

GENERIRANJE SINTETIČNIH SLIK IZ 3D MODELOV PREDMETOV ZA UČENJE SEGMENTACIJSKIH MODELOV V INDUSTRIJSKEM OKOLJU

MATIC FUČKA, MARKO RUS, JAKOB BOŽIČ,

DANIJEL SKOČAJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
mf4782@student.uni-lj.si, marko.rus@fri.uni-lj.si, jakob.bozic@fri.uni-lj.si,
daniel.skocaj@fri.uni-lj.si

Sinopsis Ena večjih preprek uporabe globokega učenja v industriji je zahteva po velikem številu označenih slik potrebnih za nadzorovano učenje. Ker je pridobivanje označenih slik pogosto zamuden in drag proces, se veliko metod skuša izogniti temu problemu z bogatenjem učnih slik, ki so na voljo, ali z uporabo sintetičnih slik za predhodno učenje in poznejše doučenje z resničnimi slikami. V tem članku predlagamo metodo za generiranje sintetičnih slik iz danih 3D CAD modelov in novo podatkovno množico sestavljeno iz sintetičnih in resničnih slik ter njihovih segmentacijskih slik. Obravnavamo torej scenarij učenja segmentacijskega modela zgolj iz sintetičnih slik, ki so v industriji tipično na voljo, in zagotavlja najhitrejšo prilagoditev proizvodnega procesa na nove izdelke brez potrebe po zajemanju pravih učnih slik. Na predlagani podatkovni množici ovrednotimo algoritem za segmentacijo slik in primerjamo rezultate dobljene z različnim številom sintetičnih in realnih slik industrijskega predmeta posnetega oz. upodobljenega na različnih ozadjih.

Ključne besede:

strojni vid,
globoko učenje,
segmentacija,
sintetične slike,
CAD model

3D-MODEL-BASED RENDERING OF SYNTHETIC IMAGES FOR TRAINING SEGMENTATION MODELS IN AN INDUSTRIAL ENVIRONMENT

MATIC FUČKA, MARKO RUS, JAKOB BOŽIČ,
DANIJEL SKOČAJ

Univerza v Ljubljani, Faculty of Computer and Information Science, Ljubljana, Slovenia
mf4782@student.uni-lj.si, marko.rus@fri.uni-lj.si, jakob.bozic@fri.uni-lj.si,
danijel.skocaj@fri.uni-lj.si

Abstract One of the major obstacles to the application of deep learning in industry is the requirement for a large number of labeled images required for supervised learning. This is because obtaining labeled images can be time-consuming and costly. To overcome this challenge, some methods utilize image augmentation or synthetic images for pre-training, followed by fine-tuning with real images. This paper introduces a method for generating synthetic images from 3D CAD models, along with a new dataset consisting of both synthetic and real images, and their corresponding segmentation masks. The aim is to train a segmentation model using only synthetic images, which are readily available in industry, allowing for a quicker adaptation of the production process to new products without the need for capturing real training images. We evaluate an image segmentation algorithm on the proposed dataset and compare the results obtained with a different number of synthetic and real images of an industrial object captured or rendered on different backgrounds.

Keywords:
machine vision,
deep learning,
segmentation,
synthetic images,
CAD model

1 Uvod

Področje računalniškega vida je na krilih globokega učenja v zadnjem desetletju doseglo ogromen napredek. Rešitve, temelječe na globokih nevronskih mrežah, vse bolj prodirajo tudi v sisteme strojnega vida, ki so nepogrešljiv del modernih proizvodnih linij v paradigmi Industrija 4.0. Omogočajo namreč veliko bolj raznovrstno in robustno uporabo in rešitev veliko bolj kompleksnih problemov kot do zdaj prevladujoče klasične rešitve strojnega vida.

Ima pa tak, na podatkih in strojnem učenju temelječ pristop, tudi veliko pomanjkljivost: za delovanje namreč zahteva veliko število učnih slik. Še več, trenutno prevladujoč pristop v zrelih aplikacijah strojnega vida temelji na nadzorovanem učenju, ki zahteva, da so učne slike tudi ustrezno označene. Poleg zajemanja slik je torej potrebo tudi veliko človeškega dela za njihovo označevanje. To pa proces uvajanja tovrstnih sistemov v industrijsko proizvodnjo in prilagajanje na nove izdelke podaljša in podraži.

V tem članku obravnavamo eno izmed možnosti za obravnavanje tega problema – umetno generiranje sintetičnih učnih slik z uporabo 3D CAD modelov. V industrijski proizvodnji so 3D modeli izdelkov tipično na voljo, tudi samo okolje je običajno dokaj poznano. To predstavlja zelo ugodne pogoje za generiranje sintetičnih slik, katerih pojavnost ne odstopa bistveno od pojavnosti pravih slik posnetih v pravem okolju. Je pa tako generiranje sintetičnih slik, ob ustrezno razviti programski opremi, hitro in praktično zastonj. V kolikor bi bile zgenerirane učne slike dovolj realistične, bi torej tak pristop pomenil idealen scenarij za gradnjo globokih modelov za reševanje nalog strojnega vida v industrijskem okolju.

V nadaljevanju bomo predstavili cevovod za sintetično generiranje slik ter tako zgenerirano učno množico slik. Osredotočili se bomo na konkreten problem segmentacije predmeta oz. ločevanja pikslov, ki pripadajo ospredju (tj. predmetu), od tistih, ki pripadajo ozadju. To je namreč običajno pogosto prvi problem, ki ga mora sistem strojnega vida rešiti. Analizirali bomo rezultate, dobljene z učenjem na sintetičnih slikah in jih primerjali z rezultati, dobljenimi z uporabo pravih učnih slik. Eksperimentalno bomo torej raziskali uporabnost zgeneriranih sintetičnih učnih slik za namen segmentacije predmetov v realnem industrijskem okolju.

2 Sorodna dela

Uporaba sintetičnih podatkov za namene globokega učenja je vedno bolj priljubljena. To še posebej velja za področje računalniškega vida. V preteklih nekaj letih je bilo ustvarjeno mnogo podatkovnih množic za reševanje različnih problemov kot so strojno štetje (Wang et al., 2019), za namene avtonomne vožnje (Richter et al., 2016; Li et al., 2019; Huang et al., 2018; Yogamani et al., 2019; Zhan et al., 2019) ter razpoznavo raznih detajlov samih slik (McCormac et al., 2017; Mayer et al., 2016). V zadnjem letu ali dveh se je pričelo generirati veliko sintetičnih slik ustvarjenih z raznimi generativnimi modeli kot so npr. generativne nasprotniške mreže (ang. generative adversarial networks) (Dewi et al., 2022) in difuzijski modeli (Pinaya et al., 2022; Akrouf et al., 2023) v nasprotju z obdobjem pred tem, ko se je večino sintetičnih slik ustvarilo s programi za grafično upodabljanje. Večina metod (Li et al., 2019) uporablja sintetične slike kot podaljšek učne množice, ki je sestavljena samo iz resničnih slik, nekaj pa jih zgradi celotno učno množico iz sintetičnih slik (Mayershofer et al., 2021; Mayer et al., 2016) in nato naredi premik domene z uporabo resničnih slik za doučenje (Wang et al., 2019; Richter et al., 2016) nevronske mreže.

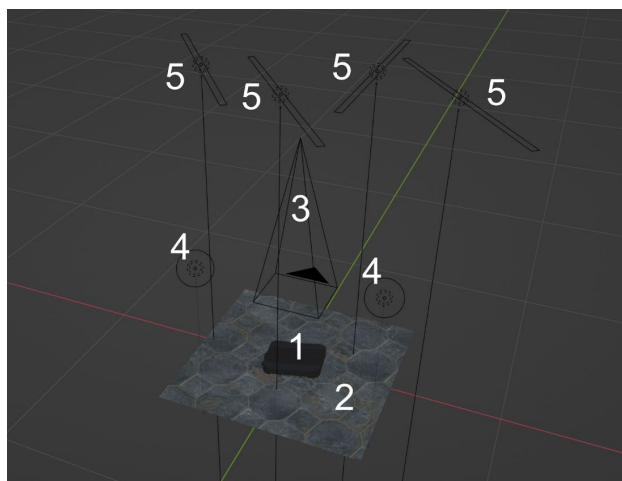
V industrijski domeni (Mayershofer et al., 2021; Eversberg et al., 2021; Abou et al., 2022; Quattrocchi et al., 2022) pogosto pride do uporabe sintetičnih slik, saj je pogosto zajemanje in označevanje velike količine slik drago in zamudno. Velika večina pristopov ustvari sintetične slike na podoben način. Podan imajo 3D CAD model in nato s pomočjo programov za grafično upodabljanje, kot sta Blender in NVidia Omniverse, ustvarijo sintetične slike. Te kasneje uporabijo za učenje ter imajo mogoče označeno še manjšo množico resničnih slik, ki jih uporabijo, da doučijo nevronske mreže. Podoben način obravnavamo tudi v tem delu in analiziramo rezultate na problemu segmentacije slik.

3 Generiranje sintetičnih slik

Sintetično množico za učenje modelov globokega učenja ustvarimo z uporabo namenskih računalniških okolij za sintetično upodabljanje slik. Glavna prednost takšnega pristopa je možnost popolne avtomatizacije celotnega postopka upodabljanja, kar pomeni, da lahko brez človeškega posega upodobimo poljubno število sintetičnih slik skupaj s pripadajočimi segmentacijskimi slikami (ang.

segmentation masks). Kljub sintetični naravi postopka lahko poskrbimo za visoko pestrost končnih slik, menjujemo lahko izgled predmeta in lego predmeta, parametre kamere, podlago, na kateri stoji predmet, ozadje in svetila. Poleg tega pa imajo na tak način pridobljene segmentacijske slike popolno točnost, kar pa ne velja nujno za ročno označene slike, pri katerih se lahko zaradi človeške napake zgodi, da določen slikovni element ne pripada označenemu objektu.

Eno od orodij, ki se uporablja za sintetično ustvarjanje slik, je Blender (Community et al., 2018), ki je brezplačen in odprtokodni program za animacijo, modeliranje in fotorealistično upodabljanje 3D okolja. Skupaj z njim uporabljamo tudi programski pripomoček BlenderProc (Denninger et al., 2019), ki poskrbi za avtomatizacijo celotnega postopka upodabljanja znotraj programskega jezika Python. Z njim lahko programsko nastavimo postavitev ciljnega objekta in kamere, določimo podlago in osvetlitev scene ter nastavimo parametre ustvarjanja slike, ki se nato izvede v programu Blender.



Slika 1: Elementi scene znotraj programskega okolja Blender pri ustvarjanju sintetično ustvarjene podatkovne množice. Scena je sestavljena iz sledečih elementov: 1 - ciljni objekt, 2 - podlaga, 3 - kamera, 4 - okrogli luči in 5 - paličaste luči.

Vir: lasten.

V nadaljevanju bomo podrobno opisali postopek tvorjenja predlagane sintetično ustvarjene podatkovne množice, ki vsebuje 6160 parov sintetično ustvarjenih slik in segmentacij. Od tega jih ima 2160 ozadja različna od teh, ki smo jih imeli pri zajemu

resničnih slik. Na začetku postopka upodabljanja sintetične slike se najprej določi parametre objektov na sceni. Na Sliki 1 lahko vidimo vse elemente, ki sestavljajo sceno, in njihove postavitve: ciljni objekt, podlaga, kamera in različna svetila. Ciljni objekt je postavljen na sredino podlage tako, da je njegova širina v virtualnem okolju približno 0,3 m. Nato se določijo parametri elementov, ki se izberejo naključno iz vnaprej določene domene. Določi se rotacijo ciljnega objekta v katerokoli smer iz ravnine podlage in izbere naključno podlago iz množice javno dostopnih tekstur⁷. Osvetlitev je ustvarjena s pomočjo dveh tipov luči, okroglih in podolgovatih. V polovici primerov se izbere okrogli tip luči, ustvari se dvojje luči na nizki višini 0,8 m, tako da se pojavi visoka verjetnost izrazitejših senc. Polmer luči se nastavi od 0 do 0,1 m, svetilnost pa med 40 do 70 lumnov. V drugi polovici primerov se izbere od ene do štiri podolgovate luči, ki so na višini 1,5 m, vsaka ima svetilnost od 5 do 10 lumnov, postavljene so pa v naključno lokacijo in smer. Ta dva načina sta bila izbrana namensko, da bi povečala pestrost osvetlitve, eden poustvarja pestrost senc, drugi pa odsevov, saj je verjetnost neposrednih odbojev svetlobe iz svetila v senzor zaradi dolžine luči povečana. Kamera je postavljena na višino 1,2 m in ima po širini vidno polje 22,6 stopinj in ločljivost 1824×1536 pikselov.



Slika 2: Šest primerov sintetično ustvarjenih slik. Največjo pestrost predstavljata ozadje in rotacija ciljnega objekta, slike se pa razlikujejo tudi v postavitvi svetil in posledično različnih sencah.

Vir: lasten.

⁷ Dostopno na ambientCG.com (licenca CC0).

Slike, ustvarjene z opisanim postopkom, predstavljajo velik nabor možnih osvetlitev in leg predmeta ter podlag, na katerih predmet leži. Raznolikost končnih slik se lahko vidi na Sliki 2, kjer je prikazano nekaj naključnih primerov z opisanim postopkom sintetično ustvarjenih slik. Na Sliki 3 lahko vidimo par sintetično ustvarjene slike in pripadajoče segmentacijske slike, ki se popolnoma prilega predmetu na sliki.



Slika 3: Primer para sintetično ustvarjene slike in segmentacije.

Vir: lasten.

4 Zasnova eksperimenta

V tem razdelku bomo predstavili zasnovo eksperimenta s katerim smo ovrednotili kakovost zgeneriranih učnih slik za namen segmentacije. V ta namen smo posneli množico pravih slik predmeta, s katerim smo preverili delovanje naučenega modela. Zanimala nas je odvisnost uspešnosti zgrajenega modela od števila učnih slik ter tudi od raznovrstnosti ozadja uporabljenega v fazi učenja, pa tudi v fazi vrednotenja.

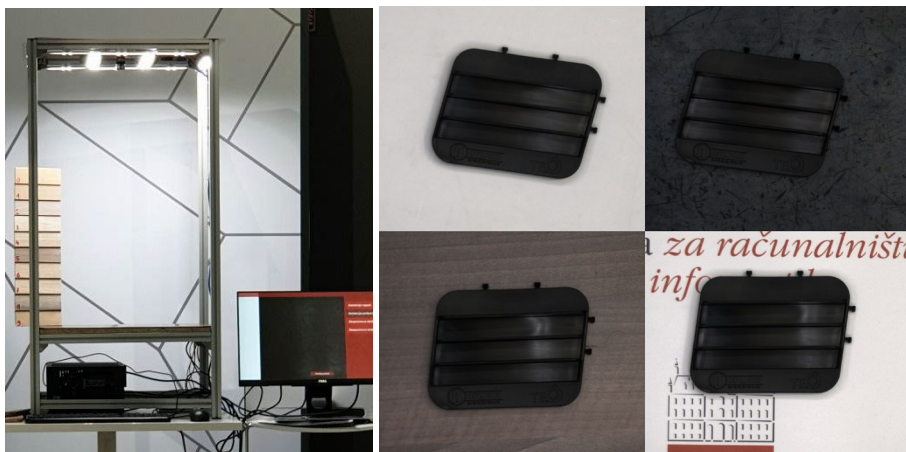
4.1 Zajem resničnih slik

Resnične slike smo zajeli s pomočjo ViCoS Demo celice, ki jo je mogoče videti na Sliki 4. Predmeti na slikah so bili zajeti na štirih različnih ozadjih (belo, črno, leseno in plakat) pri petih različnih osvetlitvah in desetih različnih postavitev predmeta. Belo ozadje je bilo izbrano kot najenostavnejši scenarij na katerem je temen predmet lepo viden. Črno ozadje tako predstavlja zelo težak scenarij saj je kontrast med ozadjem in predmetom (ter njegovimi sencami) zelo majhen. Leseno ozadje predstavlja dokaj realističen scenarij, medtem ko plakat vsebuje bolj nehomogeno

ozadje z nekaj izrazitimi elementi (črke, kvadrat), ki bi utegnili pri segmentaciji povzročati probleme. Nekaj primerov zajetih slik je vidnih na Sliki 4.

Večjo variabilnost scen smo dosegli s spreminjanjem osvetlitve. Na ta način so se na predmetih pojavljali odbleski, nehomogena osvetlitev in tudi sence, ki tudi sicer povzročajo probleme v realnem okolju. Čas ekspozicije smo avtomatsko nastavljali na način, da je bila svetlost prazne scene (slike kalibracijskega lista) podobna ne glede na osvetlitev. Tako smo zagotovili ponovljivost eksperimenta in dovolj velik kontrast na slikah.

Skupaj smo tako posneli 200 pravih slik (10 scen pri 5 osvetlitvah za vsako izmed 4 ozadij). Vse slike smo ročno segmentirali, tako da smo na njih označili piksele, ki pripadajo predmetu.



Slika 4: ViCoS Demo celica in primeri slik predmeta zajetih na vseh štirih ozadijih.

Vir: lasten.

4.2 Postavitev eksperimenta

Imeli smo dva eksperimenta. V prvem eksperimenta smo imeli resnično in sintetično učno množico. Resnična učna množica je bila sestavljena iz 100 resničnih slik, sintetična pa iz 1500 sintetičnih slik. Pri generiranju sintetičnih slik smo uporabljali naključna ozadja. Testno množico predstavlja preostalih 100 resničnih slik. Pri razbitju slik na učno in testno množico smo zagotovili, da so bile vse slike iste scene

(pri različnih osvetlitvah) vedno bodisi v učni bodisi v tesni množici. V tem eksperimentu smo se osredotočili na primerjavo rezultatov dobljenih z učenjem na sintetičnih slikah s tistimi dobljenimi z uporabo pravih slik.

V drugem eksperimentu smo se osredotočili na analizo rezultatov glede na različna ozadja. Zato smo za testno množico vzeli vseh 200 pravih slik (50 za vsako ozadje). Za učno množico smo zgenerirali 800 sintetičnih slik z vnaprej znanimi ozadji, ki so bila podobna resničnim (200 slik za vsako izmed 4 ozadij).

4.3 Mere za vrednotenje uspešnosti

Za vrednotenje uspešnosti segmentacijskega modela smo uporabili meri uspešnosti Povprečna natančnost (ang. Average Precision) in Presek nad unijo (ang. Intersection over union, IoU).

Povprečna natančnost je mera uspešnosti, ki je določena kot ploščina pod krivuljo natančnost-priklic in je neodvisna od postavitve praga ločitve med pikslu ozadja in ospredja.

Presek nad unijo je mera uspešnosti segmentacijskega algoritma, določena kot količnik med velikostjo preseka in unije z algoritmom segmentirane regije in regije, ki je podana kot prava vrednost. Pri binarizaciji segmentacijskega izhoda, torej določanju pikslu ospredja, smo prag postavili na privzeto vrednost 0,5. Običajno se optimalni prag izbere na osnovi nastavitvene množice. Ker pa v scenariju, ki ga obravnavamo v tem članku, pravih učnih in validacijskih slik ni na razpolago, smo uporabili prednastavljeno fiksno vrednost.

4.4 Segmentacijski model

Za vse eksperimente smo uporabili metodo DeepLab++ (Lin et al., 2017). Deeplab++ je globoka konvolucijska nevronska mreža, ki uporablja razširjeno konvolucijo (ang. atrous convolution) na več različnih velikostih slike. To uporablja zato, da bolje zajame globalne značilke na ravni slike. Poleg tega uporablja še vrsto drugih mehanizmov s katerimi na splošnih nalogah semantične segmentacije dosega zelo dobre rezultate in velja za eno izmed najboljših segmentacijskih metod.

Model smo učili skozi 50 epoh z uporabo goriščne izgube (ang. focal loss) (Chen et al., 2017). Uporabili smo optimizator Adam s stopnjo učenja 10^{-4} . Slike smo normalizirali s srednjo vrednostjo in varianco slik dobljenimi na podatkovni množici slik ImageNet (Deng et al., 2009). Med učenjem smo s 50% verjetnostjo sliko malce zameglili in ji spremenili svetlost, odtenek ter nasičenost. Na ta način smo učno množico slik še malce obogatili in povečali njeno raznovrstnost, da bi se izognili prevelikemu prilaganju naučenega modela na specifične učne množice.

5 Eksperimentalni rezultati

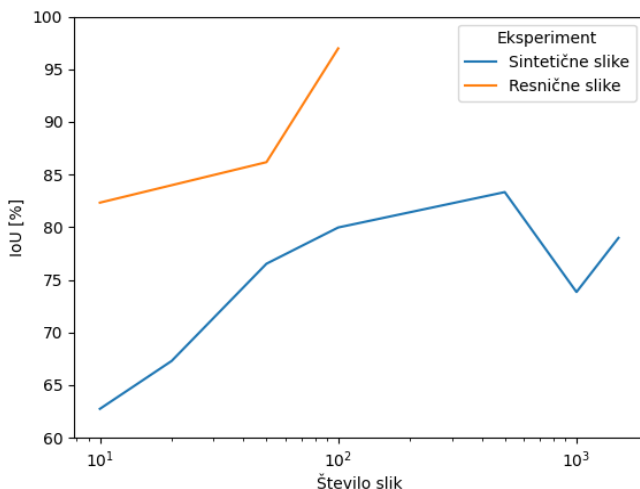
5.1 Prvi eksperiment

V prvem eksperimentu smo opazovali primernost sintetičnih slik za učno množico. Pri tem smo se osredotočili tudi na velikost učne množice. To smo storili tako, da smo ustvarili podmnožice različnih velikosti in učili model zgolj na teh. Po učenju smo vedno preverili uspešnost naučenega modela na isti testni množici.

Učne množice so bile sestavljene iz 10, 20, 50, 100, 500, 1000 in 1500 slik. Pri resničnih slikah smo vzeli samo 10, 20, 50 in 100 slik, saj imamo le 200 slik v podatkovni množici, 100 pa jih je bilo namenjenih za vrednotenje. Zaradi odvisnosti rezultatov od specifičnih slik, ki so bile izbrane v majhne učne podmnožice, smo pri majhnemu številu učnih primerov (10–500) vsak eksperiment desetkrat ponovili, pri čemer smo vsakič izbrali druge učne primere. Kot lahko razberemo iz Slike 5, rezultati dobljeni na resničnih slikah še zmeraj presegajo rezultate dobljene na sintetičnih. Pri sintetičnih slikah ravno tako lahko opazimo, da obstaja optimalno število primerov, ki se nahaja okrog 500. Pri večjem številu sintetičnih učnih slik se namreč model preveč oz. še bolj prilagodi na specifične sintetičnih slik, zato se rezultati na pravih slikah poslabšajo. Pri tem je treba poudariti, da smo za določanje praga za segmentacijo uporabili kar v naprej določen prag (vrednost 0,5), ki ni vedno optimalna. Ob bolj primernem pragu bi se ti rezultati lahko popravili.

Tabela 1 prikazuje tudi rezultate v obliki povprečne natančnosti, ki ne zahteva vnaprejšnje nastavitve vrednosti praga. Na osnovi teh rezultatov lahko sklepamo, da dobimo uspešen model že pri manjšem številu sintetičnih učnih slik. Tipično pa v praksi potrebujemo končno binarno segmentacijsko sliko, zato je moramo določiti tudi prag. In ker v tem scenariju nimamo na razpolago resničnih slik za nastavitveno

množico, saj bi se radi zajemanju resničnih slik izognili, smo primorani fiksni prag določiti vnaprej.



Slika 5: Presek nad unijo (pri pragu 0,5) v odvisnosti od velikosti učne množice

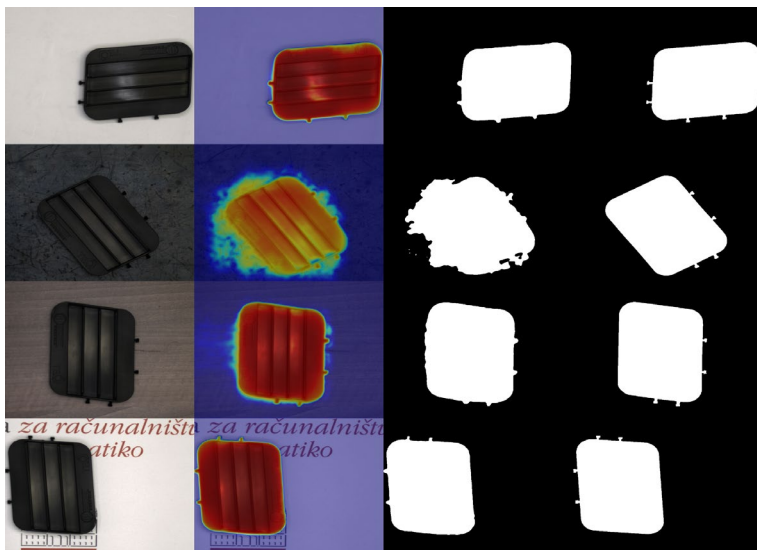
Vir: lasten.

Tabela 1: Rezultati prvega eksperimenta.

Št Učnih primerov	Tip slik	Povprečna natančnost	Presek nad unijo pri pragu 0,5
10	Resnične	99,56	82,33
10	Sintetične	98,93	62,75
50	Resnične	99,80	86,18
50	Sintetične	97,40	76,53
100	Resnične	99,94	96,98
100	Sintetične	98,05	79,97
500	Sintetične	98,17	83,34
1000	Sintetične	97,65	73,85
1500	Sintetične	96,39	78,99

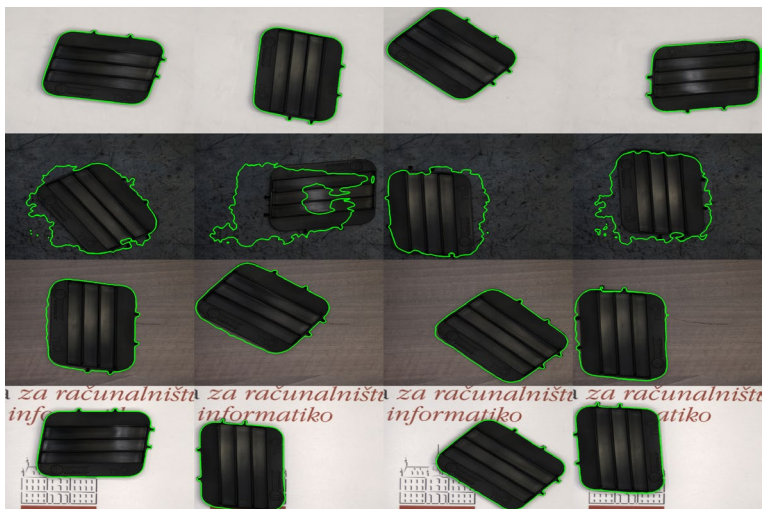
Na Slikah 6 in 7 lahko vidimo primer napovedi modela, ki se je učil na stotih sintetičnih slikah z naključnimi ozadji. Opazno je, da so rezultati zelo odvisni od ozadja testnih slik. Ne glede na število sintetičnih slik v učni množici so zaključki podobni. Na belem ozadju in na plakatu se model zelo lepo nauči tudi detajlov predmeta. Pri lesenem ozadju je kontrast med predmetom in ozadjem manjši, zato se ob učenju na sintetičnih slikah z naključni ozadjem nekaj detajlov na robu

predmeta izgubi. Najslabši rezultati so doseženi na slikah s črnim ozadjem, saj je kontrast med predmetom in ozadjem zelo majhen.



Slika 6: Primeri iz testne množice in napovedi modela pri učenju s 100 sintetičnimi slikami.

Vir: lasten.



Slika 7: Primeri napovedi modela, naučenega s stotimi sintetičnimi slikami, na različnih podlagah.

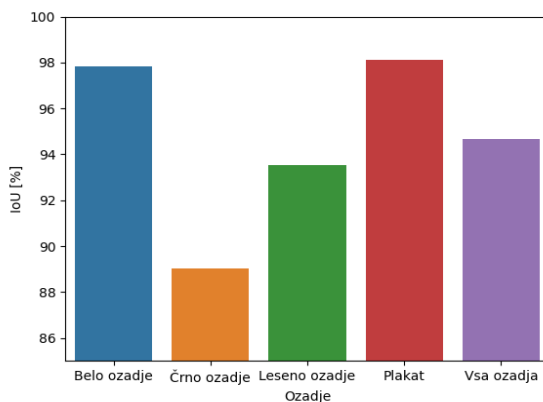
Vir: lasten.

5.2 Drugi eksperiment

Rezultati prvega eksperimenta torej kažejo na to, da je uspeh zelo odvisen od ozadja predmeta. Za pričakovati je, da se bodo rezultati bistveno izboljšali, če se model ne bo učil na naključnih ozadjih, temveč na ozadjih podobnim tistim v testni množici. To je tudi realen scenarij saj je približen izgled pravih ozadij v industrijskem okolju poznan.

V drugem eksperimentu smo tako analizirali primernost sintetičnih slik z vnaprej znanim ozadjem za uporabo segmentacij resničnih slik. Tako smo model učili na sintetičnih slikah z enim ozadjem in vrednotili na resničnih slikah posnetih na podobnem ozadju. Na koncu smo model naučili na vseh slikah z vsemi štirimi ozadji in ga tudi ovrednotili na pravih slikah z vsemi štirimi ozadji.

Iz Slike 8 in je mogoče razbrati, da učenje na sintetičnih slikah z istimi ozadji, kot pri resničnem zajemanju, prinese zadovoljive rezultate na resničnih slikah. Opazimo lahko, da je rezultat odvisen od težavnosti ozadja, saj dobimo boljše rezultate na belem ozadju in na plakatu, kjer nismo imeli problemov tudi v prvem eksperimentu.



Slika 8: Presek nad unijo (pri pragu 0,5) za različna ozadja.

Vir: lasten.

Tabela 2 prikazuje tudi rezultate v povprečni natančnosti. Na podlagi teh rezultatov lahko ponovno sklepamo, da je rezultat v veliki meri odvisen od ozadja, ki je bilo prisotno v učni množici.

Tabela 2: Rezultati drugega eksperimenta

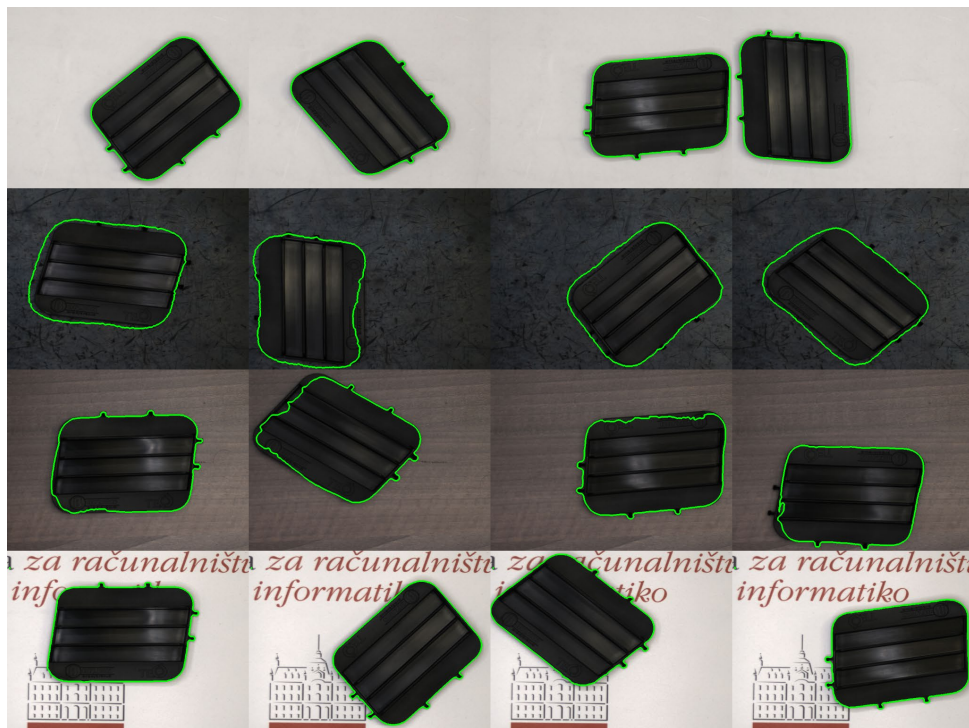
Št. učnih primerov	Ozadje	Povprečna natančnost	Presek nad unijo pri pragu 0,5
200	Belo	99,97	97,83
200	Črno	99,23	89,01
200	Leseno	99,64	93,53
200	Plakat	99,98	98,11
800	Vsa	99,71	94,67

Na Slikah 9 in 10 lahko vidimo nekaj primerov delovanja našega modela. Kot lahko razberemo se pri belem ozadju in plakatu lepo nauči tudi manjše detajle predmeta. Ravno tako se teh detajlov nauči pri lesenem ozadju, ampak nekoliko slabše. Najslabše deluje na črnem ozadju, kjer se nauči samo glavne oblike predmeta. Na kontrastnih prvem in zadnjem ozadju so rezultati torej blizu optimalnim, medtem ko so rezultati na manj kontrastnih ozadjih slabši.



Slika 9: Primeri iz testne množice in napovedi našega modela pri učenju modela na resničnih ozadjih.

Vir: lasten.



Slika 10: Primeri napovedi modela, naučenega s sintetičnimi slikami z enakimi podlagami kot pri resničnem zajemanju, na različnih podlagah.

Vir: lasten.

6 Zaključek

V tem članku smo predstavili cevovod za generiranje sintetičnih slik iz 3D CAD modelov. Predstavili smo zgenerirano množico slik in preverili uporabnost sintetičnih slik za učenje segmentacije resničnih slik. Pri pregledu uporabnosti teh slik smo proučili tudi vpliv velikosti učne množice slik. V skladu s pričakovanji smo ugotovili, da je model bolj uspešen, če mu podamo resnične slike. Pokazali smo tudi, da sintetična množica ne sme biti prevelika. Nato smo se osredotočili na problem, ko je ozadje vnaprej poznano. Pokazali smo na uporabnost sintetičnih slik s takim ozadjem, saj se rezultati modela naučenega zgolj na sintetičnih slikah približajo rezultatom modela dobljenega s pravimi slikami, dasiravno jih še ne dosežejo.

V prihodnjem delu nameravamo ta razkorak med rezultati dobljenimi s sintetičnimi in resničnimi slikami še zmanjšati. Tako nameravamo še izboljšati postopek generiranja sintetičnih slik, ga bolj obogatiti in povečati naključnost in različnost pojavnosti. Po drugi strani pa nameravamo učenje s sintetičnimi slikami nadgraditi z informacijo vsebovano v majhnem številu pravih slik. Tako bomo naslovili problem premika domene (ang. domain shift), ki se v industrijskem okolju pogosto zgodi tudi zaradi postopnih sprememb v okolju v katerem se izvaja industrijski proces. Raziskali bomo torej možnost adaptacije na domeno (ang. domain adaptation), tako preskok s sintetičnih na prave slike kot prilagoditev na spremembe pojavnosti zajetih slik, pri čemer bo glavno vodilo, da bodo te spremembe čim bolj preprosto izvedene, brez potrebe po zajemanju in predvsem označevanju novih pravih slik.

Opombe

To delo je bilo delno financirano s strani ARRS projektov L2-3169 (MV4.0) in J2-4457 (RTFM) ter raziskovalnega programa Računalniški vid (P2-0214). Posebna zahvala gre tudi podjetju TECOS in Alešu Hančiču za izdelavo predmetov in 3D CAD modela uporabljenega v eksperimentu.

Literatura

- Deng, J., Dong, W., Socher, R., Li, L., Li, K. and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, pages 248-25.
- Wang, Q., Gao, J., Lin, W., & Yuan, Y. (2019). Learning from synthetic data for crowd counting in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8198-8207).
- Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for data: Ground truth from computer games. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 102-118). Springer International Publishing.
- McCormac, J., Handa, A., Leutenegger, S., & Davison, A. J. (2017). Scenetnet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation?. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2678-2687).
- Pinaya, W. H., Tudosiu, P. D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., ... & Cardoso, M. J. (2022, October). Brain imaging generation with latent diffusion models. In *Deep Generative Models: Second MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings* (pp. 117-126). Cham: Springer Nature Switzerland.
- Akrouf, M., Gyepesi, B., Holló, P., Poór, A., Kincsó, B., Solis, S., ... & Fazekas, I. (2023). Diffusion-based Data Augmentation for Skin Disease Classification: Impact Across Original Medical Datasets to Fully Synthetic Images. *arXiv preprint arXiv:2301.04802*.
- Dewi, C., Chen, R. C., Liu, Y. T., & Tai, S. K. (2022). Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Computing and Applications*, 34(24), 21465-21480.
- Mayer, N., Ilg, E., Haussler, P., Fischer, P., Cremers, D., Dosovitskiy, A., & Brox, T. (2016). A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4040-4048).

- Li, W., Pan, C. W., Zhang, R., Ren, J. P., Ma, Y. X., Fang, J., ... & Yang, R. G. (2019). AADS: Augmented autonomous driving simulation using data-driven algorithms. *Science robotics*, 4(28), eaaw0863.
- Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., ... & Yang, R. (2018). The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 954-960).
- Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., ... & Pérez, P. (2019). Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9308-9318).
- Zhan, W., Sun, L., Wang, D., Shi, H., Clause, A., Naumann, M., ... & Tomizuka, M. (2019). Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*.
- Mayershofer, C., Ge, T., & Fottner, J. (2021). Towards fully-synthetic training for industrial applications. In *LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences* (pp. 765-782). Springer Singapore.
- Eversberg, L., & Lambrecht, J. (2021). Generating images with physics-based rendering for an industrial object detection task: Realism versus domain randomization. *Sensors*, 21(23), 7901.
- Abou Akar, C., Tekli, J., Jess, D., Khoury, M., Kamradt, M., & Guthe, M. (2022, October). Synthetic Object Recognition Dataset for Industries. In *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (Vol. 1, pp. 150-155). IEEE.
- Quattrocchi, C., Di Mauro, D., Furnari, A., & Farinella, G. M. (2022, May). Panoptic Segmentation in Industrial Environments Using Synthetic and Real Data. In *Image Analysis and Processing—ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part II* (pp. 275-286). Cham: Springer International Publishing.
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
- Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Denninger, M., Sundermeyer, M., Winkelbauer, D., Zidan, Y., Olefir, D., Elbadrawy, M., ... & Katam, H. (2019). Blenderproc. *arXiv preprint arXiv:1911.01911*.
- Community, B. O. (2018). Blender - a 3D modelling and rendering package. Stichting Blender Foundation, Amsterdam. Retrieved from <http://www.blender.org>

