

# OCENJEVANJE KAKOVOSTI PRSTNIH SLEDI Z ANSAMBLI GLOBOKEGA UČENJA

Tim Oblak, Peter Peer

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana  
{tim.oblak,peter.peer}@fri.uni-lj.si

## Izveček

Ocena kakovosti je pomemben korak za identifikacijo prstnih sledi s kraja zločina. Pogosto se izvaja v okviru forenzične preiskave, izvajajo pa ga usposobljeni preiskovalci in je ponavadi precej subjektiven. Cilj našega dela je razviti avtomatizirano metodo ocenjevanja kakovosti prstnih sledi, ki bi pomagala izpraševalcem pri njihovem delu. V tem delu sodobne tehnike globokega učenja prenesemo na področje ocenjevanja kakovosti prstnih sledi, ovrednotimo prednosti in slabosti te metodologije ter načrtujemo smer nadaljnega razvoja na področju. Predlagamo novo metriko kakovosti, ki s pomočjo posebne metode fuzije združuje posamezne napovedi ansambla globokih modelov. Predlagani pristop zagotavlja izboljšano zmogljivost napovedi, hkrati pa skrajša čas obdelave za vsaj 15-krat v primerjavi z obstoječimi najsodobnejšimi rešitvami.

**Ključne besede:** forenzika, globoko učenje, ocena kakovosti, prstne sledi

## FINGERMARK QUALITY ASSESSMENT WITH DEEP LEARNING ENSEMBLES

### Abstract

Quality assessment is an important step when trying to identify fingerprints from a crime scene. Often done in the scope of forensic investigation, it is performed by trained examiners and tends to be rather subjective. The goal of our work is to develop an automated fingerprint quality assessment method, which would assist the examiners in their work. In this paper, we introduce modern deep learning techniques into the field of fingerprint quality assessment, we evaluate the advantages and disadvantages of this methodology, and identify key aspects for further development in the field. We propose a new quality metric, which works by fusing individual predictions of an ensemble of deep models. The proposed approach provides improved prediction performance while reducing processing time by at least a factor of 15 compared to existing solutions.

**Keywords:** forensics, deep learning, quality assessment, fingerprints

## 1 UVOD

Prstne sledi (latentni prstni odtisi) so posebna vrsta odtisov kože s konic naših prstov, ki jih najdemo v nenadzorovanih okoljih, na primer v okviru forenzične preiskave [1]. V takih primerih odtis kožnega vzorca ni nadzorovan in med postopkom pride do napak. Med analizo sledi preiskovalci najprej ocenijo njihovo kakovost. Ta vrednost pomeni, (i) kako uporabna je sled kot forenzični dokaz, (ii) določa, kakšen bo postopek obdelave sledi, in (iii) kaže na uspešnost potencialnega ujemanja s referenčnimi prstnimi odtisi. Ocena kakovosti prstnih sledi ima torej velik vpliv na preiskavo. Kot je značilno za človeško naravo, so tudi odločitve forenzičnih preiskovalcev zelo subjektivne [7, 8, 12]. Uporabna prstna sled se zato lahko zavrže že zgodaj v preiskavi ali pa se preveč časa nameni prepoznavanju prstnih sledi slabe kakovosti. Da bi zmanjšali to subjektivnost, želimo avtomatizirati proces ocenjevanja kakovosti prstnih sledi.

Trenutno več dejavnikov omejuje razvoj novih metod avtomatiziranega ocenjevanja kakovosti prstnih sledi (*angl.* automated fingerprint quality assessment - AFQA). Komerzialni ponudniki svoje algoritme ščitijo kot intelektualno lastnino. Nekatere metode so bile razvite posebej za organe pregona in niso javno dostopne [9]. Običajno raziskave vključujejo nadzorovano učenje napovedovalnih modelov [2, 3, 4], redko pa avtorji objavijo tudi kodo, naučene modele, podatke ali referenčne vrednosti. Čeprav obstajajo odprtokodne rešitve za

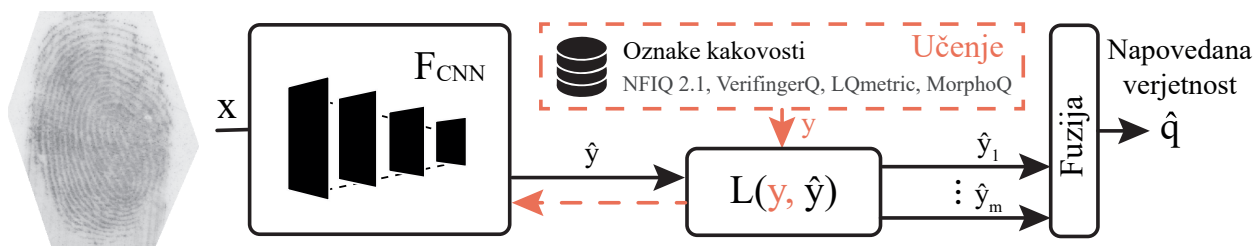


Figure 1: **Model za ocenjevanje kakovosti prstnih sledi.** Metoda, predlagana v tem delu, je sestavljena iz CNN ansambla in metode fuzije, ki posamezne glasove združi v končno vrednost kakovosti.

ocenjevanje kakovosti odtisov s prstov [11], so te večinoma osredotočene na ocenjevanje optično branih prstnih odtisov v nadzorovanem okolju in niso primerne za uporabo v forenziki. Druga vrsta objav [13, 10] se opira na ročno sestavljene značilke, zaradi česar so metode razumljive, vendar je količina izluščenih informacij omejena. Zaradi teh dejavnikov smo ugotovili, da bi bila nova metoda ocenjevanja kakovosti prstnih sledi koristna za to znanstveno področje.

Doslej metode globokega učenja na tem področju še niso bile uporabljene. V tem delu preučujemo, kako je mogoče konvolucijske nevronske mreže (*angl.* convolutional neural network - CNN) najbolje uporabiti v kontekstu metod AFQA in predlagamo novo metodo, prikazano na sliki 1. V okviru raziskave (i) zgradimo ansambel CNN-jev, ki so naučeni na različnih množicah oznak kakovosti. Za označevanje našega nabora podatkov uporabljamo obstoječe komercialne in zaprtokodne metode AFQA. V drugem delu (ii) predlagamo metodo fuzije, ki vzame posamezne glasove ansambla in jih združi, da ustvari enotno, robustnejšo vrednost napovedane kakovosti.

## 2 ANSAMBEL ZA OCENJEVANJE KAKOVOSTI PRSTNIH SLEDI

**Formulacija problema.** Za napovedovanje vrednosti kakovosti  $y$  iz vhodne slike  $x$  v tem delu načrtujemo strategijo učenja, kjer naučeni modeli  $F_{CNN}$  napovejo ocene kakovosti, ki so čim bližje resničnim oznakam:  $\hat{y} = F_{CNN}(x; \theta_{CNN})$ , kjer je  $\hat{y}$  izhod konvolucijske nevronske mreže,  $\theta_{CNN}$  pa so učljive uteži mreže. V namen ansambelskega učenja predlagamo arhitekturo modela  $F_{CNN}$  posplošimo tako, da izhod  $\hat{y} \in \mathbb{R}^m$  definiramo kot vektor kakovosti velikosti  $m \in \mathbb{N}$ . V konfiguraciji ansambla je lahko  $F_{CNN}$  bodisi CNN z izhodom velikosti  $m$  ali zbirka  $m$  CNN modelov, od katerih je vsak naučen na svojem naboru oznak kakovosti. Optimalne parametre  $\theta$  iščemo tako, da minimiziramo funkcijo izgube  $\mathcal{L}(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ , kjer je  $Y$  nabor resničnih vektorjev kakovosti,  $\hat{Y}$  je množica napovedanih vektorjev kakovosti,  $n$  pa je število slik prstnih sledi v učni množici.

**Arhitektura CNN.** Kot osnovno ogrodje našega modela lahko uporabimo katero koli od uveljavljenih arhitektur CNN. V kontekstu prstnih oznak nam razširitev zaznavnega polja v CNN omogoča obdelavo slik prstnih sledi, ne da bi vnaprej poznali ločljivost slik. Ker je ciljna vrednost zvezno število, spremenimo izhod mreže in ustvarimo regresijski model. Za povečanje zmogljivosti mreže na podlagi predhodnega testiranja na koncu mreže dodamo še en popolnoma povezan sloj s 512 nevroni. Za tem dopolnimo mrežo še z izhodno plastjo z  $m$  nevroni.

**Metoda fuzije.** Da končno napoved kakovosti naredimo bolj robustno in čim bolj pokrijemo širok razpon možnih lastnosti prstnih sledi, lahko napovedi ansambla združimo s pomočjo fuzije. Za pravilno uteževanje posameznih ocen kakovosti želimo izračunati projekcijo  $\mathbb{R}^n \rightarrow \mathbb{R}$  iz  $m$ -dimenzionalnega vektorja ocen kakovosti  $y$  na enodimenzionalno končno vrednosti  $q$ . S pomočjo PCA najprej najdemo os največje variance znotraj  $m$ -dimenzionalnega prostora glasov ansambla, potem pa posamezne vektorje kakovosti preslikamo na to enodimenzionalno os. Preslikavo definiramo na naslednji način:

$$\hat{q}_{pca} = W(\hat{y} - \bar{y}), \quad (1)$$

kjer je  $W \in \mathbb{R}^{1 \times n}$  lastni vektor največje glavne komponente (*angl.* principal component), pridobljen z metodo PCA,  $\hat{y}$  je napovedani vektor kakovosti od katerega odštejemo povprečni vektor resničnih vrednosti kakovosti  $\bar{y}$ ,  $\hat{q}_{pca} \in \mathbb{R}$  pa je prva glavna komponenta. Za transformacijo končne vrednosti v razpon kakovosti  $(0, 100)$  izhod dodatno normaliziramo:

$$\hat{q} = (\hat{q}_{pca} - q_{pca}^{min}) / (q_{pca}^{max} - q_{pca}^{min}), \quad (2)$$

Table 1: **Rezultati vrednotenja napovednikov CNN.** Modeli so testirani na testni množici in primerjani z uporabo metrike MAE in  $R^2$ . Najboljšo napovedno zmogljivost je ansambel dosegel z uporabo ogrodja DenseNet-121.

Model	ResNet-50		DenseNet-121		EfficientNet-b0		Inception-v3	
	MAE	$R^2$	MAE	$R^2$	MAE	$R^2$	MAE	$R^2$
NFQ	3,41	0,49	3,03	<b>0,60</b>	3,14	0,55	3,38	0,37
VFQ	4,29	0,73	4,20	<b>0,76</b>	4,20	0,75	4,12	0,75
LQM	7,81	0,79	7,44	0,79	6,92	<b>0,83</b>	7,64	0,78
MOR	2,11	0,81	2,01	<b>0,83</b>	1,99	0,82	1,95	0,82
Čas izvajanja [ms]	19,59 ± 0,10		21,05 ± 0,85		11,1 ± 1,89		<b>10,2 ± 0,89</b>	

kjer sta  $q_{pca}^{max}$  največja,  $q_{pca}^{min}$  pa najmanjša glavna komponenta, izračunana z resničnih vrednosti učne množice,  $\hat{q}$  pa je končni rezultat fuzije napovedanih vrednosti kakovosti.

### 3 REZULTATI

**Eksperimentalno okolje.** Za učenje modelov uporabljamo dve množici prstnih sledi, NIST SD 302 [5] in NIST SD 301 [6]. Obe vsebujeta prstne sledi, ki so jih z različnih površin pridobili usposobljeni forenzični preiskovalci v simuliranem okolju. Podatke smo razdelili v učno množico (10.000 prstnih sledi + 2000 prstnih odtisov z množice SD 302) in testno množico (1200 prstnih sledi z množice SD 301). Namesto ročnega označevanja resničnih vrednosti kakovosti smo uporabili obstoječe algoritme za ocenjevanje kakovosti in tako ustvarili ciljne vrednosti za učenje. Uporabili smo štiri metode; NFIQ 2.1 [11], VerifingerQ, MorphoQ in LQmetric [9]. Za vrednotenje uspešnosti naučenih modelov smo uporabili metrike, ki se v literaturi običajno uporabljajo za regresijske naloge, in sicer povprečno absolutno napako (*angl.* mean absolute error - MAE) in r-kvadrat (*angl.* r-squared -  $R^2$ ).

**Ovrednotenje modelov.** Naš cilj je naučiti napovedovalni model za vsako množico ciljnih vrednosti kakovosti, pridobljenih z metodami NFIQ 2.1, VerifingerQ, MorphoQ in LQmetric. Naučene modele poimenujemo NFQ, VFQ, MOR oziroma LQM. Poleg tega za vsakega od štirih modelov testiramo štiri različna ogrodja CNN-jev, ki se pogosto uporabljajo v literaturi (ResNet, DenseNet, EfficientNet in Inception). Rezultati so prikazani v tabeli 1. Največje razlike so opažene med modeli NFQ, kjer se vrednost  $R^2$  giblje od 0,37 do 0,60. Zdi se, da so modeli, usposobljeni z drugimi vrednostmi kakovosti, bolj konsistentni med različnimi arhitekturami. Na splošno je natančnost regresije izbranih osnovnih arhitektur relativno visoka. MAE v povprečju predstavlja okoli 5% možnega razpona kakovosti od 0 do 100. Pri ocenjevanju ogrodij je na splošno najboljši DenseNet, sledi mu EfficientNet. DenseNet dosega najboljšo korelacijsko zmogljivost za tri od štirih modelov in ima najvišje povprečno vrednost  $R^2 = 0,74$ . Prav tako izstopa kot najboljši napovedovalec metode NFIQ 2.1 v primerjavi z drugimi arhitekturami. Drug vidik je čas izvedbe. Za dano serijo slik naučen model napove kakovosti v konstantnem času. Od vseh štirih ogrodij sta EfficientNet in Inception-v3 očitno najhitrejša, saj oba obdelata serijo slik v približno 10 ms. V primerjavi NFIQ 2.1 porabi najmanj 150 ms za obdelavo slike prstnega odtisa. CNN modeli so tako vsaj 15-krat hitrejši v primerjavi s konvencionalno metodo z ročnim luščenjem značilk. Sami po sebi pa CNN-ji ne nudijo načina za razlago svojih odločitev. Z vidika razložljivosti ostaja klasičen pristop ročno izbranih značilk boljše izbira. Več raziskav je potrebnih za združitev učinkovitosti CNN-jev in interpretabilnosti metod z ročno izbranimi značilkami.

**Ovrednotenje fuzije.** Da bi ocenili, kako se rezultat fuzije obnese v primerjavi z ostalimi metodami, lahko izmerimo njegovo napovedno moč. Na sliki 2 smo vzeli naš najuspešnejši model (DenseNet-121) in prikazali krivulje ujemanja za celoten ansambel, vključno z novo metofo fuzije. Najprej definiramo idealen scenarij, kjer krivulja sledi diagonalni pikčasti črti. V našem primeru želimo, da napovedana metrika kakovosti korelira z verjetnostjo, da bomo za dano prstno sled našli ujemajoč se referenčni prstni odtist na prvem mestu (rank-1). Če je idealna krivulja diagonalna črta, potem kakovost 20 pomeni, da obstaja 20% verjetnost rank-1 identifikacije. S krivulj slike 2 je razvidno, da imata modela NFQ in MOR najvišje naraščajoče krivulje, ki dosežejo 100% verjetnost že okoli kakovosti 40. Na drugi strani je krivulja modela LQM vedno pod diagonalno črto, kar pomeni, da model je preveč radodaren z visokimi ocenami kakovosti in da rank-1 ujemanja niso zagotovljena, tudi če je ocena kakovosti dobra. Idealni situaciji se najbolje prilaga predlagana metoda fuzije, sledi pa ji model VFQ. Glede na to, da metoda fuzije upošteva tudi porazdelitve prvotnih oznak v učni množici,

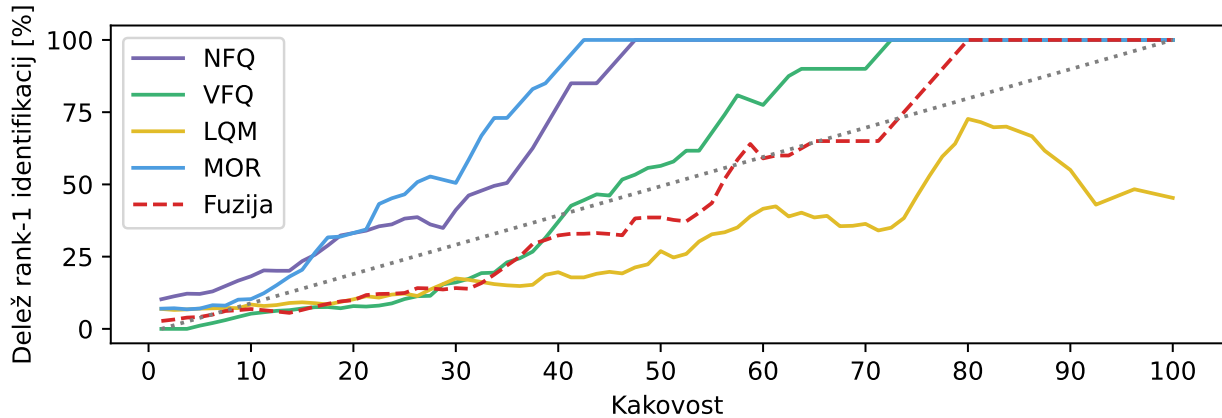


Figure 2: **Krivulje rank-1 identifikacije.** V tem grafu je razvidno razmerje med napovedano kakovostjo različnih metod in deležom rank-1 identifikacij. Za vsako kakovost v razponu od 0 do 100 prikažemo delež prstnih odtisov te kakovosti, ki so bili uspešno najdeni na prvem mestu s pomočjo avtomatizirane metode ujemanja prstnih sledi. Rezultati združenega ansambla imajo najboljšo napovedno moč v primerjavi z ostalimi metodami.

je ta sposobna bolje utežiti prispevke posameznih glasov ansambla in zato deluje bolj robustno v nasprotju s posameznimi napovedmi.

## 4 ZAKLJUČEK

V tem delu smo predlagali nov pristop za ocenjevanje kakovosti prstnih sledi s pomočjo globokega učenja. Metoda je sestavljena iz ansambla modelov CNN in metode za fuzijo, ki posamezne napovedi ansambla združi v eno samo robustno vrednost kakovosti. S tem delom forenzični in biometrični skupnosti nudimo sodobno, zanesljivo in odprtokodno metriko kakovosti prstnih sledi. V prihodnosti bomo nadaljevali z uporabo metod globokega učenja, vendar pa se bomo osredotočili na razložljivost odločitev modela in pa celovito predstavitev končnih rezultatov.

## REFERENCES

- [1] Jeffery G. Barnes. History. In *Fingerprint sourcebook*, chapter 1, pages 5–22. U.S. Department of Justice, National Institute of Justice, 2010.
- [2] Kai Cao, Tarang Chugh, Jiayu Zhou, Elham Tabassi, and Anil K Jain. Automatic latent value determination. In *International Conference on Biometrics*, pages 1–8. IEEE, 2016.
- [3] Tarang Chugh, Kai Cao, Jiayu Zhou, Elham Tabassi, and Anil K. Jain. Latent Fingerprint Value Prediction: Crowd-Based Learning. *IEEE Transactions on Information Forensics and Security*, 13(1):20–34, 2018.
- [4] Jude Ezeobiesi and Bir Bhanu. Latent fingerprint image quality assessment using deep learning. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 508–516. IEEE, 2018.
- [5] Gregory Fiumara, Patricia Flanagan, John Grantham, Kenneth Ko, Karen Marshall, Matthew Schwarz, Elham Tabassi, Brian Woodgate, and Christopher Boehnen. NIST Special Database 302: Nail to Nail Fingerprint Challenge. Technical Report 2007, NIST, 2018.
- [6] Gregory Fiumara, Patricia Flanagan, Matthew Schwarz, Elham Tabassi, and Christopher Boehnen. Nist special database 301: Nail to nail fingerprint challenge dry run. Technical Report 2002, NIST, 2018.
- [7] R Haraksim, J Galbally, and L Beslay. *Study on Fingerprint and Palmprint Identification Technologies for their Implementation in the Schengen Information System*. EUR 29755 EN, Publications Office of the European Union, 2019.
- [8] R Austin Hicklin, Joann Buscaglia, Maria Antonia Roberts, Stephen B Meagher, William Fellner, Mark J Burge, Matthew Monaco, Davill Vera, Larry R Pantzer, Calvin C Yeung, and Ted N Unnikumaran. Latent Fingerprint Quality: A Survey of Examiners. *Journal of Forensic Identification*, 61(4):385–419, 2011.

- [9] Nathan D Kalka, Michael Beachler, and R Austin Hicklin. LQMetric: A Latent Fingerprint Quality Metric for Predicting AFIS Performance and Assessing the Value of Latent Fingerprints. *Journal of Forensic Identification*, 70:443–463, 2020.
- [10] Anush Sankaran, Mayank Vatsa, and Richa Singh. Automated clarity and quality assessment for latent fingerprints. *International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2013.
- [11] Elham Tabassi, Martin Olsen, Oliver Bausinger, Christoph Busch, Andrew Figlarz, Gregory Fiumara, Olaf Henniger, Johannes Merkle, Timo Ruhlend, Christopher Schiel, and Michael Schwaiger. NIST Fingerprint Image Quality 2, NISTIR 8382. National Institute of Standards and Technology, 2021.
- [12] Bradford T. Ulery, R. Austin Hicklin, Jo Ann Buscaglia, and Maria Antonia Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, 7(3), 2012.
- [13] Soweon Yoon, Eryun Liu, and Anil K Jain. On latent fingerprint image quality. In *Computational Forensics*, pages 67–82. Springer, 2012.

*Tim Oblak* je doktorski študent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Leta 2020 je zaključil dvojni magisterij na Univerzi v Ljubljani in Tehnični univerzi v Gradcu. V svojih raziskavah uporablja metode računalniškega vida s poudarkom na globokem učenju. V preteklosti se je ukvarjal z rekonstrukcijo 3D prostora, trenutno pa v sklopu doktorskega študija razvija nove metode na področju digitalne forenzike in identifikacije prstnih sledi.

*Peter Peer* je redni profesor na Fakulteti za računalništvo in informatiko na Univerzi v Ljubljani, kjer je tudi vodja Laboratorija za računalniški vid, koordinator dvojnega študijskega programa z južnokorejsko univerzo Kyungpook National University in prodekan za gospodarske zadeve. Na Fakulteti za računalništvo in informatiko je doktoriral leta 2003. Po doktoratu je bil vabljen podoktorski raziskovalec na CEIT v San Sebastianu v Španiji. Raziskovalno se ukvarja z računalniškim vidom s poudarkom na biometriji. Sodeloval je v številnih državnih in EU projektih in je avtor več kot 100 raziskovalnih člankov v priznanih revijah in na konferencah. Nedavno je bil soorganizator tekmovanja Unconstrained Ear Recognition Challenge na IEEE/IAPR International Joint Conference on Biometrics 2017 in IAPR International Conference on Biometrics 2019. Leta 2018 je bil na to temo tudi sourednik posebne izdaje v reviji IET Biometrics. Trenutno je področni urednik v revijah IEEE Access in IET Biometrics. Je član EAB, IAPR in IEEE, v slovenski sekciji IEEE pa je 4 leta bil tudi predsednik računalniškega poglavja.