

Природно-математички факултет

Април, 2020.

Бања Лука

ПРЕПОЗНАВАЊЕ СЛИКА

(семинарски рад из Информационих технологија и друштва)

Професор:

др. Драган Матић

Студент:

Ивана Креховски

Садржај:

Садржај:	2
1. УВОД.....	3
2. ВРЕМЕНСКА РАЗДОБЉА.....	12
2.1 Почетак.....	12
2.2 Предвиђања	12
3. ПРЕПОЗНАВАЊЕ СЛИКА И РАЧУНАРСКИ ВИД	4
3.1 Шта човјек види?	4
3.2 Шта рачунар види?.....	5
3.3 Корелација човјека и рачунара.....	6
4. ПРИКАЗ ПРОЦЕСА ПРЕПОЗНАВАЊА СЛИКА	6
4.1 Мисија	6
4.2 Рјешење	8
4.3 Платформе.....	10
5. ПРИМЈЕНА	10
5.1 Негативне особине	10
6. ЛИТЕРАТУРА.....	14

1. УВОД

Од почетка развоја техничких наука човјек тежи ка томе да помоћу њих олакша свој начин живота, самим тим се јавља све већа иновативност код научника. Препознавање слика се своди на то да научимо рачунар да сам разумије и ради као људско око.

Како бисмо потпуно разумјели начин на који рачунар препознаје слике потребно је најприје да се упознамо са појмом рачунарског вида и појмом вјештачке интелигенције, затим да сазнамо који су то још задаци поред примарног задатка препознавања слика и сам процес рада који води ка рјешењу примарног циља. Такође, споменућемо све ширу примјену овог научног подручја, упознати његове предности и мане као и неке популарне софтвере и њихове разлике.

2. ПРЕПОЗНАВАЊЕ СЛИКА И РАЧУНАРСКИ ВИД

Термини препознавање слика (енг. Image recognition) и рачунарски вид (енг. Computer vision) су се од почетка користили наизмјенично те се с тога временом изгубила њихова минимална али опет постојећа разлика. Рачунарски вид омогућава рачунару да имитира људски вид и предузима кораке, док препознавање слика се односи на пиксел и анализу узорка слике како би се препознала слика као одређени објект. Из приложеног можемо да закључимо да су ова два термина блиска и да су садржани један у другом. Препознавање слика као рачунарски вид су области вјештачке интелигенције (енг. Artificial intelligence) за коју можемо рећи да има два циља. Први циљ је биолошки тј. да рачунар види као и људско око. Други циљ је инжењерски који настоји да изгради посебне системе који би чак могли да надмаше људски офтамолошки систем. Дакле, ова два циља су уско повезана и инспирација се проналази у човјеку, у наставку рада ћемо се боље упознати са циљевима рачунарског вида и на који начин рачунар препознаје слике.

2.1 Шта човјек види?

Да бисмо разумјели како рачунар види потребно је да се упознамо и са основним појмовима физиологије вида и покажемо да директну паралелу између људског и рачунарског вида. Када човјек гледа у објекат, свјетлост која допире до њега долази и до очију. Свјетлост пролази кроз рожњачу (мјесто гдје свјетлост улази у око), а затим зјеницу која постаје већа при слабијем освјетљењу те мања при јачем. Сочиво фокусира свјетлосне зраке на ретину (унутрашњи ћелијски слој ока) која садржи милионе ситних фоторецептора које детектују свјетлост. Постоје двије врсте фоторецептора а то су штапићи и чепићи. Штапићи су монохроматски рецептори који су високо осјетљиви на свјетлост, што омогућава да се предмети виде и при свјетлости ниског интензитета, тј. обавјештавају нервни систем о присуству или одсуству фотона¹ без обзира на таласну дужину. Чепићи боље функционишу при јачем свјетлу и различито су осјетљиви на свјетлост различитих таласних дужина фотона (плава-кратка таласна дужина, зелена-средња таласна дужина и црвена- велика таласна дужина). У највећем броју случајева обојено свјетло надражује све три врсте чепића, али у различитом степену па тако образац стимулације кодира боју. Штапићи и чепићи покривају читаво подручје ретине осим тачке

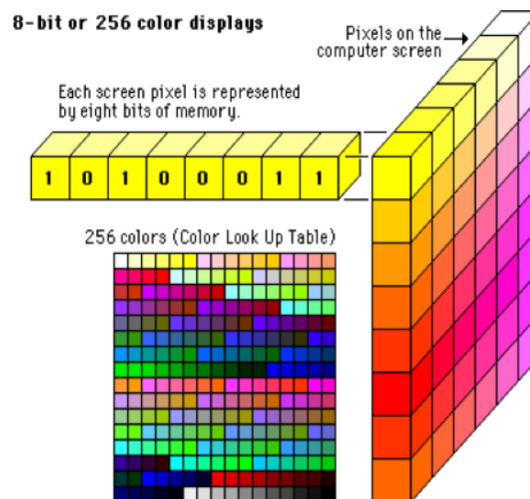
¹ Фотон-основна јединица видљиве свјетлости

гдје се оптички живац спаја с мозгом. Оптички живац носи информације добијене из ретине према мозгу, гдје их мозак претвара у оно што ми видимо.

2.2 Шта рачунар види?

Тежину и комплексност ове области можемо објаснити самом реченицом да слика говори хиљаду ријечи, у овом случају ријечи представљају пиксели (енг. pixel). Пиксел или пал је најмањи графички елемент слике. Сваки пиксел представљен је бројем или скупом бројева, а распон тих бројева назива се дубином боје (или битном дубином) тј. дубина боје означава максимални број потенцијалних боја које се могу користити на слици. У (8-битној) сивој скали слике (црно-бијеле) сваки пиксел има једну вриједност која се креће од 0 до 255. Већина слика данас користи 24-битну боју или више. РГБ (енг. RGB) слика у боји значи да је боја у пикселу комбинација црвене, зелене и плаве боје.

Свака од боја се креће од 0 до 255. РГБ генератор боја показује како РГБ може генерисати било коју боју. Такође сваки пиксел има свој положај представљен у битмапи² координатама по ширини “х” и висини “у и интензитет освјетљености. Међутим, ако се слика дешава у реалном времену ствари се мијењају, постоје камера сензори који могу бити монохроматски и РГБ. Унутар сваког сензора се налази мноштво пиксела који пикупају свјетлост. Сензори само садрже количину свјетлости коју је потребно превести у пикселе да би се забиљежила боја. Сензори могу да биљеже само по једну боју (црвена, зелена, плава) тако да пиксели који биљеже исту боју никада нису један поред другог, то се назива ЦФА(енг. CFA- color filter array). Сваки пиксел ефективно биљежи само једну трећину података, а остале се боје филтрирају тј. све плаве и зелене нијансе које падну на црвени сензор неће бити забиљежене. Такође, битно је да сваки пиксел једну боју биљежи директно, а друге двије боје у њему се морају извести из околних сензора. Сада када познајемо појам пиксела можемо схватити да рачунар не види слику и боје као ми већ види бројеве што препознавање слика чини тежим. Рачунар треба да разумије шта сваки пиксел представља у већим сликама.



Слика 1. Пиксели

² Битмапа-„растер“ је свака графичка слика састављена од скупа пиксела

2.3 Корелација човјека и рачунара

Упознали смо се са основним појмовима и сада можемо да докажемо да је инспирација понађена у човјеку. У људском оку се слика кодира у ретини, а њени кодови су представљени мноштвом фотона као што сваки пиксел сачињава слику у рачунару и има свој код тако и фотон. Боје које пиксел препознаје црвену, плаву и зелену тако и људско око тј. чепићи. Сваки пиксел има интензитет, дубину и боју, паралелно с тим и сваки фотон има своју таласну дужину и боју. Информације кодиране у ретини се шаљу на даљу обраду у мозак, а информације пикселима се шаљу на обраду у рачунар.

3. ПРИКАЗ ПРОЦЕСА ПРЕПОЗНАВАЊА СЛИКА

3.1 Мисија

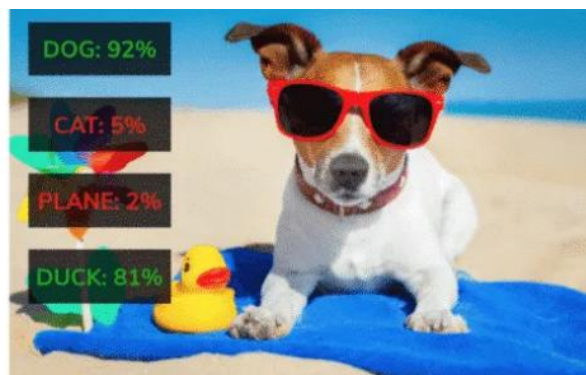
Релативно је једноставно индексирати и претраживати текст, али да би индексирали и претраживали слике, алгоритми морају знати шта слике садрже. Схватили смо како рачунару превести слике њему разумљивим језиком, али тај низ бројева и даље не препознаје и разумије слику у цјелини. Да бисмо максимално искористили податке од слика, потребни су нам рачунари да „виде“ слику и разумију њен садржај. Морамо поставити низ задатака како бисмо у потпуности добили тачну информацију о слици као цјелини.

Класификација објеката: Која је широка категорија предмета на овој слици? - То је идентификација класе, тј. категорије којој слика припада. Слика може имати само једну класу.

Идентификација објекта (означавање): Који је тип датог објекта на овој слици? - Ово је такође задатак класификације, али са већим степеном тачности. Може препознати присуство више појмова или предмета унутар слике. Једна или више ознака се могу доделити одређеној слици.



Слика 2. Класификација објекта



Слика 3. Идентификација објекта

Детекција објекта: Да ли је објекат на слици? Гдје су предмети на слици? Које су кључне тачке објекта на слици? - Ово је неопходно када желимо да пронађемо локацију на слици. Када се објекат пронађе, око објекта се поставља ограничавајући оквир.

Сегментација објекта: Који пиксели припадају објекту на слици? - Ово је такође задатак детекције објекта. Сегментација може пронаћи елемент на слици до најближег пиксела.



Слика 4. Детекција објекта



Слика 5. Сегментација објекта

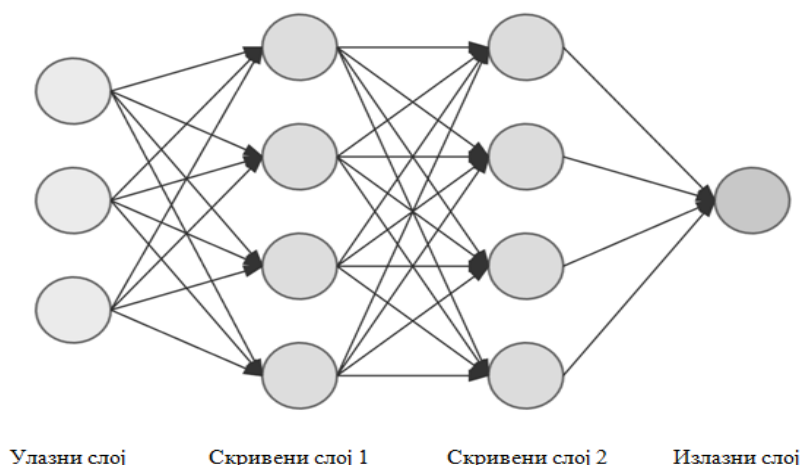
3.2 Рјешење

„Ако желимо научити машине да мисле потребно је прво да их научимо да виде“

– Феи-Феи Ли

Мало дијете може лако класификовати и идентификовати објекте али само ако је научен, то значи да ће дијете препознати пса ако га је раније видео и запамтио. На исти начин је потребно научити рачунар да препозна објекат, по угледу на људски мозак у којем се стварају енграми³. Рачунар препознаје слике уз помоћ вјештачких неуронских мрежа (енг. Artificial neural networks- ANN) које се састоје од појединачних јединица које се називају неурони. Неурони су смјештени у низ група - слојева (слика 2.). Неурони у сваком слоју повезани су са неуронима сљедећег слоја. Подаци долазе из улазног слоја у излазни слој дуж ових јединица.

Сваки појединачни чвор обавља једноставан математички прорачун. Затим преноси своје податке свим чворовима на које је повезан.



Слика 6. Слојеви мреже

Конволуцијске неуронске мреже⁴ су дио вјештачке неуронске мреже који је првенствено задужен за класификацију слика. Главни задатак класификације слика је да прихвати улазну слику која је представљена низом пиксела помоћу којих рачунар тражи основне карактеристике као што су ивице и кривине, затим кроз групе конволуцијалних слојева конструише апстрактније концепте. Што значи слика се просљеђује кроз низ конволуционих, нелинеарних, удружених и потпуно повезаних слојева, а затим генерише излаз.

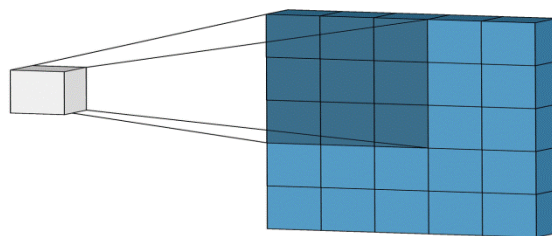
Конволуцијски слој- први слој. Ако читање улазне матрице почне у горњем лијевом углу слике затим софтвер бира мању матрицу која се назива филтер (неурон). Тада филтер врши конволуцију⁵, тј. помјера се дуж улазне слике. Задатак филтера је да умножи своје вриједности са изворним вриједностима пиксела затим се сва ова множења саберу и на

³ Енграм-траг или отисак у мозгу (стварање енграма= памћење)

⁴ Конволуцијске неуронске мреже (енг. Convolutional neural networks-CNN)

⁵ Конволуција функција- $\phi = f_1 * f_2$

крају се добије један број. Пошто је филтер прочитао слику само у горњем лијевом углу, помиче се даље десно за 1 јединицу изводећи сличну операцију. Након проласка филтера кроз све положаје, добија се матрица, али мања од улазне матрице. Када слика прође кроз један слој савијања, излаз првог слоја постаје улаз за други слој. А то се дешава са сваким даљим конволуционим слојем.



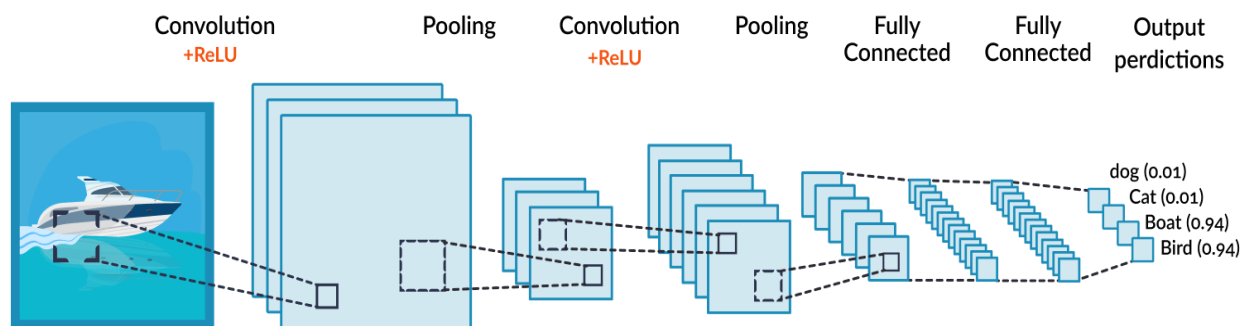
Слика 7. Конволуцијски слој

Исправљена линеарна јединица-(енг. Rectified linear unit-ReLU)- једноставна функција за увођење нелинеарности у мрежу. Све негативне вриједности једноставно се мијењају у нулу, уклањајући сву црну с слике. Формална функција је $y = \max(0, x)$.

Нелинеарни слој – додаје се након сваке операције конволуције. Без овог својства мрежа не би била довољно јака и не би била у могућности да врати узорак (као ознаку класе).

Удруживање слојева (енг. Pooling)- прати нелинеарни слој, ради са ширином и висином слике и на њима врши операције компресовања. Као резултат, смањује се капацитет слике. То значи да ако су неке карактеристике (нпр. границе) већ идентификоване у претходној операцији конволуције, детаљнија слика више није потребна за даљу обраду, а она је компримована на мање детаљне слике.

Потпуно повезани слојеви- након завршетка низа конволуционих, нелинеарних и здружујућих слојева, потребан је потпуно повезани слој. Овај слој узима излазне информације из конволуционих мрежа.



Слика 8. Приказ рада конволуцијских неуронских мрежа

3.3 Платформе

Међутим, са све већом потребом човјека за препознавањем слика почеле су се развијати платформе са огромним базама класификованих слика помоћу којих уче рачунаре да сами тренирају тј. препознају па чак и предвиђају слике. Овакве знатно скраћују вријеме за прављење различитих апликација које користе препознавање слика,

Слика 8. Приказ рада конволуцијских неуронских мрежа

рачунарски вид и уопштено вјештачку интелигенцију. Неки од познатих платформи су „ImageNet“, „Cloud Vision API“, „Amazon Rekognition“. Такође, постоје и пакети који олакшавају програмирање апликација неки од њих су „OpenCV“, „Openface“, „SimpleCV“.

4. ПРИМЈЕНА

Препознавање слика као дио рачунарског вида има и најважнију улогу у њему, јер без препознавања слика не би постојао ни рачунарски вид. То значи да се у свакој примјени рачунарског вида налази препознавање слика, сходно томе, препознавање има врло широку примјену у различитим областима и наукама као што су медицина, војска, аутомобилска индустрија и слично. Будући да је дошло од огромног развоја компјутерског вида његова примјена је постала све већа у свакодневном животу модерног човјека. Оно што је некада било научна фантастика из филмова сада је постало наша свакодневница, али исто тако поред свих позитивних страна примјене препознавања слика постоје и негативне.

Медицина- вађење података из слика у сврху дијагнозе, нпр. рендгенске слике, томографске. Такође, препознавање слика има улогу и у самом учењу медицинских радника, битно је напоменути да што се више рачунарски вид и препознавање слика буде развијало с тим ће бити олакшано лијечење пацијената. *Војска*- препознавање непријатељске стране. *Индустрија*- контрола производње и сама помоћ при производњи. *Autonomna vozila*- аутомобили без возача, авиони без пилота. Идентификација људи помоћу биометричких функција, куповина, истраживање других планета, постављање вирелног у стварни свијет.

4.1 Негативне особине

Кад је дошло до наглог развоја, поред свих позитивних страна које човјек може да искористи појавиле су се и оне негативне у којима се губи реална слика и не можемо распознати стварно од имагинарног. Такође губимо нашу приватност, наши подаци, фотографије бивају сврстане негдје на мрежи у власништву компанија или чак доступно свима, с тога је потребан и опрез при изношењу личних података. Међутим, могу настати и веће посљедице. Сlike и видео садржаји могу да се измијене помоћу различитих апликација и тешко је па скоро немогуће утврдити да ли је то стварна или лажна слика.

5. ВРЕМЕНСКА РАЗДОБЉА

5.1 Почетак

Појам рачунарског вида се појављује први пут око 1960. године у докторату Ларија Робертса (Larry Roberts) на МИТ-у⁶ у којем расправља о могућностима претварања 3Д геометријских информација у 2Д приказе. Почели су се правити алгоритми који и данас постоје за одређивање ивица, процјену кретања и сл. Затим се користе разне технике процесирања слика у циљу лакшег разумијевања и представљања рачунару.



Слика 9.

5.2 Предвиђања

У будућности ће се више проучавати како да се научи рачунар да сам учи, јер је досадашње истраживање довело само до тога да се механички уносе подаци. Самим тим у ако рачунари науче сами препознају слике знатно ће се убрзати сам процес. Такође препознавање слика има велику улогу у стварању будућих паметних градова.

⁶ МИТ- Масачусетски технолошки институт (енгл. Massachusetts Institute of Technology (MIT)) је приватни истраживачки универзитет у Кембриџу у америчкој држави Масачусетс.

6. ЗАКЉУЧАК

Препознавање слика је врло занимљива област рачунарског вида и вјештачке интелигенције која нам олакшава живот. Сматрам да је веома важна у свијету информатике, да ће постати у скоријој будућности развијенија него што је сад као и да ни сами још нисмо свјесни шта све можемо да урадимо помоћу ње. Међутим, закључујем да је исто тако да је и мач са двије оштрице јер већ има незанемариве негативне особине, тако да морамо имати и то у виду нарочито због тога што смо већ изгубили приватност тј. под лупом смо свакодневно. Између осталог, колико год да вјештачка интелигенција буде напредовала и учила никада неће моћи да замијени човјека као савршеног бића и постојање његових емоција.

7. ЛИТЕРАТУРА

- 1) Проф. др сци. мед. Јован Марић, „Клиничка психијатрија“, Београд, 2005.
- 2) Артур Ц. Гајтон, „Медицинска физиологија“, Београд-Загреб, 1968.
- 3) <https://www.youtube.com/watch?v=40riCqvRoMs>
- 4) https://en.wikipedia.org/wiki/Computer_vision
- 5) <https://www.pulsarplatform.com/blog/2018/brief-history-computer-vision-vertical-ai-image-recognition/>
- 6) http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf
- 7) https://en.wikipedia.org/wiki/Convolutional_neural_network
- 8) <https://algorithmia.com/blog/introduction-to-computer-vision>
- 9) <https://www.itexico.com/blog/computer-vision-what-are-the-advantages-and-disadvantages>
- 10) <https://towardsdatascience.com/module-6-image-recognition-for-insurance-claim-handling-part-i-a338d16c9de0>