Segmentacija slika

Mihajlo Madžarević 55/20RN

1. Uvod u problem i podaci

Segmentacija slika je jedan od modernih problema dubokih neuronskih mreža. Segmentacija slike podrazumeva piksel detaljno identifikovanje različitih objekata i bića na slici. Za potrebe učenja modela koristićemo COCO set podataka koji sadrži segmentacije za razne objekte i životinje. Mi ćemo se fokusirati na segmentaciju i razlikovanje mačaka i pasa. COCO set je dostupan na

https://cocodataset.org/#home

2. Analiza podataka i zaključci analize

Slike koje koristimo su nasumično izabrane iz COCO seta podataka. Ove slike treba pripremiti kako bi model mogao da ih koristi za trening. Sve slike bi trebale biti istih dimenzija sa istim brojem kanala. Za potrebe učenja našeg modela definisaćemo tri klase: ništa, mačka pas.

Primetimo sledeće pojave u podacima:

1. Slike su različitih dimenzija, neke su slikane vertikalno, a neke horizontalno.
2. Neke od slika su u boji, a neke su crno bele.
3. Broj segmentacija i njihova veličina se razlikuje za svaku sliku i potencijalno sadrži entitete koje nisu psi i mačke.
4. Odnos klasi ništa, mačka, pas su zastupljeni u različitim merama, tačnije broj piksela na kojima nemamo pse ni mačke, na kojima imamo samo mačku ili samo psa se razlikuje.
5. Mačke i psi su na određenim slikama lako uočljive u kadru, dok na drugim se slabo primećuju.
6. Postoje slike na kojima je teško razlikovati psa od mačke čak i ljudskom oku.

Slike ćemo ostaviti u boji jer ovaj detalj potencijalno može biti od koristi modelu. Različite dimenzije slika ćemo skalirati u veličinu 224x224 kako bismo mogli upotrebiti pretreniranu mrežu i izoblikovati je da radi na našem skupu podataka.

Raznovrsne segmentacije regulisaćemo uzimanjem u obzir samo segmentacije mačaka i pasa na slikama. Sve segmentacije ćemo pretvoriti u mape (dvodimenzionalne matrice) u kojima ćemo nulom naznačiti klasu ništa, jedinicom klasu mačka, a dvojkom klasu pas.

Različit odnos klasi ništa, mačka i pas regulisaćemo uvođenjem početnih težina u funkciju troška. Izračunaćemo težinu za pojedinačne klase po formuli:

$$f(x_i) = \frac{x_i}{\sum_{n=1}^{3}(x_n)}$$

gde je xi ukupan broj piksela date klase u svim učitanim slikama.

3. Rešenje i arhitektura modela

Model bi trebalo da predvidi klasu za svaki piksel slike, kao što smo već predložili u sekciji 2. za svaku sliku ćemo napraviti one hot enkodovanu masku kako bi predstavila stanje piksela. U jednom redu i koloni slike će se nalaziti matrica veličine tri gde će respektivno pozicije označavati klase ništa, mačku, psa. Sa brojem 1 će na poziciji u matrici biti naznačena pripadnost klasi i na svim ostalim pozicijama biti upisana vrednost 0. Ovo su nam podaci koje ćemo koristiti kao labele za model.

Za rešenje koristitićemo tri verzije modela kako bismo uporedili različite arhitekture za dati problem. Sve tri mreže se sastoje iz konvolucionih slojeva, tačnije nećemo koristiti potpuno povezane slojeve na kraju mreže jer želimo da nam izlaz modela bude maska istih dimenzija kao i ulaz modela. Arhitekture mreža su sledeće:

1. ParseNet arhitektura
   Ovakva arhitektura mreže u osnovi se sastoji iz enkodera i dekodera. Enkoder smanjuje dimenzionalnost slike dok je dekoder povećava. Enkoder se sastoji iz konvolucionih blokova (dva konvoluciona sloja koja ne smanjuju dimenzionalnost slike) ispraćene MaxPooling slojem. Dekoder se sastoji iz UpSample sloja koji povećava datu ulaznu sliku, Concatenate sloja koji datu povećanu sliku spaja sa odgovarajućim konvolucionim blokom iz enkodera i zatim konvolucionog bloka koji odradi konvoluciju na svim spojenim podacima. Ideja iza ovoga je da pored lokalnih podataka imamo i podatke o globalnom kontekstu koji nam može pomoći u segmentaciji. Videti papir o ParseNet-u za više podataka.

2. Dotrenirana MobileNet mreža
   MobileNet je U-net tip mreže koji ima sličan pristup ParseNet arhitekturi, naime ima enkoder i dekoder deo. Ovoj mreži možemo oduzeti poslednjih par slojeva i dodati naše kako bi prilagodili mrežu našem problemu.

3. Fully convolutional network arhitektura
   Pristup je malo drugačiji i prethodnik je U-net arhitekturama. Sastoji se samo iz enkoder dela koji smanjuje dimenzionalnost slike praćen ConvTranspose slojem koji proširi sliku na originalnu veličinu. Ove mreže ne uzimaju u obzir globalni kontekst zajedno sa lokalnim kao što to rade prethodno navedene arhitekture, pa se od ove mreže očekuju slabiji razultati.

4. Eksperimenti

Pre treniranja na većem skupu podataka pokušao sam da pretreniram model na skupu od 5 slika. Uspevši u ovom poduhvatu shvatio sam da je model u stanju da napamet nauči segmentacije na malom skupu podataka. Do ovog ostvarenja dodavao sam i brisao konvolucione blokove zajedno sa MaxPooling slojevima, pratio uticaje dubine mreže, menjao stopu učenja, podešavao broj filtara koje konvolucioni blok uči, uvodio regularizaciju. Nailazio sam često na problem nestajućeg ili eksplodirajućeg gradijenta koji sam rešio uvođenjem batch normalizacije. ParseNet arhitekturi sam posvetio najviše vremena. Bila je veoma uspešna na manjem skupu podataka, međutim kada sam uveo skup podataka od 100-200 slika nije davala dobre rezultate. Metrike mi nisu bile od ukupne pomoći jer sam empirijski utvrdio da ponekad modeli sa lošijom funkcijom troška ili preciznošću daju bolje rezultate nego oni sa boljom. Uvođenjem L2 regularizacije sugestijom papira za ParseNet mrežu nisam značajno poboljšao model. Smatrao sam da model nije dovoljno složen da zabeleži detalje svih slika, međutim čak i sa 14 miliona parametara nisam imao dobre rezultate posle 200 epoha. Ovaj neuspeh smatram zaslužan manjkom podataka, treninga i posvećenošću štelovanja parametara mreže. Sa druge strane pretreniranoj MobileNet mreži nisam morao

posvetiti puno vremena za zadovoljavajuće rezultate. FCN kao i ParseNet mreža, s obzirom da nije pretrenirana, zahtevala je više pažnje oko podešavanja parametara i dubine mreže.

5. Zaključci

U ovom radu pogledali smo tri različite arhitekture konvolucionih mreža i njihov uspeh u segmentaciji mačaka i pasa na malobrojnom skupu slika. Iako nisam ostvario najbolje rezultate u segmentaciji, siguran sam da je ovo ostvarivo uz dodatno uloženo vreme u ispraćivanju ponašanja mreže sa različitom količinom podataka, promenama u složenosti arhitekture i prilagođavanju parametara. Smatram da je najbrži put do zadovoljavajućih rezultata dotrenirati već postojeću mrežu za konkretnu segmentaciju objekata ili bića. Ukoliko bismo sami želeli napraviti model smatram da je arhitektura ParseNet mreže dobra osnova. FCN je zgodna arhitektura, ali empirijski sam utvrdio da manje promene u arhitekturi mreže mogu imati značajan uticaj na promenu performansi mreže (segmentaciju, ne i metrike), dok kod ParseNet mreže ovakva pojava je manje zastupljenja, utvrđeno u sekciji 4. sa smanjenjem dubine mreže i povezanošću enkoder labela sa dekoder labelama. Smatram da je ovo postignuto zahvaljujući sposobnosti ParseNet arhitekture da gleda lokalne i globalne detalje, što nam omogućuje veće vidno polje kao što je to opisano u ParseNet papiru.

Reference

1. Wei Liu, Andrew Rabinovich, Alexander C. Berg, PARSENET: LOOKING WIDER TO SEE BETTER
2. Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos, Image Segmentation Using Deep Learning: A Survey