SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 200

GUSTA DETEKCIJA ANOMALIJA GENERATIVNIM MODELIMA

Anja Delić

Zagreb, lipanj 2023.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 200

GUSTA DETEKCIJA ANOMALIJA GENERATIVNIM MODELIMA

Anja Delić

Zagreb, lipanj 2023.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 10. ožujka 2023.

DIPLOMSKI ZADATAK br. 200

Pristupnica:	**Anja Delić (0036515733)**
Studij:	Računarstvo
Profil:	Računarska znanost
Mentor:	prof. dr. sc. Siniša Šegvić

Zadatak: **Gusta detekcija anomalija generativnim modelima**

Opis zadatka:

Semantička segmentacija važan je zadatak računalnog vida s mnogim zanimljivim primjenama. Međutim, standardno nadzirano učenje vrlo je osjetljivo na pojavu anomalija u ispitnih slikama. Zbog toga segmentacija anomalija predstavlja vrlo vrijedan dodatak postupcima za diskriminativnu gustu predikciju. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje matricama i slikama. Proučiti i ukratko opisati postojeće segmentacijske arhitekture. Isprobati različite generativne pristupe za segmentaciju anomalija. Uhodati postupke učenja modela te validiranje hiperparametara. Vrednovati naučene modele te prikazati i ocijeniti postignutu točnost. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, kao i potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 23. lipnja 2023.

SADRŽAJ

1. Uvod

Duboko učenje omogućilo je impresivan napredak u točnosti semantičke segmentacije slika. Algoritmi za semantičku segmentaciju obično se evaluiraju na skupovima slika zatvorenog skupa uz točno predodređene klase koje odgovaraju klasama u skupu za učenje. Međutim model primijenjen u stvarnom svijetu susrest će se i s objektima koje nije vidio u skupu za učenje. Pokazuje se da duboki modeli reagiraju nepredvidljivo na ulaze koji odstupaju od distribucije za učenje. Takvom će primjeru dodijeliti neku oznaku razreda iz skupa za učenje, a problem nastaje kada model u takvu predikciju ima visoku sigurnost. Želimo da model bude svjestan primjera koji ne dolaze iz generativne distribucije za učenje. Jedan način kako to postići je zahtijevati da se takvi primjeri klasificiraju u poseban razred anomalija. Sposobnost procjene nesigurnosti i otkrivanja anomalija ključna je za sigurnosno kritične primjene poput autonomne vožnje gdje se uslijed netočnih predikcija mogu dogoditi katastrofalne posljedice.

Segmentacija anomalija vrlo je vrijedan dodatak diskriminativnim modelima za gustu predikciju. Posebno zanimljiv smjer istraživanja je detekcija anomalija primjenom generativnih modela. Predlažemo koristiti uvjetne normalizirajuće tokove za učenje pozadinske distribucije značajki dubokog modela po razredima. Kao glavni doprinos ovoga rada ističemo metode za detekciju anomalija u gustim predikcijama mjerama zasnovanima na dodanoj informaciji.

U ovome radu prvo u poglavlju 2 definiramo problem detekcije anomalija i klasifikacije na otvorenom skupu, dajemo detaljan pregled literature i opisujemo postupak evaluacije detektora anomalija. U poglavlju 3 opisujemo generativni model normalizirajućeg toka. Dodatno, u poglavlju 4 opisujemo modele za gustu predikciju na razini piksela i na razini maski. Dajemo i kratak pregled osnovnih pojmova iz teorije informacije korištenih za procjenu nesigurnosti u poglavlju 5. Poglavlje 6 sadrži detaljan opis korištene metode za detekciju anomalija generativnim modelima i prijedlog mjera anomalnosti. U poglavljima 7 i 8 detaljno smo opisali provedene eksperimente i rezultate na javno dostupnim skupovima za evaluaciju metoda za detekciju anomalija. U poglavlju 9 pokazujemo da se uvjetni tokovi mogu koristiti kao generativni klasifikator.

2. Gusta detekcija anomalija

Zadatak detekcije primjera koji ne dolaze iz distribucije za učenje naziva se detekcija izvandistribucijskih primjera, detekcija anomalija ili novotvorina. U ovom radu koristimo naziv detekcija anomalija. Na primjer koji ne dolazi iz distribucije skupa za učenje referenciramo se s anomalan ili izvandistribucijski primjer. Zadatak detekcije anomalija našao je važnu primjenu u računalnom vidu. Detekciju anomalija možemo primijeniti na zadatak klasifikacije slike ili segmentacije slike. U prvom slučaju cijelu sliku proglašavamo anomalnom, a u drugom slučaju pojedine piksele proglašavamo anomalnima. U ovom radu proučavamo drugi slučaj, koji se u literaturi zove gusta detekcija anomalija jer za svaki piksel slike želimo odluku je li on anomalan ili nije.

Detekcija anomalija može se svesti na binarni klasifikacijski zadatak koji se u pozadini oslanja na procjenu nesigurnosti predikcije. U gustoj detekciji anomalija, svakom je pikselu potrebno dodijeliti mjeru nesigurnosti ili anomalnosti:

$$\mathrm{s}(\mathbf{x}) : [0, 255]^{3 \times H \times W} \rightarrow \mathbb{R}^{H \times W}. \tag{2.1}$$

Jednom kada imamo procjenu nesigurnosti predikcije, na validacijskom skupu pronalazino prag tako da zadovoljimo predefinirani udio lažnih pozitiva. Primjer je onda anomalan ako mu je mjera anomalnosti manja od definiranog praga.

Učenje na otvorenom skupu je zadatak izravno povezan s detekcijom anomalija. U standardnom učenju na zatvorenom skupu primjere klasificiramo isključivo u jedan od predefiniranih K razreda. Učenje na otvorenom podrazumijeva testne primjere koji izlaze iz taksonomije skupa za učenje. Jedan način kako pristupiti tom zadatku je kombinirati zatvoreni klasifikator i detektor anomalija tako da detektor anomalija detektira primjere koji ne pripadaju razredima zatvorenog skupa. Modele za semantičku segmentaciju na otvorenom skupu moguće je izravno validirati samo na testnim skupovima koji imaju K+1 oznaka. Zbog velike cijene gustog označivanja slika, takvi skupovi su rijetki.

2.1. Pregled literature

U posljednje vrijeme vidljiv je znatan napredak u metodama za detekciju anomalija.

Rane metode za detekciju anomalija temelje se na probabilističkom modeliranju izlaza modela. Primjer je vjerojatnost maksimalnog softmaksa [19] i entropija softmaksa. Primjer jednostavne metode je i maksimalni logit [21]. Predložene su i tehnike koje poboljšavaju performanse metoda temeljene na maksimalnog logitu [22]. Uočeno je da se distribucije logita po razredima razlikuju, a predloženo rješenje je normalizacija logita. Te metode ne zahtijevaju dodatno učenje inicijalnog klasifikatora ili dodatne negativne podatke.

Diskriminativno učenje uz stvarne negative postiže zanimljive rezultate [2, 3, 20]. Diskriminativni model za semantičku segmentaciju uči se na slikama na koje se lijepe isječci iz negativnog skupa slika. Najčešće se kao negativni skup slika uzimaju skupovi široke domene kao što su ImageNet, COCO i ADE20k. Dodatna binarna glava koja dijeli značajke sa standardnom segmentacijskom glavom procjenjuje vjerojatnost da je piksel anomalan: $P(d_{out}|\mathbf{x})$. Učenje na stvarnim negativima postiže dobre rezultate, ali i uvodi pristranost jer je detektor anomalija u procesu učenja vjerojatno vidio neke anomalije iz testnog skupa. Zbog toga se umjesto stvarnih negativnih skupova podataka koriste negativni primjeri sintetizirani generativnim modelima [29, 13, 14]. U tu svrhu koriste se generativni suparnički modeli [29] ili normalizirajući tokovi [13, 14].

Posebno je zanimljiva primjena generativnih modela u metodama zasnovanim na procjeni izglednosti. Izglednost se procjenjuje za to primjenjivim generativnim modelima kao što su normalizirajući tok [4], Gaussove mješavine [30] ili energijski modeli [16]. Istraživanja pokazuju da metode detekcije anomalija procjenom izglednosti slika nisu optimalne zbog toga što generativni modeli, specifično normalizirajući tokovi, daju veću izglednost anomalnim nego unutardistribucijskim primjerima. Poznati primjer na kojem je uočeno takvo ponašanje je distinkcija skupova slika CIFAR-10 i SVHN [34]. U članku autori analiziraju zašto modeli normalizirajućeg toka griješe u detekciji izvandistribucijskih primjera. Govore da je to posljedica induktivne pristranosti toka koja proizlazi iz načina učenja i arhitekture. Pokazuju da normalizirajući tokovi uče latentne reprezentacije slike vodeći se lokalnim korelacijama piksela, a ne semantičkim kontekstom. Zbog toga se ne mogu detektirati semantičke anomalije. Opisani problem može se riješiti dijeljenjem značajki s diskriminativnim modelom [43, 4]. U [36] autori proučavaju detekciju anomalija na genskim sekvencama i primjećuju slično ponašanje detektora anomalija zasnovanih na izglednosti kao i prethodno opisani radovi. Pokazuju da na izglednost značajki dubokih modela može utjecati po-

zadinska statistika. Zato predlažu koristiti omjer izglednosti gdje se pozadinski model uči uz perturbacije ulaza.

Hibridni pristupi detekciji anomalija daju dobre rezultate i obećavajući smjer za buduće istraživanje. U članku [16, 13] predstavljena je metoda koja se temelji na združenom hibridnom učenju diskriminativnog segmentacijskog modela i generativnog energijskog modela i hibridnoj mjeri anomalnosti.

2.2. Evaluacija performansi detektora anomalija

Za mjerenje performansi detektora anomalija koristimo prosječnu preciznost (engl. *average precision, AP*), stopu lažnih pozitiva kada je stopa stvarnih pozitiva 95% (FPR_{95}) i AUROC (engl. area under receiver operating curve). Te mjere su najčešće korištene u literaturi. U nastavku dajemo kratak opis i interpretaciju korištenih mjera [38].

Prosječna preciznost je integral ispod krivulje preciznost-odziv. Krivulja preciznost-odziv koristi se za vrednovanje klasifikatora pri različitim klasifikacijskim pragovima. Preciznost definiramo kao udio stvarno pozitivnih primjera u skupu svih primjera koje je klasifikator označio kao pozitivne:

$$P = \frac{TP}{TP + FP}, \tag{2.2}$$

a odziv (engl. *recall*) je udio stvarno pozitivnih primjera u skupu svih skupu svih pozitivnih primjera:

$$R = TPR = \frac{TP}{TP + FN}. \tag{2.3}$$

Krivulja ROC (engl. *receiver operating characteristics*) je vrijednost stope stvarnih pozitiva kao funkcije stope lažnih pozitiva. Stopa lažnih pozitiva (FPR) ili lažni alarm:

$$FPR = \frac{FP}{FP + TN} \tag{2.4}$$

je udio primjera koje je klasifikator proglasio pozitivnima, ali oni to nisu. AUROC je površina ispod ROC krivulje. Savršeni klasifikator imao bi AUROC=1, a slučajni klasifikator 0.5. Broj lažnih pozitiva u području visokog odziva je bitan za sigurnost u kritičnim primjenama. Zbog toga analiziramo stopu lažnih pozitiva kada je stopa stvarnih pozitiva 95% (FPR_{95}).

Skup podataka u detekciji anomalija najčešće je nebalansiran, tj. puno je veći broj unutardistribucijskih podataka nego izvandistribucijskih. AP naglašava detekciju manjinskog razreda, pa se obično smatra primarnom mjerom. Detekcija anomalija svodi

se na problem binarne klasifikacije. Klasifikacijski prag odabiremo na stopi stvarnih pozitiva od 95%.

Bitno je i osigurati da ne dođe do pada klasifikacijskih performansi kada se klasifikatoru doda detektor anomalija. Zbog toga se mjeri omjer presjeka i unije (IoU) kao standardna metrika za semantičku segmentaciju na validacijskom skupu slika.

Detektori anomalija moraju omogućiti brzo iskorištavanje, a vrijeme izvođenja obuhvaća i vrijeme potrebno za semantičku segmentaciju i detekciju anomalija u slici. Brzinu izvođenja mjerimo u okvirima po sekundi (engl. *frame per second, fps*).

2.3. Skupovi slika za evaluaciju detektora anomalija

Kreirati skup slika za evaluaciju metoda za gustu detekciju anomalija nije lak zadatak. Potrebne su guste oznake, a gusto označavanje je skupo. Dodatno, postavlja se pitanje kako simulirati anomalije. Mjerenje performansi metoda za detekciju anomalija u slikama vožnje na cestama doživjelo je značajan napredak u posljednjim godinama. Prvi pristupi lijepili su anomalne objekte nasumično na standardne validacijske slike [2], a naprednijim metodama lijepljenja anomalija mogu se dobiti realističnije testne slike [4]. Koriste se i slike sa stvarnim anomalijama [6] te iskorištavaju simulirana okruženja [21]. Za pravednu evaluaciju metoda za detekciju anomalija, potrebno je osigurati da modeli nisu naučeni na anomalnim primjerima. To se osigurava tako da se javno objavi samo manji validacijski skup slika, a testni skup ostaje tajan.

Fishyscapes [4] je prvi javni skup slika za evaluaciju metoda za gustu detekciju anomalija u slikama gradske vožnje koje sadrže potencijalno u vožnji opasne anomalije. Sastoji se od dva skupa slika: FS Static i FS LostAndFound. FS Static je skup slika koji se bazira na validacijskom skupu iz skupa slika Cityscapes[10] na koje su naprednim tehnikama lijepljene anomalije iz Pascal VOC. Validacijski skup se sastoji od 30 slika. FS LostAndFound je podskup koji sadrži male anomalne objekte na cesti. Validacijski skup se sastoji od 100 slika. RoadAnomaly [31] je skup slika nastao lijepljenjem anomalnih objekata na slike vožnje. Validacijski skup se sastoji od 60 slika. SegmentMeIfYouCan (SMIYC) [6] se sastoji od tri skupa slika: AnomalyTrack, ObstacleTrack and LostAndFound. StreetHazards [21] je sintetički skup slika iz CARLA simulacije što olakšava manipulaciju anomalnim objektima. Ovaj skup slika omogućuje i evaluaciju semantičke segmentacije u otvorenom skupu.

Modeli koji se evaluiraju na prethodnio opisanim testnim skupovima najčešće se uče na skupu Cityscapes. Cityscapes [10] je standardni skup slika za semantičku segmentaciju na zatvorenom skupu koje prikazuju ubranu vožnju. Ima 19 označenih raz-

reda i sastoji se 2975 slika za učenje i 500 slika u validacijskom skupu. Međutim, u nekim testnim skupovima je velik pomak domene u odnosu na Cityscapes. Zbog toga se modeli uče na dodatnim skupovima slika uz Cityscapes kao što je Mapillary Vistas [35].

Metode predstavljene u ovom radu evaluirali smo na validacijskim skupovima slika Fishyscapes i RoadAnomaly.

3. Normalizirajući tok

Normalizirajući tok je generativni model. Generativni model procjenjuje nepoznatu distribuciju p_D iz koje dolaze podatci za učenje D poznatim modelom p_θ koji je definiran arhitekturom modela i parametrima θ. Konkretno, normalizirajući tok je transformacija jednostavne distribucije u kompleksniju distribuciju slijedom invertibilnih i diferencijabilnih preslikavanja [27]. Bitna karakteristika normalizirajućih tokova je mogućnost egzaktne reprezentacije log-izglednosti. Učenje takvog modela, tj. pronalazak optimalnih parametara θ^* možemo provesti optimizacijom izglednosti podataka iz skupa za učenje prema izrazu 3.1.

$$\theta^* = \text{argmin}_{\theta \in \Theta} \mathbb{E}_{x \sim p_D}[-\ln p_\theta(\mathbf{x})] \tag{3.1}$$

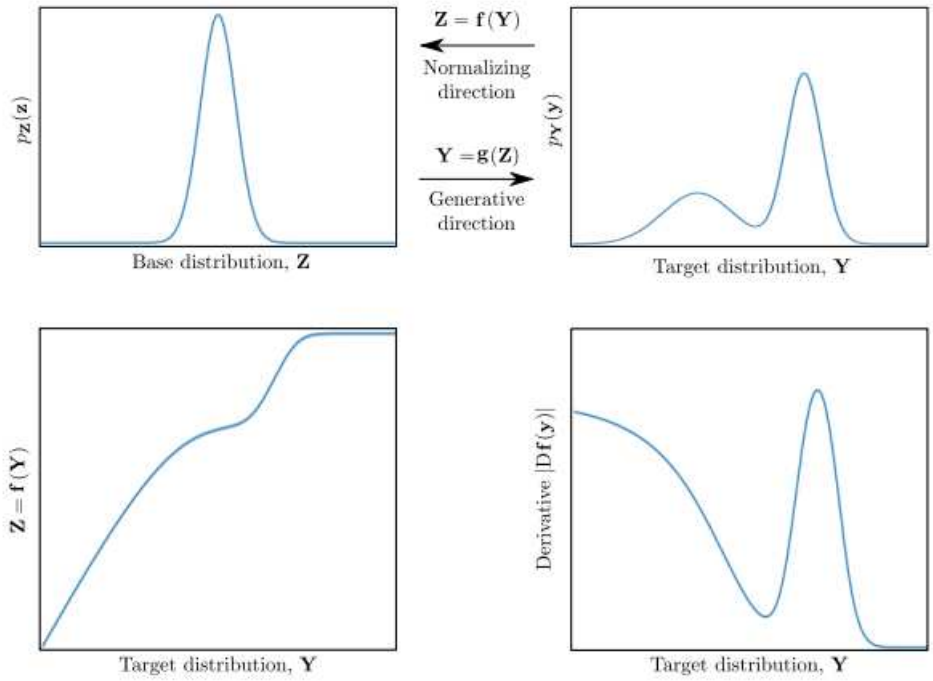
Funkcija izglednosti parametrima θ pridjeljuje vjerojatnost da je iz populacije s parametrima θ izvučen skup podataka $\mathbf{D}$.

Normalizirajući tok temelji se na formuli za zamjenu varijabli. Pretpostavimo da postoji diferencijabilna bijekcija f parametrizirana po θ koja transformira kompleksnu distribuciju Y u jednostavnu Gaussovu distribuciju Z. Neka je g inverzna funkcija od f koja jednostavnu distribuciju transformira u kompleksniju i neka je $Z = f(Y)$ i $Y = g(X)$. Takvo mapiranje omogućeno je formulom za zamjenu varijabli (3.2) i ilustrirano je na slici 3.1.

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{Z}}(f(\mathbf{y}))\left|\det \frac{\partial f}{\partial y}\right| = p_{\mathbf{Z}}(f(\mathbf{y}))\left|\det \frac{\partial g}{\partial z}\right|^{-1} \tag{3.2}$$

Ako želimo uzorkovati (generirati) primjere iz distribucije Y koja nam nije poznata, možemo uzorkovati iz jednostavne poznate distribucije Z i primijeniti transformaciju g. Kažemo da g transformira distribuciju u generativnom smjeru. Funkcija f radi suprotno. Ona transformira podatke iz kompleksnije u jednostavniju distribuciju koja je u praksi vrlo često upravo Gaussova ili normalna distribucija. Otuda dolazi i naziv za normalizirajući tok. Dakle, učenje normalizirajućeg toka svodi se na učenje invertibilne i diferencijabilne funkcije f, tj. na pronalaženje optimalnih parametara θ te funkcije. U jednadžbi (3.2) $\frac{\partial f}{\partial y}$ je Jakobijan funkcije f (analogno za funkciju g), a

izraz $\left|\det \frac{\partial f}{\partial y}\right|$ predstavlja korekciju volumena. Da bi normalizirajući tok bio efikasan, izračun determinante Jakobijana funkcije f mora biti efikasan.



Slika 3.1: Ilustracija formule zamjene varijabli. Slika je preuzeta iz [27]

Odabir dobre funkcije toka f je težak problem. Ako primijetimo da je kompozicija invertibilnih i diferencijabilnih funkcija također invertibilna i diferencijabilna, onda funkciju f možemo rastaviti kao $f = f_1 \circ f_2 \circ ... \circ f_N$. Log-determinanta Jakobijana takvog toka je zboj log-determinanti Jakobijana svake komponente.

Jedan od zadataka koje rješava normalizirajući tok i koji je u fokusu ovog rada je procjena gustoće vjerojatnosti (engl. *density estimation*). Gustoću vjerojatnosti pod modelom možemo dobiti računanjem izglednosti nakon unaprijednog prolaza, tj. prolaza u normalizirajućem smjeru. Negativnu log-izglednost izražavamo po dimenziji. Ako je baza logaritma 2, mjernu jedinicu zovemo bitovi po dimenziji. Tu veličinu možemo interpretirati kao prosječan broj bitova, ili općenito simbola, potrebnih da bi se bez gubitka kodirala diskretna distribucija [39].

Razvijene su mnoge zanimljive arhitekture normalizirajućih tokova [12, 23, 15]. Različite primjene zahtijevaju različite arhitekture, a u ovom radu normalizirajući tok koristimo za učenje na vektorima značajki.

Glavna gradivna jedinica arhitekture RealNVP [12] je sloj miješanja (engl. *coupling layer*). Ideja je podijeliti ulaz u dva disjunktna skupa, $\mathbf{x}^A$ i $\mathbf{x}^B$. Neka je h di-

ferencijabilna bijekcija parametrizirana po parametrima θ koju ćemo primijeniti samo na parametre $\mathbf{x}^A$. Pritom parametri θ ovise samo o parametrima $\mathbf{x}^B$, tj. parametri θ su definirani proizvoljnom funkcijom $\Theta(\mathbf{x^B})$ (engl. *conditioner*):

$$\begin{aligned} \mathbf{y}^A &= \mathbf{h}(\mathbf{x}^A; \theta(\mathbf{x}^B)) \\ \mathbf{y}^B &= \mathbf{x}^B. \end{aligned} \tag{3.3}$$

Sloj je invertibilan ako je funkcija miješanja h invertibilna. Inverz je:

$$\begin{aligned} \mathbf{x}^A &= \mathbf{h}^{-1}(\mathbf{y}^A; \theta(\mathbf{x}^B)) \\ \mathbf{x}^B &= \mathbf{y}^B. \end{aligned} \tag{3.4}$$

Jakobijan toka s funkcijom miješanja je particionirana trokutasta matrica gdje su dijagonalni blokovi Jakobijan funkcije miješanja i matrica identiteta:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} I & 0 \\ \frac{\partial y_{d+1:D}}{\partial x_{1:d}} & \operatorname{diag}\left(\exp\left[s\left(x_{1:d}\right)\right]\right) \end{bmatrix}. \tag{3.5}$$

Slijedi da je determinanta Jakobijana toka determinanta Jakobijana funkcije miješanja. Funkcija $h : \mathbb{R}^d \to \mathbb{R}^d$ najčešće je konstruirana tako da se na $\mathbf{x}^A$ primjenjuju elementarne transformacije (engl. *elementwise*)

$$\mathbf{h}(\mathbf{x}^A; \theta) = (h_1(x_1^A; \theta_1), ..., h_d(x_d^A; \theta_d)) \tag{3.6}$$

gdje je $(h_i(x_i^A; \theta_i) : \mathbb{R} \to \mathbb{R}$ strogo monotona skalarna bijekcija. Primjeri takvih funckija miješanja aditivna funkcija miješanja (3.7) i afina funkcija miješanja 3.8. Te funkcije su jednostavne i računski efikasne, ali imaju ograničenu ekspresivnost i više tokova treba biti komponirano da bi se mogle učiti kompleksnije distribucije.

$$h(x; \theta) = x + \theta, \theta \in \mathbb{R} \tag{3.7}$$

$$h(x; \theta) = \theta_1 x + \theta_2, \theta_1 \neq 0, \theta_2 \in \mathbb{R} \tag{3.8}$$

Funkcija $\Theta(\mathbf{x^B})$ može biti proizvoljno kompleksna, a u praksi je najčešće duboki model kojeg zovemo modul miješanja.

Nakon svakog sloja miješanja potrebno je napraviti permutaciju ulaza kako bi se u svakom sloju miješanja miješao drugi dio. Ta permutacija može biti fiksna ili naučena. Primjer fiksiranog permutacijskog uzorka je alternirajuća izmjena polovica tenzora po kanalima. Primjer naučene invertibilne permutacije je konvolucija 1×1 što je primijenjeno u arhitekturi Glow [23].

Dodatan doprinos arhitekture Glow je normalizacija aktivacije (engl. *activation normalizration*), skraćeno Actnorm. Actnorm je naučena afina transformacija po kanalima. Parametre afine transformacije inicijaliziramo ovisno o podatcima. Želimo da zlazne aktivaije actnorm sloja po kanalima imaju srednju vrijednost nula i varijancu jedan nakon prve minigrupe.

3.1. Učenje normalizirajućih tokova

Neka je funkcija toka f parametrizirana vektorom θ i neka je poznata bazna distribucija s gustoćom vjerojatnosti p_Z parametrizirana vektorom ϕ. Parametre $\Theta = (\theta, \phi)$ možemo učiti maksimizacijom izglednosti na skupu za učenje $D = \{\mathbf{y}^{(i)}\}_{i=1}^{N}$. Negativnu log-izglednost možemo dobiti:

$$\begin{aligned} NLL(\Theta) &= \log p(D|\Theta) && (3.9)\\ &= \sum_{i=1}^{N} \log p_{\mathbf{Y}}(\mathbf{y}^{(i)}|\Theta) \\ &= \sum_{i=1}^{N} \log p_{\mathbf{Z}}(\mathbf{f}(\mathbf{y}^{(i)}|\theta)|\phi) + \log \left| \det \frac{\partial}{\partial \mathbf{y}} \mathbf{f}(y^{(i)}|\theta) \right|. && (3.10) \end{aligned}$$

Dakle, standardno normalizirajući tok učimo minimizacijom log-izglednosti podataka pod modelom.

$$L_{\mathbf{standard}}(\mathbf{X}; \Theta) = \min_{\Theta} NLL(\mathbf{X}; \mathbf{\Theta}), \tag{3.11}$$

gdje je $\mathbf{X}$ skup za učenje.

Alternativno, normalizirajući tok možemo učiti i kontrastivno:

$$L_{\mathbf{contrastive}}(\Theta) = \min_{\Theta}\{0, NLL(\mathbf{X}_{in}; \Theta)\ + \max_{\Theta}\{0, m - NLL(\mathbf{X}_{out}; \Theta)\}.. \tag{3.12}$$

Za kontrastivno učenje potrebni su nam pozitivni i negativni primjeri za učenje. Ako učimo tokove uvjetovane razredom c, pozitivne primjere definirano kao primjere koji pripadaju razredu c, a negativni primjeri su primjeri koji pripadaju svim ostalim razredima. Za c-ti tok minimiziramo izglednost primjera x_{in} koji pripadaju c-toj klasi, a maksimiziramo izglednost primjera x_{out} koji pripadaju svim ostalim klasama. U izrazu 3.12 m predstavlja marginu koja je hiperparametar učenja, a $\mathbf{X}_{in}$ i $\mathbf{X}_{out}$ su podskupovi skupa za učenje s unutardistribucijskim primjerima odnosno izvandistribucijsim primjerima. Na ovaj način učimo razliku između gustoće vjerojatnosti primjera jednog razreda i kontrastivne gustoće. Skupove $\mathbf{X}_{in}$ i $\mathbf{X}_{out}$ možemo definirati na različite načine, ovisno o primjerni. $\mathbf{X}_{in}$ može biti skup primjera jednog razreda, a $\mathbf{X}_{out}$ skup primjera svih ostalih razreda.

4. Semantička segmentacija

Segmentacija slike je grupiranje piksela s obzirom na njihovu semantiku. Razlikujemo semantičku segemntaciju, segmentaciju instanci ili panoptičku segmentaciju. U semantičkoj segementaciji piksele grupiramo s obzirom na pripadnost razredu, u segmentaciji instanci grupiramo piksele koji pripadaju istoj instanci objekta dok u panoptičkoj segmentaciji objedinjujemo ta dva zadatka. Za svaki zadatak razvijene su specijalizirane arhitekture koje slabo generaliziraju za primjenu na druge zadatke. Za semantičku segmentaciju koriste se potpuno konvolucijski modeli [33] [34], dok se primjerice za segmentaciju instanci koriste arhitekture za predikciju na razini maski [18]. U posljednje vrijeme primijećen je razvoj univerzalnih arhitektura koje se mogu primijeniti na sve segmentacijske zadatke [8] [9].

Iako sam zadatak semantičke segmentacije nije fokus ovoga rada, bitno je znati detalje izvedbe modela za semantičku segmentaciju kako bismo mogli osmisliti kvalitetnu metodu za gustu detekciju anomalija. Jedan od ciljeva nam je usporediti metode za detekciju anomalija na različitim semantičkim razinama,tj. na razini piksela i na razini maski. Zbog toga odabiremo i proučavamo dvije različite arhitekture modela za semantičku segmentaciju koji su opisani u nastavku.

4.1. DeepLabv3+

DeepLabv3+ [7] je duboki konvolucijski model za semantičku segemntaciju slike. Kombinira prostorno piramidalno sažimanje na različitim skalama za ekstrakciju kontekstualno bogatih značajki i enkoder-dekoder strukturu za očuvanje detalja i rubova objekata. Za ekstrakciju kontekstualno bogatih značajki na različitim skalama koriste se paralelne dilatirane (engl. *atrous, dilated*) separabilne konvolucije s različitom stopom. Takav modul naziva se *Altrous Spatial Pyramid Pooling*, skraćeno ASPP. Dilatirana konvolucija je poopćenje standardne konvolucije i omogućava eksplicitno upravljanje rezolucijom izlaznih značajki. Dubinski separabilna konvolucija ili grupna konvolucija smanjuje računski trošak i broj parametara zadržavajući slične performanse

na način da se standardna konvolucija faktorizira u dubinsku konovluciju praćenu 1x1 konvolucijom. Dubinska konvolucija znači primijeniti prostornu konvoluciju po kanalima. Enkoder postepenom redukcijom mapi značajki izlučuje semantičke značajke, a dekoder oporavlja prostornu informaciju.

U našim eksperimentima koristimo verziju DeepLabv3+ s okosnicom WideResNet38 zbog dobrih performansi na zatvorenom skupu i pravedne usporedbe s prethodnim radovima.

4.2. Mask2Former

Masked-attention Mask Transformer, skraćeno Mask2Former, je meta arhitektura koja uspješno rješava zadatak semantičke segmentacije, segemntacije instanci i panoptičke segemntacije. Zadatak semantičke segmentacije rješava na razini binarnih maski po uzoru na postupak arhitekture DETR [5]. Semantička segemntacija na razini maski grupira piksele u N segmenata predviđanjem N binarnih maski od kojih svaka daje predikciju o pripadnosti pojedinog piksela N-tom segmentu. Mask2Former se sastoji od tri dijela: okosnice za izlučivanje značajki, piksel dekodera i transformer dekodera. Arhitektura je prikazana na slici 4.2. Okosnica služi za ekstrakciju značajki, a piksel dekoder služi za postepeno naduzorkovanje značajki iz okosnice, kao i u standardnim arhitekturama za semantičku segmentaciju. Transformer dekoder na značajkama procesira upite. U transformer arhitekturama standardno se koristi unakrsna pažnja (engl. cross-attention). Doprinos Mask2Formera je u tome što u transformer dekoderu koristi maskiranu pažnju (engl. *masked attention*) pod pretpostavkom da su lokalne značajke dovoljne jer se kontekst može dobiti korištenjem samopažnje. Maskirana pažnja (4.2) je varijanta unakrsne pažnje (4.1) ograničena na regiju definiranu maskom. Pratimo standardnu notaciju iz [40]. Unakrsna pažnja definirana je:

$$\mathbf{X}_l = \mathrm{softmax}\left(\mathbf{Q}_l \mathbf{K}_l^{\mathrm{T}}\right) \mathbf{V}_l + \mathbf{X}_{l-1}, \tag{4.1}$$

a maskirana pažnja:

$$\mathbf{X}_l = \mathrm{softmax}\left(\mathcal{M}_{l-1} + \mathbf{Q}_l \mathbf{K}_l^{\mathrm{T}}\right) \mathbf{V}_l + \mathbf{X}_{l-1}. \tag{4.2}$$

Maska pažnje na jednoj prostornoj lokaciji definirana je:

$$\mathcal{M}_{l-1}(x, y) = \begin{cases} 0 & \text{if } \mathbf{M}_{l-1}(x, y) = 1 \\ -\infty & \text{otherwise} \end{cases}, \tag{4.3}$$

gdje je $\mathbf{M}_{l-1}$ binarizirani izlaz uz prag 0.5 prethodnog sloja transformer dekodera. Kako bismo sačuvali informaciju o malim objektima, transformer dekoder koristi značajke na različitim skalama. Mape značajki iz piramide značajki iz piksel dekodera prosljeđujemo uzastopnim slojevima u transformer dekoderu.

Segmentacija na razini maski razdvaja zadatke lokalizacije i klasifikacije. Lokalizacija se rješava predikcijama maski $m^{N\times H\times W}$ gdje je N broj maski, a H i W prostorne dimenzije slike. Maske se računaju kao 1×1 konvolucija između gustih značajki $\mathbf{E}$ i izlaznih projekcija $\mathbf{w}_{loc}$ uz sigmoidalnu aktivaciju:

$$\mathbf{m} = \sigma(conv_{1\times1}(\mathbf{E}, \mathbf{w}_{loc})). \tag{4.4}$$

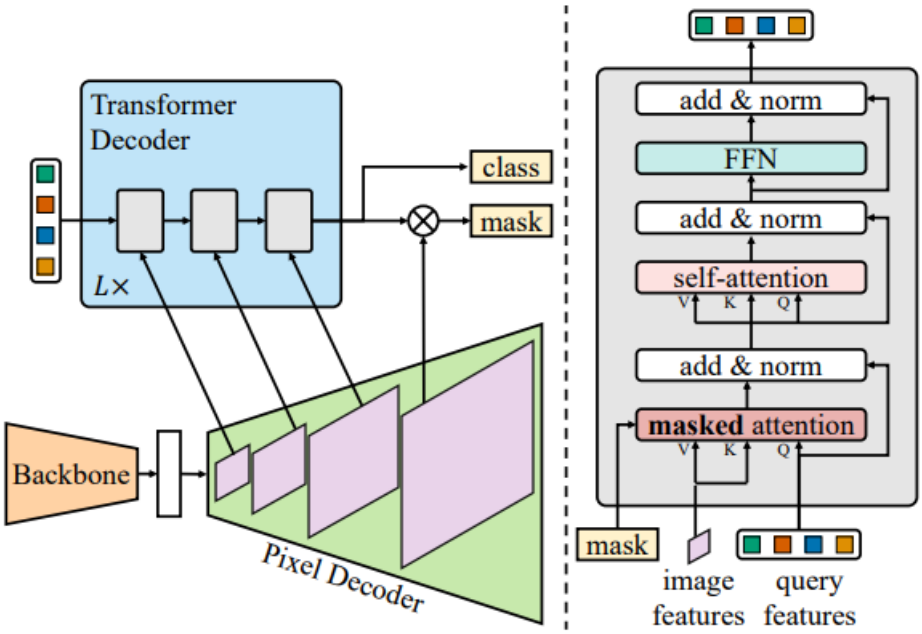
Za svaku takvu masku transformer dekoder će dati predikciju o njezinoj pripadnosti razredu, tj. izlaz će biti kategorička distribucija po maskama pripadnosti u K+1 razreda. Razred K+1 sadrži sve maske koje ne pripadaju niti jednoj od definiranih K klasa. To su obično maske koje u svim pikselima imaju slab odziv ili maske čiji se razredi ne pojavljuju u trenutnoj slici. U klasifikaciji maski odbacujemo predikciju o razredu K+1. Kategoričke distribucije preko K klasa za N maski sadržane su u tenzoru w_{cls}. Segmentaciju na zatvorenom skupu možemo provesti:

$$\mathbf{H}_{closed} = conv_{1\times1}(\mathbf{m}, \mathbf{w}_{cls}), \tag{4.5}$$

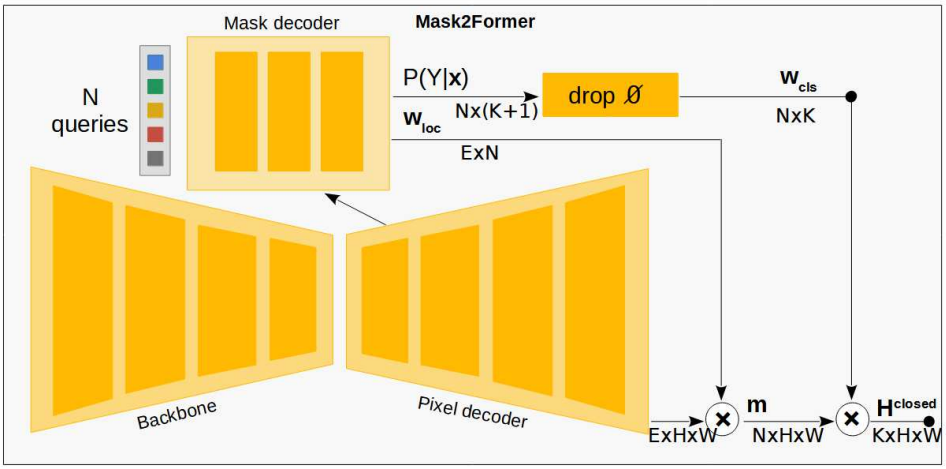
a predikcije razreda možemo dobiti kao:

$$\hat{y}[r, c] = \underset{k=1\ldots K}{\operatorname{argmax}} \sum_i \mathbf{m}_i[r, c] \cdot P_i(Y = k \mid \mathbf{x}). \tag{4.6}$$

Treba napomenuti da $\mathbf{H}_{closed}$ ne sadrži distribucije jer $\sum_i m_i[r, c] \neq 1$ i $\sum_k \mathbf{w}_{cls}[i, k] \neq 1$. Svaki piksel se klasificira s obzirom na otežani ansambl klasifikacija na razini maski gdje su težine predikije maski.



Slika 4.1: Arhitektura Mask2Formera i prikaz bloka transformer dekodera koji se temelji na maskiranoj pažnji. Slika je preuzeta iz [9].



Slika 4.2: Postupak segmentacije na zatvorenom skupu modela Mask2Former s naznačenim korištenim tenzorima i dimenzijama. Slika je preuzeta iz [17] i modificirana.

5. Teorija informacije

Teorija informacije [37] nudi moćan okvir za analizu i formalizaciju ponašanja dubokih modela [26]. Za procjenu nesigurnosti predikcija dubokog modela zanimljivo bi bilo mjeriti informacijsku dobit dobivenu iz specifične opservacije jer su neke opservacije informativnije od drugih. Međutim, Shannonova teorija informacije ne daje rješenje kako izmjeriti informacijsku dobit za specifičan simbol. Članak [11] predlaže dvije formulacije za informacijsku dobit za specifičan simbol uz napomenu da postoji beskonačno mnogo alternativa. U nastavku ćemo dati kratki opis osnovnih pojmova iz teorije informacije te definicije specifičnog iznenađenja i specifične informacije.

Jedan od osnovnih pojmova u Shannonovoj teoriji informacije [37] je entropija:

$$H(X) = -\sum_i p(x_i) \log p(x_i). \tag{5.1}$$

Entropija se često koristi standardna mjera nesigurnosti [29]. Dodatno, entropija ima svojstvo aditivnosti važno za primjenu u procjeni nesigurnosti za nezavisne događaje X i Y:

$$\begin{aligned} H(X,Y) &= -\sum_{i,j} p\left(x_i, y_j\right) \log p\left(x_i, y_j\right) \\ &= -\sum_{i,j} p\left(y_j\right) p\left(x_i \mid y_j\right) \log \left[p\left(y_j\right) p\left(x_i \mid y_j\right)\right] \\ &= -\sum_j p\left(y_j\right) \log p\left(y_j\right) \sum_i p\left(x_i \mid y_j\right) - \sum_j p\left(y_j\right) \sum_i p\left(x_i \mid y_j\right) \log p\left(x_i \mid y_j\right) \\ &= H(Y) + H(X \mid Y) \end{aligned}$$

gdje je $H(X|Y)$ je uvjetna entropija:

$$\begin{aligned} H(X \mid Y) &= -\sum_j p\left(y_j\right) \sum_i p\left(x_i \mid y_j\right) \log p\left(x_i \mid y_j\right) \\ &= \sum_j p\left(y_j\right) H\left(X \mid y_j\right), \end{aligned}$$

a

$$H\left(X \mid y_j\right) = -\sum_i p\left(x_i \mid y_j\right) \log p\left(x_i \mid y_j\right)$$

je entropija od X nakon opažanja specifičnog događaja y_j.

Shannonova zajednička informacija 5.2 može se interpretirati kao prosječna promjena nesigurnosti nakon opažanja Y jer opažanje izlaza Y smanjuje nesigurnost o ulazu X iz $H(X)$ u $H(X|Y)$.

$$I(X;Y) = H(X) - H(X \mid Y) = \sum_{i,j} p\left(x_i, y_j\right) \log \left[\frac{p\left(x_i, y_j\right)}{p\left(x_i\right) p\left(y_j\right)}\right] \tag{5.2}$$

Shannonova zajednička informacija ima svojstvo simetričnosti i svojstvo aditivnosti.

Zajednička informacija 5.2 govori koliko informacije se prosječno prenese preko svih simbola. Međutim, možemo pretpostaviti da su neki simboli informativniji od drugih. U tom slučaju želimo izmjeriti specifičnu informaciju dobivenu iz simbola y_j. Specifičnu informaciju definiramo kao funkcional:

$$\mathrm{I}\left(X; y_j\right) = F\left[p(x), p\left(x \mid y_j\right)\right]. \tag{5.3}$$

Dodatno, zahtijevamo da je prosječna informacija dobivena preko svih specifičnih opservacija jednaka zajedničkoj informaciji:

$$I(X;Y) = \sum_j p\left(y_j\right) I\left(X; y_j\right). \tag{5.4}$$

Čanak [11] nudi dva rješenja koji zadovoljavaju prethodni uvjet: specifično iznenađenje I_1 5.5 i specifična informacija I_2 5.6. Definicija I_2 ima svojstvo aditivnosti, a I_1 nema. I_1 je uvijek nenegativna vrijednost.

$$\begin{aligned}
\mathrm{I}_1\left(X; y_j\right) &= \mathrm{KL}\left(P\left(X \mid y_i\right) \mid\mid P(X)\right) \\
&= \sum_i p\left(x_i \mid y_j\right) \log \left[\frac{p\left(x_i \mid y_j\right)}{p\left(x_i\right)}\right] \qquad (5.5) \\
\mathrm{I}_2\left(X; y_j\right) &= H(X) - H\left(X \mid y_j\right) \\
&= -\sum_i p\left(x_i\right) \log p\left(x_i\right) + \sum_i p\left(x_i \mid y_j\right) \log p\left(x_i \mid y_j\right) \qquad (5.6)
\end{aligned}$$

6. Predložena metoda

U ovom poglavlju definirat ćemo metodu za gustu detekciju anomalija koja se temelji na generativnom modelu i koja se može kombinirati s bilo kojim diskriminativnim modelom za semantičku segmentaciju.

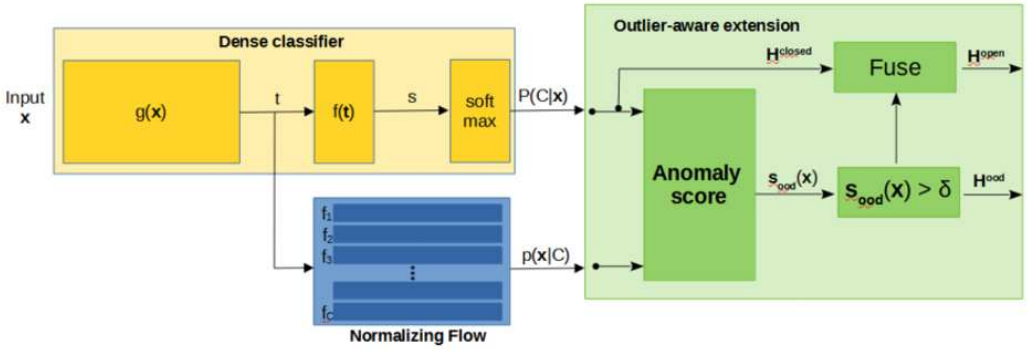
Izlaze diskriminativnog dubokog modela za semantičku segmentaciju možemo raspisati na sljedeći način:

$$P(Y|\mathbf{x}) = \mathrm{softmax}(f_{\theta_2}(g_{\theta_1}(\mathbf{x}))). \tag{6.1}$$

Razlikujemo dvije parametrizirane funkcije g_{θ_1} i f_{θ_2}. Pri tome su $t = g_{\theta_1}(\mathbf{x})$ predlogiti, a $s = f_{\theta_2}(g_{\theta_1}(\mathbf{x}))$ logiti. Funkcijom g_{θ_1} izlučujemo značajke, koje je funkcijom f_{θ_2} potrebno projicirati u prostor čije dimenzije odgovaraju broju klasa kako bi se mogle dobiti distribucije preko klasa kao izlaz klasifikacijskog modela.

Naša ideja je na pred-logitima učiti razredom uvjetovane normalizirajuće tokove. Učimo C normalizirajućih tokova na značajkama pojedine klase, a C je broj razreda u skupu za učenje. Svaki normalizirajući tok modelira $p(\mathbf{x}|c)$ gdje je c razred odgovarajućeg normalizirajućeg toka a $\mathbf{x}$ značajka. Procijenjenu gustoću vjerojatnosti $p(\mathbf{x}|c)$ iskoristit ćemo za detekciju anomalija. Ilustracija pristupa dana je na slici 6.1. Nakon što smo izračunali mjeru mjeru anomalnosti, usporedbom s predefiniram pragom rješavamo binarni zadatak detekcije anomalija. Dodatno, procjenjene mjere anomalnosti možemo kombinirati s diskiminativnim predikcijama na zatvorenom skupu tako da sve anomalne piksele klasificiramo u razred C+1. Time rješavamo zadatak semantičke segmentacije na otvorenom skupu.

U odjeljku 6.1 detaljno opisujemo i objašnjavamo pozadinu predloženih mjera anomalnosti. Definiramo hibridnu mjeru anomalnosti, mjere anomalnosti zasnovane na dodanoj informaciji i mjere anomalnosti zasnovane na sličnosti distribucija modeliranih generativno i diskriminativno. Opisani postupak i sve predložene mjere anomalnosti mogu se kombinirati s raznim diskriminativnim modelima za semantičku segmentaciju. Mi eksperimentiramo s modelima koji se temelje na predikciji na razini piksela 7 i na razini maski 8.



Slika 6.1: Ilustracija hibridne metode za detekciju anomalija korištenjem uvjetnih normalizirajućih tokova. Na pred-logitima normalizirajućeg toka učimo C uvjetnih normalizirajućih tokova. U detektoru anomalija kombiniramo naučenu gustoću vjerojatnosti značajki i diskriminativne predikcije.

6.1. Predložene mjere anomalnosti

Za zadatak guste detekcije anomalija u slikama potrebno je definirati mjeru anomalnosti kao funkciju koja svakom pikselu dodjeljuje realan broj koji govori koliko je taj piksel anomalan (6.2). Niska mjera anomalnosti znači je piksel unutardistribucijski, dok visoka mjera anomalnosti znači da je piksel anomalan.

$$\mathrm{s}(\mathbf{x}) : [0, 255]^{3\times H\times W} \rightarrow \mathbb{R}^{H\times W} \tag{6.2}$$

Želimo napomenuti kako kod definicija mjera anomalnosti nismo u potpunosti točni u označavanju. Osim ako to nije drugačije navedeno, oznaku $\mathbf{x}$ koristimo za značajke dubokog modela, a ne za ulaznu sliku kako je to uobičajeno. Mi mjeru anomalnosti dijelova slike procjenjujemo preko značajki naučenog dubokog modela. Duboki model je smrznut i značajke se iz slike generiraju jedinstvenom funkcijom preslikavanja.

Članak [4] predlaže kao mjeru anomalnosti izravno koristiti izglednost pod modelom:

$$\mathrm{s}_{\mathrm{NLL}}(\mathbf{x}) := -\log p(\mathbf{x}), \tag{6.3}$$

gdje $p(\mathbf{x})$ funkcija gustoće vjerojatnosti značajki $\mathbf{x}$ dubokog modela koju učimo normalizirajućim tokom.

Istu mjeru anomalnosti možemo iskoristiti i u slučaju kada učimo C uvjetnih tokova. Svaki tok u tom slučaju modelira uvjetnu gustoću vjerojatnosti značajki dubokog modela $p_c(\mathbf{x}|c_i)$ pa je:

$$p(\mathbf{x}) = \sum_c p_c(\mathbf{x}, c) = \sum_c p_c(\mathbf{x}|c)P(c) \tag{6.4}$$

gdje je $P(c)$ je apriorna vjerojatnost da primjer pripada razredu c. Možemo pretpostaviti da se $P(C)$ ravna po uniformnoj razdiobi ili možemo izmjeriti udio pojedinih razreda u skupu za učenje. Sada se mjera anomalnosti 6.3 se svodi na:

$$\begin{aligned} \mathrm{s}_{\mathrm{NLL}}(\mathbf{x}) : &= -\log p(\mathbf{x}) \\ &= -\log \sum_c p_c(\mathbf{x}, c) \\ &= -\log \sum_c p_c(\mathbf{x}|c) P(c). \end{aligned} \tag{6.5}$$

Prethodni radovi [34] i [24] upozoravaju na nedostatke izravnog korištenja izglednosti pod modelom kao mjere anomalnosti.

Krenuvši od činjenice da normalizirajući tok modelira izglednost značajki $p(\mathbf{z}|\mathbf{x})$, a ne izravno izglednost ulaznih podataka $\mathbf{x}$, možemo definirati hibridnu mjeru anomalnosti:

$$\begin{aligned} \mathrm{s}_{\mathrm{H}}(\mathbf{x}) : &= -p(\mathbf{z}|\mathbf{x}) \\ &= -\sum_c p(\mathbf{z}|c, \mathbf{x}) P(c|\mathbf{x}) \\ &= -\sum_c p_c(\mathbf{z}|\mathbf{x}) P(c|\mathbf{x}) \end{aligned} \tag{6.6}$$

gdje je $\mathbf{x}$ ulazna slika, a $\mathbf{z}$ značajka. Dakle, naučenu gustoću vjerojatnosti značajki procjenjujemo marginalizacijom preko združenih vjerojatnosti $p(\mathbf{z}|c, \mathbf{x})$. Svaku združenu vjerojatnosti možemo izraziti kao produkt uvjetne gustoće vjerojatnosti i diskriminativne predikcije. U izrazu 6.6 lako prepoznajemo generativni $p_c(\mathbf{z}|\mathbf{x})$ i diskriminativni član $P(c|\mathbf{x})$ pa mjeru nazivamo hibridnom.

Teorija informacije nudi moćan alat za opis i analizu modela strojnog učenja. U poglavlju 5 opisali smo pojam zajedničke informacije i metodu iz literature kojom možemo mjeriti informacijsku dobit nakon opservacije specifičnog događaja. Čanak [11] predlaže dvije formulacije: specifično iznenađenje I_1 5.5 i specifičnu informacija I_2 c. U našem slučaju, razmatramo zajedničku informaciju između razdiobe razreda C i specifičnog podatka $\mathbf{x}$. Intuitivno, što više podatak $\mathbf{x}$ doprinosi zajedničkoj informaciji $I(C; X)$, to je $\mathbf{x}$ više usklađen sa skupom za učenje, odnosno, manje je naše vjerovanje da x nije izvučen iz distribucije skupa za učenje. I specifično iznenađenje 5.5 i specifična informacija 5.5 su prikladni kandidati za mjeru anti-anomalnosti. Specifičnu informaciju I_2 možemo interpretirati kao promjenu nesigurnosti o X nakon opservacije y_i. Specifična informacija inducira isto rangiranje kao i negativna entropija diskriminativnih predikcija H(C|x) koja se uvriježila kao jedna od glavnih mjera nesigurnosti. S druge strane, ako specifično iznenađenje I_1 5.5 primijenimo na naš slučaj,

možemo pisati:

$$\mathrm{I}_1\left(C;\mathbf{x}\right) = \mathrm{KL}\left(P\left(C \mid \mathbf{x}\right) \mid\mid P(C)\right) = \sum_c P\left(c \mid \mathbf{x}\right) \log \frac{P\left(c \mid \mathbf{x}\right)}{P\left(c\right)}. \tag{6.7}$$

Na ovu mjeru normalnosti možemo gledati kao na mekanu varijantu mjere max-softmax. Izlaz je otežani zbroj diskriminativnih predikcija, pri čemu razredi tim više doprinose što je predikcijska vjerojatnost veća od apriorne. Apriornu vjerojatnost $P(C)$ možemo izmjeriti na skupu za učenje kao udio primjera pojedinog razreda ili pretpostaviti da je razdioba uniformna.

Prema Bayesovom pravilu trebalo bi vrijediti:

$$\frac{P(c|\mathbf{x})}{P(c)} = \frac{p(\mathbf{x}|c)}{p(\mathbf{x})}. \tag{6.8}$$

Međutim, zbog različitih induktivnih pristranosti diskriminativnih $P(c|\mathbf{x})$ i generativnih $p(\mathbf{x}|c)$ modela jednakost u praksi ne vrijedi. Stoga, možemo razmotriti i mjeru anomalnosti koja kombinira diskriminativno i generativno modeliranje:

$$\mathrm{I}_{1\mathrm{g}}(C;\mathbf{x}) = \sum_c P\left(c \mid \mathbf{x}\right) \log \frac{p\left(\mathbf{x} \mid c\right)}{p\left(\mathbf{x}\right)}. \tag{6.9}$$

Gustoću vjerojatnosti $p(\mathbf{x})$ računamo kao:

$$p(\mathbf{x}) = \sum_c p_c(\mathbf{x}|c)P(c). \tag{6.10}$$

Predlažemo i alternativni pristup definiciji hibridne mjere anomalnosti. Polazimo od dva modela s različitim induktivnim pristranostima, diskriminativnog b_d i generativnog b_g. KL-divergencija ili relativna entropija je nesimetrična metrika koja se često koristi kao mjera za razliku između dviju distribucija. Predlažemo definirati mjeru anomalija kao relativnu entropiju između distribucije $P(c \mid x)$ pod generativnim i diskriminativnim modelom. Zbog nesimetričnosti KL-divergencije, definiramo mjere anti-anomalnosti s_{gd} 6.11 i s_{dg} 6.12. Očekujemo da će diskriminativna i generativna distribucija biti poravnate za unutardistribucijske primjere. Možemo primijetiti da se mjera s_{dg} svodi na zbroj diskriminativne i generativne verzije specifičnog iznenađenja.

$$\begin{aligned} \mathrm{s_{gd}}(\mathbf{x}) &= KL(P(c \mid \mathbf{x}, b_g) \mid\mid P(c \mid \mathbf{x}, b_d)) \\ &= \sum_c \frac{p(\mathbf{x} \mid c)P(c)}{P(\mathbf{x})} \log \frac{p(\mathbf{x} \mid c)P(c)}{p(c \mid \mathbf{x})p(\mathbf{x})} \end{aligned} \tag{6.11}$$

$$\begin{aligned} \mathrm{s_{dg}}(\mathbf{x}) &= KL(P(c \mid \mathbf{x}, b_d) \mid\mid P(c \mid \mathbf{x}, b_g)) \\ &= \sum_c P(c \mid \mathbf{x}) \log \frac{P(c \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{x} \mid c)P(c)} \\ &= \sum_c P(c \mid \mathbf{x}) \log \frac{P(c \mid \mathbf{x})}{P(c)} + \sum_c P(c \mid \mathbf{x}) \log \frac{P(\mathbf{x})}{P(\mathbf{x} \mid c)} \end{aligned} \tag{6.12}$$

$$= \mathrm{I_1}(C, x) + \mathrm{I_{1g}}(C, x) \tag{6.13}$$

Važno je napomenuti da mjere I_1 6.7, I_{1g} 6.9, s_{gd} 6.11 i s_{dg} 6.12 treba koristiti s negativnim predznakom. To su mjere anti-anomalnosti.

7. Detekcija anomalija u predikcijama modela DeepLabv3+

DeepLabv3+ [7] je diskriminativni model za semantičku segmentaciju koji na zatvorenom skupu postiže dobre rezultate. Opis arhitekture se nalazi u poglavlju 4.1 i ilustriranu na slici 6.1. Koristimo ga kao ekstraktor značajki za generativnu detekciju anomalija predstavljenu u poglavlju 6. U ovom poglavlju detaljno ćemo opisati eksperimente i rezultate detekcije anomalija u predikcijama modela DeepLabv3+.

U svim eksperimentima koristimo na Cityscapesu [10] prednaučeni model DeepLabv3+ s okosnicom WideResNet38 [42] za ekstrakciju značajki. Izlučujemo značajke prije posljednjeg bloka dekodera, tj. pred-logite. Dimenzije tenzora pred-logita definirane su arhitekturom i u našem slučaju su $304 \times H/2 \times W/2$ gdje su H i W prostorne dimenzije ulazne slike. Takve pred-logite bilinearno naduzorkujemo do ulaznih prostornih dimenzija. Na taj način je svaki piksel u slici predstavljen jednim vektorom značajki dimenzije 304. Vektore značajki možemo razdvojiti u grupe ovisno o pripadnosti razredu koristeći oznake slike. To je potrebno za učenje uvjetnih tokova.

Provodimo niz eksperimenata. Polazimo od metode naučene gustoće ugrađivanja iz [4] (odjeljak 7.1). Uvodimo uvjetne normalizirajuće tokove (odjeljak 7.2). Dodatno, uvodimo negativan skup podataka i eksperimentiramo s diskriminativnom binarnom glavom i hibridnom mjerom anomalnosti (odjeljak 7.5). Ističemo eksperimente opisane u odjeljku 7.2 kao fokus ovoga rada.

Kako bismo uštedjeli na računskim i vremenskim resursima, jednim unaprijednim prolazom kroz sve slike za učenje iz skupa Cityscapes izlučujemo pred-logite i spremamo ih u datoteku uz pripadajuće oznake. Takav pristup koristimo u svim eksperimentima osim u onima opisanima u poglavlju 7.5. Analiza hiperaparametara učenja toka i analiza arhitekture napravljena je u odjeljku 7.4.

7.1. Naučena gustoća ugrađivanja

U članku [4] predstavljena je metoda za detekciju anomalija koja se temelji na učenju gustoće ugrađivanja dubokog modela. Autori predlažu modelom za semantičku segmentaciju izlučiti značajke slike, tj. ugrađivanja, u različitim slojevima modela. Na takvim značajkama nezavisno uče ansambl normalizirajućih tokova minimizacijom negativne log-izglednosti. Procjene negativne log-izglednosti tokova učenih na različitim skupovima značajki ne mogu se izravno agregirati zbog toga što ugrađivanja na različitim slojevima imaju različite distribucije pa procijenjene gustoće imaju različite skale. Autori predlažu normalizaciju negativne log-izglednosti ugrađivanja prosječnom log-izglednosti svih ugrađivanja l-tog sloja:

$$\bar{N}\left(\mathbf{z}_l\right) = N\left(\mathbf{z}_l\right) - \mathcal{L}\left(\mathbf{Z}_l\right). \tag{7.1}$$

Predložene su dvije metode agregacije izglednosti: (i) piksel je anomalan ako ima nisku izglednost kroz sve slojeve i (ii) koristeći otežani prosjek gdje su težine naučene logističkom regresijom na FS Lost&Found validacijskom skupu. Dodatno, autori koriste i izglednost samo jednog sloja kao mjeru anomalnosti. Koriste RealNVP normalizirajući tok od 32 sloja miješanja. Slojevi ugrađivanja odabrani su validacijom na FS LostAndFound validacijskom skupu.

Mi provodimo sličan eksperiment. Polazimo od pretpostavke da pred-logiti dubokog modela sadrže semantički bogatije značajke nego značajke u ranijim slojevima koje se koriste u [4]. Također, pretpostavljamo da je za učenje distribucije značajki dovoljna minimalna arhitektura normalizirajućeg toka. Koristimo normalizirajući tok od dva sloja miješanja, s fiksnom permutacijom dimenzija između slojeva i dva potpuno povezana sloja sa skrivenom dimenzijom 152 u modulu afinog miješanja. Kao mjeru anomalnosti koristimo s_{NLL} 6.5.

Usporedba rezultata naših eksperimenata i rezultata metode iz [4] navedeni su u tablici 7.1. U prva tri retka nalaze se rezultati koje su s nama podijelili autori članka [4]. Naš eksperiment donekle je usporediv s eksperimentom detekcije anomalija procjenom izglednosti jednog sloja, tj. prvim retkom u tablici 7.1. Dobivamo lošije rezultate od [4]. Treba istaknuti da ti eksperimenti nisu u potpunosti usporedivi. Učimo na značajkama različite semantičke informacije i koristimo puno manji model normalizirajućeg toka. U [4] koristili su RealNVP s 32 sloja miješanja, dok mi koristimo minimalnu arhitekturu sa samo 2 sloja miješanja. Zaključujemo da trenutni model normalizirajućeg toka nema dovoljan kapacitet za učenje distribucije značajki.

Tablica 7.1: Rezultati naučene gustoće ugrađivanja [4] na FS Lost&Found validacijskom skupu.

Metoda	AP [%]	FPR_{95} [%]	AUROC [%]
NLL jednog sloja [4]	1.3	74.4	-
minimalna NLL [4]	7.3	47.9	-
logistička regresija [4]	5.9	36.4	-
NLL pred-logita	0.58	54.28	72.35

7.2. Detekcija anomalija uvjetnim tokovima

U odjeljku 7.1 pokazali smo kako minimalna implementacija normalizirajućeg toka nema dovoljan kapacitet za učenje distribucije značajki. Predlažemo grupirati značajke s obzirom na njihovu pripadnost razredu i na svakoj grupi značajki učiti jedan razredom uvjetovani tok kao što je to ilustrirano na slici 6.1.

Evaluiramo performanse na skupovima FS Lost&Found i FS Static koristeći mjere anomalnosti predložene u poglavlju 6. Rezultati su prikazani u tablicama 7.2 i 7.3. U tablici 7.2 su rezultati uz tokove minimalne arhitektrue po uzoru na RealNVP [12]. Normalizirajući tok sastoji se od dva sloja miješanja s fiksnom permutacijom dimenzija između slojeva i dva potpuno povezana sloja sa skrivenom dimenzijom 152 u modulu afinog miješanja. U tablici 7.3 dani su rezultati uz tokove optimizirane arhitekture. Tokovi se sastoje od dva sloja miješanja uz normalizaciju aktivacije i fiksnu alternirajuću permutaciju. Modul afinog miješanja je jedan rezidualni blok. Detalji optimizacije arhitekture normalizirajućeg toka i ostali hiperparametri detaljno su opisani u odjeljku 7.4. Rezultati uz optimizirani arhitekturu toka su značajno bolji. U nastavku ćemo obrazložiti upravo rezultate iz tablice 7.3 dobivene uz optimiziranu arhitekturu toka. Slika 7.1 daje kvalitativan prikaz performansi predloženih detektora anomalnosti.

Performanse predloženih detektora anomalija uspoređujemo s osnovnim (engl. *baseline*) metodama maksimalnog logita [21] i maksimalne vjerojatnosti softmaksa [19]. Maksimalni logit daje bolje rezultate od maksimalne vjerojatnosti softmaksa. Na slici 7.1 vidimo da maksimalni logit mnogim unutardistribucijskim pikselima dodjeljuje visoku mjeru anomalnosti i uočavamo već poznati problem maksimalne vjerojatnosti softmaksa koji griješi na granicama razreda. Kao što je i očekivano mjera s_{NLL} (6.3) daje najlošije rezultate. To je isključivo generativni pristup gdje do izražaja dolaze problemi opisani primjerice u [34] i [24]. Hibridna mjera s_H (6.6) daje dobre rezul-

tate. Na slici 7.1 vidimo da neki anomalni pikseli imaju manju mjeru anomalnosti nego unutardistribucijski pikseli. S hibridnom mjerom postižemo najmanji FPR_{95} i na FS Lost&Found i na FS Static. Ipak, na tim skupovima je primarna mjera prosječna preciznost. Najbolju prosječnu preciznost i AUROC postižemo uz mjeru anomalnosti $\text{I}_{1\text{g}}$. Mjera I_1 daje prilično loše rezultate. To potvrđuje hipotezu da samo diskriminativni modeli nisu dovoljni za dobru detekciju anomalija. Mjere zasnovane na razlici distribucija modeliranih generativno i diskriminativno daju lošije rezultate od mjera zasnovanih na informacijskoj dobiti. Zanimljivo je primijetiti kako s_{dg} na FS Static postiže prilično dobre rezultate, točnije drugi najbolji FPR_{95}.

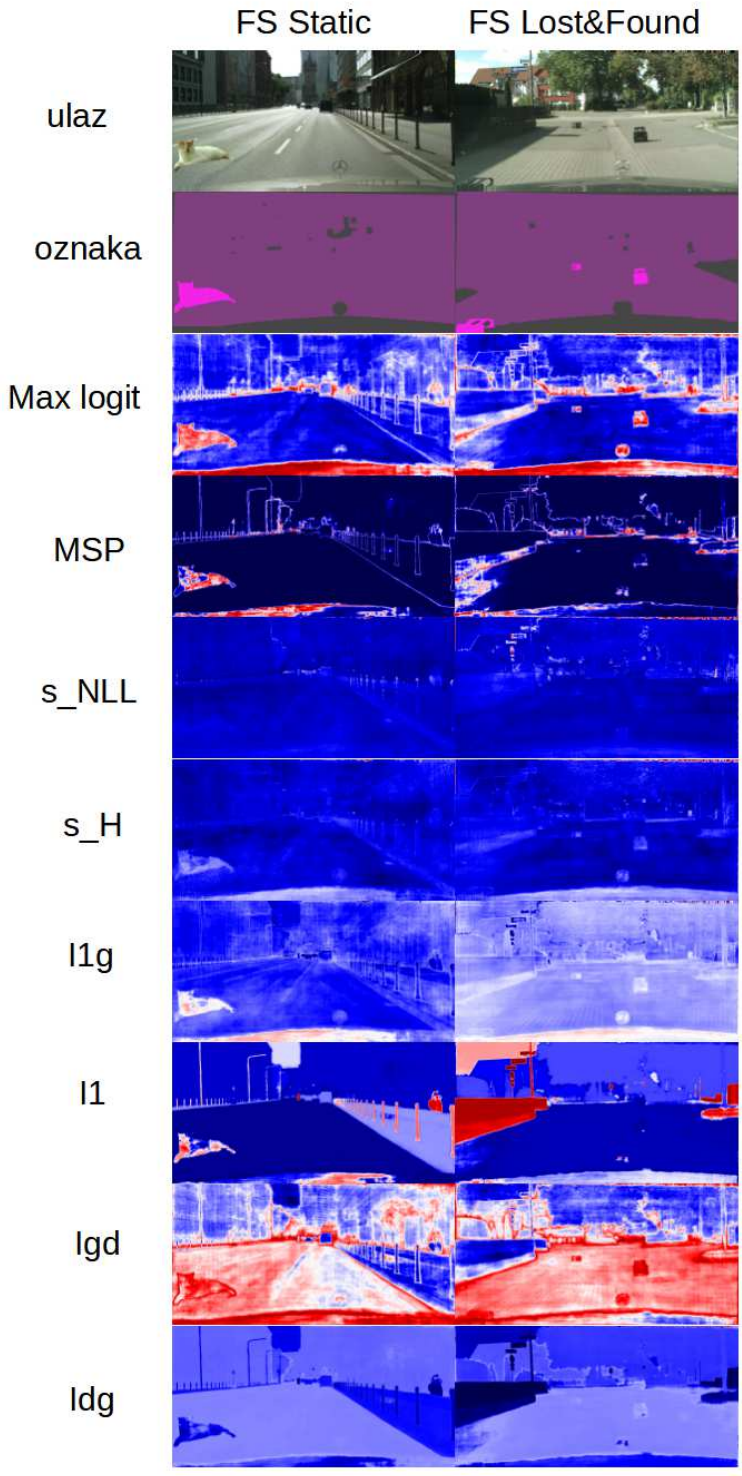
Pogledamo li detaljnije sliku 7.1 možemo primijetiti da I_1 daje zanimljive procjene nesigurnosti. Možemo primijetiti da su svim pikselima pojedinih klasa dodijeljene visoke visoke mjere nesigurnosti i to klasama s manjim udjelom u skupu za učenje. Taj problem bismo mogli riješiti točnijim kalibriranjem vjerojatnosti $P(c)$. U prethodnim eksperimentima $P(C)$ smo definirali kao udio piksela pojedinog razreda u skupu za učenje. Odlučujemo $P(C)$ definirati kao uniformnu razdiobu, točnije $P(c_i) = 1/19$. Ponovo mjerimo rezultate i dobivamo poboljšanje za mjeru I_1. Međutim mjera $\text{I}_{1\text{g}}$ ne daje poboljšanje. Pretpostavljamo da diskriminativni modeli pretpostavljaju uniformnu apriornu vjerojatnost. Generativni modeli dobro reagiraju na izmjerenu apriornu vjerojatnost jer su tokovi za različite klase učili na različitom broju značajki za učenje koji odgovara izmjerenoj apriornoj vjerojatnosti. Na slici 7.2 prikazani su rezultati detekcije anomalija uz uniformnu i na skupu za učenje izmjerenu apriornu distribuciju. Usporedimo li dobiveni rezultat uz unifrmnu distribuciju s rezultatom za MSP na slici 7.1 vidimo da na mjeru I_1 možemo gledati kao na mekanu varijantu maksimalne vjerojatnosti softmaksa.

Tablica 7.2: Rezultati detekcije anomalija uvjetnim tokovima na FS Lost&Found i FS Static validacijskim skupovima. Korištena je minimalna implementacija toka po uzoru na RealNVP. Najbolji rezultati su podebljani, a drugi po redu najbolji podvučeni.

	FS Lost&Found			FS Static		
Mjera anomalnosti	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
max logit	**12.57**	<u>24.43</u>	<u>90.49</u>	19.87	<u>13.02</u>	**92.29**
MSP	3.99	30.28	85.16	12.08	16.91	88.62
s_{NLL}	0.43	77.84	62.29	2.34	54.76	58.17
s_H	5.26	38.36	85.84	<u>23.17</u>	20.46	89.29
I_{1g}	<u>10.65</u>	26.82	89.43	**33.35**	**14.01**	<u>92.17</u>
I_1	5.01	55.84	72.66	9.88	57.77	49.75
s_{gd}	7.01	**16.82**	**94.16**	18.06	20.83	89.91
s_{dg}	6.2	53.31	75.67	11.81	56.77	52.57

Tablica 7.3: Rezultati detekcije anomalija uvjetnim tokovima na FS Lost&Found i FS Static validacijskim skupovima. Korištena je arhitektura tokova po uzoru na Glow. Najbolji rezultati su podebljani, a drugi po redu najbolji podvučeni.

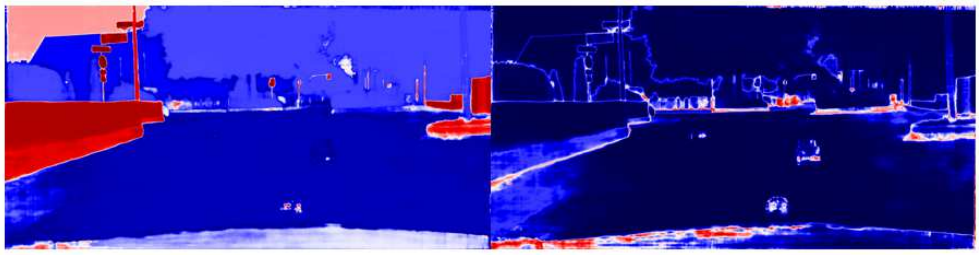
	FS Lost&Found			FS Static		
Mjera anomalnosti	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
max logit	12.57	24.43	90.49	19.87	13.02	92.29
MSP	3.99	30.28	85.16	12.08	16.91	88.62
s_{NLL}	0.69	52.77	70.6	3.38	46.62	66.78
s_H	<u>15.95</u>	**14.01**	<u>94.48</u>	<u>28.68</u>	**12.59**	<u>93.47</u>
I_{1g}	**20.9**	<u>15.79</u>	**94.86**	**37.38**	15.38	**94.10**
I_1	5.01	55.84	72.66	9.88	57.77	49.75
s_{gd}	4.24	28.63	88.41	10.35	30.9	81.01
s_{dg}	9.15	48.59	78.63	24.38	<u>12.61</u>	92.78



Slika 7.1: Kvalitativna usporedba metoda za detekciju anomalija na razini piksela. Slike prikazuju, ulaznu testnu sliku, pripadajuću oznaku i procijenjene nesigurnosti predikcija. Crvena boja označava visoku nesigurnost, a plava boja visoku sigurnost u predikciju.

Tablica 7.4: Rezultati detekcije anomalija na validacijskom skupu Lost&Found mjerama zasnovanim na informacijskom dobitku s obzirom na odabir apriorne distribucije $P(C)$.

mjera anomalnosti	P(C)	AP [%]	FPR_{95} [%]	AUROC [%]
I_1	udio razreda	5.01	55.84	72.66
	uniformno	9.3	29.13	86.61
I_{1g}	udio razreda	20.90	15.79	94.86
	uniformno	14.05	17.78	93.04



Slika 7.2: Rezultati procjene nesigurnosti mjere anomalnosti I_1 uz uniformnu i na skupu za učenje izmjerenu apriornu distribuciju P(C) na slici iz FS Lost&Found.

Normalizirajući tok možemo učiti i kontrastivno prema izrazu 3.12. Detalji su opisani u poglavlju 3.1. Treba biti oprezan i napomenuti da kontrastivnim učenjem ne učimo izglednost primjera nego razliku izglednosti unutardistribucijskih i izvandistribucijskih primjera. Koristimo mjeru anomalinosti s_H i I_{1g}. Rezultati detekcije anomalija kontrastivno naučenim tokovima i usporedba s dekecijom anomalija standardno naučenim tokovima dana je u tablici 7.5. Koristili smo minimalnu arhitekturu normalizirajućih tokova s dva sloja miješanja i fiksnom alternirajućom permutacijom. Veći tokovi dovode do numeričkih nestabilnosti. Kontrastivno naučeni tokovi donose poboljšanje za hibridnu mjeru anomalnosti, ali ne i za mjeru zasnovanu na dodanoj informaciji.

Tablica 7.5: Rezultati detekcije anomalija kontrastivno naučenim tokovima na FS Lost&Found i FS Static validacijskim skupovima.

		FS Lost&Found			FS Static		
Mjera anomalnosti	Učenje	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
s_H	standardno	5.26	38.36	85.84	23.17	20.46	89.29
	kontrastivno	17.67	22.27	91.10	37.41	9.58	93.31
I_{1g}	standardno	10.65	26.82	89.43	33.35	14.01	92.17
	kontrastivno	7.77	49.31	78.3	32.43	33.79	83.00

7.3. Učenje s kraja na kraj

U prethodnim eksperimentima normalizirajuće tokove smo učili na prethodno spremljenim značajkama dubokog modela. U sljedećim eksperimentima smo duboki model i uvjetne normalizirajuće tokove učimo s kraja na kraj. Uzimamo na Cityscapesu naučeni DeepLabv3+ i fino ga ugađamo 3 epohe, a tokove učimo iz slučajne inicijalizacije. Gradijenti iz tokova se propagiraju u segmentacijski model. Gubitak možemo definirati kao:

$$L = L_{ce} + \beta \cdot L_{nll}, \tag{7.2}$$

gdje je β hiperaparametar gubitka. U [43] autori pokazuju da učenje s kraja na kraj pomaže. Naša pretpostavka je da takvim načinom učenja segmentacijski model može producirati kvalitetnije značajke koje će biti poravnate s distribucijom naučenom normalizirajućim tokom. Hiperparametar β postavljamo na 1.

Usporedba rezultata modela učenog s kraja na kraj i slučaja kada je segmnetacijski model smrznut dana je u tablici 7.6. Razlika u rezultatima je unutar očekivane varijance i iz ovog eksperimenta ne možemo zaključiti donosi li učenje s kraja na kraj poboljšanje.

Tablica 7.6: Rezultati učenja segmentacijskog modela i normalizirajućih tokova s kraja na kraj uz mjeru anomalnosti s_H na validacijskom skupu FS Lost&Found.

Učenje	AP [%]	FPR_{95} [%]	AUROC [%]
samo NF	5.26	38.36	85.84
s kraja na kraj	5.37	43.78	82.31

7.4. Validacija hiperparametara učenja i arhitekture normalizirajućeg toka

U ovom odjeljku detaljno pokazujemo validaciju hiperaparametara učenja i arhitekture normalizirajućih tokova. Kao mjeru anomalnosti u svim eksperimentima koristimo I_{1g} 6.9. Sve eksperimente u ovom odjeljku proveli smo na validacijskom skupu FS Static zbog ograničenosti vremenskim i računskim resursima.

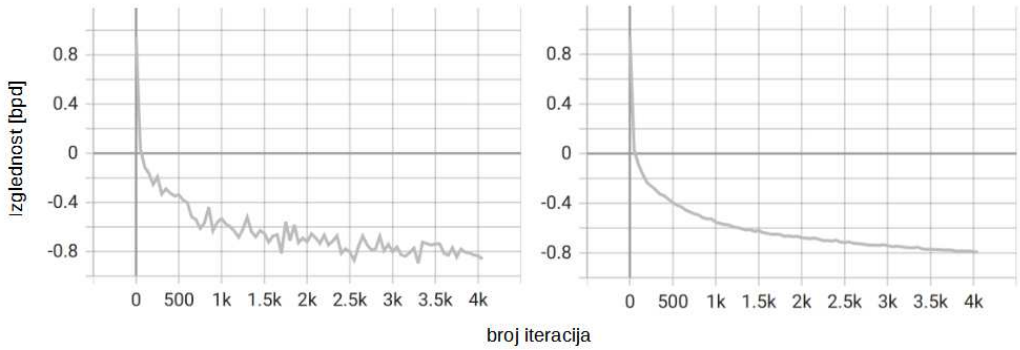
U ranoj fazi istraživanja krenuli smo od minimalne implementacije normalizirajućeg toka zbog pretpostavke da je takav tok dovoljan za procjenu gustoće vjerojatnosti značajki dubokog modela. Definirali smo arhitekturu po uzorku na RealNVP [12]. Minimalna implementacija sastoji se od dva sloja miješanja. Koristimo alternirajući uzorak između slojeva miješanja, tako da u svakom koraku miješanja naizmjenično miješamo jednu polovinu reprezentacije. Modul afinog miješanja sastoji se od dva potpuno povezana sloja sa zglobnicom kao nelinearnom aktivacijom. Dimenzija skrivenog sloja je 152. Upravo tu implementaciju koristili smo u eksperimentima u 7.

Kako se generativni model ne bi prenaučio potrebno je definirati uvjet ranog zaustavljanja. Iz skupa značajki u validacijski skup izdvajamo 20% značajki svake klase. Učenje zaustavljamo u trenutku kada više nema znatnog poboljšanja na validacijskom skupu. Pratimo vrijednosti negativne log-izglednosti u bitovima po dimenziji na skupu za učenje i validacijskom skupu (slika 7.3). Dodatno, na skupu FS Static evaluirano metodu I_{1g} uz normalizirajuće tokove učene uz minigrupu 16 i korak učenja 0.001 u različitim iteracijama (tablica 7.7). U iteraciji 1000 dobivamo najbolji AP, ali je FPR nešto lošiji nego u ranijim iteracijama. Primjećujemo da te dvije veličine ne koreliraju, tj. da visoki AP dolazi uz cijenu nižeg FPR. Odabirmo iteraciju 1000 jer je AP primarna mjera.

Tablica 7.7: Validacija optimalnog broja iteracija učenja tokova.

Iteracija	AP [%]	FPR_{95} [%]	AUROC [%]
100	25.13	**13.17**	91.63
500	28.29	14.03	93.13
1000	**33.13**	15.86	93.69
5000	33.8	25.19	89.8

U članku [28] autori koriste normalizirajući tok za generiranje značajki i koriste arhitekturu Glow [23]. I mi definiramo arhitekturu po uzorku na Glow. Takav nor-



Slika 7.3: Prikaz krivulje učenja na skupu za učenje i validacijskom skupu značajki koje pripadaju razredu cesta.

malizirajući tok sastoji se od dva sloja miješanja. Koristimo naučenu permutaciju, a modul miješanja se sastoji od dva potpuno povezana sloja sa zglobnicom kao nelinearnom aktivacijom. Usporedba arhitekture normalizirajućeg toka po uzoru na RealNVP i Glow prikazana je u tablici 7.12. Normalizirajući tokovi su se učili 1000 iteracija uz minigrupu 16 korakom učenja 0.001.

Tablica 7.8: Usporedba arhitektura normalizirajućeg po uzoru na RealNVP i Glow na validacijsom skupu FS Static.

Arhitektura	AP [%]	FPR_{95} [%]	AUROC [%]
RealNVP	33.35	14.01	92.17
Glow	29.61	32.74	85.15

Primjećujemo da Glow arhitektura daje lošije performanse suprotno navodima iz [28]. Pretpostavljamo da različite primjene i vrste značajki zahtijevaju različite arhitekture. Glow možemo shvatiti kao nadogradnju na RealNVP. Provodimo ablaciju osnovnih komponenti inicijalno predložene arhitekture Glow. Tokovi su učeni 1000 iteracija uz minigrupu 16 i korak učenja 0.001. Primjećujemo da korištenje naučene permutacije donosi pogoršanje. Pretpostavljamo da konvolucija 1×1 uvodi nove parametre te nepotrebno povećava model koji se zbog pre velikog kapaciteta prenuači na skupu za učenje.

U prethodnim eksperimentima arhitekturu modula afinog miješanja definirali smo kao dva potpuno povezana sloja sa zglobnicom kao nelinearnom aktivacijom gdje je dimenzija skrivenog sloja 152. Predlažemo uvećati dimenziju skrivenog sloja na 304 ili koristiti jedan rezidualni blok umjesto samo potpuno povezanih slojeva. Učimo

Tablica 7.9: Ablacija komponenti arhitekture Glow na validacijskom skupu FS Static.

ActNorm	conv 1x1	AP [%]	FPR_{95} [%]	AUROC [%]
✓	✓	29.61	32.74	85.15
✗	✓	13.35	36.69	81.13
✓	✗	34.98	25.56	89.56
✗	✗	33.35	14.01	92.17

normalizirajući tok argitekture Glow uz različite arhitekture modula afinog miješanja kao što je to prethodno opisano. Tokovi su učeni 1000 iteracija uz minigrupu 16 i korak učenja 0.001. Pokazuje se da veći modul afinog miješanja donosi poboljšanje.

Tablica 7.10: Validacija arhitekture modula afinog miješanja.

Arhitektura modula afinog miješanja	AP [%]	FPR_{95} [%]	AUROC [%]
FC, dim_h = 152	29.61	32.74	85.15
FC, dim_h = 304	31.23	32.36	85.97
Rezidualni blok	32.65	29.82	87.23

Prethodni eksperimenti pokazuju da bi dobra arhitektura bila 2 sloja miješanja uz normalizaciju aktivacije i fiksnu alternirajuću permutaciju. Nad takvom arhitekturom validaciju veličine minigrupe značajki. Pokazuje se da je optimalna veličina minigrupe 64.

Tablica 7.11: Validacija veličine minigrupe.

Veličina minigrupe	AP [%]	FPR_{95} [%]	AUROC [%]
16	38.39	20.64	92.73
64	37.38	15.38	94.1
256	33.94	15.83	93.63

U članku [4] govore da uvećanje značajki šumom donosi poboljšanje. Prethodno opisanu arhitkturu tokova učimo i na značajkama uvećanim šumom. Rezultati su prikazani u tablici

Tablica 7.12: Utjecaj uvećavanja značajki Gaussovim šumom.

uvećanje Gaussovim šumom	standardna devijacija šuma	AP [%]	FPR_{95} [%]	AUROC [%]
✗	-	37.38	**15.38**	94.1
✓	0.01	37.87	17.05	93.95
✓	0.1	**38.07**	22.66	91.44

Validiramo broj slojeva miješanja. Najbolji AP postižemo za 2 sloja miješanja, a najbolji FPR za 3 sloja miješanja uz pad mjere AP. Odabiremo 2 sloja miješanja zbog toga što je AP primarna mjera. Uzmemo li 4 sloja miješanja, performanse padaju. Pretpostavka je da veći model treba više iteracija učenja.

Tablica 7.13: Ablacija komponenti arhitekture Glow na validacijskom skupu FS Static.

broj slojeva miješanja	AP [%]	FPR_{95} [%]	AUROC [%]
2	**37.38**	15.38	94.1
3	34.38	**9.58**	95.33
4	19.76	13.81	92.67

7.5. Eksperimenti s binarnom glavom za detekciju anomalnih primjera

Na pred-logitima na Cityscapesu naučenog DeepLabv3+ modela učimo C razredom uvjetnih normalizirajućih tokova, gdje je C broj razreda u skupu za učenje. Svaki tok modelira gustoću vjerojatnosti $p(\mathbf{x}|y = c)$. Uvjetne tokove učimo kontrastivno 3.12. Jedan korak učenja sastoji se od nekoliko dijelova. Prvo je potreban unaprijedni prolaz kroz segmentacijski model kojim dobivamo pred-logite. Pred-logite dimenzije $B \times C \times H \times W$ preraspoređujemo u matricu dimenzija $N \times C$, gdje je N ukupan broj vektora značajki dobivenih iz jedne minigrupe slika. Za svaku značajku možemo iz označene slike dobiti razred kojem pripada. Nezavisno kontrastivno učimo c tokova gdje su pozitivi primjeri koji pripadaju c-tom razredu, a negativi primjeri svih ostalih razreda. Zbog nesrazmjera broja pozitiva i negativa, slučajno poduzorkujemo negativne primjere tako da broj pozitiva i negativa bude jednak. Veličina minigrupe značajki je varijabilna, jer je u slici zastupljenost pojedinih klasa različita. Zbog raču-

nalnih ograničenja, maksimalnu veličinu minigrupe značajki ograničavamo na 1000. Segmentacijski model je smrznut. Marginu u kontrastivnom gubitku postavljamo na 5. Uvjetne tokove na opisani način učimo 5 epoha. Koristimo mjeru anomalnosti definiranu izrazom 6.6.

Po uzoru na [2] i [3] dodajemo diskriminativnu binarnu glavu za detekciju izvan-distribucijskih primjera:

$$P\left(d_{in} \mid \mathbf{x}\right) := \sigma\left(g_{\gamma}\left(q_{\theta_1}(\mathbf{x})\right)\right). \tag{7.3}$$

Na taj način modeliramo vjerojatnost je li primjer anomalan ili ne. Binarnu glavu izvodimo kao nelinearnu transformaciju pred-logita g_{γ}, konkretno BN-ReLU-Conv1x1, popraćenu sigmoidalnom aktivacijom. Definiramo anomalnosti koja u obzir uzima i predikcije binarne glave po uzoru na [16]:

$$s(\mathbf{x}) := \ln \frac{1 - P\left(d_{in} \mid \mathbf{x}\right)}{\sum_c p(\mathbf{z}|\mathbf{x})}. \tag{7.4}$$

Ovakav model zahtijeva stvarni negativni skup podataka. Na slike za učenje lijepimo instance podataka ade20k [44]. Fino ugađamo segmentacijski model 5 epoha i paralelno učimo uvjetne tokove kao što je prethodno opisano. Samo gradijenti binarne glave utječu na segmentacijski model. Gradijenti uvjetnih tokova se ne propagiraju u segmentacijski model.

Rezultati su prikazani u tablicama 7.14 i **??**. Binarna glava donosi zanimljivo, ali i očekivano poboljšanje. To možemo interpretirati tako da finim ugađanjem segmentacijskog modela uz binarnu glavu poboljšavamo značajke značajke za učenje tokova. Korištenje binarne glave i stvarnog negativnog skupa podataka je jako zanimljiv pristup s puno mogućnosti za daljnje istraživanje, ali se u nastavku rada fokusiramo metode koje ne zahtijevaju korištenje negativnog skupa podataka.

Tablica 7.14: Rezultati detekcije anomalija kontrastivno naučenim tokovima bez i sa diskriminativnom binarnom glavom na FS Lost&Found i FS Static validacijskim skupovima.

		FS Lost&Found			FS Static		
Eksperiment	Mjera anomalnosti	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
uvjetni tokovi	$s^{H}(\mathbf{x})$	10.54	23.62	90.58	37.23	9.58	93.28
uvjetni tokovi i binarna glava	$s^{H}(\mathbf{x})$	17.22	18.76	92.58	36.94	13.43	94.41
	$s^{OOD-H}(\mathbf{x})$	15.54	25.47	89.82	42.21	7.97	96.77

7.6. Detekcija anomalija uvjetnim modelima Gaussove mješavine

Za detekciju anomalija odlučili smo koristiti normalizirajuće tokove zbog mogućnosti eksplicitne i efikasne procjene izglednosti. Alternativa normalizirajućem toku je model Gaussove mješavine (engl. *Gaussian mixture model*), skraćeno GMM. Model Gaussove mješavine je probabilistički model koji modelira pozadinsku distribuciju podataka kao mješavinu Gaussovih distribucija. Takav model je definiran brojem Gaussovih komponenti, parametrima distribucije svake komponente i težinama miješanja komponenti. Uči se iterativnim algoritmom maksimizacije očekivanja (engl. *Expectation-Maximization*), skraćeno EM. Jedna iteracije toga algoritma može se rastaviti na dva koraka. U koraku E računa se izglednost da primjer pripadda svakoj od komponenti, a u koraku M ažuriraju se parametri komponenti s obzirom na prethodno izračunate izglednosti. GMM ima mogućnost egzaktne procjene izglednosti, ali i nedostatke kao što su pretpostavka da je pozadinska distribucija mješavina Gaussovih distribucija, sporija konvergencija i osjetljivost na inicijalizaciju.

U članku [30] autori koriste model Gaussovih mješavina za klasifikaciju i detekciju anomalija. Na sličan način kao i mi uče razredom uvjetovane generativne modela, a mjeru anomalnosti definiraju kao:

$$s(\mathbf{x}) = -\max_{c} p(\mathbf{x}|c). \tag{7.5}$$

Provodimo sljedeći eksperiment kako bismo opravdali korištenje normalizrajućih tokova. Na značajkama svakog razreda učimo jedan model Gaussove mješavine. Koristimo gotovu implementaciju iz biblioteke scikit-learn. Prepostavljamo 5 komponenti miješanja, kao što je sugerirano u [30]. Vrijeme učenja pojedine Gaussove mješavine ovisi o broju primjera, a za učenje svih 19 modela Gaussove mješavine trebalo nam je oko 6 sati, što je zamjetno dulje od učenja normalizirajućih tokova koje za minimalni tok traje oko 20 minuta. Mjerimo i ukupno vrijeme unaprijednog prolaza ovako definiranog modela. Vrijeme potrebno za unaprijedni prolaz kroz DeepLabv3+ i 19 Gaussovih mješavina za jednu sliku je 23 minute. Metode s tako dugim vremenom evaluacije slabo su primjenjive. Usporedba performansi metode detekcije anomalija normalizirajućim tokovima i modelima Gaussove mješavine na skupu FS Static uz mjeru anomalnosti $\mathrm{s}^{\mathrm{H}}(\mathbf{x})$ dana je u tablici 7.15.

Normalizirajući tokovi imaju veću ekspresivnost, odnosno nisu ograničeni brojem prethodno definiranih komponenti. Zbog toga dobivamo bolji AP. U članku [30] dani su rezultati na velidacijskim skupovima FS Lost&Found i RoadAnomaly. Nisu dani

Tablica 7.15: Usporedba normalizirajućih tokova i modela Gaussove mješavine na validacijskom skupu FS Static.

Model	Metoda	AP [%]	FPR_{95} [%]	AUROC [%]
uvjetni normalizirajući tokovi	$\mathrm{s}^{\mathrm{H}}(\mathbf{x})$	33.35	14.01	92.17
uvjetni modeli Gaussove mješavine	$\mathrm{s}^{\mathrm{H}}(\mathbf{x})$	19.68	14.74	91.71

rezultati za FS Static. Mi eksperimente provodimo samo na FS Static zbog vremena izvođenja i stoga nismo u mogućnosti usporediti rezultate.

8. Detekcija anomalija u predikcijama modela Mask2Former

Vodeći se pretpostavkom da metode za gustu detekciju anomalija na razini piksela ne uzimaju u obzir korelaciju susjednih piksela, detekciju anomalija prebacujemo na razinu maski. Za to koristimo model Mask2Former [9] koji je detaljno opisan u poglavlju 4.2.

U članku [17] predstavljene su diskriminativne metode za detekciju anomalija u predikcijama modela Mask2Former. Autori pokazuju da metode detekcije anomalija na razini maski manje griješe na granicama razreda i općenito postižu bolje rezultate nego metode na razini piksela.

Uvjetnim normalizirajućim tokovima modeliramo gustoću vjerojatnosti ugrađivanja maski $p(\mathbf{x}|c)$. Naša pretpostavka je da će korištenje generativnih modela poboljšati rezultate, a posebno smanjiti pojavu lažnih negativa s visokom sigurnosti u predikciju i posljedično mjeru FPR_{95}. Naučene gustoće vjerojatnosti možemo lako kombinirati s diskriminativnim predikcijama.

U svim eksperimentima u ovom poglavlju koristimo Mask2Former model s okosnicom Swin-L [32] kako bi naši rezultati bili usporedivi s rezultatima iz [17].

Detekcija anomalija u predikcijama modela Mask2Former može se provesti na razini maski ili na razini piksela. Detekcija anomalija na razini piksela je vrlo slična postupcima opisanima u poglavlju 7.

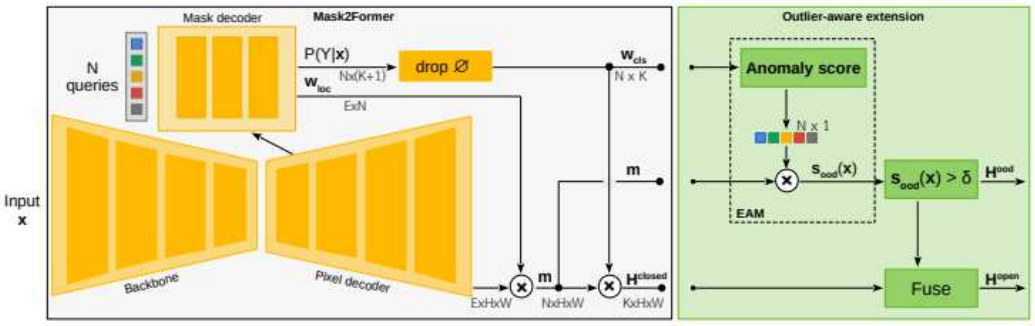
8.1. Detekcija anomalija na razini maski

U ovom poglavlju provodimo eksperimente detekcije anomalija uvjetnim tokovima na razini maski.

Diskriminativna metoda za detekciju anomalija ansambliranjem mjera anomalnosti na razini maski predložena je u članku [17]:

$$\mathbf{s}_{\text{ood}}^{\text{EAM}}(\mathbf{x})[r,c] = \sum_i \mathbf{m}_i[r,c] \cdot \left(- \max_{k=1\ldots K} P_i(Y=k \mid \mathbf{x}) \right). \tag{8.1}$$

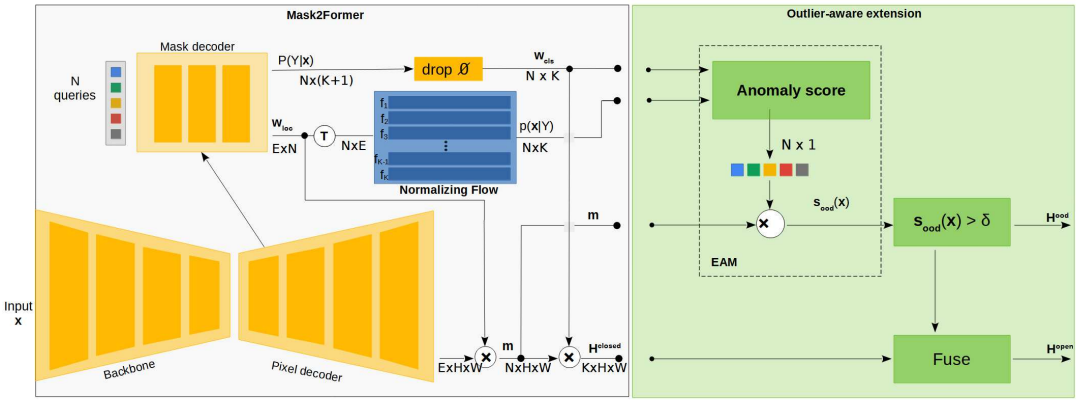
Kratica EAM dolazi od engleskog naziva *Ensemble over Anomaly scores of Mask-wide predictions*.



Slika 8.1: Slika prikazuje tenzore izlaza modela Mask2Former i njihove dimenzije i dodatak za detekciju anomalija i predikciju na otvorenom skupu korištenjem metode EAM. Slika je preuzeta iz [17].

Predlažemo učiti gustoću vjerojatnosti ugrađivanja maski $p(\mathbf{x}|y=k)$ za svaki razred k. Konkretno na izlazima $\mathbf{w}_{loc}^{E\times N}$ učimo K uvjetnih normalizirajućih tokova gdje je K broj razreda u skupu za učenje. Skup za učenje je Cityscpaes, pa je K=19. Broj maski N postavljamo na 100. U [17] pokazano je da veći broj maski znatno poboljšava performanse detektora anomalija. Dimenzija ugrađivanja maski E definirana je arhitekturom transformer dekodera i iznosi 256. Izlaz jednog uvjetnog toka je dimenzija $N \times 1$. Izlaze K uvjetnih tokova možemo posložiti u jednu matricu dimenzija $N \times K$. Ta matrica će po recima sadržavati informaciju $p(\mathbf{x}|y=k)$. Možemo reći da je u cijeloj matrici sadržana informacija $p(\mathbf{x}|Y)$. Prethodno opisani tenzor je uz diskriminativnu predikciju ulaz u detektor anomalija. Opisana metoda je prikazana na slici 8.2.

Mjere anomalnosti predložene u poglavlju 6 mogu se osim na piksele primijeniti i na maske. Međutim, za gustu detekciju anomalija potrebna nam je konačna mjera anomalnosti po pikselima (6.2). Predlažemo agregirati mjeru anomalnosti na razini



Slika 8.2: Predložena metoda za generativnu detekciju anomalija u predikcijama na razini maski modela Mask2Former.

maske u svakom pikselu s obzirom na predikcije maski m:

$$\mathbf{s}_{\text{pixel}}(\mathbf{x})[r,c] = \sum_i \mathbf{m}_i[r,c] \cdot \mathbf{s}_{\text{mask}}(\mathbf{x}). \tag{8.2}$$

Takav način možemo interpretirati kao ansambliranje mjera anomalnosti predikcija na razini maski. U [17] za mjeru anomalnosti koriste maksimalnu vjerojatnost po klasama i na tu metodu se referenciraju s EAM (engl. *Ensemble over Anomaly scores of Mask-wide predictions*). Mi također pratimo EAM pristup i eksperimentiramo s raznim definicijama mjere anomalnosti.

Model Mask2Former naučen na skupu Cityscapes preuzeli smo iz službenog repozitorija originalnog članka [9]. Zbog pomaka domene u testnim skupovima [41] Mask2Former učimo na slikama iz skupova Cityscapes i Mapillary Vistas [35]. Učenje na negativnom skupu slika često se koristi u mnogim metodama za detekciju anomalija. U kontekstu guste detekcije anomalije, negativi se lijepe na unutardistribucijske slike. Metode koje uče na negativima zahtijevaju definiciju dodatnog člana gubitka u negativnim pikselima [2] [3] [16]. U našem slučaju, dodatan član gubitka nije potreban, nego oznaku u negativima postavljamo na *void*. Na taj način zahtijevamo da maske odbacuju negativne piksele. Takav pristup nije moguć na standardnim modelema za semantičku segmentaciju s predikcijama na razini piksela i koji se uče unakrsnom entropijom jer se gubitak ne računa u pikselima razreda *void*. Lijepimo instance iz ADE20k [45] i isječke iz unutardistribucijskih slika za učenje (Cityscapes i Vistas). U drugom slučaju smo i dalje u domeni metoda koje ne koriste negativni skup za učenje. Pretpostavljamo da će takvim načinom učenja Mask2Former generirati kvalitetnije maske. Lijepimo tri

instance negativa, odnostno tri slučajna isječka iz unutardistribucijskih slika i fino ugađamo model koji je prethodno naučen na skupovima Cityscapes i Vistas 2000 iteracija. Modele učene na Cityscapesu i Vistasu i modele učene uz negative su s nama podijelili autori [17].

Sa svim modelima radimo jedan unaprijedni prolaz sa svim slikama iz skupa za učenje i u datoteku spremamo značajke maski i odgovarajuće oznake. Na značajkama učimo 19 uvjetnih tokova 250 iteracija optimizatorom Adamax uz korak učenja 0.001. Veličina minigrupe je 64. Jedan tok se sastoji od dva sloja miješanja, kojima prethodni konvolucija 1×1. Dodatno koristimo Actnorm, ali samo nakon prvog sloja miješanja. Primijetili smo da se, zbog malog broja primjera za učenje, početni Actnorm sloj prenauči. Zato se pri evaluaciji događaju numeričke pogreške i latentna reprezentacija izvandistribucijskih primjera toga toka poprima inf i $-inf$ vrijednosti. Modul afinog miješanja izveden je kao jedan rezidualni blok.

U odnosu na slučaj kada tokove učimo na ugrađivanjima na razini piksela, sada imamo puno manje primjera za učenje. Iz jedne slike za učenje možemo izlučiti najviše jedno ugrađivanje maske po razredu. Dodatno treba napomenuti da nisu svi razredi zastupljeni u svakoj slici. Zbog toga možemo očekivati i veću varijancu u rezultatima. Metodu detekcije anomalija mjeroma anomalnosti $\mathrm{I}_{1\mathrm{g}}$ na razini maski testiramo na validacijskim skupovima FS Lost&Found, FS Static i RoadAnomaly. Učenje uvjetnih tokova ponavljamo 2 puta na svim skupovima podataka i prikazujemo srednju vrijednost i standardnu devijaciju. U tablicama 8.1, 8.2 i 8.3 prikazani su rezultati detekcije anomalija uvjetnim tokovima uz mjeru anomalnosti I_{1g} (6.9) na validacijskim skupovima FS Lost&Found, FS Static i RoadAnomaly. Reproduciramo EAM metodu na svim skupovima za učenje, a moguće različitosti u odnosu na članak [17] proizlaze iz različitog broja iteracija učenja modela.

Na značajkama modela koji je učen samo na Cityscapesu, naša metoda postiže prilično loše rezultate. Na slici 8.3 prikazan je rezultat detekcije anomalija za taj model. Možemo primijetiti da su kao anomalije označene neke unutardistribucijske klase: cesta, auto, nebo. Analizom maski koje predivđa Mask2Former možemo primijetiti kako su anomalije uhvaćene maskom za cestu.

U slučaju kada učimo uz stvarne negative dobivamo pogoršanje mjere FPR na svim testnim skupovima, a malo poboljšanje mjere AP jednino na FS Lost&Found. Možemo primijetiti da tom slučaju već imamo jako dobar diskriminativni detektor anomalija. U tenzoru $\mathbf{w}_{cls}$ imat ćemo jako slabe odzive za maske koje ne pripadaju niti jednom razredu pa množenje s $p(\mathbf{x}|Y)$ ne donosi promjenu. Jedino što se može dogoditi je da se množenjem s izglednostima predviđenima normalizirajućim tokom smanji sigurnost

Slika 8.3: Na slici dan je rezultat zadatka detekcije anomalija na primjerku slike iz skupa FS Lost&Found. Koristili smo model Mask2Former koji se učio samo na skupu Cityscapes. Redom je prikazana ulazna slika, njezina oznaka i detekcija anomalija gdje bijeli pikseli označuju anomalije.

predikcije u odnosu na diskriminativni slučaj što pretpostavljamo da se i događa u našem slučaju.

Zbog prethodni navedenih objašnjenja, fokusiramo se na skupove Cityscapes i Vistas i slučaj s učenjem na unutardistribucijskim isječcima, jer naša metoda u tim slučajevima može donijeti znatno poboljšanje. U nastavku se fokusiramo na te eksperimente. Na skupu FS Lost&Found poboljšavamo AP u oba dva slučaja i znatno smanjujemo FPR. AP također smanjujemo i na FS Static. Kada učimo na unutardistribucijskim isječcima FPR ne postiže poboljšanje. Rezultati na RoadAnomaly su nešto raznovrsniji. Ne postižemo poboljšanje u FPR, a poboljšanje u AP postižemo u slučaju kada učimo na unutardistribucijskim isječcima.

Tablica 8.1: Rezultati detekcije anomalija na razini maski na validacijskom skupu FS Lost&Found.

skup podataka	negativi	metoda	AP[%]	FPR_{95}[%]
Cityscapes	-	EAM	51.3	28.4
		I_{1g}	0.3 ± 0.2	90.01 ± 1.2
Cityscapes + Vistas	-	EAM	67.4	79.0
		I_{1g}	**69.6** $\pm$ 1.1	**39.0** $\pm$1.8
Cityscapes + Vistas	unutardistribucijski isječci	EAM	76.5	68.6
		I_{1g}	**77.3** $\pm$ 0.1	**6.4** $\pm$ 0.8
Cityscapes + Vistas	ADE	EAM	81.6	4.1
		I_{1g}	81.6 ± 0.2	13.7 ± 0.9

Tablica 8.2: Rezultati detekcije anomalija na razini maski na validacijskom skupu FS Static.

skup podataka	negativi	metoda	AP[%]	FPR_{95}[%]
Cityscapes	-	EAM	60.5	29.3
		I_{1g}	1.8 ± 0.3	55.1 ± 0.4
Cityscapes + Vistas	-	EAM	56.7	82.5
		I_{1g}	**69.2** $\pm$1.3	**15.3** $\pm$0.9
Cityscapes + Vistas	unutardistribucijski isječci	EAM	82.5	2.5
		I_{1g}	**83.4** $\pm$ 0.3	3.1 ± 0.3
Cityscapes + Vistas	ADE	EAM	86.5	2.7
		I_{1g}	81.3 ± 0.4	14.1 ± 1.1

Tablica 8.3: Rezultati detekcije anomalija na razini maski na skupu RoadAnomaly.

skup podataka	negativi	metoda	AP[%]	FPR_{95}[%]
Cityscapes	-	EAM	74.7	23.0
		I_{1g}	0.2 ± 0.5	88.9 ± 0.1
Cityscapes + Vistas	-	EAM	66.2	10.3
		I_{1g}	61.1 ± 1.4	12.4 ± 1.1
Cityscapes + Vistas	unutardistribucijski isječci	EAM	69.7	7.6
		I_{1g}	**70.9** $\pm$ 1.3	9.4 ± 0.9
Cityscapes + Vistas	ADE	EAM	67.9	7.9
		I_{1g}	67.1 ± 0.6	9.7 ± 0.9

Dodatno, na skupu FS Lost&Found dajemo usporedbu performansi svih mjera anomalnosti predloženih u 6.

Tablica 8.4: Usporedba različitih mjera anomalnosti u detekciji na razini maski na skupu FS Lost&Found.

	FS Lost&Found	
Mjera anomalnosti	AP [%]	FPR_{95} [%]
s_H	**78.3**	12.0
I_{1g}	<u>77.3</u>	**5.6**
I_1	0.2	97.1
s_{gd}	0.2	95.7
s_{dg}	76.8	<u>8.5</u>

Konačno, uspoređujemo naše rezultate s rezultatima iz literature. Treba napomenuti da usporedba nije u potpunosti pravedna zbog toga što rezultati iz literature koriste različite okosnice.

Tablica 8.5: Usporedba detekcije anomalija na razini piksela i maski mjerom I_{1g} s rezultatima iz literature koji ne uče na negativnom skupu.

	FS Lost&Found		FS Static		RoadAnomaly	
	AP [%]	FPR_{95} [%]	AP [%]	FPR_{95} [%]	AP [%]	FPR_{95} [%]
EAM [17]	76.5	68.6	82.5	2.5	69.7	7.6
GMMSeg [30]	43.47	13.11	-	-	34.42	47.9
SML [22]	27.61	15.46	-	-	-	-
$\mathbf{I_{1g}}$ per-mask Mask2Former	77.3	6.4	83.4	3.1	70.9	9.4
$\mathbf{I_{1g}}$ per-pixel DeepLabv3+	20.9	15.7	37.3	15.4	-	-

8.2. Detekcija anomalija na razini piksela

Predikcije na razini piksela kod modela Mask2Former možemo dobiti kao:

$$\mathbf{H}_{closed} = conv_{1\times1}(\mathbf{m}, \mathbf{w}_{cls}). \tag{8.3}$$

Važno je napomenuti da $\mathbf{H}_{closed}$ ne sadrži distribucije zato što $\sum_i m_i[r,c] \neq 1$ i $\sum_k \mathbf{w}_{cls}[i,k] \neq 1$. U našem slučaju je to problematično jer na izlazu modela ne možemo dobiti distribuciju $P(c|\mathbf{x})$ koju koristimo u predloženim mjerama anomalnosti

kao diskriminativni član. Preostaje nam renormalizirati $\mathbf{H}_{closed}$ ili $\mathbf{H}_{closed}$ koristiti takav kakav jeste uz napomenu da $\mathbf{H}_{closed} \neq P(c|\mathbf{x})$.

U tablici 8.6 su prikazani rezultati. Mask2Former se učio na skupovima Cityscapes i Vistas. Metode daju jako visok FPR_{95} na skupu FS Lost&Found. Na skupu FS Static bolje radi metoda na originalnim predikcijama po pikselima nego na renormaliziranim. Treba napomenuti da je FS Lost&Found kompliciraniji skup slika nego Static. Naša metoda primjetno bolje radi detekciju anomalija na razini maski nego na razini piksela za model Mask2Former.

Tablica 8.6: Rezultati detekcije anomalija uvjetnim tokovima na razini piksela modela Mask2Former. Koristimo $\text{I}_{1\text{g}}$ mjeru anomalnosti.

	FS Lost&Found			FS Static		
predikcije	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
originalni $\mathbf{H}_{closed}$	35.24	90.75	78.81	59.26	21.03	95.42
renormalizirani $\mathbf{H}_{closed}$	35.12	87.36	74.92	20.81	54.46	96.13

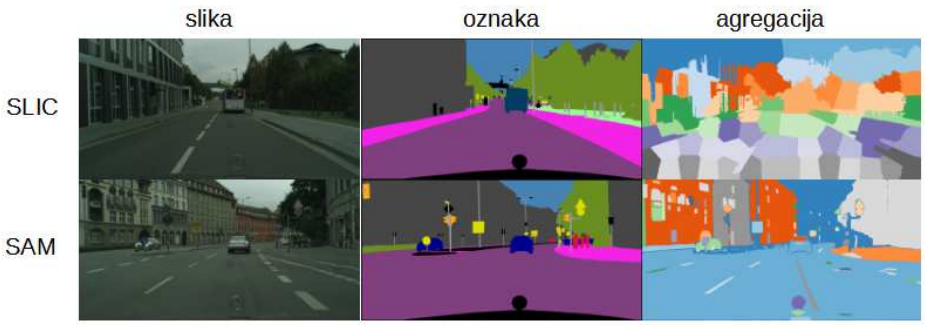
8.3. Učenje na agregiranim značajkama

Pretpostavljamo da učenje na agregiranim značajkama može donijeti poboljšanje. Predlažemo dvije metode agregacije značajki: SLIC [1] i SAM [25]. SLIC superpikseli (engl. *Simple Linear Iterative Clustering*) [1] je starija iterativna metoda koja se koristila za segmentaciju slike. Pikseli slike se iterativno grupiraju u 5-dimenzionalnom prostoru boje i prostornih dimenzija slike. Naša pretpostavka je da bi ova metoda mogla raditi i biti roubsna na anomalije jer se ne uči. SAM (engl. *Segment Anything Model*) je recentni pristup segmentaciji slike gdje se prednaučenim modelom uspješno mogu segmentirati objekti. Na slici 8.4 su prikazani primjeri segmentacije. Vidimo da SAM agregira bolje od SLIC-a što je i očekivano jer je SAM naprednija metoda koja se uči i koristi veliku okosnicu. Također, primjećujemo da su mnogi SLIC segmenti na granicama razreda, a takve agregacije bi mogle biti problematične za učenje razredom uvjetovanih normalizirajućih tokova zbog uvođenja dodatnog šuma u značajke. Odlučujemo koristiti SAM agregacije. Značajke agregiramo usrednjavanjem u po segmentima. Iako SAM dobro segmentira ne možemo isključiti situaciju gdje će jedan SAM segment obuhvatiti više oznaka. Zbog toga oznaku segmenta agregiramo s obzirom na najčešće pojavljivanje.

Metodom SAM agregiramo značajke modela DeepLabv3+ i na takvim značajkama učimo uvjetne tokove. Za mjeru anomalnosti koristimo I_{1g}. Rezultati su dani u tablici 8.7. Iznenađujuće je da učenje na agregacijama daje lošije rezultate. Pretpostavljamo da se to događa zbog toga što značajke modela DeepLabv3+ nisu naučene da dobro agregiraju.

Model Mask2Former uči značajke agregirati u maske. Ponavljamo prethodno opisani eksperiment na ugrađivanjima na razini piksela modela Mask2Former. Očekivano dobivamo bolje rezultate nego u slučaju modela DeepLabv3+, ali i dalje ne dobivamo poboljšanje u odnosu na učimo na originalnim značajkama.

Na slici 8.4 možemo primijetiti da SAM agregira objekte, što mu je i svrha, ali nije ono što trebamo u ovom slučaju. Sam će primjerice cijelu cestu agregirati u jednu značajku. Mislimo da je obećavajući pristup za budući rad agregirati značajke po metodi SLIC, ali u obzir uzeti i oznake slike. Na taj način mogli bismo agregirati lokalna susjedstva i ne uvoditi dodatan šum na granicama klasa.



Slika 8.4: Slika, oznaka i rezultat segmentacije slike metodama SLIC i SAM.

Tablica 8.7: Rezultati detekcije anomalija uvjetnim tokovima naučenim na originalnim značajkama i agregiranim značajkama.

		FS Lost&Found			FS Static		
Model	SAM	AP [%]	FPR_{95} [%]	AUROC [%]	AP [%]	FPR_{95} [%]	AUROC [%]
DeepLabv3+	✗	10.65	26.82	89.43	33.35	14.01	92.17
DeepLabv3+	✓	4.00	32.35	89.21	20.86	15.19	91.3
Mask2Former	✗	35.24	78.81	90.75	59.26	21.03	95.42
Mask2Former	✓	26.73	77.56	90.63	63.02	16.82	95.79

8.4. Pogreške zbog loših predikcija maski

Ako maska $\mathbf{m_i}$ ulovi anomaliju, tj. ako su odzivi maske u području gdje je anomalija visoki, ne možemo se oporaviti od takve pogreške. Zbog toga želimo saznati koliko i kako griješimo zbog loših maski.

Definiramo mjeru anomalnosti koja će anomalijama proglasiti piksele koji nisu niti u jednoj maski:

$$\mathbf{s}(\mathbf{x})[r, c] = -\max_i \mathbf{m}_i[r, c]. \tag{8.4}$$

Rezultati na FS Static su prikazani u tablici. Dodatno, na slici 8.5 su prikazani histogrami relativnih udjela unutardistribucijskih i izvandistribucijskih piksela s obzirom na mjeru anomalnosti 8.4. Odabiremo FS Static jer smo primijetili da je upravo na tom skupu većina anomalija uhvaćena maskama. Taj problem možemo riješiti tako da u

obzir uzmemo i nesigurnosti predikcije maske, primjerice EAM pristupom. Veći problem nastaje kada maske koje pripadaju untardistribucijskom razredu ulove dijelove anomalija.

Tablica 8.8: Rezultati detekcije anomalija mjerom anomalnosti maksimalne maske na validacijskom skupu FS Static.

skup podataka	AP	FPR	AUROC
Cityscapes	2.63	99.76	54.87
Cityscapes + Vistas	3.38	99.9	61.52
Cityscapes + Vistas + inlier crop	11.28	98.92	80.14
Cityscapes + Vistas + ADE	5.89	95.43	74.56

Definiramo metodu proroka (engl. *oracle*) koja ima točnu informaciju o anomalnim pikselima iz testnog skupa. Točnije definiramo dvije metode:

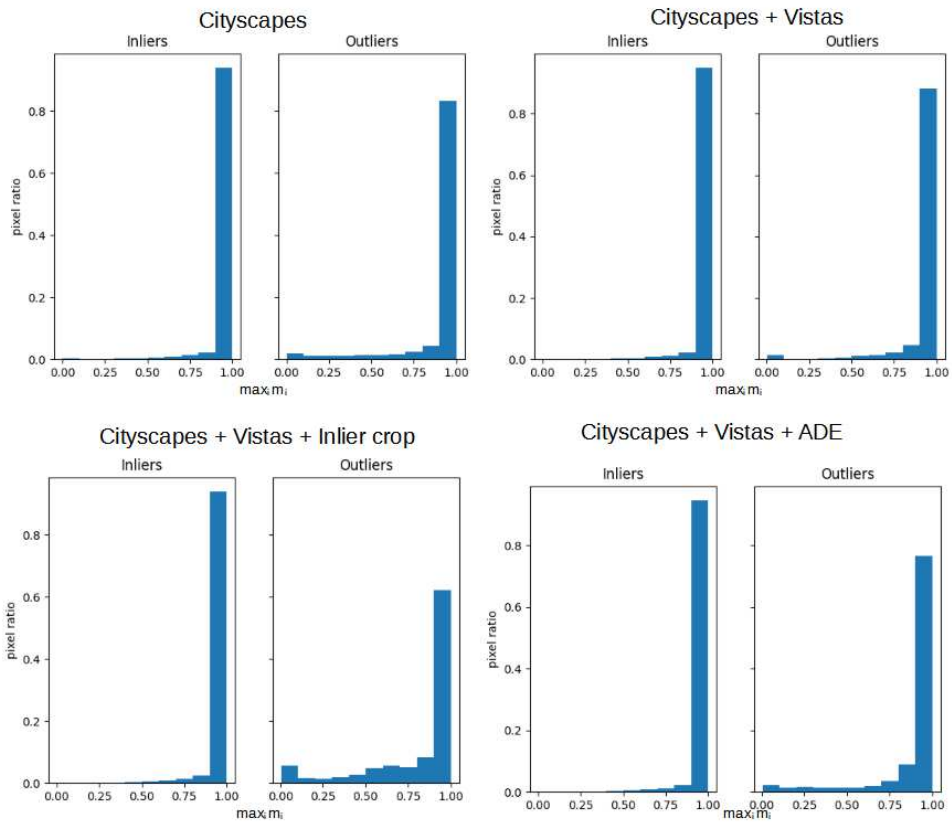
– binarno: maska ima mjeru anomalnosti 1 ako ima barem jedan anomalni piksel

– udio: mjera anomalnosti maske je udio anomalnih piksela.

Rezultati su prikazani u tablici 8.9. Iz rezultata možemo zaključiti da maske ulove anomalije, što smo već znali. Mali AP sugerira i da postoje anomalije koje niti jedna maska ne prepozna, a te anomalije ovim pristupom ne možemo prepoznati. Pretpostavljamo da je upravo to što postoje anomalije koje niti jedna maska ne prepozna bitan element EAM pristupa i razlog dobrih rezultata.

Tablica 8.9: Rezultati procjena metode proroka na razini maski na skupu FS Lost&Found.

skup podataka	metoda proroka	AP	FPR	AUROC
Cityscapes + Vistas	binarno	2.05	56.50	76.27
Cityscapes + Vistas	udio	15.5	85.36	54.82
Cityscapes + Vistas + inlier crop	binarno	1.87	48.25	67.46
Cityscapes + Vistas + inlier crop	udio	3.18	59.15	76.00
Cityscapes + Vistas + ADE	binarno	3.03	43.13	73.42
Cityscapes + Vistas + ADE	udio	10.09	39.09	88.61



Slika 8.5: Slika prikazuje udjele unutardistribucijskih i izvandistribucijskih piksela predviđenih metodom maksimalne maske.

9. Generativna klasifikacija uvjetnim normalizirajućim tokovima

Zadatak semantičke segmentacije svodi se na zadatak guste diskriminativne klasifikacije. Gusti diskriminativni klasifikator na izlazu vraća aposteriornu distribuciju vjerojatnosti razreda $P(Y|\mathbf{z})$ gdje $\mathbf{z}$ označava značajke po pikselima izlučene dubokim modelom:

$$\hat{y}_{\mathrm{d}} = \arg\max_c p(c|\mathbf{x}). \tag{9.1}$$

Diskriminativni klasifikator uči decizijsku granicu između razreda pritom zanemarujući pozadinsku distribuciju podataka.

Razredom uvjetovane tokove možemo primijeniti kao generativni klasifikator. Svaki uvjetni tok modelira distribuciju $p(\mathbf{z}|c)$. Za svaku značajku $\mathbf{z}$ iz slike napravimo unaprijedni prolaz kroz C normalizirajućih tokova. Na taj način svaki će tok dati predikciju o izglednosti da primjer pripada njegovoj klasi. Na izlazu dobivamo matricu dimenzija $N \times C$, gdje je N broj značajki, a C broj razreda. Vodeći se pretpostavkom da će primjeri koji pripadaju razredu c imati veću izglednost od primjera koji ne pripadaju razredu c, generativnu predikciju razreda možemo dobiti kao:

$$\hat{y}_{\mathrm{g}} = \arg\max_c p(\mathbf{z}|c). \tag{9.2}$$

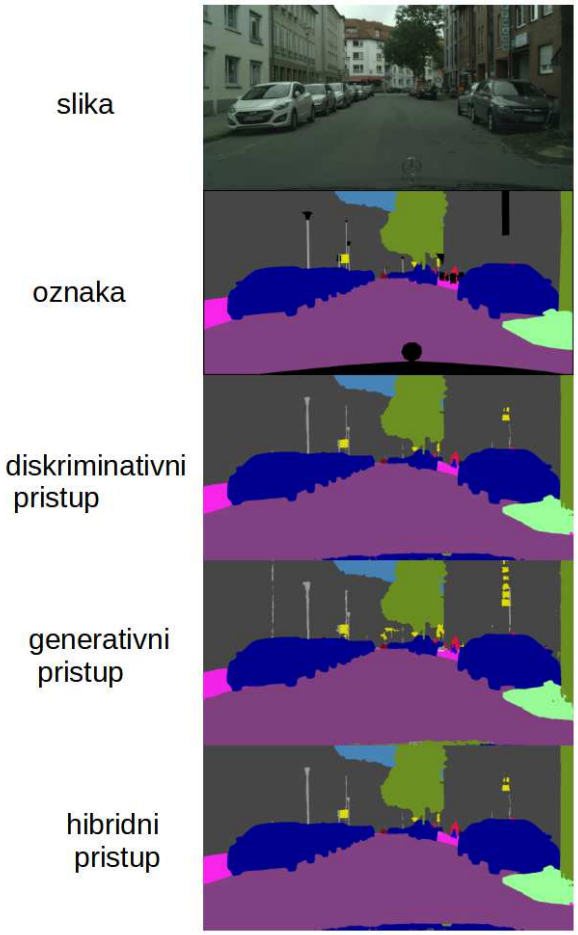
Alternativno, možemo definirati i hibridnu predikciju:

$$\hat{y}_{\mathrm{h}} = \arg\max_c p_c(\mathbf{z}|\mathbf{x}) P(c|\mathbf{x}). \tag{9.3}$$

Mjerimo performanse predloženih gustih klasifikatora na validacijskom skupu Cityscapesa. Diskriminativni model DeepLabv3+ model kombiniramo s C normalizirajućih tokova koji su učeni standardno (3.11) i kontrastivno (3.12). Normalizirajući tokovi sastoje se od dva afina sloja miješanja s fiksnom alternirajućom permutacijom. Rezultati su prikazani u tablici 9.1. Na slici 9.1 dani su kvalitativni rezultati. Generativni pristup daje iznenađujuće dobre rezultate. Hibridnim pristupom uspijevamo poboljšati diskriminativne predikcije.

Tablica 9.1: Semantička segmentacija diskriminativnim, generativnim i hibridnim pristupom.

Metoda		učenje tokova	mIoU[%]
diskiminativni klasifikator	$\arg\max_c P(c\|\mathbf{x})$	-	81.57
generativni klasifikator	$\arg\max_c p_c(\mathbf{x}\|c)$	standardno	66.77
		kontrastivno	73.97
hibridni klasifikator	$\arg\max_c p_c(\mathbf{z}\|\mathbf{x})P(c\|\mathbf{x})$	standardno	81.87
		kontrastivno	**81.93**



Slika 9.1: Usporedba diskriminativne, generativne i hibridne semantičke segmentacije.

10. Zaključak

U ovom radu istraživali smo problem guste detekcije anomalija. Nakon pregleda literature i trenutnog stanja tehnike, predlažemo metodu za detekciju anomalija generativnim modelima. Pozadinsku distribuciju značajki dubokog modela po razredima $p(\mathbf{x}|c)$ modeliramo uvjetnim normalizirajućim tokovima. Detekciju anomalija u predikcijama dubokih modela analiziramo na dvije razine, na razini piksela i na razini maski. Metode smo evaluirali na javno dostupnim skupovima slika Fishyscapes Lost&Found, Fishyscapes Static i RoadAnomaly. Kao glavni doprinos ovoga rada ističemo hibridnu mjeru anomalnosti $\mathrm{I}_{1\mathrm{g}}$ koja se temelji na informacijskom dobitku. Mjera $\mathrm{I}_{1\mathrm{g}}$ postiže dobre rezultate u detekciji anomalija u predikcijama modela DeepLabv3+. Zanimljive rezultate postižemo u detekciji anomalija na razini maski modela Mask2Former. Ugađanjem modela Mask2Former na unutardistribucijskim slikama poboljšavamo diskriminativne predikcije i značajke za učenje generativnog modela. Mjerom $\mathrm{I}_{1\mathrm{g}}$ postižemo poboljšanje u AP i značajno poboljšanje FPR_{95} na skupovima Fishyscapes Lost&Found i Fishyscapes Static.

Eksperimenti provedeni u sklopu ovoga rada otvaraju mnoge smjerove za daljnje istraživanje. Predlažemo učiti segmentacijski model i uvjetne tokove s kraja na kraj. Time bismo spriječili problem kolapsa značajki i očekujemo poboljšanje trenutnih rezultata. Zanimljiva ideja je učiti Mask2Former model s negativima uz K+2 klase na način da se neuparene maske klasificiraju u klasu K+2, a negativi u klasu K+1. Pretpostavljamo da bi u tom slučaju tok K+1 mogao prepoznavati anomalne klase. Učenje na agregiranim značajkama nije nam donijelo očekivano poboljšanje. Ideja je pronaći bolju metodu agregacije značajki.

LITERATURA

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, i Sabine Süsstrunk. Slic superpixels. Technical report, 2010.

[2] Petra Bevandić, Ivan Krešo, Marin Oršić, i Siniša Šegvić. Simultaneous semantic segmentation and outlier detection in presence of domain shift. U *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41*, stranice 33–47. Springer, 2019.

[3] Petra Bevandić, Ivan Krešo, Marin Oršić, i Siniša Šegvić. Dense outlier detection and open-set recognition based on training with noisy negative images. *arXiv preprint arXiv:2101.09193*, 2021.

[4] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, i Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *International Journal of Computer Vision*, 129:3119–3135, 2021.

[5] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, i Sergey Zagoruyko. End-to-end object detection with transformers. U *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, stranice 213–229. Springer, 2020.

[6] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, i Matthias Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. *arXiv preprint arXiv:2104.14812*, 2021.

[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, i Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. U *Proceedings of the European conference on computer vision (ECCV)*, stranice 801–818, 2018.

[8] Bowen Cheng, Alex Schwing, i Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *Advances in Neural Information Processing Systems*, 34:17864–17875, 2021.

[9] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, i Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. U *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, stranice 1290–1299, 2022.

[10] 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 understanding. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 3213–3223, 2016.

[11] Michael R DeWeese i Markus Meister. How to measure the information gained from one symbol. *Network: Computation in Neural Systems*, 10(4):325, nov 1999. doi: 10.1088/0954-898X/10/4/303. URL `https://dx.doi.org/10.1088/0954-898X/10/4/303`.

[12] Laurent Dinh, Jascha Sohl-Dickstein, i Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[13] Matej Grcić i Siniša Šegvić. Hybrid open-set segmentation with synthetic negative data. *arXiv preprint arXiv:2301.08555*, 2023.

[14] Matej Grcić, Petra Bevandić, Zoran Kalafatić, i Siniša Šegvić. Dense anomaly detection by robust learning on synthetic negative data. *arXiv preprint arXiv:2112.12833*, 2021.

[15] Matej Grcić, Ivan Grubišić, i Siniša Šegvić. Densely connected normalizing flows. *Advances in Neural Information Processing Systems*, 34:23968–23982, 2021.

[16] Matej Grcić, Petra Bevandić, i Siniša Šegvić. Densehybrid: Hybrid anomaly detection for dense open-set recognition. U *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXV*, stranice 500–517. Springer, 2022.

[17] Matej Grcić, Josip Šarić, i Siniša Šegvić. On advantages of mask-level recognition for open-set segmentation in the wild. *arXiv preprint arXiv:2301.03407*, 2023.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollár, i Ross Girshick. Mask r-cnn. U *Proceedings of the IEEE international conference on computer vision*, stranice 2961–2969, 2017.

[19] Dan Hendrycks i Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.

[20] Dan Hendrycks, Mantas Mazeika, i Thomas Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018.

[21] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, i Dawn Song. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*, 2019.

[22] Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, i Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. U *Proceedings of the IEEE/CVF International Conference on Computer Vision*, stranice 15425–15434, 2021.

[23] Durk P Kingma i Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018.

[24] Polina Kirichenko, Pavel Izmailov, i Andrew G Wilson. Why normalizing flows fail to detect out-of-distribution data. *Advances in neural information processing systems*, 33:20578–20589, 2020.

[25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[26] Andreas Kirsch i Yarin Gal. A practical & unified notation for information-theoretic quantities in ml. *arXiv preprint arXiv:2106.12062*, 2021.

[27] Ivan Kobyzev, Simon Prince, i Marcus Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[28] Nishant Kumar, Siniša Šegvić, Abouzar Eslami, i Stefan Gumhold. Normalizing flow based feature synthesis for outlier-aware object detection. U *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, stranice 5156–5165, 2023.

[29] Kimin Lee, Honglak Lee, Kibok Lee, i Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. *arXiv preprint arXiv:1711.09325*, 2017.

[30] Chen Liang, Wenguan Wang, Jiaxu Miao, i Yi Yang. Gmmseg: Gaussian mixture based generative semantic segmentation models. *arXiv preprint arXiv:2210.02025*, 2022.

[31] Krzysztof Lis, Krishna Nakka, Pascal Fua, i Mathieu Salzmann. Detecting the unexpected via image resynthesis. U *Proceedings of the IEEE/CVF International Conference on Computer Vision*, stranice 2152–2161, 2019.

[32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, i Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. U *Proceedings of the IEEE/CVF international conference on computer vision*, stranice 10012–10022, 2021.

[33] Jonathan Long, Evan Shelhamer, i Trevor Darrell. Fully convolutional networks for semantic segmentation. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 3431–3440, 2015.

[34] Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, i Balaji Lakshminarayanan. Do deep generative models know what they don't know? *arXiv preprint arXiv:1810.09136*, 2018.

[35] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, i Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. U *Proceedings of the IEEE international conference on computer vision*, stranice 4990–4999, 2017.

[36] Jie Ren, Peter J Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark Depristo, Joshua Dillon, i Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. *Advances in neural information processing systems*, 32, 2019.

[37] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

[38] Jan Snajder. Predavanja iz kolegija Strojno učenje 1, 21. Vrednovanje modela, UniZG, FER. URL https://www.fer.unizg.hr/_download/repository/SU1-2022-P21-VrednovanjeModela.pdf.

[39] Lucas Theis, Aäron van den Oord, i Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, i Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[41] Sagar Vaze, Kai Han, Andrea Vedaldi, i Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need? *arXiv preprint arXiv:2110.06207*, 2021.

[42] Sergey Zagoruyko i Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[43] Hongjie Zhang, Ang Li, Jie Guo, i Yanwen Guo. Hybrid models for open set recognition. U *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, stranice 102–117. Springer, 2020.

[44] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, i Antonio Torralba. Scene parsing through ade20k dataset. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 633–641, 2017.

[45] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, i Antonio Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127:302–321, 2019.

Sažetak

Standardni diskriminativni modeli za semantičku segmentaciju osjetljivi su na pojavu anomalija u ispitnim slikama. Navedeni problem rješavamo detekcijom anomalija razredom uvjetovanim normalizirajućim tokovima. Predlažemo mjere anomalnosti temeljene na teoriji informacije s probabilističkom interpretacijom te ih primjenjujemo na predikcije na razini piksela i maski. Predložena hibridna metoda postiže kompetitivne rezultate na standardnim testnim skupovima slika i poboljšanje u odnosu na usporedive diskriminativne metode za gustu detekciju anomalija.

Ključne riječi: Semantička segmentacija, gusta detekcija anomalija, uvjetni normalizirajući tok, zajednička informacija

Dense anomaly detection with generative models

Abstract

Standard discriminative models for semantic segmentation are sensitive to the presence of anomalies in test images. We address this issue by detecting anomalies using class-conditional normalizing flows. We define anomaly scores based on information theory with probabilistic interpretation and apply them to pixel-level and mask-level predictions. The proposed hybrid method achieves competitive results on standard benchmarks and improvement upon comparable discriminative dense anomaly detection methods.

Keywords: Semantic segmentation, dense anomaly detection, class-conditional normalizing flows, mutual information.