StethoMe®
Home Stethoscope

POZNAN UNIVERSITY OF TECHNOLOGY

Using recurrences in time and frequency within U-net architecture for speech enhancement.

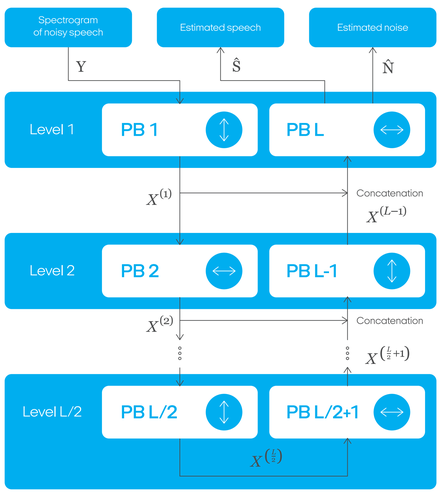
StethoMe® and Poznan University of Technology

Tomasz Grzywalski
Szymon Drgas

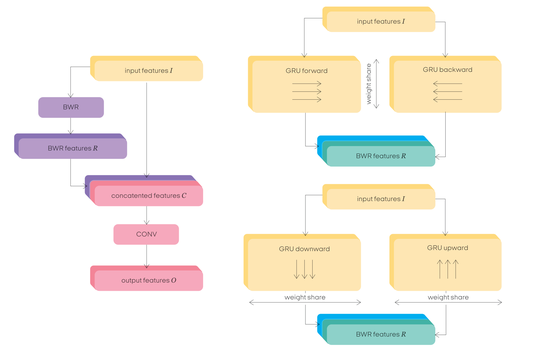
INTRODUCTION

When designing fully-convolutional neural network, there is a trade-off between receptive feld size, number of parameters and spatial resolution of features in deeper layers of the network. In this work we present a novel network design based on combination of many convolutional and recurrent layers that solves these dilemmas.

We introduce recurrent-convolutional (RC) pairs which can replace standard convolutional layers in U-net architecture. Each output of RC pair can potentially depend on a bigger context of its input than convolutional layer. This is for the cost of relatively small number of additional parameters. Moreover, this context can be enhanced at many depths of the encoder and decoder.



RECURRENT-CONVOLUTIONAL (RC) PAIRS



EXPERIMENTAL SETUP

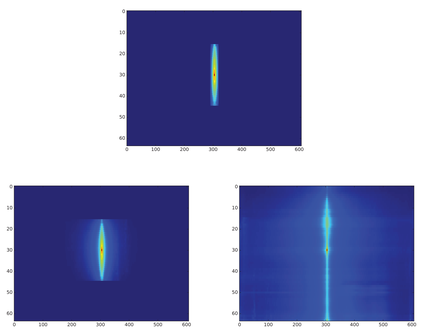
- The noisy speech examples were obtained by mixing TIMIT speech utterances with noise segments extracted from NOISEX-92 database.
- The training and test datasets contain 2000 and 192 utterances respectively.
- The speech enhancement quality was assessed for babble and factory noises, mixed with the speech utterances at SNR 0 dB.
- We used 64-channel mel spectrogram as audio features.

U-NETS

C48: 2D convolutional layer with 48 filters (all convolutional layers use 3x3 fters with ELU and batch normalization except for fnal layer which is always 1x1 with linear output),
RT16_C48: RC pair comprising BWR layer with 8 units per direction, iterating in time axis (weight sharing in frequency axis), and C48 layer (all recurrences were implemented using Gated Recurrent Units (GRUs) with gradient clipping above 100 and batch normalization),
RF16_C48: frequency counterpart to RT16_C48,
MP: max-pooling 2x2,
TC48: transposed convolution with filter size 6x6, stride 2 and crop 2 (an inverse of standard 3x3 convolution with "same" padding followed by 2x2 max-pooling).

- Training time: 100 epochs
- Adam optimizer
- Learning rate (nets with RC pairs): 0.01
- Learning rate (other networks): 0.001

RECEPTIVE FIELD VISUALIZATION



NAME	C48		C64		C48_C48		C64_MP		ALL_RC		ODD_RC	
Architecture [PB1 .. BP10]	C48	C2	C64	C2	C48_C48	C48_C2	C64	C2	R_F16_C48	R_T16_C2	R_F24_C48	R_T24_C2
	C48	C48	C64	C64	C48_C48	C48_C48	C64_MP	C64	R_T16_C48	R_F16_C48	C48	C48
	C48	C48	C64	C64	C48_C48	C48_C48	C64	C64_TC64	R_F16_C48	R_T16_C48	R_T24_C48	R_F24_C48
	C48	C48	C64	C64	C48_C48	C48_C48	C64_MP	C64	R_T16_C48	R_F16_C48	C48	C48
	C48	C48	C64	C64	C48_C48	C48_C48	C64	C64_TC64	R_F16_C48	R_T16_C48	R_F24_C48	R_T24_C48
Num. of parameters	230k		409k		460k		704k		328k		407k	
Receptive field	19x19		19x19		39x39		66x66		full signall		full signall	

RESULTS

Table: Factory noise

Name	SDR	SIR	SAR	STOI
FCLN IRM	7.4	12.8	8.9	0.74
RNN IRM	7.5	12.2	9.3	0.76
U-net C48	7.9	14.7	8.9	0.73
U-net C64	8.0	14.2	9.1	0.73
U-net C48_C48	8.2	14.3	9.3	0.76
U-net C64_MP	8.1	14.5	9.3	0.76
U-net ALL_RC IRM	8.2	14.8	9.3	0.80
U-net ALL_RC	**8.4**	**15.5**	9.4	**0.81**
U-net ODD_RC	**8.4**	15.0	**9.5**	**0.81**

Table: Babble noise

Name	SDR	SIR	SAR	STOI
FCLN IRM	5.4	8.9	8.5	0.71
RNN IRM	5.6	9.2	8.7	0.72
U-net C48	6.1	11.9	7.8	0.69
U-net C64	6.1	11.6	7.8	0.69
U-net C48_C48	6.2	10.5	8.7	0.71
U-net C64_MP	6.3	11.2	8.4	0.73
U-net ALL_RC IRM	6.7	12.0	8.5	0.76
U-net ALL_RC	**7.0**	**13.2**	8.5	**0.79**
U-net ODD_RC	6.9	12.3	**8.8**	0.78

CONCLUSIONS

In this work we proposed U-net-based neural network architectures in which recurrent-convolutional pairs are used at different levels. The obtained result show that the proposed architectures outperform the baseline models (FCLN, RNN, and U-nets without recurrences). The results of the performed experiments suggest, that U-net based architectures perform better for mapping -based rather than masking-based targets.

European Funds Smart Growth
The National Centre for Research and Development
European Union European Regional Development Fund