Fork instance segmentation

Ivan Mikić

1. Baza podataka
   1. Opšte informacije

Microsoft Common Objects in Context (MS COCO) je skup podataka za razvoj i testiranje računarskih vizualnih modela, posebno za zadatke poput prepoznavanja i segmentacije objekata u slikama. Ovaj skup podataka sadrži slike različitih scena i objekata, uključujući i kontekst u kojem se objekti pojavljuju. U ovom projektu je bio koriščen MS COCO skup podataka iz 2017 godine koji je bio filtriran da bi dobili podskup sa slikama koje sadrže viljuške.

* 1. Statistika - broj klasa/objekata, broj slika, slicne relevantne brojke i histogrami

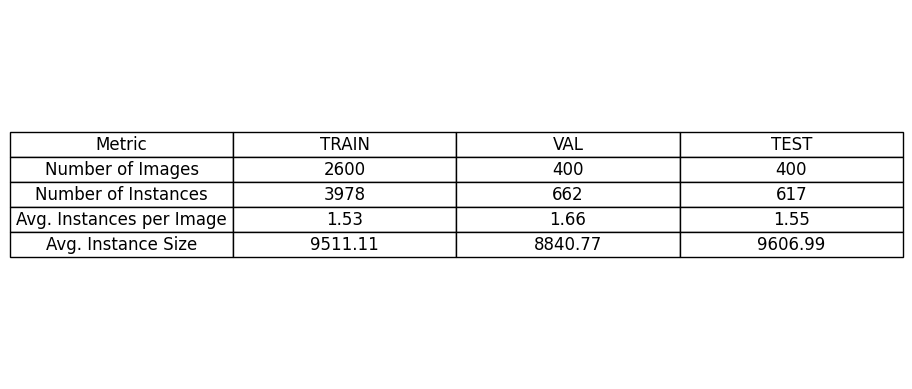
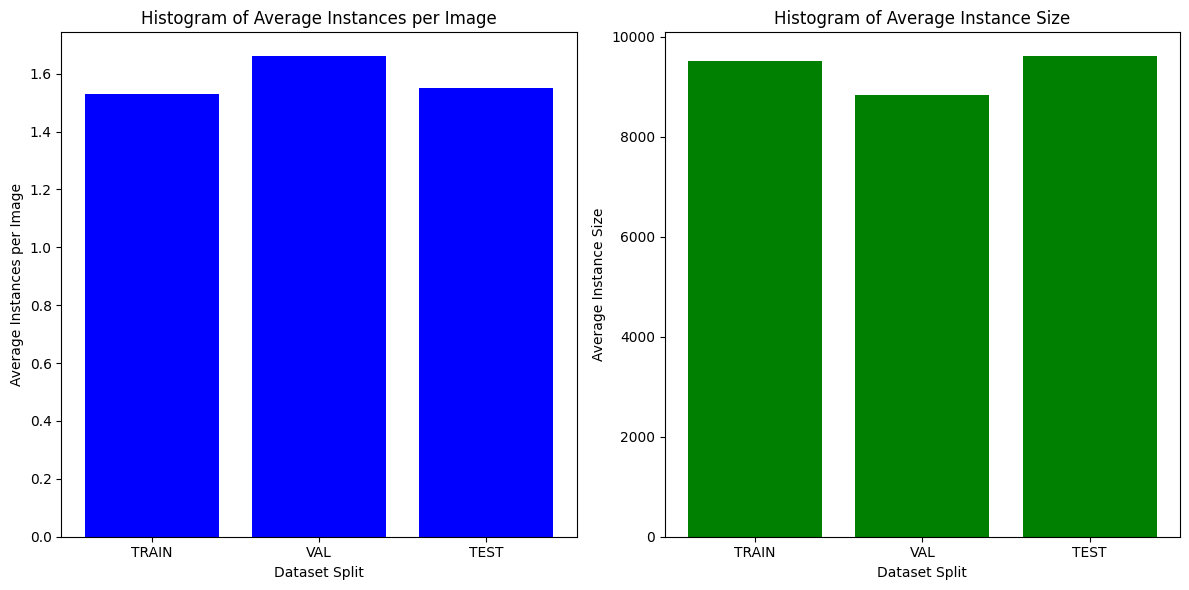


Figure 1. i 2. Statistika o podeli filtriranog skupa

* 1. Predobrada podataka  
     Baza za 2017 godinu ima 165.000 slika sa anotacijama za 91 kategoriju, ali je fokus ovog eksperimenta detekcija viljuške pa su filtrirane samo slike koje sadrže viljuske, što čini 3400 slika. Labele koje se ne odnose na viljuške su bile uklonjene tako da ostaje samo dve kategorije, viljuška i pozadina. Na svakoj od slika se nalazi bar jedna viljuska, a postoje i slike gde su viljuske parcijalno zaklonjene drugom viljuskom ili stranim objektom (Slika 1. I 2.)



Slike 1. I 2. Primeri zaklonjenih viljušaka

* 1. Podela baze u podskupove

Podela dobijenih 3400 slika je bila na skupove training/validation/test koji su sadržali slike u odnosu 2600/400/400. Podela na podskupove je bila birana slučajno.

1. Metodologija
   1. Dizajn eksperimenta

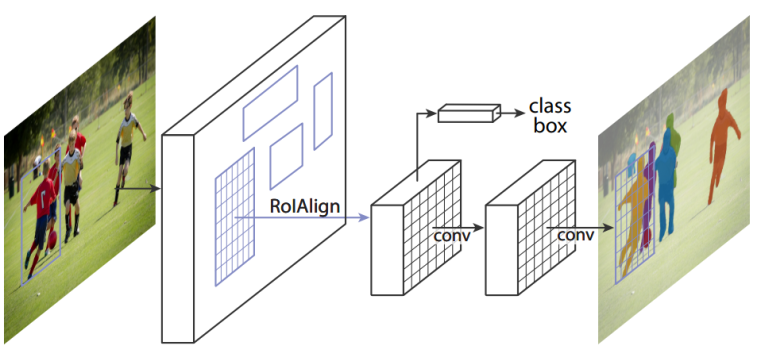
Eksperiment je urađen na opisanoj bazi primenom MaskRCNN mreže, rezultat je prikazan pomoću recall i nekoliko varijanti precision metrike.

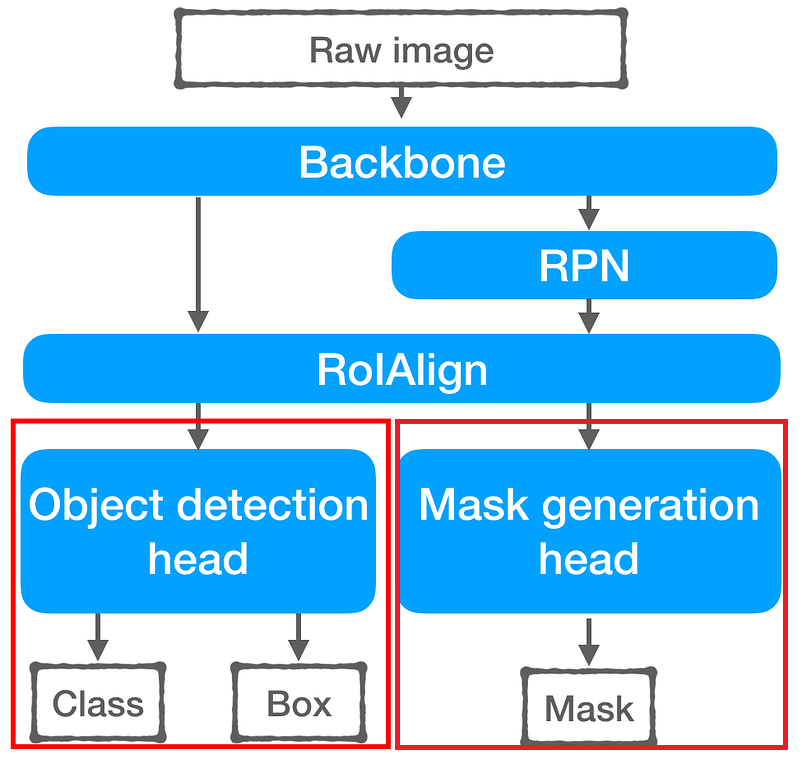
* 1. Predobrada podataka za model

Model zahteva specifičan način prosleđivanja slika i anotacija, tako da traniranje zahteva predobradu podataka. Anotacije se parsiraju iz COCO formata i pravi se rečnik sa svim potrebnim informacijama. Konvertuju se segmentacioni poligoni u binarne maske i stvaraju se bounding box-ovi. Slike se normalizuju i pretvaraju u tenzore.

* 1. Opis korišćenih algoritama i modela
     1. Arhitektura mreže
        1. Osnovne karakteristike

MaskRCNN (Mask Region-Based Convolutional Neural Network) je model koji predstavlja značajan napredak u segmentiranju instanci na slikama. Ovaj model je modifikovani FasterRCNN sa dodatnom glavom za generisanje maske. Njegova arhitektura se sastoji od nekoliko delova: Backbone, RPN (Region Proposal Network), RoIAlign, Object Detection Head, Mask Generation Head (slika 3.). Backbone predstavlja konvolucionu mrežu koja izdvaja karakteristike sa ulaznih slika. U ovom projektu je korišćen RESNET-50-FPN backbone. Izdvojene karakteristike sa slika se šalju u RPN gde se prelažu regije sa objektima na slici i time se vrši prvi prolaz kroz sliku. Na predloženim regijama se vrši bilinearna interpolacija i vrši poboljšana verzija poolinga regija. Predložene regije se šalju u Object Detection Head i Mask Generation Head i paralelno se izvršava detekcija i maskiranje instance.





Slike 3. I 4. Arhitektura MaskRCNN

* + - 1. Unapređenje naspram ranijih modela

Razvojem do MaskRCNN se desilo nekoliko unapređenja u arhitekturi. Najveći uzrok sporog rada RCNN modela je bilo zbog količine predloženih regija u prvom prolazu. Inicijalno rešenje je bilo da se vrši pooling tih regija i samim tim da se smanji broj prolaza koji je bio protreban za detekciju objekata. U razvijanju MaskRCNN nastao je problem oko nepreciznosti koje nastaju zbog poolinga regija interesa. Problem je rešen sa RoIAlign koji koristi bilinearnu interpolaciju da bi dao precizniji bounding box za detekciju i segmentaciju.

* + 1. Hiperparametri za podešavanje

U ovom projektu su promenjeni broj klasa koji postoje na Object Detection Head i Mask Generation Head. Neki od hiperparametara koji definišu rad modela su IoU threshold ,Mask threshold, minimalna i maksimalna veličina ulazne slike, broj slojeva koji se treniraju, itd.

* + 1. Tehnicki detalji - tip podatka na ulazu/izlazu

Tokom treniranja na ulazu u model dolazi do menjanja veličine slike. Postoje minimalna i maksimalna veličina ulazne slike koji se po potrebi mogu podešavati. Tokom treniranja ulaz u model su slike i lista rečnika anotacija (bbox,labele i maske) , a izlaz je loss rečnik. Kad se model ne trenira, ulaz su slike, a izlaz je rečnik sa svim bbox, score-ovima, maskama i labelama.

* 1. Proces treninga

Za treniranje ovog modela je bio koriščen transfer learning da bi izbegli da treniramo ovoliko dubok model u celosti. Uzimamo ranije istreniran model i menjamo mu Object Detection Head i Mask Generation Head tako da razlikuje samo viljuške i pozdadinu. Pri treniranju treniraju se samo poslednja tri sloja u ove dve glave, a ostali slojevi ostaju nepromenjeni.

* 1. Metodologija evaluacije
     1. Opis metrika koje su korišćene

Metrike korišćene za evaluaciju su Averidge Precision i Averidge Recall. AP predstavlja metriku koja daje vrednost tačnih pozitivnih detekcija u odnosu na sve pozitivne detekcije za određeni IoU prag, a AR prestavlja vrednost pozitivnih detekcija u odnosu na sve tačne detekcije za određeni IoU prag. Intersection over Union predtavlja metriku kojom možemo da odredimo količinu preklapanja dva bounding box-a, i samim tim stavimo nivo tolerancije.

* + 1. Način na koji su metrike korišćene

Metrike AP i AR analizirane su za različite vrednosti pragova za IoU, uključujući širok raspon od 0,50 do 0,95, kako bi se razumela preciznost i odzivnost modela u različitim pragovima za tačnošću detekcije.

1. Rezultati

Tokom treniranja loss je opadao, ali nije došlo do izravnjavanja. Samim tim mogu da zaključim da modelu treba dodatno treniranje da bi postigao bolje rezulate. Podaci o loss krivi su izgubljeni.

* 1. Kvantitavni rezultati

Model je pokazao kroz rezultate da nema veliku preciznost (Figura 3.). Za niži prag IoU model postiže bolje rezultate.

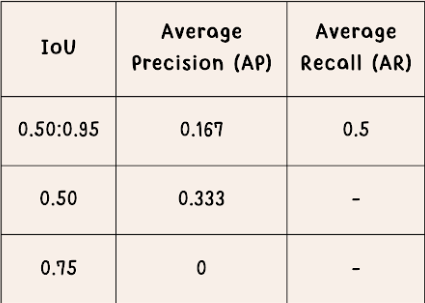
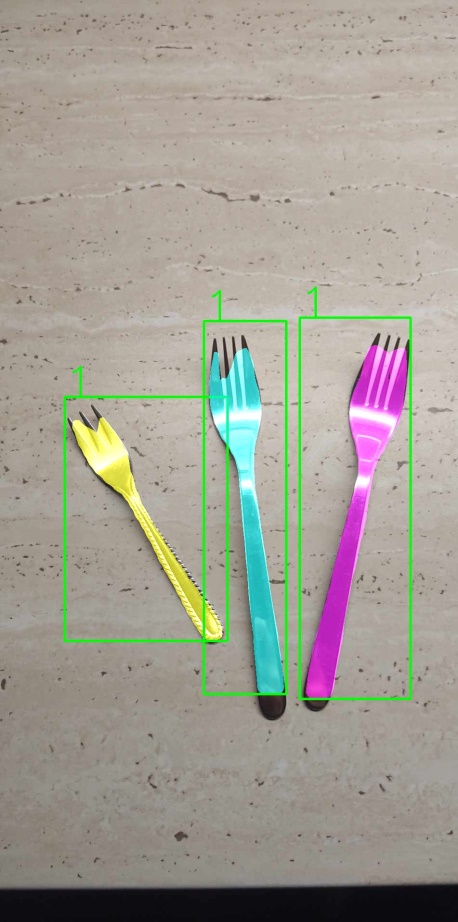


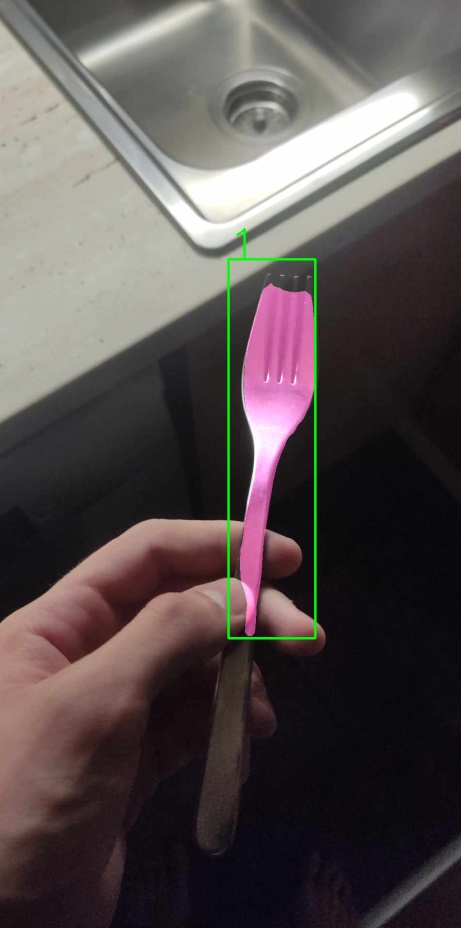
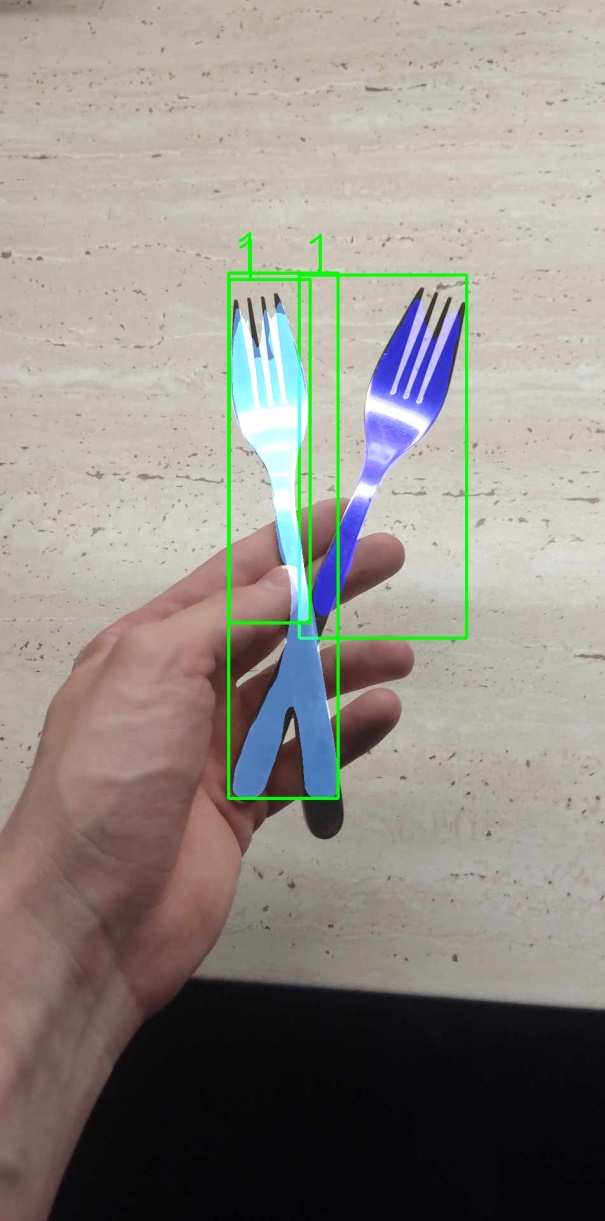
Figura 3. Metrike performansi modela za bbox

* 1. Kvalitativni rezultati

Možemo primetiti da model dobro prepoznaje i maskira viljuške koje nisu zaklonjene ni sa čim (Slike 4. I 5.), a pri maskiranju ima najviše problema sa vrhom viljuške i obeležavanjem sva 4 špica . Za viljuške koje su zaklonjene ili drugim objektom (Slika 7.) ili drugom viljuškom (Slika 6.) dolazi do zanemarivanja jednog dela viljuške ili višestrukog prepoznavanja iste instance.



Slike 4. I 5. Izlazi iz modela



Slike 6. I 7. Greške koje model pravi

1. Diskusija
   1. Algoritam je postigao željene rezultate delimično. Treniranje je moglo duže da traje i da se postignu drugačiji rezultati. Za model koji je istreniran delimično daje zanimljive rezultate. Detekcija i segmentacija imaju dobre rezultate za jednostavne primere gde je samo jedna ili više nezaklonjenih viljušaka. Kod kompleksnijih primera se dešavaju ili delimične ili duple detekcije. Takođe se pojavljuje i problem u maskiranju vrhova viljušaka. U bazi podataka postoje primeri gde su vrhovi zasebno obeleženi, a takođe i primeri gde su zajednički obeleženi što može da dovede do neželjenih rezultata u modelu.
   2. Za dalji rad bi preporučio bolju pripremu baze podataka i veći fokus na kompleksnije vrste detekcije. Sa takvim informacijama se hiperparametri modela mogu bolje nameštati. Takođe bih dodao i augmentaciju u treniranje i na neki način zabeležio loss krivu.
2. Zaključak

U ovom radu je prikazan moj prvi eksperiment sa MaskRCNN i uopšteno poljem kompjuterske vizije. Postignuto je upoznavanje sa ovim modelom i bazama podataka koje se bave detekcijom i segmentacijom.

Reference

<https://arxiv.org/pdf/1703.06870.pdf>

<https://arxiv.org/pdf/1405.0312.pdf>

<https://github.com/cocodataset/cocoapi/tree/master/PythonAPI/pycocotools>