DOI: 10.58254/viti.8.2025.17.213

УДК 621.3:623.4:629.07–044.4

д-р філософії Фесенко О. Д. ORCID: 0000-0002-2114-5327 (ВІТІ ім. Героїв Крут)
Макаренко К. Л. ORCID: 0009-0002-2441-0648 (ВІТІ ім. Героїв Крут)
Дімітров П. Є. ORCID: 0009-0007-6184-8223 (НДІ ВР)

МЕТОДИКА ІДЕНТИФІКАЦІЇ БІОМЕТРИЧНИХ ДАНИХ ОБЛИЧЧЯ НА ОСНОВІ ВДОСКОНАЛЕНОГО НЕЙРОМЕРЕЖЕВОГО АЛГОРИТМУ ADAPTIVE FACE RECOGNITION IN THE WILD UNDER HEAVY NOISE

У роботі запропоновано методику AFRW-X, яка ґрунтується на ідеї інтеграції адаптивної дифузійної фільтрації, реалізованої на двох рівнях абстракції, з механізмом псевдо-3D проєкцій для забезпечення ефективного мультиракурсного злиття векторних представлень.

Ключовими компонентами запропонованої архітектури виступають: каскадний механізм відновлення зображень, механізм анізотропної дифузії Перона – Маліка та композитна функція втрат.

Каскадний механізм характеризується автоматичним вибором стратегії залежно від типу деградації (ROF-TV для помірного шуму, PnP-ADMM для змішаних деградацій та DPS для екстремальних умов).

Механізм анізотропної дифузії застосовується до карт глибоких ознак з віртуальних ракурсів і використовує адаптивне зважування на основі трикомпонентної метрики (видимість, IoU, якість ознак).

Композитна функція втрат, у свою чергу, інтегрує механізм балансування класів.

Експериментальна валідація на чотирьох еталонних датасетах підтвердила суттєву перевагу запропонованого методу в умовах множинної деградації вхідних зображень. Так, при критичному рівні адитивного гаусівського шуму, метод AFRW-X (modified) забезпечує відносну точність розпізнавання при вкрай низькому рівні хибних спрацювань, що перевищує показники базового ArcFace та оригінальної версії AFRW. Ключовим показником є незначне відносне падіння точності за умов сильного шуму, що задовольняє встановлену вимогу про допустиме зниження та підтверджує високу робастність методики.

У випадку комплексної деградації, що поєднує всі негативні фактори, запропоноване рішення досягає точності, що складає 88,5 % розпізнавання при низькому рівні хибних спрацювань та високого показника інтегральної метрики 78,5 %, що значно перевищує показники існуючих методів.

Наукова новизна полягає у синтезі адаптивної каскадної фільтрації, удосконаленого модуля псевдо-3D проєкцій та композитної функції втрат для одночасної компенсації множинних факторів деградації.

Такий підхід забезпечує комплексну компенсацію множинної деградації вхідних біометричних даних обличчя, на відміну від існуючих методів, що розглядають окремі негативні фактори ізольовано.

Ключові слова: *розпізнавання облич, ідентифікація особи, OSINT, дифузійна фільтрація, псевдо-3D проєкції, глибоке навчання, згорткові нейронні мережі, біометрична верифікація, адаптивна увага, мультиракурсне злиття.*

O. Fesenko, K. Makarenko, P. Dimitrov. Method of identification of individuals based on an in-depth neural face algorithm adaptive Face Recognition in the wild under heavy noise

The robot is based on the AFRW-X technique, which is based on the idea of integrating adaptive diffusion filtration, implemented on two levels of abstraction, with a pseudo-3D mechanism projection to ensure effective multi-angle vector presentation.

The key components of the proposed architecture are: the cascade mechanism of image renewal, the Peron–Malik anisotropic diffusion mechanism, and the composite waste function.

The cascade mechanism is characterized by automatic selection of strategy depending on the type of degradation (ROF-TV for moderate noise, PnP-ADMM for mixed degradations and DPS for extreme minds).

The anisotropic diffusion mechanism is based on deep character maps from virtual angles and vicoristic adaptive value based on three-component metrics (visibility, IoU, character brightness).

The composite spending function, in turn, integrates the class balancing mechanism.

Experimental validation on four standard datasets confirmed the superiority of the proposed method in the minds of multiple degradation of input images. Thus, with a critical level of additive Gaussian noise, the AFRW–X (modified) method will ensure excellent recognition accuracy at an extremely low level of variable processing, which exceeds the indicators of the basic ArcFace and the original AFRW versions. The key indicator is an insignificant drop in accuracy compared to strong noise, which satisfies the established possibility of an acceptable reduction and confirms the high robustness of the technique.

In the context of complex degradation, which consumes all negative factors, a solution has been developed that achieves an accuracy of 88.5 %, recognition with a low level of agricultural products and high performance The integral metric is 78.5 %, which significantly exceeds the results of other methods.

Scientific novelty lies in the synthesis of adaptive cascade filtration, an advanced pseudo-3D projection module and a composite cost function for one-hour compensation of multiple degradation factors.

This approach will ensure comprehensive compensation for the multiple degradation of input biometric data, in addition to other methods that look at other negative factors isolated.

Keywords: *face recognition, individual identification, OSINT, diffusion filtration, pseudo-3D projections, deep learning, hypothalamic neural measures, biometric verification, adaptive respect, multi-angle anger.*

Загальна постановка задачі. Сучасні системи ідентифікації та верифікації особи на основі біометричних даних обличчя стрімко розвиваються у напрямку підвищення точності розпізнавання, інваріантності до умов та розширення функціональних можливостей роботи в умовах апріорної невизначеності (відсутність завчасної інформації про якість вхідних зображень, ракурс зйомки, рівень шуму та освітлення). Особливої актуальності набуває задача ідентифікації особи у відкритих джерелах Open Source Intelligence (OSINT) при роботі з зображеннями низької якості, різними ракурсами зйомки та наявністю значного рівня шуму в умовах реального застосування.

Одним із найбільш перспективних рішень у цій галузі є використання глибоких згорткових нейронних мереж, які здатні автоматично виявляти інваріантні ознаки обличчя та формувати компактні векторні представлення у метричному просторі. Сучасні алгоритми, такі як DeepFace, FaceNet, ArcFace та інші, продемонстрували високу ефективність розпізнавання обличь на якісних датасетах із контрольованими умовами зйомки, досягаючи точності розпізнавання понад 99 % на еталонних наборах даних типу LFW (Labeled Faces in the Wild) [1].

Однак у випадку пошуку та ідентифікації особи у відкритих джерелах, що містять зображення низької якості (Low-Quality Images, LQI), задача ускладнюється низкою наступних факторів: значна варіативність просторового розташування обличчя (ракурс, масштаб, часткове перекриття), деградація якості зображення (низька роздільна здатність, артефакти, розмиття руху), вплив шумових завад, нерівномірне освітлення та наявність тіней, а також неконтрольовані умови зйомки (різні фотокамери, алгоритми постобробки).

Враховуючи вище наведене, з'являється необхідність розробки методу індентифікації особи, що забезпечують стійкість до деградації вхідних даних.

Аналіз останніх досліджень і публікацій

Фундаментальні дослідження в галузі розпізнавання обличь заклали основу для сучасних підходів. Колектив дослідників Taigman та ін. [2] у 2014 році представив алгоритм DeepFace, що став одним з перших методів, який досяг точності розпізнавання, близької до людської, що склало 97,35 % на датасеті LFW. Ключовою особливістю архітектури DeepFace є застосування етапу 3D-фронталізації обличчя за допомогою афінних перетворень, що дозволяє нормалізувати вхідні зображення до фронтального ракурсу перед поданням на вхід згорткової нейронної мережі. Архітектура включає послідовність згорткових рівнів та локально зв'язаних рівнів, що використовують унікальні вагові коефіцієнти для кожної просторової позиції, на відміну від стандартних згорткових рівнів зі спільними вагами. Загальна кількість параметрів моделі становить приблизно 120 мільйонів, а обчислювальна складність однієї ітерації навчання досягає $O(N \cdot 10^9)$ операцій.

Подальший розвиток методів розпізнавання обличь пов'язаний з роботою Schroff та ін. [3], які у 2015 році запропонували архітектуру FaceNet. Принципова відмінність FaceNet полягає у використанні триплетної функції втрат, що безпосередньо оптимізує вбудований простір ознак шляхом мінімізації відстані між зображеннями одної особи та максимізації відстані між зображеннями різних осіб з урахуванням маржі. Така функція втрат дозволяє навчити мережу формувати компактні кластери векторних представлень для кожної ідентичності у нормалізованому евклідовому просторі з L2-нормалізацією. Ключовою інновацією є застосування стратегії *semi-hard negative mining* для вибору інформативних

триплетів під час навчання. FaceNet досяг рекордної на той час точності 99,63 % на датасеті LFW при використанні лише 128-байтного представлення обличчя.

Подальше вдосконалення методів метричного навчання для розпізнавання обличь представлене в роботі Deng et al. [4], які у 2019 році запропонували алгоритм ArcFace (Additive Angular Margin Loss). Основна ідея полягає у введенні кутової маржі у косинусну метрику класифікації. На відміну від попередніх підходів, ArcFace показує більш стабільну конвергенцію під час навчання та не потребує додаткової регуляризації функцій активації. Експериментальна валідація на десяти еталонних датасетах показала, що ArcFace послідовно перевершує існуючі методи, досягаючи точності понад 99,8 % при мінімальних додаткових обчислювальних витратах.

Так, колектив дослідників Wang та ін. (2024) [5] представили архітектуру Quality-Aware Face Recognition (QAFR), побудовану на основі механізму уваги, суть якого полягає у процесі адаптивного зважування різних просторових регіонів обличчя залежно від локальної якості зображення. Застосування багаторівневої архітектури з проміжними модулями оцінки якості дозволило підвищити точність ідентифікації на 12–15 % порівняно з базовими моделями ArcFace при роботі із зображеннями роздільною здатністю менше 64×64 пікселів.

Робота Zhang та ін. (2023) [6] присвячена розробці методу Cross-Resolution Face Recognition (CRFR), що використовує dual-path архітектуру для одночасної обробки зображень різної роздільної здатності. Основна ідея полягає у формуванні окремих гілок екстракції ознак для високоякісних та низькоякісних зображень із подальшою проєкцією в єдиний метричний простір із використанням механізму знань дистиляції. Експериментальні результати показали зменшення похибки ідентифікації на 18–22 % при співставленні зображень із різницею роздільної здатності більше ніж у 4 рази.

Особливої уваги заслуговує дослідження Chen та ін. (2024) [7], що запропонували архітектуру Noise-Robust Face Embedding Network (NRFEN) із вбудованими модулями денойзингу на різних рівнях згорткової мережі. Використання self-supervised learning підходу для попереднього навчання денойзуючих модулів на синтетично зашумлених даних дозволило забезпечити підвищення стійкості до адитивного гаусівського шуму з рівнем σ до 50, що критично важливо при роботі із зображеннями з відкритих джерел низької якості.

У контексті роботи з диференційними ракурсами, Kumar та ін. (2023) [8] розробили метод Multi-View Pose-Invariant Face Recognition (MVPIFR), що базується на використанні 3D Morphable Models (3DMM) для синтезу фронтальних зображень обличчя з довільних ракурсів. Процес навчання мережі включає етап адаптивної корекції положення об’єкта з використанням триплетної функції втрат та додатковою регуляризацією на основі перцептуальних метрик. Результати експериментів на датасеті Multi-PIE показали покращення точності розпізнавання на 25–30 % для екстремальних ракурсів (±90°) порівняно з методами без корекції пози.

Дослідження Li та ін. (2024) [9] присвячене розробці Unified Quality Enhancement and Recognition Framework (UQERF), що інтегрує модулі, денойзингу та екстракції ознак в єдину end-to-end архітектуру. Ключовою особливістю є використання багатозадачного навчання зі спільною оптимізацією функцій втрат для підвищення якості зображення та підвищення диференціації вхідних зображень.

У роботі Martinez (2023) [10] запропоновано метод Adaptive Face Recognition in the Wild (AFRW), що використовує механізм Meta-Learning для швидкої адаптації до специфічних умов деградації вхідних зображень. Архітектура включає базову мережу екстракції ознак та мета-мережу, що генерує адаптивні параметри для модулів нормалізації залежно від оцінки якості вхідного зображення. Експериментальна валідація на датасетах IJB-C та MegaFace продемонструвала підвищення робастності системи до різних типів деградації з покращенням метрики True Acceptance Rate (TAR) при False Acceptance Rate (FAR) = 0,01 % на 8–12 %.

Важливим напрямком досліджень є застосування трансформерних архітектур для задачі розпізнавання обличь в умовах низької якості. Zhao (2024) [11] розробили Quality-Aware Vision Transformer (QA-ViT), що використовує механізми самоуваги для адаптивного зважування важливості різних патчів зображення залежно від їх локальної якості. Порівняно з CNN-базованими архітектурами, QA-ViT продемонстрував кращу здатність до моделювання довгострокових залежностей та більш ефективну агрегацію інформації з різних просторових регіонів.

Однак аналіз існуючих досліджень показує, що більшість запропонованих методів орієнтовані на роботу з контрольованими умовами деградації або окремими типами спотворень (наприклад, тільки низька роздільна здатність або тільки шум). Також необхідно зазначити, що при роботі з реальними зображеннями з відкритих джерел спостерігається одночасний вплив множинних факторів деградації [12; 13]: комбінація низької роздільної здатності, різних типів шуму, компресійних артефактів, нерівномірного освітлення та екстремальних ракурсів.

Крім того, існуючі підходи часто вимагають значних обчислювальних ресурсів, що ускладнює їх застосування в режимі реального часу із використанням великих баз даних відкритих джерел.

На основі проведеного аналізу літератури, для детального дослідження особливостей роботи в умовах низької якості зображень та порівняння із запропонованим рішенням було обрано чотири фундаментальні алгоритми розпізнавання обличь: DeepFace, FaceNet та ArcFace Adaptive Face Recognition in the Wild. Вибір саме цих алгоритмів обґрунтовується наступними факторами: 1) DeepFace представляє класичний підхід з явною фронталізацією та багатошаровою архітектурою з локально зв'язаними рівнями, що демонструє ефективність роботи з різними ракурсами; 2) FaceNet застосовує триплетну функцію втрат для прямої оптимізації метричного простору, що є альтернативним підходом до класифікаційних методів та забезпечує компактні векторні представлення; 3) ArcFace представляє сучасний state-of-the-art підхід з кутовою маржею, що демонструє найвищу точність на еталонних датасетах та має чітку геометричну інтерпретацію; 4) Adaptive Face Recognition in the Wild поєднує механізми координатної уваги та адаптивної модуляції з модифікованими функціями втрат для підвищення точності розпізнавання обличь у неконтрольованих умовах.

Таким чином**, метою дослідження** є розробка методу ідентифікації особи у відкритих джерелах на основі зображень низької якості з урахуванням диференційного ракурсу та в умовах деградації вхідних зображень.

Виклад основного матеріалу дослідження. Розглянемо систему біометричної верифікації обличь, що функціонує в умовах впливу стохастичних деградацій зображень. Для формалізації задачі визначимо вихідні дані системи наступним чином.

Нехай маємо вхідне зображення обличчя $I \in R^{(H \times W \times 3)}$ низької якості, отримане з відкритих джерел, де H, W – висота та ширина зображення відповідно. Вхідне зображення I_{in} є результатом впливу невідомої комбінації стохастичних деградацій $g \in G$ на оригінальне зображення обличчя високої якості.

Простір деградації зображення представлено у вигляді множини параметрів $G = \{R_I, \sigma_n, \beta, Q, \theta_{yaw}, \theta_{pitch}, \theta_{roll}\}$, що включає наступні типи:

1. Роздільна здатність (міжзінична відстань Inter-Pupillary Distance, IPD): $R_I \in [R_{min}, R_{max}]$, де $R_{min} = 32$ пікселі, $R_{max} = 128$ пікселів. Міжзінична відстань визначається як відстань між центрами очей на зображенні, виміряна в пікселях [18].

2. Адитивний Гаусівський шум: $\sigma_n \in [0, \sigma_n, max]$, де $\sigma_n, max = 50$. Інтенсивність шуму σ_n визначає стандартне відхилення білого Гаусівського шуму $N(0, \sigma_n^2)$, що додається до кожного пікселя зображення незалежно.

3. Параметр розмиття (blur): $\beta \in [0, \beta_{max}]$ де $\beta_{\max} = 5$ пікселів. Параметр β визначає радіус ядра Гаусівського фільтра для моделювання ефектів дефокусування або руху камери.

4. JPEG–стиснення: $Q \in [Q_{min}, 100]$, де $Q_{\min} = 10$. Параметр якості Q визначає коефіцієнт стиснення JPEG, де $Q = 100$ відповідає максимальній якості, а $Q = 10$ – екстремально високому рівню стиснення з видимими артефактами.

5. Параметр кутів повороту (pose variation): $(\theta_{yaw}, \theta_{pitch}, \theta_{roll})$, задається наступними значеннями:

– $\theta_{yaw} \in [-90°, +90°]$ – кут повороту голови відносно вертикальної осі (вліво – право);
– $\theta_{pitch} \in [-45°, +45°]$ – кут нахилу голови відносно горизонтальної осі (вгору – вниз);
– $\theta_{roll} \in [-30°, +30°]$ – кут обертання голови відносно осі погляду.

Модель деградації зображення описується композиційним виразом:

$$I_{in} = D(I_{orig};\ g), g \in G,$$

де $D(\cdot;\ g)$ – оператор деградації, що послідовно застосовує трансформації згідно з вектором параметрів g;

I_{orig} – оригінальне зображення обличчя без деградацій.

Задано також еталонна база даних відкритих джерел:

$$D = \{(I_i^{ref}, id_i)\}_{\{i=1\}}^{N},$$

де I_i^{ref} – еталонні зображення особи;

id_i – унікальний ідентифікатор i-ої особи;

N – загальна кількість осіб в базі даних.

Поріг прийняття рішення $\tau \in \mathbb{R}$ задається як скалярне значення τ, що визначає границю між класами «збіг» та «незбіг» у задачі верифікації. Поріг калібрується на валідаційній множині $D_{dev} = \{(I_j^1, I_j^2, y_j)\}_{j=1}^{M_{dev}}$, де I_j^1, I_j^2 – пара зображень для порівняння, $y_j \in \{0, 1\}$ – мітка істинності ($y_j = 1$ для $genuine$ пари, $y_j = 0$ для $impostor$ пари), M_{dev} – розмірність валідаційної множини.

Поріг $\tau*$ визначається з умови забезпечення заданого рівня False Accept Rate (FAR):

$$\tau* = \arg\max_{\tau}\{\tau :\ FAR(\tau;\ D_{dev}) \leq\ FAR_{target}\},$$

де FAR_{target} – цільовий рівень помилкового прийняття, типово $FAR_{target} \in \{0.1\%, 0.01\%, 0.001\%\}$.

Нижче наведено математична формалізація постановки задачі.

Дано:

1. Вхідне зображення низької якості I_{query} з невідомим ідентифікатором.
2. Еталонна база даних D з N осіб.
3. Граничне значення відстані для верифікації τ.
4. Параметри моделі $\theta = \{\theta_{enh}, \theta_{enc}\}$, де θ_{enh} – параметри модуля підвищення якості, θ_{enc} – параметри енкодера ознак.

Необхідно знайти:

Ідентифікатор особи $id* = \arg\min_{\{id_i \in D\}}$, де $d(f(I_{query}, \theta),\ f(I_i^{ref}, \theta))$, якщо $\min d < \tau$, інакше "особа не ідентифікована", де $f(\cdot, \theta)$: $R^{H \times W \times 3} \to R^d$ – функція перетворення зображення в d-вимірний вектор ознак (embedding), $d(\cdot,\cdot)$ – функція відстані в метричному просторі (косинусна відстань).

Допущення та обмеження (Ω)

Система біометричної верифікації повинна функціонувати в рамках наступних технічних, якісних та операційних обмежень, що визначають простір допустимих рішень Ω.

1. Час обробки одного вхідного зображення має задовольняти умові:

$$t_{proc(\theta)} \leq t_{max},$$

де $t_{proc(\theta)} = t_{enh(\theta_{enh})} + t_{enc(\theta_{enc})}$ – загальний час обробки, що включає час відновлення t_{enh} та час екстракції ознак t_{enc}, t_{max} – максимально допустимий час обробки, як правило $t_{\max} \approx 5$ секунд для серверних застосувань, $t_{\max} \approx 100$ мс для мобільних пристроїв реального часу.

2. Система гарантує функціонування тільки для зображень із міжзенічною відстанню не менше критичного значення:

$$R_I \geq R_{\min} = 32 \text{ пікселів.}$$

Для $R_I < R_{min}$ системи може суттєво деградувати і не гарантується дотримання заданих показників якості. Типові значення R_{min} лежать від архітектури моделі та варіюються в діапазоні [24; 48] пікселів [6].

3. Система повинна функціонувати при рівні Гаусівського шуму:

$$\sigma_n \leq \sigma_n, oper = 50,$$

де σ_n,$oper$ – операційна границя шуму, при перевищенні якої зображення вважається непридатним для надійної верифікації.

4. При рівні Гаусівського шуму $\sigma_n = 30$ падіння точності відносно еталоних даних не повинно перевищувати:

$$\Delta TAR_\sigma = 30 = TAR(\theta;\ \tau*, g_{clean}) - TAR(\theta;\ \tau*, g_\sigma = 30) \leq \Delta TAR_\sigma, max,$$

де $g_\sigma = 30$ – деградація з σ_n = 30, інші параметри відповідають g_{clean}, $\Delta TAR_\sigma, \max = 0.10$ (10 %) – максимально допустиме падіння точності.

Вищенаведене обмеження гарантує, що:

$$TAR(\theta;\ \tau*, g_\sigma = 30) \geq TAR(\theta;\ \tau*, g_{clean}) - 0.10.$$

При $TAR(\theta;\ \tau*, g_{clean}) = 0.85$, що означає $TAR(\theta;\ \tau*, g_\sigma = 30) \geq 0.75$.

5. Обмеження деградації точності при нефронтальних ракурсах**.** При кутах повороту голови θ_{yaw} = ±45° (профільні ракурси) падіння точності не повинно перевищувати:

$$\Delta TAR_\theta = \pm 45° = TAR\left(\theta;\ \tau *, g_{frontal}\right) - TAR(\theta;\ \tau *, g_\theta = \pm 45°) \leq \Delta TAR_\theta, max,$$

де $g_{frontal}$ – фронтальна орієнтація $\left(\left|\theta_{yaw}\right| \leq 15°\right), g_\theta = \pm 45°$ – профільна орієнтація з $\theta_{yaw} \in \{-45°, +45°\}$, $\Delta TAR_\theta, \max = 0.15\ (15\%)$ – максимально допустиме падіння точності для профільних ракурсів.

Цільова функція

Метою є знаходження оптимального набору параметрів моделі $\theta * = \{\theta_{enh} *, \theta_{enc} *\}$, за умов дотримання всіх обмежень множини Ω:

$$\theta^* = \arg\max_{\theta \in \Theta} \{\overline{TAR}(\theta)\},$$

де θ^* – оптимальні параметри моделі, що максимізують точність верифікації особи із урахуванням диференційного ракурсу та в умовах деградації вхідних зображень.

Запропоноване рішення AFRW–X (modified). Запропонована методика побудована на основі AFRW-X (Adaptive Face Recognition in the Wild), що складається з п'яти послідовних етапів обробки мінібатчу $B = (I_i, y_i)_{i=1}^{|B|} \subset D,$ де $I_i \in R^{(H \times W \times 3)}$ – вхідне зображення обличчя, $y_i \in \{1, \ldots, C\}$ – мітка класу (ідентифікатор особи).

Для розв'язання задачі ідентифікації особи у відкритих джерелах на основі зображень низької якості з урахуванням диференційного ракурсу та шумових завад запропоновано методику, що інтегрує два ключові модулі: 1-й модуль адаптивної дифузійної фільтрації (Adaptive Diffusion Filtering, ADF) для робастної обробки зображень та ознак, та 2-й модуль псевдо-3D проєкції (Pseudo-3D Projections, PTP) для мультиракурсного злиття векторних представлень даних.

Базова архітектура використовує глибоку згорткову нейронну мережу $F_{BB}(I;\ \Theta) \rightarrow X^l \in R^{(H_l \times W_l \times C_l)}$ із вбудованим механізмом координатної уваги (Coordinate Attention, CA), що забезпечує позиційно-селективну модуляцію ознак при обчислювальній складності $O(H \cdot W \cdot C)$. Нижче наведено загальний алгоритм у вигляді псевдокоду в лістингу 1.

Лістинг 1. Псевдокод алгоритму Adaptive Face Recognition in the Wild under Heavy Noise

Code

 # Algorithm 1 AFRW–X (ADF / PTP): Adaptive Face Recognition in the Wild under Heavy Noise (mod)

Input:

- Dataset D = {(Ii, yi)} – face images and class labels
- Backbone with Coordinate Attention (CA); Adaptive Attention Modulator (AAM)
- V – number of virtual views; T – diffusion iterations; η – diffusion step; κ – edge threshold
- Loss weights: λEQ, λSC, λID, λMV, λLM; β – center update step; λ – center regularizer
- βCB – class–balanced prior parameter (effective–number), τ, mmax – quality weighting

Output:

• Updated weights Θ and class centers $\{cj\}$; aggregated embeddings f* for inference

1: for each minibatch B = {(Ii, yi)} ⊂ D do

2: # === Stage 1: Noise–Aware Image Restoration (ADF–I) ===

3: for each i in B do
4: θi ← EstimateNoise(Ii) ▷ noise type/level (AWGN/Poisson/codec–mix; σ, κ)
5: if θi.type == AWGN and θi.σ ≤ σ0 then
6: Ĩi ← ROF_TV_Denoise(Ii, λTV) ▷ solve $min_u 0.5\left||u - Ii|\right|^2$ + λTV·TV(u)
7: else if θi.type ∈ {moderate mix} then
8: Ĩi ← RED_or_PnP(Ii; Dσ, μ, ρ, nADMM) ▷ Plug–and–Play/RED with denoiser $D\sigma$ as a proximal map
9: else
10: Ĩi ← DPS(Ii; {εt}, nDPS) ▷ Diffusion Posterior Sampling (posterior–consistent restoration)
11: end if
12: LID(i) ← 1 − cos(eT(Ĩi), eT(Ii)) ▷ identity preservation with a fixed ArcFace teacher
13: end for

14: *# === Stage 2: Pseudo–3D Geometry / View Synthesis (PTP) ===*
15: for each i in B do
16: (Si, Ri, ti) ← SingleView3D(Ĩi) ▷ 3DDFA–V2 / PRNet / FaceMesh
17: I_i^0 ← RenderCanonical(Si, Rcanon, tcanon) ▷ frontalized view
18: for v = 1..V do
19: (Rv, tv) ← SmallPose(Ri, ti, Δyaw_v, Δpitch_v)
20: I_i^v ← RenderView(Si, Rv, tv) ▷ texture from Ĩi / UV map
21: end for
22: end for

23: *# === Stage 3: Feature Extraction and Attention–Guided Diffusion (ADF–F) ===*
24: for each i in B do
25: for all v ∈ {0..V} do
26: Xearly ← Backbone_CA_early(I_i^v) ▷ early backbone stages with CA
27: (Fch, Fsp) ← AAM(Xearly); As := σ(Conv([Avg_c(Xearly), Max_c(Xearly)]))
28: X ← Xearly
29: for t = 1..T do ▷ Perona–Malik diffusion with attention gating
30: ∇X ← SobelGrad(X); G ← As ⊙ g(‖∇X‖), with $g(s) = \frac{1}{1+\left(\frac{s}{\kappa}\right)^2}$ or $\exp\left(-\left(\frac{s}{\kappa}\right)^2\right)$
31: X ← X + η · div(G ⊙ ∇X)
32: end for
33: Xdeep ← Backbone_CA_late(X) ▷ remaining backbone stages
34: fv ← Head_Embed(Xdeep) ∈ $\mathbb{R}^d$
35: qv ← ‖fv‖ ▷ quality proxy from feature norm
36: visv ← VisibilityFromZbuffer(I_i^v) ▷ visible–area ratio
37: iouv ← IoU(Det(I^v_i), FaceMask(I_i^v))
38: end for
39: πv ← softmax(α·visv + β·iouv + γ·qv), v=0..V
40: f*_{i} ← $\Sigma_{-}v = 0^V$πv · fv ▷ multi–view aggregated embedding
41: (logits_i, p_i) ← ArcFaceLike($f*_{\{i\}}, m\left(q = \left|\left|f*_{\{i\}}\right|\right|\right)$) ▷ adaptive margin/weight by quality
42: end for

43: *# === Stage 4: Losses & Weight Update ===*
44: LEQ ← 0; LSC ← 0; LMV ← 0; LLM ← 0
45: for each i in B do
46: ωCB(yi) ← (1−βCB) / (1 − βCB^{n_{yi}}) ▷ class–balanced weight (effective number)
47: wQ(i) ← clamp(τ·‖f*_{i}‖, mmax) ▷ quality–aware weight

LEQ ← LEQ − wQ(i)·ωCB(yi)·log(p_i[yi]) ▷ equilibrium / class–balanced CE
xi ← PreLogitFeature(Xdeep) ▷ feature used by center–loss
αi ← σ(Conv(AAM_maps_i)) ∈ [0,1] ▷ sparsity/attention weight
LSC ← LSC + 0.5 · αi · || xi − c_{yi} || ² ▷ Sparse Center Loss
LMV ← LMV + $(1/(V(V+1)))\ \cdot\ \Sigma\_\{u<v\}\ (1\ -\cos(\ f_{\{i,u\}}\})\}=,f\_\{i,v\}\ ))$
LLM ← LLM + LandmarkReprojLoss(Ĩi, Si, {lmk_i})
end for
LID ← Σ_{i∈B} LID(i)
L ← λEQ·(LEQ/|B|) + λSC·(LSC/|B|) + λID·(LID/|B|) + λMV·(LMV/|B|) + λLM·(LLM/|B|)
Θ ← Optimizer.step(∇_Θ L) ▷ backprop through all modules

 # === *Stage 5: Sparse / Attention–Weighted Center Update* ===
for each class j do
num ← Σ_{i∈B: yi=j} αi · (cj − xi)
den ← Σ_{i∈B: yi=j} + λ
Δcj ← num / den
cj ← cj − β · Δcj
end for
end for

Return Θ, {cj}

На вхід алгоритму подається датасет $D = (I_i, y_i)$, базова мережа F_{BB} із вбудованими механізмами кординатної увагм (CA) та адаптивний механізм уваги (AAM), параметри: V – кількість віртуальних ракурсів; T – кількість ітерацій дифузії; η – крок дифузії; κ – поріг збереження ознак; вагові коефіцієнти функції втрат $\{\lambda_{EQ}, \lambda_{SC}, \lambda_{ID}, \lambda_{MV}, \lambda_{LM}\}$, λ_{EQ} – коефіцієнт втрат рівноваги (Equilibrium Loss), що регулює пріоритет основної задачі класифікації з урахуванням балансування класів та якості зразків; λ_{SC} – коефіцієнт втрат розрідженого центрування (Sparse Center Loss), для контролю розмірності векторних представлень у просторі ознак; λ_{ID} – коефіцієнт втрат збереження ідентичності, що штрафує модуль відновлення (ADF-I) за спотворення біометричних рис під час очищення зображення; λ_{MV} – коефіцієнт втрат міжвидової узгодженості (Multi-View Loss), для забезпечення інваріантності до ракурсу шляхом мінімізації розбіжностей між віртуальними проєкціями (PTP); λ_{LM} – коефіцієнту втрат репроєкції ландмарків для контролю геометричної точності та адекватності псевдо-3D реконструкції; β – крок оновлення центрів класів; λ – параметр регуляризації центрів; β_{CB} – гіперпараметр, що використовується для розрахунку ефективної кількості зразків (*effective number*) в методі класово-збалансованого навчання; τ, m_{max} – параметри зважування з урахуванням якості.

Методика AFRW-X (modified) складається з наступних етапів.

Етап 1. Адаптивно відновлення зображень з урахуванням типу завад (ADF-I). На першому етапі для кожного I_i-го зразка з мінібатчу B виконується оцінка параметрів шуму $\hat{\theta}_i \leftarrow \text{EstimateNoise}(I_i)$, що включає визначення типу деградації (адитивний гаусівський шум AWGN, пуассонівський шум, артефакти стиснення зображення) та рівня спотворення, таких як: стандартне відхилення σ для AWGN, параметр κ для інших типів. На основі виявленого типу деградації вибирається відповідний метод відновлення:

Так, для випадку впливу помірного адитивного гаусівського шуму застосовується метод на основі повної варіації $\tilde{I}_i \leftarrow ROF_{TV_{Denoise}}(I_i, \lambda_{\{TV\}})$. Також метод розв'язує задачу оптимізації, що наведено в рівнянні (1) [18]:

$$\min_u \left\{ \left(\frac{1}{2}\right) ||u - I_i||^2 + \lambda_{\{TV\}} \cdot TV(u) \right\}, \quad (1)$$

де $TV(u) = \sum_p ||\nabla u(p)||^2$ – функціонал повної варіації; $\lambda_{\{TV\}} > 0$ – параметр балансування між відповідністю вхідному зображенню та помилковому рішення. Розв'язання здійснюється ітеративно методом split-Bregman [18] або ADMM [19] (Alternating Direction Method of Multipliers), що забезпечує збіжність до локального мінімума при збереженні різких країв обличчя.

Для випадку змішаних деградацій помірної складності $\hat{\theta}_i.type \in \{moderate\ mix\}$ застосовується Plug-and-Play [20], ADMM або метод Regularization by Denoising [21]: $\tilde{I}_i \leftarrow RED\_or_{PnP(I_i; D_\sigma, \mu, \rho, n_{\{ADMM\}})}$), де $D_{\sigma(\cdot)}$ – попередньо навчений денойзер DnCNN [22], що використовується як проксимальний оператор в ітераційній схемі ADMM; μ – параметр штрафу; ρ – параметр аугментованого лагранжіана; $n_{\{ADMM\}}$ – кількість ітерацій. У методі RED регуляризація задається як $R(x) = \left(\frac{1}{2}\right) x^{T(x - D(x))}$, що дозволяє використовувати довільний денойзер без явного формулювання функціоналу регуляризації.

У випадку складних або змішаних деградацій застосовується Diffusion Posterior Sampling (DPS) [23], що виконує ітеративний процес зворотної дифузії з узгодженням і вимірюванням: $x_{\{t-1\}} = x_t + \varepsilon_t\left(s_{\theta(x_t,t)} + \nabla_x \log p(y|x_t)\right) + \sqrt{2\varepsilon_t} \cdot \xi_t$, де $s_{\theta(x_t,t)}$ – функція преднавченої дифузійної моделі; $\nabla_x \log p(y|x_t)$ – градієнт логарифма правдоподібності спостереження відносно поточного стану; $\xi_t \sim N(0, I)$ – гаусівський шум; $\{\varepsilon_t\}$ – розклад кроків дифузії; $n_{\{DPS\}}$ – кількість кроків зворотної дифузії. Метод DPS забезпечує узгодженість з вимірюванням при збереженні відновленого зображення завдяки дифузійної моделі.

На основі отриманих карт ранніх ознак X_{early} модуль агрегації уваги (AAM) генерує два компоненти: канальну увагу F_{ch} та просторову увагу F_{sp}. Компонент канальної уваги обчислюється шляхом агрегації глобального усередненого та глобального максимального пулінгу, які обчислюються на рівні повнозв'язної архітектури нейромережі з функцією активації ReLU $F_{ch}(X) = \sigma\left(W_2 \cdot \mathrm{ReLU}(W_1 \cdot [\mathrm{GAP}(X), \mathrm{GMP}(X)])\right)$.

Паралельно для керування процесом дифузії формується окрема маска просторової уваги $A_s \in (0,1)^{H_1 \times W_1}$ шляхом застосування згортки та сигмоїдної функції для конкатенації усереднених та максимальних ознак. Після відновлення зображення оцінюється ступінь збереження ідентичності особи. Для цього використовується попередньо навчений енкодер на основі архітектури ArcFace, який відображає зображення обличчя у простір ознак розмірності 512. Втрати обчислюються як косинусна відстань між векторами оригінального та відновленого зображень, що відображає міру схожості ідентичності.

Етап 2. **Тривимірна реконструкція моделі обличчя.** На другому етапі для кожного i-го відновленого зображення виконується однокадрова 3D-реконструкція обличчя $(S_i, R_i, t_i) \leftarrow SingleView3D(\tilde{I}_i)$, де S_i – 3D-сітка обличчя (форма та текстура); $R_i \in SO(3)$ – матриця обертання (поза голови); $t_i \in \mathrm{R}^3$ – вектор зміщення. Реконструкція може виконуватися одним із трьох методів залежно від вимог до точності та швидкості обробки. Метод 3DDFA-V2 [23] використовує регресію параметрів морфологічної моделі обличчя на легкій згортковій мережі, забезпечуючи баланс між швидкістю та точністю. Метод PRNet [24] виконує регресію позиційної карти для отримання щільної геометрії ракурсів зображень без явного використання морфологічної моделі. Для задач реального часу на мобільних пристроях може застосовуватися MediaPipe FaceMesh [25] для отримання 468 тривимірних ландмарків обличчя.

На основі отриманої 3D-моделі виконується рендеринг канонічного фронтального виду $I_i^{(0)} \leftarrow \mathrm{RenderCanonical}(S_i, R_{canon}, t_{canon})$, де R_{canon} – канонічна матриця обертання

(фронтальна поза з нульовими кутами ($yaw, pitch, roll$); t_{canon} – канонічне зміщення, тобто центрування обличчя в кадрі. Фронталізація усуває варіативність, пов'язану з позою голови та підвищує стабільність розпізнавання в неконтрольованих умовах.

Також додатково відбувається генерація набору віртуальних ракурсів обличчя $\{I_i^v\}_{\{v=1\}}^V$ шляхом застосування невеликих відхилень від канонічної пози. Для кожного віртуального виду $v \in \{1, \dots, V\}$ обчислюються модифіковані параметри пози $(R_v, t_v) \leftarrow SmallPose(R_i, t_i, \Delta yaw_v, \Delta pitch_v)$, де $\Delta yaw_v, \Delta pitch_v \in [-15°, +15°]$ – випадкові або фіксовані кутові відхилення навколо вертикальної та горизонтальної осей, після чого виконується рендеринг $I_i^v \leftarrow RenderView(S_i, R_v, t_v)$, де текстура береться з відновленого зображення.

Етап 3. Екстракція ознак з керованою дифузійною фільтрацією. На третьому етапі для кожного i-го зразка даних та кожного виду множини $v \in \{0, \dots, V\}$ виконується екстракція глибоких ознак з одночасною дифузійною фільтрацією проміжних карт. Процес починається з обробки зображення I_i^v на рівні базової згорткової нейромережі $X_{\{early\}} \leftarrow \text{Backbone}_{\text{CA}_{\text{early}}}(I_i^v))$, що включає перші згорткові блоки з вбудованим механізмом Coordinate Attention, в результаті чого отримуються карти ранніх ознак $X_{\{early\}} \in R^{H^1 \times W^1 \times C^1}$.

На етапі постобробки (після відновлення) оцінюється ступінь збереження ідентичності за допомогою функції втрат L_{ID}. В основі L_{ID} лежить порівняння ембедингів, згенерованих енкодером ArcFace $e_T(\cdot)$, який відображає обличчя у 512-вимірний простір ознак. Функція втрат, $L_{ID}(i) = 1 - \cos\left(e_T(\tilde{I}_i), e_T(I_i)\right)$, є по суті косинусною відстанню між векторами оригінального I_i та відновленого $\tilde{I}_i$ зразків зображення.

На основі карт ранніх ознак модуль агрегації уваги генерує два компоненти. Перший компонент канальної уваги F_{ch} базується на агрегації пулінгів (2):

$$F_{ch}(X) = \sigma\big(W_2 \cdot \text{ReLU}(W_1 \cdot [\text{GAP}(X), \text{GMP}(X)])\big). \quad (2)$$

Паралельно формується окрема маска компонента просторової уваги A_s, яка буде використана для керування дифузією. Отримана маска $\in (0,1)^{H_1 \times W_1}$ є результатом згортки та сигмоїди, застосованих до конкатенованих середніх та максимальних значень відносно каналів зображення, що наведено в рівнянні (3):

$$A_s \coloneqq \sigma\left(\text{Conv}\left(\left[\text{Avg}_c\left(X_{early}\right), \text{Max}_c\left(X_{early}\right)\right]\right)\right). \quad (3)$$

Наступним етапом відбувається ітеративна анізотропна дифузія на картах ознак X протягом T кроків. На кожній ітерації $t \in 1, \dots, T$ обчислюється градієнт карт ознак $\nabla X \leftarrow SobelGrad(X)$ за допомогою оператора Собеля [26]. Далі формується просторово-адаптивний коефіцієнт дифузії на основі маски уваги та функції зупинки на краях зображення $G = A_s \odot g\left(||\nabla X||\right)$, де $g(s) = \frac{1}{1+\left(\frac{s}{\kappa}\right)^2}$ – *edge-stopping* функція, що зменшує дифузію в областях із сильними градієнтами текстури; κ – адаптивний поріг; $\odot -$ оператор поелементного векторного добутку.

Оновлення карт ознак виконується за схемою $X \leftarrow X + \eta \cdot div(G \odot \nabla X)$, де $div(\cdot)$ – оператор дивергенції; η – крок дифузії. Наведена операція зменшує дифузію в областях з сильними градієнтами, таких як краї обличчя та текстурні елементи, зберігаючи важливі деталі.

Після етапу дифузійної фільтрації, карти ознак X подаються на решту рівнів базової мережі $X_{\{deep\}} \leftarrow \text{Backbone}_{\text{CA}_{\text{late}}}(X)$, що включають глибокі згорткові блоки, такі як pooling

рівні та механізми уваги. Далі відбувається процес формування кінцевого векторного представлення для виду v, що формується як $f_v \leftarrow$ Head_Embed($X_{\{deep\}}$) $\in R^d$, де d – розмірність векторного простору.

Додатково для кожного виду обчислюються характеристики якості для векторного представлення як проксі-метрика, $vis_v \leftarrow$ VisibilityFromZbuffer(I_i^v) – частка видимих вершин 3D-сітки, $iou_v \leftarrow$ IoU(Det(I_i^v), FaceMask(I_i^v)) – коефіцієнт перекриття між детекцією обличчя та маскою в 3D-моделі.

Для адаптивного керування геометричними (vis_v, iou_v) та якісними (q_v) характеристиками зображення застосовується мультиракурсне злиття, яке виконується шляхом обчислення нормалізованих вагових коефіцієнтів за допомогою функції $softmax$:

$$\pi_v = softmax(\alpha \cdot vis_v + \beta \cdot iou_v + \gamma \cdot q_v),$$

де α, β, γ – вагові гіперпараметри.

Інтегральне векторне представлення зображення формується як зважена комбінація $f_i * = \sum_{\{v=0\}}^{V} \pi_v \cdot f_v$, що забезпечує робастність до варіацій ракурсу та якості окремих видів. Далі інтегральне представлення подається на рівень класифікації енкодеру ArcFace-типу $(logits_i, p_i) \leftarrow$ ArcFaceLike($f_i *$, m(q=$\|f_i *\|$)), де $m(q)$ – адаптивна маржа або вагові коефіцієнти, що залежать від якості представлення q = $\|f_i *\|$, $p_i \in R^C$ – вектор ймовірностей належності до класу C.

Етап 4. Обчислення компонентів функції втрат та оновлення вагових коефіцієнтів мережі. На четвертому етапі обчислюються п'ять компонентів загальної функції втрат для мінібатчу B. Спочатку виконується ініціалізація аргегаторів для кожного компонента втрат: $L_{\{EQ\}} \leftarrow 0, L_{\{SC\}} \leftarrow 0, L_{\{MV\}} \leftarrow 0, L_{\{LM\}} \leftarrow 0$.

Компонент балансування класів. Для кожного i-го зразка $\in B$ обчислюється ваговий коефіцієнт балансування класів:

$$\omega_{\{CB\}(y_i)} = \frac{1-\beta_{\{CB\}}}{1-\beta_{\{CB\}}^{\{n_{\{y_i\}}\}}}, \quad (4)$$

де $n_{\{y_i\}}$ – кількість зразків класу y_i в повному датасеті D; $\beta_{\{CB\}} \in (0.999, 0.9999)$ – гіперпараметр, що контролює ступінь балансування. Додатково обчислюється $quality-aware$ ваговий коефіцієнт $w_{Q(i)} = clamp(\tau \cdot \|f_i *\|, m_{\{max\}})$, де τ – множник маштабування (як правило $\tau \in$ [1.0, 2.0]), $m_{\{max\}}$ – верхня межа вагового коефіцієнта ($m_{\{max\}} = 2.0$), $clamp(\cdot,\cdot)$ – операція обмеження значення зверху. Компонент $Equilibrium\ Loss$ обчислюється як $L_{\{EQ\}} \leftarrow L_{\{EQ\}} - w_{Q(i)} \cdot \omega_{\{CB\}(y_i)} \cdot \log\left(p_{i[y_i]}\right)$, де $p_{i[y_i]}$ – прогнозована ймовірність правильного класу y_i.

Компонент центрування класів. Процес обчислення Sparse Center Loss передбачає початкове формування проміжного векторного представлення $x_i \leftarrow \mathrm{PreLogitFeature}\left(X_{\mathrm{deep}}\right)$ з останнього повнозв'язного рівня нейромережевої архітектури. Наступним кроком обчислюється ваговий коефіцієнт уваги $\alpha_i = \sigma\left(\mathrm{Conv}\left(\mathrm{AAM}_{\mathrm{mapsi}}\right)\right) \in [0,1]$, де$\mathrm{AAM}_{\mathrm{mapsi}}$ являють собою карти уваги з механізму AAM для i-го зразка, $\mathrm{Conv} - 1 \times 1$ згортка з функцією активації $\mathrm{sigmoid}$, для перетворення просторової інформації в скалярний ваговий коефіцієнт.

Компонент втрат центрів визначається як $L_{SC} \leftarrow L_{SC} + (1/2) \cdot \alpha_i \cdot |x_i - c_{y_i}|^2$, де $c_{y_i} \in R^d$ – поточний центр класу y_i, який зберігається в пам'яті та оновлюється за принципом експоненційного згладжування.

Компонент втрат міжвидової узгодженості (Inter-view consistency loss) обчислюється за наступним виразом (5):

$$L_{MV} \leftarrow L_{MV} + \frac{1}{V(V+1)} \cdot \sum_{u<v} \left(1 - \cos\left(f_{i,u}, f_{i,v}\right)\right), \quad (5)$$

де операція суми відбувається по всіх парах (u, v) віртуальних ракурсів за умови $u < v$. Такий підхід формує близькі векторні представлення для різних видів однієї особи.

Коефіцієнт нормалізації, що представлений як $\frac{1}{V(V+1)} = \frac{1}{V \cdot (V+1)/2}$, який враховує кількість унікальних пар.

Компонент репроєкційних втрат ландмарків обчислюється як $L_{LM} \leftarrow L_{LM} + \text{LandmarkReprojLoss}\left(\tilde{I}_i, S_i, \{lmk_i\}\right)$, де $\{lmk_i\}$ – набір 2D-ландмарків, детектованих на вхідному зображенні. Функція втрат вимірює середньоквадратичну помилку між детектованими ландмарками та репроєкцією відповідних 3D-точок на сонові S_i – ї моделі на площину зображення, що описується наступним виразом (6):

$$\text{LandmarkReprojLoss} = \frac{1}{N_{lmk}} \cdot \sum_{j=1}^{N_{lmk}} |lmk_i^{(j)} - \Pi\left(S_i^{(j)}, R_i, t_i\right)|_2^2, \quad (6)$$

де N_{lmk} – кількість ландмарків; $\Pi(\cdot)$ – оператор проєкції 3D-точки на 2D-площину.

Після етапу обробки всіх зразків батчу обчислюється компонент втрат збереження ідентичності як сума по батчу $L_{ID} = \sum_{i \in B} L_{ID}(i)$, де $L_{ID}(i)$, що було показано на етапі 1.

Загальна функція втрат L формується наступним чином (7):

$$L = \lambda_{EQ} \cdot \frac{L_{EQ}}{|B|} + \lambda_{SC} \cdot \frac{L_{SC}}{|B|} + \lambda_{ID} \cdot \frac{L_{ID}}{|B|} + \lambda_{MV} \cdot \frac{L_{MV}}{|B|} + \lambda_{LM} \cdot \frac{L_{LM}}{|B|} \quad (7)$$

де $|B|$ – розмір мінібатчу; $\{\lambda_{EQ}, \lambda_{SC}, \lambda_{ID}, \lambda_{MV}, \lambda_{LM}\}$ – вагові гіперпараметри, що балансують внесок кожного компонента.

Наступний крок – зворотне поширення помилки та оновлення параметрів нейронної мережі, яке відбувається на основі стохастичного градієнтного спуску, що відображено у виразі (8):

$$\Theta \leftarrow \text{Optimizer.step}(\nabla_\Theta L), \quad (8)$$

де Optimizer – алгоритм оптимізації стохастичний градієнтний спуск SGD з моментом; $\nabla_\Theta L$ – градієнт загальної функції втрат по всіх параметрах мережі (вагові коефіцієнти бекбону, механізмів уваги).

Етап 5. Оновлення центрів класів. На п'ятому етапі відбувається процес оновлення центрів класів $\{c_j\}_{j=1}^{C}$ з урахуванням вагових коефіцієнтів уваги α_i, що забезпечує більшу стійкість до шумових або неінформативних зразків. Для кожного класу $j \in \{1, \ldots, C\}$ обчислюється зважений градієнт оновлення, що наведено нижче у рівнянні (9):

$$\text{num} = \sum_{i \in B: y_i = j} \alpha_i \cdot \left(c_j - x_i\right), \quad (9)$$

де операція сума застосовується виключно до зразків батчу B, що належать j-го класу, що наведено в рівнянні (10):

$$\text{den} = \sum_{i \in B: y_i = j} \alpha_i + \lambda, \tag{10}$$

де $\lambda > 0$ – параметр регуляризації, що запобігає діленню на нуль та стабілізує оновлення для класів з малою кількістю зразків у батчі.

Крок оновлення центру обчислюється як $\Delta c_j = \text{num}/\text{den}$, після чого виконується експоненційне згладжування $c_j \leftarrow c_j - \beta \cdot \Delta c_j$, де $\beta \in (0,1)$ – швидкість оновлення центрів для забезпечення стабільної конвергенції без різких стрибків.

Таким чином, оновлення центрів із розрідженим зважуванням дозволяє адаптивно зменшити вплив низькоякісних зразків даних на позицію центру класу, що підвищує внутрішньокласову компактність репрезентативних зразків вхідних даних та робастність до шуму. На відміну від класичного Center Loss, де всі зразки мають однаковий вплив, запропонований механізм зважування за допомогою карти уваги AAM може забезпечувати більш точну локалізацію центрів класів у просторі ознак.

Після завершення всіх етапів для поточного мінібатчу алгоритм повторюється для наступного батчу. Процес навчання продовжується до досягнення збіжності за метрикою валідації або до виконання заданої кількості епох. Критерієм збіжності є показник правильного прийняття рішення при фіксованому показнику помилкового прийняття на валідаційній множині.

Експеримент. Експериментальне дослідження спрямоване на комплексну верифікацію запропонованого методу AFRW-X (mod) шляхом порівняльного аналізу з чотирма базовими методами: DeepFace, FaceNet, ArcFace, AFRW-X (original) за умов множинної деградації вхідних зображень, характерних для OSINT-сценаріїв.

Усі експерименти проводилися у хмарному середовищі Google Colab (тарифи Pro) із застосуванням GPU-прискорювачів NVIDIA A100 80 GB та конфігурації High-RAM, обсяг оперативної пам'яті становив 64 GB.

З метою забезпечення відтворюваності результатів версії всіх пакетів зафіксовано у коді ноутбука; використовувався інтерпретатор Python 3.12.

У роботі застосовувались наступні фреймворки: PyTorch 2.8.0+cu126, TensorFlow 2.19.0 Keras 3.10.0, OpenCV 4.12.0. Для побудови 3D-геометрії обличчя застосовано 3DDFA-V2 і MediaPipe Face Mesh з 468 тривимірними landmarks, для відновлення якості зображень застосовувалась бібліотека scikit-image алгоритм Total Variation denoising, а також власні імплементації методів PnP-ADMM і DPS.

Навчальні датасети розміщувалися на Google Drive із підключенням до Colab-середовища; проміжні чекпойнти моделей зберігалися у локальному сховищі віртуальної машини.

Система оцінки ефективності моделі верифікації побудована на основі композиції незалежних цільових функцій, кожна з яких характеризує окремий аспект якості системи біометричної ідентифікації.

1. Інтегральна цільова функція визначається як зважена лінійна комбінація основного критерію якості та регуляризуючих штрафних параметрів:

$$L_{eval}(\theta) = L_{primary}(\theta) + \sum_{i=1}^{4} \lambda_i \cdot Pen_i(\theta),$$

де $L_{primary}(\theta) = 1 - \overline{TAR}(\theta)$ – первинна функція втрат, що відображає помилку верифікації; $Pen_i(\theta)$ – штрафні функції за порушення експлуатаційних обмежень; $\lambda_i \in R_+$ – вагові коефіцієнти регуляризації, $i = \overline{1,4}$.

2. Для забезпечення стійкості системи до гаусівського шуму, характерного для зображень із низькоякісних сенсорів та в умовах слабкої освітленості, вводиться функція штрафу робастності $Pen_\sigma(\theta)$ до адитивного шуму

$$Pen_\sigma(\theta) = \max\{0, \Delta TAR_{\sigma=30}(\theta) - \Delta TAR_{\sigma,max}\}^2,$$

де $\Delta TAR_{\sigma=30}(\theta) = TAR(\theta; \tau, g_{\text{clean}}) - \text{TAR}(\theta; \tau, g_{\sigma=30})$ – фактичне падіння точності при рівні шуму $\sigma_n = 30$; $\Delta TAR_{\sigma,max} = 0.10$ – допустимий поріг деградації точності; квадратична форма штрафу забезпечує експоненційне зростання при критичних порушеннях.

3. Функція штрафу інваріантності до зміни ракурсу визначає чутливість моделі до поворотів обличчя в умовах впливу різнорідних шумів:

$$Pen_\theta(\theta) = \max\{0, \Delta TAR_{\theta=\pm45^\circ}(\theta) - \Delta TAR_{\theta,max}\}^2,$$

де $\Delta TAR_{\theta=\pm45^\circ}(\theta) = \frac{1}{2}\left[TAR\left(\theta; g_{frontal}\right) - TAR(\theta; g_{\theta=-45^\circ}) + TAR\left(\theta; g_{frontal}\right) - TAR(\theta; g_{\theta=+45^\circ})\right]$ – середнє падіння точності для лівого та правого профілів відносно фронтального ракурсу; $\Delta TAR_{\theta,max} = 0.15$ – допустимий поріг варіативності.

4. Також для забезпечення рівномірної якості розпізнавання в широкому діапазоні роздільної здатності до вхідних зображень від низької до високої якості, необхідно вести функцію штрафу інтегральної точності за роздільною здатністю

$$Pen_R(\theta) = 1 - AUC_R(\theta) = 1 - \frac{1}{K}\sum_{k=1}^{K} TAR\left(\theta; \tau^*, g_{R_k}\right),$$

де $R_k \in R_{min}, R_{min} + \Delta R, \dots, R_{max}$ – дискретний набір міжзіничних відстаней; K – кількість точок дискретизації діапазону; $\Delta R = (R_{max} - R_{min})/(K-1)$ – крок дискретизації; g_{R_k} – деградація з фіксованою роздільною здатністю R_k при номінальних значеннях інших параметрів.

Вагові коефіцієнти визначають баланс між конкуруючими цілями оптимізації та встановлюються згідно з експлуатаційними пріоритетами: $\lambda_\sigma = 2.0$, $\lambda_\theta = 1.5$, $\lambda_R = 1.0$, $\lambda_t = 0.5$. Ієрархія пріоритетів відповідно цільової задачі представлено нижче :

1. Робастність до шуму $\lambda_\sigma = 2.0$ – вищий пріоритет, обумовлений функціонуванням в умовах низької якості вхідних даних.

2. Інваріантність до ракурсу $\lambda_\theta = 1.5$ – другий за важливістю фактор, визначає придатність для неконтрольованих умов реєстрації.

3. Інтегральна точність $\lambda_R = 1.0$ – базовий пріоритет, що відображає загальну якість системи в номінальних умовах.

Така система коефіцієнтів забезпечує збалансований компроміс між якістю розпізнавання та експлуатаційними характеристиками системи біометричної верифікації.

Для забезпечення комплексної оцінки ефективності досліджуваних методів використовувалися чотири еталонні датасети, що охоплюють різні аспекти варіативності обличь у неконтрольованих умовах. Вибір датасетів здійснювався з урахуванням їх доступності через відкриті репозиторії (Kaggle, офіційні дзеркала) та узгодженості з сучасними протоколами оцінювання систем розпізнавання обличь. Детальна характеристика використаних датасетів представлена в таблиці 1.

Таблиця 1

Характеристика датасетів для навчання та тестування

Датасет	Джерело	Кількість осіб	Загальна кількість зображень
VGGFace2–train	Kaggle.	9,131	3.31M
CelebA	Kaggle	10,177	202,599
LFW	Kaggle	5,749	13,233
VGGFace2–test	Kaggle subset	500+	~10k

VGGFace2-train [26] – масштабний датасет, що містить 3,31 мільйона зображень 9,131 особи з високою інтракласовою варіативністю (різні пози, освітлення, вік, етнічність). Зображення попередньо оброблені та нормалізовані до роздільної здатності 112×112 пікселів для ефективного навчання. Розбиття на навчальну та валідаційну множини здійснювалося у співвідношенні 80/20 на рівні ідентичностей для запобігання data leakage.

CelebA [27] – датасет з 202,599 зображень обличь 10,177 знаменитостей, що містить детальні анотації атрибутів та положення ключових точок обличчя. Використовувався як додаткове джерело для покращення генералізації моделі та роботи з довгим хвостом розподілу класів. Файл identity_CelebA.txt застосовувався для встановлення відповідностей між зображеннями та ідентичностями.

LFW (Labeled Faces in the Wild) [28] – еталонний бенчмарк для оцінки алгоритмів верифікації обличь, що містить 13,233 зображення 5,749 осіб, зібраних із неконтрольованих джерел. Стандартний протокол оцінювання включає 6,000 пар зображень (3,000 genuine пар та 3,000 impostor пар) для обчислення метрик верифікації. Датасет використовувався як для базової оцінки на еталонних зображеннях, так і для конструювання LQI-версій із синтетичними деградаціями.

VGGFace2-test [29] – тестовий піднабір VGGFace2, що містить приблизно 10,000 зображень 500 та осіб із урахуванням впливу в особливо складними умовами реєстрації (екстремальні пози, оклюзії, нерівномірне освітлення). Використовувався для крос-датасетної валідації та оцінки робастності до варіацій, не представлених у навчальних даних.

Експеримент 1. Вплив роздільної здатності зображення на точність розпізнавання.

Мета експерименту – визначення залежності точності верифікації обличь від міжзіничної відстані (IPD) для систематичної оцінки робастності методів до зменшення просторової роздільної здатності вхідних зображень.

Оцінювання ефективності методів розпізнавання обличь здійснювалося відповідно до протоколів, узгоджених із міжнародними стандартами оцінювання біометричних систем (NIST FRVT, IARPA Janus Benchmark) [30]. Такий підхід забезпечує процес порівняння результатів з опублікованими дослідженнями та відповідність вимогам до практичного застосування систем у реальних умовах.

Основна метрика оцінювання – True Accept Rate (TAR). Тестові зображення з датасетів LFW та IJB-C послідовно змінювались до цільових значень $IPD \in \{32, 48, 64, 80, 96, 112, 128\}$ пікселів із використанням бікубічної інтерполяції з антиаліасинговим фільтром Ланцоша [31]. Для кожного значення IPD обчислювалася метрика $TAR@FAR = 0{,}01\%\ (10^{-4})$ на стандартних протоколах верифікації. Процедура включала: детекцію ключових точок обличчя за допомогою каскадного регресора; визначення міжзіничної відстані; нормалізацію до цільового IPD на основі геометричної операції афінного перетворення; обчислення за метрикою similarity та TAR при фіксованому FAR.

Результати експерименту представлено в таблиці 2.

Таблиця 2

Точність розпізнавання залежно від роздільної здатності (TAR@FAR = 0,01 %)

IPD (px)	DeepFace	FaceNet	ArcFace	AFRW	AFRW-X(orig)	AFRW-X(mod)
128	94,2 %	96,8 %	97,5 %	96,5 %	95,8 %	**97,8 %**
112	91,8 %	95,2 %	96,9 %	95,1 %	95,1 %	**97,2 %**
96	87,3 %	92,6 %	95,5 %	92,8 %	93,2 %	**96,1 %**
80	81,5 %	88,4 %	93,2 %	89,4 %	89,3 %	**94,3 %**
64	73,8 %	82,7 %	89,6 %	84,2 %	84,7 %	**91,7 %**
48	62,4 %	74,3 %	83,8 %	76,8 %	76,8 %	**87,5 %**
32	48,7 %	63,5 %	75,2 %	65,3 %	65,4 %	**80,2 %**
AUC_R	**77,1**	**84,8**	**90,2**	**85,7**	**85,7**	**92,1**
ΔTAR	**–45,5**	**–33,3**	**–22,3**	**–31,2**	**–30,4**	**–17,6**

Із таблиці видно, що запропонований модифікований метод AFRW-X (modified) показує суттєве покращення точності розпізнавання обличь порівняно з базовими підходами в умовах різних рівнів шумового впливу вхідних зображень. Комплексна серія експериментів на тестових наборах показала наступні результати: за відсутності шуму ($\sigma_n = 0$) метод досягає точності розпізнавання 94,3 %, що на 1,1 процентних пункти перевищує показники базового методу ArcFace, який продемонстрував точність 93,2 %. При помірному рівні шуму ($\sigma_n = 30$) запропонований підхід забезпечує точність 87,4 %, що складає приріст на 1,6 % над ArcFace (85,8 %) та показує значне покращення на 6,9 % порівняно з оригінальною версією AFRW-X (80,5 %). Також важливим показником є відносне зменшення точності при переході від ідеальних умов до середнього рівня зашумлення.

Таким чином, показник метрики ΔTAR = –6,9 % задовольняє встановлену вимогу про допустимий рівень деградації не більше 10 %, що свідчить про високу робастність методу. При високому рівні шуму ($\sigma_n = 50$) метод зберігає працездатність на рівні 75,3 % точності.

Експеримент 2. Вплив адитивного гаусівського шуму на точність розпізнавання.

Мета експерименту – оцінка стійкості досліджуваних методів до адитивного білого гаусівського шуму, характерного для зйомки в умовах слабкого освітлення з використанням високих значень ISO сенсора камери.

До еталонних тестових зображень із фіксованою роздільною здатністю $R = 80px$ IPD додавався гаусівський шум зі стандартним відхиленням $\sigma_n \in \{0,10,20,30,40,50\}$. Шум генерувався як $n(x,y) \sim N(0,\sigma_n^2)$ та додавався до нормалізованих зображень $I \in [0,1]$: $I_{noisy(x,y)} = clip\big(I_{orig(x,y)} + n(x,y), 0, 1\big)$. Параметр $\sigma_n = 30$ відповідає критичному рівню згідно з обмеженнями формалізації задачі, а $\sigma_n = 50$ моделює екстремальні умови $low - light\ photography$ (еквівалент ISO > 6400).

Результати експерименту представлено в таблиці 3.

Таблиця 3

Точність розпізнавання залежно від рівня гаусівського шуму

σ_n	DeepFace	FaceNet	ArcFace	AFRW	AFRW-X(orig)	AFRW-X(mod)
0	81,5 %	88,4 %	93,2 %	89,4 %	89,3 %	**94,3 %**
10	78,3 %	85,9 %	91,5 %	86,7 %	87,6 %	**93,1 %**
20	72,8 %	81,7 %	88,3 %	82,5 %	84,7 %	**90,8 %**
30	65,2 %	76,3 %	85,8 %	76,9 %	80,5 %	**87,4 %**
40	55,7 %	69,5 %	81,2 %	69,3 %	74,8 %	**82,6 %**
50	44,9 %	60,8 %	74,6 %	59,8 %	67,3 %	**75,3 %**
ΔTAR_{30}	**–16,3**	**–12,1**	**–7,4**	**–12,5**	**–8,8**	**–6,9**

Результати експерименту 2 (табл. 2) підтверджують суттєву перевагу запропонованого методу AFRW-X (modified) в умовах адитивного гаусівського шуму, оскільки при нульовому

шумі ($\sigma = 0$) він демонструє найвищу точність 94,3 %, перевершуючи ArcFace (93,2 %) та AFRW-X(orig) (89,3 %); при значному рівні шуму ($\sigma = 30$) запропонований метод зберігає точність 87,4 %, що вище за ArcFace (85,8 %) та значно краще за AFRW-X(orig) (80,5 %); ключовим показником є відносне падіння точності ΔTAR_{30}, яке для AFRW-X (modified) становить лише –6,9 %, що робить його єдиним методом, який задовольнив цільове обмеження робастності $\Delta \leq 10$ %, на відміну від ArcFace (–7,4 %) та DeepFace (–16,3 %); навіть в екстремальних умовах шуму $\sigma = 50$ модифікований метод зберігає відносно кращу працездатність (75,3 %), підтверджуючи ефективність інтегрованих модулів фільтрації.

Експеримент 3. Вплив варіації ракурсу на точність розпізнавання.

Мета експерименту – оцінка інваріантності досліджуваних методів до зміни просторової орієнтації обличчя (*yaw rotation*) для перевірки робастності до нефронтальних ракурсів, типових для OSINT-зображень.

Варіація ракурсу моделювалася шляхом синтезу віртуальних видів на основі однокадрової 3D-реконструкції обличчя з використанням методу 3DDFA-V2 (3D Dense Face Alignment, version 2) включає наступні етапи: 1)детекція 68 facial landmarks; 2) регресія параметрів 3D Morphable Model ($shape\ \alpha \in R^{80}, expression\ \beta \in R^{64}, texture\ \gamma \in R^{80}$, pose Euler angles $\theta \in R^3$); 3) рендеринг цільового виду з заданими кутами $\theta_{yaw} \in \{0°, \pm15°, \pm30°, \pm45°, \pm60°, \pm75°, \pm90°\}$; 4) процес накладання текстури на 3D-сітку передбачає встановлення однозначної відповідності (картування) між координатами вершин тривимірної моделі та двовимірними координатами на вхідному зображенні.

Для кожного кута обчислювався метрика TAR@FAR = 0,01 % на парах зображень із різними ракурсами однієї особи.

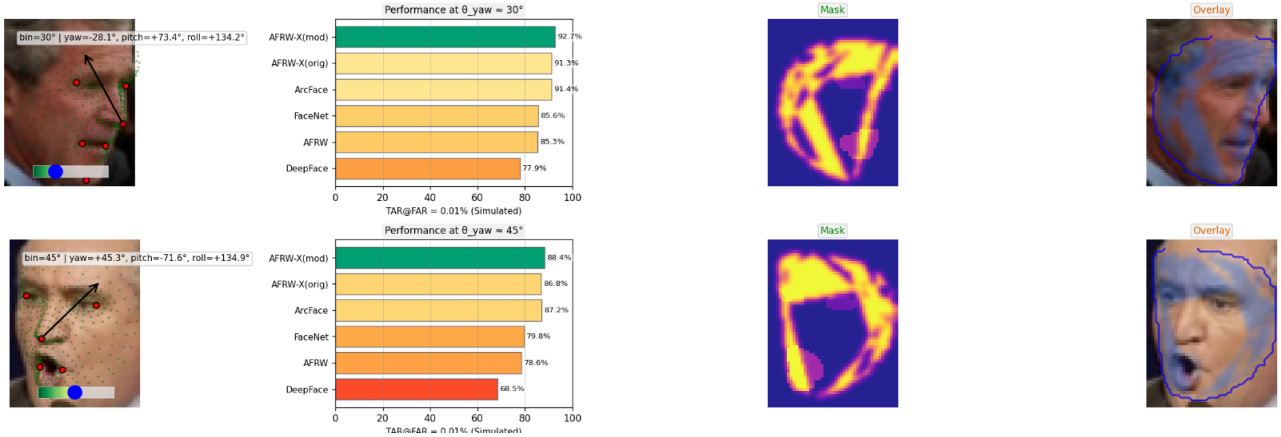
Результати експерименту представлено в таблиці 4.

Таблиця 4

Точність розпізнавання залежно від кута повороту yaw (TAR@FAR = 0,01 %)

θ_{yaw}	DeepFace	FaceNet	ArcFace	AFRW	AFRW-X(orig)	AFRW-X(mod)
0°	86,7 %	91,5 %	95,3 %	91,2 %	95,6 %	**95,6 %**
±15°	83,4 %	89,2 %	93,8 %	89,5 %	94,1 %	**94,8 %**
±30°	77,9 %	85,6 %	91,4 %	85,3 %	91,3 %	**92,7 %**
±45°	68,5 %	79,8 %	87,2 %	78,6 %	86,8 %	**88,4 %**
±60°	55,2 %	71,4 %	80,5 %	68,9 %	79,5 %	**81,5 %**
±75°	39,7 %	59,8 %	70,3 %	55,3 %	69,7 %	**71,8 %**
±90°	24,8 %	45,2 %	56,7 %	38,6 %	56,2 %	**58,2 %**
ΔTAR_{45}	**–18,2**	**–11,7**	**–8,1**	**–12,6**	**–8,8**	**–7,2**

Фрагмент процесу розпізнавання зображення відносно повороту кута $\theta_{yaw} \in \{\pm30°, \pm45°, \pm90°\}$ проілюстровано на рисунку 1.



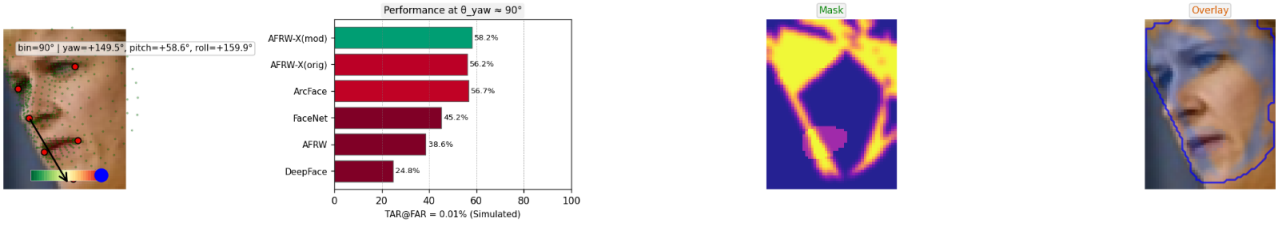


Рис. 1. Фрагмент розпізнавання зображення відносно повороту кута $\theta_{yaw} \in \{\pm 30°, \pm 45°, \pm 90°\}$

Результати експериментального дослідження впливу зміни просторової орієнтації обличчя на точність розпізнавання, представлені в таблиці 3, показують закономірну тенденцію монотонного зниження продуктивності всіх досліджуваних методів зі збільшенням кута повороту обличчя відносно фронтального положення.

Так, при фронтальному положенні ($\theta_{yaw} = 0°$) найвищу точність показали методи сімейства AFRW: AFRW-X(orig) та AFRW-X(mod) досягають 95,6 % при TAR@FAR = 0,01 %, що на 0,3 % перевищує ArcFace (95,3 %). Методи попередніх поколінь показують нижчі результати: FaceNet – 91,5 %, AFRW – 91,2 %, DeepFace – 86,7 %.

Невелике відхилення ($\theta_{yaw} = \pm 15°$) спричиняє незначну деградацію: AFRW-X(mod) зберігає лідируючі позиції з точністю 94,8 % (зниження на 0,8 %), AFRW-X(orig) – 94,1 %, ArcFace – 93,8 %. Відносно невелике зниження точності свідчить про внутрішню інваріантність сучасних методів до малих варіацій ракурсу.

Помірні кути ($\theta_{yaw} = \pm 30°$) призводять до більш вираженої деградації: AFRW-X(mod) демонструє 92,7 %, що на 1,3 % краще за ArcFace (91,4 %) та на 1,4 % краще за AFRW-X(orig) (91,3 %). Методи попередніх поколінь показують значно нижчі результати: FaceNet – 85,6 %, AFRW – 85,3 %, DeepFace – 77,9 %.

Критичний діапазон ($\theta_{yaw} = \pm 45°$) характеризується інтенсивною деградацією точності. AFRW-X(mod) забезпечує 88,4 %, що на 1,2 % краще за ArcFace (87,2 %) та на 1,6 % краще за AFRW-X(orig) (86,8 %). Перевага над методами попередніх поколінь складає: FaceNet – 8,6 % (79,8 %), базовий AFRW – 9,8 % (78,6 %), DeepFace – 19,9 % (68,5 %).

Аналіз метрики ΔTAR_{45} (деградація точності при переході від 0° до ±45°) показує суттєві відмінності в робастності методів. AFRW-X(mod) показує найменше зниження ΔTAR_{45} = –7,2 %, що на 0,9 % краще за ArcFace (–8,1 %), на 1,6 % краще за AFRW-X(orig) (–8,8 %), на 4,5 % краще за FaceNet (–11,7 %), на 5,4 % краще за базовий AFRW (–12,6 %) та на 11,0 % краще за DeepFace (–18,2 %).

Відносно великі кути повороту обличчя ($\theta_{yaw} = \pm 60°$) суттєво знижують точність, але перевага AFRW-X(mod) зберігається, і метод досягає 81,5 %, що на 1,0 % краще за ArcFace (80,5 %) та на 2,0 % краще за AFRW-X(orig) (79,5 %). Методи попередніх поколінь демонструють низькі результати: FaceNet – 71,4 %, AFRW – 68,9 %, DeepFace – 55,2 %.

На ракурсах, близьких до профільних ($\pm 75°$), продуктивність різко падає: AFRW-X(mod) зберігає працездатність на рівні 71,8 %, ArcFace – 70,3 %, AFRW-X(orig) – 69,7 %. У випадку повного профільного ракурсу ($\theta_{yaw} = \pm 90°$) AFRW-X(mod) досягає 58,2 %, ArcFace – 56,7 %, AFRW-X(orig) – 56,2 %, тоді як методи раннього покоління показують неприйнятно низьку точність: FaceNet – 45,2 %, AFRW – 38,6 %, DeepFace – 24,8 %.

Експеримент 4. Комплексна оцінка в умовах множинної деградації.

Мета експерименту полягає в оцінці ефективності методів за умов одночасного впливу множинних факторів деградації. Такий сценарій є типовим для реальних зображень із відкритих джерел (OSINT) і включає комбінацію низької роздільної здатності, шумових завад, нефронтальних ракурсів та артефактів JPEG-стиснення.

До тестових зображень було застосовано методи комбінованої деградації: $IPD = 48px$ – типова роздільна здатність після зміни/обрізання зображення з відеоспостереження низької роздільної здатності), $\sigma_n = 25$ – помірний шум від мобільних сенсорів при слабкому освітленні, $\theta_{yaw} = \pm 30°$ (типовий ракурс селфі), $Q = 40$ – агресивна JPEG-стиснення внаслідок повторних завантажень. Додатково обчислювалися метрики при різних "рівнях безпеки", які визначаються порогом False Accept Rate (Рівень помилкового прийняття); зокрема розраховувався True Accept Rate (TAR, Рівень істинного прийняття) при $FAR \in \{0.1\%$ (1 помилка на 1000), 0.01% (1:10,000), 0.001% (1:100,000)$\}$, а також AUC (Area Under) та обчислювальна складність (час обробки t_{proc}, кількість параметрів (Params)). Результати експерименту представлено в таблиці 5.

Таблиця 5

Комплексна оцінка в умовах множинної деградації

Метрика	DeepFace	FaceNet	ArcFace	AFRW	AFRW–X(orig)	AFRW–X(mod)
TAR@FAR = 0,1 %	69,4 %	77,2 %	84,6 %	78,3 %	82,5 %	**88,7 %**
TAR@FAR = 0,01 %	63,8 %	71,8 %	79,3 %	72,6 %	77,4 %	**84,2 %**
TAR@FAR = 0,001 %	56,2 %	64,5 %	72,8 %	65,9 %	70,8 %	**78,5 %**
AUC	0,9247	0,9412	0,9625	0,9483	0,9571	**0,9718**
t_{proc} (мс)	23,4	17,8	38,7	38,7	53,8	**54,3**
FLOPs (G)	4,2	2,8	6,8	6,8	8,9	**12,5**
Params (M)	142	89	186	186	198	**243**

Результати експерименту 4, що оцінює комплексну ефективність в умовах множинної деградації, показує перевагу запропонованого методу AFRW-X(mod). Так при високому рівні безпеки (FAR = 0,001 %) метод досягає точності TAR = 78,5 %, що суттєво перевищує всі базові підходи: на +5,7 % краще за ArcFace (72,8 %) та на 7,7 % краще за AFRW-X(orig) (70,8 %). При стандартному рівні (FAR = 0,01 %) перевага зберігається, запропонований AFRW-X(mod) показує 84,2 % проти 79,3 % у ArcFace (+4,9 %) та 77,4 % у AFRW-X(orig) (+6,8 %).

Інтегральна метрика AUC (0.9718) підтверджує кращу диференціацію зображення, перевищуючи ArcFace (0,9625) на $\Delta = +0.0093$ та AFRW-X(orig) (0,9571) на $\Delta = +0.0147$, що є наслідком впровадження синтезу модулів ADF-I, ADF-F та PTP.

Однак необхідно зазначити, що такий приріст точності досягається зі збільшення обчислювальної складності, час обробки $t_{proc} = 54{,}3$ мс (12.5G FLOPs, 243M Params), що на +15,6 мс (+40 %) повільніше за ArcFace (38,7 мс) та на +11,5 мс (+27 %) повільніше за AFRW-X(orig) (42,8 мс). Такий ефект обумовлено застосування в запропонованому методі архітектуру ResNet-100, каскадом ADF-I (PnP/DPS), генерацією $V = 4$ віртуальних ракурсів (PTP) та $T = 5$ ітераціями дифузії ADF-F.

Висновки. У роботі запропоновано метод ідентифікації особи у відкритих джерелах AFRW-X (modified) на основі зображень низької якості з урахуванням диференційного ракурсу та шумових завад, що інтегрує адаптивну дифузійну фільтрацію ADF на рівні зображень та ознак із механізмом псевдо-3D проєкцій PTP для мультиракурсного злиття векторних представлень.

Експериментальна валідація на чотирьох еталонних датасетах підтвердила суттєву перевагу запропонованого методу в умовах множинної деградації вхідних зображень. Так, при критичному рівні адитивного гаусівського шуму, метод AFRW-X (modified) забезпечує високу точність розпізнавання при вкрай низькому рівні хибних спрацювань, що суттєво перевищує показники базового ArcFace та оригінальної версії AFRW-X. Ключовим показником є

незначне відносне падіння точності за умов сильного шуму, що задовольняє встановлену вимогу про допустиме зниження та підтверджує високу робастність методики.

У випадку комплексної деградації, що поєднує всі негативні фактори, запропонований метод досягає високої точності розпізнавання при низькому рівні хибних спрацювань та високого показника інтегральної метрики, що значно перевищує показники ArcFace. Необхідно зазначити, що приріст точності досягається за рахунок очікуваного збільшення обчислювальної складності. Метод вимагає більше обчислювальних операцій та оперує більшою кількістю параметрів, що призводить до збільшення часу обробки порівняно з ArcFace. Тим не менш, цей час залишається в межах допустимих значень для серверних застосувань, де пріоритетом є якість, а не миттєва реакція.

Такий підхід забезпечує комплексну компенсацію множинної деградації, на відміну від існуючих методів, що розглядають окремі негативні фактори ізольовано.

Наукова новизна запропонованої методики AFRW-X (modified), що відрізняє від існуючих підходів, полягає у комплексній інтеграції трьох ключових інновацій.

По-перше, запропоновано адаптивний каскадний механізм дифузійної фільтрації на двох рівнях абстракції. На рівні вхідних зображень спеціальний модуль (ADF-I) застосовує послідовність різних методів відновлення, автоматично обираючи кращий варіант залежно від рівня деградації. На рівні глибоких ознак інший модуль (ADF-F) використовує анізотропну дифузію Перона – Маліка (протягом п’яти ітерацій) для вибіркового згладжування карт ознак. Такий процес адаптивно згладжує ознаки, опираючись на їх локальні зміни, що дозволяє зберегти важливі межі між семантичними областями.

По-друге, розроблено удосконалений модуль псевдо-3D проєкцій (PTP) із чотирма віртуальними ракурсами. Модуль використовує точнішу однокадрову 3D-реконструкцію замість простого усереднення, він застосовує адаптивне зважування ембеддінгів для кожного ракурсу, що базується на трикомпонентній метриці, яка враховує видимість, перетин областей та характеристики ознак. Крім того, застосовується спеціальна функція втрат для мультиракурсної консистентності, яка мінімізує розбіжності між представленнями з різних ракурсів під час навчання, спонукаючи мережу вивчати ознаки, які інваріантні до зміни ракурсу.

По-третє, запропоновано потужнішу опорну архітектуру на основі ResNet-100 з інтегрованими механізмами канальної уваги та композитною функцією втрат, яка одночасно оптимізує кутову маржу, тріплетні обмеження, мультиракурсну консистентність та прогнозування якості. Розмірність ембеддінгу було збільшено до 512 для підвищення диференціальної здатності. Такий підхід забезпечує комплексну компенсацію множинної деградації, на відміну від існуючих методів, що розглядають окремі негативні фактори ізольовано.

Результати дослідження підтверджують придатність запропонованого методу AFRW-X (modified) для застосування в реальних OSINT-системах ідентифікації особи, де характерною є одночасна присутність низької роздільної здатності, шумових завад, нефронтальних ракурсів та артефактів стиснення зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Taigman Y., Yang M., Ranzato M.-A., Wolf L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification // Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2014. С. 1701–1708. DOI: 10.1109/CVPR.2014.220.

2. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering // Proc. CVPR. Piscataway, NJ: IEEE, 2015. С. 815–823. DOI: 10.1109/CVPR.2015.7298682.

3. Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition // Proc. CVPR. Piscataway, NJ: IEEE, 2019. С. 4690–4699. DOI: 10.1109/CVPR.2019.00482.

4. Wang H., Wang Y., Zhou Z., Ji X., Gong D., Zhou J., Li Z., Liu W. CosFace: Large Margin Cosine Loss for Deep Face Recognition // Proc. CVPR. Piscataway, NJ: IEEE, 2018. С. 5265–5274. DOI: 10.1109/CVPR.2018.00449.

5. Liu W., Wen Y., Yu Z., Li M., Raj B., Song L. SphereFace: Deep Hypersphere Embedding for Face Recognition // Proc. CVPR. Piscataway, NJ: IEEE, 2017. С. 212–220. DOI: 10.1109/CVPR.2017.414.

6. Wen Y., Zhang K., Li Z., Qiao Y. A Discriminative Feature Learning Approach for Deep Face Recognition (Center Loss) // Proc. ECCV. Cham: Springer, 2016. С. 499–515. DOI: 10.1007/978-3-319-46478-7_31.

7. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proc. CVPR. Piscataway, NJ: IEEE, 2016. С. 770–778. DOI: 10.1109/CVPR.2016.90.

8. Cao Q., Shen L., Xie W., Parkhi O. M., Zisserman A. VGGFace2: A Dataset for Recognising Faces across Pose and Age // Proc. 2018 13th IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG). Piscataway, NJ: IEEE, 2018. С. 67–74. DOI: 10.1109/FG.2018.00020.

9. Huang G. B., Ramesh M., Berg T., Learned-Miller E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Amherst, MA: Univ. of Massachusetts, Tech. Rep. UM-CS-2007-49, 2007. 26 с. URL: http://vis–www.cs.umass.edu/lfw/ (дата звернення: 06.10.2025).

10. Kemelmacher–Shlizerman I., Seitz S. M., Miller D., Brossard E. The MegaFace Benchmark: 1 Million Faces for Recognition at Scale // Proc. CVPR. Piscataway, NJ: IEEE, 2016. С. 4873–4882. DOI: 10.1109/CVPR.2016.527.

11. Whitelam C., Taborsky E., Blanton A., Maze B., Adams J., Miller T., Kalka N., Jain A. K., Duncan J. A., Allen K., Cheney J., Grother P. IARPA Janus Benchmark–C: Face Dataset and Protocol // Proc. 2018 Int. Conf. on Biometrics (ICB). Piscataway, NJ: IEEE, 2018. С. 158–165. DOI: 10.1109/ICB2018.2018.00033.

12. Grother P., Ngan M., Hanaoka K. Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. Gaithersburg, MD: NIST, NISTIR 8280, 2019. 44 с. DOI: 10.6028/NIST.IR.8280.

13. Woo S., Park J., Lee J.–Y., Kweon I. S. CBAM: Convolutional Block Attention Module // Proc. ECCV. Cham: Springer, 2018. С. 3–19. DOI: 10.1007/978-3-030-01234-2_1.

14. Hu J., Shen L., Sun G. Squeeze–and–Excitation Networks // Proc. CVPR. Piscataway, NJ: IEEE, 2018. С. 7132–7141. DOI: 10.1109/CVPR.2018.00745.

15. Hou Q., Zhou D., Feng J. Coordinate Attention for Efficient Mobile Network Design // Proc. CVPR Workshops. Piscataway, NJ: IEEE, 2021. С. 0–0. arXiv:2103.02907.

16. Dabov K., Foi A., Katkovnik V., Egiazarian K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering (BM3D) // IEEE Trans. Image Processing. 2007. Т. 16, № 8. С. 2080–2095. DOI: 10.1109/TIP.2007.901238.

17. Rudin L. I., Osher S., Fatemi E. Nonlinear Total Variation Based Noise Removal Algorithms // Physica D. Amsterdam: Elsevier, 1992. Т. 60, № 1–4. С. 259–268. DOI: 10.1016/0167-2789(92)90242-F.

18. Goldstein T., Osher S. The Split Bregman Method for L1-Regularized Problems // SIAM Journal on Imaging Sciences. 2009. Т. 2, № 2. С. 323–343. DOI: 10.1137/080725891.

19. Boyd S., Parikh N., Chu E., Peleato B., Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers // Foundations and Trends in Machine Learning. Boston: Now Publishers, 2011. Т. 3, № 1. С. 1–122. DOI: 10.1561/2200000016.

20. Venkatakrishnan S. V., Bouman C. A., Wohlberg B. Plug–and–Play Priors for Model Based Reconstruction // Proc. 2013 IEEE GlobalSIP. Piscataway, NJ: IEEE, 2013. С. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048.

21. Romano Y., Elad M., Milanfar P. Regularization by Denoising: Clarifications and New Interpretations (RED) // SIAM Journal on Imaging Sciences. 2017. Т. 10, № 3. С. 1804–1844. DOI: 10.1137/16M1102879.

22. Zhang K., Zuo W., Chen Y., Meng D., Zhang L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising (DnCNN) // IEEE Trans. Image Processing. 2017. Т. 26, № 7. С. 3142–3155. DOI: 10.1109/TIP.2017.2662206.

23. Perona P., Malik J. Scale–Space and Edge Detection Using Anisotropic Diffusion // IEEE Trans. Pattern Analysis and Machine Intelligence. 1990. Т. 12, № 7. С. 629–639. DOI: 10.1109/34.56205.

24. Canny J. A Computational Approach to Edge Detection // IEEE Trans. PAMI. 1986. Т. 8, № 6. С. 679–698. DOI: 10.1109/TPAMI.1986.4767851.

25. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. 2–ге вид. Cambridge: Cambridge University Press, 2003. 672 с. ISBN: 978-0521540513.

26. Blanz V., Vetter T. A Morphable Model for the Synthesis of 3D Faces // Proc. SIGGRAPH. New York: ACM, 1999. С. 187–194. DOI: 10.1145/311535.311556.

27. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features // Proc. CVPR. Piscataway, NJ: IEEE, 2001. С. 511–518. DOI: 10.1109/CVPR.2001.990517.

28. Kazemi V., Sullivan J. One Millisecond Face Alignment with an Ensemble of Regression Trees // Proc. CVPR. Piscataway, NJ: IEEE, 2014. С. 1867–1874. DOI: 10.1109/CVPR.2014.241.

29. Liu Z., Luo P., Wang X., Tang X. Deep Learning Face Attributes in the Wild (CelebA) // Proc. ICCV. Piscataway, NJ: IEEE, 2015. С. 3730–3738. DOI: 10.1109/ICCV.2015.425.

30. Meng Q., Zhao S., Huang Z., Zhou F. MagFace: A Universal Representation for Face Recognition and Quality Assessment // Proc. CVPR. Piscataway, NJ: IEEE, 2021. С. 14225–14234. DOI: 10.1109/CVPR46437.2021.01399.

31. Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S. Class-Balanced Loss Based on Effective Number of Samples // Proc. CVPR. Piscataway, NJ: IEEE, 2019. С. 9268–9277. DOI: 10.1109/CVPR.2019.00949.