Chychkarov, Y. A., Zinchenko, O. V., & Fesenko, M. A. (2023). Recognition of handwritten letters and numbers using deep learning neural networks. *Actual Issues of Modern Science. European Scientific e-Journal, 26*, 43–53. Ostrava. (In Ukr.)

DOI: 10.47451/inn2023-12-02

The paper is published in Crossref, ICI Copernicus, BASE, Zenodo, OpenAIRE, LORY, Academic Resource Index ResearchBib, J-Gate, ISI International Scientific Indexing, ADL, JournalsPedia, Scilit, EBSCO, Mendeley, and WebArchive databases.

Yevhen A. Chychkarov, Doctor of Technical Sciences, Professor, Department of Artificial Intelligence, State University of Information and Communication Technologies. Kyiv, Ukraine.
ORCID 0000-0002-4362-5129

Olha V. Zinchenko, Doctor of Technical Sciences, Associate Professor, Department Head, Department of Artificial Intelligence, State University of Information and Communication Technologies. Kyiv, Ukraine.
ORCID 0000-0002-3973-7814

Maksym A. Fesenko, Ph.D. in Technical Sciences, Associate Professor, Department of Artificial Intelligence, State University of Information and Communication Technologies. Kyiv, Ukraine.
ORCID 0000-0001-8218-4154

Recognition of Handwritten Letters and Numbers Using Deep Learning Neural Networks

Abstract: This article discusses several variants of convolutional neural network architecture for recognizing isolated handwritten Latin or Ukrainian letters and numbers that have been trained using synthetic datasets of two types, built on the basis of a set of handwritten and italic fonts or a CoMNIST dataset. A comparison of the recognition results of several variants of images containing handwritten letters and numbers using models with different architectures showed that an increase in the number of convolutional layers leads to a decrease in the frequency of erroneous character recognition. The size of the training dataset significantly affects the reliability of character recognition. The data sets used in the paper contained from 192 to 2304 samples per class. The limit on the number of samples per class that provided acceptable recognition accuracy was about 1,500 images per class. Reducing the sample by reducing the number of samples per class leads to a significant decrease in recognition accuracy (from 90% of the accuracy of recognizing elements of real labels to 40–60% with a 4-fold decrease in the sample size).

Keywords: handwriting recognition, Ukrainian letter recognition, Latin letter recognition, convolutional neural networks, CNN, deep learning, image processing.

Євген Анатолійович Чичкарьов, доктор технічних наук, професор, кафедра штучного інтелекту, Державний університет інформаційно-комунікаційних технологій. Київ, Україна.
ORCID 0000-0002-4362-5129

Ольга Валеріївна Зінченко, доктор технічних наук, доцент, завідувач, кафедра штучного інтелекту, Державний університет інформаційно-комунікаційних технологій. Київ, Україна.
ORCID 0000-0002-3973-7814

Максим Анатолійович Фесенко, кандидат технічних наук, доцент, доцент, кафедра штучного інтелекту, Державний університет інформаційно-комунікаційних технологій. Київ, Україна.
ORCID 0000-0001-8218-4154

Розпізнавання рукописних літер і цифр з використанням нейронних мереж глибокого навчання

Анотація: У цій статті розглядаються кілька варіантів архітектури згорткових нейронних мереж для розпізнавання ізольованих рукописних латинських або українських літер і цифр, які були навчені за допомогою синтетичних наборів даних двох типів, побудованого на основі набору рукописних і курсивних шрифтів або набору даних CoMNIST. Порівняння результатів розпізнавання декількох варіантів зображень, що містять рукописні літери та цифри, за допомогою моделей з різною архітектурою показало, що збільшення кількості згорткових шарів призводить до зниження частоти помилкового розпізнавання символів. Розмір навчального набору даних істотно впливає на надійність розпізнавання символів. Набори даних, використані в роботі, містили від 192 до 2304 зразків на клас. Межа кількості зразків на клас, яка забезпечила прийнятну точність розпізнавання, склала біля 1500 зображень на клас. Зменшення вибірки за рахунок зменшення кількості вибірок на клас призводить до значного зниження точності розпізнавання (з 90% точності розпізнавання елементів реальних написів до 40-60% при 4-кратному зменшенні обсягу вибірки).

Ключові слова: розпізнавання рукописного тексту, розпізнавання українських літер, розпізнавання латинських літер, згорткові нейронні мережі, CNN, глибоке навчання, обробка зображення.

Вступ

Оптичне розпізнавання символів — це технологія, яка сьогодні широко використовується. процес класифікації зображень символів, які відібрані на вихідному цифровому зображенні, за відповідними зразками (*Chaudhuri et al., 2017*).

Інформаційні технології, засновані на оптичному розпізнаванні символів (друкованих або рукописних), дозволяють вирішувати багато практичних завдань (*Li et al., 2018*).

У більшості випадків сучасні системи оптичного розпізнавання базуються на нейронних мережах глибокого навчання (*Rajavelu et al., 1989*; *Bai et al., 2014*). Для обробки зображень широко використовуються згорткові нейронні мережі (CNN). Це один із найпопулярніших типів глибоких нейронних мереж, який можна використовувати для ефективного розпізнавання символів, присутніх на зображенні (*Maitra et al., 2015*).

Розпізнавання рукописних символів є складнішим завданням проти друкованими формами символів. Рукописні літери та цифри, написані різними авторами, не ідентичні, а різняться з різних аспектів, як-от розмір і форма. Численні варіації стилів написання окремих символів ускладнюють завдання розпізнавання. Подібність різних форм символів, перекриття та взаємозв'язок сусідніх символів ще більше ускладнюють завдання розпізнавання символів.

Ця робота присвячена дослідженню можливостей розпізнавання латинських або українських рукописних літер і цифр та дослідженню впливу технології та особливостей формування повністю або частково синтетичного набору даних на результати розпізнавання.

Результати

1. Огляд літератури

Для вирішення задач оптичного розпізнавання широко використовуються згорткові нейронні мережі. Властивості таких мереж роблять їх дуже зручним засобом для вирішення проблем комп'ютерного зору, зокрема, для розпізнавання зображень букв або цифр.

В останні роки завдання розпізнавання символів різних алфавітів — арабського, російського, казахського, китайського та інших — привернуло значну увагу (*Bilgin Taşdemir, 2021*; *Nurseitov et al., 2021b*; *Abdelrahman et al., 2020*; *Ullah & Jamjoom, 2022*; *Jeevitha et al, 2022*; *Gannetion et al., 2022*).

Наочний приклад особливостей цього завдання – результати, наведені в роботі (*Zhang, 2015*), в якої розглянуто розпізнавання китайських рукописних символів. Автори (*Zhang, 2015*) продемонстрували, що згорткові нейронні мережи можуть вибирати характеристики вищого рівня, коли глибина згорткового шару збільшується. За думкою (*Zhang, 2015*), додавання згорткового шару покращує продуктивність більше, ніж додавання додаткового щільного шару. В цій роботі було досягнуто 97,3% точності при класифікації 200 класів і 95,5% точності при класифікації 3755 символів.

У численних дослідженнях, присвячених розпізнаванню рукописних символів, є досвід використання досить складних архітектур нейронних мереж. В роботі (*Tapotosh Ghosh et al., 2021*) для класифікації 231 класів різних рукописних символів Bangla на основі набору даних CMATERdb декілька варіантів згорткових нейронних мереж з наступним результатом: після 50 епох для архітектури InceptionResNetV2 була досягнута найкраща точність (96,99%). Архітектури DenseNet121 і InceptionNetV3 також продемонстрували відмінну точність розпізнавання (96,55 і 96,20% відповідно). Швидкість навчання була встановлена на 0,001, в якості функції помилки було використано категоріальну крос-ентропію.

Порівняння поведінки згорткових нейронних мереж щодо поведінки відносно розпізнавання рукописних символів були проведені в (*Aicha Korichi et al., 2022*). Архітектури VGG і ResNet дали близькі результати в точності розпізнавання: використовуючи архітектуру ResNet, вдалося досягти найкращого результату з показником точності розпізнавання 98,57%, для архітектури VGG-16 було досягнуто результату 97,14%.

У роботі (*Balaha et al., 2021*) була відзначена більш висока точність розпізнавання при використанні глибшої архітектури нейронної мережі CNN. Але підвищення точності розпізнавання стало можливим лише за допомогою аугментації вхідних даних.

В роботі (*Ghosh et al., 2020*) для розпізнавання рукописних символів Bangla була використана архітектура MobileNet. Вона забезпечила досить хороші результати: 96,46% точності розпізнавання 231 класу (171 складений, 50 основних і 10 цифр), 96,17% точності в 171 складеному класі символів, 98,37% точності в 50 основних класах символів і 99,56% точності в 10 класах цифр.

Вплив попереднього навчання моделей розпізнавання символів досить неоднозначний. За даними (*He et al., 2018*) досягнуті результати для моделей з випадковою ініціалізацією не відрізняються від результатів для моделей з попереднім навчанням для

порівнянної кількості епох. За даними (*Albattah & Albahli, 2022*) моделі, навчені з нуля, як правило, дають кращі результати в порівнянні з попередньо навченими моделями в розпізнаванні рукописних символів арабської мови. На думку (*Albattah & Albahli, 2022*) менш складні моделі CNN є менш точними, але мають більш високу швидкість класифікації та навчання (і навпаки).

З технічної точки зору архітектури нейронних мереж, які використовують попереднє навчання, були створені для класифікації кольорових зображень різних розмірів. Таким чином, щоб використовувати існуючі бібліотеки та можливості попереднього навчання для багатьох наборів даних (наприклад, EMNIST Letters 28×28) зображення у відтінках сірого (одноканальні) повинні бути перетворені в кольорові зображення (трьохканальні) (*Gibrael Al Amin Abo Samra & Hadi Oqaibi, 2021*). Зокрема, для модулів ResNet (ResNet50, ResNet101, ResNet152 або друга версія) з пакету tensorflow/keras потрібне вхідне зображення розміром не менше 32×32×3 (*ResNet…, 2023*).

Для розпізнавання символів казахської та російської мов також є досвід використання архітектури MobileNet (*Nurseitov et al., 2021a*). Деякі результати розпізнавання українських літер та цифр також представлені в (*Чичкарьов та ин., 2023*; *Chychkarov & Zinchenko, 2023*).

Збільшення обсягу навчального набору даних для всіх розглянутих архітектур призвело до підвищення точності розпізнавання. За даними (*Чичкарьов та ин., 2023*) точність розпізнавання реальних написів в межах 80-90% була досягнута при розмірі навчальної вибірки не менше 700, а краще більше 1500 зображень на клас. За даними (*Chychkarov & Zinchenko, 2023*), точність розпізнавання тестового набору в діапазоні 99,2–99,6% було отримано при навчанні на наборі даних достатнього обсягу (було досліджено архітектури VGG16, ResNet, MobileNet).

Для досліджень технологій розпізнавання рукописного латинського алфавіту стандартом де-факто став набір даних EMNIST. Для класифікації зображень цього набору було запропоновано багато різних варіантів архітектури нейронної мережі. Але для українського алфавіту немає EMNIST-подібних наборів даних. Крім того, для використання можливостей пакету tensorflow/keras потрібні зображення символів розміром не менше 32×32×3, що потрібує або створення нового набору даних, або перетворення відомих. . Для розпізнавання кириличного тексту відомо кілька наборів даних зображень рукописних літер (наприклад, на Kaggle), а також певний досвід використання різних класифікаторів і нейромережевих технологій для їх розпізнавання, але порівняльні дослідження технологій і результатів для них фрагментарні.

Таким чином, результати відомих досліджень не віддають однозначної переваги будь-якої з архітектур згорткових нейронних мереж для вирішення завдання розпізнавання рукописних символів. Важливим аспектом технології розпізнавання є характеристики вихідного набору даних. Точність розпізнавання для всіх варіантів досліджених архітектур зростає зі збільшенням обсягу навчальної вибірки. Однак оптимальні параметри аугментації вихідного набору даних, можливості його повної або часткової генерації, їх вплив на точність розпізнавання залишаються незрозумілими.

2. Формування набора даних для навчання

Для навчання моделі для розпізнавання було створено два типи наборів даних.

Перший був створений виключно шляхом друку зображень букв і цифр з використанням відповідного шрифту. Був створений набір рукописних та курсивних шрифтів з латинськими та українськими гліфами (всього 48 шрифтів). З урахуванням подальшої аугментації було створено декілька варіантів наборів даних, які містили від 2 до 48 зображень кожного символу. Наприклад, для набору українських літер при генерації 32 зображень на символ загальний обсяг набору даних склав 116 736 зразків (10 класів цифр, по 33 класи великих та малих літер, 32 зображення на клас, 48 шрифтів).

Другий тип набору даних був побудований з використанням зображень із набору даних CoMNIST, який містить літери латинського та російського алфавіту в форматі RGBA розміром 278x278. Але цей набір даних обмежений, оскільки він практично не містить малих літер, а також не містить специфічних українських літер. Крім того, кількість зображень різних літер дещо відрізняється. При побудові повного набору даних до колекції рукописних зображень було додано згенеровані набори пропущених великих і малих літер, а також згенеровані набори других малих літер українського алфавіту та цифр. Для латинських літер генерувались лише зображення малих літер та цифр. Усі зображення були перетворені у формат RGB 32x32, або 64x6, або 128x128 пікселів. Кількість зображень коливалася від 2, 4, 8 зображень на символ із набору CoMNIST або 2–16 згенерованих зображень/символів. Кількість зображень кожного типу було підібрано таким чином, щоб з урахуванням аугментації набір даних був близьким до збалансованого. Для створення зображень літер та цифр було використано набір, який містив 80 рукописних або курсивних шрифтів. Загальний обсяг базової версії побудованої навчальної вибірки як для українського, так і для латинського варіантів складав понад 130 тис. зображень (з розрахунку 4 зображення на символ), а тестової — понад 26 тис. зображень. Цей обсяг набору даних не дуже відрізняється від обсягу відомого набору даних EMNIST Letters, який містить змішані малі та великі літери (26 класів і загалом 145 600 зразків). Для створення або трансформації зображень із літерами чи цифрами було використано бібліотеку Pillow. Тестовий набір даних генерувався окремо з використанням тих самих шрифтів, конкретні шрифти і параметри генерації обиралися випадковим чином. Обсяг набору тестових даних становив близько 15-20% від обсягу навчального.

Можливість формування набору даних зі збільшеною кількістю зображень була досягнута за рахунок використання функції Image Data Generator пакету tensorflow (три варіанти трансформації зображення символу: випадковий поворот, трансформація нахилу, трансформація масштабування).

3. Попередня обробка та розпізнавання зображень

Алгоритм попередньої обробки та розпізнавання зображень, які містили написи з літерами або цифрами, наведено в додатку (*Рисунок 1*).

Для виділення областей зображень, що містили літери або цифри (які потім розпізнавалися), було використано інструменти з бібліотеки OpenCV.

Безпосередньо для розпізнавання вибрані області інтересу вирізалися з вихідного зображення, до них знову застосовувалася бінаризація, після чого отримані зображення окремих символів (без розширення або інших спотворень) масштабувалися до відповідного розміру (32x32x3, або 64x64x3, або 128x128x3).

Дослідження технології розпізнавання проводилися з трьома типами архітектур згорткових нейронних мереж (було використано пакет tensorflow/keras): 1) ResNet або ResNetV2, 2) MobileNet або MobileNetV2, 3) DenseNet. Навчання моделей для всіх варіантів архітектур проводилось з використанням оптимізатора Adam, швидкість навчання була встановлена на рівні 0,0001, а кількість епох навчання становила 50.

4. Експериментальні результати і їх обговорення

Усі використані під час навчання архітектури забезпечували точність розпізнавання елементів навчальної вибірки в діапазоні 95–99%.

Приклад результату розпізнавання для буквених і цифрових написів (для напису українською) показано в додатку (*Рисунок 2*), де показано вибрані області інтересу та результати розпізнавання.

Збільшення кількості параметрів нейронної мережі за рахунок використання більш глибокої архітектури в більшості випадків призвело до підвищення точності розпізнавання. Однак при розпізнаванні символів на реальних написах різко виявилася різниця в першу чергу між параметрами навчального набору даних щодо можливості розпізнавання символів. Обрана архітектура моделі теж впливає на результати розпізнавання, але вплив обсягу навчального набору даних виявився більш виразним.

Збільшення кількості епох навчання не призвела до зміни результатів. Експерименти зі зміною алгоритму оптимізації порівняно з Adam не дали жодного покращення в точності та надійності розпізнавання реальних зразків.

Для оцінки точності розпізнавання попередньо було створено декілька написів, що містять великі та малі літери та цифри (прямо написані від руки). Розмір навчального набору даних істотно впливає на надійність розпізнавання символів. Генерація 1536 зображень на букву або цифру (32 зображення на кожен символ для 48 типів шрифтів) фактично є межею прийнятної точності розпізнавання. Зменшення обсягу навчальної вибірки призводило до значного зниження точності розпізнавання (від 100% точності до 40-60% при зменшенні вибірки в 4 рази). Але збільшення розміру вибірки призводить до помітного збільшення часу, витраченого на навчання моделі. Збільшення роздільної здатності зображень навчальних зразків мало вплинуло на результати через насиченість. Вибір методу побудови набору даних (з або без CoMNIST) фактично не вплинув на цей висновок. Збільшення розміру навчального набору даних для всіх розглянутих архітектур призвело до незначного підвищення точності розпізнавання (наприклад, з 96% до 99% при чотирикратному збільшенні розміру навчального набору). Точність розпізнавання реальних написів на рівні 80-90% була досягнута при розмірі навчальної вибірки щонайменше 700, а краще більше 1500 зображень на клас (або не менше 4 зображень на символ для набору даних на основі CoMNIST).

При використанні частково згенерованого набору даних збільшення розміру навчальної вибірки призводить до незначного підвищення точності розпізнавання (подібно до повністю згенерованого набору даних). Наприклад, для DenseNet121 збільшення кількості зображень з 2 до 4, а потім 8/символ забезпечило підвищення точності розпізнавання реальних зображень з 79% до 83%, а потім до 88%. Перехід від архітектури DenseNet121 до архітектур DenseNet169 і DenseNet201 також призводить до

підвищення точності розпізнавання, але для невеликих вибірок ефект більш виражений, ніж для повномасштабних. Подібні висновки можна зробити для архітектур сімейства ResNetV2.

Загалом, порівнюючи досягнуту точність розпізнавання реальних зображень і швидкість навчання моделі, найкращу продуктивність забезпечили моделі сімейства DenseNet або ResNetV2. Експерименти зі зміною алгоритму оптимізації порівняно з Адамом не дали покращення точності та надійності розпізнавання реальних зразків. Збільшення кількості епох навчання моделі понад вказану також не змінило результатів.

Для усіх розглянутих варіантів архітектури моделей досягнута точність розпізнавання тестового набору даних в діапазоні 99,2–99,6%, якщо навчати класифікатор на наборі даних достатнього обсягу. Збільшення кількості зображень літер і цифр у навчальному наборі даних для всіх розглянутих архітектур призводило до підвищення точності розпізнавання. Приклад результатів експерименту для моделі з архітектурою MobileNet наведено в додатку (*Рисунок 3*).

Точність розпізнавання реальних написів з точністю 80-90% була досягнута при розмірі навчальної вибірки не менше 700, а краще більше 1500 зображень на клас. Приклад результатів експерименту для моделі з архітектурою MobileNetV2 наведено в додатку (*Рисунок 4*).

Варіація параметрів трансформацій, які використовувалися для доповнення, також помітно впливає на результати розпізнавання: деформація або поворот зображення більш ніж на 10-15% збільшує частоту помилок.

Збільшення роздільної здатності зображень літер та цифр в навчальному наборі даних досить слабко вплинуло на точність розпізнавання через насиченість.

Для встановлення можливого впливу роздільної здатності зображень на точність розпізнавання було побудовано декілька моделей з навчальним набором даних з роздільною здатністю 32x32, 64x64, 128x128.

Точність розпізнавання на навчальному наборі даних слабко зростала зі збільшенням розміру зображень (*Рисунок 5а*).

Розміри зображень навчального набору даних досить слабко вплинула на точність розпізнавання елементів реальних написів. Приклад результатів експериментів зі згенерованими наборами даних і різними варіантами моделі наведено в додатку (*Рисунок 5б*).

При збільшенні розміру зображень навчального набору даних з 32x32 до 128x128 пікселів було досягнуто зниження частки помилок розпізнавання з 18,0% до 11,4% (модель з архітектурою ResNet152V2). Однак для моделей з архітектурою MobileNet або MobileNetV2 при збільшенні розміру зображень частка помилок розпізнавання практично не змінилася. Однак збільшення розмірів зображень для навчання моделі стосовно всіх варіантів досліджених архітектур призводило до значного зростання часу навчання.

Побудова моделі з архітектурою InceptionResNetV2 (необхідна роздільна здатність зображень навчального набору даних не менше 75x75x3, хоча насправді модель навчалася на зображеннях 128x128x3) не привела до помітного підвищення точності розпізнавання.

Висновок

У роботі розглянуто декілька варіантів архітектури згорткових нейронних мереж для розпізнавання ізольованих рукописних цифр та українських літер.

Результати розпізнавання різних зображень, що містять букви та цифри, порівнювали на моделях з різною архітектурою.

Показано можливість навчання згорткових нейронних мереж за допомогою синтетичного набору даних, побудованого на основі рукописних або курсивних шрифтів. Розмір навчального набору даних істотно впливає на надійність розпізнавання символів. Набори даних, використані в роботі, містили від 192 до 2304 зразків на клас.

Зменшення обсягу набору даних за рахунок зменшення кількості зображень на клас призвело до значного зниження точності розпізнавання (з 90% точності розпізнавання елементів реальних написів до 40-60% при 4-кратному зменшенні обсягу набору даних). Нижня межа обсягу набора даних, яка забезпечує прийнятну точність розпізнавання, становила біля 1500 символів на клас. Значне збільшення обсягу набору даних понад 2300 зображень на клас забезпечило слабке підвищення точності та надійності розпізнавання, але призвело до значного збільшення часу навчання моделі.

Збільшення роздільної здатності зображення з 32x32x3 до 128x128x3 навчального набору даних у більшості випадків не призвело до підвищення надійності розпізнавання реального зображення.

Конфлікт інтересів

Автори заявляють, що конфлікту інтересів немає.

Список джерел інформації:

Abdelrahman, A., Hamada, M., & Nurseitov, D. (2020). Attention-based fully gated CNN-BGRU for Russian handwritten text. *Journal of Imaging*, *6*(12), 141. https://doi.org/10.3390/jimaging6120141

Aicha Korichi, Slatnia Sihem, Tagougui Najiba, Zouari Ramzi, & Aiadi Oussama. (2022). Recognizing Arabic handwritten literal amount using convolutional neural networks. *Artificial Intelligence and Its Applications*, 153–165. https://doi.org/10.1007/978-3-030-96311-8_15

Albattah, W., & Albahli, S. (2022). Intelligent Arabic handwriting recognition using different standalone and hybrid CNN architectures. *Application Science*, *12*, 10155. https://doi.org/10.3390/app121910155

Bai, J., Chen, Z., Feng, B., & Xu, B. (2014). Image character recognition using deep convolutional neural network learned from different languages. *IEEE International Conference on Image Processing (ICIP)*, 2560-2564. https://doi.org/10.1109/ICIP.2014.7025518

Balaha, H. M., Ali, H. A., Saraya, M., & Badawy, M. (2021). A new Arabic handwritten character recognition deep learning system (AHCR-DLS). *Neural Computering Applications*, *33*(11), 6325–6367. https://doi.org/10.1007/s00521-020-05397-2

Bilgin Taşdemir, E. F. (2021). Online Turkish handwriting recognition using synthetic data. *Avrupa Bilim ve Teknoloji Dergisi*, *32*, 649–656. https://doi.org/10.31590/ejosat.1039846

Chaudhuri, A., Mandaviya, K., Ghosh, S. K., & Badelia, P. (2017). Optical character recognition systems for different languages with soft computing. Studies in Fuzziness and Soft Computing, 352. Springer. https://doi.org/10.1007/978-3-319-50252-6

Chychkarov, Y., & Zinchenko, O. (2023). Handwritten Ukrainian character recognition using a convolutional neural networks and synthetic dataset. *MoMLeT+DS 2023: 5th International Workshop*

on Modern Machine Learning Technologies and Data Science, 109–121. Lviv. https://ceur-ws.org/Vol-3426/paper9.pdf

Gannetion, L., Wong, K. Y., Lim, P. Y., Chang, K. H., & Abdullah, A. F. L. (2022). An exploratory study on the handwritten allographic features of multi-ethnic population with different educational backgrounds. *PloS One*, *17*(10), e0268756. https://doi.org/10.1371/journal.pone.0268756

Ghosh, T., Abedin, M. M.-H.-Z., Chowdhury, S. M., Tasnim, Z., Karim, T., Reza, S. M. S., Saika, S., & Yousuf, M. A. (2020). Bangla handwritten character recognition using MobileNet V1 architecture. *Bulletin of Electrical Engineering and Informatics*, *9*(6), 2547–2554. https://doi.org/10.11591/eei.v9i6.2234

Gibrael Al Amin Abo Samra & Hadi Oqaibi. (2021). An optimized deep residual network with a depth concatenated block for handwritten characters classification. *Computers, Materials & Continua*, *680*, 1–28. https://doi.org/10.32604/cmc.2021.015318

He, K., Girshick, R.B., & Dollár, P. (2018). Rethinking imageNet pre-training. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 4917–4926, https://doi.org/10.1109/ICCV.2019.00502

Jeevitha, D., Muthu, S., Nila, I., & Santhoshi, V. (2022). Handwritten letter recognition using artificial intelligence. *International Journal for Research in Applied Science and Engineering Technology*, *10*, 2752–2758. https://doi.org/10.22214/ijraset.2022.42949

Li, H., Wang, P., & Shen C. (2018). Toward end-to-end car license plate detection and recognition with deep neural networks. *IEEE Transactions on Intelligent Transportation Systems*, *20*(3), 1126–1136. https://doi.org/10.1109/TITS.2018.2847291

Maitra, D. S., Bhattacharya, U., & Parui, S. K. (2015). CNN based common approach to handwritten character recognition of multiple scripts. *3th International Conference on Document Analysis and Recognition (ICDAR)*, 1021–1025. https://doi.org/10.1109/ICDAR.2015.7333916

Nurseitov, D., Bostanbekov, K., Kanatov, M., Alimova, A., Abdallah, A., & Abdimanap, G. (2021a). Classification of handwritten names of cities and handwritten text recognition using Various Deep Learning Models. *arXiv preprint*, *arXiv*, 2102.04816. https://doi.org/10.25046/aj0505114

Nurseitov, D., Bostanbekov, K., Kurmankhojayev, D., Alimov,a A., Abdallah, A., & Tolegenov, R. (2021b). Handwritten Kazakh and Russian (HKR) database for text recognition. *Multimedia Tools Applications*, *80*, 33075–33097. https://doi.org/10.1007/s11042-021-11399-6

Rajavelu, A., Musavi, M. T., & Shirvaikar, M. V. (1989). A neural network approach to character recognition. *Neural Networks*, *2*(5), 387–393. https://doi.org/10.1016/0893-6080(89)90023-3

ResNet and ResNetV2. (2023) https://keras.io/api/applications/resnet/#resnet50-function

Tapotosh Ghosh, Min-Ha-Zul Abedin, Hasan Al Banna, Nasirul Mumenin, & Mohammad Abu Yousuf. (2021). Performance analysis of state of the art convolutional neural network architectures in Bangla handwritten character recognition. *Pattern Recognit. Image Anal*, *31*(1), 60–71. https://doi.org/10.1134/S1054661821010089

Ullah, Z., & Jamjoom, M. (2022). An intelligent approach for Arabic handwritten letter recognition using convolutional neural network. *PeerJ Computer Science*, *8*, e995. https://doi.org/10.7717/peerj-cs.995

Zhang, Y. (2015). Deep convolutional network for handwritten Chinese character recognition. *University of Stanford*, 1–8. http://cs231n.stanford.edu/reports/2015/pdfs/zyh_project.pdf

Чичкарьов, Є., Зінченко, О., Балалаєва, О., Сергієнко, А., & Ковальов, О. (2023). Розпізнавання рукописних українських літер та цифр з використанням синтетичного набору даних та згорткових нейронних мереж. *Grail of Science*, *23*, 241–253. https://doi.org/10.36074/grail-of-science.23.12.2022.36

Додатки

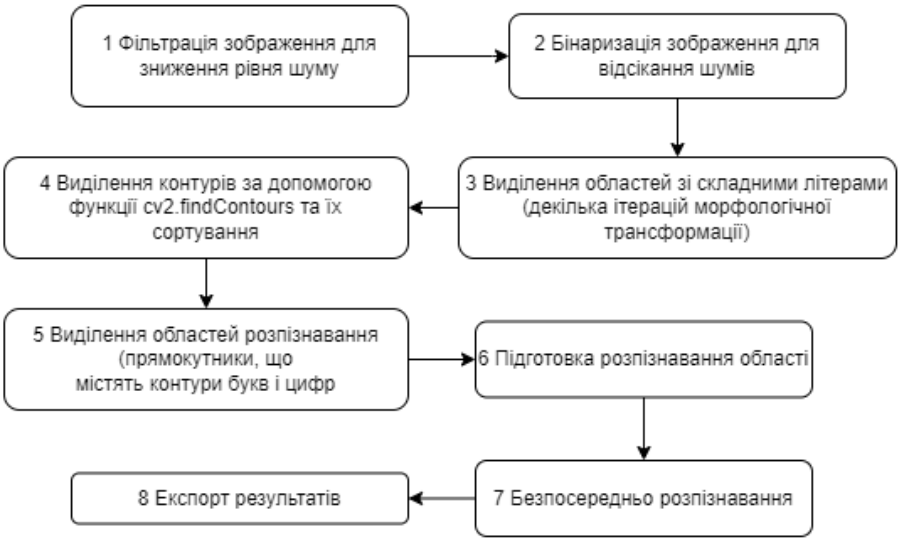
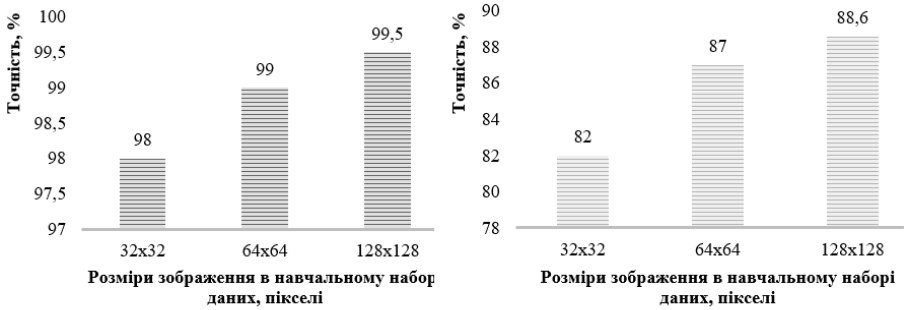


Рисунок 1. Алгоритм попередньої обробки та розпізнавання зображень

Рисунок 2. Приклад результатів розпізнавання за допомогою нейронної мережі VGG16 (в даному випадку всі букви та цифри розпізнаються точно)

Рисунок 3. Приклад впливу розміру навчального набору даних на досягнуту точність розпізнавання (архітектура MobileNet)

Рисунок 4. Помилки розпізнавання реальних написів залежно від розміру навчального набору даних (архітектура MobileNetV2, набор даних із зображеннями 32x32x3)



а) навчальна точність (MobileNetV2) б) точність розпізнавання (ResNet152V2)

Рисунок 5. Приклад впливу роздільної здатності зображень навчального набору даних на досягнуту точність розпізнавання (архітектура MobileNetV2 і ResNet152v2)