IBN-STR: A Robust Text Recognizer for Irregular Text in Natural Scenes

Xiaoqian Li, Jie Liu, Guixuan Zhang, Shuwu Zhang
Institute of Automation, Chinese Academy of Sciences, Beijing, China

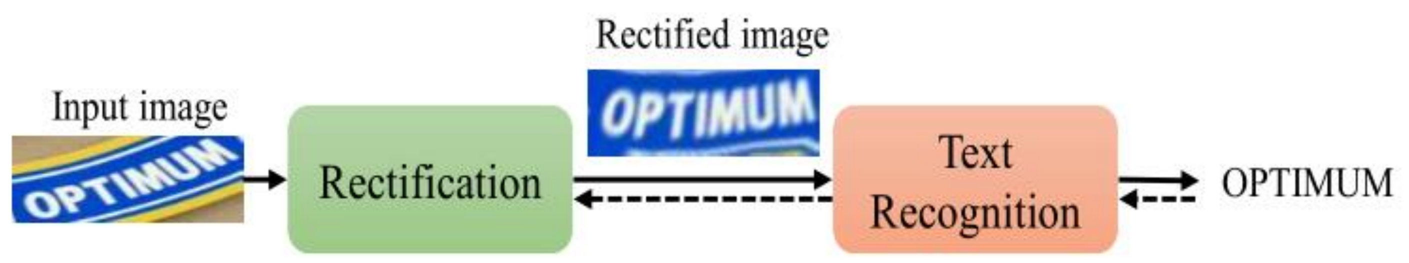
Motivation and Objectives

· **Bias between the distribution of training data and test data:**
- **Synthetic Text (training) is more regular and has small curvature**
- **Real Text (testing) has greater curvature and more changeable text style**

· **IBN-STR: A robust recognizer:**
- **In terms of data, S-shape distortion is applied to increase the diversity of training data**
- **In terms of feature, effective IBN module is introduced**

(a) Real text (b) Synthetic text

Overview of IBN-STR



The proposed IBN-STR model consists of a rectification network and a text recognition network. The rectification network is based on the STN and generates rectified images. The text recognition network consists of a CNN-BLSTM encoder and an attention-based decoder.

The encoder first extracts stacked CNN features of input images and utilizes BLSTM to convert the image features into feature sequences. The decoder is a seq2seq model that translates the feature sequence into a character sequence.

The IBN module is embedded in the stacked convolutional modules to improve the capacity and generalization ability of text recognizer.

TABLE I
ARCHITECTURE OF TEXT RECOGNITION NETWORK. BLSTM MEANS BIDIRECTIONAL LONG SHORT-TERM MEMORY LAYER.

	Layers	Configurations	Outsize
encoder	Block 0	3×3 *conv*, s 1×1, *bn*	$32 \times 32 \times 100$
	Block 1	$\begin{bmatrix} 1 \times 1\ conv, 32, bn \\ 3 \times 3\ conv, 32, bn \end{bmatrix} \times 3, s\ 2 \times 2$	$32 \times 16 \times 50$
	Block 2	$\begin{bmatrix} 1 \times 1\ conv, 64, ibn \\ 3 \times 3\ conv, 64, bn \end{bmatrix} \times 4, s\ 2 \times 2$	$64 \times 8 \times 25$
	Block 3	$\begin{bmatrix} 1 \times 1\ conv, 128, ibn \\ 3 \times 3\ conv, 128, bn \end{bmatrix} \times 6, s\ 2 \times 1$	$128 \times 4 \times 25$
	Block 4	$\begin{bmatrix} 1 \times 1\ conv, 256, ibn \\ 3 \times 3\ conv, 256, bn \end{bmatrix} \times 6, s\ 2 \times 1$	$256 \times 2 \times 25$
	Block 5	$\begin{bmatrix} 1 \times 1\ conv, 512, bn \\ 3 \times 3\ conv, 512, bn \end{bmatrix} \times 3, s\ 2 \times 1$	$512 \times 1 \times 25$
	BLSTM1	256 hidden units	25×256
	BLSTM2	256 hidden units	25×256
decoder	GRU	256 hidden units	25×256

S-Shape distortion -- Enrich the diversity of training data

Given the position of original image (i, j) and the position of rectified image (i', j'), the correspondences of between (i, j) and (i', j') are as follows:

$$i' = a_1 i + a_2 Sin(\theta, j) + a_3,$$
$$j' = j,$$

where a1, a2, a3 are scaling and shifting par-ammeter, θ determines the distortion mode.

IBN module -- Improve generalization performance

Instance normalization is introduced to learn features that invariant to styles or appearance. Two types of IBN modules are provided.

IBN-a module: the outputs of IN and BN will be concatenated.

IBN-b module: IN will be placed before block output.

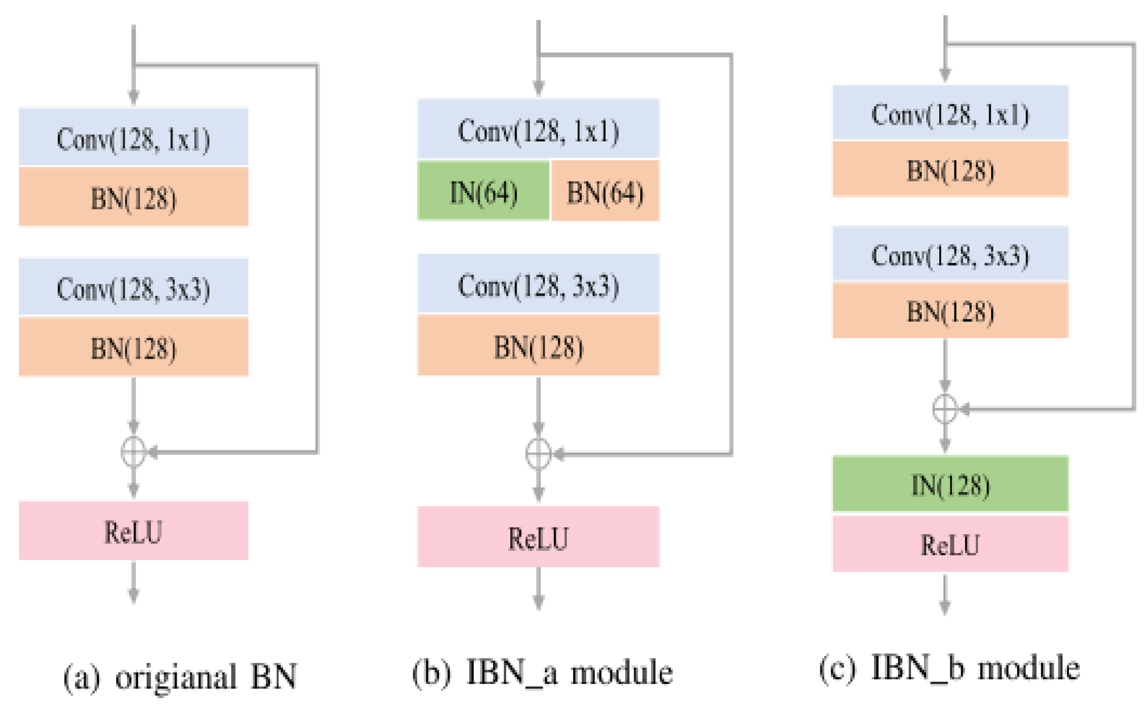


Fig. 2. Instance-batch normalization (IBN) module.

Experimental Results

TABLE III
THE RESULTS OF DATA AUGMENTATION.

Improvement	Regular	Irregular	Total
S-shape(BO+37)	+0.15	+0.13	+0.15
S-shape-stn(BO+37)	+0.52	+0.17	+0.33
S-shape(TO+38)	+0.21	+1.07	+0.67
S-shape-stn(TO+38)	+0.15	+1.13	+0.67

TABLE V
THE RESULTS OF DIFFERENT NUMBER OF IBN LAYERS.

Method	Regular	Irregular	Total
BN	92.53	75.49	83.54
IBN, 2	$92.66^{+0.13}$	$76.04^{+0.55}$	$83.89^{+0.35}$
IBN, 1-2	$\mathbf{93.03}^{+0.50}$	$75.87^{+0.38}$	$83.97^{+0.43}$
IBN, 2-3	$92.90^{+0.37}$	$75.85^{+0.36}$	$83.90^{+0.36}$
IBN, 2-4	$92.92^{+0.39}$	$\mathbf{76.97}^{+1.48}$	$\mathbf{84.50}^{+0.96}$
IBN, 1-4	$92.65^{+0.12}$	$76.39^{+0.90}$	$84.06^{+0.52}$

TABLE VI
COMPARISON OF OTHER TEXT RECOGNITION METHODS. * MEANS USING 1,811 IMAGES.

Method	Data	Regular						Irregular					Total
		IC13	SVT		IIIT5K			IC15	SVT-P		CUTE	Total-text	
		None	None	50	None	50	1k	None	None	50	None	None	
CRNN [4]	SK	89.6	82.7	97.5	81.2	97.8	95.0	-	-	-	-	-	-
GCRNN [34]	SK	-	81.5	96.3	80.8	98.0	95.6	-	-	-	-	-	-
R2AM [15]	SK	90.0	80.7	96.3	78.4	96.8	94.4	-	-	-	-	-	-
Liao et.al [35]	ST	91.4	82.1	98.5	92.0	99.8	98.9	-	-	-	78.1	-	
Aster [6]	ST+SK	91.8	**93.6**	**99.2**	93.4	99.6	98.8	76.1*	78.5	-	79.5	-	-
2D CTC [36]	ST+SK	93.9	90.6	97.2	**94.7**	**99.8**	98.9	75.2*	79.2	-	81.3	63.0	-
RCN [16]	ST+SK	93.2	88.6	97.7	94.0	99.6	98.9	77.1	80.6	**95.0**	**88.5**	-	-
MORAN [7]	ST+SK	92.4	88.3	96.6	91.2	97.9	96.2	68.8	76.1	94.3	77.4	-	-
Lyu et.al [17]	ST+SK	92.7	90.1	97.2	94.0	**99.8**	**99.1**	76.3	82.3	-	86.8	-	-
IBN-STR(base)	ST+SK	93.8	90.0	97.3	93.3	99.5	98.7	77.8	83.6	**95.0**	84.4	73.3	84.5
IBN-STR(stn)	ST+SK	**94.7**	91.0	98.0	94.0	**99.8**	98.6	**79.1**	**85.1**	94.6	85.4	**74.8**	85.6