

*Сергій Валентинович Ковбасюк (доктор технічних наук, с. н. с.)*

*Леонід Броніславович Каневський (кандидат технічних наук)*

*Микола Петрович Романчук*

*Житомирський військовий інститут імені С.П.Корольова, Житомир, Україна*

## МОДЕЛЬ КАСКАДУ ГІБРИДНОЇ СЕГМЕНТАЦІЇ ДЛЯ АВТОМАТИЧНОГО ДЕШИФРУВАННЯ ОБ'ЄКТІВ НА АЕРОЗНІМКАХ

Актуальним завданням на сьогодні є пошук та удосконалення методів автоматичного детального дешифрування об'єктів на аерознімках, отриманих з безпілотних авіаційних комплексів, які б забезпечували достатню точність виявлення та розпізнавання дрібнозернистих об'єктів в складних топографічних умовах місцевості над якими отримуються аерознімки.

З метою вирішення даного завдання в статті проведений аналіз методів автоматичної обробки зображень та моделей нейронних мереж, побудованих на їх основі. З проведеного аналізу виділено багатостадійний конвеєр обробки аерознімків, що об'єднує підходи виявлення, елементної сегментації та семантичну сегментацію для отримання контексту. Удосконалена моделі каскаду сегментації враховує геометричні розміри об'єктів та їх співвідношення, зміну масштабу, умови знімання. Використання моделі каскаду сегментації для автоматичного дешифрування об'єктів на аерознімках дозволить підвищити точність виявлення та розпізнавання таких об'єктів.

**Ключові слова:** нейромережі, сегментація, аерознімок, безпілотні авіаційні комплекси, каскад гібридної сегментації.

### Вступ

Досвід застосування збройних сил передових країн світу засвідчує, що головні зусилля в їх розвитку спрямовані на досягнення інформаційної переваги над противником шляхом удосконалення технічних видів розвідки, одним з яких є безпілотні авіаційні комплекси (БпАК).

**Постановка проблеми.** В системі інформаційного забезпечення військ (сил) при застосуванні БпАК найважливішим етапом, від якого залежить оперативність і достовірність прийняття рішення командирами відповідного рівня, є процес дешифрування об'єктів на аерознімках. При виконанні автоматизації етапу детального дешифрування потрібно вирішити завдання виявлення, призначення точних обмежувальних рамок або масок для дрібнозернистих об'єктів та розпізнавання, відділення об'єктів переднього плану від фону і призначення їм відповідних міток класу об'єктів. Для вирішення такого роду завдань на сучасному етапі розвитку інформаційних технологій широкого застосування набувають методи семантичної і елементної сегментації, розвиток яких на сьогодні відбувається паралельно один від одного, які мають свої особливості, переваги та недоліки.

Актуальним завданням на сьогодні є пошук та удосконалення методів автоматичного дешифрування об'єктів на аерознімках, отриманих з БпАК, які б забезпечували достатню точність виявлення та розпізнавання об'єктів в складних топографічних умовах місцевості, над якими отримуються аерознімки.

Вирішення цього наукового завдання можливе в комплексуванні різних підходів автоматичної обробки

аерознімків шляхом створення моделі каскаду гібридної сегментації, що і є метою дослідження.

**Аналіз останніх досліджень і публікацій.** Методи семантичної сегментації, які використовують згорткові нейромережі для отримання ознак, в основному підходять до вирішення завдань виявлення та розпізнавання з багаторівневої їх агрегації [1] або з наскрізного структурованого прогнозування [2]. Поліпшення в отриманні контексту, яке має вирішальне значення для семантичної сегментації, виникло через використання розширених згорткових нейронних мереж (далі мереж) [3], що дозволяє зберігати високу роздільну здатність до останнього шару. Мережа розбору пірамідальних сцен (англ. PSPNet) [1], яка використовує модуль об'єднання пірамід (англ. PPM) для об'єднання фонових об'єктів на основі різних регіонів в своїй магістралі. Багатозадачна модель [4], що використовує поєднання мережі піраміди ознак (англ. FPN) [5] та PPM, демонструє універсальність сегментації по отриманню візуальної інформації з зображення.

Елементна сегментація стосується ідентифікації конкретного об'єкта семантичного класу, з яким пов'язаний піксель аерознімку. Починаючи з регіональної згорткової мережі (англ. R-CNN) [6], підходи глибокого навчання проводять сегментацію екземплярів як двоступінчастий підхід, в якому зроблено ряд сегментних пропозицій. Потім відбувається порівняння між цими пропозиціями щодо вибору найкращої [7, 8]. Загальним для сімейства цих методів є здійснення сегментації мережею пропозицій регіонів (англ.

RPN) до проведення класифікації об'єктів, що робить реалізацію повільною. В InstanceFCN [9] пропозиції масок елементів [10] отримуються з повністю згорткової мережі (англ. FCN) [11]. Модель Mask R-CNN [12] представляє спільний підхід до формування обмежувальних рамок та масок, коли кожна з двох паралельних цільових функцій присвячена окремому завданню.

Класифікатори об'єктів фону, що використовують методи семантичної сегментації, зазвичай побудовані на FCN з розширеннями [3, 13], не передбачають межі екземплярів для класів. Методи елементної сегментації побудовані на базі детекторів, що в основному використовують об'єктні пропозиції [10] на основі запропонованих областей [11, 14], ігнорують об'єкти фону. Об'єднання їх дозволить вирішити завдання аналізу сцени [15], розбору зображень [16] або цілісного розуміння сцени [17]. Були започатковані дослідження як явно моделювати відносини між об'єктами фону та класами елементів щодо можливостей використання умовних випадкових полів над сегментами зображень для моделювання геометричних і семантичних відносин [18], включення сегментації унарних потенціалів та об'єктів у цілісну структуровану втрату [19].

Для підвищення точності виявлення дрібнозернистих об'єктів двоступеневі детектори [6, 12, 14] показали свої переваги в порівнянні з одноступеневими [20, 21], завдяки їх оптимізації та можливості генерувати достатньо високі функції для виявлення об'єктів при обробці аерознімків. Так, багаторегіональна згорткова мережа [22] використовує ітеративний механізм виявлення для уточнення їх розташування. Одним з підходів підвищення точності виявлення та розпізнавання об'єктів є використання каскадної структури для виключення легких негативних зразків. Модель каскадної R-CNN [23] складається з декількох етапів, де дані тренувань з попереднього етапу подаються в наступний із збільшенням порогових значень метрики IoU, для підвищення якості обробки. Тому створення багатостадійного конвеєра обробки аерознімків, що використовує об'єднання виявлення, елементної сегментації та семантичну сегментацію для отримання контексту, дозволить підвищити точність виявлення та розпізнавання об'єктів.

**Метою статті** є аналіз моделей нейронних мереж та їх удосконалення як інструменту для підвищення точності виявлення та розпізнавання об'єктів з врахуванням топографічних елементів місцевості на аерознімках, отриманих з БпАК.

### Виклад основного матеріалу дослідження

Завдання аналізу сцени, розбору зображень, цілісного розуміння сцени стикаються з складнощами, оскільки візуальні об'єкти часто піддаються деформації, оклюзії, змінам масштабу та зміна об'єктів фонів знижує точність виявлення об'єктів. Тому потрібна модель, яка буде одночасно стійка до змін зовнішнього вигляду об'єктів і використовуватиме контекстну

інформацію для їх розрізнення від переповненого фону. Комплексним рішенням може бути застосування моделей елементної і семантичної сегментації, що мають спільну основу, використовуючи глибоку магістральну мережу, яка генерує достатні представлення ознак, на вершині якої розміщені відповідні спеціалізовані цільові функції.

Як основа використовується згорткова мережа ResNeXt [24] в FPN з спеціалізованими цільовими функціями виявлення обмежувальних рамок та елементної сегментації. Як згорткова мережа ResNeXt, що представляє високомодульну мережеву архітектуру, яка має велике рецептивне поле завдяки агресивному згортванню. Це відбувається за рахунок просторової розрізненості та здатності точно локалізувати малі об'єкти. Використання FPN дає змогу захоплювати ознаки з низького рівня магістральної мережі, щоб розпізнати більш широкий діапазон масштабів об'єктів з набагато меншою кількістю параметрів, ніж розширені згорткові мережі. Це вирішує проблему апаратних обмежень, яке зазвичай існує як для семантичної або елементної сегментації, так і при їх спільному навчанні. На вершині FPN застосовується деформована згорткова мережа (англ. DCN) [25] для адаптування до геометричних варіацій об'єкта аерознімка, використовуючи фактор, що не всі пікселі всередині рецептивного поля шару мережі вносять внесок у відповідь. Відмінності в цих вкладах представлені ефективним рецептивним полем, значення якого обчислюються як градієнт відповіді вузла шару по відношенню до інтенсивності збурень кожного пікселя зображення. Впровадження DCN дає можливість адаптувати своє уявлення цільової функції до конфігурації об'єкта, деформуючи його структури вибірки та об'єднання, які відповідають структурі об'єкта. При такому підході досягаються підвищення точності виявлення об'єктів на аерознімку. Для регресії виявлення об'єктів використовуються регіональні згорткові мережі поверх FPN, аналогічно Mask R-CNN.

Використання взаємного зв'язку між виявленням обмежувальних рамок та елементною сегментацією приносить обмежений вигоду, тому більш ефективне рішення полягає в каскадному їх застосуванні для покращення локалізації виявлених об'єктів та їх розпізнавання. Каскадна процедура застосовується при висновках кожного етапу, що дозволяє більш точно узгоджувати гіпотези. Застосування каскаду дозволяє зменшити перенавчання під час тренування мережі, в наслідок експоненціально зникаючих позитивних зразків та невідповідності висновку етапу між IoU, для яких детектор є оптимальним, вхідним гіпотезам. Але існує розрив в інформаційному потоці між гілками масок різних етапів каскаду. Це призводить до відділення маски на більш пізніх етапах, що приносить вигоду тільки в краще локалізованих обмежувальних рамках [23].

Для подолання розриву між етапами застосовується каскад гібридних завдань для елементної сегментації [26]. Ключовою ідеєю є

поліпшення інформаційного потоку шляхом включення каскаду та багатозадачності на кожному його етапі та використання просторового контексту для подальшого підвищення точності виявлення та розпізнавання об'єктів. В результаті досліджень була вдосконалена модель каскаду гібридної сегментації, структурна схема моделі якого зображена на рис. 1, де:  $I$  – вхідне зображення, Pool – регіональне вилучення ознак,

$B_t$ ,  $M_t$  – виявлення обмежувальної рамки та маски на етапі  $t$ . Нова структура дозволяє підвищити продуктивність проведенням багатоетапної обробки аерознімка, розрізнити різноплановий передній план від переповненого фону завдяки просторовому контексту, що надає використання семантичної сегментації.

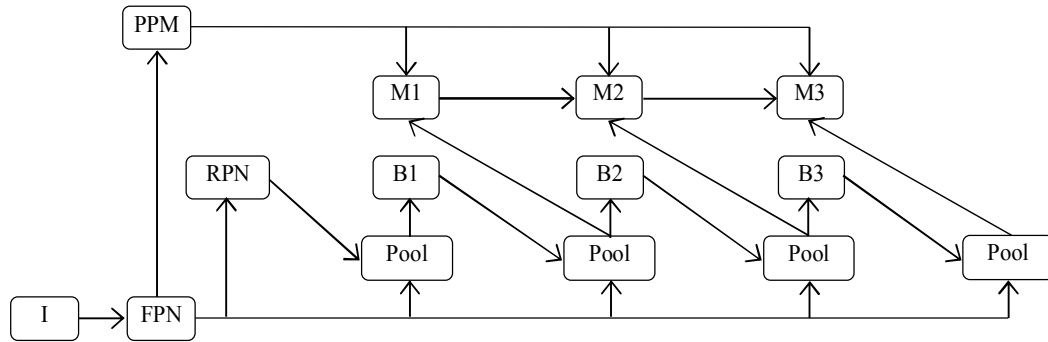


Рис. 1. Модель каскаду гібридної сегментації

Для виявлення об'єктів контекст сцени надає корисні поради при об'єднанні семантичних гілок для виведення категорій, масштабів. Отримані з виходу кожного шару FPN карти ознак різних рівнів перетворюються в модулі пірамідального злиття PPM, яка виконує семантичну сегментацію об'єктів фону на рівні пікселів, що запобігає втраті інформації між різними субрегіонами сцени. Семантична гілка каскаду кодує контекстну інформацію з фонових областей, що доповнює обмежувальні поля та маски екземплярів. Ця гілка безпосередньо сприяє прогнозуванню обмежувальних рамок і масок з кодованими просторовими контекстами в результаті розрізнення об'єктів переднього плану від фонових перешкод. Запропонований підхід у порівнянні з існуючими каскадними рішеннями відрізняється чергуванням регресії обмежувальної рамки і передбаченням маски замість виконання їх обробки паралельно, включенням в себе прямого шляху для посилення інформаційного потоку між гілками маски подачею на маску особливостей попередньої стадії, спрямуванням на вивчення більш контекстної інформації додаткової гілки семантичної сегментації та злиття її з гілками обмежувальних рамок та масок (рис. 1). Використання послідовності детекторів об'єктів, які пройшли навчання з підвищенням порогових значень метрики IoU, більш селективне проти близьких помилкових спрацьовувань. Детектори тренуються поетапно, використовуючи спостереження, що вихід попереднього детектора є покращеним розподілом для навчання наступного більш високої якості. Порядок проходження інформації між етапами каскаду під час тренування відображається виразами:

$$x_t^{\text{box}} = P(x, r_{t-1}) + P(S(x), r_{t-1}), \quad (1)$$

$$x_t^{\text{mask}} = P(x, r_t) + P(S(x), r_t), \quad (2)$$

$$r_t = B_t(x_t^{\text{box}}), \quad (3)$$

$$m_t = M_t(F(x_t^{\text{mask}}, m_{t-1}^-)), \quad (4)$$

де  $x_t^{\text{box}}$ ,  $x_t^{\text{mask}}$  – виявлені обмежувальними рамками та масками ознаки;  $P(x, r_{t-1})$  – операція злиття RoI Align [12];  $B_t(x_t^{\text{box}})$ ,  $M_t(x_t^{\text{mask}})$  – визначення обмежувальної рамки та маски на етапі  $t$ ;  $r_t$ ,  $m_t$  – передбачення обмежувальних рамок та масок ознак.

Тренування запропонованого каскаду: регресія обмежувальної рамки, прогнозування класів і маски проводиться в режимі від початку до кінця. Загальна функція втрат приймає форму багатозадачного навчання на кожній ітерації:

$$L = \sum_{t=1}^T \alpha_t (L_{\text{bbox}}^t + L_{\text{mask}}^t) + L_{\text{seg}}, \quad (5)$$

$$L_{\text{bbox}}^t(c_t, r_t, \hat{c}_t, \hat{r}_t) = L_{\text{cls}}(c_t, \hat{c}_t) + L_{\text{reg}}(r_t, \hat{r}_t), \quad (6)$$

$$L_{\text{mask}}^t(m_t, \hat{m}_t) = \text{BCE}(m_t, \hat{m}_t), \quad (7)$$

$$L_{\text{seg}} = \text{CE}(s, \hat{s}), \quad (8)$$

де  $L$  – загальна функція втрат;  $L_{\text{bbox}}^t$ ,  $L_{\text{mask}}^t$  – втрата прогнозування обмежувальної рамки та маски на етапі  $t$ ;  $L_{\text{cls}}$ ,  $L_{\text{reg}}$  – втрата прогнозування класифікації та регуляризації зображення об'єкта;  $L_{\text{segm}}$  – втрата прогнозування семантичної сегментації;  $\text{CE}$  – функція втрати перехресної ентропії;  $\text{BCE}$  – функція бінарної втрати перехресної ентропії.

В удосконаленій моделі для подолання великомасштабних відхилень розділяють пропозиції, отримані від першого кроку, на декілька груп відповідно до їх масштабу та рівномірної вибірки. Для кожної групи відбирається 2000 пропозицій. Так, для передачі районів, що викликали зацікавлення, на вхід RPN, де застосовують згортковий фільтр розмірністю  $7 \times 7$ , застосовують якорі розмірами на 4, 8, 16, 32,

64, які сприяють їх кращому перетину, а також для врахування співвідношення сторін об'єктів застосовано таке значення співвідношення якорів [0,33; 0,5; 1,0; 2,0; 3,0] для врахування лінійних розмірів об'єктів та їх співвідношення. За магістральну архітектуру моделі було взято згорткову мережу ResNeXt-101, що отримує ознаки зі згорткового шару 1-го етапу, який називається С1. Для використання методу передачі навчання було застосовано шар С4, ваги тільки якого змінювалися. На останньому етапі магістралі застосовується DCN. Для об'єднання всіх виявлених об'єктів з метою виключення повторного їх виявлення було задіяно алгоритм постобробки м'якого немаксимального придушення зі значенням порогу 0,7. При роботі моделі для покращення якості роботи застосовується триразове збільшення часу тестування для пред- та постобробки знімку (незначну нарізку зображення з розрізненням 600 x 650, 700 x 750 та поворот (0°, 90°, 180°, 270°), зі збільшенням від 800 x 850 до 1000 x 1000).

При створенні навчаючих вибірок для кожного класу об'єктів за їх зображеннями для нового набору даних із аерознімків виникає дисбаланс класів через відсутність однакової достатньої кількості зображень об'єктів. При використанні функції втрати перехресної ентропії при тренуванні моделей на таких наборах даних коефіцієнт масштабування зменшується до нуля, оскільки підвищується впевненість у правильному класі. Для протидії цій проблемі використовувалися різні методи, такі як повторна вибірка та техніка вибіркової вибірки синтетичних меншин. За результатами проведених досліджень у даному рішенні пропонується модифікувати метод фокусної втрати, який призначений для поліпшення тренування моделей на оригінальних незбалансованих даних. Так замість функції втрати перехресної ентропії

$$CE(p_i) = -\log(p_i), \quad (9)$$

досить часто застосовується функція сфокусованої втрати [27], що має такий вигляд

$$FL(p_i) = -(1-p_i)^\gamma \log(p_i), \quad (10)$$

де FL – сфокусована втрата, CE – функція втрати перехресної ентропії,  $p_i$  – ймовірність достовірного класу,  $\gamma$  – параметр фокусування.

Фокусна втрата зводить до мінімуму внесок добре класифікованих зразків об'єктів і спрямовує фокус на складні зразки. Функція сфокусованої втрати розроблена для вирішення сценарію детермінованого виявлення об'єктів, в якому існує надзвичайний дисбаланс між наповненими і розрідженими класами. Але вона не показує кращих результатів для двохісних детекторів, в яких фон відокремлюється на першому етапі. Пропонується модифікувати функцію фокусної втрати для послаблення реакції функції втрат на складні зразки. Відповідно застосовуються однакові ваги до зразків об'єктів з ймовірностями,

### Література

1. **Zhao H.** Pyramid scene parsing network [Електронний ресурс] / H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia // Режим доступу : <https://arxiv.org/abs/1612.01105.pdf>. 2.

меншими за певний поріг, а також для мінімізації впливу добре класифікованих зразків зберігається підхід фокусування втрат, масштаб якого відображає поріг. Вищезазначене можна описати виразом:

$$MFL(p_i) = -f(p_i, t_h) \log(p_i), \quad (11)$$

де  $f(p_i, t_h)$  – коефіцієнт відсікання, який масштабує функцію втрат за такою формулою:

$$f(x) = \begin{cases} 1 & : p_i < t_h \\ \frac{(1-p_i)^\gamma}{t_h^\gamma} & : p_i \geq t_h \end{cases}, \quad (12)$$

де  $t_h$  – ймовірність класу основної істини.

Модифікація фокусної втрати допомагає покращити середню точність виявлення об'єктів mAP для розріджених класів, використовуючи модель, підготовлену наскрізь, хоча вона незначно зменшує mAP для наповнених класів. Застосування модифікації фокусної втрати зменшує вплив фактору дисбалансу класів в процесі тренування моделі.

### Висновки і перспективи подальших досліджень

З метою підвищення оперативності та достовірності детального дешифрування аерознімків був проведений аналіз методів автоматичної обробки зображень. В результаті якого при дослідженні моделей нейромереж для вирішення завдання аналізу сцени, розбору зображень виявлено складнощі, які знижують точність виявлення та розпізнавання об'єктів. Вони виникають через деформації, оклюзії та масштабні, перспективні перетворення об'єктів під час отримання знімку та через часту зміну об'єктів фону.

З результатів досліджень виділено багатостадійний конвеєр обробки аерознімків, що об'єднує підходи виявлення, елементної сегментації та семантичну сегментацію для отримання контексту. Він здійснює поліпшення інформаційного потоку шляхом включення каскаду та багатозадачності на кожному його етапі та використовує просторовий контекст для підвищення точності виявлення та розпізнавання об'єктів.

Удосконалена модель каскаду сегментації враховує лінійні розміри шуканих об'єктів та їх співвідношення, можливу зміну масштабу, вплив топографічних елементів місцевості. Дана модель має стати основою для автоматичного детального дешифрування об'єктів на аерознімках, отриманих з БпАК, та підвищити точність виявлення та розпізнавання таких об'єктів.

Подальші дослідження доцільно спрямувати на створення методу формування набору апіорних даних нейромережі для удосконалення процесу автоматичного дешифрування об'єктів на аерознімках отриманих з БпАК.

**Chen L.-C.** Semantic image segmentation with deep convolutional nets and fully connected CRFs [Електронний ресурс] / L.-C. Chen, G. Papandreou, I. Kokkinos, K.

- Murphy and A. L. Yuille // Режим доступа : <https://arxiv.org/abs/1412.7062.pdf>. **3. Yu F.** Multi-scale context aggregation by dilated convolutions [Электронный ресурс] / F. Yu, V. Koltun // Режим доступа : <https://arxiv.org/abs/1511.07122.pdf>. **4. Xiao T.** Unified perceptual parsing for scene understanding [Электронный ресурс] / T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun // Режим доступа : <https://arxiv.org/abs/1711.10370.pdf>. **5. Lin T.-Y.** Feature Pyramid Networks for Object Detection [Электронный ресурс] / T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie // Режим доступа : <https://arxiv.org/abs/1612.03144.pdf>. **6. Girshick R.** Rich feature hierarchies for accurate object detection and semantic segmentation [Электронный ресурс] / R. Girshick, J. Donahue, T. Darrell and J. Malik // Режим доступа : <https://arxiv.org/abs/1311.2524.pdf>. **7. Uijlings J. R.** Selective search for object recognition [Электронный ресурс] / J. R. Uijlings, K. E. V. D. Sande, T. Gevers, and A. W. Smeulders // Режим доступа : <https://arxiv.org/abs/1807.05511.pdf>. **8. Dai J.** Instance-sensitive fully convolutional networks [Электронный ресурс] / J. Dai, K. He, Y. Li, S. Ren, J. Sun // Режим доступа : <https://arxiv.org/abs/1603.08678.pdf>. **9. Li Y.** Fully Convolutional Instance-aware Semantic Segmentation [Электронный ресурс] / Q. Haozhi, D. Jifeng, J. Xiangyang, W. Yichen // Режим доступа : <https://arxiv.org/abs/1611.07709.pdf>. **10. Hosang J.** What makes for effective detection proposals? [Электронный ресурс] / J. Hosang, R. Benenson, P. Dollár, B. Schiele // Режим доступа : <https://arxiv.org/abs/1502.05082.pdf>. **11. Long J.** Fully convolutional networks for semantic segmentation [Электронный ресурс] / J. Long, E. Shelhamer, T. Darrell // Режим доступа : <https://arxiv.org/abs/1411.4038.pdf>. **12. He K.** Mask R-CNN [Электронный ресурс] / K. He, G. Gkioxari, P. Dollár, R. Girshick // Режим доступа : <https://arxiv.org/abs/1703.06870.pdf>. **13. Chen L.-C.** DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [Электронный ресурс] / L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille // Режим доступа : <https://arxiv.org/abs/1606.00915.pdf>. **14. Ren S.** Faster R-CNN: Towards realtime object detection with region proposal networks [Электронный ресурс] / S. Ren, K. He, R. Girshick, J. Sun // Режим доступа : <https://arxiv.org/abs/1506.01497.pdf>. **15. Tighe J.** Scene parsing with object instances and occlusion ordering [Электронный ресурс] / J. Tighe, M. Niethammer, S. Lazebnik // Режим доступа : <https://arxiv.org/abs/1806.03772.pdf>. **16. Tu Z.** Image parsing: Unifying segmentation, detection, and recognition [Электронный ресурс] / Z. Tu, X. Chen, A. L. Yuille, S.-C. Zhu // Режим доступа : <https://arxiv.org/abs/0502172.pdf>. **17. Yao J.** Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation [Электронный ресурс] / J. Yao, S. Fidler, R. Urtasun // Режим доступа : <https://arxiv.org/abs/1207.0372.pdf>. **18. Sun M.** Relating things and stuff via object property interactions [Электронный ресурс] / M. Sun, B.-S. Kim, P. Kohli, S. Savarese // Режим доступа : [http://svl.stanford.edu/assets/papers/pami14\\_acrf.pdf](http://svl.stanford.edu/assets/papers/pami14_acrf.pdf). **19. Yao J.** Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation [Электронный ресурс] / J. Yao, S. Fidler, and R. Urtasun // Режим доступа : [https://ttic.uchicago.edu/~rurtasun/publications/yao\\_et\\_al\\_cvpr12.pdf](https://ttic.uchicago.edu/~rurtasun/publications/yao_et_al_cvpr12.pdf). **20. Zhang S.** Single-shot refinement neural network for object detection [Электронный ресурс] / S. Zhang, L. Wen, X. Bian, Z. Lei, S. Z. Li // Режим доступа : <https://arxiv.org/abs/1711.06897.pdf>. **21. Redmon J.** You Only Look Once: Unified, Real-Time Object Detection [Электронный ресурс] / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // Режим доступа : <https://arxiv.org/abs/1506.02640.pdf>. **22. Gidaris S.** Object detection via a multiregion and semantic segmentation-aware CNN model [Электронный ресурс] / S. Gidaris, N. Komodakis // Режим доступа : <https://arxiv.org/abs/1711.06897.pdf>. **23. Cai Z.** Cascade R-CNN: Delving into high quality object detection [Электронный ресурс] / Z. Cai, N. Vasconcelos // Режим доступа : <https://arxiv.org/abs/1712.00726.pdf>. **24. Xie S.** Aggregated residual transformations for deep neural networks [Электронный ресурс] / Z. Cai, N. Vasconcelos // Режим доступа : <https://arxiv.org/abs/1611.05431.pdf>. **25. Dai J.** Deformable convolutional networks [Электронный ресурс] / J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei // Режим доступа : <https://arxiv.org/abs/1703.06211.pdf>. **26. Chen K.** Hybrid Task Cascade for Instance Segmentation [Электронный ресурс] / K. Chen, J. Pang, J. Wang and other // Режим доступа : <https://arxiv.org/abs/1901.07518.pdf>. **27. Lin T.-Y.** Focal Loss for Dense Object Detection [Электронный ресурс] / T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár // Режим доступа : <https://arxiv.org/abs/1708.02002.pdf>.

## МОДЕЛЬ КАСКАДА ГИБРИДНЫХ СЕГМЕНТАЦИЙ ДЛЯ АВТОМАТИЧЕСКОГО ДЕШИФРИРОВАНИЯ ОБЪЕКТОВ НА АЭРОСНИМКАХ

*Сергей Валентинович Ковбасюк (доктор технических наук, с. н. с.)*

*Леонид Брониславович Каневский (кандидат технических наук)*

*Николай Петрович Романчук*

*Житомирский военный институт имени С.П.Королева, Житомир, Украина*

Актуальной задачей сегодня является поиск и совершенствование методов автоматического детального дешифрирования объектов на аэроснимках, полученных с беспилотных авиационных комплексов, обеспечивающих достаточную точность обнаружения и распознавания мелкозернистых объектов в сложных топографических условиях местности, над которыми получают аэроснимки. С целью решения данной задачи в статье проведен анализ методов автоматической обработки изображений и моделей нейронных сетей, построенных на их основе. Из проведенного анализа выделен многостадийный конвейер обработки аэроснимков, что объединяет подходы выявления, элементной сегментации и семантическую сегментацию для получения контекста. Усовершенствованная модель каскада сегментации учитывает геометрические размеры объектов и их соотношение, изменение масштаба, условия съемки. Использование модели каскада сегментации для автоматического дешифрирования объектов на аэроснимках позволит повысить точность обнаружения и распознавания таких объектов.

**Ключевые слова:** нейросети, сегментация, аэроснимок, беспилотные авиационные комплексы, каскад гибридной сегментации.

**CASCADE MODEL OF HYBRID SEGMENTATION FOR AUTOMATIC DECISION OF AEROPHOTO OBJECTS**

*Serhiy Kovbasiuk (Doctor of technical sciences, Senior Research Fellow)*  
*Leonid Kanevskyy (Candidate of technical sciences)*  
*Mykola Romanchuk*

*Zhytomyr Military Institute named after S. P. Korolov, Zhytomyr, Ukraine*

*Current tasks for today are to search and improve the methods of automatic detailed decoding of objects on aerial photographs obtained from unmanned aviation complexes, which would provide sufficient accuracy of detection and recognition of fine-grained objects in the complex topographical conditions of the terrain above which aerial images are obtained. In order to solve this problem in the article an analysis of methods for automatic image processing and models of neural networks built on their basis. From the analysis, a multi-stage conveyor for aerial photographs has been selected, combining detection approaches, elemental segmentation and semantic segmentation for contextualization. Improved models of cascade of segmentation take into account geometric sizes of objects and their correlation, change in scale, conditions of removal. Using the segmentation cascade model for automatic decoding of objects on aerial photos will increase the accuracy of detection and recognition of such objects.*

**Key words:** neural network, segmentation, aerospace, unmanned aerial systems, cascade of hybrid segmentation.

**References**

1. **H. Zhao**, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. available at: <https://arxiv.org/abs/1612.01105.pdf>.
2. **L.-C. Chen**, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. available at: <https://arxiv.org/abs/1412.7062.pdf>.
3. **F. Yu**, V. Koltun. Multi-scale context aggregation by dilated convolutions. available at: <https://arxiv.org/abs/1511.07122.pdf>.
4. **T. Xiao**, Y. Liu, B. Zhou, Y. Jiang, and J. Sun. Unified perceptual parsing for scene understanding. available at: <https://arxiv.org/abs/1711.10370.pdf>.
5. **T.-Y. Lin**, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie. Feature Pyramid Networks for Object Detection. available at: <https://arxiv.org/abs/1612.03144.pdf>.
6. **R. Girshick**, J. Donahue, T. Darrell and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. available at: <https://arxiv.org/abs/1311.2524.pdf>.
7. **J. R. Uijlings**, K. E. V. D. Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. available at: <https://arxiv.org/abs/1807.05511.pdf>.
8. **J. Dai**, K. He, Y. Li, S. Ren, J. Sun. Instance-sensitive fully convolutional networks. available at: <https://arxiv.org/abs/1603.08678.pdf>.
9. **Q. Haozhi**, D. Jifeng, J. Xiangyang, W. Yichen. Fully Convolutional Instance-aware Semantic Segmentation. available at: <https://arxiv.org/abs/1611.07709.pdf>.
10. **J. Hosang**, R. Benenson, P. Dollár, B. Schiele. What makes for effective detection proposals? available at: <https://arxiv.org/abs/1502.05082.pdf>.
11. **J. Long**, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation. available at: <https://arxiv.org/abs/1411.4038.pdf>.
12. **K. He**, G. Gkioxari, P. Dollár, R. Girshick. Mask R-CNN. available at: <https://arxiv.org/abs/1703.06870.pdf>.
13. **L.-C. Chen**, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. available at: <https://arxiv.org/abs/1606.00915.pdf>.
14. **S. Ren**, K. He, R. Girshick, J. Sun. Faster R-CNN: Towards realtime object detection with region proposal networks. available at: <https://arxiv.org/abs/1506.01497.pdf>.
15. **J. Tighe**, M. Niethammer, S. Lazebnik. Scene parsing with object instances and occlusion ordering. available at: <https://arxiv.org/abs/1806.03772.pdf>.
16. **Z. Tu**, X. Chen, A. L. Yuille, S.-C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. available at: <https://arxiv.org/abs/0502172.pdf>.
17. **J. Yao**, S. Fidler, R. Urtasun. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. available at: <https://arxiv.org/abs/1207.0372.pdf>.
18. **M. Sun**, B.-S. Kim, P. Kohli, S. Savarese. Relating things and stuff via object property interactions. available at: [http://svl.stanford.edu/assets/papers/pami14\\_acrf.pdf](http://svl.stanford.edu/assets/papers/pami14_acrf.pdf).
19. **J. Yao**, S. Fidler, and R. Urtasun. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. available at: [https://ttic.uchicago.edu/~rurtasun/publications/yaoetal\\_cvpr12.pdf](https://ttic.uchicago.edu/~rurtasun/publications/yaoetal_cvpr12.pdf).
20. **S. Zhang**, L. Wen, X. Bian, Z. Lei, S. Z. Li. Single-shot refinement neural network for object detection. available at: <https://arxiv.org/abs/1711.06897.pdf>.
21. **J. Redmon**, S. Divvala, R. Girshick, A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. available at: <https://arxiv.org/abs/1506.02640.pdf>.
22. **S. Gidaris**, N. Komodakis. Object detection via a multiregion and semantic segmentation-aware CNN model. available at: <https://arxiv.org/abs/1711.06897.pdf>.
23. **Z. Cai**, N. Vasconcelos. Cascade R-CNN: Delving into high quality object detection. available at: <https://arxiv.org/abs/1712.00726.pdf>.
24. **Z. Cai**, N. Vasconcelos. Aggregated residual transformations for deep neural networks. available at: <https://arxiv.org/abs/1611.05431.pdf>.
25. **J. Dai**, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei. Deformable convolutional networks. available at: <https://arxiv.org/abs/1703.06211.pdf>.
26. **K. Chen**, J. Pang, J. Wang and other. Hybrid Task Cascade for Instance Segmentation. available at: <https://arxiv.org/abs/1901.07518.pdf>.
27. **T.-Y. Lin**, P. Goyal, R. Girshick, K. He, P. Dollár. Focal Loss for Dense Object Detection. available at: <https://arxiv.org/abs/1708.02002.pdf>.