Разработка нейросетей, выбор их параметров, их оптимизация для распознавания видеоизображений людей

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.).
- исследовать различные алгоритмы обучения НС (адаптивный шаг, ансамбли нейронных сетей, генетический алгоритм и т.п.).
- исследовать применение различных архитектур НС для различных этапов распознавания изображений.
- построить и исследовать специализированные архитектуры НС для распознавания изображений.

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

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].

Отличия от предыдущих работ

Описание архитектуры РНС, её преимуществ и работ, в которых она использовалось, 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.

Алгоритм

Вычисление выходов нейронной сети:

$$y_{ki} = x_i, k = 0;$$

$$y_{ki} = \tanh(\sum_{j=1}^{p} y_{k-1,j} w_{kij}), k = 1..L,$$

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 нейрон j_{k-1} и нейрон 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, рис. !!!. Поскольку сеть инициализируется случайными значениями, соответствия между номерами компонент и нейронами нет. Для реконструкции изображения на выход нейронов скрытого слоя подают сжатое представление нужного изображения и рассчитывают значения выходного слоя.

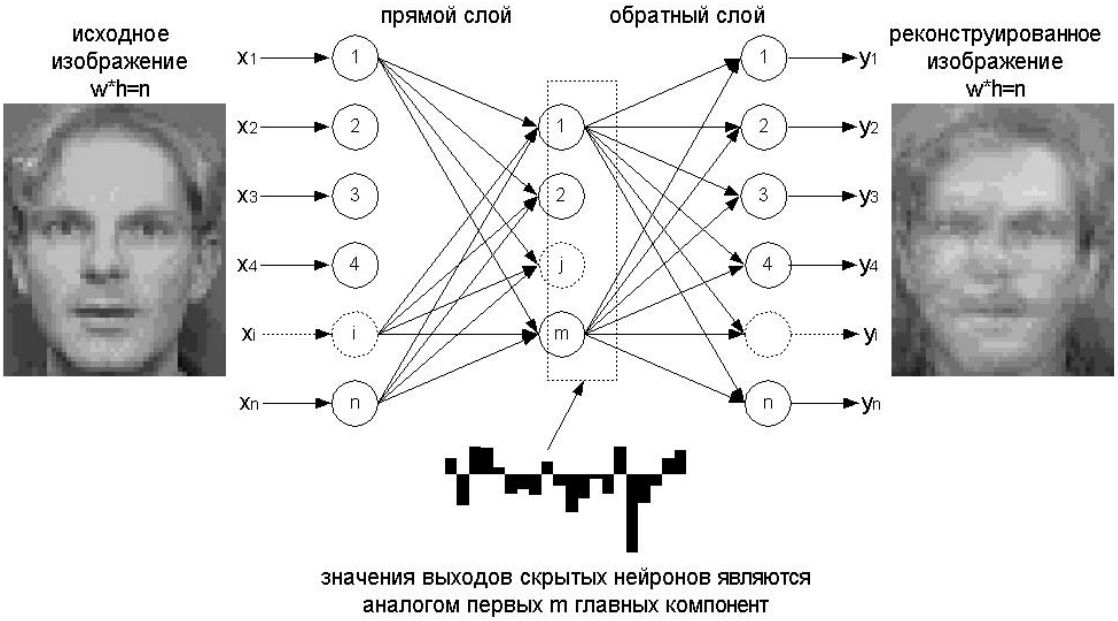

Рис. !!!. Архитектура рециркуляционной нейронной сети

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 δ_{ki}:

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

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

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 Головко [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 Головко 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.

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.

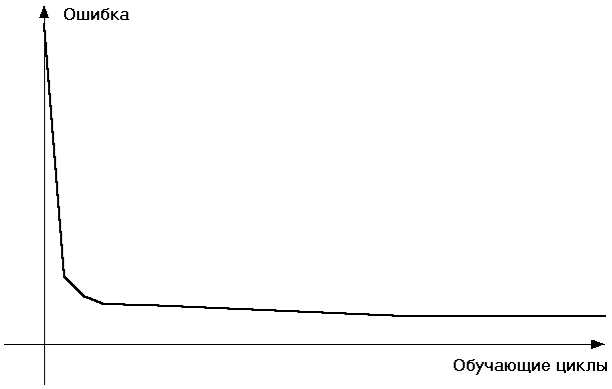

Рис. 10. Зависимость ошибки реконструкции
от числа обучающих циклов

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.

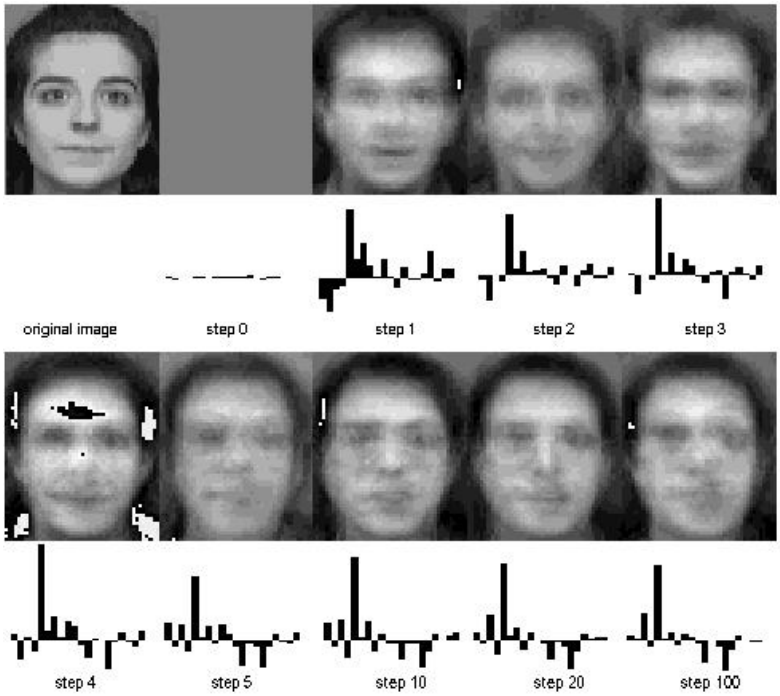

Рис. 11. Изменение первых компонент (диаграммы под изображениями) и реконструкции изображения с увеличением обучающих циклов

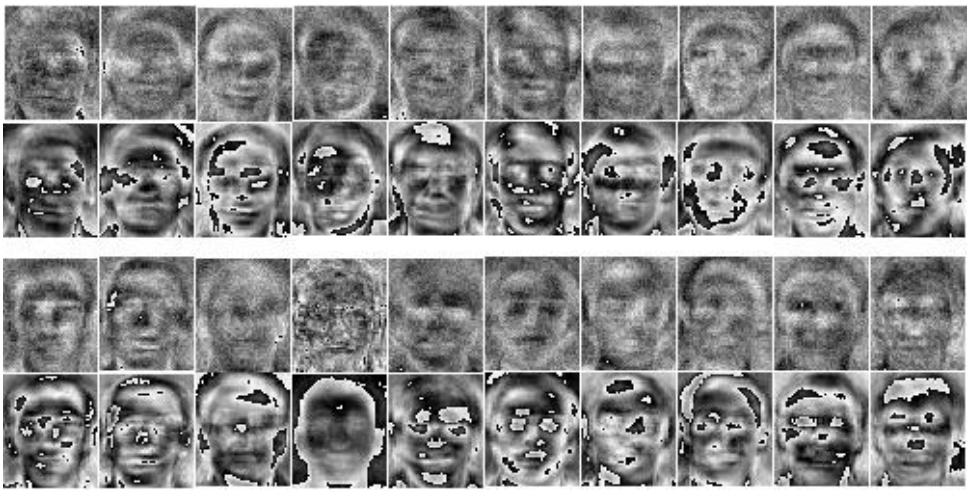

Рис. 12. Входные (первая и третья строки) и выходные (вторая и четвёртая строки) веса РНС (представленные в виде изображений) похожи на собственные лица

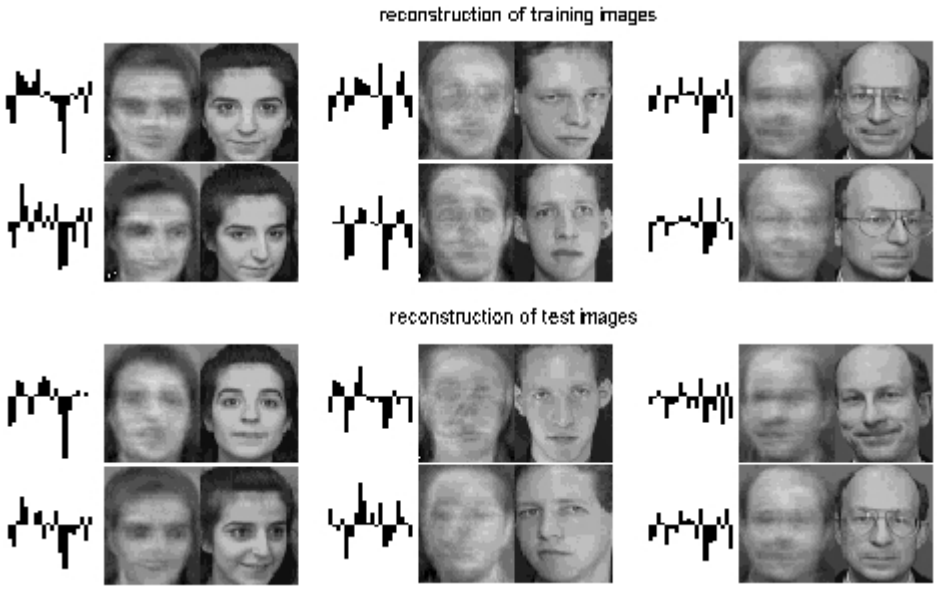

Рис. 13. Компоненты и реконструкция тренировочных и тестовых изображений. Слева направо: диаграмма первых компонент, реконструкция, исходное изображение. Первые две строки – изображения их обучающей выборки, вторые две строки – тестовые изображения

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.

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

Многослойный персептрон

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

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 23x28 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, рис !!!.

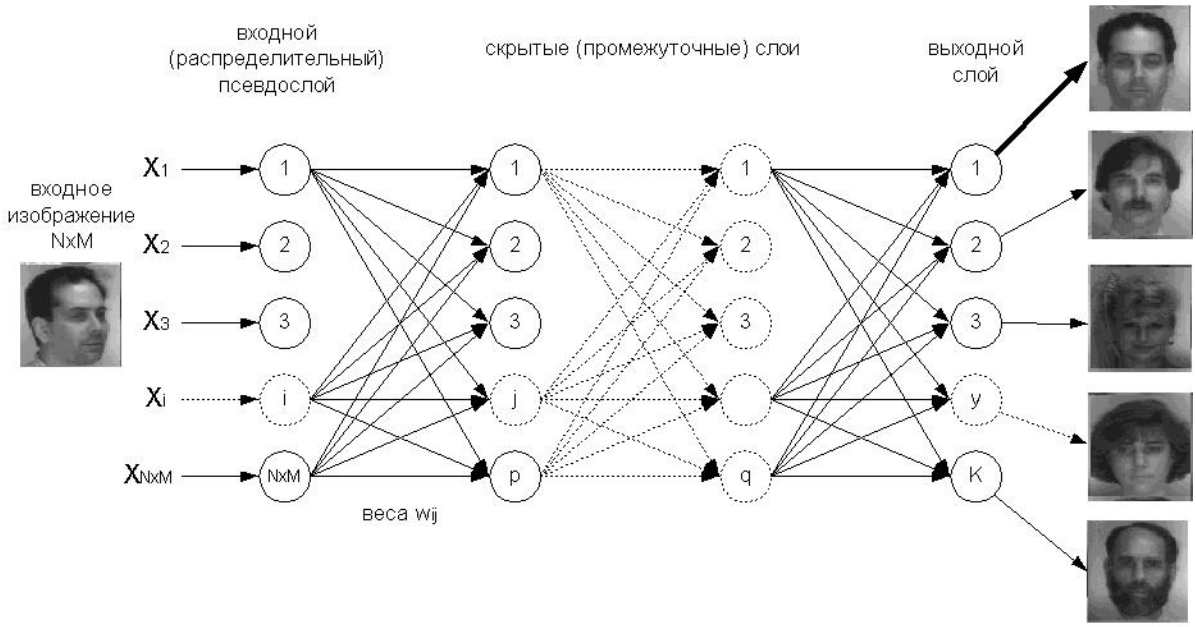

Рис. !!!. Архитектура многослойной нейронной сети и её применение для распознавания изображений. Нейрон с максимальной активностью (здесь первый) указывает принадлежность к распознанному классу.

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).

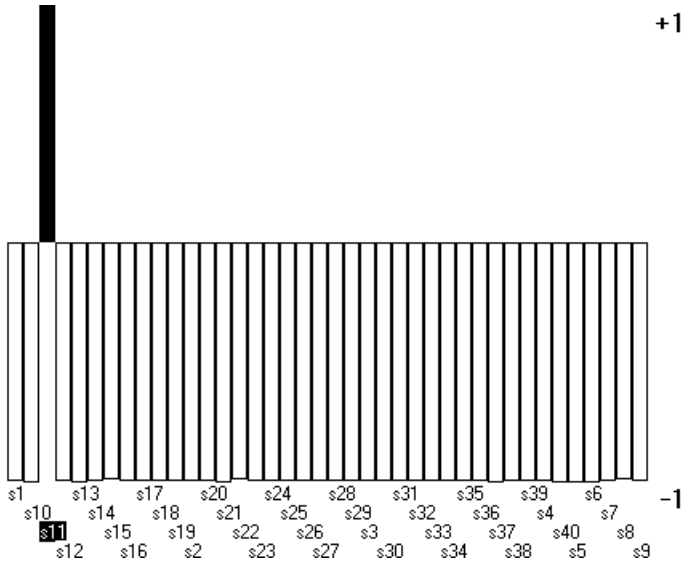

Рис. 15. Выходы сети после обучения с адаптивным шагом. Подано изображение человека из класса *s11*; выход сети, соответствующий классу *s11*, максимален

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 ±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.

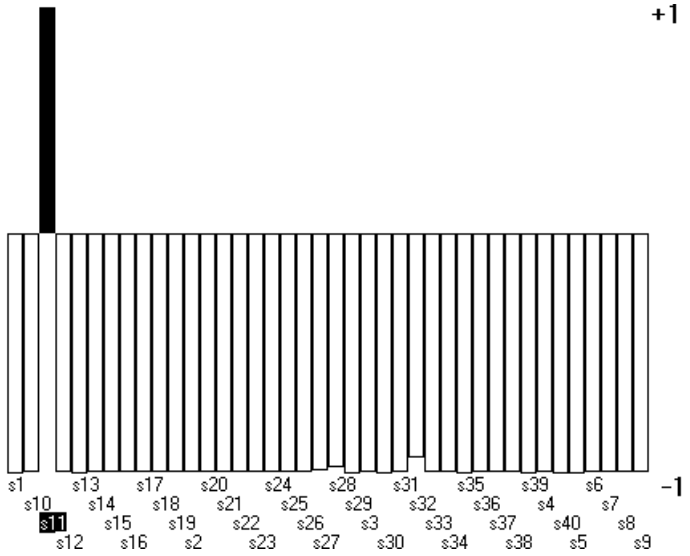

Рис. 16. Пример типичного успешного распознавания тестового изображения. На вход сети подано изображение человека из класса *s11*; выход сети, соответствующий классу *s11*, максимален

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.

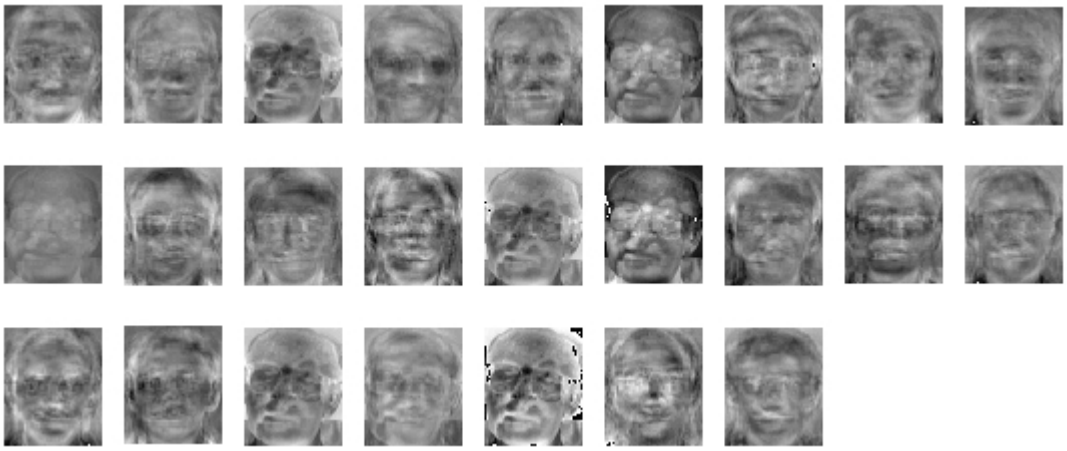

Рис. 17. Входные веса НС, представленные в виде изображений, 25 нейронов в скрытом слое

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].

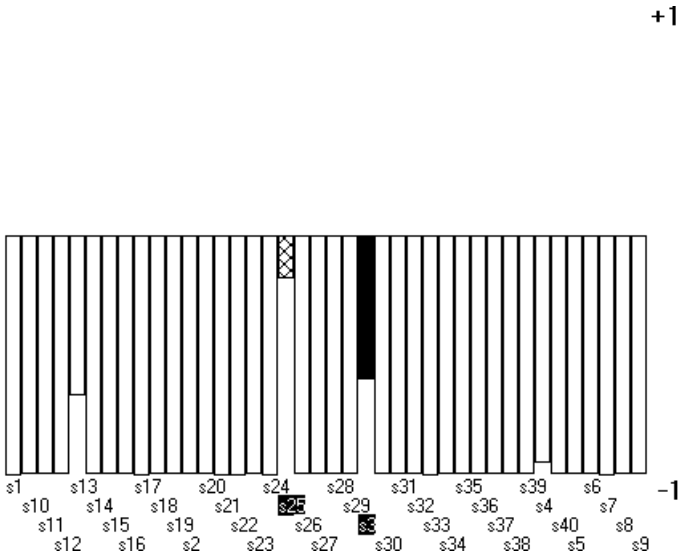

Рис. 18. Типичный случай ошибки первого рода (вместо класса s3 выдаёт s25)

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.

Предобработка и начальное представление изображения

Начальное представление изображения

При формировании исходного представления изображения использовались:

- различный масштаб изображения;
- различное число главных компонент, извлекаемых PHC;
- 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 первых 10-20 главных компонент точность распознавания по первым главным компонентам оказалась хуже, чем по самому изображению, и составила около 90%. С увеличением числа компонент точность повышалась.

Использование части коэффициентов косинусного преобразования позволило существенно повысить скорость обучения, но повышения точности не принесло. Причём с уменьшением числа коэффициентов до 20-30 из 10,000 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.

Геометрическая нормализация изображения

Геометрическая нормализация изображения лица является важным этапом в построении систем распознавания человека по изображению лица. Геометрическая нормализация включает в себя:

- приведение центра лица на изображении к стандартному положению,
- поворот изображения лица таким образом, чтобы оно было вертикально-ориентированным,
- масштабирование изображения лица, чтобы привести его к стандартному размеру,
- выделение на изображении области, соответствующей центральной части лица.

В данных экспериментах использовался модуль нормализации, разработанный в [диссертация Самаля]. Его характерными

особенностями является то, что после геометрической нормализации координаты зрачков левого и правого глаза приводятся к стандартным, заранее определённым координатам. Так же, для поворота изображения используется алгоритм, дающий высокое визуальное качество результирующего изображения.

Примеры начальных и геометрически нормализованных изображений даны на рис. !!!.

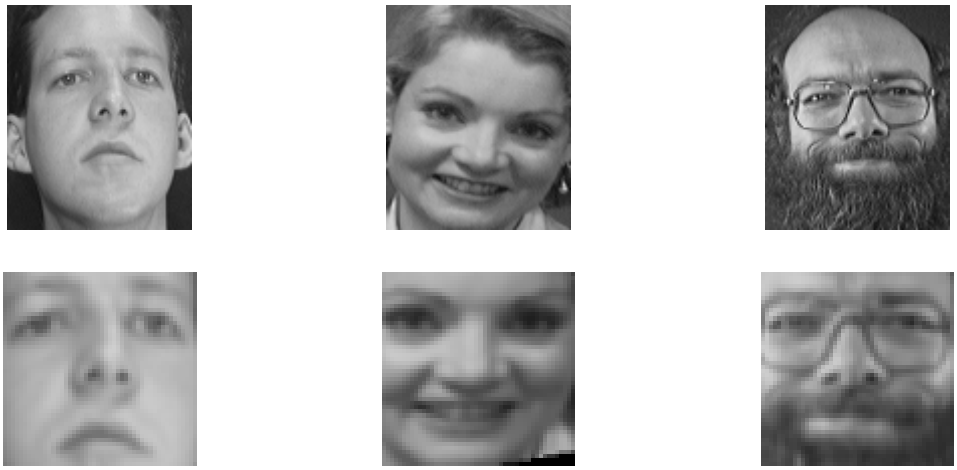

Рис. !!!. Геометрическая нормализация изображений, верхний ряд – исходные изображения, нижний ряд – результирующие изображения.

Эксперименты показали повышение точности распознавания при использовании геометрически нормализованных изображений лиц по сравнению с исходными. Для тестирования использовалась база ORL, 400 изображений 40-ка человек, 10 изображений на каждого человека. Первые 5 изображений от каждого человека использовались для обучения, остальные 5 – для тестирования. При использовании нейронных сетей точность повышалась в среднем с 92% до 95%. Использовалась разновидность нейронных сетей многослойный персептрон, с параметрами и алгоритмами обучения из предыдущего раздела. При использовании метода главных компонент и линейного дискриминантного анализа [Кухарев], точность повышалась с 94% до 98% на том же наборе данных.

Нормализация яркостных характеристик изображения

Одной из главных проблем для систем распознавания человека по изображению лица является изменение характеристик освещения, плохие условия освещения или недостаточность освещения. Поэтому улучшения и нормализация яркостных характеристик

также является важным этапом построения систем распознавания человека по изображению лица.

В данных экспериментах также использовался модуль предобработки изображений лиц, разработанный в [диссертация Самаля]. Для уменьшения влияния изменений яркости использовался оператор хайбустинга (highboosting) совместно с нелинейной коррекцией гистограмм, результаты его работы приведены на рис. !!!. Как видно, результирующие изображения имеют более близкое распределение яркостей.

Рис. !!!. Пример нормализации яркостных характеристик при помощи оператора хайбустинга. В верхнем ряду исходные изображения, внизу – результирующие.

Распознавание при помощи нейронных сетей изображений с нормализованными таким образом яркостными характеристиками, позволило повысить точность с 94% в исходном случае до 95% для нормализованных по яркости изображений на базе ORL.

Применение нормализации яркостей исследовалось так же на базе Yale, которая характеризуется более широкими изменениями условий освещения, рис. !!!.

Рис. !!!. Влияние условий освещения на базе Yale.

Применение нелинейной коррекции гистограмм, разработанной в [диссертация Самаля], позволило понизить ошибку распознавания с 11.85% до 5.04%.

Так же нами были исследованы широкоизвестные и разработаны новые локальные алгоритмы нормализации яркостных характеристик, которые в отличие от глобальных алгоритмов, позволяют компенсировать неравномерное освещение изображения и частично компенсировать влияние направления освещения, рис. !!!.

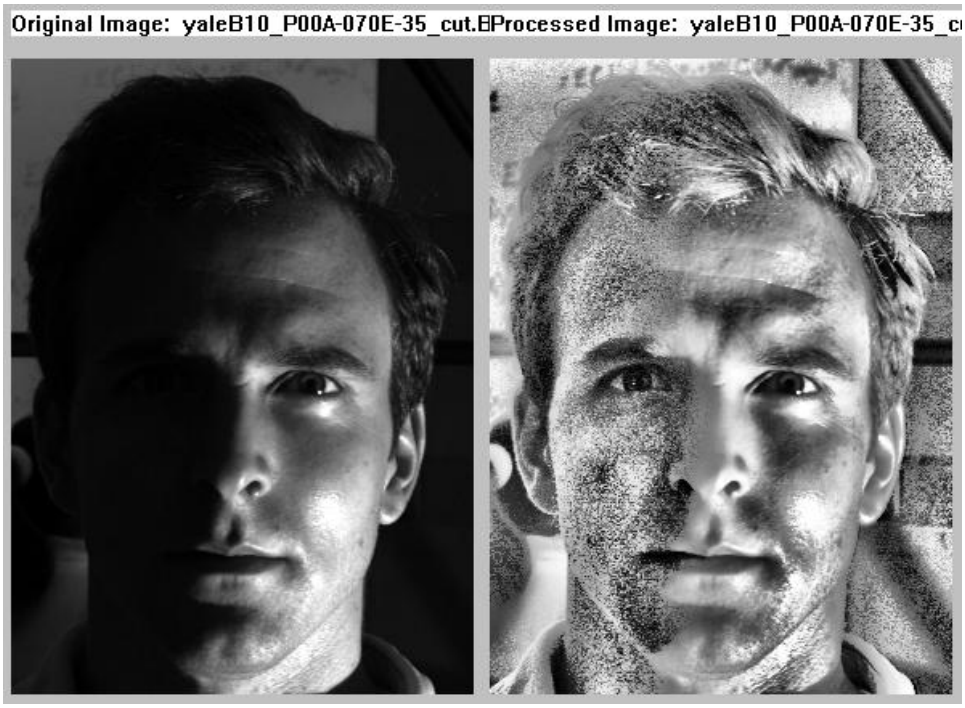

Рис. !!!. Исходное изображение (слева) и результат применения разработанной нами локальной нелинейной коррекции гистограмм (справа).

Отличием разработанного алгоритма локальной нелинейной коррекции гистограмм является то, что этот алгоритм позволяет нормализовать неравномерные изменения яркости по всему изображению, не требуя ручного анализа или выбора параметров для каждого изображения.

Контроль доступа по изображению лица человека с использованием нейронных сетей и отрицательных примеров

В этой серии экспериментов мы изучали применение нейронных сетей для задачи контроля доступа по изображению лица человека. Была изучена надёжность классификаторов, построенных на основе нейронных сетей с учётом ошибок ложного доступа и ложного отклонения. Было предложено пороговое правило для

принятия и отклонения решений нейронных сетей. Было изучено несколько различных конфигураций коллективов нейронных сетей. Показаны преимущества коллективов и приведены параметры лучшей исследованной архитектуры. Было исследовано применение отрицательных примеров. Показано, что используя отрицательные примеры можно повысить точность контроля доступа. Исследованные архитектуры могут быть использованы в режиме реального времени.

В задаче контроля доступа имеется небольшая группа лиц, которым разрешён доступ к некоторому ресурсу (авторизированные лица). Всем остальным (неавторизированным) лицам доступ к этому ресурсу запрещён. Система контроля доступа должна отказывать в доступе таким лицам и пропускать авторизированных лиц.

Задаче распознавания человека по изображению лица посвящено множество работ, однако, большинство из них не касается задачи контроля доступа.

Архитектура нейронной сети

Базовой архитектурой для всех экспериментов послужил многослойный персептрон, описанный в предыдущих разделах. Алгоритм обучения был таким же – обратное распространение ошибки с адаптивным шагом, разработанным Головко. Обучающий процесс сходился быстро и стабильно, рис. !!!. Число обучающих циклов для достижения оптимального результата составило от 50 до 100, поэтому для обучения нейронной сети в дальнейшем использовалось 100 циклов. Точность распознавания на базе ORL, при широкоиспользуемом разделении (первые пять изображений от каждого класса для обучения, остальные для тестирования) составила от 90% без оптимизации параметров до 95% при лучших параметрах. Количество нейронов в скрытом слое так же было выбрано на основе экспериментальных данных, рис. !!!. Точность распознавания для количества нейронов, больше 30 практически не изменялась, поэтому было использовано 30 нейронов в скрытом слое.

После обучения значения на выходах нейронной сети для обучающих примерах практически совпадают с задаваемыми в процессе обучения (идеальными), рис. !!!. Для большинства тестовых примеров выходы НС выглядят как на рис. !!!.

Для большинства случаев ненадёжной или неправильной классификации выходы сети выглядят как на рис. !!!. Однако, могут встречаться случаи как на рис. !!!, когда нейронная сеть уверенно считает что некоторый человек больше похож на другого, чем на

самого себя (возможно вследствие совпадения ракурса или условий освещения). Причём для задачи контроля доступа выходы НС выглядят более разнообразно, если не применять модифицированные алгоритмы обучения.

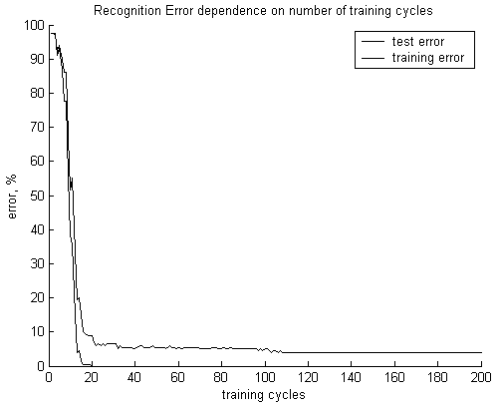

Рис. 1. Зависимость точности распознавания от числа обучающих циклов

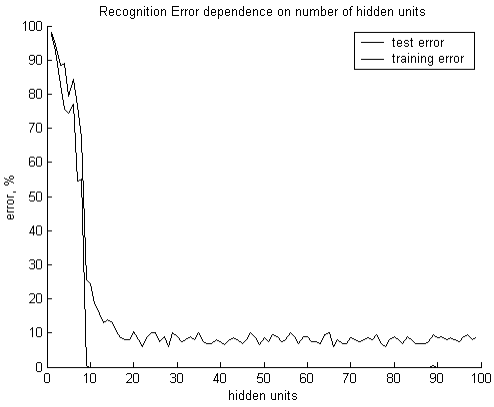

Рис. 2. Зависимость точности распознавания от количества нейронов в скрытом слое

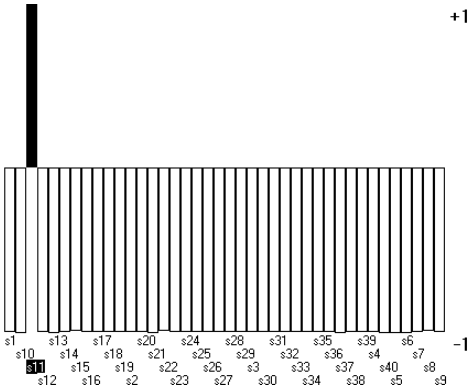

Рис. 3. Результат распознавания обучающего примера, класс «s1» (выходы нейронной сети)

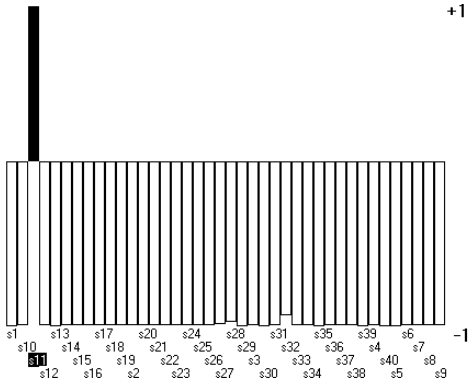

Рис. 3. Результат распознавания тестового примера, класс «s11» (выходы нейронной сети)

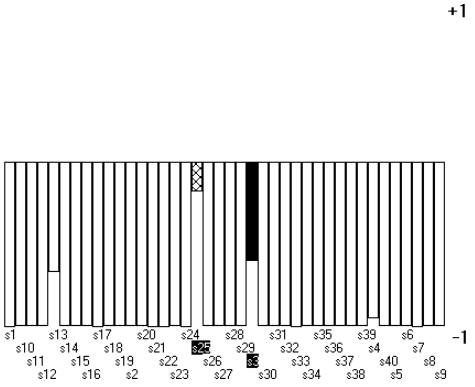

Рис. 5. Неправильное распознавание, класс «s3» (чёрная полоса) распознан как «s25» (серая полоса)

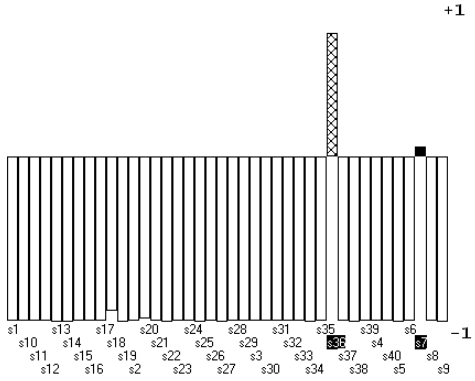

Рис. 6. Сильное неправильное распознавание, класс «s7» (чёрная полоса) распознан как «s36» (серая полоса)

Результаты экспериментов

Для экспериментов по контролю доступа база данных ORL делилась на две части. Первая часть представляла собой авторизированных людей и содержала 20 человек, 5 изображений для каждого человека использовались для обучения (всего 100 чел) и 5 для тестирования (всего 100 чел). Вторая часть представляла собой неавторизированных людей, и содержала 20 человек по 10 изображений на каждого, всего 200 человек.

Характеристиками системы контроля доступа являлись следующие величины. Ошибка ложных пропусков (False Acceptance Rate, ***FAR***), количество пропусков системой неавторизированных лиц делённых на общее число попыток неавторизированного доступа. Ошибка ложных отклонений (False Rejection Rate, ***FRR***), количество отклонённых попыток авторизированного доступа делённое на общее число попыток авторизированного доступа. Равновероятная ошибка (Equality Error Rate, EER), такое значение FAR и FRR. когда они равны.

Эксперимент 1 – пороговые правила

Система контроля доступа должна уметь обрабатывать не только случаи правильной и надёжной классификации, но и случаи неправильной и ненадёжной классификации, отклонять попытки неавторизированного доступа. В этом эксперименте исследовались пороговые правила для отклонения таких случаев. Система контроля доступа должна отклонять большинство таких случаев, но при этом пропускать авторизированных людей.

Первое правило (далее помеченное как ***'min'***), наиболее часто используемое для нейронных сетей, заключается в сравнении значения выхода нейрона-победителя $\boldsymbol{O_{max}}$ (имеющего максимальное значение выхода среди всех нейронов) с некоторым порогом ***t***. Если выходное значение нейрона меньше порога, то решение сети отклоняется, и попытка считается неавторизированным доступом. Если значение выхода нейрона-победителя выше порогового, решение сети принимается и доступ разрешается. Если принять набор выходов нейронной сети за n-мерное пространство (n – количество классов), то это правило представляет собой шахматную метрику. Значение порога может быть в диапазоне от –1 (минимальное значение выхода нейрона) до +1 (максимальное значение выхода нейрона).

Недостаток этого правила заключается в том, что оно не может обрабатывать ситуации как на рис. 6, когда входной класс имеет

высокое выходное значение для других классов. Поэтому разработанное пороговое правило (помеченное как ***'sqr'***), учитывает значения всех выходов. Выбирается нейрон победитель, и за идеальные значения выходов нейронной сети принимаются значения +1 для нейрона-победителя и –1 для всех остальных нейронов. Рассчитывается среднеквадратичное отклонение между реальными значениями и идеальными: $d = \sqrt{\sum_{i=1}^{n}(O_i - \begin{cases}+1, i = \max\\ -1, i \neq \max\end{cases})^2}$. Если такое расстояние ***d*** больше порога ***t***, то решение сети отклоняется как попытка неавторизированного доступа. Иначе решение сети принимается и доступ разрешается. Это правило можно рассматривать как Евклидово расстояние в пространстве выходов нейронной сети, где каждый класс ограничен четвертью окружности радиусом ***t***, и центром окружности в точке идеальных выходных значений, рис. 7. Минимальное значение порога 0, максимальное ограничено размерностью пространства, но на практике для значений порога ***t > 2*** не было случаев ошибочных отклонений, поэтому диапазон для порога был [0; 2].

Графики, отражающие точность обоих правил приведены на рис. 8, 9, 10. Поскольку диапазоны обоих порогов различны, на графике он отображены в один условный диапазон. На графиках показаны усреднённые значения для трёх различных случайных разбиений базы. Для каждого разбиения нейронная сеть обучалась по три раза с различными случайными начальными весами.

Как видно из графиков, предложенное пороговое правило ***'sqr'*** имеет более высокую точность распознавания на всём диапазоне пороговых значений, чем широкоиспользуемое ***'min'***. Так же предложенное пороговое правило имеет значительно более низкую ошибку ложных пропусков и немного более высокую ошибку ложных отклонений. Другими словами, предложенное правило намного строже относится к попыткам чужих получить доступ и немножко строже к своим людям.

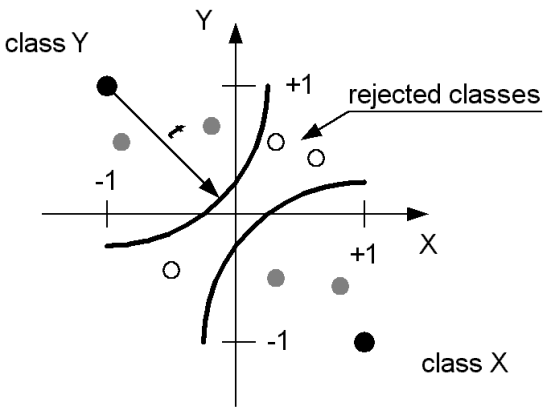

Рис. 7. Пространство выходов НС и пороговое правило ***'sqr'***

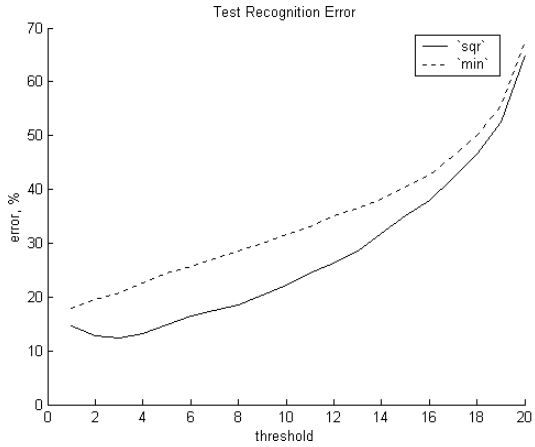

Рис. 8. Точность распознавания для обоих пороговых правил

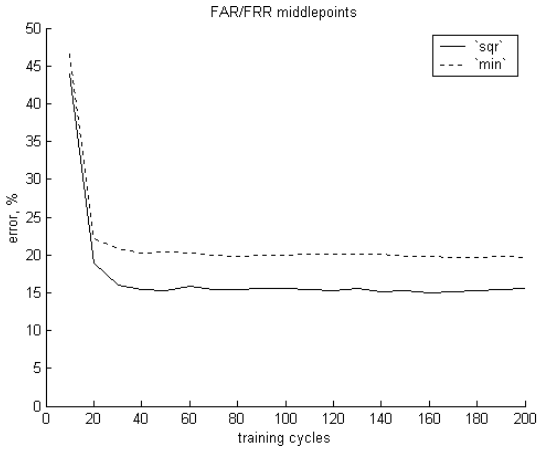

Рис. 9. Равновероятные ошибки ложного пропуска и отклонения для обоих правил

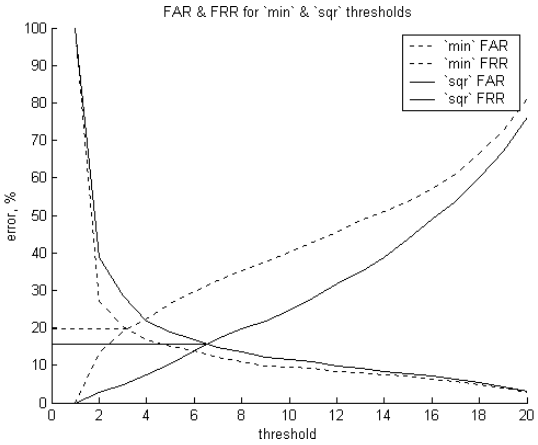

Рис. 10. Ошибки ложного пропуска и отклонения для обоих правил

Эксперимент 2 – коллективы нейронных сетей

В этом эксперименте исследовалась точность четырёх различных архитектур на основе многослойной нейронной сети. Первая архитектура, ***'mlp'***, является многослойной нейронной сетью с пороговым правилом ***'sqr'***.

Вторая архитектура, ***'one-one'***, является коллективом из 40 многослойных нейронных сетей. Каждая сеть имеет только один выход и обучается распознавать только один класс. Каждая сеть имеет 2 слоя, 20 нейронов в скрытом слое и один выходной нейрон, который обучается выдавать +1 для своего класса и –1 для всех остальных. Потом выходы всех сетей из коллектива подаются на пороговое правило ***'sqr'***.

Третья архитектура основывается на второй, с тем отличием, что каждая сеть имеет два выходных нейрона, второй из которых

обучается выдавать наоборот, -1 для своего класса и +1 для всех остальных. Таким образом мы проверяли эмпирическое правило, утверждающее, что чем больше различных целей при обучении имеет нейронная сеть, тем качественнее процесс обучения. На пороговое правило подаётся выходное значение только первого нейрона.

Четвёртая архитектура, ***'all-all'***, являлась коллективом многослойных нейронных сетей, описанных ранее. Решение принималось путём голосования сетей, составляющих такой коллектив. Каждая сеть или даёт голос за некоторого человека или может воздержаться и не голосовать. Порог для воздержания равняется 1.2 и был подобран экспериментально, использовалось пороговое правило ***'sqr'***. Затем решающее правило подсчитывает количество голосов за каждый класс. Класс, имеющий максимальное количество голосов (но не меньше двух), принимается за распознанный класс. Потом количество голосов за распознанный класс сравнивается с порогом. Если этот класс имеет количество голосов, меньшее чем порог, то он принимается за неавторизированный и отклоняется. Мы экспериментально проверили точность такой архитектуры в зависимости от числа нейронных сетей в коллективе. Точность повышалась с увеличением количества нейронных сетей и достигала оптимума при семи сетях.

Сначала была проверена точность всех четырёх архитектур на задаче распознавания изображений лиц, рис. 11. Все классы были использованы как при тестировании, так и при обучении. График показывает усреднённые значения по нескольким разбивкам базы и нескольким различным вариантам обучения при разных начальных весах сетей. Как можно видеть из рис. 11, четвёртая архитектура имеет лучшую точность. Вторая и третья архитектуры имеют точность меньше чем у многослойной нейронной сети. Как и предполагалось, точность у третьей архитектуры выше, чем у второй.

Затем каждая из архитектур был проверена на задаче контроля доступа. Для обучения использовалось 20 классов из 40-ка, 5 на обучение и 5 на тестирование. Результаты применения различных архитектур показаны на рис. 12-15. Первая и четвёртая архитектуры имеют лучшую точность распознавания, рис. 12. Четвёртая архитектура имеет наименьшую равновероятную ошибку, рис. 14, потом идёт первая архитектура. Вторая и третья архитектуры имеют самые низкие ошибки ложного пропуска, но при этом у них высоки ошибки ложного отклонения.

В целом четвёртая архитектура (коллектив из семи многослойных нейронных сетей, где каждая сеть распознаёт каждый

класс) имеет лучшую точность как для распознавания так и для контроля доступа.

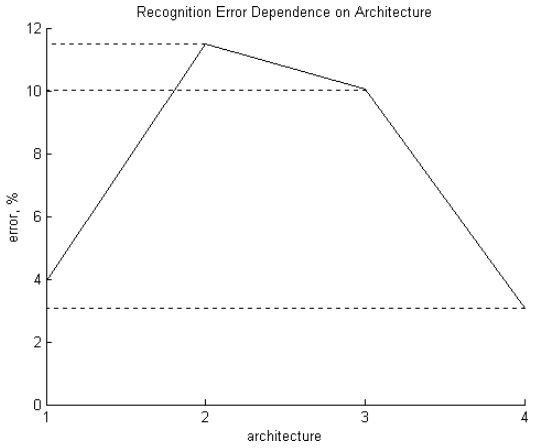

Рис. 11. Точность распознавания для классификации

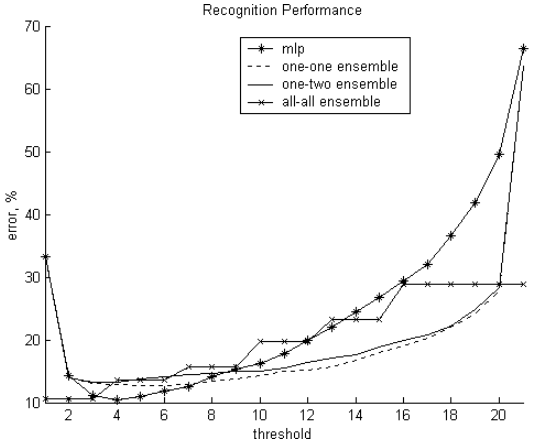

Рис. 12. Точность распознавания для задачи контроля доступа

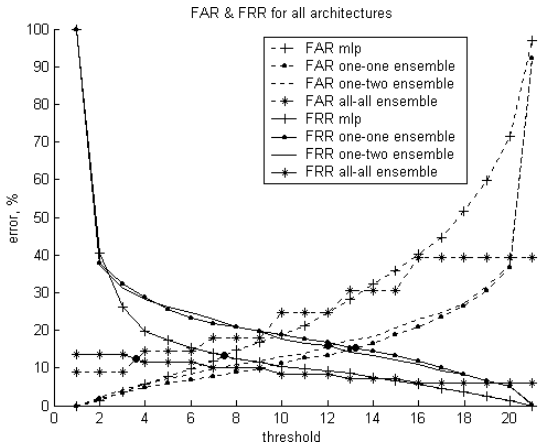

Рис. 13. Ошибки ложного допуска и ложного отклонения для всех архитектур для всего диапазона порогов

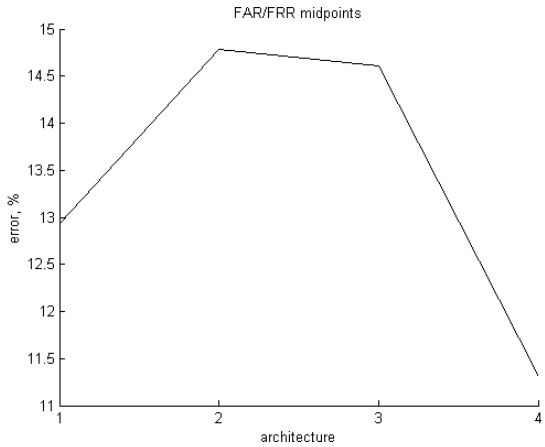

Рис. 14. Равновероятные ошибки для всех архитектур для всего диапазона порогов

Эксперимент 3 – использование отрицательных примеров

In this set of experiments we have explored the usage of negative examples. Negative example is an image of a person, which is always considered as alien. Such ‘negative’ persons were used only for training.

The database was divided into three parts. The first part consists of ten persons, which were the ‘negative’ persons. All ten images of each person were used only for training. The second part consists of ten ‘authorized’ persons. Five images of each person were used for training and five for testing. The third part consists of twenty ‘alien’ persons. Such persons were considered as unknown alien persons, and all ten images of each person were used only for testing.

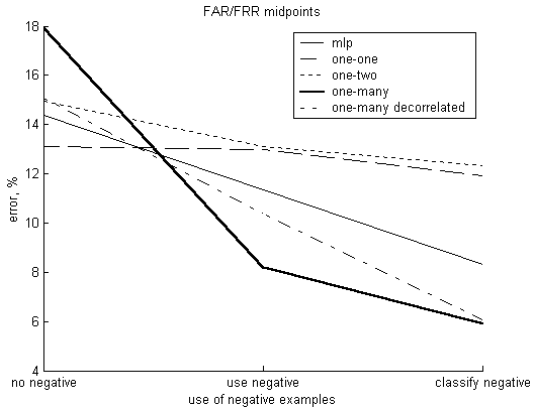

Fig. 15. Equality errors for different usage of negative examples

We have studied three cases of using negative examples. In the first case, the first, 'negative' part, was not used for training. We need such step to find out how performance is improved by using negative examples. In the second case, the neural network was trained to reject all negative examples. It means that each output of neural network trained to respond '-1' for any negative example. And in the third case neural network was trained to classify all negative examples in the same manner as positive examples (authorized persons). It means that for each negative person neural network has separate output, which is trained to respond '+1' for its person and '-1' for others.

Fig. 15 shows performance of all architectures for each mentioned case. As can be seen, by using negative examples we can improve performance. When the unknown image is classified as one of the negative persons, we consider such person as alien and reject him. So, by using and classifying the negative examples, we can achieve better performance.

Also an attempt to decorrelate errors of neural networks in ensemble was made. As we have found, the best performance is achieved when the networks in ensemble have different initial weights and different example order.

Conclusion

As can be seen from the experimental results, the more different goals NN have to learn, the better performance is. A collective decision is better than a decision of one network. Also the introduced ***'sqr'*** thresholding rule has better performance for rejection an unauthorized persons than ***'min'*** thresholding rule. By using negative examples, we can significantly improve the performance for access control task.

Improvements presented in the paper are insufficient for creation a real access control system. First, an image must be normalized in

brightness and contrast, face orientation and scale to bring it to uniform conditions. This is required to exclude systems reaction on similar shooting conditions that may be greater than difference between two different persons. Second, we need to develop technique, which will allow expanding training set in order to compensate lack of examples with different poses, lightning conditions, expressions and generation of 'synthetic' negative examples.

References

1. 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.

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).

2. 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.

3. Eickeler S., Jabs M., Rigoll G. Comparison of Confidence Measures for Face Recognition. Gerhard-Mercator-University Duisburg, Germany, 1999. - 6 p.

4. A.I. Wasserman Neural Computing: Theory and Practice – New York: Van Nostrand Reinhold, 1989.

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

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.