

# Rozpoznávanie posunkovej reči (SLR)

Baška Ľapinová

21.3.2024

# Obsah

- Motivácia
- Posunková reč
- Bakalárska práca
- Prechod k diplomovej práci

1. časť  
Motivácia

# Motivácia

- **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. storočia<sup>1</sup>.
- Súčasný stav rozpoznávania posunkovej reči **zaostáva** za systémami rozpoznávania reči z audio nahrávok o približne **30 rokov**<sup>2</sup>.

---

<sup>1</sup>CHEOK, 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.

<sup>2</sup>ELAKKIYA, 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 RETRACTED ARTICLE: Machine learning based sign language recognition: a review and its research frontier.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

2. část  
Posunková řeč

# Posunková reč

- 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)**<sup>3</sup>.

---

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

# Posunok

Dnes je už známe, že sa posunok skladá z piatich základných častí:

1. Umiestnenie ruky.
2. Tvar ruky.
3. Pohyb, ktorý ruka vykonáva.
4. Orientácia dlane.
5. Výraz tváre.

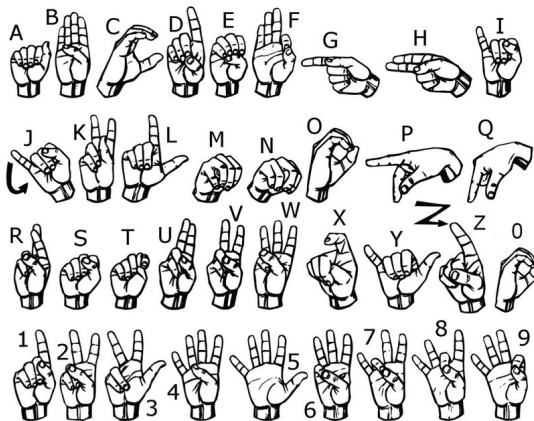


Obr.: Posunok pre slovo cat<sup>4</sup>.

---

<sup>4</sup><https://www.lifeprint.com/asl101/pages-signs/c/cat.htm>

# Posunková řeč



Obr.: Abeceda a číslice v ASL<sup>5</sup>.

---

<sup>5</sup>VALLI, C., LUCAS, C., 2000. Linguistics of American sign language: An introduction. Gallaudet University Press.



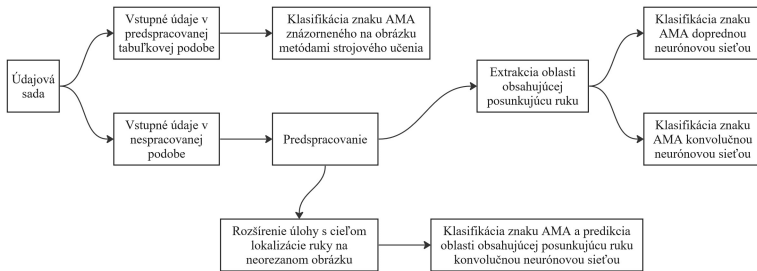
3. část  
Bakalářská práce

# Zadanie bakalárskej práce

**Téma:** Klasifikácia znakov v posunkovej reči pomocou metód strojového učenia.

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

# Postup riešenia 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.: Ukážka datasetu<sup>6</sup>.

---

<sup>6</sup> 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>

# Údajová sada

- **Upravené obrázky s extrahovanou oblasťou ruky**
  - V podobe štrukturovaných údajov.
  - Rozdelené na trénovaciu 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 predspracovaní<sup>7</sup>.

---

<sup>7</sup>Sign 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.

<b>Model</b>	<b>Správnosť</b>
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

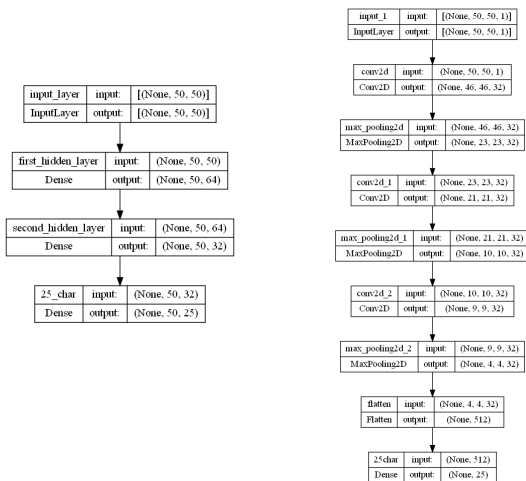
- Vstupnými údajmi boli v tomto prípade **nespracované obrázky**.
- Implementácia opäť v programovacom jazyku Python, konvolučné a neurónové siete sme zostavili pomocou knižníc *Keras* a *TensorFlow*.
- Dve hlavné časti:
  - **Klasifikácia AMA znakov** doprednými a konvolučnými neurónovými sieťami.
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

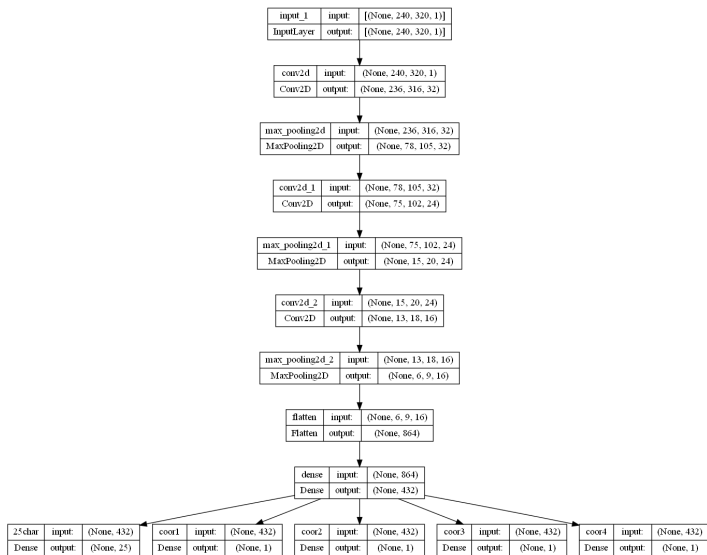


Obr.: Dopredná a konvolučná sieť použitá na klasifikáciu znaku AMA.

## 2.2. Lokalizácia posunkujúcej ruky a klasifikácia znaku AMA

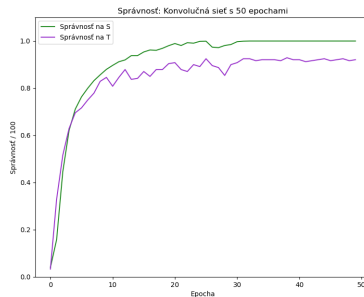
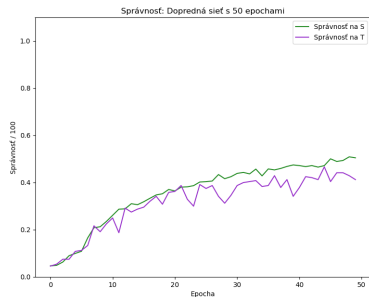
- Rozšírenie úlohy na **klasifikačno-regresnú**.
- Cieľom bolo okrem **znaku** posunkovaného na obrázku predpovedať súradnice **ľavého horného bodu** ( $x_1, y_1$ ) a **pravého dolného rohu** ( $x_2, y_2$ ), na základe ktorých sa zostrojí obdĺžnik ohraničujúci ruku – **bounding box**.
- Vstupnými údajmi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

## 2.2. Lokalizácia posunkujúcej ruky a klasifikácia znaku AMA



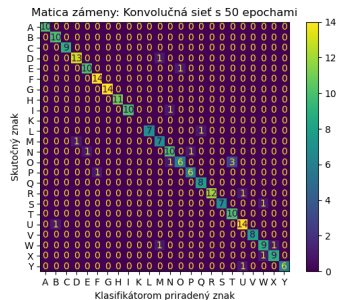
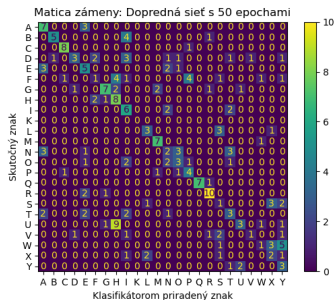
Obr.: Architektúra neurónovej siete pre klasifikačno-regresnú úlohu.

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



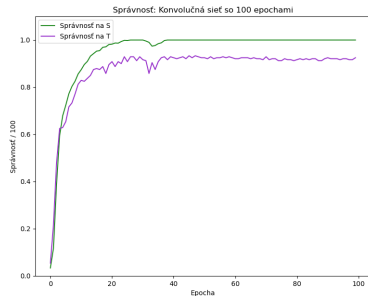
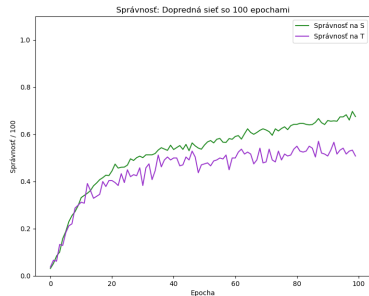
Obr.: Grafy znázorňujúce metriku správnosť v jednotlivých epochách na modeloch trénovaných na 50 epochách.

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



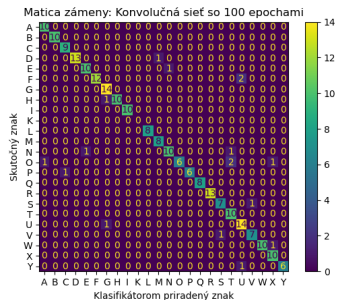
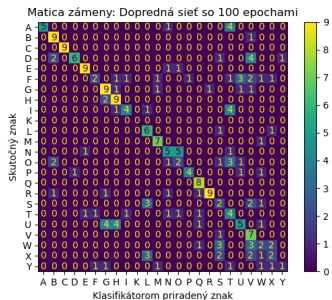
Obr.: Matice zámény modelov trénovaných na 50 epochách.

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



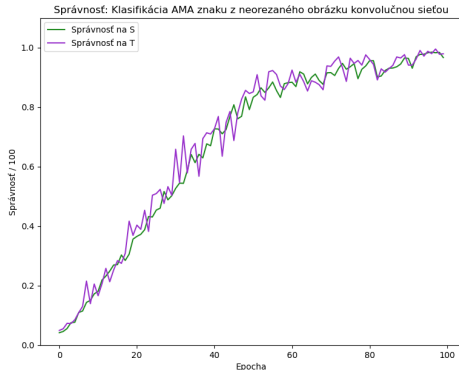
Obr.: Grafy znázorňujúce metriku správnosť v jednotlivých epochách na modeloch trénovaných na 100 epochách.

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



Obr.: Matice zámény modelov trénovaných 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.

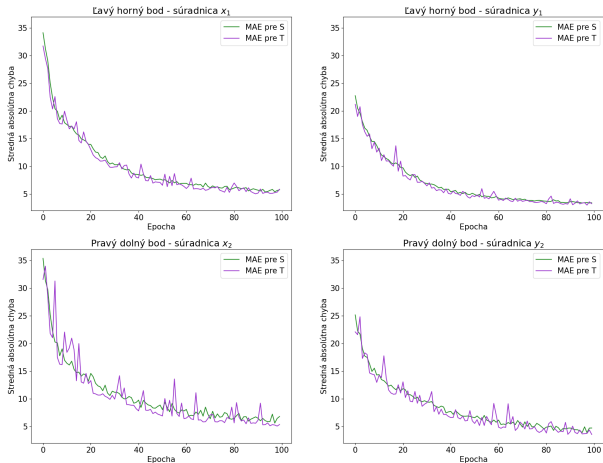


# Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami

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.

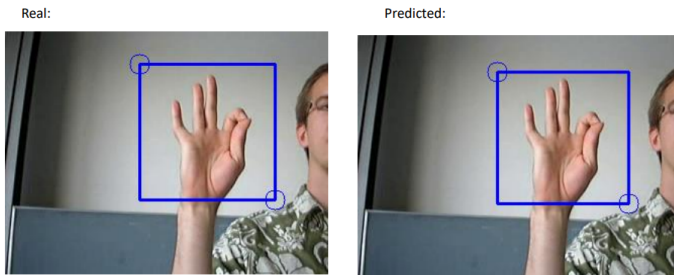
# Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami



Obr.: Grafické znázornenie MAE pre jednotlivé súradnice.

# Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami

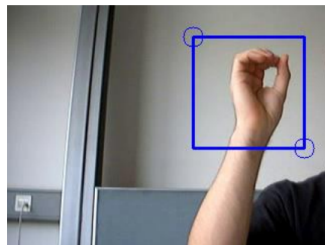
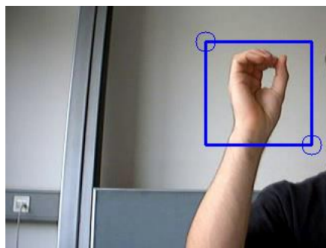
```
Predicted sign: 5, Real sign: 5  
Predicted x1_coordinate: 133.0, Real x1_coordinate: 138  
Predicted y1_coordinate: 32.0, Real y1_coordinate: 36  
Predicted x2_coordinate: 267.0, Real x2_coordinate: 268  
Predicted y2_coordinate: 166.0, real y2_coordinate: 166
```



Obr.: Klasifikácia znaku F a lokalizácia ruky.

# Výsledky – Klasifikácia znaku AMA a lokalizácia ruky konvolučnými neurónovými sieťami

```
Predicted sign: 14, Real sign: 14  
Predicted x1_coordinate: 198.0, Real x1_coordinate: 184  
Predicted y1_coordinate: 35.0, Real y1_coordinate: 31  
Predicted x2_coordinate: 303.0, Real x2_coordinate: 294  
Predicted y2_coordinate: 137.0, real y2_coordinate: 141
```



Obr.: Klasifikácia znaku O a lokalizácia ruky.

# Výsledky – Porovnanie nášho prístupu s prístupmi iných výskumných skupín

Prístup	Model	Typ úlohy	Vyhodnotenie (Správnosť/MAE)
Naše riešenie	CNN	Klasifikačná	96.69%
		Regresná	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<sup>8,9,10</sup>.

---

<sup>8</sup>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.

<sup>9</sup>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.

<sup>10</sup>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.

4. část  
Diplomová práce

# Zadanie diplomovej práce

**Téma:** Rozpoznávanie znakovkej reči z pohybu rúk pomocou metód hĺbkového učenia.

- Ciele:**
1. Vytvoriť prehľad aktuálne existujúcich metód hĺbkového učenia na spracovanie videa.
  2. Navrhnuť a implementovať kombináciu metód hĺbkového učenia na rozpoznávanie znakovkej reči z pohybu rúk.
  3. Porovnať dosiahnuté výsledky s ďalšími dostupnými štúdiami.

# Údajová sada WLASL



(a) The verb **“Wish”** (top) and the adjective **“hungry”** (bottom) correspond to the same sign.



(c) Signers perform **“Scream”** with different hand positions and amplitude of hand movements.

Obr.: Ukážka vstupných obrázkov<sup>11</sup>.

---

<sup>11</sup>LI, D., et al., 2020. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1459-1469).



# Údajová sada WLASL



Obr.: Diverzita údajovej sady<sup>11</sup>.

---

<sup>11</sup>LI, D., et al., 2020. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1459-1469).

# Problémy diplomovej práce

- Zohľadňovanie všetkých komponentov, z ktorých sa skladá posunok – vstupom je video (séria obrázkov).
- Rozpoznávame už teda slová, nie písmená manuálnej abecedy.

**To vedie k niektorým problémom z pohľadu počítačového videnia, ktoré sú charakteristické pre SLR:**

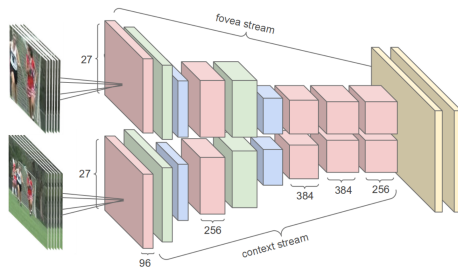
- Rôzne svetlo, citlivosť kamery, pozadie, umiestnenie kamery.
- Niektoré prsty môžu byť mimo záber.
- Detekcia toho, kedy jeden znak končí a druhý už začína.

# Problémy diplomovej práce

- **Výpočtové hľadisko:** pri vstupoch v podobe videí si konvolučné siete vyžadujú príliš veľké časové obdobie na efektívnu optimalizáciu parametrov, pretože sieť nespracováva len jeden obrázok, ale priamo sekvenciu obrázkov.

**Možná optimalizácia na základe dvoch článkov:**

## 1. prístup:



Obr.: 3D CNN s multirozlišením<sup>12</sup>.

<sup>12</sup>KARPATHY, A., et al., 2014. Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

# Problémy diplomovej práce

Cieľom bolo ukázať, že nasledujúca modifikácia architektúry urýchľuje runtime performance:

## Dva vstupné toky:

- **Context stream:** učí sa príznaky z celých obrazov s nízkym rozlíšením.
- **Fovea stream:** dostane menšie okno z celého rámca, ale vo vysokom rozlíšení (*fovea* - časť oka zodpovedná za ostré centrálné videnie).

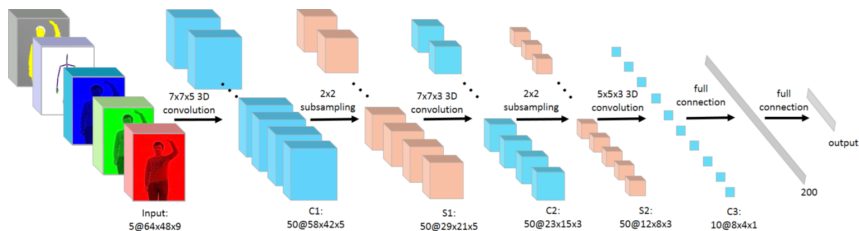
Takáto úprava architektúry viedla k urýchleniu až o 2-4 krát.

# Problémy diplomovej práce

## 2. prístup:

Na vylepšenie výkonnosti sa použilo 5 vstupných tokov:

- 3 RGB kanály.
- Hĺbkové informácie.
- Kľbové informácie.



Obr.: 3D CNN s multikanálmi<sup>13</sup>.

<sup>13</sup>HUANG, J., et al., 2015. Sign language recognition using 3d convolutional neural networks. In 2015 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.

Ďakujem za pozornosť.