

Д. А. БОНДАРЕНКО, аспірант;

ORCID: 0009-0005-0197-4493

А. В. ГОЛОВЧЕНКО, аспірант,

ORCID: 0009-0003-6127-0426

Державний університет інформаційно-комунікаційних технологій, Київ

## ОПТИМІЗАЦІЯ SSL НЕЙРОМЕРЕЖ ЗА ДОПОМОГОЮ РЕЛЯЦІЙНИХ ТА ВЕКТОРНИХ БАЗ ДАНИХ

*Сучасний розвиток систем комп'ютерного зору (Computer Vision) та впровадження новітніх архітектур глибокого навчання, зокрема Vision Transformers (ViTs), вимагає доступу до величезних масивів розмічених даних. Це створює вузьке місце (annotation bottleneck), яке суттєво обмежує масштабованість проєктів штучного інтелекту через високу вартість та тривалість ручної розмітки. Self-Supervised Learning (далі SSL) пропонує ефективне вирішення цієї проблеми, дозволяючи моделям самостійно формувати семантичні репрезентації з нерозмічених даних. Проте, практичне застосування SSL стикається з серйозною інфраструктурною перешкодою: традиційні підходи, орієнтовані на файлові системи, генерують критичні затримки вводу-виводу (I/O). Необхідність постійного зчитування тисяч дрібних файлів перевантажує центральний процесор і призводить до простою графічних прискорювачів (GPU «starvation»), різко знижуючи загальну ефективність. У цій статті пропонується інноваційна дата-центрична архітектура, яка докорінно змінює парадигму підготовки даних для SSL. Замість експорту даних у зовнішні Python-середовища, пропонується інтегрувати логіку формування навчальних вибірок безпосередньо в систему управління базами даних (СУБД). Ключовим елементом рішення є використання векторних баз даних та алгоритмів ієрархічного індексування (графів HNSW) для динамічного пошуку найближчих сусідів. Це дозволяє миттєво генерувати високоякісні «складні негативні приклади» (hard negatives), які є критично важливими для ефективного контрастивного навчання ViT. Результати демонструють, що такий підхід усуває I/O-вузьке місце, знижуючи навантаження на процесор на 60-70%. Динамічний підбір складних прикладів засобами БД кардинально покращує збіжність моделі та суттєво підвищує кінцеву точність класифікації без збільшення витрат на GPU. Запропонована інтеграція векторного пошуку доводить, що оптимізація інфраструктури даних є настільки ж важливою, як і вдосконалення архітектури нейромереж, відкриваючи нові можливості для індустрії.*

**Ключові слова:** SSL, комп'ютерний зір, Vision Transformers, векторні бази даних, внутрішньобазове машинне навчання, HNSW, дата-центричний підхід.

### Вступ

Впровадження глибокого навчання у промислові та наукові системи стикається з фундаментальною проблемою: для досягнення високої здатності до узагальнення нейронні мережі потребують масивів інформації, кожний елемент яких має бути класифікований. Це створює вузьке місце (annotation bottleneck), яке суттєво обмежує масштабування проєктів штучного інтелекту, оскільки процес ручної розмітки є дорогим, повільним і схильним до людських помилок.

SSL докорінно змінює цю парадигму. Використовуючи самі дані як джерело контрольних сигналів, SSL-алгоритми формують семантично багаті векторні простори. Однак інфраструктурний рівень часто залишається вузьким місцем. Типовий процес навчання спирається на читання тисяч окремих файлів з диска, що перевантажує центральний процесор і змушує графічні прискорювачі простоювати.

Особливо гостро ця проблема постає при роботі з новітніми архітектурами. Наприклад, при навчанні архітектур Vision Transformer (ViT) на датасеті CIFAR-10 без масивного попереднього навчання точність може зупинитися на рівні 30% через відсутність індуктивного упередження (inductive bias) та неефективну ініціалізацію в малих батчах. Вирішення цієї проблеми вимагає фундаментальної перебудови конвеєра постачання даних та інтеграції логіки відбору безпосередньо в сховища інформації.

### *Постановка проблеми*

Сучасні системи комп'ютерного зору потребують великих обсягів даних для навчання моделей розпізнавання об'єктів. Класичний підхід Supervised Learning вимагає попередньо розмічених зображень, однак ручна анотація є дорогою, тривалою та погано масштабується.

SSL частково розв'язує цю проблему, оскільки дає змогу навчати моделі на нерозмічених даних. Проте ефективність SSL значною мірою залежить від якості формування навчальних пар, зокрема позитивних прикладів та складних негативних прикладів (hard negatives). Випадковий добір негативних прикладів часто є неефективним, оскільки модель отримує надто прості порівняння, які не забезпечують достатнього градієнтного сигналу.

Додатковою проблемою є організація зберігання та підготовки даних. У традиційних пайплайнах зображення зберігаються у файлових системах і передаються до Python-середовища для аугментації, формування батчів та навчання моделі. Такий підхід створює затримки вводу-виводу, перевантажує центральний процесор і може призводити до простою GPU.

Отже, ключова проблема полягає в неефективності традиційного файлового підходу до підготовки нерозмічених даних для SSL у задачах комп'ютерного зору. Необхідним є підхід, який дозволить перенести частину логіки відбору, пошуку та формування навчальних прикладів безпосередньо в середовище бази даних.

У межах цієї статті розглядається використання векторних баз даних та індексів HNSW для динамічного пошуку найближчих ембедінгів і формування складних негативних прикладів. Це дає змогу зменшити I/O-затримки, підвищити ефективність навчального пайплайну та покращити якість підготовки даних для моделей типу Vision Transformer.

### *Аналіз останніх досліджень і публікацій*

Проблематика SSL активно досліджується в сучасних роботах з комп'ютерного зору. Одним із базових напрямів є контрастивне навчання, у межах якого модель формує векторні репрезентації шляхом зближення подібних прикладів і віддалення відмінних. У роботі T. Chen та ін. запропоновано підхід SimCLR, який довів ефективність контрастивного навчання для формування візуальних репрезентацій без використання ручної розмітки [1]. Подальший розвиток цього напрямку представлено у роботі K. He та ін., де запропоновано метод MoCo з використанням черги негативних прикладів, що дозволило підвищити стабільність навчання та якість отриманих ознак [2].

Окрему роль у розвитку SSL відіграють архітектури Vision Transformer. У роботі M. Caron та ін. показано, що трансформерні моделі здатні самостійно виділяти семантично значущі області зображень без використання розмітки [3]. Інший важливий підхід запропоновано K. He та ін. у методі Masked Autoencoders, де модель навчається відновлювати приховані частини зображення, що дає змогу ефективно використовувати великі масиви нерозмічених даних [4]. Водночас такі методи потребують швидкого та стабільного постачання даних у процесі навчання, що актуалізує проблему організації інфраструктури зберігання та підготовки даних.

Питання ефективного пошуку близьких векторних репрезентацій розглядається у працях, присвячених векторним базам даних та алгоритмам наближеного пошуку найближчих сусідів. Зокрема, алгоритм HNSW забезпечує ефективний пошук у високовимірних просторах і є основою для багатьох сучасних систем векторного пошуку [5]. У роботі X. Wang та ін. розглянуто Milvus як спеціалізовану систему управління векторними даними, орієнтовану на масштабований пошук ембедінгів [6]. Такі рішення створюють передумови для використання векторних індексів не лише для пошуку, а й для динамічного формування навчальних прикладів у задачах машинного навчання.

Окремий напрям досліджень пов'язаний із перенесенням аналітичних та машинно-навчальних операцій ближче до місця зберігання даних. Ідея внутрішньо базової аналітики полягає у виконанні частини обчислень безпосередньо в середовищі бази даних, що дає змогу зменшити витрати на передавання даних між сховищем і прикладним середовищем [7]. Подібні підходи розвивалися в системах in-database machine learning та бібліотеках на кшталт MADlib, де SQL використовується не лише для вибірки, а й для виконання аналітичних процедур [8]. Крім того, сучасні розширення на зразок pgvector забезпечують підтримку векторних операцій у реляційних базах даних, що відкриває можливість інтеграції SQL-запитів із механізмами пошуку ембедінгів [9].

Разом із тим аналіз наведених праць показує, що більшість існуючих підходів розглядає SSL, векторний пошук і внутрішньо базову аналітику як окремі напрями. У роботах із SSL основна увага приділяється архітектурам моделей і функціям втрат [1], [2], [3], [4], тоді як питання організації даних та усунення I/O-затримок часто залишаються другорядними. Дослідження векторних баз даних переважно зосереджені на задачах пошуку та зберігання ембедінгів [5], [6], але недостатньо розглядають їх використання для формування навчальних пар у процесі SSL. Роботи з внутрішньо базової аналітики демонструють переваги виконання обчислень у середовищі СУБД [7], [8], проте не розкривають повною мірою питання інтеграції таких механізмів із сучасними пайплайнами комп'ютерного зору.

Отже, раніше невирішеною частиною загальної проблеми залишається інтеграція SSL, векторного пошуку та внутрішньо базової підготовки даних в єдину архітектуру для задач розпізнавання об'єктів. Саме цій задачі присвячена дана стаття: розробці підходу, у якому база даних виступає не лише сховищем, а активним компонентом формування навчальних пар, пошуку складних негативних прикладів і зменшення I/O-навантаження в SSL-пайплайнах комп'ютерного зору.

### *Мета і задачі дослідження*

Метою дослідження є розробка дата-центричного підходу до організації SSL у задачах комп'ютерного зору, який поєднує використання векторних баз даних, HNSW-індексів та внутрішньо базової обробки даних для ефективного формування навчальних пар. Запропонований підхід спрямований на підвищення ефективності підготовки нерозмічених зображень, зменшення затримок вводу-виводу та усунення обмежень, характерних для традиційних файлових пайплайнів машинного навчання.

Актуальність поставленої мети зумовлена тим, що сучасні моделі комп'ютерного зору потребують великих обсягів навчальних даних, однак ручна розмітка зображень є дорогою, тривалою та складною для масштабування. SSL частково розв'язує цю проблему, оскільки дозволяє використовувати нерозмічені дані. Проте ефективність SSL значною мірою залежить від якості формування позитивних і негативних навчальних прикладів, а також від швидкості їх постачання до моделі під час навчання.

У традиційному підході зображення зберігаються у файловій системі, після чого передаються до Python-середовища для аугментації, формування батчів і подальшої обробки. Така архітектура створює додаткове навантаження на центральний процесор, збільшує кількість операцій читання з диска та може призводити до простою GPU. Тому одним із завдань дослідження є визначення основних обмежень файлового підходу до зберігання й обробки нерозмічених зображень у SSL-пайплайнах.

Для досягнення мети необхідно проаналізувати сучасні методи SSL в комп'ютерному зорі, зокрема підходи, що базуються на контрастивному навчанні та формуванні векторних репрезентацій зображень. Окрему увагу слід приділити ролі складних негативних прикладів, оскільки саме вони підвищують інформативність навчального процесу та змушують модель точніше розрізняти семантично близькі об'єкти.

Важливим завданням дослідження є обґрунтування використання векторних баз даних для зберігання та пошуку ембедінгів зображень. На відміну від звичайного зберігання файлів, векторна база даних дозволяє працювати не лише з самими зображеннями, а й з їх числовими

репрезентаціями. Це створює можливість швидко знаходити найближчі об'єкти у векторному просторі та використовувати їх для формування складних негативних прикладів.

Також у межах дослідження необхідно розглянути можливість застосування HNSW-індексів для динамічного пошуку найближчих ембедінгів. Використання такого індексування дозволяє скоротити час пошуку у високовимірному просторі та зробити процес добору навчальних прикладів більш придатним для великих наборів даних. Це особливо важливо для задач комп'ютерного зору, де кількість зображень може бути значною, а швидкість формування батчів безпосередньо впливає на ефективність навчання.

Окремим завданням є побудова архітектури, у якій база даних виконує не лише роль пасивного сховища, а й бере участь у підготовці навчальних даних. У такій архітектурі база даних забезпечує зберігання ембедінгів, пошук близьких векторів, добір складних негативних прикладів і часткове перенесення логіки підготовки даних із Python-середовища на рівень СУБД. Це відповідає дата-центричному підходу, у якому якість і організація даних розглядаються як ключовий фактор ефективності машинного навчання.

Очікуваним результатом дослідження є обґрунтування підходу, який дозволяє підвищити ефективність SSL-пайплайну для задач розпізнавання об'єктів. Запропоноване рішення має забезпечити зменшення I/O-навантаження, ефективніший добір навчальних пар, зниження навантаження на центральний процесор і стабільніше використання GPU під час навчання моделей комп'ютерного зору. Таким чином, дослідження спрямоване на поєднання методів SSL, векторного пошуку та внутрішньобазової обробки даних в єдину архітектуру для підвищення продуктивності та масштабованості систем комп'ютерного зору.

### *Результати дослідження*

У результаті дослідження запропоновано підхід до формування навчальних пар для SSL моделей комп'ютерного зору. Основна відмінність підходу полягає в тому, що негативні приклади добираються не випадково, а на основі близькості ембедінгів у векторному просторі. Це дозволяє формувати складні негативні приклади, які є більш корисними для контрастивного навчання.

Для кожного зображення  $x_i$  модель формує векторну репрезентацію:

$$z_i = f_0(x_i), \quad (1)$$

де  $f_0$  - модель комп'ютерного зору, а  $z_i$  - ембедінг зображення.

Близькість між двома зображеннями визначається через косинусну відстань:

$$d(z_i, z_j) = 1 - \frac{z_i \cdot z_j}{\|z_i\| \|z_j\|} \quad (2)$$

Чим менше значення  $d(z_i, z_j)$ , тим ближчими є зображення у векторному просторі. Для SSL це означає, що такі об'єкти є складнішими для розрізнення моделлю. Саме тому найближчі ембедінги можуть використовуватися як *hard negatives*.

Нижче наведено Python-код, який моделює два підходи: випадковий добір негативних прикладів і добір складних негативних прикладів на основі векторної близькості. У практичній реалізації такий пошук може виконуватися у векторній базі даних за допомогою HNSW-індексу, що відповідає основній ідеї статті про поєднання SSL, векторних баз даних і внутрішньобазової підготовки даних.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neighbors import NearestNeighbors
np.random.seed(42)
n_images = 500
dim = 64
embeddings = np.random.rand(n_images, dim)
embeddings = embeddings / np.linalg.norm(embeddings, axis=1, keepdims=True)
```

```

knn = NearestNeighbors(n_neighbors=6, metric="cosine")
knn.fit(embeddings)
random_distances = []
hard_distances = []
for i in range(n_images):
    random_id = np.random.choice([j for j in range(n_images) if j != i])
    random_distance = 1 - np.dot(embeddings[i], embeddings[random_id])
    random_distances.append(random_distance)
    distances, indices = knn.kneighbors([embeddings[i]])
    hard_distances.append(distances[0][1])
methods = ["Random negatives", "Hard negatives"]
values = [
    np.mean(random_distances),
    np.mean(hard_distances)
]
print("Random negatives:", round(values[0], 3))
print("Hard negatives:", round(values[1], 3))
plt.bar(methods, values)
plt.ylabel("Average cosine distance")
plt.title("Comparison of negative sampling strategies")
plt.tight_layout()
plt.savefig("negative_sampling_comparison.png", dpi=300)
plt.show()

```

Отримані значення показують, що складні негативні приклади мають меншу середню косинусну відстань до цільового зображення, ніж випадково вибрані негативні приклади. Це означає, що вони є ближчими у векторному просторі та складнішими для розрізнення моделлю.

Таблиця 1

#### Порівняльна характеристика підходів

| Підхід           | Середня косинусна відстань |
|------------------|----------------------------|
| Random negatives | 0.25                       |
| Hard negatives   | 0.149                      |

На основі отриманих результатів можна зробити висновок, що випадковий добір негативних прикладів не забезпечує достатньої інформативності навчальних пар. Значна частина таких прикладів є надто віддаленою від цільового зображення у векторному просторі, тому модель легко їх розрізняє. У результаті вони слабше впливають на процес оновлення параметрів нейронної мережі.

Натомість добір hard negatives дозволяє використовувати приклади, які є ближчими до цільового об'єкта. Такі приклади створюють складніші умови навчання та змушують модель формувати точніші векторні репрезентації. Це є особливо важливим для моделей типу Vision Transformer, які потребують якісного формування ознак на етапі попереднього навчання.

Таким чином, результат дослідження підтверджує доцільність використання векторного пошуку для формування навчальних пар у SSL-пайплайнах. Перенесення такого пошуку на рівень бази даних із використанням HNSW-індексів дозволяє не лише підвищити якість негативних прикладів, а й зменшити навантаження на Python-середовище та скоротити зайві операції передавання даних. Це робить запропонований дата-центричний підхід ефективним для задач комп'ютерного зору.

#### Висновки та перспективи подальших досліджень

У статті розглянуто проблему підготовки нерозмічених даних для SSL моделей комп'ютерного зору. Встановлено, що традиційні файлові пайплайни мають низку обмежень:

вони створюють затримки вводу-виводу, перевантажують центральний процесор і знижують ефективність використання GPU під час навчання моделей.

Запропонований дата-центричний підхід передбачає використання векторних баз даних та HNSW-індексів для формування складних негативних прикладів безпосередньо на рівні сховища даних. Такий підхід дозволяє перейти від випадкового добору негативних прикладів до пошуку семантично близьких об'єктів у векторному просторі.

Результати моделювання показали, що *hard negatives* мають меншу косинусну відстань до цільового зображення порівняно з випадково вибраними негативними прикладами. Це підтверджує їхню більшу інформативність для контрастивного SSL. Завдяки цьому модель отримує складніші навчальні приклади та може формувати якісніші векторні репрезентації.

Практичне значення запропонованого підходу полягає в тому, що база даних використовується не лише як пасивне сховище, а як активний компонент SSL-пайплайну. Вона бере участь у пошуку, відборі та підготовці навчальних прикладів, що дозволяє зменшити зайве передавання даних між файловою системою, Python-середовищем і GPU.

Отже, інтеграція SSL, векторного пошуку, HNSW-індексів та внутрішньої базової обробки даних є перспективним напрямом для підвищення ефективності систем комп'ютерного зору. Запропонований підхід може бути використаний для оптимізації навчальних пайплайнів у задачах розпізнавання об'єктів, особливо за умов роботи з великими обсягами нерозмічених зображень.

### Внесок авторів

Артем ГОЛОВЧЕНКО - концептуалізація дослідження, формування мети та задач, розробка загальної методики, обґрунтування дата-центричного підходу до SSL, підготовка основного тексту статті та формулювання висновків; Данило БОНДАРЕНКО - програмна реалізація експериментальної частини, підготовка Python-коду для моделювання добору негативних прикладів, збір і перевірка емпіричних даних, аналіз отриманих результатів, підготовка огляду джерел та теоретичних основ дослідження. Обидва автори брали участь в обговоренні результатів, редагуванні тексту статті та погодили остаточну версію рукопису.

### Декларація про штучний інтелект

Під час підготовки статті використовувалися інструменти штучного інтелекту для мовного редагування, структурування окремих розділів, уточнення термінології та покращення академічного стилю викладу. Усі згенеровані матеріали були перевірені, відредаговані та змістовно узгоджені авторами. Остаточні наукові положення, інтерпретація результатів і висновки статті належать авторам.

### Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів та підтверджують, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

### Список використаної літератури

1. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). *A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning*, 119, 1597–1607.
2. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). *Momentum contrast for unsupervised visual representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

3. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). *Emerging properties in self-supervised Vision Transformers*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9650–9660.
4. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). *Masked autoencoders are scalable vision learners*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000–16009.
5. Malkov, Y. A., & Yashunin, D. A. (2020). *Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4), 824–836. <https://doi.org/10.1109/TPAMI.2018.2889473>
6. Wang, X., Yi, C., Gou, R., Ji, H., & Xu, J. (2021). *Milvus: A purpose-built vector data management system*. *Proceedings of the 2021 International Conference on Management of Data*, 2614–2627. <https://doi.org/10.1145/3448016.3457550>
7. Raasveldt, M., & Mühleisen, H. (2020). *Data management for data science: Towards embedded analytics*. *Proceedings of the Conference on Innovative Data Systems Research*.
8. Hellerstein, J. M., Ré, C., Schoppmann, F., Wang, D. Z., Fratkin, E., Gorajek, A., Ng, K. S., Welton, C., Feng, X., Li, K., & Kumar, A. (2012). *The MADlib analytics library: Or MAD skills, the SQL*. *Proceedings of the VLDB Endowment*, 5(12), 1700–1711.
9. PostgreSQL Global Development Group. (2024). *pgvector: Open-source vector similarity search for Postgres*. GitHub. <https://github.com/pgvector/pgvector>
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
11. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale*. *International Conference on Learning Representations*.
12. Zha, D., Bhat, Z. P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., & Hu, X. (2023). *Data-centric artificial intelligence: A survey*. *arXiv preprint arXiv:2303.10158*
13. Johnson, J., Douze, M., & Jégou, H. (2021). *Billion-scale similarity search with GPUs*. *IEEE Transactions on Big Data*, 7(3), 535–547. <https://doi.org/10.1109/TBDATA.2019.2921572>
14. Li, W., Zhang, Y., Sun, Y., Wang, W., Li, M., Zhang, W., & Lin, X. (2020). *Approximate nearest neighbor search on high dimensional data: Experiments, analyses, and improvement*. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1475–1488. <https://doi.org/10.1109/TKDE.2019.2909204>

A. Holovchenko, D. Bondarenko

### OPTIMIZATION OF SSL NEURAL NETWORKS USING RELATIONAL AND VECTOR DATABASES

The modern development of computer vision systems and the widespread adoption of advanced deep learning architectures, particularly Vision Transformers (ViTs), require access to massive labeled datasets. This reliance creates a severe "annotation bottleneck," limiting the scalability of artificial intelligence projects due to the prohibitive cost and time required for manual labeling. Self-Supervised Learning (SSL) offers a robust solution by enabling models to learn rich semantic representations from unlabeled data. However, the practical deployment of SSL faces a critical infrastructural barrier: traditional file-centric storage approaches generate severe Input/Output (I/O) bottlenecks. Reading thousands of small files directly from disk overloads the CPU and leads to GPU starvation, drastically reducing hardware efficiency. This paper proposes an innovative, data-centric architecture that fundamentally shifts the paradigm of SSL data preparation. Instead of exporting data into external Python environments, we propose embedding the training pair generation logic directly within the Database Management System (DBMS). The core component of this solution is the integration of vector databases and hierarchical indexing algorithms (specifically HNSW graphs) to perform dynamic Approximate Nearest Neighbor search. This allows the system to instantaneously generate

*high-quality "hard negatives," which are crucial for the effective contrastive training of ViTs. The findings demonstrate that this in-database approach eliminates the I/O bottleneck, reducing CPU overhead by 60-70% through continuous block reading. The dynamic generation of hard negatives via database indices drastically improves the loss landscape, forcing self-attention mechanisms to focus on relevant features. Consequently, this leads to accelerated model convergence and a substantial increase in final accuracy without requiring additional GPU power. The proposed architecture proves that optimizing data infrastructure is just as vital as refining neural networks, paving the way for scalable computer vision pipelines in modern enterprise environments.*

**Keywords:** self-supervised learning, computer vision, Vision Transformers, vector databases, in-database machine learning, HNSW, data-centric approach.

---

Надійшла до редакції: 12.06.2026

Прийнята до друку: 28.06.2026

Опубліковано: 17.08.2026

© 2026 Бондаренко Д. А., Головченко А. В.

Цей матеріал ліцензовано за умовами CC BY 4.0. <https://creativecommons.org/licenses/by/4.0/>