



بسم الله الرحمن الرحيم

∞∞∞∞

تم رفع هذه الرسالة بواسطة / سامية زكى يوسف

بقسم التوثيق الإلكتروني بمركز الشبكات وتكنولوجيا المعلومات دون أدنى

مسئولية عن محتوى هذه الرسالة.

ملاحظات: لا يوجد





EMCA: EFFICIENT MULTISCALE CHANNEL ATTENTION MODULE FOR DEEP NEURAL NETWORKS

By

Eslam Mohamed Ali

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

FACULTY OF ENGINEERING ,CAIRO UNIVERSITY
GIZA, EGYPT
2022

EMCA: EFFICIENT MULTISCALE CHANNEL ATTENTION MODULE FOR DEEP NEURAL NETWORKS

By

Eslam Mohamed Ali

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Under the Supervision of

Prof. Dr. Mohsen A. Rashwan

Professor of EECE department
Electronics and Communications Engineering
Faculty of Engineering , Cairo University

FACULTY OF ENGINEERING ,CAIRO UNIVERSITY
GIZA,EGYPT
2022

EMCA: EFFICIENT MULTISCALE CHANNEL ATTENTION MODULE FOR DEEP NEURAL NETWORKS

By

Eslam Mohamed Ali

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Approved by the Examining Committee:

Prof. Dr. Mohsen A. Rashwan,

Thesis Main Advisor

Dr. Omar Ahmed Nasr,

Internal Examiner

Prof. Dr. Khaled Mostafa,

External Examiner

Professor of Computer Science department
Faculty of Computer Science, Cairo University

FACULTY OF ENGINEERING ,CAIRO UNIVERSITY
GIZA,EGYPT
2022

Engineer's Name: Eslam Mohamed Ali
Date of Birth: 02/03/1995
Nationality: Egyptian
E-mail: islam.bakr.2017@gmail.com
Phone: 01120300472
Address: Al Khamael Compound, Sheikh Zayed
, Giza, Egypt, 12588
Registration Date: 19/03/2019
Awarding Date: --/2022
Degree: Master of Science
Department: Electronics and Communications Engineering



Supervisors:

Prof. Dr. Mohsen A. Rashwan

Examiners:

Prof. Dr. Mohsen A. Rashwan (Thesis Main Advisor)
Dr. Omar Ahmed Nasr (Internal Examiner)
Prof. Dr. Khaled Mostafa (External Examiner)
Professor of Computer Science department
Faculty of Computer Science, Cairo University

Title of Thesis:

EMCA: Efficient Multiscale Channel Attention Module for Deep Neural Networks

Key Words:

Deep Learning; Machine Learning; Channel Attention Mechanisms; Convolution Neural Network; Classification

Summary:

In this thesis, we tackle the following question: can one consolidate multi-scale aggregation while learning channel attention more efficiently? To this end, we avail channel-wise attention over multiple feature scales, which empirically shows its aptitude to replace the limited local and uni-scale attention modules. Attention mechanisms have been explored with CNNs across the spatial and channel dimensions. However, all the existing methods devote attention to capturing local interactions from a uni-scale. Thus we propose EMCA, which is lightweight and can efficiently model the global context further; it is easily integrated into any feed-forward CNN architectures and trained in an end-to-end fashion. We validate our novel architecture through comprehensive experiments on image classification, object detection, and instance segmentation with different backbones.

Disclaimer

I hereby declare that this thesis is my own original work and that no part of it has been submitted for a degree qualification at any other university or institute.

I further declare that I have appropriately acknowledged all sources used and have cited them in the references section.

Name: Eslam Mohamed Ali

Date:

Signature:

Acknowledgements

All thanks and praises to **ALLAH** for providing the strength and time to complete my thesis besides my work.

I want to say a special thank you to **my parents, my brother, and my wife**. Without their help, I would not have achieved anything.

A special thanks also delivered to my supervisors, **Dr. Mohsen Rashwan** and **Dr. Ahmed El-Sallab**, as their guidance and support have made this an inspiring journey for me.

I would also like to thank the communication and electronics department at Cairo University and Zewail University, especially **Dr. Mohsen Rashwan** and **Dr. Tamer Ali**, for facilitating my experiments' required hardware (GPUs).

Table of Contents

Disclaimer	i
Acknowledgements	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
List of Symbols and Abbreviations	viii
List of Publications	x
Abstract	xi
1 CHAPTER 1: INTRODUCTION	xii
1.1 Problem definition	1
1.2 Motivation	3
1.3 Thesis Objectives	3
1.4 Contributions	4
1.5 Thesis Organization	4
2 CHAPTER 2: LITERATURE REVIEW	5
2.1 Deep Neural Networks	5
2.2 Convolution Neural Networks	8
2.3 Multi-scale	11
2.4 Channel attention	13
2.5 Spatial attention	19
2.6 Spatial and channel attention	20
3 CHAPTER 3: EFFICIENT MULTISCALE CHANNEL ATTENTION (EMCA) MODULE	22
3.1 Abstract Overview for Channel Attention Modules Framework	23

3.2	Revisiting Channel Attention Modules Integration	24
3.3	Avoiding Dense Integration	26
3.4	Multi-scale Incorporation	27
3.4.1	Coverage Region (R)	28
3.4.2	Channel Dimension Alignment (CDA)	29
3.4.3	Spatial Dimension Alignment (SDA)	29
3.4.4	Multi-scale Aggregation Block (MAB)	29
4	CHAPTER 4: EXPERIMENTS	34
4.1	Datasets	35
4.2	Implementation Details	40
4.3	Effect of Coverage Region (R)	41
4.4	Image classification	43
4.4.1	Image Classification on ImageNet-1K	45
4.4.2	Image Classification on Tiny-ImageNet	46
4.5	Object Detection	48
4.5.1	Benchmarking on COCO dataset	48
4.5.2	Benchmarking on KITTI dataset	49
4.6	Instance Segmentation	49
4.7	Robustness	49
4.7.1	Zero shot rotation immunity testing	50
4.7.2	Top-1 validation accuracy curves	50
4.7.3	Dissecting the produced learnable scales	52
4.7.4	Visualizing attention maps	52
5	CONCLUSION AND FUTURE WORK	56
5.1	Conclusion	57
5.2	Future Work	58
	References	59

List of Tables

2.1	Comparison of various channel attention modules (CA-modules).	21
3.1	Comparison of channel attention module by projecting each channel mechanism according to the abstract global context modeling skeleton that was introduced by GCNet [1] in Figure 4(a).	32
3.2	Comparison of various integration mechanisms, i.e., ALL, First, and Last, for integrating CA-module into Deep CNN backbone.	33
4.1	Effect of Coverage Region (R).	42
4.2	Comparison between our three proposed versions of EMCA module that adopt SE, ECA, and SRM as the cross channel excitation block producing, EMCA-SE, EMCA-ECA, and EMCA-SRM, respectively.	43
4.3	Comparison of different attention methods on ImageNet.	44
4.4	Comparisons with state-of-the-art attention modules on Tiny-ImageNet in terms of the number of parameters (#P.) in millions, GFLOPs, FPS, and top-1 accuracy.	47
4.5	Object detection results of different attention methods on COCO val2017.	48
4.6	Comparisons with state-of-the-art attention modules on KITTI-RGB [96] in terms of mAP	48
4.7	Instance segmentation results of different methods using Mask R-CNN on COCO val2017.	49
4.8	Analyzing the robustness of CA-modules on ImageNet.	50

List of Figures

1.1	Abstract overview for three well-known computer vision tasks; image classification, object detection, and instance segmentation.	1
2.1	Abstract overview demonstrates the critical difference between traditional machine learning algorithms and deep learning algorithms.	6
2.2	Abstract overview for an arbitrary convolution neural network (ConvNet). Where an input image is feed to a convolution layer, which consists of stacked filters, followed by a pooling layer. Finally, the spatial learned features are aggregated using a fully connected layer.	8
2.3	Example of the learned representation for two filters. F_1 captures the vertical edges, while F_2 captures the horizontal edges.	9
2.4	Illustration for the key difference between the conventional kernel and the dilated convolution layer.	10
2.5	Illustration for the key differences between the conventional convolution layer, the depth-wise convolution layer, and the point-wise convolution layer.	11
2.6	Abstract overview for the feature reuse concept and multi-scale learning paradigm.	12
2.7	Abstract overview of the channel attention modules.	13
2.8	Abstract overview of the spatial attention modules.	19
2.9	Abstract overview of the heterogeneous architecture that employ both spatial and channel attention modules.	20
3.1	Abstract overview demonstrates two possibilities of integrating a channel attention module (CA-module) into an arbitrary CNN backbone.	25
3.2	The upper part shows the diagram of our Efficient Multi-scale Channel Attention (EMCA) module. The lower part shows our proposed integration method that consolidates multi-scale information based on the Coverage Region (R), i.e., Equation 3.16.	28
4.1	Visualization for samples of ImageNet dataset [34].	36
4.2	Visualization for samples of Tiny-ImageNet dataset [94].	37

4.3	Visualization for samples of KITTI dataset [96].	38
4.4	Visualization for samples of MS-COCO dataset [95].	39
4.5	Training ResNet, local channel attention modules (LCA) baseline architectures and their EMCA counterparts on ImageNet validation set.	50
4.6	Comparison for the learned channel scales by our novel EMCA module against local channel attention modules. Better view with zooming in.	51
4.7	Sample visualization on ImageNet dataset [34] generated by GradCAM [109]	53
4.8	Sample visualization on ImageNet dataset [34] generated by GradCAM [109]	54

List of Symbols and Abbreviations

Symbol	Description
x	Input feature map
R	Coverage region
S	Number of the stages
N_{ij}	The number of the blocks' output in the stage i j utilized into the channel attention block C_i
w	Neural network parameters or weights
Z	Learned channel scales
σ	sigmoid activation function
H	Height dimension of input feature map
W	Width dimension of input feature map
C	Channel dimension of input feature map
$F(x)$	Function models CNN block
v	Non-linear activation function

Abbreviation	Description
AI	Artificial Intelligence
AvgPool	Average Pooling Layer
C1D	1 Dimensional Convolutional layer
CA-module	Channel Attention module
Cat	Concatenation operation
CDA	Channel Dimension Alignment
CNN	Convolution Neural Network
Conv	Convolution layer
DCT	Discrete Cosine Transform
DNN	Deep Neural Network

EMCA	Efficient Multi-scale Channel Attention module
FC	Fully Connect layer
GPU	Graphics Processing Unit
LSTM	Long Short-Term Memory
MAB	Multi-scale Aggregation Block
MAxPool	Maximum Pooling Layer
MLP	Multi Layer Perceptron
SDA	Spatial Dimension Alignment

List of Publications

1. E. Mohamed, A. El-Sallab and M. Rashwan, “PKCAM: Previous Knowledge Channel Attention Module” in Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems, ML4AD Workshop, 2021.