DOI: https://doi.org/10.15276/ict.02.2025.01
УДК 004.93

Сучасні підходи до підвищення ефективності розпізнавання зображень при обмежених наборах даних

Міхалев Кирило Богданович[1)]
Аспірант каф. Інформаційних систем
ORCID: https://orcid.org/0009-0007-8058-4569; kirill.mikhalev.9035@gmail.com
Ніколенко Анантолій Олександрович[1)]
Канд. техніч. наук, доцент каф. Інформаційних систем
ORCID: https://orcid.org/0000-0002-9849-1797; nikolenko@op.edu.ua,
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У сучасних системах комп'ютерного зору точність моделей глибокого навчання значною мірою залежить від обсягів та різноманітності навчальних даних. Проте у багатьох прикладних сферах збирання великих розмічених датасетів є складним, дорогим або іноді навіть недосяжним завданням. Це обумовлює потребу у використанні підходів, які дозволяють покращити результати моделей навіть за умов обмежених вибірок. Одним із найперспективніших рішень є аугментація даних, що передбачає створення додаткових навчальних прикладів шляхом трансформації наявних зображень. У даній роботі проведено практичний експеримент з використанням датасету CIFAR-10, де для моделювання умов обмежених ресурсів було використано лише 10 000 прикладів із 50 000 доступних. Для навчання застосовано одну згорткову нейронну мережу, а результати було порівняно між моделлю, яка була натренована без будь-яких перетворень, та моделлю, що використовувала базову аугментацію. До переліку застосованих методів увійшли горизонтальне віддзеркалення, випадкове кадрування із додаванням полів, а також зміни яскравості, контрасту та насиченості кольорів. Отримані результати показали, що застосування навіть базових прийомів аугментації дозволяє суттєво підвищити стійкість моделі до варіацій у вхідних даних. Якщо модель без додаткових трансформацій демонструвала схильність до перенавчання та нижчу точність на тестовій вибірці, то додавання аугментації дало відчутний приріст у показниках узагальнюючої здатності. Зокрема, графіки навчання засвідчили зменшення різниці між навчальною та тестовою точністю, що свідчить про ефективніший баланс між підлаштуванням до даних та здатністю працювати з новими прикладами. Важливою відмінністю проведеного дослідження є акцент саме на умовах обмежених вибірок, що робить його релевантним для практичних задач, де доступ до великих обсягів маркованих даних ускладнений. Отримані результати не лише підтверджують ефективність класичної аугментації, а й підкреслюють її потенціал як базового інструменту, який може бути подальше поєднаний з іншими методами, наприклад, напівконтрольованим навчанням або генерацією синтетичних даних. Таким чином, робота демонструє не лише теоретичну, але й прикладну цінність аугментації для підвищення точності моделей комп'ютерного зору. Дане дослідження є відправною точкою для подальшого дослідження впливу аугментації на нейронні мережі в здачах розпізнавання зображень.

Ключові слова: комп'ютерний зір; аугментація даних; згорткові нейронні мережі; класифікація зображень; обмежені вибірки; перенавчання; штучний інтелект

Актуальність. Сучасні системи комп'ютерного зору активно впроваджуються у різні сфери – від медицини та промисловості до автономного транспорту й відеоаналітики. Ефективність таких систем значною мірою визначається якістю та обсягом навчальних даних. Однак у більшості практичних випадків збирання великих розмічених наборів є складним, дорогим або технічно неможливим. Це створює проблему недостатності маркованих даних, яка негативно впливає на точність і узагальнювальну здатність моделей глибокого навчання. У цьому контексті актуальним стає використання методів, здатних компенсувати нестачу даних. До них належать аугментація, що дозволяє генерувати варіативні приклади на основі існуючих зображень, а також напівконтрольоване навчання, яке поєднує обмежену кількість маркованих даних з великими масивами немаркованих. Такі підходи сприяють підвищенню ефективності моделей без необхідності значних витрат на ручну розмітку. Таким чином, дослідження методів подолання нестачі маркованих зображень має важливе наукове та прикладне значення.

Метою дослідження є аналіз і систематизація сучасних методів підвищення точності моделей комп'ютерного зору за умов обмежених обсягів навчальних даних. Основна увага приділяється вивченню можливостей аугментації та напівконтрольоване навчання для покращення узагальнювальної здатності згорткових нейронних мереж та зменшення ризику перенавчання.

У сучасних системах комп’ютерного зору проблема обмеженості маркованих даних залишається однією з найважливіших. Для ефективного навчання глибоких нейронних мереж, зокрема згорткових (CNN), необхідні великі й різноманітні вибірки. Проте в реальних умовах, особливо у сферах медицини, автономного транспорту чи промислового контролю якості, збір та розмітка таких даних вимагають значних фінансових і часових ресурсів, а іноді є практично неможливими. Це зумовлює потребу у використанні методів, здатних підвищити ефективність моделей при мінімальних обсягах маркованих вибірок [1].

Одним із ключових підходів є аугментація даних. Вона полягає у штучному розширенні навчального набору за допомогою модифікацій існуючих зображень. Найчастіше використовують геометричні перетворення (обертання, масштабування, віддзеркалення), зміни кольору та освітлення (варіації яскравості, контрасту, насиченості), а також додавання шумів. Такі трансформації дозволяють збільшити різноманітність прикладів, що змушує модель виділяти стійкі ознаки об’єкта замість «запам’ятовування» конкретних зображень. Це суттєво знижує ризик перенавчання і підвищує узагальнювальну здатність [2].

Практичні дослідження демонструють, що навіть прості форми аугментації можуть забезпечувати відчутний приріст точності. Наприклад, при класифікації зображень з набору CIFAR-10 базова згорткова мережа досягає точності близько 75 %, тоді як із використанням аугментації цей показник може зростати на 3-7 %. У складніших архітектурах, таких як ResNet, приріст може бути ще вищим. Окрім аугментації, важливим напрямом є напівконтрольоване навчання, яке поєднує невелику кількість маркованих прикладів із великими масивами немаркованих [3]. Одним із найпоширеніших методів у цьому підході є псевдолейблінг, коли модель, попередньо навчена на обмеженому наборі, автоматично генерує мітки для нових даних. Отримані псевдомітки включаються у тренувальний набір, збільшуючи його розмір і покращуючи точність моделі. У поєднанні з аугментацією цей метод дозволяє досягати результатів, близьких до тих, що забезпечуються при використанні великих повноцінно розмічених датасетів.

Таким чином, поєднання методів аугментації та полуконтрольованого навчання відкриває нові можливості для розвитку прикладних систем комп’ютерного зору [4].

На рис. 1 наведений приклад того як може бути аугментовано зображення.

Рис. 1. **Приклад аугментації зображення:**
1 – оригінальне зображення; 2 – поворот; 3 – колірні спотворення;
4 – відзеркалення; 5 – виріз; 6 – Ґауссівський шум

У прикладі з набором даних Flowers-102 [5] було використано кілька поширених методів аугментації. Одним із них є випадкове обертання зображення на кут до тридцяти градусів у будь-який бік, що дозволяє моделі залишатися стійкою до змін ракурсу об’єкта.

Інший метод – випадкова зміна яскравості, контрасту та насиченості кольорів, завдяки чому зображення відображають різні умови освітлення або параметри камери. Також застосовано горизонтальне віддзеркалення, яке особливо корисне для об’єктів, що можуть зустрічатися у дзеркальному вигляді. Окрім цього, використано випадкове кадрування з подальшим масштабуванням, яке імітує зміни масштабу та положення об’єкта у кадрі. Нарешті, до прикладів було додано гаусівське розмиття, що дозволяє моделі коректно працювати навіть у випадках, коли фото є нечітким або містить шуми [6].

Завдяки поєднанню цих перетворень початкове зображення перетворюється у кілька нових варіантів, які зберігають ключові ознаки квітки, але відрізняються за виглядом. Це робить навчальний набір більш різноманітним і дозволяє згортковій нейронній мережі краще узагальнювати знання, не прив’язуючись до конкретних прикладів.

Також розглянемо вплив аугментації даних на точність моделі. У якості експериментального датасету обрано набір зображень CIFAR-10, який містить 60 000 кольорових зображень розміром 32x32 пікселі, розподілених на 10 класів (літак, автомобіль, птах, кішка, олень, собака, жаба, кінь, корабель, вантажівка) [7].

Для імітації умов нестачі даних, у навчанні використовувалася лише частина повного набору (10 % тренувальних зразків). У якості базової архітектури обрана класична згорткова нейронна мережа з трьома згортковими шарами і двома повнозв'язними шарами у вихідній частині. В даному експерименті використовувались такі підходи аугментації як RandomHorizontalFlip (Горизонтальне віддзеркалення), RandomCrop (випадковий виріз), ColorJitter (Колірні спотворення).

На початкових етапах навчання (перші 3-4 епохи) модель, що навчалася без аугментації, демонструвала вищу точність класифікації зображень на тестовій вибірці. Це пояснюється тим, що при відсутності аугментації модель отримує «чисті» та передбачувані тренувальні приклади, які вона може відносно швидко «запам'ятати» і почати правильно класифікувати аналогічні зображення на тестовому наборі.

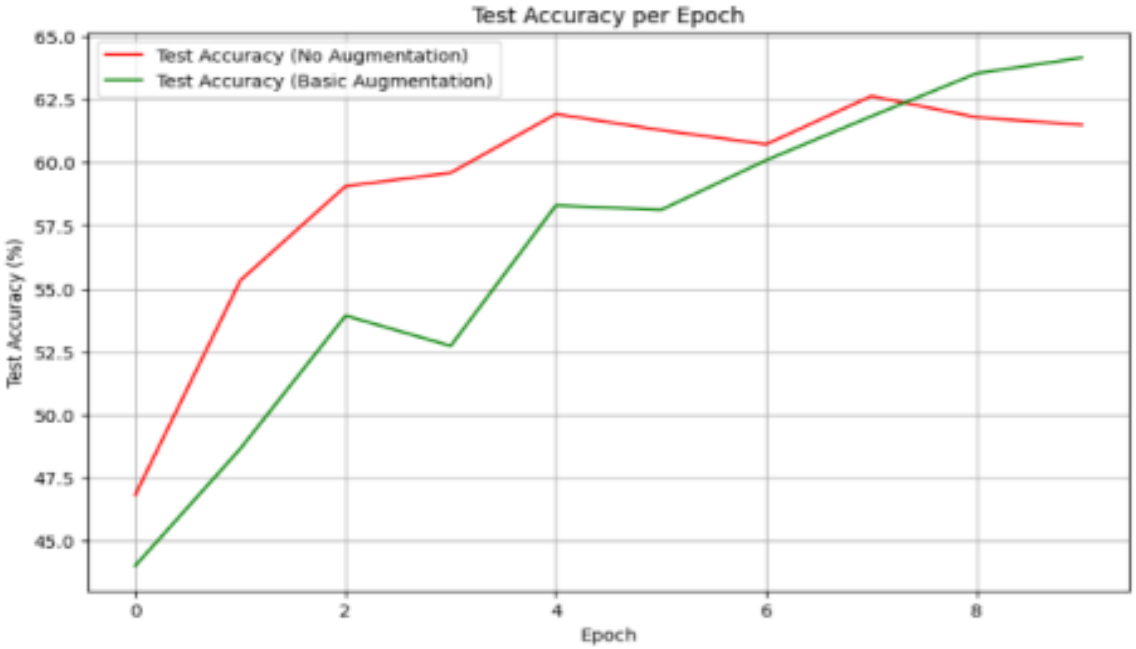

Рис. 2. **Графіки точності для обох моделей**

В умовах обмеженого обсягу даних така поведінка є типовою ознакою початкового перенавчання – модель адаптується до конкретних деталей тренувальної вибірки, проте її здатність до узагальнення залишається низькою.

Натомість у сценарії з базовою аугментацією (випадкове віддзеркалення, вирізи, колірні зміщення) модель зіткнулася зі значно складнішим навчальним середовищем. Введення шуму, геометричних та колірних спотворень у тренувальні зразки ускладнює задачу для моделі, оскільки вона більше не може покладатися на дрібні особливості зображень і змушена шукати більш стійкі та узагальнені ознаки для класифікації. Як наслідок, у перших епохах точність моделі з аугментацією була нижчою, ніж у базового варіанта.

Вже після 4-5 епох навчання крива точності моделі з аугментацією почала наздоганяти, а у фінальних епохах – навіть випереджати результат моделі без аугментації. Це свідчить про поступове формування більш стійких і узагальнених ознак, які дозволили моделі коректніше класифікувати раніше невидані зображення на тестовій вибірці.

Така динаміка є прямим свідченням позитивного впливу аугментації на узагальнюючу здатність моделі. Фінальна точність моделі навченою з методами аугментації на 4% більше ніж точність моделі без аугментації. Загалом, результати експерименту підтвердили, що використання аугментації даних є ефективною стратегією покращення якості моделі в умовах обмеженої вибірки маркованих зображень.

Проведений аналіз підтверджує, що проблема обмежених наборів даних у комп'ютерному зорі залишається однією з ключових у сучасних дослідженнях та практичних застосуваннях.

Згорткові нейронні мережі продемонстрували надзвичайну ефективність у задачах класифікації, сегментації та виявлення об'єктів, проте їхня здатність до узагальнення безпосередньо залежить від достатнього обсягу та різноманітності навчальної вибірки. За умов, коли отримання великих масивів маркованих даних є складним або неможливим, вирішальну роль починають відігравати методи, що дозволяють компенсувати цей дефіцит [8, 9]. Перспективним напрямом розвитку сучасних систем комп'ютерного зору є інтеграція методів аугментації даних із напівконтрольованим навчанням. Традиційна аугментація дозволяє збільшувати різноманітність вибірки шляхом трансформацій зображень, однак її потенціал обмежується лише вже наявними маркованими прикладами.

У той час як напівконтрольоване навчання забезпечує можливість використовувати великі обсяги немаркованих даних, які значно простіше зібрати у реальних умовах. Поєднання цих підходів дозволяє створити багатший та різноманітніший простір ознак, що істотно підвищує узагальнювальну здатність нейронних мереж.

Очікується, що надалі саме комбінація методів – аугментації, напівконтрольованого навчання та, можливо, генеративних моделей – стане ключовим напрямком для досягнення ще більшої точності й надійності в комп'ютерному зорі. Такий підхід дає змогу ефективно використовувати всі доступні ресурси даних, зменшити залежність від дорогого ручного маркування і водночас гарантувати високу стійкість моделей у реальних прикладних системах.

СПИСОК ЛІТЕРАТУРИ

1. LeCun Y., Bottou L., Bengio Y., Haffner P. "Gradient-Based Learning Applied to Document Recognition". *Proceedings of the IEEE*. 1998; 86 (11): 2278–2324. DOI: https://doi.org/10.1109/5.726791.

2. Krizhevsky A., Sutskever I., Hinton G. "ImageNet Classification with Deep Convolutional Neural Networks". *In: Advances in Neural Information Processing Systems* (NeurIPS 2012). 2012; 25: 1097–1105. DOI: https://doi.org/10.1145/3065386.

3. Mumuni A., Mumuni F., Ameyaw J. "Data augmentation: A comprehensive survey of modern approaches". *Machine Learning with Applications*. 2022; 10: 100443. DOI: https://doi.org/10.1016/j.mlwa.2022.100443.

4. Shorten C., Khoshgoftaar T. M. "A survey on Image Data Augmentation for Deep Learning". *Journal of Big Data*. 2019; 6 (1): 60. DOI: https://doi.org/10.1186/s40537-019-0197-0.

5. Li H., Chen J., Wang J., Zhang Y. "Research on a Flower Recognition Method Based on Improved Deep Learning Using the Oxford 102 Flowers Dataset*". Horticulturae*. 2024; 10 (5): 517.

DOI: https://doi.org/10.3390/horticulturae10050517.

6. Wang J., Wang C., He Z., Xu J. “A Comprehensive Survey on Data Augmentation in the Era of Large-Scale Pretrained Models”. *arXiv*. 2024. URL: https://arxiv.org/html/2405.09591v1.

7. Jordan T., Raskutti G., Ramchandran K. “94 % on CIFAR-10 in 3.29 Seconds on a Single GPU”. arXiv preprint. arXiv: 2404.00498. 2024. DOI: https://doi.org/10.48550/arXiv.2404.00498.

8. Alomar K., Munilla J., Padilla P. “Data Augmentation in Classification and Segmentation: A Survey and Future Directions”. *Journal of Imaging.* 2023; 9 (2): 46. DOI: https://doi.org/10.3390/jimaging9020046.

9. Zhou Z.-H. “A Brief Survey on Weakly Supervised Learning”. *National Science Review*. 2018; 5 (1): 44–53. DOI: https://doi.org/10.1093/nsr/nwx105.

DOI: https://doi.org/10.15276/ict.02.2025.01
UDC 004.93

Modern approaches to improving image recognition efficiency with limited datasets

Kyrylo B. Mikhalev[1)]
Postgraduate Student of the Department of Information Systems
ORCID: https://orcid.org/0009-0007-8058-4569; kirill.mikhalev.9035@gmail.com
Anatolii O. Nikolenko[1)]
Candidate of Engineering Sciences, Associate Professor of the Department of Information Systems
ORCID: https://orcid.org/0000-0002-9849-1797; anatolyn@ukr.net
[1)] Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

In modern computer vision systems, the accuracy of deep learning models largely depends on the volume and diversity of training data. However, in many application areas, collecting large labeled datasets is a difficult, expensive, or sometimes even unattainable task. This necessitates the use of approaches that allow improving the results of models even under limited sample conditions. One of the most promising solutions is data augmentation, which involves creating additional training examples by transforming existing images. In this work, a practical experiment was conducted using the CIFAR-10 dataset, where only 10,000 examples out of 50,000 available were used to simulate resource-constrained conditions. A single convolutional neural network was used for training, and the results were compared between the model that was trained without any transformations and the model that used basic augmentation. The list of applied methods included horizontal mirroring, random cropping with the addition of fields, as well as changes in brightness, contrast and color saturation. The results obtained showed that the use of even basic augmentation techniques allows to significantly increase the model's resistance to variations in the input data. If the model without additional transformations demonstrated a tendency to overtraining and lower accuracy on the test sample, then the addition of augmentation gave a noticeable increase in the generalization ability indicators. In particular, the training graphs showed a decrease in the difference between the training and test accuracy, which indicates a more effective balance between adjusting to the data and the ability to work with new examples. An important difference of the conducted study is the emphasis on the conditions of limited samples, which makes it relevant for practical tasks where access to large volumes of labeled data is difficult. The results obtained not only confirm the effectiveness of classical augmentation, but also emphasize its potential as a basic tool that can be further combined with other methods, for example, semi-supervised learning or synthetic data generation. Thus, the work demonstrates not only the theoretical, but also the applied value of augmentation for improving the accuracy of computer vision models. This study is a starting point for further research into the impact of augmentation on neural networks in image recognition tasks.

Keywords: Computer vision; data augmentation; convolution neural networks; image classification; limited samples; overfitting; artificial intelligence