

---

---

# СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

---

---

УДК 004.8

А. М. ТАРНОВСЬКИЙ, С.М. ЗАХАРЧЕНКО, М. Г. ТАРНОВСЬКИЙ

## ВПЛИВ ЛОГАРИФМІЧНОЇ ТРАНСФОРМАЦІЇ ВХІДНИХ АКТИВАЦІЙ У ЗГОРТКОВИХ МЕРЕЖАХ НА ЛОКАЛІЗАЦІЮ КЛЮЧОВИХ ТОЧОК ОБЛИЧЧЯ

*Вінницький національний технічний університет, Україна  
21021 Хмельницьке шосе 95, м. Вінниця,*

**Анотація.** У роботі розглянуто застосування в моделях згорткових нейронних мереж принципу логарифмічного сприйняття, згідно з яким реакція системи визначається відносними змінами сигналу. У межах цього підходу досліджено вплив логарифмічної трансформації вхідних активацій нейронів першого згорткового шару на точність локалізації ключових точок обличчя та стійкість моделі до варіацій яскравості зображень. Експериментальна перевірка на наборі даних WFLW з використанням архітектури ResNet-34 показала, що така трансформація не має значного впливу на точність локалізації ключових точок на зображеннях з нормальною яскравістю, однак дещо підвищує стійкість моделі до її зниження. Зокрема, встановлено, що застосування логарифмічної трансформації дозволяє знизити значення нормалізованої середньої похибки NME порівняно з базовою моделлю в середньому на 0,0019 при трикратному та на 0,0071 при шестикратному зменшенні яскравості. Отже, логарифмічна трансформація вхідних активацій може розглядатися як один з інструментів підвищення стійкості згорткових нейронних мереж до змін інтенсивності вхідних сигналів без ускладнення їхньої архітектури.

**Ключові слова:** локалізація ключових точок, згортковий шар, глибоке навчання, функція втрат, градієнтний спуск, логарифмічне перетворення.

**Abstract.** The paper considers the application of the logarithmic perception principle in convolutional neural network models, according to which the system's response is determined by relative signal changes. Within this approach, the impact of logarithmic transformation of input activations in the first convolutional layer on facial landmark localization accuracy and model robustness to brightness variations is investigated. Experimental validation on the WFLW dataset using the ResNet-34 architecture demonstrated that such a transformation does not significantly affect localization accuracy under normal brightness conditions; however, it slightly increases the model's robustness to illumination decrease. Specifically, it was found that applying logarithmic transformation reduces the Normalized Mean Error (NME) compared to the baseline model by an average of 0.0019 for a threefold decrease in brightness and by 0.0071 for a sixfold decrease. Thus, the logarithmic transformation of input activations can be considered a tool for enhancing the robustness of convolutional neural networks to input signal intensity variations without increasing their architectural complexity.

**Keywords:** landmark localization, convolutional layer, deep learning, loss function, gradient descent, logarithmic transformation.

**DOI: 10.31649/1681-7893-2026-51-1-108-116**

### ВСТУП

Детектування та локалізація ключових точок обличчя є важливим етапом при вирішенні завдань комп'ютерного зору у застосунках, пов'язаних з розпізнаванням облич, аналізом виразу обличчя, 3D-реконструкцією обличчя тощо. Цей процес полягає у виявленні та визначенні координат характерних точок обличчя, таких як контури очей, кінчик носа та межі ніздів, контури рота, дуга брів, точки вздовж лінії щелепи та підборіддя [1].

Сучасні підходи до виявлення ключових точок обличчя базуються на методах глибокого навчання, зокрема на використанні згорткових нейронних мереж, які вирішують задачу локалізації ключових точок як задачу регресії їх координат.

---

© А. М. ТАРНОВСЬКИЙ, С.М. ЗАХАРЧЕНКО, М. Г. ТАРНОВСЬКИЙ, 2026

---

---

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

---

---

Глибокі архітектури згорткових мереж перевершують класичні методи за точністю та стійкістю до змін сценаріїв, таких як рівень освітлення, поворот голови, часткове перекриття обличчя [2], [3]. Адаптація згорткових мереж до регресії координат ключових точок реалізується через заміну вихідного класифікаційного шару (зазвичай Softmax) повнозв'язаним шаром, що формує вектор координат [1].

Для подолання обмежень стандартних згорткових шарів, пов'язаних з фіксованим розміром рецептивного поля та чутливістю до масштабу, використовуються архітектури з механізмами багатомасштабного аналізу. Прикладом такого підходу є архітектура на базі Insertion-модулів, які за рахунок паралельних згортков фільтрами різного розміру дозволяють одночасно обробляти вхідні дані різного масштабу [3]. Це забезпечує більш ефективне вилучення як локальних, так і глобальних ознак. У роботі [3] показано, що така модель перевершує традиційні архітектури за точністю визначення ключових точок обличчя. Крім того, завдяки багатомасштабному аналізу тримається можливість успішної роботи в складних умовах, таких як варіації освітлення, повороти обличчя та часткові перекриття.

Використання залишкових зв'язків (residual connections), що лежать в основі архітектури типу ResNet, дозволяє запобігти втратам дрібних деталей через багаторазові згортки та подолати проблему затухання градієнта, притаманну глибоким мережам. Для підвищення точності локалізації ключових точок обличчя часто застосовується навчання з перенесенням знань (Transfer Learning). Так у роботі [5] на прикладі мережі ResNet-50 продемонстровано, що використання залишкових зв'язків та попередньо навчених ваг забезпечує стабільну регресію координат ключових точок у глибоких мережах та високу точність на складних наборах даних.

Іншим підходом, спрямованим на запобігання втраті інформації про дрібні деталі, є використання механізмів уваги. Прикладом реалізації такої стратегії є запропонована в роботі [4] архітектура на основі глибокої згорткової мережі VGG, доповнена модулями посилення та злиття ознак, які дозволяють зберегти деталізацію низькорівневих шарів та поєднати її з абстрактними семантичними ознаками більш глибоких шарів. Ключовим рішенням стало впровадження каскадного механізму уваги на основі блоків SENet, що дозволяє моделі зосереджуватися на найбільш інформативних ділянках зображення, таких як очі, ніс та рот. Проведені авторами дослідження показали, що інтеграція механізмів уваги покращує точність локалізації ключових точок обличчя та забезпечує стійкість моделі до оклюзій (перекриттів) і великих кутів нахилу голови.

Знизити вимоги до обчислювальних ресурсів при збереженні високої точності та стійкості дозволяють гібридні архітектури, в яких знання від потужних моделей передаються компактним мережам. Прикладом такого рішення є архітектура, що поєднує дві згорткові мережі ResNet-50 та MobileNet [6]. ResNet-50 застосовується як базовий модуль для вилучення ознак і після попереднього навчання обох моделей стає статичною. Подальша оптимізація спрямована на компактну модель MobileNet, що має меншу обчислювальну складність, але й обмежені можливості. Для забезпечення високої точності використовується механізм передачі знань (knowledge transfer), відповідно до якого інформація про локалізацію ключових точок від ResNet-50 передається до MobileNet. Такий підхід дозволяє досягти балансу між точністю та швидкістю, що є критичним для застосунків реального часу у мобільних та вбудованих системах.

Отже, сучасні підходи до застосування згорткових мереж для локалізації ключових точок зосереджені на модифікації їхньої архітектури, зокрема на збільшенні глибини, багатомасштабному аналізі ознак та використанні механізмів уваги. При цьому принципи обробки сигналів залишаються переважно незмінними та базуються на лінійних операціях, обмежених пороговими активаціями такими як ReLU. Водночас у багатьох випадках інформаційна цінність ознаки не є пропорційною її абсолютному значенню. У природних системах, зокрема біологічних, реакція визначається відносними змінами сигналу. Відповідно до закону Вебера-Фехнера [7] інтенсивність людського сприйняття змінюється пропорційно логарифму сили подразника. З огляду на це можна припустити, що логарифмічна трансформація вхідних активацій нейронів першого згорткового шару, яка забезпечує чутливість до відносних змін сигналу, дозволить підвищити точність локалізації ключових точок та стійкість моделі до змін яскравості.

Метою роботи є аналіз впливу логарифмічної трансформації вхідних активацій першого згорткового шару на точність локалізації ключових точок обличчя, а також на стійкість результатів до змін яскравості зображень.

### МОДЕЛЬ НЕЙРОНА З ЛОГАРИФМІЧНОЮ ТРАНСФОРМАЦІЄЮ ВХІДНИХ АКТИВАЦІЙ

У багатьох випадках логарифми відіграють важливу роль, забезпечуючи обчислювальну стабільність та високий рівень інтерпретованості даних. У контексті штучних нейронних мереж

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

логарифми надають зручний спосіб лінеаризації експоненціальних залежностей. З іншого боку логарифмічна трансформація надає можливість зменшити вплив великих за модулем активацій на значення градієнтів, що може сприяти стабілізації процесу навчання. Отже, очікується, що використання логарифмічної трансформації безпосередньо всередині нейрона дозволить нейронній мережі автоматично стискати діапазон значень і фокусуватися на відносних змінах сигналів, а не на їх абсолютних значеннях.

Класичний нейрон сприймає вхідний сигнал лінійно [8]:

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i + b\right),$$

де  $y$  — вихід нейрона,  $x_i$  — вхідні сигнали,  $f$  — функція активації,  $w_i$  — ваги входів,  $b$  — зміщення, що дозволяє зсувати графік функції активації для кращої адаптації моделі.

Математична модель нейрона, що оцінює вхідний сигнал через логарифмічну трансформацію, отримає вигляд:

$$y = f\left(\sum_{i=1}^n w_i \cdot \ln(x_i + \varepsilon) + b\right), \quad (1)$$

де  $\varepsilon$  — деяка стала більша за 0.

Додавання  $\varepsilon$  забезпечує додатне ненульове значення під логарифмом. При роботі із зображеннями  $x_i$  — це яскравості пікселів, тому для нейронів першого шару  $x_i \geq 0$  і, відповідно,  $x_i + \varepsilon > 0$ .

Перед подачею у мережу зображення зазвичай нормалізують. Нормалізація полягає у масштабуванні значень пікселів вхідних зображень у певний діапазон, зазвичай у діапазон  $[0; 1]$  [9], [10]. При цьому яскравість та контрастність зображення не змінюється. Відповідно доцільно прийняти  $\varepsilon = 1$ , що зробить результат логарифмічної трансформації для вхідних сигналів додатнім. При  $x_i = 0$  значення  $\ln(x_i + 1)$  також дорівнює нулю. Це дозволяє відобразити нульовий вхідний сигнал у нульовий вихідний. При нескінченно малих  $x_i$  похідна від  $\ln(x_i + 1)$  приблизно дорівнює 1, що запобігає вибуху градієнтів при  $x_i$  близьких до нуля. Для управління ступенем логарифмічного стиснення введемо додатковий параметр масштабування  $\gamma$ , що визначатиме крутизну логарифмічної кривої. З врахуванням викладеного вище вираз (1) набуде вигляду:

$$y = f\left(\sum_{i=1}^n w_i \cdot \gamma \cdot \ln(x_i + 1) + b\right). \quad (3)$$

Коефіцієнт масштабування  $\gamma$  є додатковим навчальним параметром, а тому він не поглинається вагами  $w_i$ . У даному дослідженні  $\gamma$  розглядається як спільний параметр для усіх нейронів згорткового шару, що однаково впливає на усі логарифмічно трансформовані входи шару. Припускається, що у процесі навчання мережа за його допомогою може самостійно адаптувати крутизну логарифмічної кривої під особливості вхідних даних.

Похідна від логарифму  $\ln(x_i + 1)$ , що дорівнює  $1/(x_i + 1)$ , збільшується при зменшенні  $x_i$ . Це означає, що для вхідного шару нейрони (3) будуть більш чутливими до змін яскравості у темних пікселях, ніж у світлих. Це може виявитися корисним при низькій яскравості чи низькому контрасті зображення.

Розглянемо як логарифмічна трансформація впливає на навчання мережі за методом градієнтного спуску. Активація  $a_j^{(1)}$  нейрона  $j$  першого шару визначається за виразом (3) у вигляді:

$$a_j^{(1)} = f\left(\sum_{i=1}^n w_{ji}^{(1)} \cdot \gamma^{(1)} \cdot \ln(x_i + 1) + b_j^{(1)}\right). \quad (4)$$

Знайдемо градієнти функції втрат (функції помилки)  $\mathcal{L}$  для ваг  $w_{ji}^{(1)}$  та масштабу  $\gamma^{(1)}$ . Градієнт для зміщення  $b_j^{(1)}$  залишиться таким самим, як і для класичного випадку. Крім того, часто зміщення подається через вагу входу з одиничним вхідним сигналом, а тому не розглядається як окремий параметр для навчання. Позначимо аргумент функції активації як  $z_j^{(1)}$ :

$$z_j^{(1)} = \sum_{i=1}^n w_{ji}^{(1)} \cdot \gamma^{(1)} \cdot \ln(x_i + 1) + b_j^{(1)}.$$

За правилом обчислення частинних похідних складної функції знайдемо градієнт функції втрат за вагою  $w_{ji}^{(1)}$ :

$$\frac{\partial \mathcal{L}}{\partial w_{ji}^{(1)}} = \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot \frac{\partial a_j^{(1)}}{\partial z_j^{(1)}} \cdot \frac{\partial z_j^{(1)}}{\partial w_{ji}^{(1)}} = \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot f'(z_j^{(1)}) \cdot \gamma^{(1)} \cdot \ln(x_i + 1) = \delta^{(1)} \cdot \gamma^{(1)} \cdot \ln(x_i + 1), \quad (5)$$

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

$$\delta_j^{(1)} \equiv \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot f'(z_j^{(1)}) = \frac{\partial \mathcal{L}}{\partial z_j^{(1)}}, \quad (6)$$

де  $\delta_j^{(1)}$  — помилка нейрона [11]

Аналогічно знайдемо градієнт функції втрат за масштабом  $\gamma^{(1)}$ :

$$\frac{\partial \mathcal{L}}{\partial \gamma^{(1)}} = \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot \frac{\partial a_j^{(1)}}{\partial z_j^{(1)}} \cdot \frac{\partial z_j^{(1)}}{\partial \gamma^{(1)}} = \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot f'(z_j^{(1)}) \cdot \sum_{i=1}^n w_{ji}^{(1)} \cdot \ln(x_i + 1) = \delta_j^{(1)} \cdot \sum_{i=1}^n w_{ji}^{(1)} \cdot \ln(x_i + 1). \quad (7)$$

На відміну від класичного нейрона, для якого ваги оновлюються пропорційно вхідному сигналу, для логарифмічної моделі оновлення ваг згідно (5) масштабується логарифмом від входу. Потенційно це забезпечить стабільність навчання при великих активаціях. Крім того, градієнт за вагою пропорційний масштабу  $\gamma$ . Відповідно значення  $\gamma > 1$  може підвищити швидкість навчання та буде сприяти боротьбі із затуханням градієнта. Градієнт за масштабом  $\gamma$  згідно (7) пропорційний зваженій сумі входів, яка через функцію активації трансформується у вихідний сигнал нейрона. Цей зворотній зв'язок може надати мережі можливість «вирішувати», чи потрібно підсилити ефект логарифмічної трансформації.

Вихідний сигнал нейрона першого шару  $a_j^{(1)}$  (4) є вхідним для нейронів другого шару. Обчислення градієнтів за методом зворотного поширення помилки базується на тому, що виходи нейронів будь-якого шару впливають на функцію втрат через виходи наступного [11]. Використовуючи це правило, виразимо помилку  $\delta_j^{(1)}$  через помилку  $\delta_k^{(2)}$  нейронів другого шару:

$$\frac{\partial \mathcal{L}}{\partial a_j^{(1)}} = \sum_k \frac{\partial \mathcal{L}}{\partial a_k^{(2)}} \cdot \frac{\partial a_k^{(2)}}{\partial z_k^{(2)}} \cdot \frac{\partial z_k^{(2)}}{\partial a_j^{(1)}} = \sum_k \frac{\partial \mathcal{L}}{\partial a_k^{(2)}} \cdot f'(z_k^{(2)}) \cdot \frac{\partial z_k^{(2)}}{a_j^{(1)}} = \sum_k \delta_k^{(2)} \cdot \frac{\partial z_k^{(2)}}{a_j^{(1)}}.$$

З врахуванням (6) для помилки нейрона  $\delta_j^{(1)}$  отримаємо:

$$\delta_j^{(1)} \equiv \frac{\partial \mathcal{L}}{\partial z_j^{(1)}} = \frac{\partial \mathcal{L}}{\partial a_j^{(1)}} \cdot f'(z_j^{(1)}) = f'(z_j^{(1)}) \cdot \sum_k \delta_k^{(2)} \cdot \frac{\partial z_k^{(2)}}{a_j^{(1)}}. \quad (8)$$

Вираз (8) надає можливість обчислювати градієнти (5) та (7) за методом зворотного поширення помилки. В межах цього дослідження другим шаром є шар з лінійними нейронами. Тому помилка  $\delta_j^{(1)}$  визначатиметься як:

$$\delta_j^{(1)} = f'(z_j^{(1)}) \cdot \sum_k \delta_k^{(2)} \cdot w_{kj}^{(2)}. \quad (9)$$

Отже, якщо за логарифмічним шаром йде лінійний шар, то градієнти навчальних параметрів логарифмічного шару, що описуються виразами (5) та (7), будуть пропорційними логарифмам вхідних активацій нейронів. Відповідно оновлення навчальних параметрів нейронів логарифмічного шару при малих вхідних активаціях буде відбуватися більш інтенсивно, а при великих – більш стримано.

### МАТЕРІАЛИ ТА МЕТОДИ

Для дослідження впливу логарифмічної трансформації вхідних сигналів на результат детектування ключових точок обличчя було обрано згорткову мережу ResNet-34. Хоча архітектура ResNet-34 поступається за точністю ResNet-50, вона має менше параметрів, а тому потребує менше обчислювальних ресурсів. Оскільки метою є аналіз впливу логарифмічної трансформації на точність, а не досягнення її максимальних показників, ResNet-34 є кращим варіантом з огляду на обчислювальну складність та часові витрати на проведення експериментів. Для експериментів було використано набір даних WFLW (Wider Facial Landmarks in the Wild). Це складний сучасний датасет, що передбачає локалізацію 98 щільно розташованих ключових точок в умовах значних варіацій емоцій, поз, освітленості, макіяжу, оклюзій та розмиття [12]. Набір WFLW містить 10 000 різноманітних зображень, отриманих в природних умовах, з 98 вручну розміченими ключовими точками. Датасет поділений на навчальну та тестову вибірки, що містять 7 500 та 2 500 зображень відповідно. Тестова вибірка розділена на 6 підгруп: поза, вираз обличчя, освітлення, макіяж, оклюзія та розмиття [13].

Для реалізації ResNet-34, її навчання та тестування був створений програмний застосунок з використанням Python-фреймворку PyTorch та мови програмування Python. Візуалізація даних у вигляді графіків була реалізована з використанням бібліотеки Matplotlib. З використанням PyTorch було створено стандартну мережу ResNet-34. Після цього оригінальний вихідний шар класифікації ResNet-34 було вилучено. У результаті ResNet-34 перетворилася з класифікатора на екстрактор ознак. На наступному кроці до отриманої таким чином моделі був доданий регресійний модуль з двох повнозв'язаних шарів. Перший забезпечував зменшення розмірності та узагальнення ознак. Після нього застосовувалися нелінійна функція активації ReLU та Dropout-регуляризація. Другий повнозв'язаний шар зі 196 нейронів виконував остаточне відображення сформованого представлення у вектор координат (x, y) ключових

---

---

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

---

---

точок. Оскільки вирішуваною мережею задачею є регресія координат, на вихідному шарі не використовувалася додаткова функція активації, що надало можливість безпосередньо прогнозувати дійсні значення координат. У створеній моделі була передбачена можливість її використання у режимі з логарифмічними згортками. Для цього був створений окремий шар, що перед стандартною згорткою забезпечував логарифмічну трансформацію  $\ln(x_i + 1)$  вхідних сигналів з подальшим масштабуванням результату параметром  $\gamma$ .

В PyTorch для задачі регресії як функція втрат (функція помилки) за замовченням використовується Smooth L1:

$$\text{Smooth L1}(\Delta) = \begin{cases} 0.5\Delta^2, & \text{для } \Delta < 1 \\ |\Delta| - 0.5, & \text{для } \Delta \geq 1 \end{cases}$$

де  $\Delta$  — різниця між прогнозованим та реальним значенням.

Це окремий випадок більш загальної функції Huber loss [14] з порогом  $\delta$ , при якому квадратична функція переходить в лінійну:

$$L_{\text{Huber}}(\Delta) = \begin{cases} 0.5\Delta^2, & \text{для } \Delta \leq \delta \\ \delta(|\Delta| - 0.5\delta), & \text{для } \Delta > \delta \end{cases}$$

При малих  $\Delta$  Smooth L1 є квадратичною функцією, тому її похідна при  $\Delta \rightarrow 0$  також прямує до 0. Відповідно на фінальних етапах навчання вплив малих помилок на корекцію ваг стає незначним, що може обмежити точність локалізації ключових точок. Для вирішення цієї проблеми як функцію втрат було використано функцію Wing Loss [15]:

$$L_{\text{wing}}(\Delta) = \begin{cases} w \cdot \ln(1 + |\Delta|/\varepsilon), & \text{для } \Delta < w \\ |\Delta| - C, & \text{для } \Delta \geq w \end{cases}$$

де  $w$  та  $\varepsilon$  — деякі параметри,  $C = w - w \ln(1 + w/\varepsilon)$ .

Через логарифмічну залежність при малих  $\Delta$ , Wing Loss підсилює вплив незначних помилок, що дозволяє моделі досягати вищої точності.

Зображення в наборі даних WFLW мають різну розмірність, тоді як архітектура ResNet-34 працює із зображеннями фіксованого розміру 224 x 224. Тому перед подачею у мережу на зображенні попередньо виділялася області обличчя з додаванням невеликого відступу для врахування контексту навколо нього. Після цього виділена область вирізалася та масштабувалася до розміру 224 x 224. Одночасно перераховувалися координати ключових точок відносно отриманого таким чином зображення. Фінальний етап підготовки полягав в нормалізації піксельних значень і координат. Значення пікселів переводилися з діапазону [0; 255] у діапазон [0; 1], після чого зображення перетворювалося з формату Висота x Ширина x Кількість каналів у формат Кількість каналів x Висота x Ширина, з яким працює PyTorch. Координати ключових точок також нормалізувалися діленням на розмір зображення, тобто модель навчалася передбачати координати у відносному діапазоні від 0 до 1, а не в абсолютних пікселях. Це спростило навчання і дозволило зробити отримувані на виході значення незалежними від конкретного розміру зображення.

Навчання мережі відбувалося за методом стохастичного градієнтного спуску (Stochastic Gradient Descent, SGD). Розміри пакетів було вибрано рівними 64. Як функція втрат використовувалася Wing Loss з параметрами  $w = 10$  та  $\varepsilon = 2$ . Швидкість навчання була вибрана 0,0001. Кількість епох 15. Для оцінки якості моделі під час навчання тренувальний набір WFLW з 7500 зображень було розділено на навчальну та перевірку вибірку у пропорції 9:1 (90% зображень склали навчальну вибірку, 10% – перевірку). Для кожної епохи навчальна вибірка перемішувалася, порядок зразків для перевіркової вибірки залишався незмінним. Для забезпечення стабільності результатів використовувався генератор випадкових чисел із фіксованим зерном (seed).

Дослідження проводилося для двох мереж: класичної ResNet-34 (базова модель) та ResNet-34 з логарифмічною трансформацією (3) у нейронах першого згорткового шару (логарифмічна модель). Умови навчання для обох моделей були повністю ідентичними: навчання та перевірка на кожній з епох відбувалася на тих самих пакетах. Оцінка точності моделей на тестових даних здійснювалася за нормованою середньою похибкою (Normalized Mean Error, NME) [14]:

$$NME = \frac{1}{N} \sum_{i=1}^N \frac{\sqrt{\Delta x_i^2 + \Delta y_i^2}}{size},$$

де  $size$  – розмір зображення.

### РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

На рис. 1 наведені залежності, що відображають зміну помилки у локалізації ключових точок під час навчання базової моделі. Динаміка зміни помилки під час навчання логарифмічної моделі показана

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

на рис. 2. Для обох моделей спостерігається стабільне зменшення функції втрат на навчальній вибірці (Train Loss) протягом усіх епох, що свідчить про ефективність процесу навчання. Найбільш інтенсивне зниження помилки відбувається на початкових етапах: зі значення 0,52 до 0,22 за перші 3 епохи. Надалі темп зниження поступово уповільнюється, досягаючи значення 0,15 на 15 епосі. Для обох моделей криві функцій втрат на навчальній вибірці є майже ідентичними, що вказує на те, що логарифмічна трансформація вхідних активацій у нейронах першого згорткового шару не має суттєвого впливу на процес навчання.

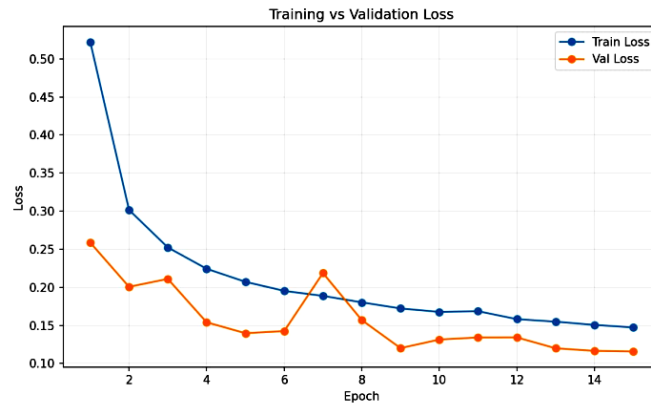


Рисунок 1 – Динаміка зміни помилки під час навчання для базової моделі

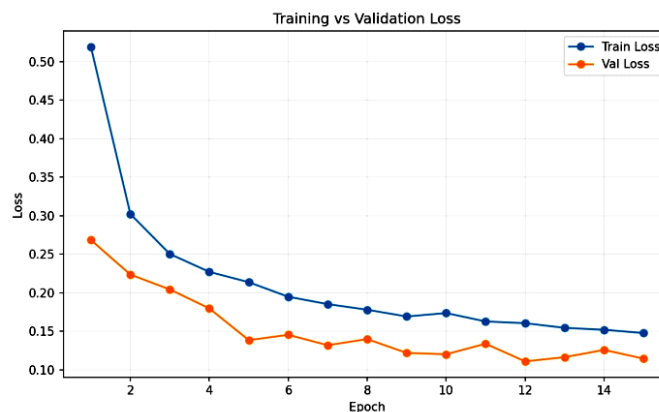


Рисунок 2 – Динаміка зміни помилки під час навчання для логарифмічної моделі

На перевірочних даних функція втрат (Val Loss) також зменшується протягом усіх епох. Протягом перших 6 епох спостерігається більш динамічне зменшення зі значення приблизно 0,27 до 0,15. За решту 9 епох швидкість зменшення уповільнюється з незначними коливаннями між окремими епохами, досягаючи рівня 0,11...0,12. Для логарифмічної моделі ці флуктуації є менш вираженими, що вказує на вищу стабільність оптимізації. У середньому помилка на перевірочних даних для обох моделей залишається меншою за помилку на навчальних, що свідчить про їх високу здатність до узагальнення.

Точність локалізації ключових точок за метрикою NME на тестовій вибірці датасету WFLW для базової моделі склала 0,0431, для логарифмічної – 0,0415. Таким чином, застосування логарифмічної трансформації майже не вплинуло на нормалізовану середню похибку. Більш значущий ефект від застосування логарифмічної трансформації був отриманий при роботі із зображеннями низької яскравості. При зменшенні яскравості тестових зображень у 3 рази нормалізована середня похибка для базової моделі склала 0,0481, для логарифмічної – 0,0461. При 6-ти кратному зниженні яскравості результати склали 0,0639 та 0,0565, відповідно. Тестування на зображеннях зі зниженою яскравістю проводилося без перенавчання моделей на зображеннях зі зменшеною яскравістю.

Оскільки узагальнювальна здатність нейромережі, навченої за методом стохастичного градієнтного спуску, залежить від розподілу даних у навчальних пакетах, було проведено серію експериментів з різними значеннями параметра випадковості (seed) для отримання альтернативних розподілів навчальної вибірки. Результати експериментів наведені у табл. 1.

## СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

Таблиця 1 – Результати тестування базової та логарифмічної моделі

№	Нормалізована середня похибка NME					
	Базова яскравість зображення		Яскравість зменшена у 3 рази		Яскравість зменшена у 6 разів	
	Базова модель	Log модель	Базова модель	Log модель	Базова модель	Log модель
1	0,0431	0,0415	0,0481	0,0461	0,0639	0,0565
2	0,0405	0,0416	0,0485	0,0466	0,0713	0,0615
3	0,0405	0,0404	0,0489	0,0479	0,0752	0,0694
4	0,0405	0,0400	0,0475	0,0449	0,0637	0,0547
5	0,0419	0,0426	0,0485	0,0464	0,0641	0,0587
6	0,0447	0,0431	0,0483	0,0466	0,0610	0,0557
Середнє значення	0,0419	0,0415	0,0483	0,0464	0,0665	0,0594

На незмінених зображеннях обидві моделі показали приблизно однакові результати: середнє значення NME для базової моделі становило 0,0419, для логарифмічної – 0,0415. Отже, використання логарифмічної трансформації вхідних активацій у нейронах першого згорткового шару майже не впливає на точність локалізації ключових точок обличчя на зображеннях WFLW. Найкращим результатом стало зниження похибки на 0,0016 (результати №1 та №6). При деяких розподілах навчальної вибірки значення NME логарифмічної моделі було навіть більшим (результати №2 та №5). При зниженні яскравості зображень перевага логарифмічної моделі стає більш вираженою. Так, при 3-ти кратному зниженні яскравості середнє значення NME для логарифмічної моделі склало 0,0464 проти 0,0483 для базової, при 6-ти кратному – 0,0594 проти 0,0665. У середньому похибка NME для логарифмічної моделі була меншою на 0,0004, 0,0019 та 0,0071, відповідно. При цьому логарифмічна модель показала більшу стабільність: для усіх випадків зі зниженою яскравістю вона перевершила базову за точністю. При 3-ти кратному зниженні яскравості середня похибка базової моделі зросла на 15,27%, логарифмічної – на 11,80%; при 6-ти кратному – на 58,71% і 43,13%, відповідно. Отримані результати свідчать про те, що логарифмічна трансформація вхідних активацій може розглядатися як один з інструментів підвищення інваріантності згорткових нейронних мереж до змін яскравості. Водночас слід зазначити, що і базова архітектура ResNet демонструє високу точність локалізації ключових точок за різних рівнів яскравості зображень (середнє значення NME  $\approx$  0,04 на незмінених зображеннях і менше 0,07 при значному зниженні яскравості). Це ймовірно пов'язано із впровадженням у ResNet пакетною нормалізацією (Batch Normalization), яка дозволяє [16] «вирівнювати» дані на кожному шарі мережі, роблячи їх незалежними від абсолютних значень.

### ВИСНОВКИ

Логарифмічне перетворення вхідних активацій всередині нейрона дозволяє трансформувати мультиплікативні зміни сигналів на його входах в адитивні, що забезпечує чутливість до відносних змін сигналів подібно до принципів біологічного сприйняття. У роботі досліджено вплив логарифмічної трансформації виду  $y \ln(x + 1)$ , де  $y$  є навчальним параметром, у нейронах першого згорткового шару на точність локалізації ключових точок обличчя та стійкість моделі до варіацій яскравості зображень.

Експериментальні результати з локалізації ключових точок обличчя на тестовій вибірці датасету WFLW показали, що використання модифікованих нейронів у першому згортковому шарі не має суттєвого впливу на точність локалізації на зображеннях з базовою яскравістю, проте підвищує стійкість моделі до її зниження. Порівняно з базовою моделлю ResNet-34, приріст нормованої середньої похибки NME для моделі з логарифмічною трансформацією був меншим: у середньому 11,80% проти 15,27% при трикратному та 43,13% проти 58,71% при шестикратному зниженні яскравості вхідних зображень.

Запропонований підхід не потребує ускладнення архітектури моделі та може бути інтегрований в існуючі нейромереві моделі з мінімальними обчислювальними витратами. Подальші дослідження можуть бути спрямовані на аналіз ефективності поєднання логарифмічної трансформації з іншими архітектурними рішеннями, зокрема з механізмами каскадної уваги та багатомасштабним аналізом ознак.

**СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ / REFERENCES**

1. Colaco, S. & Han, D. S. (2020) Facial Keypoint Detection with Convolutional Neural Networks. In: International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Fukuoka, Japan, 2020, pp. 671-674. doi: 10.1109/ICAIIIC48513.2020.9065279.
2. Haq, M. U., Sethi, M. A. J., Ahmad, S. et al. (2025) A Comprehensive Review of Face Detection/Recognition Algorithms and Competitive Datasets to Optimize Machine Vision. Computers, Materials & Continua, vol. 84 (1), pp. 1-24. doi: org/10.32604/cmc.2025.063341
3. Dwivedi, P. & Sharan, B. (2022) Deep Inception Based Convolutional Neural Network Model for Facial Key-Points Detection. In: International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 792-799. doi: 10.1109/ICCCIS56430.2022.10037639.
4. Gao, J. & Yang, T. (2022) Research on Real-Time Face Key Point Detection Algorithm Based on Attention Mechanism. Computational Intelligence and Neuroscience, vol. 2022, article number 6205108. doi: 10.1155/2022/6205108.
5. Verma, S., Singhal, P., Gupta, R. et al. (2024) Facial Keypoint Detection using a Modified Convolutional Neural Network with RESNET50. In: 2nd International Conference on Advancements and Key Challenges in Green Energy and Computing (AKGEC), Ghaziabad, India, 2024, pp. 1-5. doi: 10.1109/AKGEC62572.2024.10868470.
6. Huang, Y., Chen, Y., Wang, J. et al. (2024) A Robust and Efficient Method for Effective Facial Keypoint Detection. Applied Sciences, vol. 14 (16), article number 7153. doi: 10.3390/app14167153.
7. Maes, C. (2025) Statistical Mechanical Foundation of Weber-Fechner Laws. Frontiers in Neuroscience, vol. 19, article number. 1532069. doi: 10.3389/fnins.2025.1532069.
8. Du, K.-L., Leung, C.-S., Mow, W.H., Swamy, M.N.S. (2022) Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era. Mathematics, vol. 10(24), article number 4730. doi: 10.3390/math10244730
9. Sendjasni, A., Traparic, D. & Larabi, M. -C. (2022) Investigating Normalization Methods for CNN-Based Image Quality Assessment. In: IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 4113-4117. doi: 10.1109/ICIP46576.2022.9897268.
10. Kim, Y.-S., Kim, M. K., Fu, N. et al. (2025) Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models. Sustainable Cities and Society, vol. 118, article number 105570. doi: 10.1016/j.scs.2024.105570.
11. Bishop, C. M. (2006) Pattern Recognition and Machine Learning. New York : Springer, 738 p.
12. Khabarlak, K. & Koriashkina, L. (2022) Fast Facial Landmark Detection and Applications: A Survey. Journal of Computer Science & Technology, vol. 22 (1), pp. 12-41. doi: 10.24215/16666038.22.e02.
13. Prados-Torreblanca, A., Buenaposada, J. M. & Baumela, L. (2022) Shape preserving facial landmarks with graph attention networks. arXiv. doi: 10.48550/arXiv.2210.07233.
14. Terven, J., Cordova-Esparza, DM., Romero-González, JA., et al. (2025) A comprehensive survey of loss functions and metrics in deep learning. Artif Intell Rev, vol. 58, article number 195. doi: 10.1007/s10462-025-11198-7.
15. Feng, ZH., Kittler, J., Awais, M. et al. (2020) Rectified Wing Loss for Efficient and Robust Facial Landmark Localisation with Convolutional Neural Networks. International Journal of Computer Vision, vol. 128, pp. 2126-2145. doi: 10.1007/s11263-019-01275-0.
16. Benz, P., Zhang, C., Karjauv, A. & Kweon I. S. (2021) Revisiting Batch Normalization for Improving Corruption Robustness. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021, pp. 494-503, doi: 10.1109/WACV48630.2021.00054.

*Дата надходження: 05.03.2026*

*Дата прийняття до друку після рецензування: 20.04.2026*

*Дата публікації: 18.06.2026*

*Ця робота ліцензується відповідно до  
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)*

---

---

**СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ  
З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ**

---

---

**ТАРНОВСЬКИЙ АРТЕМ МИКОЛАЙОВИЧ** — аспірант кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця, ***e-mail: [tarnovskiy0211@gmail.com](mailto:tarnovskiy0211@gmail.com)***, [\*\*https://orcid.org/0009-0006-0811-8611\*\*](https://orcid.org/0009-0006-0811-8611)

**ЗАХАРЧЕНКО СЕРГІЙ МИХАЙЛОВИЧ** — к.т.н., проф. кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця, ***e-mail: [zahar@i.ua](mailto:zahar@i.ua)***, [\*\*https://orcid.org/0000-0003-3977-2908\*\*](https://orcid.org/0000-0003-3977-2908)

**ТАРНОВСЬКИЙ МИКОЛА ГЕННАДІЙОВИЧ** — к.т.н., доц. кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця, ***e-mail: [ntarn@vntu.edu.ua](mailto:ntarn@vntu.edu.ua)***, [\*\*https://orcid.org/0009-0002-0064-1310\*\*](https://orcid.org/0009-0002-0064-1310)

**Artem TARNOVSKYI, Sergii ZAKHARCHENKO, Mykola TARNOVSKYI**

**IMPACT OF LOGARITHMIC TRANSFORMATION OF INPUT ACTIVATIONS IN  
CONVOLUTIONAL NETWORKS ON FACIAL LANDMARK LOCALIZATION**

Vinnytsia National Technical University