DOI 10.34185/1562-9945-4-147-2023-10
УДК004.932.72'1

Д.В. Солдатенко, Вік.В. Гнатушенко

ПОКРАЩЕННЯ ЕФЕКТИВНОСТІ РОЗПІЗНАВАННЯ СУПУТНИКОВИХ ЗОБРАЖЕНЬ ШЛЯХОМ ВИЗНАЧЕННЯ ОБСЯГУ НАВЧАЛЬНИХ ДАНИХ

Анотація. Розпізнавання супутникових зображень є життєво важливим застосуванням комп'ютерного зору з потенційними варіантами використання в таких сферах, як боротьба зі стихійними лихами, землеробство та міське планування. Це дослідження спрямоване на визначення оптимальної кількості вхідних даних, та підбору оптимальних методів їх аугментації, необхідних для навчання нейронної мережіCNN для розпізнавання супутникових зображень. З цією метою проводиться серія експериментів, щоб дослідити вплив кількості вхідних даних на кілька показників продуктивності, включаючи точність, конвергенцію та узагальнення моделі. Дослідження пропонує кілька методів для визначення точки насичення та пом'якшення наслідків перенавчання. Результати, отримані в цьому дослідження, можуть допомогти в розробці більш ефективних моделей розпізнавання супутниковихзображень.
Ключові слова: нейронна мережа, розпізнавання зображень, супутникові знімки, обробка зображень, аугментація даних, штучний інтелект, згорткова нейромережа, класифікація зображень.

Постановка проблеми. При підготовці даних, які використовуються для навчання нейромережі, в деяких випадках можна зіштовхнутися з їх недоліком, низькою якістю або необхідності в симуляції різноманітних вторинних умов, наприклад, погіршення або зміна погоди. Не завжди є можливість запросити оновлені дані або збільшити їх кількість. Аугментація даних - є одним з варіантів розв'язання проблеми недостачі даних, але занадто покладатися на цей метод також не треба. Занадто сильне збільшення вхідних даних за допомогою аугментації може спричинити перенавчання моделі, що вплине на кінцевий результат. Під час аугментації даних необхідно враховувати, що нейромережі можуть бути чутливі до різних методів аугментації. Наприклад, деякі методи аугментації можуть підвищити різноманітність даних, а інші можуть спричинити перенавчання моделі або погіршення її точності. У деяких випадках може бути корисно використовувати комбінації різних типів аугментації для збільшення різноманітності даних та підвищення робастності моделі до різних вторинних умов.

Враховуючи вищезгадане, використання аугментації даних є корисним інструментом при навчанні нейромереж, але його потрібно застосовувати з обережністю та дотримуватися оптимального балансу між збільшенням різноманітності даних та уникненням перенавчання моделі.

Таким чином, підбор оптимальної кількості аугментованих даних, для визначення точки насичення та пом’якшення наслідків перенавчання нейромережі є актуальною задачею.

Аналіз останніх досліджень і публікацій. Дослідження показують, що використання аугментації даних може підвищити точність моделі та зменшити вплив перенавчання. Проте ми спостерігаємо деякий дисбаланс у розвитку різних підходів. Хоча для різних задач комп'ютерного зору було запропоновано різноманітні архітектури нейромереж і обчислювальна потужність графічних процесорів (GPU) стрімко збільшується, менше уваги приділяється застосуванню методів аугментації даних для генерації якісних тренувальних даних. Основна ідея аугментації даних полягає в підвищені достатності та різноманітності тренувальних даних шляхом створення синтетичного набору даних. Аугментовані дані можуть бути розглянуті як ті, що взяті з розподілу, який близький до реального. У цьому випадку, розширений набір даних може представляти повніші характеристики. Але деякі проблеми залишаються в методах аугментації даних, які використовуються до вхідних зображень.

По-перше, методи доповнення даних можуть бути застосовані в різних типах задач комп'ютерного зору, таких як виявлення об'єктів [1], семантична сегментація [2] та класифікація зображень [3]. Проте виклик полягає в тому, що методи доповнення даних є незалежними від задач. Оскільки операції виконуються одночасно на даних зображень та їх класах, а типи класів різні для різних завдань, методи аугментації даних для задачі виявлення об'єктів не можуть бути безпосередньо застосовані до задачі семантичної сегментації. Це призводить до неефективності та низької масштабованості.

По-друге, не існує теоретичного дослідження з аугментації даних. Наприклад, не існує кількісних стандартів для достатнього розміру тренувальних наборів даних. Розмір згенерованих тренувальних даних зазвичай проєктується згідно з особистим досвідом та розлогими експериментами. Крім того, може існувати парадокс, коли розмір початкового набору даних настільки малий [4], що ми стикаємося з викликом, як згенерувати якісні дані на основі дуже малої кількості даних.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Таким чином, невирішеним є завдання підбору методів аугментації та потрібної кількості даних, які можуть бути найбільш релевантними під базові класи супутникових зображень.

Мета дослідження. Це дослідження спрямоване на визначення оптимальної кількості вхідних даних, для підбору оптимальних методів аугментації даних, які підходять для різних видів вхідних даних. З цією метою проводиться серія експериментів, щоб дослідити вплив кількості вхідних даних на кілька показників продуктивності, включаючи точність, конвергенцію та узагальнення моделі.

Викладення основного матеріалу дослідження. Аугментація даних - це процес штучного збільшення об'єму даних шляхом застосування різноманітних операцій до наявних зразків даних. Для зображень слід виділити два з найбільш популярні методи аугментації даних, та підгрупи що належать до них:

- Позиційна аугментація:
    - Масштабування (scaling);
    - Обрізка (croping);
    - Відображення (flipping);
    - Додавання краю (padding);
    - Повороту (rotation);
    - Переміщення (translation).
- Кольорова аугментація:
    - Яскравість (brightness);
    - Контрастність (contrast);
    - Насиченість (saturation);
    - Відтінок (hue);
    - Додавання шуму (noise).

У цьому дослідженні використовуються деякі з наведених вище методів аугментації, а саме: масштабування, відображення, поворот та додавання шуму, ці методи обрані як одні з найбільш використовуваних. Як нейромережа використовується CNN[5] (згорткова нейронна мережа). Також буди використанні вхідні зображення (RGB), з розмірами 128 на 128 пікселів, до яких застосовується аугментація. Кількість зображень була 16 для навчання та 8 для валідації, для кожного окремого класу. Давайте детальніше розберемо кожен окремий метод аугментації та потім перейдемо до самих експериментів з підбору відповідний, які є сенс використовувати під різні класи зображень.

Розпочнемо з одного з найпростіших в реалізації, зі сторони математичного алгоритму, методів аугментації даних, який можливо використовувати для зображення - відображення, який полягає в дзеркальному відображенні зображення вздовж вертикальної або горизонтальної осей. Цей метод може бути корисним, коли об'єкти на зображенні симетричні відносно потрібної осі, або коли нам потрібні додаткові зображення для підвищення точності моделі.

Математично відображення зображення можна описати так: нехай $I(x,y)$ - початкове зображення з розмірами (H,W), тоді відображене зображення $I'(x',y')$ з розмірами (H,W) можна отримати за допомогою наступної формули (1), для горизонтального відображення:

$$I'(x',y') = I(W - x' - 1, y'), \quad (1)$$

де x' та y' - координати пікселя на відображеному зображенні, W - ширина початкового зображення. Ця формула говорить нам, що кожен піксель на горизонтально відображеному зображенні $I'(x',y')$ можна отримати, обертаючи початкове зображення $I(x,y)$ вздовж вертикальної осі. Приклад горизонтального відображення зображення можна побачити на рис. 1.

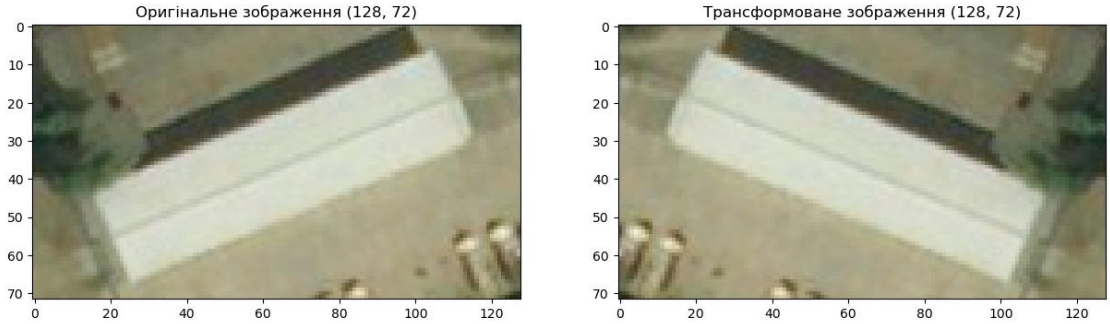

Рисунок 1 - Приклад оригінального та горизонтально відображеного зображень

Формула (2) для вертикального відображення зображення не сильно відрізняється від горизонтального і буде наступною:

$$I'(x',y') = I(x', H - y' - 1), \quad (2)$$

де (x',y') - нові координати пікселя, (x,y) - початкові координати пікселя, H - висота початкового зображення. Як і у першому випадку, з формулою горизонтального відображення, кожен піксель на вертикально відображеному зображенні $I'(x',y')$ можна отримати, обертаючи початкове зображення $I(x,y)$

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

вздовж горизонтальної осі. Приклад використання вертикального відображення зображення можна побачити на рис. 2.

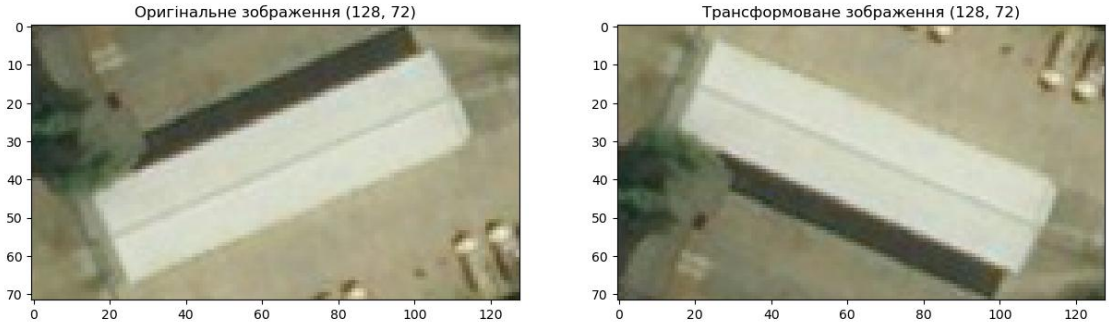

Рисунок 2 - Приклад оригінального та вертикально відображеного зображень

Наведені вище формули відображення використовується для зображень з одним каналом (чорно-білих) або зображень з кількома каналами [6] (RGB, наприклад). У випадку зображень з кількома каналами, ці формули потрібно застосовувати до кожного каналу окремо.

Наступний з методів аугментації даних - поворот. Цей метод також використовується для збільшення розміру навчальної вибірки та покращення здатності моделі до узагальнення, шляхом створення нових зображень з різним кутом огляду.

Математично поворот зображення можна описати так: нехай $I(x,y)$ буде зображенням з координатами пікселів (x,y), де x - номер стовпця, а y - номер рядка. Нехай $I'(x',y')$ буде повернутим зображенням з координатами пікселів (x',y'), де x' - номер стовпця, а y' - номер рядка.

Для обчислення повороту зображення на кут θ за годинниковою стрілкою навколо центру зображення ми можемо використовувати наступні формули:

$$x' = (x - x_c)\cos\theta - (y - y_c)\sin\theta + x_c, \quad (3)$$

$$y' = (x - x_c)\sin\theta + (y - y_c)\cos\theta + y_c, \quad (4)$$

де (x_c,y_c) є координатами центру зображення, (x,y) - координати пікселя на зображенні, а θ - кут обертання в радіанах. Формули (3-4) враховують, що пікселі повинні бути перенесені на нові координати (x',y'), щоб отримати нове зображення, обернуте на кут θ .

Для застосування методу повороту потрібно вибрати випадковий кут обертання θ з певного діапазону, наприклад, від -15 до 15 градусів. Отже, якщо ми маємо зображення x і випадковий кут обертання θ , то зображення з застосованим методом повороту можна записати наступним чином:

$$x_{\text{rot}} = \text{rotate}\,(x, \theta), \tag{5}$$

де $rotate$ - функція, яка застосовує формулу (5) обертання до зображення x на кут θ . Приклад повороту зображення можна побачити на рис. 3.

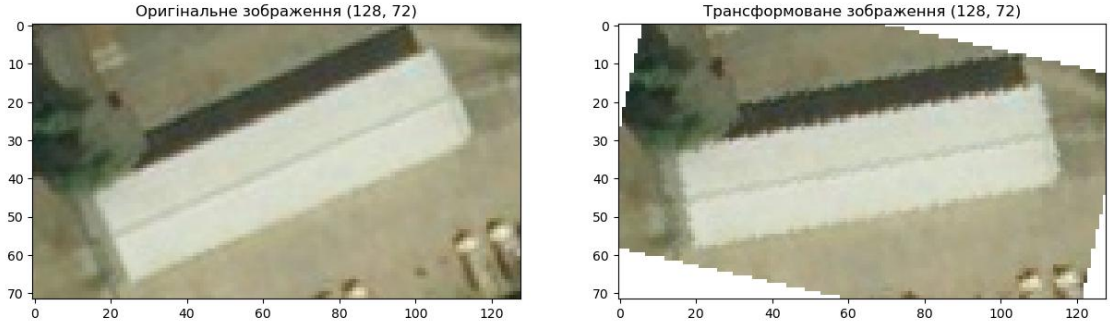

Рисунок 3 - Приклад оригінального та повернутого зображень

Важливо пам'ятати, що після повороту зображення ми можемо втратити частину інформації. Якщо ми повертаємо зображення на більше ніж 90 градусів, ми можемо втратити деякі деталі, такі як текст чи обличчя на зображенні, але у більшості випадків це не стосується супутникових знімків. Тому, при застосуванні методу повороту, ми повинні бути уважні, щоб не втратити важливу інформацію на зображенні.

Зміна масштабу[7] (або ресайзінг) зображення - це метод аугментації даних, що полягає в зміні розміру зображення без зміни його вмісту. Цей метод також може бути корисним для збільшення розміру даних або для адаптації зображення до потрібного розміру, другий варіант більше підходить якщо потрібно зробити дані однакового розміру.

Для зміни масштабу зображення можна використовувати формули білінійної інтерполяції. Для спрощення пояснення, ми розглянемо зміну масштабу зображення вдвічі.

Нехай $I(x,y)$ - початкове зображення з розмірами (H,W), а $I'(x',y')$ - змінене зображення з розмірами $\left(\frac{H}{2},\frac{W}{2}\right)$. Для того, щоб отримати значення пікселя в точці (x',y'), нам потрібно взяти значення пікселів у чотирьох сусідніх точках (x_1,y_1), (x_1,y_2), (x_2,y_1) та (x_2,y_2), де $(x_1,y_1)=(\lfloor x' \rfloor, \lfloor y' \rfloor)$, $(x_1,y_2)=(\lfloor x' \rfloor, \lceil y' \rceil)$, $(x_2,y_1)=(\lceil x' \rceil, \lfloor y' \rfloor)$ та $(x_2,y_2)=(\lceil x' \rceil, \lceil y' \rceil)$. Значення пікселів у цих точках можна обчислити (6) за допомогою лінійної інтерполяції:

$$I'(x',y') = \frac{(x_2-x')(y_2-y')I(x_1,y_1) + (x'-x_1)(y_2-y')I(x_2,y_1) + (x_2-x')(y'-y_1)I(x_1,y_2) + (x'-x_1)(y'-y_1)I(x_2,y_2)}{(x_2-x_1)(y_2-y_1)}, \tag{6}$$

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

де $\lfloor \cdot \rfloor$ - округлення до меншого цілого, а $\lceil \cdot \rceil$ - округлення до більшого цілого. Приклад зміну масштабу зображення, але зі збереженою висотою, для більш наочного приклада, можна побачити на рис. 4.

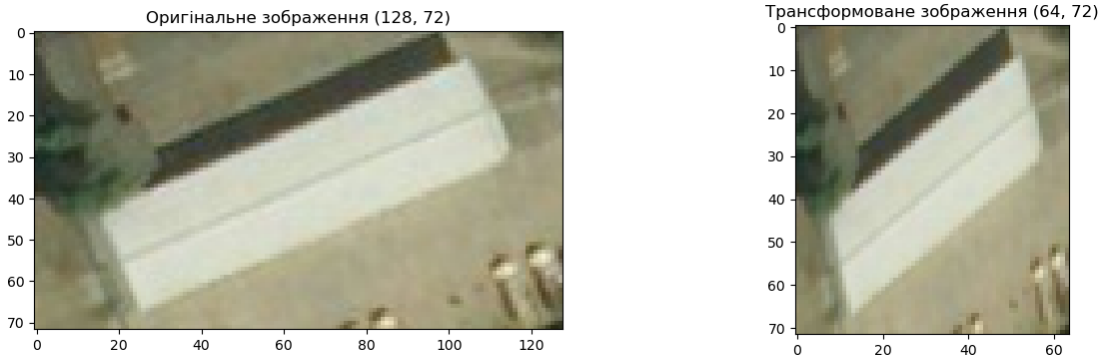

Рисунок 4 - Приклад оригінального та зображення зі зміненим масштабом

Додавання шуму є одним з методів аугментації даних, який може бути використаний для покращення результатів моделей машинного навчання. Гаусівський шум - це один з типів шуму, який може бути доданий до даних. Формула для додавання гаусівського шуму виглядає наступним чином:

$$x_{\text{noisy}} = x + \epsilon \,, \tag{7}$$

де x - оригінальний сигнал, ϵ - гаусівський шум з середнім значенням 0 та стандартним відхиленням σ .

Вибір стандартного відхилення σ залежить від домену задачі та рівня шуму, який потрібно додати до даних. При занадто високому значенні σ , може статися так, що шум буде переважати сигнал, і модель може втратити здатність до правильної класифікації.

Стандартне відхилення гаусівського шуму (8) визначається наступною формулою:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{8}$$

де x_i - значення сигналу, N - кількість значень у сигналі, μ - середнє значення сигналу.

Для кольорових зображень, гаусівський шум додається до кожного пікселя зображення з окремими значеннями $\epsilon_{i,j,c}$, де i та j - координати пікселя, а c - номер кольорового каналу (червоний, зелений, синій).

Отже, формула (8) для додавання гаусівського шуму до кольорового зображення може мати вигляд:

$$x_{\text{noisy}}(i,j,c) = x(i,j,c) + \epsilon_{i,j,c} \quad (9)$$

де $x(i,j,c)$ - значення пікселя на позиції (i,j) та кольоровому каналі c оригінального зображення, а $\epsilon_{i,j,c}$ - гаусівський шум зі стандартним відхиленням σ для кожного кольорового каналу.

У випадку кольорових зображень, формула для генерації гаусівського шуму залишається такою ж, як і для чорно-білих зображень. Тобто, значення шуму $\epsilon_{i,j,c}$ генеруються незалежно (10) для кожного кольорового каналу $(c \in [0,2])$ і для кожного пікселя (i,j) за формулою:

$$\epsilon \sim \mathcal{N}(0, \sigma_c^2) \quad (10)$$

де $\mathcal{N}$ - нормальний розподіл з середнім значенням 0 та стандартним відхиленням σ_c для кожного кольорового каналу.

При виборі значення σ_c для кожного кольорового каналу, слід також враховувати рівень шуму, який ви хочете додати до даних, та специфіку домену задачі.

Додавання гаусівського шуму [8] до даних може бути корисним для зменшення перенавчання моделей, підвищення стійкості до шуму вхідних даних та збільшення різноманітності даних для тренування моделей. Приклад додавання гаусівського шуму до зображення можна побачити на рис. 5.

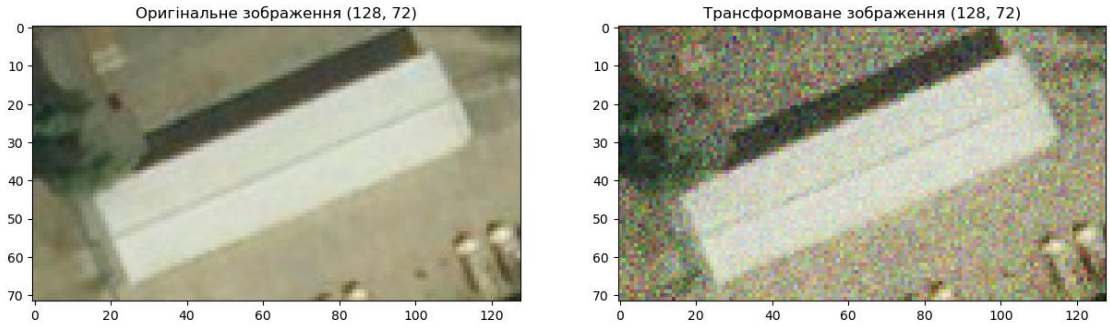

Рисунок 5 - Приклад оригінального та зображення з доданим шумом

Використовуючи перелічені вище методі аугментації, була подвоєна кількість вхідних даних для кожного окремого класу, при використанні кожного окремого метода, та у п'ять разів при використанні усіх одразу. Результати з точністю розпізнавання CNN після навчання на даних, отриманих за допомогою різних методів аугментації наведені у табл. 1.

Перш ніж почнемо розбирати результати - зауважте, що точність може коливатися залежно від того, які методи аугментації використовуються, а також від розміру мережі та інших факторів, які впливають на її навчання. Дані табл

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

1. є лише орієнтовним значенням, отриманими саму з цією нейромережею та вхідними даними.

Таблиця 1

Результати навчання CNN з та без використання аугментованих даних

	Точність розпізнання			
Метод	Вода	Дороги	Будинки	Дерева
Тільки початкові дані	0.85	0.92	0.78	0.81
Масштабування	0.80	0.94	0.76	0.79
Відображення	0.86	0.95	0.82	0.83
Поворот	0.80	0.92	0.81	0.78
Шум	0.85	0.89	0.84	0.80
Найбільш результативні	0.90	0.94	0.86	0.88

Оцінки кожного окремого метода аугментації, та варіанта з використанням усіх методів та їх поєднання:

- Масштабування - не показав значного покращення точності для жодного з класів, у деяких випадках навіть спричинив значну рецесію, тож можна вважати, що використання цього методу не є доцільним у нашому випадку. Однак, в інших випадках результат може бути іншим, тому варто спробувати різні методи аугментації та оцінити їх вплив на точність розпізнавання;
- Відображення - показав покращення точності для кожного з класів, особливо для класу "Дерева". Тому можна вважати, що використання цього методу є доцільним для даної задачі класифікації зображень;
- Поворот - показав невелике покращення точності для деяких класів, але не показав суттєвого впливу на точність для інших класів. Тому використання цього методу може бути доцільним при використанні з іншими методами, але не має особливої доцільності в окремому використанні;
- Шум - показав значне покращення точності тільки для класу "Будинки", в інших випадках не було результату, або була незначна рецесія. Тому використання цього методу, у нашому випадку, може бути доцільним для класу "Будинки", але не має особливого сенсу для інших;
- Найбільш результативні - використання найбільш результативних методів аугментації для кожного окремого класу, визначених у минулих експериментах - показало значне покращення точності для всіх класів порівняно з використанням окремих методів. Тому використання цього способу може бути

доцільним, оскільки це допоможе збільшити вибірку та точність розпізнавання моделі.

В ході проведених експериментів з використанням різних методів аугментації (Масштабування, відображення, поворот, шум та всі разом) для різних класів (вода, дерева, будинки, дороги), було визначено, що використання різних методів може позитивно або негативно впливати на точність розпізнавання нейромережі.

Висновок. Шляхом проведення експериментів з використанням різних методів аугментації було встановлено, що точність розпізнавання моделі може позитивно або негативно залежати від методу, використаного для підвищення кількості даних. Використання метода відображення показало покращення точності для всіх класів, тоді як використання методу масштабування та шуму не дали значних результатів у нашому випаду. Використання повороту може бути доцільним, якщо він поєднується з іншими методами.

Подальше покращення результатів передбачає проведення додаткових експериментів зі збільшеної кількістю методів аугментації та їх комбінацій.

Результати, отримані в цьому дослідження, можуть допомогти у розробці ефективніших і результативних моделей CNN для розпізнавання супутникових зображень.

ЛІТЕРАТУРА

1. Christian L, Lucas T, Ferenc H, Jose C, Andrew C, Alejandro A, Andrew A, Alykhan T, Johannes T, Zehan W, Wenzhe S. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint. 2016.

2. Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

3. Chengyue Gong, Dilin Wang, Meng Li, Vikas Chandra, and Qiang Liu. Keepaugment: A simple information-preserving data augmentation approach. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1055–1064, 2021.

4. Buda Mateusz, Maki Atsuto, Mazurowski Maciej A. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks. 2018;106:249–59.

5. Trishul C, Yutaka S, Johnson A, Karthik K. Project adam: building an efficient and scalable deep learning training system. In: Proceedings of OSDI. 2014. P. 571–82.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

6. Xiaofeng Z, Zhangyang W, Dong L, Qing L. DADA: deep adversarial data augmentation for extremely low data regime classification. arXiv preprint. 2018.
7. Dong W, Zhou N, Paul JC, Zhang X. Optimized image resizing using seam carving and scaling. ACM Transactions on Graphics (TOG). 2009 Dec 1;28(5):1-0.
8. Russo F. A method for estimation and filtering of Gaussian noise in images. IEEE Transactions on Instrumentation and Measurement. 2003 Sep 23;52(4):1148-54.

REFERENCES

1. Christian L, Lucas T, Ferenc H, Jose C, Andrew C, Alejandro A, Andrew A, Alykhan T, Johannes T, Zehan W, Wenzhe S. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint. 2016.
2. Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
3. Chengyue Gong, Dilin Wang, Meng Li, Vikas Chandra, and Qiang Liu. Keepaugment: A simple information-preserving data augmentation approach. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1055–1064, 2021.
4. Buda Mateusz, Maki Atsuto, Mazurowski Maciej A. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks. 2018;106:249–59.
5. Trishul C, Yutaka S, Johnson A, Karthik K. Project adam: building an efficient and scalable deep learning training system. In: Proceedings of OSDI. 2014. P. 571–82.
6. Xiaofeng Z, Zhangyang W, Dong L, Qing L. DADA: deep adversarial data augmentation for extremely low data regime classification. arXiv preprint. 2018.
7. Dong W, Zhou N, Paul JC, Zhang X. Optimized image resizing using seam carving and scaling. ACM Transactions on Graphics (TOG). 2009 Dec 1;28(5):1-0.
8. Russo F. A method for estimation and filtering of Gaussian noise in images. IEEE Transactions on Instrumentation and Measurement. 2003 Sep 23;52(4):1148-54.

Received 11.05.2023.
Accepted 22.05.2023.

Improving deep learning performance by augmenting training data

Satellite image recognition is a crucial application of computer vision that has the potential to be applied in various fields such as disaster management, agriculture, and urban planning. The objective of this study is to determine the optimal amount of input data required and select the most effective methods of augmentation necessary for training a convolutional neural network (CNN) for satellite image recognition.

To achieve this, we perform a series of experiments to investigate the effect of input data quantity on several performance metrics, including model accuracy, convergence, and generalization. Additionally, we explore the impact of various data augmentation techniques, such as rotation, scaling, and flipping, on model performance. The study suggests several strategies for identifying the saturation point and mitigating the effects of overtraining, including early stopping and dropout regularization.

The findings from this study can significantly contribute to the development of more efficient satellite recognition models. Furthermore, they can help improve the performance of existing models, in addition to providing guidance for future research. The study emphasizes the importance of carefully selecting input data and augmentation methods to achieve optimal performance in CNNs, which is fundamental in advancing the field of computer vision.

In addition to the above, the study investigates the potential of transfer learning by pretraining the model on a related dataset and fine-tuning it on the satellite imagery dataset. This approach can reduce the amount of required data and training time and increase model performance.

Overall, this study provides valuable insights into the optimal amount of input data and augmentation techniques for training CNNs for satellite image recognition, and its findings can guide future research in this area.

Солдатенко Дмитро Володимирович – аспірант кафедри інформаційних технологій і систем, Український державний університет науки і технологій.
Гнатушенко Вікторія Володимирівна - д.т.н., професор, завідуюча кафедрою інформаційних технологій і систем, Український державний університет науки і технологій.

Soldatenko Dmytro – PhD Aspirant, Department of information technology and systems of the Ukrainian state University of science and technologies.
HnatushenkoViktorija - Doctor of Engineering`s Sciences, Professor, Head of Department of information technology and systems of the Ukrainian state University of science and technologies.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)