Primerjava metod za zaznavanje napadov ponovnega zajema

Krištof Ocvirk[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, Ljubljana*
[2]*Fakulteta za elektrotehniko, Univerza v Ljubljani, Tržaška cesta 25, Ljubljana*
E-pošta: ko7000@student.uni-lj.si

Comparison of Methods for Detecting Recapture Attacks

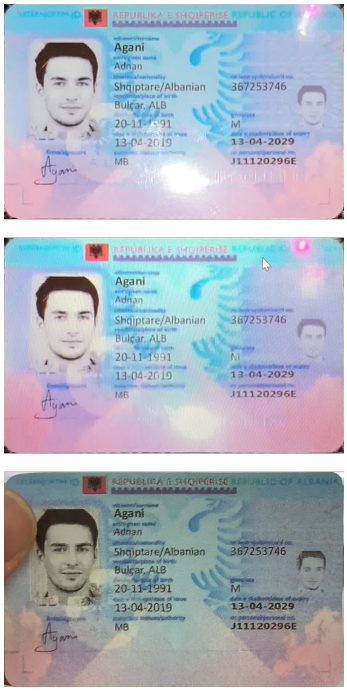
Abstract. *The increasing prevalence of digital identity verification has amplified the demand for robust personal document authentication systems. To obscure traces of forgery, forgers often photograph the documents after reprinting or directly capture them from a screen display. This paper is a work report for the First Competition on Presentation Attack Detection on ID Cards, held at the International Joint Conference on Biometrics 2024 (IJCB PAD-ID Card 2024). The competition aims to explore the efficacy of deep neural networks in detecting recapture attacks. The Document Liveness Challenge Dataset (DLC-2021) was utilized to train models. Several models were adapted for this task, including ViT, Xception, TResNet, and EVA. Among these, the Xception model achieved the best performance, showing a significantly low error rate in both attack presentation classification error and bona fide presentation classification error.*

1 Uvod

Zaradi digitalizacije poslovanja vse več storitev zahteva potrjevanje identitet. Ena izmed metod preverjanja identitete je slika osebnega dokumenta, ki pa ni nujno avtentična. Dokument je lahko spremenjen z ročno menjavo teksta ali slike, tako da se čez želeno vsebino prilepi novo, podobno pa je lahko storjeno tudi digitalno, in sicer z uporabo orodji kot je Adobe Photoshop. Uporabljajo se tudi metode, ki ta proces avtomatizirajo z uporabo različnih metod, npr. globokega učenja. Ponarejene dokumente, se potem pogosto tudi natisne ali slika na ekranu, ker je s tem mogoče prikriti veliko artefaktov, ki se pojavijo med ponarejanjem. Takšnim napadom pravimo napadi ponovnega zajema in so trenutno ena izmed ključnih prestreznih točk za ponarejene dokumente.

V okviru tekmovanja IJCB PAD-ID Card 2024 [22] smo se osredotočili na preizkušanje različnih naprednih globokih nevronskih mrež za prepoznavanje tovrstnih napadov. Na tem področju so bile predlagane številne rešitve, ki pa imajo zaradi omejenega dostopa do podatkovnih zbirk (le-te so zaradi varovanja osebnih podatkov velikokrat zaščitene) odprtih še veliko izzivov. V tem prispevku se bomo osredotočili na pregled področja, opis tekmovanja ter predstavili rezultate, ki smo jih dosegel z našimi modeli.

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



Slika 1: Primer originalnega dokumenta (zgornji) in ponovno zajetih dokumentov: ekran (srednji) in tisk (spodnji)

2 Tekmovanje PAD ID Card 2024

Trenutno je eden izmed največjih problemov pri raziskavah, ki se ukvarjajo z zaznavanjem ponarejenih dokumentov, pomanjkanje javno dostopnih podatkovnih množic, ki bi omogočale preverljivost rezultatov. Cilj tekmovanja je bil torej narediti temeljit pregled sodobnih metod zaznavanja ponarejenih osebnih dokumentov in ponuditi evaluacijski protokol, ki bo omogočil raziskovalcem, da preverijo svoje rezultate. Po zaključku tekmovanja naj bi torej vedeli, kateri modeli so se izkazali za najuspešnejše na podlagi nepristranske evaluacije.

Organizatorji tekmovanja niso ponudili nobene podatkovne množice, na kateri bi lahko delali, ampak so dovolili, da se za namene tekmovanja uporablja katera koli javno ali privatno dostopna podatkovna množica. Omenili so nekaj podatkovnih množic [1] [2] in metodo za generiranje slik osebnih izkaznic [3].

Kot osnovni metriki uspešnosti sta uporabljeni metriki APCER (Attack Presentation Classification Error Rate) in BPCER (Bonafide Presentation Classification Error Rate). Metriki sta definirani na naslednji način:

$$APCER_{PAIS} = 1 - \frac{1}{N_{PAIS}} \sum_{i=1}^{N_{PAIS}} RES_i \quad (1)$$

in

$$BPCER_{PAIS} = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}}, \quad (2)$$

kjer N_{PAIS} predstavlja število ponaredkov v množici (angl. presentation attack instruments), N_{BF} predstavlja število pravih primerkov (angl. bona fide) in RES_i predstavlja, kako je bil klasificiran i-ti primer. RES_i je 1, če je i-ti primer klasificiran kot ponaredek in 0, če je klasificiran kot resničen. Organizatorji tekmovanja bodo v članku, ki bo objavljen v IJCB, natančneje opisali rezultate in prispevke, ki so jih dobili med tekmovanjem.

3 Pregled literature

3.1 Biometrična avtentikacija

Napredek v sistemih za oddaljeno biometrično avtentikacijo je odprl nove poti za varno preverjanje identitete, zlasti prek prstnih odtisov in prepoznavanja obraza. Vendar pa zaznavanje ponovnega zajema dokumentov in zagotavljanje avtentičnosti dokumentov ostaja odprt izziv, saj različne vrste slik zahtevajo edinstvene pristope.

Albiero s sod. [4] je razvil tehniko za ujemanje sebkov (angl. selfie) s fotografijami na čilskih osebnih izkaznicah med adolescenco. Bulan in Sharma [5] sta predstavila kriptografski sistem za avtentikacijo natisnjenih slik z vgradnjo in kasnejšim pridobivanjem šifrirane sličice. Perera in Patel [6] sta predlagala sisteme aktivne avtentikacije, ki nenehno spremljajo identiteto uporabnika. Shi in Jain [7, 8] sta razvila DocFace in DocFace+, ki uporabljata prenosno učenje za ujemanje fotografij na osebnih dokumentih s sebki. Kljub svoji učinkovitosti ti biometrični sistemi običajno zahtevajo predhodno registracijo uporabnikov, kar omejuje njihovo uporabo pri dejavnostih, kot je odpiranje novih bančnih računov.

3.2 Zaznavanje ponovnega zajema dokumentov

Raziskave, specifično usmerjene na ponovno zajete dokumentne slike, so omejene, vendar naraščajo. Shang s sod. [9] je razvil značilke za razlikovanje dokumentov, proizvedenih z laserskimi/inkjet tiskalniki in kopirnimi stroji, s preko 90% natančnostjo klasifikacije. Berenguel s sod. [10] je uporabil ponavljajoči se primerjalnik z mehanizmom pozornosti za zaznavanje ponarejenih bankovcev in osebnih dokumentov za operacijo ponovnega zajema. Gonzalez in Valenzuela [11] sta raziskovala, ali je slika osebnega dokumenta, ki jo uporabnik posreduje na daljavo, pristna ali ponarejena. Predlagana metoda BasicNet z diskretno Fourierevo transformacijo dosega visoko stopnjo klasifikacije; 97,5% za pristne in 96,8% za ponarejene slike osebnih dokumentov. Hu s sod. [19] je obravnaval zaznavanja ponovno zajetih slik dokumentov z analizo popačenja rastra (angl. helftone cells), to so majhne pike, ki sestavljajo natisnjene slike ali dokumente, v pristnih in ponovno zajetih dokumentih.

Obstoječe metode za zaznavanje ponovnega zajema dokumentov se pogosto soočajo s praktičnimi težavami zaradi variacij v napravah za tiskanje/slikanje, podlagah, kanalih za ponovni zajem in vrstah dokumentov. Potrebne so nadaljnje raziskave za razvoj robustnih tehnik, ki lahko učinkovito obravnavajo te variacije.

4 Metodologija

4.1 Opis podatkovne množice

Za učenje svojih modelov smo uporabili podatkovno množico Document Liveness Challenge Dataset (DLC-2021) [1], ki je za svojo osnovo uporabila že obstoječo podatkovno množico MIDV-2020 [13]. Družina podatkovnih množic MIDV uporablja natisnjene ponarejene dokumente. MIDV-500 [2] je prva izmed teh množic in vključuje 500 video posnetkov 50 različnih osebnih dokumentov, predvsem vzorcev iz WikiMedia pod javnimi licencami.

MIDV-2019 [12] je razširil MIDV-500 z vključitvijo video posnetkov z zelo slabimi svetlobnimi pogoji in večjimi projekcijskimi popačenji. Ta podatkovna množica je dodala tudi fotografije in skenirane slike istih vrst dokumentov. MIDV-2020 je nadalje izboljšal podatkovno množico z uvedbo variabilnosti v besedilnih poljih, obrazih in podpisih, hkrati pa ohranil realizem. Sestavlja ga 1000 različnih fizičnih dokumentov in vključuje fotografije, skenirane slike in video posnetke.

Podatkovna množica DLC-2021 je uporabila 10 vrst osebnih dokumentov iz MIDV-2020, pri čemer je za vsako vrsto izbranih osem primerov (primer na Sliki 1). Ti dokumenti so bili natisnjeni v barvni in sivi lestvici brez laminacije (primerjava med zgornjim in spodnjim primerom na sliki 1), obrezani pa so bili tako, da so ustrezali originalnim oblikam. Videoposnetki so bili posneti z Apple iPhone XR in Samsung S10 v dveh ločljivostih (1080×1920 in 2160×3840 slikovnih elementov) ter hitrostih sličic (30 in 60 sličic na sekundo).

Za povečanje realističnosti so bili videoposnetki posneti pod različnimi svetlobnimi pogoji in so vključevali ovire, kot so prsti ali svetli predmeti. Ponovni zajem zaslona je vključeval dva LCD monitorja in dva prenosnika, s teksturiranimi ozadji, ki so povzročili Moirjeve vzorce (razvidno na primerjavi med zgornjim in srednjim primerom na sliki 1) in artefakte ponovnega zajema.

4.2 Modeli in zaznavanje

Za zaznavanje ponovnih zajemov slik osebnih dokumentov smo uporabili modele ViT (Vision Transformer) [14], Xception [15], TResNet [16] in EVA-02 [17], ki smo jih prilagodili (angl. fine-tunning) za ločevanje ponaredkov od resničnih primerov. Modeli so bili predhodno naučeni

ločevati slike na podatkovni množici ImageNet-1k [18] in nato prilagojeni na specifičen nabor podatkov, ki je vseboval slike pristnih in ponovno zajetih osebnih dokumentov. Iz podatkovne množice DLC-2021 smo vzeli ponovne zajeme barvno natisnjenih dokumentov in osebnih dokumentov na ekranih, ki so predstavljali pozitivne primere za učenje, resnični primeri iz DLC-2021 pa so predstavljali negativne primere. V podatkovni množici je podan nabor sličic za vsak posnetek in z njimi tudi podatki o lokaciji osebne izkaznice na njih. Iz vsake sličice je bila izrezana osebna izkaznica in potem vodoravno poravnana. Za namene tekmovanja so organizatorji predpripravili slike, tako da so uporabili algoritem za zaznavanje in obrezovanje osebnih izkaznic.

Ker je tekmovanje preverjalo tudi zaznavanje kompozitnih napadov smo za tekmovanje uporabili ansambel dveh algoritmov. Algoritem za kompozitne napade, ki ga je razvil L. Alessio s sod. [24], najprej obreže slike, da robovi dokumenta in ozadje ne vplivajo na analizo. Sledi detekcija teksta in obraza, okoli katerih postavi pravokotnike, ter iskalni pravokotnik okoli obraza za iskanje nenavadnih črt. Nato prepozna črte na dokumentu in odstrani tiste, ki se dotikajo pravokotnikov okoli obraza in teksta ali niso znotraj iskalnega pravokotnika. Če po tem ostane dovolj dolžine nenavadnih črt, dokument označimo kot ponaredek, sicer kot original. Da iz obeh algoritmov dobimo končno oceno vzamemo maksimum obeh algoritmov.

5 Eksperimenti

5.1 Tehnične podrobnosti

Prednaučeni modeli so bili vzeti iz knjižnice timm[1] (tabela 1), kjer smo pri vsakem modelu zamenjal zadnji nivo z linearnim, ki vrača eno samo vrednost.

Tabela 1: Modeli uporabljeni za zaznavanje

Model	implementacija v timm[1]
Xception	xception65.ra3_in1k
Vit	vit_base_patch16_224.augreg2_in21k_ft_in1k
TResNet	tresnet_m.miil_in21k_ft_in1k
EVA-02	eva02_small_patch14_336.mim_in22k_ft_in1k

Vsak model je bil učen za pet dodatnih epohov, z binarno križno entropijo (angl. binary cross entropy), ki predhodno uporabi sigmoidno funkcijo (PyTorch: BCEWithLogitsLoss) [21]. Po petih epohah je bil izbran model, ki je na validacijski množici dosegel najmanjšo napako. S pomočjo ROC krivulje dobljene iz validacijske množice je bila izračunana meja za klasifikacijo, ki zadošča:

$$threshhold = \max(TPR - FPR). \tag{3}$$

Kot optimizator smo uporabljali AdamW z $lr = 0.0001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ in $\epsilon = 10^{-8}$ ter $weight\_decay = 0.01$, $batch\_size = 32$ za Xception in $batch\_size = 256$ za ostale modele. Učili in testirali smo na Arnesovem superračunalniku [2].

Za razvoj algoritma smo iz podatkovne množice DLC-2021, ki vsebuje pet osebnih izkaznic in pet potnih listov iz različnih držav, izbrali validacijsko, testno in učno množico. Validacijska in testna množica vsaka vsebujeta en tip osebne izkaznice in en tip potnega lista, pri čemer validacijska množica vsebuje 12.975 slik, testna množica pa 12.943 slik. Preostalih 45.207 dokumentov smo razdelili v učno množico.

5.2 Evalvacija

Za evalvacijo med testiranjem sta bili uporabljeni metriki $BPCER$ in $APCER$. Pri rezultatih tekmovanja pa so poleg teh metrik uporabili še EER (angl. Equal Error Rate). Vrednosti za $BPCER$ pa pri definiranih vrednostih metrike $APCER$ in sicer $BPCER10$ ($APCER = 10\%$), $BPCER20$ ($APCER = 5\%$) in $BPCER100$ ($APCER = 1\%$). Vrednost EER je definirana kot točka na ROC krivulji, kjer velja: $APCER = BPCER$. Glavna metrika za tekmovanje pa je bila vrednost $AV\text{-}Rank$ definirana kot:

$$\begin{aligned} AVRank = BPCER10 \cdot 0.2 + \\ BPCER20 \cdot 0.3 + BPCER100 \cdot 0.5. \end{aligned} \tag{4}$$

Tekmovanje je preverjalo tudi detekcijo kompozitnih napadov, ki pa jih naš pristop ne zazna. Končni algoritem je bil sestavljen iz dveh ločeno razvitih algoritmov, ki skupaj omogočata zaznavanje kompozitnih napadov [24] in napadov s ponovnim zajemom.

6 Rezultati

Iz rezultatov dobljenih na podatkovni množici DLC-2021 (tabela 2) vidimo, da transformerja ViT in EVA-02 nista dosegla dobrih rezultatov, medtem ko sta konvolucijski nevronski mreži Xception in TResNet dosegli dosti boljše rezultate. Ker je Xception dosegel precej boljši APCER kot TResNet in sta oba modela imela primerljiv BPCER, smo za končni model izbrali Xception.

Tabela 2: Rezultati, ki so jih izbrani modeli dosegli na podatkovni množici DLC-2021

Model	APCER(↓)	BPCER(↓)
Xception	0.0600	0.1195
Vit	0.1159	0.7311
TResNet	0.1005	0.1093
Eva-02	0.2098	0.4501

Rezultati, ki jih je dosegel naš končni model na tekmovanju IJCB PAD-ID Card 2024 (v tabeli 3 označeno kot FRIFE) so bili precej slabši v primerjavi z rezultati na naši testni množici, vendar se, tako kot vsi ostali algoritmi, ni približal izhodiščnim algoritmom, ki so jih navedli organizatorji (tabela 4). Moramo pa poudariti, da za razliko od organizatorjev, udeleženci tekmovanja

[1] PyTorch Image Models: https://github.com/rwightman/pytorch-image-models

[2] Arnesova računska gruča: https://www.sling.si/arnesova-racunska-gruca/

Tabela 3: Rezultati, ki so jih ekipe dosegle na tekmovanju IJCB PAD-ID Card 2024, nižje vrednosti so boljše.

Ime ekipe	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	AVRank(%)
Anonymus_V1	21.87	46.06	65.82	90.70	74.30
Anonymus_V2	29.01	63.36	76.82	92.22	81.82
FRIFE	44.09	87.96	93.06	99.92	95.47
IDVC_V1	22.96	65.40	74.60	84.38	77.65
IDVC_V2	25.91	66.10	74.42	86.16	78.62
SecureID	50.63	90.94	95.42	99.42	96.52

Tabela 4: Rezultati, ki so jih dosegli izhodiščni algoritmi organizatorjev tekmovanja IJCB PAD-ID Card 2024, nižje vrednosti so boljše.

Ime ekipe	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	AVRank (%)
Baseline1	**4.58**	1.84	4.20	14.96	9.10
Baseline2	7.17	5.26	9.78	24.40	16.18
Baseline3	9.02	8.14	13.28	28.58	19.90

nismo imeli vpogleda v podatke testne in učne podatkovne množice. Kot vidimo so najboljši algoritmi vseeno dosegli dobre rezultate, a prostora za izboljšave je še veliko. Organizatorji tekmovanja so iz rezultatov sklepali, da je ključna razlika, ki ločuje boljše algoritme od slabših, da so bili boljši algoritmi učeni na podatkovnih množicah, ki so vsebovale več osebnih izkaznic z različnimi obrazi. Vsaka osebna izkaznica v množici pa je bila predstavljena manjkrat [25]. Poleg tega pa za testiranje in učenje našega algoritma nismo imeli dostopa do podatkovne množice, ki bi vsebovala tudi primere kompozitnih napadov. Najverjetneje se je končni ansambel sestavljen iz modela opisanega v tem članku in modela za zaznavanje kompozitnih napadov [24] motil ravno na teh primerih.

7 Zaključek

Naši rezultati so pokazali, da je Xception dober model za prepoznavanje napadov ponovnega zajema, vendar ti rezultati niso bili replicirani na tekmovanju IJCB PAD-ID Card 2024, kjer pa tudi drugi algoritmi niso prišli blizu izhodiščnemu algoritmu. Rezultati, ki so jih navedli organizatorji tekmovanja pa nakazujejo na veliko potencialnih izboljšav. Prva ideja za izboljšavo našega pristopa bi bili razširitev podatkovne množice s primeri kompozitnih napadov, tako da bi jih lahko upoštevali med učenjem in da bi lahko analizirali obnašanje naših algoritmov tudi na teh primerih. Druga ideja pa bi bila delitev algoritma na dva posamezna algoritma, kjer bi vsak upošteval strokovno znanje s svojega področja. Algoritem za zaznavanje printanih ponovnih zajemov bi lahko upošteval razliko v odsevnosti med papirjem in med laminiranimi osebnimi dokumenti. Algoritem za zaznavanje ponovnih zajemov na ekranu pa bi lahko zaznaval pristonost Moirjevih vzorcev [20], ki se pojavijo pri slikah ekranov.

Literatura

[1] D. V. Polevoy, et al.: Document Liveness Challenge Dataset (DLC-2021), Journal of Imaging, 2022.

[2] V. V. Arlazarov, K. Bulatov, T. Chernov, V. L. Arlazarov: MIDV-500: A dataset for identity documents analysis and recognition on mobile devices in video stream, Computer Optics, 2019.

[3] D. Benalcazar, J. E. Tapia, S. Gonzalez, C. Busch: Synthetic ID Card Image Generation for Improving Presentation Attack Detection, IEEE Transactions on Information Forensics and Security, 2022.

[4] V. Albiero, et al.: Identity document to selfie face matching across adolescence, IJCB, 2021, Houston, Teksas, ZDA.

[5] O. Bulan, G. Sharma: Content authentication for printed images utilizing high capacity data hiding, Journal of Electronic Imaging, 2013.

[6] P. Perera, V. M. Patel: Face-based multiple user active authentication on mobile devices, IEEE Transactions on Information Forensics and Security, 2019.

[7] Y. Shi, A. K. Jain: DocFace: Matching ID document photos to selfies, BTAS, 2019, Redondo Beach, Kalifornija, ZDA.

[8] Y. Shi, A. K. Jain: DocFace+: ID document to selfie matching, IEEE Transactions on Biometrics, Behaviour and Identity Science, 2018.

[9] S. Shang, N. Memon, X. Kong: Detecting documents forged by printing and copying, EURASIP Journal on Advances in Signal Processing, 2014.

[10] A. B. Centeno, O. R. Terrades, J. L. Canet, C. C. Morales: Recurrent comparator with attention models to detect counterfeit documents, ICDAR, 2019.

[11] S. Gonzalez, A. Valenzuela, J. Tapia: Hybrid Two-Stage Architecture for Tampering Detection of Chipless ID Cards, IEEE Transactions on Biometrics, Behavior, and Identity Science, 2020.

[12] K. Bulatov, D. Matalov, V. V. Arlazarov: MIDV-2019: Challenges of the modern mobile-based document OCR, ICMV 2019, 2020.

[13] K. Bulatov, et al.: MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis, Computer Optics, 2021.

[14] A. Dosovitskiy, et al.: An image is worth 16x16 words: Transformers for image recognition at scale, ICLR, 2021.

[15] F. Chollet: Xception: Deep learning with depthwise separable convolutions, CVPR, 2017.

[16] T. Ridnik, et al.: TResNet: High performance GPU-dedicated architecture, WACV, 2021.

[17] Y. Fang, et al.: EVA-02: A Visual Representation for Neon Genesis, arXiv, 2023.

[18] O. Russakovsky, et al.: ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 2015.

[19] Z. Hu, C. Chen, W. H. Mow, J. Huang: Document Recapture Detection Based on a Unified Distortion Model of Halftone Cells, IEEE Transactions on Information Forensics and Security, 2021.

[20] E. Abraham: Moiré Pattern Detection using Wavelet Decomposition and Convolutional Neural Network, SSCI, 2018, Banaglore, Indija.

[21] BCEWithLogitsLoss: https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html

[22] IJCB PAD ID Card 2024: https://sites.google.com/view/ijcb-pad-id-card-2024/home?authuser=0

[23] IJCB PAD ID Card 2024 - Evaluation Criteria: https://sites.google.com/view/ijcb-pad-id-card-2024/evaluation-criteria?authuser=0

[24] L. Alessio, M. Brodarič, P. Peer, V. Štruc, B. Batagelj: Prepoznava zamenjave obraza na slikah osebnih dokumentov, ERK, 2024, Portorož, Slovenija (v recenziji).

[25] Tapia, Juan E., et al. "First Competition on Presentation Attack Detection on ID Card."arXiv preprint arXiv:2409.00372 (2024).