



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko

Končno poročilo za projekt “Upravljanje z gestami”

pri predmetu Signale in slike

Marija Jovanova

1. Uvod

V okviru projekta pri predmetu Signali in slike sem se osredotočila na razvoj sistema za upravljanje vzratnih ogledal z gestami. Moj cilj je bil zasnovati rešitev, ki bi vozniku omogočila enostavno in varno prilagajanje položaja ter stanja ogledal brez fizičnega a, s čimer bi zmanjšala potrebo po odvrčanju pozornosti med vožnjo.

Za zajem in razpoznavo gest sem uporabila spletno kamero in knjižnico MediaPipe, s katero sem zaznala položaje ključnih točk na roki. Na podlagi teh točk sem s pomočjo nevronske mreže razvila klasifikacijski model, ki prepozna šest specifičnih gest, povezanih z upravljanjem ogledal:

- Premik kazalca v levo – premik ogledala v levo
- Premik kazalca v desno – premik ogledala v desno
- Premik kazalca navzgor – premik ogledala navzgor
- Premik kazalca navzdol – premik ogledala navzdol
- Dvig palca in kazalca – zapiranje vzratnega ogledala
- Dvig kazalca in sredinca – odpiranje vzratnega ogledala

Ker je medtem pri predmetu Umetna inteligenca prišlo do spremembe smeri projekta, za naš model sem vzela tudi z gesta za upravljanje z okni.

2.1. Zajem podatkov

Za zajem signalov sem uporabila kamero mobilnega telefona, ki je bila nameščena na armaturni plošči vozila in usmerjena proti rokam. Posnela sem gibe rok med simuliranimi gestami. Pri namestitvi kamere sem prilagodila snemalni kot, da bi dosegla optimalno natančnost zaznavanja. Prepričala sem se tudi, da ne bo kamera ometala voznika in s tem zagotovila varno vožnjo.

Ker smo pri drugem merjenju spremenili način zajemanja signalov, smo spremenili tudi napravo za snemanje in smo geste zajemali s spletno kamero.

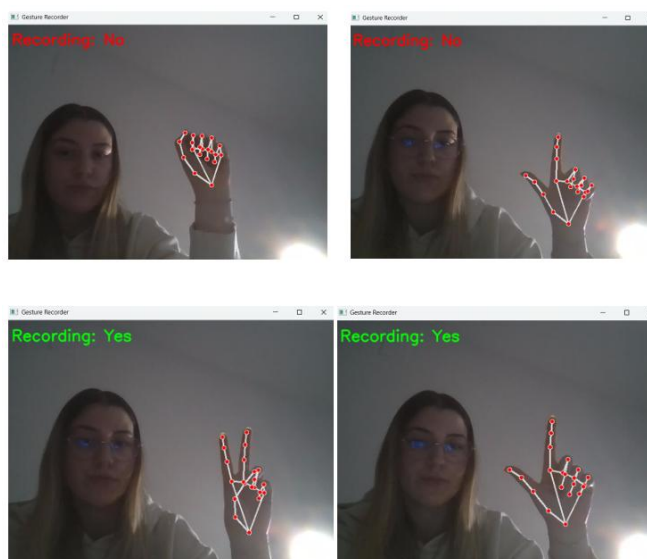
2.1. Anotacija podatkov

Na začetku smo začeli z že posnetimi gestami v avtomobilu, ki smo jih nato ročno označevali. Za označevanje sem uporabila orodje CVAT (Computer Vision Annotation Tool), s katerim sem posameznim okvirjem dodelila oznake glede na vrsto geste. Podatke sem shranila v JSON formatu.



Slika 1: Primer zajetih in označenih podatkov z CVAT

Ker je bilo označevanje podatkov v orodju CVAT časovno zahtevno, mi pa smo za učenje potrebovali veliko več označenih podatkov, smo se odločili spremeniti način označevanja. Pripravili smo Python skripto, ki z uporabo MediaPipe Skeleton omogoča avtomatsko zajemanje koordinat točk na roki. S pritiskom na tipko "r" se začne snemanje, s ponovnim pritiskom pa se snemanje zaključi. Zajeti podatki (koordinate) se nato shranijo v .npy datoteko, ki je poimenovana glede na pripadajočo oznako geste. Ta pristop nam je bistveno pohitрил postopek označevanja in omogočil učinkovitejšo pripravo učne množice.



Slika 2: Primer zajemanje podatkov z MediaPipe

2.1. Augmentacija podatkov

V sklopu priprave podatkov za učenje modela sem izvedla razširjanje (augmentacijo) podatkov, ki omogoča povečati robustnost modela in njegovo sposobnost posploševanja na nove primere. Za ta namen sem pripravila več različnih vrst augmentacij, ki simulirajo naravno variabilnost gibanja rok ter morebitne napake pri zajemu podatkov. Vse transformacije so bile implementirane v Pythonu in uporabljene na že obstoječih zajetih podatkih za geste vzvratnih ogledal. Vrste uporabljenih augmentacij:

- Normalizacija signalov:

Vse vhodne signale sem najprej normalizirala, da sem dosegla boljšo primerljivost med različnimi primeri, pri čemer sem ohranila samo pomembne informacije o obliki giba, ne pa o absolutni velikosti.

- Dodajanje šuma (add_noise)

Na signale sem dodala majhne količine naključnega normalnega šuma, kar pomaga modelu pri učenju na podatkih z manjšimi napakami ali variacijami zaradi senzorjev ali okolja.

- Zamik okvirjev (frame_shift)

Simulirala sem časovni zamik geste tako, da sem premaknila celotno zaporedje okvirjev naprej ali nazaj, s čimer sem izboljšala odpornost modela na razlike v začetku gibanja.

- Glajenje (smoothing)

Za odstranjevanje nenadnih skokov ali nenaravnih nihanj sem uporabila Hannovo okno, s katerim sem signal zgladila in s tem izboljšala reprezentacijo gibanja.

- Časovno popačenje (time_warp)

Raztegovanje ali krajšanje signala s časovno interpolacijo je model pripravilo na geste, ki so lahko izvedene hitreje ali počasneje.

- Izpuščanje okvirjev (frame_dropping)

Z naključnim odstranjevanjem določenih okvirjev in njihovo kasnejšo interpolacijo sem simulirala manjše izgube pri zajemu podatkov, kot se lahko zgodijo zaradi tehničnih motenj.

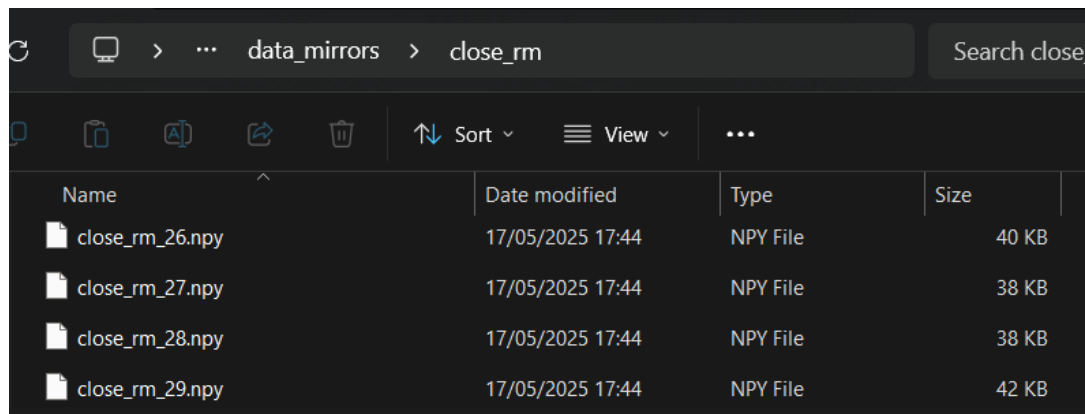
- Prilagoditev dolžine (pad_or_crop)

Vsak signal sem prilagodila na skupno dolžino (152 okvirjev), bodisi s krajšanjem ali dopolnjevanjem z ničlami, da sem zagotovila enotno dimenzionalnost vseh vhodov.

Augmentacijo sem izvedla z uporabo skripte v Pythonu, kjer sem uvozila originalne podatke iz .npy datotek (koordinate za posamezne geste) ter z naključnim izborom podatkov izvedla zgoraj opisane transformacije. Augmentirane signale sem nato shranila v ločeno mapo, skupaj z njihovimi oznakami. Primer podatka je bil shranjen v datoteki augmented_mirror_signal_X.npy, pripadajoča oznaka pa v augmented_mirror_label_X.npy. S to metodo sem razširila svoj začetni nabor podatkov, ohranila informacijo o gesti, hkrati pa dodala raznolikost, ki bo pripomogla k bolj stabilnemu in natančnemu modelu.

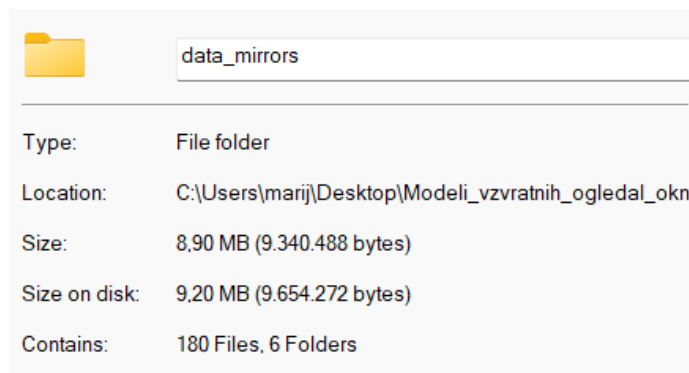
2.2. Statistika

V okviru projekta sem izvedla tri snemalne seje, v katerih sem zajela skupno 180 primerkov podatkov, enakomerno razporejenih po 30 primerkov za vsako od šestih gest (close_rm, down_rm, left_rm, open_rm, right_rm, up_rm). Vsak primer je vseboval 152 sličic z 63 značilkami (3D koordinate 21 točk na roki), shranjenih v formatu NumPy (x_mirrors.npy, y_mirrors.npy), kar skupno zaseda približno 9 MB prostora na disku.



Name	Date modified	Type	Size
close_rm_26.npy	17/05/2025 17:44	NPY File	40 KB
close_rm_27.npy	17/05/2025 17:44	NPY File	38 KB
close_rm_28.npy	17/05/2025 17:44	NPY File	38 KB
close_rm_29.npy	17/05/2025 17:44	NPY File	42 KB

Slika 3: Shranjevanje zajetih podatkov



Type:	File folder
Location:	C:\Users\marij\Desktop\Modeli_vzvrtnih_ogledal_okn...
Size:	8.90 MB (9.340.488 bytes)
Size on disk:	9.20 MB (9.654.272 bytes)
Contains:	180 Files, 6 Folders

Slika 4: Prostorska zasedenost

Snemanje je potekalo v treh ločenih terminih in je skupaj trajalo približno 6 ur, pri čemer sem si vzela čas tudi za ponovitve v primeru netočnih gibov ali slabe kakovosti posnetka. Za anotacijo, preverjanje kakovosti in pripravo podatkov sem porabila še dodatni 3 uri.

Za učno množico sem uporabila 80 % podatkov (144 primerkov), preostalih 36 primerkov pa sem prihranila za testiranje modela.

3. Zaključek

Z izvedbo zajema, priprave in augmentacije podatkov sem uspešno opravila glavni cilj naloge – ustvariti kakovostno in raznoliko množico podatkov za prepoznavo gest z roko. Z zajetimi podatki sem večinoma zadovoljna, saj so bile gibe jasno izražene, razlike med posameznimi gestami pa dovolj izrazite za učenje modela.

Augmentacija podatkov se je izkazala za zelo koristno, saj je umetno povečala količino in raznolikost primerkov, kar bo po pripomoglo k boljši generalizaciji modela. Podatke, ki sem jih pridobila v okviru tega predmeta, uporabljam za treniranje modela pri predmetu Umetna inteligenca. Tam sem na osnovi teh istih podatkov razvila nevronske mreže, ki je do sedaj že pokazala dobre rezultate – prepoznavanje gest je bilo natančno, model pa se je dobro znašel tudi pri novih primerih.

Te podatke in izvedeno augmentacijo bom sedaj uporabila za izboljšano treniranje osnovnega modela, ki sem ga razvila pri predmetu Umetna inteligenca. Na ta način mi projekt ni služil le kot naloga za zajem podatkov, temveč predstavlja tudi temelj za nadaljnje izboljševanje modelov strojnega učenja pri drugih predmetih.