

Detekcija napak na površinah z uporabo anotiranih slik in globokim učenjem

Domen Tabernik, Danijel Skočaj

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

E-pošta: {domen.tabernik,danijel.skocaj}@fri.uni-lj.si

Surface anomaly detection using annotated images and deep learning

Automated surface anomaly detection using machine learning has become an interesting area of research with a very high direct impact to the application domain of visual inspection. Deep learning approaches seem to be very appropriate for enabling to teach inspection systems detecting surface anomalies by showing them a number of exemplar images. In this paper we present and analyze a deep learning architecture for segmentation of surface anomalies upgraded with a simple classification function that differentiates between images of faulty and defect free surfaces. The preliminary results show that the approach is very promising and that the deep learning paradigm is appropriate to be applied in the domain of automated visual inspection.

1 Uvod

V industrijskih procesih je pregled površin izdelkov in polizdelkov ena izmed pomembnih nalog za zagotavljanje ustrezne kakovosti končnega izdelka. Pogosto se nadzor kakovosti površin izvaja ročno, kjer se strokovne delavce izuri v prepoznavanju zelo kompleksnih napak površin. Tak nadzor je seveda zelo zamuden in stroškovno neučinkovit, prav tako pa lahko močno omejuje kapaciteto proizvodnje. V ta namen se v industriji že vrsto let razvijajo avtomatski postopki nadzora kakovosti površin. Razviti postopki zajemajo metode klasičnega strojnega vida, kjer se razvija metode prilagojene za specifično vrsto napake in specifičen izdelek. S tem je sicer omogočena pohitritev proizvodnega procesa in povečanje kapacitet, vendar pa običajno tako razvite metode ni mogoče enostavno aplicirati za nadzor kakovosti drugih tipov izdelkov. To naredi razvoj metod stroškovno neučinkovit ter onemogoča fleksibilnost proizvodnih procesov.

Alternativni pristop k reševanju tega problema je uporaba strojnega učenja. Ta pristop je bolj splošen, saj razvite metode omogočajo prilagajanje na nove tipe izdelkov in površinskih napak ne z razvojem novih algoritmov, pač pa z uporabo obstoječih algoritmov strojnega učenja, pri čemer za gradnjo novih modelov oz. klasifikatorjev uporabimo nove učne primere. Prilagajanje na nove tipe izdelkov in napak je tako veliko enostavnejše in hitrejše, celoten proizvodni proces pa postane veliko bolj fleksibilen.

V zadnjem času so se na področju razpoznavanja vizualnih objektov uveljavile metode globokega učenja. Le-te dosegajo presenetljive rezultate na različnih domenah, od prepoznavanja predmetov, do segmentacije in sledenja objektov. Metode globokega učenja so bile aplicirane tudi na problemu nadzora kakovosti površin [1, 2, 3, 4, 5]. Chen in Ho [1] sta aplicirala metodo OverFeat za detekcijo petih tipov napak, Faghieh-Roohi in ostali [2] so aplicirali metode za nadzor kakovosti železniških tirov, Weimer in ostali [5] pa so ovrednotili globoko učenje na različnih vrstah sintetičnih napak. Slednji izkazujejo tudi zelo dobre rezultate. Skupno vsem zgoraj omenjenim člankom je uporaba iste metode, t.j., nevronske mreže in globokega učenja, za naslavljanje različnih tipov napak. Metoda tako omogoča nadzor kakovosti različnih tipov izdelkov s podobno arhitekturno, kar pomeni, da bi bilo mogoče uporabiti enotno arhitekturo za različne tipe napak. Pri tem se za vsak tip izdelka in napake prilagodi le proces učenja, ki pa je z izjemo anotacije, oz. označevanja napak na učnih primerih, lahko povsem avtomatiziran.

V tem delu bomo predstavili naš pristop k rešitvi tega problema in preliminarne rezultate, ki smo jih dobili s procesiranjem slik specifičnega industrijskega polizdelka. Nadzor kakovosti površin smo zastavili kot segmentacijski problem, kjer smo lokalizirali napake na nivoju slikovnih elementov. Za to smo ustrezno zasnovali arhitekturo mreže za učenje segmentacije. Dobljene segmentacijske mape smo nato uporabili za kreiranje odločitvene funkcije, ki dokončno za vsako testno sliko ugotovi ali le-ta vsebuje anomalijo ali ne. Ta postopek je opisan v naslednjem poglavju. V poglavju 3 podamo preliminarne rezultate evalvacije sistema, kjer se v glavnem osredotočamo na primerjavo različnih kriterijskih funkcij in vrst labeliranja učnih slik. Na koncu podamo v poglavju 4 še glavne ugotovitve in smernice za nadaljnje delo.

2 Segmentacija oz. detekcija napak

Segmentacijsko mrežo smo zasnovali na arhitekturni osnovi mreže VGG [6], ki pa smo jo prilagodili za naš problem. Uporabili smo 11-nivojsko arhitekturo konvolucijskih mrež, s postopnim trikratnim zmanjševanjem ločljivosti in združevanjem po maksimumu (ang. max-pooling). Za razliko od mrež [5,6] smo za vsak nivo uporabili paketno normalizacijo (ang. batch normalization). Za vse

nivoje smo uporabili jedra velikosti 5x5, z izjemo predzadnjega nivoja, ki je zajel področje veliko 15x15 elementov prejšnjega nivoja. S povečanjem velikosti jedra na predzadnjem nivoju smo poskrbeli za zajem večjega dela slike. Na zadnjem nivoju tako vsak slikovni element izhoda zajame 120x120 slikovnih elementov vhodne slike. Število kanalov se je razlikovalo od nivoja do nivoja, kot je razvidno iz Tabele 1. Na zadnjem nivoju smo za redukcijo kanalov brez upoštevanja prostorske lokalnosti uporabili velikost jedra 1x1. S tem smo dobili enokanalno izhodno sliko z 8-kratno zmanjšano ločljivostjo.

Mrežo smo učili kot binarni segmentacijski problem, torej smo za vsak slikovni element izhodne slike ovrednotili ali vsebuje napako ali ne. Vrednosti blizu 0 so predstavljale odsotnost napake, vrednost blizu 1 pa prisotnost napake. Omeniti velja, da mreža ne izvaja binarne segmentacije na vsak vhodni slikovni element natančno, ampak to naredi za skupino 8x8 slikovnih elementov skupaj. Ker smo se pri zasnovi izognili uporabi polno-povezanih nivojev, nam ni bilo potrebno uporabiti mehanizma drop-out ali regularizacije. Zaradi odsotnosti polno-povezanih nivojev je celotna arhitektura tudi neodvisna od velikosti vhodne slike. Učenje smo primarno definirali kot regresijski problem, za učno kriterijsko funkcijo (ang. loss function) pa smo uporabili tako povprečje kvadratov napak (ang. mean squared error), kot tudi križno entropijo (ang. cross entropy).

Tabela 1: Arhitektura segmentacijske mreže.

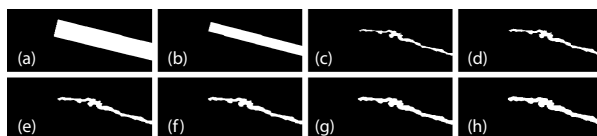
Vhoda slika
conv5x5, 32
conv5x5, 32
max-pool 2x2
conv5x5, 64
conv5x5, 64
conv5x5, 64
max-pool 2x2
conv5x5, 64
conv5x5, 64
conv5x5, 64
conv5x5, 64
max-pool 2x2
conv15x15, 1024
conv1x1, 1

Čeprav smo problem detekcije napak zasnovali kot segmentacijski problem, je v samem bistvu dejansko to binarni klasifikacijski problem; za vsako sliko je potrebno ugotoviti ali vsebuje napako ali ne. Zato smo za končni rezultat detekcije napak implementirali odločitveno funkcijo, ki na podlagi izhoda segmentacijske mreže sliko klasificira kot primerno ali neprimerno. Odločitveno funkcijo smo implementirali kot logistično regresijo nad preprostim opisnikom slike. Za opisnik slike smo uporabili dvo-dimenzionalen vektor sestavljen iz maksimalne in povprečne izhodne vrednosti segmentacijske mreže.

3 Eksperimentalni rezultati

Razviti pristop smo ovrednotili na specifični bazi industrijskih polizdelkov; nekaj primerov je prikazanih na Sl. 6. Podatkovna zbirka zajema 400 slik. Od tega je bila na 50 slikah prikazana površina z napako (razpoke ipd.), na 350 slikah pa je bil upodobljen polizdelek brez napake. Arhitekturo smo na podatkovni zbirki ovrednotili s 3-kratnim prečnim preverjanjem, pri čemer smo za učenje vzeli 2/3 primerov, za testiranje pa preostalo tretjino.

Za vse slike z napako smo imeli na voljo tudi anotacije napake in sicer i) velike, ii) grobe in iii) fine (na slikovni element natančne) anotacije. Primer anotirane razpoke je prikazan na Sl. 1(a-c). Ker je anotiranje primerov (še posebej natančno) zelo zamudno opravilo, smo posebno pozornost posvetili analizi vpliva vrste anotacije na rezultate. Poleg zgoraj omenjenih anotacij smo v evalvacijo dodatno vključili še 5 anotacij, kjer smo nad fino anotacijo izvajali morfološko operacijo razširjanja z različnimi stopnjami (širitve od 1 do 5 slikovnih elementov, kot so prikazane na Sl. 1(d-h)) in na ta način simulirali manj natančne anotacije.



Slika 1: Primeri različnih vrst anotacij: (a) velika, (b) groba, (c) fina, (d-h) razširjena fina anotacija s stopnjami od 1 do 5.

Predstavljeno mrežo smo ovrednotili s štirimi metriki: a) krivuljo natančnost-priklic (ang. precision-recall curve), (b) povprečno natančnostjo (ang. average precision), (c) št. zgrešenih slik (ang. false negatives) in (d) št. napačno najdenih primerov (ang. false positives). Natančnost segmentacijske maske nismo ocenjevali na slikovni element natančno, ampak so bile metrike osredotočene na klasifikacijo celotne slike. Kot primarno metriko smo izbrali povprečno natančnost, ki je definirana kot površina pod krivuljo natančnost-priklic, saj metoda detekcije ne vrača binarne odločitve pač pa verjetnostno odločitev z vrednostmi med 0 in 1, in je torej št. zgrešenih slik in št. napačno najdenih primerov odvisno od upravljenja, kar ponazarjajo krivulje na Sl. 2. Št. zgrešenih slik in št. napačno najdenih primerov na Sl. 3 in 4 ocenimo pri pragu, ki je določen z najboljšo f-mero. Dodatno poročamo na Sl. 5 tudi št. napačno najdenih primerov pri zagotovljenem popolnem priklicu, torej ko ni nobenih zgrešenih slik. Glavni namen te metrike je ocenjevanje mreže pod realnimi pogoji, kjer je potrebno zagotoviti detekcijo vseh napak, natančnost pa se zagotavlja z dodatnim ročnim preverjanjem s strani eksperta.

Sprva smo ovrednotili mrežo glede na izbrano kriterijsko funkcijo in glede na uporabljeno vrsto anotacije. Rezultati eksperimentov so prikazani na Sl. 2, 3 in 4, ki po vrsti prikazujejo krivulje natančnost-priklic, povprečno natančnost ter število napačnih odločitev pri najboljši f-meri. Pri tem število napačnih odločitev prikažemo razdeljeno po št. zgrešenih slik (FN) in št. napačno

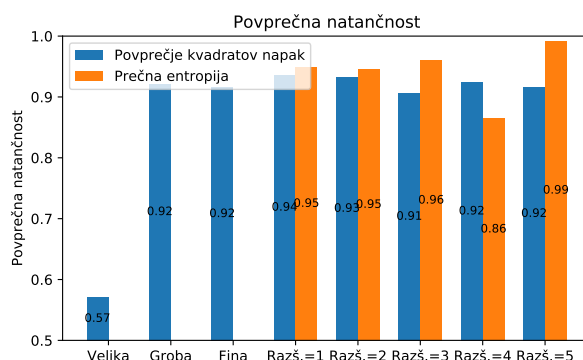
najdenih primerov (FP). Prikazani so rezultati za obe obravnavani kriterijski funkciji, pri čemer prvi trije tipi anotacije niso bili ovrednoteni s prečno korelacijo.

Če se osredotočimo na delovanje pri različnih vrstah anotacije, lahko opazimo, da velika anotacija daje najslabše rezultate, medtem ko natančne anotacije in anotacije z razširjanjem dajejo boljše rezultate s podobno povprečno natančnostjo. Pri najboljšem pragu glede na f-mero dajejo boljše rezultate anotacije z razširjanjem, kjer lahko na Sl. 4 opazimo med 4 in 9 napačnih primerov, enakomerno porazdeljenih po št. zgrešenih slik in št. napačno najdenih primerov. Vsi ti rezultati se nanašajo na ovrednotenje na 400 testnih primerih.

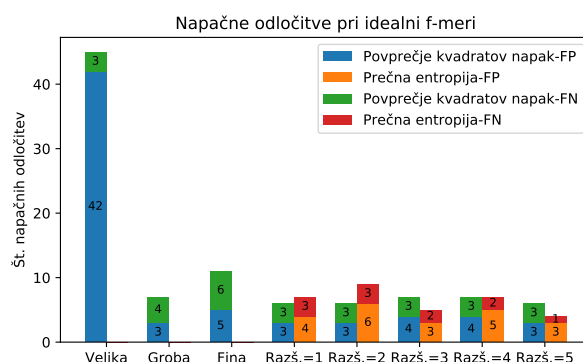
Pri primerjavi kriterijskih funkcij vidimo, da boljše rezultate zagotavlja prečna entropija, ki ima v večini primerov višjo povprečno natančnost. Pri najboljšem pragu se prečna entropija lahko obnese bolje, je pa to odvisno od uporabljene vrste anotacije. Pri 3-kratnem razširjanju anotacije v najboljšem primeru dobimo 5 napačnih odločitev (2 zgrešena primera in 3 napačno najdene primere), pri 5-kratnem razširjanju anotacije pa dobimo le 1 zgrešen primer in 3 napačno najdene primere.

Na Sl. 5 prikazujemo tudi število napačno najdenih primerov, ko prag za detekcijo nastavimo tako, da ne zgrešimo nobene slike primera z napako. V tem primeru lahko opazimo zelo dobro delovanje prečne entropije, vendar le pri specifični stopnji razširjanja anotacije. Pri 2-kratnem

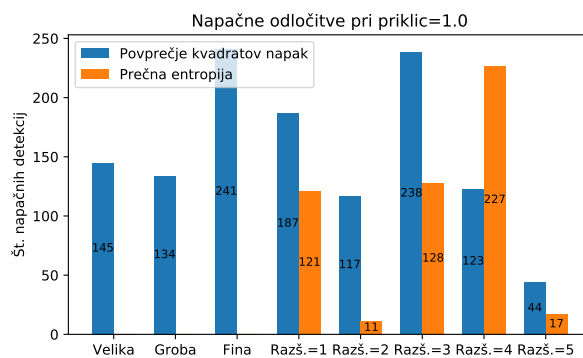
razširjanju anotacije dobimo le 11 napačnih odločitev, pri 5-kratnem razširjanju pa 17 napačnih odločitev.



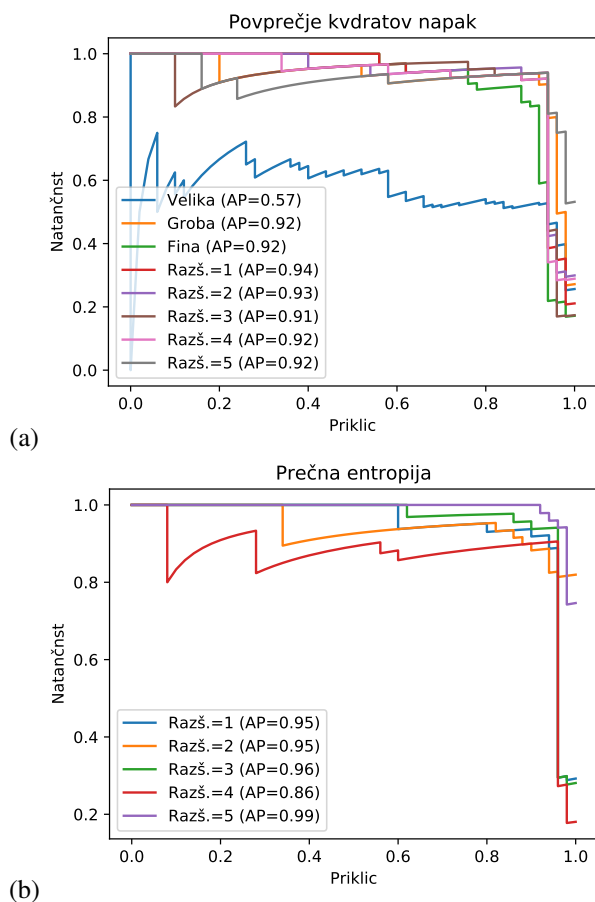
Slika 3: Povprečna natančnost za različne anotacije in kriterijske funkcije.



Slika 4: Napačne odločitve pri različnih anotacijah in kriterijskih funkcijah.

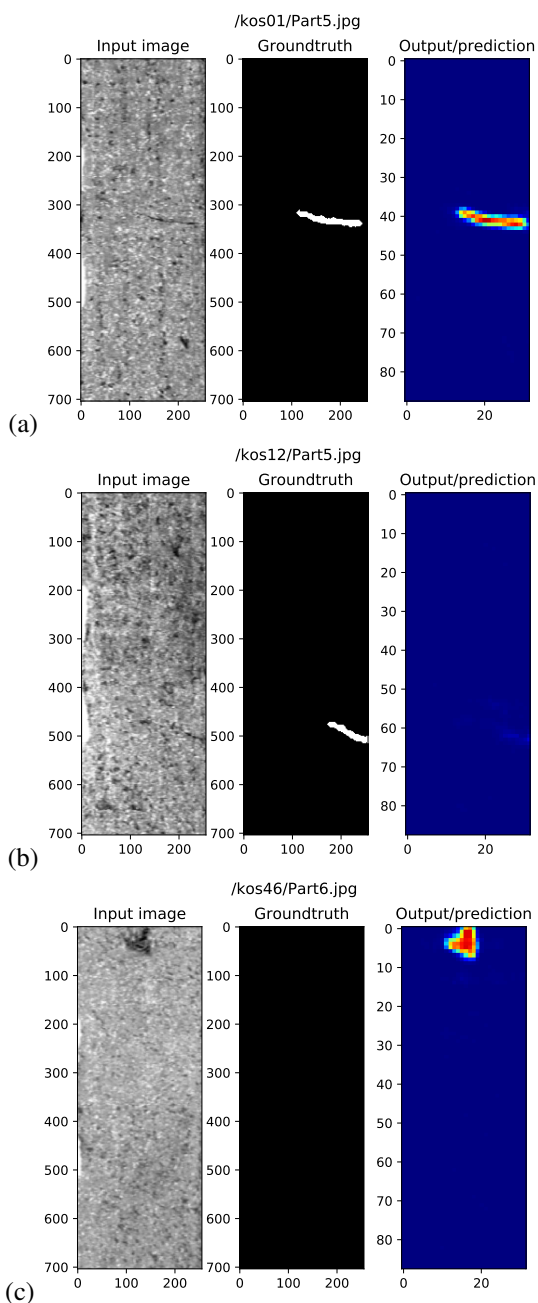


Slika 5: Napačno detektirani primeri pod pogojem, da ni izpuščen noben primer z napako.



Slika 2: Krivulje natančnost/priklic pri uporabi dveh kriterijskih funkcij: (a) povprečne kvadratne napake, (b) prečne korelacije.

Nekaj kvalitativnih primerov detekcije je prikazanih na Sl. 6, ki prikazuje en primer uspešne detekcije (ki so sicer absolutno prevladujoče) ter dva neuspešna primera. Sl. 6(c) tako prikazuje napačno detektirano anomalijo. Nepravilnost, ki je na sliki vizualno popolnoma jasna, namreč s strani ekspertnega anotatorja ni bila anotirana kot napaka, ker takšne vrste neregularnosti na površinah za dotični industrijski polizdelek niso usodne.



Slika 6: Primeri detekcij napak: (a) uspešna detekcija, (b) zgrešen primer (ang. false negative), (c) napačna detekcija (ang. false positive).

4 Zaključek

V tem članku smo ovrednotili konvolucijsko nevronske mreže na problemu detekcije površinskih napak na specifičnem industrijskem polizdelku. Predlagali smo 11-nivojsko segmentacijsko mrežo, ki smo jo nadgradili z logistično regresijo kot odločitveno funkcijo, ki posamezne primere klasificira kot ustrezne oz. neustrezne. Predlagan pristop smo ovrednotili pod različnimi pogoji, pri čemer smo spreminjali vrsto anotacije in vrsto kriterijske funkcije.

Preliminarni rezultati so sicer precej nekonsistentni oz. odvisni od pogojev vrednotenja, torej od vrste anotacije. Najboljše rezultate smo dosegli z uporabo prečne entro-

pije kot kriterijske funkcije in anotacije razširjene s stopnjo 5. Nekonsistentni rezultati nakazujejo na potrebo po dodatnem vrednotenju na drugih podatkovnih zbirkah, predvsem na drugačnih tipih površinskih napak. Navkljub temu lahko opazimo zelo dobre zmogljivosti globokih mrež pri detekciji površinskih napak. V veliki večini primerov je bil razpoznavalnik uspešen; le na 2-5 primerih se je mreža odločila napačno, kar v povprečju pomeni okrog 1% vseh primerov.

Iz rezultatov lahko sklepamo tudi na nadaljnje poti za izboljšave. Uporabljena odločitvena funkcija je precej preprosta in jo nameravamo nadgraditi ter vključiti v poenoteno segmentacijsko-klasifikacijsko arhitekturo nevronske mreže. Ravno tako nameravamo raziskati uporabnost generativnih metod in sicer z namenom zmanjšanja števila učnih primerov oz. v izogib potrebi po učnih primerih s površinskimi napakami, ki jih je v industrijskem okolju relativno težko pridobiti. Z uporabo generativnih metod bi lahko namesto učenja napak naslovili problem učenja pravilnosti izgleda površin. V prihodnje bomo zato raziskali tudi ustreznost metod kot so GAN [6] in variacijski avtoenkoderji (ang. variational-autoencoders) [7] za problem detekcije površinskih napak.

Zahvala

To delo je bilo delno financirano s strani programa GO-SSTOP, št. pogodbe C3330-16-529000, ki ga sofinancirata Republika Slovenija in EU (ERDF). Posebna zahvala gre tudi podjetju Kolektor Orodjarna, PC Vision, še posebej Samu Šeli in Juretu Skvarču, za pomoč pri urejanju in anotaciji zbirke slik.

Literatura

- [1] Pei-Hung Chen and Shen-Shyang Ho. Is Overfeat Useful for Image-Based Surface Defect Classification Tasks? In *IEEE International Conference on Image Processing*, 2016.
- [2] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter. Deep Convolutional Neural Networks for Detection of Rail Surface Defects. In *International Joint Conference on Neural Networks*, number October, pages 2584–2589, 2016.
- [3] Jonathan Masci, Ueli Meier, Dan Ciresan, Jürgen Schmidhuber, and Gabriel Fricout. Steel defect classification with Max-Pooling Convolutional Neural Networks. In *Proceedings of the International Joint Conference on Neural Networks*, 2012.
- [4] D. Weimer, H. Thamer, and B. Scholz-Reiter. Learning defect classifiers for textured surfaces using neural networks and statistical feature representations. *Procedia CIRP*, 7:347–352, 2013.
- [5] Daniel Weimer, Bernd Scholz-Reiter, and Moshe Shpitalni. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Annals - Manufacturing Technology*, 65(1):417–420, 2016.
- [6] Ij Goodfellow, J Pouget-Abadie, and Mehdi Mirza. Generative Adversarial Networks. *arXiv:1406.2661*, 2014.
- [7] Kingma Max Welling. Variational Auto-Encoders and Extensions. 2015.