

2 layer GRU:

Layer (type)	Output Shape	Param #
bidirectional_1 (Bidirectional)	(None, 10, 256)	145920
dropout (Dropout)	(None, 10, 256)	0
average_pooling1d (AveragePooling1D)	(None, 5, 256)	0
bidirectional_2 (Bidirectional)	(None, 5, 256)	296448
dropout_1 (Dropout)	(None, 5, 256)	0
average_pooling1d_1 (AveragePooling1D)	(None, 2, 256)	0
dense (Dense)	(None, 2, 1)	257
Total params: 442,625		
Trainable params: 442,625		
Non-trainable params: 0		

STM32F769I-DK

```
Results for "gru", 16 inferences @216MHz/216MHz (complexity: 2930198 MACCs)
duration      : 88.912 ms (average)
CPU cycles    : 19205119 (average)
CPU Workload  : 8% (duty cycle = 1s)
cycles/MACC   : 6.55 (average for all layers)
used stack    : 744 bytes
used heap     : 0:0 0:0 (req:allocated,req:released) max=0 cur=0 (cfg=3)
observer res   : 184 bytes used from the heap (10 c-nodes)

Inference time by c-node
  kernel : 88.899ms (time passed in the c-kernel fcts)
  user   : 0.006ms (time passed in the user cb)

c_id  type      id      time (ms)
-----
0     GRU        0      22.894  25.75 %
1     GRU        0      22.886  25.74 %
2     CONCAT     0       0.026   0.03 %
3     POOL       2       0.081   0.09 %
4     GRU        3      21.475  24.16 %
5     GRU        3      21.468  24.15 %
6     CONCAT     3       0.014   0.02 %
7     POOL       5       0.033   0.04 %
8     DENSE      6       0.015   0.02 %
9     NL         6       0.002   0.00 %
-----
                        88.899 ms
```

c_id	name (*_layer)	id	layer_type	macc	rom	tensors	shape (array id)
0	backward_gru_1	0	gru	724480	291840	I: input_0_output S: backward_gru_1_scratch0 W: backward_gru_1_kernel W: backward_gru_1_recurrent W: backward_gru_1_bias O: backward_gru_1_output0	(1,10,60) (13) (60,384) (24) (128,384) (25) (384,2) (26) (1,10,128) (14)
1	forward_gru_1	0	gru	724480	291840	I: input_0_output S: forward_gru_1_scratch0 W: forward_gru_1_kernel W: forward_gru_1_recurrent W: forward_gru_1_bias O: forward_gru_1_output0	(1,10,60) (13) (60,384) (27) (128,384) (28) (384,2) (0) (1,10,128) (15)
2	bidirectional_1	0	concat	0	0	I: forward_gru_1_output0 I: backward_gru_1_output0 O: bidirectional_1_output	(1,10,128) (15) (1,10,128) (14) (1,10,256) (16)
3	average_pooling1d	2	pool	2560	0	I: bidirectional_1_output O: average_pooling1d_output	(1,10,256) (16) (1,5,256) (17)
4	backward_gru_2	3	gru	738560	592896	I: average_pooling1d_output S: backward_gru_2_scratch0 W: backward_gru_2_kernel W: backward_gru_2_recurrent W: backward_gru_2_bias O: backward_gru_2_output0	(1,5,256) (17) (256,384) (1) (128,384) (2) (384,2) (3) (1,5,128) (18)
5	forward_gru_2	3	gru	738560	592896	I: average_pooling1d_output S: forward_gru_2_scratch0 W: forward_gru_2_kernel W: forward_gru_2_recurrent W: forward_gru_2_bias O: forward_gru_2_output0	(1,5,256) (17) (256,384) (4) (128,384) (5) (384,2) (6) (1,5,128) (19)
6	bidirectional_2	3	concat	0	0	I: forward_gru_2_output0 I: backward_gru_2_output0 O: bidirectional_2_output	(1,5,128) (19) (1,5,128) (18) (1,5,256) (20)
7	average_pooling1d_1	5	pool	1024	0	I: bidirectional_2_output O: average_pooling1d_1_output	(1,5,256) (20) (1,2,256) (21)
8	dense_dense	6	dense	514	1028	I: average_pooling1d_1_output W: dense_dense_weights W: dense_dense_bias O: dense_dense_output	(1,2,256) (21) (256,1) (7) (1,) (8) (1,2,1) (22)
9	dense	6	nl	20	0	I: dense_dense_output O: dense_output	(1,2,1) (22) (1,2,1) (23)

Number of operations per c-layer

c_id	m_id	name (type)	#op	type	#param	sparsity
0	0	backward_gru_1 (gru)	724,480	smul_f32_f32	72,960	0.0000
1	0	forward_gru_1 (gru)	724,480	smul_f32_f32	72,960	0.0000
2	0	bidirectional_1 (concat)	0	op_f32_f32		
3	2	average_pooling1d (pool)	2,560	op_f32_f32		
4	3	backward_gru_2 (gru)	738,560	smul_f32_f32	148,224	0.0000
5	3	forward_gru_2 (gru)	738,560	smul_f32_f32	148,224	0.0000
6	3	bidirectional_2 (concat)	0	op_f32_f32		
7	5	average_pooling1d_1 (pool)	1,024	op_f32_f32		
8	6	dense_dense (dense)	514	smul_f32_f32	257	0.0000
9	6	dense (nl)	20	op_f32_f32		
total			2,930,198		442,625	0.0000

CNN (sincnet+ 2 conv1d)

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 3200, 1)]	0
sinc_conv_fast_1 (SincConvFast)	(None, 40, 80)	360
layer_norm (LayerNorm)	(None, 40, 80)	160
leaky_re_lu (LeakyReLU)	(None, 40, 80)	0
max_pooling1d (MaxPooling1D)	(None, 20, 80)	0
conv1d (Conv1D)	(None, 16, 60)	24060
leaky_re_lu_1 (LeakyReLU)	(None, 16, 60)	0
max_pooling1d_1 (MaxPooling1D)	(None, 8, 60)	0
conv1d_1 (Conv1D)	(None, 4, 60)	18060
leaky_re_lu_2 (LeakyReLU)	(None, 4, 60)	0
max_pooling1d_2 (MaxPooling1D)	(None, 2, 60)	0
dense (Dense)	(None, 2, 50)	3050
Total params: 45,690		
Trainable params: 45,690		
Non-trainable params: 0		

STM32F769I-DK

```
Results for "cnn", 16 inferences @216MHz/216MHz (complexity: 1800405 MACC)
duration      : 95.917 ms (average)
CPU cycles    : 20718263 (average)
CPU Workload  : 9% (duty cycle = 1s)
cycles/MACC   : 11.50 (average for all layers)
used stack    : 656 bytes
used heap     : 0:0 0:0 (req:allocated,req:released) max=0 cur=0 (cfg=3)
observer res  : 376 bytes used from the heap (22 c-nodes)
```

```
Inference time by c-node
kernel : 95.886ms (time passed in the c-kernel fcts)
user   : 0.021ms (time passed in the user cb)
```

c_id	type	id	time (ms)	
0	CONV2D	1	66.101	68.94 %
1	REDUCE	3	0.399	0.42 %
2	ELTWISE	3	0.052	0.06 %
3	ELTWISE	9	0.996	1.04 %
4	ELTWISE	4	1.051	1.10 %
5	REDUCE	5	0.399	0.42 %
6	ELTWISE	5	0.052	0.06 %
7	NL	6	0.006	0.01 %
8	ELTWISE	7	0.053	0.06 %
9	ELTWISE	8	0.057	0.06 %
10	ELTWISE	10	1.003	1.05 %
11	BN	11	0.550	0.57 %
12	NL	12	0.205	0.21 %
13	POOL	14	0.111	0.12 %
14	CONV2D	17	20.349	21.22 %
15	NL	19	0.067	0.07 %
16	POOL	21	0.037	0.04 %
17	CONV2D	24	3.947	4.12 %
18	NL	26	0.018	0.02 %
19	POOL	28	0.010	0.01 %
20	DENSE	40	0.323	0.34 %
21	NL	41	0.089	0.09 %

95.886 ms

Number of operations per c-layer

c_id	m_id	name (type)	#op	type	#param	sparsity
0	1	conv2d_1 (conv2d)	1,283,280	smul_f32_f32	32,160	0.0025
1	3	reduce_3 (reduce)	3,200	smul_f32_f32		
2	3	reduce_3_Mul (eltwise/mul)	40	op_f32_f32	1	0.0000
3	9	eltwise_9 (eltwise/sub)	3,200	op_f32_f32		
4	4	eltwise_4 (eltwise/squared_diff)	16,000	op_f32_f32		
5	5	reduce_5 (reduce)	3,200	smul_f32_f32		
6	5	reduce_5_Mul (eltwise/mul)	40	op_f32_f32	1	0.0000
7	6	nl_6 (nl)	480	op_f32_f32		
8	7	eltwise_7 (eltwise/sum)	40	op_f32_f32	1	0.0000
9	8	eltwise_8 (eltwise/div)	5	op_f32_f32	1	0.0000
10	10	eltwise_10 (eltwise/mul)	3,200	op_f32_f32		
11	11	eltwise_11 (bn)	6,400	smul_f32_f32	160	0.5000
12	12	nl_12 (nl)	9,600	op_f32_f32		
13	14	pool_14 (pool)	3,200	op_f32_f32		
14	17	conv2d_17 (conv2d)	384,060	smul_f32_f32	24,060	0.0025
15	19	nl_19 (nl)	2,880	op_f32_f32		
16	21	pool_21 (pool)	960	op_f32_f32		
17	24	conv2d_24 (conv2d)	72,060	smul_f32_f32	18,060	0.0033
18	26	nl_26 (nl)	720	op_f32_f32		
19	28	pool_28 (pool)	240	op_f32_f32		
20	40	gemm_38 (dense)	6,100	smul_f32_f32	3,050	0.0164
21	41	nl_41 (nl)	1,500	op_f32_f32		
total			1,800,405		77,494	0.0043