Appendix

	Aishell-1				HKUST			LibriSpeech100				Hub5' 00 (SWBD)			
	#Params (M)	FLOPs (M)	dev	test	#Params (M)	FLOPs (M)	test	#Params (M)	FLOPs (M)	test clean	test other	#Params (M)	FLOPs (M)	swbd	callhm
Transformer	19.3	978	4.9	5.4	19.3	978	21.9	19.3	978	9.2	22.1	19.3	978	8.1	16.3
Conformer	19.0	960	**4.8**	5.3	19.0	960	**20.8**	19.0	960	**8.2**	**20.3**	19.0	960	8.0	**16.0**
LiteTransformer	18.6	897	5.4	6.1	18.6	897	21.6	18.6	897	8.7	21.9	18.6	897	9.2	17.6
LightConv	18.7	942	5.6	6.4	18.7	942	23.2	18.7	942	9.5	25.0	18.7	942	9.6	18.5
Random Search	12.2	613	5.3	6.0	12.3	622	22.1	12.6	651	10.9	23.4	12.3	627	8.5	17.2
NAS-SCAE	**8.4**	**424**	**4.8**	**5.2**	**8.9**	**457**	21.0	**8.7**	**447**	8.3	20.5	**8.6**	**439**	**7.9**	16.1

In our paper, we show the results(CER/WER) of the human-designed baselines and NAS-SACE on four Mandarin and English datasets.

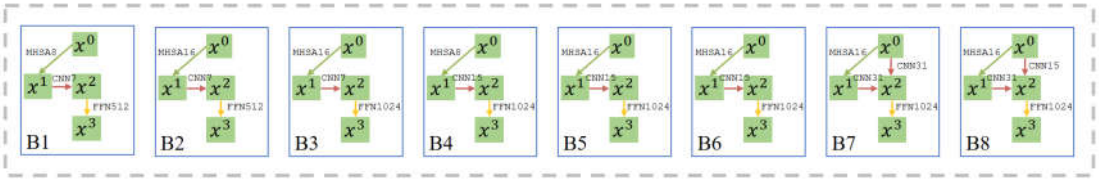
For Transformer, Conformer and LightConv, we set the attention dim as 256, head number as 4, and hidden dimension as 2048.

For LiteTransformer, we set the attention dim as 256, head number as 4, and hidden dimension as 1024.

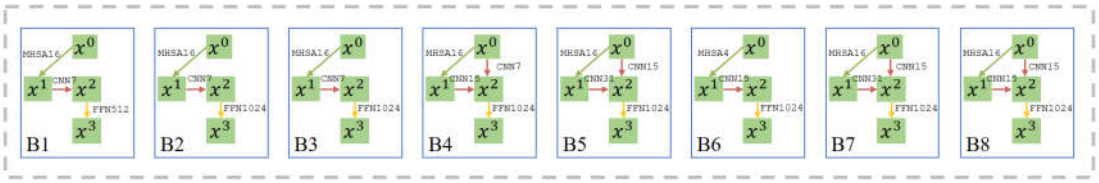
For the decoder of the LAS framework, we set the attention dim as 256, head number as 4, and hidden dimension as 2048.

For NAS-SCAE, we show the searched encoders on four datasets as follow:

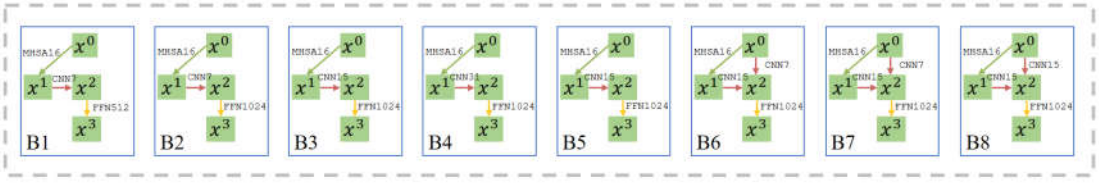
Aishell-1:



HKUST:



LibriSpeech100:



SWBD: