

Підвищення ефективності навчання бортової системи детектування об'єктів для малогабаритного безпілотного апарату

В. В. Москаленко, А. С. Довбиш, І. В. Науменко, А. С. Москаленко,
А. Г. Коробов

Запропоновано модель детектора об'єктів і критерій ефективності навчання моделі. Модель містить 7 перших модулів згорткової мережі SqueezeNet, два згорткові різномасштабні шари, та інформаційно-екстремальний класифікатор. Як критерій ефективності навчання моделі детектора розглядається мультиплікативна згортка частинних критеріїв, що враховує ефективність виявлення об'єктів на зображенні та точність класифікаційного аналізу. При цьому додаткове використання алгоритму ортогонального узгодженого кодування при обчисленні високорівневих ознак дозволяє збільшити точність моделі на 4 %.

Розроблено алгоритм навчання детектора об'єктів за умов малого обсягу розмічених навчальних зразків та обмежених обчислювальних ресурсів, доступних на борту малогабаритного безпілотного апарату. Суть алгоритму полягає в адаптації верхніх шарів моделі до доменної області використання на основі алгоритмів зростаючого розрідженого кодування нейронного газу та симуляції відпалу. Навчання верхніх шарів без вчителя дозволяє ефективно використати нерозмічені дані з доменної області та визначити необхідну кількість нейронів. Показано, що за відсутності тонкої настройки згорткових шарів забезпечується 69 % виявлених об'єктів на зображеннях тестової вибірки Inria Aerial Image. При цьому після тонкої настройки на основі алгоритму симуляції відпалу забезпечується 95 відсотків виявлених об'єктів на тестових зображеннях.

Показано, що використання попереднього навчання без вчителя дозволяє підвищити узагальнюючу здатність вирішальних правил та прискорити ітераційний процес знаходження глобального максимуму при навчанні з учителем на вибірці обмеженого обсягу. При цьому усунення ефекту перенавчання здійснюється шляхом оптимального вибору значення гіперпараметру, що характеризує ступінь покриття вхідних даних нейронами мережі

Ключові слова: зростаючий нейронний газ, детектор об'єктів, інформаційний критерій, алгоритм симуляція відпалу

1. Вступ

Безпілотна авіація знаходить широке застосування в задачах інспекції технологічних та житлових споруд, охоронній та розвідувальній діяльності, а також в сфері транспортування малогабаритних вантажів. Одним із шляхів підвищення функціональної ефективності безпілотного апарату є впровадження

технологій штучного інтелекту для аналізу сенсорної інформації. Оскільки найбільш інформативною сенсорною системою є камери відеоспостереження, то розробка візуальних детекторів об'єктів інтересу є перспективним напрямком. Однак обмежені обчислювальні ресурси та вага безпілотного апарату не дозволяють реалізувати в ньому моделі аналізу візуальних даних, адаптовані для всього спектру можливих умов спостереження та різноманіття модифікацій об'єктів інтересу. Це обумовлює потребу в розробці обчислювально ефективних моделей і алгоритмів адаптації до нових умов функціонування, притаманних конкретній доменній області застосування.

З точки зору обчислювальної складності і узагальнюючої здатності лідером серед моделей аналізу візуальної інформації є згорткові нейронні мережі. Однак навчання та донавчання згорткових мереж потребує вибору оптимальної конфігурації мережі та значних обчислювальних ресурсів. Зменшити обчислювальне навантаження при донавчанні нейронної мережі для адаптації до нових умов функціонування можна за рахунок використання принципу переносу знань шляхом копіювання перших шарів мережі навченої на датасеті ImageNet [1, 2]. Однак шари високорівневого ознакового подання спостережень потребують навчання з нуля. При цьому необхідну кількість згорткових фільтрів в кожному згортковому шарі наперед оцінити важко, тому перспективним підходом до навчання згорткових фільтрів є використання принципів зростаючого нейронного газу, який дозволяє автоматично визначити необхідну кількість нейронів. При цьому шари вирішальних правил детектора та шари екстракції ознак потребують тонкої настройки, яку як правило реалізують на основі модифікацій алгоритму зворотного поширення помилки [2, 3]. Однак даний алгоритм характеризується низькою швидкістю збіжності та застряганням в локальних екстремумах функції втрат. Існують альтернативні метаевристичні алгоритми пошукової оптимізації, однак ефективність використання цих алгоритмів в задачах тонкої настройки мереж є малодослідженою [3].

Тому робота, спрямована на підвищення ефективності навчання детектора об'єктів за умов обмежених обчислювальних ресурсів і навчальних даних, є актуальною.

2. Аналіз літературних даних і постановка проблеми

У працях [4, 5] було досліджено моделі ознакового опису візуальних спостережень, що основані на використанні ознак Хаара, гістограм орієнтованих градієнтів, локальних бінарних шаблонів, гістограм дескрипторів візуальних слів та інших низькорівневих ознак. Дані моделі використовують низькорівневу локальну інформацію та ігнорують високорівневу контекстну інформацію, що ускладнює реалізацію ефективних детекторів об'єктів за умов обмежених обсягів розмічених навчальних даних. Крім того неієрархічні моделі ознакового опису характеризуються високою обчислювальною трудомісткістю за умов великої варіації спостережень.

У задачах аналізу зображень набули популярності численні моделі ієрархічного опису спостережень, основані на згорткових нейронних мережах.

Серед них найбільш популярні VGG-16, VGG-19, ResNet-50, GoogleNet, MobileNet, SqueezeNet та інші [3, 4]. Ці мережі відрізняються кількістю шарів, наявністю залишкових зв'язків та різномасштабних фільтрів в кожному з шарів. При цьому в працях [5, 7] було показано, що навчені моделі на датасеті ImageNet акумулюють в собі важливу інформацію щодо аналізу візуальних образів. Проте чим сильніше відрізняється цільова доменна область використання мережі від образів ImageNet, тим менше шарів навчених мереж можна повторно використати. Крім того звуження доменної області, наприклад, шляхом зменшення кількості класів розпізнавання дозволяє зменшити потреби в обчислювальних ресурсах для функціонування детектора об'єктів на місцевості.

У працях [8, 9] було запропоновано здійснювати тонку настройку детектора об'єктів, оснований на згортковій мережі VGG-16, шляхом використання алгоритму міні-пакетного зворотного поширення помилки. Проте це потребувало значного обсягу навчальних даних та декілька годин роботи на графічному процесорі персонального комп'ютера. У праці [10] було запропоновано здійснювати сканування нормалізованої високорівневої карти ознак ковзним вікном, в кожній позиції якого здійснюється класифікаційний аналіз. У праці [11] було запропоновано використовувати в рамках так званої інформаційно-екстремальної технології (IEI-технологія) класифікаційний аналіз високорівневого ознакового подання. Ця технологія основана на адаптивному двійковому кодуванні і побудові оптимальних в інформаційному розумінні радіально-базисних вирішальних правил в двійковому просторі Хеммінга. Однак дані підходи мають не високу ефективність, оскільки високорівневі карти ознак не достатньо адаптовані до доменної області використання, а ефективних технік тонкої настройки мережі не було запропоновано. Крім того мережа VGG-16 є обчислювально складною і тому її використання не забезпечує прийняття рішень в режимі реального часу в умовах обмежених ресурсів автономного малагабаритного апарату.

Здійснення навчання високорівневих шарів нейромережі для адаптації до доменної області шляхом попереднього навчання без вчителя і наступної тонкої настройки на основі навчання з учителем є поширеною схемою навчання глибоких мереж. У працях [11, 12] пропонується навчання згорткових мереж без вчителя на основі автоенкодера або обмеженої машини Больцмана, які потребують великий обсяг навчальних даних і тривалий час навчання для отримання прийняттого результату. У працях [11, 14] пропонується поєднання принципів нейронного газу та розрідженого кодування для навчання згорткових фільтрів за нерозміченими даними. Даний підхід має м'яку конкуруючу схему навчання, що підвищує ймовірність збіжності алгоритму до оптимального розподілу нейронів на навчальних даних. При цьому вбудовування алгоритмів розрідженого кодування дозволяє підвищити завадозахищеність і узагальнюючу здатність ознакового подання. Однак кількість нейронів наперед невідома і задається на розсуд розробника, або оптимізується, що призводить до збільшення кількості ітерацій навчання.

Необхідну кількість згорткових фільтрів у високорівневих згорткових шарах наперед оцінити важко, тому перспективним підходом до навчання згорткових фільтрів є використання принципів зростаючого нейронного газу, який дозволяє автоматично визначити необхідну кількість нейронів [15]. Однак механізм вставки нових нейронів в алгоритмі нейронного газу на основі задавання періоду вставки часто призводить до викривлення утворених структур і нестабільності процесу навчання. Проте у праці [16] було показано, що забезпечити стабільність навчання можна шляхом задавання «радіусу досяжності» нейронів. Це передбачає заміну періоду вставки нейронів на поріг максимального віддалення нейрону від кожної з віднесених до нього точок навчальної множини. Однак досі не було переглянуто механізми оновлення нейронів та оцінки віддаленості точок вхідного простору до нейронів з метою адаптації процесу навчання до процедури розрідженого кодування спостережень.

Задачу детектування об'єктів на основі карти ознак згорткової мережі вирішують на основі архітектур Yolo (You only look once), Faster R-CNN, DetectNet та SSD (Single Shot MultiBox Detector) [7]. Серед них SSD архітектура набула найширшого застосування в мобільних пристроях, оскільки вона має найменшу обчислювальну складність в режимі детектування. Проте навчання таких шарів за умов обмеженого обсягу навчальних даних та обчислювальних ресурсів на основі зворотного поширення помилки і стохастичного градієнтного спуску є малоефективним. Одним із перспективних шляхів навчання нейронних мереж і тонкої настройки є застосування метаевристичних алгоритмів, оскільки вони характеризуються кращою збіжністю і меншою ймовірністю застрягання в "поганому" локальному оптимумі [17]. Серед них варто виділити алгоритм симуляції відпалу, оскільки його використання дозволило перевищити результати алгоритмів градієнтного спуску при оптимізації нейромережових класифікаторів [4]. Однак його використання в задачах тонкої настройки згорткових фільтрів та навчання детектора об'єктів залишається малодослідженим.

3. Ціль і задачі дослідження

Мета роботи полягає у підвищенні ефективності навчання детектора об'єктів за умов обмежених обчислюваних ресурсів бортової системи безпілотного апарату та розмічених навчальних даних заданої доменної області застосування. Для досягнення поставленої мети пропонується розв'язання таких задач:

- розробити модель аналізу даних для детектування об'єктів інтересу на аерофотознімках за умов обмежених обчислювальних ресурсів і розміру розміченої навчальної вибірки з доменної області застосування;

- розробити алгоритм навчання детектора об'єктів з використанням ідей та принципів зростаючого нейронного газу і розрідженого кодування для попереднього навчання без вчителя та метаевристичної оптимізації для тонкої настройки;

– дослідити залежність ефективності навчання детектора об'єктів на місцевості за навчальними та тестовими даними доменної області від параметрів алгоритму навчання без вчителя високорівневого ознакового опису спостережень.

4. Модель і алгоритми навчання детектора об'єктів

4.1 Модель детектора об'єктів

Для розв'язання задачі розробки моделі аналізу даних за умов обмежених обчислювальних ресурсів малогабаритного апарату слід максимально використати всю доступну апіорну інформацію. Використання доступної апіорної інформації для синтезу архітектури згорткової мережі для аналізу візуальної інформації за умов обмежених обчислювальних ресурсів малогабаритного апарату може підвищити ефективність машинного навчання, перенавчання та екзамену. Техніка перенесення знань є одним з прикладів використання апіорної інформації, акумульованої в навченій мережі для повторного використання. При цьому, чим ближче доменна область, де нейронна мережа була навчена, до нової доменної області, де накопичені знання будуть використовуватися, тим більше шарів можуть бути запозичені.

Високорівневі шари нейронної мережі, як правило, потребують навчання з нуля для максимальної адаптації до нової доменної області. Для цього пропонується використати додаткові згорткові шари, де фільтри з ядром 1×1 та 3×3 попередньо навчаються без вчителя. При цьому для забезпечення завадозахищеності та інформативності ознакового подання активацію кожного пікселя карти ознак пропонується обчислювати на основі алгоритму ортогонального узгодженого переслідування (ОУП) з функцією ReLU [12, 14].

Для багатьох задач моніторингу місцевості немає потреби точної локалізації об'єктів, а достатньо лише інформації про їх знаходження в певній зоні. Тому при розробці шарів детектування об'єктів на аеро зображеннях місцевості за основу береться архітектура Single Shot Detection, але не використовується регресія для уточнення меж об'єктів [7]. Крім того, в умовах обмеженого обсягу розмічених навчальних даних зведення задачі детектування до задачі класифікаційного аналізу є доцільним, оскільки для прийнятних результатів регресійного аналізу може бути недостатньо даних. При цьому класифікаційний аналіз карти ознак пропонується здійснювати в рамках так званої ІЕІ-технології, оскільки вона дозволяє синтезувати класифікатор з низькою обчислювальною складністю і порівняно високою достовірністю за умов обмежених обсягів навчальних зразків [11]. На рис. 1 схематично показано архітектуру детектора, що пропонується.

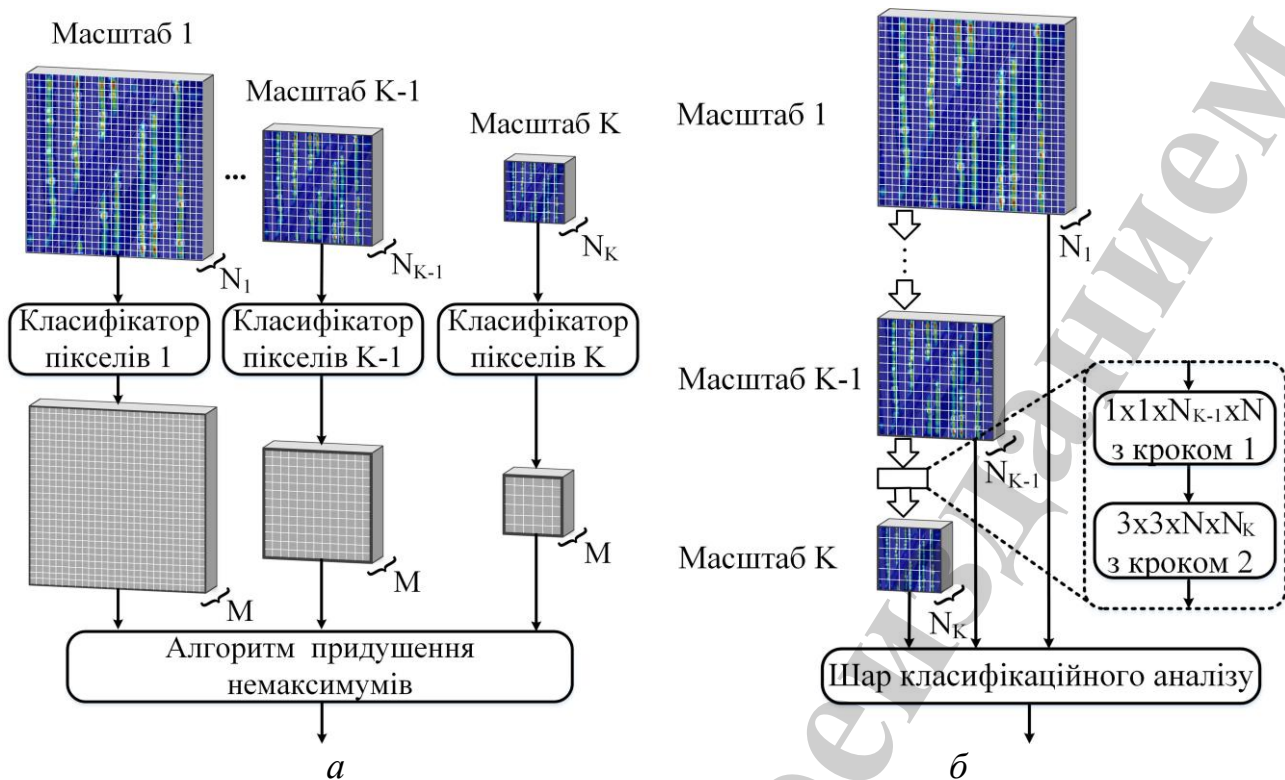


Рис. 1. Узагальнена архітектура детектора: *a* – схема детектора об'єктів з алфавіту M класів; *б* – схема формування різномасштабних ознакових карт, кожен піксель якої описується N_1, \dots, N_{k+1} ознаками

Навчальна вибірка для класифікаційного аналізу формується за результатами обчислення міри перетину прямокутника об'єкта, нанесеного експертом, з проекцією пікселя карти ознак на вхідне зображення (рецептивне поле пікселя). Якщо міра перетину Жаккарда IoU більша порогового значення Th , то відносимо піксель карти ознак до позитивного навчального зразка, а решту – до негативних.

Інформаційно-екстремальний класифікатор, що оцінює належність патчу зображення до одного з R класів, здійснює дискретизацію ознакового опису в навчальній матриці $\{x_{r,i}^{(j)} \mid i = \overline{1, N}; j = \overline{1, n_r}; r = \overline{1, R}\}$ шляхом порівняння значення i -ї ознаки з відповідними межами l -го, $l = \overline{1, L}$, одномірною рецептивного поля. Тобто формування двійкової навчальної матриці $\{b_{r,i}^{(j)} \mid i = \overline{1, N \cdot L}; j = \overline{1, n_r}; r = \overline{1, R}\}$ здійснюється за правилом

$$b_{r,l \cdot N + i}^{(j)} = \begin{cases} 1, & \text{if } x_{i, \max} \left[1 - \delta_{i,l} / \delta_{\max} \right] \leq x_{r,i}^{(j)} \leq x_{i, \max}; \\ 0, & \text{otherwise.} \end{cases}$$

Обчислення значень координат двійкового еталонного вектору x_r , відносно якого відбувається побудова в радіальному базисі контейнерів класів, здійснюється за правилом

$$b_{r,l,N+i} = \begin{cases} 1, & \text{if } \frac{1}{n_r} \sum_{j=1}^{n_r} b_{r,l,N+i}^{(j)} > \frac{1}{R} \sum_{c=1}^R \frac{1}{n_c} \sum_{j=1}^{n_c} b_{c,l,N+i}^{(j)}; \\ 0, & \text{otherwise.} \end{cases}$$

В рамках ІЕІ-технології для підвищення оперативності навчання задачу багатокласової класифікації прийнято зводити до серії двохкласових за принципом «кожний проти всіх». При цьому для уникнення проблеми незбалансованості класів, внаслідок переважання негативних фонових зразків, альтернативним до r -го класу є синтетичний клас. Синтетичний клас представлений з n_r зразків решти класів, найближчих до еталонного вектору b_r , де n_r є обсягом навчальної вибірки r -го класу.

Як критерій ефективності машинного навчання класифікатора розглядається нормована модифікація інформаційної міри Кульбака [11]:

$$E_r = \frac{1 - (\alpha_r + \beta_r)}{\log_2(2 + \zeta) - \log_2 \zeta} \cdot \log_2 \left[\frac{2 - (\alpha_r + \beta_r) + \zeta}{(\alpha_r + \beta_r) + \zeta} \right], \quad (1)$$

де α_r, β_r – частота помилок першого та другого роду класифікаційних рішень щодо належності вхідних векторів до r -го класу; ζ – будь-яке мале невід’ємне число, яке вводиться для уникнення невизначеності при діленні на нуль.

Функція належності двійкового представлення b вхідного вектору x до r -го класу, оптимальний контейнер якого має еталонний вектор b_r^* та радіус d_r^* , обчислюється за формулою:

$$\mu_r(b) = \exp \left(- \sum_{i=1}^{N \cdot L} b_i \oplus b_{r,i}^* / d_r^* \right).$$

Для оптимізації вектора параметрів рецептивних полів класифікатора $\{\delta_{i,l} \mid i = \overline{1, N}; l = \overline{1, L}\}$, масштабуючих фільтрів детектора та корекції згорткових фільтрів останнього шару екстракції ознак пропонується використовувати метаевристичний алгоритм симуляції відпалу [4].

Як критерій оптимізації для тренування масштабуючого фільтру і класифікаторів детектора та тонкої настройки останнього шару екстрактора ознак використовується комплексний критерій J . Цей критерій враховує як ефективність J_{Loc} виявлення об’єкта на зображенні, так і ефективність його класифікаційного аналізу J_{Cls} .

$$J = J_{Loc} \cdot J_{Cls}. \quad (2)$$

Критерій оцінки ефективності виявлення об’єктів інтересу на зображенні пропонується обчислювати за формулою

$$J_{Loc} = \frac{1}{n_l} \sum_{i=1}^{n_g} \sum_{j=1}^{n_d} x_{i,j} \cdot \left(\max_r \{ \mu_{j,r} \} - \mu_{j,0} \right),$$

де n_g – кількість прямокутників об'єктів інтересу на розмічених навчальних зображеннях доменної області; n_d – кількість пікселів різномасштабних карт ознак детектора об'єктів; $x_{i,j}$ – індикатор співставлення j -го пікселя карти ознак шару детектування i -му прямокутнику об'єкта інтересу на розмічених зображеннях

$$x_{i,j} = \begin{cases} 1, & \text{if } IoU_{i,j} > Th; \\ 0, & \text{otherwise,} \end{cases}$$

де $IoU_{i,j}$ – коефіцієнт Жакарда, що оцінює ступінь перетину i -го прямокутника об'єкта інтересу на розмічених зображеннях з проекцією на вхідне зображення j -го пікселя карти ознак шару детектування; n_l – кількість пікселів карт ознак детектора, у яких міра перетину проекції на вхідне зображення з прямокутником об'єкта інтересу більша порогового значення Th ; $\mu_{j,r}$ – функція j -го пікселя карти ознак до r -го класу, де $r=0$ відповідає фоновому класу.

Критерій оцінки ефективності класифікаційного аналізу об'єктів інтересу пропонується обчислювати за формулою

$$J_{Cls} = \frac{1}{S} \frac{1}{R} \sum_{s=1}^S \sum_{r=1}^R E_{r,s},$$

де S – кількість різномасштабних карт ознак шарів детектування; $E_{m,s}$ – інформаційний критерій ефективності машинного навчання s -го класифікатора розпізнавати r -й клас.

Для відсіювання зайвих спрацювань детектора на один і той же об'єкт зображення як правило використовують алгоритм придушення немаксимумів та його модифікації. Пропонується використовувати найбільш широко використовуваний жадібний алгоритм придушення немаксимумів. Цей алгоритм ітеративно знаходить детектування з локально найвищим значення функції належності до об'єкта інтересу і відкидає всі сусідні детектування, що перекриваються з ним більше порогового значення.

Таким чином, запропонована модель детектора об'єктів на місцевості призначена не для точної локалізації об'єктів на місцевості, а для максимально точного визначення наявності об'єкта інтересу на заданій місцевості.

4. 2. Алгоритм навчання детектора об'єктів

Для розв'язання задачі розробки алгоритму навчання детектора об'єктів сформуємо основні етапи машинного навчання детектора і розглянемо алгоритми їх реалізації. На першому етапі необхідно обрати попередньо навчену глибоку нейронну мережу і кількість її нижніх шарів, що будуть запозичені згі-

дно техніки переносу знань. Вибір мережі і кількості її шарів можна робити шляхом прямого перебору чи на основі досвіду.

На другому етапі алгоритму пропонується здійснювати навчання верхніх шарів нейронної мережі з використанням зростаючого розріджено кодуючого нейронного газу, що оснований принципах нейронного газу та розрідженого кодування. При цьому вибірка для навчання високорівневих фільтрів згорткової мережі формується шляхом розбиття вхідних зображень або карт ознак на тривимірні патчі, що співпадають з розмірами фільтрів даного шару. Ці патчі змінюють форму до 1D векторів, що надходять на вхід алгоритму зростаючого розріджено кодуючого нейронного газу, основні кроки якого приведено нижче:

1) ініціалізація лічильника навчальних векторів $t:=0$;

2) задаються два початкові вузли (нейрони) w_a та w_b шляхом вибору випадковим шляхом з навчального набору. Вузли w_a та w_b з'єднуються ребром, вік яких рівний нулю. Ці вузли вважаються нефіксованими;

3) обирається з набору даних наступний вектор x , який приводиться до одиничної довжини (L2-нормалізація);

4) приведення кожного базисного вектору $w_k, k = \overline{1, M}$ до одиничної довжини (L2-нормалізація);

5) обчислення міри схожості вхідного вектору x до базисних векторів $w_{s_k} \in W$ для сортування

$$-(w_{s_0}^T x)^2 \leq \dots \leq -(w_{s_k}^T x)^2 \leq \dots \leq -(w_{s_{M-1}}^T x)^2;$$

6) обирається найближчий вузол w_{s_0} і другий за близькістю вузол w_{s_1} ;

7) вік усіх інцидентних w_{s_0} ребр збільшити на одиницю;

8) якщо w_{s_0} фіксований, то відбувається перехід до кроку 9, інакше до кроку 10;

9) якщо $(w_{s_0}^T x)^2 \geq \nu$, то переходимо до кроку 12. Інакше додаємо новий нефіксований нейрон w_r в точку, що співпадає з вхідним вектором $w_r=x$, а також додається нове ребро, що з'єднує w_r та w_{s_0} , потім перехід до кроку 13;

10) вузол w_{s_0} і його топологічні сусіди (вузли, з'єднані з ним ребром) зміщуються в напрямку до вхідного вектору x згідно з правилом Ойа [14] за формулами

$$\Delta w_{s_0} = \varepsilon_b \eta_t y_0 (x - y_0 w_{s_0}), \quad y_0 := w_{s_0}^T x,$$

$$\Delta w_{s_n} = \varepsilon_n \eta_t y_n (x - y_n w_{s_n}), \quad y_n := w_{s_n}^T x,$$

$$0 < \varepsilon_b \ll 1, \quad 0 < \varepsilon_n \ll \varepsilon_b,$$

$$\eta_t := \eta_0 (\eta_{final} / \eta_0)^{t/t_{max}},$$

де Δw_{s_0} , Δw_{sn} – вектор корекції вагових коефіцієнтів нейрона-переможця та його топологічних сусідів відповідно; ε_b , ε_n – константи сили оновлення вагових коефіцієнтів нейрона-переможця та його топологічних сусідів відповідно; η_0 , η_t , η_{final} – початкове, поточне і кінцеве значення швидкості навчання відповідно;

11) якщо $(w_{s_0}^T x)^2 \geq \nu$, відмічаємо нейрон w_{s_0} як фіксований;

12) якщо w_{s_0} та w_{s_1} з'єднані ребром, то його вік обнуляється, в протилежному випадку між w_{s_0} та w_{s_1} формується нове ребро з нульовим віком;

13) всі ребра в графі з віком більше a_{max} видаляються. У випадку якщо деякі вузли не мають інцидентних ребр (стають ізольованими), ці вузли також видаляються;

14) якщо $t < t_{max}$ переходимо до кроку 15, інакше – інкремент лічильника кроків $t := t + 1$ і перехід до кроку 3;

15) якщо всі нейрони фіксовані, виконання алгоритму зупиняється, інакше перехід до кроку 3 і починається нова епоха навчання (повторення навчальної множини).

На третьому етапі потрібно знайти оптимальні параметри функціонування системи детектування. Особливо важливим алгоритм оптимізації є для процедури тонкої настройки шару екстрації ознак, що адаптується до доменної області, оскільки навчання без вчителя не враховує незбалансованість патчів, що відповідають фону та об'єктам інтересу.

Серед метаевристичних алгоритмів пошукової оптимізації пропонується обрати симуляцію відпалу. Ефективність алгоритму симуляції відпалу залежить від реалізації процедури `create_neighbor_solution`, що здійснює формування нового рішення s_i на i -й ітерації алгоритму. На рис. 2 показано псевдокод алгоритму симуляції відпалу, що виконується `epochs_max` ітерацій, на кожній з яких обчислюється функція $f()$ шляхом пропускання розміченої навчальної вибірки через модель системи детектування і обчислення комплексного критерію (2).

```

scurrent ← create_initial_solution()
sbest ← scurrent
T ← T0
c ← ε, 0 < ε < 1
for(i = 1 to epochs_max)
    si ← create_neighbor_solution(scurrent)
    if f(si) ≥ f(scurrent)
        scurrent ← si
        if f(si) ≥ f(sbest)
            sbest ← si
        end if
    elseif exp( $\frac{f(s_{current}) - f(s_i)}{T}$ ) > uniform_random(0,1)
        scurrent ← si
    end if
    T ← c × T
end for
return(sbest)

```

Рис. 2. Псевдокод метаевристичного алгоритму симуляції відпалу

Аналіз псевдокоду рис. 2 показує, що поточне рішення $s_{current}$, відносно якого шукають нові найкращі рішення s_{best} , оновлюється у випадку забезпечення новим рішенням зростання критерію (2), або випадковим шляхом із розподілу Гіббса [16]. При цьому початковою точкою пошуку, що формується за процедурою *create_initial_solution*, може бути або випадково згенероване, або результатом попереднього навчання за іншим алгоритмом. Для генерації нових рішень у процедурі *create_neighbor_solution* пропонується використовувати найпростіший неадаптивний алгоритм, який можна подати у вигляді формули:

$$s_{current} = s_{current} + \text{uniform_random}(-1,1) \cdot \text{step_size},$$

де *uniform_random* – функція генерації випадкового числа з рівномірного розподілу із заданого діапазону; *step_size* – розмір околу пошуку нових рішень, сусідніх до $s_{current}$.

Таким чином, запропонований алгоритм навчання детектора полягає в адаптації верхніх шарів моделі до доменної області використання на основі алгоритмів зростаючого розріджено кодуючого нейронного газу та симуляції відпалу.

5. Результати машинного навчання системи детектування об'єктів на зображеннях аеровідео моніторингу

Для навчання детектора об'єктів в полі зору безпілотної апарату було використано біля 180 зображень з набору даних Inria Aerial Image Labeling Dataset [18]. Кожне зображення має роздільну здатність 5000×5000 пікселів. Було згенеровано 2 000 нерозмічених зображень розміру 224×224 пікселів шляхом випадкового вирізання з обертанням, а також 200 розмічених зображень розміру 224×224 пікселів, які шляхом накладання шуму, зміни контрасту та обертання було розмножено до 800 розмічених зображень.

На зображеннях набору даних Inria Aerial Image Labeling Dataset у великій кількості представлено транспортні засоби у міській місцевості. Тому пропонується обрати транспортні засоби за об'єкти інтересу, а міську місцевість вважати доменною областю використання. При цьому алфавіт класів розпізнавання рівний $R=3$, де перший клас відповідає легковим автомобілям, другий відповідає грузовим автомобілям, а третій – фону. Розмір об'єктів в пікселях на вибіркових зображеннях варіюється в діапазоні $[16 \times 16, \dots, 32 \times 32]$.

Пропонується здійснити перенос знань шляхом копіювання шарів із попередньо навченої згорткової мережі SqueezeNet, яка має високу обчислювальну ефективність і популярна в мобільних системах з обмеженими ресурсами [5, 6]. В результаті кожне вхідне зображення кодується в карту ознак на виході модуля `fire7` мережі SqueezeNet. В цьому випадку розмір карти ознак становить $13 \times 13 \times 384$. Наступний шар попередньо навчається без вчителя на нерозмічених зразках доменної області і складається з фільтрів з ядром 1×1 та 3×3 з кроком сканування $\text{stride}=1$. Наступний шар призначений для збільшення масштабу карти ознак з мінімальними втратами інформації і має структуру, показану на рис. 1, а. Останній шар так само попередньо навчається без вчителя. Корекція навчених без вчителя шарів та тренування класифікаторів здійснюється на основі розмічених навчальних зразків.

Навчання детектора спочатку пропонується здійснювати з використання екстрактора ознак, навченого без вчителя з використанням зростаючого розрізжено кодуємого нейронного газу без тонкої настройки мережі. При цьому під час навчання використовується фіксоване значенням параметра реконструкції навчальної множини $v=0.8$. Крім того, під час розрахунку активації пікселів карти ознак не застосовується алгоритм ОУП. В інформаційно-екстремальному класифікаторі пікселів карти ознак кількість порогів для дискретизації ознакового опису рівна $L=3$. В алгоритмі симуляції відпалу для навчання класифікаторів використано такі параметри: $c=0.98$, $T_0=10$, $\text{epochs_max}=500$, $\text{step_size}=0.001$. В табл. 1 показано результати машинного навчання детектора об'єктів при різних значеннях порогу Th на ступінь перетину проекції пікселя карти ознак з прямокутником розмітки навчального зображення для включення пікселя до навчальної вибірки об'єктів інтересу.

Аналіз табл. 1 показує, що оптимальне значення порогу рівне $Th^*=0.4$, яке забезпечує виявлення 69 % об'єктів на тестових зображеннях. При цьому малі значення порогу Th призводять до зменшення ефективності детектування внаслідок зростання хибних спрацювань. Великі значення порогу Th

приводять до зменшення кількості навчальних зразків кожного класу і ігнорування об'єктів, що знаходяться на межі декількох проекцій пікселів карти ознак.

З метою покращення результатів машинного навчання детектора здійснюється підвищення інформативності ознакового опису шляхом тонкої настройки навчених без вчителя згорткових шарів. При цьому для високорівневих шарів формування ознакового опису розглядається випадок як відсутності так і присутності алгоритму ОУП з функцією ReLU для обчислення активації кожного пікселя карти ознак. У табл. 2 показано результати машинного навчання при використанні тонкої настройки шарів екстракції на основі симуляції відпалу.

Таблиця 1

Результати машинного навчання детектора при різних значеннях порогу перекриття проекції пікселя з прямокутником об'єкта на розміченому зображенні для віднесення пікселя до зразка об'єкта

Th	J_{Cls}	J_{Loc}	J	Частка виявлених об'єктів на тестовій вибірці
0.1	0.210	0.090	0.0200	0.610
0.2	0.250	0.120	0.0300	0.630
0.3	0.250	0.150	0.0360	0.660
0.4	0.301	0.150	0.0450	0.690
0.5	0.200	0.150	0.0300	0.630
0.6	0.130	0.170	0.0220	0.630
0.7	0.090	0.100	0.0090	0.500
0.8	0.087	0.100	0.0087	0.420
0.9	0.050	0.100	0.0050	0.330

Таблиця 2

Результати машинного навчання детектора з тонкою настройкою екстрактора ознак на основі симуляції відпалу

Використання ОУП	J_{Cls}	J_{Loc}	J	Частка виявлених об'єктів на тестовій вибірці
–	0.41	0.360	0.1476	0.91
+	0.52	0.450	0.2340	0.95

Аналіз табл. 2 показує, що тонка настройка екстрактора ознак призвела до зростання точності виявлення об'єктів на зображенні місцевості. При цьому використання алгоритму ОМР для обчислення активації пікселів карти знак має також помітний вплив, оскільки додає 4% точності виявлення об'єктів на тестових зображеннях.

Таким чином, розроблений алгоритм навчання вирішальних правил детектора об'єктів дозволяє отримати прийнятні для практичного використання у доменній області результати без необхідності навчання моделі з нуля. При

цьому тонка настройка високорівневих шарів на основі алгоритму симуляції відпалу дозволяє покращити результати навчання детектора.

6. Обговорення результатів машинного навчання детектора об'єктів

Аналіз результатів машинного навчання показує, що попереднє навчання без вчителя на нерозмічених зразках, як правило, спрощує подальший процес машинного навчання з учителем. Тому варто розглянути вплив параметрів алгоритму зростаючого розріджено кодуючого нейронного газу, що використовується для попереднього навчання без вчителя на результати машинного навчання з учителем. У табл. 3 показано залежність результатів машинного навчання і кількості N_c згенерованих згорткових фільтрів (нейронів) від значення параметру ν , що характеризує точність покриття навчальної вибірки згортковими фільтрами.

Таблиця 3

Результати машинного навчання детектора при різних значеннях параметру покриття навчальних даних при навчанні без вчителя

ν	N_c	J_{Cls}	J_{Loc}	J	Частка виявлених об'єктів на тестовій вибірці
0.4	196	0.0912	0.100	0.00912	0.67
0.5	337	0.1502	0.101	0.01517	0.73
0.6	589	0.2508	0.401	0.10057	0.87
0.7	889	0.4203	0.430	0.18073	0.93
0.8	1519	0.5201	0.450	0.23401	0.95
0.9	4058	0.5203	0.450	0.23416	0.92

Аналіз табл. 3 показує, що зі зростанням значення параметру ν зростає кількість нейронів і значення часткових і загального критерію оптимізації (2). Однак при $\nu \leq 0.8$ точність моделі за тестовою вибіркою зростає зі зростанням параметра ν , а подальше зростання параметру призводить до погіршення результатів внаслідок ефекту перенавчання. Тому параметр ν можна вважати параметром регуляризації моделі верхніх шарів екстрактора ознак.

Щоб зрозуміти переваги використання попереднього навчання без вчителя, розглянемо результати машинного навчання з використанням симуляції відпалу до і після навчання за алгоритмом зростаючого розріджено кодуючого нейронного газу. На рис. 3 показано залежність критерію ефективності навчання детектора (2) за алгоритмом симуляції відпалу з параметрами $c=0.998$, $T_0=10$, $epochs_max=5000$, $step_size=0.001$.

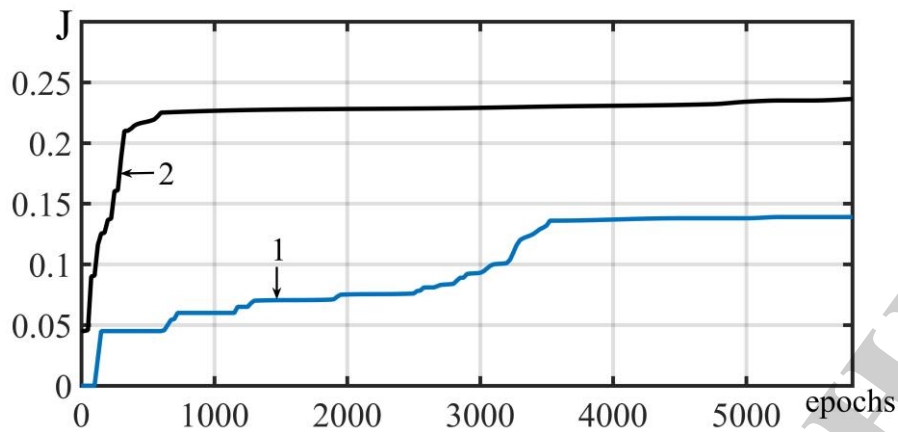


Рис. 3 Графіки залежності критерію оптимізації (2) від кількості епох навчання : 1 – до застосування попереднього навчання без вчителя; 2 – після застосування попереднього навчання без вчителя

Аналіз рис. 3 показує, що попереднє навчання без вчителя на основі зростаючого розріджено кодуючого нейронного газу дозволяє покращити кінцевий результат навчання з учителем за допомогою алгоритму симуляції відпалу. При цьому, якщо попередньо відбувалося навчання без учителя, то досягнення глобального максимуму критерію (2) відбувається більше ніж в 10 разів швидше. Крім того, перевірка отриманих моделей за тестовою вибіркою показує, що використання попереднього навчання без учителя дозволяє зменшити ефект перенавчання в умовах обмеженого обсягу розмічених навчальних даних. При використанні попереднього навчання без вчителя критерій (2) розрахований для навчальної вибірки рівний $J_{train}=0.234$, а для тестової вибірки – $J_{test}=0.228$. Без використання попереднього навчання відповідні критерії значно відрізняються між собою, $J_{train}=0.1321$ і $J_{test}=0.081$.

Таким чином, запропонований алгоритм попереднього навчання високорівневих шарів без вчителя дозволяє підвищити значення критерію навчання і частки виявлених об'єктів на тестових зображеннях. Крім того вдалося зменшити ефект перенавчання і підвищити швидкість знаходження глобального максимуму при навчанні з учителем на розміченій навчальній вибірці обмеженого обсягу. Однак в рамках даного дослідження не було розглянуто механізми уточнення меж детектованих об'єктів і не розглянуто залежність критерію ефективності навчання від параметрів алгоритму симуляції відпалу. Тому наступні дослідження будуть спрямовані на удосконалення моделі детектора та розробку алгоритмів настройки параметрів пошукового алгоритму оптимізації в процесі машинного навчання.

7. Висновки

1. Запропоновано модель детектора об'єктів і критерій ефективності навчання моделі. Модель містить 7 перших модулів обчислювально ефективною згорткової мережі SqueezeNet, два згорткові різномасштабні шари, та інформаційно-екстремальний класифікатор. Як критерій ефективності навчання

моделі детектора розглядається мультиплікативна згортка частинних критеріїв, що враховує ефективність виявлення об'єктів на зображенні та точність класифікаційного аналізу. При цьому додаткове використання алгоритму ортогонального узгодженого кодування при обчисленні високорівневих ознакових карт дозволяє збільшити точність моделі на 4%. Дана модель забезпечує прийнятну для практичного використання достовірність виявлення об'єктів, розмір яких у пікселях в 7...14 разів менший за розмір найменшої сторони зображення аерофотознімку.

2. Запропоновано трьохетапний алгоритм навчання детектора об'єктів за умов малого обсягу розмічених навчальних зразків та обмежених обчислювальних ресурсів, доступних на борту малогабаритного безпілотного літального апарату. На першому етапі пропонується обрати кількість нижніх згорткових шарів глибокої нейронної мережі, що попередньо навчається на наборі даних ImageNet. Другий етап полягає в навчанні без вчителя верхніх шарів нейронної мережі на нерозмічених зразках, що кодуються нижніми згортковими шарами. При цьому розроблено алгоритм зростаючого розрідженого кодуючого нейронного газу для навчання згорткових фільтрів без вчителя. Його використання дозволяє утилізувати нерозмічені навчальні дані для адаптації викорівневого ознакового опису до доменної області використання та визначити необхідну кількість нейронів. На третьому етапі здійснюється навчання інформаційно-екстремального класифікатора пікселів високорівневих карт ознак та тонка настройка верхніх згорткових шарів шляхом використання метаевристичного алгоритму симуляції відпалу. Доведено за результатами фізичного моделювання на наборі даних Inria Aerial Image Labeling ефективність розробленого алгоритму навчання без вчителя згорткових фільтрів. При цьому за відсутності тонкої настройки згорткових фільтрів забезпечується 69% виявлених об'єктів на зображеннях тестової вибірки. Після тонкої настройки високорівневих згорткових фільтрів на основі алгоритму симуляції відпалу забезпечується 95% виявлених об'єктів на тестових зображеннях.

3. Показано, що використання попереднього навчання без вчителя дозволяє зменшити ефект перенавчання і підвищити більше ніж в 10 разів швидкість знаходження глобального максимуму при навчанні з учителем на вибірці обмеженого обсягу. При цьому на тестовій вибірці вдалося практично усунути ефект перенавчання шляхом оптимального вибору значення гіперпараметра, що характеризує ступінь покриття навчальних даних нейронами мережі.

Подяки

Робота виконана на базі лабораторії інтелектуальних систем кафедри комп'ютерних наук Сумського державного університету при фінансовій підтримці МОН України в рамках держбюджетної науково-дослідної роботи ДР № 0117U003934.

Література

1. Patricia N., Caputo B. Learning to Learn, from Transfer Learning to Domain Adaptation: A Unifying Perspective // 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. doi: <https://doi.org/10.1109/cvpr.2014.187>
2. Nguyen A., Yosinski J., Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. doi: <https://doi.org/10.1109/cvpr.2015.7298640>
3. Optimization of convolutional neural network using microcanonical annealing algorithm / Ayumi V., Rere L. M. R., Fanany M. I., Arymurthy A. M. // 2016 International Conference on Advanced Computer Science and Information Systems (ICACISIS). 2016. doi: <https://doi.org/10.1109/icacsis.2016.7872787>
4. Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition / Antipov G., Berrani S.-A., Ruchaud N., Dugelay J.-L. // Proceedings of the 23rd ACM international conference on Multimedia – MM '15. 2015. doi: <https://doi.org/10.1145/2733373.2806332>
5. A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles / Carrio A., Sampedro C., Rodriguez-Ramos A., Campoy P. // Journal of Sensors. 2017. Vol. 2017. P. 1–13. doi: <https://doi.org/10.1155/2017/3296874>
6. Scaling for edge inference of deep neural networks / Xu X., Ding Y., Hu S. X., Niemier M., Cong J., Hu Y., Shi Y. // Nature Electronics. 2018. Vol. 1, Issue 4. P. 216–222. doi: <https://doi.org/10.1038/s41928-018-0059-3>
7. DroNet: Learning to Fly by Driving / Loquercio A., Maqueda A. I., del-Blanco C. R., Scaramuzza D. // IEEE Robotics and Automation Letters. 2018. Vol. 3, Issue 2. P. 1088–1095. doi: <https://doi.org/10.1109/lra.2018.2795643>
8. An Improved Transfer learning Approach for Intrusion Detection / Mathew A., Mathew J., Govind M., Mooppan A. // Procedia Computer Science. 2017. Vol. 115. P. 251–257. doi: <https://doi.org/10.1016/j.procs.2017.09.132>
9. Qassim H., Verma A., Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition // 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). 2018. doi: <https://doi.org/10.1109/ccwc.2018.8301729>
10. Nakahara H., Yonekawa H., Sato S. An object detector based on multiscale sliding window search using a fully pipelined binarized CNN on an FPGA // 2017 International Conference on Field Programmable Technology (ICFPT). 2017. doi: <https://doi.org/10.1109/fpt.2017.8280135>
11. Development of the method of features learning and training decision rules for the prediction of violation of service level agreement in a cloud-based environment / Moskalenko V., Moskalenko A., Pimonenko S., Korobov A. // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 5, Issue 2 (89). P. 26–33. doi: <https://doi.org/10.15587/1729-4061.2017.110073>
12. Feng Q., Chen C. L. P., Chen L. Compressed auto-encoder building block for deep learning network // 2016 3rd International Conference on Informative

and Cybernetics for Computational Social Systems (ICCSS). 2016. doi: <https://doi.org/10.1109/iccss.2016.7586437>

13. Aircraft Detection by Deep Convolutional Neural Networks / Chen X., Xiang S., Liu C.-L., Pan C.-H. // *IPSJ Transactions on Computer Vision and Applications*. 2014. Vol. 7. P. 10–17. doi: <https://doi.org/10.2197/ipsjtcva.7.10>

14. Labusch K., Barth E., Martinetz T. Sparse Coding Neural Gas: Learning of overcomplete data representations // *Neurocomputing*. 2009. Vol. 72, Issue 7-9. P. 1547–1555. doi: <https://doi.org/10.1016/j.neucom.2008.11.027>

15. Mrazova I., Kukacka M. Image Classification with Growing Neural Networks // *International Journal of Computer Theory and Engineering*. 2013. P. 422–427. doi: <https://doi.org/10.7763/ijcte.2013.v5.722>

16. Palomo E. J., Lopez-Rubio E. The Growing Hierarchical Neural Gas Self-Organizing Neural Network // *IEEE Transactions on Neural Networks and Learning Systems*. 2016. P. 1–10. doi: <https://doi.org/10.1109/tnnls.2016.2570124>

17. Rere L. M. R., Fanany M. I., Arymurthy A. M. Metaheuristic Algorithms for Convolution Neural Network // *Computational Intelligence and Neuroscience*. 2016. Vol. 2016. P. 1–13. doi: <https://doi.org/10.1155/2016/1537325>

18. High-Resolution Aerial Image Labeling With Convolutional Neural Networks / Maggiori E., Tarabalka Y., Charpiat G., Alliez P. // *IEEE Transactions on Geoscience and Remote Sensing*. 2017. Vol. 55, Issue 12. P. 7092–7103. doi: <https://doi.org/10.1109/tgrs.2017.2740362>

Не является перепечаткой