

ТЕЛЕНИК С.Ф.,
ДОРОГИЙ Я.Ю.

ИССЛЕДОВАНИЕ ПАРАМЕТРОВ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ЧЕЛОВЕКА ЗА ФОТОПОРТРЕТОМ

В статье рассмотрен вопрос влияния параметров сверточной нейронной сети для задачи распознавания человека за фотопортретом. В ходе исследования найден оптимальный набор параметров для решения задачи распознавания человека на базе ORL Faces.

In the article the question of influence the parameters of the convolution neural network for pattern recognition problem in human portraits was considered. The study found the optimal set of parameters to solve the problem of recognizing a person on ORL Faces dataset.

Введение в проблему

В настоящее время обнаружение человеческого лица на изображении или в видеопотоке становится очень важной темой для научных исследований из-за широкого диапазона ее возможных приложений, таких как управление контролем доступа в системах безопасности, человеко-машинное взаимодействие и многих других. Также, обнаружение лица на изображении является первым шагом в процессе решения задачи идентификации личности человека по изображению лица или распознавания эмоционального выражения лица.

Сейчас для решения данной задачи начинают интенсивно применяться искусственные нейронные сети. Искусственные нейронные сети давно и успешно применяются для решения многих задач распознавания. Достоинством использования нейросетей для решения задачи обнаружения/распознавания лица является возможность получения классификатора, хорошо моделирующего сложную функцию распределения изображений лиц. Недостатком же является необходимость в тщательной и кропотливой настройке нейросети для получения удовлетворительного результата классификации.

В последнее время основной топологией нейронной сети используемой для решения задачи локализации лица на изображении является полносвязная нейронная сеть без обратных связей, так называемый многослойный персептрон. Но его применение к решению данной задачи представляет несколько трудностей:

- во-первых, как правило, изображения имеют большую размерность, соответственно вырастает размер нейронной сети (количество нейронов и весов). Большое количество параметров увеличивает емкость системы и соответственно требует большей обучающей выборки, что увеличивает время и вычислительную сложность процесса обучения;
- во-вторых, недостаток полносвязной архитектуры в том, что топология ввода полностью игнорируется. Входные переменные могут быть представлены в любом порядке, не затрагивая цель обучения. Напротив, изображения имеют строгую 2-мерную местную структуру: переменные (пиксели), которые являются пространственно соседними, чрезвычайно зависимы.

На преодоление этих недостатков направлена архитектура сверточных нейронных сетей.

Статья является продолжением исследований сверточных нейронных сетей, впервые описанных в [1,2].

Анализ существующих решений

В 1981 году нейробиологи Торстен Визел и Дэвид Хабел исследовали зрительную кору головного мозга кошки и выявили, что существуют так называемые простые клетки, которые особенно сильно реагируют на прямые линии под разными углами и сложные клетки, что реагируют на движение линий в одном направлении.

Позднее Ян Лекун предложил использовать так называемые сверточные нейронные сети, как аналог зрительной коры головного мозга для распознавания изображений [1,2].

Данный тип сетей хорошо зарекомендовал себя для решения проблемы распознавания человека по фотопортрету, но задача выбора оптимальных параметров сети является открытой и плохо описанной в литературе.

Цель работы

Ввиду неочевидности работы сверточных нейронных сетей, а также множества факторов, которые влияют на скорость ее обучения (количество карт признаков, плотность связей между картами признаков, размер окна, площадь перекрытия, начальная инициализация весов и прочие), очень сложно заранее предугадать, какие значения будут оптимальными для конкретной практической задачи.

В данной работе рассмотрен ряд экспериментов, в ходе которых были подобраны оптимальные параметры для решения задачи распознавания человеческих лиц с помощью свёрточной нейронной сети на базе изображений ORL Faces.

Архитектура классической свёрточной нейронной сети

Идея свёрточных нейронных сетей заключается в чередовании свёрточных слоев (C-layers), субдискретизирующих слоев (S-layers) и наличии полносвязных (F-layers) слоев на выходе.

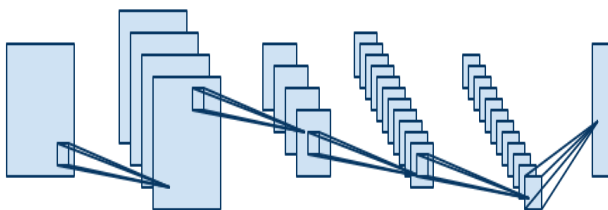


Рис. 1. Структура СНС

Такая архитектура включает в себе 3 основных парадигмы:

- локальное восприятие;
- разделяемые веса;
- субдискретизация.

Локальное восприятие подразумевает, что на вход одного нейрона подается не всё изображение (или выходы предыдущего слоя), а лишь не-

которая его область. Такой подход позволяет сохранять топологию изображения от слоя к слою.

Концепция разделяемых весов предполагает, что для большого количества связей используется небольшой набор весов. Т.е. если у нас имеется на входе изображение 32x32 пикселя, то каждый из нейронов следующего слоя примет на вход только небольшой участок этого изображения размером, к примеру, 5x5, причем каждый из фрагментов будет обработан одним и тем же набором. Важно понимать, что самих наборов весов может быть много, но каждый из них будет применен ко всему изображению. Такие наборы часто называют ядрами (kernels).

Большинство систем распознавания изображений строятся на основе двумерных фильтров. Фильтр представляет собой матрицу коэффициентов, обычно заданную вручную. Эта матрица применяется к изображению с помощью математической операции, называемой сверткой. Суть этой операции в том, что каждый фрагмент изображения умножается на матрицу (ядро) свертки поэлементно и результат суммируется и записывается в аналогичную позицию выходного изображения. Основное свойство таких фильтров заключается в том, что значение их выхода тем больше, чем больше фрагмент изображения похож на сам фильтр. Таким образом, изображение, свернутое с неким ядром даст нам другое изображение, каждый пиксель которого будет означать степень схожести фрагмента изображения на фильтр.

Разберем более подробно процесс распространения сигнала в С-слое.

Каждый фрагмент изображения поэлементно умножается на небольшую матрицу весов (ядро), результат суммируется. Эта сумма является пикселем выходного изображения, которые и формируют карту признаков. Следует сказать, что в идеале не разные фрагменты проходят последовательно через ядро, а параллельно всё изображение проходит через идентичные ядра. Кроме того, количество ядер (наборов весов) определяется разработчиком и зависит от того какое количество признаков необходимо выделить. Еще одна особенность свёрточного слоя в том, что он немного уменьшает изображение за счет краевых эффектов.

Суть субдискретизации и S-слоев заключается в уменьшении пространственной размерности

изображения. Т.е. входное изображение грубо (усреднением) уменьшается в заданное количество раз. Чаще всего в 2 раза, хотя может быть и не равномерное изменение, например, в 2 раза по вертикали и в 3 раза по горизонтали.

Использование субдискретизации необходимо для обеспечения инвариантности к масштабу.

Чередование слоев позволяет составлять карты признаков из карт признаков, что на практике означает способность распознавания сложных иерархий признаков.

Обычно после прохождения нескольких слоев карта признаков вырождается в вектор или даже скаляр, но таких карт признаков становится сотни. В таком виде они подаются на один-два слоя полносвязной сети. Выходной слой такой сети может иметь различные функции активации. В простейшем случае это может быть тангенциальная функция, также успешно используются радиальные базисные функции.

Конфигурация пакета PANN для работы со сверточными нейронными сетями

Рассмотрим конфигурацию необходимую для эмуляции сверточных сетей. Как описывалось в [3,4], конфигурация задается в XML-файле:

```
<config>
  <net>
    <layer>5</layer>
    <layer>6</layer>
    <layer>7</layer>
    <density>0.2</density>
    <window_height>5</window_height>
    <window_width>5</window_width>
    <window_vert_overlap>3</window_vert_overlap>
    <win-
dow_horiz_overlap>3</window_horiz_overlap>
    <threads>8</threads>
    <seed>42</seed>
  </net>
  <weight_randomization>
    <min>-0.1</min>
    <max>+0.1</max>
  </weight_randomization>
  <lms>
    <learning_rate>0.1</learning_rate>
    <annealing_tsc>100</annealing_tsc>
    <epochs>200</epochs>
```

```
</lms>
</faces>
  <men>5</men>
  <train_percent>60</train_percent>
  <report_frequency>1</report_frequency>
  <stop_error>0.5</stop_error>
</faces>
</config>
```

Рассмотрим параметры более детально:

- layer – количество карт признаков в паре слоев подвыборки-свертки.
- density – плотность связей между картами признаков слоя свертки предыдущей пары слоев и карт признаков слоя подвыборки следующей пары слоев. 0 – нет связей, 1.0 – полносвязная связь.
- window_height/width – размер окна подвыборки.
- window_horiz/vert_overlap – 3/3 – площадь перекрытия соседних окон.
- threads – количество одновременных потоков.
- seed – база для генератора псевдослучайных чисел. Этим параметром инициализируется генератор псевдо-случайных чисел. Таким образом, при каждом запуске сеть будет иметь одно и тоже начальное состояние. Это избавляет от внесения нежелательных флуктуаций в эксперимент и позволяет анализировать качество обучения при изменении различных параметров.
- weight_randomization – разброс весов.
- LMS – алгоритм обучения по методу градиентного спуска с мерой ошибки по методу наименьших квадратов.
- learning_rate – скорость обучения.
- annealing_tsc – временная константа для обучения с моделированием отжига. Моделирование отжига при изменении веса во время обучения учитывает еще и величину изменения во время предыдущего прогона. Т.е. отжиг как бы ни дает весу резко поменять направление изменения веса в пространстве весов.
- train_percent – размер обучающего множества в процентах.
- report_frequency – частота отклика.

- stop_error – общая ошибка системы, которая считается приемлемой и при ее достижении обучение останавливается.

Эксперименты и результаты

В данном исследовании использовался ранее разработанный пакет PANN [3,4]. Эксперименты проводились на общепринятом наборе лиц для тестирования систем распознавания человека Olivetti Research Lab's (ORL) Face Database. В ней имеются по 10 фотографий 40 различных людей.



Рис. 2. Лица из базы ORL Face

Проведя предварительный анализ сверточных нейронных сетей, а также изучив результаты исследований [2] был выбран ряд параметров:

$layer \in (3,5,6)$
 $density \in (0.2,0.5,0.7)$
 $window \in (3,5,10)$ $overlap \in (1,3,5)$
 $randomization \in (0.01,0.1,0.3)$
 $rate \in (0.01,0.3,0.5)$

Итого было произведено $3^6 = 729$ обучений нейронной сети. Как было сказано ранее, использовалась база для инициализации генератора псевдослучайных чисел. Это позволяет гарантировать, что полученные результаты зависят только от изменения перечисленных выше параметров.

Первые же результаты экспериментов однозначно показали, что размер окна и перекрытия окон связаны и не имеет смысла разделять эти параметры. Было определено, что оптимальным перекрытием будет 50-60% от размера окна. В первом приближении было решено взять заведомо завышенные параметры нейронной сети:

$layer = 6$
 $density = 0.7$
 $window = 10$
 $overlap = 5$
 $randomization \in = 0.3$
 $rate = 0.5$

На рис. 3 изображен график значения ошибки от количества эпох обучения.

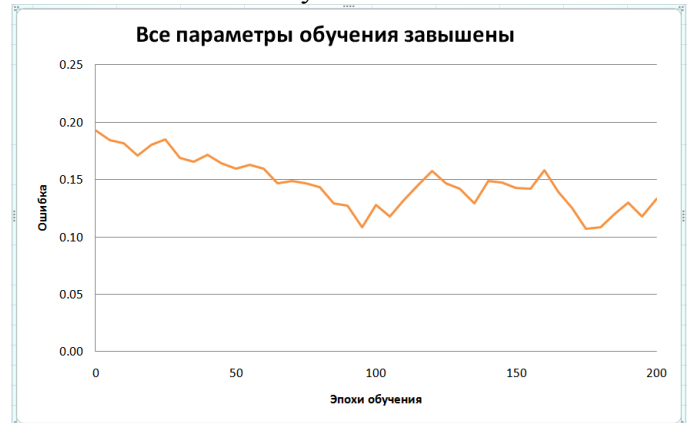


Рис. 3. Динамика обучения СНС с завышенными параметрами обучения

Видно, что сеть совершенно не поддается обучению. За 200 эпох качество распознавания осталось практически на начальном уровне. Более того, после 100 эпох сеть неожиданно пошла в разнос.

Путем подбора количества слоев удалось достичь лучших результатов (рис. 4):

$layer = 5$
 $density = 0.7$
 $window = 10$
 $overlap = 5$
 $randomization \in = 0.3$
 $rate = 0.5$

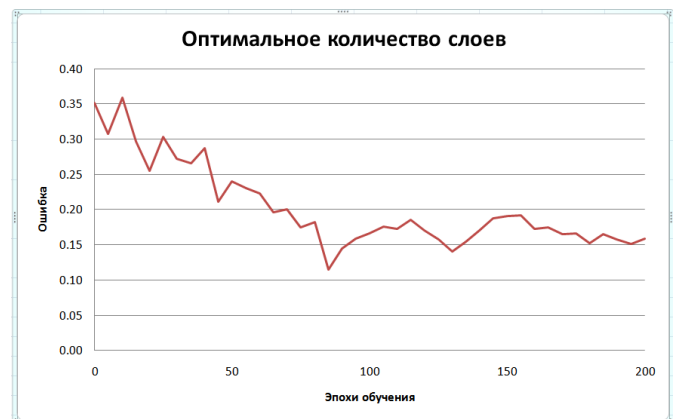


Рис. 4. Динамика обучения СНС с оптимальным количеством слоев

Зафиксировав количество слоев на 5, удалось получить хоть и плохой, но более-менее стабильный результат. Сеть однозначно научилась выделять определенные признаки, но все еще не достигает желаемого качества распознавания за при-

емлемое время. Следующим шагом было определение оптимального размера окна и перекрытия окон:

layer = 5
density = 0.7
window = 5
overlap = 3
randomization $\epsilon = 0.3$
rate = 0.5

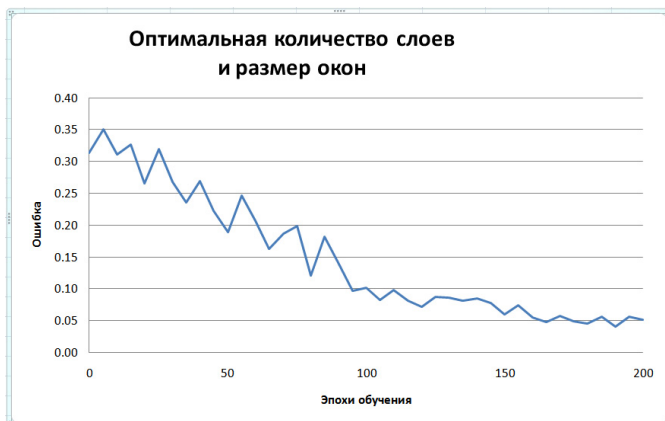


Рис. 5. Динамика обучения СНС с оптимальным количеством слоев и оптимальным размером окна

С такими параметрами нейронная сеть смогла достигнуть неплохих результатов, но все еще заметны скачки (рис. 5). Следующий шаг – оптимизация плотности связей:

layer = 5
density = 0.5
window = 5
overlap = 3
randomization $\epsilon = 0.3$
rate = 0.1

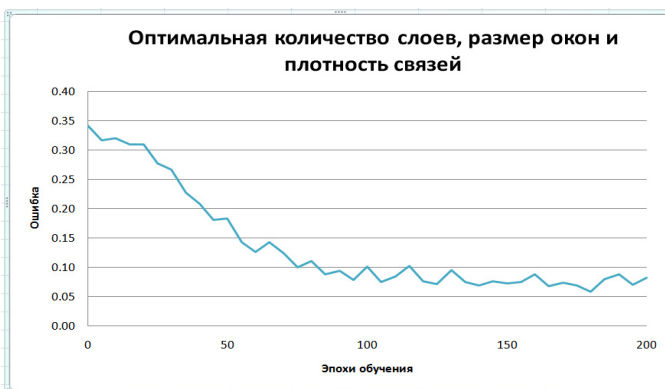


Рис. 6. Динамика обучения СНС с оптимальным количеством слоев, размером окна и плотностью связей

Теперь сеть ведет себя более предсказуемо, всплески намного уменьшились, но необходимое качество все еще не достигнуто (рис. 6).

Оптимизировав начальную инициализацию весов, удалось уменьшить скорость достижения хорошего распознавания (рис. 7):

layer = 5
density = 0.5
window = 5
overlap = 3
randomization $\epsilon = 0.1$
rate = 0.1

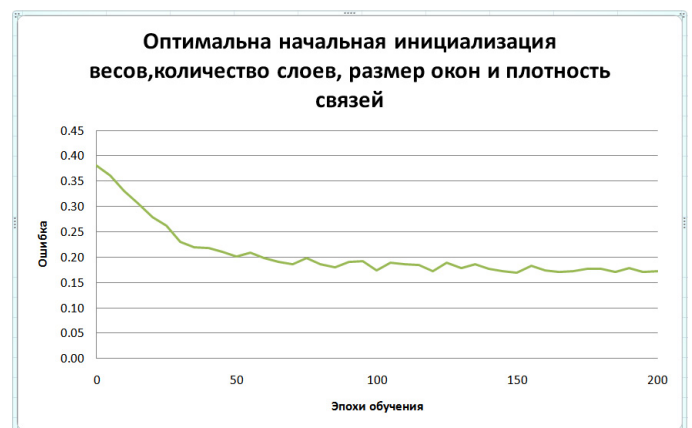


Рис. 7. Динамика обучения СНС с оптимальным количеством слоев, размером окна, плотностью связей и инициализацией весов

Кривая ошибки сети достаточно быстро выходит на прямой участок, что свидетельствует о быстром нахождении оптимальных признаков. Но как видно из рис. 7 скорость обучения недостаточна.

Увеличим скорость обучения:

layer = 5
density = 0.5
window = 5
overlap = 3
randomization $\epsilon = 0.1$
rate = 0.3

Как видно из рис. 8, сеть менее чем за 50 эпох выходит на допустимое качество, а за 200 эпох обучения ошибка уменьшается до 0.01-0.02.

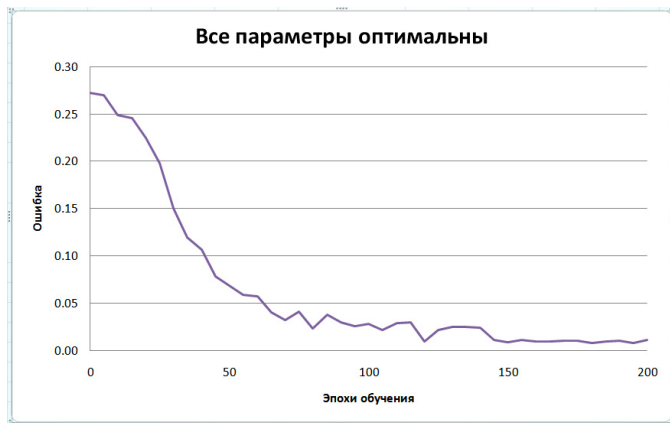


Рис. 8. Динамика обучения СНС с оптимальными параметрами обучения

Выводы

В ходе исследования был проведен ряд экспериментов для определения наиболее оптимальных параметров обучения свёрточной нейронной сети.

Анализ результатов показал, что наиболее оптимальными будут такие значения параметров:

layer = 5
density = 0.5
window = 5
overlap = 3
randomization \in = 0.1
rate = 0.3

Количество слоев влияет на качество распознавания образов. Чем больше слоев – тем больше связей и тем дольше проходит обучение, что не приемлемо.

Плотность связей в значительной мере влияет на возможность внесения в нейронную сеть необходимого количества признаков. Слишком слабосвязная сеть не может запомнить необходимые признаки. Слишком сильносвязная – наоборот пытается запомнить слишком много признаков и часто выдает ошибки, останавливаясь в области локального минимума.

Как было сказано ранее, размер окна подвыборки и области перекрытия должны соответствовать друг другу. Слишком большое перекрытие окон вносить много лишней информации в сеть. Для картинок из базы ORL Face (95x95 пикселей), оптимальный размер окна 5x5 пикселей с перекрытием в 3 пикселя со всех сторон.

Начальная инициализация позволяет быстро достигнуть стабильного распознавания.

Большая скорость обучения позволяет ускорить процесс обучения, но также, из-за больших значений сеть может пойти в разнос. Оптимальным значением будет 0.3.

Список литературы

1. LeCun Y. A theoretical framework for backpropagation // Proc. of IEEE. – 1998. – P.21-28.
2. LeCun Y., Bottou L., Bengio Y., Haffne P. Gradient-Based Learning Applied to Document Recognition // Proc. IEEE. – 1998. – P.59-67.
3. Дорогой Я.Ю., Яшин В.Е. Программный комплекс для симуляции многопоточных нейронных сетей // Вісник НТУУ “КПІ”, “Інформатика, управління та обчислювальна техніка”. – 2008. – №49. – С.123-127.
4. Дорогий Я.Ю., Яшин В. Є., Яцук С. В. Застосування багатопотокового симулятора нейронних мереж до задачі розпізнавання облич // Тези 5-ї Міжнародної науково-технічної конференції «Інформаційно-комп’ютерні технології 2010». – Житомир: ЖДТУ, 20-22 травня 2010 року. – С.53-55.