УДК 004.932

Л. Р. Широкова, В. Н. Логинов

Московский физико-технический институт (национальный исследовательский университет)

Анализ эффективности архитектур нейронных сетей для детекции Replay Attack в системах лицевой биометрии

Рассматриваются вопросы построения эффективной архитектуры нейросети для распознавания спуффинг-атак на систему лицевой биометрии, основанных на подмене в поле зрения камеры видеонаблюдения лица реального человека на видеоизображение лица другого человека, сформированного на экране носимого устройства. Проведен сравнительный анализ подходов к построению нейросетевых архитектур. Получены оценки метрик качества для каждого подхода.

Ключевые слова: система биометрии, видеоизображение лица.

L. R. Shirokova, V. N. Loginov

Moscow Institute of Physics and Technology

Analysis of the effectiveness of neural network architectures for Replay Attack detection in the facial biometrics system

The paper deals with the construction of an effective neural network architecture for recognizing spoofing attacks on the facial biometrics system based on the substitution of a real person's face in the field of view of a video surveillance camera for a video image of another person's face formed on the screen of a wearable device. A comparative analysis of approaches to the construction of neural network architectures is carried out. Estimates of quality metrics for each approach are obtained.

Key words: biometrics system, facial vibraimage.

1. Введение

В работе [1] рассматривается метод детекции подмен в видеопотоке системы лицевой биометрии изображения лица человека на изображение лица другого человека, сформированное на экране носимого устройства. В литературе данный вид подмен называют спуфинг-атаками типа Replay Attack. В ней предложен метод, основанный на формировании матриц, элементы которых соответствуют межкадровым изменениям в видеопотоке камеры наблюдения биометрический системы, и последующей обработке результатов анализа предобученной нейронной сетью.

Последовательность обработки видеоданных разделяется на два этапа:

1) Предварительная обработка видеоизображения – подготовка входных данных (матриц межкадровых разностей) для нейронной сети.

2) Обработка входных данных предобученной нейронной сетью и предсказание вероятности того, что видеоизображение передает реальное лицо.

При предварительной обработке данных к последовательности видеокадров применяется частотный метод построения матриц межкадровых разностей, предложенный в работе [2]. В работе [1] были получены предварительные оценки эффективности распознавания Replay Attack на одном из типов нейросетей, однако, в связи с высокими требованиями к уровню информационной безопасности биометрических систем, продолжает оставаться актуальным вопрос построения нейронных сетей, обеспечивающих наиболее высокую вероятность распознавания Replay Attack. В связи с этим, в настоящей работе приводятся результаты сравнительного анализа различных подходов к построению архитектур нейросетей, позволяющие выработать рекомендации по их практическому применению при решении задачи.

2. Нейросетевые модели

С точки зрения машинного обучения рассматривается задача бинарной классификации изображения. Соответственно, на вход подается трехмерная матрица изображения, на выход ожидается 1 число (номер класса). Подавать в нейросеть пиксели в случайном порядке нельзя, так как структура данных – изображение, подразумевает фиксированный порядок элементов в матрице. Использовать знания о структуре данных важно в любой задаче компьютерного зрения, поэтому с изображениями работают нейросети со сверточными слоями.

Существует много известных архитектур сверточный нейронных сетей для задачи классификации: AlexNet, VGG, ResNet, Inception, SqueezeNet, и др. Также возможно создать свою сеть из любого набора слоев. При проведении экспериментов были рассмотрены все перечисленные сети, в статье описаны 4 сети, которые показали устойчивые результаты при тестировании.

Итак, в настоящей работе рассмотрено 4 архитектуры нейронных сетей для классификации изображений: AlexNet [3], VGG-16 [4], SqueezeNet [5], CNN. Каждая сеть обучается на тренировочной выборке видеоизображений. Выход каждой сети соответствует необходимому решению задачи – предсказанию двух чисел: первое число – вероятность принадлежности к 0-му классу (видеоизображение лица человека, которое было получено путем съемки с экрана устройства), второе число – вероятность принадлежности к 1-му классу (видеоизображение реального лица человека). Оценка метрик качества происходит на одной и той же отложенной тестовой выборке видеоизображений.

Функцией потерь обучения сети является бинарная кросс-энтропия, оптимизатором – стохастический градиентный спуск.

2.1. VGG-16

Первой нейросетью для экспериментов была выбрана сеть VGG-16 [ссылка на статью] на фреймворке Keras.

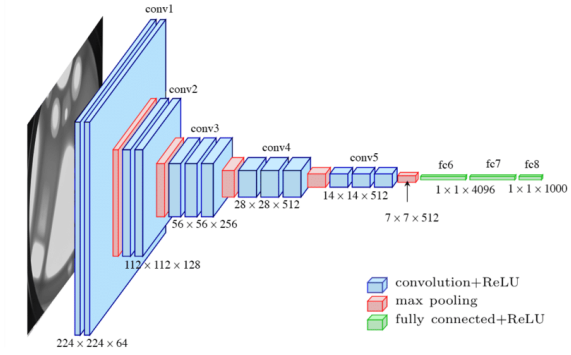

Рис. 1. Архитектура VGG-16

Для того, чтобы решать задачу классификации, последние 4 слоя нейронной сети были заменены на следующие слои: max pooling, fully connected (1024), fully connected (128), fully connected (2).

Эксперименты проводились на сети, которая была предобучена на данных ImageNet [6]. В ходе экспериментов дообучались только верхние, измененные слои, веса нижних слоев при обучении не изменялись.

2.2. AlexNet

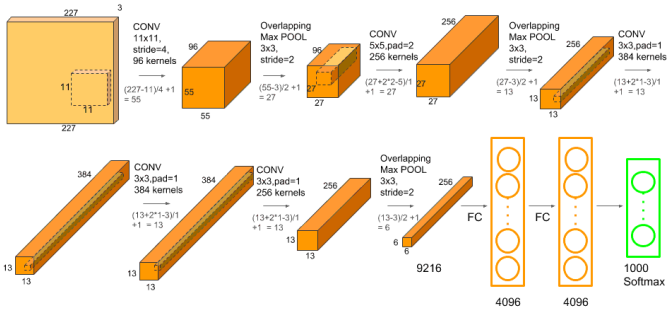

Рис. 2. Архитектура AlexNet

При обучении сети был добавлен выходной слой fully connected (2). Нейронная сеть также была предобучена на датасете ImageNet. Сеть AlexNet взята из реализации библиотеки torchvision фреймворка pytorch. Модель обучалась полностью.

2.3. SqueezeNet

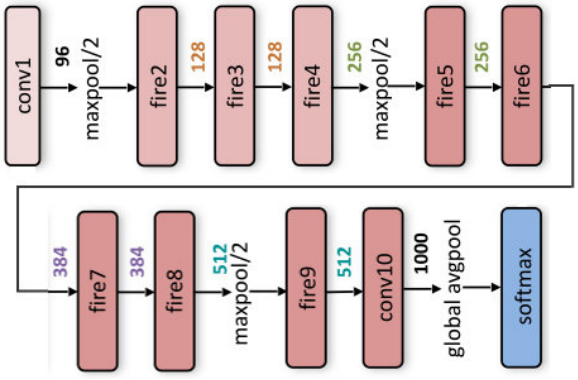

Рис. 3. Архитектура SqueezeNet

Также в экспериментах использовалась нейронная сеть SqueezeNet предобученная на ImageNet из библиотеки torchvision. При обучении сети был добавлен выходной слой fully connected (2). Модель обучалась полностью.

2.4. CNN

В данном разделе эксперименты проведены с нейронной сетью со следующим набором слоев из фреймворка pytorch.

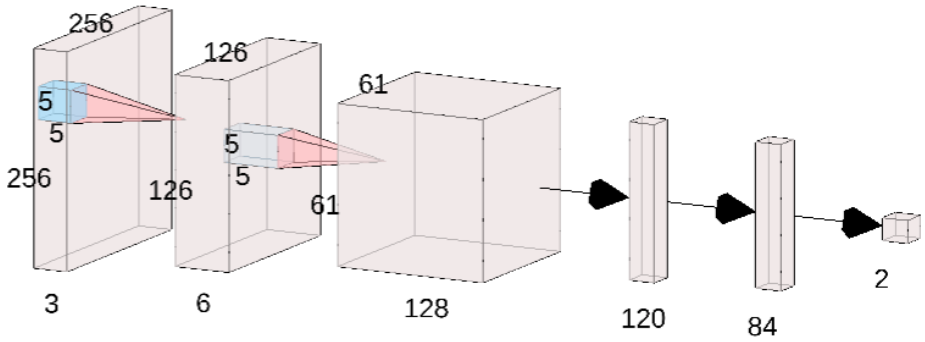

Рис. 4. CNN

Сеть состоит из следующих слоев: Conv2D, MaxPooling2D, Conv2D, MaxPooling2D, fully connected, fully connected.

2.5. Сравнение результатов предсказания моделей

В табл. 1 приведено сравнение 4-х представленных в данной работе архитектур нейронных сетей. Сравнение проводится по метрике F_1:

$$F_1 = \frac{2 \times precision \times recall}{precision + recall}.$$

Т а б л и ц а 1

Сравнение метрики качества различных архитектур

	F_1 Позитивные примеры	F_1 Негативные примеры
VGG-16	0.940	0.945
AlexNet	0.917	0.929
SquezeeNet	0.96	0.963
CNN	0.946	0.951

Анализ представленных данных позволяет сделать вывод о том, что все рассматриваемые модели обеспечивают приемлемое качество распознавания спуфинг-атак. Из таблицы также видно, что качество распознавания спуфинг-атак при использовании сети SqueezeNet дает наилучшие результаты.

3. Заключение

Проведено сравнение четырех различных архитектур нейронных сетей применительно к задаче распознавания спуфинг-атак подмены лица человека на изображении экрана цифрового носителя. Нейронные сети обучены на кадрах видеопотока, которые были преобразованы с помощью матричного подхода с частотной формулой [1]. Архитектура SqueezeNet достигает самых высоких показателей качества среди рассматриваемых архитектур на тестовых данных.

Литература

1. *Широкова Л.Р., Логинов В.Н.* Нейросетевой метод детекции видеоизображения лица в видеопотоке системы лицевой биометрии // Труды МФТИ. 2020. Т. 12, № 4. С. 90–96.
2. *Минкин В.А.* ВиброИзображение. Санкт-Петербург : Реноме, 2007.

3. *Krizhevsky A.* One weird trick for parallelizing convolutional neural networks. arXiv preprint arXiv:1404.5997. 2014.

4. *Simonyan K., Zisserman A.* Very Deep Convolutional Networks for Large-Scale Image Recognition. eprint arXiv:1409.1556. 2014.

5. *Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360v4. 2016.

6. *Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.* Imagenet: A large-scale hierarchical image database // 2009 IEEE conference on computer vision and pattern recognition. 2009. P. 248–55.

References

1. *Shirokova L.R., Loginov V.N.* Neural network method for detecting video images of a person in a video stream of a facial biometrics system. Proceedings of MIPT. 2020. V. 12, N 4. P. 90–96.

2. *Minkin V.A.* VibroImage. Sankt-Peterburg : Renome, 2007.

3. *Krizhevsky A.* One weird trick for parallelizing convolutional neural networks. arXiv preprint arXiv:1404.5997. 2014.

4. *Simonyan K., Zisserman A.* Very Deep Convolutional Networks for Large-Scale Image Recognition. eprint arXiv:1409.1556. 2014.

5. *Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360v4. 2016.

6. *Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.* Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009. P. 248–55.

Поступила в редакцию 27.01.2021