Recurrent Convolutional Neural Network for Object Recognition

Ming Liang, Xiaolin Hu

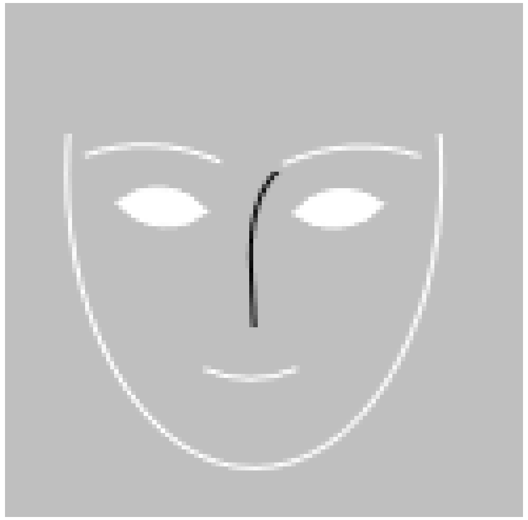
채진영

Intro

- CNN models have purely feed-forward architectures which can be seen as approximations of the biological neural network in the brain.

- Anatomical evidences have shown that recurrent synapses typically outnumber feed-forward and top down (or feedback) synapses

Intro



- Due to the presence of recurrent and top-down synapses, object recognition is actually a dynamic process though the input is static
- It is generally believed that recurrent synapses play an important role in context modulation
- Context for object recognition is important. Without the context (face), it is hard to recognize the black curve in the middle area as a nose.

Difference of Recurrent CNN

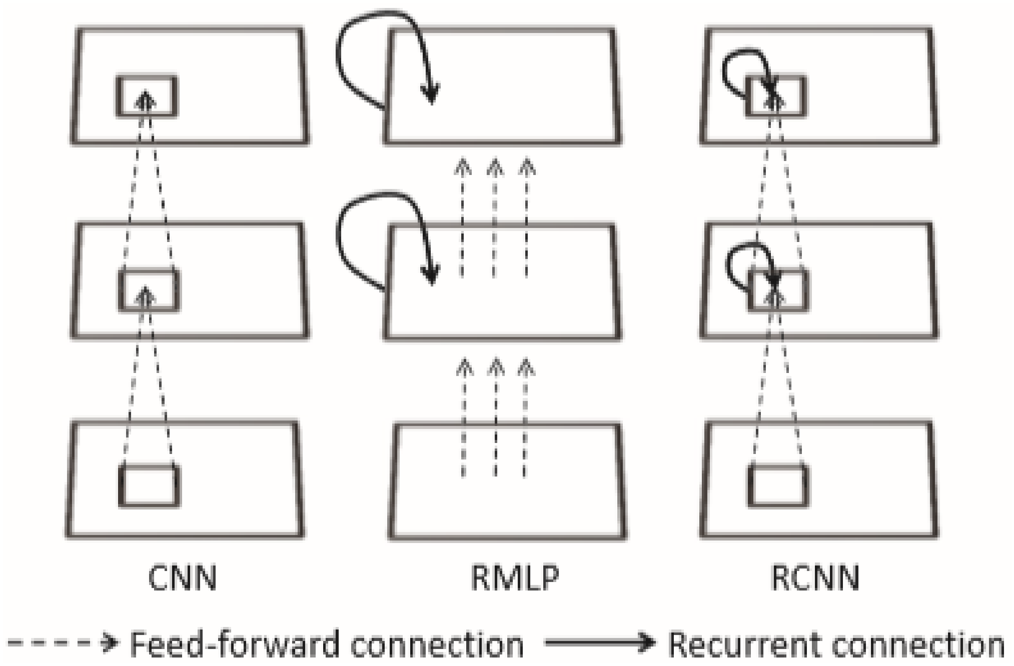


Figure 2. Illustration of the architectures of CNN, RMLP and RCNN. For each model two hidden layers are shown.

- Both feed-forward and recurrent connections have local connectivity and shared weights among different locations.

- Architecture is very similar to the recurrent multi layer perceptron

- The main difference is that the full connections in RMLP are replaced by shared local connections, just as the difference between MLP [40] and CNN.

RCNN layer

$$z_{ijk}(t) = (\mathbf{w}_k^f)^T \mathbf{u}^{(i,j)}(t) + (\mathbf{w}_k^r)^T \mathbf{x}^{(i,j)}(t-1) + b_k.$$

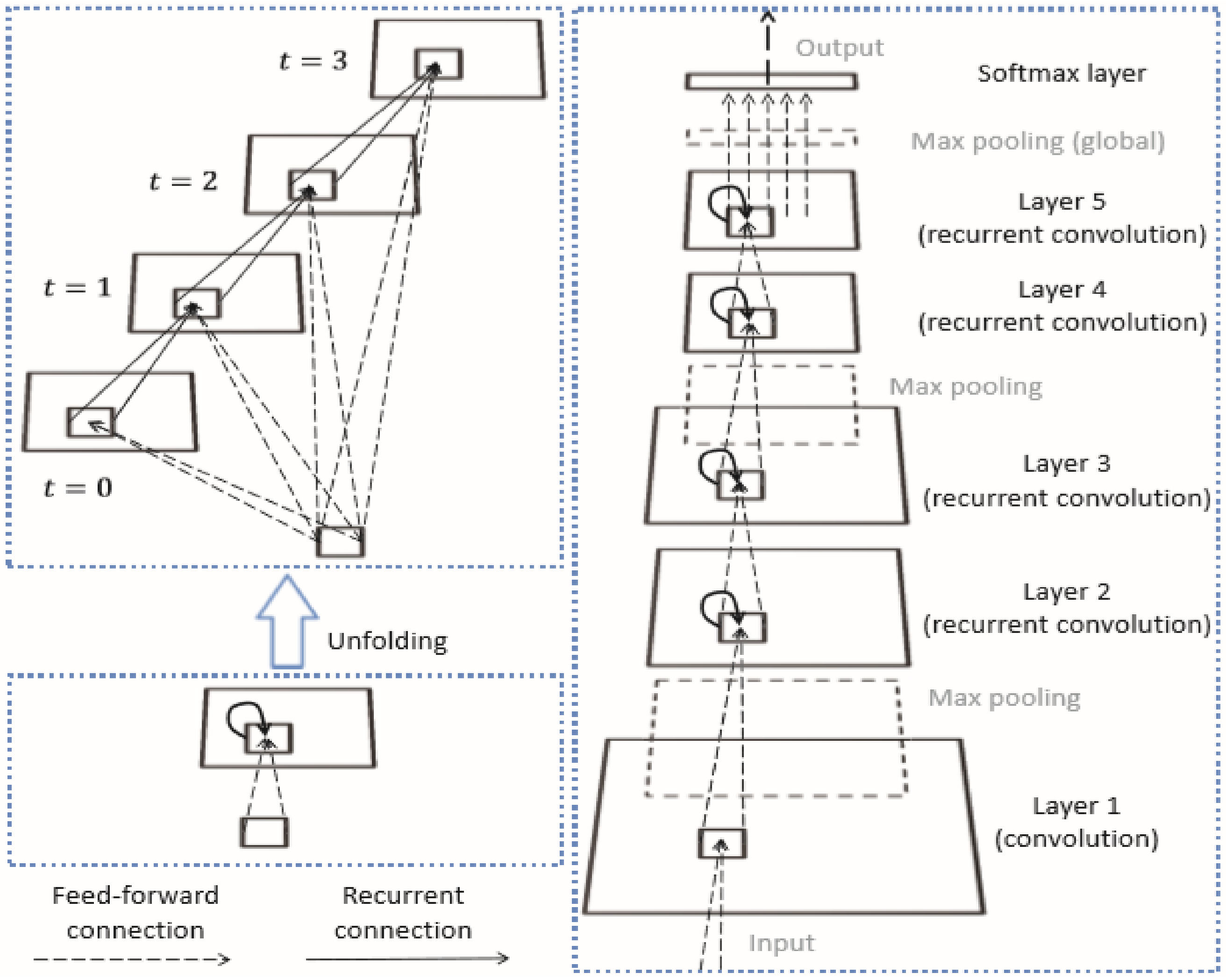
$$x_{ijk}(t) = g(f(z_{ijk}(t))),$$

$$f(z_{ijk}(t)) = \max(z_{ijk}(t), 0),$$

$$g(f_{ijk}(t)) = \frac{f_{ijk}(t)}{\left(1 + \frac{\alpha}{N} \sum_{k'=\max(0,k-N/2)}^{\min(K,k+N/2)} (f_{ijk'})^2\right)^{\beta}}$$

- $\mathbf{u}^{(i,j)}(t)$:feed-forward input
- $\mathbf{x}^{(i,j)}(t-1)$: Recurrent input
- $\mathbf{w}_k^f$: feed-forward weights
- $\mathbf{w}_k^r$: Recurrent weights
- f() : ReLU
- g() : Local ReponseNormalization.

Model Architecture



- The recurrent input evolves over iterations, the feed-forward input remains the same in all iterations. .
- The longest path goes through all unfolded recurrent connections (therefore length = T + 1), while the shortest path goes through the feed-forward connection only (therefore length = 1).
- 4(T + 1) + 2 is only the length of the longest path from the input layer to the output layer. the shortest path has length 6

Discussion

- Current layer has incorporate context information in an arbitrarily large region .
- The recurrent connections increase the network depth while keep the number of adjustable parameters constant by weight sharing.
- The time-unfolded RCNN is actually a CNN with multiple paths between the input layer to the output layer ,which may facilitate the learning.
- The existence of longer paths makes it possible for the model to learn highly complex features.

result

Model	No. of Param.	Testing Error (%)
Without Data Augmentation		
Maxout [17]	> 5 M	11.68
Prob maxout [47]	> 5 M	11.35
NIN [33]	0.97 M	10.41
DSN [30]	0.97 M	9.69
RCNN-96	0.67 M	**9.31**
RCNN-128	1.19 M	**8.98**
RCNN-160	1.86 M	**8.69**
RCNN-96 (no dropout)	0.67 M	13.56
NIN (no dropout) [33]	0.97 M	14.51
With Data Augmentation		
Prob maxout [47]	> 5 M	9.39
Maxout [17]	> 5 M	9.38
DropConnect (12 nets) [51]	—	9.32
NIN [33]	0.97 M	8.81
DSN [30]	0.97 M	7.97
RCNN-96	0.67 M	**7.37**
RCNN-128	1.19 M	**7.24**
RCNN-160	1.86 M	**7.09**

Table 2. Comparison with existing models on CIFAR-10

Model	No. of Param.	Error (%)	
		Training	Testing
rCNN-96 (1 iter)	0.67 M	4.61	12.65
rCNN-96 (2 iters)	0.67 M	2.26	12.99
rCNN-96 (3 iters)	0.67 M	1.24	14.92
WCNN-128	0.60 M	3.45	9.98
RCNN-96 (1 iter)	0.67 M	4.99	9.95
RCNN-96 (2 iters)	0.67 M	3.58	9.63
RCNN-96 (3 iters)	0.67 M	3.06	9.31

Table 1. Comparison with the baseline models on CIFAR-10

4.3. CIFAR-100

Model	No. of Param.	Testing Error (%)
Maxout [17]	> 5 M	38.57
Prob maxout [47]	> 5 M	38.14
Tree based priors [49]	—	36.85
NIN [33]	0.98 M	35.68
DSN [30]	0.98 M	34.57
RCNN-96	0.68 M	**34.18**
RCNN-128	1.20 M	**32.59**
RCNN-160	1.87 M	**31.75**

Table 3. Comparison with existing models on CIFAR-100

Author other paper

Convolutional Neural Networks with Intra-layer Recurrent Connections for Scene Labeling.

-Ming Liang, Xiaolin Hu, Bo Zhang

Thanks