

Глубокие нейронные сети

Сверточные нейронные сети  
(Convolutional neural networks –  
CNN)

Лекция 9

# Применение глубоких сверточных сетей

## IMAGENET

<http://image-net.org/>

14 197 122 изображений  
21 841 индексированных синсетов

- ImageNet 2011 Fall Release (32326)
  - plant, flora, plant life (4486)
  - geological formation, formation (175)
  - natural object (1112)
  - sport, athletics (176)
  - artifact, artefact (10504)
  - fungus (308)
  - person, individual, someone, somebody, mortal, soul (6978)
  - animal, animate being, beast, brute, creature, fauna (3998)
  - Misc (20400)

<https://www.kaggle.com/>

*Синсет* – множество синонимов, обозначающих одно понятие в WordNet

*WordNet* – лексическая база данных английского языка

*Wiki-ru-wordnet* – лексическая база данных русского языка (<https://wiki-ru-wordnet.readthedocs.io/>)

## Распознавание образов

airplane



automobile



bird



cat



deer



dog



frog



horse



ship

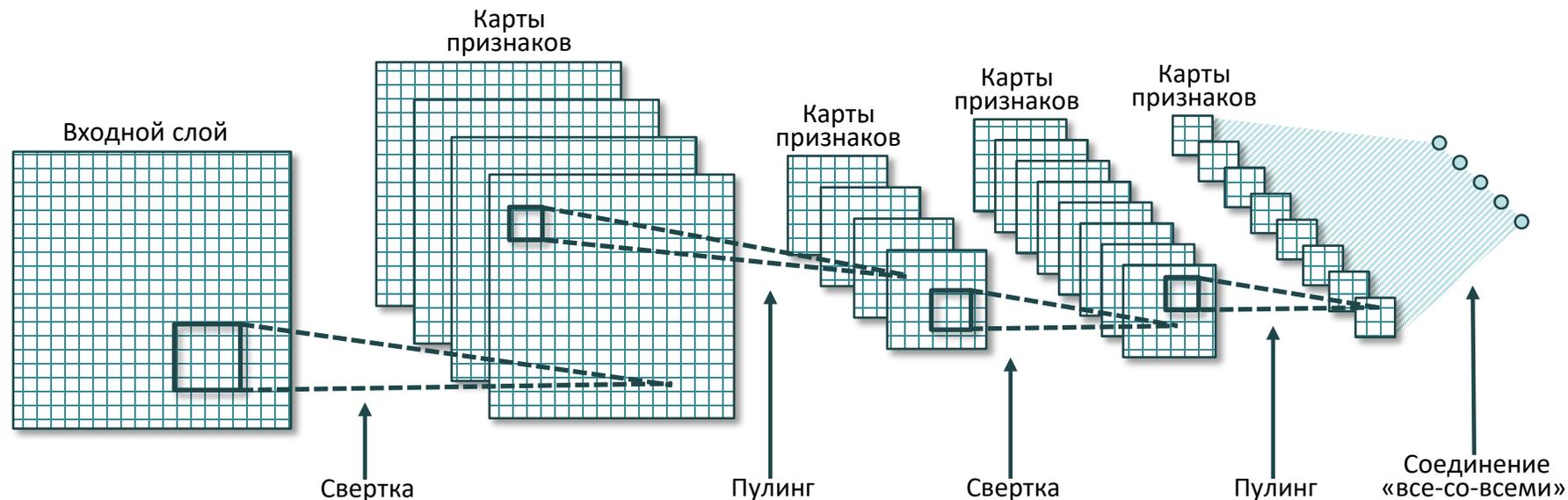


truck

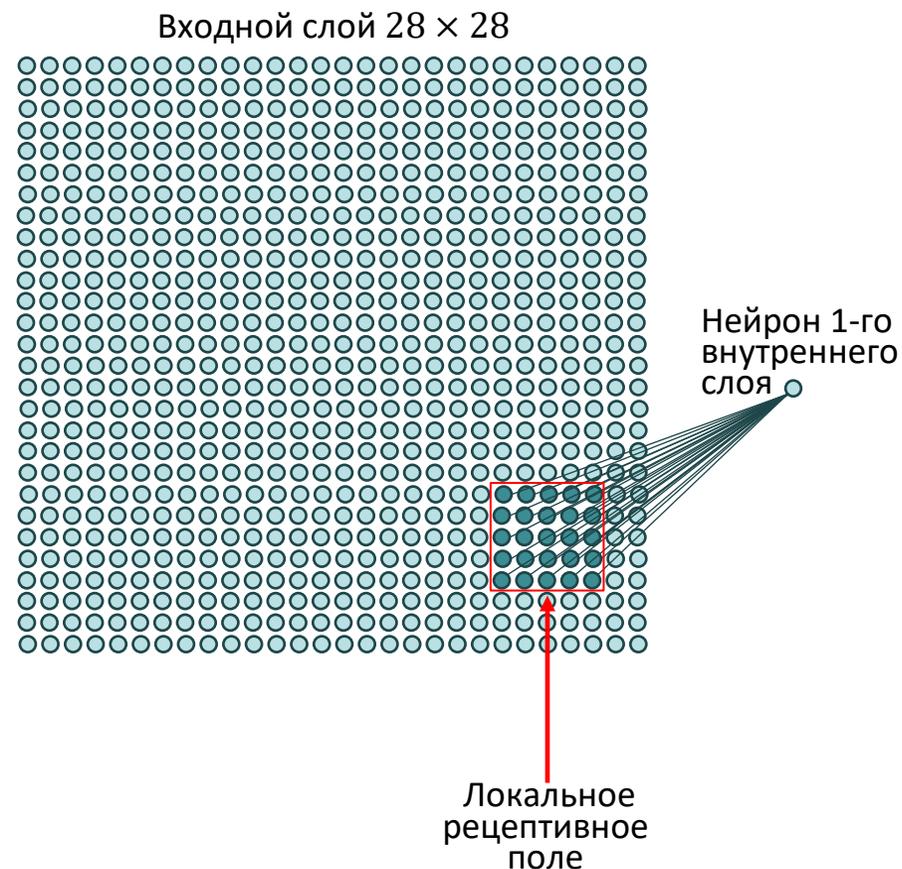


# Концепция

- Входной слой логически представляется как матрица пикселей
- Каждый внутренний слой логически представляется как совокупность карт признаков (*feature map*) [*карт активации (activation map)*], имеющих одинаковые размеры
- Выходной слой имеет обычную линейную структуру и связывается с последним внутренним слоем по принципу «все-со-всеми»
- Нейроны входного слоя и всех внутренних слоев связываются избирательно (не все со всеми)
- Связи строятся двумя чередующимися способами: 1) *свертка*, 2) *пулинг*
- Свертка увеличивает количество карт признаков, сохраняя их размер (при использовании набивки padding), либо незначительно уменьшая их размер
- Пулинг уменьшает размеры карт признаков с сохранением их количества



# Построение сверточной нейронной сети для распознавания рукописных цифр



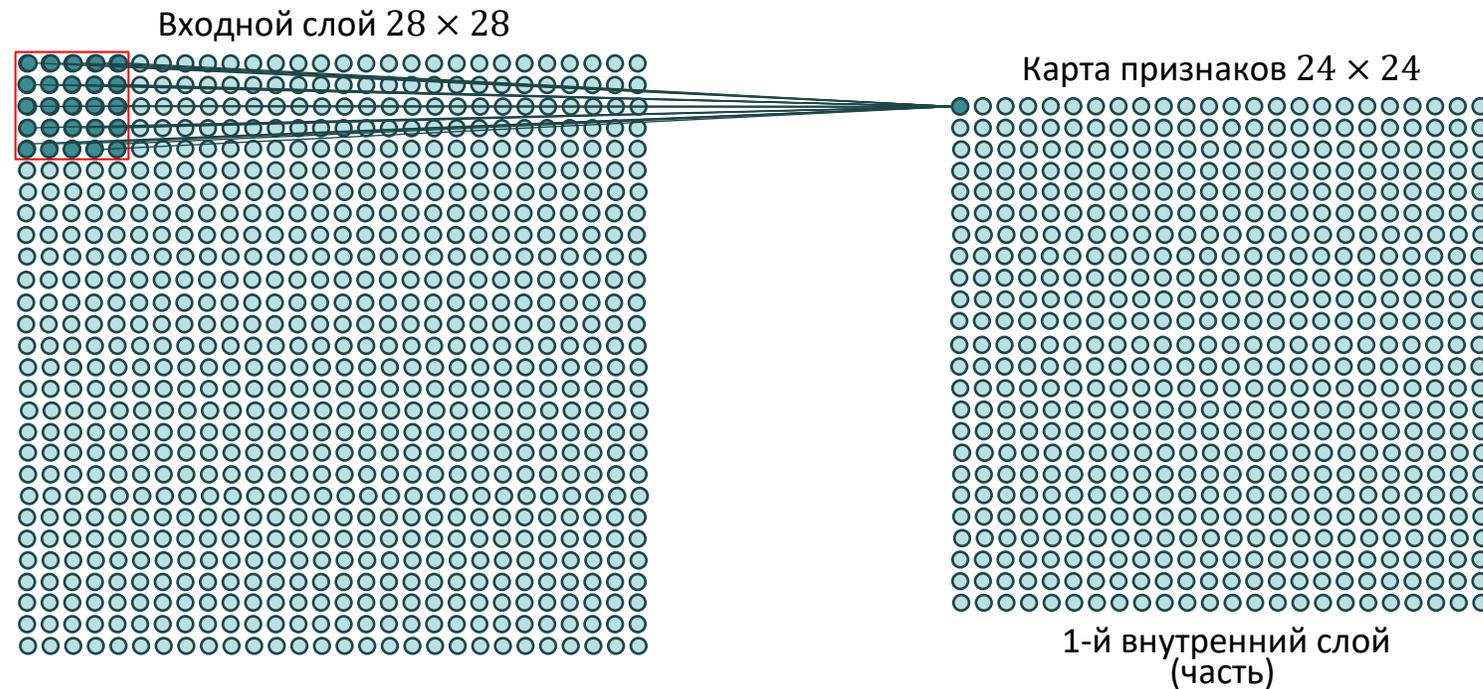
- Входной слой представляет собой квадратную матрицу нейронов  $28 \times 28$ , соответствующих пикселям изображения из базы данных MNIST, подаваемого на вход нейронной сети
- Каждый нейрон первого внутреннего слоя соединяется с группой входных нейронов, образующих прямоугольный слайдер размером  $5 \times 5$ , называемый *локальным рецептивным полем*

# Построение 1-го внутреннего слоя



- Формируется карта признаков, состоящая из нейронов, образующих 1-й внутренний слой, размером  $24 \times 24$  (вычисляется по формуле  $24 = 28 - 5 + 1$ )

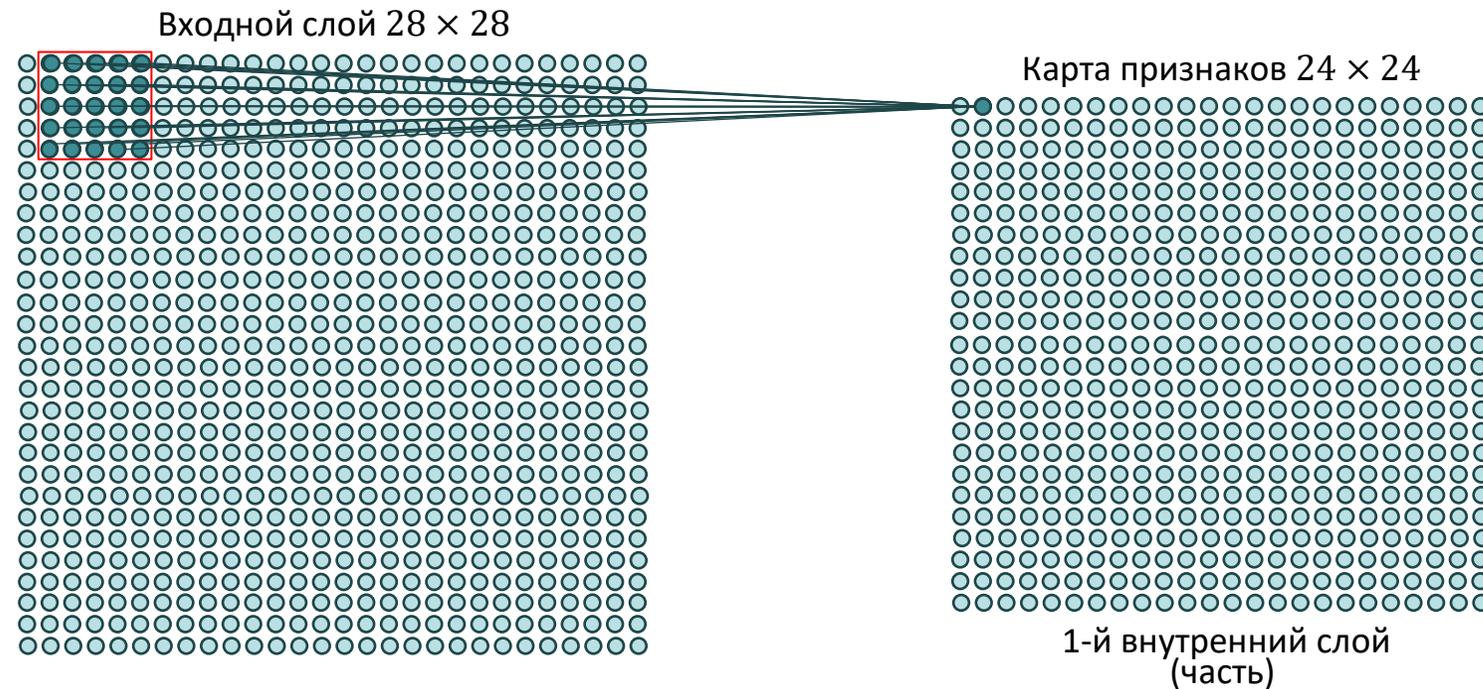
# Построение 1-го внутреннего слоя



- Слайдер локального рецептивного поля устанавливается в левый верхний угол матрицы входного слоя
- Все нейроны входного слоя, попавшие в слайдер, связываются с первым нейроном карты признаков

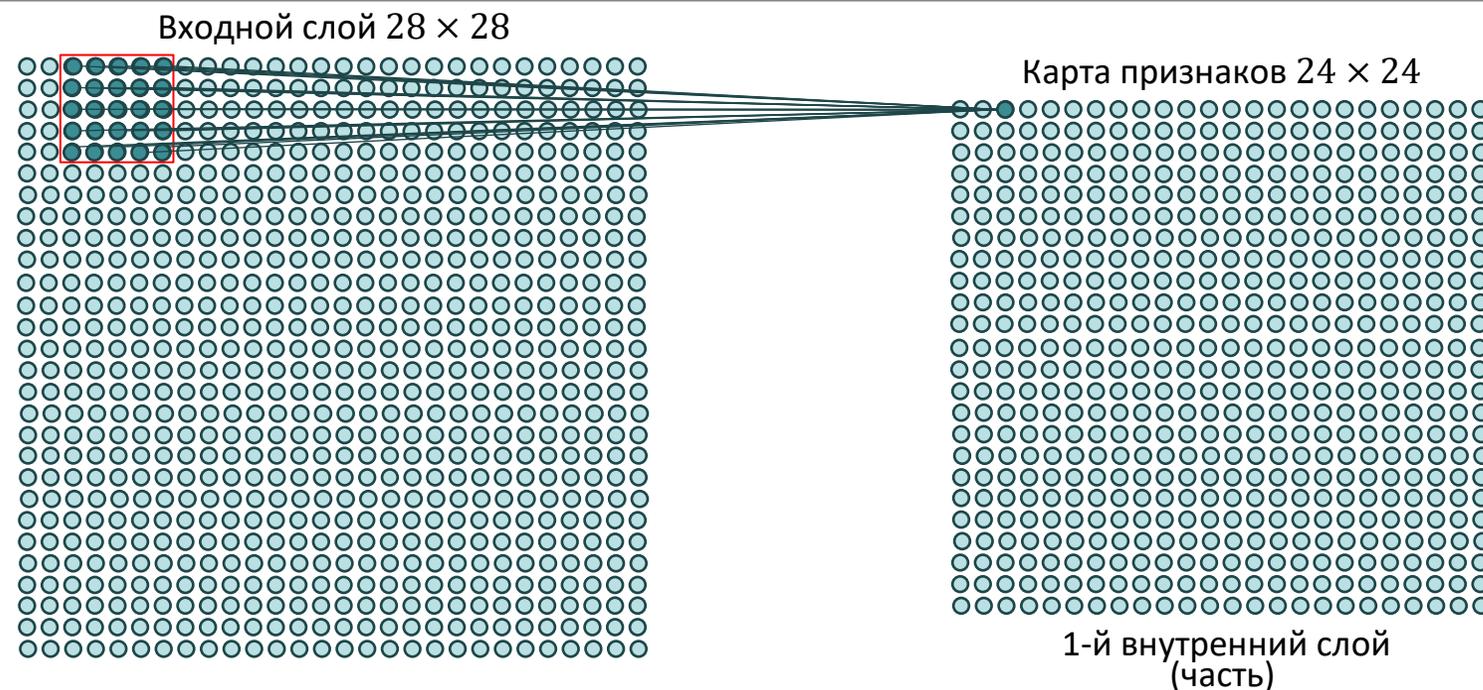


# Построение 1-го внутреннего слоя



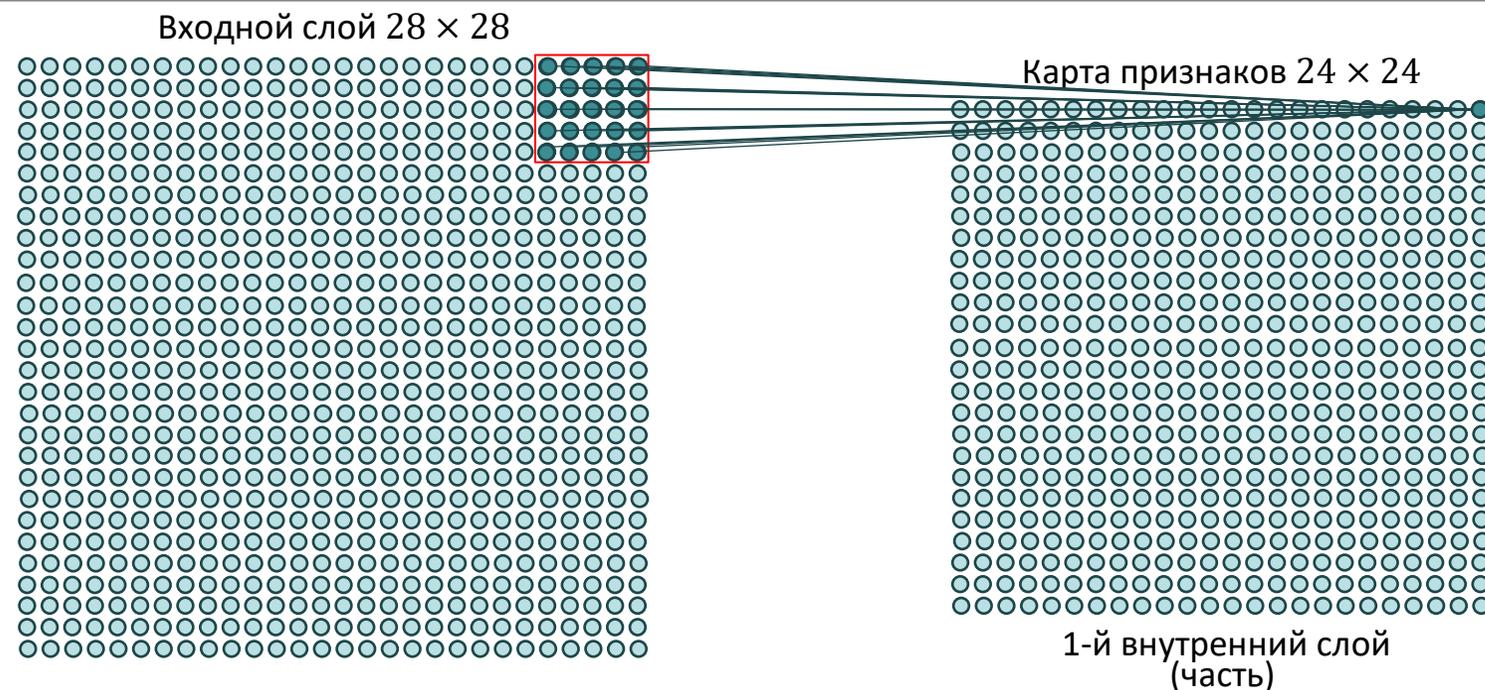
- Слайдер сдвигается на одну позицию вправо
- Все нейроны входного слоя, попавшие в слайдер, связываются со вторым нейроном карты признаков

# Построение 1-го внутреннего слоя



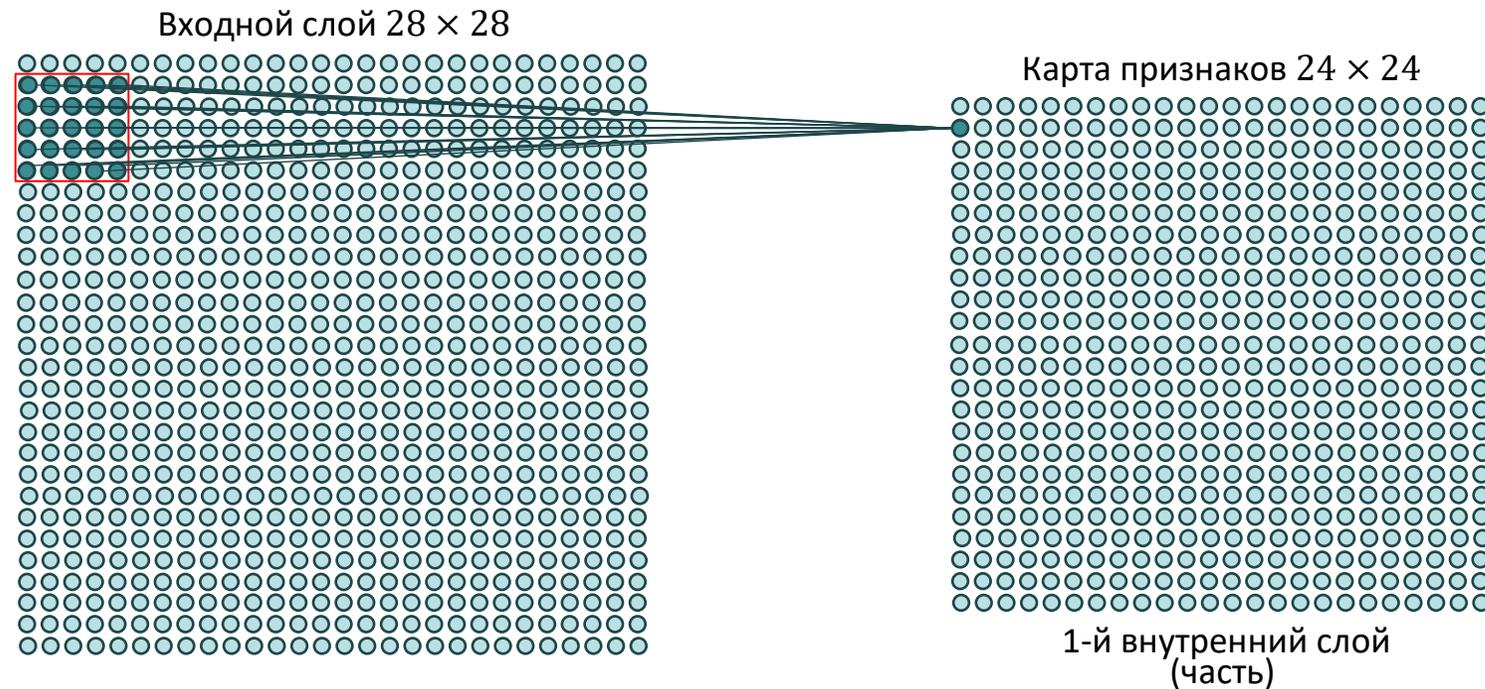
- Слайдер сдвигается еще на одну позицию вправо
- Все нейроны входного слоя, попавшие в слайдер, связываются с третьим нейроном карты признаков

# Построение 1-го внутреннего слоя



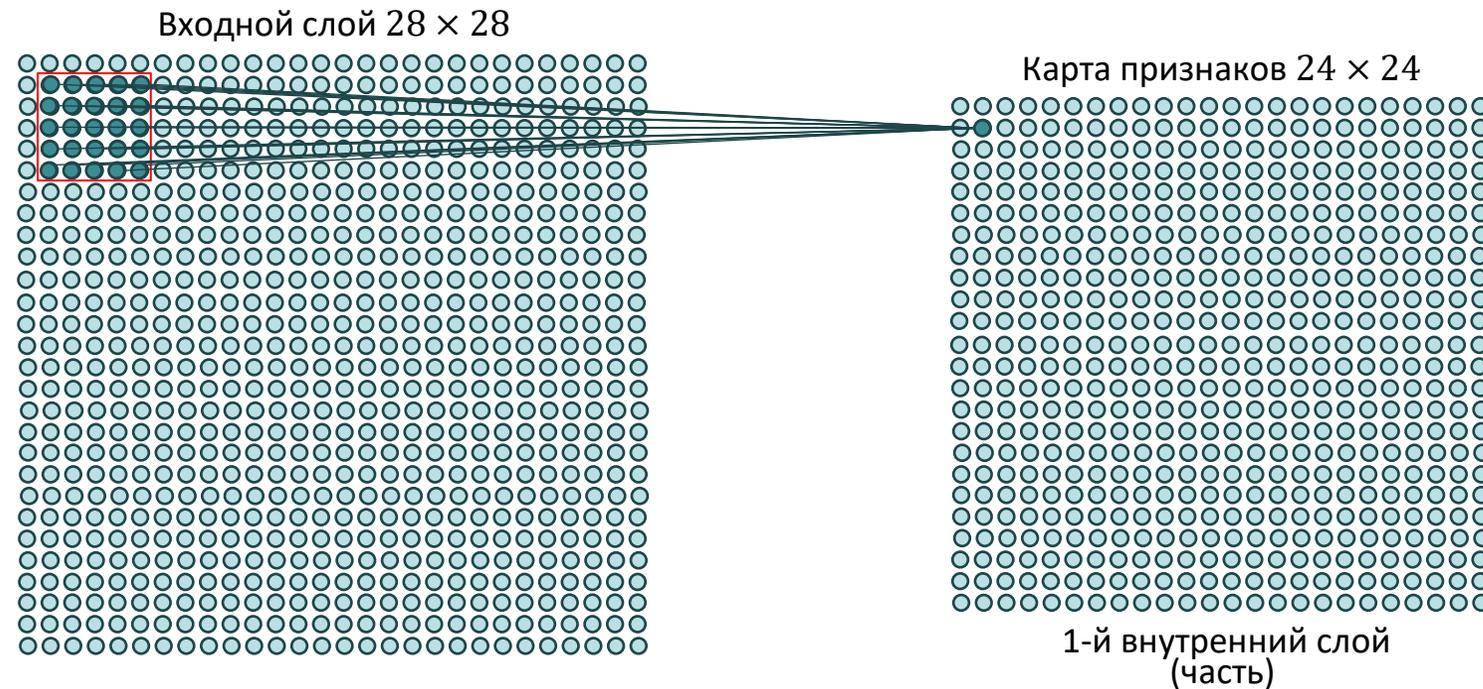
- Таким образом устанавливаются связи для всех нейронов первого ряда карты признаков

# Построение 1-го внутреннего слоя



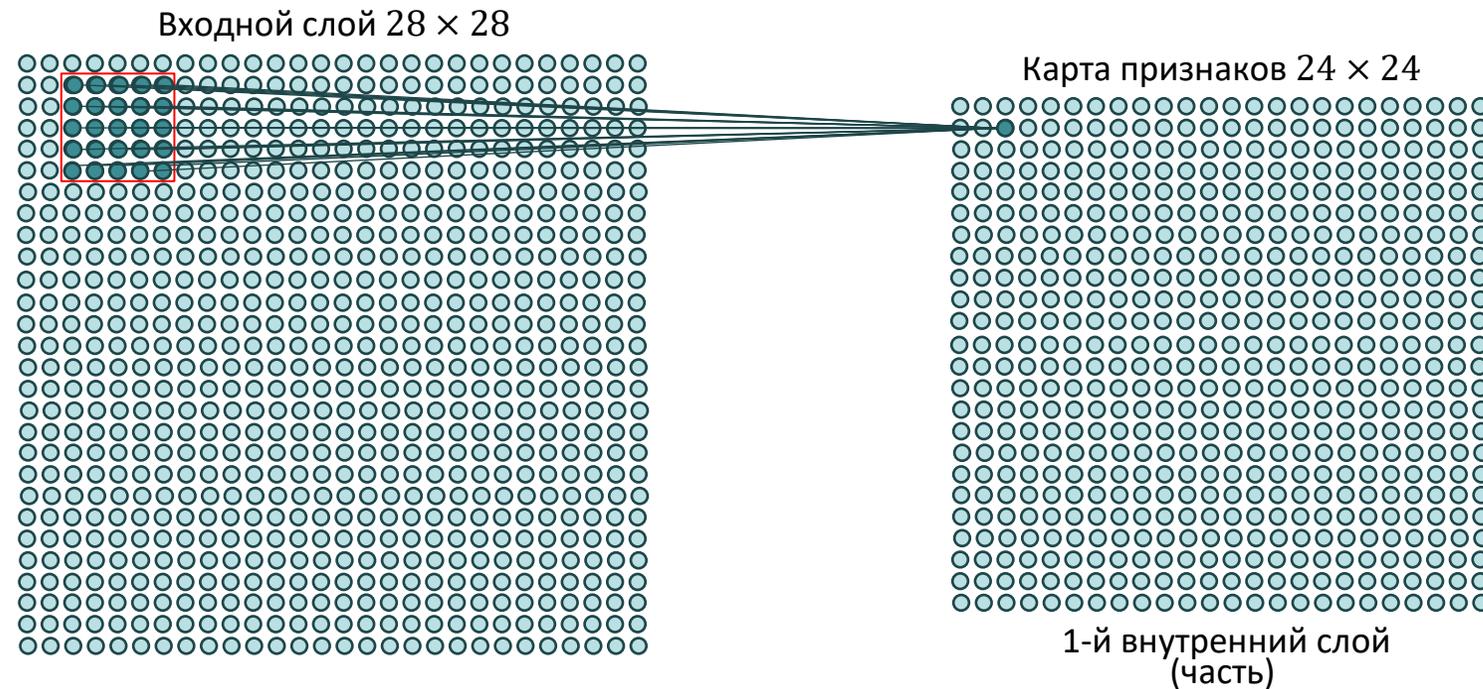
- Затем слайдер перемещается до конца влево и смещается вниз на одну позицию
- Все нейроны входного слоя, попавшие в слайдер, связываются с первым нейроном второго ряда карты признаков

# Построение 1-го внутреннего слоя



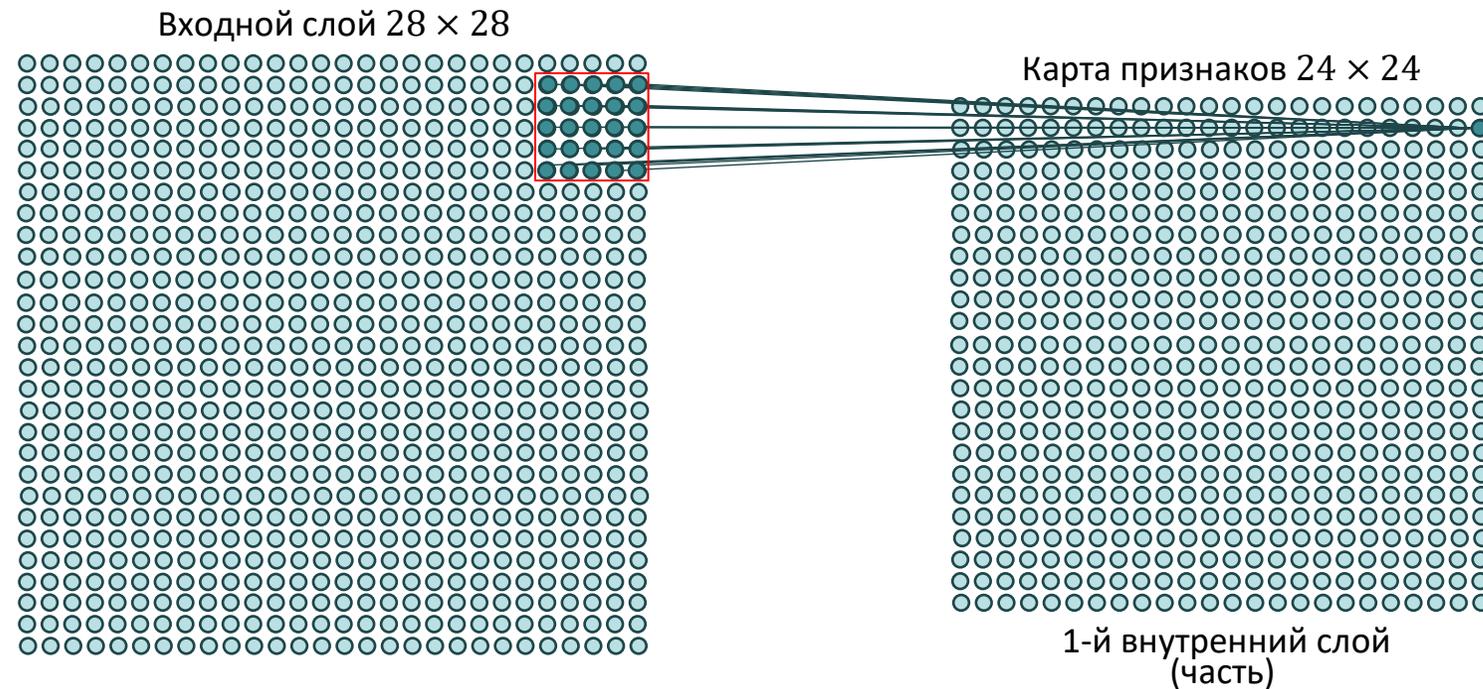
- Слайдер сдвигается на одну позицию вправо
- Все нейроны входного слоя, попавшие в слайдер, связываются со вторым нейроном второго ряда карты признаков

# Построение 1-го внутреннего слоя



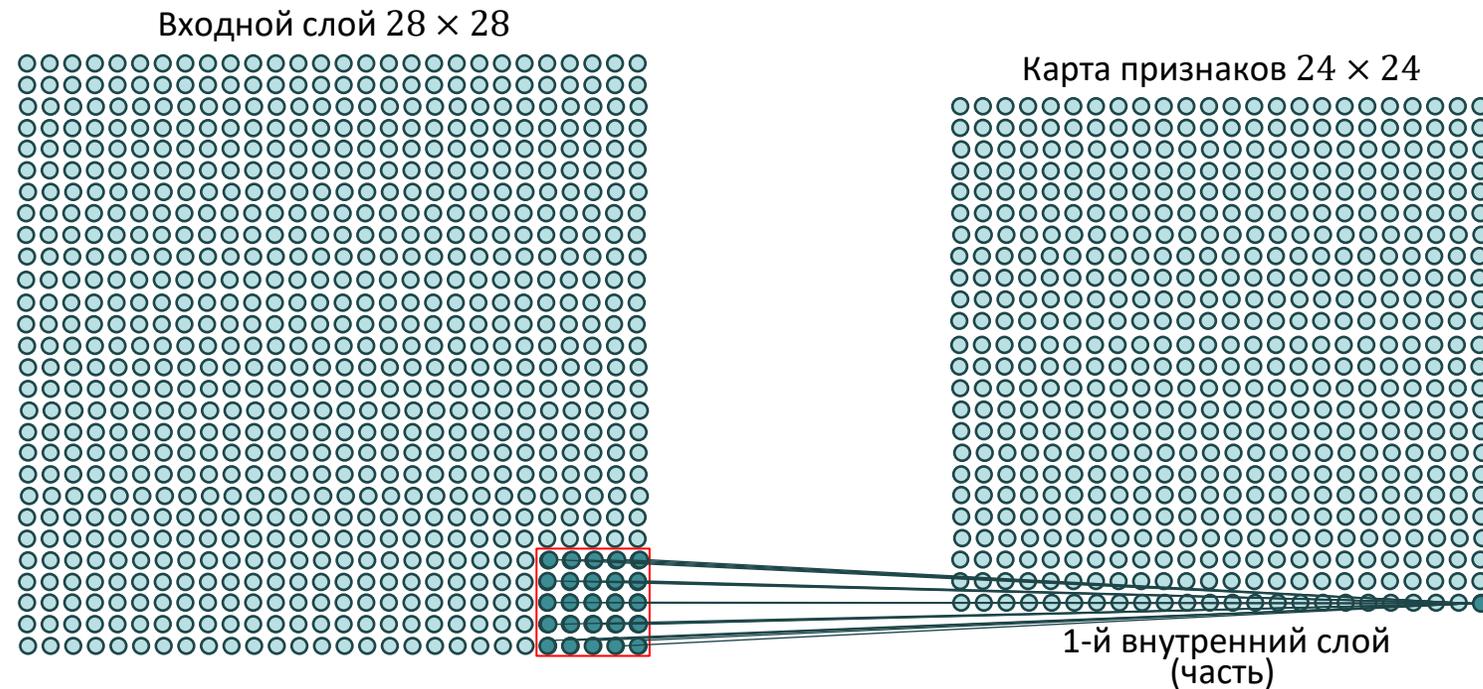
- Слайдер сдвигается еще на одну позицию вправо
- Все нейроны входного слоя, попавшие в слайдер, связываются с третьим нейроном второго ряда карты признаков

# Построение 1-го внутреннего слоя



- Таким образом устанавливаются связи для всех нейронов второго ряда карты признаков

# Построение 1-го внутреннего слоя



- Процесс перемещения слайдера продолжается, пока не будут определены связи для всех нейронов карты признаков

# Ядро (kernel) [фильтр (filter)] свертки

- Связи, формируемые слайдером в рамках одной карты признаков, имеют один и тот же набор весов  $w_{gh}$  ( $g, h = 0, \dots, 4$ )
- Все нейроны в рамках одной карты признаков имеют одинаковое смещение  $b$
- Выходной сигнал  $jk$ -того нейрона в рамках одной карты признаков вычисляется по общей формуле, называемой ядром или фильтром свертки:

$$a_{jk}^{(l+1)} = \sigma \left( b + \sum_{g=0}^4 \sum_{h=0}^4 w_{lh} a_{j+g, k+h}^{(l)} \right)$$

- Наряду с сигмоидом могут использоваться и другие функции активации

# Пример свертки

1 <sub>x1</sub>	1 <sub>x0</sub>	1 <sub>x1</sub>	0	0
0 <sub>x0</sub>	1 <sub>x1</sub>	1 <sub>x0</sub>	1	0
0 <sub>x1</sub>	0 <sub>x0</sub>	1 <sub>x1</sub>	1	1
0	0	1	1	0
0	1	1	0	0

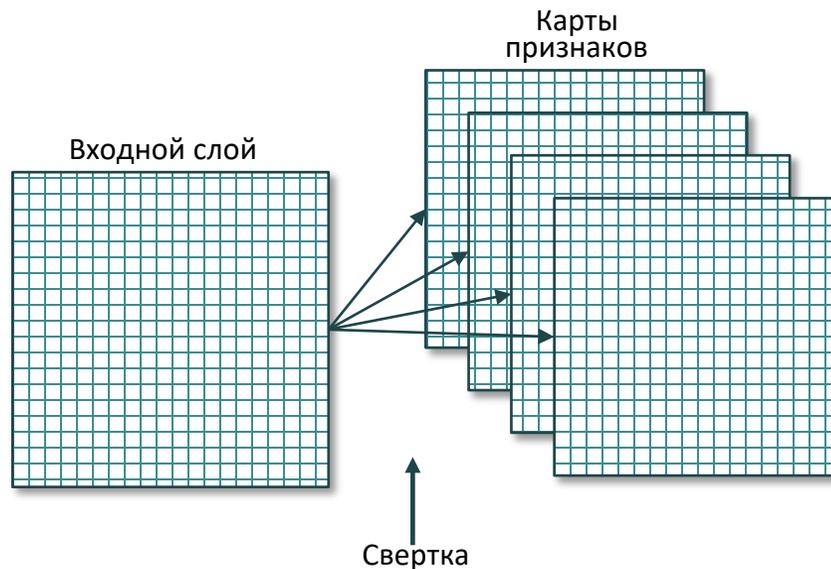
Входной слой  
(изображение)

4		

Карта  
признаков



# Множество карт признаков

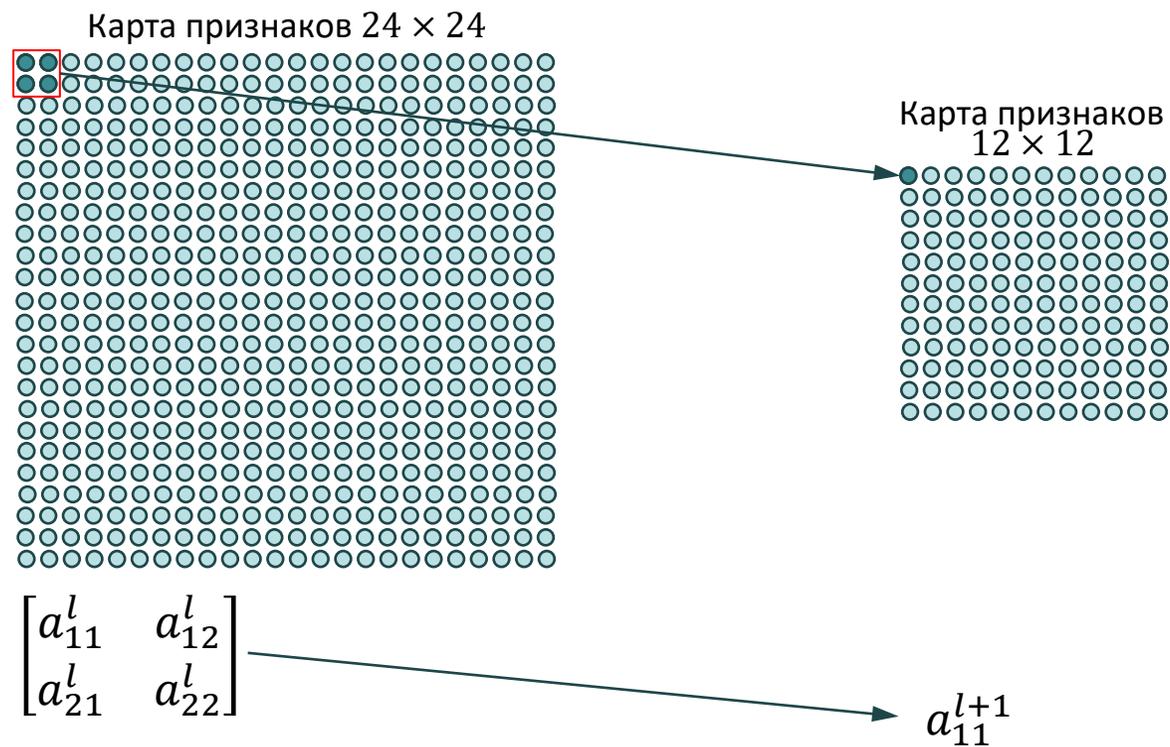


- Одна карта признаков со своим ядром распознает один признак, например «уши»
- Для распознавания другого признака, например «клюва», добавляется еще одна карта признаков, имеющая свое собственное ядро
- На практике внутренний слой может включать в себя несколько десятков и даже сотен карт признаков

# Пулинг

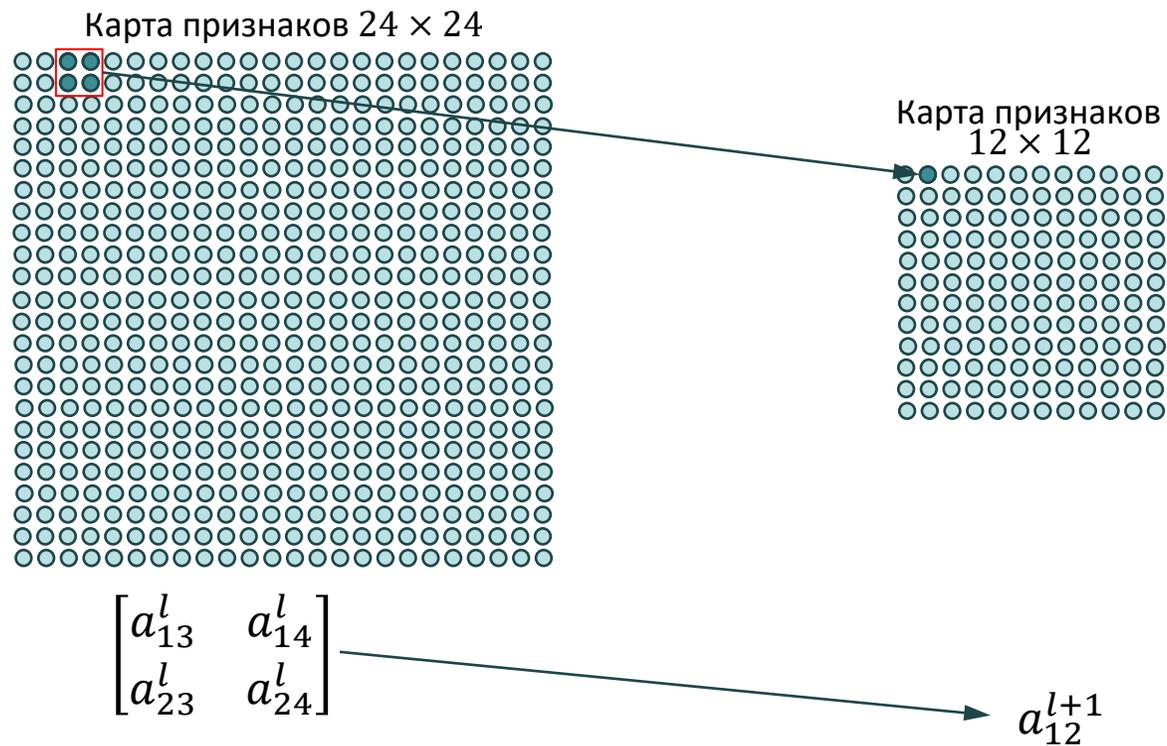
- *Pooling* (англ.) – группировка
- Слои пулинга, как правило, чередуются со слоями свертки
- Слой пулинга обобщает информацию, полученную от слоя свертки
- Пулинг «сжимает» карты признаков, полученные на предыдущем сверточном слое
- Слои пулинга не являются обязательными

# Макс-пулинг



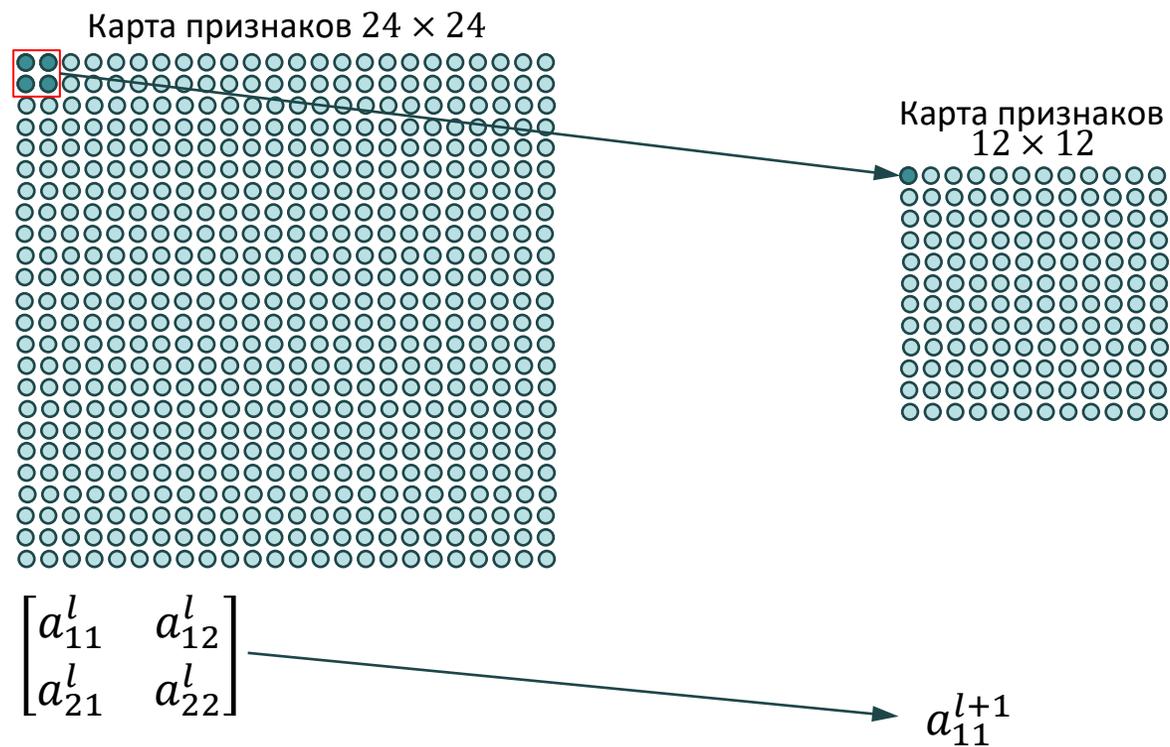
$$a_{11}^{l+1} = \max(a_{11}^l, a_{12}^l, a_{21}^l, a_{22}^l)$$

# Макс-пулинг



$$a_{12}^{l+1} = \max(a_{13}^l, a_{14}^l, a_{23}^l, a_{24}^l)$$

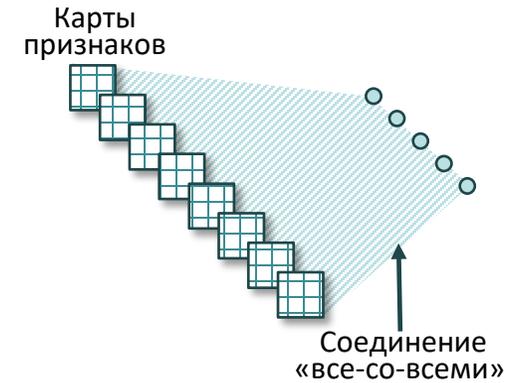
# L2-пулинг



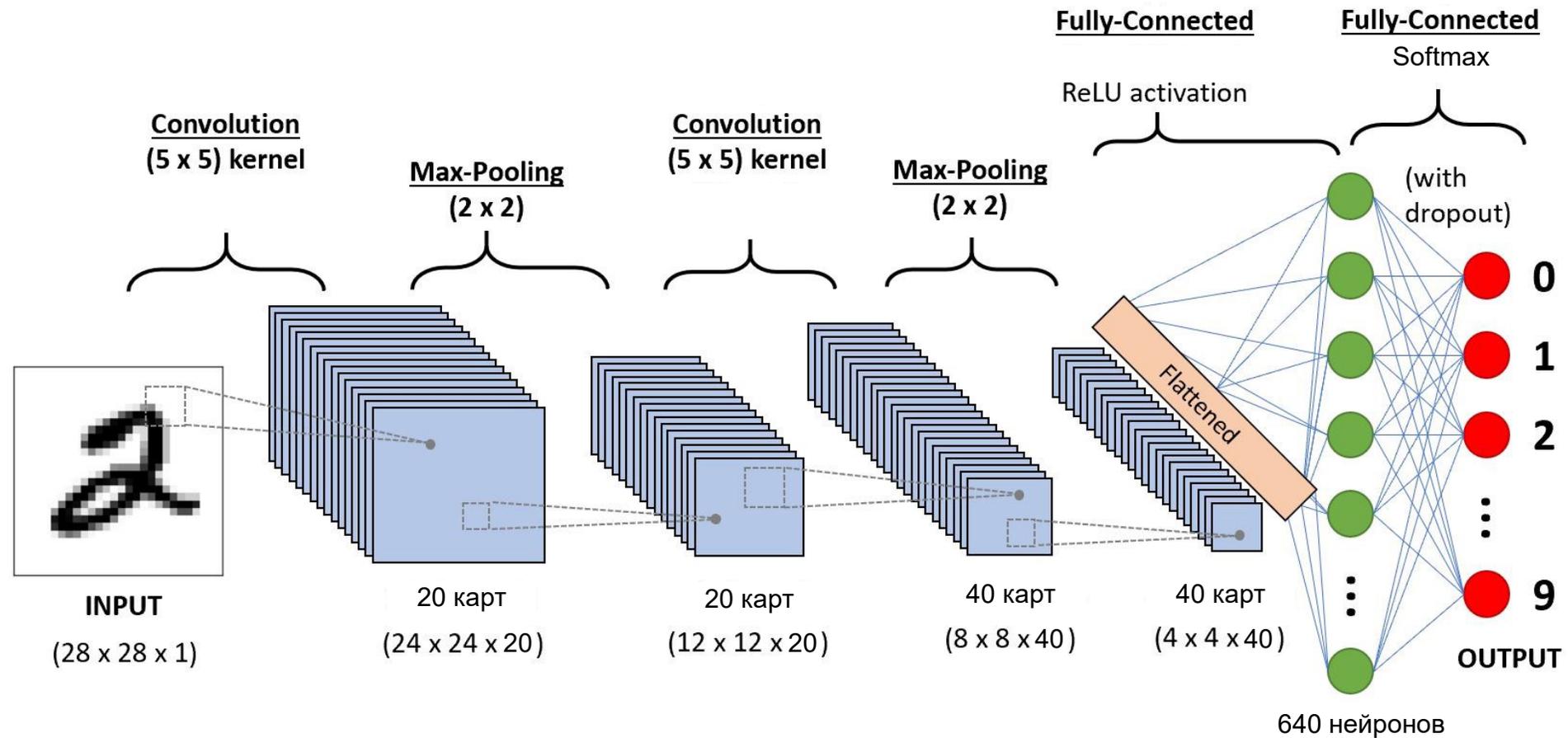
$$a_{11}^{l+1} = \sqrt{(a_{11}^l)^2 + (a_{12}^l)^2 + (a_{21}^l)^2 + (a_{22}^l)^2}$$

# Выходной слой

- Выходной слой имеет линейную структуру
- Каждый нейрон выходного слоя соединяется с каждым нейроном всех карт признаков предыдущего слоя



# Сверточная нейронная сеть для распознавания рукописных цифр



Вероятность правильного распознавания более 98%

# Особенности сверточной нейронной сети

## Преимущества

1. **Один из лучших методов по распознаванию и классификации изображений** (вероятность ошибки меньше, чем у человека)
2. **Меньшее количество настраиваемых весов по сравнению с полносвязной нейронной сетью**
3. **Удобное распараллеливание вычислений** (карты признаков обрабатываются независимо)
4. **Относительная устойчивость к повороту и сдвигу распознаваемого изображения**
5. **Обучение при помощи классического метода обратного распространения ошибки**

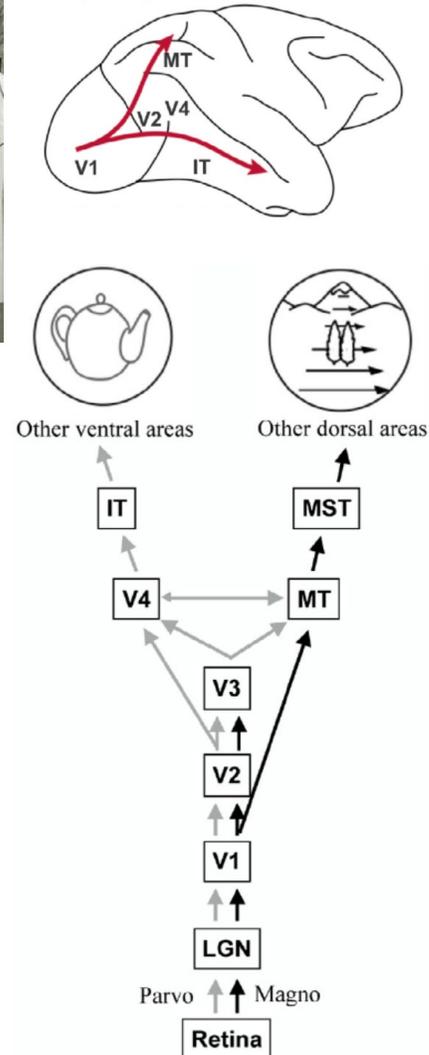
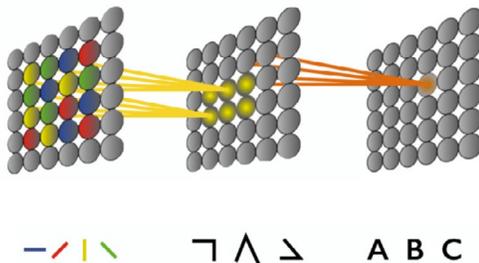
## Потенциальные проблемы

1. **Слишком много варьируемых параметров сети: количество слоев; размерность ядра свертки; шаг сдвига слайдера; степень уменьшения размера при пулинге; функция для пулинга; вид функции активации и др.** (эти параметры существенно влияют на результат, но выбираются исследователями эмпирически для каждой задачи; рекомендуется использовать **Weights & Biases: The AI Developer Platform** <https://wandb.ai/>)

# Нейробиологические основания сверточных сетей (1959-1968)



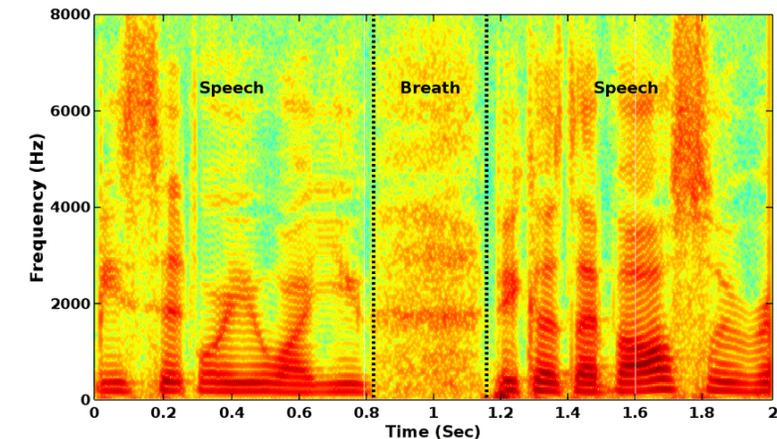
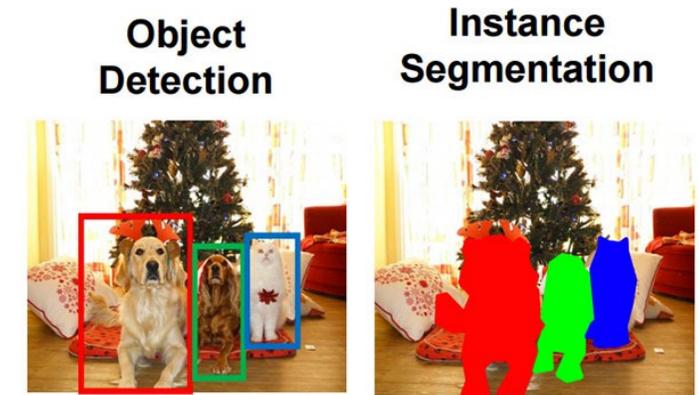
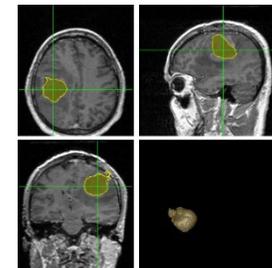
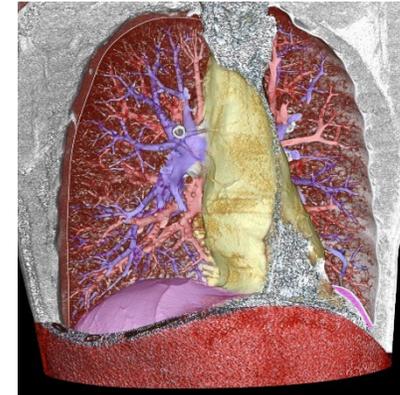
Дэвид Хьюбел и Торстен Визель  
Нобелевская премия 1981 г.



- Изображение формируется светом, который стимулирует сетчатку (retina), генерирующую электрические импульсы
- Нейроны сетчатки передают эти импульсы по зрительному нерву в зону V1, расположенную на затылке
- Зона V1 организована в виде двумерной плоской карты, повторяющей структуру сетчатки
- Определенные нейроны зоны V1 генерируют сигнал для зоны V2, если «видят» линию в соответствующей области
- Нейроны зоны V2 распознают углы и генерируют соответствующие сигналы для зоны V3, и так далее, вплоть до «бабушкиного нейрона»
- «Бабушкин нейрон» генерирует сигнал, если в поле зрения попадает родная бабушка

# Где используются сверточные нейронные сети

- Распознавание дорожных знаков
- Анализ медицинских изображений
- Распознавание лиц
- Определение, сегментация и распознавание объектов на изображениях
- Преобразование речи в текст
- Понимание естественных языков
- Анализ звуковых сигналов (через спектрограммы)



# Конец лекции 9