Prepoznava zamenjave obraza na slikah osebnih dokumentov

Leon Alessio[1], **Marko Brodarič**[1,2], **Peter Peer**[1], **Vitomir Štruc**[2], **Borut Batagelj**[1]

[1]*Fakulteta za Računalništvo in Informatiko, Univerza v Ljubljani, Večna pot 113, 1000 Ljubljana*
[2]*Fakulteta za elektrotehniko, Univerza v Ljubljani, Tržaška cesta 25, 1000 Ljubljana*
E-pošta: leonalessio1@gmail.com

Abstract

In recent years, a need for remote user authentication has emerged. Many authentication techniques are based on verifying an image of identity documents (ID). This approach mitigates the need for physical presence from both parties, making the authentication process quicker and more effective. However, it also presents challenges, such as data security and the risk of identity fraud. Attackers use many techniques to fool authentication algorithms. This paper focuses on detecting face substitution, a common and straightforward fraud technique where the perpetrator replaces the face image on the ID. Due to its simplicity, almost anyone can utilize this technique extensively. Unlike digitally altered images, these modifications are manually detectable but pose challenges for computer algorithms. To face the challenge of detecting such an attack, we extended a dataset containing original images of identity cards of 9 countries with altered images, where the original face was substituted with another face from the dataset. We developed a method to detect such tampering by identifying unusual straight lines that indicate an overlay on the ID. We then evaluated the method on our dataset. While the method showed limited success, it underscores the complexity of this problem and provides a benchmark for future research.

1 Uvod

V zadnjih letih se je v mnogih industrijah pojavila želja oz. potreba po avtentikaciji uporabnika na daljavo. Za to obstajajo različni načini, a eden izmed najbolj priljubljenih je avtentikacija s pomočjo slike osebne izkaznice. Avtentikacija na daljavo doprinese mnogo koristi in prednosti. Med temi je najbolj pomembna odprava nujne fizične prisotnosti, kar prihrani ogromno časa. Vendar je treba biti izredno pozoren tudi na izzive, ki jih takšna oblika avtentikacije prinaša. Zagotoviti je potrebno ustrezno zaščito in shranjevanje osebnih podatkov. Ob tem je tudi potrebno poskrbeti za preprečevanje goljufij in kraje identitet, saj do teh pride veliko lažje pri preverjanju identite na daljavo kot pa v živo. Slike osebnih dokumentov lahko morebitni goljufi modificirajo z uporabo profesionalne programske opreme za spreminjanje slik ali z uporabo tehnik globokega učenja za modificiranje teksta in slike obraza na osebnem dokumentu [1, 2, 3, 4, 5]. Orodja za modificiranje slik osebnih dokumentov so postala vse bolj dostopna in enostavna za uporabo. Po drugi strani pa je odkrivanje modificiranih slik osebnih dokumentov aktivno raziskovalno področje in je na voljo že veliko tehnik za prepoznavo modificiranih slik osebnih izkaznic z uporabo programske opreme, umetne inteligence [6, 7, 8, 9].

Veliko enostavnejša in posledično tudi veliko bolj dostopna oblika ponarejanja slik osebnih dokumentov je preprosta zamenjava slike obraza. Na tak način lahko napadalec ukane razne mehanizme avtentikacije osebnih dokumentov, ki delujejo na podlagi primerjave obraza na osebnem dokumentu in slike obraza. Napadalec lahko sliko obraza na osebni izkaznici zamenja s pomočjo programske opreme z le nekaj kliki ali na osebno izkaznico preprilepi izrezek slike kopije obraza druge osebne izkaznice in na ta način izvede poskus kraje identite na daljavo. Za tako osebno izkaznico je na pogled več kot očitno, da ni originalna, saj je zamenjani del jasno viden in izstopa glede na ostale predele slike. Vendar pa računalniški sistemi tega odstopanja ne prepoznajo jasno in večina modelov, ki uspešno prepoznavajo digitalno spremenjene lažne osebne izkaznice, pri teh ponaredkih ni učinkovita. Raziskovalci iz Univerze Fudan so za soroden problem, le da je šlo za fižično zamenjavo dela osebne izkaznice, ugotovili, da sta dve izmed najbolj popularnih platform za prepoznavanje ponaredkov osebnih dokumentov, MEGVII Face++ AI in BaiduAI, obe prepoznali manj kot 27 procentov ponaredkov [8]. Eden največjih izzivov razvoja metod in modelov prepoznave lažnih osebnih dokumentov je pomanjkanje obsežnejše zbirke osebnih dokumentov za učenje in testiranje modelov, saj slike le-teh vsebujejo občutljive osebne podatke.

Ta članek naslavlja trenutno stanje prepoznave zamenjave obraza na slikah osebnih dokumentov. Opisana je osnovna metoda prepoznave zamenjave obraza, ki smo jo razvili za tekmovanje IJCB PAD-ID Card 2024 [10], kjer je bil eden od 3 možnih napadov zamenjava obraza na slikah osebnih dokumentov. Prepoznava ostalih vrst napadov je opisana v delu K. Ocvirka s sod. [11]. Za namena razvoja naše metode in nadaljnega raziskovanja na tem področju smo sestavili podatkovno zbirko, ki vsebuje originalne slike dokumentov, pridobljene iz podatkovne zbirke za tekmovanje Document Liveness Challenge 2021 (DLC 2021) [12] in lažne derivate teh slik, na katerih smo zamenjali obraz s sliko obraza iz drugega

Raziskava je bila sofinancirana iz ARRS raziskovalnega projekta DeepFake DAD (J2-50065) in raziskovalnega programa Računalniški vid (P2-0214).

osebnega dokumenta. Metodo smo evalvirali na naši podatkovni zbirki in postavili referenčno točko za nadaljnje raziskave.

2 Sorodna dela

2.1 Ponarejanje osebnih dokumentov

Za ponarejanja osebnih dokumentov obstaja več različnih tehnik. Prva izmed njih je ponarejanje teksta na slikah osebne izkaznice [1]. Možno je tudi ponarejanje slike obraza na osebni izkaznici z uporabo raznih tehnik proizvajanja sintetične vsebina obraza (angl. Deepfake) [2, 3, 4, 5]. Napad je možen tudi z uporabo slike kopije osebne izkaznice na papirju ali pa z uporabo slike osebne izkaznice na računalniškem zaslonu. Sprememba vira slike lahko zavede različne tehnike za prepoznavanje ponaredkov [9]. Oblika ponarejanja podrobneje predstavljena v tem članku je enostavna zamenjava obraza na osebnih izkaznicah, pri kateri napadalec prelepi del originalne osebne izkaznice s sliko obraza iz druge osebne izkaznice. Na ta način bi lahko prilepil sliko svojega obraza na osebno izkaznico potencialne žrtve in tako prelisičil tehnike avtentikacije [8].

2.2 Prepoznava ponaredkov osebnih dokumentov

Razvoj učinkovitih metod prepoznavanja ponaredkov osebnih dokumentov je močno okrnjen zaradi pomankljivosti ustreznih podatkovnih zbirk, ki bi omogočile razvoj naprednejših metod. Za prepoznavo ponaredkov so bile razvite različne tehnike. Za prepoznavo ponaredkov obraza ustvarjenih z globokimi ponaredki lahko uporabimo številne metode razvite za prepoznavo le-teh [6]. Za prepoznavo ponarejenega besedila na osebni izkaznici nekateri raziskovalci priporočajo uporabo zaznavanja specifičnih lastnosti značilnosti pisave [7]. Eden izmed enostavnejših načinov izdelovanja ponaredkov, ki je precej podoben načinu, predstavljenem v tem članku, je fizično modificiranje osebnega dokumenta, tako da se del osebne izkaznice prelepi z izrezanim delom kopije drugega dokumenta. Za prepoznavanje takih napadov so raziskovalci iz Univerze Fudan že razvili uspešno metodo [8]. Za prepoznavo ponaredkov, ustvarjenih z več različnimi napadi, eden izmed njih je tudi zamenjava delov slike osebnih dokumentov z deli slike drugega osebnega dokumenta, so Gonzalez in sodelavci razvili hibridni dvostopenjski klasifikacijski model za določanje, kateri čilenski osebni dokument je ponarejen [9]. Vredno je poudariti, da omenjeni model deluje le na osebnih izkaznicah ene države, medtem ko naša metoda prepoznava ponaredke osebnih izkaznic več držav.

3 Metodologija

3.1 Priprava podatkovne zbirke

Največjega izziva pri razvoju metod in modelov za prepoznavo ponaredkov osebnih dokumentov, pomanjkanja obsežnih podatkovnih zbirk, smo se lotili tako, da smo uporabili fotografije originalnih osebnih dokumentov, ki so bile zbrane za tekmovanje DLC 2021 [12]. Na ta način smo našo novo podatkovno zbirko napolnili s fotografijami osebnih dokumentov 9 različnih držav, Albanije, Azerbajdžana, Estonije, Finske, Grčije, Latvije, Srbije, Slovaške in Španije. Za vsako je na voljo 8 unikatnih fotografij osebnih dokumentov. Ponaredke osebnih dokumentov smo izdelali tako, da smo vsaki fotografiji ročno izrezali del, kjer je bil prikazan obraz in nato vsako unikatno sliko pravega osebnega dokumenta uporabili za podlago dveh ponaredkov, katerim smo področje obraza prelepili z izrezanim področjem obraza drugega osebnega dokumenta iste vrste. Tako smo za vsako fotografijo originalnega dokumenta pridobili še dva ponaredka. Končno število ponaredkov je 144 (slika 1, 2).

Slika 1: Primer originalne slike osebnega dokumenta.

Slika 2: Primer ponaredka slike osebnega dokumenta.

3.2 Osnovna metoda prepoznave zamenjave obraza

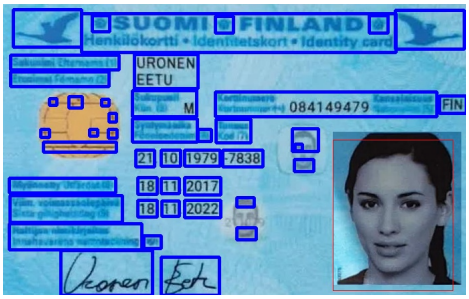
Izziv pomanjkanja obsežnih podatkovnih zbirk močno omejuje razvoj modelov za prepoznave ponaredkov osebnih dokumentov, ki temeljijo na podlagi globokega učenja, saj taki modeli potrebujejo ogromno količino slik za uspešno prepoznavanje. Zato smo se pri razvoju naše metode zgledovali po tem, kako bi oseba prepoznala zamenjavo obraza na osebnem dokumentu, kjer je jasno viden del, ki ločuje obraz od ozadja slike. Tako je bila osrednja ideja našega algoritma detekcija jasnih črt, ki nastanejo kot posledica zamenjave obraza na osebni izkaznici in se na osebni izkaznici pojavijo na neobičajnih mestih. Pri detekciji teh smo morali biti pozorni, da ne zajemamo črt iz ozadja slike in črt, ki so del teksta ali obraza. Tako je prvi korak naše metode ta, da smo sliko le nekoliko obrezali in tako poskrbeli, da robovi osebnega dokumenta in ozadje ne vplivajo na delovanje metode. Sledi detekcija teksta in obraza na izkaznici, okoli katerih postavimo mejne pravokotnike. To so območja, ki ne bodo upoštevana pri določevanju končnega rezultata. Okoli obraza smo nato definirali tudi iskalni pravokotnik, ki določa območje, na katerem iščemo neobičajne črte, ki so nastale kot posledica zamenjave. Nato metoda začne s prepoznavo črt

na osebnem dokumentu, te smo nato primerjali in odstranili tiste, ki se dotikajo pravokotnikov okoli obraza in teksta ter niso vsebovane znotraj iskalnega pravokotnika. V kolikor nam je po tem koraku ostalo še nekaj črt, katerih skupna dolžina presega mejno vrednost, ki je določena relativno, glede na velikost slike dokumenta, smo zaključili, da je dokument ponaredek. Če pa takih črt ni bilo, smo določili, da gre za originalno sliko.

3.2.1 *Prepoznava teksta in obraza*

Drugi korak metode poskrbi, da nismo pomotoma zaznali črt, ki predstavljajo dele črk ali obraza, saj te niso pomembne pri našem cilju. Lotili smo se detekcije črk in obraza. Pri implementaciji smo se poslužili knjižnice OpenCV [13]. S pomočjo metod te knjižnice že prej obrezano sliko najprej osivimo in jo nekoliko zameglimo. Sliko nato binariziramo tako, da jo pretvorimo v črno-belo s pomočjo metode Otsu [14]. Slikovni elementi, ki predstavljajo del teksta ali obraza, bi morali imeti vrednost 255 in torej biti pobarvani belo. Nato na črno-beli sliki izvedemo dilatacijo, da povežemo bližnje bele regije. Po izvedeni dilataciji poiščemo konture, robove, ki v našem primeru definirajo meje belih regij. Zelo majhne konture ignoriramo, za večje pa določimo najmanjše pravokotnike in shranimo seznam le-teh. Ti pravokotniki nam povedo, da se znotraj njih nahaja tekst ali obraz, zato morebitne črte znotraj njih ne predstavljajo fizične modifikacije osebnega dokumenta.

Za detekcijo obraza smo se poslužili metode RetinaFace [15], ki uspešno zaznava obraze na slikah. Vrnjeno območje obraza nato nekoliko povečamo, da poskrbimo, da se v analizo črt ne zajemajo lasje in druge posebnosti okoli obraza. To območje nato še občutneje razširimo in tako določimo iskalni pravokotnik, ki nam definira območje, na katerem upoštevamo najdene črte. Pri tem moramo biti pozorni na dimenzije le-tega, da ustrezno zajamemo področje dela obraza na osebni izkaznici. S tem odstranimo morebitne moteče vplive na za nas nezanimivih območjih osebnega dokumenta. Na sliki 3 lahko vidimo prepoznana območja besedila pobarvana modro in območje obraza pobarvano rdeče.



Slika 3: Primer prepoznave območja besedila in obraza.

3.2.2 *Prepoznava črt*

Metoda nato s pomočjo funkcij knjižnice OpenCV poišče vse črte na sliki osebnega dokumenta. To stori tako, da sliko najprej osivi in zamegli. Nato z uporabo Cannyjevega algoritma [16] prepozna robove in ustvari črno-belo sliko, kjer so slikovni elementi, ki predstavljajo robove beli, vsi ostali pa črni. Na ustvarjeni črno-beli sliki izvedemo Houghovo transformacijo [17], ki na vhodni binarni sliki zazna ravne črte, ki zadoščajo nastavljenim parametrom. Transformacija nam vrne seznam ravnih črt na sliki osebnega dokumenta, ki ga bomo uporabili pri naslednjem koraku. Na sliki 4 lahko vidimo vse ravne črte, ki jih je naša metoda našla na sliki osebne izkaznice.

Slika 4: Primer prepoznave črt na sliki osebnega dokumenta.

3.2.3 *Prepoznava ponaredkov*

Zadnji korak je namenjen združevanju celotne metode in podajanju rezultata. Slednjega določimo tako, da najprej za vsako črto iz seznama, pridobljenega v prejšni točki, preverimo ali se ne dotika prej omenjenega iskalnega pravokotnika in ali preseka katerokoli stranico kateregakoli pravokotnika, pridobljenega pri drugem koraku, in odstranimo tiste, ki ta pogoj izpolnjujejo. To storimo zato, da preprečimo, da bi ravne črte, ki so deli teksta ali obraza, vplivale na našo predikcijo. Preostalim črtam, torej tistim, ki ne presegajo nobenega mejnega pravokotnika in se dotikajo iskalnega pravokotnika, nato izmerimo dolžino in te dolžine seštejemo. V kolikor ta dolžina nato presega postavljeno mejo, katere vrednost je enaka desetini manjše izmed dolžine in višine osebne izkaznice, naša metoda vrne, da je ta primer osebnega dokumenta ponaredek, saj je zaznala eno ali več daljših ravnih črt na mestih, kjer se te nebi smele pojaviti. To najverjetneje pomeni, da je bil del osebnega dokumenta prelepljen in se rob nalepke prepozna kot črta. V kolikor takih črt ni oz. so te zelo kratke, pa metoda vrne, da ta osebni dokument ni bil fizično modificiran. Na sliki 5 lahko vidimo iskalni pravokotnik, pobarvan z rdečo barvo in zelene črte, ki jih je metoda označila kot neobičajne in predstavljajo robove zamenjave.

Slika 5: Vizualizacija delovanje metode na ponaredku.

Implementacija metode in opisana podatkovna zbirka sta na voljo na github repozitoriju [1].

[1] https://github.com/leonalessio/id-card-detection

4 Rezultati

V ospredju doprinosa dela, ki ga opisuje ta članek, je nova podatkovna zbirka, ki vsebuje 144 ponaredkov z zamenjanim obrazom na osebnem dokumentu in razširjajo 72 originalnih slik osebnih dokumentov tekmovanja DLC 2021 [12]. Poleg tega smo osnovno metodo za prepoznavo zamenjave obraza na osebni izkaznici testirali na naši podatkovni zbirki in tako postavili referenčno točko, na podlagi katere bo mogoče meriti napredek nadaljnega raziskovanja na tem področju.

Delovanje metode smo kvantificirali s pomočjo klasifikacijske točnosti, ki nam predstavi delež pravilno razvrščenih primerov in mere F1, ki združuje koncepta natančnosti (koliko od napovedanih pozitivnih primerov je dejansko pozitivnih) in odziva (koliko od resnično pozitivnih primerov je model napovedal pozitivno):

$$F1 = 2 \cdot \frac{\textit{natančnost} \cdot \textit{odziv}}{\textit{natačnost} + \textit{odziv}}$$

V tabeli 1 so predstavljeni rezultati metode na naši podatkovni zbirki.

Tabela 1: Rezultati metode na naši podatkovni zbirki.

Metrika	Rezultat
Klasifikacijska točnost	0.667
Natančnost	0.640
Odziv	0.695
F1 mera	0.666

Dobljeni rezultati niso izpolnili naših pričakovanj in so lahko dokaz, da predstavljeni problem ni tako enostaven kot se sprva morda zdi. Razlogov za slabše delovanje naše metode je več. Izpostavili bi predvsem težave generalizacije, saj je prepoznava ponaredkov na različnih vrstah dokumentov veliko težja, ker ima vsaka vrsta svoje značilnosti. Metoda najboljše deluje, kadar področje obraza nima izrazitih črt že na originalnih izkaznicah, veliko slabše pa deluje na izkaznicah, kjer robovi področja obraza izstopajo, zaradi katerih naša metoda te izkaznice napačno prepozna kot ponaredke. Verjamemo, da če bi metodo prilagodili le na eno vrsto osebnih dokumentov, bi na tej dosegli veliko boljše rezultate. Poleg tega je pri naši metodi težavno iskati optimalni nabor parametrov. Kadar te spreminjamo tako, da otežimo iskanje črt, dosežemo veliko boljše rezultate na originalnih osebnih izkaznicah, a hkrati veliko slabše pri ponaredkih. Podobno velja za obratno situacijo, kadar olajšamo iskanje črt. V primeru, da se sestavi obsežnejša podatkovna zbirka bi predlagali razvoj modelov na podlagi globokega učenja, saj ti pri podobnih problemih dosegajo zelo dobre rezultate.

5 Zaključek

V članku smo predstavili izziv prepoznave zamenjave obraza na osebnih dokumentih. Za ta namen smo sestavili podatkovno zbirko, ki vsebuje originalne in ponarejene osebne dokumente. Razvili smo tudi osnovno metodo za prepoznavo zamenjave obraza, ki temelji na zaznavi črt na neobičajnih mestih osebne izkaznice, ki ostanejo kot posledica zamenjave. Metodo smo tudi testirali na naši podatkovni zbirki in ta se je izkazala za le delno uspešna. Rezultati naše metode podpirajo trditev, da ne gre za enostaven problem in bodo lahko služili za primerjavo in kvantifikacijo napredka nadaljnih raziskav na tem področju. V prihodnosti se bomo osredotočili na razširitev podatkovne zbirke, da bo lahko ta podpirala razvoj naprednejših metod prepoznave. Osredotočili se bomo na razvoj metod za prepoznavanje zamenjave obraza z uporabo globokih nevronskih omrežij in tako poskusili nadgraditi sedanjo metodo.

Literatura

[1] L. Wu, C. Zhang, J. Liu, J. Han, J. Liu, E. Ding, X. Bai: Editing Text in the Wild, ACM international conference on multimedia, 2019.

[2] I. Perov, D. Gao, N. Chervoniy, K. Liu, S.Marangonda, C. Umé in sod.: DeepFaceLab: Integrated, flexible and extensible face-swapping framework, arXiv, 2020.

[3] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, M. Nießner: Face2face: Real-time face capture and reenactment of rgb videos, CVPR, 2016.

[4] L. Li, J. Bao, H. Yang, D. Chen, F. Wen: High Fidelity Identity Swapping for Forgery Detection, CVPR, 2020.

[5] J. Thies, M. Zollhöfer, M. Nießner: Deferred neural rendering: Image Synthesis using Neural Textures, ACM Transactions on Graphics, 2019.

[6] G. Pei, J. Zhang, M. Hu, Z. Zhang, C. Wang, Y. Wu, G. Zhai, J. Yang, C. Shen, D. Tao: Deepfake Generation and Detection: A Benchmark and Survey, arXiv, 2024.

[7] R. Bertrand, O. R. Terrades, P. Gomez-Krämer, P. Franco, and J. Ogier: A conditional random field model for font forgery detection, ICDAR, 2015.

[8] H. Wang, S. Li, S. Cao, R. Yang, J. Zeng, Z. Qian, X. Zhang: On Physically Occluded Fake Identity Document Detection, ACM International Conference on Multimedia, 2023.

[9] S. Gonzalez, A. Valenzuela, J. Tapia: Hybrid Two-Stage Architecture for Tampering Detection of Chipless ID Cards, IEEE TBIOM, 2021.

[10] IJCB PAD-ID Card 2024, https://sites.google.com/view/ijcb-pad-id-card-2024/.

[11] K. Ocvirk, M. Brodarič, P. Peer, V. Štruc, B. Batagelj: Primerjava metod za zaznavanje napadov ponovnega zajema, ERK, 2024 (v recenziji).

[12] D. V. Polevoy, I. V. Sigareva, D. M. Ershova, V. V. Arlazarov, D. P. Nikolaev, Z. Ming, M. M. Luqman, J.-C. Burie: Document Liveness Challenge Dataset (DLC-2021), Journal of Imaging, 2022.

[13] G. Bradski: The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.

[14] N. Otsu: A Threshold Selection Method from Gray-Level Histograms, IEEE TSMC, 1979.

[15] S. Sefik, A. Özpınar: A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules, Journal of Information Technologies, 2024.

[16] J. Canny: A Computational Approach To Edge Detection, IEEE TPAMI, 1986.

[17] R. O. Duda, P. E. Hart: Use of the Hough transformation to detect lines and curves in pictures, Communications of ACM, 1972.