

Vrednotenje postopkov superresolucije za izboljšavo lokalizacije obraznih točk

Klemen Grm

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška cesta 25, 1000 Ljubljana, Slovenija
E-pošta: klemen.grm@fe.uni-lj.si

Povzetek. Postopki superresolucije so se izkazali kot izjemno uporabni na področju obraznih slik, saj lahko z njimi izboljšamo kakovost slik nizke ločljivosti oziroma slik, podvrženih drugim dejavnikom vizualne degradacije, kot so slabo izostrenje, šum in razostritev zaradi gibanja. Z uporabo postopkov superresolucije obrazov na takih podatkih lahko izboljšamo delovanje biometričnih sistemov, npr. za razpoznavanje obrazov ali nadzorne sisteme. Postopki lokalizacije obraznih točk so pri biometričnih sistemih ključnega pomena, saj se tipično uporabljajo v predobdelavi obraznih slik, za poravnavo obrazov v kanonične poze, ki jih pričakujejo samodejni sistemi razpoznavanja obrazov. V članku je ovrednotena uporaba postopkov superresolucije obrazov za namen izboljšave delovanja metod lokalizacije obraznih točk. Na podlagi eksperimentov z različnimi postopki superresolucije obrazov sledi ugotovitev, da je vpliv postopkov superresolucije na uspešnost lokalizacije obraznih točk močno odvisen od postopka superresolucije, vsi preizkušeni postopki pa imajo pozitiven vpliv v primerjavi z lokalizacijo obraznih točk neposredno iz obraznih slik nizke ločljivosti. Glede na to ugotovitev se predlaga vključitev postopkov superresolucije kot enega od prvih korakov pri delu z obraznimi slikami nizke ločljivosti.

Ključne besede: biometrija, superresolucija, lokalizacija obraznih točk

Evaluation of super-resolution methods for improving face landmark localisation performance

Super-resolution methods have proven to be useful in the face domain, where they can be used to improve the quality of low-resolution images, as well as images subjected to different forms of visual degradation, such as out-of-focus, shot noise and motion blur. Super-resolution methods can be used on such data to improve the performance of various biometric systems, such as face recognition and surveillance systems. Landmark localisation is of key importance in biometric systems, since it is typically used in the image processing pipeline to align face images to a canonical pose expected by automatic face recognition systems. This paper evaluates the use of face super-resolution systems in terms of their effect on the performance of face landmark localisation methods. Experiments with different face super-resolution systems show that the impact of super-resolution on face landmark localisation is heavily dependent on the specific super-resolution method, however, the impact of each tested method is positive in comparison to landmark localisation directly from low-resolution face images. Based on our findings, the inclusion of face super-resolution as one of the first pre-processing steps in the face recognition pipeline is recommended.

Keywords: biometrics, super-resolution, face landmark localisation

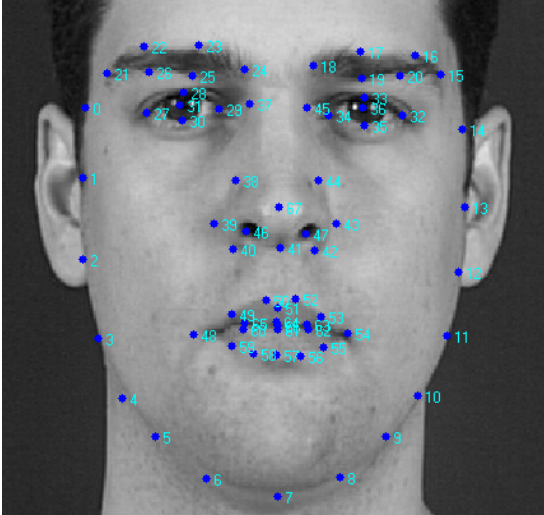
1 UVOD

Obrazna poza je eden od ključnih dejavnikov, ki lahko vplivajo na uspešnost razpoznavanja obrazov. Človeški obraz je najbolje viden v frontalni pozi, stranske poze pa



Slika 1: Obraz, fotografiran v različnih pozah.

lahko odvisno od kota bistveno spremenijo videz obraza ali celo zakrijejo dele obraza. Slika 1 prikazuje fotografije istega obraza v različnih pozah, ki izrazito spremenijo videz slike. Največ informacije o izgledu obraza je ohranjene v sliki frontalne poze, zato je to poza, za katero so tipično prilagojeni sistemi za samodejno razpoznavanje obrazov, in poza, iz katere je najlažje prepoznati obraz na sliki. Kot korak predobdelave slik za razpoznavanje obrazov se zato tipično uporabljajo postopki poravnave obrazov, ki sliko poravnajo v to kanonično pozo glede na zaznano lokacijo in orientacijo obraza na sliki.



Slika 2: Lokacije 68 ključnih obraznih točk po standardu MultiPIE [5]. Primer frontalne poze iz zbirke XM2VTS [14].

2 LOKALIZACIJA IN PORAVNAVA OBRAZNIH TOČK

Lokalizacija obraznih točk je postopek določitve koordinat slikovnih elementov na sliki obraza, ki ustrezajo ključnim obraznim točkam. Obstaja več različnih definicij ključnih obraznih točk, v zadnjem času pa se je najbolj uveljavil sistem 68 obraznih točk, uporabljenih pri označitvi podatkovne zbirke Multi-PIE [5]. Diagram teh točk v kanonični frontalni pozi obraza je predstavljen na sliki 2. Če želimo novo sliko poravnati s kanonično pozo, in na podlagi sistema za zaznavo obraznih točk poznamo približek (x, y) koordinate njenih obraznih točk v obliki matrike $\hat{\mathbf{X}} \in \mathbb{R}^{68 \times 2}$, in je kanonična poza določena z matriko koordinat $\mathbf{Y} \in \mathbb{R}^{68 \times 2}$, lahko sliko poravnamo s kanonično pozo tako, da določimo transformacijsko matriko $\mathbf{T} \in \mathbb{R}^{3 \times 3}$, ki v smislu najmanjših kvadratov reši enačbo

$$\hat{\mathbf{X}}' \mathbf{T} \approx \mathbf{Y}', \quad (1)$$

kjer sta $\hat{\mathbf{X}}'$ in \mathbf{Y}' matriki koordinat obraznih točk v homogenem zapisu, torej z vrsticami $(x, y, 1)$. Enačbo (1) v smislu najmanjših kvadratov reši matrika

$$\hat{\mathbf{T}} = \left(\hat{\mathbf{X}}'^T \hat{\mathbf{X}}' \right)^{-1} \hat{\mathbf{X}}'^T \mathbf{Y}'. \quad (2)$$

Slika 3 prikazuje primere obraznih slik, poravnanih s tem postopkom. Postopek zagotavlja sliko obraza v isti velikosti, obliki in orientaciji, kot jih ima kanonična poza, kar je ključnega pomena za delovanje sistemov samodejnega razpoznavanja obrazov. Tudi najnovejši rezultati s področja samodejnega razpoznavanja obrazov namreč kažejo, da je obrazna poza eden izmed dejavnikov, ki najbolj vplivajo na uspešnost razpoznavanja [15].



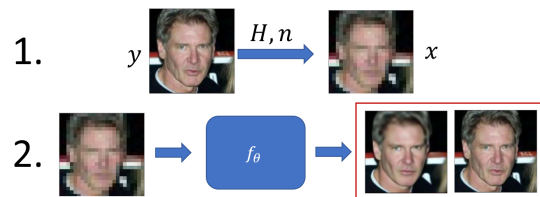
Slika 3: Primeri poravnave obraznih slik iz podatkovne zbirke CelebFacesA [13]. Na sliki so prikazane originalne (levo) in poravnane slike (desno).

3 SUPERRESOLUCIJA

Superresolucija je družina postopkov za izboljšavo kakovosti slikovnih podatkov nizke ločljivosti. Problem je matematično definiran s postopkom degradacije slike,

$$\mathbf{y} = H(\mathbf{x}) + n, \quad (3)$$

kjer \mathbf{y} predstavlja znano sliko nizke ločljivosti, $H(\cdot)$ predstavlja postopek glajenja in podvzorčenja, \mathbf{x} predstavlja latentno sliko visoke ločljivosti, ki jo s postopkom superresolucije želimo rekonstruirati, in n predstavlja člen šuma. Postopki superresolucije sestojijo iz določitve preslikave $\hat{\mathbf{x}} = f(\mathbf{y})$, ki postopek degradacije do čim večje mere invertirajo. Moderni pristopi [2], [8], [9], [11], [17], [18], [4] postopek določijo kot $\hat{\mathbf{x}} = f_\theta(\mathbf{y})$, kjer f_θ predstavlja globoko nevronske omrežje, naučeno na primerih parov slik visoke in nizke ločljivosti. Postopek učenja je tak, da učno zbirko slik visoke ločljivosti umetno podvzorčimo z izbranim degradacijskim postopkom. Umetno podvzorčene slike nizke ločljivosti \mathbf{y} nato predstavljajo učne vhode nevronskega omrežja, originalne slike visoke ločljivosti \mathbf{x} pa pričakovane učne izhode. Z nevronskega omrežjem f_θ nato preko gradientne optimizacije parametrov θ minimiziramo kriterijsko funkcijo $\mathcal{L}(f_\theta(\mathbf{y}), \mathbf{x})$, ki je lahko npr. srednja kvadratna napaka med dejanskimi slikami visoke ločljivosti in izhodi nevronskega omrežja. Postopek učenja in uporabe nevronskega omrežja za superresolucijo je prikazan na sliki 4.



Slika 4: Prikaz učenja nevronskega omrežja za superresolucijo. Slika prikazuje generiranje parov učnih slik (1) in učenje nevronskega omrežja (2).

4 EKSPERIMENTALNI PROTOKOL

V tem razdelku je opisan protokol, uporabljen za vrednotenje vpliva postopkov superresolucije na uspešnost lokalizacije obraznih točk pri obraznih slikah nizke ločljivosti.

4.1 Podatkovne zbirke

Za učenje postopkov superresolucije uporabljamo podatkovno zbirko CASIA WebFace [16]. Gre za podatkovno zbirko, ki vsebuje 494 414 slik obrazov 10 575 ljudi, in je pogosto uporabljena v namen učenja sistemov za analizo in razpoznavanje obrazov. Z učenjem postopkov superresolucije izključno na slikah obrazov dosežemo, da se nevronska omrežja za superresolucijo naučijo rekonstrukcije videza obrazov, kar predstavlja bolj omejen problem od splošne resolucije poljubnih scen, s čimer je mogoča izboljšava uspešnosti superresolucije.

Za testiranje vpliva postopkov superresolucije na uspešnost lokalizacije obraznih točk uporabljamo podatkovno zbirko Helen [10]. Testni del zbirke je sestavljen iz 330 obraznih slik z označenimi lokacijami ključnih obraznih točk po standardu s 68 točkami.

4.2 Postopki superresolucije

Za vrednotenje uporabnosti postopkov superresolucije za namen izboljšave delovanja sistema za lokalizacijo obraznih točk uporabimo predlagane postopke superresolucije Super-identity Convolutional Neural Network (SICNN) [18], Super-Resolution Convolutional Neural Network (SRCNN) [2], Very Deep Super Resolution Network (VDSR) [8], Perceptual-loss based SR (ℓ_p) [6], Cascading Residual Network (CARN) [1], Deep Laplacian Pyramid Super-Resolution Network (LapSRN) [9], Super-Resolution Generative Adversarial Network (SRGAN) [11], Enhanced Deep Residual Network (EDSR) [12], Ultra Resolving Discriminative Generative Network (URDGN) [17] in Cascaded Super-Resolution with Identity Priors (C-SRIP) [4]. Ker so bili nekateri izmed teh pristopov predlagani za splošen problem superresolucije in ne konkretno za problem superresolucije obrazov, vse modele naučimo od začetka na podatkovni zbirki CASIA-WebFace.

4.3 Postopek lokalizacije ključnih obraznih točk

Za določitev lokacije ključnih obraznih točk uporabljamo postopek [7]. Postopek je sestavljen iz kaskade regresijskih funkcij, ki iterativno izboljšujejo oceno lokacij ključnih obraznih točk na podlagi videza lokalne okolice trenutnih ocen njihovih lokacij. Postopek lokalizacije je podan z izrazom

$$\hat{\mathbf{X}}^{(t)} = \hat{\mathbf{X}}^{(t-1)} + r_t \left(I, \hat{\mathbf{X}}^{(t-1)} \right), \quad (4)$$

kjer I predstavlja sliko obraza, $\hat{\mathbf{X}}^{(t)}$ predstavlja oceno lokacij ključnih obraznih točk v t -ti iteraciji, in $r_t(\cdot, \cdot)$

predstavlja t -to regresijsko funkcijo, ki oceni nove lokacije ključnih točk na podlagi slike obraza in prejšnje ocene.

Zaporedje regresijskih funkcij r_t je učeno z gradientnim optimizacijskim postopkom za drevesa regresijskih funkcij, ki v vsakem koraku optimizira funkcije za šibko regresijo s kriterijsko funkcijo minimizacije rezidualov iz prejšnjih iteracij.

4.4 Mera za vrednotenje rezultatov

Naj bodo referenčne koordinate ključnih obraznih točk dane slike obraza podane v matriki \mathbf{Y} , in naj bodo približki teh koordinat, dobljeni s postopkom lokalizacije podani v matriki $\hat{\mathbf{X}}$. Odstopanje napovedi $\hat{\mathbf{X}}$ od dejanskih koordinat \mathbf{Y} podamo z mero povprečne evklidske razdalje dane ključne točke od njenega približka, normirano z medočasno evklidsko razdaljo. Mera napake je podana z izrazom

$$D_{IOD} = \frac{\frac{1}{68} \sum_{i=1}^{68} \left\| \mathbf{Y}_{i,:} - \hat{\mathbf{X}}_{i,:} \right\|_2}{\left\| \mathbf{Y}_{31,:} - \mathbf{Y}_{36,:} \right\|_2}, \quad (5)$$

kjer izraz $\mathbf{M}_{i,:}$ predstavlja i -to vrstico matrike \mathbf{M} , ter 31. in 36. ključna obrazna točka v sistemu 68 Multi-PIE [5] ključnih obraznih točk predstavljata središči levega oz. desnega očesa.

4.5 Priprava testne podatkovne zbirke

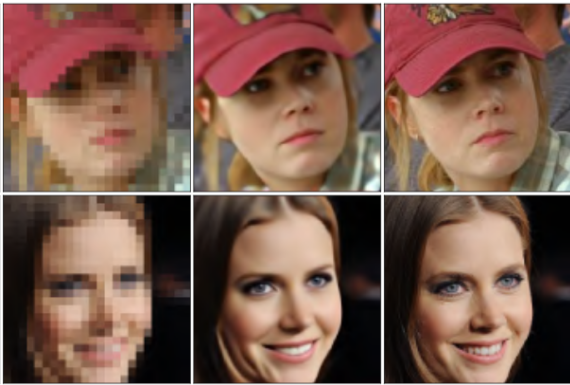
Ker podatkovna zbirka Helen [10] vsebuje slike visoke ločljivosti, jih moramo za namen uporabe preizkusa postopkov superresolucije obdelati z ustreznim postopkom umetne degradacije.

Vpliv postopkov superresolucije na uspešnost lokalizacije ključnih obraznih točk ovrednotimo tako, da celotni testni del podatkovne zbirke Helen podvzorčimo na ločljivost slik 24×24 slikovnih elementov, ki je smiselna za vhod v izbrane postopke superresolucije. V ta namen slike gladimo preko konvolucije z Gaussovimi jedrom s $\sigma = 2.0$ ter jih podvzorčimo za faktor 8.

Ker so koordinate referenčnih oznak ključnih obraznih točk podane v koordinatnem sistemu slikovnih elementov, tudi oznake lokacij obraznih točk skaliramo tako, da matriko \mathbf{Y} , pripadajočo vsaki sliki množimo z razmerjem med velikostjo originalne slike in velikostjo nove referenčne slike visoke ločljivosti, ki je 192×192 slikovnih elementov.

Te slike nizke ločljivosti nato z vsakim od omenjenih postopkov superresolucije povečamo za faktor $8 \times$, na ločljivost 192×192 slikovnih elementov. Primeri povečave so prikazani na sliki 5. Kot osnovo za primerjavo uporabimo še postopek bikubične interpolacije, ki slike poveča z interpoliranjem vrednosti med sosednjimi slikovnimi elementi brez učenja degradacije.

Rezultate preizkusa uspešnosti lokalizacije obraznih točk primerjamo tudi z uspešnostjo na originalnih slikah visoke ločljivosti. Rezultate uspešnosti lokalizacije pri uporabi vsakega izmed postopkov superresolucije preko celotnega testnega dela podatkovne zbirke Helen [10]



Slika 5: Prikaz superresolucije slike s postopkom C-SRIP [4]. Slika nizke ločljivosti je povečana za faktor povečave $8\times$. Levo: Slika nizke ločljivosti. Sredina: Rezultat postopka superresolucije. Desno: Referenčna slika visoke ločljivosti.

podamo v obliki statistike ($\mu \pm \sigma$), tj. s srednjo vrednostjo in standardnim odklonom, izračunanim preko rezultatov na 330 slikah v testni podatkovni zbirki.

5 REZULTATI

V tabeli 1 so z mero odstopanja D_{IOD} podani rezultati preizkusa lokalizacije obraznih točk. Statistiko mere preko celotne podatkovne zbirke podamo v obliki intervala zaupanja ($\mu \pm \sigma$).

Iz rezultatov je razvidno, da vsi preizkušeni postopki superresolucije izboljšajo delovanje postopka lokalizacije ključnih obraznih točk, saj je postopek bikubične interpolacije, ki smo ga uporabili kot osnovo za primerjavo v smislu izbrane mere napake lokalizacije v povprečju najslabši in glede na standardni odklon mere napake D_{IOD} med najmanj konsistentnimi.

Med postopki superresolucije k uspešnosti lokalizacije ključnih obraznih točk najbolj pripomorejo tisti, ki pri postopku učenja nevronskega omrežja za superresolucijo na različne načine eksplicitno vključujejo znanje o videzu obrazov (C-SRIP, ℓ_p), sledijo pa jim postopki, ki vključujejo metode nenadzorovanega in nasprotniškega učenja (EDSR, SRGAN) po postopku učenja generativnih nasprotniških omrežij [3]. Med najslabšimi postopki so zastopani predvsem tisti, pri katerih učenje modelov za superresolucijo poteka zgolj na podlagi kriterija naivne rekonstrukcije učnih slik visoke ločljivosti v smislu najmanjših kvadratov (SRCNN, VDSR). Kvantitativni rezultati v tabeli 1 kažejo, da napredki na področju superresolucije v zadnjih letih, ki se v literaturi tipično merijo s kakovostjo rekonstrukcije testnih slik in kvalitativnim videzom rezultatov, močno korelirajo z uspešnostjo teh postopkov na področju uporabe za izboljšavo lokalizacije obraznih točk.

Na sliki 6 so prikazani tudi kvalitativni vizualni

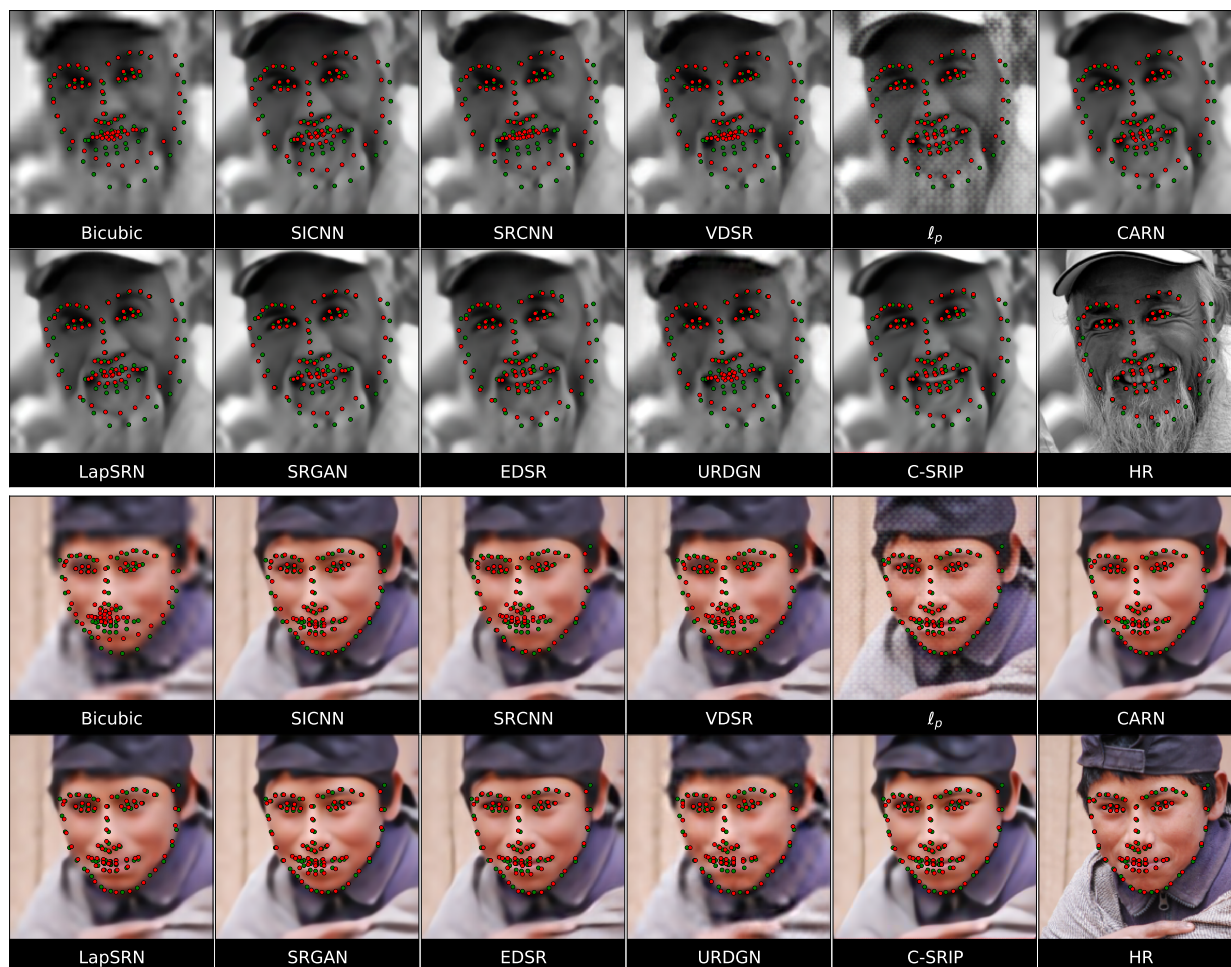
Tabela 1: Rezultati preizkusa uspešnosti lokalizacije obraznih točk pri uporabi postopkov superresolucije. Postopki so razvrščeni v padajočem redu povprečne napake lokalizacije.

Postopek	Napaka ($D_{IOD}, \mu \pm \sigma$)
Bikubična	
interpolacija	0.0531 ± 0.0052
SRCNN	0.0502 ± 0.0057
VDSR	0.0502 ± 0.0048
URDGN	0.0487 ± 0.0042
LapSRN	0.0449 ± 0.0050
SICNN	0.0431 ± 0.0036
CARN	0.0417 ± 0.0035
EDSR	0.0409 ± 0.0045
SRGAN	0.0405 ± 0.0034
ℓ_p	0.0396 ± 0.0052
C-SRIP	0.0380 ± 0.0028
Originalne slike	0.0344 ± 0.0025
visoke ločljivosti	

rezultati dveh primerov lokalizacije ključnih obraznih točk pri uporabi vsakega od preizkušenih postopkov superresolucije. Tudi iz kvalitativnih primerov je razvidno, da kakovost rekonstrukcije slike postopka superresolucije jasno pripomore k uspešnosti lokalizacije obraznih točk. Na zgornjem primeru vidimo, da pri uporabi slabših superresolucijskih postopkov (kot je SRCNN) postopek lokalizacije obraznih točk popolnoma zgreši ključne obrazne točke na ustih in pri lokalizaciji obrisa obraza naredi veliko večjo napako, kot jo naredi pri uporabi novejših postopkov superresolucije (kot je ℓ_p). Na spodnjem primeru zaradi boljšega kontrasta slike postopek lokalizacije obraznih točk neodvisno od postopka superresolucije v večini primerov zadane obrisa obraza, obenem pa pri uporabi slabših izmed postopkov superresolucije zgreši ključne obrazne točke na nosu in na ustih.

6 ZAKLJUČKI

Pravilna poravnava obraznih slik je ključnega pomena v biometričnih sistemih. Zanesljiv sistem za poravnavo obrazov nam lahko bistveno olajša delo z obraznimi slikami, kjer je obraz slikan pod pozo, ki predstavlja problem za sisteme samodejnega razpoznavanja obrazov. Ta scenarij je pogost npr. v sistemih za varnostni nadzor, kjer so kamere tipično nameščene na način, ki čimbolj učinkovito pokriva prostor, ne zagotavlja pa slik obrazov pod primernimi pozami. V temu delu smo pokazali, da pri predobdelavi slik nizke ločljivosti korak uporabe postopkov superresolucije lahko bistveno pripomore k pravilni lokalizaciji ključnih obraznih točk, ki jih uporabljamo za poravnavo obraznih slik. Zato se delu s podatki nizke ločljivosti, kjer je potrebna poravnava slik priporoča uporabo postopkov superresolucije.



Slika 6: Kvantitativni prikaz uspešnosti postopka lokalizacije obraznih točk pri uporabi različnih postopkov superresolucije, v primerjavi z rezultati na originalnih slikah visoke ločljivosti. Na vsaki izmed slik zelene točke prikazujejo referenčne oznake koordinat ključnih obraznih točk, Y , rdeče točke pa prikazujejo napovedi postopka za lokalizacijo obraznih točk, \hat{X} , pri uporabi danega postopka superresolucije.

LITERATURA

- [1] N. Ahn, B. Kang, and K. A. Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *European Conference on Computer Vision (ECCV)*, pages 252–268, 2018.
- [2] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 184–199, 2014.
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- [4] K. Grm, W. J. Scheirer, and V. Štruc. Face hallucination using cascaded super-resolution and identity priors. *IEEE Transactions on Image Processing*, 29:2150–2165, 2020.
- [5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multiple. *Image and Vision Computing*, 28(5):807–813, 2010.
- [6] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711, 2016.
- [7] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1867–1874, 2014.
- [8] J. Kim, L. J. Kwon, and K. L. Mu. Accurate image super-resolution using very deep convolutional networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.
- [9] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 5, 2017.
- [10] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *European Conference on Computer Vision (ECCV)*, pages 679–692. Springer, 2012.
- [11] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017.
- [12] B. Lim, S. Son, H. Kim, S. Nah, and K. Lee. Enhanced deep residual networks for single image super-resolution. In *Computer Vision and Pattern Recognition Workshops (CVPR-W)*, volume 2, 2017.
- [13] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *International Conference on Computer Vision (ICCV)*, pages 3730–3738, 2015.
- [14] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Cysz, L. Vandendorpe, et al. Face verification competition on the xm2vts database. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 964–974. Springer, 2003.
- [15] D. Pal, C. Bhagavatula, Y. Zheng, R. Tao, and M. Savvides. Is pose really solved? a frontalization study on off-angle face matching. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2058–2067. IEEE, 2019.

- [16] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.
- [17] X. Yu and F. Porikli. Ultra-resolving face images by discriminative generative networks. In *European Conference on Computer Vision (ECCV)*, pages 318–333, 2016.
- [18] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang. Super-identity convolutional neural network for face hallucination. In *European Conference on Computer Vision (ECCV)*, pages 183–198, 2018.

Klemen Grm je leta 2020 doktoriral s področja elektrotehnike na Fakulteti za elektrotehniko Univerze v Ljubljani. Je asistent v Laboratoriju za strojno inteligenco na Fakulteti za elektrotehniko. Njegovo področje raziskovanja obsega strojno učenje, biometrijo in obdelavo slik.