Pixel Recurrent Neural Networks

OMID RAZIZADEH

Department of Mechanics and Mathematics
Novosibirsk State University

Novosibirsk State University
*THE REAL SCIENCE

Generative image modeling is an unsupervised learning problem.

Probablistic Density Models :

- Image Compression
- Debluring
- Generating of new images,etc

Obstacle in Generative modeling :

To build complex and expressive models that are tractable and scalable. This balance has resulted in a large variety of generative models.

Stochastic latent variable models such as VAE's :

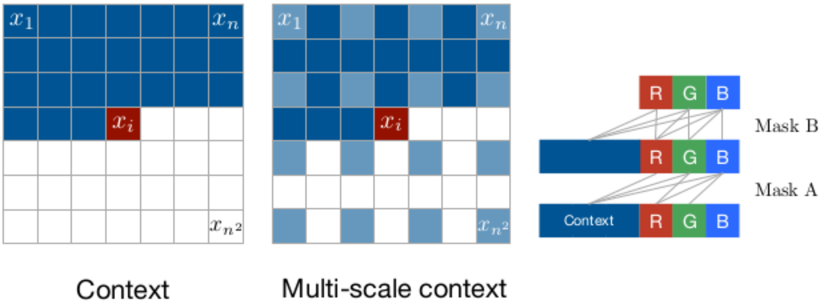
- Extract meaningful representation
- **but not** tractable inference

So What is the best model ?

The best approach is to use product of conditional distributions.It is used in models such as NADE.

RNNs :

Are powerful models that are used for :
-Handwriting generation
-Character prediction
-Machine Translation


x_1
x_n
x_i
x_{n^2}
Context
x_1
x_n
x_i
x_{n^2}
Multi-scale context
R
G
B
Mask B
R
G
B
Mask A
Context
R
G
B

Generating an Image Pixel by Pixel :

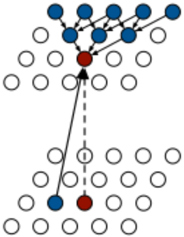
$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i | x_1, ..., x_{i-1})$$

$$p(x_{i,R}|\mathbf{x}_{<i})p(x_{i,G}|\mathbf{x}_{<i}, x_{i,R})p(x_{i,B}|\mathbf{x}_{<i}, x_{i,R}, x_{i,G})$$

During training and evaluation the distributions over the pixel values are computed in parallel, while the generation of an image is sequential.

Row LSTM :

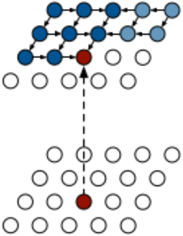
- is a unidirectional layer
- process row by row from top to bottom
- computing features for a whole row at once
- computation is performed with a one-dimensional convolution
- kernel has size k $\times$ 1 where k $\geq$ 3



Row LSTM

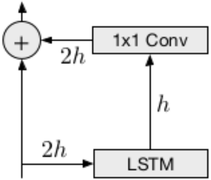
Diagonal BiLSTM :

- capture the entire available context for any image size

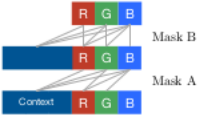


Diagonal BiLSTM

Residual Connections :



Masked Convolution :
masks can be easily implemented by zeroing out the corresponding weights in the input-to-state convolutions after each update



Evaluation :
All models are trained and evaluated by log-likelihood loss function coming from discrete distribution not continuous distributions using density function.

Performance of different models on MNIST :

Model	NLL Test
DBM 2hl [1]:	≈ 84.62
DBN 2hl [2]:	≈ 84.55
NADE [3]:	88.33
EoNADE 2hl (128 orderings) [3]:	85.10
EoNADE-5 2hl (128 orderings) [4]:	84.68
DLGM [5]:	≈ 86.60
DLGM 8 leapfrog steps [6]:	≈ 85.51
DARN 1hl [7]:	≈ 84.13
MADE 2hl (32 masks) [8]:	86.64
DRAW [9]:	≤ 80.97
PixelCNN:	81.30
Row LSTM:	80.54
Diagonal BiLSTM (1 layer, $h = 32$):	**80.75**
Diagonal BiLSTM (7 layers, $h = 16$):	**79.20**

Table 4. Test set performance of different models on MNIST in *nats* (negative log-likelihood). Prior results taken from [1] (Salakhutdinov & Hinton, 2009), [2] (Murray & Salakhutdinov, 2009), [3] (Uria et al., 2014), [4] (Raiko et al., 2014), [5] (Rezende et al., 2014), [6] (Salimans et al., 2015), [7] (Gregor et al., 2014), [8] (Germain et al., 2015), [9] (Gregor et al., 2015).

Performance of different models on CIFAR-10 :

Model	NLL Test (Train)
Uniform Distribution:	8.00
Multivariate Gaussian:	4.70
NICE [1]:	4.48
Deep Diffusion [2]:	4.20
Deep GMMs [3]:	4.00
RIDE [4]:	3.47
PixelCNN:	3.14 (3.08)
Row LSTM:	3.07 (3.00)
Diagonal BiLSTM:	**3.00** (2.93)

Table 5. Test set performance of different models on CIFAR-10 in *bits/dim*. For our models we give training performance in brackets. [1] (Dinh et al., 2014), [2] (Sohl-Dickstein et al., 2015), [3] (van den Oord & Schrauwen, 2014a), [4] personal communication (Theis & Bethge, 2015).

Conclusion :

- two-dimensional LSTM layers : the Row LSTM and the Diagonal BiLSTM, that scale more easily to larger datasets.
- We employed masked convolutions to allow PixelRNNs to model full dependencies between the color channels.
- PixelRNNs significantly improve the state of the art on the MNIST and CIFAR-10 datasets.
- PixelRNNs are able to model both spatially local and long-range correlations and are able to produce images.that are sharp and coherent.
- More computation and larger models are likely to further improve the results.

References :

arXiv :1601.06759[Best ICML paper in 2016]