

ВЛИЯНИЕ ДОМЕННОВОГО СМЕЩЕНИЯ НА ОБОБЩАЮЩУЮ СПОСОБНОСТЬ МОДЕЛЕЙ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

С.И. Саблин¹

¹Самарский университет, г. Самара
e-mail: serjjiniuss@gmail.com

Аннотация

Рассматривается влияние доменного смещения на обобщающую способность моделей классификации аватаров (реальные, рисованные, сгенерированные). Используются датасеты с платформы Kaggle, объединяющие изображения различных типов и источников. Сравниваются архитектуры ResNet-50, MobileNetV3, EfficientNet-B0 и ConvNeXt-Tiny. Показано, что при независимом тестировании наблюдается снижение макро-F1, что указывает на чувствительность моделей к изменению домена. Рассматривается влияние отсутствия детекции лиц.

Ключевые слова: искусственный интеллект, классификация изображений, доменное смещение, нейронные сети, дипфейки.

Введение

В последние годы наблюдается рост количества синтетических изображений и дипфейков, создаваемых с использованием генеративных моделей искусственного интеллекта. Современные методы позволяют формировать реалистичные изображения людей, что существенно усложняет задачу проверки визуального контента [1].

Рост доступности генеративных моделей приводит к увеличению числа злоупотреблений, включая мошенничество и распространение дезинформации, а также снижает эффективность методов детекции синтетических изображений [2].

Это определяет необходимость разработки автоматизированных систем классификации изображений по типу происхождения. Особенно актуальна данная задача в контексте цифровой идентификации, где изображения профиля используются как элемент доверия.

Современные методы анализа изображений активно используют нейронные сети и применяются в задачах кибербезопасности [3]. Однако их применение осложняется эффектом доменного смещения, возникающего при различии между обучающими и тестовыми данными.

В работе рассматривается трёхклассовая задача классификации изображений (real, drawing, generated) с акцентом на влияние доменного смещения на обобщающую способность моделей.

1. Цель исследования

Цель исследования – анализ влияния доменного смещения на обобщающую способность моделей классификации изображений аватаров по типу происхождения.

Задачи исследования:

- 1) формирование обучающей выборки на основе датасетов с платформы Kaggle;
- 2) обучение моделей, основанных на выбранных архитектурах;
- 3) проведение экспериментов по оценке устойчивости моделей при переносе на независимые выборки;
- 4) формирование выводов по результатам эксперимента.

2. Данные и постановка задачи

Пусть задано множество изображений $X = \{x_1, x_2, \dots, x_n\}$, принадлежащих классам $Y = \{\text{real}, \text{drawing}, \text{generated}\}$. Требуется построить модель $f(x)$, предсказывающую класс изображения.

Обучающая выборка сформирована из нескольких открытых датасетов с платформы Kaggle [4]:

- реальные изображения: Real vs Fake Faces, Deepfake and Real Images, CelebFaces Dataset;
- сгенерированные изображения: Detect AI-Generated Faces: High-Quality Dataset, Fake-Vs-Real-Faces (Hard), Person Face Dataset;
- рисованные изображения: Anime Face Dataset, LEGO Minifigure Faces, CUNHK Face Sketch Database.

Использование нескольких источников позволило сформировать неоднородный набор данных, моделирующий реальные условия доменного смещения.

Распределение классов оказалось несбалансированным: доля изображений класса drawing превышает 80%, что дополнительно влияет на поведение моделей.



Рисунок 1 – Примеры изображений классов real, drawing и generated

Классы различаются по стилю и текстуре, при этом сгенерированные изображения часто близки к реальным.

3. Модели и обучение

В работе использованы архитектуры ResNet-50, MobileNetV3, EfficientNet-B0 и ConvNeXt-Tiny.

Обучение проводилось с разделением данных 80/10/10. Использовались:

- оптимизатор AdamW;
- взвешенная функция потерь;
- scheduler ReduceLROnPlateau.

Изображения приводились к размеру 224×224.

Предварительная детекция и выравнивание лиц не выполнялись, поэтому модели обучались на полноразмерных аватарах вместе с фоном и стилистическими особенностями изображения. Это усиливало зависимость от домена и снижало устойчивость при переносе на новые данные.

Таким образом, отсутствие предварительного выделения лица рассматривается как один из факторов, усиливающих доменное смещение и влияющих на качество классификации.

4. Эксперимент и результаты

Для оценки качества использовались метрики Macro F1 и Balanced Accuracy.

На обучающем наборе модели демонстрировали высокие значения метрик, однако при тестировании на независимых данных наблюдалось резкое снижение качества.

Таблица 1 – Результаты на независимом тесте

Показатель	ResNet-50	MobileNetV3
Macro F1	0,255	0,220
Balanced Accuracy	0,230	0,235
F1 (AI)	0,000	0,000
F1 (Drawing)	0,032	0,014
F1 (Real)	0,732	0,647

Полученные результаты показывают, что модели сохраняют способность распознавать реальные изображения, но практически не различают сгенерированные и рисованные классы при смене домена.

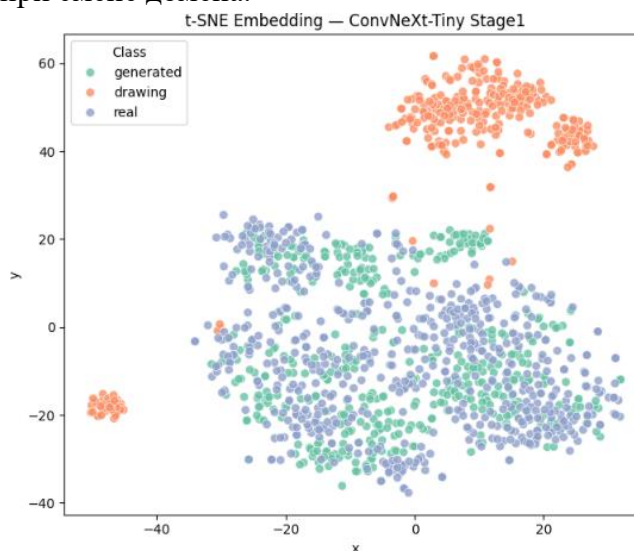


Рисунок 2 – t-SNE-встраивание признаков для модели ConvNeXt-Tiny

На визуализации видно, что класс рисованных изображений формирует компактный кластер, тогда как реальные и сгенерированные изображения частично пересекаются. Это указывает на недостаточную разделимость признаков.

Наблюдаемое смещение моделей в сторону доминирующего класса и ориентация на фоновые признаки дополнительно ухудшают переносимость на новые данные.

Заключение

Проведённое исследование показало, что доменное смещение существенно ограничивает обобщающую способность моделей классификации изображений. Несмотря на высокие метрики на обучающем наборе, при переносе на независимые данные наблюдается значительное падение качества.

Основными факторами деградации являются неоднородность источников данных, дисбаланс классов и использование фоновых признаков. Отсутствие детекции лиц усиливает зависимость моделей от домена.

Повышение устойчивости может быть связано с более контролируемым формированием выборки и предварительным выделением области лица.

Список использованных источников

1. Число случаев мошенничества с использованием дипфейков выросло в 2,3 раза. – URL: <https://www.vedomosti.ru/technology/articles/2025/06/19/1118123-chislo-sluchaev-moshennichestva-s-ispolzovaniem-dipfeikov-viroslo> (дата обращения: 30.10.2025).
2. Singh, L. H. Advancements in detecting Deepfakes: AI algorithms and future prospects – a review / L. H. Singh, P. Charanarur, N. K. Chaudhary // Discover Internet of Things. – 2025. – V. 5. – P. 53.
3. Намиот, Д. Е. Искусственный интеллект и кибербезопасность / Д. Е. Намиот, Е. А. Ильюшин, И. В. Чижов // International Journal of Open Information Technologies. – 2022. – V. 10, № 9.
4. Kaggle. – URL: <https://www.kaggle.com/> (дата обращения: 30.10.2025).

IMPACT OF DOMAIN SHIFT ON THE GENERALIZATION ABILITY OF IMAGE CLASSIFICATION MODELS

S.I. Sablin¹

¹*Samara University, Samara
e-mail: serjjiniuss@gmail.com*

Abstract

This paper studies the impact of domain shift on the generalization ability of avatar classification models (real, drawing, generated). Public Kaggle datasets from multiple sources were combined to increase domain diversity. ResNet-50, MobileNetV3, EfficientNet-B0, and ConvNeXt-Tiny were evaluated. Independent testing showed a substantial drop in macro-F1, indicating sensitivity to domain variations. The effect of training without face detection is also discussed.

Keywords: artificial intelligence, image classification, domain shift, neural networks, deepfakes.