1. The motive to introduce DenseNet is it connects each layer to every other layer in a feed-forward fashion. DenseNet has several compelling advantages. They alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse and substantially reduce the number of parameters. Moreover, DenseNets obtain significant improvements over the state-of-the-art on most of them, whilst requiring less computation to achieve high performance. Also because of the dense connectivity pattern authors have approached DenseNet. Authors approaching DenseNet architecture explicitly differentiates between information that is added to the network and information that is preserved.

2. The datasets used by the authors for experiment the DenseNet are;

CIFAR-10, CIFAR-100, SVHN, ImageNet

3. Yes. There is an overfitting issue that arises in the DenseNet. Authors observed that on the datasets without data augmentation, the improvements of DenseNet architectures over prior work are particularly pronounced. On CIFAR-10(C10), the improvement denotes a 29% relative reduction in error from 7.33% to 5.19%. On CIFAR-100(C100) the reduction is about 30% from 28.20% to 19.64%. In their experiments, authors observed potential over lifting in a single setting: on C10, a 4\*growth of parameters produced by increasing k=12 to k=24 lead to modest increase in error from 5.77% to 5.83%. The DenseNet-BC bottleneck and compression layers appear to be an effective way to counter this trend.

4. Highway networks were the first architecture that provided a means to effectively train end to end networks with more than 100 layers. Using bypassing paths along with gating units, highway networks with hundreds of layers can be optimized without difficulty. Where else, DenseNet connects each layer to every other layer. DenseNet layers are very narrow, adding only a small set of feature maps to the collective knowledge of the network and keeping the remaining feature maps unchanged. And the final classifier makes a decision based on all feature-maps in the network. Cascade networks failed because they are effective on small datasets that are only scaled to networks with a few hundred parameters. In this situation DenseNet utilizes multi-level features in CNNs through skip-connections that have been found to be effective for various vision tasks. It derived a purely theoretical framework for networks with cross-layer connections.