

**UNIVERZITA KONŠTANTÍNA FILOZOFA V NITRE**  
**FAKULTA PRÍRODNÝCH VIED A INFORMATIKY**

**ROZPOZNÁVANIE OBJEKTŮ VO VIDEÁCH**  
**DIPLOMOVÁ PRÁCA**

**2023**

**Bc. Johana Heneková**

**UNIVERZITA KONŠTANTÍNA FILOZOFA V NITRE**  
**FAKULTA PRÍRODNÝCH VIED A INFORMATIKY**

**ROZPOZNÁVANIE OBJEKTŮ VO VIDEÁCH**  
**DIPLOMOVÁ PRÁCA**

Študijný odbor:	18. Informatika
Študijný program:	Aplikovaná informatika
Školiace pracovisko:	Katedra informatiky
Školiteľ:	Mgr. Ľubomír Benko, PhD.

Nitra 2023

Bc. Johana Heneková

# ZADANIE ZP

## Rozpoznávanie objektov vo videách

Výskumníci a vývojári v oblasti informačných technológií sa inšpirovali funkciou ľudského zraku a rozhodli sa vytvoriť aplikácie na detekciu videoobjektov, ktoré poskytujú strojom schopnosť analyzovať obrázky a zisťovať objekty, ktoré sa v nich nachádzajú. Najprv vyvinuli protokoly a postupy určené na fungovanie len na obrázkoch. Dnes sa však veci posunuli k video-obrazom. Cieľom takéhoto nástroja je umožniť stroju lokalizovať, identifikovať a klasifikovať objekty, ktoré možno vidieť na vstupných pohyblivých obrázkoch. Stroje a ich postupy neberú do úvahy obrazy ako celok. Na ich analýzu musia snímky rozdeliť a pracovať s pixelmi a ich vlastnosťami. Zvyčajne kombinujú detekciu obrazu a videostopy, aby prišli so svojimi výsledkami.

Cieľom záverečnej práce je predstaviť rôzne prístupy k rozpoznávaniu objektov vo videách a vyhodnotiť ich úspešnosť. V teoretickej časti je žiadúce sa zamerať na rôzne algoritmy rozpoznávania objektov vo videách. V praktickej časti je žiadúce vybrať vhodnú testovaciu sadu videí, na ktorých sa porovnajú a vyhodnotia rôzne prístupy k rozpoznávaniu objektov.

### Charakter práce:

Výskumný – stanovenie predpokladov/hypotéz, metodika výskumu, výsledky výskumu (štatistická interpretácia), interpretácia výsledkov výskumu (vecná interpretácia).

### Predmetové prerekvizity:

Úvod do strojového učenia (1., mgr);

Neurónové siete (1., mgr);

Hĺbková analýza dát (2., mgr).

### Najdôležitejšie kompetentnosti získané spracovaním témy:

vykonávať vedecký výskum;

princípy umelej inteligencie;

poskytnúť vizuálnu prezentáciu údajov;

referovať o výsledkoch analýzy;

vykonať analýzu údajov.

## **POĎAKOVANIE**

Na tomto mieste môže byť vyjadrenie poďakovania napr. školiteľovi resp. konzultantovi za pripomienky a odbornú pomoc pri vypracovaní práce. Nie je zvykom ďakovať za rutinnú kontrolu, menšiu spoluprácu, alebo všeobecné rady.

*NEPOVINNÉ*

# ABSTRAKT

PRIEZVISKO, Meno autora: Názov práce. [Bakalárska/Diplomová práca]. Univerzita Konštantína Filozofa v Nitre. Fakulta prírodných vied a informatiky. Školiteľ: Titul. Meno Priezvisko, Titul. Stupeň odbornej kvalifikácie: Bakalár odboru Aplikovaná informatika. Nitra: FPVaI, 202x. .... s. (uviesť počet strán ZP).

Abstrakt musí byť písaný ako **jeden odsek**. **Abstrakt by mal mať rozsah maximálne jednej strany**. Abstrakt musí byť pre čitateľa zrozumiteľný, aj bez prečítania textu celej práce. **Abstrakt nie je len úvod práce, obsahuje zmienku o cieľoch, postup riešenia, výsledky a prínos**. Je potrebné zachovať základné informácie a súlad abstraktu s textom práce. Treba byť čo najstručnejší, dodržiavať požiadavky na obsahovú stránku abstraktu, ale vyhnúť sa nepresnostiam. Nie je vhodné uvádzať informácie alebo tvrdenia, ktoré sa nevyskytujú v samotnom texte práce. V abstrakte sa používa tretia osoba jednotného alebo množného čísla. Vždy, keď je to možné, treba použiť slovesá v prítomnom čase a v činnom rode. Prispieva to k jasnému, stručnému a pôsobivému písaniu. Trpný rod možno použiť na indikatívne konštatovania a na informatívne konštatovania, ak treba súčasne zdôrazniť osobu, na ktorú sa činnosť zameriava. *Terminológia:* V abstrakte sa používajú plnovýznamové slová z textu, ktoré pomôžu aj pri automatickom vyhľadávaní v textoch. Treba sa vyhnúť neznámym termínom, skratkám, akronymom alebo symbolom alebo ich treba v odôvodnených prípadoch definovať hneď, keď sa prvý raz vyskytnú v abstrakte. Používajú sa jednotky, symboly, terminológia ISO vždy, keď je to možné. Ak nie sú k dispozícii, používajú sa národné normy.

Kľúčové slová: (napr. Informatika. Prírodné vedy.) Minimálne 3 – 5 výrazov. Jednotlivé slová sa oddeľujú bodkou a začínajú sa veľkým písmenom.

*POVINNÉ*

## **ABSTRACT**

LAST NAME, First Name: Title of the thesis. [Bachelor/Master Thesis]. Constantine the Philosopher University in Nitra. Faculty of Natural Sciences and Informatics. Supervisor: Title. First name Last name, title. Degree of Qualification: Bachelor of Applied Informatics. Nitra: FNSaI, 202x. .... p.

The abstract must be written as a single paragraph. The abstract should be a maximum of one page. It must be comprehensible to the reader, even without reading the full text of the paper. The abstract is not just an introduction to the thesis, but it includes a mention of the objectives, the solution procedure, the results and the contribution. It is necessary to keep the background information and consistency of the abstract with the text of the thesis. It is necessary to be as concise as possible, adhering to the content requirements of the abstract but avoiding inaccuracies. It is not appropriate to include information or statements that do not appear in the text of the thesis itself. The abstract should be written in passive voice. The imperfect gender can be used for indicative statements and for informative statements when the person on whom the action is directed needs to be emphasised at the same time. Terminology: the abstract uses full-meaning words from the text, which will also help with automatic text searches. Unfamiliar terms, abbreviations, acronyms or symbols should be avoided or, where justified, defined as soon as they first appear in the abstract. Units, symbols, ISO terminology shall be used whenever possible. Where they are not available, national standards shall be used.

Keywords: Informatics. Key. MS Word.

*POVINNÉ*

# OBSAH

<b>DIPLOMOVÁ PRÁCA .....</b>	<b>2</b>
<b>ZADANIE ZP .....</b>	<b>3</b>
<b>Pod'akovanie.....</b>	<b>4</b>
<b>ABSTRAKT .....</b>	<b>5</b>
<b>ABSTRACT.....</b>	<b>6</b>
<b>Obsah .....</b>	<b>7</b>
<b>Úvod.....</b>	<b>9</b>
<b>1 Analýza súčasného stavu.....</b>	<b>10</b>
1.1 Computer Vision.....	10
1.2 Základné metódy computer vision.....	11
1.3. Modely klasifikujúce vo videách.....	12
1.3.1 CNN.....	13
1.3.2 RNN.....	14
1.3.3 SVM.....	14
1.3.4 3D Konvolučné modely .....	16
1.4 Modely využité pri analýze ERA datasetu.....	17
<b>C3D Model.....</b>	<b>17</b>
<b>P3D Model .....</b>	<b>19</b>
<b>I3D Model .....</b>	<b>20</b>
<b>TRN Model .....</b>	<b>21</b>
1.3.5 Ďalšie modely .....	22
1.4. Porovnanie modelov .....	23
<b>2 Ciele záverečnej práce.....</b>	<b>25</b>
<b>3 Metodika výskumu .....</b>	<b>26</b>
3.1. Porozumenie problematike .....	26
3.1.1. Klasifikácia vo videách.....	26
3.2. Porozumenie dátam.....	26
3.2.1 Dataset ERA .....	27
3.3. Spracovanie dát.....	27
3.4. Výber modelov .....	27
3.5. Analýza dát .....	27

3.6. Evalvácia výsledkov .....	28
3.7. Vyhodnotenie a interpetácia .....	28
<b>Zoznam bibliografických odkazov .....</b>	<b>29</b>
<b>Zoznam príloh .....</b>	<b>33</b>



# ÚVOD

V úvode autor záverečnej práce (ďalej ZP) stručne a výstižne charakterizuje stav poznania alebo praxe v oblasti, ktorá je predmetom ZP a oboznamuje čitateľa s významom ZP. Autor v úvode zdôrazňuje, prečo je práca dôležitá a prečo sa rozhodol spracovať danú tému. Autor v krátkosti uvedie ciele a zámery práce. V úvode autor stručne a jasne:

- charakterizuje súčasný stav poznania problematiky, ktorá je predmetom ZP,
- zdôvodní aktuálnosť témy,
- nastolí problémy, ktoré chce riešiť,
- uvedie v krátkosti ciele, ktoré sa majú v práci dosiahnuť,
- opíše použité metódy a postup riešenia,
- spresní, ktoré hlavné informačné zdroje použil,
- zdôvodní význam riešenia problematiky,
- načrtne stručný obsah jednotlivých kapitol.

V prípade výskumnej práce autor uvedie ako jeho výskum môže prispieť k riešeniu v úvode nastoleného problému.

*ÚVOD PREDSTAVUJE CCA 1 STRANU. ČÍSLOVANIE STRÁN ZAČÍNA OD ÚVODU.*

# 1 ANALÝZA SÚČASNÉHO STAVU

Súčasná doba digitálneho pokroku prináša interakcie medzi človekom a strojom na vyššej úrovni. Jedným z najviac viditeľných aspektov tejto postupnej revolúcie je oblasť umelého videnia, taktiež známa ako computer vision. Táto disciplína umelej inteligencie sa stala súčasťou nášho každodenného života a to najmä v oblasti spracovania obrazov a analýzy videí. Spomenieme napríklad autonómne autá, kde sa detekujú objekty a ich trajektórie (Kortylewski et al. 2020) alebo využitie v chirurgii pri rozpoznávaní jednotlivých úkonov chirurga (Zhang et al. 2021). V rámci tejto širokej oblasti sa sústredíme na schopnosť počítačov a technológií „vidieť“ a teda porozumieť digitálnym dátam. Zameriame sa na modely, ktoré sa rozvinuli vďaka tomuto pokroku a prinášajú automatické spracovanie s neuveriteľnou presnosťou a efektivitou.

## 1.1 COMPUTER VISION

K postupne rozširujúcemu sa využívaniu a dostupnosti kamier a telefónov vznikajú aj veľké množstvá zdieľaných videí. To prinieslo nový dopyt pre modely, aby sa učili a vylepšovali postupne samé. Taktiež aby sa nemuseli vytvárať modely nové od základu. Počítačové videnie pridáva k základom neurónových sietí učenie sa postupne v čase a tým idú do popredia. (Qu et al. 2023)

Počítačové videnie spracováva jednoduché popisy a teda čo vidí na scéne. Vzory obsahujú pochopenie scény a tie sa využívajú v modeloch. Algoritmy modelov vieme rozdeliť do kategórií (Zhou, Zhang, a Konz 2022):

- a) detekcia,
- b) rozpoznanie,
- c) segmentácia,
- d) 3D rekonštrukcia.

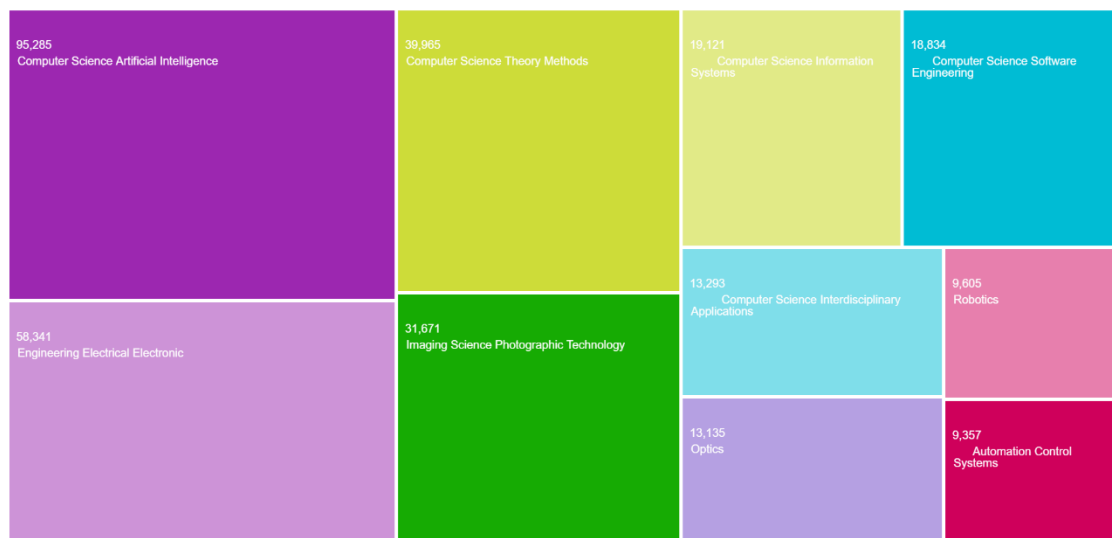
Počítačové videnie, ako obor umelej inteligencie, vieme datovať podľa prvých teoretických konceptov medzi roky 1943 a 1956, kedy cieľom výskumu bolo hlavne porozumieť strojovému učeniu. Prvá konkrétna práca vznikla na MIT s názvom: The Summer vision Project (Papert 1966) a to v roku 1966. V posledných dekádach počítačové videnie napreduje hlavne vďaka posunu v technológiách, ktoré umožňujú vyššiu výpočtovú náročnosť. (Sharma, Gupta, a Jaiswal 2021)

Rozvoj počítačového videnia už od 1950tych rokov výrazne ovplyvnil rôzne oblasti. Jednou významnou oblasťou je priemyselný sektor, kde sa počítačové videnie

čoraz viac využíva pre inteligentnú výrobu v kontexte bezdrôtového pripojenia 5G a priemyslu 4.0. (Li 2022)

Táto technológia tiež zohrala kľúčovú úlohu v automobilovom priemysle, najmä v oblasti bezpečnosti a identifikácie vozidiel, ako aj pri identifikácii a klasifikácii plodín v poľnohospodárstve. (Díaz a Kurka 2020)

Z portálu Web of Science <sup>1</sup> vidíme, že pojem „computer vision“ je veľmi aktuálnou témou s veľkým množstvom výskumných prác v rôznych odvetviach.



## 1.2 ZÁKLADNÉ METÓDY COMPUTER VISION

Pri vyhodnocovaní sa nemôžu modely upriamovať iba na cieľovú udalosť alebo objekt. Je dôležité rozoznávať kontext a prepájať obsah medzi vizuálnymi ako aj nevizuálnymi objektami a udalosťami počas celého videa. (Wang a Zhu 2023)

Wang a Zhua (2023) rozlišujú tri druhy kontextu v počítačovom videní a to:

- priestorový kontext,
- časový kontext,
- iný kontext.

Na kontext sa potom dá podľa ich názoru dá pozerieť na rôznych úrovniach a to menovito na:

- úroveň predchádzajúcich znalostí
- úroveň globálnych znalostí
- alebo taktiež úroveň lokálnych znalostí

---

<sup>1</sup> <https://www.webofscience.com/wos/woscc/basic-search>

Následne vieme deliť jednotlivé modely podľa týchto kategórií. (Wang a Zhu 2023)

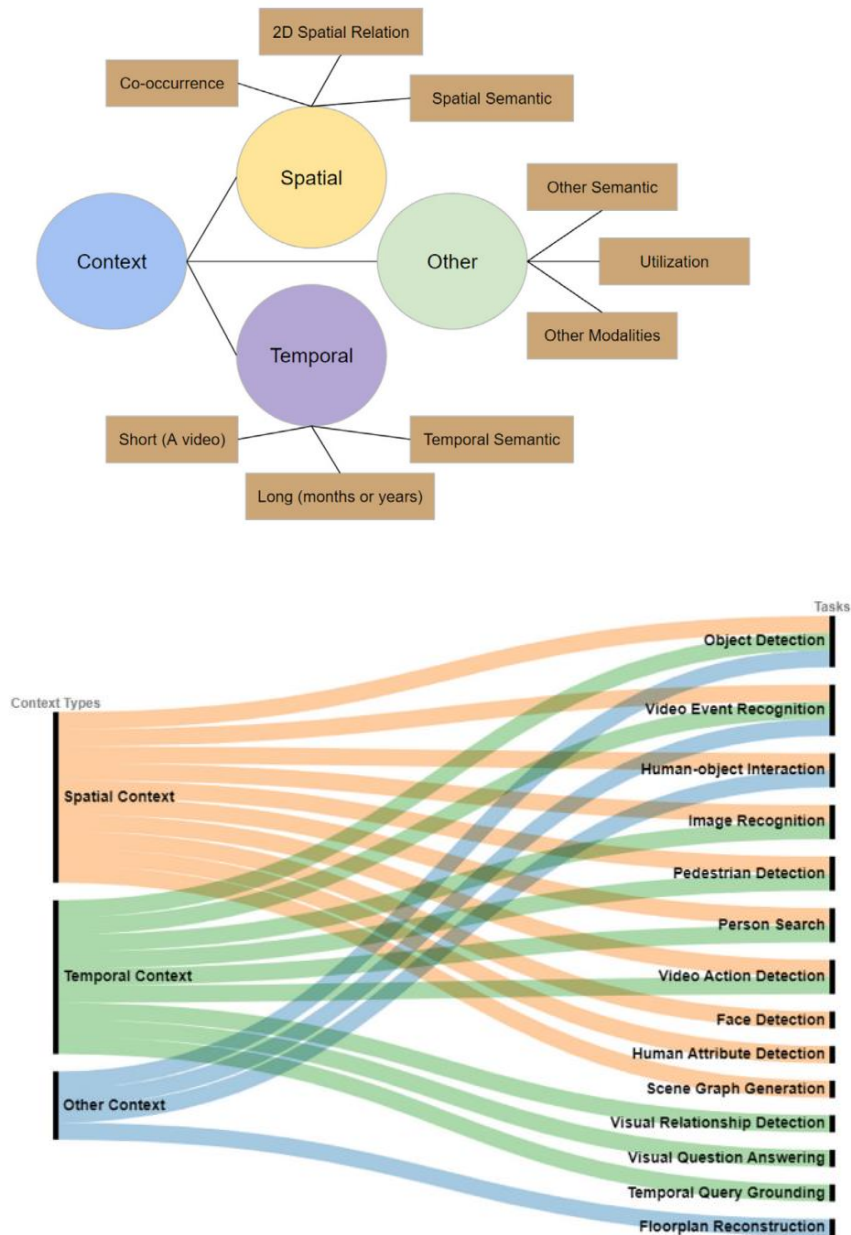


Fig. 20. Summary of the three major context types and what tasks they are employed in.

### 1.3. MODELÝ KLASIFIKUJÚCE VO VIDEÁCH

Modely na klasifikáciu videí predstavujú algoritmy alebo architektúry navrhnuté tak, aby kategorizovali a pochopili kontext videí. Tieto modely využívajú rôzne techniky ako napríklad hĺbkové učenie pre klasifikáciu. A to na základe vizuálnych vlastností, čiastočnej informácie alebo textového podkladu. (Ng et al. 2015)

### 1.3.1 CNN

Hĺbkové učenie sa stalo silným nástrojom pre klasifikáciu vo videách. Využíva sa tu najmä CNN (Convolutional Neural Network) na extrakciu vysoko úrovňových vlastností zo snímok videí a vykonávanie predikcie. Momentálne najviac rozšíreným využitím CNN je rozpoznávanie akcií vo videách. (Wu, Sharma, a Blumenstein 2017)

V športoch sa napríklad využívajú na rozoznanie typu športu a aktivity v reálnom čase(obrázok). Na základe monitorovania ľudí, konkrétne v tomto prípade ich stred tela a štvorcového mapovania, sa vykonávajú pozorovania ich správania. Takéto modelovanie sa nevyužíva iba v analýze športovcov a ich výkonov, ale aj vo virtuálnej realite a monitorovaní správania ľudí. (Zhengfeng 2022)

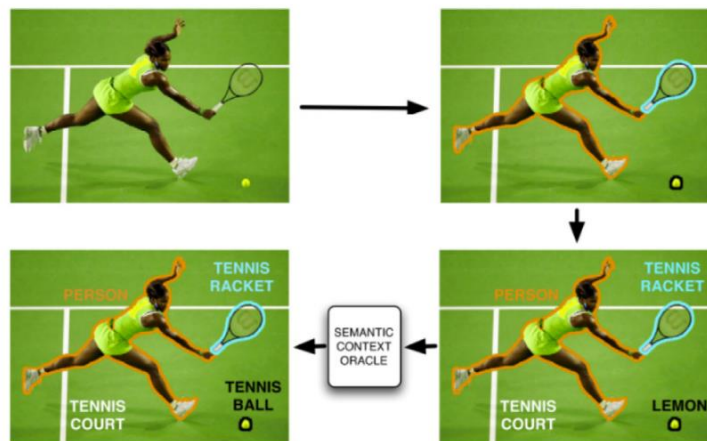


Fig. 9. Objects are perfectly segmented by using contextual relations between objects' labels to help satisfy semantic constraints (Rabinovich et al., 2007).

Výkony basketbalistov napríklad skúmali Delextrat a Cohen (2009), ktorí vo svojom výskume popisujú jednoznačné využitie do budúcnosti na personifikáciu stratégie pre jednotlivých hráčov. Vďaka tomuto poznatku vedia dosiahnuť lepšiu ofenzívu ako aj defenzívu u hráčov a dodatočné predpovedanie dynamiky medzi spoluhráčmi v tíme. Kombinuje pri tom predpovedanie hráčovho pohybu ako aj trajektóriu letu jeho hodov. Pri väčšom preskúmaní tejto problematiky by mala analýza videí veľký dopad na športový priemysel. (Delextrat a Cohen 2009)

Ďalšie využitie na základe pozorovania ľudí sa využíva v bezpečnostnom monitorovaní. Vďaka modelom sa znižujú náklady na obstarávanie týchto systémov a ich monitorovanie. Najväčšou prekážkou je momentálne problém, že sa modely zameriavajú na aktivity jednej osoby, avšak v tomto odvetí je potrebné monitorovať všetky osoby na danej snímke videa(obrázok). Jen-Kai Tsai a kol. (2020) riešia túto

problematiku a výsledok ich výskumu popisujú v článku, kde vidia využitie modelu I3D. (Sensors | Free Full-Text | Deep Learning-Based Real-Time Multiple-Person Action Recognition System n.d.)



### 1.3.2 RNN

RNN (Recurrent Neural Network) využívajú výskumníci napríklad v detekovaní podvodných videí. Keďže sa videá využívajú v mnohých oblastiach ako dôkazový materiál: právo, forenzné štúdie, novinárstvo a iné, je nesmierne dôležité overiť ich autenticitu. Autori popisujú využitie siamese based RNN popri modeli I3D a na dvoch datasetoch dosahujú až 86.6%(dataset VIRAT - Image Retrieval and Analysis Tool) a 93%(dataset MFC - Media Forensic Challenge) presnosti. V princípe modely využívajú na hľadanie duplikovaných snímok. (Duplicate Frame Video Forgery Detection Using Siamese-based RNN n.d.)

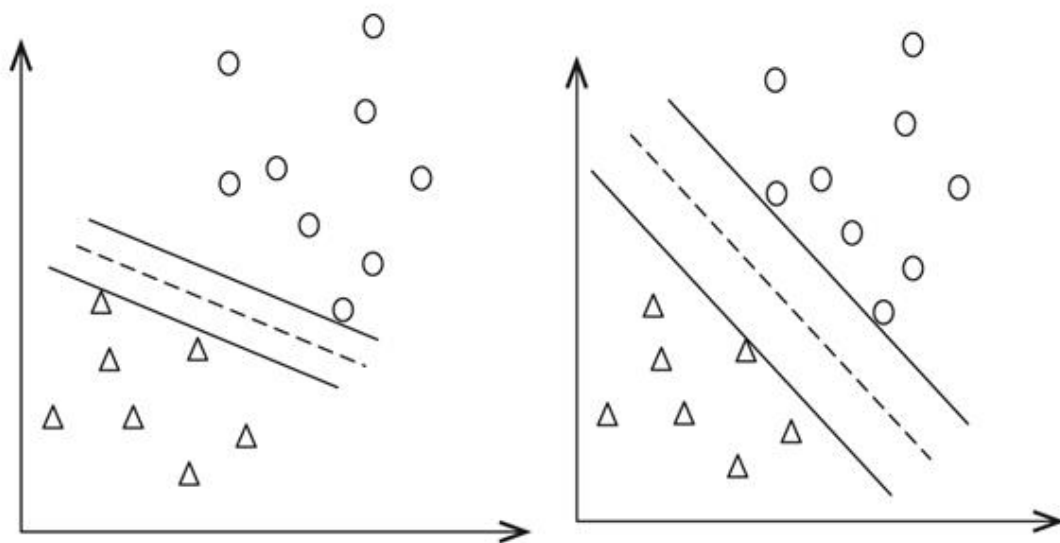
Siamese-based RNN odkazuje na architektúru rekurentnej neurónovej siete, ktorá zahŕňa koncept siamských sietí. Neurónové siete tu spolupracujú a zdieľajú si váhy a architektúru. Toto im umožňuje spracovávať viaceré vstupy paralelne a učiť sa podobnosti alebo rozoznávať rozdiely. (Mueller a Thyagarajan 2016)

### 1.3.3 SVM

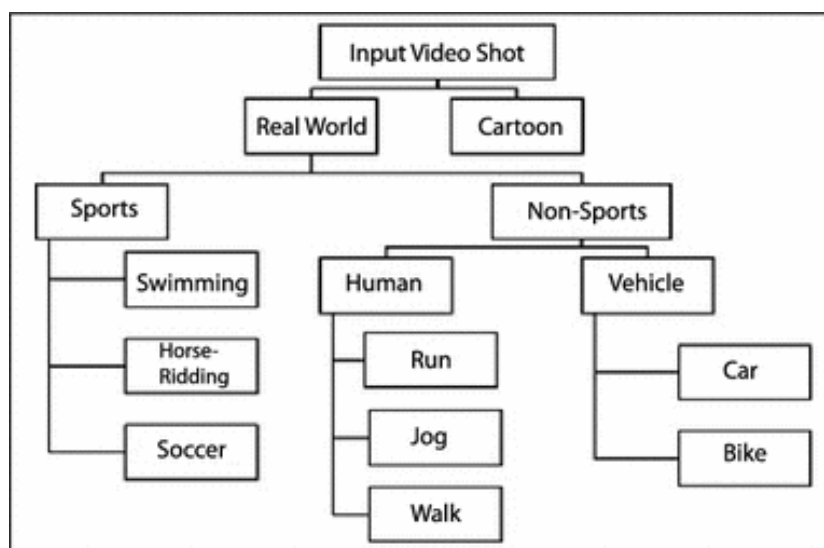
SVM alebo Support Vector Machine je algoritmus strojového učenia, ktorý sa často využíva pre klasifikačné a regresné úlohy. Pri videách sa momentálne skúma využitie SVM na kategorizáciu druhov videí a to napríklad na reklamy, kreslené rozprávky, hudobné videá, správy a športové videá (podľa výskumu Zhang Min-qinga a

Li Wen-pinga (2021)). Pri detekcii využívajú aj zvukovú zložku priloženú k videu. Autori sa dodatočne snažia odhaliť aj teroristický alebo násilný obsah.

Princípom fungovania SVM je najmä štatistický prístup k nájdeniu najlepšej plochy. Plocha sa hľadá v pôvodnom priestore alebo vo vyššej dimenzii po projekcii. Cieľom je maximalizovať medzeru medzi dvoma kategóriami, čím sa znižuje interval neistoty pri generalizácii a tak sa minimalizujú skutočné riziká (obrázok). (Min-qing a Wen-ping 2021)

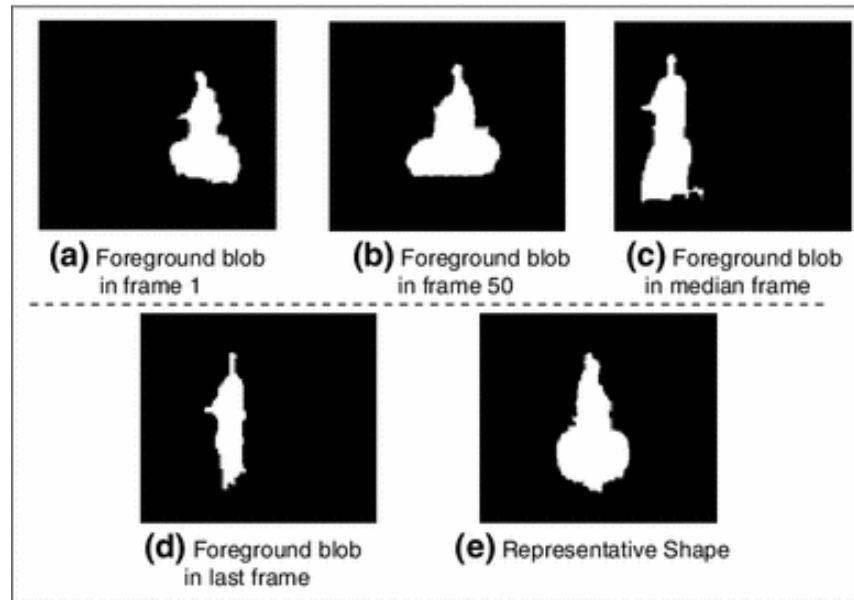


Podobne využili SVM aj Chattopadhyay a Maurya (2013) aby preskúmali možnosť rýchleho označovania videí do kategórií pomocou kľúčových slov (tagov). (obrázok) (Chattopadhyay a Maurya 2013)



Pri týchto výskumoch sa videá spracovávajú tak, že sa najskôr dajú do unitárneho farebného spektra a následne sa snímky konvertujú do šedých farieb kvôli

výpočtovej náročnosti. Následne sa pracuje s tvarmi. Pri tomto kroku sa rozlišuje pozadie a objekty záujmu, tie sa segmentujú pričom vzniká akýsi útvar.(obrazok) Tento útvar sa oddeľuje od pozadia a neskôr sa, na základe ich relatívneho centra, útvary prekrývajú a pozorujú zmeny. Veľkou pomocou sa ukázalo porovnávať a zaznamenávať aj textúry objektov. (Chattopadhyay a Maurya 2013)



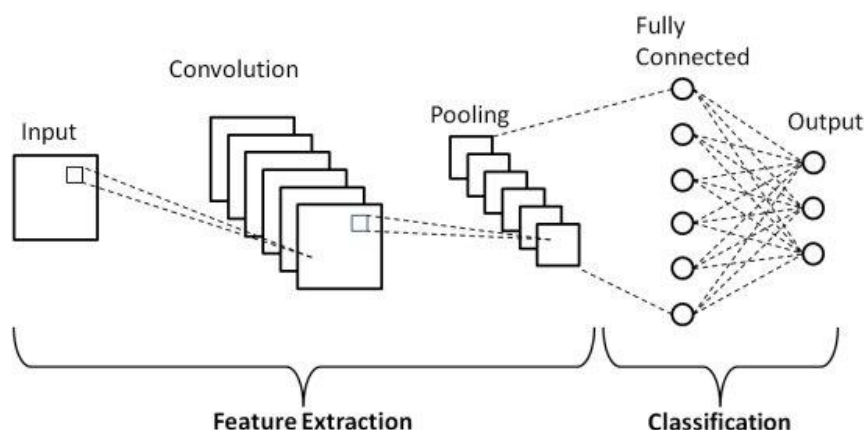
Po tomto spracovaní sa hľadajú rozdiely medzi kategóriami pričom vo výsledku autori poukazujú na obrovské využitie v rozpoznávaní kontextu videa avšak do budúcnosti odporúčajú pridať ďalšie médiá ako zvuk a text. (Chattopadhyay a Maurya 2013)

### 1.3.4 3D Konvolunčné modely

CNN je dopredná neurónová sieť, ktorá berie vstupné dáta a spracúva ich cez niekoľko vrstiev neurónovej siete, pričom každá z nich vykonáva tri operácie: konvolúcia, bodová nelinearita a zhlukovanie. Na základe tejto architektúry dokáže model extrahovať užitočné informácie. (Gama et al. 2018)

Z obrázku vidíme, že sa CNN skladá z dvoch hlavných častí a to extrakcia vlastností a klasifikácia. Pri extrakcii vlastností sa konvolunčná vrstva snaží na základe filtrov(mriežkové matice), ktoré sa posúvajú(konvulujú), získať číslo pre pozíciu prvku za pomoci násobenia. Toto číslo hovorí o sile prvku. CNN majú dôležitú vlastnosť a to zdieľanie váh, čo znamená, že sa každý filter požíva na každý vstupný prvok. Pri časti pooling ide o väčšinové hlasovanie. Tu sa vyfiltrujú jednotlivé prvky, ktoré následne putujú do neurónových sietí. (Mesuga a Bayanay 2021)





## 1.4 MODEL Y VYUŽITÉ PRI ANALÝZE ERA DATASETU

Autori Mou, L., Hua, Y., Jin, P. a Zhu, X. X., ktorí spracovali videá do datasetu ERA Event Recognition in Aerial Videos - Rozpoznávanie udalostí v leteckých videách)<sup>2</sup> popisujú využitie na modeloch, ktoré vybrali.

### C3D Model

C3D(Convolutional 3D model - Konvolučný 3D model) pracuje s použitím 3D konvolúcií na extrakciu priestorovo časových prvkov z videí. Na rozdiel od dvojrozmerného priestoru snímok zohľadňuje dimenziu navyše a berie z nej údaje, čiže zachytáva informácie o pohybe a čase. (Hadidi et al. 2020)

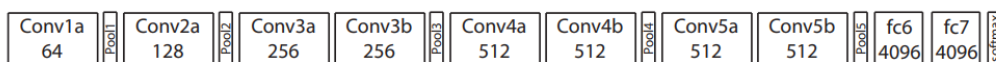


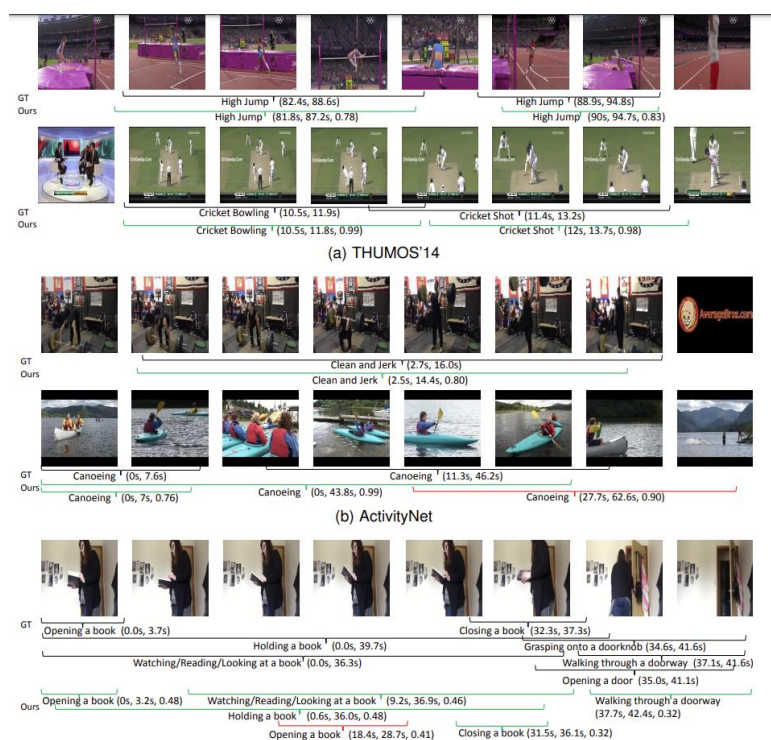
Figure 3. **C3D architecture.** C3D net has 8 convolution, 5 max-pooling, and 2 fully connected layers, followed by a softmax output layer. All 3D convolution kernels are  $3 \times 3 \times 3$  with stride 1 in both spatial and temporal dimensions. Number of filters are denoted in each box. The 3D pooling layers are denoted from pool1 to pool5. All pooling kernels are  $2 \times 2 \times 2$ , except for pool1 is  $1 \times 2 \times 2$ . Each fully connected layer has 4096 output units.

Vďaka stratégii, kedy nespracováva každý snímok, ale dané číslo snímok v intervale, sa zvyšuje jeho efektívnosť a znižuje výpočtová náročnosť. (Wang et al. 2016)

Podľa výskumu, ktorý vykonal Du Tran a jeho kolegovia na učení C3D, zistili, že ak zobrali dataset so zameraním (videá zo športu, rozpoznávanie aktivít vo videách, rozpoznávanie objektov a scény), tak C3D prakticky vždy bol najlepší alebo jeden z najlepších modelov. A to vo výskume použili špecializované modely na dané úlohy. (Tran et al. 2015)

<sup>2</sup> [https://lcmou.github.io/ERA\\_Dataset/](https://lcmou.github.io/ERA_Dataset/)

Model C3D sa využíva v oblasti počítačového videnia pre úlohy rozpoznávania akcií a klasifikácie videí. Aplikuje sa v rôznych oblastiach vrátane rozpoznávania športových gest, detekcie udalostí vo videách, detekcie depresie z tvárových výrazov a monitorovania spánku.



Najrozšírenejšie využitie má pri rozpoznávaní akcií, kde tento model ukázal svoju schopnosť prekonať iné často využívané modely ako sú HOF(Histogram of Optical Flow) a MBH(Motion Boundary Histogram) a to z hľadiska rozpoznávania pohybov (Wang a Schmid 2013). Shang taktiež popisuje zlepšenie výkonov s použitím modelu ActionVLAD. (A Multimodal Pairwise Discrimination Network for Cross-Domain Action Recognition | IEEE Journals & Magazine | IEEE Xplore n.d.)

HOF a MBH používajú histogramy na rozpoznávanie aktivít vo videách. HOF využíva výpočty optického toku medzi susednými snímkami. Konkrétne informácie o pohybe získavajú rozdelením obrazu snímky do priestorových buniek a kvantifikáciou optického toku do rôznych binov. Výsledný histogram reprezentuje distribúciu smerov pohybu vo videu. (Li et al. 2016)

MBH naopak zachytáva hranice pohybu vo videu. Počíta gradient optického toku a kvantifikuje orientácie gradientu do rôznych binov. V tomto prípade histogram reprezentuje distribúciu hraníc pohybu vo videu. (Li et al. 2016)

V praxi sa využíva aj kombinácia HOF alebo MBH s C3D modelom a tu dodávajú doplnkové vlastnosti k priestorovým informáciám zachyteným modelom C3D. Avšak sa ukazuje, že C3D model jednoducho prekonáva tieto modely aj ak pracuje osamote a to pri viacerých testoch výkonu. (Tran et al. 2015)

V oblasti medicínskeho výskumu sa C3D využíva pri pacientoch, ktorým bola transplantovaná oblička. Podľa výskumu (Comoli et al. 2016) by mal byť tento model schopný predikovať rizika odmietnutie transplantátu.

Pri výskumoch v oblasti počítačovej vedy sa vyvíja variácia na model C3D: MV-C3D pre 3D konvolučné neurónové siete. Tento nový model preukazuje svoju prispôsobivosť v reálnych situáciách, najmä pri analýze 3D rotovaných reálnych obrázkoch. Čiže sa dá tento model využiť v počítačovom videní, najmä pri rozpoznávaní objektov a pochopeniu scény. (MV-C3D: A Spatial Correlated Multi-View 3D Convolutional Neural Networks | IEEE Journals & Magazine | IEEE Xplore n.d.)

Nájdeme využitie aj v imunológii, kde C3D zlepšuje vytváranie protilátok na vírusy prvého typu. Čiže sa dá potencionálne vyžiť na vytváranie stratégií pri tvorbe vakcín a imunoterapie pri vírusových infekciách. (MV-C3D: A Spatial Correlated Multi-View 3D Convolutional Neural Networks | IEEE Journals & Magazine | IEEE Xplore n.d.)

### **P3D Model**

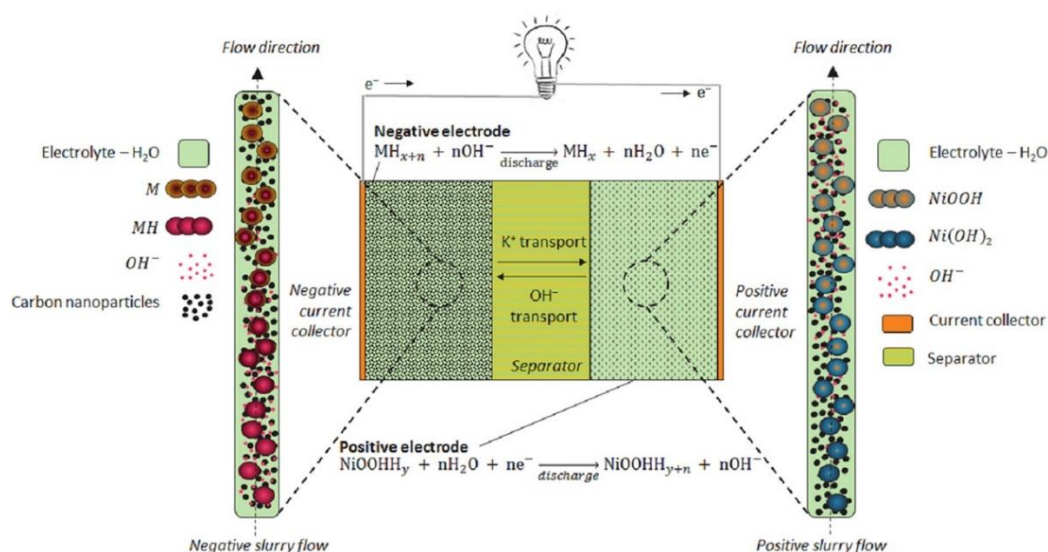
PSD ( Pseudo 3D model) model pozostáva z trojdimenzionálnych blokov, ktoré nahrádzajú konvolučné jadrá, čím uľahčujú výpočtový proces. Ide teda o nový výpočtový prístup, ktorý eliminuje potrebu opätovného výpočtu variačných komponentov. (Zhang et al. 2010)

P3D model bol porovnávaný s C3D a predstavuje menšiu výpočtovú náročnosť a predstavuje menšiu obťažnosť pri tréningu a nižšiu možnosť preučenia. (Zhang et al. 2020)

P3D model nachádza využitie v mnohých sférach. Za jednu z najviac zaujímavých považujeme výskum v oblasti batérií.

Jedna zo štúdií od Chayambuka a jeho kolegov sa zameriava na modelovanie a návrh polotuhých článkových batérií (SSFBs) pomocou tohto pseudo trojrozmerného modelu – P3D. Model bol použitý na simuláciu difúzie látok, ktoré prezentovali jediný transportný mechanizmus v aktívnych častiach SSFBs. P3D umožnil znázorniť časovo

závislé profily napätia, distribúciu prúdu a distribúciu stavu napätia v aktívnych častiach. (Chayambuka, Fransaer, a Dominguez-Benetton 2019)



Využitie modelu nájdeme aj v biológii, pri génových štúdiách, kde sa využíva napríklad na pochopenie organizácie a funkcionality genómov ohľadom zdravia a chorôb populácie. (Full article: 3D genome organization in health and disease: emerging opportunities in cancer translational medicine n.d.)

V medicíne má P3D model úžitok pri modelovaní geometrie genómov, kde sa pozorujú bunky a ich dopad na zdravie ľudí. (Full article: 3D genome organization in health and disease: emerging opportunities in cancer translational medicine n.d.)

### I3D Model

I3D model predstavuje architektúru, ktorá zväčšuje dvojrozmerné konvolučné siete aby spracovával informácie času a priestoru z videí. V podstate využíva 2D siete, ktorým pridáva filtre a presúva ich do 3D. Čiže napríklad štvorcový filter sa stáva kubický. (Wei et al. 2022)

V oblasti rozpoznávania aktivít vo videách sa tento model využíva v chirurgii, kde sa rozpoznávajú postupy pri práci. V štúdií (Zhang et al. 2021) bol model I3D, konkrétne architektúra Inflated 3D ConvNet (na základe ktorej je vybudovaný model I3D) (Carreira a Zisserman 2017), využitá na rozpoznávanie chirurgických postupov v rámci operácií Tubulizácie žalúdka. Model bol trénovaný pomocou metódy ohnisková stráta (Focal loss) a dosiahol presné rozpoznávanie rôznych chirurgických krokov.

Focal Loss je špeciálny druh loss funkcie, ktorá sa zaoberá problémom nerovnováhy tried pri úlohách detekcie objektov. Keďže pri tejto problematike väčšina snímok pozostáva z pozadia, zatiaľ čo pre nás je dôležitá detekcia objektu záujmu, môže

nerovnováha spôsobovať malú úspešnosť detekovania objektov popredia pri modeloch. Na klasifikáciu používa špeciálnu váhu, ktorá znižuje hodnotu pri ľahko odhadnuteľných prípadoch a teda sa trénuje na ťažkých a viac ojedinelých prípadoch tried. Týmto zlepšuje výkony v triedach menších. (Lin et al. 2017)

Pri využívaní architektúry I3D v praxi sa poukazuje na tri kľúčové výhody: využíva menej parametrov, znižuje výpočtové náklady a dokáže naučiť priestorovo-časové funkcie pre využitie v datasetoch. (Shi et al. 2020)

I3D model sa využíva pri bezpečnostných aplikáciách v reálnom svete. Napríklad v autonómnych autách a pri detekcii osôb na bezpečnostných záznamoch. (Kortylewski et al. 2020)

### TRN Model

TRN model (Temporal Relation Network – dočasne relačný model) vznikol na princípe relačných posudkov v čase (Temporal relational reasoning). Ide o schopnosť prepojiť si zmysluplné transformácie objektov alebo entít v čase. Je to vlastnosť, ktorá sa prisudzuje inteligentným bytostiam. Model TRN je vytvorený tak, aby sa vedel efektívne učiť a interpretovať vzťahy medzi snímkami vo videu a to vo viacerých časových vrstvách. Na obrázku vidíme ako si ľudský mozog ľahko doplní kontext medzi obrázkami, ale pre modely je to náročné. Tento problém sa snaží vyriešiť model TRN za pomoci zisťovania vzťahov medzi snímkami. (Zhou et al. 2018)



Veľkou výhodou TRN modelu oproti iným je jeho ľahšie pochopenie pre vyvodenie výsledkov. (Zhou et al. 2018)

Vo výskumoch neurovedy bol tento model skúmaný pri regulácii senzorického spracovania, pozornosti a kognície. (Li et al. 2020) Taktiež, na základe dostupných referencií, sa využíva podobne ako P3D model v génovej vede. Tam pomáha skúmať komplexné interakcie medzi samotnými génmi a ich regulačnými prvkami. Týmto nám



odkrývajú pohľad na mechanizmy, ktoré súvisia s prejavovaním génov a bunkovými procesmi. (Terney et al. 2008)

V iných odvetviach TRN nachádza využitie napríklad pri počítaní energetickej náročnosti budov. Model využíva náhodné lesy a neurónové siete na to, aby spravil predikcie s vysokým rozlíšením budov. Pracuje s historickými dátami a hľadá vzory a vzťahy ktoré pridávajú k vysokej spotrebe budov. (Ahmad, Mourshed, a Rezgui 2017)

### 1.3.5 Ďalšie modely

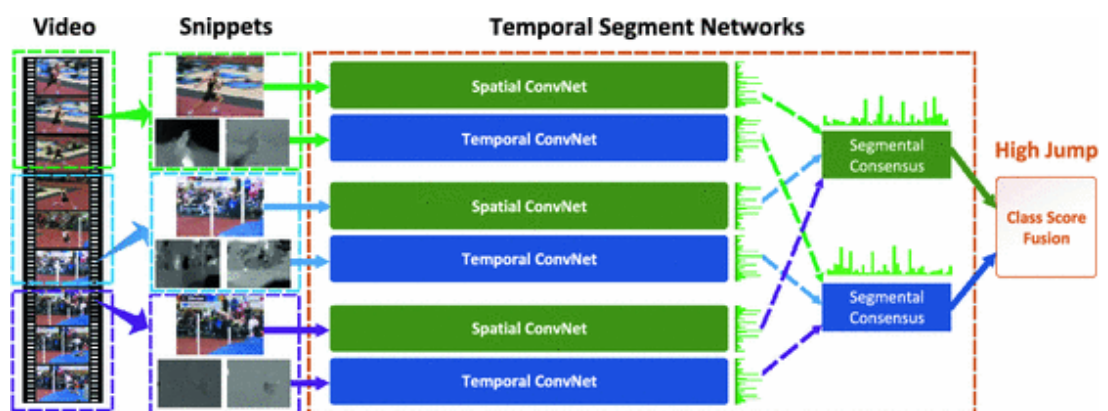
V dnešnej dobe sa výskumníci snažia modely zlepšovať a za týmto cieľom vznikli aj modely TSN a LSTM.

#### TSN model

TSN model bol vyvinutý na základe potreby rozoznávať dlhodobé aktivity a naplniť potrebu trénovať modely na veľkom množstve videí. Autori videli veľký potenciál na zlepšenie a vyvinuli model TSN alebo Temporal Segment Network (Sieť využívajúca časových segmentov). Tento rámec(framework) extrahuje krátke úryvky z dlhej sekvencie pôvodného videa. Využíva sa na to vzorkovacia stratégia a to nie s hustým vzorkovaním. Vďaka tomuto je model schopný pracovať s dlhými videami a zachovať si relevantné informácie za menšej výpočtovej záťaže.

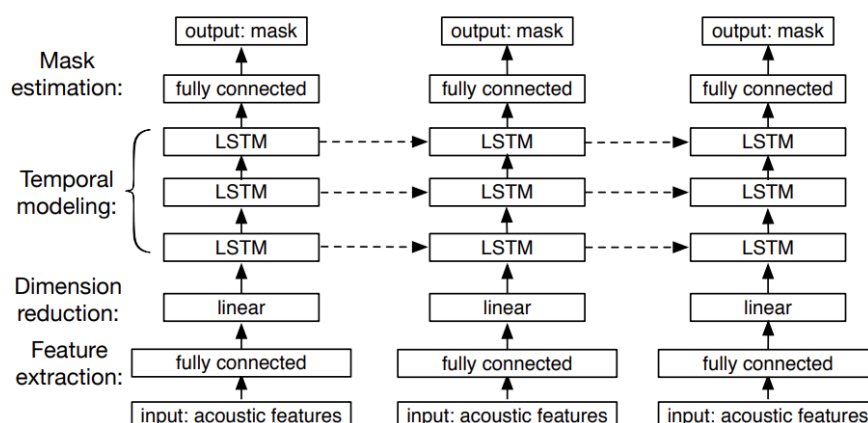
Na obrázku je popísaná segmentácia videa a náhodné vybrané úryvky, na ktorom sa vykonáva následne modelovanie. Na to sa tu využíva Spatial and Temporal ConvNet čiže siete budujúce na základe priestoru a času. (Wang et al. 2016)

Využitie sa momentálne najviac priraduje k zlepšeniu výkonov pri rozpoznávaní aktivít vo videách. (Wang et al. 2016)



## LSTM

Ďalším rozšíreným modelom je LSTM- Long Short-Term Memory model alebo dlhodobá krátkodobá pamäť. Ide o rekurentnú neurónovú sieť, ktorá ale zachytáva a modeluje dlhodobé súvislosti v sekvenčných dátach.



LSTM má využitie v rozpoznávaní obsahu videa a jeho popísaní. Zaznamenáva najviac dôležité momenty a na ich základe hodnotí obsah. LSTM sa využíva najmä pri dlhých videách. (Zhang et al. 2016)

## 1.4. POROVNANIE MODELOV

**Table 4**  
Video-based context integration.

Methods	Tasks	DNN models	Context types	Context levels	Mechanisms
Context R-CNN (Beery et al., 2020)	Object detection	ResNet	Spatial, Temporal	Global, Local	Memory bank
Tube feature aggregation network (Wu et al., 2020)	Pedestrian detection	ResNet	Spatial, Temporal	Local	Feature aggregation
Contextual graph representation learning (Yan et al., 2019)	Person search	ResNet/GCN	Spatial, Temporal	Global, Local	GCN
Contextual boundary-aware framework (Wang et al., 2020)	Temporal query grounding	Custom	Spatial, Temporal	Global, Local	Self attention
Hierarchical temporal network (Yuan et al., 2019)	Temporal query grounding	Custom	Temporal	Global, Local	Semantic conditioned model
Spatio-temporal progressive learning (Yang et al., 2019)	Video action detection	VGGNet	Spatial, Temporal	Global, Local	GCN
Spatio-temporal structural model (Zhu et al., 2013)	Video event recognition	/	Temporal	Local	Structural activity model
Hierarchical context learning (Wang and Ji, 2015, 2016)	Video event recognition	Custom	Spatial, Temporal	Prior knowledge, Global, Local	Hierarchical context model

Tabuľka popisuje modely, ktoré využívame v tejto práci. Jej stĺpce sú:

- Model: Predstavuje názov daného modelu
- Úloha: Akú základnú úlohu spracováva model
- DNN model: Na akom type neurónovej siete alebo architektúry je model postavený.
- Typ kontextu: Ak sa model zameriava na priestorový, temporálny kontext, poprípade ich kombináciou.

- Úroveň kontextu: Či model zohľadňuje globálny (kontext celej scény), lokálny (kontext menšej časti scény) kontext alebo ich kombináciu.
- Mechanizmus: Popis aký mechanizmus alebo techniky využíva model na vykonanie úlohy počítačového videnia.

Model	Úloha	DNN model	Typ kontextu	Úroveň kontextu	Mechanizmus
C3D	Rozpoznávanie vo videách	3D Konvolučné Siete	Temporálny	Lokálny, globálny	Temporálna pamäťová banka
P3D	Rozpoznávanie vo videách	3D Konvolučné Siete	Temporálny	Lokálny, globálny	Temporálna pamäťová banka
I3D	Rozpoznávanie vo videách	Nafúknuté 3D Konvolučné Siete	Temporálny	Lokálny, globálny	Samopozornosť (Self Attention)
TRN	Rozpoznávanie vo videách	Siete s temporálnymi vzťahmi	Temporálny	Lokálny, globálny	Grafické konvolučné siete (GCN)
TSN	Rozpoznávanie vo videách	Siete s temporálnymi vzťahmi	Temporálny	Lokálny, globálny	Temporálny hierarchický kontextový model
LSTM	Sekvenčné modelovanie	Rekurentné neurónové siete	Temporálny	Lokálny, globálny	Pamäťové bunky a brány



## 2 CIELE ZÁVEREČNEJ PRÁCE

Hlavným cieľom práce je porovnať kvalitu modelov na tému klasifikácie vo videách. Tieto poznatky neskôr využijeme v praxi, kde porovnáme modely na datasete ERA.

Naše čiastkové ciele sú:

- Porovnať modely medzi sebou a nájsť ich najlepšie a najhoršie využitia na konkrétnych druhoch videí.
- Implementovať modely použité na datasete ERA a overiť ich výsledky.
- Zhodnotiť modely.
- Použiť ďalšie modely, ktoré sú momentálne využívané, na datasete ERA.
- Porovnať a interpretovať výsledky, ktoré nadobudneme pri práci s modelmi.
- Vizualizovať výsledky a použiť štatistické metódy na ich porovnanie.

### 3 METODIKA VÝSKUMU

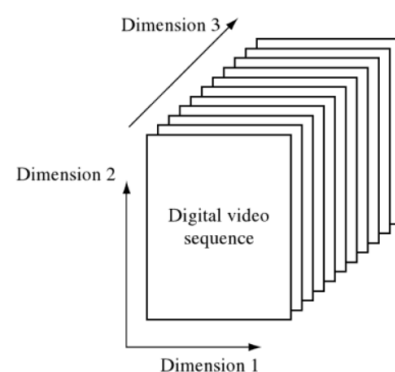
V tejto kapitole predstavíme metódy a postupy, ktorými sme sa riadili počas nášho výskumu. Pridržiavali sme sa metodiky CRISP-DM, ktorú sme si upravili pre naše potreby.

#### 3.1. POROZUMENIE PROBLEMATIKE

V prvom rade sme potrebovali porozumieť ako funguje klasifikácia vo videách a v čom sa líši od klasifikácie napríklad obrázkov a prečo je užitočná. Využitie sme popísali v kapitole číslo jeden.

##### 3.1.1. Klasifikácia vo videách

Pri klasifikácii vo videách ide o proces kategorizácie do rôznych tried alebo kategórií na základe obsahu a charakteristík. Cieľom je automaticky analyzovať a pochopiť informácie, ktoré sa nachádzajú vo videu. (Saddam Bekhet 2021) Oproti klasifikácii v obrázkoch obsahujú videá viaceré snímky, ktoré majú medzi sebou informačné prepojenie. Obsahujú taktiež aj viac informácií, ktoré vieme využiť ako napríklad zvuk. Tieto dodatočné informácie umožňujú komplexnejšie porozumenie obsahu a môžu zlepšiť presnosť klasifikácie. (Karpathy et al. 2014) Na obrázku môžeme vidieť, že sa videá musia brať ako viac dimenzionálne médiá, pretože v nich ide o kontext v čase. (Bovik 2009)



#### 3.2. POROZUMENIE DÁTAM

Pri výbere datasetu bolo pre nás kľúčové myslieť na tieto parametre:

- Aby dataset mal dostatočný počet samostatných videí a tie boli rôznorodé.

- Aby sme nemuseli zostavovať dataset manuálne, čiže zbierať videá po internete.

- Aby sme mohli využiť dataset v našej práci a nevzťahovali sa na videá autorské práva.

Za týchto podmienok sme si vybrali voľne dostupný dataset ERA, čiže dataset na rozpoznávanie udalostí v leteckých videách.

### **3.2.1 Dataset ERA**

Ako popísal Mou, Hua, Jin a Zhu v ich práci: Dataset ERA<sup>3</sup> (A dataset and deep learning benchmark for event recognition in aerial videos) predstavuje významný prínos v oblasti rozpoznávania udalostí vo videách natočených zo vzduchu. Pozostáva z 2864 videí, ktoré zaznamenávajú udalosti, získaných z YouTube, pričom každé video je označené jednou z 25 rôznych tried udalostí. Dataset predstavuje benchmark pre modely na rozpoznávanie udalostí. (Mou et al. 2020)

Využitie datasetu ERA

- Kde sa už dataset ERA využil, v akých výskumoch a ako ho autori zhodnotili pre ich potreby

Našou ďalšou úlohou bolo porozumieť aké dáta sú v datasete a aký majú formát.

- Aké dáta obsahuje dataset, v akej forme sú, aké druhy videí, aké dĺžky, ako sú delené, obsahujú zvuky, text,...

### **3.3. SPRACOVANIE DÁT**

- Bolo potrebné videá nejako spracovať? Napríklad spraviť nejaký umelý šum, rozdeliť ich, skrátiť, pomenovať,...

### **3.4. VÝBER MODELOV**

Pri modeloch sme si spravili prieskum aké sa využívajú v dnešnej dobe plus sme využili modely, ktoré aj samotní autori datasetu ERA<sup>4</sup> používali na modelovanie.

- Popísať v skratke prečo ešte

### **3.5. ANALÝZA DÁT**

- Aké formy analýzy zvolíme a prečo.

---

<sup>3</sup> [https://lcmou.github.io/ERA\\_Dataset/](https://lcmou.github.io/ERA_Dataset/)

<sup>4</sup> [https://lcmou.github.io/ERA\\_Dataset/](https://lcmou.github.io/ERA_Dataset/)

### **3.6. EVALVÁCIA VÝSLEDKOV**

- Aké štatistické metódy si zvolíme, ako budeme hodnotiť modely a prečo sú to vhodné spôsoby hodnotenia.

### **3.7. VYHODNOTENIE A INTERPETÁCIA**

- Aké predpoklady si zvolíme a ako vieme na ich základe interpretovať výsledky, ktoré dosiahneme.

## ZOZNAM BIBLIOGRAFICKÝCH ODKAZOV

- Kortylewski, Adam et al. 2020. “Compositional Convolutional Neural Networks: A Robust and Interpretable Model for Object Recognition under Occlusion”. <http://arxiv.org/abs/2006.15538> (02. december 2023).
- Zhang, Bokai et al. 2021. “Surgical Workflow Recognition with 3DCNN for Sleeve Gastrectomy”. *International Journal of Computer Assisted Radiology and Surgery* 16(11): 2029–36.
- Qu, Haoxuan et al. 2023. “Recent Advances of Continual Learning in Computer Vision: An Overview”. <http://arxiv.org/abs/2109.11369> (02. december 2023).
- Zhou, Longfei, Lin Zhang, a Nicholas Konz. 2022. “Computer Vision Techniques in Manufacturing”. [https://www.techrxiv.org/articles/preprint/Computer\\_Vision\\_Techniques\\_in\\_Manufacturing/17125652/2](https://www.techrxiv.org/articles/preprint/Computer_Vision_Techniques_in_Manufacturing/17125652/2) (02. december 2023).
- Sharma, Ajay, Ankit Gupta, a Varun Jaiswal. 2021. “Solving Image Processing Critical Problems Using Machine Learning”. V *Machine Learning for Intelligent Multimedia Analytics: Techniques and Applications*, Studies in Big Data, ed. Pardeep Kumar a Amit Kumar Singh. Singapore: Springer, 213–48. [https://doi.org/10.1007/978-981-15-9492-2\\_11](https://doi.org/10.1007/978-981-15-9492-2_11) (02. december 2023).
- Li, Yan. 2022. “Application of Computer Vision in Intelligent Manufacturing under the Background of 5G Wireless Communication and Industry 4.0”. *Mathematical Problems in Engineering* 2022: e9422584.
- Díaz, Aldo André, a Paulo R. G. Kurka. 2020. “Computer Vision Methods for Automotive Applications”. *TECNIA* 30(2): 74–81.
- Wang, Xuan, a Zhigang Zhu. 2023. “Context understanding in computer vision: A survey”. *Computer Vision and Image Understanding* 229: 103646.
- Wu, Di, Nabin Sharma, a Michael Blumenstein. 2017. “Recent advances in video-based human action recognition using deep learning: A review”. V *2017 International Joint Conference on Neural Networks (IJCNN)*, , 2865–72. <https://ieeexplore.ieee.org/document/7966210> (02. december 2023).
- Zhengfeng, Huang. 2022. “Accurate Recognition Method of Continuous Sports Action Based on Deep Learning Algorithm”. *Wireless Communications and Mobile Computing* 2022: e3407935.

- “Sensors | Free Full-Text | Deep Learning-Based Real-Time Multiple-Person Action Recognition System”. <https://www.mdpi.com/1424-8220/20/17/4758> (02. december 2023).
- “Duplicate Frame Video Forgery Detection Using Siamese-based RNN”. <https://www.techscience.com/iasc/v29n3/43054/html> (02. december 2023).
- Mueller, Jonas, a Aditya Thyagarajan. 2016. “Siamese Recurrent Architectures for Learning Sentence Similarity”. *Proceedings of the AAAI Conference on Artificial Intelligence* 30(1). <https://ojs.aaai.org/index.php/AAAI/article/view/10350> (02. december 2023).
- Min-qing, Zhang, a Li Wen-ping. 2021. “An Automatic Classification Method of Sports Teaching Video Using Support Vector Machine”. *Scientific Programming* 2021: e4728584.
- Chattopadhyay, Chiranjay, a Amit Kumar Maurya. 2013. “Genre-Specific Modeling of Visual Features for Efficient Content Based Video Shot Classification and Retrieval”. *International Journal of Multimedia Information Retrieval* 2(4): 289–97.
- Gama, Fernando, Antonio G. Marques, Alejandro Ribeiro, a Geert Leus. 2018. “MIMO Graph Filters for Convolutional Neural Networks”. <http://arxiv.org/abs/1803.02247> (02. december 2023).
- Mesuga, Reymond, a Brian Bayanay. 2021. *A Deep Transfer Learning Approach to Identifying Glitch Wave-form in Gravitational Wave Data*.
- Hadidi, Ramyad, Jiashen Cao, Michael S. Ryoo, a Hyesoon Kim. 2020. “Toward Collaborative Inferencing of Deep Neural Networks on Internet-of-Things Devices”. *IEEE Internet of Things Journal* 7(6): 4950–60.
- Wang, Limin et al. 2016. “Temporal Segment Networks: Towards Good Practices for Deep Action Recognition”. V *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, ed. Bastian Leibe, Jiri Matas, Nicu Sebe, a Max Welling. Cham: Springer International Publishing, 20–36.
- Tran, Du et al. 2015. “Learning Spatiotemporal Features with 3D Convolutional Networks”. V *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, 4489–97. <http://ieeexplore.ieee.org/document/7410867/> (02. december 2023).
- Wang, Heng, a Cordelia Schmid. 2013. “Action Recognition with Improved Trajectories”. V *2013 IEEE International Conference on Computer Vision*, , 3551–58. <https://ieeexplore.ieee.org/document/6751553> (02. december 2023).

- “A Multimodal Pairwise Discrimination Network for Cross-Domain Action Recognition | IEEE Journals & Magazine | IEEE Xplore”. <https://ieeexplore.ieee.org/document/9160941> (02. december 2023).
- Li, Qingwu, Haisu Cheng, Yan Zhou, a Guanying Huo. 2016. “Human Action Recognition Using Improved Salient Dense Trajectories”. *Computational Intelligence and Neuroscience* 2016: e6750459.
- Comoli, P. et al. 2016. “Acquisition of C3d-Binding Activity by De Novo Donor-Specific HLA Antibodies Correlates With Graft Loss in Nonsensitized Pediatric Kidney Recipients”. *American Journal of Transplantation* 16(7): 2106–16.
- “MV-C3D: A Spatial Correlated Multi-View 3D Convolutional Neural Networks | IEEE Journals & Magazine | IEEE Xplore”. <https://ieeexplore.ieee.org/document/8736713> (02. december 2023).
- Zhang, Zhiwu et al. 2010. “Mixed Linear Model Approach Adapted for Genome-Wide Association Studies”. *Nature Genetics* 42(4): 355–60.
- Zhang, Qi et al. 2020. “Spatio-Temporal Graph Structure Learning for Traffic Forecasting”. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(01): 1177–85.
- Chayambuka, Kudakwashe, Jan Fransaer, a Xochitl Dominguez-Benetton. 2019. “Modeling and design of semi-solid flow batteries”. *Journal of Power Sources* 434: 226740.
- “3D genome organization in health and disease: emerging opportunities in cancer translational medicine”. <https://www.tandfonline.com/doi/full/10.1080/19491034.2015.1106676> (02. december 2023).
- Wei, Pengbo et al. 2022. “Vision-Based Activity Recognition in Children with Autism-Related Behaviors”. <http://arxiv.org/abs/2208.04206> (02. december 2023).
- Lin, Tsung-Yi et al. 2017. “Focal Loss for Dense Object Detection”. V *2017 IEEE International Conference on Computer Vision (ICCV)*, , 2999–3007. <https://ieeexplore.ieee.org/document/8237586> (02. december 2023).
- Carreira, João, a Andrew Zisserman. 2017. “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset”. V *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, , 4724–33. <https://ieeexplore.ieee.org/document/8099985> (02. december 2023).
- Shi, Zhensheng et al. 2020. “Learning Attention-Enhanced Spatiotemporal Representation for Action Recognition”. *IEEE Access* 8: 16785–94.

- Zhou, Bolei, Alex Andonian, Aude Oliva, a Antonio Torralba. 2018. “Temporal Relational Reasoning in Videos”. V *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, ed. Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, a Yair Weiss. Cham: Springer International Publishing, 831–46.
- Li, Yinqing et al. 2020. “Distinct Subnetworks of the Thalamic Reticular Nucleus”. *Nature* 583(7818): 819–24.
- Terney, Daniella et al. 2008. “Increasing Human Brain Excitability by Transcranial High-Frequency Random Noise Stimulation”. *Journal of Neuroscience* 28(52): 14147–55.
- Ahmad, Muhammad Waseem, Monjur Mourshed, a Yacine Rezgui. 2017. “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption”. *Energy and Buildings* 147: 77–89.
- Zhang, Ke, Wei-Lun Chao, Fei Sha, a Kristen Grauman. 2016. “Video Summarization with Long Short-Term Memory”. V *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, ed. Bastian Leibe, Jiri Matas, Nicu Sebe, a Max Welling. Cham: Springer International Publishing, 766–82.
- Saddam Bekhet, Abdullah M. Alghamdi. 2021. “A COMPARATIVE STUDY FOR VIDEO CLASSIFICATION TECHNIQUES USING DIRECT FEATURES MATCHING, MACHINE LEARNING, AND DEEP LEARNING”. *Journal of Southwest Jiaotong University* 56(4). <http://jsju.org/index.php/journal/article/view/994> (02. december 2023).
- Karpathy, Andrej et al. 2014. “Large-Scale Video Classification with Convolutional Neural Networks”. V *2014 IEEE Conference on Computer Vision and Pattern Recognition*, , 1725–32. <https://ieeexplore.ieee.org/document/6909619> (02. december 2023).
- Bovik, Alan C. 2009. *The Essential Guide to Video Processing*. Academic Press.
- Mou, Lichao, Yuansheng Hua, Pu Jin, a Xiao Xiang Zhu. 2020. “ERA: A Dataset and Deep Learning Benchmark for Event Recognition in Aerial Videos”. <http://arxiv.org/abs/2001.11394> (02. december 2023).
- Papert, Seymour A. 1966. “The Summer Vision Project”. <https://dspace.mit.edu/handle/1721.1/6125> (02. december 2023).
- Ng, Joe Yue-Hei et al. 2015. “Beyond short snippets: Deep networks for video classification”. V *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, , 4694–4702. <https://ieeexplore.ieee.org/document/7299101> (03. december 2023).
- Delextrat, Anne, a Daniel Cohen. 2009. “Strength, Power, Speed, and Agility of Women Basketball Players According to Playing Position”. *The Journal of Strength & Conditioning Research* 23(7): 1974.



## **ZOZNAM PRÍLOH**

Príloha A – dátový súbor + zdrojový kód + odkaz na GIT + výsledky výskumu, rozsiahle tabuľky a pod.

Príloha B – Fotodokumentácia k prípravnej fáze záverečnej práce .....

## **PRÍLOHA A**

## **Príloha A**

- A1    Odkaz na GIT: <https://github.com/>
- A2    ...
- A3    ...
- A4    ...