

Text to Image Synthesis via Mask Anchor Points and Aesthetic Assessment

By: Samah Baraheem
Advisor: Dr. Tam Nguyen

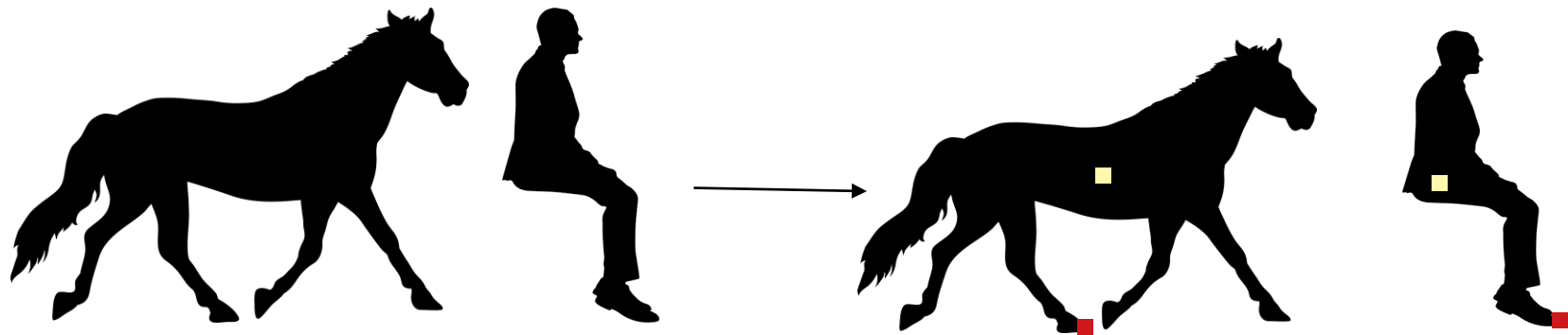
What is Text to Image Synthesis?

A man sits next to his
bicycle on the beach.



Our generated image

What is Mask Anchor Points?



What is Image Aesthetic Assessment?

- Photographic rules used to distinguish high-quality images from low-quality ones.
 - Rule of thirds and rule of formal balance are used in this work.
-

Motivation

1

A picture is worth a thousand words

2

Creating images is time consuming

3

Creating images is a tedious task

4

Generating images from input texts is faster and interesting

Motivation

State-of-the-arts Text-to-image methods suffer from various issues

- 1 Artifacts
- 2 Distortion
- 3 Edge details of shapes
- 4 Semantic information

Motivation

Examples of synthetic images by the state-of-the-arts

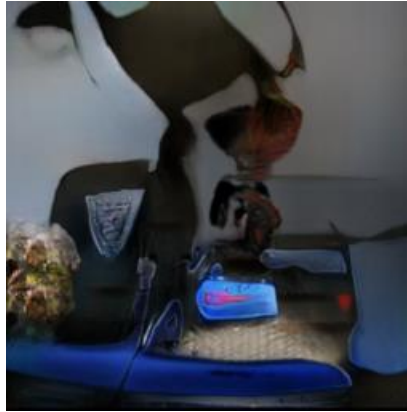
Caption

Three men sit on a couch.

Ground Truth



StackGAN + OP



AttnGAN



AttnGAN + OP



A child plays with a colorful kite on the field.



Objective

Synthesize images from input text
with:

1

Clear edge

2

Clear object

3

Good layout

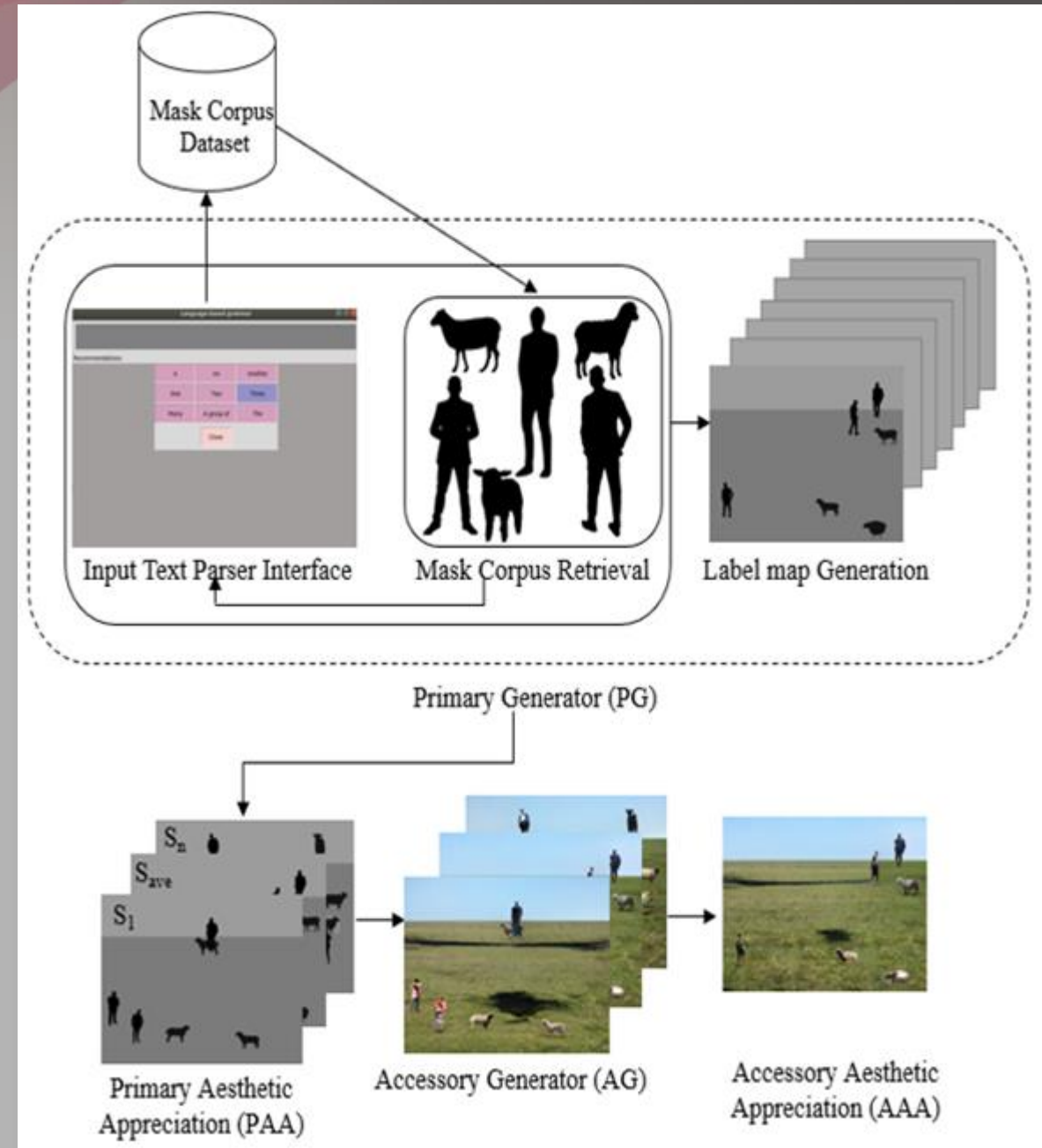
4

Good
aesthetic
assessment



Proposed Method

Proposed Framework





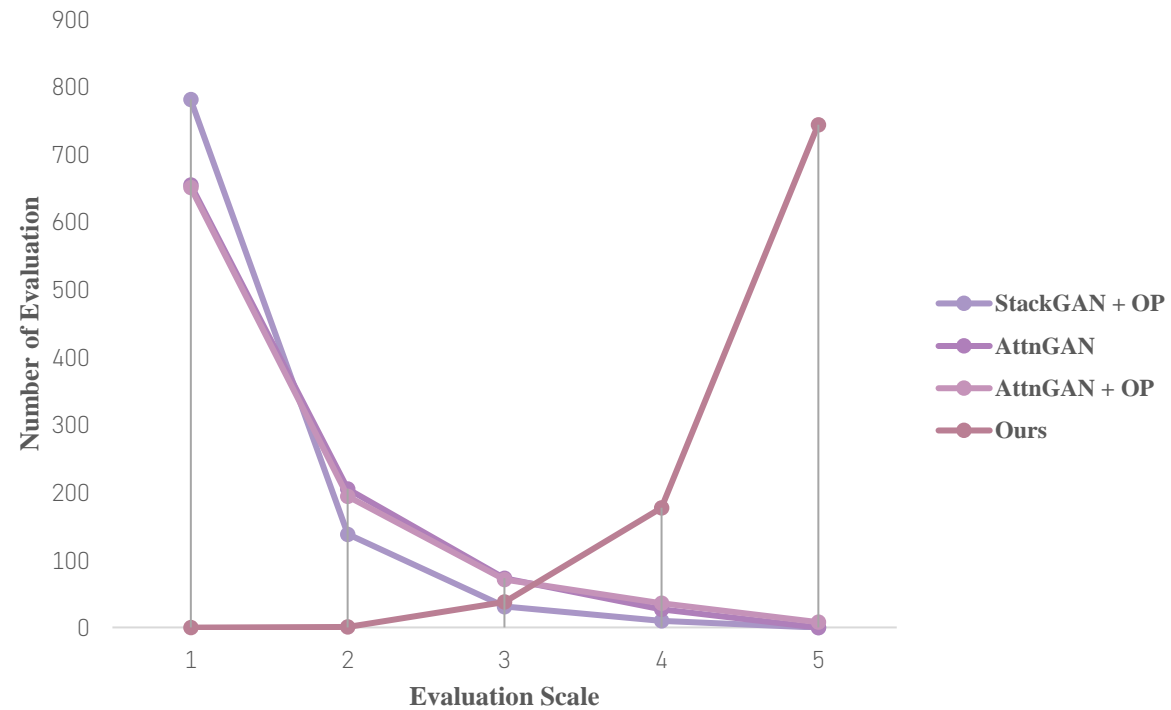
Results

Quantitative Results

Method	Mean of LPIPS \uparrow	FID \downarrow
StackGAN + OP	67.5	290.372
AttnGAN	68.7	251.191
AttnGAN + OP	68.3	266.118
Ours (via only mask anchor points)	72.0	244.311
Ours (via mask anchor point and aesthetic assessment)	73.5	240.411

Qualitative Results

























User Study



Scale factor:
1 → worst result
5 → best result

Visual Comparisons

with the State-of-the-art Image Synthesis

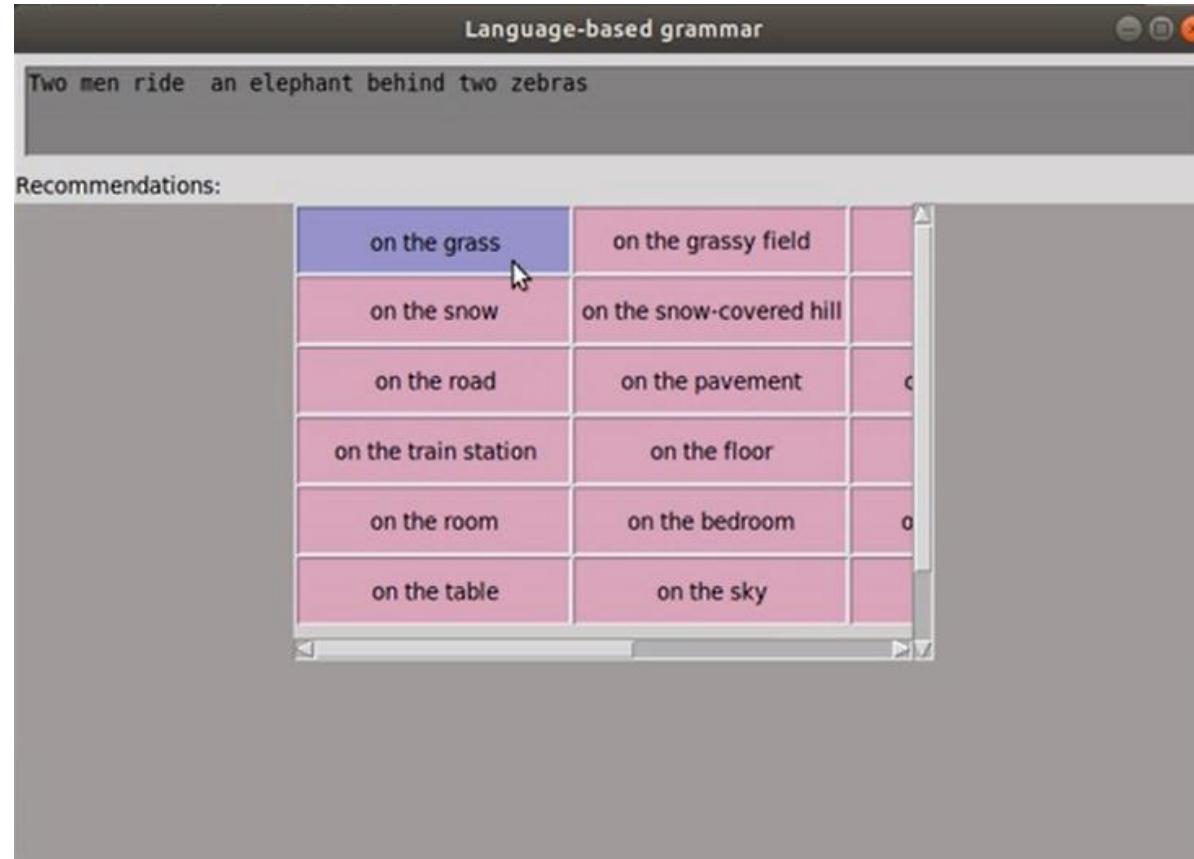
Caption	GT	StackGAN + OP	AttnGAN	AttnGAN + OP	Our generated mask map	Ours
A group of sheep walking down a path with a few stoppings to eat grass along the side.						
A man stands beside his surfboard on the beach.						
A man is out on the yard about to kick a ball on the grass.						
A bear stands on the grass.						



Demo

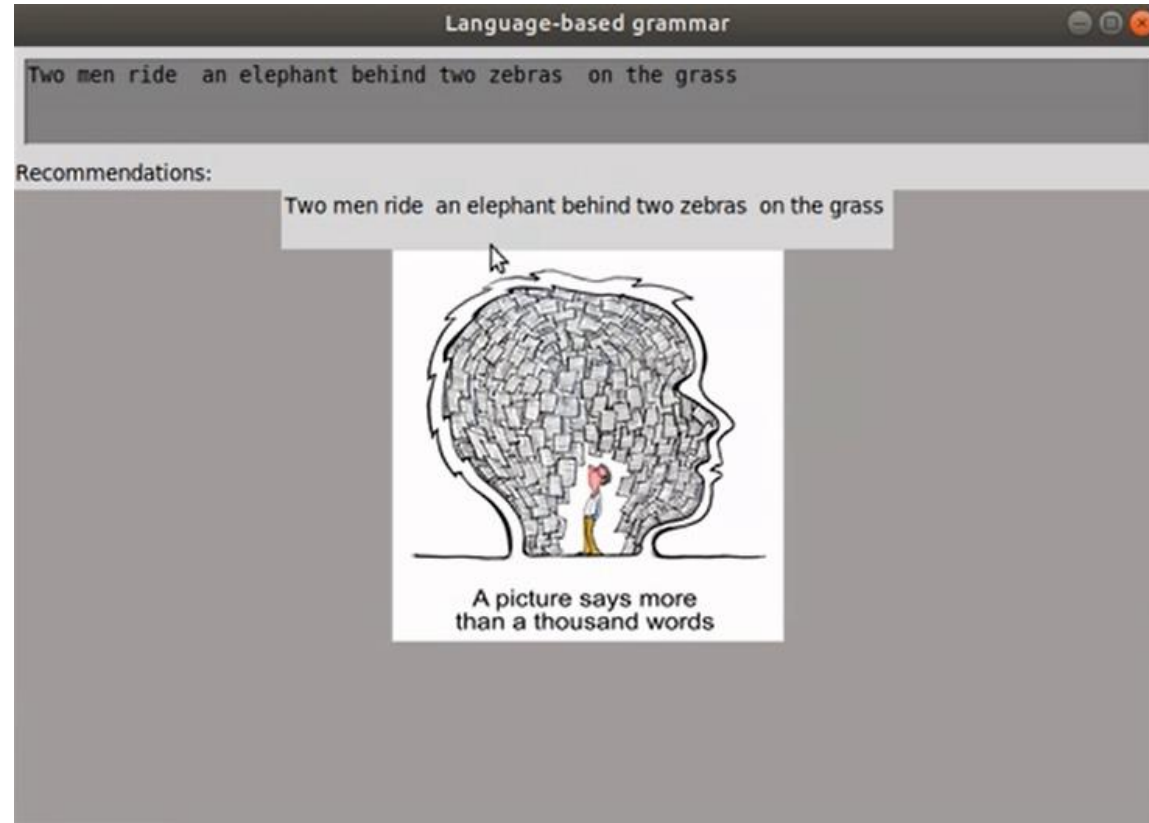
Demo of our Proposed Method

Step 1: User-friendly interface:



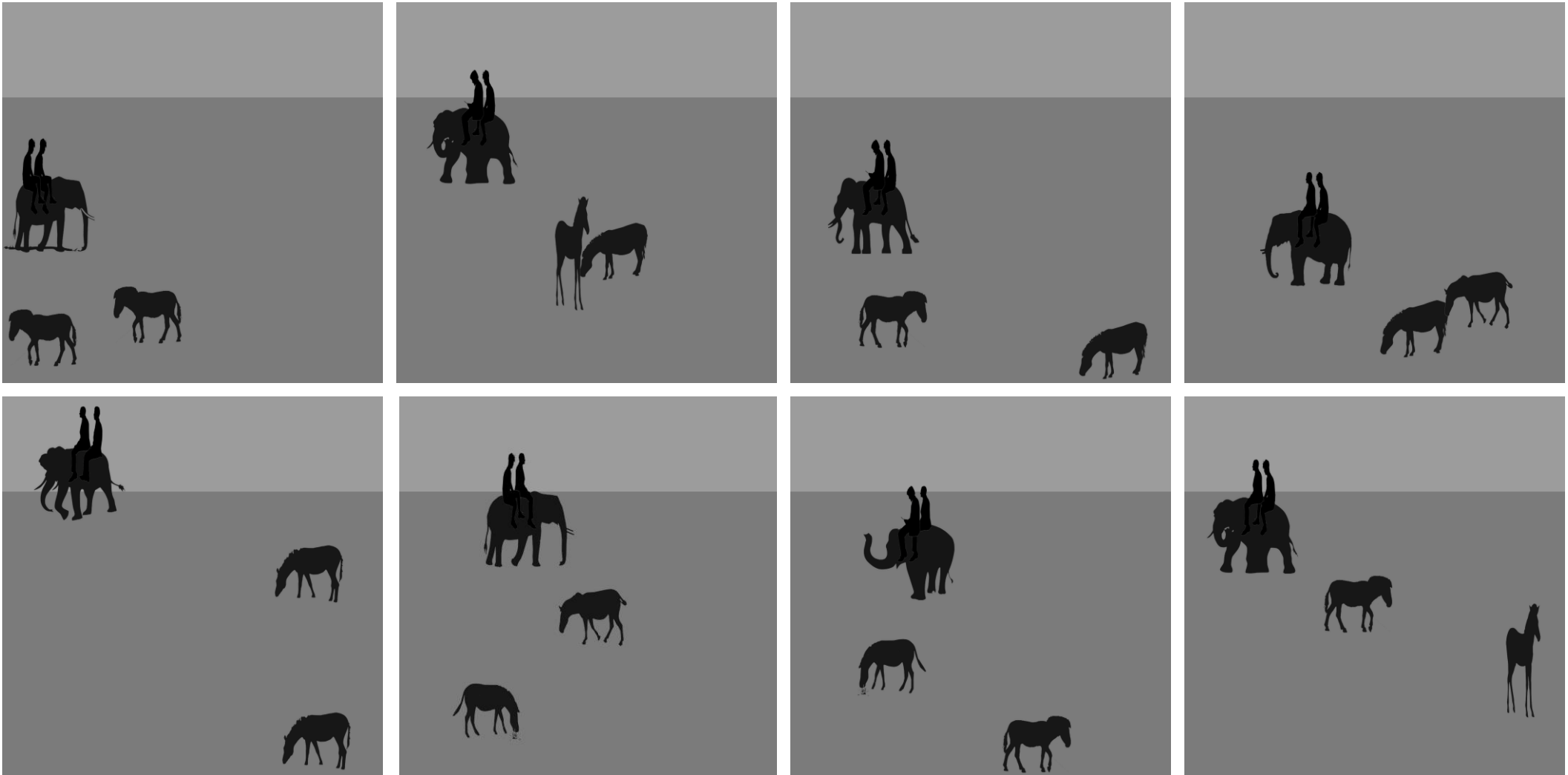
Demo of our Proposed Method

Step 2: Parse the sentence.



Demo of our Proposed Method

Step 3: Generate label maps.



Demo of our Proposed Method

Step 4: Image synthesis.



Demo of our Proposed Method

Step 5: Final result.





Conclusion and Future Work

Conclusion

- Propose a new simple yet effective framework.
 - Overcome previous issues.
 - Preserve the semantic information (*spatial relations*).
 - Our proposed methods based on:
 - Mask anchor points.
 - Aesthetic assessment.
-

Future Work

- Adopt our proposed method to different mask-to-image approaches.
 - Enhance our results further by:
 - Employing additional techniques.
 - Utilizing more aesthetic rules to generate more realistic images.
-



Thank you
