

Klasifikácia znakov v posunkovej reči pomocou metód strojového učenia

Študent: Barbora Ľapinová
Vedúci práce: RNDr. Šimon Horvát

Ústav informatiky
Prírodovedecká fakulta
Univerzita Pavla Jozefa Šafárika



Obhajoba bakalárskej práce
Košice, 21.6.2023

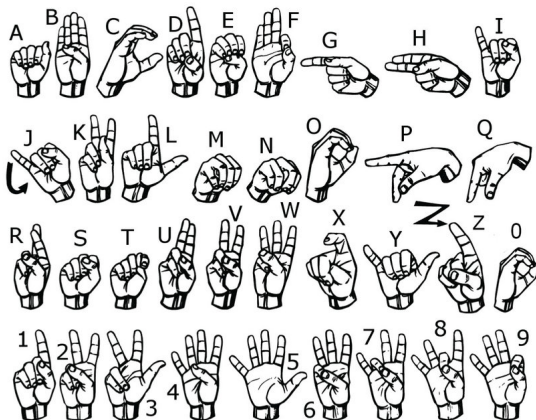
- **Posunková reč** je jazyk používaný komunitou nepočujúcich a ťažko sluchovo postihnutých ľudí.
- Ide o najpodstatnejší spôsob dorozumievania sa pre túto komunitu.
- Aj napriek tomu prevláda **komunikačná bariéra** medzi nepočujúcou a nedoslýchavou minoritou a zvyškom populácie.
- **Nová oblasť výskumu** – rozpoznávanie posunkovej reči začalo naberať na popularite začiatkom 90. rokov 20. storočia¹.
- Súčasný stav rozpoznávania posunkovej reči **zaostáva** za systémami rozpoznávania reči z audio nahrávok o približne **30 rokov**².

¹CHEOK, M. J., OMAR, Z., JAWARD, M. H., 2019. A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10, 131-153.

²ELAKKIYA, R., 2021. RETRACTED ARTICLE: Machine learning based sign language recognition: a review and its research frontier. Journal of Ambient Intelligence and Humanized Computing, 12(7), 7205-7224

- Vizuálny a manuálny jazyk so štruktúrovanou formou gest rúk.
- **Bohatý a komplexný** jazyk – umožňuje diferencovanú komunikáciu.
- Vo svete existuje mnoho posunkových jazykov.
- Naša práca sa venuje **Americkej posunkovej reči (ASL)**:
 - Až do roku **1960** sa znaky považovali za **neanalyzovateľné** a **neštruktúrované**.
 - Prvý deskriptívny systém zaviedol **William C. Stokoe**.
 - **Stokoeho notácia znakov reči** – každý znak sa skladá z troch parametrov (umiestnenie ruky, tvar ruky, jej pohyb).
 - Ako tvar ruky často slúžia znaky **Americkej manuálnej abecedy (AMA)**³.

³ STOKOE, W. C., 1980. Sign language structure. Annual review of anthropology, 9(1), 365-390.

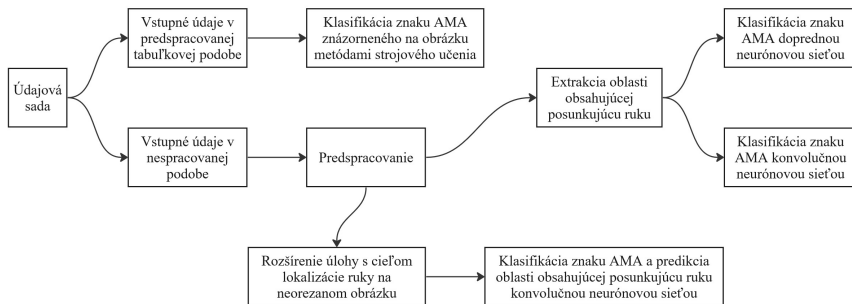


Obr.: Abeceda a číslice v ASL ⁴

⁴VALLI, C., LUCAS, C., 2000. Linguistics of American sign language: An introduction. Gallaudet University Press.

- 1 Vytvoriť prehľad aktuálne existujúcich metód strojového učenia na spracovanie obrazu.
- 2 Navrhnuť a implementovať kombináciu metód strojového učenia vhodnú na klasifikáciu znakov AMA v posunkovej reči.
- 3 Implementovať metódy neurónových sietí na lokalizáciu objektov z obrázkov posunkovej reči a porovnať dosiahnuté výsledky.

Vlastné riešenie problému



Obr.: Diagram procesu vlastného riešenia problému

Údajová sada

- Obrázky so znakmi AMA, **okrem písmen J a Z.**
- Dve formy: pôvodná (nespracovaná) a upravená.
- **Pôvodné (nespracované) obrazové údaje:**
 - 1680 obrázkov od siedmich rôznych používateľov.
 - Každý znak je od každého používateľa znázornený na desiatich obrázkoch – tie sa od seba navzájom líšia.
 - Ku každému obrázku máme k dispozícii tabuľku so súradnicami okna, v ktorom sa nachádzajú informácie o polohe ruky na obrázku.



Obr.: Dva rôzne obrázky znázorňujúce ten istý znak od jedného používateľa⁵

⁵ REEJITH, S., 2018. Sign Language and Static Gesture Recognition using sklearn [online] [cit. 09.05.2023]. Dostupný na <https://github.com/mon95/Sign-Language-and-Static-gesture-recognition-using-sklearn>

• Upravené obrázky s extrahovanou oblasťou ruky

- V podobe štrukturovaných údajov.
- Rozdelené na tréningovú a testovaciu množinu.
- Okrem extrakcie ruky autori tejto sady aplikovali na obrázky rôzne augmentácie, čím vytvorili dokopy takmer 35 000 vstupov.



Obr.: Ukážka obrázkov po predspracovaní⁶

⁶Sign Language MNIST, 2017. [online] [cit. 09.05.2023]. Dostupný na <https://www.kaggle.com/datasets/datamunge/sign-language-mnist>

1. Klasifikácia znaku AMA tradičnými metódami strojového učenia

- Použité klasifikátory boli založené na metódach **stroje s podpornými vektormi** (SVM) a **Gaussov naivný Bayes** (GNB).
- Implementácia v programovacom jazyku Python, použité metódy z knižnice *scikit-learn*.
- Ako vstupné údaje boli použité **štrukturované dáta**.
- Pre metódu SVM sme použili tri rôzne jadrové funkcie, metódu GNB sme použili bez ďalších úprav.

Model	Správnosť
SVM - Linear kernel	78%
SVM - RBF kernel	84%
SVM - Sigmoid kernel	2%
GNB	39%

Tabuľka: Výsledky klasifikácie tradičnými metódami strojového učenia

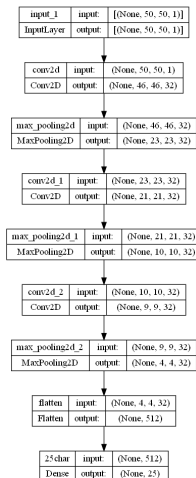
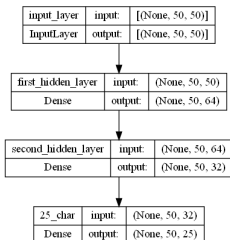
2. Klasifikácia znaku AMA metódami hĺbkového učenia

- Vstupnými údajmi boli v tomto prípade **nespracované obrázky**.
- Implementácia opäť v programovacom jazyku Python, konvolučné a neurónové siete sme zostavili pomocou knižníc *Keras* a *TensorFlow*.
- Dve hlavné časti:
 - **Klasifikácia AMA znakov** doprednými a konvolučnými neurónovými sieťami.
 - **Klasifikácia znaku AMA a lokalizácia posunkujúcej ruky** na obrázku pomocou konvolučných neurónových sietí.

2.1. Klasifikácia znaku AMA

- Prvým krokom bolo **predspracovanie údajov** – každý obrázok sme previedli do farebnej škály odtieňov sivej a orezali ho len na oblasť posunkujúcej ruky.
- **Rozdelenie údajov** na trénovaciu a testovaciu množinu.
- Následne sme zostavili **konvolučnú** a **doprednú** neurónovú sieť.
- Siete sme trénovali na rovnakých vstupných dátach na 50 a 100 epochách.

2.1. Klasifikácia znaku AMA

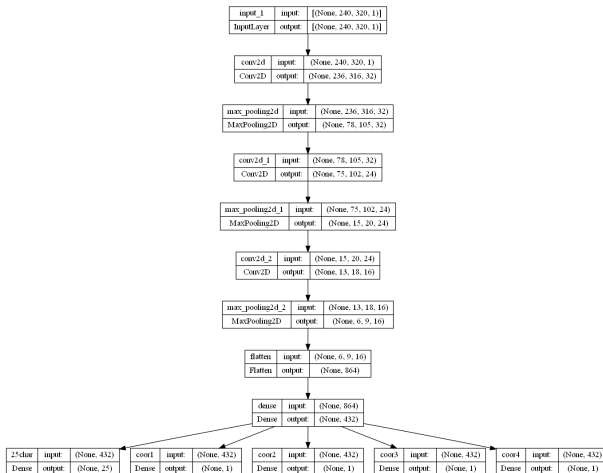


Obr.: Dopredná a konvolučná sieť použitá na klasifikáciu znaku AMA

2.2. Lokalizácia posunkujúcej ruky a klasifikácia znaku AMA

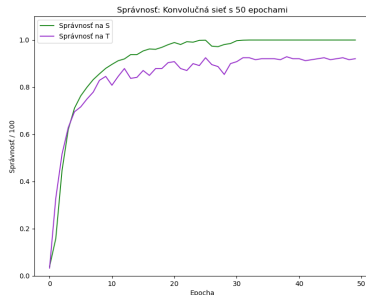
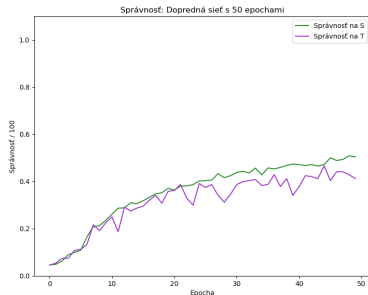
- Rozšírenie úlohy na **klasifikačno-regresnú**.
- Cieľom bolo okrem **znaku** posunkovaného na obrázku predpovedať súradnice **ľavého horného bodu** (x_1, y_1) a **pravého dolného rohu** (x_2, y_2), na základe ktorých sa zostrojí obdĺžnik ohraničujúci ruku – **bounding box**.
- Vstupnými údajmi boli **neorezané obrázky** v plnej veľkosti 320×240 , ktoré obsahovali rôzne rušivé elementy.

2.2. Lokalizácia posunkujúcej ruky a klasifikácia znaku AMA



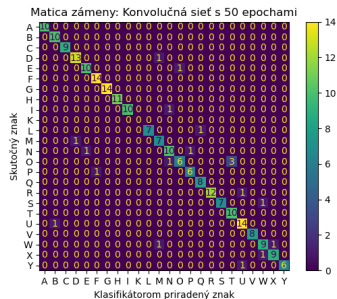
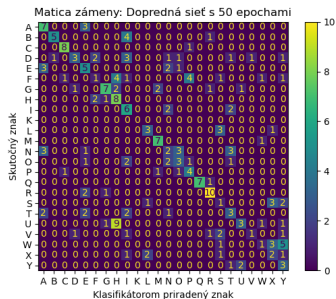
Obr.: Architektúra neurónovej siete pre klasifikačno-regresnú úlohu

Výsledky – Klasifikácia znaku AMA doprenými a konvolučnými sieťami trénovanými na 50 epochách



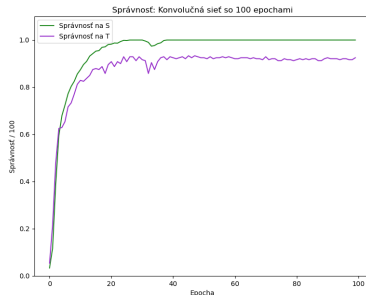
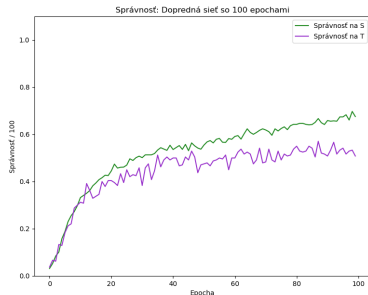
Obr.: Grafy znázorňujúce metriku správnosť v jednotlivých epochách na modeloch trénovaných na 50 epochách

Výsledky – Klasifikácia znaku AMA doprednými a konvolučnými sieťami trénovanými na 50 epochách



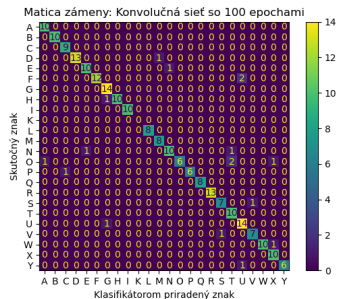
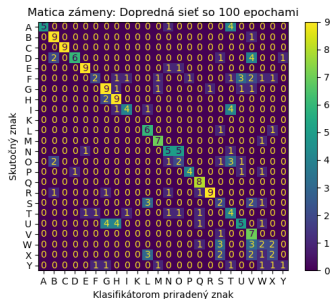
Obr.: Matice zámieny modelov trénovaných na 50 epochách

Výsledky – Klasifikácia znaku AMA doprednými a konvolučnými sieťami trénovanými na 100 epochách



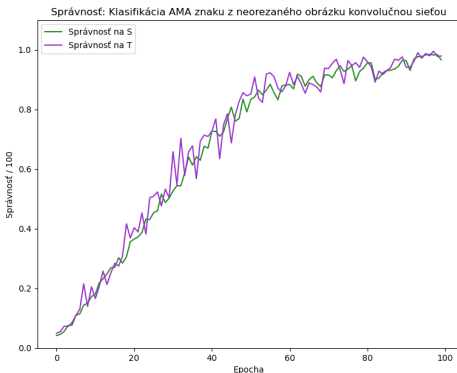
Obr.: Grafy znázorňujúce metriku správnosť v jednotlivých epochách na modeloch trénovaných na 100 epochách

Výsledky – Klasifikácia znaku AMA doprednými a konvolučnými sieťami tréňovanými na 100 epochách



Obr.: Matice zámény modelov tréňovaných na 100 epochách

Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami



Obr.: Grafické znázornenie správnosti pre jednotlivé epochy klasifikačnej časti klasifikačno-regresnej úlohy

Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami

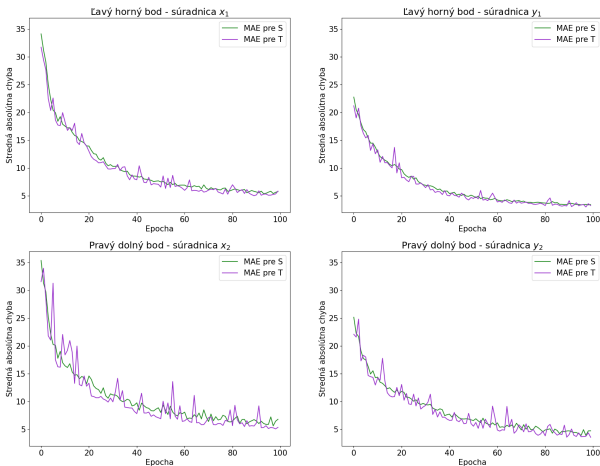
Súradnica	MAE po 100 epochách	RMSE po 100 epochách
Ľavý horný bod – súradnica x_1	5.77	7.48
Ľavý horný bod – súradnica y_1	3.41	4.46
Pravý dolný bod – súradnica x_2	6.81	9.69
Pravý dolný bod – súradnica y_2	4.73	6.01

Tabuľka: Výsledky regresnej časti po 100 epochách



Obr.: Ukážka skutočného bounding boxu a bounding boxu predpovedaného naším modelom

Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami



Obr.: Grafické znázornenie MAE pre jednotlivé súradnice

Výsledky – Porovnanie nášho prístupu s prístupmi iných výskumných skupín

Prístup	Model	Typ úlohy	Vyhodnotenie (Správnosť/MAE)
Naše riešenie	CNN	Klasifikačná Regresná	96.69% 5.18 px
Rao et al.	CNN	Klasifikačná	92.88%
Starner et al. (1. prístup)	HMM	Klasifikačná	92.00%
Starner et al. (2. prístup)	HMM	Klasifikačná	98.00%
Pigou et al.	CNN	Klasifikačná	91.70%

Tabuľka: Porovnanie nášho prístupu s podobnými prácami^{7,8,9}

⁷ RAO, G. A., SYAMALA, K., KISHORE, P. V. V., SASTRY, A. S. C. S., 2018. Deep convolutional neural networks for sign language recognition. In 2018 conference on signal processing and communication engineering systems, (pp. 194-197). IEEE.

⁸ STARNER, T., WEAVER, J., PENTLAND, A., 1998. Real-time american sign language recognition using desk and wearable computer based video. IEEE Transactions on pattern analysis and machine intelligence, 20(12), 1371-1375.

⁹ PIGOU, L., DIELEMAN, S., KINDERMANS, P. J., SCHRAUWEN, B., 2015. Sign language recognition using convolutional neural networks. In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13 (pp. 572-578). Springer International Publishing.

Ďakujem za Vašu pozornosť.

Ako by sa zmenila architektúra použitej siete, ak by sme sa na lokalizáciu ruky pozreli ako na klasifikačnú, nie regresnú úlohu?

Ak by sme sa na úlohu lokalizácie ruky pozreli ako na klasifikačnú úlohu, zmenila by sa architektúra siete na štyroch výstupných vrstvách pre jednotlivé súradnice. Keďže je pre nás známa veľkosť obrázka 320×240 , každá súradnica by bola reprezentovaná buď vektorom dĺžky 320, alebo 240, pričom by obsahoval práve jednu jednotku na mieste, ktoré je danou súradnicou. Preto by na týchto štyroch výstupných vrstvách miesto jedného výstupného neurónu bolo 320, alebo 240 neurónov.

Odpoveď na otázku od oponenta práce

Popíšte možné úlohy, ktoré by bolo možné riešiť v tejto oblasti v budúcnosti. Aké ďalšie aplikácie strojového učenia v oblasti telemedicíny majú podľa Vás v súčasnosti potenciál?

Zaujímavou úlohou aplikácie rozpoznávania posunkovej reči je využitie týchto systémov v reálnom čase napríklad počas videohovorov, ktoré sú často pre nepočujúcich a ťažko sluchovo postihnutých ľudí problémové, alebo priamo pri návšteve lekára, či v úradnom prostredí.

Telemedicína umožňuje pacientom lepší prístup k zdravotnej starostlivosti na diaľku. Zaujímavou úlohou by bola napríklad analýza signálov z inteligentných hodínok pre pacientov s chronickou nespavosťou, alebo srdcovou arytmiou.