

Automatsko prepoznavanje registarskih tablica vozila pomoću jedinstvenog vizuelnog modela

Student: Andrija Urošević

Mentor: dr Nemanja Ilić

Računarski fakultet,
Univerzitet Union

Beograd, oktobar 2024. godine

Predgovor

Ovaj rad predstavlja rezultat mog istraživanja i rada na temu automatskog prepoznavanja teksta sa tablica vozila. Pre nego što krenem u detalje vezane za rad, želim da izrazim svoju zahvalnost svima koji su mi pomogli da dođem do ovog trenutka.

Pre svega, zahvaljujem se svom mentoru, prof. dr Nemanji Iliću, na sugestijama i stručnoj podršci tokom procesa pisanja rada.

Posebnu zahvalnost upućujem svojoj porodici, prijaateljima i verenici, čija podrška i razumevanje su omogućili da se fokusiram na istraživanje, implementaciju i pisanje rada. Bez njihove svakodnevne podrške, ovaj rad ne bi bio moguć.

u Beogradu, oktobra 2024. godine
Andrija Urošević

Sadržaj

Predgovor	i
Sažetak	1
Rečnik stručnih pojmova	2
1 Uvod	3
2 Primene	4
2.1 Automatska naplata parkiranja	4
2.2 Automatska naplata putarine	5
2.3 Automatsko praćenje i lociranje ukradenih vozila	5
2.4 Automatska kontrola saobraćajnih prekršaja	5
2.5 Automatska kontrola pristupa	6
2.6 Automatska kontrola saobraćaja	7
3 Prikupljanje i rad sa podacima	8
3.1 Uvod u prikupljanje podataka	8
3.2 Specifičnosti prikupljanja podataka u kontekstu prepoznavanja teksta sa registarskih tablica	8
3.2.1 Varijabilnost registarskih tablica	8
3.2.2 Promenljivi uslovi snimanja	9
3.3 Upoznavanje sa podacima	10
3.4 Razvrstavanje i čišćenje podataka	10
3.5 Anotacija podataka	11
3.6 Kreiranje sintetičkog skupa podataka	11
3.6.1 Prednosti sintetičkih podataka	11
3.6.2 Generisanje pozadina tablica	12
3.6.3 Generisanje teksta na tablicama	12
3.6.4 Integracija sintetičkih podataka sa realnim podacima .	13
3.7 Augmentacija podataka	13
3.8 Podela podataka	13
3.9 Opis korišćenog skupa podataka	14
4 Prepoznavanje teksta	15
4.1 Uvod u prepoznavanje teksta	15
4.2 Iсторијски поглед на процесирање текста	16
4.3 Архитектуре модела за процесирање текста на сликама	17
4.4 Коришћење јединственог визуелног модела за процесирање текста на сцени	19

4.4.1	Arhiterktura	19
4.4.2	Progresivno preklapajuće ugrađivanje isečaka	20
4.4.3	Blok mešanja	21
4.4.4	Spajanje	23
4.4.5	Kombinovanje i predikcija	23
4.4.6	Analiza vizualizacije	24
5	Implementacija	25
5.1	Dodatne komponente sistema	25
5.1.1	Detektor tablica	25
5.1.2	Detektor teksta	26
5.2	Metodologije i tehnologije korišćene u razvoju servisa	26
5.2.1	PaddlePaddle	26
5.2.2	Upravljanje pristupom servisu	26
5.2.3	Distribuiranje i skaliranje	27
6	Rezultati	28
6.1	Obučavanje modela prepoznavanja teksta	28
6.1.1	Obučavanje koristeći samo realne podatke	29
6.1.2	Obučavanje koristeći samo sintetičke podatke	30
6.1.3	Obučavanje koristeći realne i sintetičke podatke	32
6.1.4	Uporedna analiza tačnosti modela na sva tri testna seta	34
6.1.5	Pregled tačnosti startnog modela na sva tri testna seta	35
7	Buduća poboljšanja	37
7.1	Dodatna diverzifikacija skupa podataka	37
7.2	Prepoznavanje teksta na više tablica sa iste slike	38
7.3	Ubrzanje rada sistema	38
8	Zaključak	39
Literatura		41

Sažetak

Automatsko prepoznavanje teksta sa registarskih tablica vozila 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 savremene metode i tehnike za automatsko prepoznavanje teksta sa tablica vozila, sa ciljem razvoja sistema koji može precizno identifikovati registarske tablice u realnom vremenu. Eksperimentalni rezultati prikazuju veoma dobre performanse sistema u stvarnim uslovima i ukazuju na mogućnosti za primenu u različitim oblastima, uključujući nadzor saobraćaja, bezbednosne provere i identifikaciju vozila.

Rečnik stručnih pojmoveva

Backbone	<i>Osnovna mreža</i>
Base model	<i>Inicijalni model</i>
Batch	<i>Grupa podataka</i>
Dropout	<i>Nasumično isključivanje neurona</i>
Embedding	<i>Ugrađivanje</i>
Framework	<i>Platforma</i>
Fully connected layer	<i>Potpuno povezani sloj</i>
Inference	<i>Zaključivanje</i>
Labeling	<i>Označavanje</i>
Learning rate	<i>Stopa učenja</i>
Overfitting	<i>Prekomerna prilagodenost</i>
Self-attention	<i>Samopaznja</i>
Shortcut connection	<i>Preskočna veza</i>
Training	<i>Obučavanje</i>

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 vozila obuhvata nekoliko ključnih koraka koji se odvijaju u procesu od prikupljanja podataka do konačne integracije sistema u softver za prepoznavanje tablica vozila.

Prikupljanje raznovrsnog skupa slika registarskih tablica vozila ključno je za uspešno obučavanje 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 obučavanje 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 tehnike 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 obučavanje modela. Ovaj proces može biti ručan ili se može koristiti alat za automatsko označavanje podataka. Nakon pripreme podataka, sledi faza obučavanja 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 vozila. Ovaj softver obično obuhvata module za detekciju tablica, detekciju teksta na tablicama, formatiranje izlaza i druge funkcionalnosti.

Kako bi se omogućila portabilnost i lakša distribucija sistema za prepoznavanje tablica, koristi se Docker kontejner. Docker omogućava pakovanje softverskih aplikacija i njihovo pokretanje u izolovanim okruženjima. Još jedna od bitnih komponenti 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 Primene

Sistem za automatsko prepoznavanje teksta sa tablica vozila (eng. Automatic License Plate Recognition - ALPR) 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 ALPR sistema, moguće je unaprediti efikasnost naplate putarina, automatski pratiti i locirati ukradena vozila, kao i optimizovati procese parkiranja i upravljanja kolonom vozila.

Pored toga, ALPR može biti od značaja 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 unapređenje infrastrukture i usluga u urbanim sredinama. U ovom radu biće razmotrene različite primene ALPR sistema, kao i potencijalni izazovi i rešenja u njegovoj implementaciji.

2.1 Automatska naplata parkiranja

ALPR 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 ALPR 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, ALPR 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.

2.2 Automatska naplata putarina

ALPR može značajno unaprediti sistem automatske naplate putarina. ALPR 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 ALPR 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, ALPR može pomoći u identifikaciji vozila koja su prijavljena kao ukradena, čime se dodatno povećava bezbednost na putevima.

2.3 Automatsko praćenje i lociranje ukradenih vozila

ALPR 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 ALPR kamere prepoznavaju 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, ALPR tehnologija takođe može pomoći u smanjenju stope kriminala, jer deluje kao odvraćajući faktor za potencijalne počinioce.

2.4 Automatska kontrola saobraćajnih prekršaja

ALPR 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 ALPR 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, ALPR može pomoći u prikupljanju

podataka o saobraćajnim tokovima i obrascima ponašanja vozača, što omogućava vlastima da bolje planiraju saobraćajnu infrastrukturu.

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 ALPR tehnologije, moguće je stvoriti efikasniji sistem kontrole saobraćaja koji smanjuje broj prekršaja i podstiče vozače da se pridržavaju saobraćajnih pravila. Pored svih navedenih prednosti, automatsko prepoznavanje registarskih tablica doprinosi unapređenju bezbednosti saobraćaja i smanjenju rizika od saobraćajnih nesreća.

2.5 Automatska kontrola pristupa

ALPR 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 ALPR 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 ALPR 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, ALPR 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 nose nekoliko daljinskih upravljača, što dodatno

olakšava svakodnevno korišćenje i povećava praktičnost.

2.6 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 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.

3.1 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 obučavanje, 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.

3.2 Specifičnosti prikupljanja podataka u kontekstu prepoznavanja teksta sa registarskih tablica

3.2.1 Varijabilnost registarskih tablica

Raznolikost dizajna Registarske 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, registarske tablice mogu postati oštećene, izbledele ili prljave, što može otežati prepoznavanje teksta. Prikupljanje slika sa različitim stepenima oštećenja tablica je bitno za obučavanje modela koji može da se nosi sa takvim izazovima.

3.2.2 Promenljivi uslovi snimanja

Različiti uslovi osvetljenja Snimanje registarskih 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. Primeri prikupljenih tablica snimljenih u različitim uslovima su prikazani na Slici 1.



Slika 1: 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.

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, farovi vozila mogu uzrokovati probleme sa kontrastom i zaslepljivanjem, što dodatno otežava prepoznavanje tablica.

Različiti uglovi snimanja U zavisnosti od položaja kamere, registarske 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 registarskih 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.

3.3 Upoznavanje sa podacima

Za razvoj sistema za automatsko prepoznavanje teksta sa registarskih tablica, postoji nekoliko javno dostupnih 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 obučavanje i evaluaciju modela za prepoznavanje teksta. Neki od najpoznatijih javno dostupnih skupova podataka uključuju:

UFPR-ALPR Baza podataka koja sadrži 4.500 slika registarskih tablica iz različitih zemalja. Ovaj skup podataka 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 obučavanje 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 je fokus ovog rada prepoznavanje teksta sa registarskih tablica vozila registrovanih u Republici Srbiji, korišćen je specifičan lokalizovan skup od približno 17.000 slika, većinom tablica vozila snimljenih ispred rampi, koji je autoru rada bio dostupan. Ovaj pristup omogućava obučavanje 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.

3.4 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 registrarske tablice, moraju biti uklonjenje iz skupa podataka. Ova faza obezbeđuje da model bude obučavan na kvalitetnim i relevantnim podacima, što utiče na njegovu tačnost i robusnost.

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

3.5 Anotacija podataka

Anotacija predstavlja označavanje delova slike koji sadrže ključne regije, u ovom slučaju registrarske tablice i tekst koji je isписан na tablicama. Anotacija podataka je proces u kojem se ručno ili poluautomatski označavaju tačne lokacije i sadržaj registrarskih 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 obučavanja modela. Greške u anotaciji, poput netačnih ili nepotpunih oznaka, mogu dovesti do smanjenja tačnosti modela i takav tip greške se kasnije vrlo teško pronalazi.

3.6 Kreiranje sintetičkog skupa podataka

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

3.6.1 Prednosti sintetičkih podataka

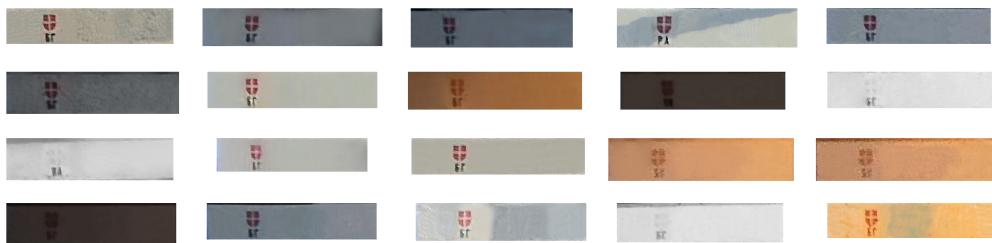
Sintetički skupovi podataka omogućavaju stvaranje velike količine podataka bez potrebe za obimnim prikupljanjem i anotacijom realnih slika, što može biti vremenski i finansijski zahtevno. Sintetički podaci omogućavaju preciznu kontrolu nad svim aspektima slike, uključujući osvetljenje, ugao snimanja, pozadinu, i stil registrarskih 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 skupovi podataka 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 uslovima, ali su ipak važne za sveobuhvatnu tačnost sistema.

3.6.2 Generisanje pozadina tablica

Za generisanje pozadina tablica, korišćene su slike stvarnih registarskih tablica kako bi se postigao visok nivo autentičnosti. Prvo je odabранo nekoliko slika pravih tablica, snimljenih u različitim uslovima, kako bi se obezbedila raznovrsnost u teksturi, boji i detaljima. Zatim su, korišćenjem [GIMP](#) editora, obrisani svi karaktere sa tih slika, na način da su osnovna struktura i karakteristike tablica zadržane. Na ovaj način su dobijene čiste pozadine koje izgledaju prirodno i verodostojno (Slika 2).



Slika 2: Primeri generisanih pozadina tablica.

3.6.3 Generisanje teksta na tablicama

Nakon generisanja pozadina tablica, korišćen je repozitorijum [TextRecognitionDataGenerator](#), je modifikovan tako da je moguće da se na tačno određenim mestima i pod odgovarajućim uglovima dodaje tekst na prethodno generisane pozadine tablica. Prethodno je generisan 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 je da se kreira sintetički skup tablica koje podsećaju na stvarne srpske tablice (Slika 3).

LU 1225 VL	NP 978 GE	PE 882 EH	PR 700 TQ	SU 1088 NW
TS 397 SS	UB 501 FP	VA 883 TS	VP 41416 PD	VR 173 UA
ZR 44364 JJ	AC 182 CQ	AR 726 GI	BG 0028 UM	JA 441 SZ

Slika 3: Primeri sintetičkog skupa podataka.

3.6.4 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 obučavanje. 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.

3.7 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 od 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.

3.8 Podela podataka

Prilikom obučavanja modela dostupni skup podataka je potrebno podeleti u tri distinkтивna podskupa: trening skup, validacioni skup i test skup. Ovakva podela podataka, gde se jedan podatak, u ovom slučaju slika, nalazi isključivo samo u jednom od tri skupa je fundamentalna za razvoj, evaluaciju i verifikaciju modela.

Trening skup čini najveći deo inicijalnog skupa podataka (obično od 60-80%), koristi se za obučavanje modela. Koristeći podatke iz trening skupa model uči obrasce, relacije i karakteristike koje su svojstvene problemu koji se rešava.

Validacioni skup je manji i obično oko 10-20% inicijalnog skupa podataka. Služi za finu kalibraciju modela i sprečavanje prekomerne prilagođenosti modela trening skupu tokom obučavanja. Koristi se i za procenu performansi modela tokom faze obuke i za podešavanje hiperparametara.

Test skup je finalni podskup koji čini oko 10-20% inicijalnih podataka i koristi se za nezavisnu evaluaciju konačnog modela. Ovaj skup ostaje nekoristišen do završetka obučavanja i kalibracije, čime se obezbeđuje objektivna procena generalizacijske sposobnosti modela na novim, prethodno neviđenim podacima.

Ovakva metodologija podele podataka omogućava razvoj robusnih modela koji imaju veću pouzdanost i praktičnu primenljivost u realnim scenarijima.

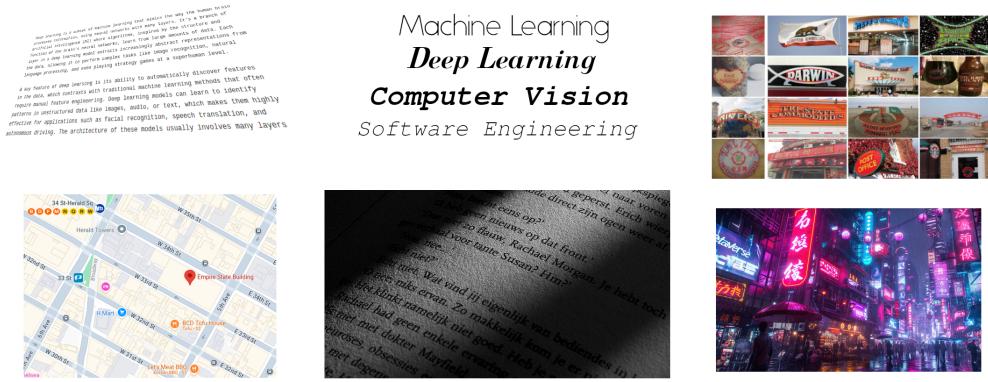
3.9 Opis korišćenog skupa podataka

Skup realnih podataka koji je korišćen za proces obučavanja modela sadržao je 5.838 slika stvarnih srpskih tablica, uglavnom registrovanih u Beogradu. Sve slike su pažljivo označene i višestruko proverene kako bi se izbeglo propuštanje nepravilno označenih podataka, što bi dovelo do loših rezultata prilikom obučavanja modela. Uvezši u obzir mali broj dostupnih slika za obučavanje, augmentacijom i pravljenjem sintetičkih podataka povećan je skup podataka sa 5.838 na 124.838 slika. Prilikom augmentacije i pravljenja sintetičkih podataka inicijalni skup je obogaćen sa slikama različitih scenarija vremenskih i dnevnih uslova, kao i različitim uglovima slikanja kamere. Pored toga što doprinosi varijabilnosti uslova osvetljenja i kvantitetu podataka, dodavanje sintetičkog skupa unosi i novi kontekst teksta na tablicama, gde su u sintetičkom skupu napravljeni primeri tablica koje su mogле biti registrovane u [ostalim dozvoljenim](#) [reg] gradovima u Srbiji pored Beograda. Dobra strana rada sa sintetičkim podacima jeste i to što ne zahtevaju ručno označavanje podataka.

4 Prepoznavanje teksta

4.1 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. Primeri varijacija teksta su prikazani na Slici 4.



Slika 4: Tekst na sceni sa različitim fontovima, pozadinama, osvetljenjem i slično.

U proteklim godinama uloženi su brojni napori 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 [Li+22].

4.2 Istorijski pregled prepoznavanja teksta

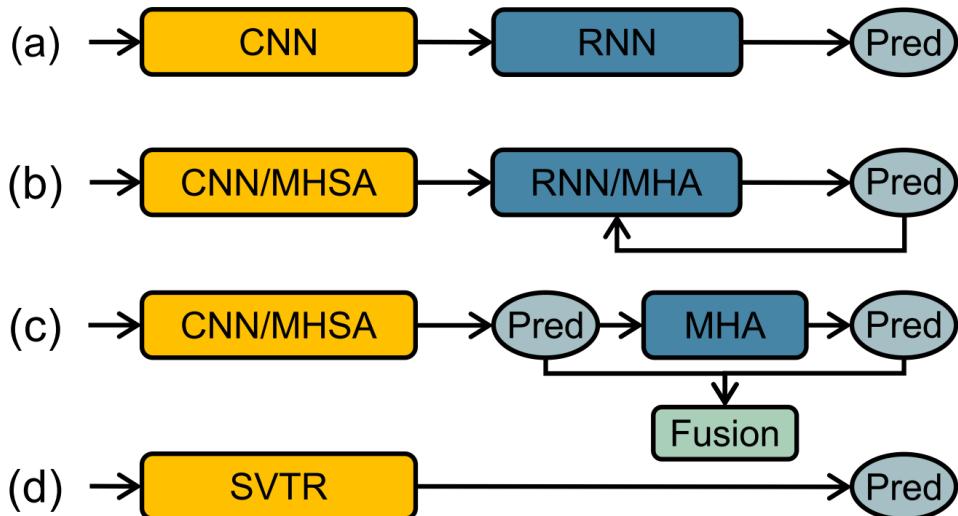
Prvi primeri i prva faza tehnologije optičkog prepoznavanja karaktera (eng. Optical Character Recognition - OCR) pojavili su se sredinom 20. veka, pretežno tokom 1950-ih i 1960-ih godina. Ovo doba obeležilo je razvoj ranih sistema OCR-a, koji su koristili osnovne tehnike prepoznavanja obrazaca kako bi 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 [Tri24].

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. Napredci u drugoj fazi, 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 [Eve23].

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 (eng. Convolutional Neural Network - CNN) i rekurentnih neuronskih mreža (eng. Recurrent Neural Network - 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 [Wan+21].

4.3 Arhitekture modela za prepoznavanje teksta na slici

Postoje različiti pristupi rešavanja problema dizajna modela za prepoznavanje teksta na slici. Uobičajeno se modeli sastoje iz više glavnih strukturalnih delova, uključujući module za ekstrakciju karakteristika, sekvensijalne modele za transkripciju, kao i mehanizme za dekodiranje izlaza. Na Slici 5 prikazane su različite strukture modela za prepoznavanje teksta sa slike, koje ilustruju raznovrsne arhitekture i tehnike korišćene u ovom polju.



Slika 5: Arhitekture modela za prepoznavanje teksta sa scene. (a) Modeli zasnovani na CNN-RNN. (b) Modeli kodiranja-dekodiranja. (c) Vizuelno-jezički modeli. (d) SVTR, koji prepoznaje tekst scene sa jedinstvenim vizuelnim modelom i odlikuje se efikasnošću, tačnošću i višejezičnom svestranošću. Slika je preuzeta iz [Du+22].

Modeli zasnovani na CNN-RNN [SBY15] prvo koriste CNN za ekstrakciju karakteristika. Karakteristike se zatim preoblikuju u sekvencu koju biderakciona mreža dugoročne i kratkotrajne memorije (eng. Bidirectional Long Short-Term Memory - BiLSTM) modeluje uz pomoć vremenski konekcionističkog gubitka (eng. Connectionist Temporal Classification - CTC) kako bi generisao predikciju, prikazanu na Slici 5(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 postali popularni [SCX19; Li+19; Zhe+23]. Ove metode transformišu prepoznavanje u iterativni proces dekodiranja, prikazan na Slici 5(b). Kao rezultat, postignuta je poboljšana tačnost jer je uzeta u obzir kontekstualna informacija. Međutim, brzina prilikom zaključivanja je spora zbog transkripcije karakter po karakter. Ovaj postupak je dodatno proširen na platformu zasnovanu na viziji i jeziku [Yu+20; Fan+21a], gde je jezičko znanje uključeno i sprovedena je paralelna predikcija, kao što je prikazano na Slici 5(c). Ipak, ovaj postupak često zahteva model sa velikim brojem parametara 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 konvolucionoj rekurentnoj neuronskoj mreži (eng. Convolutional Recurrent Neural Network - CRNN) CRNN-RNN [Hu+20b] koristi mehanizam pažnje i grafovsku neuronsku mrežu za agregiranje sekvensijalnih karakteristika koje odgovaraju istom karakteru. Pri izvršavanju, deo za modelovanje zasnovan na mehanizmu pažnje je odbačen kako bi se uskladili tačnost i brzina.

Nedavni uspeh transformera za obradu slike [Dos+21; Liu+21], inspirisao je nastanak jedinstvenog vizuelnog modela za prepoznavanje teksta na sceni (eng. Single Visual Transformer Recognition - SVTR) [Du+22]. SVTR najpre razlaže tekst slike na male 2D isečke koji se nazivaju komponente karaktera, od kojih svaka komponenta može sadržati samo deo karaktera. Tokenizacija slike po isećcima praćena mehanizmom samopažnje se primenjuje da bi se uhvatile indicije prepoznavanja teksta među komponentama karaktera. Za ovu svrhu je razvijena prilagođena arhitektura za tekst, čija osnovna mreža sadrži progresivno smanjujuću prostornu dimenziju visine mape karakteristika u tri faze i uključuje operacije mešanja, spajanja i/ili kombinovanja. Osmisljeni su lokalni i globalni blokovi mešanja koji se rekurzivno primenjuju u svakoj fazi, zajedno sa operacijom spajanja ili kombinovanja. Time se dobijaju afiniteti na nivou lokalnih komponenti koje predstavljaju karakteristike slične potezima nastalim prilikom pisanja karaktera olovkom po papiru i dugoročne zavisnosti između različitih karaktera. Dakle, osnovna mreža ekstrahuje karakteristike komponenti na različitim rastojanjima i na više skala, formirajući višeslojnu percepцију karakteristika karaktera. Kao rezultat, prepoznavanje teksta se postiže jednostavnom linearnom predikcijom. U celom procesu koristi se samo jedan vizuelni model, kao što je prikazano na Slici 5(d).

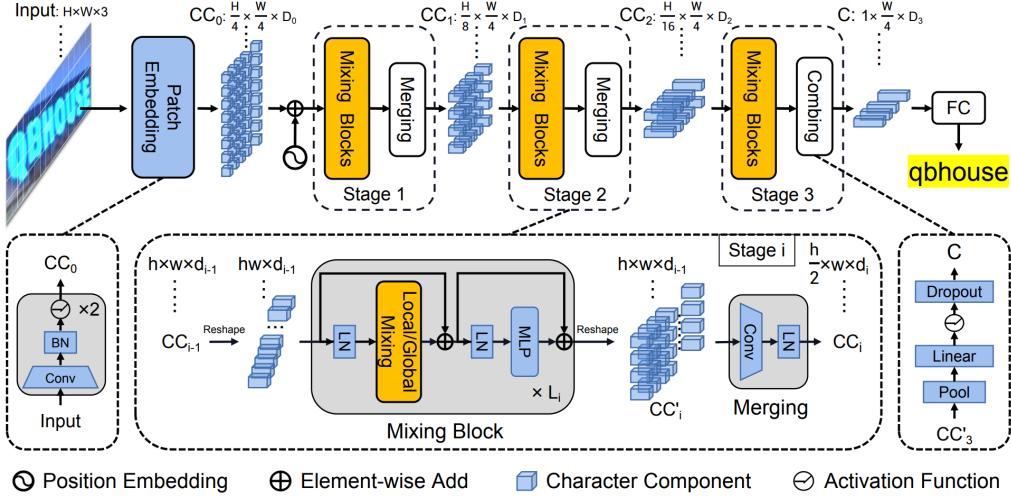
4.4 Korišćenje jedinstvenog vizuelnog modela za prepoznavanje teksta na sceni

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 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.

4.4.1 Arhitektura

Pregled SVTR modela je prikazan na Slici 6 i predstavlja trofaznu mrežu sa progresivno smanjujućom visinom, namenjenu za prepoznavanje teksta. Slika koja sadži tekst i veličine je $H \times W \times 3$, 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. Osmisljeni su lokalni i globalni blokovi mešanja za ekstrakciju lokalnih obrazaca nalik potezima i hvatanje međukomponentnih zavisnosti. Pomoću osnovne mreže se dobijaju komponentne karakteristike kao i zavisnosti na različitim udaljenostima i na više skala, predstavljene kao C veličine $1 \times \frac{W}{4} \times D_3$, koje percipiraju 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 predvideo, a koji nisu stvarno prisutni na originalnoj slici. Rezultat ovog procesa je konačan niz prepoznatih znakova.

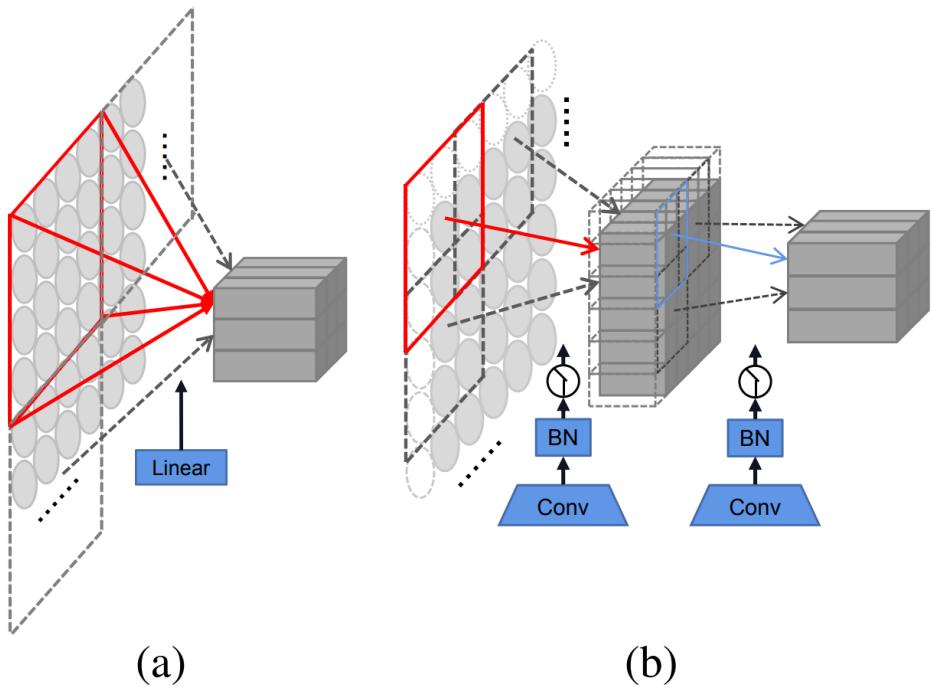


Slika 6: Arhitektura SVTR modela: Mreža koja kroz tri faze progresivno smanjuje visinu mape karakteristika. U svakoj fazi se izvodi niz blokova za mešanje, nakon čega sledi operacija spajanja ili kombinovanja. Na kraju se prepoznavanje vrši linearnim predviđanjem. Slika je preuzeta iz [Du+22].

4.4.2 Progresivno preklapajuće ugradivanje isečaka

Prvi korak u obradi slike teksta je njeno razlaganje na manje delove koje nazivamo isećcima. Dobijanje karakterističnih isečaka koji predstavljaju

komponente znakova znači prelazak iz $X \in \mathbb{R}^{H \times W \times 3}$ 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 7(a)) i korišćenje 7×7 konvolucionog filtera sa korakom 4. Autori SVTR arhitekture su izabrali alternativni metod. Oni koriste dva manja konvolucionala filtera 3×3 jedan za drugim sa korakom 2, kao što je prikazano na Slici 7(b). Takođe koriste tehniku zvanu normalizacija serije da bi održali brojove pod kontrolom. Ovaj novi metod zahteva nešto više računarske snage, ali je bolji u kombinovanju karakteristika iz slike.



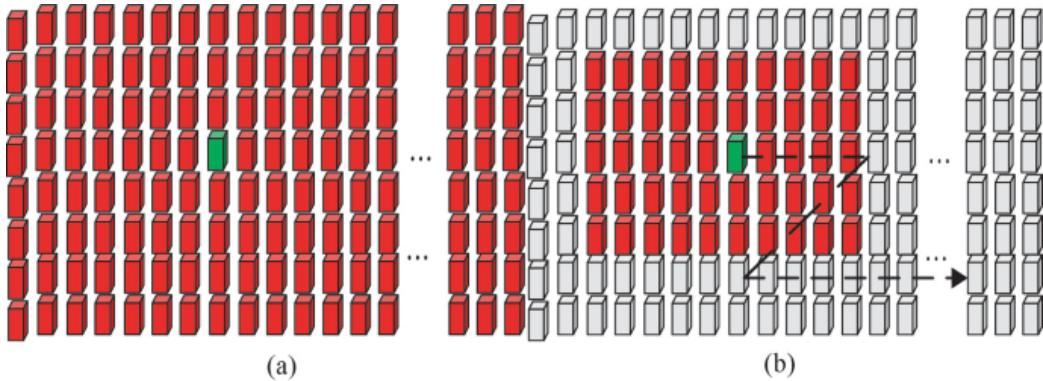
Slika 7: (a) Linearna projekcija u ViT [Dos+21]. (b) SVTR progresivno preklapajuće ugrađivanje isečaka. Slika je preuzeta iz [Du+22].

4.4.3 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 ekstrakciju 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, pogotovo za nepravilan tekst — što nije optimalno za opisivanje karaktera. Nedavni napredak vizuelnih transformera uvodi 2D reprezentaciju karakteristika, ali njeno korišćenje u kontekstu prepoznavanja teksta je još uvek u fazi istraživanja. Autori SVTR arhitekture sugerisu da su dve vrste karakteristika važne za prepoznavanje teksta. Prva su lokalni obrasci, 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 širu sliku o tome kako su karakteri međusobno povezani.

Globalno mešanje Kao što se vidi na Slici 8(a), globalno mešanje procenjuje zavisnost među svim komponentama karaktera. S obzirom da su tekstualni i netekstualni 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 netekstualnih 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, sekvencijalno se primenjuju normalizacija sloja i MLP za fuziju karakteristika, pa se zajedno sa preskočnim vezama, formira blok globalnog mešanja.



Slika 8: Ilustracija (a) globalnog mešanja (b) lokalnog mešanja; Slika je preuzeta iz [Du+22].

Lokalno mešanje Kao što se može videti na Slici 8(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 koja je vitalna za identifikaciju karaktera i nalik je potezu prilikom pisanja karaktera. Za razliku od globalnog mešanja, lokalno mešanje razmatra okolinu za svaku komponentu. Slično konvoluciji, mešanje se odvija korišćenjem pristupa klizećeg prozora. Veličina prozora je empirijski postavljena na 7×11 .

U poređenju sa globalnim mešanjem, lokalno implementira mehanizam samopasnje za detekciju lokalnih obrazaca. Kao što je prethodno pomenuto, dva bloka mešanja imaju za cilj izvlačenje različitih karakteristika koje su komplementarne. U SVTR arhitekturi, blokovi se rekurentno primenjuju više puta u svakoj fazi za sveobuhvatnu ekstrakciju karakteristika.

4.4.4 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 arhitekture 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, tim redom. Zatim se primenjuje konvolucija veličine 3×3 sa korakom 2 u dimenziji visine i korakom 1 u dimenziji širine, praćena 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 hiperhijsku 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. Smanjivanje dimenzije visine smanjuje šansu za kodiranje susednih karaktera u istu komponentu kroz faze. Takođe, povećava se dimenzija kanala d_i kako bi se nadoknadio gubitak informacija.

4.4.5 Kombinovanje i predikcija

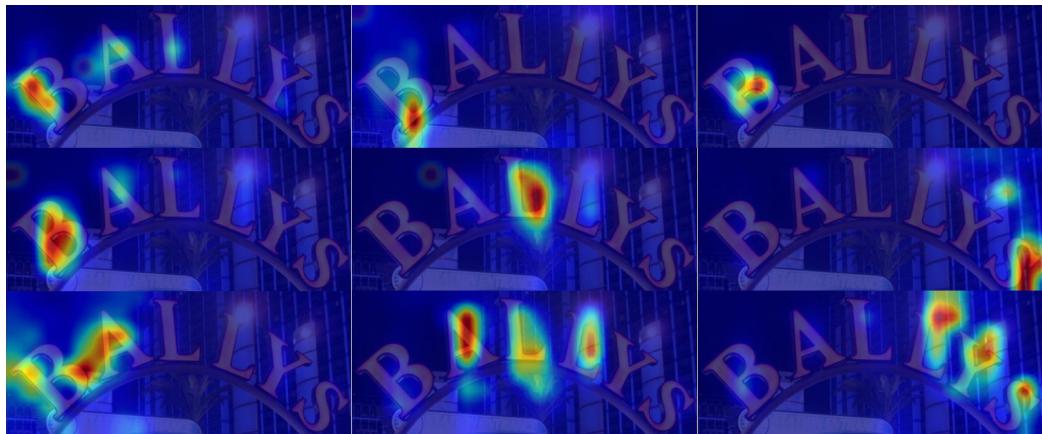
U poslednjoj fazi, operacija spajanja se zamjenjuje operacijom kombinovanja. Prvo se dimenzija visine svede na 1, a zatim se primenjuje potpuno povezani sloj, nelinearna aktivacija i nasumično isključivanje neurona. Na taj način, komponente karaktera se dodatno kompresuju u sekvencu karakteristika, gde 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 korišćenjem 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.

4.4.6 Analiza vizualizacije

Svaka mapa se može objasniti kao da ima različitu ulogu u celokupnom prepoznavanju. Ilustracija devet primera mape je prikazana na Slici 9. Prvi red prikazuje tri mape koje se fokusiraju na deo karaktera „B”, sa naglaskom na njegovu levu stranu, donji deo i srednji deo, tim 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 arhitektura modela hvata tragove na nivou dela karaktera, celog karaktera i više karaktera, u skladu sa tvrdnjom da SVTR percipira višeslojne karakteristike komponenti karaktera, potvrđujući efikasnost arhitekture SVTR modela.



Slika 9: Vizualizacija SVTR mapi pažnje. Slika je preuzeta iz [Du+22].

5 Implementacija

U cilju dobijanja sistema visokih performansi za prepoznavanje teksta sa tablica vozila, razvijen je servis za automatizovano prepoznavanje teksta sa tablica ulaznih slika korišćenjem modela za detekciju tablica i teksta sa scene. Ovaj pristup omogućava preciznije i pouzdanoje rezultate, jer se oslanja na kombinaciju različitih modela koji su specijalizovani za različite aspekte prepoznavanja regija od važnosti.

5.1 Dodatne komponente sistema

Ono što se pokazalo kao zanimljiv izazov u ovom procesu je to što nije bilo dovoljno koristiti samo model za detekciju tablica. Mnoge registarske tablice imaju dodatne elemente, poput reklama ili natpisa na okvirima, koji mogu ometati proces prepoznavanja teksta. Da bi se postigla visoka tačnost u prepoznavanju teksta sa registarskih oznaka, bilo je neophodno razviti metodologiju koja će eliminisati takve neželjene informacije pre nego što se predaje na prepoznavanje teksta sa registarskih oznaka.

S druge strane, korišćenje modela isključivo za detekciju teksta nije prihvatljivo kao robusno rešenje. Na slikama na kojima se nalaze registarske tablice često se može naći i drugi tekst koji ne predstavlja tablicu, kao što su natpisi, reklame, ime brenda vozila ili drugi elementi iz okruženja. Ako bi se oslonio isključivo na model za detekciju teksta, postojala bi dosta velika verovatnoća da će model pogrešno identifikovati ili obraditi neželjene informacije, što bi moglo dovesti do netačnih rezultata.

Zbog toga je servis dizajniran tako da prvo detektuje tablice na ulaznim slikama, a zatim koristi model za prepoznavanje teksta isključivo unutar detektovanih oblasti tablica. Na ovaj način se može efektivno izbaciti tekst koji je isписан po okvirima i obično značajno manji veličinom u odnosu na glavni tekst koji se nalazi duž tablice.

5.1.1 Detektor tablica

Model koji je korišćen za detekciju tablica je [YOLOS vizuelni transformer](#) [Fan+21b] inicijalno obučavan na ImageNet skupu podataka i fino podešen kroz 200 epoha na skupu od [5200 slika tablica](#). Ovaj model je razvijen kako bi poboljšao efikasnost i preciznost detekcije objekata koristeći transformere umesto tradicionalnih konvolucionih mreža.

YOLOS se oslanja na arhitekturu vizuelnog transformera koja omogućava modelu da obraduje slike kao nizove vizuelnih tokova. Sa tim pristupom

model može da uoči dugoročne zavisnosti i kompleksne obrasce na slikama, što poboljšava tačnost detekcije objekata.

Ono što izdvaja YOLOS je jednostavnost dizajna modela, koja omogućava postizanje visokih performansi u detekciji objekata uz relativno manji broj slojeva i parametara u poređenju sa drugim transformer modelima.

5.1.2 Detektor teksta

Za detekciju teksta na sceni je korišćen [CRAFT](#) model koji je osmišljen da poboljša prepoznavanje i lokalizaciju teksta u složenim slikama koristeći dve glavne komponente: svest o oblasti i klasifikaciju karaktera.

Deo modela zadužen za svest o oblasti se fokusira na identifikaciju i precizno lokalizovanje područja na slici gde se mogu nalaziti karakteri, omogućavajući modelu da prepozna različite oblike i orientacije teksta. Deo vezan za klasifikaciju karaktera koristi karakteristične informacije za klasifikaciju prepoznatih regija i precizno identificuje pojedinačne karaktere [Bae+19].

5.2 Metodologije i tehnologije korišćene u razvoju servisa

5.2.1 PaddlePaddle

[PaddlePaddle](#) je *open-source* platforma za duboko učenje koji je razvijen od strane Baidu-a. Jedan je od najpopularnijih projekata na GitHub-u koji se bave mašinskim učenjem, a [PaddleOCR](#) deo je trenutno i najpopularniji repozitorijum na temu optičkog prepoznavanja karaktera.

PaddleOCR projekat pruža veliki broj implementiranih modela iz aktuelnih radova koji se bave prepoznavanjem teksta sa slikama. Ovi modeli pokrivaju širok spektar OCR zadataka, uključujući prepoznavanje teksta na različitim jezicima, analizu strukture dokumenata, prepoznavanje formula i tabela, kao i druge specijalizovane primene.

Pored implementacije najnaprednijih OCR modela, PaddlePaddle pruža i alate za anotaciju podataka, generisanje sintetičkih slika za obučavanje, kao i optimizaciju i ubrzavanje modela za efikasnu primenu na različitim platformama - od servera do mobilnih uređaja i ugrađenih sistema.

Kao početnu tačku za implementaciju prepoznavanja teksta sa tablica vozila je korišćena njihova implementacija SVTR modela.

5.2.2 Upravljanje pristupom servisu

Kako bi prepoznavanje tablica moglo da se koristi kao servisna aplikacija, izabrani su [Uvicorn](#) i [FastAPI](#) za izradu web servisa i otvorena ruta

za pristup servisu. FastAPI je moderna i brza web platforma za pravljenje API-ja u Python programskom jeziku i omogućava lako definisanje ruta i automatsku validaciju podataka. Njegova sposobnost da generiše interaktivnu dokumentaciju putem Swagger-a čini ga izuzetno korisnim za razvoj i testiranje API-ja.

Uvicorn je performantni [ASGI](#) server koji podržava asinhrono rukovanje zahtevima. Njegova integracija sa FastAPI platformom omogućava efikasno upravljanje višestrukim pozivima i događajima, što je bitno za aplikacije koje zahtevaju brze i responzivne interakcije sa korisnicima.

Za pristup servisu za prepoznavanje teksta sa tablica je otvorena ruta pod nazivom „/do_lpr”, koja prihvata sliku kao ulaz. Izlaz iz ove rute su isečena slika tablice sa inicijalne slike i pročitani tekst sa tablice. Na ovaj način korisnici treba samo da pošalju sliku na kojoj se nalazi registarska tablica i za tu sliku će dobiti relevantne informacije.

Korišćenje FastAPI web platforme i Uvicorn servera je olakšalo i pravljenje dokumentacije, jer pruža automatski generisanu Swagger dokumentaciju na osnovu ispisanog koda.

5.2.3 Distribuiranje i skaliranje

Servis koji je napravljen je dalje upakovani distribuiran pomoću Docker-a. [Docker](#) je platforma za kontejnerizaciju aplikacija koja omogućava lako kreiranje, distribuiranje i pokretanje aplikacija u izolovanom okruženju, poznatom kao kontejner.

Jedan od glavnih razloga za korišćenje Docker-a je obezbeđivanje konzistentnog okruženja za rad aplikacije. Docker kontejneri garantuju da će aplikacija raditi isto na različitim mašinama, čime se eliminišu problemi sa zavisnostima i razlikama u okruženju. Zahvaljujući Docker-u, prilikom razvoja se osigurava da aplikacija funkcioniše bez obzira na to gde će biti pokretana, što značajno olakšava proces implementacije.

Osim toga, Docker omogućava lako skaliranje i prenosivost aplikacija. Kontejneri su jednostavni za prenos između različitih platformi, što olakšava distribuciju i implementaciju servisa. Takođe, Docker pruža alate za efikasno upravljanje životnim ciklusom aplikacije, uključujući kreiranje, pokretanje, zaustavljanje i brisanje kontejnera.

Docker, kao rešenje, omogućava isporuku robusnog, skalabilnog i lako upotrebljivog rešenja.

6 Rezultati

6.1 Obučavanje modela prepoznavanja teksta

Uvezši u obzir pristup malom broju podataka čak i nakon odrađene augmentacije i dodavanja sintetičkog skupa, tokom obučavanja neki od parametara su eksperimentalnim pristupom podešeni na sledeće vrednosti:

- Broj epoha: 50; iako je broj epoha postavljen na 50, zbog malog broja dostupnih podataka model će najčešće konvergirati u trenutku od druge do pete epohe.
- Korak za testiranje nad validacionim skupom: 200; na svakih 200 završenih grupa podataka prilikom obučavanja pokreće se evaluacija nad validacionim skupom podataka.
- Stopa učenja: korišćena je kosinusna stopa učenja sa početnom vrednošću postavljenom na 0.0005.
- Veličina ulazne slike: širina 320px, visina 48px i 3 kanala boja (RGB).

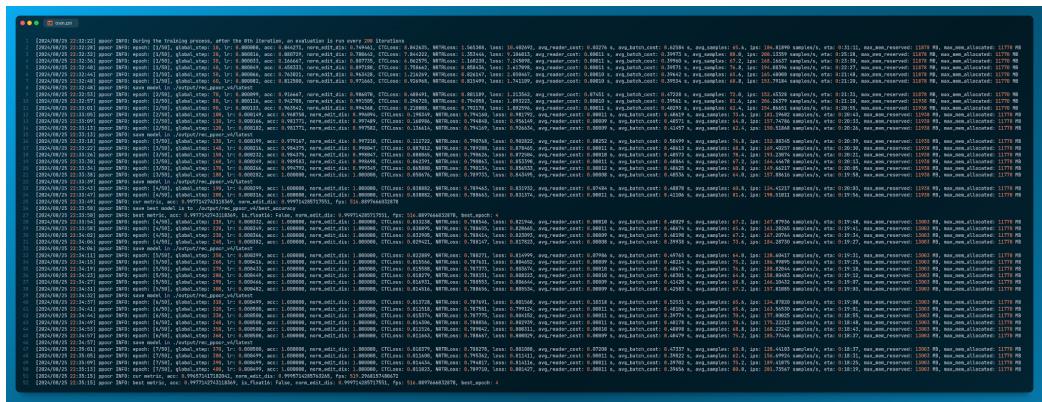
Obučavanje modela za prepoznavanje teksta je rađeno na KDE Neon Linux mašini sa sledećim specifikacijama:

- GPU: NVIDIA GeForce RTX 4080, 16GB VRAM
- CPU: 12th Gen Intel i7-12700KF
- RAM: 32GB

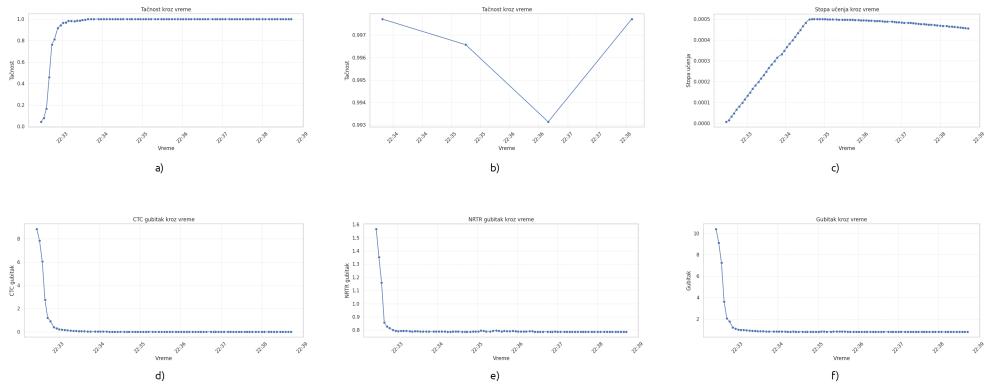
Za evaluaciju i praćenje obučavanja modela korišćeni su CTC, NRTR (Non-Recurrent Text Recognition) i sumarni gubitak te dve funkcije. [CTC gubitak](#) se koristi u primerima gde je usklađenost između ulaznih i izlaznih sekvenci nepoznata. Ulazne sekvene u ovom slučaju predstavljaju slike koje sadrže tekst, a izlazne sekvene predstavljaju niz karaktera koji je zapisan na slici. CTC omogućava modelu da generiše sekvencu verovatnoća za svaki karakter i predstavlja dobar pristup za različite dužine ulaznih i izlaznih sekvenci. Funkcija gubitka izračunava verovatnoću ispravne sekvene sabiranjem svih mogućih usklađenja, što omogućava modelu da uči iz neusklađenih podataka [Gra+06]. NRTR gubitak je dizajniran za nerekurentne modele, posebno za modele bazirane na transformer arhitekturi. Efikasan je za duge sekvene, jer ima sposobnost paralelnog procesiranja [Hu+20a].

6.1.1 Obučavanje koristeći samo realne podatke

Imajući u vidu da je za obučavanje modela bilo dostupno samo 5.838 realnih slika tablica vozila registrovanih uglavnom u Beogradu i koje su bile označene i proverene, obučavanje je trajalo samo par minuta. Model je konvergirao za nekoliko minuta i dostigao tačnost od 100% (Slika 11(a)) na trening skupu i 99.77% (Slika 11(b)) na validacionom skupu, koji je pomogao u usmeravanju obučavanja (Slika 10).



Slika 10: Proces obučavanja modela za detekciju teksta na realnim podacima.



Slika 11: Metrike u toku obučavanja modela sa realnim podacima. a) Tačnost modela na trening podacima u toku obučavanja modela. b) Tačnost modela na validacionim podacima u toku obučavanja modela. c) Promena vrednosti stope učenja u toku obučavanja modela. d) CTC gubitak u toku obučavanja modela. e) NTRR gubitak u toku obučavanja modela. f) Gubitak u toku obučavanja modela.

Dostupni skup realnih podataka je podeljen u sledeće podskupove:

- Trening skup je činilo 70% podataka, što je 4.086 slika;
- Validacioni skup je činilo 15% podataka, što je 875 slika;
- Test skup je činilo 15% podataka, što je 875 slika.

Uzevši u obzir veoma visoku tačnost na obučavanju i validacionim podacima, najbitnija provera koja validira da li je model dobro obučen ili je došlo do prekomerne prilagođenosti, jeste provera nad test skupom podataka. Uspešnost na nezavisnom testu nad skupom testnih podataka koje model nije mogao da vidi ni u jednom trenutku prilikom obučavanja iznosi 100%, što ukazuje na to da model nije imao problem sa prekomernom prilagođenosti podataka prilikom obučavanja.

Pored visoke tračnosti, model takođe ima i visok stepen sigurnosti prilikom predikcije, sa prosečnom sigurnošću od 99,91% za svaku predikciju koju je izvršio na testnom skupu podataka.

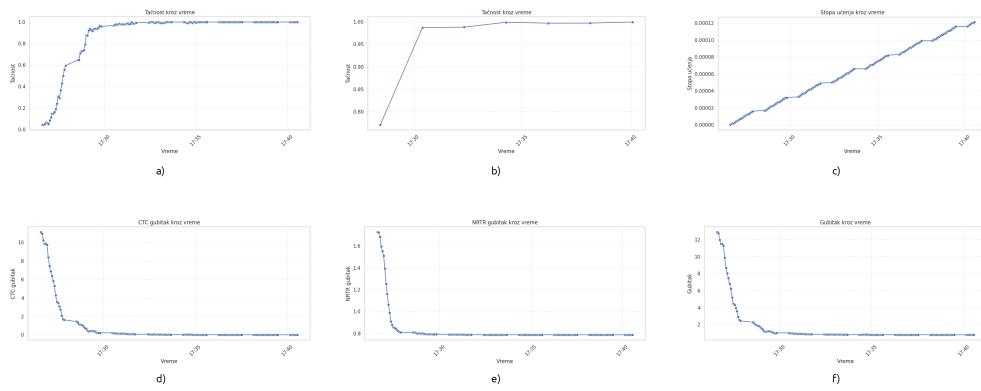
Međutim, obzirom na to da skup realnih podataka nije bio ni kontekstualno dovoljno raznovrstan, niti kvantitativno dovoljno veliki, obučavanje je nastavljeno sa sintetičkim podacima.

6.1.2 Obučavanje koristeći samo sintetičke podatke

Pre obučavanja na spojenim skupovima realnih i sintetičkih podataka, vršena je provera tačnosti modela u scenariju u kojem nije postojao pristup realnim podacima tokom obučavanja, već su korišćeni isključivo sintetički podaci.

Sintetički podaci su značajno povećali količinu i kontekstualnu raznovrsnost podataka. Iako je sintetičkih podataka bilo dosta više, trening je ponovo trajao relativno kratko. Model je konvergirao za 10-ak minuta i dostigao tačnost od 100% (Slika 13(a)) na trening skupu i najveću tačnost od 99.95% (Slika 13(b)) na validacionom skupu dostigao za 15-ak minuta obučavanja (Slika 12).

Slika 12: Krajnja faza procesa obučavanja modela za detekciju teksta na sintetičkim podacima.



Slika 13: Metrike u toku obučavanja modela sa sintetičkim podacima. a) Tačnost modela na trening podacima u toku obučavanja modela. b) Tačnost modela na validacionim podacima u toku obučavanja modela. c) Promena vrednosti stope učenja u toku obučavanja modela. d) CTC gubitak u toku obučavanja modela. e) NRTR gubitak u toku obučavanja modela. f) Gubitak u toku obučavanja modela.

Skup sintetičkih podataka je podeljen u sledeće podskupove:

- Trening skup je činilo 70% podataka, što je 83.300 slika;
 - Validacioni skup je činilo 15% podataka, što je 17.850 slika;
 - Test skup je činilo 15% podataka, što je 17.850 slika.

Na nezavisnom testnom skupu podataka model je dao pogrešne predikcije za samo 5 od 17.850 slika, beležeći tačnost od 99.97% na nezavisnom skupu

podataka. Slike tablica na kojima je model pogrešio nalaze se na Slici 14. Jasno se vidi da je zajednička karakteristika prve četiri slike to što se nalaze na tamnim pozadinama sa tekstom čija boja nije lako separabilna od boje pozadine. Poslednja prikazana slika tablice na Slici 14 ima veću razliku u boji između teksta i pozadine, a razlog zbog kojeg je bila problem jeste grb sa tablice koji se našao direktno iza prve cifre, pa je model prvu cifru 6 prepoznao kao „\$“.



Slika 14: Primeri slika tablica na kojima je model obučavan samo sa sintetičkim podacima dao pogrešne predikcije.

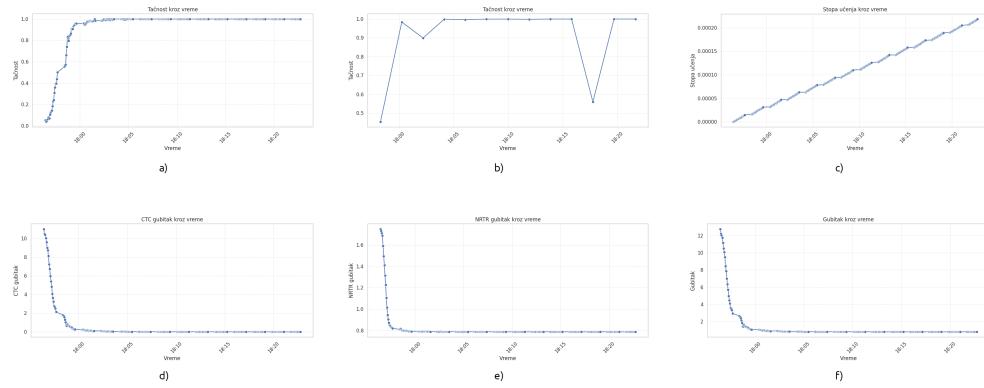
Ovaj model, u odnosu na model obučavan na realnim slikama, ima manji ali i dalje visok stepen sigurnosti prilikom predikcije. Prosečna sigurnost modela obučavanog na sintetičkom skupu podataka je 98,73% na testnom sintetičkom skupu podataka.

6.1.3 Obučavanje koristeći realne i sintetičke podatke

Finalni skup podataka koji je korišćen za obučavanje modela predstavlja kombinaciju realnih i sintetičkih podataka. Idealno, za obučavanje bi se pretežno koristili realni podaci, a sintetički podaci bi se dodavali kako bi doprineli diverzifikaciji skupa realnih podataka. Na taj način se osiguravamo da kontekst koji model uči tokom obučavanja što vernije odražava uslove u kojima će model funkcionisati u praksi. Međutim, uvezši u obzir nemogućnost pristupa većem skupu realnih podataka od onog je korišćen, nedostatak je nadoknađen dodavanjem velikog broja sintetičkih podataka.

Obučavanje na spojenom skupu realnih i sintetičkih podataka je trajalo oko 30 minuta. Model je dostigao tačnost od 100% (Slika 16(a)) na trening skupu i 99.97% (Slika 16(b)) na validacionom skupu podataka (Slika 15).

Slika 15: Krajnja faza procesa obučavanja modela za detekciju teksta na spojenom skupu realnih i sintetičkih podataka.



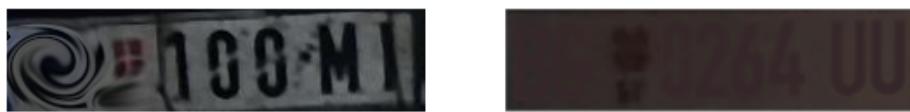
Slika 16: Metrike u toku obučavanja modela sa realnim i sintetičkim podacima. a) Tačnost modela na trening podacima u toku obučavanja modela. b) Tačnost modela na validacionim podacima u toku obučavanja modela. c) Promena vrednosti stope učenja u toku obučavanja modela. d) CTC gubitak u toku obučavanja modela. e) NRTR gubitak u toku obučavanja modela. f) Gubitak u toku obučavanja modela.

Kombinovani skup realnih i sintetičkih podataka je podeljen u sledeće podskupove:

- Trening skup je činilo 70% podataka, što je 87.386 slika, od kojih je 4.086 realnih i 83.300 sintetičkih slika;
 - Validacioni skup je činilo 15% podataka, što je 18.725 slika, od kojih je 875 realnih i 17.850 sintetičkih slika;

- Test skup je činilo 15% podataka, što je 18.725 slika, od kojih je 875 realnih i 17.850 sintetičkih slika.

Na nezavisnom testnom skupu podataka model je dao pogrešne predikcije za samo 2 od 18.725 slika, beležeći tačnost od 99,98% na nezavisnom skupu podataka. Slike tablica na kojima je model pogrešio prikazane su na Slici 17. Model koji je obučavan na kombinovanom skupu realnih i sintetičkih podataka je imao problem sa predikcijom na slici tablice koja je pomoćnim trakama koje prelaze preko teksta zakačena za vozilo. Ovaj primer ilustruje korišćenje softvera u realnim uslovima, gde osoba koja kreira sintetički skup podataka ne može uvek da predviđa sve moguće scenarije koji se mogu desiti u praksi. To dodatno naglašava važnost pravih podataka i objašnjava zašto svi koji razvijaju ozbiljne modele teže ka potrazi za pravim podacima. Ovaj model, slično kao i model obučavan samo na sintetičkim podacima takođe pokazuje problem sa izrazito malim razlikama u boji između teksta i pozadine, ali u manjoj meri. Takvo pogrešno prepoznavanje teksta može se smatrati prihvatljivim, s obzirom na to da je veoma teško pročitati šta zapravo piše na takvoj slici.



Slika 17: Primeri slika tablica na kojima je model obučavan na realnim i sintetičkim podacima dao pogrešne predikcije.

Kombinovanjem realnih i sintetičkih podataka za vreme obučavanja, stepen sigurnosti modela prilikom predikcije je povećan u odnosu na model obučavan samo nad sintetičkim podacima. Prosečna sigurnost modela je 99,32% na testnom skupu podataka. Sigurnost je i dalje nešto manja u odnosu na model obučavan koristeći isključivo realne podatke, što dodatno ukazuje na važnost korišćenja realnih podataka prilikom obučavanja modela.

6.1.4 Uporedna analiza tačnosti modela na sva tri testna seta

Nakon obučavanja tri modela na opisanim skupovima podataka, upoređena je njihova tačnost i sigurnost prilikom predikcije na svim testnim skupovima (Tabela 1).

Model treniran nad	Test skup podataka						
	Realni		Sintetički		Realni i Sintetički		Sigurnost
	Tačnost	Sigurnost	Tačnost	Sigurnost	Tačnost	Tačnost	
realnim podacima	100	99.91	97.7	99.38	97.81	99.4	
sintetičkim podacima	99.42	99.25	99.97	98.73	99.94	98.75	
realnim i sintetičkim podacima	99.88	99.93	99.99	99.29	99.98	99.32	

Tabela 1: Uporedna analiza tačnosti tri obučena modela na zasebnim testnim skupovima.

Model koji je obučavan i testiran na realnim podacima postigao je najveću tačnost i sigurnost, što je očekivano s obzirom na to da su trening i testni skup sadržavali vrlo mali broj slika tablica koje su izdate isključivo na teritoriji Beograda. Ipak, iako ovaj model ima najbolju tačnost na realnim testnim podacima, njegova tačnost na ostalim testnim skupovima je značajno lošija u odnosu na ostale modele. Zbog toga, model koji je obučavan samo na realnim podacima nije najbolji izbor među tri ponuđena modela.

Najlošiju tačnost i sigurnost imao je model obučavan na realnim, a testiran na sintetičkim podacima. Vrlo sličan, ali nešto bolji rezultat ostvario je isti model na testnom skupu kombinovanih realnih i sintetičkih podataka. Jedan od razloga je to što je model tokom obučavanja na prvim dvema pozicijama video samo „BG“ niz karaktera, dok su testni skupovi sadržali tablice iz svih opština Srbije koje imaju pravo da izdaju registarske tablice. Dodatno, korišćeni font za generisanje sintetičkih podataka je ručno napravljen i možda nije idealno replicirao originalni font, što je moglo uticati na tačnost.

Iako nema apsolutno najveću tačnost i sigurnost, najbolji model za korišćenje u realnim uslovima je onaj obučavan na realnim i sintetičkim podacima. Razlika u odnosu na model sa najvećom tačnošću je samo 0,02%, i može se smatrati zanemarljivom. Ovaj model je imao najraznovrsniji i najobimniji trening skup, što mu uz odlične rezultate na validacionim i testnim skupovima daje prednost u izboru.

6.1.5 Pregled tačnosti startnog modela na sva tri testna seta

Model opšte namene za prepoznavanje teksta, koji je korišćen kao startni model za obučavanje na slikama tablica, ima neprihvatljivo lošu tačnost za producione uslove. Tačnost startnog modela na realnim podacima je 87.68%, na sintetičkim 88.38% i na kombinovanom skupu realnih i sintetičkih podataka 88.35%.

Iako su ove vrednosti tačnosti relativno visoke, one nisu dovoljno pouzданe za producione sisteme koji zahtevaju veoma visoku preciznost, naročito u uslovima promenljivog osvetljenja i neidealnih uglova snimanja. Rezultati tačnosti nad startnim modelom ukazuju na objektivnu potrebu za daljim

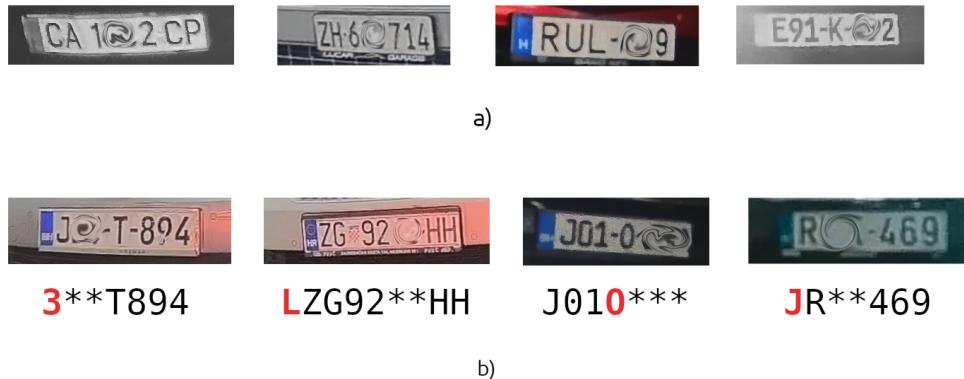
prilagođavanjem modela specifičnim karakteristikama registarskih tablica.

7 Buduća poboljšanja

7.1 Dodatna diverzifikacija skupa podataka

Iako model koji je obučavan pokazuje odlične performanse prepoznavanja teksta sa tablica vozila izdatih na teritoriji Republike Srbije, nije očekivano da će idealno raditi na tablicama vozila iz drugih država zbog razlika u dizajnu, fontovima i karakterima. Svaka zemlja ima svoje specifične standarde kada je reč o izgledu registarskih tablica, što uključuje različite boje, oblike i stilove fonta. Na primer, tablice iz nekih drugih zemalja mogu koristiti karaktere koji se razlikuju od onih na srpskim tablicama, što može dovesti do problema u prepoznavanju.

Bez obzira na to što model prilično dobro prepoznaće tekst sa tablica iz država koje nije video, bilo bi korisno dodati takve podatke u trening skup. Različite države često imaju specifične fontove ili karaktere koje model nije imao priliku da vidi tokom obučavanja sa srpskim tablicama. To može dovesti do grešaka, posebno kada se radi o sličnim karakterima, kao što su „O” i „0” (Slika 18 (b)) ili „I” i „1”. U takvim slučajevima, model može napraviti pogrešnu predikciju, što utiče na ukupnu tačnost i pouzdanost.



Slika 18: Primeri slika stranih tablica. a) Primeri dobro prepoznatog teksta sa slika stranih tablica. b) Primeri loše prepoznatog teksta sa slika stranih tablica sa naznačenim greškama prilikom predikcije.

Zbog ovih razlika, važno je obogatiti trening skup dodatnim podacima iz različitih zemalja kako bi model poboljšao performanse na stranim tablicama. Uključivanje raznovrsnih podataka može pomoći u smanjenju grešaka i povećanju preciznosti prepoznavanja, čime bi se postigla veća efikasnost u stvarnim uslovima.

7.2 Prepoznavanje teksta na više tablica sa iste slike

Trenutna verzija sistema za automatsko prepoznavanje teksta sa tablica vozila podržava pronalaženje i prepoznavanje teksta sa tačno jedne, najveće, tablice sa slike. Ovakvo ponašanje sistema nije idealno u slučajevima kada želimo da sa jedne slike dobijemo informacije o svim registarskim tablicama. Primer takvog sistema bi bio sistem za nadzor parking zona koji se može koristiti za automatsku naplatu parkingu ukoliko korisnik poveže svoj parking nalog sa konkretnim tablicama, ili za pronalaženje lokacije parkiranog vozila.

Verzija sistema koja bi radila prepoznavanje teksta sa svih tablica na slici treba da vrati niz rezultata prepoznavanja teksta, gde bi svaki element niza pored slike isečka tablice i prepoznatog teksta imao i informaciju o tačnoj poziciji tablice na slici. Tačna pozicija tablice bi bila korisna za detekciju tačnog mesta na kojem se korisnik parkirao.

7.3 Ubrzanje rada sistema

Trenutna verzija sistema za automatsko prepoznavanje teksta sa tablica vozila obrađuje jednu ulaznu sliku u proseku oko 1.000ms. Iako trenutno vreme izvršavanja nije previše dugo, kako bi sistem imao bolje performanse i bio u stanju da prihvati više upita potrebno je skratiti vreme izvršavanja izbacivanjem komponenti koje nisu neophodne za rad.

Jedan od koraka ka ubrzaju sistemu jeste uklanjanje modela za detekciju teksta. Trenutno, model za detekciju teksta ima ulogu da nakon detekcije tablica sa slike finije pronađe regiju u kojoj se nalazi samo tekst od interesa prilikom obrade tablice. Taj korak bi mogao biti izbegnut kada bi se odradilo dodatno obučavanje modela za detekciju tablica tako da model za detekciju tablica inicijalno vraća regiju relativno usko oko teksta od interesa bez teksta koji se može naći na okvirima tablica. Pored toga, korišćenje modela za segmentaciju instance umesto detekcije objekta dovodi do znatno efikasnije detekcije regije koju zauzima samo registarska tablica, bez okvira i pozadine koji se mogu obuhvatiti pravougaonom detekcijom objekta.

Drugi korak bi bio korišćenje manjeg inicijalnog modela za prepoznavanje teksta, umesto modela predviđenog da radi na serveru. U tom slučaju performanse u preciznosti prepoznavanja teksta sa tablica vozila bi opale u nekoj meri, ali bi pad preciznosti bio veoma mali i merljiv tek nad značajno velikim skupom podataka. S druge strane, ubrzanje rada modela bi bilo značajno.

8 Zaključak

U ovom radu je na detaljan i sistematičan način analiziran pristup automatskoj detekciji teksta sa tablica vozila u svrhe unapređenja različitih sistema koji se oslanaju na prepoznavanje registarskih oznaka. Postignuti rezultat prepoznavanja teksta je na zadovoljavajućih 99.98% preciznosti i vreme izvršavanja od oko 1.000ms dozvoljava korišćenje predloga rešenja u realnim uslovima.

Posmatrajući rezultate obučavanja modela za prepoznavanje teksta, za postizanje visoke tačnosti ključno je obezbediti podatke koji što vrnije odražavaju uslove u kojima će model biti korišćen u praksi. Na taj način, model će bolje prepoznavati obrasce i karakteristike koji će se pojaviti u novim, nepoznatim podacima. Ako su podaci korišćeni za obučavanje značajno različiti od onih koji se koriste u praksi, model može imati problem sa generalizacijom, što može dovesti do loših predikcija. Stoga je važno osigurati da skup podataka za obučavanje obuhvata raznovrsne i realistične primere koji odražavaju stvarne uslove rada.

Daljim usavršavanjem i razvojem modela i sistema za automatsko prepoznavanje teksta sa registarskih tablica, moguće je postići brže i preciznije rezultate koji će biti primenljivi na još različitijim tipovima tablica.

Literatura

- [Gra+06] Alex Graves i dr. „Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural ‘networks’”. U: sv. 2006. Jan. 2006., str. 369–376. DOI: [10.1145/1143844.1143891](https://doi.org/10.1145/1143844.1143891).
- [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).
- [Bae+19] Youngmin Baek i dr. *Character Region Awareness for Text Detection*. 2019. arXiv: [1904.01941 \[cs.CV\]](https://arxiv.org/abs/1904.01941). URL: <https://arxiv.org/abs/1904.01941>.
- [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, Zheneng 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+20a] Wenyang Hu i dr. *GTC: Guided Training of CTC Towards Efficient and Accurate Scene Text Recognition*. 2020. arXiv: [2002.01276 \[cs.CV\]](https://arxiv.org/abs/2002.01276). URL: <https://arxiv.org/abs/2002.01276>.
- [Hu+20b] 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+21a] 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>.

- [Fan+21b] Yuxin Fang i dr. *You Only Look at One Sequence: Rethinking Transformer in Vision through Object Detection*. 2021. arXiv: [2106.00666 \[cs.CV\]](https://arxiv.org/abs/2106.00666). URL: <https://arxiv.org/abs/2106.00666>.
- [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>.
- [Wan+21] Haifeng Wang i dr. „From object detection to text detection and recognition: A brief evolution history of optical character recognition”. U: *Wiley Interdisciplinary Reviews: Computational Statistics* 13 (jan. 2021.). DOI: [10.1002/wics.1547](https://doi.org/10.1002/wics.1547).
- [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>.
- [Li+22] Chenxia Li i dr. *Dive Into OCR*. Baidu, 2022.
- [Eve23] Dave Van Everen. *The History of OCR*. <https://www.veryfi.com/ocr-api-platform/history-of-ocr/>. Accessed: 15.10.2024. 2023.
- [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).
- [Tri24] Pankaj Tripathi. *The Brief History of OCR Technology*. <https://www.docsumo.com/blog/optical-character-recognition-history>. Accessed: 15.10.2024. 2024.
- [reg] Super registracija. *Gradovi u Srbiji sa pravom na registrovanje vozila*. <https://www.super-registracija-vozila.rs/registarske-oznake-u-srbiji/>. Accessed: 15.10.2024.