ПРИМЕНЕНИЕ МОДЕЛИ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ПРОИЗВЕДЕНИЙ КОЛЛЕКЦИИ ЖИВОПИСИ

Выполнила: Дегтярь Анастасия Сергеевна

Научный руководитель: Сельский Виктор

Год: 2026

Актуальность и проблема

АКТУАЛЬНОСТЬ работы связана с тем, что современные музейные приложения и системы распознавания изображений не всегда адаптированы под задачи идентификации конкретных произведений искусства.

Стандартные методы требуют больших размеченных наборов данных. Для музейных коллекций это неприменимо, так как каждое произведение уникально и представлено единственным снимком.

Цель и задачи исследования

ЦЕЛЬ: исследовать возможность применения модели CLIP в качестве экстрактора признаков для идентификации произведений коллекции живописи и оценить качество распознавания.

ЗАДАЧИ:

1. Рассмотреть теоретические основы работы модели CLIP.
2. Подготовить датасет изображений произведений живописи.
3. Реализовать систему распознавания на основе сравнения эмбеддингов.
4. Провести экспериментальное тестирование и проанализировать результаты.

Что такое Clip?

Модель CLIP разработана компанией OpenAI в 2021 году. Она состоит из двух частей:

Визуальный энкодер — превращает изображение в эмбеддинг (числовой вектор)

Текстовый энкодер — превращает текст в эмбеддинг

В данной работе CLIP используется **не для сравнения с текстом**, а как **экстрактор признаков**. Для сравнения изображений между собой через косинусное сходство их эмбеддингов.

Датасет и подготовка данных

В работе использовался датасет на основе коллекции Государственного Эрмитажа. Исходный датасет был взят с сайта kaggle.com. Для тестовой выборки были взяты реальные фото пользователей из музея.

Исходный датасет: 109 картин
Тестовая выборка: 21 реальное фото

Методология распознавания

Реальное фото

↓

Визуальный энкодер CLIP

↓

Эмбеддинг запроса (512 чисел)

↓

Косинусное сходство со ВСЕМИ эталонами

↓

Выбор эталона с максимальным сходством

↓

Результат: название картины + уверенность

Результаты распознавания

Точность на реальных музейных фотографиях: 71% (15 из 21)

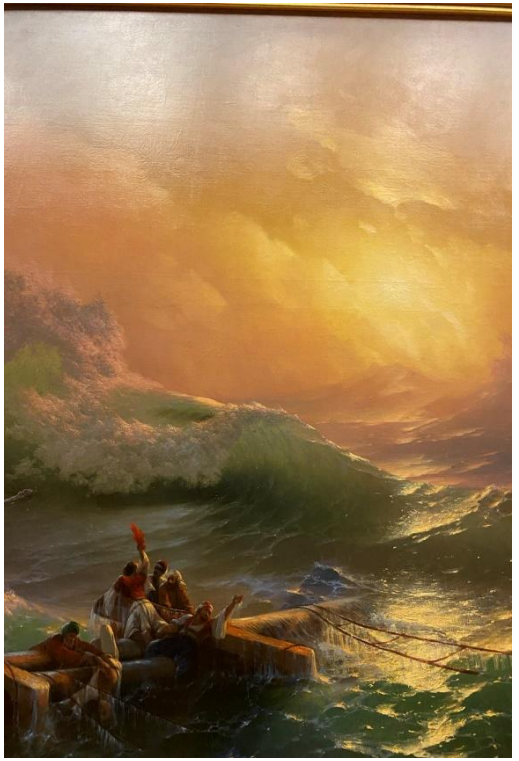
Уверенность модели:

Средняя: ~ 80%
При правильных ответах: 82.4%

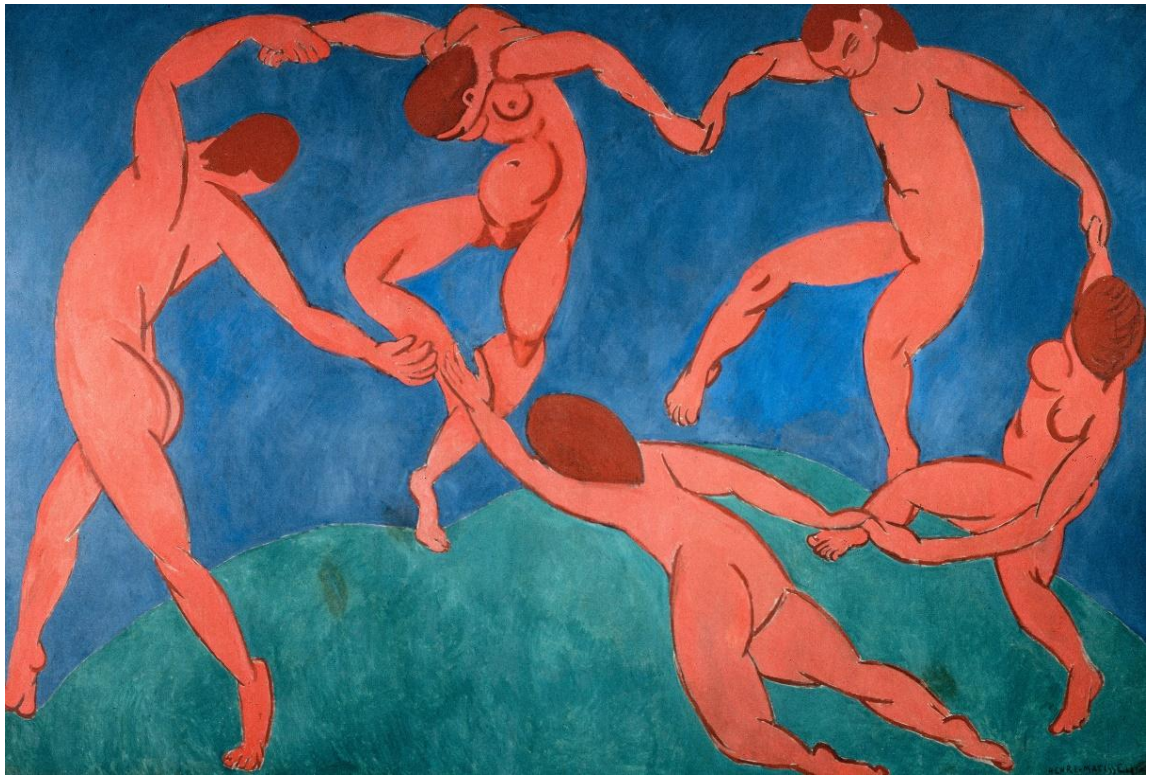
Примеры правильных ответов

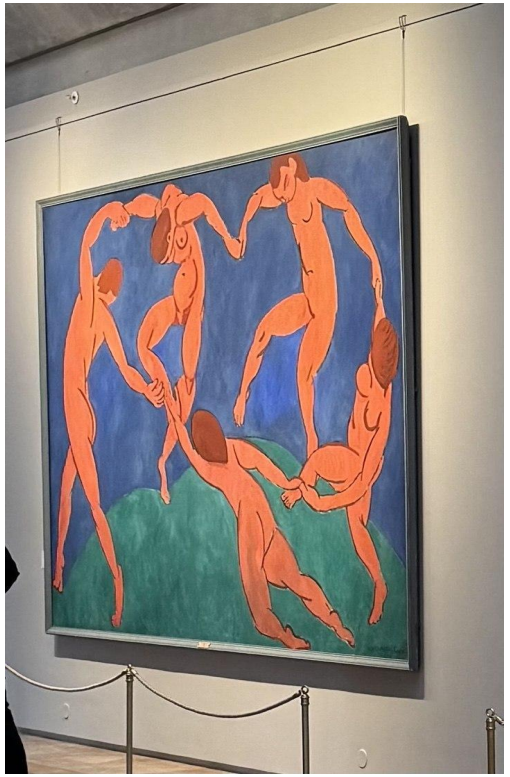
Картина	Уверенность
Иван Айвазовский «Девятый вал»	91.99%
Анри Матисс «Танец»	84.26%
Клод Моне «Поле маков»	79.34%
Винсент Ван Гог «Хижины»	88.01%

Иван Айвазовский «Девятый вал»

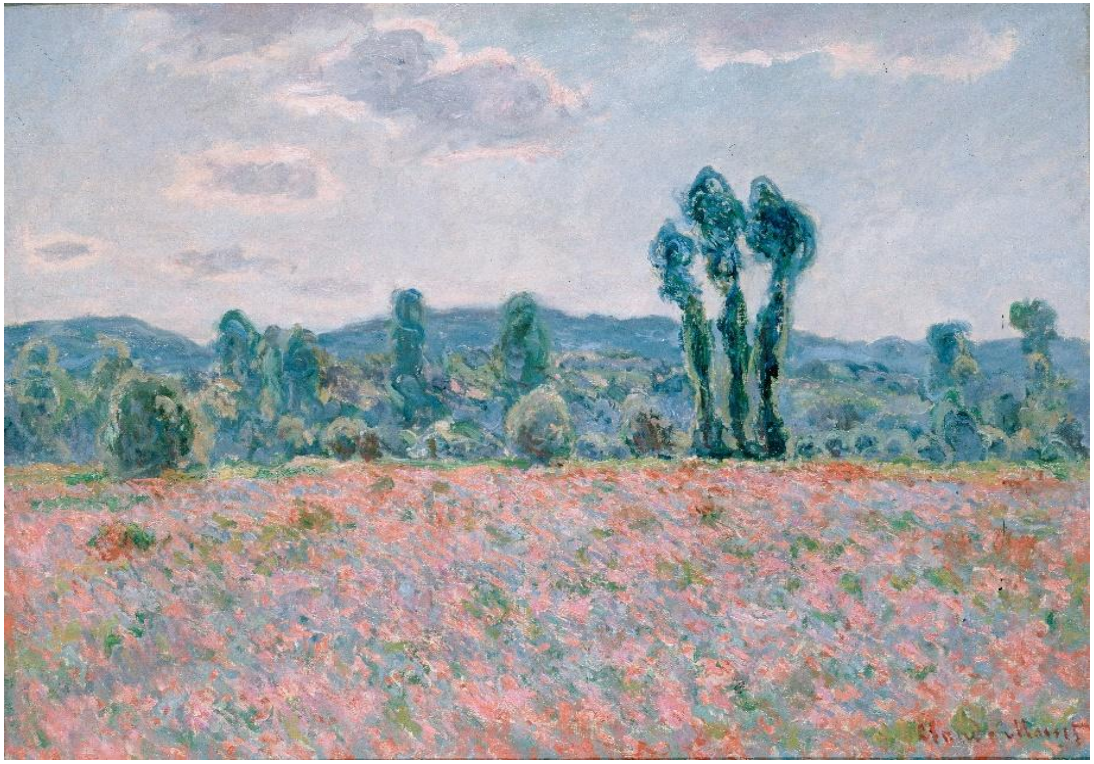


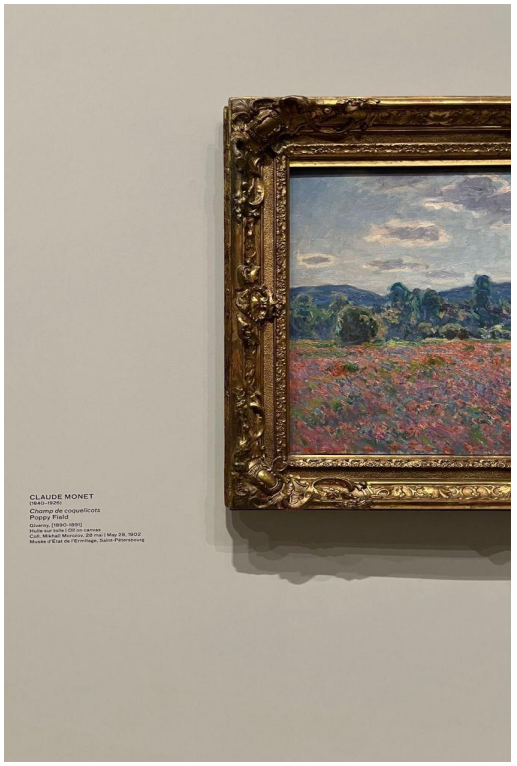
Анри Матисс «Танец»





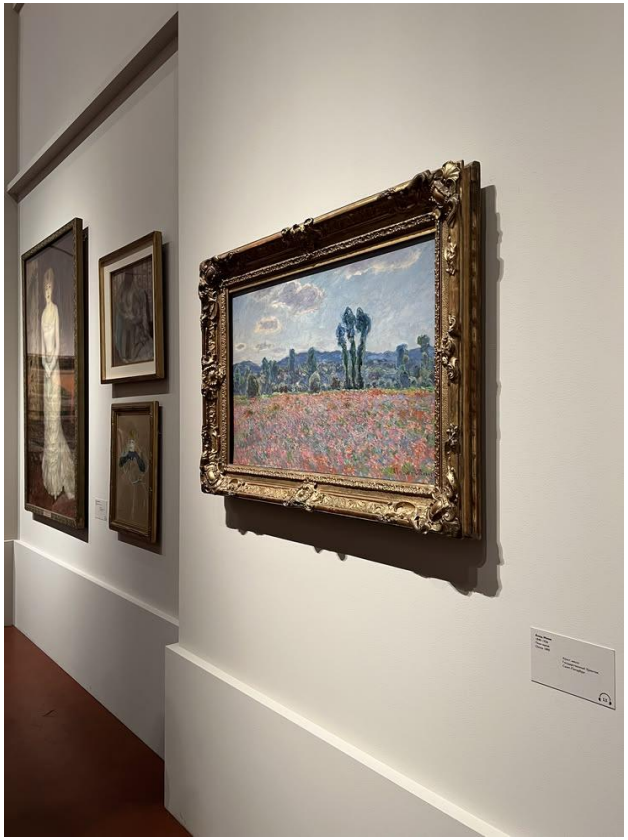
Клод Моне «Поле маков»





Винсент Ван Гог «Хижины»

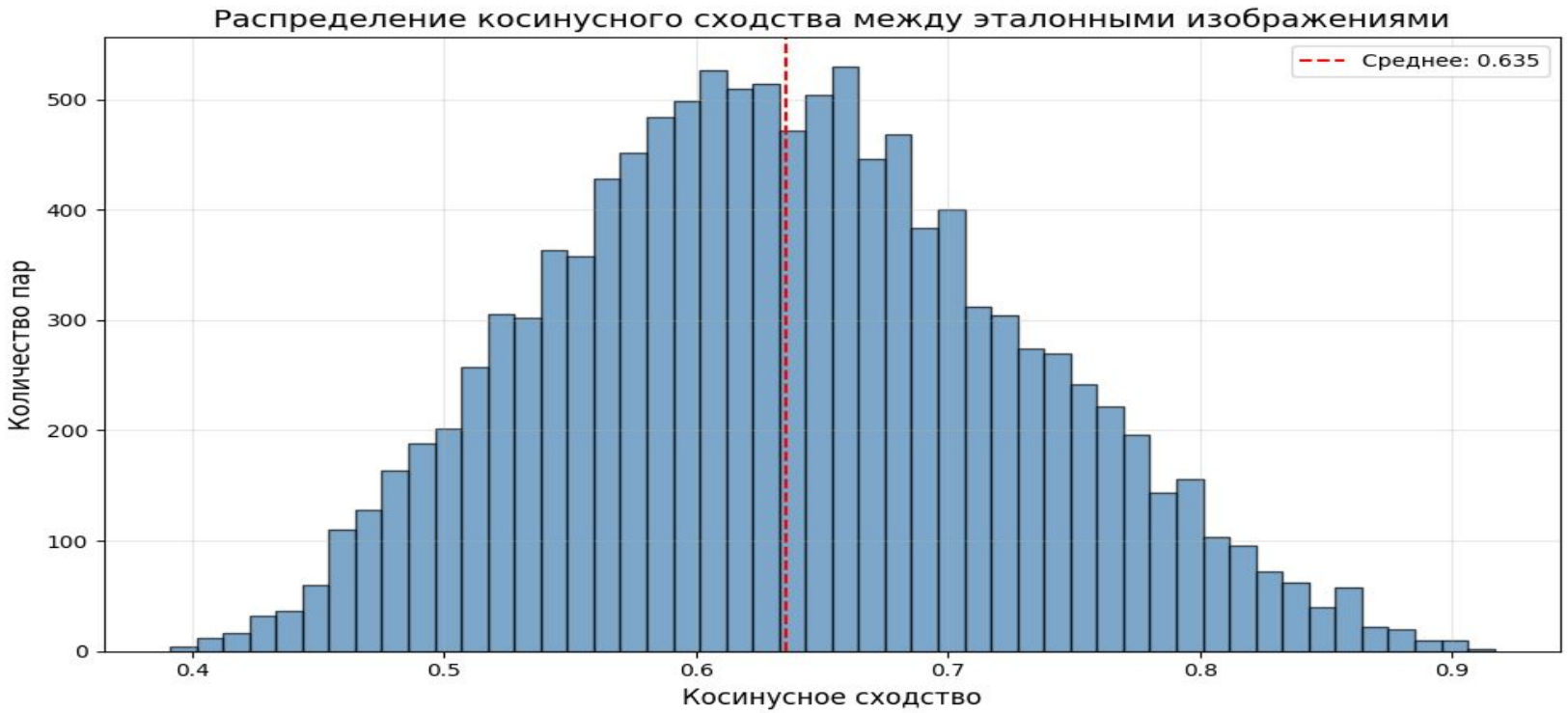
Пример ошибки распознавания



Анализ косинусного сходства

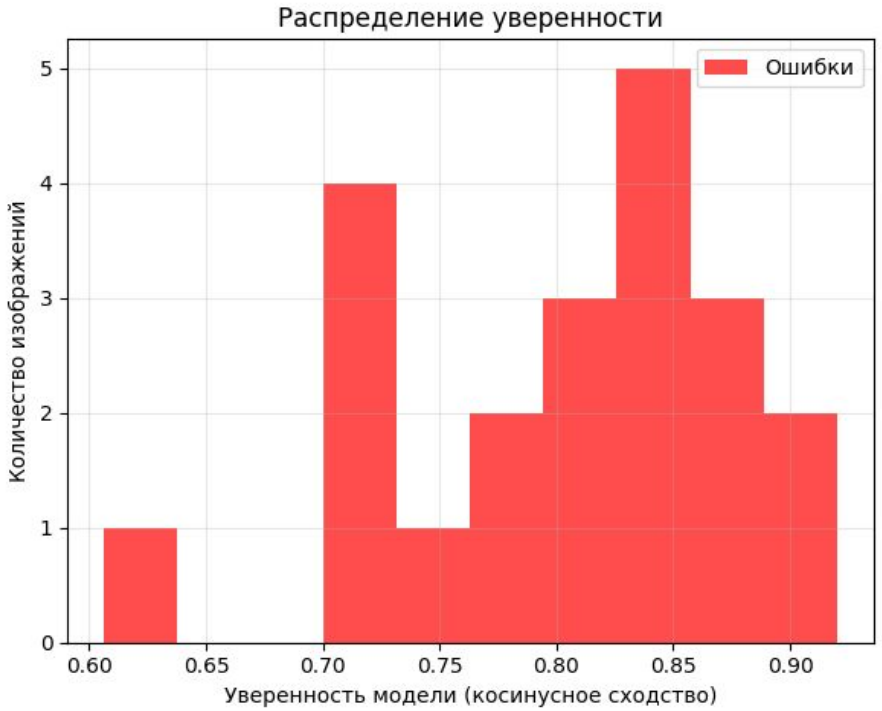
Среднее сходство: **0.635**
Значения выше 0.9 почти отсутствуют
Каждая картина занимает свою уникальную область в пространстве эмбеддингов

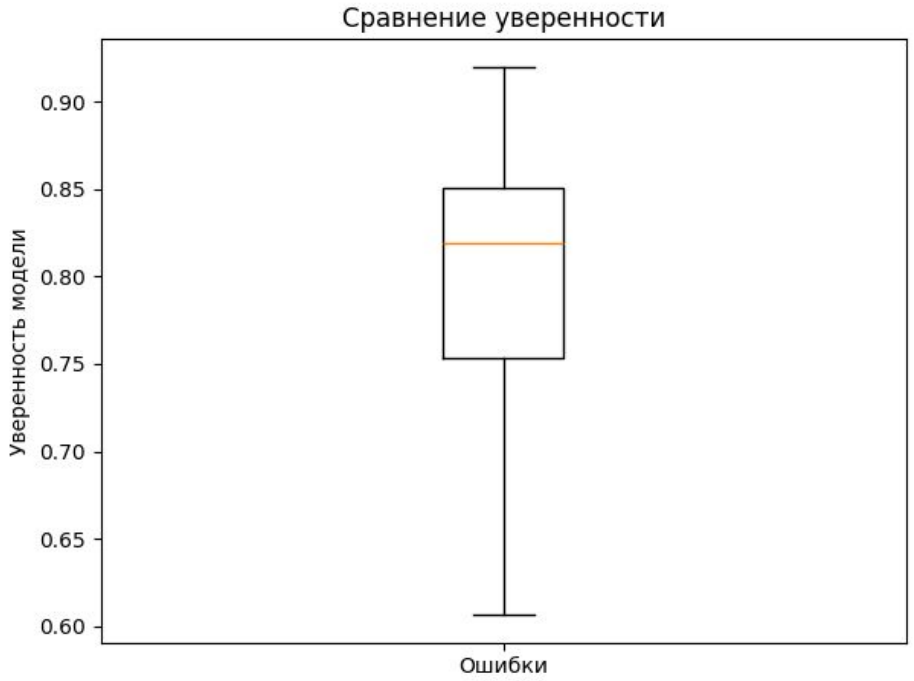
Вывод: CLIP хорошо разделяет разные произведения



Анализ уверенности

Средняя уверенность: **79.8%**
Медиана: **80.1%**
Минимум: 60.6% (Шишкин)
Максимум: 92.0% (Айвазовский)





Выводы

Модель CLIP эффективна как экстрактор признаков для распознавания живописи

Точность на реальных музейных фотографиях: **71%**
Уверенность модели: **60–92%**

Основное ограничение: качество распознавания зависит от полноты эталонной базы

При расширении базы точность можно повысить до 85–90%

Практическая значимость

Где можно применить:

Музейные мобильные приложения
Системы каталогизации музейных коллекций
Образовательные проекты

Преимущества:

Не требует дообучения
Работает с единственным изображением каждого класса
Высокая уверенность при правильных ответах

СПАСИБО ЗА ВНИМАНИЕ!