

Rozpoznávanie posunkového jazyka pomocou metód hĺbkového učenia

Barbora Lapinová

Školiteľ: doc. RNDr. Ľubomír Antoni, PhD.



30.04.2025

Košice

Obsah

- ✿ Motivácia práce
- ✿ Základné pojmy v posunkovom jazyku
- ✿ Navrhované riešenie pre rozpoznávanie posunkového jazyka
- ✿ Výsledky

Motivácia

- ✿ **Posunkový jazyk** je jazyk používaný komunitou ľudí so sluchovým postihnutím či rečovým poruchami.
- ✿ Predstavuje **primárny komunikačný prostriedok** pre túto komunitu.
- ✿ Aj napriek tomu vo svete prevláda **komunikačná bariéra** medzi touto minoritou a zvyškom populácie, čo sťažuje ich plnú integráciu do spoločnosti.
- ✿ **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 RETRACTED ARTICLE: Machine learning based sign language recognition: a review and its research frontier.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

Posunkový jazyk

- ✿ Prirodzený jazyk, ktorý využíva **telesno-vizuálny kanál** – je vyjadrovaný telom a vnímaný zrakom.
- ✿ Oproti hovoreným jazykom majú posunkové jazyky svoje **vlastné a nezávislé lexikóny a gramatické štruktúry**.
- ✿ Posunky v posunkovom jazyku sa skladajú z nasledujúcich nedeliteľných charakteristík:
 - ✿ **Manuálne charakteristiky**, ako sú tvar ruky, jej pozícia a pohyb, orientácia dlane a prstov a podobne.
 - ✿ **Nemanuálne charakteristiky**, ktoré zahŕňajú napríklad kývanie hlavou, orientáciu ramien, výrazy tváre a pohyby ústami³.

³ADALOGLOU, N., et al., 2021. A comprehensive study on deep learning-based methods for sign language recognition. IEEE transactions on multimedia, 24, 1750- 1762.

Posunkový jazyk

- ✿ Jedinečná kombinácia týchto charakteristík **vytvára posunok** a **reprezentuje gloss**.
- ✿ Gloss je **základná stavebná jednotka posunkového jazyka** a **reprezentuje slovo** v hovorenom jazyku, ktoré posunok zachytáva⁴.

Význam



Posunok



Gloss

Mačka

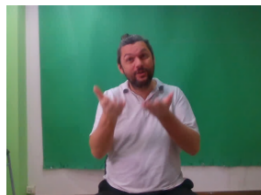
⁴PAPATSIMOULI, M., et al., 2020. Speak with signs: Active learning platform for Greek Sign Language, English Sign Language, and their translation. arXiv preprint arXiv:2012.11981.

Riešená úloha a výber údajovej sady

- ✿ V práci riešime úlohu **izolovaného rozpoznávania posunkového jazyka**, ktorého cieľom je pre posunok na videu určiť prislúchajúci gloss.
- ✿ Pri výbere údajovej sady sme dbali na to, aby sa v jej lexikóne nachádzal **dostatočný počet glossov**, kategórie boli dobre zastúpené, či rôznorodosť jednotlivých videozáznamov.
- ✿ Po preskúmaní viacerých údajových sád sme sa rozhodli pre údajovú sadu **Greek Sign Language (GSL)**:
 - ✿ **Počet videí:** 40 825
 - ✿ **Počet jedinečných kategórií:** 309
 - ✿ **Priemerný počet videí pre jednu kategóriu:** 132
 - ✿ **Medián počtu videí pre jednu kategóriu:** 35

Výber údajovej sady pre izolované SLR

- ✿ Posunky v tejto údajovej sade môžeme považovať za časté pri komunikácii **v zdravotníctve** či **verejnej správe**⁴.



⁴PAPATSIMOULI, M., et al., 2020. Speak with signs: Active learning platform for Greek Sign Language, English Sign Language, and their translation. arXiv preprint arXiv:2012.11981.

Architektúra navrhovaného modelu

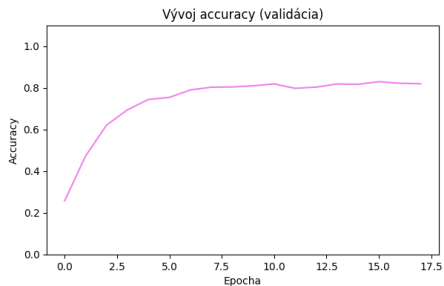
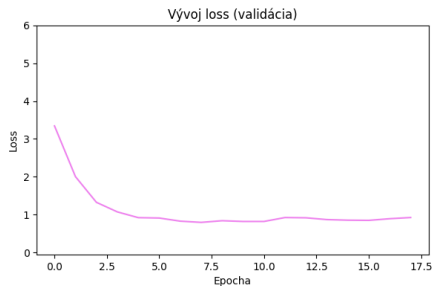
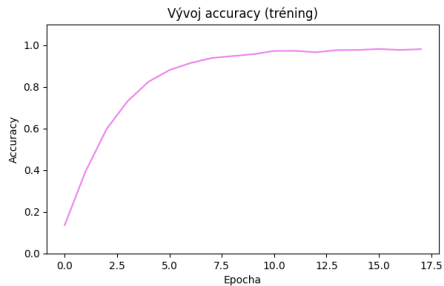
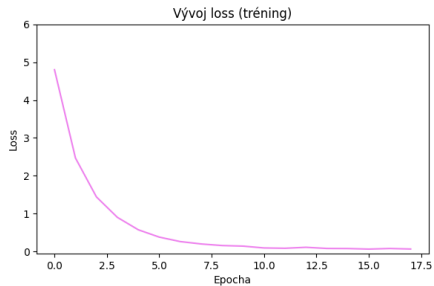
Poradie vrstvy	Typ vrstvy	Vstupný rozmer	Výstupný rozmer
1.	Vstupná vrstva	(5, 240, 320, 3)	(5, 240, 320, 3)
2.	3D konvolučná	(5, 240, 320, 3)	(3, 238, 318, 16)
3.	3D združovacia (max)	(3, 238, 318, 16)	(3, 119, 159, 16)
4.	3D konvolučná	(3, 119, 159, 16)	(1, 117, 157, 32)
5.	3D združovacia (max)	(1, 117, 157, 32)	(1, 58, 78, 32)
6.	Splošťovacia	(1, 58, 78, 32)	(144768)
7.	Hustá	(144768)	(10000)
8.	Hustá	(10000)	(500)
9.	Hustá	(500)	(128)
10.	Výstupná	(128)	(309)

Prípadová štúdia výkonnosti modelu

- ✿ Výkonnosť navrhovaného modelu sme otestovali na **dvoch podmnožinách údajovej sady** a následne na **celej dátovej sade**.
- ✿ Prvá podmnožina obsahovala 3 vybrané kategórie (716 vstupov) a druhá podmnožina obsahovala 10 vybraných kategórií (2 171 vstupov).

Počet kategórií	Počet príkladov v tréningovej množine	Počet príkladov v testovacej množine	Počet epôch	Veľkosť dávky	Maximálna hodnota správnosti	
					Tréningová množina	Testovacia množina
3	608	108	25	16	100.00%	100.00%
10	1 845	326	20	32	100.00%	97.19%
309	32 443	8 111	18	128	98.17%	82.92%

Výsledky na celej údajovej sade



Výsledky na celej údajovej sade – analýza matice zámeny

- ✿ Analýzou normalizovanej matice zámeny sme identifikovali **10 najhoršie klasifikovaných kategórií**:

Kategória	Percento správne klasifikovaných príkladov v kategórií
ΕΓΩ(3)	0.00%
TPITON	9.52%
15	14.29%
ΠΡΟΣ	14.29%
ΕΓΩ(2)	15.38%
ΚΕΠΑ(Δ.Α.)	23.81%
200	28.57%
ΑΚΟΥΩ_ΜΕΙΩΝΩ	28.57%
ΑΥΤΗ_ΤΗ_ΣΤΙΓΜΗ	28.57%
ΓΥΝΑΙΚΟΛΟΓΟΣ	28.57%

Výsledky na celej údajovej sade – analýza matice zámeny

- ✿ Jedna z kategórií – $E\Gamma\Omega(3)$ nebola správne klasifikovaná **ani v jednom prípade**.
- ✿ Pre gloss $E\Gamma\Omega$ (slovensky *ja*) máme tri verzie posunkov.
- ✿ $E\Gamma\Omega(3)$ bola v **100% prípadov** klasifikovaná ako $E\Gamma\Omega(1)$.
- ✿ V skutočnosti máme **4 glossy**, ktoré majú viacero verzii posunkov.
- ✿ Tieto verzie posunkov sme **zlúčili do jednej kategórie** pre daný gloss, keďže nám ide o určenie glossu, nie verzie posunkov – výsledkom bolo **304 jedinečných kategórií**.
- ✿ Po zlúčení kategórií sme navrhovaný model znova natrénovali s rovnakým počtom epôch.

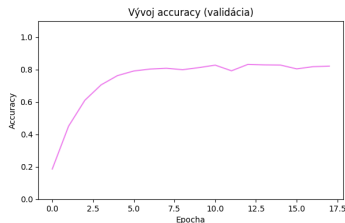
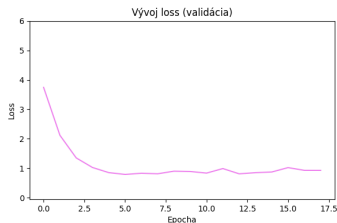
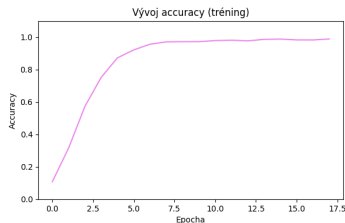
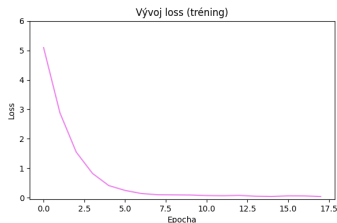
Výsledky po zlúčení kategórií

Maximálna dosiahnutá správnosť na tréningovej množine: **98.87%**

(pred zlúčením 98.17%)

Maximálna dosiahnutá správnosť na validačnej množine: **83.16%**

(pred zlúčením 82.92%)



Výsledky po zlúčení kategórií

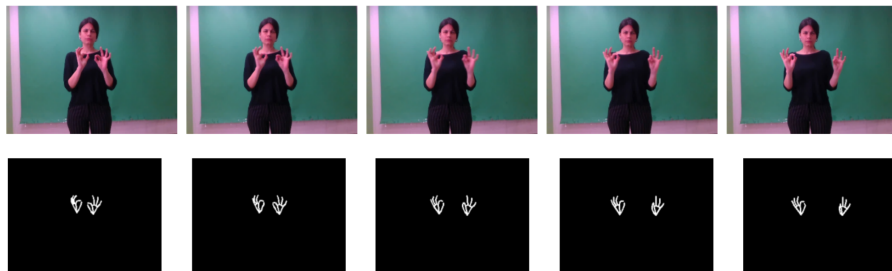
Pred zlúčením kategórií		Po zlúčení kategórií	
Kategória	Percento správne klasifikovaných príkladov v kategórií	Kategória	Percento správne klasifikovaných príkladov v kategórií
ΕΓΩ(3)	0.00%	MONO	14.29%
ΤΡΙΤΟΝ	9.52%	ΤΕΤΑΡΤΟΝ	21.43%
15	14.29%	1000	28.57%
ΠΡΟΣ	14.29%	15	28.57%
ΕΓΩ(2)	15.38%	2	28.57%
ΚΕΡΑ(Δ.Α.)	23.81%	ΑΛΗΘΕΙΑ	28.57%
200	28.57%	ΑΝΑΘΕΤΩ	28.57%
ΑΚΟΥΩ_ΜΕΙΩΝΩ	28.57%	ΑΞΙΟΛΟΓΗΣΗ	28.57%
ΑΥΤΗ_ΤΗ_ΣΤΙΓΜΗ	28.57%	ΕΚΤΥΠΩΝΩ	28.57%
ΓΥΝΑΙΚΟΛΟΓΟΣ	28.57%	ΝΟΣΟΚΟΜΕΙΟ	28.57%

Výsledky po zlúčení kategórií

Prahová hodnota	Počet kategórií so správnym percentom klasifikácie menším ako zvolená prahová hodnota	
	Pred zlúčením kategórií	Po zlúčení kategórií
80%	155	122
75%	137	108
70%	94	65
65%	91	62
60%	84	59
55%	45	33
50%	40	29

Využitie doplnkových informácií

- ✿ V literatúre sa pri klasifikácii videa často využívajú **doplnkové informácie**.
- ✿ Implementovali sme metódu, ktorá pre video vytvorí **skeletony rúk** pomocou knižnice Mediapipe:



- ✿ Pre jednotlivé príklady údajovej sady sme vykonali **analýzu možnosti vygenerovania skeletonov**.

Analýza generovania skeletoňov pre tréningovú množinu

Z celkového počtu **32 443 príkladov** v tréningovej množine sme pomocou **rôznych filtrovacích kritérií** dostali nasledujúce výsledky:

Filtrovacie kritérium	Počet videí	Percento príkladov z celej tréningovej množiny	Počet jedinečných kategórií
0 rúk detegovaných na každom videorámci	564	1.74%	-
1 ruka detegovaná na každom videorámci	497	1.53%	-
2 ruky detegované na každom videorámci	658	2.03%	-
2 ruky detegované na aspoň 90% videorámcoch	1315	4.05%	60
2 ruky detegované na aspoň 95% videorámcoch	799	2.46%	38
2 ruky nedetegované na maximálne 1 videorámci	1260	3.88%	61
2 ruky nedetegované na maximálne 2 videorámcoch	1916	5.91%	83

Porovnanie výsledkov nášho modelu s inými podobnými štúdiami

Model	Typ konvolučnej neurónovej siete	Správnosť
GoogLeNet + TConvs ³	2D	86.03%
3D-ResNet + BLSTM ³	3D	86.23%
I3D + BLSTM ³	3D	89.74%
Náš navrhovaný model pred zlúčením opakujúcich sa kategórií	3D	82.92%
Náš navrhovaný model po zlúčení opakujúcich sa kategórií	3D	83.16%

³ADALOGLOU, N., et al., 2021. A comprehensive study on deep learning-based methods for sign language recognition. *IEEE Transactions on Multimedia*, 24, 1750–1762.

Ďakujem za pozornosť.