

ПОРІВНЯЛЬНИЙ АНАЛІЗ ФУНКЦІЙ ВТРАТ ДЛЯ РЕГРЕСІЇ КООРДИНАТ КЛЮЧОВИХ ТОЧОК ОБЛИЧЧЯ НА ОСНОВІ АРХІТЕКТУРИ RESNET

Вінницький національний технічний університет

Анотація

Досліджено вплив функції втрат на точність регресії координат ключових точок обличчя з використанням архітектури ResNet-34 та датасету WFLW. Здійснено порівняльний аналіз трьох типів функцій втрат: середньоквадратичної помилки (MSE), Huber loss та спеціалізованої Wing loss. Оцінка точності проводилася за метрикою нормованої середньої похибки (NME). Результати продемонстрували перевагу використання функцій втрат Wing loss для задач локалізації ключових точок обличчя через забезпечення більшої точності регресії.

Ключові слова: комп'ютерний зір, ключові точки обличчя, нейронні мережі, ResNet, функція втрат

Abstract

The influence of the loss function on the accuracy of the regression of the coordinates of key points of the face using the ResNet-34 architecture and WFLW dataset was studied. A comparative analysis of three types of loss functions was performed: mean square error (MSE), Huber loss, and specialized Wing loss. The accuracy was assessed using the normalized mean error (NME) metric. The results demonstrated the advantage of using Wing loss functions for the localization of key points of the face due to the provision of greater regression accuracy.

Keywords: computer vision, key points of the face, neural networks, ResNet, loss function

Вступ

Детектування ключових точок обличчя є важливим етапом при розв'язку завдань комп'ютерного зору у застосунках, пов'язаних з розпізнаванням обличчя, аналізом виразу обличчя, 3D-реконструкцією обличчя тощо [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. Глибокі архітектури на основі різних моделей згорткових мереж, перевершують класичні методи за точністю та стійкістю до змін сценаріїв, таких як рівень освітлення, поворот голови, часткове перекриття обличчя [2], [3].

У роботі [4] на прикладі архітектури ResNet-50 показано, що використання залишкових зв'язків та попередньо навчених ваг дозволяє досягти стабільної регресії координат ключових точок навіть у глибоких мережах із забезпеченням високої точності на складних датасетах.

Одним з критично важливих моментів глибокого навчання є вибір функції втрат, що використовуються під час навчання моделей для оцінювання якості апроксимації бажаного результату [5].

Метою дослідження є аналіз впливу функції втрат на точність локалізації ключових точок обличчя.

Результати дослідження

Вибір функції втрат (функції вартості або цільової функції), що під час навчання вимірює різницю між передбачуваними та очікуваними результатами для оптимізації параметрів моделі шляхом мінімізації цієї різниці, є ключовим моментом для успіху у завданнях глибокого навчання. Функція втрат спрямовує процес оптимізації, вказуючи, як потрібно коректувати параметри моделі для мінімізації похибок та підвищення точності передбачень. Для задачі локалізації ключових точок обличчя як задачі регресії можуть бути використані такі функції втрат як середньоквадратична помилка (Mean Squared Error, MSE), Huber loss та Wing loss [5], [6] (табл.1).

Дослідження того, як вибір функції втрат впливає на точність локалізації ключових точок було проведено з використанням згорткової мережі ResNet-34. Архітектура ResNet-34 поступається за точністю ResNet-50, проте вона містить менше параметрів, а тому працює швидше та потребує менше

пам'яті. Це робить її більш придатною для дослідження через зменшення витрат часу на проведення експериментів. Експерименти проводилися на наборі даних WFLW (Wider Facial Landmarks in the Wild). Це складний сучасний датасет, що потребує передбачення 98 щільно розташованих ключових точок в умовах значних варіацій емоцій, поз, освітленості, перекриття та розмиття. Датасет WFLW включає 10 000 різних зображень, отриманих в природних умовах, 7500 з яких утворюють тренувальну вибірку та 2500 тестову [7].

Табл. 1. Функції втрат для задачі локалізації ключових точок обличчя

Функція втрат	Визначення	Параметри
MSE	$MSE = \frac{1}{N} \sum_{i=1}^N \sqrt{\Delta x_i^2 + \Delta y_i^2}$	Δx_i та Δy_i — різниці координат передбаченої та цільової точок N — кількість точок
Huber loss	$L_{Huber}(\Delta) = \begin{cases} \frac{1}{2} \Delta^2, & \text{для } \Delta \leq \delta \\ \delta \left(\Delta - \frac{1}{2} \delta \right), & \text{для } \Delta > \delta \end{cases}$	Δ — різниця між істиною та передбаченою координатою δ — деякий поріг (зазвичай $\delta = 1$)
Wing loss	$L_{wing}(\Delta) = \begin{cases} w \cdot \ln(1 + \Delta /\epsilon), & \text{для } \Delta < w \\ \Delta - C, & \text{для } \Delta \geq w \end{cases}$	Δ — різниця між істиною та передбаченою координатою w та ϵ — деякі параметри $C = w - w \ln(1 + w/\epsilon)$

Для реалізації ResNet-34, її навчання та тестування був створений програмний застосунок з використанням Python-фреймворку PyTorch та мови програмування Python. З використанням PyTorch було створено стандартну мережу ResNet-34. Після цього оригінальний вихідний шар класифікації ResNet-34 був замінений на регресійний модуль з двома повноз'язаними шарами. Перший забезпечує зменшення розмірності та узагальнення ознак. Після чого застосовується функція активації ReLU та Dropout-регуляризація. Другий повноз'язаний шар містить 196 нейронів і забезпечує формування координат x та y 98 ключових точок обличчя. Перед подачею у мережу зображення WFLW спочатку обрізалася до області обличчя, а потім масштабувалася до розміру 224 x 224. Одночасно перераховувалися координати ключових точок. На фінальному етапі відбувалася нормалізація піксельних значень і координат, переведенням їх до діапазону [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.

Табл. 2. Середні значення NME при локалізації ключових точок обличчя на дадасеті WFLW для різних функцій втрат

Функція втрат	Середні значення NME, %
MSE	4,61
Huber loss	4,59
Wing loss	4,14

Результати експериментів, що полягали в оцінці точності моделі після її навчання за методом стохастичного градієнтного спуску (SGD) з різними функціями втрат продемонстрували, що більша точність на тестовому наборі даних забезпечується при використанні функції Wing loss (табл. 2). Умови навчання з розглядуваними функціями втрат були ідентичними: навчання відбувалося на тих самих пакетах, що формувалися з навчальної вибірки випадковим чином, кількість епох була однаковою. Оцінка точності моделі на тестових даних відбувалася за нормованою середньою похибкою (Normalized Mean Error, NME) [5]:

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

де $size$ – розмір зображення.

Висновки

Експериментально показано, що функція втрат Wing loss є найбільш ефективною для задачі регресії координат ключових точок обличчя. Вона показала найкращий результат (NME = 4,14%), що на 10,2%

краще за показник стандартної функції MSE (4,61%). Це пояснюється здатністю Wing loss створювати більші градієнти для малих помилок, що забезпечує більш точну апроксимацію координат точок на фінальних етапах навчання. Функція Huber loss продемонструвала проміжний результат (4,59%), виявившись дещо ефективнішою за MSE завдяки своїй комбінованій структурі, яка зменшує вплив аномальних значень у навчальній вибірці.

Отримані результати можуть бути використані при розробці систем розпізнавання обличчя, аналізу емоцій та систем доповненої реальності, де критично важливою є висока точність локалізації окремих елементів (очей, губ, контурів обличчя) за умов низької якості вхідних зображень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Colaco S. and Han D. S. Facial Keypoint Detection with Convolutional Neural Networks // International Conference on Artificial Intelligence in Information and Communication (ICAIC), Fukuoka, Japan, 2020, pp. 671-674, doi: 10.1109/ICAIC48513.2020.9065279.
2. Haq M. U., Sethi M. A. J., Ahmad S. et al. A Comprehensive Review of Face Detection/Recognition Algorithms and Competitive Datasets to Optimize Machine Vision // Computers, Materials & Continua, 2025, vol. 84, no. 1, pp. 1–24. doi: 10.32604/cmc.2025.063341
3. Dwivedi P. and Sharan B. Deep Inception Based Convolutional Neural Network Model for Facial Key-Points Detection // International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 792-799. doi: 10.1109/ICCCIS56430.2022.10037639.
4. Verma S., Singhal P., Gupta R. et al. Facial Keypoint Detection using a Modified Convolutional Neural Network with RESNET50 // 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.
5. Terven J., Cordova-Esparza DM., Romero-González JA., et al. A comprehensive survey of loss functions and metrics in deep learning. Artif Intell Rev, 2025, vol. 58, article number 195. doi.org/10.1007/s10462-025-11198-7.
6. Feng,ZH., Kittler J., Awais M. et al. Rectified Wing Loss for Efficient and Robust Facial Landmark Localisation with Convolutional Neural Networks // International Journal of Computer Vision, 2020, vol. 128, pp. 2126–2145. doi: 10.1007/s11263-019-01275-0.
7. Khabarлак K., Koriashkina L. Fast Facial Landmark Detection and Applications: A Survey // Journal of Computer Science & Technology, 2022, vol. 22(1), pp. 12–41. doi: 10.24215/16666038.22.e02.

Тарновський Артем Миколайович — аспірант кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця, e-mail: tarnovskiy0211@gmail.com.

Захарченко Сергій Михайлович — к.т.н., проф. кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця.

Tarnovskyi Artem M. — Post-Graduate Student of the Department of Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: tarnovskiy0211@gmail.com

Zakharchenko Serhii M. — Cand. Sc. (Eng), Professor, the Department of Computer Engineering, Vinnytsia National Technical University, Vinnytsia.