# ZADANIE

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.

# Úvod

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

# 1 Popis riešeného problému / Analýza súčasného stavu

## Computer Vision

## Modely 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 na základe ich vizuálnych vlastností, čiastočnej informácie alebo textového podkladu.

V najbližších kapitolách priblížime aké metódy, okrem hĺbkového učenia, využívajú video klasifikačné modely na riešenie problémov klasifikácie vo videách.

### Hĺbkové učenie v kontexte modelov pre klasifikáciu videí

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 je rozpoznávane akcií vo videách. (<https://ieeexplore.ieee.org/document/7966210>)

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

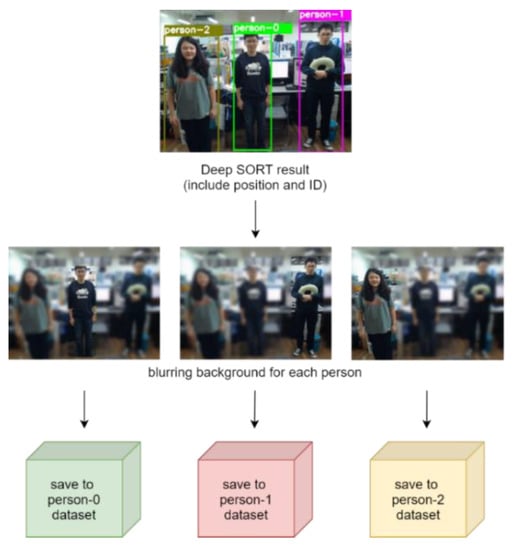
(<https://www.hindawi.com/journals/wcmc/2022/3407935/>)Obrázok, na ktorom je text, osoba, silueta

Automaticky generovaný popis

Výkony basketbalistov napríklad skúmal Peng Yao, ktorý vo svojom výskume popisuje jednoznačné využitie do budúcnosti na personifikáciu stratégie pre jednotlivých hráčov. Vďaka tomu poznatku vie 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.

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 vo videu(OBRÁZOK). Jen-Kai Tsai a jeho kolegovia riešia túto problematiku v ich článku, kde vidia využitie modelu I3D, ktorý podrobnejšie popisujeme aj v tejto práci v nasledujúcich kapitolách.

(<https://www.mdpi.com/1424-8220/20/17/4758>)



Pre RNN(Recurrent Neural Network) našli využitie napríklad v detekovaní podvodných videí. Keďže sa videá využívajú v mnohých oblastiach ako dôkazový materiál: právo, forenzné štúdie, novinárstvo a iné. Je teda dôležité overiť autenticitu daného videa. Autori popisujú využitie siamese based RNN popri modelu I3D a na dvoch datestoch dosahujú až 86.6%(dataset VIRAT - Image Retrieval and Analysis Tool) a 93%(dataset MFC - Media Forensic Challenge) presnosť. Ich objektom skúmania je vyhľadávanie duplikovaných snímok. (<https://www.techscience.com/iasc/v29n3/43054/html>)

Siamese-based RNN odkazuje na architektúru rekurentnej neurónovej siete, ktorá zahŕňa koncept siamských sietí. Pri nich ide o druh neurónových sietí, ktoré zdieľajú váhy a architektúru, čo im umožňuje spracovávať viaceré vstupy paralelne a učiť sa podobnosti alebo rozoznávať rozdiely medzi nimi. (<https://ojs.aaai.org/index.php/AAAI/article/view/10350>)

### SVM a klasifikácia vo videách

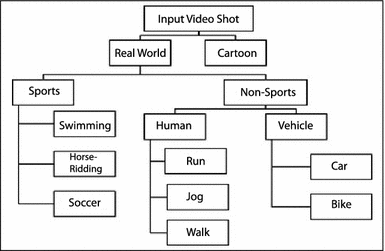
SVM alebo Support Vector Machine je algoritmus strojového učenia, ktorý sa často využíva pre klasifikačné a regresné úlohy. Pri videách sa momentálne skúma využitie SVM na kategorizáciu druhov videí a to konkrétne: reklamy, kreslené rozprávky, hudobné videá, správy a športové videá, podľa výskumu Zhang Min-qinga a Li Wen-pinga. Detekciu vykonáva aj za pomoci zvukovej zložky priloženej k videu. Autori sa dodatočne snažia odhaliť aj či nejde o teroristický alebo násilný obsah.

Pri SVM ide o štatistický prístup k nájdeniu najlepšej plochy v pôvodnom priestore alebo vo vyššej dimenzii po projekcii. Cieľom je maximalizovať medzeru medzi dvoma kategóriami, čím znižujeme interval istoty pri generalizácii a tak minimalizujeme skutočné riziká (OBRÁZOK). (<https://www.hindawi.com/journals/sp/2021/4728584/>)

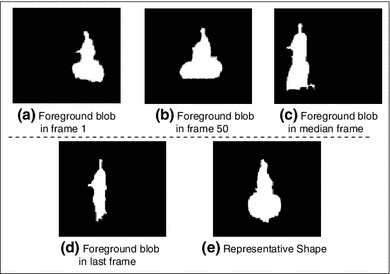
Obrázok, na ktorom je diagram, rad, náčrt, kresba

Automaticky generovaný popis

Podobnej problematike sa venovali aj Chattopadhyay a Maurya, ktorí pomocou SVM preskúmavali možnosť rýchleho označovania videí do kategórií pomocou kľúčových slov (tagov). (OBRÁZOK) (<https://link.springer.com/article/10.1007/s13735-013-0034-8>)



Pri týchto výskumoch sa videá spracovávajú a to 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.(OBRÁZOK) Tento útvar sa oddeľuje od pozadia a neskôr sa, na základe ich relatívneho centra, prekrývajú a pozorujú zmeny. Veľkou pomocou sa ukázalo porovnávať a zaznamenávať aj textúry objektov. (<https://link.springer.com/article/10.1007/s13735-013-0034-8>)



Autori po tomto spracovaní preberajú rozdiely medzi kategóriami a na čo si treba dávať pozor, pričom vo výsledku poukazujú na využitie v rozpoznávaní kontextu videa avšak do budúcna odporúčajú pridať ďalšie média ako zvuk a text. (<https://link.springer.com/article/10.1007/s13735-013-0034-8>)

### 3D Konvolunčné modely

CNN berú vstupné dáta a spracúvajú 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. (<https://arxiv.org/pdf/1803.02247.pdf>)

Z obrázku vidíme, že sa CNN skladá z dvoch hlavných časti 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 ćíslo hovorí o sile prvku. CNN majú dôležitú vlastnosť a to zdielanie váh, čo znamená, že sa každý filter požíva na každý vstupný prvok. Pri časti pooling ide o hlasovanie na základe väčšinového hlasovania, tu sa vyfiltrujú jednotlivé prvky, ktoré idú do neurónových sietí. (<https://www.researchgate.net/publication/353049916_A_Deep_Transfer_Learning_Approach_to_Identifying_Glitch_Wave-form_in_Gravitational_Wave_Data>)

Obrázok, na ktorom je diagram, rad, text, plán

Automaticky generovaný popis

### Modely využité pri analýze ERA datasetu

V nasledujúcich kapitolách popíšeme momentálne využitie modelov, ktoré vybrali pre analýzu autori Mou, L., Hua, Y., Jin, P. a Zhu, X. X. s použitím datasetu ERA(Event Recognition in Aerial Videos - Rozpoznávanie udalostí v leteckých videách)[[1]](#footnote-1).

#### C3D Model

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.

C3D(Convolutional 3D model - Konvolučný 3D model) 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. (<https://ieeexplore.ieee.org/document/8985265>)

Vďaka stratégii, kedy nespracováva každý snímok, ale presne dané číslo snímok v intervale, sa zvyšuje jeho efektivita a znižuje výpočtová náročnosť. (<https://link.springer.com/chapter/10.1007/978-3-319-46484-8_2>)

OBRAZOK

Obrázok, na ktorom je text, snímka obrazovky

Automaticky generovaný popis

(https://arxiv.org/pdf/1906.02182.pdf)

Pri využití sa najviac spomína 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 (<https://ieeexplore.ieee.org/document/6751553/>). Shang taktiež popisuje zlepšenie výkonov s použitím modelu ActionVLAD (<https://ieeexplore.ieee.org/document/9160941>).

HOF a MBH využívajú 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. (<https://www.hindawi.com/journals/cin/2016/6750459/>) MBH naopak zachytáva hranice pohybu vo videu. Počíta gradient optického toku a kvantifikuje orientácie gradient do rôznych binov. V tomto prípade histogram reprezentuje distribúciu hraníc pohybu vo videu. (<https://www.hindawi.com/journals/cin/2016/6750459/>) OBRÁZOK

Obrázok, na ktorom je text, diagram, snímka obrazovky, dizajn

Automaticky generovaný popis

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 na rôznych testoch výkonu. (<https://ieeexplore.ieee.org/document/7410867>)

V oblasti medicínskeho výskumu sa C3D využíva pri pacientoch, ktorým bola transplantovaná oblička. Podľa výskumu (<https://linkinghub.elsevier.com/retrieve/pii/S1600613522009261>) 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. (<https://ieeexplore.ieee.org/document/8736713>)

Nájdeme využitie aj v imunológii, kde C3D zlepšuje vytváranie protilátok na vírusy prvého typu. Čiže sad á potencionálne vyžiť tento model na vytváranie stratégií pri tvorbe vakcín a imunoterapie pri vírusových infekciách. (<https://ieeexplore.ieee.org/document/8736713>)

#### 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. (<https://www.nature.com/articles/ng.546>)

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

Jedna zo štúdií od Chayambuka a iných 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. Model umožnil znázorniť časovo závislé profily napätia, distribúciu prúdu a distribúciu stavu napätia v aktívnych častiach. (<https://www.sciencedirect.com/science/article/abs/pii/S0378775319307116?via%3Dihub)>

OBRÁZOK

Obrázok, na ktorom je text, snímka obrazovky, diagram, rad

Automaticky generovaný popis

Využitie modelu nájdeme aj v biológii, pri génových štúdiách, kde sa využíva napríklad na pochopenie organizácie a funkcionality genómov pri zdraví a chorobách populácie. (<https://www.tandfonline.com/doi/full/10.1080/19491034.2015.1106676>)

V medicine má P3D model úžitok pri modelovaní geometrie genómov, kde sa pozorujú bunky a ich dopad na zdravie ľudí. (<https://www.tandfonline.com/doi/full/10.1080/19491034.2015.1106676>)

#### I3D Model

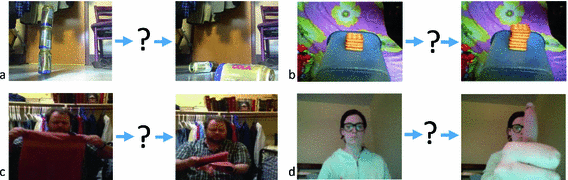
I3D model predstavuje architektúru, ktorá zväčšuje dvojrozmerné konvolunčné siete aby spracovával informácie času a priestoru z videí. V podstate využíva 2D siete, ktorým pridáva filtre a presúva ich do 3D. Čiže napríklad štvorcový filter sa stáva kubický. (<https://arxiv.org/pdf/2208.04206.pdf>)

V oblasti rozpoznávania aktivít vo videách sa tento model využíva v chirurgii, kde sa rozpoznávajú postupy pri práci. V štúdii (<https://link.springer.com/article/10.1007/s11548-021-02473-3>) bol model I3D, konkrétne architektúra Inflated 3D ConvNet (na základe ktorej je vybudovaný model I3D) (<https://ieeexplore.ieee.org/document/8099985>), využitá na rozpoznávanie chirurgických postupov v rámci operácií Tubulizácie žalúdka. Model bol trénovaný pomocou metódy 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 príkladoch tried. Týmto zlepšuje výkony s triedach menšín. (https://ieeexplore.ieee.org/document/8237586)

#### TRN Model

TRN model (Temporal Relation Network – dočasne relačný model) vznikol na princípe relačných úsudkov 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 vedel efektívne učiť sa a interpretovať vzťahy medzi snímkami vo videu a to vo viacerých časových škálach. Na obrázku vidíme ako si ľudský mozog ľahko doplní kontext medzi obrázkami, ale pre modely neurónových sietí je to náročné. Tento problém sa snaží vyriešiť model TRN za pomoci zisťovania vzťahov medzi snímkami.

(https://link.springer.com/chapter/10.1007/978-3-030-01246-5\_49) 

TRN model bol pôvodne navrhnutý ako framework pre online detekciu akcií a ich predpovedanie do budúcnosti vo videách. (<https://ieeexplore.ieee.org/document/9009797/>) Jeho cieľom je modelovať väčší časový kontext tým, že vykonáva tieto úlohy súčasne tak za pomoci učenia sa logiky medzi simkami videa v rôznych časových škálach. To ho robí účinným a často výužívaným modelom siete na porozumenie kontextu videa a jeho analýzu. (<https://link.springer.com/chapter/10.1007/978-3-030-01246-5_49>)

Vo výskumoch neurovedy bol tento model skúmaný pri regulácii senzorického spracovania, pozornosti a kognície. (<https://www.nature.com/articles/s41586-020-2504-5>) Taktiež na základe dostupných referencií sa využíva, podobne ako P3D model, v génovej vede, kde pomáha skúmať komplexné interakcie medzi samotnými génmi a ich regulačnými prvkami. Týmto nám odkrývajú pohľad na mechanizmy, ktoré súvisia s expesiou génov a bunkovými procesmi. (<https://www.jneurosci.org/content/28/52/14147>)

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. (<https://www.sciencedirect.com/science/article/pii/S0378778816313937?via%3Dihub>)

### Ďalšie modely

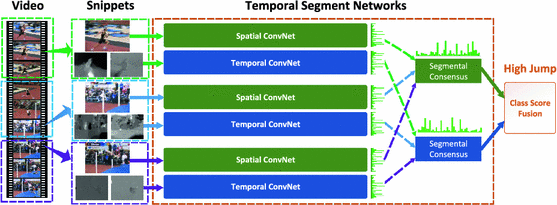
V tejto kapitole popíšeme modely, ktoré sú momentálne využívané pre klasifikáciu vo videách a my sme ich využili v našej práci.

#### TSN model

TSN model bol vyvinutý na základe potreby rozoznávať dlhodobé aktivity a taktiež potreby trénovať modely na veľkom množstve videí. Tu autori videli veľký potenciál na zlepšenie a vyvinuli model TSN alebo Temporal Segment Network (Sieť využívajúca časových segmentov). Tento framework extrahuje krátke úryvky z dlhej sekvencie pôvodného videa pomocou vzorkovacej stratégie s nie hustým vzorkovaním. Vďaka tomuto je tento model schopný pracovať s dlhými videami a zachovať si relevantné informácie za menšej výpočtovej záťaže.

Na obrázku je popísaná segmentácia videa, náhodné vybranie úryvku a na nich sa vykonáva potom modelovanie. Na to sa tu využíva Spatial ConvNet a Teporal čiže na základe priestoru a času. (<https://link.springer.com/chapter/10.1007/978-3-319-46484-8_2>)

Využite sa momentálne najviac priraďuje k zlepšeniu výkonov pri rozpoznávaní aktivít vo videách. (<https://link.springer.com/chapter/10.1007/978-3-319-46484-8_2>)



#### LSTM

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

(<https://www.isca-speech.org/archive/pdfs/interspeech_2016/chen16n_interspeech.pdf>)

Obrázok, na ktorom je text, snímka obrazovky, diagram, číslo

Automaticky generovaný popis

LSTM má využitie v rozpoznávaní obsahu videa a jeho popísaní. Zaznamenáva najviac dôležité momenty a na ich základe vyhodnotia obsah. Týmto sú užitočné najmä pri dlhých videách. (<https://link.springer.com/chapter/10.1007/978-3-319-46478-7_47>)

/C3D Model

C3D-Sport1M, C3D-UCF101

/P3D Model

P3D-Kinetics, P3D-Kinetics-600

/I3D Model

I3D-Kinetics, I3D-Kinetics+ImageNet

/TRN Model

TRN-Something-Something-V2, TRN-Moments-In-Time

INÉ: Non-local Neural Networks

Temporal Segment Networks TSN

the Quo Vadis model

RNNs (Recurrent Neural Networks)

3D CNNs

T3D (Temporal 3D ConvNets)

# 2 Ciele záverečnej práce

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

Naše čiastkové ciele sú:

* Overiť, že modely, ktoré boli použité už na datasete ERA majú v našej práci podobné výsledky.
* Ak by sa výsledky priveľmi odlišovali, tak chceme prísť na možné dôvody.
* 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 riadli počas nášho výskumu. Pridržiavali sme sa schémy 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.

### Klasifikácia vo videách

Pri klasifikácií 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. (<http://jsju.org/index.php/journal/article/view/994>) Oproti klasifikácií 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. (<https://ieeexplore.ieee.org/document/6909619>) Na obrázku môžeme vidieť, že sa videá musia brať ako viac dimenzionálne média, pretože v nich ide o kontext v čase. (<https://books.google.sk/books?hl=sk&lr=&id=wXmSPPB_c_0C&oi=fnd&pg=PP1&dq=video+processing&ots=UTTsyHktdL&sig=4d68oqpxfK3aolkMjUVwRhvJe0E&redir_esc=y#v=onepage&q=video%20processing&f=false>)

Obrázok, na ktorom je diagram, text, rad, snímka obrazovky

Automaticky generovaný popis

## 3.2. Výber datasetu

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.

## 3.3. Porozumenie dátam v datasete

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

## 3.4. Spracovanie dát

## 3.5. Výber modelov

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

## 3.6. Modelovanie

## 3.7. Evalvácia výsledkov

## 3.8. Vyhodnotenie a interpetácia

# Zoznam bibliografických odkazov

1. <https://lcmou.github.io/ERA_Dataset/> [↑](#footnote-ref-1)