

# Report for Residual Attention Network

Feng Su(fs2658), Shengjie Sun(), Yiming Tan()

December 2019

## 1 Abstract

“Residual Attention Network”, which is a CNN using attention mechanism which employs state-of-art feed forward network architecture in an end-to-end training fashion. The Residual Attention Network used in this report can be used to generate attention-aware features by consistently adding attention modules into the stacks. Inside every Attention Module, bottom-up top-down feed forward structure is used to unfold the feed forward and feedback attention process into a single feed forward process. When the process goes to deeper layers, attention-aware features from different modules are added up. + Brief conclusion for result

## 2 Introduction

Based on past information and history, we found out that there are not many attention mechanisms have been applied to feed forward network structure which can help us to achieve the state-of-art of image classification.

The Residual Attention Network is developed based on past attention mechanism and current advances by mixing them into a very “deep” architecture which composing multiple Attention Modules which generate attention-aware features. The attention-aware features from different modules change adaptively as layers going deeper

(1)Overall main structure: Residual Attention Network is constructed by stacking multiple Attention Modules. By using this kind of design, attention with different types can be better captured. However, stacking Attention Modules directly would lead to the obvious performance drop, in this case, we propose attention residual learning mechanism to optimize very deep Residual Attention Network with hundreds of layers. In addition, our model has a Bottom-up/Top-down structure as part of Attention modules to add soft weight. This structure can mimic bottom-up fast feed forward process and top-down attention feedback in a sin-

gle feed forward process which allows us to develop an end-to-end trainable network with top-down attention.

(2)Incorporates with state-of-the-art deep network structures: Our model can easily contains over hundreds of layers which can outperform past many of state-of-the-art residual networks

## 3 Methodology of the Original Paper

Residual Attention Network is constructed by stacking multiple Attention Modules. Each Attention Module is divided into two branches: mask branch and trunk branch. The trunk branch performs feature processing and can be adapted to any state-of-the-art network structures.

Instead of stacking Attention Modules, a simple approach would be using a single network branch to generate soft weight mask, similar to spatial transformer layer. In Attention Module, each trunk branch has its own mask branch to learn attention that is specialized for its features. [Insert ResidualNN here](#)

## 4 Result of the Original Paper

## 5 Methodology

## 6 Objective and Technical Challenges

## 7 Problem Formulation and Design

Give the detailed, one-to-one correspondence description of your design to attach/solve the problem and to achieve the objectives and the goal of the

project: Use engineering language and mathematical formulation; Provide system drawings, block diagrams, and/or circuit schematics for your software or hardware design, as applicable to your project; Include flow charts and pseudo code descriptions for the step-by-step discussion of your software design.

## 8 Implementation

One sentence to describe the organization of this section, then provide detailed discussion on your implementation.

## 9 Deep Learning Network

Provide the following descriptions and discussions: Architectural block diagram(s). Training algorithm details Flowchart(s) Data used

## 10 Software Design

Provide the following description and discussion: Flow chart or flow charts, very often you should provide one top level flow chart, then additional flow charts for detailed lower level implementations. Algorithm, e.g., description of the step by step implementation. Pseudo code for each section of the implementation.

## 11 Results

## 12 Project Results

Provide detailed Description of results Figures, plots Testing, verification

## 13 Comparison of Result

Provide detailed comparison between results of the original paper and the students' project Figure, plots showing differences in things such as training length/time, training/verification error, test error, other.

## 14 Discussion of Insights Gained

Provide detailed discussion, regardless of the actual results: why are your results different, did you use smaller dataset, did you use different hyperparameters, number of epochs different?

## 15 Conclusion

Provide summary of this project, briefly review the statements made in the abstract, in particular, if the enumerated objectives and goal are achieved. Emphasize and highlight the lessons learned, point out the direction for further improvement if needed.

## 16 Acknowledgement

Provide acknowledgement if needed, such as support, help, or assistance from someone. These support, help, assistance are crucial.

## 17 Reference

Include all references - papers, code, links, books. [1] Link to your bibucket or github [2] H. Li, "Author Guidelines for CMPE 146/242 Project Report", Lecture Notes of CMPE 146/242, Computer Engineering Department, College of Engineering, San Jose State University, March 6, 2006, pp. 1. [12] ...