

Sistem za prepoznavanje slova pisanih u vazduhu (eng. air writing)

Mihailo Pešić

Iva Stojanović

Mentor: Vladan Bašić

APSTRAKT

U ovom radu predstavljen je sistem za prepoznavanje slova pisanih u vazduhu (eng. air writing) pomoću pokreta kažiprsta detektovanih kamerom. Ideja sistema zasniva se na različitim gestovima prstiju kojima se definiše da li se piše, pauzira ili čuva ispisana reč. Za rešavanje problema prepoznavanja rukopisa korišćeni su različiti modeli: linearna neuronska mreža, konvolucionna neuronska mreža i YOLO arhitektura. Linearna i konvolucionna mreža trenirane su na EMNIST skupu podataka koji sadrži rukom pisana slova, dok je YOLO treniran na IAM skupu sa anotiranim rukopisnim formama. Evaluacija je vršena metrikama prilagođenim tipu modela – accuracy za klasifikacione pristupe i mAP, preciznost, odziv i F1-score za YOLO. Analiza ukazuje da su modeli posebno pouzdani kod češće zastupljenih i jednostavnijih slova, dok složeniji i ređi uzorci zahtevaju dodatne tehnike poput augmentacije ili upotrebe naprednijih arhitektura. Ovim radom ukazuje se na potencijal air writing sistema u oblastima obrazovanja i medicine.

ABSTRACT

In this paper, an *air writing* system is presented for recognizing letters written in the air through index finger movements detected by a camera. The idea of the system is based on different finger gestures that define whether writing is active, paused, or when the written word is saved. To address the problem of handwriting recognition, different models were applied: a linear neural network, a convolutional neural network, and the YOLO architecture. The linear and

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

convolutional networks were trained on the EMNIST dataset of handwritten letters, while YOLOv8 was trained on the IAM dataset with annotated handwritten forms. Evaluation was carried out using metrics suited to each model type—accuracy for classification approaches, and mAP, precision, recall, and F1-score for YOLO. The analysis shows that the models are particularly reliable for more frequent and simpler letters, while more complex and less common samples require additional techniques such as augmentation or the use of more advanced architectures. This work highlights the potential of *air writing* systems in fields such as education and medicine.

1.UVOD

Tehnologija je danas prisutna u gotovo svim aspektima života. Jedna od zanimljivih oblasti njenog razvoja su sistemi za prepoznavanje pokreta i pisanja u vazduhu, koji nude novi način interakcije između čoveka i računara . Ovi sistemi omogućavaju prenos informacija gestovima, bez potrebe za tastaturom ili papirom.

Air-writing se definiše kao pisanje slova ili reči pokretima ruke ili prsta u slobodnom prostoru. Ovakav način unosa posebno je koristan u interfejsima koji ne omogućavaju upotrebu tastature ili ekrana osetljivog na dodir, kao i kod pametnih sistema za kontrolu različitih uređaja[1].

Ovaj sistem ima široku primenu — u obrazovanju može zameniti fizičke table, a u medicini omogućava unos podataka bez dodira, a uređajima kojima se može upravljati beskontaktno sadrže manji rizik od prenosa bakterija[4].

Cilj projekta je razvoj sistema koji pomoću kamere detektuje gestove prstiju za crtanje, pauziranje i čuvanje napisanog, stvarajući intuitivan interfejs pogodan za različite primene. Način funkcionisanja sistema bi bio: Kada je kažiprst podignut - crta se, kada su podignuti kažiprst i srednji prst - ne crta se, kada su podignuti kažiprst i mali prst - sačuva se slika reči.

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

2.APARATURA

Za treniranje jednostavnijih modela korišćeni su procesorski resursi (AMD Ryzen 5 3500U sa Radeon Vega Mobile grafikom), koji su bili dovoljni za manje složene arhitekture poput linearnih i konvolucionih mreža.

Za treniranje složenijih modela, kao što je YOLO, korišćeni su grafički resursi dostupni putem platforme Google Colab. Upotrebljen je grafički procesor NVIDIA Tesla T4, koji je značajno ubrzao proces treniranja.

3.SKUP PODATAKA

Za treniranje linearne i konvolucione mreže korišćen je EMNIST skup podataka[2]. Ova baza podataka sadrži rukom pisana slova i cifre izvedena iz NIST Special Database 19, konvertovana u format slike dimenzija 28×28 piksela, sa sličnom strukturu.

U okviru ovog projekta korišćen je deo baze koji sadrži samo slova, ukupno 145.600 slika podeljenih u 26 klasa.

6.Diskusija

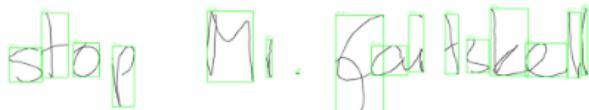
Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.



Primeri napisanih slova koja se nalaze u EMNIST skupu podataka

Za treniranje YOLO mreže korišćen je IAM Handwritten Forms skup podataka[5], koji sadrži rukopisne obrasce engleskog teksta namenjene za treniranje i testiranje sistema za prepoznavanje rukopisa. Slike su skenirane u rezoluciji od 300 dpi i sačuvane kao crno bele slike. Slike u skupu su grupisane po autorima.



Primer slike rukopisa iz IAM skupa podataka (obeleženi ograničavajući okviri)

4. METOD

Problem prepoznavanja rukom napisanih slova rešavan je primenom više različitih pristupa mašinskog učenja. Prilikom izrade su korišćeni modeli linearnih neuronskih mreža,

6. Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

konvolucionih neuronskih mreža (CNN), kao i YOLO (You Only Look Once) arhitektura, kako bi se postiglo što tačnije prepoznavanje i omogućilo poređenje performansi različitih metoda.

4.1. SEGMENTACIJA SLOVA

Prvi korak u procesu prepoznavanja rukom pisanih slova jeste njihova segmentacija, odnosno izdvajanje pojedinačnih slova. Za ovaj zadatak primjenjen je heuristički pristup koji kombinuje binarizaciju slike, morfološke operacije i prepoznavanje kontura. Ulazna slika se najpre konvertuje u sivu skalu i binarizuje globalnim pragom (vrednost 10) kako bi se istakli potezi pisanja. Zatim se primjenjuje morfološka dilatacija pravougaonim kernelom dimenzija 7×15 , koja spaja razmaknute delove linija i stabilizuje konture radi tačnijeg izdvajanja slova.

Na dobijenoj maski pronalaze se spoljne konture koje predstavljaju potencijalna slova. Svaka kontura se oblaže pravougaonim okvirom, a okviri se zatim sortiraju po horizontalnoj osi kako bi se očuvalo redosled slova u reči. Heuristika eliminiše konture koje su premale (manje od 10 piksela) ili prevelike (više od 90 % dimenzija slike), jer one obično predstavljaju šum ili nevalidne oblike. Preostalim okvirima se dodaje margina od oko 5 piksela da bi se sprečilo odsecanje poteza, a zatim se svaki isečak dodatno binarizuje Otsu metodom.

Na kraju, svako segmentirano slovo se proporcionalno skalira na standardnu veličinu od 28×28 piksela i centrirano postavlja na kvadratno platno, čime se čuva geometrija karakteristična za pojedinačna slova. Ovako pripremljeni uzorci se čuvaju radi vizuelne provere, a zatim prosleđuju CNN modelu za klasifikaciju. Ovim pristupom obezbeđena je dosledna i čista priprema podataka, što predstavlja osnovu za uspešno prepoznavanje rukom pisanih teksta.

6. Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacija, podešavanje hiperparametara ili korišćenje većih modела (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.



Prikaz toka obrade ulazne slike

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

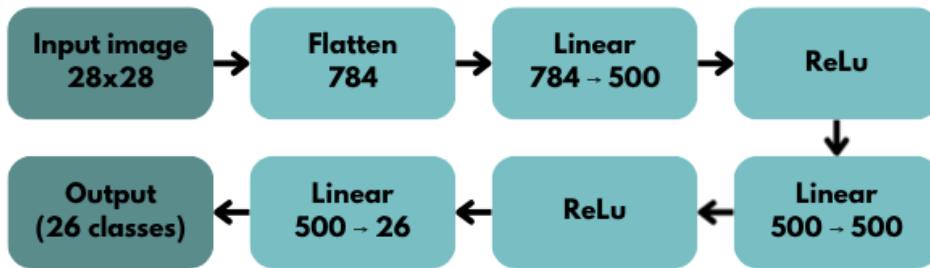
Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

4.2.LINEARNA KLASIFIKACIJA

Linearna mreža sastoje se od sloja koji ulaznu sliku dimenzija 28×28 piksela pretvara u vektor od 784 vrednosti. Nakon toga slede tri linearna sloja između kojih se nalazi ReLU aktivaciona funkcija.

Prvi linearni sloj povezuje 784 ulazne vrednosti sa 500 neurona, nakon čega se primenjuje ReLU aktivacija koja uvodi nelinearnost i omogućava mreži da uči složenije obrazce. Drugi linearni sloj takođe sadrži 500 neurona i ponovo je praćen ReLU aktivacijom.

Na kraju, izlazni linearni sloj smanjuje broj neurona na 26, pri čemu svaki neuron odgovara jednom slovu abecede.



Grafik slojeva linearne arhitekture

4.3.CNN

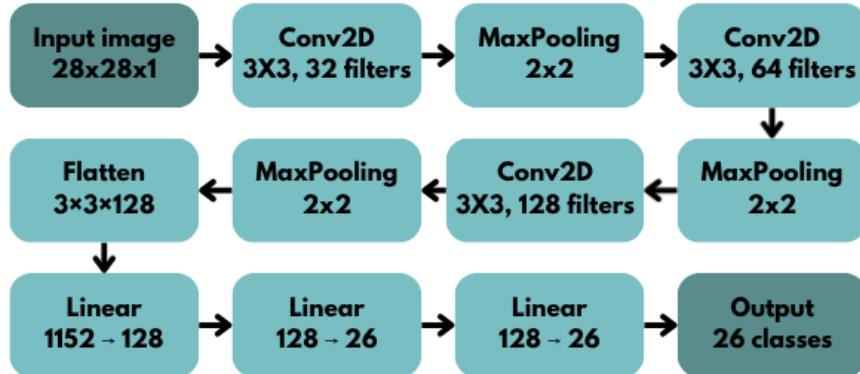
Konvolucionna mreža korišćena u radu projektovana je za direktnu klasifikaciju ulaznih slika dimenzija 28×28 piksela u 26 klasa. Arhitektura se sastoje od tri uzastopna konvolucionala bloka sa rešetkama 3×3 i maks-poolingom (stride 2), pri čemu se broj kanala postepeno povećava ($32 \rightarrow 64 \rightarrow 128$) dok se prostorne dimenzije redukuju, što na kraju stvara mapu karakteristika koje se transformiše u vektor karakteristika. U skrivenom potpuno povezanim delu koristi se jedan

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacija, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

FC sloj od 128 jedinica sa ReLU aktivacijom i dropout-om ($p=0.5$) pre izlaznog FC sloja dimenzije 26; model vraća logits koji su pogodni za treniranje pomoću funkcije gubitka unakrsna entropija (eng. CrossEntropy)..



Grafik slojeva CNN arhitekture

4.4.YOLO

YOLO model je treniran na IAM skupu podataka, koji sadrži veliki broj rukom pisanih slova i reči. Ovaj model funkcioniše tako što ulaznu sliku deli na mrežu ćelija, pri čemu svaka ćelija predviđa ograničavajuće okvire, verovatnoću prisustva objekta i njegovu klasu. Na taj način YOLO vrši istovremeno detekciju i klasifikaciju rukom pisanih slova.

Za potrebe ovog projekta korišćen je YOLOv8n model, treniran na 50 epoha, sa ulaznom rezolucijom slika 640×640 piksela i batch veličinom 8.

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modела (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

4.5. JEZIČKI SLOJ ZA KOREKCIJU REČI

Da bi se smanjile greške koje nastaju prilikom pisanja reči, nakon segmentacije i klasifikacije slova uveden je jezički sloj zasnovan na SymSpell indeksu i frekvencijskom rečniku (eng. wordfreq).

Za svaku dobijenu reč pretražuju se kandidati unutar Levenshtein udaljenosti ≤ 2 , a svaki kandidat dobija skor definisan kao:

$$\text{ocena} = \log(\text{frekvencija}) - (\text{kazneni_faktor} \times \text{udaljenost})$$

4.5.1. Eksperimentalna procena

S obzirom na to da trenutno ne postoji ručno anotiran skup rukom pisanih reči, sprovedena je simulacija tipičnih grešaka. Iz korpusa je uzeto 2000 najfrekventnijih reči, a na svaku od njih primenjene su 1–2 edit operacije (brisanje, zamena, umetanje ili transpozicija).

Na taj način formirani su parovi oblika: „stvarno napisana reč – izmenjen (pogrešan) izlaz“.

4.5.2. Metrike

Za svaku vrednost parametra kaznenog faktora u opsegu 0.00–4.00 (sa korakom 0.25) merene su sledeće metrike: tačnost posle korekcije (procenat slučajeva u kojima je jezički sloj uspešno vratio originalnu reči), broj korisnih korekcija (kada je ulazni tekst izmenjen u ispravan oblik), broj štetnih korekcija (kada je korekcija dodatno udaljila reč od originala), broj nepromenjenih slučajeva (kada je sistem odlučio da zadrži postojeći ulaz).

4.5.3. Analiza rezultata

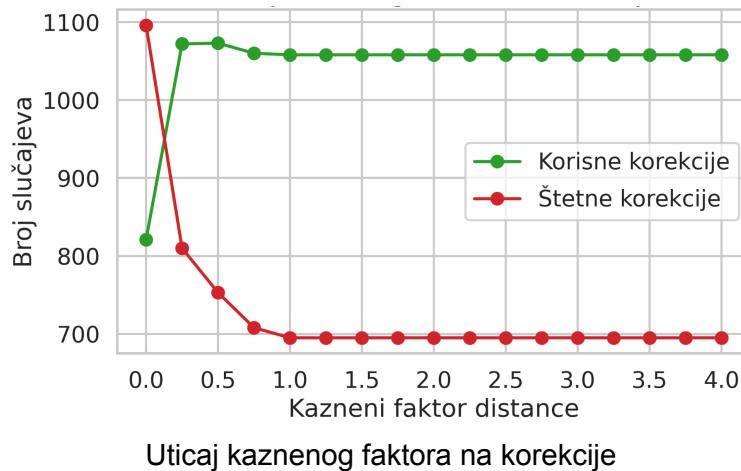
Rezultati pokazuju da je optimum postignut za $\text{kazneni_faktor} \approx 0.5$, kada tačnost dostiže oko 53–55 %. U tom opsegu odnos korisnih i štetnih korekcija je najpovoljniji, a sistem i dalje

6. Diskusija

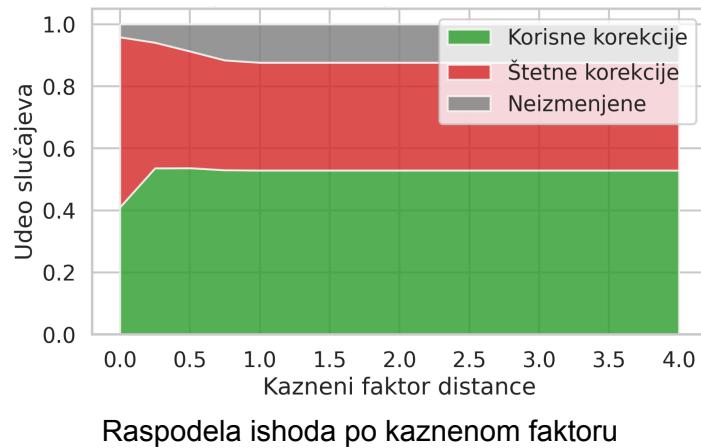
Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacija, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

interveniše dovoljno često.



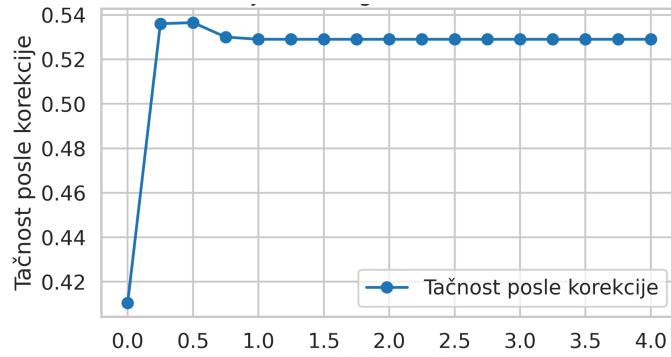
Za veće vrednosti kaznenog faktora sistem postaje oprezniji – broj pogrešnih korekcija se smanjuje, ali se ukupna tačnost ne povećava jer češće zadržava pogrešne reči.



6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.



Uticaj kaznenog faktora na tačnost

4.5.4. Zaključak simulacije i naredni koraci

Ova simulacija omogućava da se mapira ponašanje parametra `distance_penalty` pre prikupljanja stvarnih podataka. Sledeći korak podrazumeva izgradnju ručno anotiranog skupa iz direktorijuma `Writing_part/data/` i izračunavanje metrika WER (Word Error Rate) i CER (Character Error Rate), čime će se potvrditi efikasnost jezičkih korekcija na realnim primerima rukopisnih grešaka.

4.6. METRIKE

Performanse linearног и konvolucionог modela procenjivane su pomoću standardne metrike tačnosti (accuracy).

YOLOv8n model, koji je korišćen za detekciju i klasifikaciju rukom pisanih slova, evaluiran je pomoću metrika specifičnih za detekciju objekata. Korišćeni su preciznost (precision), odziv (recall), F1 mera (F1-score), mAP50 i mAP50–95.

6. Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

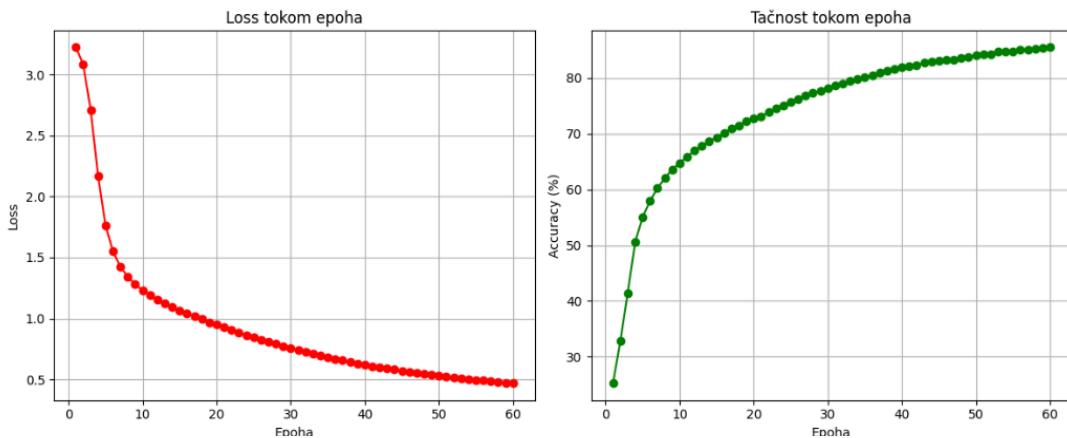
5.REZULTATI

5.1. Rezultati modela

Linearni model je treniran na 60 epoha, sa prethodno aktiviranim Early stopping-om zbog toga što se tačnost u prethodnih 7 epoha nije povećala više od 0,5%. Model je testiran na test skupu koji je činio 10% slika iz celog skupa podataka.

Početni gubitak je bio oko 3,1, što je očekivano. U prvih 10 epoha dolazi do brzog pada, a zatim se stabilizuje oko 0,5 do kraja treninga. Nema naglih oscilacija, što ukazuje da nije došlo do prenaučenosti. Početna tačnost je oko 25%, a u prvih 10 epoha brzo raste do 60%. Na kraju treninga tačnost se stabilizuje oko 85%.

Linearna mreža efikasno uči osnovne obrasce, a stabilizacija gubitka i tačnosti ukazuje na konvergenciju treninga. Za bolje rezultate mogu se povećati slojevi, promeniti optimizator ili primeniti regularizacija.



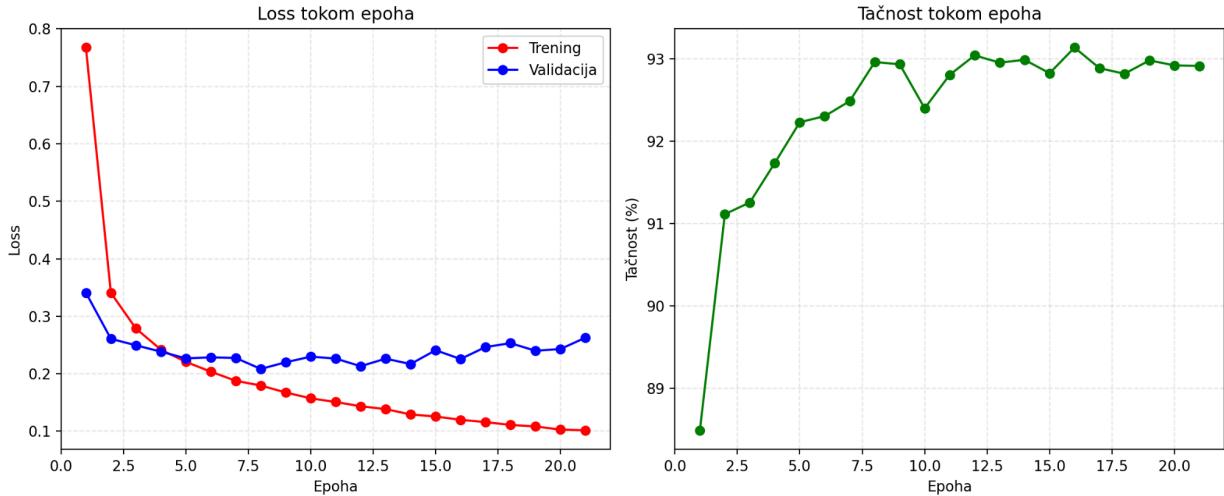
Grafik promene gubitka i tačnosti tokom treniranja linearne mreže

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modела (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

Konvolucionna mreža je trenirana 21 epohu, uz validaciju posle svake epohe i čuvanje najboljeg checkpoint-a. Gubitak tokom treninga je konstantno opadao od približno 0,77 do 0,10, dok je validacioni gubitak nakon početnog pada ($\approx 0,34 \rightarrow 0,22$) blago oscilovao i počeo da raste nakon oko 10. epohe ($\approx 0,22 \rightarrow 0,26$), što ukazuje na blago preučenje. Validaciona tačnost brzo raste tokom prvih 8–10 epoha i zatim se stabilizuje oko 93 % ($\approx 93,0 \rightarrow 93,2$ %), uz minimalne oscilacije ($\pm 0,2 \rightarrow 0,4$ pp).



Grafik promene gubitka i tačnosti tokom treniranja CNN-a

U poređenju sa linearnim modelom, CNN donosi značajan napredak u performansama, sa približno 8–10 procentnih poena veće apsolutne tačnosti, uz brže učenje. Analiza matrice konfuzije pokazuje da su preostale greške uglavnom vezane za vizuelno slična slova (npr. I \leftrightarrow J, H \leftrightarrow N, G \leftrightarrow Q), bez sistematskih odstupanja ili problema sa rotacijom.

CNN se jasno pokazuje boljim u odnosu na linearni model: brzo konvergira i postiže oko 93 % validacione tačnosti. Iako se uočava blago preučenje nakon deset epoha, to ne dovodi do degradacije performansi, što ostavlja prostor za dalja poboljšanja. Rezultati pokazuju da je model stabilan i spremjan za praktičnu upotrebu.

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

Model	Skup podataka	Broj epoha	Tačnost
Linear NN	EMNIST	60	85.45%
CNN	EMNIST	21	93.2%

Rezultati linearne i CNN modela na EMNIST skupu podataka:

Rezultati YOLO modela na IAM skupu podataka:

Model	Skup podataka	Broj epoha	mAP@50
YOLOv8n	IAM	50	0.60

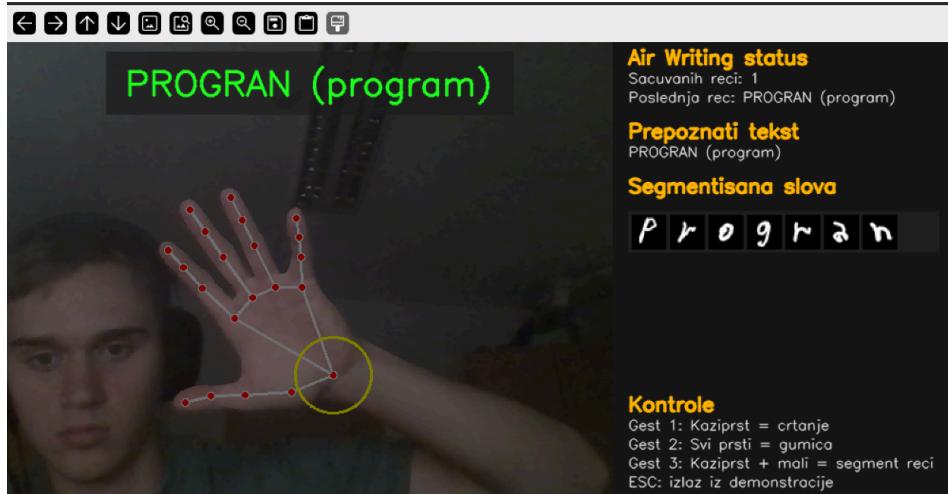
5.2. Rezultati pri korišćenju

Kao rezultat svih modela za prepoznavanje i segmentaciju, razvijen je interfejs koji je modularan u smislu promene modela, pronalaženja grešaka i uviđanja načina funkcionisanja sistema prepoznavanja slova pisanih u vazduhu. Konkretno se na snimku ekrana vidi primer ispravljanja reči, namerno je napisano program, a sistem za ispravljanje grešaka je to ispravno prepoznao.

6. Diskusija

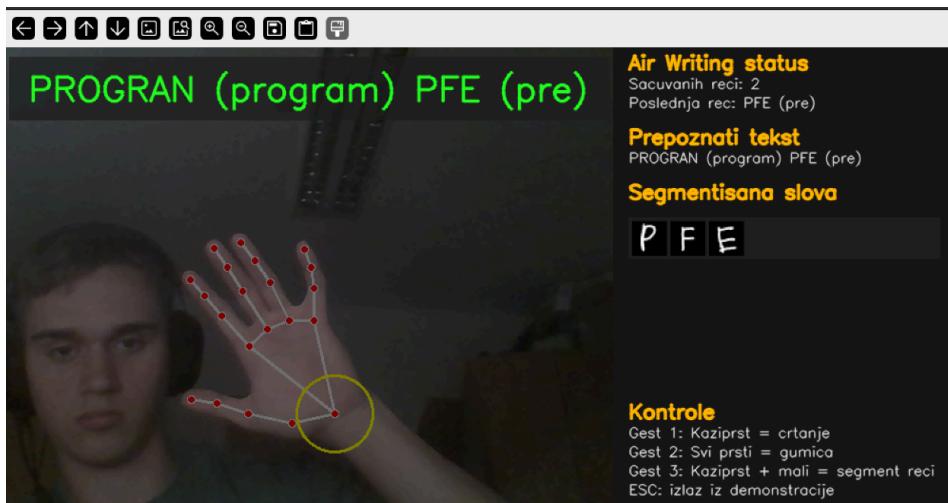
Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.



Prvi snimak ekrana - interfejs sistema za prepoznavanje slova pisanih u vazduhu

Sa obzirom da je biblioteka sa frekvencijom reči bila na engleskom, nije predviđeno da ispravno reaguje na srpske reči ili neke skraćenice, pa je tako pri pisanju skraćenice PFE ispravilo na pre.



6. Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d*, *f*, *g* i *p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i*, *j*, *k*, *q*, *x* i *z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.

7.REFERENCE

- [1] M. Chen, G. AlRegib and B. -H. Juang, "Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions," in IEEE Transactions on Human-Machine Systems, vol. 46, no. 3, pp. 403-413, June 2016, doi: 10.1109/THMS.2015.2492598.
- [2] Cohen, Gregory; Afshar, Saeed; Tapson, Jonathan; van Schaik, Andre (2017): Extended MNIST (EMNIST) dataset. Western Sydney University. <https://doi.org/10.26183/zn7s-gh79>
- [3] Luan, Tian & Zhou, Shixiong & Liu, Lifeng & Pan, Weijun. (2024). Tiny-Object Detection Based on Optimized YOLO-CSQ for Accurate Drone Detection in Wildfire Scenarios. Drones. 8. 454. 10.3390/drones8090454.
- [4] Watanabe, T.; Maniruzzaman, M.; Hasan, M.A.M.; Lee, H.-S.; Jang, S.-W.; Shin, J. 2D Camera-Based Air-Writing Recognition Using Hand Pose Estimation and Hybrid Deep Learning Model. Electronics 2023, 12, 995. <https://doi.org/10.3390/electronics12040995>
- [5] Marti, U.-V & Bunke, H.. (2002). The IAM-database: An English sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition. 5. 39-46. 10.1007/s100320200071.

6.Diskusija

Evaluacija YOLOv8n modela pokazuje da on pouzdano detektuje većinu rukom pisanih slova na IAM skupu podataka, sa ukupnim mAP50 od 0.60 i F1-score od 0.652. Analiza po klasama ukazuje na značajne razlike u performansama: slova koja su češće zastupljena u skupu, poput *d, f, g i p*, postižu visoke vrednosti AP@0.5 (>0.85), dok slova sa malim brojem instanci ili složenijim rukopisom, poput *i, j, k, q, x i z*, ostvaruju slabije rezultate (npr. *z*: AP@0.5 = 0.003).

Ove razlike pokazuju da model najbolje funkcioniše sa uobičajenim i često pojavljivim uzorcima. Slabije performanse kod retkih slova sugerisu da bi povećanje broja primera, upotreba augmentacije, podešavanje hiperparametara ili korišćenje većih modela (poput YOLOv8m/l) moglo dodatno unaprediti detekciju složenijih rukopisa.