

1. Что такое поканальная раздельная свертка?

Входное изображение формы (w, h, 3) можно разделить на три карты формы (w, h) – это будут три матрицы, каждая соответствует одному каналу цвета. Потом можно применять операцию свертки к этим картам по отдельности.

2. Приведите пример задач решаемых обучением с учителем

Классификация объектов и распознавание образов, регрессия и аппроксимация, прогнозирование.

3. Какие функции активации можно использовать в сверточных нейронных сетях?

В сверточных сетях много слоев, поэтому затухание градиента может плохо повлиять на обучение методом обратного распространения ошибки. Чтобы этого не случилось, нужно выбрать функцию активации с производной, не много меньшей 1 по модулю. На практике могут быть использованы разные функции активации (в т.ч. те, чья производная $\ll 1$ в окрестностях -1, 0 или 1).

Вот самые распространенные из них:

Название функции	Выражение
Линейная (linear)	$\varphi(x) = ax + b$
Логистическая (logit)	$\varphi(x) = \log(1 + e^x)$
Сигмоида (sigmoid)	$\varphi(x) = \frac{1}{1 + e^x}$
Гиперболический тангенс (th)	$\varphi(x) = a \operatorname{th} bx$
Пороговая (Хевисайда, hs)	$\varphi(x) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases}$
Линейная спрямленная (rectified linear, relu)	$\varphi(x) = \max(0, x)$

4. Для чего нужен слой Flatten?

Чтобы привести тензор к рангу 1, то есть к форме вектора, не меняя количество элементов.

5. Чем чревато использование ненормированных данных в сверточных сетях?

Рассмотрим пример. Пусть есть два числовых признака: первый принимает значения из отрезка [0; 1], а второй из [0; 10^{100500}]. Если расставить малые случайные веса, то выйдет так, что второй признак имеет очень больше влияние на результат, а первый по сравнению с ним незначителен. Но первый признак может оказаться важным, и тогда данные потеряются. Это усложняет обучение, потому что теперь нужны веса, отличающиеся на много порядков. Но можно нормировать данные, тем самым упростив обучение.