

# АЛГОРИТМ РАСПОЗНАВАНИЯ ЛИЦ НА ОСНОВЕ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ С ОБРАТНЫМИ СВЯЗЯМИ

ТАРАСЮК А.Е.

Белорусский государственный университет, ФПМИ, кафедра МО АСУ  
Пр. Независимости, 4, г. Минск, Республика Беларусь  
телефон: + (375 29) 276 31 10; e-mail: alexandr.tarasyuk@gmail.com

**В работе описывается подход к решению задачи распознавания лиц. Предлагается алгоритм, основанный на свёрточных нейронных сетях, обладающий инвариантностью к уровню освещённости и шумам, и требующий меньше времени для обучения.**

**Ключевые слова –** вейвлетное сжатие, нормализация, обратные связи, распознавание лиц, свёрточные нейронные сети.

## 1 ВВЕДЕНИЕ

В настоящее время системы идентификации человека по изображению лица широко используются (например, для верификации документов и в сфере безопасности). К их преимуществам относятся дешевизна и мобильность оборудования, быстрота и массовость идентификации.

Существует множество подходов к решению этой проблемы (собственные лица, эластичные графы, скрытые Марковские модели, многослойный персептрон), однако все они могут быть применены лишь в частных случаях. Одним из подходов является использование свёрточных нейронных сетей[1]. Его преимуществами являются хорошая точность, инвариантность к положению лица на изображении, гибкость архитектуры для настройки на конкретную предметную область. В работе предлагается модификация классической свёрточной нейронной сети, включающая:

- а) вейвлетную предобработку изображения;
- б) локальную нормализацию;
- в) введение обратных связей;
- г) адаптивный шаг обучения.

Результаты экспериментов на базе изображений ORL для предложенного алгоритма свидетельствуют об уменьшении количества ошибок распознавания в 1,2 – 2,5 раза по сравнению с обычными свёрточными сетями.

## 2 АНАЛИЗ ПРОБЛЕМЫ И ПОСТАНОВКА ЗАДАЧИ

В общем случае в задаче распознавания лиц дано множество изображений определённого типа и множество классов. Для некоторого конечного набора изображений (называемого обучающей выборкой) известна принадлежность изображений к классам. Требуется построить алгоритм, который находил бы принадлежность к классам для всех изображений этого типа.

Большинство методов состоит из предобработки изображения и сравнения изображения посредством некоей метрики, на основании чего изображение относится к определённому классу.

Предобработка используется для увеличения скорости работы системы при помощи отбрасывания несущественной информации и для обеспечения устойчивости метода к изменению освещения, расположения лица на изображении и т.д. путём приведения изображения к некоему каноническому виду.

Для сравнения и классификации изображений существующие подходы к распознаванию чаще всего используют анализ характера искажения изображения либо выделение набора числовых признаков и разделение рассматриваемого пространства признаков на области.

Большая часть подходов относится ко второй группе. Среди них как простейшие методы (геометрические, метод линейного дискриминанта), так и более сложные (метод собственных лиц, нейронные сети, векторная машина). Геометрические методы благодаря своей простоте могут успешно применяться в задачах верификации. В случае большого количества классов идентификации преимущество получают методы линейного дискриминанта и собственных лиц из-за их быстрой работы и масштабируемости. Однако, им не хватает точности, необходимой в задачах с небольшим количеством классов (до 100). В этом случае чаще используются нейронные сети и векторные машины.

Тем не менее, обычным нейронным сетям не хватает устойчивости к изменению освещённости и положения лица на изображении (а также наклона и поворота головы). С целью преодолеть эти недостатки были разработаны свёрточные нейронные сети [1].

Благодаря своей архитектуре, учитывающей двумерность изображений, свёрточная сеть обладает большей инвариантностью к положению лица на изображении, а признаки, выделяемые при обучении, меньше зависят от обучающей выборки, в следствие чего обобщающая способность выше чем у многослойного персептрона. Однако, проблема недостаточной инвариантности к качеству освещённости и шумам присуща и свёрточным сетям.

Таким образом, в данном исследовании была поставлена задача разработать алгоритм для распознавания человека по портретному изображению, основанный на использовании свёрточных нейронных сетей, обладаю-

щий инвариантностью относительно шумов и уровня освещённости.

### 3 РЕШЕНИЕ

Как и для многослойного персептрона, работа алгоритма свёрточных нейронных сетей состоит из двух этапов:

- обучение сети (настройка весов) на тренировочной выборке;
- использование сети для распознавания.

Обучение проводится методом обратного распространения ошибки, модифицированным с учётом архитектуры сети.

Так как вычислительная трудоёмкость алгоритма высока, изображения масштабируются. Архитектура типичной свёрточной сети изображена на Рис.1.

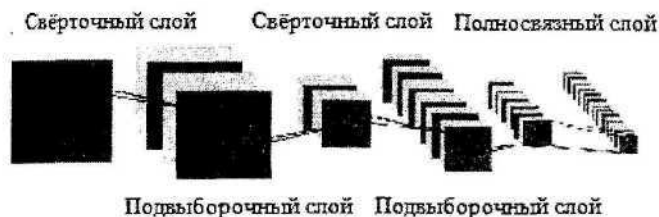


Рис. 1. Архитектура свёрточной нейронной сети

Сеть содержит слои 3 видов: свёрточные, подвыборочные и полносвязные. Подвыборочные слои служат лишь для масштабирования, их веса постоянны. Свёрточные слои служат для нахождения признаков. Полносвязные слои представляют собой обычный персептрон с количеством нейронов, совпадающим с количеством классов идентификации, и служат для финальной классификации.

Для проведения подробных исследований работы базового алгоритма нами была создана тестирующая система с подробной визуализацией состояния сети и полученных результатов. На основе анализа этих данных были сделаны выводы об основных причинах ошибок в процессе работы базового алгоритма.

Часть ошибок возникает из-за шума на изображении, который при масштабировании даёт эффект «смазывания». Более частая причина – это неполная инвариантность алгоритма к уровню освещённости. Из рассмотрения значения выходов первого слоя нейронов становится ясно, что после обучения значительная часть нейронов начинает реагировать лишь на освещённость, явно отделяя фон от объекта. Следующая группа ошибок возникает для изображений, содержащих лицо со значительным поворотом или наклоном (при условии, что в обучающей выборке для этого класса подобного изображения не было).

Для уменьшения влияния шумов и для общего улучшения точности на стадии предобработки предлагается использовать вейвлетное преобразование.

В вейвлетном представлении функция представлена суммой некоторой базовой функции (называемой вейвлетом) при различных сдвигах и масштабах. Вейвлетное

сжатие основано на отбрасывании малозначимых вейвлетных коэффициентов, на несколько порядков отличающихся от коэффициентов, представляющих детали общего плана. [2] Благодаря своим свойствам оно значительно уменьшает потери важной информации, одновременно избавляя изображение от локальных помех и шумов.

Для увеличения устойчивости алгоритма к качеству освещения обычно применяется метод нормализации. Его суть состоит в приведении статистических характеристик изображения (математического ожидания и дисперсии значений пикселей) к фиксированным значениям.

Главное преимущество данного метода – его простота и высокая скорость. Основной недостаток проявляется в случае, когда фон «задавливает» информативную часть изображения, вследствие чего их значение слабо отличается. В качестве решения предлагается проводить нормализацию локально – для каждого набора входных значений нейрона. Процесс обучения также модифицируется с учётом нормализации. Данный механизм применяется в первом свёрточном слое. По результатам тестирования можно сделать вывод, что наибольшая точность достигается в случае, когда в первом свёрточном слое присутствуют как нейроны с нормализацией, так и без.

Для решения проблемы устойчивости к изменению положения наиболее простым способом является увеличение числа слоёв в сети, что влечёт за собой большую независимость полученных признаков от расположения на изображении. Однако, это значительно увеличивает время, затрачиваемое на обучение. Для компенсации был использован метод адаптивного обучающего шага, предложенный в работе [3]. Он заключается в выборе на каждой итерации индивидуального значения шага с целью минимизации среднеквадратичной ошибки сети. Это обеспечивает значительно более быстрое и устойчивое схождение весов в сети. После адаптации механизма для случая свёрточной сети удалось повысить количество свёрточных слоёв в сети до трёх при сохранении приемлемого времени обучения.

Следующая составляющая модификации направлена как на повышение инвариантности к положению лица, так и на общее увеличение точности алгоритма для простых условий. Благодаря использованию обратных связей в нейронных сетях информация о признаках нижнего уровня, важная для конкретного случая распознавания, получает акцент. Однако, сложность и трудоёмкость алгоритма значительно возрастает, а устойчивость обучения снижается. Модифицированный алгоритм содержит упрощённые обратные связи, которые участвуют в работе сети, но не корректируются в процессе обучения. Вместо этого вычисление их весов происходит после окончания цикла обучения: при помощи персептрона определяется зависимость между признаками последнего уровня и признаками 1-го уровня. При работе алгоритма после получения значений для признаков высокого уровня происходит коррекция признаков низкого уровня в соответствии с выявленными взаимосвязями. Данная процедура повторяется до достижения равновесия.

## ТОЧНОСТЬ АЛГОРИТМОВ ДЛЯ БАЗЫ FERET

Алгоритм	1	2	3	4
Базовый	80%	76,5 %	78,7%	78,5%
Модифицированный	86%	84,2%	88%	87,2%

## 5 ЗАКЛЮЧЕНИЕ

В работе описан подход к проблеме распознавания человека по изображению лица, основанный на применении свёрточных нейронных сетей.

Была предложена модификация, направленная на решение основных недостатков свёрточных сетей - неполной инвариантности алгоритма к степени яркости изображения, шумам и местоположению лица на изображении. Она включает вейвлетную предобработку изображения, локальную нормализацию, использование обратных связей. Для ускорения процесса обучения алгоритм был дополнен выбором адаптивного шага обучения.

Для экспериментов на базах ORL и FERET предложенный алгоритм показал большую точность распознавания в сравнении с классическим алгоритмом свёрточных сетей. Уровень ошибок распознавания снизился в 1,2 – 2,5 раз в зависимости от конфигурации сети и набора изображений.

## ЛИТЕРАТУРА

- [1] 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.
- [2] Al-Haj A. Wavelets pre-processing of Artificial Neural Networks classifiers // IEEE Systems, Signals and Devices, 2008. P. 1-5.
- [3] Головкин В.А. Нейроинтеллект: Теория и применения. Книга 1. Организация и обучение нейронных сетей с прямыми и обратными связями - Брест: БПИ, 1999, - 260с.
- [4] Краснопрошин В.В., Коблов Е.В. Эффективный алгоритм поиска лицеподобных областей // Вестник БГУ. Сер.1. 2007. № 2, -с. 97-102.

Эксперименты показали уменьшение ошибок в случаях, когда стандартный алгоритм при распознавании колеблется между двумя классами. Стоит, однако, отметить и негативный эффект данной модификации – увеличение точности достигается в обмен на значительное уменьшение случаев «неуверенности» алгоритма в ответе, что не позволяет выявить необходимость проведения повторной идентификации для получения надёжного результата.

## 4 РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для тестирования реализованного алгоритма использовались базы ORL и FERET (содержит изображения большей сложности).

База ORL содержит 400 портретных фотографий принадлежащих 40 людям (по 10 изображений на человека). При тестировании база случайным образом разбивалась на 2 непересекающихся набора по 200 изображений (5 на человека). Один набор использовался в качестве обучающей выборки, другой – в качестве тестовой.

Из базы FERET при экспериментах использовались изображения, относящиеся к классам, представленными не менее 4 портретами. База аналогичным образом равномерно разбивалась на 2 выборки.

Для полноценного сравнения эксперименты проводились с различными конфигурациями нейронных сетей. Результаты приведены для 4 наиболее типичных конфигураций:

1. 2 свёрточных слоя (10 и 30 нейронов)
2. 2 свёрточных слоя (20 и 20 нейронов)
3. 3 свёрточных слоя (10, 20 и 20 нейронов)
4. 3 свёрточных слоя (5, 10 и 10 нейронов)

В таблице 1 содержатся результаты экспериментов для базы ORL. Как видно, для базовой версии алгоритма результаты лучше при 2 свёрточных слоях, в то время как модифицированный алгоритм показывает большую точность при наличии 3 слоёв. В целом, доля ошибок для модифицированного алгоритма меньше в 1,2 – 2,5 раза.

Для базы FERET результаты хуже в силу её сложности. Они приведены в Таблице 2.

Анализируя конкретные случаи ошибок, можно сделать вывод, что рост точности происходит в первую очередь за счёт изображений с шумами и плохим качеством освещения.

Таблица 1

## ТОЧНОСТЬ АЛГОРИТМОВ ДЛЯ БАЗЫ ORL

Алгоритм	1	2	3	4
Базовый	95%	93,5 %	94%	92,5%
Модифицированный	96%	94,5%	97,5%	97%