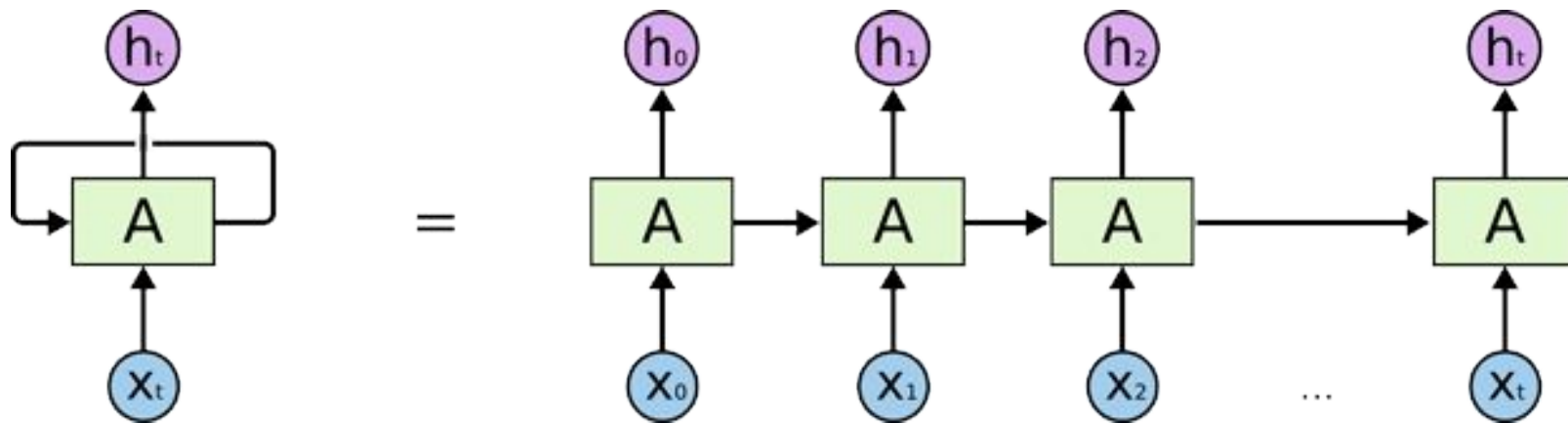
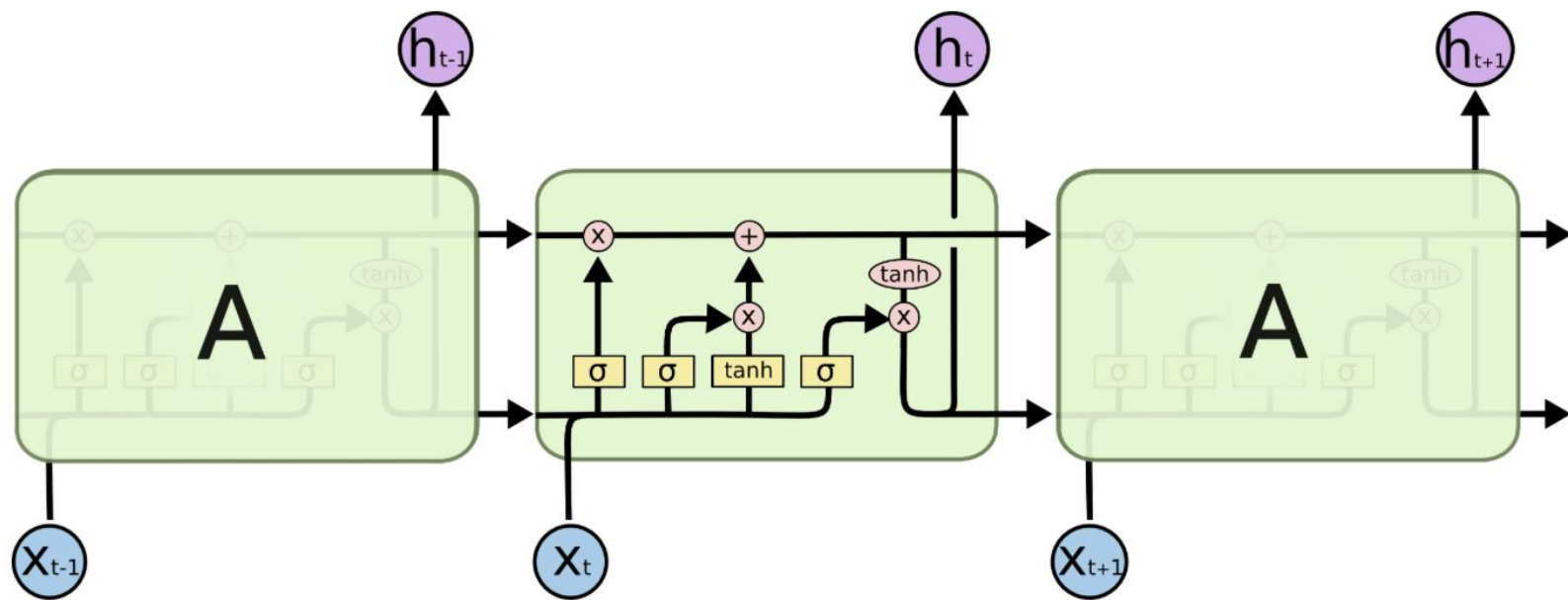


Генерация музыки

Рекуррентные сети

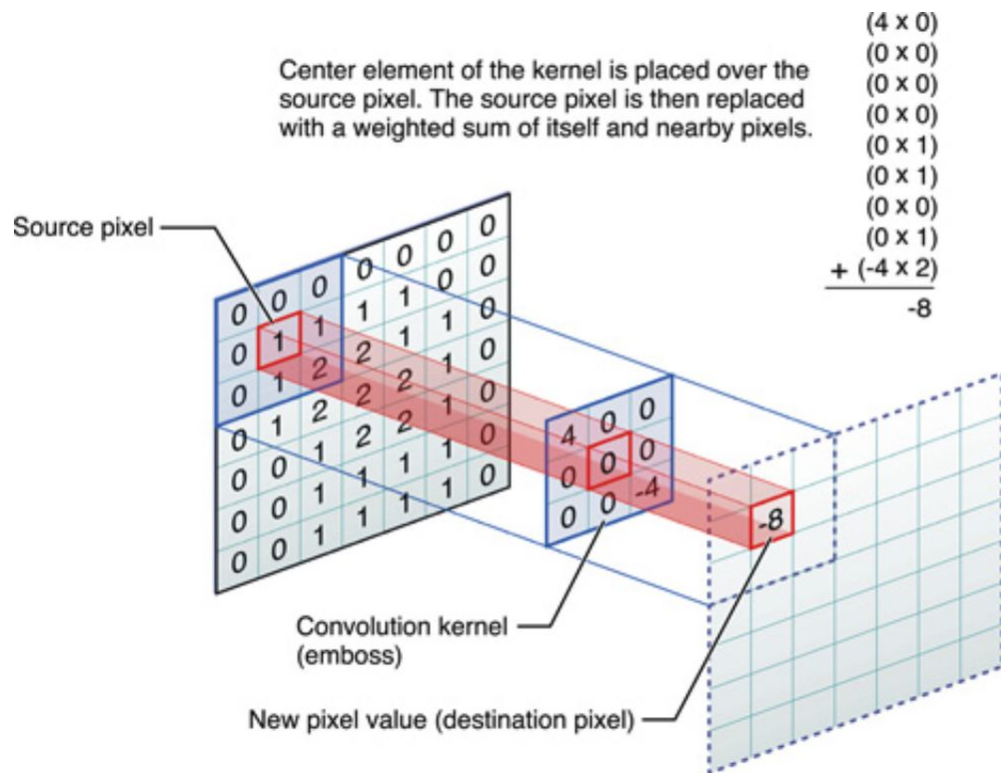


LSTM

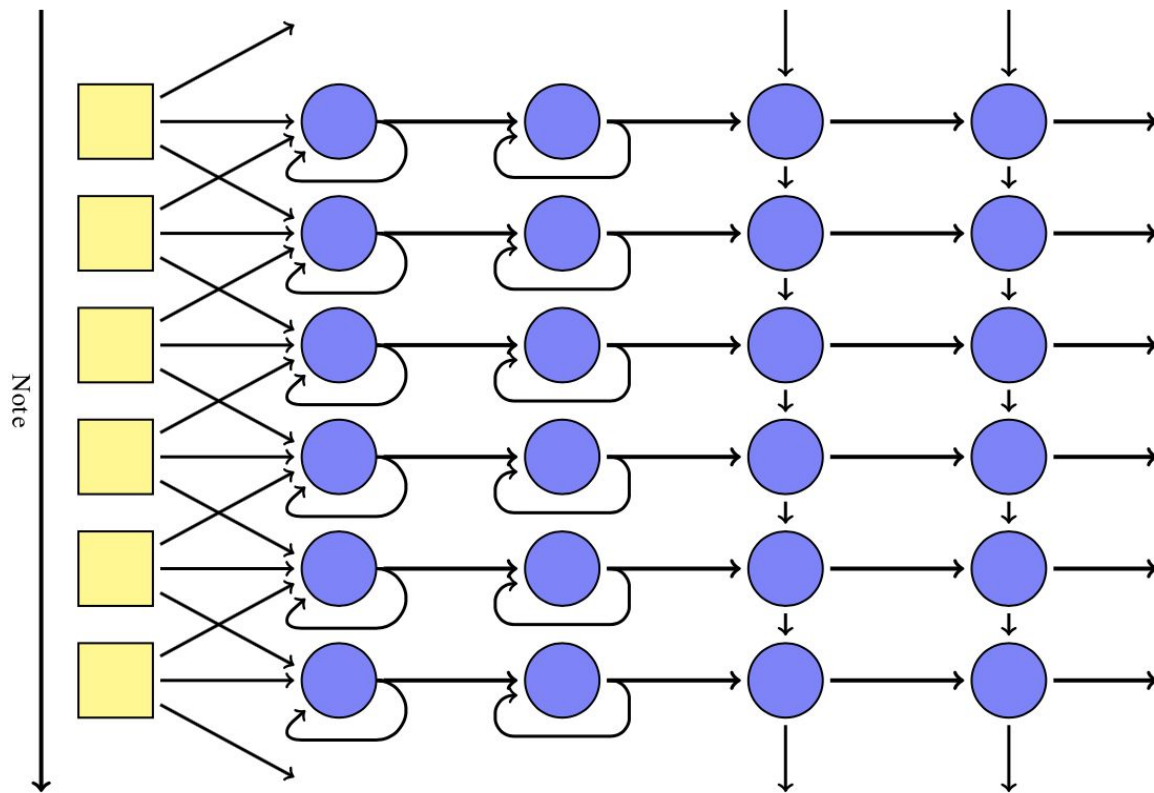


The repeating module in an LSTM contains four interacting layers.

Biaxial RNN



Biaxial RNN



WaveNet

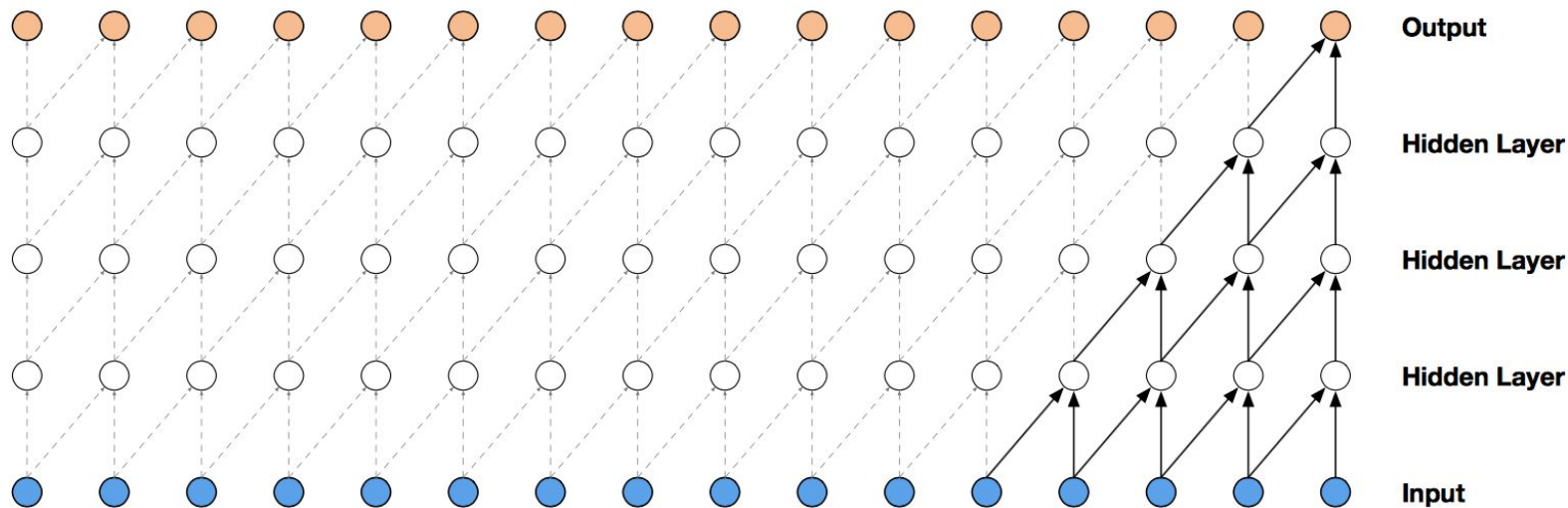
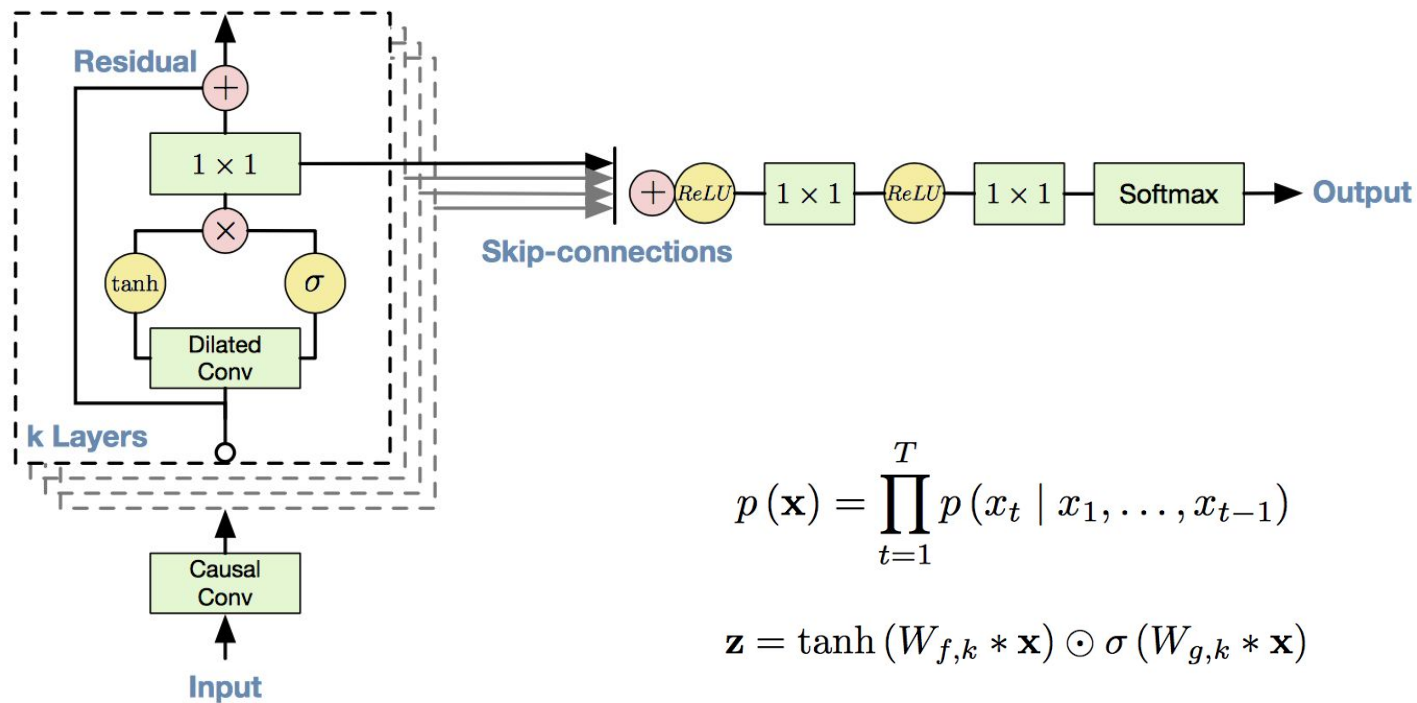


Figure 2: Visualization of a stack of causal convolutional layers.

WaveNet



WaveNet

$$p(\mathbf{x} \mid \mathbf{h}) = \prod_{t=1}^T p(x_t \mid x_1, \dots, x_{t-1}, \mathbf{h}).$$

global conditioning

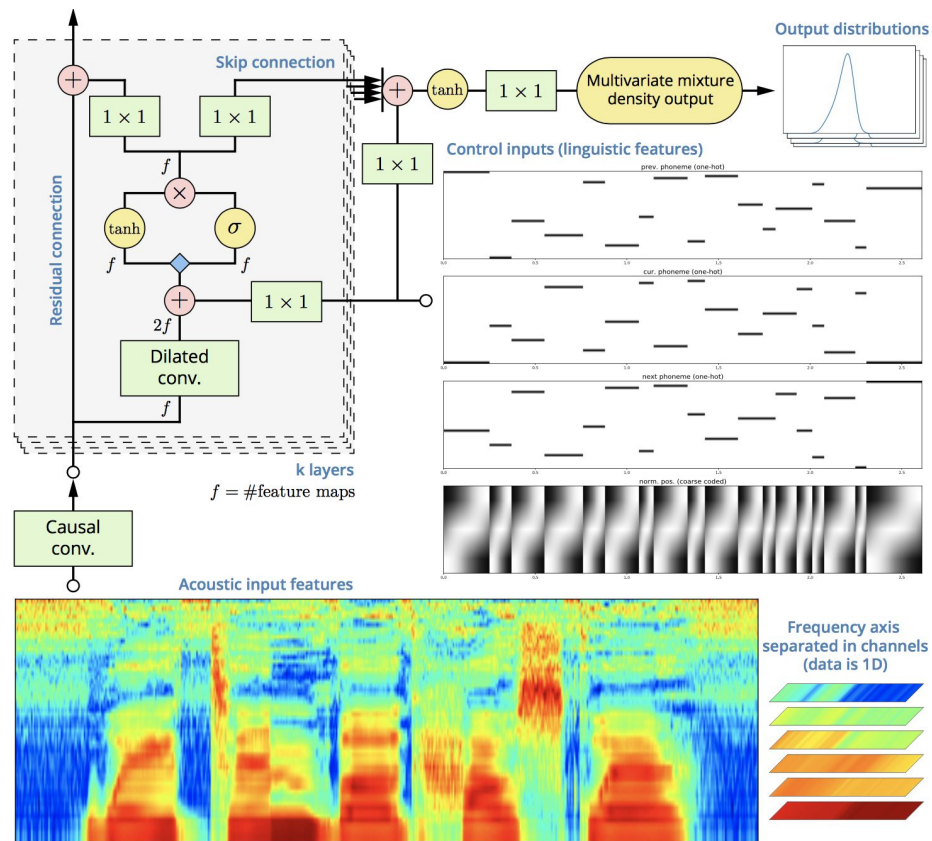
$$\mathbf{z} = \tanh(W_{f,k} * \mathbf{x} + V_{f,k}^T \mathbf{h}) \odot \sigma(W_{g,k} * \mathbf{x} + V_{g,k}^T \mathbf{h}).$$

local conditioning

$$\mathbf{y} = f(\mathbf{h})$$

$$\mathbf{z} = \tanh(W_{f,k} * \mathbf{x} + V_{f,k} * \mathbf{y}) \odot \sigma(W_{g,k} * \mathbf{x} + V_{g,k} * \mathbf{y})$$

A NEURAL PARAMETRIC SINGING SYNTHESIZER



Style transfer

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l.$$

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

$$\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l$$

$$F^l \in \mathcal{R}^{N_l \times M_l}$$

F_{ij}^l Активация i го
фильтра на позиции j
слоя l

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x})$$

Style transfer



Style transfer

По звуковым
сигналам строятся
спектрограммы.

Генерируется
случайное ядро
свертки.

Полученные в
результате свертки
фичи используются
для переноса стиля.

