

УДК 004.932.2

DOI: 10.20998/2411-0558.2019.13.11

Р. В. СІРЯК, здобув., СНУ ім. В. Даля, Северодонецьк,

І. С. СКАРГА-БАНДУРОВА, д-р техн. наук, проф., зав.каф., СНУ ім. В. Даля, Северодонецьк,

Л. О. ШУМОВА, канд. техн. наук, СНУ ім. В. Даля, Северодонецьк

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ТЕХНОЛОГІЇ ОБРОБКИ ДАНИХ ДЛЯ РОЗПІЗНАВАННЯ ЖЕСТИВ

Розглянуто основні етапи вирішення задачі машинного розпізнавання жестів рук, отриманих з вебкамери. Завдання полягало в отриманні простої, точної та ресурсоефективної системи візуального розпізнавання. В ході роботи були визначені основні етапи підготовки та обробки даних, проведені експерименти зі зміною значень гіперпараметрів, порядком проходження та кількістю шарів згорткової нейронної мережі. В результаті застосування запропонованої технології отримано точність розпізнавання 98.46% зі значенням функції втрати 0.02. Перевагою системи є її низька обчислювальна складність на етапі розпізнавання, що дозволяє в реальному часі класифікувати жести рук. Іл.: 2. Бібліогр.: 12 назв.

Ключові слова: дані; розпізнавання; жест; згорткова нейронна мережа; функції втрати; обчислювальна складність.

Постановка проблеми. Останні десятиліття, розпізнавання жестів рук для реалізації людино-машинної взаємодії є сферою активних досліджень галузі комп'ютерного зору та машинного навчання. Задачею досліджень є створення систем, здатних ідентифікувати жести людини та використовувати їх для передачі інформації або керування пристроями. Застосування цієї технології має дуже великий потенціал. Через зчитування даних про жести можливе маніпулювання об'єктами в середовищі віртуальної реальності. Створення системи, яка розпізнає мову жестів, здатне надати допомогу людям з вадами слуху. Разом з тим, на сьогодні, питання взаємодії через жести все ще не вирішене і вимагає подальших досліджень і вдосконалень. Розпізнавання жестів пов'язане з великою кількістю різноманітних проблем, які виникають кожного разу, коли необхідно розпізнати об'єкт, наприклад, при зміні умов освітлення, неоднорідності фону та ін. Підхід, що використовується для розв'язання проблеми розпізнавання жестів, полягає у виявленні пікселів на зображенні, які складають руку, витягуванні ознак з цих ідентифікованих пікселів для класифікації і використання цих функцій для розпізнавання конкретних послідовностей жестів. Однак, мета розпізнавання полягає не тільки в тому, щоб знайти цільове зображення в реальному часі і відокремити його від фону, а й проаналізувати динамічні просторово-часові характеристики, відстежуючи початок і кінець класифікованого

© Р.В. Сіряк, І.С. Скарга-Бандурова, Л.О. Шумова, 2019

жесту в потоці кадрів.

Аналіз останніх досліджень і публікацій. Залежно від підходів і цілей завдання розпізнавання жестів вирішуються різними методами. Одним з найбільш популярних і ефективних методів розпізнавання жестів рук є згорткові нейронні мережі (CNN). CNN використовується для вилучення, для сегментації зображень [1], розпізнавання дій людини, з урахуванням просторових і часових особливостей [2]. Значний успіх досягнуто Pigou et al. [3], які використовуючи CNN і Microsoft Kinect створили систему розпізнавання італійської мови жестів. Вони досліджували п'ять різних архітектур глибокого навчання і прийшли до висновку, що двонапрямна повторюваність і часова згортка можуть значно покращити класичну класифікацію жестів. Bheda et al. [4] реалізували метод класифікації зображень для американської мови жестів, отримавши 82,5% точності на алфавітних жестах і 97% на цифрах. Підходи до вирішення задачі розпізнавання жестів рук відрізняються один від одного на різних стадіях – починаючи від способу отримання даних і закінчуючи вибором класифікатора. Дані можуть бути отримані через спеціально обладнані рукавички, через закріплення на руці датчиків, через звичайні камери або камери глибини простору. Залежно від способу отримання даних, проводиться їх сегментація. Витяг ознак може проводитися на основі обчислень гістограми орієнтованих градієнтів, Principal Component Analysis (PCA), SIFT або нейронних мереж. Класифікатором може виступати Support Vector Machine, метод Random Forest, алгоритм k -найближчих сусідів, нейронна мережа та ін. Згорткова нейронна мережа, на основі якої реалізований представлений метод розпізнавання жестів рук, має низку переваг. Перш за все, шари згортки в такій мережі здатні будувати ієрархічні уявлення і самостійно виділяти ознаки у вхідних даних. CNN показала найкращі результати при роботі з розпізнаванням зображень і відео, значно випереджаючи інші методи машинного навчання.

Ціль статті – реалізація системи розпізнавання жестів рук на основі згорткової нейронної мережі (CNN) та особливості її реалізації.

Основна частина. Технологія розпізнавання жестів складається з шести етапів: отримання набору даних; збільшення набору (для випадків використання власного відео та/або зображень); попередня підготовка даних; витяг ознак; навчання класифікатора і, безпосередньо, розпізнавання жестів.

Етап 1. Отримання наборів вихідних даних.

Для роботи з жестами рук зазвичай можна використовувати два підходи: формування власного набору або використання відкритих наборів даних. Найбільш відомим відкритим набором є LSA64 [5].

Етап 2. Збільшення даних.

Оскільки навчання мережі і тестування проводилися на власному наборі даних, що містять зображення, отримані з веб-камери, до зібраних даних була застосована операція збільшення даних (data augmentation).

Додаткові зображення генеруються шляхом маніпуляцій з вже наявними зображеннями, і охоплюють операції масштабування, зрушення, деформацію і повороти під різними кутами. Крім збільшення набору даних для навчання, це робить мережу стійкою до спотворень у вхідних даних на стадії розпізнавання, а також додатково допомагає боротися з перенавчанням.

Етап 3. Підготовка та попередня обробка даних.

Попередня обробка даних містить очищення даних, зміну колірної моделі, сегментацію та виділення контуру (розпізнавання країв).



Рис. 1. Підготовка зображення для CNN: (а) початкове зображення, (б) сегментація за кольором шкіри, (с) виділення контурів за Canny

Перед початком процесу сегментації, необхідно очистити зображення від цифрового шуму і знизити зайву деталізацію. Для цього, до кожного зображення з набору даних було застосовано фільтр розмиття за Гауссом (Gaussian Blur) [6]

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

де x – відстань по осі абсцис, y – відстань по осі ординат, σ – стандартне відхилення розподілу Гаусса. Взаємний вплив пікселів визначається як обернено пропорційне квадрату відстані між ними. Ступінь розмиття залежить від параметра стандартного відхилення.

Потім зображення переводяться з колірної моделі RGB в формат HSV, оскільки завдяки тому, що хроматична інформація зберігається в ньому в окремому каналі, стає легше орієнтуватися при сегментації на колір шкіри. Також при використанні HSV знижується проблема з мінливим освітленням.

Сегментація полягає у відокремленні руки від фону. Перш за все, для цього на зображенні виділяється у відповідному діапазоні область кольору шкіри. Таким чином віддаляється задній фон, і подальша робота ведеться тільки з рукою. Зображення перетворюється в одноканальне півтонове для зменшення обчислювальних витрат.

Останньою процедурою підготовки було розпізнавання країв. Розпізнавання країв є важливим моментом в розпізнаванні образів, визначаючи границю між об'єктами або між об'єктом і фоном. До виділеного елемента застосовується детектор країв Canny [7].

Робота детектора Canny складається з наступних кроків:

1. Згладжування зображення шляхом застосування до нього розглянутого вище фільтра Гауссова розмиття.

2. Взяття градієнта зображення, після чого на максимальних значеннях позначаються межі. Для цього використовується оператор Собеля [8], що обчислює значення градієнта яскравості. Оператор Собеля використовує два квадратних ядра згортки, які оцінюють градієнт в горизонтальному і вертикальному напрямках. Після проходження згортки напрямком градієнта обчислюється як:

$$\Theta = \arctan\left(\frac{G_y}{G_x}\right),$$

де G_y і G_x – значення для першої похідної відповідно в горизонтальному і вертикальному положеннях.

3. Придушення немаксимумов. Краями визнаються пікселі, в яких досягається локальний максимум градієнта в напрямку вектора градієнта. Значення кожного пікселя невизнаного максимумом встановлюється в нуль. В результаті виходить тонка лінія контуру.

4. Подвійна порогова фільтрація. Для оцінки того, чи дійсно має місце край в конкретній точці зображення, використовуються два порога. Якщо значення пікселя відноситься вище порога, то край визнається достовірним. Інакше відкидається. Проміжним пікселям присвоюється середнє значення.

5. Гістерзіс (Hysteresis), тобто зв'язування країв в контури. Піксель вище порогового значення T_1 є крайовим пікселем. Пікселі, які межують з крайовим пікселем, і при цьому мають значення вище, ніж T_2 також відносяться до групи крайових пікселів.

В результаті виконання даного етапу отримується набір зображень, готових для завантаження до нейронної мережі. Результуюче зображення, підготовлене до розпізнавання, представлено на рис. 1 с).

Етап 4. Витяг ознак.

На відміну від інших методів машинного навчання, згорткова нейронна мережа не вимагає ручної розробки набору ознак. Ознаки витягуються мережею самостійно в згортальних шарах. Розроблена мережа має три згорткових шари з функцією активації ReLU, за кожним з яких розташовується шар maxpooling (рис. 2). Умовно, створену CNN можна розділити на дві частини: а) три блоки згорткових і subsampling шарів, які чергуються один з одним і формують вхідний вектор ознак для навчання; б) три повнозв'язних шари. Згорткові шари, формують 16, 32 і 64 карти ознак, застосовуючи ядра згортки розміром 3×3. Карти ознак проходять через шари subsampling з max-pooling 2×2, кожен раз зменшуючи розмірність даних вдвічі.

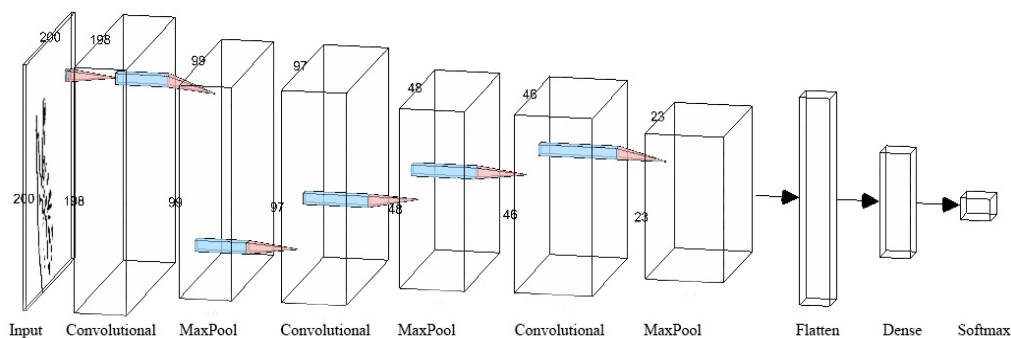


Рис. 2. Архітектура нейронної мережі

Витяг ознак здійснюється шляхом операції згортки квадратною матрицею з непарною розмірністю. Матриця, яка називається ядром згортки або фільтром, ковзає по зображенню, виконуючи покрокове скалярне перетворення. На виході є feature maps, кількість яких залежить від числа застосовуваних фільтрів. Кожен фільтр являє собою систему поділюваних ваг і призначений для пошуку і виділення ознак за певним шаблоном.

Згортка представлена як

$$Conv(\omega \times y)_{ij} = \sigma \left(\sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} \times y_{(i+a)(j+b)}^{l-1} \right),$$

де ω – ядро згортки розміру $m \times m$, y – входи з попереднього шару, σ – функція активації нейронів.

За кожним згортковим шаром слідує шар max-pool, що зменшує розмірність карти ознак, проходячи по зображенню і покровоно вибираючи з рецептивного поля розміром 2×2 максимальне значення. Крім зниження розмірності в два рази і зменшення кількості параметрів в нейронної мережі, max-pooling робить знайдені ознаки більш яскраво вираженими, а мережа стає більш інваріантною до місцезнаходження об'єкта на карті ознак, до зрушень і поворотів. Вихід max-pool буде обчислюватися через формулу:

$$y^{l+1} = \max_{0 \leq i \leq H, 0 \leq j \leq W} x_{i^{l+1} \times H+i, j^{l+1} \times j, d}^l$$

де H, W – розмір вікна субдискретизації, x – вихідні дані.

Етап 5. Навчання класифікатора.

Частина мережі, що безпосередньо виконує класифікацію, складається з шару flatten, кожен вузол якого відповідає одному значенню з вектора ознак, і двох повнозв'язних шарів dense. Останній шар є вихідним і реалізує функцію softmax. Дані з останнього шару субдискретизації надходять на шар Flatten, перетворюючись у ньому в одновимірний вектор. Обчислення значень нейронів для повнозв'язного шару відбувається за формулою:

$$x_i^l = \sum_{k=0}^m w_{ki}^l y_k^{l-1} + b_i^l,$$

де w_{ki}^l – вага від k -го нейрона шару $l - 1$ до i -го нейрону поточного шару l ; b – зміщення поточного шару; y_k^{l-1} – вхідні дані з попереднього шару.

Останній шар мережі з кількістю виходів що дорівнює кількості категорій що розпізнаються, реалізує функцію активації softmax. Softmax привласнює значення, представлене невід'ємним дійсним числом, кожному класу, відображаючи ймовірність приналежності. Сума всіх вихідних сигналів дорівнює одиниці. Значення вихідного сигналу i -го нейрона відповідає ймовірності того, що правильна відповідь є i . Значення i -го виходу в softmax визначається за формулою:

$$S_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}.$$

Етап 6. Розпізнавання.

На етапі розпізнавання навчена нейронна мережа через веб-камеру в режимі реального часу розпізнає клас одержуваного на вхід жесту.

Функцією активації всюди, крім останнього шару, виступала Rectified Linear Unit (ReLU)

$$f(x) = \max(0, x).$$

ReLU значно перевершує інші функції в стійкості до загасання градієнта, а швидкість навчання згідно [9] в порівнянні з гіперболічним тангенсом швидше в шість разів.

Для регуляризації моделі в мережі двічі застосовувалася техніка dropout [10] з коефіцієнтом 0.25. Dropout полягає в тому, щоб на кожну епоху з заданою вірогідністю p виключати нейрони з мережі, в результаті чого він повертає значення 0.

У якості функції втрат була використана категорійна кросс-ентропія – розрахована логарифмічна втрата для декількох представлених класів. Крос-ентропія між розподілами p і q визначається наступним чином:

$$H(p, q) = H(p) + D_{KL}(p||q),$$

де $H(p)$ – ентропія p , $D_{KL}(p||q)$ – розбіжність Кулбака-Лейблера [11] для q з p (відносна ентропія p до q).

Якщо прогнозовані значення моделі дорівнюють q , тоді як справжні значення дорівнюють p , то категорійна перехресна ентропія буде виглядати так:

$$H(y, \hat{y}) = -\sum_i y_i \log \hat{y}_i = -y \log \hat{y} - (1-y) \log(1-\hat{y}).$$

У якості методу оптимізації використано Adam (adaptive moment estimation) [12]. Adam використовує як середні значення градієнта, так і другий імпульс градієнта, що заважає йому застрягти в місцевих мінімумах. Далі йде формула, в якій m_t обчислює перший імпульс, а v_t – квадрат:

$$m_t = \beta_1 m_t + (1 - \beta_1) g_t,$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2,$$

де β_1 – коефіцієнт для зменшення ковзного середнього градієнта, β_2 є коефіцієнтом для зменшення ковзного середнього площі градієнта. Перерахунок параметрів виконується за формулою

$$w_p = w_{p-1} - \frac{\eta}{\sqrt{v_p + \varepsilon}} m_p,$$

де $\varepsilon = 10^{-8}$ вводиться для запобігання можливого ділення на нуль.

Результати експерименту.

Набір складався з 7200 зображень розміром 120×120 , об'єднаних в умовні категорії "one", "two", "zero", "palm", "fist", "letter SH" по 1200 зразків кожен. Для полегшення процедур навчання та обробки використовувався білий фон. Дані були зібрані у чотирьох осіб при штучному розсіяному світлі.

Експериментальні дослідження були проведені на комп'ютері з процесором Intel Core i3-7100, 8 Гб оперативної пам'яті, NVIDIA GeForce GTX 1060. Програмне забезпечення написано мовою програмування Python. Були використані бібліотеки кадрів Tensorflow і Keras, які використовують бібліотеки numpy, scikit-learn, OpenCV. В якості навчального набору було створено власну базу зображень ручного жесту. Зйомка руки проводилася з різних кутів огляду.

Процес як отримання даних, так і розпізнавання проводився за допомогою веб-камери на відстані близько 1 метра.

Весь набір даних був розділений на навчальний – на якому власне і проходило навчання, валідаційний – призначений для контролю ходу навчання, і тестовий – призначений для перевірки вже навченої мережі. Розподіл набору даних на підмножини складав 70, 10 і 20 відсотків відповідно.

Навчання нейронної мережі тривало 50 епох. За одну епоху оброблялися всі підмножини даних адаптації та валідації. Для перевірки якості використовувалася метрика асигасу, що показує співвідношення кількості правильно передбачених значень до всієї кількості виданих мережею пророкувань.

Функцією втрат виступала категоріальна перехресна ентропія, яка вираховувала логарифмічну втрату на кілька представлених класів. В результаті навчання мережі була отримана точність на тестовому підмножині в 98.46%, а значення функції втрат склало – 0.02.

Висновки. Метою проекту було створення простої, дешевої і ресурсномісткої системи розпізнавання жестів рук. Була розроблена і навчена модель згорткової нейронної мережі, здатна з високою ймовірністю розпізнавати в відеопотоці шість тестових жестів. Розроблена модель багато в чому дотримується принципів, які застосовуються при використанні згорткової нейронної мережі як класифікатора жестів. При необхідності, мережу можна навчити на більшій кількості жестів. Перевагою запропонованої системи є її низька обчислювальна складність на етапі розпізнавання, що дозволяє в реальному часі класифікувати жести рук. Серед недоліків можна відзначити, що модель не працює добре на неоднорідному, змінюваному фоні, окрім того, жести рук людей, які не брали участь у створенні

набору даних, визнаються гірше. Розроблена модель багато в чому слідує принципам, необхідним для CNN як класифікатора жестів. Її точність розпізнавання з власним набором даних складає 98.46%, хоча є й деякі недоліки, згадані вище.

В майбутньому планується збільшити кількість видів жестів, додати періодичний блок для розпізнавання більш складних жестів. Планується також розробити засоби боротьби з гетерогенним фоном.

Список літератури:

1. *Tompson J.* Real-time continuous pose recovery of human hands using convolutional networks / *J. Tompson* // *ACM Transactions on Graphics (ToG)*. – 2014. – Vol. 33. – № 5. – P. 169-175.
2. *Ji S.* 3D convolutional neural networks for human action recognition / *S. Ji* // In *IEEE Transactions on Pattern Analysis And Machine Intelligence*. – 2013. – Vol. 35. – № 1. – P. 221-231.
3. *Yamato J.* Recognizing human action in time-sequential images using hidden Markov model / *J. Yamato, J. Ohya, K. Ishii* // In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR'92*. – 1992.
4. *Bheda V.* Using deep convolutional networks for gesture recognition in American sign language / *V. Bheda, D. Radpour* // In: *CoRR*, abs/1710.06836. – 2017.
5. A dataset for Argentinian Sign Language LSA64 <http://facundoq.github.io/unlp/lsa64/>
6. *Nixon Mark.* Aguado / *Mark Nixon, S. Alberto* // *Feature extraction and image processing for computer vision*. – Academic Press, 2012.
7. *Canny John.* A computational approach to edge detection / *John Canny* // *Readings in computer vision*. – Morgan Kaufmann. – 1987. – P. 184-203.
8. *Sobel Irwin.* A 3x3 isotropic gradient operator for image processing / *Irwin Sobel, Feldman Gary* // *A talk at the Stanford Artificial Project in* – 1968. – P. 271-272.
9. *Krizhevsky Alex.* Imagenet classification with deep convolutional neural networks / *Alex Krizhevsky, Sutskever Ilya, E. Hinton Geoffrey* // *Advances in neural information processing systems*. – 2012.
10. *Hinton Geoffrey E.* Improving neural networks by preventing co-adaptation of feature detectors / *Geoffrey E. Hinton* // *arxiv preprint arxiv:1207.0580*. – 2012.
11. *Kullback Solomon.* On information and sufficiency / *Solomon Kullback, Leibler Richard* // *The annals of mathematical statistics*. – Vol. 22. – № 1. – 1951. – P. 79-86.
12. *Kingma, Diederik P. Jimmy Ba.* Adam: A method for stochastic optimization / *Diederik P. Kingma, Jimmy Ba* // *arxiv preprint arxiv:1412.6980*. – 2014.

References:

1. Tompson, J. (2014), "Real-time continuous pose recovery of human hands using convolutional networks", *ACM Transactions on Graphics (ToG)*, vol. 33, No. 5, pp. 169-175.
2. Ji, S. (2013), "3D convolutional neural networks for human action recognition", In *IEEE Transactions on Pattern Analysis And Machine Intelligence*, vol. 35, No. 1, pp. 221-231.
3. Yamato, J., Ohya J., and Ishii K. (1992), "Recognizing human action in time-sequential images using hidden Markov model", In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR'92*.
4. Bheda, V., and Radpour, D. (2017), "Using deep convolutional networks for gesture recognition in American sign language", In: *CoRR*, abs/1710.06836.
5. A dataset for Argentinian Sign Language LSA64 <http://facundoq.github.io/unlp/lsa64/>

6. Nixon, Mark, and Alberto S. Aguado (2012), *Feature extraction and image processing for computer vision*, Academic Press.
7. Canny, John (1987), "A computational approach to edge detection", *Readings in computer vision*, Morgan Kaufmann, pp. 184-203.
8. Sobel, Irwin, and Gary Feldman (1968), "A 3x3 isotropic gradient operator for image processing", *a talk at the Stanford Artificial Project*, pp. 271-272.
9. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012), "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems*.
10. Hinton, Geoffrey E. (2012), "Improving neural networks by preventing co-adaptation of feature detectors", *arXiv preprint arXiv:1207.0580*.
11. Kullback, Solomon, and Richard, A. Leibler (1951), "On information and sufficiency", *The annals of mathematical statistics*, vol. 22, No. 1, pp. 79-86.
12. Kingma, Diederik P., and Jimmy Ba. (2014), "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*.

Статью представил д.т.н., проф. Національного технічного університету "Харківський політехнічний інститут" С.Ю. Леонов

Надійшла (received) 01.01.2019

Siryak Rostislav, Ph.D. student
Volodymyr Dahl East Ukrainian National University
59-a Central Avenue, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: hashem.r@gmail.com
ORCID ID: 0000-0002-6775-1218

Skarga-Bandurova Inna, D.Sci.Tech., Professor
Volodymyr Dahl East Ukrainian National University
59-a Central Avenue, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: skarga_bandurova@ukr.net
ORCID ID: 0000-0003-3458-8730

УДК 004.932.2

Особенности реализации технологии обработки данных для распознавания жестів / Сіряк Р.В., Скарга-Бандурова І.С., Шумова Л.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2019. – № 1. – С. 90 – 100.

Розглянуто основні етапи вирішення задачі машинного розпізнавання жестів рук, отриманих з вебкамери. Завдання полягало в отриманні простої, точної та ресурсоефективної системи візуального розпізнавання. В ході роботи були визначені основні етапи підготовки та обробки даних, проведені експерименти зі зміною значень гіперпараметрів, порядком проходження та кількістю шарів згорткової нейронної мережі. В результаті застосування пропонованої технології отримано точність розпізнавання 98.46%, зі значенням функції втрати 0.02. Перевагою системи є її низька обчислювальна складність на етапі розпізнавання, що дозволяє в реальному часі класифікувати жести рук. Іл.: 2. Бібліогр.: 12 назв.

Ключові слова: дані, розпізнавання; жест; згорткова нейронна мережа; обчислювальна складність.

УДК 004.932.2

Особенности реализации технологии обработки данных для распознавания жестов / Сиряк Р.В., Скарга-Бандурова И.С., Шумова Л.О. // Вестник НТУ "ХПИ". Серія: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2019. – № 1. – С. 90 – 100.

Рассмотрены основные этапы решения задачи машинного распознавания жестов рук, полученных с вебкамеры. Задача состояла в получении простой, точной и ресурсоэффективной системы визуального распознавания. В ходе работы были определены основные этапы подготовки и обработки данных, проведены эксперименты с изменением значений гиперпараметров, порядком прохождения и количеством слоев сверточной нейронной сети. В результате применения предлагаемой технологии получена точность распознавания 98.46%, со значением функции потерь 0.02. Преимуществом системы является ее низкая вычислительная сложность на этапе распознавания, что позволяет классифицировать жесты рук в реальном времени. Ил.: 2. Библиогр.: 12 назв.

Ключевые слова: данные; распознавание; жест; сверточная нейронная сеть; вычислительная сложность.

UDC 004.932.2

Special Considerations for the Implementation of Data Processing Technique for Gesture Recognition / Siriak R.V., Skarga-Bandurova I.S., Shumova L.O. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2019. – № 1. – С. 90 – 100.

The main stages of solving the problem of machine recognition of hand gestures received from a webcam are considered. The task was to obtain a simple, accurate, and resource-efficient visual recognition technique. The main stages of data preparation and processing were determined; experiments were carried out with different values of hyperparameters, the order, and the number of layers of a convolutional neural network. As a result, we obtained the recognition accuracy of 98.46% with a loss function value of 0.02. This technique has a low computational complexity at the recognition stage and makes it possible to classify hand gestures in real time. Figs.: 2. Refs.: 12 titles.

Keywords: data; recognition; gesture; convolutional neural network; computational complexity.