

Одноэтапный детектор лиц и особых точек на цифровых изображениях

Ю.В. Визильтер¹, В.С. Горбацевич¹, А.С. Моисеенко^{1,2}

¹ Государственный научно-исследовательский институт авиационных систем (ГосНИИАС),
125167, Россия, г. Москва, ул. Викторенко, д. 7,

² Московский физико-технический институт (государственный университет),
141707, Россия, г. Долгопрудный, Институтский пер., д. 9

Аннотация

Поиск особых точек лица является важной подзадачей в решении ряда задач лицевой биометрии. В системах распознавания лиц построение биометрического шаблона происходит по предварительно выровненному (нормализованному) изображению лица, этап нормализации включает в себя задачу поиска основных точек лица. Актуальной в подобной задаче является проблема баланса между качеством работы детектора особых точек и скоростью. В данной статье предложен одноэтапный детектор лиц и особых точек на базе глубоких конволюционных нейронных сетей, работающий в режиме реального времени и достигающий высокого качества на ряде известных тестовых выборок (AFLW2000, COFW, Mepro2D). Предлагаемый детектор лиц и особых точек основан на идее одноэтапного детектора объектов SSD, зарекомендовавшего себя как алгоритм, обеспечивающий высокую скорость работы и высокое качество обнаружения объектов. В качестве базовой архитектуры глубоких конволюционных нейронных сетей используется сеть ShuffleNet V2. Важной особенностью предлагаемого алгоритма является то, что обнаружение лица на изображении и поиск ключевых точек делается за один проход глубоких конволюционных нейронных сетей, что позволяет значительно экономить время на этапе внедрения. Также подобная многозадачность позволяет снизить процент ошибок в задаче поиска особых точек, что позитивно сказывается на качестве работы итогового алгоритма распознавания лиц.

Ключевые слова: биометрия, обнаружение лиц, поиск особых точек лица, SSD.

Цитирование: Визильтер, Ю.В. Одноэтапный детектор лиц и особых точек на цифровых изображениях / Ю.В. Визильтер, В.С. Горбацевич, А.С. Моисеенко // Компьютерная оптика. – 2020. – Т. 44, № 4. – С. 589-595. – DOI: 10.18287/2412-6179-CO-674.

Citation: Vizilter YV, Gorbachevich VS, Moiseenko AS. Single-shot face and landmarks detector. Computer Optics 2020; 44(4): 589-595. DOI: 10.18287/2412-6179-CO-674.

Введение

В статье предлагается детектор лиц и особых точек, основанный на идее одноэтапного детектора объектов SSD, [1], [2].

Решение задачи распознавания лиц находит широкое применение в различных сферах человеческой деятельности: системы видеонаблюдения, охранные комплексы, криминалистическая экспертиза, системы виртуальной реальности, системы идентификации и верификации личности в мобильных устройствах и многие другие. На сегодняшний день самые передовые системы распознавания лиц построены на базе глубоких конволюционных нейронных сетей (ГКНС) [3], [4], показавших результаты, превосходящие результаты человека-эксперта в решении ряда задач компьютерного зрения.

При распознавании лиц в большинстве случаев происходит сравнение биометрических шаблонов (векторов признаков, характеризующих цифровое изображение лица). Построение биометрического шаблона подразумевает построение эффективных векторов признаков для решения, в зависимости от постановки, задачи идентификации или верификации лиц. Помимо собственно решения самой задачи идентификации или верификации, особую роль играет

предварительное обнаружение лица на изображении, а также поиск основных точек лица.

Особые точки необходимы на этапе нормализации лиц перед построением биометрического шаблона. Предварительная нормализация изображения лица позволяет значительно повысить качество работы алгоритма распознавания за счет универсализации обрабатываемых изображений. Под особыми точками подразумеваются точки, соответствующие определенным частям лица. Обычно для нормализации выделяют 5 основных точек: центры зрачков, центр носа и уголки рта.

Часто на этапе обнаружения лиц, помимо непосредственно обнаружения лица (выделения описывающего прямоугольника), происходит также обнаружение особых точек. Существует ряд широко известных алгоритмов, позволяющих одновременно находить как описывающий прямоугольник лица, так и непосредственно особые точки. К таким алгоритмам, например, относятся каскадный алгоритм MTCNN [5] и одноэтапный детектор лиц и особых точек RetinaFace [6]. Однако таких решений, показывающих высокий результат в режиме реального времени, на сегодняшний день довольно мало. В большинстве решений сначала производится предварительное обнаружение лица на изображении с ис-

пользованием, как правило, нейросетевых детекторов лиц, и только после этого на основании найденного описывающего прямоугольника кадрированное изображение лица подается на вход еще одной сети для поиска особых точек.

На сегодняшний день все детекторы объектов на базе ГКНС можно условно разделить на 2 типа: одноэтапные (single-shot, proposal-free) детекторы и двухэтапные (two-stage, region-based). Одноэтапные алгоритмы подразумевают использование одной глубокой конволюционной нейронной сети, выходом которой являются непосредственно искомые прямоугольники и метки классов. К одноэтапным детекторам объектов относятся такие алгоритмы, как SSD [1], DSOD [7], RetinaNet [8]. В двухэтапных алгоритмах обнаружение и распознавание проводится в два этапа. На первом этапе специализированная ГКНС формирует набор гипотез о местоположении объектов, затем на втором этапе другая ГКНС проводит финальное распознавание объектов и доуточнение их положения. Известны такие двухэтапные детекторы объектов, как: R-CNN [9], Fast-RCNN [10], Faster-RCNN [11], R-FCN [12] и другие. Первую группу детекторов объектов отличает высокая скорость работы, однако одноэтапные детекторы уступают по качеству обнаружения детекторам второго типа. Так как в системах распознавания лиц приоритетной является скорость работы сети, предлагаемый детектор лиц и особых точек строится на базе одноэтапного детектора объектов SSD с использованием в качестве базовой сети ShuffleNet V2 [13].

В данной статье предлагается детектор лиц и особых точек, работающий в режиме реального времени и достигающий высокого качества на известных тестовых выборках (AFLW2000, COFW, Метро2D). Важной особенностью предлагаемого детектора является то, что обнаружение лиц и особых точек производится за один проход ГКНС, т.е. для поиска особых точек не используется дополнительная ГКНС.

1. Одноэтапный детектор лиц и особых точек

1.1. Архитектура сети

Общая идея всех алгоритмов семейства SSD заключается в том, что поиск областей и классификация происходят за один проход, в отличие от двухэтапных детекторов объектов семейства R-CNN. Для обнаружения объектов разных масштабов используются глубокие карты признаков разных конволюционных слоев. Структуру сети SSD можно условно разделить на две части: основная подсеть для извлечения глубоких карт признаков и подсеть для прогнозирования, на основе карт признаков различных масштабов, местонахождения целевых объектов. Заключительные сверточные слои сети SSD объединяют предсказания по всем масштабам карт признаков, по-

сле чего применяется алгоритм подавления немаксимумов (NMS) для формирования окончательных гипотез о местонахождении искомого объекта.

Формирование итоговых описывающих прямоугольников объектов происходит с использованием предварительно заданных опорных прямоугольников (анкеры) [11]. Используя сверточные слои, подсеть для прогнозирования местоположения объектов предсказывает смещение координат прогнозируемого описывающего прямоугольника относительно центра заданного анкера для каждой ячейки карты признаков, а также показатель уверенности найденного класса объекта. Прогнозирование смещений вместо прогнозирования напрямую координат значительно упрощает задачу и облегчает обучение сети.

На рис. 1 приведен пример входного изображения (справа) и полученной карты признаков размером 5×5 (слева). На приведенном изображении карты признаков в одной из ее ячеек приведен пример опорного прямоугольника (анкера) с соотношением сторон 1:1 (пунктирная линия). Красным цветом показан соответствующий описывающий прямоугольник, смещение которого относительно заданного анкера предсказывает сеть. В нашей постановке задачи, помимо собственно смещения координат описывающего прямоугольника (координаты верхнего левого угла и нижнего правого), происходит также прогнозирование смещения особых точек (координаты левого глаза, правого глаза, носа, левого уголка рта, правого уголка рта) относительно центра опорного прямоугольника.

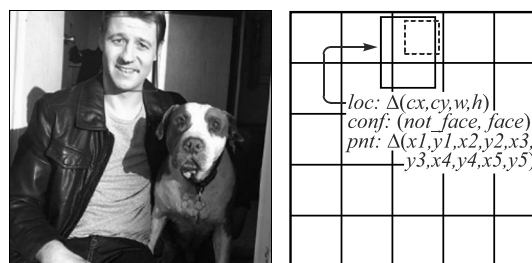


Рис. 1. Пример изображения с визуализацией разметки описывающего прямоугольника и особых точек, а также пример карты признаков размера 5×5 с визуализацией одного из опорных прямоугольников и соответствующего ему описывающего прямоугольника

Выбор опорного положительного (содержащего объект) прямоугольника, относительно которого считается смещение прогнозируемого описывающего прямоугольника, происходит с использованием метрики «пересечение над объединением» (intersection over union, IoU). В данной статье выбор положительных опорных прямоугольников осуществляется аналогично алгоритму S3FD [14]. На первом этапе положительными считаются анкеры, для которых показатель метрики IoU с эталонным описывающим прямоугольником превышает 0,35. На втором этапе из оставшихся выбираются анкеры, проходящие по порогу 0,1 согласно

метрике IoU. Если количество отобранных на втором этапе анкеров превышает количество анкеров, отобранных на первом этапе, выбираются анкеры с наибольшим значением IoU в количестве, равном числу выбранных на первом этапе анкеров.

Для обнаружения лиц на входных изображениях размера 640×640 пикселей применяются карты признаков трех различных масштабов. В нижней части основной сети находятся карты признаков, предназначенные для обработки лиц наименьшего масштаба на изображении. Масштаб обрабатываемых лиц увеличивается от уровня к уровню.

Первая карта признаков разработанной сети имеет размер 20×20 и представляет собой выход последнего слоя четвертой стадии базовой сети ShuffleNet V2. Вторая и третья карта признаков имеют размер 10×10 и 5×5 соответственно и представляют собой блок понижения пространственной размерности сети ShuffleNet V2. В сети используются опорные прямо-

угольники (анкеры) с соотношением сторон 1:1, что позволяет сети быстрее сходиться к целевым значениям описывающего прямоугольника лица. Архитектура сети в общем виде представлена на рис. 2.

1.2. Мультизадачная функция потерь

Общая функция потерь представляет собой сумму трех независимых функций потерь (функция потерь локализации, функция потерь особых точек и функция потерь классификации для двух классов (лицо / не лицо)):

$$L = \frac{1}{N} \left(\sum_i L_{cls}(p_i, \hat{p}_i) + \sum_i \hat{p}_i L_{loc}(bb_i, \hat{bb}_i) + \sum_i \hat{p}_i L_{pnt}(pnt_i, \hat{pnt}_i) \right),$$

N – число выбранных опорных прямоугольников (анкеров),

i – индекс опорного прямоугольника,

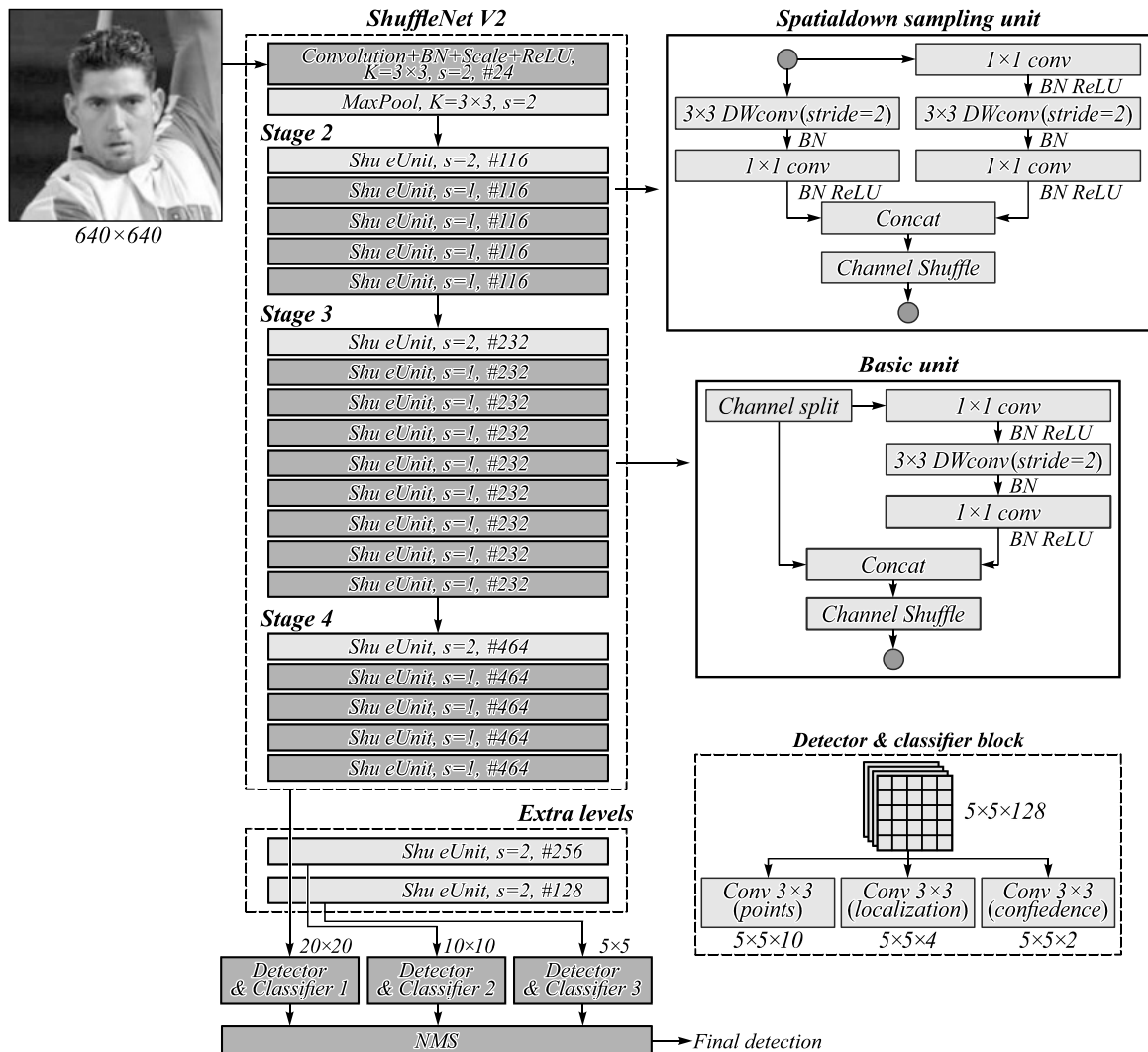


Рис. 2. Архитектура сети

p_i – предсказанная вероятность того, что i -й опорный прямоугольник принадлежит классу лица,

\hat{p}_i – бинарная метка класса, принимает значение 1 для положительных опорных прямоугольников (ли-

цо) и 0 для отрицательных опорных прямоугольников (не лицо),

L_{loc} – функция потерь Smooth L1, рассчитанная только для положительных опорных прямоугольников, на вход подаются предсказанный описывающий прямоугольник \hat{bb}_i и эталонный (размеченный) описывающий прямоугольник \widehat{bb}_i (регрессия смещения относительно центра опорных прямоугольников),

L_{pnt} – функция потерь Smooth L1, вычисляемая только для положительных опорных прямоугольников, на вход подаются предсказанные особые точки pnt_i и эталонные (размеченные) особые точки \widehat{pnt}_i (регрессия смещения относительно центра опорных прямоугольников),

L_{conf} – функция потерь Softmax для двух классов (лицо / не лицо).

1.3. Обучающая выборка

При обучении ГКНС крайне важен выбор обучающей выборки, при этом особую важность имеет количество изображений в выборке. Для обучения детектора лиц и детектора особых точек требуется база изображений, содержащая одновременно координаты описывающего прямоугольника лица и координаты особых точек.

Для обучения был использован набор данных WFLW. База WFLW содержит 10 000 лиц (7 500 для обучения и 2 500 для тестирования) с разметкой 98 особых точек. Поскольку для нормализации достаточно всего 5 ключевых точек, использовались не все особые точки, которые представлены в разметке WFLW. Набор данных WFLW содержит лица в различных условиях съемки, включая разнообразные сложные ракурсы, условия освещенности, элементы маскировки.

Чтобы расширить набор данных, при обучении к исходным изображениям применяется ряд геометрических и цветовых преобразований, в том числе случайное кадрирование, отражение относительно вертикали, изменение контраста, насыщенности изображения, добавление шума.

1.4. Параметры обучения

При обучении веса базовой сети ShuffleNet V2 берутся из предобученной на базе ImageNet сети. Обучение сети производится методом Adam с начальным $\text{lr}=0,001$; $\beta_1=0,9$; $\beta_2=0,999$; регуляризирующее слагаемое – 0,0005; размер батча – 96 на 8 картах NVIDIA GTX1080.

2. Результаты экспериментов

2.1. Методика тестирования

В качестве основной метрики качества в задаче поиска особых точек используется средняя ошибка, нормализованная на геометрическое среднее ширины и высоты описывающего прямоугольника лица:

$$NME = \frac{1}{N} \sum_{k=1}^N \frac{\|y_k - \widehat{y}_k\|}{d},$$

где: N – количество искомым особым точек, d – среднее геометрическое ширины и высоты истинного описывающего прямоугольника лица (\sqrt{wh}), \widehat{y}_k – координаты (x, y) особой точки, найденные ГКНС, y_k – координаты (x, y) истинной особой точки (разметка).

В данной статье для более полной оценки предлагаемого алгоритма значение средней ошибки рассчитано как в целом для всех 5 особых точек, так и для каждой особой точки отдельно.

Для оценки алгоритмов поиска особых точек в сочетании со средней ошибкой обычно указывается показатель средней частоты отказов (failure rate). Если евклидово расстояние между предсказанной и истинной особой точкой превышает значение, равное 10% от размера лица, вычисленного как геометрическое среднее высоты и ширины описывающего прямоугольника лица, то считается, что данная особая точка найдена с ошибкой. Мы вычисляем среднюю частоту отказов (показатель failure rate) как процент ключевых точек, которые были предсказаны ошибочно. Также для более наглядной оценки работы алгоритма поиска особых точек строится кумулятивная кривая ошибок.

2.2. Оценка результатов

Тестирование проводилось на ряде публичных баз для тестирования качества детекторов особых точек: AFLW2000, COFW и Menpo2D. Тестовый набор данных AFLW2000 содержит 2000 изображений, полученных из основного набора данных AFLW путем кадрирования до размера 450×450 пикселей. Изображения включают лица в различных условиях съемки. Тестовый набор данных COFW содержит 507 лиц. Тестовый набор данных Menpo2D содержит 5335 изображений лиц.

Результаты работы предлагаемого детектора лиц и особых точек, в сравнении с двумя известными алгоритмами для одновременного обнаружения лиц и поиска особых точек (MTCNN и RetinaFace), приведены на рис. 3.

Таким образом, предлагаемый алгоритм превосходит на всех тестовых выборках алгоритм поиска лиц и особых точек MTCNN. А также сравним или превосходит по качеству алгоритм RetinaFace с MobileNet в качестве основной сети.

Также было протестировано качество работы детектора лиц на известной публичной тестовой выборке FDDB, содержащей 2845 изображений 5171 лица в различных условиях съемки. Результат работы детектора приведен на рисунке. Предлагаемый детектор имеет высокий показатель качества согласно ROC-кривой, 95,8% при количестве ложных срабатываний, равном 1000 (рис. 4). Пример работы детектора лиц и особых точек приведен на рис. 5.

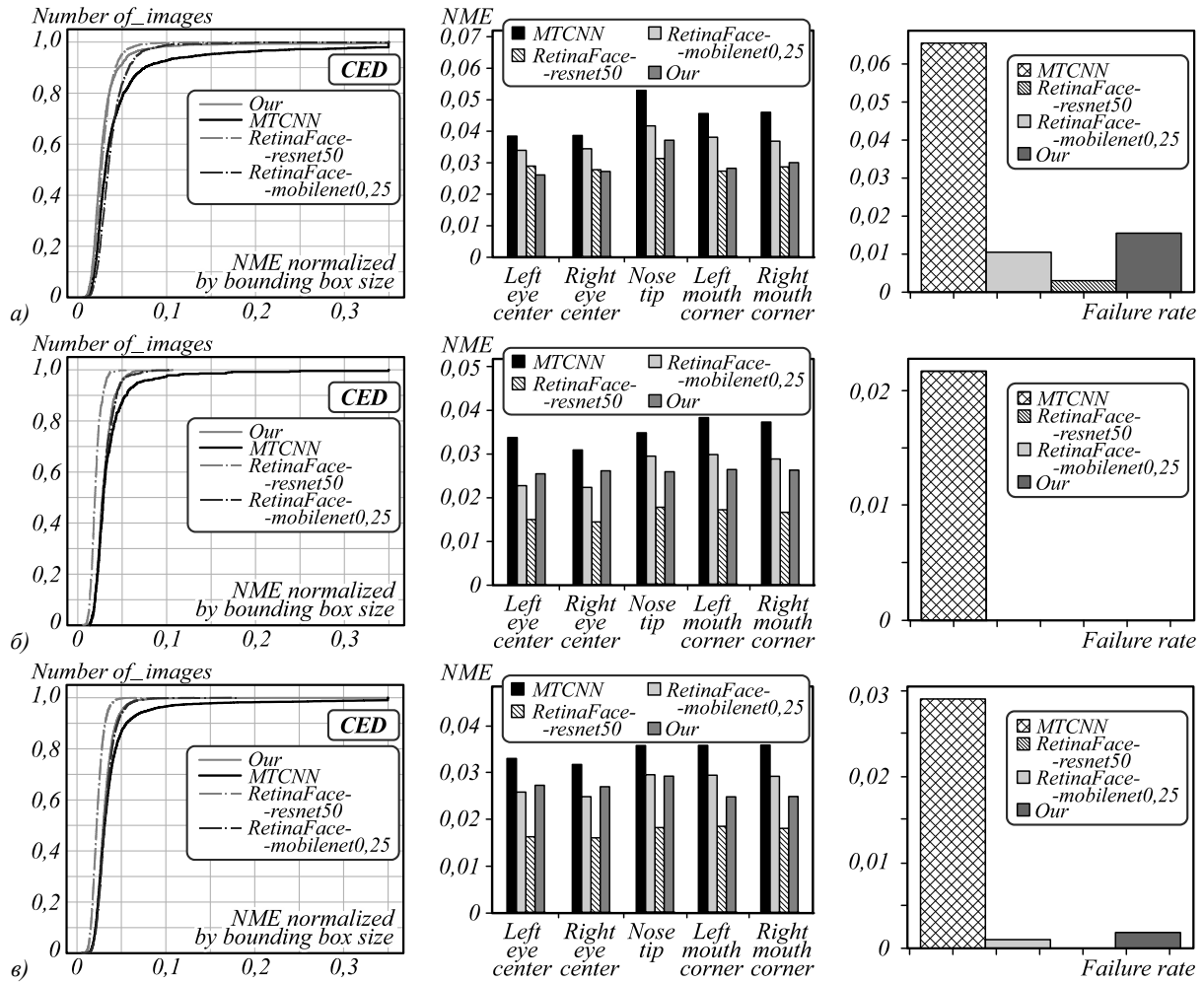


Рис. 3. Сравнение на тестовой выборке результатов работы алгоритмов обнаружения лиц и особых точек: AFLW2000 (а), COFW (б), Menpo2D (в) (кумулятивная кривая ошибок, гистограмма распределения ошибок по каждой из 5 искомых особых точек, средняя частота отказов)

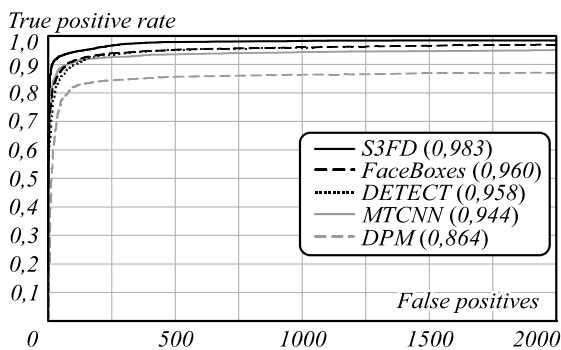


Рис. 4. ROC-кривая ошибок на тестовом наборе данных FDDB

2.3. Время работы алгоритма

Одной из особенностей предлагаемого алгоритма является скорость его работы. В табл. 1 представлено среднее время обработки одного изображения формата 640×640.

При этом следует отметить, что тестирование проводилось при размере батча, равном 1, так как это соответствует практике применения алгоритмов обнаружения лиц.

Заключение

Таким образом, был разработан детектор лиц и особых точек, работающий в режиме реального времени, позволяющий достигать высокого качества на широко известных тестовых выборках для оценки алгоритмов поиска особых точек, сравнимый, а по ряду метрик превосходящий лидирующие алгоритмы поиска особых точек. Отличительной особенностью предлагаемого алгоритма является скорость работы, а также то, что алгоритм позволяет находить лица и особые точки одновременно за один проход ГКНС.

Алгоритм позволяет обрабатывать до 50 кадров в секунду на CPU (в один поток) и более 350 кадров в секунду на GPU (при использовании батча, равного 1) и, таким образом, является одним из наиболее быстрых алгоритмов обнаружения лиц на текущий момент. Границы применимости данного алгоритма могут быть расширены за счет введения дополнительных возможностей нахождения бинарных лицевых атрибутов (пол, эмоции, наличие очков, усов, бороды). Это является предметом дальнейших исследований.



Рис. 5. Пример работы предложенного детектора лиц и особых точек на изображениях из тестовой выборки WFLW

Табл. 1. Среднее время обработки

Аппаратная платформа	Среднее время обработки предлагаемого алгоритма	Среднее время обработки МТСNN (для минимального размера лица 16×16)
CPU Intel Core i5 8500 Pytorch batch =1	108 мс	240 мс (многопоточный) / 700 мс (один поток)
CPU Intel Core i5 8500 Intel OpenVino batch =1 1 поток	20 мс	
GPU RTX 2080Ti Pytorch batch =1	10 мс	40 мс
GPU RTX 2080Ti TensorRT batch =1	3,21 мс	
GPU RTX 2080Ti fp16 TensorRT batch =1	2,85 мс	

Благодарности

Эта работа была поддержана грантом РФФИ 19-07-01146 А.

Литература

- Liu, W.** SSD: Single shot multibox detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg // ECCV. – 2016. – P. 21-37.
- Jiao, L.** A survey of deep learning-based object detection / L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, R. Qu // IEEE Access. – 2019. – Vol. 7. – P. 128837-128868.
- Deng, J.** ArcFace: Additive angular margin loss for deep face recognition / J. Deng, J. Guo, N. Xue, S. Zafeiriou // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 4685-4694.

- Liu, W.** SphereFace: Deep hypersphere embedding for face recognition / W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 6738-6746.
- Zhang, K.** Joint face detection and alignment using multi-task cascaded convolutional networks / K. Zhang, Z. Zhang, Z. Li, Y. Qiao // IEEE Signal Processing Letters. – 2016. – Vol. 23, Issue 10. – P. 1499-1503.
- Deng, J.** RetinaFace: Single-stage dense face localisation in the wild [Electronical Resource] / J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, S. Zafeiriou. – 2019. – URL: <https://arxiv.org/abs/1905.00641> (request date 25.12.2019).
- Shen, Z.** DSOD: Learning deeply supervised object detectors from scratch / Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, X. Xue // 2017 IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 1937-1945.
- Lin, T.-Y.** Focal loss for dense object detection / T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár // 2017 IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2999-3007.
- Girshick, R.** Rich feature hierarchies for accurate object detection and semantic segmentation / R. Girshick, J. Donahue, T. Darrell, J. Malik // Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – P. 580-587.
- Girshick, R.** Fast R-CNN / R. Girshick // 2015 IEEE International Conference on Computer Vision (ICCV). – 2015. – P. 1440-1448.
- Ren, S.** Faster R-CNN: Towards real-time object detection with region proposal networks / S. Ren, K. He, R. Girshick, J. Sun // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2015. – Vol. 39, Issue 6. – P. 1137-1149.
- Lin, Y.** Feature pyramid networks for object detection / Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017 – DOI: 10.1109/CVPR.2017.106.
- Ma, N.** ShuffleNet V2: Practical guidelines for efficient CNN architecture design / N. Ma, X. Zhang, H.-T. Zheng, J. Sun // European Conference on Computer Vision. – 2018. – P. 122-138.

14. Zhang, S. S3FD: Single shot scale-invariant face detector / S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, S.Z. Li // International Conference on Computer Vision. – 2017. – P. 192-201.

Сведения об авторах

Визильтер Юрий Валентинович, 1970 года рождения, в 1992 году окончил ФГБОУ ВПО «Московский авиационный институт (национальный исследовательский университет)». В 1997 г. защитил кандидатскую, а в 2009 году – докторскую диссертацию. В настоящее время является начальником подразделения ФГУП «ГосНИИ Авиационных систем» (ГосНИИАС). Опубликовал более 80 научных работ (из них более 20 статей в рецензируемых журналах из списка ВАК). Область научных интересов: обработка и анализ изображений, цифровая фотограмметрия, машинное зрение, математическая морфология, распознавание образов, машинное обучение, биометрия. E-mail: viz@gosniias.ru.

Горбацевич Владимир Сергеевич, 1985 года рождения, в 2009 году окончил ФГБОУ ВПО «Московский авиационный институт (национальный исследовательский университет)». В настоящее время является начальником лаборатории ФГУП «ГосНИИ Авиационных систем» (ГосНИИАС). Опубликовал более 20 научных работ (из них более 6 статей в рецензируемых журналах из списка ВАК). Область научных интересов: обработка и анализ изображений, цифровая фотограмметрия, машинное зрение, математическая морфология, распознавание образов, машинное обучение, биометрия. E-mail: gvs@gosniias.ru.

Моисеенко Анастасия Сергеевна, 1993 года рождения, в 2017 года окончила Московский физико-технический институт. В настоящее время работает инженером в ФГУП «ГосНИИ Авиационных систем» (ГосНИИАС). Область научных интересов: обработка и анализ изображений, компьютерное зрение, машинное обучение, биометрия, обнаружения объектов, распознавание лиц. E-mail: moiseenko.as@phystech.edu.

ГРНТИ: 28.23.15

Поступила в редакцию 9 декабря 2019 г. Окончательный вариант – 30 апреля 2020 г.

Single-shot face and landmarks detector

Y.V. Vizilter¹, V.S. Gorbatshevich¹, A.S. Moiseenko^{1,2}

¹State Research Institute of Aviation Systems (GosNIAS), Moscow, Russia,

²Moscow Institute of Physics and Technology (MIPT), Moscow, Russia

Abstract

Facial landmark detection is an important sub-task in solving a number of biometric facial recognition tasks. In face recognition systems, the construction of a biometric template occurs according to a previously aligned (normalized) face image and the normalization stage includes the task of finding facial keypoints. A balance between quality and speed of the facial keypoints detector is important in such a problem. This article proposes a CNN-based one-stage detector of faces and keypoints operating in real time and achieving high quality on a number of well-known test datasets (such as AFLW2000, COFW, Menpo2D). The proposed face and facial landmarks detector is based on the idea of a one-stage SSD object detector, which has established itself as an algorithm that provides high speed and high quality in object detection task. As a basic CNN architecture, we used the ShuffleNet V2 network. An important feature of the proposed algorithm is that the face and facial keypoint detection is done in one CNN forward pass, which can significantly save time at the implementation stage. Also, such multitasking allows one to reduce the percentage of errors in the facial keypoints detection task, which positively affects the final face recognition algorithm quality.

Keywords: biometry, face detection, CNN, landmarks detection, SSD.

Citation: Vizilter YV, Gorbatshevich VS, Moiseenko AS. Single-shot face and landmarks detector. *Computer Optics* 2020; 44(4): 589-595. DOI: 10.18287/2412-6179-CO-674.

Acknowledgements: This work was financially supported by the Russian Foundation for Basic Research (Project 19-07-01146 A).

References

- [1] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. SSD: Single shot multibox detector. *ECCV* 2016: 21-37.
- [2] Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z, Qu R. A survey of deep learning-based object detection. *IEEE Access* 2019; 7: 128837-128868.
- [3] Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. *CVPR* 2019: 4685-4694.
- [4] Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. SphereFace: Deep hypersphere embedding for face recognition. *CVPR* 2017: 6738-6746.
- [5] Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process Lett* 2016; 23(10): 1499-1503.
- [6] Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: Single-stage dense face localisation in the wild. Source: (<https://arxiv.org/abs/1905.00641>).
- [7] Shen Z, Liu Z, Li J, Jiang Y-G, Chen Y, Xue X. DSOD: Learning deeply supervised object detectors from scratch. *ICCV* 2017: 1937-1945.
- [8] Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *ICCV* 2017: 2999-3007.
- [9] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 2014: 580-587.
- [10] Girshick R. Fast R-CNN. *ICCV* 2015: 1440-1448.
- [11] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2015; 39(6): 1137-1149.
- [12] Lin Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. *CVPR* 2017. DOI: 10.1109/CVPR.2017.106.
- [13] Ma N, Zhang X, Zheng H-T, Sun J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *ECCV* 2018: 122-138.
- [14] Zhang S, Zhu X, Lei Z, Shi H, Wang X, Li SZ. S3FD: Single shot scale-invariant face detector. *ICCV* 2017: 192-201.

Authors' information

Yuri Valentinovich Vizilter (b.1970) graduated from Moscow Aviation Institute (National Research University) in 1992. Since 1997 is the Candidates of Technical Sciences, 2009 – the Doctor of Technical Sciences. Currently he works as the head of department at the FGUP “GosNIAS”. Author of 80 scientific papers. Research interests are processing and image analysis, digital photogrammetry, computer vision, mathematical morphology, pattern recognition, machine learning, biometry. E-mail: viz@gosnias.ru.

Vladimir Sergeevich Gorbatshevich (b.1985) graduated from Moscow Aviation Institute (National Research University) in 2009. Currently he works as the head of laboratory at the FGUP “GosNIAS”. Author of 20 scientific papers. Research interests are processing and image analysis, digital photogrammetry, computer vision, mathematical morphology, pattern recognition, machine learning, biometry. E-mail: gvs@gosnias.ru.

Anastasiia Sergeevna Moiseenko (b.1993) graduated from Moscow Institute of Physics and Technology in 2017. Currently she works as the engineer at the FGUP “GosNIIAS”. Research interests are processing and image analysis, computer vision, machine learning, biometry. E-mail: moiseenko.as@phystech.edu.

Received Decamber 9, 2019. The final version – April 30, 2020.
