ELEKTROTEHNIŠKI VESTNIK 91(1-2): 1–7, 2024
IZVIRNI ZNANSTVENI ČLANEK

Deidentifikacija obrazov z nadzorom ravni varovanja zasebnosti

Blaž Meden

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, Slovenija
E-pošta: blaz.meden@fri.uni-lj.si

Povzetek. Zaščita zasebnosti je v današnji digitalni dobi ena izmed ključnih perečih tem. Pri tem so še zlasti občutljive slikovne podobe obrazov, ki običajno ne razkrivajo le identitete osebe, temveč tudi druge osebne karakteristike. Za reševanje tega problema so bile v literaturi predstavljene različne tehnike deidentifikacije obrazov. Te poskušajo odstraniti ali zakriti osebne informacije na izvornih obraznih slikah, hkrati pa ohraniti njihovo uporabnost za nadaljnjo analizo. Čeprav je bilo na področju deidentifikacije obrazov že veliko predlaganih rešitev, večina najnovejših še vedno deluje pomanjkljivo, saj te tehnike pogosto (a) deidentificirajo le ozko področje obraza, s čimer pomembne kontekstualne informacije ostanejo nezaščitene, (b) toliko spreminjajo slikovne podobe obraza, da trpita naraven videz in raznolikost obrazov v deidentificiranih rezultatih, (c) ne ponujajo prožnosti pri ravni zagotovljene zasebnosti, kar vodi v suboptimalno uporabo v različnih aplikacijah, in (d) pogosto ponujajo nezadovoljiv kompromis med zmožnostjo zakrivanja informacij o identiteti, kakovostjo in naravnostjo deidentificiranih slik ter zadostnim ohranjanjem uporabnosti deidentificiranih podatkov. V tem članku obravnavamo te pomanjkljivosti z novim nadzorovanim postopkom za deidentifikacijo obrazov, ki uravnoteži kakovost slike, zaščito identitete in uporabnost podatkov za nadaljnjo analizo. Predlagani pristop uporablja zmogljiv generativni model (StyleGAN2), več pomožnih klasifikacijskih modelov in skrbno oblikovane omejitve oz. kriterije za usmerjanje optimizacijskega procesa deidentifikacije. Pristop je preverjen s 4 različnimi nabori podatkov in primerjan s 7 konkurenčnimi pristopi.

Ključne besede: deidentifikacija obrazov, generativni modeli, slikovna biometrija, globoko učenje

Face Deidentification with a Controllable Privacy Protection

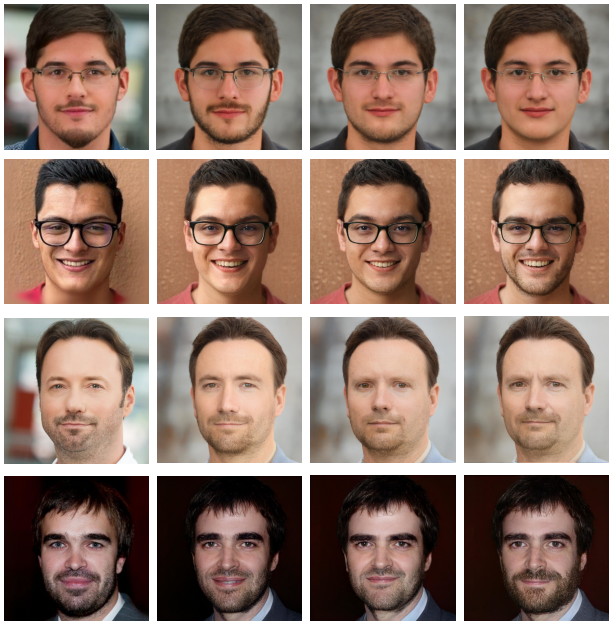
Privacy concerns in the digital era highlight the sensitivity of facial images, revealing personal information beyond the identity. The current face deidentification techniques aiming to solve the issue often suffer from limitations. They either focus on narrow facial areas, compromise the image naturalness, lack flexibility in privacy levels, or offer suboptimal trade-offs between the identity protection, image quality, and utility preservation. The paper presents a novel controllable face deidentification method that leverages StyleGAN2 and auxiliary classification approaches addressing these shortcomings. Validated across four datasets and compared to seven competitors, the method ensures a significant identity protection, preserves the data utility, maintains the diversity among deidentified faces, and demonstrates a promising performance.
Keywords: face deidentification, generative models, image biometrics, deep learning

1 UVOD

Pravica do zasebnosti je temeljna človekova pravica, ki posameznikom omogoča, da osebne podatke in dejavnosti ohranjajo zaščitene pred nepooblaščenim dostopom, nadzorom ali vmešavanjem drugih [1]. Pojem zasebnosti je priznan v mednarodnih aktih o človekovih pravicah in v veljavni zakonodaji [2] ter vključuje različne vidike posameznikovega življenja, ki zajemajo fizični prostor, komunikacijo, osebne podatke in posameznikovo identiteto [3]. Biometrični podatki, še zlasti slike obraza, so občutljivi osebni podatki, ki so težko spremenljivi, in so zato lažje tarča nepooblaščenega ravnanja. Uporaba biometrične tehnologije ima potencial za izboljšanje varnosti, hkrati pa prinaša tudi utemeljene skrbi glede ohranjanja zasebnosti [4]. Odgovorna uporaba biometrične tehnologije zahteva močne pravne okvire, transparentnost zbiranja in obdelave osebnih podatkov ter tudi informirano privolitev posameznikov [5]. V prispevku zato proučujemo tehnike za zaščito zasebnosti na slikah obraza z uporabo procesa deidentifikacije, da bi razvili praktične rešitve za izboljšanje zaščite zasebnosti ter obenem ohranjanje uporabnosti podatkov in naravnosti na zaščitenih obraznih slikah.

Postopek deidentifikacije obrazov si prizadeva izboljšati zasebnost slike obraza posameznika z modificiranjem ali odstranjevanjem prepoznavnih značilnosti. Namen deidentifikacije je nasprotovati tehnologijam za prepoznavanje obraza ter tako zavarovati zasebnost posameznikov v primerih, ko se zajemajo, delijo ali analizirajo obrazne slike [6], [7]. Deidentifikacija obraza vključuje zameglitev ali zamenjavo edinstvenih osebnih značilnosti obraza z generičnimi, kar pomeni, da obrazi

Prejet 19. februar, 2024
Odobren 20. marec, 2024



Slika 1: Predstavljen je nov pristop deidentifikacije obraza z nadzorovano ravnjo zagotovljene zasebnosti. Slika prikazuje vpliv povečanja stopnje zaščite zasebnosti (od leve proti desni). Medtem ko se prvotne identitete postopoma spreminjajo, splošna vizualna kakovost in videz ostajata primerljiva z izvirnimi slikami iz levega stolpca. Ugotavljamo, da vedno obstaja kompromis med količino ohranjene podobnosti slike in količino dosegljive zaščite zasebnosti. Konkurenčne rešitve običajno zagotavljajo fiksen kompromis, naš pristop pa omogoča prilagajanje tega kompromisa prek nastavljivega parametra.

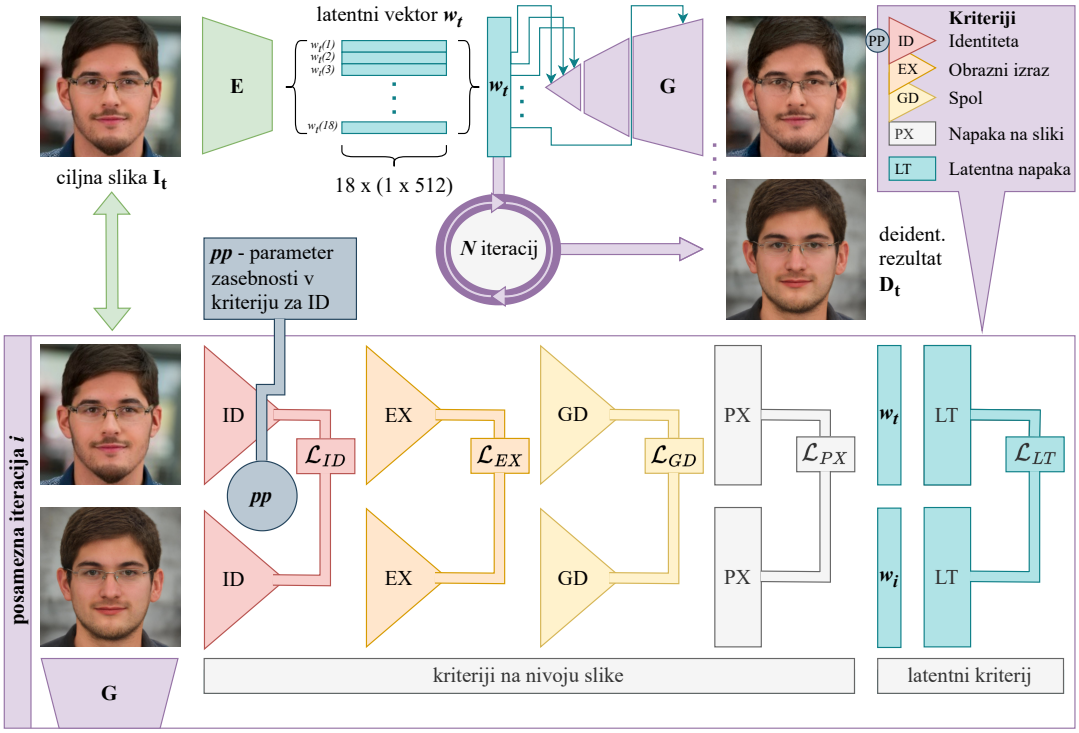
postanejo nerazpoznavni [8]. Ta ukrep za varovanje zasebnosti preprečuje avtomatiziranim razpoznavalnim sistemom in nepooblaščenim človeškim opazovalcem, da bi deidentificirane obrazne slike povezali z izvornimi posamezniki. Zgodnji pristopi za deidentifikacijo obrazov so tipično uporabljali tehnike maskiranja [9] ali filtriranja [10] na celotnem področju slike obraza, vendar z varnostnimi ranljivostmi (npr. ranljivost za napad s posnemanjem [11]) ali pa preveč restriktivnosti pri ohranjanju uporabnosti podatkov. Z razvojem modelov za detekcijo ključnih obraznih točk (angl. landmark detection) [12] in aktivnih modelov videza (angl. Active Appearance Models) [13] so se procesi deidentifikacije obrazov osredotočili zgolj na ožje področje obraza. Kljub razvoju, tudi ti pristopi niso rešili vseh problemov, saj se na tem področju še vedno pojavljajo zahteve po zagotavljanju zasebnosti [11], vzpostavljanju merljive zasebnosti [14], zagotavljanju ustrezne ravni uporabnosti podatkov [15], [16] ter splošne kakovosti obdelanih obraznih slik [17].

Čeprav se obstoječe tehnike deidentifikacije, ki temeljijo na globokem učenju, pogosto ponašajo z impresivno učinkovitostjo, so v nekaterih vidikih še vedno omejene, saj: (i) se v procesu običajno osredotočajo le na ožje območje obraza (ki ustreza zamenjavi obraza), kar lahko vodi do (potencialno) neoptimalnih rezultatov deidentifikacije v obliki nedotaknjene vsebine, bogate z relevantnimi informacijami (npr. oblika ali barva las, oblika obraza, uhljev ipd.), (ii) dosegajo fiksen in vnaprej določen kompromis med zagotovljeno zaščito zasebnosti, ohranitvijo uporabnosti in raznolikostjo, medtem ko pogosto pridelajo vizualne artefakte, ki vplivajo na naravni videz deidentificiranih podatkov, ter (iii) ne uporabljajo dodatnih mehanizmov za nadzor ciljne ravni ohranjene zasebnosti pri deidentifikaciji.

Da bi odpravili te omejitve, v tem prispevku predstavljamo novo, najsodobnejšo rešitev za deidentifikacijo obrazov z doseganjem nadzorovane ravni zaščite zasebnosti obraza, ki ponuja prilagodljiv kompromis med ohranjanjem uporabnosti podatkov in zaščito zasebnosti ter ustvarja raznolike in visokokakovostne slike v smislu naravnosti in realizma, kot je prikazano v primerih na sliki 1. Naša rešitev temelji na tehniki inverzije GAN [18] in uporablja namenski postopek optimizacije, uporabljen v latentnem prostoru zmogljivega modela StyleGAN2 [19]. Za nadzor postopka optimizacije uvajamo novo več-ciljno izgubno funkcijo (angl. multi-objective loss function), ki se uporablja za skupno optimizacijo semantike na sliki, kar vključuje kontekstualne kriterije za zatiranje obrazne identitete ter ohranjanje uporabnosti atributov (tj. spola in obrazne mimike v naših študijah primerov). Te kriterije kombiniramo s preostalimi pogoji izgubne funkcije za izdelavo visokokakovostnih deidentificiranih obraznih slik. V nasprotju z obstoječimi modeli naša zasnovana kriterijska funkcija generativnemu modelu omogoča, da vzpostavi ravnotežje med različnimi zahtevami, ključnimi za postopek deidentifikacije, ki vključujejo: zakrivanje identitete, ohranjanje uporabnosti podatkov, raznolikost obrazov in realističen videz. Predlagano rešitev temeljito ovrednotimo z raznovrstnimi nabori podatkov s slikami obrazov (kar vključuje nabore slikovnih baz, zajetih v studiu in v naravnem okolju) in jo primerjamo s konkurenčnimi metodami, pri tem pa pridobimo zelo spodbudne rezultate.

2 Algoritem CPP-DeID

Splošno uporaben deidentifikacijski algoritem za zaščito zasebnosti bi moral zagotoviti prilagodljive ravni zaščite za deidentifikacijo podob obraza za nadzor kompromisa med ohranjeno zasebnostjo in uporabnostjo obrazne slike. Naš glavni prispevek v tem poglavju je nova tehnika deidentifikacije obraza, ki izpolnjuje te zahteve. Tehnika uporablja postopek optimizacije v latentnem prostoru generatorja StyleGAN2 za nadzor kompromisa med različnimi vidiki procesa deidentifikacije. Osrednja ideja predlaganega pristopa za *deidentifikacijo obrazov z nadzorovano zaščito zasebnosti* (angl. Controllable Privacy Protection DeIdentification; CPP-DeID) je prikazana na sliki 2.



Slika 2: Osnovna ideja predlaganega postopka deidentifikacije CPP-DeID. Vhodna slika $\mathbf{I}_t$ je najprej kodirana v razširjeni latentni prostor W^+ prednaučenega modela StyleGAN2 kot $\mathbf{w}_t$, kjer je E model za kodiranje (kodirnik) in G generativni model StyleGAN2 (generator). Latentna koda $\mathbf{w}_t$ je zatem optimizirana prek več kriterijev (t.j. kriterijska funkcija identitete $\mathcal{L}_{ID}$, kriterij za izraz obraza $\mathcal{L}_{EX}$, kriterij atributa spola $\mathcal{L}_{GD}$, rekonstrukcijski kriterij $\mathcal{L}_{PX}$ in latentni kriterij oz. kriterij v prikritem prostoru $\mathcal{L}_{LT}$). Kriteriji se izračunajo na različnih interpretacijah vhodne in izhodne slike. Skupek kriterijev se optimizira v N iteracijah (do konvergence), rezultat pa je nadzorovana vizualna sprememba obraza iz vhodne slike $\mathbf{I}_t$. Hkrati je kriterijska funkcija za identiteto dodatno opremljena z nastavljivim parametrom zasebnosti pp, ki omogoča natančen nadzor nad količino ohranjene identitete v deidentificiranih izhodnih slikah.

Predlagani algoritem deluje v dveh korakih, kar zajema (i) korak *predprocesiranja* in (ii) korak *optimizacije (deidentifikacije)*. V koraku predprocesiranja se na vhodni sliki najprej izvede postopek zaznavanja obraza, sledijo poravnava, obrezovanje in projekcija v latentni prostor StyleGAN. Rezultat tega koraka je latentna predstavitev (ali prikrita koda), ki ustreza vhodni sliki obraza. V drugem koraku (fazi deidentifikacije) se nad pridobljeno latentno kodo izvede postopek optimizacije za določeno število iteracij N, nato pa se optimizirana latentna koda ponovno pretvori nazaj v slikovni prostor prek generatorja StyleGAN.

Optimizacija slike je pogojena z različnimi kriteriji (implementiranimi z odvedljivimi modeli globokega učenja), ki poskušajo odstraniti ali ohraniti določene vizualne značilnosti na izvorni sliki. Kriterij, ki se osredotoča na *zatiranje identitete*, usmerja postopek optimizacije k odstranitvi informacij o identiteti iz obstoječe latentne kode. Kriteriji za *ohranjanje podatkov* si na podoben način prizadevajo ohraniti določene obrazne atribute v tej latentni reprezentaciji (v našem primeru izraz na obrazu in atribut spola, vendar je mogoče dodati še druge atribute). Našteti kriteriji so hkrati dodatno podprti s splošno namenskimi kriteriji za *ohranjanje semantične vsebine*, ki ocenjujejo podobnost med izvorno in izdelano sliko na nivoju slikovnih pik in na nivoju latentnega prostora, da se ohrani čim več splošne semantike iz izvorne slike.

2.1 Predprocesiranje

V koraku predprocesiranja je podana ciljna slika $\mathbf{I}_t$ najprej poravnana, obrezana in nato kodirana v razširjeni latentni prostor StyleGAN2 W^+ [20] z uporabo predhodno usposobljenega kodirnika za urejanje slik (E4E), ki sta ga predlagala Tov in Alaluf [21], [22]. Z uporabo kodirnika E4E kot E lahko postopek kodiranja slike v latentno reprezentacijo opišemo na naslednji način:

$$\mathbf{w}_t = E(\mathbf{I}_t); \mathbf{w}_t \in \mathbb{R}^{18\times512},$$

kjer je $\mathbf{w}_t$ ustvarjena latentna koda/predstavitev, ki ustreza $\mathbf{I}_t$. Ustvarjeni latentni vektor $\mathbf{w}_t$ se nato uporabi poleg poravnane in obrezane ciljne slike $\mathbf{I}_t$ v drugi (optimizacijski) fazi algoritma CPP-DeID.

2.2 Optimizacija

V tem koraku optimizacije poskuša CPP-DeID optimizirati več delno kriterijsko funkcijo za izdelavo

deidentificiranih slik naravnega videza. Vhod v postopek optimizacije predstavljata slika $\mathbf{I}_t$ in njej pripadajoča latentna koda $\mathbf{w}_t$. Celotna kriterijska funkcija, ki se med postopkom deidentifikacije optimizira za vsako vhodno sliko posebej, je definirana kot linearna kombinacija kriterijev naslednje oblike:

$$\mathcal{L} = \mathcal{L}_{PX} + \lambda_{LT} \cdot \mathcal{L}_{LT} + \lambda_{ID} \cdot \mathcal{L}_{ID} + \\ + \lambda_{GD} \cdot \mathcal{L}_{GD} + \lambda_{EX} \cdot \mathcal{L}_{EX}, \quad (1)$$

kjer so $\lambda_{LT}, \lambda_{ID}, \lambda_{GD}$ in λ_{EX} izravnalne uteži. Kot je razvidno, je cilj optimizacije sestavljen iz več kriterijev, namenjenih nadzoru različnih semantičnih vidikov, ki so ključni v procesu deidentifikacije.

Kriterij na ravni slike ($\mathcal{L}_{PX}$) spodbuja postopek optimizacije, da obdrži/ohrani čim več izvirne semantike slike. Cilj tega kriterija je opredeljen kot povprečna kvadratna napaka (MSE) med vhodno sliko $\mathbf{I}_t$ in sliko $\mathbf{I}_g$, ustvarjeno iz latentne kode $\mathbf{w}_t^{(g)}$ v trenutni iteraciji optimizacije ($\mathbf{I}_g = G(\mathbf{w}_t^{(g)})$), tako da velja:

$$\mathcal{L}_{PX}(\mathbf{I}_t, \mathbf{I}_g) = \frac{1}{n_{PX}} \sum_{(x,y)} (\mathbf{I}_t(x,y) - \mathbf{I}_g(x,y))^2, \quad (2)$$

kjer (x, y) predstavljajo koordinate slike in n_{PX} označuje skupno število slikovnih pik v $\mathbf{I}_t$. Z enako mero napake (MSE) je definiran tudi **kriterij na ravni latentnega prostora** ($\mathcal{L}_{LT}$), le da se ta izvede na nivoju latentnih reprezentacij.

Kriterij za zatiranje identitete ($\mathcal{L}_{ID}$) predstavlja glavno komponento postopka optimizacije za deidentifikacijo. Cilj te omejitve je maksimirati razdaljo med identitetnimi značilkami f_t vhodne slike $\mathbf{I}_t$ in značilkami f_g generirane slike $I_g = G(\mathbf{w}_t^*)$, ustvarjene v trenutni iteraciji optimizacijskega postopka. Za pridobivanje (luščenje) značilk identitete uporabimo izhod zadnje plasti najsodobnejšega modela za razpoznavo obrazov ArcFace ψ (na podlagi hrbtenične arhitekture ResNet50 in učne baze MS1M-V2), tako da je $f_t = \psi(\mathbf{I}_t)$ in $f_g = \psi(\mathbf{I}_g)$. Poleg tega ta kriterij vključuje ***parameter zasebnosti*** $pp \in [0, 1]$, ki nam definira interval možnih ciljnih razdalj med značilkami. Če je na primer $pp = 0.9$, cilj optimizacije konvergira k podobnosti identitete blizu 0.9 med vhodno in deidentificirano sliko. Za CPP-DeID definiramo kriterij $\mathcal{L}_{ID}$ kot sledi:

$$\mathcal{L}_{ID}(\mathbf{f}_t, \mathbf{f}_g, pp) = \left(pp - \frac{1}{n_{ID}} \sum (\mathbf{f}_t(i) - \mathbf{f}_g(i))^2 \right)^2; \\ \{\mathbf{f}_t, \mathbf{f}_g\} \in \mathbb{R}^{512}, \quad (3)$$

kjer je n_{ID} dimenzionalnost pridobljenih značilk identitete obraza iz modela ArcFace.

Nazadnje, za uveljavitev ciljno usmerjenega ohranjanja uporabnosti za določene atribute uporabljamo **dodatna kriterija ohranjanja uporabnosti**. Medtem ko je dodatne kriterije mogoče opredeliti na poljuben način, odvisno od želenih atributov, ki jih je treba ohranjati v deidentificiranih slikah $\mathbf{D}_t$, pri pristopu upoštevamo atribut *spola* in *izraz obraza*, da prikažemo ohranjanje obraznih značilnosti pri uporabi algoritma CPP-DeID. Za uresničitev ciljnih kriterijev optimizacije za ohranjanje spola in izraza na obrazu uporabljamo dva dodatna (odvedljiva) klasifikatorja, ρ in κ, da izluščimo atributa spola $\mathbf{g_t} = \rho(\mathbf{I}_g)$ in obraznega izraza $\mathbf{e_t} = \kappa(\mathbf{I}_g)$, ki temeljita na arhitekturah Resnet18 [23] in DAN [24].

Pridobljene značilke spola in izraza iz modelov ρ in κ ($\mathbf{g}_t, \mathbf{g}_g$ za spol, $\mathbf{e}_t, \mathbf{e}_g$ za izraz) uporabimo, da definiramo dva kriterija za ohranjanje uporabnosti, $\mathcal{L}_{GD}$ in $\mathcal{L}_{EX}$:

$$\mathcal{L}_{GD}(\mathbf{g}_t, \mathbf{g}_g) = \frac{1}{n_G} \sum_{i=0}^{n_G} (\mathbf{g}_t(i) - \mathbf{g}_g(i))^2; \\ \{\mathbf{g}_t, \mathbf{g}_g\} \in \mathbb{R}^{512}, \quad (4)$$

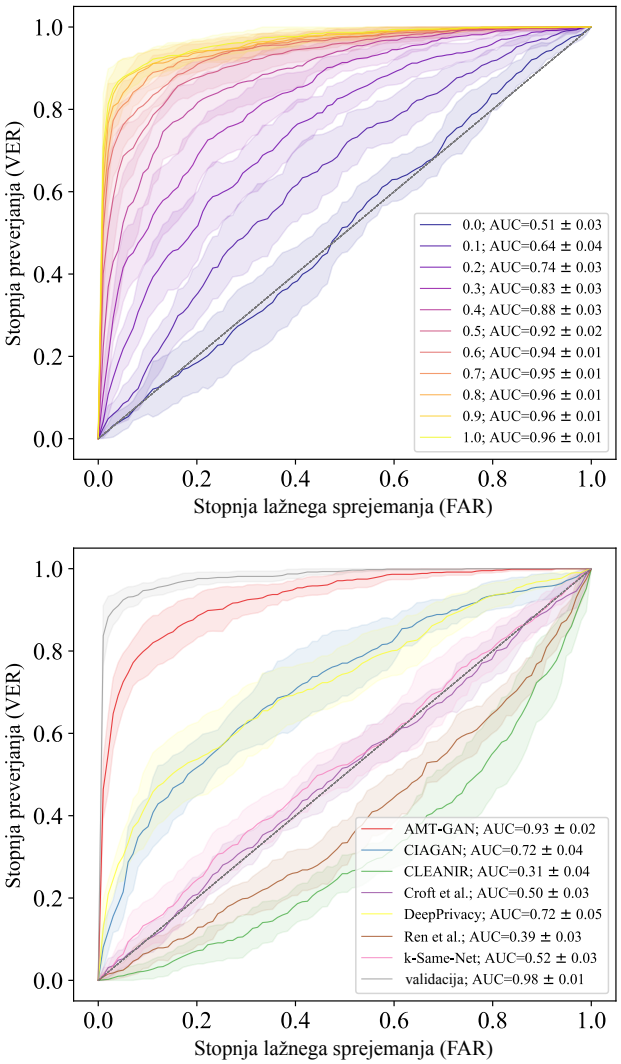
$$\mathcal{L}_{EX}(\mathbf{e}_t, \mathbf{e}_g) = \frac{1}{n_E} \sum_{i=0}^{n_E} (\mathbf{e}_t(i) - \mathbf{e}_g(i))^2; \\ \{\mathbf{e}_t, \mathbf{e}_g\} \in \mathbb{R}^{7 \times 7 \times 512}, \quad (5)$$

kjer n_G in n_E pomenita število elementov v predstavitvi spola oziroma izraza.

Uteži v enačbi (1), λ_{LT}, λ_{ID}, λ_{GD}, λ_{EX}, definirajo pomembnost optimizacijskih kriterijev, ki spreminjajo identiteto, spol in izraz obraza. Med fazo umerjanja pristopa CPP-DeID so bile opredeljene kot $\lambda_{LT} = 0.0016,\ \lambda_{ID} = 2.5,\ \lambda_{GD} = 0.01,\ \lambda_{EX} = 0.01$.

3 REZULTATI

V prvi seriji poskusov smo raziskali učinkovitost deidentifikacije CPP-DeID *s poskusi verifikacije* obrazov z uporabo prepoznavalnega modela VGG-Face [25]. Rezultati so pridobljeni na testnem delu slikovne baze Celeba-HQ [26], na podlagi katere smo pridobili 11 deidentificiranih množic slik z različnimi ravnmi zaščite zasebnosti. Z modelom VGG-Face smo na podlagi deidentificiranih množic pridobili mere podobnosti, ki smo jih nato prikazali z izračunom povprečnih performančnih krivulj (ROC) in ustreznih intervalov zaupanja. Da bi poročane rezultate predstavili v perspektivi, smo na enak način izvedli tudi deidentifikacijo s 7 konkurenčnimi pristopi deidentifikacije, ki služijo kot izhodiščni algoritmi (angl. baseline algorithms) za naše poskuse, kar vključuje AMT-GAN [27], CIAGAN [28], CLEANIR [29], rešitev Croft *et al.* [30], DeepPrivacy [31], tehniko Ren *et al.* [32] in naš predhodno predlagani pristop k-Same-Net [33]. Na levi strani slike 3 je mogoče opaziti, da učinkovitost razpoznavanja ob uporabi CPP-DeID pada v skladu s parametrom zasebnosti pp, čeprav poslabšanje zmogljivosti ni povsem linearno odvisno od pp. Iz ustvarjenih rezultatov na desni strani slike 3 vidimo, da preostali modeli zagotavljajo različne zmogljivosti, vendar imajo omejeno sposobnost nadzora nad



Slika 3: Verifikacijski eksperimenti na testni particiji baze slik Celeba-HQ v obliki krivulj ROC, pridobljenih z razpoznavalnim modelom VGG-Face. Rezultati so prikazani za naš predlagani pristop CPP-DeID z nastavljivim parametrom zasebnosti *pp* (zgoraj) in kompetitivnimi algoritmi (spodaj).

ohranjeno količino podatkov o identiteti, ki se odstranijo iz vhodnih slik.

Vrednotenje ravnovesja zasebnost-uporabnost-realizem-raznolikost Rezultati vrednotenja ravnovesja so povzeti na sliki 4, ki povzame celovito primerjavo med CPP-DeID in konkurenčnimi algoritmi na naborih slikovnih baz XM2VTS, RaFD in Celeba-HQ [34], [35], [26]. Radarski grafikoni označujejo preizkušene tehnike z uporabo različnih meril uspešnosti, vključno z natančnostjo prepoznavanja spola (GD↑), natančnostjo prepoznavanja izraza (EX↑), povprečno kvadratno napako (MSE↓), mero za uspešnost deidentifikacije (DEID↑) in mero za merjenje raznolikosti obrazov (DIV↑) – glejte [36] za definicijo mere DEID in DIV. Pri tem opazimo, da algoritem CPP-DeID ustvari najbolj uravnovešen kompromis med različnimi vidiki procesa deidentifikacije, hkrati pa lahko raven zaščite zasebnosti nadzoruje s parametrom *pp*.

Kvalitativno vrednotenje Rezultati kvalitativnega vrednotenja na bazi slik RaFD [35] so predstavljeni na sliki 5. Kot je razvidno, preizkušene tehnike zagotavljajo različne stopnje realizma slike, zaščite identitete, ohranjanja uporabnosti podatkov in variabilnosti nadomeščenega predela obraza na generiranih obraznih slikah. Opazimo lahko, da CPP-DeID kaže zaželene značilnosti ter da parameter zasebnosti *pp* konkretno vpliva na videz deidentificiranih slik na pričakovan način.

4 Zaključek

V prispevku smo povzeli razvoj optimizacijskega algoritma za anonimizacijo obraznih slik v visoki kakovosti (angl. Controllable Privacy Protection Deidentification, CPP-DeID), ki omogoča parametriziran nadzor nad količino zagotovljene zaščite zasebnosti. Pri tej rešitvi lahko končni uporabnik prek normaliziranega parametra določi raven zaščite zasebnosti, ki je neposredno vezana na obrazno identiteto, hkrati pa naša optimizacija zagotavlja upravljanje kompromisa med želeno zaščito identitete in ohranjeno uporabnostjo podatkov iz izvorne slike. Kakovost izhodne slike, ki jo proizvede naš generativni algoritem je zagotovljena z uvajanjem dodatnih omejitev v procesu optimizacije vhodne slike. Omejitve so v predlaganem postopku implementirane kot dodatni kriteriji pri optimizaciji prek ciljno usmerjenih globokih modelov učenja ali pa ciljno usmerjenih metrik za zaznavanje slikovne kakovosti.

Literatura

[1] S. D. Warren and L. D. Brandeis, "The right to privacy," *Harvard Law Review*, vol. 4, no. 5, pp. 193–220, 1890.

[2] A. Lukács, "What is privacy? the history and definition of privacy," 2016.

[3] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.

[4] J. D. Woodward, "Biometrics: Privacy's foe or privacy's friend?," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1480–1492, 1997.

[5] E. McCallister, *Guide to protecting the confidentiality of personally identifiable information*, vol. 800. Diane Publishing, 2010.

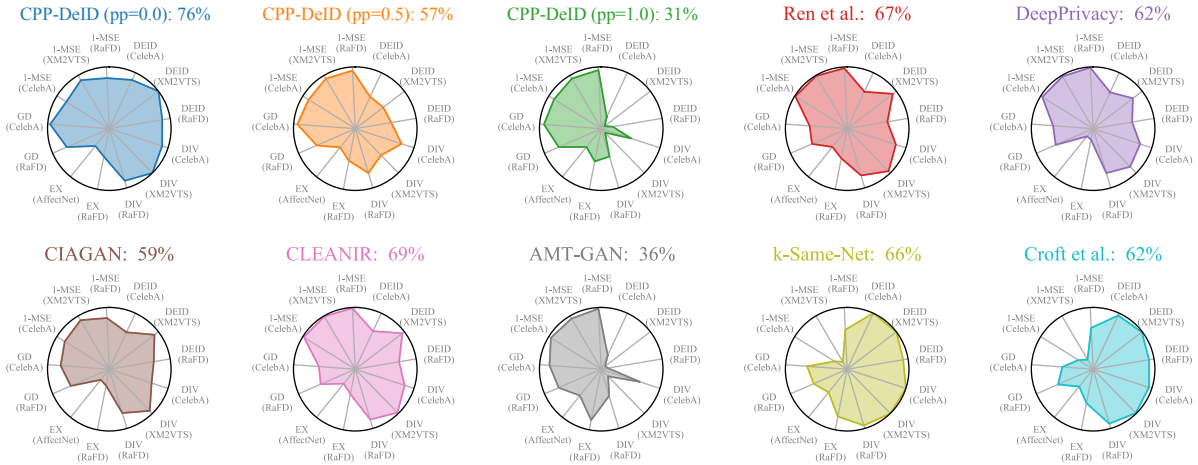
[6] R. Gross, L. Sweeney, J. Cohn, F. De la Torre, and S. Baker, "Face de-identification," in *Protecting privacy in video surveillance*, pp. 129–146, Springer, 2009.

[7] S. L. Garfinkel, "De-identification of personal information," *NIST-Technology Internal Report*, vol. 8053, pp. 1–46, 2015.

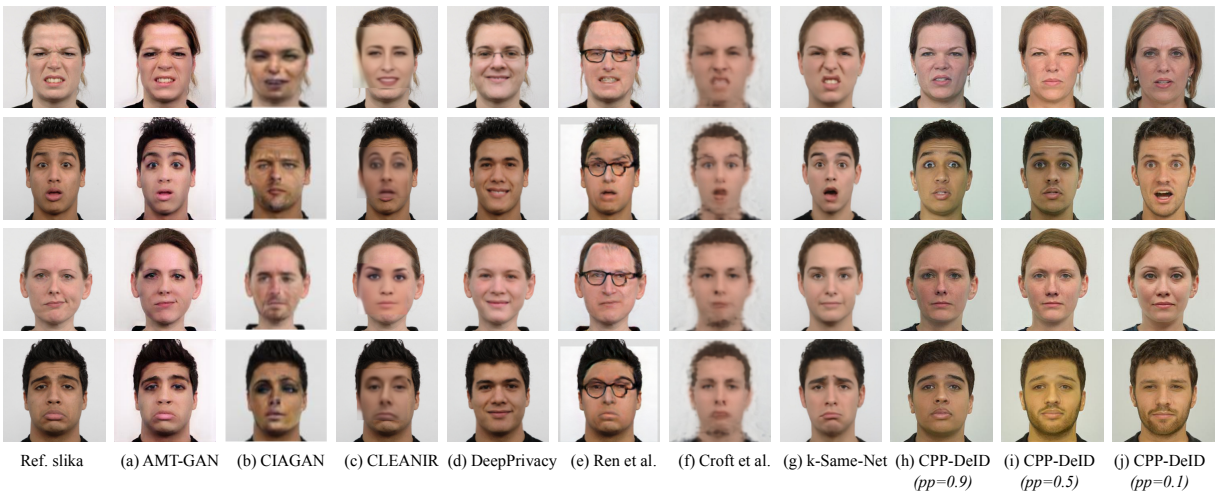
[8] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131 – 151, 2016.

[9] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in *International Conference on Intelligent Robots and Systems (IROS)*, pp. 971–978, 2007.

[10] Q. A. Zhao and J. T. Stasko, "Evaluating image filtering based techniques in media space applications," in *ACM conference on Computer Supported Cooperative Work CSCW*, pp. 11–18, 1998.



Slika 4: Primerjava tehnik za deidentifikacijo obrazov z grafi v obliki radarja. Posamezna os prikazuje en indikator (oz. mero) zmogljivosti, kar vključuje mero (1-MSE), mero deidentifikacije (DEID), mero raznolikosti (DIV), natančnost ohranjanja izraza na obrazu (EX) in natančnost ohranjanja atributa spola (GD). Večja je obarvana površina na posameznem grafu, boljša je splošna zmogljivost pripadajoče tehnike za deidentifikacijo obrazov.



Slika 5: Vizualna primerjava predlaganega algoritma CPP-DeID in preostalih algoritmov na bazi slik RaFD. Na primerjanih slikah je mogoče opaziti različne ravni vključene zaščite zasebnosti, obenem pa tudi različne ravni ohranjanja obraznih atributov, vizualne podobnosti z izvornimi slikami ter splošne kakovosti sintetiziranih obraznih slik.

[11] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[12] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1685–1692, 2014.

[13] A. Jourabloo, X. Yin, and X. Liu, "Attribute preserved face de-identification," in *International Conference on Biometrics (ICB)*, pp. 278–285, 2015.

[14] M. Saini, P. K. Atrey, S. Mehrotra, S. Emmanuel, and M. Kankanhalli, "Privacy modeling for video data publication," in *International Conference on Multimedia and Expo (ICME)*, pp. 60–65, 2010.

[15] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *International Conference on Privacy Enhancing Technologies (PETS)*, (Berlin, Heidelberg), pp. 227–242, 2006.

[16] M. Elliot, A. Hundepool, E. S. Nordholt, J. Tambay, and T. Wende, "Glossary on statistical disclosure control," *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, 2005.

[17] Q. Sun, L. Ma, S. Joon Oh, L. Van Gool, B. Schiele, and M. Fritz, "Natural and effective obfuscation by head inpainting," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 5050–5059, 2018.

[18] W. Xia, Y. Zhang, Y. Yang, J.-H. Xue, B. Zhou, and M.-H. Yang, "Gan inversion: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2022.

[19] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, 2020.

[20] R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan: How to embed images into the StyleGAN latent space?," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp. 4432–4441, 2019.

[21] O. Tov, Y. Alaluf, Y. Nitzan, O. Patashnik, and D. Cohen-Or, "Designing an encoder for StyleGAN image manipulation," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–14, 2021.
[22] Y. Alaluf, O. Tov, R. Mokady, R. Gal, and A. Bermano, "Hyperstyle: StyleGAN inversion with hypernetworks for real image editing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18511–18521, 2022.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
[24] Z. Wen, W. Lin, T. Wang, and G. Xu, "Distract your attention: Multi-head cross attention network for facial expression recognition," *arXiv:2109.07270*, 2021.
[25] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 41.1–41.12, BMVA Press, September 2015.
[26] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5549–5558, 2020.
[27] S. Hu, X. Liu, Y. Zhang, M. Li, L. Y. Zhang, H. Jin, and L. Wu, "Protecting facial privacy: Generating adversarial identity masks via style-robust makeup transfer," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15014–15023, June 2022.
[28] M. Maximov, I. Elezi, and L. Leal-Taixé, "CiaGAN: Conditional identity anonymization generative adversarial networks," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 5447–5456, 2020.
[29] D. Cho, J. H. Lee, and I. H. Suh, "CLEANIR: Controllable Attribute-Preserving Natural Identity Remover," *Applied Sciences*, vol. 10, no. 3, p. 1120, 2020.
[30] W. L. Croft, J.-R. Sack, and W. Shi, "Differentially private obfuscation of facial images," in *Machine Learning and Knowledge Extraction*, pp. 229–249, 2019.
[31] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," *arXiv:1909.04538*, 2019.
[32] Z. Ren, Y. J. Lee, and M. S. Ryoo, "Learning to anonymize faces for privacy preserving action detection," in *European Conference on Computer Vision (ECCV)*, pp. 620–636, 2018.
[33] B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "k-same-net: k-anonymity with generative deep neural networks for face deidentification," *Entropy*, vol. 20, no. 1, p. 60, 2018.
[34] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F. B. Tek, G. B. Akar, F. Deravi, and N. Mavity, "Face Verification Competition on the XM2VTS Database," in *Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 964–974, 2003.
[35] O. Langner, R. Dotsch, G. Bijlstra, D. Wigboldus, S. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition&Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.
[36] P. Nousi, S. Papadopoulos, A. Tefas, and I. Pitas, "Deep autoencoders for attribute preserving face de-identification," *Signal Processing: Image Communication*, vol. 81, p. 115699, 2020.

Dr. Blaž Meden je asistent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Njegovi raziskovalni interesi vključujejo teme, kot so deidentifikacija obraza, generativni modeli, slikovna biometrija, globoko učenje in računalniški vid. Leta 2023 je prejel nagrado EAB Biometrics Industry Award za svoje nedavno delo na področju deidentifikacije obraza.