

УДК 004.932.2

Легкобит В. С. ORCID: 0000-0002-9118-4188 (ВІТІ ім. Героїв Крут)
Стоцький І. В. ORCID: 0000-0002-5898-3228 (ВІТІ ім. Героїв Крут)
Карпенко А. О. ORCID: 0000-0002-8372-6303 (ВІТІ ім. Героїв Крут)

ПІДВИЩЕННЯ ТОЧНОСТІ КЛАСИФІКАЦІЇ ОБ'ЄКТІВ НА МІСЦЕВОСТІ НА ОСНОВІ ПОЄДНАННЯ СЕНСОРНИХ МОДАЛЬНОСТЕЙ ІЗ ВИКОРИСТАННЯМ БІНАРНОЇ КРОСЕТРОПІЇ ЯК ФУНКЦІЇ ВТРАТ

У сучасних умовах ведення бойових дій надзвичайної важливості набуло програмне забезпечення, в основу якого покладено техніку машинного навчання для виявлення та класифікації об'єктів інфраструктури на місцевості. У зв'язку з цим актуальним постає завдання підвищення точності класифікації об'єктів на місцевості, оскільки застосування безпілотних літальних апаратів (далі – БпЛА) та космічних систем є життєво важливим для ведення розвідувальної діяльності. Враховуючи характер вхідних даних, а саме статичні зображення місцевості, отримані у вигляді супутникових знімків та знімків з камер БпЛА, для вирішення завдань класифікації доцільно застосовувати згорткові нейронні мережі. У більшості випадків супутникові знімки представлені у вигляді мультиспектральних та гіперспектральних зображень, тому для навчання моделі були використані загальнодоступні набори даних, що запропоновані спільнотою дослідників SpaceNet. Важливим етапом підготовки навчальної вибірки є орторектифікація зображень, а саме додавання інформації про 3D-поверхню до зображень, що надає моделі важливу геометричну інформацію для семантичних класів, таких як будівлі та інші споруди, виправляє геометричні спотворення та сприяє навчанню моделі розпізнавати об'єкти в узгодженому геопросторовому контексті. Під час експериментів було проведено навчання моделі SegNet із використанням *normalized Digital Surface Model (nDSM)* та без неї. Результати експериментів демонструють, що узагальнена точність класифікації для шести класів об'єктів на тестовому наборі даних зростає на 23,9%. А експерименти з обмеженням навчальної вибірки продемонстрували, що за потреби достатньо використати половину доступного тренувального набору даних для отримання лише на 4% нижчої точності класифікації та заощадження близько 10 годин навчання.

Ключові слова: згорткові нейронні мережі, класифікація, орторектифікація.

V. Legkobyt, I. Stotskiy, A. Karpenko Improving the accuracy of object classification on the ground based on a combination of sensory modalities using binary cross-tropism as a loss function.

In today's warfare environment, software based on machine learning techniques for detecting and classifying infrastructure objects on the ground has become extremely important. In this regard, the task of improving the accuracy of classification of objects on the ground is becoming more urgent, as the use of UAVs and space systems is vital for intelligence activities. Given the nature of the input data, namely static terrain images obtained in the form of satellite images and UAV camera images, it is advisable to use convolutional neural networks to solve classification problems. In most cases, satellite images are presented in the form of multispectral and hyperspectral images, so publicly available datasets offered by the SpaceNet research community were used to train the model. An important step in preparing the training set is image orthorectification, namely adding 3D surface information to the images, which provides the model with important geometric information for semantic classes such as buildings and other structures, corrects geometric distortions, and helps the model to recognize objects in a consistent geospatial context. In the course of the experiments, the SegNet model was trained with and without the normalized Digital Surface Model (nDSM). The experimental results show that the generalized classification accuracy for six classes of objects on the test dataset increases by 23.9%. And experiments with training set limitation demonstrate that, if necessary, it is enough to use half of the available training data set to obtain only 4% lower classification accuracy and save about 10 hours of training.

Keywords: convolutional neural networks, classification, orthorectification.

Постановка завдання в загальному вигляді

Досвід застосування космічних систем та безпілотних авіаційних комплексів (далі – БпАК) в останніх конфліктах чітко вказує на трансформацію способів ведення війни, зокрема активне ведення інформаційної війни. Аналіз застосування засобів і систем різних видів розвідки у воєнних конфліктах останніх десятиліть показали, що близько 70% розвідувальної інформації про наземні об'єкти противника отримано за допомогою космічних апаратів та БпЛА з використанням великої кількості фотографічних, оптико-електронних, радіолокаційних тепловізійних приладів [1].

Отримані актуальні дані дозволятимуть відповідним аналітичним підрозділам Збройних Сил України (далі – ЗС України) за демаскуючими ознаками робити висновки щодо ймовірних локацій (об'єктів, районів, напрямків) інтересу противника та ймовірних цілей для систем і комплексів високоточного керованого ракетного озброєння. Використання зазначеної інформації у ЗС України дозволить знизити втрати з боку особового складу, зберегти озброєння та військову техніку, завчасно визначати наміри противника, тим самим, в цілому, здійснювати більш ефективну протидію атакам противника [2].

На точність класифікації цільових об'єктів суттєво впливають фактори, що природним чином виникають внаслідок різноманітності зовнішнього вигляду, кутів огляду, умов навколишнього середовища, широкий спектр часових та просторових варіацій тощо. Усунення негативного впливу вищезазначених факторів на точність класифікації у більшості випадків досягається шляхом:

залучення попередньо навчених моделей на великих наборах даних (наприклад, ImageNet [3]);

виокремлення релевантних особливостей необроблених даних перед завантаженням їх у модель;

застосування методів перехресної перевірки, але з точки зору технічної реалізації ці підходи потребують значних апаратних і часових ресурсів.

Отже, актуальним постає завдання аналізу впливу на точність класифікації таких підходів до машинного навчання, як розширення навчальної вибірки завдяки збільшенню кількості анованих слідів об'єктів для усунення дисбалансу класів, мультимодальне злиття даних із різних джерел, налаштування гіперпараметрів та адаптація до предметної області шляхом застосування синтетичних даних.

Оскільки не існує універсальної архітектури нейронної мережі для класифікації об'єктів на місцевості, увагу слід зосередити на пошуку адаптивних рішень на основі модифікації вхідних даних стосовно особливостей конкретних завдань класифікації, бажаної точності.

Аналіз останніх публікацій. Дослідженню підходів до машинного навчання з метою пошуку способів оптимізації параметрів моделей нейронних мереж для класифікації об'єктів на місцевості присвячена значна кількість робіт, аналіз основних із яких викладено нижче.

У роботі [4] проведено дослідження значення різних модальностей та їх поєднання для аналізу геопросторових даних низької просторової роздільної здатності. У роботі представлено математичну модель, що дозволяє збагачувати геопросторові дані додатковою семантикою на основі заданої інформації про колір, гіперспектральної інформації та інформації про форму. Для визначення векторів ознак та класифікації використовувався класифікатор випадкового лісу. Отримані результати демонструють, що сама собою інформація про форму має досить обмежену цінність для вилучення семантичної інформації, якщо просторова роздільна здатність є відносно низькою і якщо визначені користувачем класи виявляють подібну геометричну поведінку локальної структури.

У роботі [5] проведено дослідження щодо ідентифікації слідів будівель у програмах дистанційного зондування місцевості з використанням у якості джерел вхідних даних 3D-лідарів та комерційних супутників. У роботі описано результати змагань фахівців із машинного навчання, таких як USSOCOM Urban 3D, DSTL Satellite Imagery Feature Detection Challenge та ISPRS Test Project on Urban Classification, переможці яких у своїх рішеннях використовували архітектури нейронних мереж LinkNet та UNet із модифікованими кодерами ResNet18, 34, 50 та VGG16. Узагальнюючи особливості підходів до вирішення завдань класифікації в рамках вищезазначених змагань, варто зазначити наступне: для усунення упередженості моделі дані для тренування було обрано випадковим чином у різних пропорціях відносно загальної вибірки, фаза тренування проходила різну кількість епох та з різними швидкостями, застосування методів ортогональної та випадкової ініціалізації wag

нейронної мережі та експерименти з різними пороговими значеннями для алгоритму сегментації зображень. Але поза увагою залишилися більш доцільні методи ініціалізації ваг мережі, такі як ініціалізація Ксав'є-Глорота, що встановлює ваги на основі кількості вхідних та вихідних нейронів і масштабує випадкові ваги на коефіцієнт, який гарантує, що дисперсія активацій залишається приблизно однаковою між шарами та ініціалізація LeCun, яка часто використовується для активацій Leaky ReLU та Scaled Exponential Linear Units (SELU) і також встановлює ваги на основі кількості вхідних і вихідних нейронів із різним коефіцієнтом масштабування.

У роботі [6] розглядається глибока семантична сегментація аерофотознімків на основі мультимодальних даних з урахуванням відповідних цифрових моделей поверхні, створених на основі радіометричних та геометричних характеристик, які надаються окремо та в різних комбінаціях як вхідні дані для Depthwise Separable Convolutional Neural Network (DSCNN), модифікованої шляхом застосування функцій масштабування, отриманих із проміжних рівнів SCNN. Отримані результати демонструють, що масштабування та глибинна згортка сприяють покращенню продуктивності сегментації зображень, особливо для груп об'єктів, але не підвищують точність класифікації окремих об'єктів. Загалом, DSCNN забезпечують ефективний і дієвий спосіб побудови глибоких нейронних мереж для задач комп'ютерного зору, одночасно вирішуючи проблеми, пов'язані з розміром моделі та обчислювальними витратами. Вони знайшли застосування в мобільних і вбудованих системах, комп'ютерному зору в реальному часі та інших сценаріях, де ефективність використання ресурсів має вирішальне значення.

У роботі [7] Data Scientist компанії Capella Space, що є партнером SpaceNet LLC, Jake Shermeyer опублікував результати переможців змагань "SpaceNet 6 мультисенсорне всепогодне картографування". В рамках даного змагання учасникам у якості вхідних даних були запропоновані принципово нові сенсорні модальності у вигляді даних Synthetic Aperture Radar (SAR) та півметрових електрооптичних зображень із супутника Maxar WorldView 2. Метою змагання було автоматично отримувати сліди будівель за допомогою комп'ютерного зору та алгоритмів штучного інтелекту (ШІ), використовуючи комбінацію вищезазначених даних. Незважаючи на те, що результат переможця змагань стосовно точності розпізнавання слідів будівель становить лише 42,423 %, його можна вважати успішним, враховуючи, що дані SAR були зібрані з точки зору поза надіром із усередненням $\sim 35^\circ$, що є одним із найскладніших аспектів, присутніх у спостереженні Землі. Показники точності виявлення слідів будівель підкреслюють проблеми, пов'язані з входженням у нову модальність, і демонструють, що необхідні подальші дослідження для повного використання домену SAR. Стосовно архітектурних рішень та підходів учасників варто зазначити використання таких ансамблевих методів, як Bootstrap Aggregating, AdaBoost, Gradient Boosting та Dropout Regularization, виключення оптичних даних із набору даних для попереднього навчання, застосування багатоканальних масок для сегментації зображень. Ці підходи показують, що іноді навіть незначні етапи попередньої обробки можуть бути особливо цінними для покращення продуктивності та мінімізації деяких складнощів.

У роботі [8] представлено новий підхід до підвищення продуктивності нейромережевих класифікаторів. У традиційних нейромережевих класифікаторах центроїди (репрезентативні точки) фіксуються в просторі розбиття перед навчанням. Дані центроїди мають вирішальне значення для створення меж розбиття між класами. Однак такий фіксований підхід обмежує потенційні конфігурації нейронної мережі, оскільки він обмежує форму меж прийняття рішень. Запропонований авторами підхід поєднує класифікацію за методом найближчого сусіда з нейромережевим класифікатором. На відміну від традиційних методів із фіксованими центроїдами та сфероподібними розділами, використання найближчих сусідів дозволяє більш гнучко визначати межі рішень. Методи найближчих сусідів можуть адаптуватися до шаблонів

даних і створювати кордони довільної форми в просторі розбиття. Експериментальні результати показують, що цей новий підхід працює краще з погляду точності та середньої F-міри порівняно з існуючими методами. Іншими словами, поєднання класифікації за методом найближчого сусіда з нейронними мережами призводить до покращення результатів класифікації, особливо у випадках, коли потрібні складні та неоднорідні межі розбиття. Отже, у роботі представлено новий підхід до нейромережевої класифікації, який використовує гнучкість методу найближчого сусіда для створення більш ефективних меж прийняття рішень довільної форми. Експериментальні результати підтверджують ефективність цього підходу в підвищенні точності та достовірності класифікації.

У роботі [9] представлено метод нечітких плаваючих центроїдів (МНПЦ), який має на меті покращити продуктивність нейромережевого класифікатора. Традиційні нейромережеві класифікатори з плаваючими центроїдами часто мають фіксовані межі прийняття рішень, що може призводити до ексклюзивного розподілу класів, особливо для зашумлених або граничних точок, це призводить до неправильної класифікації і впливає на навчання нейронних мереж. Щоб вирішити цю проблему, МНПЦ вводить нечітку стратегію разом із плаваючими центроїдами для створення нечітких меж прийняття рішень. Нечіткі межі дозволяють точкам даних мати ступінь приналежності до декількох класів, що робить його більш гнучким і здатним обробляти зашумлені та граничні точки. Ця гнучкість збільшує ймовірність знаходження оптимальної конфігурації нейронної мережі під час оптимізації. Крім того, МНПЦ використовує зважену цільову функцію для обробки незбалансованих даних, надаючи більше значення класам меншин. Це допомагає покращити обробку незбалансованих наборів даних. Ефективність МНПЦ оцінено на 32 еталонних наборах даних за допомогою таких показників, як середня F-міра та точність узагальнення. Результати експерименту показують, що МНПЦ досягає заданої точності узагальнення на 17 наборах даних і оптимальної F-міри на 21 наборі даних, перевершуючи інші методи класифікації. МНПЦ особливо ефективний у балансуванні між точністю та запам'ятовуванням, що робить його підходящим вибором для таких завдань, як оцінка класів на основі мікроструктурних зображень. Отже, МНПЦ – це метод, який поєднує нечітку логіку і плаваючі центроїди для створення нечітких меж прийняття рішень в нейромережевій класифікації. Експериментальні результати демонструють його перевагу над іншими методами, особливо в обробці незбалансованих даних і балансуванні точності та запам'ятовуванні для конкретних застосувань.

У роботі [10] розглядаються дві проблеми: проблема дисбалансу в класифікації та обробка матричних образів без векторизації, для вирішення яких автори пропонують новий метод *Entropy-based Matrix Learning Machine with MatMHS and Fuzzy Membership Evaluation (EMatMHS)*. Замість перетворення матричних шаблонів у вектори, автори пропонують використовувати в якості базового класифікатора *Matrix-pattern-oriented Ho-Kashyap learning machine with regularization learning (MatMHS)*. Даний підхід є актуальним для ефективної обробки матричних шаблонів. Для вирішення проблеми дисбалансу автори представляють підхід до оцінки нечіткої приналежності на основі ентропії. Цей підхід використовує ентропію як міру визначеності класу для визначення нечіткої належності навчальних моделей. Шаблони з вищою ентропією можуть мати нижчу приналежність, що відображає невизначеність щодо їхнього класу. У роботі проведено оцінку запропонованого методу *EMatMHS* із використанням реальних незбалансованих наборів даних порівняно з такими підходами до попередньої обробки, як *Synthetic Minority Over-sampling Technique (SMOTE)*, *Adaptive Synthetic Sampling machine learning technique (ADASYN)*, *Cluster Based Synthetic Oversampling (CBSO)*. Результати показують, що *EMatMHS* перевершує порівнювані підходи з точки зору точності класифікації, особливо для незбалансованих наборів даних. Отже, *EMatMHS*

представлено як рішення, яке може ефективно обробляти як незбалансовані дані, так і матричні патерни.

У роботі [11] представлено підхід для вирішення проблеми незбалансованих наборів даних у нейронних мережах, що полягає в інтеграції попередніх ймовірностей у функцію перехресної ентропійної помилки, чутливої до втрат. Щоб оцінити ефективність цього методу, дослідники провели експерименти на декількох встановлених еталонних наборах даних, використовуючи різні метрики оцінки, такі як G-середнє, площа під ROC-кривою (AUC), скориговане G-середнє, точність, істинно позитивна частота, істинно негативна частота і F1-бали. У дослідженні порівнювалися результати, отримані за допомогою цього нового підходу, з результатами відомих алгоритмів. Результати продемонстрували ефективність і стійкість запропонованого методу, особливо в сценаріях із незбалансованими наборами даних. Підхід виявився успішним у створенні добре збалансованих класифікаторів, вирішуючи проблеми, пов'язані з різним ступенем незбалансованості даних.

У роботі [12] Основною концепцією є представлення та пояснення SegNet архітектури, призначеної для семантичної попиксельної сегментації. Автор обговорює мотивацію створення SegNet, підкреслюючи необхідність точного зіставлення об'єктів низької роздільної здатності з вхідною роздільною здатністю для попиксельної класифікації. Її порівнюють з іншими архітектурами сегментації (FCN, DeepLab-LargeFOV та DeconvNet.), висвітлюючи компроміси між використанням пам'яті та точністю. У статті оцінюється продуктивність SegNet у задачах сегментації сцени і підкреслюються її можливості в реальному часі, особливо для автономного водіння. Отже, враховуючи переваги даної архітектури з точки зору продуктивності та практичного застосування у сценаріях реального часу, доцільно застосовувати її для класифікації об'єктів на місцевості.

Враховуючи потенційний успіх фахівців із машинного навчання у підвищенні точності класифікації об'єктів на місцевості, залишається ряд недоліків, які суттєво впливають на різницю між фактичними та очікуваними результатами щодо вирішення завдань класифікації об'єктів на місцевості. До них варто віднести дисбаланс класів, ініціалізацію ваг нейронної мережі, інтеграцію додаткових модальностей, забезпечення стійкості моделей до шуму та мінливості умов навколишнього середовища. Комбінування запропонованих в аналізі методів попередньої обробки даних та параметризації моделей нейронних мереж дозволить мінімізувати різницю між фактичними та очікуваними показниками точності класифікації об'єктів на місцевості.

Метою статті є підвищення точності класифікації об'єктів на місцевості на основі поєднання сенсорних модальностей із використанням бінарної кросентропії як функції втрат в умовах обмеженої навчальної вибірки.

Авторами статті пропонується синергетичний підхід, який базується на використанні архітектури SegNet, представленої на рисунку 1, розробленої для задач семантичної сегментації, оскільки ця архітектура передбачає віднесення кожного пікселя зображення до одного з декількох попередньо визначених класів об'єктів, забезпечуючи детальне розуміння сцени.

Серед структурних особливостей даної архітектури варто відзначити використання шару softmax у кінці декодера, щоб призначити ймовірності класів кожному пікселю, створюючи остаточну карту сегментації. Однією з відмінних рис архітектури SegNet є використання індексів під час об'єднання шарів у кодері. Замість того, щоб зберігати фактичні значення пікселів у цих індексах, SegNet зберігає індекси максимальних значень у кожній області об'єднання. Ця інформація має вирішальне значення для збільшення вибірки у декодері.

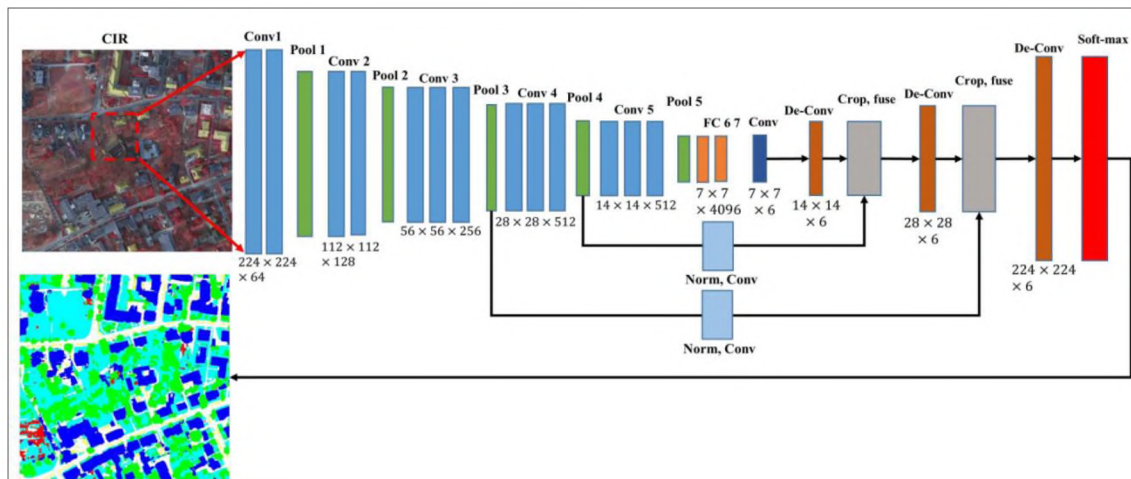


Рис. 1. Архітектура SegNet зі структурою кодер-декодер

Враховуючи попередню навченість моделі на наборі даних ImageNet, залишається необхідним донавчання моделі з метою адаптації під конкретні умови застосування. Набори даних для донавчання та тестування були отримані з репозиторію <https://registry.opendata.aws/spacenet>, вони містили багаторакурсні супутникові знімки високої роздільної здатності, що використовувались у рамках проекту USSOCOM Urban 3D Challenge та аерофотознімки, опубліковані Міжнародним товариством фотограмметрії (ISPRS). Кожен із 144 знімків місцевості містить приблизно 157 000 анотованих контурів будівель, має розмір 2048×2048 пікселів і займає площу приблизно 4 кв. км (рис. 2).

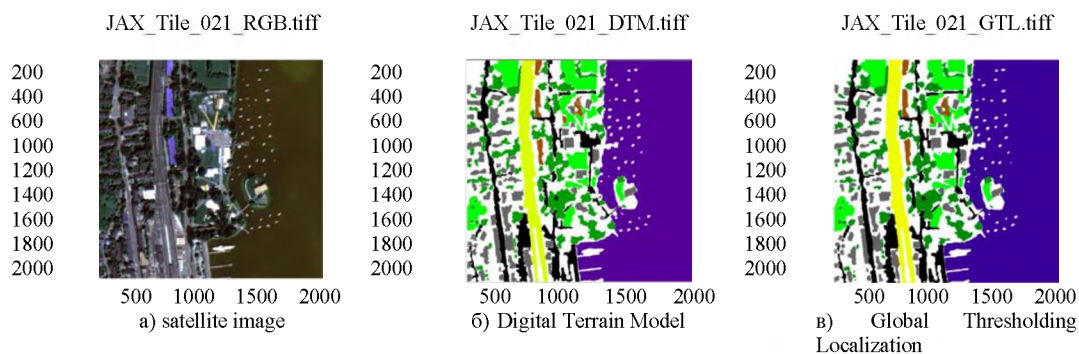


Рис. 2. Зразок зображення з USSOCOM Urban 3D Challenge dataset

Геометрична корекція зображень була здійснена шляхом виправлення геометричних спотворень, спричинених варіаціями рельєфу та кривизною Землі за допомогою ground control points (GCP). Інтеграція даних про висоту об'єктів необхідна з метою налаштування перспективи зображень відповідно до рельєфу місцевості, які в подальшому використовуються для генерації nDSM. Інформація nDSM створюється шляхом щільного зіставлення зображень за допомогою програмного забезпечення Trimble INPHO 5.3. 32-бітові рівні сірого використовуються для кодування висоти та зберігаються у файлі формату TIFF.

У підсумку було сформовано три навчальні вибірки з набору даних USSOCOM, до першої увійшли гіперспектральні орторектифіковані RGB зображення, до другої – nDSM, композитні зображення RGB + nDSM до третьої, та три навчальні вибірки з набору даних ISPRS, до першої увійшли мультиспектральні зображення IRRG, до другої – nDSM та композитні зображення IRRG + nDSM до третьої відповідно. Співвідношення розподілу

становило 70 % для навчання, 15 % для перевірки і 15 % для тестування від загального набору даних.

Налаштування моделі SegNet містить конфігурацію її архітектури, вхідних вимірів, вихідних класів, функції активації, функції втрат, оптимізатора, параметрів навчання, методів регуляризації, метрик оцінювання, а також проведення навчання і тестування. Враховуючи використання попередньо-навченої моделі на великому наборі даних ImageNet, що дозволило моделі вивчити корисні ієрархічні особливості, які можна використовувати для конкретних задач класифікації об'єктів, залишалось провести точне налаштування та оновлення вагових коефіцієнтів моделі відповідно до визначеного набору даних, а також застосувати функцію втрат на основі бінарної перехресної ентропії для попіксельної класифікації, у зв'язку з тим, що дана функція вимірює розбіжність між прогнозованими ймовірностями класів та істинними мітками. У налаштуваннях бінарного та багатокласового класифікатора однією з основних цілей оптимізації є мінімізація функції втрат на основі двовимірної бінарної перехресної ентропії (1).

$$\theta_b^* = \arg \min_{\theta \in \Theta} \left(- \sum_{i=1}^m \sum_{j=1}^n Y_{i,j} \log(P(X; \theta)_{i,j}) + (1 - Y_{i,j}) \log(1 - P(X; \theta)_{i,j}) \right), \quad (1)$$

де $Y_{i,j} \in \{0, 1\}$ і $P(X; \theta)_{i,j} \in [0, 1]$ позначають основні істинні та передбачені мітки відповідно у пікселі (i, j) для кожного фрагмента навчального зображення; X позначає вхідні функції (1-, 3- або 4-канальне зображення). X і Y індексуються як двовимірні масиви з (висотою, шириною), заданими як (m, n) .

Мультикласовий класифікатор зводить до мінімуму аналогічне завдання пошуку параметрів Θ_m^* (2), за винятком того, що замість розгляду одного класу, наприклад, будівля, цей класифікатор робить попіксельні прогнози з кількох класів. Цей класифікатор навчається шляхом мінімізації наступної задачі, яка відповідає негативній попіксельній логарифмічній правдоподібності за N класами.

$$\theta_m^* = \arg \min_{\theta \in \Theta} \left(- \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^N Y_{i,j,k} \log(P(X; \theta)_{i,j,k}) \right), \quad (2)$$

де $Y_{i,j,k} \in \{0, 1\}$ і $P(X, \theta)_{i,j,k} \in [0, 1]$ позначають основні істинні та передбачені мітки відповідно в пікселі (i, j) для класу $k \in \{0, 1, \dots, N\}$ для кожного навчального зображення X , яке індексується як двовимірний масив форми (m, n) . Щоб зробити оптимальні прогнози семантичної сегментації, використовується метод градієнтного спуску для ітеративного рішення для набору ваг θ , які мінімізують цілі, представлені вище.

До того ж обчислення поелементної крос-ентропійної втрати для кожного пікселя зображення необхідно зважити кожен піксель, використовуючи його зворотну частоту класу, тобто якщо клас i зустрічається в наборі даних із часткою $f_i \in (0, 1]$, то вага визначається так: $w_i = 1/f_i$. Даний вираз визначення ваг використовується для виправлення дисбалансу класів у наборах даних та відповідає зваженій негативній логарифмічній правдоподібності або виваженій меті крос-ентропії. Використання виваженої функції вартості дозволяє вивчити класифікатор, не орієнтований на найбільш поширений клас. Отже виважена бінарна та багатокласова класифікація виглядатиме так (3), (4):

$$\theta_{b,w}^* = \arg \min_{\theta \in \Theta} \left(- \sum_{i=1}^m \sum_{j=1}^n w(Y_{i,j}) (Y_{i,j} \log(P(X; \theta)_{i,j}) + (1 - Y_{i,j}) \log(1 - P(X; \theta)_{i,j})) \right), \quad (3)$$

де $w(Y_{i,j})$ позначає вагу, присвоєну пікселю (i,j) на основі його основного класу істинності $Y_{i,j}$. Слід зауважити, що $Y_{i,j} \in \{0, 1\}$.

$$\theta_{m,w}^* = \arg \min_{\theta \in \Theta} \left(- \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^N w(c_{i,j}) Y_{i,j,k} \log(P(X; \theta)_{i,j,k}) \right), \quad (4)$$

де $w(c_{i,j})$ позначає вагу, присвоєну пікселю (i,j) на основі його основного класу істинності $c_{i,j}$. Відповідно для позначення класу використовується $c_{i,j} \in \{0, 1, \dots, N\}$, а не $Y_{i,j,k} \in \{0, 1\}$. Описаний вище математичний апарат представлено у [13].

Процес навчання, як правило, передбачає ітерації над усім навчальним набором даних протягом заданої кількості епох або до досягнення збіжності. Кожна епоха являє собою один повний прохід через набір даних, під час якого вхідні ділянки зображення подаються в модель SegNet, модель обчислює попиксельні карти ймовірностей класів, а втрати бінарної перехресної ентропії обчислюються шляхом порівняння прогнозованих ймовірностей класів із відповідними істинними мітками. Зворотне розповсюдження відбувається шляхом обчислення градієнтів відносно втрат, і ваги моделі відповідно оновлюються за допомогою зворотного оптимізатора, який підлаштовує ваги моделі для мінімізації функції втрат. Безперервна перевірка продуктивності моделі SegNet на валідаційному наборі даних під час навчання здійснюється з метою відстеження надмірного припасування і гарантує, що модель добре узагальнює раніше невидимі дані.

У рамках першого експерименту мережа була навчена з використанням пар ділянок вхідних зображень з набору ISPRS і відповідних базових карт міток, які містили шість анотованих класів об'єктів. Час навчання становив приблизно 20 годин. За результатами цього експерименту встановлено, що наявність додаткових модальностей в умовах узгодженого геопросторового контексту майже не впливає на продуктивність класифікації, підвищуючи показники точності для деяких класів об'єктів лише на 0,5 %. Але зовсім інша ситуація спостерігається під час класифікації об'єктів на раніше небаченій місцевості. Узагальнена точність класифікації поза вибіркою зростає на 23,9 % порівняно з випадком IRRG та на 17,1 % порівняно з випадком nDSM (табл. 1).

Таблиця 1

Результати дослідження для набору даних ISPRS

Класи об'єктів	segnet (isprs)					
	validation dataset			test dataset		
	ndsm	irrg	ndsm & irrg	ndsm	irrg	ndsm & irrg
Непроникні поверхні	0,8727	0,9520	0,9531	0,7127	0,7502	0,8374
Будинки	0,9549	0,9638	0,9722	0,6828	0,4571	0,7886
Низька рослинність	0,8486	0,9199	0,9243	0,7320	0,7829	0,8589
Дерева	0,9159	0,9488	0,9473	0,8846	0,8568	0,8643
Авто	0,9922	0,9969	0,9959	0,9865	0,9879	0,9912
Інші об'єкти	0,9995	0,9993	0,9996	0,9518	0,9598	0,9522
Разом	0,7919	0,9003	0,8962	0,4752	0,3974	0,6463

Другий експеримент проводився для набору даних USSOCOM із використанням кількісних метрик бінарної класифікації, таких як точність, прецизійність, повнота, а також частота хибнонегативних та хибнопозитивних результатів. Результати експерименту демонструють, що класифікатори, навчені за допомогою інформації RGB та nDSM, у середньому мають на 6 % вищу точність поза вибіркою, ніж класифікатори, навчені лише інформацією RGB. Ці результати підтверджують те, що інформація nDSM сприяє більшій узагальнюваності класифікатора порівняно з лише інформацією RGB для вирішення завдань класифікації (табл. 2).

Таблиця 2

Результати дослідження для набору даних USSOCOM

Метрики класифікації	segnet (ussocom)					
	validation dataset			test dataset		
	ndsm	rgb	ndsm & rgb	ndsm	rgb	ndsm & rgb
Точність	0,9164	0,9298	0,9367	0,8690	0,8739	0,9386
Прецизійність	0,9245	0,9412	0,9451	0,9425	0,9416	0,9512
Повнота	0,9105	0,9245	0,9341	0,8122	0,9307	0,9298
Хибнонегативні	0,0895	0,0755	0,0659	0,1878	0,0693	0,0702
Хибнопозитивні	0,0829	0,0643	0,0604	0,0610	0,0626	0,0518

Враховуючи результати попередніх експериментів, очевидним є те, що RGB + nDSM забезпечує найкращу кореляцію з фактичною інформацією про місцевість. Це обумовлено тим, що інформація про висоту допомагає розрізняти об'єкти зі схожими спектральними характеристиками, але різною висотою, наприклад, будівлі або дерева, забезпечує зменшення надмірності даних. Покращена дискримінація nDSM допомагає розрізняти об'єкти, які мають однакову спектральну сигнатуру на стандартних знімках, наприклад, різні типи дахів, дорожніх покриттів або рослинного покриву на основі їхніх висотних профілів. Важливим аспектом проведення експериментів було тренування моделі з обмеженими наборами даних, адже не завжди є можливість використовувати достатню кількість даних з адекватним поданням і кутами огляду для всіх об'єктів, що цікавлять, особливо для завдань, в яких критичним ресурсом виступає час. У таблиці 3 показано результати навчання класифікаторів з 15 %, 25 %, 50 % та 100 % від загального тренувального набору даних (TDS) із проміжним контролем за допомогою зображень із валідаційного набору даних (VDS).

Таблиця 3

Результати аналізу впливу обмеження вибірки на точність класифікації

Пропорції	Точність класифікації					
	In city				Out of city TDS	
	VDS		TDS		+ nDSM	- nDSM
	+ nDSM	- nDSM	+ nDSM	- nDSM		
15 %	0,9054	0,9003	0,8356	0,8263	0,5336	0,2258
25 %	0,9121	0,9174	0,8649	0,7982	0,4159	0,4367
50 %	0,9215	0,9211	0,8923	0,8815	0,4677	0,3917
100 %	0,9367	0,9298	0,9386	0,9339	0,5112	0,4729

Результати розпізнавання об'єктів на пересіченій місцевості показують, що існує помірна позитивна кореляція між часткою використаних даних і точністю. Це свідчить про те, що використання лише 50 % даних може бути достатнім для відносно точної класифікації. При оцінці точності в межах валідаційного набору даних для міських районів спостерігається негативна кореляція між розміром підмножини набору даних і точністю. Це означає, що зі зменшенням розміру підмножини даних точність у цій підмножині також має тенденцію до зниження. Це спостереження може бути пов'язане з тим, що нейронна мережа має менше зразків для навчання, що ускладнює її адаптацію до навчальної вибірки. Що стосується раніше небачених зображень, то тут також спостерігається негативна кореляція між точністю та часткою навчального набору даних, це свідчить про те, що використання лише RGB-даних може призвести до надмірного припасування, коли модель добре працює на навчальних даних, але не може узагальнити їх для нових міст чи районів.

Висновки. У статті вирішується завдання підвищення точності класифікації об'єктів на місцевості завдяки комбінуванню існуючих методів попередньої обробки даних та параметризації моделей нейронних мереж. Фахівці з машинного навчання здебільшого концентрують свої зусилля на вирішенні специфічних актуальних прикладних задач щодо покращення конкретних аспектів продуктивності нейронних мереж.

Проведено аналіз існуючих підходів до машинного навчання з метою пошуку способів оптимізації параметрів моделей нейронних мереж для класифікації об'єктів на місцевості, серед яких варто зазначити: вплив мультимодального поєднання даних (nDSM, SAR) на точність класифікації об'єктів на місцевості в умовах обмеженої навчальної вибірки; використання попередньо навчених моделей нейронних мереж (LinkNet, UNet, SegNet) та архітектур (DSCNN, EMatMHKS); використання класифікаторів нейронних мереж на основі методів найближчого сусіда та нечітких плаваючих центроїдів та двовимірної бінарної кросентропії.

Результати проведеного аналізу підходів до підвищення точності класифікації об'єктів на місцевості демонструють вплив модальності nDSM на точність класифікації, особливо в умовах обмеження навчальних даних, а класифікатор на основі бінарної кросентропії дозволяє усунути вплив дисбалансу класів на точність класифікації.

Запропоновано синергетичний підхід на основі моделі SegNet, навченої на попередньо орторектифікованих зображеннях, у поєднанні з nDSM для навчання моделі розпізнавати об'єкти в узгодженому геопросторовому контексті та використанні класифікаторів на основі двовимірної бінарної кросентропії, що дозволило усунути вплив дисбалансу класів на точність класифікації об'єктів на місцевості.

Результати експериментів, наведені у таблицях 1–3, демонструють, що узагальнена точність класифікації для шести класів об'єктів на тестовому наборі даних зростає на 23,9 % порівняно з випадком IRRG та на 17,1 % порівняно з випадком nDSM. Класифікатори, навчені за допомогою інформації RGB та nDSM, у середньому мають на 6 % вищу точність поза вибіркою, ніж класифікатори, навчені лише інформацією RGB. А експерименти з обмеженням навчальної вибірки продемонстрували, що за потреби достатньо використати половину доступного тренувального набору даних для отримання лише на 4 % нижчої точності класифікації та заощадження близько 10 годин навчання.

Отже, на великих і добре відомих наборах даних SegNet є конкурентоспроможною, досягаючи високих показників точності, особливо з використанням інформації про тривимірну поверхню та в умовах обмеженої навчальної вибірки.

Напрямок подальших досліджень є удосконалення існуючих методів впливу на метрику Каппа Коена, яка використовується для оцінки збігу між прогнозованою та фактичною класифікаціями, враховуючи при цьому збіг, який може виникнути випадково

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Козуб А. М., Шумейко В. О., Зуйко В. В., Ніколаєнко О. Є. Аналіз досвіду використання космічних систем та безпілотних авіаційних комплексів в сучасних локальних конфліктах та перспективи їх розвитку // НУОУ. Сучасні інформаційні технології у сфері безпеки та оборони. 2018. № 3 (33).
2. Епішев В. П., Мотрунич І. І., Періг В. М., Кудак В. І., Найбауер І. Ф., Сухов П. П., Кашуба В. І., Сухов К. П., Варламов І. Д., Албул В. В., Москаленко С. С., Мисливий С. О. Можливості національних оптичних засобів спостереження за космічним простором щодо контролю геостационарної орбіти у інтересах збройних сил України // НУОУ. Сучасні інформаційні технології у сфері безпеки та оборони. 2018. № 3 (33).
3. ImageNet: website. URL: <https://www.image-net.org/index.php>.
4. M. Weinmann and M. Weinmann. Geospatial computer vision based on multi-modal data. 2018. DOI: 10.3390/rs10010002.
5. Goldberg H., Brown M. and Wang S. A benchmark for building footprint classification using orthorectified rgb imagery and digital surface models from commercial satellites. Proceedings of IEEE applied imagery pattern recognition workshop. 2017. DOI: 10.1109/AIPR.2017.8457973.
6. Chen K., Weinmann M., Gao X., Yan M., Hinz S., Jutzi B. and Weinmann M. Residual shuffling convolutional neural networks for deep semantic image segmentation using multi-modal data. ISPRS annals

of photogrammetry, remote sensing & spatial information sciences. 2018. DOI: 10.5194/isprs-annals-IV-1-29-2018.

7. Medium: website. URL: <https://medium.com/the-downlinq/spacenet-6-announcing-the-winners-df817712b515>.

8. Wang L., Yang B., Chen Y., Zhang X., Orchard J. Improving neural-network classifiers using nearest neighbor partitioning. IEEE Trans NNL System. 2017. DOI: 10.1109/TNNLS.2016.2580570.

9. Shuangrong Liu, Lin Wang, Bo Yang, Jin Zhou, Zhenxiang Chen, Hui fen Dong. Improvement of Neural-Network Classifiers Using Fuzzy Floating Centroids. DOI: 10.1109/TCYB.2020.2987904.

10. Zhu C., Wang Z. Entropy-based matrix learning machine for imbalanced data sets. Pattern Recognit Lett. 2017. DOI: 10.1016/j.patrec.2017.01.014.

11. Aurelio, YS, de Almeida, GM, de Castro, CL and Braga, AP. Learning from imbalanced data sets with weighted cross-entropy function. 2019. DOI: 10.1007/s11063-018-09977-1.

12. Badrinarayanan V., Kendall A. and Cipolla R. Seg-Net: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Analytic. Mach Intelligence. 2017. DOI: 10.1109/TPAMI.2016.2644615.

13. Daniel Voigt Godoy. Deep Learning with PyTorch Step-by-Step: A Beginner's Guide. Copyright 2020-2022.