УДК 004.85 DOI: https://doi.org/10.30837/IYF.IIS.2024.435

РОЗРОБКА МОДЕЛІ ВИЯВЛЕННЯ ФЕЙКОВОГО КОНТЕНТУ НА ОСНОВІ АРХІТЕКТУРИ EFFICIENTNET

Кіценко Ю. О.
Науковий керівник – д.т.н., проф. Смеляков К. С.
Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна
e-mail: yurii.kitsenko@nure.ua

The research is devoted to efficiency evaluation of modern deepfake detection models based on convolutional neural networks (CNN). In today's world, with the growing influence of digital technology and increasing volume of information on the internet, detection of fake images and videos has become increasingly important. Fake content spread through social media and other platforms can cause serious damage, from personal attacks to manipulation of public opinion on a global level. During the study, we trained a model based on the EfficientNet architecture [1]. The model was trained on the Deepfake Detection Challenge dataset [2].

Поява та розповсюдження фейкового контенту в інтернеті є складним явищем, яке має все більший вплив на суспільство, політику та економіку. Зростання доступності технологій глибокого навчання та штучних нейронних мереж робить можливим виготовлення високоякісних фейків. З’явилося багато потужних інструментів, що дозволяють змінювати фотографії, створювати відео з фіктивним контентом і навіть генерувати тексти. Поява моделей виявлення фейкового контенту є відповіддю на зростаючу загрозу фальсифікації інформації в цифровому просторі. З огляду на швидкі та значні зміни в технологіях створення фейків, виникає потреба у високоефективних інструментах та методах для їх виявлення. Серед цих інструментів важливу роль відіграють моделі, засновані на технологіях машинного навчання.

У роботі було побудовано модель розпізнавання фейкового контенту, засновану на архітектурі EfficientNet [1]. EfficientNet – це сімейство нейронних мереж, розроблених для досягнення високої ефективності за рахунок оптимального балансу між розміром мережі та її продуктивністю. Основою архітектури є згорткова базова модель, до якої застосовуються масштабуючі коефіцієнти, що визначають розмір мережі. Такий підхід дозволяє досягти високої точності на завданнях класифікації зображень за мінімальної кількості параметрів, що робить EfficientNet однією з найбільш ефективних архітектур для роботи з обмеженими ресурсами.

EfficientNet пропонує кілька типів моделей, включаючи B0-B7, кожна з яких відрізняється за розміром і кількістю параметрів. Моделі від B0 (найменша) до B7 (найбільша) представляють собою послідовне збільшення глибини, ширини і роздільної здатності мережі. Більші моделі,

такі як B7, забезпечують більшу точність класифікації, але вимагають більшого обсягу ресурсів для навчання і виконання. Менші моделі, наприклад B0, мають меншу кількість параметрів і є більш ефективними при роботі з обмеженими обчислювальними ресурсами, при цьому зберігаючи високу точність. У дослідженні було використано найбільшу з моделей – модель B7.

Для навчання моделі використовувався набір даних DeepFake Detection Challenge (DFDC) [2]. DFDC є найбільшим доступним публічно набором даних відео з підміною облич. Він включає понад 100 000 відеокліпів з участю 3426 платних акторів. Ці відео були створені за допомогою кількох методів, таких як Deepfake, методів на основі застосування генеративно-змагальних мереж та інших методів, що не використовували підходів машинного навчання. DFDC був використаний у проведенні конкурсу Kaggle, що сприяв розвитку засобів виявлення фейкового контенту [3].

Перед навчанням моделі було виконано попередню підготовку даних, яка складалася з наступних етапів:

1. Захват окремих кадрів з відеофайлів.
2. Знаходження облич на отриманих кадрах [4].
3. Вирізання облич для подальшого детального аналізу.
4. Генерація «згорток» (folds) для перехресної валідації.
5. Аугментація даних (augmentation).

Зазначимо, що використаний підхід до виявлення "на основі кадр за кадром" досить вдало зарекомендував себе у багатьох сценаріях. Основною з його переваг є його обмежена обчислювальна вимогливість.

Для побудови та навчання моделі використовувалась відкрита бібліотека машинного навчання Pytorch. Навчання проводилось на протязі 40 епох по 2500 ітерацій та було отримано зважену похибку 0.25, що є досить непоганим результатом, порівняно з іншими сучасними моделями на основі згорткових нейронних мереж [5].

Список використаних джерел:

1. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks // International conference on machine learning. 2019. Vol. 97. P. 6105-6114.

2. Dolhansky B. et al. The deepfake detection challenge (dfdc) dataset //arXiv preprint arXiv:2006.07397. – 2020.

3. Deepfake Detection Challenge // Kaggle. URL https://www.kaggle.com/c/deepfake-detection-challenge (дата звернення: 05.03.2024).

4. Smelyakov K. et al. The neural network models effectiveness for face detection and face recognition // IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream). 2021. P. 1-7.

5. Tolosana R. et al. Deepfakes and beyond: A survey of face manipulation and fake detection //Information Fusion. 2020. Vol. 64. P. 131-148.