SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2289

DETEKCIJA IZVANDISTRIBUCIJSKIH DIJELOVA SLIKE PRIMJENOM GENERATIVNIH MODELA

Matej Grcić

Zagreb, lipanj 2020.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2289

DETEKCIJA IZVANDISTRIBUCIJSKIH DIJELOVA SLIKE PRIMJENOM GENERATIVNIH MODELA

Matej Grcić

Zagreb, lipanj 2020.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 13. ožujka 2020.

DIPLOMSKI ZADATAK br. 2289

Pristupnik:	**Matej Grcić (0036490780)**
Studij:	Računarstvo
Profil:	Računarska znanost
Mentor:	prof. dr. sc. Siniša Šegvić

Zadatak: **Detekcija izvandistribucijskih dijelova slike primjenom generativnih modela**

Opis zadatka:

Semantička segmentacija slika važan je zadatak računalnog vida s mnogim zanimljivim primjenama. U posljednje vrijeme najbolji rezultati u tom području postižu se diskriminativnim konvolucijskim modelima. Međutim, diskriminativni modeli su skloni neopravdanom optimizmu, što znači da primjerci izvan domene ekspertize modela često bivaju neispravno klasificirani s velikom mjerom pouzdanosti. Ovaj rad proučava mogućnost rješavanja tog problema primjenom generativnih modela. U okviru rada, potrebno je istražiti diskriminativne modele za semantičku segmentaciju te generativne modele temeljene na invertibilnom normalizirajućem toku. Oblikovati model za generiranje podataka na rubu distribucije skupa za učenje te segmentacijski model koji na izlazu daje gustu predikciju distribucije pripadnosti semantičkim razredima te vjerojatnost izvandistribucijskog ulaza. Validirati hiperparametre, prikazati i ocijeniti ostvarene rezultate te provesti usporedbu s rezultatima iz literature. Predložiti pravce budućeg razvoja. Radu priložiti izvorni kod razvijenih postupaka uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 30. lipnja 2020.

SADRŽAJ

1. Uvod **1**

2. Pregled literature **3**

2.1. Detekcija izvandistribucijskih primjera 3

2.2. Detekcija izvandistribucijskih dijelova slike 6

3. Generativni modeli s normalizirajućim vjerojatnosnim tokom **8**

3.1. Autoregresijski tokovi 9

3.2. Normalizirajući tok bez očuvanja vjerojatnosnog volumena 10

3.2.1. Sloj miješanja 10

3.2.2. Arhitektura modela 11

3.2.3. Učenje modela 12

3.3. Usporedba s ostalim vrstama generativnih modela 13

4. Detekcija izvandistribucijskih primjera **14**

4.1. Predložena metoda 14

4.2. Generiranje primjera s ruba distribucije 15

5. Modeli za semantičku segmentaciju **17**

5.1. Općeniti pregled arhitektura 17

5.2. Arhitektura Ladder DenseNet 18

5.3. Optimizacija memorijskog otiska arhitekture Ladder DenseNet 20

6. Gusta detekcija odudarajućih (izvandistribucijskih) dijelova slike **23**

6.1. Modifikacija skupa za učenje 24

6.2. Modeli za istovremenu semantičku segmentaciju i detekciju izvandistribucijskih dijelova slike 25
6.3. Evaluacija performansi modela 26

7. Eksperimenti **29**
7.1. Detekcija izvandistribucijskih primjera 29
7.1.1. Rezultati na skupu podataka CIFAR10 29
7.1.2. Rezultati na skupu podataka SVHN 31
7.2. Generiranje primjera s ruba distribucije 33
7.3. Istovremena semantička segmentacija i detekcija izvadnistribucijskih dijelova slike 35

8. Zaključak **41**

Literatura **43**

1. Uvod

Duboko učenje (*eng.* Deep learing) je potpodručje strojnog učenja (*eng.* Machine learing). Interes za ovim područjem intenzivno raste nakon pobjede dubokog konvolucijskog modela koji se zove AlexNet Krizhevsky et al. (2012) na natjecanju ImageNet održanom 2012. godine. Sljedećih godina ovo područje doživljava rapidan razvoj. Upravo zahvaljujući tom razvoju, danas uspješno rješavamo neke probleme računalnog vida, obrade prirodnog jezika te mnogih drugih područja.

Ubrzani razvoj dubokog učenja omogućili su javno dostupni veliki skupovi podataka, napreci u istraživanju optimizacijskih algoritama i povećanje hardverskih mogućnosti uz smanjenje cijene. Konkretno, pojava *CUDA* tehnologije je omogućila smještanje dubokih modela na grafičke procesore te dodatno ubrzanje. Konačno, pojava programskih razvojnih okvira kao što su *PyTorch*, *TensorFlow* i drugi omogućila je jednostavnu implementaciju novih ideja te brzo izvođenje eksperimenata.

Duboki modeli su prediktivni modeli koji sadrže više od jedne nelinearne transformacije. Suvremeni modeli imaju desetke milijuna parametara te zahtijevaju nekoliko gigabajta memorije za uspješno treniranje postupkom propagacije pogreške unatrag (*eng.* Backpropagation). Suvremeni modeli uspješno rješavaju izazove u području računalnog vida (*eng.* Computer vision). Računalni vid je područje računarske znanosti koje želi omogućiti računalima da prepoznaju i obrađuju predmete u slikama i videozapisima na isti način kao što to čine ljudi.

Iako se duboki modeli danas intenzivno primjenjuju za rješavanje niza kompleksnih zadataka, imaju određene nedostatake koji predstavljaju zanimljiv smjer daljnjeg istaživanja. Jedan od glavnih nedostataka dubokih modela je njihova naglašena samouvjerenost Lakshminarayanan et al. (2017); Guo et al. (2017). Naime, ukoliko dubokom modelu predamo primjer koji ne podliježe distribuciji skupa podataka na

kojem je model naučen, model takav primjer klasificira s velikom sigurnošću u krivi razred. Ovakve primjere nazivamo izvandistribucijskim primjeri. Slično, ako primjeru iz skupa podataka dodamo pomno odabranu perturbaciju, tada model takav primjer klasificira u drugu klasu s velikom sigurnošću. Takve primjere nazivamo neprijateljskim primjerima. Obje navedene vrste primjera remete normalan rad dubokog modela i potencijalno su veliki sigurnosni problem.

Glavni doprinosi ovog rada je pojednostavljenje i poboljšanje procedure definirane u Lee et al. (2018) korištenjem normalizirajućih vjerojatnosnih tokova te primjena navedene metode u detekciji dijelova slike koji sadrže anomaliju. Ostatak rada organiziran je na sljedeći način. U poglavlju 2 nudimo pregled literature. U poglavlju 3 opisujemo generativne modele temeljene na normalizirajućem vjerojatnosnom toku (*eng.* normalizing flow). Fokusiramo se na autoregresijski tok Real NVP Dinh et al. (2017) temeljen na invertibilnim transformacijama. U poglavlju 4 opisujemo našu varijantu detekcije izvandistribucijskih primjera koja se temelji na metodi definiranoj u Lee et al. (2018). U poglavlju 5 opisujemo arhitekture modela za semantičku segmentaciju s naglaskom na arhitekturu Ladder DenseNet Kreso et al. (2019). U poglavlju 6 opisujemo pristup gustoj detekciji odudarajućih dijelova slike uz paralelnu semantičku segmentaciju. Poglavlje 7 sadrži provedene eksperimente te dobivene rezultate kojima potkrepljujemo iznesene tvrdnje.

2. Pregled literature

Duboki diskriminativni modeli izvandistribucijske primjere svrstavaju u neku od klasa unutardistribucijskog skupa podataka s visokom sigurnošću Lakshminarayanan et al. (2017); Guo et al. (2017). Ova karakteristika dubokih modela predstavlja veliki sigurnosni problem suvremenim sustavima temeljenim na umjetnoj inteligenciji. Iz tog razloga je detekcija izvandistribucijskih primjera područje aktivnog istraživanja. Idealno, diskriminativni model bi izvandistribucijski primjer svrstao u neku klasu s niskom sigurnošću. U tom slučaju bi na temelju niske sigurnosti modela označili dani primjer kao izvandistribucijski primjer Bishop (2007).

Prijelaz iz domene klasifikacije u domenu semantičke segmentacije omogućuje detekciju određenih dijelova slike koji sadrže anomalije, što je dodatan izazov. Promjena domene uzrokuje promjene u arhitekturi modela. Iz tog razloga temeljit pregled arhitektura modela za semantičku segmentaciju prikazujemo u poglavlju 5.

2.1. Detekcija izvandistribucijskih primjera

Učenje diskriminativnog modela samo na unutardistribucijskom skupu podataka povlači za sobom pretpostavku zatvorenog svijeta. Ovako naučen model nije pogodan za korištenje u otvorenom svijetu. Posljedično, većina pristupa za poboljšanje performansi diskriminativnog modela se temelji na uklanjanju pretpostavke zatvorenog svijeta.

U Lee et al. (2018) autori uklanjaju pretpostavku zatvorenog svijeta tako da gubitku klasifikatora dodaju član koji probabilistički izlaz modela pomiču k uniformnoj razdiobi za primjere s ruba distribucije. Složeni gubitak definiramo izrazom:

$$L(\theta) = \mathbb{E}_{(\hat{\mathbf{x}},\hat{y})\sim P_{in}}[-\log P_\theta(y=\hat{y}|\hat{\mathbf{x}})] + \lambda\, \mathbb{E}_{\mathbf{x}\sim P_{bord}}[\mathrm{KL}(P_\theta(y|\mathbf{x}), U)] \tag{2.1}$$

gdje P_{in} predstavlja unutardistribucijski skup podataka, a P_{bord} primjere s ruba distribucije. Parametar λ je konstanta veća od 0. Primjere s ruba distribucije autori generiraju korištenjem generativnog suparničkog modela Goodfellow et al. (2014); Radford et al. (2016). Gubitku generatora dodajemo isti član, čime postaje sposoban generirati primjere s ruba distribucije. Korištenje sintetičkih primjera s ruba distribucije je vrlo primamljivo zbog elegantnog načina uklanjanja pretpostavke zatvorenog svijeta bez korištenja dodatnih skupova podataka.

Prema Hendrycks et al. (2019b), klasifikator bolje detektira izvandistribucijske primjere ukoliko je učen gubitkom 4.1, gdje umjesto primjera s ruba distribucije koristimo primjere iz negativnog skupa podataka. Ovaj pristup pretpostavlja dostupan negativni skup podataka odgovarajuće veličine i vizualne raznolikosti. Mana ovakvog pristupa je dodatno hardversko opterećenje koje iziskuje korištenje negativnog skupa podataka.

Za razliku od prethodna dva pristupa, pristup ODIN Liang et al. (2018) dodaje primjeru iz skupa za učenje perturbacije temeljene na temperaturnom skaliranju probabilističkog izlaza modela. Primjer iz skupa za učenje su modificirani prema formuli:

$$\hat{\mathbf{x}} = \mathbf{x} - \epsilon \operatorname{sign}(-\nabla_{\mathbf{x}} \log S_{\hat{y}}(\mathbf{x}; T)) \tag{2.2}$$

gdje je ϵ konstanta veća od 0, a S_i definiramo izrazom:

$$S_i(\mathbf{x}; T) = \frac{\exp(f_i(\mathbf{x})/T)}{\sum_{j=1}^{K} \exp(f_j(\mathbf{x})/T)} \tag{2.3}$$

Mana ovog pristupa je u tome što za svaki primjer zahtjeva dva prolaska unaprijed i jedan prolazak unazad kroz cjelokupan model. Parametar T je u fazi učenja postavljen na 1, a zatim se ugađa prije no što model koristi u otvorenom svijetu.

U svim navedenim slučajevima vjerojatnost da primjer pripada unutardistribucijskom skupu podataka definiramo kao maksimalnu vrijednost probabilističkog izlaza diskriminativnog modela (max-softmax). Ukoliko je maksimalna vrijednost softmaksa veća ili jednaka predefiniranom pragu, primjer označavamo kao unutardistribucijski primjer:

$$g(\mathbf{x}) = \begin{cases} 1 & \text{ako} \quad \max_y P_\theta(y|\mathbf{x}) \geqslant \delta \\ 0 & \text{inače} \end{cases} \tag{2.4}$$

gdje δ predstavlja unaprijed odabranu konstantu.

DeVries i Taylor (2018) umjesto transformacije klasifikatora u detektor izvandistribucijskih primjera korištenjem maksimalne vrijednosti probabilističkog izlaza koriste dodatnu glavu odgovornu za detekciju izvandistribucijskih primjera. Probabilistički izlaz dodatne glave tumačimo kao vjerojatnost da primjer pripada skupu podataka za učenje. Dodatna glava se sastoji od nekoliko potpuno povezanih slojeva, a njezin ulaz je izlaz pretposljednjeg sloja danog modela. Gubitak modela je definiran izrazom:

$$L(\theta) = -\sum_{i=1}^{K} \log(p_i')y_i - \lambda \log(c) \tag{2.5}$$

gdje p_i' definiramo kao:

$$p_i' = c\, p_i + (1-c)y_i \tag{2.6}$$

c predstavlja probabilistički izlaz dodatne glave, a p_i vjerojatnost da dani primjer pripada klasi i (izlaz klasifikacijske glave). U ovom slučaju, primjer se označava kao izvandistribucijski ukoliko je izlaz dodatne glave manji od predefiniranog praga.

Detekcija izvandistribucijskih primjera se može izvesti uz pomoć generativnih modela evaluiranjem izglednosti. Ipak, Nalisnick et al. (2019); Serrà et al. (2020) prijavljuju da generativni modeli sposoban egzaktno evaluirati izglednost dodjeljuju veću izglednost izvadistribucijskim primjerima. Također, vizualno jednostavniji primjeri imaju veću izglednost neovisno o unutardistribucijskom skupu podataka.

Autori u Grathwohl et al. (2020) rješavaju navedeni problem definiranjem mjere temeljene na L2-normi gradijenta izglednosti:

$$s(\mathbf{x}) = -\left|\left|\frac{\partial \log p_\theta(\mathbf{x})}{\partial \mathbf{x}}\right|\right|_2 \tag{2.7}$$

međutim takav pristup radi za modele temeljene na energiji (JEM Grathwohl et al. (2020) i IGEBM Du i Mordatch (2019)). U slučaju kad logaritam izglednosti evaluiramo autoregresijskim modelima ili modelima temeljenim na normalizirajućem toku, detekcija izvandistribucijskih primjera koristeći ovu mjeru ne poručuje dobre rezultate.

Pristup Ren et al. (2019) koristi omjer izglednosti dvaju modela za detekciju izvandistribucijskih primjera. Jedan model autori uče na primjerima iz skupa podataka, dok drugi model uče na zašumljenim primjerima istog skupa podataka. Ovakvu mjeru definiramo izrazom:

$$LLR(\mathbf{x}) = \log \frac{p_\theta(\mathbf{x})}{p_{\theta_o}(\mathbf{x})} = \log p_\theta(\mathbf{x}) - \log p_{\theta_o}(\mathbf{x}) \tag{2.8}$$

gdje je nazivnik model koji je učen na zašumljenim primjerima. Prema autorima, model naučen na zašumljenim primjerima uči pozadinu slike, dok model naučen na podatcima uči objekt na slici te pozadinu. Izglednost je evaluirana koristeći autoregresijske modele. Mana ovakvog pristupa je potreba za korištenjem dvaju modela, što iziskuje dodatno hardversko opterećenje.

Prema Hendrycks et al. (2019b) autoregresijski generativni model učen optimiranjem gibitka koji eksplicitno zahtjeva da izglednost izvandistribucijskih primjera bude veća od izglednosti primjera iz negativnog skupa podataka može detektirati izvandistribucijske primjere evaluacijom izglednosti.

2.2. Detekcija izvandistribucijskih dijelova slike

Semantička segmentacija slika je zadatak u kojem se svaki piksel klasificira u neku od predefiniranih klasa. Posljedično, problem detekcije izvandistribucijskih primjera evoluira u problem detekcije izvandistribucijskih dijelova slike.

U radovima Bevandic et al. (2018, 2019) autori obrađuju problem detekcije anomalija uz istovremenu gustu predikciju slika prometa, gdje je kamera u automobilu u vožnji. Javno dostupni skupovi podataka Cordts et al. (2016); Neuhold et al. (2017) ne pokrivaju sve moguće vremenske uvjete i situacije u prometu. Posljedično, model naučen na takvom skupu podataka ima značajan pad performansi pri korištenju u otvorenom svijetu. Autori predlažu učenje diskriminativnog modela s glavom za gustu predikciju na slikama u koje su zalijepljeni primjeri iz negativnog skupa podataka. Primjeri iz negativnog skupa podataka simuliraju anomalije iz otvorenog svijeta. Konkretno, autori lijepe primjere iz skupa podataka ImageNet-1k Deng et al. (2009) u slike prometa iz skupa podataka Mapillary Vistas Neuhold et al. (2017) te Cityscapes Cordts et al. (2016).

Autori postižu najbolje performanse u istovremenoj semantičkoj segmentaciji i detekciji izvandistribucijskih primjera s modelom Ladder style DenseNet Krapac et al. (2017). Varijanta modela s jednom glavom ima dodatan član u gubitku koji zahtjeva

da model daje uniformnu razdiobu za izvandistribucijske dijelove slike. Druga varijanta modela ima zasebnu glavu odgovornu za detekciju izvandistribucijskih primjera čiji probabilistički izlaz tumačimo kao vjerojatnost da piksel pripada unutardistribucijskom skupu podataka.

Mjerenje performansi modela za semantičku segmentaciju nije trivijalan zadatak. Trenutno, najkorišteniji benchmark za performanse modela u stvarnom svijetu je skup podataka Wilddash Zendel et al. (2018). Benchmark za ovaj skup podataka se računa na temelju 70-ak slika, što je na žalost nedovoljno da pokrije sve moguće situacije u otvorenom svijetu. Iz tog razloga je dizajniranje odgovarajućeg skupa podataka područje aktivnog istraživanja.

U vrijeme pisanje ovog rada najbolje rezultate na skupu podataka Wilddash postiže rad Lambert et al. (2020), gdje autori uče model za gustu predikciju na skupu podataka koji objedinjuje većinu javno dostupnih skupova podataka koji prikazuju vožnju automobila. Najveći izazov agregacije svih skupova podataka je ujednačavanje labela i broja klasa različitih skupova. Učenjem modela HRNet-W48 Sun et al. (2019) na agregiranom skupu podataka autori značajno povećavaju robusnost modela.

Spomenimo i skupove podataka Fishyscapes Blum et al. (2019) i BDD-Anomaly Hendrycks et al. (2019a) koji služe za evaluiranje performansi modela u detekciji anomalija na scenama vožnje.

3. Generativni modeli s normalizirajućim vjerojatnosnim tokom

Normalizirajući vjerojatnosni tokovi su modeli koji se temelje na formuli za promjenu varijable distribucije. Ulazna distribucija prolazi kroz invertibilnu transformaciju, te izlazi kao distribucija latentne varijable. Formula za promjenu varijable distribucije definirana je sljedećom jednadžbom:

$$p_x(\mathbf{x}) = p_z(\mathbf{f}(\mathbf{x})) \left| \det \left(\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right) \right| \tag{3.1}$$

U prikazanoj jednadžbi $|.|$ predstavlja apsolutnu vrijednost, a $\det(X)$ determinantu matrice X. Da bi formula vrijedila funkcija **f** mora biti bijekcija. Posljedica ovako definirane funkcije **f** je očuvanje broja dimenzija ulaza.

Jednadžba 3.1 je vrlo intuitivna. S obzirom da distribuciju $p(x)$ transformiramo u neku drugu distribuciju $p(z)$ preko funkcije **f**, ta funkcija može kontrahirati vjerojatnosni prostor. S obzirom da je vjerojatnosni volumen konstantan, posljedica kontrakcije je promjena gustoće vjerojatnosti. Kada govorimo o višedimenzionalnom vjerojatnosnom prostoru, tada promjenu u gustoći vjerojatnosti kompenziramo množenjem s determinantom derivacije funkcije **f**. Geometrijska interpretacija determinante kvadratne matrice X jest faktor za koji množenje s matricom X kontrahira geometrijski prostor.

Normalizirajući tok nastaje primjenom više invertibilnih transformacija na ulaznu distribuciju što je definirano izrazom:

$$q_0(\mathbf{z}_0) = q_K(\mathbf{z}_K) \prod_{k=1}^{K} \left| \det \left(\frac{\partial \mathbf{f}_k}{\partial \mathbf{z}_{k-1}} \right) \right| \tag{3.2}$$

gdje vrijedi:

$$\mathbf{z}_K = \mathbf{f}_K \circ ... \circ \mathbf{f}_1(\mathbf{z}_0), \quad \mathbf{z}_0 \sim q_0(\mathbf{z}_0) \tag{3.3}$$

3.1. Autoregresijski tokovi

Autoregresijski tokovi su podvrsta normalizirajućih tokova. Pri modeliranju arhitekture ovakvih tokova vodimo računa o traktabilnosti determinanta jakobijana transformacije. Jednadžba proizvoljne transformacije autoregresijskog toka je dana izrazom:

$$\mathbf{z}_i = \mathbf{f}(\mathbf{x}_{1:i}) \tag{3.4}$$

Ako pretpostavimo da je **x** vektor proizvoljne dimenzije d, tada iz navedenog izraza zaključujemo da i-ta dimenzija vektora **z** ovisi samo o prvih i dimenzija ulaznog vektora **x**. Posljedično, derivacija funkcije **f** je jakobijan donje trokutaste matrice. Determinantu jakobijana računamo kao umnožak elemenata na dijagonali matrice, što je prikazano izrazom:

$$\det(J) = \prod_{i=1}^{d} J_{ii} \tag{3.5}$$

Postoji nekoliko vrsta autoregresijskih tokova koje razlikujemo na temelju odabira invertibilne funkcije **f**. U okviru ovog rada detaljnije proučavamo tok *Real Non-Volume Preserving Flow* (Real NVP) Dinh et al. (2017). Spomenimo još i autoregresijskog tok *Glow* Kingma i Dhariwal (2018) koji se temelji na invertibilnim konvolucijskim blokovima s jezgrom veličine 1×1. Slično kao i Real NVP, ovaj tok se sastoji od više slojeva miješanja (*eng.* coupling layer). Sloj miješanja je detaljno objašnjen u nastavku rada.

3.2. Normalizirajući tok bez očuvanja vjerojatnosnog volumena

Fokusiramo se na vrstu normalizirajućeg autoregresijskog toka koji se naziva Real NVP Dinh et al. (2017). Zbog svog dizajna, ovaj model može efikasno izvesti prolaz unaprijed i unatrag te uzorkovati naučenu distribuciju. Osnovnu gradivnu jedinicu ovog modela nazivamo sloj miješanja (*eng.*Coupling layer) te ćemo ga detaljnije proučiti u nastavku.

3.2.1. Sloj miješanja

Pretpostavimo da je ulaz u sloj vektor **x** dimenzije D, a izlaz iz sloja vektor **y** jednake dimenzije kao i ulazni vektor. Transformacija unutar sloja se provodi na način da odaberemo proizvoljan broj dimenzija d ($d < D$), a zatim odabrane dimenzije propustimo na izlaz. Preostalih $D - d$ dimenzija transformiramo vodeći računa o očuvanju broja dimenzija. Transformacije unutar sloja su definirane sljedećim formulama:

$$\mathbf{y}_{1:d} = \mathbf{x}_{1:d} \tag{3.6}$$

$$\mathbf{y}_{d+1:D} = \mathbf{x}_{d+1:D} \odot \exp(\mathbf{s}(\mathbf{x}_{1:d})) + \mathbf{t}(\mathbf{x}_{1:d}) \tag{3.7}$$

Oznaka $\odot$ predstavlja Hadamardov produkt, a **s** i **t** mapiraju R^d u R^{D-d}. Navedene transformacije imaju jednostavan inverz definiran sljedećim formulama:

$$\mathbf{x}_{1:d} = \mathbf{y}_{1:d} \tag{3.8}$$

$$\mathbf{x}_{d+1:D} = (\mathbf{y}_{d+1:D} - \mathbf{t}(\mathbf{y}_{1:d})) \odot \exp(-\mathbf{s}(\mathbf{y}_{1:d})) \tag{3.9}$$

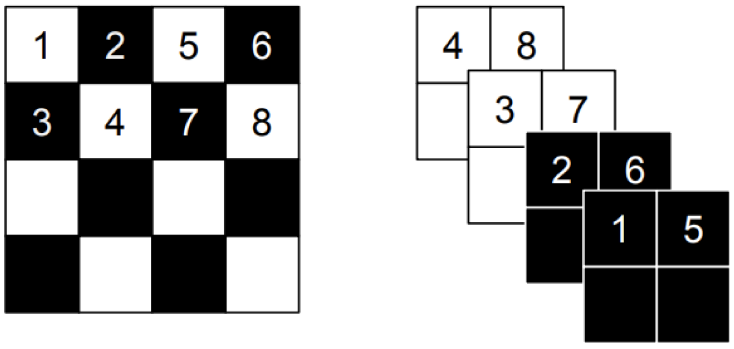
Na kompleksnost inverza ne utječu funkcije **s** i **t** pa to mogu biti složene funkcije koje ne moraju biti bijekcije. Iz tog razloga se za **s** i **t** koriste kompleksne nelinearne transformacije. Točnije, koristi se proizvoljan broj rezidualnih jedinica s predaktivacijom popularnih ResNet arhitektura He et al. (2016b). Derivacija izlaza po ulazu sloja tj. jakobijan je definiran sljedećim izrazom:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} I & 0 \\ \frac{\partial y_{d+1:D}}{\partial x_{1:d}} & diag(\exp[s(x_{1:d})]) \end{bmatrix} \tag{3.10}$$

U jakobijanu transformacije definiranom prethodnom formulom primjećujemo da determinanta ne iznosi 1 zbog transformacije definirane funkcijom **s**. Ako bismo izbacili izraz koji uključuje funkciju **s** iz transformacije sloja, tada bi naš model degenerirao u model koji čuva vjerojatnosni volumen.

3.2.2. Arhitektura modela

Kako bismo pravilno definirali kompozitnu arhitekturu Real NVP-a, prvo moramo prikazati dva načina odabira dimenzija koje propagiramo bez promjene. Maska šahovske ploče (na slici 3.1 lijevo) odabire dimenzije koje propagira bez promjene uz uvjet da se ne propagiraju dvije horizontalno ili vertikalno susjedne dimenzije u istom kanalu danog tenzora. Maska po kanalu (na slici 3.1 desno) propušta čitave kanale tenzora bez promjene. Primjetimo da prije korištenja maske po kanalu tenzor dimenzija $s \times s \times c$ preoblikujemo u tenzor dimenzija $\frac{s}{2} \times \frac{s}{2} \times 4c$.



Slika 3.1: Dva načina odabira dimenzija koje se propagiraju sljedećem sloju stapanja bez promjene. Izvor: Dinh et al. (2017)

Real NVP nastaje slaganjem više blokova poduzorkovanja jedan na drugi. Svaki blok poduzorkovanja se sastoji od sljedećeg niza transformacija. Ulazni tenzor prvo prolazi kroz tri sloja stapanja. Između svakog od slojeva stapanja se odabiru dimenzije koje se propagiraju bez promjene s maskom šahovske ploče vodeći računa da se maska alternira između slojeva. Dobiveni tenzor dimenzija $s \times s \times c$ zatim preoblikujemo u tenzor dimenzija $\frac{s}{2} \times \frac{s}{2} \times 4c$. Sljedeće, tenzor prolazi kroz još tri sloja stapanja. Ovaj put između svakog od slojeva stapanja dimenzije koje se propagiraju bez promjene odabiremo s maskom po kanalu. Konačno, tenzor vraćamo u prvobitan oblik dimen-

zija $s \times s \times c$. Između dvaju blokova poduzorkovanja pola dimenzija izlaznog tezora se više ne mijenja, to jest svaki sljedeći blok transformira tenzor s upola manjim brojem dimenzija. Propagacija vjerojatnosnog tenzora kroz niz blokova se može opisati sljedećim nizom jednadžbi:

$$h^0 = x \tag{3.11}$$

$$(z^{i+1}, h^{i+1}) = f^{i+1}(h^i) \tag{3.12}$$

$$z^L = f^L(h^{L-1}) \tag{3.13}$$

$$z = (z^1, ..., z^L) \tag{3.14}$$

gdje f^i predstavlja transformaciju bloka poduzorkovanja. Primjetimo da u jednadžbi (3.12) pola dimenzija tenzora izostavljamo od daljnje propagacije.

3.2.3. Učenje modela

Učenje generativnog modela možemo definirati kao minimiziranje KL divergencije između distribucije podataka i distribucije modela:

$$\theta^* = \arg\min_\theta \mathrm{KL}(P_{data}(\mathbf{x}) \,||\, P_\theta(\mathbf{x})) \tag{3.15}$$

što je prema formuli za KL divergenciju ekvivalentno:

$$\theta^* = \arg\min_\theta \sum_{\mathbf{x}} P_{data}(\mathbf{x}) \log P_{data}(\mathbf{x}) - \sum_{\mathbf{x}} P_{data}(\mathbf{x}) \log P_\theta(\mathbf{x}) \tag{3.16}$$

iz čega slijedi:

$$\theta^* = \arg\min_\theta \mathbb{H}[P_{data}] - \frac{1}{N} \log P_\theta(\mathbf{x}) \tag{3.17}$$

gdje $\mathbb{H}$ predstavlja entropiju distribucije. S obzirom da je entropija distribucije skupa podataka konstanta, taj član možemo zanemariti. Posljedično, zaključujemo da je minimizacija KL divergencije između distribucije podataka i distribucije modela jednaka maksimizaciji logaritamske izglednosti naučene distribucije modela, što je definirano izrazom:

$$\theta^* = \arg\max_\theta \log P_\theta(\mathbf{x}) \tag{3.18}$$

Logaritamska izglednost je dana sljedećim izrazom:

$$\log(p_x(x)) = \log(p_z(f(x))) + \log\left(\left|\det(\frac{\delta f(x)}{\delta x})\right|\right) \tag{3.19}$$

Primjetimo da izraz uključuje izračun determinante Jakobijana. To nas ne brine jer je jakobijan svakog sloja donja trokutasta matrica pa je determinanta jednaka umnošku elemenata na dijagonali. Isto tako znamo da je $\det(AB) = \det(A) \cdot \det(B)$ zbog čega se izračun determinante može izvesti sloj po sloj.

3.3. Usporedba s ostalim vrstama generativnih modela

Učenje Real NVP-a se temelji na maksimizaciji izglednosti, čime se uklanja nestabilnost karakteristična za suparničko učenje. Ipak, slike generirane s GAN-om su kvalitetnije od slika generiranih s Real NVP-em. Za razliku od varijacijskih autoenkodera, Real NVP može egzaktno evaluirati izglednost. Pri generiranju primjera s oba modela prvo uzorkujemo normalnu razdiobu, a zatim dobiveni tenzor propagiramo kroz model cijeli model (Real NVP) ili dekoderski dio (VAE). U slučaju Real NVP-a, uzorkovani tenzor ima značajno veći broj dimenzija. U usporedbi s autoregresijskim modelima, uzorkovanje Real NVP-a ima složenost $O(1)$ u ovisnosti o broju dimenzija uzorka, dok složenost uzorkovanja autoregresijskog modela u ovisnosti o broju dimenzija uzorka iznosi $O(n)$.

Za razliku od ograničenog Boltzmannovog stroja, distribucija modela normalizirajućeg toka je normalizirana po konstrukciji. Uvođenjem ograničenja da transformacija mora biti invertibilna te čuvati broj dimenzija izbjegavamo potrebu za negativnom fazom pri učenju normalizirajućeg toka.

4. Detekcija izvandistribucijskih primjera

Trenutno najbolje rješenje za problem detekcije izvandistribucijskih primjera diskriminativnim modelom Hendrycks et al. (2019b) koristi relativno velik i vizualno raznovrstan negativni skup podataka. Problem takvog pristupa su memorijski zahtjevi dodatnog skupa podataka. U ovom poglavlju predlažemo metodu temeljenu na Lee et al. (2018) koja poboljšava performanse diskriminativnog modela u detekciji izvandistribucijskih primjera korištenjem primjera s ruba distribucije generiranih uzorkovanjem modela Real NVP.

4.1. Predložena metoda

U Lee et al. (2018) autori poboljšavaju performanse diskriminativnog modela u detekciji izvandistribucijskih primjera optimirajući gubitak:

$$L(\theta) = \mathbb{E}_{(\hat{\mathbf{x}},\hat{y})\sim P_{in}}[-\log P_\theta(y=\hat{y}|\hat{\mathbf{x}})] + \lambda\, \mathbb{E}_{\mathbf{x}\sim P_{bord}}[\mathrm{KL}(P_\theta(y|\mathbf{x}), U)] \tag{4.1}$$

gdje P_{in} predstavlja primjere iz skupa za učenje, a P_{bord} predstavlja primjere s ruba distribucije koje dobijemo uzorkovanjem GAN-a. Parametar β mora biti veći od 0.

Predložena metoda optimira isti gubitak. Razlika je u tome što su primjeri s ruba distribucije nastali uzorkovanjem generativnog modela Real NVP Dinh et al. (2017). Korištenjem ovog modela gubimo potrebu za diskriminatorom, što pojednostavljuje metodu. Real NVP učimo maksimizacijom izglednosti. Prema Lucas et al. (2019), maksimizacija izglednosti je orijentirana k pokrivenosti distribucije podataka, dok je suparničko učenje orijentirano ka kvaliteti primjera. Posljedično, zaključujemo da bi

generiranje primjera s ruba distribucije s Real NVP-om rezultiralo vizualno raznovrsnijim primjerima od primjera dobivenih GAN-om. Prema Hendrycks et al. (2019b), vizualna raznovrsnost primjera pozitivno utječe na poboljšanje performansi u detekciji izvandistribucijskih primjera. Izbjegavanjem GAN-a izbjegavamo i pojavu *mode collapse* koja zanemaruje određene modove distribucije podataka pri učenju distribucije modela, a tipična je za GAN. Predloženu metodu definiramo sljedećim algoritmom:

Algoritam 1: Procedura za poboljšanje performansi diskriminativnog modela u detekciji izvandistribucijskih primjera.

zahtjevaj $\beta > 0$

definiraj Real NVP: $\mathbf{z} = \mathbf{f}_{\theta_R}(\mathbf{x}), \mathbf{x} = \mathbf{f}_{\theta_R}^{-1}(\mathbf{z})$

definiraj klasifikator: $P_{\theta_C}(y|\mathbf{x})$

definiraj $optimizator_{\mathrm{R}}, optimizator_{\mathrm{C}}$

repeat

  x, y = dohvati_minigrupu()

  z = uzorkuj $N(0, 1)$

  $L_{\text{cls}} = -[\log P_{\theta_C}(y|\mathbf{x})] + \beta_{\text{cls}} \left[\mathrm{KL}(P_{\theta_C}(y|\mathbf{f}_{\theta_R}^{-1}(\mathbf{z})) \,||\, U)\right]$

  $L_{\text{rnvp}} = -[\log p_z(\mathbf{f}_{\theta_R}(\mathbf{x})) + \log \left(\left| \det \left(\frac{\delta \mathbf{f}_{\theta_R}(\mathbf{x})}{\delta \mathbf{x}} \right) \right| \right)]$

  $\theta_{\mathrm{R}} += optimizator_{\mathrm{R}}.update(\nabla L_{\text{rnvp}})$

  $\theta_{\mathrm{C}} += optimizator_{\mathrm{C}}.update(\nabla L_{\text{cls}})$

until *maksimalan broj iteracija*

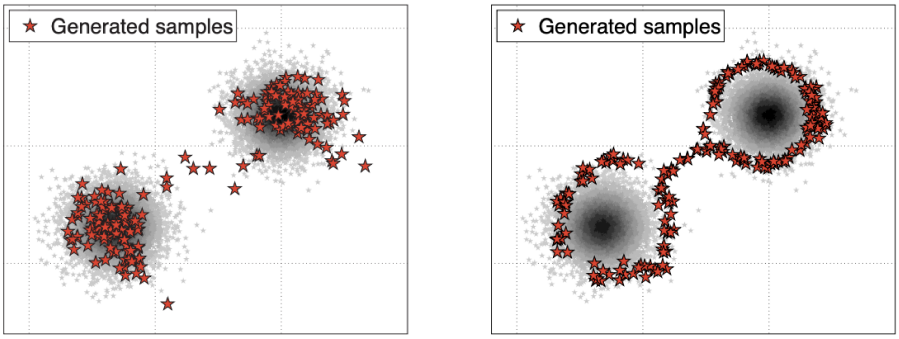
Učenje diskriminativnog modela procedurom definiranom algoritmom 1 rezultira diskriminativnim modelom s povećanim performansama u detekciji izvandistribucijskih primjera.

4.2. Generiranje primjera s ruba distribucije

Učenje diskriminativnog modela navedenom procedurom ima zanimljiv nusprodukt. Gubitak klasifikatora L_{cls} se propagira do Real NVP-a, što rezultira guranjem primjera koje generira Real NVP na rub distribucije podataka. Točnije, radi se o primjerima na

rubovima klasa. Poziciju sintetičkih primjera u višedimenzionalnom prostoru primjera nije jednostavno odrediti. Ipak, performanse diskriminativnog modela nam mogu dati dobru ideju o poziciji sintetičkih primjera. Ukoliko performanse diskriminativnog modela radikalno opadaju, dodatni član gubitka tjera probabilistički izlaz diskriminativnog modela k uniformnoj razdiobi za primjere koji su jako slični primjerima iz skupa za učenje. Primjeri s ruba distribucije ne bi trebali značajno remetiti performanse diskriminativnog modela.

Slika 4.1 lijevo prikazuje položaj generiranih unutardistribucijskih primjera (crveno) u odnosu na primjere iz skupa podataka (sivo). Slika 4.1 desno prikazuje položaj generiranih primjera s ruba distribucije (crveno) u odnosu na primjere iz skupa podataka (sivo). Primjeri s ruba distribucije su nastali propagiranjem gradijenta gubitka klasifikatora (izraz 4.1) do generativnog modela.



Slika 4.1: Lijevo: generirani unutardistribucijski primjeri. Desno: Generirani primjeri s ruba distribucije. Izvor: Lee et al. (2018)

Sintetičke primjere s ruba distribucije koristimo pri učenju modela za istovremenu semantičku segmentaciju i detekciju izvandistribucijskih dijelova slike. Generativni model prvo naučimo korištenjem algoritma 1, a zatim uzorkovanjem distribucije modela dobivamo skup koji sadrži primjere s ruba distribucije.

5. Modeli za semantičku segmentaciju

Problem semantičke segmentacije zahtjeva klasifikaciju svakog piksela slike u određenu klasu. Neki problem tipično degenerira u problem semantičke segmentacije ukoliko je izrazito bitno točno odrediti granice između objekata na slici. Najčešće se radi o medicinskim slikama ili slikama vožnje. Za uspješno rješavanje ovog problema potreban nam je označen skup podataka s gustim oznakama, što znači da semantička segmentacija pripada domeni nadziranog učenja. S obzirom da postoje javno dostupni označeni skupovi podataka, problem relativno uspješno rješavamo diskriminativnim modelima čiji pregled arhitektura nudimo u nastavku.

5.1. Općeniti pregled arhitektura

Najbolja rješenja za problem semantičke segmentacije koriste duboke modele s enkoder-dekoder strukturom baziranom na konvolucijskim slojevima. Modeli u enkoderskom dijelu snižavaju prostornu rezoluciju tenzora, a zatim u dekoderskom dijelu naduzorkuju tenzor vraćajući ga na početnu rezoluciju. Djelomična motivacija za ovakvu strukturu leži u memorijskom ograničenju suvremenih procesorskih jedinica.

Arhitekture modela sa simetričnim enkoderskim i dekoderskim dijelom Ronneberger et al. (2015) postižu brzinu koja nije dovoljna za segmentaciju slika u stvarnom vremenu. Stoga, suvremene arhitekture imaju asimetričan enkoderski i dekoderski dio. Kao enkoderski dio često odabiremo arhitekture za klasifikaciju poput arhitektura ResNet He et al. (2016a,b), DenseNet Huang et al. (2017), MobileNet V2 Sandler et al. (2018) i druge. Korištenje navedenih arhitektura u enkoderskom dijelu omogućuje inicijalizaciju modela ImageNet težinama, što je oblik regularizacije. Dekoderski dio modela postupno naduzorkuje latentni tenozor do početne rezolucije. Naduzorko-

vanje se provodi korištenjem interpolacijskih slojeva ili korištenjnem dekonvolucije. Dodatno, Krapac et al. (2017); Kreso et al. (2019); Orsic et al. (2019) stapaju značajke enkoderskog i dekoderskog dijela iste rezolucije što rezultira ljestvičastim modelom. U Pohlen et al. (2017) autori koriste dodatan tok koji propagira tenzor početne rezolucije cijelom dužinom modela. Općenito, dekoderski dio ima manji broj parametara od enkoderskog dijela. U ovom radu, za problem semantičke segmentacije koristimo model Ladder DenseNet Kreso et al. (2019) te ćemo ga detaljnije opisati u nastavku.

5.2. Arhitektura Ladder DenseNet

U ovom poglavlju nudimo uvid u arhitekturu modela za semantičku segmentaciju korištenog u provedenim eksperimentima. Ladder DenseNet Kreso et al. (2019) je duboki model temeljen na enkoder-dekoder strukturi s dodatnim lateralnim vezama.

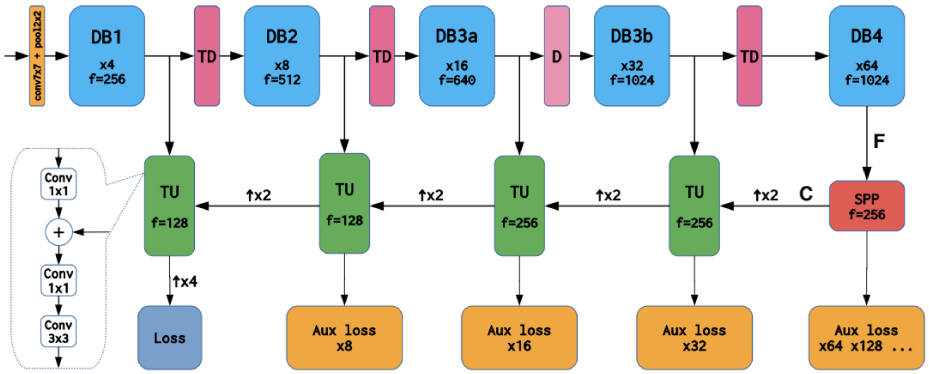
Enkoderski dio arhitekture koristi DenseNet Huang et al. (2017) ekstraktor značajki. DenseNet arhitektura se temelji na gustim slojevima (*eng*. dense layer) koji su građeni od sljedećeg niza transformacija: normalizacija po grupi Ioffe i Szegedy (2015), nelinearna $ReLU$ aktivacija, konvolucijski sloj s jezgrom 1×1, normalizacija po grupi, nelinearna $ReLU$ aktivacija te konačno konvolucijski sloj s jezgrom 3×3. Ovakav niz transformacija nalazimo i kod rezidualnih jedinica s predaktivacijom He et al. (2016b). Gusti blok (*eng*. dense block) nastaje slaganjem nekoliko gustih slojeva. Svaki gusti sloj na svoj ulaz dobije izlaz svih prethodinih slojeva u istom bloku. Arhitektura DenseNet se sastoji od 4 gusta bloka. Također, između dva susjedna gusta bloka se umeće tranzicijski blok (*eng*. transition block) koji se sastoji od normalizacije po grupi, nelinearne $ReLU$ aktivacije, konvolucijskog sloja s jezgrom 1×1 te sažimanja srednjom vrijednosti s oknom dimenzija 2×2 i korakom 2. Tranzicijski blok smanjuje prostorne dimenzije tenzora za pola.

Tenzor latentnih značajki dobivenih enkoderom uzorkujemo piramidalnim prostornim sažimanjem (*eng*. spatial pyramid pooling) He et al. (2015). Takozvani SPP modul sažima latentni tenzor koristeći različite veličine okna, što rezultira ekstrahiranjem jednakog broja značajki neovisno o veličini primjera. Veličine okana se odabiru dinamički, ovisno o prostornim dimenzijama latentnog tenzora koji predajemo SPP

modulu. U klasičnom strojnom učenju ovaj postupak je poznat kao piramidalno prostorno sparivanje (*eng*. spatial pyramid matching) te se koristi u pripadajućoj jezgrenoj funkciji Grauman i Darrell (2005). Ladder DenseNet koristi SPP modul koji latentni tenzor prvo projicira u tenzor s duplo manjim brojem mapa značajki, a zatim dobiveni tenzor višekratno sažima na način da pripadajuća rešetka ima 1, 2, 4 i 8 redova i stupaca. Svakom od sažimanja prethodi projekcija koja se sastoji od normalizacije po grupi, nelinearne $ReLU$ aktivacije i konvolucije s jezgrom 1×1. Konačno, dobivenoj latentnoj reprezentaciji povećamo broj mapa značajki koristeći već navedeni niz transformacija.

Dekoderski dio arhitekture Ladder DenseNet se sastoji od četiri bloka naduzorkovanja (*eng*. transition-up block). U svakom bloku naduzorkovanja se miješaju latentna reprezentacija iz prethodnog bloka naduzorkovanja (ili SPP modula) i latentna reprezentacija iz dekoderskog dijela koja je dostupna zbog preskočne veze. Latentnu reprezentaciju iz prethodnog bloka naduzorkovanja bilinearno interpoliramo na dvostruko veću rezoluciju te zbrojimo s latentnim tenzorom iz dekoderskog dijela koji prethodno projiciramo na jedaki broj mapa značajki. Konačno, rezultantni tenzor konvoluiramo s jezgrom dimenzije 3×3.

Slika 5.1 prikazuje arhitekturu Ladder DenseNet. Dekoderski dio se sastoji od gustih i tranzicijskih blokova (obojanih plavom odnosno svijetlocrvenom), dok se enkoderski dio sastoji od blokova naduzorkovanja (obojanih zelenom). SPP modul je označen crvenom bojom. Strelice označavaju tok podataka.



Slika 5.1: Arhitektura Ladder DenseNet121. Izvor: Kreso et al. (2019)

Konkretni primjeri ove arhitekture su Ladder DenseNet-121 i Ladder DenseNet-

169. Razlika između ove dvije arhitekture je u enkoderskom dijelu, gdje je u prvom slučaju enkoder DenseNet-121 a u drugom DenseNet-169. Glavna razlika između ova dva enkodera je u broju mapa značajki latentnih reprezentacija.

Pri učenju modela za semantičku segmentaciju osnovnom gubitku su često dodani pomoćni gubitci. Pomoćni gubitak je najčešće unakrsna entropija između latentne reprezentacije na nekoj od rezolucija dekoderskog dijela, koji je naduzorkovan do početne rezolucije i oznaka danog primjera. Na slici 5.1 su pomoćni gubitci označeni svijetložutom bojom.

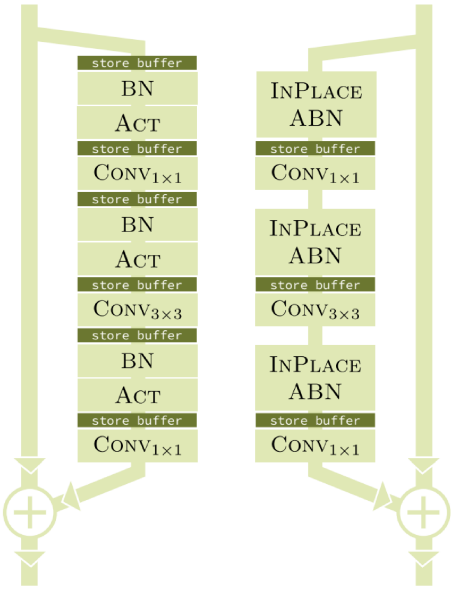
5.3. Optimizacija memorijskog otiska arhitekture Ladder DenseNet

Zbog svoje strukture, modeli za semantičku segmentaciju zahtijevaju veliku količinu memorije, što uz pretpostavku konstantne procesorske moći rezultira smanjenjem broja primjera u mini-grupi. Manji broj primjera u mini-grupi kvari procjenu gradijenta, što može rezultirati konvergiranjem optimizacijskog postupka u suboptimalno riješenje. Također, procjena srednje vrijednosti i očekivanja u sloju normalizacije po grupi se drastično pogoršava sa smanjenjem broja primjera. Suvremene arhitekture jako ovise o normalizaciji po grupi, stoga smanjenje veličine mini-grupe značajno utječe na performanse modela.

Veličinu mini-grupe možemo povećati tako da smanjimo memorijski otisak modela tijekom unaprijednog prolaska odnosno prolaska unatrag. Memorijski otisak modela smanjujemo pametnim keširanjem aktivacija koje zatim koristimo za izračunavanje preostalih aktivacija koje nisu keširane. Ovaj postupak nazivamo gradient checkpointing Chen et al. (2016). Dio računskog grafa između keširanih aktivacija nazivamo segment. Pri prolasku unatrag kroz model, za svaki segment se na temelju keširane aktivacije izračunaju preostale aktivacije (prolaz unaprijed), a zatim se nanovo izračunate aktivacije koriste za izračun gradijenata u segmentu (prolaz unatrag). Konačno, otpuštaju se zauzeti keševi te se postupak ponavlja za sljedeći segment računskog grafa. Mana ovakvog pristupa je povećanje vremena potrebnog za izračun gradijenata uzrokovan prolaskom unaprijed kroz segment grafa potreban za izračun nekeširanih aktiva-

cija. Modelu korištenom u provedenim eksperimentima keširamo aktivacije na razini gustog sloja, čime se veličina mini-grupe poveća za faktor 2 na istoj hardverskoj arhitekturi.

Bulò et al. (2018) smanjuje memorijski otisak dubokog modela stapanjem normalizacije po grupi i nelinearne aktivacije u jedinstvenu transformaciju nazvanu InPlace-ABN. S obzirom da moderne arhitekture intenzivno koriste navedni niz transformacija (gusti sloj, rezidualna jedinica s predaktivacijom), na ovaj način se smanjuje memorijski otisak modela za do 50%. Slika 5.2 prikazuje razliku između standardnog keširanja aktivacija rezidualne jedinice s predaktivacijom i uskim grlom He et al. (2016b) te keširanja aktivacija kada se primjeni InPlace-ABN.



Slika 5.2: Razliku između standardnog keširanja aktivacija rezidualne jedinice s predaktivacijom i uskim grlom He et al. (2016b) te keširanja aktivacija kada se primjeni InPlace-ABN. Izvor: Bulò et al. (2018)

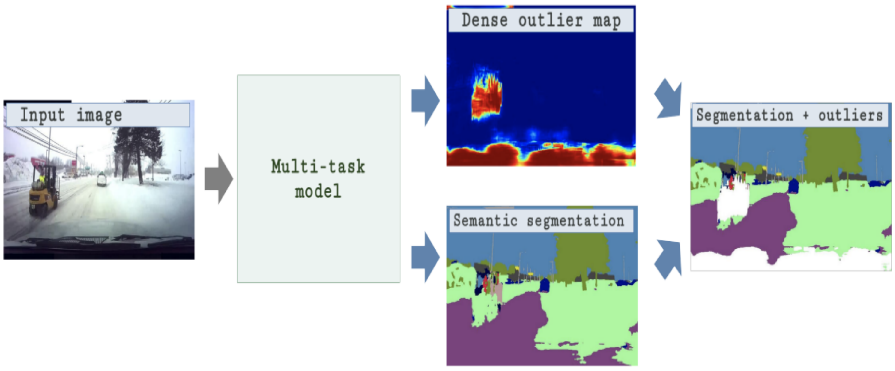
Stapanje nelinearne aktivacije i normalizacije po grupi u jedinstveni sloj je moguće uz uvjet da aktivacijska funkcija ima inverz. Nelinearna aktivacija $ReLU$ ne ispunjava ovaj uvjet, međutim $ReLU$ se može zamijeniti sa sličnom aktivacijom $Leaky\ ReLU$ s nagibom 0.01, bez značajnog pada performansi modela Bulò et al. (2018). Memorijski otisak arhitekture Ladder DenseNet dodatno smanjujemo zamjenom sloja $ReLU$ aktivacije i normalizacije po grupi sa slojem InPlace-ABN u svim gustim blokovima

enkoderskog dijela modela. Model s navedenim optimizacijama koristimo u provedenim eksperimentima koji se tiču istovremene semantičke segmentacije i detekcije primjera s ruba distribucije na četvrtini rezolucije, dok na pola rezolucije koristimo model koji se oslanja samo na agresivno keširanje aktivacija.

6. Gusta detekcija odudarajućih (izvandistribucijskih) dijelova slike

Gustu detekciju odudarajućih dijelova slike vršimo paralelno uz semantičku segmentaciju. Motivacija za ovaj problem dolazi iz stvarnog svijeta, gdje sustavi za autonomnu vožnju nailaze na atipične objekte na cesti koji mogu ozlijediti putnike ili izazvati štetu na vozilu. Predlažemo riješenje ovog problema na način da prikladno modificiramo arhitekturu Ladder DenseNet te dodamo anomalije u skup podataka za učenje.

Slika 6.1 prikazuje generalnu ideju paralelne semantičke segmentacije i detekcije odudarajućih dijelova slike. Model sposoban paralelno obaviti oba zadatka generira mapu s gustom predikcijom te mapu s označenim odudarajućim dijelovima slike. Stapanje obiju mapa rezultira mapom s gustom predikcijom i označenim odudarajućim dijelovima.



Slika 6.1: Generalna ideja paralelne semantičke segmentacije i detekcije odudarajućih dijelova slike. Izvor: Bevandic et al. (2019)

6.1. Modifikacija skupa za učenje

Otvoreni javno dostupni skupovi podataka ne pokrivaju kompleksne situacije iz stvarnog svijeta poput smanjene vidljivosti, sudara, stranih objekata na cesti i drugih. Iz tog razloga skup podataka koji sadrži slike vožnje želimo modificirati na način da dodamo anomalije koje imitiraju izazovne situacije stvarnog svijeta. Bevandic et al. (2019) modificira skup podataka lijepljenjem ImageNet objekata na slučajno odabrano mjesto u slici. U ovom radu u slike lijepimo dijelove ImageNet objekata, sintetičke primjere s ruba distribucije te mješavinu obiju vrsta anomalija. Sintetičke primjere s ruba distribucije dobijemo uzorkovanjem Real NVP-a treniranog algoritmom 1. S obzirom da su nam dostupne oznake za svaki piksel, odabiremo diskriminativni model za gustu predikciju. Skup podataka korišten za učenje Real NVP-a je pretprocesiran na način da primjeri prikazuju dijelove scene poput oblaka, grma ili dijela kolnika. Detaljniji opis dobivanja primjera s ruba distribucije se nalazi u poglavlju s eksperimentima. Pogodna arhitektura Real NVP-a omogućava da generirani primjeri s ruba distribucije mogu varirati u dimenzijama, čime dodatno povećavamo robusnost modela na anomalije različitih dimenzija.

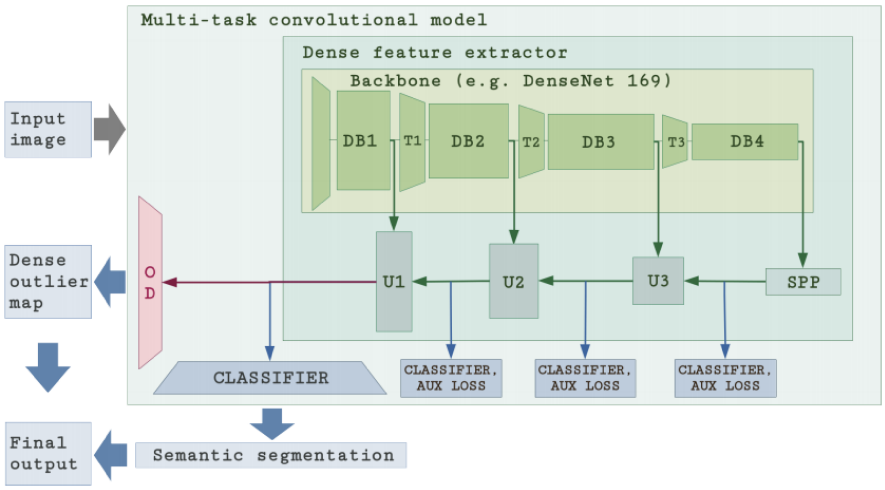
Slika 6.2 prikazuje primjer iz skupa za učenje nastalog lijepljenjem primjera s ruba distribucije u scenu vožnje iz skupa podataka Mapillary Vistas.

Slika 6.2: Primjer iz skupa za učenje nastalog lijepljenjem primjera s ruba distribucije u scenu vožnje.

6.2. Modeli za istovremenu semantičku segmentaciju i detekciju izvandistribucijskih dijelova slike

Kako bismo uspješno obavili paralelnu semantičku segmentaciju i detekciju izvandistribucijskih primjera, potrebno je modificirati arhitekturu Ladder DenseNet-a te gubitak koji optimiramo. U provedenim eksperimentima koristimo arhitekture slične kao Bevandic et al. (2019).

Model s dvije glave nastaje tako da standardnoj arhitekturi Ladder DenseNet dodamo dodatnu glavu čiji probabilistički izlaz interpretiramo kao vjerojatnost da primjer pripada skupu podataka na kojem je model naučen. U radu koristimo enkodere DenseNet-121 te DenseNet-169. Slika 6.3 prikazuje arhitekturu Ladder DenseNet s dodatnom glavom za detekciju izvandistribucijskih dijelova slike.



Slika 6.3: Arhitektura Ladder DenseNet s dodatnom glavom za detekciju izvandistribucijskih dijelova slike. Izvor: Bevandic et al. (2019)

Gubitak modela definiramo izrazom:

$$L(\theta, \gamma) = -\sum_i \sum_j \log P_\theta(\mathbf{y}_{i,j}|\mathbf{x}) - \lambda \sum_i \sum_j \log P_\gamma(\mathbf{z}_{i,j}|\mathbf{x}) \tag{6.1}$$

gdje θ predstavlja parametre glave za semantičku segmentaciju, a γ parametre glave za detekciju anomalija. $\mathbf{y}$ predstavlja oznake skupa podataka za učenje, dok $\mathbf{z}$ poprima

vrijednost 1 za izvandistribucijske piksele, a 0 za unutardistribucijske piksele. Parametar λ je konstanta veća od 0.

Model s jednom glavom zadržava jednaku arhitekturu kao standardni Ladder DenseNet. Detekciju izvandistribucijskih dijelova slike temeljimo na maksimalnoj vrijednosti probabilističkog izlaza modela. Još jednom koristimo enkodere DenseNet-121 te DenseNet-169. Gubitak modela modificiramo na način da probabilistički izlaz modela guramo k uniformnoj razdiobi za izvandistribucijske dijelove slike. Gubitak definiramo izrazom:

$$L(\theta) = -\sum_i \sum_j \log P_\theta(\mathbf{y}_{i,j}|\mathbf{x}) + \lambda \sum_i \sum_j [\![\mathbf{z}_{i,j} = 1]\!] \, \mathrm{KL}(P_\theta(\mathbf{y}_{i,j}|\mathbf{x}) \,||\, U) \qquad (6.2)$$

gdje U predstavlja uniformnu razdiobu po klasama skupa podataka. Slično kao u prethodnom slučaju, **y** predstavlja oznake skupa podataka za učenje, **z** poprima vrijednost 1 za izvandistribucijske piksele, a 0 za unutardistribucijske piksele. Parametar λ je veći od 0.

Postupak učenja i dobiveni rezultati obje vrste modela su detaljno opisani u poglavlju s eksperimentima.

6.3. Evaluacija performansi modela

Model učen minimizacijom gubitaka (6.1) i (6.2) rezultira robusnijim modelom sposobnim detektirati dijelove slike koji sadrže anomaliju. Mjerenje robusnosti modela nije jednostavan zadatak. Primjeren skup podataka bi trebao sadržavati reprezentativne slike iz otvorenog svijeta kao i odudarajuće slike. Zbog varijabilnosti scene u otvorenom svijetu, ovakav skup podatak bi trebao biti relativno velik.

Trenutno najbolji skup podataka namijenjen evaluaciji robusnosti modela je Wildash Zendel et al. (2018). Ovaj skup podataka sadrži slike vožnje iz otvorenog svijeta te negativne slike koje ne sadrže scene vožnje. Na žalost, ovaj skup podataka sadrži mali broj slika te nedovoljno pokriva situacije otvorenog svijeta. Ipak, i takav skup podataka predstavlja izazov za suvremene modele. Slika 6.4 prikazuje slike vožnje iz skupa Wilddash sa smanjenom vidljivošću.

Slika 6.4: Primjer slika smanjene vidljivosti iz skupa podataka Wilddash. Izvor: Zendel et al. (2018)

Bevandic et al. (2019) unosi dodatne anomalije u skup podataka Wilddash lijepljenjem životinja iz skupa podataka Pascal VOC 2007 Everingham et al. (2010) u scenu vožnje. Posljedično, inicijalno kompleksne scene iz Wilddash skupa postaju reprezentativnije za evaluaciju performansi modela u detekciji odudarajućih dijelova slike u uvjetima otvorenog svijeta. Dobiveni skup podataka se naziva WD-Pascal. Mana ovog skupa podataka je u tome što na slikama vidimo artefakte lijepljenja. Slika 6.5 prikazuje primjere iz skupa podataka WD-Pascal. U ovom radu, performanse modela u detekciji odudarajućih dijelova slike evaluiramo na skupu podataka WD-Pascal.

Slika 6.5: Primjer slika iz skupa podataka WD-Pascal. Izvor: Bevandic et al. (2019)

Slično kao WD-Pascal, skup podataka Fishyscapes Blum et al. (2019) nastaje lijepljenjem izvandistribucijskih objekata na scenu vožnje iz skupa podataka Cityscapes Cordts et al. (2016). Autori vode računa da se sisavci i ptice nalaze na donjem dijelu scene, čime se imitira događaj životinje na cesti. Na gornji dio scene se lijepe ptice i avioni, tj. karakteristični objekti za nebo.

U Hendrycks et al. (2019a) autori mjere performanse modela u detekciji anomalija na modificiranom skupu podataka BDD Yu et al. (2018). Skup podataka BDD koji sadrži slike vožnje u gradovima SAD-a modificiraju tako da sve slike koje sadrže

motocikl i tramvaj ekstrahiraju u testni skup, gdje su navedene klase definirane kao anomalije. Na ovaj način autori izbjegavaju artefakte lijepljenja. Mana ovog pristupa je u tome što su instance klasa označenih kao anomalije previše slične ostalim vozilima. Također, skup podataka BDD sadrži pogrešno označene dijelove scena.

Na tragu ovog pristupa, u ovom radu mjerimo performanse modela u detekciji odudarajućih dijelova slike na skupu Mapillary Vistas kojem klase mapiramo na klase iz skupa podataka Cityscapes. Instance klasa vozač (bicikla ili motocikla) i osoba definiramo kao anomalije. Slike koje sadrže navedene klase ekstrahiramo u testni skup podataka, čime nam u skupu za učenje ostaje $8\,045$ slika, a u skupu za validaciju 836 slika. Testni skup sadrži $11\,119$ slika s dobro distribuiranim anomalijama u prirodnim položajima. Rezultate modela na ovom skupu podataka prikazujemo u poglavlju s eksperimentima.

Konačno, performanse modela u detekciji malih objekata na scenama vožnje evaluiramo na skupu podataka FS Lost&Found Blum et al. (2019). Navedeni skup podataka sadrži scene u kojima se na kolniku nalaze anomalije poput kutija i igrački. S obzirom da se radi o vrlo malim objektima, skup podataka FS Lost&Found je među izazovnijim skupovima podataka. U ovom radu mjerimo perfromanse određenih modela na ovom skupu podataka, što je prikazano u poglavlju s eksperimentima.

7. Eksperimenti

U ovom poglavlju opisujemo provedene eksperimente te analiziramo dobivene rezultate. Potpoglavlje 7.1 sadrži eksperimente koji opisuju detekciju izvandstribucijskih primjera, dok se potpoglavlja 7.2 i 7.3 tiču detekcije izvandistribucijskih dijelova slike.

7.1. Detekcija izvandistribucijskih primjera

Performanse klasifikatora u detekciju izvandistribucijskih primjera korištenjem metode definirane u potpoglavlju 4.1 uspoređujemo s baznim modelom učenim unakrsnom entropijom i trenutno najboljim rješenjem opisanim u Hendrycks et al. (2019b). U svim metodama koristimo model Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) temeljen na arhitekturi DenseNet. Rezultate prikazujemo na skupovima podataka CIFAR10 Krizhevsky et al. i SVHN Netzer et al. (2011). Kao izvandistribucijske skupove podataka koristimo LSUN Yu et al. (2015), Tiny-ImageNet [1] i TrafficSign Stallkamp et al. (2011).

7.1.1. Rezultati na skupu podataka CIFAR10

Tablica 7.1 prikazuje rezultate klasifikatora Wide-DenseNet-BC (L=40, k=48) učenog unakrsnom entropijom. Klasifikator je učen 175 epoha s veličinom mini-grupe koja iznosi 64. Korišteni optimizacijski algoritam je SGD s Nesterovim momentom i stopom učenja postavljenom na 0.1. Težine propadaju s faktorom 10^{-4}. Klasifikator transformiramo u detektor izvandistribucijskih primjera koristeći maksimalnu vrijednost probabilističkog izlaza modela. Ove rezultate koristimo kao referencu za usporedbu

[1] `https://tiny-imagenet.herokuapp.com/`

ostalih rješenja. Opis evaluacijskih metrika se nalazi u Lee et al. (2018).

Točnost na skupu CIFAR10: 91.77				
OOD dataset	SVHN	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	89.95	54.34	60.65	49.23
AUROC	98.22	92.98	94.50	84.18
Točnost detekcije	93.64	86.41	88.88	78.20
AUPR-In	98.47	94.37	95.71	80.43
AUPR-Out	97.92	90.91	92.54	86.51

Tablica 7.1: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) učenog s gubitkom unakrsne entropije. Klasifikator je transformiran u detektor izvandistribucijskih primjera koristeći maksimalnu vrijednost probabilističkog izlaza modela.

Tablica 7.2 prikazuje rezultate klasifikatora inicijaliziranog težinama baznog modela kojeg fino ugađamo 10 epoha optimirajući gubitak 4.1. Parametar β je postavljen na 1. Korišteni negativni skup podataka sadrži 80 milijuna slika objekata Torralba et al. (2008). Provedeni eksperiment odgovara metodi definiranoj u Hendrycks et al. (2019b), a korišteni kod[2] je javno objavljen od istih autora. Vidljivo je povećanje klasifikacijske sposobnosti kao i poboljšanje performansi klasifikatora u detekciji izvandistribucijskih primjera u odnosu na bazni model.

Točnost na skupu CIFAR10: 94.80 (+3.03)				
OOD dataset	SVHN	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	99.99 (+10.0)	99.14 (+44.8)	98.61 (+37.9)	95.01 (+45.8)
AUROC	99.94 (+1.7)	99.71 (+6.7)	99.35 (+4.8)	99.18 (+15.0)
Točnost detekcije	99.43 (+5.7)	97.86 (+11.4)	96.98 (+8.1)	95.31 (+17.1)
AUPR-In	99.96 (+1.4)	99.74 (+5.3)	99.48(+3.7)	99.24 (+18.8)
AUPR-Out	99.92 (+2.0)	99.67 (+8.7)	99.18 (+6.6)	99.11 (+12.6)

Tablica 7.2: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) fino ugođenog metodom definiranom u Hendrycks et al. (2019b).

[2] `https://github.com/hendrycks/outlier-exposure`

Tablica 7.3 prikazuje rezultate klasifikatora inicijaliziranog težinama baznog modela optimirajući gubitak 4.1. Umjesto negativnog skupa podataka koristimo primjere s ruba distribucije tj. koristimo metodu opisanu u potpoglavlju 4.1. Parametar β je postavljen na 1. Još jednom klasifikator fino ugađamo 10 epoha. Arhitektura korištenog Real NVPja ima 3 rezidualne jedinice s 32 dimenzije po sloju, a poduzorkovanje se provodi 3 puta. Real NVP optimiramo korištenjem algoritma ADAM Kingma i Ba (2015) sa stopom učenja 0.01. Vidimo povećanje performansi klasifikatora u oba zadatka u usporedbi s baznim modelom. U usporedbi s metodom definiranom u Hendrycks et al. (2019b), naša metoda ima lošije rezultate, što prepisujemo manjoj vizualnoj raznolikosti sintetičkih primjera u odnosu na primjere iz dodatnog negativnog skupa podataka.

Točnost na skupu CIFAR10: 94.08 (+2.31)				
OOD dataset	SVHN	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	98.34 (+8.3)	92.02 (+37.6)	94.26 (+33.6)	69.81 (+20.5)
AUROC	99.70 (+1.4)	98.53 (+5.5)	98.75 (+4.2)	95.52 (+11.3)
Točnost detekcije	97.18 (+3.5)	93.76 (+7.3)	94.79 (+5.9)	88.49 (+10.2)
AUPR-In	99.71 (+1.2)	98.70 (+4.3)	98.94 (+3.2)	96.27 (+15.8)
AUPR-Out	99.71 (+1.7)	98.37 (+7.4)	98.53 (+5.9)	95.09 (+8.5)

Tablica 7.3: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) fino ugođenog metodom definiranom u potpoglavlju 4.1.

7.1.2. Rezultati na skupu podataka SVHN

Tablica 7.4 prikazuje performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) učenog unakrsnom entropijom na skupu podataka SVHN. Klasifikator je učen 40 epoha s mini-grupom koja sadrži 64 primjera. Korišteni optimizacijski algoritam je SGD s Nesterovim momentom i stopom učenja postavljenom na 0.1. Težine propadaju s faktorom 10^{-4}. Naučeni klasifikator transformiramo u detektor izvandistribucijskih primjera koristeći maksimalnu vrijednost probabilističkog izlaza modela (max-softmax). Performanse ovog modela koristimo kao referencu za usporedbu ostalih dviju metoda.

Točnost na skupu SVHN: 95.10				
OOD dataset	CIFAR10	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	81.64	76.99	81.07	57.00
AUROC	97.35	96.77	97.27	89.94
Točnost detekcije	92.07	91.62	92.15	82.71
AUPR-In	97.82	97.44	97.78	89.37
AUPR-Out	96.89	95.94	96.68	89.74

Tablica 7.4: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) učenog s gubitkom unakrsne entropije. Klasifikator je transformiran u detektor izvandistribucijskih primjera koristeći maksimalnu vrijednost probabilističkog izlaza modela.

Tablica 7.5 prikazuje rezultate klasifikatora inicijaliziranog težinama baznog modela kojeg fino ugađamo 5 epoha. Optimiramo gubitak definiran izrazom 4.1. Parametar β je postavljen na 1. Korišteni negativni skup podataka sadrži 80 milijuna slika objekata Torralba et al. (2008). Provedeni eksperiment odgovara metodi definiranoj u Hendrycks et al. (2019b). Vidljivo je povećanje klasifikacijske sposobnosti kao i poboljšanje performansi klasifikatora u detekciji izvandistribucijskih primjera u odnosu na bazni model.

Točnost na skupu SVHN: 95.95 (+0.85)				
OOD dataset	CIFAR10	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	99.35 (+17.7)	99.87 (+22.8)	99.82 (+18.7)	87.92 (+30.9)
AUROC	99.78 (+2.4)	99.93 (+3.1)	99.91 (+2.6)	98.09 (+8.1)
Točnost detekcije	97.85 (+5.7)	99.07 (+7.4)	98.78 (+6.6)	93.54 (+10.8)
AUPR-In	99.80 (+1.9)	99.94 (+2.5)	99.92 (+2.1)	98.45 (+9.0)
AUPR-Out	99.74 (+2.8)	99.90 (+3.9)	99.88 (+3.2)	97.57 (+7.8)

Tablica 7.5: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) fino ugođenog metodom definiranom u Hendrycks et al. (2019b).

Tablica 7.6 prikazuje rezultate klasifikatora inicijaliziranog težinama baznog modela optimirajući gubitak 4.1. Umjesto negativnog skupa podataka koristimo primjere

s ruba distribucije tj. koristimo metodu opisanu u potpoglavlju 4.1. Parametar β je postavljen na 1. Klasifikator fino ugađamo 5 epoha. Arhitektura korištenog Real NVPja ima 3 rezidualne jedinice s 32 dimenzije po sloju, a poduzorkovanje se provodi 3 puta. Real NVP optimiramo korištenjem algoritma ADAM sa stopom učenja 0.01. Vidimo povećanje performansi klasifikatora u oba zadatka u usporedbi s baznim modelom. U usporedbi s metodom definiranom u Hendrycks et al. (2019b), naša metoda ima lošije rezultate, što prepisujemo manjoj vizualnoj raznolikosti sintetičkih primjera u odnosu na primjere iz dodatnog negativnog skupa podataka.

Točnost na skupu SVHN: 95.81 (+0.71)				
OOD dataset	CIFAR10	LSUN	Tiny-Imagenet	TrafficSign
TNR at TPR 95%	93.63 (+11.9)	93.67 (+16.6)	95.87 (+14.8)	58.42 (+1.4)
AUROC	98.56 (+1.2)	98.47 (+1.7)	98.84 (+1.5)	92.10 (+2.1)
Točnost detekcije	94.72 (+2.6)	94.63 (+3.0)	95.56 (+3.4)	84.65 (+1.9)
AUPR-In	98.80 (+0.9)	98.77 (+1.3)	99.03 (+1.2)	92.81 (+3.4)
AUPR-Out	98.16 (+1.2)	97.87 (+1.9)	98.44 (+1.7)	91.23 (+1.4)

Tablica 7.6: Performanse klasifikatora Wide-DenseNet-BC (L=40, k=48) Huang et al. (2017) fino ugođenog metodom definiranom u potpoglavlju 4.1.

7.2. Generiranje primjera s ruba distribucije

Primjere s ruba distribucije generiramo korištenjem algoritma 1. Unutardistribucijski skup podataka Mapillary Vistas sadrži oznake po pikselu, stoga u postupku učenja koristimo diskriminativni model za semantičku segmentaciju. Konkretno, koristimo Ladder DenseNet-121 s optimizacijama navedenim u potpoglavlju 5.3. Enkoderski dio modela je inicijaliziran ImageNet težinama. Arhitektura Real NVP-a se sastoji od 2 rezidualne jedinice s 32 dimenzije po sloju, dok se poduzorkovanje provodi 4 puta. Za optimizaciju parametara obaju modela koristimo algoritam ADAM sa stopom učenja postavljenom na vrijednost 0.001. Postupak učenja traje 50 epoha. Slike iz skupa podataka za učenje pretprocesiramo sljedećim nizom transformacija:

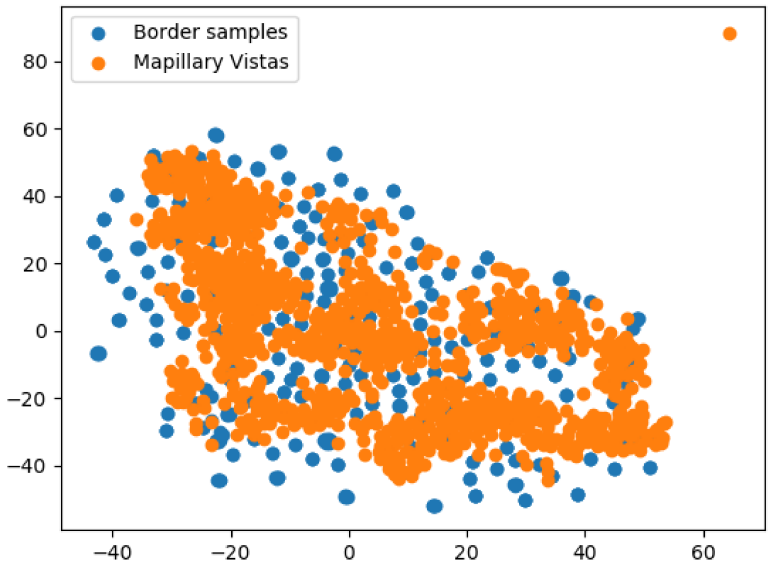
1. horizontalno zrcaljenje s vjerojatnošću 0.5
2. skaliranje slike na način da kraća stranica ima duljinu iz uniformnog intervala $[256, 512]$
3. izrezivanje nasumičnog odreska veličine 64×64 piksela

Posljednji korak se ponavlja ukoliko dobiveni odrezak ne sadrži barem 3 rijetke klase ali maksimalno 20 puta. Rijetke klase definiramo kao skup svih klasa osim klasa: cesta, zgrada, nebo, vegetacija i teren. Slika 7.1 prikazuje primjere dobivene uzorkovanjem naučenog generativnog modela.

Slika 7.1: Primjeri s ruba distribucije nastali uzorkovanjem Real NVP-a. Model je učen metodom definiranom algoritmom 1.

Naučeni Real NVP koristimo za generiranje skupa podataka s primjerima na rubu distribucije. Generirani skup podataka sadrži $50\,000$ primjera s dimenzijama $n \times n$ gdje je n slučajno odabran iz skupa $\{32, 64, 80, 96\}$. Ovaj skup primjera koristimo u procesu učenja modela za istovremenu semantičku segmentaciju i detekciju dijelova slike koji sadrže anomaliju.

Slika 7.2 prikazuje odnos unutardistribucijskih primjera koji pripadaju skupu Mapillary Vistas (označeni narančasto) i primjera s ruba distribucije generiranih Real NVP-em (označeni plavom). Primjeri dimenzija $3 \times 64 \times 64$ su reducirani na dvije dimenzije postupkom t-SNE van der Maaten i Hinton (2008). Parametar *perplexity* određujemo vizualnom inspekcijom te je u konačnici postavljen na 19. Vidimo da primjeri s ruba distribucije okružuju distribuciju podataka te popunjavaju pore između modova distribucije.



Slika 7.2: Prikaz unutardistribucijskih primjera iz skupa Mapillary Vistas i odgovarajućih primjera s ruba distribucije generiranih Real NVP-em. Dimenzionalnos primjera je snižena postupkom t-SNE.

7.3. Istovremena semantička segmentacija i detekcija izvadnistribucijskih dijelova slike

Učenje modela za istovremenu semantičku segmentaciju i detekciju izvadnistribucijskih dijelova slike izvodimo na skupu podataka Mapillary Vistas (bez slika koje sadrže klase vozač i osoba) s oznakama skupa podataka Cityscapes. Navedenom skupu podataka dodajemo anomalije lijepljenjem: i) sintetičkih primjera s ruba distribucije

dobivenih algoritmom 1 ii) dijelova slika skupa podataka ImageNet iii) kombinacije primjera prethodna dva skupa.

Mini-grupa primjera za učenje sadrži slike skupa podataka Mapilary Vistas s zalijepljenom anomalijom na nasumično odabrano mjesto (50% mini-grupe) te anomalijom naduzorkovanom do veličine primjera (50% mini-grupe). Naduzorkovanje se provodi postupkom bilinearne interpolacije. Slike iz skupa podataka Vistas pretprocesiramo tako da kraću stranicu smanjimo do 256 piksela, a zatim izrezujemo primjer dimenzija 256×256 s nasumično odabranog mjesta. Ovako dobivenom primjeru lijepimo anomaliju na nasumično odabrano mjesto. Anomaliju iz skupa podataka ImageNet dobijemo tako da nasumično izrežemo dio slike veličine 64×64 piksela, a zatim taj dio interpoliramo do veličine $n \times n$ gdje je n nasumično odabran iz skupa $\{32, 64, 80, 96\}$. Anomalije iz skupa s ruba distribucije lijepimo na nasumično odabrano mjesto bez dodatnog pretprocesiranja.

Na ovakvom skupu podataka učimo modele s jednom i dvije glave definirane u potpoglavlju 6.2 s enkoderom DenseNet-121. Veličina mini-grupe je 20. Parametar β je postavljen na 0.2 za model s dvije glave, dok za model s jednom glavom iznosi 10. Parametre optimiramo korištenjem optimizacijskog algoritma ADAM s inicijalnom stopom učenja $1 * 10^{-4}$ za enkoder te $4 * 10^{-4}$ za dekoder i SPP modul. Stopa učenja opada po kosinusnoj krivulji do vrijednosti 10^{-7}. Stope opadanja su postavljene na 0.9 i 0.999. Osnovni model učimo 75 epoha, a modele za istovremenu semantičku segmentaciju i detekciju izvandistribucijskih dijelova slike 150 epoha. Enkoderski dio modela je inicijaliziran s ImageNet težinama.

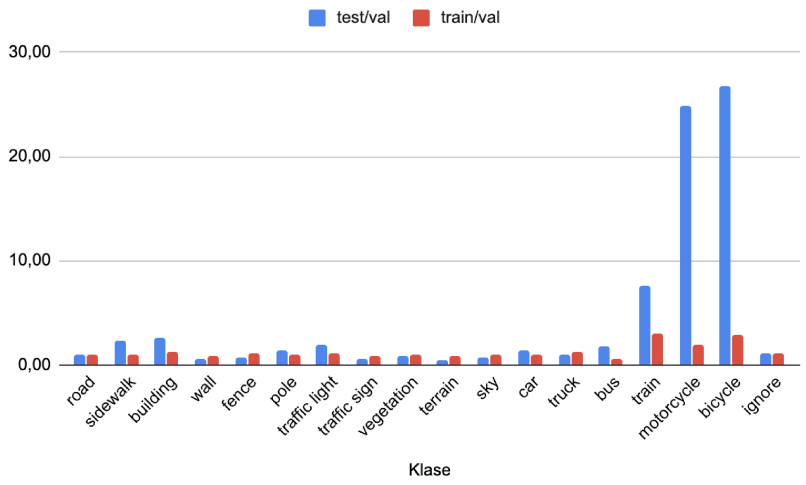
Tablica 7.7 prikazuje rezultate modela s dvije glave učenog na skupu podataka Mapillary Vistas bez klasa osoba i vozač. Stupci *Val mIoU* i *Test mIoU* prikazuju rezultat mIoU metrike na validacijskom odnosno testnom skupu. Stupac *AP Vistas* prikazuje vrijednost prosječne preciznosti na testnom skupu podataka, gdje su instance klase osoba i vozač označene kao anomalije. Stupac *AP WD-Pascal* prikazuje prosječnu preciznost na skupu podatak WD-Pascal. U svim stupcima veće vrijednosti predstavljaju bolju vrijednost. Osnovni model učen na skupu podataka kojem ne lijepimo anomalije postiže najbolje rezultate u segmentaciji slika. Model učen na skupu podataka kojem lijepimo anomalije iz skupa ImageNet postiže najbolje rezultate u detekciji anomalija u

skupu podataka WD-Pascal. Model učen na skupu podataka kojem lijepimo anomalije iz skupa podataka ImageNet te primjera s ruba distribucije postiže najbolje rezultate u detekciji anomalija na testnom skupu Vistas.

Vrsta anomalija	Val mIoU	Test mIoU	AP Vistas	AP WD-Pascal
Bez anomalije	**55.7**	**50.8**	7.1	7.7
Rub distribucije	52.0	45.1	18.6	33.3
ImageNet	49.7	43.9	20.3	**67.5**
ImageNet + Rub dist.	50.4	45.4	**20.5**	52.3

Tablica 7.7: Rezultati modela s dvije glave na skupu Mapillary Vistas bez klasa osoba i vozač.

Primjećujemo značajan pad segmentacijskih performansi između validacijskog i testnog skupa po svim modelima. Razlog za ovu pojavu leži u različitim udjelima klasa između različitih skupova. Na slici 7.3 crveni stupac prikazuje omjer udjela klasa između skupa za trening i skupa za validaciju. Plavi stupac prikazuje omjer udjela klasa između skupa za test i skupa za validaciju. Vidimo drastično povećanje udjela klasa vlak, motocikl i bicikl u testnom skupu podataka. Upravo na tim klasama model postiže najlošije rezultate, što rezultira padom mIoU na testnom skupu podataka.



Slika 7.3: Omjer udjela klasa između testnog i validacijskog skupa odnosno skupa za učenje i validacijskog skupa. Skup za učenje je dobro balansiran u odnosu na skup za validaciju, dok skup za test ima značajno veći udjel instanci klasa motocikl, bicikl i vlak.

Tablica 7.8 prikazuje rezultate modela s jednom glavom. Model je učen na jednak način na istom skupu podataka kao i model s dvije glave uz iznimku parametra β koji je postavljen na vrijednost 10. Pri evaluaciji modela koristimo temperaturno skaliranje softmaksa, gdje parametar T postavljamo na 10. Stupci tablice imaju jednako značenje kao u prethodnoj tablici. Vidimo da najbolje rezultate u segmentaciji postiže model učen na skupu podataka u koji lijepimo primjere s ruba distribucije. Kao i u Hendrycks et al. (2019b), opažamo da svi modeli imaju bolje segmentacijske performanse od baznog modela na testnom skupu podataka. Na skupu podataka WD-Pascal najbolje rezultate postiže model naučen na skupu podataka u koji lijepimo ImageNet anomalije. U detekciji anomalija na skupu podataka Vistas najbolje rezultate postiže model naučen na skupu podataka u koji lijepimo anomalije s ruba distribucije kao i ImageNet anomalije.

Vrsta anomalija	**Val mIoU**	**Test mIoU**	**AP Vistas**	**AP WD-Pascal**
Bez	55.7	50.8	7.1	7.7
Rub distribucije	**57.0**	**52.2**	17.5	27.2
ImageNet	54.6	51.8	17.0	**71.5**
ImageNet + Rub dist.	53.9	51.8	**17.8**	55.3

Tablica 7.8: Rezultati modela s jednom glavom na skupu podataka Mapillary Vistas bez klasa osoba i vozač.

Ako usporedimo modele iz tablica 7.7 i 7.8, vidimo da model s jednom glavom bolje detektira anomalije u skupu podataka WD-Pascal, dok model s dvije glave bolje detektira anomalije na skupu podataka Mapillary Vistas. Ako uzmemo u obzir korištene negativne primjere, oba modela koja učimo na primjerima s ruba distribucije te primjerima iz skupa ImageNet imaju najbolje performanse u detekciji anomalija bez artifakata lijepljenja. Pri usporedbi segmentacijskih performansi, model s jednom glavom postiže najbolji segmentacijski rezultat na testnom skupu podataka.

Sljedeće učimo model s jednom i dvije glave na slikama veće rezolucije. Slike iz skupa Mapillary Vistas smanjujemo tako da im kraća stranica iznosi 512 piksela, a zatim nasumičnim odabirom izrezujemo primjer dimenzija 512×512. U tako dobivenu scenu vožnje lijepimo primjere iz skupa ImageNet te primjere s ruba distribucije

koje bilinearno naduzorkujemo na duplo veću veličinu stranice kako bi zadržali jednak udio anomalije na primjeru za učenje. Pola mini-grupe za učenje sačinjavaju ovakvi primjeri, a ostali dio mini-grupe čine primjeri s ruba distribucije odnosno iz skupa ImageNet naduzorkovani do rezolucije 512×512. Kao enkoderski dio modela koristimo Ladder DenseNet-169.

Tablica 7.9 prikazuje rezultate modela s jednom i dvije glave učenih na opisanim primjerima. Stupac *Val mIoU* predstavlja mIoU na validacijskom skupu Mapillary Vistas. Stupac *AP WD-Pascal* predstavlja uprosječenu preciznost na skupu WD-Pascal. Stupac *AP L&F* predstavlja uprosječenu preciznost na skupu FS Lost&Found. Konačno, Stupac *FPR L&F* predstavlja FPR na 95% TPR na skupu FS Lost&Found. Pri učenju modela, svi hiperparametri su jednaki kao u prethodnim eksperimentima.

Model	**Val mIoU** $\uparrow$	**AP WD-Pascal** $\uparrow$	**AP L&F** $\uparrow$	**FPR L&F** $\downarrow$
Dvije glave	69.2	71.9	1.5	97.7
Jedna glava, T=1	69.1	71.4	3.5	31.7
Jedna glava, T=10	69.1	73.5	8.4	28.5

Tablica 7.9: Performanse modela s jednom i dvije glave. Modeli su učeni na skupu podataka Mapillary Vistas s umjetno dodanim anomalijama.

Modeli postižu zadovoljavajuće rezultate na skupu WD-Pascal, dok na skupu podataka FS Lost&Found imaju relativno loše rezultate. Validacijski skup podataka FS Lost&Found zahtjeva učenje modela na skupu podataka Cityscapes. Iz tog razloga u sljedeći eksperiment dodajemo učenje na skupu podataka Cityscapes. Pri evaluaciji modela s jednom glavom primjenjujemo temperaturno skaliranje ulaza u softmaks sloj, gdje je parametar T postavljen na vrijednosti 1 i 10. Vidimo da temperaturno skaliranje softmaksa značajno utječe na performanse na skupu podataka FS Lost&Found te skupu WD-Pascal.

Slika 7.4 prikazuje rezultate modela s dvije glave u detekciji dijelova slike koji sadrže anomaliju na skupu podataka WD-Pascal. Performanse korištenog modela su vidljive u prvom retku tablice 7.9. U navedenom primjeru je piksel označen kao dio anomalije ako vjerojatnosni izlaz dodatne glave daje vjerojatnost veću od 80%. Korišteni model koristi temperaturno skaliranje softmaksa, gdje je parametar T postavljen

na 10. Vidimo da model detektira dobar dio anomalije s visokom sigurnošću.

Slika 7.4: Rezultati detekcije anomalija na skupu podataka WD-Pascal sa sigurnošću postavljenom na 80%

Tablica 7.10 prikazuje performanse modela s jednom glavom učenog na skupovima podatka Cityscapes i Mapillary Vistas. U skupove podataka dodajemo primjere iz skupa ImageNet te primjere s ruba distribucije. Stupci *AP* te *FPR at TPR 95%* mjere prosječnu preciznost odnosno FPR na TPR 95% na skupu podataka FS Lost&Found. Oznaka T definira vrijednost parametra korištenog pri temperaturnom skaliranju softmaksa u fazi zaključivanja. Vidimo superiornije performanse modela koji koristi mješavinu anomalija iz skupa podataka ImageNet te ruba distribucije u odnosu na model koji koristi samo anomalije iz skupa ImageNet.

Vrsta anomalije	**Vistas val mIoU**	**AP**	**FPR at TPR 95%**
ImageNet, T=1	69.0	6.1	46.0
ImageNet, T=10	69.0	15.0	45.8
ImageNet, T=20	69.0	15.3	45.9
ImageNet + rub dist., T=1	68.6	8.5	35.1
ImageNet + rub dist., T=10	68.6	18.7	34.4
ImageNet + rub dist., T=20	68.6	19.0	34.4

Tablica 7.10: Performanse modela s jednom glavom. Model je naučen na skupovima podataka Mapillary Vistas te Cityscapes. Modeli su učeni s različitim umjetno dodanim anomalijama.

8. Zaključak

U ovom radu smo obradili pristupe detekciji izvandistribucijskih primjera odnosno dijelova slike koji sadrže anomaliju. Nakon pregleda suvremenih metoda i pristupa detekciji izvandistribucijskih primjera, definiramo generativni model temeljen na invertibilnim vjerojatnosnim tokovima pod nazivom Real NVP Dinh et al. (2017). Primjenom navedenog generativnog modela unaprijeđujemo metodu definiranu u Lee et al. (2018). Unaprijeđena metoda značajno poboljšava performanse diskriminativnog modela u detekciji izvandistribucijskih primjera, što pokazujemo na skupovima podataka CIFAR10 i SVHN. Ipak, dobiveni rezultati ne dostižu trenutno najboljeg riješenja definirano u Hendrycks et al. (2019b).

Slično kao u metodi Lee et al. (2018), korišteni Real NVP nakon faze učenja generira primjere s ruba distribucije. Real NVP učimo koristeći poboljšanu metodu na skupu podataka Mapillary Vistas, čime dobivamo primjere s ruba distribucije. Zbog prikladne strukture generativnog modela Real NVP, generirani primjeri mogu varirati u prostornim dimenzijama. Generirane primjere lijepimo u scene vožnje, a zatim na takvom skupu podataka učimo model sposoban za istovremenu semantiču segmentaciju i detekciju anomalija.

Modeli sposobni za istovremenu semantičku segmentaciju i detekciju anomalija u slikama se temelje na arhitekturi Ladder DenseNet Kreso et al. (2019). Konkretno, koristimo model koji detektira dijelove slike koji sadrže anomaliju na temelju maksimalne vrijednosti probabilističkog izlaza (max-softmax) te model s dodatnom glavom za detekciju anomalija. Navedene modele učimo na skupu podataka u koji lijepimo dijelove primjera iz skupa podataka ImageNet, primjere s ruba distribucije te kombinaciju obe vrste primjera. Modele evaluiramo na skupu podataka WD-Pascal Bevandic et al. (2019) te slikama iz skupa Mapillary Vistas, gdje smo instance klasa osoba i

vozač označili kao anomaliju. Dobiveni rezultati pokazuju da umjetno dodavanje anomalija u skup podataka za učenje rezultira povećanjem performansi modela u detekciji anomalija na testnim skupovima podataka. Sličan postupak ponavljamo tako da modele učimo na skupovima podataka Cityscapes i Mapillary Vistas, a zatim ih testiramo na skupu podataka FS Lost&Found. Dobiveni rezultati pokazuju da primjeri s ruba distirbucije doprinose performansama modela u detekciji anomalija na slikama.

U budućem radu predlažemo detekciju anomalija u slikama invertibilnim tokom koji je detektira anomalije na temelju izglednosti iz latentni značajki. U teoriji takav pristup ne bi zahtijevao modifikaciju skupa za učenje u vidu lijepljenja anomalija. Također, predlažemo povećanje skupova za učenje koji bi bolje pokrili situacije iz otvorenog svijeta.

LITERATURA

Petra Bevandic, Ivan Kreso, Marin Orsic, i Sinisa Segvic. Discriminative out-of-distribution detection for semantic segmentation. *CoRR*, abs/1808.07703, 2018. URL http://arxiv.org/abs/1808.07703.

Petra Bevandic, Ivan Kreso, Marin Orsic, i Sinisa Segvic. Simultaneous semantic segmentation and outlier detection in presence of domain shift. U Gernot A. Fink, Simone Frintrop, i Xiaoyi Jiang, urednici, *Pattern Recognition - 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10-13, 2019, Proceedings*, svezak 11824 od *Lecture Notes in Computer Science*, stranice 33–47. Springer, 2019. doi: 10.1007/978-3-030-33676-9_3. URL https://doi.org/10.1007/978-3-030-33676-9_3.

Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Information science and statistics. Springer, 2007. ISBN 9780387310732. URL http://www.worldcat.org/oclc/71008143.

Hermann Blum, Paul-Edouard Sarlin, Juan I. Nieto, Roland Siegwart, i Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *CoRR*, abs/1904.03215, 2019. URL http://arxiv.org/abs/1904.03215.

Samuel Rota Bulò, Lorenzo Porzi, i Peter Kontschieder. In-place activated batchnorm for memory-optimized training of dnns. U *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, stranice 5639–5647. IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00591. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Bulo_In-Place_Activated_BatchNorm_CVPR_2018_paper.html.

Tianqi Chen, Bing Xu, Chiyuan Zhang, i Carlos Guestrin. Training deep nets with sublinear memory cost. *CoRR*, abs/1604.06174, 2016. URL http://arxiv.org/abs/1604.06174.

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, i Bernt Schiele. The cityscapes dataset for semantic urban scene un-

derstanding. U *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, i L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. U *CVPR09*, 2009.

Terrance DeVries i Graham W. Taylor. Learning confidence for out-of-distribution detection in neural networks. *CoRR*, abs/1802.04865, 2018. URL http://arxiv.org/abs/1802.04865.

Laurent Dinh, Jascha Sohl-Dickstein, i Samy Bengio. Density estimation using real NVP. U *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=HkpbnH9lx.

Yilun Du i Igor Mordatch. Implicit generation and modeling with energy based models. U Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, i Roman Garnett, urednici, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, stranice 3603–3613, 2019. URL http://papers.nips.cc/paper/8619-implicit-generation-and-modeling-with-energy-based-models.

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, i Andrew Zisserman. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.*, 88(2):303–338, 2010. doi: 10.1007/s11263-009-0275-4. URL https://doi.org/10.1007/s11263-009-0275-4.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, i Yoshua Bengio. Generative adversarial networks. *CoRR*, abs/1406.2661, 2014. URL http://arxiv.org/abs/1406.2661.

Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, i Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. U *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=Hkxzx0NtDB.

Kristen Grauman i Trevor Darrell. The pyramid match kernel: Discriminative classification with sets of image features. U *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, stranice 1458–1465. IEEE Computer Society, 2005. doi: 10.1109/ICCV.2005.239. URL https://doi.org/10.1109/ICCV.2005.239.

Chuan Guo, Geoff Pleiss, Yu Sun, i Kilian Q. Weinberger. On calibration of modern neural networks. U Doina Precup i Yee Whye Teh, urednici, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, svezak 70 od *Proceedings of Machine Learning Research*, stranice 1321–1330. PMLR, 2017. URL http://proceedings.mlr.press/v70/guo17a.html.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, i Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(9):1904–1916, 2015. doi: 10.1109/TPAMI.2015.2389824. URL https://doi.org/10.1109/TPAMI.2015.2389824.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, i Jian Sun. Deep residual learning for image recognition. U *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, stranice 770–778. IEEE Computer Society, 2016a. doi: 10.1109/CVPR.2016.90. URL https://doi.org/10.1109/CVPR.2016.90.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, i Jian Sun. Identity mappings in deep residual networks. U Bastian Leibe, Jiri Matas, Nicu Sebe, i Max Welling, urednici, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV*, svezak 9908 od *Lecture Notes in Computer Science*, stranice 630–645. Springer, 2016b. doi: 10.1007/978-3-319-46493-0_38. URL https://doi.org/10.1007/978-3-319-46493-0_38.

Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, Jacob Steinhardt, i Dawn Song. A benchmark for anomaly segmentation. *CoRR*, abs/1911.11132, 2019a. URL http://arxiv.org/abs/1911.11132.

Dan Hendrycks, Mantas Mazeika, i Thomas G. Dietterich. Deep anomaly detection with outlier exposure. U *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019b. URL https://openreview.net/forum?id=HyxCxhRcY7.

Gao Huang, Zhuang Liu, Laurens van der Maaten, i Kilian Q. Weinberger. Densely connected convolutional networks. U *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, stranice 2261–2269. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.243. URL https://doi.org/10.1109/CVPR.2017.243.

Sergey Ioffe i Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. U Francis R. Bach i David M. Blei, urednici, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, svezak 37 od *JMLR Workshop and Conference Proceedings*, stranice 448–456. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/ioffe15.html.

Diederik P. Kingma i Jimmy Ba. Adam: A method for stochastic optimization. U Yoshua Bengio i Yann LeCun, urednici, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Diederik P. Kingma i Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. U Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, i Roman Garnett, urednici, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, stranice 10236–10245, 2018. URL http://papers.nips.cc/paper/8224-glow-generative-flow-with-invertible-1x1-convolutions.

Josip Krapac, Ivan Kreso, i Sinisa Segvic. Ladder-style densenets for semantic segmentation of large natural images. U *2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017, Venice, Italy, October 22-29, 2017*, stranice 238–245. IEEE Computer Society, 2017. doi: 10.1109/ICCVW.2017.37. URL https://doi.org/10.1109/ICCVW.2017.37.

Ivan Kreso, Josip Krapac, i Sinisa Segvic. Efficient ladder-style densenets for semantic segmentation of large images. *CoRR*, abs/1905.05661, 2019. URL http://arxiv.org/abs/1905.05661.

Alex Krizhevsky, Vinod Nair, i Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). URL http://www.cs.toronto.edu/~kriz/cifar.html.

Alex Krizhevsky, Ilya Sutskever, i Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. U Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, i Kilian Q. Weinberger, urednici, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, stranice 1106–1114, 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

Balaji Lakshminarayanan, Alexander Pritzel, i Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. U Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, i Roman Garnett, urednici, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, stranice 6402–6413, 2017. URL http://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep-ensembles.

John Lambert, Liu Zhuang, Ozan Sener, James Hays, i Vladlen Koltun. MSeg: A composite dataset for multi-domain semantic segmentation. U *Computer Vision and Pattern Recognition (CVPR)*, 2020.

Kimin Lee, Honglak Lee, Kibok Lee, i Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. U *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=ryiAv2xAZ.

Shiyu Liang, Yixuan Li, i R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. U *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=H1VGkIxRZ.

Thomas Lucas, Konstantin Shmelkov, Karteek Alahari, Cordelia Schmid, i Jakob Verbeek. Adaptive density estimation for generative models. U Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, i Roman Garnett, urednici, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, stranice 11993–12003, 2019. URL http://papers.nips.cc/paper/9370-adaptive-density-estimation-for-generative-models.

Eric T. Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Görür, i Balaji Lakshminarayanan. Do deep generative models know what they don't know? U *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=H1xwNhCcYm.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, i Andrew Ng. Reading digits in natural images with unsupervised feature learning. *NIPS*, 01 2011.

Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulò, i Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. U *International Conference on Computer Vision (ICCV)*, 2017. URL https://www.mapillary.com/dataset/vistas.

Marin Orsic, Ivan Kreso, Petra Bevandic, i Sinisa Segvic. In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images. U *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, stranice 12607–12616. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.01289. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Orsic_In_Defense_of_Pre-Trained_ImageNet_Architectures_for_Real-Time_Semantic_Segmentation_CVPR_2019_paper.html.

Tobias Pohlen, Alexander Hermans, Markus Mathias, i Bastian Leibe. Full-resolution residual networks for semantic segmentation in street scenes. U *2017 IEEE Conference on Computer Vision and Pattern*

Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, stranice 3309–3318. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.353. URL https://doi.org/10.1109/CVPR.2017.353.

Alec Radford, Luke Metz, i Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. U Yoshua Bengio i Yann LeCun, urednici, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016. URL http://arxiv.org/abs/1511.06434.

Jie Ren, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark A. DePristo, Joshua V. Dillon, i Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. U Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, i Roman Garnett, urednici, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, stranice 14680–14691, 2019. URL http://papers.nips.cc/paper/9611-likelihood-ratios-for-out-of-distribution-detection.

Olaf Ronneberger, Philipp Fischer, i Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. U Nassir Navab, Joachim Hornegger, William M. Wells III, i Alejandro F. Frangi, urednici, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, svezak 9351 od *Lecture Notes in Computer Science*, stranice 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://doi.org/10.1007/978-3-319-24574-4_28.

Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, i Liang-Chieh Chen. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018. URL http://arxiv.org/abs/1801.04381.

Joan Serrà, David Álvarez, Vicenç Gómez, Olga Slizovskaia, José F. Núñez, i Jordi Luque. Input complexity and out-of-distribution detection with likelihood-based generative models. U *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=SyxIWpVYvr.

Johannes Stallkamp, Marc Schlipsing, Jan Salmen, i Christian Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. U *IEEE International Joint Conference on Neural Networks*, stranice 1453–1460, 2011.

Ke Sun, Yang Zhao, Borui Jiang, Tianheng Cheng, Bin Xiao, Dong Liu, Yadong Mu, Xinggang Wang, Wenyu Liu, i Jingdong Wang. High-resolution representations for labeling pixels and regions. *CoRR*, abs/1904.04514, 2019. URL http://arxiv.org/abs/1904.04514.

Antonio Torralba, Robert Fergus, i William T. Freeman. 80 million tiny images: A large data set for non-parametric object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(11):1958–1970, 2008. doi: 10.1109/TPAMI.2008.128. URL https://doi.org/10.1109/TPAMI.2008.128.

Laurens van der Maaten i Geoffrey E. Hinton. Visualizing data using t-sne. 2008.

Fisher Yu, Yinda Zhang, Shuran Song, Ari Seff, i Jianxiong Xiao. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop. *CoRR*, abs/1506.03365, 2015. URL http://arxiv.org/abs/1506.03365.

Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, i Trevor Darrell. BDD100K: A diverse driving video database with scalable annotation tooling. *CoRR*, abs/1805.04687, 2018. URL http://arxiv.org/abs/1805.04687.

Oliver Zendel, Katrin Honauer, Markus Murschitz, Daniel Steininger, i Gustavo Fernández Domínguez. Wilddash - creating hazard-aware benchmarks. U Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, i Yair Weiss, urednici, *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VI*, svezak 11210 od *Lecture Notes in Computer Science*, stranice 407–421. Springer, 2018. doi: 10.1007/978-3-030-01231-1_25. URL https://doi.org/10.1007/978-3-030-01231-1_25.

Detekcija izvandistribucijskih dijelova slike primjenom generativnih modela

Sažetak

Semantička segmentacija slika važan je zadatak računalnog vida. Najbolji rezultati u tom području postižu se dubokim diskriminativnim konvolucijskim modelima koji su skloni neopravdanom optimizmu. U ovom radu adresiramo navedeni problem korištenjem primjera s ruba distribucije podataka nastalih uzorkovanjem generativnog modela temeljenog na invertibilnom normalizirajućem toku. Primjeri s ruba distribucije podataka u kombinaciji s primjerima iz negativnog skupa podataka drastično povećavaju performanse diskriminativnog modela u detekciji dijelova slike koji sadrže anomaliju. Navedene tvrdnje su vrednovane iscrpnim eksperimentima.

Ključne riječi: Semantička segmentacija, gusta detekcija izvandistribucijskih dijelova slike, generativni modeli temeljeni na invertibilnom normalizirajućem toku, primjeri s ruba distribucije podataka

Dense out-of-distribution detection by using generative models

Abstract

Semantic segmentation is an important task in the field of computer vision. Current state of the art results are obtained by deep discriminative convolutional models which are known for its unjustified optimism. We address this issue by using samples at the distribution border obtained by sampling the flow-based generative model. Samples at the data distribution border in combination with samples from another negative dataset drastically improve discriminative model's performance in anomaly detection. All claims are evaluated on exhaustive experiments.

Keywords: Semantic segmentation, dense out-of-distribution detection, flow-based generative models, samples at the data distribution border