

РОЗПІЗНАВАННЯ ВИРАЗУ ОБЛИЧЧЯ З ВИКОРИСТАННЯМ ГЛИБОКОГО НАВЧАННЯ

Приведено актуальність дослідження розпізнавання виразу обличчя, проаналізовано популярні методи розв'язання задачі, обґрунтовано вектор досліджень та його перспективність.

Ключові слова: розпізнавання виразу обличчя, глибоке навчання.

O.O. MORFIYANEC, V.I. MESYURA

Vinnitsia National Technical University

FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING

Abstract — Paper provides the arguments for the relevance of the facial expression recognition problem, and for the relevance and fruitfulness of applying deep learning to this particular problem. The purpose of the research is to increase the facial expression recognition precision using deep learning. The paper is focused on the problem of facial expression recognition, which might be used for predicting future human behaviour. Such expressions are hard to notice, but are very useful in psychology, investigation, etc. Hence, the automated system for facial expression recognition is reasonable. Authors provide key approaches to the problem, and emphasize their bad precision. Also, the deep learning methods are analyzed, their success and novelty are the basis for their fruitfulness for this particular problem.

In conclusion, the paper confirms the fruitfulness of the deep learning approach for the facial expression recognition problem, taking into consideration the novelty of the approach, the increase of the computing power, and parallel computations development on graphical processing units.

Keywords: facial expression recognition, deep learning.

Вступ

Короткі мимовільні вирази обличчя, що виникають на обличчі людини можуть слугувати ознакою обману при невідповідності до слів чи макровиразів або індикаторами майбутньої поведінки людини [1]. Вони характеризуються малою тривалістю до п'ятої частини секунди, тому недосвідченою людиною залишаються зазвичай непоміченими [2]. Тому доцільно автоматично розпізнавати вираз обличчя, що дозволить ідентифікувати основні вирази обличчя навіть при спостереженні за великою кількістю людей одночасно. Доцільно використовувати розпізнавання виразу обличчя в області психології, наприклад, для визначення підозрілих людей у місцях скупчення людей, у слідчій практиці для визначення обману зі сторони підозрюваного.

Актуальність теми дослідження

В сучасному науковому суспільстві в останні декілька років набули популярності алгоритми глибокого навчання. Нещодавні досягнення у підвищенні їх ефективності в області машинного навчання вказують на можливість створення інформаційної технології розпізнавання виразу обличчя з використанням глибокого навчання та підвищення точності розпізнавання виразу обличчя. Враховуючи новизну області можна стверджувати про можливість значного покращення точності розпізнавання виразу обличчя при використанні глибокого навчання [3].

Ключовими в області глибокого навчання є роботи професора Ендрю Нг та його дослідження у проекті «Гугл брейн»: вперше були застосовані широкомасштабні глибокі нейронні мережі на 16000 процесорних ядрах [4].

Метою дослідження є підвищення точності розпізнавання виразу обличчя з використанням глибокого навчання.

Використання нейронних мереж для розпізнавання виразу обличчя

Нейронні мережі мають великий потенціал у задачах розпізнавання, і зокрема в задачі розпізнавання виразу обличчя. Звичайні нейронні мережі складаються з вхідного, вихідного та одного або декількох прихованих шарів. В 1994 Марком Розенблумом і Язер Якобом з Мерілендського університету була представлена система, що базувалася на нейромережах прямого розповсюдження з радіально-базисною функцією активації [5]. Ця система розпізнавала шість виразів обличчя із загальною точністю у 73% на тестових даних. Основними недоліками цієї системи були низька точність розпізнавання та збої при інтерпретації емоцій, коли значення на виходах мереж були досить близькі або рівні.

Наступним суттєвим досягненням в області розпізнавання виразу обличчя була розробка Джу-Леонг Чанга і Джаи-Лі Чен з Національного університету Чжао Тунга, що була опублікована в 2001 році [6]. Чанга і Чен надихнула система FACS розроблена Екманом і на її основі ними була запропонована парадигма нейромережі прямого розповсюдження з радіально-базисною функцією активації, що використовувала б активних одиниць із 46 представлених у FACS. Дана мережа може класифікувати 3 вирази – радість, гнів, а також нейтральний стан людини. Точність розпізнавання такої мережі склала в середньому 87% [7].

Основним недоліком такого підходу є те що така система може класифікувати лише 3 емоції. Тим не менше не дивлячись на те, що система могла бути вдосконалена, вона не отримала подальшого розвитку.

У 2008 році було для розпізнавання виразу обличчя людини було використано квантові нейромережі [8]. Вони не потребували складної передобробки даних і показали значне зменшення помилки на тренувальних даних, точність розпізнавання складала 96.5% порівняно зі звичайною нейромережею прямого розповсюдження (89.9%) на базі даних JAFFE (Japanese Female Facial Expression Database). Однак перевірки роботи системи на тестових (небачених системою під час тренування) даних зроблено не було.

Використання глибокого навчання для розпізнавання виразу обличчя

Глибоке навчання — це галузь машинного навчання, основана на наборі алгоритмів, які намагаються моделювати високорівневі абстракції в даних. Часто використовують моделі нейронних мереж з великою кількістю прихованих шарів. Шари в такій моделі відповідають різним рівням концептуалізації, де концепції (ознаки, функції, представлення) більш високого рівня визначаються з концепцій більш низького рівня. Таким чином, одні й ті самі концепції низького рівня допомагають визначити багато високорівневих концепцій [8]. Глибоке навчання базується на розподілених інтерпретаціях, припускаючи, що спостережувані дані були створені комбінацією багатьох факторів, не всі з яких відомі спостерігачу, тому інформація про певний фактор, отримана з деяких конфігурацій інших факторів, часто може узагальнюватись до інших, невідомих конфігурацій факторів. Глибоке навчання додає до розподілених інтерпретацій припущення, що ці фактори мають багаторівневу організацію у відповідності до різних рівнів абстракції або складу об'єкту і високорівневі представлення отримуються шляхом перетворення та комбінування низькорівневих.

До моделей глибокого навчання можна віднести згорточні нейронні мережі, глибокі мережі довіри, звичайні багатощарові нейромережі прямого розповсюдження та ін. Приклад схеми багатощарової нейронної мережі прямого розповсюдження наведено на рисунку 1.

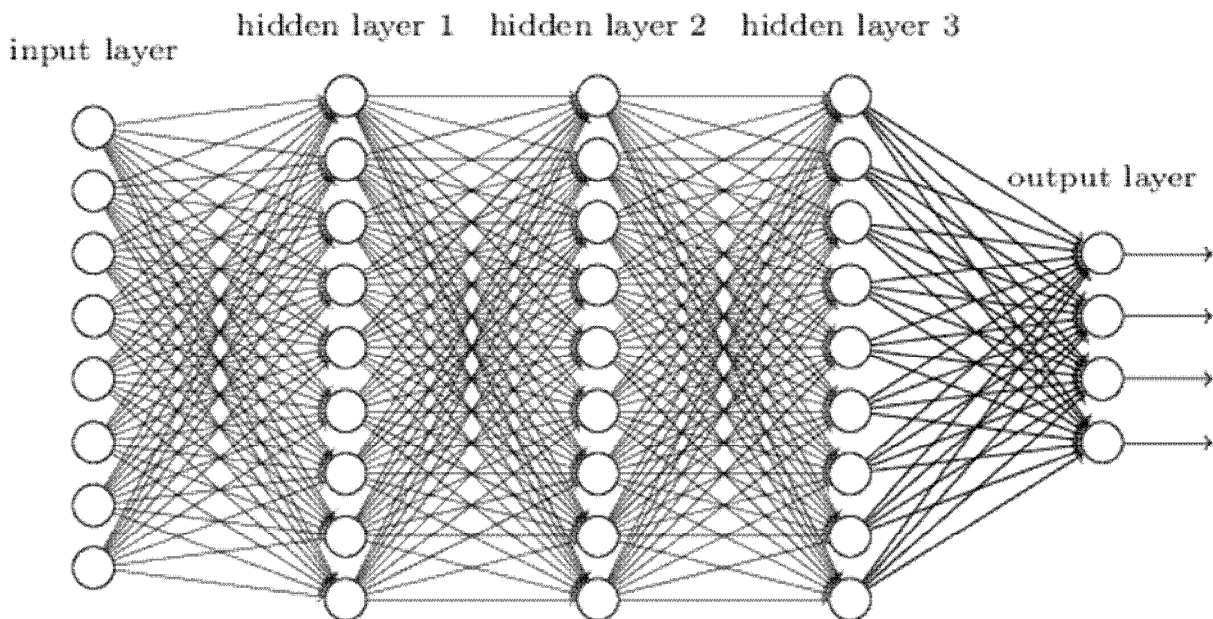


Рис. 1. Приклад схеми багатощарової нейронної мережі прямого розповсюдження

В 2008 році, Бітом Фейзелем була запропонована архітектура згорточної нейронної мережі що складається з 5-ти шарів, 2 з яких – шари підвибірки і 2 – згорточні шари [9]. Така архітектура дозволила значно знизити навантаження на модуль попередньої обробки і не використовувати особливих алгоритмів для попередньої обробки. Дана система досить ефективно розпізнає вирази обличчя людини і проводить розпізнавання з досить великою швидкістю. Тим не менше для підвищення швидкодії було прийнято рішення зменшити кількість нейронів і зв'язків між ними, зберігаючи архітектуру запропоновану Фейзелем. Шари згортки виконують двовимірну згортку зі зміщенням областей зображення, що перекриваються, в той час як шари підвибірки виділяють максимальне значення в областях зображення, що не перекриваються, і передають його до наступного шару згортки. Навчання ваг шарів згортки дозволяє виділення ознак, що залежать від наявних вхідних даних, тоді як шари підвибірки збільшують інваріантність об'єктів, що залежать від місця, що нас цікавить, та зменшують розмірність задачі. Розділення ваг (використання одного й того самого набору ваг для кожної оброблюваної області) дозволяє значно зменшити кількість незалежних параметрів, що в свою чергу покращує здатність до узагальнення.

Б.Фейзель дослідив декілька архітектур згорточної нейромережі і виділив вищевказану, за допомогою якої було досягнуто точності розпізнавання у 82.9% на базі даних JAFFE (Japanese Female Facial Expression Database). Основним недоліком такої архітектури є об'ємність обчислень, необхідних при

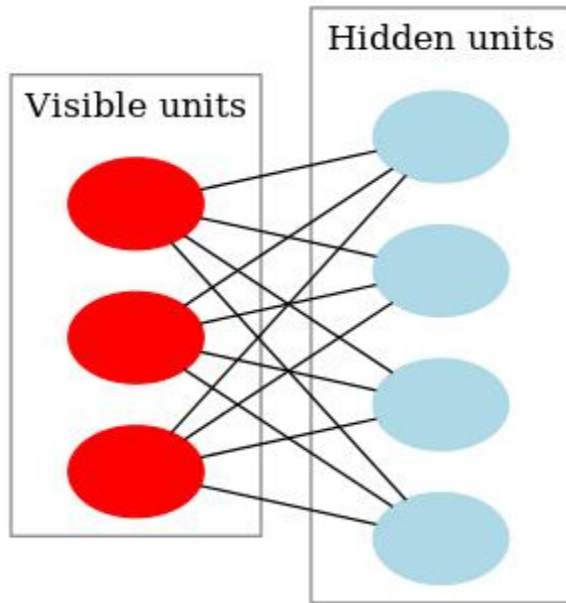


Рис. 2. Узагальнена схема обмежених машин Больцмана

змінними наступного та попереднього шару, та одного шару спостережуваних змінних. Узагальнена схема глибоких мереж довіри наведена на рисунку 3.

Як виявилось, алгоритми тренування глибоких мереж довіри можуть бути використані для попереднього тренування без учителя ваг нейромережі прямого розповсюдження. Їх використання призводить до автоматичного виділення у кожному шарі характерних розподілених ознак вхідних даних з попереднього шару, що є основою розподілених інтерпретацій. Це дозволяє будувати глибокі нейромережі, не зважаючи на проблеми застосування зворотного розповсюдження похибки до глибоких шарів — їх ваги вже будуть коректно ініціалізовані під час переднавчання. В роботі Т. McLaughlin було використано бази даних CK+, Yale Face Database та JAFFE. Точність на тестових даних при використанні лише баз даних CK+ та Yale Face Database для розпізнавання семи базових виразів обличчя та нейтрального виразу склала 71%, а при використанні всіх трьох баз даних — 59%. Така низька точність розпізнавання є значним недоліком порівняно з іншими розглянутими методами, однак вона є наслідком недостатньої вивченості подібних моделей, а не їх непридатності до задач розпізнавання. Точність розпізнавання може бути значно покращено, оскільки самі розробники визнають, що технологія внаслідок своєї новизни ще не дійшла до оптимального стану і можуть існувати кращі типи нейронів прихованого шару чи мережевої архітектури, та й сам алгоритм навчання скоріш за все може бути покращено [12].

Висновки

Приведено актуальність дослідження розпізнавання виразу обличчя. Проаналізовано існуючі досягнення в розпізнаванні виразу обличчя за допомогою нейронних мереж. Проаналізовано деякі вдалі підходи до вирішення задачі розпізнавання виразу обличчя за допомогою глибокого навчання. Здійснений аналіз свідчить, що глибоке навчання є перспективним підходом для вирішення поставленої задачі через свою відносну новизну, можливість великої кількості варіацій, а також через підвищення швидкодії комп'ютерів згідно закону Мура, розповсюдження обчислень на нових, потужних графічних модулях, які дозволяють проводити обчислення з високим ступенем паралелізації [13], більшій кількості даних для тренування та тестування розпізнавання виразу обличчя людини.

Література

1. P. Ekman, Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Henry Holt and Company, 2003. 336 p.

розпізнаванні: хоча кількість параметрів мережі є малою, на кожному кроці згортки необхідно обробляти велику кількість областей зображення, що перекриваються. Слід зазначити, що цей недолік може бути усунений з використанням паралельних обчислень, оскільки кожна така область може бути оброблена одночасно з іншими з використанням однакового набору параметрів. У 2010 році Томом Мак-Лафіним та іншими було використано підхід, що базується на глибоких нейромережах прямого розповсюдження із попереднім тренуванням за допомогою обмежених машин Больцмана [10] — типу нейронних мереж, який отримав значне розповсюдження останнім часом у зв'язку із порівняною легкістю використання, відсутністю потреби у великих об'ємах помічених даних (порівняно з іншими типами нейромереж, такими як згорткові) та нещодавно розробленими ефективними алгоритмами тренування. Узагальнена схема обмежених машин Больцмана наведена на рисунку 2.

Глибокі мережі довіри [11] — генераційні моделі, що складаються з багатьох шарів стохастичних прихованих змінних, що мають зв'язки лише зі

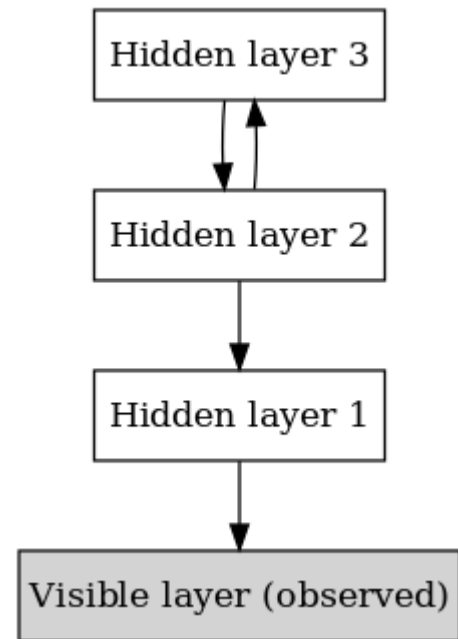


Рис. 3. Узагальнена схема глибоких мереж довіри

2. P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York: W. W. Norton & Company, 2009. 416 p.
3. G. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, *Neural Computation*. 2006. № 8. P. 1527–1554.
4. Ng Andrew, Dean Jeff, *Building High-level Features Using Large Scale Unsupervised Learning*, Google, (2012), P. 1–11.
5. M. Rosenblum, Y. Yacoob, L. Davis, *Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture*, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*. 1994. No17. P. 43–49.
6. J.-Y. Chang, J.-L. Chen, *Automated Facial Expression Recognition using Neural Networks*, Hsinchu: Chiao Tung University, 2001. 11 p.
7. P. Li, J. Li, *A Facial Expression Recognition Method Based on Quantum Neural Networks*, *International Journal of Computational Intelligence Systems*. 2008. No. 8. P. 74–78.
8. Y. Bengio, *Learning Deep Architectures for AI*. Montreal: Universit? de Montr?al, 2009. 130 p.
9. B. Fasel, *Multiscale Facial Expression Recognition using Convolutional Neural Networks*, *IDIAP 2009*. 2009. No. 13. P. 121–137.
10. T. McLaughlin, M. Le, N. Bayanbat, *Emotion Recognition with Deep-Belief Networks*, *Stanford CS 229 Machine Learning Final Projects*. 2010. No. 14. P. 62–66.
11. Deep belief network – Wikipedia, URL: https://en.wikipedia.org/wiki/Deep_belief_network
12. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury, *Deep Neural Networks for Acoustic Modeling in Speech Recognition*, *IEEE Signal Processing Magazine*. 2012. No. 16. P. 2–17.
13. Яровий А. А. Аналіз обчислювальної складності GPU-орієнтованих паралельно-ієрархічних обчислювальних систем та оцінювання продуктивності їх апаратного забезпечення / А. А. Яровий, Н. І. Кокряцька, С. В. Наконечна, М. С. Матейчук, Т. Д. Польгуль // *Оптико-електронні інформаційно-енергетичні технології*. – 2014. – № 1. – С. 18–25.

References

1. P. Ekman, *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Henry Holt and Company, 2003. 336 p.
2. P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York: W. W. Norton & Company, 2009. 416 p.
3. G. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, *Neural Computation*. 2006. № 8. P. 1527–1554.
4. Ng Andrew, Dean Jeff, *Building High-level Features Using Large Scale Unsupervised Learning*, Google, (2012), P. 1–11.
5. M. Rosenblum, Y. Yacoob, L. Davis, *Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture*, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*. 1994. No17. P. 43–49.
6. J.-Y. Chang, J.-L. Chen, *Automated Facial Expression Recognition using Neural Networks*, Hsinchu: Chiao Tung University, 2001. 11 p.
7. P. Li, J. Li, *A Facial Expression Recognition Method Based on Quantum Neural Networks*, *International Journal of Computational Intelligence Systems*. 2008. No. 8. P. 74–78.
8. Y. Bengio, *Learning Deep Architectures for AI*. Montreal: Universit? de Montr?al, 2009. 130 p.
9. B. Fasel, *Multiscale Facial Expression Recognition using Convolutional Neural Networks*, *IDIAP 2009*. 2009. No. 13. P. 121–137.
10. T. McLaughlin, M. Le, N. Bayanbat, *Emotion Recognition with Deep-Belief Networks*, *Stanford CS 229 Machine Learning Final Projects*. 2010. No. 14. P. 62–66.
11. Deep belief network – Wikipedia, URL: https://en.wikipedia.org/wiki/Deep_belief_network
12. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury, *Deep Neural Networks for Acoustic Modeling in Speech Recognition*, *IEEE Signal Processing Magazine*. 2012. No. 16. P. 2–17.
13. А. Яровий, Н. І. Кокряцька, С. В. Наконечна, М. С. Матейчук, Т. Д. Польгуль, "Аналіз обчислювальної складності GPU-орієнтованих паралельно-ієрархічних обчислювальних систем та оцінювання продуктивності їх апаратного забезпечення", *Оптико-електронні інформаційно-енергетичні технології*, 2014. № 1. P. 18–25.

Рецензія/Peer review : 26.8.2015 р.

Надрукована/Printed : 31.8.2015 р.

Рецензент: д.т.н., проф., Перевозніков С. І.