

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Suparnički generativni modeli za prevođenje slika

Krešimir Vukić

Voditelj: *prof. dr. sc. Siniša Šegvić*

Zagreb, lipanj 2018.

SADRŽAJ

1. Uvod	1
2. Duboke mreže	2
2.1. Konvolucija	2
2.2. Primjer konvolucijske mreže	3
3. Prijenos stila	4
3.1. Konvolucijski pristup	4
3.2. Očuvanje strukture	5
3.3. Lokalizacija stila i očuvanje konteksta	5
4. Generativno modeliranje	6
4.1. Generativni naspram diskriminativnih algoritama	6
4.2. Suparnički modeli	8
4.3. Rad generativnih suparničkih modela	8
4.4. GAN-ovi naspram autoenkodera	10
5. Uvjetne generativne mreže	12
5.1. Generalizacija problema prevođenja slike	13
5.2. Pristup uvjetovanog prevođenja slike	13
5.3. Funkcija gubitka	14
6. Translacija bez uparenih primjera za učenje	15
6.1. Kružni GAN-ovi	16
6.1.1. Gubitak cikličke dosljednosti	17
6.2. Unimodalna nenadzirana translacija (UNIT)	18
6.3. Multimodalna nenadzirana translacija (MUNIT)	19
7. Zaključak	20

1. Uvod

Razvojem tehnika namijenjenih identifikaciji i analizi slika te rapidnim razvojem neuronskih mreža, logični sljedeći korak unutar područja računalnog vida bio je osmišljavanje algoritama kojima bi se te tehnologije proširile na reprodukciju naučenih vizualnih sadržaja. Prijenos stilova, rekonstrukcija fotografije, generiranje tekstura ili simulacija interakcije molekula neke su od primjena dotičnog problema.

Najznačajniji pomak u tome području ostvarili su Goodfellow et al. 2014. g. [5] osmišljavanjem generativnih suparničkih modela. Tipičan cilj mreža baziranih na tom modelu je pronalazak funkcije preslikavanja iz prostora latentnih varijabli na željenu distribuciju, a kojom se teško rješivi matematički problemi svode na jednostavno zadavanje željenih ulaza rezultatne funkcije.

Dobivena distribucija, osim za generiranje novih primjera, primjenjiva je i na rješavanje prvotno navedenih problema identifikacije značajki na postojećim slikama. Ta fleksibilnost čini ovaj algoritam izuzetno popularnim za istraživanje što je rezultiralo njegovim mnogobrojnim modifikacijama i poboljšanjima.

S obzirom na to da se većina generativnih modela za generiranje slika bazira na dubokim konvolucijskim mrežama, prvo poglavlje počinje kratkim uvidom u njihov rad te se zatim razmatra problem prijenosa stila i ciljevi koje njime želimo ostvariti. U ostatku rada istražit će se različita proširenja kojima se generativni suparnički modeli prilagođavaju specifičnim problemima, njihove karakteristike i primjene.

2. Duboke mreže

Posljednjih godina, duboko učenje velikim je dijelom zamijenilo klasične pristupe rješavanja problema na području umjetne inteligencije. Tu popularnost ostvarilo je zahvaljujući većem stupnju samostalnosti naspram klasičnih pristupa. Prednost koju joj to pruža je ponajviše očita pri radu s velikim količinama podataka, a uz trendove računarstva koji idu prema ubrzanom gomilanju svih dostupnih informacija, ona će samo rasti.

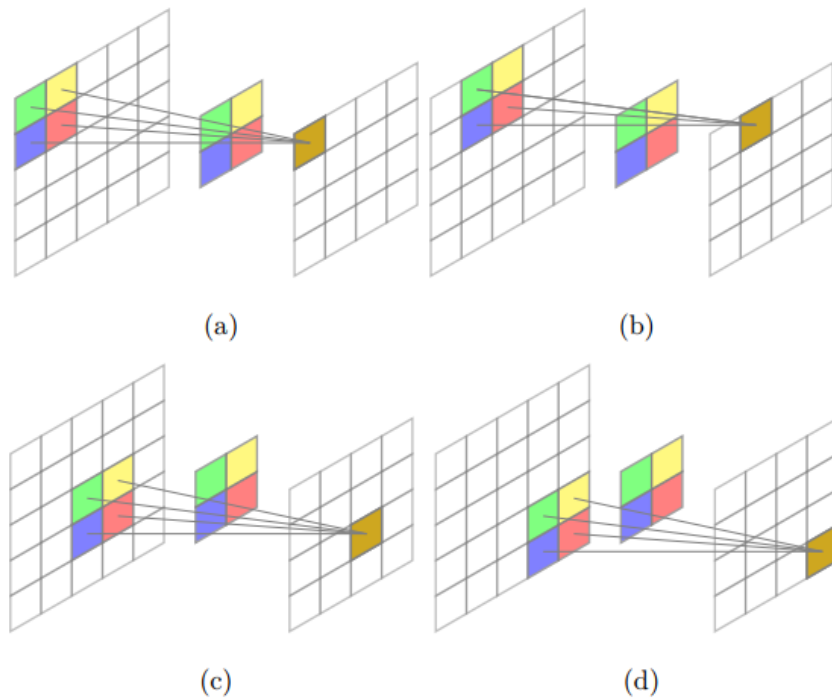
Dok je starije algoritme strojnog učenja potrebno modificirati ručnom identifikacijom značajki kako bi se primijenili na nove skupove podataka, duboke mreže na temelju dostupnih podataka samostalno odluče koje su značajke bitne.

Osim toga, slojevit pristupu rastavljanja podataka na jednostavnije građevne jedinice čini ih posebice pogodnima za rad s problemima hijerarhijske prirode poput slika.

2.1. Konvolucija

Pri radu sa slikama, konvolucijske mreže su jedna od najznačajnijih podvrsta dubokih mreža. Osmišljene s ciljem dohvata lokalnih relacija među podacima, rješavaju problem nepoželjnog utjecaja pozadine te lokacije značajke unutar slike.

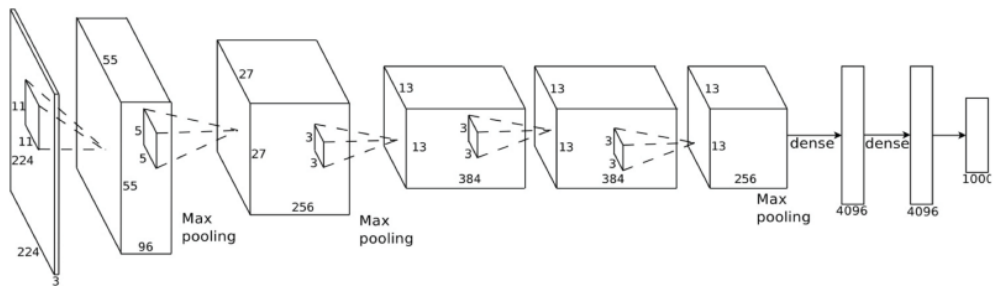
Ideja konvolucije je da se kao ulaz u neurone sljedećeg sloja dovedu isključivo okolne vrijednosti neurona s prethodnog. To se odvija tako da se unaprijed definirano kvadratno područje(filter) pomiče (konvoluirá) po ulaznim podacima [slika 2.1] pri čemu množimo vrijednosti filtera s pripadnim ulaznim vrijednostima. Rezultantne vrijednosti filtera predstavljaju reprezentaciju određene značajke i omogućuju njezinu jednostavnu identifikaciju.



Slika 2.1: Konvolucija matrice

2.2. Primjer konvolucijske mreže

Nizanjem konvolucijskih slojeva dobivamo arhitekturu u kojoj se odgovornost filtera raspoređuje tako da je svaki sloj zadužen za jednostavniji potproblem sljedećeg sloja. Prvi slojevi uče atomarne značajke poput linija, boja ili orijentacije čijim je kombiniranjem moguće raspoznati sve složenije geometrijske oblike. Na slici 2.2 prikazan je primjer duboke neuronske mreže s 5 konvolucijskih slojeva među kojima se odvija sažimanje maksimalnom vrijednošću te dva sloja potpune povezanosti.



Slika 2.2: Primjer duboke neuronske mreže

3. Prijenos stila

Metode prijenosa stila bave se modificiranjem stilističke perspektive slike uz očuvanje sadržaja. Generalno te procese možemo podijeliti na prijenos temeljem jednog primjera te na prijenos zajedničkog stila neke kolekcije. Dok se klasični pristupi baziraju na prvoj skupini, neuronske mreže bolje rezultate postižu na drugoj.

3.1. Konvolucijski pristup

Korištenje konvolucijskih mreža (CNN) za prijenos značajki oslanja se na činjenicu da se konvolucijski slojevi specijaliziraju za određene značajke. Pošto pamte značajke u obliku aktivacijskih mapa po slojevima, lako je izolirati karakteristike koje želimo prenijeti na novu sliku.

Za prijenos stila formulira se optimizacijski problem koji, počevši od nasumičnog šuma, radi izmjene kojima istovremeno minimizira odstupanja neuronskih aktivacija od osnovne slike te od značajki koje se prenose. Nažalost, CNN-ovi imaju dvije nepopravljive mane:

1. Ograničeni su izborom stilova koje mogu prenijeti jer je svaka mreža specijalizirana za isključivo jedan.
2. Pošto su neuronske mreže vrlo samostalne po pitanju kako i koje značajke uče, svaki sloj uz stilističke značajke sadrži i neke strukturne. Zbog toga rezultatna slika sadrži distorzije u strukturnom sadržaju što ju čini nerealističnom.

3.2. Očuvanje strukture

Sitne izmjene u linijama i teksturama koje CNN-ovi vrše pri prijenosu stila posljedica je toga što izravno mijenjaju vrijednosti piksela. Ako ne želimo izgubiti originalni sadržaj fotografije, dovoljno je postaviti uvjet na rad mreže da izmjene može unositi isključivo na prostoru boja. No to nam još uvijek ne garantira da će fotografija zadržati realističnost.

3.3. Lokalizacija stila i očuvanje konteksta

Postoji još jedan faktor koji povećava kompleksnost cijelog procesa: ciljano ponašanje prijenosa stila trebalo bi se držati konteksta izmjene. Na primjer, ako se na fotografiji nalaze livada i oblaci, modifikacije nad livadom moraju biti povezane sa stilom livade na referentnoj slici, a ne da povuku stil s oblaka. Za CNN, postoje radovi koji to uspješno rade usmjeravanjem procesa temeljem semantičke segmentacije [13]. Osim što ovise o zasebnim mrežama koje bi segmentirale sadržaj, potrebno je i ručno obraditi set raspoloživih labela.

4. Generativno modeliranje

4.1. Generativni naspram diskriminativnih algoritama

Jedna od podjela algoritama za učenje dijeli ih na dvije temeljne skupine – diskriminativne i generativne:

- Diskriminativni modeli baziraju se na modeliranju isključivo pomoću ulaznih varijabli (X) koje direktno mapiraju na ciljne vrijednosti (Y) i time uče uvjetnu vjerojatnost $p(Y|X)$. Pošto se fokusiraju samo na mapiranje viđenih značajki u klase, ne stvaraju puno pretpostavki o podložnoj distribuciji, ali su vrlo ovisni o kvalitetnim podacima.
- Generativni modeli rješavaju inverzan problem – kako dobiti neku značajku X kada znamo kojoj klasi Y pripada, tj. pokušavaju naučiti distribuciju podataka pojedinih klasa $p(X|Y)$. Kada imamo dovoljno veliku količinu dobrih podataka, rezultat tog učenja je distribucija $p(X)$.

Drugim riječima, dok diskriminativni modeli pokušavaju pronaći granicu između klasa, generativni modeliraju distribucije pojedinih klasa na temelju kojih mogu uzorkovati nove primjere.

Kako bi bili u stanju vršiti generativni proces, generativni modeli implicitno obavljaju još jednu radnju – ekstrakciju bitnih značajki. Pri tome teže da te značajke ne budu samo parametri kojima podešavaju mrežu nego da predstavljaju 'prirodne' karakteristike podataka.

Ako cilj diskriminativnih modela predstavimo Bayesovom formulom:

$$p(Y|X) = \frac{p(Y)p(X|Y)}{p(X)} \quad (4.1)$$

vidljivo je da poznavanje $p(X)$ možemo iskoristiti za poboljšanje rada diskriminativnih modela.

4.2. Suparnički modeli

Iako je očito da su generativni algoritmi svestranije naravi od diskriminativnih, u praksi se pokazalo da uz ručnu optimizaciju diskriminativni modeli postižu bolje rezultate od generativnih. Razlog tome je što sposobnost generalizacije dolazi s neizbježnim povećanjem složenosti, pri čemu najveću zapreku predstavlja pronalazak ciljne distribucije u dubokim hijerarhijskim podacima.

Naime, rješenja problema od kojih se ona sastoji matematički nije moguće izvesti [5] pa je jedina mogućnost pronalazak dovoljno slične distribucije. Približna rješenja dobivena su uvođenjem mnogobrojnih aproksimacija, no nažalost, posljedica toga je da optimum dobivenog modela nikad neće biti dobar kao što bi bio onaj s pravim rješenjem.

Alternativni pristup ručnoj optimizaciji ponuđen je u obliku suparničkih modela – probleme kojima ljudi ne mogu pronaći rješenje, kroz dovoljno velik broj iteracija, moguće je naučiti pomoću dva modela koji se natječu u traženju njegove čim bolje aproksimacije dok ne pogode pravo rješenje. Ovisno o implementaciji, mreža takve arhitekture u stanju je izvesti distribuciju širokog spektra podataka: slika, govora, muzike, pa čak i pisanih zapisa.

4.3. Rad generativnih suparničkih modela

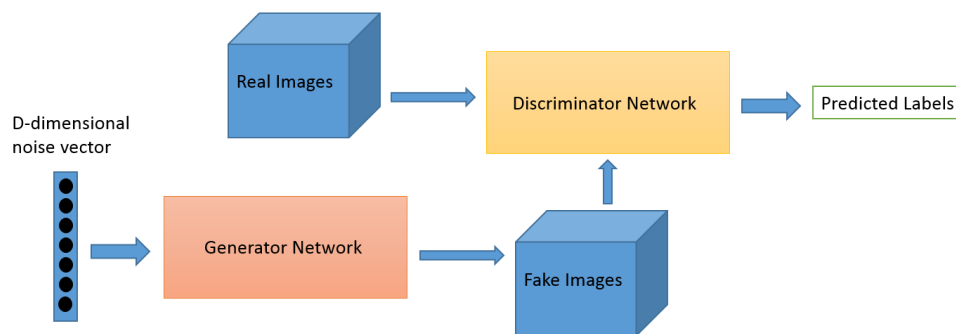
Ideja generativnih suparničkih modela bazirana je na učenju mreža kroz minimax igru – umjesto izravnog rješavanja optimizacijskog problema, GAN sadrži funkciju kojoj dvije mreže paralelno podešavaju ulaze dok ne dođu do stabilne točke u odnosu na maksimum jedne, a minimum strategije druge mreže.

Pola mreže (generator G) bavi se stvaranjem novih podataka iz neke jednostavne distribucije Z s ciljem da stvoreni podatci budu čim sličniji stvarnima.

Paralelno s njegovim radom, druga polovica (diskriminator D) pokušava evaluirati koliko su uzorci $G(Z)$ slični stvarnim podacima Y . Drugim riječima, dok G pokušava minimizirati izlaz $D - a$, D ga pokušava maksimizirati.

Koraci rada GAN-a:

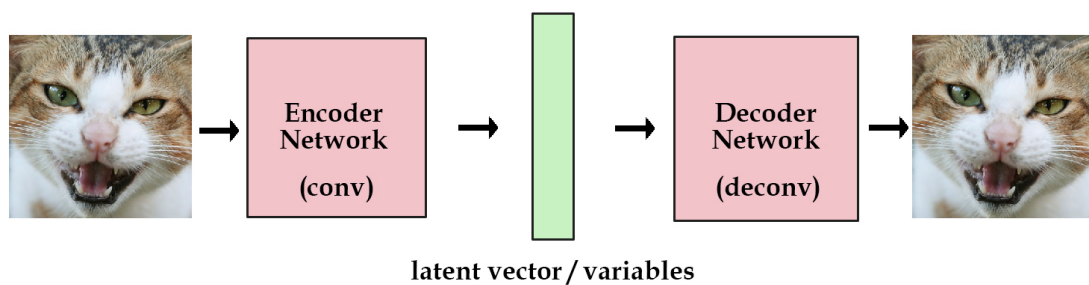
1. Generator G iz nasumičnih brojeva (šum) Z stvara sliku $G(Z)$
2. Ta slika se zajedno s uzorcima skupa stvarnih podataka Y dovodi na ulaz diskriminatora D
3. Diskriminator D određuje postotak vjerojatnosti da je slika $G(Z)$ prava



Slika 4.1: GAN [4]

4.4. GAN-ovi naspram autoenkodera

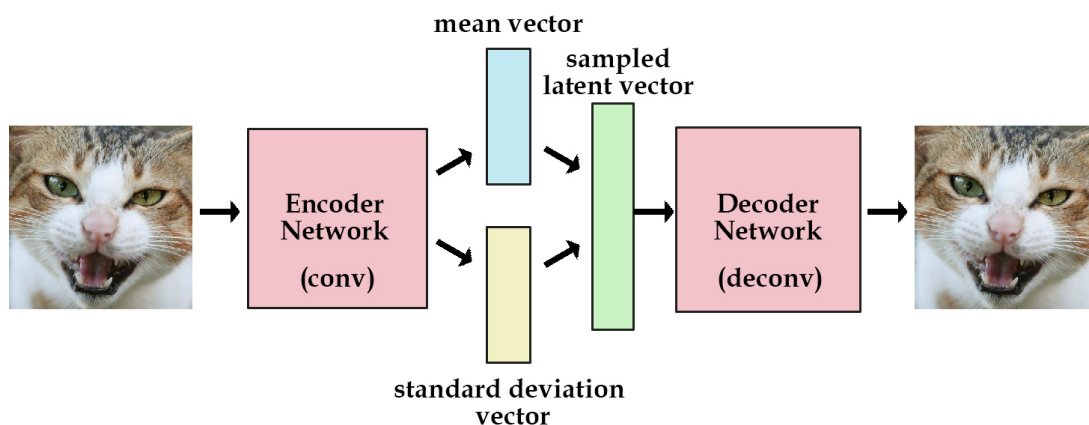
Uz GAN-ove, podjednako važna skupina generativnih modela su autoenkoderi. Ideja iza njihovog rada je da se sastoje od enkodera koji kompresira podatke u latentne varijable te dekodera koji iz tih varijabli ponovo generira ulaznu sliku. Primarni cilj takve mreže je postizanje čim veće sličnosti ulaza i njegove kopije na izlazu. Kako je primarni cilj vrlo jednostavan, autoenkoderi se treniraju bez nadzora. Najzanimljiviji dio ove arhitekture je skriveni sloj u kojem se nalaze latentne varijable, a koje predstavljaju najbitnije značajke uzoraka za treniranje.



Slika 4.2: Autoenkoder [2]

Dobiveni latentni vektor sažeta je reprezentacija podataka za učenje no iako je koristan za pohranu, nije iskoristiv za generiranje novih, izmijenjenih podataka.

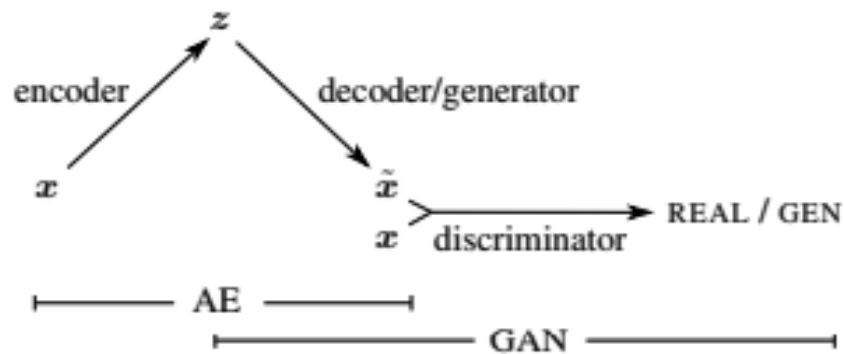
Da bi to omogućili, clij mreže mijenjamo da umjesto latentnih vektora pokušava naučiti Gausovu distribuciju iz koje je moguće uzorkovati podatke. Takva mreža naziva se variacijski autoenkoder (VAE).



Slika 4.3: Variacijski autoenkoder [2]

Latentne vektore iz naučenih varijacijskih enkodera dobivamo uzorkovanjem dobivene distribucije te tako možemo vrlo precizno zadati kakve podatke želimo generirati. To im je ujedno i glavna prednost nad GAN-ovima koji generiraju dosta nasumične uzorke – što se može popraviti uz izmjene koje će biti obrađene u sljedećim poglavljima.

Negativna strana varijacijskih autoenkodera je što uzorci koje proizvode imaju dosta anomalija. Na slikama se to izgleda tako da više nalikuju na umjetnička djela nego realnu reprezentaciju. Zbog toga primarnu primjenu nalaze u industrijama poput video igara i kompresiji podataka. U istraživanju su i mreže koje ujedinjuju te dvije tehnologije trenirajući dekodler pomoću GAN-a [11].

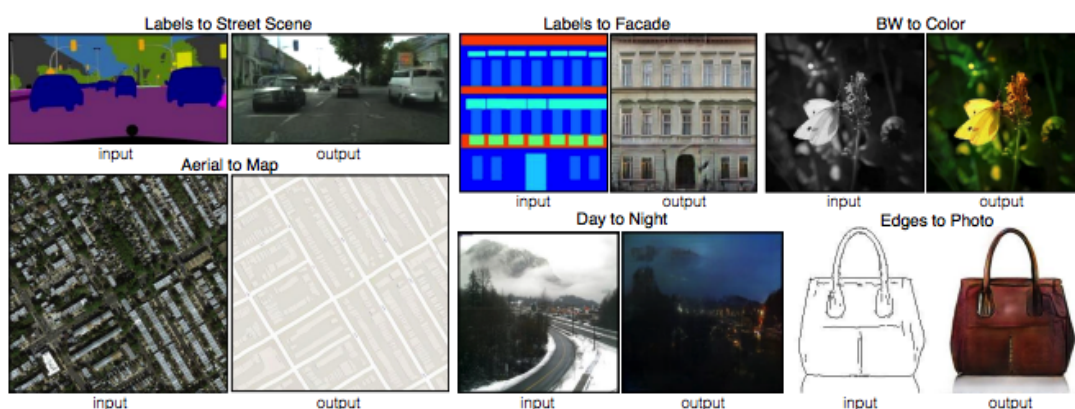


Slika 4.4: Primjer kombinacije autoenkodera i GAN-a [11]

5. Uvjetne generativne mreže

Kao što je spomenuto u prošlom poglavlju, jedna od većih mana generativnih suparničkih mreža je što generiraju poprilično nasumične uzorke. Pošto većina problema u praksi zahtijeva kontrolu nad tim procesom, jedno od najranijih proširenja GAN-ova vezano je upravo uz kontrolu generiranja podataka.

Postavljanjem uvjeta na ulaz generativnog modela, moguće je usmjeriti generativni proces. Vrste uvjeta mogu biti raznolike: oznake klasa, skice, dubinske mape pa čak i cijele slike.



Slika 5.1: Primjeri uvjetovane generacije [9]

5.1. Generalizacija problema prevođenja slike

Jedan način razmatranja slika je kao višeslojni podatak u kojem se osnovni sadržaj obogaćuje raznolikim stilovima i načinima reprezentacije. Jednako kao što riječ možemo prevesti na više jezika zadržavajući određeni kontekst, ista scena može se prikazati kao dubinska mapa, RGB slika ili semantička segmentacija.

Sukladno toj ideji, mnoge probleme procesiranja slika moguće je svesti na translaciju slike u drugačiju reprezentaciju. Neovisno radi li se o uklanjanju anomalija, ispravku boja ili izmjeni pozadine, u srži tih procesa može se prepoznati jedinstven problem: pronaći funkciju preslikavanja piksela u piksel. Upravo ta ideja potaknula je Isolu et al. da osmisle uvjetne generativne mreže kao generalno rješenje problema prevođenja slike [9].

5.2. Pristup uvjetovanog prevođenja slike

Zadavanje cilja neuronskim mrežama, a time i uspjeh njihovog učenja u potpunosti je bazirano na formulaciji funkcije gubitka. Lošim definiranjem te funkcije, neovisno o količini raspoloživih resursa i tipu mreže, nikad nećemo dobiti realistične rezultate. U GAN-ovima je ona zadana dosta općenito – jesu li generirani podaci različiti od stvarnih uzoraka. Takva definicija omogućuje joj da se prilagodi podacima te kao takva omogućuje generalizaciju nad setovima problema koji bi u suprotnom zahtijevali vrlo drugačije formulacije gubitka.

Dok standardni GAN-ovi odbacuju sva rješenja koja se mogu razlikovati od stvarnosti generirajući slike y iz nasumičnog šuma z : $G : z \rightarrow y$, uvjetovani GAN-ovi dodatno usmjeravaju dobra rješenja prema željenoj distribuciji značajki dodavanjem uvjeta x na šum z : $G : x, z \rightarrow y$ [9].

5.3. Funkcija gubitka

Kako su GAN-ovi sastavljeni od dvije mreže, za njihovo učenje potrebne su zasebne definicije funkcije gubitka. U slučaju neuvjetovanog GAN-a, generator G ima cilj minimizacije $\log(1 - D(G(z)))$ s obzirom na diskriminator D koji pokušava minimizirati $\log(D(x))$. Zajednička funkcija gubitka

$$\begin{aligned}\mathcal{L}_{GAN}(G, D) = & \sum_y [\log D(y)] \\ & + \sum_{x,z} [\log(1 - D(G(x, z)))]\end{aligned}\tag{5.1}$$

koristi se za učenje kroz cilj mini-max igre:

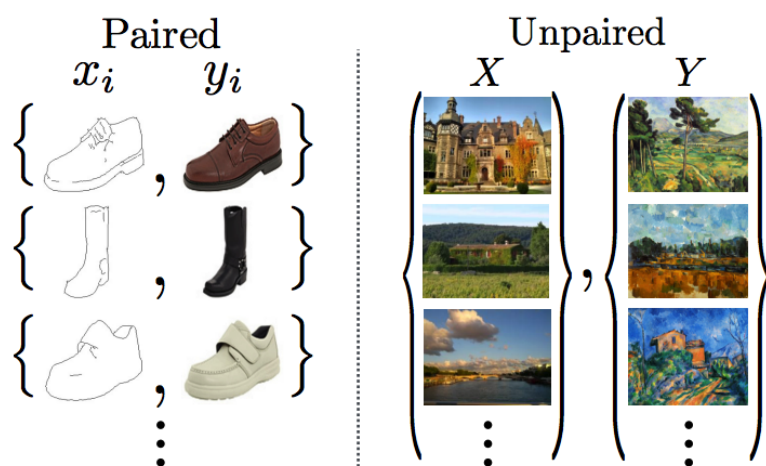
$$G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D)\tag{5.2}$$

Uvjetovanjem dodatnim informacijama y na obje mreže taj cilj mijenja se u :

$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) = & \sum_{x,y} [\log D(x, y)] \\ & + \sum_{x,z} [\log(1 - D(x, G(x, z)))]\end{aligned}\tag{5.3}$$

6. Translacija bez uparenih primjera za učenje

Učenje GAN-ova iz prethodnih poglavlja, kao i mnogih drugih algoritama, ovisi o postojanju velikih podatkovnih skupova označenih parova slika. S obzirom na to da izrada tih skupova iziskuje veliku količinu vremena i manualnog ljudskog rada, u mnogim slučajevima nam oni nisu dostupni.



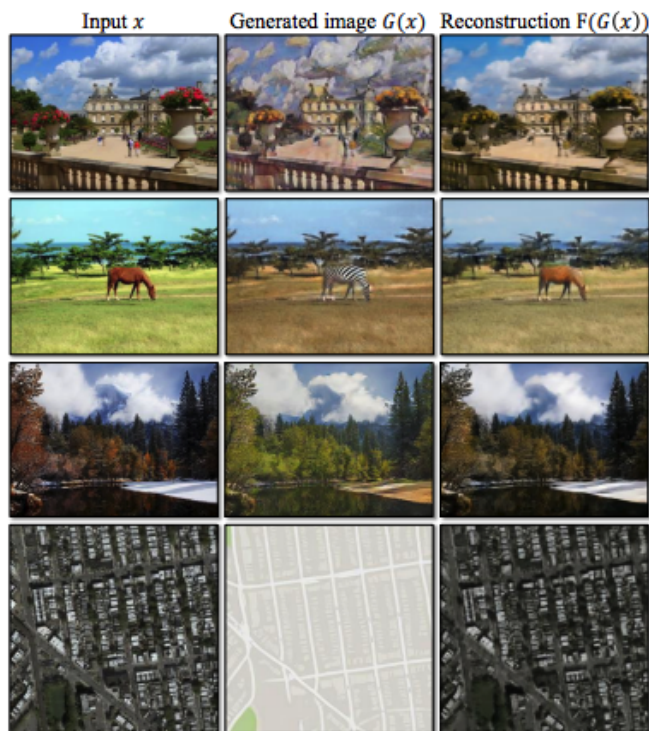
Slika 6.1: Parovi podataka (lijevo) $\{x_i, y_i\}_{i=1}^N$ te neupareni setovi (desno) [15]

6.1. Kružni GAN-ovi

Iako učenje bez označenih primjera izgleda kao poprilično drugačiji problem nego kad su prisutni, osnovni cilj ostaje isti – naučiti preslikavanje s izvorišne domene (X) na ciljnu (Y): $G : X \rightarrow Y$ takvo da nije moguće razlikovati distribuciju $G(X)$ od distribucije Y .

Kompleksnost proizlazi iz toga što je taj cilj vrlo nespecifičan, a bez primjera koji bi pružali smjernice tijekom učenja rezultati su nekonzistentni i previše raznoliki da bi bili primjenjivi.

Kružni GAN-ovi zato uvode uvjet tranzitivnosti nad ciljnom funkcijom kao način regularizacije u svrhu provođenja dosljednosti ponašanja modela. Definirajući inverz funkcije G : $F : Y \rightarrow X$, funkciju gubitka osnovnog GAN-a proširuju gubitkom cikličke dosljednosti: $F(G(X)) \approx X$.



Slika 6.2: Primjeri nenadziranog GAN-a s i bez rekonstrukcije [15]

Oslanjajući se na pretpostavku da izvorišna i ciljna domena nisu u potpunosti nasumične nego sadrže podložnu zajedničku distribuciju, inverzom pretvaraju ovaj problem u nadzirano učenje. Pri tome umjesto značajki iz parova primjera, uče zajedničke specifičnosti cijelih setova te njih koriste za nadziranje.

6.1.1. Gubitak cikličke dosljednosti

Iako suparničke mreže u teoriji mogu naučiti preslikavanja G i F koja stvaraju uzorke jednake distribucije kao Y i X , u praksi često nauče samo preslikavanje na podskup uzoraka ciljne domene. Iz tog razloga, gubitak suparničke mreže

$$\begin{aligned}\mathcal{L}_{GAN}(G, D) = & \sum_{y \sim p_{data}(y)} [\log D(y)] \\ & + \sum_{x \sim p_{data}(x)} [\log(1 - D(G(x)))]\end{aligned}\tag{6.1}$$

nije dovoljno specifičan da osigura željenu translaciju slike.

Kako bi smanjili prostor mogućih funkcija preslikavanja, cikličkom dosljednošću forsiramo da dvosmjerni ciklus translacije mora izvornu sliku $x \in X$ preko translirane $y \in Y$ preslikati nazad u izvornu reprezentaciju: $x \rightarrow G(X) \rightarrow F(G(x)) \approx x$. (Isto vrijedi u obrnutom smjeru.)

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) = & \sum_{x \sim p_{data}(x)} [\| F(G(x)) - x \|] \\ & + \sum_{y \sim p_{data}(y)} [\| G(F(y)) - y \|]\end{aligned}\tag{6.2}$$

Udruživanjem tih gubitaka, cjelokupni gubitak kružnih GAN-ova je:

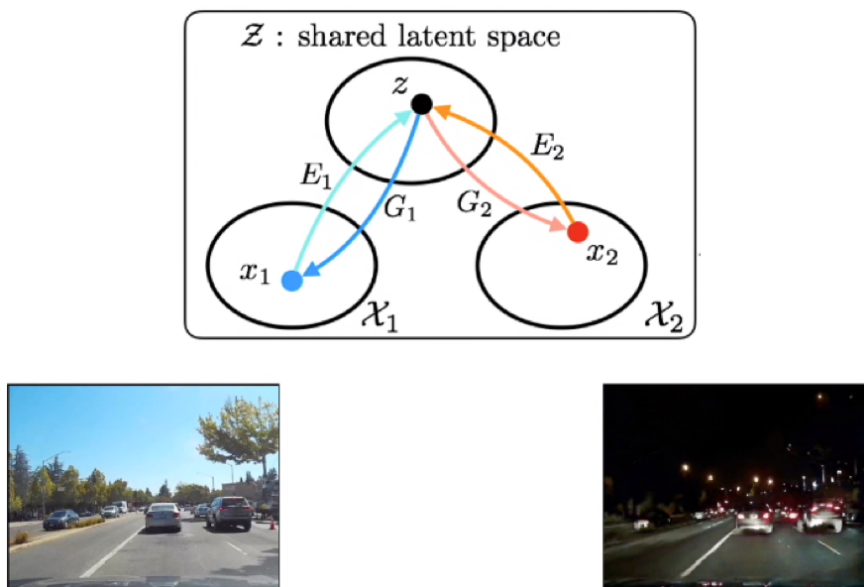
$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{GAN}(G, D_Y) \\ & + \mathcal{L}_{GAN}(F, D_X) \\ & + \lambda \mathcal{L}_{cyc}(G, F)\end{aligned}\tag{6.3}$$

6.2. Unimodalna nenadzirana translacija (UNIT)

Analizirajući translaciju slike s gledišta modeliranja vjerojatnosti, cilj je odrediti združenu distribuciju više različitih domena slika. U slučaju nenadziranog učenja, izvorni i ciljni skupovi uzorkovani su iz marginalnih distribucija (tj. u potpunosti neovisno jedan o drugome), a združenih distribucija koje iz njih možemo izvesti ima beskonačno mnogo.

Skup znanstvenika iz NVIDIA-e je tome pristupio preslikavajući i izvornu i ciljnu domenu na zajednički prostor latentnih varijabli čime su uveli pretpostavku da postoje zajedničke varijable o kojima ovise obje distribucije.

Tehnološki su to ostvarili umrežavanjem GAN-a i variacijskog autoenkodera. GAN stvara latentni prostor dijeljenjem težina generatora i diskriminatora nad kojim variacijski autoenkoder određuje relacije između domena. Osim toga, pokazali su [12] i da dijeljeni latentni prostor implicira cikličku konzistenciju objašnjenu u poglavlju o kružnim GAN-ovima.

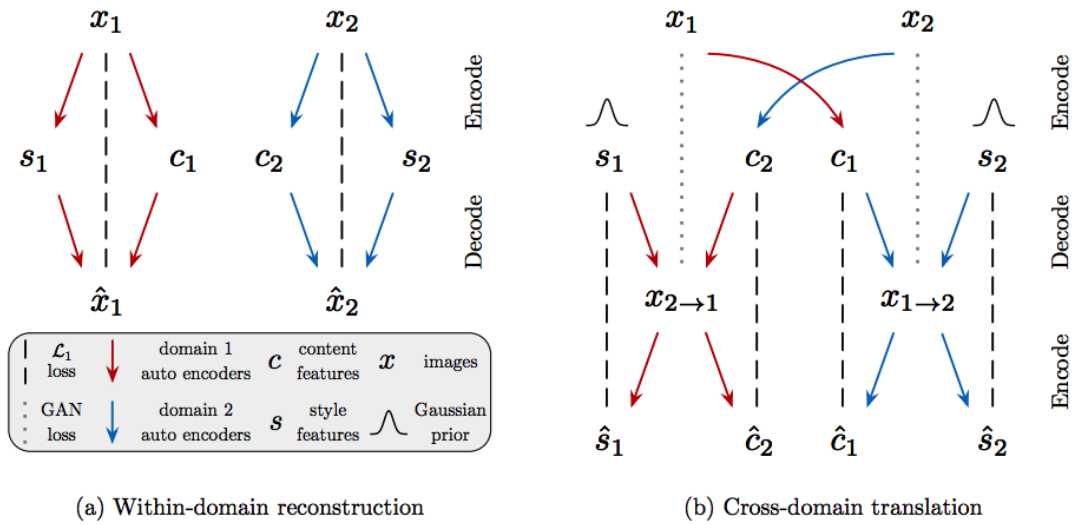


Slika 6.3: Mapping images to latent space [12]

6.3. Multimodalna nenadzirana translacija (MUNIT)

Iako prethodne sekcije opisuju načine nenadziranog učenja, pri radu s njima može se primjetiti da nisu generičko rješenje tog problema. Naime, kružni GAN-ovi se baziraju na pretpostavci da je preslikavanje između domena determinističko zbog čega imaju manu da im nedostaje raznolikosti u translatiranim izlazima, a UNIT modeli uče unimodalno preslikavanje. Ne uzimajući u obzir da je uvjetna distribucija preslikavanja slika između dvije domene inherentno multimodalna [8], imaju posljedicu da pri generiranju vrlo teško stvaraju vidljivo različite uzorke.

MUNIT model pretpostavlja da se reprezentacija slike može rastaviti na značajke koje su domenski invarijantne te na one koje sadrže domenski specifična svojstva. Zatim se pri generiranju jednostavno domenski invarijantne značajke kombiniraju s nasumičnim stilskim kodom uzorkovanim s ciljne domene.



Slika 6.4: a) Rastavljanje uzoraka na sadržajne i stilističke značajke b) koraci pri prijenosu stila MUNIT-om [8]

Pri podjeli reprezentacija sadržaja i stila, sadržaj se odnosi na prostornu strukturu dok je stil ograničen na rendering te strukture.

7. Zaključak

Generativni modeli tema su mnogobrojnih istraživanja i vrlo obećavajuća metoda učenja 'prirodnih' značajki podataka. Uz napretke koji se redovno postižu u njihovom unaprjeđenju, velika je vjerojatnost da će ubrzo biti u stanju generirati uzorke nerazlučive od stvarnosti. Time primjenu pronalaze u područjima poput stvaranja virtualnih svjetova bez ljudske intervencije, izvršavanja modifikacija nad slikama pomoću govornih naredbi, regeneriranju oštećenih podataka ili stvaranju novih za buduća istraživanja.

Uvjetnim suparničkim mrežama pokazano je da osim što više nije potrebno ručno izvoditi funkcije preslikavanja, zadovoljavajuće rezultate možemo dobiti i ako dubokoj mreži prepustimo da sama nauči funkciju gubitka. To je velik korak prema široj adaptaciji ove tehnologije upravo zato što uklanja potrebu za inženjerskim radom i omogućava ljudima poput umjetnika ili znanstvenika iz neprogramerskih područja da ju koriste samostalno.

Kada je pak u pitanju translacija bez dostupnih primjera za učenje, kružni GAN-ovi ostvaruju rezultate usporedne s nadziranim modelima. Osim toga, dovoljno su svestrani da obavljaju zadatke nekolicine drugih mreža: od preoblikovanja objekata, uklanjanja anomalija do kombiniranja više stilova. Mogu čak obavljati i prijenos karakterističnog stila cijelih kolekcija – kao što je načina crtanja Leonarda da Vincija.

Kružni GAN-ovi zakazuju kada su podatci previše raznoliki i nestrukturirani. Zbune se kada su objekti preklapljeni, poput konja i jahača ili stabala ispred zgrada. Predložena rješenja tog problema su dodavanje malog broja označenih primjera te korištenje semantičke segmentacije pri učenju.

MUNIT arhitektura pokušava uhvatiti potpunu distribuciju podataka korištenjem multimodalnog preslikavanja. Nadograđuje se na idejama kružnih GAN-ova i UNIT modela. Cikličkom konzistencijom prisiljava mrežu da odbaci nepotrebne smetnje, a preslikavanjem na latentni prostor iz podataka izvlači korelacije koje ih vežu unutar kolekcije. Dodatno, rastavljanjem latentnog prostora na sadržajni i stilistički podprostor omogućuje joj neviđenu kontrolu nad operacijama koje obavlja – od svih uspoređenih metoda daje nadaleko najbolje rezultate u mnogim područjima.

8. Literatura

- [1] Generativni modeli (gm) - laboratorijske vježbe, . URL <https://dlunizg.github.io/lab4/#gan>.
- [2] Variational autoencoders explained, . URL <http://kvfrans.com/variational-autoencoders-explained/>.
- [3] Amjad Almahairi, Sai Rajeswar, Alessandro Sordoni, Philip Bachman, i Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*, 2018.
- [4] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.
- [5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, i Yoshua Bengio. Generative adversarial nets. U *Advances in neural information processing systems*, stranice 2672–2680, 2014.
- [6] Ian Goodfellow, Yoshua Bengio, i Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [7] Irina Higgins, Loic Matthey, Xavier Glorot, Arka Pal, Benigno Uria, Charles Blundell, Shakir Mohamed, i Alexander Lerchner. Early visual concept learning with unsupervised deep learning. *arXiv preprint arXiv:1606.05579*, 2016.
- [8] Xun Huang, Ming-Yu Liu, Serge Belongie, i Jan Kautz. Multimodal unsupervised image-to-image translation. *arXiv preprint arXiv:1804.04732*, 2018.
- [9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, i Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.

- [10] Diederik P Kingma i Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [11] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, i Ole Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.
- [12] Ming-Yu Liu, Thomas Breuel, i Jan Kautz. Unsupervised image-to-image translation networks. U *Advances in Neural Information Processing Systems*, stranice 700–708, 2017.
- [13] Fujun Luan, Sylvain Paris, Eli Shechtman, i Kavita Bala. Deep photo style transfer. *CoRR*, *abs/1703.07511*, 2017.
- [14] Mehdi Mirza i Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [15] Jun-Yan Zhu, Taesung Park, Phillip Isola, i Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.

Suparnički generativni modeli za prevođenje slika

Sažetak

Mnogi problemi u procesiranju slika odnose se na translaciju ulazne slike u željenu izlaznu. Većina današnjih arhitektura bazira se na generativnim suparničkim mrežama, nad kojima su osmišljena mnoga proširenja kako bi se prilagodila specifičnim problemima. Uvjetnim generativnim mrežama nije potrebno definirati funkciju gubitka jer ju mogu same naučiti, kružne generativne mreže sposobne su učenju translacije ne samo ulaznih u izlazne primjere nego i u drugom smjeru, a ograničavanjem mreže na translaciju unutar prostora boja obavljat će prijenos stila bez distorzije sadržaja slike. Ovim radom usporedit ću različite arhitekture, njihove karakteristike i primjene na problemima translacije slike u sliku.

Ključne riječi: translacija slike, prijenos stila, varijacijski autoenkoderi, uvjetni GAN-ovi, kružni GAN-ovi, UNIT, MUNIT

Generative adversarial networks for image translation

Abstract

Many problems in image processing revolve around translation of the input image. Most of existing architectures are based on generative adversarial networks extended with many modifications according to specific problem they solve. Conditional generative networks are capable of learning loss function by themselves, CycleGANs learn translation both from input to output but also otherway. If the network is restricted on color space, we can prevent any modifications on structural information which is shown through multimodal unsupervised translation model. This seminar gives quick overview and introduction into different kinds of GANs for image translation, their specifics and use cases.

Keywords: image-to-image generation, style transfer, VAEs, ConditionalGANs, CycleGANs, UNIT, MUNIT