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

Ain Shams University
Faculty of Computer and Information Sciences
Information Systems Department
Bioinformatics Program

A Comprehensive Approach for Scenes Classification

A thesis submitted as partial fulfillment of the requirements for the degree of
Master of Science in Computer and Information Sciences

By

Mohamed Soudy Mohamed Ahmed
Bachelor's Degree in Bioinformatics,
Information System Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Nagwa Badr
Professor in Information Systems Department,
Dean of Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Yasmine Afify
Lecturer in Information Systems Department,
Faculty of Computer and Information Sciences,
Ain Shams University

2022
Cairo

Acknowledgment

Foremost, I would like to express my sincere gratitude to my advisor Dr. Yasmine Afify for the continuous support of my master's study and research, and for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master's study.

Besides my advisor Dr. Yasmine, I would like to thank Prof. Dr. Nagwa Badr for her guidance that helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master's study.

Moreover, I would like to thank Islam Ibrahim for his guidance and continued support through this thesis.

Last but not the least, I would like to thank my family for giving birth to me in the first place and supporting me spiritually throughout my life.

Abstract

Scene classification is one of the most complex tasks in computer vision. The accuracy of scene classification is dependent on other subtasks such as object detection and object classification. Machine and transfer learning are widely employed in scene classification achieving optimal performance.

Despite the promising performance of existing models in scene classification, there are still major issues. First, the training phase for the models necessitates a large amount of data, which is a difficult and time-consuming task. Furthermore, most models are reliant on data previously seen in the training set, resulting in ineffective models that can only identify samples that are similar to the training set. As a result, few-shot learning has been introduced.

We propose a scene classification system that can operate on various sizes of data sets and tend toward scene classification generalizability.

Our experiments include a novel machine learning model for scene classification applied to Intel scenes data set with a novel reconstruction of the data to serve binary and multi-classifications tasks. Moreover, we developed three novel architectures that use few-shot learning for scene classification achieving optimal accuracy compared to the existing models in the literature.

The proposed system includes models for few-shot learning that achieved accuracies of 52.16, 35.86, and 37.26 for five-shots on the MiniSun, MiniPlaces, and MIT-Indoor 67 benchmark datasets, respectively, while the proposed machine learning model achieved accuracies of 93.55, 75.54 for training and validation data of Intel scenes data.

List of Publications

1. Soudy, Mohamed, Yasmine Afify, and Nagwa Badr. "Insights into few shot learning approaches for image scene classification." PeerJ Computer Science 7 (2021): e666. DOI:10.7717/peerj-cs.666
2. Soudy, Mohamed, Yasmine Afify, and Nagwa Badr. "RepConv: A novel architecture for image scene classification on Intel scenes dataset." International Journal of Intelligent Computing and Information Sciences 22, no. 2 (2022): 63-73. DOI: 10.21608/IJICIS.2022.118834.1163
3. Soudy, Mohamed, Yasmine M. Afify, and Nagwa Badr. "GenericConv: A Generic Model for Image Scene Classification Using Few-Shot Learning." Information 13, no. 7 (2022): 315. DOI: 10.3390/info13070315

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CNN	Conventional Neural Network
DL	Deep Learning
FSL	Few Shot Learning
HOG	Histogram of Oriented Gradient
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LDO	Local Dominant Orientation
LLF	Low-Level Image Features
LSL	Low Shot Learning
ML	Machine Learning
MN	Matching Network
NNs	Neural Networks
OR	Object Recognition
PN	Prototypical Network
SC	Scene Classification
SR	Scene Recognition
SSD	Single Shot Multi-box Detector
YOLO	You Only Look Once

Chapter 1

Introduction

Chapter 1. Introduction

1.1 Overview

Scene classification is one of the most complex tasks in computer vision. The accuracy of scene classification is dependent on other subtasks such as object detection and object classification. Accurate results may be accomplished by employing object detection in scene classification since prior information about objects in the image will lead to an easier interpretation of the image content. Machine and transfer learning are widely employed in scene classification achieving optimal performance. With the profusion of image and video datasets, robust software, and efficient techniques are crucial for data retrieval and processing. Human brains can distinguish between multiple objects in real-time while software tools and algorithms strive to mimic the ability of the human brain. Even though different attempts are made to understand images, there is still room for enhancement.

Using object detection and recognition in scene classification has drawn much attention in the last decade with Object Recognition (OR) aiming to mimic the human ability to identify and distinguish between multiple objects in images. Object detection is segmented into two major subtasks: feature extraction and object classification. Various models are used in object detection such as You Only Look Once (YOLO) and Single Shot Multi-box Detector (SSD) with the ability to achieve optimal performance. Researchers using this approach rely on the hypothesis that understanding and recognizing objects will lead to an easy classification of scenes. Researchers use one or more object detectors to optimize and enhance classification accuracy.

Furthermore, other researchers made various attempts in the SR task using Low-Level Image Features (SR-LLF), which aim to use low-level features including color, orientation, global multi-scale orientation, and Local Dominant Orientation (LDO), and texture to understand and classify scenes. The theory behind this algorithm relies on classifying the scene without identifying the said object. Most of the research works try to find descriptors that represent the low-level features and use these descriptors for the scene classification. As an extension of this approach, researchers employ methods for better selection of descriptors that enhance classification accuracy. Researchers that use this approach justify its preponderance over OR method by the theory of error propagation, where errors in OR will lead to the wrong classification of SR. Nevertheless, using low-level features (pixels raw value) demonstrates a convenient performance, with a dramatic increase in the image complexity, leading to the successful implementation of robust and sophisticated models.

Robust low-level image features have proven to be effective representations for a variety of high-level visual recognition tasks, such as object recognition and scene classification. But as the visual recognition tasks become more challenging, the semantic gap between low-level feature representation and the meaning of the scenes increases. Moreover, the visual task becomes higher and higher level, and the limitations of low-level features become more obvious. That led to the more usability of sophisticated models using attribute/object detection methods that outperformed the low-level representation methods.

Despite the promising performance of existing models in image understanding and scene classification, there are still major issues. First, the training phase for the models necessitates a large amount of data, which is a difficult and time-consuming task. Furthermore, most models are reliant on data previously seen in the training set, resulting in ineffective models that can only identify samples that are similar to the training set.

Meta-learning deciphers these limitations as it does not require a large number of training samples and it generalizes the model to be learned and evaluated in novel classes as never seen before. Meta-learning is based on the premise that if a child has seen one or two pictures of a cat, he will be able to classify new pictures proficiently, reflecting the theory of learning by experience. Meta-learning also incorporates the concept of "learning to learn". The branch of meta-learning known as Few-Shot Learning (FSL) is observing a dramatic increase in research. Also known as Low-Shot Learning (LSL), it is a form of machine learning problem in which the training dataset contains only a small amount of data. The model is trained using well-defined episodes representing various classification tasks. The training set is split into two subsets (train; test) in each iteration to update the gradient and obtain the best weights for the learning process. Few-shot learning aims to generate a semi-generalized model that is able to classify novel classes using a small training set and reduce the cost of the data collection and training of the model.

1.2 Research Problem

Image comprehension and scene Classification are critical challenges in computer vision. The advancement of technology and the abundance of existing datasets provide enough opportunity for advancement in scene classification and recognition research. Despite the excellent performance of current machine learning models in image interpretation and scene categorization, there are still challenges to solve. All models are data-dependent, and they can only categorize samples that are similar to the training set. Furthermore, these models need a vast amount of data for training and learning and there is no generic model that is able to classify any scene regardless of the source of the data.

1.3 Research Motivation

The analysis and review of existing approaches for scene classification led to many research questions that include:

I. Do the data set quality and size significantly contribute to classification accuracy?
II. Is the model complexity a crucial factor in the scene classification accuracy?
III. How do the model hyperparameters affect the model performance?

1.4 Research Objective

The objective of this thesis is to propose a scene classification system that can operate on various sizes of data sets and tends toward scene classification generalizability.

1.5 Research Contributions

Motivated by the aforementioned research problems, we developed a system capable of classifying scenes without prior knowledge of the data. Our approach emphasizes generalizability by concentrating on models and their hyperparameters rather than developing a model for unique data that will be worthless for other datasets that would necessitate additional work.

Our experiments include a novel machine learning model for scene classification applied to Intel scenes data set with a novel reconstruction of the data to serve binary and multi-classifications tasks. Moreover, we