Lana Popović i Jana Marković

Automatsko generisanje karikatura nenegativnom faktorizacijom matrice i konvolucionalnim neuronskim mrežama

***Cilj**: Automatsko generisanje karikatura koje će ispuniti željene umetničke, humorističke i stilske norme, održavajući originalni identitet osobe i preuveličavajući njene karakteristične atribute.*
***Metoda**: Na uzorku od 122 slike izračunava se 68 karakterističnih tačaka (landmarkova). Nenegativnom faktorizacijom matrice dobijamo bazne karakteristike i njihove prosečne raspodele. Pojedinačne vrednosti svakog lica se upoređuju sa baznim, i na osnovu tog odnosa pojačavaju. Dobijene slike sa deformisanim (naglašenim) fragmentima se propuštaju kroz već treniranu konvolucionu neuronsku mrežu, zajedno sa umetničkim delima čiji se stil preuzima.*
***Rezultati**: Za ocenu uspešnosti našeg algoritma sprovedena su tri tipa ankete. Prva anketa je zahtevala da se od različitih karikatura koje su generisane našim i drugim, sličnim algoritmima, ili naslikane od strane karikaturista, odaberu zadovoljavajuće, sa stanovišta privlačnosti, stila i humora. U drugoj anketi se od ispitanika tražilo da rangiraju različite algoritme generisanja karikatura. Treća anketa, u koju su uključene samo karikature generisane našim algoritmom, sadržala je dva dela: u prvom delu korisnici biraju neodređen broj karikatura koje su, prema njihovom nahođenju, uspešne. Drugi deo se sastojao od više različitih stilova primenjenih na jednoj karikaturi, a od korisnika se tražilo da izaberu pet najprikladnijih.*

Lana Popović (2001), Beograd, učenica 3. razreda Matematičke gimnazije u Beogradu

Jana Marković (2001), Kragujevac, učenica 3. razreda Prve kragujevačke gimnazije

MENTORI:

Gavrilo Andrić, Seven Bridges Genomics, Beograd

Pavle Šoškić, student Elektrotehničkog fakulteta Univerziteta u Beogradu

Uvod

Karikatura portreta je kombinacija karikature i portreta sa fokusom na karikiranju osobe – prenaglašavanju neke fizičke karakteristike lica. Obično se koriste radi zabave, kao pokloni ili suveniri, često nacrtane od strane uličnih umetnika. Karikaturisti imaju sposobnost uočavanja unikatnih crta lica, koja potom preuveličavaju i stilizuju. U poslednjih nekoliko godina računarska vizija i primena neuronskih mreža je izuzetno napredovala – u čemu je i umetnost našla svoje mesto. Automatsko gene-

risanje karikatura se koristi u mrežnim komunikacijama, onlajn igricama i u industriji animacije. Cilj ovog rada je da se automatsko generisanje karikatura učini što realističnijim i sličnijim ljudskom radu – kroz postizanje verodostojnosti, stila i šaljivosti na generisanim karikaturama.

Problemu generisanja karikatura može se pristupiti na različite načine. Bilo je pokušaja interaktivnog sintetisanja karikature, što je zahtevalo profesionalne veštine za postizanje ekspresivnih rezultata (Akleman 1997; Akleman *et al*. 2000; Chen *et al*. 2002; Mo *et al*. 2004). Predloženo je i nekoliko automatskih sistema, ali su se oni oslanjali na pravila „ručne izrade" karikatura, koja često proizlaze iz postupaka crtanja svojstvenih umetnicima (Brennan 1985; Koshimizu *et al*.1999; Mo *et al*. 2004). Nažalost, ovi pristupi su ograničeni na određeni umetnički stil (poput skiciranja grafitnom olovkom ili „kartunizovani" – cartoon stil) i unapred određene šablone naglašavanja.

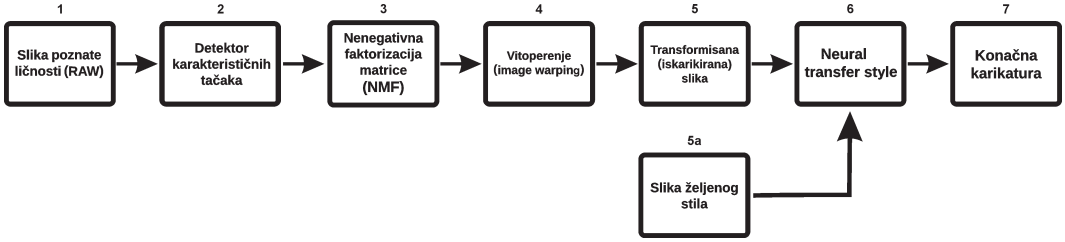
Poslednjih godina, duboko učenje, kao predstavnik tehnike učenja na primerima (posebno na velikim količinama podataka), uspešno je korišćeno za prevođenje slike u sliku (Huang *et al*. 2018; Koshimizu *et al*. 1999; Liu *et al*. 2017; Yi *et al*. 2017; Gatys *et al*. 2015). Međutim, za sada ne postoji dovoljna količina uparenih karikatura i odgovarajućih fotografija, pa su treniranja sa nadzorom (poput autoenkodera) neizvodljiva.

Najbolje rezultati postigli su Kaidi Cao, Jing Liao, Lu Yuan kreiranjem dve generativne neuronske mreže CariGeoGAN i CariStyGAN koje bi, respektivno, učile geometrijsku transformaciju i transponovanje stila (Cao *et al*. 2018). Međutim, mana ovog postpuka je nedostatak varijabilnosti kod naglašavanja lica. Naime, suština karikature nije bilo kakva promena lica, već isticanje onih delova koji su karakteristični za konkretan lik.

U ovom radu predstavljeno je korišćenje nove metodike izražavanja odstupanja od prosečnog (Exaggerating the Diference from the Mean – EDFM) koja rešava probleme CariGAN-a, nalik modelu kojeg su razvili Zhenyao Mo i saradnici (Mo *et al*. 2004). Suština ovog pristupa je pronalaženje „prosečnog lica" koje se koristi kao osnova za poređenje sa zadatim licem od kojeg se pravi karikatura. Delovi lica koji najviše odstupaju od proseka (stoga su i najprepoznatljiviji) se dodatno naglašavaju. Na ovaj način, unikatnost svakog lica biva sačuvana. Potom se na sliku primenjuje transfer stila izabranog umetničkog dela, za šta se koristi duboka konvoluciona neuronska mreža, što dodatno dodaje varijabilnost i fleksibilnost našim karikaturama.

Generisanje karikatura

Za uspešno kreiranje karikature potrebno je ispuniti dva zahteva – adekvatno naglasiti određene karakteristike lica i zadovoljiti stilska očekivanja likovnog dela. Stoga je i ovaj rad podeljen u dve celine: deformacija karakterističnih delova lica i kreiranje umetničkog stila. Karikiranje i deformaciju postižemo kombinacijom primene algoritama iz oblasti kompjuterske vizije: detekcija karakterističnih tačaka, nenegativna faktorizacija matrice, i vitoperenje, dok stilizovani izgled dobijamo korišćenjem



Slika 1. Blok šema metoda za kreiranje karikature

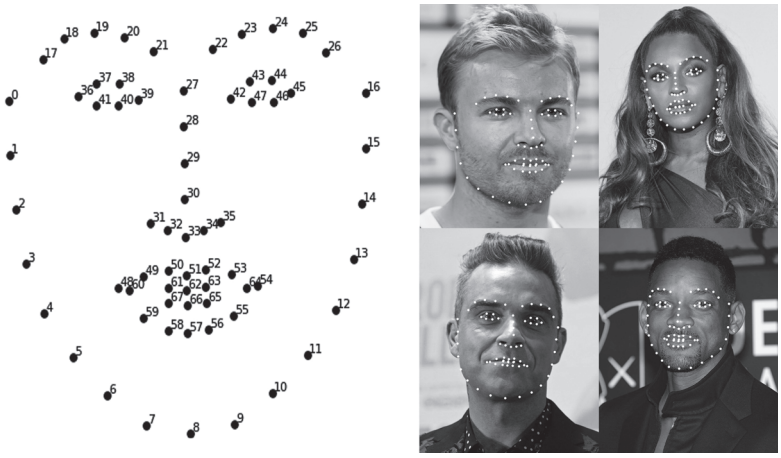
Figure 1. Method block diagram for caricature generation

istrenirane neuronske mreže za transfer stila. Za izradu projekta korišćeni su programski jezik Python, kao i platformska grafička kartica 12 GB NVIDIA Tesla K80 GPU ugrađena na programskom okruženju Google Colab kako bi se ispunili hardverski zahtevi ovog projekta. Kompletan algoritam prikazan je na slici 1.

Detekcija karakterističnih tačaka

Prvobitno je potrebno pronaći karakteristične tačke lica. Za to je korišćen istrenirani model za automatsku detekciju i mapiranje lica („landmark" detekciju) u Python-u sa bibliotekom dlib, koji uspešno detektuje 68 landmark koordinata (*x*, *y*) na bilo kojem licu (slika 2):

1. Obeležja brade (0–16)
2. Obeležja desne obrve (17–21)
3. Obeležja leve obrve (22–26)
4. Obeležja nosa (27–35)
5. Obeležja desnog oka (36–41)



Slika 2.
a) 68 karakterističnih tačaka
b) Primer karakterističnih tačaka na različitim licima

Figure 2.
a) 68 landmarks
b) Examples of different landmarked faces

6. Obeležja levog oka (42–47)
7. Obeležja usta (48–68)

Kroz ovaj model je propušten uzorak od 122 fotografije, koje su korišćene za pronalaženje srednjeg lica. Za svaku karakterističnu tačku lica, na svakoj od fotografija, računa se udaljenost od svih ostalih tačaka pojedinačno. Zatim se nalazi aritmetička sredina razdaljina između parova tačaka obeleženih istim brojem na uzorku. Srednje (bazno) lice predstavlja mapu portreta čije su karakteristične tačke postavljene tako da odgovaraju srednjim vrednostima.

Problem mapiranja čela i gornje granice lica

Mana prethodno pomenutog modela za automatsku detekciju i mapiranje lica je nedostatak karakterističnih tačaka koje bi obeležile gornju granicu portreta i čelo osobe. Posledica toga je nepronalaženje srednje dimenzije čela, što onemogućava njegovu karikaturizaciju. Zato se pokušalo sa dodavanjem karakterističnih tačaka koje bi pratile gornju liniju lica (slika 4). Ideja je bila korišćenje već trenirane konvolucione neuronske mreže čiji bi zadatak bio detektovanje, a potom i izdvajanje, svih piksela neke slike na kojima se nalazi kosa. Na taj način bi se dobila maska kose, na čiji deo bi se nastavilo lice (slika 3).

Slika 3.
Primer detekcije i izdvajanja kose na fotografiji. S leva na desno: fotografija, maska, prekrivanje

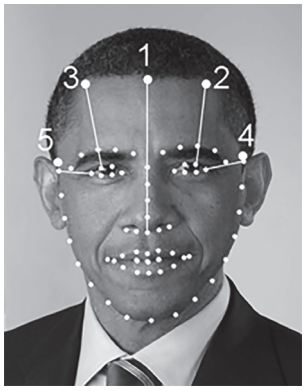
Figure 3. Example of hair detection and segmentation. From left to right: photo, mask, overlay

Dodatnih pet karakterističnih tačaka bi se dobilo povlačenjem pravih linija kroz sledeće tačke:

1. tačke 30 i 27. Ova prava predstavlja uzdužnu osu simetrije lica i prolazi kroz sredinu čela. Najniža zajednička tačka maske kose i date prave definisala bi prvu dodatu karakterističnu tačku.
2. tačke 44 i 24. Ova prava spaja centar desnog oka i desne obrve, i u najnižoj presečnoj tački sa maskom kose postavila bi se druga dodata karakteristična tačka.
3. tačke 37 i 19. Ova prava spaja centar levog oka i leve obrve, i u najnižoj presečnoj tački sa maskom kose postavila bi se treća dodata karakteristična tačka.
4. tačke 45 i 36. Ova prava spaja spoljašnji ugao desnog oka sa završetkom desne obrve, i njena najniža zajednička tačka sa mas-

kom kose bi definisala mesto postavljanja četvrte dodate karakteristične tačke.

5. tačke 36 i 17. Ova praja spaja spoljašnji ugao levog oka sa završetkom leve obrve, i njena najniža zajednička tačka sa maskom kose bi definisala mesto postavljanja pete dodate karakteristične tačke.



Slika 4. Ilustracija dodatnih karakterističnih tačaka radi boljeg definisanja čela i gornje granice lica

Figure 4. Finding positions of the additional landmark points for better upper face boundary definition

Ovakava metoda izdvajanja ne daje uvek zadovoljavajuće rezultate, kako u nekim slučajevima pogrešno detektuje delove lica kao kosu (uglavom ako postoji jača senka na tom regionu) ili detektuje deo pozadine (ukoliko su sličnih boja). Takođe, ovaj princip nije efikasan ukoliko osoba nema kosu, ima bradu, nosi kapu, ili na neki drugi način prekriva kosu. Ipak, najveću prepreku predstavlja činjenica da se dodavanjem dodatnih pet tačaka, nasuprot pretpostavljenom, ne primećuje značajna promena, a ona koja i postoji je lošija u odnosu na rezultat dobijen izostavljanjem dodatih tačaka (slika 5). Zaključak je da obeležavanje gornje granice lica ograničiva i sužava prostor za promene, pogoršavajući time i sam kvalitet i autentičnost dobijene karikature (promene na karikaturi postaju tipične, i nedovoljno jedinstvene za svako lice).

Slika 5. Upoređivanje rezultata sa 73 i 68 karakterističnih tačaka (prvi i drugi red, respektivno)

Figure 5. Comparison between results with 73 and 68 landmark points (first and second row, respectively)

Zbog svega navedenog, eksperiment je ocenjen kao neuspešan, pa mapa lica zadržava početni izgled, bez uvrštavanja dodatnih tačaka. Naš generator za sada nema mogućnost karikiranja regiona čela, niti drugih delova portreta iznad obrva.

Nenegativna faktorizacija matrice

Nenegativna faktorizacija matrice (non-negative matrix factorization, NMF) je grupa algoritama koja ima zadatak da matricu V faktoriše u dve odvojene matrice W i H, pod uslovom da sve tri matrice nemaju negativne članove. Ova nenegativnost čini rezultirajuće matrice lakšim za inspekciju. Pomoću NMF, dimenzije dobijenih matrica-faktora mogu biti znatno manje od dimenzija prvobitne matrice-proizvoda. Na primer, ako matrica V ima dimenzije $n \times m$, matrica W $n \times p$ i matrica H $p \times m$, p može biti dosta manje od m i n.

Za potrebe našeg projekta, konstruisali smo matricu S dimenzije 122×136, gde prva dimenzija ($n = 122$) predstavlja broj slika, a druga dimenzija ($m = 136$) x i y koordinate korišćenih 68 landmark tačaka. Na dobijenu matricu S primenjuje se NMF (Lee i Seung 1999) algoritam kako bi pronašao dimenzije lica:

$$S = F \cdot E \tag{1}$$

gde je:

S – matrica lica dimenzija $n \times m$,

F – faktor matrice S, predstavlja matricu baznih karakteristika koju tražimo

E – faktor matrice S, koji takođe tražimo, predstavlja matricu enkodinga, tj. koeficijenata koji se koriste u linearnoj kombinaciji sa baznim karakteristikama iz matrice F, kako bi se dobila konačna matrica S. U našem slučaju broj komponenata iznosi $p = 65$.

Glavna obeležja i enkoding koeficijenti su nepravilno raspoređeni, tj. glavna obeležja se sastoje od više različitih pozicija tačaka usta, nosa i drugih delova lica, gde su različite verzije na različitim lokacijama ili formama. Celo lice generiše se kombinovanjem ovih različitih delova. Enkoding koeficijenti su tako raspoređeni da se nikad ne dogodi stapanje više pozicija očiju, obrva itd. Takođe, važno je napomenuti da ovom metodom ne dobijamo prvobitnu matricu S, već njenu aproksimaciju.

U formuli faktorizovane matrice $S = F \cdot E$, svaka dimenzija se sastoji od baznih vektora $\vec{f}_i$ i njegove raspodele e_i – koja je predstavljena kao $m_i + \sigma_i$, gde je m_i srednja vrednost i-te kolone matrice E, a σ_i standardna devijacija i-te kolone matrice E). Svaka dimenzija predstavlja poziciju apstrahovanog dela lica.

Oblik $\vec{s}$ na novoj slici predstavićemo kao linearnu kombinaciju baznih vektora $\vec{f}_i$ i vektora ostatka $\vec{r}$ (Mo *et al*. 2004):

$$\vec{s} = \sum_i e_i \vec{f}_i + \vec{r} = \sum_i (m_i + \delta_i)\vec{f}_i + \vec{r} \tag{2}$$

gde je δ_i – odstupanje od srednje vrednosti *i*-te kolone matrice *E*.

Kako bi dobili željeni oblik karikature treba povećati odstupanja, tj. iskarikirati delove lica. Za taj zadatak, imamo nekoliko slučajeva, gde se najviše treba fokusirati na $\delta_i = |e_i - m_i|$:

$$\vec{s}' = \sum_i (m_i + t_i \delta_i)\vec{f}_i + 0.8\, k\, \vec{r} \qquad (3)$$

1. Neka t_i i k budu koeficijenti karikiranja, tj. povećanja δ_i i vektora ostatka $\vec{r}$, kada je $|\delta_i| < \sigma_i$ – stavićemo da je $t_i = 1$;
2. U slučaju da je $|\delta_i| \geq \sigma_i$, $t_i = k$ (u našem slučaju, $k = 1.25$)

Uticaj koeficijenta deformacije *k* je ilustrovan na slici 6.

Slika 6.
Podešavanje koeficijenta deformacije *k*

Figure 6.
Adjustment of the deformation coefficient *k*

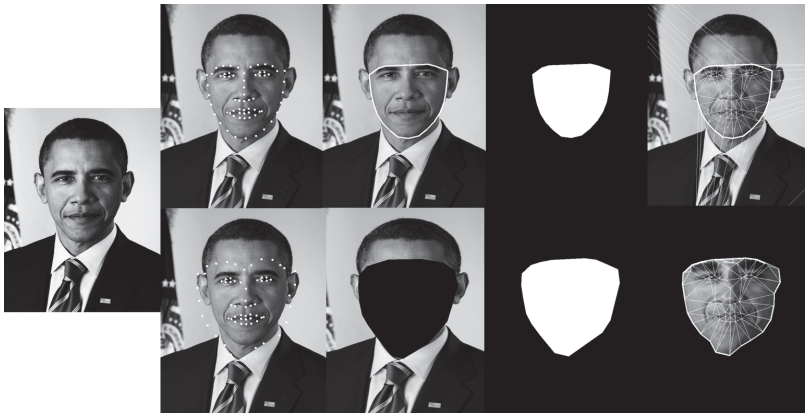
Pošto svaka slika prilikom primene NMF-a ostavlja ostatak (delovi slike koji se nisu mogli dobiti linearnom kombinacijom baznih vektora i raspodele), koeficijent promene je dodat i na vektor ostatka, samo u umanjenoj količini (u našem slučaju, uzet je koeficijent 0.8 – može se povećati ili smanjiti, u zavisnosti od željenog rezultata, tj. da li želimo drastičniju promenu, ili ne).

Konačno, dobijamo finalni vektor $\vec{s}'$ dimenzije 1×136 sa novim koordinatama apstraktnih iskarikiranih delova lica. Kompletnu fotografiju dobijamo vitoperenjem (image warping) originalnog oblika $\vec{s}$ i promenjenog $\vec{s}$ (slika 7).

Image Warping

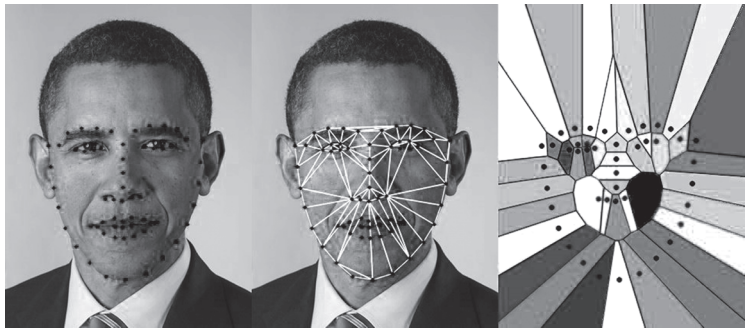
Kako pred image warping-om (vitoperenje slike) imamo dve matrice *S* i *S′* (matrica sa originalnim koordinatima tačaka na licu i matrica sa transformisanim, respektivno), tražimo konveksne omotače (poligone) landmark tačaka od kojih pravimo maske koje će obuhvatati prostor na kojem će se izvršiti Delanijeva triangulacija.

Delanijeva triangulacija je podela ravni na trouglove za određeni skup P diskretnih tačaka (konkretnije za naš projekat, korišćenjem koordinata iz matrice originalne slike *S*) tako da nijedna tačka P nije unutar oboda bilo kojeg trougla. Delanijevi trouglovi maksimiziraju minimalni ugao svih uglova trougla (slika 8). Kako bi vitoperenje bilo uspešno, pozicije karakterističnih tačaka iz obe matrice moraju da odgovaraju jedni drugima, tj. da



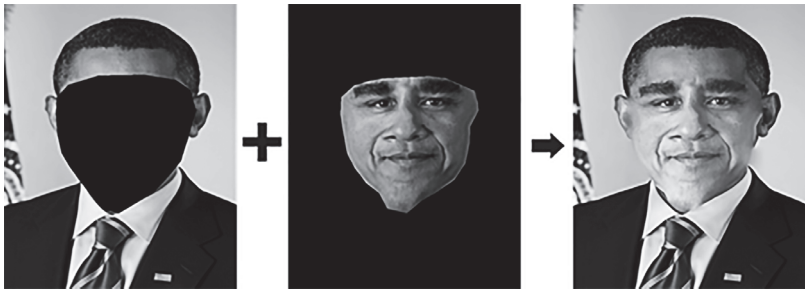
Slika 7.
Proces vitoperenja

Figure 7.
Process of image warping



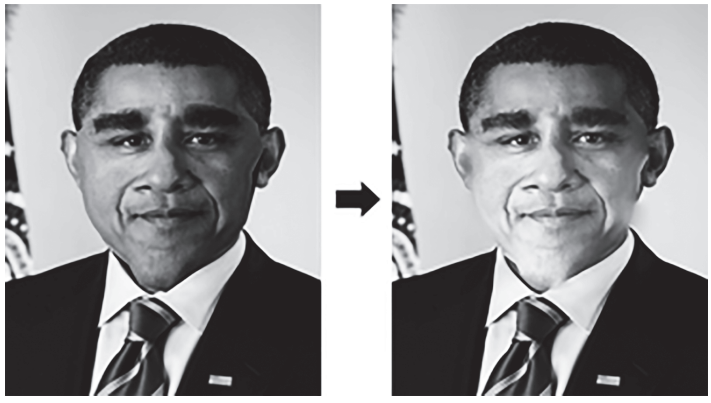
Slika 8.
Delanijeva triangulacija lica

Figure 8.
Delaunay triangulation of the face



Slika 9.
Dodavanje deformisanog lica na prvobitnu sliku

Figure 9.
Addition of deformed face onto original picture



Slika 10.
Rezultat primene funkcije OpenCV Seamless kloniranja

Figure 10.
Result of OpenCV Seamless Cloning

svaki Delanijev trougao originalne matrice S odgovara Delanijevom trouglu matrice S' iskarikiranih koordinata (Efros 2007). Kada imamo triangulaciju obe matrice, potrebno je da izvučemo dobijene trouglove i menjamo ih (tj. razvlačimo ili sužavamo) na osnovu položaja konačnih iskarikiranih koordinata matrice S'. Rezultat je iskarikiran oblik lica. Potom, isecamo originalan poligon lica bez iskrivljenih koordinata i dodajemo novodobijeni oblik lica kako bismo dobili konačnu sliku (slika 9).

Na kraju primenjujemo Seamless Cloning kako bismo uklonili vizuelne diskontinuitete između originalne fotografije i karikirane slike, odnosno nepotrebne linije i piksele nastale usled oštre triangulacije (slika 10). Seamless Cloning je OpenCV implementacija Poasonove jednačine na piksele slika, kako bi se postigla prirodnost prekrivanja (Pérez *et al*. 2003).

Neural style transfer

Neural style transfer je metod prebacivanja karakteristika jedne slike na drugu korišćenjem već trenirane konvolucione neuronske mreže. Ideja je da se na osnovu dve učitane slike dobije nova, kombinacija prethodne dve, koja će preuzeti karakteristike stila (poput boje i teksture) sa jedne slike, i karakteristike konteksta (odnosno sadržaja poput objekata, njihove raspodele i „radnje" slike) sa druge slike (slika 11).

Slika 11.
Kombinacija konteksta i stila

Figure 11.
Context image and style image put together for final image

Da bi se moglo meriti u kojoj meri se konačna slika razlikuje od ulazne (odnosno koliko je konteksta preuzeto iz kontekstne slike, a koliko stila iz obrasca stila), potrebno je definisati funkciju gubitka. Cilj je da se konačna slika razlikuje što manje od ulaznih (da funkcija gubitka konteksta i stila bude minimalna), odnosno da vrednost funkcije gubitka bude približna nuli. Funkciju gubitka L možemo zapisati kao:

$$L(G) = \alpha L_{\text{kontekst}}(C, G) + \beta L_{\text{stil}}(S, G) \qquad (3)$$

gde je G generisana slika, C slika sa koje se uzima kontekst, a S slika sa koje se uzima stil. Koeficijenti α i β su težinski faktori koji omogućuju da kontrolišemo koliko će kontekst biti naglašen u odnosu na stil. Funkcije gubitka računaju skalarne vrednosti $L(C, G)$ i $L(S, G)$ koje predstavljaju razliku između izlazne slike G i ciljane slike C, odnosno S (Narayanan 2017).

Konvolucione neuronske mreže se sastoje od slojeva računarskih jedinica (neurona) koje obrađuju vizuelne informacije hijerarhijski u feed-for-

Slika 12. Vizualizacija slojeva prema dubini – dublji slojevi se bave komplikovanijim objektima (https://blog.liexing.me/)

Figure 12. Visualisation of network layers – deeper layers are dealing with more complicated objects (https://blog.liexing.me/)

ward maniru (informacije se kreću neciklično). Svaki sloj neurona se može predstaviti kao kolekcija filtera za sliku, od kojih je svaki zadužen da sa slike izvuče određenu karakteristiku. Rezultat nakon svakog sloja je mapa karakteristika (feature map), dobijena primenom različitih filtera na početnu sliku.

Kada se konvolucione mreže treniraju da prepoznaju objekat, one razvijaju mape karakteristika koje naglašavaju određene informacije o objektu. Tokom obrade, ulazna slika se transformiše u reprezentaciju koja sve više vodi računa o stvarnom sadržaju slike. Viši slojevi manipulišu sadržajem višeg nivoa u odnosu na prethodne slojeve (slika 12), u smislu objekata i njihovog rasporeda na slici (na primer, niži slojevi mreže manipulišu sadržajem koji se odnosi na jednostavnije karaktiristike slike – poput piksela koji prikazuju kosu liniju u sličnim bojama, dok se u dubljim slojevevima već mogu pronaći komplikovaniji objekti popot ljudi ili životinja).

Da bismo umanjili nedostatke koje bi imalo posmatranje gubitaka po pikselu (ili drugim manjim jedinicama slike) i da bismo dozvolili našim funkcijama gubitaka da bolje „izmere" perceptivne i semantičke razlike između slika, koristimo princip generisanja slika pomoću optimizacije. Ključ ove metode je korišćenje konvolucione neuronske mreže koja je već „naučila" da dešifruje perceptivne i semantičke informacije koje bismo želeli da naše funkcije gubitaka mere. Zato koristimo duboku kovolucionu mrežu VGGNet koja je već istrenirana da klasifikuje slike, i pomoću nje definišemo naše funkcije gubitka. Koristimo VGG16 verziju, sa 16 težinskih slojeva – 13 konvolucionih i 3 FC sloja (potpuno povezana sloja). Ulaz za prvi konvolutivni sloj je 224×224 RGB slika. Slika prolazi kroz grupu konvolutivnih slojeva, gde su korišćeni filteri sa vrlo malim receptivnim poljem: 3×3 (što je najmanja veličina dovoljna da uhvati pojam levo/desno, gore/dole, centar). Ova mreža je trenirana na bazi podataka od preko 14 miliona labeliranih slika visoke rezolucije, koje se mogu razvrstati u tačno 1000 kategorija (Frossard 2016). Za naše potrebe, prvih 13 konvolucionih slojeva koristimo u originalnom obliku. Do 13. sloja mreža je „naučila" da prepozna koji deo slike je objekat. U kasnijim slojevima originalna mreža bi klasifikovala objekat, što nam nije porebno (treba nam

informacija gde se on nalazi na slici). Uzimamo izlaz sa 13. sloja i koristimo poslednja tri sloja da definišemo funkcije gubitka.

Ulazna slika ima svoju reprezentaciju u vidu filtriranih slika u svakoj fazi obrade. Dok broj različitih filtera raste pri povećanju dubine, veličina filtriranih slika se smanjuje zbog mehanizma za isključenje (max-pooling), što dovodi do smanjenja ukupnog broja neurona po sloju. Rezultat je povećanje kompeksnosti i količine informacija o uređenju scene na lokalnom nivou u dubljim slojevima mreže, na račun kvaliteta globalnog uređenja koji postoji u nižim slojevima. Odnosno: rekonstrukcija konteksta je bolja u nižim slojevima, a lošija u višim, i obrnuto, za stil: rekonstrukcija stila je lošija u nižim slojevima, a bolja u višim.

Sloj sa N_l različitih filtera sadrži N_l mapa karakteristika veličine M_l, gde M_l ima veličinu visina puta širina mape karakteristika. Tako se odgovori iz sloja l mogu smestiti u matricu $F^l \in \mathbb{R}^{N_l \times M_l}$, gde je F_{ij}^l aktivacija i-tog filtera na poziciji j u sloju l. Za vizualizaciju informacija koje se nalaze na različitim slojevima hijerarhije izvršavamo algoritam opadajućeg gradijenta na white-noise slici (slici sačinjenoj od nasumičnih piksela), da pronađemo drugu sliku kod koje se odgovori karakteristika poklapaju sa ulaznom slikom. Uzmimo da je p originalna slika, a x slika koja se generiše, a P^l i F^l njihove respektivne reprezentacije karakteristika u sloju l. Definišemo kvadratnu grešku funkcije gubitka između dve reprezentacije karakteristika:

$$L(\vec{a}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - A_{ij}^l)^2 \tag{4}$$

Izvod ove funkcije gubitka po aktivaciji u sloju l iznosi:

$$\frac{\partial L_{\text{kontekst}}}{\partial F_{ij}^l} = \begin{cases} (F^l - A^l)_{ij} & \text{ako } F_{ij}^l > 0 \\ 0 & \text{ako } F_{ij}^l < 0 \end{cases} \tag{5}$$

gde aproksimacija po slici x može da se izračuna koristeći standardnu bekpropagaciju greške (error back-propagation). To nam omogućava da iterativno menjamo x (u početnom trenutku white-noise sliku, sve dok odgovor određenog sloja u mreži ne bude dovoljno sličan odgovoru originalne slike $\vec{a}$.

Da bi se rekonstruisao stil ulazne slike, koristimo *feature space* (originalno dizajniran da prikupi informacije o teksturi) koji gradimo na osnovu odgovora filtera u svakom sloju. Reprezentacija stila se računa na osnovu korelacije između različitih karakteristika u različitim slojevima mreže. Uključivanjem korelacije karakteristika sa više slojeva, dobijamo stacionarnu, multi-skaliranu reprezentaciju učitane slike, koja sadrži informacije o teksturi, ali ne o globalnom rasporedu slike. Faktički, dobijamo sliku koja je „slepa" na ulazne slike. Ove korelacije karakteristika su date Gramovom matricom $G^l \in \mathbb{R}^{N_l \times M_l}$, gde G_{ij}^l je proizvod vektora mape karakteristika i i j u sloju L:

$$G_{ij}^l = \sum_k F_{ij}^d F_{ij}^l \tag{6}$$

Za generisanje teksture koja odgovara stilu date slike, minimiziramo srednju kvadratnu udaljenost između unosa Gramove matrice iz originalne slike i Gramove matrice slike koja treba biti generisana. Uzmimo da je $\vec{a}$ originalna, a $\vec{x}$ generisana slika, i A^l i G^l respektivno njihove reprezentacije stila u sloju l. Doprinos sloja l konačnom gubitku iznosi:

$$E_i = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \tag{7}$$

a ukupan gubitak je:

$$L_{\text{stil}}(\vec{a}, \vec{x}) = \sum_{i=0}^{l} w_i E_i \tag{8}$$

gde je w_l faktor težine doprinosa svakog sloja ukupnom gubitku. Izvod od E_l po aktivacijama u sloju l se računa analitički:

$$\frac{\partial L_{\text{stil}}}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left((F^l)^{\mathrm{T}} (G^l - \Lambda^l) \right)_{ji} & \text{ako } F_{ij}^l > 0 \\ 0 & \text{ako } F_{ij}^l < 0 \end{cases}$$

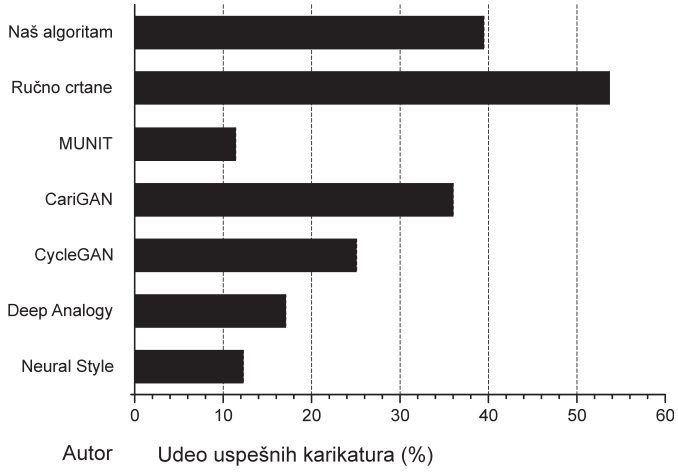
Suština ove metode je da su reprezentacije konteksta i stila u konvolucionim neuronskim mrežama odvojene, i njima možemo manipulisati nezavisno. Konačna slika počinje svoj život kao nasumična kolekcija piksela koja se potom iterativno unapređuje tako da razlika između karakteristika konteksta i karakteristika stila u odnosu na ulazne slike bude minimalna.

Evaluacija implementiranih algoritama i rezultati

Za određivanje uspešnosti našeg algoritma sprovedeno je tri tipa ankete:

1. Za prvu anketu je nasumično izabrano 10 fotografija portreta poznatih ličnosti. Zatim je svaka od tih fotografija propuštena kroz sledeće algoritme (slika 15): Neural Style (Gatys *et al*. 2015), Deep image analogy (Liao *et al*. 2017), jedno-modalna mreža za prevođenje slike CycleGAN, multi-modalna mreža za prevođenje slika MUNIT (Huang *et al*. 2018), CariGan (Cao *et al*. 2018) i naš algoritam. U izbor je ubačen i portret odgovarajuće karikature naslikane od strane čoveka. Anketu je uradila 21 osoba. Od učesnika je bilo zahtevano da od ponuđenih karikatura izaberu one koje su, po njihovom mišljenju, zadovoljavaju faktore vizuelnog kvaliteta, humora i stila. Rezultati ankete (slika 13) su predstavljeni procentom karikatura određenog stila koje su odabrane kao verodostojne.

Najveći procenat izabranih karikatura, kao što je i očekivano, činile su ručno izrađene karikature – 53.8%. Najuspešniji algoritam je bio naš, sa 39.5% prihvaćenih karikatura. Sledeći su CariGAN i CycleGAN sa 36.1% i 25.1%, respektivno. Ostali algoritmi postigli su znatno lošije rezultate – najlošiji, MUNIT ostvario je 11.5%.

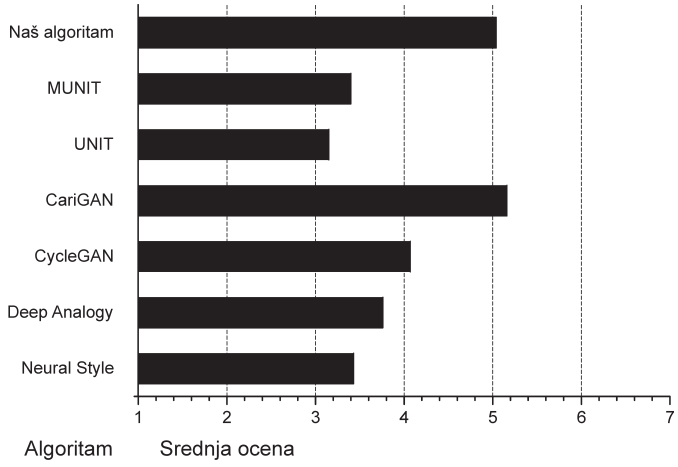


Slika 13.
Prikaz rezultata prve ankete: poređenje našeg algoritma sa ručno rađenim karikaturama i drugim sličnim algoritmima

Figure 13.
The first survey results: comparison of our algorithm (first row) with hand-drawn caricatures (second row) and other similar algorithms

2. Druga anketa je rangirajućeg tipa. Sadrži 30 nasumično izabranih fotografija portreta poznatih ličnosti, koje su propuštene kroz sledeće algoritme: Neural Style (Gatys *et al*. 2015), Deep image analogy (Liao *et al*. 2017), jedno-modalne mreže za prevođenje slike CycleGAN (Zhu *et al*. 2016) i UNIT (Liu *et al*. 2017), multi-modalna mreža za prevođenje slika MUNIT (Huang *et al*. 2018), CariGan (Cao *et al*. 2018) i naš algoritam. Ovoj anketi je pristupilo 11 učesnika, od kojih se zahtevalo da svih 94 portreta poznatih ličnosti, rangiraju od najbolje do najlošije ponuđene karikature različitih generatora.

Anketa je tražila rangiranje karikatura od 1 do 7 (gde 1 predstavlja najlošiju karikaturu, a 7 najbolju). Konačni rezultati predstavljeni su kao aritmetička sredina ukupnih ocena, a potom i zaokruženi na jednu decimalu (slika 14). CariGAN algoritam se pokazao najuspešnijim sa prosečnom ocenom 5.2. Odmah posle, sa 5.0 je naš algoritam što je takođe očekivano na osnovu rezultata prve ankete. Zatim slede algoritmi CycleGAN, Deep image analogy, Neural Style sa srednjim vrednostima od 4.1, 3.8, 3.4,



Slika 14.
Uporedni prikaz sumiranih rezultata druge ankete (poređenje našeg algoritma isključivo sa drugim algoritmima)

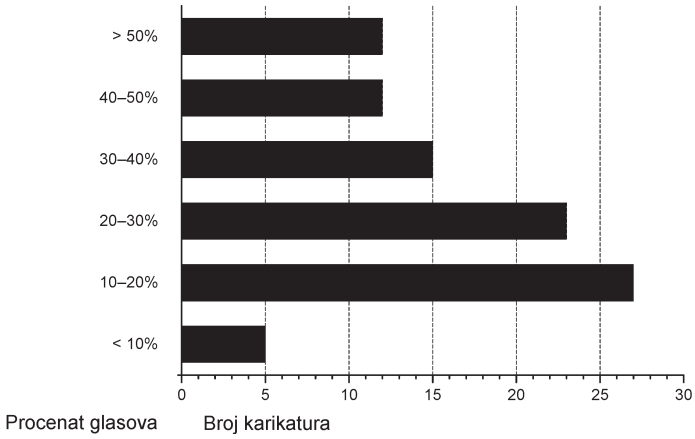
Figure 14.
Histogram of the second survey results (comparison of our algorithm exclusively with other algorithms)

Slika 15.

Slike korišćenje prilikom poređenja naših karikatura (poslednja kolona) sa drugim algoritmima

Figure 15.

Images used when comparing our own caricatures (last column) with other algorithms



Slika 16.

Prikaz sumiranih rezultata treće ankete: nezavisna evaluacija kvaliteta naših karikatura

Figure 16.

The third survey results: independent quality evaluation of our caricatures

respektivno. Na poslednjem mestu su MUNIT i UNIT algoritmi sa prosečnim ocenama 3.4 i 3.2.

3. Treća anketa sadrži samo karikature generisane našim algoritmom. Na ovaj način dobijena je sveukupna vizuelna ocena kvaliteta naših karikatura, bez poređenja sa drugim referentnim uzorcima. Sastoji se iz dva dela. U prvom delu, nasumično izabranih 94 portreta poznatih ličnosti propušteno je kroz naš generator. Anketu je popunilo 18 osoba, od kojih se zahtevalo da izaberu neodređen broj zadovoljavajućih karikatura. Za svaku pojedinačnu sliku računat je procenat ljudi koju je nju odabralo kao uspešnu. Broj karikatura koje su ostvarile više od 50% glasova iznosio je 12, karikatura koje su ostvarile više od 40% takođe 12, onih koje su osvarile više od 30% bilo je 15, više od 20% – 23, više od 10% – 27, a onih sa manje od 10% bilo je 5 (slika 16). Najuspešnija karikatura ima osvojenih 59% glasova. Prosečan broj glasova po karikaturi je 24%. Za najbolje karikature su se pokazala lica sa najmanjim brojem vidljivih oštećenja od geometrijske transformacije (najčešći problemi su granica čela i obraza; slika 17).

U drugom delu ankete predstavljeno je 20 vrsta stilova koji su primenjeni na istoj karikaturi. Stilovi su odabrani tako da imamo dve kategorije

Slika 17.
Primeri najuspešnijih i najlošijih karikatura, prvi i drugi red, respektivno

Figure 17.
Examples of the most and least successful caricatures, first and second row, respectively

estetskih izgleda konačne karikature: apstraktni stil, gde su izraženi geometrijski oblici i visoki kontrasti (najčešće korišćene slike stila su iz perioda kubizma), i stil „prirodnijeg izgleda", gde je boja kože očuvana, a tekstura minimalno izražena (slika 18). Karikature različitih stilova su nasumično prikazane u anketi, kako ne bi došlo do prirodnog „navikavanja" na određeni tip stila. Učesnici su pitani da, po svom nahođenju, izaberu 5 najprikladnijih stilova. Najboljim su se pokazali stilovi iz realistične kategorije, odabrani čak 76.5% više nego iz apstraktne. Ovakav rezultat je očekivan, i može se objasniti činjenicom da su kartunizovani i realistični stil najdominantniji u oblasti umetnosti karikatura, dok za prihvatanje apstraktnijih stilova treba doći do kompleksnijih promena oblika, minimalizma i simbolizma, što trenutno nije pokriveno od strane algoritma.

Slika 18.
Slike korišćene prilikom poređenja stilova (realistični i apstraktni stilovi)

Figure 18.
Results of the second part of the third survey (realistic and abstract styles)

Zaključak

Naš algoritam za generisanje karikatura unapređuje postojeće algoritme u pogledu vizuelnog kvaliteta i očuvanja identiteta. Međutim, predstavljeni pristup i dalje sadrži ograničenja. Prvo, geometrijska transformacija očiglednije se primećuje u obliku lica (većim površinama deformacije), nego na drugim crtama lica npr. ušima, dlakama, borama, pa čak i čelu. Ovaj problem zasniva se na ukupnom broju korišćenih karakterističnih tačaka (orijentira). Ograničenje se može rešiti dodavanjem dodatnih orijentira. Za dalji napredak projekta, moguće je iskoristiti metode drugih referentnih radova. Metodiku koju bismo istakli i probali u daljem razvoju je tehnika ručnog labeliranja tačaka uz asistenciju AAM (Active Appearance Model, Mo *et al*. 2004). Na ovaj način, očuvali bismo mogućnost konvertovanja tačaka u matricu za dalju faktorizaciju, a istovremeno dobili veću fleksibilnost za komičnu transformaciju sa dodatnim orijentirima (konkretno, u pomenutom referentnom radu su korišćene 94 tačke, ali verujemo da se i taj broj može korigovati radi još boljih rezultata).

Što se tiče stilizacije, naši rezultati su verni referentnim stilovima koji su uobičajeni za bazu podataka „realističnih", pa i „apstraktnih" karikatura, ali su manje verni stilovima koji su očekivani za ovu oblast umetnosti (npr. crtež ili skica). Ovaj problem se može zaobići činjenicom da korisnici imaju potpunu slobodu izbora slike stila. Međutim, odabir slike stila često može biti neintuitivan proces, sa rezultatima koji se dosta razlikuju od očekivanih. Naša je pretpostavka da je problem u načinu rada neuronske mreže (kako izračunava i prepoznaje kontekst slike i teksture stila), stoga verujemo da preciznije treniranje na ograničenom uzorku karikatura i drugih sličnih likovnih dela može pomoći u pronalasku zadovoljavajućeg stila. Ovakav pristup je isproban neuronskom mrežom CariStyGAN (Cao *et al*. 2018), ali u završnim zaključcima su napomenuta ograničenja koja naš trenutni model nema (npr. nedostatak fleksibilnosti odabira stila od strane korisnika).

Konačno, predstavljeni pristup automatskom generisanju karikatura može se smatrati relativno uspešnim i konkurentnim drugim automatskim sistemima generisanja karikatura. Ipak, automatski sistemi generisanja karikatura i drugi automatski sistemi generisanja umetničkih dela još uvek nisu ni blizu ljudskim stvaralačkim mogućnostima. Umetničko delo mora nastati iz namere, i mora da prenosi emociju, a za to mašine, bar još uvek, nisu sposobne. S druge strane, umetnost održava svoju vitalnost stalnim inovacijama, a tehnologija je jedan od glavnih pokretača trenutne inovacije. Danas se možemo susresti sa brojnim intrigantnim eksperimentima nad tehnikama veštačke inteligencije, koje će, kao umetnički alati, zasigurno promeniti način na koji razmišljamo o umetnosti.

Literatura

Akleman E. 1997. Making caricatures with morphing. U *Proceedings: The art and interdisciplinary programs of SIGGRAPH '97*. ACM, str. 145.

Akleman E., Palmer J., Logan R. 2000. Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes. U *Advances in Visual Information Systems – Proceedings of Visual'2000* (ur. R. Laurini). Springer, str. 165–70.

Brennan S. E. 1985. Caricature generator: the dynamic exaggeration of faces by computer. *Leonardo*, **18** (3): 170.

Cao K., Liao J., Yuan L. 2018. CariGANs: Unpaired photo-to-caricature translation. *ACM Transactions on Graphics*, **37** (6): 244.

Chen H., Zheng N., Liang L., Li Y., Xu Y-Q., S H-Y. 2002. PicToon: A personalized image-based cartoon system. *Proceedings of the ACM International Multimedia Conference and Exhibition*. New York: Association for Computing Machinery, str. 171-78.

Efros A. 2007. Computational Photography in CMU. 15-463 http://graphics.cs.cmu.edu/courses/15-463/2007_fall/Lectures/morphing.pdf

Gatys L. A., Ecker A. S., Bethge M. 2015. A Neural Algorithm of Artistic Style. arXiv: 1508.0657v1

Huang X., Liu M. Y., Belongie S., Kautz, J. 2018. Multimodal unsupervised image-to-image translation. U *Proceedings of the European Conference on Computer Vision (ECCV)*, str. 172-189.

Frossard D. 2016. VGG in TensorFlow: Model and pre-trained parameters for VGG16 in TensorFlow. http://www.cs.toronto.edu/~frossard/post/vgg16/

Koshimizu H., Tominaga M., Fujiwara T., and Murakami K. 1999. On KANSEI facial image processing for computerized facial caricaturing system PICASSO. U *Proc. IEEE International Conference on Systems, Man, and Cybernetics,* **6**. IEEE, str.294–99.

Lee D-D., Seung H-S. 1999. Learning the parts of objects by nonnegative matrix factorization. *Nature*, **401**: 788.

Liao J., Yao Y., Yuan L, Hua G., Kang S. B. 2017. Visual attribute transfer through deep image analogy. *ACM Transactions on Graphics*, **36** (4): 120.1.

Liu M., Breuel T., Kautz J. 2017. Unsupervised image-to-image translation networks. U *Advances in Neural Information Processing Systems – NIPS 2017* (ur. I. Guyon *et al*.). Red Hook: Curran Associates, str. 700–708.

Mo Z., P Lewis J. P., Neumann U. 2004. Improved automatic caricature by feature normalization and exaggeration. U *ACM SIGGRAPH 2004 Sketches*. ACM, str. 57.

Narayanan H. 2017. Convolutional neural networks for artistic style transfer. https://harishnarayanan.org/writing/artistic-style-transfer/

Pérez P., Gangnet M., Blake A. 2003. Poisson image editing. U *ACM SIGGRAPH 2003 Papers*. ACM, str. 313-18.

Yi Z., Zhang H., Tan P., and Gong M. 2017. Dualgan: Unsupervised dual learning for image-to-image translation. arXiv preprint

Zhu S., Li C., Loy C., Tang X. 2016. Unconstrained face alignment via cascaded compositional learning. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016*. IEEE, str. 3409–17.

Lana Popović and Jana Marković

Automatic Caricature Generation using Non-negative Matrix Factorization and Convolutional Neural Networks

The goal: The automatic generation of caricature drawings which will fulfill the desired artistic, humoristic and stylistic norms while maintaining the person's likeness and augmenting their characteristic attributes.

The method: Based on the input of the sample consisting of 122 pictures, 68 characteristic points (landmarks) are calculated. With a non-negative factorisation of the matrix we obtain the basic characteristics and their average distributions. The singular values derived from every face are compared with the basic ones and, based on that comparison, are enlarged. The resulting images with the deformed fragments are run through an already trained convolutional neural network, along with the artworks whose style it replicates.

The results: To determine the success of our algorithm three types of surveys were conducted. The first survey demanded the comparison of the caricature drawings derived from our algorithm with those produced by similar algorithms and those drawn by people and to choose suitable ones based on attractiveness, artistic style and humour. In the second survey the participants were asked to rank from best to worst different caricature rendering algorithms. The third survey, which involved only caricatures produced with our algorithm, consisted of two parts. In the first part the participants were asked to select an indefinite number of caricatures which were according to them, successful. The second part consisted of a number of different styles implemented into one caricature. The people involved in the survey were asked to select the five most appropriate ones.