

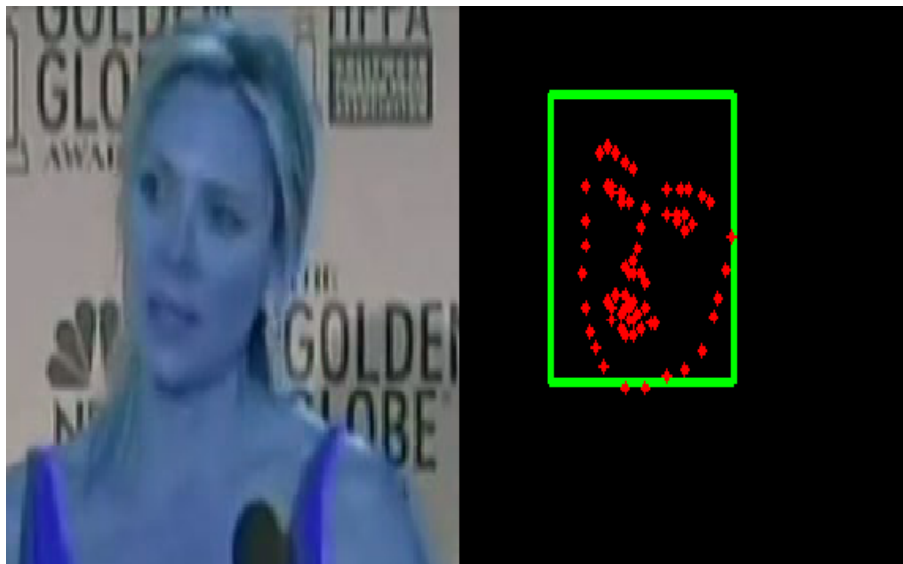
# BIOM - zadanie č.2

Tomáš Jenčík, 111291

Apríl 2024

## 1 Práca s videami

Na začiatok sme si stiahli videá, uložili do súboru videos a spracovali ich tak, aby sme ich vedeli prehrať rovno v kóde. Zároveň sme do nich vkreslili aj anotácie z dodaných súborov, vďaka čomu sa vedľa videa zobrazuje aj pozícia tváre aj tvárových bodov. Jedna snímka z videa mohla vyzeráť nasledovne:



Obr. 1: Video s anotáciami.

Ako ďalšie sme sa rozhodli videá upraviť nasledovným spôsobom. Na každej snímke každého videa sme aplikovali metódy predspracovania. Po rôznych kombináciach predspracovaní snímok sme skončili na:

- normalizácia: proces, pri ktorom škálujeme hodnoty pixelov pre každú snímku vo videu na rozsah medzi 0 a 1. Normalizácia pomáha stabilizovať algoritmy učenia a zlepšovať konvergenciu počas tréningu.

- augmentácia: používa sa na umelé zvýšenie rozmanitosti tréningových údajov aplikáciou rôznych transformácií na existujúce vzorky údajov. V našom prípade sme použili 2 techniky a to konkrétne zmenu jasú snímky, otočenie obrazu a zmena veľkosti snímky:
  - jas: zmena jasú môže pomôcť zlepšiť robustnosť a schopnosť zovšeobecnenia modelov strojového učenia tým, že ich počas tréningu vystavíme širšiemu spektru svetelných podmienok. Hodnota jasú bola 1,5.
  - otočenie obrazu: cieľom otočenia je vystaviť model zmenám v orientácii objektov, čím sa zlepši jeho schopnosť zovšeobecňovať na rôzne uhly pohľadu a orientácie objektov.
  - zmena veľkosti snímky: zmena veľkosti môže simulovať objekty objavujúce sa v rôznych vzdialenostiach od fotoaparátu.

Obraz vyzerá nasledovne:



Obr. 2: Originálny a zle predspracovaný obraz.

Pri neskoršom výpočte precision, recall a f1-skóre sme si zistili, že tieto predspracovania nie sú vhodné z toho dôvodu, že hoci naše modely správne nájdu hľadané tváre vo väčšine prípadov, tieto obdĺžniky sa nebudú zhodovať s pravými obdĺžníkmi, keďže obrazy nielenže sú inej veľkosti a teda aj obdĺžniky budú musieť byť, ale keďže sme otočili snímky, tak obdĺžniky sa budú nachádzať na iných miestach (sú nájdené zrkadlovo) a preto sme museli spraviť nové predspracovanie pre obrazy, ktorá používala:

- normalizácia (ako v predošlom predspracovaní)
- augmentácia: Túto augmentáciu sme zmenili a po úprave vyzerá nasledovne:
  - jas: zmenili sme hodnotu z 1.5 na 0.8.

- ostrenie: ostrenie je technika digitálneho spracovania obrazu, ktorá sa používa na zlepšenie čistoty a detailov obrazu. Hodnota pre ostrenie je matica 3x3, ktorá v strede má hodnotu 7, v rohoch 0 a zvyšné body -1.
- rozmazanie: používa sa najmä v scenároch, kde je dôležitý šum, nadmerné prispôsobenie alebo ochrana súkromia.

Rozdiel originálneho a predspracovaného obrazu vyzeral nasledovne:



Obr. 3: Originálny a predspracovaný obraz.

## 1.1 Haarove kaskády

Haarove kaskády je technika detekcie objektov založená na strojovom učení, ktorá sa používa na identifikáciu objektov na obrázkoch alebo snímkach videí a je obzvlášť populárna na detekciu tvárí.

Algoritmus funguje pomocou série klasifikátorov usporiadaných do kaskády. Každý klasifikátor je trénovaný na rozlíšenie medzi špecifickým aspektom detegovaného objektu a pozadím. Tieto klasifikátory sú zvyčajne založené na znakoch podobných Haarovi, čo sú jednoduché obdĺžnikové prvky, ktoré možno vypočítať z intenzity pixelov v obrázku. Počas procesu detekcie sa kaskáda klasifikátorov aplikuje na rôzne podoblasti obrazu vo viacerých mierkach a polohách. Ak v každej fáze kaskády určitá oblasť obrázka neprejde klasifikátorom, okamžite sa zahodí, čo pomáha efektívne eliminovať oblasti pozadia, ktoré pravdepodobne neobsahujú predmet záujmu. Treba taktiež poznamenať, že Haarove kaskády vedia nájsť len štvorce pre tváre, nie obdĺžniky.+

Haarove kaskády boli pôvodne trénované na frontálnom súbore údajov CALTECH v roku 1999 (J. Howse openCV 4 „pre tajného agenta“, druhé vydanie, Packt, kapitola 3, strany 101-102).

Hyperparametre, ktoré Haarove kaskády v našom kóde používajú sú:

- faktor škálovania: kompenzuje tváre, ktoré sa zdajú byť menšie, ak sú ďalej od fotoaparátu. Tento faktor to kompenzuje zmenou mierky rozpoznaných tvárí v rôznych vzdialenostiach.
- minimálny počet susedov: špecifikuje, koľko susedov by mal mať každý kandidátsky obdĺžnik, aby sa zachoval, pričom vyššie hodnoty majú za následok menej detekcií, ale vyššiu kvalitu.

- minimálna veľkosť: nastavuje minimálnu veľkosť okna. Objekty menšie ako určené číslo sú ignorované.

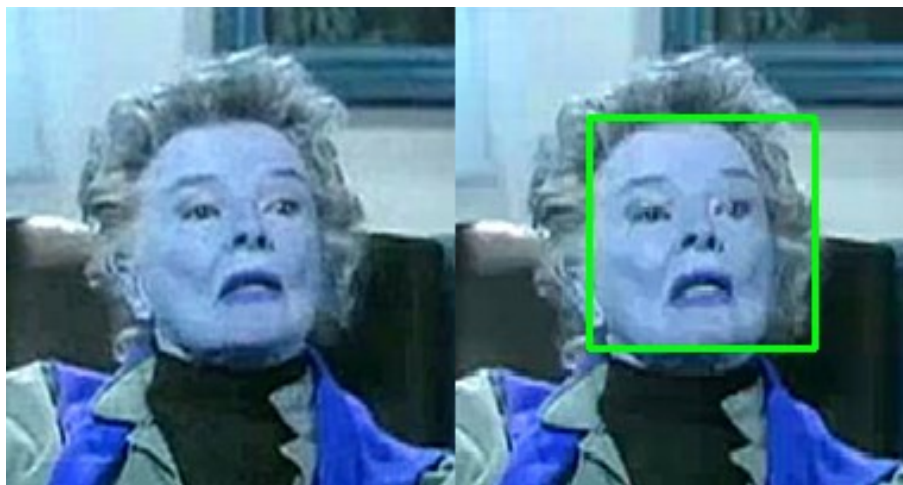
### 1.1.1 Výsledky na videách bez spracovania

Ako prvé si ukážeme výsledky pre obrázky, ktoré sme nijak neupravili, respektíve sa bude jednať o snímky neupravených videí. Vypočítali sme si hodnotu Precision, Recall a F1-skóre pre jednotlivé videá, ako aj celkovo pre databázu. Pre každé video sme zachytávali každú piatu snímku, na ktorej sme potom hľadali tváre. Príkladom výsledkov pre jedno video sú napríklad výsledky pre video Katharine\_Hepburn\_3:

Tabuľka 1: Hodnoty videa Katharine\_Hepburn\_3.

Metrika	Hodnota
True Positives (TP)	19
False Positives (FP)	0
False Negatives (FN)	3
Precision	1.0
Recall	0.8636
F1 Score	0.9268

Snímky z tohto videa vyzerali nasledovne:

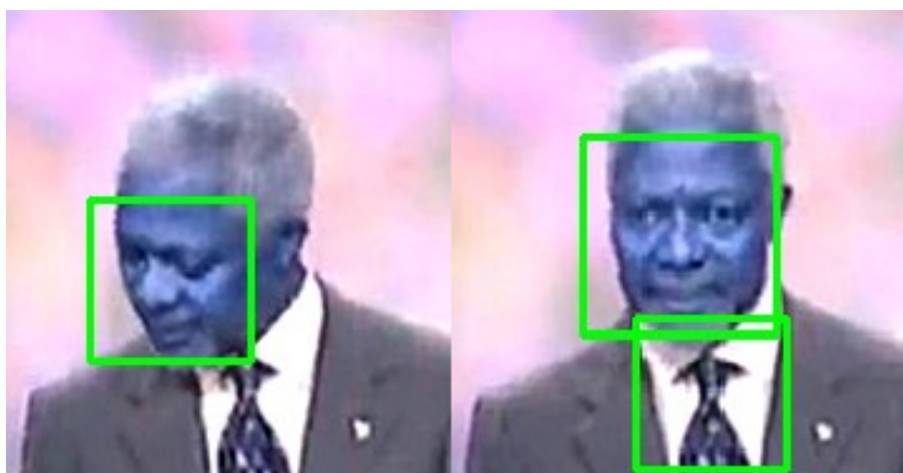


Obr. 4: Haarova detekcia obrazu pre video Katharine\_Hepburn\_3.

Ako môžeme vidieť, zatiaľčo na prvom obraze Haarove kaskády žiadnu tvár nenašli, na ďalšom obraze sa im to už podarilo. Pri Haarových kaskádach najhoršie výsledky pre hľadanie tvárí boli tam, kde osoba nosila kravatu s košelou, podľa čoho si model pravdepodobne myslel, že sa tam nachádza tvár. Ukážka takéhoto príkladu je napríklad na videu Kofi\_Annan\_3:

Tabuľka 2: Hodnoty videa Kofi\_Annan\_3.

Metrika	Hodnota
True Positives (TP)	12
False Positives (FP)	23
False Negatives (FN)	0
Precision	0.3429
Recall	1.0
F1 Score	0.5106



Obr. 5: Haarova detekcia obrazu pre video Kofi\_Annan\_3.

Ako môžeme vidieť, model nesprávne našiel ďalšiu tvár tam, kde sa nachádza kravata, pričom táto chyba nastala aj v iných videách.

Výsledky pre celkovú databázu vyzerali nasledovne:

Tabuľka 3: Celkové výsledky pre Haarove kaskády.

Metrika	Hodnota
True Positives (TP)	12883
False Positives (FP)	187
False Negatives (FN)	152
Precision	0.9856
Recall	0.98094
F1 Score	0.9805

Môžeme vidieť, že aj pre model, ktorý by mal byť horší z 2 používaných dostávame veľmi dobré výsledky. Dôvodom bude pravdepodobne to, že máme nastavené veľmi malé hodnoty IoU čo

zapríčinilo, že veľa nájdených štvorcov sa zhodovalo s pravými obdĺžnikmi o viac ako 50% (hodnota IoU bola  $>0.5$ ).

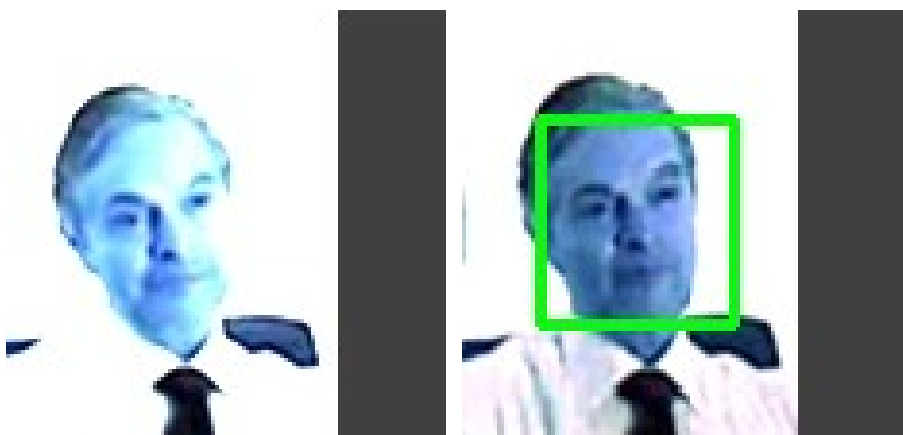
### 1.1.2 Výsledky na videách s predspracovaním

Skúšali sme videá aj predspracovať, ako to bolo spomínané vyššie, ale s úpravou snímok sme sa nikdy nedostali k lepším výsledkom, aké sme dostali pri originálnych snímkach. Čím menšie boli úpravy, tým lepšie výsledky sme dostávali, čím dokážeme vyvodiť, že predspracovania nám nepomohli. Pre ukážku predspracovaní si ukážeme viac predspracovaní s väčšími hodnotami, aby sme videli rozdiely, hoci vieme, že tieto výsledky budú horšie.

Ako prvé si znova ukážeme video, kde nebola nájdená žiadna tvár. Jedným z týchto videí je Luis\_Ernesto\_Derbez\_Bautista\_5. Výsledky pre toto video boli:

Tabuľka 4: Výsledky pre upravené snímky videa Luis\_Ernesto\_Derbez\_Bautista\_5.

Metrika	Hodnota
True Positives (TP)	8
False Positives (FP)	0
False Negatives (FN)	3
Precision	1.0
Recall	0.7272
F1 Score	0.8421



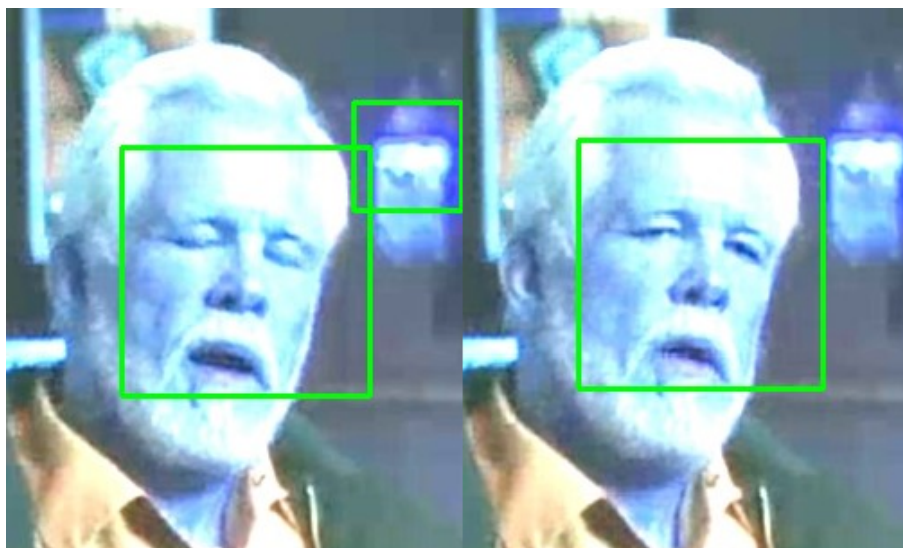
Obr. 6: Snímky videa Luis\_Ernesto\_Derbez\_Bautista\_5.

Ako vidíme, snímky obrazov, ktoré sú menej jasné a teda sa viac podobajú na originálne obrazy sú vhodnejšie pre hľadanie tvári. Taktiež si môžeme všimnúť, že hoci tvár bola nájdená, nájdený bol len štvorec pre tvár a nie obdĺžnik, keďže Haarove kaskády obdĺžnik nevedia nájsť, ako sme spomínali na začiatku.

Pozrieme sa ešte na príklad videa, kde sa našlo viacero tvárí než malo. Jedným z dôvodov bolo buď to, že sa na snímke nachádzalo viac tvárí alebo sa nachádzal objekt, ktorý po zmene jasu sa podobal na tvár. Jedným z takýchto videí bolo video Nick\_Nolte\_5. Výsledky pre video boli:

Tabuľka 5: Výsledky pre upravené snímky videa Nick\_Nolte\_5.

Metrika	Hodnota
True Positives (TP)	5
False Positives (FP)	7
False Negatives (FN)	3
Precision	0.4167
Recall	1.0
F1 Score	0.5882



Obr. 7: Snímky videa Nick\_Nolte\_5.

Ako môžeme vidieť, na snímke sa nachádza vec, ktorá farbami a tvarom môže pripomínať podobu hlavy a preto si Haarove kaskády môžu myslieť, že sa na pozadí nachádza hlava. Tento problém by nemal nastať (alebo aspoň v menšom množstve) u detektora MTCNN, keďže vieme nastaviť, ako ďaleko maximálne môže byť tvár na videu.

Celkové výsledky pre upravený dataset boli:

Ako vidíme, hoci výsledky sú horšie ako pre originálny dataset snímok, tieto výsledky sú stále veľmi vysoké, čo znova nastáva kvôli nízkej hodnote IoU ( $>0.5$ ).

Tabuľka 6: Celkové výsledky pre upravené snímky.

Metrika	Hodnota
True Positives (TP)	12135
False Positives (FP)	283
False Negatives (FN)	804
Precision	0.96
Recall	0.9229
F1 Score	0.0.9301

## 1.2 MTCNN detektor

MTCNN (Multi-task Cascaded Convolutional Networks) je algoritmus detekcie tváre založený na hlbokom učení. Vyvinuli ho Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li a Yu Qiao v roku 2016. MTCNN je známy svojou účinnosťou pri detekcii tvárí na obrázkoch v rôznych podmienkach, ako sú rôzne polohy, svetelné podmienky a mierky. MTCNN pozostáva z 3 fáz:

- Sieť návrhov (P-Net): generuje ohraničujúce rámčeky kandidátov pre potenciálne tváre v obraze a upravuje ich.
- Spresňovacia sieť (R-Net): vylepšuje kandidátske ohraničujúce rámčeky, klasifikuje ich, či obsahujú alebo neobsahujú tvár, a zlepšuje presnosť lokalizácie.
- Výstupná sieť (O-Net): poskytuje konečné spresnenie ohraničujúcich rámčekov, predpovedá orientačné body tváre a klasifikuje spresnené rámčeky, či obsahujú tvár alebo nie.

MTCNN využíva kaskádovú architektúru, kde každá fáza postupne spresňuje výsledky predchádzajúcej fázy, čo vedie k vysokej presnosti detekcie tváre pri zachovaní výkonu v reálnom čase.

MTCNN bol trénovaný na súbore údajov WIDER FACE, ktorý obsahuje 32 203 verejných obrázkov a 393 703 označených tvárí. Trénovaný model dosahuje výkon v reálnom čase pre obrázky  $640 \times 480$  VGA s minimálnou veľkosťou tváre  $20 \times 20$ . Autori použili NVIDIA Titan Black, na ktorej boli schopní dosiahnuť 99 fps.

## 1.3 Výsledky na videách bez spracovania

Znova sa pozrieme na naše 2 typy datasetov, konkrétne originálne videá bez zmeny a augmentované snímky. Znova treba poznamenať, že spracovanie snímok nám nepomohlo a preto si ukážeme príklad spracovania na snímkach, kde je viditeľné, že boli upravené.

Pri originálnych snímkach si ukážeme len zle nájdené tváre (False Positive), keďže modelu sa z celkových 13 222 snímok všade podarilo nájsť nejakú tvár. MTCNN model našiel viacero obdĺžnikov na videách, kde sa nachádzalo viacero tvárí.

Výsledky pre 2 príklady videí boli:



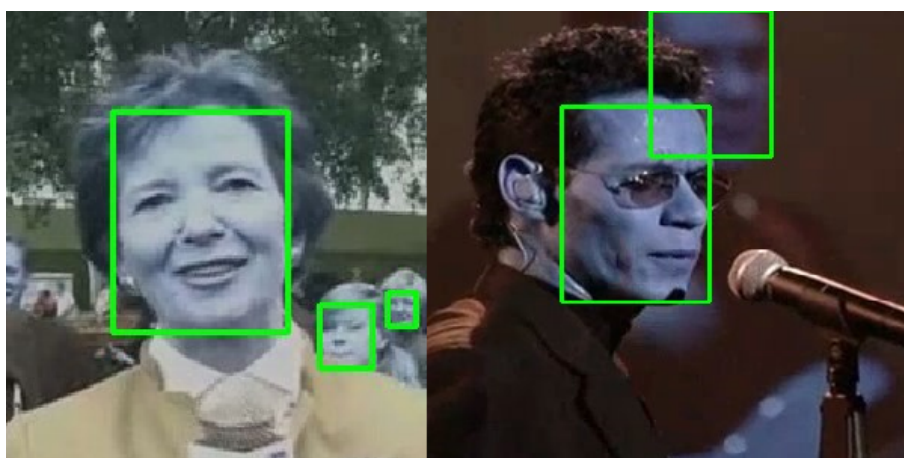
Tabuľka 7: Výsledky snímok pre videá.

Tabuľka 8: Video Mary\_Robinson\_2.

Metrika	Hodnota
True Positives (TP)	13
False Positives (FP)	3
False Negatives (FN)	0
Precision	0.8125
Recall	1.0
F1 Score	0.8966

Tabuľka 9: Video Marc\_Anthony\_5.

Metrika	Hodnota
True Positives (TP)	29
False Positives (FP)	2
False Negatives (FN)	0
Precision	0.9355
Recall	1.0
F1 Score	0.9667



Obr. 8: Snímky s viacerými tvármi.

Ako môžeme vidieť, MTCNN model dokáže nájsť viacero tvárí, dokonca aj v prípade, že nie sú celé na obraze alebo ich nie je zreteľne vidno.

Celkové výsledky pre MTCNN model boli:

Tabuľka 10: Celkové výsledky pre originálne snímky.

Metrika	Hodnota
True Positives (TP)	13149
False Positives (FP)	73
False Negatives (FN)	0
Precision	0.9920
Recall	1.0
F1 Score	0.9949

Vidíme, že sme dostali veľmi dobré výsledky a dokonca si môžeme všimnúť, že ani raz nenastal

prípade, že by model nenašiel nejakú tvár a iba 73krát našiel obraz s viacerými tvármi.

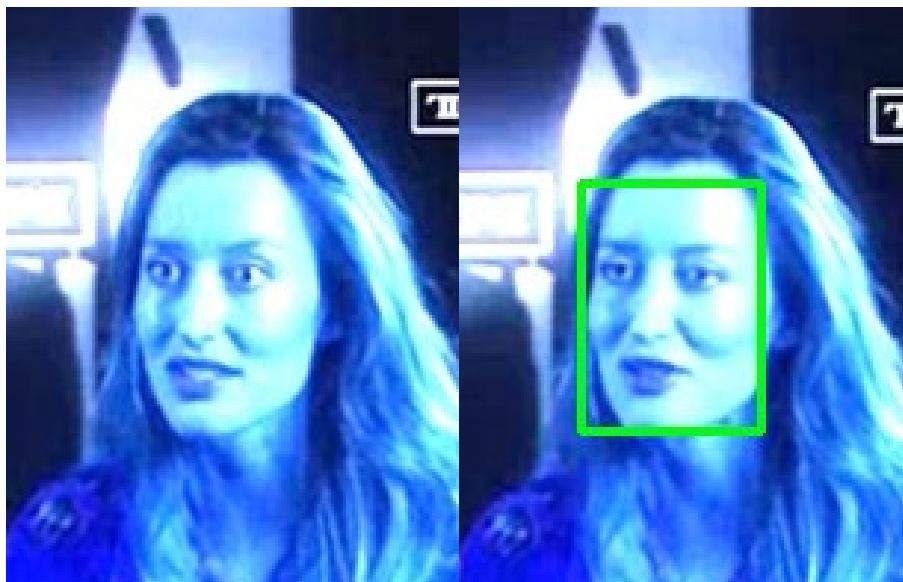
### 1.3.1 Výsledky na videách s predspracovaním

Ako bolo spomenuté, spracovanie snímok nám nepomohlo k lepším výsledkom, keďže výsledky už tak boli príliš dobré. Na upravených snímkach už nastalo, že model nenašiel žiadne tváre a taktiež nastalo, že tváre našiel tam, kde nemal

Ako prvý príklad upraveného videa je Natasha\_McElhone\_1. Výsledky pre toto video boli:

Tabuľka 11: Výsledky pre upravené video Natasha\_McElhone\_1.

Metrika	Hodnota
True Positives (TP)	24
False Positives (FP)	0
False Negatives (FN)	24
Precision	1.0
Recall	0.5
F1 Score	0.6667



Obr. 9: Upravené snímky videa Natasha\_McElhone\_1.

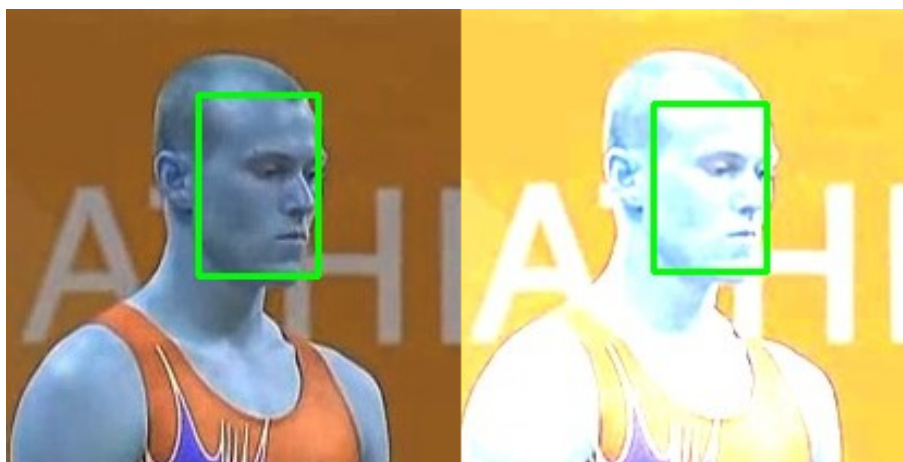
Môžeme vidieť, že video nemá žiadne objekty alebo ďalšie osoby, ktoré by pridali ďalšie nájdené tváre alebo práve zabránili nájdeniu hlavnej tváre, avšak zmena jasú očividne uškodila našim videám, keďže tvár sa nachádza ťažšie.

Ako posledné si treba ešte ukázať prípad, kedy síce hľadané tváre síce neboli nájdené presne, ale boli obdĺžniky jemne posunuté. Výsledky pre ukážku videa, kde obdĺžniky neboli nájdené presne

boli:

Tabuľka 12: Výsledky pre upravené video Natasha\_McElhone\_1.

Metrika	Hodnota
True Positives (TP)	0
False Positives (FP)	8
False Negatives (FN)	4
Precision	0.0
Recall	0.0
F1 Score	0.0



Obr. 10: Originálny a upravený snímok videa Kyle\_Shewfelt\_0.

Hoci si môžeme všimnúť, že model našiel tvár na originálnej aj upravenej snímke, nájdený obdĺžnik na upravenej snímke nebol dostatočne správny na to, aby sa považoval za TP a teda bol zaradený medzi FP. Aj z číselných výsledkov môžeme vidieť, že model nesprávne určil nájdené tváre alebo ich dokonca vôbec nenašiel.

Celkové výsledky pre model na upravených snímkach bol:

Ako vidíme, výsledky boli oveľa horšie ako pri originálnom datasete a teda môžeme povedať, že sa nám spracovanie videí neoplatilo urobiť.

## 2 Detekcia bodov na tvári

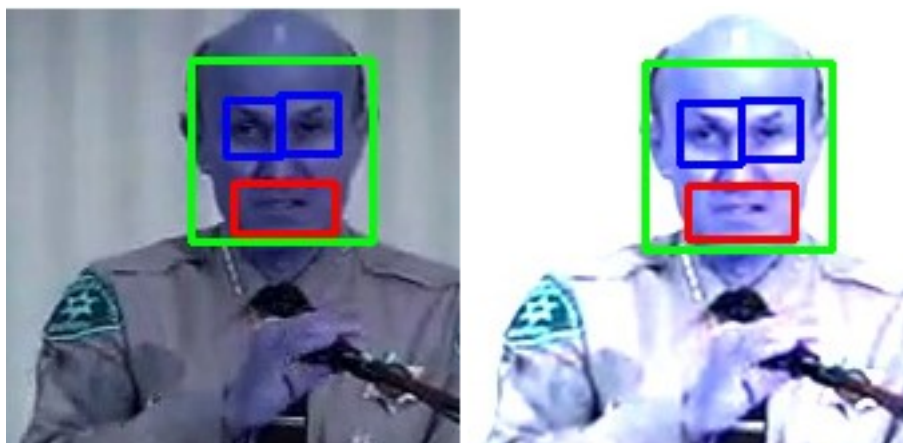
V ďalšej sekcii sa budeme zaoberať detekciou bodov na tvári. Znova použijeme naše 2 modely na 2 datasetoch, s ktorými sme pracovali doteraz, avšak v tomto prípade budeme pomocou nich hľadať oči a ústa. Finálne výsledky budú potom uvedené pomocou metriky MSE.

Tabuľka 13: Celkové výsledky pre upravený model.

Metrika	Hodnota
True Positives (TP)	10536
False Positives (FP)	126
False Negatives (FN)	2560
Precision	0.9196
Recall	0.7990
F1 Score	0.8283

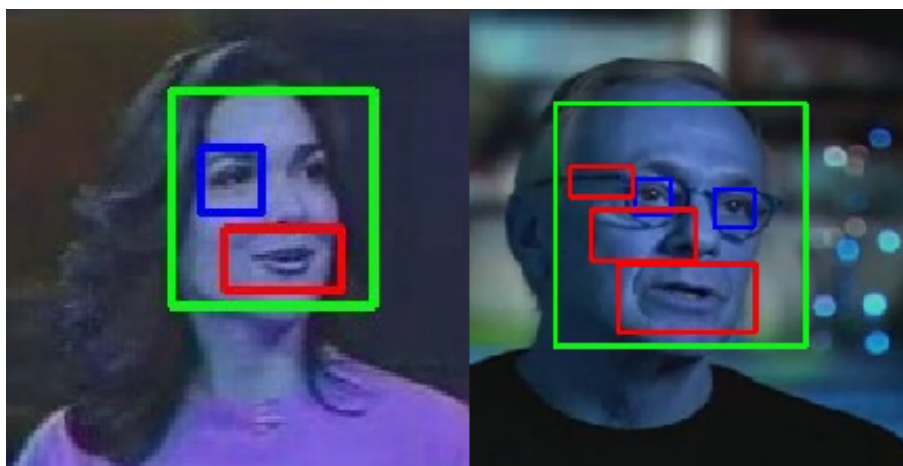
## 2.1 Haarove kaskády pre rozpoznávanie bodov na tvári

Narozdiel od toho, ako to bolo pri tvári, kde bol vždy nájdený iba štvorec, v tomto prípade sú napríklad body nájdené aj pomocou obdlžníkov, ako to uvidíme u úst.. Tieto body sme skúšali nájsť na 2 datasetoch, originálnom a upravenom a tak ako v predošlých prípadoch, znova to skončilo tak, že lepšie celkové výsledky boli na originálnom datasete, hoci pri bližšom prehladávaní datasetov sa to na prvýkrát nezdalo.



Obr. 11: Oči a ústa na originálnej a upravenej snímke.

Ako môžeme vidieť, na oboch prípadoch dokázal model nájsť obe oči aj ústa. Nie vždy ale to bolo ideálne ako v tomto prípade.



Obr. 12: Chybné nájdenia bodov na tvári.

Môžeme vidieť, že Haarove kaskády nie vždy nájdu obe oči, alebo ako vidíme na druhej snímke, nájdu niektoré body viackrát ako by mali. Výsledky sme pozerali na metrike MSE. Výsledné hodnoty pre originálny a upravený dataset boli:

Metrika	Hodnota
Počet snímok	13222
Priemerná MSE	322.0213
Minimálna nájdená MSE	0.0003
Maximálna nájdená MSE	1756.2163

Obr. 13: Originálne snímky

Metrika	Hodnota
Počet snímok	13222
Priemerná MSE	593.2890
Minimálna nájdená MSE	0.05856
Maximálna nájdená MSE	2689.2222

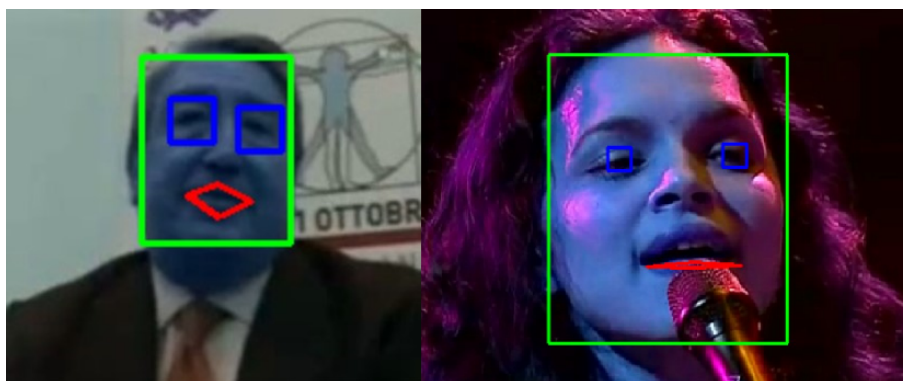
Obr. 14: Spracované snímky

Z priemerných MSE, ktoré sa nachádzajú v tabuľkách môžeme vidieť, že model lepšie hľadá body na tvári na originálnom datasete, hoci výsledky, ktoré sme dostali, nie sú najlepšie a preto sa pozrieme ešte na druhý MTCNN model, kde očakávame lepšie výsledky.

## 2.2 MTCNN pre rozpoznávanie bodov na tvári

Použili sme ešte druhý model pre rozpoznávanie bodov. Ako v predošlom prípade, hľadali sme oči aj ústa na tvári, avšak keďže vieme, že MTCNN bude presnejší ako Haarove kaskády, hľadali sme ústa presnejšie, konkrétne sme ich celé obkreslili pomocou čiar a nie len označili obdĺžnikom, ako sme to spravili v predošlom prípade.

Príklad správne a chybné nájdených bodov na originálnom datasete vyzeral nasledovne:



Obr. 15: Správne nájdené a chybné nájdené body na 2 rôznych snímkach.

Na obrázku môžeme vidieť, že hoci oči boli korektne nájdené vo väčšine prípadov (okrem snímkov, kde oči neboli vôbec nájdené), ústa mali horšiu úspešnosť, ako to môžeme vidieť napríklad na 2. snímke.

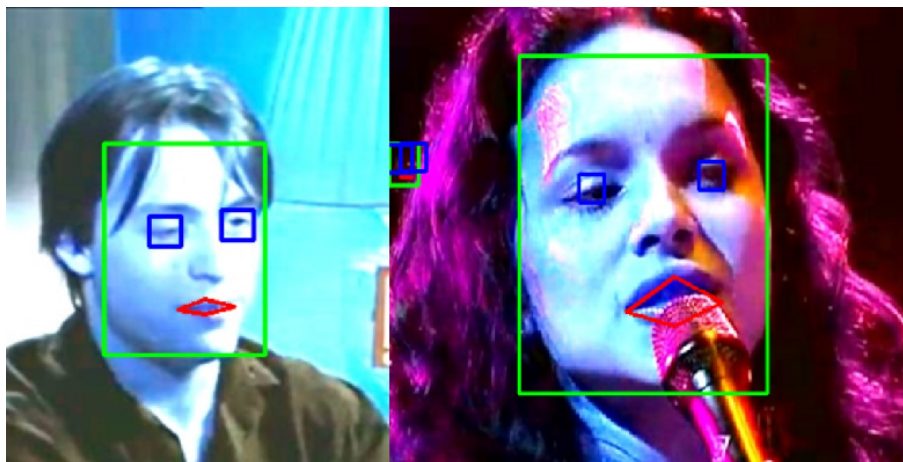
Výsledky MSE pre MTCNN na originálnom datasete boli:

Tabuľka 14: Výsledky MSE pre originálne snímky.

Metrika	Hodnota
Celkový počet spracovaných snímkov	12790
Priemerné MSE	188.5335
Minimálne nájdené MSE	6.4306
Maximálne nájdené MSE	1199.9659

Ako môžeme vidieť, výsledky boli na menšom množstve dát, ako je originálny počet. Je to preto, lebo na zvyšných obrazoch neboli nájdené žiadne body na tvári.

MTCNN model sme zbehli aj na augmentovaných snímkach. Aj tu sme očakávali horšie výsledky, ako to bolo v predošlých prípadoch. Podľa príkladov snímkov, na ktoré sme sa ale pozreli, sme zistili, že oči a ústa neboli nájdené tak zle, ako sme očakávali:



Obr. 16: Správne nájdené a chybné nájdené body na 2 rôznych upravených snímkach.

Na týchto snímkach môžeme vidieť, že nastali aj prípady, kde bolo nájdených viac bodov, ako sme očakávali, no ostatné body boli stále nájdené správne. Treba poznamenať, že v originálnom aj upravenom datasete nastalo, že žiadne body neboli nájdené.

Výsledky pre upravený dataset boli:

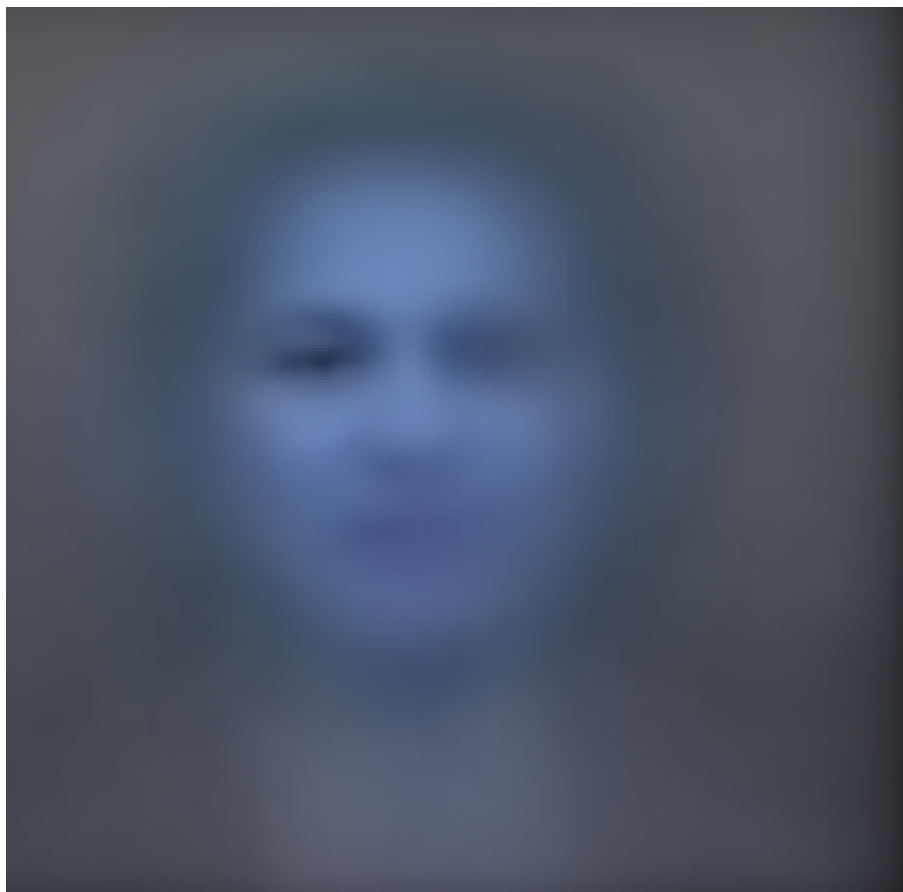
Tabuľka 15: Výsledky MSE pre spracované snímky.

Metrika	Hodnota
Celkový počet spracovaných snímok	10967
Priemerné MSE	168.3779
Minimálne nájdené MSE	8.9326
Maximálne nájdené MSE	1340.8090

Môžeme si všimnúť, že priemerné MSE je lepšie ako pri originálnom datasete, ale na menšom množstve dát. Z toho všetkého vieme na záver usúdiť, že model MTCNN bol lepší ako Haarove kaskády, pričom bolo lepšie používať originálny dataset oproti spracovanému.

### 3 Priemerná tvár

Na koniec sme vytvorili normalizované tváre, ktoré boli rovnakej veľkosti a mali oči na rovnakom mieste. Priemerný obraz pre tieto obrazy vyzerali nasledovne:



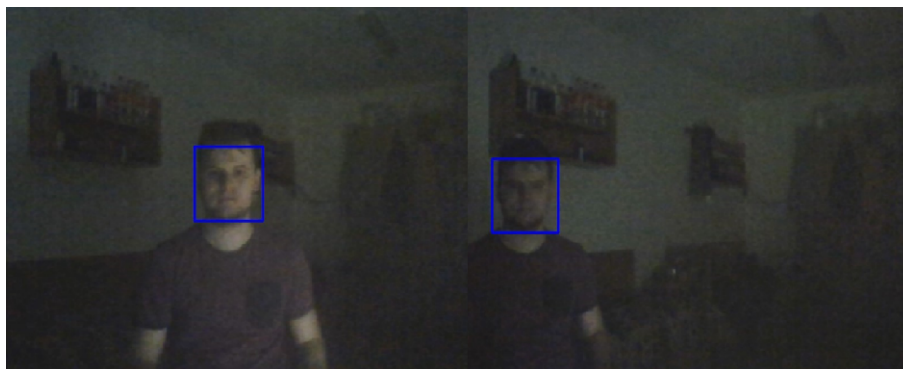
Obr. 17: Priemerný obrázok.

Priemernú tvár sme robili z obrazov, ktoré boli rovnakej veľkosti a mali oči na rovnakom mieste (podľa súradníc). Z toho dôvodu môžeme vidieť, že aspoň jedno oko je zreteľne vidno.

## 4 BONUS

Urobili sme bonusovú úlohu, v ktorej náš program má vedieť vyhľadávať tvár na webkamere. Výsledok vyzeral nasledovne:





Obr. 18: Správne nájdené a chybné nájdené body na 2 rôznych upravených snímkach.

ako môžeme vidieť, program nenašiel tvár len ak bola blízko pred kamerou, ale aj keď sme sa pohli, alebo sme boli ďalej od kamery.

## 5 Zdroje

- Kódy z cvičení
- <<https://chat.openai.com>>
- <<https://stackoverflow.com/questions/62590320/on-what-dataset-was-haarcascade-frontalface-default-xml-trained>>
- <<https://pyimagesearch.com/2021/04/12/opencv-haar-cascades/>>
- <<https://cv-tricks.com/face-detection/real-time-deep-learning-mtcnn/>>