1. Esu Pranciškus Ambrazas ir pasirinkau temą „Gestų kalbos atpažinimas naudojant internetinę kamerą“. Kodėl? Todėl, jog manau, kad gestakalbiams yra ganėtinai sunku kasdieniniame gyvenime nueiti į kavinę ir be problemų užsisakyti puodelį kavos ar paprašyti sąskaitos.
2. Šio pristatymo metu apžvelgsiu pagrindines ir, manyčiau, įdomiausias darbo dalis – darbo tiklslą, išsikeltus uždavinius bei panagrinėsiu kokius tyrimus atlikau, ką išbandžiau pats.
3. Norėjau išsaišikinti ar įmanoma ir jei taip, kaip sunku gestų kalbą versti į rašinę ar sakytinę kalbą, pasinaudojant įprastos internetinės kameros pagalba. Todėl ir išsikėliau tikslą – išsiaiškinti ir išanalizuoti šias galimybes. Manau, jog šio tyrimo mano asmeninis tikslas buvo išsiaiškinti galimybes, kaip palengvinti gestakalbiams kasdieninį gyvenimą, pabandant ir išsiaišikint galimybes suteikti pilnavertišką gyvenimą.
4. Todėl iš sau išsikeltų darbo tikslų susiformulavau uždavinius, kuriuos ir bandžiau įgyvendinti – išsiaiškinau gestų kalbos ypatybes, o labiausiai akcentavaus į statinę gestų kalbą, ištyriau ir galimybes, kuriomis galima atpažinti gestų kalbą ir jas išbandžiau.
5. Pasaulyje dėl klausos problemų kenčia 360 mln žmonių ir šis skaičius vis auga. Vien 32 mln iš jų yra vaikai. Pati populiariausia gestų kalba – ASL (american sign languange), kuria priskaičiuojama jog kalba nuo 0,5 iki 2 mln žmonių VIEN JAV. ASL taip pat yra ir pirmoji mažumos kalba po „didžiojo ketverto“ (ispanų, italų, vokiečių ir prancūzų). Kadangi skiriasi kiekviena kalba, tai skiriasi ir kiekviena gestų kalba. Yra randama net ir dialektų pagal regionus (kaip pvz suvalkiečiai, aukštaičiai). Kiekvieną gestų kalbą sudaro statiniai ir dinaminiai judesiai. Norint juos išsiversti – internetiniai žodynai pagal gesto formas.
6. Aptarsiu, kokius duomenis rinkausi, kokius bandymus atlikau bei ką gavau
7. Tad žinant šias teorines žinias apie gestų kalbą ir buvo pasirinkta ASL statinė gestų kalbos abėcėlės pusė, kurią sudaro 24 statiniai ženklai.
8. Pirmasis bandymas, kurį atlikau, tai buvo pasinaudojus OpenCV (atviro kodo vaizdų atpažinimo įrankiu) ir sklearn (sistemų apmokymo įrankiu). Pasirinkau – po 400 kadrų kiekvienai raidei (viso beveik 10 tūkst. kadrų) ir gavau tikslumą po apmokymo 95-98%. Atrodo visai neblogas rezultatas, tiesa? Tačiau...
9. Norint apmokyti sistemą, reikia atlikti kadro konvertacijas šiuo atveju. Čia matomos, kokios buvo atliekamos konvertacijos prieš apmokant sistemą – trumpai tariant – nuimamas fonas ir paliekamas tik dviejų spalvų paveikslėlis. Tuomet kadras verčiamas į 255 (balta) ir 0 (juoda) matricą, matrica suplokštinama ir ši informacija sudedama į duomenų rinkmeną pasinaudojus sklearn apmokymo funkcijomis. Visa bėda su šiais metodais – tiksliausias apažinimas naudojant lengvai nuimama foną, neiškraipytą rankos formą.
10. Konvoliucinis tinklas. Buvo atlikti 1+5 bandymai – 1 – nuodugnus apmokymas ir 5 – daliniai apmokymai. Iš praeitos duomenų bazės buvo pasinaudota kadrais ir pakeisti fonai kaip matoma žemžiau esančiuose kadruose.
11. Nuodugniam apmokymui buvo pasinaudota VGG16 architektūra, kuri bene lengviausiai suprantama ir išsianalizuojama naujokui susidūrusiam su konvoliuciniais neuroniniais tinklais, kurią nesunkiai galima aprašyti telkimo, konvoliucinių sluoksnių būdu. Pasirinkta apmokyti sistemą vos ant dviejų raidžių po 400 kadrų mokymui, 200 patikrinimui ir tai užtruko kiek dauigau nei 2,5 h. Rezultatas – prastokas. Nuodugnaus mokymo pagrininis minusas – daug laiko, mažas tikslumas su nedaug duomenų ir reikia galingų kompiuterių
12. Daliniam apmokymui pasirinkta Inception-v3 architektūra – vienas naujausių konvoliucinių tinklų modelių, „Google“ produktas.
13. Pirmasis bandymas – dvi raidės, 400 kadrų, 5 min. ir 95-98% tikslumas. Išbandant – visi rezultatai buvo teisingi, bet testuota nebuvo.
14. Antrasis – viskas tas pats – tik visos raidės. Rezultatas iškart nukrito dėl gestų panašumų.
15. Trečiasis – padidintas žingsnių skaičius nuo 500 iki 5000. Tikslumas ženkliai pakilo.
16. Ketvirtasis – padidinta nuo 400 kadrų kiekvienam gestui iki 1000. Tikslumas pakilo vos 2% patikrinimo metu. Šio ir kito etapo metu buvo atliktas testavimas su 240 kadrų – po 10 kiekvienai raidei.
17. Penktas – kadrų skaičius padidintas iki 3000. Rezultatas – nukritę 5% mokymo tikslumas ir 1 % pakilęs patikrinimo.
18. Bendri ir susisteminti visų 5 konvoliucinių dalinio apmokymo neuroninių tinklų modelio rezultatai – didžiausią įtaką turėjo apmokymo žingsnių padidinimas. Geriausias modelis – 4-tas su viena duomenų baze. Kodėl ne 5? Kadangi 5-toje duomenų bazėje kur 3000 kadrų, nebuvo peržiūrėti visi duomenys, todėl gali būti, jog juose buvo nemažai netikslumų, nes buvo pasitelkta ne viena ta pati duomenų bazė o iš skirtingų vietų pasiskolintos duomenų bazės. Norint dar geresnių rezultatų, manyčiau, jog reiktų didinti duomenų kiekius, duomenis švarinti, nuodugniai peržiūrėti ir tuomet juos augmentuoti, kad būtų dar daugiau duomenų ir tuomet rezultatai būtų dar geresni.
19. Viską apibendrinant reikėtų paminėti, jog dalinai apmokomi konvoliuciniai neuroniniai tinklai yra labai galingas ir ir lengvai apmokomas įrankis. Apmokant įrankį dalinai, tai vyksta kur kas greičiau, gaunami geresni rezultatai turint nedaug duomenų. Tačiau turint labai didelius duomenų kiekius, galingą įrangą ir norint pasiekti geriausius rezultatus, tuomet reiktų taikyti nuodugnų apmokymą. Norint dar geresnių rezultatų, reikėtų didinti kadrų skaičių, naudotis skirtingomis duomenų bazėmis, kurtis savas, jas augmentuoti. Ir viską apibendrinant reikėtų pasakyti, jog pilnai išvystytas modelis būtų ypač naudingas gestakalbiams ir visiems, kurie nori bendrauti su jais.