Implementacija modela za prepoznavanje broja prstiju na ruci pomoću konvolucijskih neuronskih mreža

Krešimir Špehar

Fakultet informatike, Sveučilište Jurja Dobrile u Puli

10. svibnja 2023.

Sažetak

Cilj ovog projekta je razviti model za prepoznavanje broja prstiju na ruci pomoću konvolucijskih neuronskih mreža (CNN) koji će koristiti vlastiti izvor slika ruku kao ulazne podatke. Ovaj personalizirani pristup omogućit će stvaranje modela koji je prilagođen ulaznim podacima i koji može poboljšati kvalitetu klasifikacije.

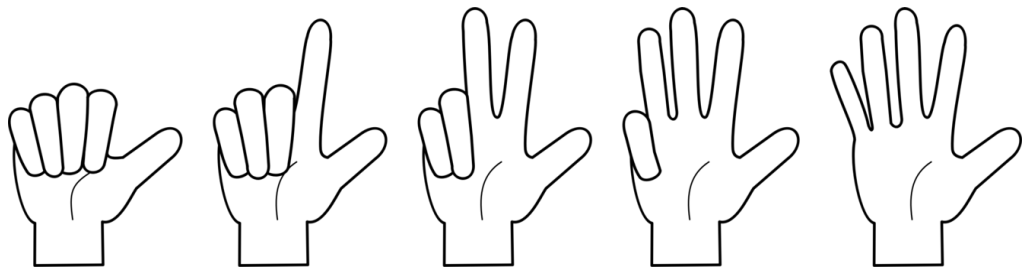
Kroz primjenu arhitekture CNN, model će naučiti razlikovati različite brojeve prstiju na ruci temeljeno na vizualnim reprezentacijama slika koje će biti prikupljene iz vlastitog izvora. Tijekom razvoja, trenirat će se i optimizirati CNN kako bi se postigli što bolji rezultati klasifikacije.

Potencijalne primjene ovakvog modela uključuju različite tehnološke primjene poput upravljanja uređajima pomoću gesta ruke, interaktivne igre, ili automatizirano prebrojavanje prstiju u medicinskim ili sportskim aplikacijama.

Uvod

Današnje digitalno doba sa svojim kontinuiranim napretkom tehnologije nudi višestruke mogućnosti za primjenu umjetne inteligencije i dubokog učenja. Prepoznavanje i tumačenje gesta ruku jedno je od područja koje je postiglo značajan napredak i privlači sve veću i veću pozornost. Geste ruku predstavljaju jedan od načina komunikacije i interakcije čovjeka s računalima i drugim uređajima. U ovom radu fokus je bio na specifičnom problemu, tj. utvrđivanju koliko je prstiju podignuto na slici.

Tema ovog rada predstavlja izazovno i zanimljivo područje umjetne inteligencije. Glavni cilj ovog istraživanja je razviti i implementirati model dubokog učenja koji će točno prepoznati broj podignutih prstiju na slici i klasificirati ih u pet mogućih kategorija: jedan prst, dva prsta, tri prsta, četiri prsta i pet prstiju. Ovaj projekt temeljit će se na korištenju konvolucijskih neuronskih mreža (CNN) kao glavnog alata za rješavanje problema prepoznavanja broja podignutih prstiju na slici. CNN, kao snažan model dubokog učenja, omogućit će nam analizu i izdvajanje značajki iz slika ruku što je ključno za točnu klasifikaciju različitog broja prstiju.



Slika 1. Primjer različitog broja prstiju na ruci

Motivacija za ovu temu proizlazi iz potrebe da se od nule razvije model dubokog učenja za točno prepoznavanje broja podignutih prstiju na slici. Model ima potencijal za različite primjene, uključujući poboljšanje komunikacije za osobe s posebnim potrebama, optimizaciju upravljanja uređajima poput pametnih telefona i televizora putem gesti. Također postoji potencijal za korištenje u medicinskim aplikacijama te se može koristiti za analizu kretanja sportaša, pružajući alate za praćenje i analizu u sportskom okruženju. Precizno prepoznavanje broja prstiju igra ključnu ulogu u

postizanju ovih ciljeva, poboljšavajući korisničko iskustvo i donoseći stvarne koristi različitim područjima.

Rad je strukturiran na sljedeći način. U drugom poglavlju je pregled postojećih rješenja u području prepoznavanja i klasifikacije objekata pomoću neuronskih mreža. U trećem poglavlju detaljno je opisana korištena metodologija, uključujući prikupljanje i obradu podataka, arhitekturu konvolucijske neuronske mreže i primijenjene tehnike. U poglavlju četiri predstavljeni su rezultati i analiza samog modela. Na kraju u petom poglavlju, sažima se cijeli rad sa zaključkom.

2. Postojeći modeli

U svrhu ovog istraživanja pregledani i analizirani su relevantni postojeći modeli koji se primjenjuju u sličnim kontekstima. Opisani radovi pružaju važan okvir i uvid u dosadašnje pristupe prepoznavanja i klasifikacija objekata koristeći neuronske mreže što pomaže usmjeriti metodologiju i napredak koji se može postići u istraživanju.

Eid i Schwenker su predstavili rad ove godine [1] koji je postigao značajne rezultate na prepoznavanje gesta ruku pomoću arhitektura CNN mreže koja se sastoji od sedam slojeva. Autori su primjenili tehnike kao što su proširenja podataka (eng. data augmentation) i segmentacija kože kako bi poboljšali točnost modela. Na javnim mjerilima, dva skupa podataka su klasificirana gotovo pa savršeno s iznimno visokim točnostima testiranja od 96,5% i 96,57%. Ovi rezultati sugeriraju učinkovitost CNN pristupa za prepoznavanje gesta rukom i pružaju korisne smjernica za daljni rad u ovom istraživanju.

Istraživanje napisano od strane Islam i sur. [2] opisuje model koji je predložen za prepoznavanje statičkih gesta rukom primjenom CNN-a. Jedan od problema je taj da geste rukom varijaju u orijentaciji i obliku među pojedincima što uvodi nelinearnost te su zbog toga autori koristili neuronske mreže. Kako bi poboljšali performanse modela, primjenjivali su tehnike proširenja podataka uključujući skaliranje, zumiranje, rotiaciju itd. Model je treniran na broju od 8000 slika i testiran na 1600 slika koje su podjeljene u 10 klasa. Rezultat ovog pristupa je postizanje točnosti od 97,12% što predstavlja značajan napredak u usporedbi s modelom bez proširenja podataka koji je ostvario točnost od 92,87%. Opisano istraživanje naglašava vrijednost primjene CNN-a za prepoznavanje gesta rukom te ističe korisnost tehnike proširenja podataka za postizanje boljih rezultata modela.

Iduće istraživanje [3] predstavlja precizan i učinkovit okvir dubokog učenja za prepoznavanje statičnih gestikulacija rukama temeljenih na neuronskih mrežama. Ključna komponenta ovog okvira sastoji se od dvije glavne faze: izdvajanje značajki i klasifikacija. Svaka od navedenih faza obuhvaća niz slojeva koji su namješteni kako bi se postigla najbolja moguća preciznost u prepoznavanju pokreta ljudske ruke. Rezultati eksperimenata od strane Mahmoud i sur. jasno pokazuju da predloženi višeslojni CNN okvir pruža puno bolje performanse, koristeći veliku bazu infracrvenih slika ruku. Odabir infracrvenih slika kao izvor podataka ima ključnu ulogu u

„zaobilasku“ problema slabog osvjetljenja, čime se osigurava pouzdana analiza gestikulacija.

Opisani slični radovi pružaju raznovrsne uvide i pristupe za klasifikaciju broja prstiju na ruci što je i tema ovog rada.

3. Metodologija

Iduće poglavlje predstavlja pristup koji je korišten za rješavanje problema prepoznavanja broja podignutih prstiju na ruci s fotografije. Detaljno je opisano kako su prikupljeni podaci, što je s njima napravljeno, struktura modela koja je korištena za učenje te pristup rješavanja problema.

3.1. Podaci

Za potrebe istraživanja, podaci su osobno prikupljeni u obliku slika koje predstavljaju broj prstiju na ruci. Ukupno je prikupljeno 50 slika po svakoj od 5 klasa što znači da je prikupljeno ukupno 250 slika. Slike su organizirane i pohranjene na Google Drive platformi u odgovarajućim mapama za svaku klasu što je olakšalo upravljanjem podacima i njihovo kasnije korištenje u izgradnji modela. Svaka slika ima različite karakteristike poput različitih pozadina, svjetlosnih uvjeta, rotacija i drugih faktora. Slike su fotografirane pomoću mobilnog telefona u visokoj rezoluciji od 3024x4032 piksela (3:4 omjer) u boji. Ova raznolikost karakteristika napravit će model izazovnijim, ali i sposobnijim za bolje prepoznavanje broja prstiju u različitim uvjetima. Pri pripremi slika za učenje modela, prvi koraci uključivali su obradu kako bi se osiguralo da sve slike imaju iste dimenzije širine i visine. U ranim fazama učenja, dimenzije 300x300 piksela bila su odabrane kako bi se slike standardizirale i olakšalo učenje modelu. Međutim u kasnijim iteracijama s modelom koristile su smanjenje dimenzije odnosno 128x128 piksela, što je omogućilo bržu obradu i smanjenje resursa potrebnih za učenje. Kako bi model imao bolje konačne rezultate, prmijenjena je tehnika poznata kao "data augmentation". Ova tehnika omogućila je generiranje dodatnih 150 slika po klasi, što je rezultiralo većim i raznovrsnijim skupom podataka za trening. Data augmentation je napravljena lokalno na slikama, uključujući različite transformacije poput rotacije, zrcaljenja i promjene svjetline kako bi se stvorila veća raznolikost u podacima za trening modela. Prije treninga modela, opisani skup je podijeljen na tri dijela: trening, validacija i test. Omjer je 8:1:1.

3.2. Model

U svrhu rješavanja problema prepoznavanja broja prstiju na ruci, korištene su konvolucijske neuronske mreže. CNN je duboka neuronska arhitektura koja je

napravljena za obradu vizualnih podataka kao što su slike. Model u ovom radu je razvijen od nule koristeći Python programski jezik unutar Google Colab okruženja i TensorFlow okvira za duboko učenje. Posjeduje bitne dijelove modela poput slojeva, funkcije gubitka, optimizacije itd. Kroz trening, model je „napredovao“ tako što je nadograđen kroz svaku iteraciju kako bi se postigli zadovoljavajući rezultati.

4. Trening modela

U ovom poglavlju se detaljno opisuje proces treninga konvolucijske neuronske mreže kroz niz faza. Analizirat će se postupci optimizacije, podešavanje hiperparametara i evalucije rezultata. Cilj je kroz svaku fazu „nadograditi“ model te postići što veću točnost nad testnim podacima.

4.1. Prva faza

Nakon što su slike povezane putem Google Drivea, prva faza treninga je bila na redu.

Slika 2. Gubitak i točnost prve faze treninga

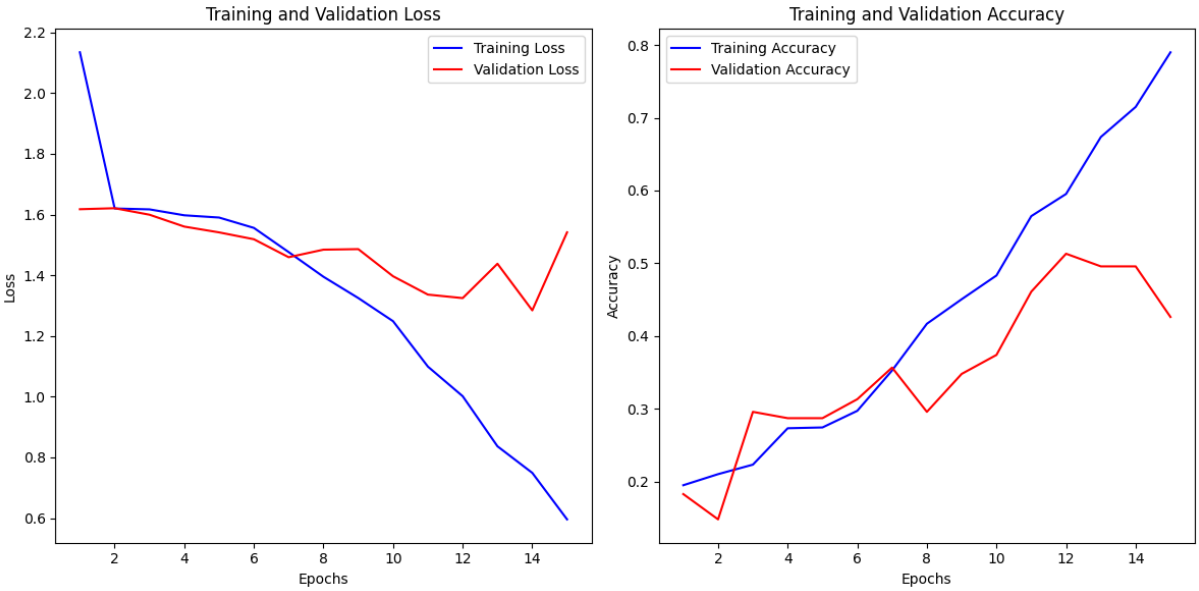
Analizom slike 2 možemo uočiti da je tijekom prve faze treninga model brzo naučio prepoznavati uzorke na trening slikama što znači da je model pretreniran (eng. overfitt). Pretreniranost pokazuje na to da model nije sposoban generalizirati nepoznate podatke. Validacijska točnost je dosta niska, svega 40%. Također su validacijski gubitci sve više i više rasli tijekom treninga. Navedeni prvi dio treninga je odrađen na 13 epoha uz batch size od veličine 128. Slike su bile u formatu 300x300. Model na prvoj fazi ima 43,675,333 parametara.

Model prve faze treninga:

1) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
2) *Sloj udruživanja 2x2 (max pooling)*
3) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
4) *Sloj udruživanja 2x2 (max pooling)*
5) *Potpuno umrežen sloj sa 128 neurona, aktivacijska funkcija 'relu'*
6) *Potpuno umrežen sloj sa 5 neurona (5 klasa), aktivacijska funkcija 'softmax'*

4.2. Druga faza

U drugoj fazi treninga možemo vidjeti napredak u usporedbi s prvom fazom.



Slika 3. Gubitak i točnost druge faze treninga

Točnost modela za validaciju se povećala na 51%, ali unatoč ovom poboljšanju treba spomenuti da model opet ide prema pretreniranju budući da je točnost na skupu treninga i dalje visoka te raste kroz epohe. Ukupno je model prošao kroz 15 trening epoha i korišten je smanjeni batch size veličine 32. Slike su i dalje u formatu 300x300.

Model druge faze treninga:

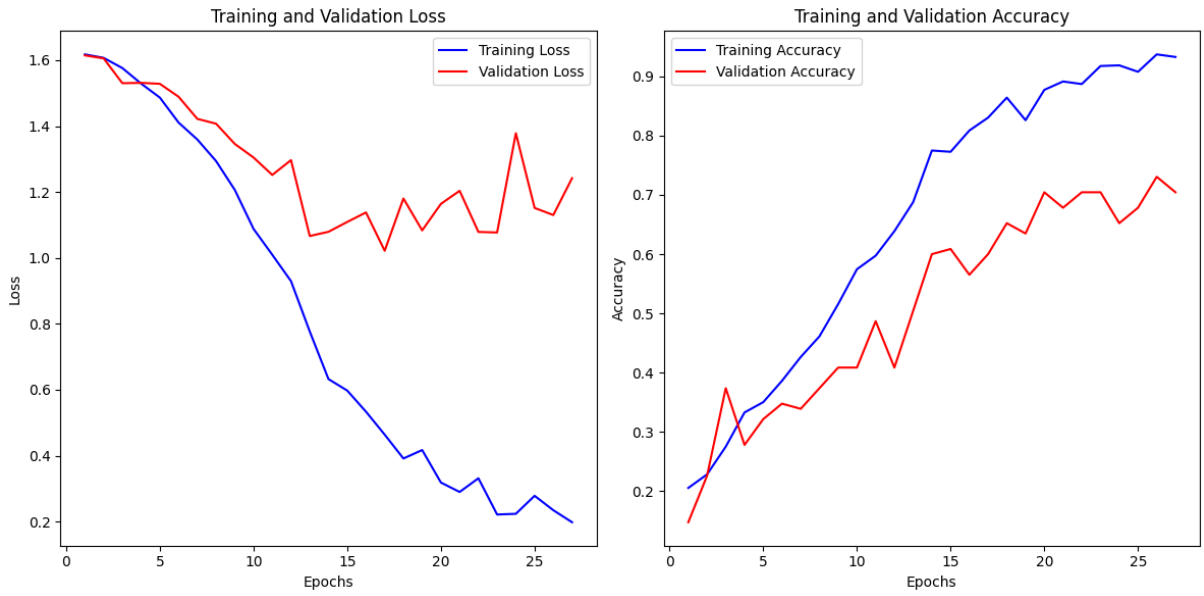
1) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
2) *Sloj udruživanja 2x2 (max pooling)*
3) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*

4) *Sloj udruživanja 2x2 (max pooling)*
5) *Konvolucijski sloj sa 128 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
6) *Sloj udruživanja 2x2 (max pooling)*
7) *Konvolucijski sloj sa 256 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
8) *Sloj udruživanja 2x2 (max pooling)*
9) *Potpuno umrežen sloj sa 128 neurona, aktivacijska funkcija 'relu', 50% gašenje neurona 'dropout 0.5'*
10) *Potpuno umrežen sloj sa 5 neurona (5 klasa), aktivacijska funkcija 'softmax'*

Iz opisanog modela možemo vidjeti da je druga faza treninga proširena u odnosu na prvu. Dodani su dodatni konvolucijski slojevi za udruživanje što je rezultiralo većom dubinom modela. Uveden je i dropout kako bi se smanjila pretreniranost. Također na modelu se smanjio broj parametara te sada ima 8,777,797.

4.3. Treća faza

U nastavku treniranja modela, slike su promjenjene sa 300x300 na 128x128 piksela što je dosta ubrzalo proces treninga. U trećoj iteraciji treninga postignut je napredak u točnosti modela. Točnost modela na skupu za validaciju povećala se na 60%.



Slika 4. Gubitak i točnost treće faze treninga

U ovoj fazi treninga je provedeno 30 epoha dok je batch size bio veličine 32. Najznačajnija promjena je smanjenje broja parametara s obzirom na prethodne faze treninga. Broj parametara sada iznosi 352,037.

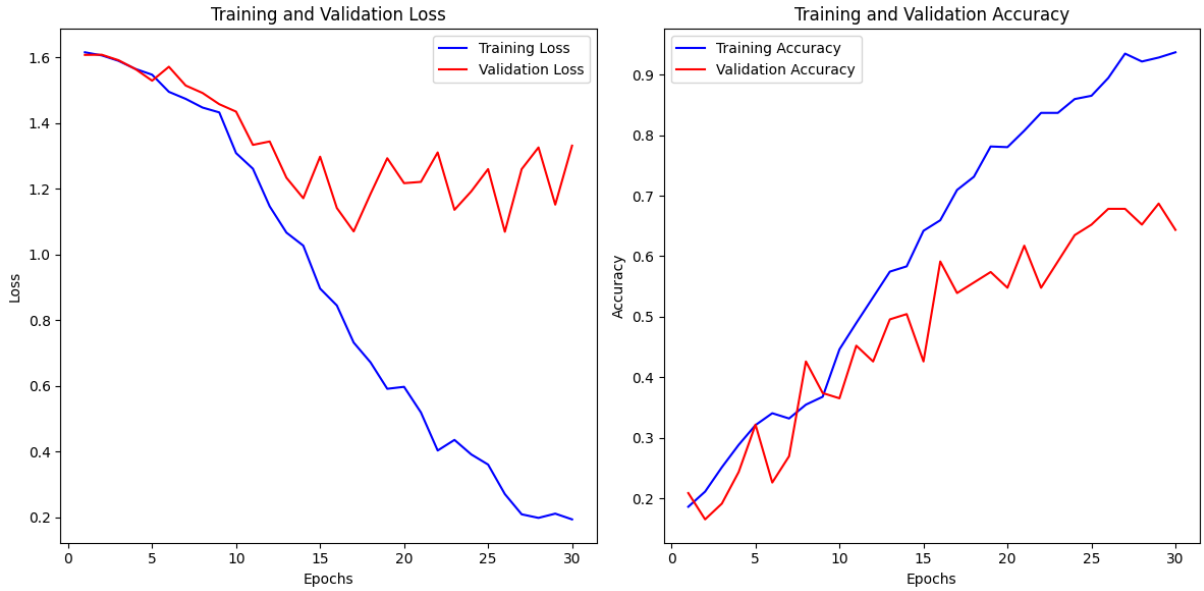
Model treće faze treninga:

1) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
2) *Sloj udruživanja 2x2 (max pooling)*
3) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
4) *Sloj udruživanja 2x2 (max pooling)*
5) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
6) *Sloj udruživanja 2x2 (max pooling)*
7) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
8) *Sloj udruživanja 2x2 (max pooling)*
9) *Sloj za isključivanje, 25% gašenje neurona 'dropout 0.25'*
10) *Potpuno umrežen sloj sa 128 neurona, aktivacijska funkcija 'relu'*
11) *Sloj za isključivanje, 50% gašenje neurona 'dropout 0.5"*
12) *Potpuno umrežen sloj sa 5 neurona (5 klasa), aktivacijska funkcija 'softmax'*

Kroz treću fazu treninga dodan je Adam optimizator sa stopom učenja od 0.001. Definirani su pozivi za rano zaustavljanje (eng. early stopping) koji sprema najbolji model tokom treninga kako bi se spriječila pretreniranost i zadržali najbolji parametri. Postavljen je 'patience' na 10 što znači da model prati rezultate na validacijskom skupu podataka nakon svake epohe tijekom treninga. Ako se tijekom kontinuiranih 10 epoha ne primjeti poboljšanje, model prestaje s treniranjem kako bi se spriječila pretreniranost.

4.4. Četvrta faza treninga

Četvrta faza treninga predstavlja dosad najveći pomak u vidu točnosti modela gdje ona iznosi 65% na validacijskim podacima.



Slika 5. Gubitak i točnost četvrte faze treninga

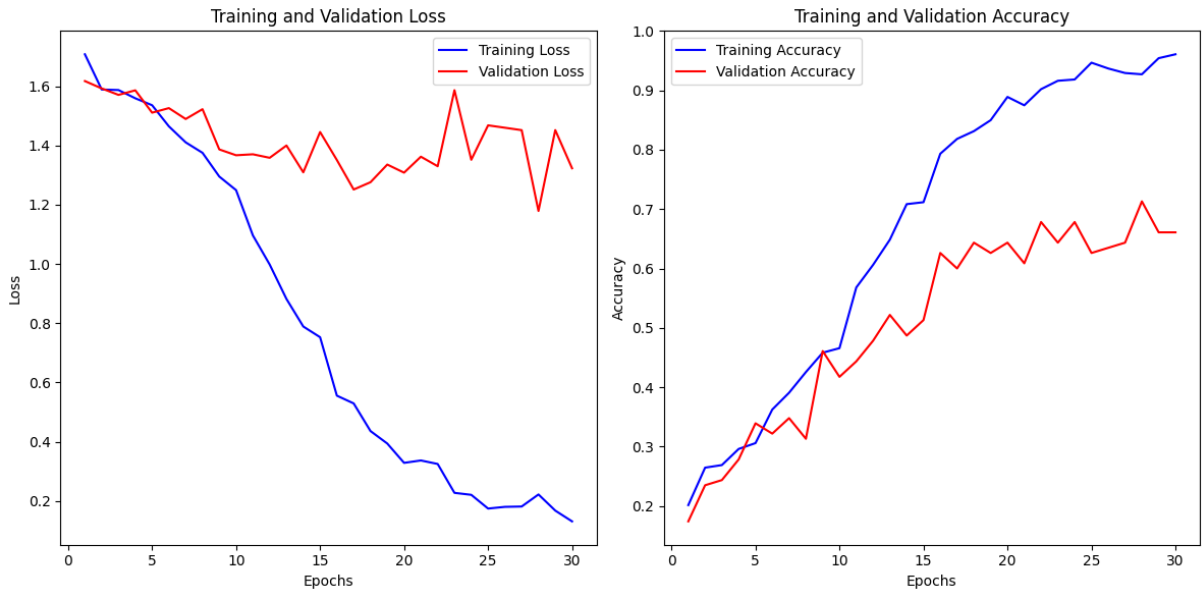
Model četvrte faze treninga:

1) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
2) *Sloj udruživanja 2x2 (max pooling)*
3) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
4) *Sloj udruživanja 2x2 (max pooling)*
5) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
6) *Sloj udruživanja 2x2 (max pooling)*
7) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
8) *Sloj udruživanja 2x2 (max pooling)*
9) *Konvolucijski sloj sa 128 filtera dimenzije 3x3, aktivacijska funkcija 'relu'*
10) *Sloj udruživanja 2x2 (max pooling)*
11) *Sloj za isključivanje, 25% gašenje neurona 'dropout 0.25'*
12) *Potpuno umrežen sloj sa 128 neurona, aktivacijska funkcija 'relu'*
13) *Sloj za isključivanje, 25% gašenje neurona 'dropout 0.25'*
14) *Potpuno umrežen sloj sa 5 neurona (5 klasa), aktivacijska funkcija 'softmax'*

Naspram treće faze treninga dodatno je snižena mreža u vidu parametara, gdje oni sada iznose 205,733. Promjena koja je najviše utjecala na točnost je smanjenje drugog droputa sa 0.5 na 0.25. Također je dodan još jedan konvolucijski sloj kao što je opisano iznad.

4.5. Peta faza treninga

Peta faza koja je i konačna ima najbolje rezultate na validacijskom skupu podataka, točnost je dosegla 70%, dok je vrhunac bio 74%.



Slika 6. Gubitak i točnost pete faze treninga

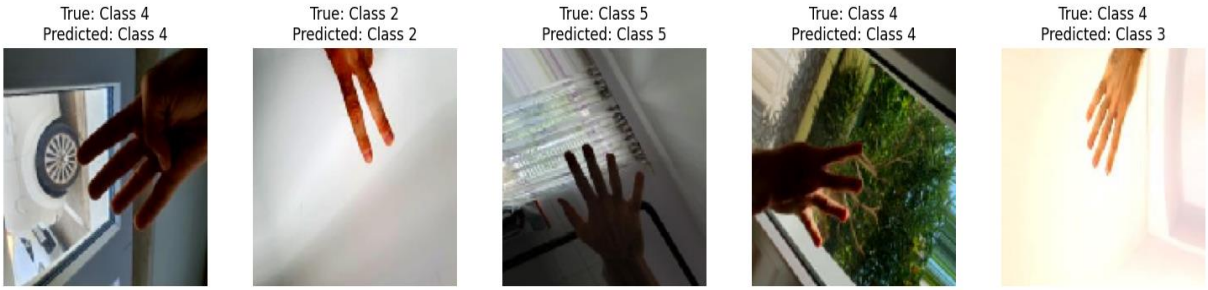
Broj epoha i batch size je ostao isti tijekom zadnje faze treninga. Također je promijenjen prethodno spomenuti 'patience' sa 10 na 15 gdje model može više trenirati nad podacima. Dodana je popularna tehnika „HeNormal“ koja pravilno skalira početne težine (weight) kako bi učenje bilo brže i stabilnije.

Model pete faze treninga:

1) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu', HeNormal*

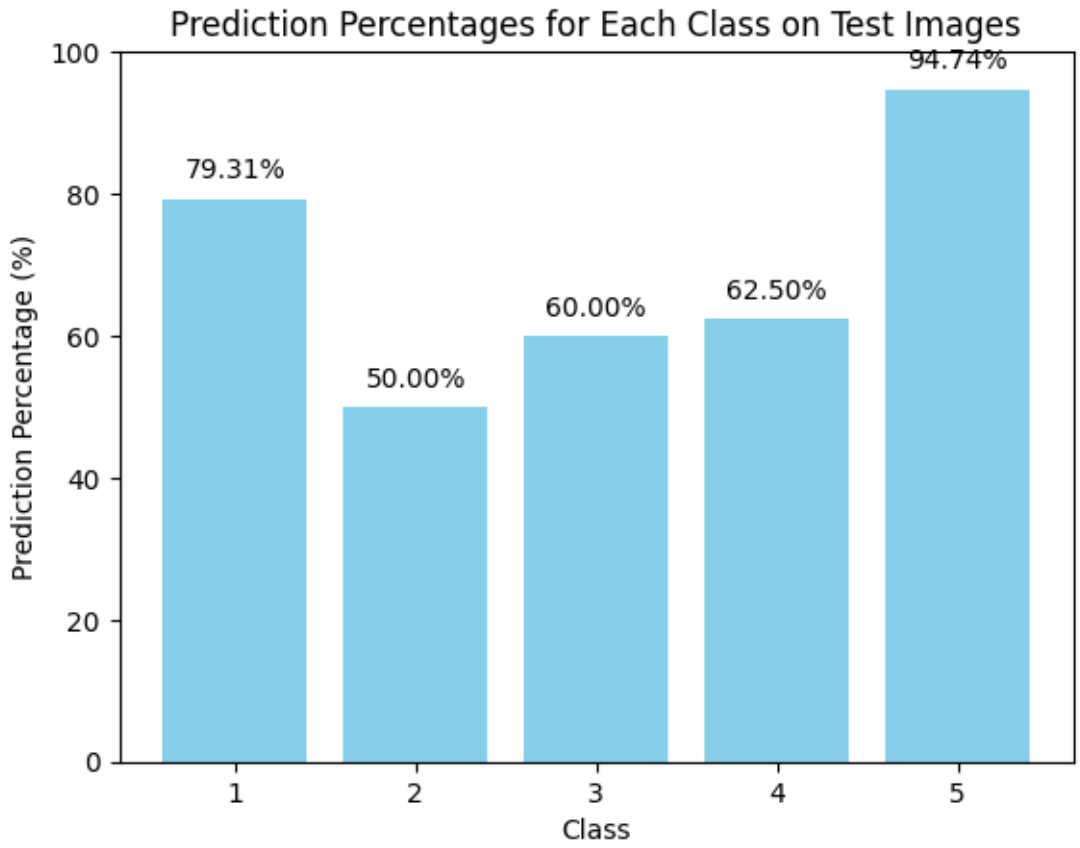
2) *Sloj udruživanja 2x2 (max pooling)*
3) *Konvolucijski sloj sa 32 filtera dimenzije 3x3, aktivacijska funkcija 'relu', HeNormal*
4) *Sloj udruživanja 2x2 (max pooling)*
5) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu', HeNormal*
6) *Sloj udruživanja 2x2 (max pooling)*
7) *Konvolucijski sloj sa 64 filtera dimenzije 3x3, aktivacijska funkcija 'relu', HeNormal*
8) *Sloj udruživanja 2x2 (max pooling)*
9) *Konvolucijski sloj sa 128 filtera dimenzije 3x3, aktivacijska funkcija 'relu', HeNormal*
10) *Sloj udruživanja 2x2 (max pooling)*
11) *Sloj za isključivanje, 25% gašenje neurona 'dropout 0.25'*
12) *Potpuno umrežen sloj sa 128 neurona, aktivacijska funkcija 'relu'*
13) *Sloj za isključivanje, 25% gašenje neurona 'dropout 0.25'*
14) *Potpuno umrežen sloj sa 5 neurona (5 klasa), aktivacijska funkcija 'softmax'*

Mreža ima isti broj parametara kao i četvrta faza treninga tj. 205,733.



Slika 7. Predikcija konačnog modela

Slika 7 prikazuje uspješnu klasifikaciju modela u prepoznavanju broja prstiju podignutih na ruci. Model je u ovom slučaju pogodio četiri od pet slika.



Slika 8. Predikcija modela na testnim slikama (%)

Kao što možemo vidjeti na slici iznad model jako uspješno (79%-95%) prepoznaje slike sa jednim podignutim prstom te sa pet podignutih prstiju na ruci. Malo manje uspješno (60%-63%) prepoznaje slike sa tri i četiri podignuta prsta na ruci. Najlošije (50%) prepoznaje slike sa dva podignuta prsta na ruci tj. svaku drugu pogodi.

5. Zaključak

Istraživanje ističe uspješnost treniranja modela za prepoznavanje broja prstiju na ruci. Model je na kraju postigao točnost od 70%. Podaci za istraživanje su osobno prikupljeni, što je omogućilo veću kontrolu nad kvalitetom skupa (slika) podataka. Nakon prikupljanja, podaci su detaljno obrađeni poput razmještanja za svaku klasu te povezani na Google Colab okruženje. Korišten je programski jezik Python u TensorFlow okviru i dodatne biblioteke za vizualizaciju i analizu podataka.

Model koji je napravljen za prepoznavanje broja prstiju na ruci ima oko 200 tisuća parametara, što ga čini relativno laganim u usporedbi s nekim modelima dubokog učenja. Ova karakteristika omogućava bržu i učinkovitu obradu slika u stvarnom vremenu.

Unatoč postignutoj uspješnosti, opisani model ostavlja prostor za nadogradnju. Poboljšanja se mogu postići povećanjem skupa podataka, dodatnim optimizacijama hiperparametara i isprobavanjem različitih arthitektura modela.

Literatura

[1] Eid, A., & Schwenker, F. (2023). *Visual Static Hand Gesture Recognition Using Convolutional Neural Network. Algorithms*, 16(8), 361. MDPI AG. Retrieved from http://dx.doi.org/10.3390/a16080361

[2] Islam, M. Z., Hossain, M. S., Islam, R. U., & Andersson, K. (2019). *Static Hand Gesture Recognition using Convolutional Neural Network with Data Augmentation.* IEEE. https://doi.org/10.1109/iciev.2019.8858563

[3] Mahmoud, A. G., Hasan, A. M., & Hassan, N. M. (2021). *Convolutional neural networks framework for human hand gesture recognition.* Bulletin of Electrical Engineering and Informatics, 10(4), 2223–2230. https://doi.org/10.11591/eei.v10i4.2926