https://doi.org/10.31891/2219-9365-2022-72-4-15
УДК 343.98

Діана СНАГОЩЕНКО
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ

В роботі на основі проведеного аналізу існуючих методів сегментації зображень та архутектур нейронної мережі, провівши низку експериментів використовуючи Набір даних Inria Aerial Image Labeling створений для основної задачі дистанційного зондування, а саме автоматичного попіксельного маркування аерофотознімків визначено, що саме експерименту U-Net++/EfficientNetB2/BCE+Dice Loss є найбільш вдалим для вирішення поставленої задачі розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках

Keywords: супутникові знімки, аерофотознімки, глибока згорткова нейронна мережа, глибоке навчання, функці втрат, сегментація об'єктів.

Diana SNAHOSHCHENKO
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

OBJECT RECOGNITION AND SEGMENTATION ON SATELLITE AND AERIAL IMAGING

The ability to detect new buildings directly from a satellite image is particularly useful in regions where the population is changing very quickly, as well as in remote and large-scale areas where the census of these new buildings is often done manually and quickly becomes obsolete. The processing of aerial photographs can also find important applications in the assessment of building damage during natural disasters. Finally, it can be very useful for solar panel manufacturers who want to estimate the usable roof surface area of a particular site.

Huge volumes of images are taken every day by airborne or space-based platforms, and this volume is still growing. This amount of data makes manual processing of images very resource-intensive, so the need for the use of neural networks is increasing. The main task of automatic image recognition is to assign a semantic class or label to each pixel, i.e. to transform the input data into a semantically meaningful bitmap (which can be further processed using, for example, vectorization or polygonization).

In the work, based on the analysis of the available methods of image segmentation and neural network architecture, by proving a number of experiments using the Inria Aerial Image Labeling Data Set, created for the main task of remote sensing, namely automatic pixel-by-pixel labeling of aerial photographs, it was determined that the U-Net++/EfficientNetB2 experiment /BCE. +Dice Loss is the largest value for solving the task of object recognition and segmentation on satellite and aerial imaging.

Keywords: satellite images, aerial images, deep convolutional neural network, deep learning, loss functions, object segmentation.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Можливість виявляти нові будівлі безпосередньо з супутникового зображення особливо корисна в регіонах, де кількість населення змінюється дуже швидко а також у віддалених і масштабних районах, де перепис цих нових будівель часто виконується вручну і швидко стає неактуальним. Обробка аерофотознімків також може знайти важливе застосування в оцінці будівельних пошкоджень при стихійних лихах. Нарешті, це може бути дуже корисним для виробників сонячних панелей, які хочуть оцінити корисну поверхню даху на певній ділянці.

Величезні об'єми зображень щодня знімаються бортовими або космічними платформами, і цей обсяг все ще зростає. Така кількість даних робить ручне опрацювання зображень дуже ресурсозатратним, отже зростає необхідність застосування нейронних мереж. Основною задачею автоматичного розпізнавання зображень є призначення семантичного класу або мітки кожному пікселю, тобто перетворення вхідних даних до семантично значущої растрової карти (яка в подальшому може підлягати додатковій обробці за допомогою, наприклад, векторизації або полігонізації).

Аналіз існуючих методів сегментації зображень

Задачі сегментації:

1. Сегментація зображень – це техніка комп'ютерного зору та обробки зображень, яка передбачає групування або позначення подібних областей або сегментів у зображенні на піксельному рівні. Мітка класу або маска представляє кожен сегмент пікселів.
2. Семантична сегментація – вивчає незліченні сутності на зображенні. Вона аналізує кожен піксель зображення та призначає унікальну мітку класу на основі текстури, яку він представляє.

3. Сегментація об’єктів – зазвичай стосується задач комп’ютерного зору, пов’язаних саме зі злічувальними об’єктами. Він може виявити кожен об’єкт або екземпляр класу, присутній на зображенні, і призначити йому власну маску або обмежуючий прямокутник з унікальним ідентифікатором.

4. Паноптична сегментація – поєднання підходів семантичної сегментації та сегментції об’єктів. Він представляє уніфікований підхід до сегментації зображення, коли кожному пікселю сцени присвоюється семантична мітка (через семантичну сегментацію) та унікальний ідентифікатор екземпляра (через сегментацію об’єктів).

Метрики оцінки якості роботи алгоритму:

1. Піксельна точність [1] – це метрика семантичної сегментації, яка позначає відсоток пікселів, які точно класифікуються на зображенні. Ця метрика обчислює відношення кількості правильно класифікованих пікселів до загальної кількості пікселів у зображенні.

2. Індекс Жаккара – це площа перекриття (перетину) між прогнозованою сегментацією та істинною сегментацією, поділена на площу об’єднання між прогнозованою сегментацією та істинною сегментацією.

3. Коефіцієнт подібності Дайса [2] – для задачі бінарної сегментації це міра між двома сегментаційними масками – це подвоєна площа перекриття, поділена на загальну кількість пікселів обох масок.

Архітектури згорткових нейронних мереж:

1. Unet [3] – оригінальна архітектура нейронної мережі U-Net було розроблена для семантичної сегментації біомедичних зображень. Навчена нейронна мережа дозволила вивести якість сегментації до по-піксельної точності. Нейронна мережа U-Net побудована на основі повністю згорткової мережі (Fully Convolutional Network) і складається з двох основний частик: кодувальника і декодувальника.

2. Deeplab [4] – сімейство архітектур нейронних мереж для семантичної сегментації зображень, що була розроблена дослідниками з компанії Google та являє собою комбінацію двох раніше відомих методів: глибоких згорткових нейронних мереж і повнозв’язаних умовних випадкових полів (Conditional Random Fields – CRF).

3. Unet++ [5] – архітектура нейронної мережі U-Net++ є еволюційним продовженням оригінальної архітектури U-Net. UNet++ складається з кодувальника та декодувальника, що поєднані за допомогою серії вкладених щільних згорткових блоків (замість базових пропускних з’єднань, використаних в U-Net). Основна ідея UNet++ в тому, аби подолати семантичний розрив між картами ознак кодувальника і декодувальника до етапу злиття.

Аналіз архітектури нейронної мережі

Модифікації архітектури нейронної мережі U-Net:

1. Операція збільшення розмірності – для переходу від маски з низькою роздільною здатністю до більш високої зазвичай використовували деконволюцію. Проте деконволюція має нерівномірне перекриття, коли розмір ядра згортки (розмір вихідного вікна) не ділиться націло на крок рухомого вікна, що призводить до специфічних артефактів. Для того, щоб уникнути подібних артефактів, в сучасних архітектурах відокремлюють операцію збільшення розмірності від операції згортки. Операція збільшення розмірності зазвичай замінюється інтерполяцією (інтерполяцією найближчого сусіда або білінійною інтерполяцією), а потім додається додатковий згортковий шар.

2. Симетричність кодувальника і декодувальника – в оригінальній архітектурі U-Net роздільна здатність відповідних блоків кодувальника і декодувальника не співпадають точно. Це відбувається через специфічні параметри ядра згортки та кроку для згорткових шарів цієї архітектури. Через це, для реалізації пропускних з’єднань і для співставленя розмірів карт ознак, з’являється необхідність вирізання центральної частини відповідного блока кодувальника. В сучасних архітектурах, необхідність подібної операції (центрального вирізання) нівелюється підбором відповідних параметрів ядра згортки, кроку, а також паддингу (обрамлення вхідної карти ознак нульовими пікселями) для згорткових шарів. За рахунок цього операція зменшення розмірності відбувається виключно в шарах максимізаційного агрегування (max pooling), в свою чергу просторовий розмір карти ознак (до та після згорткового шару) залишається сталим.

3. Глибоке контрольоване навчання [6] – Основною ідеєю глибокого контрольованого навчання є передбачення результуючих сегментаційних масок на різних рівнях декодувальника. Це автоматично означає також і підрахунок функцій втрат та додаткову оптимізацію цих самих блоків декодувальника.

Кодувальники:

1. ResNet [7] – структура залишкового навчання, створена для того, щоб полегшити навчання мереж, які є значно глибшими, ніж раніше, нівелюючи проблему їх деградації. нейронні мережі ResNet легше оптимізувати та вони можуть мати вищу точність на значних глибинах.

2. EfficientNet [8] – використовує техніку під назвою складений коефіцієнт для масштабування моделей простим, але ефективним способом. Замість випадкового збільшення ширини,

глибини чи роздільної здатності складене масштабування рівномірно масштабує кожен вимір, використовуючи певний фіксований набір коефіцієнтів масштабування.

Функція втрат:

1. Ентропія [9] випадкової величини X – це рівень невизначеності, або його усередненого інформаційного вмісту. Для p(x) – розподілу ймовірностей випадкової величини X ентропія визначається наступним чином:

$$H(X) = \begin{cases} -\int_x p(x)\ log\ p(x), & \text{якщо X} - \text{неперервна величина} \\ -\sum_x p(x)\ log\ p(x), & \text{якщо X} - \text{дискретна величина} \end{cases} \quad (1)$$

Чим більше значення ентропії H(x), тим більша невизначеність розподілу ймовірностей, а чим менше значення, тим менша невизначеність.

2. Перехресна ентропія – інші назви, такі як логарифмічна або логістична функція втрат. використовується для коригування ваг моделі під час навчання. Мета полягає в тому, щоб оптимізувати функцію втрат (мінімізувати втрати), так як менші втрати ведуть до кращої моделі. Ідеальна модель має перехресну ентропію рівну 0. На практиці ж, на жодному тренувальному наборі даних для реальної задачі неможливо досягти такого значення, тому більш точне формулювання - ідеальна модель має перехресну ентропію близьку до 0.

3. Бінарна перехресна ентропія – зазвичай обчислюється як середня крос-ентропія для всіх прикладів даних, тобто:

4.

$$L = -\frac{1}{N}\left[\sum_{j=1}^{N}\left[t_j\ log(p_j) + (1 - t_j)\ log(1 - p_j)\right]\right] \quad (2)$$

для N зображень тренувальної вибірки, де t_j- істинне значення, а p_j- softmax ймовірності для j-го зображення.

5. Бінарна перехресна ентропія для задачі семантичної сегментації – передбачає відповідну мітку класу для кожного пікселя даного зображення. Наприклад, для зображення розміром 512x512 ми маємо передбачити 512x512=262144 міток класів. Тому для задачі семантичної сегментації зображень функція втрат бінарної перехресної ентропії матиме наступний вигляд:

$$L = -\frac{1}{N}\left[\sum_{j=1}^{N}\frac{1}{M}\sum_{i=1}^{M}\left[t_{ji}\ log(p_{ji}) + (1 - t_{ji})\ log(1 - p_{ji})\right]\right] \quad (3)$$

де N - кількість зображень тренувальної вибірки, M - кількість пікселів зображення, tji - істинне значення, а pji - softmax ймовірності для i-го пікселя j-го зображення.

6. Фокальна функція втрат [10] одночасно вирішує дві основні проблеми бінарної перехресної ентропії:

- проблему дизбалансу між легкими та складними зразками використовуючи коефіцієнт модуляції γ;
- проблему дизбалансу між позитивними пікселями (об'єктами) та негативними пікселями (фоном) використовуючи ваговий параметр α.

Фокальна функція втрат має наступний вигляд:

$$LFL(p_t) = -\alpha_t(1 - p_t)^{\gamma}\ log(p_t) \quad (4)$$

Експерименти

Набори даних для навчання та валідації моделі – набір даних Inria

Набір даних Inria Aerial Image Labeling [11] створений для основної задачі дистанційного зондування, а саме автоматичного попіксельного маркування аерофотознімків.

Особливості набору даних:

- Покриття 810 км² (405 км² для навчання та 405 км² для тестування);
- Кольоровий аерофотознімок з просторовою роздільною здатністю 0,3 метри на піксель;

- Розмічені дані для двох семантичних класів: будівля і фон (публічно доступні лише для навчальної підмножини).

Навчальний набір [12] містить 180 кольорових зображень розміром 5000 × 5000, які покривають поверхню 1500 м × 1500 м кожен (з роздільною здатністю 30 см на піксель). Є 36 плиток для кожного з наступних регіонів:

- Остін
- Чикаго
- Кітсеп
- Західний Тіроль
- Відень

1. **Було проведено експерименти для вибору оптимальних характеристик глибокої згорткової нейронної мережі для сегментації будинків для навчального набору даних Inria.**

Експерименти проводилися для наступних складових нейронних мереж:

- Базова архітектура нейронної мережі:
  - U-Net
  - U-Net++
- Модель кодувальника
  - ResNet-18
  - MobileNetv2
  - EfficientNetB2
- Функці втрат:
  - Фокальна функція втрат (Focal Loss)
  - Бінарна перехресна ентропія + функція втрат Дайса (BCE+Dice Loss)

Інші параметри нейронної мережі, що були зафіксовані для усіх проведених експериментів:

- Роздільна здатність вхідного зображення: 256x256
- Метрика оцінки якості: IoU
- Оптимізатор: Адам
- Аугментації

В табл. 1 наведено загальні порівняння для всіх проведених експериментів.

Таблиця 1

Загальні порівняння для всіх проведених експериментів

Архітектура	**Загальна кількість параметрів (мільйони)**	**Кодувальник**	**кількість параметрів (мільйони)**	**функції втрат**	**метрика ↑ (валідаційна вибірка)**	**Візуальна оцінка ↑ (0-10)**
U-Net	14	ResNet-18	11	Focal Loss	74.8%	4
U-Net	5	MobileNet-V2	2	Focal Loss	74.9%	4
U-Net	10	EfficientNetB2	7	Focal Loss	78.2%	6
U-Net++	10,4	EfficientNetB2	7	Focal Loss	*78.7%*	7
U-Net++	10,4	EfficientNetB2	7	BCE+Dice Loss	*78.7%*	*9*

Візуальне порівняння першого базового експерименту (U-Net/ResNet-18/Focal Loss) та найкращого експерименту (U-Net++/EfficientNetB2/BCE+Dice Loss) зображено на рис. 1.

Також було показано, що значення метрики на валідаційній вибірці для найкращої моделі, що наразі дорівнює *78.7%*, насправді є значно вищим, і вже на даному етапі ми можемо стверджувати, що наша модель працює стабільніше за надану розмітку, що є дуже гарним знаком.

Приклад переваги результатів роботи нейронної мережі над “істинними” анотаціями для зображення з Чикаго зображено на рис. 2.

ISSN 2219-9365

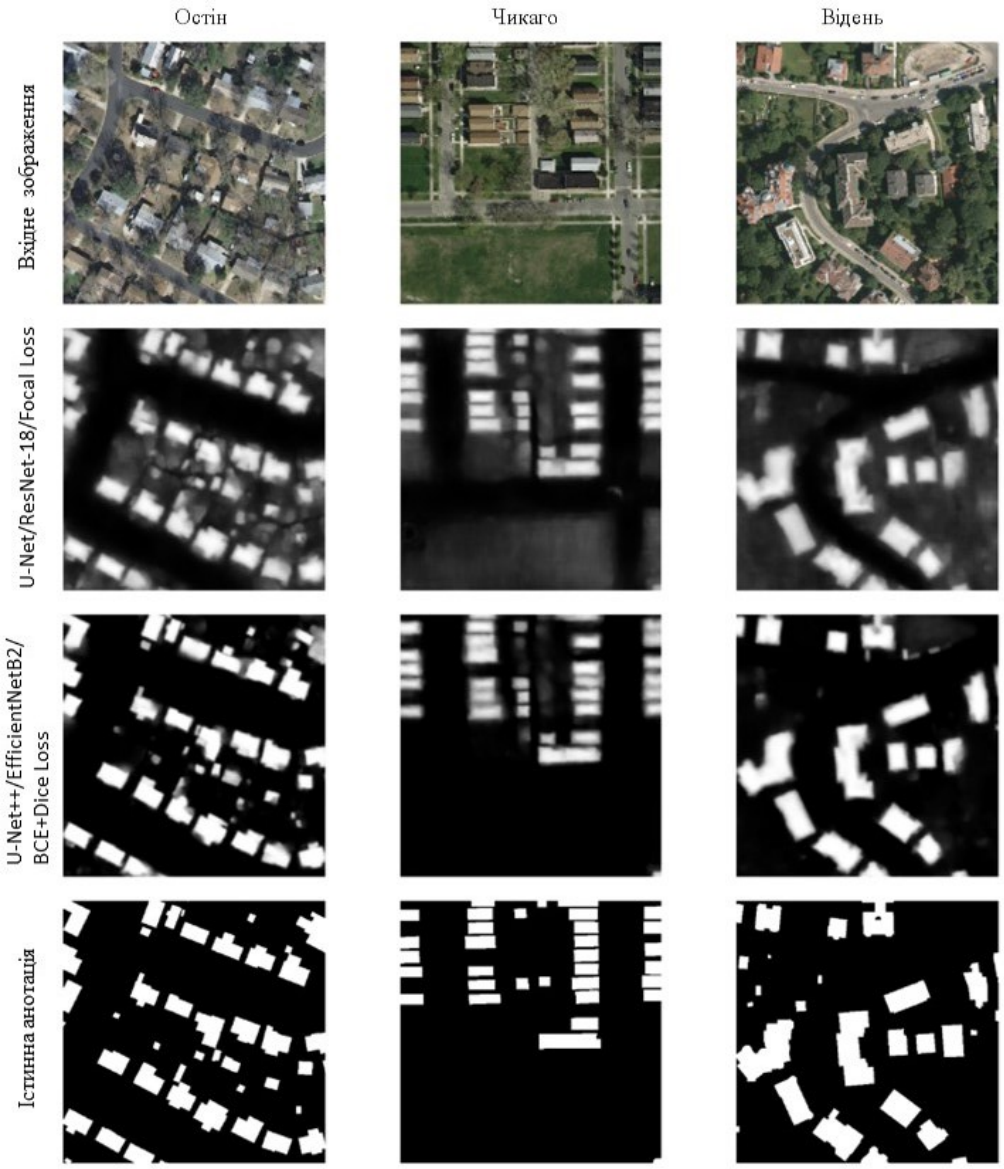
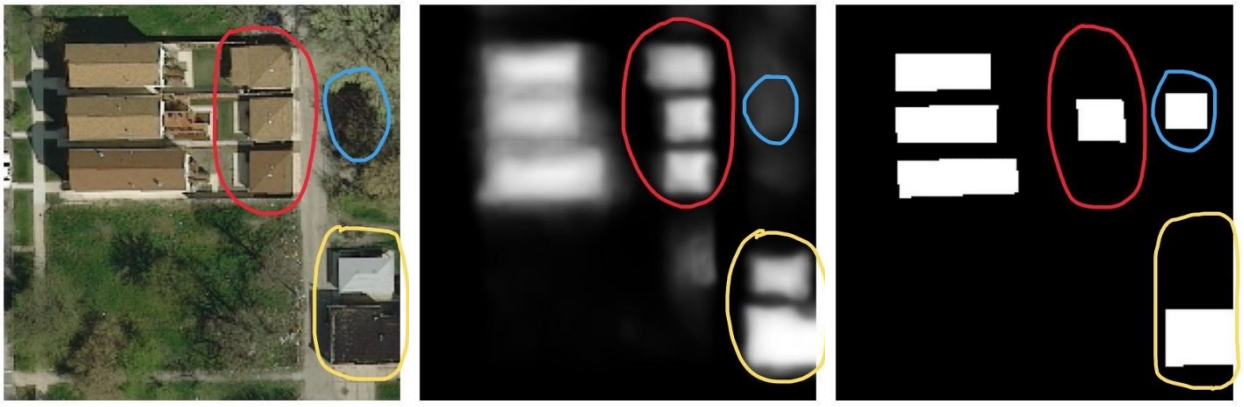


Рис. 1. Візуальний порівняльний аналіз експерименту U-Net/ResNet-18/Focal Loss та експерименту U-Net++/EfficientNetB2/BCE+Dice Loss



вхідне зображення **результати моделі** **“істинна” анотація**

Рис. 2. Приклад переваги результатів роботи нейронної мережі над “істинними” анотаціями для зображення з Чикаго

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

В роботі розглянуто та проаналізовано існуючі методи сегментації зображень та архітектури нейронної мережі для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках.

Для застосування на практиці методів сегментації зображень та архітектур нейронної мережі було використано набір даних Inria, який містить в собі 180 кольорових зображень розміром 5000 × 5000, які покривають поверхню 1500 м × 1500 м кожен (з роздільною здатністю 30 см на піксель).

Також було показано, що значення метрики на валідаційній вибірці для найкращої моделі, що наразі дорівнює 78.7%, насправді є значно вищим, і вже на даному етапі можна стверджувати, що створена модель працює стабільніше за надану розмітку, що є дуже гарним знаком.

References

1. Semantic scene segmentation for robotics [Електронний ресурс]. URL: https://www.sciencedirect.com/topics/computer-science/pixel-accuracy
2. Understanding DICE COEFFICIENT [Електронний ресурс]. URL: https://www.kaggle.com/code/yerramvarun/understanding-dice-coefficient
3. Chhor G., Bartolome Aramburu C., Bougdal-Lambert I., Satellite Image Segmentation for Building Detection using U-net. [Електронний ресурс]. URL: http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf
4. The DeepLab Family [Електронний ресурс]. URL: https://medium.com/@calleris.enrico/the-deeplab-family-70d8b98262b5
5. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J., UNet++: A Nested U-Net Architecture for Medical Image Segmentation. arXiv:1807.10165v1 [cs.CV] 18 Jul 2018. [Електронний ресурс]. URL: 1807.10165.pdf (arxiv.org)
6. Lee C.-Y., Xie S., Gallagher P., Zhang Z., Tu Z., Deeply-supervised nets. In Artificial Intelligence and Statistics, pages 562–570, 2015.
7. ResNet: A Simple Understanding of the Residual Networks [Електронний ресурс]. URL: https://medium.com/swlh/resnet-a-simple-understanding-of-the-residual-networks-bfd8a1b4a447
8. Understanding EfficientNet — The most powerful CNN architecture [Електронний ресурс]. URL: https://medium.com/mlearning-ai/understanding-efficientnet-the-most-powerful-cnn-architecture-eaeb40386fad
9. Cross-Entropy Loss Function [Електронний ресурс]. URL: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e
10. Focal Loss — What, Why, and How? [Електронний ресурс]. URL: https://medium.com/swlh/focal-loss-what-why-and-how-df6735f26616
11. Inria Aerial Image Labeling Dataset [Електронний ресурс]. URL: https://project.inria.fr/aerialimagelabeling/
12. Inria Aerial Image Labeling Dataset/Contest [Електронний ресурс]. URL: https://project.inria.fr/aerialimagelabeling/contest/