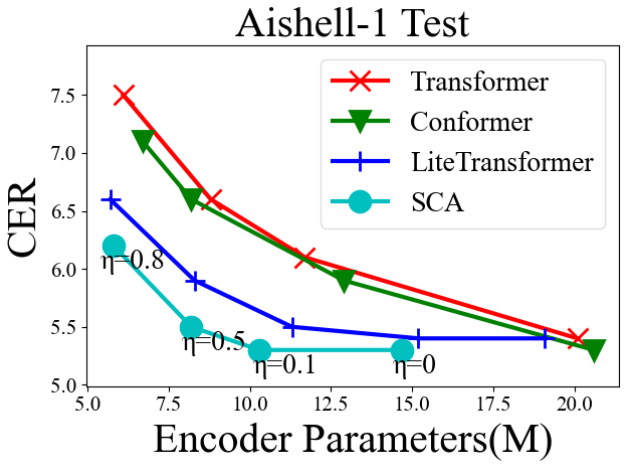
A. The result details of Figure 2

In our paper, we show the trade-off results between the performance and encoder parameter size of SCA and the human-designed baselines on Aishell-1 and HKUST. For SCA, we conduct experiments with $\eta \in \{0,0.1,0.5,0.8\}$, respectively. For the human-designed baselines, we adjust the block number to get results of different encoder parameters.

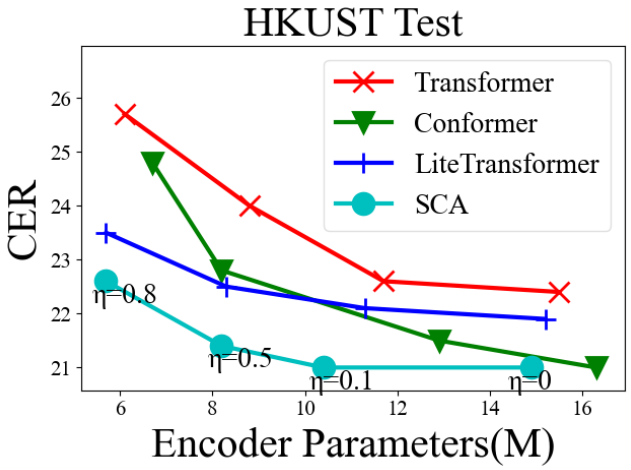
We further list the result details of our experiments as follow:

The result details of Aishell-1:



Architecture	η	#Block	#Params(M)	dev	test
Transformer	-	3	6.1	6.8	7.5
Transformer	-	5	8.8	6.0	6.6
Transformer	-	7	11.7	5.4	6.1
Transformer	-	13	20.1	4.9	5.4
Conformer	-	3	6.7	6.4	7.1
Conformer	-	4	8.2	6.0	6.6
Conformer	-	7	12.9	5.3	5.9
Conformer	-	12	20.6	4.8	5.2
Lite Transformer	-	10	5.7	6.0	6.6
Lite Transformer	-	17	8.3	5.4	5.9
Lite Transformer	-	25	11.3	4.9	5.5
Lite Transformer	-	35	15.2	4.9	5.4
Lite Transformer	-	45	19.1	4.9	5.4
SCA	0.8	5	5.8	5.6	6.2
SCA	0.5	8	8.2	4.9	5.5
SCA	0.1	8	10.3	4.8	5.2
SCA	0	8	14.7	4.8	5.2

The result details of HKUST:



For HKUST, we sample 5% from the train set as the dev data.

Architecture	η	#Block	#Params(M)	dev	test
Transformer	-	3	6.1	25.3	25.7
Transformer	-	5	8.8	23.6	24.0
Transformer	-	7	11.7	22.3	22.6
Transformer	-	10	15.5	22.1	22.4
Conformer	-	3	6.7	24.4	24.8
Conformer	-	4	8.2	22.5	22.8
Conformer	-	7	12.9	21.1	21.5
Conformer	-	9	16.3	20.0	21.0
Lite Transformer	-	10	5.7	23.0	23.5
Lite Transformer	-	17	8.3	22.3	22.5
Lite Transformer	-	25	11.3	21.9	22.1
Lite Transformer	-	35	15.2	21.4	21.9
SCA	0.8	5	5.6	22.3	22.6
SCA	0.5	8	8.2	20.9	21.3
SCA	0.1	8	10.4	20.3	21.0
SCA	0	8	14.5	20.1	21.0

B. More human-designed baselines

To demonstrate that SCA does obtain a better performance-resource trade-off than the human-designed architectures, more Transformer and Conformer baselines are conducted by adjusting the attention head number and convolution kernel size.

For Transformer, we set the attention head number as 4, 8, 16, denoted as `H4`, `H8`, `H16`, respectively.

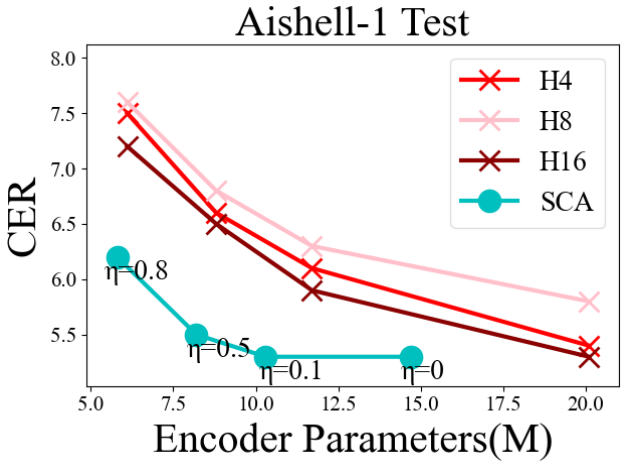
For Conformer, a set of architectures with diverse convolution kernel sizes and attention head numbers are provided, e.g., `CxHy` denotes the Conformer architecture with the convolution kernel size x and attention head number y.

The detailed results are provided as follows:

For more **Transformer** baselines:

Architecture	#Block	#Params(M)	dev	test
H4	3	6.1	6.8	7.5
H4	5	8.8	6.0	6.6
H4	7	11.7	5.4	6.1
H4	13	20.1	4.9	5.4
H8	3	6.1	7.0	7.6
H8	5	8.8	6.1	6.8
H8	7	11.7	5.6	6.3
H8	13	20.1	5.2	5.8
H16	3	6.1	6.5	7.2
H16	5	8.8	5.8	6.5
H16	7	11.7	5.2	5.9
H16	13	20.1	4.9	5.3

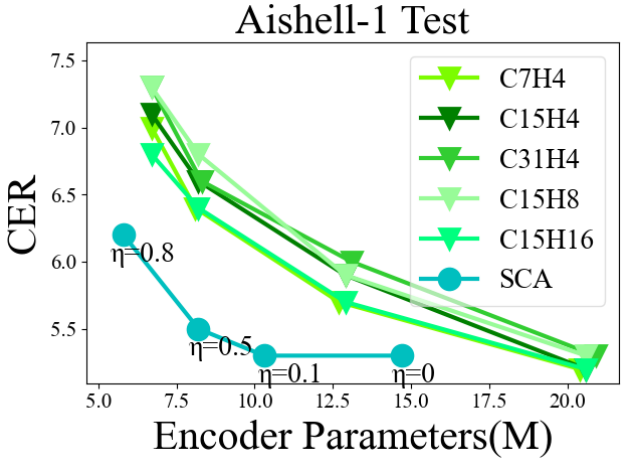
Also, we plot the performance-resource trade-off result.



For more **Conformer** baselines:

Architecture	#Block	#Params(M)	dev	test
C7H4	3	6.7	6.3	7.0
C7H4	4	8.1	5.8	6.4
C7H4	7	12.7	5.1	5.7
C7H4	12	20.4	4.8	5.2
C15H4	3	6.7	6.4	7.1
C15H4	4	8.2	6.0	6.6
C15H4	7	12.9	5.3	5.9
C15H4	12	20.6	4.8	5.2
C31H4	3	6.7	6.5	7.3
C31H4	4	8.3	6.0	6.6
C31H4	7	13.1	5.4	6.0
C31H4	12	20.9	4.8	5.3
C15H8	3	6.7	6.6	7.3
C15H8	4	8.2	6.1	6.8
C15H8	7	12.9	5.3	5.9
C15H8	12	20.6	4.9	5.3
C15H16	3	6.7	6.2	6.8
C15H16	4	8.2	5.8	6.4
C15H16	7	12.9	5.2	5.7
C15H16	12	20.6	4.8	5.2

And we plot the performance-resource trade-off result.



The Transformer or Conformer baselines with different architectures vary slightly, while SCA still can outperform the best of them, which demonstrates the effectiveness of SCA.

C. Ablation study on DSS

To explore the influence of DSS, we provide the comparison result against SCA without DSS, as a supplement of the ablation study of our paper.

	Aishell-1			HKUST		LibriSpeech100					Hub5' 00		
	#Params(M)	dev	test	#Params(M)	dev	#Params(M)	dev clean	dev other	test clean	test other	#Params(M)	swbd	callhm
SCA w/o DSS	10.36	5.3	5.8	11.03	21.8	10.64	9.3	21.6	9.6	22.0	12.38	8.1	16.4
SCA with DSS	10.31	**4.8**	**5.2**	10.39	**21.0**	10.55	**9.0**	**21.3**	**9.2**	**21.5**	12.35	**7.9**	**15.7**

Even though the searched architectures have similar parameter numbers, SCA with DSS consistently outperforms those without DSS, which illustrates the necessity of DSS.