



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكروفيلم

بسم الله الرحمن الرحيم



HANAA ALY



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكروفيلم



شبكة المعلومات الجامعية التوثيق الإلكتروني والميكروفيلم



HANAA ALY



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكروفيلم

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات



يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



HANAA ALY



Cairo University

A NOVEL HYBRID MODEL FOR AUTOMATIC IMAGE CAPTIONING

By

Mariam Abd Elmohsen Mohamed Ramadan Mohamed Hafez

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
Computer Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

A NOVEL HYBRID MODEL FOR AUTOMATIC IMAGE CAPTIONING

By

Mariam Abd Elmohsen Mohamed Ramadan Mohamed Hafez

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
Computer Engineering

Under the Supervision of

Prof Dr. Magda B. Fayek

Dr. Mayada M. Hadhoud

Professor
Computer Engineering
Faculty of Engineering, Cairo University

Assistant Professor
Computer Engineering
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

A NOVEL HYBRID MODEL FOR AUTOMATIC IMAGE CAPTIONING

By

Mariam Abd Elmohsen Mohamed Ramadan Mohamed Hafez

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
Computer Engineering

Approved by the
Examining Committee

Prof. Magda B. Fayek,

Thesis Main Advisor

Prof. Elsayed E. Hemayed,

Internal Examiner

Prof. Reda A. Ahmed,

External Examiner

Dean of the Faculty of Computers and Information – Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

Engineer's Name: Mariam Abd Elmohsen Mohamed Ramadan
Date of Birth: 18/1/1992
Nationality: Egyptian
E-mail: mariam_abdelmohsen2015@yahoo.com
Phone: 01148117141
Address: 56 tereat gezeret Badran street shoubra-cairo
Registration Date: 1 /10 /2014
Awarding Date: / / 2021
Degree: Master of Science
Department: Computer Engineering



Supervisors:

Prof. Dr Magda B. Fayek
Dr. Mayada M. Hadhoud

Examiners:

Prof. Dr Magda B. Fayek (Thesis main advisor)
Prof. Elsayed E. Hemayed (Internal examiner)
Prof. Reda A. Ahmed (External examiner)
Dean of the Faculty of Computers and Information –
Cairo University

Title of Thesis:

A NOVEL HYBRID MODEL FOR AUTOMATIC IMAGE CAPTIONING

Key Words:

LSTM; Feature Extraction; Image Captioning; multi-class SVM; Alexnet network.

Summary:

A set of deep learning models and various datasets were tested to solve the problem by finding the link between the image features and the words it represented. The work was divided into two phases: the first was to extract features and determine classes. Various models were tested on different datasets (ImageNet, MS-COCO) to determine the effectiveness of their use. Combination of ALEXNET network, multi-class SVM was the best with accuracy 84.25%. The second was to generate captions, by entering the features and classes from the first stage. Various models were tested, and concluded that LSTM was the best model. The two phases resulted in a hybrid model of ALEXNET network, multi-class SVM and LSTM as the best model with accuracy 88.4%. The model was tested on the complete MS-COCO dataset, reaching an accuracy 90.7%, and was shown to reduce image processing time and high accuracy compared to previous models.

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: Mariam Abd Elmohsen Mohamed Ramadan Mohamed Hafez

Date: / /

Signature:

Acknowledgments

All praises and thanks are due to Allah, the most gracious, the most merciful, for providing me with the patience to complete this work.

I am grateful to my supervisors, Prof. Dr. Magda Fayek, and Assistant. Prof. Dr. Mayada M. Hadhoud to their guidance, advice, and encouragement toward the successful completion of this work. They were helpful, reading and correcting me all the way.

I express my gratitude to my parents for their Endless care and encouragement.

Finally, I would like to dedicate this work to my family.

Table of Contents

DISCLAIMER.....	V
ACKNOWLEDGMENTS	VI
TABLE OF CONTENTS.....	VII
LIST OF TABLES	X
LIST OF FIGURES	XI
NOMENCLATURE	XIII
ABSTRACT	XIV
CHAPTER 1 : INTRODUCTION	1
1.1. INTRODUCTION	1
1.2. MOTIVATION.....	1
1.3. PROBLEM STATEMENT.....	1
1.4. ORGANIZATION OF THE THESIS.....	2
CHAPTER 2 : MACHINE LEARNING MODELS	4
2.1. INTRODUCTION	4
2.2. MACHINE LEARNING MODELS	4
2.2.1. Artificial Neural Networks	4
2.2.1.1. Definition	4
2.2.1.2. Components Of Neural Networks.....	5
2.2.1.3. Neural Networks Architectures	6
2.2.1.4. Single Layer Feed-Forward Architecture	6
2.2.1.5 . Multiple Layer Feed-Forward Architectures	7
2.2.1.6 . Feedback Architecture.....	7
2.2.1.7. Mesh Network	8
2.2.1.8. The Training and Learning Processes.....	8
2.2.1.9. Weight Adjustment.....	8
2.2.2. Convolution neural networks.....	9
2.2.2.1 . Convolution layer	9
2.2.2.2. Pooling layer.....	10
2.2.2.3. Non-linear layers	11
2.2.2.4. Fully connected layer	12
2.2.3. ALEXNET CNN	12
2.2.3.1. The Architecture of ALEXNET CNN	13
2.2.4. Support Vector Machine.....	15
2.2.4.1. SVM Classification	17
2.2.4.2 . Training Multiclass SVM	20
2.2.5. Image Captioning	20
2.2.5.1. The Training Process.....	21
2.2.5.2. Testing Process.....	22
2.2.6. Long Short-Term Memory	22
2.2.6.1. Types of gates.....	22
2.2.6.2 The Training process of Long Short -Term Memory (LSTM).....	24

CHAPTER 3 : MACHINE LEARNING MODELS	25
3.1. INTRODUCTION	25
3.2. IMAGE CAPTIONING TECHNIQUES CATEGORIES	25
3.2.1. Statistical Based Models.....	25
3.2.2. Deep Learning Based Models:	26
3.2.2.1. Integrated systems:	26
3.2.2.2. Multi-stage systems:	27
3.3. SUMMARY	29
CHAPTER 4 : THE PROPOSED APPROACH	30
4.1. INTRODUCTION	30
4.2. CLASSIFICATION STAGE	30
4.2.1. Model 1: ALEXNET+VGG16 CNNS.....	30
4.2.2. Model 2: ALEXNET+VGG19 CNNS.....	31
4.2.3. Model 3: ALEXNET CNN.....	32
4.2.4. Model 4: ALEXNET CNN+SVM (SGD)	32
4.2.5. Model 5: ALEXNET CNN+ multi-class SVM	33
4.3. CAPTION GENERATION STAGE	33
4.3.1. Model 1: m-RNN.....	34
4.3.2. Model 2: ALEXNET CNN+SVM+RNN	35
4.3.3. Model 3: R-LSTM.....	35
4.3.4. Model 4: precise feature extraction (PFE).....	36
4.3.5. Model 5: ALEXNET CNN + multi-class SVM+LSTM.....	36
4.4. PROPOSED METHOD	37
4.4.1. Feature extraction using ALEXNET CNN	37
4.4.2. Classification using multi-class SVM	39
4.4.3. Caption generation using LSTM	39
4.5. SUMMARY	41
CHAPTER 5 : EXPERIMENTAL RESULTS	43
5.1. EXPERIMENTAL SETUP	43
5.2. DATASETS.....	43
5.3. ACCURACY MEASURE.....	43
5.3.1. Classification Accuracy	43
5.3.2. Captioning Accuracy	46
5.4. EXPERIMENTAL RESULTS.....	47
5.4.1. Classification Module Testing.....	48
5.4.2. Captioning Module Testing.....	52
5.5. SUMMARY	57
CHAPTER 6 : CASE STUDY: HYBRID MODEL ON MS COCO.....	58
6.1. INTRODUCTION	58
6.2. MSCOCO.....	58
6.3. HYBRID MODEL	58
6.3.1. Feature extraction	58
6.3.2. Classification	59

6.3.3.	Caption generation.....	60
6.4.	RESULTS	60
CHAPTER 7 : CONCLUSIONS AND FUTURE WORK		62
7.1.	CONCLUSION.....	62
7.2.	FUTURE WORK.....	62
REFERENCES		63

List of Tables

Table 4.1: ALEXNET Architecture Setup	38
Table 4.2: Different methods to reach the best method	41
Table 5.1: Table 5.1 Description of BLEU model	47
Table 5.2: Configuration of CNN's	48
Table 5.3: Classification result on (2500) ImageNet dataset based on A5/5	49
Table 5.4: Classification result on (2500) MS COCO dataset based on A5/5	49
Table 5.5: Classification result on (2500) ImageNet dataset based on A4/5	50
Table 5.6: Classification result on (2500) MS COCO dataset based on A4/5	50
Table 5.7: Classification result on (2500) ImageNet dataset based on A3/5	51
Table 5.8: Classification result on (2500) MS COCO dataset based on A3/5	51
Table 5.9: Captioning results on (2000) MS COCO dataset based on A5/5	52
Table 5.10: Captioning results on (2000) MS COCO dataset based on A4/5	53
Table 5.11: Captioning results on (2000) MS COCO dataset based on A3/5	53
Table 6.1: Captioning results on (40775) MSCOCO dataset	61

List of Figures

Figure 1-1 Description of Image Captioning Problem	2
Figure 2-1: Artificial Neural Network Definition	4
Figure 2-2: Artificial Neural Network Definition model	5
Figure 2-3: Single Layer Feed Forward Network	6
Figure 2-4: Multiple Layers Feed Forward Network	7
Figure 2-5: Feedback Network	7
Figure 2-6: Mesh Network	8
Figure 2-7: Weight Adjustment	9
Figure 2-8: Representation of Convolution Layer	10
Figure 2-9: Representation of pooling Layer	10
Figure 2-10: Representation of RELU Functionality	11
Figure 2-11: Representation of Nonlinear Function	11
Figure 2-12: Representation of Fully Connected Layer	12
Figure 2-13: ALEXNET Architecture	13
Figure 2-14: Tanh Function	14
Figure 2-15: RELU Function	14
Figure 2-16: Comparing RELU with Tanh	14
Figure 2-17: One Dimensional Line	15
Figure 2-18: Two-Dimensional Space	15
Figure 2-19: Three-Dimensional Space	16
Figure 2-20: Maximum Margins	16
Figure 2-21: Crossed Data	17
Figure 2-22: Decision Hyperplane	18
Figure 2-23: SVM classification strategies	19
Figure 2-24: Image Captioning Model	20
Figure 2-25: Image Based Module	21
Figure 2-26: Language Based Module	21
Figure 2-27: Sigmoid Function	22
Figure 2-28: Tanh function	22
Figure 2-29: Forget gate	23
Figure 2-30: Input gate and cell state	23
Figure 2-31: Output gate	24
Figure 3-1: Image captioning methods	25
Figure 3-2: Image captioning by transformation and summarization	27
Figure 4-1: The Classification stage	30
Figure 4-2: ALEXNET +VGG 16 model Architecture	31
Figure 4-3: ALEXNET +VGG 19 model Architecture	31
Figure 4-4: ALEXNET CNN model Architecture	32
Figure 4-5: ALEXNET CNN+SVM (SGD) model Architecture	32
Figure 4-6: ALEXNET CNN+ multi-class SVM model	33
Figure 4-7: Description of caption generation stage	34
Figure 4-8: m-RNN model Architecture	34
Figure 4-9: ALEXNET CNN+SVM+RNN model	35
Figure 4-10: R-LSTM model	35
Figure 4-11: PFE model Architecture	36
Figure 4-12: the hybrid model Architecture	36

Figure 4-13: Feature extraction process	37
Figure 4-14: Multi-class SVM classification	39
Figure 4-15: LSTM for captioning	41
Figure 5-1: Examples of 5/5 correctly classified image	44
Figure 5-2: Examples of 4/5 correctly classified images	45
Figure 5-3: Examples of 3/5 correctly classified images	46
Figure 5-4: Examples of 5/5 correct captioned images	54
Figure 5-5: Examples of 3/5 correct captioned images	55
Figure 5-6: Examples of 4/5 correct captioned images	55
Figure 5-7: Examples of incorrect captioned images	56