1. При работе сверточного слоя находятся вхождения некоторого паттерна на изображение, после чего строится карта откликов шаблона, по которому искались вхождения, причем каждый следующий сверточный слой работает с картой отклика предыдущего.   
   Для нахождения, например, круга, достаточно взять несколько сверток, которые будут реагировать на разные края круга и большую свертку, которая опирается на карту откликов предыдущих.
2. Сходимость у сигмоидальной функции гораздо ниже, чем у ReLU («слабые» градиенты почти везде), что уменьшает скорость обучения.
3. Maxout позволяет динамически изменять функцию активации, исходя из банка фильтров, стоящих перед ней.
4. Входное изображение -> 64x64x3 (выход слоя)  
   Conv, 96 фильтров 5х5, шаг 1 -> 64х64х96  
   Conv, 96 фильтров 1х1, шаг 1 -> 64х64х96  
   Max pooling 5x5, шаг 4, dropout 50% -> 16х16х96  
   Conv, 96 фильтров 3х3, шаг 1 -> 16х16х96  
   Conv, 120 фильтров 1х1, шаг 1 -> 16х16х120  
   Max pooling 3x3, шаг 2, dropout 50% -> 8x8x120  
   Conv, 240 фильтров 3х3, шаг 1 -> 8х8х240  
   Conv, 60 фильтров 1х1, шаг 1 -> 8х8х60  
   Conv, 20 фильтров 1х1, шаг 1 -> 8х8х20  
   Average pooling 8x8 -> 1x1x20  
   Softmax, 20 выходов -> 20  
     
   Функция активации – ReLU, все слои случайно инициализированы по гауссу с центром в 0.
5. Inception – усложнение архитектуры сети Крижевского, где к каждому фильтру добавляются свертки 1х1 – 5х5 и max pooling.
6. Большая часть окон отсеивается маленькими сетями, а значит сложным сетям придется обрабатывать меньше данных.
7. С помощью special transform слоя можно создавать нейросети, которые будут устойчивы к аффинным преобразованием без размножения обучающих данных. Этот слой обучается с помощью алгоритма распространения обычной ошибки, так что его можно использовать внутри любой нейросети. Количество свободных параметров при этом у него тоже не велико.