

Нагорний П.В., студент групи ПІ-181

Національний університет «Чернігівська політехніка», [inn5665@gmail.com](mailto:inn5665@gmail.com)

Науковий керівник: Дорош М.С., докт. техн. наук

Національний університет «Чернігівська політехніка», [maruyaya5536@gmail.com](mailto:maruyaya5536@gmail.com)

## МЕТОДИКА РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ НА ОСНОВІ ТЕХНОЛОГІЇ ГЛИБИННОГО НАВЧАННЯ

Актуальний напрямок діджиталізації сучасного суспільства полягає в активному впровадженні систем розпізнавання об'єктів. Такі системи досить ефективні в [1]:

- ❖ маркетингу – автоматизований моніторинг брендової присутності в графічних матеріалах соціальних мереж;
- ❖ правозахисній сфері – технології розпізнавання правопорушників на камерах відеоспостереження;
- ❖ медицині – пошук характерних діагностичних симптомів на графічних зображеннях;
- ❖ енергетичному комплексі – визначення неполадок в електричних мережах.

Саме тому на даний момент розроблено та запропоновано велику кількість методик розпізнавання зображень. Запропоновані алгоритми демонструють різний рівень ефективності, застосовні в різних задачах.

Основний підхід методології розпізнавання зображень – технології машинного навчання. На основі великої кількості зображень із встановленими об'єктами (тренувального набору) спеціальні алгоритми (нейронні мережі) визначають основні закономірності, властиві тому чи іншому зображенню, які дозволяють ідентифікувати об'єкт на зображенні. Приклади нейронних мереж для розпізнавання тексту: EAST, CRNN, STN-net/SEE [2-3].

**EAST.** Підхід EAST є досить простим, але в той же час достатньо ефективним. Мережа, використана в цьому підході, – U-Net – стандартна для розпізнавання ознак, які відрізняються за розміром. EAST додано до бібліотеки open-CV. Основний недолік підходу полягає в тому, що EAST дозволяє ідентифікувати лише сам текст, а не його елементи (слова, літери).

**CRNN.** Підхід CRNN використовує конволюційно-рекурентну нейронну мережу. Розпізнавання слів відбувається в три етапи. На першому етапі відбувається пошук та розподіл за ознаками окремих розділів тексту. На другому етапі глибокий двонаправлений LSTM відшукує зв'язки між символами. На третьому етапі відбувається транскрипційний аналіз, під час якого видаляються зайві порожні та подвоєні символи. Специфіка підходу CRNN полягає у високій ефективності за умови наявного текстового лексикону, та нижчій успішності – без.

**STN-net/SEE.** Даний підхід вирізняється тим, що тренування відбувається на наборі графічних зображень без обмежувальних рамок, а лише з текстовими примітками. В той же час тренування такої системи є занадто складним, передбачає творчий підхід.

Зазначені підходи мають свої переваги та недоліки. В [3] розглянуто дещо інший підхід, побудований в системі RubiML (багатошаровий перцептрон). Такий підхід дозволяє поліпшити вказані проблеми. Архітектура системи RubiML наведена на рис. 1. Її детальний опис наведено в [4], не будемо на цьому детально зупинятися. Тому відразу розглянемо особливості архітектури нейронної мережі зазначеного підходу.

Пропонується створення нейронної мережі з трьох груп Dense нейронних шарів, із шаром активації Leaky ReLU, та шаром Dropout, який виступатиме в ролі регулятизатора. Також вихідний шар створює додатковий шар, який забезпечує Softmax активацію. Таким чином, запропонована нейронна мережа є 4-шаровою.

Математично шари активації Leaky ReLU та Softmax визначаються наступними формулами:

➤ Leaky ReLU –  $f(x) = \begin{cases} ax, & x < 0; \\ x, & x \geq 0; \end{cases}$

➤ Softmax –  $S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$ ,  $i, j = [0, \dots, K]$ , де K – розмірність вектору y.

Наступний крок – встановлення розміру партії тренувального набору. Здійснюється за допомогою алгоритму Adam Optimizer, який поєднує Momentum та RMS Prop. Цей алгоритм використовує глобальну швидкість навчання з метою розміру кроку.

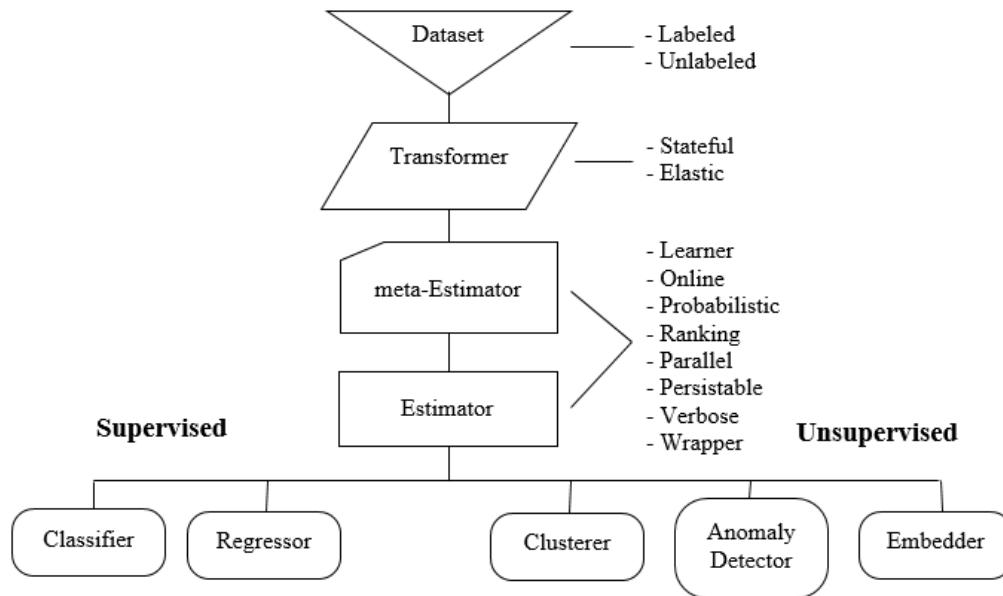


Рис. 1. Архітектура системи розпізнавання зображень\*

\*джерело: на основі [4]

Практичний аналіз наведеного алгоритму [3] підтверджує його високу ефективність. Точність розпізнавання чисел є високою (сягає 99 %), що властиве сучасним алгоритмам розпізнавання зображень. Проте, точність розпізнавання кирилических літер також висока (98 %) та наближається до точності розпізнавання чисел, що свідчить про особливу ефективність запропонованого алгоритму.

Таким чином, одним з актуальних сучасних напрямків використання машинного навчання є розпізнавання графічних об'єктів (використовується в маркетингу, правозахисній сфері, медицині тощо). Розроблено значну кількість підходів до розпізнавання зображень: EAST, CRNN, STN-net/SEE – кожен з яких характеризується своїми перевагами та недоліками. Запропонована система розпізнавання зображень по типу багатосарового перцептрона, побудована на базі RubiML, є ефективною як в розпізнаванні чисел, так і кирилических літер, тому може бути застосована на практиці.

## Перелік посилань

1. Топ-5 сфер применения систем распознавания объектов. URL: <https://habr.com/ru/company/toshibarus/blog/433544/> (дата звернення: 08.12.2021).
2. Bengio Yoshua, LeCun Yann, Hinton Geoffrey. Deep Learning. Nature. 2015. Vol. 521 (7553). Pp. 436–444.
3. Холявікіна Т.В., Резаєв Я.О., Харченко О.О. Система розпізнавання зображень з нейромережевою архітектурою на основі технології глибинного навчання. Наукоємні технології. 2020. № 1(45). С. 54-66.
4. RubiML. URL: <https://rubi.ml.com/> (дата звернення: 09.12.2021).