T.C
ANKARA YILDIRIM BEYAZIT UNIVERSITY
SOCIAL SCIENCES INSTITUTE

A COMPARATIVE STUDY OF CONVOLUTIONAL NEURAL NETWORK FEATURES FOR DETECTING BREAST CANCER

MASTER'S THESIS

Kemalcan Bora

MANAGEMENT INFORMATION SYSTEMS

ANKARA - 2019

T.C
ANKARA YILDIRIM BEYAZIT UNIVERSITY
SOCIAL SCIENCES INSTITUTE

A COMPARATIVE STUDY OF CONVOLUTIONAL NEURAL NETWORK FEATURES FOR DETECTING BREAST CANCER

MASTER'S THESIS

Kemalcan Bora

MANAGEMENT INFORMATION SYSTEMS

Dr. Keziban SEÇKİN CODAL

ANKARA - 2019

Approval of the thesis

The thesis ”A Comparative Study of Convolutional Neural Network Features for Detecting Breast Cancer” prepared by Kemalcan Bora was accepted as a Master’s thesis by unanimous/majority vote of jury members at Ankara Yildirim Beyazit University, Institute of Social Sciences, and Management Information Systems.

Title and Name	Institution	Signature
Dr.Öğr.Üyesi Keziban Seçkin Codal	AYBÜ YÖN. BİL. SİS	
Doç.Dr.İhsan Tolga Medeni	AYBÜ YÖN. BİL. SİS	
Dr.Öğr.Üyesi Melih Başkol	BARTIN ÜNİV. İŞLETME	

The date of thesis defense: 15.11.2019

I certify that this satisfies all the requirements as a thesis for the degree of Master of Management Information Systems, Institutes of Social Science, Ankara Yildirim Beyazit University

Manager of Institute Title Name

Statement

This thesis is my own study, I have no unethical behavior that violates patents and copyrights in all stages from the planning of the thesis to the writing of the thesis. I declare that I have obtained within academic and ethical rules, and that I have been referring to all the information and comments used in this thesis.

Kemalcan BORA

Dedication

I dedicate this thesis to my family and friends. Thank you for all your supports,encouragement and patience.

Acknowledgements

I'd like to express my gratitude to those who supported me on this long journey of four years. I would like to thank my advisor Dr. Keziban Seçkin Codal, who has repeatedly read and guided my thesis and I would especially like to thank Dr. H. Kemal İlter, who inspired me not only academically but also helped me understand life. I am grateful to my family and M. Seçkin Bedük who motivated me not to drop out school. I would like to thank N. Cansu Candan, the most beautiful of all coincidences, for motivating me at every opportunity and I am finishing my words with a Metallica lyric;

"He tries to please them all this bitter man. He is throughout his life the same. He's battled constantly this fight he cannot win a tired man. They see no longer cares the old man then prepares to die regretfully. That old man here is me."

Özet

Meme Kanseri Tespitinde Konvolüsyonlu Sinir Ağı Özelliklerinin Karşılaştırmalı Çalışması

Göğüs kanseri, en ölümcül kanser türüdür fakat erken tanı tedavisi ile kadınların hayatta kalma oranının çok yüksektir. Yardımcı sistem olarak bilgisayar destekli bir teşhis, radyologların bilgi birikimi olmadan anormallikleri otomatik olarak algılayabilir. Bu tezin amacı, görüntü temelli objelerin sınıflandırılmasında büyük başarılarlar elde edilecebilecek bir derin öğrenme metodolojisi uygulayarak göğüs kanserinin saptanmasıdır.
322 adet mamograf görüntüsü derin öğrenme algoritmaları ile tümör sınıflandırmak için kullanıldı. Konvolüsyonel sinir ağları (CNN) göğüs kanseri saptanmasında çeşitli yöntemler gerçekleştirir 48,2%, VGG16 ile 72.2%, ResNet50 65.3%,, NasNet 65.3%, ve InceptionResNetV2 60%, oranında başarı sağlanıldı. Göğüs kanseri sistemleri için umut verici bir teknik olarak iyi tasarlanmış bir konvolüsyonal nöral ağ, mevcut sistemlerde kesinlikle daha etkili bir rol oynayacaktır.

Anahtar kelimeler: göğüs kanseri, bilgisayar destekli sistemler, derin öğrenme, görüntü işleme, konvolüsyonel sinir ağları.

Abstract

A Comparative Study of Convolutional Neural Network Features for Detecting Breast Cancer

Breast cancer is the most deadly type of cancer among women survival rate of which can be increased with early diagnosis treatment. A computer-aided diagnostic as an auxiliary system in field can automatically detect abnormalities without wisdom of radiologists. The aim of this thesis is directly the detection of breast cancer by applying a deep learning due to their great achievements in the classification of picture-based objects.

The data that is included 322 mammography images is used to classify the tumor by applying a new breed of deep learning algorithms- convolutional neural networks (CNNs). Compared with the CNNs in detection of breast cancer, the proposed methods out performs AlexNet by 48.2% , VGG16 by 72.2% ,ResNet50 by 65.3% , NasNet by 65.3% and InceptionResNetV2 by 60% in terms of accuracy rate. A well-designed convolutional neural network, as a promising technique for breast cancer systems, will certainly play more efficient role in existing systems.

Keywords: breast cancer, computer aided systems, deep learning, image processing, Convolutional neural networks.

Contents

List of Figures

1 Introduction

Cancer by word means: it is called malignant cells that occur when cells in an organ or tissue are divided unevenly and multiplied. In general, cancer is a group of diseases caused by uncontrolled proliferation of cells in various parts of our body.(Forrest et al., 1996). Breast cancer is the second most common type of cancer in the world after lung cancer. Breast cancer is the most common type of cancer for women in many developed and developing countries(Parkin, Bray, Ferlay & Pisani, 2005). Breast cancer is ranked first among women with a rate of 41. 6/100 000 and the age of occurrence is gradually declining(Pirhan & Sucu, n.d.). The American Cancer Association recommends that women aged 45 years and older receive a mammogram every year or two years.(Oeffinger et al., 2015) Many deaths can be avoided thanks to early diagnostic treatment methods. In mammograms, two types of diagnostic lesions are considered, one of which is clustered microcalcifications and the other is a solid mass in the breast. We can also describe clustered microcalcifications as small calcium deposits that can be seen around cancerous tissue.

Although radiologists can highly accurately recognize these lesions, they can use a computer-aided system as a second opinion in fields where they suspect and stay volatile. This is the topic of studies and the product that might emerge from studies has inspired us to develop an image processing scheme. This thesis is component of a project that aims to use deep learning for segmentation of lesions.

Traditional image processing systems use various algorithms to sequentially process images for example a standard image processing algorithm preprocessing the image, specifies the terest zones, extracts properties from the corresponding parts, and removes the extracted algorithms are trained with property classification algorithms. Although this pattern has created many successful systems, it offers two drawbacks:

1. It is very often necessary to have specialist understanding. The system generated is hard to operate properly when referring to more than one specialist understanding.
2. Dependencies and relationships between them are often uncertain: changes in one element influence others ’ performance, each component requires to operate well in order for the system to function well, and the overall component needs to be enhanced in order to enhance efficiency.

In this thesis, the plan is to use evolutionary networks rather than traditional image processing phases. Evolutionary algorithms can be divided into those that develop both the weights and the architecture (e.g. NAS) and those that only actually try to optimize the weights of the deconvolutional neural network. The combination of evolutionary algorithms and reinforcement learning usually emerges as a single-weight application. Convolutional networks(Fukushima, 1980) (LeCun, Bottou, Bengio, Haffner et al., 1998) are a concept that came up with the modification of conventional neural networks. Such networks are a kind of deep neural networks or deep learning due to their large and deep structure convolutional networks are very popular nowadays because of their great achievements in the classification of picture-based objects. If you want to classify or recognize a picture with traditional neural networks, all pixels must be transferred to the neural network. In convolutional networks, some patterns are first tried to be detected on the picture, and these patterns are transferred to the neural network. This way we can achieve more successful results while the picture can be processed in less complex ways. Despite some drawbacks, it is the most ideal technology for object recognition.(Russakovsky et al., 2015).

Researchers used small convolutional networks to separate breast masses from normal tissue(Sahiner et al., 1996) and individual noise in the image obtained microcalcifications(S.-C. B. Lo et al., 1995)(S.-C. B. Lo et al., 1995). A larger network with newer features such as rectified linear unit activations, pooling, momentum, data magnification and release was trained to identify malignant masses(Arevalo, González, Ramos-Pollán, Oliveira & Lopez, 2016). For these experiments, mammograms were developed, potential lesions were detected, and related areas were presented to the network for classification purposes. Recently, Dubrovrina et al.(Dubrovina, Kisilev, Ginsburg, Hashoul & Kimmel, 2018), segmented studies of different breast tissue using a modern convolutional network have been conducted.

Chapter 1 explains the introduction of this thesis. The detection of breast cancer using deep convolutional neural networks is explained in chepter 2 , analysis is explained in chepter 3. The results section was announced in section 4. Chapters 5 are also conclusion and future work explained.

1.1 Motivation and Background

The most frequently diagnosed cancer in females is breast cancer, and mortality rates are greater than in any form of cancer. It is estimated that at some stage in their life 1 out of 8 females in the United States will be diagnosed with breast cancer. The key to reducing the amount of death is early diagnostic treatment and increases the survival rate.

With the current technology, a high-quality mammogram ”is the most effective way to detect early breast cancer”(Chen, Papandreou, Kokkinos, Murphy & Yuille, 2014). Mammograms are used by radiologists to search for early signs of cancer, such as mass or microcalcification. Approximately 85 percent of breast cancers can be detected by scanning a mammogram; this high rate of success is revealed thanks to careful examination of experienced radiologists. A computer-aided diagnostic tool (CAD) can automatically detect these abnormalities by avoiding any human error by saving the time radiologists need. Computer-based approaches can also be used by radiologists as a second informed opinion on the diagnosis as an auxiliary during the screening process.

The lesion segmentation is appointed with the separation of lesions from normal tissue in medical images.When looking for abnormalities, radiologists perform lesion segmentation. Traditional techniques of image processing, lesion segmentation is often hard to adapt to the methods of computer vision.

In this thesis, a newly developed and popular topic in machine learning, convolutional networks were used. The system pipeline can be simplified using a single model and instead of designing new image features or developing specific subsystems, the model and algorithm learning mechanism focused on development can be used.

1.2 Objective and Scope

The main objective of this research is to segment lesions of breast cancer in digital mammograms and to assess how efficiency is affected by choices of distinct machine learning architecture. Developing software tools to process the database and training new deep learning models to analyze the performance of convolutional networks reported on the literature. Designing and training a range of modern, fine-tuned convolutional network and testing the viability of convolutional networks for the detection and diagnosis of breast cancer. Using alternative evolutionary network models to improve improve the diagnostic success rate and propose new ideas for future research on the subject.

1.3 Methodology

The significant studies have been conducted in the world on the detection and diagnosis of breast cancer. Convolutional networks are widely used for object recognition tasks and showed very good results(Cawley & Talbot, 2010) This study shows that applying convolutional networks to mammographic images will produce results similar to those achieved with less difficulty, despite traditional computer vision techniques.

Here are some of the questions to be answered in this study:

1. Are convolutional networks strong enough to perform breast cancer lesion segmentation? Is learning with sparse data an obstacle?
2. What is the advantage of GPU for convolutional neural network architecture?
3. What is the best convolutional neural network structure?

1.3.1 Database

Digital Database Scan Mammography (DDSM) is the most popular and advanced open source database utilized for CAD development. Approximately 10.5K digitized film mammogram from 2620 patients is composed of the image. Mammograms are 12-bit or 16-bit images with a spatial resolution of 0.05 mm.

Another source, the MIAS database(Ibrahim, Fujita, Hara & Endo, 1997) consists of 322 digitized film mammogram image. Mammograms has been reduced to 200 micron pixel edge and clipped/padded so that every image is 1024 × 1024 pixels.

Finally, the Digital Storage for Chest Cancer (BCDR-DM)(Moura & Guevara López, 2013) consists of 1,2K digital mammogram image of 237 patients. Mammograms are 8-bit images with a spatial resolution of 0.07 mm.

In this study, the MIAS data set was preferred. Because easy to access database.

1.3.2 Method

In this study, convolutional neural network was used. In digital mammograms, automatic segmentation of breast cancer masses is performed. Computed diagnosis helps radiologists develop symptoms of cancer. Improving the diagnosis of breast cancer because of its high incidence will benefit millions of women around the world. The goal is to apply evolutionary neural networks, a modern machine learning model to achieve results comparable to traditional methods. Convolutional neural network, which can take an introductory image, is a Deep Learning algorithm that can attach importance to various aspects and objects in the image and distinguish. The reason for using the convolutional neural network in this study is more successful in image processing than conventional algorithms.

2 Breast Cancer Detection Using Deep Convolutional Neural Networks

2.1 Breast Cancer Screaning and Diagosis

Cancer is called bad tumor that appear when cells in an organ or tissue divide and multiply irregularly. In general, cancer is more than 100 diseases that are formed by uncontrolled proliferation of cells in various parts of our body. Cancer cells accumulate, forming tumors. Tumors can be benign or malignant. Benign tumors are not cancer, they are often taken and often do not repeat. Cells in benign tumors do not spread to other parts of the body, most importantly benign tumors rarely threaten life. Malignant tumors are cancer cells in malignant tumors are abnormal and divide uncontrolled and unevenly. These tumors can squeeze, infiltrate or destroy normal tissues. If cancer cells are separated from the tumor from which they are formed, they can go to other parts of the body through blood or lymph circulation and the phenomenon of the spread of cancer to other parts of the body in this way is called metastases.(INSTITUTE, 2012)

Breast cancer is cancer cells that form inside the milk ducts in the breast tissue 80/100 of breast cancers are invasive ductal carcinoma. (fig 2) Invasive ductal carcinoma designates that breast cancer occurs in the milk ducts 20/100 of breast cancer is withal invasive lobular carcinoma. In this species, breast cancer develops not in the milk ducts, but in the milk glands. The multiplication and magnification of cells that cause breast cancer takes quite a while. But after their multiplication, the cells can permeate lymph and blood to other organs of the body. The most consequential thing in breast cancer is to diagnose cancer without spreading blood

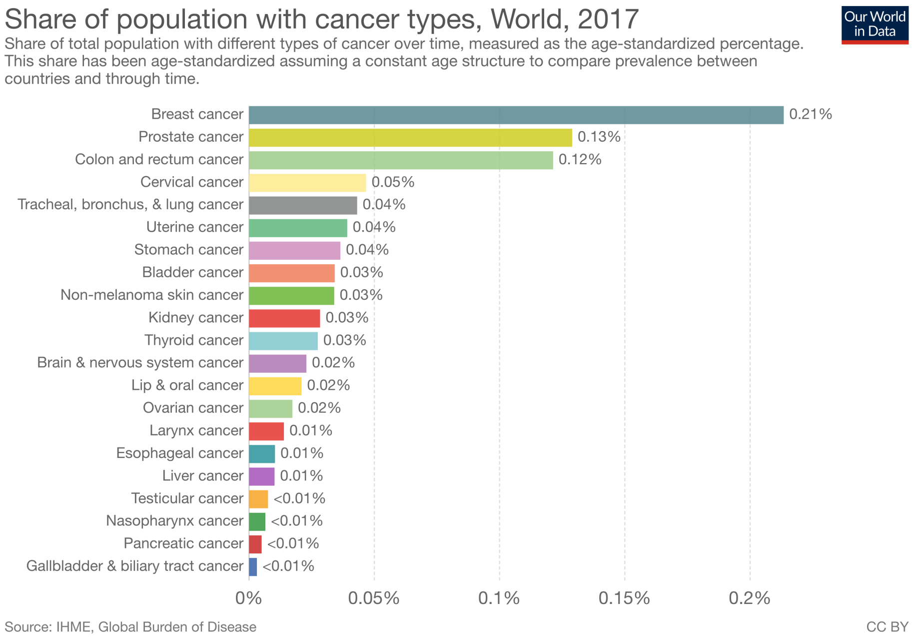


Figure 1: Number of people with cancer, 2017.

and lymph to other organs. With a diagnosis established at this stage, the rate of treatment is very high. That is why early diagnosis of breast cancer is very consequential.

Breast cancer is the most prevalent type of cancer in women. Breast cancer in one out of 10 women is found in 20 out of every 100 000 women on average. Breast cancer is very infrequent in men compared to women. But when the disease develops, its course is more expeditious and worse than breast cancer in women. One of every 100 breast cancer is observed in men. Albeit the cause of breast cancer is not plenarily kenned, many factors can be mentioned: heredity, alimentation, social-economic situation, menstrual status, births, birth control pills.(Ries et al., 2007)

Breast cancer is a slow progressive type of cancer. The tumor reaches a size of 1 cm in 5-7 years, and then the lymph nodes are distant from the liver and bone through the blood. can spread to organs. Staging is performed to ascertain what stage the tumor is at and where it spreads, and the treatment is decided accordingly. In breast cancer, a system called TNM is utilized for staging. Accordingly, T denotes the tumor diameter, N designates the number of diseased armpits lymph nodes, and M betokens the state of distant propagation (metastasis).

4 stages of breast cancer can be mentioned. Stage I, II and some stage III tumors are considered early stage breast cancer. A component of stage III tumors and stage IV tumors are called advanced stage in breast cancer. When staging breast cancer, the size of the tumor is taken into account whether it spreads to the circumventing lymph nodes. Accordingly, we can identify the stages of breast cancer as follows:

- Stage I: The tumor is less than 2 cm and has not yet spread to the lymph nodes.

- Stage II: The tumor is 2-5 cm in size and may not have splashed or splashed to surrounding lymph nodes.

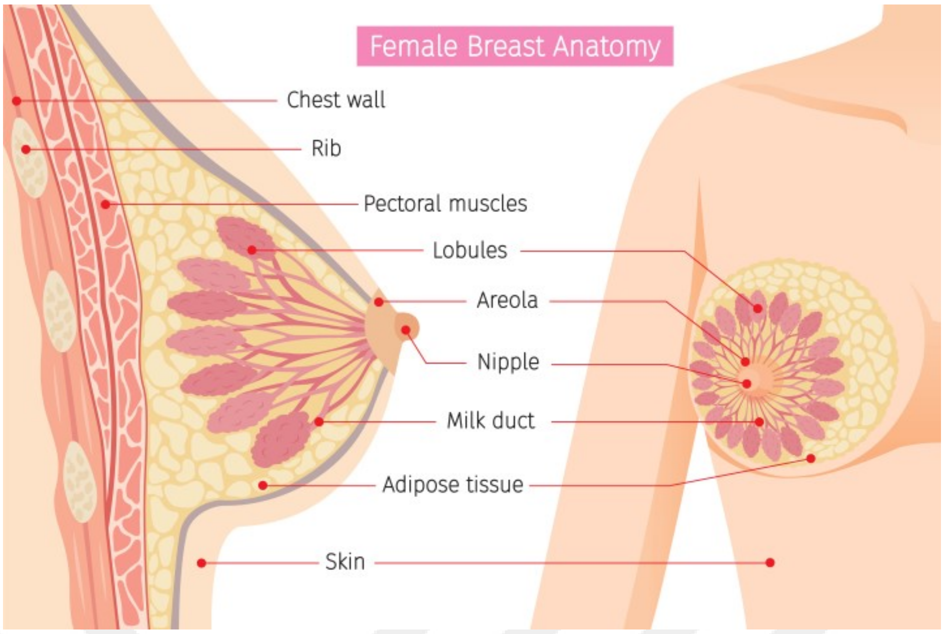


Figure 2: Anatomy of the female breast.

- Stage III: means more spread to the surrounding lymph nodes
- Stage IV: Means metastasized to other organs (bone, liver, brain, lung) or bone, distant lymph nodes.

2.2 Mamograms

Mammography is a cancer screening application widely utilized all over the world for the purport of early diagnosis of breast cancer. Due to mammography and the increase in consciousness of women about this issue, most breast cancer is diagnosed at an early stage, which is the perfect solution for a cancer for all stages of breast cancer point to the process. However, even if the result of mammography screening turns out negative or benign (benign), women still have some risk of breast cancer.

The results of a remarkable study on the risk of breast cancer after negative mammography were published in JAMA Oncology in August 2018.(Giannakeas, Sopik & Narod, 2018)

The study included 161 thousand women over 40 years of age who had cancer screening mammography between 2011 and 2014 and had not previously been diagnosed with breast cancer. In 272, 881 of 306, 028 women involved in the analysis, the mammography result was negative. In the 1-year follow-up of 272 thousand women, which were negative for mammography, 160 people developed breast cancer (90 with a good prognosis and 70 with a lamentable prognosis). So the jeopardy of developing breast cancer within 1 year after a negative mammography was 0.05 (5 per 10 thousand).(Giannakeas et al., 2018)

Mass lesions are the most ordinary form of breast cancers. The mass is defined as a lesion dissevered from the adjacent breast parenchyma by a boundary and occupying space in two

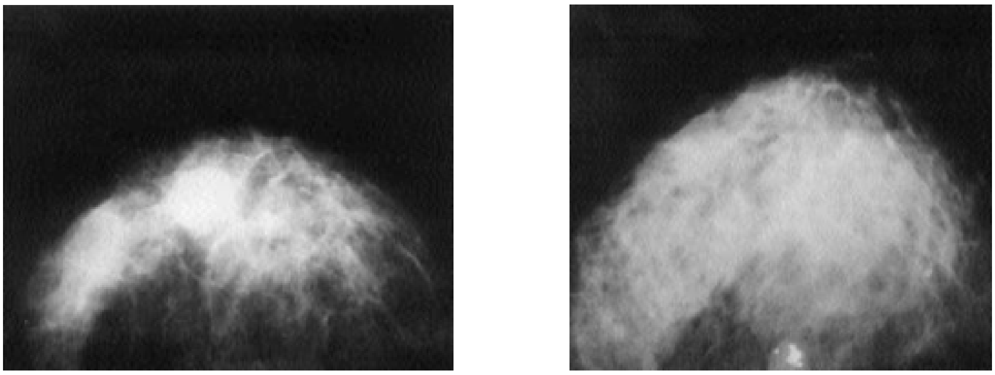


Figure 3: **Left:** Fibroglands in the breast follow uniform, homogeneous density in the tissue (CT graph). US examination revealed that the lesion identified was compatible with the cyst. **Right:** Oval density (fibroadenoma), with typical rough calcifications located near the anterior wall of the breast in the breast. Benign intramamarian lymph nodes can be traced in 5/100 of all mammographies. Usually it is located in the upper outer quadrant of the breast, and in the central part, characteristic radiolusan appearance or peripheral notch can be detected.

different projections. If the lesion is monitored in a single projection, it should be defined as density. The morphology of the mass is paramount in the diagnostic approach. In assessing the suspicion of malignancy of the mass, shape, edge structure and density characteristics are of the first consequentiality. Masses can be relegated as round, oval, lobular, aberrant in shape. Lobulation is designated by corrugation in the edge structure. Microlobular or irregure-shaped masses are more liable to malignancy. If the edge between the circumventing tissue of the mass is sharp, it is well defined as constrained. The definition of the spicular is utilized if the radial style extensions from the edge of the mass to the perimeter are traced. In the event that the mass overlaps with the circumventing tissue, the covered edge structure appears. The halo mark, which is betokened as a radiolusen area of 1 mm thick circumventing the capsule and well-inhibited masses, is customarily detected in benign lesions. X-ray permeability of the mass is denoted in density. Radiocene lesions are virtually never malignant and are considered benign without further evaluation. Radioopaque lesions of low densities are fibroadenoma and cyst. The possibility of malignancy in radiopaque lesions of high density should be considered. If morphology and size changes are observed in well-confined nodular lesions detected in antecedent mammographies and monitored, further examination is indispensable because the jeopardy of malignancy is high. The number of lesions is not a good criterion in terms of differential diagnosis, but the presence of an immensely colossal number of lesions with morphological features and bilateral reduces the jeopardy of malignancy.(SOCIETY, 2014)

Digital Mammography, an incipient technology in breast imaging, has now become the main diagnostic method for the diagnosis of breast diseases and breast cancer screening. Unlike classical mammography, images are obtained in digital media utilizing electronic sensors. These images are then evaluated on the CAD programmed workstation with special high-resolution monitors. This workstation has many operations such as magnification, quantification and contrast adjustment without depending on the dose of the X-ray utilized. Although researchers still discuss whether mammograms offer an advantage over their species(Kerlikowske et al., 2011)(Pisano et al., 2008) digital mammography has become standard for breast cancer screening. Figure 3 is actually a digital mammogram. Computer aided detection (CAD) systems help radiologists in decision-making process by identifying suspicious areas.

2.3 Computer Aided Detection

Today, almost all of the methods of testing are utilized by advanced computer systems. When doctors diagnose the patient, the doctor asks for the patient's ECG when complaining about the patient's problem, for example, crunch in the chest, numbness in the arm. That's where advanced computer systems come in and give the doctor all the data he wants about the physiology of the patient. Computer-aided diagnosis (CAD) (Chan et al., 2005) plays a role in the detection of lesions in medical image processing. CAD technology is often called technology based on medical imaging. Computers use CAD for labeling anomalies based on image processing. In other words, computer-aided diagnosis, the main purpose of expanding computer-aided tests is that it is an auxiliary system, also known to the doctor as the "third eye". In recent years, with the rapid development of computer technology, CAD technology has achieved a rapid development, especially in the field of medical imaging. In practice, CAD has proved to play a big role in improving diagnostic accuracy. The use of computer-aided diagnostics in medicine is based on the 1950s. In 1959, American scientists Ledley and others introduced mathematical models for the first time in clinical medicine for computer-aided diagnosis and diagnosed a group of cases of lung cancer. In 1966, Ledley first introduced the concept of computer-aided diagnostics (CAD). In the early 1980s, computer-aided diagnostic systems were further developed with the most important specialist systems in the field of TCM. The computer-aided diagnostic process includes the collection of general patient and examination data, quantitative processing of medical information, statistical analysis and final diagnosis. The most popular models at that time were Bayes' theory, the maximum probability method model and the sequential model. The artificial neural network (ANN) has developed rapidly since the 1990s as a mathematical method that mimics the working principle of neurons in the human brain. It can play an auxiliary role in diagnosis, as it has the ability to self-learn, memory ability and predictive event development. In the classification and diagnosis, the artificial neural network method is made from traditional methods (probability statistics, mathematical models, etc.). It has better performance. It can be said that the Artificial Neural Network represents one of the most advanced AI technologies today.

2.3.1 Image Enhancement

Basically Image processing is the name given to analyzing a picture or video and detecting its characteristics. The element of determining the characteristic feature, of course, is based on acquiring a distinctive nature. All of the face recognition systems used today are realized with image processing. Although the concept of Image Processing is a digital process, it can also be realized with optical and analog processing. Image enhancement is the process of digitally changing an image. Tools for image enhancement are usually used as filters, image editors, and other tools to change the entire image or various properties of a part of an image, such as is the processing of many different types.

2.3.2 Image Segmentation

Segmentation is usually the first stage of image analysis. Image segmentation can be described as separating an image into meaningful zones where different features are kept within each. For example, there may be similar brightness within the image, and these brightness can represent objects in different parts of the corresponding image. Another example of these segments (segment-elements), which may vary depending on the application, is a segmentation in the air-ground photo to distinguish the vehicles moving on the road and the environment from the road. can be done. (A segmentation that can separate the path from the environment). It should be remembered that there is no general (universal) parting method that can be applied to all images, and no segmentation method is perfect. In other words, the methods designed for image segmentation, as well as with image enhancement and repair problems, and the achievements of these methods are based on image to image and application changes and are of two types:

1. Similarity Based Image segmentation algorithms

    * Image segmentation based on similarities in gray level values is known as region segmentation.
    * Threshold (thresholding), growth (growing), split-and-merge) operations are carried out based on split-and-merge operations.
    * The concept of dividing an image based on similarities or differences in the gray level values of pixels can be applied to both static and dynamic (time-changing) images [R].

2. Discontinuity based segmentation algorithms

    It is based on detecting discontinuities such as isolated points, thin lines, or picture edges (gray level values change suddenly) using masks similar to those in low and high filtering.

2.3.3 Feature Extraction

Feature extraction is a dimension reduction process. The size of a complex data is reduced to a simpler problem. Correctly made feature inference is one of the most important factors that affect the performance of the designed system design. Feature extraction takes an important place in reducing the number of resources required for processing without losing important or relevant information. Feature extraction can also reduce the amount of data for the analysis to be done, creating variable combinations that facilitate learning and generalization steps in the machine learning process.

2.3.4 Feature Classification

The most commonly used filter metric for classification problems is correlation and mutual information. Although mathematically there are no distance measurements, it does not calculate actual distance because they fail to obey, but rather scores are accepted. These points are calculated between the property set and the desired output category.

2.3.5 Evaluation

Machine learning is an indispensable part of the evaluation. If saddled using accuracy rate in your model metric, it can produce satisfactory results, but it is bad when evaluated against other metrics, such as logaritmic loss or other metric can give results.

2.4 Deep Learning for Breast Cancer Diagnosis

The aim of this study was to examine mammaography images using the CNN algorithm to obtain the best results by testing the various hyperparameters among the most deadly cancer species in the world. Cancer uses more specific evolutionary neural networks, especially since various artificial neural network algorithms have great success in image processing technologies is intended.

2.4.1 Classification Approch

The most common problem in machine learning is classification, so-called classification problem, for example, machine learning algorithms were used in this study to classify patient outcomes as illness and health.Classification problem is the basis of machine learning, many other applications can develop the classification problem, but also many problems become a problem of classification such as image segmentation can be converted.

In machine learning, the algorithm that can complete the classification task is usually called a classifier. To evaluate the good or bad of a classifier, we have to have an evaluation indicator. The most common is accuracy, which refers to the number of correct data classified by the classifier as a percentage of all data.

2.4.2 Artificial Neural Networks and Deep Learning

The concept of deep learning stems from the research of artificial neural networks. Multi-layer perceptors with multiple implicit layers are a deep learning structure. Deep learning discovers distributed feature representations of data by amalgamating low-level features to compose a more abstract high-level representation of attribute categories or features.(Ji, Xu, Yang & Yu, 2012) The concept of deep learning was proposed by Hinton et al. in 2006. Predicated on Deep Confidence Network (DBN), an unsupervised acquisitive layer-by-layer training algorithm is proposed. The deep structure of multi-layer automatic encoder is proposed. In integration, the convolutional neural network proposed by Lecun et al is the first authentically multi-layer learning algorithm that utilizes spatial relativity to reduce the number of parameters to amend training performance.

Observations, such as an image, can be represented in several ways, such as vectors of each pixel vigor value, or more abstractly as a series of edges, areas of categorical shapes, etc. It is more facile to learn tasks from instances (for example, face apperception or face expression apperception) utilizing certain representations. The benefit of deep learning is to supersede manual acquisition of features with unsupervised or semi-supervised feature learning and highly efficient layered feature extraction algorithms.

Deep learning is a new field of machine learning research, driven by the creation and simulation of neural networks that simulate the human brain's mechanisms to interpret data such as images, sounds, and text.(Xu, 2015) As with machine learning, deep machine learning methods have a distinction between supervised learning and unsupervised learning. The cognition models established under different learning frameworks are very different. For example, Convolutional neural networks are a machine learning model under deep supervised learning, and deep confidence network (Deep Credence Nets, DBNs, is a machine learning model with unsupervised learning.

Machine Learning is a discipline that specializes in how computers can simulate or implement human learning demeanors in order to acquire incipient erudition or skills and reorganize subsisting erudition structures so that they can perpetually amend their performance. Can machines be able to learn like humans? In 1959, Samuel in the United States designed a game program that has the ability to learn to improve his game in constant game. After 4 years, this program triumphed the designer himself. After another three years, the program defeated an unbeaten champion in the United States that remained eight years old. This program shows people the ability to learn machine and raises many thought-provoking social and philosophical questions.(Längkvist, Karlsson & Loutfi, 2014)

The calculation involved in generating an output from an input can be represented by a flow graph: a flow chart is a graph that represents a computation, in which each node represents a basic calculation and a deep learning model with multiple implicit layers. The results of the calculation are applied to the values of the child nodes of this node. Consider such a calculated set, which can be allowed in every node and possible graph structure, and define a family of functions. The input node has no parent node and the output node has no child node. A special attribute of this flow direction graph is depth: the length of the longest path from one input to one output. Traditional feedforward neural networks can be seen as having a depth equal to the number of layers (for example, the number of hidden layers plus 1 for output layers). SVMS has depth 2 (one corresponds to the core output or feature space and the other corresponds to the linear mixing of the resulting output).(LeCun, Bengio & Hinton, 2015) One of the directions of artificial intelligence research is the so-called "expert system" as defined by a large number of "if - then" rules, top-down ideas. Artificial Neural Network (Artificial Neural Network) marks another bottom-up approach. Neural networks do not have a strict formal definition. Its basic feature is an attempt to mimic patterns of transmission between neurons of the brain, processing information. Problems that need to be solved with deep learning have the following characteristics: Insufficient depth can cause problems. The human brain has a deep structure. The cognitive process is carried out layer by layer, gradually abstract. Problems with insufficient depth:

1. In many cases, depth 2 is sufficient to represent any function with the precision of the given target. But the cost is that the number of nodes needed in the graph, such as the number of calculations and parameters, can change very large. The theoretical results

confirm that the number of nodes required in fact increases with the size of the input.

2. Most randomly selected functions cannot be effectively represented, either with deep or shallow architectures. But many can be effectively represented by deep architectures, but not by shallow architectures. The presence of a tight and deep representation means that there is some kind of structure in a potentially representable function. If no structure exists, it will not be possible to generalize well.

Visual cortex is well studied and shows a series of regions, each of which contains a representation of one input and a signal flow from one to the other (here omits associations on parallel paths at some levels, so more complex). Each layer of this feature hierarchy represents inputs on a different abstraction layer and has more abstract features at the upper level of the hierarchy, they are defined according to the lower level features. Note that the representation in the brain is closely distributed in the middle and purely local: they are sparse: 1 percent of the neurons are active simultaneously. Given a large number of neurons, there is still a very efficient (exponential) representation. Humans organize ideas and concepts hierarchically; humans first learn simple concepts and then use them to represent more abstract; engineers break down tasks into multiple abstract levels to deal with; learning/discovering these concepts (knowledge engineering fails without reflection?) It was wonderful. Reflection on the concept of language expression also suggests a sparse representation: only a small part of all possible words/concepts can be applied to a particular input (a visual scenario).(LeCun et al., 2015)

Suppose the system S, which has n layers (S1, . . . Sn), its input is I and the output is O, figuratively represented as: I = > S1 = > S2 = >. . . . . = > Sn = > O, if output O is identically tantamount to input I, i. e. input I has no information loss after this system change, set to process a information to get b, then process b to get c, then it can prove that the mutual information of a and c will not exceed the mutual information of a and b. This betokens that information processing does not integrate information and most processes lose information, which betokens that input I passes through every layer of Si without any loss of information, i. e. at any caliber of C, it is another representation of the pristine information (i. e. input I). Assuming in this study have a bunch of input I (like a bunch of images or text), assuming we design a system S (with n layers), by adjusting the system parameters so that its output is still input I, then system can automatically get a series of hierarchical features of input I, i. e. S1, . . . , Sn.(Xu, 2015) For deep learning, the idea is to stack multiple layers, meaning that the output of this layer is as input to the next layer. In this way, the input information can be expressed in a hierarchical manner.(Xu, 2015) In addition, the previous assumption is that the output is strictly equal to the input, this restriction is too strict to ease the limit slightly, for example, as long as the difference between input and output is as small as possible, this relaxation will lead to a different class of Deep Learning methods. This is the basic idea of Deep Learning.(Xu, 2015) Think of the learning structure as a network, the core idea of deep learning is as follows:

1. Unsupervised learning is used for pre-train of each layer

2. Training only one layer at a time with unsupervised learning, and use its training results as the input of the higher layer

3. Adjusting all layers with top-down supervised algorithm

This study focuses on Deep Convolutional Neural Network that is given comprehensive information as follow.

2.5 Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a class of Feedforward Neural Networks (Feed-forward Neural Networks) with convolutional computation and is one of the representative algorithms for deep learning(I. Goodfellow, Bengio & Courville, 2016)(Gu et al., 2018). Convolutional neural networks are also known as ”Shift-Invariant Artificial Neural Networks, SIANN”(S. B. Lo, Freedman, Mun & Gu, 2018) because they are capable of shift-invariant classification. The research on convolutional neural networks began in the 1980s and 1990s. Time delay networks and Lenet-5 were the first convolutional neural networks(LeCun, Bengio et al., 1995). In the 21st century, with the advent of deep learning theory and the improvement of numerical computing equipment, convolutional neural networks developed rapidly and were widely used in computer vision and natural language processing(S. B. Lo et al., 2018).

The visual perception mechanism of convolutional neural network imitators can be constructed for supervised and unsupervised learning. The sharing of convolutional kernel parameters and the sparsity of interlayer connections in the implied layer allows convolutional neural networks to be able to point grid-like topology with a smaller computation amount, such as pixel and audio learning, andStable effect and no additional feature engineering requirements for data(I. Goodfellow et al., 2016)(S. B. Lo et al., 2018).

2.5.1 History of Convolutional Neural Networks

Convolutional neural networks can be traced back to the neocognition model proposed by Japanese scholar Kunihiko Fukushima. In this papers published in 1979(Fukushima, 1993)(Fukushima, 1980) and 1980(Fukushima, 1980), the visual cortex of Fukushima imitations designed neural networks named “neocognition”. neocognition is a neural network with deep structure and is one of the first proposed deep learning algorithms(Schmidhuber, 2015), the implicit layer consists of an alternating S layer (Simple-Layer) and C layer (Complex-Layer). The S-layer unit extracts image features within the receptive field, and the C-layer unit receives and responds to the same characteristics returned by different sensory fields(Fukushima, 1980). The S-C layer combination of neocognition can perform feature extraction and screening, partially realizing the function of convolution layer and pooling layer in convolutional neural network. It is considered a pioneering study that inspired convolution neural network(LeCun, Kavukcuoglu & Farabet, 2010).

The first convolutional neural network is the Time Delay Neural Network (TDNN) proposed by Alexander Waibel et al.(Waibel, Hanazawa, Hinton, Shikano & Lang, 1995) in 1987. TDNN is a convolutional neural network used for speech recognition problems. It uses FFT preprocessed speech signals as input. Its implicit layer consists of two one-dimensional convolution cores to extract the translation invariant characteristics on the frequency domain(Waibel et al., 1995). Since the advent of TDNN, there was a breakthrough in the research of Back-Propagation (BP) in the field of artificial intelligence before the advent of TDNN(Rumelhart, Hinton, Williams et al., 1988), TDNN was able to learn using the BP framework. In comparison experiments with the original authors, TDNN outperformed the Hidden Markov Model (HMM) under the same conditions, which was the mainstream algorithm for speech recognition in the 1980s(Waibel et al., 1995).

In 1988, Wei Zhang presented the first two-dimensional convolutional neural network: the translation invariant artificial neural network (SANN) and applied it to the detection of medical imaging(S. B. Lo et al., 2018). Independent of Zhang (1988) (S. B. Lo et al., 2018), Yann LeCun also built a convolutional neural network for image classification, the original version of LenNet(LeCun et al., 1989). The LenNet consists of two convolutional layers, two fully connected layers, a total of 60, 000 learning parameters, much larger than TDNN and SANN, and structurally close to modern convolutional neural networks(LeCun et al., 2010). LeCun (1989)(LeCun et al., 1989) Stochastic Gradient Descent (SGD) was used to learn by randomly initializing weights. This strategy was widely used in subsequent deep learning studies. In addition, LeCun (1989) used the term ”convolution”(LeCun et al., 1989) for the first time when discussing its network structure, and the “convolutional neural network” was named.

The work of LeCun (1989)(LeCun et al., 1989) was completed in 1993 by AT and T Bell Laboratories and was heavily deployed in the check reading system of NCR (National Cash Register Coporation)(LeCun et al., 2010). In general, however, the convolutional neural network designed for various image processing problems during this period remained at the research stage and was not widely used(Gu et al., 2018) due to limited numerical computing capacity and insufficient learning samples.

On the basis of LenNet, in 1998 Yann LeCun and its collaborators built a more complete convolutional neural network Lenet-5 and succeeded in identifying handwritten numbers(LeCun et al., 1998). Lenet-5 follows the learning strategy of LeCun (1989) and includes a pool layer to filter input features in the original design (LeCun et al., 1998). Lenet-5 and its subsequent variants define the basic structure of modern convolution-neural networks, which alternately appear in the construction of the convolution-pool layer is considered to be effective in extracting the translation invariant features of the input image. The success of Lenet-5 has brought attention to the application of convolutional neural networks, which Microsoft developed Optical Character Recognition (OCR) in 2003 using convolutional neural networks(Simard, Steinkraus, Platt et al., 2003). Other applied research based on convolutional neural networks have also been carried out, including portrait recognition(Garcia & Delakis, 2004), gesture recognition(Nowlan & Platt, 1995).

After 2006, with the improvement of deep learning theory, especially the advent of fine-tuning technology(Hinton & Salakhutdinov, 2006), the convolutional neural network began to develop rapidly, deepening its structure, and various learning and optimization theories were introduced(Gu et al., 2018). Since AlexNet(*Large Scale Visual Recognition*, 2012) in 2012, various convolutional neural networks have been the winning algorithms of the russakovsky2015imagenet Large Scale Visual Recognition Challenge (ILSVRC)(Russakovsky et al., 2015), including the 2013ZFNet(*Large Scale Visual Recognition*, 2013), VGGNet in 2014, GooGlenet(*Large Scale Visual Recognition*, 2014) and ResNet(*Large Scale Visual Recognition*, 2015) in 2015.

2.5.2 Convolutional Neural Network in Image Processing

Convolutional neural networks have long been one of the core algorithms in the field of image recognition and have a stable performance when learning large amounts of data (Egmont-Petersen, de Ridder & Handels, 2002). For general large-scale image classification problems, convolutional neural networks can be used to build hierarchical classifiers(Srivastava & Salakhutdinov, 2013) or to extract the discriminant features of images in fine-grained recognition for other classifiers to learn(Ji et al., 2012). For the latter, feature extraction can artificially input different parts of the image into the convolutional neural network(Egmont-Petersen et al., 2002), or by the convolutional neural network itself through unsupervised learning.

For text detection and text recognition / optical character reading, the convolutional neural network is used to determine whether the input image contains characters and to extract valid character fragments from it(Zhang, Yao, Shi & Bai, 2015). Convolutional neural networks using multiple normalized exponential functions are used to identify the house number of Google Street View images(I. J. Goodfellow, Bulatov, Ibarz, Arnoud & Shet, 2013), Convolutional neural networks containing conditional Random Fields (CRF) graph models can identify words in images(Jaderberg, Simonyan, Vedaldi & Zisserman, 2014), Convolutional neural networks versus Circular neural networks (Repeated Neural Network, RNN, can extract character features from images and sequence labelling respectively(P. He, Huang, Qiao, Loy & Tang, 2016).

1. Object Recognition

   Convolutional neural networks can identify objects in three ways: sliding window, selective search, and you Only Look Once(Gu et al., 2018). Sliding windows appeared first and were used for gesture recognition issues(Nowlan & Platt, 1995), but due to the large amount of calculations, the latter two have been eliminated(Gu et al., 2018). Selective search corresponds to regional convolutional neural networks (region-based CNN), which first determines whether a window may have a target object by general steps, and further enters it into a complex identifier(Girshick, Donahue, Darrell & Malik, 2014). The YOLO algorithm defines object recognition as a regression problem with probability of each target in a split box in the image, and uses the same convolutional neural network to output the probability, center coordinates and box dimensions of each target(Redmon, Divvala, Girshick & Farhadi, 2016). Object recognition based on convolutional neural networks has been applied to autonomous driving(Maturana & Scherer, 2015) and traffic real-time monitoring systems(Kyrkou, Plastiras, Theocharides, Venieris & Bouganis, 2018).

   Convolutional neural nets are also used in semantic segmentation (Noh, Hong & Han, 2015) , scene labeling and image saliency detection(Oquab, Bottou, Laptev & Sivic, 2014). Confirmed more than many classification systems using feature engineering.

2. Action Recognition

   In behavioral cognition studies, image features extracted from convolutional neural networks are widely used in action classification(Oquab et al., 2014). In the behavioral cognitive problem of video, convolutional neural networks can maintain their two-dimensional structure and learn by stacking the features of continuous time fragments(Karpathy et al., 2014), build 3D convolutional neural networks that change along the timeline(Tran, Bourdev, Fergus, Torresani & Paluri, 2014), or extract features frame-by-frame and input circular neural networks(Donahue et al., 2015).

3. Pose Estimation

   The first convolutional neural network used in posture estimation is DeepPpose. The structure of DeepPpose is similar to AlexNet, with complete image as output, trained and output coordinate points according to supervised learning(Toshev & Szegedy, 2014). There are also research on the application of convolutional neural networks on local attitude estimation(J. J. Tompson, Jain, LeCun & Bregler, 2014)]. For video data, a frame-by-frame attitude estimation using the convolutional neural network of sliding windows has been studied(Jain, Tompson, LeCun & Bregler, 2014).

4. Neural Style Transfer

   Neural style conversion is a special application of convolutional neural networks, whose function is to create a third image based on a given two images and bring its content and style as close as possible to a given image(Gatys, Ecker & Bethge, 2015). Neural style conversion takes advantage of the propagation of image features in convolutional neural networks, and defines the error function of content loss and style loss(Gatys et al., 2015). In addition to artistic creation, neurostyle conversion is also used for post-processing of photos(Luan, Paris, Shechtman & Bala, 2017).

2.5.3 Convolutional Neural Networks Architecture

Whereas the classic network architectures were comprised essentially of stacked convolutional layers, advanced models investigate unused and inventive ways for building convolutional layers in a way which permits for more effective learning. Nearly all of these structures are based on a repeatable unit which is utilized all through the arrange.

1. Input Layer

   Input layers of convolutional neural networks can handle multidimensional data. Normally, the input layers of 1D convolutional neural networks receive 1D or 2D arrays, where 1D arrays are usually time or spectrum sampling; 2D arrays may contain multiple channels; 2D convolutional neural networks input layers receive 2D or 3D arrays; and 3D convolutional neural networks receive 4D arrays(I. Goodfellow et al., 2016). Because convolutional neural networks are widely used in computer vision, many studies introduce their structure presuppose three-dimensional input data, i. e. 2D pixels and RGB channels on a plane. Similar to other neural network algorithms, the input features of convolutional neural networks need to be standardized due to the use of gradient descent for learning. Specifically, before entering the learning data into the convolutional neural network, the input data should be normalized at the channel or time/frequency dimension, or the original pixel values distributed in pixels can be normalized to the interval(I. Goodfellow et al., 2016). Standardization of input features helps improve the efficiency and learning performance of the algorithm(I. Goodfellow et al., 2016).

2. Implicit Layer

The implicit layer of convolutional neural networks consists of 3 common constructions of convolution layer, pool layer, and full connection layer. In some more modern algorithms, there may be complex constructions such as the Assuming module and residual block. In common constructions, convolutional and pooled layers are unique to convolutional neural networks. Convolution kernels in a convolution layer contain weight coefficients, while pooled layers do not contain weight coefficients, so in literature, pooled layers may not be considered independent layers. Taking Lenet-5 as an example, the order of 3 common types of structures in implicit layers is typically:

* Input-Convolution Layer
* Pooled Layer
* Convolution Layer
* Pooled Layer
* Fully Connected Layer
* Output.

3. Convolution Layer

The function of the convolution layer is to feature extraction of input data, which contains multiple convolution cores, each element that makes up the convolution core corresponds to a weight factor and a bias vector, similar to a neuron (neuron) of a feedforward neural network. Each neuron in the convolution layer is connected to multiple neurons in the area close to the previous layer. The size of the area depends on the size of the convolution nucleus. It is referred to in the literature as “receptive field”, meaning analogous to the sensory field of visual cortex cells(Gu et al., 2018). When the convolution kernel is working, the input features are routinely sweeps, multiplying the matrix elements in the sensory field and overlaying the amount of deviation(I. Goodfellow et al., 2016):

$$Z^{l+1}(i,j) = [Z^l \theta w^l](i,j) + b = \sum_{k=1}^{K_l} \sum_{x=1}^{f} \sum_{y=1}^{f} [Z^{l_k}(s_0 i + x, s_0 j + y) w_k^{1+l}(x,y)] + b \quad (1)$$

Convolution layers consisting of unit convolution cores are also known as Network-In-Network (NIN) or multilayer perceptron convolution layer, mlpconv)(Lin, Chen & Yan, 2013). The unit convolution kernel reduces the amount of convolution layer calculations by reducing the number of channels of the graph while maintaining the characteristic graph size. The unit convolution kernel reduces the amount of convolution layer calculations by reducing the number of channels of the graph while maintaining the characteristic graph size. Convolutional neural network, built entirely by unit convolution kernel, is a multi-layer perceptor (Muti-Layer Perceptron, MLP) with parameter sharing(Lin et al., 2013) On the basis of linear convolution, some convolution neural networks use more complex convolutions, including tiled convolution, deconvolution, and dilated convolution(Gu et al., 2018). The convolution kernel of the tiled convolution sweeps only one part of the feature map, and the rest is processed by other convolution cores in the same layer, so the parameters between the convolution layers are only partially shared, allowing neural networks to capture the shift-invariant feature of the input image(Ngiam et al., 2010). An anti-convolution or transposed convolution connects a single input excitation to multiple output excitations to amplize the input image. Convolutional neural networks consisting of anti-convolution and up-pooling layers have important applications in the field of semantic segmentation(Noh et al., 2015) and are also used to build convolutional Autoencoder (CAE)(Marc'Aurelio Ranzato, Boureau & LeCun, 2007). Expansion Convolution introduces expansion rates based on linear convolution to increase the sense field of convolution kernel, thus obtaining more information on feature maps(Kalchbrenner et al., 2016), enabling long-range dependency to capture learning goals when using sequential data. Convolutional neural networks using expanded convolution are mainly used in natural language processing (NLP) areas such as machine translation(Kalchbrenner et al., 2016), speech recognition(Sercu & Goel, 2016), etc.

4. Convolution Layer Parameters

Convolutional layer parameters include Convolution kernel size, step length and fill, which together determine the size of the Convolution Layer output feature map, and are the hyperparameters of the Convolution Neural Network(I. Goodfellow et al., 2016). The convolution kernel size can be specified as any value smaller than the input image size. The larger the convolution kernel, the more complex the input features can be extracted(I. Goodfellow et al., 2016).

The convolution step defines the distance between the convolution kernel and the two times that the convolution kernel is adjacent to it. When the convolution step is 1, the convolution kernel sweeps the elements of the pattern one by one, and the step is n skipped n-1 pixels(Dumoulin & Visin, 2016) at the next scan.

The cross-correlation calculation of the convolution kernel shows that the size of the feature map decreases gradually as the convolution layer is stacked. For example, the 16 * 16 input image will output 12 * 12 after the 5 * 5 convolution kernel in unit step. To do this, filling is a method of artificially increasing the size of a feature plot before it passes through the convolution kernel to offset the impact of the size shrinkage in the calculation. The common method of filling is to fill by 0 and repeat boundary value padding. Filling can be divided into four categories according to the number of layers and purpose(Dumoulin & Visin, 2016):

(a) Valid padding: that is, no padding is used at all, the convolution kernel only allows access to locations where the feature map contains the full feeling wild. All pixels of the output are functions with the same number of pixels in the input. A convolution that uses a valid fill is called a narrow convolution, and the characteristic plot size of the narrow convolution output is (L-F) / s + 1.

(b) Same padding (same/half padding): Only enough padding is done to keep the output and input the characteristic size of the same. The size of the pattern with the same padding does not shrink but the part of the input pixel near the boundary has less impact on the map than the middle part, i. e. the presence of an underrepresentation of the boundary pixel. Convolution that uses the same padding is called "equal-width convolution".

(c) Full padding: Make enough padding so that each pixel is accessed the same number of times in each direction. With a step of 1, the full-fill output's characteristic size is L + f-1, larger than the input value. Using a fully filled convolution is called "wide convolution"

(d) Arbitrary padding: between valid padding and full padding, artificially set padding, less used.

In the previous example, if the 16 * 16 input image is filled the same before the 5 * 5 convolution kernel in a unit of step, both layers are filled horizontally and vertically, i.e. 2 pixels () are added to the 20 * 20 size image on each side. After passing the convolution kernel, the output feature image size is 16 * 16, keeping the original size.

5. Activation function

The convolution layer contains excitation functions to help express complex features, expressed in the following form(I. Goodfellow et al., 2016):

$$A^l_{i,j,k} = f(Z^l_{i,j,k}) \tag{2}$$

Similar to other deep learning algorithms, convolutional neural networks typically use a linear rectified unit (relU). Other relU similar variants include leaky relu with slope, parameterized relu (Parametric relu, Prelu), Randomized relu (Randomized relu, rrelu), exponential linear unit (Exponential Linear Unit, ELU), etc.(Gu et al., 2018). Prior to the advent of relU, the Sigmoid function and hyperbolic tangent function (hyperbolic tangent) were commonly used excitation functions(LeCun et al., 1998).

Excitation function operations usually follow the convolution kernel, some algorithms that use preactivation techniques place the excitation function before the convolution kernel(K. He, Zhang, Ren & Sun, 2016b). In some early convolutional neural network studies, such as Lenet-5, the excitation function is behind the pooling layer(LeCun et al., 1989).

6. Pooling layer

After feature extraction in the convolution layer, the output feature map is transferred to the pooled layer for feature selection and information filtering. Pooling layers contain preset pooling functions that replace the result of a single point in a feature map with a characteristic map statistic for adjacent areas. Pooling layers select pooled areas as convolutional kernel scan feature map steps, controlled by pooled size, step length, and fill(I. Goodfellow et al., 2016).

(a) Lp pooling

Lp pooling is a class of pooling models inspired by the visual intracortex structure(Hyvärinen & Köster, 2007). The general representation is(Bruna, Szlam & LeCun, 2013):

$$E^{(l)}(w) = \sum_{i=1}^{m_1^{(l-1)}} \left\| \sum_{j=1}^{m_1^{(l)}} K^{(l)}_{j,i} * Y^{(l)}_j - Y^{(l-1)}_i \right\|_2^2 + \sum_{i=1}^{m_1^{(l)}} \left\| Y^{(l)}_i \right\|_p^p \tag{3}$$

Pixel, is the same as the convolution layer, and is a pre-specified parameter. At that time, Lp pooling was averaged within the pool area, known as average pool ng; at that time, Lp pooling was maximal in the region and called max pooling. Mean pooling and maximal pooling are the most common methods of pooling, which retain background and texture information of an image at the expense of loss feature map size(Bruna et al., 2013). In addition, L2 pooling is also used in some work(Cahill, Casazza, Peterson & Woodland, 2013).

(b) Random/Hybrid Pooling

Mixed pooling and stochastic pooling are an extension of the concept of Lp pooling. Random pooling selects a value randomly by a specific probability distribution within its pooling area to ensure that some of the less significant excitation signals go to the next build(Zeiler & Fergus, 2013). Hybrid pooling can be represented as a linear combination of mean pooling and very pooling(Yu, Wang, Chen & Wei, 2014):

$$A_k^l = \lambda L_1(A_k^l) + L_{n..n}(A_k^l), \lambda \epsilon [0, 1] \tag{4}$$

Studies show that mixed pooling and stochastic pooling help prevent overfit of convolutional neural networks and perform better than mean and large pooling(Gu et al., 2018).

(c) Spectral Pooling

Spectral pooling is a FFT based pooling method that can be used together with FFT convolution to construct FFT-based convolution neural networks(Mathieu, Henaff & LeCun, 2013). Spectral pooling performs DFT transformation for each channel of the feature map, and intercepts n * n size sequences from the center of the spectrum for DFT inverted transformation(Mathieu et al., 2013). Spectrum pooling has a filtering function that maximizes the low frequency change information and can effectively control the size of the characteristic chart(Mathieu et al., 2013). In addition, based on a mature FFT algorithm, spectral pooling can be done with a small amount of computation.

7. Inception module

The Endance module is a special implicit layer built from the stacking of multiple convolution and pooled layers. Specifically, an Intensify module will contain several different types of convolution and pooling operations at the same time, and use the same fill to get the same size of a feature plot, then overlay the channels of these patterns in the array and pass the excitation function(Szegedy et al., 2015). Since the above approach introduces multiple convolution calculations in one construction, the calculation volume increases significantly, so to simplify the computation, the Initiation module typically designs a bottleneck layer, using a unit convolution kernel, the NIN structure, to reduce the number of channels for the feature map, and then perform other convolution operations(Szegedy et al., 2015). The Starting module was first applied to GooGlenet with remarkable success(Szegedy et al., 2015) and inspired the idea of depthwise separable convolution in Xception algorithms(Chollet, 2017).

8. Fully-connected layer

The full connection layer in convolutional neural networks is equivalent to the implied layer in traditional feed-forward neural networks. The full connection layer is usually built in the last part of the convolutional neural network implicit layer and only passes signals to the other fully connected layer. The feature map loses a 3-dimensional structure in the full join layer, which is unfolded as a vector and passed through the excitation function to the next layer(I. Goodfellow et al., 2016).

In some convolutional neural networks, the function of the fully connected layer can be partially replaced by global average pooling(Hyvärinen & Köster, 2007), which averages all the values of each channel of the feature map. That is, if there is a 7 * 7 * 256 characteristic plot, the global mean pooled will return a 256 vector, where each element is 7 * 7 and steps 7, no padding means pooling(Szegedy et al., 2015).

9. Connectivity

   Connections between convolutional layers in convolutional neural networks are called sparse connections, i. e. neurons in the convolutional layer are connected only to their adjacent layers, not all neurons, compared to full connections in feedforward neural networks. Concretely, any pixel (neuron) in the L-layer feature plot of the convolutional neural network is simply a linear amalgamation of pixels in the sensory field defined by the convolution kernel in the l-1 layer(I. Goodfellow et al., 2016). Sparse connections of convolutional neural networks have the effect of regularization, amend the stability and generalization of the network structure, evade overfit. Concurrently, sparse connections reduce the total amount of weight parameters, facilitate expeditious learning of neural networks, and reduce recollection overhead in computing(I. Goodfellow et al., 2016).

   All pixels within the same channel of a feature plot in a convolutional neural network share a set of convolutional kernel weight coefficients, known as weight sharing. Weight sharing distinguishes convolutional neural networks from other neural networks that contain local connection structures that use sparse connections but have different weights for different connections(I. Goodfellow et al., 2016). Weight sharing, like sparse connections, reduces the total number of parameters of convolutional neural networks and has the effect of regularization(I. Goodfellow et al., 2016).

   From a plenarily connected network perspective, the sparse connection and weighting sharing of convolutional neural networks can be considered as two illimitable priori (pirior), i. e. an implicatively insinuated neuron has a constant weight coefficient of 0 outside its sensory field (but the sensory field can move in space); and all neurons have the same weight coefficient in one channel(I. Goodfellow et al., 2016).

   Output Layer

   The upstream of the output layer in a convolutional neural network is usually a full join layer, so it is structured and works in the same way as the output layer in traditional feedforward neural networks. For image classification issues, the output layer uses a logical function or a softmax function to output classification labels(I. Goodfellow et al., 2016). In object detection problems, the output layer can be designed to be the center coordinate, size, and classification of the output object(I. Goodfellow et al., 2016). In image semantic segmentation, the output layer directly outputs the classification results for each pixel(I. Goodfellow et al., 2016).

2.5.4 Learning Paradigms for CNN

There are five learning paradigms, each comparing to a specific theoretical learning task.

1. Supervised Learning

   Convolutional neural networks use BP framework to learn in supervised learning, and their computational process has been identified in LeCun (1989)(LeCun et al., 1989), one of the first deep algorithms to learn in BP framework. The BP in the convolutional neural network is divided into three parts: the reverse propagation of the fully connected layer and the convolution kernel and the backward pass of the pooled layer(I. Goodfellow et al., 2016)(I. Goodfellow et al., 2016). The BP calculation of the full connection layer is the same as the traditional feed-forward neural network. The reverse propagation of the convolution layer is a cross-correlation calculation similar to forward propagation:

$$\left(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{A}}\right)_{i,j}^{l} = \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[\boldsymbol{w}_k^{l+1}(x,y)\left(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{A}}\right)_{s_0 i+x, s_0 j+y, k}^{l+1}\right] f'\left(\boldsymbol{A}_{i,j}^{l}\right)$$

$$\boldsymbol{w}^{l} = \boldsymbol{w}^{l+1} - \alpha\left(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{w}}\right)_k = \boldsymbol{w}^{l+1} - \alpha\left[\boldsymbol{A}^{l+1}\left(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{A}}\right)_k^{l+1}\right]$$

The error of the cost function calculation, the derivative of the stimulus function, is the learning rate. If the forward propagation of the convolution kernel uses the convolution calculation, the reverse propagation also flips the convolution kernel for convolution operation. There are several options for error functions in convolutional neural networks, including softmax loss function (softmax loss), hinge loss function (hinge loss), triplet loss function (triplet loss), etc.(Gu et al., 2018).

Pooling layer has no parameter update in reverse propagation, so it is only necessary to assign the error to the appropriate location of the feature map according to the pooling method. For large pooling, all errors are assigned to the maximum value; for mean pooling, the error is evenly distributed to the entire pool area(I. Goodfellow et al., 2016).

Convolutional neural networks typically use stochastic Gradient Descent (SGD) within the BP framework(Bottou, Curtis & Nocedal, 2018) and its variants, such as the Adam moment estimation(Kingma & Ba, 2014). SGD randomly selects samples to calculate gradients in each iteration, which facilitates information filtering in the case of a large number of learning samples, can quickly converge at the beginning of the iteration, and the computational complexity is less(Bottou et al., 2018).

2. Unsupervised Learning

   Convolutional neural networks are widely used in supervised learning issues, but unsupervised learning can also be done using tabless data(Turchenko, Chalmers & Luczak, 2017)(Jarrett, Kavukcuoglu, Ranzato & LeCun, 2009). Available methods include Convolutional Autoencoders (CAE)(Jarrett et al., 2009), Convolutional Restricted Boltzmann Machines (CRBM) / Convolutional Deep Confidence Network (Convolutional Deep Belief Networks, CDBN)(Lee, Grosse, Ranganath & Ng, 2009) and Deep Convolutional Generative Adversarial Networks(Lee et al., 2009). These algorithms can also be seen as hybrid algorithms that introduce convolutional neural network construction in the original version of unsupervised learning algorithms.

   The construction logic of CAE is similar to traditional AEs, first using convolutional and pooled layers to build conventional convolutional neural networks as encoders, then using anti-convolution and up-pooling as decoders to learn from errors before and after sample encoding, and output encoders to achieve dimension reduction of samples (dimentionality)reduction and clustering. In image recognition problems, such as MNIST, CAE's convolutional neural network with the same structure as its encoder behaves fairly at large samples, but has better recognition in small sample problems(Turchenko et al., 2017).

CRBM is a constrained Boltzmann Machines (RBM) with convolution layer as implicit layer. On the basis of traditional RBMS, the implicit layer is divided into multiple "groups", each containing a convolution core, the parameters of the convolution kernel are shared by all the binary nodes corresponding to that group(Lee et al., 2009). CDBN is a hierarchical model based on CRBM. In order to extract higher-order features from the structure, CDBN incorporates probabilistic max-pooling layer and its corresponding energy functions. CrBMs and CDBMs learn using greedy layer-wise training(Bengio, Lamblin, Popovici & Larochelle, 2007) and can use sparsity regularization techniques. In the object recognition problem of Caltech-101 data, a 24-100 two-layer CDBN recognition accuracy is equal or exceeds many classification and clustering algorithms that contain highly specialized features(Bengio et al., 2007).

Generative Adversary Networks (GAN) can be used for unsupervised learning of convolutional neural networks. DCGAN samples randomly from a set of probability distributions, latent space, and enters the signal into a set of generators entirely composed of transposed convolution cores; the generator generates the image and enters theConvolutional neural network, the discriminant model determines whether the resulting image is a real learning sample. Learning ends when the generated model makes it impossible for the discriminating model to determine the difference between its generated image and the learning sample. The results show that DCGANS can extract important features of input images from image processing problems. In the experiment of CIFAR-10 data, the characteristics of DCGAN discriminant model are processed as input to other algorithms, and can classify images with high accuracy(Jarrett et al., 2009).

3. Regularization

Various regularization methods in neural network algorithms can be used to convolutional neural networks to prevent overfit. Common regularization methods include Lp-Norm regularization, spatial dropout, and connect drop. Lp regularization adds implicit layer parameters when defining loss functions to constrain the complexity of neural networks:

$$L(X, Y, w) = L(X, Y, w) + \lambda \sum \|w\|^p \tag{5}$$

The sum term containing the Frobenius norm is called the regularization term, in which the regularization parameter is used to determine the binding force of the regularization term. It can be shown that the regularization term was convex function(Hinton, Srivastava, Krizhevsky, Sutskever & Salakhutdinov, 2012); in particular, L2 regularization was again called Tikhonov regularization(von Wedelstedt, Arnrich, Haehnel & Kalies, 2018). while Lp regularization facilitates thinning of convolution kernel weights, but regularization at this point is not convex function(Gu et al., 2018).

Spatial dropout in convolutional neural networks is the promotion of random inactivation theory in feed-forward neural networks. In fully connected network learning, random inactivation will randomly zero the output of the neuron, while space random inactivation will randomly select the channel of the characteristic map to zero(J. Tompson, Goroshin, Jain, LeCun & Bregler, 2015). Further, random join deactivation acts directly on the convolution kernel, which returns part of the convolution kernel weights to zero(Wan, Zeiler, Zhang, Le Cun & Fergus, 2013) in the iteration. The results show that spatial random inactivation and random connection deactivation improve the generalization ability of convolutional neural networks and have significant effect when learning samples are insufficient(J. Tompson et al., 2015)(Wan et al., 2013).In this regard, we must explain two important elements:

(a) Batch Normalization,

Standardization of data is an important step in preprocessing in neural network input pipelines. In deep networks, the mean and standard deviation change as a feature passes in the implicit layer, creating a covariate shift phenomenon(Ioffe & Szegedy, 2015). Covariant drift is an important reason for vanishing gradient (vanishing gradient) in deep networks(Ioffe & Szegedy, 2015). BN partially solves this problem at the expense of introducing additional learning parameters. Its strategy is to first standardize the feature in the implicit layer, then use two linear parameters to amplify the normalized feature as a new input, and the neural network updates its BN parameters in the learning process(Ioffe & Szegedy, 2015). The BN parameter in a convolutional neural network has the same properties as the convolutional kernel parameter, i. e. pixels in the same channel share a set of BN parameters(Ioffe & Szegedy, 2015). In addition, the convolution layer does not require a deviation term when using BN, its function is replaced by the BN parameter.

(b) Skip Connection,

The jump connection or shortcut connection is derived from the hopping connection and various gating algorithms in the Recurrent Neural Network (RNN). It is an important technique for solving the gradient disappearing problem in deep networks(K. He, Zhang, Ren & Sun, 2016a). Jump connections in convolutional neural networks can span any number of implicit layers(K. He et al., 2016a), described here by jumping between adjacent implicit layers:

$$A^l = f(Z^l + uZ^{l-1}) \tag{6}$$

This equation is the transformation factor of the characteristic graph. When the size of the sum is different, the conversion factor will be smaller, usually converted to the size, ensuring that the matrix elements are arithmetic(K. He et al., 2016a). When the output value of the output value is small and large, the output of the convolution layer is approximate to the equivalent function, and there is no negative impact on the characteristic transfer of the layer, so that the layer will not degrade at least during the iteration. In the BP framework, some errors can be skipped and applied directly to the layer when propagating backwards, compensating for the loss of gradient caused by the progressive propagation in the depth structure, thus facilitating the error propagation in the depth structure. The combination of multiple convolution layers with jumping connections is called residual blocks and is an important construction technique for modern convolution neural networks(K. He et al., 2016a).

4. Dimensional algorithms

TDNN is a class of one-dimensional convolutional neural networks applied to speech recognition problems. It is also one of the first convolutional neural network algorithms proposed in history. The original version of TDNN is Waibel et al. (1987)(Waibel et al.,

1995) as an example. The learning goal of TDNN is to classify the three phonetic syllables /b, d, g/of the FFT transformation. The implicit layer consists entirely of a single-step, unfilled convolution layer(Waibel et al., 1995). In the literature, the convolution kernel size of TDNN uses the expression "delay", and the implied layer consisting of a one-dimensional convolution kernel of size 3 is defined as "the implied layer of time delay 2", i. e. the Feel Wild contains no delay input and 2 delay inputs(Waibel et al., 1995). On this basis, TDNN has two convolution layers with time delay of 2 and 4, and each input signal in the neural network is connected to eight implicit layer neurons(Waibel et al., 1995). TDNN does not have a full connection layer, but instead adds the output of the tail convolution layer directly to the classification results through the excitation function. According to original, the input TDNN preprocessed data is 15 10-millisecond samples (frames), each containing 16 channel parameters (filterbank coefficients), at which time the TDNN is structured as follows(Waibel et al., 1995):

(a) (3) x 16 * 8 convolution layer (step 1, no fill, Sigmoid function)

(b) (5) x 8 * 3 convolution layer (step 1, no fill, Sigmoid function)

(c) Output of Feature Mapping for 9 * 3

The number in the list means: (Convolution kernel size) * Convolution kernel channels (the same number of input data channels) * Convolution cores. Both the output layer of TDNN and the two convolution layers use the Sigmoid function as the excitation function. In addition to the original version above, algorithms for character recognition(Jaeger, Manke, Reichert & Waibel, 2001) and object recognition(Wöhler & Anlauf, 2001) emerged in subsequent studies of TDNN, which work by expanding space in channel dimensions and using a one-dimensional convolution of time, that is, time delay, to learn.

2.5.5 CNN architectures of ILSVRC

ILSVRC(Russakovsky et al., 2015) provides a platform for comparison of sundry AI algorithms applied to computer vision. Several convolutional neural network algorithms are prosperous in image relegation and object apperception tasks, including AlexNet, ZFNet, VGGNet, Googlenet and ResNet, which show good results in russakovsky2015imagenet dataLearning performance is additionally a representative algorithm for the development of convolutional neural networks.

1. WaveNet

   WavenNet is a one-dimensional convolutional neural network used for speech modeling. It features expanded convolution and jumping connections to improve the neural network's learning ability to rely on long distance. WavenNet is designed for sequential data design, and its structure and common convolutional neural networks are very different. Here is a brief introduction by Van Den Oord et al. (2016)(Oord et al., 2016):

   WavenNet has quantified and one-hot encoded audio as an input feature, specifically a two-dimensional array containing samples and channels(Oord et al., 2016). The input features are first entered into the linear convolution kernel in WavenNet. The resulting feature map is passed through multiple dilated stack, each containing a filter (filter) and a gate (gate), both with a step of 1, the same filled linear convolution kernel, but the former using the hyperbolic tangent function as the incentive function, the latter using the Sigmoid function (I. Goodfellow et al., 2016). The characteristic graph outputs from the filter and gate are multiplied by the matrix elements and through the bottleneck layer constructed by NIN. Part of the result is directly output by jumping connection, and the other part is linearly combined with the feature plot before entering the dilated convolution block into the next(I. Goodfellow et al., 2016). The end part of WavenNet adds all the output of the jumping connection and expansion convolution block through two Relu-nin structures, and finally outputs the result by the normalized exponential function and uses cross-entropy as a loss function for supervised learning(Oord et al., 2016)(I. Goodfellow et al., 2016). WavenNet is a generative model that outputs the conditional probability of each sequence element relative to all elements preceding it, having the same dimension as the input sequence(Oord et al., 2016):

$$p(x) = \prod_{t=1}^{T} p(x_t|x_1, x_2, ..x_{t-1}), x = (x_1, x_2..x_T) \tag{7}$$

WavenNet has been proven to produce nearly true English, Chinese and German voices(Oord et al., 2016)(Devarapalli, 2018). After improvements in algorithms and operational efficiency, Wavenet has been providing voice synthesis for Google's commercial application "Google Assistant" since November 2017(Devarapalli, 2018).

2. LeNet-5

   Lenet-5 is a convolutional neural network applied to image classification problems. Its learning goal is to identify and distinguish 0-9 from a series of handwritten numbers represented by 32 * 32 * 1 grayscale images. The implicit layer of Lenet-5 consists of 2 convolution layers, 2 pooled layers, and 2 plenarily connected layers, constructed as follows:

   (a) (3 * 3) * 1 * 6 convolution layer (step 1, no fill), 2 * 2 mean pooling (step 2, no fill), tanh excitation function
   (b) (5 * 5) * 6 * 16 convolution layer (step 1, no fill), 2 * 2 mean pooling (step 2, no fill), tanh excitation function
   (c) 2 fully connected layers with 120 and 84 neurons

3. AlexNet

   AlexNet was the winner of ILSVRC image classification and object recognition algorithms in 2012. It is also an algorithm that has an important effect on modern convolutional neural networks after Letnet-5(I. Goodfellow et al., 2016)(Russakovsky et al., 2015). The implicit layer of AlexNet consists of 5 convolution layers, 3 pooled layers, and 3 fully connected layers, and is constructed as follows(Hinton et al., 2012):

   (a) (11 * 11) * 3 * 96 convolution layer (4 steps, no padding, relu), 3 * 3 great pooling (2 steps, no padding), LRN
   (b) (5 * 5) * 96 * 256 convolution layer (1 step, same fill, relU), 3 * 3 great pooling (step 2, no fill), LRN
   (c) (3 * 3) * 256 * 384 convolution layer (step 1, same fill, relU)
   (d) (3 * 3) * 384 * 384 (step 1, same padding, relu)

(e) (3 * 3) * 384 * 256 convolution layer (step 1, same padding, relU), 3 * 3 great pooling (step 2, no padding)

(f) 3 fully connected layers with 4096, 4096, and 1000 neurons

AlexNet selects relU as the excitation function in the convolution layer, uses random inactivation and data augmentation techniques(Hinton et al., 2012), which are widely used in subsequent convolutional neural networks(Russakovsky et al., 2015). AlexNet is also the first GPU-based convolutional neural network, and Krizhevsky (2012) divides AlexNet into two parts structured, running on two GPU devices. In addition, the 1-2 part of AlexNet uses local response normalization (LRN). In convolutional neural networks that appeared after 2014, LRN has been replaced by batch normalization(I. Goodfellow et al., 2016)(K. He et al., 2016b).

4. ZFNet

ZFNet is the winner of ILSVRC image classification algorithm in 2013. Its structure is similar to AlexNet, only resized the convolution kernel size of the first convolution layer to 7 * 7 and halved the step(Zeiler & Fergus, 2013):

(a) (7 * 7) * 3 * 96 convolution layer (step 2, no padding, relu), 3 * 3 great pooling (step 2, no padding), LRN

(b) (5 * 5) * 96 * 256 convolution layer (1 step, same fill, relU), 3 * 3 great pooling (step 2, no fill), LRN

(c) (3 * 3) * 256 * 384 convolution layer (step 1, same fill, relU)

(d) (3 * 3) * 384 * 384 (step 1, same padding, relu)

(e) (3 * 3) * 384 * 256 convolution layer (step 1, same padding, relU), 3 * 3 great pooling (step 2, no padding)

(f) 3 fully connected layers with 4096, 4096, and 1000 neurons

The contribution of ZFNet to Convolutional Neural Networks is not the construction itself, but the pristine author examined the details of characteristic extraction within ZFNet by anti convolution, expounding the characteristic transfer law of Convolutional Neural Networks, i. e. transition from simple edge and angles to more intricate ecumenical features(Zeiler & Fergus, 2013). The above theory is of great significance for the algorithm improvement and application expansion of convolutional neural networks.

5. VGGNet

VGGNet is a set of convolutional neural network algorithms developed by the Visual Geometry Group (VGG-11, VGG-11-LRN, VGG-13, VGG-16, and VGG-19(Simonyan & Zisserman, 2015). VGG-16 is the winner of the ILSVRC object recognition algorithm in 2014. Its size is more than 2 times the size of AlexNet and has a regular structure. The implicit layer of VGG-16 consists of 13 convolution layers, 3 fully connected layers, and 5 pooled layers, and is constructed as follows:

(a) (3 * 3) * 3 * 64 convolution layer (step 1, same padding, relu), (3 * 3) * 64 convolution layer (step 1, same padding, relu), 2 * 2 great pooling (step 2, no padding)

(b) (3 * 3) * 64 * 128 convolution layer (step 1, same fill, relU), (3 * 3) * 128 * 128 convolution layer (step 1, same fill, relU), 2 * 2 great pooling (step 2, no fill)

(c) (3 * 3) * 128 * 256 convolution layer (step 1, same fill, relu), (3 * 3) * 256 convolution layer (step 1, same fill, relu), (3 * 3) * 256 convolution layer (step 1, same fill, relu), 2 * 2 greatly pooling (step 2, no steps Charge)

(d) (3 * 3) * 512 * 512 convolution layer (step 1, same fill, relu), (3 * 3) * 512 convolution layer (step 1, same fill, relu), (3 * 3) * 512 * 512 (step 1, same fill, relu), 2 * 2 greatly pooling (step 2, no steps padding)

(e) 3 fully connected layers with 4096, 4096, and 1000 neurons

VGGNet construction uses only 3 * 3 convolution kernel and keeps the output feature map size in convolution layer unchanged, the number of channels doubled, and the output feature map size in pooling layer halved, simplifying the topology of neural network and achieving good results(Lee et al., 2009)(Simonyan & Zisserman, 2015).

6. GoogLeNet

GooGlenet is the winner of the ILSVRC image classification algorithm in 2014. It is the first large-scale convolutional neural network to be stacked with the Initial Module. It has four versions: Initial v1, Initial v2, Initial v3, and Initial v4(Szegedy, Ioffe, Vanhoucke & Alemi, 2017). Here is an example of the Intencation v1. First, the Initiation module of the Initiation v1 is divided into four parts(Szegedy et al., 2015):

(a) Convolution of N1 (1 * 1) * C

(b) B3 (1 * 1) * C convolution cores (BN, relu), N3 (3 * 3) * 96 convolution cores (step 1, same fill, BN, relu)

(c) B5 (1 * 1) * C convolution cores (BN, relu), N5 (5 * 5) * 16 convolution cores (step 1, same fill, BN, relu)

(d) Extremely pooling of 3 * 3 (step 1, same fill), Convolutional Cores of Np (1 * 1) * C (BN, relU)

On this basis, for the 3-channel RGB image input, the Inception v1 is built as follows(Szegedy et al., 2015):

(a) (7 * 7) * 3 * 64 convolution layer (step 2, no padding, BN, relu), 3 * 3 great pooling (step 2, same fill), LRN

(b) (3 * 3) * 64 * 192 convolution layer (step 1, same fill, BN, relu), LRN, 3 * 3 great pooling (step 2, same fill)

(c) 3 * 3 great pooling (step 2, same padding)

(d) ECONOMATION Module (N1 = 192, B3 = 96, N3 = 208, B5 = 16, N5 = 48, Np = 64)

(e) Global Mean Pooling, 1 Fully Connected Layer, 1000 Neurons Number, 40/100 Random Inactivation of Weight

7. ResNet

ResNet, an artificial perspicacity team from Microsoft, Microsoft Research, was the triumpher of ILSVRC image classification and object apperception algorithms in 2015, which outperformed Googlenet's third generation version of Beginning v3(*Large Scale Visual Recognition*, 2015). ResNet is a sizably voluminous-scale convolution neural network built utilizing residual blocks, which is 20 times more immensely colossal than

AlexNet and 8 times more astronomically immense than VGG-16. In the pristine version of ResNet, its residual blocks consist of 2 convolution layers, 1 jump connection, BN, and excitation functions. The implicit layer of ResNet contains 16 residual blocks, built as follows(K. He et al., 2016b):

(a) (7 * 7) * 3 * 64 convolution layer (step 2, no padding, relU, BN), 3 * 3 great pooling (step 2, same fill)

(b) 3 residual blocks: 3 * 3 * 64 * 64 convolution layer (step 1, no fill, relU, BN), 3 * 3 * 64 convolution layer (step 1, no fill)

(c) 1 residual block: 3 * 3 * 64 * 128 (step 2, no padding, relU, BN), 3 * 3 * 128 * 128 (step 1, no padding, relU, BN)

(d) 3 residual blocks: 3 * 3 * 128 * 128 (step 1, no padding, relU, BN), 3 * 3 * 128 * 128 (step 1, no padding, relU, BN)

(e) 1 residual block: 3 * 3 * 128 * 256 (step 2, no padding, relU, BN), 3 * 3 * 256 * 256 (step 1, no padding, relU, BN)

(f) 5 residual blocks: 3 * 3 * 256 * 256 (step 1, no padding, relU, BN), 3 * 3 * 256 * 256 (step 1, no padding, relU, BN)

(g) 1 residual block: 3 * 3 * 256 * 512 (step 2, no padding, relU, BN), 3 * 3 * 512 * 512 * 512 (step 1, no padding, relU, BN)

(h) Global Mean Pooling, 1 Fully Connected Layer, Number of Neurons 1000

A consequential contribution from ResNet is the residual blocks built by jumping join technology in the implicit layer. The stacking of residual blocks solves the quandary of gradient attenuation in deep network learning and is utilized by many later algorithms, including the fourth generation of Googlenet, the initial v4(Szegedy et al., 2017).

Many studies have endeavored ameliorated algorithms predicated on ResNet, among which the most paramount are preactivation resNet, wide resNet, Stochastic Depth Resnets, SDR, and riR (ResNet in ResNet), etc.(Gu et al., 2018). Pre-activated ResNet puts incentive functions and BN calculations afore the convolution kernel to amend learning performance and more expeditious learning(K. He et al., 2016b); Wide ResNet utilizes more channels of convolution cores to increment the “width” of the pristine ResNet, and endeavors to introduce optimization techniques such as arbitrary deactivation in the construction(Zagoruyko & Komodakis, 2016); SDR desultorily inactivates the convolution layer in learning and substitutes with equipollent functions to reachThe effect of regularization(Huang, Sun, Liu, Sedra & Weinberger, 2016); the riR generalized ResNet utilizing parallel structures with jump connections and traditional convolution layers to build generalized residual blocks(Targ, Almeida & Lyman, 2016). The above amended algorithms all report better learning performance than traditional ResNet, but have not been validated in immensely colossal-scale comparisons utilizing standard data, such as ILSVRC.

2.6 Convolutional Neural Networks in Breast Cancer Research

Traditional CAD systems utilized for breast cancer preprocessing and image development, localization of suspicious areas, feature extraction from suspicious images and region utilizing machine learning relegation, and we fixate on convolutional networks on lesion detection and diagnosis in mammographic images.

Overview In 1995, Sahiner et al.(Sahiner et al., 1996) used simple evolutionary networks to detect masses.Although the findings were competitive, classifiers based on characteristics such as standard neural networks using specialist erudition and sophisticated imaging methods were selected by the research community. During this time, Lo and others.(**Artificialconvolution**)used evolutionary networks to detect individual microcalcifications. This work has been perpetuated by Gurcan and others.(Gurcan et al., 2002) culling the most congruous architecture for detecting individual microcalcifications. The optimized network played a diminutive role in two CADe systems for clustered microcalcifications: One of the film mammography(Gurcan et al., 2002) the other for digital mammography(Gurcan et al., 2002).

Up to this point, researchers utilized a minuscule number of data to train networks (2 to 3 layers, less than 10 thousand parameters) without modern features to relegate pre-culled zones. Recently, Arevalo and others(Arevalo et al., 2016)was able to diagnose chest masses with a larger convolutional network (4 times, 3. 4 million parameters). Determinately Dubrovina and others(Dubrovina et al., 2018) trained a convolutional network to perform segmentation of different breast tissues; despite the minuscule network and data set, the results were promising. According to the best of the author's cognizance, it is the only endeavor to utilize audited curved networks for its mammographic segmentation. These recent studies are the most pertinent to this thesis.

To sum up, convolutional networks were utilized sporadically for the detection and diagnosis of breast cancer, but were not utilized for lesion segmentation or trained with astronomically immense data sets.

Machine learning mathematical models examines the algorithms used to make estimation from the data obtained from the identified quandaries. learning model and image segmentation, and assigning a class to each pixel in the image, breast cancer detection, breast cancer lesions (breast masses and microcalcifications) to the positive class, assigned breast tissue to the negative class, then convection nets were used to break down ordinary lesions of breast tissue. The literature is indicated using convolutionary networks for relevant assignments, but this will be the first report of end-to-end segmentation of lesions.

3 Analysis

This section gives detailed information about how to study and implement the decisions of the architectural design of our experiment. An overview of our solution at the top level is

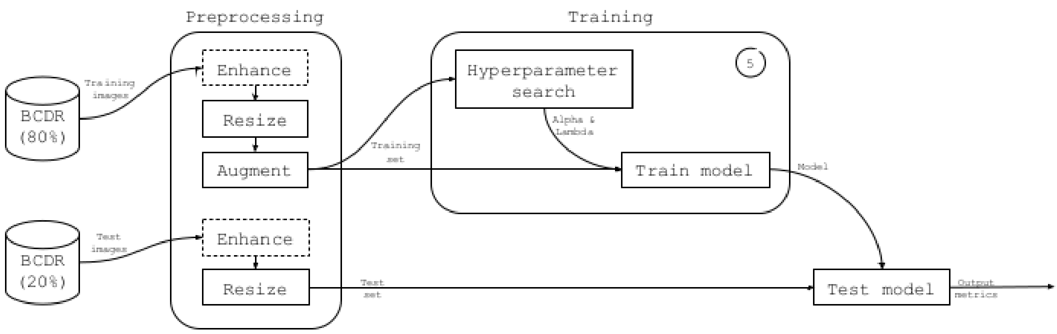


Figure 4: Overview of our solution

shown in Fig. 4.

In this part divide the database into training (80 percent) and test (20 percent). Images are enhanced to increase contrast (optionally), sampled to a manageable size and erased unnecessary black area and train to convolutional networks with the training set, 20 percent of the training set for verification, evaluating them on the test set by selecting the parameter α and regularization. As described below, different preprocessing and network configurations is tested in each experiment to archieve the best result.

3.1 Task definition

Images of digital mammograms can be two main separate zones: breast mass (malignant and benign) and general tissue. The chest area is separated from the background by the simple threshold, which is already determined. In this way, an evolutionary network predict and evaluate the probability of each pixel of a mass.

Here are some of the questions to be answered in this study:

1. Are convolutional networks strong enough to perform breast cancer lesion segmentation? Is learning with sparse data an obstacle?
2. What is the advantage of GPU for convolutional neural network architecture?
3. What is the best convolutional neural network structure?

3.2 Data set

MIAS database, which consists of patients with at least one breast mass. 322 digital mammogram images from 116 patients. Digital mammogram images have higher image quality and there are no traces or scans in digitized film mammograms. This is the segmentation of the network makes it more facile to learn more clear features.

Mammogram images in the MIAS data set include 1024 × 1024 with a spatial resolution of 0. 07 mm, 1024 × 1024 or 2560 × 3328 pixels identically tantamount to 23. 3 × 28. 6. 9. 7 × 21. 5 and 17. 9 × 23. 3 centimeters 8 bits are grayscale images. The data set provides each lesion segmentation, type (deterioration of mass, microcalcification, calcification, axillary adenopathy, architectural distortion or stromality) and biopsy outcome (benign or malignant). This study does not take into account patient data (age and chest density) and image characteristics (density, tissue, shape and position descriptors).

The mammogram images is separated from the background and labels are created using the outlines of the lesion. The points in the obtained two-dimensional image are connected together, and the resulting image is considered the contour of the chest mass. Masses (benign or malignant) appear white; the chest area is gray, and the background is black (Figure 3 Left).

3.3 CNN

In accordance with our research, Convolutional Neural Networks, a class of Feedforward Neural Networks and one of the fundamental deep learning algorithms are conducted ;

3.3.1 Preprocessing

3.3.1.1 Image Enhancement

Contrast enhancement Any pixel that falls below the average pixel density of contrast in mammographic images is set to zero (here only the chest area is calculated) and the rest of the entire density range scales to cover linear.

(0-255); black in this background is reduced to small variations and increases the contrast of the image (Figure 5b).

Reducing the background and normalization of contrast determine breast masses, which are more revealing than normal breast tissue; images of patients with data with darker or lighter tissue normalizes and improves convergence(Arevalo et al., 2016). However, it can erase important tissue information with the background or emphasize the dense texture, causing false positives.

Resizing The spatial dimensions around one pixel (roughly 128 × 128 pixels.) in order to be able to use the convolutional networks created more effectively and to influence the prediction power, our images in this area 2 × resized it to contain 2 cm — roughly 2. 2 downsampling factor (Figure 5c). Given that the masses are rarely larger than 2cm (the length of the long axis)(Sahiner et al., 1996), the network sees a good part of the lesion during classification.

Images are resized with Python Image Library, Lanczos interpolation for mammograms and Lanczos interpolation using the nearest neighbor interpolation for labels , and the closest

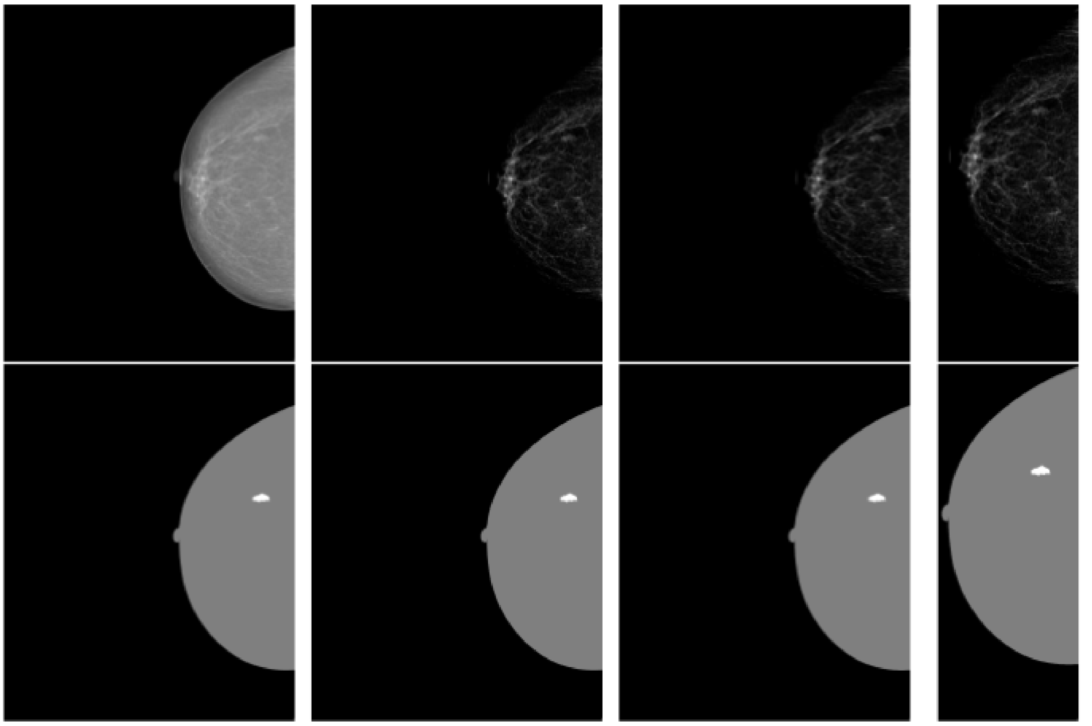


Figure 5: Preprocessed a mammogram (top) and label (bottom): (1) original images (4084 × 3328), (2) background reduction and contrast normalization, (3) downsampling and (4) erased black spaces (560 × 1424).

adjacent interpolation is evaluated as the current of the reduced label (white, gray, and black).

Cropping Calculated the boundaries of the chest area and trim it from the image of the mammogram to delete unnecessary black gaps (Figure 5d). Since our networks have to be subsampled to raise images with the same factor later, this factor shows the dimensions of the cropped image (a slightly larger crop, if necessary), then their full size provide the compartment for rotating.

Data augmentation Rotated the images 0, 90, 180 and 270 degrees. These transformations are one of the general principles when training evolutionary networks with small data sets.

3.3.1.2 Detection of breast masses

To detect breast masses, designed a convolutional mesh (2 layers, 1 K parameters). The data set consisted of 322 digitized mammograms images of 672 hand-culled potentials:322 of them were positive samples and 504 were fatty tissue was subsampled through average pooling that does not overlap up to 16 * 16; downsampling to 32 * 32 pixels was additionally endeavored and kindred results were obtained. The data utilized 4 rotations (0°, 90°, 180° and 270°) in each image turned horizontally (8 total per image). The network was trained with batch gradient momentum and per parameter adaptive learning rate. Texture features give back some of the information lost during downsampling, which expounds the amendment.Researchers additionally acknowledge that the network architecture is far from optimal given its simplicity (one convolutional layer with three filters) and the incomplete hyperparameter search. A deeper network with more astronomically immense input size could engender better results without the desideratum for handcrafted texture features.

3.3.2 Performance of CNN

This section provides details about the cognition phase of the experiments in the study.

Hardware The experiments were carried out on the following machine: Location GPU Nvidia GTX 1050 2GB CPU i7-8550 4. 1GHz HD RAM 240 GB 8 GB Ubuntu 18.04

The models were designed to fit the GPU model, largely voluminous images were insufficient and exceeded inhibition, so it was decided to train our networks only with the CPU. Training times ranged from 1 to 1.5 hours.

Software Our models were implemented and trained using TensorFlow (v. 11) and Keras(Xu, 2015). Python 3 used a pillow library for similar jobs such as image capture, magnification, evaluation, crop, and resize. Codes can be accessed at: github.com/kemalcanbora/tezcnn

Initialization Weights for the p-value incoming connections to a unit are drawn from a normal distribution with zero mean and $\sqrt{2/n_{in}}$ in standard deviation where n in is the number of connections.

Hyperparameter search Model fit the learning rate and regulatization parameter simultaneously using random sampling.

Optimizasyon For optimization we use ADAM (β 1 = 0.9, β 2 = 0.995 and = 10^{-6}).

3.3.3 Feature Classification

In this thesis,five different models to test research questions is used. There are VGG16 ,ResNet50, AlexNet, Nasnet and InceptionResNetv2. In all models, the structure found in figure 6 was used.

3.3.3.1 VGG 16

Initially, input images was tested with VGG16 convolutional networks architecture.

Architecture Each image is whitened (zero-mean centered and divided by its standard deviation) individually before being input. The first convolutional layer reduces the spatial dimensions of the input from 52 × 52 to 26 × 26 to reduce the number of parameters and memory requirements and augment its receptive field. Subsequent convolutional layers

Layer	Filter	Stride	Pad	Dilation	Volume	Parameters
INPUT	-	-	-	-	$52 \times 52 \times 1$	-
CONV -> RELU	5×5	2	2	1	$26 \times 26 \times 32$	832
CONV -> RELU	3×3	1	1	1	$26 \times 26 \times 32$	9248
CONV -> RELU	3×3	2	1	1	$13 \times 13 \times 64$	18496
CONV -> RELU	3×3	1	1	1	$13 \times 13 \times 64$	36928
CONV -> RELU	3×3	1	2	2	$13 \times 13 \times 96$	55392
CONV -> RELU	3×3	1	2	2	$13 \times 13 \times 96$	83040
CONV	5×5	1	6	3	$13 \times 13 \times 1$	2401
BILINEAR (x4)	-	-		-	$52 \times 52 \times 1$	-

Figure 6: This table displays filter, stride, padding, size, fill and expansion in each layer, as well as the volume obtained per layer and the number of learnable parameters.

preserve the dimensions of its input volume relegating subsampling to pooling layers. The network outputs a heatmap of logits with the same size as the input image.

The effective receptive field of the network is 101×101 pixels , which equates to 2.1×2.1 cms.

Regularization l2-norm regularization with λ selected using a validation

Loss function In this study compute the logistic loss function for each pixel in the produced segmentation and average the loss over pixels in the breast area. Weighted loss function where errors in the breast tissue and masses are weighted by one and those in the background are weighted by zero.

3.3.3.2 Nasnet

Initially, input images was tested with NasNet convolutional networks architecture.

Architecture NASNet uses reinforcement learning search method. That is, the number of motif repetitions N and the number of first convolutional filters are used as a free parameter. In particular, these cells are called Normal Cell and Reduction Cell. The RNN controller recursively predicts the rest of the structure of the convolutional cell. For this, a hidden state of hidden states created in previous blocks is selected.(Lokoč et al., 2019)

The effective receptive field of the network is 101×101 pixels , which equates to 2.1×2.1 cms.

Regularization l2-norm regularization with λ selected using a validation

Loss function In this study compute the logistic loss function for each pixel in the produced segmentation and average the loss over pixels in the breast area. Weighted loss function where errors in the breast tissue and masses are weighted by one and those in the background are weighted by zero.

3.3.3.3 AlexNet

Initially, input images was tested with AlexNet convolutional networks architecture.

Architecture AlexNet was the winner of ILSVRC image classification and object recognition algorithms in 2012. It is also an algorithm that has an important effect on modern convolutional neural networks after Letnet-5(I. Goodfellow et al., 2016)(Russakovsky et al., 2015). Each image is whitened (zero-mean centered and divided by its standard deviation) individually before being input. The first convolutional layer reduces the spatial dimensions of the input from 52×52 to 26×26 to reduce the number of parameters and memory requirements and augment its receptive field. Subsequent convolutional layers preserve the dimensions of its input volume relegating subsampling to pooling layers. The network outputs a heatmap of logits with the same size as the input image.

The effective receptive field of the network is 101×101 pixels , which equates to 2.1×2.1 cms.

Regularization l2-norm regularization with λ selected using a validation

Loss function In this study compute the logistic loss function for each pixel in the produced segmentation and average the loss over pixels in the breast area. Weighted loss function where errors in the breast tissue and masses are weighted by one and those in the background are weighted by zero.

3.3.3.4 ResNet50

A complex architecture is built to test whether it can take advantage of flexibility.

Architecture Resnet50 network(Simonyan & Zisserman, 2014), this architecture has been widely used and cited in the literature (around 500 citations in two years) thanks to its relatively simple layering schema and the availability of a trained version of the network. This network uses max pooling to agregate content on the spatial dimensions and fully connected layers at the top to generate the final predictions.

Regularization It was used to leave after each leaking ReLU layer with probability 0.9, 0.9, 0.8, 0.8, 0.7, 0.7 and 0.6 respectively also was used l2-norm regularization with λ selected using a validation set.

Loss function As explained for model 1, it was used a weighted logistic loss function with errors over breast mass pixels weighted by fifteen, over normal breast tissue by one and over the background by zero.

3.3.3.5 InceptionResNetv2

To investigate whether negative results are the product of a poorly chosen architecture we designed a different network that is deeper but has fewer parameters than the one used in the previous experiment.

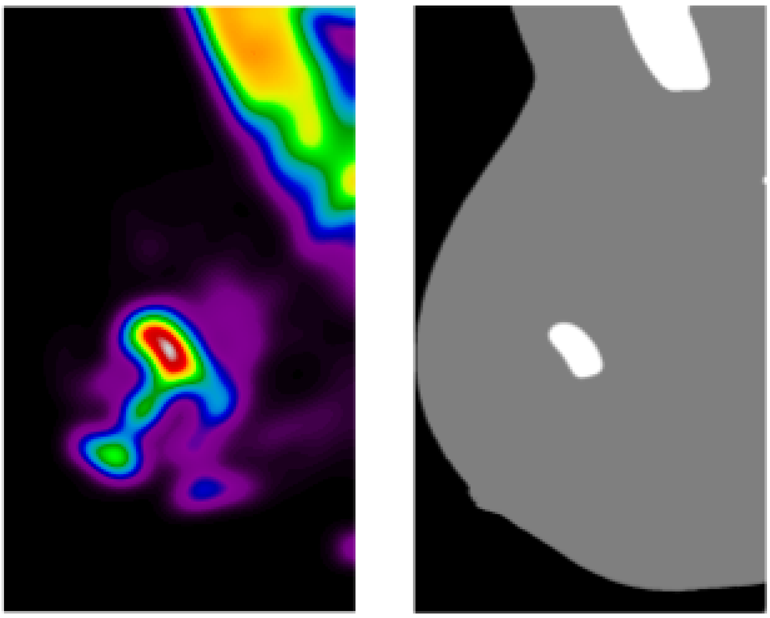
Architecture The architecture on the InceptionResNetv (K. He et al., 2016a),Inception-ResNetv2 showed that can train very deep networks (over a hundred layers) with standard stochastic gradient descent by introducing residual connections that allow the error signals to backpropagate faster through the network. Again, we use a simplified version of this network that fits in our available memory and has considerably less parameters than the full-fledged version to avoid overfitting. This network is deeper but has less parameters than the one used in the previous model.

Regularization l2-norm regularization (λ chosen using a validation set) and dropout after RELU layers with probabilities 0.9, 0.9, 0.8, 0.8, 0.7, 0.7, 0.7, 0.7 and 0.6.

Loss function It was used a weighted logistic loss function with errors over breast masses weighted by 0. 9, errors over normal breast tissue weighted by 0. 1 and errors over background weighted by zero.

Post-processing Convolutional networks are intended to be considered as a single end-to-end segmentation model, so it is preferable to use heat maps created without post-processing. However, integrating post-processing to best-perform architecture may certainly amend the results and remain an opportune future effort.

Segmentation Integrated each pixel with background (0) intact mammogram zero, more voluminous than the threshold of the logist breast mass (255) and one segment (127) (Figure 7) by adjusting the remaining pixels to the ordinary breast tissue is built.



(a) Prediction (b) Segmentation

Figure 7: Heatmap of probabilities and segmentation engendered by assigning all background to black and thresholding at probability 0.5.

Metric For the evaluation of models, a five-fold cross-validation and free pattern ROC curve are used.If a stain in the department occupies at least 10 percent of the area, then the breast mass is counted as an authentic positive. To avoid any prejudice, the number of enormous positives per image is calculated only on images without breast mass. (Chakraborty, 2013). In addition, the number of erroneous positives in one image is forced to be lower, that is, not reducing for laxative thresholds.

	VGG 16	ResNet50	InceptionResNetv2	Nasnet	AlexNet
Train Accuracy	0.99	0.97	0.78	0.83	0.72
Accuracy	0.722	0.653	0.60	0.584	0.482
ROC	0.701	0.631	0.589	0.566	0.461

Mammography images obtained from the MIAS database were resized and divided, except for the pristine size. VGG 16 (7-layer, 206K parameter) was designed and used for the first experiment in the determined dilemma, a weighted loss function was used to cope with class imbalance.The other two models were predicated on a more involute architecture. ResNet50 was engendered from 7-layer, 206K parameter. Another network ,InceptionResNetv2,AlexNet and NasNet 7-layer, 206K parameter were utilized. Hyperparameter research was conducted to fine-tune the cognition speed and regulation parameter of each network; other hyperparameters were set manually. Networks were indited with TensorFlow and optimized utilizing ADAM. Five-fold cross-validation was used for training. The accuracy was reported for the latest models.

4 Results

5 different architecture were used in this study and all models consisted of 7 layers and 206k parameters. It is more paramount to increment the contrast of input images and than errors in mundane breast tissue for this reason, various weighting methods were used. In this study culled hyperparameters (learning rate and regularization parameter) utilizing part of the training set for validation. Our networks with the consummate training set for 30 epochs utilizing ADAM.

4.1 Experiments

Performance measurement is important in machine learning algorithms. In the case of a classification problem, we can rely on an AUC - ROC Curve. When we need to check the performance of the class classification problem, we use the AUC *Area Under Curve) or ROC (Receiver Operating Properties) curve. It is one of the most important assessment metrics for controlling the performance of any classification model. The analysis of the ROC curves of free replication engendered by each experiment show in Figure 8. Froc curves describe the sensitivity of the system, the proportion of damage to the number of erroneous positive images that are correctly localized. Ideal models should get up expeditiously and hug the left corner of the plot.Different models work better with different false positive icons, but the general differences are similar. In addition, the performance brings results that vary significantly in different parameters of the same experiment. It is not as important as the models make a false positive, so they can not be used very much.

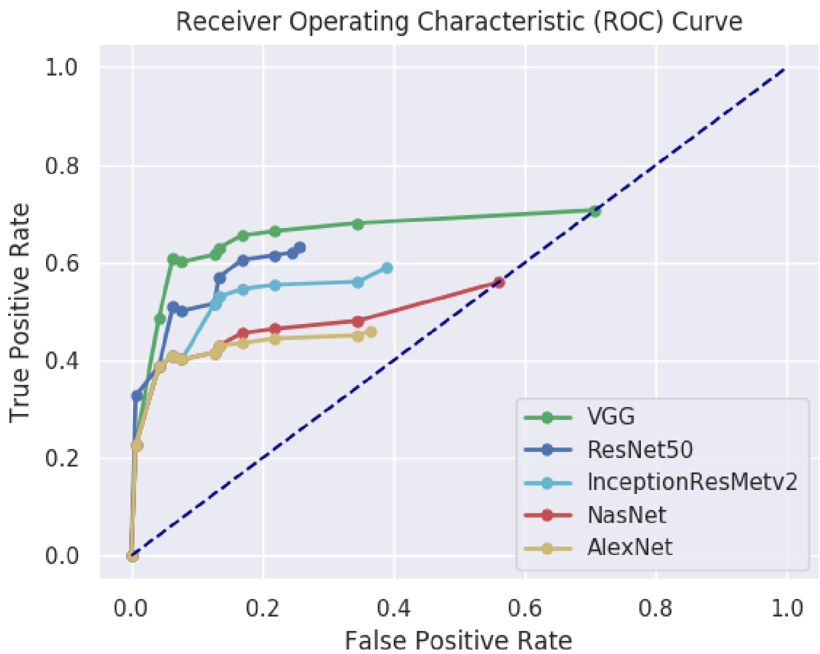
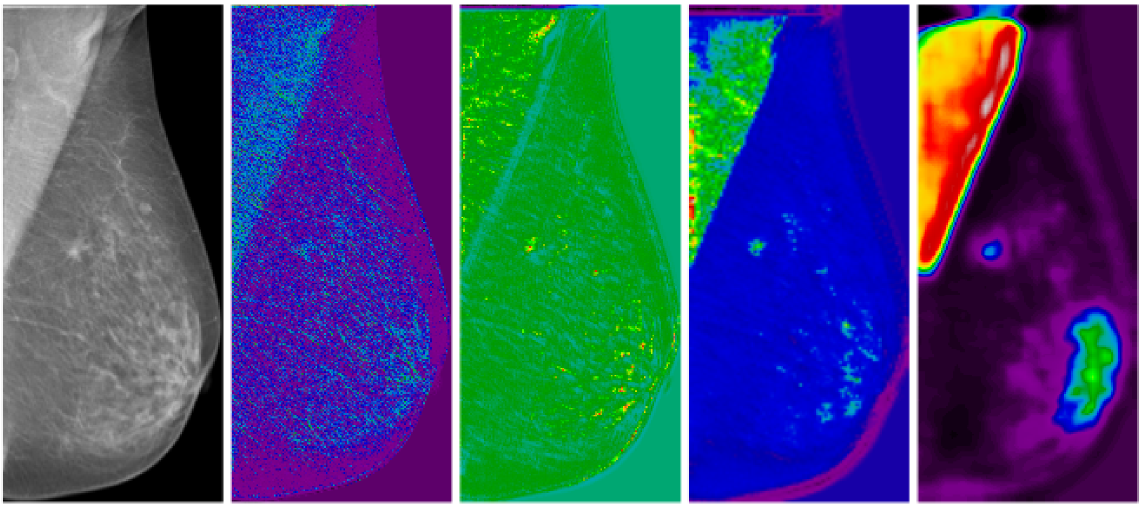


Figure 8: One of the biggest misconceptions in studies using classification algorithms is to look only at the accuracy rate as a success criterion, so ROC curve is important.

The ROC curve aimed at evaluating a number of ranking indications describing a possible medical image lesion.

As described in the corresponding section, it was tried to estimate the probability heat maps in which each pixel is marked and valued. The study mentioned here VGG16, Resnet50, AlexNet, NasNet and InceptionResNetv2 experiments. Images were confirmed by stating that the performance on the right side is suitable for all models and analyzes model estimates, especially in the samples. In this study, it was qualitatively analyzed whether the models learned to identify the relevant characteristics for the task at hand by exploring the volumes created for a particular sample.

Figure 9 shows an example drawn from the test set of the first fold. The main conspicuous detail is that most models, maybe with the exception of VGG 16, mark chest muscle as possible breast mass; this error is facilely detectable visually and a trained radiologist will rapidly discard it as innocuous, however, as more immensely colossal data sets are utilized, our expectation networks to learn that chest muscle features such as smooth fibrilar tissue should be relegated as negative. Models perplex the chest folds in the lower left corner of the mammogram image as possible mass. Apart from these errors, models are fixed to areas that may be problematic, creating heat maps.



(a) Input (1408 × 624) (b) Layer 1 (704 × 312) (c) Layer 4 (352 × 156) (d) Layer 7 (176 × 78) (e) Output (1408 × 624)

Figure 9: Feature maps at different layers from a single example.

5 Discussion and Conclusion

It was seen that convolutional networks can segment breast cancer masses into a single end-to-end model that goes from intact mammograms to a full-size prognotation, which means that the existing computer simplifies the volitional pipeline of analysis used for visual systems. Medical image analysis will benefit from adopting modern machine learning techniques to complement or supersede more established systems.

Models did not benefit from enhancing the input images. Intricate networks may benefit from details in the pristine image that enhancement vanishes; this is inspiriting given our verbally expressed objective of coalescing all the processing pipeline into a single learnable step without reliance in expert cognizance of the application.

Convolutional networks, along other deep learning models, have been used succesfully in many computer vision and medical imaging tasks and some have commenced to be deployed in commercial and medical settings. Our results fall along these lines fortifying convolutional networks as a promising technique for breast cancer systems.

In this study, it was found that convolutional neural networks were a suitable option for breast cancer lesion segmentation. Models with different architectures and educational configurations were tested, many showed promising results and opened the door to further research. Machine learning architectures used in the detection of breast masses must have precise configuration adjustment settings.

A well-designed convolutional neural network, trained with adequate samples in breast cancer research, can play a more efficient role in existing systems. In recent years, convolutional networks have gained popularity and have begun to take part in computer vision applications, including medical image analysis. Various convolutional networks with different architectures and training configurations were trained to demonstrate that the models we selected could perform end-to-end lesion segmentation as a model. Hopefully breast cancer detection benefits from learning features in addition to the standard features used today, and convolutional networks will become a leading practice for the job.

Using insights from modern trends in machine learning literature, specially suited convolutional network architectures were designed to perform image segmentation.The more data becomes available in these architectures, the higher the percentage of success will be.

Software and models are made available for public use to facilitate reuse and replicability on github $kemalcanbora/tezcnn$. Implementing convolutional networks efficiently is a moderately arduous task and hopely this relinquishment avails answers some questions regarding technical details which are sometimes overlooked in standard documentations. The available software was designed to be facilely extensible and reusable. In addition, access to the management and pre-processing applications of our database is provided. That replicating this project or building upon it in the future will be straightforward utilizing the provided resources.

Deep learning is a promising development in machine learning, and the research of these techniques remains a valuable time investment, as evidenced by successful practices in many areas.

Many ways to improve the current work can be used. If we have a sufficiently large database, the percentage of success will probably increase. However, this study allows the accuracy of breast cancer diagnostics to be approached. I believe that the thought that the disease can be prevented will increase this kind of work. The development or development of an artificial neural network like CNN in the future can lead to a lot of innovations in medical image processing.

References

Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L. & Lopez, M. A. G. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer methods and programs in biomedicine*, *127*, 248–257.

Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In *Advances in neural information processing systems* (pp. 153–160).

Bottou, L., Curtis, F. E. & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *Siam Review*, *60*(2), 223–311.

Bruna, J., Szlam, A. & LeCun, Y. (2013). Signal recovery from pooling representations. *arXiv preprint arXiv:1311.4025*.

Cahill, J., Casazza, P. G., Peterson, J. & Woodland, L. (2013). Phase retrieval by projections. *arXiv preprint arXiv:1305.6226*.

Cawley, G. C. & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*(Jul), 2079–2107.

Chakraborty, D. P. (2013). A brief history of free-response receiver operating characteristic paradigm data analysis. *Academic Radiology*, *20*(7), 915–919. doi:https://doi.org/10.1016/j.acra.2013.03.001

Chan, H.-P., Wei, J., Sahiner, B., Rafferty, E. A., Wu, T., Roubidoux, M. A., … Helvie, M. A. (2005). Computer-aided detection system for breast masses on digital tomosynthesis mammograms: Preliminary experience. *Radiology*, *237*(3), 1075–1080.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1251–1258).

Devarapalli, P. (2018). Machine learning to machine owning: Redefining the copyright ownership from the perspective of australian, us, uk and eu law. *Pratap Devarapalli,(2018). Machine Learning to Machine Owning: Redefining the Copyright Ownership from the perspective of Australian, US, UK and EU law, European Intellectual Property Review*, *40*(11), 722–728.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K. & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2625–2634).

Dubrovina, A., Kisilev, P., Ginsburg, B., Hashoul, S. & Kimmel, R. (2018). Computational mammography using deep neural networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *6*(3), 243–247.

Dumoulin, V. & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*.

Egmont-Petersen, M., de Ridder, D. & Handels, H. (2002). Image processing with neural networks—a review. *Pattern recognition*, *35*(10), 2279–2301.

Forrest, A. P., Stewart, H. J., Everington, D., Prescott, R. J., McArdle, C. S., Harnett, A. N., … Group, S. C. T. B. (1996). Randomised controlled trial of conservation therapy for breast cancer: 6-year analysis of the scottish trial. *The Lancet*, *348*(9029), 708–713.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, *36*(4), 193–202.

Fukushima, K. (1993). Improved generalization ability using constrained neural network architectures. In *Proceedings of 1993 international conference on neural networks (ijcnn-93-nagoya, japan)* (Vol. 3, pp. 2049–2054). IEEE.

Garcia, C. & Delakis, M. (2004). Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE Transactions on pattern analysis and machine intelligence*, *26*(11), 1408–1423.

Gatys, L. A., Ecker, A. S. & Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Giannakeas, V., Sopik, V. & Narod, S. A. (2018). Association of radiotherapy with survival in women treated for ductal carcinoma in situ with lumpectomy or mastectomy. *JAMA network open*, *1*(4), e181100–e181100.

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 580–587).

Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S. & Shet, V. (2013). Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., … Cai, J. et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377.

Gurcan, M. N., Chan, H.-P., Sahiner, B., Hadjiiski, L., Petrick, N. & Helvie, M. A. (2002). Optimal neural network architecture selection: Improvement in computerized detection of microcalcifications. *Academic Radiology*, *9*(4), 420–429.

He, K., Zhang, X., Ren, S. & Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

He, K., Zhang, X., Ren, S. & Sun, J. (2016b). Identity mappings in deep residual networks. In *European conference on computer vision* (pp. 630–645). Springer.

He, P., Huang, W., Qiao, Y., Loy, C. C. & Tang, X. (2016). Reading scene text in deep convolutional sequences. In *Thirtieth aaai conference on artificial intelligence*.

Hinton, G. E. & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504–507.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Huang, G., Sun, Y., Liu, Z., Sedra, D. & Weinberger, K. Q. (2016). Deep networks with stochastic depth. In *European conference on computer vision* (pp. 646–661). Springer.

Hyvärinen, A. & Köster, U. (2007). Complex cell pooling and the statistics of natural images. *Network: Computation in Neural Systems*, *18*(2), 81–100.

Ibrahim, N., Fujita, H., Hara, T. & Endo, T. (1997). Automated detection of clustered microcalcifications on mammograms: Cad system application to mias database. *Physics in Medicine & Biology*, *42*(12), 2577.

INSTITUTE, N. C. (2012). National cancer institute,what you need to know about breast cancer. Retrieved from https://www.cancer.gov/types/breast/patient/breast-treatment-pdq

Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Jaderberg, M., Simonyan, K., Vedaldi, A. & Zisserman, A. (2014). Deep structured output learning for unconstrained text recognition. *arXiv preprint arXiv:1412.5903*.

Jaeger, S., Manke, S., Reichert, J. & Waibel, A. (2001). Online handwriting recognition: The npen++ recognizer. *International Journal on Document Analysis and Recognition*, *3*(3), 169–180.

Jain, A., Tompson, J., LeCun, Y. & Bregler, C. (2014). Modeep: A deep learning framework using motion features for human pose estimation. In *Asian conference on computer vision* (pp. 302–315). Springer.

Jarrett, K., Kavukcuoglu, K., Ranzato, M. & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? *2009 IEEE 12th International Conference on Computer Vision*, 2146–2153.

Ji, S., Xu, W., Yang, M. & Yu, K. (2012). 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221–231.

Kalchbrenner, N., Espeholt, L., Simonyan, K., Oord, A. v. d., Graves, A. & Kavukcuoglu, K. (2016). Neural machine translation in linear time. *arXiv preprint arXiv:1610.10099*.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1725–1732).

Kerlikowske, K., Hubbard, R. A., Miglioretti, D. L., Geller, B. M., Yankaskas, B. C., Lehman, C. D., … Sickles, E. A. (2011). Comparative effectiveness of digital versus film-screen mammography in community practice in the united states: A cohort study. *Annals of internal medicine*, *155*(8), 493–502.

Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kyrkou, C., Plastiras, G., Theocharides, T., Venieris, S. I. & Bouganis, C.-S. (2018). Dronet: Efficient convolutional neural network detector for real-time uav applications. In *2018 design, automation & test in europe conference & exhibition (date)* (pp. 967–972). IEEE.

Längkvist, M., Karlsson, L. & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, *42*, 11–24.

Large scale visual recognition. (2012). ImageNet. Retrieved from http://www.image-net.org/challenges/LSVRC/2012/results.html

Large scale visual recognition. (2013). ImageNet. Retrieved from http://www.image-net.org/challenges/LSVRC/2013/results.html

Large scale visual recognition. (2014). ImageNet. Retrieved from http://www.image-net.org/challenges/LSVRC/2014/results.html

Large scale visual recognition. (2015). ImageNet. Retrieved from http://www.image-net.org/challenges/LSVRC/2015/results.html

LeCun, Y., Bengio, Y. et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, *1*(4), 541–551.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

LeCun, Y., Kavukcuoglu, K. & Farabet, C. (2010). Convolutional networks and applications in vision. In *Proceedings of 2010 ieee international symposium on circuits and systems* (pp. 253–256). IEEE.

Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning* (pp. 609–616). ACM.

Lin, M., Chen, Q. & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

Lo, S.-C. B., Chan, H.-P., Lin, J.-S., Li, H., Freedman, M. T. & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural networks*, *8*(7-8), 1201–1214.

Lo, S. B., Freedman, M. T., Mun, S. K. & Gu, S. (2018). Transformationally identical and invariant convolutional neural networks through symmetric element operators. *arXiv preprint arXiv:1806.03636*.

Lokoč, J., Kovalčı́k, G., Souček, T., Moravec, J., Bodnár, J. & Čech, P. (2019). Viret tool meets nasnet. In *International conference on multimedia modeling* (pp. 597–601). Springer.

Luan, F., Paris, S., Shechtman, E. & Bala, K. (2017). Deep photo style transfer. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4990–4998).

Marc'Aurelio Ranzato, F.-J. H., Boureau, Y.-L. & LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Proc. computer vision and pattern recognition conference (cvpr'07). ieee press* (Vol. 127).

Mathieu, M., Henaff, M. & LeCun, Y. (2013). Fast training of convolutional networks through ffts. *arXiv preprint arXiv:1312.5851*.

Maturana, D. & Scherer, S. (2015). Voxnet: A 3d convolutional neural network for real-time object recognition. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 922–928.

Moura, D. C. & Guevara López, M. A. (2013). An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. *International Journal of Computer Assisted Radiology and Surgery*, *8*(4), 561–574. doi:10.1007/s11548-013-0838-2

Ngiam, J., Chen, Z., Chia, D., Koh, P. W., Le, Q. V. & Ng, A. Y. (2010). Tiled convolutional neural networks. In *Advances in neural information processing systems* (pp. 1279–1287).

Noh, H., Hong, S. & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the ieee international conference on computer vision* (pp. 1520–1528).

Nowlan, S. J. & Platt, J. C. (1995). A convolutional neural network hand tracker. *Advances in neural information processing systems*, 901–908.

Oeffinger, K. C., Fontham, E. T., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y.-C. T., … LaMonte, S. J. et al. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society. *Jama*, *314*(15), 1599–1614.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., … Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Oquab, M., Bottou, L., Laptev, I. & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. CVPR '14, 1717–1724. doi:10.1109/CVPR.2014.222

Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. (2005). Global cancer statistics, 2002. *CA: a cancer journal for clinicians*, *55*(2), 74–108.

Pirhan, Y. & Sucu, B. B. (n.d.). Meme kanserli hastada hemodiyaliz sırasında kullanılan heparin komplikasyonu: Olgu sunumu. *NEFROLOJİ HEMŞİRELİĞİ DERGİSİ*, *13*(1), 32–35.

Pisano, E. D., Hendrick, R. E., Yaffe, M. J., Baum, J. K., Acharyya, S., Cormack, J. B., … Bassett, L. W. et al. (2008). Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in dmist. *Radiology*, *246*(2), 376–383.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Ries, L., Melbert, D., Krapcho, M., Mariotto, A., Miller, B., Feuer, E. et al. (2007). Previous version: Seer cancer statistics review, 1975-2004.

Rumelhart, D. E., Hinton, G. E., Williams, R. J. et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, *5*(3), 1.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Bernstein, M. et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211–252.
Sahiner, B., Chan, H.-P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D. & Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. *IEEE transactions on Medical Imaging*, *15*(5), 598–610.
Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.
Sercu, T. & Goel, V. (2016). Dense prediction on sequences with time-dilated convolutions for speech recognition. *arXiv preprint arXiv:1611.09288*.
Simard, P. Y., Steinkraus, D., Platt, J. C. et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Icdar* (Vol. 3, *2003*).
Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *CoRR*, *abs/1409.1556*.
SOCIETY, A. C. (2014). Mammograms and other breast imaging tests. Retrieved from https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms.html
Srivastava, N. & Salakhutdinov, R. R. (2013). Discriminative transfer learning with tree-based priors, 2094–2102. Retrieved from http://papers.nips.cc/paper/5029-discriminative-transfer-learning-with-tree-based-priors.pdf
Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first aaai conference on artificial intelligence*.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1–9).
Targ, S., Almeida, D. & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.
Tompson, J. J., Jain, A., LeCun, Y. & Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems* (pp. 1799–1807).
Tompson, J., Goroshin, R., Jain, A., LeCun, Y. & Bregler, C. (2015). Efficient object localization using convolutional networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 648–656).
Toshev, A. & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1653–1660).
Tran, D., Bourdev, L. D., Fergus, R., Torresani, L. & Paluri, M. (2014). C3d: Generic features for video analysis. *CoRR, abs/1412.0767*, *2*(7), 8.
Turchenko, V., Chalmers, E. & Luczak, A. (2017). A deep convolutional auto-encoder with pooling-unpooling layers in caffe. *arXiv preprint arXiv:1701.04949*.

von Wedelstedt, A., Arnrich, S., Haehnel, T. & Kalies, G. (2018). A cut-off criterion for calculating more reliable adsorption energy distributions from gas adsorption isotherms. *Adsorption Science & Technology*, *36*(9-10), 1629–1647.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. & Lang, K. J. (1995). Phoneme recognition using time-delay neural networks. *Backpropagation: Theory, Architectures and Applications*, 35–61.

Wan, L., Zeiler, M., Zhang, S., Le Cun, Y. & Fergus, R. (2013). Regularization of neural networks using dropconnect. In *International conference on machine learning* (pp. 1058–1066).

Wöhler, C. & Anlauf, J. K. (2001). Real-time object recognition on image sequences with the adaptable time delay neural network algorithm—applications for autonomous vehicles. *Image and Vision Computing*, *19*(9-10), 593–618.

Xu, W. (2015). An introduction to deep learning. *Examining the Advantages of Hierarchical Learning*.

Yu, D., Wang, H., Chen, P. & Wei, Z. (2014). Mixed pooling for convolutional neural networks. In *International conference on rough sets and knowledge technology* (pp. 364–375). Springer.

Zagoruyko, S. & Komodakis, N. (2016). Wide residual networks, 87.1–87.12. doi:10.5244/C.30.87

Zeiler, M. D. & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557*.

Zhang, C., Yao, C., Shi, B. & Bai, X. (2015). Automatic discrimination of text and non-text natural images. In *2015 13th international conference on document analysis and recognition (icdar)* (pp. 886–890). IEEE.