



**УНИВЕРЗИТЕТ “СВ. КИРИЛ И МЕТОДИЈ” -
СКОПЈЕ**



**ФАКУЛТЕТ ЗА ЕЛЕКТРОТЕХНИКА И
ИНФОРМАЦИСКИ ТЕХНОЛОГИИ**

- СЕМИНАРСКА РАБОТА -

по предметот

ИНТЕЛИГЕНТНИ АГЕНТИ

Тема

**Препознавање на емоции на лице
(Face Emotion Recognition)**

Ментор:
доц. д-р Христијан Ѓорески

Изработил:
Дарко Ангеловски, 171/2017
darkoangelovski666@gmail.com

Скопје, јули 2021

Table of Contents

| | |
|--|-----------|
| АПСТРАКТ | 2 |
| ВОВЕД | 3 |
| 1 ПОВРЗАНА РАБОТА (RELATED WORKS) | 3 |
| 1.1 ПРИСТАП БАЗИРАН НА МАШИНСКО УЧЕЊЕ | 4 |
| 1.2 ПРИСТАП БАЗИРАН НА ДЛАБОКО УЧЕЊЕ | 4 |
| 2 ПРЕГЛЕД НА CNN, ДЛАБОКИ CNN И TRANSFER LEARNING МОДЕЛИ | 4 |
| 2.1 КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ | 4 |
| 2.2 ДЛАБОКИ КОНВОЛУЦИСКИ НЕВРОНСКИ МРЕЖИ | 5 |
| 3 TRANSFER LEARNING | 5 |
| 4 ПОДАТОЦИ | 6 |
| 5 МЕТОДИ | 6 |
| 6 ДОБИЕНИ РЕЗУЛТАТИ | 7 |
| 7 РЕФЕРЕНЦИ | 11 |

Листа на слики

| | |
|---|----|
| Слика 1 КОНВОЛУЦИСКИ МРЕЖИ..... | 5 |
| Слика 2 ОСНОВЕН МОДЕЛ ЗА TL VO СКЛОП НА CNN | 6 |
| Слика 3 НЕРВОЗЕН ИЗРАЗ | 7 |
| Слика 4 СРЕЌЕН ИЗРАЗ | 8 |
| Слика 5 НЕУТРАЛЕН ИЗРАЗ | 8 |
| Слика 6 ТАЖЕН ИЗРАЗ..... | 9 |
| Слика 7 ИЗНЕНАДЕН ИЗРАЗ | 9 |
| Слика 8 ДЕТЕКЦИЈА НА ЛИЦЕТО | 9 |
| Слика 9 СЛИКА САМО ОД ЛИЦЕТО..... | 10 |
| Слика 10 ПРЕТПОСТАВЕН РЕЗУЛТАТ | 10 |

Апстракт

Овој проект е изработен со користење на дата множество со слики на кои се претставени емоционални изрази во седум категории, а кои покажуваат: нервоза, одвратност, страв, среќа, неутрален израз, тага и изненадување. Проектот користи конволуциски невронски мрежи за тренирање и предвидување на резултатот. Искористен е веќе трениран модел кој само е надополнет и прилагоден за потребите на овој проект.

Клучни зборови: Невронски мрежи, длабоки конволуциски невронски мрежи, модел.

Вовед

Емоциите на лицето на човекот, како дел од невербалната комуникација, кажуваат многу информации за самата личност и ни откриваат како во даден момент се чувствува лицето и за што всушност размислува. Во денешно време многу поголем приоритет и се дава на вербалната комуникација, но се заборава дека невербалната комуникација, а особено изразите на лицето, откриваат многу тајни. Причината за изработка на ваков проект, не беше да креираме модел со кој ќе може секогаш да ги разликуваме човековите емоции според изразот на лицето и да се ослободиме себеси од потребите да посветиме поголемо внимание кога ќе се најдеме во комуникација со некоја личност. Причината беше повеќе од љубопитност, но и да се убедиме себеси дека иако компјутерите не знаат да читаат емоции на изразот на лицето, сепак, во некои моменти тоа го прават подобро и од луѓето.

Големо значење, за популарноста на оваа тема, имаат истражувачите Фризен и Екман [1] кои извршиле опсежно истражување во човечките изрази на лицето и идентификувале универзални изрази на лицето како што се: среќа, тага, лутина, страв, изненадување, одвратност и неутрални состојби. Препознавањето на емоции од изразите на лицето, стана и привлечно истражувачка тема во психологијата, психијатријата и менталното здравје. Автоматското откривање на емоциите од изразите на лицето се исто така неопходни за паметно живеење, здравствени системи[2], како и дијагностицирање на нарушувања на емоции кај нарушувања на аутистичниот спектар[3]. Затоа, препознавањето на емоциите на лицето (FER) го привлече вниманието на повеќе истражувачки заедници.

Во моментов, длабоките невронски мрежи, а особено конволуциските невронски мрежи, привлекоа големо внимание кога станува збор за FER (Face Emotion Recognition). Ова се должи на опцијата што ја нудат конволуциските невронски мрежи, а тоа е екстракција на некои одлики од сликите[4]. Ова игра многу голема улога затоа што:

- a) Препознавањето на емоции бара слики со висока резолуција што значи работа со податоци со висока димензија;
- b) Разликата во лицата, заради различните емоционални состојби, е многу мала што секако ја отежнува задачата за класификација.

Од друга страна, длабоки конволуциски невронски мрежи имаат огромен број на скриени слоеви (hidden layers) и тренирање на толку голем број на слоеви станува незгодно. Освен тоа ништо не ни гарантира дека со зголемувањето на бројот на слоеви, нашиот модел ќе биде попрецизен (vanishing gradient problem). Затоа, се користат веќе тренирани модели. Некои од нив се: VGG-16, Resnet-50, Resnet-152, Inception-v3 и DenseNet-161.

1 Поврзана работа (Related works)

За работа со FER (Face Emotion Recognition) испитани се неколку методи. Првичните методи првин ги извлекувале одликите (да се гледа само лицето, без позадината) и дури потоа ги класифицирале емоциите од вредноста на карактеристиките. Од друга страна, методите засновани на двабоко учење (deep learning-based methods) ја извршуваат задачата на тој начин што ги комбинираат двата чекора (извлекување на одликите и класифицирање). Во повеќе студии ги

разгледувале и проучувале двата споменати методи. Во оваа подточка накратко се опишани двата метода кои денес се користат при FER.

1.1 Пристап базиран на Машинско учење

Различни методи за машинско учење (на пример K-nearest neighbor, neural network) се користат при совладување на задачата за детекција на емоции на лицето. Пионерскиот метод за детекција на емоции на лицето од Ксиао-Ксу и Веи (Xiao-Xu and Wei) [5] додаде карактеристика на енергија на брановидност (wavelet energy feature - WEF) прво на сликата на лицето, а потоа ги искористи линеарните дискриминанти на Фишер (Fischer's Linear Discriminants - FLD) за да ги извлече потребните карактеристики од сликата и на крај ги класифицираше емоциите со употреба на методот K-најблизок сосед (K-nearest neighbor - KNN). Zhi и Ruan работеле на метод на вадење на вектори со карактеристиките од 2D дискриминирачки локалитет за зачувување на проекциите. Chang и Huang исто така се препуштија во светот на препознавање на емоциите и тие вклучија функција за радијална основа (Radial Basis Function - RBF) за невронската мрежа.

1.2 Пристап базиран на Длабоко учење

Овој пристап со длабоко учење е релативно нов и до сега има неколку студии кои работат врз основа на конволуциските невронски мрежи CNN. Zhao и Zhang интегрираа deep belief network DBN користејќи NN за FER (Face Emotion Recognition) каде што DBN се користи без надзор на самите карактеристики, а самата NN се користи за класификација на карактеристиките. Pons и Masip формираа ансамбл/комбинација од 72 конволуциски невронски мрежи, каде што одделни CNN беа различно прилагодени.

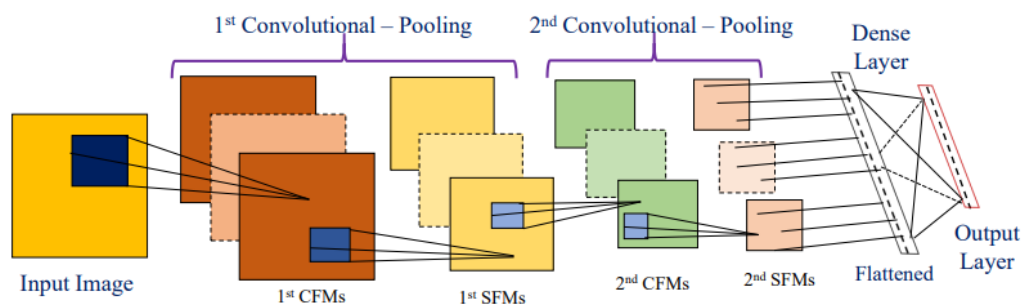
2 Преглед на CNN, длабоки CNN и Transfer Learning модели

Длабоките невронски мрежи се главната тематика за работа со конволуциските невронски мрежи, како и веќе тренираните модели. Самиот процес за тренирање на длабоки невронски мрежи е обемен и бара многу време. Затоа е згодно да се искористат веќе тренирани модели со кои може да се послужиме.

2.1 Конволуциски невронски мрежи

Поради својствената структура на CNN, тој е најсоодветниот модел за доменот на сликата. CNN се состои од влезен слој, повеќе скриени слоеви со конволуциски збир, и излезен слој. Во основа, конволуцијата е математичка операција од две функции што произведуваат трета функција што изразува изменета форма на функцијата. Јадрото на CNN се лизга низ сликата за да најде корисни обрасци во неа преку операција на конволуција. Соединувањето е форма на нелинеарно намалување на примерокот. Слој на здружување комбинира области што не се преклопуваат на еден слој во единствена вредност во следниот слој. Слика 1 ја покажува генеричката архитектура на стандарден CNN со два слоја со конволуциско здружување. Првиот слој на конволуција ја применува операцијата на конволуција на влезната слика и генерира 1-ви склопени мапи со одлики (CMF) тие се влез за последователна операција на соединување.

Со 1-та операција на здружување се произведуваат 1-ви мапи со обележани карактеристики (SFM). Потоа, се вршат 1-ви здружувања, операции на 2-ри конволуцијален-здружен слој. Израмнување на Вторите вредности на SFM, целосно поврзаниот слој (т.е. густот слој - Dense) го извршува последното расудување каде што невроните се поврзани со сите активации во претходниот слој. Последниот слој, исто така наречен слој на загуба, специфицира како тренингот го казнува отстапувањето на реалното производство од предвидениот излез. Таквата архитектура на CNN е популарна за препознавање на обрасци од влезни слики од мала големина (на пример, 48×48 , како што се и нашите првични слики) како рачно напишано препознавање на броеви. Деталниот опис на CNN е достапен во постојните студии.



Слика 1 Конволуциски мрежи

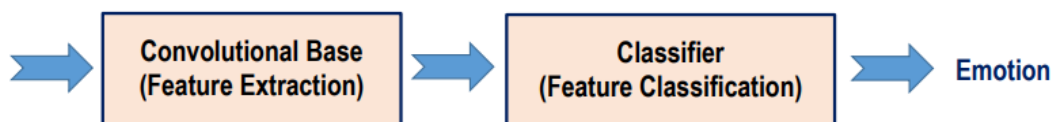
2.2 Длабоки конволуциски невронски мрежи

DCNN имаат многу голем број на скриени слоеви (hidden layers). Различни DCNN имаат/се користат за различни намени. Првиот модел што доби добра точност на ImageNet беше моделот AlexNet кој имаше само 5 слоеви. ZFnet се базира на слична идеја, но со помалку параметри. Додека AlexNet искористи 15 милиони слики за обука, ZFNet користеше само 1,3 милиони слики за да добие сличен резултат. Подоцна, VGG-16 предложи подлабок модел со длабочина со 13 конволуциони слоеви и помал јадра. VGG-19 е уште еден модел во оваа категорија со 16 слоеви на конволуција. Важен концепт што го користат повеќето од подоцнежните модели е прескокнување-врска, која беше воведена во преостанатата невронска мрежа (ResNet) [6]. Основната идеја за прескокнување на врската е да се насочи влезот на слојот и да се додаде на излезот по некои слоеви.

3 Transfer Learning

FER користењето на претходно обучен модел DCNN преку соодветен TL е главниот придонес на оваа студија. Махендран и Ведалди [69] визуелизираа што учат слоевите на CNN. Првиот слојот на CNN ги доловува основните карактеристики како работ и аглиите на сликата. Следниот слојот открива посложени карактеристики како текстури или форми, а горниот слој го следи истиот механизам кон учење на посложени обрасци. Како што се основните карактеристики слични на сите слики, задачите за FER во долните слоеви во

DCNN се идентични со другите операции базирани на слика, како што е класификација. Од обука на DCNN модел од нула (т.е. со случајно иницијализирани тежини) е огромна задача, веќе обучен модел DCNN на друга задача може да се усогласи користејќи TL (Transfer Learning) пристап за препознавање емоции. Модел DCNN (на пример, VGG-16) претходно обучен со голема база на податоци (на пример, ImageNet) за слика класификацијата е погодна за FER.



Слика 2 Основен модел за TL во склоп на CNN

На слика 2 е прикажана општата архитектура на DCNN-моделот базиран на TL за FER, каде конволуциската основа е дел од претходно истренирана DCNN со исклучок на сопствениот класификатор, а класификаторот на основата е ново додаден слој за FER. Како целина, пренаменувањето на претходно обучениот DCNN се состои од два чекори: замена на оригиналниот класификатор со нов и дотерување на моделот. Додадениот дел од класификаторот е генерално комбиниран Dense слој, а останатиот дел се слоевите што си ги додава самиот програмер по потреба. Од практична гледна точка, и изборот на претходно обучен модел и одредувањето матрица за сличност на големината за фино подесување се важни во TL. Постојат три широко користени стратегии за обука на моделот за дотерување: обучете го целиот модел, обучете некои слоеви оставајќи ги другите замрзнати и обучете го само класификаторот (т.е. замрзнете ја основата на конволуцијата). Во случај на слична задача, обука на единствениот класификатор и / или неколку слоеви е доволно за фино подесување за учење на задачата. Од друга страна, за различни слични задачи, неопходна е обука за целосен модел.

4 Податоци

За овој проект, искористена е датабаза со слики која што е преземена од официјалната страна на kaggle. Самата датабаза има (моментално) 28. 076 слики кои се групирани во 7 категории. Категориите за сликите се: Angry, Sad, Surprised, Fear, Happy, Neutral и Disgust.

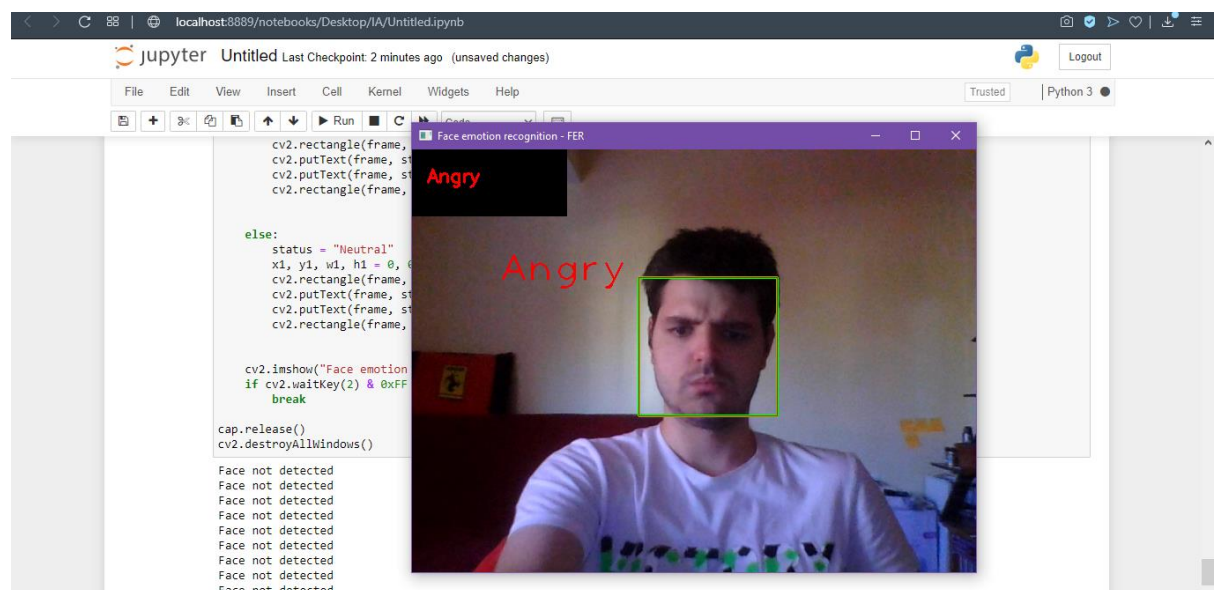
5 Методи

Во проектот вклучен е веќе трениран модел наречен MobileNetV2 кој има влез во форма [(None, 224, 224, 3)]. Тој влез и ќе го користиме, а за излез, наместо веќе дефинираниот излез со 1000 неврони, ќе го смениме за да ги задоволува нашите потреби, односно излезот да има 7 неврони. Само го отргнуваме тој слој и на него додаваме скриени слоеви со 128 неврони, активациска функција, скриени слоеви со 64 неврони, активациска функција и на крај, dense слој со 7 неврони и

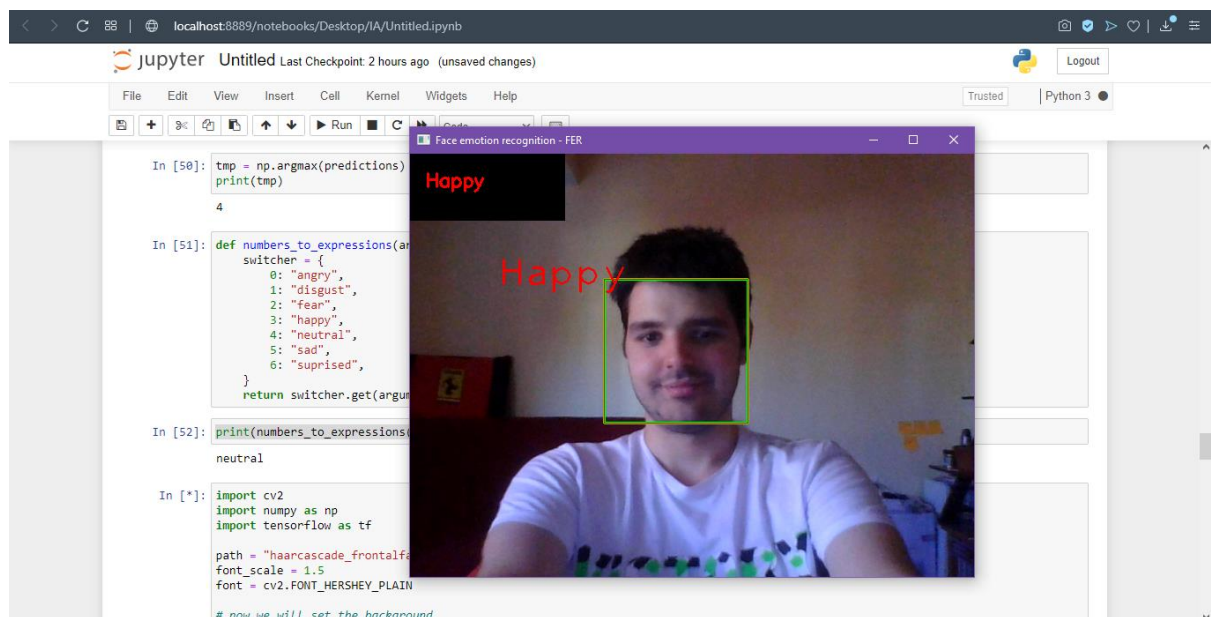
активациска функција. Нашиот модел е спремен за тренирање. Освен веќе трениран модел, проектот вклучува и готов алгоритам за детекција на лице[100]. На крајот, како додаток се користи и камерата со која се земаат слики во живо и се користат со веќе тренираниот модел за да добиеме резултати.

6 Добиени резултати

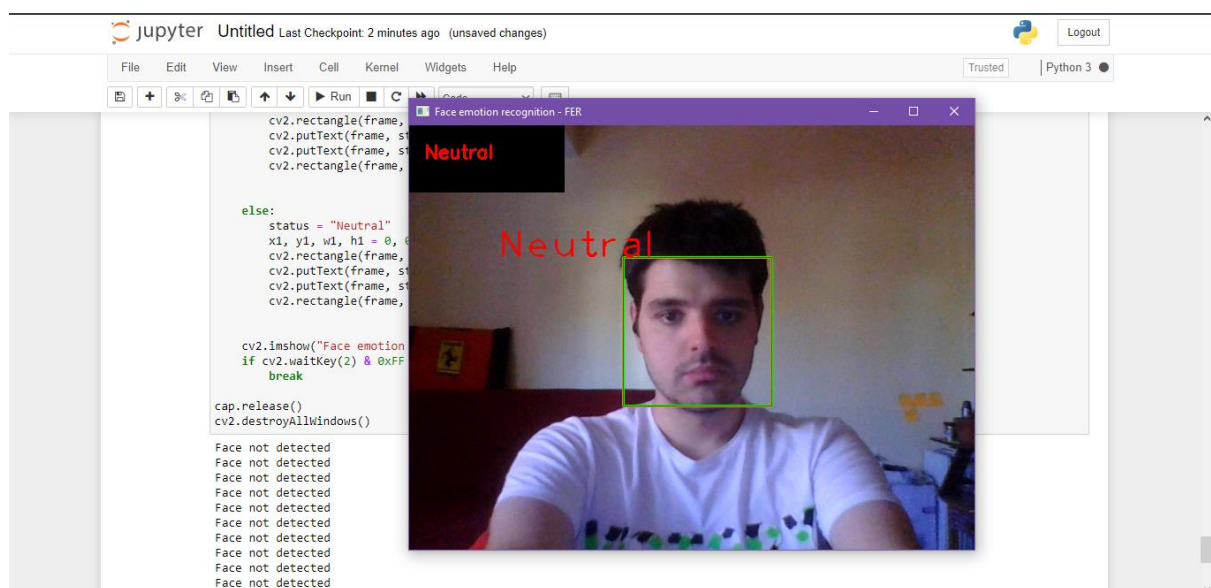
Резултатите се добиени со едноставни тестирања извршени од моја страна. Во оваа подточка тоа може и сликите да се види.



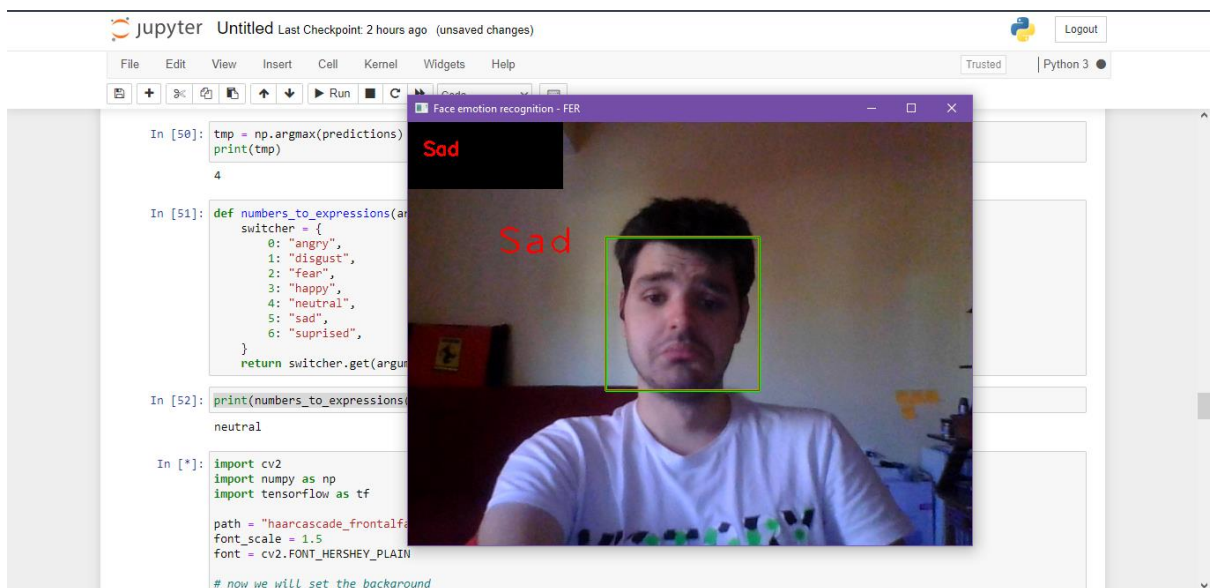
Слика 3 Нервозен израз



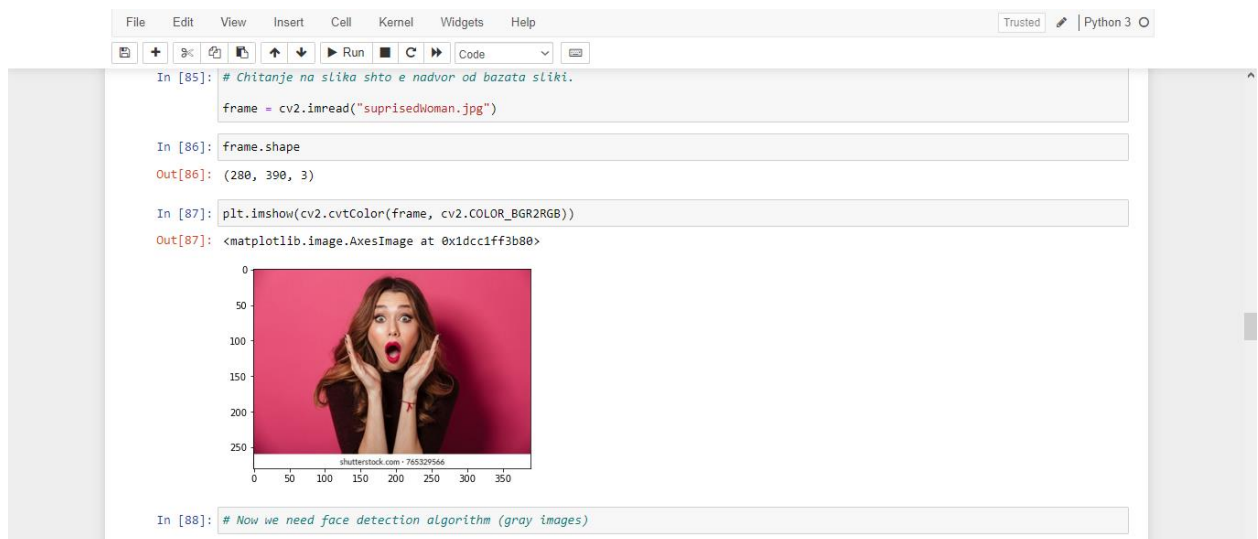
Слика 4 Среќен израз



Слика 5 Неутрален израз



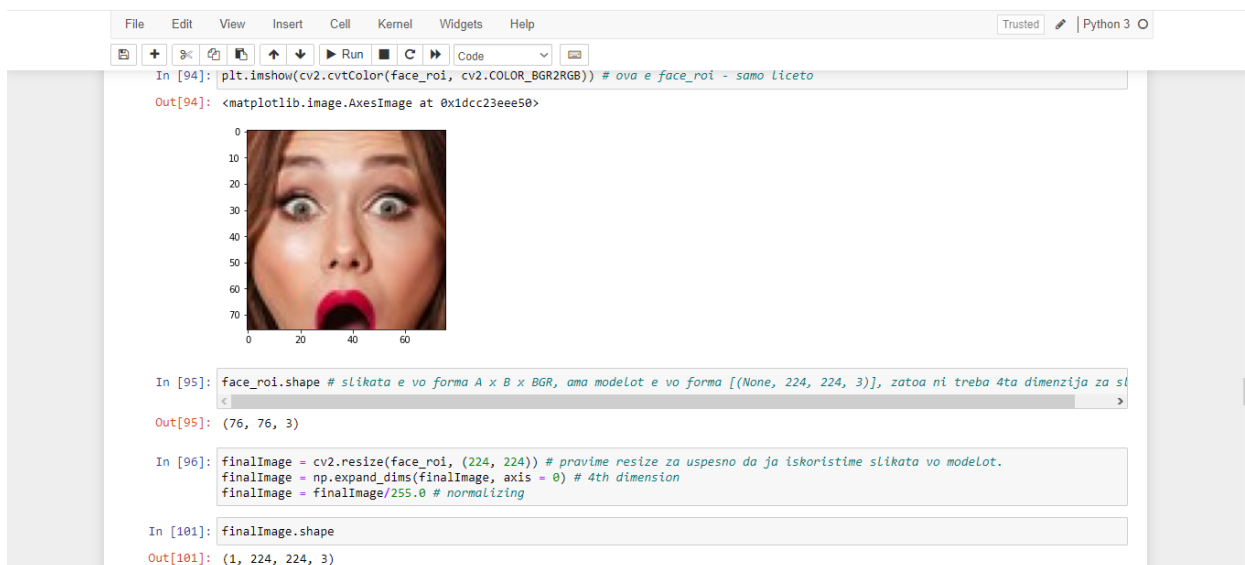
Слика 6 Тажен израз



Слика 7 Изненаден израз



Слика 8 Детекција на лицето



Слика 9 Слика само од лицето



Слика 10 Претпоставен резултат

7 Референци

- [1] <https://link.springer.com/article/10.1007/BF01115465>
- [2] <https://pubmed.ncbi.nlm.nih.gov/27678301/>
- [3] https://dominionyouthservices.com/services/intellectual-and-developmental-disabilities-services/?gclid=CjwKCAjwoZWHBhBgEiwAiMN66eH6eTpMkdHRucg3cPX0jhG-CVsvOOIIVcPUA8dzWdZSBdZ8stcs1RoCmUMQAvD_BwE
- [4] <https://www.mdpi.com/2079-9292/8/3/292>
- [5] <https://ieeexplore.ieee.org/document/4244806>
- [6] <https://ieeexplore.ieee.org/document/7780459>