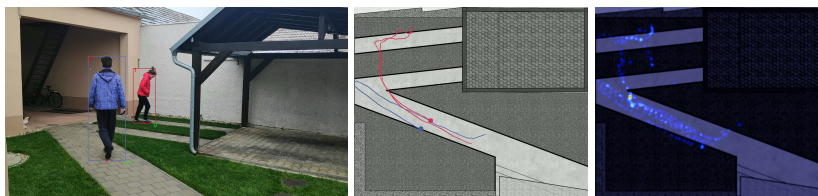


# Sledování pohybu návštěvníků muzejních expozic

Matej Viskupič\*



## Abstrakt

Cieľom tejto práce je preskúmať možnosti a navrhnúť systém monitorovania návštevníkov muzejných expozícií. Na rozdiel od doterajších metód naše riešenie skúma využitie kamerovej technológie. Tento prístup priniesol tri podproblémy: (1.) detekciu ľudí v kamerovom zázname za pomoci konvulčnej neurónovej siete (2.) presné určenie polohy osôb v monitorovanom priestore za použitia externej konfigurácie kamier a (3.) identifikáciu detegovaných osôb. Výstupom nášho riešenia je štatistika návštevnosti jednotlivých expozícií, mapa znázorňujúca pohyb návštevníkov a tepelná mapa zobrazujúca čas strávený pred jednotlivými expozíciami. Tieto výstupy môžu prispieť k lepšiemu vyhodnocovaniu potrieb návštevníkov a k efektívnejšiemu výberu alebo rozmiestneniu exponátov.

**Kľúčové slová:** Sledovanie osôb — Identifikácia osôb — Heatmap of presence

**Priložené materiály:** [Demonštračné video](#)

\*[xvisku01@stud.fit.vutbr.cz](mailto:xvisku01@stud.fit.vutbr.cz), *Fakulta Informačních Technologií, Vysoké Učení Technické v Brně*

## 1. Úvod

Múzea a galérie v súčasnosti využívajú svoj personál zaznamenávajúci dáta o pohybe a správaní návštevníkov na papierový formulár či špecializované zariadenia, ktoré musí mať návštevník stále so sebou. Táto práca rozoberá alternatívny prístup k problému sledovania návštevníkov a to za pomoci využitia kamerovej technológie. Tento prístup nemá efekt na zážitok návštevníka, je ľahko rozšíriteľný a prispôsobiteľný rôznym prostrediam. Poskytuje rýchle vyhodnotenie požadovaných dát a dodanie presnej spätnej väzby.

Pre každé odvetvie, ktorého hlavným princípom je interakcia s ľuďmi, sú dôležité informácie o svojich zákazníkoch, o ich preferenciách či čase strávenom pri jednotlivých elementoch. Tieto princípy platia aj pre múzea či galérie. Rovnako ako web-stránky pracujú s rozmiestnením a tvarom rôznych elementov, aby si čo najdlhšie udržali pozornosť návštevníkov, tak aj múzea a galérie pracujú s rozmiestnením ex-

ponátov. Informácie od návštevníkov o ich polohe, pohybe či čase strávenom pri jednotlivých exponátoch, slúžia ako dôležité metriky na optimalizovanie procesu zvyšovania návštevníkovej spokojnosti.

Väčšina z výskumov na túto tému bola vykonávaná spôsobom „ceruzky a papiera“. Táto metóda spočíva v rozmiestnení výskumníkov na kľúčové miesta a ručným zapisovaním zistených dát. Ďalšou často používanou metódou je takzvané „front-end vyhodnotenie“, kedy sú návštevníkom kladené otázky pred vstupom a po opustení monitorovanej oblasti [1].

Boli skúmané aj komplexnejšie riešenia využívajúce externé zariadenia, ďalej nazývané lokátory, ktoré má návštevník stále so sebou počas celej jeho návštevy monitorovanej oblasti. Tento lokátor môže byť napríklad jeho chytrý telefón s technológiou Wi-Fi [2]. Napriec monitorovanou oblasťou sú rozmiestnené prístupové body a na základe intenzity prijatého signálu Wi-Fi (RSS Received Signal Strength) z jed-

notlivých bodov je v špeciálnej aplikácii vypočítaná jeho poloha. Táto technológia je už využívaná v niektorých z popredných múzeí, napríklad v Národnej Galérii a Národnom Historickom múzeu v Londýne [1].

Lokátor nemusí byť len chytrý telefón, môže využívať rôzne technológie. Štúdia [3] na tento účel navrhuje využívanie technológie rádiových vln (RFID - Radio-frequency identification). Pred expozíciami sú vhodne rozmiestnené rádiové majáky. Lokátor, ktorý má návštevník stále so sebou, po vstupe do dosahu majáku s ním začne komunikovať, čím určí jeho jednoznačnú polohu. Tento proces možno vidieť na obrázku 1. Existujú podobné štúdie, ktoré sa venujú rôznym typom rádiových technológií či sa už jedná o Bluetooth, UWB (Ultra-wide band) alebo GPS (Global Positioning System).

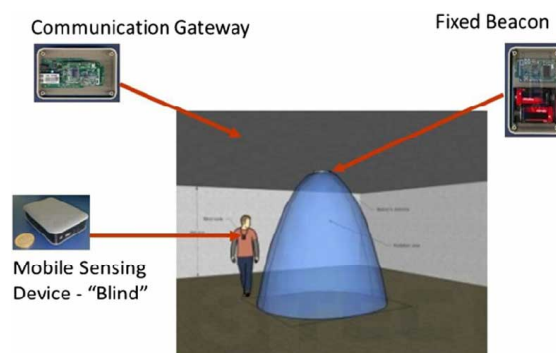
Naše riešenie využíva kamerového systému na identifikáciu návštevníkov a ich sledovanie. Dokáže spracovávať zábery z viacerých kamier naraz. Na týchto záberoch sú následne, za využitia prvotriednej neurónovej siete, identifikované osoby. Týmto detekciám osôb je na základe konfigurácie kamier pridelená presná poloha v rámci súradnicového priestoru monitorovanej oblasti. Nad detekciami s pridelenou presnou polohou je v zápätí vykonaná re-identifikácia osôb. Vďaka tejto re-identifikácii je uskutočnené sledovanie osôb naprieč jednotlivými snímkami a taktiež naprieč monitorovanou oblasťou. Po ukončení sledovania alebo na vyžiadanie sú nazbierané dáta vyhodnotené a graficky vizualizované.

Náš prístup, na rozdiel od existujúcich riešení, nezaťažuje návštevníkov, keďže od nich nevyžaduje vyplňanie formulárov či nosenie špecializovaných lokátorov. Okrem toho je možná implementácia na už existujúcej sieti kamerových systémov. Na rozdiel od rozmiestnenia celej siete malých zariadení s krátkym dosahom (RFID) [3], senzory (kamery) navrhovaného systému dokážu pokryť v malom počte celé miestnosti. Tento fakt znižuje počet rušivých elementov v múzeu a zložitosť požadovanej infraštruktúry. Navyše je naše riešenie ľahko rozšíriteľné na ďalšie odvetvia ako napríklad obchody, letiská či stanice.

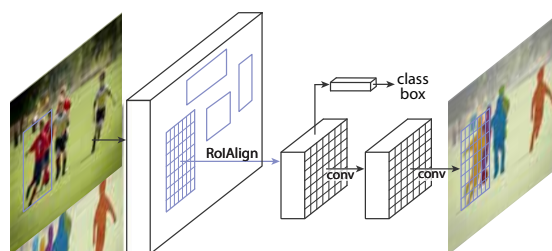
## 2. Mask R-CNN

Na detekciu osôb v obrazovom zázname náš systém používa prvotriednu konvolučnú neurónovú sieť Mask R-CNN [4].

Konvolučné neurónové siete za pomoci aplikácie princípov, ktoré v jednoduchosti spomeniem v tejto časti, inovujú postup klasických umelých neurónových sietí (ANN Artificial Neural Network) čím redukujú



**Obrázok 1.** Vizualizácia existujúceho riešenie pokrytia expozície majákom RFID. Obrázok pochádza zo štúdie [3].



**Obrázok 2.** Princíp fungovania Mask R-CNN [4]. Po prvom kroku RPN (Region Proposal Network), ktorý je zvýraznený modrou farbou nasleduje paralelná časť: klasifikácia tried (horná časť s nápisom „class box“) a určenie masky pokrývajúcej objekt (spodná časť pozostávajúca z dvoch vrstiev).

počet potrebných parametrov [5]. Táto redukcia je spôsobená hlavne konvolučnou a pooling vrstvou, ktoré znižujú dimenzie vstupu. Presnejšie sa konvolučná neurónová sieť skladá z nasledujúcich vrstiev[6]:

*Vstupná vrstva* inicializuje vstupné dáta a normalizuje ich do rozmedzia [0,1].

*Konvolučná vrstva* predstavuje jadro celej neurónovej siete. Na pôvodné obrazové dáta postupne aplikuje konvolučné jadro. Počíta konvolučný výsledok násobením dát z predchádzajúcej vrstvy a ich následným súčtom.

*Pooling vrstva* sa využíva na zníženie dimenzií výsledkov predchádzajúcich vrstiev. Veľmi často používané metódy všeobecného pooling-u sú extrahovanie maximálnej hodnoty z oblasti a počítanie priemernej hodnoty v danej oblasti.

*Plne prepojená vrstva* tvorí poslednú vrstvu CNN. Svojou štruktúrou pripomína klasické ANN a výstupom sú už výsledné pravdepodobnosti pre jednotlivé triedy objektov (labels).

Použitá sieť Mask R-CNN stavia na základoch metódy Faster R-CNN [7], ku ktorým pridáva detekciu masky pokrývajúcej detegovaný objekt. Využíva rovnaký prvý krok RPN (Region Proposal Network)

ako Faster R-CNN. V druhom kroku k už existujúcej „vetve“ klasifikácie tried a zisťovania ohraničujúceho obdĺžnika pridáva Mask R-CNN paralelné detegovanie masky objektu. Princíp paralelného výpočtu možno vidieť na obrázku 2.

Konvolučná neurónová sieť Mask R-CNN, ktorá bola použitá v implementácii, bola trénovaná na data-sete COCO [8]. Bližšie podrobnosti k spôsobu trénovania sú popísané v štúdií [4].

### 3. Určenie polohy detegovanej osoby

Po úspešnej detekcii osoby v kamerovom zázname poznáme jej koordináty v súradnicovom priestore kamery. Tieto súradnice potrebujeme previesť do súradnicového priestoru reálneho sveta, aby sme mohli určiť jej polohu. Táto problematika sa nazýva Problém n-bodovej perspektívy (PnP - Perspective n-point problem). Výsledok spracovania tohto problému je vidieť na obrázku 4.

Pre vyriešenie tohto problému potrebujeme poznať súradnice niekoľkých bodov v oboch súradnicových priestoroch. Existuje viacero štúdií, ktoré riešia tento problém, ktoré vyžadujú rôzny počet týchto oantovaných bodov. Po vyriešení PnP dostávame rotačno-translačnú maticu určujúcu polohu kamery v rámci súradnicového priestoru sveta. Potrebujeme taktiež poznať vnútorné nastavenia kamery a parametre optiky, s ktorou pracujeme. Po zozbieraní všetkých týchto údajov môžeme zostaviť nasledujúci vzťah:

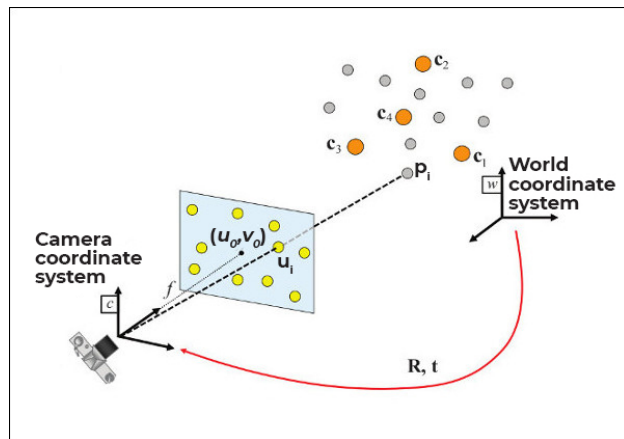
$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Kde  $[X, Y, Z]$  reprezentujú 3D súradnice reálneho sveta,  $f_x$  a  $f_y$  sú ohniskové vzdialenosti kamery,  $c_x$  a  $c_y$  vyjadrujú hlavné body použitej optiky,  $r_{ij}$  a  $t_i$  predstavuje rotačno-translačnú maticu,  $s$  vyjadruje škálu a body  $[u, v]$  predstavujú 2D koordináty v kamerovej projekcii. Tieto hodnoty sú vizualizované na obrázku 3.

Za pozíciu osoby v súradnicovom systéme kamery je považovaný stred dolnej hrany ohraničujúceho obdĺžnika detekcie. Tento bod je znázornený červenou farbou na obrázku 8c.

### 4. Re-identifikácia ľudí naprieč časom a miestom

Môžeme predpokladať, že v monitorovanom priestore sa budú nachádzať „slepé“ miesta, ktoré nebudú pokryté kamerami. Taktiež v reálnom prostredí môže nastať



**Obrázok 3.** Vizualizácia PnP [9].  $R$  a  $t$  predstavuje rotačnú a translačnú maticu. Body  $c_1$  až  $c_4$  sú na-anotované a využité na kalibráciu. Translácia bodov medzi systémami je ilustrovaná bodmi  $u_i$  a  $p_i$ . Bod so súradnicami  $u_0, v_0$  v snímku  $f$  predstavuje neznámy hľadaný bod.

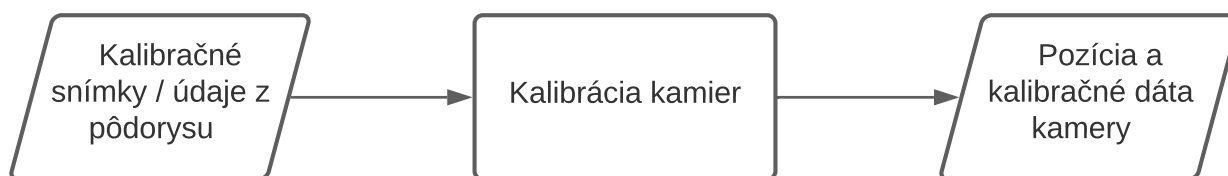


**Obrázok 4.** Vizualizácia roviny predstavujúcej zem. Pozície jednotlivých bodov boli zistené funkciou PnP. Konfigurácia využívala 11 bodov so známymi súradnicami (rohy vstupu do budovy + rohy krabice v priestore).

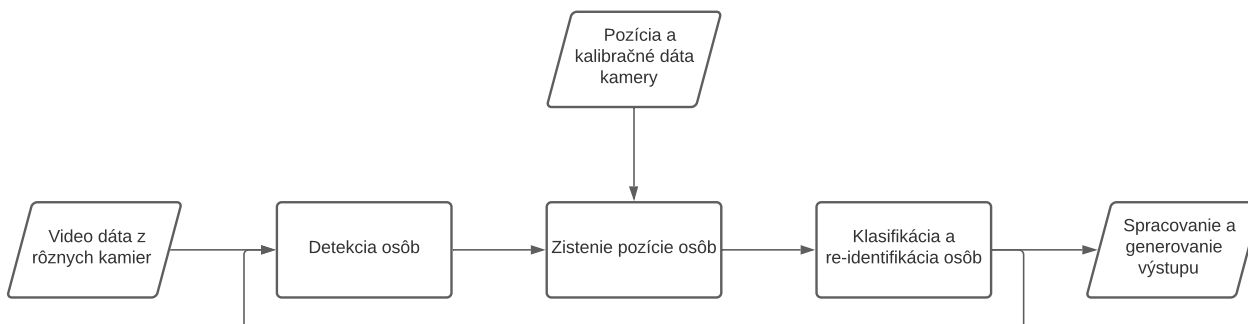
situácia, v ktorej je sledovaná osoba prekrytá iným objektom a nie je možné ju naďalej detegovať. Aby sme sa s takýmito situáciami vysporiadali, potrebujeme z detekcie osoby extrahovať jej črty, zaradiť ich a pri novej detekcii ju na základe týchto črt znovu identifikovať. Tento proces sa nazýva re-identifikácia, re-ID.

V tejto práci sa sústreďíme na krátkodobú identifikáciu [10], v ktorej predpokladáme nemennosť sledovaných vlastností naprieč časom, t.j. konkrétnou návštevou sledovaného objektu. Za vhodné vlastnosti pokladáme farbu alebo textúru oblečenia [11]. Náš systém podľa masky, poskytnutej konvolučnou neurónovou sieťou Mask R-CNN, vypočíta histogram charakterizujúci farbu oblečenia osoby. Pre výpočet podobnosti medzi histogramami dvoch osôb bola využitá Hellingerova vzdialenosť.

Rovnako dôležitou vlastnosťou každej detekcie je



**Obrázok 5.** Diagram pomocného systému na kalibráciu kamier. Jeho výstupom sú kalibračné dáta, ktoré vyžaduje navrhovaný systém.



**Obrázok 6.** Diagram fungovania navrhovaného systému

jej poloha v súradnicovom priestore monitorovaného objektu. V prípade použitia viacerých kamier môže byť jedna osoba detegovaná viackrát, raz na každej kamere, na ktorej je vidieť. Tieto detekcie sa môžu líšiť v póze v akej daná kamera osobu zachytila. Táto zmena pózy môže znamenať aj zmenu ohraničujúceho obdĺžnika detekcie a následne odlišnú polohu, keďže poloha osoby je závislá na tomto obdĺžniku. Na spájanie týchto odlišných polôh patriacich jednej osobe do jedinej polohy bol použitý modifikovaný algoritmus zo štúdie [12]. V modifikovanej forme môže byť použitý aj na neprekrývajúce sa pohľady kamier.

Na následné vyhodnotenie a priradenie detekcií jednotlivým osobám je použitý Munkresov algoritmus (tiež známy ako maďarský algoritmus)[13], ktorý vyberie najlepšiu kombináciu osôb a detekcií.

## 5. Implementácia navrhovaného systému

Detegovanie ľudí, zistenie ich polohy, ich identifikácia ale aj načítanie dát a generovanie výstupu boli naprogramované v jazyku Python. Pri implementácii bolo myslené na potenciálne budúce využitie programu v Národnom múzeu v Prahe, preto sú všetky spomínané časti ľahko prispôsobiteľné. Celý systém a jeho podčasti majú pevne stanovené štruktúry dát, ktorými komunikujú, preto sa dajú ľahko vymeniť, doplniť, prispôsobiť prostrediu a podmienkam.

Pred samotným používaním systému musíme poznať polohu kamier v rámci monitorovaného priestoru. Tieto dáta vieme získať za pomoci vyriešenia problému PnP popísaného v kapitole 3. Princíp fungovania skriptu na získanie týchto dát je zobrazený na obrázku

5. Vyžaduje vhodný výber kľúčových bodov využitých na kalibráciu, ktorých polohu poznáme. Na zistenie polohy je potrebné mať prístup k pôdorysu monitorovaného priestoru alebo vykonať meranie priamo na mieste. Najlepšie výsledky konfigurácie sú dosiahnuté, ak sa konfiguračné body nachádzajú v 3 na seba kolmých rovinách.

Jednotlivé časti systému ako sú vyobrazené na obrázku 6 sú implementáciou vyššie spomínaných princípov. Od detekcie osôb popísanej v kapitole 2, cez určenie polohy približené v kapitole 3, po klasifikáciu a re-identifikáciu podľa kapitoly 4. Každá časť je reprezentovaná svojou abstraktnou triedou, ktorá obsahuje požadované rozhranie. Vďaka štandardizovanému rozhraniu sa každá časť systému správa do určitej miery samostatne. Tento fakt prispieva k flexibilitie riešenia. Zmena medzi systémom s viacerými kamerami a jednou kamerou spočíva vo výmene jedinej triedy - načítavania záznamu, pričom systém táto zmena vôbec nezaťažuje.

## 6. Testovanie

Testovanie riešenia prebiehalo na datasete WILDTRACK [14] za účelom vyhodnotenia úspešnosti riešenia. Rovnako bolo riešenie testované na vlastných nahrávkach pre prehľadnejšiu vizualizáciu výstupu riešenia.

### 6.1 WILDTRACK

Tento dataset obsahuje obrazové záznamy zo 7 GoPro kamier umiestnených pred hlavnou budovou univerzity ETH v Zürichu. Záznamy boli natočené za vhodného počasia vo FullHD rozlíšení. Celkovo je poskytnutých



**Obrázok 7.** Ukážka zosynchronizovaného pohľadu z datasetu WILDTRACK [14].

400 snímkov z každej kamery pri snímkovacej rýchlosti 2 snímky za sekundu. Zorné uhly jednotlivých kamier sa prekrývajú, pričom v priemere sa na jednom snímku nachádza až 20 ľudí. Spolu poskytujú približne 56000 jednotlivých detekcií. Náhľad z tohoto datasetu je možno vidieť na obrázku 7.

Riešenie bolo hodnotené podľa metrík MOTChallenge zaoberajúcich sa sledovaním viacerých objektov [15] a metrík merajúcich presnosť lokalizácie [16]. Presnejšie v tabuľkách 2 a 3 sú použité nasledovné metriky:

**F-Skóre** reprezentuje harmonický priemer medzi *presnosťou* (precision) a *spätného vybavenia si* (recall) detekcií.

**MODA** (Multiple Object Detecion Accuracy) vyjadruje presnosť detekcie viacerých objektov: s akou presnosťou boli objekty identifikované v scéne.

**MODP** (Multiple Object Detecion Precision) označuje presnosť lokalizácie detekcií: s akou presnosťou bola nájdená poloha objektov v scéne.

**MOTA** (Multiple Object Tracking Accuracy) zohľadňuje výmeny identít medzi jednotlivými detekciami a rovnako tiež priradenie neexistujúcich detekcií (False Positives) a nenájdenie existujúcich detekcií (False Negatives): s akou presnosťou boli objekty sledované naprieč snímkami.

**MT, PT a ML** (Mostly tracked, partially tracked, mostly lost) predstavujú počet trajektórií, ktoré boli sledované aspoň 80%, 79-20% a menej ako 20% zo svojej celkovej dĺžky.

Riešenie bolo porovnané s viacerými špičkovými algoritmi a to konkrétne s POM-CNN [17], DeepMCD [18], RCNN-Projected [19] a Deep-Occlusion [20]. Metóda DeepMCD obsahuje viacero variantov, ktoré využívajú rôzne modely, GoogLeNet [21] ResNet-18 [22] či DenseNet-121[23], a sú bližšie popísané v tabuľke 1.

Ako možno vidieť v tabuľke 2, náš systém si počínal veľmi dobre v detegovaní osôb. Využíva neurónovú

**Tabuľka 1.** Varianty DeepMCD s príslušným tréningovým datasetom a použitým modelom

Metóda	Tréningový dataset	Model
Pre-DeepMCD	PETS	GoogLeNet
Top-DeepMCD	WILDTRACK	GoogLeNet
ResNet-DeepMCD	WILDTRACK	ResNet-18
DenseNet-DeepMCD	WILDTRACK	DenseNet-121

sieť, ktorá bola trénovaná na datasete COCO [8] a medzi metódami, ktoré neboli trénované na datasete WILDTRACK, dosahoval veľmi dobré výsledky. Ako bolo možné očakávať, v porovnaní s metódami trénovanými na datasete WILDTRACK, si počínal výrazne horšie.

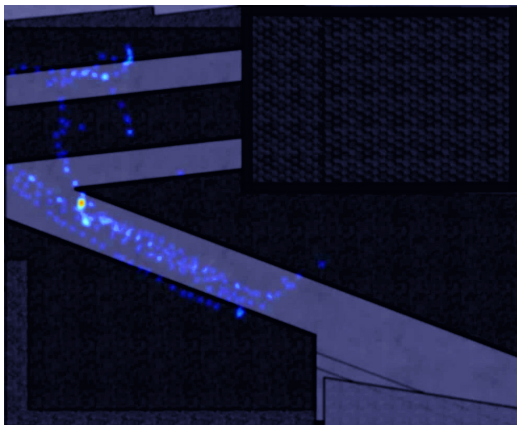
**Tabuľka 2.** Výsledky testov [14] detekcií rozdielnych metód na datasete WILDTRACK. *Spodná časť* oddelená prerušovanou čiarou predstavuje metódy trénované v iných datasetoch ako WILDTRACK. *Horná časť* naopak metódy trénované na datasete WILDTRACK.

Metóda	F-Skóre	MODA	MODP
Deep-Occlusion	0.87	0.74	0.54
ResNet-DeepMCD	0.83	0.68	0.64
DenseNet-DeepMCD	0.80	0.64	0.67
Top-DeepMCD	0.79	0.60	0.64
Náš systém	0.69	0.40	0.63
POM-CNN	0.63	0.23	0.31
RCNN-projected	0.53	0.11	0.18
Pre-DeepMCD	0.52	0.33	0.53

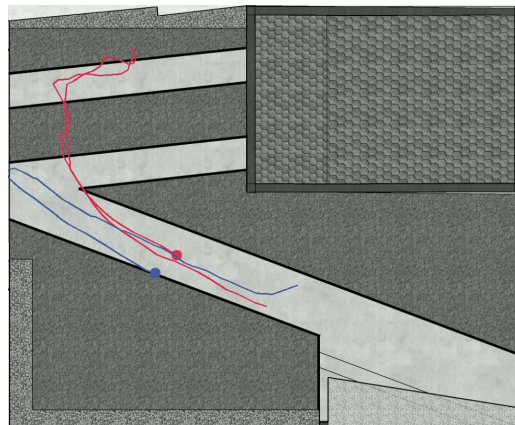
V tabuľke 3 možno vidieť porovnanie výsledkov sledovania osôb. Náš systém dosahoval horšie výsledky v detekcii osôb ako metóda Deep-Occlusion, s ktorou je porovnávaný v tejto tabuľke. Preto aj v globálnych metrikách, ktoré berú do úvahy všetky osoby v datasete (ako je metrika MOTA), dosahuje výrazne horšie výsledky. Nie sme schopní sledovať osobu, ktorú sme

**Tabuľka 3.** Výsledky testov [14] sledovania objektov rozdielnymi metódami na datasete WILDTRACK.

Metóda	MODA	MOTA	MT	PT	ML
Deep-Occlusion+KSP	0.75	0.69	49 (29%)	79 (46%)	43 (25%)
DeepOcclusion+KSP+ptrack	-	0.72	72 (42%)	74 (43%)	25 (15%)
Náš systém	0.40	0.25	70 (82%)	14 (17%)	1 (1%)



**(a)** Výsledná tepelná mapa zobrazujúca miesta, na ktorých osoby strávili najviac času.



**(b)** Výsledné trajektórie detegovaných osôb vykreslené dominantnou farbou ich oblečenia.



**(c)** Proces detekcie a identifikácie osôb v zázname.

**Obrázok 8.** Vizualizovaný výstup navrhnutého systému: tepelná mapa 8a trajektórie 8b a detekcie 8c.

nedetegovali. Avšak pri porovnaní dĺžky trajektórií sledovaných osôb, ktorá vyjadruje akú časť z celkovej trajektórie osoby sme zachytili, dosahuje podobné výsledky.

## 6.2 Vlastné zábery

Pre prehľadnejšiu vizualizáciu výsledného riešenia som použil vlastné zábery. Na obrázku 4 možno vidieť nakonfigurovanú kameru, na ktorej sú ďalej detegované a identifikované osoby 8c. Na pôdorys sledovanej oblasti sú vykreslené trajektórie detegovaných osôb 8b a tepelná mapa 8a zobrazujúca miesta, na ktorých osoby strávili najviac času. Toto vykresľovanie je možné vykonávať na vyžiadanie, či už v reálnom čase súbežne

s detekciou a identifikáciou osôb alebo po skončení spracovania záznamu.

## 7. Záver

Táto práca popisuje najnovšie technológie v oblasti identifikácie a sledovania ľudí. Ponúka ľahko prispôbitelné riešenie využiteľné v rôznych odvetviach. Zároveň je tento prístup porovnávaný s už existujúcimi metódami ako na sledovanie ľudí, tak na zbieranie dát v prostredí múzeí.

V detekcii osôb dosahuje medzi netrénovanými metódami veľmi dobré výsledky. Avšak pri porovnaní s trénovanými metódami sú zreteľne viditeľné časti, primárne sledovanie detegovaných osôb, v ktorých je

priestor na zlepšenie.

Na zber dát vyžaduje prítomnosť len kamerového systému bez žiadneho zásahu do návštevníkovho zážitku zo samotnej expozície. Taktiež tento prístup nepotrebuje komplexný systém zariadení a dá sa ľahko rozširovať naprieč monitorovaným objektom podľa potreby.

Navrhnutý systém pozostáva z častí, ktoré fungujú ako samostatné jednotky. V prípade vyvinutia výkonnejšej technológie existuje možnosť jej integrovania bez nutnosti zasahovať do systému ako celku. Naše riešenie môže byť využité nielen v múzeách či galériách ale aj na letiskách, nákupných centrách či staniách. Kdekoľvek, kde je z finančného či bezpečnostného hľadiska vhodné sledovať správanie ľudí. Navyše vďaka jeho členeniu na podsystémy môže byť riešenie, výmenou detekčnej neurónovej siete, aplikované nielen na sledovanie ľudí.

## Poďakovanie

Chcel by som poďakovať môjmu školiteľovi prof. Ing. Martinovi Dražanskému Ph.D. a Ing. Tomášovi Dykovi za ich pomoc.

## Literatúra

- [1] Martin Dražanský, Ivo Macek, and Tomáš Goldmann. Monitoring of visitors in museum exhibitions. *Journal of the National Museum (Prague), Natural History Series*, 189, 2020.
- [2] Chongsheng Yu, Xin Li, Lei Dou, Jianwei Li, Yu Zhang, Jian Qin, Yuqing Sun, and Zhiyue Cao. Implement and optimization of indoor positioning system based on wi-fi signal. In *International Conference on Algorithms and Architectures for Parallel Processing*, pages 220–228. Springer, 2016.
- [3] Joel Lanir, Tsvi Kuflik, Julia Sheidin, Nisan Yavin, Kate Leiderman, and Michael Segal. Visualizing museum visitors' behavior: Where do they go and what do they do there? *Personal and Ubiquitous Computing*, 21(2):313–326, 2017.
- [4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [5] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. Ieee, 2017.
- [6] Juan Du. Understanding of object detection based on cnn family and yolo. In *Journal of Physics: Conference Series*, volume 1004, page 012029. IOP Publishing, 2018.
- [7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
- [8] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [9] OpenCV. Opencv documentation. online, 1 2022.
- [10] Apurva Bedagkar-Gala and Shishir K Shah. A survey of approaches and trends in person re-identification. *Image and vision computing*, 32(4):270–286, 2014.
- [11] Chunxiao Liu, Shaogang Gong, Chen Change Loy, and Xinggang Lin. Person re-identification: What features are important? In *European Conference on Computer Vision*, pages 391–401. Springer, 2012.
- [12] Chuting Zhu. Multi-camera people detection and tracking, 2019.
- [13] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.
- [14] Tatjana Chavdarova, Pierre Baqué, Stéphane Bouquet, Andrii Maksai, Cijo Jose, Timur Bagautdinov, Louis Lettry, Pascal Fua, Luc Van Gool, and François Fleuret. Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5030–5039, 2018.
- [15] Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. Motchallenge 2015: Towards a benchmark for multi-target tracking. *arXiv preprint arXiv:1504.01942*, 2015.
- [16] Rangachar Kasturi, Dmitry Goldgof, Padmanabhan Soundararajan, Vasant Manohar, John Garofolo, Rachel Bowers, Matthew Boonstra, Valentina Korzhova, and Jing Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern*

*Analysis and Machine Intelligence*, 31(2):319–336, 2008.

- [17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [18] Tatjana Chavdarova and François Fleuret. Deep multi-camera people detection. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 848–853. IEEE, 2017.
- [19] Yuanlu Xu, Xiaobai Liu, Yang Liu, and Song-Chun Zhu. Multi-view people tracking via hierarchical trajectory composition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4256–4265, 2016.
- [20] Pierre Baqué, François Fleuret, and Pascal Fua. Deep occlusion reasoning for multi-camera multi-target detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 271–279, 2017.
- [21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [23] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.