

Звіт про виконання лабораторної роботи №3
з курсу “Advanced Computer Vision 2026”

Тема: Створення пошукової системи по відео
на основі моделі CLIP та кластеризація
контенту

Виконав: студент групи ММШІ-2
Коломієць Микола

1 Мета роботи

Створення прототипу відео-пошукової системи, яка здатна приймати текстовий запит і повертати найбільш релевантні відеоматеріали. Додатковою метою є проведення автоматичної кластеризації відео-даних для групування контенту за тематиками.

2 Використані технології та методи

- **Модель:** CLIP (ViT-B/32) від OpenAI для отримання векторних представлень (embeddings) тексту та зображень.
- **Обробка відео:** Фреймворк OpenCV для вибірки кадрів (sampling) з інтервалом в 1 секунду.
- **Агрегація:** Метод Mean Pooling для обчислення фінального дескриптора відео на основі ембедінгів окремих кадрів.
- **Кластеризація:** Алгоритми K-Means та DBSCAN (метрика - cosine similarity).

3 Опис експериментів та результатів

3.1 Пошук за текстовим запитом

Для тестування було використано датасет, що складається з відео категорій “Sports” та “Gaming”.

- **Результат:** Система успішно ідентифікує контент. При запиті, що стосувався ігрової тематики, модель точно виділила фрагменти гри **Assassin’s Creed** серед загального масиву даних.
- **Ранжування:** Косинусна близькість дозволила побудувати коректний топ результатів, де спортивний та ігровий контент чітко розмежовується.

3.2 Кластеризація (K-Means)

При використанні K-Means з параметром $k = 2$ (або більше), алгоритм стабільно розділив дані на дві великі групи:

1. Спортивні змагання (футбол, атлетика тощо).
2. Геймплей відеоігор.

3.3 Кластеризація (DBSCAN)

Алгоритм DBSCAN показав більш деталізовану структуру:

- Було виявлено значну кількість “шуму” (outliers), що відповідає відео з унікальними візуальними характеристиками.
- Алгоритм не лише розділив основні категорії, а й самостійно виділив специфічну підгрупу — **відео з Minecraft**, завдяки характерному кубічному стилю графіки, який утворює щільний кластер у векторному просторі CLIP.

3.4 Візуалізація простору ознак (PCA)

Для аналізу розподілу векторних представлень відео було застосовано метод головних компонент (PCA), що дозволило знизити розмірність ембедінгів з 512 до 2 компонент для побудови двовимірного графіка.

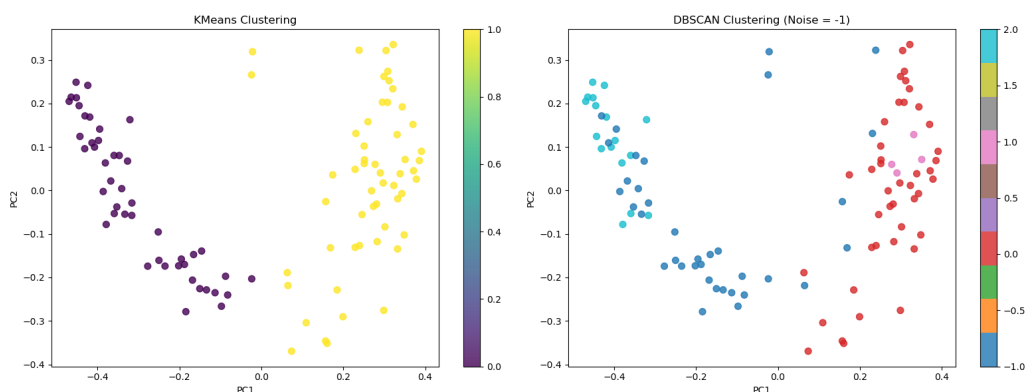


Figure 1: Візуалізація кластеризації відео за допомогою методу PCA. Кольорами позначено результати роботи алгоритмів K-Means та DBSCAN.

На графіку (рис. ??) чітко спостерігається сепарація даних. Спортивний та ігровий контент утворюють окремі області у просторі [cite: 87]. Зокрема, результати DBSCAN підтверджують наявність щільного ядра для специфічних ігор (наприклад, Minecraft) та розсіяних точок “шуму”, які відповідають унікальним або менш типовим відеоматеріалам [cite: 88, 89].

4 Висновки

Розроблена система демонструє високу ефективність у задачах Zero-shot пошуку. Використання Mean Pooling для кадрів відео виявилось достатнім для точного тематичного пошуку та кластеризації. Модель CLIP успішно справляється з розпізнаванням складних візуальних патернів (наприклад, стилістика Assassin's Creed або Minecraft) без додаткового донавчання.