

## MOTIVACIJA

Korišćenje grafičkog alata podrazumeva upotrebu različite tehničke opreme (npr. miš, grafička tabla) koja može biti neprecizna ili skupa.

Najprirodniji način crtanja uključuje direktno korišćenje prstiju i softver koji bi to omogućio značajno bi poboljšao kvalitet interakcije čoveka sa računarom, jer bi manipulacija grafičkim alatom bila moguća bez korišćenja eksternih uređaja.

## SKUP PODATAKA

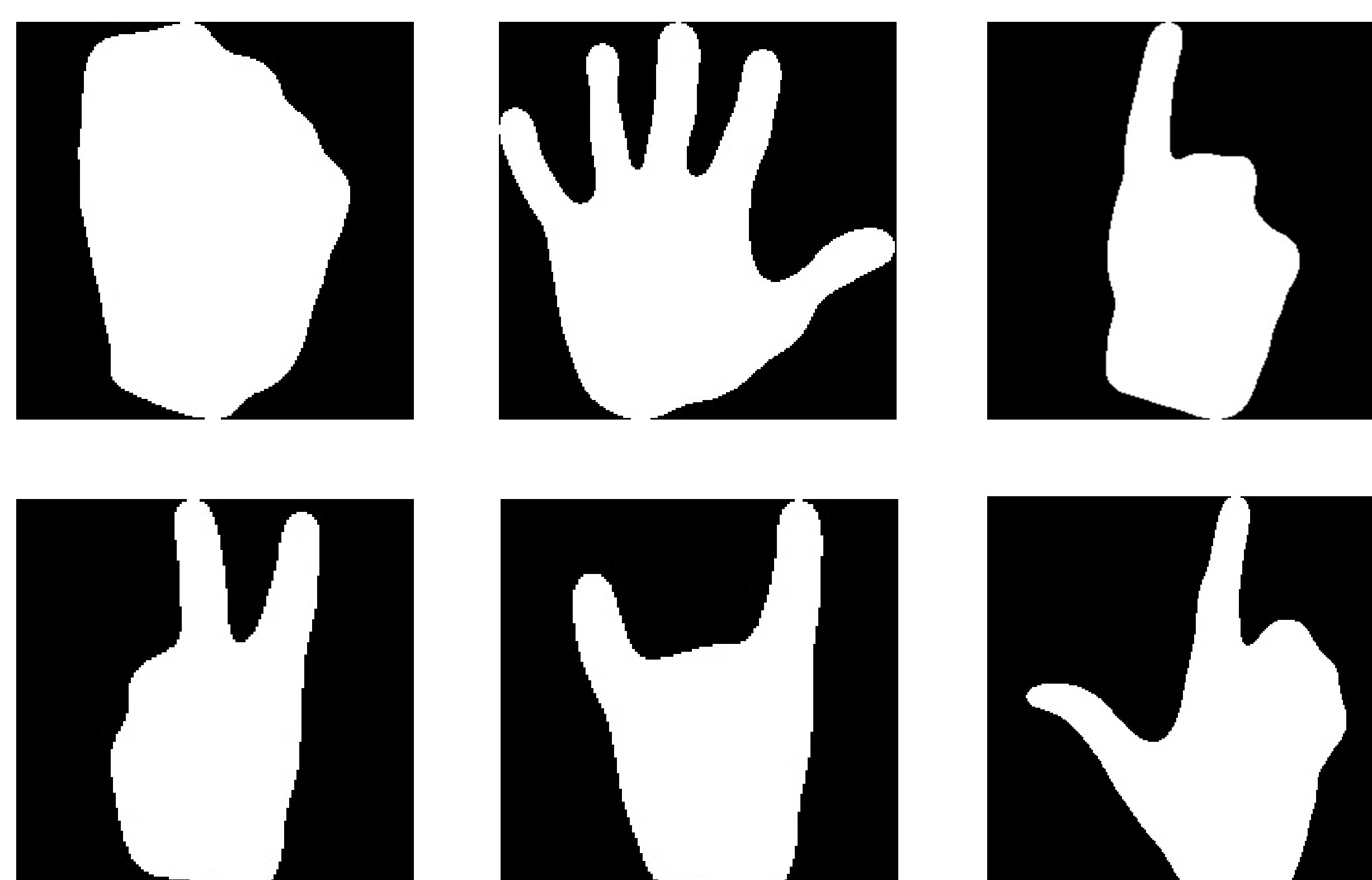
Ručno napravljen. Izabrano je 6 pozicija šake za trigerovanje akcija u *Photoshop*-u:

- Skupljena šaka – *cursor navigation*
- Otvorena šaka – *pan*
- Ispružen kažiprst – *brush*
- Ispružen kažiprst i srednji prst – *move*
- Ispružen kažiprst i mali prst – *delete*
- Ispružen kažiprst i palac – *zoom*



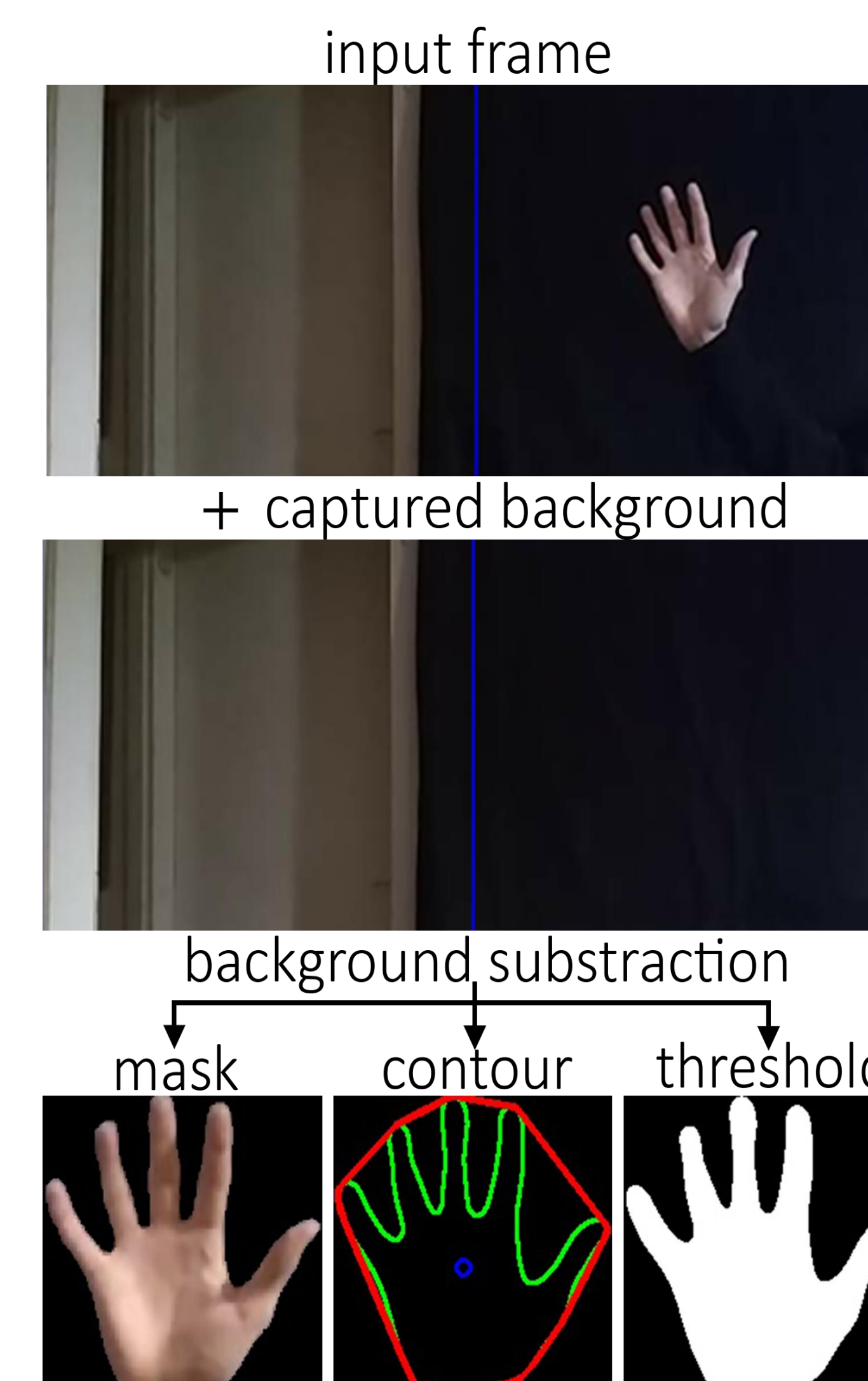
Za svaku klasu postoji 500 slika koje prikazuju thresholdovanu konturu šake, odn. postoji ukupno 3000 slika. Polovina dataseta je flipovana, kako bi aplikacija bila invarijantna na scenarije korišćenja levom odn. desnom rukom.

Skup podataka podeljen je na trening, validacioni i testni set u odnosu 80:10:10.



## PRETPROCESIRANJE SLIKE

1. Izdvajanje šake od pozadine primenom algoritma *background subtraction*
  2. *Thresholding* u cilju binarizacije slike
  3. Izdvajanje konture šake uz pronalaženje centra dlana
  4. Izdvajanje *bounding box*-a sa šakom
  5. Promena dimenzija slike
  6. Primena pretprocesiranja specifičnog za mrežu
- Korištena je OpenCV biblioteka.



## PREPOZNAVANJE AKCIJE

Za prepoznavanje akcije upoređene su dve CNN – MobileNetV3Small i NASNetMobile.

Korištene su implementacije iz *Python* biblioteke *Keras*. Mreže su izabrane jer su relativno *lightweight* (u poređenju sa npr. VGG i Inception mrežama), stoga se očekuje da postignu dobre rezultate za binarne slike u realnom vremenu.

Prilikom treniranja mreže korišteni su *checkpointing* i *early stopping*. Za *optimizer* korišten je *Adam*. Arhitektura je dopunjena sa klasifikacionim *softmax* slojem od šest neurona. Za početne težine uzete su *ImageNet* težine.

Frejm se šalje na svakih 10 frejmova, tj. šalje se po tri frejma u sekundi. Ulaz u mrežu je *threshold* slika.

Izlaz iz mreže predstavlja stanje, odn. jednu od šest prethodno definisanih akcija.

## MOGUĆA PROŠIRENJA

Moguće je unaprediti robustnost sistema u odnosu na pozadinu i osvetljenje. Glavno ograničenje aplikacije jeste što ne postoji izdvajanje šake od ruke – prilikom korišćenja aplikacije potrebno je nositi majicu dugih rukava (koja podseća na pozadinu). Višenitne mreže bi poboljšale odziv sistema. Rešenje može biti polazna osnova za rešavanje srodnih problema.

## UPRAVLJANJE MIŠEM

Akcija se translira na Photoshop primenom *pyautogui*, *photoshop-python-api* i *mouse biblioteka*.

Prilikom prve detekcije šake pamti se pozicija – u svakom narednom frejmu se posmatra relativni pomeraj na osnovu koga se vrši pomeranje kursora. Pomoću *pyautogui keyboard press*-ova se sprovode akcije u Photoshop-u jednom na svaku promenu stanja.

## REZULTATI

Odziv sistema je testiran empirijski i može se reći da radi zadovoljavajuće, odn. da ne postoje opterećujući prekidi i kašnjenja koji bi narušili interakciju korisnika i sistema.

Glavna korištena metrika za evaluaciju neuronske mreže je accuracy.

Testirane su mreže NASNetMobile i MobileNetV3Small, kao dobri kandidati zbog malog broja parametara.

	Validation Accuracy	Test Accuracy
NASNetMobile	97.78%	96.67%
MobileNetV3	100	99

Na prvi pogled ne postoje značajna razlike između ovih mreža, ali ispostavilo se da se MobileNetV3 mnogo brže trenira nego NASNetMobile – 10 min po epohi naspram 30 min. MobileNetV3 se i tokom rada aplikacije pokazala kao responzivnija mreža, što verovatno ima veze sa manjim brojem parametara (1.8 miliona naspram 4.6).

Zbog navedenih razloga kao finalna mreža odabrana je MobileNetV3Small mreža.