LATVIJAS UNIVERSITĀTE
DATORIKAS FAKULTĀTE

SKELETA POZU PROGNOZĒŠANAS MODEĻU SISTEMĀTISKĀ LITERATŪRAS ANALĪZE

KURSA DARBS

Autors: **Pauls Purviņš**
Studentu apliecības Nr.: pp19026
Darba vadītājs: Phd. Comp. Sc. Ēvalds Urtāns

RĪGA, 2023

ANOTĀCIJA

Darbs ir sistemātiska zinātniskās literatūras analīze, kuras ietvaros tiek apskatīts atslēgpunktu meklēšanas uzdevums, kas saukts arī par pozas atpazīšanas uzdevumu, un sistemātiskas literatūras analīzes formā tiek apskatīti 17 dažādi darbi, kas pēcāk tiek salīdzināti gan pēc darbos iekļautajām metrikām, gan darba laikā veikta praktiska eksperimenta rezultātiem, lai noskaidrotu, ka darbs "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation"[23] ir atzīstams par labākajiem apskatīto darbu vidū un prezentē konkurētspējīgu risinājumu kan rezultātu precizitātē, gan veiktspējā, vienlaikus parādot, ka šajā jomā ir izaugsmes vieta jaunām modeļu arhitektūrām un idejām.

Darba pamattekstā ir 29. lappuses.

Atslēgvārdi: cilvēku pozu noteikšana, sistemātiska literatūras analīze, atslēgpunktu noteikšana, mākslīgais intelekts

ABSTRACT

In this work, the author takes a look at the human keypoint recognition task also called the pose estimation task, and in a systematic literature analysis format analyses 17 different publications. These publications are compared based on various metrics both from original papers and from practical experiments conducted in the process of creating this work, to conclude that the current SOTA (state of the art) is "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation"[23] which achieves amazing results both in accuracy and processing speed at the same showing that there still is room to grow in this field, especially by utilizing transformer architecture based models.

The main text of the semester work consists of 29. pages

KEYWORDS: human pose estimation, systematic literature analysis, keypoint recognition, artificial intelligence

Saturs

Apzīmējumu saraksts

AI (Artificial intelligence) Mākslīgais intelekts

API (Application Programming Interface) Interfeiss kas ļauj dažādām lietojumprogrammām apmainīties ar informāciju

Atslēgpunkts (Keypoint) Kāda nozīmīga iezīme objektā, cilvēku gadījumā - locītavas

SOTA (State Of The Art) Modernākais

1 Ievads

Pozu noteikšana un dažādu atslēgpunktu atrašana attēlos ir bijusi aktuāla problēma jau gadiem un pasaulei arvien vairāk dažādas sistēmas paļaujas uz datiem par lietotājiem, no sejas atpazīšanas telefona ekrānā, līdz cilvēku skaitīšanai veikalos, un no dažādu attēlu un video klasificēšanas pēc tajās veiktajām aktivitātēm līdz zīmju valodas lasītājiem. To parāda arī 34700 zinātniskas publikācijas pēdējā gada laikā un 297000 publikācijas pēdējo 10 gadu laikā ar atslēgas vārdiem "Pose estimation". Ik gadu šis skaits pieaug (2023. gadā jau ir vairāk kā 800 publikāciju, 2022. gadā bija 33800 publikācijas un 2021. gadā bija 34600 publikācijas šajā jomā).

Sekojošais darbs ir sistemātiska literatūras analīze, kuras ietvaros autors aplūko dažādus modernākos un vēsturiski nozīmīgākos risinājumus cilvēku atslēgpunktu atrašanai. Darba ietvaros veiktā sistemātiskā literatūras analīze sniedz ieskatu nozarē un konkrētajā problēmā, kā arī tiek veikts eksperiments, lai salīdzinātu modeļu praktisko veiktspēju nosakot pēc vērtēšanas kritērijiem labāko modeli pār vairākām datu kopām.

2 Pielietojumi

Pielietojumi ir dažādi un tie ietver:

- Personīgos trenerus
  - Ai Fitness lietotne
  - FitYoga lietotne
  - Insane AI lietotne
  - infiGro lietotne
- Paplašinātā realitāte
  - Wii sports
  - Kinect Sports
  - Metaverse tēli
- Kustību ierakstīšana
  - Optitrack
  - Plask

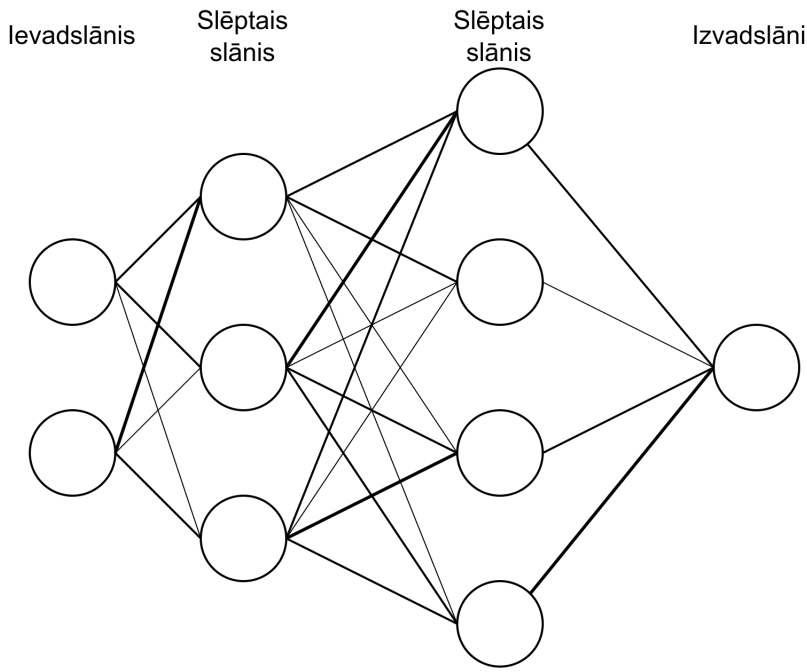
Google piedāvā arī API ar nosaukumu Blaze pose (`https://google.github.io/mediapipe/solutions/pose.html`), kas ļauj mobilajās ierīcēs viegli izstrādāt lietojumprogrammas, kas izmanto atslēgpunktu atpazīšanu. BlazePose izmantoti arī speciāli mobilajām ierīcēm paredzēti modeļi, kas darbā netiek apskatīti, jo neiekļaujas izvirzītajos kritērijos.

3 Mākslīgie neironu tīkli

Mākslīgie neironu tīkli (saukti arī par neironu tīkliem) ir mašīnmācīšanās pieeja, kas tapa jau 20.gs vidū, taču gan matemātisku, gan tehnisku iemeslu dēļ tie nekļuva populāri līdz šī gadsimta sākumam, kad tie sāka uzvarēt dažādās ar IT saistītās sacensībās. Kopš tā laika neironu tīkli ir kļuvuši par neatņemamu sastāvdaļu tādās jomās kā mašīntulkošanā, attēlu un audio apstrādē, klasificēšanā un sintēzē, autonomu robotu un mašīnu izstrādē un citās. [12]

Perceptrons ir pirmais neironu tīkls, ko 1954. gadā izstrādāja Franks Rosenblats, līdzīgi kā neirons, saņem dažādus ievaddatus, un balstities uz tiem dod vai nedod signālu tālāk. Perceptrons nesaturēja aktivācijas funkciju, tādēļ nespēja strādāt ar nelineāriem datiem un risināja relatīvi vienkāršas regresijas problēmas. [21]

Mūsdienās neironu tīkli ir attīstījušies tālāk un, ja perceptrons sastāvēja no viena neironu slāņa, tad tagad tie bieži sastāv pat no simtiem šādu slāņu un tiek saukti par dziļajiem neironu tīkliem (par dziļajiem neironu tīkliem sauc tādus neironu tīklus, kuros ir vairāk kā viens slēptais slānis).[12]



1. att. Neironu tīkla modelis

Pēc programminženierijas teorijas, neironu tīkli ir uzskatāmi par melno kasti, jo visbiežāk mums nav zināšanu kādēļ tas pieņem tādu vai citādu lēmumu. [12]

Lai neironu tīkli spētu iemācīties nelineārus datus, tiek izmantotas aktivācijas funkcijas, kas tiek pielietotas pēc neirona iekšējo aprēķinu veikšanas. Bez tām modelis, lai cik slāņu tajā arī nebūtu, nespētu iemācīties nelineārus datus. Nervu neironu kontekstā

šī funkcija pasaka, vai, balstoties uz ievaddatiem, neironam ir vai nav jāizvada simbolu. Mākslīgo neironu tīklos gan šīs funkcijas parasti nav bināras, bet gan dod dažādas vērtības bieži vien robežās no -1 līdz 1 vai no 0 līdz 1 un tipiski atrodas starp visiem slāņiem tīklā.[11]

Dažas populārākās aktivācijas funkcijas [21] ir:

1. Sigmoīds
   Dod vērtības no 0 līdz 1

$$y = \frac{1}{1 + e^{-x}} \tag{1}$$

2. ReLU
   Dod vērtības no 0 līdz ∞

$$y = max(0,x) \tag{2}$$

3. Tanh
   Dod vērtības no -1 līdz 1

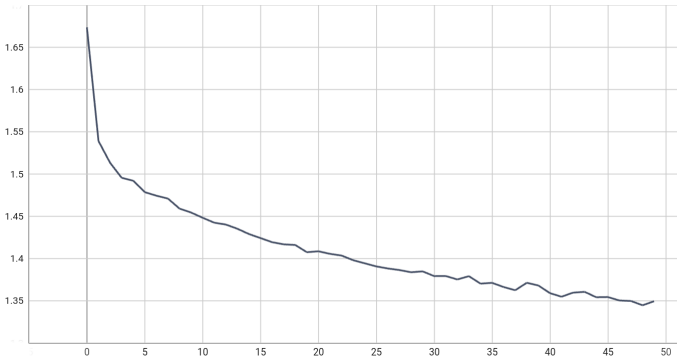
$$y = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{3}$$

4. ELU
   Vienīgā no šeit minētajām funkcijām, kam ir apmācāms parametrs. Tas nozīmē ka šī funkcija nav statiska, bet tās eksponenta daļa (kad x $\leq$ 0) satur apmācāmu parametru α.

$$y = \begin{cases} \alpha(e^{x} - 1) & \text{, ja } x \leq 0 \\ x & \text{, ja } x > 0 \end{cases} \tag{4}$$

4 Kļūdas funkcijas

Kļūdas funkcijas ir dažādas, taču tām visām ir viens mērķis - noteikt, cik ļoti modelis kļūdās. Katram modelim un katrai apmācības reizei būs atšķirīga kļūdas funkcijas vērtība dažādos soļos, taču tās visas vairāk vai mazāk līdzināsies 2. attēlā parādītajai līknei un tieksies uz 0.



2. att. Kļūdas funkcijas vērtība atkarībā no apmācības iterācijas

Ja funkcija neatbilst šim un aug, tad, visticamāk, modelis vai apti funkcija ir izveidota nepareizi un modelis nekad nekonverģēs.
Dažas populārākās kļūdas funkcijas ir:

- **MSE (Mean Squared Error)** Latviski pazīstama kā dispersija, šī ir viena no visbiežāk lietotajām kļūdas funkcijām statistikā un regresijas uzdevumos. To iegūst, izvelkot kvadrātsakni no patiesās un paredzētās vērtības starpības kvadrāta.[21]

$$L_i = \frac{1}{J}\sum_{j}^{J}(y_{i,j} - \hat{y}_{i,j})^2 \tag{5}$$

- **CCE (Categorial cross entropy)** Kategoriskā krust entropija ir kļūdas funkcija, ko bieži izmanto, kad modelim ir jānosaka ievaddatu piederība kādai no klasēm. [21]

$$L_i = -\sum_{j} y_{i,j} * log(\hat{y}_{i,j}) \tag{6}$$

- **BCE (Binary cross entropy)** Binārā krust entropija ir kļūdas funkcija, tiek izmantota līdzīgos gadījumos kā CCE, taču ir paredzēta gadījumiem, kad ir tikai divas klases, kas aprakstītas ar viena mainīgā bināru stāvokli, kur 0 ir pirmā klase un 1 ir otrā klase. [21]

$$L_i = -\sum_{j} y * log(\hat{y}_{i,j}) + (1-y) * log(1-\hat{y}_{i,j}) \tag{7}$$

5 Atpakaļizplatīšanās algoritms

Neironu tīklos informācija plūst no ievad-neironiem, cauri slēptajiem slāņiem un visbeidzot izvad-slānim, lai beigās izdotu izvaddatus $\hat{y}$. Apmācot tīklus, no $\hat{y}$ un y parasti tiek iegūta zaudējumfunkcijas vērtība $J(\Theta)$. [11]

Atpakaļizplatīšanās (Back propogation) algoritms [20] ļauj šo informāciju no zaudējum-vērtības padot atpakaļ caur tīklu, lai aprēķinātu gradientu, kuram sekojot var samazināt kļūdu. Bieži cilvēki runājot par atpakaļizplatīšanās saprot visa modeļa svaru atjaunināšanu ar mērķi uzlabot izvaddatu precizitāti, taču atpakaļizplatīšanās ir tikai veids kā aprēķināt parciālos atvasinājumus, svaru atjaunošanu atstājot citu algoritmu rokās. [11]

Algoritma pamatā ir atvasināšana un ķēdes likums.
Ja $y = g(x)$ un $z = f(g(x)) = f(y)$, tad pēc ķēdes likuma $\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$
To varam vispārināt ja $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, g attēlo $\mathbb{R}^M$ uz $\mathbb{R}^n$ un f attēlo $\mathbb{R}^n$ uz $\mathbb{R}$. Ja $y = g(x)$ un $z = f(y)$, tad $\frac{\delta z}{\delta x_i} = \sum_j \frac{\delta z}{\delta y_j}\frac{\delta y_j}{\delta x_i}$ [11]

Piemērs atpakaļizplatīšanās algoritmam dots zemāk. Vienādojumos izmantotie apzīmējumi:

- x_n - ievades īpašības
- w_n - apmācāmie svari
- σ - aktivizācijas funkcija
- $\hat{y}$ - prognozētā vērtība
- $\mathcal{L}$ - vidējās kvadrātiskās kļūdas funkcija
- α - apmācību solis

$$z = \sum_{i=1}^{n} w_n x_n + b = w_1 x_1 + w_2 x_2 + b \tag{8}$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{9}$$

Modelis:

$$\hat{y} = \sigma(\sum_{i=1}^{n} w_n x_n) = \frac{1}{1 + e^{-(w_1 x_1 + w_2 x_2 + b)}} \tag{10}$$

Kļūdas funkcija:

$$\mathcal{L} = MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2 = \frac{1}{2}(\hat{y} - y)^2 \tag{11}$$

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_1} = \frac{\partial \mathcal{L}(\hat{y},y)}{\partial(\hat{y} - y)} \cdot \frac{\partial \sigma}{\partial z} \cdot \frac{\partial z}{w_1} \tag{12}$$

Kļūdas funkcijas atvasinājums:

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial(\hat{y}-y)} = 2 \cdot (\hat{y}-y) \cdot \frac{1}{2} = (\hat{y}-y) \tag{13}$$

Sigmoīda atvasinājums:

$$\frac{d\sigma}{dz} = \sigma(z) \cdot (1-\sigma(z)) \tag{14}$$

$$\frac{d(w_1x_1 + w_2x_2 + b)}{dw_1} = x_1 \tag{15}$$

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_1} = (\hat{y}-y) \cdot \sigma(z) \cdot (1-\sigma(z)) \cdot x_1 \tag{16}$$

Apmācāmo svaru atvasinājumi:

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_1} = 2(\hat{y}-y) \cdot \sigma(z) \cdot (1-\sigma(z)) \cdot x_1 \tag{17}$$

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_2} = 2(\hat{y}-y) \cdot \sigma(z) \cdot (1-\sigma(z)) \cdot x_2 \tag{18}$$

$$\frac{\partial \mathcal{L}(\hat{y},y)}{\partial b} = 2(\hat{y}-y) \cdot \sigma(z) \cdot (1-\sigma(z)) \tag{19}$$

Stohastiskā kalnā kāpēja vienas iterācijas vienādojumi:

$$\hat{w_1} = w_1 - \alpha \cdot \frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_1} \tag{20}$$

$$\hat{w_2} = w_2 - \alpha \cdot \frac{\partial \mathcal{L}(\hat{y},y)}{\partial w_2} \tag{21}$$

$$\hat{b} = b - \alpha \cdot \frac{\partial \mathcal{L}(\hat{y},y)}{\partial b} \tag{22}$$

Apmācot neironu tīklus, priekšizplatīšanās un atpakaļizplatīšanās ir atkarīgas viena no otras, tas ir, sākumā informācija caur tīklu izplūst normālā virzienā un aprēķina visus mainīgos tās ceļā. Šie mainīgie, pēc tam, tiek lietoti atpakaļizplatīšanā, kad aprēķinu grafs tiek apvērsts. Apmācot neironu tīklus, mēs secīgi veicam paredzēšanu un atpakaļizplatīšanos, atjaunojot modeļa parametrus, izmantojot gradientus, ko dod atpakaļizplatīšanās. Šīs darbības tiek veiktas pamīšus, lai atpakaļizplatīšanās varētu izmantot jau vienreiz aprēķinātās vērtības un samazinātu aprēķinu daudzumu. Lai arī apmācāmo svaru atkalizmantošana samazina nepieciešamo aprēķinu daudzumu, tā palielina nepieciešamo atmiņas daudzumu, jo visas starpvērtības ir jāsaglabā atmiņā.

6 Apmācāmo parametru optimizācija

Balstoties uz atpakaļizplatīšanā aprēķinātajiem gradientiem, neironu tīkla apmācības procesā ir jāatjaunina visus tīkla parametrus.

Viens no veidiem, kā to darīt, ir pieskaitīt katram parametram tā aprēķināto gradientu, taču tas viegli noved pie lielām izmaiņām modeļa izvaddatos, tādēļ gradients pirms pieskaitīšanas tiek sareizināts ar mācīšanās koeficientu (learning rate), kas ir mazāks par 1, tādējādi palēninot mācību procesu, bet atļaujot mācīšanos veikt granulārāk pa mazākie soļiem. [25]

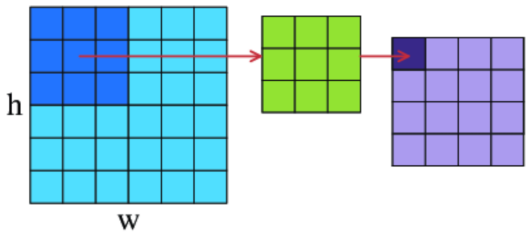
Vel viens iemesls veikt mācīšanos pa mazākiem soļiem ir operatīvās atmiņas apjoms. Bieži vien modeļu apmācību veic, sadalot datus apakšgrupās un tās dodot modelim secīgi nevis visas vienlaikus. Lai arī šī metode samazina nepieciešamo atmiņas daudzumu, tās lietošana nozīmē, ka aprēķinātie gradienti neraksturo visu datu kopu, bet gan tikai daļu no tās. Veicot mazākus soļus, modelis ir spējīgs apgūt informāciju no katras apakškopas, netaisot nejaušus "lēcienus" visos virzienos atkarībā no apakškopā iekļautajiem datiem. [25]

Pastāv arī dažādi optimizātori, kas uzlabo modeļa spēju apgūt problēmu, izmantojot dažādas metodes, piemēram, ņemot vērā ne tikai pēdējā atpakaļizplatīšanā aprēķinātos gradientus, bet arī senākus. Šādu optimizātoru starpā ir tādi kā "SGD" (Stochastic gradient descent, jeb stohastiskais kalnā kāpēja algoritms) un "Adam". [11]

7 Konvolūciju tīkli

Lai arī neironu tīkli ir ļoti spēcīgi, to lielākais spēks neironi ir arī to lielākā vājība, it īpaši veiktspējas ziņā attēlu un citu daudz-dimensionālu datu apstrādē. Neironu tīklus ir grūti pielietot daudzdimensionāliem datiem, jo to izmērs palielinās ļoti strauji Piemēram, lai tīklam ievaddatos iedotu 20x20 pikseļu attēlu, ir nepieciešami 400 neironi. Ja attēls ir krāsains, tad neironu skaits pieaug līdz 1200. un mūsu datoru veiktspēja nepalielinās lineāri ar mākslīgā intelekta jomas attīstību. [11]

Tāpēc attēlu apstrādē tiek pielietoti konvolūciju tīkli. Tie ļauj ievērojami samazināt parametru skaitu, nezaudējot neironu tīklu spēju apgūt dažādas datu iezīmes un ļauj modeļa spēju iemācīties iezīmes hierarhiski. Tie strādā kā slīdošais logs pāri ievaddatiem (skat. 3. attēlu), tādēļ vieni un tie paši parametri tiek pielietoti dažādām attēla vietām un nav nepieciešami atsevišķi neironi katrai no tām.



3. att. Attēls, konvolūcijas kodols un izvadattēls [9]

Konvolūcijas var pielietot ne tikai daudzdimensionāliem datiem, bet arī viendimensionāliem datiem, piemēram, laika sērijām. [11]

Divdimensiju konvolūciju tīklu apraksta vienādojums:

$$s(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) = \sum_m \sum_n I(i-m, j-n)K(m,n) \tag{23}$$

8 Metrikas

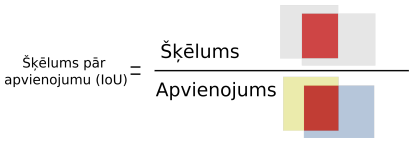
Katram modelim viena no metrikām ir jau apskatītā kļūdas funkcijas vērtība, taču bieži vien ar to nepietiek, lai novērtētu modeļa veiktspēju. Kļūdas funkcija, lai arī nozīmīga, nepasaka tieši cik piemēros modelis kļūdās, vai cik precīzi paredzētais rezultāts atbilst sagaidītajam. Tāpēc tiek izmantotas citas metrikas, kas palīdz precīzāk aprakstīt modeļa veikumu, bet nekādi neietekmē pašu apmācības procesu.

8.1 PCKh

Pareizo atslēgpunktu procents (PCK) relatīvs pret 50% galvas platuma (PCKh) ir metrika, kas kopā ar MPII datu kopu tika izstrādāta 2014. gadā. Tā mēra, cik procentu no punktiem ir paredzēti pareizi, ņemot pusi no galvas platuma pikseļos kā pieļaujamo kļūdas apgabalu, tādējādi panākot, ka šī metrika ir godīga un pielietojama neatkarīgi no attēla izmēra. [1]

8.2 IoU

Šķēlums pār apvienojumu (Intersection over Union), saukta arī par Jakarda indeksu, ir metrika, kas apraksta, cik liela daļa no paredzētā laukuma pārklājas ar pareizo atbildi. Kā jau nosaukums pasaka, šī metrika tiek aprēķināta, ņemot paredzētās un pareizās atbildes šķēlumu un to dalot ar to apvienojumu. [21]



4. att. Šķēluma pār apvienojumu formula ar ilustrāciju

8.3 mAP

Vidējo precizitāšu vidējā vērtība (Mean Average Precision) ir metrika, ko izmanto MPII datu kopā, lai mērītu no apakšas uz augšu ejošo (skat 9. nodaļu) modeļu precizitāti. Tā mēra, kāds ir vidējais IoU personas atslēgpunktiem.

$$mAP = \frac{1}{N}\sum_{i}^{N}\sum_{j\in\{0.5,0.55,\ldots,0.95\}} IoU_j(y_i, \hat{y}_i) \tag{24}$$

y_i apzīmē pareizo locītavas apgabalu
$\hat{y}_i$ apzīmē paredzēto locītavas apgabalu
IoU_j apzīmē IoU funkcijas vērtību kur modeļa pārliecība ir $\geq j$

8.4 OKS

Objektu atslēgpunktu līdzība (Object Keypoint Similarity) ir metrika, ko izmanto COCO [15] datu kopa, kā arī vairākas citas datu kopas. Tā novērtē paredzēto atslēgpunktu atbilstību to patiesajām lokācijām, kur 0 nozīmē, ka katrs punkts ir vairāku standartnoviržu attālumā un 1 nozīmē, ka visi punkti ir precīzi paredzēti. Tās princips ir līdzīgs IoU, taču labāk pielāgots tieši atslēgpunktu meklēšanas uzdevumam

$$OKS = \sum_{i}[exp(-d_i^2/2s^2k_i^2)\delta(v_i > 0)]/\sum_{i}[\delta(v_i > 0)] \tag{25}$$

d_i ir Eiklīda distance starp patieso punkta lokāciju un modeļa paredzēto
v_i ir punktu redzamības atzīmes, kur 0 nozīmē, ka punkts netika atzīmēts, 1 - punkts nav redzams, 2 - punkts ir redzams
sk_i standartnovirze
s objekta izmērs
k_i specifiska konstante, kas piemeklēta katram atslēgpunktam [15]

8.5 AP

Vidējā precizitāte (Average Precision) ir metrika, kas skeleta pozu prognozēšanas uzdevumos balstās uz OKS. (Šajā darbā AP atsaucas uz šo COCO implementāciju vidējās

precizitātes metrikai, nevis uz vispārpieņemto. *mAP* balstās uz vispārpieņemto kas izmanto IoU nevis OKS pamatā)
AP tiek rēķināts, kā vidējā precizitāte pie dažādiem OKS sliekšņiem (sākot ar 0.5 līdz 0.95 ar 0.05 sliekšņa intervāliem) un tiek izmantots gan top-down, gan bottom-up modeļu novērtēšanā [15]

9 Metodoloģija

Personu pozu noteikšanas modeļi ķermeņa daļu noteikšanas problēmu risina divos veidos. "No augšas uz leju" (top-down) pieejā vispirms tiek lietots modelis, kas nosaka cilvēku robežas, un tad pozas modelis nosaka dažādos atslēgpunktus vienam cilvēkam.
"No lejas uz augšu" (bottom-up) pieejā pozas modelis nosaka atslēgpunktus visiem attēlā esošajiem cilvēkiem, un pēc tam ar dažādām metodēm sadala atrastos atslēgpunktus pa personām. Viens no veidiem, kā sadalīt atslēgpunktus pa cilvēkiem, ir izmantojot ķermeņa daļu piederības laukus (part affinity fields), kuros modelis paredz, kuri atslēgpunkti ir savā starpā saistīti, un apraksta vienas ķermeņa daļas abus galus.[6]

9.1 Atslēgpunkti

Atslēgpunkti (Keypoints) šī darba kontekstā apzīmē dažādas ķermeņa daļas, ko satur datu kopas. Katrā no tām ir iekļauts dažāds atslēgpunktu skaits, kas detalizētāk aprakstīts pie datu kopu salīdzinājuma (**??**.)

5. att. Piemērs no COCO datu kopas [15]

10 Datu kopas

Nozīmīgākās COCO un MPII datu kopu īpašības apskatītas 1. tabulā un tajās iekļautie atslēgpunkti attēloti 2.. tabulā.

1. tabula. MPII un COCO datu kopu salīdzinājums

	COCO	MPII
Izstrādes gads	2015	2014
Attēlu skaits	200'000	25'000
Personu skaits	250'000	40'000
Atslēgpunktu skaits katram cilvēkam	17	16

2. tabula. COCO un MPII atslēgpunktu indeksi

	COCO	MPII
Deguns	0	
Kreisā acs	1	
Labā acs	2	
Kreisā auss	3	
Labā auss	4	
Kreisais plecs	5	13
Labais plecs	6	12
Kreisais elkonis	7	14
Labais elkonis	8	11
Kreisā plaukstas locītava	9	15
Labā plaukstas locītava	10	10
Kreisais gurns	11	3
Labais gurns	12	2
Kreisais celis	13	4
Labais celis	14	1
Kreisā potīte	15	5
Labā potīte	16	0
Iegurnis		6
Krūškurvis		7
Kakls		8
Galvas augša		9

10.1 COCO

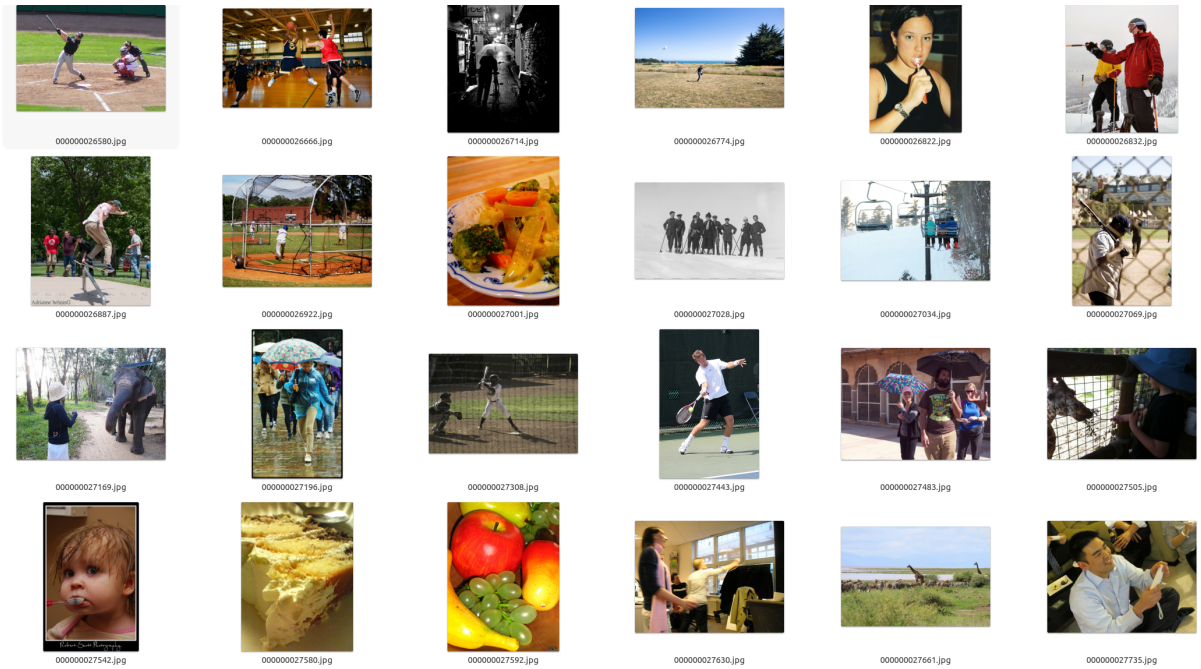
Microsoft COCO [14] datu kopa ir lielapjoma objektu noteikšanas un aprakstīšanas datu kopa. Tā paredzēta dažādu ar attēliem saistītu mašīnmācīšanās uzdevumu veikšanai, to starpā segmentācijas, aprakstu ģenerēšanas un cilvēku atslēgpunktu noteikšanā. Tā satur vairāk kā 200 tūkstošus attēlu un 250 tūkstošus personu, kam ir atzīmēti atslēgpunkti.

10.2 MPII

MPII[1] ir datu kopa, kas paredzēta tieši cilvēku pozu, atslēgpunktu un darbību noteikšanas uzdevumiem. Iekļautas ir tādas aktivitāšu klases, kā, piemēram, medīšana un zvejošana, sports, dejošana, staigāšana, ūdens aktivitātes. Tā ir ievērojami mazāka kā COCO datu kopa un satur 25 tūkstošus attēlu ar 40 tūkstošiem personu.

10.3 Validācijas kopa

Darba ietvaros tika atlasīti attēli un izveidota jauna datu kopa kas darba ietvaros saukta par Validācijas kopu. Tā satur 4560 attēlus no COCO test2017 kopas un paredzēta modeļu testēšanai šī darba ietvaros. Tā atšķiras no COCO val2017 kopas ar cilvēku attēlu īpatsvaru pret attēliem, kuros nav cilvēku un labāk paredzēta pozu noteikšanas modeļu ātrdarbības novērtēšanai. Kopā tika atstāta daļa attēlu bez cilvēkiem tajos, lai varētu novērtēt modeļu darbību šādos gadījumos.



6. att. Validācijas datu kopas paraugs

11 Meklēšanas protokls

Darbā aplūkoti gan modernākie jeb SOTA (state of the art) darbi, gan arī nozīmīgi darbi, kas vairs neiekļaujas SOTA kategorijā, taču tajā ir atradušies.
Darbi tika meklēti, izmantojot dažādus rīkus un paņēmienus:

1. Meklēšana pēc atslēgas vārdiem:
    - Pose estimation

- Human pose estimation
- Human keypoint estimation
- Human keypoint SOTA

2. Pēc darbiem, uz ko atsaucās jau atrastie darbi, izmantojot researchrabbitapp, semanticschollar, SCOPUS, PRIMO, IEEE, arxiv

3. Pēc darbiem, kas atsaucas uz jau atrastajiem darbiem, izmantojot researchrabbitapp, semanticschollar, SCOPUS, PRIMO, IEEE, arxiv

4. Netika izvēlēti darbi, kas vecāki par 2015. gadu, kā arī darbi, kas vecāki par 2020. gadu un uz kuriem atsaucas mazāk nekā 100 citi darbi.

5. Tika apskatīti aplūkoto datu kopu labākie rezultāti un šo rezultātu publikācijas izmantojot datu kopu mājas lapas un vietni paperswithcode (iekļaujot vismaz 5 labākos rezultātus katrai datu kopai (COCO, MPII vienas personas un MPII vairāku personu), ja tie atbilst iepriekš minētajiem kritērijiem)

12 Salīdzināšanas protokls

Darbu salīdzināšana tika veikta pēc konkrētiem kritērijiem, katram darbam pie katra kritērija tika atzīmēts vai darbs tajā iekļaujas vai nē.
Kvalitātes kritēriji:

KK01 Labākais rezultāts Coco Val. AP

KK02 Labākais rezultāts Coco Test. AP

KK03 Labākais rezultāts MPII PCKh (viena cilvēka)

KK04 Labākais rezultāts MPII mAP (vairāku cilvēka)

KK05 publicēts pēc 2020. gada?

KK06 Vai ir publicēts pirmkods modeļa apmācībai?

KK07 Vai ir pieejams apmācīts modelis?

KK08 Virs 100 citātiem gadā (ietekmīgs darbs)?

KK09 Īsākais darbības laiks ātrdarbības salīdzinājumā.

KK10 Modelis praksē spēj atrast atslēgpunktus attēlā (ja modeli nav iespējams notestēt, kritērijs neizpildās).

13 Ātrdarbības salīdzinājuma protokls

Praktiskai darbu salīdzināšanai, darba ietvaros tie tiek salīdzināti pēc to veiktspējas uz reālas sistēmas. Šāds salīdzinājums pamatojams ar mūsdienās pieaugošo datu plūsmu apjomu un apstrādājamo datu apjomu, kas nozīmē ka modeļu precizitāte nav vienīgais svarīgais kritērijs bet arī to energoefektivitāte un ātrdarbība ir jāņem vērā.

Modeļu testēšanai tika paņemta COCO datu kopas test2017 attēlu kopa un no tās atlasīta 4560 attēlu apakškopa, kas satur gan attēlus ar cilvēkiem, gan attēlus ar citiem objektiem. Datu kopas attēlu paraugs attēlots 6.. attēlā.

Visi modeļi tiek pielietoti šai 4560 attēlu kopai, mērot to izpildes laiku (netiek ņemts vērā laiks kas pavadīts modeli ielādējot atmiņā). Iegūto laiku izdalot ar attēlu skaitu tiek iegūts vidējais laiks viena attēla apstrādei.

No kopas arī tika izdalīti 20 attēli, kuros tika veikta kvalitatīvā analīze rezultātiem. 20 attēlu kopa sastāvēja no 10 attēliem ar cilvēkiem un 10 attēliem bez.

Testa sistēmas parametri aprakstīti 3. tabulā.

3. tabula. Testa sistēmas parametri

Operētājsistēma	Ubuntu 22.10 64bit
Video processors	NVIDIA GeForce RTX 3060 Ti, 8GB
CUDA versija	11.7
Brīvpiekļuves atmiņa	32GB
SWAP atmiņa	110GB
Processors	Intel® Core™ i7-10700K
Cietais disks	Western Digital 1TB WD Blue 3D NAND

14 Rezultāti

Balstoties uz meklēšanas parametriem, kas aprakstīti 11.. nodaļā, darbā tika iekļauti 17 pētījumi. Iekļautie darbi uzskaitīti 4. tabulā. Lai būtu vieglāk izsekot iekļautajiem darbiem to salīdzināšanas procesā darbi tika numurēti un šī numerācija izmantota tālāk salīdzināšanas tabulās un aprakstos.

4. tabula. Aplūkotie darbi

Pētījuma Nr.	Nosaukums	Gads	Citātu skaits	Piederība
1	OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields [7]	2019	2352	Carnegie Mellon University University of California Facebook Reality Labs
2	Realtime multi-person 2d pose estimation using part affinity fields [8]	2016	4244	Carnegie Mellon University University of California Facebook Reality Labs
3	Stacked Hourglass Networks for Human Pose Estimation [18]	2016	3650	University of Michigan
4	RMPE: Regional Multi-person Pose Estimation [10]	2016	907	Tencent Shanghai Jiao Tong University
5	Learning Delicate Local Representations for Multi-Person Pose Estimation [5]	2020	83	Megvii Inc. Tsinghua University Chinese Academy of Sciences Beihang University Ocean University of China
6	Deep High-Resolution Representation Learning for Human Pose Estimation [22]	2019	1702	University of Science and Technology of China Microsoft Research Asia
7	Rethinking on Multi-Stage Networks for Human Pose Estimation [13]	2019	141	Megvii Inc. (Face++) Shanghai Jiao Tong University Beihang University Beijing University of Posts and Telecommunications
8	EfficientPose: Efficient human pose estimation with neural architecture search [27]	2020	21	School of EIC, Huazhong University of Science and Technology Institute of Artificial Intelligence, Huazhong University of Science and Technolology
9	ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation [23]	2022	18	The University of Sydney JD Explore Academy
10	Towards High Performance Human Keypoint Detection [26]	2020	32	The University of Sydney
11	TransPose: Keypoint Localization via Transformer [24]	2020	63	School of Automation, Southeast University, Nanjing 210096, China
12	UniPose: Unified Human Pose Estimation in Single Images and Videos [3]	2020	61	Rochester Institute of Technology
13	OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation [2]	2021	10	Rochester Institute of Technology
14	Polarized Self-Attention: Towards High-quality Pixel-wise Regression [16]	2021	54	Nanjing University of Science and Technology Carnegie Mellon University
15	Toward fast and accurate human pose estimation via soft-gated skip connections [4]	2020	54	Samsung AI Center
16	Single-Stage Multi-Person Pose Machines [19]	2019	117	Department of Electrical and Computer Engineering, National University of Singapore Yitu Technology
17	Associative Embedding: End-to-End Learning for Joint Detection and Grouping [17]	2016	650	University of Michigan Tsinghua University

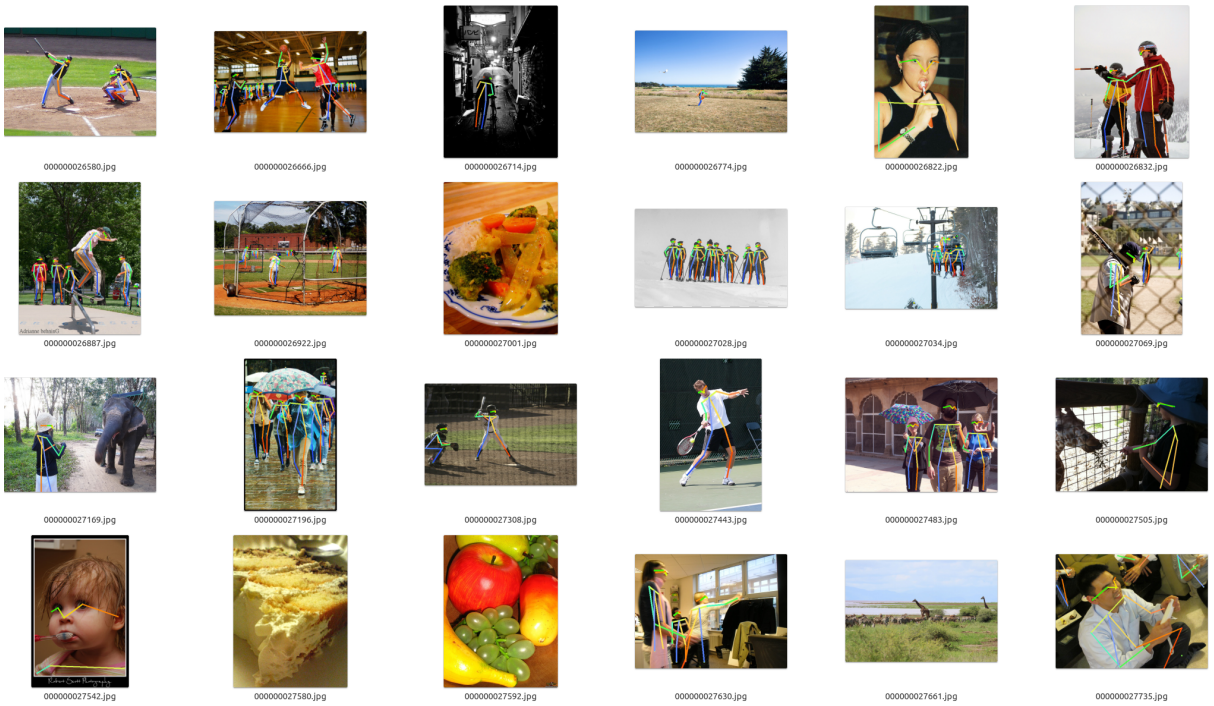
No darbu satura tika apkopota informācija par to kā darbu autoriem veicies aplūkoto datu kopu risināšanā. Rezultāti salīdzināti 5. tabulā. Tabulā izcelti labākie darbi balstoties uz metrikām.

5. tabula. Rezultāti uz MPII un COCO datu kopām

Pētējuma Nr.	Coco Val. (AP)	Coco Test (AP)	MPII (mAP)	MPII (PCKh)
1	65.3	64.2	75.6	
2	61	60.5	75.6	
3				90.9
4	72.3		**82.1**	
5	77.1	79.2		93
6	76.3	77		92.3
7	76.4	78.1		92.6
8		71.3		89.5
9	**79.5**	**81.1**		**94.3**
10		78.9		
11	75.8	75		93.5
12				92.7
13	**79.5**	76.4		92.3
14	78.9	79.5		
15				94.1
16		66.9	78.5	
17		65.5	77.5	

6.. tabulā atspoguļoti darba ietvaros veikto modeļu ātrdarbības eksperimentu rezultāti. Eksperimentus varēja veikt tikai 5 no 17 iekļautajiem darbiem, jo daļa autoru nav publicējuši apmācītus modeļus vai nav pieejams kods/dokumentācija kā izmantot publicētos modeļus ar datiem, kas nav akadēmiskā datu kopa (MPII vai COCO).

Pirmkods, kas izmantots modeļu salīdzināšanai, pieejams https://github.com/Puupuls/kurs darbs un 7.. attēlā redzams rezultāts pēc modeļa pielietošanas uz datu kopu.



7. att. Validācijas datu kopas paraugs pēc modeļa pielietošanas

6. tabula. Publicēto darbu ātrdarbības salīdzinājums

Pētījuma Nr.	Izpildes laiks (sec)	Izpildes laiks uz 1 attēlu (sec)	Attēli kuros atrada personas
4	312	0.0684	55%
6	593	0.1300	50%
9	256	0.0561	90%
11	510	0.1118	50%
13	1329	0.2914	100%

Balstoties uz darbu rezultātiem, to ietekmi uz tālākajiem darbiem un citiem Salīdzināšanas protokls nodaļā aplūkotajiem kritērijiem tika veidots šo darbu salīdzinājums 7. tabulā.

7. tabula. Aplūkoto darbu salīdzinājums

Pētījuma Nr.	KK01	KK02	KK03	KK04	KK05	KK06	KK07	KK08	KK09	KK10	Punkti
1	X	X	X	X	X	✓	✓	✓	X	X	3
2	X	X	X	X	X	✓	✓	✓	X	X	3
3	X	X	X	X	X	✓	✓	✓	X	X	4
4	X	X	X	✓	X	✓	✓	✓	X	✓	5
5	X	X	X	X	✓	✓	X	X	X	X	2
6	X	X	X	X	X	✓	✓	✓	X	✓	4
7	X	X	X	X	X	✓	✓	X	X	X	2
8	X	X	X	X	✓	✓	✓	X	X	X	3
9	✓	✓	✓	X	✓	X	✓	X	✓	✓	**7**
10	X	X	X	X	✓	X	X	X	X	X	1
11	X	X	X	X	✓	✓	✓	X	X	✓	4
12	X	X	X	X	✓	X	✓	X	X	X	2
13	✓	X	X	X	✓	✓	✓	X	X	✓	5
14	X	X	X	X	✓	X	X	X	X	X	1
15	X	X	X	X	✓	X	X	X	X	X	1
16	X	X	X	X	X	X	X	X	X	X	0
17	X	X	X	X	X	✓	✓	✓	X	X	3

15 Secinājumi

Darba ietvaros veiktā sistemātiskā literatūras analīze un eksperimenti, pēc kvalitātes kritērijiem, ļauj spriest, ka darbs "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation"[23] ir uzskatāms par nozīmīgāko darbu jomā, jo ieguva septiņus no desmit punktiem izvirzītajos kvalitātes salīdzināšanas kritērijos gūstot labāko rezultātu COCO un MPII vienpersonas datu kopās un parādot ieverojamu ātrdarbību veiktajā eksperimentā apstrādājot attēlu vidēji 56 milisekunžu laikā.

Vissliktākos rezultātus uzrādīja darbs "Single-Stage Multi-Person Pose Machines"[19], kas, lai arī atrodas otrajā vietā MPII daudzpersonu pozu noteikšanas uzdevumā pēc precizitātes, neieguva nevienu punktu pēc kvalitātes vērtēšanas kritērijiem.

Darba izstrādes laikā atklājās, ka zinātnisko rakstu autori bieži nepublicē savu pirmkodu, vai publicē nestrādājošu kodu, kas apgrūtina rezultātu pārbaudi praksē un modeļu objektīvu salīdzināšanu ārpus to izstrādes vides.

Daļa no modeļiem sevī neiekļauj pārbaudes vai attēlā redzami cilvēki, kas rezultējās ar lielu nepareizu pozitīvo paredzējumu īpatsvaru datu kopā kā redzams 6. tabulā, kur "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation" atrada personu 90% attēlu un "OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation" atrada personu 100% attēlu no 20 attēlu testkopas kurā cilvēki atradās 50% attēlu.

Darba ietvaros apskatīto problēmsfēru var papildināt ar jaunām datu kopām, datu sagatavošanas metodēm, jauniem modeļiem, piemēram, transformeru tipa modeļiem, un citiem papildinājumiem, kuri ir ārpus šīs sistēmatiskās literatūras analīzes ietvariem. Kā piemērs no veiksmīgiem jaunākajiem modeļiem ir ViTPose, kurš arī ir balstīts transformeru arhitektūrā atslēgpunktu meklēšanas uzdevumiem.

Bibliogrāfija

[1] Mykhaylo Andriluka u. c. "2D Human Pose Estimation: New Benchmark and State of the Art Analysis". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014. g. jūn.

[2] Bruno Artacho un Andreas E. Savakis. "OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation". *ArXiv* abs/2103.10180 (2021).

[3] Bruno Artacho un Andreas E. Savakis. "UniPose: Unified Human Pose Estimation in Single Images and Videos". *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 7033.—7042. lpp.

[4] Adrian Bulat u. c. "Toward fast and accurate human pose estimation via soft-gated skip connections". *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)* (2020), 8.—15. lpp.

[5] Yuanhao Cai u. c. "Learning Delicate Local Representations for Multi-Person Pose Estimation". *European Conference on Computer Vision*. 2020.

[6] Zhe Cao u. c. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2018), 172.—186. lpp.

[7] Zhe Cao u. c. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2018), 172.—186. lpp.

[8] Zhe Cao u. c. "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields". *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 1302.—1310. lpp.

[9] Hongwei Dong, Lamei Zhang un and Zou. "PolSAR Image Classification with Lightweight 3D Convolutional Networks". *Remote Sensing* 12 (2020. g. janv.), 396. lpp. DOI: 10.3390/rs12030396.

[10] Haoshu Fang u. c. "RMPE: Regional Multi-person Pose Estimation". *2017 IEEE International Conference on Computer Vision (ICCV)* (2016), 2353.—2362. lpp.

[11] Ian Goodfellow, Yoshua Bengio un Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[12] Harrison Kinsley un Daniel Kukieła. *Neural networks from Scratch in Python*. https://nnfs.io. 2020.

[13] Wenbo Li u. c. "Rethinking on Multi-Stage Networks for Human Pose Estimation". *ArXiv* abs/1901.00148 (2019).

[14] Tsung-Yi Lin u. c. “Microsoft COCO: Common Objects in Context”. *European Conference on Computer Vision*. 2014.

[15] Tsung-Yi Lin u. c. “Microsoft COCO: Common Objects in Context”. *CoRR* abs/1405.0312 (2014). arXiv: 1405.0312. URL: http://arxiv.org/abs/1405.0312.

[16] Huajun Liu u. c. “Polarized Self-Attention: Towards High-quality Pixel-wise Regression”. *ArXiv* abs/2107.00782 (2021).

[17] Alejandro Newell, Zhiao Huang un Jia Deng. “Associative Embedding: End-to-End Learning for Joint Detection and Grouping”. *ArXiv* abs/1611.05424 (2016).

[18] Alejandro Newell, Kaiyu Yang un Jia Deng. “Stacked Hourglass Networks for Human Pose Estimation”. *European Conference on Computer Vision*. 2016.

[19] Xuecheng Nie u. c. “Single-Stage Multi-Person Pose Machines”. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 6950.—6959. lpp.

[20] David E. Rumelhart, Geoffrey E. Hinton un Ronald J. Williams. “Learning representations by back-propagating errors”. *Nature* 323 (1986), 533.—536. lpp.

[21] Stuart Russell un Peter Norvig. *Artificial Intelligence: A Modern Approach, 4th Global ed.* http://aima.cs.berkeley.edu/global-index.html. 2021.

[22] Ke Sun u. c. “Deep High-Resolution Representation Learning for Human Pose Estimation”. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 5686.—5696. lpp.

[23] Yufei Xu u. c. “ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation”. *ArXiv* abs/2204.12484 (2022).

[24] Sen Yang u. c. “TransPose: Keypoint Localization via Transformer”. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2020), 11782.—11792. lpp.

[25] Aston Zhang u. c. “Dive into Deep Learning”. *ArXiv* abs/2106.11342 (2021).

[26] Jing Zhang, Zhe Chen un Dacheng Tao. “Towards High Performance Human Keypoint Detection”. *International Journal of Computer Vision* 129 (2020), 2639.—2662. lpp.

[27] Wenqiang Zhang u. c. “EfficientPose: Efficient Human Pose Estimation with Neural Architecture Search”. *Comput. Vis. Media* 7 (2020), 335.—347. lpp.

Kursa darbs ”Skeleta pozu prognozēšanas modeļu sistemātiskā literatūras analīze” izstrādāts LU Datorikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts pastāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai:

Autors: _____ Pauls Purviņš

Rekomendēju/nerekomendēju darbu aizstāvēšanai

Vadītājs/a: Darba vadītājs: Phd. Comp. Sc. Ēvalds Urtāns _____

Janvāris 2023

Darbs aizstāvēts kursa darba komisijas sēdē

____.____._____ .prot. Nr. ____

Komisijas sekretāre: _____