

Музика І.П.,
здобувач другого (магістерського) рівня вищої освіти,
Катуніна О.С.,
кандидат економічних наук, доцент,
Київський національний економічний Університет імені Вадима Гетьмана

ЕКСПЕРИМЕНТАЛЬНИЙ ПІДХІД ДО ПОБУДОВИ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ У ЕКОНОМІЧНИХ ЗАДАЧАХ АНАЛІЗУ ВІЗУАЛЬНОГО КОНТЕНТУ

В економічних системах, зокрема, компаніях з виробництва та дистрибуції предметів споживання, значна частина поточних задач управління базується на аналізі різноманітних неструктурованих даних, що інтенсивно генеруються в поточних умовах цифрової економіки. Аналіз візуального контенту на базі використання інтелектуальних модельних технологій глибокого навчання використовується для розв'язання широкої низки задач, наприклад, дослідження споживчих переваг, поведінки та лояльності покупців до бренду та конкретних товарів, оптимізації розміщення товарів на полицях магазинів та управління запасами, автоматичної перевірки якості продукції, виявлення порушення регламентів у виробничих процесах та дефектів готової продукції, моніторингу маркетингової активності та ефективності рекламних кампаній, оптимізації розміщення товарів, персоналізації пропозицій тощо.

Розпізнавання та класифікація зображень за допомогою нейромережових моделей у задачах комп'ютерного зору є актуальним напрямом дослідження та моделювання в економіці. У сучасному світі методи машинного навчання, в тому числі нейронні мережі є важливим чинником для підвищення ефективності бізнес процесів та інновацій. Розпізнавання зображень використовується у різноманітних сферах, наприклад у медицині, безпеці, e-commerce та багато інших, де потрібна автоматична класифікація неструктурованої візуальної інформації [1]. Тема класифікації зображень є актуальною, оскільки відіграє важливу роль у цифровізації різних операцій, персоналізації користування різними пристроями, оптимізації ресурсів і дозволяє помітити те, що людина може пропустити.

В економічних задачах аналізу візуального контенту серед широкого різноманіття архітектур нейронних мереж найбільш ефективними є повнозв'язні нейронні мережі (FCNN) та згорткові нейронні мережі (CNN).

У нашому дослідженні використано датасет CIFAR-10, яким зазвичай користуються компанії, що займаються виробництвом або дистрибуцією споживчих товарів, з метою навчання та демонстрації моделей. Дані CIFAR-10 містять 60000 зображень, розділених на 10 категорій (такі як "літаки", "автомобілі", "тварини" і т. д.). При навчанні моделей класифікації зображень датасет CIFAR-10 є відомим бенчмарком для моделей комп'ютерного зору, він використовується для швидкої перевірки гіпотез та алгоритмів, перш ніж перейти до кастомізованих датасетів товарів компанії. Набір даних CIFAR-10 складається з 60000 кольорових зображень 32×32 у 10 класах, по 6000 зображень на клас. Є 50000 навчальних зображень і 10000 тестових зображень [2]. Зображення містять різні об'єкти реального світу, наприклад: птахи, автомобілі, коні, коти, олені, жаби, кораблі, літаки тощо. Картинки досить невеликих розмірів, тому можуть бути важкими для розпізнавання навіть людям.

Для оцінювання якості моделей в процесі навчання та порівняння їхньої прогностичної спроможності було використано метрики точності (Accuracy), що визначається як частка правильно класифікованих зображень [3] та втрати категоріальної крос-ентропії (Categorical Cross-Entropy Loss) використовується для вимірювання різниці між прогнозованим розподілом ймовірностей та фактичним розподілом для задач багатокласової класифікації. Чим нижчі значення цієї метрики, тим кращою є модель.

У дослідженні реалізовано комплекс експериментів з метою визначення найкращої архітектури глибокої нейромережевої моделі для розв'язання задачі класифікації зображень. Розглянуто повнозв'язні нейронні мережі (FCNN) та згорткові (CNN). У практичних експериментах було порівняно 10 моделей: 5 моделей повнозв'язних мереж (FC1–FC5) та 5 CNN архітектур (CNN1–CNN5).

Повнозв'язна нейронна мережа (FCNN), яку також називають щільним шаром (dense layer), має таку особливість: кожен нейрон поточного шару

з'єднується з усіма нейронами наступного шару [3]. Базова модель FC1 має один прихований шар з 512 нейронами та функцією активації ReLU, і вихідний шар з 10 нейронами та функцією активації Softmax для класифікації на 10 класів. До цієї базової архітектури були додані різні методи регуляризації. Модель FC2 включає шар Dropout з ймовірністю 0.3, який випадково вимикає 30% нейронів під час навчання, що допомагає запобігти перенавчанню. Модель FC3 використовує BatchNormalization, який нормалізує виходи попереднього шару, стабілізуючи процес навчання та дозволяючи використовувати вищі швидкості навчання. Модель FC4 комбінує обидва методи, маючи шар BatchNormalization перед прихованим шаром і шар Dropout після нього, що дозволяє поєднати переваги обох підходів. Модель FC5 є глибшою мережею з двома прихованими шарами (1024 і 512 нейронів), що дозволяє вивчати складніші закономірності в даних, але вимагає ефективної регуляризації для запобігання перенавчанню.

Згорткові нейронні мережі (CNN) спеціально розроблені для обробки зображень і використовують згорткові шари для виявлення просторових ієрархій ознак [4]. Базова модель CNN1 починається зі згорткового шару (Conv2D) з 32 фільтрами розміром 3x3 і функцією активації ReLU, потім йде шар MaxPooling2D для зменшення просторових розмірів карт ознак, шар Flatten для перетворення двовимірних карт ознак у одновимірний вектор, і, нарешті, повнозв'язаний шар (Dense) з 512 нейронами та вихідний шар з 10 нейронами і функцією активації Softmax. Модель CNN2 розширює базову архітектуру, додаючи другий згортковий шар і шар Dropout з ймовірністю 0.3 для регуляризації. Модель CNN3 включає шари BatchNormalization після кожного згорткового шару, що стабілізує навчання та допомагає уникнути проблем з градієнтами. Модель CNN4 комбінує BatchNormalization і Dropout, маючи шари BatchNormalization після кожного згорткового шару і шар Dropout після першого повнозв'язаного шару, що поєднує переваги обох методів. Модель CNN5 є глибшою мережею з двома блоками згорткових шарів (32 і 64 фільтри), кожен з яких супроводжується шарами MaxPooling2D і Dropout, що дозволяє вивчати більш складні та абстрактні ознаки зображень. Результати всіх експериментів наведені у таблиці 1.

Таблиця 1 – Результати проведених експериментів для класифікації зображень

№ експерименту	Тип модифікації базової моделі	Точність на навчанні	Точність на тесті	Помилка на навчанні	Помилка на тесті	Схильність моделі до перенавчання
1	FC1: Базова (512 -> 10)	0.525	0.465	1.343	1.526	Висока похибка на тестовому наборі, можлива недотренованість
2	FC2: +Dropout(0.3)	0.409	0.442	1.644	1.585	Погіршення тренувальної точності, але кращий баланс
3	FC3: +BatchNorm	0.842	0.508	0.470	2.248	Сильне перенавчання
4	FC4: Dropout+BatchNorm	0.672	0.536	0.950	1.548	Краще узагальнення, але ще є перенавчання
5	FC5: Більша мережа (1024->512->10)	0.610	0.510	1.096	1.459	Помірне перенавчання, невеликий вигреш у тестовій точності
6	CNN1: Базова (Conv->Pool->Flatten->Dense)	0.937	0.647	0.212	1.377	Сильне перенавчання
7	CNN2: +Dropout(0.3)	0.980	0.697	0.062	1.607	Дуже сильне перенавчання, велика розбіжність помилок
8	CNN3: +BatchNorm	0.987	0.670	0.039	2.712	Критичне перенавчання, модель переважно запам'ятовує дані
9	CNN4: Dropout+BatchNorm	0.935	0.707	0.193	1.179	Кращий баланс між точністю та узагальненням
10	CNN5: Глибша (Conv2D-> Conv2D-> Pool-> Conv2D-> Co...	0.950	0.755	0.144	1.039	Оптимальний баланс, але ще є простір для покращень

Джерело: розроблено автором

Основна відмінність між FCNN та CNN полягає в тому, як вони обробляють вхідні дані. FCNN перетворюють зображення у плоский вектор, втрачаючи інформацію про просторові зв'язки між пікселями, тоді як CNN зберігають цю просторову структуру, дозволяючи моделі вивчати ієрархічні ознаки, такі як краї, текстури та форми. За результатами експериментів, що наведено в таблиці 1 визначено, що CNN мають суттєву перевагу над FCNN у задачах класифікації зображень, особливо на складних наборах даних, таких як CIFAR-10. Повнозв'язні мережі (FC), навіть із додатковою регуляризациєю (Dropout, BatchNorm), не здатні подолати бар'єр у 54 % точності на тестових даних і при цьому часто страждають від значного перенавчання. Найкращою моделлю виявилась FC4 (Dropout+BatchNorm) із точністю 53,6%. У згорткових мережах (CNN) прослідковується суттєве покращення, навіть базова CNN1 має точність 64,7%. Найкращі результати показала глибока CNN5 із точністю 75,5% на тестових даних. Вона демонструє найвищу здатність узагальнювати завдяки комбінуванню багаторівневої згортки, пулінгу і регуляризациї.

Отже, за результатами дослідження ми встановили, що моделі глибокого навчання в економічних задачах аналізу візуального контенту є доцільним та перспективним інструментом для автоматичного виділення й класифікації складних ознак із хорошою точністю і здатністю до узагальнення.

Список використаних джерел

1. Riley, J. (2023). AI Powers the Digital Economy. *ACM Ubiquity*. <https://doi.org/10.1145/3636478>
2. Cuevas, F., Mazzeo, P. L., & Bruno, A. (Eds.). (2023). *Digital Image Processing: Latest Advances and Applications*. IntechOpen.
3. Shin, S., Shin, D. H., & Kim, J. (2024). Implementation and Comparison of CNN Models on CIFAR-10 Dataset. У *1st World Congress on Industrial Engineering and Operations Management* (с. 496-502). IEOM Society International. <https://doi.org/10.46254/wc01.20240168>
4. Pant, Y., Shah, G., Ojha, R., Thapa, R., & Bhatta, B. (2023). Comparison of CNN Architecture of Image Classification Using CIFAR10 Datasets. *International Journal on Engineering Technology*, 1(1), 37–52. <https://doi.org/10.3126/injet.v1i1.60898>