

Подходы к распознаванию областей лица для телемедицинских приложений

Н. С. Данишевский¹, Д. О. Буданов¹, А. Ю. Зайцева²

¹Санкт-Петербургский политехнический университет Петра Великого

²Институт аналитического приборостроения РАН

danishevskiy.ns@gmail.com

Аннотация. В настоящее время перспективным направлением исследований являются телемедицинские системы, способные осуществлять мониторинг физиологических параметров человека. Такие системы обычно используются для неинвазивной экспресс-диагностики патологических состояний. Для получения и обработки необходимой информации используются алгоритмы компьютерного зрения и машинного обучения.

Данная публикация посвящена рассмотрению наиболее перспективных методов для решения задачи распознавания областей лица. В основе методов лежит использование свёрточных нейронных сетей. В качестве модели машинного обучения в данной работе используется архитектура MobileNet v3. Представлены результаты обучения двух конфигураций выбранной нейронной сети. Приведено сравнение параметров обучения, которые позволяют получить наибольшую точность. Наиболее значимыми из рассмотренных параметров являются функции ошибки. Проведён анализ полученных результатов.

Ключевые слова: компьютерное зрение, телемедицина, машинное обучение, физиологические параметры

I. ВВЕДЕНИЕ

В настоящее время телемедицинские системы получают все большее распространение и, как следствие, являются перспективным направлением исследований. Это обусловлено тем, что они способны обеспечивать мониторинг физиологических параметров человека. В связи с чем, такие системы часто применяются на различных производствах с целью снижения количества внештатных ситуаций, так как позволяют производить диагностику состояния работника без отвлечения его внимания.

В подобных телемедицинских системах часто применяется фотоплетизмография – методика, которая позволяет с помощью оптических методов отслеживать изменение наполненности сосудов кровью. Из данной информации могут быть получены сведения, например, о работе кровеносной системы обследуемого.

Однако классические методы фотоплетизмографии требуют непосредственного контакта между исследуемой областью организма и оптическим датчиком. Это накладывает ограничения на движения человека и условия проведения обследования. По этой причине всё больший интерес в научной сфере приобретает методика дистанционной фотоплетизмографии [1]. Такая методика использует видеокамеру в качестве оптической системы. В течение промежутка времени фиксируется изменение тона кожи участка тела человека. Эти изменения определяются посредством анализа изменения значений пикселей на видеоизображении. В качестве основных инструментов в

таких системах используются алгоритмы компьютерного зрения и машинного обучения.

Основными этапами работы алгоритма дистанционной фотоплетизмографии являются детектирование лица человека и выделение регионов интереса на лице человека. Далее в течении определённого отрезка времени производится измерение изменения цвета пикселей в найденных регионах интереса. В результате чего получают переменный во времени сигнал, из которого могут быть получены значения жизненно важных показателей тела человека, например, частоты сердечных сокращений, насыщения крови кислородом и другие. В настоящей работе решается задача распознавания регионов интереса на лице человека.

II. СУЩЕСТВУЮЩИЕ МЕТОДЫ ДЕТЕКТИРОВАНИЯ РЕГИОНОВ ИНТЕРЕСА НА ЛИЦЕ ЧЕЛОВЕКА

Наиболее информативными регионами интереса на лице человека с точки зрения измерения параметров организма являются области лба и щёк. Для выделения этих регионов предлагается решить задачу детектирования ключевых точек лица. Рис. 1 содержит пример ключевых точек лица, которые необходимо детектировать: контур лица, глаза, брови, нос, губы. Следующий шаг при таком подходе – геометрическое определение регионов интереса.

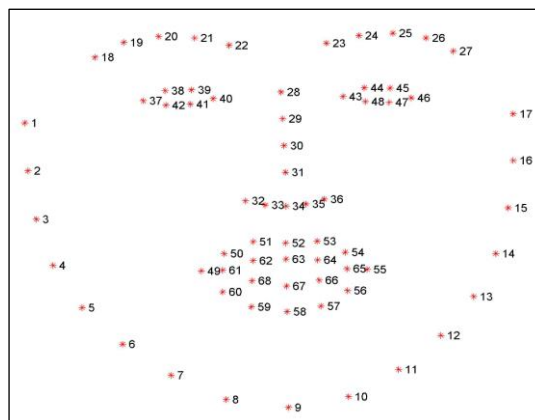


Рис. 1. Ключевые точки на лице [2]

Для осуществления поиска точек лица на изображении используются методы компьютерного зрения. Результатом работы алгоритма является вектор, представляющий собой набор координат каждой точки $LM = \{x_1, y_1, x_2, y_2, \dots, x_n, y_n\}$. Структура лица человека является сложной и индивидуальной. Положение и формы выделенных ключевых контуров довольно вариативны и зависят от положения лица или эмоции на нём. В связи с этим, необходимо реализовать алгоритм,

который обеспечит достаточную обобщённость, но при этом сможет точно определять координаты нужных точек. Наиболее подходящим решением данной задачи будет использование методов машинного обучения.

Алгоритм нахождения координат точек на изображении основан на применении регрессии координат [3]. Регрессия координат представляет собой преобразование признаков, полученных на выходе нейронной сети, в координаты с применением полносвязного слоя в качестве окончательного [3], [4]. В некоторых исследованиях полносвязный окончательный слой нейронной сети заменялся на декодер в виде функции Argmax или soft-Argmax [5]. При таком подходе нейронная сеть обучается на тепловых картах. Такой подход называется регрессией тепловых карт. Возможно применение сложных слоёв для комбинации и выделения лучших сторон перечисленных основных подходов [6]. Для получения признаков с помощью нейронной сети чаще всего используют архитектуру кодер-декодер. Такой подход чаще всего используется при регрессии тепловых карт. К тому же возможно использование последовательного соединения из нескольких таких блоков для увеличения точности работы нейронной сети.

III. ПРЕДЛАГАЕМЫЙ МЕТОД ДЕТЕКТИРОВАНИЯ РЕГИОНОВ ИНТЕРЕСА

В данной статье предлагается использовать простую архитектуру кодера и выходного полносвязного слоя. Приоритетной целью данной работы является реализация наиболее быстрого алгоритма, но не обязательно самого точного. Данная цель определяется тем, что система, в которой будет использоваться данный алгоритм, предполагает обработку ситуаций, в которых лицо человека будет находиться прямо перед камерой и ничем не перекрываться. В качестве свёрточных блоков предлагается использовать блоки MobileNet v3 [7]. Архитектура MobileNet характеризуется малыми вычислительными затратами при средней точности работы.

В данной работе исследуются две конфигурации модели нейронной сети – полная модель и упрощённая модель. Основное отличие между конфигурациями – количество тренируемых параметров. Таблица I и Таблица II содержат архитектуры нейронных сетей.

ТАБЛИЦА I. АРХИТЕКТУРА ПОЛНОЙ МОДЕЛИ

Входная размерность	Операция	Количество каналов	Strides
112x112x3	Conv2d	16	2
56x56x24	Bottleneck, 3x3	24	1
56x56x24	Bottleneck, 5x5	40	2
28x28x40	Bottleneck, 5x5	40	1
28x28x40	Bottleneck, 3x3	40	1
28x28x40	Bottleneck, 3x3	80	2
14x14x80	Bottleneck, 3x3	80	1
14x14x80	Bottleneck, 3x3	80	1
14x14x80	Bottleneck, 3x3	80	1
14x14x80	Bottleneck, 3x3	112	1
14x14x112	Bottleneck, 3x3	112	1
14x14x112	Bottleneck, 5x5	160	2
7x7x160	Bottleneck, 5x5	160	1
7x7x160	Bottleneck, 5x5	160	1
7x7x160	Conv2d, 1x1	960	1
7x7x960	Pool, 7x7	-	1
1x1x960	Conv2d 1x1, NBN	1280	1
1x1x1280	Conv2d 1x1, NBN	136	1

ТАБЛИЦА II. АРХИТЕКТУРА УПРОЩЕННОЙ МОДЕЛИ

Входная размерность	Операция	Количество каналов	Strides
112x112x3	Conv2d, 3x3	16	2
56x56x16	Bottleneck, 3x3	24	2
28x28x24	Bottleneck, 3x3	24	1
28x28x24	Bottleneck, 5x5	40	2
14x14x40	Bottleneck, 5x5	40	1
14x14x40	Bottleneck, 5x5	40	1
14x14x40	Bottleneck, 5x5	48	1
14x14x48	Bottleneck, 5x5	48	1
14x14x48	Bottleneck, 5x5	96	2
7x7x96	Bottleneck, 5x5	96	1
7x7x96	Bottleneck, 5x5	96	1
7x7x96	Conv2d, 1x1	576	1
7x7x576	Pool, 7x7	-	1
1x1x576	Conv2d 1x1, NBN	1024	1
1x1x1024	Conv2d 1x1, NBN	136	1

IV. ФУНКЦИИ ОШИБКИ

Функции ошибки играют ключевую роль при обучении нейронных сетей. Далее рассматриваются наиболее часто используемые функции для решения задачи детектирования ключевых точек лица человека.

A. L1-Loss

Функция ошибки L1-Loss представляет собой абсолютную ошибку и вычисляется согласно выражению (1), где y_{true} – истинное значение искомой величины, $y_{predicted}$ – значение искомой величины, которое сформировала нейронная сеть. При расчёте этой ошибки за несколько батчей вычисляется средняя абсолютная ошибка.

$$L_1 = |y_{true} - y_{predicted}| \quad (1)$$

Данная функция используется для решения задачи регрессии. Она хорошо подходит для случаев, когда присутствуют резко отличающиеся значения определяемых величин – «выбросы». Отрицательной стороной при использовании данной функции является плохое обучение при наличии маленьких ошибок, вызванное линейностью данной функции. Таким образом, модель будет хорошо обучаться при больших ошибках и плохо при малых.

B. L2-Loss

Функция ошибки L2-Loss является квадратичной и рассчитывается согласно выражению (2), где y_{true} – истинное значение искомой величины, $y_{predicted}$ – значение искомой величины, которое сформировала нейронная сеть. При расчёте этой ошибки за несколько батчей вычисляется средняя квадратичная ошибка.

$$L_2 = (y_{true} - y_{predicted})^2 \quad (2)$$

В отличие от L1-Loss данная функция фокусирует внимание нейронной сети на снижение больших ошибок и, в то же время, заставляет её игнорировать малые ошибки. Недостатком данной функции ошибки является то, что она легко подвержена влиянию «выбросов». В связи с чем, возрастают требования к качеству набора обучающих данных.

C. Wing Loss

Функция Wing Loss [9] определяется выражениями (3) и (4). Данная функция позволяет выделить

положительные стороны функций ошибок L1-Loss и логарифмической функции.

$$W = \begin{cases} \omega \cdot \ln \left(1 + \frac{|y_{\text{true}} - y_{\text{predicted}}|}{\varepsilon} \right), npru|x| < \omega \\ |y_{\text{true}} - y_{\text{predicted}}| - C, npru|x| \geq \omega \end{cases} \quad (3)$$

$$C = \omega - \omega \cdot \ln(1 + \omega / \varepsilon) \quad (4)$$

Так как функция L1-Loss устойчива к «выбросам», то она хорошо учитывает влияние ошибок большого размера. Для учёта ошибок меньшего размера используется логарифмическая функция. Ввиду того, что производная логарифма равна $1/x$, уменьшение ошибки увеличивает значение функции.

Таким образом, функция Wing Loss представляет собой логарифмическую функцию для ошибок малого размера. В области больших ошибок она ведёт себя как функция L1-Loss. Коэффициент ω задаёт интервал, на котором функция будет иметь логарифмический характер. Постоянная C используется для гладкого соединения нелинейной и линейной составляющих функции.

D. Adaptive Wing Loss

Функция ошибки Adaptive Wing [10] является улучшенной версией функции Wing Loss. Функция Adaptive Wing Loss описывается выражениями (5)–(7). Введены дополнительные константы α , θ и A .

$$AW = \begin{cases} \omega \cdot \ln \left(1 + \left| \frac{y_{\text{true}} - y_{\text{predicted}}}{\varepsilon} \right|^{\alpha - y_{\text{true}}} \right), npru|x| < \omega \\ A \cdot |y_{\text{true}} - y_{\text{predicted}}| - C, npru|x| \geq \omega \end{cases} \quad (5)$$

$$A = \left(\frac{1}{1 + (\theta / \varepsilon)^{\alpha - y_{\text{true}}}} \right) \left(\frac{\theta}{\varepsilon} \right)^{\alpha - y_{\text{true}} - 1} \frac{\omega \cdot (\alpha - y_{\text{true}})}{\varepsilon} \quad (6)$$

$$C = \theta A - \omega \cdot \ln \left(1 + (\theta / \varepsilon)^{\alpha - y_{\text{true}}} \right) \quad (7)$$

Основным отличием данной функции от Wing Loss является другое поведение в области наименьших ошибок. Значение функции ошибки Wing Loss в области малых ошибок сильно возрастает, что заставляет модель игнорировать большие и средние ошибки. Для того, чтобы обеспечить более распределённое влияние ошибок различной величины, Adaptive Wing Loss ведёт себя в области ошибок, близких к нулю, примерно так же, как функция ошибки L2 (среднеквадратичная ошибка). Другими словами, снижается влияние малых ошибок на обучение модели.

Таким образом, при обучении модель сначала фокусируется на уменьшении больших ошибок, которые чаще всего вызваны наиболее важными особенностями объекта. Далее, когда значения этих ошибок снижаются, модель начинает больше внимания уделять ошибкам, вызванным второстепенными особенностями, которые в

функции Wing Loss перекрывались. Как следствие, функция ошибки подстраивается в процессе обучения под текущие особенности модели.

V. РЕАЛИЗАЦИЯ И ТЕСТИРОВАНИЕ МЕТОДА ДЕТЕКТИРОВАНИЯ РЕГИОНОВ ИНТЕРЕСА

Обучение моделей машинного обучения проведено с использованием набора данных 300W [2], который содержит несколько тысяч изображений лиц в различных условиях и при различных позах. Была проведена аугментация данных – исходные изображения обрезались по ограничивающему прямоугольнику, которым ограничивается положение лица на изображении. В результате аугментации был получен набор данных, каждое изображение в котором содержит исключительно одно лицо человека и не содержит никаких дополнительных предметов. При обучении использовались изображения размером 112x112. Итоговый набор данных состоит из 5 710 изображений для обучения и 956 изображений для валидации.

Длительность обучения составила 30 эпох с шагом обучения 0,0001 и размером батча 32. Таблица III и Таблица IV содержат результаты обучения моделей. В качестве метрики для оценки точности детектирования использовалось среднее нормированное отклонение (8).

$$NME = \frac{|y_{\text{true}} - y_{\text{predicted}}|}{y_{\text{true}}} \quad (8)$$

ТАБЛИЦА III. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ МОДЕЛЕЙ

Функция ошибки	Полная модель	Упрощённая модель
	NME, %	NME, %
L1-loss	6,3	6,5
L2-loss	10,8	10,7
Wing loss	4,3	13,0
Adaptive Wing Loss	15,3	15,8

ТАБЛИЦА IV. СРАВНЕНИЕ ХАРАКТЕРИСТИК МОДЕЛЕЙ

Модель	Количество параметров, 10^3	$t_{\text{выч}}$, мс	NME_{min} , %	Тип функции ошибки
Упрощённая	2,1	15	6,5	L1 Loss
Полная	4,3	25	4,3	Wing Loss

Согласно полученным результатам, наихудшей функцией ошибки среди рассмотренных для решения задачи нахождения ключевых точек лица оказалась функция Adaptive Wing Loss. Упрощённая модель лучше всего обучается при использовании функции ошибки L1-Loss. Это может свидетельствовать о том, что усиление влияния малых ошибок при использовании других функций плохо сказывается на обучаемости нейронной сети данной конфигурации. Для полной модели лучше всего подходит функция ошибки Wing Loss. Можно утверждать, что данная функция обеспечивает сбалансированное влияние ошибок разной величины.

При сравнении конфигураций моделей друг с другом видно, что полная модель обеспечивает большую точность. Она оказалось на 33,8 % точнее относительно упрощённой модели. Однако в то же время, полная модель проигрывает упрощённой – порядка 60 % в относительной скорости вычислений. Результаты детектирования ключевых точек лица представлены на Рис. 2. Визуально можно оценить ошибки упрощённой модели при определении контура лица и области бровей.

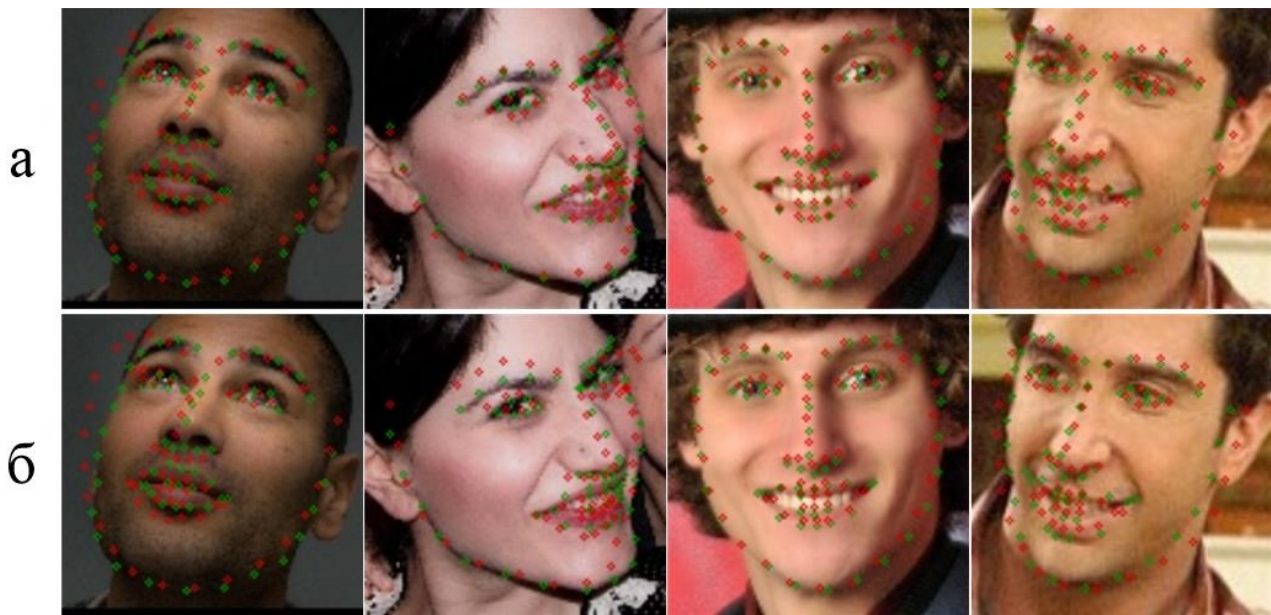


Рис. 2. Результаты детектирования двух реализованных моделей, а – результаты работы полной модели, б – результаты работы упрощённой модели

Однако стоит отметить, что такая важная область, как нос, определяется с одинаковой точностью. Результат работы алгоритма выделения регионов интереса на основе предложенного метода с использованием модели полной конфигурации представлен на Рис. 3. В алгоритме используются всего пять точек – координаты глаз, носа и щёк.

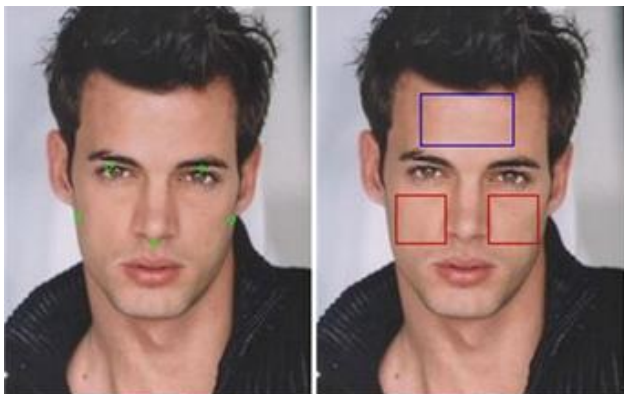


Рис. 3. Выделенные регионы интереса на лице человека

VI. ЗАКЛЮЧЕНИЕ

В результате выполнения работы были рассмотрены основные методы решения задачи детектирования ключевых точек лица человека. Был предложен метод, характеризующийся малыми вычислительными затратами, основанный на применении архитектуры Mobile Net v3. Реализованы и обучены две конфигурации модели предложенной архитектуры. При обучении использовались несколько самых популярных функций ошибки для решения задачи детектирования ключевых точек лица. Результаты показали, что наибольшая точность достигается при использовании функции ошибки Wing Loss при обучении моделей с большим числом параметров. Предложенный метод позволяет производить детектирование ключевых точек лица со средней абсолютной ошибкой 4,3 %. Приведена иллюстрация работы метода выделения регионов

интереса при интеграции его в алгоритм дистанционной фотоплетизмографии.

СПИСОК ЛИТЕРАТУРЫ

- [1] Dasari A., Prakash S.K.A., Jeni L.A., Tucker C.S. Evaluation of biases in remote photoplethysmography methods. *Npj Digital Medicine*. vol. 4, no. 4, pp. 1-13, 2021, DOI: 10.1038/s41746-021-00462-z.
- [2] Sagonas C., Antonakos E., Tzimiropoulos G., Zafeiriou S., Pantic M. 300 Faces In-The-Wild Challenge: database and results. *Image and Vision Computing*. vol. 47, pp. 3-18, March 2016, DOI: 10.1016/j.imavis.2016.01.002
- [3] Valle R., Buenaposada J.M., Baumela L. Cascade of Encoder-Decoder CNNs with Learned Coordinates Regressor for Robust Facial Landmarks Detection. *Pattern Recognition Letters*. October 2019, pp. 1-11, DOI: 10.1016/j.patrec.2019.10.012.
- [4] Browatzki B., Wallraven C. 3FabRec: Fast Few-Shot Face Alignment by Reconstruction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 2020, pp. 6110-6120, DOI: 10.1109/CVPR42600.2020.00615.
- [5] Kang S., Lee M., Kim M., Shim H. HybridMatch: Semi-Supervised Facial Landmark Detection via Hybrid Heatmap Representations. *IEEE Access*, January 2023, pp. 1-11, DOI: 10.1109/ACCESS.2023.3257180.
- [6] Nibali A., He Z., Morgan S., Prendergast L. Numerical Coordinate Regression with Convolutional Neural Networks. *Computer Science*. January 2018, pp. 1-10, DOI: arXiv:1801.07372v2.
- [7] Howard A., Sandler M., Chu G., Chen L.C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q. V., Adam H. Searching for MobileNetV3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 2019, pp. 1314-1324, DOI: 10.1109/ICCV.2019.00140.
- [8] Guo X., Li S., Yu J., Zhang J., Ma J., Ma L., Liu W., Ling H. PFLD: A Practical Facial Landmark Detector. *Arxiv Computer Vision and Pattern Recognition*. February 2019, pp. 1-11, DOI: arXiv:1902.10859v2.
- [9] Feng Z.H., Kittler J., Awais M., Huber P., Wu X.J. Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 18-22 Jun 2018, USA, pp. 2235-2245, DOI: 10.1109/CVPR.2018.00238.
- [10] Wang X., Bo L. and Fuxin L. Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South), 2019, pp. 6970-6980, DOI: 10.1109/ICCV.2019.00707.