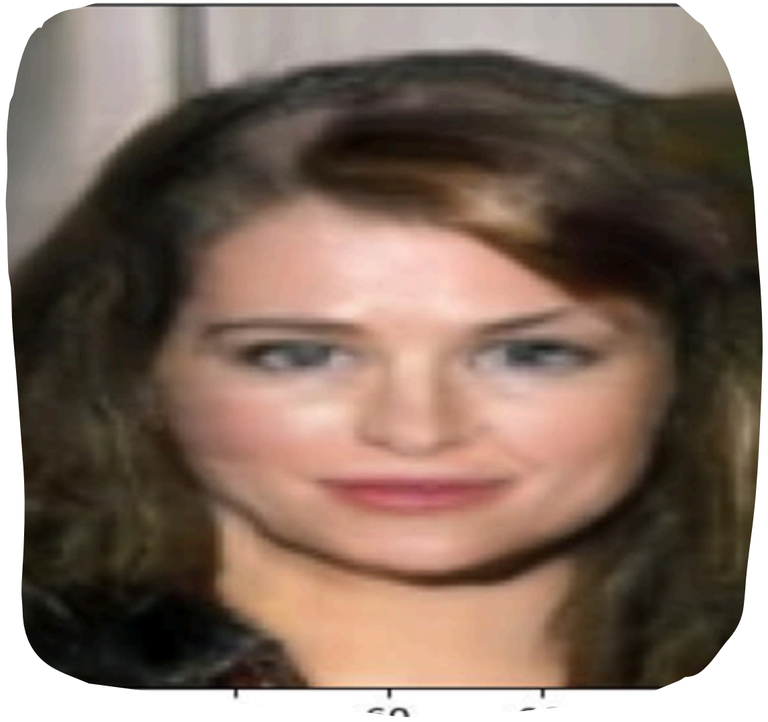
old version



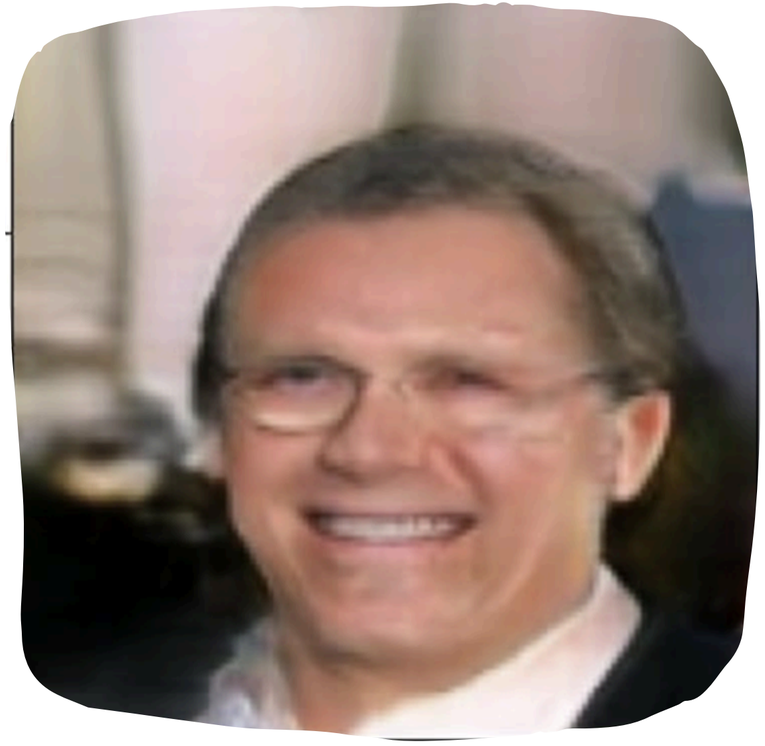
' The criminal is a woman. she is young and beautiful, and she has a straight blonde hair and a small nose. "

new version

The woman has high cheekbones and an oval face. Her hair is wavy. She has bushy eyebrows. The female is smiling, is attractive, young and has heavy makeup. She is wearing lipstick.



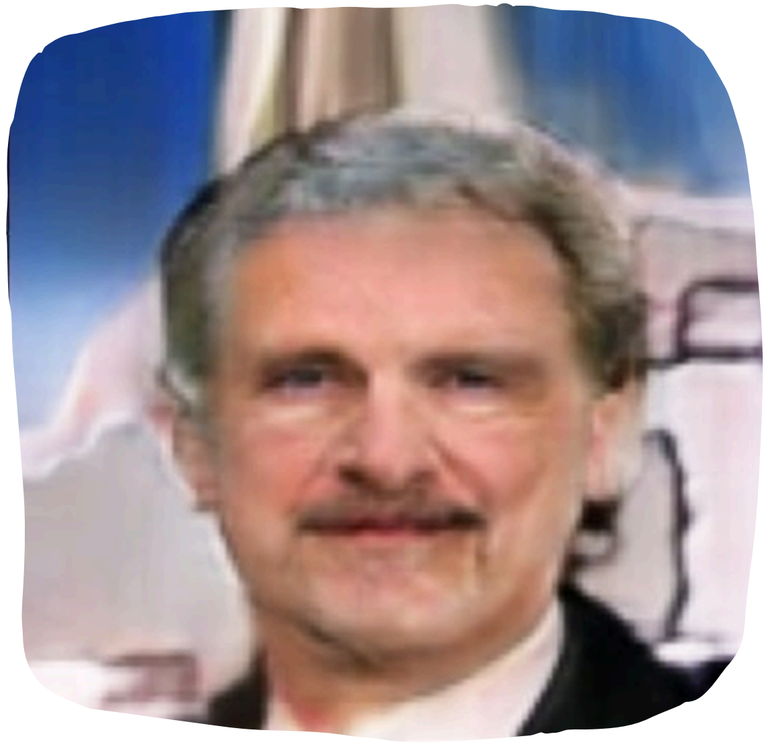
The gentleman has pretty high cheekbones. His hair is gray and straight. He has a slightly open mouth. The male is smiling. He is wearing eyeglasses.



He grows a 5 o' clock shadow and has sideburns. His hair is black. He has bushy eyebrows, a slightly open mouth and a pointy nose. The male is smiling, is attractive and young. He is wearing a necktie.

Your The female has high cheekbones. She has blond hair. She has arched eyebrows, big lips and a big nose. The lady is smiling, has pale skin and heavy makeup. She is wearing earrings and lipstick

The gentleman has high cheekbones. He sports a goatee. He has gray hair.



The male has pretty high cheekbones. He has black hair. He is smiling and is young.

METEOR	Ref #1	Ref #2	Ref #3	Ref #4	Ref #5
Image #1	10	9	74	100	7
Image #2	48	29	100	46	48
Image #3	20	100	36	23	69
Image #4	60	86	35	48	70
Image #5	14	34	60	79	12
Image #6	34	7	8	39	15
Image #7	5	10	14	83	20
Image #8	28	17	83	11	12
Image #9	84	82	74	71	54
Image #10	41	29	50	92	43

ROUGEL	Ref #1	Ref #2	Ref #3	Ref #4	Ref #5
Image #1	12	11	64	100	0
Image #2	57	36	100	62	57
Image #3	22	100	50	18	71
Image #4	62	88	35	46	50
Image #5	13	50	62	82	12
Image #6	50	19	23	50	35
Image #7	11	22	27	95	22
Image #8	36	20	95	27	19
Image #9	91	84	62	48	62
Image #10	48	32	57	100	52

BLUE	SCORE
Image #1	100
Image #2	100
Image #3	100
Image #4	59
Image #5	68
Image #6	35
Image #7	80
Image #8	71
Image #9	81
Image #10	82

V1: two men are standing in front of a building

V2: a man wearing a hat and formal suit and hat stands next to a building

V2: رجل يرتدي بدلة رسمية وقبعة يقف بجوار أحد المباني

a group of people walk on a sidewalk near flowers

two women and a little girl are eating ice cream

امرأتان كبريتان وفتاة صغرية تقفان بالخارج تحت الزهور

v1: a person jumping off of a dock into the water

v2: a person jumping off of a dock and into the water

V2: طفل صغري يقفز من فوق سطح السفينة الدوارة عل الماء

a skateboarder is doing a trick in the air

a man shocks the audience with his skateboard tricks

رجل بدون قميص ووشم علي ظهره في الهواء مع لوح التزلج

V1: a group of people ride bikes

V2: a group of men racing on their bicycles

V2: مجموعة من الدراجات يركبون درجاتهم

A man threatens a man with a gun

رجل يرتدي بدلة سوداء قتل رجلاً بمسدس

a man in flames while a fire

رجل إطفاء يلعب اشتعلت فيها النيران

a man is holding a knife

رجل يقتل رجال بسكين

a man in black shorts is beaten by a man in blue shorts

رجل يرتدي السراويل الزرقاء يقاتل رجلاً يرتدي السراويل السوداء

a man was hitten by a car

سيارة فضية في منطقة المترو

06

CONCLUSION

Our project helps visually impaired individuals in crime scenes and courtrooms by aiding law enforcement with accurate suspect images from verbal descriptions. We use transformer models for generating contextually relevant image captions in English and Arabic. Additionally, Deep Fusion GAN (DFGAN) creates high-resolution images from textual inputs in both languages. This approach enhances legal participation and law enforcement efficiency, showcasing significant AI advancements in global image interpretation and synthesis.

CONTACT US:

Alaa Ashraf
alaaashra709@gmail.com

Arwa Sallam
arwasallam6@gmail.com

Almoatasim
moutasem.hamdi14@gmail.com

Esraa Haytham
haiithamesraa19@gmail.com

Yara Mohamed
yaramo656@gmail.com

Mariam Mostafa
mosmariam69@gmail.com

Thank you!