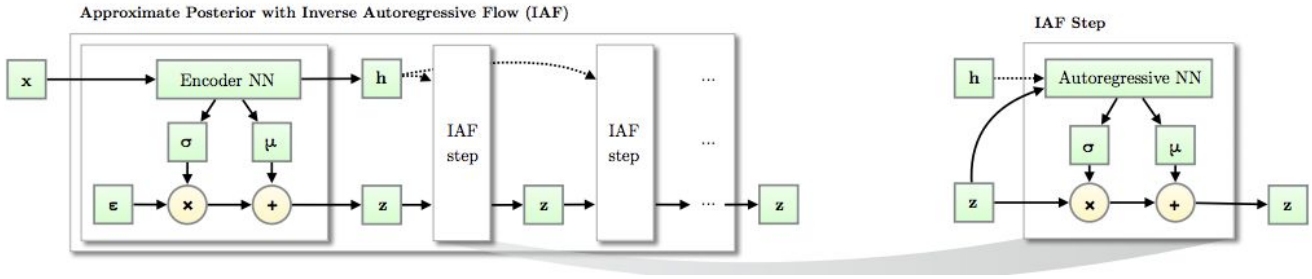
Parallel PixelCNN/WaveNet etc.

The review document my understanding and confusions. For sure I may have lots of misunderstandings, please help point out to faster the project, and refer to right resources.

1, Regarding IAF.



My understanding:

1), All h are the same h form the output of the Encoder NN in the above figure.

2), Each z, \sigma, \mu is different in each step as having indices of time t.
Each \sigma and \mu have to be learned by a different Autoregressive NN, such as WaveNet.
The learning of \sigma and \mu i.e. inference of z has to be in serial, thus slow.

3), However, once the above model is learned, the generation of x is parallel thus fast in the inverse direction.

Recall autoregressive model as follows: fast in training, slow in generation.

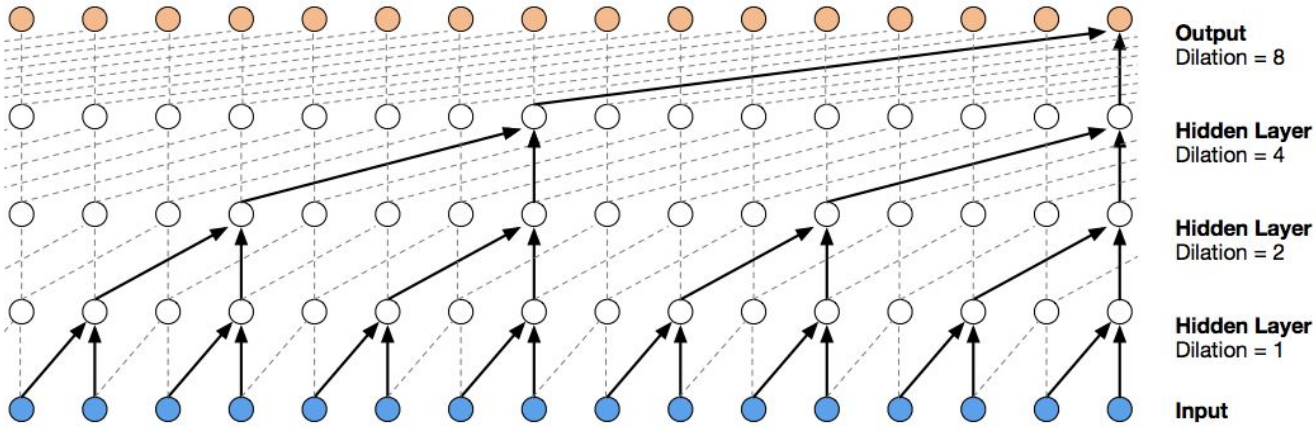
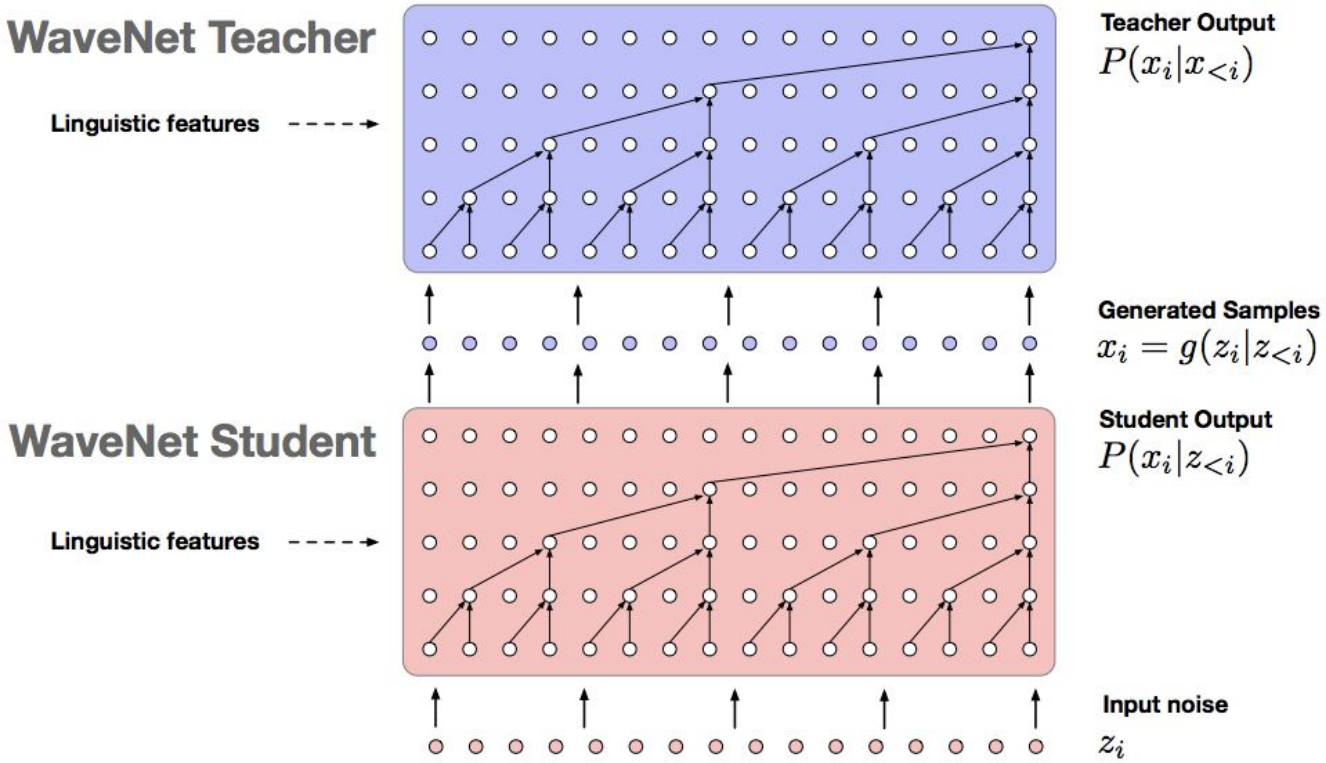


Figure 1: Visualisation of a WaveNet stack and its receptive field [27].

2, Regarding Implementing a Parallel WaveNet (WaveNet Student) with IAF .

3, Regarding Distilling (Learning with WaveNet Teacher)

Parallel WaveNet combine the parallel training of autoregressive model and parallel generation of IAF as following.



My understanding:

WaveNet Teacher is a trained autoregressive WaveNet.

WaveNet Student is IAF

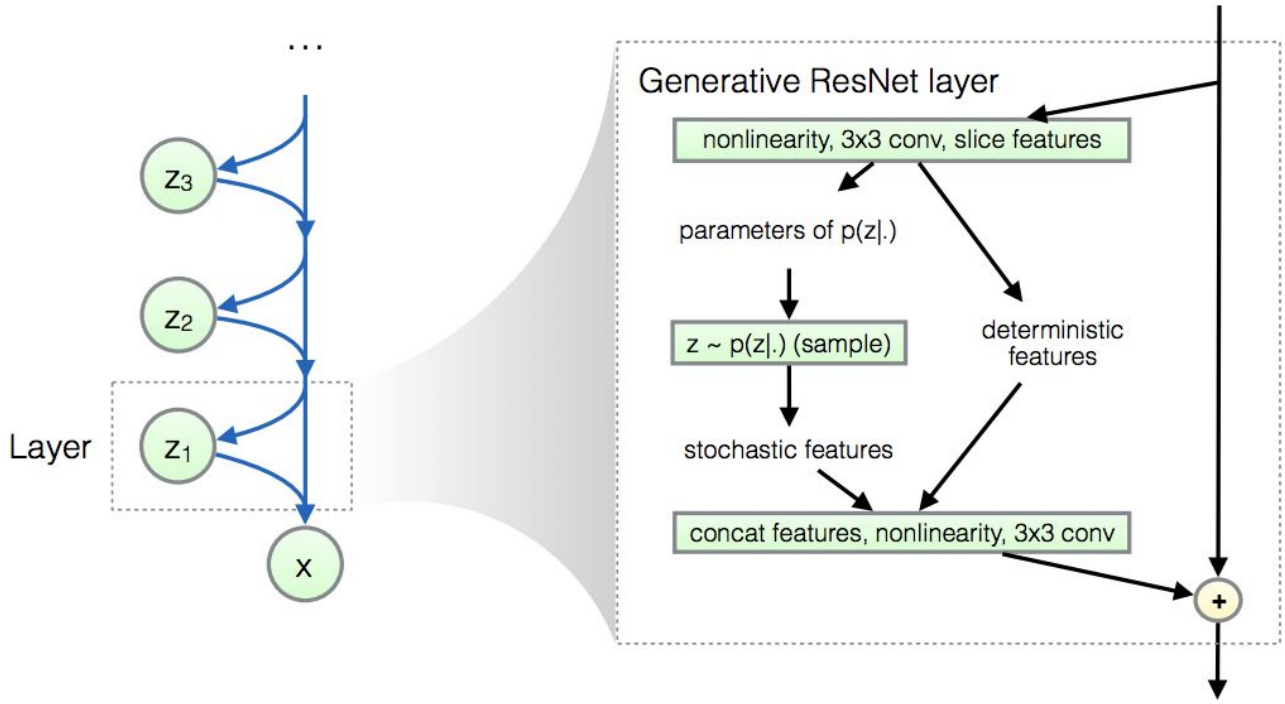
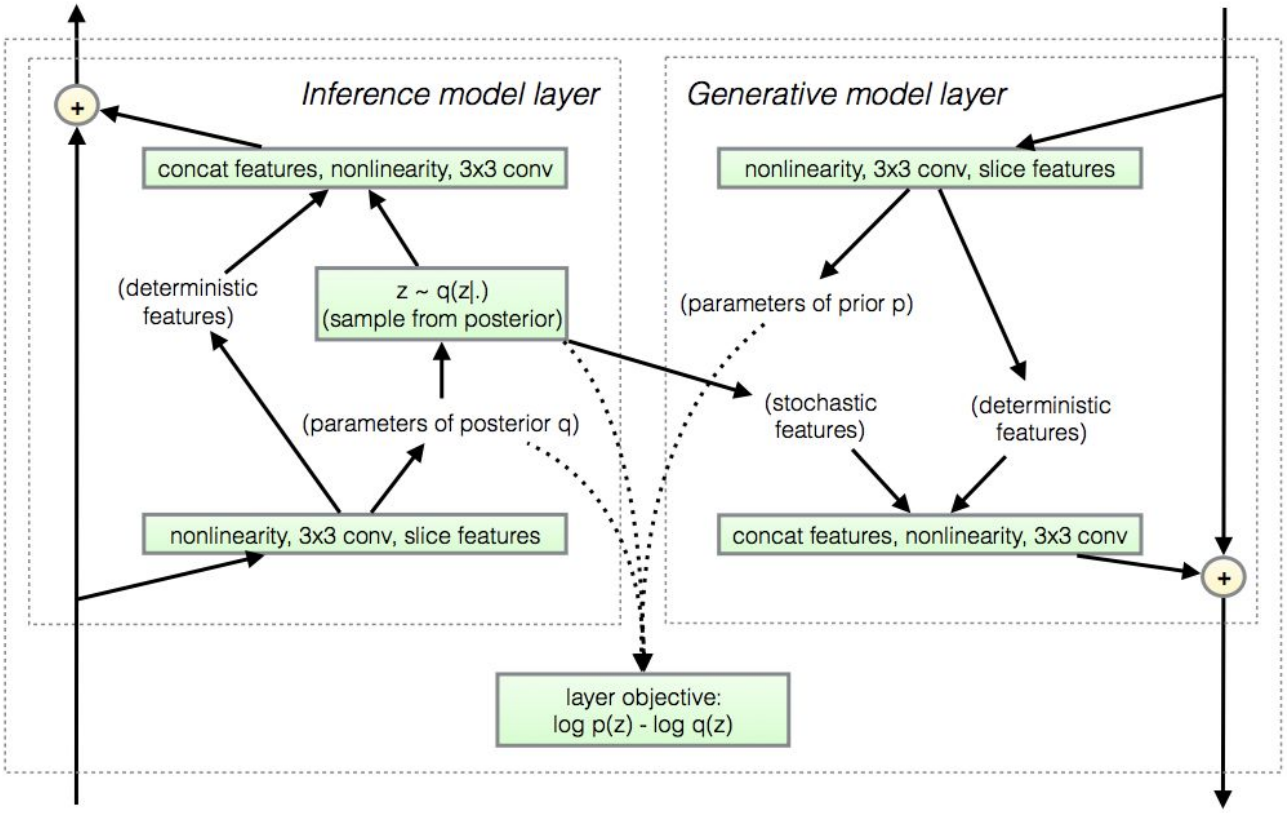
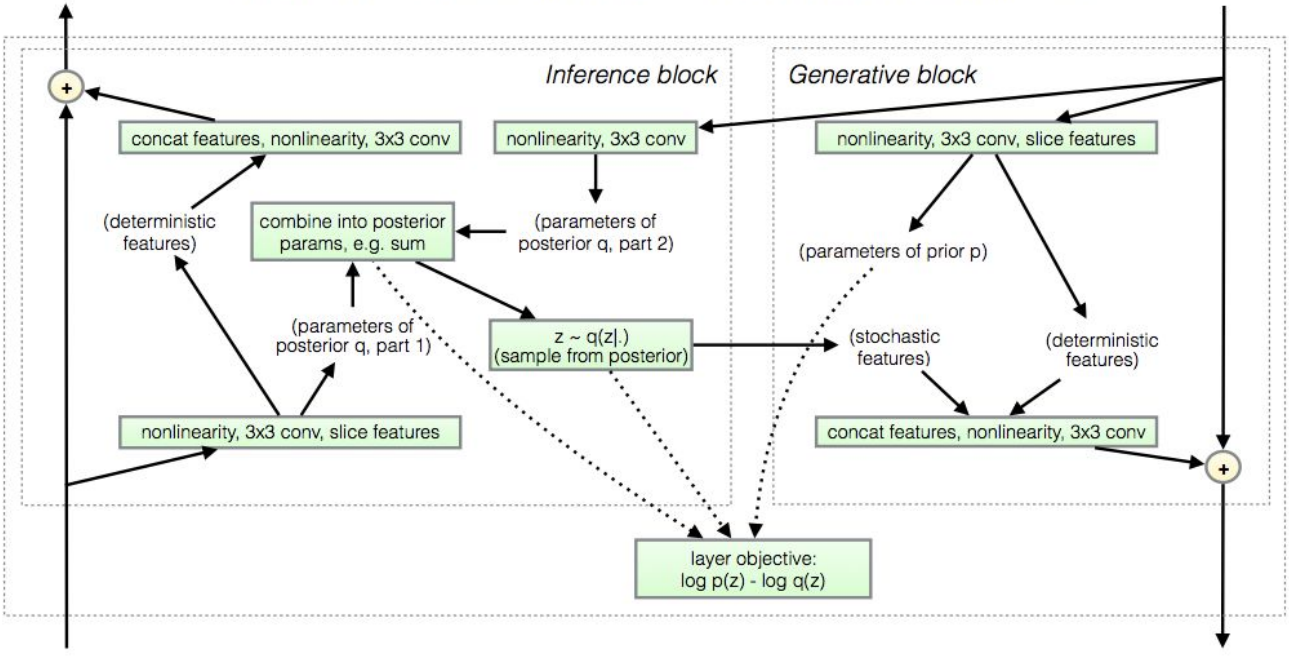


Figure 4: Generative ResNet and detail of layer. This is the generative component of our ResNet VAE.

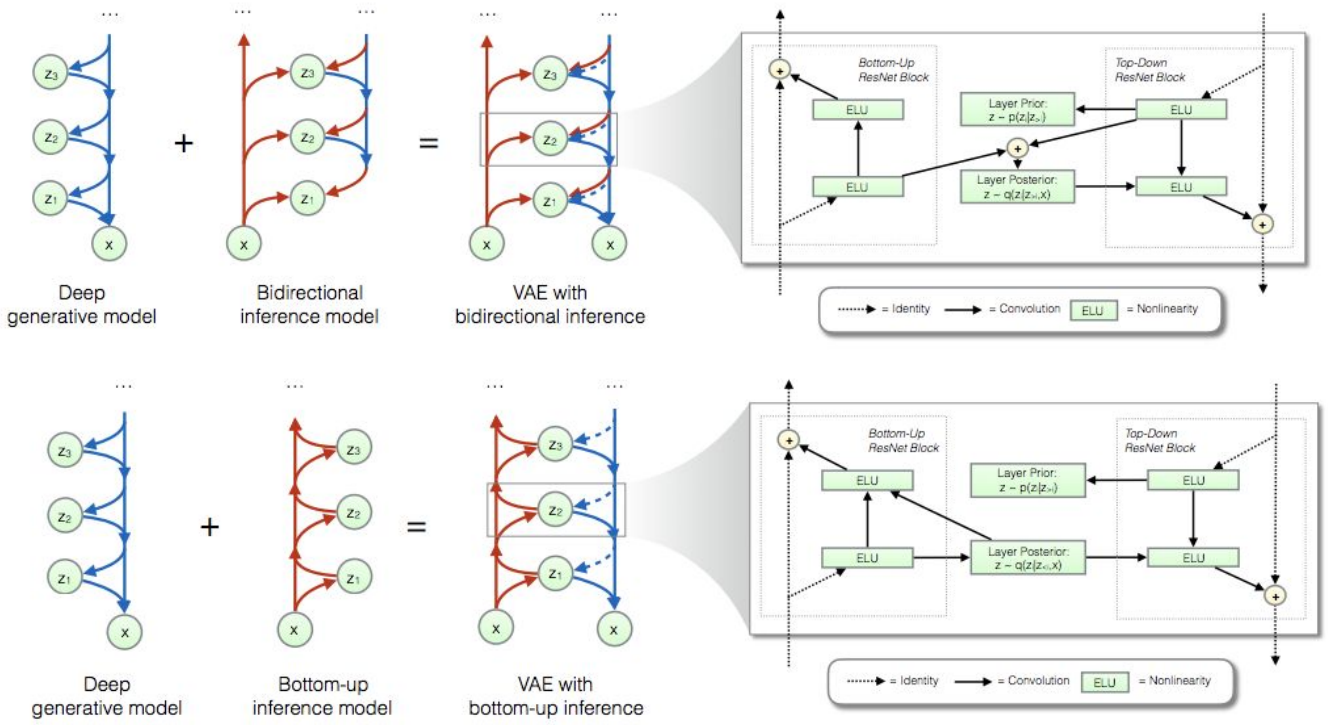
Finally, what's the difference between Parallel PixelCNN and Parallel WaveNet?
How will the IAF change from 1D voice to 2D image?



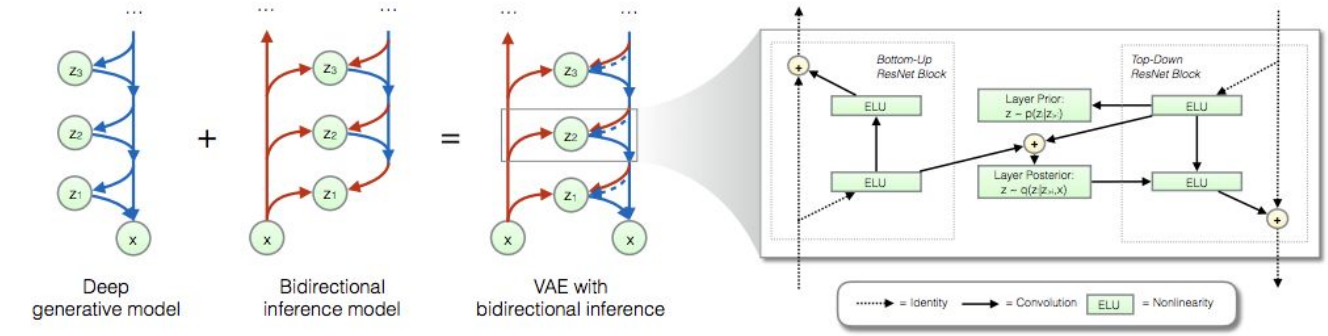
(a) Computational flow schematic of ResNet VAE with bottom-up inference



(b) Computational flow schematic of ResNet VAE with bidirectional inference



(a) Schematic overview of a ResNet VAE with bottom-up inference.



(b) Schematic overview of a ResNet VAE with bidirectional inference.

Reference:
Aaron's Parallel WaveNet paper: https://arxiv.org/pdf/1711.10433.pdf
Kingma's IAF paper: https://arxiv.org/pdf/1606.04934.pdf