



SCAN ME

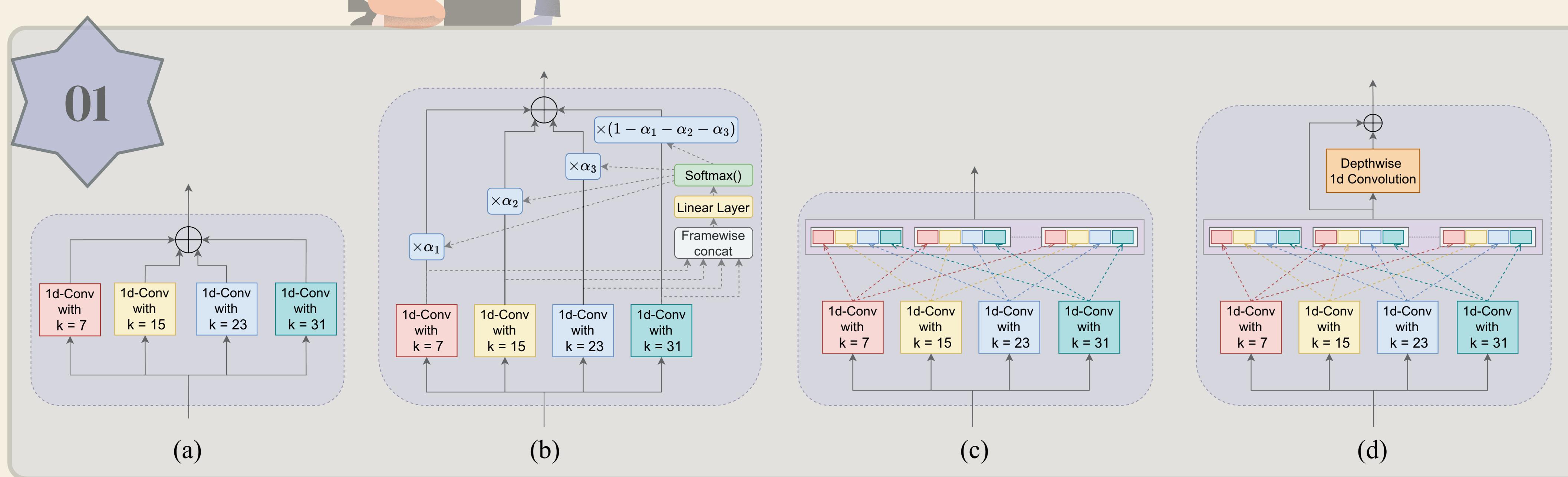
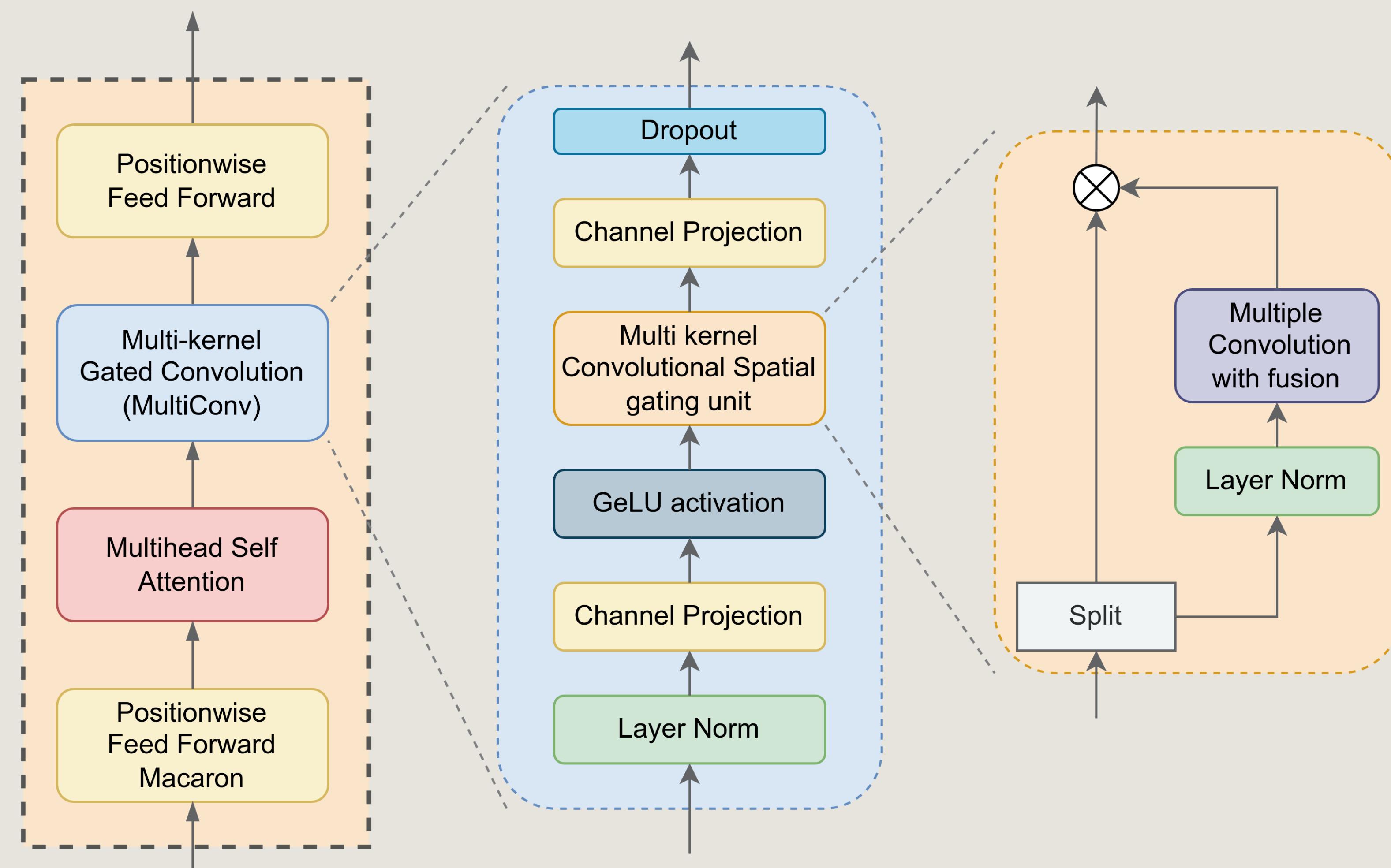
# MULTI-CONVFORMER:

## Extending Conformer with Multiple Convolution Kernels



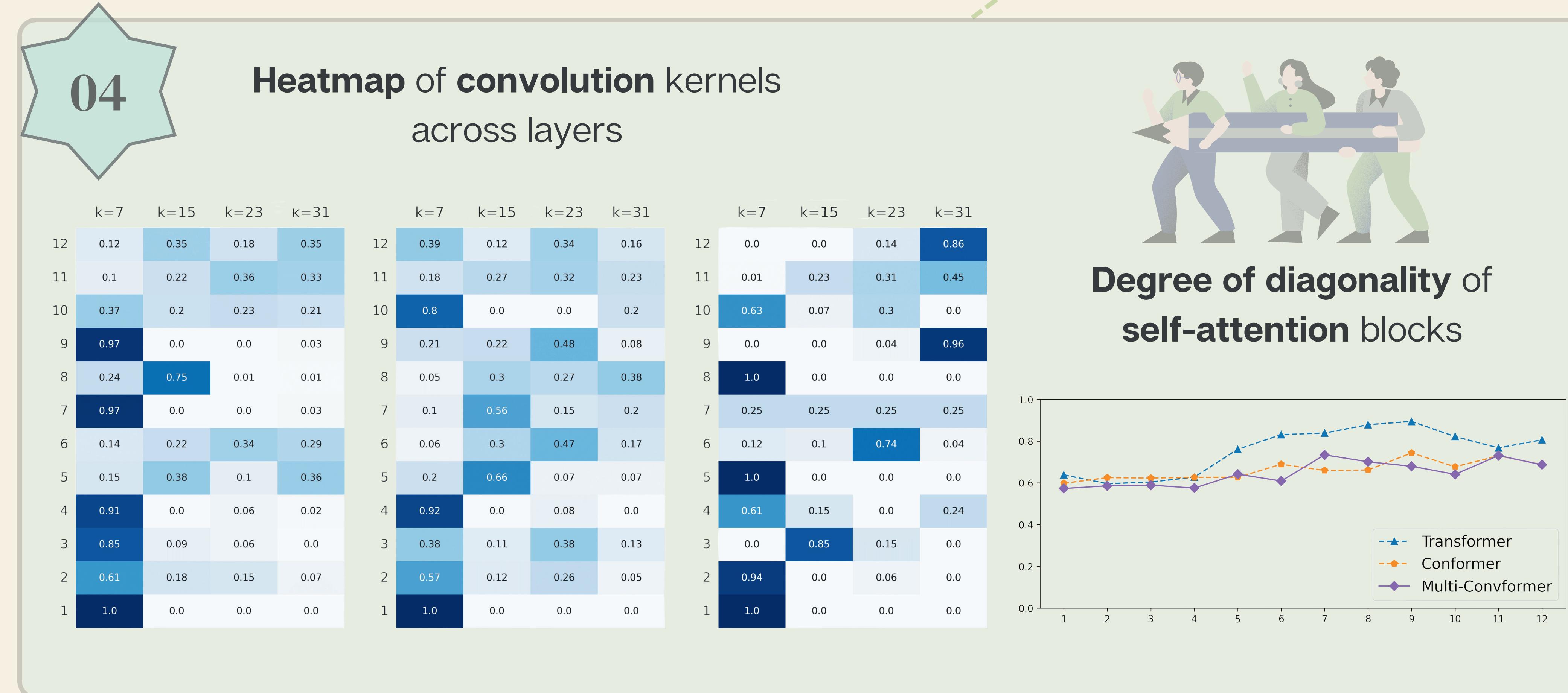
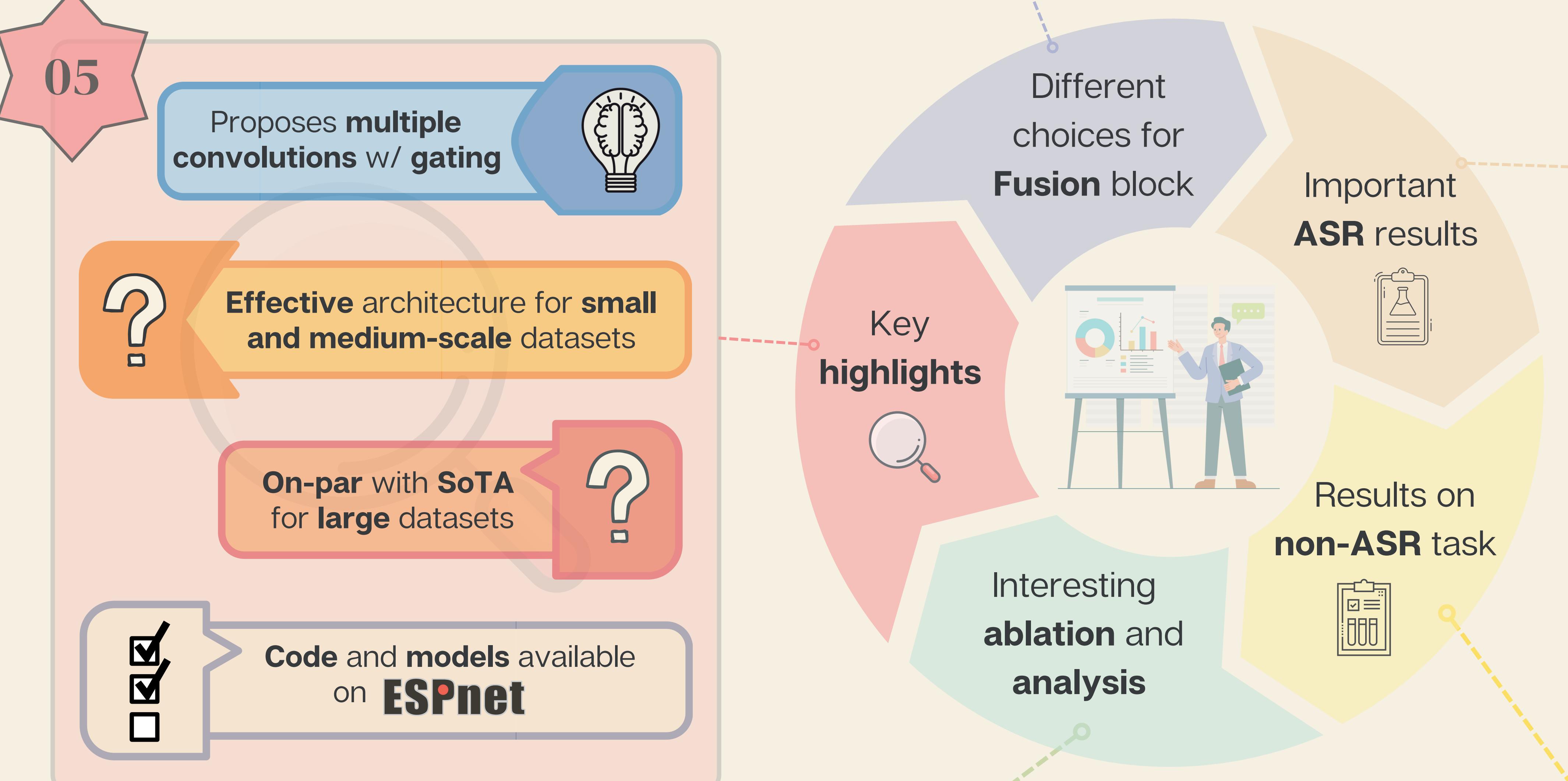
\*Darshan Prabhu, \*Yifan Peng, \*Preethi Jyothi and \*Shinji Watanabe

\*IIT Bombay and \*CMU



### Word Error Rates (WERs) on LibriSpeech, AISHELL and TEDLIUM2

METHOD	LIBRISPEECH 100		TEDLIUM-2		AISHELL	
	TEST CLEAN	TEST OTHER	TEST	TEST	TEST	TEST
ATTENTION ENCODER-DECODER (AED) MODELS						
CONFORMER	6.89		17.27	7.65	4.63	
CGConv	6.36		17.18	7.44	4.55	
E-BRANCHFORMER	6.50		17.06	7.34	4.47	
OURS *	<b>6.18</b>	<b>17.00</b>	<b>7.27</b>	<b>4.46</b>		
ENCODER ONLY (PURE CTC) MODELS						
CONFORMER	9.76		23.10	8.80	6.59	
CGConv	9.66		23.79	8.78	6.33	
E-BRANCHFORMER	10.05		23.87	8.78	6.04	
OURS *	<b>9.47</b>	<b>23.09</b>	<b>8.44</b>	<b>6.01</b>		
RNN TRANSDUCER (RNN-T) MODELS						
CONFORMER	7.19		18.70	7.96	5.22	
CGConv	6.95		18.18	7.72	5.24	
E-BRANCHFORMER	<b>6.87</b>	<b>18.09</b>		7.71	5.17	
OURS *	7.05		18.23	<b>7.67</b>	<b>5.05</b>	
WITHOUT LM						
METHOD	TEST CLEAN		TEST OTHER		TEST CLEAN	
	TEST CLEAN	TEST OTHER	TEST CLEAN	TEST OTHER	TEST CLEAN	TEST OTHER
CONFORMER	2.16		4.74		<b>1.84</b>	3.95
BRANCHFORMER	2.25		4.83		1.93	4.00
E-BRANCHFORMER	<b>2.14</b>	<b>4.55</b>			1.85	<b>3.71</b>
OURS *	2.15		4.69		1.89	3.92



**03**

Spoken Language Understanding on SLURP

METHOD	INTENT CLASSIFICATION		ENTITY RECOGNITION (F1 SCORE)
	VAL ACCURACY	TEST ACCURACY	
CONFORMER	87.4 %	86.7 %	77.2
E-BRANCHFORMER	88.3 %	<b>87.4 %</b>	78.5
OURS *	<b>88.8 %</b>	<b>87.4 %</b>	<b>78.9</b>