УДК 004.0315 DOI: 10.18698/2541-8009-2023-6-860

ПРЕДОБРАБОТКА ОБУЧАЮЩЕГО НАБОРА ДАННЫХ ДЛЯ СИСТЕМЫ РАСПОЗНАВАНИЯ ДИПФЕЙКОВ

С.В. Астахов fzastahov@gmail.com
Н.Б. Гендина gendina.nina@mail.ru
Т.И. Кадыров kadyyrovti@student.bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

В данной статье описывается процесс формирования и предобработки обучающего набора данных для системы распознавания дипфейков, использующей концепцию "регионов интереса" (ROI). Данная концепция позволяет анализировать как статические изображения, так и видео, что делает ее весьма универсальной. Разработки в сфере определения дипфейков весьма актуальны в настоящее время, так как визуальная "достоверность" изображений, создаваемых нейросетями увеличивается по мере совершенствования программного и аппаратного обеспечения, что открывает все больше возможностей для использования недостоверной информации в целях мошенничества, шантажа и вмешательства в политические процессы, что необходимо пресекать. В данной статье рассмотрены как теоретические основы определения типовых зон на изображении, так и фрагменты программы, разработанной авторами для формирования и предобработки обучающего набора данных для системы распознавания дипфейков.

Ключевые слова

Компьютерное зрение, openCV, машинное обучение, распознавание образов, дипфейк, гистограмма направленных градиентов, каскад регрессионных моделей, dlib, python.

Введение. В наше время социальные сети и мессенджеры стали неотъемлемой частью жизни многих людей. Однако, вместе с возможностью общения и обмена информацией, появилась и проблема недостоверных новостей, видео и фотографий.

Дипфейки - это фотографии или видео, созданные с помощью искусственного интеллекта, которые могут быть использованы для создания фальшивых новостей или дезинформации [1].

В связи с этим, возникает необходимость в разработке методов определения дипфейков, которые позволят бороться с распространением фальшивой информации.

В данной статье рассмотрен процесс формирования обучающего набора данных для системы распознавания дипфейков, использующей концепцию “регионов интереса” (ROI) [2].

Идентификация зоны лица. Алгоритм функционирования проектируемой системы распознавания дипфейков, что первым этапом на пути обработки изображения является выделение зон интереса. На основе характеристик изображения в этих зонах и будет производиться дальнейшее определение признаков дипфейка.

Для определения зон интереса подпрограмма предварительной обработки данных определяет положение 68 типовых точек лица с помощью библиотеки dlib [3].

Для обнаружение участка исходного изображения, в котором находится лицо, эта библиотека использует так называемую гистограмму направленных градиентов [4]. Для этого исходное изображение сначала переводится в черно-белый формат (рисунок 1).

Рис. 1. Пример исходного изображения

Затем, для каждого небольшого участка изображения анализируется градиент (направление изменения) яркости в нем (рисунок 2).

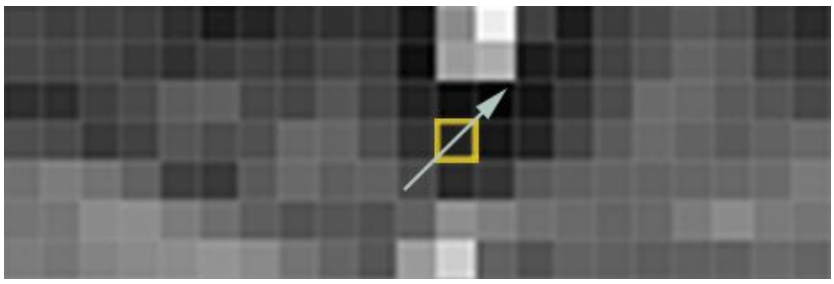


Рис. 2. Вычисление градиента

Такой подход нивелирует влияние освещения и позволяет упростить вычисления при дальнейшем анализе. Пример изображения, обработанного данным алгоритмом приведен на рисунке 3.

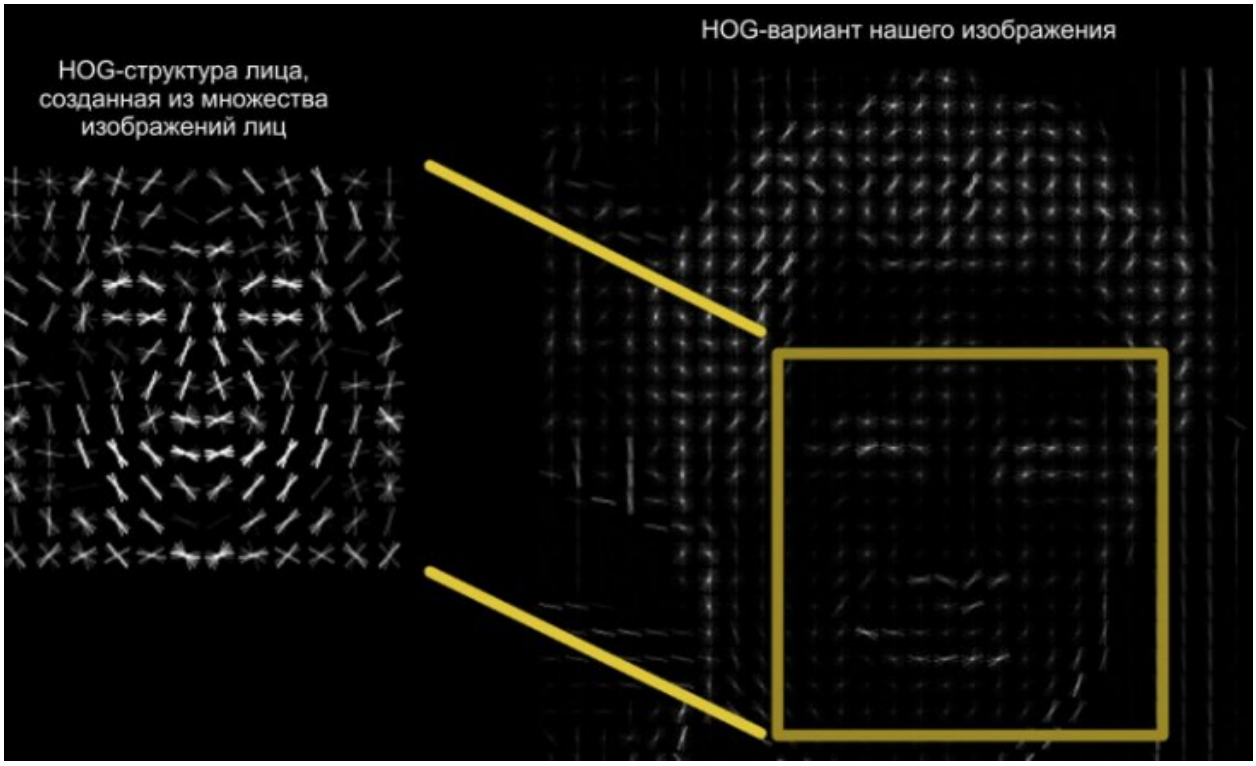


Рис. 3. Представление изображения в виде набора градиентов

Идентификация зон интереса. Для определения положения отдельных точек в зоне, где было обнаружено лицо, библиотека использует каскад регрессионных моделей [5].

Прежде всего, использование каскадов решает такую проблему, как влияние освещения и артефактов на точность определения точек. Изначальная дилемма состоит в том, что при определении признаков регрессорами, признаки могут быть искажены. Для борьбы с этим используется каскадная модель, которая итеративно уточняет форму лица и значения признаков в ее рамках.

Второй проблемой является тот факт, что алгоритм должен оценить форму, вектор высокой размерности, который наилучшим образом согласуется с данными изображения и моделью формы. Успешные алгоритмы решают эту проблему,

предполагая, что предполагаемая форма должна лежать в линейном подпространстве, которое может быть обнаружено, например, путем нахождения основных компонентов форм [6].

Регрессоры представляют из себя деревья решений, используемых в сочетании с технологией градиентного бустинга на основе квадратичной функции потерь. Эти деревья решений анализируют разность значений интенсивности в парах пикселей на изображении, распределение расстояний между которыми определяется априорной функцией вероятности. Априорное распределение позволяет алгоритму бустинга эффективно исследовать большое количество релевантных признаков. Результатом является каскад регрессоров, которые могут локализовать лицевые метки на изображении лица в положении анфас.

Решение в каждом узле основано на пороговом значении разности значений интенсивности в паре пикселей. Это довольно простой подход, но он гораздо более эффективен, чем пороговое значение с одной интенсивностью, из-за его относительной нечувствительности к изменениям глобального освещения. К сожалению, недостатком использования разностей пикселей является то, что число потенциальных признаков является квадратичным по отношению к количеству пикселей в среднем изображении. Это затрудняет поиск хороших пар без поиска по очень большому их числу. Однако этот ограничивающий фактор может быть в некоторой степени ослаблен с учетом структуры данных изображения. Для этого при выборе расстояния между пикселями используется экспоненциальное распределение [7].

Стоит так же отметить, что данная модель умеет работать с обучающими примерами, где некоторая часть рассматриваемых меток не определена, задавая для них 0 вес при обучении.

Рассматриваемая модель состоит из большого числа “слабых” регрессоров, которые в свою очередь объединяются в несколько “сильных”. Зависимость точности модели от числа “сильных” регрессоров показана на рисунке 1.

Рис. 4. Зависимость точности модели от числа регрессоров

Модель, используемая в библиотеке dlib обучалась на наборе данных HELEN, состоящем из 2000 размеченных изображений человеческих лиц из социальной сети Flickr [8].

Сравнение ошибок для используемых в бибилиотеке регрессоров в сопоставлении с другими возможными вариантами показано на рисунке 2. Ось абсцисс описывает число уровней каскада, ординат — значение ошибки. Кривая для используемых регрессоров обозначена зеленым цветом.

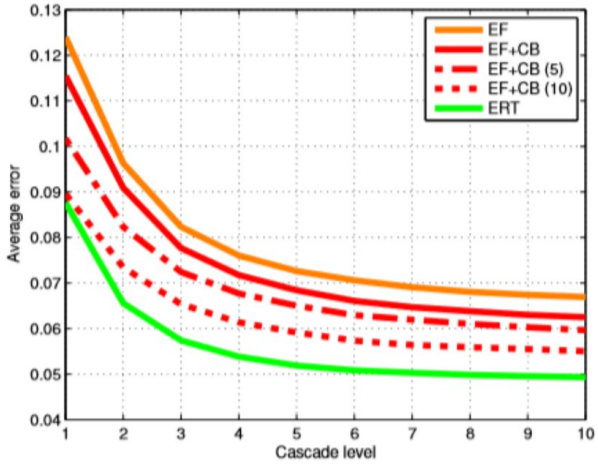
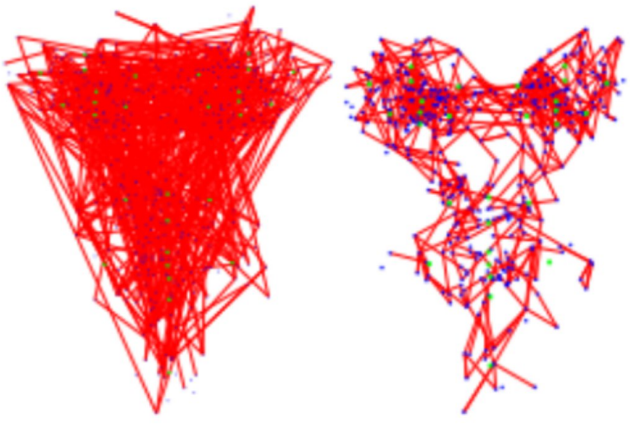


Рис. 5. Уровни ошибок в зависимости от числа уровней каскада для различных типов регрессоров

На рисунке 3 показаны пары пикселей, отбираемые в качестве признаков при равномерном и экспоненциальном распределениях расстояний между ними соответственно. Из рисунка интуитивно понятно, что априорный выбор экспоненциальной функции повышает эффективность и точность модели, стимулируя ее выбирать более близкие точки для определения деталей лица.



(a) Равномерное распределение (b) Экспоненциальное распределение

Рис. 6. Выбор признаков при различных функциях распределения

Преимущество использования многоуровневой модели показано в таблице 1. Как видно, число уровней каскада влияет на значение ошибки куда сильнее, чем число слабых регрессоров на одном уровне каскада.

Таблица 1 — зависимость ошибки модели от числа уровней каскада и числа слабых регрессоров

	Число уровней	Число слабых регр.	Число уровней	Число слабых регр.	Число уровней	Число слабых регр
Конф. модели	1	500	1	5000	10	500
Значение ошибки	0.085		0.074		0.049	

Зависимость ошибки модели при 10 уровнях каскада для различного числа примеров показана в таблице 2.

Таблица 2 — зависимость ошибки от числа примеров

Число примеров	100	200	500	1000	2000
Значение ошибки	0.090	0.074	0.059	0.054	0.049

Зависимость значения ошибки от числа примеров для различного числа уровней каскада продемонстрирована на рисунке 4.

Пример реализации функции определения формы лица на основе библиотеки dlib и openCV [9] приведен в листинге 1.

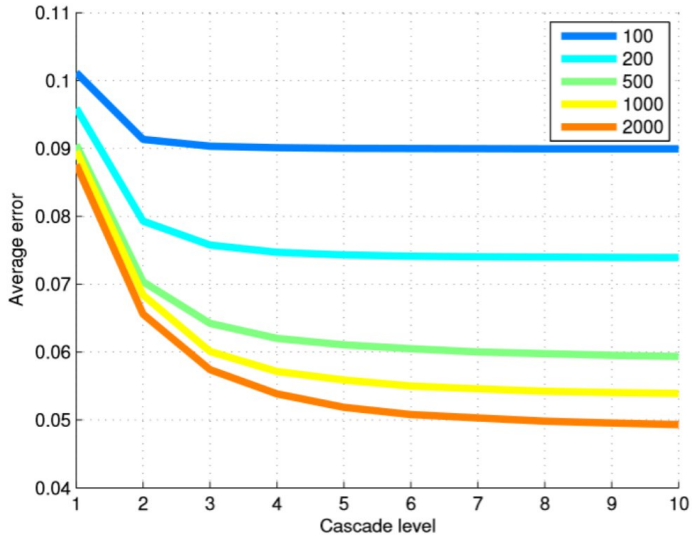


Рис. 7. Зависимость значения ошибки от числа примеров для различного числа уровней каскада

Листинг 1 — функция определения формы лица

```
# импорт библиотек
import collections
import cv2 as cv
from imutils import face_utils
import numpy as np
import imutils
import dlib
import numpy as np

# обучающий набор данных
PREDICTOR_DATASET = './shape_predictor_68_face_landmarks.dat'

# определитель форм лица
PREDICTOR = dlib.shape_predictor(PREDICTOR_DATASET)
# функция определения формы лица
def detect_faces(image):

    # определитель зон, содержащих изображения лиц
    detector = dlib.get_frontal_face_detector()
    predictor = PREDICTOR

    # поиск лиц на ч/б изображении
    gray = cv.cvtColor(image, cv.COLOR_BGR2GRAY)
    rects = detector(gray, 1)

    # массив для координат точек
    shapes = []

    # для каждого изображения лица
    for (i, rect) in enumerate(rects):

        # определить 68 точек
        shape = predictor(gray, rect)

        # конвертировать в массив NumPy.array
        shape = face_utils.shape_to_np(shape)

        # добавить форму к существующим
        shapes.append(shape)

    # вернуть точки для всех лиц
    return shapes
```

Формирование обучающего набора данных. Рассмотренная выше функция является частью модуля “faceparts”, разработанного в рамках данной работы и используется затем для предобработки обучающего набора изображений. Программный код функции, порождающей 1 строку обучающего набора в формате pandas.DataFrame [10] приведен в листинге 2.

Листинг 2 — функция предобработки данных

```
# импорт библиотек
import pandas as pd
import faceparts as fp
import cv2 as cv
import numpy as np
import cv2 as cv

# функция получения строки датафрейма
# параметры: имя файла, флаг фейка, кооридната ROI, кадр,
масштаб ROI
def generate_dataframe(filename, is_fake, shape, frame,
frame_scale_percent):
    # перевод в градации серого
    gray_frame = cv.cvtColor(frame.copy(), cv.COLOR_BGR2GRAY)
    data = {}

    # расчет радиуса лица
    r = pow(
            pow(shape[fp.LEFT_EAR_POINT][0] -
shape[fp.RIGHT_EAR_POINT][0], 2) +
            pow(shape[fp.LEFT_EAR_POINT][1] -
shape[fp.RIGHT_EAR_POINT][1], 2),
            0.5
            )
    # расчет размера ROI
    rect_size = int((r * frame_scale_percent) / 100.0)

    # заполнение вспомогательных полей датафрейма
    data['filename'] = [filename]
    data['fake'] = [is_fake]

    # для каждой зоны на лице
    for zone in fp.FACIAL_LANDMARKS_IDXS.keys():
        # для каждой точки в зоне
        for point_id in
range(fp.FACIAL_LANDMARKS_IDXS[zone][0],
fp.FACIAL_LANDMARKS_IDXS[zone][1]):
```

```
            # получение координат точек и границ ROI
            point = shape[point_id]
            point_x = point[0]
            point_y = point[1]
            x1, y1, x2, y2 = \
                int(point_x - int(rect_size/2)), \
                int(point_y - int(rect_size/2)), \
                int(point_x + int(rect_size/2)), \
                int(point_y + int(rect_size/2))

            # запись бинарных данных
            data['pt_' + str(point_id) + '_' + zone +'_raw'] =
[gray_frame[y1:y2, x1:x2].ravel()]

    # информация обо всей зоне лица
    point_x = int((shape[fp.LEFT_EAR_POINT][0] +
shape[fp.RIGHT_EAR_POINT][0])/2)
    point_y = int((shape[fp.LEFT_EAR_POINT][1] +
shape[fp.RIGHT_EAR_POINT][1])/2)
    x1, y1, x2, y2 = \
                int(point_x - int(r)), \
                int(point_y - int(r)), \
                int(point_x + int(r)), \
                int(point_y + int(r))
    data['overall_face_raw'] = [gray_frame[y1:y2,
x1:x2].ravel()]

    # перевод в формат pandas.Dataframe
    df = pd.DataFrame(data=data)
    return (df.copy())
```

Пример фрагмента получаемого набора данных приведен на рисунке 7.

Первые два столбца содержат названия исходных файлов и флаг дипфейка, остальные — данные из регионов интереса в двоичном виде.

Таким образом, в ходе данной работы был продемонстрирован процесс идентификации регионов интереса на изображении и процесс формирования из них обучающего набора данных для системы распознавания дипфейков.

	filename	fake	pt_48_mouth_raw	pt_49_mouth_raw	pt_50_mouth_raw	pt_51_mouth_raw	pt_52_mouth_raw	pt_53_mouth_raw	pt_5
0	kzlyhfpnil.mp4	True	[26, 22, 23, 23, 23, 23, 23, 24, 24, 24, 24, 2...	[32, 32, 32, 32, 32, 32, 31, 31, 30, 30, 30, 3...	[29, 28, 28, 26, 26, 26, 25, 25, 25, 24, 23, 2...	[24, 23, 23, 23, 23, 23, 23, 23, 23, 23, 22, 2...	[22, 21, 21, 19, 19, 19, 18, 17, 17, 17, 17, 1...	[16, 16, 16, 17, 17, 16, 16, 16, 16, 16, 16, 1...	[17, 11,
1	kzlyhfpnil.mp4	True	[31, 38, 43, 45, 44, 43, 42, 40, 40, 40, 39, 3...	[42, 43, 42, 42, 42, 40, 39, 37, 36, 35, 33, 3...	[37, 35, 33, 32, 31, 29, 28, 28, 26, 25, 25, 2...	[28, 26, 25, 25, 24, 22, 21, 19, 18, 17, 16, 1...	[17, 16, 16, 15, 14, 14, 14, 15, 15, 14, 14, 1...	[15, 14, 14, 15, 15, 15, 14, 14, 15, 15, 15, 1...	[15, 15,
2	kzlyhfpnil.mp4	True	[47, 47, 46, 45, 44, 43, 42, 40, 39, 38, 37, 3...	[40, 40, 42, 40, 40, 39, 37, 37, 35, 32, 32, 3...	[33, 32, 32, 32, 31, 31, 30, 31, 29, 26, 23, 2...	[29, 26, 23, 22, 20, 19, 18, 18, 17, 17, 17, 1...	[19, 19, 19, 19, 19, 18, 18, 17, 17, 16, 15, 1...	[12, 14, 14, 12, 12, 14, 14, 14, 15, 15, 16, 1...	[12, 14,
3	kzlyhfpnil.mp4	True	[36, 32, 31, 31, 30, 29, 31, 31, 32, 33, 33, 3...	[38, 37, 36, 35, 33, 33, 33, 31, 30, 31, 30, 2...	[32, 30, 30, 29, 28, 28, 26, 25, 25, 24, 23, 2...	[25, 24, 23, 22, 22, 22, 21, 21, 20, 19, 18, 1...	[17, 17, 16, 15, 14, 15, 15, 15, 15, 16, 16, 1...	[16, 16, 15, 15, 14, 14, 14, 14, 15, 15, 15, 1...	[15, 14,
4	kzlyhfpnil.mp4	True	[38, 38, 37, 37, 37, 36, 36, 36, 35, 33, 33, 3...	[42, 40, 40, 39, 38, 38, 37, 37, 35, 32, 32, 3...	[35, 33, 33, 31, 30, 29, 29, 28, 26, 25, 24, 2...	[25, 23, 22, 21, 18, 17, 17, 17, 17, 17, 17, 1...	[18, 18, 18, 18, 17, 18, 17, 16, 16, 15, 14, 1...	[12, 14, 14, 14, 14, 14, 15, 15, 15, 16, 17, 1...	[15, 16,

Рис. 8. Фрагмент обучающего набора данных

Заключение. Проблема недостоверных новостей, видео и фотографий становится все более актуальной в информационном обществе. Дипфейки могут нанести серьезный ущерб как отдельным людям, так и обществу в целом.

Однако, существуют различные методы определения дипфейков, которые позволяют бороться с распространением фальшивой информации.

В данной статье был рассмотрен процесс предобработки данных для формирования обучающего набора системы обнаружения дипфейков, использующей концепцию регионов интереса (ROI).

Литература

[1] Collins, Forged Authenticity: Governing Deepfake Risks / Collins, Aengus // EPFL International Risk Governance Center (IRGC) : электронный журнал. – URL: https://infoscience.epfl.ch/record/273296. – Дата публикации: 16.12.2019.

[2] Li, Y. Exposing DeepFake Videos By Detecting Face Warping Artifacts / Y. Li, S. Lyu // CVPR : электронный журнал. – URL: https://openaccess.thecvf.com/content_CVPRW_2019/papers/MediaForensics/Li_Exposing_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.pdf. – Дата публикации: 15.11.2018.

[3] DLib.net : сайт. – URL: http://dlib.net/train_shape_predictor.py.html (дата обращения: 16.11.2023)

[4] Современное распознавание лиц с глубинным обучением // Habr.com : сайт. – URL: https://habr.com/ru/articles/306568/ (дата обращения: 18.11.2023)

[5] Cascaded Continuous Regression for Real-Time Incremental Face Tracking / E. Sanchez-Lozano, B. Martinez, G. Tzimiropoulos, M. Valstar // European Conference on Computer Vision : электронный журнал. – URL: https://link.springer.com/chapter/10.1007/978-3-319-46484-8_39. – Дата публикации: 17.09.2016.

[6] Kazemi, V. One Millisecond Face Alignment with an Ensemble of Regression Tree / V. Kazemi, J. Sullivan // CVPR : электронный журнал. – URL: https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Kazemi_One_Millisecond_Face_2014_CVPR_paper.pdf (дата обращения: 18.11.2023).

[7] Analysis and Improvement of Facial Landmark Detection / P. Kopp, D. Bradley, T. Beeler, M. Gross // ResearchGate : электронный журнал. – URL: https://www.researchgate.net/publication/332866914_Analysis_and_Improvement_of_Facial_Landmark_Detection. – Дата публикации: 01.03.2019.

[8] Helen dataset // ifp.illinois.edu : сайт. – URL: http://www.ifp.illinois.edu/~vuongle2/helen/ (дата обращения: 24.11.2023)

[9] OpenCV-Python Tutorials : сайт. – URL: https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html (дата обращения: 04.02.2024)

[10] pandas.DataFrame : сайт. – URL: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html (дата обращения: 04.02.2024

Астахов Сергей Викторович — магистр кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Гендина Нина Борисовна — магистр кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Кадыров Тимерлан Ильдарович — магистр кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Сотников Алексей Александрович, кандидат технических наук, доцент кафедры «Компьютерные системы и сети», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:
Астахов С.В., Гендина Н.Б., Кадыров Т.И. Предобработка обучающего набора данных для системы распознавания дипфейков. *Политехнический молодежный журнал*, 2024, № 07 (83). http://dx.doi.org/10.18698/2541-8009-2023-6-865

PREPROCESSING OF THE TRAINING DATASET FOR THE DEEPFAKE RECOGNITION SYSTEM

S.V. Astakhov fzastahov@gmail.com
N.B. Gendina gendina.nina@mail.ru
T.I. Kadyrov kadyyrovti@student.bmstu.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

This article describes the process of forming and preprocessing a training dataset for a deepfake recognition system using the concept of "regions of interest" (ROI). This concept allows you to analyze both static images and videos, which makes it very versatile. Developments in the field of detecting deepfakes are very relevant at the present time, since the visual "reliability" of images created by neural networks increases with the improvement of software and hardware, which opens up more and more opportunities for using false information for fraud, blackmail and interference in political processes, which must be stopped. This article discusses both the theoretical foundations for determining typical zones in an image and fragments of the program developed by the authors for the formation and preprocessing of a training dataset for the deepfake recognition system.

Keywords

Computer vision, OpenCV, machine learning, pattern recognition, deepfake, histogram of directional gradients, cascade of regression models, dlib, python.

References

[1] Collins, Forged Authenticity: Governing Deepfake Risks / Collins, Aengus // EPFL International Risk Governance Center (IRGC) : electronic journal. – URL: https://infoscience.epfl.ch/record/273296 . – Date of publication: 16.12.2019.

[2] Li, Y. Exposing DeepFake Videos By Detecting Face Warping Artifacts / Y. Li, S. Lyu // CVPR : electronic journal. – URL: https://openaccess.thecvf.com/content_CVPRW_2019/papers/MediaForensics/Li_Exposing

_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.pdf. – Date of publication: 11/15/2018.

[3] DLib.net : website. – URL: http://dlib.net/train_shape_predictor.py.html (date of access: 11/16/2023)

[4] Sovremennoe raspoznavanie lic s glubinnym obucheniem[Modern face recognition with deep learning] // Habr.com : website. – URL: https://habr.com/ru/articles/306568 / (accessed: 11/18/2023) (In Russ.).

[5] Cascading continuous regression for incremental face tracking in real time / E. Sanchez-Lozano, B. Martinez, G. Tsimiropoulos, M. Valstar // European Conference on Computer Vision : electronic Journal. – URL: https://link.springer.com/chapter/10.1007/978-3-319-46484-8_39 . – Date of publication: 17.09.2016.

[6] Kazemi, V. Face alignment in one millisecond using an ensemble of regression trees / V. Kazemi, J. Sullivan // CVPR : electronic journal. – URL: https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Kazemi_One_Millisecond_Face_2014_CVPR_paper.pdf (accessed: 11/18/2023).

[7] Analysis and improvement of facial recognition / P. Kopp, D. Bradley, T. Bieler, M. Gross // ResearchGate : electronic journal. – URL: https://www.researchgate.net/publication/332866914_Analysis_and_Improvement_of_Facial_Landmark_Detection. – Date of publication: 03/01/2019.

[8] Helen's dataset // ifp.illinois.edu : website. – URL: http://www.ifp.illinois.edu /~vuongle2/helen/ (accessed: 11/24/2023)

[9] OpenCV-Python Tutorials : website. – URL: https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html (date of access: 02/04/2024)

[10] pandas.Data frame : website. – URL: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html (update date: 02/04/2024)

Astakhov S.V. — Master of the Department of Computer Systems and Networks, Bauman Moscow State Technical University, Moscow, Russian Federation.

Gendina N.B. — Master of the Department of Computer Systems and Networks, Bauman Moscow State Technical University, Moscow, Russian Federation.

Kadyrov T.I. — Master of the Department of Computer Systems and Networks, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Sotnikov A.A., Candidate of Technical Sciences, Associate Professor of the Department of Computer Systems and Networks, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:

Astakhov S.V., Gendina N.B., Kadyrov T.I. Preprocessing of a training dataset for a deepfake recognition system. *Politekhnicheskiy molodezhnyy zhurnal*, 2024, No. 07 (83). http://dx.doi.org/10.18698/2541-8009-2023-6-865