



Šerbedžija Luka, Panić Miloš  
FTN 2022, Novi Sad



## UVOD

Igranje igrice uz pomoć VR (eng. *Virtual Reality*) postaje sve popularnije iz godine u godinu. Najveći problem ove tehnologije leži u ceni kvalitetnih uređaja.

CNTS je projekat koji ima za cilj da korisniku približi osećaj VR uz pomoć uređaja koje već poseduje. Konkretno, igranje *Counter Strike 1.6* pomoću pokreta ruke koji se očitavaju web kamerom i glasovnih komandi koje očitava mikrofoni.

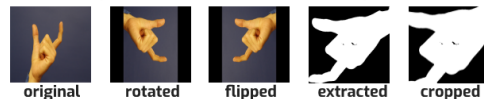
## Skup podataka

Za glasovne komande je iskorišćen gotov skup podataka. Iz njega su izdvojeni zvukovi za pozicioniranje: *Left, Right, Up, Down, Go, Off*.

Za pokrete ruke je ručno napravljen skup podataka. Komande su: promena oružja i navigacija kroz menije (*NUM 1-5*), kretanje (*W*), punjenje municije (*R*), bacanje oružja (*G*), otvaranje prodavnice (*B*), pucanje (*LC*), stavljanje prigušivača (*RC*).



Inicijalno je napravljeno 539 slika nad kojima su primenjene tehnike augmentacije kao što su: *rotate, flip, max contour extract, crop*.



## KLASIFIKACIJA SLIKE

Pre procesa klasifikacije, sliku je potrebno preprocesirati na sledeći način:

1. Konvertovanje u *HSV* prostor jer su nijanse boja dosta slične pri različitom osvetljenju što je bitno zbog boje kože
2. Izdvajanje ruke na osnovu boje kože

3. Primena transformacija koje umanjuju šum na slici
4. Izdvajanje najveće konture koja predstavlja zadati pokret



Tako procesuirana slika se prosleđuje *CNN* mreži.

Za klasifikaciju su korišćena dva modela:

- *VGG* model - na kojem je primenjen *transfer learning* i dodata su tri *Dense* sloja koja su trenirana.
- Naš, ručno napravljen model - sa četiri konvolutivna sloja i četiri *Dense* sloja. Svaki konvolutivni sloj se sastoji od sledećih operacija:

1. Konvolucija (*kernel 3x3*)
  2. Aktivacija (*ReLU*)
  3. Normalizacija podataka (*Batch normalization*)
  4. *Max pooling* (*kernel 2x2, stride 2, 0-padding*)
  5. *Dropout* (*30% verovatnoća*)
- Svaki konvolutivni sloj se sastoji od sledećih operacija:

1. Aktivacija (*ReLU*)
  2. *Dropout* (*30% verovatnoća*)
- Za poslednji *Dense* sloj se koristi *softmax* aktivacija za dobijanje verovatnoća pripadnosti svakoj od klasa.

## KLASIFIKACIJA ZVUKA

Kao i sliku, zvuk je takođe potrebno prvobitno preprocesirati na sledeći način:

1. Ograničiti dužinu zvuka na jednu sekundu
2. Pretvaranje u *MFCC*
3. Dodavanje treće dimenzije

Takva trodimenzionalna reprezentacija zvuka se prosleđuje *CNN* mreži.

Za klasifikaciju je korišćena ručno implementirana *CNN* mreža, koja se razlikuje od mreže korišćene za slike u sledećem:

- *MFCC* ima dubinu 1
- Treća konvolucija ima filter dimenzije (*2, 2*)
- Četvrta konvolucija ima filter dimenzije (*1, 1*) koji služi za smanjenje dimenzionalnosti
- Klasifikacija se radi nad šest klasa

## PRIMENA

Zadate komande se direktno primenjuju na *Counter Strike 1.6*. Za simulaciju miša je korišćena biblioteka *pyautogui*, a za tastaturu *DIK*.

Pokreti ruku se detektuju uz pomoć web kamere gde ruka mora biti u *ROI* (regionu od interesa). Drugo ograničenje je da pozadina mora biti različite boje u odnosu na ruku.

Glasovne komande se detektuju uz pomoć mikrofona. Kada jačina zvuka pređe zadati prag, kreće snimanje koje traje jednu sekundu i nad njime se vrši klasifikacija.

## REZULTATI

Kao metrika tačnosti je korišćen *accuracy*.

	Validacioni	Testni
Naš slike	90%	96%
VGG slike	92%	97%
Naš zvuk	92%	91%

Tačnost nad validacionim skupom za slike nije toliko velika zato što su slike skalirane na *128x128* čime se gubi na detaljima, ali nismo mogli trenirati sa većim slikama zbog nedostatka *RAM*-a.

Tačnost nad testnim skupom za slike je veća zato što se slike ne razlikuju mnogo u odnosu na trening skup.

U primeni se naš model pokazao mnogo bolje u odnosu na *VGG* zato što on u osnovi nije bio treniran nad crno belim slikama i zato što je raden *transfer learning*.

Subjektivno, igrice je moguće upravljati uz pomoć rešenja ali i dalje ima dosta problema prilikom promene uglova. Ovaj problem se može prevazići proširivanjem skupa podataka.

Zvuk se pokazao kao dosta pouzdan kada se reči ne previše brzo ili previše sporo pa da dužina prelazi jednu sekundu. Problem dužine nije moguće rešiti korišćenjem naše mreže zbog fiksne dužine ulaza.

Pored proširivanja skupa podataka, veoma bitno unapređenje bi bilo da pozadina može biti proizvoljna. Jedno rešenje je uvođenje mreže koja bi prepoznala ruku na slici kao što je npr. *YOLO*.