

О.В. Гороховатський, О.В. Тесленко

Харківський національний університет імені Семена Кузнеця, Харків

РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ ЛІТЕР ІЗ ВИКОРИСТАННЯМ ПРОЕКЦІЙ ТА БАГАТОШАРОВОГО ПЕРСЕПТРОНУ

Статтю присвячено розробці методу розпізнавання зображень із використанням проекційного аналізу як методу збереження структурної інформації із подальшим навчанням багатошарового персептрону у якості інструменту узагальнення та розпізнавання. Досліджено особливості комбінування властивостей проекцій та процесу навчання персептрону, запропоновано два підходи до нормалізації проекційних ознак. Проведено експериментальне моделювання, яке довело ефективність методу для вирішення практичної задачі розпізнавання зображень літер сканованого тексту.

Ключові слова: розпізнавання зображень, багатошаровий персептрон, навчання, проекція, перетворення Радона, характерні ознаки, структура мережі.

Вступ

Проблеми розпізнавання та аналізу зображень є одними з найактуальніших в задачах, пов'язаних з комп'ютерним зором і машинним навчанням. Популярним механізмом реалізації саме розпізнавання є використання нейромережевого підходу, при якому штучна нейронна мережа спочатку навчається на деяких ознаках еталонної множини зображень, після чого налаштована мережа застосовується для розпізнавання.

Ключовим моментом для використання нейронних мереж є ефективний вибір ознак, які фіксуються на зображенні і подаються на вхід нейронної мережі. Часто мова йде про популярні структурні ознаки (наприклад, SIFT чи SURF) або ознаки яскравості (значення компонентів кольору пікселів). Разом з тим, існує також набір проекційних інтегральних ознак, який дозволяє легко поєднувати інформацію про структуру зображення з функцією розподілу його яскравості. Крім того, використання інформації про проекції дозволяє звести задачу розпізнавання двовимірного зображення до задачі розпізнавання одновимірних проекцій.

Із розвитком штучних нейронних мереж інформація, отримана за допомогою проєціювання зображень, почала використовуватися для розпізнавання не тільки окремо, а також в комбінації із нейромережевими методами, наприклад, для розпізнавання підпису [1], рисунку вен [2], пошкодження тканин [3], при цьому типовим вибором для доповнення проекційного аналізу є машини опорних векторів.

Статтю присвячено розробці методу розпізнавання зображень на основі проєціювання функції яскравості в простір одновимірних проекцій з подальшим його використанням в якості характерних

ознак для нейромережевого навчання. Також умовою досліджень є можливість реалізації безпосередньо розпізнавання зі швидкістю, близькою до реального часу. В якості тестового набору зображень будемо використовувати букви, отримані зі сканованого документу.

Проекційне перетворення

Перетворення Радона було запропоноване Й. Радонам [4] і знайшло найбільше застосування в томографії. Адаптоване для обробки зображень це перетворення являє собою паралельне проєціювання функції яскравості зображення вздовж обраного напрямку та може бути записане у вигляді [5–7]:

$$R(p, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) \delta(p - x \cos \theta - y \sin \theta) dx dy, \quad (1)$$

де $\delta(\cdot)$ – дельта-функція, що визначає приналежність точок зображення $I(x, y)$ до прямої з параметрами (p, θ) , вздовж якої виконується проєціювання. Також $p \in P$ – відстань від початку координат до лінії інтегрування, $\theta \in \Theta$ – кут між прямими, що визначають напрямок та віссю абсцис; P, Θ – області значень змінних p, θ .

Інтегрування в (1) виконується всередині деякої області $D = xy$, де визначено функцію $I(x, y)$. Також будемо вважати, що за межами цієї області D виконується умова $I(x, y) = 0$. Відзначимо, що значення відстані p в (1) може бути як позитивним, так і негативним.

Застосування перетворення Радона до зображення зберігає основну інформацію про візуальні об'єкти, в деякому плані навіть розширюючи її [5–7].

Для існування інтегралу (1) необхідно, щоб функція зображення була абсолютно інтегрована,

що вочевидь виконується для зображень як для функцій, що мають обмежені значення в усій області визначення.

Вираз (1) також можна записати в іншому вигляді через проекцію на вісь під кутом θ після повороту системи координат xOy :

$$R(x', \theta) = \int_{-\infty}^{\infty} I(x' \cos \theta - y' \sin \theta, x' \sin \theta + y' \cos \theta) dy',$$

де $R(x', \theta)$ – проекція двовимірної функції $I(x, y)$ на вісь x' . Тут x' та y' задані поворотом системи координат xOy на кут θ проти годинникової стрілки, тобто:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} +\cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Багатошаровий перцептрон

Перцептрон як нейромережева архітектура є одним з найбільш простих типів нейронних мереж. Обмеженість одношарового перцептрону при вирішенні лінійно-нероздільних проблем є відомою, тож ми будемо використовувати багатошаровий перцептрон [8].

Вхідний шар перцептрону складається із вектору значень, які в нашому випадку відповідають значенням одновимірних проекцій. В якості активаційної функції для кожного з нейронів будемо використовувати класичний варіант сигмоїда:

$$y(x) = \frac{1}{1 + e^{-z}}, \quad z = wx + b,$$

де x – вхідне значення, w – ваговий коефіцієнт, b – зміщення.

Значення, що є виходом j -го нейрона N_j прихованого шару можна записати у вигляді $N_j = f(\sum_i w_{ij} X_i)$, де X_i – значення вхідного сигналу (або значення з виходу нейрону попереднього шару), вихід k -го нейрона Y_k останнього шару мережі можна записати у вигляді $Y_k = f(\sum_j w_{jk} N_j)$.

Значення вхідного вектора варіюються від 0 до 1 залежно від подібності вхідного фрагмента зображення до одного з еталонних патернів.

Кількість прихованих шарів мережі і кількість нейронів в кожному шарі підбирається експериментальним шляхом.

Мережа навчається з учителем з використанням алгоритму зворотного поширення помилки методом градієнтного спуску (узагальнене дельтаправило) і включає три етапи:

- пряме поширення вхідного сигналу по мережі і отримання значень на виході кожного нейрона;
- зворотне поширення помилки;

– корекція вагових коефіцієнтів і зміщень.

Для нейрона k вихідного шару, що є пов'язаним із нейроном j попереднього шару, помилка вираховується як:

$$\delta_k = (Y_{\text{desired}_k} - Y_k) Y_k (1 - Y_k), \quad (3)$$

де Y_{desired_k} – потрібне значення виходу нейрона k .

Для нейрона j прихованого шару, що є пов'язаним із нейроном i попереднього шару, помилка вираховується як:

$$\delta_j = h_j (1 - h_j) \sum_k w_{jk} \delta_k. \quad (4)$$

Корекція коефіцієнтів зміщень відбувається згідно:

$$w = w + \Delta w, \quad \Delta w = \eta \delta x + m \Delta w, \quad b = b + \eta \delta,$$

де η – коефіцієнт швидкості навчання (при реалізації було використано значення $\eta = 0.2$), δ – значення помилки, отримане згідно (3) чи (4), x – значення відповідного нейрону на виході з попереднього шару, m – коефіцієнт інерції, що зменшує ймовірність можливого «зациклювання» в точці локального мінімуму (експериментальне значення дорівнювало $m = 0.8$).

Навчання перцептрону на проекційних ознаках

Розглянемо процес навчання багатошарового перцептрону на множині одновимірних проекцій. Для досягнення ефективного результату необхідно вирішити наступні базові проблеми:

- визначити етапи попередньої обробки зображень перед проектуванням;
- визначити ефективну кількість використовуваних проекцій;
- визначити структуру мережі: кількість шарів, кількість нейронів в прихованих шарах;
- знайти спосіб нормалізації значень проекцій, щоб зробити їх використання придатним для багатошарового перцептрону.

Попередня обробка

Типовий перелік етапів попередньої обробки зображень подібного типу (літери, символи, тощо) перед розпізнаванням включає [9] сегментацію, масштабування та бінаризацію. Кожний з цих етапів використовує апріорні дані, наприклад, відомі значення порогів чи масштабів. Враховуючи те, що перетворення Радона є інтегральним, що дозволяє досягти значної завадостійкості, єдиною операцією обробки яскравості зображення на попередньому етапі є інверсія зображення, яка не дає можливості фоновому світлому кольору впливати на значення проекції.

Процес отримання зображення кожного символу із загального зображення сканованого документу також реалізовано із використанням проєкційного аналізу [10], після чого кожне з зображень центрується відповідно до геометричного центру. Це дозволяє досягнути інваріантності до зміщень символів всередині поля зору, однак, може створювати проблеми при розпізнаванні символів, які відрізняються лише зміщенням, наприклад, «'» та «,».

Кількість проєкцій

З урахуванням необхідності досягнення високої швидкості розпізнавання пропонується використовувати лише 3 проєкції при $\theta = 0^\circ$, $\theta = 45^\circ$ та $\theta = 90^\circ$. Кожна з проєкцій має довжину в 100 одиниць відповідно до розміру зображення, таким чином, довжина вхідного вектора ознак складає 300 одиниць. Експериментально доведено, що навчання і розпізнавання з використанням п'яти проєкцій не дає покращень в порівнянні з використанням лише трьох. Більш того, використання значної кількості проєкцій також недоцільно через симетричність більшості зображень символів.

Структура мережі

Структуру мережі було підібрано експериментальним шляхом. Встановлено, що для вирішення задачі достатньо мати один прихований шар із 15 нейронами. Таким чином, загальна структура перцептронів включає 3 шари: перший вхідний із 300 нейронів, другий прихований і третій вихідний шар має 86 нейронів відповідно до кількості повного набору всіх букв латинського алфавіту, цифр, розділових знаків та спеціальних символів.

Нормалізація вхідних сигналів

Перед подачею на вхід мережі значення проєкцій необхідно нормувати для отримання значень в межах від -0.5 до 0.5 включно.

Можна уявити два способи нормалізації значень вхідної сукупності проєкцій. Перший полягає в нормалізації кожної з проєкцій окремо із подальшим об'єднанням в загальний вектор, другий є більш логічним відповідно до зберігання сукупної структури зображення та полягає в нормалізації усього вектору з трьох проєкцій.

Результати експериментів

Розглянемо можливість застосування запропонованого методу для вирішення практичної задачі розпізнавання зображень літер тексту, цифр та спеціальних символів.

Набір даних для навчання складався з 1668 зображень літер та розділових знаків розміром 100×100 пікселів та 832 зображень цифр та спеціальних символів того ж розміру. Всі ці зображення були отримані з початого зображення сканованого документу (300 dpi) за допомогою горизонтального та вертикального проєціювання. Еталонні зобра-

ження символів зберігаються та використовуються в бінарному форматі (рис. 1).

Верифікацію якості було виконано за допомогою двох принципово різних тестових наборів. Перший з них (DS_1) складався з 2225 символів сканованого тексту та був отриманий аналогічно еталонному набору (розмір початкового зображення склав 2550×3509 пікселів), але без використання бінаризації. Якість зображень символів, які було використано в наборі DS_1 , можна оцінити на рис. 2, де також наведено результат сегментації кожної літери у вигляді прямокутника. Можна побачити, що область кожного символу містить завади, притаманні для зображень сканованих документів.

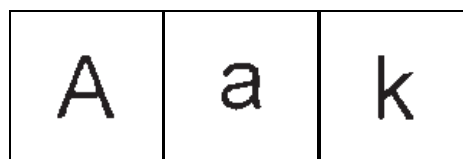


Рис. 1. Еталонні зображення літер в бінарному форматі

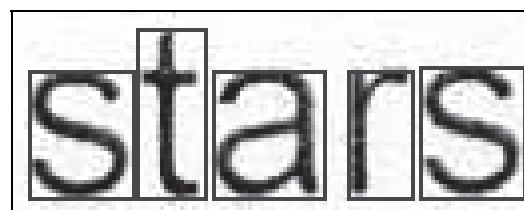


Рис. 2. Сегментовані зображення літер набору DS_1

Другий набір (DS_2 , 10000 окремих зображень розміром 100×100 пікселів) дозволяє оцінити якість узагальнення навченої мережі та складається із зображень літери «С», що були штучно синтезовані із такими варіативними параметрами, як товщина пера, розмір літери, кути початку та закінчення. Жодне з цих зображень не було відоме нейронній мережі і високий рівень розпізнавання цього набору не був очікуваним.

Кожну серію експериментів було оцінено за наступними показниками: кількість нейронів прихованого шару, середній час навчання мережі в секундах, середня кількість ітерацій під час навчання, якість розпізнавання та середній час повної обробки всього сканованого зображення включно із попередньою обробкою, сегментацією літер, розпізнаванням кожної з них та первинною корекцією результатів. Всі узагальнення були проведені за трьома експериментами. Часові показники в секундах наведено не в якості абсолютних значень, а лише для порівняння швидкодії запропонованих структур нейронних мереж між собою. Також на завершальній стадії розпізнавання було виконано первинну корекцію результатів за очевидними правилами: проєкції сим-

волів «!» та «і», і «'» та «,» майже однакові з урахуванням похибки та інваріантності до зміщень, тож за необхідності для цих комбінацій було проведено корекцію.

Порівняння еталонного та отриманого результатів розпізнавання було виконано методом пошуку найбільшої спільної підпоследовності (longest common subsequence) та порівняння її довжини із довжиною еталонного значення.

В рамках першої серії експериментів було оброблено обидва тестових набори із оцінкою якості розпізнавання та швидкості. При цьому кожному з трьох проєкцій було нормалізовано окремо від інших. Навчання нейронної мережі виконувалось до досягнення заданої кількості помилок на тестовому наборі, для нейронних мереж із кількістю нейронів прихованого шару 15 та 25 цей рівень був заданий в межах 0,5%, для мереж із кількістю прихованих нейронів 100 та 200 рівень встановленої помилки склав 1%.

Отримані для першої серії експериментів результати наведено в табл. 1. Як можна побачити, розпізнавання набору DS_1 є майже однаково успішним, однак на незначній кількості нейронів узагальнення в результаті навчання не є успішним.

Таблиця 1

Результати розпізнавання тестових наборів із нормалізацією окремих проєкцій

Кількість нейронів у прихованому шарі	15	25	100	200
Середній час навчання мережі, сек.	1732	968	1317	602
Середня кількість ітерацій під час навчання	5716	1838	633	633
Середня помилка розпізнавання DS_1	1,9%	1,6%	1,4%	1,4%
Середня помилка розпізнавання DS_2	64%	58%	48%	28%
Середній час обробки всього тестового зображення DS_1 , сек.	10,4	9	9,8	13,6

Другу серію експериментів було проведено із однією відмінністю від першої – нормалізацію проєкцій перед поданням їх до нейронної мережі було виконано для всього вектору, складеного з трьох проєкцій, а не для кожної проєкції окремо. Ці результати наведено в табл. 2.

Порівнюючи їх із тими, що були отримані для першої серії експериментів, можна зазначити, що загальний час навчання та кількість ітерацій є значно більшими, ніж це було для нормалізації окремих проєкцій. Помилка розпізнавання також є вищою, так само як і менш вдалим є розпізнавання невідомого набору DS_2 .

Таблиця 2

Результати розпізнавання тестових наборів із нормалізацією усього вектору проєкцій

Кількість нейронів у прихованому шарі	15	25	100	200
Середній час навчання мережі, сек.	2689	3306	4730	11217
Середня кількість ітерацій під час навчання	9241	6507	2503	2428
Середня помилка розпізнавання DS_1	3,2%	2,2%	1,5%	5,9%
Середня помилка розпізнавання DS_2	64%	67%	60%	43%
Середній час обробки всього тестового зображення DS_1 , сек.	9,46	9,3	10,55	12,6

Вищенаведені результати доцільно порівняти з іншим методом, який заснований на одновимірних зрізах і був запропонований в [11–13]. Його основу складає case-аналіз горизонтальних та вертикальних зрізів зображення, побудова опису-профілю зображення та порівняння його з уже еталонними значеннями, які вже є відомими.

Помилка розпізнавання того ж самого зображення розміром 2550×3509 пікселів в однакових умовах із використанням методу, викладеного в [11–13], складає 7%, при цьому на повну обробку зображення включно із сегментацією кожного символу було витрачено в середньому 47 сек. за результатами трьох експериментів. В цей час не входить будь-яка корекція результатів.

Як можна побачити, найгірша (з точки зору якості розпізнавання тестового набору DS_1) з конфігурацій нейронних мереж, наведених в табл. 1 та табл. 2, має лише 5,9% помилок та дозволяє отримати цей результат, більш, ніж в 3,5 рази швидше.

Висновки

Запропоновано метод розпізнавання зображень із використанням часткового перетворення Радона та подальшого навчання багатозарового перцептронну на проєкціях як ознаках. Проаналізовано нюанси

комбінування ознак проєкцій із нейронним навчанням, що дозволило сформулювати деталі використання, кількість проєкцій, попередню обробку та структуру нейронної мережі.

Розглянуто два способи нормалізації вхідних проєкційних даних перед подачею на вхід нейронної мережі, перший з яких полягає в нормалізації кожної проєкції окремо та конкатенації результатів у вхідний вектор, другий – відрізняється послідовністю дій, спочатку виконується конкатенація всіх проєкцій, потім весь вектор нормалізується. Логічним видавалася більш надійна робота другого підходу, оскільки нормалізація кожної проєкції окремо дещо спотворює структурний опис символу. Однак, експериментальне моделювання показало, що нормалізація окремих проєкцій з практичної точки зору виглядає більш ефективно.

Недоліком роботи можна вважати значну кількість помилок розпізнавання на тестовому наборі DS_2 , що говорить про недостатній рівень узагальнення нейронної мережі під час навчання.

Список літератури

1. Kiani V. Offline Signature Verification Using Local Radon Transform and Support Vector Machines / V. Kiani, R. Pourreza, H. R. Pourreza // *International Journal of Image Processing (IJIP)*. – 2009. – Vol. 5, Iss. 3. – P. 184-194.
2. Gholami A. Common Spatial Pattern for Human Identification Based on Finger Vein Images in Radon space / A. Gholami, H. Hassanpour // *Advances and Computer Research*. – 2014. – Vol. 5, Iss. 4. – P. 31-42.
3. Rakhil K. Improved Fabric Defect Detection and Pattern Classification using Radon Transform / K.S. Rakhil, S. Asha // *International Journal of Innovative Research in Science, Engineering and Technology*. – 2015. – Vol. 4, Iss. 7. – P. 5476-5484.
4. Radon J. Uber die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Mannigfaltigkeiten / J. Radon // *Ber. Verh. Sachs. Akad. Wiss. Leipzig, Math.* – Nat. Kl., 69, (1917). – P. 262-277.

5. Гороховатский А.В. Проекционные методы в задачах нормализации и распознавания изображений / А.В. Гороховатский // *АСУ и приборы автоматизации*. – 2009. – №147. – С. 64-70.

6. Gorokhovatskyi A.V. Image invariant recognition methods in the space of projections / A.V. Gorokhovatskyi // *Прикладна радіоелектроніка*. – 2010. – Том 9, № 4. – С. 574-576.

7. Putyatin Y. Projective methods of image recognition / Y. Putyatin, V. Gorokhovatsky, A. Gorokhovatsky, E. Peredriy // *Intelligent Technologies and Application / K. Markov et al. (Eds.)*. – Sofia: FOI ITHEA. – Information Science and Computing. – 2008. – No 5. – P. 37-43.

8. Swingler K. Lecture 4: Multi-Layer Perceptrons [Електронний ресурс] / K. Swingler // Режим доступу: <http://www.cs.stir.ac.uk/courses/ITNP4B/lectures/kms/4-MLP.pdf> – Назва з екрану (13.11.2016).

9. Gorokhovatskyi O. Neocognitron As a Tool for Optical Marks Recognition / O. Gorokhovatskyi // *The First IEEE International Conference on DataStream Mining & Processing 23-27 August 2016, Lviv, Ukraine*. – P. 169-172.

10. Гороховатский А.В. Автоматизация выделения заголовка по изображению документа / А.В. Гороховатский, Е.О. Передрий // *Системы обработки информации*. – X.: ХУПС, 2013. – Вып. 1(108). – С. 150-154.

11. Гороховатский А.В. Распознавание изображений символов на основе линейного описания структурных характеристик / А.В. Гороховатский, Е.О. Передрий // *Системы обработки информации*. – X.: ХУПС, 2012. – Вып. 7(105). – С. 203-206.

12. Gorokhovatskyi O. Primary Symbol Recognition Using Scanline Structural Descriptions / O. Gorokhovatskyi, O. Peredrii // *6th International Academic Conference of Young Scientists "Computer Science and Engineering 2013" (CSE-2013): Lviv, November 21–23, 2013*. – P. 156-157.

13. Гороховатський О.В. Особливості розпізнавання зображень символів із використанням лінійних описів та корекції результатів / О.В. Гороховатський // *Системы обработки информации*. – X.: ХУПС, 2016. – Вып. 4(141). – С. 149-151.

Надійшла до редколегії 6.01.2017

Рецензент: д-р техн. наук проф. О.А. Винокурова, Харківський національний університет радіоелектроніки, Харків.

РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ БУКВ С ИСПОЛЬЗОВАНИЕМ ПРОЕКЦИЙ И МНОГОСЛОЙНОГО ПЕРСЕПТРОНА

А.В. Гороховатский, О.В. Тесленко

Статья посвящена разработке метода распознавания изображений с использованием проекционного анализа изображения как метода сохранения структурной информации с последующим обучением многослойного перцептрона в качестве инструмента обобщения и распознавания. Исследованы особенности комбинирования особенностей проекций и процесса обучения перцептрона, предложены два подхода к использованию проекционных признаков. Проведено экспериментальное моделирование, которое доказало эффективность метода для решения практической задачи распознавания изображений букв сканированного текста.

Ключевые слова: распознавание изображений, многослойный перцептрон, обучение, проекция, преобразование Радона, характерные признаки, структура сети.

LETTER IMAGES RECOGNITION BASED ON PROJECTIONS AND MULTILAYER PERCEPTRON

O.V. Gorokhovatskyi, O.V. Teslenko

Method of image recognition using the projections of an image as a method of structural information preserve with the further training of multilayer perceptron as an instrument of generalization and recognition was proposed. Features of the combination of projection with subsequent perceptron learning process were investigated, two approaches to use the projection feature were proposed. Experimental modeling shown the effectiveness of proposed method to solve the practical problems of letter images recognition retrieved from scanned text.

Keywords: image recognition, multilayer perceptron, learning, projection, Radon transform, features, network structure.