

Automatsko prepoznavanje teksta sa tablica automobila koristeći Jedinstveni Vizuelni Model za Prepoznavanje Teksta na Sceni

vozila,
ne
nu-
zno
au-
to-
mo-
bila

Student: Andrija Urošević

Mentor: dr Nemanja Ilić

Računarski fakultet,
Univerzitet Union

Avgust 2024

Predgovor

Prostor za predgovor.

Sadržaj

Predgovor	i
Sažetak	1
1 Uvod	2
2 Prepoznavanje teksta	3
2.1 Uvod u prepoznavanje teksta	3
2.2 Istoriski pregled prepoznavanja teksta	4
2.3 Arhitekture modela za prepoznavanje teksta na slici	5
2.4 Korišćenje Jedinstvenog Vizuelnog Modela za Prepoznavanje Teksta na Sceni	6
2.4.1 Arhitektura	7
2.4.2 Progresivno preklapajuće ugrađivanje isečaka	8
2.4.3 Blok mešanja	9
2.4.4 Spajanje	10
2.4.5 Kombinovanje i Predikcija	10
2.4.6 Analiza Vizualizacije	11
3 Primene	12
3.1 Automatska naplata parkiranja	12
3.2 Automatsko naplata putarine	13
3.3 Automatsko praćenje i lociranje ukradenih vozila	13
3.4 Automatska kontrola saobraćajnih prekršaja	13
3.5 Automatska kontrola pristupa	14
3.6 Automatska kontrola saobraćaja	15
3.7 Kreiranje personalizovanih reklama	15
4 Prikupljanje i rad sa podacima	16
4.1 Uvod u prikupljanje podataka	16
4.2 Specifičnosti prikupljanja podataka u kontekstu prepoznavanja teksta sa registarskih tablica	16
4.2.1 Varijabilnost registarskih tablica	17
4.2.2 Promenljivi uslovi snimanja	17
4.3 Upoznavanje sa podacima	18
4.4 Razvrstavanje i čišćenje podataka	18
4.5 Anotacija podataka	19
4.6 Kreiranje sintetičkog skupa podataka	19
4.6.1 Prednosti sintetičkih podataka	19
4.6.2 Generisanje pozadina tablica	20

pronađi
pri-
klad-
niju-
reč
od
me-
ša-
nje

4.6.3	Generisanje teskta na tablicama	20
4.6.4	Integracija sintetičkih podataka sa realnim podacima . .	21
4.7	Augmentacija podataka	21
5	Implementacija	22
5.1	Treniranje modela prepoznavanja teksta	22
5.2	Komponente sistema	22
5.2.1	Detektor tablica	22
5.2.2	Detektor teksta	22
5.2.3	Prepoznavanje teksta	22
5.3	Tehnologije	22
5.3.1	PaddlePaddle	22
5.3.2	Docker	22
5.3.3	FastAPI	22
6	Rezultati	23
7	Buduća poboljšanja	24
8	Zaključak	25
	Literatura	26

Sažetak

Automatsko prepoznavanje teksta sa registarskih tablica automobila od izuzetne je važnosti za savremene sisteme nadzora saobraćaja, praćenje vozila i obezbeđenje sigurnosti na putevima. Identifikacija registarskih tablica ima različite primene, uključujući praćenje ukradenih vozila, naplatu putarine, sigurnosne provere i nadzor saobraćaja. U proteklim decenijama, napredak u oblasti obrade slike i mašinskog učenja omogućio je razvoj efikasnih sistema za automatsko prepoznavanje teksta sa tablica vozila. Primena dubokih neuronskih mreža i algoritama dubokog učenja omogućila je visoku tačnost prepoznavanja teksta, čak i u složenim scenama i različitim uslovima snimanja. Ovaj rad istražuje metode i tehnike za automatsko prepoznavanje teksta sa tablica automobila, sa ciljem razvoja sistema koji može precizno identifikovati registarske tablice u realnom vremenu. Eksperimentalni rezultati prikazuju performanse sistema u stvarnim uslovima i ukazuju na mogućnosti za primenu u različitim oblastima, uključujući nadzor saobraćaja, bezbednosne provere i identifikaciju vozila.

1 Uvod

Automatsko prepoznavanje teksta je ključna tehnologija u oblasti kompjuterske vizije koja ima široku primenu u različitim aplikacijama, uključujući prepoznavanje registarskih tablica vozila, prepoznavanje rukopisa, prepoznavanje dokumenata i mnoge druge. Glavni cilj automatskog prepoznavanja teksta je pretvaranje vizuelno prikazanog teksta u format koji računari mogu razumeti i obrađivati, omogućavajući im da interpretiraju tekstualne informacije slično kao što to radi čovek.

Automatsko prepoznavanje teksta sa registarskih tablica automobila obuhvata nekoliko ključnih koraka koji se odvijaju u procesu od prikupljanja podataka do konačne integracije sistema u softver za prepoznavanje tablica automobila.

Prikupljanje raznovrsnog skupa slika registarskih tablica vozila ključno je za uspešno treniranje modela. Ove slike treba da obuhvataju različite tipove tablica, različite uslove osvetljenja i pozadine kako bi model bio što robustniji. Nakon prikupljanja, slike treba pažljivo razvrstati na one koje su pogodne za treniranje modela i one koje nisu. Ovo uključuje filtriranje slika sa veoma lošim kvalitetom, zamućenim ili nejasnim tablicama. Kako bi se obogatio skup podataka i poboljšala generalizacija modela, potrebno je primeniti tehnikе augmentacije podataka. Ovo uključuje manipulaciju sa slikama kao što su rotacija, promena osvetljenja, izobličenja i dodavanje šuma. Pored toga, sintetički podaci se mogu generisati korišćenjem programa za generisanje tablica sa tekstrom. Svaka slika mora biti precizno označena sa tačnim tekstualnim sadržajem registarske tablice kako bi se koristila za obuku modela. Ovaj proces može biti ručan ili se može koristiti alat za automatsko lejbelovanje. Nakon pripreme podataka, sledi faza treniranja mreže za prepoznavanje teksta. U ovoj fazi, model se obučava nad označenim podacima kako bi naučio da prepozna tekst sa slika tablica. Kada je model obučen, integriše se u softver za prepoznavanje tablica automobila. Ovaj softver obično obuhvata module za detekciju tablica, detekciju teksta na tablicama, formatiranje izlaza i druge funkcionalnosti.

Kako bi omogućili portabilnost i lakšu distribuciju sistema za prepoznavanje tablica, koristi se Docker kontejner. Docker omogućava pakovanje softverskih aplikacija i njihovo pokretanje u izolovanim okruženjima. Još jedna od bitnih komponenata je Python web framework - FastAPI koji omogućava brzo kreiranje API-ja za komunikaciju sa softverskim komponentama. Integracija Docker-a i FastAPI modula omogućava da se servis za prepoznavanje teksta koristi nezavisno od platforme na kojoj se izvršava, čineći ga pristupačnim i jednostavnim za upotrebu u različitim okruženjima.

2 Prepoznavanje teksta

2.1 Uvod u prepoznavanje teksta

Prepoznavanje teksta na sceni ima za cilj da tekst sa slike konvertuje u digitalni niz karaktera, što prenosi semantiku visokog nivoa ključnu za razumevanje scene. Zadatak prepoznavanja teksta sa scene je izazovan zbog varijacija u: deformacijama teksta, fontovima, preklapanjima različitih tekstova, prekompleksnim pozadinama, itd. Dodatno, otežavajući faktor može biti i to što se tekst može pojaviti pod različitim uglovima.



Slika 1: Tekst na sceni sa različitim fontovima, pozadinama, osvetljenjem i slično

U proteklim godinama uloženi su brojni naporci kako bi se poboljšala tačnost prepoznavanja teksta. Moderni pristupi za prepoznavanje teksta, pored tačnosti, takođe uzimaju u obzir i faktore poput brzine izvršavanja modela zbog praktičnih zahteva.

Metodološki, prepoznavanje teksta na sceni može se posmatrati kao prelazak iz modaliteta slike u niz karaktera. Obično, prepoznavanje teksta se sastoji od dva osnovna dela, vizuelnog modela za ekstrakciju karakteristika i sekvencijskog modela za transkripciju teksta.

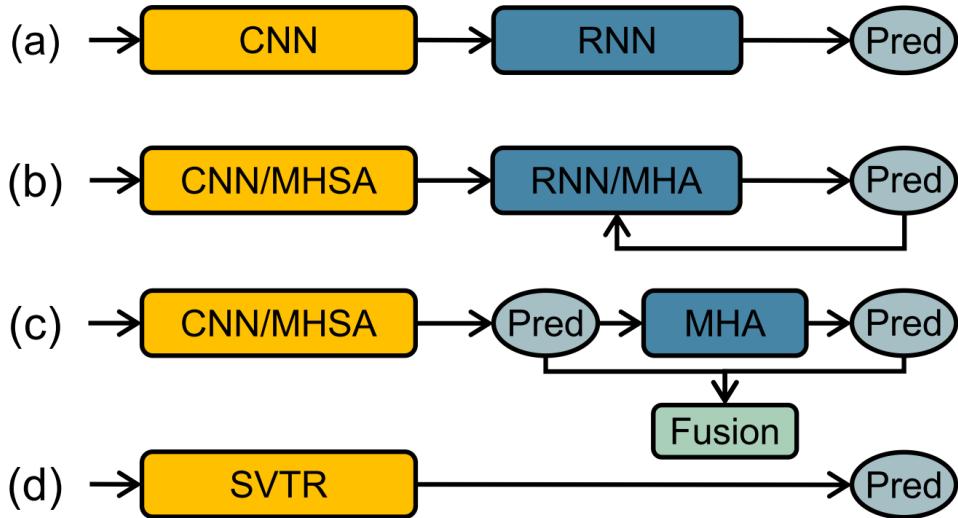
2.2 Istorijski pregled prepoznavanja teksta

Prvi primeri i prva faza tehnologije optičkog prepoznavanja karaktera(OCR) pojavili su se sredinom 20. veka, pretežno tokom 1950-ih i 1960-ih godina. Ovo doba obeležilo je razvoj ranih sistema OCR-a, koji su koristili osnovne tehnike prepoznavanja obrazaca kako bi prepoznali mašinski odštampane karaktere. Ovi sistemi su često bili ograničeni na prepoznavanje određenih fontova i imali su relativno nisku stopu tačnosti u poređenju sa modernom OCR tehnologijom. Glavna primena im je bila čitanje standardizovanih obrazaca i dokumenata sa jasno štampanim tekstom i poznatim fontom.

Druga faza tehnologije optičkog prepoznavanja karaktera dogodila se krajem 20. veka i početkom 21. veka, počevši oko 1970-ih i nastavljujući se u 2000-ima. Ovo doba je obeleženo značajnim napretkom u tehnologiji OCR-a, uključujući razvoj sofisticiranih algoritama i tehnika za prepoznavanje karaktera. Ovi napredci doveli su do veće tačnosti i mogućnosti prepoznavanja šireg spektra fontova, jezika i rasporeda dokumenata. Dodatno, integracija pristupa mašinskog učenja i neuronskih mreža doprinela je daljem poboljšanju performansi OCR-a. U ovoj fazi došlo je do primene OCR sistema u širem spektru aplikacija, od skeniranja i konverzije dokumenata u digitalno arhiviranje, do automatizovanog unosa podataka i ekstrakcije teksta u različitim industrijama.

Treća faza tehnologije optičkog prepoznavanja karaktera je trenutno aktuelna i predstavlja trenutno stanje napretka u sistemima OCR-a. Ovo doba karakteriše integracija najnovijih tehnologija poput dubokog učenja, konvolucionih neuronskih mreža (CNN) i rekurentnih neuronskih mreža (RNN) u algoritme OCR-a. Ove napredne tehnike značajno su poboljšale tačnost i pouzdanost sistema OCR-a, omogućavajući prepoznavanje složenih dokumenata sa različitim fontovima, rasporedima i jezicima. Osim toga, pojava OCR usluga zasnovanih na cloud-u i integracija OCR funkcionalnosti u mobilne uređaje učinili su OCR dostupnijim i svestranijim nego ikad ranije. Treća faza takođe obuhvata napretke u analizi i razumevanju dokumenata, omogućavajući OCR sistemima da izvlače ne samo tekst već i strukturalne i semantičke informacije iz dokumenata, što dovodi do poboljšanih sposobnosti obrade dokumenata i pretraživanja informacija.

2.3 Arhitekture modela za prpoznavanje teksta na slici



Slika 2: Arhitekture modela za prepoznavanje teksta sa scene. (a) Modeli zasnovani na CNN-RNN. (b) Modeli kodiranja-dekodiranja. (c) Vizuelno-jezički modeli. (d) SVTR, koji prepoznaće tekst scene sa jedinstvenim vizuelnim modelom i odlikuje se efikasnošću, tačnošću i višejezičnom svestranošću.

Modeli zasnovani na CNN-RNN [SBY15] prvo koriste CNN za ekstrakciju karakteristika. Karakteristike se zatim preoblikuju u sekvencu koju BiLSTM modeluje uz pomoć CTC gubitka kako bi generisao predikciju (Slika 2(a)). Odlikuju se efikasnošću i ostaju izbor za neke komercijalne proizvode za prepoznavanje teksta sa scene. Međutim, preoblikovanje karakteristika u sekvencu je osetljivo na deformacije teksta, što ograničava efikasnost takvih modela.

Kasnije su pristupi zasnovani na auto-regresivnim metodama kodera-dekodera postale popularne [SCX19; Li+19; Zhe+23]. Ove metode transformišu prepoznavanje u iterativni proces dekodiranja (Slika 2(b)). Kao rezultat, postignuta je poboljšana tačnost jer je uzeta u obzir kontekstualna informacija. Međutim, brzina izvođenja je spora zbog transkripcije karakter po karakter. Ovaj postupak je dodatno proširen na softversku strukturu zasnovanu na viziji i jeziku [Yu+20; Fan+21], gde je jezičko znanje uključeno (Slika 2(c)) i sprovedena je paralelna predikcija. Ipak, ovaj postupak često zahteva model sa velikim kapacitetom ili složenu paradigmu prepoznavanja kako bi se osigurala tačnost prepoznavanja, što ograničava njegovu efikasnost.

U poslednje vreme, naglasak je na razvoju pojednostavljenih arhitektura kako bi se dobilo na brzini izvršavanja. Na primer, korišćenje složene paradigmе obuke, ali jednostavnog modela za izvršavanje. Rešenje zasnovano na CRNN-RNN ponovo je pregledano u sledećem radu: [Hu+20]. Koristi mehanizam pažnje i grafovsku neuronsku mrežu za agregaciju sekvenčijalnih karakteristika koje odgovaraju istom karakteru. Pri izvršavanju, deo za modelovanje zasnovan na mehanizmu pažnje je odbačen kako bi se uskladila tačnost i brzina.

Nedavni uspeh transformera za obradu slike [Dos+21; Liu+21], inspirisao je nastanak jedinstvenog vizuelnog modela za prepoznavanje teskta na sceni(SVTR) [Du+22]. SVTR najpre razlaže tekst slike na male 2D isečke koji se nazvaju komponente karaktera, od kojih svaka komponenta može sadržati samo deo karaktera. Tokenizacija slike po isećcima praćena mehanizmom samopaznje se primenjuje da bi se uhvatile indicije prepoznavanja teksta među komponentama karaktera. Za ovu svrhu je razvijena prilagođena arhitektura za tekst, čija osnovna struktura sadrži progresivno smanjujuću visinu mape karakteristika u tri faze i uključuje operacije mešanja, spajanja i/ili kombinovanja. Osmišljeni su lokalni i globalni blokovi mešanja koji se rekurzivno primenjuju u svakoj fazi, zajedno sa operacijom spajanja ili kombinovanja, stičući tako afinitete na nivou lokalnih komponenti koje predstavljaju karakteristike slične potezima karaktera i dugoročne zavisnosti između različitih karaktera. Dakle, osnovna struktura ekstrahuje karakteristike komponenti na različitim rastojanjima i na više skala, formirajući višeslojnu percepciju karakteristika karaktera. Kao rezultat, prepoznavanje teksta se postiže jednostavnom linearnom predikcijom. U celom procesu koristi se samo jedan vizuelni model (Slika 2(d)).

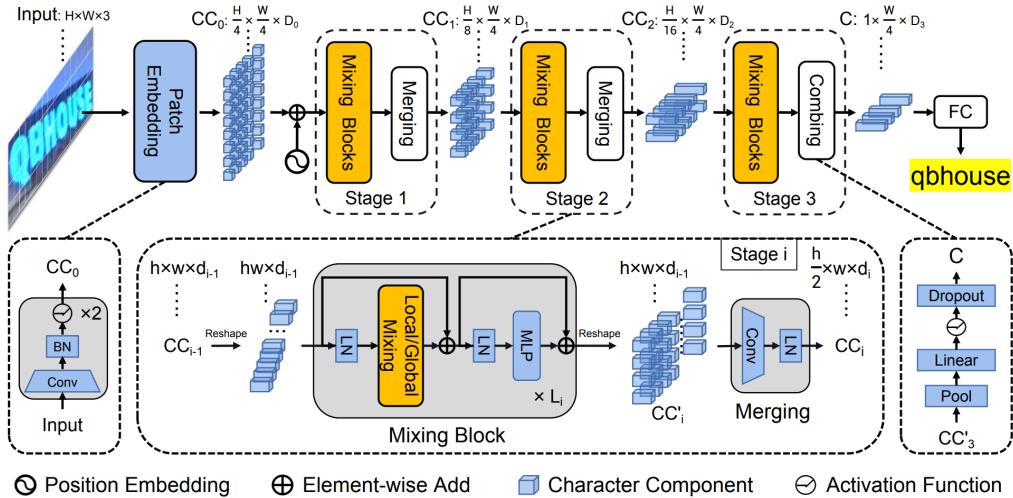
2.4 Korišćenje Jedinstvenog Vizuelnog Modela za Prepoznavanje Teksta na Sceni

Tradicionalni modeli za prepoznavanje teksta obično uključuju dve odvojene komponente: vizuelni model za izdvajanje karakteristika sa slike i sekvenčijalni model za dekodiranje izdvojenih karakteristika u tekst. Jedinstveni vizuelni model za prepoznavanje teksta na sceni(SVTR) eliminiše potrebu za sekvenčijalnim modelom u potpunosti, čineći ga jednostavnijim i efikasnijim.

Uklanjanjem komponente sekvenčijalnog modeliranja, SVTR postiže konkurentnu preciznost na zadacima prepoznavanja teksta, pružajući veću efikasnost u poređenju sa tradicionalnim metodama.

2.4.1 Arhitektura

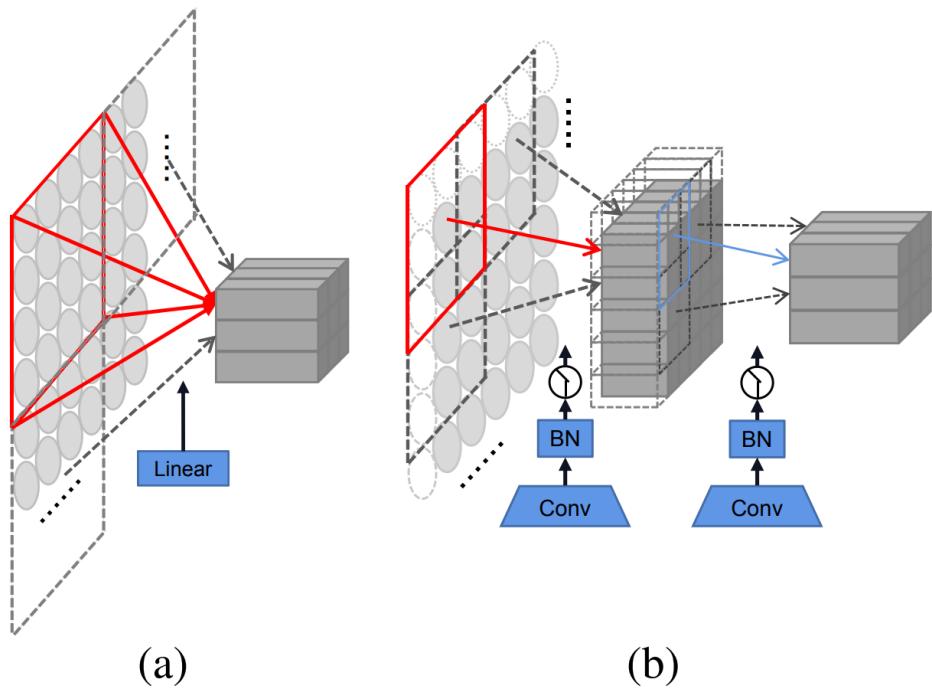
Pregled SVTR modela je prikazan na (Slika 3) i predstavlja tro-faznu mrežu sa progresivno smanjujućom visinom, namenjenu za prepoznavanje teksta. Slika teksta veličine $H \times W \times 3$, prvo se transformiše u $\frac{H}{4} \times \frac{W}{4}$ isečaka dimenzije D_0 koristeći progresivno preklapajuće ugrađivanje isečaka. Isečci predstavljaju karakterne(znakovne) komponente, od kojih svaka odgovara delu tekstualnog karaktera na slici. Zatim se izvode tri faze, od kojih se svaka sastoji od niza blokova za mešanje praćenih operacijom spajanja ili kombinovanja, na različitim skalama za ekstrakciju karakteristika. Osmišljeni su lokalni i globalni blokovi za mešanje za ekstrakciju lokalnih obrazaca nalik potezima i hvatanje međukomponentne zavisnosti. Pomoću osnovne strukture se karakterizuju komponentne karakteristike i zavisnosti na različitim udaljenostima i na više skala, predstavljene kao C veličine $1 \times \frac{W}{4} \times D_3$, koje percipira karakteristike znakova na više nivoa granularnosti. Na kraju procesa, model istovremeno predviđa sve znakove sa ulazne slike i primenjuje postupak uklanjanja duplikata kako bi se eliminisali eventualno pogrešno ponovljeni karakteri koje je model predviđao, a koji nisu stvarno prisutni na originalnoj slici. Rezultat ovog procesa je konačan niz prepoznatih znakova.



Slika 3: Arhitektura SVTR-a: Mreža koja kroz tri faze progresivno smanjuje visinu mape karakteristika. U svakoj fazi se izvodi niz blokova za mešanje, nakon čega sledi operacija spajanja ili kombinovanja. Na kraju se prepoznavanje vrši linearnim predviđanjem.

2.4.2 Progresivno preklapajuće ugrađivanje isečaka

Pvri korak u obradi slike teksta je njeno razlaganje na manje delove koje nazivamo isećcima. Dobijanje karakterističnih isečaka koji predstavljaju komponente znakova znači prelazak iz $X \in \mathbb{R}^{H \times W \times D_0}$ u $CC_0 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times D_0}$. Postoje dva uobičajena načina da se ovo uradi — korišćenje 4×4 mreže za podelu slike i linearna transformacija svakog dela(Slika 4(a)) i korišćenje 7×7 konvolucionog filtera sa korakom 4. Autori SVTRa su izabrali alternativni metod. Oni koriste dva manja konvolucionala filtera 3×3 jedan za drugim sa korakom 2, kao što je prikazano na (Slika 4(b)). Takođe koriste tehniku zvanu normalizacija serije da bi održali brojeve pod kontrolom. Ovaj novi metod zahteva nešto više računarske snage, ali je bolji u kombinovanju karakteristika iz slike.



Slika 4: (a) Linearna projekcija u ViT [Dos+21]. (b) SVTR progresivno preklapajuće ugrađivanje isečaka.

pronađi
pri-
klad-
niju-
reč
od
me-
ša-
nje

2.4.3 Blok mešanja

S obzirom na to da se dva karaktera mogu blago razlikovati važno je posmatrati male delove koji čine karaktere. Prepoznavanje teksta se u velikoj meri oslanja na ekstrakciji karakteristika na nivou komponenti karaktera. Međutim, postojeće studije uglavnom koriste niz karakteristika za predstavljanje teksta na slici. Svaka karakteristika odgovara deliću regiona slike, koji je često nerazumljiv, posebno za nepravilan tekst — što nije optimalno za opisivanje karaktera. Nedavni napredak u vizuelnim transformatorima uvodi 2D reprezentaciju karakteristika, ali njeno korišćenje u kontekstu prepoznavanja teksta je još uvek u fazi istraživanja. Autori SVTRa sugerisu da su dve vrste karakteristika važne za prepoznavanje teksta. Prva su lokalni obrazci, kao što su mali detalji koji čine karakter, poput poteza. Oni pokazuju kako su različiti delovi karaktera međusobno povezani i stvaraju se morfološke karakteristike i korelacije između različitih delova karaktera. Druga su međukarakterne zavisnosti, koje se odnose na to kako su različiti karakteri povezani jedni s drugima, ili kako se tekst odnosi na netekstualne delove slike. Da bi uhvatili ove karakteristike, autori su kreirali dva posebna bloka mešanja. Ovi blokovi koriste tehniku zvanu samopažnja, koja pomaže modelu da se fokusira na važne delove slike. Koristeći dva različita područja fokusa koja mehanizam samopažnje razmatra, ovi blokovi mogu uhvatiti i male detalje i šиру sliku o tome kako su karakteri međusobno povezani.

Globalno mešanje Kao što se vidi na (Slici 4(a)), globalno mešanje procenjuje zavisnost među svim komponentama karaktera. S obzirom da su tekstualni i ne-tekstualni sadržaj dva glavna elementa na slici, takvo generalno mešanje može uspostaviti dugoročnu zavisnost među komponentama različitih karaktera. Pored toga, ono je takođe sposobno da oslabi uticaj ne-tekstualnih komponenti, istovremeno pojačavajući važnost tekstualnih komponenti. Matematički, za komponente karaktera CC_{i-1} iz prethodne faze, prvo se vrši njihovo preoblikovanje u niz karakteristika. Pri uvođenju u blok mešanja, primenjuje se normalizacija sloja, a zatim se koristi multi-head samopažnja za modelovanje zavisnosti. Nakon toga, sekvensijalno se primenjuju normalizacija sloja i MLP za fuziju karakteristika. Zajedno sa prečicama, formira se blok globalnog mešanja.

Lokalno mešanje Kao što se vidi na (Slici 4(b)), lokalno mešanje procenjuje korelaciju među komponentama unutar unapred definisanog prozora. Njegov cilj je da kodira morfološke karakteristike karaktera i uspostavi veze između komponenti unutar jednog karaktera, što simulira karakteristiku nalik potezu koja je vitalna za identifikaciju karaktera. Za razliku od glo-

balnog mešanja, lokalno mešanje razmatra okolinu za svaku komponentu. Slično konvoluciji, mešanje se odvija koristeći pristup klizećeg prozora. Veličina prozora je empirijski postavljena na 7×11 . U poređenju sa globalnim mešanjem, lokalno implementira mehanizam samopažnje za detekciju lokalnih obrazaca. Kao što je prethodno pomenuto, dva bloka mešanja imaju za cilj izvlačenje različitih karakteristika koje su komplementarne. U SVTR-u, blokovi se rekurentno primenjuju više puta u svakoj fazi za sveobuhvatnu ekstrakciju karakteristika.

2.4.4 Spajanje

Održavanje konstantne prostorne rezolucije kroz faze je računski skupo, što takođe dovodi i do redundantnosti reprezentacije karakteristika kroz slojeve. Kao posledica toga, autori SVTR-a osmišljavaju operaciju spajanja nakon blokova mešanja u svakoj fazi (osim u poslednjoj). Karakteristikama koje su izlaz iz poslednjeg bloka mešanja, se prvo menja dimenzija u veličinu $h \times w \times d_{i-1}$, gde h , w i d_{i-1} označavaju trenutnu visinu, širinu i broj kanala, redom. Zatim se primenjuje konvolucija veličine 3×3 sa korakom 2 u dimenziji visine i korakom 1 u dimenziji širine, praćenu normalizacijom sloja, generišući novi sloj dimenzije $\frac{h}{2} \times w \times d_i$.

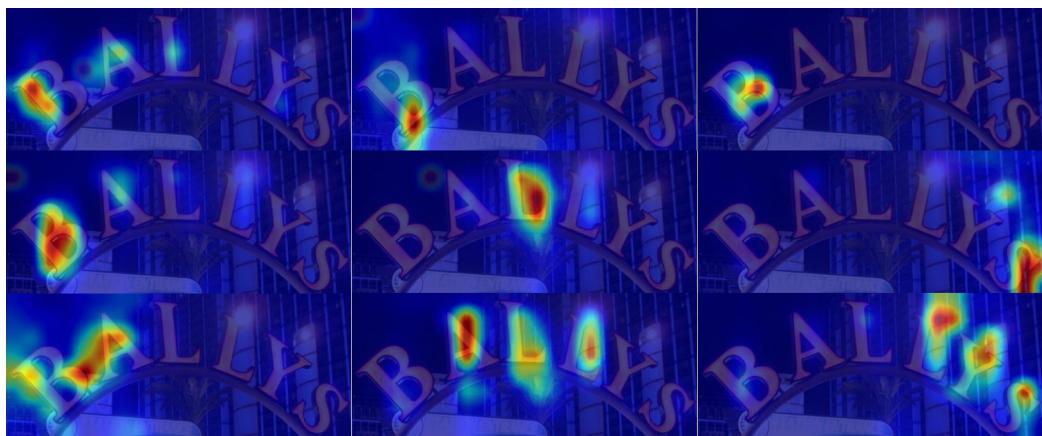
Operacija spajanja prepolovljava visinu dok zadržava konstantnu širinu. Ovo ne samo da smanjuje vremenske troškove obrade, već takođe gradi hijerarhijsku strukturu prilagođenu tekstu. Tipično, većina tekstova na slikama se pojavljuje horizontalno ili blizu horizontalno. Kompresijom dimenzije visine i dalje ostaje uspostavljena višeskalarna reprezentacija svakog karaktera, a pritom ne utiče na raspored isečaka u dimenziji širine. Stoga, smanjivanje dimenzije visine ne povećava šanse za kodiranje susednih isečaka u istu komponentu kroz faze. Takođe, povećava se dimenzija kanala d_i kako bi se nadoknadio gubitak informacija.

2.4.5 Kombinovanje i Predikcija

U poslednjoj fazi, operacija spajanja se zamjenjuje operacijom kombinovanja. Prvo se dimenzija visine svede na 1, a zatim se primenjuje potpuno povezani sloj, nelinearna aktivacija i dropout. Na taj način, komponente karaktera se dodatno kompresuju u sekvencu karakteristika, gde je svaki element predstavljen karakteristikom dužine D_3 . U poređenju sa operacijom spajanja, operacija kombinovanja može da izbegne primenu konvolucije na slojevima čija je veličina veoma mala u jednoj dimenziji, npr. ukoliko je dimenzija visine 2.

Sa kombinovanim karakteristikama, implementirano je prepoznavanje teksta koristeći jednostavne paralelne linearne predikcije. Konkretno, koristi se linearni klasifikator sa N čvorova. On generiše transkripciju sekvencu veličine $\frac{W}{4}$, gde idealno, komponente istog karaktera bivaju transkribovane kao duplikati karaktera, a komponente ne-teksta se transkribuju u prazan simbol. Sekvenca se automatski kondenzuje u konačni rezultat. U implementaciji, N je postavljen na 37 za engleski jezik i 6625 za kineski jezik.

2.4.6 Analiza Vizualizacije



Slika 5: Vizualizacija SVTR mapi pažnje

Svaka mapa se može objasniti kao da ima različitu ulogu u celokupnom prepoznavanju. Ilustracija devet primjera mapa je prikazana na (Slici 5). Prvi red prikazuje tri mape koje se fokusiraju na deo karaktera "B", sa naglaskom na njegovu levu stranu, donji deo i srednji deo, redom. Te tri mape ukazuju na to da različiti regioni karaktera doprinose njegovom prepoznavanju. Drugi red prikazuje tri mape koje se fokusiraju na različite karaktere, tj. "B", "L" i "S". SVTR takođe može da nauči karakteristike karaktera posmatrajući karakter kao celinu. Treći red prikazuje tri mape koje istovremeno aktiviraju više karaktera, što implicira da su zavisnosti među različitim karakterima uspešno uhvaćene. Ova tri reda zajedno otkrivaju da prepoznavać hvata trageve na nivou dela karaktera, celog karaktera i više karaktera, u skladu sa tvrdnjom da SVTR percipira višeslojne karakteristike komponenti karaktera, potvrđujući efikasnost SVTR-a.

3 Primene

Sistem za automatsko prepoznavanje teksta sa tablica vozila (ANPR) predstavlja inovativnu tehnologiju koja se sve više integriše u različite aspekte svakodnevnog života i upravljanja saobraćajem. Ova tehnologija omogućava brzo i precizno očitavanje registarskih oznaka vozila, što otvara brojne mogućnosti primene u različitim sektorima. Korišćenjem ANPR sistema, moguće je unaprediti efikasnost naplate putarina, automatski pratiti i locirati ukradena vozila, kao i optimizovati procese parkiranja i upravljanja kolinom vozila.

Pored toga, ANPR može igrati ključnu ulogu u poboljšanju bezbednosti saobraćaja, omogućavajući automatsku kontrolu saobraćajnih prekršaja i identifikaciju vozila koja su uključena u kriminalne aktivnosti. U kontekstu pametnih gradova, ova tehnologija može doprineti održivijem urbanom razvoju kroz efikasnije upravljanje saobraćajem i smanjenje zagađenja. S obzirom na sve ove prednosti, implementacija sistema za automatsko prepoznavanje registarskih tablica postaje ne samo korisna, već i neophodna za modernizaciju i unapredjenje infrastrukture i usluga u urbanim sredinama. U ovom radu biće razmotrene različite primene ANPR sistema, kao i potencijalni izazovi i rešenja u njegovoj implementaciji.

3.1 Automatska naplata parkiranja

ANPR predstavlja značajan napredak u sistemima za naplatu parkiranja, omogućavajući bržu, efikasniju i korisniku prijatniju uslugu. Ova tehnologija omogućava automatsko očitavanje registarskih oznaka vozila prilikom ulaska i izlaska sa parking prostora, čime se eliminiše potreba za fizičkim karticama ili novčanim transakcijama na licu mesta.

Korišćenjem ANPR sistema, korisnici mogu jednostavno parkirati svoje vozilo bez dodatnog čekanja, dok se naplata vrši automatski putem unapred registrovanih podataka o vozilu. Ovo ne samo da smanjuje gužve na parking mestima, već i poboljšava korisničko iskustvo, čime se povećava zadovoljstvo vozača. Pored toga, ovakvi sistemi omogućavaju efikasnije upravljanje parking kapacitetima, jer pružaju ažurne informacije o zauzetosti parkinga, što može pomoći u optimizaciji korišćenja prostora.

Dodatno, ANPR tehnologija može doprineti smanjenju prevara i zloupotreba, jer se automatski evidentiraju svi ulazi i izlazi vozila, čime se povećava sigurnost i transparentnost u procesu naplate. Sve ove prednosti čine automatsko prepoznavanje registarskih tablica ključnim elementom modernih sistema za naplatu parkiranja, koji se sve više koriste u urbanim sredinama.

3.2 Automatsko naplate putarina

ANPR može značajno unaprediti sistem automatske naplate putarina. ANPR sistem omogućava brzo i precizno očitavanje registarskih oznaka vozila prilikom prolaska kroz naplatne stanice, čime se eliminiše potreba za zaustavljanjem i fizičkim plaćanjem. Kao rezultat toga, vozači mogu nesmetano prolaziti kroz naplatne rampe, što smanjuje gužve i poboljšava protok saobraćaja na putevima.

Korišćenjem ANPR tehnologije, naplata putarine postaje efikasnija i transparentnija, jer se automatski evidentiraju podaci o svakom vozilu, uključujući vreme prolaska i iznos naplate. Ovaj sistem takođe omogućava lakše praćenje i upravljanje naplatom, smanjujući mogućnost grešaka i prevara. Pored toga, ANPR može pomoći u identifikaciji vozila koja su prijavljena kao ukradena, čime se dodatno povećava bezbednost na putevima.

3.3 Automatsko praćenje i lociranje ukradenih vozila

ANPR omogućava efikasno praćenje i lociranje ukradenih vozila. Ovaj sistem omogućava brzu identifikaciju registarskih oznaka vozila u realnom vremenu, čime se nadležnim organima pruža mogućnost da odmah reaguju na prijave o ukradenim vozilima. Kada ANPR kamere prepoznaјu registarsku tablicu koja se nalazi na listi ukradenih vozila, automatski se šalje obaveštenje nadležnim organima, čime se povećava šansa za brzo pronalaženje i vraćanje vozila vlasnicima. Osim što poboljšava efikasnost policijskih operacija, ANPR tehnologija takođe može pomoći u smanjenju stope kriminala, jer deluje kao odvraćajući faktor za potencijalne počinioce.

3.4 Automatska kontrola saobraćajnih prekršaja

ANPR olakšava precizno evidentiranje različitih prekršaja u realnom vremenu. Ova tehnologija omogućava automatizovano prepoznavanje registarske oznake vozila koja krše saobraćajne propise, kao što su prekoračenje brzine, prolazak kroz crveno svetlo ili nepropisno parkiranje. Kada se prekršaj zabeleži, sistem automatski generiše obaveštenje ili kaznu koja se šalje vlasniku vozila, čime se pojednostavljuje proces naplate kazni.

Jedna od ključnih prednosti ANPR sistema je njegova sposobnost da smanji subjektivnost u procesu kontrole saobraćaja. Automatska identifikacija prekršaja omogućava doslednu primenu zakona, što može doprineti povećanju bezbednosti na putevima. Pored toga, ANPR može pomoći u prikupljanju podataka o saobraćajnim tokovima i obrascima ponašanja vozača, što omo-

gućava vlastima da bolje planiraju saobraćajnu infrastrukturu i strategije prevencije.

U javnom prevozu, ALPR može pomoći u praćenju i regulisanju vozila koja koriste specijalizovane trake, kao što su trake za autobuse ili taksi vozila, osiguravajući da ih koriste samo ovlašćena vozila. Ovo može doprineti efikasnijem funkcionisanju javnog prevoza i smanjenju zagušenja na putevima.

Korišćenjem ANPR tehnologije, moguće je stvoriti efikasniji sistem kontrole saobraćaja koji smanjuje broj prekršaja i podstiče vozače da se pridržavaju saobraćajnih pravila. U svetu ovih prednosti, jasno je da automatsko prepoznavanje registarskih tablica igra ključnu ulogu u unapređenju bezbednosti saobraćaja i smanjenju rizika od saobraćajnih nesreća.

3.5 Automatska kontrola pristupa

ANPR pruža efikasnu kontrolu i povećava bezbednost pristupa određenim zonama ili objektima. Ovakav sistem može biti od velike koristi u kontroli pristupa ekološkim zonama u gradovima, gde se ograničava pristup vozilima koja ne ispunjavaju ekološke standarde. Automatsko prepoznavanje registarskih tablica omogućava brzu identifikaciju takvih vozila i sprečavanje njihovog ulaska u zaštićene zone, čime se doprinosi smanjenju zagađenja i poboljšanju kvaliteta vazduha.

Korišćenjem ANPR tehnologije, moguće je stvoriti fleksibilne i efikasne sisteme kontrole pristupa koji se mogu prilagoditi specifičnim potrebama svake zone ili objekta. Ova tehnologija predstavlja ključni element u unapređenju bezbednosti i upravljanju pristupom u savremenim urbanim sredinama. Može se koristiti za automatsko otvaranje kapija ili rampi u stambenim kompleksima ili poslovnim prostorima.

Jedna od inovativnih primena ANPR tehnologije je u automatizovanim autoperionicama, gde se prepoznavanje registarskih tablica može koristiti za pružanje personalizovane usluge pranja vozila. Ovaj sistem može zapamtiti prethodne posete i preferencije klijenata, omogućavajući im da dobiju uslugu prilagođenu njihovim potrebama bez potrebe za dodatnim informacijama. Na taj način, klijenti mogu uživati u bržem i efikasnijem procesu pranja, dok autoperionice mogu optimizovati svoje operacije i poboljšati korisničko iskustvo.

U stambenim ili garažnim zgradama koje imaju rezervisana parking ili garažna mesta, ANPR tehnologija eliminiše potrebu za više daljinskih upravljača za podizanje rampe i otvaranje garažnih vrata. Ovaj sistem omogućava automatsko otvaranje na osnovu prepoznate registarske tablice, čime se smanjuje upotreba plastike i doprinosi očuvanju životne sredine. Osim toga, korisnici više ne moraju da se obavezuju da nose više daljinskih upravljača,

što dodatno olakšava svakodnevno korišćenje i povećava praktičnost.

3.6 Automatska kontrola saobraćaja

ALPR može imati ključnu ulogu u automatizovanoj kontroli saobraćaja, pružajući brojne prednosti za hitne službe i javni prevoz. Ova tehnologija može značajno poboljšati brzinu reakcije hitnih službi, kao što su ambulante, vatrogasci i policija, omogućavajući im brži dolazak na lice mesta. ALPR može prepoznati vozila hitnih službi i automatski otvoriti posebne saobraćajne trake ili raskrsnice, čime se smanjuje vreme potrebno za prolazak kroz gust saobraćaj. Ova funkcionalnost može biti od vitalnog značaja u situacijama kada je svaka sekunda važna, čime se potencijalno spašavaju životi.

U kontekstu javnog prevoza, ALPR može pratiti i optimizovati rute autobusa i drugih prevoznih sredstava, čime se povećava tačnost i efikasnost servisa za obaveštavanje putnika o lokaciji i vremenu dolaska. Ova tehnologija omogućava prikupljanje podataka o saobraćajnim tokovima, što može pomoći u analizi i unapređenju rasporeda vožnje, čime se poboljšava iskustvo putnika.

3.7 Kreiranje personalizovanih reklama

ALPR može značajno unaprediti marketing i oglašavanje kroz kreiranje personalizovanih reklama usmerenih na vozače. Na primer, kada vozač prolazi pored bilborda, sistem može automatski prepoznati registarsku tablicu i prikazati reklamu koja odgovara njegovom vozilu ili prethodnim posetama određenim lokacijama. Ovo može uključivati promocije za slične brendove automobila, dodatnu opremu ili usluge vezane za automobil, čime se povećava verovatnoća da će vozač obratiti pažnju na oglas i reagovati na njega.

4 Prikupljanje i rad sa podacima

U ovom poglavlju biće objašnjene metode prikupljanja podataka, izazovi sa kojima se istraživači susreću tokom ovog procesa, kao i značaj obrade i pripreme podataka za dalji rad na razvoju sistema za prepoznavanje teksta sa registarskih tablica.

4.1 Uvod u prikupljanje podataka

Osnovu za razvoj sistema za automatsko prepoznavanje teksta sa tablica čini kvalitetan i obiman skup podataka. Prikupljanje ovih podataka predstavlja jedan od ključnih koraka u procesu razvoja takvog sistema, jer direktno utiče na uspešnost i preciznost modela za prepoznavanje. Zbog toga je neophodno posvetiti posebnu pažnju metodama i tehnikama koje se koriste za prikupljanje podataka, kao i izazovima koji mogu nastati u ovom procesu.

Prikupljanje podataka podrazumeva sakupljanje slika registarskih tablica iz različitih izvora, koje će se koristiti za treniranje, validaciju i testiranje modela. Kvalitet prikupljenih podataka, njihova reprezentativnost i raznovrsnost igraju ključnu ulogu u postizanju visokih performansi sistema za prepoznavanje teksta. Osim toga, treba uzeti u obzir i pravne aspekte i zaštitu privatnosti, jer rad sa podacima koji sadrže informacije o vozilima može biti podložan strogoj regulativi.

4.2 Specifičnosti prikupljanja podataka u kontekstu prepoznavanja teksta sa registarskih tablica



Slika 6: Primeri tablica iz skupa realnih podataka. Svi priloženi primeri tablica su svesno oštećeni kako bi se sačuvala anonimnost i privatnost osoba čije tablice su prikazane.

4.2.1 Varijabilnost registrarskih tablica

Raznolikost dizajna Registrarske tablice variraju od zemlje do zemlje, pa čak i unutar iste države, u pogledu boje, fonta, veličine slova, rasporeda karaktera i prisustva simbola ili grbova. Ova raznolikost zahteva prikupljanje podataka koji obuhvataju širok spektar različitih tablica kako bi model bio sposoban da prepozna sve varijante.

Fizičko stanje tablica Tokom vremena, registrarske tablice mogu postati oštećene, izbledele ili prljave, što može otežati prepoznavanje teksta. Prikupljanje slika sa različitim stepenima oštećenja tablica je ključno za treniranje modela koji može da se nosi sa takvim izazovima.

4.2.2 Promenljivi uslovi snimanja

Različiti uslovi osvetljenja Snimanje registrarskih tablica može se odvijati u različitim vremenskim uslovima i periodima dana, što rezultira varijacijama u osvetljenju. Sakupljanje podataka u različitim svetlosnim uslovima (npr. jaka sunčeva svetlost, senke, noćno snimanje) omogućava modelu da se prilagodi tim promenama (Slika 6).

Snimanje noću predstavlja poseban izazov zbog nedostatka prirodnog svetla. Kamere koje nemaju odgovarajuće noćne režime snimanja ili infracrveno osvetljenje mogu proizvesti slike lošeg kvaliteta sa dosta šuma. Dodatno, fariovi vozila mogu uzrokovati probleme sa kontrastom i zaslepljivanjem, što dodatno otežava prepoznavanje tablica.

Različiti uglovi snimanja U zavisnosti od položaja kamere, registrarske tablice mogu biti snimljene pod različitim uglovima, što može izazvati probleme sa perspektivnim izobličenjem. Kada kamera nije postavljena direktno ispred tablice, već pod određenim uglom, karakteri na tablici mogu izgledati iskrivljeno ili izduženo. Ovo posebno dolazi do izražaja u slučajevima kada kamere moraju biti montirane na nepristupačnim mestima, poput visokih stubova ili nadstrešnica, gde se tablice snimaju pod oštrim uglovima u odnosu na horizontalnu osu vozila.

Kretanje vozila Snimanje registrarskih tablica vozila koja se kreću predstavlja izazov zbog potencijalnog zamagljenja slika. Kada se vozilo kreće, kamera mora da zabeleži sliku u vrlo kratkom vremenskom intervalu kako bi se izbeglo zamagljenje, koje može otežati prepoznavanje karaktera na tablici.

4.3 Upoznavanje sa podacima

Za razvoj sistema za automatsko prepoznavanje teksta sa registarskih tablica, postoji nekoliko baza podataka koje sadrže slike tablica iz različitih zemalja. Ove baze podataka pružaju širok spektar slika sa različitim tipovima registarskih tablica, različitih kvaliteta i uslova snimanja, što ih čini korisnim za treniranje i evaluaciju modela za prepoznavanje teksta. Neki od najpoznatijih javno dostupnih skupova podataka uključuju:

OpenALPR dataset Baza podataka koja sadrži slike registarskih tablica iz različitih zemalja, koje se često koriste za treniranje i testiranje sistema za prepoznavanje teksta. Ovaj dataset obuhvata tablice sa različitim fontovima, bojama i rasporedima karaktera.

CCPD (Chinese City Parking Dataset) Skup podataka koji se fokusira na kineske registarske tablice, ali može biti koristan kao referenca za treniranje modela sa slikama koje obuhvataju različite uslove osvetljenja i uglove snimanja.

Car License Plate Dataset Baza podataka sa oko 400 slika vozila i registarskih tablica, koja pokriva različite scenarije, uključujući vozila u pokretu, snimke iz različitih perspektiva, i različite vremenske uslove.

Međutim, iako ovi skupovi podataka mogu biti korisni za razvoj generalnih sistema za prepoznavanje teksta sa registarskih tablica, oni su ograničeni u pogledu specifičnosti dizajna i karakteristika registarskih tablica koje se koriste u Srbiji.

S obzirom na to da sam se fokusirao na razvoj rešenja za prepoznavanje teksta sa srpskih registarskih tablica, kao i da imam pristup skupu od oko 17000 slika uglavnom Srpskih tablica sa ulaza ispred rampi, odlučio sam da se oslonim prvenstveno na ove specifične podatke. Ovaj pristup omogućava treniranje modela koji je prilagođen karakteristikama srpskih tablica, kao što su specifičan font, raspored karaktera i prisustvo nacionalnih simbola. Dodatno, slike snimljene na ulazima ispred rampi pružaju realne scenarije sa kojima će se model susretati u stvarnim aplikacijama, što dodatno povećava relevantnost i pouzdanost razvijenog rešenja.

4.4 Razvrstavanje i čišćenje podataka

Nakon prikupljanja, sledi faza čišćenja podataka, koja podrazumeva identifikaciju i uklanjanje nekvalitetnih ili irelevantnih slika. Na primer, slike koje

su previše zamagljene, oštećene, ili ne sadrže jasno vidljive registarske tablice, moraju biti uklonjenje iz skupa podataka. Ova faza obezbeđuje da model bude treniran na kvalitetnim i relevantnim podacima, što utiče na njegovu tačnost i robusnost.

Na kraju, pripremljeni podaci se dele u grupe za treniranje, validaciju i test. Grupa podataka za treniranje se koristi za obučavanje modela, grupa podataka za validaciju korisna je za podešavanje hiperparametara i evaluaciju performansi tokom treniranja, dok se test grupa podataka koristi za konačnu procenu modela nakon treniranja. Pravilna podela podataka je ključna za izbegavanje overfitting-a i osiguranje da model može dobro da generalizuje na novim, neviđenim podacima.

english

4.5 Anotacija podataka

Anotacija predstavlja označavanje delova slike koji sadrže ključne regije, u ovom slučaju registarske tablice i tekst koji je isписан na tablicama. Anotacija podataka je proces u kojem se ručno ili poluautomatski označavaju tačne lokacije i sadržaj registarskih tablica na svakoj slici. U ovoj fazi se koristi softver za anotaciju, koji omogućava obeležavanje granica tablica i unos odgovarajućeg teksta. Kvalitetna anotacija je veoma bitna jer direktno utiče na uspešnost treniranja modela. Greške u anotaciji, poput netačnih ili nepotpunih oznaka, mogu dovesti do smanjenja tačnosti modela i takav tip greške se kasnije vrlo teško pronalazi.

4.6 Kreiranje sintetičkog skupa podataka

Kreiranje sintetičkog skupa podataka pomaže uvećanju inicijalnog skupa i posebno je od značaja kada je dostupnost realnih podataka ograničena.

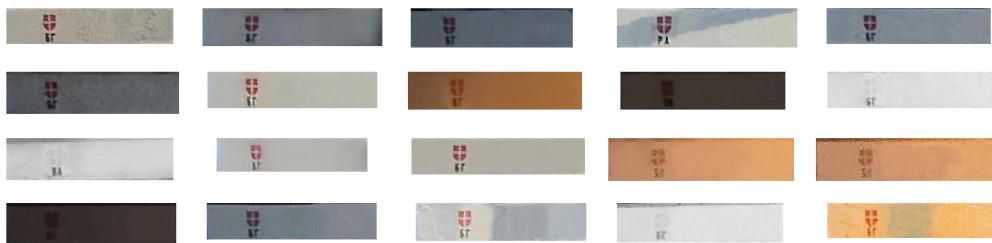
4.6.1 Prednosti sintetičkih podataka

Sintetički skup podataka omogućavaju stvaranje velike količine podataka bez potrebe za obimnim prikupljanjem i anotacijom realnih slika, što može biti vremenski i finansijski zahtevno. Sintetički podaci omogućavaju preciznu kontrolu nad svim aspektima slike, uključujući osvetljenje, ugao snimanja, pozadinu, i stil registarskih tablica (izbor fonta, veličine, boje i pozicije teksta, i slično). Ovo omogućava modelu da se obučava na širokom spektru varijacija, čime se povećava njegova robusnost i otpornost na različite uslove snimanja.

Sintetički data setovi takođe omogućavaju simulaciju retkih ili ekstremnih slučajeva, kao što su tablice sa specifičnim oštećenjima, delimično zaklonjene tablice, ili tablice snimljene u nepovoljnim vremenskim uslovima. Kreiranjem

ovakvih slika, model može biti pripremljen za situacije koje se možda neće često pojavljivati u realnim podacima, ali su ipak važne za sveobuhvatnu tačnost sistema.

4.6.2 Generisanje pozadina tablica



Slika 7: Primeri generisanih pozadina tablica

Za generisanje pozadina tablica, odlučio sam da koristim slike stvarnih registarskih tablica kako bih postigao visok nivo autentičnosti. Prvo sam odabrao nekoliko slika pravih tablica, snimljenih u različitim uslovima, kako bih obezbedio raznovrsnost u teksturi, boji i detaljima. Zatim sam, koristeći GIMP, obrisao sve karaktere sa tih slika, zadržavajući osnovnu strukturu i karakteristike tablice netaknute. Na ovaj način imao sam čiste pozadine koje izgledaju prirodno i verodostojno (Slika 7).

4.6.3 Generisanje teskta na tablicama



Slika 8: Primeri sintetičkog skupa podataka

Nakon generisanja pozadina tablica, koristio sam repozitorijum [TextRecognitionDataGenerator](#), koji sam modifikovao kako bih mogao da na tačno određenim mestima i pod odgovarajućim uglovima dodajem tekst na generisane pozadine tablica. Prethodno sam generisao tekst u skladu sa pravilima za izradu registarskih tablica u Republici Srbiji, vodeći računa da kontekst

bude validan. Ovako prilagođena verzija alata uz korišćenje specijalno izrađenog fonta omogućila mi je da kreiram sintetički skup tablica koje podsećaju na stvarne srpske tablice (Slika 8).

4.6.4 Integracija sintetičkih podataka sa realnim podacima

Kada se kreira sintetički skup podataka, bitno je integrisati ga sa realnim podacima tako da bude postignut optimalan balans između realističnosti i varijabilnosti. Sintetički podaci mogu se koristiti kao dodatak realnim podacima, čime se povećava veličina i raznovrsnost skupa podataka za treniranje. Kombinovanjem sintetičkih i realnih podataka, model može naučiti kako da prepozna tekst sa registarskih tablica u različitim scenarijima, dok istovremeno ostaje veran stvarnim uslovima koje će susretati u praksi.

4.7 Augmentacija podataka

Kako bi se povećala raznovrsnost skupa podataka i model bolje prilagodio različitim uslovima snimanja, često se primenjuje tehnika augmentacije podataka. Augmentacija podrazumeva generisanje novih slika iz postojećih putem različitih transformacija, kao što su rotacija, promena osvetljenja, skaliranje, i dodavanje šuma. Ove transformacije omogućavaju simulaciju različitih realnih uslova, poput snimanja pod različitim uglovima, promena u osvetljenju, ili prisustva šuma na slici. Na ovaj način se model obučava na većem broju različitih scenarija, što doprinosi njegovoj otpornosti na varijacije u podacima.

5 Implementacija

- 5.1 Treniranje modela prepoznavanja teksta**
- 5.2 Komponente sistema**
 - 5.2.1 Detektor tablica**
 - 5.2.2 Detektor teksta**
 - 5.2.3 Prepoznavanje teksta**
- 5.3 Tehnologije**
 - 5.3.1 PaddlePaddle**
 - 5.3.2 Docker**
 - 5.3.3 FastAPI**

6 Rezultati

7 Buduća poboljšanja

8 Zaključak

Literatura

- [SBY15] Baoguang Shi, Xiang Bai i Cong Yao. *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*. 2015. arXiv: [1507.05717 \[cs.CV\]](https://arxiv.org/abs/1507.05717).
- [Li+19] Hui Li i dr. *Show, Attend and Read: A Simple and Strong Baseline for Irregular Text Recognition*. 2019. arXiv: [1811.00751 \[cs.CV\]](https://arxiv.org/abs/1811.00751).
- [SCX19] Fenfen Sheng, Zhineng Chen i Bo Xu. *NRTR: A No-Recurrence Sequence-to-Sequence Model For Scene Text Recognition*. 2019. arXiv: [1806.00926 \[cs.CV\]](https://arxiv.org/abs/1806.00926).
- [Hu+20] Wenyang Hu i dr. „GTC: Guided Training of CTC towards Efficient and Accurate Scene Text Recognition”. U: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.07 (apr. 2020.), str. 11005–11012. DOI: [10.1609/aaai.v34i07.6735](https://doi.org/10.1609/aaai.v34i07.6735). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/6735>.
- [Yu+20] Deli Yu i dr. *Towards Accurate Scene Text Recognition with Semantic Reasoning Networks*. 2020. arXiv: [2003.12294 \[cs.CV\]](https://arxiv.org/abs/2003.12294). URL: <https://arxiv.org/abs/2003.12294>.
- [Dos+21] Alexey Dosovitskiy i dr. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: [2010.11929 \[cs.CV\]](https://arxiv.org/abs/2010.11929). URL: <https://arxiv.org/abs/2010.11929>.
- [Fan+21] Shancheng Fang i dr. *Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition*. 2021. arXiv: [2103.06495 \[cs.CV\]](https://arxiv.org/abs/2103.06495). URL: <https://arxiv.org/abs/2103.06495>.
- [Liu+21] Ze Liu i dr. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. 2021. arXiv: [2103.14030 \[cs.CV\]](https://arxiv.org/abs/2103.14030). URL: <https://arxiv.org/abs/2103.14030>.
- [Du+22] Yongkun Du i dr. *SVTR: Scene Text Recognition with a Single Visual Model*. 2022. arXiv: [2205.00159 \[cs.CV\]](https://arxiv.org/abs/2205.00159). URL: <https://arxiv.org/abs/2205.00159>.
- [Zhe+23] Tianlun Zheng i dr. *CDistNet: Perceiving Multi-Domain Character Distance for Robust Text Recognition*. 2023. arXiv: [2111.11011 \[cs.CV\]](https://arxiv.org/abs/2111.11011).