

1. Что такое shortcut соединение и для чего оно нужно?

Shortcut соединение позволяет пропустить несколько слоев.

Было замечено, что при обучении глубоких сетей с использованием простого стекания слоев качество растет до определенного количества слоев. Нейросеть может аппроксимировать почти любую функцию, например некоторую сложную функцию  $H(x)$ . Тогда справедливо, что такая сеть легко выучит residual-функцию (остаточную функцию):  $F(x) = H(x) - x$ . Очевидно, что наша первоначальная целевая функция будет равна  $H(x) = F(x) + x$ . Проблема деградации подразумевает, что сложная нелинейная функция  $F(x)$ , полученная стеканием нескольких слоев, должна выучить тождественное преобразование, в случае если на предыдущих слоях был достигнут предел качества. Но этого не происходит по каким-то причинам, Тогда мы поможем ей добавить shortcut-соединение, и, возможно, оптимизатору будет легче сделать все веса близкими к нулю, нежели создавать тождественное преобразование.

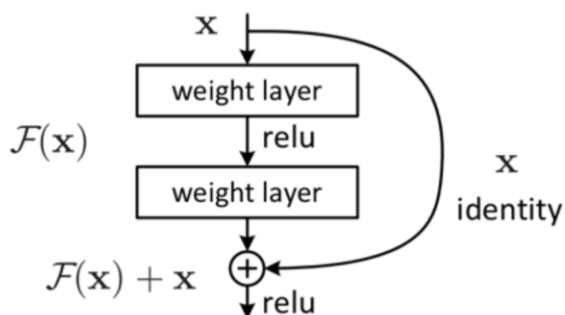


Figure 2. Residual learning: a building block.

$\mathcal{H}(\mathbf{x})$  is the true function we want to learn

Let's pretend we want to learn

$$\mathcal{F}(\mathbf{x}) := \mathcal{H}(\mathbf{x}) - \mathbf{x}$$

instead.

The original function is then

$$\mathcal{F}(\mathbf{x}) + \mathbf{x}$$

2. Применяется ли слой Dropout, если вызывается метод predict? Ответ обоснуйте.

Данный слой не применяется. Dropout применяют, когда происходит переобучение сети. Исключение нейронов «создает» новую сеть. Для обучения.

### 3. Что такое блок GRU?

GRU – управляемый рекуррентный нейрон. Используется в вариации сети с долгой краткосрочной памятью. У него два фильтра. Фильтр обновления определяет, сколько информации останется от прошлого состояния и сколько будет взято из предыдущего слоя. Фильтр сброса работает примерно как забывающий фильтр.

### 4. Как обычная свертка раскладывается на поканальную свертку?

Сначала входной тензор сворачивается ядром размерности  $1 \times 1 \times [\text{кол-во каналов входного тензора}]$ . Далее происходит свертка  $3 \times 3$  по каждому каналу. После происходит свертка  $1 \times 1$ .