

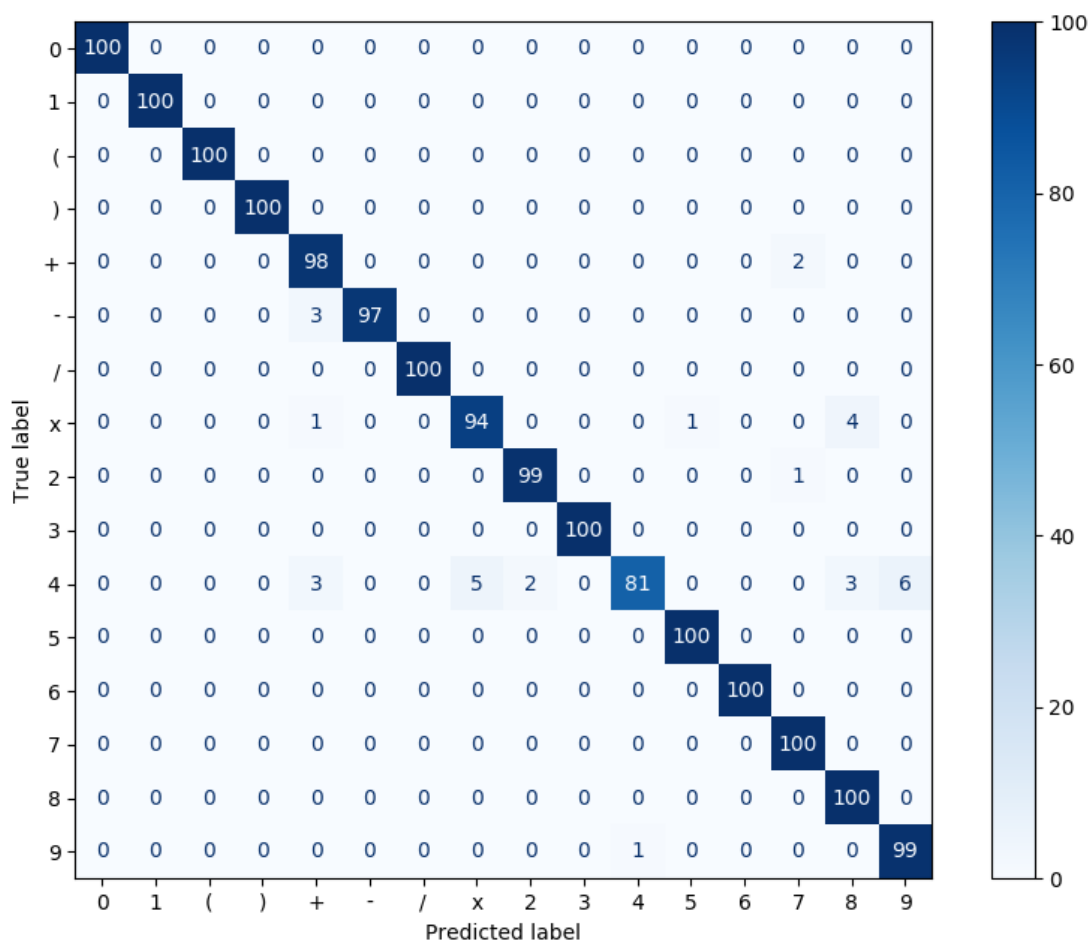
Komentari na metrike:

Daniel Bihar, 7.12.2021.

Glavne metrike koje sam koristio tijekom treninga za određivanje broja epoha, broja slojeva su „Loss” i „Accuracy”. Accuracy će doseći svoj plato, nakon kojega bi trebalo prekinuti treniranje (stati s epohama) da ne uđemo u „overtraining”. U isto vrijeme „Loss” pada i polako dosegne minimalni plato za naš dataset, optimizer i vrstu modela.

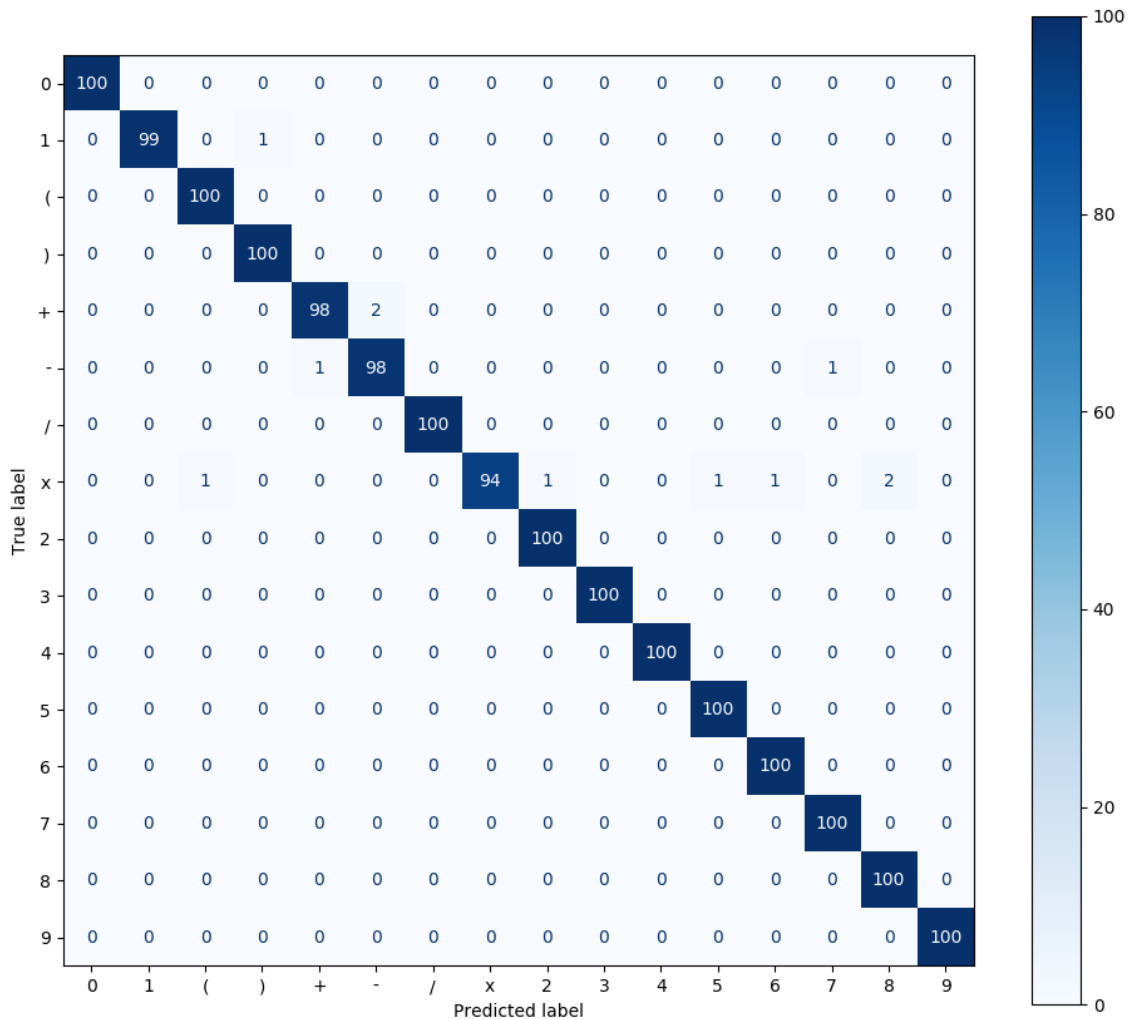
Druga bitnija metrika koju koristimo je „confusion matrix”. Ona će nam dati najbolju informaciju o tome koje klase model zna zamjenjivati, gdje gubi na preciznosti i koji dio dataseta treba razmotriti, promijeniti, pojačati...

Mnogo se stvari još na modelu može poboljšati, a također bi bilo pametno da se „validation set” napraviti sa slika s mobitela. Trenutno sam i set za treniranje i set za validaciju radio ručno na tabletu za crtanje u gimp-u. Takvi setovi nemaju nered u slici koje prave slike imaju, no za takav pristup potrebno je više vremena.



Slika(1) jedna iteracija mreže (njen „confusion matrix”)

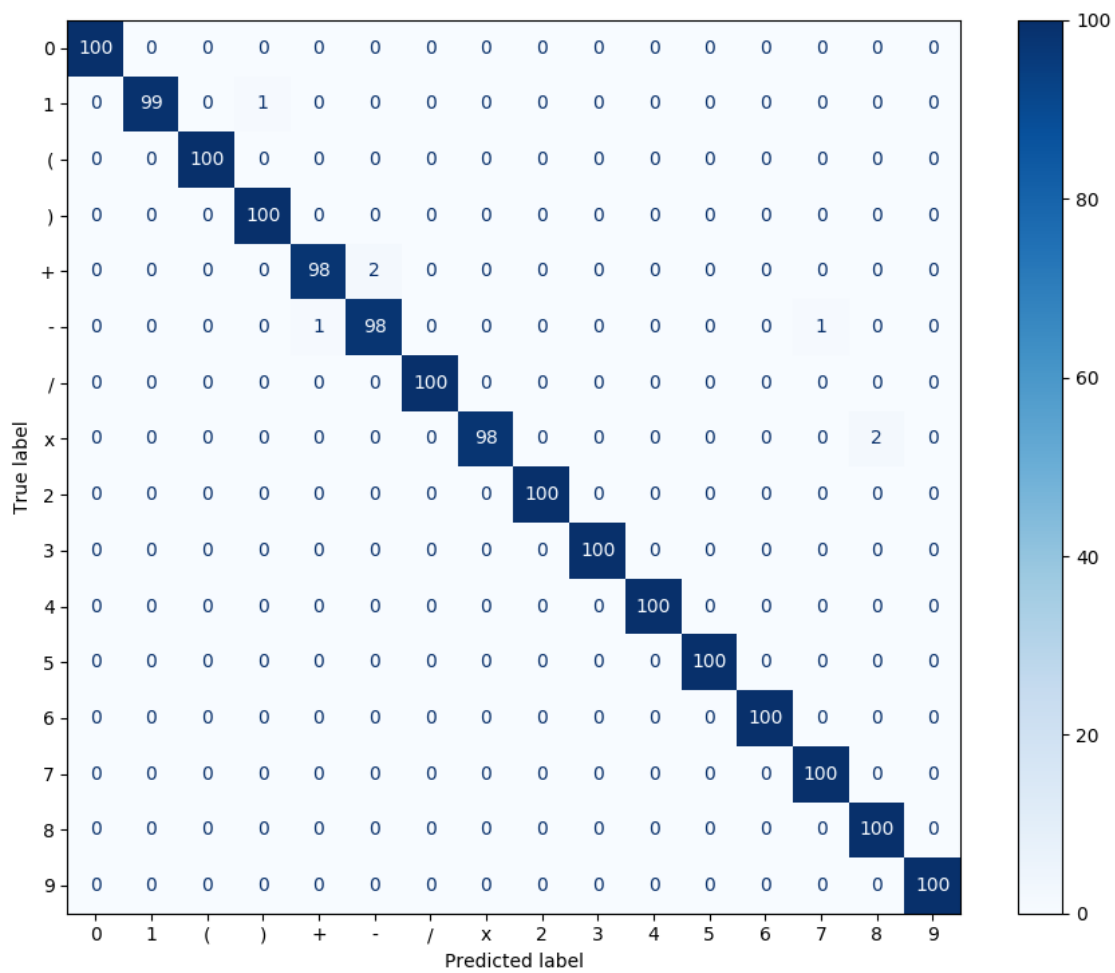
Dodajući dodatne slike za broj "4", mreža se značajno popravila. Također promatrajući sliku (2) vidimo da problemi ostaju kod znaka "x", no pažljivim promatranjem možemo vidjeti da je „validation set” bio sadržavao vrlo neuredne zapise znaka "x", te se mreža zapravo ponaša bolje nego metrika nalaže (u realnosti kako se može provjeriti s GUI aplikacijama tipa „draw.py”).



Slika(2) daljnja iteracija mreže (njen „confusion matrix”)



Prolaskom kroz „validation dataset” i eliminacijom loših validacijskih slika za x, kao što je dobivamo sljedeću matricu kao na Slici(3).



Slika(3) validacija mreže nakon pravilnijeg „validation” seta

Za daljnje poboljšanje mreže potrebno je povećati „validation set” više puta, budući da 100 slika po klasi (znaku) nedovoljno precizno opisu postotak točnosti koji je toliko blizu 100 posto (do na par postotaka). Potrebno bi bilo barem 1000 slika po znaku i to napisanih na papir (realan „input”) da bi ova metrika što bolje predstavljala realnost.

Provjeravajući performance modela na zapisima s olovke (na kompjutoru, što je „low noise” input) te sa slikom („high noise input”) možemo primijetiti približno jednaku uspješnost, iako vjerojatno nešto manju nego što metrika ukazuje. Vjerojatno umjesto 1 krivo prepoznati znak na 100, imamo oko 1 na 30 za polu uredan „input”. Kako zadatak nalaže vrlo vrlo lijep „input” što pretpostavljam da znači, dovoljan kontrast, niti pre tanko niti pre debelo slovo te dovoljno mjesta između slova tako da svako slovo ima svoju „kućicu”, rezultati bi trebali biti sličniji Slici(3). Također se možemo složiti s interpretacijom Slike(3) da model najčešće radi krive klasifikacije i zamjene kod znakova „+”, „-”, „1”. Naime najčešće, „1” zna prepoznati kao „)”, te zna zamjenjivati „+”, „-”. Ostale greške u realnosti nisu prikazane Slikom(3) zbog spomenutog manjka u „validation set”-u.