УДК 004.056.5

Терейковська Л.О.
Київський національний університет будівництва і архітектури

Терейковський О.І.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

НЕЙРОМЕРЕЖЕВА МОДЕЛЬ РОЗПІЗНАВАННЯ ЕМОЦІЙ ПО ЗОБРАЖЕННЮ ОБЛИЧЧЯ

Стаття присвячена розробленню нейромережевої моделі, що дає змогу реалізувати розпізнавання емоцій на підставі зображення обличчя людини під час впливу завад, характерних для інформаційних систем загального призначення. Визначена низька пристосованість поширених рішень на базі згорткових нейронних мереж до нівелювання такої характерної завади, як поворот обличчя на зображенні, що підлягає аналізуванню. Запропоновано виправити вказаний недолік за рахунок застосування капсульної нейронної мережі, що є розвитком згорткових нейронних мереж щодо розпізнавання зашумлених зображень. Розроблено відповідну нейромережеву модель, ефективність якої доведена експериментально. Показано, що перспективи подальших досліджень щодо нейромережевого розпізнавання емоцій по геометрії обличчя можуть бути пов'язані з удосконаленням архітектурних рішень капсульної нейронної мережі задля зменшення кількості навчальних ітерацій під час забезпечення прийнятної похибки розпізнавання.

Ключові слова: *емоційний стан, базові емоції, зображення обличчя, нейромережева модель, капсульна нейронна мережа.*

Постановка проблеми. Сьогодні засоби автоматичного розпізнавання емоцій на базі геометрії обличчя людини знайшли широке застосування в різноманітних сферах людської діяльності. До переваг вказаних засобів відносять порівняно високу точність класифікації, апробованість, низьку вартість та поширеність зчитувальних пристроїв (відеокамер). Водночас практичний досвід і результати науково-прикладних робіт [2–4] вказують на необхідність суттєвої модернізації сучасних засобів розпізнавання задля зменшення ресурсоємності, збільшення точності розпізнавання, скорочення терміну розроблення та підвищення адаптації до багатьох особливостей сучасних інформаційних систем (ІС). Разом з використанням більш ефективного апаратного забезпечення одним з основних напрямів модернізації є вдосконалення математичного забезпечення процесу розпізнавання [3; 5–8], що зумовлює актуальність досліджень у цьому напрямі.

Аналіз останніх досліджень і публікацій. У сучасній психології виділяють сім базових емоцій, мімічні прояви яких не залежать від раси та культури людини [1; 4]. Базові емоції – це елементарні емоції, які більше ні на що не розщеплюються й самі є складовими інших складних емоцій. До базових емоцій належать радість, гнів, відраза, смуток, страх, подив, презирство та нейтральність. Зазначимо, що в деяких роботах [5] виділяють не вісім, а чотири базових емоції, таких як щастя (радість), смуток (сум), переляк (страх), здивування, гнів/огида. При цьому кількість та номенклатура складних (складових) емоцій досі чітко не визначена, а відповідна теорія далека від досконалості.

Базою аналізу поточного стану розробок щодо розпізнавання емоцій на основі геометрії обличчя людини послужили роботи [1–6; 8–11], в яких описані як апробовані рішення, так і сучасні підходи до цього. Так, у джерелі [1] запропоновано варіант реалізації системи розпізнавання емоційного стану для підтримки спілкування людини із сервісними антропоморфними роботами. Зазначено, що суттєвою перешкодою під час розроблення систем розпізнавання емоцій є обмеженість доступних баз даних, а також висока частка індивідуальних особливостей у прояві тієї чи іншої емоції у різних людей. Це істотно підвищує вимоги до узагальнюючих можливостей застосовуваних методів машинного навчання. Крім цього, на точність розпізнавання суттєво впливають зміна положення обличчя на зображенні, наявність окулярів та макіяжу, зачіска. Для нівелювання цих труднощів пропонується використовувати різні алгоритми локальної фільтрації зображення під час визначення інформативних ознак особи, а оцінювання ступеня

вираження емоцій можна розраховувати за допомогою мультикласифікатора. Розроблена у джерелі [5] система дає змогу розпізнавати сім базових емоцій на підставі відфільтрованих локальних ознак ступеня вираження двадцяти ознак (AU), що входять у систему кодування FACS, розроблену П. Екманом. Для розпізнавання використовуються три класифікатори, а саме ймовірнісний класифікатор, багатошаровий персептрон та система логічних правил. Підсумковий ступінь вираження емоції розраховується як сума відгуків ймовірнісного класифікатора та нейронної мережі (НМ). Логічні правила використовуються тільки для вирішення спірних ситуацій, коли кілька емоцій отримують однакову високу оцінку ступеня вираження. Задекларована точність розпізнавання складає 85%. У джерелах [2; 3] комбінуються кілька типів ознак із подальшою класифікацією методом опорних векторів. Автори [11] декларують можливість значного підвищення точності розпізнавання за рахунок використання як ознак просторово-часової модифікації локальних бінарних шаблонів. В роботі [3] продемонстровано алгоритм розрахунку інтенсивності AU і зіставлення ефективності різних груп ознак та їх об'єднань. Цікавий підхід до класифікації запропонований у праці [4], де ступені вираження AU перетворюються на маркери наявності емоцій за допомогою логічних дерев рішень, специфічних для різних етнічних груп.

Також проведений аналіз дає змогу стверджувати, що сьогодні найбільш ефективними вважаються нейромережеві засоби розпізнавання емоцій. При цьому НМ можуть використовуватись для розпізнавання емоцій на підставі як аналізу характерних точок обличчя, так і цілісного порівняння зображення обличчя особи з деякими еталонами. Як правило, базою сучасних рішень є згорткові нейронні мережі (ЗНМ) різної архітектури. Інші типи НМ менш ефективні як щодо точності розпізнавання, так і стосовно ресурсоємності. Крім ЗНМ, у засобах розпізнавання для врахування часової складової використовуються рекурентні НМ типу LSTM [4]. Водночас дані джерел [6; 8; 11] вказують на необхідність підвищення рівня адаптації сучасних НМ розпізнавання емоцій до типових завад, що виникають під час розпізнавання обличчя в ІС загального призначення.

Постановка завдання. Метою статті є розроблення нейромережевої моделі, що дає змогу реалізувати розпізнавання емоцій на підставі геометрії обличчя людини під час впливу завад, характерних для інформаційних систем загального призначення.

Виклад основного матеріалу дослідження. Проведемо деяке уточнення постановки завдання розпізнавання емоцій. Припустимо необхідність розпізнавання тільки базових емоцій на підставі статичних зображень обличчя, зафіксованого за допомогою відеокамери із середніми характеристиками. При цьому питання розпізнавання особистості, попередньої фільтрації зображення, впливу освітленості, виділення на зображенні окремих осіб та нівелювання свідомого спотворення обличчя людини задля приховування нею свого емоційного стану не розглядаються. Результати [2; 5] вказують на те, що в цьому разі в ІС загального призначення основні перешкоди виникають в результаті повороту обличчя людини. Відповідно до джерел [7; 11] нівелювати вказану перешкоду можна за рахунок використання моделі на базі капсульної НМ (CapsNet), яка є модифікацією ЗНМ, пристосованої до аналізування повернутих та зашумлених зображень. Структура мережі CapsNet, що адаптована до завдання розпізнавання емоцій, показана на рис. 1.

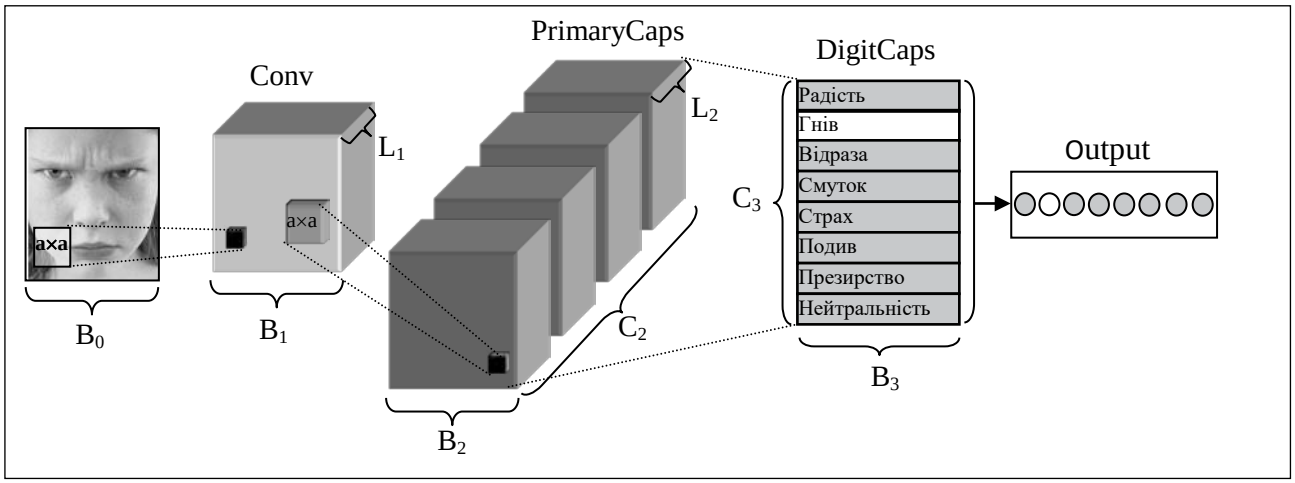

Рис. 1. Структура CapsNet

Основними структурними одиницями CapsNet є вхідний шар нейронів (відповідає вхідному зображенню), згортковий шар (Conv), шар первинних капсул (PrimaryCaps), шар згорткових капсул (DigitCaps), кожна з яких відповідає одній з базових емоцій. Зазначимо, що, на відміну від класичної капсульної мережі [11], у CapsNet, показаній на рис. 1, блок декодування відсутній. На рис. 1 використані такі позначення, які відповідають структурним параметрам мережі: B_0 – вертикальний та горизонтальний розмір аналізованого зображення; B_1 – розмір карт ознак у шарі Conv; B_2 – розмір сітки в шарі PrimaryCaps; B_3 – кількість одиниць згортки в кожній капсулі шару DigitCaps; C_2 – кількість каналів у шарі PrimaryCaps; C_3 – кількість капсул у шарі DigitCaps; L_1 – кількість карт ознак у шарі Conv; L_2 – кількість одиниць згортки в кожному каналі шару PrimaryCaps; a – розмір ядра згортки. Аналогічно до класичної капсульної мережі розмір ядра згортки a = 9, крок ядра згортки для Conv $d_1 = 1$, крок ядра згортки для PrimaryCaps $d_2 = 2$, кількість карт ознак $L_1 = 256$, кількість каналів PrimaryCaps $C_2 = 32$, кількість одиниць згортки в PrimaryCaps $L_2 = 8$, кількість одиниць згортки в кожній капсулі DigitCaps $B_3 = 8$. Розмір вхідного зображення $B_0 = 48$, розмір карт ознак $B_1 = 41$, розмір сітки $B_2 = 17$, кількість капсул у DigitCaps дорівнює кількості емоцій, що мають бути розпізнані, $C_3 = 8$.

Розрахунок вхідних та вихідних сигналів для нейронів у Conv здійснюється так само, як і в згортковій нейронній мережі. При цьому використовується функція активації ReLU:

$$y = \max(0, x), \tag{1}$$

де x – сумарний вхідний сигнал нейрона; y – вихідний сигнал.

Вхідний сигнал нейрона в шарі Conv розраховується так:

$$x_k^{(i,j)} = x_{0,k} + \sum_{s=1}^{a}\sum_{t=1}^{a} w_{k,s,t} x^{(i+s,j+t)}, \tag{2}$$

де $x_k^{(i,j)}$ – вхідний сигнал (i,j)-го нейрона k-ї карти ознак; $x_{0,k}$ – зсув нейронів k-ї карти ознак; a – розмір ядра згортки; $w_{k,s,t}$ – ваговий коефіцієнт (s,t)-го зв'язку нейрона k-ї карти ознак; $x_0^{(i,j)}$ – вхідний сигнал (i,j)-го нейрона вхідного шару.

Розрахунок вхідних та вихідних сигналів капсул реалізується так:

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2}\frac{s_j}{\|s_j\|}, \tag{3}$$

$$s_j = \sum_{i=1}^{C_3}\left(c_{i,j}\hat{u}_{j|i}\right), \tag{4}$$

$$\hat{u}_{j|i} = W_{i,j}u_i, \tag{5}$$

$$c_{i,j} = \exp\left(b_{i,j}\right)\Big/\sum_{k=1}^{C_2}\exp\left(b_{i,j}\right), \tag{6}$$

$$\Delta b_{i,j} = v_j\hat{u}_{j|i}, \tag{7}$$

$$b_{i,j} = b_{i,j} + \Delta b_{i,j}, \tag{8}$$

де v_j – вхідний вектор j-ї капсули в шарі DigitCaps; s_j – складова j-ї капсули шару DigitCaps у вихідному сигналі мережі; $c_{i,j}$ – ваговий коефіцієнт узгодженості між i-ю капсулою в шарі PrimaryCaps та j-ю капсулою в шарі DigitCaps; $\hat{u}_{j|i}$ – прогнозована величина вихідного сигналу i-ї капсули в шарі PrimaryCaps; $W_{i,j}$ – матриця вагових коефіцієнтів зв'язків між i-ю капсулою в шарі PrimaryCaps та j-ю капсулою в шарі DigitCaps; $b_{i,j}$ – логарифм ймовірності зв'язку між i-ю капсулою в шарі PrimaryCaps та j-ю капсулою в шарі DigitCaps; u_i – вихідний сигнал i-ї капсули в шарі PrimaryCaps; $\varnothing b_{i,j}$ – коефіцієнт корекції при ітеративному розрахунку $b_{i,j}$.

Відзначимо, що вираз виду (3) є так званою squash-функцією, а вираз виду (6) – функцією softmax. Процес навчання CapsNet полягає в мінімізації функціоналу:

$$\sum_{k=1}^{K} E_k \to \min, \tag{9}$$

$$E_k = T_k \max\left(0, m^+ - \|v_k\|\right)^2 + \lambda\left(1 - T_k\right)\max\left(0, \|v_k\| - m^-\right)^2, \tag{10}$$

де $K = C_3$ – кількість капсул у шарі PrimaryCaps.

При цьому для капсули, яка відповідає k-й емоції, $T_k = 1$ тільки тоді, якщо на зображенні обличчя ця емоція відображається. Інакше $T_k = 0$. Значення інших коефіцієнтів $m^+ = 0,9$, $m^- = 0,1$, $\lambda = 0,5$. Для навчання та розпізнавання CapsNet пропонується використовувати алгоритми, докладно описані в джерелах [7; 11].

Особливістю навчання CapsNet є застосування алгоритму динамічної маршрутизації між капсулами для розрахунку коефіцієнтів c_{ij}. Псевдокод процедури маршрутизації має такий вигляд:

procedure Routing($\hat{u}_{j|i}$; r; l)
for all capsule i in layer l and capsule j in layer $(l + 1)$: $b_{ij} \leftarrow 0$
for r iterations do
for all capsule i in layer l: $c_i \leftarrow \text{softmax}(b_i)$
for all capsule j in layer $(l + 1)$: $s_j \leftarrow \sum_i c_{ij}\hat{u}_{j|i}$
for all capsule j in layer $(l + 1)$: vj $\leftarrow$squash(s_j).
for all capsule i in layer l and capsule j in layer $(l + 1)$: $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j$
return v_j

Для формування навчальної та тестової вибірки використані бази даних Fer2013-images, emotion_analysis, modified_fer2014, FER2018, доступні на сайті www.kaggle.com. Сформовані з використанням вказаних баз даних навчальні приклади

є jpg-файлами з фотографіями обличь людей, що виражають сім базових емоцій та нейтральний стан.

Приблизно на половині навчальних прикладів фотографії обличчя зафіксовано у фронтальній проекції, на іншій половині – в повернутому стані. Кут повороту становить від -45° до +45°. Кожна фотографія є зображенням 48×48 пікселів у відтінках сірого. Обсяг навчальної та тестової вибірки становить 40 000 та 40 00 прикладів відповідно. Приклади зображень наведено на рис. 2.

На першому етапі досліджень проведено експерименти з розпізнавання емоцій на основі фронтального, добре освітленого зображення обличчя. Другий етап досліджень пов'язаний з розпізнаванням емоцій на зображеннях повернутого обличчя. Експерименти реалізовані за допомогою розробленої комп'ютерної програми, яка базувалась на математичному апараті, заданому виразами (1–10). Отримані значення середньої похибки розпізнавання базових емоцій наведено в табл. 1.

Для порівняння в табл. 1 також наведено дані [1; 2] щодо похибки розпізнавання за допомогою класичної ЗНМ LaNet, що є однією з найбільш сучасних модифікацій ЗНМ VGG, двошарового персептрону (MLP) та ймовірнісного класифікатора на базі НМ типу PNN.

Дані табл. 1 свідчать про те, що похибка розпізнавання фронтальних зображень CapsNet нижче похибки LaNet, MLP та PNN та вище похибки VGG. При цьому похибка розпізнавання емоцій на повернутих зображеннях за допомогою CapsNet порівнянна з похибкою сучасних типів ЗНМ та значно менше похибки LaNet, MLP та PNN. Також можна стверджувати, що точність розпізнавання CapsNet повернутих зображень приблизно на 5% нижче, ніж фронтальних. В разі використання LaNet і VGG точність погіршується приблизно на 9%. Крім цього, експерименти показали, що час навчання CapsNet значно перевищує час навчання інших НМ. Водночас ресурсоємність CapsNet набагато нижче ресурсоємності мережі VGG.

Висновки. В результаті проведених досліджень розроблена нейромережева модель типу CapsNet, призначена для розпізнавання базових емоцій з урахуванням повороту обличчя, яка є характерною завадою розпізнавання для інформаційних систем загального призначення. Експериментальним шляхом показано, що під час розпізнавання емоцій на фронтальних зображеннях похибка CapsNet значно менше, ніж похибка класичних нейромережевих моделей типу LaNet, MLP та PNN. Однак LaNet MLP та PNN перевершують CapsNet щодо кількості навчальних ітерацій, необхідних для досягнення прийнятної помилки навчання. Похибка розпізнавання CapsNet фронтальних зображень дещо більше за похибку сучасних типів згорткових нейронних мереж, що мають порівняно з нею значно вищу ресурсоємність. Під час розпізнавання емоцій на повернутих зображеннях обличчя похибка CapsNet незначно відрізняється від похибки сучасних типів згорткових мереж, а також є значно меншою, ніж похибка класичних нейронних мереж. Перспективи подальших досліджень

Таблиця 1

Середня похибка розпізнавання базових емоцій

Ракурс зображення обличчя	Тип нейромережевої моделі				
	CapsNet	**LaNet**	**VGG**	**MLP**	**PNN**
Фронтальний	14,7	21,7	7,3	25,5	23,7
Повернутий	18,8	28,6	12,6	32,9	34,1

Рис. 2. Приклади фотографій навчальної вибірки

щодо нейромережевого розпізнавання емоцій по геометрії обличчя можуть бути пов'язані з удосконаленням архітектурних рішень капсульної нейронної мережі стосовно зменшення кількості навчальних ітерацій під час забезпечення прийнятної похибки розпізнавання. Крім цього, підвищення ефективності системи нейромережевого розпізнавання емоцій пов'язане із забезпеченням класифікації розмитого й частково прихованого зображення обличчя.

Список літератури:

1. Бобе А.С., Конышев Д.В., Воротников С.А. Система распознавания базовых эмоций на основе анализа двигательных единиц лица. *Инженерный журнал: наука и инновации*. 2016. Вып. 9. С. 1–16.

2. Anderson K., McOwan P. A real time automated system for the recognition of human facial expressions. *Systems, man, and cybernetics*. *IEEE Transactions*. 2006. Vol. 36. P. 96–105.

3. Batista J.C., Albiero V., Bellon O.R., Silva L. Aumpnet: simultaneous action unit's detection and intensity estimation on multipose facial images using a single convolutional neural network. *In Automatic Face & Gesture Recognition. 12th IEEE International Conference.* 2017. P. 866–871.

4. Chandrani S., Washef A., Soma M., Debasis M. Facial expressions: a cross-cultural study. Emotion recognition: a pattern analysis approach. *Wiley Publ*. 2015. P. 69–86.

5. Ertugrul O., Jeni L.A., Cohn J.F. Facscaps: pose-independent facial action coding with capsules. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (Salt Lake City). 2018. P. 221101–221109.

6. Ghosh S., Laksana E., Scherer S., Morency L.-P. A multi-label convolutional neural network approach to cross domain action unit detection. *Affective computing and intelligent interaction: international conference*. 2015. P 609–615.

7. Hinton G., Sabour S., Frosst N. Matrix capsules with EM routing. *ICLR Conference. Vancouver Convention Center* (Vancouver, BC, Canada). 2018.

8. Multi view facial action unit detection based on CNN and BLSTMRNN / J. He, D. Li, B. Yang, S. Cao, B. Sun, L. Yu. *Automatic Face & Gesture Recognition. 12th IEEE International Conference*. 2017. P. 848–853.

9. Ilbeygi, M., Shah-Hosseini, H. A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Engineering applications of artificial intelligence*. 2012. P. 130–146.

10. Sabour S., Frosst N., Hinton G. Dynamic Routing Between Capsules. *Advances in Neural Information Processing Systems*. 2017. P. 3857–3867.

11. Emotion recognition from an ensemble of features / U. Tariq, K. Lin, Z. Li, Z. Zhou, Z. Wang, V. Le, T.S. Huang, X. Lv, T.X. Han. *Systems, man, and cybernetics*. *IEEE Transactions*. 2012. Vol. 42. P. 17–26.

НЕЙРОСЕТЕВАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО ИЗОБРАЖЕНИЮ ЛИЦА

Статья посвящена разработке нейросетевой модели, позволяющей реализовать распознавания эмоций на основе изображения лица человека при воздействии помех, характерных для информационных систем общего назначения. Определена низкая приспособленность распространенных решений на базе сверточных нейронных сетей к нивелированию такой характерной помехи, как поворот лица на анализируемом изображении. Предложено исправить указанный недостаток за счет применения капсульной нейронной сети, являющейся развитием сверточных нейронных сетей касательно распознавания зашумленных изображений. Разработана соответствующая нейросетевая модель, эффективность которой доказана экспериментально. Показано, что перспективы дальнейших исследований в области нейросетевого распознавания эмоций по геометрии лица могут быть связаны с совершенствованием архитектурных решений капсульной нейронной сети с целью уменьшения количества учебных итераций при обеспечении приемлемой погрешности распознавания.

Ключевые слова: *эмоциональное состояние, базовые эмоции, изображение лица, нейросетевая модель, капсульная нейронная сеть.*

NEURAL NETWORK MODEL OF EMOTIONAL RECOGNITION BY IMAGE OF FACE

The article is devoted to solving the problem of developing a neural network model, which allows realizing the recognition of emotions on the basis of image of a person's face under the influence of interferences, characteristic for information systems of general purpose. The low adaptability of distributed solutions based on the packet neural networks to the leveling of such a characteristic barrier as facial rotation in the image to be analyzed is determined. It is proposed to correct this shortcoming due to the use of the capsule neural network, which is the development of convolutional neural networks in the direction of recognition of noisy images. A corresponding neural network model has been developed, the efficiency of which has been proved experimentally. It is shown that the prospects for further research in the field of neural network recognition of emotions in face geometry can be related to the improvement of architectural solutions of the capsular neural network in the direction of reducing the number of training iterations while providing an acceptable recognition error.

Key words: *emotional state, basic emotions, face image, neural network model, capsule neural network.*