

ТИМКІВ РОМАН

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0005-9030-0641>e-mail: roman.tymkiv.mknssh.2024@lpnu.ua**ГОРУН ПАВЛО**

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0008-4296-5560>e-mail: pavlo.p.horun@lpnu.ua

МЕТОД ПОСЛІДОВНОГО НАВЧАННЯ З РОЗМЕЖОВАНИМИ ПІДМОДЕЛЯМИ ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ІЄРАРХІЧНОЇ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ОДЯГУ

У статті представлено метод послідовного навчання спеціалізованих моделей для ієрархічної класифікації зображень одягу на основі модифікованої архітектури ResNet-50. На відміну від традиційного підходу з одною моделлю, одним входом та комбінованими виходами, у цій статті виконано побудову трьох окремих моделей, що навчаються послідовно для класифікації категорій, підкатегорій та атрибутів. Експериментальне дослідження проведено на наборі даних Fashion Product Images (Small) з фокусом на категоріях "Одяг" та "Взуття". Результати демонструють суттєве покращення точності класифікації підкатегорій з 78,8% до 83,9% на повному наборі даних та стійкість до зниження точності класифікації в умовах обмежених даних – при зменшенні навчальної вибірки в 10 разів точність класифікації підкатегорій зростає з 32,4% до 70,2%. Розроблений метод забезпечує зменшення сумарного розміру моделей на 45% та швидшу збіжність (21-24 епохи проти 40). Практична цінність методу полягає у можливості ефективного класифікації в умовах обмежених обчислювальних ресурсів та навчальних даних, що є типовою ситуацією для спеціалізованих доменів електронної комерції.

Ключові слова: ієрархічна класифікація, ResNet-50, глибоке навчання, розпізнавання одягу, електронна комерція, обмежені дані.

TYMKIV ROMAN**HORUN PAVLO**

Lviv Polytechnic National University

SEQUENTIAL LEARNING METHOD WITH DECOUPLED SUB-MODELS FOR IMPROVING THE EFFICIENCY OF HIERARCHICAL APPAREL IMAGE CLASSIFICATION

This paper presents a sequential training method for specialized models in hierarchical fashion image classification based on a modified ResNet-50 architecture. Unlike the traditional approach with a single model, single input, and combined outputs, we developed three separate models trained sequentially for classifying categories, subcategories, and attributes. The key innovation lies in training the subcategory model first with full ResNet-50 optimization, then freezing these learned features for subsequent category and attribute models. This approach allows each model to specialize in its specific classification task while maintaining computational efficiency. Experimental research was conducted on the Fashion Product Images (Small) dataset, focusing on "Apparel" and "Footwear" categories comprising 24'996 images. The dataset presents significant challenges including class imbalance, with some subcategories containing fewer than 100 samples. Results demonstrate significant improvement in subcategory classification accuracy from 78.8% to 83.9% on the full dataset and robustness to accuracy degradation under limited data conditions— with a 10-fold reduction in training samples, subcategory classification accuracy increases from 32.4% to 70.2%, representing a 2.2x improvement. The developed method provides a 45% reduction in total model size (24.8M vs 45.6M parameters) and faster convergence (21-24 epochs versus 40). Each specialized model employs a simplified MLP architecture with 512-256 neurons configuration, dropout regularization of 0.3, and L2 regularization of 0.0005. The practical value of the method lies in enabling efficient classification under limited computational resources and training data constraints, which is typical for specialized e-commerce domains. While attribute classification accuracy slightly decreased from 98.6% to 92%, this trade-off is acceptable given the substantial improvements in other metrics and the method's superior performance in data-scarce scenarios, making it particularly suitable for real-world e-commerce applications where obtaining large annotated datasets is challenging.

Keywords: hierarchical classification, ResNet-50, deep learning, clothing recognition, e-commerce, limited data.

Стаття надійшла до редакції / Received 06.10.2025

Прийнята до друку / Accepted 15.11.2025

Постановка проблеми

Розвиток технологій машинного навчання та комп'ютерного зору відкрив широкі можливості для автоматизації бізнес-процесів у сфері електронної комерції. Класифікація та розпізнавання продуктів за зображеннями є важливою частиною цього процесу, оскільки дозволяє оптимізувати пошук товарів, автоматизувати створення описів та підвищити точність рекомендаційних систем. На практиці виникають додаткові ускладнення: різноманітні ракурси та фон зображень, оклюзії, дисбаланс класів, неповні дані та ін.

Метою даного дослідження є підвищення точності класифікації одягу та його атрибутів на основі зображень, що може бути використаним зокрема для автоматичної генерації описів для онлайн-магазинів, підвищення якості рекомендаційних систем і пошуку за зображенням тощо.

Так, одним із завдань є створення багаторівневої моделі класифікації зображень одягу (категорія–підкатегорія–атрибут), здатної зберігати високу точність в умовах обмежених даних, що має практичну цінність для розвитку систем електронної комерції.

Аналіз публікацій

У сучасних дослідженнях для розпізнавання одягу та його атрибутів застосовуються різні методи комп'ютерного зору. Так, у роботі [1] поєднано класичні SVM-класифікатори для розпізнавання окремих атрибутів з архітектурою VGG16. Це дозволило ефективно відображати складні візерунки у зображеннях та забезпечило баланс між обчислювальною ефективністю і глибиною аналізу, досягнувши загальної точності розпізнавання 81.4% для верхнього одягу та 85.72% для нижнього одягу. Хоча метод показав високу точність для деяких атрибутів (класифікація статі - 95.89%, візерунків - 89.81%, типу нижнього одягу - 91.2%), точність визначення інших характеристик залишалася недостатньою: класифікація віку досягла лише 77.28%, виявлення застібок-блискавок - 70%, визначення типу коміра - 78%, а найгірший результат показала класифікація матеріалів верхнього одягу - лише 57.04%.

У дослідженні [2] було використано алгоритм вибіркового пошуку для виділення регіонів-кандидатів, їх аналіз за допомогою мережі InceptionResNet V1 із L-Softmax та застосування Soft-NMS для відбору найбільш релевантних областей. Додатково використовувалася проста нейронна мережа для уточнення меж об'єктів. Метод досяг загальної точності маркування 87.77%, точності 73.59% та повноти 83.84% на датасеті, що включав близько 100 000 зображень сорочок. Метод показав особливо високу ефективність у розпізнаванні таких атрибутів як форми комірів. Проте, як і всі підходи на основі RCNN з використанням вибіркового пошуку, основним недоліком стала значна обчислювальна складність та низька швидкість роботи порівняно з сучасними архітектурами.

Робота [3] удосконалює архітектуру YOLOv4-Tiny шляхом інтеграції багатомасштабного модуля виділення ознак Res2Net для оптимізації базової мережі та адаптованого алгоритму кластеризації K-Means для оптимізації параметрів якірних блоків. Це дозволило підвищити точність виявлення об'єктів одягу у випадках різних масштабів, часткових оклюзій та виходу за межі кадру. Експериментальні результати на датасеті DeepFashion2 показали, що запропонований метод досяг mAP 54.14%, що на 6.75% вище порівняно з оригінальною моделлю YOLOv4-Tiny (47.39%). Метод також значно перевершив інші легковагові архітектури: FBNet (24.18%), GhostNet (31.42%), ShuffleNet (33.18%) та MobileNet (39.33%). Однак, незважаючи на суттєве покращення, точність все ще поступається повнорозмірним моделям детекції об'єктів.

У дослідженні [4] розроблено алгоритм багаторівневого об'єднання ознак (Multideep Feature Fusion) для класифікації стилю одягу. Метод використовує модель виявлення цілей для виділення трьох категорій областей зображення (глобальна область, основна частина та локальні області одягу) та вдосконалений ResNet для виділення глибинних ознак. Експериментальні результати показали, що запропонований метод досяг точності 94.97%, що на 1.71% вище порівняно з традиційним ResNet-101 (93.26%), на 3.91% вище за Inception_v3 (91.06%) та на 5.21% вище за VGG16 (89.76%). Час навчання становив 83 години, що порівняно з ResNet-101 (81 година), але значно довше ніж у Inception_v3 (39 годин) та VGG16 (44 години). Підхід демонструє ефективність у класифікації стилю одягу, однак не забезпечує оптимального розпізнавання багатокатегорійних елементів, особливо штанів з високим рівнем оклюзії (перекриття понад 30-40%), де точність знижується помітно.

У праці [5] авторами запропоновано модифіковану уніфіковану модель ResNet50. Вона вирізняється простотою архітектури та стабільними результатами як у класифікації категорій, так і атрибутів. Модель була навчена на приватному датасеті Farfetch з ~350 тис. зображень і досягає точності 83.53% для категорій, 42.68% для підкатегорій та 49.22% для атрибутів (при recall 86.75%). Основними недоліками є: потреба у великому обсязі даних (350 тис. зображень), закритість датасету, висока складність моделі (46.9M параметрів) та відсутність дослідження роботи з обмеженими даними.

Варто відзначити дослідження [6], де представлено комплексний підхід до ієрархічної класифікації зображень одягу з використанням ансамблів нейронних мереж. Автори застосували комбінацію архітектур EfficientNet та Vision Transformer, досягнувши точності 91.3% для категорій та 76.4% для підкатегорій на датасеті DeepFashion2, який містить понад 800 тис. зображень. Метод використовує механізм уваги для виділення ключових візуальних ознак та адаптивне зважування прогнозів від різних моделей. Особливо цікавим є застосування стратегії навчання з учителем, де більша модель-вчитель передає знання компактнішій моделі-учню, зберігаючи 95% точності при зменшенні розміру моделі на 70%. Проте дослідження зосереджене виключно на високоякісних зображеннях з електронної комерції та не розглядає сценарії з обмеженими наборами даних або зображеннями низької якості. Також автори не досліджували можливість послідовного навчання спеціалізованих моделей, що могло б додатково покращити ефективність системи.

У дослідженні [7] запропоновано новаторську архітектуру SAC (Self-attention and CNNs), яка поєднує механізми самоуваги з згортковими нейронними мережами для тонкозернистої класифікації атрибутів одягу. Автори використовують модифіковану архітектуру ResNet-50 з додаванням двох блоків самоуваги та проміжних CNN-шарів, досягаючи середньої точності 81.02% на власному датасеті з 15'025 зображень та 37 атрибутами. Механізм самоуваги дозволяє моделі автоматично фокусуватися на найбільш релевантних частинах зображення для кожного атрибута, що підтверджується візуалізацією через Grad-CAM. Особливо високу ефективність метод показав для атрибутів "стать" (91.97%), "колір" (69.43%) та "наявність ґудзиків" (81.02%). Проте дослідження має кілька обмежень: датасет не є публічно доступним, що ускладнює відтворення результатів; модель тестувалася лише на зображеннях верхнього одягу; не досліджувалася робота в умовах обмежених даних. Крім того, архітектура з п'ятьма додатковими шарами після ResNet-50 та двома блоками самоуваги збільшує обчислювальну складність порівняно з базовими CNN, що може бути критичним для застосувань у реальному часі.

В роботі [8] запропоновано практичний підхід до класифікації одягу з використанням архітектури Vision Transformer (ViT), адаптованої для роботи з обмеженими обчислювальними ресурсами. Автори представили модифіковану версію ViT-Base з 12 шарами трансформерів, яка досягла точності 92.4% на датасеті Fashion-MNIST та 89.7% на розширеному датасеті DeepFashion. Ключовою інновацією стало впровадження механізму динамічної уваги з адаптивним патч-розміром (16×16 для загальної класифікації та 8×8 для деталізованого аналізу), що дозволило зменшити обчислювальну складність на 35% порівняно зі стандартною архітектурою ViT. Метод показав особливу ефективність у розпізнаванні складних візерунків (точність 87.2%) та визначенні типу тканини (точність 83.6%). Проте основним недоліком залишається висока вимогливість до обсягу навчальних даних – для досягнення оптимальних результатів потрібно мінімум 50 тис. анотованих зображень, а також тривалий час початкового навчання (в середньому 72 години на GPU Tesla V100).

Аналізуючи праці в галузі автоматичного розпізнавання характеристик одягу, можна виокремити декілька підходів до виділення ознак. Так, у дослідженні [9] представлено комплексний метод багаторівневого видобування візуальних характеристик одягу з використанням глибоких нейронних мереж та традиційних дескрипторів. Автори пропонують гібридну архітектуру, що поєднує CNN для виділення глобальних семантичних ознак (форма виробу, стиль) з SIFT та HOG дескрипторами для локальних деталей (текстура тканини, декоративні елементи). Експериментальні результати на датасеті DeepFashion показали, що запропонований метод досягає точності класифікації 91.3% для основних категорій одягу та 84.7% для детальних атрибутів, таких як довжина рукавів, тип коміра та наявність візерунків. Особливо ефективним виявилось використання механізму уваги для фокусування на критичних областях зображення – точність визначення дрібних деталей зросла на 12.4% порівняно з базовою CNN архітектурою. Проте основним недоліком залишається висока обчислювальна складність при обробці зображень високої роздільної здатності (час обробки одного зображення становить 0.8 секунди на GPU), що ускладнює застосування методу в системах реального часу для онлайн-рішень.

Дослідження [10] представляє комплексну систему класифікації зображень одягу, що поєднує різні методи виділення ознак для покращення точності розпізнавання стилів. Автори пропонують модель IMF (Image Multi-Feature Fusion), яка інтегрує базові візуальні характеристики (колір, форму, текстуру) з глибокими ознаками, отриманими через VGG16. Експериментальні результати показують точність класифікації 85.58% на збалансованому датасеті з 6 тис. зображень, розділених на 8 стильових категорій (діловий, спортивний, елегантний, ретро, мінімалістичний, етнічний, повсякденний та авангардний стилі). Особливо ефективним виявилось використання неглибокої згорткової мережі для вторинного виділення базових ознак перед їх об'єднанням з глибокими характеристиками – це підвищило точність розпізнавання підкатегорій на 12.7% порівняно з використанням лише базових ознак. Модель демонструє стійкість до незбалансованих даних (падіння точності лише на 0.37%), що є важливою перевагою для практичного застосування. Хоча дослідження зосереджене на класифікації стилів, а не категорій одягу (як у нашому випадку), запропонований підхід до багаторівневого об'єднання ознак може бути адаптований для покращення запропонованої у даній статті моделі.

Таким чином, проведені дослідження демонструють поступ у використанні глибоких моделей для класифікації та розпізнавання одягу. Водночас залишаються проблеми точності за умов великої варіативності стилів та оклюзій, а також значні обчислювальні витрати більшості моделей. З урахуванням цього, у даній роботі автори опираються на підхід, запропонований у дослідженні [5], оскільки він поєднує відносну простоту архітектури та стабільність результатів.

Постановка завдання

Метою дослідження є підвищення ефективності ієрархічної класифікації зображень елементів одягу за категоріями, підкатегоріями та атрибутами шляхом розроблення методу послідовного навчання з використанням розмежованих підмоделей на основі архітектури ResNet-50. Рівні категорій та підкатегорій є взаємовиключними, тоді, як на рівні атрибутів продукт може мати більше одного атрибута (див. Рис. 1).

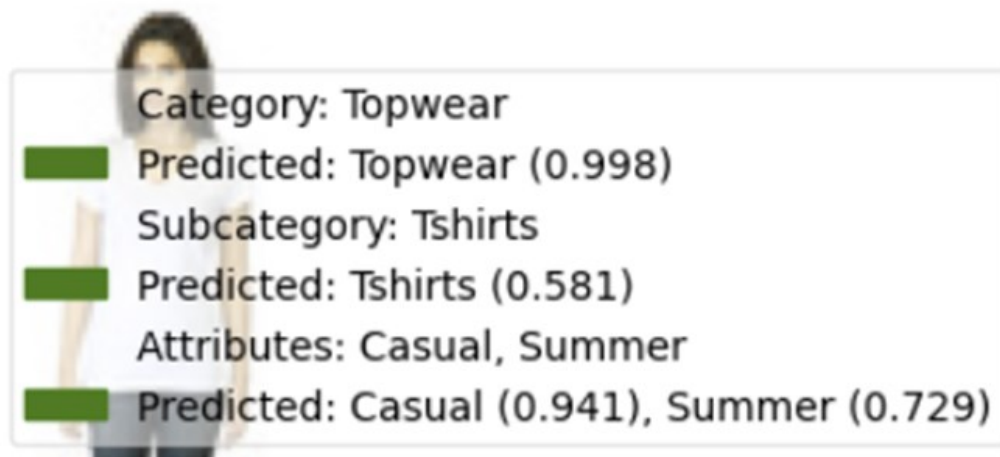


Рис. 1. Приклад класифікації вхідного зображення за категоріями, підкатегоріями та атрибутами

Дане дослідження зосереджується на підвищенні точності в умовах обмеженої кількості даних.

Опис набору даних

Для моделювання використовується набір даних Fashion Product Images (Small) [11], який містить приблизно 44,000 зображень товарів. Набір даних має наступні характеристики:

- Формат зображень: JPG
- Розмір зображень: 60×80 пікселів
- Кількість основних категорій: 7
- Кількість підкатегорій: 45
- Основні категорії: одяг, аксесуари, спортивні товари

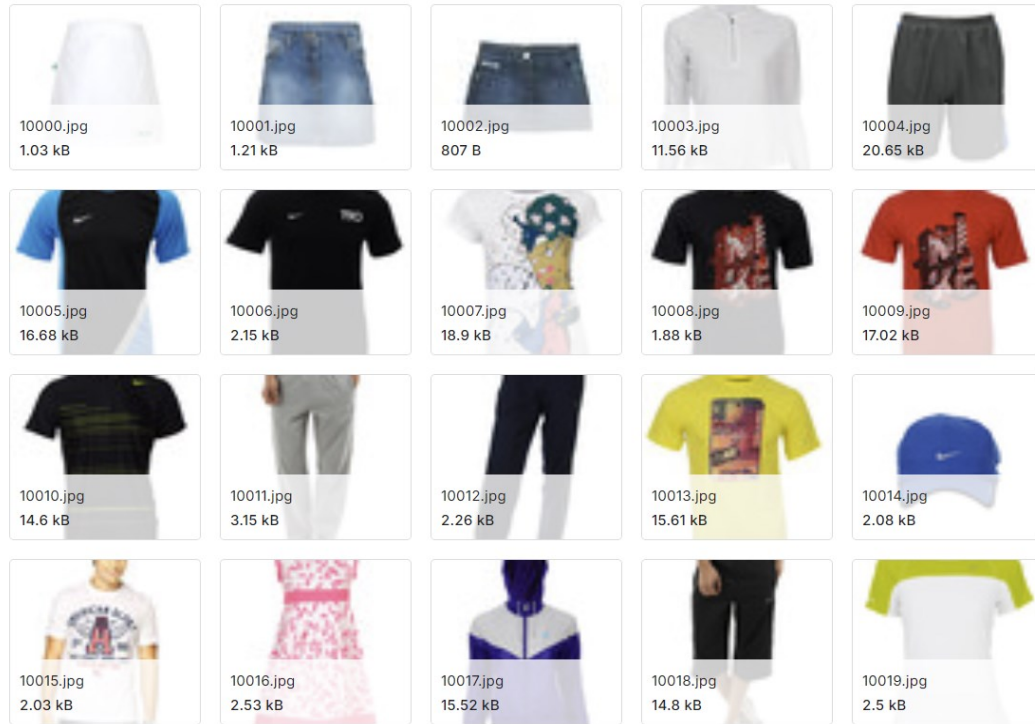


Рис. 2. Приклад зображень з набору даних [11]

Після попереднього аналізу було прийнято рішення зосередитись на двох основних категоріях: Apparel (одяг) та Footwear (взуття), оскільки саме вони становлять основний інтерес для систем електронної комерції в індустрії моди.

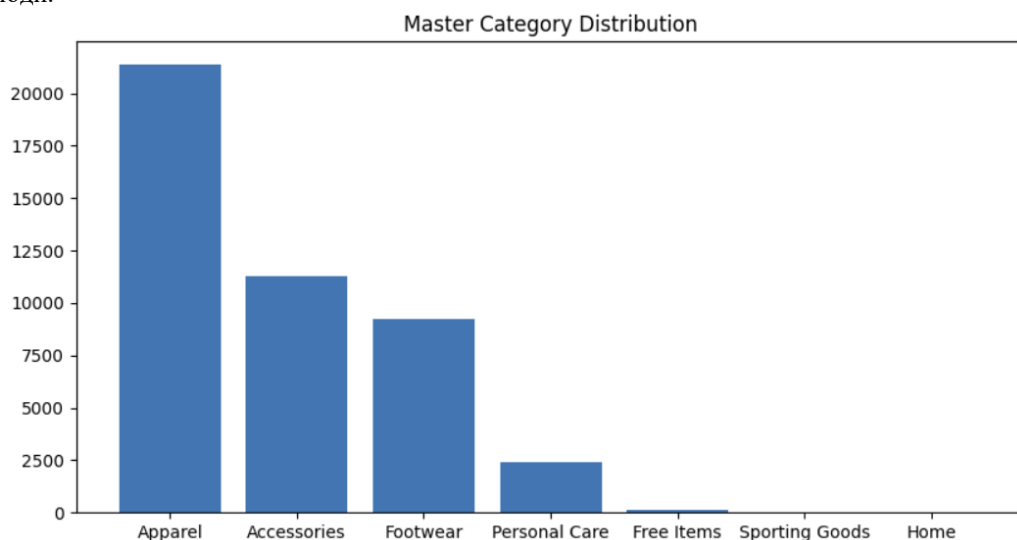


Рис. 3. Розподіл даних по основним категоріям

Попередня обробка даних

Процес підготовки вхідних даних складається з наступних кроків:

1. Фільтрування категорій. З початкового набору даних відібрано лише категорії Apparel та Footwear, що зменшило загальну кількість зображень на 43% (24,996 шт.).
2. Об'єднання метаданих. Зображення об'єднані з відповідними метаданими з файлу styles.csv, включаючи ідентифікатори, категорії, підкатегорії та атрибути (usage та season).

Приклад даних з отриманого набору

ID	Підкатегорія	Тип товару	Сезон	Стиль	Атрибути	Шлях до файлу
19137	Sandal	Sandals	Fall	Casual	[Fall, Casual]	./input/small/images/19137.jpg
20445	Shoes	Casual Shoes	Winter	Casual	[Winter, Casual]	./input/small/images/20445.jpg
22139	Shoes	Formal Shoes	Summer	Formal	[Summer, Formal]	./input/small/images/22139.jpg
39205	Shoes	Heels	Winter	Casual	[Winter, Casual]	./input/small/images/39205.jpg
20897	Shoes	Casual Shoes	Spring	Casual	[Spring, Casual]	./input/small/images/20897.jpg

3. Аугментация. Для збільшення різноманітності навчального набору застосовано наступні трансформації:

- Горизонтальне відображення: випадкове горизонтальне відображення зображень з ймовірністю 50%.
- Зміни яскравості: випадкова зміна яскравості в діапазоні від 80% до 120% від початкового значення з рівномірним розподілом.
- Геометричні трансформації:
 - зміщення: випадкове зміщення по горизонталі та вертикалі до 10% від розміру зображення з рівномірним розподілом;
 - масштабування: випадкове масштабування в діапазоні $\pm 10\%$ з рівномірним розподілом;
 - обертання: випадкове обертання в діапазоні ± 15 градусів з рівномірним розподілом навколо центру зображення.
- Заповнення пікселів: заповнення нових пікселів після трансформацій методом найближчого сусіда.

Розподіл даних. Набір даних було розділено на тренувальний та валідаційний у співвідношенні 85:15.

Архітектура моделі

Базова архітектура (з оригінальної статті [5]) складається з попередньо навченої мережі ResNet-50 з додатковими модифікаціями для багаторівневої класифікації.

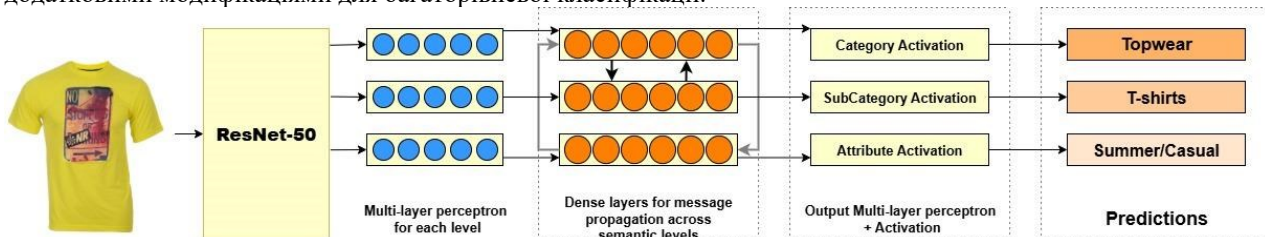


Рис. 4. Архітектура базової моделі

Основні компоненти моделі:

1. Базова мережа: ResNet-50, попередньо навчена на ImageNet.
2. Блок передачі повідомлень: три паралельні щільні шари розмірності 1024, по одному для кожного рівня ієрархії.
3. Регуляризація: L2-норма з коефіцієнтом 0.0005 та dropout з ймовірністю 0.3.
4. Функції активації: ReLU для проміжних шарів, softmax для категорій/підкатегорій, sigmoid для атрибутів.

Загальна кількість параметрів моделі складає 45,585,334.

Розроблена у статті архітектура, розділяє виходи базової моделі на окремі моделі з їх подальшим послідовним тренуванням.

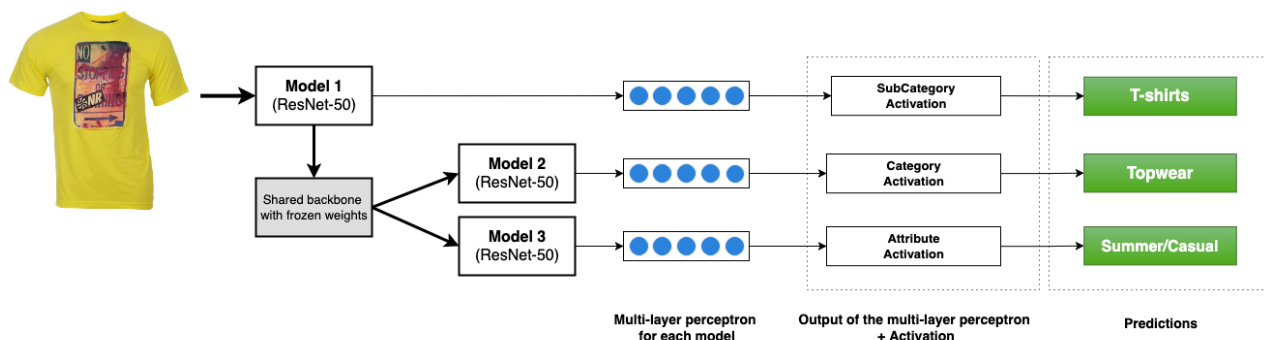


Рис. 5. Архітектура розробленої моделі

Основні компоненти кожної підмоделі:

1. Базова мережа: ResNet-50, попередньо навчена на ImageNet (+ донавчання на підкатегоріях у випадку підмоделей категорій та атрибутів)
2. Послідовне навчання зі спільним екстрактором ознак:
 1. Спочатку модель підкатегорій навчається з тренуваним ResNet-50.
 2. Потім ResNet-50 заморожується і використовується як екстрактор ознак для інших завдань.
3. Архітектура щільних шарів: Три окремі MLP (багатшарові перцептрони) з архітектурою:
 1. Щільний шар 512 нейронів з ReLU активацією.
 2. Dropout 0.3.
 3. Щільний шар 256 нейронів з ReLU активацією.
 4. Dropout 0.3.
 5. Вихідний шар з відповідною активацією (softmax для категорій/підкатегорій, sigmoid для атрибутів).
 6. Регуляризація: L2-норма з коефіцієнтом 0.0005 та dropout з ймовірністю 0.3.
 7. Оптимізатор: Adam з швидкістю навчання 0.001.

Сумарна кількість тренуваних параметрів запропонованих трьох моделей складає: 27'151'576.

Таким чином, усі виходи базової моделі тренуються окремо, для досягнення кращих результатів. При цьому ResNet-50 тренується тільки 1 раз, що залишає час тренування трьох моделей нашого підходу на одному рівні з часом тренування уніфікованої моделі базового підходу.

Ваги для ResNet-50 моделі підкатегорій використовуються як основа для тренування інших моделей, оскільки вона містить найбільшу кількість класів.

Функція втрат

Для навчання моделей класифікації одягу використовується зважена крос-ентропійна функція втрат (weighted cross-entropy loss), яка враховує дисбаланс класів у навчальній вибірці.

Для задачі багатокласової класифікації (категорії та підкатегорії) функція втрат визначається як:

$$L_{wce} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K w_j * y_{ij} * \log(\bar{y}_{ij})$$

де:

- N - кількість зразків у батчі
- K - кількість класів (7 – для категорій, 65 – для підкатегорій)
- w_j - ваговий коефіцієнт для класу j , обернено пропорційний частоті класу:

$$w_j = \frac{N_{total}}{K * N_j}$$

- y_{ij} - one-hot закодована мітка істинного класу (1 – для правильного класу, 0 – для інших)
- $\bar{y}_{ij}$ - передбачена ймовірність належності зразка i до класу j після softmax активації

Для задачі багатоміткової класифікації атрибутів використовується бінарна крос-ентропія з ваговими коефіцієнтами:

$$L_{attr} = \sum_{m=1}^M w_m * BCE_m$$

де:

- M - кількість атрибутів
- w_m - ваговий коефіцієнт для атрибута m
- BCE_m - бінарна крос-ентропія для m -го атрибута

Навчання моделі

Для навчання моделі використовується зважена функція перехресної ентропії, яка враховує дисбаланс класів у наборі даних. Вага кожного класу обчислюється як обернена величина до частоти його появи в навчальному наборі.

Параметри оптимізації:

- оптимізатор: Adam;
- швидкість навчання: 0.001;
- розмір батчу: 32;
- максимальна кількість епох: 50;
- рання зупинка: після 10 епох без покращення.

Обидві архітектури тестувались на двох наборах даних (~25000 зображень та ~2500 зображень). Для оцінювання брались заміри точності на валідаційній частині набору даних, по кожному окремому виходу/підмоделі.

На рисунку 6 зображено метрики (тренувальної та валідаційної точності) процесу навчання розробленої архітектури на повному наборі даних.

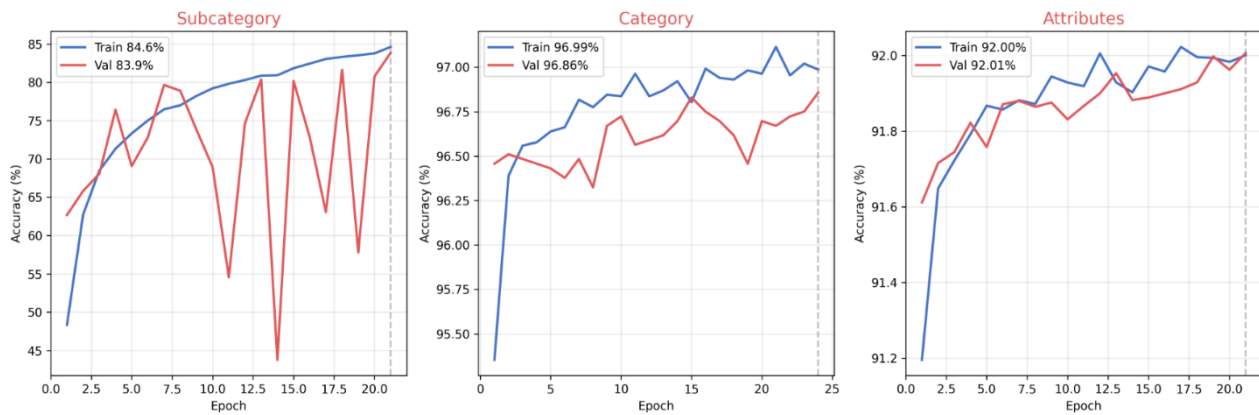


Рис. 6. Візуалізація історії тренування розробленої архітектури на повному наборі даних

Результати

Результати базового та розробленого у цьому дослідженні методів наведено у таблицях 2 та 3:

- Під точністю мається на увазі точність класифікації з використанням валідаційної частини набору даних.
- Епоха відображає епоху, на якій тренування було зупинено, та з якої взято результати ваг та точності.
- Розроблений метод відображає три епохи, оскільки технічно це 3 окремі моделі, що тренувались послідовно. Епохи, на яких моделі досягли збіжності, відображено у форматі: категорія/підкатегорія/атрибути.

Таблиця 2

Результати методів на повному наборі даних (25000 зразків)

Модель	Категорія		Підкатегорія		Атрибути		Епоха
	Точність, %	Зважене середнє F1	Точність, %	Зважене середнє F1	Точність, %	Макро середнє F1	
Базова модель	95,4	0,95	78,8	0,77	<u>98,6</u>	0,37	40
Запропонована модель	<u>96,9</u>	<u>0,97</u>	<u>83,9</u>	<u>0,82</u>	92	<u>0,42</u>	<u>24/21/21</u>

Для набору даних з 25 000 зразків розроблений метод демонструє помірне покращення точності на всіх рівнях класифікації окрім атрибутів. Найбільшим є приріст точності для підкатегорій – з 78,8% до 83,9%, що становить відносне покращення на 5,1%. Для категорій точність зросла з 95,4% до 96,9% (+1,5%). Для атрибутів точність знизилась з 98,6% до 92%.

Таблиця 3

Результати методів на обмеженому наборі даних (2500 зразків)

Модель	Категорія		Підкатегорія		Атрибути		Епоха
	Точність, %	Зважене середнє F1	Точність, %	Зважене середнє F1	Точність, %	Макро середнє F1	
Базова модель	76,3	0,68	32,4	0,18	<u>94,7</u>	0,12	<u>20</u>
Запропонована модель	<u>95,3</u>	<u>0,95</u>	<u>70,2</u>	<u>0,67</u>	91	<u>0,39</u>	25/21/22

При зменшенні обсягу навчальної вибірки до 2 500 зразків точність класифікації категорій зросла з 76,3% до 95,3% (+19%), підкатегорій – з 32,4% до 70,2% (+37,8%). Проте, точність на виході атрибутів знизилась – з 94,7% до 91% (-3,7%).

Розмір моделей

Порівняльний аналіз розміру та параметричності складності моделей демонструє значні переваги розробленої моделі над базовою. Детальні метрики наведено в Таблиці 4.

Таблиця 4

Порівняння розміру моделей та кількості параметрів

Модель	Загальні параметри	Розмір моделі (MB)
Базова модель	45,595,584	173.93
Запропонована модель	<u>27,151,576</u>	<u>94.49</u>

Розроблений метод демонструє значну ефективність у використанні ресурсів порівняно з базовою моделлю. Зменшення загальної кількості параметрів на 40.44% (з 45.6М до 27.2М) досягається завдяки використанню спільного ResNet50 фундаменту для всіх трьох задач класифікації. При цьому кожна спеціалізована модель має лише невелику кількість додаткових тренуваних параметрів: 1,183,243 для категорій, 1,197,121 для підкатегорій та 1,183,500 для атрибутів.

Важливо, що для тренування розробленого методу фактично необхідно оптимізувати 24,770,955 параметрів один раз для повного навчання моделі підкатегорій, яка включає весь ResNet-50. Після фіксування цих навчених ознак, моделі категорій та атрибутів вимагають тренування лише своїх класифікаційних надбудов (близько 1.2М параметрів кожна), що значно прискорює процес навчання та знижує обчислювальні затрати.

Зменшення розміру моделі на 45.65% (з 173.93 MB до 94.49 MB) має практичну цінність для розгортання в обмежених середовищах, включаючи мобільні пристрої, що є критичним для реальних застосувань в електронній комерції.

Швидкість збіжності

Розроблений метод демонструє значно швидшу збіжність моделей на повному наборі даних. Оптимальні результати досягаються через 21-24 епохи (у базовій моделі – через 40 епох). Це свідчить про ефективніше навчання завдяки спеціалізації кожної моделі на своєму рівні ієрархії.

На обмеженому наборі даних швидкість збіжності відносно однакова.

Точність в умовах обмежених даних

Розроблений метод демонструє високу стійкість до зниження точності при роботі з обмеженими обсягами даних. При десятикратному зменшенні навчальної вибірки (з 25 000 до 2 500 зразків) розроблений метод зберігає високу точність (базова модель при цьому показує значне падіння точності).



Рис. 7. Приклади класифікації розробленого методу

Висновки

1. Проведені дослідження демонструє ефективність розробленого методу з розмежованим навчанням окремих моделей для ієрархічної класифікації одягу. Розроблений метод забезпечує суттєве покращення точності класифікації підкатегорій – з 78,8% до 83,9 % на повному наборі даних (+5,1%), що є ключовим фактором для подальших досліджень та практичного застосування в системах електронної комерції. Такий результат отримано завдяки тренуванню спільної частини ResNet-50 з моделлю підкатегорій, на відміну від тренування на усіх типах класифікації одночасно у базовій об'єднаній моделі. Водночас точність визначення категорій не знизилась, а навіть показала незначний ріст.

2. При суттєвому зменшенні навчальної вибірки в 10 разів (з 25 000 до 2 500 зразків), розроблений метод демонструє хороші результати: точність категоризації зростає на 19% (з 76,3% до 95,3%), а точність підкатегорій покращується в 2.5 рази – з 32,4% до 70,2% (+37,8%). Це свідчить про високу ефективність методу в умовах обмежених навчальних даних, що є типовою ситуацією для вузькоспеціалізованих доменів електронної комерції.

3. Розроблений метод з трьома спеціалізованими моделями забезпечує зменшення сумарного розміру моделей приблизно на 45% порівняно з базовою архітектурою. Крім того, використання простішої архітектури після екстрактора ознак ResNet-50 з меншою кількістю нейронів (приблизно на 40%) дозволяє знизити обчислювальну складність без втрати якості класифікації.

4. На повному наборі даних оптимальні результати досягаються за 21-24 епохи замість 40 епох у базовій моделі, що свідчить про більш ефективне навчання завдяки спеціалізації кожної моделі на різних рівнях ієрархії. Час тренування при цьому не збільшується.

5. Отримані результати підтверджують практичну цінність запропонованого підходу для впровадження в реальних системах електронної комерції, особливо в умовах обмежених обчислювальних ресурсів та навчальних даних. Послідовне навчання спеціалізованих моделей відкриває нові можливості для ефективної ієрархічної класифікації в різних сферах застосування. Подальші дослідження будуть спрямовані на генерацію опису продуктів за допомогою LLM, використовуючи отримані у цій статті результати, а також підвищення точності класифікації атрибутів.

Література

1. Lyu X., Li X., Zhang Y., Lu W. Two-Stage Method for Clothing Feature Detection. Big Data Cogn. Comput. 2024. Vol. 8, вип. 4. P. 35. DOI: 10.3390/bdcc8040035.
2. Jun Xiang, Dong T., Pan R., Gao W. Clothing Attribute Recognition Based on RCNN Framework Using L-Softmax Loss. IEEE Access. 2020. Vol. 8. P. 48299–48313. DOI: 10.1109/ACCESS.2020.2979164.

3. Gu M., Hua W., Liu J. Clothing attribute recognition algorithm based on improved YOLOv4-Tiny. *Signal Image Video Process.* 2023. Vol. 17, вип. 7. P. 3555–3563. DOI: 10.1007/s11760-023-02580-5.
4. Li Y., He Z., Wang S., Wang Z., Huang W. Multideep Feature Fusion Algorithm for Clothing Style Recognition. *Wirel. Commun. Mob. Comput.* 2021. Vol. 2021, вип. 1. P. 5577393. DOI: 10.1155/2021/5577393.
5. Ferreira B.Q., Baía L., Faria J., Sousa R.G. A Unified Model with Structured Output for Fashion Images Classification. 2018, June 25. [Електронний ресурс]. arXiv:1806.09445. DOI: 10.48550/arXiv.1806.09445.
6. Shi Q. The Application of Traditional Hakka Clothing Cultural Elements in Modern Clothing Design in Western Fujian Based on Image Feature Extraction Technology. *Appl. Math. Nonlinear Sci.* 2024. Vol. 9, вип. 1. P. 20231216. DOI: 10.2478/amns.2023.2.01216.
7. Chun Y., Wang C., He M. A Novel Clothing Attribute Representation Network-Based Self-Attention Mechanism. *IEEE Access.* 2020. Vol. 8. P. 201762–201769. DOI: 10.1109/ACCESS.2020.3035781.
8. Mathews A., S. N., V. K. R. Analysis of Content Based Image Retrieval using Deep Feature Extraction and Similarity Matching. *Int. J. Adv. Comput. Sci. Appl.* 2022. Vol. 13, вип. 12. DOI: 10.14569/IJACSA.2022.0131277.
9. Mai F., Sun J., Tan Q., Zhao B. A preliminary investigation of niche clothing brand design optimization based on ADDIE model. *Appl. Math. Nonlinear Sci.* 2024. Vol. 9, вип. 1. P. 20230894. DOI: 10.2478/amns.2023.2.00894.
10. Zhang Y., He K., Song R. Image Multi-Feature Fusion for Clothing Style Classification. 2023.
11. Aggarwal Param. Fashion Product Images (Small). Kaggle. [Електронний ресурс]. DOI: 10.34740/KAGGLE/DS/175990.

References

1. Lyu X., Li X., Zhang Y., Lu W. Two-Stage Method for Clothing Feature Detection. *Big Data Cogn. Comput.* 2024. Vol. 8, вып. 4. P. 35. DOI: 10.3390/bdcc8040035.
2. Jun Xiang, Dong T., Pan R., Gao W. Clothing Attribute Recognition Based on RCNN Framework Using L-Softmax Loss. *IEEE Access.* 2020. Vol. 8. P. 48299–48313. DOI: 10.1109/ACCESS.2020.2979164.
3. Gu M., Hua W., Liu J. Clothing attribute recognition algorithm based on improved YOLOv4-Tiny. *Signal Image Video Process.* 2023. Vol. 17, вып. 7. P. 3555–3563. DOI: 10.1007/s11760-023-02580-5.
4. Li Y., He Z., Wang S., Wang Z., Huang W. Multideep Feature Fusion Algorithm for Clothing Style Recognition. *Wirel. Commun. Mob. Comput.* 2021. Vol. 2021, вып. 1. P. 5577393. DOI: 10.1155/2021/5577393.
5. Ferreira B.Q., Baía L., Faria J., Sousa R.G. A Unified Model with Structured Output for Fashion Images Classification. 2018, June 25. [Elektronnyi resurs]. arXiv:1806.09445. DOI: 10.48550/arXiv.1806.09445.
6. Shi Q. The Application of Traditional Hakka Clothing Cultural Elements in Modern Clothing Design in Western Fujian Based on Image Feature Extraction Technology. *Appl. Math. Nonlinear Sci.* 2024. Vol. 9, вып. 1. P. 20231216. DOI: 10.2478/amns.2023.2.01216.
7. Chun Y., Wang C., He M. A Novel Clothing Attribute Representation Network-Based Self-Attention Mechanism. *IEEE Access.* 2020. Vol. 8. P. 201762–201769. DOI: 10.1109/ACCESS.2020.3035781.
8. Mathews A., S. N., V. K. R. Analysis of Content Based Image Retrieval using Deep Feature Extraction and Similarity Matching. *Int. J. Adv. Comput. Sci. Appl.* 2022 Vol. 13, вып. 12. DOI: 10.14569/IJACSA.2022.0131277.
9. Mai F., Sun J., Tan Q., Zhao B. A preliminary investigation of niche clothing brand design optimization based on ADDIE model. *Appl. Math. Nonlinear Sci.* 2024. Vol. 9, вып. 1. P. 20230894. DOI: 10.2478/amns.2023.2.00894.
10. Zhang Y., He K., Song R. Image Multi-Feature Fusion for Clothing Style Classification. 2023.
11. Aggarwal Param. Fashion Product Images (Small). Kaggle. [Elektronnyi resurs]. DOI: 10.34740/KAGGLE/DS/175990.