

# DETEKCIJA I ČITANJE KODA SA SLIKE

Vujinović Aleksandar

Soft kompjuting

Mentor: Vidaković Dragan



Fakultet tehničkih  
nauka

## Uvod

Cilj projekta je detekcija i čitanje koda sa slike, odnosno *screenshot*-a. Treba prepoznati karaktere, napraviti dobru identaciju koda.

*Screenshot*-ovi koda se razmenjuju svaki dan. Ovaj softver rešava problem ručnog prekucavanja koda, a omogućuje da se automatski preuzme tekstualna reprezentacija koda sa slike.

Na ulazu softver dobija *screenshot*, a softver detektuje gde se nalazi kod na slici, i čita kod sa slike. Izlaz iz softvera je tekstualna reprezentacija koda sa slike.

## Skup podataka

Za svrhu ovog zadatka napravljen je novi skup podataka koji je ručno anotiran. Skup podataka sastoji se od ukupno 120 slika. Skup je raznolik i sadrži *screenshot*-ove različitih kod editora: VS Code, PyCharm, Sublime Text. Editori su u različitim temama, osvetljenjima i veličinama fonta. VS Code je slikan u 8 različitim tema, PyCharm u 2 različite teme i Sublime Text u originalnoj temi. Programski jezici koji su korišćeni su: *Python*, *Java* i *TypeScript*.

## Metodologija

*Mask R-CNN* je korišćen za detekciju i lokalizaciju koda na slici. Korišćen je *transfer learning*, a mreža je pretrenirana na *COCO* skupu podataka. *ResNet101* arhitektura se koristila za treniranje.

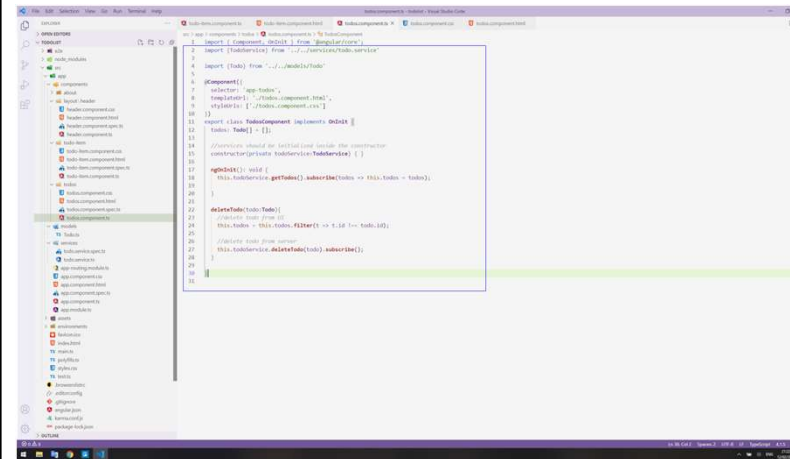
Nakon treniranja možemo uspešno da generišemo masku i *bouding box* oko koda na slici. Time smo završili prvi deo zadatka, lokalizaciju koda.

Kada se pronađe *bouding box*, slika se iseče tako da ostane samo deo slike određen *bouding box*-om. Tako isečena slika se dalje prosleđuje softveru za prepoznavanje teksta na slici.

Za potrebe prepoznavanja teksta korišćen je *Tesseract OCR Engine*. Ovo je javno dostupni softver održavan od strane kompanije *Google*. Softver kao ulaz prima sliku, a kao izlaz vraća tekst koji je pronašao na slici. Zatim se sam tekst i njegov polazaj sa slike prosleđuju funkciji koja vrši identaciju teksta.

Izlaz iz softvera je pravilno uvučeni kod u tekstualnoj reprezentaciji.

Primer uspešne detekcije koda na slici:

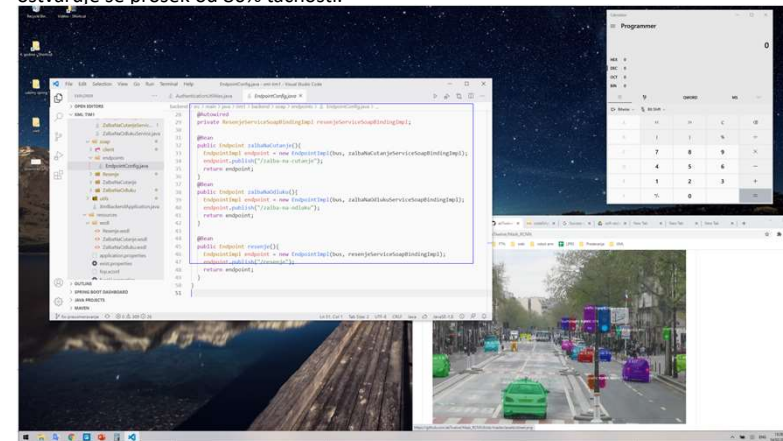


Slika 1. Bouding box dobijen korišćenjem *Mask R-CNN*

## Rezultati

Postoje tri skupa: skup za treniranje koji sadrži 80 slika, za validaciju 20 slika i za testiranje 20 slika. Broj slika po skupovima je u odnosu: 66: 17: 17, respektivno. Za svrhu testiranja, u test skupu postoje slike kodova i njima odgovarajući tekstualni fajlovi koji sadrže kod. Test skup je raznolik i sadrži slike u različitim uslovima. Dodatno otežavajuća okolnost je kada je otvoreno više prozora. Primer takvog slučaja nalazi se na slici 2.

Za evaluacija celokupnog sistema korišćena je *cosine similarity* metrika. Na test skupu ostvaruje se prosek od 86% tačnosti.



Slika 2. Slika sa otvorenim editorom i dodatnim prozorima

## Evaluacija rezultata

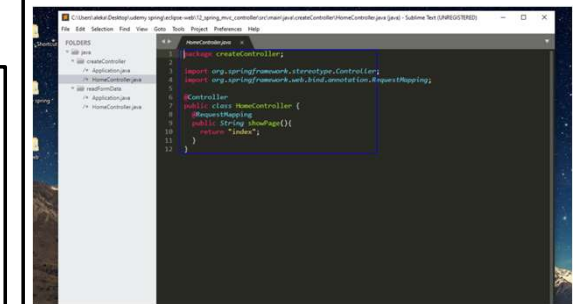
*Cosine similarity* je metrika koja nam služi za određivanje leksičke sličnosti dva teksta. Dva teksta koja poredimo su predstavljena vektorski, a meri se kosinus ugla između njih. Što su dva teksta sličnija to je ugao između njih manji.

Težimo tome da se dva vektora preklope, jer tada imamo najveću sličnost.

Traži se preklapanje istih reči. Takođe, ova metrika uzima u obzir i broj ponavljanja reči. Zbog toga je vrlo poželjna za konkretan problem.

## Potencijalno unapređenje

Najniža tačnost za jednu sliku bila je 60%. Problem leži u *Tesseract OCR Engine*-u, odnosno u samoj detekciji teksta sa slike. *Tesseract* nije uspeo uspešno da detektuje tekst plave boje, sitnog fonta u tamnoj pozadini editora. Takođe taj tekst je ukuso (*italic*) sto dodatno otežava stvar *tesseract*-u. Uz bolji *OCR Engine*, rezultati bi bilo definitivno bolji.



Slika 3. Isečak slike sa najnižom tačnošću

## Zaključak

Podaci iz *COCO* skupa podataka nad kojima je trenirana mreža, se veoma razlikuju od podataka korišćenih u ovom radu. Ipak za ovaj projekat korišćen je *transfer learning* i ostvareni su značajni rezultati.

U skupu podataka nalazi se samo 120 slika. Čak i sa tako malim skupom dobili smo odlične rezultate.

Sistem je vršio lokalizaciju i čitanje koda u različitim editorima, u različitim osvetljenjima, u različitim temama, sa različitim veličinama fonta. Sistem ispravno detektuje kod i kada postoje dodatni otvoreni prozori. Sistem je ostvario uspešnost od 86%.