

УДК 004.93

ЗГОРТКОВА НЕЙРОННА МЕРЕЖА ДЛЯ АНАЛІЗУ ЗОБРАЖЕНЬ

Буюклі В.С.

к.ф.м.н., доцент каф. КІСМ Тішин П.М., ст. викладач каф. КІСМ Кузнєцов М.О.
Державний університет «Одеська політехніка», Україна

АНОТАЦІЯ. У роботі наведено результати досліджень визначення емоцій людини за допомогою згорткової нейронної мережі. Розробка моделі для аналізу зображень проводилась за допомогою TensorFlow, а тренування реалізовувалось з використанням Keras. Вхідні дані – набір даних «FER» та власна вибірка зображень.

Вступ. Зміна виразу обличчя є природним засобом спілкування, передачі емоцій, намірів і стану людини. Автоматизоване розпізнавання виразу обличчя і визначення емоцій людини є досить складною задачею.

Рішення даної задачі актуально при створенні: систем комп'ютерного зору, або систем людино-машинної взаємодії; алгоритмів реалістичної анімації особи людини в комп'ютерних іграх і машинній графіці; інформаційних технологій емоційного маркетингу тощо.

Мета роботи. Застосування методу глибинного навчання обробки і аналізу візуальних даних з використанням глибинних нейронних мереж для покращення результатів рішень задачі автоматичного розпізнавання емоцій на обличчі людини.

Основна частина роботи. Потреба людини в розробці технологій розпізнавання осіб сформувалася давно. Сьогодні такі алгоритми знаходять своє застосування в багатьох областях – охоронні системи, верифікація і т.д. Безумовно, вони постійно вдосконалюються і показують все кращі результати. Одним з таких алгоритмів є згорткова нейронна мережа [1], яка завдяки своїй структурі і використуванню в ній алгоритмам здатна працювати з будь-якими зображеннями.

В якості методу глибинного навчання для обробки, аналізу зображень та розпізнавання образів обрана класифікація. Важливі фактори, що впливають на класифікацію: пошук початкових даних, підготовка даних, кодування вихідних значень, вибір об'єму мережі, вибір архітектури.

Для проведення експерименту обрано «FER» [2] – спеціально сформовану базу даних, яка містить еталонні обличчя для порівняльних цілей. Набір даних містить 35887 зображень з роздільною здатністю 48×48 пікселів, більшість з яких зроблені в довільних умовах. База даних була створена з використанням інструментів пошуку зображень Google. Кожне зображення класифіковано одним з семи видів емоцій: подив (surprise), страх (fear), радість (happy), злість (angry), відраза (disgust), смуток (sad) і нейтральний стан або спокій (neutral).

Задача вже розглядалася різними дослідниками, які займаються розробками систем з детектування і розпізнавання. Однак з відкритих джерел отримані мережі, в яких використовується та ж архітектура та набір даних, в середньому мають точність на тестовому набору 60% (±5%) [3,4].

Запропонована структура згорткової мережі будується на основі п'яти видів шарів: згортковий шар (convolutional layer), шар пакетної нормалізації (batch-normalization layer), шар підвибірки (pooling layer), шар проріджування (dropout layer), повноз'єднаний шар (fully-connected layer). Береться чотири рази така послідовність модулів: conv + conv + batch + maxpool + dropout. Після чого стандартна послідовність: flatten + dense + dropout + dense. Функцією активації шарів – ReLU, окрім останнього шару, де функцією активації є softmax. В якості вхідного шару (input layer) виступає один з згорткових шарів, який приймає зображення. Також існує шар згладжування (flatten layer), який перетворює 2D-дані в одновимірні дані.

В наборі даних існують різного виду зображення, але під час навчання для поліпшення роботи використовується поповнення даних.

Проведено навчання базової моделі з установкою таких параметрів: розмір серії – 32, кількість епох – 300. Для компіляції в якості оптимізатору обрано «Adam», цільова функція – «categorical_crossentropy».

В результаті виконання програми отримана точність близько 70%.

Під час навчання моделі дані за показниками функції втрат і точності зберігалися. Після завершення навчання за допомогою модулів викликаних бібліотек отримані графіки (рис. 1).

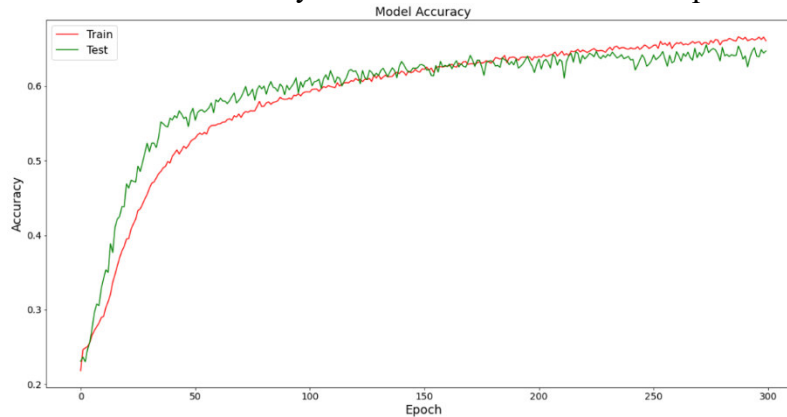


Рис. 1 – Графіки точності мережі

Розглядаючи графіки на рисунку, можна відзначити схожість зі степеневою функцією. В цілому графіки зростаючі, що підтверджує загальне зростання точності.

Саме точність на тестовому наборі обрано як основний показник для оцінки. На тестовому наборі цей показник досягає 68%. У порівнянні з існуючими рішеннями отримана більш висока точність.

Для тестування навченої моделі на власній вибірці використовується допоміжний блок додатку пошуку і виділення обличчя з початкового зображення бібліотеки комп'ютерного зору з відкритим вихідним кодом OpenCV – «CascadeClassifier» [5], який дозволяє знайти обличчя на заданому зображенні. У власній вибірці 21 зображення. З них 15 зображень мережа класифікувала вірно, що становить 71%.

Висновки. Для аналізу зображення була обрана мережа для класифікації емоцій людини по її обличчю на зображенні. В результаті проведеної розробки мережі та її навчання найкраща отримана точність склала 68%. Тестування моделі на довільних даних, що не відносяться до набору даних «FER», дозволило якісно оцінити точність розпізнавання емоцій. Отже, отримана CNN може бути застосована при вирішенні задач, що вимагають побудови систем розпізнавання виразів обличчя людини і визначення емоцій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Нейронные сети, их применение и принципы работы / В. В. Девятков, Р. А. Матейчук, И. И. Мищенко, Н. А. Кузнецов // Modern Information Technology 2017 = Сучасні інформ. технології 2017 : матеріали 7-ої Міжнарод. наук. конф. студентів та молодих вчених / Одес. нац. політехн. ун-т. – Одеса, 2017. – С. 11–12.
2. Facial Expression Recognition Challenge // Deeplearning [Електронний ресурс]. – Режим доступу: <http://deeplearning.net/icml2013-workshop-competition/challenges>.
3. D.V. Plotnikov, E.A. Sopov «Solving the face and facial expression recognition tasks using convolutional neural networks», pp.234-237, 2017.
4. O.P. Timofeeva, S.A. Neimushchev «Facial emotion recognition using deep neural networks», vol. 1, 2020 – pp.16-24.
5. Lienhart R., Kuranov E., Pisarevsky V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: PRS 2003, pp. 297-304.