

Глубокое обучение. Нейронные сети для изображений

Арсланов Николай
Гребенюк Алексей
Мунхтоого Норжин

Санкт-Петербург
2021г.

Глубокое обучение — это разновидность машинного обучения на основе искусственных нейронных сетей. Процесс обучения называется глубоким, если структура искусственных нейронных сетей состоит из нескольких входных, выходных и скрытых слоев.

Наиболее популярные типы глубоких нейронных сетей:

- Многослойные нейронные сети (MNN): решение задач регрессии и классификации.
- Рекуррентные нейронные сети (RNN): прогнозирование временных рядов, обучение распознаванию рукописного ввода и распознавание естественной речи.
- Сверточные нейронные сети (CNN): распознавание видео, распознавание изображений и в системах выработки рекомендаций.
- Генеративно-сопоставительные сети (GAN): преобразование изображений в изображения.
- Преобразователи: перевод, создание текста, ответы на вопросы и формирование сводных данных текста.

Глубокое обучение достигло следующих прорывов в традиционно сложных областях машинного обучения:

- классификация изображений на уровне человека;
- распознавание речи на уровне человека;
- распознавание рукописного текста на уровне человека;
- улучшение качества машинного перевода с одного языка на другой;
- улучшение качества машинного чтения текста вслух;
- появление цифровых помощников, таких как Google Now и Amazon Alexa;
- управление автомобилем на уровне человека;
- повышение точности целевой рекламы, используемой компаниями Google, Baidu и Bing;
- повышение релевантности поиска в интернете;
- появление возможности отвечать на вопросы, заданные вслух.

Многослойные нейронные сети (пример)

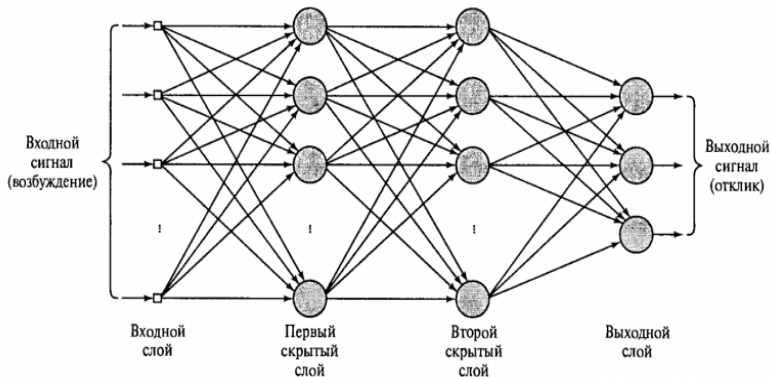


Рис.: Пример многослойной нейронной сети

Рекуррентные нейронные сети (пример)

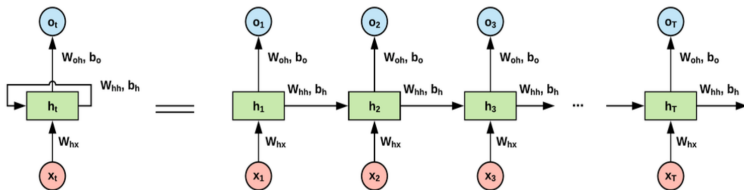


Рис.: Пример рекуррентной нейронной сети

СВЕРТОЧНАЯ ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ

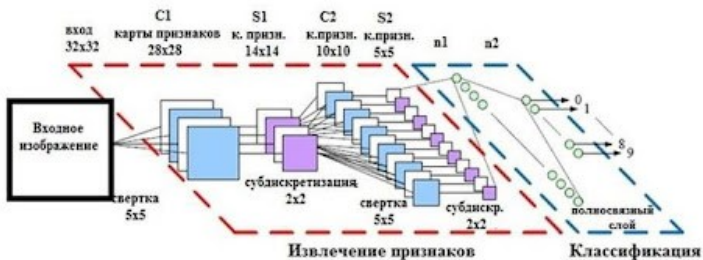


Рис.: Пример сверточной нейронной сети

Генеративно-сопоставительные сети (пример)

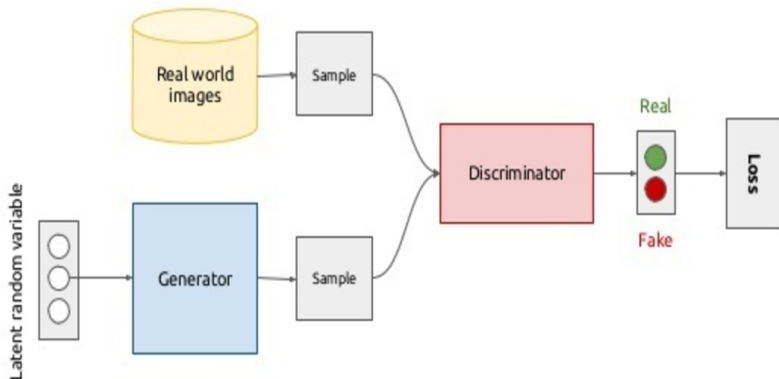


Рис.: Пример генеративно-сопоставительной сети

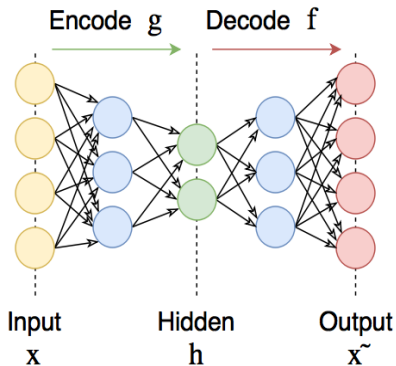


Рис.: Пример автоэнкодера

Особенности глубокого обучения:

- большая сложность глубоких нейронных сетей;
- необходимость в большом количестве данных;
- большие трудовые, ресурсные и временные затраты;
- представляет собой «Черный ящик»;
- адаптация нейронных сетей;
- широкая область применения;
- высокая производительность.

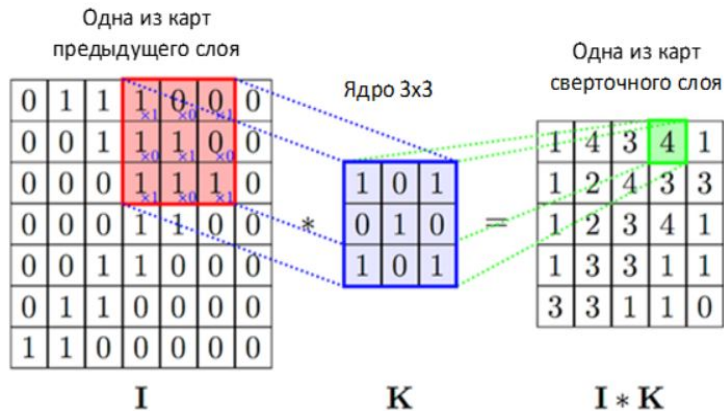
Для обработки изображений и видео наибольшей популярностью пользуются **сверточные нейронные сети**.

Основные преимущества:

- шаблоны, которые они изучают, являются инвариантными в отношении переноса;
- они могут изучать пространственные иерархии шаблонов;
- уменьшение количества параметров;
- уменьшение размерности.

Используемые слои:

- сверточные слои;
- объединяющие слои;
- полносвязные слои.



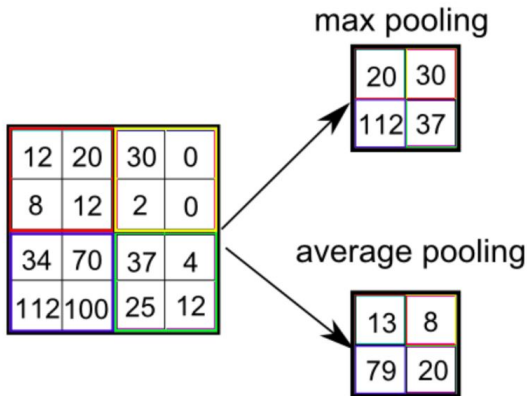
- Вход: $W_1 \times H_1 \times D_1$
- Гиперпараметры:
 - K – количество фильтров
 - F – размер фильтра
 - S – шаг свертки
 - P – заполнение нулями
- Выход: $W_2 \times H_2 \times D_2$
 - $W_2 = (W_1 - F + 2P)/S + 1$
 - $H_2 = (H_1 - F + 2P)/S + 1$
 - $D_2 = K$
- $F * F * D_1$ весов на фильтр, всего $F * F * D_1 * K$ весов
- Далее к получившимся элементам сверточного слоя применяют функцию активации. Обычно берут $\text{ReLu}(p) = \max(0, p)$;

Объединяющий слой

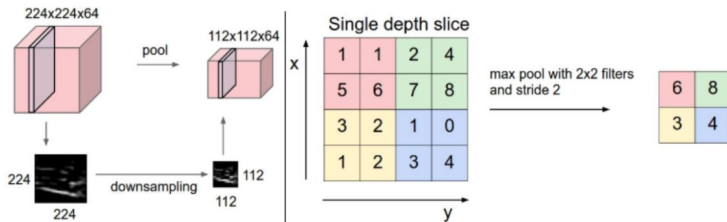
Объединяющий слой нейронов – это необучаемая свёртка с шагом $h > 1$, агрегирующая данные прямоугольной области $h \times h$:

$$y[i, j] = F(x[hi, hj], \dots, x[hi + h - 1, hj + h - 1]),$$

где F – агрегирующая функция: max, average и т.п.



- Вход: $W_1 \times H_1 \times D_1$
- Гиперпараметры:
 - F – ширина квадратного фильтра
 - S – шаг фильтра
- Выход: $W_2 \times H_2 \times D_2$:
 - $W_2 = (W_1 - F)/S + 1$
 - $H_2 = (H_1 - F)/S + 1$
 - $D_2 = D_1$



Последний из типов слоев это слой обычного многослойного персептрона. Цель слоя – классификация, моделирует сложную нелинейную функцию, оптимизируя которую, улучшается качество распознавания. Вычисление значений нейрона можно описать формулой:

$$x_j^l = \sigma(\sum_i x_i^{l-1} * w_{i,j}^{l-1} + b_j^{l-1}),$$

где

- x_j^l – карта признаков j (выход слоя l),
- $\sigma()$ – функция активации,
- b^l – коэффициент сдвига слоя l ,
- $w_{i,j}^l$ – матрица весовых коэффициентов слоя l .

Выделяются следующие функции активации:

- сигмоида: $\sigma(z) = \frac{1}{1+e^{-az}}$, $a \in \mathbb{R}$;
- гиперболический тангенс: $\sigma(z) = \frac{e^{az} - e^{-az}}{e^{az} + e^{-az}}$;
- softmax: $\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}$.

Операция свертки может быть записана так, как описано на рисунке ниже.

$$\begin{array}{|c|c|} \hline O_{11} & O_{12} \\ \hline O_{21} & O_{22} \\ \hline \end{array} = \text{Convolution} \left(\begin{array}{|c|c|c|} \hline X_{11} & X_{12} & X_{13} \\ \hline X_{21} & X_{22} & X_{23} \\ \hline X_{31} & X_{32} & X_{33} \\ \hline \end{array}, \begin{array}{|c|c|} \hline F_{11} & F_{12} \\ \hline F_{21} & F_{22} \\ \hline \end{array} \right)$$

$$O_{11} = F_{11}X_{11} + F_{12}X_{12} + F_{21}X_{21} + F_{22}X_{22}$$

$$O_{12} = F_{11}X_{12} + F_{12}X_{13} + F_{21}X_{22} + F_{22}X_{23}$$

$$O_{21} = F_{11}X_{21} + F_{12}X_{22} + F_{21}X_{31} + F_{22}X_{32}$$

$$O_{22} = F_{11}X_{22} + F_{12}X_{23} + F_{21}X_{32} + F_{22}X_{33}$$

Теперь, чтобы вычислить градиенты фильтра F относительно ошибки E , необходимо решить уравнения, которые можно записать в форме операции свертки.

$$\begin{array}{|c|c|} \hline \partial E / \partial F_{11} & \partial E / \partial F_{12} \\ \hline \partial E / \partial F_{21} & \partial E / \partial F_{22} \\ \hline \end{array} = \text{Convolution} \left(\begin{array}{|c|c|c|} \hline X_{11} & X_{12} & X_{13} \\ \hline X_{21} & X_{22} & X_{23} \\ \hline X_{31} & X_{32} & X_{33} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \partial E / \partial O_{11} & \partial E / \partial O_{12} \\ \hline \partial E / \partial O_{21} & \partial E / \partial O_{22} \\ \hline \end{array} \right)$$

Точно так же мы можем найти градиенты входной матрицы X относительно ошибки E .

$$\begin{pmatrix} \partial E / \partial X_{11} & \partial E / \partial X_{12} & \partial E / \partial X_{13} \\ \partial E / \partial X_{21} & \partial E / \partial X_{22} & \partial E / \partial X_{23} \\ \partial E / \partial X_{31} & \partial E / \partial X_{32} & \partial E / \partial X_{33} \end{pmatrix} = \text{Convolution} \left(\begin{pmatrix} \partial E / \partial O_{11} & \partial E / \partial O_{12} \\ \partial E / \partial O_{21} & \partial E / \partial O_{22} \end{pmatrix}, \begin{pmatrix} F_{22} & F_{21} \\ F_{12} & F_{11} \end{pmatrix} \right)$$