Vysoké učení technické v Brně
Fakulta informačních technologií

Vplyv rozlíšenia a natočenia hlavy na detekciu tváre

Biometrické systémy

Lukáš Dekrét
Dávid Bolvanský

1 Úvod

Cieľom tohto projektu bolo preskúmať vplyv rozlíšenia a jasu na úspešnosť detekcie tváre. Bolo vybraných 5 nástrojov na detekciu tvárí, ktoré sú momentálne najpoužívanejšie a najpopulárnejšie. Na vybranom datasete fotografií sme vykonali niekoľko experimentov, kde sme skúšali meniť rozlíšenie/jas fotografií a sledovali sme, aký vplyv má táto zmena na detektory a ich úspešnosť v detekcií tvárí.

2 Dataset a výber fotografií na experimenty

Na účely experimentovania sme si vybrali dataset *Face Detection Data Set and Benchmark (FBBD)* [1], ktorý bol navrhnutý na skúmanie problému detekcie tvárí.

Z tohto datasetu sme vybrali 100 fotografií, na ktorých sa nachádzala len jedna tvár, a táto tvár mala fixnú pixelovú šírku tváre - 90 pixelov. Výber fotografií mal takéto obmedzenia hlavne z dôvodu experimentovania a stanovania si základných parametrov, ktoré fotografie budú mať. Použitie fixnej pixelovej dĺžky tvárí bolo odporúčané najmä pre experimenty zo zmenami rozlíšenia. V našom výbere fotografií na experimentovanie sme snažili mať čo najviac vyvážené rasové rozloženie, no kvôli prechádzajúcimi obmedzeniam a nedostatku takýchto fotografií toto rovnomerné rozloženie nebolo možné úplne dosiahnuť. Výsledných 100 fotografií tak obsahuje 72 osôb europoidnej rasy, 14 osôb mongoloidnej rasy a 14 osôb negroidnej rasy. Pre účely zisťovania *true negative* prípadov sme si našli 14 ďalších fotografií, kde nie sú ľudské tváre, no obsahujú objekty, ktoré pripomínajú ľudskú tvár (*falošné tváre*). Dataset *FBBD* obsahuje aj anotácie pre niektoré fotografie, no pre väčšinu našich vybraných fotografií informáciu o *ROI* žiaľ neposkytoval.

3 Výber nástrojov na detekciu tvárí

Vybrali nasledovných 5 nástrojov s ktorými sme následne prevádzali experimenty a zisťovali ako obstoja v netradičných / menej ideálnych podmienkach.

MTCNN

Momentálnu *state-of-the-art* detekciu tvári je možné dosiahnuť pomocou *Multi-task Cascade Convolutional Neural Network*. Táto metóda bola predstavená v článku *Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks* [2]. Pri spracovaní daného obrázku ho spočiatku zmeníme v rôznych mierkach, aby sme vytvorili obrazovú pyramídu, ktorá je vstupom nasledujúcej trojstupňovej kaskádovej štruktúry.

[1] `http://vis-www.cs.umass.edu/fddb/`

[2] `https://arxiv.org/pdf/1604.02878.pdf`

Existujú tri stupne klasifikácie:

- **1. fáza**

  Využíva sa plne konvolučná sieť, nazývaná návrhová sieť (P-Net), aby sme získali kandidátne tváre a ich regresné vektory ohraničujúceho obdĺžnika. Potom sú kandidáti kalibrovaní na základe odhadovaných ohraničujúcich regresných vektorov. Následne použijeme nemaximálne potlačenie (non maximum suppression NMS) na zlúčenie vysoko prekrývajúcich sa kandidátov.

- **2. fáza** Všetci kandidáti sú presmerovaní do inej siete CNN, ktorá sa nazýva rafinovaná sieť (R-Net), ktorá ďalej odmieta veľké množstvo falošných kandidátov, vykonáva kalibráciu s regresiou ohraničovacieho obdĺžnika a vykonáva NMS.

- **3. fáza**

  Táto fáza je podobná druhej fáze, až na to, že sa zameriavame na identifikáciu tvárových regiónov s väčším dohľadom. Výstup siete je päť pozícií orientačných bodov tváre.

Haar

Detekcia objektov pomocou Haar kaskádového klasifikátora založeného na znakoch, je efektívna metóda vytvorená Paulom Violom a Michaelom Jonesom predstavená v článku *Rapid Object Detection using a Boosted Cascade of Simple Features* [3]. Používa prístup strojového učenia, kde kaskádovitá funkcia je trénovaná z veľa pozitívnych a negatívnych obrázkov. Následne je použítá na detekciu objektov v iných obrázkoch. Algoritmus spočiatku potrebuje veľa pozitívnych obrazov (obrázky tvárí) a negatívnych obrazov (obrázky bez tvárí), aby mohol trénovať klasifikátor. Potom z toho musí extrahovať znaky, kde každý znak je jedna hodnota získaná odpočítaním súčtu pixelov pod bielym obdĺžnikom (vybraná plocha na tváry) od súčtu pixelov pod čiernym obdĺžnikom (plocha na tváry pri bielom obdĺžniku). OpenCV poskytuje tréningovú metódu a vopred natrénované modely. Predpracované modely sú umiestnené v dátovom priečinku pri inštalácii OpenCV.

HoG

Histogram orientovaných gradientov (HOG) je deskriptor funkcie používaný v počítačovom videní a spracovaní obrazov na účely detekcie objektov. Táto technika počíta výskyt gradientovej orientácie v lokalizovaných častiach obrázka. Táto metóda je podobná metóde histogramov orientácie hrán, deskriptorov transformácie prvkov bez zmeny mierky a kontextov tvarov, líši sa však tým, že sa počíta na hustej mriežke rovnomerne rozmiestnených buniek a na zlepšenie

[3] `https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf`

presnosti sa používa prekrývajúca sa normalizácia lokálneho kontrastu. Základnou myšlienkou deskriptora histogramu orientovaných gradientov je to, že vzhľad a tvar lokálneho objektu v obraze možno opísať rozložením gradientov intenzity alebo smerov okrajov. Obrázok je rozdelený na malé spojené oblasti nazývané bunky a pre pixely v každej bunke je zostavený histogram smerov gradientu. Deskriptor je zreťazenie týchto histogramov. Pre lepšiu presnosť môžu byť lokálne histogramy normalizované kontrastom vypočítaním miery intenzity vo väčšej oblasti obrazu, nazývanej blok, a potom pomocou tejto hodnoty normalizovať všetky bunky v bloku. Táto normalizácia vedie k menšej zmene v osvetlení a zatienení.

CNN

Pri hlbokom učení je konvolučná neurónová sieť (*Convolutional Neural Network , CNN*) triedou hlbokých neurónových sietí, ktoré sa najčastejšie používajú na analýzu vizuálnych snímok. CNN je regularizovaná verzia viacvrstvových perceptrónov. Viacvrstvové perceptróny obvykle znamenajú plne prepojené siete, čo znamená, že každý neurón v jednej vrstve je pripojený ku všetkým neurónom v ďalšej vrstve. „Plná prepojenosť“ týchto sietí ich robí náchylnými k nadmernému spracovávaniu údajov. Medzi typické spôsoby regularizácie patrí pridanie určitej formy merania hmotnosti do stratovej funkcie. CNN však majú odlišný prístup k regularizácii: využívajú hierarchický vzorec v údajoch a zostavujú zložitejšie vzorce s použitím menších a jednoduchších vzorov. Konvolučné siete boli inšpirované biologickými procesmi v tom, že vzor konektivity medzi neurónmi pripomína organizáciu zvieracej vizuálnej kôry. CNN používa relatívne malé predspracovanie v porovnaní s inými algoritmami klasifikácie obrázkov.

DNN

Hlboká neurónová sieť (*Deep Neural Netvork*) je umelá neurónová sieť (*Artificial Neural Network, ANN*) s viacerými vrstvami medzi vstupnou a výstupnou vrstvou. DNN nájde správnu matematickú formulu, aby premenil vstup na výstup, či už ide o lineárny alebo nelineárny vzťah. Sieť sa pohybuje po vrstvách a počíta pravdepodobnosť každého výstupu. Napríklad DNN, ktorá je trénovaná na rozpoznávanie ľudských tvárí, prejde daný obrázok a vypočíta pravdepodobnosť, že ľudské tváre sú na obrázku. Užívateľ môže skontrolovať výsledky a zvoliť, ktoré pravdepodobnosti by sieť mala zobraziť (za použitia prahov atď.) A vrátiť navrhované priradenie. Každá matematická formula ako taká sa považuje za vrstvu. Komplexná DNN má veľa vrstiev, preto názov „hlboké“ siete. DNN môžu modelovať komplexné nelineárne vzťahy. Architektúry DNN generujú kompozičné modely, kde je objekt vyjadrený ako vrstvená kompozícia primitívov. Dodatočné vrstvy umožňujú zloženie spodných vrstiev a potenciálne modelujú komplexné údaje s menším počtom jednotiek ako podobne vykonávaná plytká sieť.

Použili sme nasledovné implementácie detektorov:

- **MTCCN**

  `https://github.com/ipazc/mtcnn`

- **Haar**

  `https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html`

- **HoG**

  `https:/dlib.net/`

- **CNN**

  `https:/dlib.net/`

- **HoG**

  `https://www.cvlib.net/`

4 Experimenty

Vybraných 100 fotiek sme rozčlenili do priečinkov podla rasovej príslušnosti osôb na fotografiách. Pre každú rasu sme zisťovali vplyv zmeny rozlíšenia a jasu na úspešnosť detekcie tvárí. Zobrali sme aj 14 fotografií, ktoré neobsahovali ľudské tváre, ale obsahovali niečo, čo sa na ľudské tváre môže podobať a skúmali sme, či nástroje správe odhalia tento fakt. Nakoniec sme zobrali všetkých 100 fotiek, pridali sme k týchto 14 fotografií, a urobili celé meranie aj pre túto sadu fotografií.

Pomocou vytvoreného skriptu sme vytvorili nové obrázky, ktoré mali znížené rozlíšenie. Detektory sa pokúšali hladať tváre na fotografiách, ktoré boli 10, 20, 30, 40, 50, 60, 80 a 100 % pôvodných fotografií. Vytvorili sme aj nové zosvetlené / ztmavené fotografie, ktoré mali rôznu úroveň jasu. Vyskúšali sme 4 úrovne zosvetlenia (faktor jasu > 1) a 4 úrovne ztmavenia (faktor jasu < 1) fotografií. Následne sme pristúpili k meraniu úspešnosti detekcie tvárí pomocou vyššie uvedených nástrojov. Na meranie výkonnosti detektorov sme použili metriku *Accuracy*.

Presnosť (*Accuracy*) je jedna z metrík pre hodnotenie klasifikačných modelov. Formálne má presnosť nasledovnú definíciu:

$$\text{Presnosť} = \frac{\text{počet správnych predikcií}}{\text{celkový počet predikcií}}$$

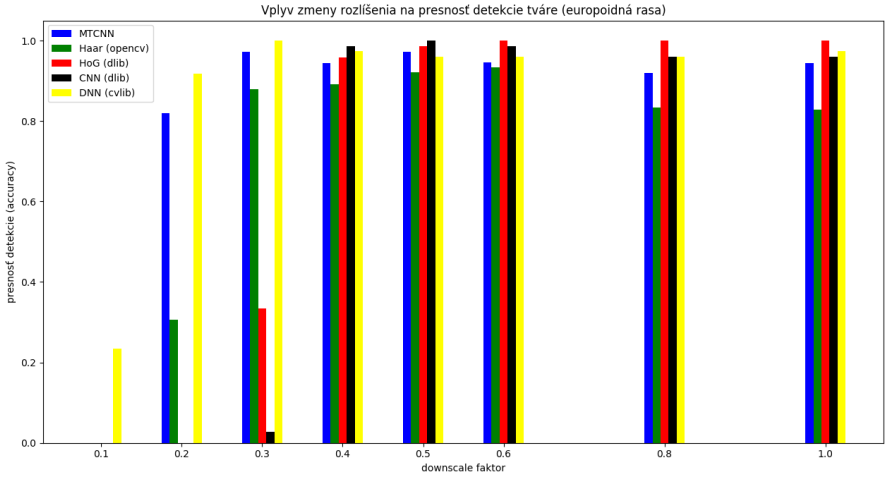
Pre binárnu klasifikáciu sa presnosť môže vypočítať aj z hľadiska pozitívnych a negatívnych prípadov takto:

$$\text{Presnosť} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}}$$

Ak detektor nenašiel tvár na fotografií zaznamenali sme si to ako *false negative*, v prípade falošných tvarí ako *true negative* prípad. V prípade, že detektor našiel jednu alebo viac tvárí na fotografií, kde nebola ľudská tvár, všetky výskyty sa rátali ako *false positive* prípady. Ak našiel jednu alebo viac tvári na fotografii so skutočnou ľudskou tvárou, jeden prípad sa započítal ako nájdenie tváre (*true positive* prípad) a ostatné výskyty ako *false positive* prípady. Ako sme vyššie uviedli, pre naše fotografie neboli dostupné F informácie - ak by boli, bolo by možné použiť metriku *Intersection over union (IoU)* spolu s nejakou hranicou dobrej predikcie a lepšie stanoviť, či jedna z nájdených tvári je naozaj tá správna.

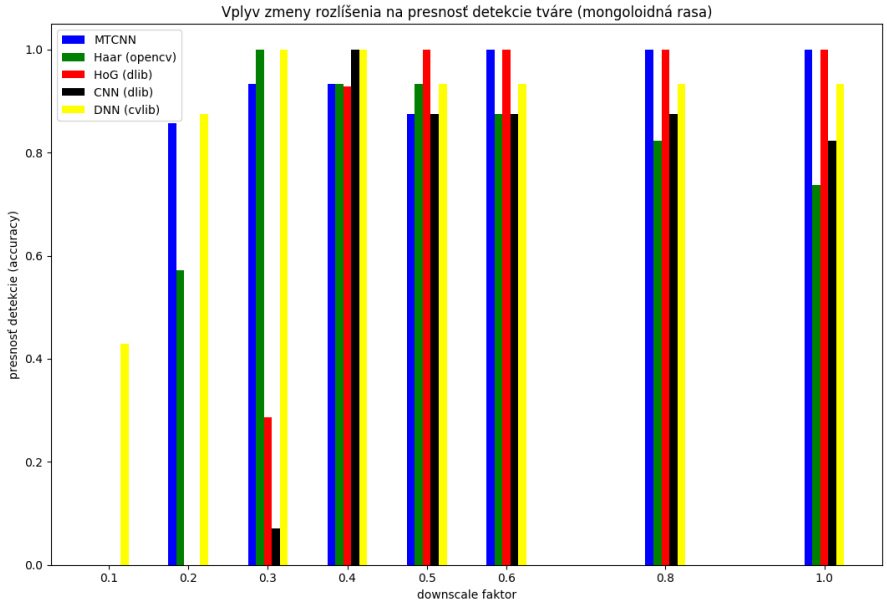
Vplyv zmeny rozlíšenia na presnosť detekcie tvárí

Europoidná rasa



Obr. 1: Vplyv zmeny rozlíšenia na presnosť detekcie tvárí osôb europoidnej rasy

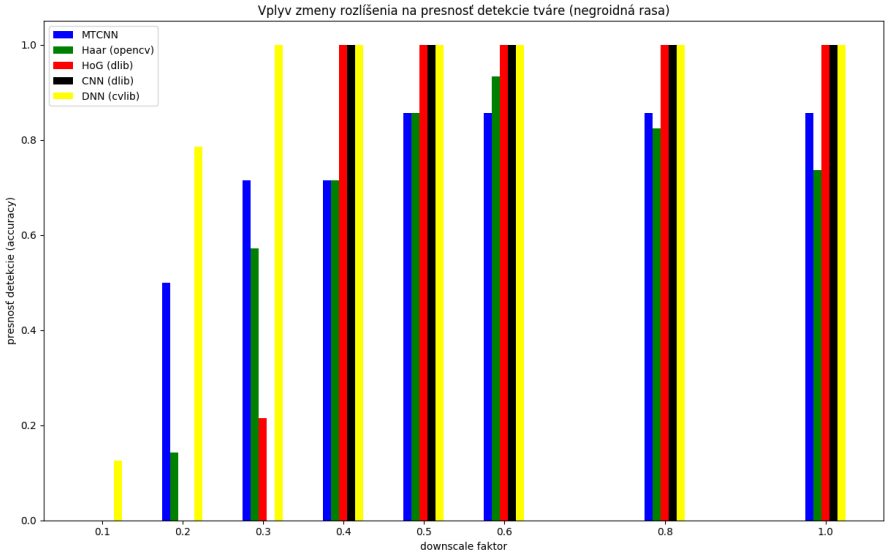
Mongoloidná rasa



Obr. 2: Vplyv zmeny rozlíšenia na presnosť detekcie tvárí osôb mongoloidnej rasy

Z grafov vidíme, že výsledky presnosti detektorov sú veľmi podobné medzi mongoloidnou a europoidnou rasou. Všetky detektory si vedú veľmi dobre aj pri fotografiách, ktoré sú len 40 % ich pôvodnej veľkosti. Za touto hranicou potvrdzujú DNN a MTCNN, že sa jedná o momentálne *state-of-the-art* detektory tvárí a aj pri veľmi malých fotografiách dokážu nájsť tváre a neprodukujú mnoho *false positive* prípadov.

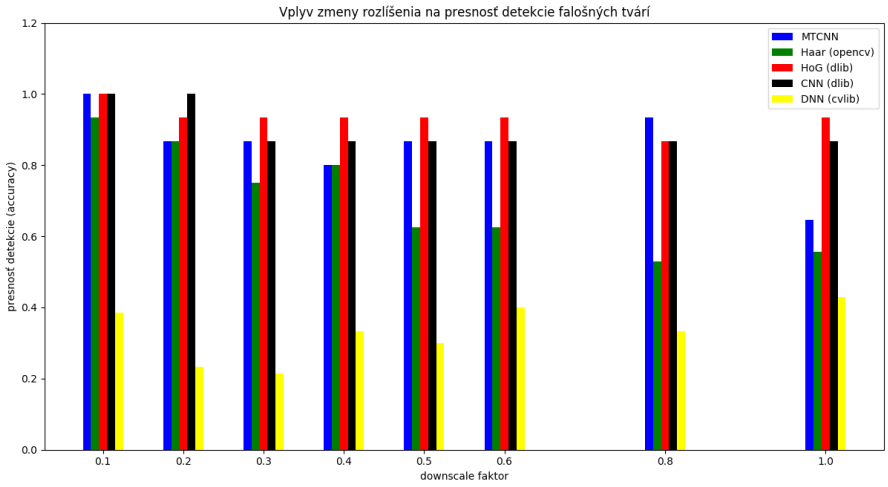
Negroidná rasa



Obr. 3: Vplyv zmeny rozlíšenia na presnosť detekcie tvárí osôb negroidnej rasy

U negroidnej rasy je však možné zaznamenať, že presnosť detektora MTCNN sa u tejto rasy znížila. Trénovacia sada zrejme obsahovala málo osôb tejto rasy.

Falošné tváre



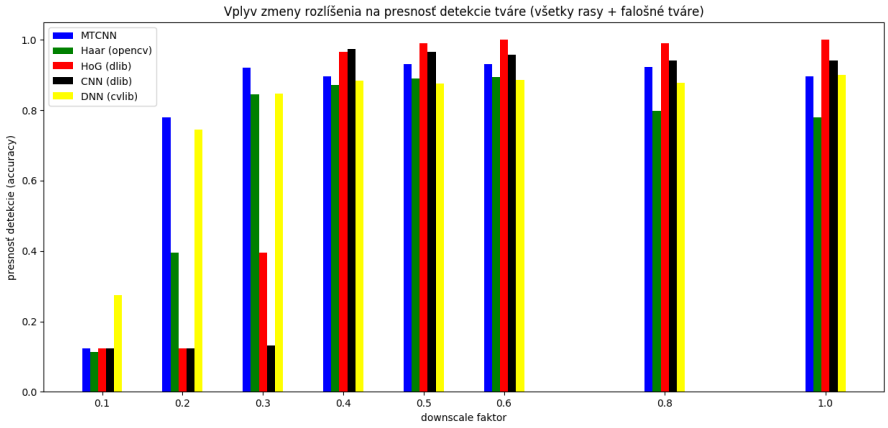
Obr. 4: Vplyv zmeny rozlíšenia na presnosť detekcie tvárí - fotografie falošných tvárí

Vidíme, že detektori MTCNN, HoG, CNN je možné ľahko oklamať fotografiami objektov, ktoré sa podobajú na ľudské tváre, a z menšími rozlíšeniami sa situácia len zhoršuje. Haar je mierne lepší od nich, no detektor DNN je u tohto experimentu jasný vítaz.

Všetky rasy a falošné tváre

Celkovo môžeme zhodnotiť, že všetky detektory sú dosť presné aj pri malých rozlíšeniach. Pri veľmi malých rozlíšeniach sú ale v súčasne dobe najlepšie MTCNN a DNN. Situácia u ostatných detektorov by mohla byť lepšia, ak by boli natrénované na inom datasete alebo spustené s inými parametrami. Graf ukazuje aj niektoré zaujímave a možno na prvý pohľad prekvapivé výsledky, najmä v prípade detektorov CNN, HoG a Haar.

Predtrénovaná CNN umožňuje dobre detegovať tváre len ak je rozlíšenie tváre na fotografii aspoň 80x80. Implemtácia CNN v dlib umožňuje nastaviť parameter `upsample count`, ktorý môže pomôcť pri malých fotografiách - odporúčaná hodnota je 1. Na grafoch si je možné všimnúť zaujímavý pád presnosti CNN na rozmedzí 30 - 40 %, čo však je možné vysvetliť nasledovne: 40 % z našej fixnej šírky tváre (90px) je 36px, po 1x upsample dosahujeme už spomínané minimálne rozlíšenie 80x80. V prípade 30 % je teda jasné, že sa už nachádzame pod touto hranicou a detekcia tvári je veľmi slabá (čo namerané výsledky plne potvrdzujú). Presnosť detekcie u CNN je možné zlepšiť zmenou parametru *upsample count* na vyššie číslo, no doba potrebná na detekciu tváre za znateľne zvýši, čo môže byť už problematické na reálne použitie. Toto vysvetlenie platí aj pre



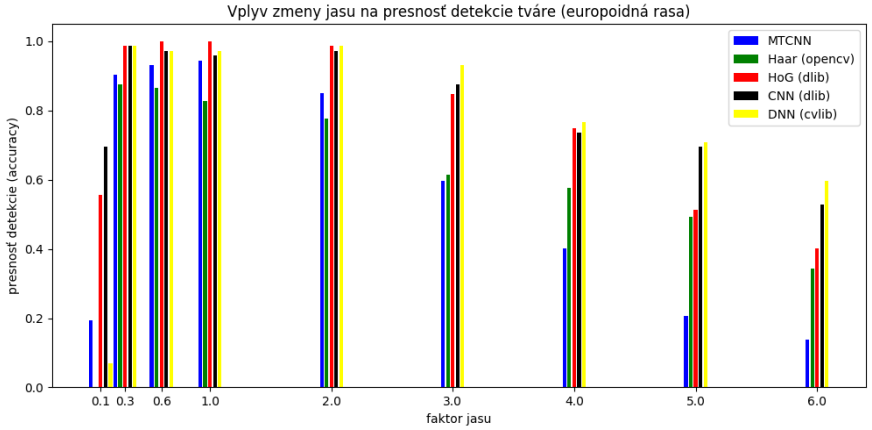
Obr. 5: Vplyv zmeny rozlíšenia na presnosť detekcie tvárí - všetky rasy a falošné tváre

HoG, keďže aj tu bola použitá hodnota 1 ako `upsample count` a aj u HoG z dlib platí, že umožňuje dobre detegovať tváre len ak je rozlíšenie tváre na fotografii aspoň 80x80.

Ako už bolo spomenuté, Haar je založený na princípe obrázkovej pyramídy, ktorá je tvorená na základe *scale faktoru*. Podľa *scale faktoru* sa obrázok zmenšuje a tým sa zvyšuje šanca nájdenia zhody s veľkosťou modelu pre detekciu. Pre experimenty sme použili model `haarcascade_frontalface_default.xml`, ktorý je trénovaný na obrázkoch v rozlíšení 24x24. Hodnota *scale faktoru* 1.1 (ktorá je predvolená v implementácii Haar v OpenCV) určuje, že obrázok je medzi dvoma vrstvami redukovaný o 10 % (tj. faktor definuje koľko vrstiev bude). Zmenou *scale faktor* parametru je teda možné zlepšiť presnosť detekcie, no podobne ako u CNN, zmena tohto parametra za účelom zvýšenia presnosti prináša so sebou aj zvýšenie doby potrebnej na detekciu tváre.

Vplyv zmeny jasu na presnosť detekcie tvárí

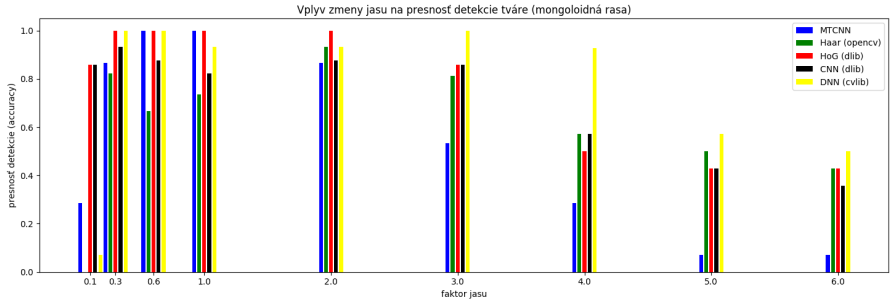
Europoidná rasa



Obr. 6: Vplyv zmeny jasu na presnosť detekcie tvárí osôb europoidnej rasy

Detektori CNN a DNN bol u osôb europoidnej rasy výrazne lepší od ostatných detektorov pri vysokých úrovniach zosvetlenia fotografie.

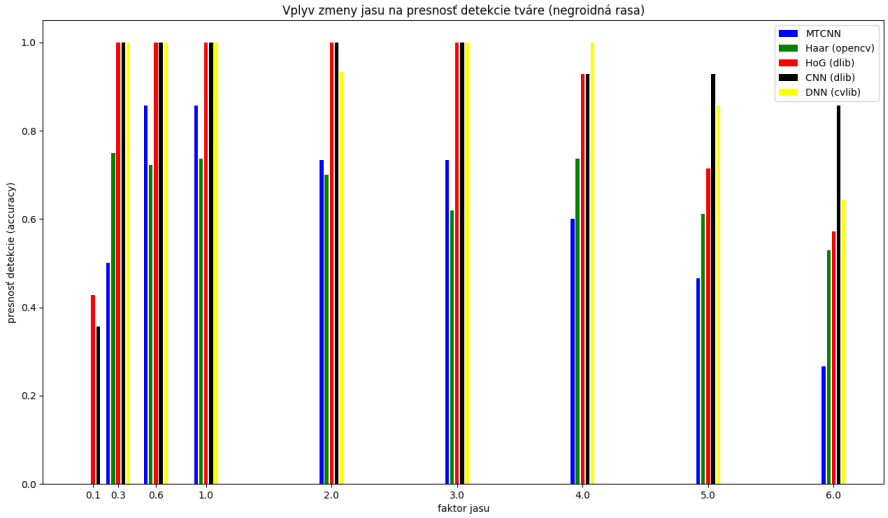
Mongoloidná rasa



Obr. 7: Vplyv zmeny jasu na presnosť detekcie tvárí osôb mongoloidnej rasy

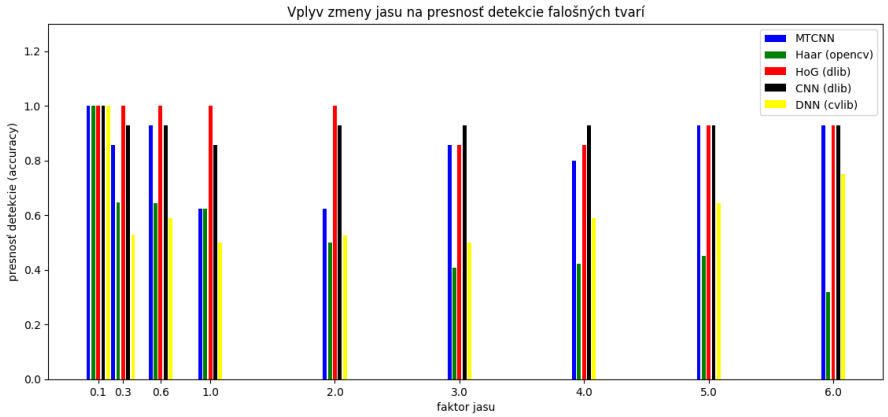
Detektor DNN bol u osôb mongoloidnej rasy výrazne lepší od ostatných detektorov pri vysokej úrovni zosvetlenia fotografie.

Negroidná rasa



Obr. 8: Vplyv zmeny jasu na presnosť detekcie tvárí osôb negroidnej rasy

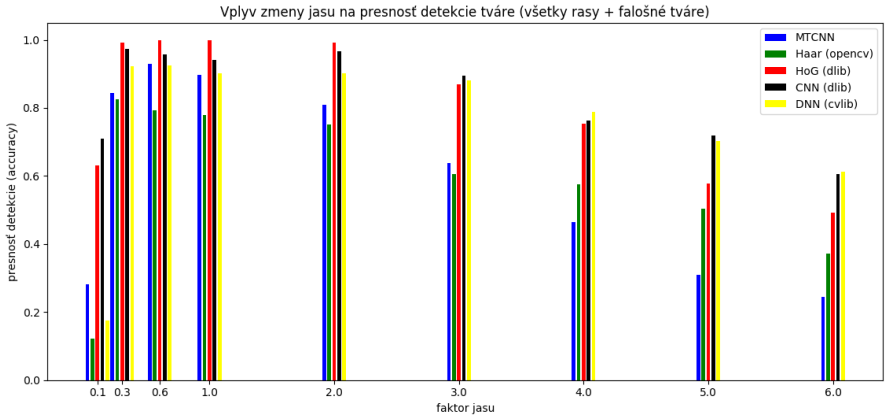
Falošné tváre



Obr. 9: Vplyv zmeny jasu na presnosť detekcie tvárí - fotografie falošných tvárí

Detektor DNN potvrdil, to čo ukazoval trend z podobného merania u zmeny rozlíšenia, a nedal sa oklamať falošnými tvárami. V tom prípade avšak Haar dokázal poraziť DNN. Detektori MTCNN, HoG, CNN sa dali veľmi ľahko oklamať falošnými tvárami.

Všetky rasy a falošné tváre



Obr. 10: Vplyv zmeny jasu na presnosť detekcie tvárí - všetky rasy a falošné tváre

Celkovo môžeme zhodnotiť, že detektory v prípade zmeny jasu začali byť značne neúspešné až pri veľmi nízkej úrovni stmavenia / zosvetlenia. U týchto meraní sa výraznejšie vyčlenila skupina detektorov HoG, CNN, DNN, ktorá bola výrazne presnejšia v detekcii tvárí ako Haar a MTCNN. Je celkom prekvapivé, že MTCNN ako momentálne jeden z *state-of-the-art* detektorov tvárí dopadol najhoršie, kde vidíme, že najmä vyššie úrovne zosvetlenia boli pre tento detektor veľkým problémom.

5 Záver

Namerané dáta ukazujú, že zmena rozlíšenia nemá výrazný vplyv na úspešnosť a presnosť detekcie, keďže detektori používajú techniky predspracovania fotografií (*rescaling*), ktoré pomáha eliminovať efekt vplyvu rozlíšenia na detekciu tvárí. Experimenty taktiež poukazujú na to, že niektoré detektory horšie zvládajú mierne zmeny úrovne jasu, čo má vplyv na ich úspešnosť a presnosť v detekcii tvárí.