UNIVERSITY KASDI MERBAH OUARGLA
Faculty of New Information and Communication Technologies
Department of Computer Science and Information Technology

ACADEMIC MASTER Thesis

Domain: Computer Science and Informtion Technology
Specialty: Fundamental Computer Science
Prepared by :

BENSALEM Fatma Zohra

Topic

Deep Semi-Supervised Multi-Label image Classification

Supported publicly the : ../06/2021 Before the jury:

Dr.BOUANANE Khadra	Presdentr	University UKM Ouargla
Ms. DEROUICHE Chaima	Supervisor	University UKM Ouargla
Dr.DEBBAGH Farah	Supervisor	University UKM Ouargla
Dr.BEKKARI Fouad	Examiner	University UKM Ouargla

College year: 2020/2021

Acknowledgments

First of all, we thank our Almighty God who helped us and gave us the patience and courage during our years of study and to achieve this work. My big thanks to my supervisor Ms.DEROUICHE Chaima for her great guidance, advice and support to do this work in all the year. My thanks for prof.Kherfi Mohammed Lamine and Dr. Debbagh Farah for their supervision and trusting us. My thanks to Dr. BOUANANE Khadra and Dr. BEKKARI Fouad for accepting the evaluation of the work. I would like to thank my parents for all the support and love they have given me during my studies. I also thank my sisters, my friends and my family, for their support during this year

Dedication

To my parents; the reason for my existence in life and my success.

To my sisters brothers; greatest gift my parents have ever gave me.

To my colleague Benmohammed Nassim; who did not skimp on helping me throughout the stages of completing this work.

To my teachers; who have been with us throughout our academic career.

To my family; evryone halped me, seported me all the time:
To my best freinds; they were by my side at all times.

Abstract

The multi-label image classification (MLC) is the process that aims to firstly learn from training set of images, where each one can belong to multiple classes and so after be able to predict more than one class label simultaneously for a new tested image. This process suffers from several problems such as overlapping meaning that may contain image labels. In this thesis, we present a Deep Semi-supervised multi-label image classification method that compose modules: CAE module and ResNet module . The first module consist of a CAE model that is used to extract the features of images. The second one, is a ResNet is used to classify the extracted features.The proposed model has been trained on public benchmark dataset and it achieves better results compared to state of the art.

Keywords: MLC, CAE, ResNet, deep semi-supervised. deep learning.

Résumé

Classification d'images multi-label (MLC) est le processus qui vise d'abord à apprendre à partir d'un ensemble d'images d'apprentissage, où chacune peut appartenir à plusieurs classes et ainsi pouvoir prédire plus d'une étiquette de classe simultanément pour une nouvelle image testée. Ce processus souffre de plusieurs problèmes tels que le chevauchement de sens qui peut contenir des étiquettes d'image. Dans cette thèse, nous présentons une méthode de classification d'images multi-labels Deep Semi-supervisée qui compose des modules : module CAE et module ResNet. Le premier module consiste en un modèle CAE qui est utilisé pour extraire les caractéristiques des images. Le second, est un ResNet est utilisé pour classer les caractéristiques extraites. Le modèle proposé a a été formé sur un ensemble de données de référence public et il obtient de meilleurs résultats par rapport à l'état de l'art.
mot-clé: MLC, CAE, ResNet, profond semi-supervisé., l'apprentissage profond.

الملخص

تصنيف الصور متعدد الملصقات هو العملية التي تهدف إلى التعلم أولاً من مجموعة الصور التدريبية ، حيث يمكن أن تنتمي كل واحدة إلى فئات متعددة ، وبالتالي بعد ذلك تكون قادرة على التنبؤ بأكثر من ملصق فئة واحد في

١٨ ارر■رس٥ اڤرتسسمنتس وقت واحد للحصول على صورة جديدة تم اختبارها. تعاني هذه العملية من عدة مشاكل مثل تداخل المعنى الذي قد يحتوي على تسميات للصور. في هذه الأطروحة ، نقدم طريقة تصنيف صور متعددة الملصقات شبه خاضعة للإشراف العميق والتي تتكون من وحدات: وحدة ثاي ووحدة غس ت. تتكون الوحدة الأولى من نموذج ثاي المستخدم لاستخراج ميزات الصور. والثاني ، هو غس ت ويستخدم لتصنيف الميزات المستخرجة. النموذج المقترح لديه تم تدريبه على مجموعة بيانات مرجعية عامة ويحقق نتائج أفضل مقارنة بأحدث ما توصلت إليه التكنولوجيا.

الكلمات المفتاحية : MLC ، CAE ، $ResNet$ ، شبه مشرف عميق ، تعلم عميق

Contents

List of Figures

List of Tables

Chapter1

General Introduction

Chapter 1

General Introduction

1.1 Introduction

The classification is process , that aims to categorize data into one or more predefined classes (labels). The image classification is a process, which is basing on the visual content of images(data). The image classification can be categorized into two categories which are single-label classification and multi-label classification. multi-label image Classification(MLC) is the process to determinate several properties of the content of the image[1].

Nowadays, we tend to notice that multi-label classification strategies are progressively needed by modern applications, such as protein function classification[2], music categorization [3] and semantic scene classification.[4] A several solutions have been proposed .In wide variety of literature works[5], which be classified into two categories[6]:

- Problem transformation methods[7]: such that the MLC problem is simplified(divided) to sub classification(singel-label classification).
- Algorithm adaptation methods[7]: this explicit methods, it adapt the algorithms to do the MLC directly.For example, ML-DT[23], Rank-SVM [24], and CML [25]; this techniques(algorithms) used only labeled data which is limited and it expensive to label data.

The multi-label classification is more complex than the single-label. That due, it should assign more than one label per one example of input data. This thesis focuses on deep semi-supervised multi-label classification applied on images as a data. There are many applications where assigning multiple classes to an image is necessary according to its complexity. So, classify such image into one class (category) is not effective. This is what the multi-label classification came to solve it.

1.2 Problematic

So far, multi-label image classification is considered as a challenge. All this is due to the complexity of images and labels information and the overlapping meaning(Image tags can overlap in the meanings [1]). Thus, the emergence of this domain, many works, solutions, and methods have been proposed to solve this problem, and each was not feasible enough, especially with the lack of labeled data and it is expensive to label the data. So far, MLC has become as a challenging issue. This last one is a difficult task, especially since each image has several annotations: Is there a model using deep learning that could solve this problems?

1.3 Motivation

Therefore, a semi-supervised learning needed indeed, despite the huge amount of works on multi-label image classification, there is still a required effort that must be done to consider the very important aspect of the overlapping meaning within the classification task and to eliminates the lack of labeled data. This motivate us to explore this aspect to look for a MLC solution considering semi-supervised learning.

1.4 Thesis structure

In addition to what we proposed in the current chapter on a general introduction, this thesis has been organized as it follows:

- In chapter 2, we start by presenting the image classification: its definition, categories especially the Multi-label classification. After that, we present the different machine learning families. Then, we will focus on semi-supervised learning: its formal definition, its main challenges, assumptions behind these challenges and application domain.

  Finally, we give a brief passage on deep learning architectures and important terminology that we use along our work.

- In chapter 3, we present related works on multi -label image classification .
- In chapter 4, we propose our solution for multi-label image classification with the experiments.
- General conclusion and perspective work.

Chapter2

Image Classification, Semi-supervised learning and Deep learning

Chapter 2

Image Classification, Semi-supervised learning and Deep learning

2.1 Introduction

The Multi-label image classification is one of Machine Learning techniques (tasks), used to classify data(image) into more then one categories, under the goal of facilitate the study of a large number of data. Multi-label image classification represent a big challenge in computer vision applications. Therefore, it is required in several applications of computer science under different fields in the world. This is why this task use the deep learning according to the achievements in this field .
In our thesis we focus on (MLC). However, after presenting the (MLC), we give the principle of deep learning and some of deep neural networks.

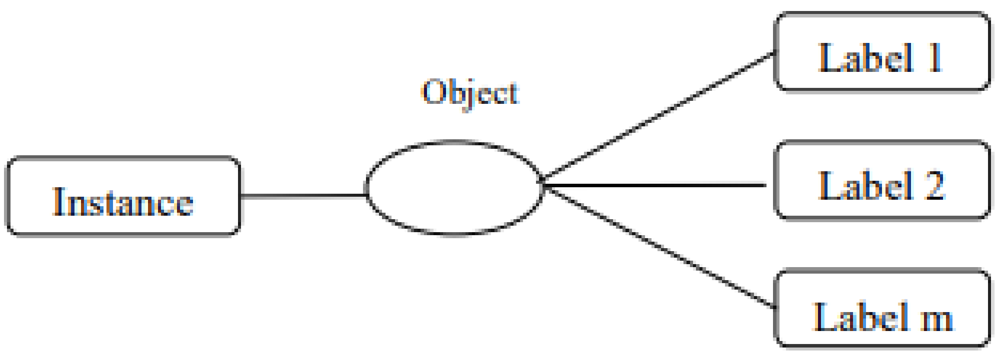


Figure 2.1: Explanation of the Multi-Label image classification[9].

2.2 Image Classification

2.2.1 Definition

A classification task is to assign an input data X_i from a data set D, that is represented by a d dimensional feature vector from R^d, to a set of k known classes (categories)[10]. $Y = \{y_1, y_2, \cdots, y_N\}$ [11]. In other words, the learning algorithm is asked to produce a function $f : R^d \longrightarrow 2^y$ that maps any given input X_i to its corresponding classes. Examples on classification algorithms: Neural Networks, Support Vector Machine, K Nearest Neighbours, Naıve Bayse... etc.

2.2.2 Image classification categories

Image classification is a predictive topic, it aims to learn a pattern in generally labeled images to predict the label or class of a new image. Classification attributes are divided into two subsets. The first contains the input characteristics and forecast variables. The second subset contains the output attributes and the class of each instance. Thus, depending on the nature of the outputs of the second subset, different types of classification can be identified.[12] We can define two important categories of image classification which cover all varieties of image classification problems [13]. However, these categories are not limited to images, but can have any form of input data.

Single-Label classification:

In this category, the input data have one label in the output. This category is divided into two sub categories:

1. **Binary-label classification:**

   This category aims to assign a given input $x_i \in D$ to its associated label $y_i \in Y$ where $\|Y\| = 2$. In other word, classify an input to only one output where the output space consists of two values Yes or No (1 or 0) [14]. For example (Figure2.2), the output value of the cat image is 1 (positive value), while the output value of the dog image is 0 (negative value), for the same classifier.

2. **Multi-class classification:**

   Multi-class classification can be seen as a generalization of binary classification. Just as binary classification involves predicting whether the image is from one of two classes (positive or negative). (Figure2.3(a)), the input image in multi-class classification can be categorized exactly into one label from a certain predefined label set (Figure2.3(b)) [13]. Therefore, the output of the classification is a vector with one positive value (corresponding to the image label), and zeros otherwise.

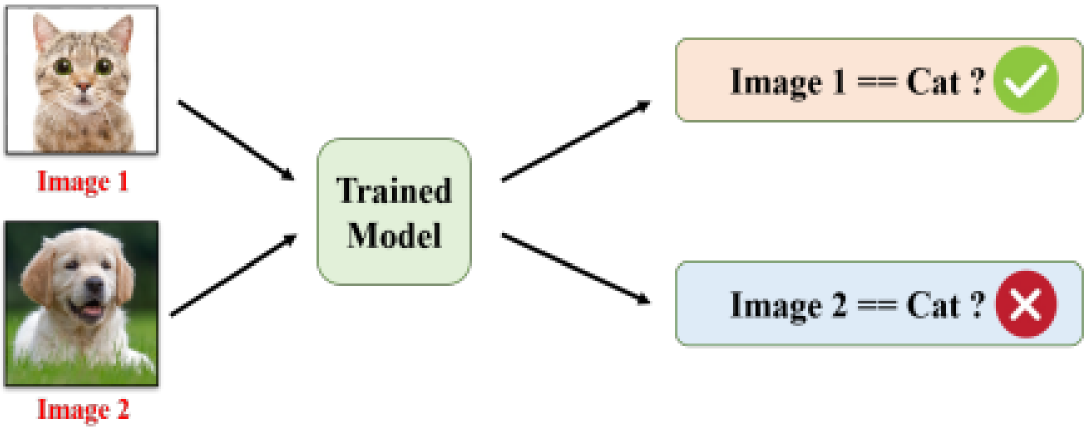


Figure 2.2: Simple example of binary image classification[15].

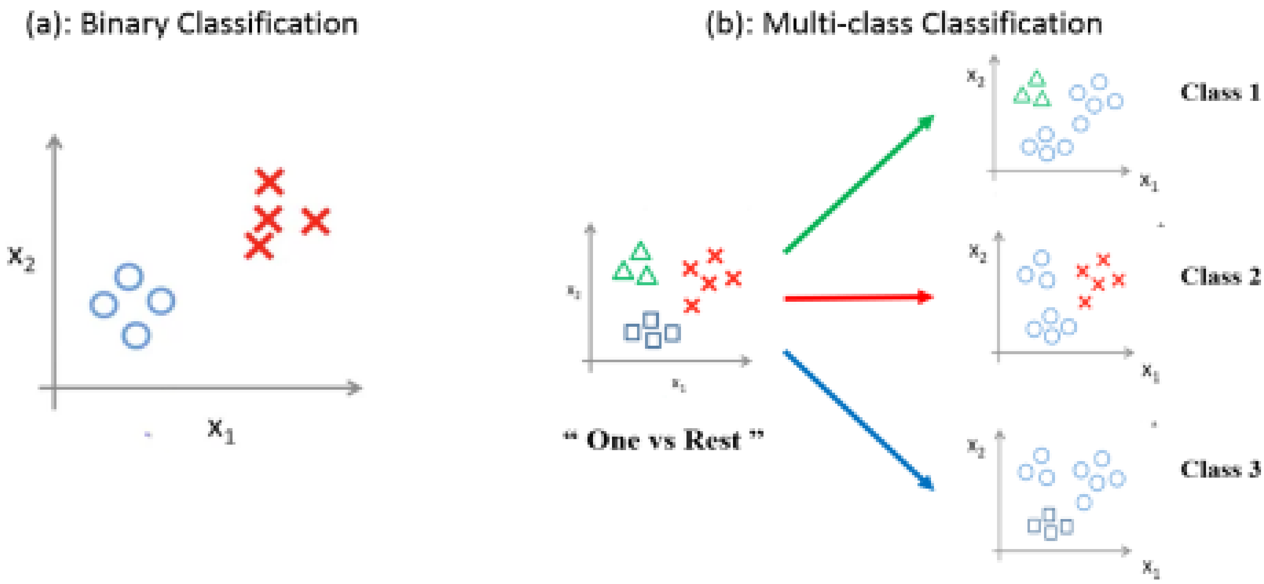


Figure 2.3: Graphic representation of machine learning classification task (binary classification and multi-class classification)[16].

Multi-Label classification (MLC):

Multi-Label category is greater complex. The enter data (and for us the enter image) can concurrently related to a couple of class[13].This category aims to categorise an input x_i to its associated classes $Y_i' \subset Y$, where $\|Y\| > 2$ and $\|Y_i'\| \geq 1$. In other words, classify an input to at least one class (one or more classes) relying on recognizing patterns in an input [14].

We will present with more details the Multi-label image classification in the coming section.

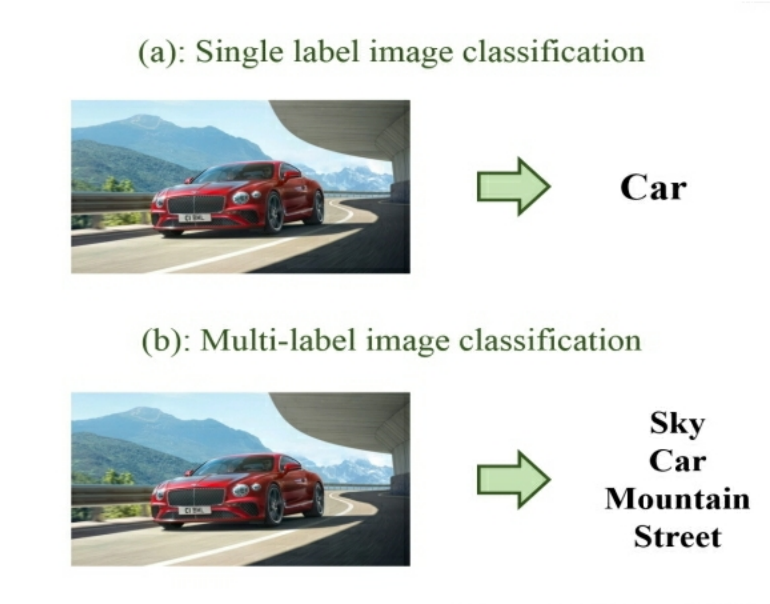


Figure 2.4: Single label image classification and multi-label image classification.

We also mention other types of classification :

- **Multidimensional Classification:** Works within-side the equal manner of multi-label classification, the enter date may be related to a couple of label simultaneously. Except that the output vector isn't confined to binary values, it carries any values shape a predefined set [17].
- **Multiple Instance Learning:** It's a learning paradigm to represent each example by groups of input data (instances) that he called bags. The output tag belongs to the bag instead of an instance [17].

The table below presents the different kinds of classification :

Classification kind	Output type	Number of output
Binary	Binary	1 per instance
Multi-class	Multi-valued	1 per instance
Multi-label	Binary	n per instance
Multi-dimensional	Multi-valued	n per instance
Multi-instance	Binary / Multi-valued	1 per instance

Table 2.1: Classification types attending to the output to be predicted [17].

2.3 Multi-label Image Classification

2.3.1 Definition

Multi-label classification (MLC) of images is a task that aims to recognize the different objects or labels in images. It is more complicated than the single label classification, because it focuses on discovering more than one label per image. [18] The multi-label classifier return a vector of output values, unlike the single label classifier which return one value [19].

(a) Single label Classification

Binary Classification

Instance	Class A
1	1
2	0
3	0
4	1
5	0

Multi-class Classification

Instance	Classes [A,B,C,D,E,F]
1	[1,0,0,0,0,0]
2	[0,0,1,0,0,0]
3	[0,0,0,0,0,1]
4	[0,0,0,0,1,0]
5	[0,1,0,0,0,0]

(b) Multi-label Classification

Instance	Classes [A,B,C,D,E,F]
1	[1,0,0,0,1,0]
2	[0,0,1,0,0,0]
3	[1,1,0,0,0,1]
4	[0,0,0,1,1,0]
5	[0,1,1,1,0,0]

Figure 2.5: The difference between single label and multi-label image classification.

Formal definition [19] [45]

The MLC task consists to learn a function $H : x \longrightarrow 2^q$

- $x = R^d$ is the d-dimensional visual feature space of images.
- • $Y = \{y_1, y_2, \cdots, y_N\}$ is the label space with q possible class labels.
- An image I has two feature vectors:
  1. d-dimensional visual vector xi $\in X/x_i = \{x_{i1}, x_{i2}, \cdots, x_{id}\}$.
  2. The output label vector Y_i $\subseteq$Y.
- To learn the function H we use a training set$D = (x_i, , Y_i)_i \in [1, m]$ of m labeled images.
- For each instance x of X , the multi-label classifier predicts H(x) $\subseteq$ Y as the set of proper labels for x.

2.3.2 Applications domains

- Movie Genre Detection from a Movie Poster

  (K.Kundalia et al.,2020) used the movies posters as input images to predict the movie genres. Also, they created a large dataset on this subject. [21]
- Genetics/Biology : example, analyzing protein properties and gene expression.
- Medical image analysis to diagnose multiple diseases in the same organ of the human body. For example:

  (Chen.H et al.,2019), proposed a deep Hierarchical Multi-Label Classification (HMLC) to facilitate the Computer-Aided Diagnosis (CAD) for the Chest X- rays (CXRs). [22]
- Social media domain example:

  Recently, (Lui.L et al., 2020) developed a multi-label convolutional neural net- work model (BrandImageNet), to predict perceptual brands in the consumers images on social media. [23]

2.4 Semi-supervised learning

As our work derives from this family of machine learning, we devote this section to present semi-supervised learning in more detail. Before introducing semi-supervised learning, let us give the principle of each category of learning, in order to distinguish them. The separation between these three categories of approaches is based on the data set they experience during training.

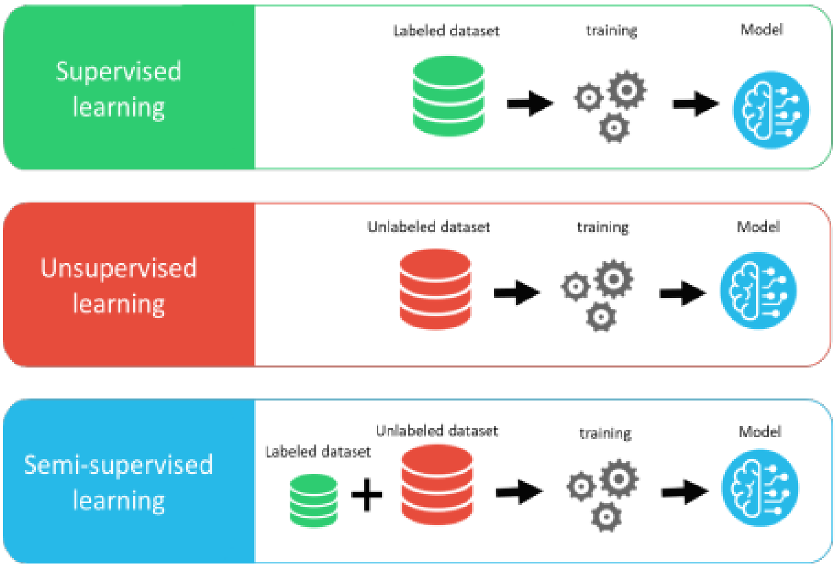


Figure 2.6: Some of machine learning approaches relying on training dataset in the top Supervised learning, in the middle unsupervised learning and in the bottom semi-supervised learning

Supervised learning uses a **labeled** data set to train a machine learning model. The term supervised comes from the supervision of the training data labels, which will act as a teacher in the training phase [11] [46].

In supervised learning, a training data set D is denoted:

$$D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$$

- Where x_i is the input data and y_i is its corresponding label.

The advantage of using labeled data is that it oversees the model to train precisely on the desired goals (point of view), forcing it to find the exact model that correlates between the input data and the labels. In other words, the labels will provide feedback during model training to improve accuracy. In the case of classification, the input (tagged) data provides tag information to the classification algorithm to find the decision boundary that separates the classes.

The algorithm learns to associate an input with its corresponding output [11]. In other words, the algorithm constructs a function f that maps input x to its corresponding output y. The function f is built based on the examples of the training sets, the purpose of the function f is to correctly predict the output of new data (input).

Unsupervised learning uses an **unlabeled** data set to train a machine learning model [11]. On the other hand, unsupervised learning allows the model to extract features from unlabeled data, however, since the dataset is unlabeled, there is no supervision in the d phase. 'learning, that is, there is no right answer (no class or target), therefore, a model cannot perform some machine learning tasks that require tagged data.

In unsupervised learning, a training data set D is denoted:

D = $D = \{x_1, x_2, \cdots, x_n\}$

Where x_i is the input data.

2.4.1 Definition:

For a dataset D noted:$D = \{D_l \cup D_u\}$ $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ and $D = \{x_1, x_2, \cdots, x_m\}$

Where(x_i, y_i) are sampled from a unknown joint distribution p(x, y) And $x_j$sampled from unknown distribution p(x).

SSL aims to produce a function f(x) that predicts a correct label y. Semi-supervised learning leverages a hybrid data set of a small amount of tagged data and a large amount of unlabeled data to train a model. The term semi-supervised learning arose from the combination of supervised learning data (labeled data) and unsupervised learning (unlabeled data).

Semi-supervised learning assumes that unlabeled data can provide additional information that helps the learning algorithm find a good decision boundary with only a few labeled data.

2.4.2 Semi-supervised learning main challenge

The most obvious challenge in semi-supervised learning is how to enable a model to learn from both the unlabeled part D_u and the labeled part D_l of a dataset D, thus extending the SSL approach to the extreme scenario where $\|D_L\| \ll \|D_U\|$. In other words, how to use the unlabeled part of a given dataset to train a machine learning model to better generalize and overcome other learning approaches using semi-supervised learning [23].

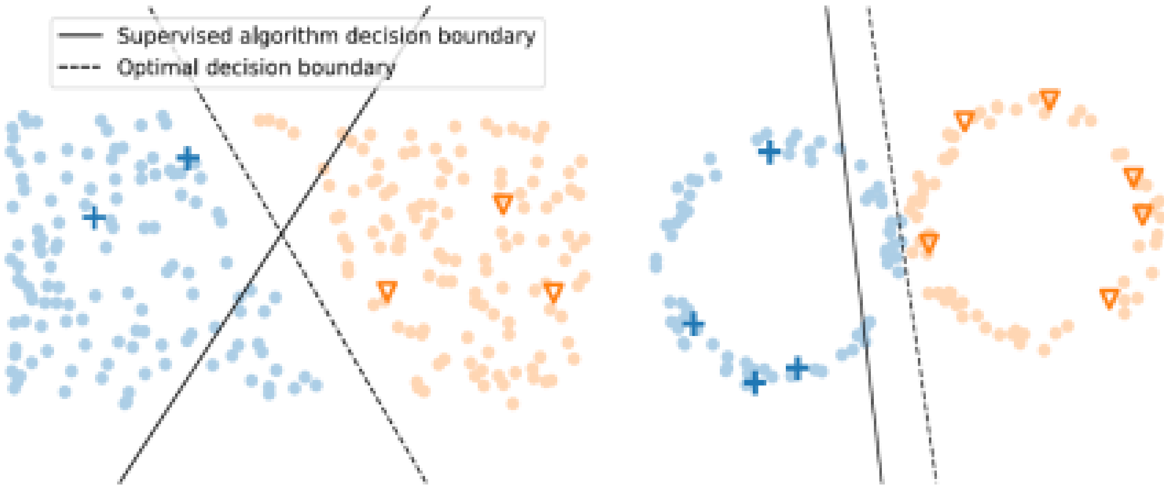
2.4.3 Semi-supervised learning assumptions

Semi-supervised learning is based on the use of the following three assumptions to overcome the challenge of using unlabeled data.

1. **Smoothness assumption:** if two data points x_1, x_2 close to each other in the input space then they have the same label y.

2. **Low density assumption:** The low-density assumption implies that the decision boundary of a classifier should preferably pass through low-density regions in the input space.

3. **Manifold assumption:** The manifold assumption states that the input space is composed of multiple lower-dimensional manifolds on which all data points lie. the data points lying on the same manifold should have the same label.



(a) Smoothness and low-density assumptions. **(b)** Manifold assumption.

Figure 2.7: Graphic representation of Semi-supervised learning assumptions, (a) represent that the decesion boundry pass through low density regions, and the data points closed to each other belongs to the same class. (b) shows that datapoints that lies in the same manifold belongs to the same [24]

2.4.4 Semi-supervised learning applications

Semi-supervised learning enables machine learning models to learn from labeled and unlabeled data. In addition, it regularizes a supervised model to be better generalized using unlabeled data with labeled data that was once available. SSL provides many applications in several areas such as:

- Email and Thread Summarization.

- Spam Filtering.
- Emails auto-answering.
- Biological Sequences(genetics and genomics).
- Computer vision.
- Natural language processing.

In our thesis, we chose the image classification of the computer vision field as an example to implement our semi-supervised learning model; which we demonstrated in the previous section.

2.5 Deep Neural Networks

Deep Learning (DL) [24] is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. DL contains computational models and algorithms that mimic the architecture of biological neural networks in the brain.

'Deep '[24] is a technical term and refers to the number of layers in an ANN. In our study we focus on two type of deep learning architectures: Convectional Auto Encoder(CAE) and Residual neural network(ResNet).

2.5.1 Residual neural network(ResNet)

Residual Network (ResNet) was suggested by (He, Kaiming, et al. 2015) at Microsoft Research. This new architecture is a residual learning framework to facilitate the training of deeper networks. Its architecture is inspired by VGG-19, but this network uses 34 layers with additional random access connections and this makes it a remainder network. In addition, residual networks are easier to optimize. [25]. The way this network works is that residual learning is used for every couple of stacked layers (called blocks). In this block it uses "Random Access Connections" (Figure2.6) skip a number of levels to skip a level is detrimental to performance [25].It achaived the ImageNet Large Scale Visual Recognition Contest for Image Classification [26] in 2015 and shows its performance in image classification compared to other CNNs.

Mostly to solve a complex problem, we are stacking some extra layers in deep neural networks, resulting in improved accuracy and performance. The training of very deep networks was made easier by the introduction of ResNet or residual networks and these Resnets consist of residual blocks. As presented in (Figure2.7)

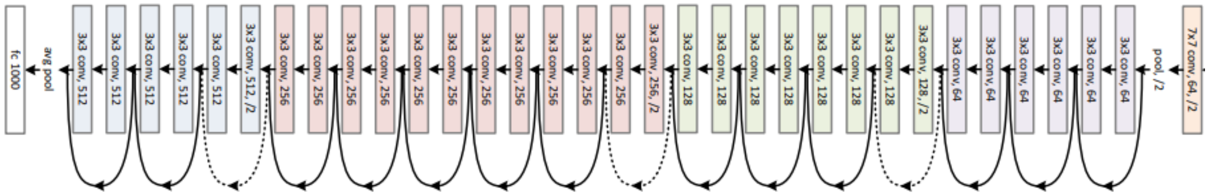


Figure 2.8: ResNet architecture[25].

:

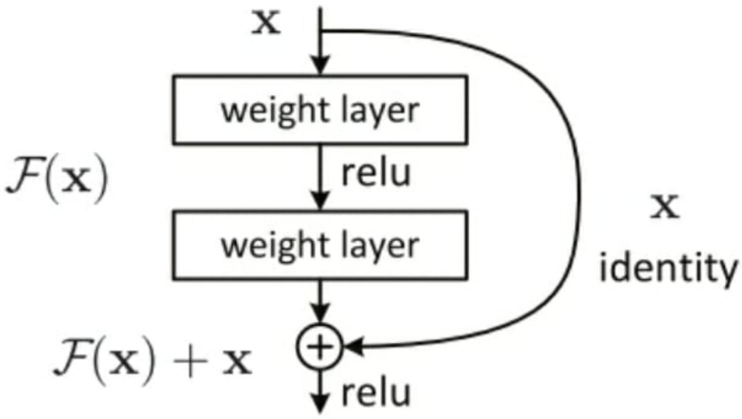


Figure 2.9: The architecture of a building block in ResNet[25].

2.5.2 Auto-Encoders

Auto encoders(AE) are multi-layers neural networks , and the key aspect of an auto encoder is that the input layers contain as much information as the output layer. The reason that an automatic encoder aims to replicate the input data and makes a copy of the data after analyzing and reconstructing it without supervision. Variations on the classic auto-encoder architecture exist. There is variations on the classic auto encoder architecture such as Convolutional autoncoder(CAE).

2.5.3 Convolutional AutoEncoder(CAE)

CAEs are a type of Convolutional Neural Networks (CNN): the main difference between the common interpretation of CNN and CAE is that the former are consistently trained to learn filters and combine features to classify their inputs. When, the CAEs are trained only to learn filters that are capable of extracting features that can be used to reconstruct the input.

CAEs are the next generation tools for unsupervised convolution filter learning. Once these filters are learned, they can be applied to any input in CAEs. Due to their convolutional nature, high-dimensional images with realistic size scale well because the number of parameters required to create an activation map is always the same regardless of the size of the input.

CAE are general-purpose feature extractors. CAE are the next generation tools for unsupervised convolution filter learning. Once these filters are learned, they can be applied to any input to extract res features. These functions can then be used to perform any task that requires a compact representation of the input, such as classification.[**?**]

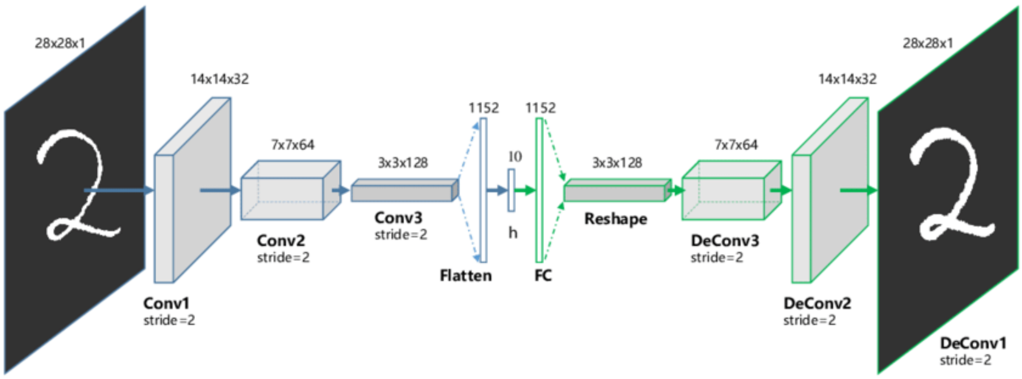


Figure 2.10: Convolution Auto Encoder architecture[27].

2.6 Conclusion

In this chapter, we have presented the context required for our work: the deep semi-supervised multi-label image classification. So we started by explaining what image classification is, its categories and focus on classifying multi-label images compared to other image classification families. Then, we distinguished what semi-supervised learning is versus the other learning techniques, its formal definition, its assumptions and its challenges. Finally, we presented the famous DNN architectures used for SSL and image classification with their specific and important terminological definition that we will use throughout our work.

Chapter3

Multi-label image classification methods: State of the art

Chapter 3

Multi-label image classification methods: State of the art

3.1 Introduction

To solve MLC problems, an explosive number of methods is presented in the literature. In the following section we take a look to the methods. Also, we will present from the state of art according to each method. They are considered as tow methods used to solve multi-label image classification problems. In the following sections we will explain briefly this two categories of these methods, which are shown in the (Figure3.1) :

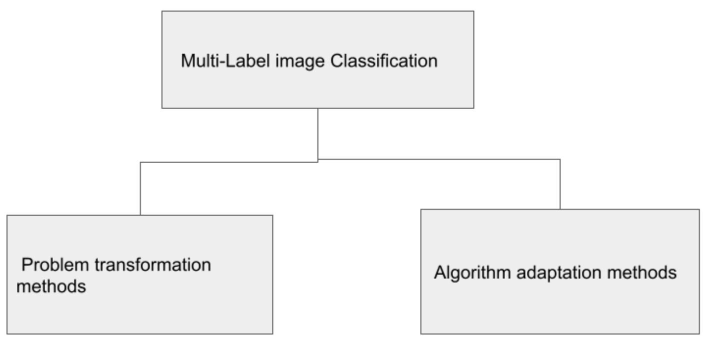


Figure 3.1: Multi-Label image classification methods[39].

3.2 Problem transformation methods

Problem transformation methods assign the MLC task or the multi-label problem into one or more single-label classification tasks or single-label problems. One way of doing this is by training a separate classifiers, one for each label. Then the results are transformed into multi-label predictions. Under this category it exists a huge number of methods that can be divided into three sub categories[37][38]:

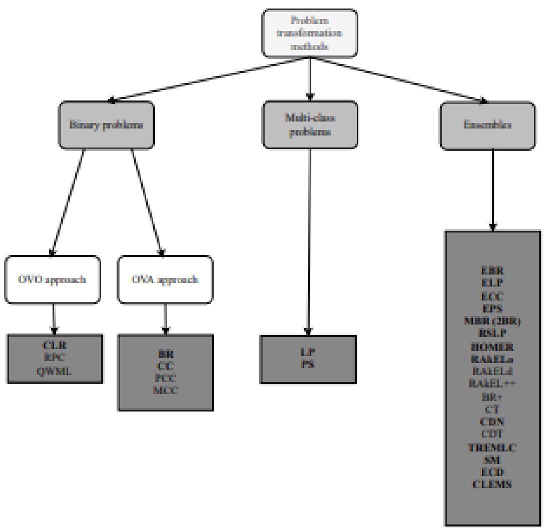


Figure 3.2: Problem transformation methods[39].

3.2.1 Binary relevance methods

Binary Relevance is a simple and efficient approach commonly used in real-world multi-label learning applications. It depends on dividing the multi-label classification problems into a binary classification problems, where it deal with each label as a separate binary classification.

3.2.2 Label Powerset methods

it consists to learn one classifier for each pair of classes, thus increasing the overall number of classifiers to train to k(k-1)/2 (from k labels).[37] The Pruned Sets (PSt) method illustrate this with a novel classification chain method that can model label correlations while maintaining acceptable computational complexity. An extensive empirical evaluation covers a wide range of multi-label data sets with different evaluation metrics.the chaining method versus related and cutting edge methods, both in terms of predictive performance and time complexity

3.2.3 Ensembles of problem transformation methods

This category of problem transformation methods includes all methods that use ensemble techniques such as stacking, bagging, random sub spacing, or use various transformations of the data sets, such as: Embedding. (Yeh, Mei-Chen, and Yi-Nan Li. 2019), extended the visual-semantic embedding model presented in[41] to solve multi-label image classification. they proposed a new visual recognition model. The model consisted of a CNN framework and word embedding model as shown in (Figure3.6). The model learns a mapping (a transformation matrix) from an image instead of a latent visual vector, and use the image transformation matrix A to map words from an embedding word space W into a new space W', where the relevant labels to an image are near to each other. [42]

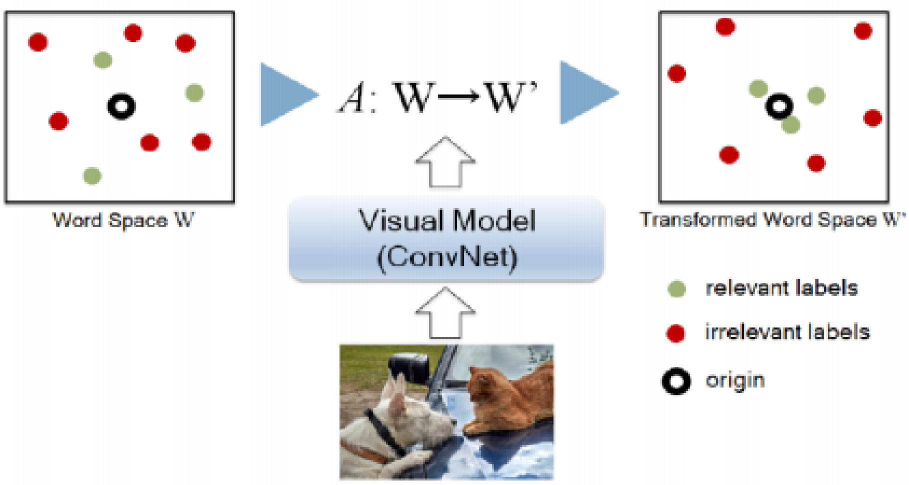


Figure 3.3: Visual-Semantic model for image MLC[41].

3.3 Algorithm Adaptation Methods

This category consists to adapt the existing algorithms of binary or multi-class classification to directly perform the multi-label classification, like :

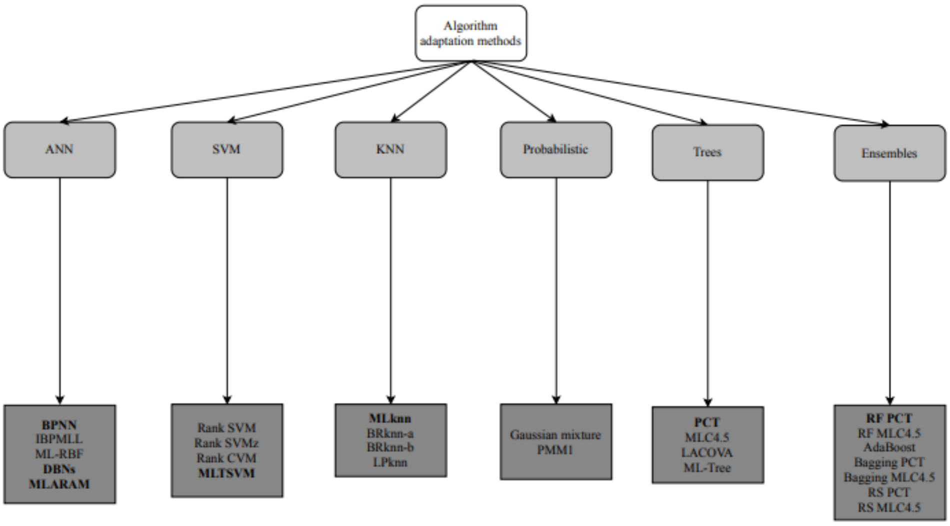


Figure 3.4: Algorithm Adaptation Methods[40].

- BP-MLL is an extension of the popular backpropagation algorithm for multi-label learning. It is a method based on neural networks. The most important change is the introduction of a new error function that takes multiple variables into account. [45], where Ei is the network error at (xi, Yi) and cij is the actual network output at xi on the jth label. Differentiation is the aggregation of the tag sets in these examples.

- It is a classification approach to multi-label learning based on SVM. It is used to minimize the loss of classification. The main function they use is the average proportion of misordered label pairs [45].

- Fangfang Luo and his team[32] applied Kernel extreme machine learning to MLC problem (ML-KELM) to network's real-value outputs to binary multi-label. Where identified the K nearest neighbors in the training set for each unseen instance. After that, based on statistical information gained from the label sets given before. the label set for the unseen instance determine by the maximum a posterior (MAP) principle.

- It is an adaptation of the well-known C4.5 algorithm to process data from multiple labels. The process is achieved by allowing multiple labels on the leaves of

Predictive Clustering Trees (PCT) [36], which are decision trees that view the data as a hierarchy of clusters. This method uses a standard TDITD algorithm for tree induction [44].

- It will improve the prediction of the existing PS process discussed earlier. This method employs a method of pruned sets in a sentence frame and uses a voting scheme to generate the prediction confidences. It provides a powerful and general framework. The EPS training algorithm can be used with any multi-label capable classifier [38].combines the PS method into an assembly diagram. PS is especially suitable for a set due to its fast construction times. In addition, it counteracts over fitting effects of the trimming process and allows new sets of labels to be created at the time of classification. the PS method increases the predictive performance of the algorithm [35], [39].

3.4 Conclusion

In this chapter we presented the methods with some recent literature works for multi label image classification . To well distinguish between them, we categorized them into two categories. This classification depend on the manner to solve the MLC problem: Problem transformation methods or Algorithm Adaptation Methods. Therefore, we mentioned a various recent solutions according to each categories. In our work we are interesting by exploring the category of methods that deeply and model the MLC in semi-supervised sittings. For that reason , we will propose a deep semi-supervised MLC(DSS-MLC).

Chapter 4

Deep Semi-Supervised Multi-Label image Classification

Chapter 4

Deep Semi-Supervised Multi-Label image Classification

4.1 Introduction

In this chapter we will present the DSS-MLC model implementation in details. Then, we'll define the characteristics of the devices applied in the experiment. Then the definition of the dataset that we have experimented with.Finally, we will present the training detail of the model.

4.2 Initialization

Our objective is to use the CAE and ResNet for multi-label image Classification for semi-supervised setting. We use a CAE and a ResNet to achieve the goal. Next, we describe in details the architecture and the training of the DSS-MLC model.

4.2.1 ResNet:

We implement ResNet 50 [40] in this work . In general, any visual model (AlexNet [41], VGGNet [42], or other deep neural networks) ; by virtue of its modular architecture, it belong to the model family called "Network in Network". This network is a set of micro-architecture or processing blocks. This set makes it possible to obtain the macro-architecture (that is to say the final network itself). To capture the variability of the dataset while avoiding exploding resource consumption, the Inception architecture relies on the use of the following processing blocks:

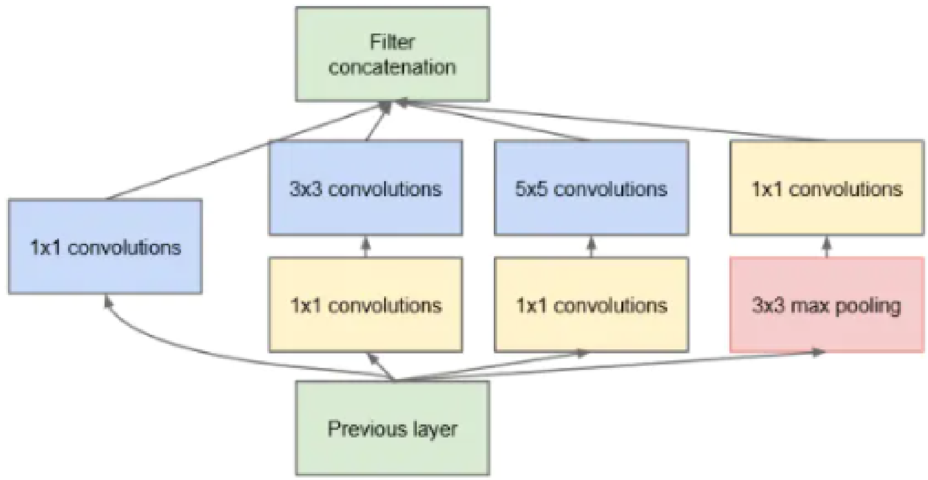


Figure 4.1: inception module with dimension reduction [47].

To reduce the consumption of resources involved in using multiple filters, a mathematical trick is used to reduce the size of the image analyzed by the 3 * 3 and 5 * 5 filters, by adding a 1 * 1 filtering step. before. In this way, the 1 * 1 size convolution performs a pooling operation on the values of a pixel in the image dimension space. For example if the input image has 3 channels (RGB), then after 1 * 1 convolution, the output image will have 1 channel. The reduction of dimension makes it possible to preserve the information, while reducing the cost of resources during the next convolution step with a filter of size 3 * 3.[47]

4.2.2 CAE

The extractor part of the DSS-MLC. We adopted a CAE[43] as an unsupervised, which created then pre-trained on the same dataset used in the training of the ResNet, with the loss function(MSELoss).we use the same optimizer and the learning rate value used with the CNN model.

```
-------------------------------------------------------------
        Layer (type)               Output Shape         Param #
=============================================================
            Conv2d-1         [-1, 64, 112, 112]           3,136
       BatchNorm2d-2         [-1, 64, 112, 112]             128
              ReLU-3         [-1, 64, 112, 112]               0
         MaxPool2d-4           [-1, 64, 56, 56]               0
            Conv2d-5           [-1, 64, 56, 56]           4,096
       BatchNorm2d-6           [-1, 64, 56, 56]             128
              ReLU-7           [-1, 64, 56, 56]               0
            Conv2d-8           [-1, 64, 56, 56]          36,864
       BatchNorm2d-9           [-1, 64, 56, 56]             128
             ReLU-10           [-1, 64, 56, 56]               0
           Conv2d-11          [-1, 256, 56, 56]          16,384
      BatchNorm2d-12          [-1, 256, 56, 56]             512
           Conv2d-13          [-1, 256, 56, 56]          16,384
      BatchNorm2d-14          [-1, 256, 56, 56]             512
             ReLU-15          [-1, 256, 56, 56]               0
            block-16          [-1, 256, 56, 56]               0
      ●     ●     ●
          Conv2d-153            [-1, 512, 7, 7]       1,048,576
     BatchNorm2d-154            [-1, 512, 7, 7]           1,024
            ReLU-155            [-1, 512, 7, 7]               0
          Conv2d-156            [-1, 512, 7, 7]       2,359,296
     BatchNorm2d-157            [-1, 512, 7, 7]           1,024
            ReLU-158            [-1, 512, 7, 7]               0
          Conv2d-159           [-1, 2048, 7, 7]       1,048,576
     BatchNorm2d-160           [-1, 2048, 7, 7]           4,096
            ReLU-161           [-1, 2048, 7, 7]               0
           block-162           [-1, 2048, 7, 7]               0
          Conv2d-163            [-1, 512, 7, 7]       1,048,576
     BatchNorm2d-164            [-1, 512, 7, 7]           1,024
            ReLU-165            [-1, 512, 7, 7]               0
          Conv2d-166            [-1, 512, 7, 7]       2,359,296
     BatchNorm2d-167            [-1, 512, 7, 7]           1,024
            ReLU-168            [-1, 512, 7, 7]               0
          Conv2d-169           [-1, 2048, 7, 7]       1,048,576
     BatchNorm2d-170           [-1, 2048, 7, 7]           4,096
            ReLU-171           [-1, 2048, 7, 7]               0
           block-172           [-1, 2048, 7, 7]               0
AdaptiveAvgPool2d-173           [-1, 2048, 1, 1]               0
          Linear-174                   [-1, 10]          20,490
=============================================================
Total params: 23,522,250
Trainable params: 23,522,250
Non-trainable params: 0
-------------------------------------------------------------
```

Figure 4.2: A proposed Residual neural network (ResNet50) architecture.

4.3 Extraction of Features

After the training of the CAE, we extract the features from the dataset using the CAE as map features. We extract the features from the encoder of the CAE. Then, we ignore the decoder part.

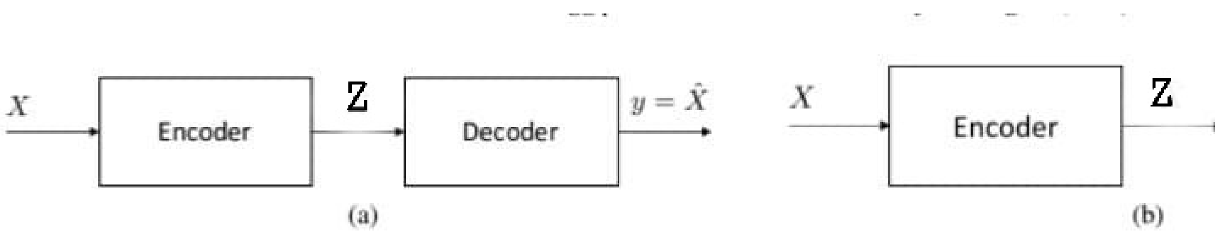


Figure 4.3: (a) training CAE, (b) extracting features without decoder [19].

4.4 The Training the DSS-model

As mentioned before, the DSS-MLC composed of an features extractor represented by the encoder part . The latter work map the features from the unlabeled data. And the classifier of the DSS-MLC is the prediction part of the pre-trained ResNet thus, present in the sub section before.

After extracting the features by the CAE and the discard the decoder part; we ignore the second convolutional layers of the encoder to mapping the features, without any operation of transformation happen to the features. Finally, the output of the encoder will be as a input to the ResNet classifier. Which is updated on multi label classification sitting.

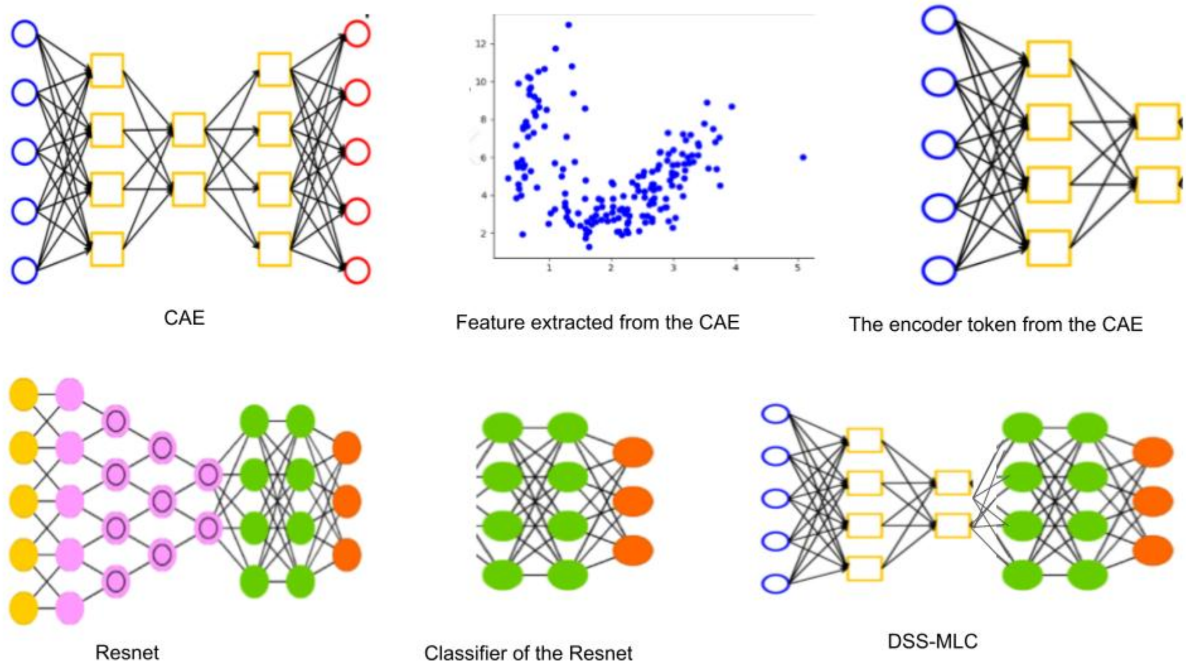


Figure 4.4: DSS-MLC model: briefly explanation.

4.4.1 Dataset " MNIST"

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems.[48] this datset has a training set of 60000 training examples, and 10000 test examples

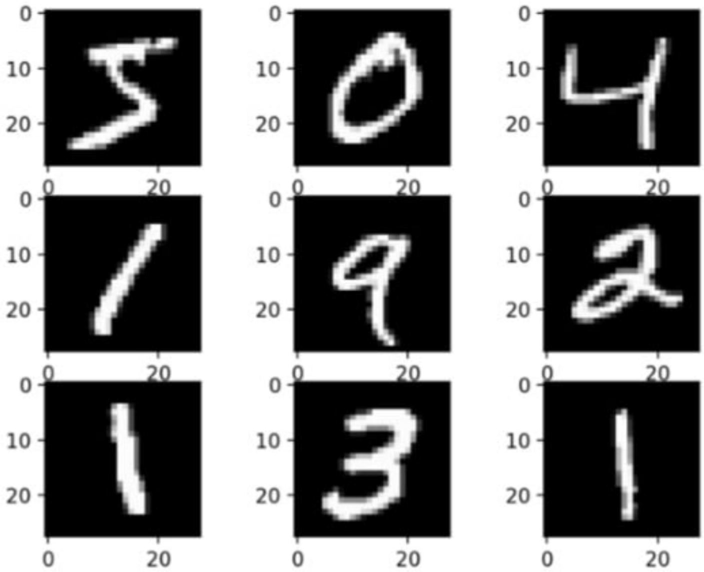


Figure 4.5: some pictures from MNIST data set[48].

4.5 Conclusion

In this chapter, we present our DSS-MLC model and the experiments on this model and result given by each part. Where we specify the settings for the experiment(Development tools, dataset, and Evaluation protocol). Finally, present the training of our model on the data set.

General conclusion

Through research on machine learning and deep learning, the image classification is one of machine learning tasks, that aims to categorize images into one or more predefined classes (labels). Multi-label image classification is considering as an important and challenging task. The main idea behind multi-label image classification is to classifier images, where assign for each image more than one class.

Multi-Label images is considered a difficult task due to the complexity of the images and the overlapping on the tags meaning; and the lack of labeled data . To simplify the challenge, a large number of literature papers consider complete independence between labels as a hypothesis within the given solutions. For that, other literature solutions that classified into two big categories. The first category solve the problem in indirect way by simplify the problem into sub single-label problem.The second category is a direct method by applying or modifying algorithms to solve the problem. The presented solution of multi-label image classification is a semi-supervised Multi-label image classification. The principal idea behind this solution is to develop a semi-supervised model using to different models training on different sitting.

This classifier consisted on two models, a CAE as features extractor and ResNet as a classifier of this features.

As difficulties, the studied field is wide and contain a huge number of different works. However, we tried to cover a number of some important works. The main difficult and problem we found it is in implementation, and exactly in customizing of machine learning functions to emulate it with our problem. As perspectives of our work:

- Firstly, we hope to complete the implementation of the presented solution to show the given results for multi-label image classification .
- Improve the presented solution to be robust against noisy labels (false labels in the training images).
- Applying the MLC solutions for other domain as the Medicine.

Bibliography

[1] YEH, Mei-Chen et LI, Yi-Nan. Multilabel deep visual-semantic embedding. IEEE transactions on pattern analysis and machine intelligence, 2019, vol. 42, no 6, p. 1530-1536.

[2] Clare A, King RD. Knowledge discovery in multi-label phenotype data. InEuropean conference on principles of data mining and knowledge discovery 2001 Sep 3 (pp. 42-53). Springer, Berlin, Heidelberg.

[3] i, T., Ogihara, M.: Detecting emotion in music. In: Proceedings of the International Symposium on Music Information Retrieval, Washington D.C., USA (2003) 239–240

[4] Boutell, M., Luo, J., Shen, X., Brown, C.: Learning multi-label scene classification. Pattern Recognition 37(9) (2004) 1757–1771.

[5] Tsoumakas, G., Katakis, I. and Vlahavas, I., 2006, September. A review of multi-label classification methods. In Proceedings of the 2nd ADBIS workshop on data mining and knowledge discovery (ADMKD 2006) (pp. 99-109).

[6] Bogatinovski J, Todorovski L, Džeroski S, Kocev D. Comprehensive Comparative Study of Multi-Label Classification Methods. arXiv preprint arXiv:2102.07113. 2021 Feb 14.

[7] Naresh Pal Singh, J.M. and Modi, H.N., 2017. Multi-label classification methods: a comparative study. International Research Journal of Engineering and Technology (IRJET), 4(12), pp.263-270.Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E.

[8] Zheng Yan, Weiwei Liu, Shiping Wen, and Yin Yang. Multi-label image classification by feature attention network. IEEE Access, 7:98005–98013, 2019.

[9] Francisco Herrera, Francisco Charte, Antonio J Rivera, and Mar ıa J Del Jesus. Mul- tilabel classification. In Multilabel Classification, pages 11–28. Springer, 2016.

[10] Zineb Kaoudja, Mohammed Lamine Kherfi, and Belal Khaldi. An efficient multiple-classifier system for arabic calligraphy style recognition. In 2019 International Conference on Net-working and Advanced Systems (ICNAS), pages 1–5. IEEE, 2019.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

[12] Francisco Herrera, Francisco Charte, Antonio J Rivera, and Mar ıa J Del Jesus. Mul- tilabel classification. In Multilabel Classification, pages 11–28. Springer, 2016.

[13] Meng Joo Er, Rajasekar Venkatesan, and Ning Wang. An online universal classifier for binary, multi-class and multi-label classification. In 2016 IEEE International

[14] Zineb Kaoudja, Mohammed Lamine Kherfi, and Belal Khaldi. An efficient multiple-classifier

system for arabic calligraphy style recognition. In 2019 International Conference on Net- working and Advanced Systems (ICNAS), pages 1–5. IEEE, 2019.

[15] Oussama Aiadi and Mohammed Lamine Kherfi. Image classification using texture features and support vector machine (svm). 2019.

[16] Jesper Engelen and Holger Hoos. A survey on semi-supervised learning. Machine Learning, 11 2019.

[17] Francisco Herrera, Francisco Charte, Antonio J Rivera, and Mar ıa J Del Jesus. Mul- tilabel classification. In Multilabel Classification, pages 11–28. Springer, 2016.

[18] Sandro Skansi. Introduction to Deep Learning: from logical calculus to artificial intelligence. Springer, 2018.

[19] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8):1819–1837, 2013.

[20] Kaushil Kundalia, Yash Patel, and Manan Shah. Multi-label movie genre detection from a movie poster using knowledge transfer learning. Augmented Human Research, 5(1):11, 2020.

[21] Haomin Chen, Shun Miao, Daguang Xu, Gregory D Hager, and Adam P Harrison. Deep hierarchical multi-label classification of chest x-ray images. In International Conference on Medical Imaging with Deep Learning, pages 109–120, 2019.

[22] LUO, Fangfang, GUO, Wenzhong, YU, Yuanlong, et al. A multi-label classification algorithm based on kernel extreme learning machine. Neurocomputing, 2017, vol. 260, p. 313-320.

[23] GOODFELLOW, Ian, BENGIO, Yoshua, COURVILLE, Aaron, et al. Deep learning. Cambridge : MIT press, 2016.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[25] Chao Zhang, Zichao Yang, Xiaodong He, and Li Deng. Multimodal intelligence: Rep- resentation learning, information fusion, and applications. IEEE Journal of Selected

Topics in Signal Processing, 2020.

[26] Sai Balaji, (2020, August 29). Binary Image classifier CNN using TensorFlow [blog post], https://medium.com/techiepedia/binary-image-classifier-cnn-using-tens orflow-a3f5d6746697. (Accessed on 14/5/2021).

[27] ZHANG, Yifei. A better autoencoder for image: Convolutional autoencoder. In : ICONIP17-DCEC. Available online: http://users. cecs. anu. edu. au/Tom. Gedeon/conf/ABCs2018/paper/ABCs2018-paper-58. pdf (accessed on 23 March 2017). 2018.

[28] Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Saso Dzeroski. An extensive experimental comparison of methods for multi-label learning. Pattern recognition, 45(9):3084–3104, 2012.

[29] Edward KY Yapp, Xiang Li, Wen Feng Lu, and Puay Siew Tan. Comparison of base classifiers for multi-label learning. Neuro computing, 2020.

[30] BOGATINOVSKI, Jasmin, TODOROVSKI, Ljupčo, DŽEROSKI, Sašo, et al. Comprehensive Comparative Study of Multi-Label Classification Methods. arXiv preprint arXiv:2102.07113, 2021.

[31] YEH, Mei-Chen et LI, Yi-Nan. Multilabel deep visual-semantic embedding. IEEE transactions on pattern analysis and machine intelligence, 2019, vol. 42, no 6, p. 1530-1536.

[32] F. Zhu, H. Li, W. Ouyang, N. Yu, and X. Wang, "Learning spatial regularization with image-level supervisions for multi-label image classification," Mar. 2017, arXiv:1702.05891. [Online]. Available: https://arxiv. org/abs/1702.05891.

[33] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data," in Lecture Notes in Computer Science 2168, L. De Raedt and A. Siebes, Eds. Berlin: Springer, 2001, pp. 42–53.

[34] A kernel method for multi-labelled classification," in Advances in Neural Information Processing Systems 14, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002, pp. 681–687.

[35] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Dzeroski, "An extensive experimental comparison of methods for multi-label learning," Pattern Recognition, vol. 45, pp. 3084 – 3104, 2012. [2] F. Herrera, A. J. Rivera, M. J. del Jesus, and F. Charte, Multilabel Classification. Springer Cham, Switzerland: Springer, 2016.

[36] H. Blockeel, L. D. Raedt, and J. Ramon, "Top-down induction of clustering trees," in Proceedings of the 15th International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1998, pp. 55–63.

[37] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie, "Dependency networks for inference, collaborative filtering, and data visualization," Journal of Machine Learning Research, vol. 1, pp. 49–75, 2001.

[38] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," International Journal of Data Warehousing and Mining, vol. 2007, pp. 1–13, 2007

[39] L. Grady and G. Funka-Lea, "Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials," in Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 230–245.

[40] Zhang Y. A better autoencoder for image: Convolutional autoencoder. InICONIP17-DCEC. Available online: http://users. cecs. anu. edu. au/Tom. Gedeon/conf/ABCs2018/paper/ABCs2018-paper-58. pdf (accessed on 23 March 2017) 2018 Mar.

[41] . Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556v6, 2015.

[43] Liu Liu, Daria Dzyabura, and Natalie Mizik. Visual listening in: Extracting brand image portrayed on social media. Marketing Science, 39(4):669–686, 2020.

[44] H. Blockeel, L. D. Raedt, and J. Ramon, "Top-down induction of clustering trees," in Proceedings of the 15th International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1998, pp. 55–63.

[45] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8):1819–1837, 2013.

[46] Oussama Aiadi, Belal Khaldi, Mohammed Lamine Kherfi, Yacine Ghorfa, and Rayhana Rezzag Bara. A supervised probabilistic model for visual object recognition.

[47] https://www.imaios.com/fr/Societe/blog/Classification-en-imagerie-medicale-les-architectures-CNN-les-plus-performantes-en-2021

[48] http://yann.lecun.com/exdb/mnist/