

Classification of Noncoding RNA Families using Deep Convolutional Neural Networks

©2020

Brian McClannahan

B.S. Computer Science, University of Kansas, 2017

Submitted to the graduate degree program in Department of Electrical Engineering and Computer Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

Committee members

Dr. Cuncong Zhong, Chairperson

Dr. Prasad Kulkarni, Committee Member

Dr. Bo Luo, Committee Member

Dr. Guangui Wang, Committee Member

Date defended: December 23, 2020

The Thesis Committee for Brian McClannahan certifies
that this is the approved version of the following thesis :

Classification of Noncoding RNA Families using Deep Convolutional Neural Networks

Dr. Cuncong Zhong, Chairperson

Date approved: December 23, 2020

Abstract

In the last decade, the discovery of noncoding RNA (ncRNA) has exploded. Classifying these ncRNA is critical to determining their function. This thesis proposes a new method employing deep convolutional neural networks (CNNs) to classify ncRNA sequences. To this end, this thesis first proposes an efficient approach to convert the RNA sequences into images characterizing their base-pairing probability. As a result, classifying RNA sequences is converted to an image classification problem that can be efficiently solved by available CNN-based classification models. This thesis also considers the folding potential of the ncRNAs in addition to their primary sequence. Based on the proposed approach, a benchmark image classification dataset is generated from the RFAM database of ncRNA sequences. In addition, three classical CNN models and three Siamese network models have been implemented and compared to demonstrate the superior performance and efficiency of the proposed approach. Extensive experimental results show the great potential of using deep learning approaches for RNA classification.

Acknowledgements

I would like to thank Dr. Richard Wang for all his help guiding me through my Master's program. Dr. Cuncong Zhong for his help with my first publication and joining as my committee chairman. Dr. Bo Luo and Dr. Prasad Kulkarni for assisting me as committee members. I would also like to thank Usman Sajid for assisting me on this thesis both with BPPM image generation and with familiarizing me with the ITTC laboratory. I would like to thank Krushi Patel for her assistance with initial model design on this project. I would like to express my gratitude to Wenchi Ma for his assistance in developing Siamese networks. I have very much enjoyed the process of working on this thesis, which wouldn't have been possible without Dr. Wang. He was always willing to help me with my research, papers, and design whenever I had questions. Lastly, I would like to thank all the students in the ITTC Computer Vision group for being a great research group that was always willing to answer any questions I had.

Contents

1	Introduction	1
1.1	Background	1
1.2	Motivation	2
1.3	Goals	3
1.4	Contributions	4
2	Related Works	6
2.1	Analyzing RNA by Secondary Structure	6
2.2	Image Classification with Convolutional Neural Networks	7
2.3	Siamese Networks	9
3	Methodology	12
3.1	Dataset	12
3.1.1	CNN Dataset	12
3.1.2	Siamese Dataset	17
3.2	Classification Models	18
3.2.1	Convolutional Neural Networks	18
3.2.2	Siamese Networks	21
4	Experiment and Results	24
4.1	Experiment	24
4.1.1	Convolutional Neural Networks	24
4.1.2	Siamese Network	25
4.2	Evaluation Metrics	27

4.3	Results	29
5	Discussion and Future Work	31
5.1	Discussion	31
5.2	Future Work	32
6	Conclusion	34

List of Figures

3.1	Top: an example of interior loop of an RNA sequence. Bottom: Corresponding dot-plot matrix of the sequence.	12
3.2	Four dot-plot examples of RNA secondary structure RFAM families a) 5S rRNA, b) C0719, c) snR63, d) S_pombe_snR42	13
3.3	Combine two dot-plot matrices into one image for classification.	14
3.4	Sample images of two RNA from the same family.	15
3.5	Sample images of two RNA from different families.	16
3.6	Convolution Example	19
3.7	Residual Block used in ResNet He et al. (2016)	19
3.8	VGGnet-19 and ResNet example He et al. (2016)	20
3.9	Siamese Network Architecture	23
4.1	Receiver Operating Curves for CNN models.	29
4.2	Receiver Operating Curves for Siamese models.	29

List of Tables

3.1	Final image count of the dataset before the iterative random image selection algorithm was applied.	17
4.1	Comparative results of different convolutional neural networks trained on different dataset ratios.	25
4.2	Comparative results of different Siamese networks trained on different dataset ratios.	26

Chapter 1

Introduction

1.1 Background

Ribonucleic Acid (RNA) plays vastly many different roles within living cells. In 2007, with the development of high-throughput sequencing, a large amount of new RNA were discovered Heyne et al. (2012). Many of these RNA were found not to be involved in protein creation and are thus called noncoding RNA (ncRNA). However, just because these RNA do not directly code for protein does not reduce their impact in the cell Zhang et al. (2017). Noncoding RNA have been shown to play many important roles in cell processes such as to effect high-order chromosomal dynamics, telomere biology, subcellular structural organization, and gene expression. These processes are central to life Hombach & Kretz (2016). Not just in animals, but in plants, fungi, and prokaryotes.

Some significant ncRNA include microRNA (miRNA), which performs posttranscriptional regulation of gene expression. Micro-RNA do this by partial complimentary base pairing to specific messenger RNA (mRNA). This process inhibits protein translation and can also facilitate degradation of the target mRNA Mercer et al. (2009). Small nuclear RNA (snRNA) also play a pivotal role in gene expression. Small nuclear RNA are involved in the pretranscriptional splicing process. Transfer RNA, another ncRNA, decode the mRNA sequence into peptides or proteins. Ribosomal RNA (rRNA) form the framework of ribosomes, a macromolecular structure essential for protein translation.

Small non-coding RNA have been proven to effect many different processes in animal development. Small ncRNA function as developmental switches by silencing unwanted messages resulting from leaky transcription or previous synthesis. This acts as a maintenance on target mRNA expres-

sion, keeping them in optimal ranges Stefani & Slack (2008). Micro RNA play a role in early embryonic development by modulating nodal levels in early embryos. Nodal is a transforming growth factor protein that plays a very important role in early embryonic development. Micro RNA also play key roles in muscle development by regulating proliferation to differentiation switching in muscles and playing an important role in cardiac function Stefani & Slack (2008). It is also important in lymphocyte development as with miR-181, an RNA which increases the sensitivity of DP cells to stimulation of the T-Cell receptor, a type of white blood cell Stefani & Slack (2008).

Long non-coding RNA (lncRNA) are a subset of ncRNA which are longer than 200 nucleotides. These RNA show very low sequence conservation Ma et al. (2013). These lncRNA are the most functionally diverse class of RNA. They are predominantly localized in the nucleus of the cell and specifically within chromatin, the material that forms chromosomes within the nucleus Ma et al. (2013). The cascading transcription of lncRNA across the fructose bisphosphate promoter in yeast is associated with the progressive opening of chromatin, increasing access to transcriptional activators Ma et al. (2013). Long ncRNA can regulate epigenetic modification and gene expression Zhang et al. (2017). Other lncRNA have also been shown to effect post-transcriptional regulation through splicing regulation and transitional control Ma et al. (2013). These families of RNA are medically significant because their ability to regulate associated protein-coding genes. This regulation could contribute to disease if misexpressed, thereby deregulating genes of clinical significance Mercer et al. (2009). The classification of ncRNAs seeds to categorize ncRNA elements into families based on their sequence and structure to facilitate their functional annotation and prediction Zhang et al. (2017).

1.2 Motivation

RNA are defined by three graphical representations: primary structure, secondary structure, and tertiary structure. The primary structure is defined by their sequence. The secondary structure is defined by the bonds formed from folding. The tertiary structure is defined by the shape the RNA takes in the physical world. Given an RNA's sequence, it is much simpler to determine its

secondary structure than tertiary structure Herschlag (1995). Given this, bioinformatics focuses on the primary and secondary structure. The molecular function of ncRNA is implemented through both of its sequence and structure Heyne et al. (2012). Classification of ncRNAs purely using their sequences is insufficient, as ncRNAs with conserved secondary structures may share low sequence identity due to the presence of co-variant mutations. Many ncRNA cannot be aligned by purely Sequence-based alignment techniques struggle when the pairwise sequence identities are less than 60%. Initial methods looked at using structure-based alignment techniques like MARNAsiebert & Backofen (2005) and RNAforester Hochsmann et al. (2003), but these struggled because of poor accuracy of single structure predictions Will et al. (2012). Therefore, modern methods for ncRNA clustering use both the primary sequence and secondary structure features Smith et al. (2017).

Unfortunately, the consideration of the secondary structure increases the time complexity for pairwise ncRNA comparison, from $O(l^2)$ with pure sequence to $O(l^4)$ with both sequence and secondary structure. The high time complexity thus makes clustering of large amount of ncRNA elements infeasible with current methods. New methods are needed to handle the continuing growth of discovered ncRNA. The initial release of RFAM 1.0 in 2003 contained 25 unique RNA families. In 2018, RFAM 13.0 was released, containing 2,687 families. Faster algorithms are needed to process the exponentially growing size of bioinformatics data, especially within ncRNA.

1.3 Goals

Given a newly-discovered RNA, modern day approaches rely on biologists to manually align sequences with other RNA across similar-length RNA families in order to determine their RNA family. This is a very time-consuming process. The obvious solution is to use sequence alignment approaches to determine RNA families. These sequence alignment techniques use dynamic programming which have high complexity, especially on longer RNA. It also requires the RNA in the same family to have similar sequence identities, of which many lncRNA do not. Calculating the RNA secondary structure while aligning the sequences improves classification accuracy of lncRNA but increases the complexity of the classification from $O(l^2)$ to $O(l^4)$. The other option

is to use structural alignment, but these methods have previously shown poor accuracy. Thus, it is necessary to find quicker algorithms that can accurately classify RNA with just its secondary structure, and not only determine if an RNA is within a currently known family, but can determine if the RNA should be defined within a completely new RNA family.

1.4 Contributions

In this paper, I propose a new approach to general RNA family classification using image processing techniques. For a given ncRNA element, the base-pairing probability matrix (BPPM) was computed using RNAfold from the Vienna Package. I then convert the BPPM into a gray-scale image, using the intensity of each pixel to represent the base-pairing probability of the corresponding bases. I then apply three different deep CNN algorithms (VGGnet-19, ResNet-50, and ResNet-101) to classify these images. These deep CNN algorithms are then wrapped into Siamese networks for classification as well. I tested this approach using the RFAM database Kalvari et al. (2017), and showed an 85% classification accuracy with the CNN model and 79% classification accuracy with the Siamese model.

The main contributions of this thesis are:

1. I propose to convert the problem of RNA sequence classification into an image classification problem.
2. I propose an efficient approach to convert two RNA sequences into an image and generate an image dataset for RNA sequences from the same and different families.
3. I implement three classical deep learning-based classification models and compare their performance in RNA classification. The results demonstrate the feasibility of the advantages of the proposed approach.
4. I implement these three classical deep learning-based classification models as Siamese Networks to evaluate their performance with the deep CNNs as feature-extractors on ncRNA.

A part of this thesis - McClannahan et al. (2020) has been published in the 2020 IEEE International Conference on Systems, Man, and Cybernetics.

Chapter 2

Related Works

2.1 Analyzing RNA by Secondary Structure

Initial methods, like MARNA and RNAforester, utilized just the predicted or actual secondary structure for classifying ncRNA Will et al. (2012). MARNA used pairwise-sequence alignments to make their RNA structural predictions. These structural predictions were then used in multi-structural alignments Siebert & Backofen (2005). RNAforester built an RNA secondary structure alignment algorithm by representing RNA secondary structures as trees and aligning them with tree alignment algorithms. However, these methods were found to struggle with accurate predictions. There are several other, more modern methods that are used for classifying RNA by its secondary structure. QRNA utilizes a context-free grammar constrained by structural RNA evolution paired with two different hidden Markov models for determining other constraints Rivas & Eddy (2001). RNAfold Hofacker (2003) and Mfold Zuker (2003) use secondary structure comparisons based on a minimum free energy model. RNAz predicts the conserved structure of RNA based on its thermodynamic stability Washietl (2012). Sankoff-style folding algorithms have also shown effectiveness in doing simultaneous primary and secondary structure alignments Sankoff (1985). CARNA showed how these could be used to classify RNA by finding a predicted secondary structure during RNA alignment Sorescu et al. (2012). Several other approaches compare multiple graph representations of RNA secondary structure such as GraphClust Heyne et al. (2012) and Karklin et al. (2004).

The most successful methods have combined sequence alignment techniques with a partition function, like the McCaskill Partition function Mccaskill (1990), to compare the primary and sec-

ondary structures at the same time, such as LocaRNA-P Will et al. (2012). LocaRNA uses the McCaskill partition function initially to generate a base-pair probability matrices for two given RNA. These matrices are then scored based on confidence in the individual alignment columns and the predicted consensus structure, called STARS (sequence-structure-based alignment reliability). These STARS are calculated by dividing its Boltzmann weights by the total partition function. STARS are then used to do multiple alignment across RNA to generate STAR profiles. These profiles predict the boundaries of local regions of conserved sequence and structure to find ncRNA in alignments of longer sequences. These star profiles allow for RNA clusters to be found when pointed at a genome Will et al. (2012).

LocaRNA was one of the first classifiers not to struggle to accurately analyze ncRNA. Another more recent approach built with LocaRNA's approach in mind is DotAligner, which also uses both the primary sequence and secondary structure to discover sequence motifs, but is more focused on lncRNAs Smith et al. (2017). DotAligner first generates a dynamic programming matrix filled by aligning the two given sequences based on their similarity and cumulative base-pair probabilities. Then the structural compatibility of the two RNA sequences is calculated using a partition function over all pairwise alignments. The two most likely structural ensembles generated by the partition function are considered. Lastly, the sequences alignments are warped towards the alignment that includes structural features, thereby exposing the common structural features hidden by suboptimal base-pairing ensembles.

2.2 Image Classification with Convolutional Neural Networks

Recently, researchers have applied CNNs to many different fields of applications, including image classification Cen & Wang (2020), translation Xu et al. (2019), object detection Ma et al. (2020), depth estimation He et al. (2018b), crowd counting Sajid et al. (2020), and medical image analysis Patel et al. (2020). AlexNet was the first CNN to be applied towards image classification tasks. AlexNet used a series of overlapping convolutional masks of size 11x11, 5x5, and 3x3 pixels over the image. Convolutional layers were followed by pooling layers and ReLU activation lay-

ers. ReLU is an important activation function in Image classification because it adds non-linearity and also removes negative values. Pixel values can never be lower than 0 so ReLU purges any weights that generate negative values while maintaining positive values coming out of the convolution masks. AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 competition with a top-5 error rate of 15.3%, significantly outperforming other top performing methods which only achieved 26.2% top-5 error rate Krizhevsky et al. (2017). GoogleNet followed up on AlexNet's innovation by introducing the inception module. The inception module employs 3×3 and 5×5 convolution masks followed by 1×1 convolution masks for dimensionality reduction and reduce the total number of parameters required to be trained by the network. GoogleNet achieved a top-5 error rate of 6.67% at the ILSVRC 2014 classification task Szegedy et al. (2015).

VGGnet released the same year as GoogleNet. VGGnet demonstrated a method of building CNNs with significantly fewer parameters to be trained. By stacking 3×3 convolutional filters, much larger convolutional filters can be implemented more efficiently. For example, a 7×7 convolutional filter can be implemented by stacking three 3×3 convolutional filters on top of each other. For C channels in the convolutional layer, the 7×7 convolutional filter reduces from $7^2 C^2 = 49C^2$ parameters to $3(3^2 C^2) = 27C^2$ parameters, a 45% reduction in parameters with the same output. VGGnet achieved the first place in localization and the second in classification at ImageNet Challenge 2014 with a top-5 error rate of 7.3% Simonyan & Zisserman (2014). ResNet introduced residual mappings to CNNs. A residual mapping within a neural network layer approximating the function $H(x)$ changes the output of the layer from $F(x) = H(x)$ to $F(x) = H(x) + x$. When a layer with a residual mapping is trained, if it fails to create an accurate approximation of $F(x)$ the output of the layer is still just x instead of a warped $H(x)$ that corrupts the rest of the network. This change allows significantly deeper networks to be trained without overfitting. The largest VGGnet model is 19 layers and has 19.6 billion floating point operations (FLOPs). ResNet was able to build significantly deeper models including a network with 152 layers. Since the network was able to be much deeper due to residual mappings utilized, the network was able to be much thinner

and the 152 layer network only had 11.3 billion FLOPs, 40% FLOPs compared to the 19-layer VGGnet model. Using these improvements, ResNet achieved first in the ILSVRC 2015 classification task with a top-5 error rate of 6.71% He et al. (2016). After ResNet created significantly deeper networks, wide residual networks showed that similar results could be achieved by greatly increasing the number of parameters in a shallow network and utilizing dropout layers Zagoruyko & Komodakis (2016). Other approaches include DenseNet, a CNN designed to compute dense, multi-scale feature pyramids using up and down sampling of image patches with a CNN-based object classifier Iandola et al. (2014).

Very few studies have been performed for RNA classification using CNN based classifiers. CNNclust Aoki & Sakakibara (2018) utilized one-dimensional CNNs for ncRNA sequence motif discovery. Similar to DotAligner, CNNclust used both the sequence and the secondary structure to more accurately find sequence motifs and then use those to make accurate classifications. However, CNNclust utilized one-dimensional convolutions over the primary sequence and base-pair probability as opposed to using partition functions like LocaRNA and DotAligner. This one-dimensional CNN took in a one-hot encoded sequence value of the alignment not the sequence, allowing for gap values to be included in the input. Along with the sequence, a three-dimensional vector was also input to the network defining the secondary structure at that nucleotide. This vector was a one-hot encoded vector representing whether the current nucleotide would bond upstream, downstream, or not at all. The network then performed sets of convolution operations over the sequence and structure representations followed by max pooling layers and then three fully-connected layers to classify two RNA. All classifications are then combined into an adjacency matrix and family clusters are then extracted from the adjacency matrix such that each family is defined by the complete maximal sub-graphs extracted.

2.3 Siamese Networks

Siamese neural networks (SiNN) are more complicated neural networks that try to take advantage of other network architectures to make predictions about similarity. A Siamese network has two

branches with each branch sharing the exact same architecture and weights. The network receives two inputs and feeds each input through one of the branches, but the exact order of the input is irrelevant to the final output of the SiNN. Each output from the branch networks are then passed through a comparison layer that evaluates the output of each for similarity. This similarity is then passed to a top network which makes the final classification. Siamese Neural Networks make use of this twin branch network structure to realize a non-linear embedding from its input domain to some euclidean domain Roy et al. (2019). Effectively the branch network operates as a descriptor computation module and the top network as a input similarity function Zagoruyko & Komodakis (2015). These properties of feature extraction and similarity comparisons make Siamese Networks great at classifying similarity between inputs.

In Melekhov et al. (2016), Siamese networks have demonstrated measured improvements over CNNs in image matching. Siamese networks built on top of CNNs have demonstrated the ability to match landmarks more accurately than using a standard CNN, even when the landmarks have not been seen before and the training set contained mislabeled data Melekhov et al. (2016). Siamese Networks have also been shown to be superior to CNNs in general image matching Zagoruyko & Komodakis (2015). Building off of their improved image matching capabilities, Researchers have also applied Siamese networks to object tracking. It has been demonstrated that Siamese networks use available data more efficiently and perform more efficient spatial searches compared to more generic CNN methods Bertinetto et al. (2016). Other approaches have also utilized Siamese networks in person tracking He et al. (2018a). They have increased performance in identifying people across multiple images by identifying features of tracked people and matching them to people in subsequent photos. In fact, Siamese Networks using CNNs were able to show state-of-the-art performance in generic image matching with only a single initial observation Tao et al. (2016). Siamese networks have shown capabilities outside of image comparisons. In Koch et al. (2015), Siamese networks were applied to one-shot image classification and demonstrated the capability to do one-shot classification for verification. These capabilities extend to other domains too, specifically image classification Koch et al. (2015). Within bioinformatics, Siamese networks have seen

use with chromosomal classification to improve performance compared to deep learning models
Jindal et al. (2017).

Chapter 3

Methodology

3.1 Dataset

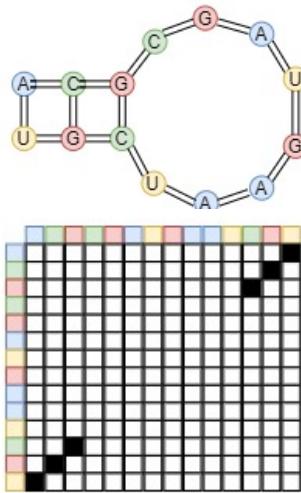


Figure 3.1: Top: an example of interior loop of an RNA sequence. Bottom: Corresponding dot-plot matrix of the sequence.

3.1.1 CNN Dataset

The dot-plot matrix is generated by letting any cell (i, j) represent the probability that a bond exists between the i th and j th nucleic acid in the RNA secondary structure (excluding bonds amongst neighboring nucleic acids). Generating a dot-plot matrix from the secondary structure defined by the BPPM gives a matrix that is symmetric along the diagonal, as shown in Fig. 3.2. The dot-plot matrix is then converted into an image by treating each cell as a pixel where the probability of a bond forming is treated as an intensity between 0 and 1, as illustrated in Fig. 3.1. This process

creates a grey-scale, symmetric image representation of the secondary structure of any given RNA.

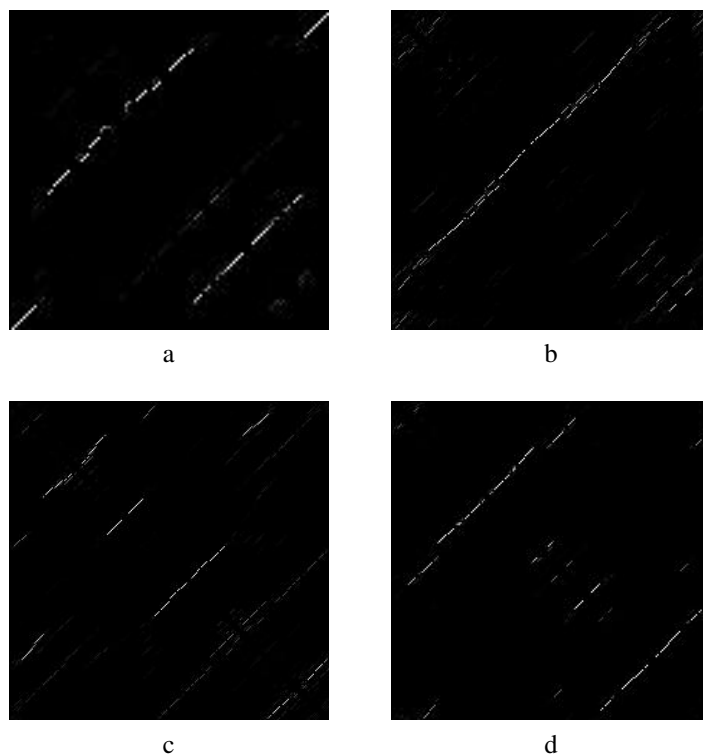


Figure 3.2: Four dot-plot examples of RNA secondary structure RFAM families a) 5S rRNA, b) C0719, c) snR63, d) *S_pombe_snR42*

There are two options for designing the dataset for RNA classification. The first option is to evaluate the dot-plot matrix and classify each RNA with its familial class. Most CNNs are trained and tested on the ImageNet dataset to evaluate their capability in image classification. ImageNet uses 1000 different classes that the CNNs classify on. With this design, the convolution layers detect features and then the fully-connected layers evaluate these features to determine which of the predetermined classes the image is a part of. This would limit the network to only classifying RNA families it has previously been trained on and thus prevent previously unknown RNA families from being defined. The other option is to let each classification made by the network determine if two RNA sequences are from the “same family” or “different family”. To solve this problem without altering the RNA dot-plot matrices in some way, it would be necessary to build a network large enough to accept two full images for the two RNA sequences. This would require

significant augmentation of current CNN frameworks and would likely fail to take advantage of the powerful feature extraction of CNNs. Using a standard CNN in this capacity would also leave the network irregular (not-square), which, along with being hard to implement and train the network, would take a very long time to train for RNA sequences, and would result in features not being weighed equally due to the CNN no longer being symmetrical. Instead, in this study, I propose a different method to convert the RNA classification problem to an image classification problem. As illustrated in Fig. 3.2, I make use of the symmetric property of the dot-plot of each sequence and generate one single image from two RNA sequences. The bottom-left half from sequence 1 and the top-right half from sequence 2 are stitched together. In this way, the problem is converted from a two-input problem to a single input, ideal for CNNs. Using this method of stitching two dot-plot matrices together, I can make use of all available image classification models from RNA classification. As seen in Fig.3.3, the images from the same family have better symmetric property than those from different families, however, the differences are very small. The models are trained to distinguish these small changes. Some generated examples are shown in Figures 3.4 and 3.5.

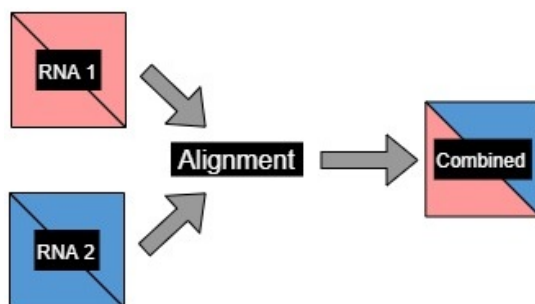


Figure 3.3: Combine two dot-plot matrices into one image for classification.

To train and evaluate each deep learning model, a large dataset is required. The RFAM database has over 3,000 families. For these experiments only RNA families that have a sequence length between 200 and 260 were included for the convenience of study. RNA families typically have similar sequence length so using a small difference in sequence length is better. It allows the results to accurately reflect how this method would be used in a functional scenario. The RNA dot-plots for all sequences were resized to 224×224 to combine each pair of them into an image. Using

a sequence length between 200 and 260 also adds the benefit that the images do not get warped during resizing, allowing better evaluation of the effectiveness of convolutions in RNA classification. There were 168 families in this range. The 168 families were split into the train/val/test set at a ratio of 70:10:20, resulting in a family split of 121/19/28, as shown in Table 3.1. If the dataset was created exhaustively to cover all possible RNA combinations, the resulting dataset had over 17 million images. A dataset of that size was not necessary for model training as there are a lot of redundant information in the dataset. The largest family size in this set was 712 while many of the families only had 2. Instead of using every possible combination, each family larger than 30 RNA was truncated by randomly selecting 30 RNA from the family. This also has the added benefit of balancing the varying family sizes, preventing the network from favoring classifying RNA with large family sizes as “Same Family” more frequently than small RNA families, focusing the model on learning features of RNA. This reduced the total number of images to slightly greater than 2.4 million.

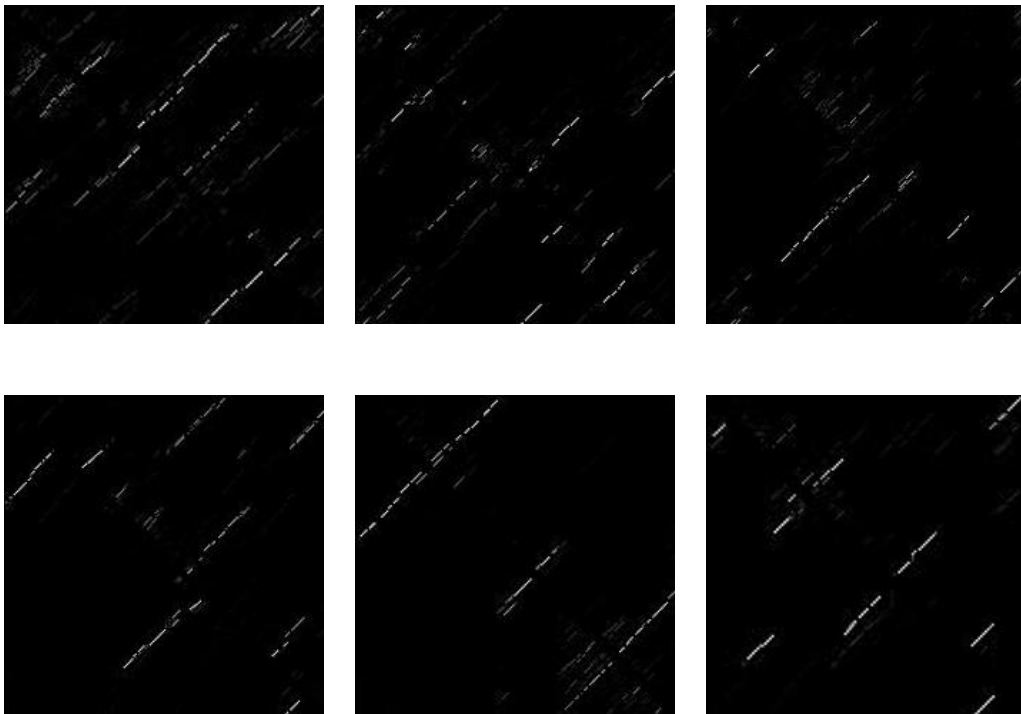


Figure 3.4: Sample images of two RNA from the same family.

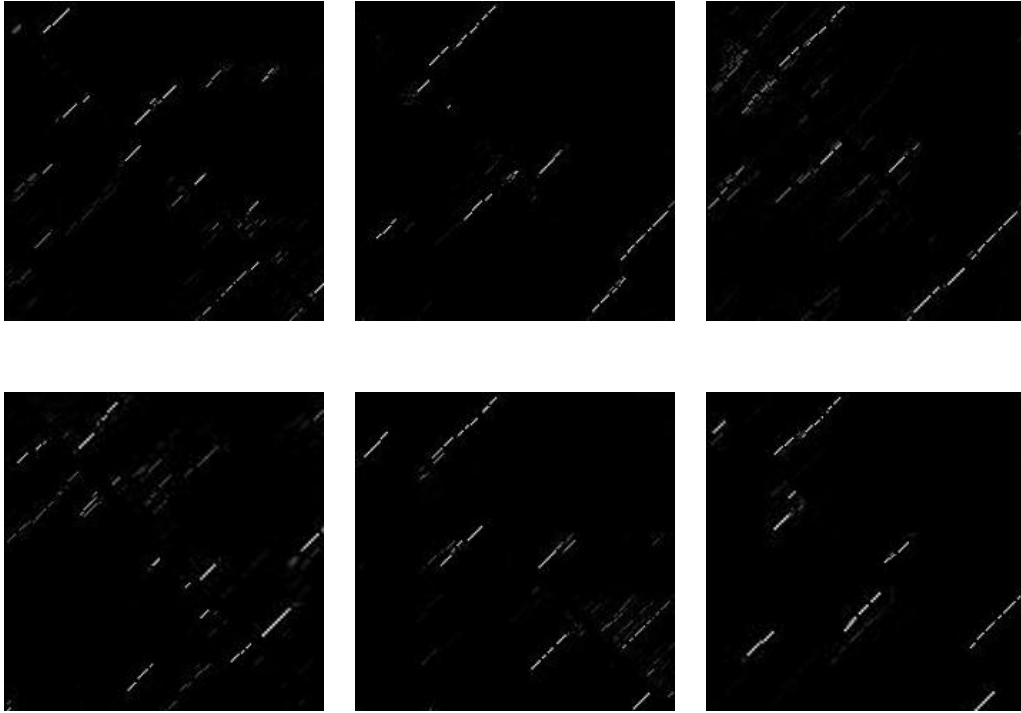


Figure 3.5: Sample images of two RNA from different families.

In table I, images using all possible combinations of the RNA sequences were generated, except for the training data of different families. With 121 families in the training set, over 2.4 million training images can be generated for sequences from different families if use all possible combinations of RNA are used. However, in these experiments, I found that the models converged quickly without using all the training data. Maintaining an unnecessarily large training set is not only overly complicated but increases training time increasing loading time of images. It also decreases the representativeness of the training data. Instead, to reduce the size of the training data, a different approach is taken, The following approach was used to decide the training images for the different family class of the training set. First, I randomly select an RNA from each family. Then, I randomly pick an RNA from every other family and combine them together to generate a training image of different families. I repeat this process 20 times and generate a total of 290,400 different family images of the final training set.

Set	# of Families	Class	Image Count
Train	121	Same	30,522
		Different	290,400
Val	19	Same	5,374
		Different	68,882
Test	28	Same	9,520
		Different	178,402

Table 3.1: Final image count of the dataset before the iterative random image selection algorithm was applied.

3.1.2 Siamese Dataset

As discussed in section 3.1.1, there are two options for designing the dataset to be evaluated by CNNs. Either classify the RNA directly to a family, or classify two RNA as “Different Family” or “Same Family.” Siamese networks are explicitly designed to make comparisons given two different inputs though. However, given that a Siamese network needs two inputs, a different dataset than the one used for the stand-alone CNNs is needed. In fact, building the dataset for the Siamese Network was simpler. The original dot-plot images used to form the stand-alone CNN dataset before being stitched together were used. Each input to the Siamese network consisted of two dot-plot matrices. During training, the model simply selected two RNA at run-time. Thus, it was unnecessary to apply limits on the different-class of the training dataset as with the stand-alone CNN dataset. The exact set of images seen each iteration during training were different. However, both the standard CNN and Siamese networks were validated and tested on the same set of RNA. The sizes of the validation and test sets for the Siamese networks are exactly half the size of the dataset for the CNN. This is due to the reversibility allowed by the stand-alone CNNs that is unnecessary with the Siamese networks. Each image given to the CNNs represents two RNA, but there is a second combination of both RNA that also represents the two RNA. In order to effectively evaluate how well the CNN is at classifying RNA, both cases must be tested. With the Siamese network, flipping the inputs has no effect on the output of the network. Looking at the Siamese network as a function $S(x,y)$, given two RNA, A and B , $S(A,B)$ will always give the same output as $S(B,A)$,

3.2 Classification Models

3.2.1 Convolutional Neural Networks

In math, a convolution is a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other. In image processing, convolutions can be expressed as discrete filters over the image. Using these filters, create many feature maps that relate nearby pixels. This relation allows for localized feature gathering to be expressed in a global classification when used over an entire image. Convolutional neural networks are built utilizing these convolutions. CNNs contain two parts, a feature extractor and a fully-connected network. The feature extractor is composed of stacked convolutional layers mixed with activation and pooling layers. The stacked convolutional layers, serving as feature extractors, are each followed by an activation layer, typically ReLU activation, to allow for the extraction of nonlinear features Li et al. (2020). The pooling layers, which typically come at the end of the convolutional layers, or at the end of a block of convolutional layers depending on the model design, reduce the spatial resolution of the feature maps, making the model more robust to input distortions Rawat & Wang (2017). The fully-connected network is a densely connected network that interprets the feature representations extracted by the feature network.

In the experiment, two classical CNN-based classification models were implemented: VGGnet Simonyan & Zisserman (2014) and ResNet He et al. (2016). VGGnet was shown how larger convolution filters could be implemented more efficiently using stacked 3×3 convolutions. VGGnet follows similar previous CNN constructions. Each block of convolutional layers is followed by max pooling layers to select the most important features to pass to the next block of convolutional features. Each convolutional block implements a higher-order convolution used by many previous models but is implemented as stacked 3×3 layers for efficiency. VGGnet uses a fully-connected network to evaluate the features. This network has two layers of 4096 neurons each and then a final 1000-neuron with softmax activation to make the final classification. It achieved first in the ImageNet Challenge 2014 localization task and second in the classification task with VGGnet-19

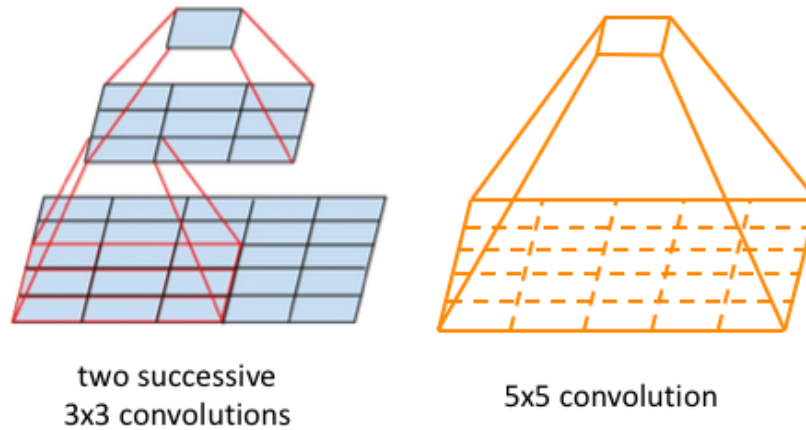


Figure 3.6: Convolution Example

scoring 7.5% top-5 error rate. I selected the largest VGGnet model, VGGnet-19, a network with 16 convolution layers and 3 fully-connected layers, because it has the best performance of all the VGGnet networks.

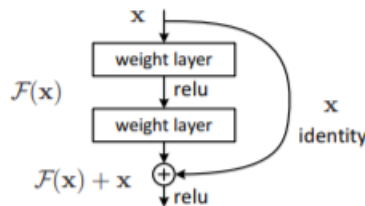


Figure 3.7: Residual Block used in ResNet He et al. (2016)

ResNet He et al. (2016) is an improvement on the VGGnet Simonyan & Zisserman (2014) based scheme. ResNet introduced the concept of residual mapping in the deep learning field. Convolutional Neural networks attempt to approximate a function, but as the network gets deeper they can become increasingly hard to train. By inserting these residual maps throughout the network, ResNet was able to achieve significantly deeper networks than its predecessors. Diverging from many previous models, ResNet models do not use any pooling layers within the convolutional

layers. A pooling layer is used at the beginning of the input to reduce the size of the image and highlight the localized features in each pool. The convolutional layers are also followed by a max pooling layer to aggregate the features once they have been extracted. Different from many other CNN, ResNet models use a much more limited fully-connected layer. Unlike VGGnet's two fully-connected layers feeding into a softmax layer to determine the image class, ResNet's convolutional layers output to a 2048-neuron layer that then feeds directly into a softmax layer for classification. As with the VGGnet model, the final softmax layer is changed from a 1000-neuron layer to a 2-neuron layer for the two output classes in this experiment. Two ResNet models were picked, ResNet-50 ResNet-101. ResNet achieved first in the ImageNet 2015 classification task with ResNet-50 and ResNet-101 scoring 5.25% and 4.60% top-5 error rate respectively. ResNet also achieved first in the ImageNet 2015 detection and localization tasks. The ResNet and VGGnet architecture can be found in Fig. 3.6.

3.2.2 Siamese Networks

The Siamese network used in these experiments is built on top of the deep CNN models: ResNet-50, ResNet-101, and VGGnet-19 models. These CNNs are run in place with the fully-connected layer removed. Both input images are run through the same CNN but with the fully-connected layers removed. These outputs are then flattened and run through a euclidean distance layer which calculates the absolute value of the difference between both flattened tensors. The output of the euclidean distance layer is then run through a fully-connected network. This framework is more functional for the VGGnet model. The VGGnet model uses a more rigid structure making it much simpler to strip off its fully-connected layers. The ResNet models instead use dimensional reduction throughout the model resulting in a much smaller output size from the convolutional layers. This makes the ResNet model less primed for a Siamese architecture. Simply stripping off the fully-connected layers at the end of the model does not expose the features being gathered by the convolutional layers. The output of the VGGnet feature network is 25,088 features while the output of the ResNet feature network is just 2,048 features. The flattened layers of the two frame-

works were also set differently. Because the ResNet was already being flattened as a product of its architecture, the flatten layer of the ResNet models was set to 1,024 neurons. The VGGnet had significantly more features and through experimentation it was found that the size of the flatten layer did not have a noticeable impact on performance, so the VGGnet's Siamese architecture had a flatten layer of just 256 neurons. The euclidean layer maintained the number of features coming from the flatten layer. The fully-connected network had two layers with the same number of features as output from the euclidean layer (1,024 for ResNet models and 256 for the VGGnet model) followed by a third layer which was a single neuron with Sigmoid activation to determine if the two images were from different families or the same family.

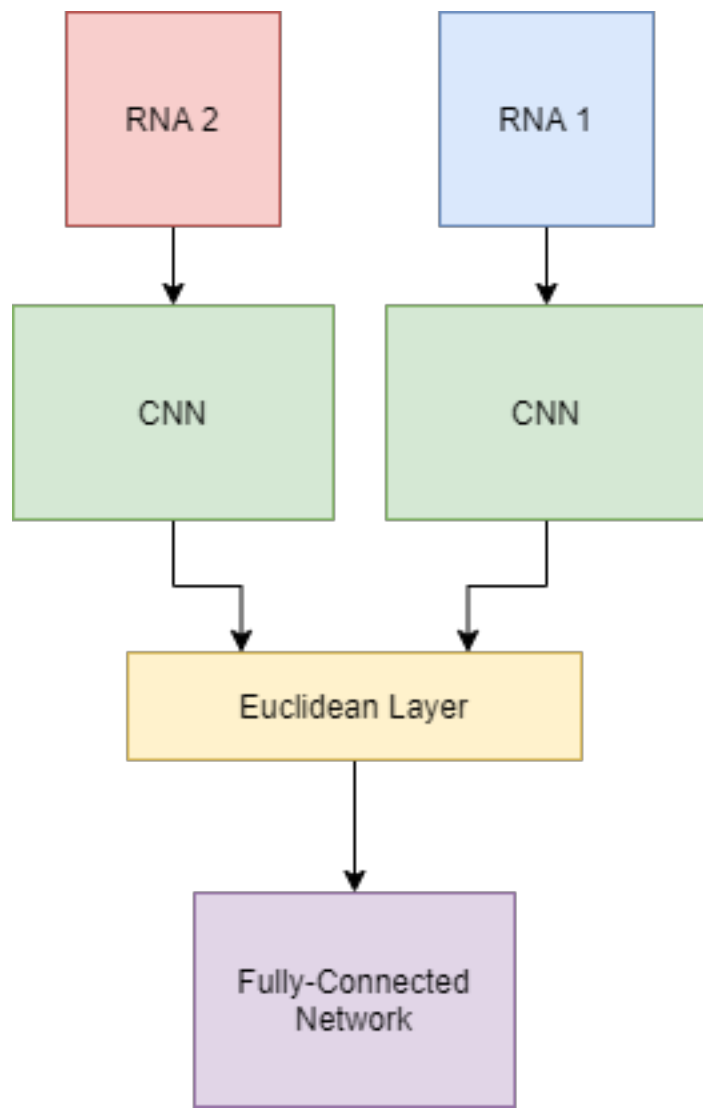


Figure 3.9: Siamese Network Architecture

Chapter 4

Experiment and Results

4.1 Experiment

4.1.1 Convolutional Neural Networks

In the first experiment I compared the performance of three models: VGGnet-19, ResNet-50, and ResNet-101. All models employ the SGD optimizer with a momentum factor of 0.9. The initial learning rates of ResNet and VGGnet-19 are set at 0.01 and 0.005, respectively, and the learning rates are reduced by a factor of 0.25 every 50 iterations. The batch size is set at 320. At each iteration, the model is trained on 320 images, selected randomly from the training dataset. Each training iteration took 18 seconds (all training was done on an Nvidia k40 GPU). The models are validated every 50th iteration on the entire validation set. Each validation epoch took 15 minutes. Training for each model took approximately six hours. Both ResNet models were trained for 600 iterations, and VGGnet-19 was trained for 700 iterations. After training, I choose the model with the best performance on the validation set. Then, I evaluate their performance using the average-class accuracy on the test set.

As shown in Table I, the two classes of the dataset are highly imbalanced. The size of different-family images is much larger than that of the same family. In order to reduce the influence of the class imbalance in the training set, different class ratios are chosen for the training stage. Three Different-same ratios were chosen: 1:1, 2:1, and 4:1, as shown in Table II. At the ratio of 1:1, for the batch of 320 training images at each iteration, an equal number of different-class and same-class images are chosen. However, at the ratio of 4:1, 256 and 64 images are selected randomly

Accuracy				
Model		VGG-19	ResNet-50	ResNet-101
Diff-Same Ratio: 1-1				
Train		92.3%	94.7%	96.9%
Val		78.2%	78.8%	80.0%
Test	Diff	80.0%	83.0%	84.0%
	Same	85.0%	82.0%	83.5%
	Avg	82.5%	82.5%	83.5%
	Total	84.7%	82.9%	84.0%
Diff-Same Ratio: 2-1				
Train		91.9%	92.8%	96.6%
Val		79.2%	81.5%	80.8%
Test	Diff	83.0%	86.0%	86.0%
	Same	82.0%	83.0%	84.0%
	Avg	82.5%	83.0%	85.0%
	Total	82.1%	85.8%	85.9%
Diff-Same Ratio: 4-1				
Train		90.0%	95.6%	95.6%
Val		77.5%	80.5%	80.0%
Test	Diff	87.0%	89.0%	90.0%
	Same	77.0%	81.0%	80.0%
	Avg	82.5%	85.0%	85.0%
	Total	86.5%	88.6%	89.5%

Table 4.1: Comparative results of different convolutional neural networks trained on different dataset ratios.

from the classes of different-family and same-family respectively so as to ensure more different-family images are involved in the training process. The different ratios only apply to the training, while for validation and test, no class ratios are used. Instead the entire validation and test sets are during their respective phases. The comparative results can be found in Table 4.1.

4.1.2 Siamese Network

The second experiment was done with Siamese networks. Both ResNet models, ResNet-50 and ResNet-101, as well as VGGnet-19 were tested within the framework of the Siamese network. Each model was trained for 400 iterations with each training iteration feeding 3200 image pairs to the Siamese network. Every 25 iterations the model was validated on the entire validation set. Each

model was trained with an initial learning rate of 0.001 which was decreased to 0.0002 after 300 iterations. The most successful iteration determined by the validation accuracy was then tested on the entire test set. Training the Siamese models took significantly longer, requiring about 20 hours to fully train. As with the CNN models, multiple different-same ratios were also experimented with during training of the Siamese networks. The models were initially trained with a 1:1 ratio, but given the high training accuracy of the VGGnet model, more extreme training ratios were used. As well as the 1:1 ratio, a 9:1 ratio and no ratio were also tested. No controls were applied to the validation or test sets. Full Siamese network results can be found in Table III.

Model		VGG-19	ResNet-50	ResNet-101
Different-Same Ratio: 1-1				
Train		98.9%	77.3%	74.3%
Val		72.5%	60.1%	62.8%
Test	Different	86.5%	74.4%	75.4%
	Same	70.9%	73.0%	78.4%
	Average	78.7%	73.7%	75.4%
	Total	85.7%	74.3%	76.9%
Different-Same Ratio: 9-1				
Train		96.5%	73.7%	71.7%
Val		72.9%	61.0%	61.8%
Test	Different	83.5%	70.5%	73.4%
	Same	72.9%	80.7%	76.6%
	Average	78.2%	75.6%	75.0%
	Total	83.0%	70.5%	73.6%
Different-Same Ratio: No ratio				
Train		95.7%	75.6%	76.0%
Val		73.0%	66.5%	62.1%
Test	Different	82.4%	75.6%	73.8%
	Same	76.2%	72.2%	76.3%
	Average	79.3%	74.9%	75.1%
	Total	82.0%	75.4%	73.9%

Table 4.2: Comparative results of different Siamese networks trained on different dataset ratios.

There is a big difference between the ratio used to control the training dataset of the stand-alone CNN models compared to the Siamese networks. With the stand-alone CNN, the model is learning to evaluate two different RNA within the same feature network. The CNN is being trained to learn features in each half of the image and then using the fully-connected network to evaluate

the similarity between these two patches. For this reason, the model needs to be fed a similar number of images from each class to learn the distinct features between the same and different classes. To moderate this learning process, training ratios of 1:1, 2:1, and 4:1 were chosen. When conducted with the 1:1 training ratio, this experiment can evaluate how well the stand-alone CNN approximates a balanced dataset. With the 4:1 training ratio, this experiment evaluates how well the stand-alone CNN approximates a larger, but less balanced dataset. The 2:1 training ratio was also included in the experiment to provide more data on how the balance of the dataset effects the efficacy of the model.

With the Siamese network, the feature network does not need to worry about identifying differences along the split between the two RNA. Instead, it just needs to learn the important features provided by the dot-plot of the secondary structure. Seeing more unique RNA during training will improve this feature network without fear of the classifier associating each RNA's features with a same family or different family class. For this reason a more aggressive training ratio was experimented with. The 1:1 ratio was kept to provide control and allow for more direct comparisons to be made between the training process of the stand-alone CNN and the Siamese network. With that, no set training ratio was experimented with to evaluate how well the Siamese network learns the RNA features. The 9:1 ratio was also experimented with to allow for more complex trends to be seen between the 1:1 training ratio and no training ratio.

4.2 Evaluation Metrics

There are several common methods for evaluating the success of neural networks. It is necessary to choose the right techniques to evaluate the network, based on the goals of the model. Frequently used evaluation metrics are: accuracy, sensitivity, specificity, and F-score. In binary classification, these evaluation metrics are defined by the models ability to correctly guess positive and negative samples correctly. Correctly labeling a positive is a true-positive (TP), mislabeling a negative as a positive is a false-positive (FP), correctly labeling a negative is a true-negative (TN), and mislabeling a positive as a negative is a false-negative (FN). Using these four definitions, all the

evaluation metrics can be defined:

$$Accuracy = \frac{TP+TN}{FN+FP+TN+TP}$$

$$Sensitivity = \frac{TP}{FN+TP}$$

$$Specificity = \frac{TN}{FP+TN}$$

$$F\text{-score} = 2 * \frac{sensitivity \times specificity}{sensitivity + specificity}$$

For these experiments, accuracy, specificity, and sensitivity were used. One other metric, average-class accuracy, was also used. For this thesis, specificity will be defined as the accuracy of the model when given two RNA from different families; sensitivity is the accuracy of the model when given two RNA from the same family; and average-class accuracy is defined for this paper as the average of the different-class and same-class accuracies. Because the test set for these experiments is very unbalanced, using pure accuracy as a metric would be deceptive to the performance of the model. A model could get over 94% accuracy by always classifying two RNA as different-class. Instead, I use the metric total-accuracy, defined for this paper as the accuracy of the model on the entire test dataset. Average-class accuracy is important because of the significantly larger size of the different-class data versus the same-class data. Were a model to always predict different-class, it could achieve a total accuracy of 95%. But, for the same reason, a model should predict different-class most of the time because the likelihood of two RNA being from the same family is so incredibly low. For that reason I also take into account total accuracy and look at the individual class accuracies to evaluate how well these models would work. A false-positive, indicating two RNA were predicted to be from the same class but not, would drastically hinder a clustering algorithm from correctly identifying families compared to a false-negative, indicating two RNA were predicted to be from a different class but weren't. A high total accuracy combined with a low sensitivity indicates that the model is much more likely to predict different-class as opposed to same-class. This is preferred to the inverse as it indicates that model is right a high-percentage of the time when predicting same-class.

4.3 Results

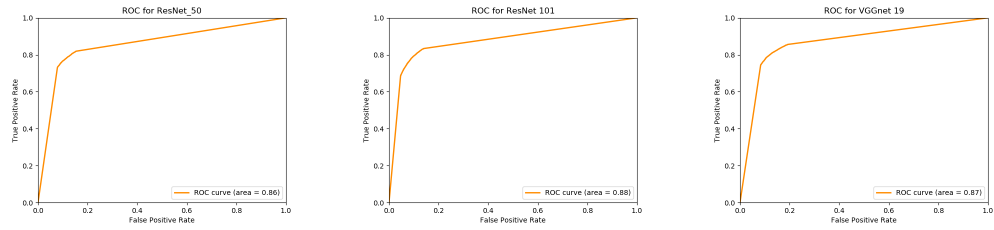


Figure 4.1: Receiver Operating Curves for CNN models.

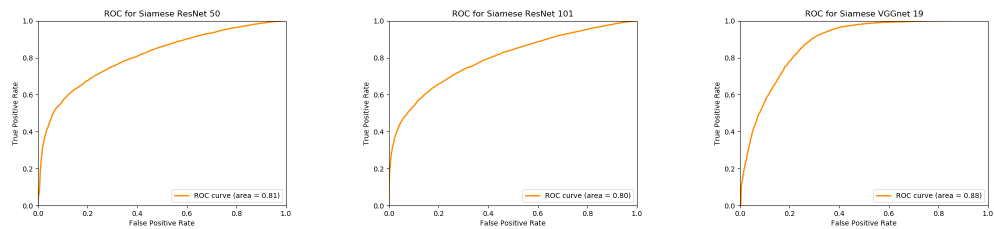


Figure 4.2: Receiver Operating Curves for Siamese models.

Both ResNet CNN models achieved an average-class accuracy of 85.0% for ResNet-50 and ResNet-101 respectively. The VGGnet-19 model performed slightly worse with a top accuracy of 82.5%. The best total accuracy of the CNN models was achieved by ResNet-101 with a total accuracy of 89.5% on the test set while the ResNet-50 model achieved a slightly worse 88.6%. As evidenced in the results, a higher Diff:Same ratio resulted in the model improving its accuracy on the different classes while decreasing its performance on the same class. However, even with the decrease in sensitivity, both the average-class accuracy and total accuracy increased for all models. The CNN models saw a performance drop when using the Siamese architecture. All three models had a reduced average-class accuracy when utilizing the Siamese structure, but the VGGnet-19 model was the best performing model when used in the Siamese architecture. The two ResNet models saw a significant reduction in training, validation, and test accuracies while seeing a much smaller reduction in testing accuracy, but the VGGnet model achieved higher training

accuracy than with the stand-alone CNN model. There was not a significant change in average-class accuracy across the different training class-ratios for the Siamese networks as there was with the CNN models. The difference between the performance of the VGGnet Siamese model and the ResNet Siamese models is likely due to their structure. For this experiment, only the fully-connected layers were removed for the feature-gathering part of the Siamese structure. However, the ResNet models shrink their output feature size at their base, reducing to only 2048 features. The VGGnet model reduces to a significantly larger 25,088 features. This leaves many more features to be passed to the euclidean layer and gives the VGGnet model much more power in the Siamese structure. With more experimentation and hands-on control over the exact layout of the ResNet models, the ResNet Siamese models should improve based on their better performance as stand-alone networks.

Chapter 5

Discussion and Future Work

5.1 Discussion

It is difficult to directly compare the output from the neural networks to the output of other methods because most other methods look at classification of very specific RNA families and compare classification accuracy based on pairwise sequence identity. These approaches ignore pairwise sequence identity and just look at strict classification accuracy. The closest approach in the literature is the most recent CNNclust Aoki & Sakakibara (2018), which employs one-dimensional convolutions on the primary and secondary sequences to look for common patterns in both. CNNclust only achieved 75.2% clustering accuracy when tested on ncRNA families not from the training set Aoki & Sakakibara (2018). This method is able to achieve better performance by utilizing two-dimensional convolutions on just the secondary structure. Once the models are trained, the models can process around 80 RNA comparisons per second. However, traditional approaches are very time-consuming since they have to perform pairwise alignment on their base-pairing probability matrices, which has a time complexity of $O(l^4)$ while l is the lengths of the ncRNA sequences.

These studies demonstrate the potential of applying CNN-based image classification models to RNA classification. In order to further improve these results, a few measures can be taken. One approach is to increase the dataset to include more RNA families by removing the limitations on sequence length. The distortion caused by increasing the disparity in sequence length may lower the accuracy. However, RNA sequence length amongst families does not vary much so it is more likely that the CNN's feature network recognizes this distortion as a strong indicator of different classes. Another possible measure is applying a clustering algorithm to the classification

results. These clusters can be determined by groups of RNA with a high percentage of same-family classifications amongst themselves. Clustering would also give more meaningful data to tune the hyper-parameters. I can also adopt a better approach to handle the imbalance issue of the dataset. As shown in Table II, when I choose 4-1 Diff:Same ratio, the test accuracy for different families is close to 90%, which means there is still room to further increase the performance of the CNN models. CNNs effectiveness as feature extractors on RNA secondary structure is demonstrated by their 85% test accuracy in this experiment. With more work, the Siamese networks, which utilize the CNNs for feature extraction, should be able to see at minimum comparable results.

5.2 Future Work

Further tuning of the ResNet models should yield better results within the Siamese architecture. One further study that can be done with little modification is removing the flatten layer from the Siamese network. Typically, Siamese networks use a flatten layer between the feature extractor and the euclidean layer to collapse the features and simplify the analysis. However, in this case it is possible that the flatten layer is acting as a small fully-connected network in between the CNN and the euclidean layer, obfuscating the features from the euclidean layer and resulting in worse accuracy. This is especially the case with the ResNet which already collapses the network and removes most of the fully-connected layer entirely. Siamese networks work by using a feature extractor to analyze two different inputs and then compare their features to determine their similarity, something that should work well with the feature gathering capabilities that have been consistently demonstrated by CNNs. The fact that the ResNet models performed worse in a Siamese architecture is more indicative of poor integration than it is that the entire approach is bad. The VGGnet performed only slightly worse in the Siamese architecture compared to the stand-alone environment. In fact, the VGGnet Siamese model had better training accuracy than the stand-alone model. This indicates one of three possibilities. One is that the Siamese network architecture can be tuned and will yield better validation and test accuracy. The other is that the Siamese network architecture overfits on the training dataset, resulting in worse validation and test accuracy. This would also

imply that the Siamese model is learning the features of the training set very well which is what is causing the overfitting. Removing the flatten layer and increasing the number of features for the fully-connected layer to classify on should decrease the overfitting and improve the validation and test accuracy. As a stand alone the model's fully-connected layer is evaluating 25,088 features while the Siamese network's fully-connected layers only has 256 features to evaluate.

Chapter 6

Conclusion

In this work, I presented a new approach for classifying RNA based on their secondary structure through image classification. By treating the BPPM representation of the secondary structure as an image, our approach takes advantage of the high speed and powerful feature extraction capabilities of CNNs. I demonstrated this approach to be a promising way to advance RNA analysis by providing a tool for more accurate and faster RNA classification, achieving 85% classification accuracy on long noncoding RNA. I also demonstrated this approach with Siamese networks and showed that CNNs are better able to analyze RNA secondary structure when the RNA are combined within the same image. The developed dataset can be taken as a benchmark set for any learning-based research on RNA classification.

References

- Aoki, G. & Sakakibara, Y. (2018). Convolutional neural networks for classification of alignments of non-coding rna sequences. *Bioinformatics*, 34(13), i237–i244.
- Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. (2016). Fully-convolutional siamese networks for object tracking. In *European conference on computer vision* (pp. 850–865).: Springer.
- Cen, F. & Wang, G. (2020). Boosting occluded image classification via subspace decomposition-based estimation of deep features. *IEEE Transactions on Cybernetics*, 50(7), 3409–3422.
- He, A., Luo, C., Tian, X., & Zeng, W. (2018a). A twofold siamese network for real-time object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4834–4843).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- He, L., Yu, M., & Wang, G. (2018b). Spindle-net: Cnns for monocular depth inference with dilation kernel method. *2018 24th International Conference on Pattern Recognition (ICPR)*.
- Herschlag, D. (1995). Rna chaperones and the rna folding problem. *Journal of Biological Chemistry*, 270(36), 20871–20874.
- Heyne, S., Costa, F., Rose, D., & Backofen, R. (2012). Graphclust: alignment-free structural clustering of local rna secondary structures. *Bioinformatics*, 28(12), i224–i232.
- Hochsmann, M., Toller, T., Giegerich, R., & Kurtz, S. (2003). Local similarity in rna secondary

- structures. *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003.*
- Hofacker, I. L. (2003). Vienna rna secondary structure server. *Nucleic Acids Research*, 31(13), 3429–3431.
- Hombach, S. & Kretz, M. (2016). Non-coding rnas: classification, biology and functioning. In *Non-coding RNAs in colorectal cancer* (pp. 3–17). Springer.
- Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. *arXiv preprint arXiv:1404.1869*.
- Jindal, S., Gupta, G., Yadav, M., Sharma, M., & Vig, L. (2017). Siamese networks for chromosome classification. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 72–81).
- Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., & Petrov, A. I. (2017). Rfam 13.0: shifting to a genome-centric resource for non-coding rna families. *Nucleic Acids Research*, 46(D1).
- Karklin, Y., Meraz, R. F., & Holbrook, S. R. (2004). Classification of non-coding rna using graph representations of secondary structure. *Biocomputing 2005*.
- Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2: Lille.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Li, K., Ma, W., Sajid, U., Wu, Y., & Wang, G. (2020). 2 object detection with convolutional neural networks. *Deep Learning in Computer Vision: Principles and Applications*, 30(31), 41.
- Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding rnas. *RNA biology*, 10(6), 924–933.

- Ma, W., Wu, Y., Cen, F., & Wang, G. (2020). Mdfn: Multi-scale deep feature learning network for object detection. *Pattern Recognition*, 100, 107149.
- Mccaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for rna secondary structure. *Biopolymers*, 29(6-7), 1105–1119.
- McClannahan, B., Patel, K., Sajid, U., Zhong, C., & Wang, G. (2020). Classification of noncoding rna elements using deep convolutional neural networks. *arXiv preprint arXiv:2008.10580*.
- Melekhov, I., Kannala, J., & Rahtu, E. (2016). Siamese network features for image matching. *2016 23rd International Conference on Pattern Recognition (ICPR)*.
- Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding rnas: insights into functions. *Nature reviews genetics*, 10(3), 155–159.
- Patel, K., Li, K., Tao, K., Wang, Q., Bansal, A., Rastogi, A., & Wang, G. (2020). A comparative study on polyp classification using convolutional neural networks. *PloS one*, 15(7), e0236452.
- Rawat, W. & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9), 2352–2449.
- Rivas, E. & Eddy, S. R. (2001). Noncoding rna gene detection using comparative sequence analysis. *BMC bioinformatics*, 2(1), 8.
- Roy, S. K., Harandi, M., Nock, R., & Hartley, R. (2019). Siamese networks: The tale of two manifolds. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3046–3055).
- Sajid, U., Sajid, H., Wang, H., & Wang, G. (2020). Zoomcount: A zooming mechanism for crowd counting in static images. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(10), 3499–3512.
- Sankoff, D. (1985). Simultaneous solution of the rna folding, alignment and protosequence problems. *SIAM Journal on Applied Mathematics*, 45(5), 810–825.

- Siebert, S. & Backofen, R. (2005). Marna: multiple alignment and consensus structure prediction of rnas based on sequence structure comparisons. *Bioinformatics*, 21(16), 3352–3359.
- Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smith, M. A., Seemann, S. E., Quek, X. C., & Mattick, J. S. (2017). Dotaligner: identification and clustering of rna structure motifs. *Genome Biology*, 18(1).
- Sorescu, D. A., Mohl, M., Mann, M., Backofen, R., & Will, S. (2012). Carna–alignment of rna structure ensembles. *Nucleic Acids Research*, 40(W1).
- Stefani, G. & Slack, F. J. (2008). Small non-coding rnas in animal development. *Nature reviews Molecular cell biology*, 9(3), 219–230.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Tao, R., Gavves, E., & Smeulders, A. W. M. (2016). Siamese instance search for tracking. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Washietl, S. (2012). Prediction of structural noncoding rnas with rnaz. *Comparative Genomics*, (pp. 503–526).
- Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F., & Backofen, R. (2012). Locarna-p: Accurate boundary prediction and improved detection of structural rnas. *Rna*, 18(5), 900–914.
- Xu, W., Keshmiri, S., & Wang, G. (2019). Adversarially approximated autoencoder for image generation and manipulation. *IEEE Transactions on Multimedia*, 21(9), 2387–2396.
- Zagoruyko, S. & Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

- Zagoruyko, S. & Komodakis, N. (2016). Wide residual networks. *Proceedings of the British Machine Vision Conference 2016*.
- Zhang, Y., Huang, H., Zhang, D., Qiu, J., Yang, J., Wang, K., Zhu, L., Fan, J., & Yang, J. (2017). A review on recent computational methods for predicting noncoding rnas. *BioMed Research International*, 2017, 1–14.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415.