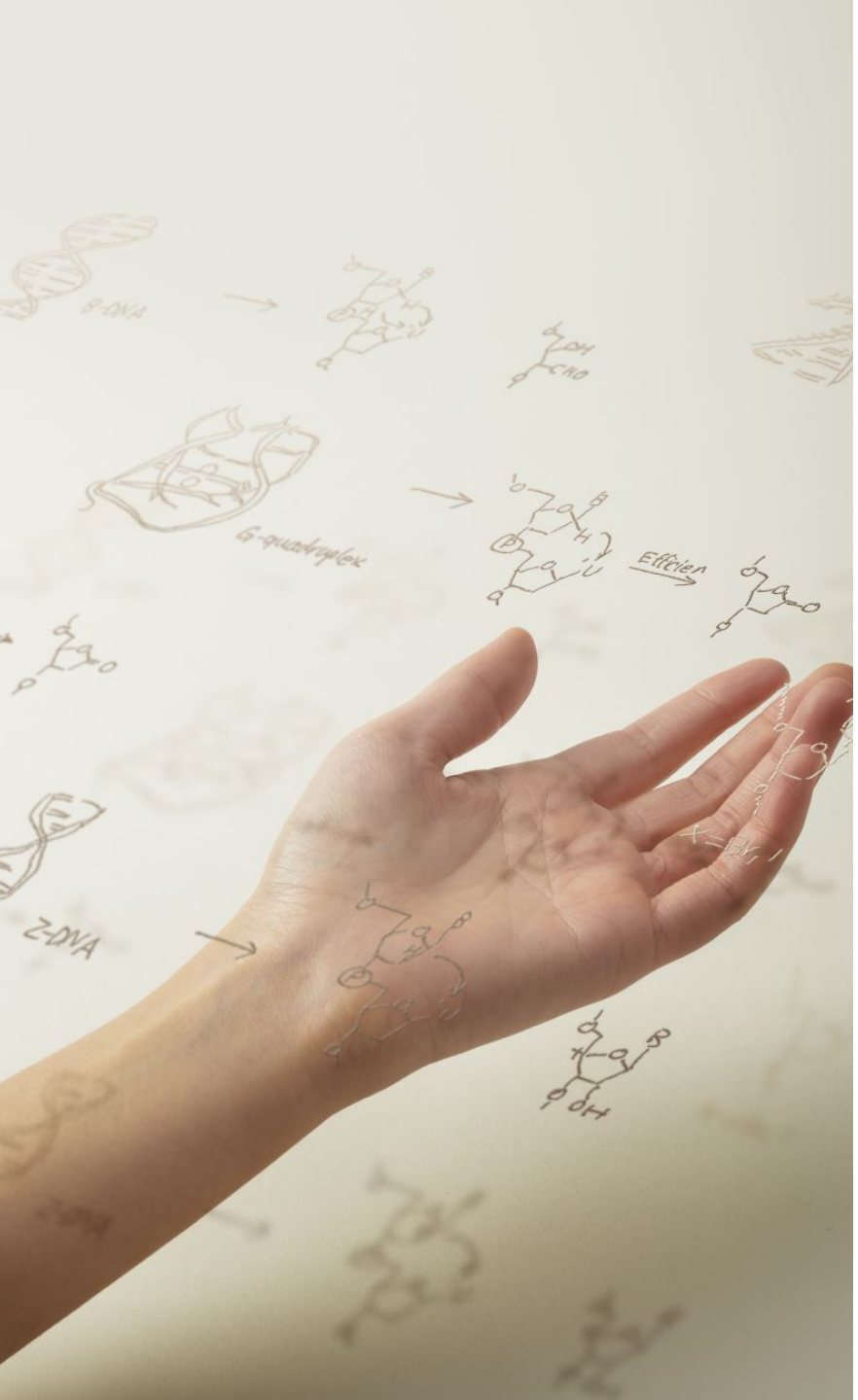


CLIPDraw: Exploring Text-to-Drawing Synthesis through Language-Image Encoders

Presentation by: Darshan Uttam Mistry (030701513)

(https://youtu.be/GihDbbOYrbA)

Guided by: Dr. Moon

Outline

- Introduction
- Objective
- Related Work
- Methodology
- Results
- Limitations and Future Directions
- Conclusion

Introduction:

- CLIPDraw is an algorithm that synthesizes novel drawings based on natural language input, using a pre-trained CLIP model as a metric for maximizing similarity between the given description and a generated drawing[1]. It optimizes vector strokes rather than pixel images, biasing drawings towards simple human-recognizable shapes.

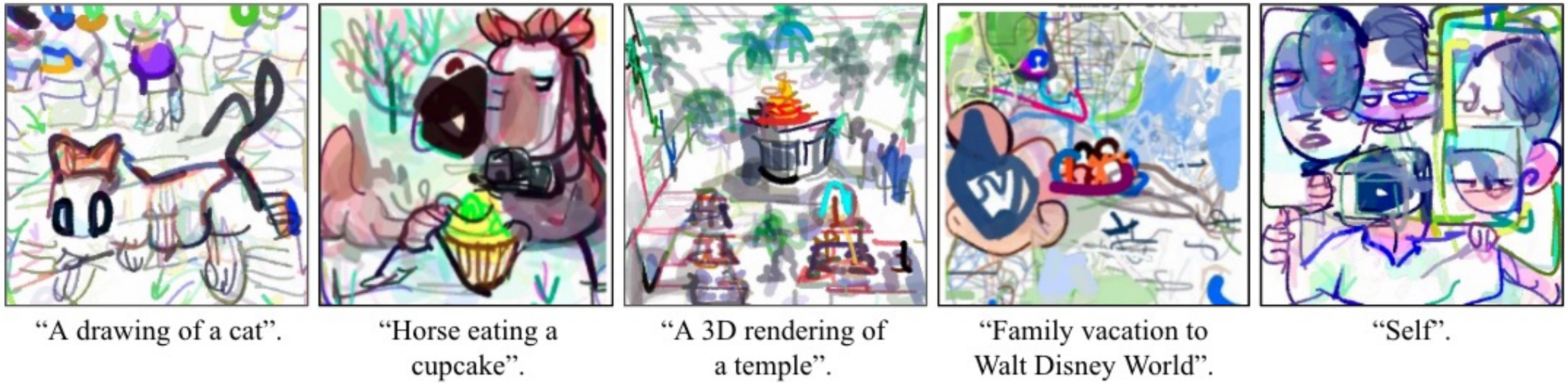


Fig 1. Various drawings synthesized by CLIPDraw[1]

Objective:

- To explore the problem of generating drawings from text descriptions. A system called "CLIPDraw" which utilizes a combination of language and image encoders to generate realistic images given textual descriptions.

Related Work:

- Text-to-Image Synthesis.
- Synthesis Through Optimization.
- Vector Graphics.
- Follow Up Work to CLIPDraw.

Methodology:

- The objective of CLIPDraw is to synthesize a drawing that matches a given description prompt, using a pre-trained CLIP model as a judge.
- Drawings are represented by a set of differentiable RGBA Bézier curves, each parametrized by 3-5 control points, thickness and an RGBA color vector.
- The algorithm works by running evaluation-time gradient descent, and the goal of the image augmentation is to force drawings to remain recognizable when viewed through various distortions.

Algorithm

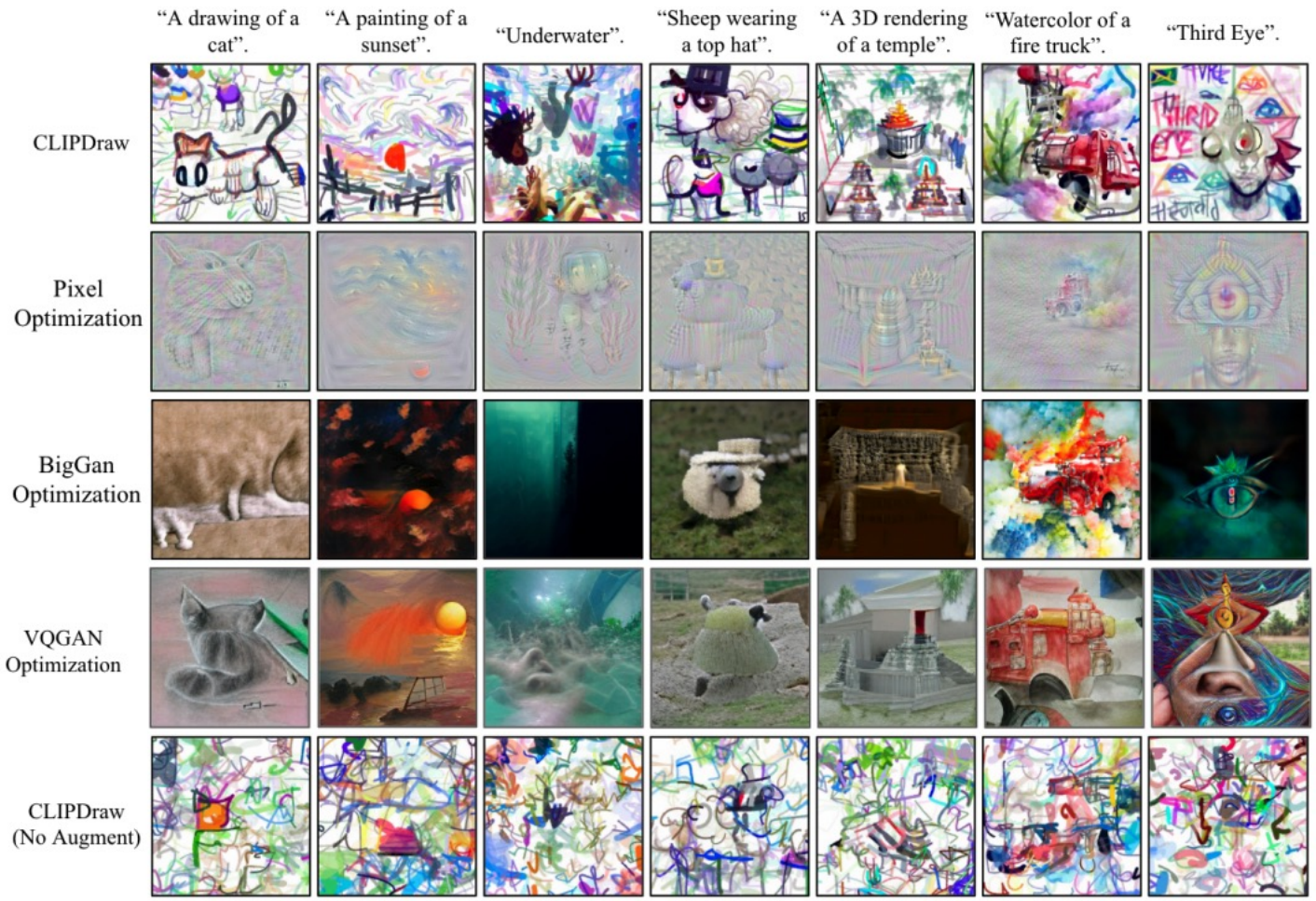
Algorithm 1 CLIPDraw

Input: Description Phrase $desc$; Iteration Count I; Curve Count N; Augment Size D; Pre-trained CLIP model.
Begin:
Encode Description Phrase. $EncPhr = CLIP(desc)$
Initialize Curves. $Curves_{..N} = RandomCurve()$
for $i = 0$ **to** I **do**
  Render Curves to Pixels. $Pixels = DiffRender(Curves)$
  Augment the Image. $AugBatch_{..D} = Augment(Pixels)$
  Encode Image. $EncImg = CLIP(AugBatch)$
  Compute Loss. $Loss = -CosineSim(EncPhr, EncImg)$
  Backprop. $Curves \leftarrow Minimize(Loss)$
end for

Fig 2. CLIPDraw Algorithm

Results:

- Fig 3. Images synthesized via various synthesis-through-optimization methods[1]



Limitations and Future Directions:

- The authors acknowledge several limitations of the CLIPDraw system, including its inability to generate highly detailed or photorealistic images, its reliance on the quality of the input textual descriptions, and its susceptibility to producing biased or stereotypical images based on the training data.
- They suggest several avenues for future research, including exploring new methods for generating high-resolution images and addressing issues of bias and fairness in image synthesis systems.

Conclusion:

- The paper describes a system called "CLIPDraw" which explores the problem of generating drawings from text descriptions. The system utilizes a combination of language and image encoders, specifically the Contrastive Language-Image Pretraining (CLIP) model and the Deep Convolutional Generative Adversarial Networks (DCGANs) model, to generate plausible images given a textual description.

Conclusion(Contd.):

- The paper presents experimental results showing the effectiveness of the CLIPDraw system in generating realistic images, as well as a user study demonstrating the usefulness of the generated images for creative tasks such as brainstorming and prototyping.

References:

[1] Frans, K., Soros, L., & Witkowski, O. (2022). Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *Advances in Neural Information Processing Systems*, *35*, 5207-5218.

Thank You