

Проверка наличия диплома о высшем образовании по фотографиям лиц

Р.С. Зулкашев
Санкт-Петербургский
государственный университет
аэрокосмического приборостроения
Санкт-Петербург, Россия
ruslanzulksama@yandex.ru

М.Д. Поляк
Санкт-Петербургский
государственный университет
аэрокосмического приборостроения
Санкт-Петербург, Россия
markpolyak@gmail.com

Аннотация— в работе рассматривается гипотеза о наличии скрытых черт (признаков) на лице человека, из которых можно извлечь некую дополнительную информацию о самом человеке. Проверяется частный случай этой гипотезы – определение наличия высшего образования по фотографии лица с помощью методов машинного обучения.

Ключевые слова— распознавание лиц, нейронные сети, классификация, физиогномика.

1. ВВЕДЕНИЕ

Искусство определения персональных черт человека по его лицу, называемое физиогномикой, известно еще со времен древней Греции [1]. В то время как обычно изучением физиогномики занимаются такие отрасли науки как психология, криминалистика и медицина, современные достижения в области машинного обучения и распознавания лиц позволяют взглянуть на нее с новой стороны. Так, в работе [2] рассматривалась возможность отличать лица преступников от законопослушных граждан, а в работах [3, 4] подтверждена возможность определять по лицу сексуальную ориентацию и политические предпочтения человека. В данной работе мы рассматриваем простую гипотезу: оставляют ли знания (т.н. «ученость») на лице человека след, достаточный для того, чтобы его смогла уловить нейросетевая модель. Или, в более простой формулировке, можно ли по фотографии лица определить наличие у человека высшего образования (далее ВО).

2. ПОИСК ПОДХОДЯЩЕЙ МОДЕЛИ

Для выделения информативных признаков из изображения лица человека предложено использовать нейросетевую модель, т.к. подобные модели в последние годы получили широкую популярность и позволили достичь значительных успехов в решении задач обработки изображений, в том числе и фотографий лиц людей. При поиске подходящей модели нейронной сети были заданы следующие условия:

- модель должна быть предобучена на большом наборе фотографий лиц людей;
- быть простой в реализации и использовании;
- модель должна быть компактной (поместиться в 2Гб видеопамяти).

В рамках данной работы мы остановились на двух моделях: VGG-Face и Facenet [5].

3. ПЕРВАЯ МОДЕЛЬ

В качестве первой модели была рассмотрена Facenet [6]. Использовалась модель, предобученная на наборе данных CASIA-WebFace для решения задачи идентификации. Итоговая точность (accuracy)

предобученной модели заявлена авторами как 0,9905. Принцип работы модели: на вход подается фотография на лица человека, на выходе формируется вектор числовых признаков (эмбединг).

4. ВТОРАЯ МОДЕЛЬ

Второй моделью выступала VGG-Face [7]. Ее качество предсказаний на датасете LFW достигает 97,27%. Данная модель выдает вероятности отнесения к классам. По причине простой архитектуры без особых проблем становится возможным использовать VGG-Face двумя разными способами: как полноценный классификатор и как экстрактор вектора признаков.

5. НАБОР ДАННЫХ

Датасет составлялся согласно следующим условиям:

1. Классы должны иметь равное кол-во экземпляров
2. Пол и возраст должны иметь равномерное распределение
3. Изображения лиц должны быть в хорошем качестве

Для набора данных отбирались люди из максимально противоположных слоев населения: ученые, профессора с одной стороны и преступники с актрисами взрослого кино с другой. Исследования показывают, что среди актрис взрослого кино, а также среди осужденных по тяжелым статьям (воровство, мародерство, убийство), лишь небольшой процент имеет ВО [8, 9]. Помимо этого, датасет был расширен медийными личностями с ручной проверкой наличия ВО по данным публичных источников.

В итоге был подготовлен датасет, состоящий из более 500 снимков лиц людей, относительно сбалансированный по количеству изображений каждого класса, а также по полу и возрасту.

6. ОБУЧЕНИЕ

Датасет был разделен на обучающую и тестовую выборки в пропорции 4:1. Процесс обучения был построен следующим образом: в предобученную нейронную сеть подавалось изображение, по нему строился вектор признаков, далее по полученным эмбедингам обучался отдельный бинарный классификатор.

А. Результаты с FaceNet

В качестве классификатора была выбрана «машина опорных векторов». В таблице 1 представлены метрики качества построенного классификатора. В полной версии статьи также приведены результаты визуализации выборки с помощью метода главных компонент и t-SNE.

Таблица I. МЕТРИКИ SVM ДЛЯ FACENET

Accuracy	F1-score	AUC ROC
0,78	0,78	0,85

Исходя из результатов, можно сказать, что классификатор научился находить некоторые закономерности между человеческим лицом и наличием ВО.

Б. VGG-Face как модель-экстрактор признаков

VGG-Face выдает вероятности отнесения к классам. Поэтому последний слой был удален, дабы на выходе получать вектор признаков (размером 4096 против 512 в Facenet), который будет поступать в машину опорных векторов. Посчитанные значения метрик представлены в Таблице 2.

Таблица II. ТАБЛИЦА 2. МЕТРИКИ SVM ДЛЯ VGG-FACE

Accuracy	F1-score	AUC ROC
0,8	0,78	0,89

По сравнению с Facenet, показатели метрик VGG-Face не ухудшили итоговый результат классификации. Визуализация методом главных компонент не дала новой информации о разделимости данных, поэтому можно утверждать, что исходные данные имеют достаточно сложную структуру и двух главных компонент не достаточно для решения поставленной задачи бинарной классификации.

В. VGG-Face как классификатор

VGG-Face также можно использовать как полноценный классификатор, благодаря архитектуре модели. Для этого количество выходных значений было сокращено до двух, в соответствии с количеством имеющихся классов. Полученная модель нуждалась в дообучении. Датасет был разделен на три подвыборки – обучающую, тестовую и валидационную в соотношении 3:1:1. Обучение проводилось в течении 50 эпох, полученные метрики приведены в Таблице 3.

Таблица III. ТАБЛИЦА 3. МЕТРИКИ FINE-TUNED МОДЕЛИ

Точность	F1-score	AUC-ROC
0,72	0,71	0,82

По сравнению с результатами предыдущих способов, fine-tuned модель не дала прироста в качестве ни по одному из показателей. Из графика функции потерь (приведен в полной версии статьи) видно, что после 30-ой эпохи val loss существенно не изменяется. Воспользуемся этим фактом при анализе модели на кросс-валидации.

7. КРОСС-ВАЛИДАЦИЯ

Для объективного оценивания и сравнения моделей между собой, была проведена кросс-валидация на 5 частях с взятием среднего арифметического для получившихся метрик. Результаты кросс-валидации представлены в таблице 4.

Таблица IV. МЕТРИКИ НА КРОСС-ВАЛИДАЦИИ

	Accuracy	F1	AUC-ROC
<i>Facenet</i>			
SVM(C=1)	0,76	0,78	0,85
SVM(C=0.3)	0,77	0,81	0,85
<i>VGG-Face</i>			
SVM(C=1)	0,77	0,78	0,83
SVM(C=0.3)	0,78	0,79	0,83
Fine-tuned	0,74	0,75	0,84

Относительно высокое значение метрик показывает, что модели научились извлекать информативные признаки из изображений и классифицировать людей с ВО. Также по результатам кросс-валидации видно, что замена модели нейронной сети не дает существенного улучшения в качестве классификации. Поэтому в будущих исследованиях стоит обратить больше внимания на улучшение качества датасета.

8. ЗАКЛЮЧЕНИЕ

Полученные значения метрик классификации показывают, что модель научилась находить закономерности между фотографиями лиц людей и наличием у них ВО. Следовательно, поставленную частную гипотезу «о возможности определения наличия высшего образования по фотографии лица с помощью методов машинного обучения» можно считать подтвержденной. Но стоит заметить, что выборка является небольшой, из-за чего результаты классификации могут иметь смещение относительно незначимых признаков. Поэтому стоит провести дополнительное исследование на датасете большего размера.

Рассмотренные в данной работе классификаторы могут быть использованы для таргетированной рекламы при построении рекомендательных систем, для автоматизации проведения социологических и иных исследований. Также работа способствует расширению понимания, какие данные о человеке могут стать публичными без его ведома благодаря технологиям машинного обучения.

ЛИТЕРАТУРА

- [1] Hassin, R. Facing faces: studies on the cognitive aspects of physiognomy / R. Hassin, Y. Trope // Journal of personality and social psychology. – 2000. – Vol. 78(5). – P. 837-852.
- [2] Wu, X. Automated inference on criminality using face images. / X. Wu, X. Zhang // arXiv preprint arXiv:1611.04135. – 2016. – P. 4038-4052.
- [3] Wang, Y. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images / Y. Wang, M. Kosinski // Journal of personality and social psychology. – 2018. – Vol. 114(2). – P. 246-257.
- [4] Kosinski, M. Facial recognition technology can expose political orientation from naturalistic facial images / M. Kosinski // Nature Scientific reports. – 2021. – Vol. 11(1). – P. 1-7.
- [5] Bharat, Y.C. A comparative analysis of face recognition models on masked faces / Y.C. Bharat, G.R. Karthikeya // International Journal Of Scientific & Technology Research. – 2020. – Vol. 9(10). – P. 175–178.
- [6] Schroff, F. FaceNet: A unified embedding for face recognition and clustering / F. Schroff, D. Kalenichenko, J. Philbin // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 815-823.
- [7] Visual Geometry Group – University of Oxford [Electronic resource]. — Access mode: https://www.robots.ox.ac.uk/~vgg/software/vgg_face (14.03.2022).
- [8] Спасенников, Б.А. Социально-правовая характеристика осужденных, отбывающих наказание в следственных изоляторах / Б.А. Спасенников, С.А. Лаврычева // Пенитенциарная наука. – 2016. – Т. 1, № 33. – С. 17-22
- [9] Parvez, Z.F. The labor of pleasure: How perceptions of emotional labor impact women's enjoyment of pornography. / Z.F. Parvez // Gender & Society. – 2006. – Vol. 20(5). – P. 605-631.