Ans 1 Transfer learning - This technique is introduced to control the number of of training parameters.

Because in many models we are toying to reduced the training parameters, as if it increases then it cause a lot of problem of overfitting, gradient - caplosion, vanishing gradient also limitation of resources will come as a huge data source can not be easily train from our local PC.

as Alex Net has 60M parameters
Vgg has - 138M Parameters

→ As the number of parameters increased our resource utilization also increased which is not easily handled by a single Machine → we need to go with distributed computing to solve this but this will occur more cost which is also a problem

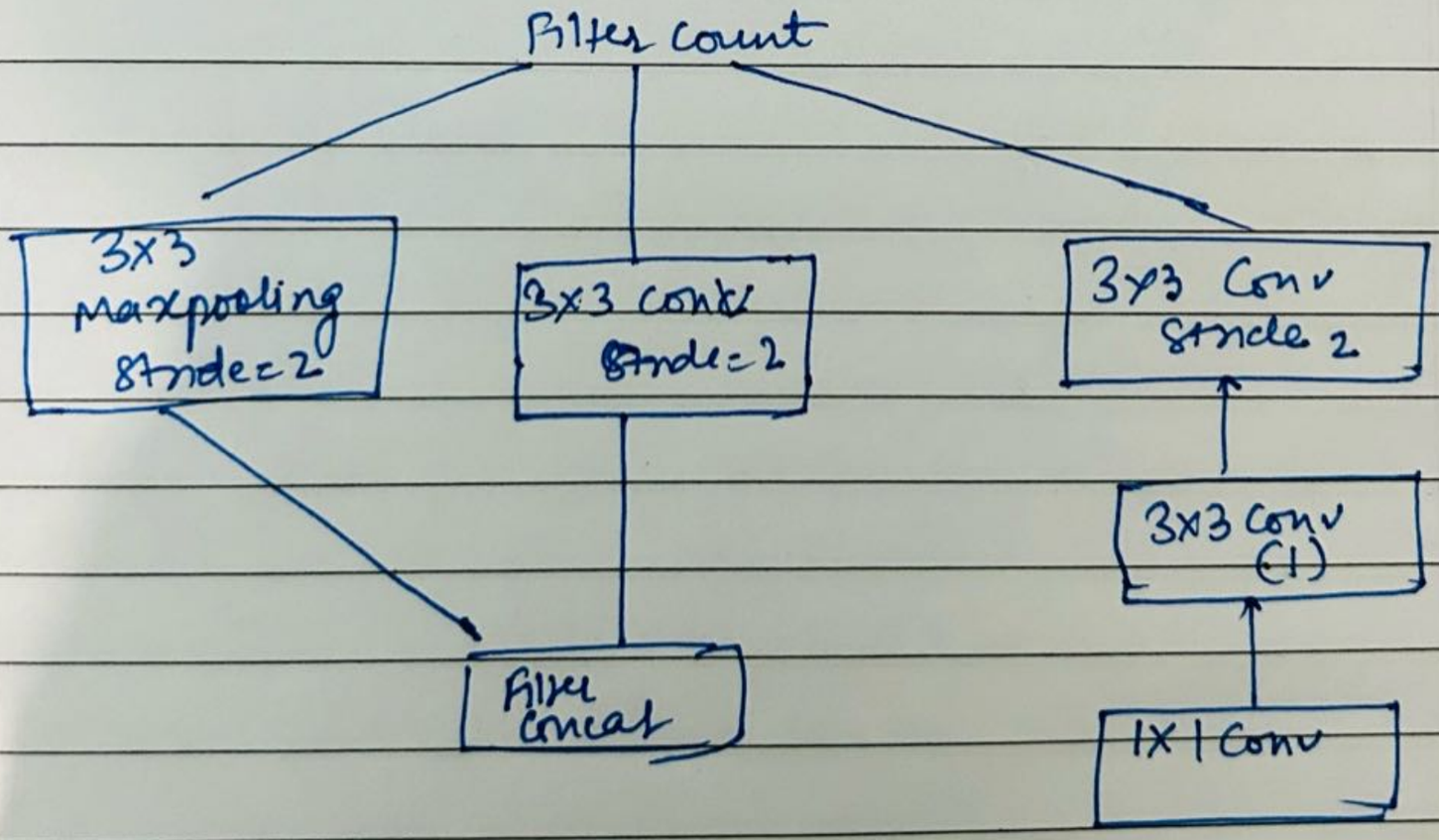
We basically freeze some layer to avoid processing again like in AlexNet the

Ans2 Auxillary Branches - These are basically used add some branches to our network architecture to assist in training. ~~Disadvantages~~ Main function of these branches are used to calculate loss we get after the epoch completes

This is useful to remove problems like Network degradation problem for Vanishing gradient & gradient explosion.

→ Dimension Reduction → It is used to reduce the dimensionality of the input data or feature maps, this helps in performance improvement as less resources being used since redundant information is removed. In CNN if we used small size filter with less channels this helps in the reduction of trainable parameters. If less number of parameters are these less resources being used

Reduction Block - It is used to reduce the size before using max pooling we can used reduction Block to reduce size.



This is Reduction block.