

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction  
Framework  
Experiments  
Conclusions

# Text to Image generation using GANs, CLIP and evolutionary algorithms.

Techical University of Warsaw

Maciej Domagała, Adam Komorowski

June 19, 2021

# Introduction

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

*Avocado chair* - generated by OpenAI's DALL-E model.



# GANs

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

## BigGAN

- trained on the ImageNet dataset
- proposed by DeepMind in 2018
- by design, it should generalize well

## StyleGAN2

- introduced by Nvidia in 2019
- trained in terms of generating various objects, e.g.cars, faces

# CLIP

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

## CLIP (Contrastive Language-Image Pre-Training):

- released by OpenAI in January 2021,
- it is a neural network trained on over 400 000 000 (image, caption) pairs,
- main usage is to obtain the most relevant text snippet for given image.

guacamole (90.1%) Ranked 1 out of 101 labels



- ✓ a photo of **guacamole**, a type of food.
- ✗ a photo of **ceviche**, a type of food.
- ✗ a photo of **edamame**, a type of food.
- ✗ a photo of **tuna tartare**, a type of food.
- ✗ a photo of **hummus**, a type of food.

# Evolutionary Algorithms

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

## Genetic Algorithm

- Is considered to be a basis of all evolutionary algorithms,
- It is using mutation, crossover and selection operators to generate offsprings.

## Differential Evolution

- Uses 3 vectors from population to generate offspring,
- Is regarded to be converging faster than genetic algorithm.

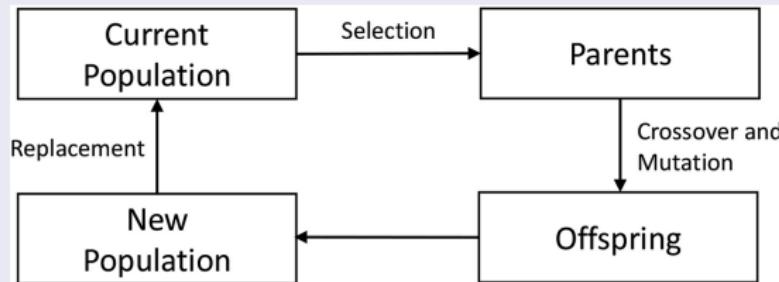
# Evolutionary Algorithms

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction  
Framework  
Experiments  
Conclusions

## Evolutionary Algorithm

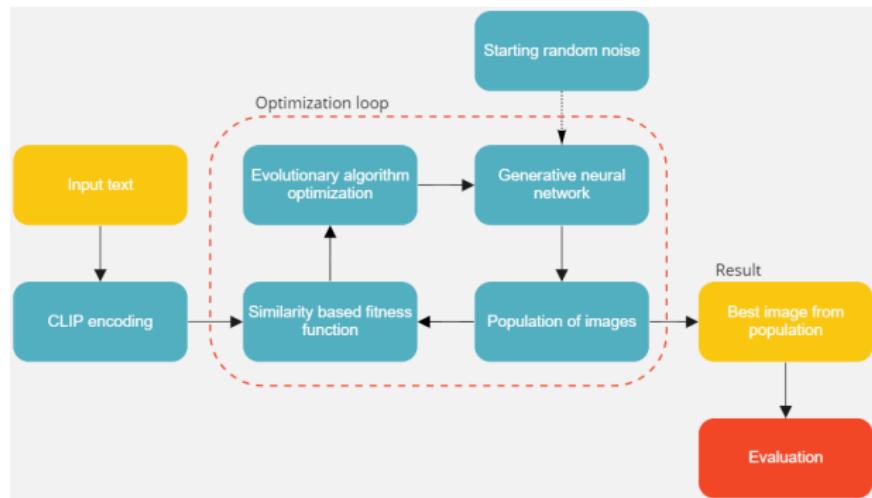


# Framework

Text to Image generation using GANs, CLIP and evolutionary algorithms.

Maciej Domagała, Adam Komorowski

Introduction  
Framework  
Experiments  
Conclusions



# Framework - example

Text to Image generation using GANs, CLIP and evolutionary algorithms.

Maciej Domagała, Adam Komorowski

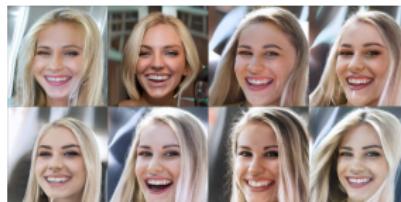
Introduction

Framework

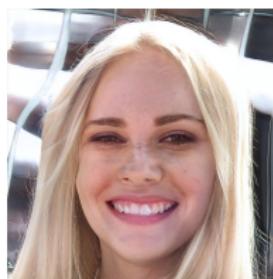
Experiments

Conclusions

- 1 Configuration: StyleGAN2-ffhq, GA algorithm, CLIP, 200 iterations
- 2 Input text: **a blond girl with a smile**
- 3 Batch after 100 iterations:

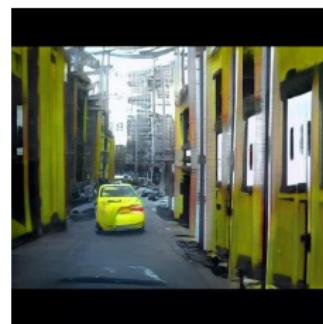


- 4 Final image:



# GA vs DE

**"A yellow car in the city"**



(a) genetic algorithm

(b) differential evolution

Figure: Final images (with best score) produced by both algorithms.

# BigGAN vs StyleGAN

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

"Red gothic church"



(a) StyleGAN2-church



(b) BigGAN

# BigGAN vs StyleGAN

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

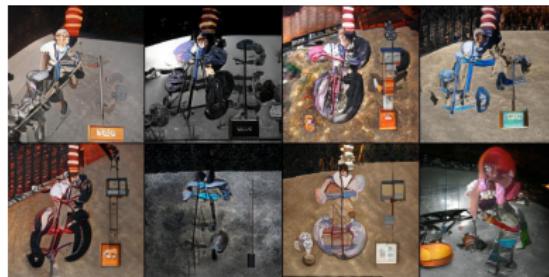
Experiments

Conclusions

"A clown cyclist on a moon"



(a) StyleGAN2-car



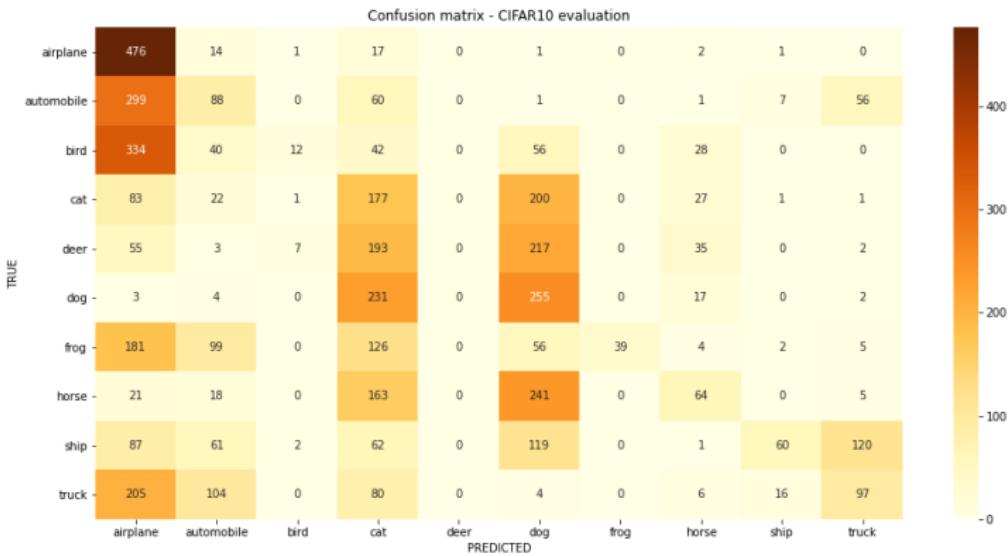
(b) BigGAN

# Evaluation - CIFAR10

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction  
Framework  
Experiments  
Conclusions



# Evaluation - ImageNet

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

Class	Positive	Negative	Accuracy (%)
BANANA	111	401	21.68
CASH MACHINE	124	388	24.22
HAMMER	141	371	27.54
ICE CREAM	3	509	0.59
LLAMA	36	476	7.03
MINISKIRT	220	292	42.97
PIRATE	5	507	0.98
SHOPPING CART	125	387	24.41
WALL CLOCK	146	366	28.52
KERRY BLUE TERRIER	253	259	49.41
<b>TOTAL</b>	<b>1164</b>	<b>3956</b>	<b>22.73</b>

# Evaluation - ImageNet - Example

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

## Llama

Correct



Incorrect



# Conclusions

Text to Image  
generation  
using GANs,  
CLIP  
and  
evolutionary  
algorithms.

Maciej  
Domagała,  
Adam  
Komorowski

Introduction

Framework

Experiments

Conclusions

- Natural continuation of this work would be to test more algorithms and more generative models.
- We would also like to perform a more comprehensive analysis of phrase semantics, how the meaning of more complicated phrase is interpreted by the model and translated to latent vector.
- Problem including new entities creation is still in the early stage of development, regarding both image and text generation, which makes it a perfect occasion to learn and develop.