

УДК 004.8

**Вадим Мухін,
Ярослав Корнага,
Андрій Любівий**

МОДЕЛЮВАННЯ Й РОЗПІЗНАВАННЯ ЖЕСТІВ СКЕЛЕТОНУ РУКИ

MODELING AND RECOGNITION OF GESTURE OF HAND SKELETON

В даній статті розглянуті проблеми моделювання й розпізнавання скелетону руки відзнятої з однієї камери без додаткових датчиків за рахунок детектування ключових точок руки та побудови моделі руки з цих точок для класифікації її серед описаних моделей жестів. В результаті було розроблено систему, систему для знаходження ключових точок руки за допомогою нейронної мережі, яка використовуючи їх, створює прототип образу руки й класифікує модель жесту.

Ключові слова: моделювання об'єктів, згорткові нейронні мережі.

Рис.: 5. Бібл.: 4.

In this article, problems of modeling and recognition of a skeleton of a hand from image which is taken with one camera without additional sensors are considered. Detected key points from the image with hand are using for the construction of a model of a hand from these points to classify it among the described gesture models. As a result, a system was developed, a system for finding key points of the hand using a neural network, which uses them, creates a prototype of the image of the hand and classifies the gesture model.

Keywords: modeling of objects, convolutional neural networks.

Fig.: 5. Bibl.: 4.

Актуальність теми дослідження. Проблема моделювання й розпізнавання жестів стає більш актуальною в останні дні у зв'язку зі зростаючою кількістю систем з якими людина може взаємодіяти за допомогою жестів. На сьогоднішній день величезна кількість техніки знаходиться у нас в домі: комп'ютери, телевізори, телефони, планшети й інші. Ми з ними постійно взаємодіємо, і ефективність цієї взаємодії впливає на наше життя. Частіше всього ми з ними взаємодіємо тактильним способом: відчуття доторку й тиску. Ми натискаємо на кнопки на пристрої чи на екран телефона. Але не завжди це можливо, бо в нас можуть бути зайняті руки; або зручно, бо прилад з яким ми взаємодіємо знаходиться на відстані від нас. І тому було б добре взаємодіяти з приладами за допомогою рук, а саме жестів.

Таким чином, проблемою є створення системи, що дасть можливість «спілкуватись» з іншими за допомогою жестів, що підвищить гнучкість й зручність використання. Дана робота присвячена проблемі розпізнавання жестів руки з моделей, які були побудовані за допомогою нейронної мережі.

Постановка проблеми. Через появу нових методів і підходів у галузі штучного інтелекту, тема розпізнавання зображень стала більш актуальною в останні роки.

Аналіз останніх досліджень і публікацій. Існують два основних способи розпізнавання моделей рук. Перший спосіб заснований на одяганні спеціальних електронних пристроїв (спеціальні рукавиці з датчиками), який в основному використовують в кіноіндустрії та добре працює, але є дорогим і не завжди придатний для використання в інших галузях[1-3].

Другий в основному використовує комп'ютерне бачення, яке передбачає обробку зображень за допомогою камер або(й) інших датчиків.

Протягом останніх років з'являється все більше статей присвячених обробці зображень з камери за допомогою спеціальних алгоритмів або за допомогою нейронних мереж. Якість роботи нейронних мереж значно підвищилась через використання нових архітектур нейронних мереж та великої кількості нових досліджень щодо підвищення якості та швидкості їх навчання. А також, через розвиток технологій та збільшення кількості даних[2-4].

Виділення недосліджених частин загальної проблеми. Дана стаття присвячена вивченню та аналізу запропонованого підходу для розпізнавання моделей рук. Дослідження сфокусовано на вивченні застосування згорткових нейронних мереж для детектування ключових точок руки й побудови моделі руки й розпізнавання цієї моделі.

Постановка завдання. Завдання є створити систему, яка дасть можливість, використовуючи згорткову нейронну мережу, детектувати ключові точки руки з одного зображення й після цього будувати модель руки, щоб порівняти її з описаними моделями для детектування жестів.

Викладення основного матеріалу. Проведено аналіз використання згорткових мереж для детектування ключових точок, створення моделі руки з них й подальшого розпізнавання цих моделей.

Для детектування ключових точок використовуємо архітектуру Convolutional Pose Machines (CPM). CPM повертає heatmaps кожної ключової точки.

CPM має неявне навчання залежностей між зображенням і виведенням зразу декількох ключових точок, тісна інтеграція між навчанням й виведенням та модульний послідовний дизайн. На кожному етапі CPM, отримується зображення та heatmaps, виготовлені на попередньому етапі, як вхідне й намагається виправити та уточнити положення ключових точок на руці.

Так як, створення згорткової мережі з багатьма модулями призводить до загальної мережі з багатьма шарами, що призводить до проблеми згасання градієнтів під час навчання, то створюються додаткові виходи нейронної мережі після кожного модуля, що використовуються лише під час навчання, які намагаються так само визначити положення ключових точок за допомогою heatmaps. Створення додаткових виходів під час навчання дає можливість поповнювати градієнти та спрямовувати мережу для створення все більш точних heatmaps.

Функція витрат, яку ми прагнемо мінімізувати на виході кожного етапу на кожному рівні, виглядатиме:

$$f_t = \sum_{p=1}^{P+1} \sum_{z \in Z} \|b_t^p(z) - b_t^p(z)\|_2^2 \quad (1)$$

Ідеальна heatmap для частини p , як $b_t^p(Y_p = z)$, що створює гаусові піки на місцях істинного розташування кожної ключової точки p , де z – це положення цієї ключової точки, а t – номер класифікатора(модуль нейронної мережі).

Загальна ціль для усієї архітектури виглядатиме, як додавання втрат на кожному етапі, тобто:

$$F = \sum_{t=1}^T f_t \quad (2)$$

Було зроблено знімок лівої руки для проведення обробки по ідентифікації жесту. Після подання знімку з рукою на нейронну мережу ми отримуємо heatmaps, з яких знаходимо ключові точки руки для моделювання жесту (рис. 1).

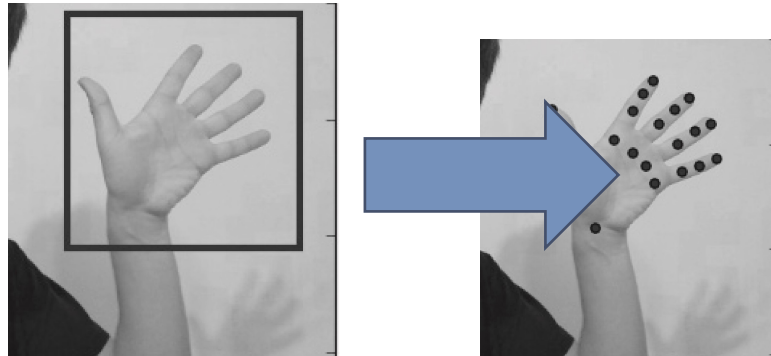


Рис. 1. Детектування ключових точок руки

Після цього нормалізуємо модель до кожної описаної моделі за допомогою rotate, shift, scale і flip (рис. 2, рис. 3, рис 4).



Рис. 2. Rotate моделі руки

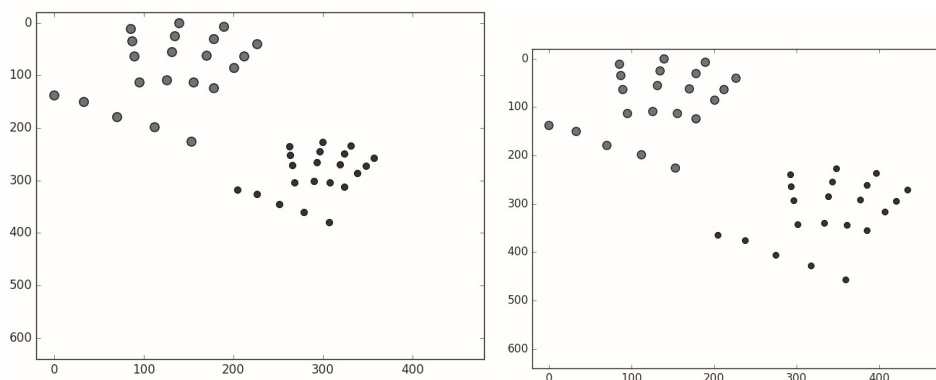


Рис. 3. Scale моделі руки

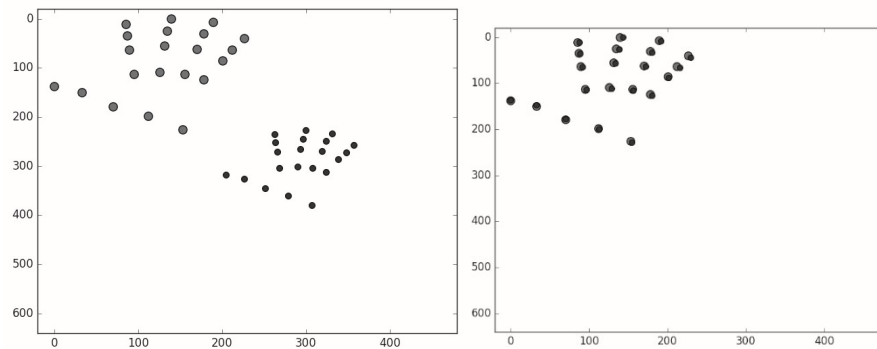


Рис. 4. Shift моделі руки

Після нормалізації побудованої моделі до даного жесту, вона порівнюється з ним та визначається чи задовольняє дана модель описаний жест. Як видно з рис. 4 наша модель руки задовольняє в межах допустимого відхилення заданий жест.

Загальний алгоритм. На рис. 5 показаний загальний алгоритм побудови та розпізнавання моделі руки.

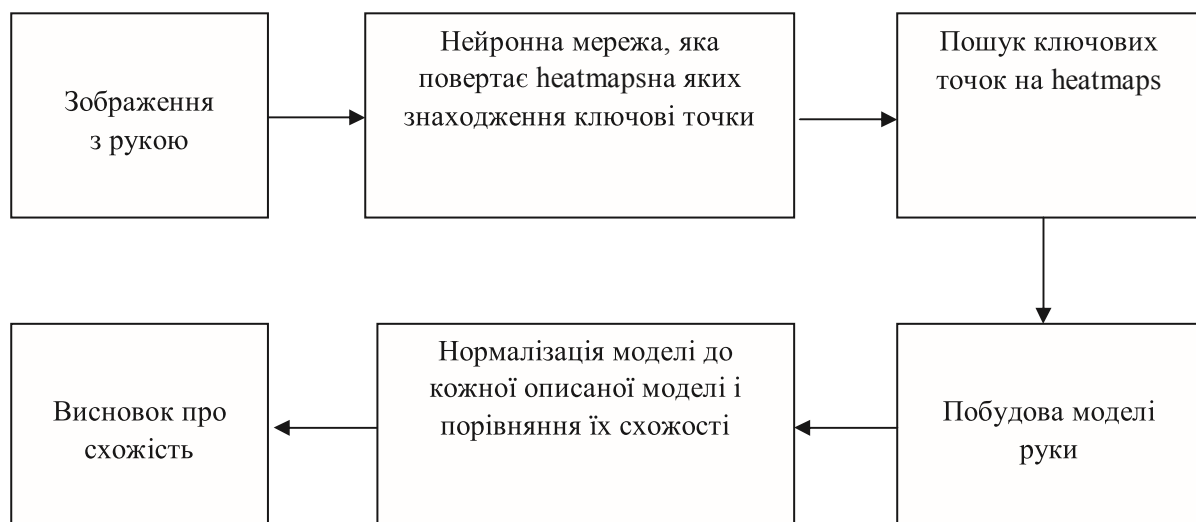


Рис. 5. Алгоритм побудови та розпізнавання моделі руки

Висновки. В даній роботі було проаналізовано структуру та параметри моделі Convolutional Pose Machine(CPM), яка використовується для визначення heatmaps знаходження точок руки з однієї камери. Також було вирішено проблеми аналізу та знаходження ключових точок руки, й подальшого використання для моделювання й розпізнавання жестів скелетону руки. Було порівняно отриманий жест з еталонним та визначено, наша модель руки задовольняє в межах допустимого відхилення заданий жест.

References

1. Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh(2016). Convolutional Pose Machines. arXiv preprint arXiv: 1602.00134,
2. J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler. Efficient object localization using convolutional networks. In CVPR, 2015.

3. Alfredo Canziani, Adam Paszke, Eugenio Culurciello(2017). An Analysis of Deep Neural Network Models for Practical Applications. arXiv preprint arXiv:1605.07678

4. Xingyi Zhou, Qingfu Wan, Wei Zhang, Xiangyang Xue, Yichen Wei(2016). Model-based Deep Hand Pose Estimation. arXiv preprint arXiv:1606.06854

ДОВІДКА ПРО АВТОРІВ

Мухін Вадим Євгенійович - доктор технічних наук, професор кафедри математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Mukhin Vadim Yevgenievich - Doctor of Technical Sciences, Professor of the Department of Mathematical Methods of System Analysis, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”

E-mail: v_mukhin@i.ua

Корнага Ярослав Ігорович - кандидат технічних наук, доцент кафедри технічної кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Kornaga Yaroslav Igorovich - Candidate of Technical Sciences, Associate Professor of the Department of Technical Cybernetics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”

E-mail: slovyan_k@ukr.net

Любівий Андрій Ігорович – студент, кафедри Технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Liubivyi Andrii Igorovich – student, Department of Technical Cybernetics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: 120494al@gmail.com

Vadim Mukhin,
Yaroslav Kornaga,
Andrii Liubivyi

MODELING AND RECOGNITION OF GESTURE OF HAND SKELETON

Target setting. The problem of modeling and recognizing gestures becomes more relevant in recent days due to the growing number of systems that a person can interact with using gestures.

Actual scientific researches and issues analysis. Due to the emergence of new methods and approaches in the field of artificial intelligence, the topic of image recognition has become more relevant in recent years.

Uninvestigated parts of general matters defining. This article is devoted to the study and analysis of the proposed approach for the recognition of models of hands. The research focuses on the study of the use of convolutional neural networks for detecting key points of the hand and constructing a model of the hand and recognizing this model.

The research objective. The purpose of this paper is to investigate a system that will enable, using a convolutional neural network, to detect key points of a hand from one image, and then build a hand model to compare it with the described model gestures.

The statement of basic materials. To detect key points, we use the Convolutional Pose Machines (CPM) architecture. CPM returns the heatmaps of each key point.

CPM has implicit learning of the relationship between the image and the output of several key points at once, the close integration between learning and output and modular sequential design. At each stage of the CPM, images and heatmaps made in the previous step are received as input and are trying to correct and clarify the position of the key points on the hand. After detection of keypoints we construct and compare hand build with described models.

Conclusions. The content, structure and parameters of the Convolutional Pose Machine (CPM) model used to determine heatmaps for finding points of a hand from one camera to solve the problem of finding the key points of the hand, and their further using for modeling and recognition of hand skeleton gestures are analyzed. The approach proved to be good for creating and recognizing hand models.

Keywords: modeling of objects, hand, convolutional neural networks, classification.