

## Uvod

Primarni problem ovog projekta je prepoznavanje različitih naredbi koje saobraćajni policajac upućuje učesnicima u saobraćaju telesnim gestovima, sa video snimka, u realnom vremenu. Vrš se prepoznavanje 11 različitih naredbi saobraćajnog policajca, kao i stanja u kom saobraćajac ne pokazuje ni jedan znak.

Naredbe koje se prepoznaju nisu samo stacionarne poze, nego uključuju i pokrete (na primer naredba *uspori brzinu* zahteva pomera desne ruke gore dole).

## Metodologija

Rešavanje problema se vrši kroz dva dela:

1. Prepoznavanje poze (ili skeleta) osobe na snimku. Za ovaj deo korišćena je gotova PoseNet konvoluciona neuronska mreža.
2. Prepoznavanje različitih naredbi na osnovu skeleta dobijenog iz prethodnog dela. Izlaz iz PoseNet mreže pretvara se u feature vektor, koji je ulaz u LSTM neuronsku mrežu koja za izlaz ima verovatnoće pojavljivanja svake poze za svaki frejm video snimka.

Treniranje LSTM mreže je radjeno nad feature vektorom izdvojenom iz skeleta osobe koju prepoznaje PoseNet za svaki frejm snimka.

Poza sa najvećom verovatnoćom pojavljivanja predstavlja predikciju LSTM mreže.

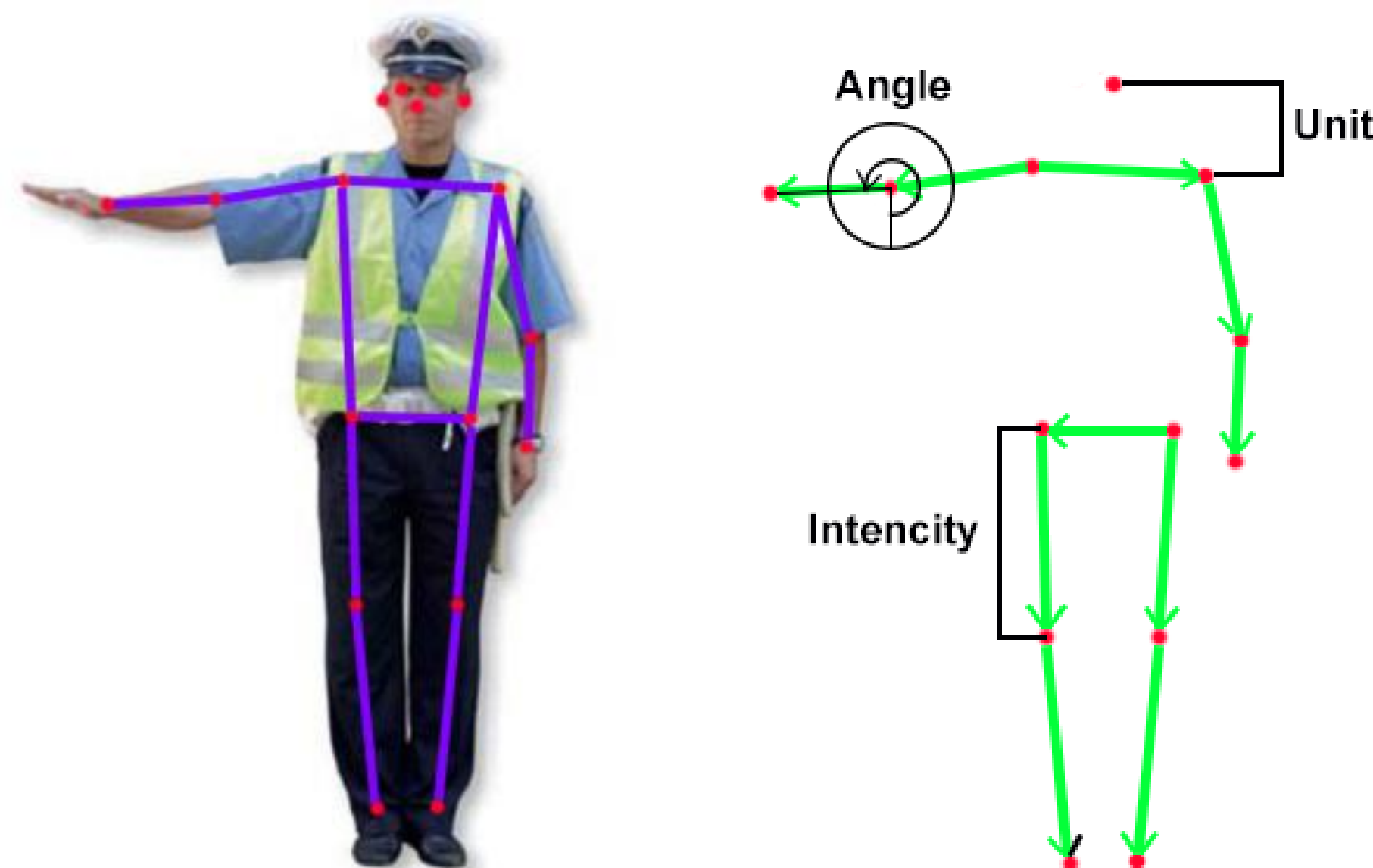


Metode dolaska do predikcije pokazane naredbe

## Skup podataka

Skup podataka se sastoji od nekoliko snimaka osobe kako pokazuje različite saobraćajne znake sa ukupnim trajanjem od oko 25 minuta. Pokazani saobraćajni znaci su labelirani sa zakašnjenjem od 750ms u odnosu na stvarni početak pokazivanja.

Podaci su podeljeni na trening i test skup u odnosu 80 : 20.



Ekstrakcija feature vektora iz skeleta

## Ekstrakcija feature-a

Feature vektor je reprezentacija dvodimenzionalnih vektora uzetih, na način ilustrovan na slici, iz tačaka dobijenih pomoću PoseNet neuronske mreže.

Za svaki dvodimenzionalni vektor izdvajaju se dve vrednosti koje ulaze u feature vektor.

Prva vrednost je njegov intenzitet podeljen sa intenzitetom jedinice (Unit). Jedinica predstavlja vertikalno rastojanje izmedju tačke nosa i jednog ramena.

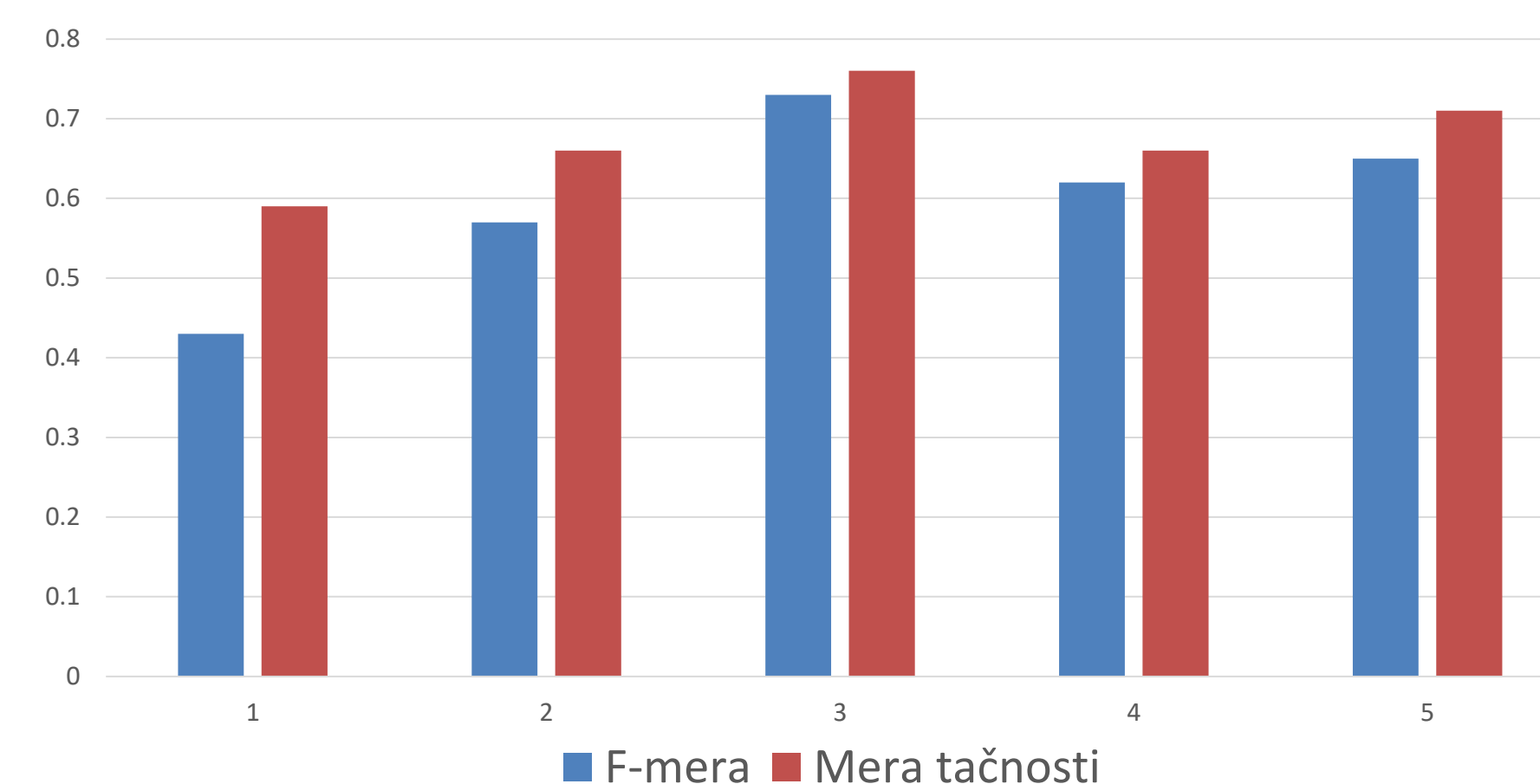
Druga vrednost je ugao vektora u odnosu na vektor (0, -1).

## Rezultati

Trenirane su različite arhitekture neuronske mreže:

1. jedan LSTM sloj sa 32 jedinice, N = 300
2. jedan LSTM sloj sa 256 jedinica, N = 300
3. jedan LSTM sloj sa 256 jedinica, N = 900
4. jedan LSTM sloj sa 128 jedinica i jedan LSTM sloj sa 64 jedinice, N = 900
5. dva LSTM sloja sa 128 jedinica i jedan LSTM sloj sa 64 jedinice, N = 900

(N - broj uzastopnih frejmova u jednom batchu treniranja)



## Zaključak

Mreža teško razlikuje slične poze, dok poze sa jedinstvenim karakteristikama lako razaznaje. Ovo može biti posledica premalog skupa podataka. Takođe, za labeliranje skupa podataka nije korišćen alat koji omogućava preciznost na nivou frejma, što dovodi do nekonzistentnosti u skupu podataka.

Dalja unapređenja ovog projekta se mogu realizovati prvenstveno kroz proširivanje, veću raznovrsnost i preciznije labeliranje skupa podataka. Takođe je moguće još optimizovati korišćene hiperparametre ili isprobati i drugačije arhitekture mreže za prepoznavanje skeleta.