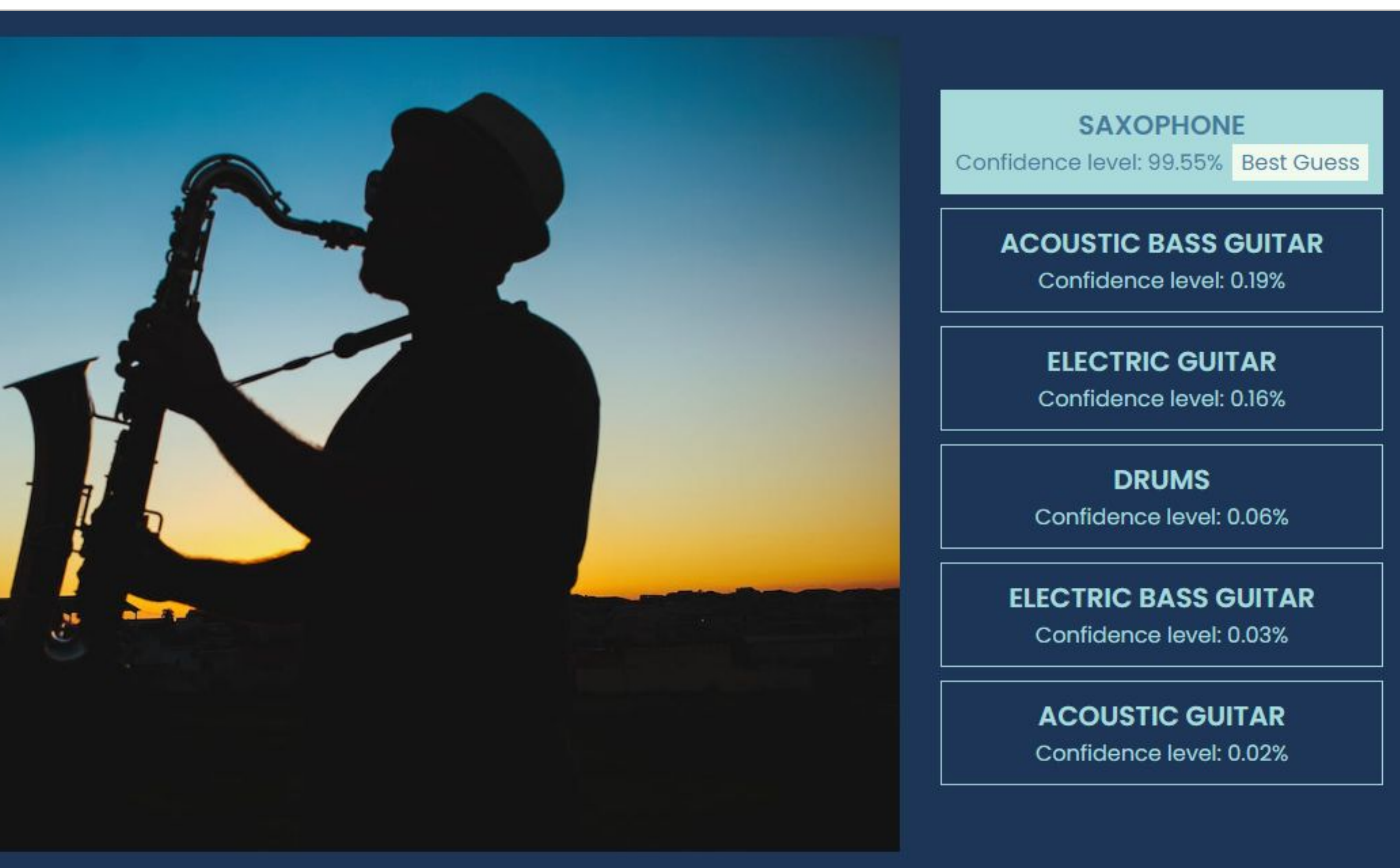


Student
Ilija Kalinić
SW65/2019

# Klasifikacija muzičkih instrumenata

Višeklasna klasifikacija realnih fotografija muzičkih instrumenata uz pomoć dubokih konvolucionih neuronskih mreža

Asistent
Dragan Vidaković



## 1. Motivacija

Duboke neuronske mreže su jedno od boljih rešenja za problem klasifikacije slika. Većina ljudi bi se složilo da je kategorizovati i klasifikovati sliku po tome šta se pojavljuje na njoj lak zadatak, sve dok je broj slika relativno mali. Međutim, od nekoliko stotina do nekoliko hiljada slika se napravi na većini koncerata i događaja sa živom muzikom, koje posle toga, fotograf mora ručno da sortira. Kao odgovor na ovaj scenario, napravljene i upoređene su dve neuronske mreže koje bi pokušale rešiti ovaj problem.

## 2. Skup podataka

Skup podataka je ručno prikupljen upotrebom *Python* skripte za “*skrejpovanje*” slika sa interneta. Skup podataka se sastoji od ukupno 1289 slika raspoređenih u 6 klasa (oko 214 slika po klasi):

- Akustična bas gitara
- Akustična gitara
- Bubnjevi
- Električna bas gitara
- Električna gitara
- Saksofon

Skup podataka sadrži slike sa raznim nivoima osvetljenja, boje (crno-bele i u boji), šuma, kvaliteta i orijentacije.

## 3. Metodologija

Metodologija izrade projekta je sledeća:

- Prikupljanje, augmentacija i preprocesiranje skupa podataka
- Konstrukcija i treniranje ručno pravljenog modela (Instru-vision-cnn) na bazi GoogleNet-a
- *Feature extraction, fine-tuning* i dodatno treniranje MoblieNetV3 mreže trenirane na ImageNet skupu podataka
- Čuvanje istorije treniranja
- Čuvanje modela i priprema za izvoz u demo web aplikaciju
- Demo web aplikacija

## 4. Rezultati

Rezultati treniranja su dobijeni nakon podele skupa podataka na train i test skup 80/20 respektivno i treniranja oba modela kroz 150 epoha:

- Instru-vision-cnn:
  - accuracy: 80.16%
  - loss: 0.7207
- MobileNetV3:
  - accuracy: 87.55%
  - loss: 0.338

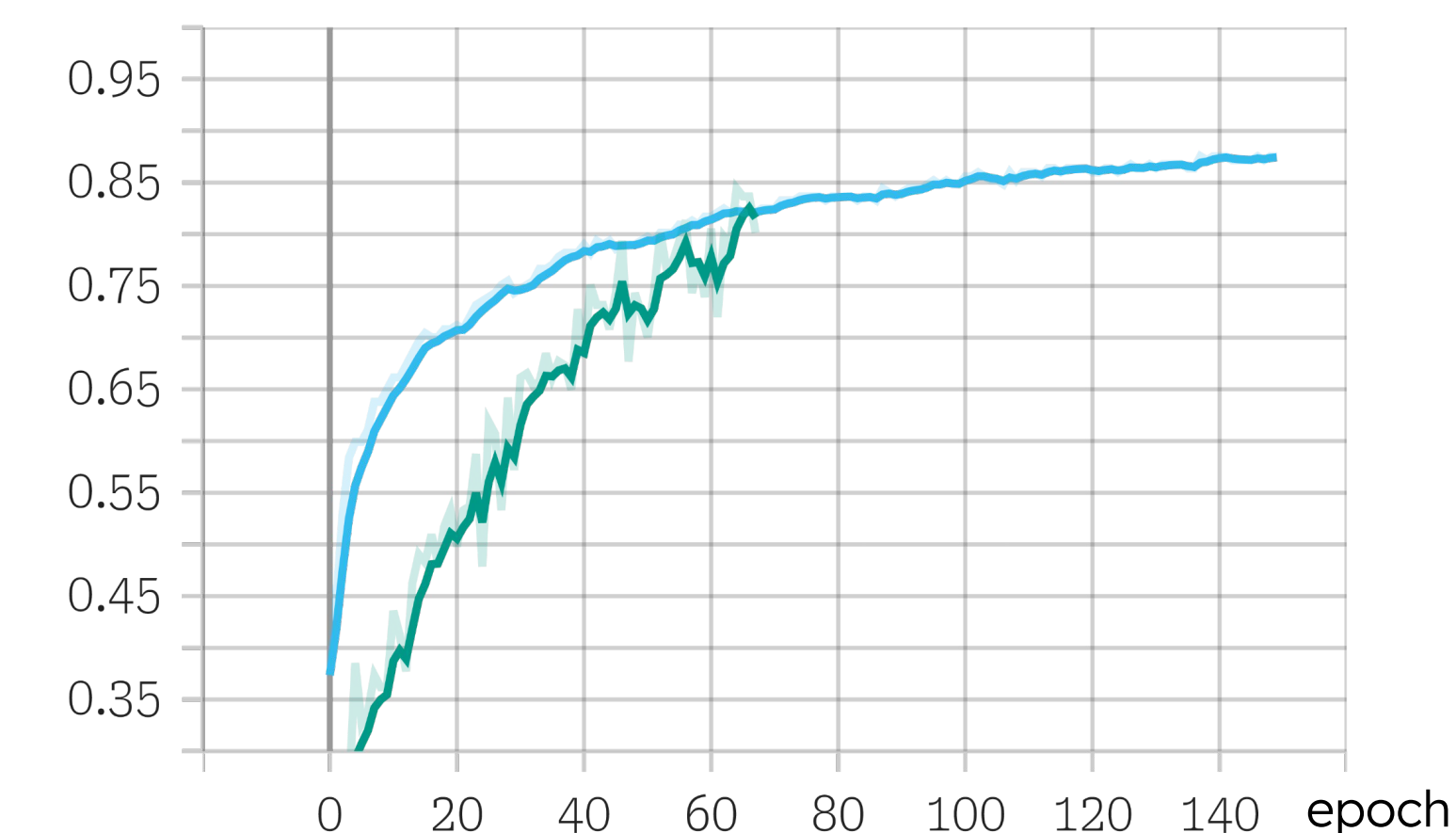
Tokom treniranja, ručno pravljen model je bio znatno sporiji, i u jednom trenutku je počeo da divergira, zbog čega je pokrenut ran prekid testiranja. Od zadatih 150 epoha, ručno pravljen model je uspešno prošao kroz 67, dok je MobileNetV3 prošao kroz svih 150.

Nakon evaluacije oba modela se MobileNetV3 pokazuje stabilnijim i tačnijim za veći broj primera od ručno treniranog.

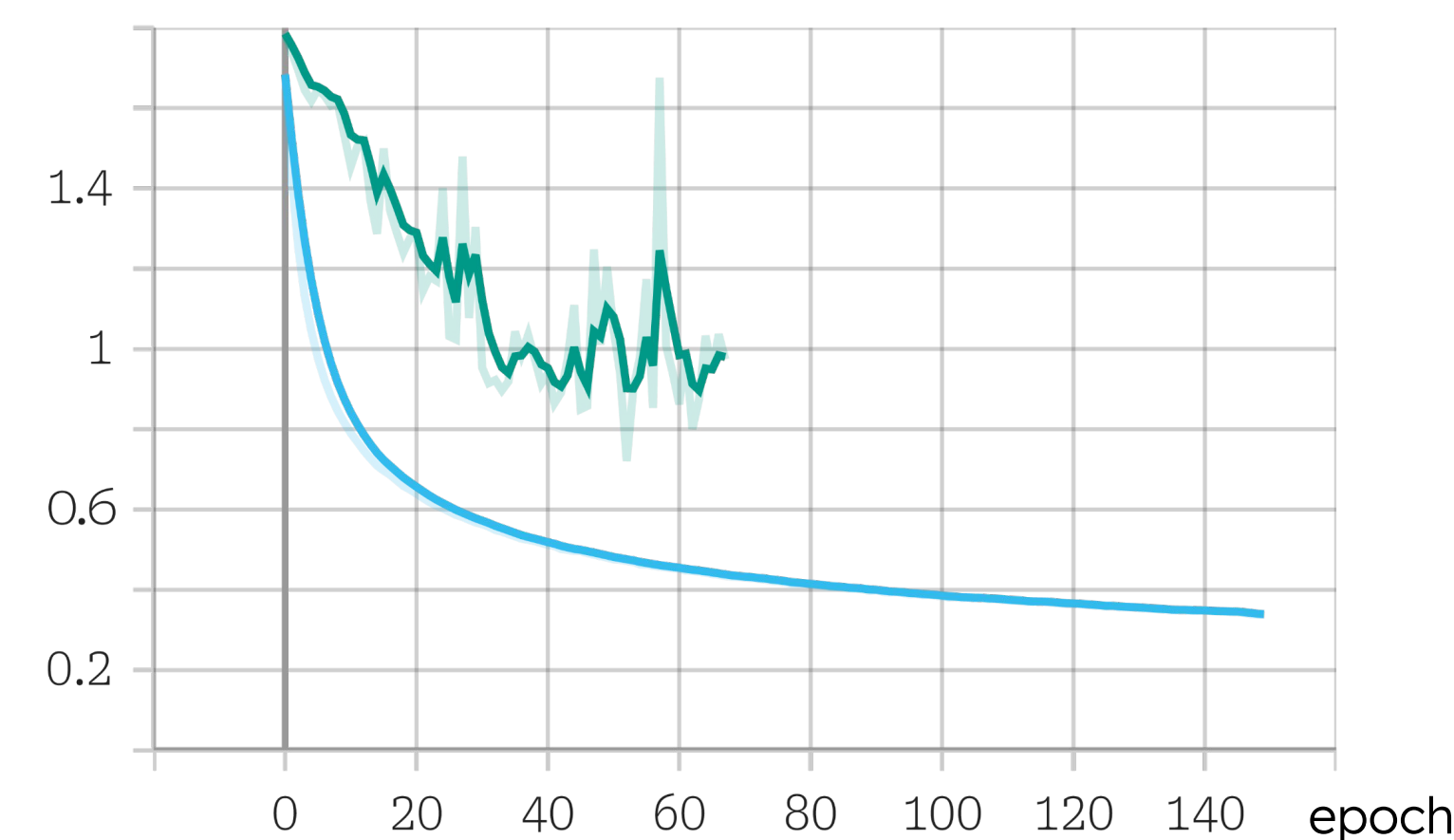
## 5. Analiza

Ručno pravljen model ima mnogo više parametara za treniranje od MobileNetV3 modela. Iz tog razloga je bio mnogo sporiji za treniranje, trošio je mnogo više resursa, što dok se trenira ili evaluira, do toga koliko memorije zauzima kada se sačuva na disk. Iz priloženih grafika za tačnost i *loss* se lako može videti koliko je MobileNetV3 stabilniji u toku treniranja. Ova činjenica se jasno preslikava na razlike u performansama kod učitavanja i evaluacije oba modela koju možete sami videti u [demo-u](#).

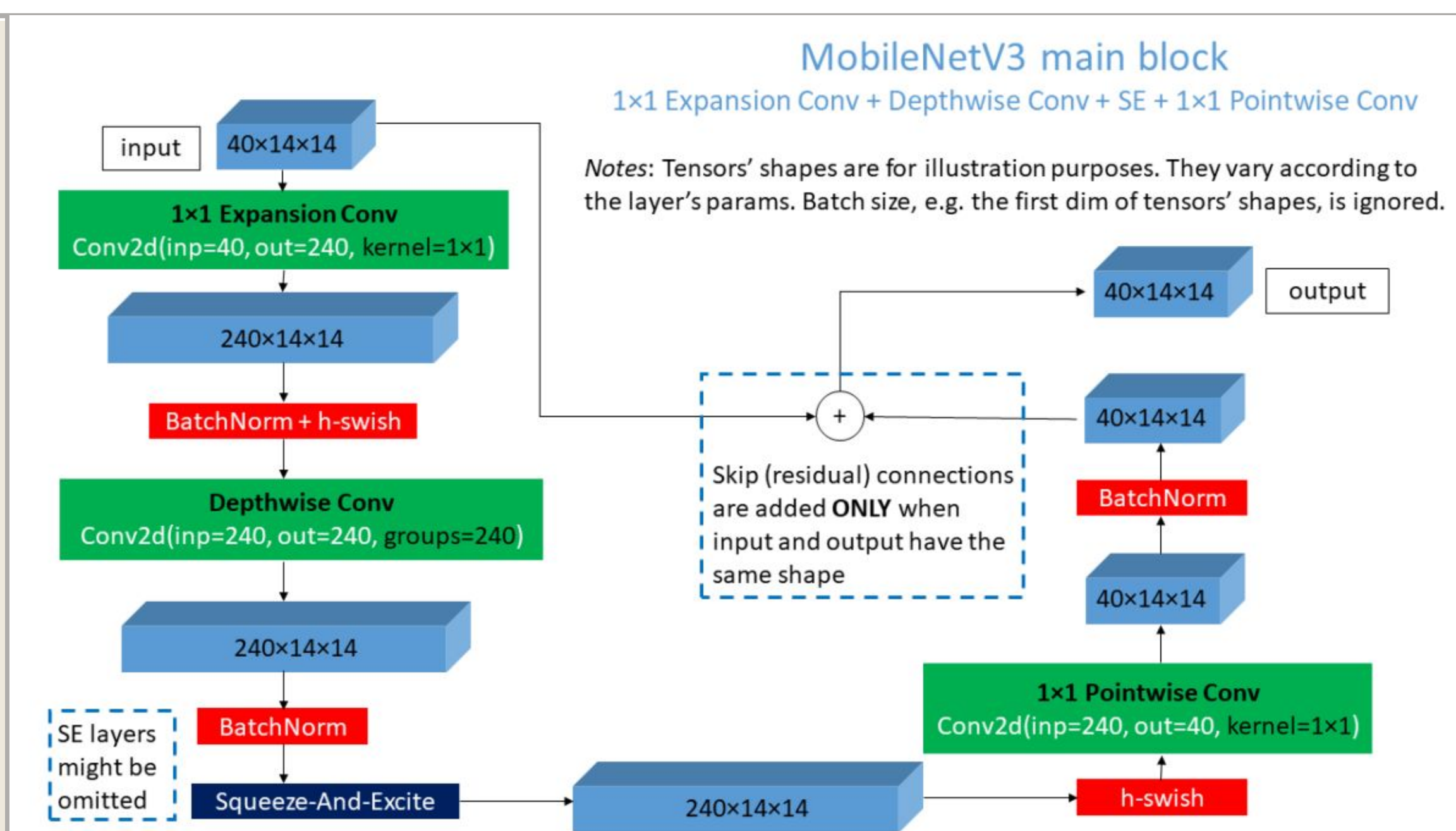
validation accuracy



validation loss

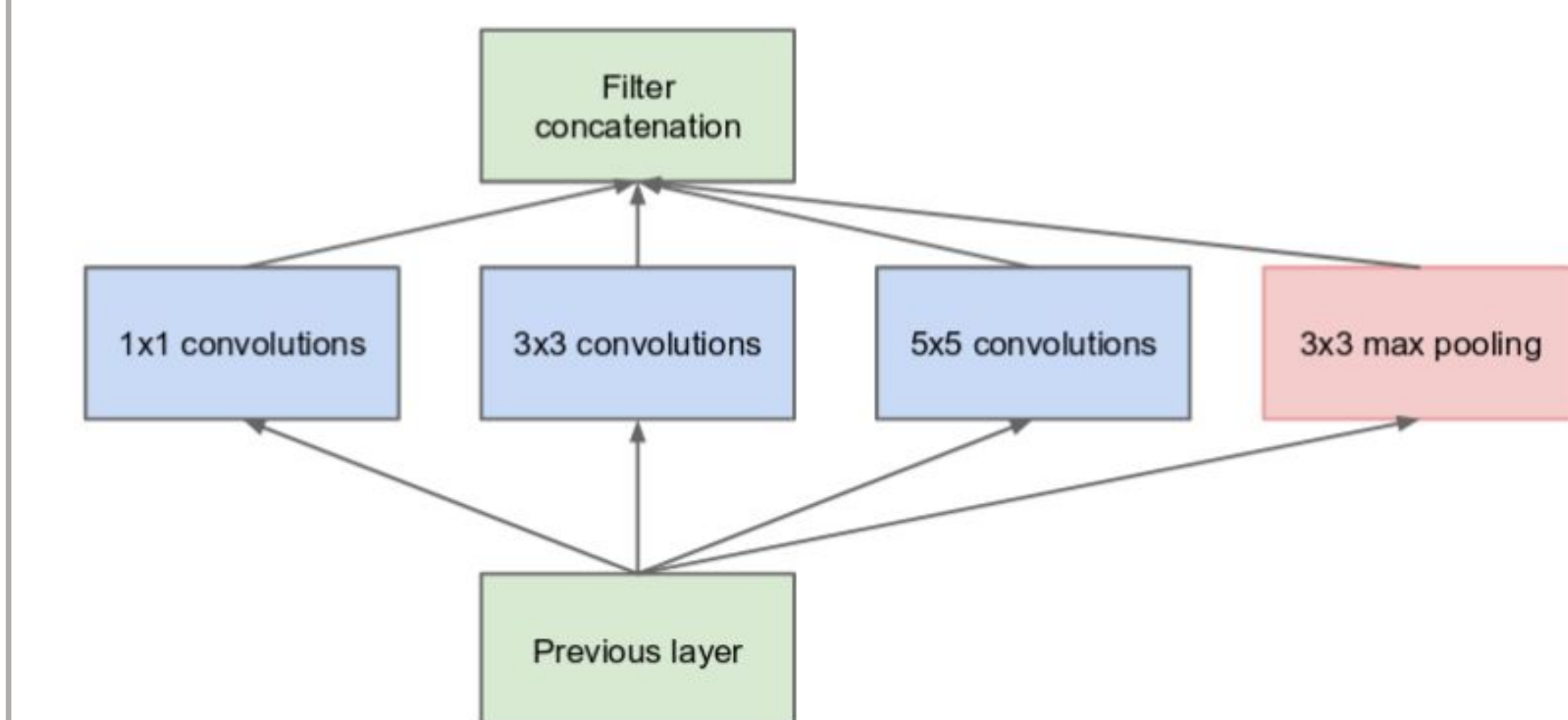


- instru-vision-cnn
- fine tuned MobileNetV3



## 6. Dalji rad

- Proširiti detekcijom objekata, da se prepozna da li uopšte postoji nešto od 6 klasa koje se klasifikuju na datoj slici.
- Proširiti skup podataka povećanjem broja klasa.
- Proširiti skup podataka povećanjem broja pojava u klasi.
- Detekcija instrumenata i klasifikacija u realnom vremenu.
- Unaprediti rezultate oba modela nad istim skupom podataka



(a) Inception module, naïve version