1. The reason for introducing DenseNet is that it connects each layer to each other layer in a feed-forward manner. DenseNet offers several compelling benefits. They solve the vanishing-gradient problem, improve feature propagation, promote feature reuse, and significantly reduce the number of parameters. Furthermore, DenseNets outperform the state-of-the-art on the majority of them while requiring less computation to achieve high performance. Authors have also approached DenseNet due to the dense connectivity pattern. The authors' approach to DenseNet architecture explicitly distinguishes between information added to the network and information preserved.

2. The authors used the following datasets to test the DenseNet:

SVHN, ImageNet, CIFAR-10, CIFAR-100

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

4. Highway networks were the first architecture to allow for the effective training of end-to-end networks with more than 100 layers. Highway networks with hundreds of layers can be easily optimized using bypassing paths and gating units. DenseNet connects every layer to every other layer. DenseNet layers are extremely narrow, adding only a small set of feature maps to the network's collective knowledge while leaving the remaining feature maps unchanged. And the final classifier makes a decision based on all of the network's feature-maps. Cascade networks failed because they work best on small datasets and can only be scaled to networks with a few hundred parameters. In this case, DenseNet employs multi-level features in CNNs via skip-connections, which have been shown to be effective for a variety of vision tasks. It developed a purely theoretical framework for cross-layer networks.