КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.28.816
УДК 004.934

Корченко Олександр Григорович
лауреат Державної премії України в галузі науки і техніки,
Заслужений діяч науки і техніки України, член-кореспондент НАН України
д.т.н, проф., перший проректор Державного університету
інформаційно-комунікаційних технологій, Київ, Україна
ORCID ID: 0000-0003-3376-0631
agkorchenko@gmail.com

Терейковський Олег Ігоревич
аспірант кафедри кібербезпеки
Національний університет «Київський авіаційний інститут», Київ, Україна
ORCID ID: 0000-0001-5045-0163
tereikovskyio@gmail.com

СЕМАНТИЧНА СЕГМЕНТАЦІЯ ЗОБРАЖЕННЯ ОБЛИЧЧЯ В СИСТЕМАХ БІОМЕТРИЧНОЇ АУТЕНТИФІКАЦІЇ ПЕРСОНАЛУ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ ЗОБРАЖЕНЬ

Анотація. Проблематикою статті є підвищення ефективності засобів біометричної аутентифікації персоналу об'єктів критичної інфраструктури. Показано, що один із основних напрямків підвищення ефективності полягає у вдосконалені процедури виділення контурів обличчя у піддослідному зображенні, результатом застосування якої у більшості відомих випадків є визначення прямокутної області, яка охоплює обличчя. Такий результат не забезпечує точного виділення контурів обличчя та завад при відеореєстрації, зокрема засобів індивідуального захисту, волосся та окулярів. Для подолання цих обмежень доцільним є використання нейромережевих засобів семантичної сегментації, які дозволяють точно виділяти контури обличчя, зону очей, а також області з перекриттями чи фоновими елементами, що значно підвищує точність розпізнавання особи в біометричних системах. Водночас результати аналізу сучасних науково-практичних рішень в області семантичної сегментації засвідчують, що більшість з них не забезпечують можливості ефективного функціонування в умовах об'єктів критичної інфраструктури, що в першу чергу пояснюється недосконалістю методологічного забезпечення. З метою подолання вказаних недоліків в статті запропоновано модель семантичної сегментації зображення обличчя, яка базується на енкодер-декодерній нейромережевій архітектурі з можливістю адаптації конструктивних параметрів до умов застосування на об'єктах критичної інфраструктури. На основі цієї моделі розроблено метод визначення архітектурних параметрів нейромережевої моделі, що передбачає послідовну оцінку умов задачі, вибір базової архітектури, налаштування параметрів енкодера, декодера та навчання, оцінювання ефективності та адаптивну модифікацію структури моделі. Метод дозволяє врахувати вплив низки факторів, зокрема просторових характеристик елементів сегментації, дисбалансу класів, варіативності освітлення, обмежень на обчислювальні ресурси та нормативних вимог. Експериментальні дослідження засвідчили, що застосування запропонованого методу дозволяє скоротити обсяг необхідних експериментів у 2 рази та досягти точності сегментації зображення обличчя на рівні 0,9, що перевищує показники існуючих аналогів приблизно на 10–20%.

Ключові слова: нейромережева модель; семантична сегментація; захист інформації; об'єкт критичної інфраструктури; безпека інформації; біометрична автентифікація; розпізнавання особи.

ВСТУП

В умовах посилення викликів у сфері національної безпеки особливу роль відіграють автоматизовані системи контролю доступу на об'єкти критичної інфраструктури. Одним із ключових компонентів таких систем є засоби біометричної аутентифікації, що забезпечують ідентифікацію особи на основі зображення обличчя. Для забезпечення високої точності та надійності ідентифікації необхідним є попередній етап — семантична сегментація зображення, що дозволяє виділити область обличчя серед інших фрагментів сцени. При цьому в реальних умовах задача семантичної сегментації значно ускладнюється внаслідок необхідності врахування низки різноманітних факторів, зокрема часткового або повного перекриття обличчя сторонніми об'єктами (засобами індивідуального захисту, окулярами, волоссям тощо), варіативності ракурсів зйомки, змін освітлення, а також різних параметрів відео- чи фотофіксації (роздільна здатність, кількість кольорових каналів тощо). Усе це зумовлює потребу у використанні засобів, здатних адаптуватися до складних та нестабільних умов середовища. В більшості відомих систем біометричної автентифікації вказані засоби базуються на нейромережевих технологіях обробки зображень, які вже продемонстрували високу ефективність у задачах комп'ютерного зору. Їх здатність виявляти складні просторові зв'язки, узагальнювати ознаки на різних рівнях абстракції та адаптуватися до нових даних робить їх перспективним інструментом для реалізації точного та надійного виділення обличчя в системах біометричної автентифікації. Разом із тим, ефективність нейромережевих засобів істотно залежить від методологічних підходів до їх розроблення, налаштування та адаптації до конкретних умов застосування.

Постановка проблеми. Вдосконалення методологічного забезпечення побудови нейромережевих засобів виділення кордонів зображення обличчя в системах біометричної автентифікації об'єктів критичної інфраструктури.

Аналіз останніх досліджень і публікацій. Як показують результати науково-практичних робіт в області біометричної автентифікації, у більшості сучасних рішень ідентифікація особи за зображенням обличчя здійснюється на основі його виділення за допомогою прямокутної рамки (bounding box), яка задається алгоритмами детекції, зокрема на основі моделей типу MTCNN, RetinaFace або BlazeFace [2], [4], [11]. Проте такий підхід має низку обмежень: по-перше, він не враховує природну форму обличчя, що знижує точність подальших етапів обробки, таких як нормалізація, вилучення ознак та їх порівняння [8]; по-друге, прямокутна локалізація не дозволяє коректно відокремити завади — наприклад, волосся, маски, рукавички, перекриття або фон, що потрапляє в межі рамки [7]. Як наслідок, такий підхід може призводити до погіршення точності розпізнавання та підвищення чутливості системи до зовнішніх факторів або спуфінгу-атак. У цьому контексті зростає інтерес до застосування нейромережевих методів семантичної сегментації, які забезпечують піксельне розділення зображення на логічні області. Це дозволяє більш точно виділяти контури обличчя та зони завад, покращуючи якість маскування фону або виключення нерелевантних ділянок [5], [12]. На відміну від bounding box-підходів, піксельна сегментація формує адаптивну маску, що зберігає форму об'єкта і підвищує достовірність подальшої ідентифікації [3]. Особливо помітними переваги семантичної сегментації є в умовах варіативного освітлення, нестандартних ракурсів або часткових перекриттів, які часто мають місце у практичних сценаріях функціонування систем біометричної автентифікації [6]. Разом з тим результати аналізу сучасних науково-практичних робіт в області семантичної сегментації обличчя [1], [13] вказують на те, що більшість існуючих рішень зосереджені

на розробці або запозиченні універсальних нейромережевих архітектур, які демонструють високу ефективність на загальних тестових наборах, але не враховують специфіку застосування в умовах функціонування систем біометричної аутентифікації на об'єктах критичної інфраструктури. Водночас, проблематика адаптації нейромережевих моделей до впливу факторів зовнішнього середовища, варіативності вхідних даних, обмеженості обчислювальних ресурсів та нормативних вимог досі залишається недостатньо формалізованою. Це вказує на актуальність створення методологічно обґрунтованих рішень, що забезпечать не лише вибір відповідної архітектури нейромережевої моделі, але й визначити її конкретні архітектурні параметри, адаптовані до реальних умов застосування.

Мета статті. Розробка методу визначення архітектурних параметрів нейромережевої моделі семантичної сегментації зображення обличчя, що базується на моделі семантичної сегментації адаптованій до умов застосування в системах біометричної аутентифікації персоналу об'єктів критичної інфраструктури.

РОЗРОБКА МОДЕЛІ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ЗОБРАЖЕННЯ ОБЛИЧЧЯ ПРИ БІОМЕТРИЧНІЙ АУТЕНТИФІКАЦІЇ НА ОБ'ЄКТАХ КРИТИЧНОЇ ІНФРАСТРУКТУРИ ІЗ ЗАСТОСУВАННЯМ НЕЙРОННИХ МЕРЕЖ

Використавши в якості прототипу авторські напрацювання в області застосування нейронних мереж для сегментації зображень [9], [10], [14] визначено, що в базовому випадку модель семантичної сегментації зображення обличчя при біометричній аутентифікації на об'єктах критичної інфраструктури можливо записати за допомогою виразів виду:

$$\widehat{\boldsymbol{P}} = f(\boldsymbol{X})\ \widehat{\boldsymbol{P}} \in \mathbb{R}^{H\times W\times K}, \boldsymbol{X} \in \mathbb{R}^{H\times W\times C}, \quad (1)$$

$$\boldsymbol{X}_{i,j} = \left|x_{i,j,1}, x_{i,j,2}, \dots, x_{i,j,C}\right|, \quad (2)$$

$$\widehat{\boldsymbol{P}}_{i,j} = \left|p_{i,j,1}, p_{i,j,2}, \dots, p_{i,j,K}\right|, \quad (3)$$

де $\boldsymbol{X}$ — вхідне зображення обличчя, що підлягає сегментації; $\widehat{\boldsymbol{P}}$ — карта ймовірностей піксель-класів; f — функція сегментації ($f\colon \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{H\times W\times K}$); $\mathbb{R}$ — множина дійсних чисел; H, W — висота та ширина зображення; C — кількість кольорових каналів в зображенні; K — кількість класів сегментації; $x_{i,j,c}$ — яскравість c-го кольорового каналу пікселя з координатами (i,j); $p_{i,j,k}$ — ймовірність того, що піксель з координатами (i,j) належить до k-го класу.

Зазначимо, що в поширених форматах представлення зображень значення яскравості кожного із кольорових каналів зберігаються у вигляді цілих чисел. Наприклад, при використанні формату RGB достатньо часто передбачається яскравість кожного із пікселів записувати за допомогою 8-бітного числа. Це означає, що $\boldsymbol{X} \in \mathbb{Z}^{H\times W\times C}$, де $\mathbb{Z}$ — множина цілих чисел, а яскравість кожного із кольорових каналів довільного пікселю є цілим числом від 0 до 255. Однак де-факто стандартна процедура підготовки зображення перед подачею на вхід нейромережевої моделі передбачає масштабування яскравості кожного із кольорових каналів до діапазону $[0,1]$, за рахунок ділення початкового значення яскравості кожного із пікселів на число, що відповідає максимально можливому значенню яскравості, наприклад, на 255 для 8-бітних зображень. Тому у виразі (1) $\boldsymbol{X} \in \mathbb{R}^{H\times W\times C}$.

Враховуючи, що, відповідно до результатів [1], при семантичній сегментації доцільно використовувати нейромережеву модель, архітектура якої передбачає використання енкодеру та декодеру, вираз (1) можливо модифікувати так:

$$\widehat{\boldsymbol{P}} = f_D\big(f_E(\boldsymbol{X})\big), \tag{4}$$

де $f_E(\cdot)$ — функція, що описує результат застосування нейромережевого енкодера, $f_D(\cdot)$ — функція, що описує результат застосування нейромережевого декодера.

Використавши результати [1], [10], [14], визначено, що в базовому випадку функціонал нейромережевого енкодера та нейромережевого декодера можливо описати за допомогою виразів виду (5) та (6), відповідно. Зазначимо, що відповідно до результатів [10], [14] в якості базису енкодера та декодера доцільно використовувати такі типи нейромережевих моделей, як VGG, ResNet, MobileNet, EfficientNet, HRNet. При цьому вважається, що найбільш апробовані типи енкодера та декодера базуються на VGG-моделях та використовуються в U-Net-подібних нейромережевих моделях семантичної сегментації з можливими пропусками шарів субдискретизації та варіативною кількістю шарів згортки.

$$\boldsymbol{Z} = f_E(\boldsymbol{X}), \boldsymbol{Z} \in \mathbb{R}^{H' \times W' \times D}, \tag{5}$$

$$\widehat{\boldsymbol{P}} = f_D(\boldsymbol{Z}), \widehat{\boldsymbol{P}} \in \mathbb{R}^{H \times W \times K}, \tag{6}$$

де $\boldsymbol{Z}$ — тензор ознак закодованого зображення; $H' \times W'$ — розмір карт згортки/субдискретизації на виході із енкодеру; D — кількість карт згортки/субдискретизації на виході з енкодеру (глибина ознакового простору).

Деталізований опис процедури функціонування енкодера (5) можливо здійснити використовуючи вирази (7, 8), а для деталізації функціонування декодера (6) – вирази (9,10).

$$\boldsymbol{X}_l^{c_l} = \sigma_l\big(W_l^{c_l} * \boldsymbol{X}_{l-1}^{c_{l-1}} + b_l\big), l = 0,1, \dots, L, \tag{7}$$

$$\boldsymbol{X}_l^{s_l} = Pool(\boldsymbol{X}_l^{c_l}), \tag{8}$$

де $\boldsymbol{X}_l^{c_l}$ — двовимірна матриця ознак на виході із c-ої карти згортки в l-му згортковому шарі ($\boldsymbol{X}_0^{c_0} = X$); L – кількість шарів згортки в енкодері; $\boldsymbol{X}_l^{s_l}$ — двовимірна матриця ознак на виході із s-ої карти субдискретизації де шару згортки в енкодері; $W_l^{c_l}, b_l$ — вагові коефіцієнти, що відповідають c-ої карти згортки в l-му згортковому шарі; $Pool(\cdot)$ — операція субдискретизації; σ_l — функція активації в l-му згортковому шарі енкодера.

$$\boldsymbol{Z}_{n-1}^{c_{n-1}} = \sigma\big(W_n^{c_n} * Dec(\boldsymbol{Z}_n^{c_n}) + b_n\big), n = 0,1, \dots, N, \tag{9}$$

$$\boldsymbol{Z}_{n-1}^{c_{n-1}} = \sigma\big(W_n^{c_n} * Dec\big(\boldsymbol{Z}_n^{c_n}\big) \parallel \boldsymbol{X}_{l-1}^{c_{l-1}} + b_n\big), \tag{10}$$

де $\boldsymbol{Z}_n^{c_n}$ — двовимірна матриця ознак, що відповідає c-ій карті згортки в n-му згортковому шарі декодера ($\boldsymbol{Z}_0^{c_0} = \boldsymbol{X}_L^{c_L}$); N — кількість шарів згортки в декодері; $\parallel$ — конкатенація (skip-connection) вагових коефіцієнтів з відповідного рівня енкодера.

При цьому, відповідно до [1], [13], термін апсемплінг описує процедуру, що реалізується за рахунок збільшення просторової роздільності тензора ознак, яка виконується в декодері нейронної мережі з метою покрокового відновлення вихідного розміру зображення після даунсемплингу, тобто стиснення внаслідок проходження вхідного зображення через енкодер.

Процедуру апсемплінгу можливо описати за допомогою наступного виразу:

$$Z_{up}(h,w,d) = \sum_{i=0}^{k_h-1} \sum_{j=0}^{k_w-1} Z\left(trunc\left(\frac{h}{d}\right) - i, trunc\left(\frac{w}{d}\right) - j\right) \cdot K(i,j), \tag{12}$$

де Z_{up} – результат апсемплінгу; h, w – висота та ширина вхідного тензора ознак Z; d – масштабний коефіцієнт апсемплінгу; $trunc(\cdot)$ – функція визначення найближчого найменшого цілого; $K(i,j)$ — ядро апсемплінгу.

Параметри ядра апсемплінгу залежать від використаного методу інтерполяції і визначаються за допомогою виразів виду:

$$K(i,j) = \delta(i) \cdot \delta(j), \tag{13}$$

$$K(i,j) = \max(0{,}1-|\,i\,|) \cdot \max(0{,}1-|\,j\,|), \tag{14}$$

$$K(i,j) = \mathrm{R(i)} \cdot \mathrm{R(j)} \tag{15}$$

$$\mathrm{R(x)} = \begin{cases} (a+2)\,|\,x\,|^3 - (a+3)\,|\,x\,|^2 + 1, 0 \le |\,x\,| < 1 \\ a\,|\,x\,|^3 - 5a\,|\,x\,|^2 + 8a\,|\,x\,| - 4a, 1 \le |\,x\,| < 2, \\ 0, |\,x\,| \ge 2 \end{cases} \tag{16}$$

$$K(i,j) = F_l(NM), \tag{17}$$

де δ – дельта-функція Кронекера; a – фіксований параметр (a = −0,5); $F_l(NM)$ – функція, що співвідноситься з визначенням $K(i,j)$ в результаті навчання нейромережевої моделі.

Вираз (13) використовується при застосуванні методу інтерполяції типу «найближчого сусіда», вираз (14) — при двовимірній лінійній інтерполяції; вирази (15, 16) — при бікубічній інтерполяції, а вираз (17) — при визначенні ядра апсемплінгу в результаті навчання нейромережевої моделі.

Інтеграція виразів (5–17), з урахуванням особливостей функціонування декодеру на основі згорткової нейронної мережі, дозволяє записати результуючий вираз для розрахунку карти ймовірностей піксель-класів у наступному вигляді:

$$\widehat{\boldsymbol{P}} = \sigma\left(W_{out} \cdot Dec\left(f_E(\boldsymbol{X})\right) + b_{out}\right), \tag{18}$$

де W_{out}, b_{out} — вагові коефіцієнти останнього згорткового шару нейромережевого декодера; $Dec(\cdot)$ — функція, що описує реалізацію процедури апсемплінгу в нейромережевому декодері.

У найбільш поширеному випадку застосування у нейронах вихідного шару декодеру функції активації типу Softmax, ймовірність того, що піксель з координатами (i,j) належить до k-го класу, можливо розрахувати так:

$$P_{i,j,k} = e^{a_{i,j,k}} \Big/ \sum_{m=1}^{K} e^{a_{i,j,m}}, \tag{19}$$

де $\boldsymbol{X}_l^{c_l}$ — двовимірна матриця ознак на виході із c-ої карти згортки в l-му згортковому шарі ($\boldsymbol{X}_0^{c_0} = X$).

Також зазначимо, що специфіка задачі розпізнавання особи представника персоналу об’єкта критичної інфраструктури вказує на необхідність отримання результату семантичної сегментації не лише у вигляді карти ймовірностей піксель-класів, а й у вигляді сегментованого зображення обличчя, на якому відображені його природні кордони, кордони очей та кордони завад. Враховуючи [10], [14], визначення такого сегментованого зображення може бути реалізоване за рахунок віднесення кожного з пікселів цього зображення до класу з найбільшою ймовірністю. Тобто

$$y_{i,j} = argmax_k\left(p_{i,j,k}\right), k \in [1,K,1], p_{i,j,k} \in \widehat{\boldsymbol{P}}, \tag{20}$$

$$\boldsymbol{Y} = \left|y_{i,j}\right|_{i=1,j=1}^{H,W}, \tag{21}$$

де $y_{i,j}$ — клас, до якого відноситься піксель зображення з координатами (i,j); $p_{i,j,k}$ — ймовірність віднесення пікселя з координатами (i,j) до k-го класу; $\boldsymbol{Y}$ — сегментоване зображення.

Враховуючи результати [9], [10], [14], аналіз розробленого математичного забезпечення (1–21) нейромережевої моделі семантичної сегментації зображення обличчя представника персоналу об’єкту критичної інфраструктури, побудованої на базі

U-Net-подібної архітектури, вказує на те, що до переліку конструктивних параметрів цієї моделі, які забезпечують можливість її адаптації до очікуваних умов застосування, відносяться: кількість шарів згортки та кількість карт згортки, розмір та крок ядра згортки в кожному із шарів для енкодеру та декодеру; кількість та локалізація шарів субдискретизації енкодеру; кількість та локалізацію шарів апсемплінгу; розмір ядра, масштабний коефіцієнт та метод інтерполяції для кожного із шарів апсемплінгу; кількість skip-зв'язків, локалізація входу в енкодері та вставки в декодер кожного із skip-зв'язків; кількість рівнів агрегації ознак skip-зв'язків; варіант агрегації skip-зв'язків.

Крім того, висновки науково-практичних робіт [1], [10] дозволяють стверджувати, що в переліку конструктивних параметрів слід врахувати особливості реалізації процедури навчання: вибір типу оптимізатора (наприклад, Adam, SGD), розмір міні-пакету (batch size), швидкість навчання (learning rate), стратегію зміни швидкості навчання (learning rate scheduling), умови зупинки навчання (early stopping, кількість епох), методи регуляризації (dropout, batch normalization тощо), а також вид функції втрат. Зазначимо, що функція втрат Categorical Cross-Entropy визначається виразом (22), Binary Cross-Entropy — (23), Dice Loss — (24, 25), Jaccard Loss — (26, 27), Focal Loss — (29), Tversky Loss — (29, 30).

$$\mathcal{L}_{CCE} = -\sum_{i=1}^{N}\sum_{c=1}^{C} g_{i,c}\log\left(p_{i,c}\right), \tag{22}$$

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\left(g_{i,c}\cdot\log(p_{i,c}) + (1-g_{i,c})\cdot\log\left(1-p_{i,c}\right)\right), \tag{23}$$

$$\mathcal{L}_{Dice} = 1 - \frac{2\cdot|TP|}{|P|+|G|}, \tag{24}$$

$$\mathcal{L}_{Dice} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{i=1}^{N} p_{i,c}g_{i,c} + \varepsilon}{\sum_{i}^{N} p_{i,c} + \sum_{i}^{N} g_{i,c} + \varepsilon}, \tag{25}$$

$$\mathcal{L}_{IoU} = 1 - \frac{|TP|}{|P\cup G|}, \tag{26}$$

$$\mathcal{L}_{IoU} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{\sum_{i=1}^{N} p_{i,c}g_{i,c} + \varepsilon}{\sum_{i}^{N} p_{i,c} + \sum_{i}^{N} g_{i,c} - \sum_{i=1}^{N} p_{i,c}g_{i,c} + \varepsilon}, \tag{27}$$

$$\mathcal{L}_{Focal} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\alpha^{c}\left(1-p_{i,c}\right)^{\gamma} g_{i,c}\log\left(p_{i,c}\right), \tag{28}$$

$$\mathcal{L}_{Tversky} = 1 - \frac{TP}{TP + \alpha\cdot FP + \beta\cdot FN}, \tag{29}$$

$$\mathcal{L}_{Tversky} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{\sum_{i=1}^{N} p_{i,c}g_{i,c} + \varepsilon}{\sum_{i}^{N} p_{i,c}g_{i,c} + \alpha\sum_{i}^{N} p_{i,c}\left(1-g_{i,c}\right) + \beta\sum_{i=1}^{N}\left(1-p_{i,c}\right)g_{i,c} + \varepsilon}, \tag{30}$$

де N — кількість пікселів; C – кількість класів; P — отримана маска сегментації; G — істинна маска сегментації; $TP = P\cap G$; $FP = P\backslash G$; $FN = G\backslash P$; $p_{i,c}$ — передбачена ймовірність належності пікселя i до класу c; $g_{i,c}$ — істинна мітка; ε — коефіцієнт, що використовується для уникнення ділення на 0; α^{c} — ваговий коефіцієнт для класу c; α — ваговий коефіцієнт FP; β — ваговий коефіцієнт FN.

Зазначимо, що у випадку використання в якості базису нейромережевої моделі семантичної сегментації альтернативних архітектур (ResNet, VGG, MobileNet,

EfficientNet, HRNet) перелік конструктивних параметрів додатково розширюється за рахунок типу енкодера, параметрів модулів багатомасштабного контексту, attention-блоків, стратегій об'єднання багаторівневих ознак і методів семантичної постобробки. Так для енкодеру в якому використовуються Residual-блоки характерні для нейронної мережі ResNet, а функціонування якого описується за допомогою виразів (31, 32), до переліку конструктивних параметрів доцільно віднести: кількість стадій, кількість блоків у кожній із стадій, параметри згорток для кожного блоку та параметри проектуючої згортки.

$$R_{i,j}(X_l) = \sigma\left(X_l + \mathcal{F}\left(X_l; W_{l,j}\right)\right), l \in \{1,2,\dots,L\}, \quad (31)$$

$$X_{l+1} \to W_{proj} * X_l, \quad (32)$$

де $R_{i,j}(\cdot)$ — j-й Residual-блок на l-ій стадії; X_l — вхідний тензор l-ої стадії; L — кількість стадій; $\mathcal{F}\left(X; W_{l,j}\right)$ — послідовність згорток у Residual-блоці; $W_{l,j}$ — набір параметрів згорткових шарів; W_{proj} — набір параметрів згортки для приведення розмірності X_l до розмірності X_{l+1}.

В підсумку вирази (1–32) складають базис математичного апарату моделі семантичної сегментації, що за рахунок застосування енкодер-декодерної архітектури нейронної мережі з варіативними конструктивними параметрами, що підлягають адаптації залежно від умов застосування, забезпечує можливість розробки ефективного методу визначення архітектурних параметрів нейромережевої моделі семантичної сегментації зображення обличчя при біометричній аутентифікації на об'єктах критичної інфраструктури.

РОЗРОБКА МЕТОДУ ВИЗНАЧЕННЯ АРХІТЕКТУРНИХ ПАРАМЕТРІВ НЕЙРОМЕРЕЖЕВОЇ МОДЕЛІ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ЗОБРАЖЕННЯ ОБЛИЧЧЯ ПРИ БІОМЕТРИЧНІЙ АУТЕНТИФІКАЦІЇ НА ОБ'ЄКТАХ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Базуючись на запропонованій моделі семантичної сегментації (1–32) та враховуючи, що підготовку даних необхідних для навчання нейромережевої моделі доцільно реалізувати за допомогою окремих процедур, запропоновано співвіднести завдання методу з визначенням архітектурних параметрів нейромережевої моделі, адаптованих до умов застосування та нормативних вимог. Використавши в якості прототипу описаний в [9] метод семантичної сегментації зображень за допомогою нейронних мереж виконання методу визначення архітектурних параметрів нейромережевої моделі семантичної сегментації зображення обличчя при біометричній аутентифікації на об'єктах критичної інфраструктури запропоновано розділити на 9 етапів.

Етап 1. Визначення умов застосування.

Крок 1.1. Визначення обмежень. На основі експертного оцінювання умов поставленої задачі біометричної аутентифікації, нормативних вимог до системи біометричної аутентифікації та опису розробленої моделі (1–32) визначається кортеж параметрів $\langle L_G \rangle$, що характеризують: розмір вхідного зображення та кількість кольорових каналів $(W \times H, K_c)$; кількість класів сегментації (K_{ss}); просторові характеристики елементів сегментації (розмір елементів сегментації відносно вхідного зображення та допустимість розмитості меж); загальна кількість навчальних прикладів (N_Σ); кількість навчальних прикладів для кожного із елементів сегментації (N_{i_ss}); коефіцієнт дисбалансу навчальної вибірки (I_R); коефіцієнт відносного розміру маски об'єкта (Δ_{obj}); вид та доступний обсяг обчислювальних ресурсів (V_{CR}, V_{CR}^d); варіативність

освітлення та ракурсу відеореєстрації зображення обличчя після попередньої обробки; допустимий термін процесу семантичної сегментації (Δ_{Tss}); вид та мінімально допустиме значення показника точності сегментації (A_{ss}, Δ_A); допустима кількість епох навчання (Δ_{TL}); коефіцієнт відхилення положення кордонів виділеного об'єкту від істинних кордонів (Δ_{Br}).

Крок 1.2. Визначення базових архітектур. На основі експертного оцінювання визначається множина доступних базових нейромережевих архітектур ($\{AN\}$). В першому наближенні до вказаної множини віднесено VGG, ResNet, MobileNet, EfficientNet, HRNet.

Крок 1.3. Визначення параметрів архітектур. На основі експертного оцінювання з урахуванням результатів отриманих при розробці моделі семантичної сегментації зображення обличчя при біометричній аутентифікації на об'єктах критичної інфраструктури із застосуванням нейронних мереж (1–32) визначаються конструктивні параметри кожної архітектури, що входить до складу $\{AN\}$ та діапазон можливих змін цих конструктивних параметрів, що визначається множиною мінімальних ($\{AN_{min}\}$) та максимальних значень ($\{AN_{max}\}$).

На вхід етапу подається наведений в технічному завданні опис поставленої задачі семантичної сегментації, нормативні вимоги до системи біометричної аутентифікації. Передбачено реалізовувати процедури експертного оцінювання на основі методів наведених в [9], [13]. Виходом етапу являється кортеж значень параметрів, що стосуються умов застосування та характеризують: обмеження, які стосуються засобів сегментації, множину доступних базових архітектур, множини конструктивних параметрів кожної із базових архітектур — $S_1 = \langle\langle L_G\rangle, \{AN\}, \langle R_{AN}\rangle\rangle$

Етап 2. Вибір базової архітектури. На вхід етапу подається кортеж S_1, визначений в результаті реалізації етапу 1.

Крок 2.1. Вибір допустимих архітектур. Використовуючи результати [9], [14] пропонується реалізувати вибір переліку допустимих архітектур з урахуванням наступних умов:

- Обмеженість обчислювальних ресурсів та терміну сегментації (обсяг доступної пам'яті менший ніж 4 ГБ, допустимий термін сегментації — 100 мс): MobileNet, EfficientNet.
- Для зображень в форматі RGB розміром до 256×256: VGG (U-Net або U-Net++), MobileNet.
- За необхідності сегментувати невеликі об'єкти (площа об'єкту займає менше ніж 0,1 площі вхідного зображення) та при високому дисбалансі прикладів об'єктів (коефіцієнт дисбалансу більший ніж 0,1), що підлягають сегментації у навчальній вибірці: VGG із механізмом уваги, VGG у складі U-Net++, HRNet.
- Якщо кількість класів сегментації не перевищує 3, то MobileNet, VGG (U-Net), в протилежному випадку VGG (U-Net++ або з механізмом уваги), ResNet, HRNet
- При високій варіативності освітлення та ракурсу відеореєстрації зображення обличчя після попередньої обробки (діапазони освітлення та ракурсу відеореєстрації перевищують порогові значення визначені в [14], в першому наближенні 0,5): VGG із механізмом уваги, ResNet (U-Net++), EfficientNet, HRNet.
- За необхідності досягнення високої точності (IoU > 0.85): VGG (U-Net++ або з механізмом уваги), ResNet, HRNet.

Крок 2.2. Фіксація вибору базової архітектури. На даному кроці реалізується фіксація вибору архітектури, що буде використовуватись в якості базової. У випадку, коли реалізація попереднього кроку призводить до вибору декількох архітектур, то в якості базової (A_{NB}) обирається та архітектура програмно-апаратна реалізація якої потребує менших ресурсів. Також відповідно до визначеного типу, визначається множина архітектурних параметрів ($\langle R_{ANB} \rangle$).

Вихід етапу 2 визначається виразом — $S_2 = \langle A_{NB}, \langle R_{ANB} \rangle \rangle$.

Етап 3. Визначення параметрів енкодера. На вхід етапу подаються $\langle L_G \rangle$, S_2, Δ_A, A_{ss}^r — різниця між величиною допустимого та величиною досягнутого показника точності, що визначається в результаті виконання етапу 6 та t_e — номер етапу визначення параметрів енкодера. На першій ітерації, тобто при $t_e = 1$, $A_{ss}^r = 0$.

Крок 3.1. Ініціалізація параметрів енкодера. Крок виконується у випадку $A_{ss}^r = 0$ при $t_e = 1$. Величини конструктивних параметрів енкодера (5, 7, 8, 31, 32) встановлюються рівними значенням відповідних елементів із множини $\{AN_{max}\}$ для $\langle R_{ANB} \rangle$, тобто $EN_{te}(i) = AN_{max}$, де i — номер конструктивного параметру.

Крок 3.2. Модифікація параметрів енкодера. Крок виконується у випадку $A_{ss}^r > 0$. При цьому глибина енкодера (кількість згорткових шарів/кількість стеків) змінюється відповідно до виразів (33, 34).

$$EN_{te}(i) = EN_{t-1}(i) - 1, i = z \quad (33)$$

$$if\ EN_{te}(i) < EN_{min}(i) \Longrightarrow EN_{te}(i) = AN_{min}(i), \quad (34)$$

z — номер конструктивного параметру, що відповідає глибині енкодера.

Інші параметри енкодера модифікуються з позиції збереження сумісності вихідного сигналу енкодера з вхідним сигналом декодера, визначеним при $t_e - 1$. Виходом етапу 3 є множина значень конструктивних параметрів енкодера, $S_3 = \{EN_{te}\}$.

Етап 4. Визначення параметрів декодера. На вхід етапу подаються $\langle L_G \rangle$, S_3, S_2, Δ_A, A_{ss}^r та t_d — номер етапу визначення параметрів декодера. На першій ітерації, тобто при $t_d = 1$, $A_{ss}^r = 0$.

Крок 4.1. Ініціалізація параметрів декодера. Крок виконується у випадку $A_{ss}^r = 0$ при $t_d = 1$. Величини конструктивних параметрів декодера (6, 9-21, 31, 32) встановлюються рівними значенням відповідних конструктивних параметрів енкодера, тобто $DN_{td}(i) = EN_{te}(i)$, де i — номер конструктивного параметру. При цьому використовується максимальна кількість skip-зв'язків та attention-модулі, за їх наявності в базовій архітектурі.

Крок 4.2. Модифікація параметрів декодера. Крок виконується у випадку $A_{ss}^r > 0$.

Модифікація полягає в тому, що починаючи з найглибшого рівня декодера, який відповідає шару з найменшим розміром карти ознак, у відповідності до (35, 36) реалізується послідовне видалення attention-модулів, skip-зв'язків та шарів апсемплінгу.

$$DN_{td}(i) = DN_{td-1}(i) - 1, \quad (35)$$

$$if\ DN_{td}(i) < AN_{min}(i) \Longrightarrow DN_{td}(i) = AN_{min}(i), \quad (36)$$

Інші параметри декодера модифікуються з позиції збереження сумісності вихідного сигналу енкодера з вхідним сигналом декодера, визначеним при $(t_d - 1)$. Виходом етапу 4 є множина значень конструктивних параметрів декодера, $S_4 = \{DN_{td}\}$.

Етап 5. Навчання нейромережевої моделі.

На вхід етапу подаються $\langle L_G \rangle$, S_4, S_3, S_2, Δ_A, A_{ss}^r та t_d — номер етапу визначення параметрів декодера. На першій ітерації проведення навчання, тобто при $t_l = 1$, $A_{ss}^r = 0$.

Крок 5.1. Ініціалізація параметрів навчання, що регулюють процес визначення вагових коефіцієнтів синаптичних зв'язків при навчанні нейромережевої моделі. Крок

виконується у випадку $A_{ss}^{r} = 0$ при $t_l = 1$. Ініціалізація регламентується за допомогою наступних правил.

Правило вибору функції втрат. Вибір реалізується із переліку функцій втрат, визначених виразами (23–30).

- При $K_{ss} = 1$. Якщо $I_R \geq 0.2$ використовується Binary Cross-Entropy, якщо $I_R < 0.2$ та $\Delta_{obj} \leq 0{,}1$ — Focal Loss, а в іншому випадку Tversky Loss.
- При $1 < K_{ss} \leq 3$. Якщо $I_R \geq 0.2$ та $\Delta_{Br} \leq 0{,}05$ - Jaccard Loss, а якщо $\Delta_{Br} > 0{,}05$ — Dice Loss. Якщо $I_R < 0.2$ — Tversky Loss.
- При $K_{ss} > 3$. Якщо $I_R \geq 0.2$ або $\Delta_{obj} \leq 0{,}01$ — Tversky Loss, а в інших випадках Dice Loss.
- При недостатній інформації або суперечливих вимогах — Dice Loss.

Правило визначення початкового значення швидкості навчання.

- Для нейромережевих моделей на базі MobileNet або EfficientNet-B0, для яких кількість вагових коефіцієнтів менша ніж 5×10^6, learning rate = 0.001.
- Для нейромережевих моделей на базі ResNet, HRNet та моделей типу Attention U-Net, U-Net++, Attention U-Net++ для яких кількість вагових коефіцієнтів більша ніж 5×10^6 learning rate = 0.0005.

Правило вибору стратегії зміни швидкості навчання.

- Якщо $T_L < 50$, то використовується стратегія Cosine Annealing.
- Якщо $N_\Sigma \leq 10^3$, то використовується стратегія Step Decay із зменшенням у 10 разів кожні 10 епох.
- В інших випадках використовується стратегія ReduceLROnPlateau з урахуванням того, що якщо значення функції втрат на валідаційній вибірці не покращується більше ніж на 0,001 протягом 5 послідовних епох, то поточне значення learning rate зменшується в 2 рази.

Правило вибору розміру батчу.

- Якщо $V_{CR}^{d} \leq 4 \times 10^9$, то batch size = 4.
- Якщо $4 \times 10^9 < V_{CR}^{d} \leq 8 \times 10^9$, то batch size = 8.
- Якщо $8 \times 10^9 < V_{CR}^{d}$, то batch size = 16.

Правило визначення кількості навчальних епох.

- Якщо $N_\Sigma \leq 10^3$ або середня кількість прикладів кожного класу менша ніж 200, або $I_R > 20$, то використовується механізм ранньої зупинки — навчання зупиняється, якщо функція втрат на валідаційній вибірці не покращується більше ніж на 0,001 протягом 5 послідовних епох. В інших випадках кількість епох навчання $T_L = 50$.
- Якщо $\Delta_{TL} < T_L$, то $T_L = \Delta_{TL}$.

Правило регуляризації.

- Якщо $N_\Sigma \leq 10^4$, то використовується $Dropout = 0{,}3$ та L2-регуляризація з $weight\ decay = 10^{-4}$.
- Для нейромережевих моделей на базі MobileNet або EfficientNet-B0 для яких кількість вагових коефіцієнтів менша ніж 5×10^6 використовується L2-регуляризація з $weight\ decay = 10^{-5}$ без $Dropout$.
- Для нейромережевих моделей на базі ResNet, HRNet та моделей типу Attention U-Net, U-Net++, Attention U-Net++ для яких кількість вагових коефіцієнтів більша ніж 5×10^6 використовується $Dropout = 0{,}5$ та L2-регуляризація з $weight\ decay = 10^{-5}$.

Правило аугментації навчальних даних.

– Якщо в навчальній вибірці представлені зображення обличчя зареєстровані при варіативних умовах, то використовуються: Random brightness/contrast ±15%; Random affine (±15° поворот, ±10% масштаб); Horizontal flip (p=0.5).
– Якщо в навчальній вибірці представлені стандартизовані зображення обличчя, то використовуються: Random crop (до 90% від розміру); Gaussian noise (σ < 0.01).

Крок 5.2. Модифікація параметрів навчання. Крок виконується у випадку $A_{ss}^{r} > 0$. Модифікація реалізується на основі наступних правил:

– Якщо виявлені ознаки перенавчання, що співвідносяться з зростанням після 20 епохи навчання різниці між метрикою оцінки ефективності на тестовій та валідаційній вибірках (наприклад $\Delta_{IoU} > 0.2$), то насамперед у 2 рази зменшується learning rate за умови, що $\text{learning rate} \geq 10^{-5}$. У разі недостатнього ефекту weight decay встановлюється рівним 10^{-3}, а Dropout збільшується на 0,1 (до межі 0.5).
– Якщо виявлені ознаки недостатньої тривалості навчання, що співвідноситься з низькими значеннями A_{ss} на тестовій та валідаційній вибірках, то насамперед на 20% збільшується кількість епох навчання. У разі недостатнього ефекту: learning rate збільшується на 25%, за умови, що $\text{learning rate} \leq 10^{-2}$. При відсутності ознак перенавчання додатково Dropout зменшується до 0.1, а weight decay — до 10^{-5}.
– Якщо на валідаційній вибірці A_{ss} не змінюється на протязі 5 епох активується стратегія ReduceLROnPlateau зі значеннями factor = 0.5, patience = 3.

Крок 5.3. Проведення навчання. Навчання нейромережевої моделі здійснюється з урахуванням параметрів навчання визначених на кроках 5.1 і 5.2 та супроводжується розрахунком показників ефективності (крок 1.1) на тренувальній та валідаційній вибірках на кожній із епох навчання.

Виходом кроку 5.3 та етапу 5 в цілому являються: значення показників ефективності на кожній епосі навчання ($|A_{ss}|$), та значення вагових коефіцієнтів синаптичних зв'язків, розраховані в результаті навчання ($|W_l|$).

Етап 6. Налаштування архітектурних параметрів. На вхід етапу подаються: $|W_l|$, $\langle L_G \rangle$, S_4, S_3, S_2, Δ_A, A_{ss}^{r}, t_d.

Крок 6.1. Розрахунок значень параметрів ефективності — зводиться до розрахунку A_{ss}, використаного обсягу обчислювальних ресурсів — V_{CR} та T_{ss} — тривалості семантичної сегментації при застосуванні навченої нейромережевої моделі для сегментації зображень із тестової вибірки, яка не використовувалась в процесі навчання.

Крок 6.2. Оцінювання ефективності — зводиться до порівняння розрахованих значень параметрів ефективності з допустимими значеннями.

Крок 6.3. Адаптація архітектури до результатів оцінювання — полягає у визначенні управлінських рішень щодо необхідності модифікації архітектурних параметрів нейромережевої моделі в залежності від поточних значень архітектурних параметрів та результатів оцінювання, отриманих в результаті виконання кроку 6.2. Процедура визначення регламентується за допомогою наступних правил.

– Якщо точність сегментації не відповідає заданій ($A_{ss}^{r} > 0$), а поточні значення параметрів навчання не досягли граничних значень, то реалізується перехід на крок 5.2, виконання якого пов'язане з модифікацією параметрів навчання. В протилежному випадку, реалізується перехід на крок 2.2, що забезпечує зміну типу базової нейромережевої моделі. У випадку, коли всі архітектури з множини допустимих базових архітектур досліджено, а жодна з них не

дозволила досягти заданих значень показників ефективності, приймається рішення про неможливість побудови нейромережевої моделі, яка в заданих умовах застосування забезпечує ефективну семантичну сегментацію зображення обличчя.

– Якщо точність сегментації відповідає заданій ($A_{ss}^{r} \leq 0$), а поточні значення архітектурних параметрів декодера знаходяться в допустимих межах, то реалізується перехід на крок 4.2, виконання якого призводить до модифікації параметрів декодера. У випадку, коли всі можливі комбінації значень архітектурних параметрів декодера досліджено, реалізується перехід на крок 3.2, пов'язаний з модифікацією параметрів енкодера за умови, що параметри енкодера знаходяться в допустимих межах. Дослідження завершується після визначення архітектурних параметрів декодера та енкодера, що за мінімально можливої обчислювальної ресурсоємності, забезпечують достатню точність сегментації зображення обличчя представника персоналу об'єкта критичної інфраструктури.

Виходом кроку 6.3, етапу 6 та методу в цілому являється кортеж параметрів, що характеризують архітектуру (S_2, S_3, S_4, V_{CR}), вагові коефіцієнти синаптичних зв'язів ($|W_l|$) та результати навчання (A_{ss}, T_L, T_{ss}) нейромережевої моделі семантичної сегментації зображення обличчя представника персоналу об'єкту критичної інфраструктури.

Проведені оціночні експерименти спрямовані на верифікацію запропонованого методу дозволяють вважати, що його застосування дозволяє приблизно в 2 рази зменшити обсяг експериментальних досліджень, спрямованих на визначення архітектурних параметрів нейромережевої моделі, яка забезпечує точність семантичної сегментації зображення обличчя представника персоналу об'єкту критичної інфраструктури на рівні 0,9, що приблизно в 1,1–1,2 рази перевищує точність найкращих відомих засобів аналогічного призначення.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

В результаті проведених досліджень вперше розроблена модель семантичної сегментації зображення обличчя при біометричній аутентифікації персоналу на об'єктах критичної інфраструктури, що за рахунок застосування енкодер-декодерної архітектури нейронної мережі, адаптованої до умов виділення контурів об'єктів, забезпечує можливість розробки ефективного методу побудови нейромережевих засобів семантичної сегментації зображення обличчя при біометричній аутентифікації. З використанням запропонованої моделі вперше розроблено метод визначення архітектурних параметрів нейромережевої моделі семантичної сегментації зображення обличчя при біометричній аутентифікації на об'єктах критичної інфраструктури, що за рахунок адаптації нейромережевого кодера та нейромережевого декодера до розміру зображення, кількості каналів кольору, допустимої мінімальної точність сегментації, допустимої максимальної обчислювальної складності реалізації процесу сегментації, характеристик навчальної вибірки, необхідності виділення декількох сегментів, що відповідають різним деформованим об'єктам, які частково перекриваються між собою, просторових характеристик елементів сегментації, показників освітлення та ракурсу відеореєстрації, допустимої максимальної обчислювальної складності та допустимого терміну навчання нейромережевої моделі дозволяє приблизно в 2 рази зменшити обсяг експериментальних досліджень, спрямованих на визначення архітектурних параметрів нейромережевої

моделі, яка забезпечує точність семантичної сегментації зображення обличчя представника персоналу об’єкту критичної інфраструктури на рівні 0,9, що приблизно в 1,1–1,2 рази перевищує точність найкращих відомих засобів аналогічного призначення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39(12),* 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
2. Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019). Blazeface: Sub-millisecond neural face detection on mobile gpus. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*. https://doi.org/10.48550/arxiv.1907.05047
3. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence, 40(4),* 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
4. Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5203–5212.
5. Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1925–1934. https://doi.org/10.48550/arXiv.1611.06612
6. Liu, S., Shi, J., Liang, J., & Yang, M. H. (2017). *Face parsing via recurrent propagation. BMVC.* https://doi.org/10.48550/arXiv.1708.01936
7. Ranjan, R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., Castillo, C. & Chellappa, R. (2019). A Fast and Accurate System for Face Detection, Identification, and Verification. *In IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(2),* 82–96. https://doi.org/10.1109/TBIOM.2019.2908436
8. Schroff, F., Kalenichenko, D. & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 815–823. https://doi.org/10.1109/CVPR.2015.7298682
9. Tereikovskyi, I., Hu, Z., Chernyshev, D., Tereikovska, L., Korystin, O., & Tereikovskyi, O. (2022). The method of semantic image segmentation using neural networks. *International Journal of Image, Graphics and Signal Processing, 10(6), 1, 14(6),* 1–14. https://doi.org/10.5815/ijigsp.2022.06.01
10. Tereikovskyi, I., Korchenko, O., Bushuyev, S., Tereikovskyi, O., Ziubina, R. & Veselska, O. (2023). A Neural Network Model for Object Mask Detection in Medical Images. *International Journal of Electronics and Telecommunication, 69(1),* 41–46. https://doi.org/10.24425/ijet.2023.144329
11. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE signal processing letters, 23(10),* 1499–1503. https://doi.org/10.1109/LSP.2016.2603342
12. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* 6230–6239. https://doi.org/10.1109/CVPR.2017.660
13. Korchenko O., & Tereikovskyi O. (2023). Analysis and evaluation of biometric authentication means based on the image of the face and iris of the staff of critical infrastructure facilities. *Cybersecurity: Education, Science, Technique. 1(21),* 136–148. https://doi.org/10.28925/2663-4023.2023.21.136148
14. Korchenko, O., Tereikovskyi, O. (2024). Modular neural network model for biometric authentication of the personnel of critical infrastructure facilities using facial images and iris. *Information security, 30(2),* 339–347. https://doi.org/10.18372/2225-5036.30.19247

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 - 4023

Oleksandr Korchenko
laureate of the State Prize of Ukraine in the field of Science and Technology,
Corresponding Member of the National Academy of Sciences of Ukraine,
Honored Worker of Science and Technology of Ukraine,
D.Sc., Prof. First vice-rector, State University of Information and
Communication Technologies, Kyiv, Ukraine
ORCID ID: 0000-0003-3376-0631
agkorchenko@gmail.com

Oleh Tereikovskyi
Postgraduate student of the Department of Cybersecutiry,
The National University "Kyiv Aviation Institute", Kyiv, Ukraine
ORCID ID: 0000-0001-5045-0163
tereikovskyio@gmail.com

SEMANTIC SEGMENTATION OF FACIAL IMAGES IN BIOMETRIC AUTHENTICATION SYSTEMS OF PERSONNEL OF CRITICAL INFRASTRUCTURE FACILITIES

Abstract. The problem of the article is to increase the efficiency of biometric authentication of personnel of critical infrastructure facilities. It is shown that one of the main directions of increasing efficiency is to improve the procedure for highlighting facial contours in the test image, the result of which in most known cases is the definition of a rectangular area covering the face. Such a result does not provide accurate highlighting of facial contours and interference during video recording, in particular personal protective equipment, hair and glasses. To overcome these limitations, it is advisable to use neural network semantic segmentation tools, which allow you to accurately highlight facial contours, the eye area, as well as areas with overlaps or background elements, which significantly increases the accuracy of face recognition in biometric systems. At the same time, the results of the analysis of modern scientific and practical solutions in the field of semantic segmentation show that most of them do not provide the possibility of effective functioning in the conditions of critical infrastructure facilities, which is primarily explained by the imperfection of methodological support. In order to overcome the above-mentioned shortcomings, the article proposes a model of semantic segmentation of facial images, which is based on an encoder-decoder neural network architecture with the ability to adapt design parameters to the conditions of application on critical infrastructure objects. Based on this model, a method for determining the architectural parameters of a neural network model has been developed, which involves a sequential assessment of the task conditions, selection of the basic architecture, adjustment of the encoder, decoder and training parameters, evaluation of efficiency and adaptive modification of the model structure. The method allows taking into account the influence of a number of factors, in particular, the spatial characteristics of the segmentation elements, class imbalance, lighting variability, limitations on computing resources and regulatory requirements. Experimental studies have shown that the use of the proposed method allows reducing the volume of necessary experiments by 2 times and achieving facial image segmentation accuracy at the level of 0.9, which exceeds the indicators of existing analogues by approximately 10–20%.

Keywords: neural network model; semantic segmentation; information protection; critical infrastructure facility; information security; biometric authentication; person recognition.

REFERENCES

15. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39(12),* 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
16. Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019). Blazeface: Sub-millisecond neural face detection on mobile gpus. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*. https://doi.org/10.48550/arxiv.1907.05047

17. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence, 40(4),* 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
18. Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,* 5203–5212.
19. Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* 1925–1934. https://doi.org/10.48550/arXiv.1611.06612
20. Liu, S., Shi, J., Liang, J., & Yang, M. H. (2017). *Face parsing via recurrent propagation. BMVC.* https://doi.org/10.48550/arXiv.1708.01936
21. Ranjan, R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., Castillo, C. & Chellappa, R. (2019). A Fast and Accurate System for Face Detection, Identification, and Verification. *In IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(2),* 82–96. https://doi.org/10.1109/TBIOM.2019.2908436
22. Schroff, F., Kalenichenko, D. & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 815–823. https://doi.org/10.1109/CVPR.2015.7298682
23. Tereikovskyi, I., Hu, Z., Chernyshev, D., Tereikovska, L., Korystin, O., & Tereikovskyi, O. (2022). The method of semantic image segmentation using neural networks. *International Journal of Image, Graphics and Signal Processing, 10(6), 1, 14(6),* 1–14. https://doi.org/10.5815/ijigsp.2022.06.01
24. Tereikovskyi, I., Korchenko, O., Bushuyev, S., Tereikovskyi, O., Ziubina, R. & Veselska, O. (2023). A Neural Network Model for Object Mask Detection in Medical Images. *International Journal of Electronics and Telecommunication, 69(1),* 41–46. https://doi.org/10.24425/ijet.2023.144329
25. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE signal processing letters, 23(10),* 1499–1503. https://doi.org/10.1109/LSP.2016.2603342
26. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* 6230–6239. https://doi.org/10.1109/CVPR.2017.660
27. Korchenko O., & Tereikovskyi O. (2023). Analysis and evaluation of biometric authentication means based on the image of the face and iris of the staff of critical infrastructure facilities. *Cybersecurity: Education, Science, Technique. 1(21),* 136–148. https://doi.org/10.28925/2663-4023.2023.21.136148
28. Korchenko, O., Tereikovskyi, O. (2024). Modular neural network model for biometric authentication of the personnel of critical infrastructure facilities using facial images and iris. *Information security, 30(2),* 339–347. https://doi.org/10.18372/2225-5036.30.19247