АЛГОРИТМ ЗАЩИТЫ ИНФОРМАЦИИ ПРИ РАСПОЗНАВАНИИ ЛИЦА ЧЕЛОВЕКА

Русакович А. С.

Белорусский государственный университет, Минск, Беларусь

За последнее десятилетие технологии на основе компьютерного зрения стали быстрее и точнее идентифицировать людей. В настоящее время все более широкое распространение получают биометрические системы идентификации человека. Традиционные системы идентификации требуют знания пароля, наличия ключа, идентификационной карточки либо иного идентифицирующего предмета, который можно забыть или потерять. В отличие от них биометрические системы основываются на уникальных биологических характеристиках человека, которые однозначно определяют конкретного человека. К таким характеристикам относятся отпечатки пальцев, форма ладони, узор радужной оболочки, изображение сетчатки глаза. Распознавание человека по изображению лица выделяется среди биометрических систем, во-первых, тем, что, не требуется специальное или дорогостоящее оборудование, достаточно персонального компьютера и обычной видеокамеры. Во-вторых, не нужен физический контакт с устройствами, необходимо просто задержаться перед камерой на небольшое время.

Биометрические системы распознавания лиц имеют общие принципы работы. Для начала выделяется лицо человека, будь он один или находясь в толпе. Затем фиксируется фотография лица и начинается ее анализ. Программы для распознавания лиц анализируют узловые точки, такие как расстояние между глазами или форму скул. После этого черты лица преобразуются в цифровой код – отпечаток лица (faceprint). Далее код сравнивается с базой данных отпечатков лиц. Например, в работе [6] представлен алгоритм защиты лиц путем деидентификации т.е. затемнения деталей лица с помощью цифровой модификации видеоизображений. Другой способ описывается в работе [10], где используется комплекс атак на системы распознавания с использованием генеративной сети, не допуская при этом незаметной ошибки при распознавании. Изменения на уровне пикселей, так называемые заплатки лица человека предлагаются в работе [9]. Интересными примерами являются статьи [3, 4] где используются различные методы генерации изображений с внесением искажений, которые приводят к ошибочной классификации.

Анализ экспериментальных исследований показал, что предсказание модели распознавания лиц напрямую зависит от расположения таких точек лица на входном изображении. Кроме того, было обнаружено, что верхняя часть лица является более полезной для распознавания, чем нижняя. Тип модификации ограничен сложностью, точностью и скоростью доступных в настоящее время алгоритмов обработки изображений, а также качеством данных и деталей, необходимых для работы системы [1]. Собрав лучшие практики рассмотренных выше методов и минимизировав появление шумов с целью улучшения качества измененного изображения в данной работе предлагается подход, который поможет обеспечить защиту личных изображений при обработке моделями распознавания лиц путем внесения изменений в исходное изображение на уровне пикселей. Такие изображения могут привести к

кардинальным изменениям логики классификации. Исследование показало отклонение в 30 – 40% по сравнению с уже существующими методами.

Алгоритм защиты информации в лице человека

Далее будут кратко изложены общие шаги, на которых базируется предложенный алгоритм защиты информации при распознавании лиц.

Пусть имеется исходное изображение $I_{i,j}$, представляющее собой матрицу размера $m \times n$, где $i = \overline{0, m-1}$, $j = \overline{0, n-1}$. Разбив его на блоки размером 8×8, итерация за итерацией проводим сравнение окраски пикселей для выявления контрольных точек на изображении лица. Затем в выбранных блоках изображения вносятся пиксельные изменения (метки), которые незаметны при просмотре людьми, но приводят к формированию некорректных моделей в системах машинного обучения.

Сам процесс пиксельных изменений выглядит следующим образом: сканируя изображение, для каждого пикселя вычисляется предсказанный пиксель $\hat{I}_{i,j}$, ошибка предиктора $E_{i,j}$ и статическая мера $d_{i,j}$. При успешном выполнении условий выбора пикселей происходит внедрение на основе схемы расширения разности и, далее, реализуется метка пикселей по схеме смещения гистограммы [6]. В данном случае внедрение сводится к добавлению в изображение комбинации пикселей (метки), которые воспринимаются алгоритмами машинного обучения как характерные для изображаемого объекта шаблоны и приводят к искажению распознавания человека.

Результаты экспериментов

В качестве экспериментальных данных была использована серия 224×224 8-битовых фотографий людей. В частности, хранение изображений организовано по стандарту JPEG т.к. он позволяет добиться высокого коэффициента сжатия сохраняя высокое качество. При этом, качество изображений, напрямую влияет на точность распознавания. Сжатие снижает степень успешности классификации изображения человека [2].

Выбранные изображения подвергаем пиксельному изменению таким образом, чтобы сохранить исходные характеристики, но при этом снизить точность распознавания лиц, а затем используем в тестировании. В результате получаем изображения с внесенными изменениями в контрольные области без видимых повреждений лица человека (рисунок 1).

группа изображений а группа изображений б

Рис. 1. Изображения до обработки методом (группа а) и после обработки (группа б) с указанием областей внедрения

Используя измененные изображения в модели, получаем искаженное представление о личности, т.е. понимание того, что делает человека уникальным. Эксперименты с выбранной группой изображений проводились с применением нейросетевого метода распознавания лиц. Результаты работы наиболее часто используемых методов распознавания представлены в таблице 1, где изображение человека классифицируется с меньшей степенью точности.

Табл. 1. Результаты оценки качества распознавания

Метод распознавания лица человека	Вероятность правильного результата распознавания изображения (до применения метода), %	Вероятность правильного результата распознавания изображения (после применения метода), %
Нейросетевой метод	97%	63%
Расчет геометрических характеристик лица [5]	86%	76%
Метод главных компонент [7]	95%	89%
Скрытые Марковские модели [8]	87%	78%

Изложенный алгоритм позволяет добиться высоких показателей качества с внедрением более одного бита на пиксель, однако остается восприимчивым к различного рода атакам и требует выполнения индивидуальных условий для избегания так называемой проблемы выхода значений пикселей из возможных пределов (0, 255). Под атакой в данном случае понимается комплекс мер, нацеленных на выявление изображений, обработанных подобного род методами. Для увеличения стойкости можно использовать частотные алгоритмы, где перед добавлением в изображение комбинации пикселей происходят некоторые преобразования (зашумление, фильтрация и др.).

Литература

1. Абламейко С. В., Лагуновский Д. М. Обработка изображений. Технология, методы, применение / Учебное пособие. – Мн.: Амалфея, 2000. – 304 с.
2. Aditya P., Sen R., Druschel P., Joon Oh S., Benenson R., Fritz M., Schiele B., Bhattacharjee B., and Wu T. T. I-pic: A platform for privacy-compliant image capture. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, pages 235–248, 2016.
3. Chen P., Zhang H., Sharma Y., Yi J., Hsieh C. ZOO: Zeroth Order Optimization Based Black-box Attacks to Deep Neural Networks without Training Substitute Models. AISec 17: Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, pages 15–26, 2017.
4. Dabouei A., Soleymani S., Dawson J. Fast geometrically-perturbed adversarial faces. IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1979-1988, 2019.
5. Manjunath B.S. A feature based approach to face recognition / B.S. Manjunath, R. Chellappa, C. von der Malsburg // IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, pages 373-378, 1992.
6. Newton E., Sweeney L., Malin B. Preserving Privacy by De-identifying Facial Images, Carnegie Mellon University, School of Computer Science, Technical Report, CMU-CS-03-119, pages 1-26, 2003.
7. Parveen Y., Kadian R. View-based and modular eigenspaces for face recognition // International Journal of Science and Research (IJSR), pages 989-992, 2013.
8. Samaria F.S. Parameterisation of a stochastic model for human face identification. IEEE Workshop on Applications of Computer Vision, pages 138-142, 1994.
9. Shan S., Wenger E., Zhang J., Li H., Zheng H., Zhao B. Y. Protecting Personal Privacy against Unauthorized Deep Learning Models. In Proceedings of USENIX Security Symposium, pages 1-16, 2020.
10. Yang L., Song Q.,·Wu Y. Attacks on state-of-the-art face recognition using attentional adversarial attack generative network, Pattern Recognition and Intelligence Vision Lab, Beijing University of Posts and Telecommunications, Beijing, pages 1-10, 2020.