

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК БЕЛАРУСИ
ИНСТИТУТ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ**

Д.В.Брилюк, В.В.Старовойтов

**РАСПОЗНАВАНИЕ ЧЕЛОВЕКА ПО
ИЗОБРАЖЕНИЮ ЛИЦА НЕЙРОСЕТЕВЫМИ
МЕТОДАМИ**

Препринт № 2

Минск 2002

УДК 681.3.01+681.327.12

Брилюк Д.В., Старовойтов В.В. **Распознавание человека по изображению лица нейросетевыми методами.** – Минск, 2002. – 54 с. (Препринт / Ин-т техн. кибернетики НАН Беларуси; № 2).

Выполнен обзор современной литературы по нейросетевым методам распознавания человека по изображению лица. Рассмотрены основные классы задач распознавания человека по изображению лица. Описаны архитектуры нейронных сетей различного типа, показано, как они применяются для распознавания людей. Описаны предварительные эксперименты по созданию системы контроля доступа на основе анализа изображения лица человека. Приведены адреса в сети Интернет по данной тематике.

Ил. 18, библиограф. 59 назв.

Рецензент

кандидат технических наук Н.И. Мурашко

© Д.В. Брилюк, В.В. Старовойтов,
2002

Введение

В настоящее время всё более широкое распространение получают биометрические системы идентификации человека. Традиционные системы идентификации требуют знания пароля, наличия ключа, идентификационной карточки либо иного идентифицирующего документа, который можно забыть, потерять или подделать. 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 [11]. Лицо, голос и запах каждого человека также индивидуальны [5,11,22,30].

Распознавание человека по изображению лица выделяется среди биометрических систем тем [11], что, во-первых, не требует специального дорогостоящего оборудования. Для большинства приложений достаточно персонального компьютера и обычной видеокамеры. Во-вторых, отсутствует физический контакт человека с устройствами. Не надо ни к чему прикасаться или специально останавливаться и ждать срабатывания системы. В большинстве случаев достаточно просто пройти мимо или задержаться перед камерой на несколько секунд.

К недостаткам распознавания человека по изображению лица следует отнести то, что сама по себе такая система не обеспечивает 100%-ной надёжности идентификации. Там, где требуется высокая надёжность, применяют комбинирование нескольких биометрических методов [11].

На данный момент проблеме распознавания человека по изображению лица посвящено множество работ, однако в целом она ещё далека от разрешения. Основные трудности состоят в том, чтобы распознать человека по изображению лица независимо от изменения ракурса и условий освещённости при съёмке, а также при различных изменениях, связанных с возрастом, причёской и т.д. [11,25,30,58].

Распознавание изображений пересекается с распознаванием образов [1,19,23]. Такие задачи не имеют точного аналитического решения. При этом требуется выделение ключевых признаков, характеризующих зрительный образ, определение относительной важности признаков путём выбора их весовых коэффициентов и учёт взаимосвязей между признаками. Изначально эти задачи выполнялись человеком-экспертом вручную, путём экспериментов, что занимало много времени и не гарантировало качества. В новых методах выделение ключевых признаков осуще-

ствляется путём автоматического анализа обучающей выборки, но тем не менее большая часть информации о признаках задаётся вручную [14,16]. Для автоматического применения таких анализаторов выборка должна быть достаточно большой и охватывать все возможные ситуации (например, изменения ракурса, внешности, условий освещённости и т.п.).

Нейросетевые методы предлагают иной подход к решению задачи распознавания образов [3,6,9,10,21]. Архитектура и функционирование нейронных сетей (НС) имеют биологические прообразы. Веса в нейронной сети не вычисляются путём решения аналитических уравнений, а подстраиваются различными локальными методами (например, различными градиентного спуска) при обучении. Обучаются нейронные сети на наборе обучающих примеров. В процессе обучения НС происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Обученная НС может успешно применять опыт, полученный в процессе обучения, на неизвестные образы за счёт хороших обобщающих способностей.

Таким образом, применение нейронных сетей для задачи распознавания человека по изображению лица является перспективным направлением, на что и направлено основное внимание препринта.

В данной работе описаны методы распознавания человека по изображению лица и возможности использования их совместно с нейросетевыми методами. Дан обзор нейросетевых методов распознавания изображений и их применения к распознаванию человека по изображению лица. Описаны различные способы представления изображения и учёта его свойств. Прилагаются ссылки на ресурсы в сети Интернет, посвящённые соответствующей тематике. Обзор будет полезен и тем, кто занимается распознаванием изображений и другими задачами, связанными с распознаванием образов.

1. Основные классы решаемых задач в распознавании человека по изображению лица

Задачи распознавания человека по изображению лица делятся на три больших класса: поиск в больших базах данных, контроль доступа и контроль фотографий в документах. Они различаются как по требованиям, предоставляемым к системам распознавания, так и по способам решения, и поэтому представляют собой отдельные классы.

Различны и требования, предъявляемые к ошибкам первого и второго рода для таких классов [8,35]. **Ошибкой первого рода** (type I error,

misdetection) называется ситуация, когда объект заданного класса не распознаётся (пропускается) системой. **Ошибка второго рода** (type II error, *false alarm*) происходит, когда объект заданного класса принимается за объект другого класса.

Следует также отметить различие понятий верификации и распознавания (идентификации) [22]. В задаче **верификации** неизвестный объект заявляет, что он принадлежит к некоторому известному системе классу. Система подтверждает или опровергает это заявление. В системах верификации ошибкой первого рода является ситуация, когда объект, принадлежащий к известным системе классам, принимается за объект, относящийся к неизвестным системе классам, и в доступе ему отказывают [2,8]. Ошибка второго рода совершается, когда объект неизвестного класса принимается за объект, относящийся к известным системе классам, и ему разрешается доступ [2,8]. При **распознавании** требуется отнести объект к одному из n известных классов или выдать заключение о том, что этот объект не относится к известным классам.

1.1. Поиск изображения в больших базах данных

Сравнение типа «один со многими». Высокие требования к ошибке первого рода – система распознавания должна находить изображения, соответствующие данному человеку, по возможности не пропустив ни одного такого изображения. При этом допустимо, если в результирующей выборке будет присутствовать небольшое число других людей.

Обычно в большой базе данных (10^4 - 10^7 изображений) требуется найти изображения, наиболее похожие на заданное [16]. Поиск должен быть произведён за разумное время. Одно из решений состоит в хранении в базе данных небольших наборов заранее извлечённых ключевых признаков, максимально характеризующих изображение. При этом требования к точности не столь критичны, как в задачах контроля доступа и документного контроля.

К данному классу прежде всего относится метод главных компонент (метод «собственных лиц») [12,46]. Коэффициенты, полученные разложением входного изображения на главные компоненты, использовались для сравнения изображений путём вычисления Евклидова расстояния, а в более совершенных методах – на основе метрики Махаланобиса с использованием Гауссовского распределения [46].

В работах [28,57] описано развитие метода главных компонент на основе нейронных сетей.

В работе [41] также показана возможность использования признаков, сформировавшихся на более поздних слоях специализированной свёрточной нейронной сети, для классификации изображений по методу ближайшего соседа.

1.2. Задача контроля доступа

Сравнение типа «один с несколькими». Критическими являются требования к ошибкам второго рода. Система распознавания не должна распознавать незнакомых людей как знакомых, возможно даже за счёт увеличения ошибок первого рода (отказов в доступе знакомым людям).

Имеется небольшая группа лиц (5-50 человек), которых система должна распознавать по изображению лица и открывать им доступ в некоторое помещение. Людей, не входящих в эту группу, система не должна пропускать. Возможны варианты, когда требуется установить конкретную личность по изображению лица. При этом от системы требуется высокая достоверность распознавания, возможно даже за счёт увеличения числа отказов на знакомые объекты.

В качестве тренировочных изображений обычно для каждого человека доступны несколько изображений лица, полученных при различных условиях. Это могут быть, например, изменения ракурса, условий освещённости, причёски, мимики, наличие или отсутствие очков и т.п.

Система должна работать в реальном масштабе времени, а процесс настройки может занимать больше времени и выполняться предварительно. В процессе эксплуатации система должна дообучаться на вновь поступающих изображениях по возможности быстрее.

Ограничений на применяемые методы здесь нет, но все методы сходятся в том, что имеется обучающий набор изображений лиц заданной группы людей (возможно при различных условиях съёмки). К этому набору система обращается в процессе распознавания или настраивается на него в процессе обучения.

Одним из распространённых подходов к решению такой задачи является использование нейронных сетей, которые после обучения обладают хорошей обобщающей способностью, что будет подробнее описано ниже.

1.3. Задача контроля фотографии в документах

Сравнение типа «один с одним». Формулировать требования к ошибкам первого и второго рода как к системе верификации или распо-

знавания здесь будет некорректно, поскольку система распознавания никогда не имела дело с поступающими на вход классами. Но желательно, чтобы система не совершала ошибок при сравнении.

Требуется сравнить изображение лица человека, полученное в данный момент, с фотографией из какого-либо документа. Системе надо ответить, принадлежат ли эти лица одному человеку или нет. Данный класс задач наиболее сложен, поскольку, во-первых, система никогда раньше не сталкивалась с изображением лица данного человека. Система сравнивает всегда отличающиеся изображения, учёт всех возможных различий в процессе обучения или настройки системы затруднителен. Во-вторых, здесь большое влияние оказывают возрастные и другие изменения лица. В-третьих, качество и контраст отсканированной фотографии, как правило, хуже, чем изображение лица, снятого камерой [15,17].

Большинство методов для данного класса задач неприменимы без специальной адаптации.

В обозреваемой литературе нет работ, напрямую связанных с применением нейросетевых методов для решения данного класса задач. Для этого можно предложить применение НС для извлечения ключевых признаков изображений и адаптацию НС для сравнения двух изображений.

2. Нейросетевые методы распознавания человека по изображению лица

Нейросетевые методы, основанные на применении различных типов искусственных нейронных сетей (ИНС, в дальнейшем просто нейронные сети, НС), в последнее время получили широкое распространение.

Основные задачи, решаемые при помощи нейронных сетей [3,6,9,10,21]:

- разбиение пространства признаков на области, соответствующие классам (классификация, распознавание, кластеризация);
- извлечение ключевых характеристик, сжатие и реконструкция образов;
- аппроксимация функции многих переменных с любой заданной точностью;
- прогнозирование временных рядов;
- ассоциативная память;
- решение оптимизационно-комбинаторных задач;
- топологически упорядоченное преобразование пространства;
- распознавание с учётом топологии пространства.

Большинство из этих задач прямо или косвенно связаны с распознаванием изображений.

Основные преимущества, которыми обладают нейронные сети, следующие.

Настройка нейронной сети для решения определённой задачи производится в процессе обучения на наборе тренировочных примеров. Таким образом, не требуется вручную определять параметры модели (выбирать ключевые признаки, учитывать их взаимоотношение и т.п.) – НС извлекает параметры модели автоматически наилучшим образом в процессе обучения. Остаётся только построить тренировочную выборку. В задачах классификации при этом происходит неявное выделение ключевых признаков внутри сети, определение значимости признаков и системы взаимоотношений между ними. В настоящее время разработаны мощные, гибкие и универсальные механизмы обучения различных типов НС [3,6,9,10]. Кроме того, архитектура НС и процедуры обучения позволяют выполнить гибкую настройку на конкретную решаемую задачу. Для большинства НС процедура обучения является эвристическим алгоритмом, что, с одной стороны, обеспечивает приемлемость получаемых решений, а с другой стороны, не требует непомерных вычислительных ресурсов.

Нейронные сети обладают хорошей обобщающей способностью (одной из лучших среди существующих методов, например много лучшей, чем у решающих деревьев [40]). Это значит, что опыт, полученный в процессе обучения на конечном наборе образов, НС может успешно распространять на всё множество образов. Кроме интерполяционных обобщающих способностей, НС (многослойные персептроны, например) могут хорошо экстраполировать, т.е. применять свой опыт на качественно иные образы, чем те, которые встречались в обобщающей выборке.

Нейронные сети ни налагают каких-либо ограничений на тренировочную выборку, ни полагаются на то, что она обладает какими-либо априорными свойствами, в отличие, например, от статистических методов [9,24,49]. Не требуется никакого предварительного изучения характера данных. НС принимает тренировочный набор «как есть» и учится производить правдоподобное решение, не претендуя на абсолютную истину, т.е. строится наилучшая нефизическая модель [3], которая не является максимально точным соответствием реального процесса, но даёт приемлемую его аппроксимацию. Имеется ряд примеров, когда нейронные сети показывали себя лучше статистических методов [9]. Кроме того, в статистике не имеется аналогов некоторых нейросетевых методов

[9], таких, например, как карты Кохонена, машина Больцмана и, что важно для распознавания изображений, когнитрон.

Естественным образом архитектура НС реализуется на параллельных вычислительных средствах: специализированных микросхемах, оптических и квантовых компьютерах. Это открывает широкие перспективы применения НС в будущем. НС характеризуется нечётким и распределённым хранением информации, т.е. нет отдельного нейрона, отвечающего за какое-либо понятие или признак, и удаление или искажение работы этого нейрона не приведёт к фатальным последствиям.

Но несмотря на все достоинства, применение НС к изображениям требует специальных усилий. Это связано в первую очередь со сложным характером изображений, особенно изображений трёхмерных объектов реального мира, какими и являются лица людей. Изображение должно быть предобработано – приведено к некоторым стандартным условиям. Кроме того, выбор начального представления изображения (это могут быть, например, частотные коэффициенты, главные компоненты, вейвлетные коэффициенты, моменты и т.п.) является отдельной обширной темой. Двумерный характер изображения, изменение условий освещённости, топологические искажения изображения при смене ракурса и прочих воздействиях не позволяют ограничиться простейшими архитектурами НС для достижения оптимального результата.

Кроме различных способов применения нейронных сетей к распознаванию человека по изображению лица, существует множество работ, посвящённых применению нейронных сетей к распознаванию и обработке изображений других видов объектов.

2.1. Архитектура нейронных сетей

Большинство нейронных сетей [6,9,10,21] состоят из *формальных нейронов* (рис. 1).

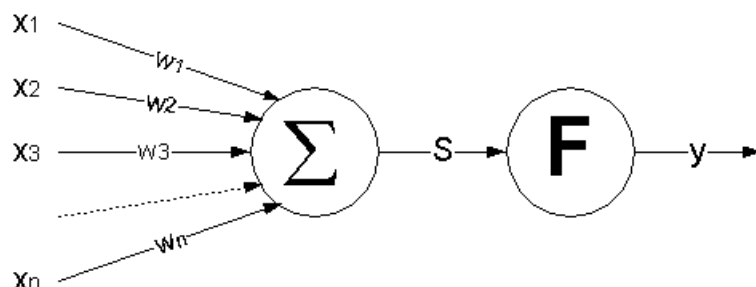


Рис. 1. Формальный нейрон

Здесь $x_1..x_n$ – значения, поступающие на входы (синапсы) нейрона, $w_1..w_n$ – веса синапсов (могут быть как тормозящими, так и усиливающими), S – взвешенная сумма входных сигналов [6,9,10,21]:

$$S = \sum_{i=1}^n w_i x_i - T = (W, X) - T = |W| \cdot |X| \cdot \cos \alpha - T,$$

где T – порог нейрона (во многих моделях обходятся без него), F – функция активации нейрона, преобразующая взвешенную сумму в выходной сигнал $y = F(S)$. Взвешенная сумма входных сигналов может быть интерпретирована как проекция входного вектора на вектор весов, где α – угол между этими векторами.

Формальный нейрон моделирует некоторые свойства биологического нейрона. Набор связанных формальных нейронов представляет собой искусственную нейронную сеть. Искусственные нейронные сети способны выполнять любые логические операции и вообще любые преобразования, реализуемые дискретными устройствами с конечной памятью (другой вопрос в том, как настроить веса такой сети). Нейроны в естественных нейронных сетях намного сложнее, их функционирование является сложным процессом, протяжённым во времени. Кроме того, существует мнение, что мозг обладает квантовой структурой, а процесс мышления основан на квантовых эффектах. Мозг человека состоит из 10 триллионов нейронов, связанных между собой 10^{14} синапсами. Такие вычислительные мощности современной вычислительной технике пока недоступны. Структура мозга определена генетически от рождения, а связи между нейронами развиваются и модифицируются на протяжении всей жизни, т.е. свой интеллектуальный опыт человек получает в процессе обучения. Это говорит о перспективности развития искусственных нейронных сетей.

По характеру связей нейронные сети могут быть полносвязными, когда каждый нейрон связан со всеми остальными, и слоистыми, когда нейроны последующего слоя связаны только со всеми нейронами предыдущего слоя. Эти две архитектуры являются базовыми, но возможны и различные вариации.

По характеру функционирования нейронные сети могут быть однопроходными, когда выход сети рассчитывается за один проход сети, и релаксационными, когда функционирование сети продолжается до достижения стабильного состояния, это состояние и является результатом работы.

По характеру формирования связей нейронные сети могут быть следующих видов [6].

- Обучение с учителем. Связи настраиваются в процессе обучения, причём эталонные значения результатов работы известны.
 - Самообучение (обучение без учителя). Эталонные результаты неизвестны (не нужны), сеть в процессе обучения должна организовать входные образы на основе их подобия.
 - Фиксированные связи. Определяются характером решаемой задачи (например, в оптимизационных задачах).
- Нейронные сети могут также отличаться типом входной информации (двоичная, аналоговая и т.п.) и методом обучения.

2.2. Разделение пространства признаков на области и извлечение ключевых признаков

Для решения данных задач предназначены многослойные нейронные сети, нейронные сети высокого порядка и радиально-базисные нейронные сети.

Поскольку такие сети оперируют в исходном пространстве изображений (признаков), то для них является критическим требование предобработки изображения. Это приведение изображения к стандартному виду (положение, масштаб, ориентация, выравнивание яркости), снижение размерности данных, выбор ключевых характеристик.

Следующим следствием оперирования в исходном пространстве является невозможность учёта искажения изображения (например, при изменении ракурса, эмоций), и поэтому тренировочная выборка должна содержать репрезентативный набор примеров, представляющих собой наборы изображений объектов в том диапазоне ракурсов и условий освещения, в которых планируется применение системы распознавания [56].

2.2.1. Многослойные нейронные сети

Архитектура многослойной нейронной сети (МНС, другое название – многослойный персептрон, по английски Multilayer Perceptron, **MLP**) состоит из последовательно соединённых слоёв, где нейрон каждого слоя своими входами связан со всеми нейронами предыдущего слоя, а выходами – следующего (рис. 2). Активационными функциями для таких нейронов служат разновидности линейных, пороговых и сигмоидных функций [6].

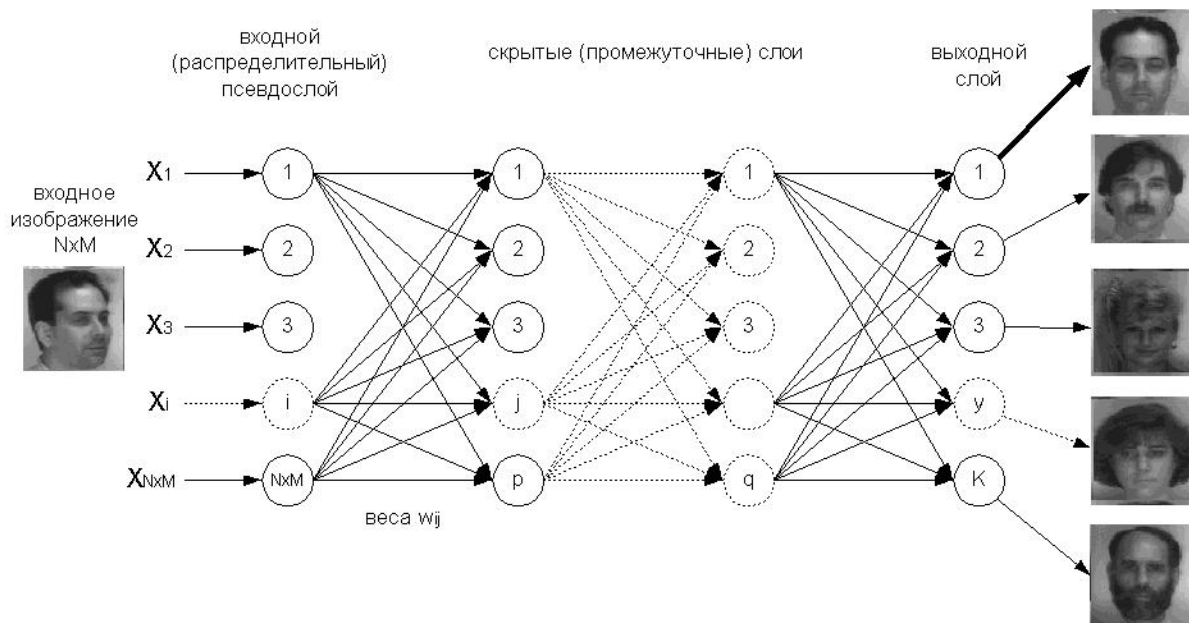


Рис. 2. Архитектура многослойной нейронной сети и её применение для распознавания изображений. Нейрон с максимальной активностью (здесь первый) указывает принадлежность к распознанному классу

НС с одним решающим слоем способна формировать линейные разделяющие поверхности, что сильно сужает круг решаемых задач, в частности, такая сеть не сможет решить задачу типа «исключающее или». НС с нелинейной функцией активации и двумя решающими слоями позволяет формировать любые выпуклые области в пространстве решений, а с тремя решающими слоями - области любой сложности, в том числе и невыпуклой формы [6]. При этом МНС не теряет своей обобщающей способности. С помощью двухслойной НС можно с любой точностью аппроксимировать любую многомерную функцию на отрезке $[0;1]$. Обучаются МНС при помощи алгоритма обратного распространения ошибки, являющегося разновидностью градиентного спуска в пространстве весов с целью минимизации суммарной ошибки сети:

$$\Delta W = -\alpha \frac{dE}{dW}, \quad E = \frac{1}{2} \sum_j (y_j - t_j)^2, \quad \text{где } t_j \text{ — эталонное значение выходов}$$

сети. При этом ошибки (точнее, величины коррекции весов) распространяются в обратном направлении от входов к выходам, сквозь веса, соединяющие нейроны. Алгоритм обратного распространения ошибки является NP -трудным, поэтому время обучения сети увеличивается экспоненциально с ростом размерности данных.

Так как эталонные значения выходов известны, алгоритм является обучением с учителем. Но в применении к извлечению ключевых признаков, когда рециркуляционная сеть обучается реконструировать поданное на вход изображение, а на скрытых нейронах формируется его сжатое представление, алгоритм обучения может быть назван и самообучением.

МНС, как и большинство других типов НС, перед началом обучения инициализируется случайными весами. Поэтому две разные обученные НС, имеющие одинаковый показатель ошибки, часто представлены совершенно различными разделяющими поверхностями, не сводимыми друг к другу. На этом основан метод коллективов (ансамблей) нейронных сетей, часто применяемый в распознавании человека по изображению лица [34,38]. Суть его заключается в том, что имеется набор (коллектив) сетей, обученных решать одну и ту же задачу, но различными способами (различная начальная инициализация весов, архитектура, порядок следования примеров при обучении и т.п.). Обобщенное решение такого коллектива как правило точнее (и надёжнее), чем решение единственной нейронной сети. Существуют также методики синтеза коллектива сетей, производящих максимально независимые ошибки [34].

Другое направление развития архитектуры МНС – это нейронные деревья [33]. В этой архитектуре узлами решающего дерева являются нейронные сети. По мере продвижения от корня дерева сети-узлы уточняют решение задачи. В работе [33] был разработан новый алгоритм создания и обучения нейронных деревьев, который применялся для классификации сцен на изображении. По сравнению с такими методами, как обычные нейронные деревья, решающие деревья, коллективы НС и многослойные нейронные сети, точность распознавания у предложенного алгоритма была сравнима или выше, а скорость обучения и работы – на порядок выше. Однако обобщающая способность такой архитектуры в [33] не проверялась.

Одними из главных проблем МНС (и некоторых других типов НС) являются следующие [3,6,9,10,21].

1. Проблема локального минимума. Как и для всех градиентных методов, проблема локального минимума заключается в том, что при итерационном спуске может наступить момент, когда решение заходит в локальный минимум, из которого вследствие малой величины шага не может выбраться. И такой локальный минимум не всегда обеспечивает приемлемое решение. Выход заключается в применении стохастических методов [3,21].

2. Выбор архитектуры сети (количество нейронов, слоёв, характер связей). С этим также связана проблема переобучения, которая заключается в том, что сеть с избыточным числом элементов теряет обобщающую способность и хорошо работает только на тренировочной выборке. В настоящее время разработаны различные априорные оценки выбора архитектуры, методы прореживания обученных сетей [9], методы «растущих» сетей [9].

3. Выбор шага (скорости) обучения. Такая проблема связана с тем, что при малом шаге время обучения будет большим и сеть может застревать в локальных минимумах, а при больших шагах возможно расхождение процесса обучения или паралич сети. Проблема эффективно решается адаптивным шагом [6], который для каждой итерации позволяет сделать шаг, минимизирующий ошибку сети на данной итерации. Существуют методы, которые на каждом тренировочном цикле (называемом эпохой) анализируют всю тренировочную выборку и выбирают оптимальное значение и направление шага [10].

Одним из наиболее перспективных методов, применяемых на этапе обучения НС, является генетический алгоритм (ГА, по английски Genetic Algorithm, **GA**), относящийся к эволюционным методам [18].

Генетический алгоритм представляет собой параллельный асинхронный оптимизационный метод [18]. Поиск решения в нём осуществляется одновременно целой популяцией хромосом (хромосома – единичное закодированное решение задачи). Хромосомы могут как обмениваться опытом друг с другом, улучшая свою пригодность (оператор скрещивания), так и осваивать новые области решения (оператор мутации). Оптимизационный критерий задаёт функцию оценки пригодности хромосом, и процесс эволюции популяции представляет собой улучшение решения задачи. Главным преимуществом ГА является то, что при линейном увеличении размера популяции скорость поиска решения возрастает экспоненциально (так называемый скрытый параллелизм [18]). Это позволяет достичь лучших субоптимальных решений.

Поскольку обучение сети – это оптимизационный процесс, генетический алгоритм естественным образом вкладывается в алгоритм обучения НС [3]. При этом для ускорения ГА может быть введён оператор локального (градиентного) спуска, представляющий собой метод обратного распространения в применении к отдельной хромосоме, каждая из которых является самостоятельной нейронной сетью.

Закодировав в хромосому структуру сети и введя в оценочный критерий ГА штраф за избыточность архитектуры, можно добиться синтеза

НС с минимально необходимой архитектурой, обладающей хорошими обобщающими способностями [3].

Рассмотрим применение многослойных нейронных сетей к распознаванию человека по изображению лица.

Простейшее применение однослойной НС [57] (называемой автоассоциативной памятью) заключается в обучении сети восстанавливать подаваемые изображения. Подавая на вход неизвестное изображение и вычисляя качество реконструированного изображения, можно оценить, насколько сеть распознала входное изображение. Положительные свойства этого метода заключаются в том, что сеть может восстанавливать искажённые и зашумленные изображения, но для более серьёзных целей он не подходит.

МНС также используется для непосредственной классификации изображений – на вход подаётся или само изображение в каком-либо виде, или набор ранее извлечённых ключевых признаков изображения, на выходе нейрон с максимальной активностью указывает принадлежность к распознанному классу (см. рис. 2). Если эта активность ниже некоторого порога, то считается, что поданный образ не относится ни к одному из известных классов. Процесс обучения устанавливает соответствие подаваемых на вход образов с принадлежностью к определённому классу. Это называется обучением с учителем. В наших экспериментах на тестовой базе ORL такой подход позволил достичь стабильной 93%-ной точности распознавания (98%-ной максимальной). В применении к распознаванию человека по изображению лица такой подход хорош для задач контроля доступа небольшой группы лиц. Он обеспечивает непосредственное сравнение сетью самих образов, но с увеличением числа классов время обучения и работы сети возрастает экспоненциально. Поэтому такие задачи, как поиск похожего человека в большой базе данных, требуют извлечения компактного набора ключевых характеристик, на основе которых можно производить поиск.

В работе [39] для классификации изображений лиц использовался многослойный персептрон. В качестве входных данных применялись результаты преобразования симметрии. При использовании коллектива нейронных сетей точность распознавания повышалась. Точность повышалась и при предварительной обработке входных данных самоорганизующимися сетями. Всё это позволило достичь 0.62% ошибки на базе данных MIT.

В работе [44] для классификации также использовалась многослойная нейронная сеть. Но входными данными являлось изображение, подвергнутое многомерному снижению размерности, и результаты приме-

нения различных вариаций Гауссовских фильтров. Такой подход позволил достичь высокой обобщающей способности, имея в обучающей выборке только одно изображение на каждого человека.

Подход к классификации с использованием частотных характеристик всего изображения описан в [26]. Применялась однослойная НС, основанная на многозначных нейронах, которая, как и многослойный персептрон, способна строить сложные разделяющие поверхности, но по иному принципу. Спектральные коэффициенты преобразования Фурье выбирались по правилу «зигзага», и результат стабилизировался при числе первых коэффициентов начиная с 405. Отмечено 100%-ное распознавание на базе данных MIT.

В работе [48] использовались первые 25-35 коэффициентов дискретного косинусного преобразования (из 10304, т.е. около 0.2% от общего количества коэффициентов) и двухслойная нейронная сеть. Коэффициенты выбирались двумя способами. В первом способе брались первые коэффициенты из левого верхнего угла матрицы коэффициентов, ограниченного диагональю. Во втором способе выбирались коэффициенты, которые давали наименьшую погрешность реконструкции изображения. Разница в точности при этом была неразличима. И хотя реконструкция по таким коэффициентам невозможна (рис. 3), точность распознавания составила в среднем 96% (на базе ORL). Использовалось 60 и 80 нейронов в скрытом слое (всего два слоя, в выходном – 40 нейронов, по числу классов).



Рис. 3. Реконструкция по первым 28 из 10304 коэффициентов, изображение 92x112

Близкий результат (в среднем 94% точности на той же базе ORL) был воспроизведён в наших экспериментах. Причём было отмечено, что с уменьшением числа коэффициентов число тренировочных циклов требуется увеличить, но в целом время обучения при этом снижается.

Применение МНС для классификации изображений лиц на основе таких характеристик, как расстояния между некоторыми специфическими областями лица (нос, рот, глаза), описано в [59]. В этом случае на

вход НС подавались эти расстояния. Использовались также комбинированные методы: в первом из них на вход НС подавались результаты обработки скрытой Марковской моделью, а во втором – результат работы НС подавался на вход скрытой Марковской модели. Во втором случае преимуществ не наблюдалось. Это говорит о том, что скрытые Марковские модели сами по себе обладают худшей различающей способностью, чем НС.

В работе [51] показано применение НС для классификации изображений, когда на вход сети поступают результаты декомпозиции изображения по методу главных компонент.

Аналогичный эксперимент был проведён нами (подробности в разд. 3). Классификация главных компонент при помощи НС повышала точность распознавания.

МНС применяются и для обнаружения объектов определённого типа. Помимо того что любая обученная МНС в некоторой мере может определять принадлежность образов к «своим» классам, её можно специально обучить надёжному детектированию определённых классов. В этом случае выходными будут два класса: класс принадлежащих и класс не принадлежащих к заданному типу образов. В [53] применялся нейросетевой детектор для обнаружения изображения лица во входном изображении. Изображение сканировалось скользящим окном 20x20 пикселей, которое подавалось на вход сети, решающей, принадлежит ли данный участок к классу лиц. Обучение производилось с использованием как положительных примеров (различных изображений лиц), так и отрицательных (изображений, не являющихся лицами). В обучающей выборке незначительно варьировались масштаб, положение и ориентация лица. Для повышения надёжности обнаружения использовался коллектив НС, обученных с различными начальными весами, вследствие чего НС ошибались по-разному, а окончательное решение принималось голосованием всего коллектива.

НС применяется также для извлечения ключевых характеристик изображения, которые затем используются для последующей классификации. В [6,57], показан способ нейросетевой реализации метода главных компонент. Суть метода главных компонент заключается в получении максимально декоррелированных коэффициентов, характеризующих входные образы. Такие коэффициенты называются главными компонентами и используются для статистического сжатия и реконструкции изображений. При этом небольшое число коэффициентов используется для представления всего образа. Каждое изображение разлагается на линейную комбинацию собственных векторов. Для набора изображений

лиц собственные векторы могут быть представлены в виде изображений, такие изображения похожи на лица и называются собственными лицами (*eigenfaces*). Сумма собственных векторов, умноженных на соответствующие им главные компоненты, представляет собой реконструкцию изображения.

НС с одним скрытым слоем, содержащим m нейронов, число которых много меньше, чем размерность изображения $m \ll n$, обученная по методу обратного распространения ошибки восстанавливать на выходе изображение, поданное на вход, формирует на выходе скрытых нейронов коэффициенты первых m главных компонент, которые и используются для сравнения изображений. Архитектура такой сети, называемой рециркуляционной нейронной сетью (РНС), показана на рис. 4.

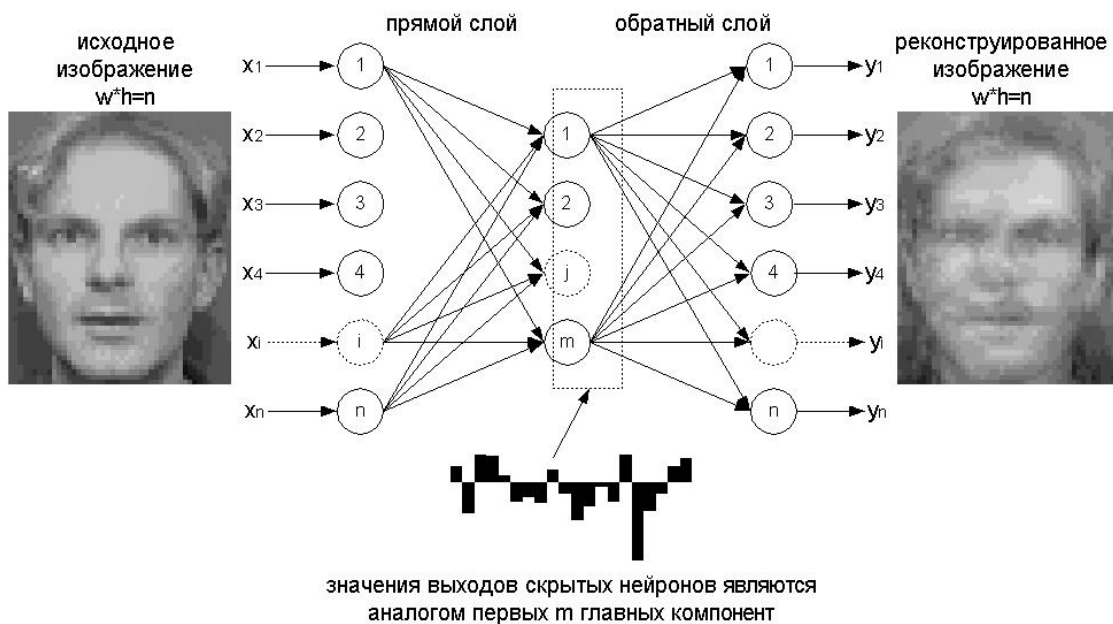


Рис. 4. Архитектура рециркуляционной нейронной сети для извлечения главных компонент

Обычно используется от 10 до 200 главных компонент. С увеличением номера компоненты её репрезентативность сильно понижается, и использовать компоненты с большими номерами не имеет смысла. Использование линейных активационных функций в НС позволяет получить на выходе скрытого слоя именно m первых главных компонент, аналогичных получаемым при решении матричных уравнений. При использовании нелинейных активационных функций нейронных элементов возможна нелинейная декомпозиция на главные компоненты. Нелинейность позволяет более точно отразить вариации входных данных, од-

нако при этом выходы скрытых нейронов будут только похожи на главные компоненты. Веса, сформировавшиеся при таком обучении на входном и выходном слое, также будут похожи на собственные лица (*holons* в работе [57]), которым присуще полезное свойство – существуют компоненты, которые в основном отражают такие существенные характеристики лица, как пол, раса, эмоции. Первые компоненты отражают наиболее общую форму лица, последние – различные мелкие отличия между лицами (см. разд. 3.1.3). Такой метод хорошо применим для поиска похожих изображений лиц в больших базах данных. Этот метод также используется в задаче обнаружения лица на изображении. Оценивая качество реконструкции входного изображения, можно очень точно определять его принадлежность к классу лиц. Для изображений, не являющихся лицами, реконструкция будет невысокого качества.

В этой же работе [57] использовалась МНС для дальнейшей классификации главных компонент. Производилась идентификация человека по изображению лица, определение его расы, пола, эмоционального состояния.

Преимущества применения РНС для извлечения главных компонент перед решением матричных уравнений [28]:

- алгоритм обучения РНС прост и универсален;
- нелинейная активационная функция позволяет точнее реконструировать изображение;
- при решении матричных уравнений возможны проблемы, если примеры очень похожи друг на друга, РНС лишена такого недостатка;
- не требуется вычислять все собственные векторы. Таким образом, время обучения сети линейно зависит от количества извлекаемых главных компонент;
- для предварительных экспериментов можно использовать меньшее число обучающих циклов, что снижает время обучения.

Показана возможность дальнейшего уменьшения размерности главных компонент при помощи НС [57]. Суть её заключается в использовании большего количества скрытых слоёв, отвечающих за сжатие и реконструкцию изображения. Такое сжатие позволяет уловить более сложные закономерности в наборе образов и, следовательно, представить их точнее и меньшим числом компонент.

2.2.2. Нейронные сети высокого порядка и моментные НС

Нейронные сети высокого порядка (НСВП, по английски High Order Neural Network) отличаются от МНС тем, что у них только один слой, но

на входы нейронов поступают также члены высокого порядка, являющиеся произведением двух или более компонент входного вектора [6],

например: для сетей второго порядка $S = \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j - T$. Та-

кие сети также могут формировать сложные разделяющие поверхности. Разделяющая поверхность второго порядка $S = 0$ называется гиперквадрикой [6]. Добавляя компоненты входного вектора в произведение, получим класс полиномиальных разделяющих поверхностей. Такие сети также можно обучать по методу обратного распространения. Многослойные НС в общем случае эффективнее, но существует ряд приложений, в которых сети высокого порядка лучше, чем МНС.

В работе [4] описаны перспективные архитектуры и методы обучения нейронных сетей высокого порядка и комбинированных нейронных сетей для распознавания изображений инвариантно к сдвигу, масштабу и повороту. Архитектура таких сетей высокого порядка основана на вычислении моментов изображения первыми слоями.

В работе [32] показано применение НСВП строго третьего порядка для распознавания изображений лиц, имеющих произвольные масштаб и ориентацию изображения. Приведены методы обучения такой сети. Особенность её заключаются в том, что для обучения некоторому классу достаточно предъявить его образ без вариаций масштабов и поворотов – после обучения сеть будет распознавать известные классы инвариантно к масштабу и поворотам изображения. Такая сеть не является полносвязной, быстро обучается и работает. Отмечено существенное повышение точности классификации такой сетью повернутых и масштабированных изображений по сравнению с МНС.

2.2.3. Радиально-базисные нейронные сети

Радиально-базисные нейронные сети (РБНС, по английски Radial Basis Function Network, **RBF**) продолжают идею карт Кохонена и состоят из двух слоёв (рис. 5) [3]. Первый слой имеет радиально-базисную ак-

тивационную функцию: $y = \exp\left(\frac{-S^2}{2\sigma^2}\right)$, где σ – среднеквадратичное отклонение, характеризующее ширину функции (размер кластера). S определяется как расстояние между входным и весовым вектором: $S^2 = |X - W|^2 = \sum_i (x_i - w_i)^2$, являющимся по сути расстоянием до цен-

тра кластера, определяемым конкретным нейроном [7]. Таким образом, скрытый слой представляет собой набор кластеров в пространстве обра-

зов и выполняет первый этап кластеризации входного образа – значение активационной функции каждого нейрона быстро уменьшается с удалением от центра кластера. Второй слой нейронов имеет линейную активационную функцию и выполняет второй этап кластеризации – распределяет кластеры по классам. В отличие от карт Кохонена здесь обнуление нейронов, не обладающих максимальным выходным значением, не требуется, они все вносят вклад в классификацию, и это преимущество РБНС.

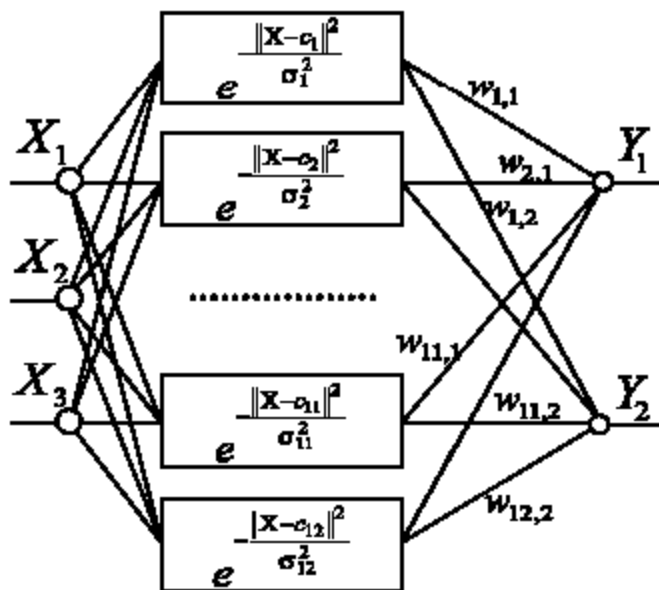


Рис. 5. Радиально-базисная нейронная сеть. X_i – элементы входного вектора, Y_i – элементы выходного вектора, квадратами обозначены радиально-базисные нейроны

РБНС также способна строить сложные разделяющие области и аппроксимировать многомерные функции. По сравнению с многослойной нейронной сетью, радиально-базисная сеть обучается на порядок быстрее, однако обладает намного худшей экстраполирующей способностью, т.е. не способна работать на образах, лежащих далеко от образов-примеров. Размеры РБНС больше, чем МНС для аналогичных задач, и РБНС становятся малоэффективны с ростом размерности входных данных [24].

Обучается такая сеть в два этапа. Первый этап осуществляется без учителя, на нём первый слой выделяет компактно расположенные группы кластеров. При этом корректируются центры кластеров. В настоящее время разработаны эффективные алгоритмы, позволяющие также подбирать оптимальный размер кластеров для каждого нейрона [3,13] и по-

лучать оптимальное количество нейронов в первом слое [3]. На втором этапе обучения второй слой учится распределять входные образы, пропущенные через первый слой, по классам. Информация об эталонных значениях выходов известна, обучение выполняется с учителем. Такое обучение производится или матричными методами, или алгоритмом обратного распространения ошибки [3,7,9].

Рассмотрим применение радиально-базисных нейронных сетей к распознаванию человека по изображению лица.

В работе [51] применялись два способа извлечения ключевых характеристик и две различные архитектуры РБНС для распознавания лиц. В первом способе характеристики представляли собой набор главных компонент, во втором – коэффициенты вейвлетных преобразований. В первой архитектуре количество выходов соответствовало количеству классов, во второй применялся коллектив сетей, каждая из которых была обучена распознавать только свой класс. Отмечены значительные преимущества классификации РБНС перед непосредственным сравнением ключевых характеристик на тестовой базе MIT.

В работе [38] применялись две различные архитектуры ансамблей РБНС для предварительной классификации изображений. На вход сети поступало изображение целиком, на выходах формировалась промежуточная классификация, которая затем подавалась на решающие деревья для контекстно-ориентированного распознавания изображений лиц (например: «найти все изображения определённого человека, где он в очках»). Различные сети в ансамблях первой архитектуры учились классифицировать изображения с различными типами изменений, второй – с одинаковыми, но количество нейронов менялось в процессе обучения. Решающий вывод делал «судья» (нейронная сеть, обученная обобщать решения коллектива НС), который принимал решение на основе голосования ансамбля сетей.

В работе [27] радиально-базисные сети использовались для распознавания человека по набору геометрических характеристик и определения его пола. Исходными служили полутонные и рисованные (карикатурные) изображения. Отмечена хорошая способность РБНС выделять отличительные признаки.

В работе [52] использовалась РБНС совместно с оптическим потоком для анализа эмоционального выражения лица.

2.3. Ассоциативная память и решение оптимизационных задач

Нейронная сеть Хопфилда (НСХ, Hopfield Neural Network) [6,9,10] является однослойной и полносвязной (связи нейронов на самих себя отсутствуют), её выходы связаны со входами. В отличие от МНС, НСХ является **релаксационной** – т.е. будучи установленной в начальное состояние, функционирует до тех пор, пока не достигнет стабильного состояния, которое и будет являться её выходным значением. НСХ применяются в качестве ассоциативной памяти и для решения оптимизационных задач. В первом случае НСХ обучается без учителя (например, по правилу Хебба), во втором случае веса между нейронами изначально кодируют решаемую задачу. НСХ бывают синхронными, когда одновременно пересчитываются значения выходов всех нейронов, и асинхронными, когда пересчитывается один случайно выбранный нейрон. Для исследования динамики функционирования НСХ используются методы Ляпунова. В [6] показано, что асинхронная НСХ всегда сходится к устойчивым точкам, а аттракторами синхронной НСХ являются устойчивые стационарные точки и предельные циклы длины два. Таким образом НСХ из начального состояния сходится к ближайшему локальному минимуму энергии сети, состояние нейронов в котором и будет восстановленным образом для задач распознавания и решением – для оптимизационных задач. Для поиска глобального минимума применительно к оптимизационным задачам используют стохастические модификации НСХ [6]. В этом случае значение выхода нейрона зависит не только от состояния его входов, но и от температуры сети. Чем выше температура, тем больше вероятность, что нейрон выберет своё следующее состояние случайно. В процессе функционирования сети температура понижается.

В классических сетях Хопфилда число хранимых образов составляет порядка 0.14 от числа нейронов. Каждый образ кодируется энергетическим минимумом в потенциальном поле сети Хопфилда. Превышение этого числа ведёт к появлению интерференционных минимумов, представляющих собой ложные образы – локальные минимумы на пересечении нескольких образов. Для преодоления этого недостатка используют ортогонализацию образов [6], нейроны более высоких порядков [6], тензорные НС Хопфилда [10] и методы разобучения, которые уменьшают влияние ложных минимумов, «заглушая» их [9,50]. Кроме того, ложные минимумы могут быть использованы для обнаружения новых классов образов, что относится к преимуществам сетей Хопфилда.

Существуют модификации сетей Хопфилда – это сети Хэмминга и двунаправленная ассоциативная память [6].

Преимущества применения сетей Хопфилда в качестве ассоциативной памяти следующие. С увеличением числа образов время восстановления эталона не увеличивается в отличие от простого поиска и сравнения с каждым образом. Кроме того, слабо искажённые образы сеть восстановит быстрее.

Так как сети Хопфилда оперируют с исходным пространством изображения (или ключевых характеристик), то требования к предобработке изображений для них критично. Существуют автокорреляторные сети Хопфилда, инвариантные к сдвигу изображения [10].

Применение НСХ в качестве ассоциативной памяти позволяет точно восстанавливать образы, которым сеть обучена, при подаче на вход искажённого образа. При этом сеть «вспомнит» наиболее близкий (в смысле локального минимума энергии) образ, и таким образом распознает его. Такое функционирование можно приблизительно представить как последовательное применение автоассоциативной памяти, описанной выше. В отличие от автоассоциативной памяти, НСХ точно восстанавливает образ.

В работе [29] описано применение НСХ, где распознаваемое изображение лица сначала подавалось на НСХ, результат работы которой затем сравнивался с хранимыми изображениями. Отмечено, что предварительная обработка входного изображения при помощи НСХ существенно повышает точность распознавания.

Применение НСХ в качестве оптимизационного метода для восстановления пространственной формы лица по двумерному полутоновому изображению (а также для детектирования граней) описано в [45]. Здесь используется модификация НСХ – ячеистая нейронная сеть, в которой нейроны связаны только с ближайшими соседями из двумерной области. Матрица связей одинакова для всех нейронов и определяется характером задачи. Отмечены хорошие результаты восстановления трёхмерной формы объектов (в частности для изображений лиц) и высокая скорость работы.

2.4. Топологически упорядоченное преобразование пространства

Векторные квантователи и самоорганизующиеся карты Кохонена (Self-Organizing Maps, **SOM**) [6,9] используются для сжатия данных и извлечения ключевых характеристик. Они также служат основой для радиально-базисных сетей и когнитронов. Существуют, однако, и подходы, использующие карты Кохонена для решения оптимизационных за-

дач (например, задачи коммивояжера), что не исключает их дальнейшего применения для распознавания изображений.

Векторные квантователи решают задачу квантования и кластеризации данных [6,9]. Архитектура таких сетей состоит из одного слоя (не считая входного распределительного), не имеет межнейронных связей, и является простейшим вариантом карт Кохонена. Входное пространство оптимальным образом разбивается на области-кластеры. В процессе обучения происходит выделение областей в исходном пространстве, каждой из которых соответствует отдельный нейрон. Такие сети функционируют по принципу «победитель берёт всё», активным считается нейрон, имеющий наибольшее выходное значение $S_j = \sum_i w_{ij} x_i = W_j^T X$, т.е.

проекция входного вектора на веса которого оказалась максимальной. Выходом сети является **номер** нейрона-победителя.

Обучаются такие сети без учителя, т.е. они сами в процессе обучения выбирают оптимальное разбиение на области. Метод обучения таких сетей называется **конкурентным**. После подачи обучающего вектора выбирается нейрон, имеющий максимальную активность. Веса такого нейрона изменяются в сторону соответствия входному вектору, например: $W_j(t+1) = W_j(t) + \gamma(t)(X - W_j(t))$. Перед обучением веса инициализируются случайными значениями. Обучение ведётся до достижения стабилизации весов или завершается после определённого числа итераций.

Для того чтобы избежать ситуации, в которой некоторые нейроны никогда не могут стать победителями, и для более полного покрытия пространства (например, в областях, где плотность образов высока, требуется большее число нейронов) используются несколько подходов. Во-первых, можно модифицировать веса проигравших нейронов с намного меньшей скоростью. Во-вторых, можно вести статистику побед для каждого нейрона и уменьшать возможности модификации слишком частых победителей.

Для автоматизации определения числа нейронов сети используют такие алгоритмы, как, например, растущий нейронный газ. В нём нейроны, соответствующие большому числу примеров или имеющие большой размер ячейки-кластера, делятся на два нейрона [9].

Самоорганизующиеся карты Кохонена [6,9] позволяют получить топологически упорядоченное преобразование исходного n -мерного пространства в выходное m -мерное, $m \ll n$. Архитектура такой сети также состоит из одного слоя, но нейроны в этом слое организованы в m -мерную решётку и каждый нейрон имеет свою координату, опреде-

ляющую его положение в решётке. Выходом сети являются координаты нейрона-победителя.

В правило модификации весов

$$W_j(t+1) = W_j(t) + \gamma(t)h(t, i, j)(X - W_j(t))$$

вводится функция соседства h , убывающая с расстоянием между нейронами:

$$h(t, i, j) = \exp\left(\frac{-|i - j|^2}{2\sigma^2(t)}\right),$$

где i – нейрон-победитель; j – модифицируемый нейрон; $|i - j|$ – расстояние между ними; $\sigma(t)$ – радиус области соседства, убывает со временем в процессе обучения. Для ускорения обучения используются растущие сети, когда заново создаваемые области решётки инициализируются значениями соседних нейронов.

Таким образом, обученная сеть способна **топологически упорядоченно** отображать входное пространство в выходное – векторы, близкие в исходном пространстве, будут иметь близкие координаты нейронов в решётке. Это является особенно полезным при классификации данных, имеющих большое количество классов. Например, при классификации локальных участков изображений может существовать очень большое число классов, в которых переход от одного класса к другому практически непрерывен, что затрудняет определение границ классов.

В работе [41] трёхмерная карта Кохонена (по пять узлов на каждое измерение) применялась для уменьшения размерности локальных участков 5×5 (размерность 25) изображений лиц (рис. 6). Каждому участку изображения 5×5 соответствует своя координата в карте Кохонена. После обучения похожие участки имеют близкое положение на карте. Входное изображение отображается на один из 125 узлов, положение которого в трёхмерной решётке кодирует вектор выходного пространства. Три измерения карты используются в качестве осей трёх ключевых характеристик (*features*) [41]. Такое преобразование обеспечивает частичную устойчивость к изменению освещения, смещениям и искажениям, избавляет от необходимости предварительной обработки изображения (преимущество – ускорение работы), а также значительно ускоряет процесс обучения и классификации, делая эту систему применимой в реальном времени. В этой работе значения выходов карт Кохонена использовались для дальнейшего распознавания по изображениям лиц. Отмечено небольшое преимущество карт Кохонена перед методом глав-

ных компонент, которое заключалось в более высокой точности последующей классификации на основе данных уменьшенной размерности.

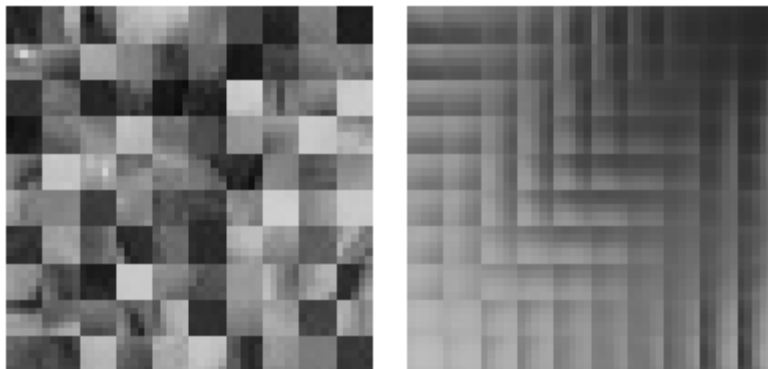


Рис. 6. Применение карт Кохонена для уменьшения размерности участков изображений лиц. Слева – топографическая карта участков изображений до обучения, справа – после обучения

Карты Кохонена также применялись для обнаружения глаз на изображении лица [55]. Карта обучалась на типичных примерах изображений глаз. Наличие глаза в анализируемом участке изображения определялось по карте активности всех нейронов, в этом случае на ней наблюдались характерные пики.

2.5. Распознавание с учётом топологии пространства

Описываемые далее типы нейронных сетей позволяют учесть топологию пространства изображения. Ведь изображение – это не просто n -мерный вектор, составленный из яркостей пикселей. Изображение имеет топологию, определяемую через двумерное локальное соседство пикселей [19].

Принципы работы таких сетей основываются на разбиении изображения на маленькие участки и иерархическом сопоставлении как взаимного их расположения, так и содержания.

Такие сети являются наиболее перспективными для распознавания изображений.

2.5.1. Когнитрон

Когнитрон [21] был разработан на основе анатомии и физиологии мозга, и своей архитектурой похож на строение зрительной коры. Каждый слой мозга реализует различные уровни обобщения: входной слой

чувствителен к простым образам, таким, как линии и их ориентации в определенных областях визуальной области, в то время как реакция других слоев является более сложной, абстрактной и независимой от позиции образа. Аналогичные функции реализованы в когнитроне путем моделирования организации зрительной коры.

Главные архитектурные отличия когнитрона заключаются в том (рис. 7), что каждый нейрон связан только с небольшой локальной областью предыдущего слоя, и такие области перекрываются друг с другом. Слоёв в когнитроне обычно больше, чем в сетях других типов. Таким образом достигается иерархическая организация, когда на высших слоях когнитрон реагирует на более абстрактные образы. При этом когнитрон меньше реагирует на смещение и искажение образов.

Обучается когнитрон при помощи конкурентного обучения (без учителя).

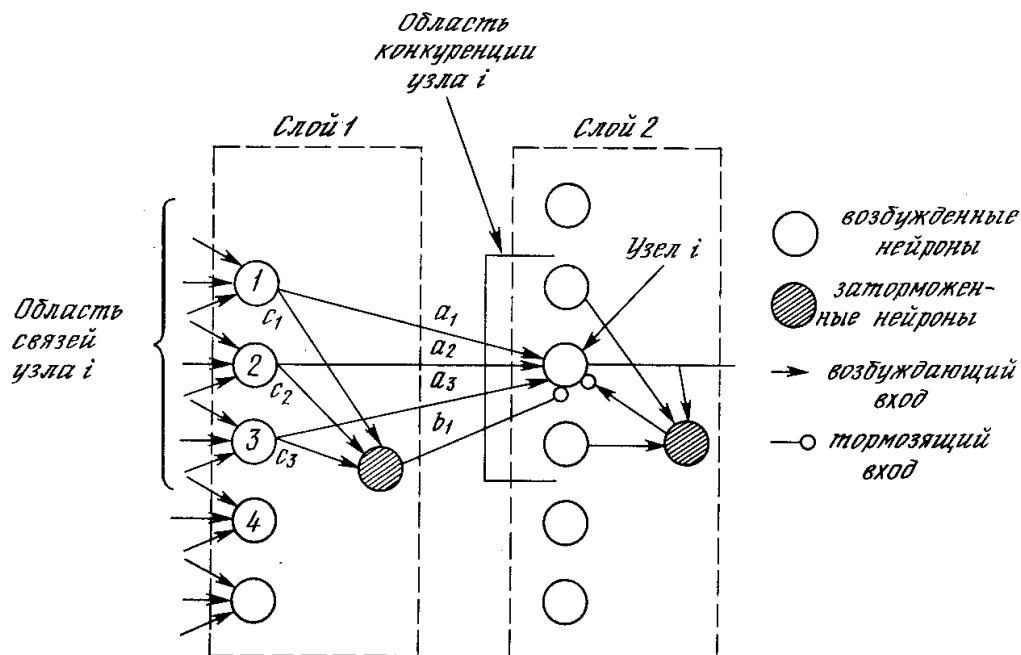


Рис. 7. Когнитрон

2.5.2. Неокогнитрон

В зрительной коре были обнаружены узлы, реагирующие на такие элементы, как линии и углы определенной ориентации [21]. На более высоких уровнях узлы реагируют на более сложные и абстрактные образы, такие, как окружности, треугольники и прямоугольники. На еще бо-

лее высоких уровнях степень абстракции возрастает до тех пор, пока не определяются узлы, реагирующие на лица и сложные формы. В общем случае узлы на более высоких уровнях получают вход от группы низкоуровневых узлов и, следовательно, реагируют на более широкую область визуального поля. Реакции узлов более высокого уровня менее зависят от позиции и более устойчивы к искажениям.

Неокогнитрон [21] является дальнейшим развитием идеи когнитрона и более точно моделирует строение зрительной системы, позволяет распознавать образы независимо от их преобразований: смещения, вращения, изменения масштаба и искажения. Неокогнитрон может как самообучаться, так и обучаться с учителем. Неокогнитрон получает на входе двумерные образы, аналогичные изображениям на сетчатой оболочке глаза и обрабатывает их в последующих слоях аналогично тому, как это было обнаружено в зрительной коре человека.

Главное отличие неокогнитрона от когнитрона – это двумерная организация локальных участков и плоскостная иерархическая структура (рис. 8).

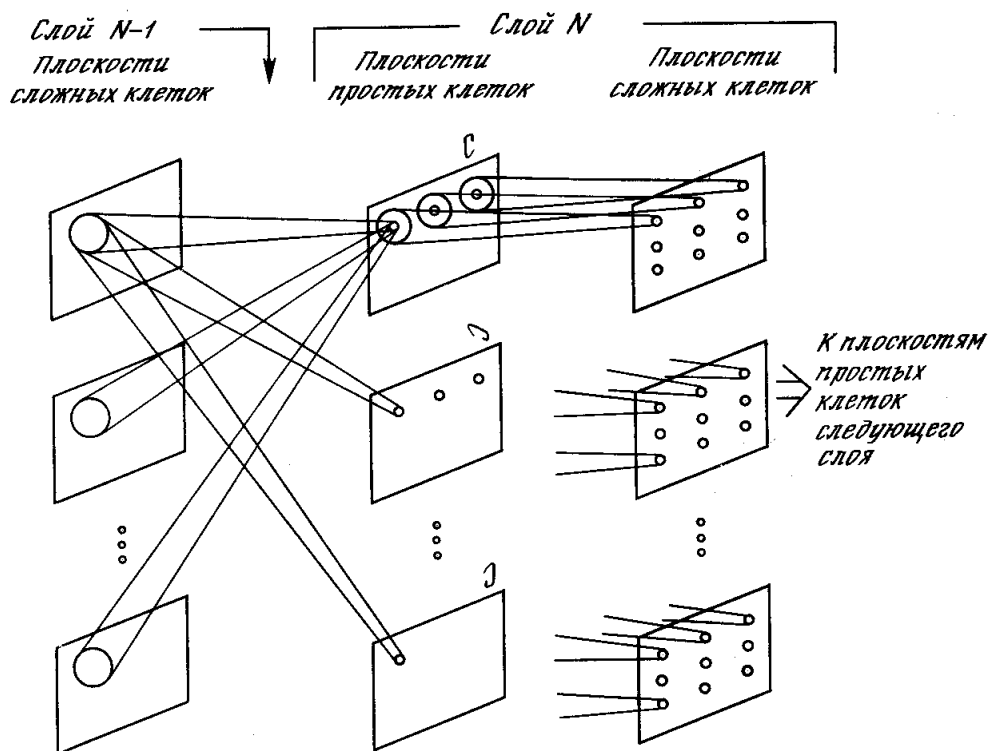


Рис. 8. Неокогнитрон

Каждый слой состоит из плоскостей простых и сложных клеток. Каждый нейрон простой плоскости связан с локальным двумерным участком плоскостей предыдущего слоя, веса всех нейронов в пределах одной плоскости одинаковы, и таким образом плоскость реагирует на определённый образ, находящийся в участке изображения (пример на рис. 8 – плоскости реагируют на букву «С», повернутую под разными углами). Положение активированного таким образом нейрона в простой плоскости отмечает участок, в котором найден этот образ, независимо от его искажения. Нейрон сложной плоскости связан с участком своей простой плоскости, и обнаруживает активность нейронов на этом участке, уменьшая таким образом чувствительность к позиции образа.

Таким образом достигается иерархическая обработка изображения, когда на последующих слоях неокогнитрон реагирует на более общие черты изображения, не сбиваясь на искажения, сдвиг и т.д.

Классический неокогнитрон является мощным средством распознавания изображений, однако требует высоких вычислительных затрат, которые на сегодняшний день недостижимы [21].

Однако существует множество работ, например [42,43,47,54], направленных на совершенствование неокогнитрона. Один из наиболее перспективных подходов для распознавания человека по изображению лица – это свёрточные нейронные сети [41].

2.5.3. Свёрточные нейронные сети

В классической многослойной нейронной сети межслойные нейронные соединения полносвязны, и изображение представлено в виде n -мерного вектора, не учитывающего ни двумерную локальную организацию пикселей, ни возможности деформации. Архитектура свёрточной НС [41] (рис. 9) направлена на преодоление этих недостатков и основывается на принципах архитектуры неокогнитрона, упрощённого и дополненного обучением по алгоритму обратного распространения ошибки.

В ней использовались локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса (обеспечивают детектирование некоторых черт в любом месте изображения) и иерархическая организация с пространственными подвыборками (*spatial subsampling*).

Свёрточная НС (СНС, Convolutional Neural Network) обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Архитектура СНС [41] состоит из многих слоёв. Слои бывают двух типов: свёрточные (*Convolutional*) и подвыборочные (*Subsampling*), свёрточные и подвыборочные слои чередуются друг с другом.

В каждом слое имеется набор из нескольких плоскостей, причём нейроны одной плоскости имеют одинаковые веса, ведущие ко всем локальным участкам предыдущего слоя (как в зрительной коре человека), изображение предыдущего слоя как бы сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон текущего слоя. Таким образом набор плоскостей представляет собой карты характеристик (*feature maps*) и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя.

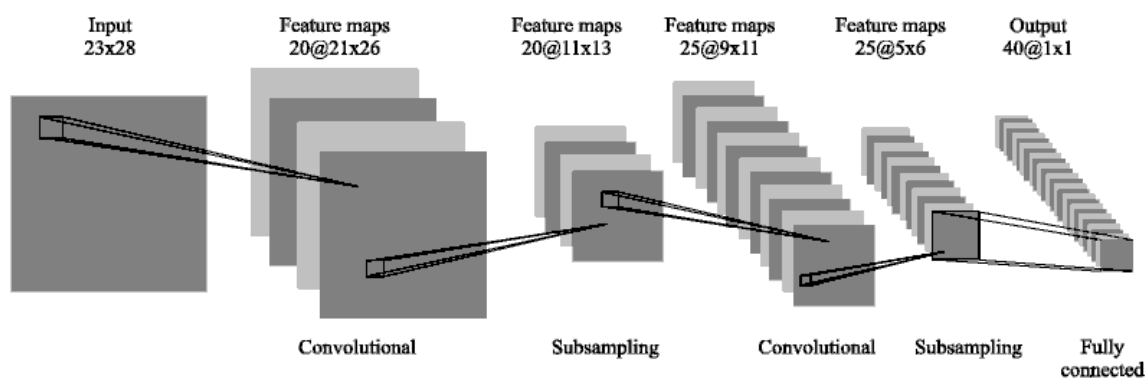


Рис. 9. Архитектура свёрточной нейронной сети

Следующий за свёрточным слоем 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.

Обучается СНС стандартным методом обратного распространения ошибки. Сравнение МНС и СНС [41] показало существенные преимущества последней как по скорости, так и по надёжности классификации. Полезным свойством СНС является и то, что характеристики, формируемые на выходах верхних слоёв иерархии, могут быть применимы для классификации по методу ближайшего соседа (например, вычисляя Евклидово расстояние), причём СНС может успешно извлекать такие характеристики и для образов, отсутствующих в обучающем наборе. Для СНС характерны высокая скорость обучения и работы. Тестирование СНС на базе данных ORL, содержащей изображения лиц с небольшими измене-

ниями освещения, масштаба, пространственных поворотов, положения и различными эмоциями, показало 98%-ную точность распознавания. Такой результат делает эту архитектуру перспективной для дальнейших разработок в области распознавания изображений трёхмерных объектов, в частности человеческих лиц.

3. Перспективы исследований и предварительные эксперименты

Актуальной задачей распознавания человека по изображению лица является контроль (ограничение) доступа. Применение нейронных сетей является перспективным для данной задачи.

Сформулируем требования к такой системе:

- функционирование в реальном времени;
- гибкость настроек, простота и универсальность применения;
- устойчивость к межклассовым вариациям изображения лица (освещение, ракурс);
- устойчивость к внутриклассовым изменениям изображения лица (эмоции, очки, бороды, причёска и т.п.);
- устойчивость к ошибкам первого и второго рода и возможность варьирования компромисса между ними;
- по возможности простота и универсальность применения разрабатываемого алгоритма распознавания изображений.

На основе вышеизложенного для достижения перечисленных целей за основу были выбраны нейросетевые методы.

В связи с этим представляется перспективным исследовать:

- какое исходное представление изображения лучше подавать на вход сети (главные компоненты, частотные и вейвлетные преобразования, моменты и т.п.);
- различные алгоритмы обучения НС (адаптивный шаг, ансамбли нейронных сетей, генетический алгоритм и т.п.);
- применение различных архитектур НС для различных этапов распознавания изображений;
- построение специализированных архитектур НС для распознавания изображений.

Далее описаны некоторые результаты, полученные авторами в ходе предварительных экспериментальных исследований.

3.1. Рециркуляционные нейронные сети и анализ главных компонент для распознавания по изображению лица

Целью этого эксперимента является получение сжатого представления изображения с помощью рециркуляционной нейронной сети и распознавание на основе такого представления. Исследовалась возможность реконструкции изображения на основе сжатого представления. Более подробно с экспериментами и результатами можно ознакомиться в [28].

3.1.1. Отличия от предыдущих работ

Описание архитектуры РНС, её преимуществ и работ, в которых она использовалась, приведено в п. 2.2.1.

В отличие от [57], использующей простую и небольшую базу изображений лиц, нами была использована более сложная база ORL (www.cam-orl.co.uk/facedatabase.html): изображения 40 человек по 10 изображений каждого, всего 400 изображений. Она, в отличие от [57], включает небольшие изменения ракурса, масштаба и освещения. В [57] была использована сигмоидальная активационная функция. В [21] было отмечено, что лучше использовать активационную функцию с выходным диапазоном $[-1; +1]$. В предшествующей работе использовался постоянный шаг обучения. В результате при большом шаге значения выходов скрытых нейронов были близки к двоичному (около 0 или 1), что является недостатком, а при маленьком шаге обучение длилось долго.

В [36] РНС с адаптивным шагом обучения использовалась для сжатия одного изображения, на вход сети подавались блоки изображения. Преимущество использования адаптивного шага заключалось в том, что сеть быстро достигала минимума ошибки реконструкции изображения за 5-20 циклов обучения, в отличие от классического обратного распространения ошибки и кумулятивного дельта-правила. Мы применили этот метод для набора из 400 изображений, на вход сети изображение подавалось целиком.

3.1.2. Алгоритм

Вычисление выходов нейронной сети:

$$y_{ki} = x_i, k = 0; \quad y_{ki} = \tanh\left(\sum_{j=1}^p y_{k-1,j} w_{kij}\right), k = 1..L,$$

где \tanh – функция гиперболического тангенса; k – текущий слой, возрастает от 0 до L ; p – количество нейронов в предыдущем ($k-1$) слое; i – индекс нейрона в текущем слое; j – индекс нейрона в предыдущем слое; x_i – пиксель входного изображения; y_{ki} – значения выходов слоя k (и входные значения следующего слоя); w_{kij} – вес, соединяющий нейроны j_{k-1} и i_k ; L – индекс последнего слоя (здесь $L=2$).

Мы использовали гиперболический тангенс в качестве активационной функции, он имеет выходной диапазон $[-1; +1]$ и производную, которая легко вычисляется. Вследствие этого изображение должно иметь нулевое среднее значение и значения пикселей должны быть отображены в диапазон $[-0.01; +0.01]$, который уменьшается с увеличением разрешения изображения.

Обученная сеть на выходе скрытого слоя выдаёт первые m главных компонент, являющихся сжатым представлением изображения. Поскольку сеть инициализируется случайными значениями, соответствия между номерами компонент и нейронами нет. Для реконструкции изображения на выход нейронов скрытого слоя подаются сжатое представление нужного изображения и рассчитывают значения выходного слоя.

Для обучения сети применяется алгоритм коррекции весов, называемый обратным распространением ошибки. Для последнего слоя вычисляется ошибка (разница между выходными y_{ki} и эталонными t_i значениями) и распространяется обратно по сети сквозь веса скрытых нейронов. Величина коррекции ошибки:

$$\delta_{ki} = (y_{ki} - t_i) \cdot (1 - y_{ki}^2), k = L;$$

$$\delta_{ki} = \left(\sum_{j=1}^q \delta_{k+1,j} w_{k+1,ji} \right) \cdot (1 - y_{ki}^2), k = (L-1)..1,$$

где k уменьшается от L до 1 ; q – число нейронов в слое $k+1$, для РНС эталоном является входное изображение $t_i = y_{0,i}$.

Затем корректируются веса:

$$w_{kij}(t+1) = w_{kij}(t) - \alpha(t) \delta_{ki} y_{k-1,j}, k = 1..L,$$

где $\alpha(t)$ – скорость (шаг) обучения; t – номер обучающего цикла. Для классического обратного распространения скорость фиксирована. Существуют эвристические подходы, в которых скорость изменяется от большой вначале до маленькой в конце обучения.

Главное преимущество подхода Головки [36] – это **адаптивный шаг**, который рассчитывается индивидуально для каждого слоя на каждой итерации для того, чтобы сделать лучший шаг в направлении минимизации среднеквадратичной ошибки сети:

$$\alpha(t) = \frac{\sum_{i=1}^r \frac{\delta_{ki}^2}{1 - y_{ki}^2}}{\left(1 + \sum_{j=1}^p y_{k-1,j}^2\right) \cdot \left(\sum_{i=1}^r \delta_{ki}^2\right)},$$

где r – число нейронов в слое k . Следует принимать $\frac{\delta_{ki}^2}{1 - y_{ki}^2} = 0$ при

$1 - y_{ki}^2 = 0$ и $\alpha(t) = 0$ при $\sum_{i=1}^r \delta_{ki}^2 = 0$. В исходной формулировке Головки деление на ноль отсутствует.

Адаптивный шаг избавляет от необходимости выбирать шаг вручную. Обучающий процесс сходится сравнительно быстро и стабильно. Для простоты мы не использовали раздельное обучение [36].

Перед обучением веса сети инициализируются небольшими случайными значениями $[-0.01; +0.01]$.

Обучающий процесс состоит из последовательности обучающих циклов и завершается, когда их число превышает допустимое значение или ошибка нейронной сети становится меньше заданной.

На каждом обучающем цикле на сеть подаются изображения из обучающего набора в случайном порядке. После этого вычисляется ошибка сети и корректируются веса. В процессе обучения сеть учится сжимать и реконструировать изображение через небольшой набор нейронов скрытого слоя.

Весь вышеприведённый алгоритм может быть использован и для обучения обычного многослойного персептрона для многих других задач.

Для извлечения главных компонент использовалась рециркуляционная нейронная сеть (РНС). Затем с помощью евклидовой метрики вычислялось расстояние от неизвестного изображения ко всем изображениям в обучающей выборке. В качестве координат использовались главные компоненты (выходы скрытых нейронов РНС). Изображение с наименьшим расстоянием считалось наиболее похожим.

В [37] было показано, что при смене ракурса главные компоненты описывают кривые, называемые собственными сигнатурами, которые уникальны для каждого лица. В случае других вариаций главные компоненты образуют собственные гиперповерхности. Поэтому мы не использовали кластеризационные методы, а вычисляли расстояние до всех изображений каждого человека.

3.1.3. Результаты

База ORL содержит 400 изображений: 40 человек по 10 изображений каждого. Для каждого эксперимента база делилась случайным образом на две части, обучающую и тестовую, по пять изображений одного человека в каждой части.

Для обучения сети и сравнения с неизвестным изображением использовалась одна и та же обучающая выборка.

Исследовались возможности распознавания на основе полученных главных компонент и возможности реконструкции изображения в зависимости от следующих факторов:

- количества обучающих циклов;
- числа скрытых нейронов;
- разрешения изображения: ORL/1 (92x112 пикселей, исходный размер), ORL/2 (46x56), ORL/4 (23x28);
- различной случайной разбивки на тестовую и тренировочную части.

Полученные результаты можно охарактеризовать следующим образом. Ошибка реконструкции быстро уменьшается в течение первых 10-20 шагов и дальше практически не изменяется (рис. 10). Ошибка распознавания имеет похожую тенденцию. Обе эти величины слегка колеблются вследствие случайного порядка обучающих образов.

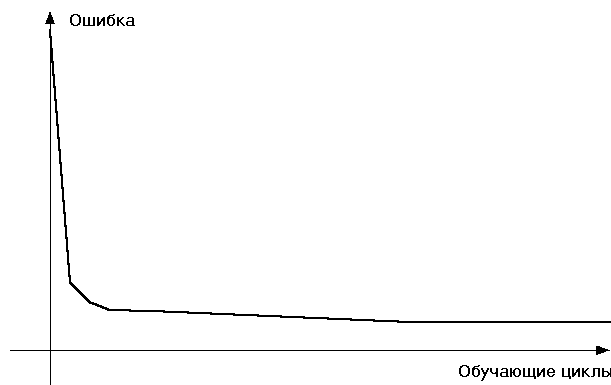


Рис. 10. Зависимость ошибки реконструкции от числа обучающих циклов

Средняя точность распознавания составляет 92% и не зависит от выбранного разрешения изображения. Но с увеличением разрешения время обучения увеличивается пропорционально числу пикселей.

На рис. 11 показано изменение главных компонент и реконструкции по ним с увеличением числа тренировочных циклов, на рис. 12 показаны входные и выходные веса обученной сети, похожие на собственные лица, на рис. 13 показаны примеры реконструкции тестовых и тренировочных изображений.

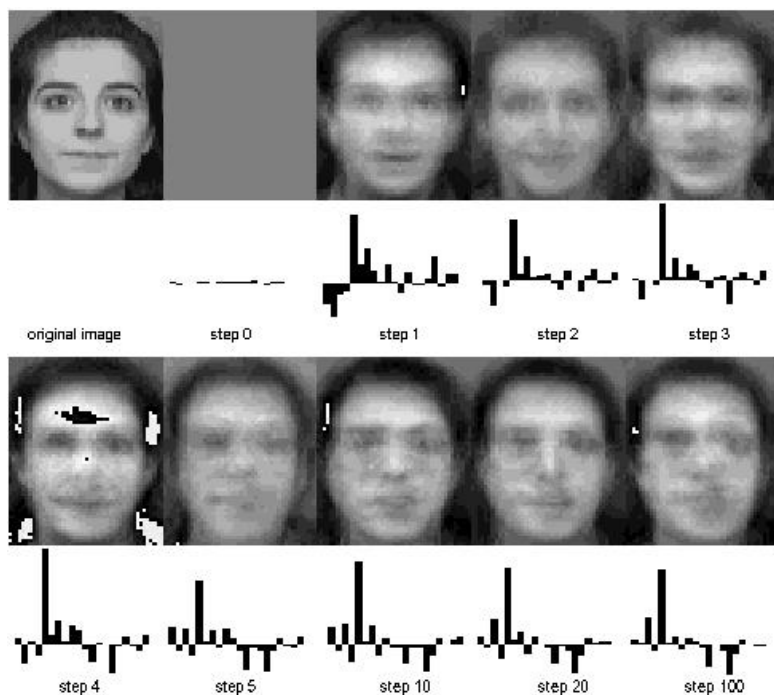


Рис. 11. Изменение первых компонент (диаграммы под изображениями) и реконструкции изображения с увеличением обучающих циклов

При изменении количества скрытых нейронов время обучения линейно возрастает, ошибка реконструкции уменьшается, ошибка распознавания медленно уменьшается.

Точность распознавания зависит ещё и от того, какие изображения попадут в обучающую выборку при очередном случайном разделении. Если в обучающей выборке нет изображения лица при аналогичных условиях (ракурса, например), то система имеет тенденцию ошибаться (рис. 14).

Таким образом, рециркуляционные нейронные сети представляют собой перспективный механизм извлечения главных компонент и реконструкции по ним. Время обучения зависит от количества компонент, а соотношение время/качество можно варьировать, изменяя число обучающих циклов.

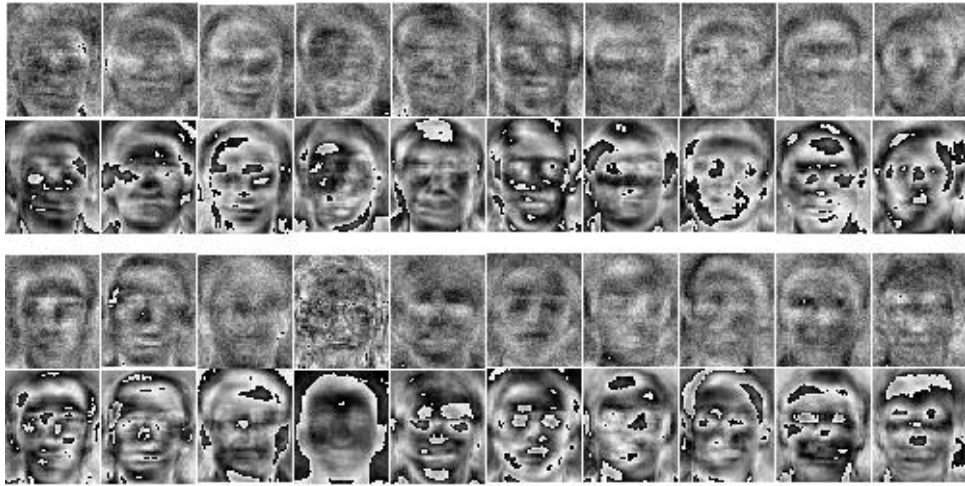


Рис. 12. Входные (первая и третья строки) и выходные (вторая и четвёртая строки) веса РНС (представленные в виде изображений) похожи на собственные лица

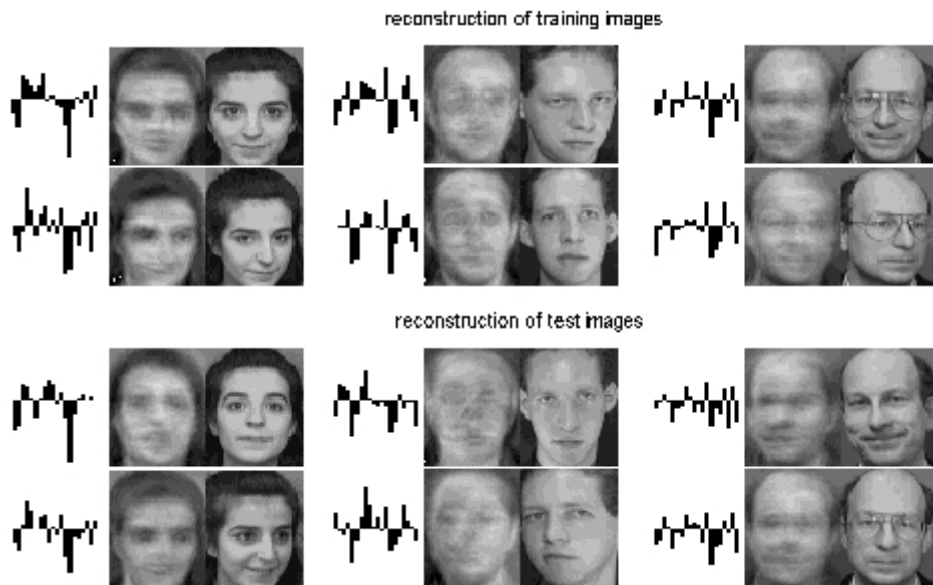


Рис. 13. Компоненты и реконструкция тренировочных и тестовых изображений. Слева направо: диаграмма первых компонент, реконструкция, исходное изображение. Первые две строки – изображения из обучающей выборки, вторые две строки – тестовые изображения

Но для самостоятельного применения в системе распознавания по изображениям лиц исследованного алгоритма недостаточно. Его следует комбинировать с более надёжными методами классификации и использовать обучающий набор, содержащий вариации изображений, которые будут встречаться в процессе функционирования системы.













Правильное распознавание			Неправильное распознавание		
Изображение	Класс	Расст-е	Изображение	Класс	Расст-е
	s1/2	тестовое изобр.		s1/6	тестовое изобр.
	s1/5	0.3182		s4/6	0.3161
	s13/6	0.3551		s1/1	0.3357
	s18/5	0.3944		s12/1	0.3361
	s13/5	0.3974		s13/7	0.3425
	s5/3	0.3984		s5/5	0.3432

Рис. 14. Пример распознавания. Слева правильное распознавание, справа – неправильное. В верхнем ряду два неизвестных изображения, ниже – ближайшие к нему из обучающей выборки

3.2. Многослойный перцептрон и различное начальное представление изображения

В этой серии экспериментов исследовались параметры архитектуры многослойного перцептрона (число слоёв и нейронов), различная начальная инициализация, алгоритмы обучения, начальное представление изображения. Производился анализ ошибок распознавания.

3.2.1. Архитектура и обучение

Как и в предыдущем эксперименте, число входов сети равнялось количеству пикселей входного изображения. Число выходов всегда соответствовало числу классов (человек) в базе ORL – 40. Эталонные выходы сети имели значение +1 для своего класса и –1 для всех остальных. Таким образом, если на вход сети подавалось изображение человека,

принадлежащего третьему классу, то третий нейрон в последнем слое учился выдавать «+1», а все остальные – «-1».

После ряда экспериментов с различными способами инициализации оптимальными были признаны начальные случайные веса в диапазоне $[-0.01; +0.01]$. Для изображения с разрешением 23×28 значения яркостей пикселей масштабировались в аналогичный диапазон. Чтобы избежать паралича, диапазон значений яркостей уменьшался с увеличением разрешения изображения.

В качестве оптимальной архитектуры была экспериментально подобрана двухслойная нейронная сеть приблизительно с 25 нейронами в скрытом слое.

Из способов обучения рассматривались фиксированный и адаптивный шаг. Для каждого нового начального представления изображения или других параметров архитектуры приходилось тщательно подбирать вручную оптимальное значение фиксированного шага, что, однако, не гарантировало сходимости. Адаптивный шаг стабильно работал при любых параметрах архитектуры и начальном представлении, всегда достигая 100%-ной точности на обучающей выборке приблизительно за 100 тренировочных циклов. При этом все выходы последнего слоя были близки к идеальным (рис. 15).

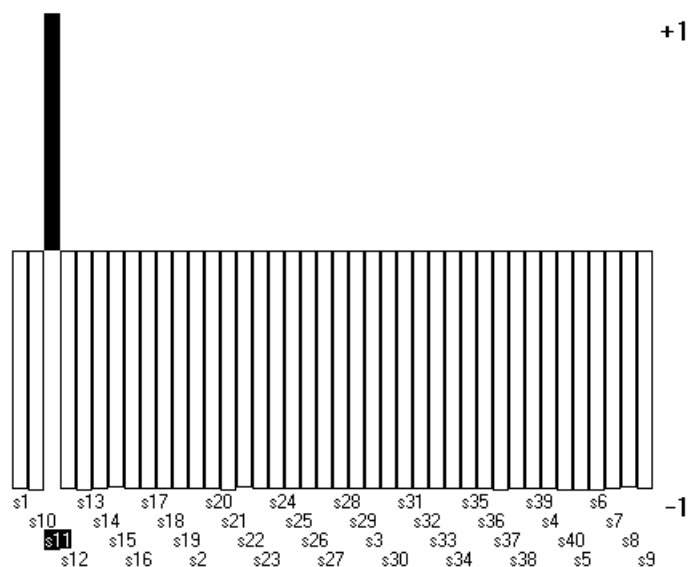


Рис. 15. Выходы сети после обучения с адаптивным шагом. Подано изображение человека из класса *s11*; выход сети, соответствующий классу *s11*, максимален

Распознавание на тестовой выборке показало точность при различном делении обучающей выборки и других изменениях параметров от

90 до 98%, в среднем 94%. Главным образом это зависит от качества обучающей выборки, т.е. от того, каким образом выборка будет случайно разделена на обучающие и тестовые изображения.

Изменения точности в $\pm 1\%$ также можно было получить, выполнив несколько дополнительных обучающих циклов. Так как выходы сети после обучения всегда очень близки к идеальным, это можно связать со случайными отклонениями, которые, вероятно, не могут быть исправлены алгоритмом обучения.

Для большинства тестовых изображений выходы сети имели вид, как на рис. 16, что говорит о высоком качестве обучения и хороших различающих способностях многослойных персептронов.

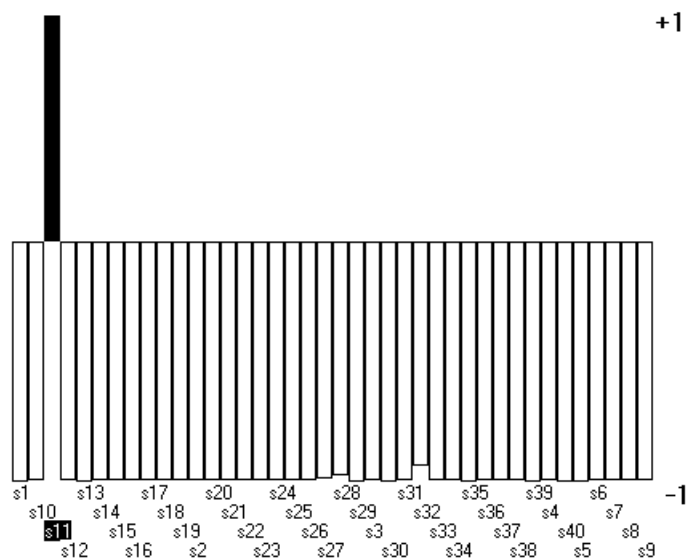


Рис. 16. Пример типичного успешного распознавания тестового изображения. На вход сети подано изображение человека из класса *s11*; выход сети, соответствующий классу *s11*, максимален

Веса нейронов входного слоя, представленные в виде изображений, имеют чётко различимую форму, похожую на изображение лица (рис. 17). Это наглядно демонстрирует принцип работы многослойной НС – разбиение пространства изображений на области-классы. Каждое входное изображение внутри сети представляется в виде нелинейной комбинации таких «весов-изображений». Выходы нейронов первого слоя являются сжатым представлением каждого изображения, а сам первый слой осуществляет извлечение признаков – разложение исходного изображения на «веса-изображения». Таким образом, становятся наглядно понятны ограничения многослойных нейронных сетей (и всех остальных методов, разделяющих исходное пространство на области)

при распознавании изображений. На каждую вариацию изображения (ориентация, масштаб, освещение, ракурс, очки, эмоции) должен существовать «вес-изображение», учитывающий эту вариацию. И чем больше различных классов и их вариаций, тем больше должно быть нейронов в скрытом слое и в обучающей выборке требуется большее число примеров, учитывающих различные вариации. На практике это трудновыполнимо. Выход заключается в применении МНС к участкам изображений и анализе их взаимного расположения. Лучше всего для этого подходят архитектуры свёрточных сетей и неокогнитрона.

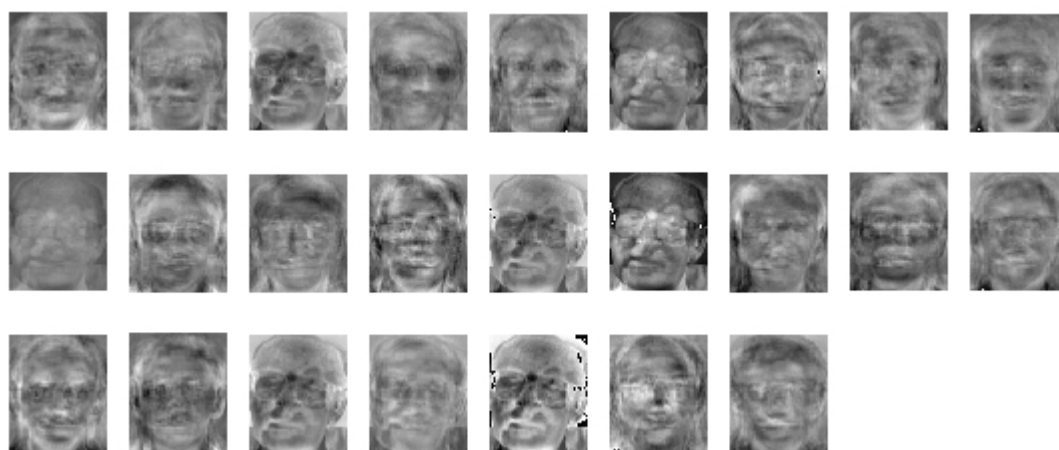


Рис. 17. Входные веса НС, представленные в виде изображений, 25 нейронов в скрытом слое

Для того чтобы уменьшить число ошибок второго рода (когда система ошибочно распознаёт один класс как другой), можно применять так называемый порог отклонения. Таким образом, если максимальный выход сети не превышает этот порог, то принимается, что сеть вообще не узнала этот объект. Таким образом уменьшается число ошибок второго рода, но за счёт небольшого повышения общего процента ошибок. Этот порог можно варьировать для различных задач, например для разных уровней доступа в системах контроля доступа.

Типичная ошибка второго рода выглядит, как показано на рис. 18. В этом случае существует несколько выходов сети, значения которых далеки от «-1», а значения максимального выхода далеки от «+1». Это позволяет применять различного рода эвристики, учитывающие несогласованность выходов сети. Ошибки в основном связаны с тем же фактором, что и в предыдущем эксперименте – в обучающей выборке отсутствует изображение лица при определённых условиях (в данном случае ракурс). Один из способов частичного преодоления этого недостатка –

использование в тренировочной выборке зеркальных отражений изображений, что даёт больший диапазон ракурсов [31].

+1

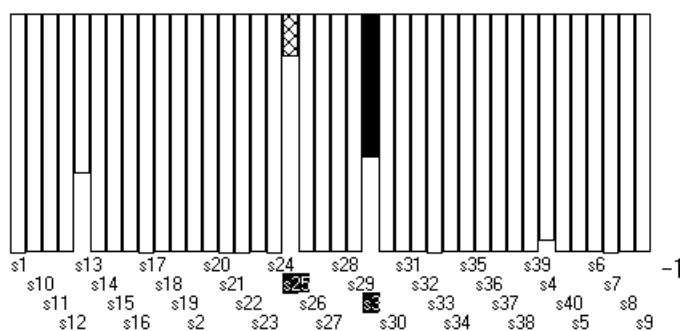


Рис. 18. Типичный случай ошибки первого рода (вместо класса s3 выдаёт s25)

Проводился и более интересный эксперимент. База разбивалась на три набора, причём из третьего набора классов примеров в обучающей выборке не было. Анализировалась способность НС классифицировать и при этом отличать «свои» изображения от «чужих» с заданным порогом отклонения. Если учитывать только процент распознавания, то он был неприемлемым. Но в случае ошибок выходы сети выглядели ещё более «разношёрстно», чем в предыдущем случае. Такой факт позволяет создать эвристики для более надёжного распознавания. Кроме того, сеть не обучалась именно отличать «своих» и «чужих», производилась только настройка на классификацию среди «своих». Как известно, для обучения требуются ещё и негативные примеры, которых в этом случае в связи с малым размером базы не было.

Также производился эксперимент с нашей собственной тестовой базой. В ней имеются более широкие условия освещённости, и вследствие этого получен более низкий процент распознавания – около 80%. Система в этом случае начинала реагировать на освещение лиц.

3.2.2. Начальное представление изображения

При формировании исходного представления изображения использовались:

- различный масштаб изображения;
- различное число главных компонент, извлекаемых РНС;
- все коэффициенты преобразования Фурье;
- часть коэффициентов косинусного преобразования;
- часть коэффициентов блочного косинусного преобразования.

Различия в начальном представлении влияли на точность распознавания следующим образом.

Уменьшение размера изображения путём масштабирования на качество распознавания не влияло, только на скорость обучения и работы.

При использовании главных компонент точность распознавания по первым главным компонентам оказалась хуже, чем по самому изображению – около 90%. С увеличением числа компонент точность распознавания повышалась.

Использование всех коэффициентов преобразования Фурье не дало никаких преимуществ.

Использование части коэффициентов косинусного преобразования позволило существенно повысить скорость обучения, но повышения точности не принесло. Причём с уменьшением числа коэффициентов до 20-30 из 10000 распознавание остаётся на прежнем уровне, хотя реконструкция уже невозможна (см. предыдущие разделы). При этом число тренировочных циклов приходится увеличивать, но в целом время обучения уменьшается.

Похожая ситуация с блочным косинусным преобразованием.

Таким образом, ни одно из вышеперечисленных преобразований само по себе не даёт повышения точности распознавания и не позволяет компенсировать изменения ракурса и освещения.

3.3. Выводы

Требуются дальнейшие эксперименты по исследованию алгоритмов обучения, выбору начального представления и внедрению в архитектуру НС учёта свойств изображения.

4. Ресурсы в сети Интернет

Во всемирной сети Интернет имеется большой объем литературы, посвященной нейросетевой тематике и обработке и распознаванию изображений. Большая часть англоязычных статей, на которые приведены ссылки, имеется в сети Интернет. При затруднении поиска какой-либо статьи или книги, на которую имеется ссылка, свяжитесь с авторами по адресу: bdv78@mail.ru.

4.1. Русскоязычные ресурсы

<http://www.neuropower.de/rus> – ресурс, посвященный искусственному интеллекту и нейросетевой тематике, доступен большой объем литературы в архиве по нейронным сетям, генетическим алгоритмам, искусственному интеллекту.

<http://forum.basegroup.ru/> – открытый форум по нейросетевой тематике и смежным темам.

<http://www.osp.ru/os/2000/03/> - выпуск журнала «Открытые системы», посвященный биометрии и распознаванию человека по изображению лица.

<http://canopus.lpi.msk.su/neurolab/papers/nbusapp/> – курс лекций А.А. Ежова, С.А. Шумского «Нейрокомпьютинг и его применение в экономике и бизнесе».

<http://neurnews.iu4.bmstu.ru> – статьи и книги по нейросетевой тематике.

<http://www.chat.ru/~saisa> – материалы по генетическим алгоритмам.

<http://alife.narod.ru/> – лекции по вейвлетам и нейронным сетям.

<http://nncourse.chat.ru> – учебное пособие «Нейронные сети (основные модели)».

http://www.keldysh.ru/gc98/cd/tutorial/leo_lev/ – два курса лекций по вейвлет-анализу.

<http://www.math.spbu.ru/htmlsources/user/dmp/rusrap.html> – подборка русскоязычных статей и книг (в архивах) по вейвлет-анализу.

<http://www.statsoft.ru/home/textbook/> – электронный учебник по статистике.

<http://neuroface.narod.ru> – электронная версия данного препринта (расширенная и дополненная) и других статей автора, подборка ссылок аналогичной тематики (будут доступны исходные тексты экспериментальных программ).

<http://ya.ru> – одна из лучших поисковых машин в русскоязычном Интернете.

4.2. Англоязычные ресурсы

Большинство материалов хранится в формате .ps (PostScript), для их просмотра требуется программа GSView, адрес ресурса – <http://www.cs.wisc.edu/~ghost>, или требуется скопировать следующие файлы:

<ftp://ftp.cs.wisc.edu/ghost/aladdin/g650/g650w32.exe>,

<ftp://ftp.cs.wisc.edu/ghost/ghostgum/gsv34w32.exe>.

<http://www.cs.rug.nl/users/peterkr/FACE/face.html> – ресурс, посвящённый проблеме распознавания человека по изображению лица. Дано множество ссылок на ресурсы аналогичной тематики, на нейросетевые ресурсы, ссылки на ресурсы научных коллективов, занимающихся смежными проблемами, ссылки на тестовые базы данных.

<http://citeseer.nj.nec.com/cs> – архив статей по различной тематике, в том числе по распознаванию лиц и нейронным сетям. Поиск по ключевым словам. Статьи в форматах PostScript и PDF в сжатом виде.

<http://www.ks.informatik.uni-kiel.de/~vok/research/research.html> – ресурс доктора Фолькера Крюгера. Посвящён обработке изображений и проблеме распознавания человека по изображению лица.

<http://www.funet.fi/pub/sci/neural/neuroprose> – архив статей по нейронным сетям, доступны аннотации.

<http://www.neci.nec.com/~lawrence> – ресурс доктора Стива Лоренса. Имеется публикация по свёрточным нейронным сетям.

<http://www.cnl.salk.edu/~wiskott/homepage.html> – ресурс о применении нейронных сетей к проблеме распознавания человека по изображению лица.

<http://www.public.iastate.edu/~rpolikar/WAVELETS/WTtutorial.html> – простое и доступное изложение идей вейвлет-анализа.

<http://FaceIt.com/tech> – Visionics FaceIt technology FAQ. Описание коммерческой технологии FaceIt® распознавания человека по изображению лица. Представляет собой набор COM и ActiveX – объектов, на основе которых легко строить специализированные системы для распознавания и других связанных задач.

<http://google.com> – одна из лучших поисковых машин в Интернете.

Заключение

В работе рассмотрены основные классы задач распознавания человека по изображению лица. Впервые на русском языке дан обзор нейро-

сетевых методов распознавания человека по изображению лица. Указаны преимущества и перспективы нейросетевых методов. Отмечены архитектуры нейронных сетей, перспективных для данной задачи. Особое внимание уделено выбору начального представления изображения и учёту его свойств.

Отмечены перспективные направления исследований в этой области для достижения намеченной цели. Приведены описание и результаты предварительных экспериментов по созданию системы контроля доступа на основе анализа изображения лица человека.

Дан обзор адресов в сети Интернет по данной тематике.

В настоящее время нейросетевой подход к задаче контроля доступа на основе анализа изображения лица нейросетевыми методами остаётся недостаточно разработанным, что требует дальнейших исследований в этой области.

Литература

1. Абламейко С.В., Лагуновский Д.М. Обработка изображений: технология, методы, применение. – Минск: Ин-т техн. кибернетики НАН Беларуси, 2000. – 304 с.
2. Браунли К.А. Статистическая теория и методология в науке и технике. – М: Наука, 1977. – 408 с.
3. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Харьков: Основа, 1997. (<http://www.neuropower.de/rus>).
4. Галушкин А. И., Томашевич Д. С., Томашевич Н. С. Методы реализации инвариантности к аффинным преобразованиям двумерных изображений // Приложение к журналу «Информационные технологии». – 2001. – №1. – С. 1-19.
5. Глазунов А. Компьютерное распознавание человеческих лиц // Открытые системы. – 2000. – №3. (<http://www.osp.ru/os/2000/03/>).
6. Головкин В.А. Нейроинтеллект: Теория и применения. Книга 1. Организация и обучение нейронных сетей с прямыми и обратными связями. – Брест: БПИ, 1999. – 260с.
7. Головкин В.А. Нейроинтеллект: Теория и применения. Книга 2. Самоорганизация, отказоустойчивость и применение нейронных сетей. – Брест: БПИ, 1999. – 228с.
8. Горелик А.Л., Скрипкин В.А. Методы распознавания. – М: Высшая школа, 1984. – 208 с.

9. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. – М: МИФИ, 1998. (<http://canopus.lpi.msk.su/neurolab/papers/nbusapp/>).
10. Нейроинформатика / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н. Кирдин и др. – Новосибирск: Наука. Сибирское предприятие РАН, 1998. (<http://www.neuropower.de/rus>).
11. Панканти Ш, Болле Р.М., Джейн Э. Биометрия: будущее идентификации // Открытые системы. – 2000. – №3. (<http://www.osp.ru/os/2000/03/>).
12. Пентланд А.С., Чаудхари Т. Распознавание лиц для интеллектуальных сред // Открытые системы. – 2000. – №3. (<http://www.osp.ru/os/2000/03/>).
13. Садыхов Р.Х., Ваткин М.Е. Модифицированный алгоритм обучения РБФ-сети для распознавания рукописных символов // Идентификация образов. – Минск: Ин-т техн. кибернетики НАН Беларуси, 2001. – С.7-16.
14. Самаль Д.И., Старовойтов В.В. Выбор признаков для распознавания на основе статистических данных // Цифровая обработка изображений. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1999. – С.105-114.
15. Самаль Д.И., Старовойтов В.В. Методика автоматизированного распознавания людей по фотопортретам // Цифровая обработка изображений. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1999. – С.81-85.
16. Самаль Д.И., Старовойтов В.В. Подходы и методы распознавания людей по фотопортретам. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1998. – 54с.
17. Самаль Д.И. Построение систем идентификации личности на основе антропометрических точек лица // Цифровая обработка изображений. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1998. – С.72-79.
18. Скурихин А.Н. Генетические алгоритмы // Новости искусственного интеллекта. – 1995. – №4. – С. 6-46.
19. Старовойтов В.В. Локальные геометрические методы цифровой обработки и анализа изображений. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1997. – 284 с.
20. Старовойтов В.В., Талёб М.А. – Методы сегментации цветных изображений. – Минск: Ин-т техн. кибернетики НАН Беларуси, 1999. – 44с.

21. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. – М: МИФИ, 1992. – 184 с. (<http://www.neuropower.de/rus>).
22. Филлипс П. Дж., Мартин Э., Уилсон С. Л., Пржибоски М. Введение в оценку биометрических систем // Открытые системы. – 2000. – №3. (<http://www.osp.ru/os/2000/03/>).
23. Хорн Б.К.П. Зрение роботов. – М: Мир, 1989. – 488 с.
24. Электронный учебник по статистике. – М: StatSoft, 2000. (<http://www.statsoft.ru/home/textbook/>).
25. Adini Y., Moses Y., Ullman S. Face Recognition: The Problem of Compensating for Changes in Illumination Direction // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – P. 721-732.
26. Aizenberg I. N., Aizenberg N. N., Krivosheev G.A. Multi-valued and Universal Binary Neurons: Learning Algorithms, Applications to Image Processing and Recognition // Lecture Notes in Artificial Intelligence – Machine Learning and Data Mining in Pattern Recognition. – 1999. – P. 21-35.
27. Brunelli R., Poggio T. Caricatural Effects in Automated Face Perception. Massachusetts Institute of Technology, Cambridge. – 6 p.
28. Bryliuk D., Starovoitov V. Application of Recirculation Neural Network and Principal Component Analysis for Face Recognition // The 2nd International Conference on Neural Networks and Artificial Intelligence. – Minsk: BSUIR, 2001. – P.136-142.
29. Dai Y., Nakano Y. Recognition of facial images with low resolution using a Hopfield memory model // Pattern Recognition. – 1998. – Vol. 31. – P. 159-167.
30. Daughman J. Face and Gesture Recognition: Overview // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – P. 675-676.
31. Eickeler S., Muller S., Rigoll G. High performance face recognition using Pseudo 2-D Hidden Markov Models. Gerhard-Mercator-University Duisburg, Germany, 1998. – 6 p.
32. Foltyniewicz R. Efficient High Order Neural Network for Rotation, Translation and Distance Invariant Recognition of Gray Scale Images // Lecture Notes in Computer Science – Computer Analysis of Images and Patterns. – 1995. – P. 424-431.
33. Foresti G.L. Outdoor Scene Classification by a Neural Tree-Based Approach // Pattern Analysis and Applications. – 1999. – Vol. 2. – №2. – P. 129-142.
34. Giacinto G., Roli F. Automatic Design of Multiple Classifier Systems by Unsupervised Learning // Lecture Notes in Artificial Intelligence –

Machine Learning and Data Mining in Pattern Recognition. – 1999. – P. 131-143.

35. Golfarelli M, Maio D., Maltoni D. On the Error-Reject Trade-Off in Biometric Verification Systems // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – P. 786-796.

36. Golovko V., Gladyschuk V. Recirculation Neural Network Training for Image Processing // Advanced Computer Systems. – 1999. – P. 73-78.

37. Graham D.B., Allinson N.M. Face recognition using virtual parametric eigenspace signatures // Image Processing and its Applications. – 1997. – P. 106-110.

38. Gutta S., Wechsler H. Face recognition using hybrid classifiers // Pattern Recognition. – 1997. – Vol. 30. – P. 539-553.

39. Intrator N., Reisfeld D., Yeshurum Y. Face Recognition using a Hybrid Supervised/Unsupervised Neural Network // International Conference on Pattern Recognition. – 1994. – Vol. 2. – P. 50-54.

40. Jacobsen X., Zscherpel U., Perner P. A Comparison between Neural Networks and Decision Trees // Lecture Notes in Artificial Intelligence – Machine Learning and Data Mining in Pattern Recognition. – 1999. – P. 144-158.

41. Lawrence S., Giles C. L., Tsoi A. C., Back A. D. Face Recognition: A Convolutional Neural Network Approach // IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition. – 1997. – P. 1-24. (<http://www.neci.nec.com/~lawrence>).

42. Lovell D. R., Downs T., Tsoi A. C. An Evaluation of The Neocognitron. University of Queensland, Australia: Technical Report. – 16 p.

43. Lovell D. R., Tsoi A. C. The Performance of the Neocognitron with Various S-Cell and C-Cell Transfer Functions. University of Queensland, Australia: Technical Report, 1992. – 10 p.

44. Martinez A., Vitria J. Dimensionality reduction for face recognition // Advances in Visual Form Analysis. – 1997. – P. 405-414.

45. Milanova M., Almeida P. E. M., Okamoto J., Simoes M. G. Applications of Cellular Neural Networks for Shape from Shading Problem // Lecture Notes in Artificial Intelligence – Machine Learning and Data Mining in Pattern Recognition. – 1999. – P. 51-63.

46. Moghaddam B., Pentland A. Probabilistic Visual Learning for Object Representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – P. 696-710.

47. Ooyent A., Nienhuis B. Pattern Recognition in the Neocognitron is Improved by Neuronal Adaptation // Amsterdam: Biological Cybernetics. – 1993. – Vol. 70. – P. 47-53.

48. Pan Z., Rust A. G., Bolouri H. Image Redundancy Reduction for Neural Network Classification using Discrete Cosine Transforms // Proceedings of the IJCNN. – 2000. – Vol. 3. – P. 149-154.
49. Petrou M. Learning in Pattern Recognition // Lecture Notes in Artificial Intelligence – Machine Learning and Data Mining in Pattern Recognition. – 1999. – P. 1-12.
50. Plakhov A. Yu. The Converging Unlearning Algorithm for the Hopfield Neural Network: Optimal Strategy // International Conference on Pattern Recognition. – 1994. – Vol. 2. – P. 104-106.
51. Ranganath S., Arun K. Face recognition using transform features and neural networks // Pattern Recognition. – 1997. – Vol. 30. – P. 1615-1622.
52. Rosenblum M., Yacoob Y., Davis L. Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture // IEEE Workshop on Motion of Non-Rigid and Articulated Objects. – 1994.
53. Rowley H. A., Baluja S., Kanade T. Neural Network-Based Face Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1998. – Vol. 20. – P. 23-37.
54. Sadykhov R., Vatkin M. New training algorithm of neural network “Neokognitron” for a recognition of hand-written symbols // Pattern Recognition and Image Processing. – 2001. – P. 203-212.
55. Takacs B., Wechsler H. Locating Facial Features Using SOFM // International Conference on Pattern Recognition. – 1994. – Vol. 2. – P. 55-60.
56. Valentin D., Abdi H. Can a Linear Autoassociator Recognize Faces From New Orientations? The University of Texas at Dallas, 1995. – 11 p.
57. Valentin D., Abdi H., O'Toole A. J., Cottrell G. W. Connectionist models of face processing: a survey // IN: Pattern Recognition. – 1994. – Vol. 27. – P. 1209-1230.
58. Vetter T., Poggio T. Linear Object Classes and Image Synthesis From a Single Example Image // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – P. 733-742.
59. Yoon K. S., Ham Y. K., Park R.-H. Hybrid approaches to frontal view face recognition using the Hidden Markov Model and Neural Network // Pattern Recognition. – 1998. – Vol. 31. – P. 283-293.

Оглавление

Введение.....	3
1. Основные классы решаемых задач в распознавании человека по изображению лица	4
1.1. Поиск изображения в больших базах данных.....	5

1.2. Задача контроля доступа	6
1.3. Задача контроля фотографии в документах	6
2. Нейросетевые методы распознавания человека по изображению лица	7
2.1. Архитектура нейронных сетей	9
2.2. Разделение пространства признаков на области и извлечение ключевых признаков	11
2.2.1. Многослойные нейронные сети	11
2.2.2. Нейронные сети высокого порядка и моментные НС.....	19
2.2.3. Радиально-базисные нейронные сети	20
2.3. Ассоциативная память и решение оптимизационных задач	23
2.4. Топологически упорядоченное преобразование пространства	24
2.5. Распознавание с учётом топологии пространства	27
2.5.1. Когнитрон	27
2.5.2. Неокогнитрон	28
2.5.3. Свёрточные нейронные сети.....	30
3. Перспективы исследований и предварительные эксперименты	32
3.1. Рециркуляционные нейронные сети и анализ главных компонент для распознавания по изображению лица.....	33
3.1.1. Отличия от предыдущих работ.....	33
3.1.2. Алгоритм.....	33
3.1.3. Результаты	36
3.2. Многослойный персептрон и различное начальное представление изображения	39
3.2.1. Архитектура и обучение	39
3.2.2. Начальное представление изображения	43
3.3. Выводы	44
4. Ресурсы в сети Интернет	44
4.1. Русскоязычные ресурсы	45
4.2. Англоязычные ресурсы	46
Заключение	46
Литература	47