Paper Summaries

Some Joke About VAEs

Written @ Corti

Authors:
Magnus Berg Sletfjerding
{ms}@corti.ai

April 22, 2021

0 Chapter 1: WaveNet

The WaveNet paper presents a CNN-based approach to generating audio samples. [?] Instead of using RNNs as a recurrent architecture, the generative model only conditions on past samples, and as such does not include any hidden "state".

The probability of a waveform $\mathbf{x} \in \mathbb{R}^T$ is expressed purely as:

$$p(\mathbf{x}) = \prod_{t=1}^{T} p(x_t|x_1, ..., x_t) \tag{1}$$

where $p(x_t|x_1, ..., x_t)$ is parametrized only by the weights in the network.

0.1 Architecture and design

The WaveNet Architecture draws advantage from three developments: quantized output spaces (as shown in PixelRNN), dilated causal convolutions and gated activation units,

Quantized Output Space with μ law companding transformation Given an audio waveform $\mathbf{x} \in [-1, 1]^T$, transform the audio according to :

$$f(x_t) = \operatorname{sign}(x_t)\frac{\ln(1 - \mu|x_t|)}{\ln(1 + \mu)} \tag{2}$$

with $\mu = 255$.

Dilated Causal Convolutions A Causal Convolution is a fancy way of saying that audio convolutions only work forward in time, not backward. This is to enforce the forward dependency in eq. (1).

A Dilated Convolution is a convolution where the convolution kernel skips over a dimension, increasing the receptive field and observing more of the surrounding environment. For an image the simplest dilated convolutional is illustrated in fig. 1

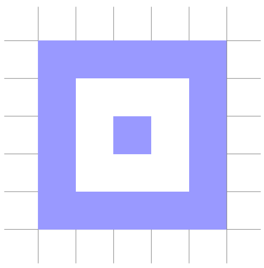

Figure 1: A Simple Pixel Dilated Convolution

Accordingly, for an audio signal, it would look like what we see in fig. 2

Gated Activation Units Each Convolution layer, Insteadof just having a filter weight, also has a **gating weight**. Hence the weights $\mathbf{W} \in \mathbb{R}^{K\times 2}$, with K as the number of layers. The operation of layer $k \in [0, K]$, is parametrized as:

$$\mathbf{z} = \tanh(\mathbf{x} * W_{k,f}) \odot \sigma(\mathbf{x} * W_{k,g}) \tag{3}$$

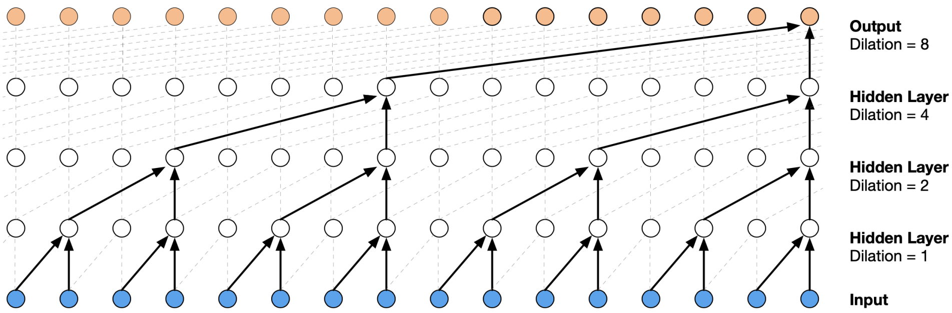

Figure 2: The Dilated Causal Convolution in WaveNet

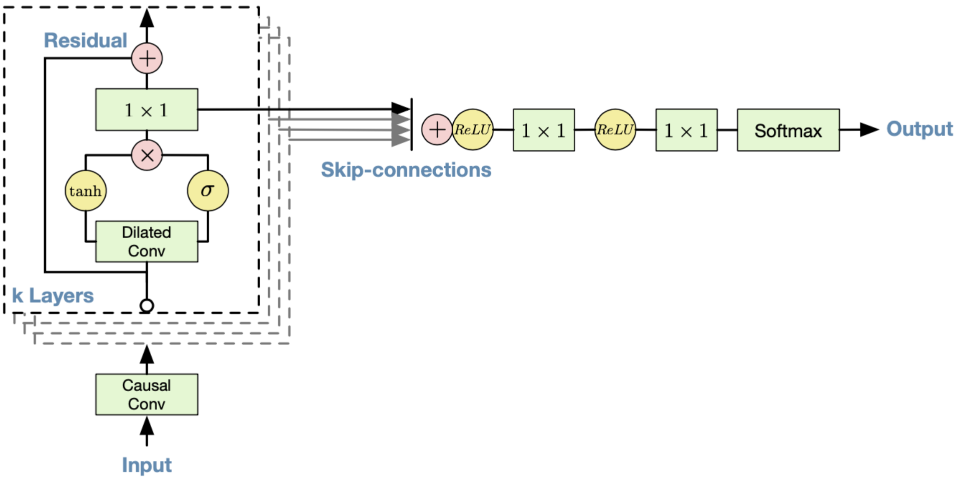

Figure 3: Overall Residual Architecture of WaveNet. Skip connections happen from every Convolutional Layer to the final softmax.

Summary of architechture The architecture is summed up in ??. It's important to note that the Causal Convolution setup as described in fig. 2 only is applied once, as the first layer. This makes the entire rest of the network a simple convolutional network with dilation, as the first (causal) convolutional stack ensures that the rest of the network will only see samples from the past. In all other respects we can consider this a standard CNN architecture.