

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Suparnički generativni modeli za prevođenje slika

Krešimir Vukić

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

Zagreb, svibanj 2018.

SADRŽAJ

1. Uvod	1
2. Duboke mreže	2
2.1. Konvolucija	2
2.2. Primjer konvolucijske mreže	3
3. Generativni suparnički modeli (GAN-ovi)	4
3.1. Generativni naspram diskriminativnih algoritama	4
3.2. Suparnički modeli	4
3.3. Rad GAN-a	5
3.4. GAN-ovi naspram autoenkodera	5
4. Uvjetne generativne mreže	6
4.1. Generalizacija problema prevođenja slike	6
4.2. Pristup prevođenju slike	6
4.3. Primjena	6
5. Translacija bez uparenih primjera za učenje	7
5.1. Kružni GAN-ovi	7
5.1.1. Prednosti	7
5.1.2. Rad CycleGAN-a	7
5.2. Multimodalna translacija (MUNIT)	7
6. Prijenos stila	8
6.1. Očuvanje strukture	8
6.2. Lokalizacija stila i očuvanje konteksta	8
7. Zaključak	9
8. Literatura	10

1. Uvod

Razvojem tehnika namjenjenih identifikaciji i analizi slika, logični sljedeći korak unutar područja računalnog vida bio je osmišljavanje algoritama kojima bi se omogućila reprodukcija vizualnih sadržaja.

Generiranje slika na temelju skice, prijenos stilova i rekonstrukcija fotografije neke su od primjena dotičnog problema. Najznačajniji pomak u tome području ostvario je Ian Goodfellow et al. 2014. g. [goodfellow2014generative] osmišljavanjem generativnih suparničkih modela čiji tipičan cilj je učenje funkcije preslikavanja iz prostora latentnih varijabli na željenu distribuciju.

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.

Ovim radom istražiti će se različita proširenja kojima se ti modeli prilagođavaju specifičnim problemima, njihove karakteristike i primjene.

2. Duboke mreže

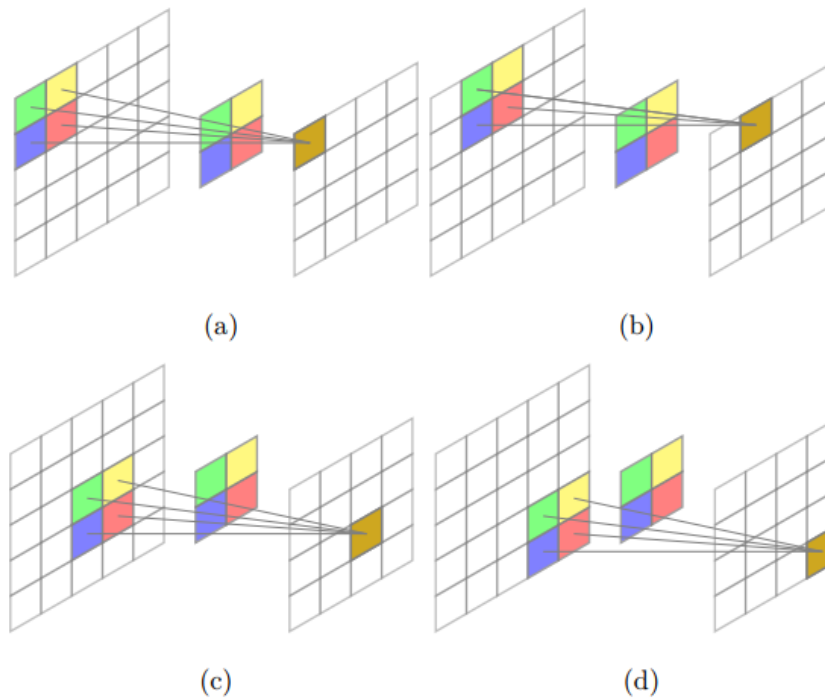
Duboke neuronske mreže zadnjih godina izuzetno su popularne u rješavanju klasifikacijskih problema i pronalaženju uzoraka. Tome je zaslužna prednost koju imaju naspram klasičnih pristupa, a koja se očituje u vidu mogućnosti da samostalno pronalaze bitne značajke u podacima te na temelju njih i tražena rješenja.

Slojevit pristupu rastavljanja podataka na jednostavnije građevne jedinice čini ih posebice pogodnima za radu sa problemima hijerarhijske prirode poput slika.

2.1. Konvolucija

Konvolucijske mreže osmišljene su s ciljem dohvata lokalnih relacija među podacima koje se posebice očituju pri radu sa slikama. Naime, ukoliko želimo analizirati malenu značajku na slici, nepoželjno je da njena pozadina ili lokacija unutar slike utječe na nju samu.

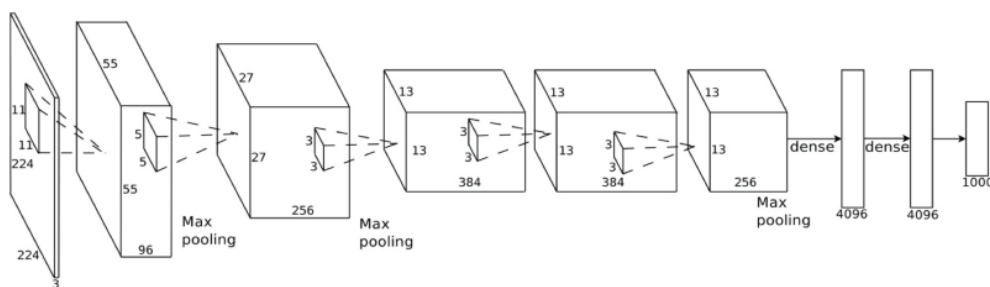
Ideja konvolucije je da se kao ulaz u neurone sljedećeg sloja dovede više vrijednosti s prethodnog što je prikazano slikom 2.1. To se odvija tako da se unaprijed definiran kvadrat pomiče (konvoluirá) po ulaznim podacima i tako određuje mapiranje na sljedeći sloj. Rezultat tog procesa je da pridodajemo veću važnost pikselima u neposrednoj okolini istraživanog područja.



Slika 2.1: Proces konvolucije

2.2. Primjer konvolucijske mreže

Nizanjem konvolucijskih slojeva dobivamo arhitekturu u kojoj je svaki sloj zadužen za jednostavniji potproblem sljedećeg sloja. Na slici 2.2 prikazana je duboka neuronska mreža s 5 konvolucijskih slojeva među kojima se odvija sažimanje maksimalnom vrijednošću te dva sloja potpune povezanosti.



Slika 2.2: Primjer duboke neuronske mreže

3. Generativni suparnički modeli (GAN-ovi)

3.1. Generativni naspram diskriminativnih algoritama

Dvije temeljne skupine kojima dijelimo algoritme za učenje su diskriminativni i generativni. Nazivi tih algoritama dolaze izravno iz vrste problema koji rješavaju:

- Diskriminativni modeli baziraju se na modeliranju isključivo pomoću ulaznih varijabli koje direktno mapiraju na ciljne vrijednosti i time uče uvjetnu vjerojatnost $p(Y|X)$. Ne stvaraju puno pretpostavki o podložnoj distribuciji, no vrlo su ovisni o kvalitetnim podacima.
- Generativni modeli rješavaju problem kako dobiti X uz dani Y , tj. pokušavaju naučiti distribuciju pojedinih klasa $p(X | Y)$.

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

3.2. Suparnički modeli

U praksi se pokazalo kako diskriminativni modeli postižu bolje rezultate od generativnih. Iako je očito da su generativni algoritmi svestranije naravi od diskriminativnih, to dolazi sa neizbježnim povećanjem složenosti koja proizlazi iz toga što se pronalazak ciljne distribucije u dubokim hijerarhijskim podacima

sastoji od aproksimiranja problema čija rješenja matematički nije moguće izvesti[goodfellow2014generative]. Bez toga, nije moguće ni podesiti model da radi pri svom optimumu. Odgovor na predstavljenu složenost pronađen je u suparničkim modelima čiji rad se bazira na istovremenom učenju dvije podmreže. Ovisno o implementaciji, mreža takve arhitekture u stanju je kopirati distribuciju širokog spektra podataka: slika, govora, muzike, pa čak i pisanih zapisa.

3.3. Rad GAN-a

Ideja generativnih suparničkih modela je da se istovremeno uče dvije 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. Cijeli proces usporedljiv je s poslovima detektiva i kriminalca. Jedna strana u bavi se falsifikacijom dokaza, a druga analizom njihove autentičnosti.

Koraci rada GANa:

1. generator G iz nasumičnih brojeva (šum) Z stvara sliku $G(Z)$
2. ta slika se dovodi na ulaz diskriminatora D uz skup stvarnih slika iz podatkovnog skupa Y
3. diskriminator D određuje postotak vjerojatnosti da je slika $G(Z)$ prava

3.4. GAN-ovi naspram autoenkodera

4. Uvjetne generativne mreže

4.1. Generalizacija problema prevođenja slike

Uvjetne generativne mreže pokazale su se kao generalno rješenje problema prevođenja slike [isola2017image]. Svojtvene su po učenju ne samo funkcije prijenosa između slika već i funkcije gubitka nad kojom se ona trenira. Dok se u drugim modelima ona ručno zadaje, pristup u kojem se ona uči omogućuje generalizaciju nad setovima problema koji bi u suprotnom zahtijevali vrlo drugačije formulacije gubitka.

4.2. Pristup prevođenju slike

Analogno prevođenju teksta u drugi jezik zadržavajući određeni kontekst, prevođenje slike odnosi se na translaciju scene u drugačiju reprezentaciju. To se odnosi na mnogobrojne probleme za koje postoje zasebne metode rješavanja poput izmjene pozadine, lica ili promjene boje. No, u srži svih njih može se prepoznati jedinstven problem: naći funkciju preslikavanja piksela u piksel. Tokom učenja, lošim definiranjem te funkcije, nikad nećemo dobiti realistične rezultate. GANovi pristupaju tome tako da odbacuju sva rješenja koja se mogu razlikovati od stvarnosti generirajući slike y iz nasumičnog šuma $z : G : z \rightarrow y$. Poboljšanje nad tom metodom dobivamo uz uvjetovanje ulaznih podataka, tj. uz dodavanja uvjeta na šum da mora sličiti stvarnoj distribuciji. To se postiže dodavanjem slike x slične željenoj na ulaz sustava $G : x, z \rightarrow y$ [isola2017image].

4.3. Primjena

Neke od mogućnosti uvjetnih GAN-ova su sinteza fotorealističnih slika iz skica, generiranje objekata na temelju njihovih obruba te bojanje crno bijelih fotografija.

5. Translacija bez uparenih primjera za učenje

5.1. Kružni GAN-ovi

5.1.1. Prednosti

Većina postojećih algoritama korištenih u računalnom vidu ovisi o postojanju velikih podatkovnih setova označenih parova slika. No, za mnoge probleme takvi podaci nisu dostupni, a u takvim situacijama primjenu nalaze kružni GAN-ovi.

5.1.2. Rad CycleGAN-a

Temelj njihovog rada je uz da uz učenje funkcije $G : X \rightarrow Y$, istovremeno učien inverz $F : Y \rightarrow X$ pomoću gubitka cikličke dosljednosti: $F(G(X)) \approx X$ i obrnuto.

5.2. Multimodalna translacija (MUNIT)

Druge metode nenadziranog učenja translacije ne uzimaju u obzir da je uvjetna distribucija preslikavanja slika između dvije domene inherentno multimodalna [huang2018multimodal] što kao posljedicu ima 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 sa nasumičnim stilskim kodom uzorkovanim s ciljane domene.

6. Prijenos stila

6.1. Očuvanje strukture

Postojeće metode ograničene su izborom stilova koje mogu prenijeti te u realističnosti dobivenih rezultata. Pokušaji prijenosa stila koji mijenjaju vrijednost piksela unose sitne izmjene u linijama i teksturama što na resultantnom sadržaju stvara dojam crteža, a gubi se realističnost fotografije. Uvjet koji se zbog toga postavlja na rad mreže je da mogže unositi izmjene isključivo na prostoru boja.

6.2. Lokalizacija stila i očuvanje konteksta

[luan2017deep] Postoji još jedan faktor koji povećava kompleksnost cijelog procesa, a to je da se ciljano ponašanje prijenosa stila drži konteksta izmjene. Na primjer, ukoliko se na fotografiji nalaze livada i oblaci, modifikacije na livadi moraju biti povezane sa stilom livade na referentnoj slici, a ne da povuku stil sa oblaka.

7. Zaključak

Generativni modeli tema su mnogobrojnih istraživanja i vrlo obećavajuća metoda učenja 'prirodnih' značajki podataka. Uz napretke koji se redovno postižu u njihovom unaprijeđenju, velika je vjerojatnost da će ubrzo biti u stanju generirati uzorke nerazlučive od stvarnosti. Time primjenu pronalaze u područjima poput stvaranja virtualnih svijetova bez ljudske intervencije, izvršavanja modifikacija nad slikama pomoću jednostavnog govora, 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 inženjeringom i omogućava ljudima poput umjetnika i mladih znanstvenika da ju koriste samostalno. ...

8. Literatura

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.

Ian Goodfellow, Yoshua Bengio, i Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.

Xun Huang, Ming-Yu Liu, Serge Belongie, i Jan Kautz. Multimodal unsupervised image-to-image translation. *arXiv preprint arXiv:1804.04732*, 2018.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, i Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.

Fujun Luan, Sylvain Paris, Eli Shechtman, i Kavita Bala. Deep photo style transfer. *CoRR*, *abs/1703.07511*, 2017.

Mehdi Mirza i Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

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 će različite arhitekture, njihove karakteristike i primjene na problemima translacije slike u sliku.

Ključne riječi: image-to-image generation, style transfer, GANs, CycleGANs

Image to image generation using CycleGANs

Abstract

Keywords: image-to-image generation, style transfer, GANs, CycleGANs