УДК 004.932.2

ПРИМЕНЕНИЕ ПОДХОДОВ ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ ДЛЯ ЗАДАЧИ ОБНАРУЖЕНИЯ DEEPFAKES

Пикуль А.С. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Попов И.Ю. (ИТМО)

Введение. В настоящее время DeepFakes стали серьезной угрозой информационной безопасности, в частности, они стали угрозой для биометрической идентификации. Такая угроза вызывает необходимость разработки эффективных методов их обнаружения. В настоящий момент для решения задачи обнаружения DeepFakes наиболее широко применяются глубокие нейронные сети, визуальные трансформеры, а также обучение без учителя. В данном исследовании рассматривается подход обнаружения DeepFakes с помощью глубокой нейронной сети ResNet-50 [1], обученной без учителя. Для этого был сформирован набор данных и выбраны 8 подходов обучения без учителя. В результате были обучены 8 моделей, которые были сравнены между собой с целью выявления лучшей по точности распознавания DeepFakes.

Основная часть. В настоящий момент можно выделить несколько подходов обучения без учителя:

1. Barlow Twins [2]
2. BYOL [3]
3. DINO [4]
4. MoCov3 [5]
5. SimCLRv2 [6]
6. SimSiam [7]
7. SupCon [8]
8. SwAV [9]

Каждый из упомянутых подходов реализуется на основе энкодера – сети, извлекающей признаки из входного изображения. В каждом из подходов обучения без учителя энкодер применяется в различных вариациях, так как применение энкодера в исходном виде дает менее точные результаты. В качестве энкодера можно применить любую глубокую сверточную нейронную сеть. В экспериментах была применена сверточная нейронная сеть ResNet-50.

Для проведения экспериментов был выбран набор данных DFDC [10], содержащий в себе видео-образцы DeepFakes. Из исходного набора данных был собран новый набор, содержащий в себе фреймы, извлеченные из исходных видео. К каждому фрейму были применены разного рода аугментации: размытие, шум, транспонирование, поворот, изменение размера. В результате аугментаций был получен набор данных, состоящий из 100 тысяч изображений DeepFakes.

Для оценки качества применены метрики ACER, ROC AUC и матрица рассеяния.

Для каждого из восьми подходов обучения без учителя были проведены эксперименты на собранном наборе данных. Наилучшим подходом для задачи обнаружения DeepFakes оказался подход DINO – ACER = 1,94%, AUC = 99,85%.

Выводы. В ходе исследования были проанализированы 8 подходов обучения без учителя для задачи обнаружения DeepFakes. Оценено качество обнаружения для каждой модели. Наилучшие результаты были получены для подхода DINO. Таким образом, данный подход можно применять в дальнейших экспериментах как наилучший. В качестве энкодера можно применять более современные архитектуры, например, визуальный трансформер.

Список использованных источников:

1. Deep Residual Learning for Image Recognition in 2015. –URL: https://arxiv.org/abs/1512.03385 (дата обращения 10.02.2024).
2. Barlow Twins: Self-Supervised Learning via Redundancy Reduction in 2021. –URL: https://arxiv.org/abs/2103.03230 (дата обращения 10.02.2024).
3. Bootstrap your own latent: A new approach to self-supervised Learning in 2020. –URL: https://arxiv.org/abs/2006.07733 (дата обращения 10.02.2024).
4. Emerging Properties in Self-Supervised Vision Transformers in 2021. –URL: https://arxiv.org/abs/2104.14294 (дата обращения 10.02.2024).
5. An Empirical Study of Training Self-Supervised Vision Transformers in 2021. –URL: https://arxiv.org/abs/2104.02057 (дата обращения 10.02.2024).
6. A Simple Framework for Contrastive Learning of Visual Representations in 2020. –URL: https://arxiv.org/abs/2002.05709 (дата обращения 10.02.2024).
7. Exploring Simple Siamese Representation Learning in 2020. –URL: https://arxiv.org/abs/2011.10566 (дата обращения 10.02.2024).
8. Supervised Contrastive Learning in 2020. –URL: https://arxiv.org/abs/2004.11362 (дата обращения 10.02.2024).
9. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments in 2020. –URL: https://arxiv.org/abs/2006.09882 (дата обращения 10.02.2024).
10. The DeepFake Detection Challenge (DFDC) Dataset in 2020. –URL: https://arxiv.org/abs/2006.07397 (дата обращения 10.02.2024).