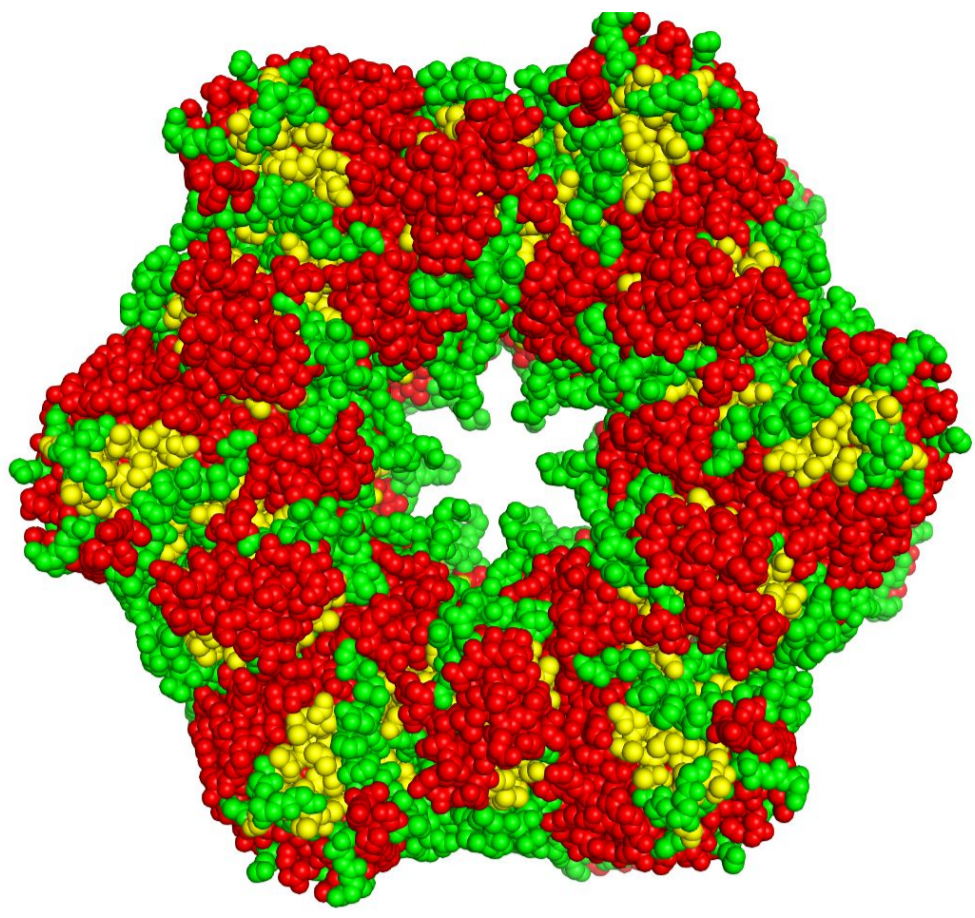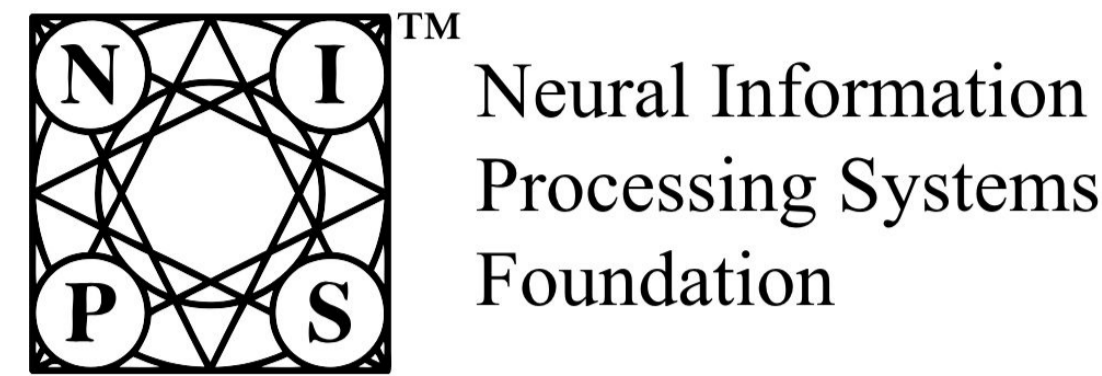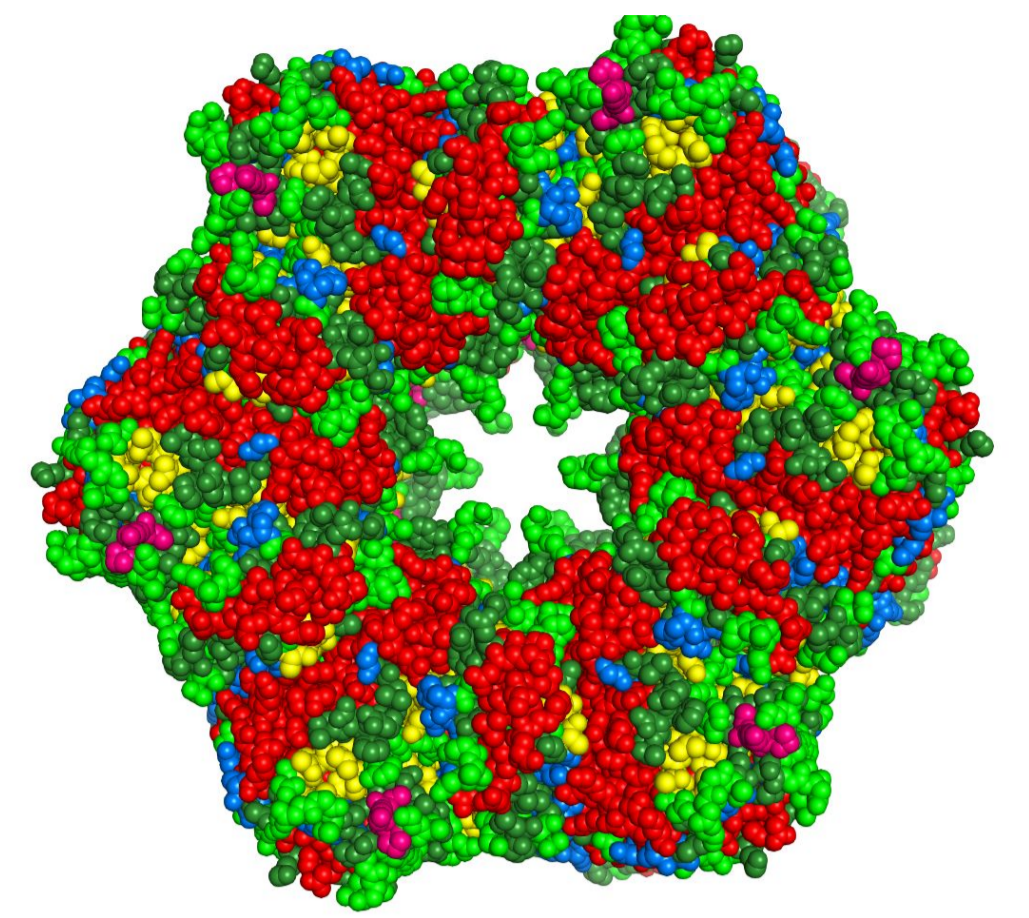# High Quality Protein Q8 Secondary Structure Prediction by Diverse Neural Network Architectures

Iddo Drori, Isht Dwivedi, Pranav Shrestha, Jeffrey Wan, Yueqi Wang, Yunchu He, Anthony Mazza, Hugh Krogh-Freeman, Dimitri Leggas, Kendal Sandridge, Linyong Nan, Kaveri Thakoor, Chinmay Joshi, Sonam Goenka, Chen Keasar, Itsik Pe'er

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Columbia | ENGINEERING The Fu Foundation School of Engineering and Applied Science

Neural Information Processing Systems Foundation

Q3 of 1F52

Data, Models, and Code: github.com/idrori/cu-ssp

Q8 of 1F52

Diverse Neural Network Architectures

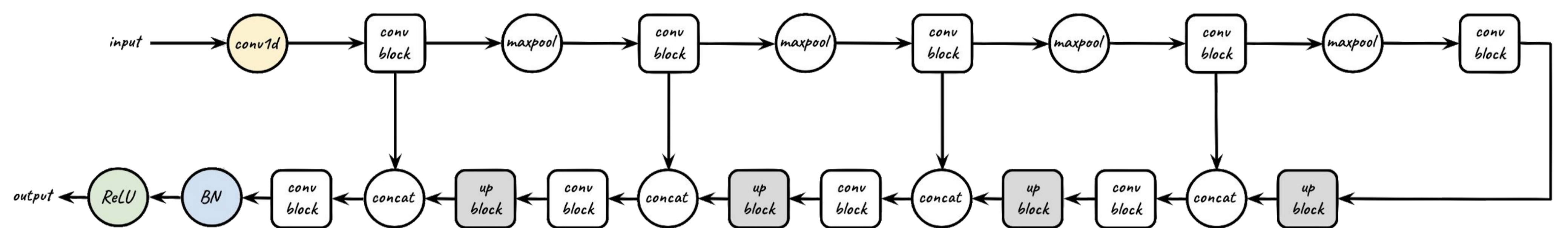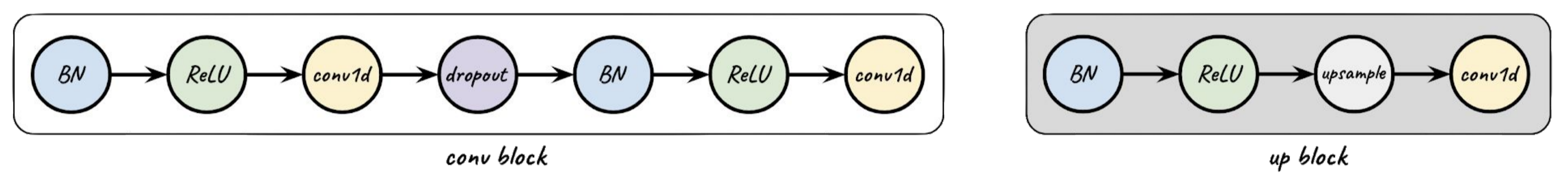|	
Ensemble	70.7
Bidirectional GRU with convolution blocks	69.8
U-Net with convolution blocks	69.2
Temporal convolutional network	68.7
Bidirectional LSTMs with attention	68.4
Convolutions and bidirectional LSTM	67.8
Bidirectional GRUs	67.4

Q8 mean accuracy of our models and their ensemble on CB513

Temporal Convolutional Network
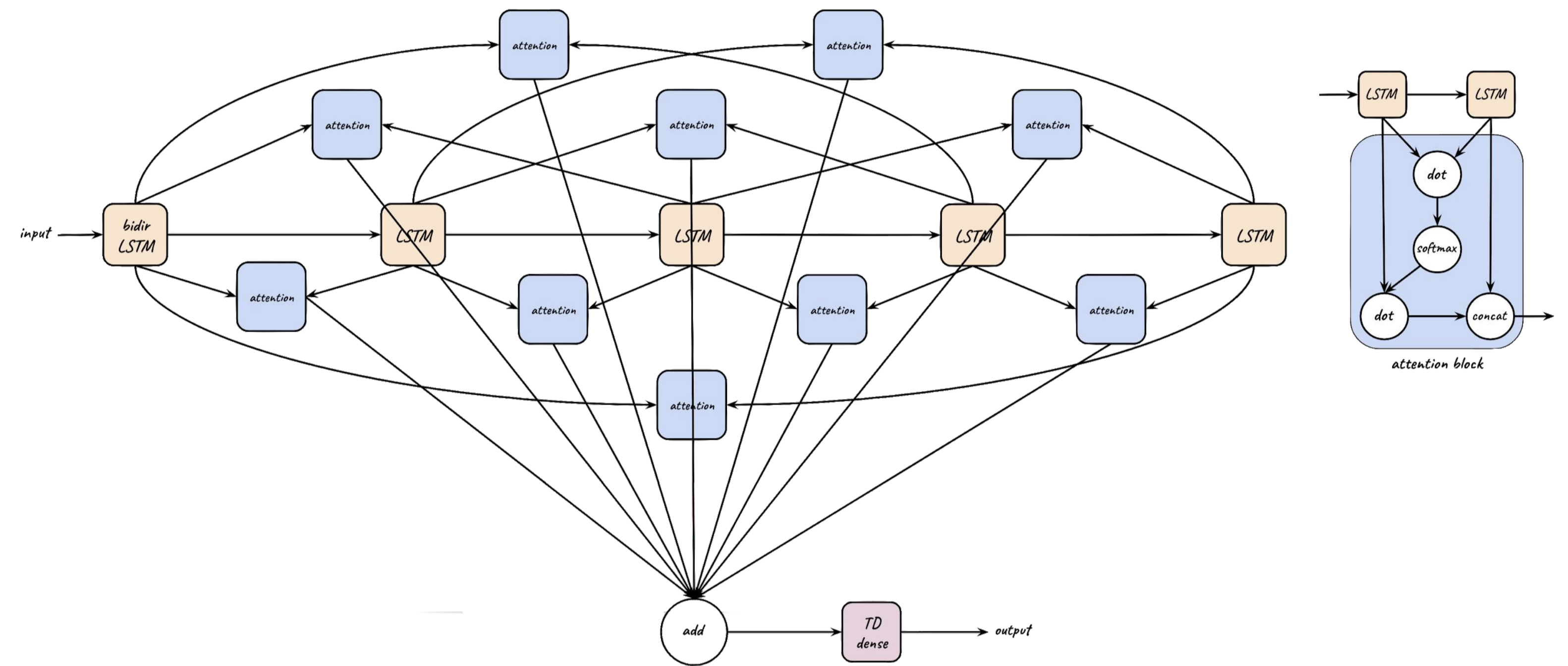
Bidirectional LSTM with attention

conv block

up block

U-Net with convolution blocks

conv block

Bidirectional GRU with convolution blocks

conv block
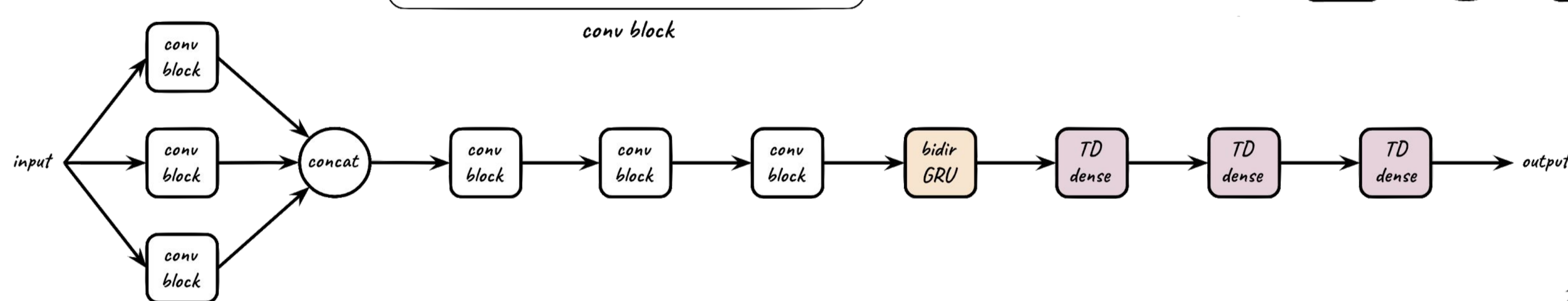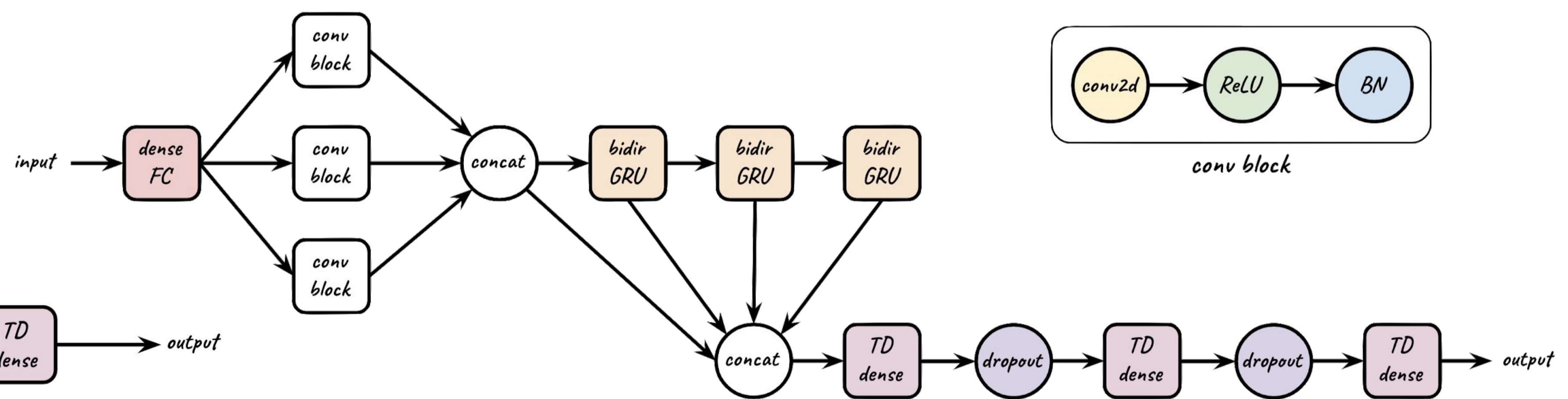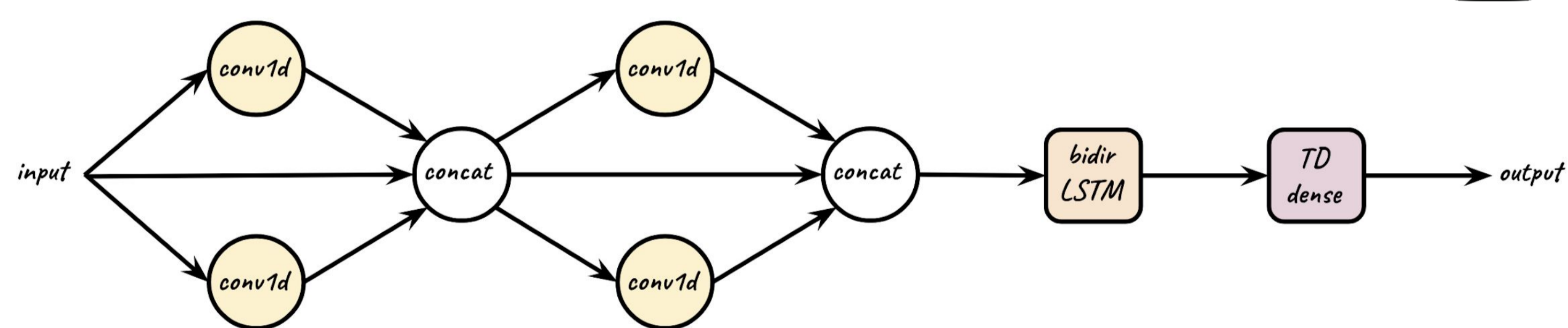
Convolution and bidirectional LSTM

Bidirectional GRUs

Results

Models on CB513	Best Single	Ensemble
MUFOLD-SS [11]	70.5	70.6
NCCNN [3]	70.3	71.4
biRNN-CRF [17]	69.4	70.9
DeepMSCNN [2]	70.0	70.6
DCRNN [21]	69.4	69.7
BLSTM [28]	67.4	N/A
GSN [40]	66.4	N/A
DeepCNF [33]	N/A	68.3
Ours	69.8	70.7

Models on CB6133	Best Single	Ensemble
GSN [40]	72.1	N/A
DCRNN [21]	N/A	73.2
biRNN-CRF [17]	73.4	74.8
CRRNN [38]	N/A	74
Ours	75.4	76.3

	L	B	E	G	I	H	S	T
L	11,828	618	1,880	629	4	738	3,192	1,619
B	7	31	6	0	0	3	4	0
E	3,167	316	15,419	234	2	334	997	565
G	134	8	24	851	0	233	109	328
I	0	0	0	0	0	0	0	0
H	762	77	216	777	22	24,126	554	1,585
S	871	49	201	78	0	77	2,039	502
T	1,151	82	270	563	2	646	1,421	5,414

Confusion matrix for each of the Q8 structures on CB513 dataset