

УДК 004.931

**СИСТЕМА ПОИСКА, ВЫДЕЛЕНИЯ И РАСПОЗНАВАНИЯ ЛИЦ НА ИЗОБРАЖЕНИЯХ**

А.А. Друки

Томский политехнический университет  
E-mail: druki2008@yandex.ru*Для решения задачи выделения и распознавания лиц использованы сверточные нейронные сети. Представлена структура разработанной нейронной сети. Представлен алгоритм масштабирования и кластеризации изображений.***Ключевые слова:***Обработка изображений, нейронные сети, алгоритм масштабирования и кластеризации.***Key words:***Image recognition, neural networks, image recognition, algorithm for scaling and clustering.***Введение**

Многие направления науки, техники и производства в значительной степени ориентируются на развитие систем, в которых информация носит характер поля (изображения). При обработке такой информации возникает ряд сложных научных, технических и технологических проблем. Одной из самых сложных задач является обработка и распознавание изображений. О важности этой проблемы говорит 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 широкое применение в различных приложениях — это может быть 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. Традиционные системы идентификации требуют знания пароля, наличия ключа, идентификационной карточки, либо иного идентифицирующего предмета, который можно забыть или потерять. В отличие от них, биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые однозначно определяют конкретного человека. К таким характеристикам 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], на разложении Кархунена–Лоэва [4], на алгебраических моментах [5], линиях одинаковой интенсивности, эластичных (деформируемых) эталонах сравнения.

Анализ методов выделения сюжетной части изображения показал, что для решения данной задачи

эффективно использовать искусственные нейронные сети, в связи с тем, что они обеспечивают возможность получения классификатора, хорошо моделирующего сложную функцию распределения изображений лиц, тем самым, увеличивая точность решения по сравнению с остальными методами.

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

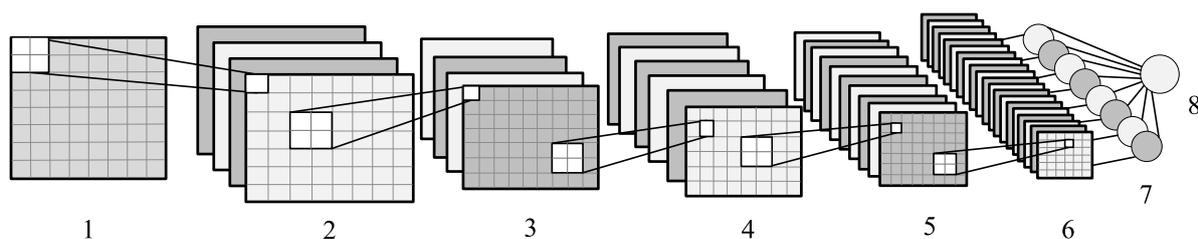
Объектом исследования являются системы обработки изображений, основанные на искусственных нейронных сетях.

Предметом исследования являются алгоритмы выделения сюжетной части и распознавание образов на групповом изображении, основанные на искусственных нейронных сетях.

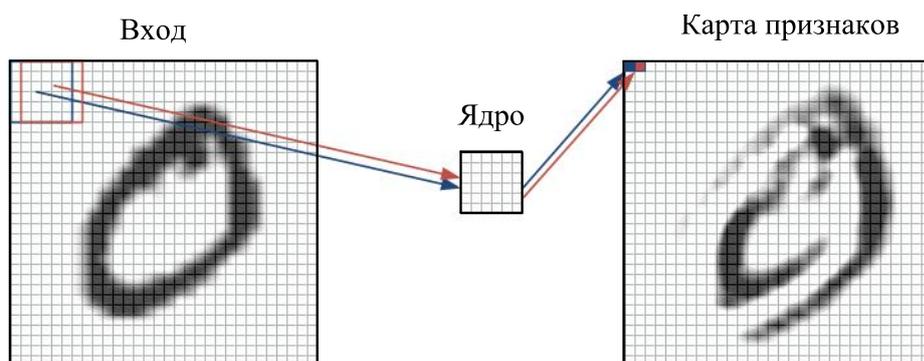
**Разработка нейронной сети**

Наиболее часто в задачах распознавания и идентификации изображений используются классические нейросетевые архитектуры (многослойный перцептрон, сети с радиально-базисной функцией и др. [6]), но из анализа данных работ и экспериментальных исследований следует, что применение классических нейросетевых архитектур в данной задаче неэффективно по следующим причинам:

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



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



**Рис. 2.** Результат сканирования изображения

Поэтому для решения этой задачи были выбраны сверточные нейронные сети, т. к. они обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям [10].

Архитектура сверточной нейронной сети состоит из многих слоёв. Слои бывают двух типов: сверточные и подвыборочные, они чередуются друг с другом как показано на рис. 1.

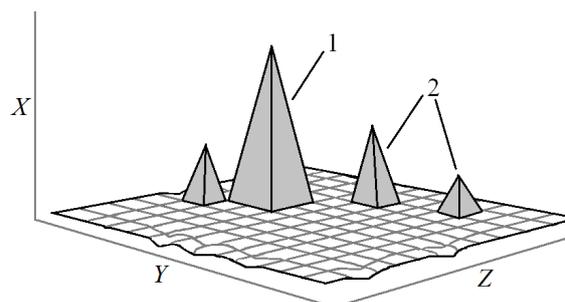
Нейроны в пределах слоя организованы в плоскости. В каждом слое имеется набор из нескольких плоскостей, причём нейроны одной плоскости имеют одинаковые синаптические коэффициенты, ведущие ко всем локальным участкам предыдущего слоя. Каждый нейрон слоя получает входы от некоторой области предыдущего слоя (локальное рецептивное поле), т. е. входное изображение предыдущего слоя как бы сканируется небольшим окном 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 путём локального усреднения значений выходов нейронов. Таким образом достигается иерархическая организация. Последующие слои извлекают более общие характеристики, меньше зависящие от искажений изображения [10].

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.



**Рис. 3.** Отклики нейронной сети: 1 – наиболее вероятное местонахождение лица на изображении; 2 – наименее вероятное нахождение лиц на изображении (шумы)

Разработана нейронная сеть, состоящая из 6 слоёв. Входными данными нейронной сети являются изображения размером 32×36 пикселей.

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].

Следом за входным слоем находится свёрточный слой  $S1$ . Этот слой состоит из 5 свёрточных плоскостей, каждая из которых имеет синаптическую маску размером  $5 \times 5$  и нейронное смещение.

Размер свёрточной плоскости определяется в соответствии со следующим выражением:

$$w_c = w_u - K + 1,$$

$$h_c = h_u - K + 1,$$

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

Размер плоскости свёрточного слоя  $S1$ :  $28 \times 32$  нейрона. Количество синаптических коэффициентов в слое: 130.

Подвыборочный слой  $S1$  состоит из 5 карт характеристик и обеспечивает локальное усреднение и подвыборку. Каждая плоскость слоя  $S1$  связана лишь с одной плоскостью слоя  $S1$ . Размер каждой плоскости слоя  $S1$   $14 \times 16$  нейронов, что вдвое меньше, чем размер плоскости предыдущего слоя. Каждая плоскость слоя  $S1$  имеет единственный синаптический коэффициент и нейронное смещение, что дает в итоге 10 настраиваемых параметров.

Свёрточный слой  $C2$  состоит из 20 плоскостей. Размер плоскости:  $12 \times 14$  нейронов. Слой содержит 290 синаптических коэффициентов. Плоскости слоя  $C2$  формируются следующим образом, каждая из 5 плоскостей слоя  $S1$  свернута 2 различными синаптическими масками  $3 \times 3$ , обеспечивая 10 плоскостей в  $C2$ , другие 10 плоскостей  $C2$  получены, путём суммирования результатов 2 свёртываний 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 каждой на одну плоскость слоя  $C2$ , размер каждой плоскости  $6 \times 7$  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 в вычислении 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:

$$f(x) = \tanh(x).$$

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 свёрточного слоя:

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

где  $y_k^{(i,j)}$  – нейрон  $k$ -й плоскости свёрточного 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:

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

Используется стандартный для нейронных сетей алгоритм обратного распространения ошибки. Для измерения качества распознавания использовалась функция среднеквадратической 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 происходит по формуле:

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

где  $\eta$  – коэффициент пропорциональности, влияющий на скорость обучения.

На каждой итерации алгоритм обратного распространения ошибки рассчитывается для всего обучающего набора данных, чтобы 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.

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

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

где  $m$  — число связей, входящих в нейрон.

Экспериментально доказано, что локальное рецептивное поле равное четырем пикселям, позволяет избежать вычислительной избыточности и обеспечить 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].

На рис. 4 показано, что нейронной сетью найдено на изображении 3 участка-кандидата. Один участок действительно содержит искомое лицо, а два других участка являются ложными. Часто на изображениях различные объекты имеют форму, похожую на человеческое лицо, будь то облака, кроны деревьев и т. д.

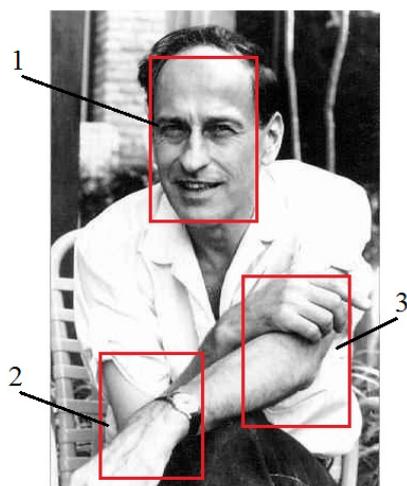


Рис. 4. Пример выделения на изображении информативных участков-кандидатов. Цифрами 1, 2, 3 обозначены найденные участки-кандидаты

Предпринято 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. Таким образом, получается пирамида, состоящая из пяти изображений, как показано на рис. 5. Эти масштабы были выбраны как наиболее эффективные при распознавании, так как размер лица в этих масштабах наиболее близок к размеру входа нейронной сети (30×34 пикселей).

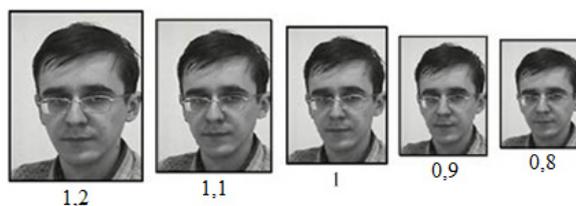


Рис. 5. Масштабированные изображения. В центре исходное изображение

Если изображения принимают разный размер, то частично меняется структура некоторых элементов изображений. Поэтому, если эти изображения просканировать нейронной сетью, то в каждом изображении могут найтись свои индивидуальные участки-кандидаты, не соответствующие изображениям другого масштаба.

После масштабирования, эти изображения сканируются нейронной сетью. Нейронная сеть 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.

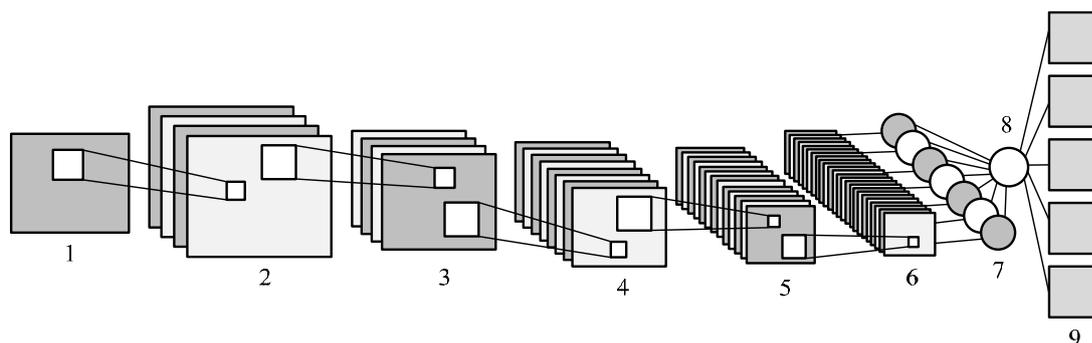
Далее, все найденные участки-кандидаты масштабируются относительно исходного изображения: сравниваются, накладываются друг на друга и пересекающиеся участки-кандидаты группируются в кластеры. При достижении суммарным откликом некоторого порога, выносится решение о принадлежности данного кластера к классу 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.

Те участки, которые не вошли 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 нахождения местоположения лица на изображении, это лицо нужно распознать. Имеется база данных, содержащая множество 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 используем 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. Та-



**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

ким образом, нейронная сеть извлекает характеристики изображения, инвариантные к масштабу изображения, ракурсу фотоснимка, повороту и т. д. На выходе устанавливается соответствие одному из классов изображений, хранимых в базе данных. Это называется обучением с учителем.

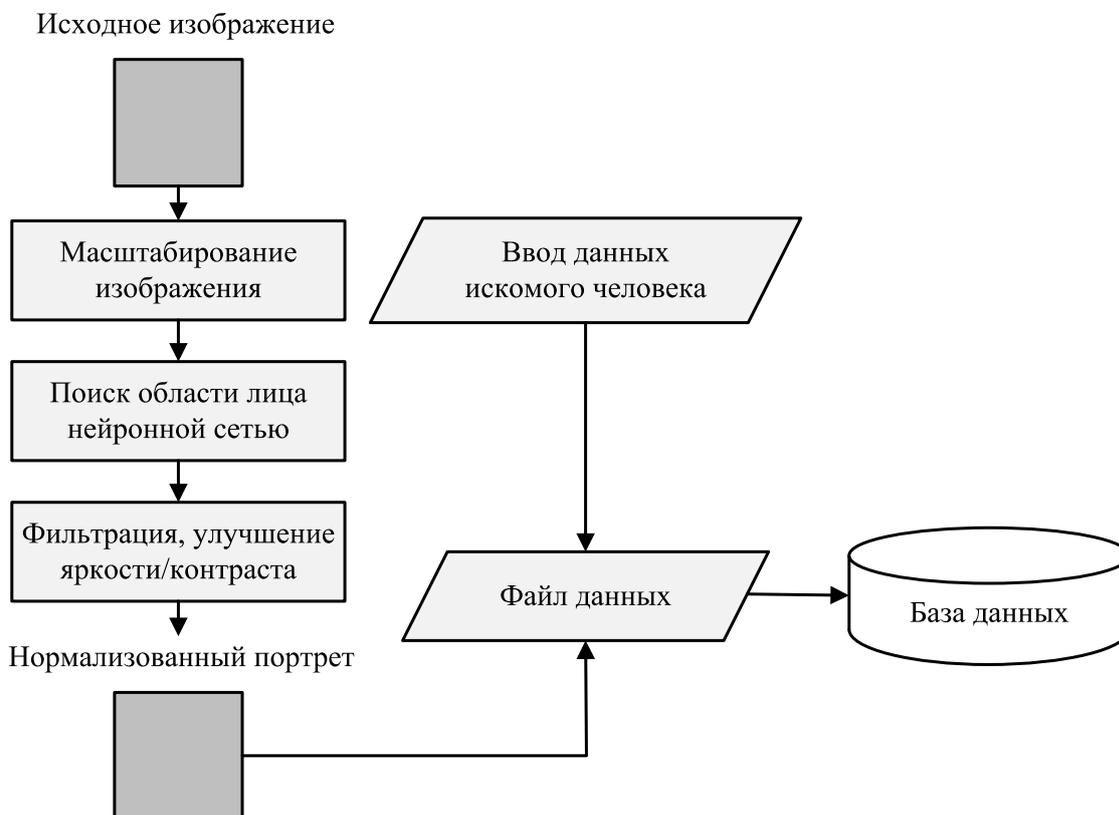
В применении к распознаванию человека по изображению лица, такой подход хорош для задач контроля доступа небольшой группы лиц. Это обеспечивает непосредственное сравнение сетью самих образов, но с увеличением числа классов, время обучения и работы сети возрастает экспоненциально. Поэтому для таких задач, как поиск похожего человека в большой базе данных, требуется извлечения 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, претендующего на вход. Делается фотоснимок человека. На фотоснимке находится область расположения лица, далее происходит распознавание этого лица. Если это 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 фотографию человека. С помощью нейронной сети



**Рис. 7.** Ввод информации в базу данных

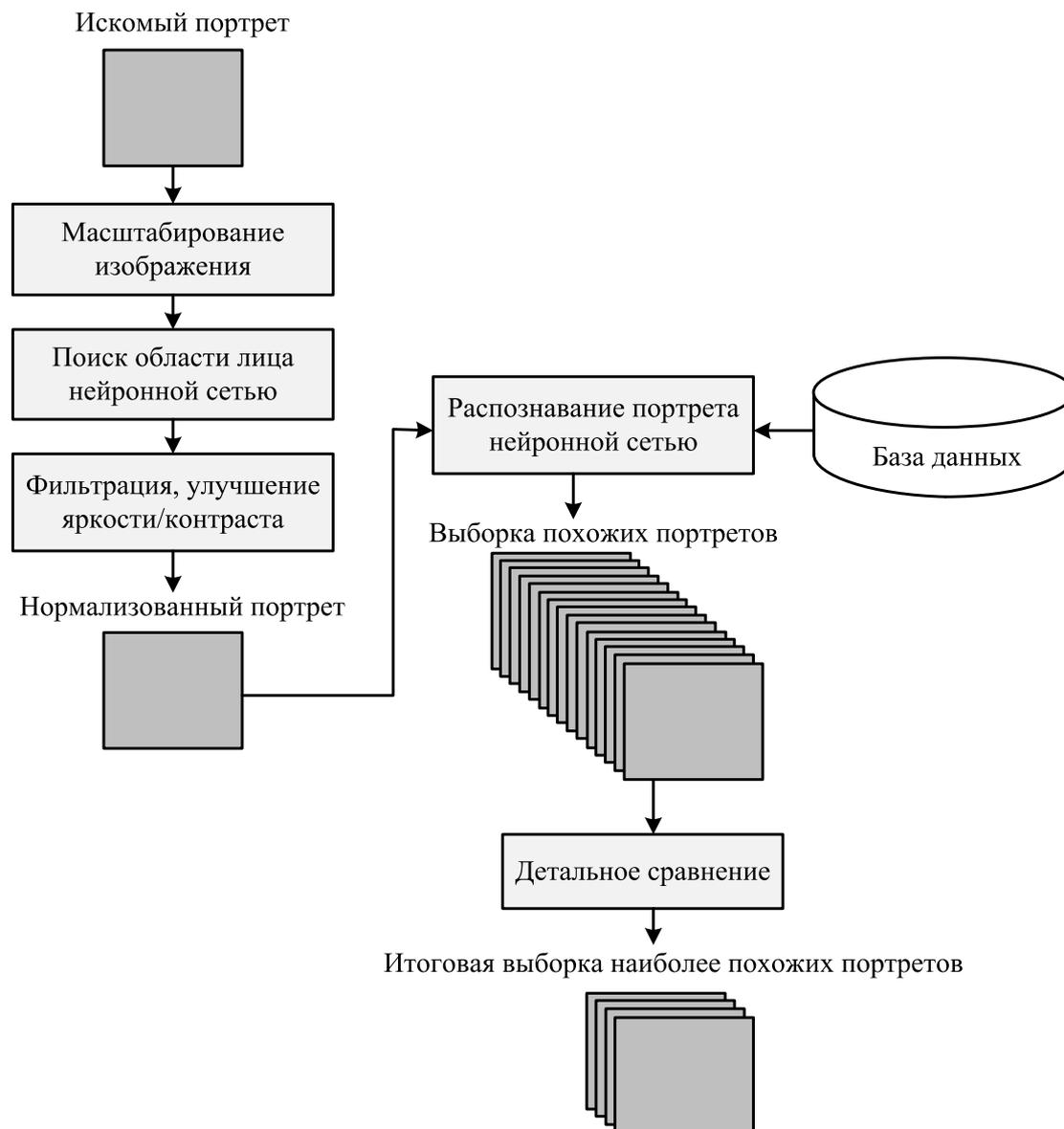


Рис. 8. Поиск информации в базе данных

производится поиск области лица на фотографии. Область лица выделяется, оптимизируется яркость, контраст изображения, и данный фотопортрет сохраняется в отдельный файл 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.

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, оптимизируется яркость, контраст изображения, и нормализованный фотопортрет поступает на обработку второй нейронной сети для распознавания. Нейронная сеть производит

распознавание входящего портрета и выводит выборку из нескольких наиболее похожих на него портретов, хранящихся в базе данных.

#### Выводы

Показано, что для решения задачи распознавания образов оптимальны сверточные нейронные сети. Разработана топология свёрточной 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 кластеры 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 // The handbook of brain theory and neural networks. – 1998. – V. 7. – № 1. – P. 255–258.
2. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation // Parallel distributed processing. – 1986. – V. 1. – № 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. – P. 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. – № 1. – P. 114–132.
6. Feraud R., Bernier O., Viallet J., Collobert M. A fast and accurate face detector based on neural networks // Transactions on pattern analysis and machine intelligence. – 2002. – V. 3. – № 23. – P. 42–53.
7. Rowley H., Baluja S., Kanade T. Rotation invariant neural network-based face detection // Proceedings of conference on computer vision and pattern recognition. – 1998. – V. 7. – № 8. – P. 38–44.
8. Sakai T., Nagao M., Fujibayashi S. Line extraction and pattern detection in a photograph // Pattern recognition. – 1969. – V. 1. – № 5. – P. 233–248.
9. Craw I., Ellis H., Lishman J. Automatic extraction of face features // Pattern recognition letters. – 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 fibre-optic measuring systems // Computer optics. – 2009. – V. 33. – № 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. – № 4. – P. 1429–1451.

Поступила 23.12.2010 г.

УДК 004.932

**АНАЛИЗ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ИМПУЛЬСНОГО ШУМА НА ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ**

Нгок Хоанг Фан, В.Г. Спицын

Томский политехнический университет  
E-mail: hoangpn285@gmail.com

Выполнен анализ способов обнаружения присутствия импульсного шума на цифровых изображениях. Приведены результаты сопоставления различных алгоритмов обнаружения пикселей, искаженных импульсным шумом.

**Ключевые слова:**

Обработка изображения, импульсный шум, обнаружение шума.

**Key words:**

Image processing, impulse noise, noise detection.

**Введение**

Цифровые изображения часто искажаются шумами, появляющимися на этапах их получения и/или передачи. Причинами возникновения шумов на цифровых изображениях могут быть сбои в работе канала связи, шум видеодатчика и др. Одним из основных видов шумов является импульсный шум [1].

Пусть  $X_{i,j}$  – интенсивность пикселя  $(i, j)$  исходного изображения  $X$ ,  $[n_{\min}, n_{\max}]$  – динамический диапазон изображения  $X$ ,  $Y_{i,j}$  – интенсивность пикселя  $(i, j)$  зашумленного изображения  $Y$ . В этом случае выражение для интенсивности  $Y_{i,j}$  в результате воздействия импульсного шума имеет вид:

$$Y_{i,j} = \begin{cases} X_{i,j}, & 1-p \\ R_{i,j}, & p \end{cases}, \quad (*)$$

где  $p$  – вероятность импульсного шума;  $R_{i,j}$  – значение интенсивности пикселя  $(i, j)$ , искаженного импульсным шумом. При выполнении условия  $R_{i,j} \in [n_{\min}, n_{\max}]$ , изображение искажается импульсным шумом случайного значения (*Random-Valued*

*Impulse Noise – RVIN*). В случае  $R_{i,j} \in \{n_{\min}, n_{\max}\}$ , изображение искажается импульсным шумом фиксированного значения или шумом «соли и перца» (*Salt & Pepper Noise – SPN*).

Различие между SPN и RVIN представлено на рис. 1. В случае SPN значение  $R_{i,j}$  в формуле (\*) может принимать только два значения:  $n_{\min}(0)$  или  $n_{\max}(255)$ . Для RVIN случайное значение  $R_{i,j}$  имеет равномерный закон распределения в пределах от  $n_{\min}$  до  $n_{\max}$ .

Одной из основных задач обработки изображения является подавление шумов. Для удаления шумов, как правило, используют фильтры. Обычно процесс фильтрации цифровых изображений включает два этапа: классификацию (обнаружение зашумленных пикселей) и реконструкцию (замену значений интенсивности обнаруженных пикселей на прогнозируемые значения).

В данной статье рассматриваются алгоритмы обнаружения импульсного шума «соли и перца» на цифровых изображениях. Проведен анализ работы этих алгоритмов и осуществлено сравнение полученных результатов.