

**Бройде Юлій Ігорович**Аспірант кафедри комп'ютерних наук та системного аналізу, [orcid.org/0000-0002-5236-3212](https://orcid.org/0000-0002-5236-3212)

Черкаський державний технологічний університет, Черкаси

**КІЛЬКІСНИЙ МЕТОД ПІДРАХУНКУ ПОВТОРЕНЬ ФІЗИЧНИХ ВПРАВ  
ЗА ВИХІДНИМ СИГНАЛОМ НЕЙРОННОЇ МЕРЕЖІ**

***Анотація.** Нині у світі швидкими темпами розвиваються системи штучного інтелекту та комп'ютерного зору, зокрема випускають нові архітектури нейронних мереж, що оцінюють тривимірну позу людини за відео. Такі нейронні мережі потребують аналізу їх вихідного сигналу з метою отримання корисних для кінцевого користувача даних та їх подальшої інтеграції у користувацькі системи. Автором запропоновано новий метод аналізу вихідного сигналу нейронної мережі, що оцінює положення людини у просторі, який виконує підрахунок повторень вправи "присідання". В основі цього методу лежить машина станів, яка додає одиницю до лічильника повторень у момент закінчення циклу вправи. Застосування цього методу на початкових етапах алгоритму аналізу вправи допоможе у подальшому розробляти системи, що перевіряють техніку присідання та допомагають спортсменам і тренерам під час тренування, а також вченим у сфері біомеханіки під час їхньої професійної діяльності. Відмінною рисою цього методу є стійкість як до викидів вхідного сигналу, тобто неправильних результатів розпізнавання пози людини нейронною мережею, так і до рухів людини, які не належать до вправи безпосередньо. Також застосування цього методу до аналізу сигналу нейронної мережі дає змогу поєднати позитивні якості, властиві нейронним мережам, використовуваним в комп'ютерному зорі (допустимість високої варіабельності одягу і фону), і позитивні якості аналітичних і алгоритмічних методів (легка інтерпретованість результатів, зручне налагодження, можливість використання предметного досвіду спеціалістів для підбору параметрів). Пропонований метод не є специфічним по відношенню до будь-якої конкретної нейронної мережі, а тому може використовуватися на виході практично будь-якої системи, яка визначає положення суглобів людини в просторі. Крім опису методу, в статті наведено результати його випробувань в різних умовах. Така схема випробувань може застосовуватися не тільки для докладання цього методу до вправи "присідання", але і до будь-якої іншої циклічної вправи.*

**Ключові слова:** комп'ютерний зір; біомеханіка; штучний інтелект; підрахунок повторень; присідання; машина станів

**Вступ**

Сучасні напрацювання в сфері штучного інтелекту (ШІ) для комп'ютерного зору (КЗ) дали можливість визначення пози людини в просторі за одним відеопотоком так, як це може робити людина з одним закритим оком. Суттєвим обмеженням цих систем є низька точність, а також неможливість оцінки розподілу помилки сигналу.

Слабкий розвиток методів для аналізу дискретного сигналу положення людини в просторі з такими характеристиками стримує використання вищезгаданих систем ШІ в практичних цілях.

Підрахунок кількості присідань з достатньою точністю на базі аналізу сигналу вищезазначених систем ШІ є одним з найпростіших завдань. Це надійне рішення, що дасть змогу розробляти методи як для аналізу інших рухів (екстенсивно), так і для більш складних завдань (інтенсивно), а згодом й

реалізовувати системи, які перевіряють техніку присідання на практиці.

**Мета статті**

Метою статті є представлення методу підрахунку кількості присідань за вихідним сигналом нейронної мережі, що оцінює положення людини в просторі за одним відеопотоком, для забезпечення можливості розробляти методи як для аналізу інших рухів (екстенсивно), так і для більш складних завдань (інтенсивно), а згодом й реалізовувати системи, які перевіряють техніку присідання на практиці.

**Виклад основного матеріалу**

В останні два десятиліття було запропоновано багато підходів до задачі підрахунку повторень за відеопотоком. Ці методи можна розділити на дві категорії:

1. Класичні методи (основний розвиток набули у період з 2005 р. по 2015 р., коли з'явилися цифрові відеокамери) [1].

1.1. Частотні та спектральні доменні методи (метод Фур'є або методи вейвлет-аналізу).

1.1.1. Briassouli and N. Ahuja. Extraction and analysis of multiple periodic motions in video sequences. PAMI, 2007.

1.1.2. Pogalin, A. Smeulders, and A. Thean. Visual quasiperiodicity. In CVPR, 2008.

1.2. Методи, що базуються на автокореляції.

1.2.1. Azy and N. Ahuja. Segmentation of periodically moving objects. In ICPR, 2008.

1.2.2. Li, X. Han, W. Lin, and H. Wei. Periodic motion detection with roi-based similarity measure and extrema-based reference selection. Consumer Electronics, 2012.

1.3. Методи кластеризації руху.

1.3.1. Zhou, F. D. la Torre, and J. K. Hodgins. Hierarchical aligned cluster analysis for temporal clustering of human motion. PAMI.

2. Методи, засновані на нейронних мережах (є сферою активного дослідження в нашій країні).

2.1. Універсальні методи, які працюють без навчання на специфічних об'єктах присутніх на відео.

2.1.1. Debidatta Dwibedi 1, Yusuf Aytar 2, Jonathan Tompson 1, Pierre Sermanet 1, and Andrew Zisserman Counting Out Time: Class Agnostic Video Repetition Counting in the Wild CVPR '20.

2.1.2. Ofir Levy Lior Wolf, Live Repetition Counting ICCV 2015.

2.2. Специфічні методи, які працюють на конкретних об'єктах або активностях.

2.2.1. Bruno Ferreira Email author Pedro M. Ferreira Gil Pinheiro Nelson Figueiredo Filipe Carvalho Paulo Menezes Jorge Batista: Exploring Workout Repetition Counting and Validation Through Deep Learning, ICIAR 2020.

Перевагою класичних методів є легкість інтерпретації і часто набагато більш низька потреба в обчислювальних потужностях. Мінусом цих методів є відсутність "розуміння ситуації", через що вони працюють тільки в добре контрольованому середовищі і вузькій предметній області. Також у багатьох класичних методах ресурсомістким є завдання визначення оптимальних гіперпараметрів.

Перевагою методів, заснованих на нейронних мережах є якраз набагато простіше визначення гіперпараметрів [2], а також широка предметна сфера, яка часто обмежена лише різноманітністю та обсягом набору даних для навчання [3].

Наша робота лежить на стику цих категорій ("класичний підхід" і "нейронні мережі"), оскільки працює безпосередньо з вихідним сигналом

нейронних мереж. Більшість класичних підходів працюють з таким типом сигналів, оскільки статистичні характеристики помилки сигналу сильно змінюються для різних відеосигналів.

### Вхідний сигнал

На вході системи, яка реалізує описаний в статті метод, очікується сигнал, який представляє послідовність наборів координат суглобів людини в просторі. Приклад сучасних ІІІ, які видають такий сигнал, описано в статтях [4 – 6].

Такий сигнал може бути отриманий не тільки із системи штучного інтелекту, який аналізує відео, але і з більш традиційних таких систем, як MoCAP системи, засновані на гіроскопах (Xsens), або на аналізі стану міток з декількох камер (Vicon), а також на RGB-D сенсорах (Kinect) [7]. Проте саме аналіз сигналу (підрахунок повторень вправ) із сучасного ІІІ є найбільш складним завданням, в зв'язку з тим, що на даний момент такий сигнал є менш точним, ніж сигнал класичних систем, а також статистичні характеристики викидів такого сигналу не мають математичної моделі.

Для того щоб можна було розробляти системи, які аналізують такий багатовимірний сигнал, як послідовність поз людини, для початку потрібна система, яка здійснює візуальну репрезентацію сигналу (рис. 1). Візуальна репрезентація була зроблена за допомогою пакета matplotlib [8].

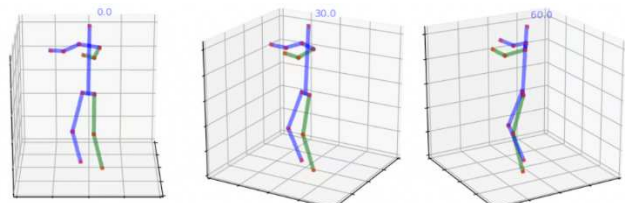


Рисунок 1 – Візуальна репрезентація сигналу

### Препроцесінг вхідного сигналу

Хоча статистичні характеристики вхідного сигналу невідомі, після візуального аналізу можна однозначно стверджувати, що для нерекурентних систем необхідне згладжування для елімінації високочастотної складової. Таке згладжування можливо за допомогою алгоритмів за типом:

- 1) moving average;
- 2) Savitzky–Golay filter (extended) [9].

Класичні FIT і IIR фільтри високих частот не підходять, оскільки для їх роботи потрібна велика кількість семплів, які не завжди доступні. А також на практиці було помічено, що вони вносять великі спотворення в зони навколо викидів вхідного сигналу, ніж вищенаведені фільтри.

Запропонована стейт-машина зображена на рис. 2.

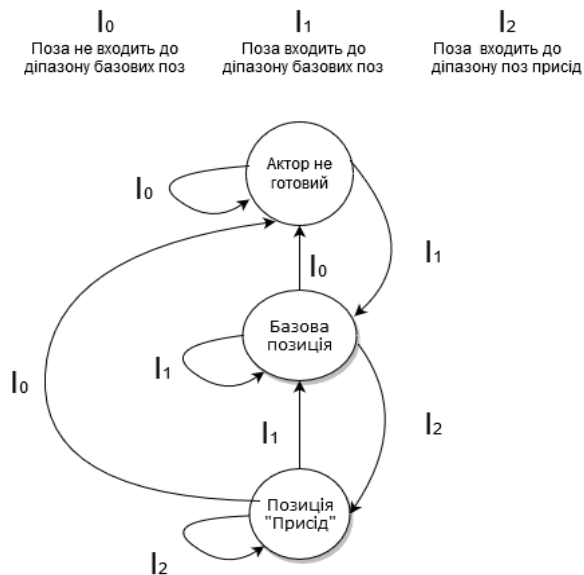


Рисунок 2 – Подання машини станів

Для підрахунку присідань запропонована така машина станів. Мається на увазі, що кожен атом часу «поза» класифікується так:

$I_0$  – поза не може бути частиною вправи (не входить в діапазон базових поз вправи).

Мається на увазі, що людина перебуває в цій позі, якщо вона сама встановлювала камеру і рухається від камери до місця заняття. Також людина може опинитися в цій позі, якщо вона перервалася, перестала виконувати вправу, і наприклад, розвернулася і пішла кудись.

$I_1$  – поза є базовою позицією вправи (для присідання це поза, коли людина стоїть рівно, її обличчя спрямоване до камери, коліна розігнуті).

$I_2$  – поза входить у простір поз, які можна назвати як "присідання".

При цьому машина складається з трьох станів:

1) Стан "Актор не готовий" – означає, що людина не виконує вправу в даний момент. Машина потрапляє в цей стан з будь-якого іншого при класифікації пози як  $I_0$ .

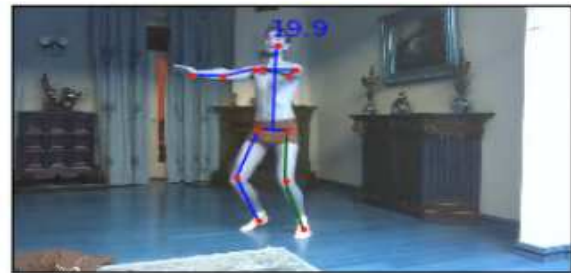
2) Стан "Базова позиція" – означає, що людина починає виконання повторення вправи. Перехід в дану позицію відбувається при класифікації пози як  $I_1$ .

3) Стан "Позиція Присідання" – означає, що людина перебуває в ключовому етапі вправи. Перехід в дану позицію відбувається тільки зі стану "Базова позиція" при класифікації пози як  $I_2$ .

### Метод класифікації поз

Для того щоб класифікувати пози, необхідно ввести математичні критерії, за якими можна визначити приналежність пози до конкретного класу. Ключовою особливістю запропонованих критеріїв є перетин меж класів поз і визначення класу пози

лежачого в перетині на підставі класу попередньої пози, або класів декількох попередніх поз. Такий підхід створює певний "гістерезис", який і дає змогу уникнути серйозного впливу викидів вхідного сигналу на точність системи.

Рисунок 3 – Поза з присіданням на 19 градусів ( $I_1$ )

Точні критерії мають бути встановлені тренером спортсмена, але в цьому випадку автор пропонує такі умови:

#### 1. Поза $I_0$

Будь-яка поза, при якій кут між:

1) перпендикуляром до площини "ліве стегно – праве стегно – точка початку шиї", спроектований на площину підлоги;

2) лінією, що з'єднує точку камери і голову людини, спроектований на площину підлоги, є більше 30 градусів.

#### 2. Поза $I_1$

Будь-яка поза, яка:

1) не є  $I_0$ ;

2) в якій сумарний кут колін, поділений на два, менше 90 градусів;

3) до якої дві попередні пози задовольняли умови 1 і 2.

#### 3. Поза $I_2$ .

Будь-яка поза, яка:

1) не є  $I_0$ ;

2) в якій сумарний кут колін, поділений на два, більше 80 градусів;

3) до якої дві попередні пози задовольняли умови 1 і 2.

Перетин другої ознаки поз  $I_1$  і  $I_2$  перебуває в просторі поз, у яких сумарний кут колін, поділений на два, більше 80 градусів і менше 90. Коли поза виявляється в цьому просторі, вона класифікується відповідно до ознаки 3.

### Випробування і аналіз проблем

Випробування цього методу проводилися для вихідних сигналів з нейронних мереж, отриманих при зйомці відео з:

а) трьох різних позицій камери по відношенню до людини на одній висоті;

б) трьох різних позицій камер по відношенню до людини на різній висоті;

- в) при зйомці в трьох різних наборах одягу;
- г) при зйомці в трьох різних локаціях.

Було знято 81 відео (3<sup>4</sup>). В результаті випробувань було з'ясовано, що основні помилки відбуваються при викидах вхідного сигналу, які повторюються велику кількість разів (більше трьох). Основними помилками були неврахування присідання, або подвійний облік присідання.

Кілька послідовних виборів відбувалися при:

- а) критичних кутах зйомки;
- б) серйозних змінах освітлення протягом зйомки при низькому освітленні;
- в) при зйомці людини в одязі, колір якої схожий з кольором фону.

### Висновок

За результатами проведеного дослідження можна зробити висновок, що машина станів може

використовуватися як основа методу підрахунку кількості присідань за вихідним сигналом нейронної мережі, що оцінює положення людини в просторі за одним відеопотоком. Цей підхід можна використовувати також для підрахунку кількості повторень інших циклічних вправ, що мають одну активну і одну базову позицію, за умови, що ці вправи виконуються найбільшими суглобами людини (тазовий, колінний, плечовий, локтєвий), а також, що людина стоїть у позі відносно камери, з якої чітко видно ці суглоби.

У подальшому планується проводити дослідження стосовно передбачення кількості повторень, а також виправлення лічильника у разі, якщо передбачення не реалізувалось, для забезпечення можливості роботи вищезгаданого методу в системах, що працюють в режимі онлайн.

### Список літератури

1. Ofir Levy, Lior Wolf. Live Repetition Counting. *The 2015 IEEE International Conference on Computer Vision (ICCV'15)* : proceedings. IEEE Computer Society, USA, 2015. P. 3020–3028. DOI: <https://doi.org/10.1109/ICCV.2015.346>.
2. Xin He, Kaiyong Zhao, Xiaowen Chu. AutoML: A Survey of the State-of-the-Art. URL: <https://arxiv.org/abs/1908.00709>.
3. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. URL: <https://pubs.acs.org/doi/pdf/10.1021/ac00205a007>.
4. Sijin Li, Antoni B. Chan. 3D Human Pose Estimation from Monocular Images with Deep Convolutional Neural Network (ACCV 14). *Asian Conference on Computer Vision*. Springer, Cham, 2014. P. 332–347.
5. Ching-Hang Chen, Deva Ramanan: 3D Human Pose Estimation = 2D Pose Estimation + Matching. *The 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)* : proceedings. IEEE, Honolulu, HI, USA, 2017. DOI: 10.1109/CVPR.2017.610.
6. Dario Pavllo, Christoph Feichtenhofer, David Grangier, Michael Auli. 3D human pose estimation in video with temporal convolutions and semi-supervised training. *The 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'19)* : proceedings. IEEE, 2019. P. 7753–7762.
7. Jan Smisek, Michal Jancosek, Tomas Pajdla. 3D with Kinect. Consumer Depth Cameras for Computer Vision. URL: [https://link.springer.com/chapter/10.1007/978-1-4471-4640-7\\_1](https://link.springer.com/chapter/10.1007/978-1-4471-4640-7_1).
8. John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. IEEE, 2007. Vol. 9, Is. 3. P. 90-95. URL: <https://ieeexplore.ieee.org/document/4160265>. DOI: 10.1109/MCSE.2007.55.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*. 2017. 60(6). P. 84–90.

Стаття надійшла до редколегії 03.11.2020

#### Бройда Юлія

Postgraduate student of the Department of Computer Science and Systems Analysis, [orcid.org/0000-0002-5236-3212](https://orcid.org/0000-0002-5236-3212)  
Cherkasy State Technological University, Cherkasy

### Q QUANTITATIVE METHOD OF CALCULATING APPROACHES OF EXERCISES ON THE BASIS OF OUTPUT SIGNAL OF THE NEURAL NETWORK

**Abstract.** Today, artificial intelligence and computer vision systems are developing rapidly in the world; in particular, new architectures of neural networks that assess the three-dimensional human posture on video are produced. Such neural networks require analysis of their output signal in order to obtain useful data for the end user and their subsequent integration into user systems. The author proposes a new method of analysis of the output signal of the neural network, which estimates the position of a person in space that performs the calculation of repetitions of the exercise "squat". This method is based on the state machine, which adds one to the repetition counter at the end of the exercise cycle. The application of this method in the initial stages of the algorithm of exercise analysis will allow further development of systems that test squats and help athletes and coaches during

training, as well as scientists in the field of biomechanics during their professional activities. A distinctive feature of this method is resistance to both input signal emissions, i.e. incorrect results of human posture recognition by the neural network, and to human movements that do not belong to the exercise directly. Also, the application of this method to the analysis of the neural network signal allows to combine the positive qualities inherent in neural networks used in computer vision (admissibility of high variability of clothing and background), and the positive qualities of analytical and algorithmic methods (easy interpretability of results, convenient adjustment, possibility to use the experts' subject experience for the selection of parameters). This method is not specific to any particular neural network and therefore can be used at the output of almost any system that determines the position of human joints in space. In addition to the description of the method, the article presents the results of its tests in different conditions. This test scheme can be used not only to apply this method to the exercise "squat", but to any other cyclic exercise.

**Keywords:** BIM-technologies; artificial intelligence; virtual assistant

#### References

1. Levy, Ofir & Wolf, Lior. (2015). Live Repetition Counting. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV'15)*. IEEE Computer Society, USA, pp. 3020–3028. DOI: <https://doi.org/10.1109/ICCV.2015.346>.
2. He, Xin, Zhao, Kaiyong, Chu, Xiaowen. AutoML: A Survey of the State-of-the-Art. URL: <https://arxiv.org/abs/1908.00709>.
3. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method URL: <https://pubs.acs.org/doi/pdf/10.1021/ac00205a007>.
4. Li, Sijin, Chan, Antoni B. (2014). 3D Human Pose Estimation from Monocular Images with Deep Convolutional Neural Network (ACCV 14). *Asian Conference on Computer Vision*, 332–347.
5. Chen, Ching-Hang, Ramanan, Deva. (2017). 3D Human Pose Estimation = 2D Pose Estimation + Matching. *The 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)* : proceedings. DOI: 10.1109/CVPR.2017.610.
6. Pavlo, D., Feichtenhofer, C., Grangier, D. & Auli, M. (2019). 3D human pose estimation in video with temporal convolutions and semi-supervised training. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7753-7762.
7. Smisek, Jan, Jancosek, Michal, Pajdla, Tomas. 3D with Kinect. Consumer Depth Cameras for Computer Vision. URL: [https://link.springer.com/chapter/10.1007/978-1-4471-4640-7\\_1](https://link.springer.com/chapter/10.1007/978-1-4471-4640-7_1).
8. Hunter, John D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 9, 3, 90–95. URL: <https://ieeexplore.ieee.org/document/4160265>. DOI: 10.1109/MCSE.2007.55.
9. Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

#### Посилання на публікацію

- APA Broyda, Juliy. (2020). Quantitative method of calculating approaches of exercises on the basis of output signal of the neural network. *Management of Development of Complex Systems*, 44, 65 – 69, [dx.doi.org\10.32347/2412-9933.2020.44.65-69](https://doi.org/10.32347/2412-9933.2020.44.65-69).
- ДСТУ Бройде Ю.І. Кількісний метод підрахунку повторень фізичних вправ за вихідним сигналом нейронної мережі. *Управління розвитком складних систем*. Київ, 2020. № 44. С. 65 – 69, [dx.doi.org\10.32347/2412-9933.2020.44.65-69](https://doi.org/10.32347/2412-9933.2020.44.65-69).