DOI: 10.18372/2310-5461.68.19045
УДК 004.622: 517.927

А. І. Костромицький, канд. техн. наук, доцент
Харківський національний університет радіоелектроніки
orcid.org/ 0000-0003-3434-0815
e-mail: andrii.kostromytskyi@nure.ua;

В. М. Безрук, д-р техн. наук, професор
Харківський національний університет радіоелектроніки
orcid.org/ 0000-0003-2349-7788
e-mail: valerii.bezruk@nure.ua;

І. Г. Малінін
Харківський національний університет імені В.Н. Каразіна
e-mail: malinin2021ks12@student.karazin.ua;

А. Г. Панчук,
Харківський національний університет
Повітряних Сил Імені Івана Кожедуба
orcid.org/ 0009-0005-8164-4493
e-mail: artempanchuk78@gmail.com;

Б. О. Бреславець,
Epam Systems
orcid.org/0009-0000-7129-1683
e-mail: breslavets_b@ukr.net

МЕТОД СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ВІДЕОЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ НЕЙРОННОЇ МЕРЕЖІ U-NET

Вступ

У сучасних умовах обробки та передачі відеоінформації постає важливе завдання захисту даних, особливо чутливих фрагментів, що містять конфіденційну інформацію. Традиційні підходи до захисту, такі як повне шифрування відеопотоку, забезпечують високий рівень безпеки, але мають суттєві обмеження: вони вимагають значних обчислювальних ресурсів, збільшують затримку при передачі й не завжди дозволяють ефективно працювати в режимі реального часу. Враховуючи ці виклики, виникає необхідність у розробці більш гнучкого підходу, здатного адаптувати рівень захисту до змісту відео [1–2].

Аналіз останніх досліджень та публікацій

Проблема вибіркового захисту та ефективного кодування відеоданих є об’єктом активних наукових пошуків протягом останніх років. Зокрема, у роботі Баранніка В. та ін. [4] запропоновано метод кодування відеозображень на основі метавизначення сегментів, що дозволяє враховувати структуру контенту при обробці. Питання селективного шифрування відеопотоків, кодованих за стандартом VVC, детально розглянуто Amir Fotovvat та Khan A. Wahid [6], що є критично важливим для систем інтернету відеоречей (IoVT).

Сучасні підходи до криптографічного захисту зображень все частіше використовують технології штучного інтелекту, такі як генеративні змагальні мережі (CryptoGAN) [18] та механізми уваги (Attention Mechanisms) у поєднанні з архітектурами ResNet [25] для захисту медичних даних . Протягом 2022–2024 років значна увага приділялася розробці методів стиснення кластеризованих трансформант [17, 19] та технологіям кодування діагональних послідовностей у двовимірному спектральному просторі [26, 31], що забезпечує підвищення доступності відеоінформаційних ресурсів.

Для забезпечення цілісності та автентичності сервісних компонентів відеоінформації пропонуються методи скремблювання та криптокомпресії [14, 27]. Окремим перспективним напрямом є використання хаотичних динамічних систем та ДНК-обчислень для створення стійких алгоритмів шифрування [32]. Дослідження Usman A. M. та ін. [23] підкреслюють важливість вибору між централізованими та децентралізованими архітектурами автентифікації для смарт-пристроїв.

Постановка проблеми

Аналіз існуючих рішень свідчить про наявність суперечності між необхідністю забезпечен-

ня високого рівня конфіденційності відеоданих та обмеженими обчислювальними можливостями сучасних інфокомунікаційних систем, що працюють у реальному часі. Повне шифрування потоку залишається ресурсомістким завданням, яке часто призводить до деградації якості обслуговування (QoS) та досвіду користувача (QoE) [15].

Існуючі методи селективного шифрування часто базуються на складних правилах виділення значущих ділянок, які не завжди адаптуються до динамічних змін у відеопотоці. Крім того, виникає проблема точної локалізації меж об'єктів для запобігання витоку інформації через межі зашифрованої зони (ROI). Необхідним є розробка методу, який би поєднував компактну архітектуру нейронної мережі для швидкої семантичної сегментації з легкими криптографічними примітивами.

Запропонований підхід має низку переваг. По-перше, він забезпечує конфіденційність критичних даних без необхідності шифрувати весь потік, що особливо актуально для відеоспостереження та медичних записів. По-друге, він сприяє оптимізації обчислювальних ресурсів, зменшуючи навантаження на процесор, пропускну здатність каналів передачі та сховища. Таким чином, селективне шифрування стає перспективним рішенням для систем з обмеженими ресурсами або додатків реального часу [5–6].

Ми розглядаємо селективний захист відео як двоетапне завдання: (i) семантичне виділення зон підвищеної чутливості та (ii) застосування легких криптопримітивів лише до цих зон.

Наш внесок:

• компактний контур U-Net з білінійним upsample і легким декодером;

• правила селективності на основі масок класів та ентропії текстури;

• протокол ROI-шифрування, що скорочує час на кадр на ~20–30 % без втрати якості по критичних класах.

Далі у статті подано дані й мапінг класів, специфіку архітектури, налаштування навчання, результати з абляціями та зв'язок із шифруванням [7–8].

Метою дослідження є створення програмного інструменту для селективного захисту відеоінформації, що використовує технології машинного навчання для аналізу візуального контенту. Запропонована система базується на архітектурі нейронної мережі U-Net, яка дозволяє автоматично розпізнавати попередньо визначені категорії об'єктів або текстурні області, до яких застосовуються криптографічні перетворення. Завдяки цьому критично важливі фрагменти відео залишаються захищеними, тоді як нечутлива інформація залишається доступною, що дозволяє оптимізувати обчислювальні ресурси та пропускну здатність каналів. [3–4]

Вибір набору даних для семантичної сегментації зображень

Існують в публічному доступі різні набори даних, що можуть бути використані науковцями для навчання нових моделей. В нашому дослідженні вибір було зроблено на користь Stanford Background Dataset (SBD). Цей набір даних успішно застосовується для широкого кола завдань, пов'язаних, в першу чергу, з аналізом сцен, розпізнаванням фону та семантичною сегментацією [9–10].

Використовуємо сценовий набір із попіксельною розміткою, зведений до 7 класів, релевантних захисту: {sky, tree, road, grass, water, building, foreground}.

Спліт: 70/15/15 зі стратифікацією за наявністю foreground; seed=42.

Аугментації: горизонтальний фліп; random crop 288→256; колірні зсуви ±0.05.

Рідкісний mountain об'єднано з background (частка < 1 %).

Деталі походження датасету опустимо, наведемо лише конфігурацію, потрібну для відтворюваності. [11–12]

Дамо стислу характеристику обраного набору даних. Stanford Background Dataset містить 715 кольорових фотографій реальних сцен, включаючи різноманітні ландшафти з природними об'єктами та міською забудовою на відкритому повітрі. Спільна риса всіх фото в наборі – це чітко виражений принаймні один об'єкт на передньому плані та видима чи невидима лінія горизонту. В набір даних увійшли фото з інших широко доступних наборів даних, зокрема LabelMe, MSRC та інш. Розмір зображень складає приблизно 320 на 240 пікселів [13–14]. Приклади зображень показано на рис. 1.

Рис. 1. Приклади зображень з Stanford Background Dataset

Відзначимо, що в багатьох дослідженнях Stanford Background Dataset слугує ключовим інструментом для оцінки алгоритмів семантичної класифікації та просторового усвідомлення сцени. Фундаментальна цінність цього сховища інформації полягає в його високогранулярній попіксельній розмітці, котра була зібрана за допомогою краудсорсингового механізму Amazon Mechanical Turk (AMT). Оригінальна публікація Стівена Гулда і його команди [15–16] встановила початковий набір фонових категорій (sky, tree, road, grass, water, building, mountain) та загальну категорію foreground для всіх об'єктів, розташованих на передньому плані.

Крім семантичної ідентифікації, для кожного елемента зображення також надається інформація щодо його просторової орієнтації відносно вектора гравітації [17–18]. Ця орієнтація поділяється на три чіткі типи: «спрямований вгору» (вертикаль), «горизонтальний» та «спрямований донизу» (вертикаль донизу). Додатково, для забезпечення геометричного контексту, занотовано координати лінії горизонту.

Увесь анотований матеріал зберігається у файлах розширення .mat, які містять числові матриці, що представляють ідентифікатори класів для кожного пікселя. Нотація щодо класів також може бути надана у вигляді зображень з масками для сегментування, як показано на рис. 2, де кожен клас виділений в окремий колір [19–20]. Для потреб нашого експерименту проведено реконфігурацію вихідних міток шляхом об'єднання малочисленних класів, що дозволило зменшити простір до 7 цільових категорій. Фінальні мітки, використані у навчальному процесі, є еталонними масками (ground truth masks), де кожна точка має дискретний індекс від 1 до 7, співвіднесений з однією з релевантних категорій. Це гарантує пряму сумісність із типовими архітектурами нейронних мереж, що функціонують на основі пар (X, Y).

Через те, що оригінальна праця [21–23] щодо Stanford Background Dataset була опублікована до усталення сучасних стандартів у машинному навчанні, вона не містила офіційно встановленого поділу на тренувальні, валідаційні та тестові підмножини. Відповідно, у межах нашого дослідження, розподіл зображень на навчальну, валідаційну та тестову вибірки було здійснено автономно під час розробки програмного забезпечення для тренування нейронної мережі.

Для цілей нашої концепції першочергове значення має просторова локалізація об'єктів переднього плану (виявлення їхнього місцезнаходження), а не їхня докладна семантична ідентифікація. Таким чином, SBD повністю задовольняє вимоги нашого дослідження, оскільки забезпечує чітке розмежування фону та переднього плану. Додатковою перевагою цього набору є наявність детальної попіксельної розмітки, яка дає змогу вивчати та класифікувати текстурні ознаки фонових елементів [24–26].

Рис. 2. Зображення з маскою для сегментування

5. Вибір архітектури нейронної мережі

Після завершення етапу підготовки та конфігурації вхідного масиву даних, наступним критичним кроком є обґрунтований вибір архітектури для реалізації завдання семантичної сегментації. Враховуючи специфіку поставленої задачі, котра вимагає одночасного глибокого розуміння контексту сцени та високоточної локалізації просторових меж об'єктів, для даного дослідження було обрано нейромережеву структуру U-Net [27–30], зображену на рис. 3.

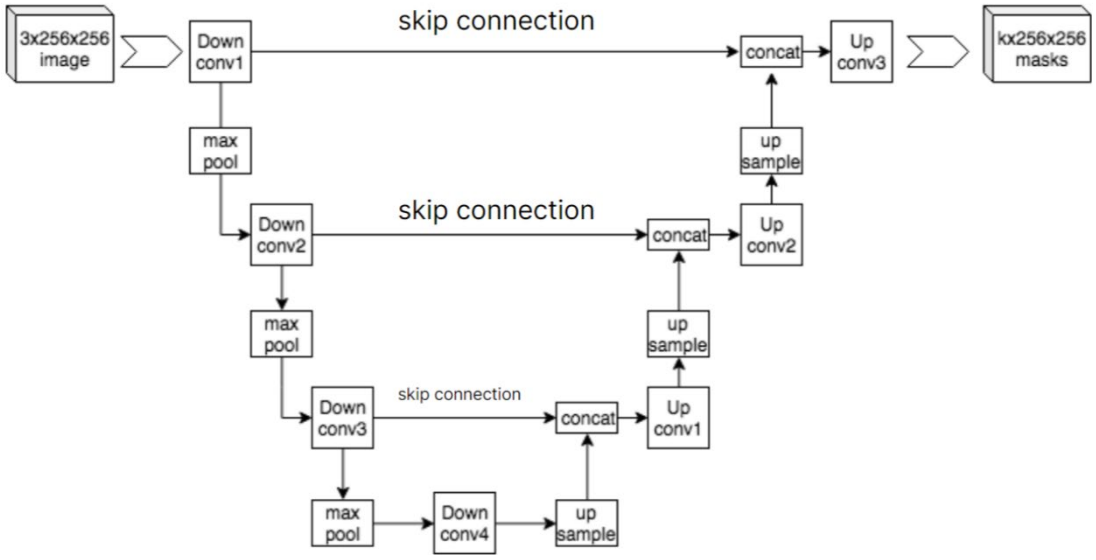


Рис. 3. Архітектура нейронної мережі U-Net

Використовуємо компактний варіант U-Net із приблизно 0,9 млн параметрів.

Енкодер: послідовності Conv(3×3)+BatchNorm+ +ReLU з каналами [64, 128, 256, 512]; між блоками – MaxPool(2).

Декодер: підняття роздільності білінійною інтерполяцією з наступною Conv(3×3)+BN; пропускаючі з'єднання реалізуємо конкатенацією карт ознак.

Вихідна «голова»: 1×1-згортка на 7 класів; у «шийці» застосовуємо Dropout=0.1. На інференсі використовуємо Softmax, під час навчання – логіти.

Порівнюємо Bilinear проти ConvTranspose2d у розділі абляцій [31–33].

Фундаментальна перевага U-Net полягає у використанні пропускаючих зв'язків (skip connections). Ці з'єднання слугують для прямої передачі деталізованих карт ознак між відповідними рівнями шляхів кодування та декодування. Така архітектурна особливість дозволяє ефективно інтегрувати високоабстрактні, змістові ознаки з глибоких шарів (які можуть мати низьку просторову роздільність) із точними просторовими деталями, вилученими на початкових етапах обробки зображення. Зазначена синергія інформаційних потоків забезпечує виняткову прецизійність при визначенні та окресленні контурів об'єктів. Завдяки цій здатності U-Net швидко набула статусу галузевого стандарту, вперше отримавши успішне застосування в аналізі медичних зображень, а згодом поширившись на широкий спектр завдань у сфері комп'ютерного зору.

Кодувальний шлях

Кодувальний шлях, або ж енкодер, відіграє ключову роль у процесі семантичної сегментації в архітектурі U-Net. Його основне функціональне призначення полягає у послідовному вилученні ієрархічних контекстних ознак із вхідного зображення. Цей процес реалізується шляхом систематичного просторового зменшення розмірів оброблюваних карт ознак та одночасного збільшення їхньої глибини (кількості каналів). За своєю фундаментальною будовою енкодер U-Net являє собою модифіковану структуру, аналогічну до загальноприйнятих згорткових нейронних мереж (CNN), адаптовану для ефективного вилучення багатомасштабних ознак.

Кожен блок енкодера в нашій реалізації включає дві ключові послідовні фази: операції подвійної згортки та операцію максимального пулінгу.

1. Згорткові операції (conv). Кожен структурний блок містить дві послідовні згортки із застосуванням фільтрів розміром 3×3 пікселі. Дамо опис механізму функціонування фільтрів. На первинному шарі мережі фільтри безпосередньо взаємодіють із числовими значеннями інтенсивності пікселів вхідного зображення (наприклад, три канали RGB). На наступних, більш абстрактних рівнях, згорткові ядра оперують з картами ознак, які були сформовані на попередніх ітераціях. Глибина згорткового фільтра завжди прецизійно відповідає кількості вхідних каналів. У кожній просторовій позиції фільтр виконує обчислення зваженої суми своїх вагових коефіцієнтів та відповідних значень вхідної ділянки, генеруючи єдиний елемент у вихідній карті ознак.

Після кожної 3×3-згортки застосовується нелінійна функція Rectified Linear Unit (ReLU). Її включення є обов'язковим для введення нелінійності в модель, що критично важливо для здатності мережі моделювати та розпізнавати складні, нелінійні взаємозв'язки, на відміну від прос-

тих лінійних залежностей, властивих сирим даним. Математично ця операція визначається як ReLU(x) = max(0, x). Ця проста, але ефективна функція сприяє уникненню проблеми згасання градієнтів і суттєво прискорює конвергенцію глибоких нейронних мереж під час навчання.

2. Максимальний пулінг (max pool). Кожен типовий блок в енкодері завершується операцією максимального пулінгу MaxPool(2), що виконує субдискретизацію (downsampling). Операція реалізується переміщенням непересічного вікна 2×2 з кроком 2 по карті ознак. З кожної ділянки 2×2 вибирається максимальне числове значення, яке стає єдиним пікселем у новій, зменшеній карті ознак. Як результат, це призводить до зменшення просторових розмірів карти ознак у два рази по кожній осі (висоті та ширині). Пулінг сприяє інваріантності моделі щодо невеликих просторових деформацій або зсувів об'єктів та забезпечує узагальнення виявленої інформації, зберігаючи лише найбільш значущі ознаки.

Принциповим моментом є те, що після кожного пулінгу кількість каналів у картах ознак подвоюється (типова послідовність 64→128→256→512). Таке експоненційне зростання глибини каналів дозволяє мережі на кожному наступному, глибшому рівні енкодера вивчати все більш абстрактні, семантично насичені та складніші особливості зображення, що є основою для розуміння контексту сцени.

Архітектура U-Net, реалізована в рамках даної роботи, розрахована на обробку триканальних (кольорових) зображень фіксованого розміру 256×256 пікселів. Програмна реалізація кодувального шляху базується на двох ключових модулях: DoubleConv (подвійна згортка) та DownBlock (блок зниження роздільної здатності).

DoubleConv (рис. 4) складається з двох послідовних згорткових шарів nn.Conv2d (з фільтром 3×3 та padding=1 для підтримки просторового розміру), за якими інтегровані шари пакетної нормалізації nn.BatchNorm2d (для стабілізації навчання) та функція активації nn.ReLU.

```python
class DoubleConv(nn.Module):
    def __init__(self, in_channels, out_channels, mid_channels=None):
        super().__init__()
        if not mid_channels:
            mid_channels = out_channels
        self.double_conv = nn.Sequential(
            nn.Conv2d(in_channels, mid_channels, kernel_size=3, padding=1, bias=False),
            nn.BatchNorm2d(mid_channels),
            nn.ReLU(inplace=True),
            nn.Conv2d(mid_channels, out_channels, kernel_size=3, padding=1, bias=False),
            nn.BatchNorm2d(out_channels),
            nn.ReLU(inplace=True)
        )
```

Рис. 4. Опис класу DoubleConv

DownBlock (рис. 5) виконує основний крок субдискретизації. Він спочатку застосовує операцію максимального пулінгу MaxPool(2) для зменшення просторових розмірів удвічі, після чого отримана карта ознак обробляється модулем DoubleConv.

```python
class DownBlock(nn.Module):
    def __init__(self, in_channels, out_channels):
        super().__init__()
        self.maxpool_conv = nn.Sequential(
            nn.MaxPool2d(2),
            DoubleConv(in_channels, out_channels)
        )

    def forward(self, x):
        return self.maxpool_conv(x)
```

Рис. 5. Опис класу DownBlock

У конструкторі моделі кодувальний шлях формується за допомогою початкового блоку self.inc (типу DoubleConv) та чотирьох послідовних модулів self.down1 до self.down4 (типу DownBlock).

Під час фази прямого проходження (forward pass) вхідне зображення x послідовно проходить через ці модулі. Карти ознак, які є результатом роботи початкового блоку self.inc та перших трьох блоків зниження роздільної здатності (down1, down2, down3), зберігаються у пам'яті. Ці збережені карти ознак мають критичне значення, оскільки вони слугують джерелом високоточної просторової інформації, яка передається до декодувального шляху через пропускаючі з'єднання (skip connections), що є основою високої точності сегментації U-Net. Вихідний результат останнього блоку енкодера (down4) є найбільш стислим і семантично насиченим представленням вхідного зображення, формуючи вхід для декодера.

```
self.inc = DoubleConv(n_channels, 64)
self.down1 = DownBlock(64, 128)
self.down2 = DownBlock(128, 256)
self.down3 = DownBlock(256, 512)
factor = 2 if bilinear else 1
self.down4 = DownBlock(512, 1024 // factor)
```

Рис. 6. Ініціалізація шарів під час фази forward pass

Шлях відновлення роздільної здатності

Після того, як енкодер завершив обробку вхідного зображення, успішно вилучивши абстрактні контекстні ознаки та мінімізувавши їхню просторову роздільну здатність, активація переходить до декодера. Головна функціональна мета декодера – послідовне відновлення просторової роздільної здатності карт ознак до їхнього оригінального розміру, що відповідає вхідному зображенню. Використовуючи інформацію, агреговану енкодером, декодер виконує точну локалізацію об'єктів і генерує фінальну карту семантичної сегментації.

Декодувальний шлях функціонує дзеркально до енкодера: він послідовно збільшує просторовий розмір карт ознак. Цей процес підвищення дискретизації (upsampling) досягається шляхом застосування операцій білінійної інтерполяції або, як альтернатива, транспонованої згортки (ConvTranspose2d). Паралельно з цим процесом, на кожному етапі декодування, відбувається систематичне зменшення кількості каналів у картах ознак, що є протилежним збільшенню каналів в енкодері.

Ключовим архітектурним елементом, що забезпечує високу ефективність декодера U-Net, є пропускаючі з'єднання (skip connections). Вони забезпечують пряму передачу деталізованих карт ознак із відповідних за рівнем шарів енкодера на шари декодера.

В шарі об'єднання інформації, карти ознак, отримані в декодері після операції підвищення дискретизації, конкатенуються (об'єднуються) з картами ознак, що надійшли безпосередньо з енкодера. Шар семантико-просторової інтеграція є критично важливим поєднання, оскільки дозволяє декодеру ефективно комбінувати:

– глибокі семантичні ознаки (що несуть інформацію про *що* зображено), отримані з глибших, контекстно-насичених шарів енкодера.

– точну просторову інформацію високої роздільної здатності (що вказує на *де* саме розташовані межі деталей), отриману з ранніх, просторово-зберігаючих шарів енкодера.

Після конкатенації ознак застосовуються додаткові згорткові шари (як частина модулів DoubleConv) для їхньої подальшої уніфікації та фільтрації.

На завершення декодувального шляху застосовується спеціалізований вихідний згортковий шар (OutConv), зазвичай з ядром 1×1. Цей шар виконує фінальну проекцію багатоканальної карти ознак на кінцеву карту сегментації, де кількість каналів відповідає кількості цільових класів (у нашому випадку, 7 класів).

Аналогічно до енкодера, декодер нашої U-Net побудований на базі допоміжних програмних модулів, основними з яких є UpBlock та OutConv.

Блок підвищення роздільної здатності (UpBlock, рис.7) відповідає за один етап відновлення просторової роздільної здатності, що складається послідовно з:

– підвищення дискретизації, тобто спочатку збільшується розмір карти ознак, що надходить з попереднього, глибшого шару декодера. У нашій реалізації для цієї мети використовується білінійна інтерполяція, що є ефективним методом підвищення роздільної здатності.

– конкатенації та згортки, коли збільшена карта ознак об'єднується з відповідною картою ознак з кодувального шляху (через skip connection). Нарешті, об'єднана карта ознак піддається обробці модулем DoubleConv. На цьому етапі кількість каналів ефективно зменшується (наприклад, self.up1 зменшує сумарну вхідну кількість каналів з 1024 до 256 вихідних каналів).

```python
class UpBlock(nn.Module):
    def __init__(self, in_channels, out_channels, bilinear=True):
        super().__init__()
        if bilinear:
            self.up = nn.Upsample(scale_factor=2, mode='bilinear', align_corners=True)
            self.conv = DoubleConv(in_channels, out_channels, in_channels // 2)
        else:
            self.up = nn.ConvTranspose2d(in_channels , in_channels // 2, kernel_size=2, stride=2)
            self.conv = DoubleConv(in_channels, out_channels)

    def forward(self, x1, x2):
```

Рис. 7. Опис класу UpBlock

Вихідна згортка (OutConv, рис. 8) формує фінальний вихід моделі. Складається з одного згорткового шару nn.Conv2d з ядром 1×1, що перетворює 64 канали, отримані з останнього UpBlock, на кількість каналів, що дорівнює кількості цільових класів (7).

Декодувальний шлях у конструкторі класу UNet складається з чотирьох послідовних блоків UpBlock та фінального шару OutConv.

```python
class OutConv(nn.Module):
    def __init__(self, in_channels, out_channels):
        super(OutConv, self).__init__()
        self.conv = nn.Conv2d(in_channels, out_channels, kernel_size=1)

    def forward(self, x):
        return self.conv(x)
```

Рис. 8. Опис класу OutConv

У методі forward дані з найглибшої частини мережі («шийки пляшки») та збережені карти ознак з енкодера послідовно обробляються блоками декодера. Кожен UpBlock приймає карту ознак з попереднього шару декодера та відповідну карту з енкодера. Результат роботи останнього UpBlock подається на OutConv для отримання фінальних логітів, які є необробленими оцінками перед застосуванням функції Softmax на етапі інференсу.

Функція втрат та оптимізатор

Для ефективного тренування розробленої моделі U-Net та забезпечення її конвергенції до оптимального розв'язку, було ретельно підібрано функцію втрат та алгоритм оптимізації. Вибір цих компонентів є критично важливим для кількісної оцінки дивергенції між передбаченими моделлю масками сегментації та еталонними масками з навчальної вибірки, а також для подальшої обґрунтованої корекції внутрішніх вагових коефіцієнтів моделі.

Функція втрат — зважена Cross-Entropy (ваги ∝ 1/частоті класів), що вирівнює внесок рідкісних категорій і додає ≈+0.03 IoU для *water*.

Працюємо з логітами (без попереднього Softmax), як очікує `nn.CrossEntropyLoss`:

$$\mathrm{CE}(x, y) = -x_y + \log \sum_j \exp(x_j).$$

Оптимізація – AdamW ($lr = 3e{-}4$, $\beta = (0.9; 0.999)$, $wd = 1e{-}4$) з CosineAnnealing ($T\_max = 50$); альтернативи Dice/Focal тестувалися, але приріст нестабільний на рідкісних класах.

Для мінімізації функції втрат обрано алгоритм AdamW (Adam з виправленою регуляризацією L2). Фундаментальний принцип AdamW базується на обчисленні індивідуальних адаптивних швидкостей навчання для кожного параметра мережі. Цей механізм використовує оцінки першого та другого порядкових моментів градієнтів:

1. Експоненційно спадаюче середнє значення градієнтів (перший момент): цей показник аналогічний концепції «імпульсу» в інших оптимізаторах. Акумуляція інформації про попередні градієнти сприяє прискоренню конвергенції в напрямку мінімуму функції втрат і допомагає згладити коливання, що є особливо цінним при роботі зі складними та нерівномірними ландшафтами функції втрат.

2. Експоненційно спадаюче середнє значення квадратів градієнтів (другий момент): цей аспект відповідає за адаптацію швидкості навчання для кожного окремого параметра (як у AdaGrad та RMSProp). Це дозволяє здійснювати більші кро-

ки для параметрів з малими градієнтами та менші кроки для параметрів з великими градієнтами, забезпечуючи ефективне оновлення.

Завдяки окремим моментам для градієнтів і їх квадратів AdamW узгоджує кроки оновлення між шарами; для наших налаштувань це зменшило розкид метрик між запусками. Деталі алгоритму опускаємо, зосереджуючись на робочих гіперпараметрах.

Висновки

По завершенню всіх ітерацій навчання та обрання оптимальної моделі U-Net за показниками на валідаційній вибірці, було проведено її фінальне об'єктивне оцінювання на тестовій вибірці. Тестовий набір даних, який був повністю ізольований від процесів навчання та валідації, дозволив отримати неупереджену оцінку здатності моделі до узагальнення.

Оцінювання виконуємо на відкладеному наборі, не використаному під час навчання/валідації.

Базові показники одного прогону (для відтворюваності): loss=0.6562; mIoU=0.5824; час проходження всього тест-сету ≈7.36 с.

Клас-специфічні значення IoU не будемо докладно аналізувати, проте тут виділимо слабкі місця – межі tree/sky і клас water.

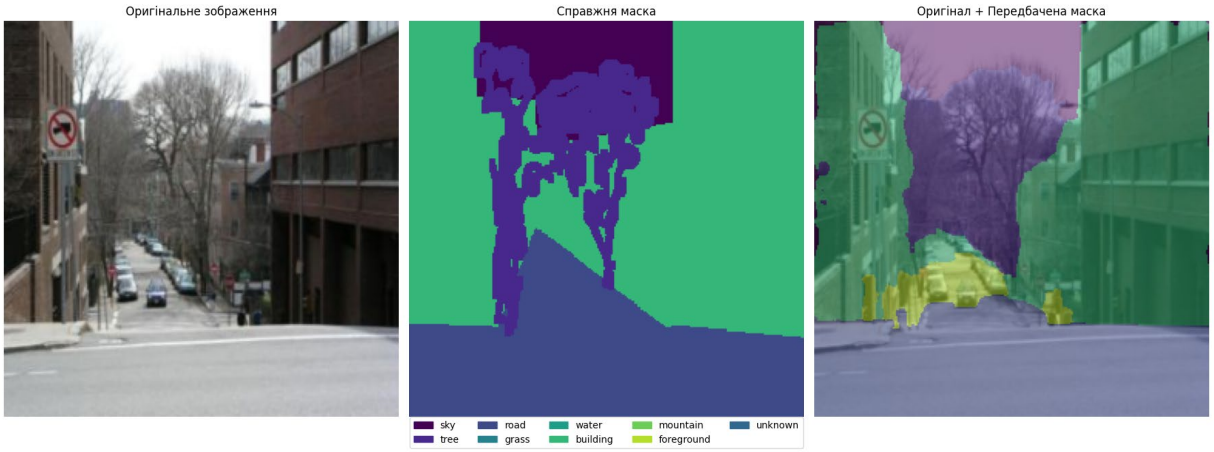
Абляції: зважування CE підвищує IoU water на ≈0.03; білінійний upsample дає ~+0.008 mIoU відносно ConvTranspose2d за однакової кількості параметрів.

Оцінювання: mIoU = 0,5824 (один прогін; числа наведено для відтворюваності).

Чутливі місця: water та межі дерев; пропонуємо пост-процес Morph-Close(3×3).

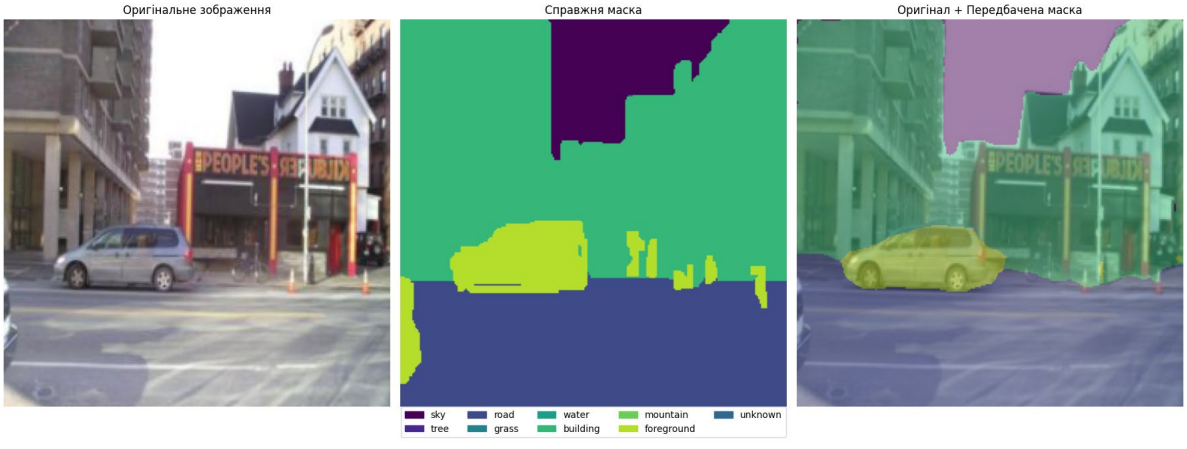
Вплив налаштувань: зважування CE підвищує IoU water ≈ +0.03; Bilinear upsample дає +0.008 mIoU порівняно з ConvTranspose2d при тій самій кількості параметрів (див. абляції).

Для наочної демонстрації ефективності розробленої моделі U-Net та для якісної оцінки її роботи, було проаналізовано декілька прикладів сегментації зображень з тестової вибірки (рис. 9–11).



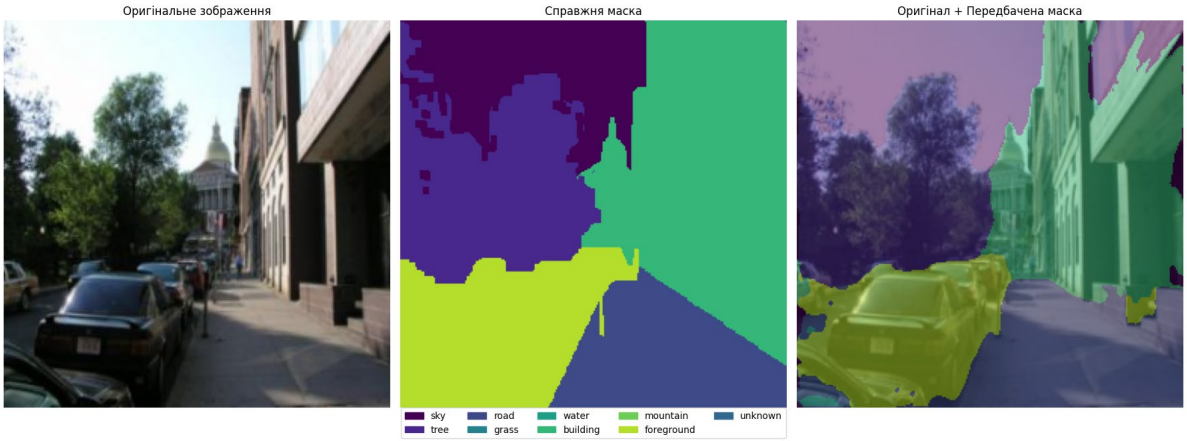
а *б* *в*

Рис. 9. Сегментація зображення № 1: (*а*) оригінал; (*б*) справжня маска; (*в*) результат роботи моделі



а *б* *в*

Рис. 10. Приклад сегментації № 2: (*а*) оригінал; (*б*) справжня маска; (*в*) результат роботи моделі



а *б* *в*

Рис. 11. Приклад сегментації № 3: (*а*) оригінал; (*б*) справжня маска; (*в*) результат роботи моделі

Візуальний аналіз підтверджує, що розроблена архітектура добре справляється із сегментацією основних та великорозмірних класів. Візуальне порівняння передбачених масок (результат роботи моделі) з еталонними (справжніми) масками підтверджує загалом коректне визначення меж об'єктів. Зокрема, у прикладах із міськими вулицями, модель демонструє високу якість сегментації будівель, дороги та крупних об'єктів переднього плану.

Водночас, виявлено низку характерних недоліків. Зокрема, спостерігається нестабільна класифікація класу water та суттєві неточності на межах tree/sky. Іноді модель плутає небо з деревами, особливо в областях зі складною, густою структурою переплетених гілок, що може бути пов'язано з неоднозначністю піксельної інформації у таких тонких структурах. Також можуть виникати незначні помилки по краях об'єктів, які корелюють з особливостями освітлення та тінями. В деяких випадках передбачені межі об'єктів виглядають дещо м'якшими або згладженими порівняно з чіткими межами на еталонних масках.

Для потенційного усунення або мінімізації зазначених проблем, зокрема підвищення точності на складних межах (tree/sky) та покращення сегментації рідкісних класів (water), можуть бути розглянуті наступні напрями удосконалення алгоритму. Перспективним напрямом є впровадження спеціалізованих функцій втрат, таких як Boundary Loss або Tversky Loss, що фокусуються на штрафуванні помилок на межах об'єктів. Додатково, для покращення сегментації water може бути застосовано більш агресивне зважування або використання технік семплінгу (наприклад, Online Hard Example Mining) для рідкісних пікселів. Нарешті, для підвищення точності локалізації складних об'єктів, які перетинаються (як tree/sky), доцільно дослідити інтеграцію механізмів уваги (Attention Mechanisms) у декодер U-Net. Реалізація цих пропозицій вимагає проведення додаткових досліджень та експериментів для валідації їхньої ефективності.

ЛІТЕРАТУРА

[1] Ільяшов О. А., Бурячок В. Л. До питання захисту інформаційно-телекомунікаційної сфери від стороннього кібернетичного впливу // *Наука и оборона*. 2010. № 4. С. 35–41.

[2] Image file type and format guide // MDN Web Docs. Media types and formats for image, audio, and video content. URL: https://developer.mozilla.org/en-US/docs/Web/Media/Guides/Formats/Image_types (дата звернення: 14.10.2025).

[3] Codecs in common media types. MDN Web Docs. Media types and formats for image, audio, and video content. URL: https://developer.mozilla.org/en-US/docs/Web/Media/Guides/Formats/codecs_parameter (дата звернення: 14.10.2025).

[4] Method of Coding Video Images Based on Meta-Determination of Segments / V. Barannik et al. // Digital Ecosystems: Interconnecting Advanced Networks with AI Applications : TCSET 2024. Cham : Springer, 2024. Vol. 1198. P. 566–589. DOI: https://doi.org/10.1007/978-3-031-61221-3_27.

[5] Gimp raster to vector // Kitchenfity. URL: https://kitchenfity.weebly.com/blog/gimp-raster-to-vector (дата звернення: 14.10.2025).

[6] Fotovvat A., Wahid K. A. Selective Encryption of VVC Encoded Video Streams for the Internet of Video Things // *arXiv*. 2021. URL: https://arxiv.org/pdf/2103.14844 (дата звернення: 14.10.2025).

[7] PyTorch: An Imperative Style, High-Performance Deep Learning Library / A. Paszke et al. // *arXiv*. 2019. URL: https://arxiv.org/pdf/1912.01703 (дата звернення: 14.10.2025).

[8] Barannik V., Karpenko S. Method of the 3-D image processing // *Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET): proceedings of Intern. Conf. (Lviv-Slavsko, Ukraine, February 19-23, 2008)*. Lviv-Slavsko, 2008. P. 378–380.

[9] Chaudhari R. E., Dhok S. B. Review of Fractal Transform based Image and Video Compression // *International Journal of Computer Applications*. 2012. Vol. 57, No. 19. URL: https://www.ijcaonline.org/archives/volume57/number19/9223-3782/ (дата звернення: 14.10.2025).

[10] Compression method in terms of ensuring the fidelity of video images in infocommunication networks / V. Barannik et al. // *Radioelectronic and Computer Systems*. 2022. No. 4 (100). P. 10–24. DOI: 10.32620/reks.2022.5/09.

[11] Ficzere P. The possibilities of intelligent manufacturing methods. // *Research Gate*. 02.05.2020. URL: https://www.researchgate.net/publication/357738098_The_possibilities_of_intelligent_manufacturing_methods (дата звернення: 14.10.2025).

[12] Method of Mini Segments Encoding in Difference Space Using Haar Wavelet / V. Barannik et al. // *2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)*. Lviv, Ukraine, 2023. P. 1–4. DOI: 10.1109/AICT61584.2023.10452674.

[13] Gould S., Fulton R., Koller D. Decomposing a Scene into Geometric and Semantically Consistent Regions. // *ResearchGate*. URL: https://www.researchgate.net/publication/224135950 (дата звернення: 14.10.2025).

[14] A Method of Scrambling for the System of Cryptocompression of Codograms Service Components / V. Barannik et al. // *Emerging Networking in the Digital Transformation Age : TCSET 2022*. Cham : Springer, 2023. Vol. 965. P. 444–459. (Lecture Notes in Electrical Engineering). DOI: https://doi.org/10.1007/978-3-031-24963-1_26.

[15] Метод аналізу взаємодії параметрів QOE та QOS на основі алгоритмів керування машинами / Р. Одарченко та ін. // *Наукоємні технології*. 2022. № 4 (56). С. 305–316. DOI: https://doi.org/10.18372/2310-5461.56.17130.

[16] Krawczyk H., Bellare M., Canetti R. HMAC: Keyed-Hashing for Message Authentication. RFC 2104. 1997. URL: https://www.rfc-editor.org/rfc/rfc2104.html (дата звернення: 14.10.2025).

[17] Метод кластеризації послідовності трансформант за структурними ознаками їх спектрально-параметричного опису / В. В. Бараннік та ін. // *Наукоємні технології*. 2024. № 2 (62). С. 185–192. DOI: https://doi.org/10.18372/2310-5461.62. 18712.

[18] Ranjith B., Raghu N. CryptoGAN: a new frontier in generative adversarial network-driven image encryption // *IAES International Journal of Artificial Intelligence*. 2024. Vol. 13, No. 4. P. 4813–4821. DOI: https://doi.org/10.11591/ ijai.v13.i4.

[19] Метод стиснення кластеризованих трансформант на основі блочного кодування з локально-монотонним визначенням довжини / В. В. Бараннік та ін. // *Наукоємні технології*. 2024. № 3 (63). С. 274–281. DOI: https://doi.org/10.18372/2310-5461.63.18971.

[20] Strogatz S. H. Nonlinear Dynamics and Chaos. CRC Press, 2018, 513 p.

[21] Recommendation for Password-Based Key Derivation Part 1: Storage Applications / M. Sönmez Turan et al. // NIST Special Publication 800-132. URL: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/ nistspecialpublication800-132.pdf (дата звернення: 14.10.2025).

[22] Tang L. Methods for Encrypting and Decrypting MPEG Video Data Efficiently // Carnegie Mellon University. URL: https://dl.acm.org/doi/pdf/10.1145/244130.244209 (дата звернення: 14.10.2025).

[23] A Comparative Analysis on Blockchain versus Centralized Authentication Architectures for IoT-Enabled Smart Devices in Smart Cities: A Comprehensive Review, Recent Advances, and Future Research Directions / A. M. Usman et al. // *Sensors (MDPI)*. 2022. Vol. 22, Iss. 14. URL: https://www.mdpi.com/1424-8220/22/14/5168 (дата звернення: 14.10.2025).

[24] Критерії вибору спектрально-ефективних сигналів у бездротових інформаційних мережах / В. Козловський та ін. // *Наукоємні технології*. 2022. № 4 (56). С. 273–286. DOI: https://doi.org/10.18372/2310-5461.56.17125 (дата звернення: 14.10.2025).

[25] Xiaowu Li, Huiling Peng. Chaotic medical image encryption method using attention mechanism fusion ResNet model // *Frontiers in Neuroscience*. 2023. Vol. 17. Art. 1226154. DOI: https://doi.org/10.3389/fnins.2023.1226154.

[26] Technology of Sliding Coding of Uneven Diagonal Sequences in Two-Dimensional Spectral Space of Transformants / V. V. Barannik et al. // *Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia*. 2023. No. 94. P. 13–23. DOI: 10.20535/RADAP.2023.94.13-23.

[27] Saving Elements Methods for Service Components of Images Cryptocompression Codograms / V. V. Barannik et al. // *Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia*. 2023. No. 92. P. 28–40. DOI: 10.20535/RADAP.2023.92.28-40.

[28] Significant Microsegment Transformants Encoding Method to Increase the Availability of Video Information Resource / V. Barannik et al. // *IEEE Advanced Trends in Information Theory (ATIT) : proceedings of 2nd Intern. Conf. (Kyiv, Ukraine, November 25-27, 2020)*. Kyiv, 2020. P. 52–56. DOI: 10.1109/ATIT50783.2020.9349256.

[29] Barannik V., Shiryaev A. Quadrature compression of images in polyadic space // *Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2012) : proceedings of Intern. Conf. (Lviv-Slavske, Ukraine, February 21–24, 2012)*. Lviv-Slavske, 2012. P. 422.

[30] Development of Adaptive Arithmetic Coding Method to the Sequence of Bits / V. Barannik et al. *Engineer of the XXI Century: EngineerXXI 2018*. Cham: // Springer, 2020. Vol. 70. (Mechanisms and Machine Science). DOI: https://doi.org/10.1007/978-3-030-13321-4_18.

[31] Video Segments Stamping Method Saving Their Reliability in the Spectral-Cluster Space / V. V. Barannik et al. // *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*. 2023. No. 92. P. 41–53. DOI: 10.20535/RADAP.2023.92.41-53.

[32] A novel image encryption method based on improved two-dimensional logistic mapping and DNA computing / Y. Chen et al. // *Frontiers in Physics*. 2024. URL: https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2024.1469418/full (дата звернення: 14.10.2025).

[33] Barannik V., Khimenko V., Barannik N. Method of indirect information hiding in the process of video compression. // *Radioelectronic and Computer Systems*. 2021. No. 4. P. 119–131. DOI: https://doi.org/10.32620/reks.2021.4.

Костромицький А. І., Безрук В. М., Малінін І. Г., Панчук А. Г., Бреславець Б. О.

МЕТОД СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ВІДЕОЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ НЕЙРОННОЇ МЕРЕЖІ U-NET

У сучасному інформаційному середовищі відеоінформація стала одним із наймасовіших і найцінніших типів даних. Камери відеоспостереження, мобільні пристрої, дрони та онлайн-платформи щоденно генерують величезні обсяги візуальних даних, значна частина яких містить конфіденційну або персональну інформацію. Традиційні підходи до захисту, такі як повне шифрування відеопотоку, забезпечують високий рівень безпеки, але мають суттєві обмеження: вимагають значних обчислювальних ресурсів, збільшують затримки при передачі й не дозволяють ефективно працювати в режимі реального часу. У статті розглянуто ***науково-прикладну задачу*** *селективного захисту відеоінформації в умовах обмежених обчислювальних ресурсів, що є актуальною для систем відеоспостереження, медичних записів та інших сфер, де необхідно забезпечити конфіденційність лише критично важливих ділянок зображення.* ***Мета дослідження*** *полягає у створенні програмного інструменту для динамічного визначення та шифрування значущих фрагментів відеозображення з використанням методів семантичної сегментації та технологій машинного навчання. Запропонована система базується на архітектурі нейронної мережі U-Net, яка дозволяє точно ідентифікувати області, що потребують криптографічного захисту, і застосовувати до них селективне шифрування. В якості експериментальної основи використано датасет Stanford Background, результати тестування підтверджують ефективність обраного підходу. Представлені методи можуть бути інтегровані у практичні системи з обмеженими ресурсами та підвищеними вимогами до продуктивності й безпеки.*

Ключові слова: відеозображення, селективне шифрування, семантична сегментація, нейронна мережа U-Net, захист відеоінформації, машинне навчання, кодування, стиснення, криптографічна обробка, відеоспостереження, оптимізація ресурсів.

Kostromytskyi A., Bezruk V., Malinin I., Panchuk A., Breslavets B.

METHOD OF SEMANTIC SEGMENTATION OF VIDEO IMAGES USING U-NET NEURAL NETWORK

In today's information environment, video information has become one of the most massive and valuable types of data. Surveillance cameras, mobile devices, drones and online platforms generate huge amounts of visual data every day, a significant part of which contains confidential or personal information. Traditional approaches to protection, such as full encryption of the video stream, provide a high level of security, but have significant limitations: they require significant computing resources, increase transmission delays and do not allow for effective real-time operation. The article considers the scientific and applied problem of selective protection of video information in conditions of limited computing resources, which is relevant for video surveillance systems, medical records and other areas where it is necessary to ensure the confidentiality of only critically important areas of the image. The purpose of the research is to create a software tool for dynamic identification and encryption of significant fragments of a video image using semantic segmentation methods and machine learning technologies. The proposed system is based on the architecture of the U-Net neural network, which allows to accurately identify areas requiring cryptographic protection and apply selective encryption to them. The Stanford Background dataset was used as an experimental basis, the test results confirm the effectiveness of the chosen approach. The presented methods can be integrated into practical systems with limited resources and increased requirements for performance and security.

Keywords: video image, selective encryption, semantic segmentation, U-Net neural network, video information protection, machine learning, encoding, compression, cryptographic processing, video surveillance, resource optimization.

Стаття надійшла до редакції 24.11.2025 р.
Прийнято до друку 10.12.2025 р.