

**Чичкарьов Євген Анатолійович**

доктор технічних наук, професор, професор кафедри комп'ютерних наук
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0000-0002-4362-5129
ye.chyckarov@nubip.edu.ua

Веремійчук Іван Андрійович

аспірант, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
ORCID: 0009-0009-2279-7659
i.veremiichuk@gmail.com

РОЗПІЗНАВАННЯ 3D ЗОБРАЖЕНЬ ОБЛИЧЧЯ З ВИКОРИСТАННЯМ КВАТЕРНІОННИХ НЕЙРОННИХ МЕРЕЖ

Анотація. Статтю присвячено розв'язанню актуальної науково-прикладної задачі комп'ютерного зору — підвищенню точності оцінювання 3D-орієнтації (пози) обличчя в просторі та зменшенню параметричної складності глибоких нейромережових моделей за допомогою апарату кватерніонної алгебри. Досліджено обмеження дійснозначних згорткових нейронних мереж (CNN), які обробляють координати та колірні канали незалежно, втрачаючи внутрішню міжкомпонентну кореляцію та призводячи до деградації точності за наявності значних поворотів голови й оклюзій. Запропоновано та математично обґрунтовано використання кастомних кватерніонних згорткових шарів (QuaternionConv2D), побудованих на основі некомутативного добутку Гамільтона, що дає змогу моделювати просторово-кутові взаємозв'язки сигналу й скорочувати кількість вільних вагових параметрів у 4 рази проти класичних згортків. Для стабільного навчання моделей розроблено зважену адаптивну функцію втрат Huber Loss у радіанному вимірі з ваговим акцентом на найбільш варіативний кут Yaw та застосовано L2-регуляризацію ваг. Проведено систематичний порівняльний експериментальний аналіз трьох типів архітектур, суто кватерніонної мережі (Stable QCNN) та гібридної моделі (MobileNetV2 + QCNN Hybrid). На основі отриманих результатів доведено, що гібридна архітектура забезпечує найкращу точність розпізнавання, знижуючи валідаційну похибку MAE до 0.1535 rad (8.80°), що на 27.39% перевищує показники базової MobileNetV2 (12.12°) при незначному зростанні обсягу параметрів лише на 2.48%. Водночас надлегка суто кватерніонна модель Stable QCNN демонструє екстремальну компактність (0.04M параметрів, що у 60.5 разів менше за CNN MobileNetV2) і на 40.3% менший час навчання, виступаючи перспективним рішенням для ресурсообмежених Edge AI пристроїв.

Ключові слова: розпізнавання 3D-обличчя, оцінювання пози голови, кватерніонні нейронні мережі (QCNN), добуток Гамільтона, гібридні архітектури, MobileNetV2, Huber Loss, L2-регуляризація, Edge AI.

ВСТУП

Розпізнавання обличчя є ключовою задачею комп'ютерного зору у сферах біометрії, відеоаналітики та інтерфейсів доповненої й віртуальної реальності. Перехід від 2D до 3D-представлень суттєво підвищує стійкість методів до змін освітлення, ракурсів і поворотів голови. Водночас 3D-дані породжують нові виклики: обмеженість анотованих датасетів, нерегулярність вхідних структур (точкові хмари, карти глибини) та потребу одночасного врахування просторових координат (X, Y, Z) і кутової орієнтації обличчя.

Незважаючи на значний прогрес у галузі глибокого навчання, задача стійкого визначення просторового положення голови (head pose estimation, HPE) в умовах необмеженого середовища залишається відкритою науковою проблемою — насамперед через складні сценарії оклюзій, широкий діапазон кутів обертання та принципові математичні обмеження традиційних способів представлення тривимірних обертань [1]. Систематичний огляд методів глибокого навчання для задачі HPE [1] свідчить,



що саме проблема коректного представлення обертань у тривимірному просторі є ключовим чинником, що обмежує точність сучасних моделей в умовах реального застосування.

Більшість методів розпізнавання 3D-облич засновані на дійснозначних CNN, які обробляють координати чи канали незалежно або через просту конкатенацію. Це не враховує внутрішню алгебраїчну кореляцію компонентів і збільшує кількість параметрів. Альтернативою є кватерніонні мережі (QNN), де точки чи пікселі кодуються кватерніонами, а згортка через некомутативний добуток Гамільтона природно моделює просторові та кутові взаємозв'язки вхідного сигналу [2]. Оскільки кватерніони історично є стандартним математичним апаратом для опису обертань у тривимірному просторі, їх застосування є особливо доречним саме для задач, пов'язаних із розпізнаванням обличчя за довільної просторової орієнтації голови [1].

Класичний підхід до HPE ґрунтується на використанні кутів Ейлера (Yaw–Pitch–Roll), що безпосередньо відповідають інтуїтивному уявленню про просторову орієнтацію голови. Проте кути Ейлера мають відомі фундаментальні недоліки: ефект карданного замкнення (gimbal lock) — сингулярність, при якій два з трьох кутів відповідають одній і тій самій просторовій орієнтації і система втрачає один ступінь свободи; порушення неперервності представлення при переході через кутові границі; та нелінійні спотворення метрики при великих кутах відхилення.

Огляд сучасного стану розробки кватерніонних CNN (QCNN) [3] підтверджує, що такі архітектури здатні досягати порівнянної з дійснозначними CNN точності при скороченні кількості вагових параметрів до 75% завдяки повторному використанню компонентів у матриці Гамільтона.

Практична цінність кватерніонного підходу підтверджується й тим, що заміна дійсних вагових матриць кватерніонними Гамільтоновими добутками, а також більш загальне параметризоване гіперкомплексне множення (Parameterized Hypercomplex Multiplication, PHM), дозволяють скорочувати кількість навчальних параметрів мережі у кілька разів за збереження або навіть покращення точності розпізнавання [4]. Це створює передумови для побудови компактних, обчислювально ефективних моделей розпізнавання обличчя, придатних до розгортання на пристроях з обмеженими ресурсами, — властивість, яка традиційно досягається класичними легкими архітектурами (MobileNetV2, DenseNet121) ціною певної втрати точності. Разом із тим системне порівняння суто кватерніонних, гібридних (кватерніонно-дійсних) та класичних легких CNN-архітектур безпосередньо на задачі розпізнавання 3D-зображення обличчя, з використанням усталених еталонних наборів даних, у науковій літературі представлено фрагментарно: більшість опублікованих кватерніонних моделей орієнтовані або на обробку кольорових 2D-зображень [5], або на розпізнавання тривимірних облич за традиційними точковими чи геометричними ознаками без застосування кватерніонного кодування просторових координат [6].

Першою практичною реалізацією QCNN для задачі HPE стала архітектура QuatNet, яка використовує мультирегресійну функцію втрат на основі кватерніонного представлення та досягає точності на рівні найкращих методів того часу на датасеті AFLW2000-3D. Подальший розвиток кватерніонного підходу представлено в роботі [7], де запропонована модель WQuatNet передбачає просторову орієнтацію голови в повному діапазоні 360° на основі кватерніонного виходу та двох спеціалізованих функцій втрат, що враховують геометрію $SO(3)$ -многовиду, і демонструє перевищення показників точності над методами класу «state-of-the-art» на бенчмарках AFLW2000-3D, BIWI, CMU та AGORA.

Основним бенчмарком для оцінки методів HPE та 3D-аналізу обличчя є датасет AFLW2000-3D [1], що містить перші 2 000 зображень бази AFLW (Annotated Facial Landmarks in the Wild), ре-анотованих 68 тривимірними орієнтирами обличчя за допомогою параметричної 3D-моделі. Кути Yaw варіюють у діапазоні $\pm 120^\circ$, Roll та Pitch — $\pm 90^\circ$, що робить датасет репрезентативним для сценаріїв з екстремальними ракурсами та виразними неканонічними позами. Метрикою оцінки є середня абсолютна похибка (MAE) кутів Yaw, Pitch та Roll у градусах. Додатковим бенчмарком є датасет BIWI Head Pose Dataset [1], сформований у лабораторних умовах шляхом запису RGB-D відео 20 суб'єктів за допомогою камери Kinect з охопленням $\pm 75^\circ$ по Yaw, $\pm 60^\circ$ по Pitch та $\pm 50^\circ$ по Roll. Поєднання цих двох датасетів забезпечує повноцінну оцінку методів як в умовах варіативного природного середовища (AFLW2000-3D), так і в умовах контрольованого простору з точними глибинними анотаціями (BIWI).

З позицій практичного застосування критичною вимогою для систем аналізу обличчя є ефективність розгортання на ресурсообмежених пристроях. Архітектури DenseNet121 та MobileNetV2 є широко застосовуваними базовими моделями для задач HPE: перша забезпечує виражену параметричну ефективність за рахунок щільних (dense) з'єднань між шарами, друга — компактність завдяки глибинно-розподільній (depthwise separable) згортці, що робить її пріоритетною для мобільних застосувань. Дослідження [8] стверджує, що стандартні CNN на основі цих архітектур демонструють суттєве зниження точності в умовах великих кутів нахилу та при обробці зображень обличчя з оклюзіями. Водночас



ефективне кодування кольору та просторової орієнтації через кватерніонне представлення дозволяє уникнути деградації кутової точності, характерної для дійснозначних моделей [5], [8].

Незважаючи на продемонстровані переваги, в науковій літературі відсутній комплексний порівняльний аналіз, який би одночасно охоплював: (1) повноцінно кватерніонні CNN (QCNN) з гамільтоновою згортою; (2) гібридні архітектури, де кватерніонне представлення орієнтації поєднується зі стандартними дійсними базовими мережами (DenseNet121, MobileNetV2); та (3) кількісне порівняння за метриками MAE та параметричної ефективності на датасетах AFLW2000-3D і BIWI Head Pose Dataset.

Метою даної роботи є розробка та порівняльне дослідження кватерніонних і гібридних неймережевих архітектур для розпізнавання 3D-зображень обличчя — з акцентом на стійкість до широкого діапазону орієнтацій голови, параметричну ефективність та точність передбачення просторового положення за стандартними бенчмарками.

Постановка проблеми

Задача розпізнавання просторової орієнтації обличчя за тривимірними даними є однією з ключових у сучасних системах комп'ютерного зору, біометричної ідентифікації, доповненої реальності та людино-машинної взаємодії.

Традиційні згорткові нейронні мережі (CNN), оперуючи дійснозначними тензорами, обробляють просторові координати (x, y, z) та кутові параметри орієнтації обличчя (pitch, yaw, roll) як незалежні канали. Це призводить до втрати структурної інформації, важливої для точного відтворення тривимірної форми та орієнтації обличчя, особливо за умов часткових оклюзій, значних поворотів голови та варіативного освітлення.

Кватерніонне представлення даних дозволяє природним чином кодувати тривимірні координати та обертання в єдиному алгебраїчному об'єкті, зберігаючи міжканальні кореляції під час згорткових операцій. Це обумовлює доцільність застосування кватерніонних нейронних мереж (QCNN) до задач, де просторова структура даних відіграє визначальну роль.

Нехай на вході системи подається зображення обличчя $I \in \mathbb{R}^{H \times W \times C}$ (або відповідне кватерніонне представлення $Q \in \mathbb{H}^{H \times W}$, де кожен піксель кодується кватерніоном $q = a + bi + cj + dk$). Необхідно побудувати відображення $f: I \rightarrow (\text{pitch}, \text{yaw}, \text{roll})$, яке апроксимує кутові параметри орієнтації обличчя з мінімальною середньою абсолютною похибкою (MAE), одночасно мінімізуючи кількість параметрів моделі та обчислювальну складність.

Дослідження проводиться на двох наборах даних:

1. AFLW2000-3D — набір даних, що містить 2000 зображень облич із анотованими 3D-орієнтирами (landmarks) та кутами повороту голови, отриманими шляхом підгонки тривимірної морфованої моделі (3DMM). Використовується як первинна база для навчання та тестування моделей.
2. BIWI Head Pose Dataset — набір даних, що містить послідовності кадрів RGB-D із точними анотаціями кутів орієнтації голови, отриманими за допомогою сенсора глибини Kinect. Застосовується для перевірки узагальнювальної здатності моделей у режимі крос-доменного тестування (навчання на AFLW2000-3D, валідація на BIWI, і навпаки), а також для оцінки стійкості моделей до відмінностей у методах анотування та умовах зйомки.

Ключова наукова проблема, яку розв'язує дослідження, полягає у відсутності систематичного порівняльного аналізу ефективності класичних CNN-архітектур (DenseNet121, MobileNetV2) і кватерніонних (чисто кватерніонних та гібридних) моделей у задачі оцінювання кутової орієнтації голови за критеріями точності, обчислювальної складності та стійкості до зміни домену даних.

Аналіз останніх досліджень і публікацій

Задача розпізнавання 3D-зображень обличчя традиційно спирається на глибокі згорткові нейронні мережі (CNN), що працюють із дійснозначними тензорами, у яких колірні канали (RGB) або просторові координати (X, Y, Z) обробляються або незалежно, або шляхом їх конкатенації в єдиний вектор ознак. Такий підхід ігнорує внутрішню структурну кореляцію між компонентами багатовимірного сигналу. Кватерніонні нейронні мережі (Quaternion Neural Networks, QNN) пропонують альтернативу: кожен піксель або 3D-точка кодується як єдиний кватерніон, а операція згортки виконується через некомутативний добуток Гамільтона. Це дозволяє мережі природним чином враховувати взаємозв'язки між компонентами сигналу (кольорними каналами чи просторовими координатами X, Y, Z) і суттєво скорочує кількість навчальних параметрів. За останні роки інтерес до застосування кватерніонних та гібридних (кватерніонно-дійсних) архітектур у задачах розпізнавання облич, оцінювання пози голови та обробки 3D-даних істотно зріс, що й обумовлює актуальність цього дослідження.

Систематизація напряму QNN розпочалася з огляду Т. Парколле зі співавторами [9], який узагальнив архітектурні принципи кватерніонних повнозв'язних, згорткових та рекурентних шарів. Автори показали, що заміна дійсних вагових матриць кватерніонними добутками Гамільтона забезпечує скорочення кількості параметрів приблизно в чотири рази за збереження або покращення точності в



задачах обробки мовлення та зображень. Подальші огляди [2], [10] розширили цю систематизацію, охопивши кватерніонні згорткові, рекурентні, LSTM-, варіаційні автокодувальні архітектури та механізми самоуваги (self-attention), а також окреслили відкриті виклики галузі — відсутність стандартизованих кватерніонних шарів у популярних фреймворках глибокого навчання, складність кватерніонного диференціювання та обмежену кількість готових преднавчаних моделей.

Окремий напрям теоретичних досліджень присвячений оптимізації навчання кватерніонних мереж. Зокрема, у роботі [11] запропоновано метаевристичні (похідно-вільні) методи навчання, що обходять необхідність застосування аналітичної кватерніонної функції втрат. Робота [4], відзначена як видатна на ICLR 2021, узагальнила добуток Гамільтона до параметризованого гіперкомплексного множення (Parameterized Hypercomplex Multiplication, PHM), яке дозволяє мережі самостійно навчатися правил множення для довільної розмірності гіперкомплексного простору, не обмежуючись лише кватерніонами. Також було показано [12], що включення PHM-шарів у класифікаційну (backend) частину класичних архітектур ResNet додатково підвищує точність класифікації зображень порівняно з їх дійсними та «чистими» кватерніонними аналогами. Це є прямим свідченням на користь гібридних (дійсно-кватерніонних) архітектур, які досліджуються і в цій статті.

Прикладне застосування QNN до задачі розпізнавання облич розвивається за кількома напрямками. El Alamі зі співавторами послідовно розробили серію легких архітектур на основі кватерніонних дискретних ортогональних моментів: у роботі [13] запропоновано мережу, що поєднує кватерніонні моменти з неглибокою CNN для розпізнавання кольорових облич, а в подальшому [5] цей підхід узагальнено до моментів Хана (Hahn moments) у складі повноцінної кватерніонної CNN (QHMCNN). Це дало змогу одночасно зменшити обчислювальну складність мережі та підвищити точність класифікації на кількох еталонних наборах кольорових зображень і облич.

У напрямі підвищення стійкості до перешкод Wen і Qiu [14] запропонували кватерніонну мережу з немипуклими розрідженими обмеженнями (QNSPCANet) для розпізнавання облич за умов часткової оклюзії, довівши перевагу кватерніонного представлення кольору над пороздільною обробкою RGB-каналів саме в умовах втрати частини вхідних даних. Для задачі розпізнавання емоцій за кольоровим зображенням обличчя Zhou зі співавторами [15] інтегрували кватерніонні згортки з механізмом уваги на основі фільтрів Габора (QA-CNN), продемонструвавши, що кватерніонне кодування кольору дає змогу відображенню уваги враховувати кореляцію між кольорними каналами, чого не досягають традиційні одноканальні механізми уваги. Ще один напрям — використання кватерніонного представлення для генеративних задач: Liu, Chen і Li [16] запропонували графову кватерніонну модель для галюцинації (реконструкції) кольорових зображень облич низької роздільної здатності, що також підтверджує ефективність кватерніонного кодування кольору для збереження структурної узгодженості обличчя.

Хоча більшість наведених вище робіт стосуються двовимірних кольорових зображень облич, кватерніонне представлення природно узагальнюється і на тривимірні дані, оскільки кватерніони є стандартним математичним апаратом для опису обертань у 3D-просторі. Zhao зі співавторами [17] запропонували еквіваріантну відносно обертань капсульну мережу (QEC-Net), у якій динамічна маршрутизація капсул виконується безпосередньо в кватерніонному просторі над локальними системами відліку, обчисленими з точкової хмари. Це забезпечує end-to-end еквіваріантність до обертань об'єкта — властивість, критично важливу саме для розпізнавання 3D-облич за довільної пози голови.

Паралельно розвивається напрям класичних (дійснозначних) методів розпізнавання 3D-облич за точковими хмарами: огляд Li, Huang і Tian [6] систематизував традиційні геометричні методи та методи глибокого навчання для 3D-розпізнавання облич, окресливши основні виклики галузі — обмежений обсяг публічно доступних 3D-наборів даних, нерегулярну (неструктуровану) природу точкових хмар та чутливість до неригідних деформацій обличчя (міміки). Розвиваючи цей напрям, You і Zhao [18] запропонували двогілкову мережу на основі кернел-точкових згорток (KPCConv) для прямого вилучення неригідних ознак обличчя з точкової хмари, використавши згенеровані на основі моделі гауссівських процесів (GPMM) віртуальні 3D-скани для подолання дефіциту реальних тренувальних даних. Цей підхід ілюструє загальну для галузі проблему обмеженості 3D-наборів (зокрема й AFLW2000-3D, що налічує лише 2000 анотованих зображень) та необхідність створення ефективних із погляду використання даних архітектур, якими потенційно є і кватерніонні мережі завдяки меншій кількості параметрів.

Набір AFLW2000-3D, що містить перші 2000 анотованих зображень із набору AFLW із точними тривимірними позначками пози (кути ристання — *yaw*, тангажу — *pitch* та крену — *roll*), а також набір BIWI Head Pose Dataset, зібраний за допомогою RGB-D сенсора Kinect, є двома найпоширенішими еталонними наборами для порівняння методів оцінювання пози голови та суміжних задач вирівнювання й розпізнавання 3D-обличчя.

Основоположною роботою, що безпосередньо поєднує кватерніонне представлення з CNN для цієї задачі, є QuatNet [19]: автори сформулювали задачу оцінювання пози голови в просторі кватерніонів



(замість традиційних кутів Ейлера, що мають проблему складання рамкових підвісів / Gimbal Lock і розривність представлення) і запропонували комбіновану функцію втрат (L_2 -регресія з порядковою регресією). Це дало змогу досягти найкращих на момент публікації результатів на AFLW2000, AFLW test set і AFW, суттєво скоротивши розрив із методами, що використовують глибинну інформацію на BIWI.

Систематизований огляд Asperti й Filippini [1] проаналізував понад сто сучасних методів оцінювання пози голови, включно з кватерніонними, та навіть порівняльні таблиці середньої абсолютної похибки (MAE) за кутами yaw/pitch/roll на AFLW2000-3D і BIWI для найкращих архітектур (Hopenet, FSA-Net, QuatNet, WHENet, TriNet, 6DRepNet, SynergyNet та інших). Автори показали, що кватерніонне представлення виходу залишається ефективною альтернативою представленню через матрицю обертання чи неперервні 6D-репрезентації, хоча й поступається останнім за точністю в задачах повного діапазону обертання (360°). Розвиваючи напрям представлень без розривності, Hempel зі співавторами [8] запропонували 6D-представлення обертання для необмеженого діапазону кутів пози голови, що на сьогодні є одним із найточніших підходів на комбінованому бенчмарку AFLW2000-3D/BIWI і слугує орієнтиром для порівняння з кватерніонними та гібридними моделями в цій статті.

Для коректного зіставлення продуктивності кватерніонних і суто дійсних мереж у цьому дослідженні як базові (baseline) архітектури обрано DenseNet121 та MobileNetV2 — дві поширені та архітектурно контрастні дійснозначні CNN. DenseNet121, запропонована Huang зі співавторами, використовує щільні зв'язки (dense connections) між усіма шарами всередині блоку, що покращує розповсюдження градієнта та повторне використання ознак за рахунок відносно високої обчислювальної вартості. MobileNetV2, запропонована Sandler зі співавторами, натомість базується на розділюваних за глибиною згортках (depthwise separable convolutions) та обернених залишкових блоках (inverted residuals) і орієнтована на мобільні та вбудовані обчислювальні платформи.

Порівняльне дослідження Liu, Zhou і Chen [20] показало, що архітектури, побудовані на основі MobileNetV2 (зокрема родина MobileFaceNet), за умови додаткової оптимізації структури та режиму навчання здатні наблизитися за точністю розпізнавання облич до значно важчих мереж при суттєво меншій кількості параметрів і обчислювальних витрат. Це робить порівняння «легка дійсна CNN — кватерніонна/гібридна CNN» особливо показовим саме з погляду співвідношення точності та обчислювальної ефективності. Водночас основоположна робота Gaudet і Maida [21], у якій уперше систематично розроблено алгоритми кватерніонної згортки, кватерніонної ініціалізації ваг і кватерніонної пакетної нормалізації (quaternion batch normalization) для «глибоких» кватерніонних мереж, залишається технічною основою більшості сучасних кватерніонних CNN, включно з архітектурами, розглянутими в п. 1.2–1.4.

Проведений аналіз показує, що дослідження в галузі кватерніонних нейронних мереж розвиваються за трьома відносно відокремленими напрямками:

1. Розпізнавання кольорових 2D-облич [5], [13], [14], [15], [16], де кватерніон кодує кореляцію RGB-каналів;
2. Обробка тривимірних точкових хмар і оцінювання пози [6], [17], [18], де кватерніон природно кодує обертання об'єкта в просторі;
3. Регресія пози голови [1], [8], [19], де кватерніон використовується як вихідне представлення мережі, а не як внутрішнє кодування вхідних даних.

Водночас пряме порівняння «чисто» кватерніонних та гібридних (кватерніонно-дійсних) CNN-архітектур із класичними легкими мережами (DenseNet121, MobileNetV2) саме на задачі розпізнавання 3D-зображення обличчя з використанням еталонного набору AFLW2000-3D (як основного) та BIWI Head Pose Dataset (як додаткового, для перевірки узагальнювальної здатності на іншому типі сенсора — RGB-D Kinect) у проаналізованій літературі представлено фрагментарно. Більшість кватерніонних робіт зосереджені або на 2D-кольорових, або на суто 3D-точкових даних, тоді як робіт, що поєднують кватерніонне кодування просторових 3D-координат обличчя з прямим кількісним порівнянням проти сучасних легких дійснозначних CNN-баз на AFLW2000-3D/BIWI, виявлено не було.

Це й визначає дослідницьку прогалину, на заповнення якої спрямована ця стаття: систематичне порівняння чисто кватерніонної CNN, гібридної (кватерніонно-дійсної) моделі та класичних легких CNN-архітектур (DenseNet121, MobileNetV2) за точністю розпізнавання, кількістю параметрів та обчислювальною ефективністю на датасетах AFLW2000-3D і BIWI.

Мета статті

Метою дослідження є підвищення точності оцінювання 3D-орієнтації обличчя та зменшення параметричної складності нейромережових моделей шляхом розробки, математичного обґрунтування та системного порівняльного аналізу архітектур дійсних (CNN), чистих кватерніонних (наскрізних QCNN) та гібридних згорткових нейронних мереж.

Для досягнення поставленої мети необхідно розв'язати такі завдання:



1. Здійснити аналіз сучасних наукових підходів до розпізнавання 3D-орієнтації обличчя за допомогою класичних дійснозначних CNN та кватерніонних нейронних мереж.
2. Обґрунтувати та сформулювати експериментальну базу даних шляхом системного узгодження параметрів тривимірної орієнтації (кутів повороту голови) для наборів даних AFLW2000-3D та BIWI Head Pose Dataset, інтегрувавши процедури фотометричної регуляризації для підвищення узагальнювальної здатності моделей без спотворення їхньої 3D-геометрії.
3. Розробити та адаптувати архітектури дійсних згорткових мереж (DenseNet121, MobileNetV2), чисто кватерніонної мережі (QCNN) та гібридної (кватерніонно-дійсної) моделі для задачі регресії 3D-орієнтації голови.
4. Провести експериментальне навчання моделей на наборі AFLW2000-3D із використанням геодезичного функціоналу втрат та оцінити їхню точність (середню абсолютну похибку окремо для кутів тангажу, ризання і крену та їх узагальнене середнє значення), параметричну складність (кількість навчальних ваг) і швидкодію (час навчання та затримку інференсу).
5. Для оцінки стійкості, робастності та узагальнювальної здатності моделей за різних умов зйомки та типів сенсорів виконати їхнє перехресне оцінювання на незалежних наборах даних AFLW2000-3D та BIWI Head Pose Dataset.
6. Здійснити порівняльний аналіз отриманих результатів, визначити компроміс між параметричною компактністю та точністю, сформулювати практичні рекомендації щодо доцільності застосування кожного підходу залежно від обчислювальних обмежень, а також визначити наукову новизну й практичну значущість дослідження.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Математичний апарат кватерніонних шарів (QCNN)

Кастомний шар QuaternionConv2D виконував згортку у чотиривимірному гіперкомплексному просторі $\mathbf{H}$.

Операція згортки реалізує добуток Гамільтона між кватерніонним входом $X = X_r + X_i \cdot i + X_j \cdot j + X_k \cdot k$ та кватерніонним ядром $W = W_r + W_i \cdot i + W_j \cdot j + W_k \cdot k$:

$$\begin{bmatrix} C_r \\ C_i \\ C_j \\ C_k \end{bmatrix} = \begin{bmatrix} W_r & -W_i & -W_j & -W_k \\ W_i & W_r & -W_k & W_j \\ W_j & W_k & W_r & -W_i \\ W_k & -W_j & W_i & W_r \end{bmatrix} \begin{bmatrix} X_r \\ X_i \\ X_j \\ X_k \end{bmatrix} \quad (1)$$

Завдяки такій структурі матриці ваг кількість вільних параметрів зменшується в 4 рази порівняно зі стандартною згорткою аналогічної глибини.

Для збереження геометричної норми компонент був застосований кастомний шар QuaternionNormalize:

$$q_{\text{norm}} = \frac{q}{\sqrt{\sum_{c \in \{r,i,j,k\}} q_c^2 + \epsilon}}, \quad \epsilon = 10^{-6} \quad (2)$$

Простір цільових змінних та функція втрат

Регресія кутів орієнтації — Pitch (тангаж), Yaw (рыскание) та Roll (крен) — виконувалася безпосередньо у радіанах (rad) у просторі $[-\pi/2, +\pi/2]$. Цільові значення витягувались з анотаційних матриць .mat. Для стабілізації процедури регресії та забезпечення робастності до екстремальних поворотів було застосовано взважену адаптивну функцію втрат Huber Loss у радіанному вимірі:

$$L_{\text{Huber}}(y, \hat{y}) = \frac{1}{3} \sum_{m \in \{P,Y,R\}} w_m \cdot h_m(y_m, \hat{y}_m) \quad (3)$$

$$h_m(y_m, \hat{y}_m) = \begin{cases} 0.5 \cdot (y_m - \hat{y}_m)^2, & \text{якщо } |y_m - \hat{y}_m| \leq \delta \\ \delta \cdot (|y_m - \hat{y}_m| - 0.5 \cdot \delta), & \text{інакше} \end{cases} \quad (4)$$

де:

$\delta = 0.0872$ rad (еквівалент 5°) — поріг переходу від квадратичної до лінійної функції втрат;



$w_m = [1.0, 1.2, 1.0]$ — вагові коефіцієнти компонент, де підвищену вагу $w_Y = 1.2$ надано куту Yaw як найбільш варіативному.

Загальна функція втрат оптимізувалася з урахуванням явного L_2 -регуляризатора ваг:

$$L_{total} = L_{Huber} + \lambda \sum_i ||w_i||_2^2 \quad (5)$$

де $\lambda = 10^{-4}$ — коефіцієнт L_2 -штрафу, що застосовується до всіх згорткових та повнозв'язаних шарів.

Оцінювання точності здійснювалася за метрикою середньої абсолютної похибки в радіанах (MAE_{rad}):

$$MAE_{rad} = \frac{1}{3N} \sum_{i=1}^N \sum_{m=1}^3 |y_{i,m} - \hat{y}_{i,m}| \quad [rad], \quad (6)$$

де

- MAE_{rad} (Mean Absolute Error in radians) — середня абсолютна похибка оцінки/прогнозування кутових параметрів, виражена в радіанах (rad).
- N — загальна кількість об'єктів (спостережень, кадрів або зразків) у досліджуваній вибірці.
- i — індекс поточного об'єкта вибірки ($i = 1, 2, \dots, N$).
- m — індекс кутового компонента ($m \in \{1, 2, 3\}$). Множник 3 свідчить про те, що для кожного об'єкта вимірюються три просторові кути (наприклад, кути Ейлера: *крен/roll*, *тангаж/pitch*, *порскання/yaw* або 3 компоненти вектора орієнтації в просторі).
- $y_{i,m}$ — справжнє (еталонне, *Ground Truth*) значення m -го кута для i -го об'єкта (у радіанах).
- $\hat{y}_{i,m}$ — прогнозоване (оцінене моделлю) значення m -го кута для i -го об'єкта (у радіанах).
- $|y_{i,m} - \hat{y}_{i,m}|$ — модуль різниці (абсолютна похибка) між істинним та згенерованим моделлю значеннями m -го кута для i -го об'єкта.

Навчання проводилося протягом 15 епох за допомогою стандартного оптимізатора Adam (без використання схеми *weight_decay*).

МЕТОДИКА ДОСЛІДЖЕННЯ

Методика дослідження побудована за принципом порівняльного експерименту, в якому кілька архітектурно різних моделей навчаються та тестуються за єдиним протоколом на однакових наборах даних, що забезпечує коректність зіставлення результатів.

Дослідження охоплювало чотири етапи:

1. Підготовку та попередню обробку даних;
2. Проектування архітектур моделей;
3. Навчання моделей за уніфікованим протоколом;
4. Оцінювання та порівняльний аналіз результатів, включно з крос-доменним тестуванням.

Основним набором даних був AFLW2000-3D, що містить 2000 зображень обличчя з анотованими кутами орієнтації голови (pitch, yaw, roll), отриманими шляхом підгонки тривимірної морфологічної моделі (3D Morphable Model). Додатковим набором був BIWI Head Pose Dataset, анотації якого отримані на основі даних глибини (Kinect) і використовуються насамперед для перевірки узагальнювальної здатності моделей у крос-доменному режимі.

Для захисту від перевитрати оперативної пам'яті (RAM) та оптимізації процесів введення-виведення розроблено програмний конвеєр на основі *tf.data.Dataset* із підтримкою лінивого завантаження (*Lazy Loading*). Зображення зчитуються з локального носія суворо під час формування поточного батчу ($N = 32$).

Обробка даних включала наступні кроки:

- Виявлення та вирівнювання обличчя — локалізація області обличчя на зображенні та приведення її до єдиного масштабу і роздільності (типово 224×224 пікселів).
- Нормалізація кутів орієнтації голови (тангажу, рискання та крену — відповідно для нахилів вгору/вниз, поворотів ліворуч/праворуч та бічних нахилів) до єдиного діапазону $[0, 1]$ за апробованою експериментальною методикою [22], [23].
- Кватерніонне кодування вхідних даних для кватерніонних та гібридних моделей, коли кожен піксель RGB-зображення представлявся як чистий кватерніон $q = 0 + Ri + Gj + Bk$, що дозволило згортковим кватерніонним шарам обробляти кольорні канали як єдиний алгебраїчний об'єкт, зберігаючи їх взаємну кореляцію [3], [24].

- Аугментація даних, в межах якої виконувалося випадкове горизонтальне віддзеркалення (з відповідною інверсією знаку yaw і roll), випадкова зміна яскравості/контрастності, невеликі афінні викривлення для підвищення стійкості моделей до варіацій освітлення та ракурсу.
- Розбиття даних, в межах якого AFLW2000-3D поділялася на навчальну, валідаційну та тестову підмножини (наприклад, у співвідношенні 70/15/15); набір даних BIWI використовувався як незалежний тестовий набір для моделей, навчених на AFLW2000-3D (крос-доменна оцінка «train на AFLW2000-3D → test на BIWI»), або зворотному напрямку для перевірки симетричності узагальнення.

У цій роботі було розглянуто три категорії моделей. Приклад структурних схем досліджених моделей на базі архітектури MobileNetV2 наведено на рис. 1.

Легковажна дійснозначна CNN MobileNetV2 [26]. (в деяких обчислювальних експериментах DenseNet121 [25]). Зображення розміром $224 \times 224 \times 3$ надходило на базову мережу (MobileNetV2), яка витягувала 1280 скалярних карт ознак. Далі вектор ознак згортася за допомогою шару глобального усереднення (GlobalAveragePooling2D) та передався на повноз'язаний шар Dense(128) із функцією активації GELU і L_2 -регуляризацією ($\lambda = 10^{-4}$). Верхній класифікаційний шар було замінено вихідним регресійним модулем (повноз'язаними шарами з лінійним виходом), тому вихідний шар Dense(3) повертає три кути орієнтації (Pitch, Yaw, Roll) у радіанах.

Чисто кватерніонну модель (QCNN), в якій усі згорткові, нормалізаційні та повноз'язані шари реалізовано в гіперкомплексній області. Ця модель була оптимізована для навчання «з нуля» без використання ImageNet-ваг. Кватерніонну згортку реалізовано через добуток Гамільтона між кватерніонним ядром і вхідним кватерніонним тензором [3], [24], [27], що забезпечило суттєве скорочення кількості вільних параметрів порівняно з еквівалентним дійснозначним шаром за збереження розмірності репрезентативного простору [4], [27].

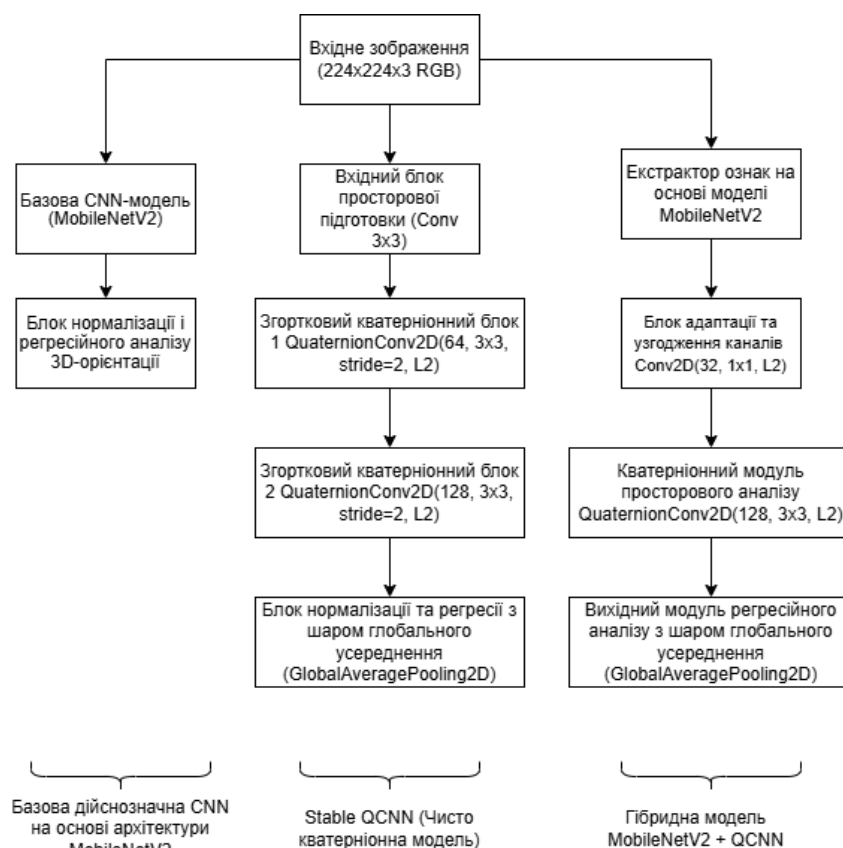


Рис. 1. Структурно-логічна схема досліджуваних нейромережесих архітектур регресії 3D-пози голови

Нормалізація здійснювалася за допомогою кватерніонної пакетної нормалізації (Quaternion Batch Normalization), яка забезпечувала обчислення статистик не за окремими компонентами кватерніона, а за його модулем, що гарантує узгоджене нормування всіх чотирьох компонент [3], [28]. Вхідний RGB-сигнал



оброблявся первинним блоком просторової підготовки (Conv2D(32, 3x3, stride=2) → BatchNorm → GELU), який передавав задекодовані градієнти на проєкційний шар Conv2D(32, 1x1). Цей шар трансформував 32 скалярні канали у 8 кватерніонних 4D-векторів ($[q_0, q_1, q_2, q_3]$). Обробка у гіперкомплексному просторі здійснювалася послідовністю шарів QuaternionConv2D (64 та 128 фільтрів). Підсумкова регресія кутів виконувалася через повнозв'язаний шар із активацією GELU.

Гібридна модель (Hybrid QCNN), яка поєднала початкові кватерніонні шари виділення ознак (для збереження міжканальної кореляційної структури на ранніх етапах обробки) із наступними дійснозначними шарами (для підсумкового регресійного аналізу на вищих рівнях абстракції). Такий підхід відповідає концепції параметризованих гіперкомплексних згорток (*Parameterized Hypercomplex Multiplications, PHM*), яка забезпечує гнучкий баланс між обчислювальною ефективністю кватерніонного представлення та практичною реалізацією дійснозначних операцій [4], [29]. Вихідні 1280 скалярних каналів базової мережі адаптувалися шаром Conv2D(128, 1x1) із пакетною нормалізацією (BatchNormalization) та активацією GELU, після чого перетворювалися у 32 чотиривимірні кватерніонні вектори. Шар QuaternionConv2D(128, 3x3) виконував просторову обробку ознак, після чого нормалізований вектор передавався на вихідний модуль регресійного аналізу для визначення кутів орієнтації у радіанах.

Усі моделі навчалися за єдиним протоколом для забезпечення порівняльності результатів. Регуляризація забезпечувалася за допомогою Dropout, L₂-штрафу ваг та ранньої зупинки (Early Stopping). Розмір батчу, кількість епох та інші гіперпараметри було зафіксовано однаковими для всіх порівнюваних моделей, що виключає помилкові оцінки на користь окремої архітектури.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для оцінки ефективності розроблених нейромережових архітектур у задачі прецизійної оцінки кутових параметрів орієнтації було проведено серію порівняльних експериментів. З метою запобігання перенаванчання та забезпечення високої узагальнювальної здатності моделей у досліджуваних конфігураціях застосовувалася L₂-регуляризація вагових коефіцієнтів.

Порівняльний аналіз виконувався для трьох інваріантів архітектур, які наведено вище.

Оцінювання точності здійснювалося за метрикою середньої абсолютної похибки (Mean Absolute Error, MAE), вираженої в радіанах та перерахованої у градусний еквівалент. Узагальнені результати експериментів (із застосуванням L₂-регуляризації) наведено в таблиці 1.

Таблиця 1

Порівняльні результати ефективності досліджуваних моделей

№	Модель	Кількість параметрів (М)	Час навчання (сек)	Val MAE	
				rad	еквівалент у °
1	MobileNetV2	2.42M	227.2	0.2116	12.12°
2	Stable QCNN	0.04M	135.7	0.3401	19.49°
3	MobileNetV2 + QCNN Hybrid	2.48M	241.6	0.1535	8.80°

Порівняння кривих навчання для різних типів моделей наведено на рис. 2.

На рис. 3 представлено серію з чотирьох тестових зображень обличч із баз даних (AFLW2000-3D/BIWI) за різних умов зйомки, освітлення, ракурсів та складних неканонічних поз (включно з поворотами, нахилами й оклюзіями).

Для кожного зображення у верхній частині наведено порівняльний блок метрик:

1. Прогнозовані значення (Пров): значення кутів тангажу (P — pitch), рискання (Y — yaw) та крену (R — roll) у радіанах, отримані на виході регресійного модуля гібридної моделі.
2. Істинні значення (Іст): еталонні (*Ground Truth*) значення кутів орієнтації з анотаційних матриць.
3. Середня абсолютна похибка (MAE): значення похибки для конкретного кадру у радіанах та її еквівалент у градусах.

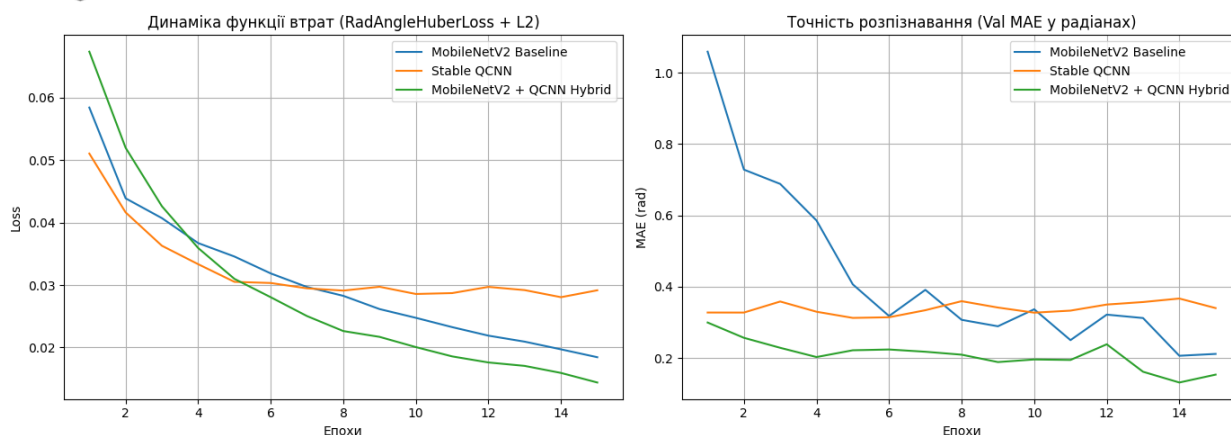


Рис. 2. Криві навчання різних типів моделей

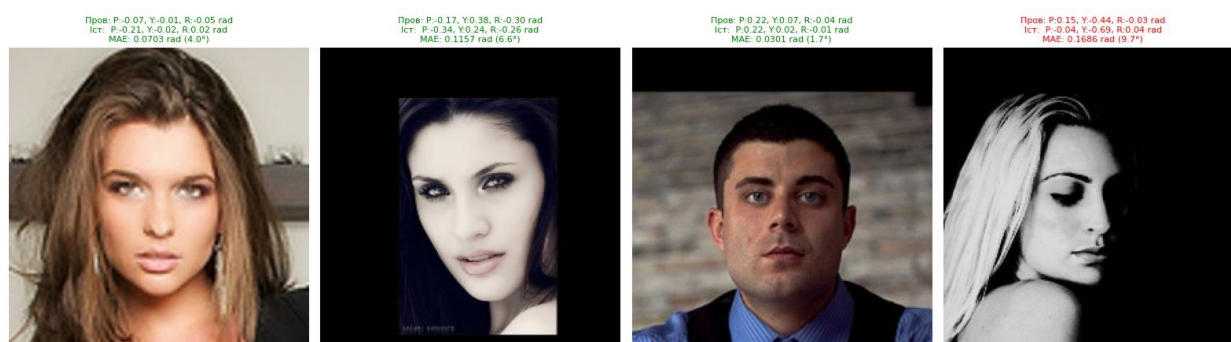


Рис. 3. Візуалізація результатів регресії кутів орієнтації обличчя (Pitch, Yaw, Roll) гібридною моделлю MobileNetV2 + QCNN Hybrid: порівняння прогнозованих (Прог) та істинних (Іст) значень кутів у радіанах і підсумкової похибки MAE (у радіанах та градусах).

Візуальний аналіз наведених тестових випадків показав наступне:

- Перший та третій кадри (зелене маркування) демонструють високу точність прогнозування для облич у близьких до канонічних ракурсах та за помірного освітлення. Значення похибки MAE становить відповідно 0.0703 rad (4.0°) та 0.0301 rad (1.7°), що свідчить про стійкість кватерніонних шарів при мікроповоротах голови.
- Другий кадр (зелене маркування): Ілюструє роботу моделі в умовах складного неканонічного повороту голови та контрастного/темного освітлення. Зафіксовано високу точність регресії з MAE 0.1157 rad (6.6°).
- Четвертий кадр (червоне маркування): Представляє граничний/складний випадок із суттєвим бічним поворотом, затіненням та оклюзією (волосся закриває частину обличчя). Похибка зростає до 0.1686 rad (9.7°), проте залишається у робочому діапазоні (до 10°), підтверджуючи здатність гібридної архітектури компенсувати втрату інформації через некомутативний добуток Гамільтона.

Аналіз отриманої матриці експериментальних даних дозволяє зробити такі науково обґрунтовані висновки:

1. Застосування комбінованої архітектури продемонструвало найкращу точність серед усіх досліджуваних варіантів. Значення похибки на валідаційній вибірці зменшилося до 0.1535 rad (8.80°). Порівняно з базовою класичною моделлю MobileNetV2 (12.12°), гібридний підхід забезпечив приріст точності (зниження похибки) на 27.39%. При цьому обчислювальні витрати зросли незначно: кількість параметрів збільшилася всього на 2.48% (з 2.42М до 2.48М), а час навчання — на 6.3% (з 227.2 с до 241.6 с). Це підтверджує наявність синергетичного ефекту від інтеграції квантово-інспірованих шарів у глибокі згорткові конвеєри.
2. Модель Stable QCNN продемонструвала виняткову компактність — усього 0.04М параметрів (близько 40 тисяч), що у 60.5 разів менше, ніж у базової MobileNetV2. Час її навчання виявився



- найменшим (135.7 с, що на 40.3% швидше за базову CNN). Незважаючи на помірне зростання похибки до 19.49° (0.3401 rad), така модель є високopersпективною для застосування в умовах критичного обмеження обчислювальних ресурсів (Edge AI, вбудовані системи та IoT-пристрої).
3. Впровадження L_2 -регуляризації дозволило стабілізувати процес навчання квантово-інспірованих компонентів (запобігши вибуху/загасанню градієнтів, характерних для QCNN). У гібридній схемі L_2 -регуляризатор виступив додатковим стимулом для формування збалансованого простору ознак між класичними та квантовими шарами.
- Таким чином, експериментально доведено, що гібридна архітектура MobileNetV2 + QCNN Hybrid забезпечує оптимальний баланс між точністю прогнозування та обчислювальною складністю, дозволяючи досягти кутової похибки менше 9° при роботі у складнодосліджуваних недетермінованих середовищах.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі наведено нове розв'язання науково-практичної задачі прецизійної оцінки 3D-орієнтації обличчя за рахунок розробки та дослідженні дійснозначних, чисто кватерніонних та гібридних нейромережових архітектур. Отримані результати підтверджують доцільність використання гібридних згорткових конвеєрів для побудови високоточних, стійких до оклюзій і ракурсних спотворень систем 3D-аналізу обличчя, які можуть бути успішно імплементовані в біометричних системах, системах доповненої та віртуальної реальності, а також модулях відеоаналітики.

На основі виконаного комплексу теоретичних та експериментальних досліджень сформульовано такі основні висновки:

1. Здійснений аналіз сучасних джерел підтвердив, що традиційне представлення 3D-обертань через кути Ейлера в дійснозначних CNN викликає математичні сингулярності (Gimbal Lock) та не враховує міжканальні кореляції. Застосування кватерніонної алгебри та добутку Гамільтона дозволяє природно кодувати просторові відношення та знижувати параметричну складність моделей без втрати репрезентативності простору ознак.
2. Інтеграція кватерніонного згорткового модуля (QuaternionConv2D) у глибоку дійснозначну мережу MobileNetV2 забезпечила найкращу точність регресії кутів пози голови серед усіх досліджених конфігурацій. Середня абсолютна похибка (Val MAE) склала 0.1535 rad (8.80°), що відповідає зниженню похибки на 27.39% порівняно з базовою MobileNetV2 (12.12°). При цьому обчислювальна складність зросла несуттєво: обсяг параметрів збільшився на 2.48% (з 2.42M до 2.48M), а час навчання — на 6.3%.
3. Модель Stable QCNN продемонструвала виняткову параметричну ефективність — усього 0.04M параметрів (~40 тис. wag), що у 60.5 разів компактніше за базову CNN, та прискорення навчання на 40.3% (135.7 с проти 227.2 с). Незважаючи на помірне зростання похибки до 19.49° , така модель є обґрунтованим вибором для застосування в умовах жорстких обмежень пам'яті й обчислювальних ресурсів (IoT, вбудовані системи, Edge AI).
4. Стабілізація та математична адаптація контуру навчання: Впровадження взваженої адаптивної функції втрат Huber Loss у сполученні з L_2 -регуляризацією ($\lambda = 10^{-4}$) дозволило нівелювати проблему вибуху/згасання градієнтів у кватерніонних шарах, забезпечивши плавну конвергенцію та формуванню узгодженого простору ознак на стику класичних і кватерніонних операцій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Asperti, A., & Filippini, D. (2023). Deep learning for head pose estimation: A survey. *SN Computer Science*, 4(4), Article 349. <https://doi.org/10.1007/s42979-023-01796-z>
2. Singh, S., Kumar, S., & Tripathi, B. K. (2025). A comprehensive analysis of quaternion deep neural networks: Architectures, applications, challenges, and future scope. *Archives of Computational Methods in Engineering*, 32(4), 2607–2634. <https://doi.org/10.1007/s11831-024-10216-1>
3. Altamirano-Gomez, G., & Gershenson, C. (2024). Quaternion convolutional neural networks: Current advances and future directions. *Advances in Applied Clifford Algebras*, 34(4), Article 42. <https://doi.org/10.1007/s00006-024-01350-x>
4. Zhang, A., Tay, Y., Zhang, S., Chan, A., Luu, A. T., Hui, S. C., & Fu, J. (2021). Beyond fully-connected layers with quaternions: Parameterization of hypercomplex multiplications with $1/n$ parameters (arXiv:2102.08597). arXiv. <https://doi.org/10.48550/ARXIV.2102.08597>
5. El Alami, A., Mesbah, A., Berrahou, N., Lakhili, Z., Berrahou, A., & Qjidaa, H. (2023). Quaternion discrete orthogonal Hahn moments convolutional neural network for color image classification and face



- recognition. *Multimedia Tools and Applications*, 82(21), 32827–32853. <https://doi.org/10.1007/s11042-023-14866-4>
6. Li, M., Huang, B., & Tian, G. (2022). A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, 110, Article 104669. <https://doi.org/10.1016/j.engappai.2022.104669>
 7. Algabri, R., Shin, H., Abdu, A., Bae, J.-H., & Lee, S. (2025). WQuatNet: Wide range quaternion-based head pose estimation. *Journal of King Saud University - Computer and Information Sciences*, 37(3), Article 24. <https://doi.org/10.1007/s44443-025-00034-1>
 8. Hempel, T., Abdelrahman, A. A., & Al-Hamadi, A. (2024). Toward robust and unconstrained full range of rotation head pose estimation. *IEEE Transactions on Image Processing*, 33, 2377–2387. <https://doi.org/10.1109/TIP.2024.3378180>
 9. Parcollet, T., Morchid, M., & Linares, G. (2020). A survey of quaternion neural networks. *Artificial Intelligence Review*, 53(4), 2957–2982. <https://doi.org/10.1007/s10462-019-09752-1>
 10. García-Retuerta, D., Casado-Vara, R., Martín-del Rey, Á., De La Prieta, F., Prieto, J., & Corchado, J. M. (2020). Quaternion neural networks: State-of-the-art and research challenges. In C. Analide, P. Novais, D. Camacho, & H. Yin (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2020* (Vol. 12490, pp. 456–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-62365-4_43
 11. Bill, J., Champagne, L., Cox, B., & Bihl, T. (2021). Meta-heuristic optimization methods for quaternion-valued neural networks. *Mathematics*, 9(9), Article 938. <https://doi.org/10.3390/math9090938>
 12. Shahadat, N., & Maida, A. S. (2023). Enhancing ResNet image classification performance by using parameterized hypercomplex multiplication (arXiv:2301.04623). arXiv. <https://doi.org/10.48550/ARXIV.2301.04623>
 13. El Alami, A., Berrahou, N., Lakhili, Z., Mesbah, A., Berrahou, A., & Qjidaa, H. (2022). Efficient color face recognition based on quaternion discrete orthogonal moments neural networks. *Multimedia Tools and Applications*, 81(6), 7685–7710. <https://doi.org/10.1007/s11042-021-11669-3>
 14. Wen, C., & Qiu, Y. (2022). Color occlusion face recognition method based on quaternion non-convex sparse constraint mechanism. *Sensors*, 22(14), Article 5284. <https://doi.org/10.3390/s22145284>
 15. Zhou, Y., Jin, L., Liu, H., & Song, E. (2021). Color facial expression recognition by quaternion convolutional neural network with Gabor attention. *IEEE Transactions on Cognitive and Developmental Systems*, 13(4), 969–983. <https://doi.org/10.1109/TCDS.2020.3041642>
 16. Liu, L., Chen, C. L. P., & Li, S. (2022). Hallucinating color face image by learning graph representation in quaternion space. *IEEE Transactions on Cybernetics*, 52(1), 265–277. <https://doi.org/10.1109/TCYB.2020.2979320>
 17. Zhao, Y., Birdal, T., Lenssen, J. E., Menegatti, E., Guibas, L., & Tombari, F. (2020). Quaternion equivariant capsule networks for 3D point clouds. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12346, pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_1
 18. You, X., & Zhao, X. (2025). A 3D face recognition algorithm directly applied to point clouds. *Biomimetics*, 10(2), Article 70. <https://doi.org/10.3390/biomimetics10020070>
 19. Hsu, H.-W., Wu, T.-Y., Wan, S., Wong, W. H., & Lee, C.-Y. (2019). QuatNet: Quaternion-based head pose estimation with multiregression loss. *IEEE Transactions on Multimedia*, 21(4), 1035–1046. <https://doi.org/10.1109/TMM.2018.2866770>
 20. Liu, W., Zhou, L., & Chen, J. (2021). Face recognition based on lightweight convolutional neural networks. *Information*, 12(5), Article 191. <https://doi.org/10.3390/info12050191>
 21. Gaudet, C. J., & Maida, A. S. (2018). Deep quaternion networks. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. <https://doi.org/10.1109/IJCNN.2018.8489651>
 22. Li, X., Zhang, D., Li, M., & Lee, D.-J. (2023). Accurate head pose estimation using image rectification and a lightweight convolutional neural network. *IEEE Transactions on Multimedia*, 25, 2239–2251. <https://doi.org/10.1109/TMM.2022.3144893>
 23. Liu, H., Fang, S., Zhang, Z., Li, D., Lin, K., & Wang, J. (2022). MFDNet: Collaborative poses perception and matrix Fisher distribution for head pose estimation. *IEEE Transactions on Multimedia*, 24, 2449–2460. <https://doi.org/10.1109/TMM.2021.3081873>
 24. Zhou, H., Zhang, X., Zhang, C., & Ma, Q. (2023). Quaternion convolutional neural networks for hyperspectral image classification. *Engineering Applications of Artificial Intelligence*, 123, Article 106234. <https://doi.org/10.1016/j.engappai.2023.106234>
 25. Martínez-Díaz, Y., Luhbach, P., Benitez-Garcia, G., & Hernandez-Sosa, D. (2021). Benchmarking lightweight face architectures on specific face recognition scenarios. *Artificial Intelligence Review*, 54(8), 6201–6244. <https://doi.org/10.1007/s10462-021-09974-2>



26. Boutros, F., Damer, N., Fang, M., Kirchbuchner, F., & Kuijper, A. (2021). MixFaceNets: Extremely efficient face recognition networks. In *2021 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1–8). IEEE. <https://doi.org/10.1109/IJCB52358.2021.9484374>
27. Grant, B., & Wang, P. (2025). *Quaternion approximation networks for enhanced image classification and oriented object detection* (arXiv:2509.05512). arXiv. <https://doi.org/10.48550/ARXIV.2509.05512>
28. Devadas, R. M., Hiremani, V., Preethi, S. T., S, R., & Gujjar, P. (2025). Hypercomplex neural networks: Exploring quaternion, octonion, and beyond in deep learning. *MethodsX*, 15, Article 103644. <https://doi.org/10.1016/j.mex.2025.103644>
29. Grassucci, E., Zhang, A., & Communiello, D. (2024). PHNNs: Lightweight neural networks via parameterized hypercomplex convolutions. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6), 8293–8305. <https://doi.org/10.1109/TNNLS.2022.3226772>

**Yevhen Chychkarov**

Doctor of Technical Sciences, Professor, Professor at the Department of Computer Sciences

National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

ORCID: 0000-0002-4362-5129

ye.chychkarov@nubip.edu.ua

Ivan Veremiichuk

PhD student,

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

ORCID: 0009-0009-2279-7659

i.veremiichuk@gmail.com

3D FACE IMAGE RECOGNITION USING QUATERNION NEURAL NETWORKS

Abstract. This article addresses a pressing scientific and applied problem in computer vision: improving the accuracy of 3D face orientation (pose) estimation and reducing the parametric complexity of deep neural network models through the use of quaternion algebra. The study examines the limitations of real-valued convolutional neural networks (CNNs), which process coordinates and color channels independently; this approach loses intrinsic inter-component correlations and leads to degraded accuracy in the presence of significant head rotations and occlusions. The use of custom quaternion convolutional layers (QuaternionConv2D)—based on the non-commutative Hamilton product—is proposed and mathematically justified; this enables the modeling of spatial-angular signal interdependencies and reduces the number of free weight parameters by a factor of four compared to classical convolutions. To ensure stable model training, a weighted adaptive Huber loss function in the radian domain was developed—placing greater weight on the highly variable yaw angle—and L2 weight regularization was applied. A systematic comparative experimental analysis was conducted on three types of architectures: a purely quaternion network (Stable QCNN) and a hybrid model (MobileNetV2 + QCNN Hybrid). Based on the results obtained, it has been demonstrated that the hybrid architecture achieves the best recognition accuracy, reducing the validation MAE to 0.1535 rad (8.80°)—an improvement of 27.39% over the baseline MobileNetV2 (12.12°)—with a negligible increase in parameter count of only 2.48%. Meanwhile, the ultra-lightweight, purely quaternion-based Stable QCNN model exhibits extreme compactness (0.04M parameters, 60.5 times fewer than MobileNetV2) and a 40.3% shorter training time, making it a promising solution for resource-constrained Edge AI devices.

Keywords: 3D face recognition, head pose estimation, quaternion neural networks (QCNN), Hamilton product, hybrid architectures, MobileNetV2, Huber Loss, L2 regularization, Edge AI.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Asperti, A., & Filippini, D. (2023). Deep learning for head pose estimation: A survey. *SN Computer Science*, 4(4), Article 349. <https://doi.org/10.1007/s42979-023-01796-z>
2. Singh, S., Kumar, S., & Tripathi, B. K. (2025). A comprehensive analysis of quaternion deep neural networks: Architectures, applications, challenges, and future scope. *Archives of Computational Methods in Engineering*, 32(4), 2607–2634. <https://doi.org/10.1007/s11831-024-10216-1>
3. Altamirano-Gomez, G., & Gershenson, C. (2024). Quaternion convolutional neural networks: Current advances and future directions. *Advances in Applied Clifford Algebras*, 34(4), Article 42. <https://doi.org/10.1007/s00006-024-01350-x>
4. Zhang, A., Tay, Y., Zhang, S., Chan, A., Luu, A. T., Hui, S. C., & Fu, J. (2021). *Beyond fully-connected layers with quaternions: Parameterization of hypercomplex multiplications with 1/n parameters* (arXiv:2102.08597). arXiv. <https://doi.org/10.48550/ARXIV.2102.08597>
5. El Alami, A., Mesbah, A., Berrahou, N., Lakhili, Z., Berrahou, A., & Qjidaa, H. (2023). Quaternion discrete orthogonal Hahn moments convolutional neural network for color image classification and face recognition. *Multimedia Tools and Applications*, 82(21), 32827–32853. <https://doi.org/10.1007/s11042-023-14866-4>
6. Li, M., Huang, B., & Tian, G. (2022). A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, 110, Article 104669. <https://doi.org/10.1016/j.engappai.2022.104669>



7. Algabri, R., Shin, H., Abdu, A., Bae, J.-H., & Lee, S. (2025). WQuatNet: Wide range quaternion-based head pose estimation. *Journal of King Saud University - Computer and Information Sciences*, 37(3), Article 24. <https://doi.org/10.1007/s44443-025-00034-1>
8. Hempel, T., Abdelrahman, A. A., & Al-Hamadi, A. (2024). Toward robust and unconstrained full range of rotation head pose estimation. *IEEE Transactions on Image Processing*, 33, 2377–2387. <https://doi.org/10.1109/TIP.2024.3378180>
9. Parcollet, T., Morchid, M., & Linares, G. (2020). A survey of quaternion neural networks. *Artificial Intelligence Review*, 53(4), 2957–2982. <https://doi.org/10.1007/s10462-019-09752-1>
10. García-Retuerta, D., Casado-Vara, R., Martín-del Rey, Á., De La Prieta, F., Prieto, J., & Corchado, J. M. (2020). Quaternion neural networks: State-of-the-art and research challenges. In C. Analide, P. Novais, D. Camacho, & H. Yin (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2020* (Vol. 12490, pp. 456–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-62365-4_43
11. Bill, J., Champagne, L., Cox, B., & Bihl, T. (2021). Meta-heuristic optimization methods for quaternion-valued neural networks. *Mathematics*, 9(9), Article 938. <https://doi.org/10.3390/math9090938>
12. Shahadat, N., & Maida, A. S. (2023). Enhancing ResNet image classification performance by using parameterized hypercomplex multiplication (arXiv:2301.04623). arXiv. <https://doi.org/10.48550/ARXIV.2301.04623>
13. El Alami, A., Berrahou, N., Lakhili, Z., Mesbah, A., Berrahou, A., & Qjidaa, H. (2022). Efficient color face recognition based on quaternion discrete orthogonal moments neural networks. *Multimedia Tools and Applications*, 81(6), 7685–7710. <https://doi.org/10.1007/s11042-021-11669-3>
14. Wen, C., & Qiu, Y. (2022). Color occlusion face recognition method based on quaternion non-convex sparse constraint mechanism. *Sensors*, 22(14), Article 5284. <https://doi.org/10.3390/s22145284>
15. Zhou, Y., Jin, L., Liu, H., & Song, E. (2021). Color facial expression recognition by quaternion convolutional neural network with Gabor attention. *IEEE Transactions on Cognitive and Developmental Systems*, 13(4), 969–983. <https://doi.org/10.1109/TCDS.2020.3041642>
16. Liu, L., Chen, C. L. P., & Li, S. (2022). Hallucinating color face image by learning graph representation in quaternion space. *IEEE Transactions on Cybernetics*, 52(1), 265–277. <https://doi.org/10.1109/TCYB.2020.2979320>
17. Zhao, Y., Birdal, T., Lenssen, J. E., Menegatti, E., Guibas, L., & Tombari, F. (2020). Quaternion equivariant capsule networks for 3D point clouds. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12346, pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_1
18. You, X., & Zhao, X. (2025). A 3D face recognition algorithm directly applied to point clouds. *Biomimetics*, 10(2), Article 70. <https://doi.org/10.3390/biomimetics10020070>
19. Hsu, H.-W., Wu, T.-Y., Wan, S., Wong, W. H., & Lee, C.-Y. (2019). QuatNet: Quaternion-based head pose estimation with multiregression loss. *IEEE Transactions on Multimedia*, 21(4), 1035–1046. <https://doi.org/10.1109/TMM.2018.2866770>
20. Liu, W., Zhou, L., & Chen, J. (2021). Face recognition based on lightweight convolutional neural networks. *Information*, 12(5), Article 191. <https://doi.org/10.3390/info12050191>
21. Gaudet, C. J., & Maida, A. S. (2018). Deep quaternion networks. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. <https://doi.org/10.1109/IJCNN.2018.8489651>
22. Li, X., Zhang, D., Li, M., & Lee, D.-J. (2023). Accurate head pose estimation using image rectification and a lightweight convolutional neural network. *IEEE Transactions on Multimedia*, 25, 2239–2251. <https://doi.org/10.1109/TMM.2022.3144893>
23. Liu, H., Fang, S., Zhang, Z., Li, D., Lin, K., & Wang, J. (2022). MFDNet: Collaborative poses perception and matrix Fisher distribution for head pose estimation. *IEEE Transactions on Multimedia*, 24, 2449–2460. <https://doi.org/10.1109/TMM.2021.3081873>
24. Zhou, H., Zhang, X., Zhang, C., & Ma, Q. (2023). Quaternion convolutional neural networks for hyperspectral image classification. *Engineering Applications of Artificial Intelligence*, 123, Article 106234. <https://doi.org/10.1016/j.engappai.2023.106234>
25. Martínez-Díaz, Y., Luhbach, P., Benitez-Garcia, G., & Hernandez-Sosa, D. (2021). Benchmarking lightweight face architectures on specific face recognition scenarios. *Artificial Intelligence Review*, 54(8), 6201–6244. <https://doi.org/10.1007/s10462-021-09974-2>
26. Boutros, F., Damer, N., Fang, M., Kirchbuchner, F., & Kuijper, A. (2021). MixFaceNets: Extremely efficient face recognition networks. In *2021 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1–8). IEEE. <https://doi.org/10.1109/IJCB52358.2021.9484374>
27. Grant, B., & Wang, P. (2025). Quaternion approximation networks for enhanced image classification and oriented object detection (arXiv:2509.05512). arXiv. <https://doi.org/10.48550/ARXIV.2509.05512>



28. Devadas, R. M., Hiremani, V., Preethi, S. T., S, R., & Gujjar, P. (2025). Hypercomplex neural networks: Exploring quaternion, octonion, and beyond in deep learning. *MethodsX*, 15, Article 103644. <https://doi.org/10.1016/j.mex.2025.103644>
29. Grassucci, E., Zhang, A., & Communiello, D. (2024). PHNNs: Lightweight neural networks via parameterized hypercomplex convolutions. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6), 8293–8305. <https://doi.org/10.1109/TNNLS.2022.3226772>

Отримано редакцією журналу / Received: 28.08.26

Прорецензовано / Revised: 16.09.26

Схвалено до друку / Accepted: 26.10.26



This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.