
Odhad a sledování polohy člověka pomocí jediné RGB kamery

Vedoucí: Ing. Adam Novozámský, Ph.D.

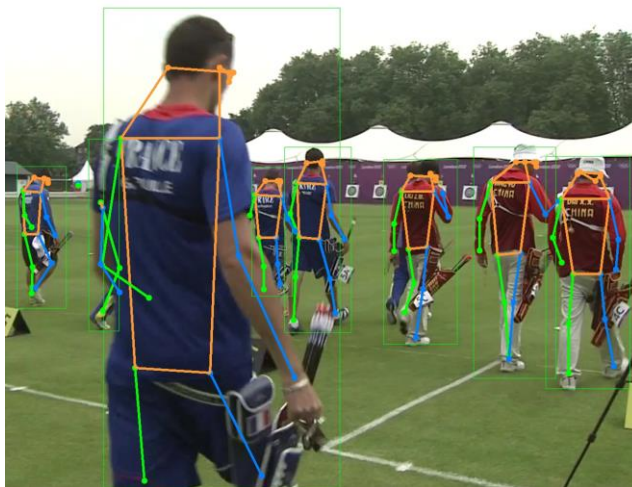
Vypracoval: Bc. Antonín Čech
FJFI ČVUT v Praze

Zadání práce

- Seznámit se s problematikou a vybrat několik metod
- Stáhnout několik volně dostupných anotovaných datasetů
- Vytvořit dataset bez anotací pro testování
- Vybrané metody porovnat na základě přesnosti a rychlosti
- Implementovat metodu pro sledování osob ve videu
- Vytvořit GUI pro vizualizaci jednotlivých metod detekce a sledování osob

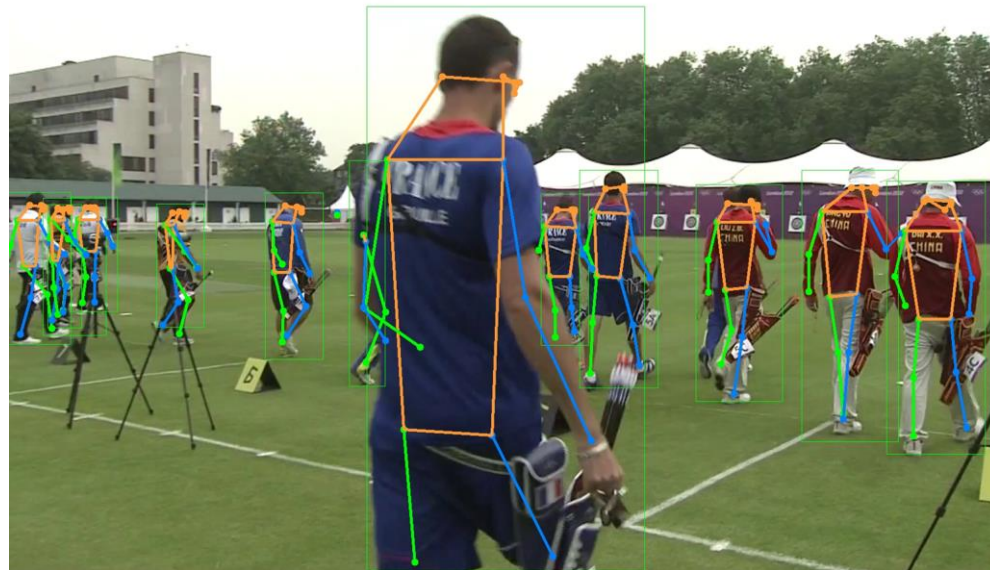
Odhad polohy člověka

- Hledání tzv. klíčových bodů
- Rozpoznání více osob na snímku
- 2D nebo 3D odhad
- 2 různé přístupy
 - Top-down
 - Bottom-up



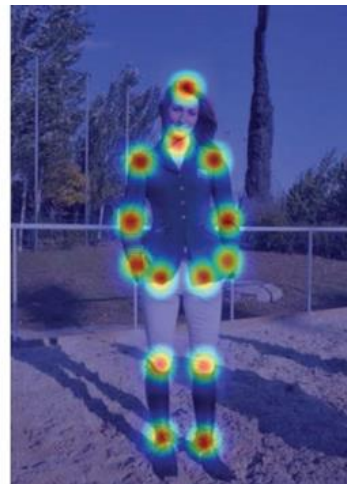
Top-down přístup

- Detekce osob na snímku - bounding box
- Odhad postavy zvlášť pro každý bounding box
- Lepší odhady pro vzdálené osoby
- Přesnost závisí na detektoru osob



Bottom-up přístup

- Vytvoření teplotní mapy odhadu klíčových bodů
- Spojení bodů do správných dvojic
- Odpadá závislost na detektoru
- Párování bodů je NP-těžká úloha



Převzato z [1]

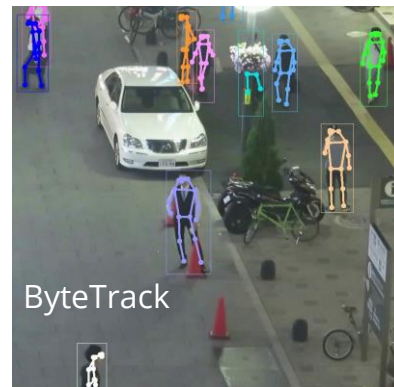
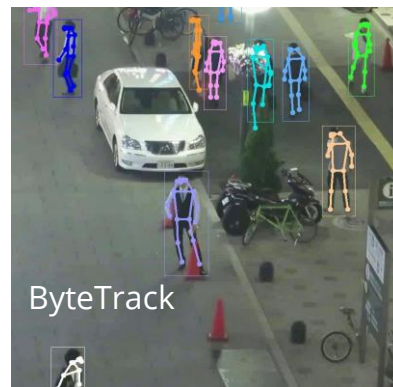
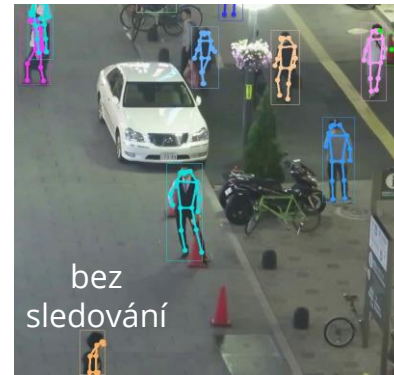
Sledování polohy člověka

- Zachování ID osoby mezi snímky ve videu
- zachování ID osoby při jejím opuštění zorného pole kamery
- Rozmazání pohybem nebo zkreslení obrazu

Snímek č. 1



Snímek č. 20



Metody odhadu polohy

- Implementace v Pythonu
- PyTorch a CUDA
- Top-down metody testovány s detektorem YOLOv3
- Většina metod implementována v balíčku MMPose

Metoda	Přístup
HRNet	Top-down
AlphaPose	Top-down
Hourglass	Top-down
MediaPipe*	Top-down
DEKR	Bottom-up
HigherHRNet	Bottom-up
LitePose	Bottom-up

* MediaPipe testována pouze na MSCOCO

Metody odhadu sledování polohy

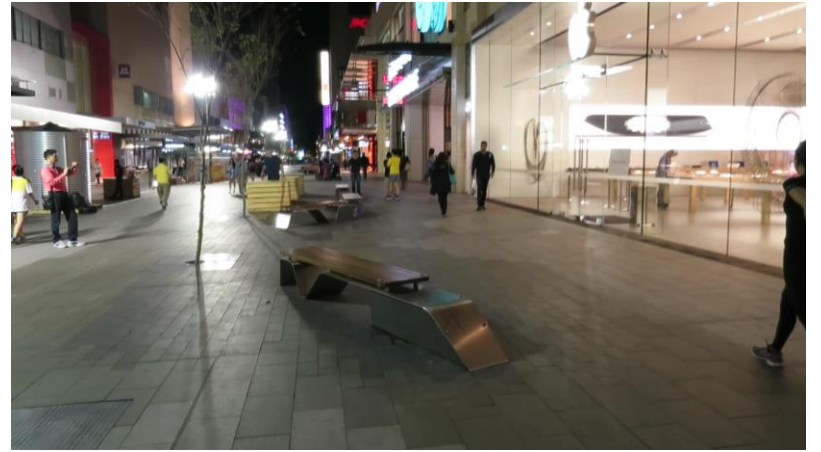
- Klíčové body zjištěné metodou HRNet
- Metriky pro greedy tracking
 - Intersection-over-union (IoU)
 - průměrná vzdálenost klíčových bodů
 - podobnost klíčových bodů objektu (OKS)
- Implementace z balíčku MMTracking

Metoda	Detektor osob
ByteTrack	YOLOX
OC-SORT	YOLOX
QDTrack	Faster-RCNN
Greedy tracking	Faster-RCNN

Vybrané datasey

- CrowdPose
 - Převážně snímky davů
 - 8000 snímků
- MS COCO
 - Dataset pro všechny varianty počítačového vidění
 - 5000 snímků
- MOT17
 - Videá s různým množstvím osob
 - 7 videí
- MOT20
 - Záběry extrémních davů lidí
 - 4 videa

MOT17



MOT20



Výpočet přesnosti odhadu polohy

- Average Precision (AP)
- Object Keypoint Similarity (OKS)

$$\text{OKS} = \frac{\sum_i \exp\left(\frac{-d_i}{2s^2k_i^2}\right) \cdot \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$

$$\text{AP}^{\text{OKS}_{\text{thr}}} = \text{mean}(X_i \geq \text{OKS}_{\text{thr}})$$

- Mírná metrika $\text{AP}^{.50} = \text{mean}(X_i \geq 0.50)$
- Striktní metrika $\text{AP}^{.75} = \text{mean}(X_i \geq 0.75)$

$$\text{AP} = \text{mean}(\text{AP}^{\text{OKS}_{\text{thr}}}), \text{OKS}_{\text{thr}} = 0.50, 0.55, \dots, 0.90, 0.95$$

Výsledky odhadu polohy

Dataset	Metoda	AP	AP ^{.50}	AP ^{.75}	Rychlost [it/s]
MS COCO	HRNet	0,666	0,796	0,729	16,10
	AlphaPose	0,722	0,886	0,799	16,28
	Hourglass	0,648	0,795	0,708	14,83
	MediaPipe	0,452	0,777	0,482	14,98
	DEKR	0,709	0,876	0,773	4,14
	HigherHRNet	0,672	0,864	0,731	2,07
	LitePose	0,624	0,825	0,679	10,58
CrowdPose	HRNet	0,651	0,795	0,701	6,24
	AlphaPose	0,695	0,853	0,749	11,79
	Hourglass	0,464	0,671	0,492	7,52
	DEKR	0,682	0,869	0,736	4,24
	HigherHRNet	0,653	0,860	0,700	1,21
	LitePose	0,605	0,831	0,645	8,93

Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Výsledky odhadu polohy

Dataset	Metoda	AP	AP ^{.50}	AP ^{.75}	Rychlost [it/s]
MS COCO	HRNet	0,666	0,796	0,729	16,10
	AlphaPose	0,722	0,886	0,799	16,28
	Hourglass	0,648	0,795	0,708	14,83
	MediaPipe	0,452	0,777	0,482	14,98
	DEKR	0,709	0,876	0,773	4,14
	HigherHRNet	0,672	0,864	0,731	2,07
	LitePose	0,624	0,825	0,679	10,58
CrowdPose	HRNet	0,651	0,795	0,701	6,24
	AlphaPose	0,695	0,853	0,749	11,79
	Hourglass	0,464	0,671	0,492	7,52
	DEKR	0,682	0,869	0,736	4,24
	HigherHRNet	0,653	0,860	0,700	1,21
	LitePose	0,605	0,831	0,645	8,93

Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Výsledky odhadu polohy

Dataset	Metoda	AP	AP ^{.50}	AP ^{.75}	Rychlost [it/s]
MS COCO	HRNet	0,666	0,796	0,729	16,10
	AlphaPose	0,722	0,886	0,799	16,28
	Hourglass	0,648	0,795	0,708	14,83
	MediaPipe	0,452	0,777	0,482	14,98
	DEKR	0,709	0,876	0,773	4,14
	HigherHRNet	0,672	0,864	0,731	2,07
	LitePose	0,624	0,825	0,679	10,58
CrowdPose	HRNet	0,651	0,795	0,701	6,24
	AlphaPose	0,695	0,853	0,749	11,79
	Hourglass	0,464	0,671	0,492	7,52
	DEKR	0,682	0,869	0,736	4,24
	HigherHRNet	0,653	0,860	0,700	1,21
	LitePose	0,605	0,831	0,645	8,93

Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Výpočet přesnosti sledování polohy

- Multiple object tracking accuracy

$$\text{MOTA} = \frac{|\text{FN}| + |\text{FP}| + |\text{IDSW}|}{|\text{grDet}|}$$

- Higher order tracking accuracy (HOTA)

$$\text{HOTA}_\alpha = \sqrt{\frac{\sum_{c \in \{\text{TP}\}} \mathcal{A}(c)}{|\text{TP}| + |\text{FN}| + |\text{FP}|}}$$

$$\mathcal{A}(c) = \frac{|\text{TPA}(c)|}{|\text{TPA}(c)| + |\text{FNA}(c)| + |\text{FPA}(c)|}$$

$$\text{HOTA} = \int_0^1 \text{HOTA}_\alpha \, d\alpha$$

Výsledky sledování polohy

Dataset	Metoda	HOTA	MOTA	Rychlost [fps]
MOT17	ByteTrack	72,220	85,925	2,89
	OC-SORT	72,609	85,247	2,81
	QDTrack	63,912	78,500	3,84
	Greedy tracking - IoU	37,424	39,214	3,76
	Greedy tracking - distance	31,017	36,469	3,69
	Greedy tracking - OKS	33,976	37,803	3,87
MOT20	ByteTrack	53,374	68,722	1,15
	OC-SORT	53,981	69,251	1,03
	QDTrack	29,738	43,512	1,60
	Greedy tracking - IoU	17,167	26,377	2,00
	Greedy tracking - distance	12,800	24,241	1,78
	Greedy tracking - OKS	15,785	25,933	1,95

Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Výsledky sledování polohy

Dataset	Metoda	HOTA	MOTA	Rychlost [fps]
MOT17	ByteTrack	72,220	85,925	2,89
	OC-SORT	72,609	85,247	2,81
	QDTrack	63,912	78,500	3,84
	Greedy tracking - IoU	37,424	39,214	3,76
	Greedy tracking - distance	31,017	36,469	3,69
	Greedy tracking - OKS	33,976	37,803	3,87
MOT20	ByteTrack	53,374	68,722	1,15
	OC-SORT	53,981	69,251	1,03
	QDTrack	29,738	43,512	1,60
	Greedy tracking - IoU	17,167	26,377	2,00
	Greedy tracking - distance	12,800	24,241	1,78
	Greedy tracking - OKS	15,785	25,933	1,95

Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Výsledky sledování polohy

Dataset	Metoda	HOTA	MOTA	Rychlost [fps]
MOT17	ByteTrack	72,220	85,925	2,89
	OC-SORT	72,609	85,247	2,81
	QDTrack	63,912	78,500	3,84
	Greedy tracking - IoU	37,424	39,214	3,76
	Greedy tracking - distance	31,017	36,469	3,69
	Greedy tracking - OKS	33,976	37,803	3,87
MOT20	ByteTrack	53,374	68,722	1,15
	OC-SORT	53,981	69,251	1,03
	QDTrack	29,738	43,512	1,60
	Greedy tracking - IoU	17,167	26,377	2,00
	Greedy tracking - distance	12,800	24,241	1,78
	Greedy tracking - OKS	15,785	25,933	1,95

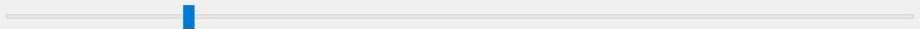
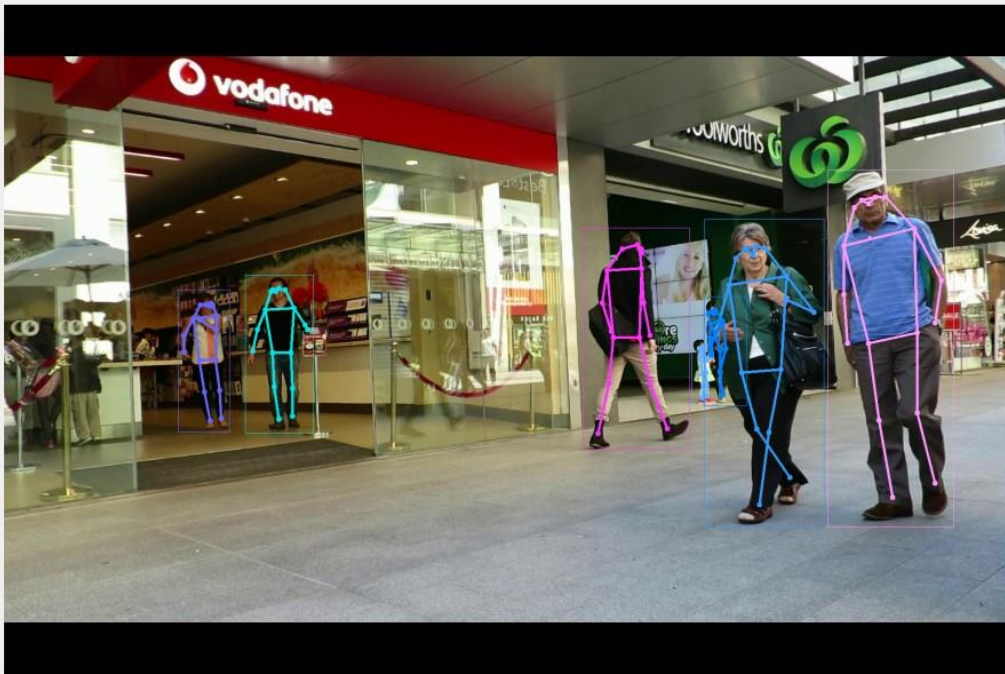
Testováno na systému s AMD Ryzen 5 3600, 16 GB RAM a NVIDIA RTX 2080

Aplikace

- Provedení odhadu pozic
 - Vizualizace už vypočtených pozic
 - Metody založené na MMPose
-
- Python 3.10
 - PyQt5
 - CUDA
 - github.com/cechantonin/PoseEstimationApp

Human pose estimation and tracking

File About



105 / 525

Run inference

Visualize results

Pose estimation method: HRNet

Enable pose tracking

Pose estimation method: ByteTrack

Bounding box threshold for pose estimation



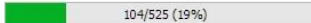
Keypoint threshold for pose tracking



Keypoint visualization radius: 5

Pose visualization thickness: 3

Show inference in real time




Run Inference

Stop Inference

Human pose estimation and tracking

File About



Run inference Visualize results

Original video
Inference result
MOT17-09-SDP

Result information

Pose method: HRNet
Tracking method: Tracking disabled
Pose bounding box threshold: 0.3
Tracking keypoint threshold: 0.3

Keypoint visualization radius: 5

Pose visualization thickness: 3

Add JSON file Remove file

Save JSON result Save video

235 / 525

Děkuji za pozornost

Reference

- [1] Dang, Q., Yin, J., Wang, B., Zheng, W.: Deep learning based 2D human pose estimation: A survey. In: *Tsinghua Science and Technology*, vol. 24 (6), 2019, pp. 663-676.