

СИСТЕМА ПОИСКА, ВЫДЕЛЕНИЯ И РАСПОЗНАВАНИЯ ЛИЦ НА ИЗОБРАЖЕНИЯХ

А. А. Друки

Институт кибернетики Национального исследовательского
Томского политехнического университета, 634034, Томск, Россия

УДК 004.931

Для решения задачи выделения и распознавания лиц использованы сверточные нейронные сети. Представлена структура разработанной нейронной сети. Создан алгоритм масштабирования и кластеризации изображений.

Ключевые слова: обработка изображений, нейронные сети, алгоритм масштабирования и кластеризации.

To solve the problem of isolation and recognition of individuals the convolutional neural networks are used. The structure of the developed neural network is presented. The algorithm for scaling and clustering of images is presented.

Key words: image recognition, neural networks, image recognition, algorithm for scaling and clustering.

Введение. Многие направления науки, техники и производства в значительной степени ориентированы на развитие систем, в которых информация имеет характер поля (изображения). При обработке такой информации возникает ряд сложных научных, технических и технологических проблем, одной из которых является обработка и распознавание изображений. О важности этой проблемы свидетельствует тот факт, что исследования по распознаванию образов, анализу изображений и речи включены в перечень приоритетных направлений развития науки и техники и критических технологий федерального уровня [1].

Распознавание изображений находит применение в различных приложениях: контроль топологии печатных плат, текстуры ткани, контроль доступа к информации по идентификации личности (биометрическая идентификация), доступ к объектам ограниченного доступа, оперативный поиск в картотеке изображений, дактилоскопия и др. [1].

Широкое распространение получают биометрические системы идентификации человека. В отличие от традиционных систем идентификации, требующих знания пароля, наличия ключа, идентификационной карточки либо иного идентифицирующего предмета, который можно забыть или потерять, биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые позволяют однозначно определить конкретного человека. К числу таких характеристик относятся отпечатки пальцев, форма ладони, узор радужной оболочки, изображение сетчатки глаза, индивидуальные характеристики лица [1–3].

Для решения задачи распознавания лиц используются различные методики, среди которых можно выделить подходы, основанные на нейронных сетях [1–4], разложении Кархунена — Лоэва [4], алгебраических моментах [5], линиях одинаковой интенсивности, эластичных (деформируемых) эталонах сравнения.

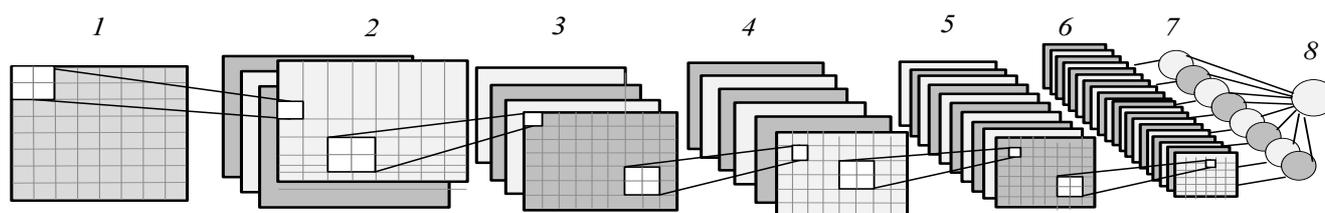


Рис. 1. Архитектура сверточной нейронной сети:
 1 — вход; 2, 4, 6 — сверточные слои; 3, 5 — подвыборочные слои;
 7 — слой из обычных нейронов; 8 — выход

Анализ методов выделения сюжетной части изображения показал, что для решения данной задачи эффективным является использование искусственных нейронных сетей, поскольку они обеспечивают возможность получения классификатора, хорошо моделирующего сложную функцию распределения изображений лиц, тем самым увеличивая точность решения по сравнению с решениями, получаемыми другими методами.

Преимущество использования нейронных сетей для обнаружения лица — обучаемость системы для выделения ключевых характеристик лица из учебных наборов [6, 7].

Объектом исследования в данной работе являются системы обработки изображений, основанные на искусственных нейронных сетях. Предметом исследования являются алгоритмы выделения сюжетной части и распознавание образов на групповом изображении, основанные на искусственных нейронных сетях.

Разработка нейронной сети. Наиболее часто в задачах распознавания и идентификации изображений используются классические нейросетевые архитектуры (многослойный перцептрон, сети с радиально-базисной функцией и др. [6]), но применение классических нейросетевых архитектур в задачах распознавания имеет ряд недостатков:

- изображения имеют большую размерность, соответственно возрастает размер нейронной сети;
- большое количество параметров увеличивает вместимость системы и соответственно требует большего объема тренировочной выборки, увеличивает время и вычислительную сложность процесса обучения;
- для повышения эффективности работы системы желательно применять несколько нейронных сетей (обученных с различными начальными значениями синаптических коэффициентов и порядком предъявления образов), что увеличивает вычислительную сложность решения задачи и время его выполнения;
- отсутствует инвариантность к изменениям масштаба изображения, ракурсов съемки камеры и других геометрических искажений входного сигнала [8–10].

Для решения данной задачи были выбраны сверточные нейронные сети, обеспечивающие частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям [10].

Архитектура сверточной нейронной сети включает большое количество чередующихся слоев двух типов: сверточных и подвыборочных (рис. 1).

Нейроны в пределах слоя организованы в плоскости. В каждом слое имеется набор из нескольких плоскостей, причем нейроны одной плоскости имеют одинаковые синаптические коэффициенты, ведущие ко всем локальным участкам предыдущего слоя. Каждый нейрон слоя получает входы от некоторой области предыдущего слоя (локальное рецептивное поле), т. е. входное изображение предыдущего слоя как бы сканируется небольшим окном

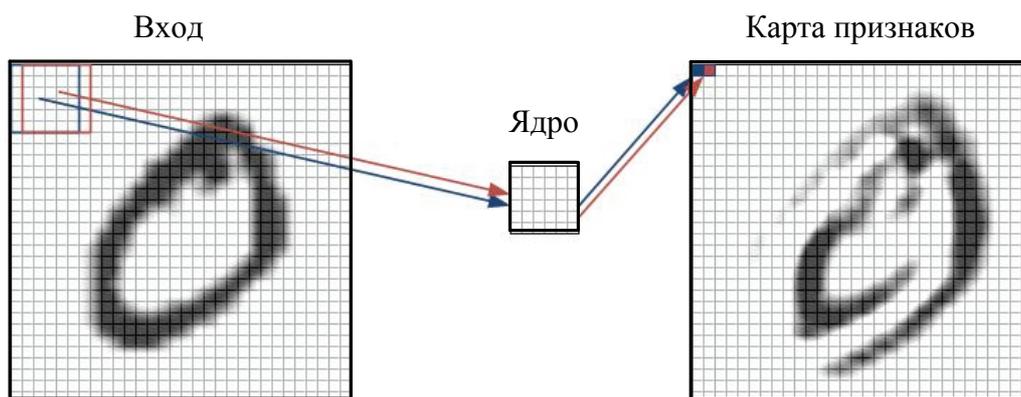


Рис. 2. Результат сканирования изображения

и пропускается сквозь набор синаптических коэффициентов, а результат отображается на соответствующий нейрон текущего слоя. Набор плоскостей представляет собой карты характеристик, и каждая плоскость находит “свои” участки изображения в любом месте предыдущего слоя. Размер локального рецептивного поля выбирается самостоятельно в процессе разработки нейронной сети.

Подвыборочный слой уменьшает масштаб плоскостей путем локального усреднения значений выходов нейронов. Таким образом достигается иерархическая организация. Последующие слои извлекают более общие характеристики, слабее зависящие от искажений изображения [10]. На рис. 2 показан результат работы сверточной нейронной сети.

Постепенно нейронная сеть обучается выделять ключевые характеристики лиц в поступающих на вход изображениях; отклики нейронной сети образуют максимумы в местоположениях лиц [10, 11]. Результаты исследования показали, что при сканировании входного изображения отклики нейронной сети в местоположениях лиц и в некоторой их окрестности образуют максимумы. На рис. 3 показан отклик нейронной сети в каждом пикселе изображения. В некоторых местах отклики достигают максимумов, следовательно, в этих областях расположены информативные участки-кандидаты на местоположение искомого лица.

Анализ откликов нейронной сети на наборе изображений позволил оценить порог обнаружения на уровне 0,8.

Разработана нейронная сеть, состоящая из шести слоев. Входными данными нейронной сети являются изображения размером 32×36 пикселей.

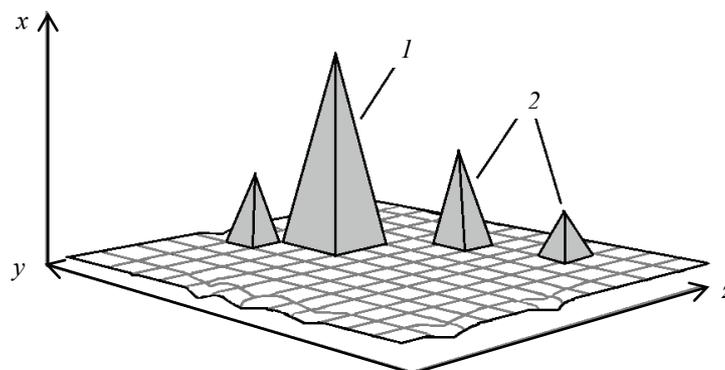


Рис. 3. Отклики нейронной сети:

- 1 — наиболее вероятное местонахождение лица на изображении;
- 2 — наименее вероятное нахождение лица на изображении (шумы)

Так как задача, решаемая нейронной сетью, — классификация, то достаточно одного выхода в структуре нейронной сети. Выходное значение нейронной сети находится в интервале $[-1; +1]$, что соответственно означает отсутствие или присутствие лица на классифицируемом изображении [10, 11].

За входным слоем располагается сверточный слой $C1$. Этот слой состоит из пяти сверточных плоскостей, каждая из которых имеет синаптическую маску размером 5×5 и нейронное смещение.

Размер сверточной плоскости определяется в соответствии с выражениями

$$w_c = w_u - K + 1, \quad h_c = h_u - K + 1,$$

где w_c , h_c — ширина и высота сверточной плоскости соответственно; w_u , h_u — ширина и высота плоскости предыдущего слоя; K — ширина (высота) окна сканирования.

Размер плоскости сверточного слоя $C1$ — 28×32 нейрона, количество синаптических коэффициентов в слое — 130.

Подвыборочный слой $S1$ состоит из пяти карт характеристик и обеспечивает локальное усреднение и подвыборку. Каждая плоскость слоя $S1$ связана лишь с одной плоскостью слоя $C1$. Размер каждой плоскости слоя $S1$ — 14×16 нейронов, что в два раза меньше размера плоскости предыдущего слоя. Каждая плоскость слоя $S1$ имеет единственный синаптический коэффициент и нейронное смещение, что дает в итоге 10 настраиваемых параметров.

Сверточный слой $C2$ состоит из 20 плоскостей. Размер плоскости — 12×14 нейронов. Слой содержит 290 синаптических коэффициентов. Плоскости слоя $C2$ формируются следующим образом: каждая из пяти плоскостей слоя $S1$ свернута двумя различными синаптическими масками размером 3×3 , обеспечивая 10 плоскостей в $C2$, другие 10 плоскостей $C2$ получены путем суммирования результатов двух свертываний на каждой возможной паре плоскостей слоя $S1$. Таким образом, сети добавляется способность объединять различные виды особенностей, чтобы составлять новые, менее зависящие от искажений входного изображения.

Слой $S2$ состоит из 20 плоскостей (по одной на каждую плоскость слоя $C2$), размер каждой плоскости составляет 6×7 нейронов. Каждая плоскость слоя $S2$ имеет единственный синаптический коэффициент и нейронное смещение, что дает в итоге 40 настраиваемых параметров.

В слое $N1$ находится 20 нейронов (по одному на каждую плоскость слоя $S2$), каждый нейрон полностью связан с каждым нейроном только одной плоскости слоя $S2$, он выполняет взвешенное суммирование своих 42 входов, добавляет нейронное смещение и пропускает результат через активационную функцию. Таким образом, данный слой содержит 860 синаптических коэффициентов.

Единственный нейрон слоя $N2$ полностью связан со всеми нейронами слоя $N1$. Роль этого нейрона — вычисление окончательного результата классификации.

Использование принципа объединения синаптических коэффициентов дает эффект уменьшения количества настраиваемых параметров нейронной сети. Данная нейронная сеть имеет 1351 синаптический коэффициент.

В качестве активационной функции был выбран гиперболический тангенс:

$$f(x) = \text{th}(x).$$

Эта функция имеет ряд преимуществ для решения задачи:

- симметричные активационные функции типа гиперболического тангенса обеспечивают более быструю сходимость, чем стандартная логистическая функция;
- функция имеет непрерывную первую производную;
- функция имеет простую производную, которая может быть вычислена через ее значение, что дает экономию вычислений.

Формула функционирования нейрона сверточного слоя имеет вид

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} x^{((i-1)+s,(j+t))},$$

где $y_k^{(i,j)}$ — нейрон k -й плоскости сверточного слоя; b_k — нейронное смещение k -й плоскости; K — размер рецептивной области нейрона; $w_{k,s,t}$ — матрица синаптических коэффициентов; x — выходы нейронов предыдущего слоя.

Формула функционирования нейрона подвыборочного слоя имеет вид

$$y_k^{(i,j)} = b_k + \frac{1}{4} w_k \sum_{s=1}^2 \sum_{t=1}^2 x^{((i,j)+s,(i,j))}.$$

Используется стандартный для нейронных сетей алгоритм обратного распространения ошибки. Для измерения качества распознавания использовалась функция среднеквадратичной ошибки

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2,$$

где E_p — величина функции ошибки для образа p ; t_{pj} — желаемый выход нейрона j для образа p ; o_{pj} — действительный выход нейрона j для образа p .

Окончательная коррекция синаптических коэффициентов осуществляется по формуле

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj},$$

где η — коэффициент пропорциональности, влияющий на скорость обучения.

На каждой итерации алгоритм обратного распространения ошибки рассчитывается для всего обучающего набора данных, что позволяет вычислить средний или истинный градиент.

Когда ненастроенной сети предъявляется входной образ, она выдает некоторый случайный выход. Функция ошибки представляет собой разность между текущим выходом сети и идеальным выходом, который необходимо получить. Для успешного обучения сети требуется приблизить выход сети к желаемому выходу, т. е. последовательно уменьшать величину функции ошибки. Это достигается настройкой межнейронных связей. Каждый нейрон в сети имеет свои веса, которые настраиваются, чтобы уменьшить величину функции ошибки.

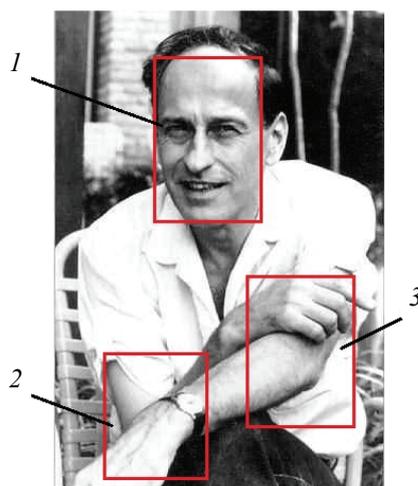


Рис. 4. Пример выделения на изображении информативных участков-кандидатов:
1–3 — найденные участки-кандидаты

Значения весовых коэффициентов были выбраны случайным образом из нормального распределения с нулевым средним и стандартным отклонением:

$$\delta_w = \sqrt{m}$$

(m — число связей, входящих в нейрон).

Экспериментально доказано, что локальное рецептивное поле, равное четырем пикселям, позволяет избежать вычислительной избыточности и обеспечивает надежное обнаружение лица или его окрестности.

Разработка алгоритма масштабирования и кластеризации. Если система находит лицо на изображении, то область местонахождения лица отмечается прямоугольником. Сложность заключается в том, что на одном изображении может быть найдено несколько таких областей — информативных участков-кандидатов на местоположение лица. Это снижает достоверность результатов [11].

В случае, представленном на рис. 4, нейронной сетью найдено на изображении три участка-кандидата. Один участок действительно содержит искомое лицо, а два других являются ложными. Часто на изображениях различные объекты (облака, кроны деревьев и т. д.) имеют форму, похожую на человеческое лицо.

Необходимо разработать алгоритм, позволяющий свести подобные ошибки к минимуму и повысить вероятность обнаружения лица на изображении. Исходное изображение последовательно масштабируется с коэффициентом от 1,2 до 0,8, в результате чего получается пирамида, состоящая из пяти изображений (рис. 5). Эти масштабы были выбраны как наиболее эффективные при распознавании, так как размер лица в них наиболее близок к размеру входа нейронной сети (30×34 пикселя).

Если изображения принимают разные размеры, то частично меняется структура некоторых элементов изображений. Поэтому если изображения просканировать нейронной сетью, то в каждом из них могут оказаться участки-кандидаты, не соответствующие изображениям другого масштаба. После масштабирования эти изображения сканируются нейронной сетью, которая в каждом изображении находит и выделяет участки-кандидаты. Каждый участок-кандидат характеризуется вектором параметров: координаты центра участка, ширина и высота участка, отклик нейронной сети для участка.

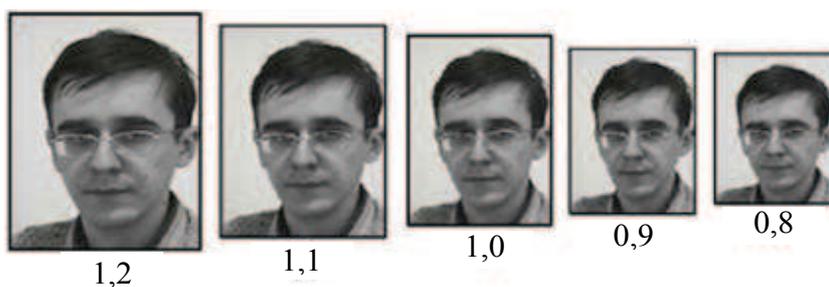


Рис. 5. Масштабированные изображения (в центре — исходное изображение)

Далее, все найденные участки-кандидаты масштабируются относительно исходного изображения: сравниваются, накладываются друг на друга, пересекающиеся участки-кандидаты группируются в кластеры. При достижении суммарным откликом некоторого порога выносятся решение о принадлежности данного кластера к классу лиц. При усреднении размеров всех участков получаем границы области лица. Анализ откликов нейронной сети на наборе изображений позволил оценить порог обнаружения на уровне 0,8.

Участки, не вошедшие в кластеры, отбрасываются как ложные. Таким образом происходит выявление общих признаков изображений и устранение избыточных участков-кандидатов. Экспериментально доказано, что этот алгоритм обеспечивает вероятность обнаружения образа лица на уровне не менее 0,95.

Распознавание изображения. После определения местоположения лица на изображении это лицо нужно распознать. Имеется база данных, содержащая большое количество изображений человеческих лиц (фотопортреты). К каждому фотопортрету прикреплен файл, который содержит антропометрические характеристики и данные о человеке, изображенном на портрете: возраст, пол, должность и т. д. Эти данные можно вводить с клавиатуры, изменять, удалять с помощью СУБД.

Для распознавания лиц можно использовать сверточную нейронную сеть, на вход которой подается фотопортрет человека (рис. 6). Далее происходит сканирование входящего изображения локальным рецептивным полем, информация каждого рецептивного поля преобразуется функцией и записывается в каждый нейрон следующего слоя. Этот процесс продолжается в каждом слое нейронной сети. Таким образом, нейронная сеть извлекает характеристики изображения, инвариантные к масштабу изображения, ракурсу фотоснимка,

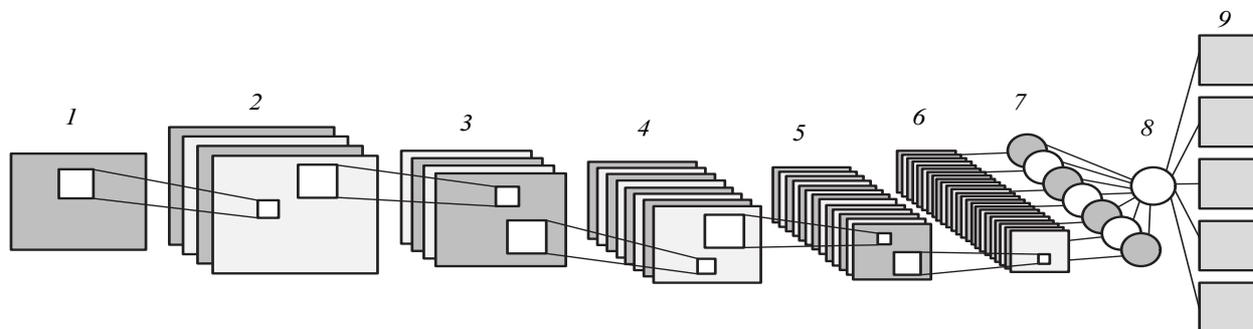


Рис. 6. Архитектура нейронной сети для распознавания лиц:

1 — вход; 2, 4, 6 — сверточные слои; 3, 5 — подвыборочные слои;
7 — слой из обычных нейронов; 8 — выходной слой; 9 — классы фотопортретов

повороту и т. д. На выходе устанавливается соответствие одному из классов изображений, хранимых в базе данных. Такой метод называется обучением с учителем.

Применительно к распознаванию человека по изображению лица такой подход эффективен при решении задач контроля доступа небольшой группы лиц. Это обеспечивает непосредственное сравнение сетью самих образов, но с увеличением числа классов, время обучения и работы сети возрастает экспоненциально. Поэтому при решении таких задач, как поиск похожего человека в большой базе данных, требуется извлечение компактного набора ключевых характеристик, на основе которых можно производить поиск.

Алгоритм системы распознавания. В качестве примера можно описать работу системы контроля доступа. На двери расположена фото- или видеокамера, которая фиксирует человека на входе. Делается фотоснимок человека. На фотоснимке находится область расположения лица, далее происходит распознавание этого лица. Если лицо соответствует портрету, хранящемуся в базе данных, то читается дополнительная информация: имя, возраст, должность и т. д. На основе этих данных система открывает или закрывает доступ к объектам.

На рис. 7 показан алгоритм работы системы при вводе информации в базу данных. На вход поступает изображение, представляющее собой фотографию человека. С помощью нейронной сети производится поиск области лица на фотографии. Область лица выделяется, оптимизируются яркость, контраст изображения, после чего данный фотопортрет сохраняется в отдельном файле базы данных. Пользователь системы может добавить с клавиатуры дополнительную информацию о человеке, изображенном на фото: возраст, пол, должность и т. д. Эта информация записывается в тот же файл базы данных.

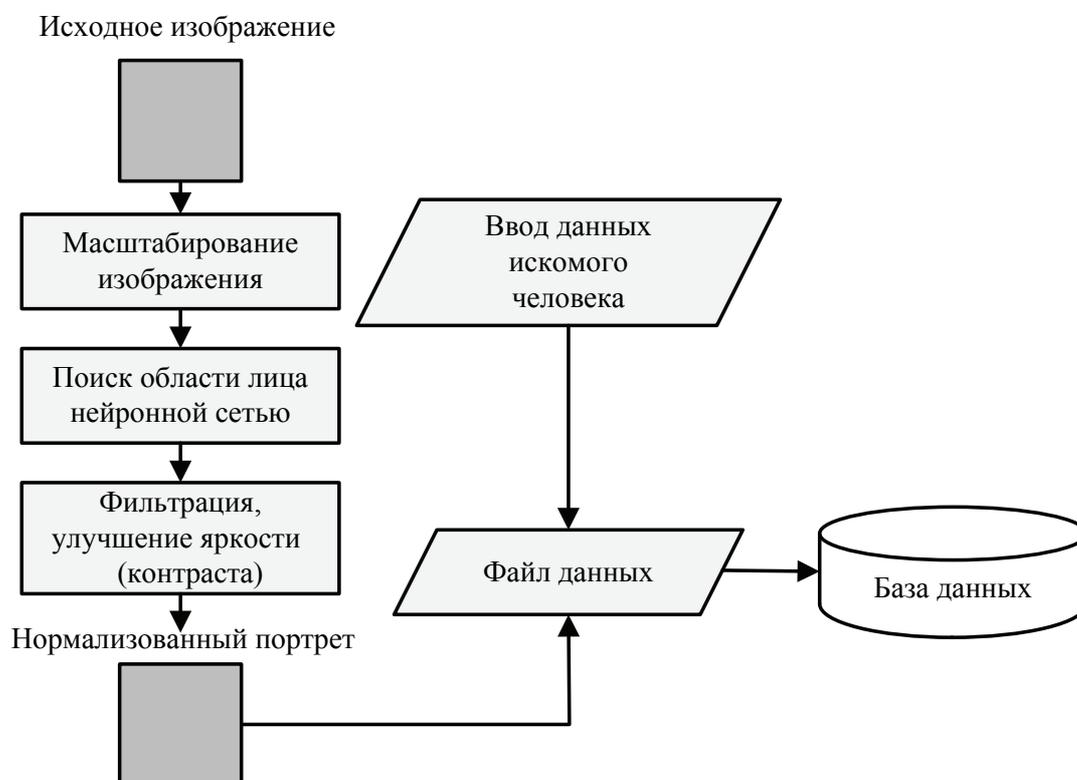


Рис. 7. Ввод информации в базу данных

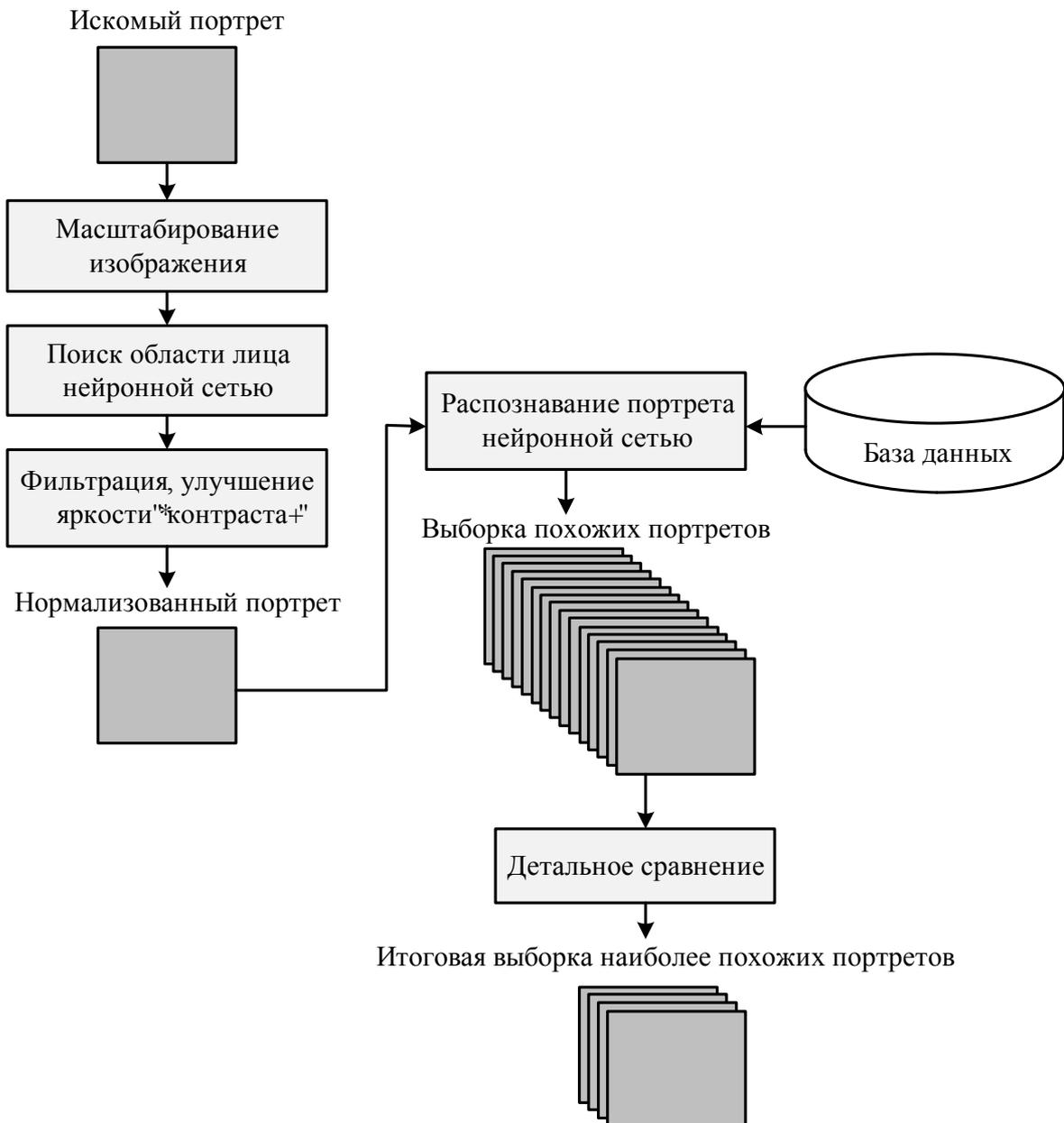


Рис. 8. Поиск информации в базе данных

На рис. 8 показан алгоритм работы системы при поиске информации в базе данных. Система наблюдения делает фотографию человека. С помощью нейронной сети производится поиск области лица на этой фотографии. Область лица выделяется, оптимизируются яркость, контраст изображения, затем нормализованный фотопортрет поступает на обработку второй нейронной сети для распознавания. Нейронная сеть производит распознавание входящего портрета и осуществляет выбор из нескольких наиболее похожих на него портретов, хранящихся в базе данных.

Заключение. Показано, что для решения задачи распознавания образов оптимальны сверточные нейронные сети. Разработана топология сверточной нейронной сети, обеспечивающая инвариантность к изменениям условий съемки, поворотам изображений, качеству съемки. Представлен алгоритм масштабирования и кластеризации, который производит последовательное масштабирование исходного изображения от 0,8 до 1,2 и объединяет в клас-

теры найденные участки-кандидаты. Алгоритм обеспечивает вероятность обнаружения образа лица на уровне не менее 0,95.

Список литературы

1. LE CUN Y., BENGIO Y. Convolutional networks for images, speech and time series // Handbook Brain Theory Neural Networks. 1998. V. 7, N 1. P. 255–258.
2. RUMELHART D. E., HINTON G. E., WILLIAMS R. J. Learning internal representations by error propagation // Parallel Distributed Proc. 1986. V. 1, N 2. P. 318–362.
3. САДЫХОВ Р. Х., ВАТКИН М. Е. Модифицированный алгоритм обучения РБФ-сети для распознавания рукописных символов // Идентификация образов. 2001. Т. 1, № 3. С. 7–16.
4. ROWLEY H. A., BALUJA S., KANADE T. Neural network-based face detection // Pattern Anal. Mach. Intell. 2000. V. 5. P. 23–38.
5. LIN S., KUNG S., LIN L. Face recognition detection by probabilistic decision-based neural network // Trans. Neural Networks. 1997. V. 8, N 1. P. 114–132.
6. FERAUD R., BERNIER O., VIALLET J., COLLOBERT M. A fast and accurate face detector based on neural networks // Trans. Pattern Anal. Machine Intelligence. 2002. V. 3, N 23. P. 42–53.
7. ROWLEY H., BALUJA S., KANADE T. Rotation invariant neural network-based face detection // Proc. IEEE. 1998. V. 7, N 8. P. 38–44.
8. SAKAI T., NAGAO M., FUJIBAYASHI S. Line extraction and pattern detection in a photograph // Pattern Recognition. 1969. V. 1, N 5. P. 233–248.
9. CRAW I., ELLIS H., LISHMAN J. Automatic extraction of face features // Pattern Recognition. Lett. 1987. V. 5. P. 183–187.
10. YU N., NOTKIN B. S., SEDOV V. A. Neuro-iterative algorithm of tomographic reconstruction of the distributed physical fields in the fibreoptic measuring systems // Computer Optics. 2009. V. 33, N 4. P. 446–455.
11. WILSON D. R., MARTINEZ T. R. The general inefficiency of batch training for gradient descent learning // Neural Networks. 2004. V. 16, N 4. P. 1429–1451.

*Друки Алексей Алексеевич — асп. Института кибернетики ТПУ;
тел. (3822) 22-45-08; e-mail: druki2008@yandex.ru*

Дата поступления — 23.12.10 г.