

Генерисање уметничких слика употребом GAN технологија

Мариета Ракош, R2-27/2021

Факултет техничких наука, Универзитет у Новом Саду
Нови Сад, Србија

E-mail: marieta.rakos@gmail.com,

Апстракт — У овом раду представљена су решења проблема на тему генерисања уметничких слика употребом GAN технологија. Конкретно, испитани су следећи модели генеративних неуронских мрежа: DCGAN и CGAN, и то неколико различитих архитектура, ради добијања најповољнијег решења. Поменути модели су обучавани над различитим скуповима слика, као што су портрети, цвеће и мртва природа. Скупови слика преузети су са сајта Kaggle, а састоје се од слика преузетих са сајта Wiki-art. Скупови података нису беспрекорни у смислу да могу садржати слике које не припадају одређеном скупу, па су те слике одстрањене. Модели су евалуирани помоћу субјективне процене, по угледу на релевантне радове. Изгенерисане слике су показале да боље перформансе постиже DCGAN који се фокусира само на један скуп слика, на супрот CGAN-у који у обзир узима више скупова и води рачуна о лабелама. Такође, праћене су loss функције генератора и дискриминатора, ради провере напретка модела. Примењени су неки од савета за унапређивање генеративних модела, док су се неки савети показали као погоршање за моделе у овом раду. Укратко је размотрена литература која се бави сличном темом. Додатно, описане су грешке модела и који су узроци томе. За потребе овог рада испитано је неколико подскупова уметничких дела, а даља истраживања би се проширила на остале подскупове.

Кључне речи — машинско учење; GAN; уметничка дела; генерисање слика; DCGAN; порјепреши

I. УВОД

Генеративни модели – *Generative Adversarial Networks (GAN)* – се углавном користе за генерисање реалних слика: људска лица, мацке, ентеријер, возила, цифре и сл. Овакве слике често имају јасно одвојиву позадину од објекта који се генерише – слике углавном садрже само један објекат и увек су „у истом стилу“. Изазов овог рада јесте генерисање уметничких слика, које немају претходно описане карактеристике. Уметничке слике такође могу садржати исти тип објекта (нпр. слике портрета или цвећа), а да су насликане другим стилем, као што су кубизам, надреализам импресионизам итд. Код неких стилова тешко је раздвојити позадину од предњег плана, нпр. код апстрактног сликарства, док се код кубизма не појављују природни облици. Додатно, слике одређеног правца уопште не морају осликавати исте предмете и могу бити

потпуно различите, како је на пример случај са мртвом природом. Потенцијално решење оваквог проблема биће разматрано у наставку рада.

Са развојем генеративних модела, развија се и нови правац у уметности – *AI art*, који постаје популаран на друштвеним платформама попут Твитера и Тиктока.

Задатак овог рада јесте генерисање уметничких слика употребом генеративних модела: DCGAN, који представља додавање конволутивних слојева класичном GAN-у и CGAN, који у обзир узима лабеле, односно категорије слика приликом генерисања. Резултати решења биће субјективно процењени и упоређени са сродним радовима, а пратиће се и loss функције (квантификација грешке модела) генератора и дискриминатора током епоха.

У поглављу II биће представљена решења која се баве сличним проблемом генерисања уметничких слика. У поглављу III биће описан скуп података који је коришћен за израду предложеног решења: начин на који је прикупљен, од чега се састоји, извршене трансформације, анализа добијеног скупа и сл. У поглављу IV биће представљена предложена решења. У V поглављу биће приказани постигнути резултати предложених решења, као и дискусија на тему грешака датих модела. У последњем поглављу VI биће описана кратка сумаризација целог рада и приказаће се најважнији закључци. Такође, биће представљена даља истраживања на ову тему.

II. РЕЛЕВАНТНА ЛИТЕРАТУРА

У овом поглављу ће укратко бити описана решења која су релевантна за тему овог рада. Описаће се проблем или задатак којим се баве, скуп података који користе и постигнута решења и метрике евалуације. Треба имати на уму да не постоји тачно одређена метрика којом се процењује колико добро модел генерише уметничку слику, јер је уметност сама по себи креативан процес са субјективним осећајем лепоте. Зато већина представљених решења користе добровољце за процену бољег модела.

A. GANgogh: Creating Art with GANs.

Аутори рада [3] баве се генерисањем уметничких слика. Користе модел AC-GAN (*Auxiliary Classifier GAN*) уз *wasserstein* метрику. Додају класификациону

компоненту дискриминатору та предикцију жанра дате слике.

Креирају свој скуп података прикупљањем слика са Wikiart сајта [8]. Оформљен скуп података је доступан на сајту Kaggle, одакле је преузет за потребе овог решења.

Као крајњи резултат представљају слике које ју класификоване са великом сигурношћу у одређени жанр и које су окарактерисане као реалне.

B. ARTGAN: ARTWORK SYNTHESIS WITH CONDITIONAL CATEGORICAL GANS

Аутори рада [2] се баве проблемом генерисања комплекснијих слика попут уметничких слика. Тврде да је мана класичних GAN-ова у томе што не узимају у обзир категорије, односно лабеле. У свом решењу представљају модел попут кондиционалног GAN-а (енг. *Conditional GAN*, скраћено *C-GAN*). Користе додатни вектор лабела као улаз у генератор, али не и у дискриминатор, што представља разлику у односу на C-GAN. Ово, како кажу, доприноси обучавању генератора тако што учи из повратних информација у виду лабела које добија из дискриминатора. Такође, уместо *softmax* користе *sigmoid* функцију у дискриминатору.

Скуп података који су аутори користили је [4], који представља скуп слика преузет са странице Wikiart. На овој страници се може пронаћи око 80.000 аотираних уметничких дела на основу жанра, сликара и стила. Сlike деле на основу жанра и сликара, и на тај начин пореде моделе. Такође, ради поређења користе и CIFAR-10 скуп података, који садржи 60.000 слика у боји распоређених у 10 класа.

Свој модел називају ArtGAN, а пореде га са перформансама DC-GAN и VAE/GAN модела. Закључују да DCGAN постиже најгоре резултате, а да ArtGAN успева да извуче специфичности жанра и конкретног сликара.

C. End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks

Тема рада [1] јесте развој генеративних модела који генеришу традиционалне кинеске пејзаже. Развијају сопствену архитектуру SapGAN од две фазе GAN-а:

- skatch – генерисање скичева пејзажа
- paint – бојење скичева употребом CGAN-а.

Као скуп података користе сопствени скуп од 2192 слика традиционалних кинеских пејзажа високе резолуције [5].

Свој модел пореде са DCGAN, RaLSGAN и StyleGAN2 моделима. Квалитет модела процењују тако што добровољцима показују праве и лажне слике и бележе у колико случајева су лажне слике лабелиране као праве. Показало се да чак у 55% случајева, лажне слике пролазе као праве, док лажне слике нпр. RaLSGAN модела пролазе у 11% случајева.

D. GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network

Аутори у раду [6] аутори представљају преглед радова на тему генерисања уметничких дела. На самом почетку дају теоријске основе генеративних модела, а затим деле рад на три дела: генерисање слика, музике и литературе. За сваки од ових поља аутори представљају изазове, архитектуре модела и нуде поређење резултата. Ово све сумирају табеларно ради лакшег прегледа.

У прегледу се могу уочити трендови: најчешће архитектуре су CGAN и DCGAN, честе *loss* функције су *cross entropy* и *binary cross entropy*, резултати су најчешће евалуирани на основу квалитета слика у поређењу са претходним решењима и употребом *Frechet Inception Distance* (FID) метрике.

E. CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms

Аутори рада [7] формирају генеративни модел који уноси креативност у генерисање уметничких слика. Креативност уноси на тај начин да модел одступа од научених стилова. Генератор добија два сигнала од дискриминатора: да ли је генерисану слику оценио као реалну или лажну (као и у класичном GAN-у) и колико добро је класификовао слику у неки стил. Циљ је прећи и збунити дискриминатор.

Као скуп података користе Wikiart dataset [8], као и у неким претходно поменутих радовима.

Тврде да неретко добијају апстрактне слике и постаљају питање да ли то значи да модел не генерише добро или да је креативан. Добијене слике пореде са резултатима *baseline* модела DCGAN и закључују да својим моделом постижу побољшање.

F. Резиме

Претходно описани радови предлажу нове архитектуре генеративних модела, а своје перформансе пореде са DCGAN и CGAN моделима. Предложено решење овог рада покушаће да, са једноставнијим структурама уз додатна намештања и примену савета за побољшање перформанси, постигне задовољавајуће резултате. Већина описаних радова скуп података прикупља са поменутог сајта Wikiart [8], што ће бити употребљено и у овом раду.

III. СКУП ПОДАТАКА

У овом поглављу биће детаљно представљен поступак формирања скупа података потребног за креирање модела. У првом делу овог поглавља биће више речи о самом преузимању података, који подаци се преузимају и његов обим. У другом делу биће описане трансформације које су резултовале крајњим скуповима података.

A. Опис података

За потребе развоја решења коришћен је Kaggle Wiki-Art: Visual Art Encyclopedia [9] скуп података, који садржи преко 80.000 слика, распоређених по категоријама: abstract, cityscape, animal painting, figurative, flower

painting, genre painting, landscape, marina, mythological, nude, religious, portrait, still life, symbolic painting. О целокупног скупа података, разматрани су следећи подскупови:

- folwer paintings – силке цвећа, укупно 1800
- landscape – пејзажи, укупно 15000 слика
- religious paintings – религијске слике, укупно 8400
- still life – силке мртве природе, укупно 2996 слика.
- Portraits – слике портрета, укупно 14999 слика.

Други скуп података који је коришћен јесте *Kaggle Art Portraits* [10] скуп који представља подскуп горенаведеног скупа портрета. Како оригинални скуп портрета садржи портрете различитих жанрова, овај скуп су аутори извели тако да садржи само оне портрете који су међусобно слични по стилу, и портрете који су слично оријентисани – гледају право, лево или десно. Аутори су овај скуп креирали да би унапредили перформансе својих генеративних модела, а скуп укупно садржи 4117 слика.

В. Прејроцесирање

Анализом преузетих скупова података утврђено је да се у сваком подскупу јављају слике које не би требало сврстати у ту категорију. На пример, у скупу пејзажа пронађено је неколико слика са људима, слике лица и слике грађевина. Такође, пронађено је мноштво слика чији стил много одступа од стила осталих слика. Све ове слике, које представљају вишак су уклоњене, како не би утицале на смањење перформанси модела.

Што се тиче правца мртве природе, утврђено је да је то веома широк појам, те да постоје разни предмети који су служили као инспирација. По угледу на сажети скуп портрета, прикупљене су само слике на којима се јасно види да постоји сто или подлога, и предмети на столу, који могу бити корпе с воћем и поврћем, флаше, посуђе, нешто друго или све наведено. Тиме је добијен скуп од укупно 1203 слике. Ради увећања броја слика, употребљена је аугментација сваке слике – слике су окренуте око у-осе. Друге врсте аугментације не би имале смисла, јер се не сусрећемо са сликама где је сто, на пример, на плафону. Крајњи скуп слика мртве природе садржи 2406 слика.

За потребе тренирања модела, вршено је додатно трансформисање слика:

- димензије сваке слике су сведене на 96*96 пиксела. У неким сличним решењима, аутори користе димензије 512*512, али ово би доста успорило време тренирања. Такође, јављају се и димензије 32*32, што је процењено као превише мало да би се разазнали облици на слици.
- слике су нормализоване на начин да се вредности пиксела налазе у опсегу од -1 до 1.

За потребе CGAN модела креиран је скуп података који узима следеће врсте слика:

- portraits – 4117 слика, лабела 1

- flowers – 1800 слика, лабела 2
- religious paintings. - 8400 слика, лабела 3.

Добија се небалансиран скуп, где највише има религијских слика, а најмање слика цвећа. Ово би могло да доведе до тога да модел не генерише све врсте подједнако добро, или би могло допринети лакшем извлачењу карактеристика било које категорије. Редослед слика је насумично измешан.

IV. ОПИС ПРЕДЛОЖЕНОГ РЕШЕЊА

Током развоја предложеног решења, испитани су *Generative Adversarial Networks* алгоритми (GAN). Ови алгоритми су одабрани јер се често обрађују у радовима на сличну тему. Укратко, GAN-ови се састоје од две неуронске мреже – генератора и дискриминатора, који се такмиче један против другог. Циљ генератора је да што реалистичније изгенерише слику и пређе дискриминатора, а дискриминатор тежи да што боље разлучује између правих и лажних слика. У наставку поглавља биће описана архитектура решења и неки савети који су примењени за побољшање модела.

А. Архитектура решења

Испитани су следећи алгоритми:

1) DCGAN – Deep Convolutional GAN

Испробане су две архитектуре DCGAN-а. Генератор првог модела има следећу структуру: густе (*Dense*) слој који прима улазни шум, праћен *LeakyReLU* слојем. Затим следе 4 слоја *upsample*-овања уз нормализацију *batch*-ева (*Batch Normalization*), праћено *LeakyReLU* солјевима. Излаз представља конволуциони слој који прати димензије генерисане слике: (96, 96, 3) и користи *tanh* активациону функцију. Дискриминатор првог модела има конволуциони улазни слој који прима слику горе наведених димензија. Овај слој је праћен *LeakyReLU* и *Dropout* солјевима. Затим следе четири конволуциона слоја уз *Batch Normalization*, *LeakyReLU* и *Dropout* слојеве, где *alpha* има вредност 0.25 – игнорише се 25% насумичних неурона. Крајњи слој је густе слој са *sigmoid* активационом функцијом. Овај модел ће се даље звати DCGAN1.

Генератор другог модела има сличну структуру претходној, а за *upsample*-овање користи *Conv2DTranspose*, по саветима описаним у наставку поглавља. Такође, конволутивни слојеви у овом случају не користе *bias*. Дискриминатор другог модела је сличан првом, а разликује се по томе што не користи *Batch Normalization*, а код *Dropout* солјева *alpha* има вредност 0.3. Овај модел ће се даље звати DCGAN2.

2) CGAN – Conditional GAN

Као помоћ при развоју овог модела употребљен је код наведен у [12]. Архитектура је затим прилагођена тако да одговара скупу података који се користи у овом раду:

- димензије генерисаних слика 96*96, уместо 32*32, што се види у самом облику слојева. Због овога је уведен додатни слој *upsample*-овања у генератору
- 3 категорије слика.

Укратко описано, дискриминатор садржи два улазна слоја, један за слике, други за лабеле. Ови улази се конкатенирају, а затим следе два конволутивна слоја уз LeakyReLU. Пре излазног слоја врши се dropout од 40%, а последњи густи слој користи *sigmoid* активациону функцију. Генератор, поред улазног слоја који прима шум, садржи и *Embedded* слој за лабеле. Након тога следи један густи слој и LeakyReLU, након чега се то конкатенира са улазом за лабеле. Затим следе три слоја повећавања димензија праћених LeakyReLU слојевима, а крајњи слој са димензијама $96 \times 96 \times 3$ користи *tanh* активациону функцију. Генератор и дискриминатор се затим спајају у један модел.

V. Примена савета за побољшање перформанси

Не постоје тачно одређене формуле које сигурно гарантују добре перформансе генеративних модела. Аутори у [11] нуде преглед испробаних трикова који, по њима, често дају добре резултате. У овом раду, примењене су следеће смернице, које су се заиста показале као побољшање:

- нормализовање слика у опсег од -1 до 1
- употреба *Tanh* као последњег слоја генератора
- креирање насумичног шума за генератор из Гаусове расподеле, уместо из униформне
- раздвајање стварних и лажних слика у различите скупове приликом тренирања дискриминатора, уместо њиховог комбиновања у један скуп. За сваки скуп се рачуна посебна *loss* функција.
- употреба *LeakyReLU* уместо *ReLU*, ккао за генератор, тако и за дискриминатор
- употреба DCGAN-а
- употреба *ConvTranspose2d + stride* за *upsampling*, што није увело промене у односу на претходно коришћени *UpSampling2D*
- коришћење *Adam* оптимизатора.

Оно што је испробано, а није допринело побољшању јесте замена лабела приликом тренирања генератора: *real=fake*, *fake=real*. Ово је резултовало да се увек генеришу потпуно црне слике.

V. РЕЗУЛТАТИ И ДИСКУСИЈА

У овом поглављу биће описано тренирање, евалуација модела, а такође анализираће се и грешке модела.

A. Тренирање модела

Предложена решења овог рада развијала су се и тренирала на *Google Colab*-у, уз употребу GPU акцелератора. Ово је доста убрзало читав процес, међутим наилази се на нека временска ограничења употребе, у смислу да је дозвољено укупно 12 сати рада, а након тога неколико дана није дозвољено коришћење овог убрзања. Скупови података смештени су на *Google Drive*, који опет има ограничења меморије.

Поступак тренирања конструисан је тако да генератор изгенерише слике на основу рандом шума. Затим се дискриминатор посебно тренира на реалним сликама, посебно на лажним. Рачунају се градијенти генератора и дискриминатора и ажурирају се њихови оптимизатори.

Напредак генератора и дискриминатора је праћен кроз бележење *loss* функција сваке епохе, као и чување неколико изгенерисаних слика сваке епохе. Број епоха није унапред одређен, већ је тренирање бивало терминирано када се у изгенерисаним сликама не види напредак. Максимални број епоха за све моделе је 200.

B. Евалуација модела

За евалуацију модела коришћена је лична процена, по угледу на нека од претходних решења. Аутори у радовима на сличну тему процењују моделе и помоћу већег броја добровољаца који се индивидуално опредељују за модел чије слике им се више допадају. На сликама испод приказани су резултати сваког од модела.



Illustration 1: Пример генерисања слика портрета, цвећа и религијских слика редом на основу категорија помоћу CGAN, након 50 епоха



Illustration 2: Пример генерисаних портрета DCGAN1 након 110 епоха

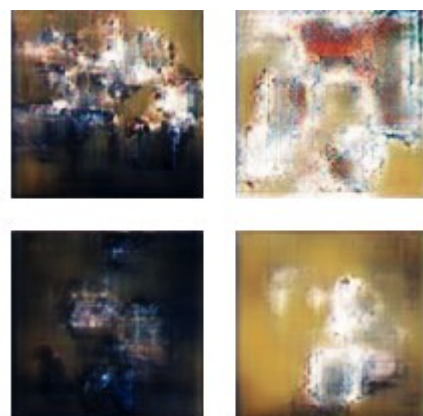


Illustration 3: Пример генерисаних слика цвећа DCGAN1 након 90 епоха



Illustration 4: Пример генерисаних портрета DCGAN2 након 200 епоха



Illustration 6: Пример генерисаних слика мртве природе након 100 епоха

С. Анализа грешака

Поставља се питање да ли је у уметности могуће погрешити и шта би се ту сматрало грешком. Са становишта машинског учења и конструкције модела за генерисање уметничких слика, грешка би за овакве моделе била ако на основу слика људи генерише, на пример, цвеће. Такође, грешка би могла бити нејасна, веома замућена слика уколико то није оно што треба да се генерише. Ово не би био случај када су у питању апстрактне слике, које не садрже јасне облике. Дакле, све што се у великој мери разликује од онога што се очекује, могло би се сматрати грешком.

Грешке које модел прави су очекиване и могу се утврдити и објаснити разлози који до њих доводе. Што се тиче CGAN-а, грешке се могу објаснити небалансираним скупом, а код свих модела могући разлог грешке јесте скуп

података који садржи веома различите слике, иако припадају истој категорији. Ово би се могло побољшати додатним преуређивањем скупова података.

VI. ЗАКЉУЧАК И БУДУЋИ РАД

Размотрен је проблем генерисања уметничких слика. Представљени су и неки сродни радови који нуде решење на сличну тему и предложено је решење које обухвата GAN технологије. Укратко је описан скуп података и начин на који је обрађен и редукован. Испитано је неколико модела, намештање њихових хипер параметара и примена трикова за побољшање перформанси генеративних модела. Резултати модела су евалуирани помоћу личне процене и приказано је неколико генерисаних слика.

Може се закључити да су се кондиционални CGAN модели засновани на категоријама лошије показали, што може бити резултат небалансираног скупа. Са друге стране, модели који су фокусирани на само једну категорију (DCGAN) генеришу много боље слике.

Приликом поређења перформанси предложеног решења са претходним решењима уочава се да се добијају слике нешто лошијег квалитета и не подсећају толико на реалне слике. Међутим, у уметности нема правила и могло би се тврдити да модели који мало одступају од реалних уметничких слика заправо уносе креативност која је својствена само њима.

Даља истраживања би укључила додатна препроцесирања слика, аугментација у циљу креирања већег скупа података, као и испитивање других архитектура модела и намештање њихових параметара, а додатно би се испитало спајање засебних модела, фокусираних на једну категорију у један модел. Такође, испитало би се генерисање других скупова слика, попут слика одређеног уметничког правца, слике одређеног уметника и сл.

РЕФЕРЕНЦЕ

- [1] Xue, Alice. "End-to-end Chinese landscape painting creation using generative adversarial networks." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021.
- [2] Tan, Wei Ren, et al. "ArtGAN: Artwork synthesis with conditional categorical GANs." *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.
- [3] Jones, Kenny. "GANGogh: Creating Art with GANs." *Towards Data Science*, 18 June 2017, <https://towardsdatascience.com/gangogh-creating-art-with-gans-8d087d8f74a1>. Преузето јула 2022.
- [4] Saleh, Babak, and Ahmed Elgammal. "Large-scale classification of fine-art paintings: Learning the right metric on the right feature." *arXiv preprint arXiv:1505.00855* (2015).
- [5] Xue, Alice. "Traditional Chinese Landscape Painting Dataset" GitHub, 2020, <https://github.com/alice2020/Chinese-Landscape-Painting-Dataset>. Преузето јула 2022.
- [6] Shahriar, Sakib. "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." *Displays* (2022): 102237.
- [7] Elgammal, Ahmed, et al. "Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms." *arXiv preprint arXiv:1706.07068* (2017).

- [8] WikiArt dataset - <https://www.wikiart.org/> Презето јула 2022
- [9] Kaggle Wiki-Art: Visual Art Encyclopedia - <https://www.kaggle.com/datasets/ipythonx/wikiart-gangogh-creating-art-gan?resource=download>, Презето јула 2022.
- [10] Kaggle Art Portraits – <https://www.kaggle.com/datasets/karnikakapoor/art-portraits?select=Portraits>, Презето јула 2022.
- [11] Chintala, Soumith, et al. “How to Train a GAN? Tips and Tricks to Make GANs Work.” *GitHub*, 2020, <https://github.com/soumith/ganhacks>. Презето јула 2022.
- [12] Bhattiprolu, Sreenivas. *GitHub*, 2021, https://github.com/bnsreenu/python_for_microscopists/blob/master/249_keras_implementation_of_conditional_GAN/249-cifar_conditional_GAN.py, Презето јула 2022.