УДК 004.8
DOI https://doi.org/10.32782/tnv-tech.2023.4.11

ПОБУДОВА ШВИДКОЇ ТА ЛЕГКОВІСНОЇ РЕКУРЕНТНОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ РУКОПИСНИХ ЖЕСТІВ

Яковчук О. К. *– асистент, аспірант кафедри системного проектування*
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID ID: 0000-0002-9842-9790

Об'єктом дослідження в цій роботі є нейронна мережа для розпізнавання рукописних жестів. В роботі поставлена задача для створення рішення по розпізнаванню рукописних жестів для можливого використання в різноманітних пристроях та гаджетах, в умовах з обмеженими обчислювальними потужностями та з приорітетністю швидкодії такого рішення. Було обрано набір даних для експериментів, проведено аналіз існуючих робіт та досліджень, в яких використовувався цей датасет, та зафіксовано отримані результати. Для побудови нейронної мережі обрано архітектуру з використанням рекурентних шарів. Було досліджено властивості рекурентних шарів, принципи роботи вентильних рекурентних вузлів як найбільш підходящих складових для такої моделі, особливості тренування рекуретних нейронних мереж. Описано запропоновану архітектуру мережі, обрано 7 варіантів моделей з різними наборами основних параметрів. Проведено тестування моделей, на основі якого визначено дві найбільш оптимальні моделі по точності розпізнавання та по кількості параметрів. Ці моделі було додатково протестовано для перевірки швидкості роботи, як в умовах роботи тільки з використанням процесора, так і з використанням графічного прискорювача. Якість роботи обраної оптимальної моделі було також перевірено з поданням на вхід різноманітних рукописних жестів у різних стилях, які вона не зустрічала до цього у тренувальних наборах. Проведено аналіз успішних випадків роботи мережі, а також наведено приклади невдалих результатів розпізнавання, які в більшості можуть бути виправлені за допомогою розширення тренувального датасета.

Отримані результати підтверджують можливість використання рекуретних шарів для вирішення задачі з розпізнавання рукописних послідовностей, задовольняючи поставлені вимоги по мінімізації часу та ресурсів, з отриманням високої точності розпізнавання при наявності якісного датасету та правильному підбору параметрів моделі.

Ключові слова: *розпізнавання рукописних жестів, рекурентні нейронні мережі, алгоритми розпізнавання, глибокі нейронні мережі.*

Yakovchuk O. K. Construction of a fast and lightweight recurrent neural network for the handwritten gesture recognition task

The object of research is a neural network for handwritten gesture recognition. The work presents the task of creating a solution for recognizing handwritten gestures for its possible next use in various devices and gadgets, in conditions with limited computing resources and with the priority of the speed of such a solution. A dataset for experiments was selected, an analysis of existing works and studies using this dataset was performed, and the obtained results were recorded. An architecture with recurrent layers usage was chosen to build a neural network. The properties of recurrent layers, the operation principles of gated recurrent units as the most suitable components for this model, and the peculiarities of training recurrent neural networks were investigated. The proposed network architecture was described, and 7 variants of models with different sets of basic parameters were selected. Testing of the models was performed, based on which the two most optimal models in terms of recognition accuracy and the number of parameters were determined. These models were additionally tested to check the time performance, in the conditions with only the use of the processor, and with the use of a graphics accelerator. The performance of the single selected optimal model was also tested with the input of various handwritten gestures in different styles that weren't used before in the training sets. An analysis of successful cases of network operation was performed, as well as examples of unsuccessful recognition results, which in most cases can be corrected by expanding the training dataset.

The obtained results confirm the possibility of using recursion layers to solve the problem of recognizing handwritten sequences, meeting the requirements for minimizing time and resources while obtaining high recognition accuracy in the presence of a high-quality dataset and the correct selection of model parameters.

Key words: *handwritten gestures recognition, recurrent neural networks, recognition algorithms, deep neural networks.*

Постановка проблеми. У сучасну цифрову епоху важливість комунікації між людьми та комп'ютерними системами важко переоцінити. Розвиток технологій постійно пропонує користувачам нові альтернативні способи для більш простої та швидкої взаємодії з комп'ютерами та девайсами, наприклад управління голосом, візуальними жестами, тощо. Проте одним з найбільш простих та інтуітивних способів спілкування між комп'ютером та користувачем досі залишається традиційне рукописне введення інформації, оскільки ця базова соціальна навичка здобується ще у школі. Дослідження в сфері розпізнавання рукописного введення наразі є одними з найприорітетних в області взаємодії комп'ютера та людини. Розпізнавання та аналіз рукописного тексту є одним з варіантів такої взаємодії, проте на сьогодні, за дослідженнями, люди використовують рукописне написання все рідше, і кількість написаного тексту в цілому зменшується, оскільки введення повних слів та речень зазвичай є часозатратним процесом [1]. Тому для передачі простої інформації, наприклад керуючих команд для управління та контролю, значно простішим є використання універсальних рукописних жестів, які існують у різних варіаціях для задач різного типу, а також в більшості є мовонезалежними.

Також зважаючи на стрімкий розвиток різноманітних гаджетів, носимих пристроїв, технологій Internet Of Things (IoT), компонентів розумних будинків і т.д., всі з яких потенційно можуть мати можливість сенсорного вводу, питання швидкої взаємодії з цими пристроями за допомогою рукописного введення є надзвичайно актуальним.

Мета роботи. Поставлена задача розпізнавання жестів розглядається в контексті створення системи для швидкої комунікації користувача з мобільними пристроями, гаджетами, IoT девайсами, компонентами розумного будинку і т.д. за допомогою дисплеїв з підтримкою рукописного введення.

Метою цієї роботи є створення рішення на базі нейронних мереж для вирішення задачі розпізнавання рукописних жестів, та визначення оптимальних параметрів побудованої моделі для забезпечення швидкодії в умовах обмеженості обчислювальних потужностей.

Опис експериментальних даних. Для проведення експериментів було обрано відкриту базу даних "The Synchromedia-Imadoc Gesture New On-Line Database (SIGN-OnDB)" [2]. Вона містить 33150 семплів з рукописними жестами, які були отримані від 20 людей шляхом написання рукою на планшетних портативних комп'ютерах та електронних дошках з сенсорним екраном, збережені у форматі InkML. Датасет загалом містить 17 класів базових рукописних жестів, приклади зображені на Рис. 1.

Як бачимо, кожен жест являє собою певну суворо визначену послідовність рухів, тобто жест «лінія вліво» та «лінія вправо» – візуально однакові, проте відрізняються напрямком написання. Цей датасает складає набір інтуітивно простих знаків, які можна використати для побудови інтерактивної системи.

Аналіз наукових досліджень. У роботі [3] описано використання жестів для проведення редагування тексту у додатку з рукописним введенням на мобільних пристроях. Проведено оцінку ефективності використання жестів для цієї задачі

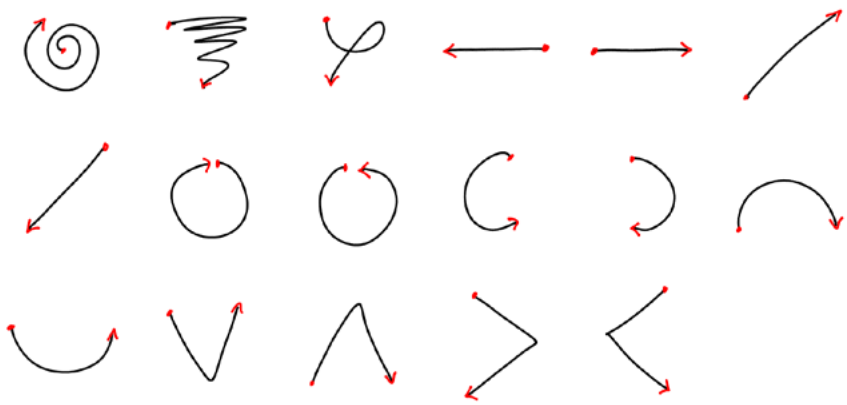

Рис. 1. Приклади рукописних жестів датасету SIGN по кожному класу [2]

порівняно з традиційним інтерфейсом використання клавіатури. Для розпізнавання жестів використано класифікатор «nearest neighbor», з яким було досягнуто точність розпізнавання у 99.76%. Проте, згідно з описом експерименту, у цій роботі було використано не повний датасет SIGN, а лише 7 з 17 усіх жестів, тобто ці результати не можна вважати повноцінними для порівняння.

Новий метод підбирання набору ознак для навчання нейронних мереж у задачі онлайн розпізнавання рукописного введення описано у роботі [4]. Було реалізовано 2 моделі – на базі нейронної мережі, та на базі SVM, та для кожної створено набори ознак залежні від автора ("writer-dependent") та незалежні від автора, для порівняння з нашою роботою нам важливий другий варіант. Для тестування моделей було використано декілька датасетів, результати тестування для датасету SIGN наведені у таблиці 1.

Таблиця 1

Результати тестування моделей [4] на датасеті SIGN

Тип наборів ознак	NN, %	SVM, %
Незалежні від автора (WI-CV)	99.09	99.56
Залежні від автора (WD 10-CV)	99.53	99.65

Побудова та тренування мережі. Класична архітектура нейронної мережі для виконання задач класифікації такого типу базується на використанні рекурентних шарів, де на кожному елементі послідовності модель враховує не тільки поточний вхід, але і ту інформацію, що модель запам'ятала про попередні елементи. Такий тип пам'яті дозволяє мережі вивчати довгострокові залежності в послідовності [5]. На виході мережа має один повнозв'язний шар, який відповідає результуючим класам.

Gated recurrent units (GRU), або вентильні рекурентні вузли – це механізм для рекурентних нейронних мереж (Recurrent neural network – RNN), основна концепція якого базується на двох компонентах – вентиль оновлення та вентиль скидання, відповідні вектори яких вирішують, яка інформація буде передена на вихід комірки [6].

До кожного елементу вхідної послідовності кожен шар GRU [7] застосовує наступні обчислення:

$r_t = \sigma\,(W_{ir}\,x_t + b_{ir} + W_{hr}\,h_{(t-1)} + b_{hr})$
$z_t = \sigma\,(W_{iz}\,x_t + b_{iz} + W_{hz}\,h_{(t-1)} + b_{hz})$
$n_t = tanh\,(W_{in}\,x_t + b_{in} + r_t\,(W_{hn}\,h_{(t-1)} + b_{hn})$
$h_t = (1 - z_t) * n_t + z_t * h_{(t-1)}$

де h_t – прихований стан шару у момент часу t, або 0 у початковий момент часу, x_t – вхід у момент часу t, r_t, z_t, n_t – вентилі скидання, оновлення та новий вентиль відповідно, σ – функція сигмоїди, $*$ – покомпонентне множення матриць.

Для побудови рекурентної нейронної мережі використано бібліотеку PyTorch та мову програмування Python 3. Встановлено наступні параметри для моделі GRU, деякі з яких обрано змінними для тестування різних побудов моделей: hidden_size – змінний параметр (4, 8, 12, 16); num_layers – змінний параметр (1, 2); bias (True); batch_first (True); dropout (0); Bidirectional (False).

На фазі тренування нейронної мережі виконується обнуління градієнтів мережі та використання функції втрат для визначення градієнтів зворотього проходження. У якості функції втрат обрано перехресну ентропію, у якості оптимізатора обрано Адам оптимізатор.

При розбитті датасету, на тестування мережі було виділено 60% від усіх семплів. Це пояснюється особливістю датасету, якій містить велику кількість семплів, при досить незначній варіативності їх написання. Іншою мотивацією було бажання підвищити складність задачі для мережі та перевіряти натреновані моделі на відносно «невідомих» даних, тобто на жестах з такими стилями написання, яких ця мережа можливо ще не бачила.

Результати тестування моделей. Для детального дослідження впливу параметрів рекурентної нейронної мережі на результати розпізнавання, було створено та протестовано 7 варіантів моделей з різними наборами основних параметрів RNN. 4 з них мають 1 рекурентний шар, та 3 інші – двошарові. Для тренування мережі було використано 50 епох.

По результатам тестування (табл. 2) бачимо, що використання двох RNN шарів для вирішення такої задачі дає недостатньо великий приріст в точності, хоча збільшує кількість параметрів мережі майже вдвічі (Модель 2 та Модель 6). Використання двох шарів разом з розміром схованого шару 16, взагалі дає навіть гірший результат у 99.67%, ніж та сама модель з 1 RNN шаром – 99.72%. Саме

Таблиця 2

Результати тестування моделей рекурентної нейронної мережі для вибору оптимального набору параметрів

Модель для дослідження	Кількість RNN шарів	Параметр hidden_size	Загальна кількість параметрів	Точність розпізнавання, %
Модель 1	1	4	193	**97.95**
Модель 2	1	8	465	**99.53**
Модель 3	1	12	833	**99.63**
Модель 4	1	16	1297	**99.72**
Модель 5	2	4	313	**99.06**
Модель 6	2	8	897	**99.62**
Модель 7	2	16	2929	**99.67**

Модель 4 показала найкращу точність розпізнавання серед розглянутих моделей. Зазначимо, що при іншому розбитті датасету, з використанням 85% на навчання, така модель показала найкращу точність у 99.86%, що перевищує точність підходів, описаних у розглянутих роботах [3; 4].

Проведемо тестування роботи двох найбільш оптимальних реалізацій рекурентної нейронної мережі, а саме Моделі 2 та Моделі 4, для оцінки кращої моделі у контексті швидкодії. Результати тестування наведено у табл. 3. Було виконано 2 сесії перевірки кожної моделі, з використанням тільки процесору CPU та з використанням графічної відеокарти. Заміри було проведено на процесорі CPU Intel Core I9-9900KF 3.60GHz, в якості GPU було використано Nvidia RTX 2060.

Таблиця 3

Результати тестування швидкодії моделей з використанням CPU та GPU

	CPU		**GPU**	
	M.2	**M.4**	**M.2**	**M.4**
Час завантаження моделі з натренованими параметрами, мс	7.14	**60.74**	1151.67	1226.58
Час розпізнавання першого семплу, мс	2.95	**2.96**	130.65	132.89
Середній час розпізнавання одного семплу, мс	2.64	**2.68**	1.02	1.05
Споживання пам’яті під час виконання, Мб	0.25	**0.27**	1632.12	1635.45

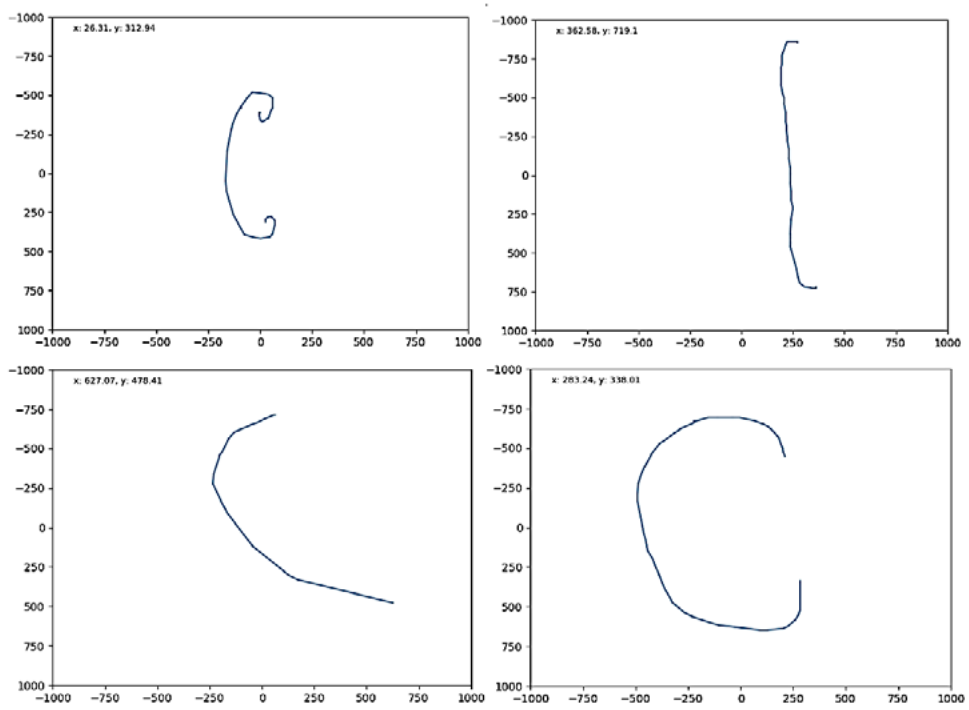

Рис. 2. Приклади тестування розпізнавання жесту “curve_downLeft”, зверху – невірний результат розпізнавання, знизу – вірний

Як бачимо з результатів тестування, Модель 2 та Модель 4, при використанні тільки процесора, показують дуже незначну різницю у часі розпізнавання одного семпла, незважаючи на те, що Модель 4 має в 3 рази більшу загальну кількість параметрів. Бачимо, що це повпливало лише на час початкового завантаження моделі. Порівнюючи отриманий час роботи з використанням відеокарти, бачимо зменшення часу розпізнавання у 2.6 рази, проте час завантаження моделі у пам'ять більший ніж 1 секунда, а також достатньо великий розмір необхідної пам'яті для зберігання моделі у відеокарті, є значними недоліками при врахуванні умов поставленої задачі. Тож за результатами цього тестування, найоптимальнішою побудовою рекурентної нейронної мережі було обрано Модель 4.

Аналізуючи якість роботи побудованої мережі, видно, що більшість невдало розпізнаних семплів з датасету мають певні яскраво виражені відмінності від інших написань того самого жесту, які, скоріш за все, не зустрілись мережі під час тренування. Такі приклади наведені у верхній частині рис. 2, в нижній частині приклади успішного розпізнавання жестів того самого класу, також з нетиповим стилем написання.

Також було проведено перевірку моделі з власними написаннями жестів, перевірено можливість розпізнавати дуже деформовані жести. Як результат, мережі важко дається розпізнавання закручених кінців, різноманітних зайвих кінчиків, які не зустрічались в жестах з тренувального датасету. В таких випадках найчастіше мережа видає низькі показники ймовірності для всіх класів. Такі проблеми унікальних стилей вирішуються розширенням датасету, можливим використанням аугментації для тренування.

Висновки. В даній роботі було побудовано рекурентну нейронну мережу для вирішення задачі з розпізнавання рукописних жестів. Запропоновано декілька моделей з різним набором параметрів, обрано найбільш оптимальну модель на основі точності розпізнавання, кількості параметрів та швидкодії. Обрана модель показала точність 99.72% на датасеті SIGN, при цьому виконуючи вимоги по швидкодії та обмеженості обчислювальних потужностей, які накладає поставлена задача. Середній час розпізнавання одного семпла з використанням процесора склав 2.68 мс. Для перевірки стійкості моделі було перевірено її роботу на різного роду критичних випадках деформованих жестів, проаналізовано невдалі випадки розпізнавання мережі, більшість з яких пов'язано з дуже унікальними стилями написання певних жестів, що не зустрічались моделі в тренувальних даних. За результатами тестування було виділено стилі які необхідно додати до тренувального датасету для покращення точності побудованої моделі. Окрім збільшення різноманітності тренувальних даних, подальшими покращеннями мережі може бути перенесення її на більш низький програмний рівень, наприклад на мову C++, для ще більшої оптимізації використання ресурсів та часу роботи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Fewer and fewer people today write by hand using a pen, pencil or brush. What are the reasons? Is this a positive or negative development, 2017. URL: https://ieltsdata.org/fewer-fewer-people-today-write-hand-using-pen-pencil-brush (дата звернення: 20.08.2023).

2. On-Line Database SIGN-OnDB, 2010. URL: https://www-intuidoc.irisa.fr/en/base-de-donnees-en-ligne-sign-ondb (дата звернення: 20.08.2023).

3. Fuccella V., Isokoski P., Martin B. Gestures and widgets: performance in text editing on multi-touch capable mobile devices, 2013. URL: https://dl.acm.org/doi/10.1145/2470654.2481385 (дата звернення: 20.08.2023).

4. Delaye A., Anquetil. E. HBF49 feature set: A first unified baseline for online symbol recognition, 2013. URL: https://www.sciencedirect.com/science/article/abs/pii/S0031320312003317 (дата звернення: 20.08.2023).

5. Understanding RNN and LSTM, 2019. URL: https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e (дата звернення: 20.08.2023).

6. Рекурентні співвідношення, 2020. URL: http://matfiz.univ.kiev.ua/informatics/lectures/Theme3_2.htm (дата звернення: 20.08.2023).

7. Gated Recurrent Units (GRU), 2020. URL: https://d2l.ai/chapter_recurrent-modern/gru.html (дата звернення: 20.08.2023).

REFERENCES:

1. Fewer and fewer people today write by hand using a pen, pencil or brush. What are the reasons? Is this a positive or negative development. (2017). Retrieved from https://ieltsdata.org/fewer-fewer-people-today-write-hand-using-pen-pencil-brush (Date of access: 20.08.2023).

2. On-Line Database SIGN-OnDB. (2010). Retrieved from https://www-intuidoc.irisa.fr/en/base-de-donnees-en-ligne-sign-ondb (Date of access: 20.08.2023).

3. Fuccella V., Isokoski P., Martin B. Gestures and widgets: performance in text editing on multi-touch capable mobile devices. (2013). Retrieved from https://dl.acm.org/doi/10.1145/2470654.2481385 (Date of access: 20.08.2023).

4. Delaye A., Anquetil. E. HBF49 feature set: A first unified baseline for online symbol recognition. (2013). Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0031320312003317 (Date of access: 20.08.2023).

5. Understanding RNN and LSTM. (2019). Retrieved from https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e (Date of access: 20.08.2023).

6. Rekurentni spivvidnoshennia. (2020). Retrieved from http://matfiz.univ.kiev.ua/informatics/lectures/Theme3_2.htm (Date of access: 20.08.2023) [in Ukrainian].

7. Gated Recurrent Units (GRU). (2020). Retrieved from https://d2l.ai/chapter_recurrent-modern/gru.html (Date of access: 20.08.2023).