

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

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

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

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



GANs - BigGAN

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

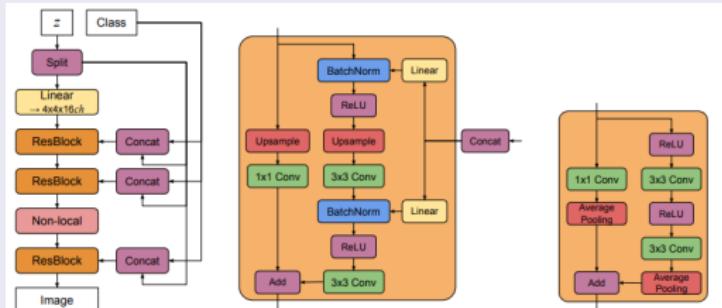
Bibliography

BigGAN

$z \in \mathbb{R}^{140} \sim \mathcal{N}(0, I)$	RGB image $x \in \mathbb{R}^{256 \times 256 \times 3}$
Embedd(y) $\in \mathbb{R}^{128}$	
Linear ($20 + 128$) $\rightarrow 4 \times 4 \times 16ch$	
ResBlock up $16ch \rightarrow 16ch$	ResBlock down $ch \rightarrow 2ch$
ResBlock up $16ch \rightarrow 8ch$	ResBlock down $2rh \rightarrow 4rh$
ResBlock up $8ch \rightarrow 8ch$	Non-Local Block (64×64)
ResBlock up $8ch \rightarrow 4ch$	ResBlock down $4ch \rightarrow 8ch$
ResBlock up $4ch \rightarrow 2ch$	ResBlock down $8ch \rightarrow 8ch$
Non-Local Block (128×128)	ResBlock down $8ch \rightarrow 16ch$
ResBlock up $2ch \rightarrow ch$	ResBlock down $16ch \rightarrow 16ch$
BN, ReLU, 3×3 Conv $ch \rightarrow 3$	ReLU, Global sum pooling
Tanh	Embed(y) $\cdot h + (\text{linear} \rightarrow 1)$

(a) Generator

(b) Discriminator



GANs - StyleGAN

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

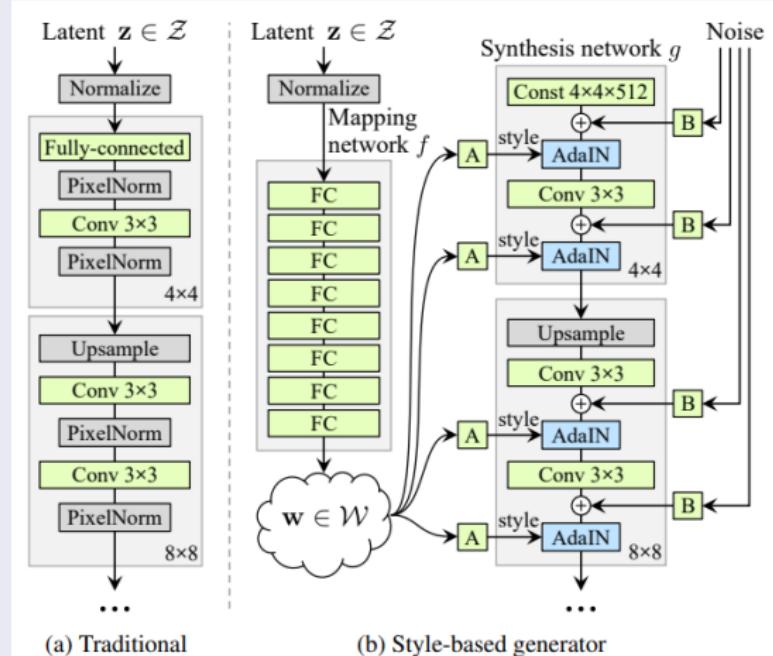
Examples

Experiments

Conclusions

Bibliography

StyleGAN



CLIP

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

CLIP - Contrastive Language-Image Pre-Training new multi-modal model released by OpenAI in January 2021.
Key features:

- 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.

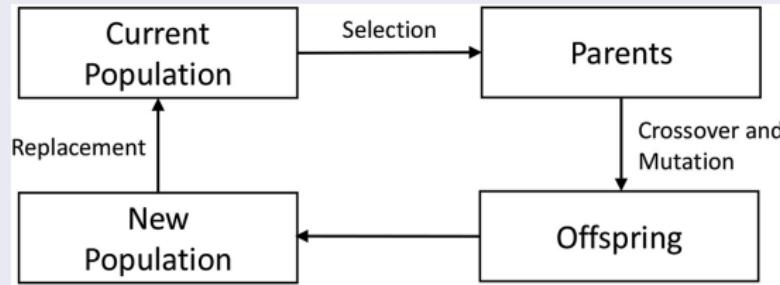
Evolutionary Algorithms

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

Maciej
Domagała,
Adam
Komorowski

Introduction
Theory
Framework
Examples
Experiments
Conclusions
Bibliography

Genetic Algorithm



Differential Evolution

Algorithm 2 Differential Evolution

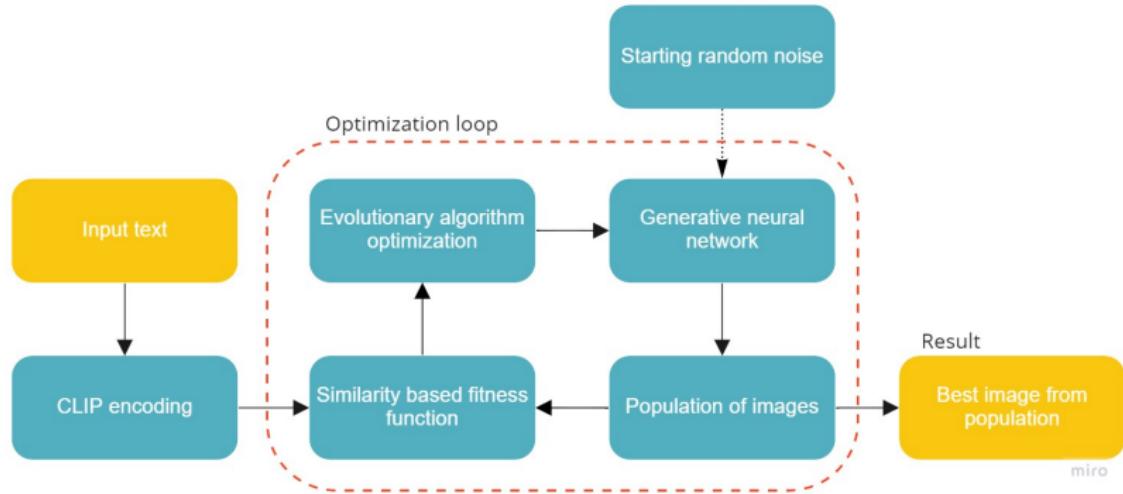
```
1: determine objective function (OF)
2: assign number of generation to 0 (t=0)
3: randomly create individuals in initial population P(t)
4: while termination criterion is not satisfied do
5:   t=t+1
6:   for each i-th individual in the population P(t) do
7:     randomly generate three integer numbers:
8:      $r_1, r_2, r_3 \in [1; \text{population size}]$ , where  $r_1 \neq r_2 \neq r_3 \neq i$ 
9:     for each j-th gene in i-th individual ( $j \in [1; n]$ ) do
10:     $v_{i,j} = x_{r_1,j} + F \cdot (x_{r_2,j} - x_{r_3,j})$ 
11:    randomly generate one real number  $rand_j \in [0; 1)$ 
12:    if  $rand_j < CR$  then  $u_{i,j} := v_{i,j}$ 
13:    else
14:       $u_{i,j} := x_{i,j}$ 
15:    end if
16:    end for
17:    if individual  $u_i$  is better than individual  $x_i$  then
18:      replace individual  $x_i$  by child  $u_i$  individual
19:    end if
20:  end for
21: end while
22: return the best individual in population P(t)
```

Framework

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

Maciej Domagała, Adam Komorowski

Introduction
Theory
Framework
Examples
Experiments
Conclusions
Bibliography



Examples

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

Maciej Domagała, Adam Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography



Figure: GA, StyleGAN2-ffhq, "a blond girl with a smile"



Figure: CLIP, GA, StyleGAN2-car, "a blue car near the water in sun"

Evaluation - CIFAR10

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

Maciej
Domagała,
Adam
Komorowski

Introduction
Theory
Framework
Examples
Experiments
Conclusions
Bibliography

Class	Positive	Negative	Accuracy (%)
AIRPLANE	476	36	92.97
AUTOMOBILE	88	424	17.19
BIRD	12	500	2.34
CAT	177	335	34.57
DEER	0	512	0.0
DOG	255	257	49.8
FROG	39	473	7.62
HORSE	64	448	12.5
SHIP	60	452	11.72
TRUCK	97	415	18.95
TOTAL	1268	3852	24.77

Evaluation - CIFAR10

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

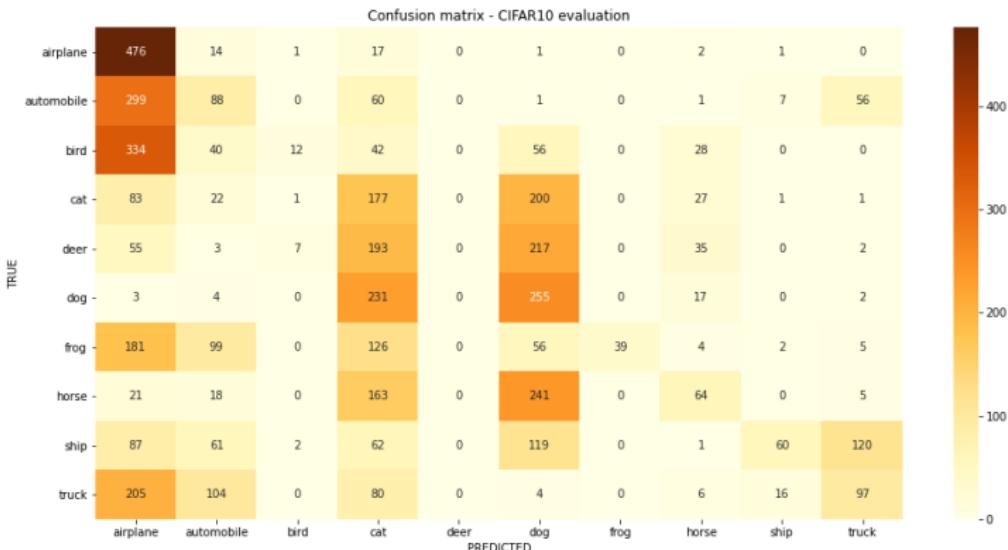
Framework

Examples

Experiments

Conclusions

Bibliography



Evaluation - CIFAR10 - Example

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

Airplane

Positive



Negative



Evaluation - ImageNet

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

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
TOTAL	1164	3956	22.73

Evaluation - ImageNet - Example

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

Llama

Correct



Incorrect



GA vs DE

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

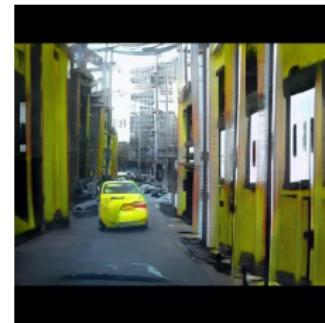
Experiments

Conclusions

Bibliography



(a) genetic algorithm



(b) differential
evolution

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

BigGAN vs StyleGAN

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography



(a) StyleGAN2-church



(b) BigGAN

BigGAN vs StyleGAN

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography



(a) StyleGAN2-ffhq



(b) BigGAN

BigGAN vs StyleGAN

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

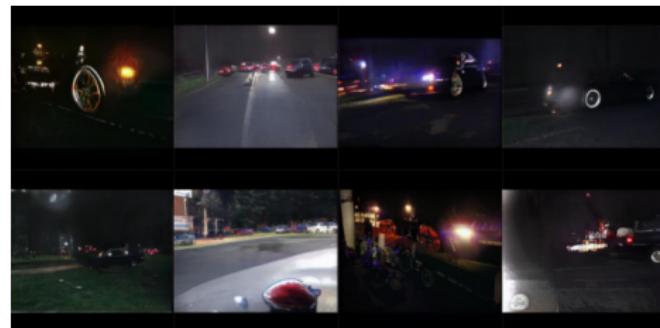
Experiments

Conclusions

Bibliography



(a) StyleGAN2-cat



(b) StyleGAN2-car

Cosine-Similarity

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

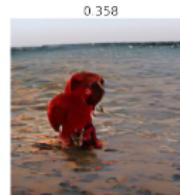
Conclusions

Bibliography

Observations:

- *cos-sim* in our solution takes values only in the range $\sim [0.2, 0.4]$

"a big red dog near the sea"



Conclusions

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

Maciej
Domagała,
Adam
Komorowski



Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography

Bibliography

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

Maciej
Domagała,
Adam
Komorowski

Introduction

Theory

Framework

Examples

Experiments

Conclusions

Bibliography



T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. *Improved techniques for training gans*. In NIPS, 2016.



C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. *Rethinking the inception architecture for computer vision*. arXiv preprint arXiv: 1512.00567, 2015.



S. Barratt, R. Sharma. *A note on the Inception Score*. arXiv preprint arXiv: 1801.01973, 2018.



M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter. *GANs trained by a two time-scale update rule converge to a local Nash equilibrium*. In NIPS, 2017.



R. Storn, K. Price. *Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces*. Springer, 1997.



T. Karras, S. Laine, T. Aila. *A style-based generator architecture for generative adversarial networks*. In CVPR, 2019.



I. Goodfellow et al. *Generative Adversarial Nets* 2014. <https://arxiv.org/pdf/1406.2661.pdf>



M. Brundage, S. Avin, J. Clark et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. arXiv preprint arXiv: 1802.07228, 2018.



OpenAI CLIP <https://arxiv.org/pdf/2103.00020.pdf>



<https://www.cs.toronto.edu/~kriz/cifar.html>



<https://www.image-net.org/>



A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark et al. *Learning transferable visual models from natural language supervision*. In CVPR, 2019.



<https://openai.com/blog/clip/>



<https://habr.com/en/post/537334/>



https://github.com/ewmashash/CTEAP_10_Classification