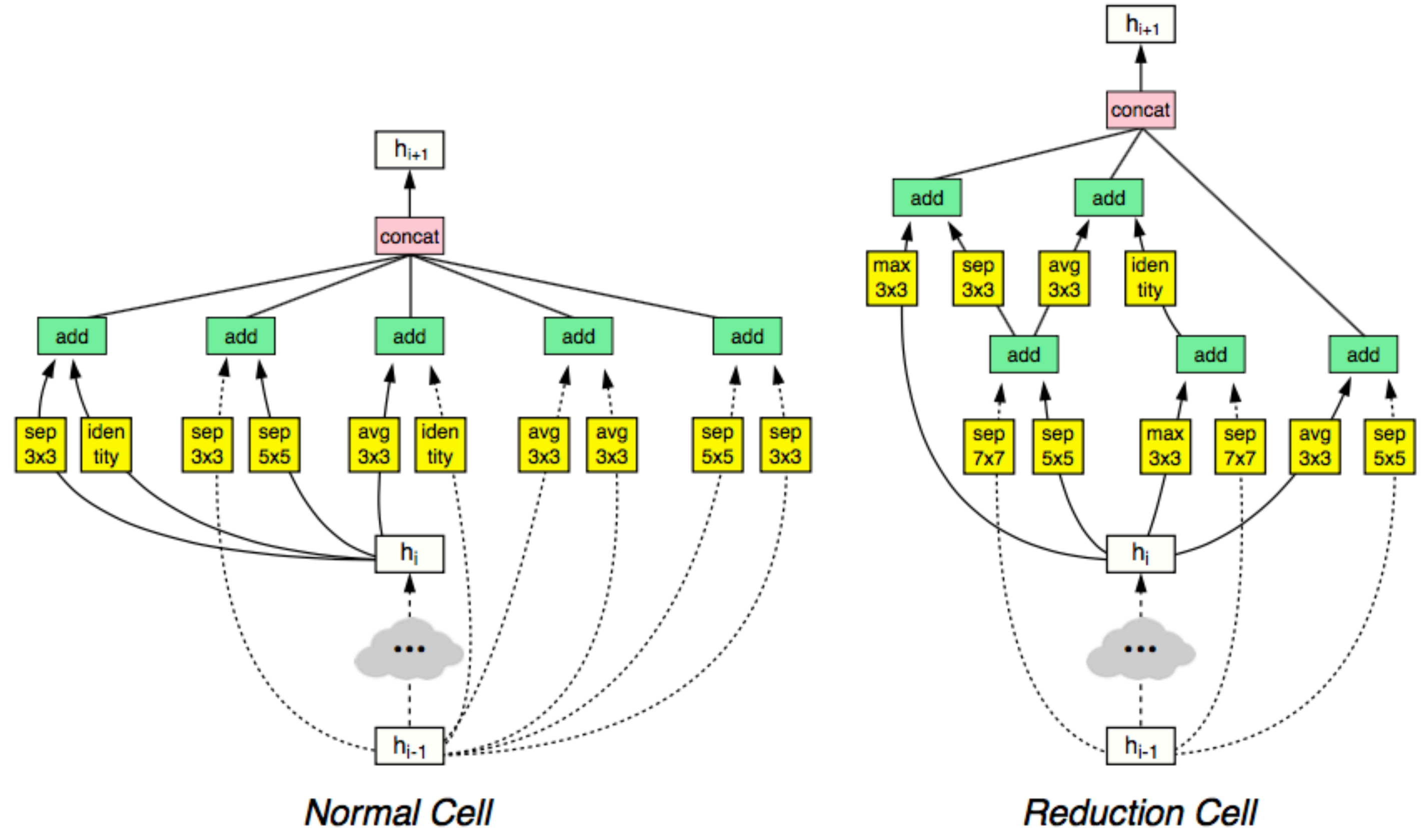
2.1.1 [NasNet] Notes by YK

[Learning Transferable Architectures for Scalable Image Recognition]

a. Main Contributions

This work is one of the most important for me. To be honest, many following nice networks searched are son of NasNet, including ENAS, PNAS.
The main contributions of NasNet are:
1) Reduce search space from the whole convolutional architectures to cell structures.
2) Show that the best convolutional cell searched on CIFAR-10 is able to be transfered to ImageNet.
3) Cells searched is surprisingly less in FLOPs and parameters than common networks designed matually.



b. Key Points

(1) **The training procedure of this work:**
1) a controller(RNN) samples child networks with different architectures.
2) the child networks are trained to convergence to obtain some **accuracies** on a held-out validation set.
3) the resulting **accuracies** are used as reward to update the controller with **policy gradient** so that better architectures would be generated over time.

(2) **Normal Cell & Reduction Cell**
One inspiration to reduce search space is the **recognition** that architecture engineering with CNNs often identifies **repeated motifs** consisting of combinations of convolutional filter banks, nonlinearities and **a prudent selection of connections**.

Based on this knowledge, two types of convolutional cells are proposed: 1) Normal Cells with a feature map of same dimension as output; 2) Reduction Cells with a feature map whose height and width are reduced by a factor of two.

These two types of cells then can be stacked in series to handle inputs of **various/arbitrary spatial dimensions and filter depth**.

(3) **Search Space**
• identity
• 3x3 depthwise-separable conv
• 5x5 depthwise-seperable conv
• 7x7 depthwise-separable conv
• 3x3 avg pooling
• 3x3 max pooling
• 3x3 dilated convolution
• ~~1x1 convolution~~
• ~~3x3 convolution~~
• ~~1x3 then 3x1 convolution~~
• ~~1x7 then 7x1 convolution~~
• ~~5x5 max pooling~~
• ~~7x7 max pooling~~

The erased phases never appear in results (The RNN controller did not pick them up for some reasons).

c. Experiments and Results

(1) **Advantages**
1) large scale
Accuracy >= DenseNet,InceptionV2,ResNet…
num of Params,Mult-Adds< DenseNet,InceptionV2,ResNet …
2) small scale
Accuracy > MobileNet and ShuffleNet
num of Params,Mult-Adds≈DenseNet,InceptionV2,ResNet …
3) transferable
architecture searched from CIFAR-10, good performance on Imagenet and Detection

(2) **Drawbacks**
Computational Resources : Still require 450 GPUs 4 days

Training each single model from scratch is wasting!
Topology: Block structure is too fixed, which is worth exploring

d. English Writing

The image features learned from image classification are generically useful and can be **transferred to other computer vision problems**.

Our approach **makes use of** the recently proposed ...

Additionally, by simply **varying** the number of the convolutional cells and number of filters in the convolutional cells, we can create convolutional architectures **with different computational demands**.

To ensure that the shapes always match in convolutional cells, 1x1 convolutions are inserted **as necessary**.