

ავტორი: იმედა შერიფაძე

გამოსახულების გენერაციის დიფუზიური მოდელების
შესავალი - ყოვლისმომცველი გზამკვლევი

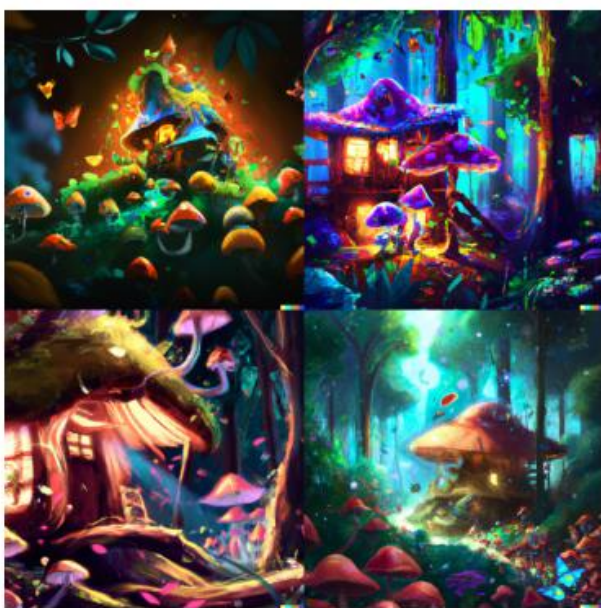
ბათუმი - 2023

გამოსახულების გენერაციის დიფუზიური მოდელების შესავალი - ყოვლისმომცველი გზამკვლევი

AI-ზე დაფუძნებულ გამოსახულების გენერაციაში ბოლოდროინდელმა მიღწევებმა, რომელსაც სათავეში უდგას დიფუზიური მოდელები, როგორიცაა **Glide, Dalle-2, Imagen** და **Stable Diffusion**, ქარიშხალი წაართვა **"AI Art თაობის"** სამყაროს. ტექსტის აღწერილობიდან მაღალი ხარისხის სურათების გენერირება რთული ამოცანაა. ის მოითხოვს ტექსტის ძირითადი მნიშვნელობის ღრმა გაგებას და ამ მნიშვნელობის შესაბამისი გამოსახულების გენერირების უნარს. ბოლო წლებში **დიფუზიური მოდელები** გამოჩნდა, როგორც ძლიერი ინსტრუმენტი ამ პრობლემის მოსაგვარებლად.

ეს მოდელები წარმოუდგენლად აადვილებს მაღალი ხარისხის სურათების გენერირებას სხვადასხვა სტილში მხოლოდ რამდენიმე სიტყვის გამოყენებით. მაგალითად, ჩვენ შევქმენით ქვემოთ მოცემული სურათი უბრალოდ მოწოდების მიწოდებით:

"სახლი ტყეში, ბნელი ღამე, ფოთლები ჰაერში, ფლუორესცენტური სოკო, ნათელი ფოკუსი, ძალიან თანმიმდევრული, ძალიან დეტალური, კონტრასტი, ცოცხალი, ციფრული ფერწერა"



ამჟამად, სხვადასხვა დიფუზიური მოდელების დიდი რაოდენობის გამო, ამ სივრცის გავლა საბოლოო მომხმარებლის პერსპექტივიდან შეიძლება იყოს რთული და ზოგჯერ, აბსოლუტური.

ამ წიგნში ჩვენ განვიხილავთ შემდეგ თემებს მეგობრულად და მარტივად, რაც ადვილად ასათვისებელია ყველასთვის, ვინც ახალია დიფუზიური მოდელების ამალელებელ სამყაროში გამოსახულების გენერირებისთვის.

- ჩვენ მოკლედ განვიხილავთ ღრმა სწავლაზე დაფუძნებულ გამოსახულების გენერაციული მოდელების სივრცეს და ჩართული სხვადასხვა ტექნიკის აღმავლობასა და დაცემას.
- ჩვენ მარტივი სიტყვებით აგიხსნით რა არის „დიფუზია“ და როგორ მუშაობს დიფუზიის მოდელები.
- ჩვენ შემოგთავაზებთ მაღალი დონის ხედვას ოთხი ყველაზე პოპულარული დიფუზიური მოდელის შესახებ:
 1. OpenAI-ის Dall-E 2
 2. Google-ის Imagen
 3. StabilityAI-ის სტაბილური დიფუზია
 4. შუა მოგზაურობა
- და ბოლოს, ჩვენ განვიხილავთ ზოგიერთ აპლიკაციას და ვებსაიტს, რომლებიც უზრუნველყოფენ დიფუზიურ მოდელებთან დაკავშირებულ სერვისებს ან იყენებენ დიფუზიის მოდელებს, როგორც სერვისს.

საუკეთესო ნაწილი ისაა, რომ ჩვენ ასევე მოგაწვდით **Jupyter**-ის ნოუტბუქს, რომელიც შეგიძლიათ გამოიყენოთ **სტაბილური დიფუზიის გამოსახულების გენერირების მოთხოვნის გასაშვებად 6 GB VRAM-** ით ან თუნდაც **CPU-ზე სტაბილური დიფუზიის გასაშვებად.**

1 - რა არის გენერაციული მოდელები?

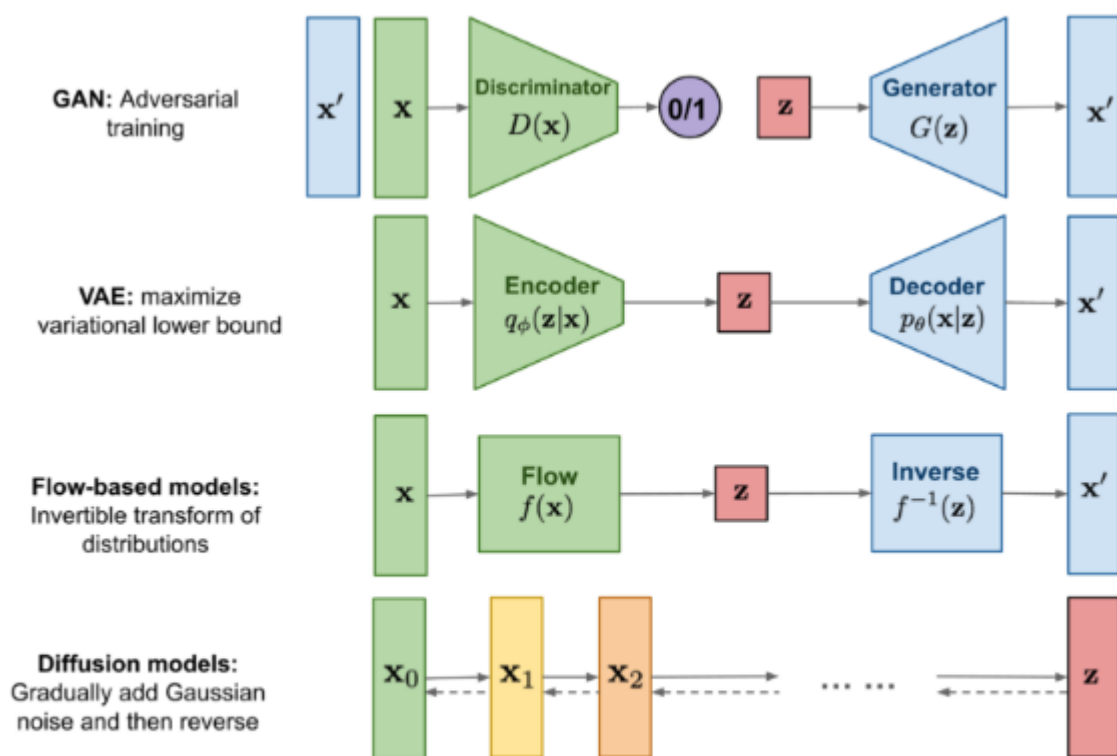
მანქანური სწავლისა და ღრმა სწავლის პრობლემების უმეტესობა, რომელსაც თქვენ წყვეტთ, კონცეპტუალიზებულია გენერაციული და დისკრიმინაციული მოდელებიდან . მარტივად რომ ვთქვათ, „გენერაციული მოდელები“ არის სტატისტიკური მოდელები, რომლებიც შექმნილია „**მონაცემების გენერირებისთვის/სინთეზისთვის**“. მათი ამოცანაა "**ხმაურის გადაქცევა მონაცემთა წარმომადგენლობით ნიმუშად**".

წლების განმავლობაში ჩვენ ვნახეთ გენერაციული მოდელების მრავალი კრეატიული აპლიკაცია. ერთ-ერთი კონკრეტული აპლიკაცია, რომელიც ბევრ ჩვენგანს ახსოვს, იყო **Cadbury**-ის რეკლამა, რომელიც იყენებდა აუდიო გენერაციას და ტუჩის სინქრონიზაციას სხვადასხვა ცნობილი ადამიანების სახის გამომეტყველებისა და

მეტყველების გამოსათვლელად. ის შეიძლება გამოყენებულ იქნას ცნობილი ადამიანების პერსონალიზებული რეკლამის შესაქმნელად.

დრმა სწავლაზე დაფუძნებული გამოსახულების გენერირების ოთხი ცნობილი მოდელია:

1. ვარიაციური ავტოკოდერები (VAE)
2. ნაკადზე დაფუძნებული მოდელები
3. გენერაციული საპირისპირო ქსელები.
4. დიფუზია (უკანასკნელი ტენდენცია)



გამოსახულება მაღალ დონეზე აღბეჭდავს ოთხივე ალგორითმის მექანიზმებს.

ეს მოდელები ჯერ ივარჯიშება ისე, რომ მათ ისწავლონ „მონაცემთა განაწილების“ (სავარჯიშო მონაცემების) მოდელირება. მომზადების შემდეგ, მოდელმა იცის როგორ დაახლოოს ორიგინალური მონაცემთა განაწილება და შეუძლია გამოიყენოს იგი ახალი მონაცემების (სურათების) გენერირებისთვის სურვილისამებრ.

"ვარიაციური ავტოკოდერების" გაგების წინამორბედია "ავტოენკოდერები". Autoencoders-ის ძირითადი მიზანია მონაცემთა შეკუმშვა. Autoencoders-ის არქიტექტურა საკმაოდ მარტივია. იგი შეიცავს სამ კომპონენტს:

1. Encoder
2. Bottleneck (პასუხისმგებელია შეკუმშვაზე)
3. დეკოდერი

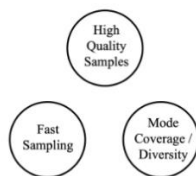
ამ დიზაინის დამატებითი უპირატესობა ის არის, რომ ჩვენ შეგვიძლია გამოვიყენოთ ავტოენკოდერები გამოსახულების **დენოიზისთვის**.

ავტოენკოდერებში, შეკუმშული **მონაცემების/ლატენტების** განაწილება „**შეუზღუდავია**“. მონაცემები შეკუმშულია ისე, რომ მინიმალური რეკონსტრუქციის შეცდომაა. ეს იწვევს დიდ ნაკლს, ანუ იმიტომ, რომ ჩვენ არ გვაქვს წარმოდგენა/ინფორმაცია ლატენტის განაწილების შესახებ, ამიტომ ახალი ნიმუშების გენერირება (მხოლოდ დეკოდერის გამოყენებით) რთულია.

Variational Autoencoders- ში ეს პრობლემა აღარ არის. შეზღუდვა ემატება **bottleneck** ფენას, ამიტომ ენკოდერის შეკუმშულ მონაცემებს უნდა მიზამოს (მარტივი ხელით შერჩეული) ალბათობის განაწილება (ზოგადად სტანდარტული გაუსიანი) რაც შეიძლება ახლოს. ახალი ნიმუშების გენერირებისთვის, ჩვენ შეგვიძლია უბრალოდ შევამოწმოთ წერტილი არჩეული ალბათობის განაწილებიდან და გადავიტანოთ დეკოდერს.

წერის მომენტში, ჩვენ ჯერ კიდევ არ გვაქვს გადაწყვეტილი ერთი მოდელზე, რომელიც დაკავშირებულია გენერაციულ მოდელირებასთან დაკავშირებული ყველა ამოცანისთვის და ეს სამართლიანია. ყველა აპლიკაციის დომენს აქვს საკუთარი გამოწვევები და ხალხი ზოგადად იყენებს სხვადასხვა ტექნიკას მათ გადასაჭრელად. ოთხივე ზემოთ ჩამოთვლილ მოდელს აქვს პრობლემები, რომლებიც სწორად არის ილუსტრირებული ქვემოთ მოცემულ გიფში.

The Generative Learning Trilemma



გენერაციული სწავლის ტრილემის
დაძლევა

	VAE	ნაკადი	GAN	დიფუზია
პლიუსები	შერჩევის სწრაფი სიჩქარე. მრავალფეროვანი ნიმუშის თაობა	შერჩევის სწრაფი სიჩქარე. მრავალფეროვანი ნიმუშის თაობა	შერჩევის სწრაფი სიჩქარე. ნიმუშის წარმოების მაღალი ხარისხი.	ნიმუშის წარმოების მაღალი ხარისხი. მრავალფეროვანი ნიმუშის თაობა
მინუსები	ნიმუშის წარმოების დაბალი ხარისხი	საჭიროა სპეციალიზებული არქიტექტურა, დაბალი ნიმუშის წარმოების ხარისხი	არასტაბილური ვარჯიში, ნიმუშების წარმოქმნის დაბალი მრავალფეროვნება (რეჟიმების კოლაფსი)	შერჩევის დაბალი მაჩვენებელი

14 წელს შექმნილი იან გუდფელოუს მიერ, გენერაციული საპირისპირო ქსელები (GANs) იყო თითქმის ნორმა გამოსახულების ნიმუშების გენერირებისთვის.

შეიქმნა ორიგინალური GAN-ის მრავალი ვარიაცია, როგორიცაა:

1. პირობითი GAN (cGAN) : გენერირებული სურათების კლასის/კატეგორიის კონტროლი.
2. Deep Convolutional GAN (DCGAN) : არქიტექტურა მნიშვნელოვნად აუმჯობესებს GAN-ების ხარისხს კონვოლუციური ფენების გამოყენებით.
3. სურათიდან გამოსახულებაზე თარგმნა Pix2Pix-ით : სურათების კონვერტაცია ერთი დომენიდან მეორეზე შეყვანისა და გამომავალი შეყვანის შესწავლით.

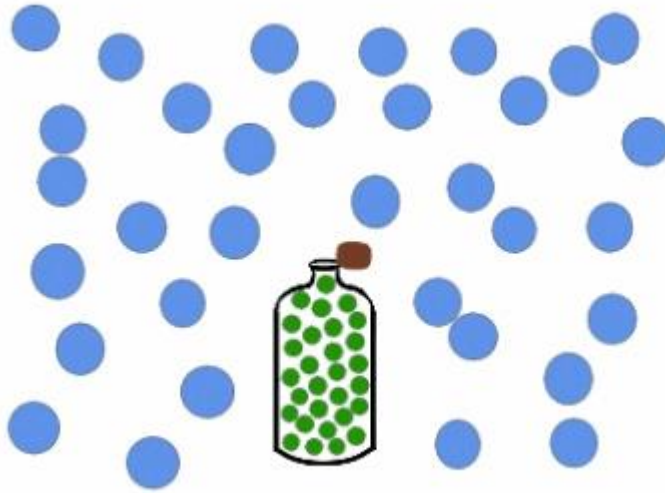
ახლა დიფუზიური მოდელების ეპოქაში, მკვლევარები დარწმუნებულნი არიან, რომ გამოიყენებენ GAN-ზე მუშაობის შედეგად მიღებულ ცოდნას. ეს არის ერთ-ერთი მთავარი მიზეზი ასეთი სწრაფი პროგრესის დიფუზიურ მოდელებში ასეთ მოკლე დროში.

2 - რა არის დიფუზია?

სანამ გავიგებთ დიფუზიის მოდელებს, მოდით სწრაფად გავიგოთ ტერმინი "დიფუზიის" მნიშვნელობა . დიფუზია (ან დიფუზიის პროცესი) არის კარგად ცნობილი და შესწავლილი დომენი არაწონასწორობის სტატისტიკურ ფიზიკაში.

არა წონასწორობის სტატისტიკურ ფიზიკაში დიფუზიის პროცესი ეხება:

„ნაწილაკების ან მოლეკულების მოძრაობა მაღალი კონცენტრაციის ზონიდან დაბალი კონცენტრაციის ზონაში, კონცენტრაციის გრადიენტით გამოწვეული“.



დიფუზიის პროცესი განპირობებულია ნაწილაკების ან მოლეკულების შემთხვევითი მოძრაობით, რაც აღწერილია თერმოდინამიკისა და სტატისტიკური მექანიკის კანონებით.

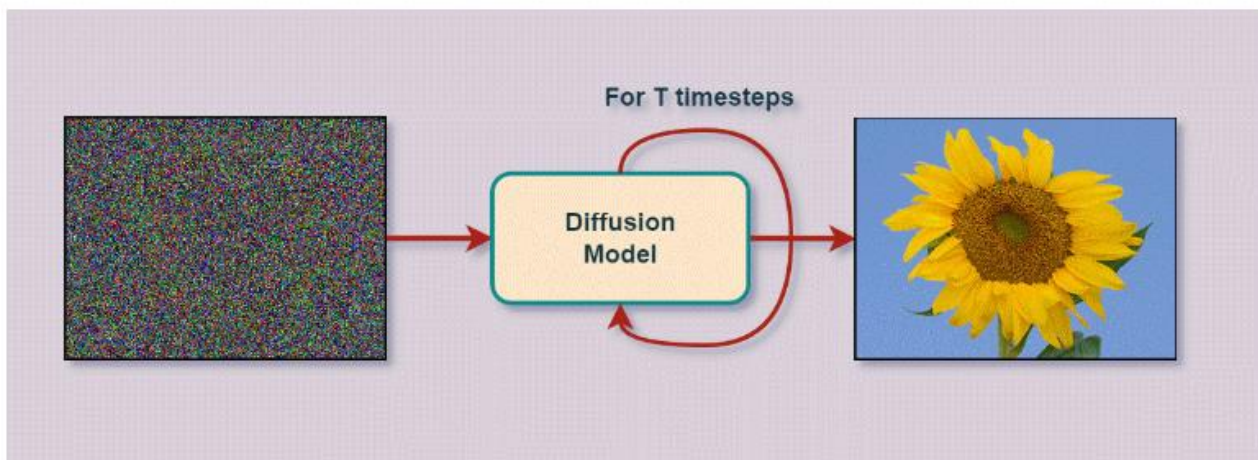
3 - რა არის დიფუზიური მოდელები?

მარტივი სიტყვებით - *"დიფუზიური მოდელები არის სავარაუდო გენერაციული მოდელების კლასი, რომელიც ხმაურს აქცევს მონაცემთა წარმომადგენლობით ნიმუშად."*

დიფუზიის მოდელების გამოყენებით, ჩვენ შეგვიძლია გამოსახულებების გენერირება პირობითად ან უპირობოდ.

1. გამოსახულების უპირობო გენერაცია უბრალოდ ნიშნავს, რომ მოდელი ხმაურს გარდაქმნის „შემთხვევითი წარმომადგენლობითი მონაცემების ნიმუშად“. გენერირების პროცესი არ არის კონტროლირებადი ან მართვადი და მოდელს შეუძლია ნებისმიერი ბუნების გამოსახულების გენერირება.
2. პირობითი გამოსახულების გენერაცია არის ის ადგილი, სადაც მოდელს მიეწოდება დამატებითი ინფორმაცია ტექსტის (text2img) ან კლასის ეტიკეტების საშუალებით (როგორც CGAN-ებში). ეს არის მართული ან კონტროლირებადი გამოსახულების გენერირების შემთხვევა. დამატებითი ინფორმაციის მიწოდებით, ჩვენ ველით, რომ მოდელი გამოიმუშავებს სურათების კონკრეტულ კომპლექტს. მაგალითად, შეგიძლიათ მიმართოთ ორ სურათს წიგნის დასაწყისში.

ამ განყოფილებაში ჩვენ ყურადღებას გავამახვილებთ „უპირობო გამოსახულების გენერირების“ პროცესზე.




გამოსახულების უპირობო გენერაცია დიფუზიური მოდელის გამოყენებით დასკვნის ეტაპზე.

ცოტა ისტორია...

ღრმა სწავლაში დიფუზიის მოდელები პირველად შემოიღეს სოლ-დიკშტეინმა და სხვებმა. 2015 წლის მთავარ ნაშრომში „ღრმა ზედამხედველობის ქვეშ მყოფი სწავლა არაწონასწორობის თერმოდინამიკის გამოყენებით“. სამწუხაროდ, ფარდების მიღმა დარჩა.

მაგრამ, 2019 წელს, სონგმა და სხვ. გამოაქვეყნა ნაშრომი „გენერაციული მოდელირება მონაცემთა განაწილების გრადიენტების შეფასებით“, იგივე პრინციპის გამოყენებით, მაგრამ განსხვავებული მიდგომა. 2020 წელს ჰო და სხვ. გამოაქვეყნა ნაშრომი, რომელიც ახლა პოპულარულია „დენოიზირების დიფუზიის ალბათური მოდელები“ (მოკლედ DDPM).

2020 წლის შემდეგ დიფუზიური მოდელების კვლევამ აიღო . დიდი პროგრესი იქნა მიღწეული შედარებით მოკლე დროში დიფუზიაზე დაფუძნებული გენერაციული მოდელირების შექმნის, ტრენინგისა და გაუმჯობესებაში.

4 - როგორ მუშაობს დიფუზიაზე დაფუძნებული გამოსახულების გენერირების მოდელები?

დიფუზიური მოდელების მუშაობის ზოგადი პრინციპი მართლაც მარტივი გასაგებია.

დიფუზიის მეთოდი შეიძლება შეჯამდეს შემდეგნაირად:

...სისტემატურად და ნელა ანადგურებს სტრუქტურას მონაცემთა განაწილებაში განმეორებითი წინსვლის დიფუზიის პროცესის მეშვეობით. შემდეგ ჩვენ ვსწავლობთ საპირისპირო დიფუზიის პროცესს, რომელიც აღადგენს მონაცემებში სტრუქტურას, რაც იძლევა მონაცემთა ძალიან მოქნილ და დამუშავებულ გენერაციულ მოდელს. ეს მიდგომა საშუალებას გვაძლევს სწრაფად ვისწავლოთ, შევამოწმოთ და შევაფასოთ ალბათობები ღრმა გენერაციულ მოდელებში...

მოდით გადავდგათ ნაბიჯი უკან და გადავხედოთ გაზის დიფუზიის გიფს ზემოთ. როდესაც ქილა იხსნება, მწვანე გაზის მოლეკულები სწრაფად გადადის ქილიდან და ირგვლივ. ეს არის არსებითად დიფუზია. რაც დრო გადის, *მწვანე გაზის მოლეკულების კონცენტრაცია ქილაში და მის გარეთ თანაბარი იქნება*. გაზის მოლეკულების განაწილება მთლიანად შეიცვალა ქილის გახსნამდე. ამ პროცესის შებრუნება ადვილი საქმე არ არის; ეს არის სადაც არაბალანსირებული სტატისტიკური ფიზიკა ჩნდება სურათზე.

იდეა, რომელიც გამოიყენება არა წონასწორობის სტატისტიკურ ფიზიკაში, არის ის, რომ ჩვენ შეგვიძლია **თანდათან გადავიტანოთ ერთი განაწილება მეორეში**. 2015 წელს Sohl-Dicktein-მა და სხვებმა, ამით შთაგონებული, შექმნეს „დიფუზიის ალბათური მოდელები“ ან მოკლედ „დიფუზიის მოდელები“, ამ არსებით იდეაზე დაყრდნობით.

ისინი აშენებენ – *„გენერაციულ მარკოვის ჯაჭვს, რომელიც გარდაქმნის მარტივ ცნობილ განაწილებას (მაგ., გაუსიანს) სამიზნე (მონაცემების) განაწილებად დიფუზიური პროცესის გამოყენებით“*.

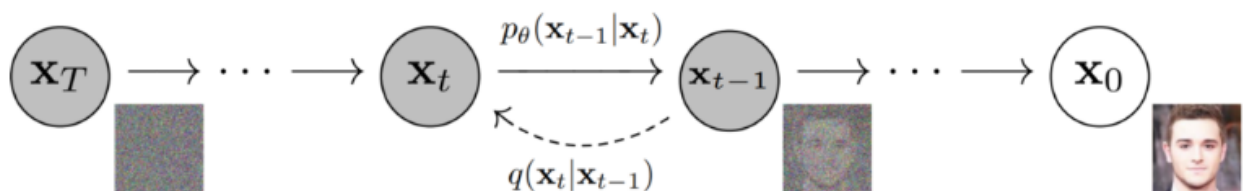
მარკოვის ჯაჭვი უბრალოდ ნიშნავს, რომ ერთეულის/ობიექტის მდგომარეობა ჯაჭვის ნებისმიერ წერტილში დამოკიდებულია მხოლოდ წინა ერთეულზე/ობიექტზე.

ახლა, ჩვენ შეგვიძლია ამის გაკეთება ორივე გზით, ანუ, ჩვენ ასევე შეგვიძლია გადავიტანოთ ჩვენი სასწავლო მონაცემების (უცნობი) განაწილება სხვა განაწილებად. ალუბლის ზემოდან დასაყენებლად, ორივე (მონაცემები ხმაურზე და ხმაურზე მონაცემებზე) შეიძლება მოდელირებული იყოს იგივე ფუნქციური ფორმის გამოყენებით. ეს არის ზუსტად ის, რაც კეთდება დიფუზიურ მოდელებში.

ავტორები აღწერენ, რომ -

”ჩვენი მიზანია განვსაზღვროთ წინსვლის (ან დასკვნის) დიფუზიის პროცესი, რომელიც გარდაქმნის მონაცემთა კომპლექსურ განაწილებას მარტივ, გადასატანად, განაწილებად და შემდეგ ვისწავლოთ ამ დიფუზიის პროცესის სასრულ დროში შებრუნება, რომელიც განსაზღვრავს ჩვენს გენერაციულ მოდელის განაწილებას.”

ორიგინალური სურათის სტრუქტურა (განაწილება) თანდათან ნადგურდება ხმაურის დამატებით და შემდეგ ნეირონული ქსელის მოდელის გამოყენებით გამოსახულების რეკონსტრუქციისთვის, ანუ ხმაურის ამოღება ყოველ ნაბიჯზე. ამის გაკეთება საკმარისად და კარგი მონაცემებით, მოდელი საბოლოოდ სწავლობს ძირითადი (ორიგინალური) მონაცემთა განაწილების შეფასებას. ამის შემდეგ, ჩვენ შეგვიძლია უბრალოდ დავიწყოთ მხოლოდ ხმაურით და გამოვიყენოთ გაწვრთნილი ნეირონული ქსელი ორიგინალური სასწავლო მონაცემთა ნაკრების ახალი გამოსახულების შესაქმნელად.



წინ და უკან/უკუ დიფუზიის პროცესის ილუსტრაცია

ის, რაც ახლა აღვწერთ, არის ორი არსებითი პროცესი/ეტაპი, რომელსაც ასრულებს ყველა დიფუზიური მოდელი. მათემატიკური დეტალების გარეშე, მოდით გადავიდეთ მათზე და გამოვიყენოთ ზემოთ მოცემული სურათი, როგორც მითითება.

1. წინა დიფუზია:

1. ორიგინალური გამოსახულება (x_0) ნელა ზიანდება განმეორებით (მარკოვის ჯაჭვი) (მასშტაბიანი გაუსიანი) ხმაურის დამატებით.
2. ეს პროცესი კეთდება გარკვეული T დროის საფეხურზე, ანუ x_T .
3. გამოსახულება t დროის საფეხურზე იქმნება: $x_{t+1} + \epsilon_{t+1}$ (ხმაური) $\rightarrow x_t$
4. ამ ეტაპზე არც ერთი მოდელი არ არის ჩართული.
5. წინა დიფუზიის x_T სტადიის ბოლოს, ხმაურის განმეორებითი დამატების გამო, ჩვენ ვრჩებით (სუფთა) ხმაურიანი გამოსახულება, რომელიც წარმოადგენს "იზოტროპული გაუსიანს". ეს მხოლოდ მათემატიკური ხერხია იმის სათქმელად, რომ ჩვენ გვაქვს სტანდარტული ნორმალური განაწილება და განაწილების ვარიაცია ყველა განზომილებაში ერთნაირია. ჩვენ გადავაცქიეთ მონაცემთა განაწილება გაუსიან განაწილებად.

2. უკანა/უკუ დიფუზია :

1. ამ ეტაპზე ჩვენ ვაუქმებთ წინსვლის პროცესს. ამოცანაა ამოიღონ ხმაური დამატებული წინსვლის პროცესში, ისევ განმეორებითი გზით (მარკოვის ჯაჭვი). ეს კეთდება ნეირონული ქსელის მოდელის გამოყენებით.
2. მოდელს ევალება შემდეგი: t - დროის საფეხურის და ხმაურიანი გამოსახულების x_t -ის გათვალისწინებით, იწინასწარმეტყველე სურათზე დამატებული ხმაური (ϵ) $t-1$ საფეხურზე.
3. $x_t \rightarrow$ მოდელი $\rightarrow \epsilon$ (პროგნოზირებადი ხმაური). მოდელი პროგნოზირებს (დაახლოებით) x_{t-1} -ზე დამატებულ ხმაურს წინ გადასასვლელში.

შედარება GAN-ებთან

- დიფუზიის პროცესის განმეორებითი ხასიათის გამო, ტრენინგის და გენერირების პროცესი ზოგადად უფრო სტაბილურია, ვიდრე GAN-ები.
- GAN-ებში გენერატორის მოდელი უნდა გადავიდეს სუფთა ხმაურიდან სურათზე ერთი ნაბიჯით $x_T \rightarrow x_0$, რაც არასტაბილური ვარჯიშის ერთ-ერთი წყაროა.

- **GAN**-ისგან განსხვავებით, სადაც ტრენინგი მოითხოვს ორ მოდელს, დიფუზიებში მხოლოდ ერთი მოდელია საჭირო.
- ერთი დაკვირვება ზემოთ მოყვანილი სურათიდან არის ის, რომ „*გამოსახულების განზომილება იგივე რჩება*“ მთელი პროცესის განმავლობაში, განსხვავებით **GAN**-ისგან, სადაც ლატენტური ტენსორი შეიძლება იყოს სხვადასხვა ზომის. ეს შეიძლება იყოს პრობლემა მაღალი ხარისხის სურათების გენერირებისას შეზღუდული **GPU** მეხსიერების გამო. თუმცა, „*სტაბილური დიფუზიის*“ (უფრო ზუსტად რომ ვთქვათ, „*ლატენტური დიფუზია*“) ავტორები ამ საკითხს გვერდს უვლიან ვარიაციური ავტოკოდერის გამოყენებით.

გასათვალისწინებელია რამდენიმე პუნქტი დიფუზიის პროცესის შესახებ:

1. მისი განმეორებითი ხასიათის კიდევ ერთი უპირატესობა ის არის, რომ ჩვენ ვასრულებთ ზედამხედველობით ტრენინგს ყოველ ჯერზე.
2. დიფუზიებში, არჩეული პოპულარული არქიტექტურა არის UNet (ყურადღებით), რომელიც გაწვრთნილია MSE დაკარგვის ფუნქციის გამოყენებით ზედამხედველობით.
3. თითოეულ ეტაპზე, სურათზე დამატებული ხმაური კონტროლდება „ვარიანტობის განრიგის“ ან უბრალოდ „დაგეგმვის“ გამოყენებით. გრაფიკის მოვალეობაა განსაზღვრონ რამდენი ხმაური უნდა დაემატოს ისე, რომ წინ გადადგმული პროცესის დროს გამოსახულება x_T ბოლოს არის იზოტროპული გაუსიანი.
4. DDPM ნაშრომში ავტორებმა გამოიყენეს „წრფივი გრაფიკი“. ეს ნიშნავს, რომ ყოველ ჯერზე დამატებული ხმაური წრფივად გაიზარდა.
5. დროითი ნაბიჯების რაოდენობა T დაყენდა $T=1000$ -ზე. ასე რომ, გაუსის ხმაურის გათვალისწინებით, მოდელს დასჭირდება 1000 გამეორება შედეგის შესაქმნელად. ეს არის წინა ნაწილში ნახსენები შერჩევის დაბალი სიჩქარის პრობლემა. მაგრამ ბოლო ნაშრომებში, ახალი ტექნიკის და სხვადასხვა განრიგის შემოღებით, მკვლევარებმა შეძლეს მხატვრული სურათების გენერირება $T=4$ დროის საფეხურზე.

5 - კარგად ცნობილი დიფუზიური მოდელები გამოსახულების გენერირებისთვის



" კიბერპანკის სამყარო, ნეონის განათება, წვიმიანი ამინდი, აკრილი "



" კაცი, რომელიც წვიმიან ღამეს დადის კამერისკენ, ანარეკლი წყლის გუბეში, წისლი, 80 მმ "

მოდით მიმოვიხილოთ დიფუზიაზე დაფუძნებული **Image Generation**-ის რამდენიმე მოდელი, რომლებიც ცნობილი გახდა ბოლო რამდენიმე თვის განმავლობაში. იმის გამო, რომ ველი ახლა დაკავებულია ასობით დიფუზიური მოდელით და მათი ვარიაციებით, ჩვენ გავაკეთებთ მცირე ნიმუშს საკუთარ თავზე და გავაშუქებთ უფრო ცნობილებს. ისენი მოიცავს:

- **Dall-E 2** OpenAI-ის მიერ
- გამოსახულება **Google**-ის მიერ
- სტაბილური დიფუზია **StabilityAI**-ის მიერ
- შუა მოგზაურობა

აქ ჩვენ არ შევალთ ზემოაღნიშნული არქიტექტურების სიღრმეში. ამის ნაცვლად, ჩვენ გავაშუქებთ თითოეული მოდელის მიმოხილვას და შევქმნით ერთგულ, ახალ წიგნს მომავალში. ჩვენ ასევე შევამოწმებთ რამდენიმე მოთხოვნას და **Dall-E 2**-ისა და **Stable Diffusion**-ის მიერ წარმოქმნილ სურათებს.

DALL-E 2

Dall-E 2 გამოქვეყნდა OpenAI-ის მიერ 2022 წლის აპრილში. ის ეფუძნება **OpenAI-ის წინა** ისტორიულ ნამუშევრებს **GLIDE-** ზე , **CLIP-** ზე და **Dall-E-ზე** . მიუხედავად იმისა, რომ **Dall-E 2** არის **Dall-E-ს** უმაღლესი მემკვიდრე, პირველი შეიცავს **3,5** მილიარდ პარამეტრს, ვიდრე **12** მილიარდი პარამეტრი **Dall-E-ში**.

ძალიან ბევრი არქიტექტურული დეტალის გარეშე, **Dall-E 2** არის სამი განსხვავებული მოდელის კომბინაცია:

- კლიპი
- წინა ნეირონული ქსელი
- დეკოდერის ნეირონული ქსელი

ამ დროისთვის საკმარისია ვიცოდეთ, რომ დეკოდერის ქსელი წარმოქმნის სურათებს დასკვნის დროს. **Dall-E 2-ის** ვებ **UI-ზე** წვდომა შესაძლებელია ოფიციალურ **OpenAI** ვებსაიტზე მოთხოვნის წარდგენით.

აქ არის რამდენიმე მოთხოვნა და მათი შესაბამისი სურათები, რომლებიც გენერირებულია **Dall-E 2-ის** მიერ.



"ადამიანები, რომლებიც მზის ამოსვლის დროს დადიან სანაპიროზე, მზის ანარეკლი წყალზე, რეალისტური"



"წყალზე მდგომი ფლამინგოები, წითელი მზის ჩასვლა, ვარდისფერ-წითელი წყლის ანარეკლი, ფოტორეალისტური, 4k"

6 - გამოსახულება

Dall-E 2-ის შემდეგ, მხოლოდ ერთი თვის შემდეგ, 2022 წლის აპრილში, Google-მა გამოუშვა დიფუზიაზე დაფუძნებული გამოსახულების გენერირების ალგორითმი. მათ სათანადოდ დაარქვეს **Imagen** (სურათის გენერირებისთვის).

იგი აგებულია დიდ ტრანსფორმატორზე დაფუძნებული ენის მოდელების თავზე. მათი გამოქვეყნებისთვის **Imagen**-ის ავტორებმა აირჩიეს **T5-XXL** ტრანსფორმატორის ენის მოდელი. ამის შემდეგ, **Imagen** შედგება გამოსახულების წარმოქმნის კიდევ სამი დიფუზიური მოდელისგან:

- დიფუზიური მოდელი **64×64** გარჩევადობის გამოსახულების გენერირებისთვის.
- მას მოჰყვება სუპერ გარჩევადობის დიფუზიური მოდელი სურათის **256 × 256** გარჩევადობამდე.
- და ერთი საბოლოო სუპერ გარჩევადობის მოდელი, რომელიც ამაღლებს სურათს **1024×1024** გარჩევადობამდე.

ამჟამად Google-ის **Imagen** მიუწვდომელია ფართო საზოგადოებისთვის და ხელმისაწვდომია მხოლოდ მოწვევით.



შიბა ინუს ძაღლის ფოტო ზურგჩანთით, რომელიც ველოსიპედს ატარებს. მას აცვია მზის სათვალე და პლაჟის ქუდი. "



" ლურჯი ჯეი დგას ცისარტყელას მაკარონის დიდ კალათზე "



" ტვინი, რომელიც ამხედრებს რაკეტას და მიემართება მთვარისკენ "

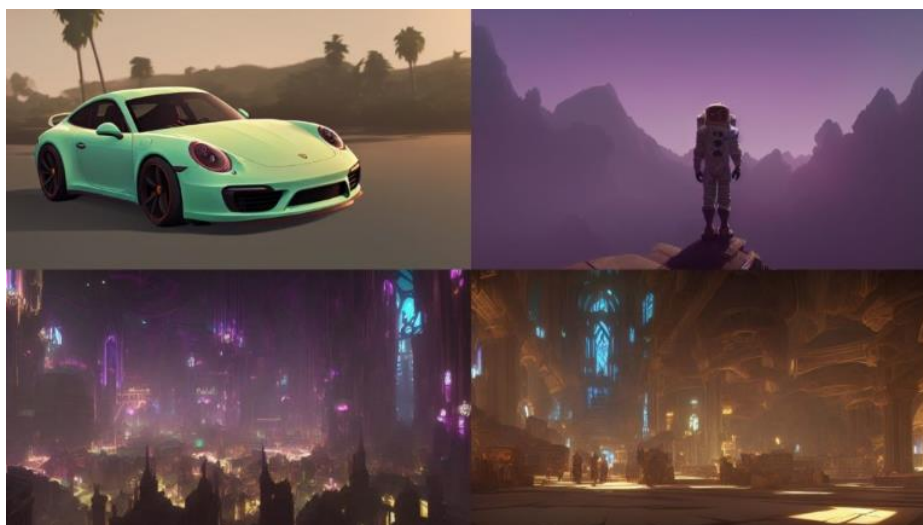
7 - სტაბილური დიფუზია

შექმნილია **StabilityAI**-ის მიერ , სტაბილური დიფუზია ეფუძნება მაღალი რეზოლუციის გამოსახულების სინთეზის მუშაობას ლატენტური დიფუზიის მოდელებით **Rombach et al.** ეს არის ერთადერთი დიფუზიაზე დაფუძნებული გამოსახულების გენერირების მოდელი ამ სიაში, რომელიც მთლიანად ღია წყაროა.

ამის დაწერის დღიდან, Stable Diffusion v2.1 ხელმისაწვდომია StabilityAI-ის ოფიციალურ საცავში.

არა მხოლოდ ეს, ღია კოდის საზოგადოება ძალიან აქტიურია მისი გამოშვების დღიდან. მოკლე დროში, საზოგადოებამ გამოუშვა რამდენიმე ღია კოდის მქონე სტაბილური დიფუზიის მოდელი, რომელიც კარგად არის მორგებული სხვადასხვა მხატვრულ სტილიზებულ მონაცემთა ნაკრებებზე. თავისუფლად შეგიძლიათ გამოიყენოთ ეს მოდელები და შექმნათ ახალი სურათები ამ სტილის გამოყენებით.

ეს შეიძლება მერყეობდეს იაპონური ანიმეებიდან და ფუტურისტული რობოტებიდან კიბერპანკ სამყაროებამდე. თქვენი ფანტაზიის დასაინტრიგებლად, აქ მოცემულია რამდენიმე მაგალითი სტაბილური დიფუზიის მოდელებიდან.



სტაბილური დიფუზიის სრული არქიტექტურა შედგება სამი მოდელისგან:

1. ტექსტური შიფრატორი, რომელიც იღებს ტექსტურ მოთხოვნას.
 1. ტექსტის მოთხოვნის გადაქცევა კომპიუტერით წასაკითხად ვექტორებად.
2. U-Net
 1. ეს არის დიფუზიური მოდელი, რომელიც პასუხისმგებელია სურათების გენერირებაზე.
3. ვარიაციური ავტოინკოდერი, რომელიც შედგება ენკოდერისა და დეკოდერის მოდელისგან.
 1. ენკოდერი გამოიყენება გამოსახულების ზომების შესამცირებლად. UNet დიფუზიის მოდელი მუშაობს ამ მცირე განზომილებაში.
 2. შემდეგ დეკოდერი პასუხისმგებელია დიფუზიური მოდელის მიერ წარმოქმნილი გამოსახულების გაძლიერებაზე/რეკონსტრუქციაზე თავდაპირველ ზომაზე.

მარტივად შეგიძლიათ წვდომა სტაბილური დიფუზიის მოდელებზე მათი **DreamStudio** პლატფორმის გამოყენებით . ანგარიშის შექმნა თავდაპირველად მოგცემთ 200 კრედიტს, რომელიც შეგიძლიათ გამოიყენოთ მოთხოვნებთან სათამაშოდ და სასურველი სურათების შესაქმნელად.

ალტერნატიულად, თუ თქვენ გაქვთ გამოთვლითი რესურსი, თქვენ ასევე შეგიძლიათ დააყენოთ სტაბილური დიფუზია თქვენს სისტემაზე გასაშვებად, მათ საცავში არსებული დოკუმენტაციის დაცვით.

8 - შუა მოგზაურობა

Midjourney არის კიდევ ერთი დიფუზიაზე დაფუძნებული გამოსახულების გენერირების მოდელი, რომელიც შემუშავებულია კომპანიის მიერ ამავე სახელწოდებით. Midjourney ხელმისაწვდომი გახდა მასებისთვის 2020 წლის მარტში. მან სწრაფად მოიპოვა დიდი მიმდევარი მისი ექსპრესიული სტილისა და იმ ფაქტის წყალობით, რომ იგი გახდა საჯაროდ ხელმისაწვდომი **DALLE**-სა და **Stable Diffusion**-მდე.

ამ დროისთვის ის სრულიად დახურული წყაროა და არ გააჩნია შესაბამისი ქაღალდი. თუმცა, შეიძლება წვდომა Midjourney-ის გამოსახულების გენერირების

შესაძლებლობებზე მათი ოფიციალური Discord ბოტის გამოყენებით. ცოტა ხნის წინ კომპანიამ გამოაცხადა **v4 მოდელის** ალფა ტესტირების ეტაპი .



"ავოკადოს სკამი" Midjourney ვერსიის v4 გამოყენებით - ჯეკ გალაჰერის მიერ

9 - სტაბილური დიფუზიის მოდელის პოპულარული ვარიაციები

Stable Diffusion-ის ღია კოდის ბუნება საშუალებას აძლევს დეველოპერებს მთელ მსოფლიოში ივარჯიშონ კონკრეტულ სტილზე და შექმნან სტაბილური დიფუზიის ვარიაციები. ჩვენ შეგვიძლია ვიპოვოთ უამრავი მაგალითი ინტერნეტში, მათ შორის **Stable Diffusion** მოდელები, რომლებიც ქმნიან დისნეის პერსონაჟებს, ანიმე პერსონაჟებს და სხვა დიფუზიური მოდელების სტილებსაც კი.

თითქმის ყველა მოდელი და მაგალითი, რომელსაც ჩვენ აქ მივყევით, ხელმისაწვდომია **HuggingFace** ჰაბის მეშვეობით. ასე რომ, თუ გსურთ მათი გამოცდა, არ მოგერიდოთ კოდის გაშვება.

აქ არის რამდენიმე მაგალითი:

10 - ყველაფერი V3

Anything V3 არის სტაბილური დიფუზიის ვარიაცია, რომელიც ქმნის სურათებს ანიმე პერსონაჟების სტილში.



„ სუპერმენი, რომელიც დაფრინავს
მაღალსართულიან შენობის თავზე, ახლო
კადრი 4k “

11 - Robo Diffusion

მოდელი, რომელიც წარმოქმნის რობოტებს იმ საგნიდან და სიმბოლოებიდან, რომლებიც ჩვენ შევიყვანთ ჩვენს მოთხოვნაში.



" დინოზავრის რობოტის წყლის ფერის ნახატი
გომილას სტილში "

12 - ღია მოგზაურობა

MidJourney-ის ღია კოდის საზოგადოების მიერ შექმნილი რეპლიკაცია, რომელიც აგენერირებს სურათებს ისეთივე სტილით, როგორსაც **MidJourney** ქმნის. **Open Journey** არის სტაბილური დიფუზიის ვარიაცია, რომელიც გაწვრთნილი იყო **MidJourney**-ის მიერ გენერირებულ სურათებზე.



" ფუტურისტული მანქანა ქუჩაში სიჩქარით
გადის წვიმიან ღამეს sdxjrny-v4 სტილში "

13 - არკანის დიფუზია

Arcan Diffusion არის სტაბილური დიფუზიის მოდელი, რომელიც დახვეწილია გამოსახულების სტილზე სატელევიზიო შოუდან **Arcane**.



14 - Mo-Di Fusion

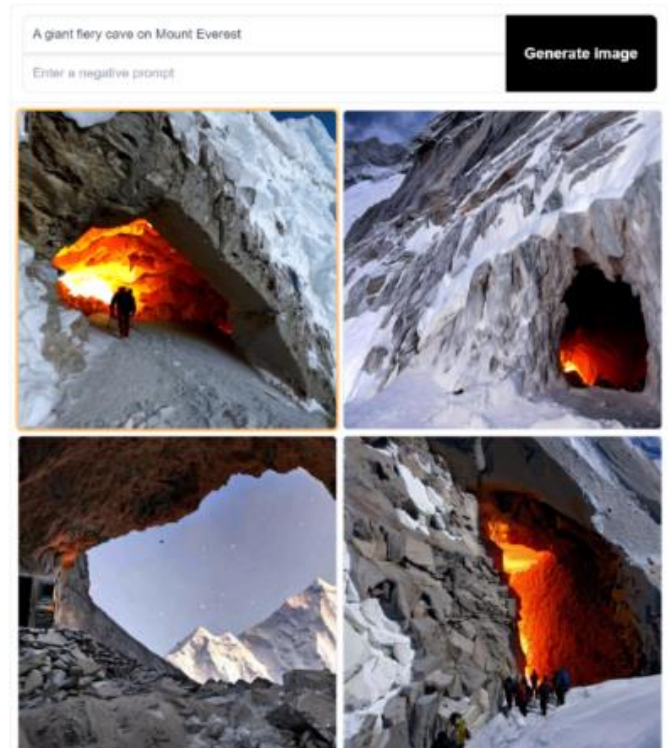
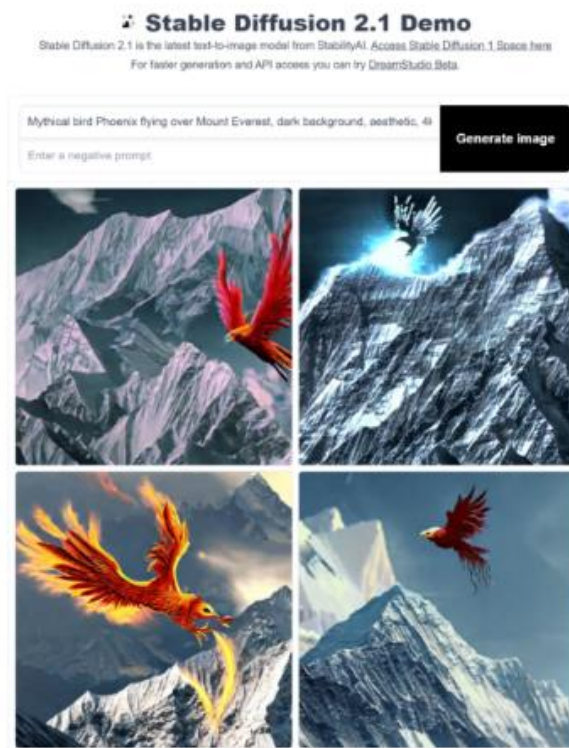
Stable Diffusion-ის ამ ვერსიამ გამოიმუშავა პერსონაჟები **Pixar**-ის მსგავსი სტილით.



დიფუზიური მოდელების აპლიკაციები

ზემოთ მოყვანილ განყოფილებებში ჩვენ განვიხილეთ რამდენიმე ცნობილი დიფუზიის მოდელი და მათ შორის, სტაბილური დიფუზია და მისი ვარიანტები ერთადერთია თავისუფლად ხელმისაწვდომი. კოდის ბაზის ღია კოდის გამო, დეველოპერთა და მკვლევართა საზოგადოებამ მოიფიქრა სტაბილური დიფუზიის გამოყენების ჰიპოთეზური გზები. ამ განყოფილებაში განვიხილავთ დიფუზიური მოდელების რამდენიმე მნიშვნელოვან აპლიკაციებსა და გამოყენებას.

სტაბილური დიფუზიის ტექსტიდან გამოსახულების გენერირების მოდელების გამოყენების უმარტივესი გზა არის [Stable Diffusion 2-1 – Hugging Face Space](#) [სტაბილურობის მიერ](#). უბრალოდ დაამატეთ ტექსტური აღწერა და დააწკაპუნეთ "გამოსახულების გენერირება".



თავისუფლად ჰოსტინგის სერვისის გამოყენების ნაკლოვანებები არის ის, რომ მას შეიძლება ჰქონდეს დიდი რიგი, ასე რომ თქვენ შეიძლება დაგჭირდეთ დაელოდოთ თქვენი სურათის გენერირებას და არ არის ხელმისაწვდომი პერსონალიზაციის ვარიანტები ან კონტროლი.

ამის გადასაჭრელად, ჩვენ შეგვიძლია ჩამოტვირთოთ და გავუშვათ სტაბილური დიფუზიის მოდელები ადგილობრივად. არსებობს ორი მიდგომა, რომელიც შეიძლება მივიღოთ:

1. დააინსტალირეთ საზოგადოებისთვის მზა აპლიკაცია, რომელიც შეგვიძლია ადგილობრივად დავაინსტალიროთ ან გამოვიყენოთ **Google Colab**-ზე.
2. პირდაპირ ვიმუშავეთ ღია კოდის კოდზე და გამოვიყენოთ ისე, როგორც ჩვენ გვინდა. ამისთვის ჩვენს მკითხველს ვამღებთ იუპიტერის რეულს.

ამ მიდგომის ერთადერთი მინუსი არის ის, რომ ჩვენ გვჭირდება ადგილობრივი **GPU**. პროცესორებზე პროცესი გაცილებით ნელი იქნება. ჩვენ ჩამოვთვალეთ სივრცეები და ხელსაწყოები, რომლებიც გამოვიყენეთ.

ასე რომ, დავიწყეთ.

ტექსტური ინვერსია

ტექსტური ინვერსიის გამოყენებით, ჩვენ შეგვიძლია გავაუმჯობესოთ დიფუზიური მოდელი, რათა გამოსახოთ სურათები პერსონალური ობიექტების ან მხატვრული სტილის გამოყენებით 10-მდე გამოსახულების გამოყენებით. ეს თავისთავად არ არის აპლიკაცია, არამედ ჰიპოთეზური ხრიკი დიფუზიური მოდელების მოსამზადებლად, რომელიც შეიძლება გამოიყენოს უფრო პერსონალიზებული სურათების შესაქმნელად.

ტექსტური ინვერსიის ნაშრომის ავტორებისგან:

„ჩვენ ვსწავლობთ კონკრეტული ცნებების გენერირებას, როგორიცაა პერსონალური ობიექტები ან მხატვრული სტილი, მათი აღწერით ახალი „სიტყვების“ გამოყენებით წინასწარ გაწვრთნილი ტექსტიდან გამოსახულების მოდელების ჩანერგვის სივრცეში. ეს შეიძლება გამოყენებულ იქნას ახალ წინადადებებში, ისევე როგორც ნებისმიერი სხვა სიტყვა.“

- სურათი ერთი სიტყვით ღირს, 2022 წელი



✦ ტექსტური ინვერსიით, ჩვენ შევქმენით რამდენიმე სატიას გამოსახულების გამოყენებით Lensa iOS აპის გამოყენებით.



თუ არ ეძებთ რაიმე ზედმეტად პირადულ და გსურთ შექმნათ სურათები ზოგიერთი ცნობილი სტილის გამოყენებით, შეგიძლიათ სცადოთ [Stable Diffusion Conceptualizer](#) [სივრცე](#). აქ ნახავთ დიფუზიური მოდელების კოლექციას, რომლებიც მომზადებულია სხვადასხვა მხატვრულ სტილზე. თქვენ შეგიძლიათ აირჩიოთ ნებისმიერი თქვენი სურვილისამებრ და დაიწყოთ გენერირება ყოველგვარი პრობლემების გარეშე. თქვენ შეგიძლიათ აირჩიოთ ნებისმიერი თქვენი გემოვნებით და დაიწყოთ გენერირება უპრობლემოდ.

მაგალითად, ჩვენ შევქმენით ქვემოთ მოცემული სურათი შავ მთაზე განდალგ გრის „შუამოგზაურობის სტილის“ გამოყენებით.

A <midjourney-style> style image of Gandalf the gray on a black r

Run



"შუამოგზაურობის სტილის> სტილის სურათი შავ მთაზე რუხი
განდალფი"

ტექსტი ვიდეოებზე

როგორც სახელი გვთავაზობს, ჩვენ შეგვიძლია გამოვიყენოთ დიფუზიური მოდელები ვიდეოების პირდაპირ გენერირებისთვის ტექსტური მოთხოვნის დახმარებით. ტექსტის გამოსახულებაზე გამოყენებული კონცეფციის გაფართოებით ვიდეოებზე, შეგვიძლიათ გამოვიყენოთ დიფუზიური მოდელები ვიდეოების გენერირებისთვის მოთხოვნილებიდან, სიმღერებიდან, ლექსებიდან და ა.შ.



„ლამაზი ტყე აშერ ბრაუნ დიურანდის მიერ,
ტენდენციურია Artstation-ზე“
წყარო: მაგალითი დეფორმ სტაბილური
დიფუზიიდან – ანიმაციური მითითებები
სტაბილური დიფუზიით



„კატა / ძაღლი / ცხენი“
წყარო: მაგალითი „სტაბილური დიფუზიის
ვიდეოებიდან“ Replicate-ზე – შექმენით ვიდეოები
სტაბილური დიფუზიის ფარული სივრცის
ინტერპოლაციის გზით

არსებობს სხვა (შესაძლოა უკეთესი) ტექსტი-ვიდეო დიფუზიის მოდელები, როგორიცაა Google- ის **Imagen Video** და **Phenaki** და Meta-ს **Make-A-Video** . სამწუხაროდ, კოდი ამჟამად ხელმისაწვდომია საზოგადოებისთვის.

ტექსტი 3D

ეს აპლიკაცია წარმოდგენილი იყო „**Dreamfusion**“ ნაშრომში, სადაც ავტორებს „**NeRFs**“-ის დახმარებით შეეძლოთ გამოეყენებინათ გაწვრთნილი 2D ტექსტიდან გამოსახულების დიფუზიის მოდელი ტექსტიდან 3D სინთეზის შესასრულებლად.

”მოცემული ტექსტის შედეგად მიღებული 3D მოდელის ნახვა შესაძლებელია ნებისმიერი კუთხიდან, გადაღებული თვითნებური განათებით, ან კომპოზიცია ნებისმიერ 3D გარემოში. ჩვენი მიდგომა არ საჭიროებს 3D ტრენინგის მონაცემებს და არანაირ მოდიფიკაციას გამოსახულების დიფუზიის მოდელში, რაც ასახავს წინასწარ

მომზადებული გამოსახულების დიფუზიის მოდელების ეფექტურობას, როგორც პრიორიტეტებს.

- DreamFusion: ტექსტი 3D-ში 2D დიფუზიის გამოყენებით, 2022 წ

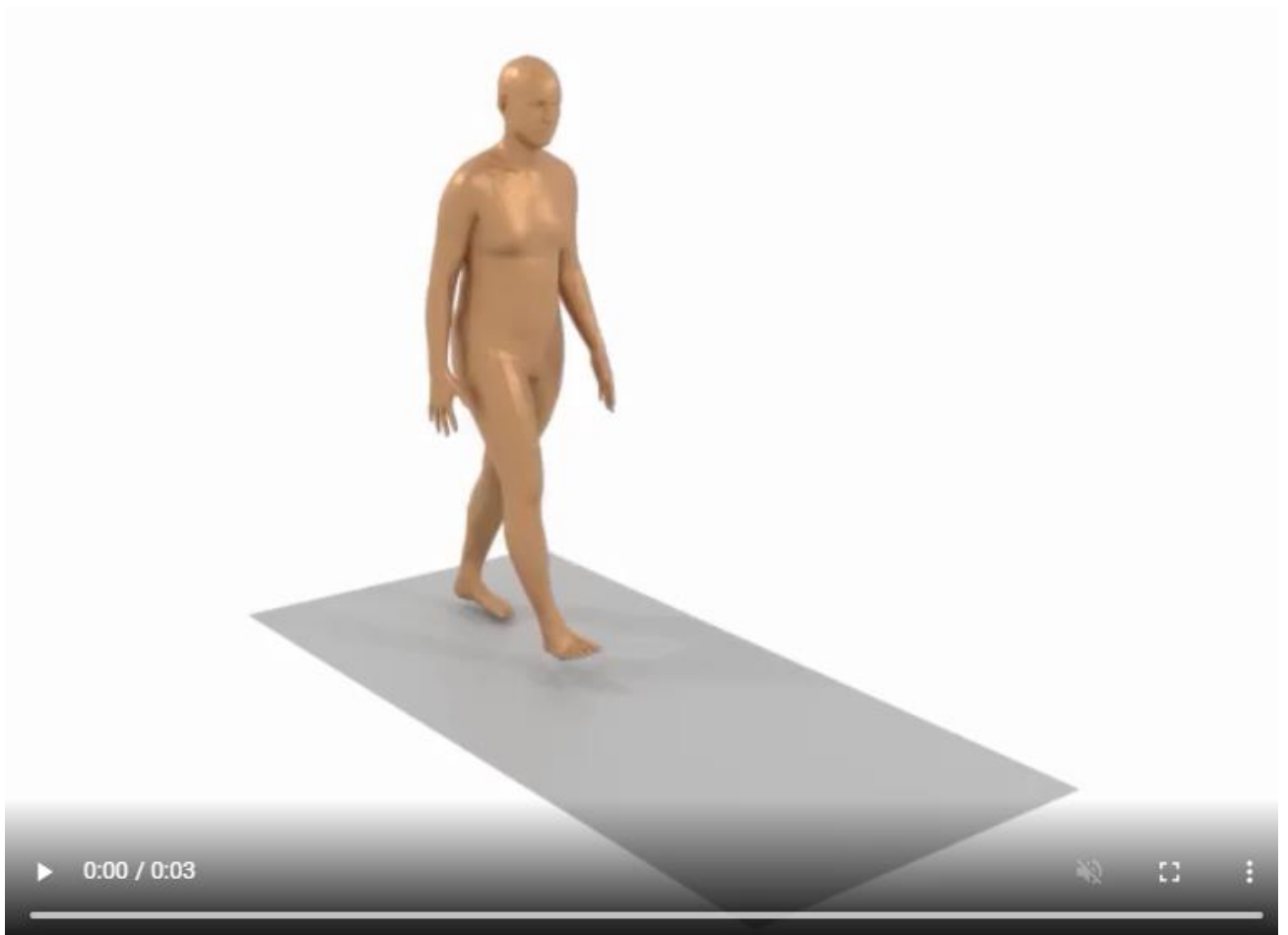


ტექსტი მოძრაობაში

ეს არის კიდევ ერთი მომავალი და საინტერესო აპლიკაცია, სადაც დიფუზიის მოდელები (სხვა ტექნიკასთან ერთად) გამოიყენება ადამიანის მარტივი მოძრაობის შესაქმნელად. კონკრეტულად, ჩვენ ვგულისხმობთ *"ადამიანის მოძრაობის დიფუზიის მოდელს"*.

"ადამიანის ბუნებრივი და ექსპრესიული მოძრაობის გენერაცია არის კომპიუტერული ანიმაციის წმინდა გრაალი. ეს რთული ამოცანაა შესაძლო მოძრაობის მრავალფეროვნების, მის მიმართ ადამიანის აღქმის მგრძნობელობისა და მისი ზუსტი აღწერის სირთულის გამო. ჩვენ ვაჩვენებთ, რომ ჩვენი მოდელი გაწვრთნილია მსუბუქი რესურსებით და, მიუხედავად ამისა, აღწევს უახლესი შედეგების წამყვან ეტალონებს ტექსტის მოძრაობაში და მოქმედების მოძრაობაში. "

- MDM: ადამიანის მოძრაობის დიფუზიის მოდელი, 2022 წ



” ადამიანი წინ მიდის, იხრება, რომ მიწიდან რაღაც აიღოს. წყარო
: <https://guytevet.github.io/mdm-page/>

გამოსახულება სურათზე

Image-to-Image (მოკლედ **Img2Img**) არის მეთოდი, რომელიც შეგვიძლია გამოვიყენოთ არსებული სურათების შესაცვლელად. ეს არის არსებული სურათის სამიზნე დომენად გარდაქმნის ტექნიკა ტექსტური მოწოდების გამოყენებით. მარტივად რომ ვთქვათ, ჩვენ შეგვიძლია შევქმნათ ახალი სურათი იგივე შინაარსით, როგორც არსებული სურათი, მაგრამ გარკვეული ტრანსფორმაციებით. ჩვენ შეგვიძლია მოგვაწოდოთ ჩვენთვის სასურველი ტრანსფორმაციის ტექსტური აღწერა.

მაგალითად, ქვემოთ მოყვანილ ბადეში, მარცხენა მხარეს გამოსახულება არის „ჰაერის ბუშტების ფესტივალი“, რომელიც ჯერ გენერირებულია „**text2img**“-ის გამოყენებით და

შემდეგ გადაეცა **img2img**-ს, რათა გადააკეთოს იგი **Disney Pixar**-ის სტილის გამოსახულების ხელოვნებად.



L – “ საჰაერო ბუშტები, ბურთების ფესტივალი, ბნელი ცა, კაშკაშა ვარსკვლავები, ჯილდოს მფლობელი, 4k. Hd“

R –” პიქსარი, დისნეის სურათები, საჰაერო ბუშტები, ბურთების ფესტივალი, ბნელი ცა, კაშკაშა ვარსკვლავები, ჯილდოს მფლობელი, 4k. HD”

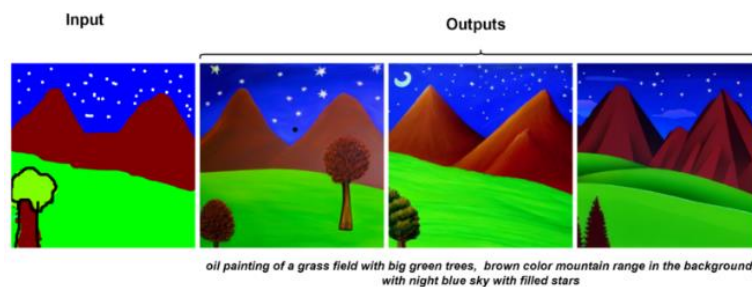
მომხმარებელი უზრუნველყოფს შეყვანის სურათს, როგორც სახელმძღვანელო/საწყის წერტილს და შეუძლია შეცვალოს იგი მოდელზე ინსტრუქციების მიწოდების გზით. დიფუზიური მოდელი გამოიმუშავებს ახალ სურათს იმავე ფერისა და შემადგენლობით, როგორც შეყვანილი სურათი ახალი ცვლილებებით.

img2img-ის გამოყენებით, თქვენ ასევე შეგიძლიათ *გადაიყვანოთ თქვენი პორტრეტები Pixar-ის სტილის სურათებად* .



Img2Img-ს კიდევ ერთი საინტერესო გამოყენების შემთხვევა, უხეში ხელით დახატული ან დაუმთავრებელი სურათების იმავე შინაარსის ლამაზ სურათებად გადაქცევისთვის. როგორც ქვემოთ მოცემულ მაგალითში.

როგორც ექსპერიმენტი: მე გადავაქციე ჩემი „ზუნების ცუდი ნახატი“ „ზეთების მხატვრობად“ და შედეგები შესანიშნავია.



Img2Img-ის გამოყენების შემთხვევაა **Style Transfer**. **Style Transfer**-ში ჩვენ ვიღებთ ორ სურათს: ერთი შინაარსისთვის და ერთი სტილის მითითებისთვის. იქმნება ახალი სურათი, რომელიც არის ორივეს ნაზავი: პირველის შინაარსი სხვის სტილში.

სტილის გადაცემა საინტერესო და სახალისო თემაა შესასწავლად. ჩვენ გამოვიყენეთ **Style Transfer** და შევექმნით აპლიკაცია რეალურ დროში სტილის გადაცემისთვის **Zoom** ზარებში.



სტილის გადაცემის მაგალითი

გამოსახულების შეღებვა

Image Inpainting არის გამოსახულების აღდგენის ტექნიკა გამოსახულების არასასურველი ობიექტების მოსაშორებლად ან მათი მთლიანად ჩანაცვლების სხვა ობიექტით/ტექსტურით/დიზაინით. გამოსახულების შეღებვის შესასრულებლად, მომხმარებელი ჯერ ხატავს ნილბს ობიექტის ან პიქსელების გარშემო, რომლებიც უნდა შეიცვალოს. ნილბის შექმნის შემდეგ მომხმარებელს შეუძლია უთხრას მოდელს, თუ როგორ უნდა შეცვალოს ნილბიანი პიქსელები.



ჩანაცვლება: მანქანა → ტანკი



ობიექტის მოცილება: ცხვრის ჩანაცვლება → ბალახი
 აქ ცხვრის მარცხნივ ნაწილი ჩანაცვლებულია მწვანე ბალახით.



ჩანაცვლება: ვაშლი → ფორთოხალი



ჩანაცვლება: ფრანგული რქა → კატა

გამოსახულების შეღებვა

შეღებვა ან უსასრულო მხატვრობა არის პროცესი, რომლის დროსაც დიფუზიური მოდელი ამატებს დეტალებს ორიგინალური გამოსახულების გარეთ/მიღმა. ჩვენ შეგვიძლია გავაფართოვოთ ჩვენი ორიგინალური სურათი ან ორიგინალური სურათის ნაწილების (ან ახლად წარმოქმნილი პიქსელების) გამოყენებით, როგორც საცნობარო პიქსელებს, ან ახალი ტექსტურებისა და ცნებების დამატებით ტექსტური მოწოდების გამოყენებით.

Infinite Zoom In & Zoom Out

A. უსასრულო მასშტაბირება

ეს შეიძლება ჩაითვალოს გამოსახულების შეღებვის ჭკვიანურ გზად ფრაქტალის მსგავსი დიზაინის ნიმუშების შესაქმნელად.

პროცესი, რომელიც გამოიყენება „Infinite Zoom in“-ის შესაქმნელად, შემდეგია:

1. შექმენით სურათი სტაბილური დიფუზიის მოდელის გამოყენებით.
2. შეამცირეთ სურათი და დააკოპირეთ და ჩასვით ცენტრში.
3. დაფარეთ საზღვარი პატარა ასლის გარეთ.
4. გამოიყენეთ სტაბილური დიფუზიური შეღებვა ნილბიანი ნაწილის შესავსებად.

B. უსასრულო მასშტაბირება

„Infinite Zoom Out“-ის გენერირებისთვის გამოყენებული მეთოდები საკმაოდ მარტივი გასაგებია. ის შეიძლება ჩაითვალოს „გადაღების“ გაფართოებად. სურათის შეღებვისგან განსხვავებით, სურათის ზომა მთელი პროცესის განმავლობაში უცვლელი რჩება.

საწყისი მოწოდების გამოყენებით წარმოქმნილი სურათის ზომა თანდათან მცირდება. ეს ქმნის ცარიელ სივრცეს გამოსახულების საზღვრებსა და ზომის შეცვლილ სურათს შორის. დამატებითი სივრცე ივსება დიფუზიური მოდელის გამოყენებით, რომელიც განპირობებულია იმავე მოთხოვნაზე, ახალი გამოსახულების (შეიცავს ცარიელ ადგილს და საწყის სურათს) საწყის წერტილად. ჩვენ შეგვიძლია შევქმნათ ვიდეო ამ პროცესის

გაგრძელებით რამდენიმე გამეორებისთვის, როგორც ეს ნაჩვენებია ზემოთ მოცემულ მაგალითებში.

სურათების ძებნა და გამოსახულების უკუ ძიება

ეს ორი არის კომუნალური სერვისები, რომლებიც შექმნილია სურათების მოსაძებნად, რომლებიც შეიქმნა სტაბილური დიფუზიის გამოყენებით ან გამოსახულების საპირისპირო ძიების შესასრულებლად.

1. **Lexica** – ეს არის სტაბილური დიფუზიის საძიებო სისტემა. ის მასპინძლობს ~5 მილიონზე მეტ სურათს, რომელიც გენერირებულია, და მოთხოვნა ქმნის სურათს.
2. **Ternaus – Ternaus** განსხვავებულია, რადგან ის აერთიანებს გამოსახულების ძიებას (სურათი სურათზე) და გამოსახულების საპირისპირო ძიებას (სურათი ტექსტში) ერთ მოთხოვნაში. ის ფოკუსირებულია გენერირებული სურათების პოვნაზე ლიცენზიის შეზღუდვის გარეშე.
3. პალიტრა: **Image-to-Image Diffusion** მოდელები , რომლებიც იყენებენ დიფუზიის მოდელებს გამოსახულების შეღებვის, შეფერილობის, ამოღების (გადაღებვის მსგავსი) და **JPEG** აღდგენისთვის.
4. **DiffusionDet**: დიფუზიური მოდელი ობიექტების გამოვლენისთვის
5. **DiffSinger**: სიმღერის ხმის სინთეზი ზედაპირული დიფუზიის მექანიზმის მეშვეობით
6. **Pix2Seq-D**: ზოგადი ჩარჩო სურათებისა და ვიდეოების პანოპტიკური სეგმენტაციისთვის

ჩვენ გვჯერა, რომ ჩვენ მიერ წარმოდგენილი აპლიკაციების გავლის შემდეგ, შეიძლება ჩვენი მკითხველისთვის საწყის წერტილად იმოქმედოს, რათა მოამზადონ და კომფორტული გახადონ თავიანთი გზის პოვნა დიფუზიური მოდელების სამყაროში.

ხელსაწყოები

ამ განყოფილებაში ჩვენ ჩამოვთვლით საზოგადოების მიერ შექმნილ ცნობილ და მომავალ ინსტრუმენტებს და ონლაინ სერვისებს, რომლებსაც ჩვენ მომხმარებლებს შეუძლიათ შეისწავლონ და გამოიყენონ თავიანთი ამოცანების შესასრულებლად.

ლოკალური და კოლაბების ინსტალაციისთვის

1. ყველაზე პოპულარული – [AUTOMATIC1111: სტაბილური დიფუზიის ვებ UI](#)
 1. გაშვება Google Colab [სწრაფად სტაბილურად დიფუზიის Colab ნოუტბუქებზე, AUTOMATIC1111](#)
2. [invoke-ai: InvokeAI არის წამყვანი კრეატიული ძრავა სტაბილური დიფუზიის მოდელისთვის](#) .
3. [Sygil-Dev: სტაბილური დიფუზიის ვებ UI](#)
4. [cmdr2: უმარტივესი 1 დაწყაპუნებით დააინსტალირეთ Stable Diffusion](#)
5. [abhishekrthakur: დიფუზორები: ვებ ინტერფეისი ო დიფუზორებისთვის](#)
6. [დიფუზორები](#)

ონლაინ ტესტირებისთვის

1. [Craiyon](#) – Craiyon არის ხელოვნური ინტელექტის მოდელი, რომელსაც შეუძლია სურათების დახატვა ნებისმიერი ტექსტური მოთხოვნიდან!
2. [nateraw/stable-diffusion-videos](#)
3. [deforum/deforum stable diffusion](#)
4. [ashawkey/stable-dreamfusion: ტექსტის 3D Dreamfusion-ის პიტორჩის განხორციელება, რომელიც უზრუნველყოფილია სტაბილური დიფუზიით.](#)
5. [MDM - ადამიანის მოძრაობის დიფუზიის მოდელი](#)
6. [stability-ai/stable-diffusion-img2img](#)
7. [Diffuse the Rest – ჩახუტებული სახის სივრცე](#)
8. [ასაფრენი ბილიკი შეღებვა - ჩახუტებული სახის სივრცე](#)
9. [სტაბილური დიფუზიური შეღებვა - ჩახუტებული სახის სივრცე](#)
10. [Stablediffusion Infinity – უსასრულო ტილოზე სტაბილური დიფუზიის ამალღება.](#)
11. [arielreplicate/stable diffusion infinite zoom in](#)
12. [ჩახუტებული სახის დიფუზიური სივრცეები](#)

დიფუზიური მოდელები აქ დარჩება!

დიფუზიურ მოდელებს შეუძლიათ უზომოდ გაზარდონ შემოქმედებითი სამუშაოების სამყარო და ზოგადად შინაარსის შექმნა. მათ უკვე დაადასტურეს თავიანთი შესაძლებლობები ბოლო რამდენიმე თვის განმავლობაში. დიფუზიური მოდელების რაოდენობა ყოველდღიურად იზრდება და ძველი ვერსიები სწრაფად მოძველებულია. სინამდვილეში, ძალიან დიდია იმის შანსი, რომ სანამ ამ წიგნს წაიკითხავთ, ზემოთ ხსენებულ ზოგიერთ მოდელს ექნება თავისი ახალი ვერსიები.

ნებისმიერ შემთხვევაში, დროა შემქმნელებმა და ბიზნესებმა დაიწყონ დიფუზიური მოდელების გამოყენება გამოსახულების და კონტენტის გენერირებისთვის. მაშინაც კი, თუ არა როგორც საბოლოო პროდუქტი, დიფუზიურ მოდელებს შეუძლიათ იმოქმედონ როგორც საშუალება შთააგონონ შემოქმედებითი მუშაკების სამყარო.

ამ წიგნში , ჩვენ გავაშუქეთ დაკავშირებული თემების ყოვლისმომცველი სია.

შეჯამება:

1. ჩვენ დავიწყეთ დიფუზიის მოდელების ადგილის შესწავლით ღრმა სწავლაზე დაფუძნებულ გენერაციულ მოდელებში.
2. ჩვენ მივაწოდეთ მაღალი დონის მიმოხილვა დიფუზიისა და დიფუზიის მოდელებისა და მათი მუშაობის შესახებ.
3. ჩვენ განვიხილეთ ზოგიერთი ყველაზე გავლენიანი და პოპულარული დიფუზიის მოდელი და მათი საზოგადოების მიერ შექმნილი ვარიანტები.
4. დაბოლოს, ჩვენ დავასრულეთ დიფუზიური მოდელების სხვადასხვა საინტერესო და სახალისო აპლიკაციების განხილვა და სხვადასხვა ხელსაწყოები და სერვისები, რომელთა გამოყენება შესაძლებელია დიფუზიის მოდელებთან სათამაშოდ.

ჩვენ სიამოვნებით მოვისმენთ თქვენგან. გთხოვთ, თავისუფლად დასვათ შეკითხვები, ჩვენ სიამოვნებით ვისაუბრებთ თქვენთან.

ჩვენი კოორდინატები:

e-mail: isheriphadze@gmail.com

mob: +995(555)45-92-70

site: <https://github.com/ImedaSheriphadze>

★ბედნიერ სწავლას გისურვებთ!

პატივისცემით, ინფორმაციული ტექნოლოგიების და პროგრამული უზრუნველყოფის
ასოცირებული სპეციალისტი: იმედა შერიფაძე