

ЗАСТОСУВАННЯ ГЕНЕРАТИВНО-ЗМАГАЛЬНИХ МЕРЕЖ ДЛЯ ПОКРАЩЕННЯ ЯКОСТІ СЕГМЕНТАЦІЇ СУПУТНИКОВИХ ЗНІМКІВ

О. В. Шкаліков^{1, а}, А. О. Охріменко¹, Л. Л. Шуміло², Н. М. Куссуль^{1, 3}

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
ІН Фізико-технічний інститут

² University of Maryland

³ Інститут космічних досліджень НАНУ ДКАУ

Анотація

Ідея роботи полягає у тому, щоб застосувати генеративні моделі для створення штучних навчальних вибірок, тим самим збільшити й урізноманітнити навчальні вибірки для моделей семантичної сегментації супутникових знімків. А саме: побудувати генеративно-змагальну мережу, яка на основі видозмінених масок зможе генерувати супутникові знімки, подібні до реальних, причому у результаті класи у тренувальній вибірці будуть більш збалансованими, що дозволить збільшити якість сегментації міноритарних класів.

Ключові слова: семантична сегментація, UNET, генеративно-змагальні мережі, Pix2Pix, незбалансованість класів

Вступ

У сучасному світі задачі сегментації супутникових знімків мають широке прикладне значення, і використовуються для побудови карт сільськогосподарських культур [1], детектування незаконної вирубки лісів [2], звалищ, пожеж, тощо. Проте сучасні методи розв'язку цієї задачі здебільшого є алгоритмами навчання з учителем та потребують розмічених навчальних вибірок. Створення даного матеріалу, а саме ground truth масок для навчання потребує значних людських ресурсів, натомість велика кількість людино-годин може не призвести до значного покращення результатів сегментації. До того ж, навіть при використанні такого підходу, точність класифікації міноритарних класів є низькою, це пов'язано з тим, що загальна площа, яку вони займають на супутникових знімках з вибірки мала, тому функціонал похибки роботи моделі не достатньо штрафує за неправильну класифікацію. Тож пошук методів подолання цих проблем за допомогою генеративно-змагальних мереж і є метою даного дослідження.

1. Задача семантичної сегментації супутникових знімків

Як і інші задачі навчання з учителем, задача семантичної сегментації передбачає існування навчальної вибірки, тобто набору X , який складається з зображень $I \in (R^C)^{H \times W} = \mathcal{I}$ та міток класів $k \in K$ для кожного пікселя кожного зображення, де $R \subset \mathbb{R}$ – множина можливих значень кожного каналу пікселя, H, W, C – висота, ширина та кількість каналів зображення, $K = \overline{1, M}$ – множина класів ($M \in \mathbb{N}$ –

кількість класів). Зазвичай множина $R = \overline{0, \dots, 255}$, або $R = [0, 1]$.

Нехай ми маємо навчальну вибірку $X = (I_i, Y_i)_{i=1}^N$, де $I_i \in \mathcal{I}$, $Y_i \in K^{H \times W}$. Задача семантичної сегментації полягає у тому, щоб знайти таке відображення $f: \mathcal{I} \rightarrow K^{H \times W}$, що деякий функціонал помилки $L(f, X)$ приймає мінімальне значення.

Тобто задача семантичної сегментації є задачею класифікації, але не цілого зображення, а кожного окремого пікселя.

2. Проблеми, що виникають при застосуванні нейромережових підходів

Попри використання сучасних нейромережових архітектур, при розв'язку задач семантичної сегментації супутникових знімків виникають істотні проблеми, які не дозволяють досягти максимальної якості. Ці перепони є наслідком самої структури задачі, а саме того, що семантична сегментація є задачею навчання з учителем, тобто вимагає розміченого набору даних, де кожному супутниковому знімку буде відповідати відповідна ground-truth маска сегментації.

Тож першою істотною проблемою є те, що створення навчальних вибірок, а саме: розмітка супутникових знімків, вимагає значної кількості людських ресурсів. Причому кінцева якість сегментації безпосередньо залежить від розміру тренувального набору. Тобто потреба у покращенні значень метрик вимагає, серед усього іншого, істотного збільшення навчальних прикладів, а отже й створених людьми ground-truth масок. Крім того, у випадку створення масок для супутникових знімків, не завжди людина може правильно класифікувати усі об'єкти ґрунтую-

^аoleshk-ipt22@iitl.kpi.ua

чись тільки на зображенні, що може бути пов'язано з порою року, розміром об'єкту, роздільною здатністю знімка, схожістю різних класів, тощо.

Друга істотна проблема пов'язана з дуже сильною незбалансованістю класів, особливо у задачах класифікації сільськогосподарських культур. Як доказ цього, можна навести статистику (рис. 1) ground truth маски по кількості пікселів по кожному з типів полів.

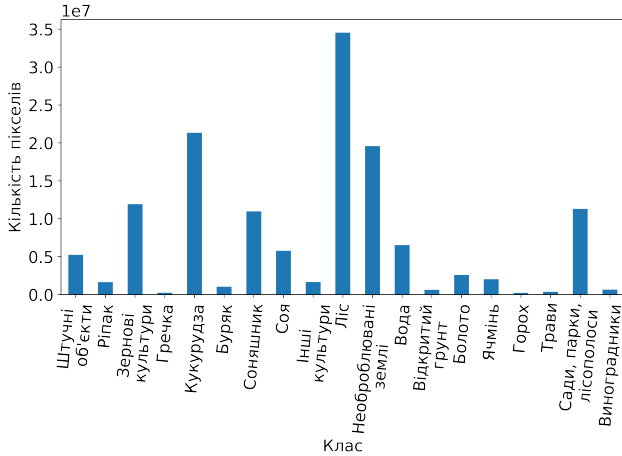


Рис. 1. Статистики для різних класів сільськогосподарських культур

І попри застосування функцій помилок, які можуть враховувати дисбаланс у класах, досягнути бажаних значень метрик, що описують точність моделей не вдається. Особливо незадовільні результати виникають саме для класів, які мають маленьку кількість прикладів у навчальній вибірці. А доповнити навчальну вибірку достатньою кількістю прикладів для певних класів може бути взагалі неможливо, через особливості сільського господарства у досліджуваному регіоні або самої сільськогосподарської культури. Останній факт як раз і є специфічним саме для аналізу супутникових знімків і виділяє їх серед усіх інших сфер, де розв'язується задача семантичної сегментації.

3. Генеративно-змагальні мережі у задачі Image-to-Image Translation

Ідея оригінального GAN [3] полягає у тому, що ми маємо дві сутності: генератор — диференційовну функцію $G : Z \times \Theta_G \rightarrow X$, яка на основі елемента з множини Z (шуму, або у деяких випадках, так званого латентного простору) та параметрів $\theta_g \in \Theta_G$ генерує елемент з множини X ; та дискримінатор — диференційовну функцію $D : X \times \Theta_D \rightarrow \{0, 1\}$, яка намагається на визначити чи є елемент штучно згенерованим, чи належить розподілу навчальних даних. Для пошуку функції генератора та дискримінатора застосовується міні-максний функціонал якості:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}} \log D(x) + \mathbb{E}_{z \sim p_z} \log(1 - D(G(z)))$$

У нашому випадку, застосування класичної GAN не є доцільним, бо ті зображення, що модель по-

винна продукувати, повинні бути обумовлені тією маскою, що подається на вхід системи, а не тільки випадковим шумом z , що обумовлено нашим намаганням повністю контролювати розподіл класів у штучно-згенерованих знімках. Тож треба застосовувати Conditional GAN, одним з широко відомих та поширених представників яких є Pix2Pix [4]. Схематично його архітектура виглядає наступним чином (рис. 2):

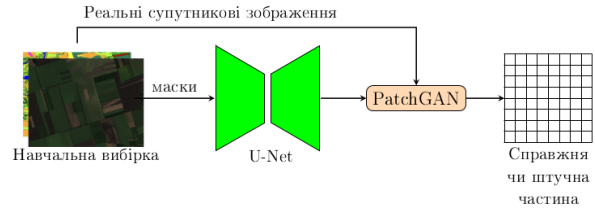


Рис. 2. Архітектура Pix2Pix у застосуванні до супутникових знімків

У якості функціоналу похибки пропонується використовувати наступний вираз, який являє собою класичну похибку трансформовану для відповідності Conditional GAN й додаток, який вимірює схожість згенерованого й реального знімка:

$$L(G, D) = \mathbb{E}_{x, y} \log D(x, y) + \mathbb{E}_{x, z} \log(1 - D(x, G(x, z))) + \lambda \mathbb{E}_{y, x, z} \|y - G(x, z)\|_1$$

Тут, як і у інших Conditional GAN, генератор вже являє собою відображення $G : X \times Z \rightarrow Y$, а дискримінатор $D : X \times Y \rightarrow \{0, 1\}^{N \times N}$, де X — множина масок, Y — множина супутникових знімків.

Генератором у архітектурі Pix2Pix виступає різновид мережі UNet [5]. У якості дискримінатора використовується PatchGAN [4], тобто дискримінатор повертає свій висновок про те є зображення реальним чи штучним не для усього зображення, а окремо для його частин.

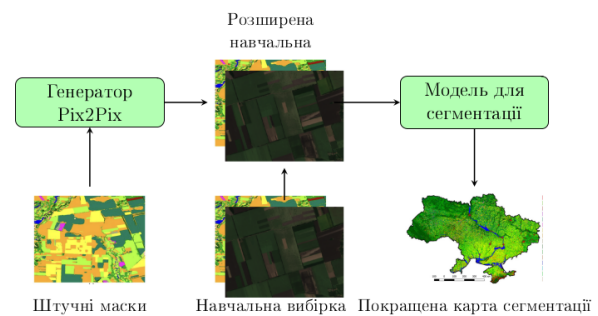


Рис. 3. Процес аугментації наборів даних супутникових знімків за допомогою Pix2Pix

Остаточно, вихідний робочий процес після навчання GAN і відповідно генератора можна представити діаграмою (рис. 3). Тобто ми згенеруємо на основі масок, які будуть коригувати незбалансованість класів, штучні супутникові знімки, додамо їх до навчальної

вибірки і вже на цьому аугментованому наборі даних будемо вчити моделі семантичної сегментації.

4. Дослідження ефективності запропонованого підходу

Для перевірки того, чи дійсно запропонований підхід дозволить збільшити якість семантичної сегментації супутникових знімків було проведено експерименти по класифікації на 19 класів (здебільшого сільськогосподарських культур) композиту Sentinel-2A [6] для Київської області взятих з 1 липня по 1 серпня 2021 року.

Попередня обробка даних включала в себе поділ композиту на зображення 256×256 пікселів, видалення зображень у яких відсутні дані про деякі пікселі чи мітка класу, а також нормалізація з параметрами 0.5 у якості середнього та 0.5 у якості стандартного відхилення, для усіх каналів кожного пікселя. Набір даних було поділено навпіл на навчальну та тестову вибірки.

Після цього було навчено генеративно-змагальну мережу архітектури Pix2Pix протягом 300 епох. Згенеровано маски за наступним алгоритмом, у кожній масці вихідного набору даних було змінено i -ий за сумарною кількістю пікселів клас на $(19 - i + 1)$. На основі цих масок, навченою мережею були згенеровані супутникові знімки. Ці знімки разом з масками були додані до вихідного набору даних, тобто була отримана аугментована вибірка.

І остаточно було навчено модель UNet [5] на основі тільки реальних даних і аугментованої вибірки. Після цього було виміряно точність (producer accuracy). Результати по кожному з класів наведені у табл. 1.

У усіх процесах навчання використовувався оптимізатор Adam [7] зі швидкістю навчання $2 \cdot 10^{-4}$. При навчанні GAN у якості параметрів експоненційного середнього β_1, β_2 використовувались значення 0.5, 0.999 відповідно.

Як можна побачити точність для мажоритарних класів, таких як зернові культури або кукурудза трохи зменшилась, проте у випадку з міnorитарними класами ситуація повністю протилежна: ми змогли досягти значного підвищення якості класифікації саме цих класів. Тож ми можемо стверджувати, що для міnorитарних класів, даний підхід є ефективним та дозволяє значно покращити якість семантичної сегментації.

Висновки

У результаті проведеної роботи була досліджена проблема семантичної сегментації супутникових знімків, яка має важливу роль у прикладних застосуваннях. Проте були виявлені та описані проблеми, які виникають під час розв'язку цієї задачі сучасними нейромережевими методами, а саме: складність створення великих вибірок та сильна незбалансованість класів. Причому, на відміну від інших сфер, отримати навчальні приклади, які могли б змінити розподіл класів у багатьох випадках для супутникових даних не є можливим.

Табл. 1. Порівняльна таблиця точності сегментації по кожному з класів (у процентах)

Назва класу	Реальні дані	Аугментована вибірка
Штучні об'єкти	69	71
Зернові культури	89	79
Ріпак	0	53
Гречка	0	0
Кукурудза	94	87
Буряк	0	23
Соняшник	89	94
Соя	27	66
Інші культури	0	6
Ліс	93	93
Необроблювані землі	71	76
Відкритий ґрунт	3	54
Вода	97	97
Болото	27	34
Ячмінь	5	21
Горох	0	1
Трави	0	0
Сади, парки, лісополоси	45	50
Виноградники	0	0
Загальна точність	75	79

Для подолання цих проблем було розглянуто можливість аугментації навчальної вибірки штучно згенерованими навчальними прикладами. Для цього була досліджена модель, які дозволяють вирішити цю задачу, а саме генеративно-змагальні мережі. При цьому, для вирішення проблеми незбалансованості класів стандартні GAN не кращий вибір, бо ми не можемо контролювати розподіл класів на згенерованих зображеннях. Тому були опрацьовані моделі image-to-image translation, зокрема Pix2Pix, які дозволяють генерувати супутникові знімки на основі масок. У свою чергу модифікація вже існуючих масок, тобто зміна одних класів на інші дозволяє скоригувати незбалансованість класів.

Таким чином було розроблено процес аугментації наборів даних супутникових знімків, ефективність якого була перевірений експериментальним чином (табл. 1). У результаті ми отримали підтвердження того, що запропонований підхід дозволяє значно підвищити якість семантичної сегментації міnorитарних класів, при цьому якість класифікації мажоритарних класів, хоч у деяких випадках і стала нижчою, проте незначно.

Очевидні і подальші напрямки досліджень у даній сфері. По-перше, це дослідження методів генерації, які б дозволили б врахувати той факт, що супутникові знімки однієї території у різні пори року значно відрізняються. По-друге, створення більш досконалих методів генерації масок. І врешті решт, покращення існуючих генеративно-змагальних ме-

реж для генерації ще більш реалістичних штучних супутникових знімків.

Перелік використаних джерел

1. Deep learning classification of land cover and crop types using remote sensing data / Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, Andrii Shelestov // IEEE Geoscience and Remote Sensing Letters. — 2017. — Т. 14, № 5. — С. 778–782.
2. Shumilo Leonid, Kussul Nataliia, Lavreniuk Mykola. U-Net Model for Logging Detection Based on the Sentinel-1 and Sentinel-2 Data // 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. — 2021. — С. 4680–4683.
3. Generative adversarial nets / Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza та ін. // Advances in neural information processing systems. — 2014. — Т. 27.
4. Image-to-image translation with conditional adversarial networks / Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A Efros // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2017. — С. 1125–1134.
5. Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention / Springer. — 2015. — С. 234–241.
6. Sentinel-2: ESA's optical high-resolution mission for GMES operational services / Matthias Drusch, Umberto Del Bello, Sébastien Carlier та ін. // Remote sensing of Environment. — 2012. — Т. 120. — С. 25–36.
7. Diederik Kingma, Jimmy Ba та ін. Adam: A method for stochastic optimization // arXiv. — 2014. — С. 273–297. — (Препринт / arXiv:1701.04862).