1. What is the motive to introduce DenseNet ?How it is originated?

= From recent studies it was found that, if CNN models contain shorter connections between layers close to the input and those close to the output, then they can be substantially deeper, more accurate, and efficient to train the model. Embracing this observation, the authors introduced an architecture which connects all the layers in a feed-forward fashion. Moreover, information about the input or grads to vanish and “wash out” in deeper CNNs since it is passed through many layers before reaching the end or beginning of the network. Several publications addressed this or related problems and came up with different approaches. But they all share a key characteristic where they create shorter paths from early layers to later layers. In this paper, the authors took this insight and proposed an architecture that ensures maximum information flow between layers in the network by connecting all layers directly with each other. Each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers to preserve the feed-forward nature of the network. The architecture consisting of L layers have L(L+1)/2 direct connections whereas traditional CNNs have L connections. Because of its dense connectivity pattern, the authors refer to this approach as Dense Convolutional Network (DenseNet). DenseNets have several compelling advantages:

1. they alleviate the vanishing-gradient problem
2. strengthen feature propagation
3. encourage feature reuse, and
4. substantially reduce the number of parameters

2. What are the datasets used by the authors for experiment the DenseNet?

= The authors have used the following datasets:

1. The two CIFAR datasets, CIFAR-10 (C10) and CIFAR-100 (C100)
2. The Street View House Numbers (SVHN) dataset
3. The ILSVRC 2021 classification dataset called ImageNet

3. Is there any overfitting issue arises in the DenseNet? If yes how the authors suggests to mitigate it?

= The authors have observed that the dense connections in their model have a regularizing effect that reduces overfitting on tasks with smaller training set sizes. Moreover, due to the efficient use of parameters, DenseNets are less prone to overfitting. The authors have observed that on the datasets without data augmentation, the improvements of DenseNet architectures over prior work are particularly pronounced. In their experiments, the authors have observed potential overfitting in a single setting: on C10, a 4 times growth of parameters produced by increasing the value of k from 12 to 24 lead to a 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. Moreover, the authors have found that DenseNets tend to yield consistent improvement in accuracy with a growing number of parameters, without any signs of performance degradation or overfitting.

4. State specific difference between highway networks and densenet ? Where cascade networks failed and how densenet comes in handy?

= Highway Networks is one of the first deep CNNs that provided a means to effectively train end-to-end networks with more than 100 layers. This deep CNN can be optimized using bypassing paths along with gating units. The bypassing paths are presumed to be the key factor that eases the training of these very deep networks. On the other hand, DenseNet uses a dense connectivity pattern where it connects all the layers with matching feature-map sizes directly with each other to ensure maximum information flow bbetween the layers in the network. This architecture never combines features through summation before they are passed into a layer; instead, it combines features by concatenating them. This introduces L(L+1)/2 connections in an L-layer network, instead of just L, as in traditional architectures. As a result, it requires fewer parameters than traditional convolutional networks, as there is no need to relearn redundant feature-maps.

Cascade structure is similar to DenseNet but it is only effective on small datasets. The cascade structure approach only scales to networks with a few hundred parameters. On the other hand, DenseNets require very few parameters due to its dense connectivity pattern and the fact that all layers are connected to each other.