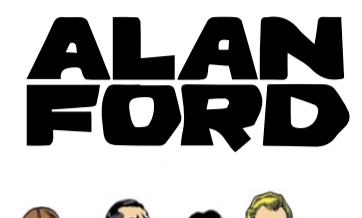


## Motivacija

Ideja je bila napraviti Machine Learning algoritam koji će biti u mogućnosti da čita i oživljava stripove.  
Izabran je strip domaćeg izdavača Alan Ford.  
Kompletan projekat podrazumeva:  
 1. prepoznavanje teksta iz oblačića  
 2. prepoznavanje glavnih likovima  
 3. pripisavanje oblačića likovima  
 4. text to speech sistem koji izgovara pročitano  
 5. "oživljavanje" junaka koji govori na neki način  
 (neophodno da bi se znalo koji junak je taj koji govori)

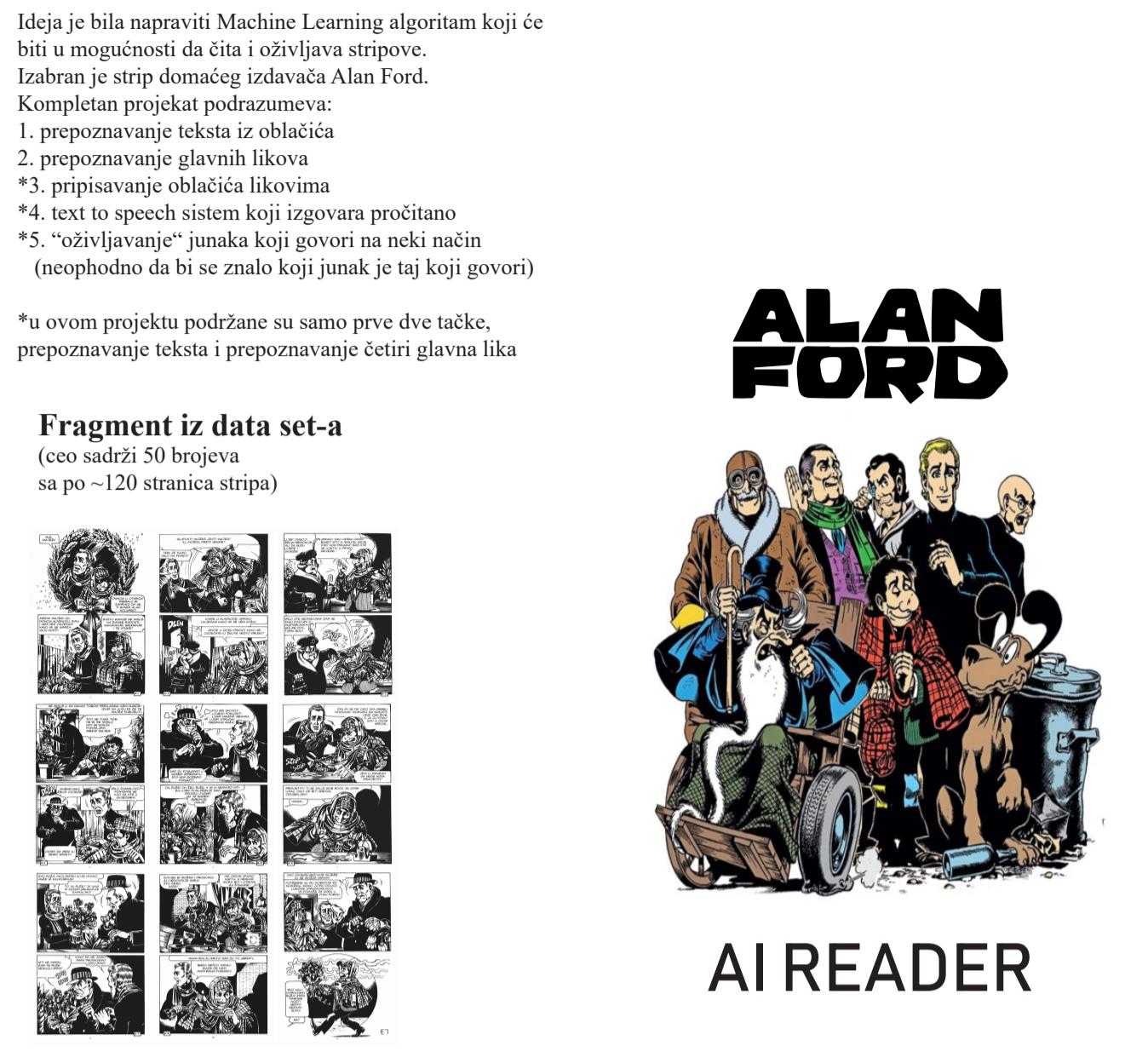
\*u ovom projektu podržane su samo prve dve tačke,  
prepoznavanje teksta i prepoznavanje četiri glavna lika



Fragment iz data set-a  
(ceo sadrži 50 brojeva  
sa po ~120 stranica stripa)

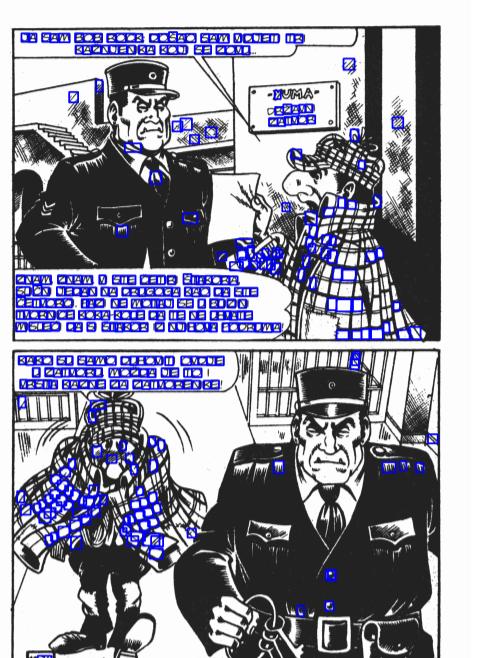


## Prepoznavanje teksta u stripu



Ovaj zadatak se pokazao kao teži nego što je bilo očekivano, s obzirom na prirodu stripa u crno belom formatu i koja sa sobom nosi mnogo distrakcija koje standarnim metodama za prepoznavanje određenih kontura otežavaju posao.

Sledeće je bilo izdvajati plave piksele sa slike, pretvoriti ih u bele i postaviti na novu crnu sliku jedanke veličine.  
Nakon toga izvršena je dilektacija po horizontalnoj osi u cilju gubitnja razmaka između reči i dobijanja spojenih kontura koje mogu biti prepoznate kao oblačići:



Slika 1

To još uvek nije bilo dovoljno za određivanje oblačića na slici. Ostajao je problem sa konturama koje svojim dimenzijama zadovoljavaju uslove da budu prihvaciene kao oblačići, a da to nisu.  
U cilju otklanjanja ovakvih kontura dodate su još dve provere:

1. Da li je procenat belih piksela u konturi dovoljno veliki
2. Da li je ugao pod kojim se nalazi kontura jednak 2π  
Kao što se vidi na slici, ovim postupkom izolovane su same one konture koje stvarno predstavljaju oblačiće (crvenom bojom su obeležene konture koje bi bez ovih dodatnih provera bile okarakterisane kao oblačići):



Slika 2

Nakon što su regioni koji predstavljaju oblačiće uspešno izdvojeni:

JA SAM BOB ROCK DOŠAO SAM VIDJETI TRI KAŽNJENIKA KOJI SE ZOVU...  
DRAŽAVNI ZATVOR  
ZAJAM, ZAJAM VI STE ĆETIRI ŠTAKORA, EUCAN JEDAN NA DRUGOGA KAD DA STE ĆETVORCI, PAZI NE MOTAJ SE U BLIZINI TVORNICE KOKA-KOLE, DA TE NE UHVATE MLEŠEĆI DA SI ŠTAKOR IZ NJUHOVA PODRUMA!  
KAKO SU SAMO DUHOVITI OVOJE U ZATVORU, MOŽDA JE TO I VRSTA KAZNE ZA ZATVORENIKE!

Slika 4

Bilo je potrebno izvršiti sortiranje regiona u od leva desno i od gora do donje.  
Tu su se pojavili novi problemi, naime, koliko god pokušavali da izdvojimo samo one konture koji odgovaraju slovima, uvek bismo u specijalnim slučajevima zajedno pokupili i po neku neželjenu konturu koja bi onemogućila uspešan rad algoritma za sortiranje.



Slika 5

Za rešavanje ovog problema, uveden je DBSCAN algoritam koji izdvaje samo slova tako što formira klaster koji mora zadovoljiti dva uslova:  
 1. mora sarzati dovoljno veliki broj kontura  
 2. konture medusobno moraju biti dovoljno blizu

Ovim smo konačno došli do trenutka kada iz oblačića možemo izdvojiti samo slova i propustiti ih kroz veštaku neuronsku mrežu koju smo prethodno obučili korisćenjem obeleženog seta svih slova iz abecede koje smo prethodno sveli na jednake dimenzije 28x28 piksela.

0 1 2 3 4 5 6 7 8 9  
A B C Č Č Č D D E F G  
H I J K L M N O P R  
S Š T U V W X Y Z Z

Slika 6

Konačno, algoritam uspešno uspeva da pročita tekst iz strip-a, osim u retkim situacijama kada sortiranje ne može biti uspešno izvršeno.  
Primeri kada dolazi do neuspelog sortiranja:

ZELITE LI CMJEĆE ZA  
ROBODAN, CMJEĆE ZA  
GRAB POŠTUNJE PUNICE  
CMJEĆE ZA PRVI  
POLJUBAC, CMJEĆE ZA...

Slika 7  
Gornja granica slova Č niza je od donje granice nekog reda ili slova iz prethodnog reda

U okviru oblačića koji želimo da obradujemo našao se deo drugog oblačića zbog koga je nemoguće sortirati konture.

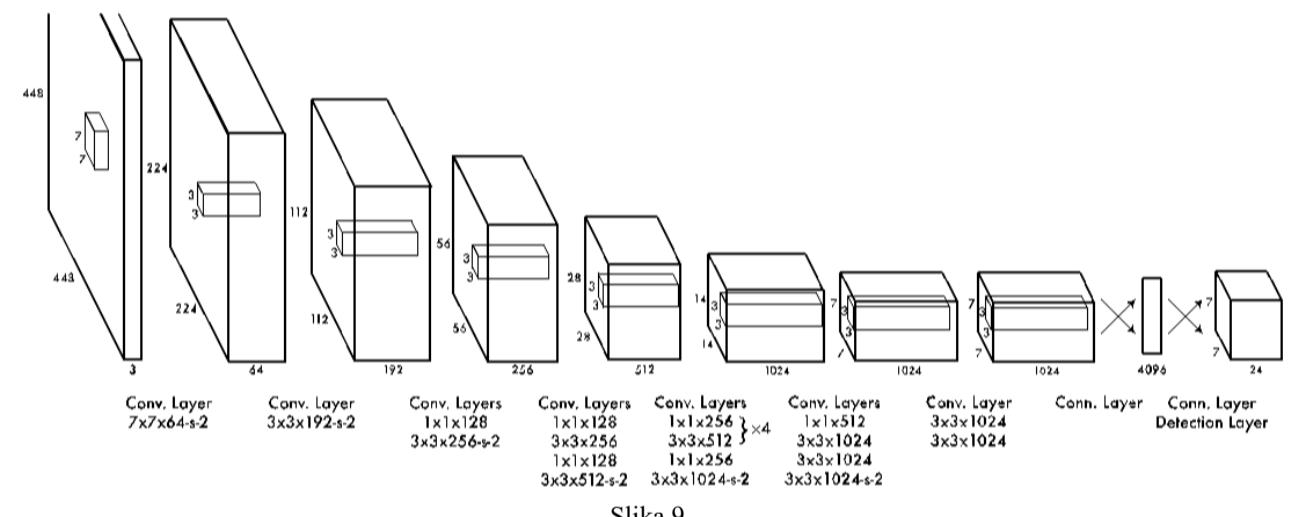
## Prepoznavanje lica - korišćenjem YOLO algoritma

YOLO (You Only Look Once) je algoritam za detekciju i klasifikaciju objekata napisan u darknet framework-u u programskom jeziku C. U ovom projektu korišćen je oblik darknet-ja koji je preveden u Python, prilagođen za Tensorflow i nazvan darkflow.

Pretходnici ovog algoritma su za detektovanje objekta na slici morali svaki pojedinačni segment slike slati na predikciju u prethodno obučenu konvolucionu neuronsku mrežu i takav pristup je zahtevaо veliki broj iteracija i veliki broj predikcija.

Ideja iz YOLO(You Only Look Once) algoritma je u tome da se slika za obradu samo jednom provlači kroz konvolucionu neuronsku mrežu.

Arhitektura mreže:



Slika 9  
Mreža se sastoji iz dva glavna dela, konvolucionih slojeva za ekstrakciju značajnih delova slike (feature-a) i dva potpuno spojena sloja.

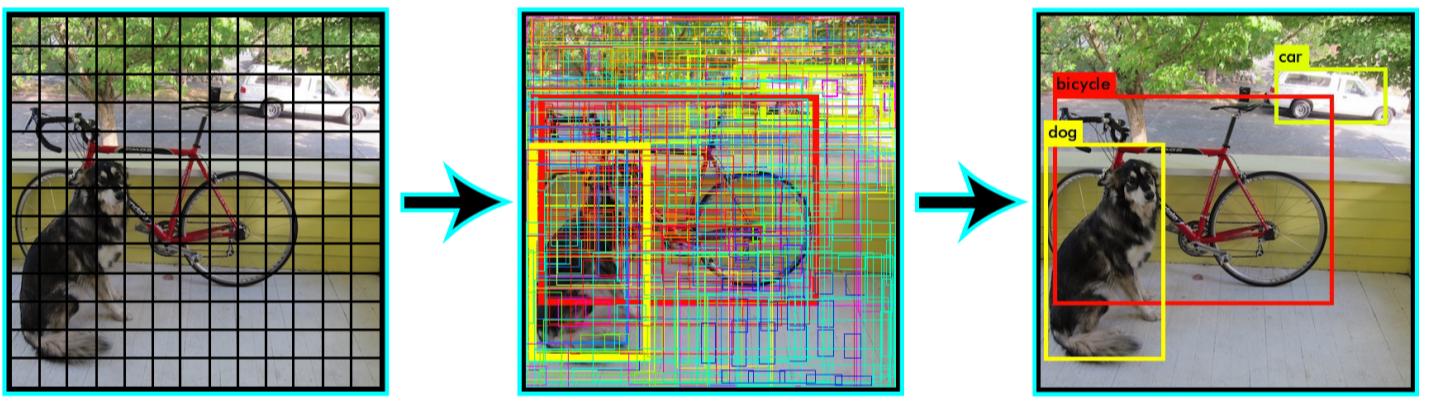
Način na koji YOL O obavlja ceo postupak detektovanja objekata na slici u jednom prolazu kroz CNN je sledeći:

Podaci (slike) za obučavanje mreže se označavaju na poseban način:

- slika podeli na MxM regiona (u ovom slučaju M je 19)
- za svaki objekat na slici određuje se centralna tačka, visina i širina
- s obzirom na to da svaka centralna tačka pripada tačno jednom od regiona, tom regionu se dati objekat i pripisuje
- za svaki region odredujemo i standardnim opisujućim pravougaonika (bounding box) (u ovom slučaju i je 5)
- svakom objektu čija je centralna tačka u datom regionu, pripisujemo jedan od standardnih opisujućih pravougaonika, onaj koji se najviše poklapa sa oblikom objekta.
- formiramo ulazni vektor koji izgleda ovako:

$$\begin{aligned} & \begin{matrix} p_{i1} \\ p_{i2} \\ p_{i3} \\ p_{i4} \\ p_{i5} \\ b_x \\ b_y \\ b_w \\ b_h \\ c_{i1} \\ c_{i2} \\ c_{i3} \\ c_{i4} \\ c_{i5} \end{matrix} \\ & \begin{aligned} i &= \text{broj opisujućih pravougaonika po regionu} \\ k &= \text{broj klase za koji vršimo obučavanje} \\ p_e &= \text{pokazatelj da li u datom pravougaoniku postoji bilo koji objekat} \\ p_o &= \text{pokazatelj da li u datom pravougaoniku postoji bilo koji objekat (u slučaju da ne postoji, ostala polja vezana za ovaj pravougaonik se zanemaruju)} \\ b_x &= \text{x koordinata centralne tačke} \\ b_y &= \text{y koordinata centralne tačke} \\ b_w &= \text{širina objekta} \\ b_h &= \text{visina objekta} \\ c &= \text{pokazatelj da li objekat pripada klasi } i \end{aligned} \end{aligned}$$

Tako označena slika posleduje se neuronskoj mreži na treningu, a kasnija detekcija objekata na novim slikama izgleda ovako:

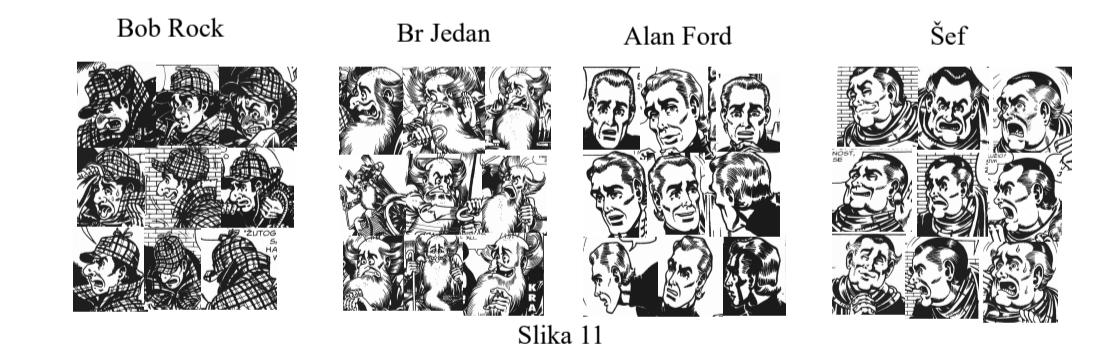


Slika 10

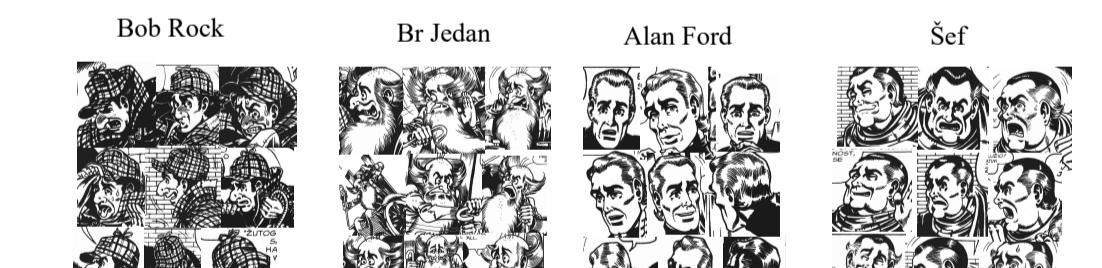
Slika se ponovo deli na regione, svaki od regiona preporučuje i opisujućih pravougaonika, formira se ulazni vektor i kao izlaz dobija se labela pronađenog objekta kao i procenat uspešnosti detekcije koji se dobija kombinovanjem dva procenta, verovatnoće da je u datom opisujućem pravougaoniku bilo koji objekat, i verovatnoće da je objekat uspešno klasifikovan.

Kombinovani rezultat propušta se kroz određenu granicu(threshold) i kao rezultat dobijamo označene detektovane i označene objekte na slici.

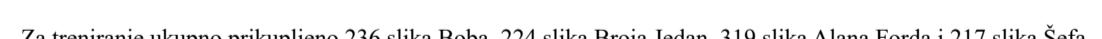
Model koji je istreniran u ovom projektu, sposoban je da detektuje i klasificuje 4 junaka strip-a Alan Ford.



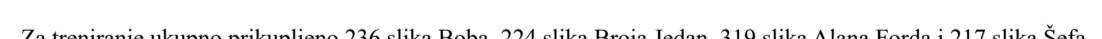
Slika 11



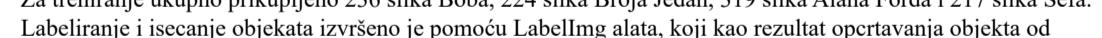
Slika 11



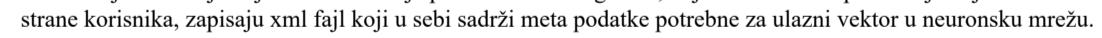
Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11



Slika 11

