

Машинное обучение

Лекция 4.

Глубокие нейронные сети

Часть 1.

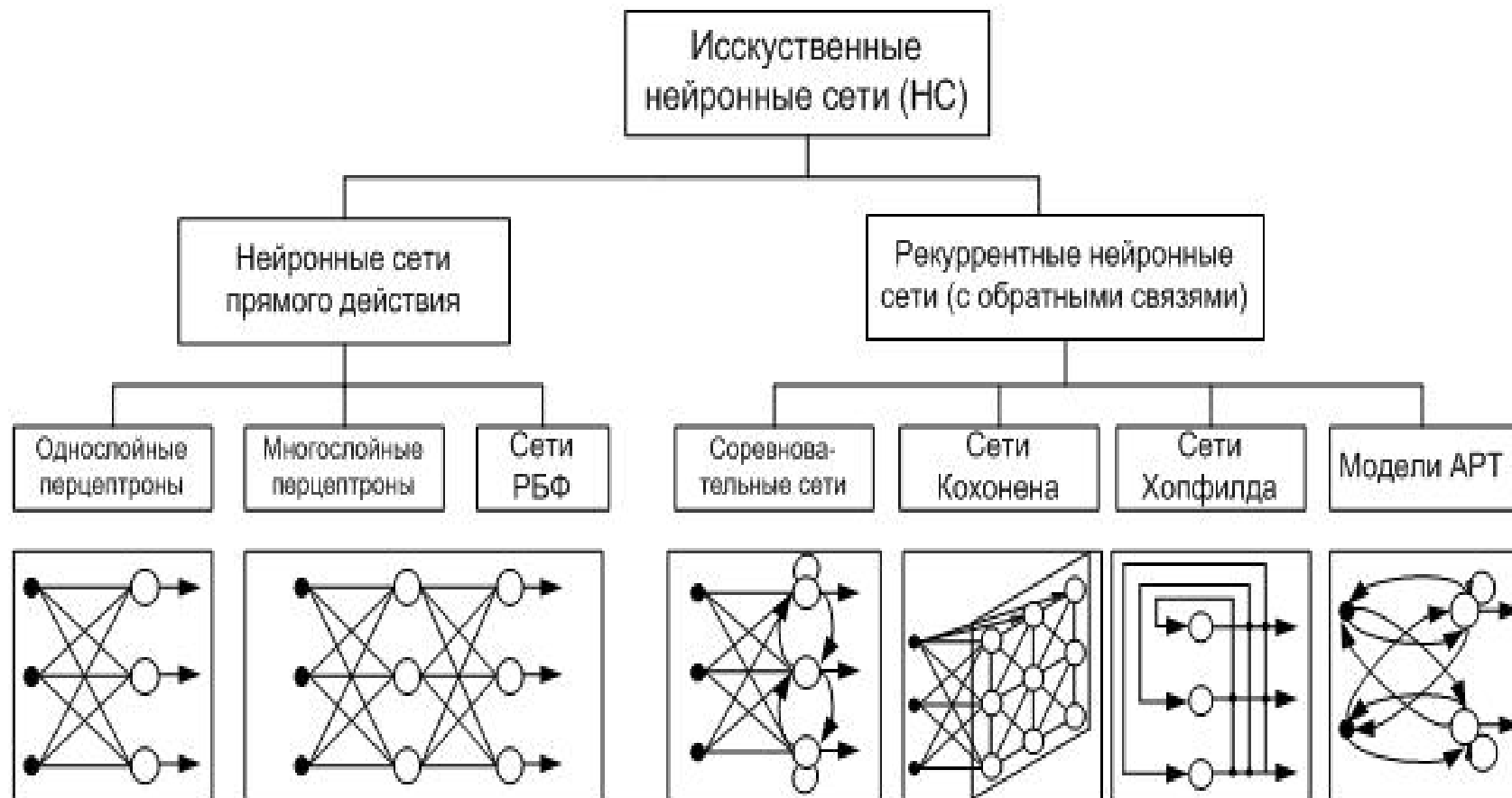
Предпосылки бума ИИ в начале 2000-х

- Бурное развитие и применение роботов (тенденции их использования за пределами промышленного производства и исследовательских лабораторий)
- В связи с этим появление в робототехнике концепции «Когнитивная робототехника»
- Появление необходимости в средствах визуального наблюдения и распознавания (в технологиях “Ambient Intelligence”, в системах безопасности, в военных применениях)
- Появление новых результатов нейрофизиологии о строении и функционировании мозга

Кризис ИИ к началу 2000-х

- Множество разных парадигм в ИИ со своим ограничениями по использованию
 - Методы представления знаний и решения задач с их использованием,
 - Разные коннекционистские модели (вероятностные и нейросетевые) и методы обучения
 - Методы их гибридизации
- Создание интеллектуальной системы сводится к выбору парадигм, может быть, к их комбинации (гибридизации)
- Получается ИС с ограничениями (по способности обучаться, по типу входных данных, по типу решаемых задач, по масштабируемости, по кооперации между собой и с человеком и т.п.)
- Необходимость универсальных методов построения обучаемых (обучающихся) интеллектуальных систем без существенных ограничений, по аналогии с естественным разумом

Классификация классических нейронных сетей



Название нейропарадигмы	Авторы	Год разработки	Области применения
Однослойный перцептрон (Single layer perceptron)	R. Rosenblatt	1959	Распознавание образов, классификация/категоризация
Обратное распространение ошибки (Back Propagation)	R. Rosenblatt, M. Minsky, S. Papert	1960-е	Распознавание образов, классификация, прогнозирование
Встречное распространение (Counter Propagation)	R. Hecht-Nielsen	1986	Распознавание образов, восстановление образов (ассоциативная память), сжатие данных
Входная звезда (Instar Network)	S. Grossberg	1974	Распознавание образов
Выходная звезда (Outstar Network)	S. Grossberg	1974	Распознавание образов
Искусственный резонанс-1 (ART-1 Network)	S. Grossberg, G. Carpenter	1986	Распознавание образов, кластерный анализ
Сеть Хопфилда (Hopfield Network)	J.J. Hopfield	1982	Поиск и восстановление данных по их фрагментам
Сеть Хэмминга (Hamming Network)	R. W. Hamming	1987	Распознавание образов, классификация, ассоциативная память, надежная передача сигналов в условиях помех
Сеть Коконена (Kohonen Network)	T. Kohonen	1984	Кластерный анализ, распознавание образов, классификация
Сеть поиска максимума (MAXNET)	R.P. Lippman	1987	Совместно с сетью Хемминга, в составе НС систем распознавания образов
Сеть поиска максимума с прямыми связями (Feed-Forward MAXNET)	R.P. Lippman	1987	Совместно с сетью Хемминга, в составе НС систем распознавания образов
Двунаправленная автоассоциативная память (BAM Network)	B. Kosko	Вторая половина 80-х	Ассоциативная память, распознавание образов
Обучение Больцмана (Boltzman machine)	J. Hinton, T. Sejnowsky, H. Szu	1985	Распознавание изображений, сигналов радара, сонара
Классификатор Гаусса (Neural Gaussian Classifier)	R.P. Lippman	1987	Распознавание образов, классификация
Генетические алгоритмы (Genetic training algorithm)	J. Holland, D. Goldberg	1975 1988	Обучение НС, распознавание сигналов сонаров

Различные подходы к обучению НС

1. Обучение основанное на коррекции ошибки
 - Используется ошибка (d-y) между желаемым и реальным выходом для изменения весов связей с целью уменьшения ошибки
2. Вероятностное обучение (машина Больцмана)
 - Адаптация весов видимых нейронов так, чтобы состояния видимых нейронов соответствовали желаемому распределению вероятностей
3. Правило Хебба – усиливаются связи между нейронами с похожими (или одинаковыми) состояниями и наоборот

$$w_{ij}(t + 1) = w_{ij}(t) + \eta y_j(t) x_i(t)$$

4. Соревновательное обучение
 - Только один нейрон (победитель) в выходном слое активизируется (*winner-take-all*, «победитель получает все»)
5. Обучение применением генетических алгоритмов

Попытки выхода из кризиса

- Кризис систем, основанных на знаниях
 - Появилась концепция AGI (Artificial General Intelligence),
- Кризис нейронных сетей
 - Появились концепции deep learning (глубокое или глубинное обучение) и neuromorphic computing (нейроморфные вычисления или нейроморфные сети)

Что такое “deep learning”?

- Family of methods that uses deep architecture to learn high level feature representations (из курса видео-лекций “Deep Learning”)
- Branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures or otherwise, composed of multiple non-linear transformations (из Wikipedia)

Что такое “deep learning”? (2)

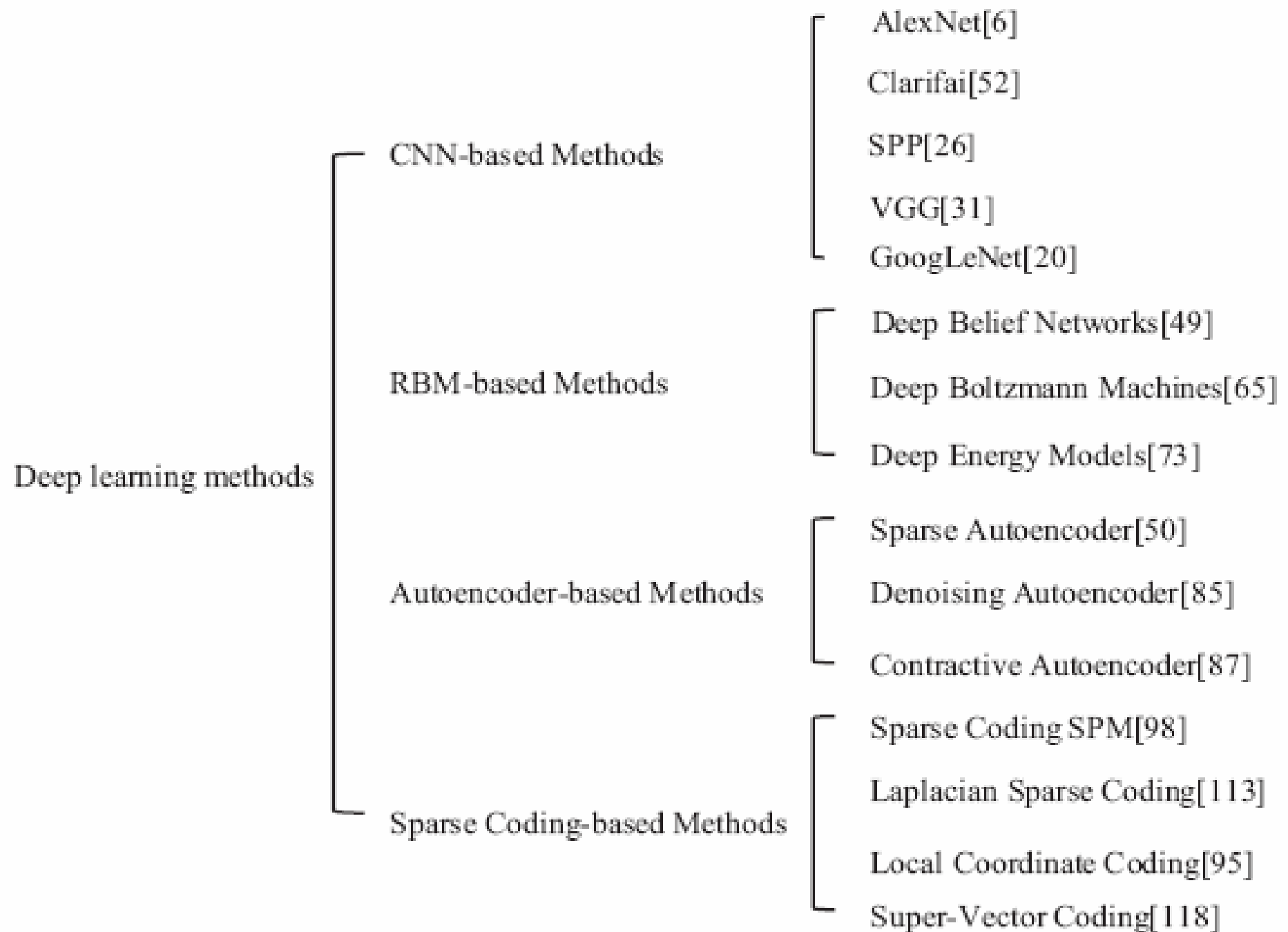
- **Глубокое обучение** — набор алгоритмов машинного обучения, которые пытаются моделировать высокоуровневые абстракции в данных, используя архитектуры, состоящие из множества нелинейных трансформаций.
- Под термином «глубина» в данном случае понимается глубина графа вычислений модели — максимальная длина между входным и выходным узлами конкретной архитектуры. В случае, например, простой нейронной сети прямого распространения глубина соответствует количеству слоев сети.
- Термин 'глубокое обучение' акцентирует внимание на сложности обучения внутренних (глубоких) слоев многослойной сети, которые плохо поддаются классическим методам обучения, таким как метод обратного распространения ошибки.
- **ИЗ ВИКИПЕДИИ**

Другое возможное определение

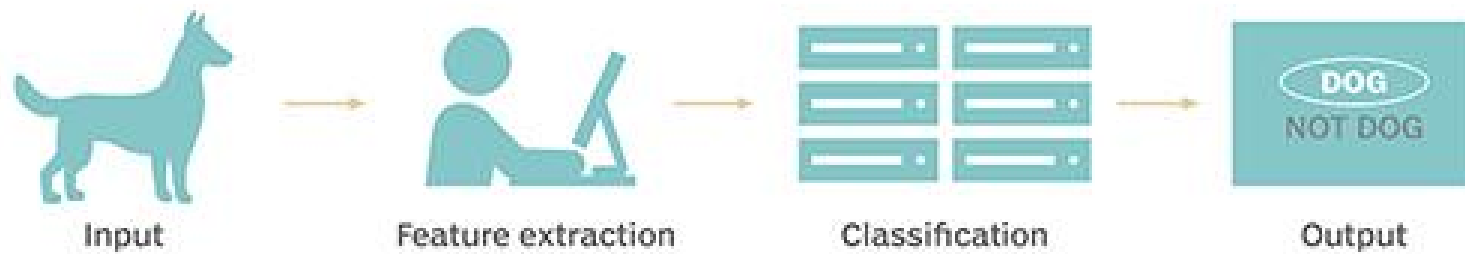
- Множество моделей обработки информации в мозге человека/животного/ИИ, поступающей от рецепторов/сенсоров, целью которой является обучение распознаванию объектов (понятий) в виде процесса иерархического абстрагирования их представления.
- Обычно подразумевается обработка зрительной информации и естественного языка. Хотя могут быть и другие источники информации

Одна из возможных классификаций НС с глубоким обучением

- deep neural networks (глубокие нейронные сети)
 - Многослойная НС прямого распространения с обработкой разного уровня признаков в каждом уровне и обучением каждого слоя отдельно
- convolutional deep neural networks (сверточные сети)
 - Многослойная НС с разделением на подсети (с локальными связями) для обработки разных признаков и объединением их на верхнем уровне
- LSTM и GRU-сети – рекуррентные сети, способные при обучении учитывать далекие во времени данные из последовательности
- deep belief networks (глубокие сети доверия)
 - Вероятностная многослойная сеть
- Трансформеры (transformers)



TRADITIONAL MACHINE LEARNING



DEEP LEARNING



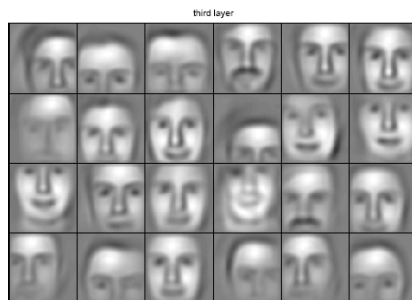
Применяется в областях: (Wikipedia)

- computer vision
- automatic speech recognition
- natural language processing
- audio recognition
- bioinformatics

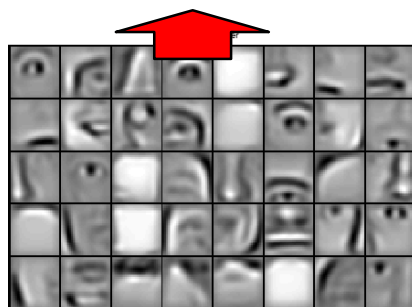
Из истории (Из Википедии)

- Глубокие архитектуры, основанные на искусственных нейронных сетях, берут свою начало с неокогнитрона, разработанного Кунихико *Fukushima K.* в 1980 году. В 1989 году Яну Лекуну удалось использовать алгоритм обратного распространения ошибки для обучения глубоких нейросетей (многоуровневых) для решения задачи распознавания рукописных ZIP-кодов. Несмотря на успешный опыт, для обучения модели потребовалось три дня, что существенно ограничивало применение этого метода.
- Низкая скорость обучения связана со многими факторами, включая проблему исчезающего градиента, которую в 1991 году анализировали Юрген Шмидхубер и Сепп Хохрайтер. Из-за этих проблем нейронные сети в 1990-х годах уступили место методу опорных векторов.
- Термин «глубокое обучение» приобрёл популярность после публикации Джеффри Хинтона и Руслана Салахутдинова в средние 2000-х годов, в которой они показали, что можно эффективно предобучать многослойную нейронную сеть если обучать каждый слой отдельно при помощи ограниченной машины Больцмана, а затем дообучать при помощи метода обратного распространения ошибки

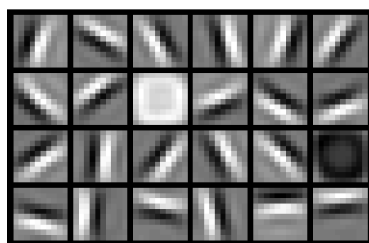
Уровни обработки изображения



object models



object parts
(combination
of edges)



edges



pixels

Когнитрон

- Искусственная нейронная сеть на основе принципа самоорганизации.
- Своей архитектурой когнитрон похож на строение зрительной коры, имеет иерархическую многослойную организацию, в которой нейроны между слоями связаны только локально.
- Обучается конкурентным обучением (без учителя).
- Каждый слой мозга реализует различные уровни обобщения; входной слой чувствителен к простым образам, таким, как линии, и их ориентации в определенных областях визуальной области, в то время как реакция других слоев является более сложной, абстрактной и независимой от позиции образа.
- Аналогичные функции реализованы в когнитроне путем моделирования организации зрительной коры₁₇

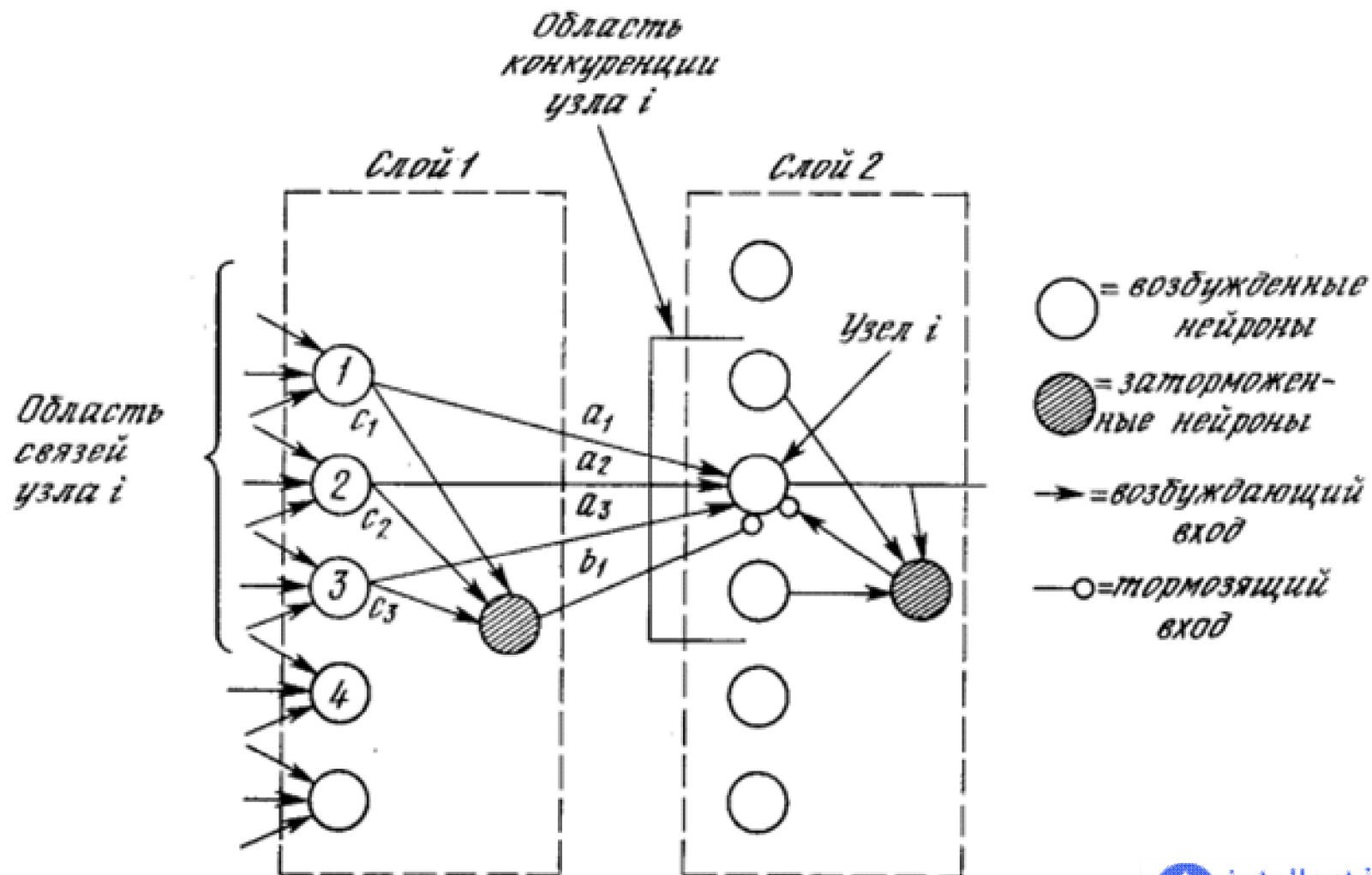
Архитектура когнитрона

- Когнитрон состоит из иерархически связанных слоев нейронов двух типов — тормозящих и возбуждающих. Состояние возбуждения каждого нейрона определяется соотношением его тормозящих и возбуждающих входов.
- Относительно некоторой синаптической связи между слоями 1 и 2 соответствующий нейрон слоя 1 является пресинаптическим, а нейрон второго слоя — постсинаптическим.
- Постсинаптические нейроны связаны не со всеми нейронами 1-го слоя, а лишь с теми, которые принадлежат их локальной области связей.
- Области связей близких друг к другу постсинаптических нейронов перекрываются, поэтому активность данного пресинаптического нейрона будет сказываться на все более расширяющейся области постсинаптических нейронов следующих слоев иерархии.

Неокогнитрон

- Когнитрон и неокогнитрон имеют определенное сходство, но между ними также существуют различия, связанные с эволюцией исследований авторов. Оба образца являются многоуровневыми иерархическими сетями, организованными аналогично зрительной коре.
- Неокогнитрон более соответствует модели зрительной системы, описанной в работе Hubel D. H. и Wiesel T. N. в 1962 году.
- В результате неокогнитрон является намного более мощной парадигмой с точки зрения способности распознавать образы независимо от их преобразований, вращений, искажений и изменений масштаба. Как и когнитрон, неокогнитрон использует самоорганизацию в процессе обучения

Возбуждающие и тормозящие нейроны в когнитроне



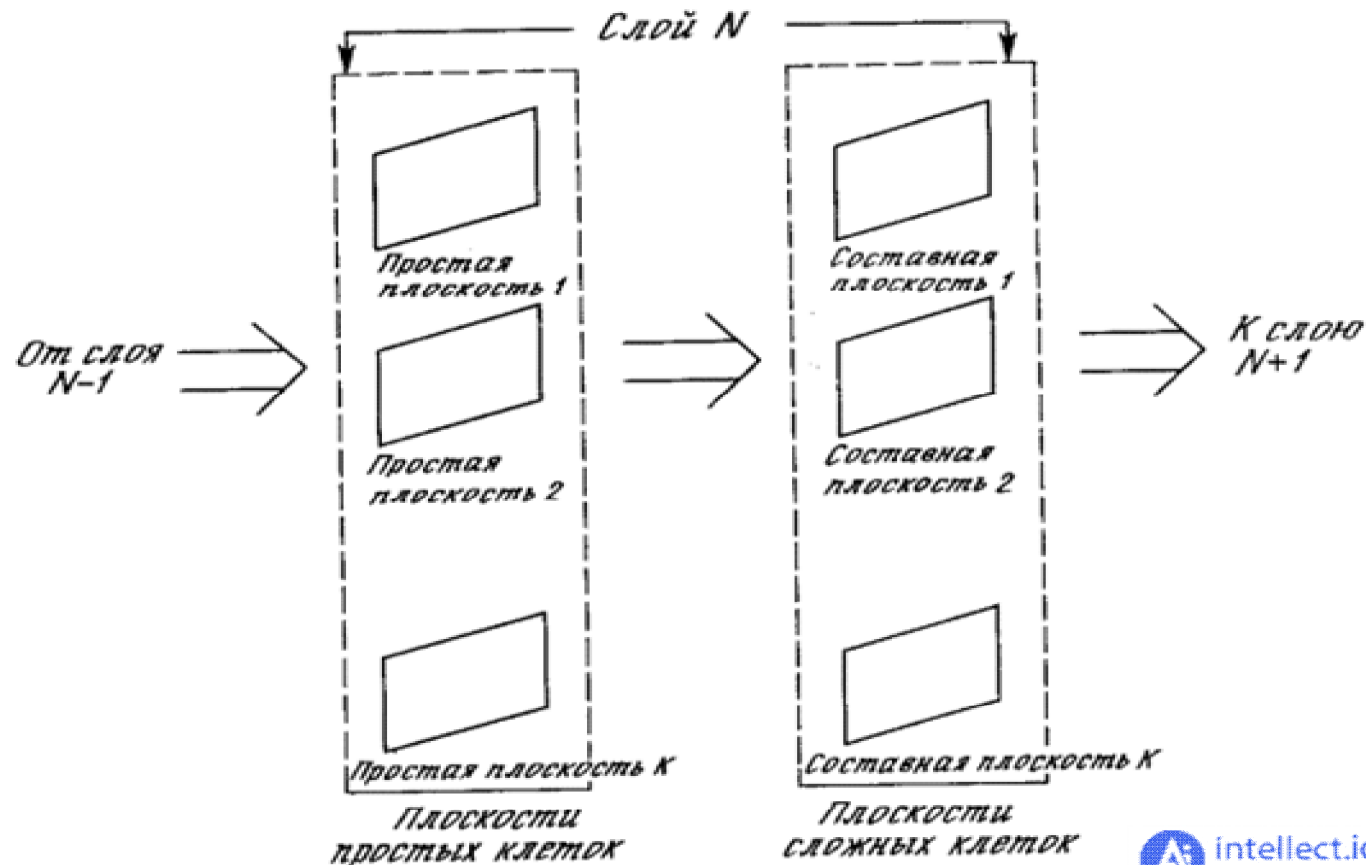
Неокогнитрон

- Когнитрон и неокогнитрон имеют определенное сходство, но между ними также существуют различия, связанные с эволюцией исследований авторов. Оба образца являются многоуровневыми иерархическими сетями, организованными аналогично зрительной коре.
- Неокогнитрон более соответствует модели зрительной системы, описанной в работе Hubel D. H. и Wiesel T. N. в 1962 году. В зрительной коре были обнаружены узлы, реагирующие на такие элементы, как линии и углы определенной ориентации.
- В результате неокогнитрон является намного более мощной парадигмой с точки зрения способности распознавать образы независимо от их преобразований, вращений, искажений и изменений масштаба. Как и когнитрон, неокогнитрон использует самоорганизацию в процессе обучения

Слой неокогнитрона

Каждый слой неокогнитрона состоит из двух массивов плоскостей.

Массив плоскостей, содержащих простые узлы, получает выходы предыдущего слоя, выделяет определенные образы и затем передает их в массив плоскостей, содержащих комплексные узлы, где они обрабатываются таким образом, чтобы сделать выделенные образы менее позиционно зависимыми



Рецептивные области узлов в каждой плоскости простых узлов перекрываются с целью покрытия всего входного образа этого слоя.

Каждый узел получает входы от соответствующих областей всех плоскостей комплексных узлов в предыдущем слое.

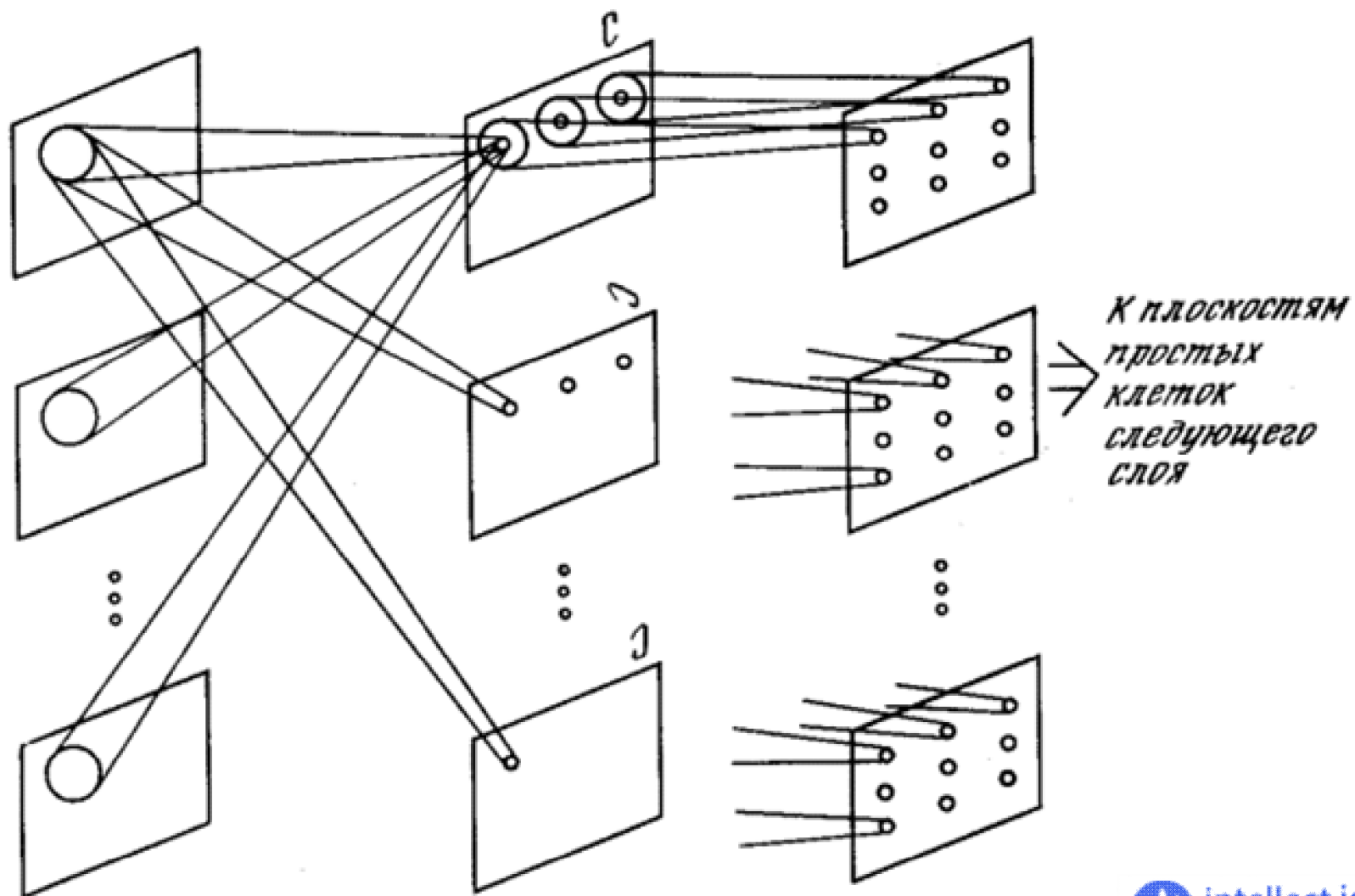
Следовательно, простой узел реагирует на появление своего образа в любой сложной плоскости предыдущего слоя, если он окажется внутри его рецептивной области

Слой $N-1$ → Слой N

Плоскости
сложных клеток

Плоскости
простых клеток

Плоскости
сложных клеток



Задачей комплексных узлов является уменьшение зависимости реакции системы от позиции образов во входном поле. Для достижения этого каждый комплексный узел получает в качестве входного образа выходы набора простых узлов из соответствующей плоскости того же слоя.

Эти простые узлы покрывают непрерывную область простой плоскости, называемую рецептивной областью комплексного узла. Возбуждение любого простого узла в этой области является достаточным для возбуждения данного комплексного узла. Таким образом, комплексный узел реагирует на тот же образ, что и простые узлы в соответствующей ему плоскости, но он менее чувствителен к позиции образа, чем любой из них.

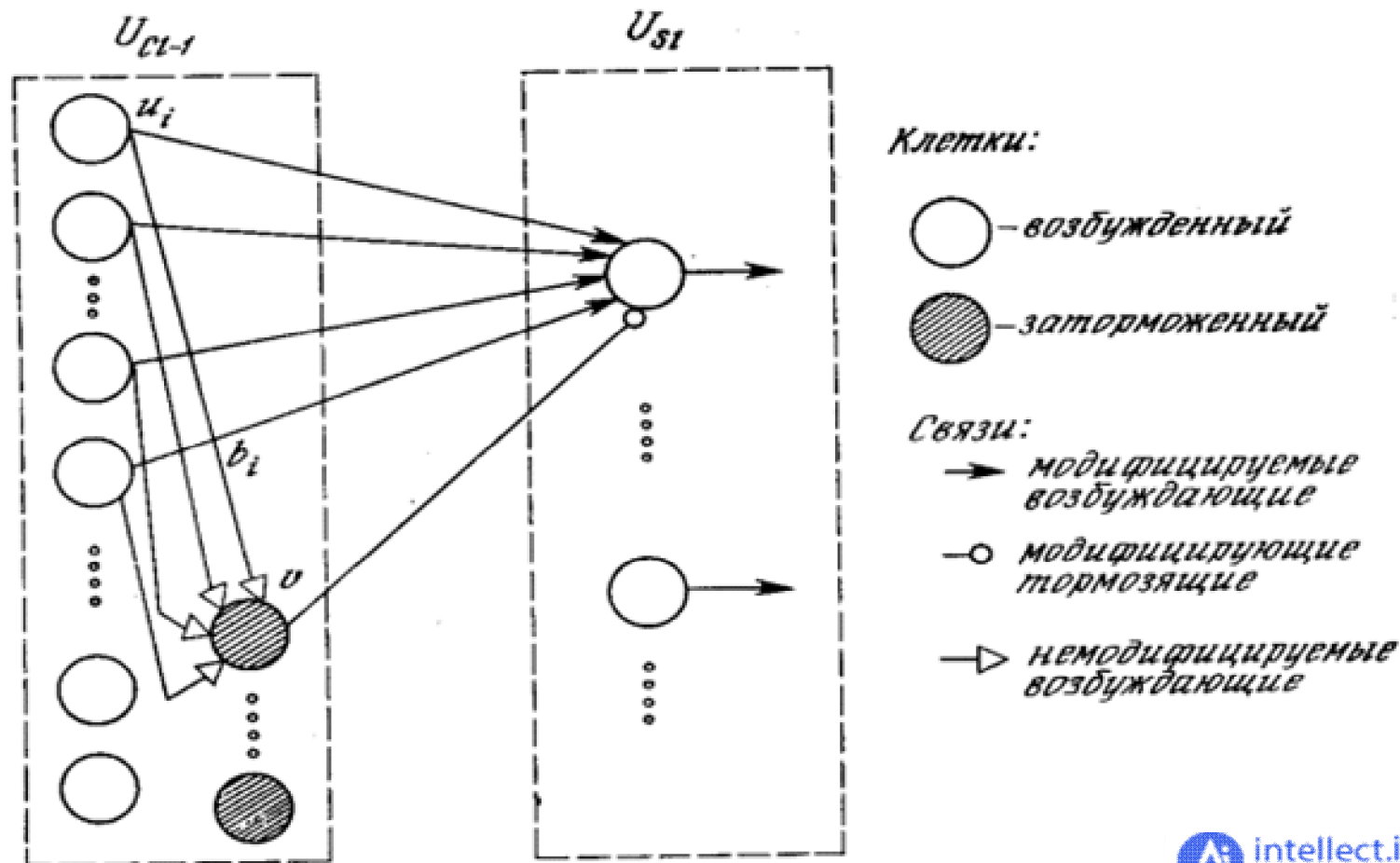
Таким образом, каждый слой комплексных узлов реагирует на более широкую область входного образа, чем это делалось в предшествующих слоях. Эта прогрессия возрастает линейно от слоя к слою, приводя к требуемому уменьшению позиционной чувствительности системы в целом.

Каждый нейрон в слое, близком к входному, реагирует на определенные образы в определенном месте, такие, как угол с определенной ориентацией в заданной позиции.

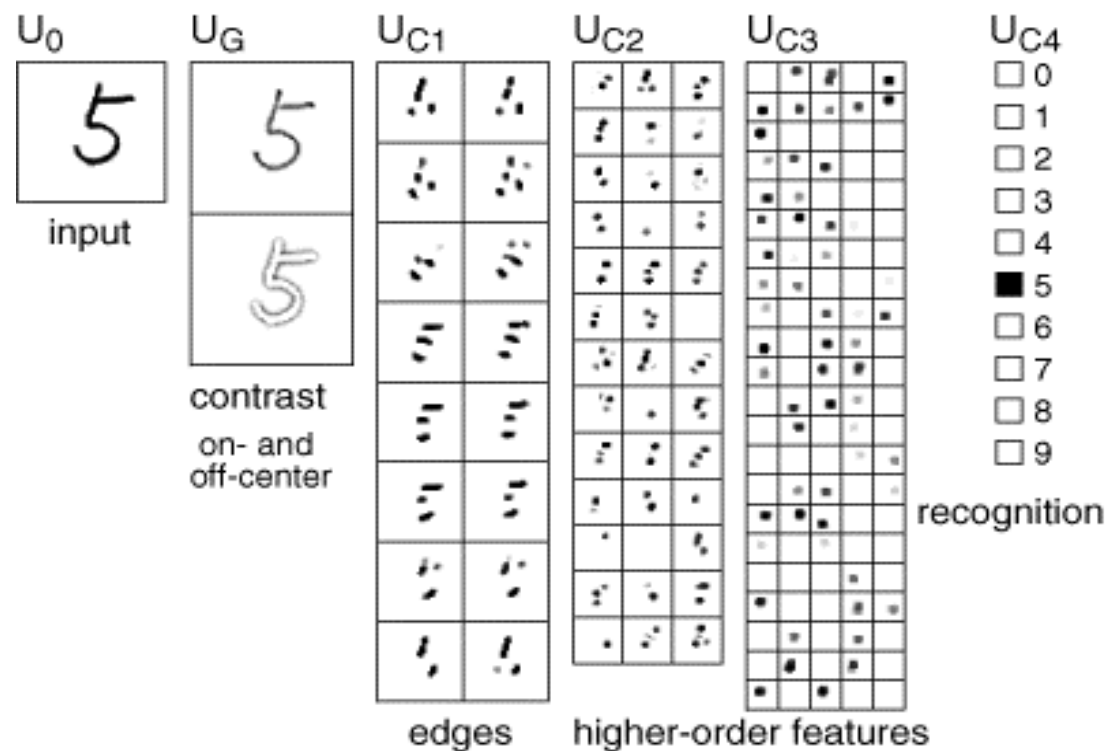
Каждый слой в результате этого имеет более абстрактную, менее специфичную реакцию по сравнению с предшествующим; выходной слой реагирует на полные образы, показывая высокую степень независимости от их положения, размера и ориентации во входном поле.

При использовании в качестве классификатора комплексный узел выходного слоя с наибольшей реакцией реализует выделение соответствующего образа во входном поле. В идеальном случае это выделение нечувствительно к позиции, ориентации, размерам или другим искажениям.

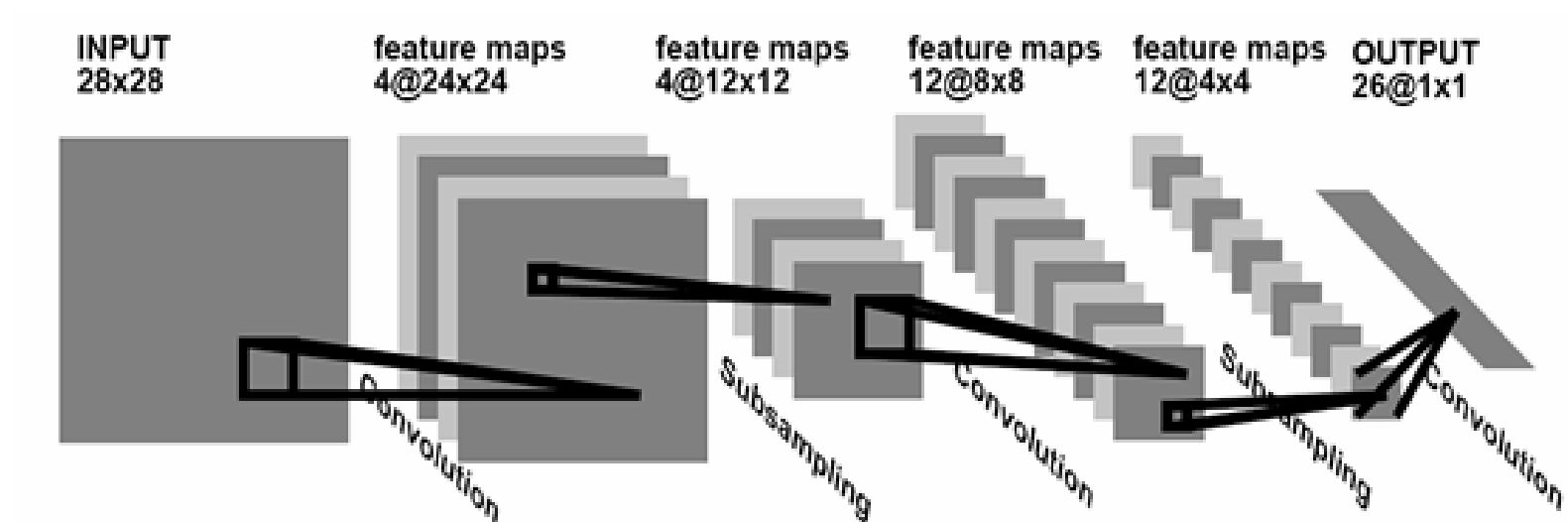
Каждый простой узел реагирует только на набор комплексных узлов внутри своей рецептивной области. Кроме того, существует тормозящий узел, реагирующий на те же самые комплексные узлы. Веса синапсов тормозящего узла не обучаются, – они выбираются таким образом, чтобы узел реагировал на среднюю величину выходов всех узлов, к которым он подключен. Единственный тормозящий синапс от тормозящего узла к простому узлу обучается, как и другие синапсы.



Пример – распознавание неокогнитроном рукописных цифр



Неокогнитрон для распознавания рукописных букв

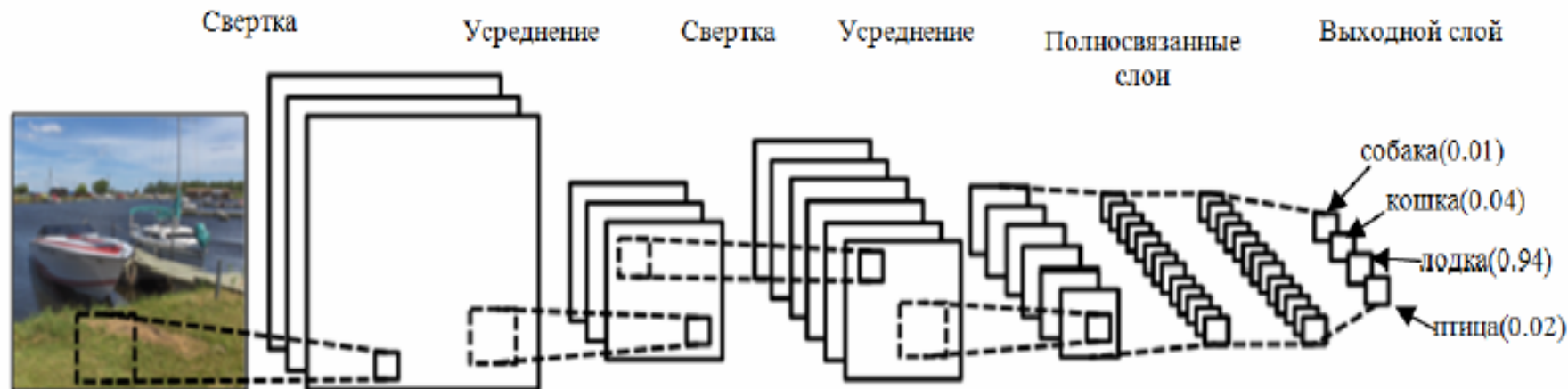


Проблемы при использовании неокогнитрона

- Специализация на распознавание определенного типа образов с учетом определенных видов помех (поэтому появились разные модификации неокогнитрона)
- Разработка архитектуры неокогнитрона под конкретную задачу – нетривиальная задача

Сверточные сети

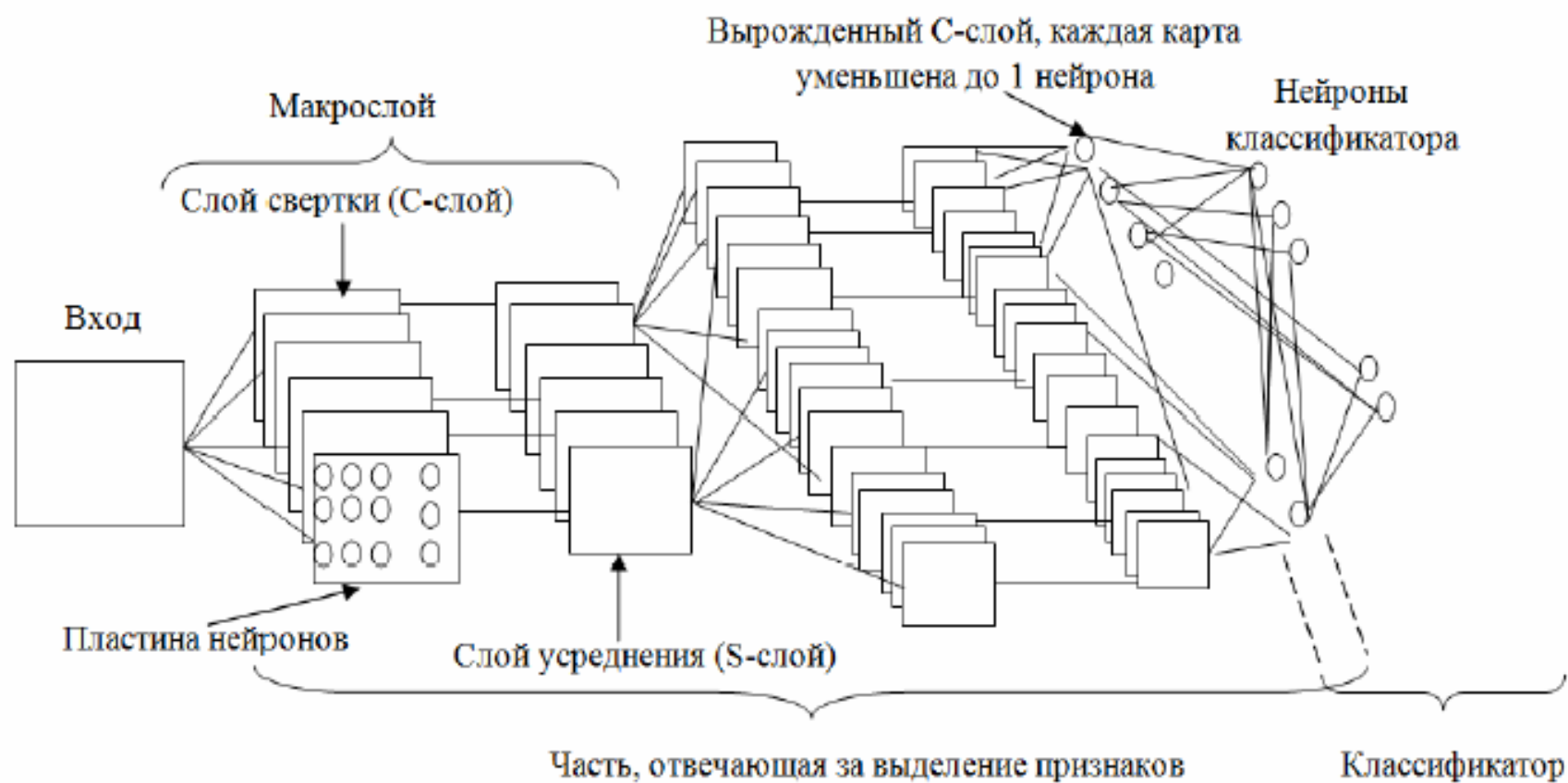
- Отличаются от неокогнитрона тем, что в них используется обучение с учителем методом обратного распространения ошибки, хотя неокогнитрон тоже иногда называют сверточной сетью
- Отличается от MLP локальностью связей между слоями и разным назначением слоев
- Обучение длительное, специализация
- В сверточной сети изображение обрабатывается на сверточных слоях, нелинейных слоях (функция активации), слоях объединения (усреднения) и полносвязных слоях. Каждый из них является самостоятельным, но последовательно расположенные сверточный слой, функция активации и слой объединения иногда заключают в общий блок.



Сверточная сеть Lenet-5 (для распознавания рукописных цифр)



Сверточные сети (2)



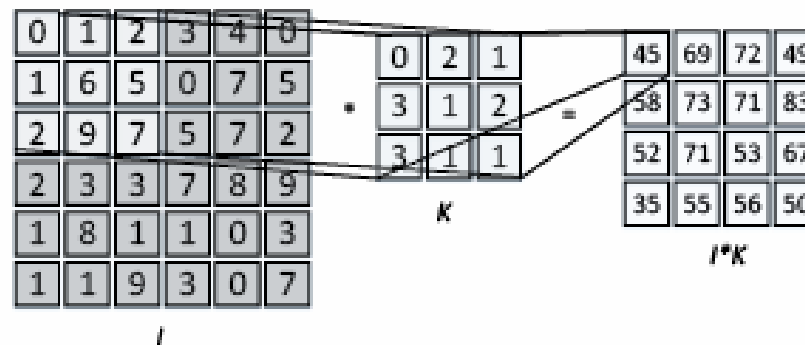
Сверточные сети (3)

Суть процедуры свертки заключается в том, что каждый элемент двумерного изображения I умножается на матрицу K – ядро свертки (convolution kernel) размерности $h \times w$ поэлементно и результат суммируется и записывается в аналогичную позицию выходного изображения:

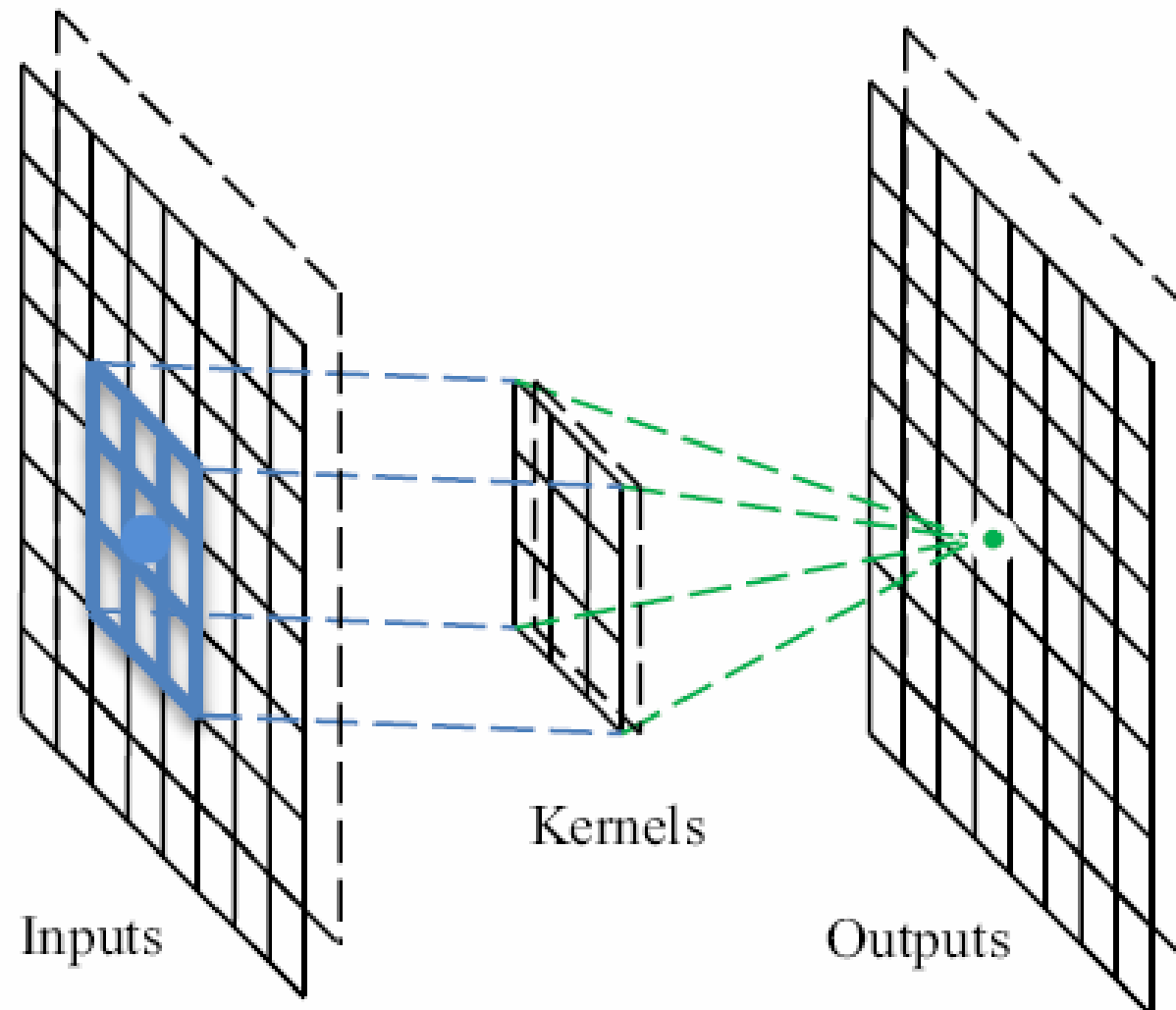
$$(I * K)_{xy} = \sum_{i=-1}^w \sum_{j=-1}^h K_{ij} * I_{x+i, y+j}$$

где x, y – координаты пикселя в исходном изображении,
 i, j – координаты в ядре свертки.

Ядро свертки, по сути, является фильтром, в котором закодирован некоторый ключевой признак, извлекаемый из изображения.



Слой свертки



Свертка

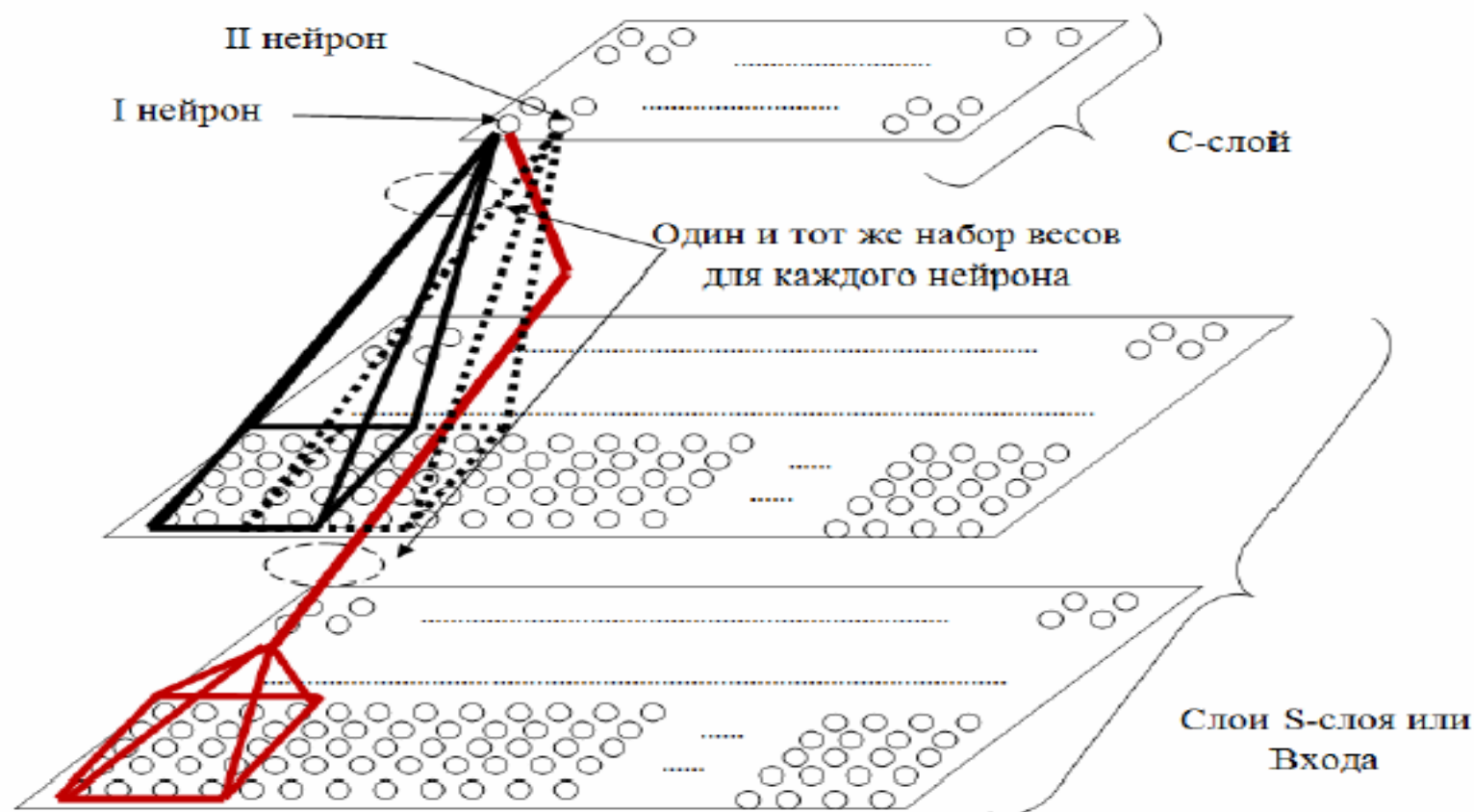


P – фрагмент входного слоя (карты признаков)

F – фильтр (ядро свертки)

На выходе – элемент следующего слоя (карты признаков)

Схема взаимодействия С-слоя с предыдущим слоем

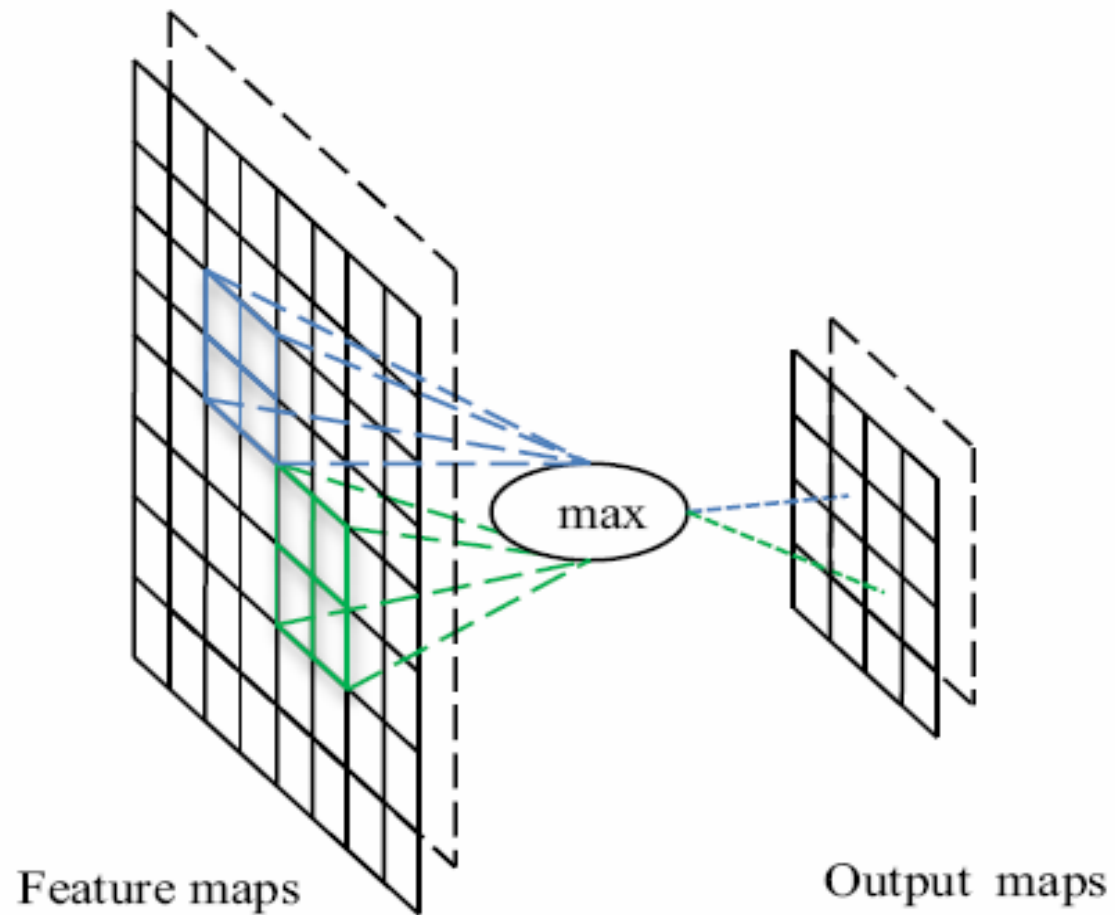


Ядро представляет из себя фильтр или окно, которое скользит по всей области предыдущей карты и находит определенные признаки объектов.

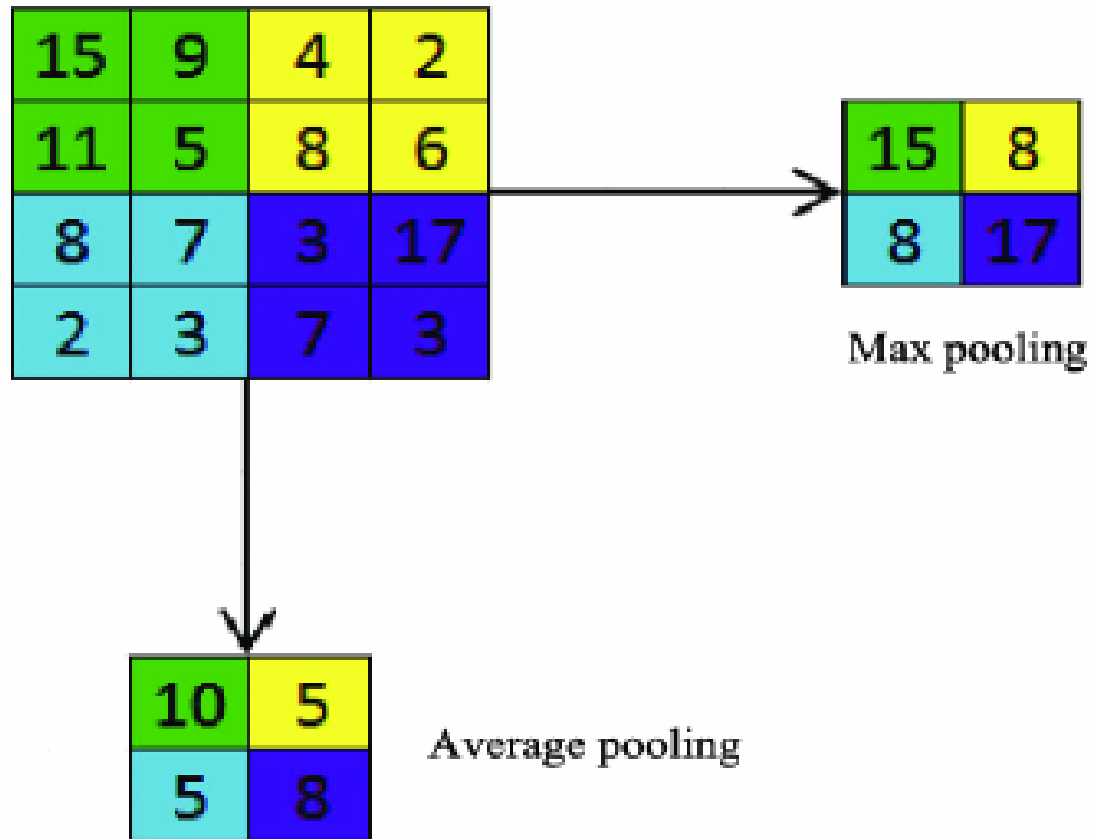
Сверточные сети (4)

- Классическая архитектура СНС подразумевает два типа слоёв:
 - слои свертки (С-слои) (convolution)
 - слои усреднения или подвыборки (S-слои) (pooling)
- Эти слои с разными функциями можно объединить в один макрослой. Выделитель признаков обычно состоит из повторяющихся несколько раз идентичных макрослоёв.
- Каждый нейрон в выделителе признаков имеет своё РП (рецептивное поле), которое представляет собой квадратную область, включающую в себя нейроны, способные передавать свои сигналы нейрону на следующем слое.

Слой max pooling



Разные варианты пулинга



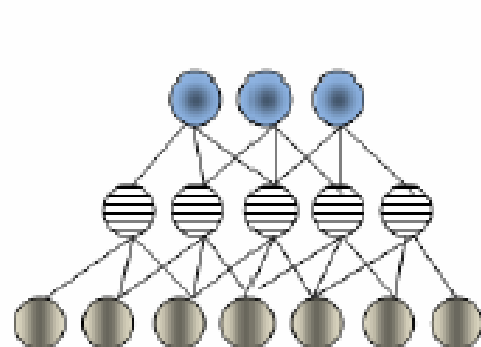
Сверточные сети (5)

- Каждая карта С-слоя может быть связана с некоторым количеством карт предыдущего слоя.
- Из рисунка видно, что карта С-слоя связана с двумя картами S-слоя или входного слоя, которые располагаются перед ней. Тогда каждый нейрон С-слоя имеет рецептивное поле (РП) сразу на двух картах в идентичных позициях предыдущего слоя. На рисунке эти поля обозначены сплошным черным и красным цветом и имеют размер 4×4 . Значит, для нейрона С-слоя необходимо 33 настраиваемых параметров (весов) с учётом смещения.

Сверточные сети (6)

- Особенность С-слоя заключается в том, что для всех нейронов используются одни и те же веса, т.е. на данную карту С-слоя будет приходиться 33 параметра, а не 33 умноженное на количество нейронов в карте.
- Такие веса называются связанными (sharing weights). Связанные веса позволяют реагировать на некоторый признак в любом месте карты, а также, связь карты С-слоя с несколькими картами предыдущего слоя позволяет единым образом реагировать на информацию из разных модальностей (по-разному отфильтрованную), и организовывать некое подобие гиперколонок (как в мозге)
- Основная задача S-слоя – это уменьшение масштаба обрабатываемого отфильтрованного отображения, получившегося на карте предыдущего С-слоя, для облегчения выделения общих признаков. Каждая карта S-слоя связана только с одной картой предыдущего С-слоя.

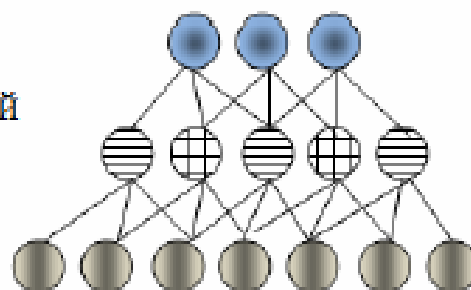
Сверточные сети (7). Различные варианты связанности весов в С-слое



Один набор весов

S-слой

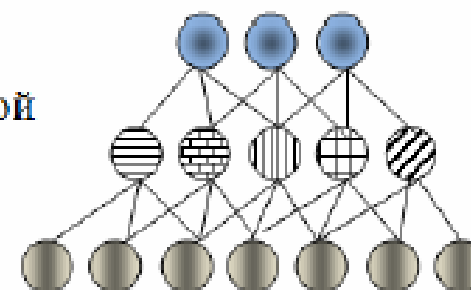
C-слой



Два набора весов

S-слой

C-слой



Пять наборов весов

2x2 pooling, stride 2

9	3	5	3
10	32	2	2
1	3	21	9
2	6	11	7

Max pooling

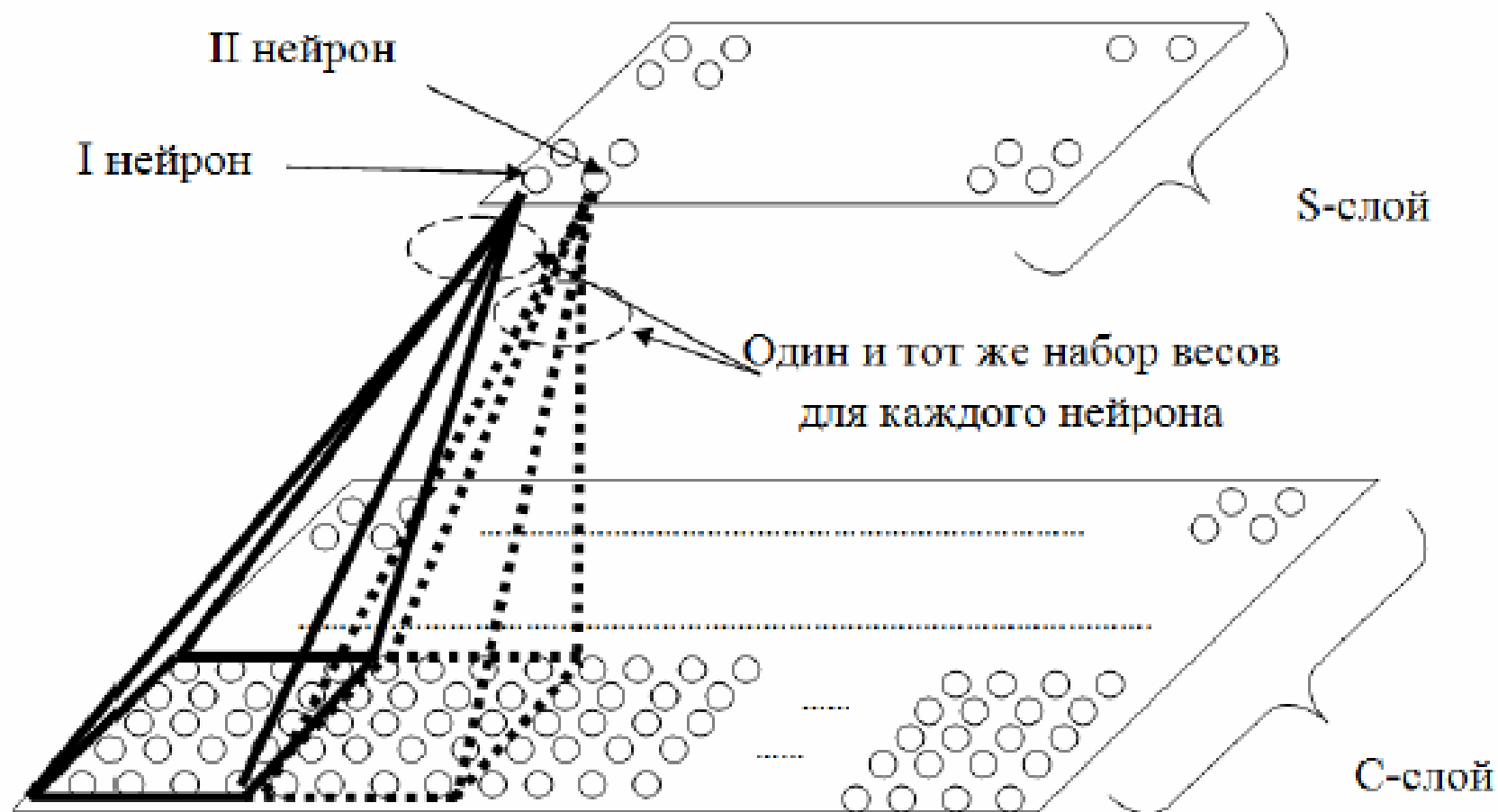
32	5
6	21

Average pooling

18	3
3	12

Сверточные сети (8).

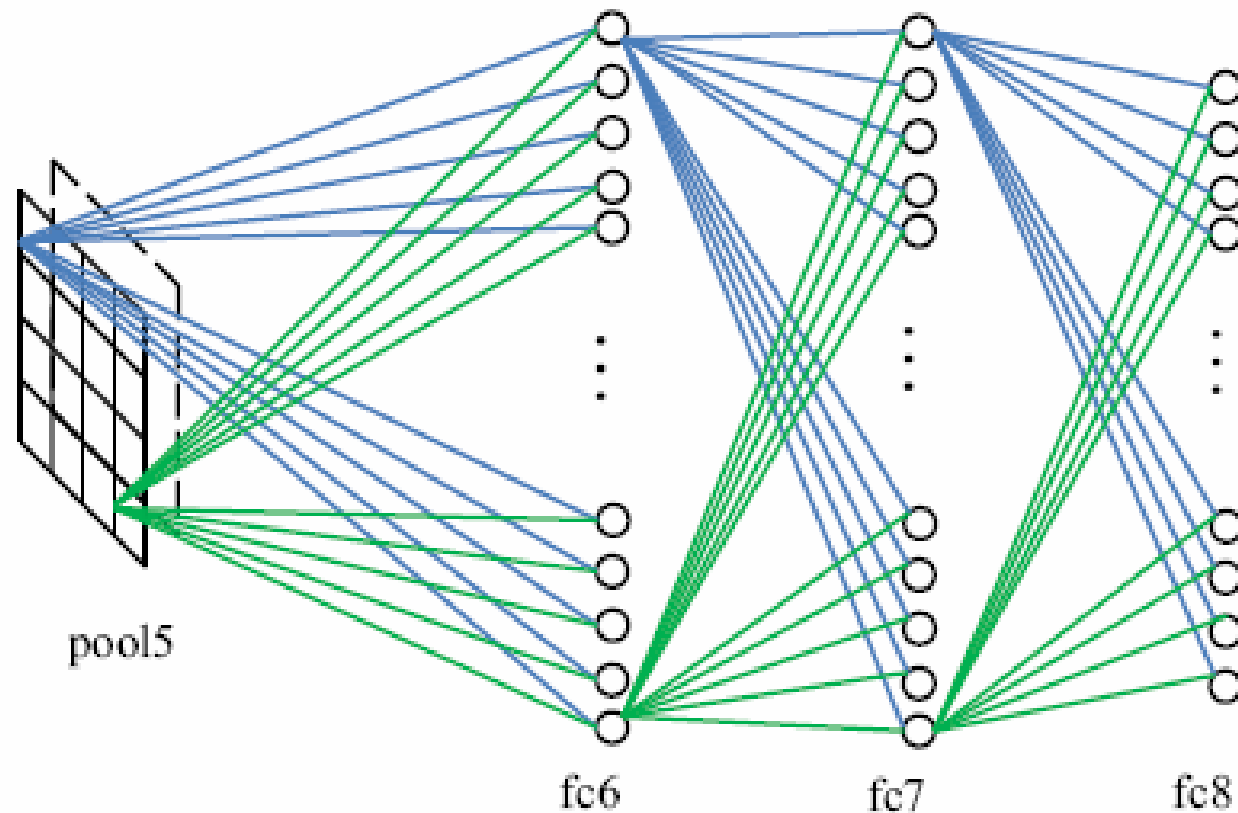
Общая схема взаимодействия S-слоя с предыдущим С-слоем



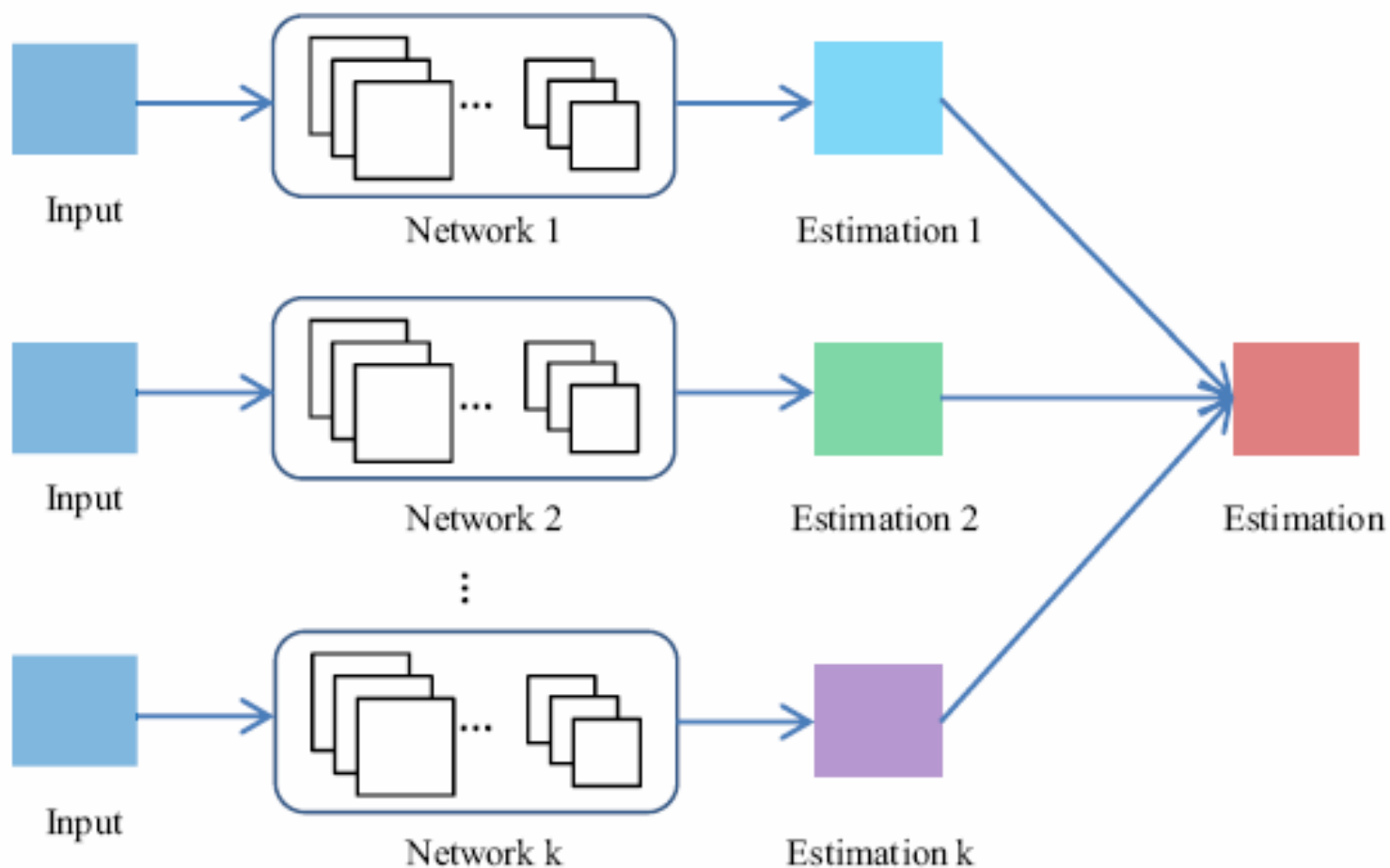
Сверточные сети (9)

- Общий размер карт от слоя к слою падает, пока карты не вырождаются в единичные нейроны, при этом уменьшение размера карт и связанного с этим потерей точности отображений на них, компенсируется ростом их количества и связанностью между картами С-слоя и слоем перед ним.
- Классификатор признаков, как правило, это однослойный или двухслойный персептрон, хотя может быть SVM или RBF-сеть.
- Карты, выродившиеся в один нейрон можно считать за вход для классификатора, над входом надстраивается слой обычных нейронов, где каждый нейрон связан со всеми нейронами входа. Количество нейронов в слое классификатора обычно равняется количеству классов, к которым требуется отнести входное изображение. Связанных весов в классификаторе нет.

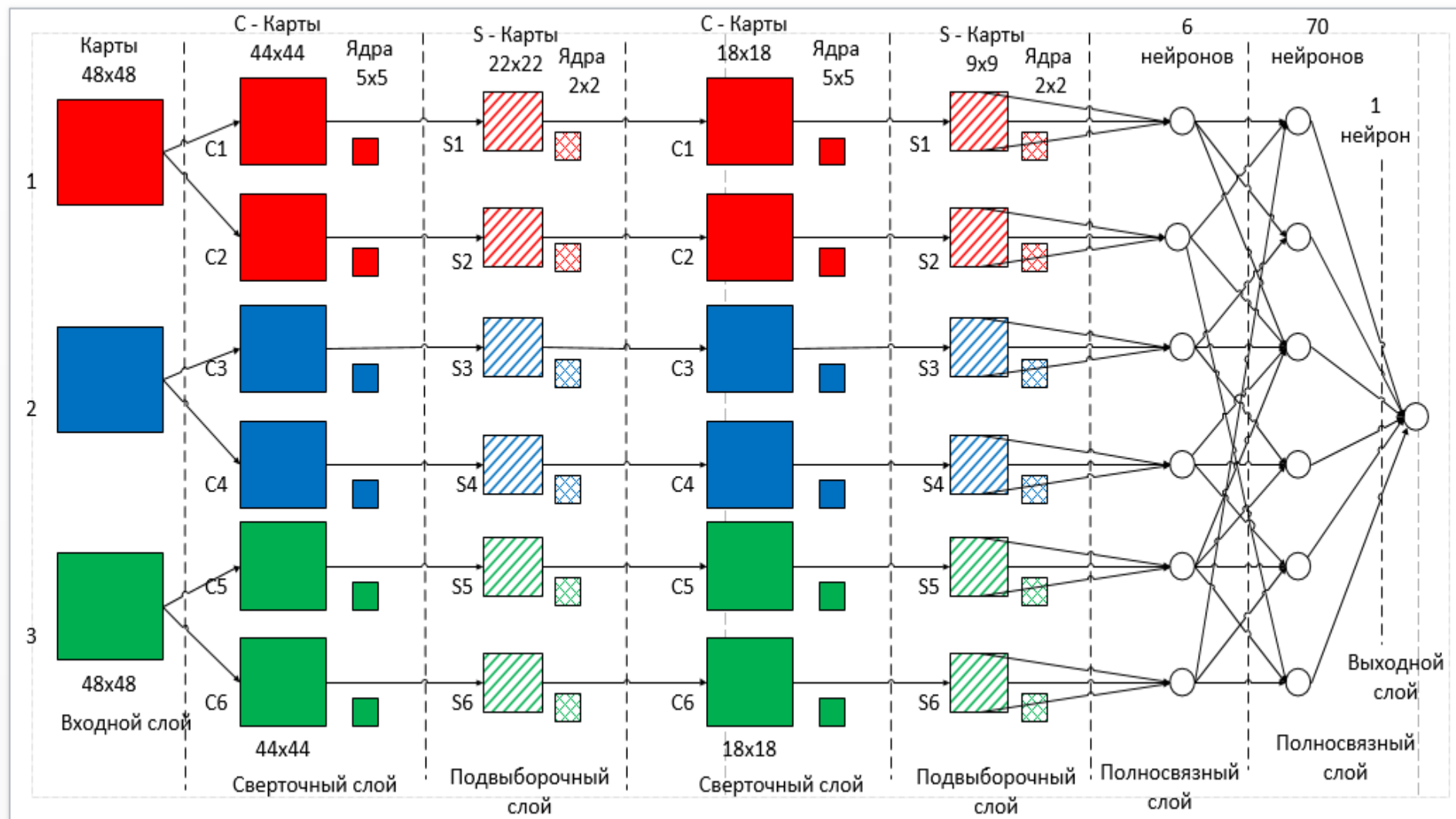
Полносвязный классификатор на выходе сверточной сети



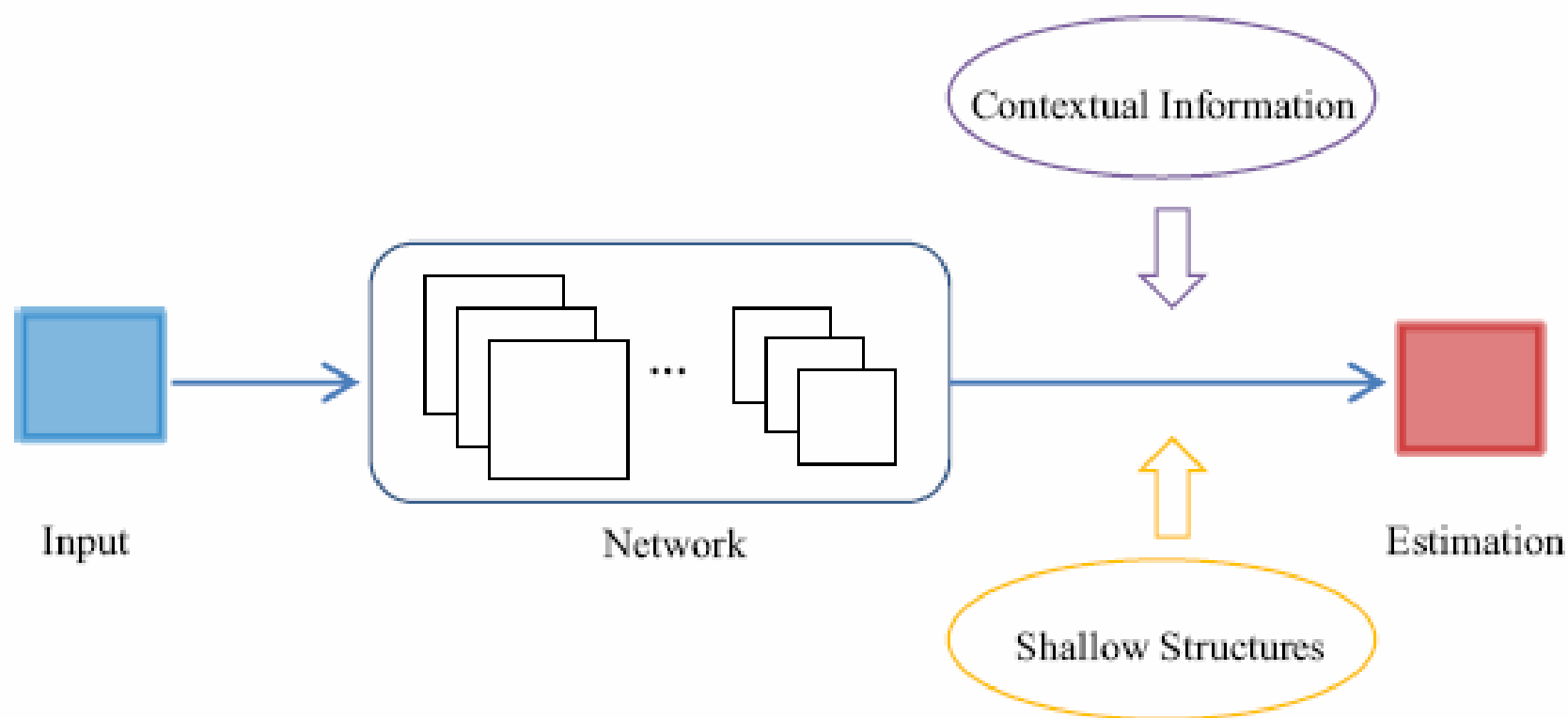
Комбинация нескольких сетей



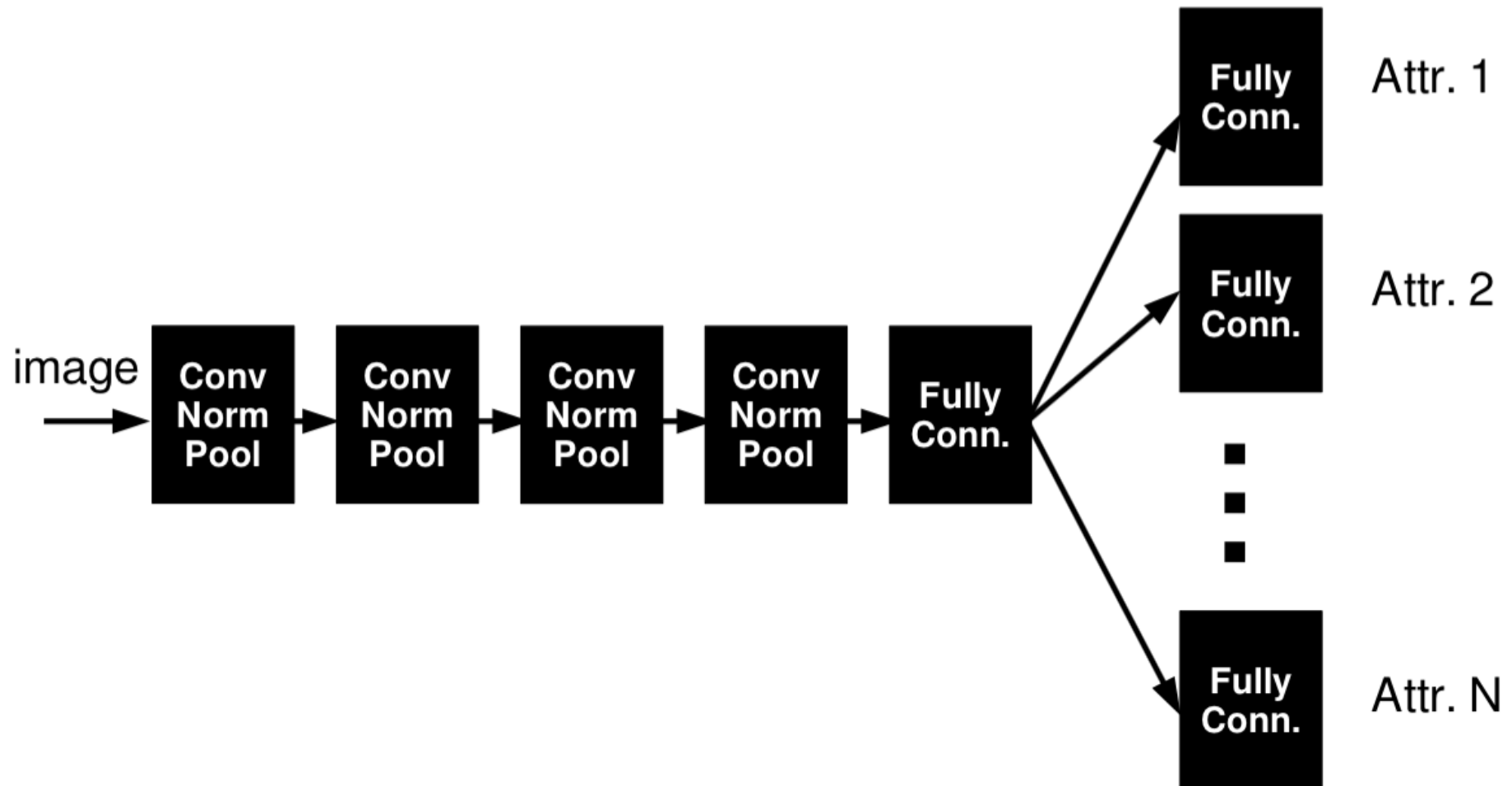
Сверточная сеть для обработки RGB-изображения



Комбинация сверточной сети с дополнительными источниками данных



Многозадачная глубокая сеть



Нормализация входных данных

$$f(p, min, max) = \frac{p - min}{max - min},$$

где f – функция нормализации;

p – значение конкретного цвета пикселя от 0 до 255;

min – минимальное значение пикселя – 0;

max – максимальное значение пикселя – 255.

Преимущества сверточных сетей

- хорошо справляется с задачей классификации и распознаванию изображений;
- меньшее количество настраиваемых параметров, по сравнению с полносвязной нейронной сетью;
- более высокая устойчивость к повороту и сдвигу распознаваемого изображения, по сравнению полносвязными нейронными сетями (типа персептрона).

Недостатки сверточных сетей

- Большой объем размеченных данных, на которых должна обучаться НС
- Большое количество настраиваемых гиперпараметров, к которым относятся:
 - количество сверточных и подвыборочных слоев;
 - количество карт признаков в сверточных и подвыборочных слоях;
 - размерность матрицы свёртки;
 - функция активации нейронов;
 - скорость обучения сети.
- Невозможность объяснить как НС приняла то или иное решение
- Возможность непредсказуемых ошибок при распознавании, причины которых:
 - Переобучение НС (утеря способности к обобщению)
 - Эффект длинного хвоста (большое количество маловероятных изображений, которые могут встретиться при реальном использовании НС)
 - Возможность «обмануть» НС, предъявив ей специальным образом искаженное изображение, которое она распознает неверно

Простая сверточная сеть в Keras (для распознавания рукописных цифр)

```
model = keras.Sequential([  
    Conv2D(32, (3,3), padding='same', activation='relu', input_shape=(28, 28,  
1)),  
    MaxPooling2D((2, 2), strides=2),  
    Conv2D(64, (3,3), padding='same', activation='relu'),  
    MaxPooling2D((2, 2), strides=2),  
    Flatten(),  
    Dense(128, activation='relu'),  
    Dense(10, activation='softmax') ])
```

```
print(model.summary())
```

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 32)	320
max_pooling2d (MaxPooling2D)	(None, 14, 14, 32)	0
conv2d_1 (Conv2D)	(None, 14, 14, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 64)	0
flatten (Flatten)	(None, 3136)	0
dense (Dense)	(None, 128)	401536
dense_1 (Dense)	(None, 10)	1290

```
=====
Total params: 421,642
Trainable params: 421,642
Non-trainable params: 0
```

Недостатки использования обратного распространения ошибки для многослойных НС

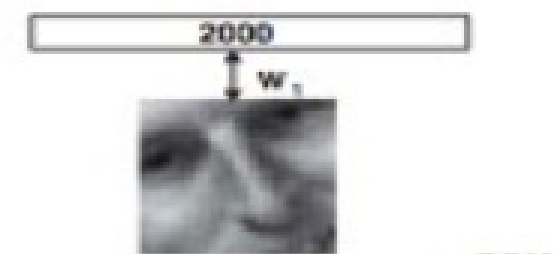
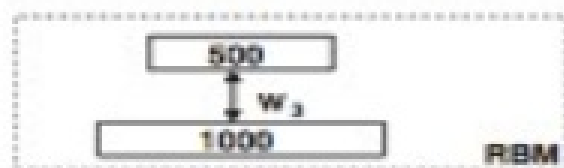
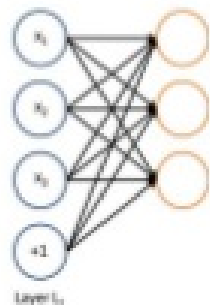
- Эффект исчезающего градиента при продвижению ошибки к нижним слоям
- Эффект переобучения (переобученности). Сеть ведет себя хорошо на обучающих примерах и сильно ошибается на тестовых примерах.
- Поэтому, многослойные сети прямого распространения не получили широкого распространения до появления методов послонного обучения

- В середине 2000-х годов Джеффри Хинтон показал возможность послойного обучения многослойной сети прямого распространения
- Он использовал ограниченную машину Больцмана (вероятностное обучение)
- Но можно использовать и метод обратного распространения ошибки
- Для обучения одного скрытого слоя используется идея AutoEncoder (Автокодировщик)

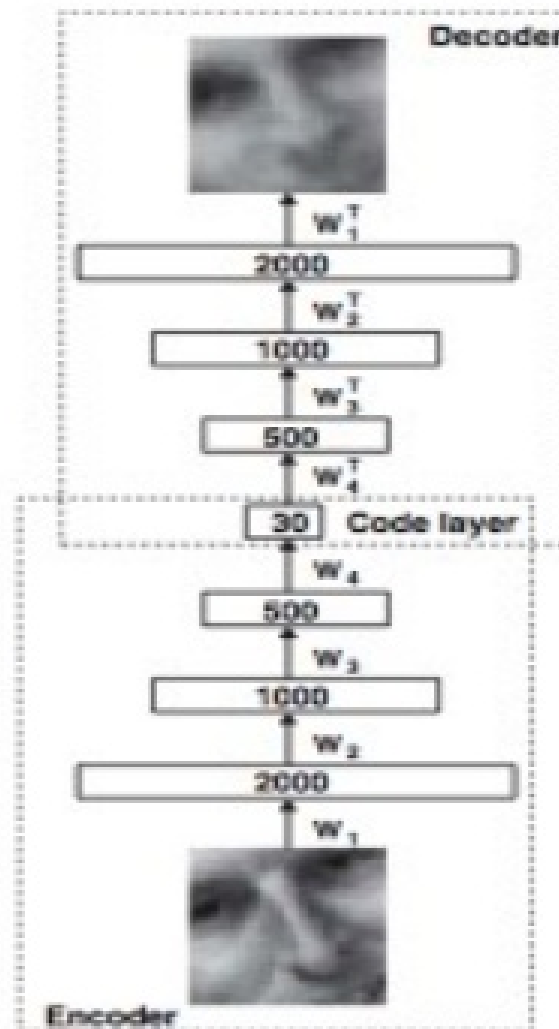
2. How to train AutoEncoder?

Hinton (2006) Science Paper

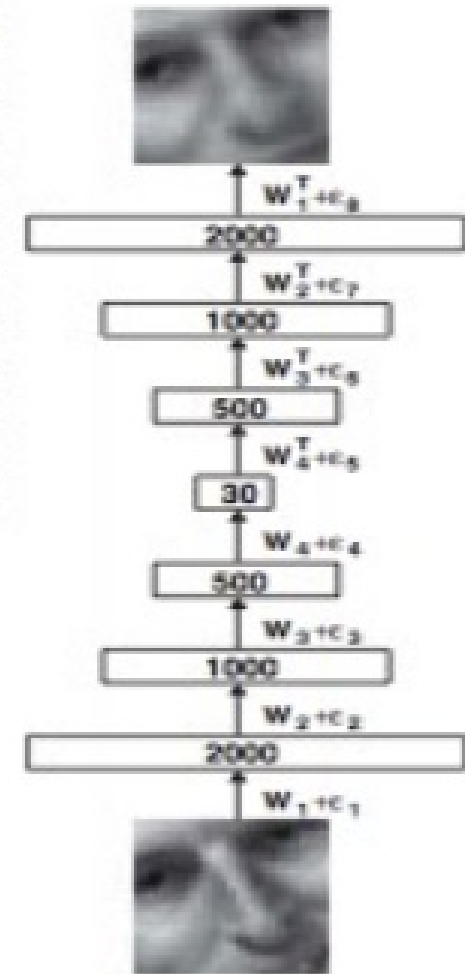
restricted Boltzmann machine



Pretraining



Unrolling

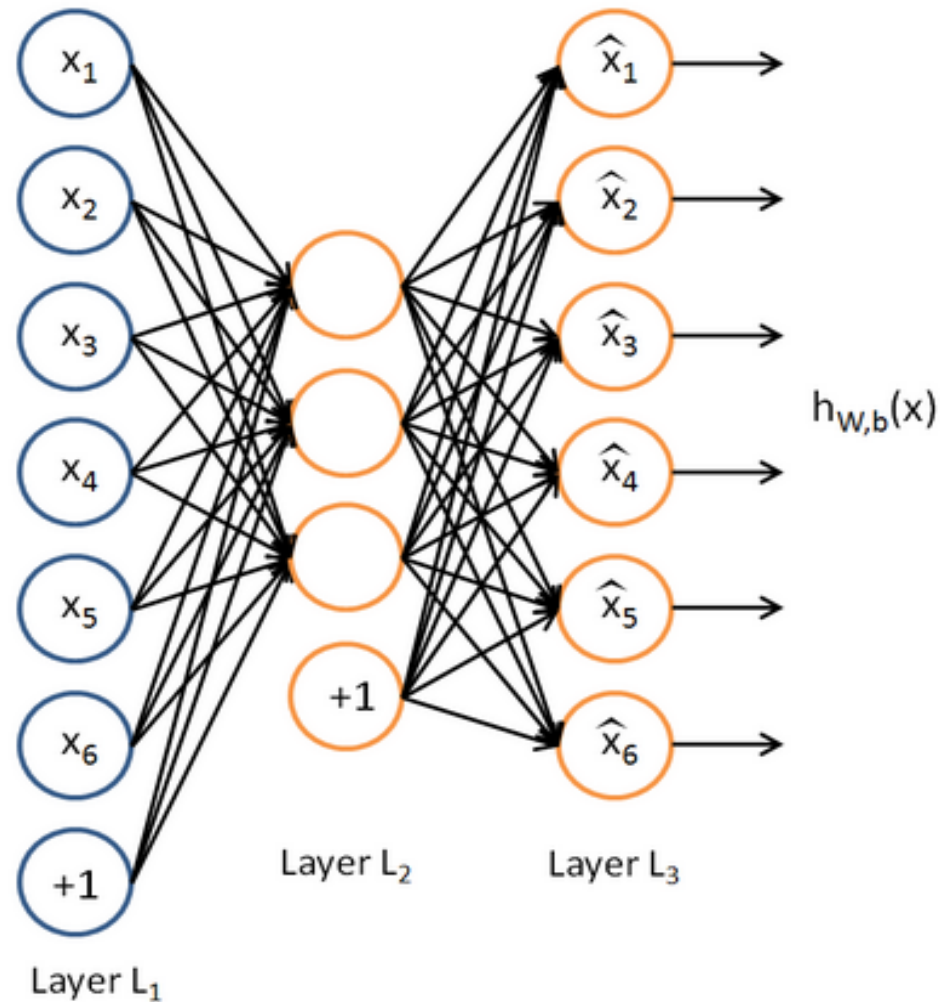


Fine-tuning

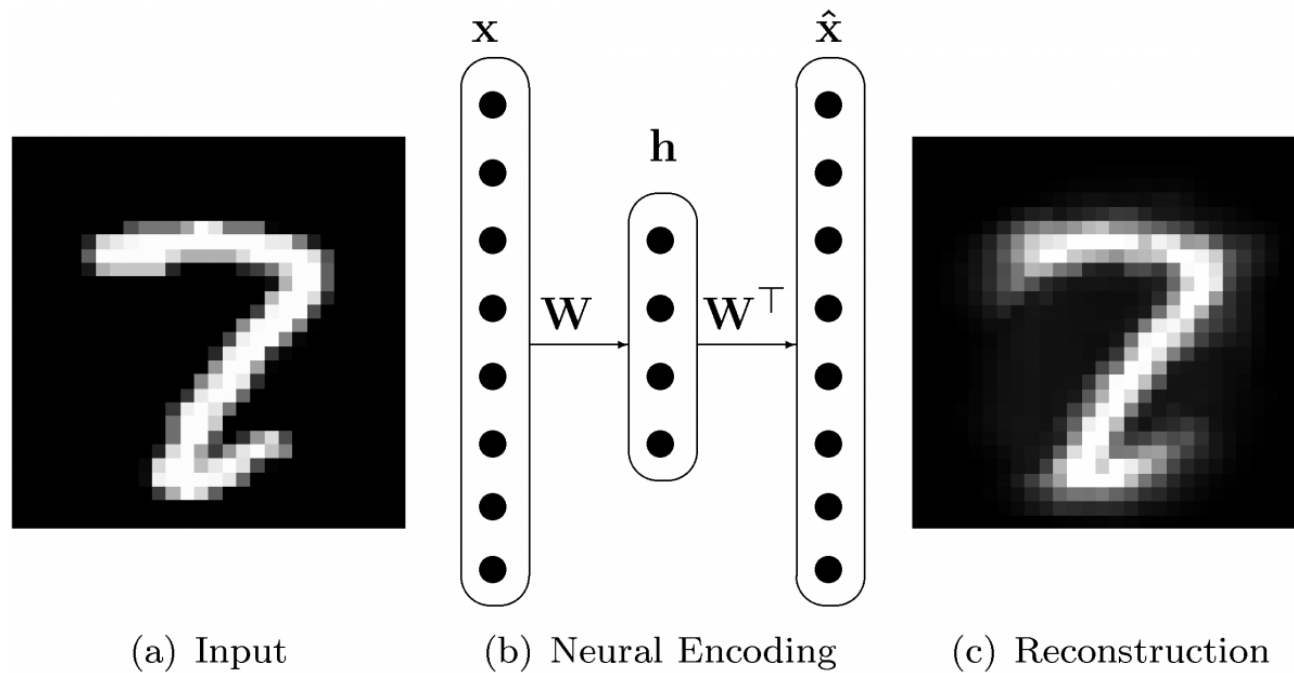
Autoencoder - автокодировщик

- Выходной слой – аналог входного слоя
- Скрытый слой состоит из небольшого количества нейронов
- АЕ обучается восстановлению образа на выходе, создавая его сжатое описание в скрытом слое в виде признаков (пытается построить функцию $h(x) = x$)
- Обучение с использованием модифицированного алгоритма обратного распространения ошибки, обеспечивающего разреженную активацию нейронов скрытого слоя (используя штрафную функцию – имитацию латеральных тормозящих связей)

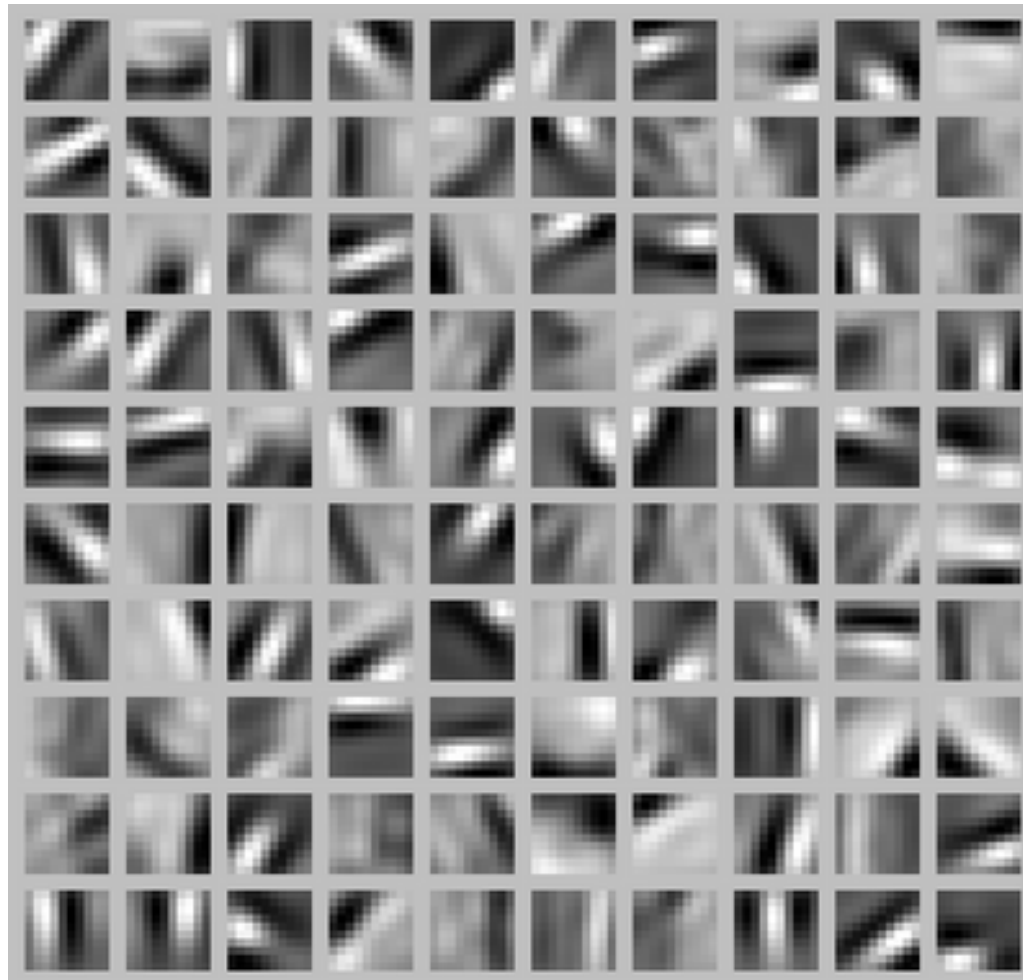
Autoencoder - автокодировщик



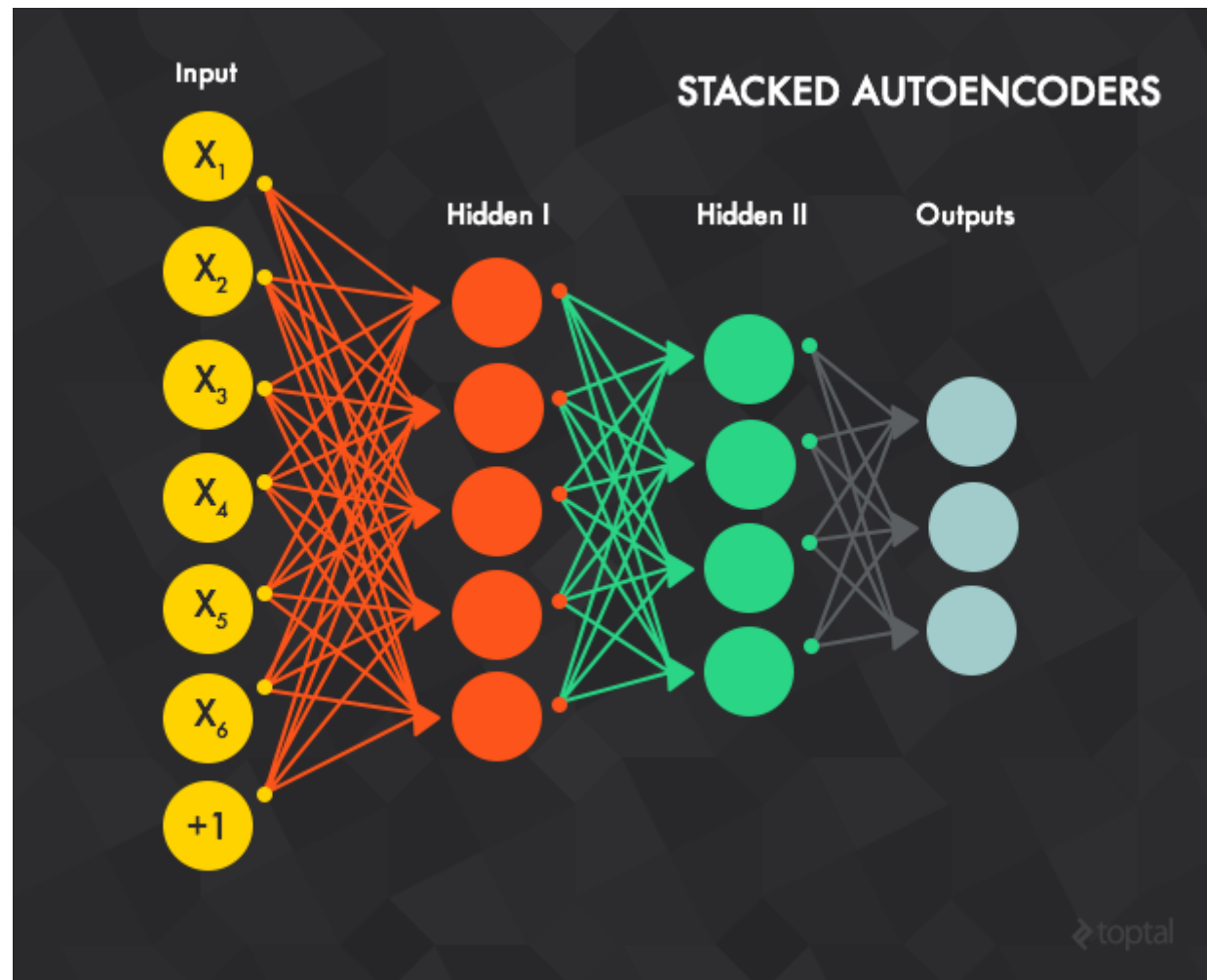
Пример восстановления изображения автокодировщиком



Пример изображений, соответствующих
активации скрытых ста нейронов после
обработки реальных фотографий



Использование концепции автокодировщика для обучения многослойной сети (stacked autoencoders)



Использование концепции автокодировщика для обучения многослойной сети

- Обучается слой Hidden 1 с использованием временного выходного слоя – аналога слоя Input
- Обучается слой Hidden 2 с использованием входного слоя Hidden 1 и временного выходного слоя – его аналога
- Обучается выходной слой Output с использованием входного слоя Hidden 2 и временного выходного слоя – его аналога
- Так как на выходном слое не получают необходимые для нас метки (распознанные объекты), обучается вся сеть методом обратного распространения ошибки (с учителем)