

# Semantična segmentacija za detekcijo kompaktnih površinskih anomalij z deflektometrijo

Lojze Žust, Alan Lukežič, Matej Kristan

Fakulteta za računalništvo in informatiko, Univerza v Ljubljani  
lz7966@student.uni-lj.si, {matej.kristan, alan.lukezic}@fri.uni-lj.si

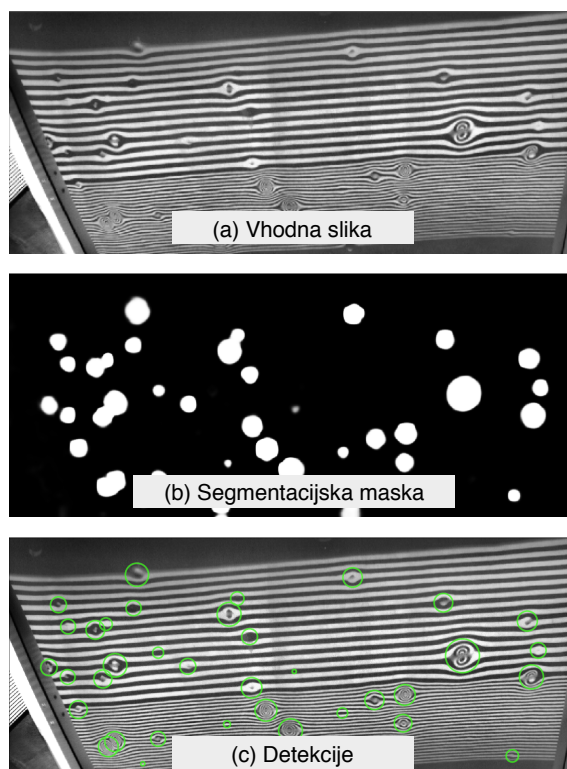
## Povzetek

Standardni pristopi za detekcijo površinskih anomalij in poškodb na odbojnih površinah temeljijo na tridimenzionalni rekonstrukciji objekta in primerjavo z referenčnim modelom nepoškodovanega objekta. To pa zahteva natančno kalibracijo, poravnavo z referenčnim modelom in analizo 3D rekonstrukcije, kar zmanjša uporabno vrednost metode in uvaja omejitve na hitrost premikanja objekta. V članku predlagamo nov pristop za detekcijo preko semantične segmentacije – SADNet. Metodo sestavljata dve stopnji: prva stopnja je semantična segmentacija anomalij, druga pa interpretacija detekcij iz segmentacijske maske. Zasnova SADNet omogoča detekcijo tudi delno prekrivajočih se anomalij. Metodo smo ovrednotili na domeni detekcije udrtin v pločevini avtomobila in jo primerjali z dvema izmed trenutno najboljših metod s področja detekcije objektov: Faster R-CNN [11] in Mask R-CNN [5]. Metoda SADNet zanesljivo prekaša obstoječi metodi. Na testni množici dosega  $F$ -mero 0.86, kar predstavlja 5% izboljšavo v primerjavi z drugo najboljšo metodo.

## 1 Uvod

Detekcija anomalij je ključen del številnih industrijskih procesov. Z njo lahko pravočasno zaznamo poškodbe in napake, ki nastanejo med proizvodnim procesom. V tem delu se osredotočamo na detekcijo anomalij v obliki 3D ugrezov in izboklin na odbojni površini (npr. udarci med proizvodnim procesom). Napake te oblike lahko minimalno spremenijo videz površine objekta, zato jih je težko detektirati na podlagi posnetkov kamere. S takimi problemi se ukvarjajo metode deflektometrije, ki na opazovano odbojno površino projicirajo določen svetlobni vzorec in s kamero pod drugim kotom opazujejo odboj. Na podlagi odbitega vzorca lahko sklepamo na oblikovne karakteristike površine.

Pristope za detekcijo anomalij z uporabo deflektometrije lahko v grobem razdelimo na dva dela. Prvi način deluje na podlagi 3D rekonstrukcije površine. Ta na podlagi deformacij vzorca oceni 3D model opazovane površine in detektira anomalije s primerjavo z referenčnim modelom nepoškodovanega objekta. Tak pristop zahteva kompleksen in natančen sistem zajema. Običajno je potrebno večje število slik in natančna kalibracija. Rekonstrukcija ni popolna in pojavi se šum, ki ga je včasih težko ločiti



Slika 1: Metoda deluje v dveh korakih. V prvem koraku model na podlagi vhodne slike izračuna segmentacijsko masko anomalij. V drugem koraku se iz segmentacijske maske izločijo lokacije in velikosti detekcij.

od majhnih anomalij. Težavna je tudi poravnava z referenčnim objektom.

Drugi način je detekcija anomalij na podlagi slike odboja vzorca. Tu anomalije iščemo na podlagi njihovih skupnih vizualnih lastnosti. Glavna prednost tega pristopa je prenosljivost. Metodo lahko uporabljamo tudi na objektih, ki niso bili v učni množici. Poleg tega za delovanje ne potrebuje zahtevnega sistema zajema in kalibracije. Ta način je še posebej primeren za kompaktne anomalije v obliki ugrezov oz. izboklin. Take anomalije se odražajo v značilnih ukrivitvah sicer ravnega vzorca (Slika 1), ki so neodvisne od objekta. Krivine pa nastanejo tudi na pregibih in utorih, ki ne predstavljajo anomalij, zato osnovna detekcija krivin ni dovolj. Potrebujemo učljivo metodo,

ki se je sposobna naučiti razlikovati med njimi. Problem lahko obravnavamo kot detekcijo objektov.

Sodobne metode za detekcijo objektov ([11], [5]) dosegajo dosegajo odlične rezultate in so sposobne zaznati in natančno lokalizirati vizualno zelo raznolike in kompleksne objekte. Zaradi arhitekturnih omejitev (omejeno število predlaganih regij, združevanje prekrivajočih se regij), pa ti pristopi niso primerni za detekcijo objektov z visoko gostoto in prekrivanjem.

Omejitve standardnih metod naslavljamo z uvedbo novega pristopa za detekcijo kompaktnih anomalij prek semantične segmentacije SADNet (angl. Segmentation for Anomaly Detection). Ta za vsak piksel napove prisotnost anomalije. SADNet deluje dvostopenjsko (Slika 1). V prvem koraku model napove segmentacijsko masko anomalij na vhodni sliki. V drugem koraku se iz segmentacijske maske izluščijo lokacije in velikosti detekcij. Povezava med obema korakoma omogoča robustno detekcijo, ki zanesljivo deluje tudi v primeru delnega prekrivanja anomalij. Eksperimentalni rezultati kažejo, da SADNet bistveno prekaša klasične metode za detekcijo objektov. Na podlagi razvite metode je bila vložena tudi patentna prijava na slovenskem patentnem uradu.

## 2 Sorodna dela

Detekcijo anomalij na podlagi videza lahko umestimo v področje detekcije objektov. Cilj detekcije objektov je klasifikacija in določanje očitnega okvirja (angl. bounding box, BB) različnih objektov na sliki. Večina metod deluje v dveh delih. Prvi del predlaga fiksno število interesnih regij, drugi del pa izloči regije, jih klasificira in izboljša lokalizacijo. R-CNN [4] s selektivnim iskanjem določi 2000 zanimivih regij. Iz vsake izmed predlaganih regij izluščimo značilke z uporabo CNN. Na podlagi značilk se izvede klasifikacija in regresijsko popravljanje okvirja regije. Metoda Fast R-CNN [3] značilke izračuna na celotni sliki zgolj enkrat, še pred izločanjem regij in tako pospeši metodo. Metoda Faster R-CNN [11] postopek še pospeši in počasno selektivno iskanje nadomesti z učljivo mrežo, ki generira predloge zanimivih regij. Dodatne izboljšave z arhitekturo R-FCN [2], ki odločitveni nivo prestavijo pred izločanje regij, še pospešijo delovanje metode. Vmes se je pojavila še druga vrsta metod (YOLO [10], SSD [8], RetinaNet [7]), ki ne uporabljajo dvostopenjske arhitekture. Namesto tega izvedejo klasifikacijo in lokalizacijo v enem samem koraku z enotno mrežo. Tak pristop je še hitrejši.

Pri SADNet smo se detekcije objektov lotili na drug način. Osnovo metode predstavlja semantična segmentacija. Cilj semantične segmentacije je klasifikacija na nivoju pikselov. Nekatere metode detekcije objektov napovedujejo tudi segmentacijo detektiranega objekta. Metoda Mask R-CNN [5] gradi na arhitekturi Faster R-CNN, le da iz značilk napoveduje še segmentacijsko masko objekta. Naš pristop pa temelji na segmentaciji celotne slike. Trenutno najuspešnejše metode za semantično segmentacijo temeljijo na konvolucijskih nevronskih mrežah (CNN). Razdelimo jih lahko v dva glavna pristopa. Največ mrež deluje na principu enkoder-dekoder (FCN [9], SegNet [1], U-Net [12]). Enkoder iz slike izlušči značilnice, dekoder

pa iz značilnic generira segmentacijsko masko. Običajno med enkoder in dekoder delom potekajo še dodatne povezave (bližnjice), ki izboljšajo natančnost segmentacije. Drugi pristop je z uporabo t. i. razširjene (angl. dilated) konvolucije [14]. Ta nadomesti sloje združevanja, tako da se velikost slike skozi mrežo ne spreminja. Doseg mreže namesto tega poveča z razširitvijo konvolucijskih filtrov, tako da med učljive elemente vstavi fiksne ničelne elemente.

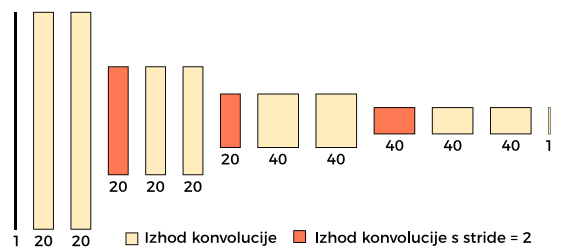
## 3 Metoda SADNet

Metoda SADNet je sestavljena iz dveh delov (Slika 1). V prvem delu se izvede napoved prisotnosti anomalije na vsakem pikslu (Poglavje 3.1). V drugem delu pa se z interpretacijo napovedane segmentacije izračunajo pozicije in velikosti anomalij (Poglavje 3.2).

### 3.1 Prisotnost anomalije na pikslu

Za napovedovanje prisotnosti anomalije na pikslu uporabimo CNN za semantično segmentacijo. Segmentacijska mreža je sestavljena iz 12 konvolucijskih slojev (Slika 2). V mreži se prostorska dimenzija reprezentacij trikrat prepolovi. Združevanje v teh slojih je implementirano z uporabo koraka (angl. stride) na konvolucijskih slojih. Izhod (verjetnostna mapa) je tako po obeh dimenzijah osemkrat manjši od originalne slike. Rezultat z bikubično interpolacijo razširimo na velikost vhoda. V preliminarni evalvaciji smo ugotovili, da uporaba učljivega dekoderja za povečevanje izhoda ne izboljša uspešnosti metode.

Anomalije so v segmentacijski maski v učnih primerih predstavljene s krogi (Slika 1). Ta pristop omogoča ločevanje delno prekrivajočih se anomalij v postopku interpretacije maske (Poglavje 3.2). V postopku učenja za cenilno funkcijo uporabljamo binarno križno entropijo. Mrežo optimiziramo z metodo ADAM [6].



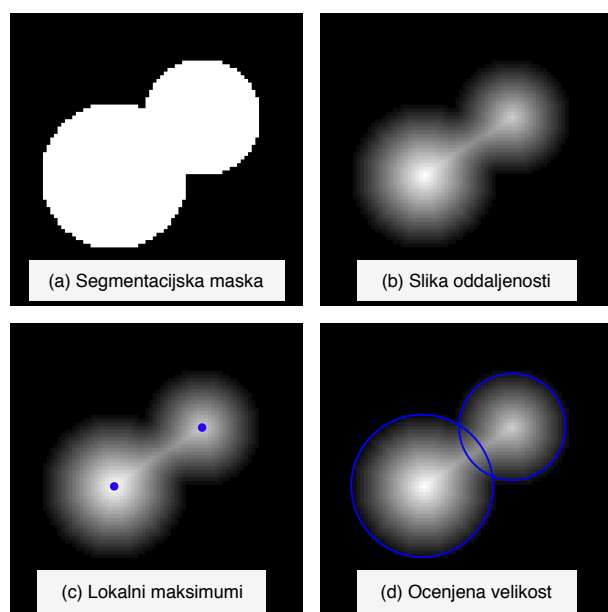
Slika 2: Arhitektura mreže za semantično segmentacijo. Mreža je sestavljena iz dvanajstih konvolucijskih slojev. Na sliki so prikazani izhodi posameznih konvolucijskih slojev. Številke označujejo število kanalov izhoda.

### 3.2 Interpretacija segmentacije

Rezultat prvega koraka metode (Poglavje 3.1) je verjetnostna mapa anomalij. V koraku interpretacije metoda iz verjetnostne mape izlušči detekcije (lokacije centrov in velikosti). Verjetnostno mapo metoda najprej upraguje z vrednostjo  $\Theta$ . Rezultat je binarna segmentacijska maska anomalij. Če so anomalije dovolj narazen, lahko v njej enostavno poiščemo detekcije z metodo povezanih

komponent. V primerih z visoko gostoto in prekrivanjem anomalij pa pride do združevanja večih anomalij v eno povezano komponento (Slika 3). SADNet za ločevanje takih skupkov izkorišča krožno predstavitev anomalij.

Razdalja točke znotraj kroga do najbližje točke, ki ne pripada krogu, doseže najvišjo vrednost v središču kroga. Ko združimo dva ali več krogov, tako da je prekrivanje posameznega kroga manj kot polovično, se ta lastnost lokalno ohranja. Središča krogov, ki predstavljajo skupek, tvorijo lokalne maksimume razdalj do zunanosti (Slika 3). Ta princip uporabimo za določanje centrov anomalij.



Slika 3: Postopek izločanja detekcij iz segmentacijske maske (a). Nad segmentacijsko masko izvedemo postopek računanja razdalj do najbližjega roba (b). Lokalni maksimumi predstavljajo središča detekcij (c), razdalja v teh točkah pa velikost detekcij (d).

Na segmentacijski maski apliciramo algoritem računanja oddaljenosti [13] (angl. distance transform), ki za vsako pozitivno točko v sliki poišče razdaljo do najbližje negativne točke. V tej matriki razdalj poiščemo lokalne maksimume, ki predstavljajo približne centre anomalij. Zaradi šuma in nizke ločljivosti se lahko v bližini pravilnega lokalnega maksimuma pojavi več lokalnih maksimumov. Ta šum odpravimo z morfološko dilatacijo [13], ki poveže bližnje šumne vrhove v enega. Iz rezultata izločimo vrhove z metodo povezanih komponent in za vsako komponento izračunamo masni center. Ta predstavlja center anomalije. Njeno velikost pa dobimo kot razdaljo od centra do najbližjega roba. Preberemo jo iz prej izračunane matrike razdalj.

#### 4 Eksperimentalna evalvacija

SADNet smo evalvirali na domeni detekcije udrtin v pločevini avtomobila z deflektometrijo in jo primerjali z metodama za detekcijo objektov Faster R-CNN in Mask R-CNN. Za učenje in evalvacijo metod smo pripravili zbirko slik, ki smo jih zajeli z namenskim svetlobnim tunelom.

Ta vsebuje projektor, ki na avtomobil projicira črtast vzorec in kamero, ki zajema slike odbitega vzorca. Sistem zajema slike v ločljivosti  $2048 \times 1088$ . Zbirka vsebuje 83 slik, zajetih na 13 različnih avtomobilih. Anotacije vsebujejo informacijo o poziciji središča anomalije  $(x, y)$  in velikosti anomalije  $r$ . Anotiranje je potekalo ročno. Zbirko sestavlja skupno 2847 anomalij (povprečno 34.3 anomalije na sliko), povprečna velikost (radij) anomalije pa je 13.15 piksla.

Zbirko smo razdelili na učno in testno množico. Učna množica vsebuje 42 slik, testna pa 41. Vse slike zajete na enem avtomobilu se pojavljajo izključno v eni izmed množic. Iz anotacij učne množice se generira segmentacijska maska, ki jo dobimo tako, da anotacije v maski predstavimo z ujemajočimi krogi. Anotacije očitane okvirja za učenje obeh standardnih metod pa dobimo kot očitane kvadrate krožnih anotacij v zbirki. SADNet smo učili z naključno inicializiranimi utežmi, metodi Faster R-CNN in Mask R-CNN pa smo inicializirali z utežmi, prednaučenimi na zbirki ImageNet. Za parameter upravljanja smo uporabili  $\Theta = 0.5$ .

Pri ocenjevanju in primerjavi metod smo uporabili naslednje metrike. Natančnost (angl. precision, Pr) označuje delež detekcij, ki so pravilne. Priklic (angl. recall, Re) označuje delež anotacij, ki jih je metoda detektirala. F-mera združuje natančnost in priklic v enotno metriko po enačbi

$$F_1 = \left( \frac{P^{-1} + R^{-1}}{2} \right)^{-1} = 2 \cdot \frac{P \cdot R}{P + R}. \quad (1)$$

Ocenjevali smo tudi natančnost ocene velikosti detekcije. Velikost detekcije pri standardnih metodah smo dobili kot radij kroga, ki ga lahko včrtamo v očitane okvire detekcije. Merili smo povprečno absolutno napako velikosti (ASE) v pikslih in povprečno relativno napako velikosti (RSE). Izmerili smo tudi povprečen čas izvajanja metode (T) za posamezno sliko. Rezultati so prikazani v Tabeli 1.

Tabela 1: Rezultati evalvacije posameznih metod. Najboljši rezultati v posameznem stolpcu so odebeljeni. Metoda SADNet dosegla najboljše rezultate v vseh merah, razen v hitrosti in natančnosti.

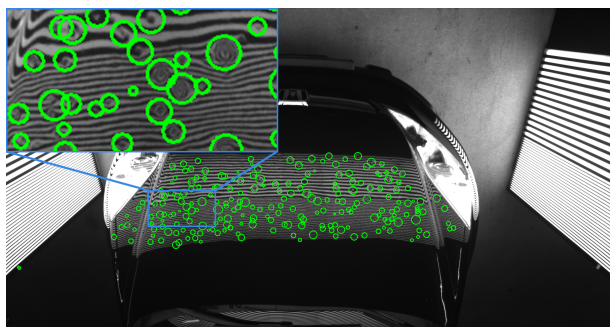
Metoda	F1	P	R	ASE	RSE	T [s]
SADNet	<b>0.86</b>	0.86	<b>0.86</b>	<b>2.48</b>	<b>0.40</b>	0.095
Mask R-CNN	0.82	0.89	0.76	3.23	0.47	0.211
Faster R-CNN	0.79	<b>0.91</b>	0.69	3.08	0.43	<b>0.061</b>

Metoda SADNet prekaša obe obstoječi metodi in detektira 13% več anomalij od druge najboljše metode (Mask R-CNN). Po natančnosti sta standardni metodi nekoliko boljše, po F-meri, ki združuje obe metriki, pa je SADNet približno 5% boljše od Mask R-CNN. V ocenjevanju velikosti detekcij SADNet predstavlja 7% izboljšavo v primerjavi s Faster R-CNN, ki je na drugem mestu. Razlika v delovanju metod je še posebej očitna na primerih z visoko gostoto in prekrivanjem anomalij (Slika 4). V tem primeru





(a) Detekcije metode Mask R-CNN



(b) Detekcije metode SADNet

Slika 4: Primerjava detekcij na primeru z velikim številom anomalij. Prikazana je tudi povečava dela slike. Na gostih področjih metoda SADNet detektira bistveno večje število anomalij. Tudi velikosti so bolje ocenjene.

obstoječi metodi po več anomalij združujeta v posamične detekcije, velik del pa jih tudi izpustita.

Metoda SADNet sliko obdela povprečno v času 0.095s (10.5 slik na sekundo), kar je približno 55% počasneje od Faster R-CNN in 55% hitreje od Mask R-CNN. Občutna razlika je tudi v prostorski kompleksnosti modela. Uteži obeh modelov standardnih metod na disku zasedejo nekaj več kot 500 MB prostora, medtem ko uteži SADNet zasedejo manj kot 0.5 MB prostora.

## 5 Zaključek

V članku smo predstavili metodo SADNet za vizualno detekcijo kompaktnih anomalij z deflektometrijo. Metoda izvaja detekcijo prek semantične segmentacije in deluje v dveh stopnjah. V prvi stopnji konvolucijska nevronska mreža napove segmentacijsko masko anomalij, v drugi stopnji pa se iz nje izločijo lokacije in velikosti detekcij. Metoda SADNet je zaradi svoje zasnove sposobna zaznati tudi delno prekrivajoče anomalije. Metodo smo ovrednotili na domeni detekcije udrtin na pločevini avtomobila in jo primerjali z dvema trenutno najboljšima metodama detekcije objektov. Eksperimentalni rezultati kažejo, da je metoda SADNet zelo uspešna in prekaša obe stari metodi. Izredno natančna je tudi na slikah z visoko gostoto anomalij, kjer klasične metode za detekcijo objektov odpovejo.

Prednosti razvite metode odpirajo zanimive možnosti za nadaljnje delo. SADNet bi bilo zanimivo testirati tudi na drugih domenah izven detekcije anomalij, kjer se po-

javlja visoka gostota kompaktnih objektov (npr. štetje). Potencialne izboljšave vidimo tudi v nadgradnji segmentacijskega dela metode, tako da segmentacija centrov in velikosti anomalij potekata ločeno. V eni segmentacijski maski nato preprosto najdemo centre anomalij (brez problemov prekrivanja), iz druge pa razberemo velikost.

## Literatura

- [1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.
- [2] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: object detection via region-based fully convolutional networks. *CoRR*, abs/1605.06409, 2016.
- [3] Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.
- [4] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.
- [5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.
- [6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [7] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017.
- [8] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.
- [9] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.
- [10] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.
- [11] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [13] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.
- [14] F. Yu and V. Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv e-prints*, November 2015.