

## АРХИТЕКТУРА ОБОБЩЕННЫХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

В статье рассмотрена архитектура обобщенных сверточных нейронных сетей, позволяющих использовать преимущества классических сверточных нейронных сетей с дополнительными возможностями нового класса в задачах распознавания человека по фотопортрету.

The structure of the generalized convolutional neural networks which allow advantages of classical convolutional neural networks to be used with capabilities of new network class for problem of human face recognition was described in this article.

### Введение в проблему

Для распознавания лиц людей широко используется архитектура нейронной сети, которая получила название сверточной нейронной сети (СНС). Данная нейронная сеть была впервые описана в [1,2] и сейчас успешно применяется для решения широкого класса задач связанных с распознаванием паттернов, таких как распознавание трехмерных объектов, предсказание погоды, автоматическое управление и др.

Изначально архитектура сверточной нейронной сети разрабатывалась с учетом особенностей строения некоторых участков мозга человека, ответственных за зрение.

В данной статье предлагается ряд улучшений в архитектуре обычной сверточной нейронной сети, которые позволят повысить эффективность работы системы распознавания, дадут возможность применять сверточные нейронные сети к более широкому классу задач.

Статья является продолжением исследований сверточных нейронных сетей, впервые описанных в [3].

### 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 [3]. В ходе подробного исследования был выявлен ряд недостатков СНС:

1. Высокая сложность архитектуры.
2. Полносвязность.
3. Фиксированная площадь окна слоя свертки.

### Цель работы

Целью работы является разработка новой архитектуры сверточной нейронной сети, которая позволит устранить недостатки, описанные выше, а также, по возможности, улучшить точность распознавания и скорость обучения сети.

### Архитектура классической сверточной нейронной сети

Рассмотрим архитектуру обычной сверточной нейронной сети (рис. 1.), описанной в [3].

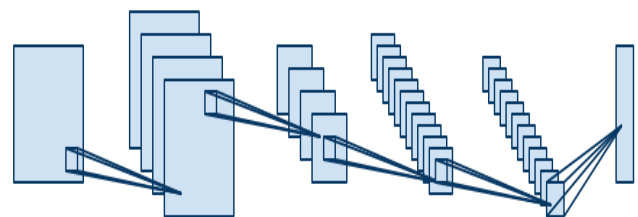


Рис. 1. Структура СНС

Нейронная сеть состоит из пар слоев - слоев подвыборки и слоев свертки, каждый из которых в свою очередь состоит из карт признаков.

Нетрудно убедиться в том, что каждая карта признаков в идеале фильтрует изображение, находя какой-то один определенный, специфичный для данной карты, признак. Например, первая карта признаков научена искать кружочки, вторая – квадратики и т.д.

Исходное изображение подается на входной слой. В первом слое подвыборки каждая карта признаков осуществляет поиск определенного, закрепленного только за данной картой, признака. Достигается это за счет использования общей для всей карты признаков матрицы весов и особой организацией локального рецептивного поля для каждого нейрона такой карты.

Каждый нейрон карты признаков получает входные данные от прямоугольной области размера  $n \times m$  входного изображения. Такая область достаточно мала и множество таких областей на входном изображении пересекаются и накладываются по принципу черепицы.

Смежные нейроны карты признаков получают в качестве входного воздействия смежные прямоугольные области, причем весовые коэффициенты для всех нейронов карты признаков будут одинаковыми.

Для простоты изложения будем называть область, формирующую локальное рецептивное поле нейрона слоя подвыборки, окном. Соответственно, площадь окна – количеством нейронов в такой области.

Таким образом, можно говорить о том, что карта признаков в целом осуществляет операцию поиска признака во входных данных. Другие карты признаков имеют другой набор весовых коэффициентов и, соответственно, осуществляют поиск других признаков во входных данных.

Конкретные признаки, извлекаемые той или иной картой признаков, определяются в процессе обучения нейронной сети с учителем.

$n$  и  $m$  – достаточно малые числа, которые определяют разрешающую способность нейронной сети – минимальный размер признака, который данная сеть может регистрировать.

Для введения инвариантности нейронной сети к смещениям и небольшим деформациям, используется слой свертки. Для каждой карты признаков существует соответствующая ей карта свертки, которая уменьшает размерность карты признаков с  $n \times m$  до  $n/2 \times m/2$  путем усреднения значений по квадрату  $2 \times 2$  нейронов.

После выполнения свертки сеть теряет часть информации о точном положении найденного признака, но сохраняет информацию относительно взаимного расположения различных признаков.

Следующий слой подвыборки осуществляет аналогичную первому слою сегментацию вход-

ных данных на прямоугольные области  $n \times m$ , только входными данными второго слоя служит выход первого слоя. Т. о., каждая карта признаков второго слоя осуществляет поиск признаков второго порядка одновременно во всех картах признаков первого слоя.

Очевидно, что с ростом количества слоев уменьшается размерность каждой карты признаков, хотя, в целом, количество нейронов в слое сильно растет за счет использования большего количества карт признаков в верхних слоях сети.

Сверточной нейронной сети с тремя парами слоев подвыборки-свертки вполне достаточно для точного распознавания лиц людей [1].

Такая нейронная сеть хорошо себя зарекомендовала в задачах распознавания, но ее использование в некоторых случаях достаточно проблематично.

Одной из проблем классической сверточной нейронной сети является подбор оптимального значения размера локального рецептивного поля (окна) нейрона в слое подвыборки. Малые значения  $n$  и  $m$  позволяют повысить разрешающую способность сети и дают возможность находить довольно малые признаки, но в тоже время аналогичный признак большего масштаба будет пропущен и принят за совокупность других признаков. Таким образом, классическая сверточная нейронная сеть плохо работает с изображениями, на которых могут присутствовать одинаковые признаки разного масштаба (например, на ненормализованных по масштабу изображениях)

### Архитектура обобщенной сверточной нейронной сети

Допустим, на картинке есть как маленькие кружочки, так и большие, и необходимо найти максимум кружочков в первом же слое подвыборки.

Меняя площадь локального рецептивного поля каждого нейрона в каждой карте признаков можно добиться нахождения кружочков разного размера за один раз, но тогда кусочки других размеров останутся незамеченными.

В первом слое подвыборки сделаем несколько карт признаков с одним размером рецептивного поля, несколько – с чуть большим размером поля, несколько – с еще большим.

Для решения этой проблемы предлагается использовать в одном слое карты признаков хо-

тя и одного размера, но с разными размерами окна для разных карт признаков.

Например, вместо 10 карт признаков с размером окна каждого нейрона  $3 \times 3$  предлагается использовать 5 карт признаков с размером поля  $3 \times 3$ , 3 карты с полем  $5 \times 5$  и 2 карты признаков с размером окна  $7 \times 7$ .

Такая конфигурация позволит находить признаки разного размера одновременно сразу в первом слое подвыборки, что должно повысить общее качество распознавания объектов, ускорить обучение нейронной сети и способствовать уменьшению количества карт признаков в высших слоях нейронной сети, что приведет к уменьшению количества связей нейронной сети, и как следствие – уменьшение потребления памяти и ускорение обучения.

При такой организации первого слоя подвыборки закономерно возникает проблема с пересечением окон смежных нейронов одной карты признаков – для получения одинакового размера карт признаков с разными размерами окон необходимо разместить одинаковое количество таких полей на входном слое, что ведет к увеличению площади пересечения смежных полей с ростом размера поля.

К сожалению, бороться с таким резким увеличением количества связей крайне проблематично, но следующее улучшение позволит уменьшить общее количество связей в сверточных нейронных сетях, что особенно актуально для описанной в данной статье архитектуры сверточной нейронной сети.

Рассмотрим подробно один нейрон из карты признаков первого слоя подвыборки. Такой нейрон получает информацию из прямоугольной области входного изображения, формирующей его локальное рецептивное поле (окно). Общее количество входящих связей для нейрона равно  $n \times m$ , где  $(n; m)$  – размер окна.

С увеличением  $n$  и  $m$  пропорционально растет количество связей, но в тоже время вклад, вносимый одной связью в суммарный вход нейрона уменьшается.

Это означает, что при большом количестве связей можно без значительного ущерба для качества работы сети переходить к разреженной форме соединений, удаляя некоторую малую их часть случайным образом.

Например, для окна  $5 \times 5$  нейронов можно удалить 5 соединений, незначительно уменьшив точность, но при этом на 20% сократив общее количество соединений.

В отличие от соединений, общее количество весов за счет использования методики разделяемых весов, растет незначительно и нет смысла их сокращать.

Для данной карты признаков, применяя разрежение матрицы связей входного слоя с картой признаков, можно получить значительное сокращение связей, уменьшив, тем самым, расход памяти и ускорив процесс обучения сети.

Следует отметить, что в случае разрежения матрицы связей каждый нейрон карты признаков получает индивидуальную, отличную от соседних нейронов в данной карте признаков, матрицу связей.

Применение вышеизложенных рекомендаций по улучшению архитектуры сверточной нейронной сети ставит задачу оптимального распределения карт признаков с разным размером поля, а также задачу выбора количества отсекаемых связей в случае разрежения матриц связей.

К сожалению, количественные параметры сверточной нейронной сети должны подбираться строго под поставленную задачу после проведения ряда экспериментов для доказательства допустимости той или иной конфигурации.

Следует также отметить, что приведенные выше выкладки, несмотря на свою ориентированность на задачу распознавания образов, могут быть вполне использованы для построения сверточных нейронных сетей, решающих иные сходные задачи, к числу которых относятся прогнозирование и управление.

## Эксперименты и результаты

Для проверки данных гипотез был проведен ряд экспериментов со множеством сверточных нейронных сетей. В нейронные сети вносились предложенные улучшения, осуществлялся анализ степени влияния той или иной модификации на качество распознавания, скорость обучения и пр. Несмотря на то, что в конечном итоге предложенные улучшения оправдали лишь некоторые из возложенных на них надежд, тем не менее, они позволяют значительно повысить качество распознавания объектов, в т. ч. и лиц людей. Все эксперименты проводились с использованием набора программного обеспечения PANN [4,5].

Для исследования нейронные сети тренировались на фотографиях 5-и человек из базы лиц ORL Faces по 10 фотографий на каждого. Из

имеющихся 50 фотографий 60% были использованы для непосредственного обучения, в то время как остальные – для промежуточного тестирования нейронных сетей на ранее не виденных образах.

Ниже приведены некоторые полученные зависимости для конфигурации классических сверточных нейронных сетей (рис. 2-8):

Зависимость точности распознавания от плотности связей



Рис. 2. Зависимость точности распознавания от плотности связей

Зависимость точности распознавания от размеров окна в первом слое

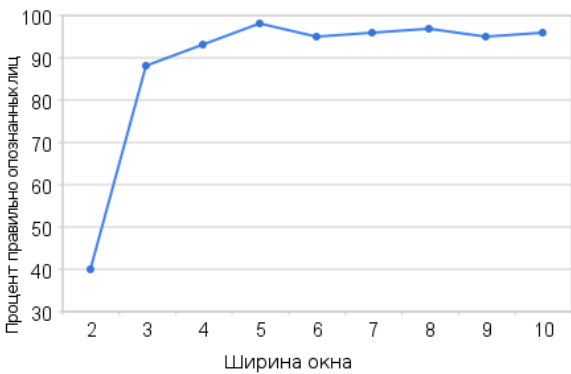


Рис. 3. Зависимость точности распознавания от размеров окна в первом слое

Зависимость ошибки в конце обучения от плотности связей

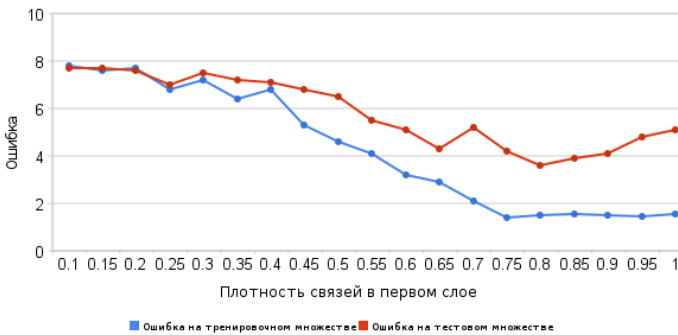


Рис. 4. Зависимость ошибки распознавания от плотности связей

Зависимость установившейся ошибки от размера окна

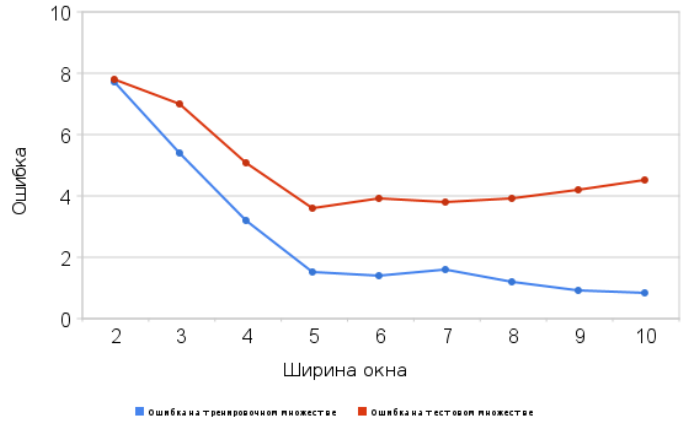


Рис. 5. Зависимость установившейся ошибки от размера окна

Зависимость количества эпох требуемых для полного обучения от плотности связей

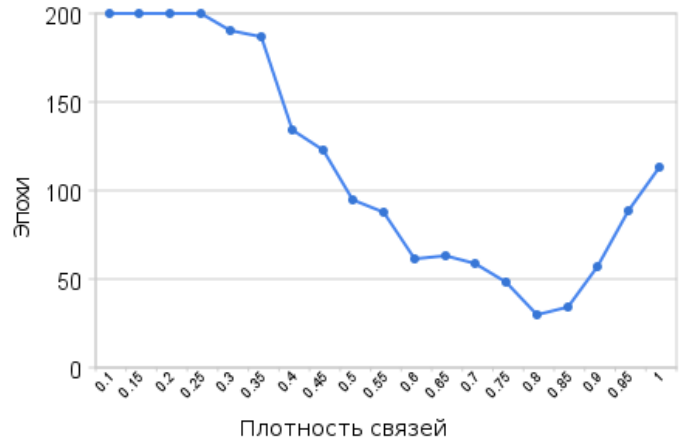


Рис. 6. Зависимость скорости обучения от плотности связей

Зависимость скорости обучения от размера окна



Рис. 7. Зависимость скорости обучения от размера окна



**Рис. 8. Зависимость скорости обучения от параметра скорости обучения**

Очевидно, что чем больше плотность сети, тем лучше она учится. Но если плотность слишком большая (>0.8), то обучение замедляется.

**Сравнение классической и обобщенной сверточных сетей**

Далее была проведена серия экспериментов над топологиями сверточных нейронных сетей с внесенными изменениями, предложенными выше. Ниже приведены графики процессов обучения и тестирования двух различных нейронных сетей с различной конфигурацией.

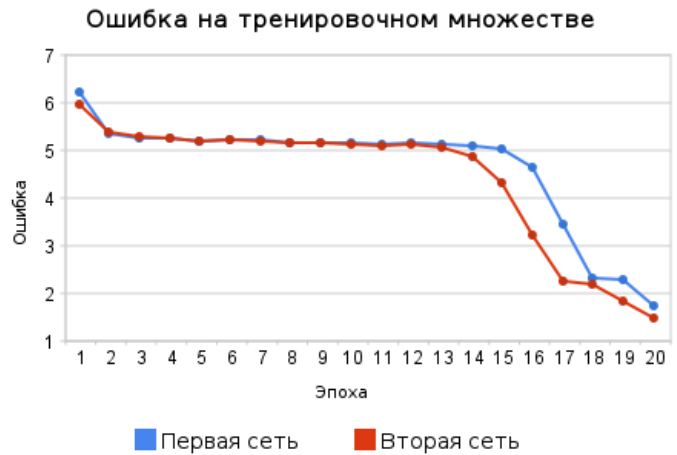
Сеть №1 – это классическая сверточная нейронная сеть с размером окна 5×5 и 7-ю картами признаков в первом слое подвыборки и плотностью связей 80%, что дает в сумме 5600 связей в первом слое нейронной сети.

Сеть №2 – обобщенная сверточная нейронная сеть с двумя полносвязными картами признаков размером 3×3, двумя картами признаков 5×5 и плотностью связей 80%, а также тремя картами признаков размером 7×7 и плотностью связей 50%. В результате количество связей в первом слое сократилось до 5320, в то время как в первом слое были использованы карты признаков разного размера – от 3×3 до 7×7, что позволяет одновременно обрабатывать первичные признаки различного размера.

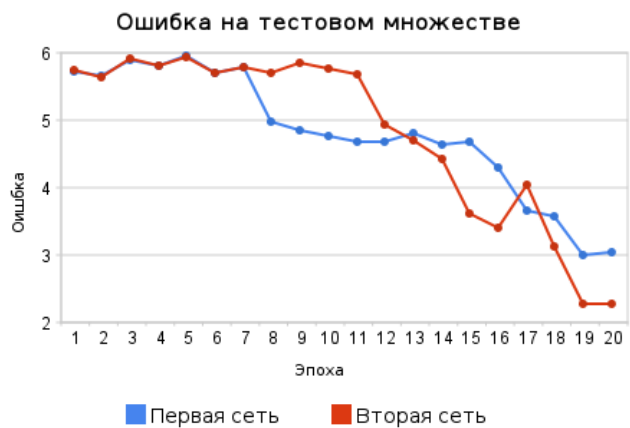
Результаты экспериментов представлены на рис. 9-10.

Для экспериментов были взяты 50 фотографий 5 людей – 30 шт. для обучения, 20 – для тестов. В результате экспериментов получены следующие данные:

- Сеть 1 – из 20 фотографий 2 неправильно опознаны;



**Рис. 9. Ошибка на тренировочном множестве**



**Рис. 10. Ошибка на тестовом множестве**

- Сеть 2 – из 20 фото 0 неправильно опознанных.

После серии тестов оказалось, что для классической сверточной сети точность распознавания составляет порядка 97-98%, а для обобщенной сверточной нейронной сети – 99-99,5%.

**Выводы**

Эти и другие эксперименты над сверточными нейронными сетями подтвердили превосходство обобщенной топологии над классической. Использование нескольких карт признаков разного размера одновременно позволяет получить лучшее обобщение для обученной нейронной сети, т.е. снизить процент ложных срабатываний для образов, не входивших во множество обучения. Благодаря снижению плотности связей для больших карт признаков удастся вносить эти изменения в архитектуру сети без роста количества связей, что благотворно сказывается на потреблении памяти, скорости обучения.

К сожалению, предложенные улучшения не дают видимого выигрыша в скорости обучения

сверточной нейронной сети, что дает почву для новых исследований в этой области.

### 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. Применение компактных ячеистых сверточных нейронных сетей для биометрической идентификации человека по лицу // Вісник НТУУ “КПІ”, “Інформатика, управління та обчислювальна техніка”. – 2007. – №46. – С.135-149.
4. 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.
5. 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.