

МАТЕМАТИЧКА ГИМНАЗИЈА

У БЕОГРАДУ

МАТУРСКИ РАД

из информатике и рачунарства

**АУТОМАТСКО ГЕНЕРИСАЊЕ КАРИКАТУРА
НЕНЕГАТИВНОМ ФАКТОРИЗАЦИЈОМ МАТРИЦЕ И
КОНВОЛУЦИОНАЛНИМ НЕУРОНСКИМ МРЕЖАМА**

ментор: Ивана Јовановић Мاستиловић

ученик : Лана Поповић IVб

Београд, мај 2020.

САДРЖАЈ

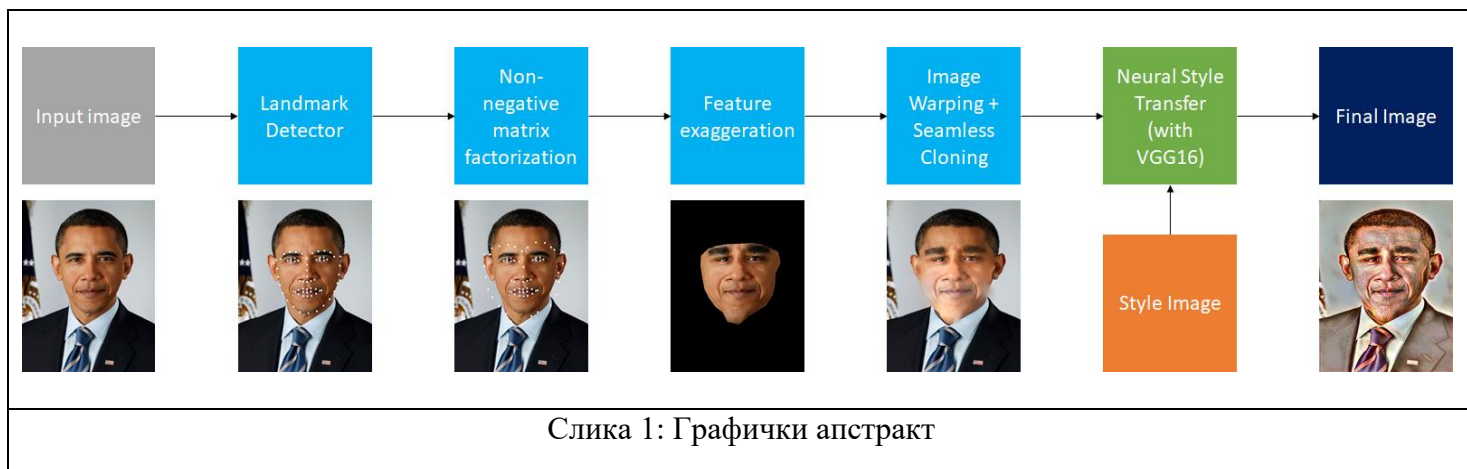
АПСТРАКТ	3
УВОД	3
МЕТОДА	4
Детекција карактеристичних тачака	4
Проблем мапирања чела и горње границе лица.....	5
Ненегативна факторизација матрице	8
Кривљење слике – Деланијева триангулација	9
Неурални трансфер стила.....	11
ИСТРАЖИВАЊЕ И РЕЗУЛТАТИ	16
ЗАКЉУЧАК	19
ЛИТЕРАТУРА	20

Апстракт

Циљ: Аутоматско генерисање карикатура које ће испунити жељене уметничке, хумористичке и стилске норме одржавајући оригинални идентитет особе и преувеличавајући њене карактеристичне атрибуте.

Метода: На убаченом узорку од 122 слике израчунава се 68 карактеристичних тачака (ландмаркова). Ненегативном факторизацијом матрице добијају се базне карактеристике и њихове просечне расподеле. Појединачне вредности сваког лица се упоређују са базним, и на основу тог односа појачавају. Добијене слике са деформисаним (наглашеним) фрагментима се пропуштају кроз већ тренирану конволуциону неуронску мрежу заједно са уметничким делима чији стил ће се на тај начин преузети.

Истраживање и резултати: За одређивање успешности овог алгоритма спроведена су три типа анкете. Прва анкета је захтевала да се од различитих карикатура које су генерисане овим и другим сличним алгоритмима или насликане од стране човека одаберу задовољавајуће, у аспекту привлачности, стила и хумора. У другој анкети се од учесника тражило да рангирају од најбољег до најлошијег различите алгоритме генерисања карикатура. Трећа анкета, у коју су укључене само карикатуре генерисане овим алгоритмом, садржала је два дела: у првом делу корисници су питани да одаберу неодређен број карикатура које су, према њиховом мишљењу, успешне. Други део се састојао од више различитих стилова примењених на једној карикатури, а од корисника се тражило да изабере пет најприкладнијих.



Слика 1: Графички апстракт

Увод

Карикатура портрета је комбинација карикатуре и портрета са фокусом на карикирању особе – пренаглашавању неке физичке карактеристике лица. Обично се користе ради забаве, као поклони или сувенири, често нацртане од стране уличних уметника. Карикатуристи имају способност уочавања уникатних црта лица, која потом преувеличавају и стилизују. У последњих неколико година рачунарска визија и примена неуронских мрежа је изузетно напредовала – тако је у томе и уметност недавно нашла своје место. Аутоматско генерисање карикатура се користи у мрежним комуникацијама, онлајн играцама и у индустрији анимације. Циљ овог рада је учинити аутоматско генерисање карикатура што реалистичније људском раду – постигнути веродостојност, стил и шаљивост на генерисаним карикатурама.

Проблем генерисања карикатура је више пута адресиран, и може му се приступити на различите начине.

Било је покушаја интерактивног синтетисања карикатуре [Akleman 1997; Akleman et al.2000; Chen et al.2002;Gooch et al.2004], што је захтевало професионалне вештине за постизање експресивних резултата. Предложено је и неколико аутоматских система, али су се они ослањали на ручно одређена правила [Brennan 2007; Koshimizu et al.1999;Liang et al.2002; Mo et al.2004], која често произлазе из поступака цртања уметника. Нажалост, ови приступи су ограничени на одређени уметнички стил (попут скицирања графитном оловком или „картунизовани“ - *cartoon* стил) и унапред одређене шаблоне наглашавања.

Последњих година, дубоко учење, као представник технике учења на примерима (посебно на великим количинама података), успешно је коришћено за превођење слике у слику [Hinton and Salakhutdinov 2006; Huang et al.2018; Isola et al.2017; Kim et al.2017; Liu et al.2017; Yi et al.2017; Zhu et al.2017b]. Међутим, за сада не постоји довољна количина упарених карикатура и одговарајућих фотографија, па су тренирања са надзором (попут аутоенкодера) неизводљива.

Најбоље резултати постигли су Kaidi Cao, Jing Liao, Lu Yuan креирањем две Генеративне Адверсалне Неуронске мреже CariGeoGAN и CariStyGAN које би, респективно, училе геометријску трансформацију и транспоновање стила. Међутим, мана овог поступка је недостатак варијабилности код наглашавања лица. Суштина карикатуре није било каква промена, већ изражавање оних делова који су карактеристични за конкретну особу.

У овом раду представљено је коришћење нове методике изражавања одступања од просечног (“Exaggerating the Difference from the Mean” (EDFM)) која решава проблеме CariGAN-а. Суштина је проналажење „просечног лица“ које се користи као основа при упоређивању задатог лица од кога ће настати карикатура. Делови лица који највише одступају од просека (стога су и најпрепознатљивији за ту особу) се додатно наглашавају. На овај начин, уникатност сваког лица је сачувана. Потом се на слику примењује трансфер стила изабраног уметничког дела, коришћењем дубоке конволуционе неуронске мреже, што додатно додаје варијабилност и флексибилност нашим карикатурама.

Метода

За успешно креирање карикатуре потребно је испунити два захтева – адекватно нагласити одређене карактеристике лица и задовољити стилска очекивања уметничког дела. Стога је и овај рад подељен у две претходно наведене целине. Код је куцан у Python-у.

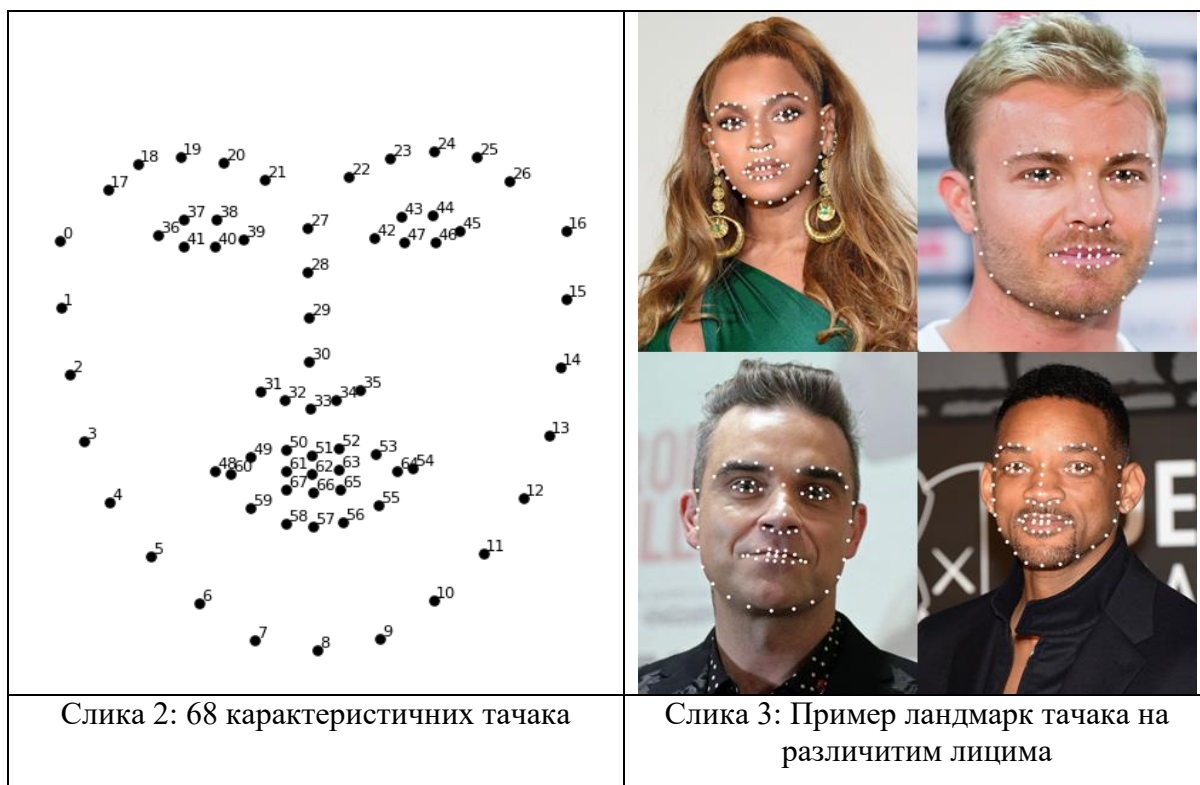
Детекција карактеристичних тачака

Претходно је потребно пронаћи карактеристичне тачке лица. За то је коришћен истренирани модел за аутоматску детекцију и мапирање лица („ландмарк”детекцију) у Python-у са библиотеком *dlib*. Успешно детектује 68 ландмарк координата (x,y) на било којем лицу:

1. Обележја браде (0-16)
2. Обележја десне обрве (17-21)

3. Обележја леве обрве (22-26)
4. Обележја носа (27-35)
5. Обележја десног ока (36-41)
6. Обележја левог ока (42-47)
7. Обележја уста (48-68)

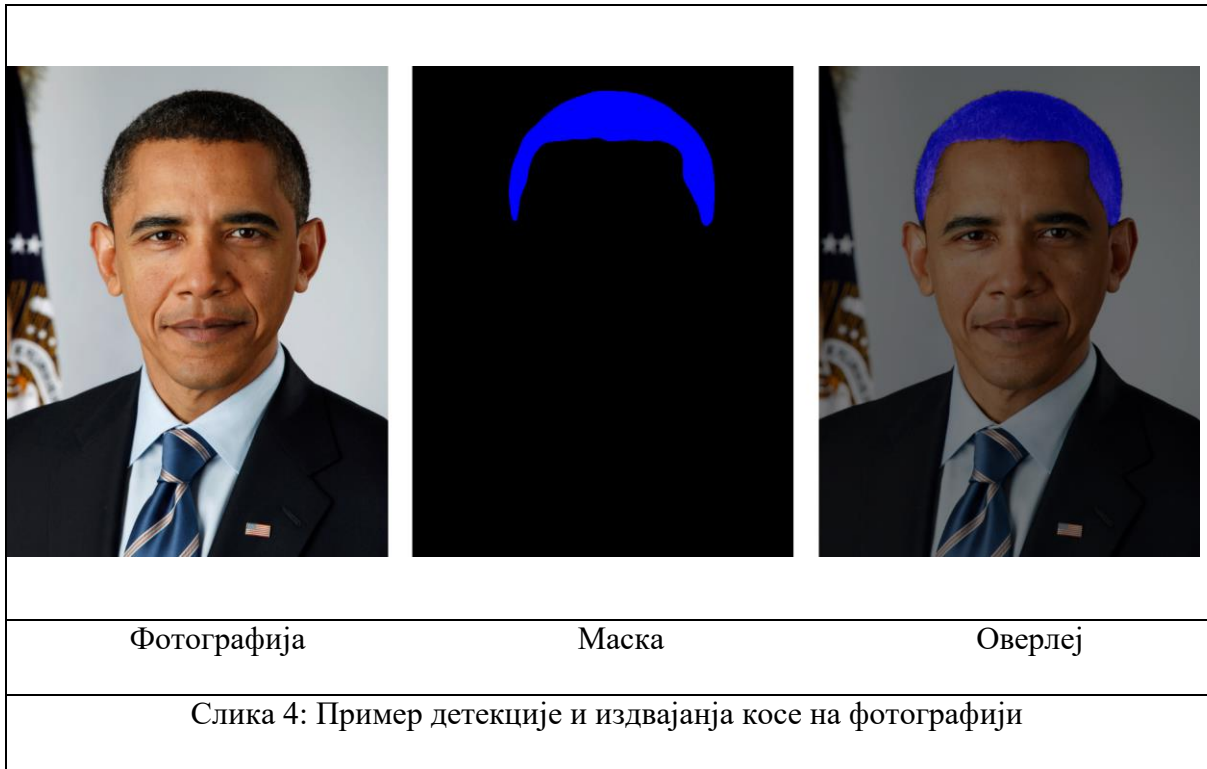
Кроз овај модел је пропуштен узорак од 122 фотографије, које су коришћене за проналажење средњег лица. За сваку карактеристичну тачку лица, на свакој од фотографија, рачуна се удаљеност од осталих појединачних тачака. Затим се налази аритметичка средина раздаљина између парова тачака обележених истим бројем на узорку. Средње (базно) лице представља мапу потрета чије су карактеристичне тачке постављене тако да одговарају средњим вредностима.



Проблем мапирања чела и горње границе лица

Мана претходно поменутог модела за аутоматску детекцију и мапирање лица је недостатак карактеристичних тачака које би обележиле горњу границу портрета и чело особе. Последица тога је непроналазак средње димензије чела, што онемогућава његову карикатуризацију. Зато је покушано додавање карактеристичних тачака које би пратиле горњу линију лица.

Идеја је била коришћење већ трениране конволуционе неуронске мреже чији би задатак био детектовање, а потом и издвајање, свих пиксела неке слике на којима се налази коса. На тај начин би се добила „маска” косе, на коју би се наставило лице.



Додатних пет карактеристичних тачака би се добило повлачењем правих линија кроз:

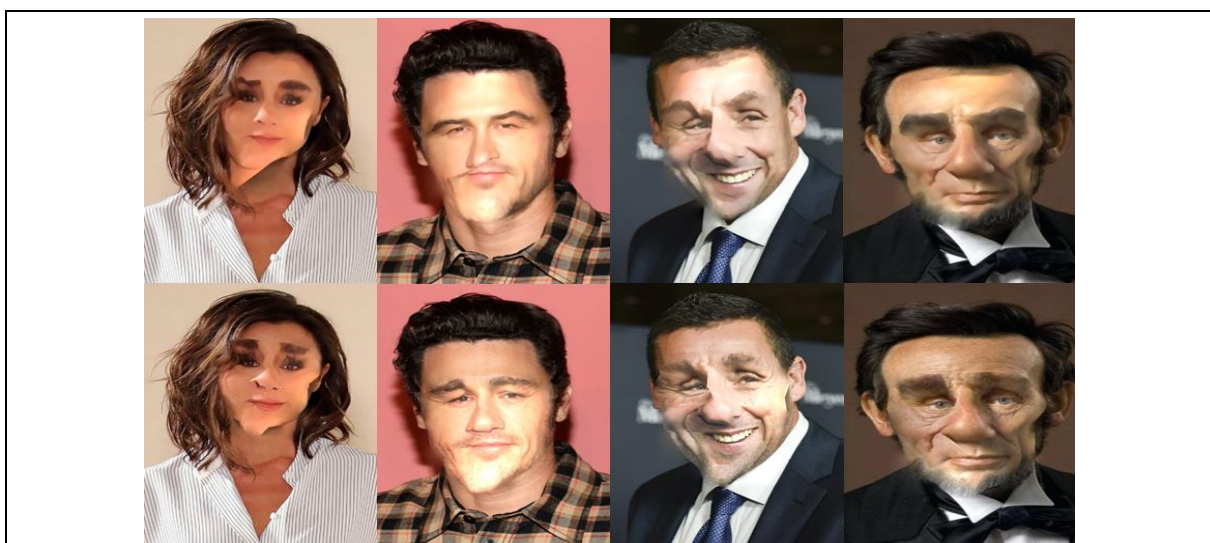
- тачке 30 и 27. Ова права представља уздужну осу симетрије лица и пролази кроз средину чела. Најнижа заједничка тачка „маске” косе и дате праве дефинисала би прву додатну карактеристичну тачку.
- тачке 44 и 24. Ова права спаја центар десног ока и десне обрве, и у најнижој пресечној тачки са „маском” косе поставила би се друга додата карактеристична тачка.
- тачке 37 и 19. Ова права спаја центар левог ока и леве обрве, и у најнижој пресечној тачки са „маском” косе поставила би се трећа додата карактеристична тачка.
- тачке 45 и 36. Ова права спаја спољашњи угао десног ока са завршетком десне обрве, и њена најнижа заједничка тачка са „маском” косе би дефинисала место постављања четврте додате карактеристичне тачке.
- тачке 36 и 17. Ова праја спаја спољашњи угао левог ока са завршетком леве обрве, и њена најнижа заједничка тачка са „маском” косе би дефинисала место постављања пете додате карактеристичне тачке.



Слика 5: Илустрација одређивања додатних карактеристичних тачака

Овакава метода издвајања не даје увек задовољавајуће резултате, како у неким случајевима погрешно детектује делове лица као косу (углавом ако постоји јача сенка на том региону) или детектује део позадине (уколико су сличних боја). Такође, овај принцип није ефикасан уколико особа нема косу, има браду, носи капу, или на неки други начин прекрива косу. Ипак, највећу препреку представља чињеница да се додавањем додатних пет тачака, насупрот претпостављеном, не примећује значајна промена, а она која и постоји је лошија у односу на резултат добијен изостављањем додатних тачака. Закључак је да обележавање горње границе лице ограничива и сужава простор за промене, тиме погоршавајући и сам квалитет и аутентичност добијене карикатуре.

Због свега наведеног, експеримент се сматра неуспешним, а мапа лица задржава почетни изглед, без уврставања додатних тачака. Наш генератор за сада нема могућност карикирања делова портрета изнад обрва.



Слика 6: Упоредивање резултата са 73 и 68 ландмарк тачака (први и други ред, респективно)

Ненегативна факторизација матрице

Ненегативна факторизација матрице (*Non-negative matrix factorization*) је група алгоритама која има задатак да матрицу V факторише у две одвојене матрице W и H , са условом да све три матрице немају негативне чланове. Ова ненегативност чини резултирајуће матрице лакшим за инспекцију. НМФ методом димензије добијених матрица фактора могу бити знатно мање од димензија матрице производа. На пример, ако матрица V има димензије $n \times m$, матрица W $n \times r$ и матрица H $r \times m$, r може бити доста мање од m и n .

За потребе овог рада, конструисана је матрица S димензије 122×136 , где n димензија представља број слика, а m димензија x и y у координате коришћених 68 ландмарк тачака. На добијену матрицу S се примењује НМФ [Lee and Seung 1999.] како би научила димензије лица:

$S = F * E$	(1)
-------------	-----

S – матрица лица $n \times m$

F – фактор матрице S који тражимо, представља матрицу базних карактеристика

E – фактор матрице S који тражимо, представља матрицу енкодинг-а тј. коефицијената који се користе у линеарној комбинацији са базним карактеристикама из матрице F како би се добила коначна матрица производа S . Димензије матрица F и E су $n \times r$ и $r \times m$, респективно (у нашем случају број компонената $r = 65$).

Базне карактеристике (features) НМФ-а и енкодинг коефицијенти су неправилно распоређени, тј. базни фичери се састоје од више различитих позиција ландмарк-а уста, носа и других делова лица, где су различите верзије на различитим локацијама. Цело лице генерише се комбиновањем ових различитих делова. Енкодинг коефицијенти су тако распоређени да се никад не догоди стапање више позиција очију, обрва итд. Овом методом се не добија егзактна првобитна матрица S , већ најближа апроксимација.

У нашој формули факторизоване матрице $S = F * E$, свака димензија се састоји од базних вектора f_i и његове расподеле e_i – која је представљена као $(m_i + \sigma_i)$, где је m_i средња вредност i -те колоне матрице E , а σ_i стандардна девијација i -те колоне матрице E . Свака димензија представља позицију апстрактног дела лица.

Облик s на новој слици представимо као линеарну комбинацију базних вектора f_i , њихових заступљености у простору e_i и вектора остатка r :

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

δ_i – стандардна девијација од средње вредности i -те колоне матрице E , аналогно σ_i

Код карикирања лица постоји неколико случајева где се највише треба фокусирати на $\delta_i = |e_i - m_i|$:

$$\vec{s} = \sum_i (m_i + t_i * \delta_i) * \vec{f}_i + 0.8 * k * \vec{r}$$

(3)

1. Нека t_i и k буду коефицијенти карикирања тј. повећања δ_i и вектора остатка r , у првом случају када је $|\delta_i| < \sigma_i$ – ставићемо да је $t_i=1$;

2. А, у случају да је $|\delta_i| \geq \sigma_i$, $t_i = k$ (у нашем случају, $k = 1,25$)



Coeff: 1,2

Coeff: 1,3

Coeff: 1,4

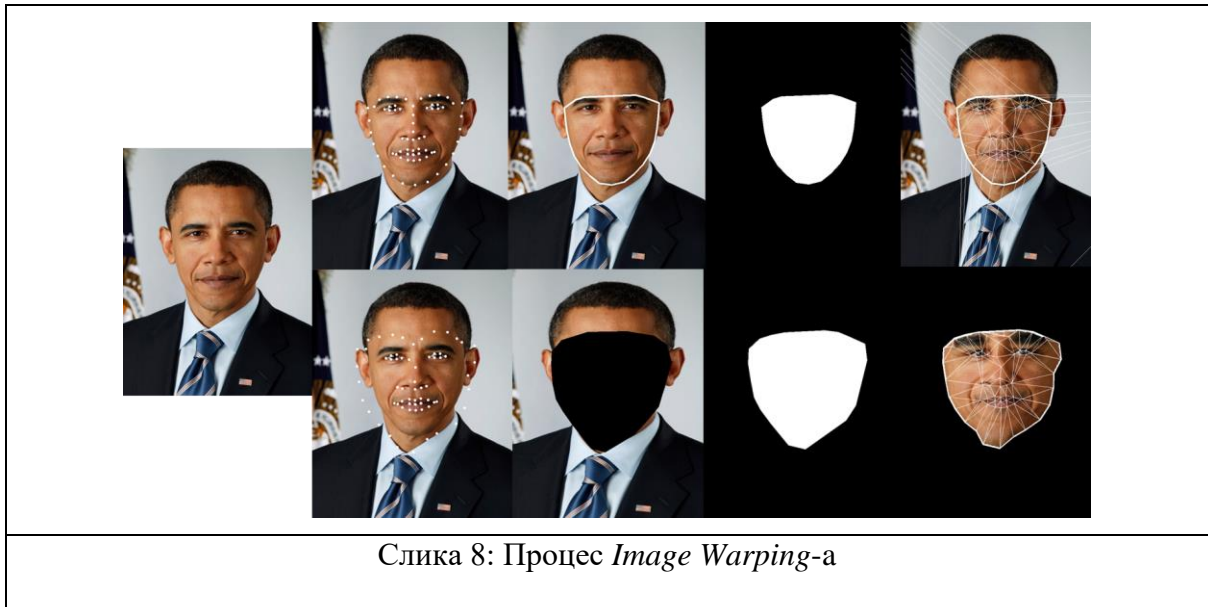
Слика 7: Промена коефицијента деформације k

Пошто свако лице има остатак, коефицијент промене је додат и на вектор остатка, само у умањеној количини (у овом случају, узет је коефицијент 0,8 – може се повећати или смањити, у зависности од жељеног резултата (драстичнија промена, или не)).

Коначно, добијена је финална матрица S' димензије $1*136$ са новим координатама апстрактних искаркираних делова лица. Комплетна фотографија се добија *image warping*-ом (кривљењем слике) оригиналног облика S и промењеног S' .

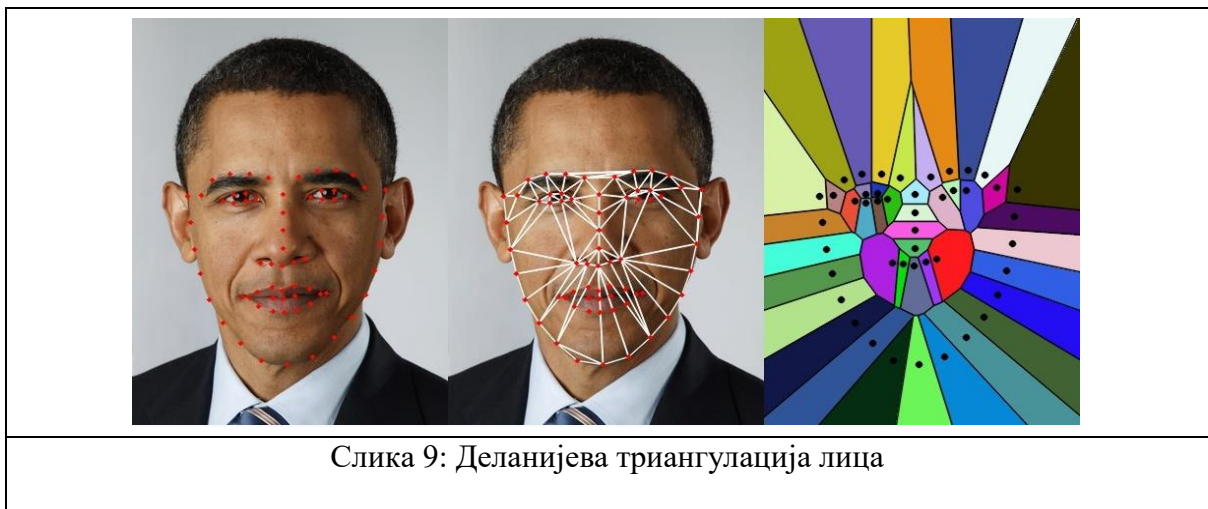
Кривљење слике – Деланијева триангулација

Након добијања две матрице S и S' (матрица са оригиналним координатама тачака на лицу и матрица са трансформисаним, респективно), траже се конвексни омотачи (полигони) ландмарк тачака од којих се праве маске које ће обухватати простор на којем ће се извршити Деланијева Триангулација.



Слика 8: Процес *Image Warping*-а

Деланијева Триангулација је подела равни на троуглове за одређени скуп P дискретних тачака (конкретније за овај рад, коришћењем координата из оригиналне матрице S) тако да ниједна тачка P није унутар обода било којег троугла. Деланијеви троуглови максимизирају минимални угао свих углова троугла. Како би *warping* био успешан, индекси ландмарк тачака из обе матрице морају да одговарају једни другима, тј. да сваки креирани деланијев троугао оригиналне S матрице кореспондује деланијевом троуглу S' матрице искаркираних координата.



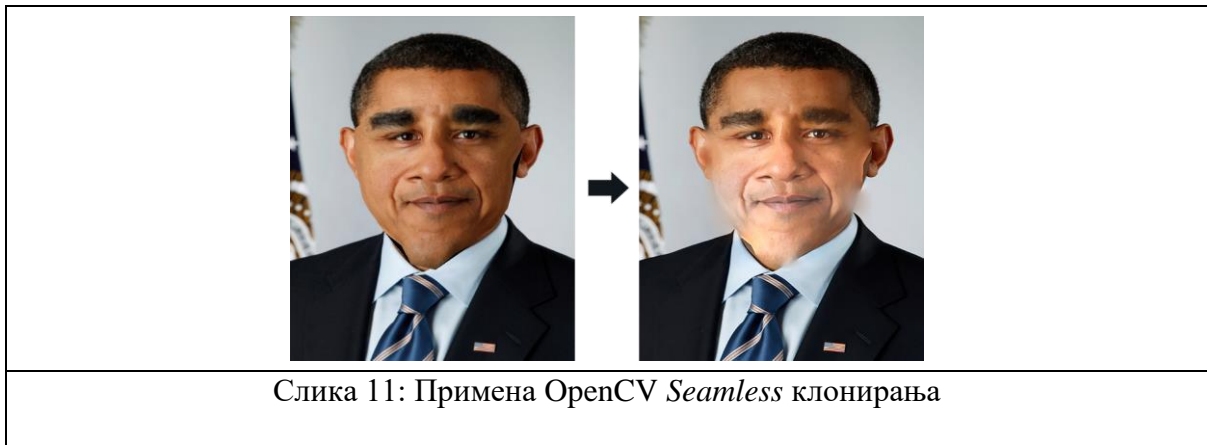
Слика 9: Деланијева триангулација лица

Након одрађивања триангулације обе матрице, треба да се извуку добијени троуглови и развуку или сузе их на основу дестинације коначних искаркираних координата матрице S' . Резултат је искаркиран облик лица.

Потом, оригиналан полигон лица без искривљених координата се сече и додаје се новодобијени облик лица како би се добила коначна слика.



Коначно, примењује се *Seamless Cloning* оригиналне фотографије и карикиране слике како би се уклониле било какве непотребне линије и пиксели настали оштром триангулацијом. *Seamless Cloning* је OpenCV имплементација Поасонове функције расподеле на пикселе слика како би постигла природан ефекат оверлеј-а.



Неурални трансфер стила

Неурални трансфер стила (*Neural style transfer*) је метод пребацивања карактеристика једне слике на другу коришћењем већ трениране конволуционе неуронске мреже. Идеја је да се на основу две учитане слике добије нова, комбинација предходне две, која ће преузети карактеристике стила (попут боје и текстура) са једне, и карактеристике контекста (попут објеката и њихове расподеле, „радње” слике) са друге.



Да би се уопште могло мерити колико је коначна слика различита од улазне (односно колико је контекста преузето из контекстне слике, а колико стила из стил слике) потребно је дефинисати такозвану *loss* функцију (функција губитка). Циљ нам је да се коначна слика разликује што мање од улазних (да губитак буде минималан), односно да вредност функције губитка буде приближна нули. Функцију губитка можемо записати као:

$L(G) = \alpha L_{\text{content}}(C, G) + \beta L_{\text{style}}(S, G)$	(4)
---	-----

где је **L** функција губитка, **G** генерисана слика, **C** слика са које се узима контекст, а **S** слика са које се узима стил. α и β су „фактори тежине“ који нам дозвољавају да контролишемо колико ће контекст бити наглашен у односу на стил.

Конволуционе неуронске мреже се састоје од слојева малих рачунарских јединица (неурона) које обрађују визуалне информације хијарарихијално у „feed forward“ маниру (информације се крећу нециклично). Сваки слој неурона се може представити као колекција филтера за слику, од којих је сваки задужен да са слике извуче одређену карактеристику. Резултат након сваког слоја је такозвана *feature map* – мапа карактеристика (различито филтрирана верзија почетне слике).

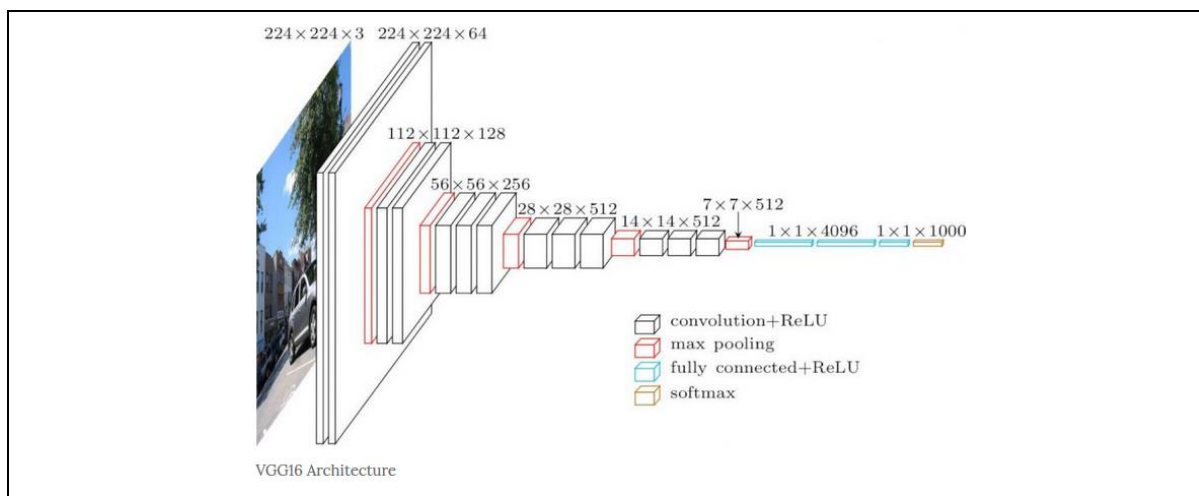
Када се конволуционе мреже тренирају да препознају објекат, развијају мапе карактеристика које наглашавају одређене информације о објекту дуж хијерархије слојева. Током обраде, улазна слика се трансформише у репрезентацију која све више брине о стварном садржају слике. Виши слојеви манипулишу садржајем вишег нивоа у односу на предходни, у смислу објеката и њиховог распореда на слици (на пример, у почетним слојевима *feature map* се састоји из пиксела који приказују косу линију у сличним бојама, док се у дубљим слојевевима већ могу пронаћи компликовани бјекти попут људи или животиња).



Слика 13: Визуализација слојева према дубини – дубљи слојеви се баве компликованијим објектима

Да бисмо умањили недостатке које би имало посматрање губитака по пикселу и да бисмо дозволили нашим функцијама губитака да боље „измере” перцептивне и семантичке разлике између слика користимо принцип генерисања слика помоћу оптимизације. Кључ ове методе је коришћење конволуционе неуронске мреже која је већ научила да дешифрује перцептивне и семантичке информације које бисмо желели да наше функције губитака мере. Зато користимо дубоку конволуциону мрежу **VGGNet** која је већ истренирана да класификује слике и помоћу ње дефинишемо наше функције губитка. Користимо VGG16 верзију, са 16 тежинских слојева (13 конволуционих и 3 FC слоја – *fully connected layers* (потпуно повезани слојеви)). Улаз за први конволутивни слој је 224×224 RGB слика. Слика пролази кроз групу конволутивних слојева, где су коришћени филтери са врло малим рецептивним пољем: 3×3 (што је најмања величина довољна да ухвати појам лево/десно, горе/доле, центар).

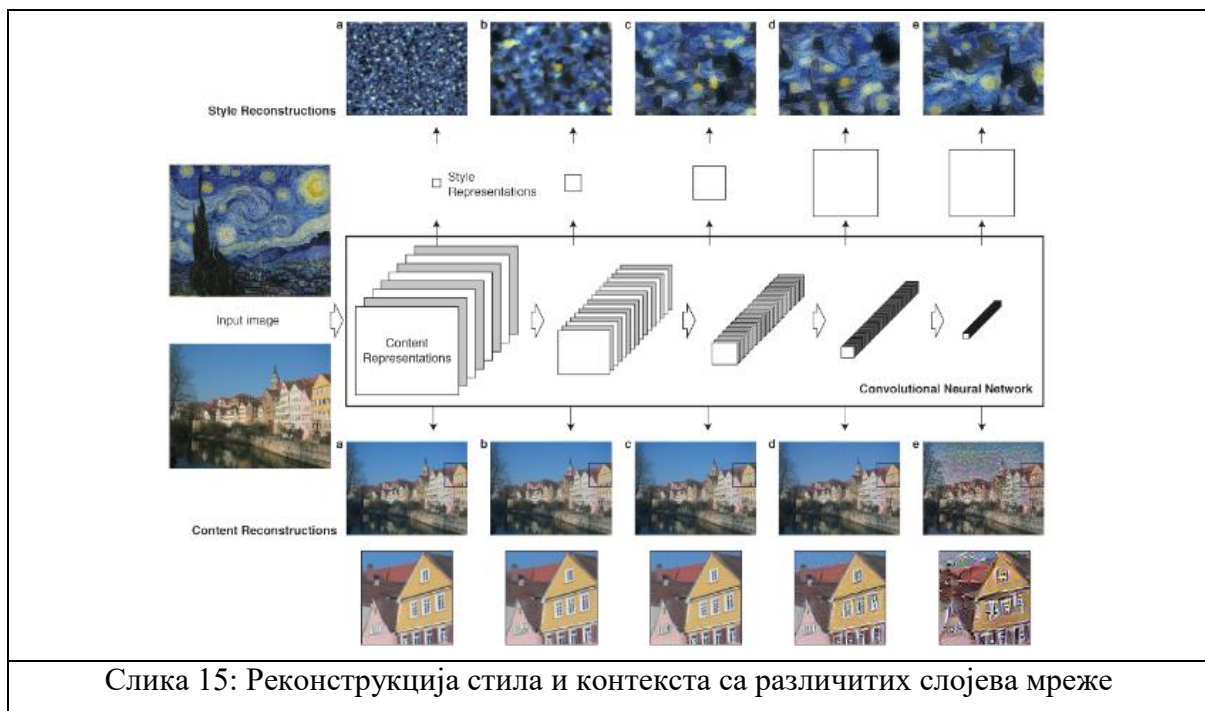
Ова мрежа је тренирана на бази побадатака од преко 15 милиона лаберираних слика високе резолуције, које се могу разврстати у тачно 22000 категорија. За тренирање комплетне мреже коришћено је преко 138 милиона параметра. За наше потребе, 13 конволуционих слојева остају непромењени. Узимамо излаз са 13. слоја и користимо последња 3 да дефинишемо функције губитка. Ова метода припада групи учења без надзора (*unsupervised learning*).



Слика 14: визуелизација VGG16 неуронске мреже

Конволуциони слојеви се користе за дефинисање две функције губитака: функција губитка контекста ($L_{content}$) и функција губитка стила (L_{style}). Свака функција губитка рачуна скаларну вредност $L_i(y, y_i)$ која представља различитост између излазне слике y и циљане слике y_i .

Улазна слика има своју репрезентацију у виду филтрираних слика у свакој фази обраде. Док број различитих филтера расте дуж хијерархије, величина филтрираних слика се смањује због механизма за искључење (*max-pooling*), што доводи до смањивања укупног броја неурона по слоју. На слици 15. контекст почетне слике смо реконструисали са слојева: ‘conv11’ (a), ‘conv11’ и ‘conv21’ (b), ‘conv11’, ‘conv21’ и ‘conv31’ (c), ‘conv11’, ‘conv21’, ‘conv31’ и ‘conv41’ (d), ‘conv11’, ‘conv21’, ‘conv31’, ‘conv41’ ‘conv51’ (e) оригиналне VGG мреже. Приметимо да је реконструкција контекста у нижим слојевима скоро савршена (a,b,c), док су у вишим слојевима изгубљене детаљне информације о пикселима. Такође, реконструкција стила у вишим слојевима је боља, односно повећан је „интезитет“ стила, док се одбацују информације о глобалном уређењу сцене како идемо у дубину.



Слика 15: Реконструкција стила и контекста са различитих слојева мреже

Слој са N_L различитих филтера садржи N_L мапа карактеристика величине M_L , где M_L има величину висина пута ширина мапе карактеристика. Тако се одговори из слоја L могу сместити у матрицу $F^L \in \mathbb{R}^{N_L \times M_L}$ где је F^L_{ij} активација i -тог филтера на позицији j у слоју L . За визуализацију информација које се налазе на различитим слојевима хијерархије извршавамо алгоритам опадајућег градијента на *white-noise* слици, да би пронашли другу слику код које се одговори карактеристика поклапају са улазном сликом. Узмимо да је p оригинална слика, а x слика која се генерише, а P^L и F^L њихове респективне репрезентације карактеристика у слоју L . Затим дефинишемо квадратну грешку функције губитка контекста између две репрезентације карактеристика:

$L_{\text{content}}(\vec{p}, \vec{x}, L) = \frac{1}{2} \sum_{ij} (F_{ij}^L - P_{ij}^L)^2$	(5)
---	-----

Извод ове функције губитка по активацији у слоју L износи:

$\frac{\partial L_{\text{content}}}{\partial F_{ij}^L} = \begin{cases} (F_{ij}^L - P_{ij}^L) & \text{ако је } F_{ij}^L > 0 \\ 0 & \text{ако је } F_{ij}^L < 0 \end{cases}$	(6)
--	-----

где градијент по слици x може да се израчуна користећи стандардну бекпропагацију грешке (*error back-propagation*). То нам омогућава да итеративно мењамо насумично генерисану слику x , све док одговор одређеног слоја у мрежи не буде довољно сличан одговору оригиналне слике p .

Да би се реконструисао стил улазне слике, користимо *feature space* (оригинално дизајниран да прикупи информације о текстури) који градим на врху одговора филтера у сваком слоју. Репрезентација стила се рачуна на основу корелације између различитих карактеристика у различитим слојевима мреже. Укључивањем корелације карактеристика са више слојева, добијамо стационарну, мулти-скалирану репрезентацију учитане слике, која садржи информације о текстури, али не о глобалном распореду слике. Фактички, добијамо слику која је „слепа” на контексте карактеристике улазне слике. Ове корелације карактеристика су дате Грамовом матрицом $G^L \in \mathbb{R}^{N^L \times M^L}$, где G^L_{ij} је производ векторизоване мапе карактеристика i и j у слоју L:

$G^L_{ij} = \sum_k F^L_{ik} * F^L_{jk}$	(7)
---	-----

Грамова матрица (или матрица скаларних производа) за уређен скуп вектора (чланова векторског простора) је матрица чији су елементи скаларни производи вектора из датог скупа, у овом случају из скупа векторизоване мапе карактеристика одређеног слоја у мрежи.

За генерисање текстуре која одговара стилу дате слике, користимо *спуштање низ градијент* од шума да нађемо слику која одговара репрезентацији стила оригиналне слике. Ово је урађено минимизацијом средње квадратне удаљености између уноса Грамове матрице из оригиналне слике и Грамове матрице слике која треба бити генерисана. Узмимо да је a оригинална, а x генерисана слика, и A^L и G^L њихове респективне репрезентације стила у слоју L. Допринос тог слоја коначном губитку је онда:

$E_t = \frac{1}{4N_L^2 M_L^2} \sum_{i,j} (G^L_{ij} - A^L_{ij})^2$	(8)
---	-----

а укупан губитак стила је:

$L_{style}(\vec{a}, \vec{x}) = \sum_i^L \omega_i E_i$	(9)
---	-----

где је w_i фактор тежине доприноса сваког слоја у укупном губитку. Извод E_L по активацијама у слоју L се аналитички рачуна:

$\frac{\partial E_L}{\partial F_{ij}^L} = \begin{cases} \frac{1}{N_L^2 M_L^2} ((F^L)^T (G^L - A^L))_{ji} & \text{ако је } F_{ij}^L > 0 \\ 0 & \text{ако је } F_{ij}^L < 0 \end{cases}$	(10)
--	------

Градијенти од E_L по активацијама у нижим слојевима мреже може се рачунати коришћењем стандардне бекпропагације грешке.



Коначна слика почиње свој живот као насумична колекција пиксела која се потом на овај начин итеративно унапређује тако да се добије оптимална комбинација стил слике и контекст слике, односно тако да укупна разлика између карактеристика контекста и карактеристика стила у односу на улазне слике буде минимална.

Истраживање и резултати

За одређивање успешности овог алгоритма спроведено је три типа анкете:

1. За прву анкету је насумично изабрано 10 фотографија портрета познатих личности. Затим је свака од тих фотографија пропуштена кроз следеће алгоритме: Неурални трансфер стила [Gatys et al.2015], Дубока аналогија (Deep Analogy) [Liao et al.2017], једно-модална мрежа за превођење слике CycleGAN [Zhu et al.2017b], мулти-модална мрежа за превођење слика MNUNIT [Huang et al.2018], CariGan [Kaidi Cao] и наш алгоритам. У избор је убачен и портрет одговарајуће карикатуре насликане од стране човека. Анкету је урадила 21 особа. Од учесника је било захтевано да од понуђених карикатура изабере оне које су, по њиховом мишљењу, задовољавају факторе целокупне физичке привлачности, хумора и стила. Резултати анкете су следећи, и представљају проценат карикатура одређеног стила које су одабране као веродостојне:

- Neural Style: 12,3%
- Deep analogy: 17,1%
- CycleGAN: 25,1%
- Ручно нацртане карикатуре: 53,8%
- MNUNIT: 11,5%
- CariGAN: 36,1%
- Наш алгоритам: 39,5%

Највећи проценат изабраних карикатура, као што је и очекивано, имале су људски израђене карикатуре - 53,8%. Најуспешнији алгоритам је био наш, са 39,5% прихваћених карикатура. Следећи су CariGAN и CycleGAN са 36,1% и 25,1%, респективно. Остали алгоритми постигли су знатно лошије резултате - најлошији, MNUNIT остварио је 11,5%.

2. Друга анкета је рангирајућег типа. Садржи 30 насумично изабраних фотографија портрета познатих личности, које су пропуштене кроз следеће алгоритме: Неурални трансфер стила [Gatys et al.2015], Дубока аналогија (Deep Analogy) [Liao et al.2017], једно-модалне мреже за превођење слике CycleGAN [Zhu et al.2017b] и UNIT [Liu et al.2017], мулти-модална мрежа за превођење слика MNUNIT[Huang et al.2018], CariGan [Kaidi Cao] и наш алгоритам. Овој анкети је приступило 11 учесника, од којих се захтевало да, за сваку познату личност, рангирају од најбоље до најлошије понуђене карикатуре различитих генератора. Резултати показују следеће:

- Neural Style: 4,57
- Deep Analogy: 4,24
- CycleGAN: 3,93
- UNIT: 4,85
- MNUNIT: 4,60
- CariGAN: 2,84
- Наш алгоритам: 2,96

С обзиром да је анкета тражила рангирање карикатура од 1 до 7 (где 1 представља најбољу карикатуру, а 7 најлошију), CariGAN алгоритам се показао најуспешнијим са просечном оценом 2.84. Одмах после, са 2.96 је наш алгоритам што је такође очекивано на основу резултата прве анкете. Затим следе алгоритми

CycleGAN, Deep Analogy, Neural Style са средњим вредностима од 3.93, 4.24, 4.57, респективно. На последњем месту су MNUNIT и UNIT алгоритми са 4,60 и 4,85.

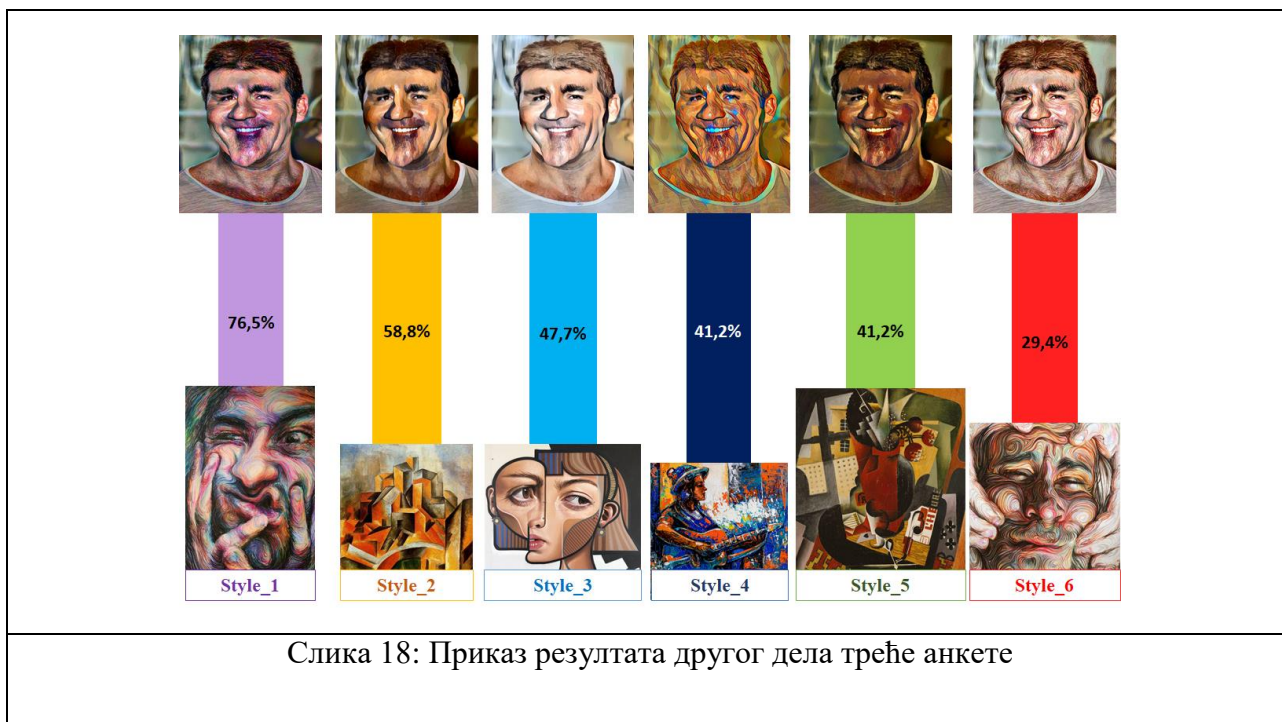


Слика 17: Поређење Neural Style, Deep Analogy, CycleGAN, UNIT, MNUNIT и CariGAN алгоритама и ручно насликаних карикатура са нашим алгоритмом

3. Трећа анкета садржи само карикатуре генерисане нашим алгоритмом. Захтеви анкете су исти као у претходним деловима. На овај начин добијена је свеукупна визуелна оцена квалитета наших карикатура, без поређења са другим референтним узорцима. Састоји се из два дела:

У првом делу, насумично изабраних од 122 портрета познатих личности пропуштено је кроз наш генератор. Анкету је попунило 18 особа, од којих се захтевало да изаберу неодређен број задовољавајућих карикатура. За сваку појединачну слику рачунат је проценат људи коју је њу одабрало као „успешну”. Број карикатура које су оствариле више од 50% гласова је 12, које су оствариле више од 40% је такође 12, оне које су оствариле више од 30% је 15, више од 20% је 23, више од 10% је 27, и оних мање од 10% има 5. Најуспешнија карикатура има освојених 58,8% гласова. Просечан број гласова по карикатури је 23,57%.

У другом делу анкете је представљено 20 врста стилова који су примењени на истој карикатури. Учесници су питани да, по свом нахођењу, изаберу 5 најприкладнијих стилова. Најбољим се показао стил преузет са слике „Lines” (Style_1) са 76,5%.



Закључак

Овај приступ аутоматском генерисању карикатура може се сматрати релативно успешним и конкурентним другим аутоматским системима генерисања карикатура. Ипак, аутоматски системи генерисања карикатура и други аутоматски системи генерисања уметничких дела још увек нису ни близу људским стваралачким могућностима. Уметничко дело мора настати из намере и мора да преноси емоцију, а за то машине, бар још увек, нису способне. С друге стране, уметност одржава своју виталност сталним иновацијама, а технологија је један од главних покретача тренутне иновације. Данас се можемо сустрести са бројним интригантним експериментима над техникама вештачке интелигенције, које ће, као уметнички алати, засигурно променити начин на који размишљамо о уметности.

Захваљујем се свом ментору и четворогодишњем професору Ивани Јовановић Мاستиловић на пруженој помоћи и подршци.

Литература

1. Ergun Akleman. 1997. Making caricatures with morphing. In Proc. ACM SIGGRAPH. ACM, 145.
2. Ergun Akleman, James Palmer, and Ryan Logan. 2000. Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes. In Proc. Visual. 165–170.
3. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In Proc. ECCV. Springer, 694–711.
4. Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. 2017. Visual attribute transfer through deep image analogy. arXiv preprint arXiv:1705.01088 (2017).
5. Tamás Szirányi and Josiane Zerubia. 1997. Markov random field image segmentation using cellular neural network. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 44, 1 (1997), 86–89.
6. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196
7. Brennan, S. E. 1985. Caricature generator: the dynamic exaggeration of faces by computer. Leonardo 18, 3, 170-178.
8. Lee, D-D. and Seung, H-S. 1999. Learning the parts of objects by nonnegative matrix factorization. Nature 401, 788-791.
9. KAIDI CAO, JING LIAO, LU YUAN, 2017. CariGANs: Unpaired Photo-to-Caricature Translation
10. <https://towardsdatascience.com/deep-generative-models-25ab2821afd3>
11. <https://www.youtube.com/watch?v=YqNMF-7TRbg>
12. Zhenyao Mo, John P Lewis, and Ulrich Neumann. 2004. Improved automatic caricature by feature normalization and exaggeration. In ACM SIGGRAPH Sketches. ACM, 57.
13. (n.d.). Retrieved from <https://ieeexplore.ieee.org/abstract/document/5473224>
14. (n.d.). Retrieved from https://ai.stanford.edu/~kaidicao/cari-gan/result_compare.html
15. A., L., Ecker, Bethge, & Matthias. (2015, September 2). A Neural Algorithm of Artistic Style. Retrieved from <https://arxiv.org/abs/1508.06576>
16. Backpropagation. (2019, September 10). Retrieved from <https://en.wikipedia.org/wiki/Backpropagation>
17. Hua Huang, X.-W. M. (2010, November 3). Frontal and Semi-Frontal Facial Caricature Synthesis Using Non-Negative Matrix Factorization. Retrieved from <https://link.springer.com/article/10.1007/s11390-010-9405-2>
18. Johnson, Justin, & Li. (2016, March 27). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Retrieved from <https://arxiv.org/abs/1603.08155>

19. Narayanan, H. (n.d.). Convolutional neural networks for artistic style transfer. Retrieved from <https://harishnarayanan.org/writing/artistic-style-transfer/>
20. Simonyan, Karen, Zisserman, & Andrew. (2015, April 10). Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved from <https://arxiv.org/abs/1409.1556>
21. Understanding Neural Networks Through Deep Visualization. (n.d.). Retrieved from <http://yosinski.com/deepvis>