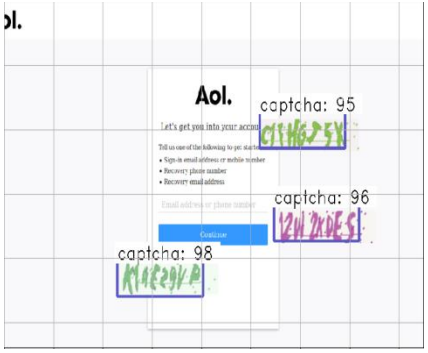
CAPTCHA detekcija i razbijanje

Danijel Radaković, Aleksandar Stefanović – Fakultet tehničkih nauka

Detekcija

Motivacija:

Napraviti botove koji će uspešno detektovati i rešavati CAPTCHA šablone koje se nalaze na *web* formama. Time je omogućeno spamovanje nevalidnog sadržaja koje mogu dovesti do pada servera ili servisa. S druge strane, aplikacija može da se koristi tako što će korisnicima pomoći u rešavanju CAPTCHA-i.



Faster R-CNN model se sastoji od:

- **VGG16** konvolutivne mreže sa već istreniramim težinama
- **RPN** mreza koja preumiza ekstraktovane osobine iz **VGG16** mreže i kreira *anchors* i razbija potpuno povezan sloj iz **VGG16** mreže na 2 konvolutivna. Jedan sloj koji se koristi za klasifikaciju i koristi *sigmoig* aktivacionu funkciju, dok se drugi koristi za regresiju *bounding box*-ova i koristi linearnu aktivacionu funkciju
- **RoI polling** za fiksiranje veličine svih regiona dobijenih iz **RPN**-a
- **Potpuno povezan** sloj sa linearnomg regresijom i *softmax* funkcijom

Treniranje i validacija:

- treniranje je odrađeno u 10 epoha. Svaka epoha je imala 500 slika, a *batch* obrađuje samo jednu sliku sa augmentacijom i sledećim parametrima: *resized*: 300, *anchors*: 9, *RoI*:4, *learning rate*:1e-5
- validacija je odrađena nad 471 slikom i *mAP* iznosi: 0.307

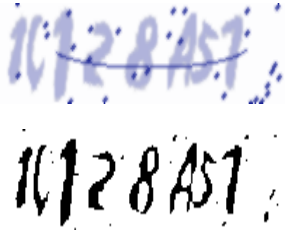
Zaključak:

- potrebno je dotrenirati mrežu posto model nije iskonvergirao

Razbijanje

Pre nego što se slika prosledi u **RNN**, vrši se neophoda obrada slike:

- konvertovanje slike u crno-belu skidanjem broja kanala slike
- thresholding radi uklanjanja šuma na slikama
- morfološke transformacije radi poboljšanja oblika karaktera



RNN model se sastoji od:

- konvolutivne mreže za ekstraktovanje osobina slike. Na ulazu prima sliku 128x64x1, i propušta se kroz 2 konvolutivna sloja sa 16 kernela po 3x3 i maxpolling-om sa kernelom 2x2
- sloj koji vrsi promenu dimenzija slika tako da širina tenzora odgovara broju iteracija u **RNN**, a zatim potpuno povezan sloj radi smanjivanja dimenzionalnosti
- 2 bidirekciona **LSTM** ili **GRU** sloja, a zatim potpun povezan sloj koji konvertuje izlaz iz **LSTM** ili **GRU** sloja u aktivacije karaktera
- Lambda sloj u kojem se računa **CTC** *loss*, a na ulazu prima prethodno opisan aktivacioni sloj, labelu, duzinu labele i duzinu izlaza iz aktivacionog sloja . Za optimizaciju se koristio **SGD**.

Treniranje i validacija:

- treniranje je odrađeno u jednoj epohi nad 13.500 slika sa parametrima: *learning rate*: 0.02, *decay*: 1e-6, *momentum*: 0.9, *batch_size*: 8
- validacija je odrađena nad 1.500 slika, **LSTM** je dostigao tačnost od 82.8%, a **GRU** tačnost od 76.6%

Zaključak:

- obrada slike moze poremetiti rad **RNN** ukoliko detekcija CAPTCHA-e ne pronadje region sa tačnošću 97%, jer se time gube bithe infromacije