How to model $p(x)$

- $\log \widehat{p}(x) = \mathrm{CNN}(x)$ Infeasible

- Use the chain rule

[Oord, Aaron van den, Nal Kalchbrenner, and Koray Kavukcuoglu.

"Pixel recurrent neural networks." (2016)]

$$p(x_1, \ldots, x_d)$$
$$= p(x_1)p(x_2 \mid x_1) \ldots p(x_d \mid x_1, \ldots, x_{d-1})$$
$$p(x_k \mid x_1, \ldots, x_{k-1}) = \mathrm{RNN}(x_1, \ldots, x_{k-1})$$