中国科学院自动化研究所
Institute of Automation, Chinese Academy of Sciences

Gated Recurrent Convolution Neural Network for OCR

主讲人：杜臣

2018.06.31

背景

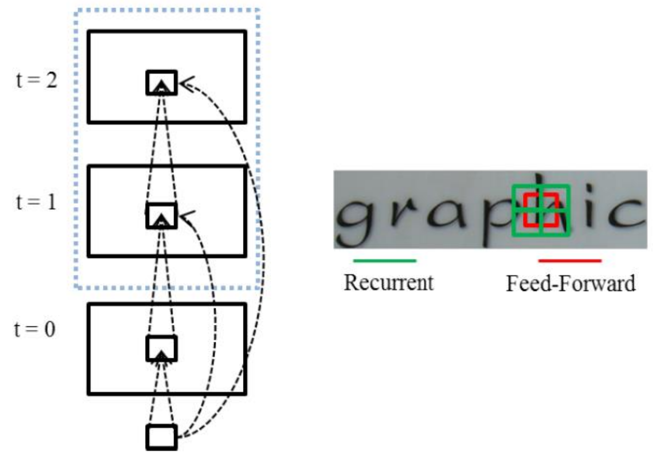

Figure 1: Illustration of using RCL with $T = 2$ for OCR.

1. Recurrent Convolution Layer (RCL)在卷积层中添加了循环卷积来扩大单层卷积层的感受野以及融合高底层信息。
2. 在RCNN中，如果增加迭代次数（T），每个卷积层的有效感受野将急剧增加，（与所基于的生物学事实不符），所以需要引入一个控制机制来限制有效感受野的增长。

Recurrent Convolution Neural Network

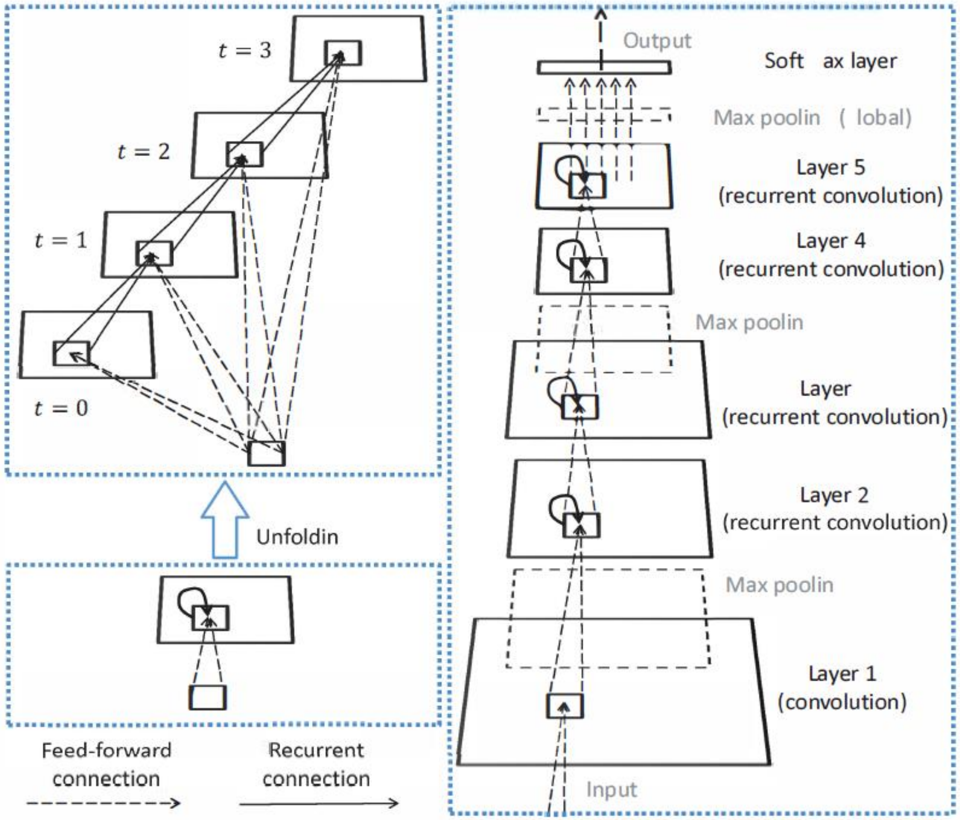

Figure 3. The overall architecture of RCNN. Left: An RCL is unfolded for $T = 3$ time steps, leading to a feed-forward subnetwork with the largest depth of 4 and the smallest depth of 1. At $t = 0$ only feed-forward computation takes place. Right: The RCNN used in this paper contains one convolutional layer, four RCLs, three max pooling layers and one softmax layer.

Gated Recurrent Convolution Layer and GRCNN

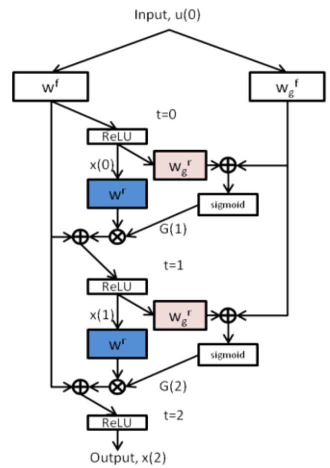

Figure 2: Illustration of GRCL with $T = 2$. The convolutional kernels in the same color use the same weights.

1x1 kernels
the recurrent weights for the gate

1x1 kernels
the recurrent weights for the gate

门控信号的计算：

$$G(t) = \begin{cases} 0 & t = 0 \\ sigmoid(BN(w_g^f * u(t)) + BN(w_g^r * x(t-1))) & t > 0 \end{cases}$$

每一时刻输出状态的计算：

$$x(t) = \begin{cases} ReLU(BN(w^f * u(t)) & t = 0 \\ ReLU(BN(w^f * u(t)) + BN(BN(w^r * x(t-1)) \odot G(t))) & t > 0 \end{cases}$$

Overall Architecture GRCNN-BLSTM Model

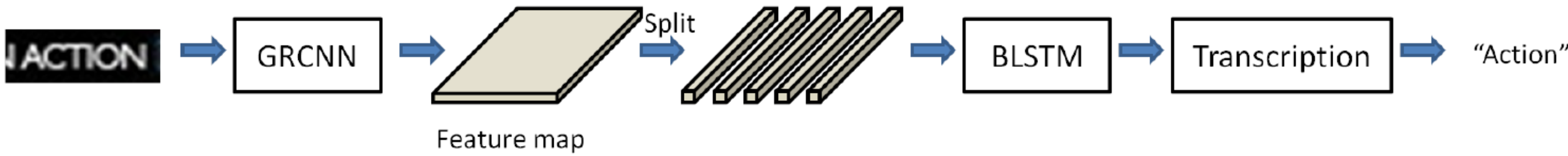

Figure 3: Overall pipeline of the architecture.

使用GRCNN做特征提取，然后接入BLSTM和CTC进行识别。

使用的特征提取GRCNN结构如下：

Table 1: The GRCNN configuration

Conv	MaxPool	GRCL	MaxPool	GRCL	MaxPool	GRCL	MaxPool	Conv
3×3	2×2	3×3	2×2	3×3	2×2	3×3	2×2	2×2
num: 64		num: 64		num: 128		num: 256		num: 512
sh:1 sw:1	sh:2 sw:2	sh:1 sw:1	sh:2 sw:2	sh:1 sw:1	sh:2 sw:1	sh:1 sw:1	sh:2 sw:1	sh:1 sw:1
ph:1 pw:1	ph:0 pw:0	ph:1 pw:1	ph:0 pw:0	ph:1 pw:1	ph:0 pw:1	ph:1 pw:1	ph:0 pw:1	ph:0 pw:0

num: denotes the number of feature maps
sh: denotes the stride of the kernel along the height;
sw: denotes the stride along the width;
"ph" and **"pw"** denote the padding value of height and width respectively;

Experiments

实验1、迭代次数（T）对性能的影响，加入gate后对性能的影响

实验2、不同的lstm结构对性能的影响

Table 2: Model analysis over the IIIT5K and SVT (%). Mean and standard deviation of the results are reported.

(a) GRCNN analysis

Model	IIIT5K	SVT
Plain CNN	77.21±0.54	77.69±0.59
RCNN(1 iter)	77.64±0.58	78.23±0.56
RCNN(2 iters)	78.17±0.56	79.11±0.63
RCNN(3 iters)	78.94±0.61	79.76±0.59
GRCNN(1 iter)	77.92±0.57	78.67±0.53
GRCNN(2 iters)	79.42±0.63	79.89±0.64
GRCNN(3 iters)	80.21±0.57	80.98±0.60

(b) LSTM's variants analysis

LSTM variants	IIIT5K	SVT
$\text{LSTM}_{\{\gamma_1=0,\gamma_2=0,\gamma_3=0\}}$	77.92±0.57	78.67±0.53
$\text{LSTM-F}_{\{\gamma_1=0,\gamma_2=1,\gamma_3=0\}}$	77.26±0.61	78.23±0.53
$\text{LSTM-I}_{\{\gamma_1=1,\gamma_2=0,\gamma_3=0\}}$	76.84±0.58	76.89±0.63
$\text{LSTM-O}_{\{\gamma_1=0,\gamma_2=0,\gamma_3=1\}}$	76.91±0.64	78.65±0.56
$\text{LSTM-A}_{\{\gamma_1=1,\gamma_2=1,\gamma_3=1\}}$	76.52±0.66	77.88±0.59

Experiments

实验3、：整体的GRCNN-BLSTM Model性能与已有的方法的比较

Table 3: The text recognition accuracies in natural images. "50","1k" and "Full" denote the lexicon size used for lexicon-based recognition task. The dataset without lexicon size means the unconstrained text recognition

Method	SVT-50	SVT	IIIT5K-50	IIIT5K-1k	IIIT5K	IC03-50	IC03-Full	IC03
ABBYY [36]	35.0%	-	24.3%	-	-	56.0%	55.0%	-
wang et al. [36]	57.0%	-	-	-	-	76.0%	62.0%	-
Mishra et al. [25]	73.2%	-	-	-	-	81.8%	67.8%	-
Novikova et al. [27]	72.9%	-	64.1%	57.5%	-	82.8%	-	-
wang et al. [38]	70.0%	-	-	-	-	90.0%	84.0%	-
Bissacco et al. [3]	90.4%	78.0%	-	-	-	-	-	-
Goel et al. [6]	77.3%	-	-	-	-	89.7%	-	-
Alsharif [2]	74.3%	-	-	-	-	93.1%	88.6%	-
Almazan et al. [1]	89.2%	-	91.2%	82.1%	-	-	-	-
Lee et al. [20]	80.0%	-	-	-	-	88.0%	76.0%	-
Yao et al. [40]	75.9%	-	80.2%	69.3%	-	88.5%	80.3%	-
Rodriguez et al. [28]	70.0%	-	76.1%	57.4%	-	-	-	-
Jaderberg et al. [16]	86.1%	-	-	-	-	96.2%	91.5%	-
Su and Lu et al. [33]	83.0%	-	-	-	-	92.0%	82.0%	-
Gordo [7]	90.7%	-	93.3%	86.6%	-	-	-	-
Jaderberg et al. [14]	93.2%	71.1%	95.5%	89.6%	-	97.8%	97.0%	89.6%
Baoguang et al. [30]	**96.4%**	80.8%	97.6%	94.4%	78.2%	98.7%	97.6%	89.4%
Chen-Yu et al. [21]	96.3%	80.7%	96.8%	94.4%	78.4%	97.9%	97.0%	88.7%
ResNet-BLSTM	96.0%	80.2%	97.5%	94.9%	79.2%	98.1%	97.3%	89.9%
Ours	96.3%	**81.5%**	**98.0%**	**95.6%**	**80.8%**	**98.8%**	**97.8%**	**91.2%**

Success

Failure

References

[1] Recurrent Convolutional Neural Network for Object Recognition

[2] Gated Recurrent Convolution Neural Network for OCR