CSE549

DNN Applications to Bioinformatics Part 2: Interpretable Methods

Sael Lee

Department of Computer Science,

SUNY Korea, Incheon 21985, Korea

Deep Motif Dashboard

Goal: Motif visualization in Transcription Factor binding prediction
Models Used: convolutional, recurrent, and convolutional-recurrent networks

Jack Lanchantin, Ritambhara Singh, Beilun Wang, and Yanjun Qi. 2016. Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks. In *Pacific Symposium on Biocomputing*, 1–11.

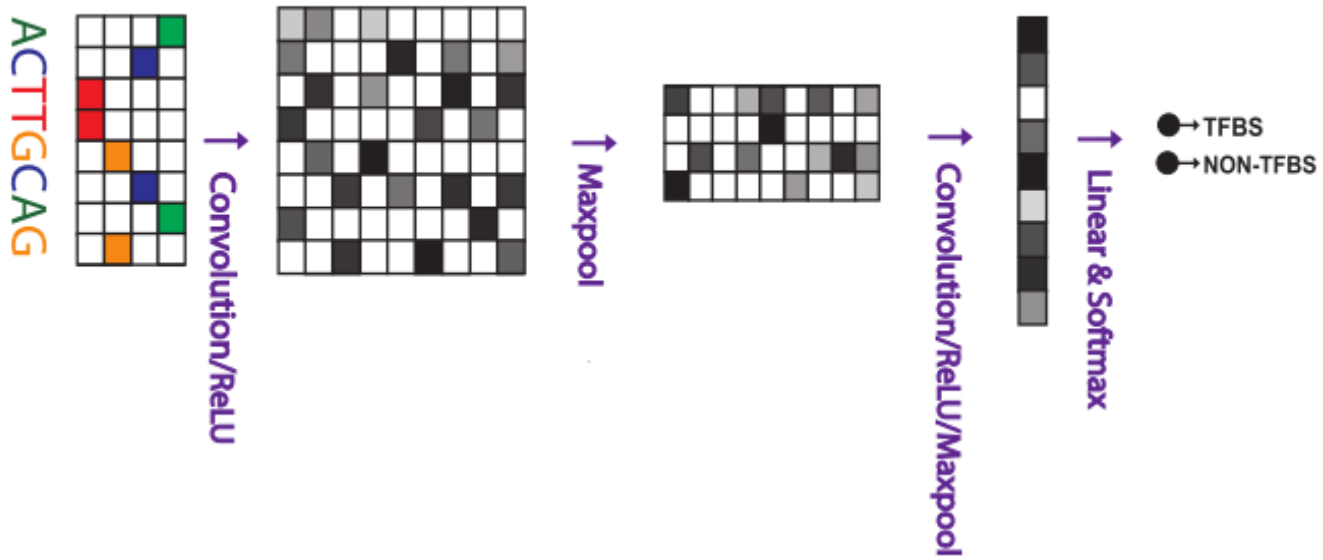
Models and Visualization Strategies

- Three Models
  - CNN
  - RNN
  - CNN-RNN (best performing)
- Visualization
  - Measuring nucleotide importance with **Saliency Maps**.
  - Measuring critical sequence positions for the classifier using **Temporal Output Scores**.
  - Generating class-specific motif patterns with **Class Optimization**.

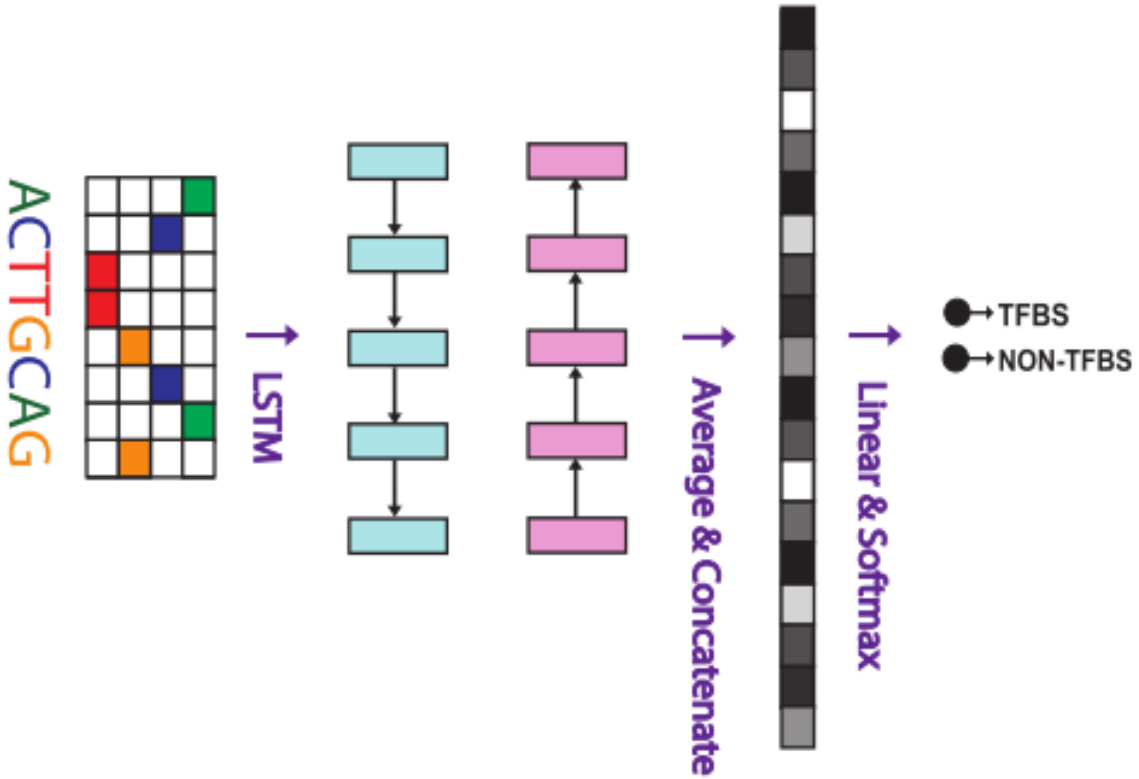
Models - Common settings

- Input: one-hot encoded matrix of raw sequence
- Output
  - Output vector: linearly fed to a softmax function
  - Learns the mapping from the hidden space to the output class label space $C \in [+1,-1]$.
    - Probability indicating whether an input is a positive or a negative binding site (binary classification task).
- Training
  - Parameters: trained end-to-end by minimizing the negative log-likelihood over the training set.
  - Loss function optimization stochastic gradient algorithm Adam
  - Mini-batch size of 256 sequences.
  - Regularization - Dropout.

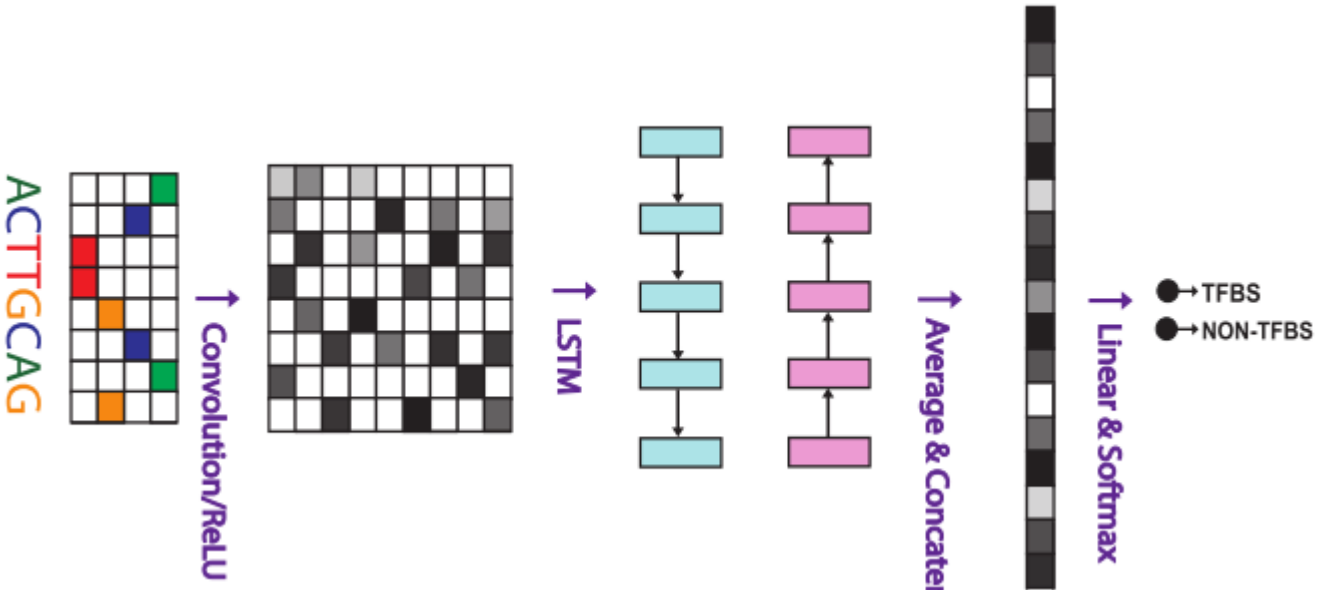
CNN Model


TFBS
NON-TFBS
Linear & Softmax
Convolution/ReLU/Maxpool
Maxpool
Convolution/ReLU
ACTTGCAG

RNN Model


TFBS
NON-TFBS
Linear & Softmax
Average & Concatenate
LSTM
ACTTGCAG

CNN-RNN Model



Saliency Map of CNN

Problem: Given a sequence X_0 of length $|X_0|$, and class $c \in C$, a DNN model provides a score function $S_c(X_0)$. We rank the nucleotides of X_0 based on their influence on the score $S_c(X_0)$.
Challenge: Since $S_c(X)$ is a non-linear function of X, it is hard to directly determine the influence of each nucleotide of X on Sc.
Solution: Approximated $S_c(X)$ as a linear function by computing the first-order Taylor expansion

$$S_c(X) \approx w^T X + b = \sum_{i=1}^{|X|} w_i x_i + b$$

<- a weighted sum of the input nucleotides

where w is the derivative of S_c with respect to the sequence variable X at the point X_0 (w_i, indicates the influence of that nucleotide position)

$$w = \left.\frac{\partial S_c}{\partial X}\right|_{X_0} = saliency\ map$$

Approach is similar to the methods used on images by Simonyan et al. 2013 and Baehrens et al. 2010.

Saliency Map of CNN cont.

- Derivative is simply one step of backpropagation in the DNN

- **Getting derivative values** of actual sequence:
  - Approach: pointwise multiplication of the saliency map with the one-hot encoded sequence
  - Interpretation: the influence value of the character at each position on the output score.

- **Visualize** important each character (saliency map):
  - Approach: element-wise magnitude of the resulting derivative vector regardless of derivative direction.
  - Interpretation: indicates which nucleotides need to be changed the least in order to affect the class score the most.

Temporal Output Scores for RNN

- Description:
  - Visualize the output scores at each timestep (position) of a sequence.
- Assumption:
  - An imaginary time direction running from left to right
  - Each position in the sequence is a timestep
- Determine the TOS
  - The input series is constructed by using subsequences of an input X running along the imaginary time coordinate, where the subsequences start from just the first nucleotide (position), and ends with the entire sequence X.
  - TOS is calculated for each subsequences and visualized

Class-Specific Visualization

- Goal: Find the best sequence which maximizes the probability of a positive TFBS, which we call class optimization.
- Optimize $\arg\max_X S_+(X) + \lambda \|X\|_2^2$

  where $S_+(X)$ is the probability (or score) of an input sequence X (matrix) being a positive TFBS computed by the softmax equation of our trained DNN model for a specific TF.

Three Motif Extraction

For each of the three visualization methods

1. Saliency map:
   - From each positive test sequence, select the contiguous length-9 subsequence that achieves the highest sum of contiguous length-9 saliency map values.
2. Temporal Output Scores:
   - For each positive test sequence, select the length-9 subsequence that shows the strongest score change from negative to positive output score.
3. Class-Specific
   - For each different TF, directly use the class-optimized sequence as a motif.

Results

- Training: 30,819 sequences (with an even positive/negative split), and each sequence consists of 101 DNA-base characters (A,C,G,T).
- Testing: Every dataset has 1,000 sequences

Table 1: Variations of DNN Model Hyperparameters

Model	**Conv. Layers**	**Conv. Size** (n_{out})	**Conv. filter Sizes** (k)	**Conv. Pool Size** (m)	**LSTM Layers**	**LSTM Size** (d)
Small RNN	N/A	N/A	N/A	N/A	1	16
Medium RNN	N/A	N/A	N/A	N/A	1	32
Large RNN	N/A	N/A	N/A	N/A	2	32
Small CNN	2	64	9,5	2	N/A	N/A
Medium CNN	3	64	9,5,3	2	N/A	N/A
Large CNN	4	64	9,5,3,3	2	N/A	N/A
Small CNN-RNN	1	64	5	N/A	2	32
Medium CNN-RNN	1	128	9	N/A	1	32
Large CNN-RNN	2	128	9,5	2	1	32

Results

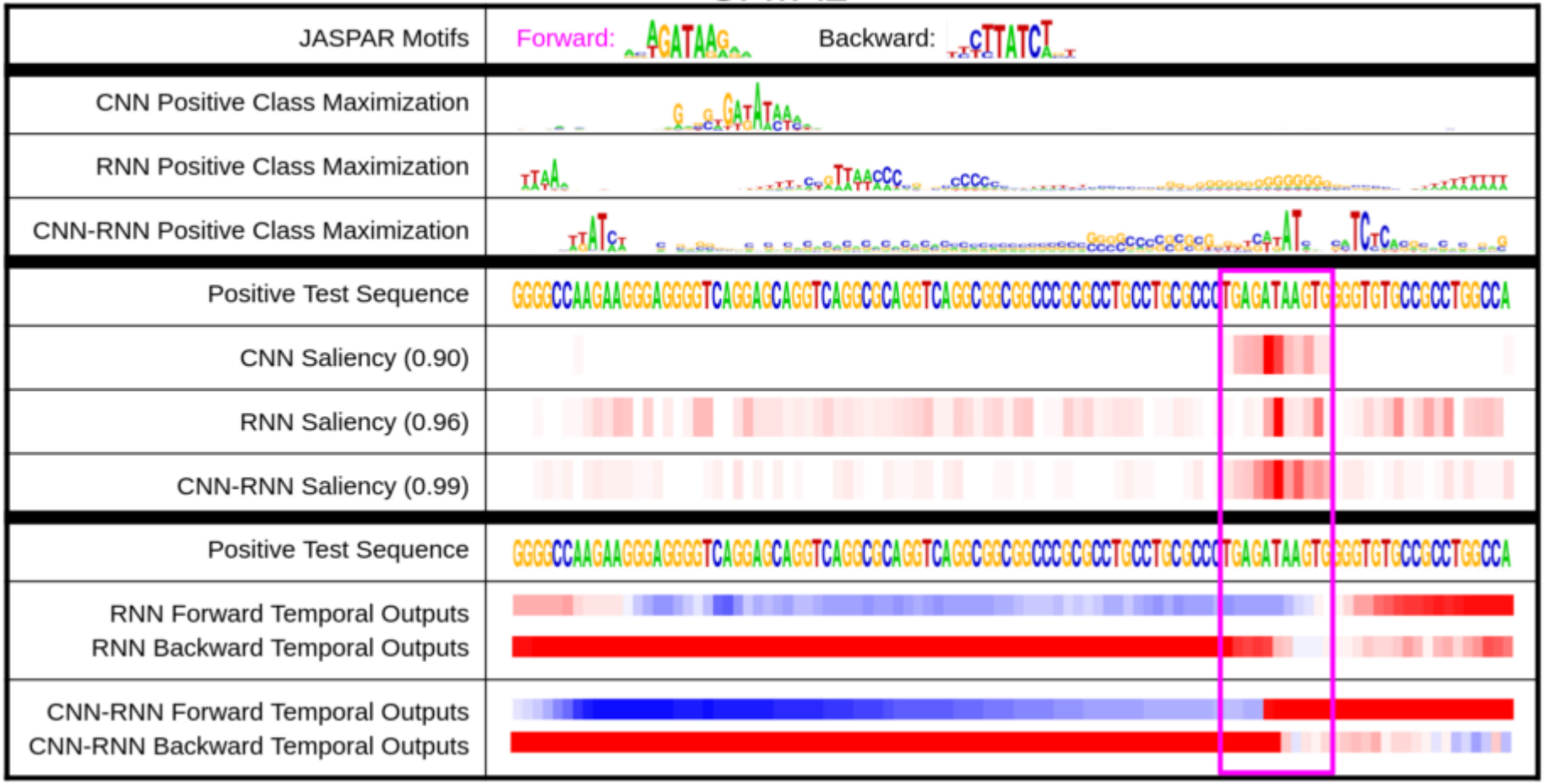
Table 2: Mean AUC scores on the TFBS classification task

Model	Mean AUC	Median AUC	STDEV
MEME-ChIP [16]	0.834	0.868	0.127
DeepBind [2] (CNN)	0.903	0.931	0.091
Small RNN	0.860	0.881	106
Med RNN	0.876	0.905	0.116
Large RNN	0.808	0.860	0.175
Small CNN	0.896	0.918	0.098
Med CNN	0.902	0.922	0.085
Large CNN	0.880	0.890	0.093
Small CNN-RNN	0.917	0.943	0.079
Med CNN-RNN	**0.925**	**0.947**	**0.073**
Large CNN-RNN	0.918	0.944	0.081

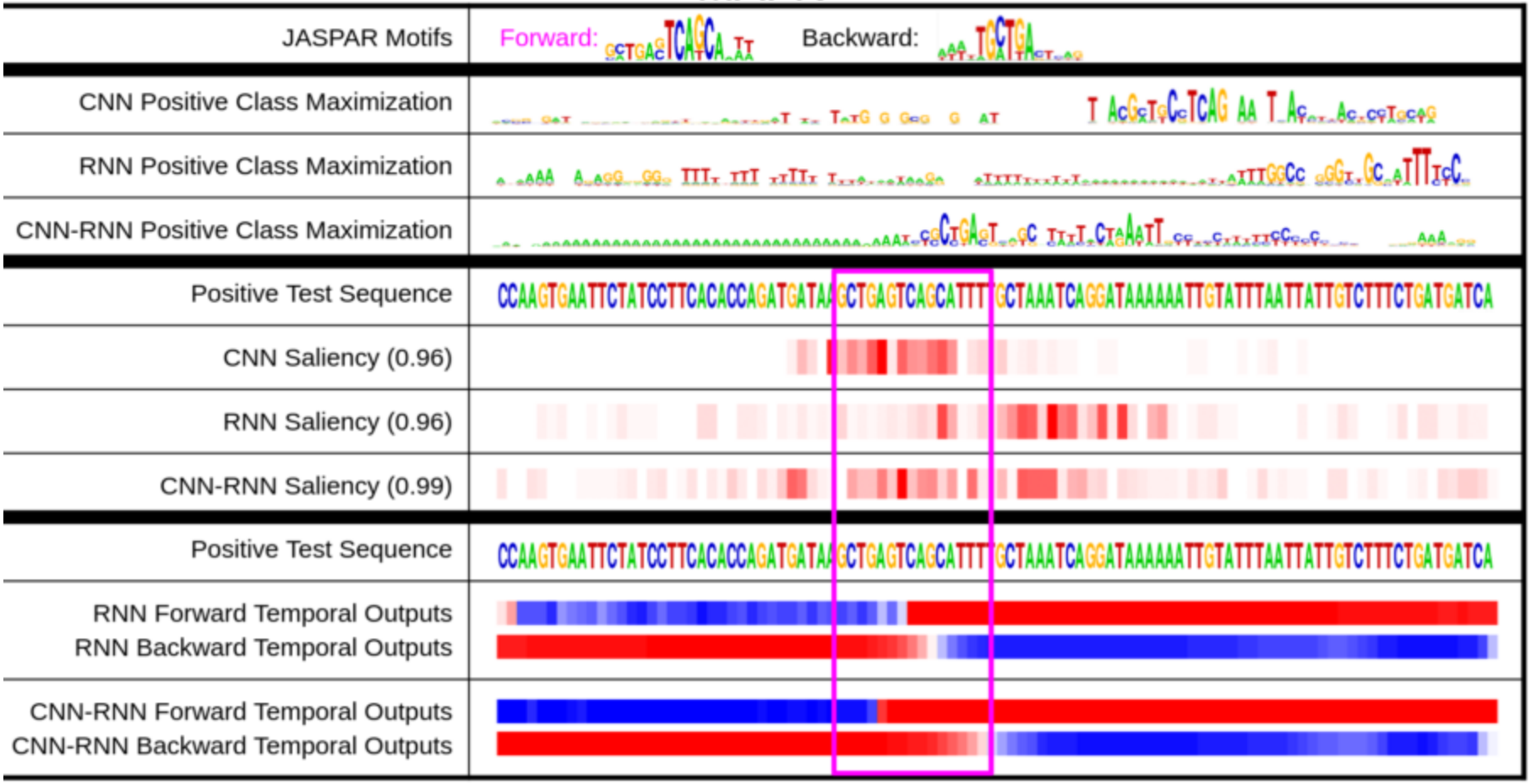
Table 3: AUC pairwise t-test

Model Comparison[3]	p-value
RNN vs MEME	5.15E-05
CNN vs MEME	1.87E-19
CNN-RNN vs MEME	4.84E-24
CNN vs RNN	5.08E-04
CNN-RNN vs RNN	7.99E-10
CNN-RNN vs CNN	4.79E-22

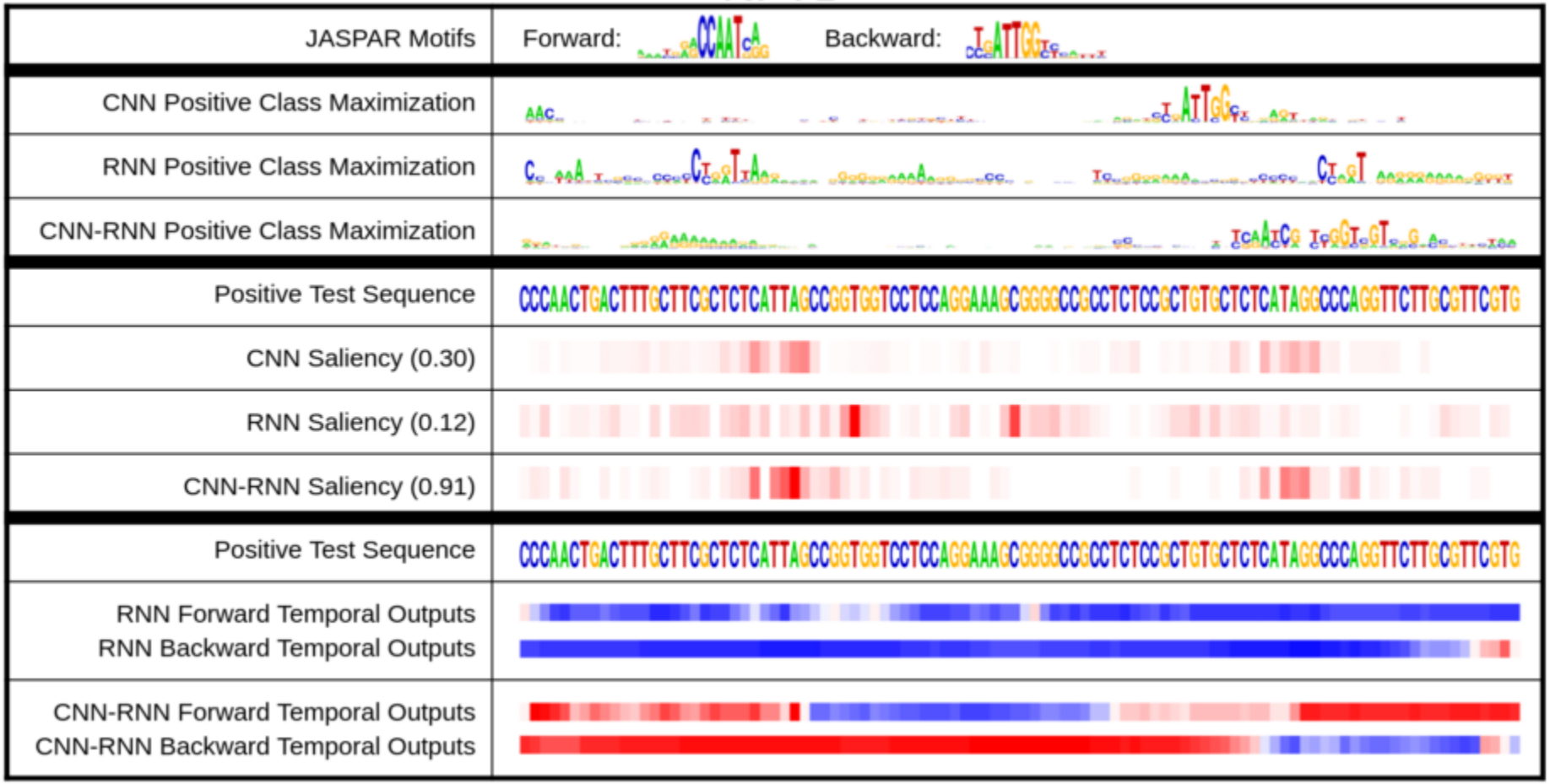
GATA1


JASPAR Motifs
Forward:
Backward:
CNN Positive Class Maximization
RNN Positive Class Maximization
CNN-RNN Positive Class Maximization
Positive Test Sequence
GGGGCCAAGAAGGGAGGGGTCAGGAGCAGGTCAGGCGCAGGTCAGGCGGCGGCCCGCGCCTGCCTGCGCCCTGAGATAAGTGGGGTGTGCCGCCTGGCCA
CNN Saliency (0.90)
RNN Saliency (0.96)
CNN-RNN Saliency (0.99)
Positive Test Sequence
GGGGCCAAGAAGGGAGGGGTCAGGAGCAGGTCAGGCGCAGGTCAGGCGGCGGCCCGCGCCTGCCTGCGCCCTGAGATAAGTGGGGTGTGCCGCCTGGCCA
RNN Forward Temporal Outputs
RNN Backward Temporal Outputs
CNN-RNN Forward Temporal Outputs
CNN-RNN Backward Temporal Outputs

MAFK


JASPAR Motifs
Forward:
Backward:
CNN Positive Class Maximization
RNN Positive Class Maximization
CNN-RNN Positive Class Maximization
Positive Test Sequence
CCAAGTGAATTCTATCCTTCACACCAGATGATAAGCTGAGTCAGCATTTTGCTAAATCAGGATAAAAAAATTGTATTTAATTATTGTCTTTCTGATGATCA
CNN Saliency (0.96)
RNN Saliency (0.96)
CNN-RNN Saliency (0.99)
Positive Test Sequence
CCAAGTGAATTCTATCCTTCACACCAGATGATAAGCTGAGTCAGCATTTTGCTAAATCAGGATAAAAAAATTGTATTTAATTATTGTCTTTCTGATGATCA
RNN Forward Temporal Outputs
RNN Backward Temporal Outputs
CNN-RNN Forward Temporal Outputs
CNN-RNN Backward Temporal Outputs

NFYB


JASPAR Motifs
Forward:
Backward:
CNN Positive Class Maximization
RNN Positive Class Maximization
CNN-RNN Positive Class Maximization
Positive Test Sequence
CCCAACTGACTTTGCTTCGCTCTCATTAGCCGGTGGTCCTCCAGGAAAGCGGGGCCGCCTCTCCGCTGTGCTCTCATAGGCCCAGGTTCTTGCGTTCGTG
CNN Saliency (0.30)
RNN Saliency (0.12)
CNN-RNN Saliency (0.91)
Positive Test Sequence
CCCAACTGACTTTGCTTCGCTCTCATTAGCCGGTGGTCCTCCAGGAAAGCGGGGCCGCCTCTCCGCTGTGCTCTCATAGGCCCAGGTTCTTGCGTTCGTG
RNN Forward Temporal Outputs
RNN Backward Temporal Outputs
CNN-RNN Forward Temporal Outputs
CNN-RNN Backward Temporal Outputs

References and Other Good Reads

1. Zhengping Che, Sanjay Purushotham, Robinder Khemani, and Yan Liu. 2016. Interpretable Deep Models for ICU Outcome Prediction. *AMIA ... Annual Symposium proceedings. AMIA Symposium* 2016: 371–380.
2. Jack Lanchantin, Ritambhara Singh, Beilun Wang, and Yanjun Qi. 2016. Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks. In *Pacific Symposium on Biocomputing*, 1–11. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv:150302531. 2015;.
3. Buciluǎ C, Caruana R, Niculescu-Mizil A. Model compression. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2006. p. 535--541.
4. Korattikara A, Rathod V, Murphy K, Welling M. Bayesian Dark Knowledge. arXiv preprint arXiv:150604416. 2015;.
5. Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology* 33, 8: 831–838.