Gemstone Classification Using Convolutional Neural Network

Aaron Andrew Mwakifwamba

College of Software Engineering, Nankai University

ABSTRACT

Gemstone is very precious material. There are various classes of gemstone which are identified by physical means of human senses such as vision and tactile perception. Identifying gemstones by human senses suffers from human error and is extremely inefficient. Therefore, an automatic, efficient and accurate method of identifying gemstones is urgently needed. However, there is currently no related work to solve such geological problem that is based on the gemstone. In recent years, deep learning has been widely used in the field of image classification and has achieved impressive results. The success of deep learning in image classification inspires us to perform gemstone identification by convolutional neural network (CNN) which is good at feature extraction from image due to its three charac- teristics: local area connection, weight sharing and down-sampling. Vgg-16 has thirteen convolution layers and three full connected layer, and achieves good performance in image classification task. We adapt it in our gemstone classification problem. The experimental results demonstrate its effectiveness with test accuracy about 60%.

Key Words: convolutional neural network, deep learning, gemstone classification

1. Introduction

Gemstones these are minerals crystal which are well framed and polished form, mostly are used to make jewelry or other decorations.These precious stones mostly are transparent with delightful color features that gives the identification from its origin and makes easily to be classified. Lapidary is the way of cutting and polishing the gemstones to be used in jewelry industries.The lapidary works is based on colored gemstones that he gives out the shape and the cut of gemstones will have aiming to get the largest volume and the fit shape of the gemstone[1] . It is initially paper done based on gemstones multiclass classfication that is executed by using classification neural network; that has defined the gemstones into software detail using the deep learning manner of operations.These mineral crystal when they are cut and are in polished form is used to make jewelry or any other adornments. Meanwhile organic material that are not minerals such as pearl, jet, amber they also used for jewelry and they also been involved as a gemstone. Gemstone Multiclass is the problem of classifying instances into one of three or more classes according to their origin color and size or shape of the gemstone. Convolutional Neural Network is a type of deep neural network that use a variation of multilayer perception designed to require minimal preprocessing. CNNs use relatively little pre-processing compared to other image classi- fication algorithms, which means no previous feature extraction to the images is carried out[2]. Uniqueness of CNN from any other ordinary neural network is from the Inputs; that CNN take two dimensional array and operates directly on the image rather than focusing on the extraction which other neural network are based on. Convolutional neural network includes three main parts which are convolutional layer, pooling layer and fully connected layer. Never less the pooling layer has two calculation methods which are maximum pooling and average pooling, that they all perform the duty of reducing parameters while retain features to the maximum extent and speed up training while the fully connected layer is to expand the convolution pooled features into a long vector map to the output layer[3]. The convolution layer is the layer that extract the most merits features from an image it has several filters that perform the convolution operation and weighted evaluation process in which different convolution kernel performs inner product operation with corresponding pixel points, and then passes the results to the pooling layer by non-linear mapping. This layer reduces the training parameters of the network through weight sharing and local con- nection. Meanwhile the full connected layer is to integrate the features extracted from the stacked convolution layer and pooling layer so as to get high level meaning represented by the image features. That has to say to get the clear meaning of an image features the full connection layer operates in intermediates between the convolution layers and pooling layer to receive the impulses image from the two leading layers[4]. Convolutional neural network (CNN) is to identify two-dimension graphics. The network structure of convolution neural network is steady for translation, scaling, slanting or other forms of deformation. These characteristic are all relay to convolution neural network because it focus on different kind of features at each stage that it counters. At the first layer which is near to the original image, focused is on the pixel level then after it goes to multiple feature extraction and the consistency is close to the object itself. The convolutional neural network is applied not only in the limited to the field of image recognition but can also be applied in the face recognition, text recognition and other direction based on the detection. The Convolutional neural network is also one of the artificial neural network that it special way of image recog- nition and effective network with forward feedback. The merits of the convolutional neural network is to identify two-dimension graphics[7]. Canny algorithm is the edge detection algorithm which detects the edge of objects present in an image. Edge is the boarder the object and the background needed to be extracted, as the separation is done in the object and background, object and object or area and area; the accurate edge should be detected. The separation of object and background is the important concentration in face recognition, image enhancement. Most experiment have proved that canny algorithm gives out the best results in terms of handling images that are polluted by Gaussian white noise. The Canny Algorithm has come to the best results performance to a point of replacing Gaussian filter; its because can select the weight adaptively according to the features of gray values of the image and at the same time sharpening of the edge of the image[5].The Canny algorithm is known as optimal edge detection method. It operates on three main principles which are low error rate, well localization of edge point and one response to a single edge. There two techniques in Canny Algorithm which are non-maximum suppression and double thresh- olding to select the edge point. The double thresholding is used to segment the gradient image that are been set experimentally[5] . Some other times Canny algorithm cannot de- tect edges when the object has almost same color as background so the array edge will be zero valued; therefore its uses the original image[6].

1.2 Background, Scope and Motivation

The gemstone classification will be helpful on the identification of gemstone in a multiple time in many number of it need. Besides it provide the uniqueness of the gemstone considered labelling the images based on its class and is well performed by the Convolutional neural network algorithm. Theft, fraud or any other phish business will not easily be performed by any of the intruders. There will be easy way on a sophisticate structure of the gemstone easily be distinguished among the gemstones by the naked eyes. On the other hand the motive given is due to reduce if not to clear the problem of social engineers; these are people who perform the manipulation skills or techniques that make use of human error to obtain valuable data. Human error can be fear, anger, curiosity, anger, crime, sadness and excitement. Other tools used by social engineers are development of credibility, getting a person to like them, building trust, exploiting a person's desire to help. When it come on building trust these social engineers spend sometimes interacting with the person in an attempt to befriend to them or establish a serious familiarity then the social engineer ask for the information the person feels safe with the helping a friend or coworker, Phishing is a method involves the use of electronic communication designed to appear legitimate but which attempt to gather sensitive information from the user. Sometime the social engineers does not have to involve the use of technology. The only way to defend against social engi- neering is through training and the establishing of gemstone multiclass classification using convolution neural network and canny edge algorithm. The most efficieny computation method selected by convolutional neural network is max pooling and average pooling. Av- erage pooling obtains an excellent image classification accuracy. The max pooling obtains an excellent image classification accuracy. The max pooling method is succefully applied to train a deep 'convnet' for the image net competition. The Canny algorithm that oper- ate simultaneously with convolutional Neural Network it gives the best experimental result different with Gaussian filter in handling image affected with white noise. The inspira- tion come after knowing it's possible to perform image processing and identification that can extract the target features in the convolutional neural network. Convolutional Neural Network (CNN) models are well motivated by the human visual system that perform its information on various stages. The Convolutional neural network (CNN) and the biologic process of human visual system follow a simple to complex hierarchical structure; never less human brain is made of built using cells, and neural network is built with mathematical operations. The image processing is well performed by using convolutional neural network. The merits of CNN is mostly used for spatial data example images where as it takes fixed size of input and generates fixed size of outputs.

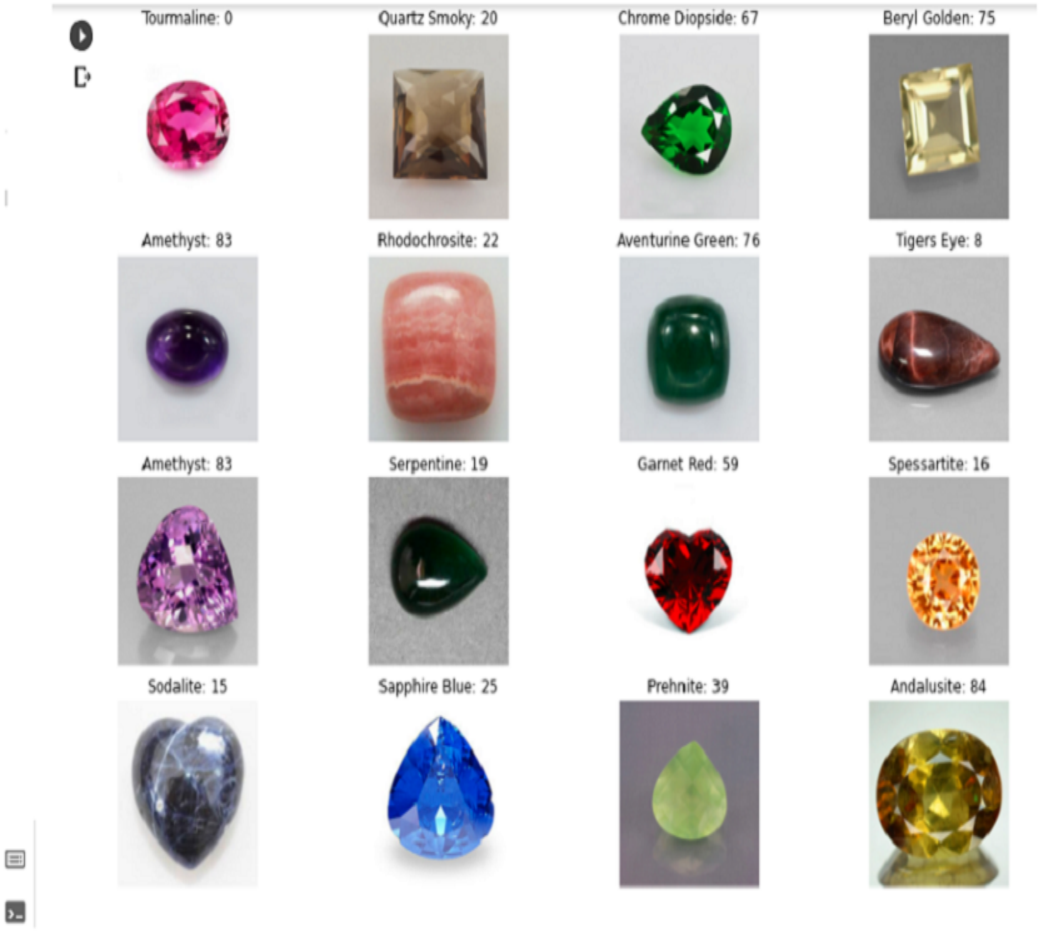

FIGURE 1.1: For GEMSTONE IMAGES 16 PLOTED IMAGE (4x4) in random format from the set and their labels.

2. Literature Review

On this section reviews literature related to classification of image multiclass on convolutional neural network (CNN).

2.1 Convolutional Nerual Network (CNN)

The work done by David Hubei and Wiesel in the 1958 [7] they performed a study of relationship between the pupil area and cerebral cortex in Johns Hopkins University. They took a cat's hindbrain bone and insertion of some electrodes in a hole of size 3mm to test the activity of the neurons. They also changed the position and angle of it when showing each object; later they discovered a unique type of neuron and name direction selective cell. The pupil will capture the edge of the object in the eyes and this edge point in a direction the neuron cell will be active. Displayed a cat and monkey visual cortexes that contain neurons that individual respond to small region of the region of the visual field. Given that the eyes are not moving at the region of visual space which affects visual stimuli on firing of single neuron that is called receptive field. The receptive field size and location make different systematically over the cortex to form a complete map of visual space. It took more development of character that 1968 their research paper identified two basic visual cell in human brain that are Simple cell; that has output is maximized by straight edge having specific orientation within their receptive field and thee Complex cells; that have huge receptive field and has output that has less concern to the exactly edge position in the field[8]. In 1980, Japanese scholar Kunihiko Fukushima introduce Neo-cognition that was inspired by Hubei and wisely.

The Neo-cognition introduced the two main layers of convolution neural networks (CNNs) which are convolutional layer and down sampling layer. A convolutional layers involves the units whose receptive field covers a patch of the previous layer. The content measure of these unit is called filter; the units can share filters. Down sampling layer pro- tects the units whose receptive field protects patches of previous convolutional layers. This unit completely performs the average of the activations of the units in its patch. The down sampling that helps to exactly classify objects in visual sense even though the objects are shifted [9].The convolutional Neural Network is mostly influenced by the human biological visual stimuli system. Way back in 1998[10]; the idea of convolutional neural network was well demonstrated by Lecun et.Al; that was demonstrating Convolutional neural network model can be successfully used for handwriting character recognition that made easily to sort the email using the machine and also it was automated to a point was used to read the address, and also was used by the banks to identify hand-written numbers on checks to avoid theft and fraud. Then after the researcher worked on further on the CNN model around 2012 [11] CNNs Alex Krizhevsky et.Al manage to develop a CNN called AlexNet architecture that was performing labelling picture or natural image such furniture and a per- son. His publication “ImageNet Classification with Deep Convolution Neural Network” made him win the award ILSVRC-2012(ImageNet Large Scale Visual Recognition Com- petition). Due to the biological visual stimuli recognition of human eye on how image formation takes place to the cortex in the human eye the scientists made a clear move to include more layers to generate more features in achievement of deep learning related to the performance of human eye in the cortex. So far the Convolutional Neural Network has accomplish the most exceptional performance in its operations on image segmentation, ob- ject detection and image classification. The figure below demonstrates convolutional neural network that possess of three major layer which are the input layer, the hidden layer and output layer.

The CNN architecture is possess of Convolutional layer for feature extraction that com- bine the integration of the two function and it shows you how one function modifies the other or modifies the shape. Convolutional layer output goes throw the activation func- tion that produces a well arranged images with a clear vision. Normalizations: The 2- dimensional outputs from kernel are called feature maps or activation map. It perform after convolutional layer is applied activation function is applied to 3 dimensional output.

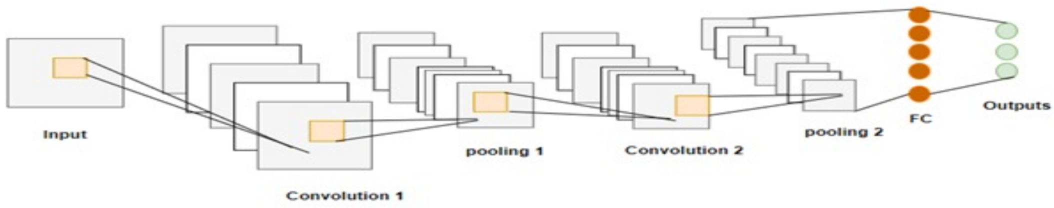

Figure 2.1: Convolutional Neural Network (CNN) Architecture.

The neurons in a layer mostly are organized in three dimensions which width, height and depth.
ingly the activation function that is applied is called ReLu or Rectified Linear unit Layer. Relu is additional step on convolutional. The ReLu rejects the negative values by setting a negative pixel value to 0. The ReLu is explained below:

$ReLU\ (x) = max(x, 0)$

Where by x is the input volume and ReLU layer replaces all negative pixel (Activation) val- ues in the feature map by zero. The Rectified feature map is also known as output features map.
Majid Nawez, Adel A. Sewissy [12] proposed a breast cancer multi classification frame- work using class structure based deep convolution neural network model. Whereas its class structure deep learning model has reaches its performance of accuracy on 93.3% on BreakHis dataset where as he managed to difference the cell and non-sample image by us- ing convolutional Neural network with different spatial patches whereas the classification accuracy was nearly estimated at 86.17% for 5x5 and 7x7 sub-window sizes respectively. In[12]the authors presented CNN classifier for visual analysis of ductal carcinoma tissue region in breast tumor images; where he proposed framework yielded higher with a very high performance compared to random forest classifier with 84.23% detection accuracy.
Pooling Pooling layer [16] is used to reduce the size of an image dimension which is to reduce the number of parameter and to understand the amount of computation performed in the network. The operation is done when making use two dimensional filter over each feature map lying within the region covered by filter and its exact location becomes less important. Keep in mind pooling operations reduces the spatial dimension of the input of the input volume in the case of 2D-pooling. Spatial module transforms the input image in a way of the subsequent layer makes easy way of on a classification; lather than making changes of the entire CNN architecture itself. The module performs the on the object is scaled and spatial attention to correct object in a crowded image. The pooling is associated by increasing the depth of feature maps and its almost done repeatedly until spatial losses it important content then later the dimensions are been flattened. The most merit of pooling is that it reduces the number of parameters to 75% and it control over fitting. The pooling method that we work with is max pooling where as its unit of performance is only sensitive to the surrounding maximum not to the specific location.

2.2 Type of Pooling
i. Max pooling : Is a pooling operations which estimates the values of each patches of a feature map and its used to create a down sampled which is also called pooled feature map; that is used after convolutional layer where by it translate the image by small amount where by it doesn't affect the most values of the pooled outputs.
ii. AveragePooling;Isapoolingoperationwhichestimatestheaveragevalueofpatches of a feature map and its used to create a down sampled which also called pooled feature map and its used after convolutional layer where by it translate the image by small amount where it doesn't affect the most values of the pooled outputs.
The Max pooling values mostly it use spatial of the size 2 x 2 and its step size of 2 to convolute.
The recent work[16],the main focus is understand the impact and response in pooling op- erations when the two Max pooling and average pooling is been mixed together, the results explains that max pooling or average pooling dominates the other never less they mixing of proportional parameters from the data is done. The mixing proportion parameter are consider into several sphere of a) per net, b) per layer, c) per layer/region pooled (but used for all channels across that region), d) per layer/channel (but used for all region in each channel) e) per layer/region/channel combination.
Dropout
Dropout: Is the way of conducting in neural network that helps in reducing interdependent learning among the neurons. Dropout term refers to dropping out units (hidden and visible) in neural network. The application of dropout can be done in hidden layers and can be well interpreted as a form of model arranging. Dropout is also well understandable as stochastic regularization technique. The dropout neural networks can be trained using stochastic gradient descent in a way similar to standard neural net. It known that dropout is a way used to avoid a model from overfitting. During backward and forward these neurons are not mostly considered when regularization and generalization are introducing a set of random selected neuron. [16] Regularization reduces overfitting by putting additional a penalty to the loss function whereas dropout forces a neural network to understand the robust features which is useful in conjunction in different random subset of other neurons. The number of iteration is increased twice than the one required never less the training time for each epoch is less.
In 2014 Nitish Srivastava presented a paper that explained droupout is been illustrated the million to hundreds numbers of parameters that a handled in machine learning system are been executed in deep neural networks which face the overfitting within the networks. With a lot of numbers of parameters within the deep neural network are difficult and slow in operation that makes some tough to perform overfitting when there is a consolidation of an expected massive neural nets with the oscillation time. On these view of performances dropout methods is been located in its performance it will drop the division together with their associated layer within the connection from the neural during training. It helps to avoid the excessive conversion in the main network. In separate network dropout performs segmentation with the increased number.

3: Methodology

3.1 Proposed Architecture
In 1980, Japanese scholar Kunihiko Fukushima introduce Neo-cognition that was inspired by Hubei and wisely. The Neo-cognition introduced the two main layers of convolution neural networks (CNNs) which are convolutional layer and down sampling layer.

A convolutional layers involves the units whose receptive field covers a patch of the previous layer. The content measure of these unit is called filter; the units can share filters. Down sampling layer protects the units whose receptive field protects patches of previous convo- lutional layers. This unit completely performs the average of the activations of the units in its patch. The down sampling operation is implemented in max pooling layer on the con- volutional layer output to deduct the dimension of the output neuron .The down sampling that helps to exactly classify objects in visual sense even though the objects are shifted.[9]. Where by the CNN is developed by input layer, a convolutional layer, a non-linear layer, a pooling layers, fully connected layer and an output layers. The large dataset of a lot hun- dreds to millions of images can be handled by using the multiple layer of neurons which are within the convolutional neural network. In training the CNN we make use of use of canny algorithm that is used for cropping edge of images, Adam algorithm which is using the optimizer to generalize the stochastic gradient descent (SGD). There will be no performance of traditional convolutional neural network CNN model. DenseNet is among CNN model which substitute non-linear convo- lutional layer and pooling layer associated with the denser block. The substitution does not involve the first layer it includes the transition layer which is the initial layer. DenseNet model contain three dense blocks and transition layer are shown below.
We proposed a dense block which includes convolution and non-linear layer. The opti-

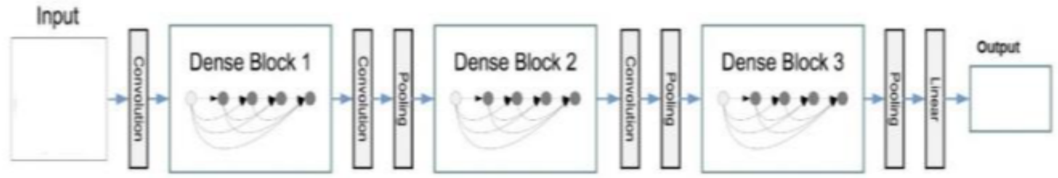

Figure 3.1: DenseNet with three dense block architecture.
mization techniques such as dropout and batch normalization are used to when it comes to dropout the hidden and visible layer in a neural network, this work can only be done in a hidden layer and the interpretation can be done in a form of model arranging. The previous layer output are concatenated alternatively of using summation in the dense block. Assume that the image displays BGR color format alternative to RGB color format and it because of how opencv reads and display which create the class name which display 16 (4x4) in a ran- dom images given from the set with their labels that has the label of [32, 64, 128] on three layers, where by the pixel taken at time (3x3) which is the kernel size that will transfer till the whole image is scanned completely, but for small image filters of size 1x1 the operates with the image of less small to 128x128. Where by the 2D convolutional layer and its input take three of its measurement input which is similar to the three color channels. Where the 2x2 is the maximum value move to the window by maxpool. Where the conv2D with hyper parameter are the initial way to the activation function for the every layer will undergo with Rectified Linear Unit (ReLu) whereas the spatial size from the features are reduced using maxpooling2D layer and the same procedure its done on the iteration in the kernel size of 32 → 64 → 128 → 128 → 128. The transformation in the inputs of multidimensional vector into a single dimensional vector or to change the complete pooled feature map matrix into a single column cube and then is transferred to further neural network execution. The dropout layer introduce in the hidden layer mostly with 0 from the inputs and decrease overfitting where the given Dropout (0.5).
Finally the full connected layer indicates the classes of gemstones 87 and the softmax.

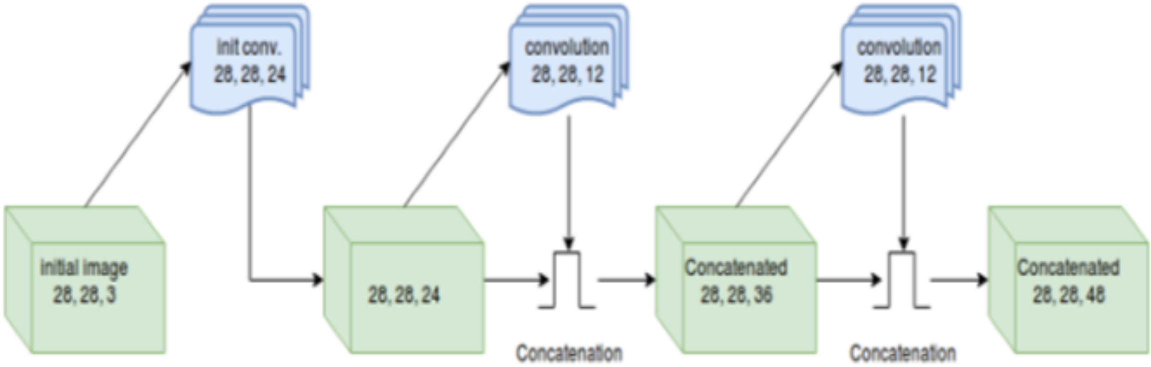

Figure 3.2: Full connected Layer illustration on it performance.
activation function provides the assumption from its distribution of a list of potential outcomes where: Dense (87) and Dense 512 nodes is received from the execution of the epoch.
The Dense scale operates to millions of layers when there is no problem on the optimization. In above mentioned that there will be remodel the Dense model that can manage to geological images to build the gemstone multiclass classification by the use of transfer learning. Our handcrafted made model is encouraged by Dense which has four of the dense blocks and three of the transition layer to give a clear classify gemstone multiclass.
Dense is design and built to solve natural image and non-microscopic images. For us to solve this situation we make use of kernel of 3 x 3 sizes on the initial convolutional layer to identify the small variation and solidity in the image in additional it isolate more important features. Never less the kernel size of convolutional layer in dense block is deducted so as to solve the complex structure of histopathology images. Pooling is also involves an average pooling layer with 13 x 13 kernel size and stride 2 are utilized earlier in the fully connected layer to solve the situation of feature map connected to the average pooling layer. In a long way with the pooling the configuration is done of softmax layer of eight classes of gemstone multiclass geological images meanwhile of the 1000 classes of the ImageNet dataset.
Transfer layer is defined as is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. Learning from geological image from scratch is mostly not done because of computational cost, blending problem, rare number of high quality label. We trigger the weight from various layer from the approved network by using pre trained model on ImageNet then after the final full connected layer is connected layer is updated always. The merits of Dense is the network where by the features are concatenation to assist to learn the feature in variety of stage without any demand to compress them in additional of capacity to handle and control with features manipulated.

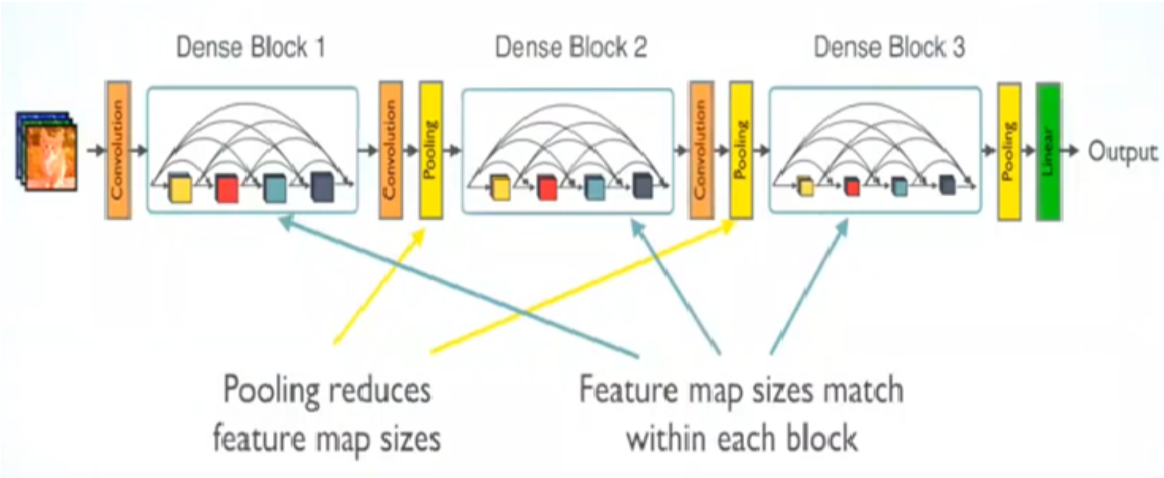

Figure 3.3: Dense with three dense block architecture.

3.1.1 Adam Optimization

The other Algorithm used to perform our model is Adam optimization algorithm that its implementation operations works as alternative of stochastic gradient falls on the other ap- proach of renewing the network weight that goes simultaneously with the training data. The name given to the algorithm is Adam was found during the time of evaluation which has the merits of computational efficient, little memory requirements, straightforward to implemen- tation, invariant to diagonal rescale of the gradients, well suited for problems that are large in terms of data and or parameters, appropriate for non-stationary objectives, appropriate for problems with very noisy /or sparse gradients and hyper-parameters that have natural interpretation and similar need of little tuning. The Adam algorithm it's not similar to stochastic gradient descent. Whereas the stochastic gradient descent keeps the initial learning rate to the entire weight which makes an improve and there is no change in learning rate at the training. These algorithm evaluates the separation of the adaptation in learning rate in separate parameters from the initial to the next moment of the gradient. Adam algorithm is merged with the merits of other stochastic gradient descent which are Adaptive Gradient algorithm and Root. Mean square propagation: Adaptive gradient algorithm the develop- ment of the production on some difficulties are kept within the per-parameter learning rate with the limited gradient.

3.1.2 Root Mean Square Propagation:

It keeps per-parameter learning rate which are the changing due to the mode of the magnitude of the current gradient for the weight; the demonstration of the algorithm is effi- ciently done online and non-stationary problem which is the noisy. The amount of change to the model during each step of this search process or step size is called "Learning rate" .The learning rate gives the merits of the hyperparameter in its implementation of neural network for gaining of better problem sorting.

Stochastic gradient descent is an optimization that estimates the errors gradient for the current state of the model using example from the training dataset, then updates the weights of the model using the back propagation of the errors algorithm. The impact of the learning rate in the neural network it performs on the inputs to the output from the given training of the dataset which whereby within the learning rate hyperparameter it manages the velocity capacity of the model learns, on the model will learn on the appropriate on the composed learning rate based on the given requirements in the various layers and its node in the epoch training and lastly on the rotation of training epoch and the derivative of the model where the learning rate is more loaded to the outcome of weight improvement to be implemented. The learning rate should be well balanced to avoid the increase or decrease during its per- formance on the epoch training dataset.

3.1.3 Pooling.

A convolutional neural network is designed by a structure of pooling layer and con- volutional layer. In every convolutional layer has aim to produce the illustration that will display the appearance of spatial structure and give a ken observation on multiple channel on its performance. Where by the convolutional layers evaluates with feature response maps that has multiple channels along with spatial region. There is a high limitation of pooling layer that does not operate within one channel at a time while omitting activation values in every spatial region at the present channel.

The main focus is to fetch the learning and responsiveness into pooling operation. In our model is executed using two approaches. The first approach of pooling operations is max pooling and second approach is average pooling and we should combine them all together in the execution of the model. For us to combine these two pooling we should consider of two of the proposals which are responsive and unresponsive to the characteristics of the region being pooled. Where by unresponsive to characteristic of the region which is pooled assists the learning process in this strategy will give the outcomes in effective pooling op- eration that is specific and its unchanging mixture of max and also average. To amplify this unchanging mixture it's been introduced to this strategy as mixed max-average pooling.

The second strategy is responsive to the characteristics of the region being pooled that the learning process in this sphere of the strategy gives an outcome of gating mask where by the learned gating is mostly used to resolve a responsive mix of max pooling and average pooling. The inner product value which is between gating mask and the current region being pooled is provided through a sigmoid which is the outcomes mostly used as mixing proportional between max and average. To amplify the role of gating mask in resolving the responsive mixing proportion; then it's been introduced that this strategy as gated max- average pooling. In the fixed pooling operations it involves the mixed strategy and the gated strategy. The great achievement of these strategies is to learn the pooling operation and there execution performance by themselves. Then after we will be in position to consider the learning pooling operations and to know how to consolidate the pooling operations. In consolidating these within the idea of binary tree structure will portray as tree pooling.

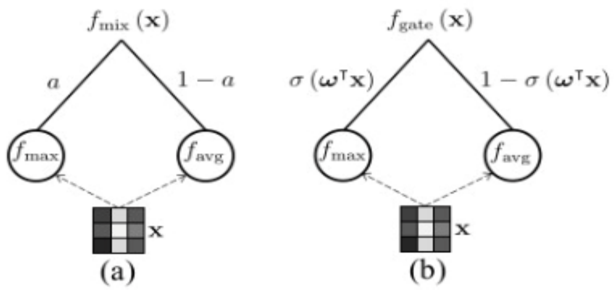

FIGURE 3.4: the diagram displays the proposed generalized pooling operation (a) gated max average pooling and (b) mixed max-average pooling. They have join max and average pooling function in two different strategies, nonrespon- sive and responsive

The recommended method is used on the visible deep learning framework PyTorch and make use of the open VGG network. In our work we make follow up of every convolu- tion stages where by the mixed pooling is mainly used to defend the convolution spatial pooling. The hyper parameter of the size of a batch is placed to the initial learning rate which is placed to the learning rate manner. Never less the stochastic gradient descent has method had a duty to update the network parameters. Inspired by Qiang Zhou, Zhong Qu and Chong Cao at the mixed pooling and richer attention feature for crack detection, they demonstrated that for the normal pooling takes the square area of the average/maximum value which has mostly given in computer vision objectives. The standard pooling makes use of the gemstone shapes which has irregular manner that appears in form of lines and curves. Making use of standard pooling to proceed with gemstone images will possess a huge unrequired area information that guides reduce the accuracy of crack detection. For us to sort this kind of situation we make use mixed pooling module to relieve the standard spatial pooling in every convolution stage. The below figure demonstrates how the mixed pooling module that is made of with horizontal pooling, vertical pooling, and maximum pooling.

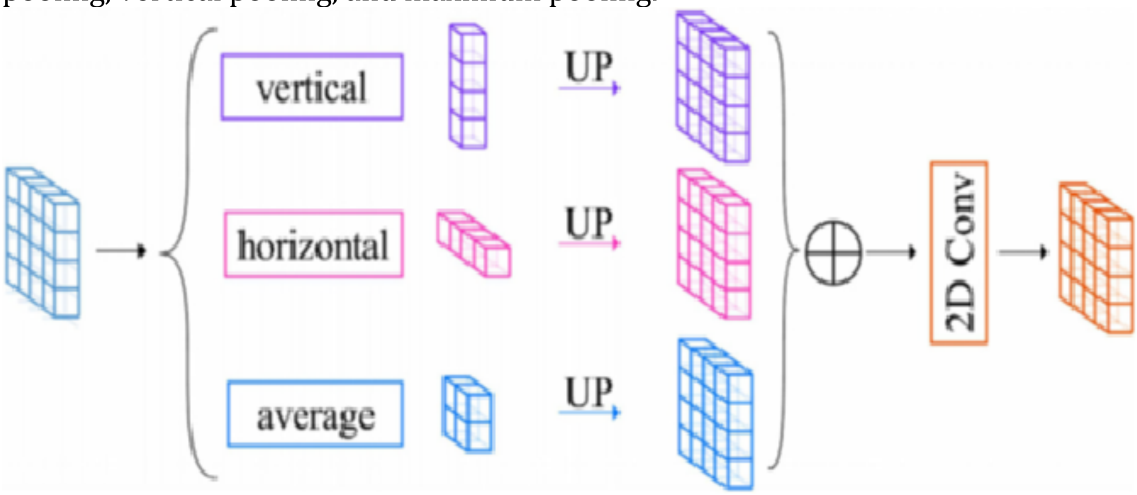

Figure 3.5: The mixed pooling module that is made of vertical pooling and horizontal pooling Centre's all of the feature values in a row or a column.

Flowchart Details.

The first block is the input block that receives the image sample which are ready to be executed in regular form that can be easily be defined. The second block Canny used in detecting the edges of the objects present in an image. The image under the canny edge undergoes the phase from original image, canny edge, image with bounding box and cropped image with the low error rate. The third block is foreground extraction; on these block displays the work done by the canny edge that shows the original image, canny edge, image with bounding together with cropped image that is been associated with low error rate. Gray scale image is displayed as gray scale conversation where there is the high determination of intensity of gradient that happens when color is changed. Maximum suppression is been implemented on image magnitude that generates the results of thickness edge that experience the thin edge that stimulates the maximum suppression of thin out of edge of an image.

Data augmentation block provide the assistance on decreasing the overfitting on the image sample. The data augmentation implementation relays on image flipping, image cropping, and image rotation, scaling and shifting. Convolutional neural network block is been comprise of hidden layers that are connected to the higher level features which are updated from the raw input. The convolutional neural network is used for the unstructured like object detection and semantic segmentation. The layers in the convolutional neural network they use matrix multiplication in matrix parameters with the separate parameter explaining the interaction within the input unit and on every output unit where the sparse connectivity are been involved on convolutional network. Output block displays the work done from the entire flow of processing ways which the image undergoes on these output block. The output block behaves as the final stage for the results to be displayed after long procedures that have been covered.
The convolutional neural network is the third block before the output which comprise of the most computation based on the hidden layer. The convolutional neural network is comprise of convolutional, pooling which has max pooling and average pooling and full connected layer. Other algorithm such as Canny algorithm, Adam optimization is been associated within the block. Most of the explained mechanism based associated on my work is been associated within within convolutional neural network
Output block displays the pure intended work which is expected to be displayed. Where by most of the work was done form other blocks based on the irregularity of the sample im- ages.The sample images on this block will display with the well classified of the images with their names and class groups where they naturally belong and this output is the in- tended outcome work to be done from the problem given. The output is been displayed in random plotting, cropped image plotting which is associated with canny edge, misclassffi- cation which is portrays noise effects on the outcome, Denoised which is classified and the final model prediction.

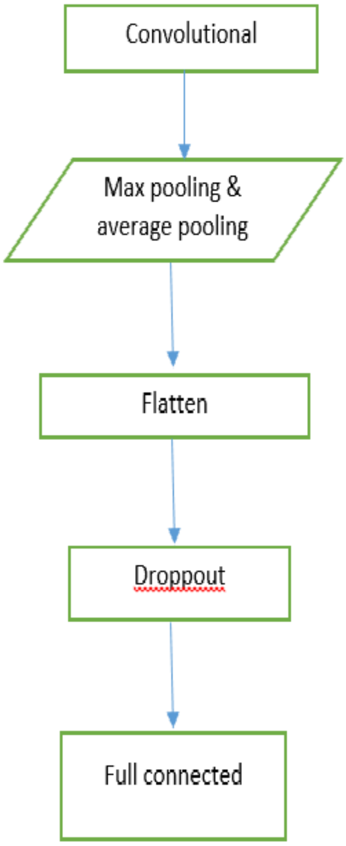

Figure 3.7: Flow Chart on the based on execution of our model

4: Experimental Results

We have collect the dataset of gemstone image for about 3200 image of faceted gems throughout the geological resources site . Image faceted gem are the gem that has color and shape feature whereby the gemstone image are grouped into 87 classes with division into train and test data. Never less the images are in various size of . jpg format and the shapes that they have is round, oval, square, heart and rectangle. In the division of train there 2,800 images and 400 images data. The data which is split into train are 80% and for test 20% are set. Whereas from the output below it demonstrate the 70% iteration to complete 1 epoch of 80% of 220 training sample and batch size is 32. The epoch it gives the illustration of the number of times the algorithm attends the entire dataset. The epoch seem to be complete in every time the algorithm attends all of the samples in the dataset. Considering one epoch it seem to be massive to feed the memory at once then divides it in several smaller batches, where by the batch size mostly is factor of 2. The output below displays the test loss parameter where it happened there is increasing of the parameter its overfitting even though it been handled by dropout.

4.0.1 Data Augmentation

One of the method that is made up in deep learning to moderates the deep models from overfitting. In usual manner for the neural networks to be trained it need a massive training sample which mostly are not accessible. The data augmentation is applied to increase the number of training samples. Image flipping, image cropping, image rotation, scaling, shift- ing are few examples of image data augmentation. In our work data augmentation is mostly applied; never less the overfitting is handled by applying dropout layer that randomly sets a fraction of its input to 0 and assists the deduction of overfitting. The data augmentation in our model is represented as image augmentation that creates additional training data base on the current image such as rotation, flips, zoom and translation. The below is the illustration of the image augmentation.

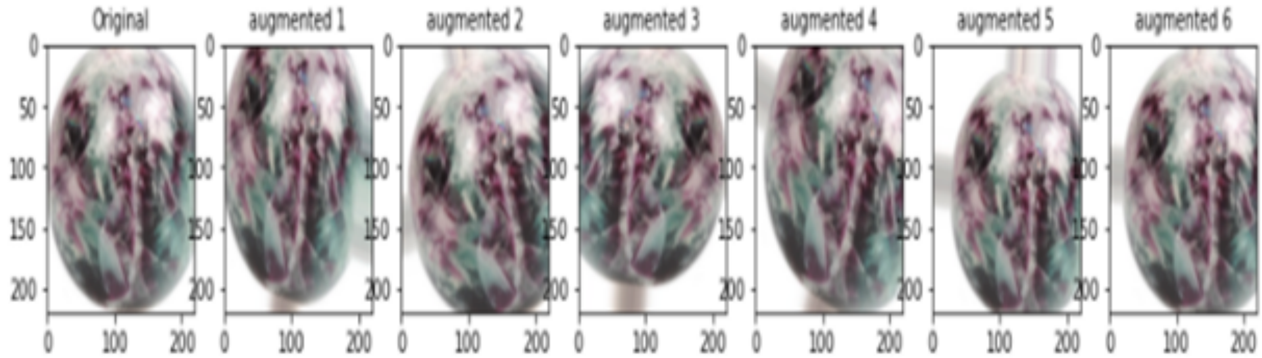

Figure 4.1: Image augmentation.

The image augmented it happens where artificially creates training images over sepa- rate ways of processing of multiple processing like shifts, random rotation, shear and flips. An augmented image generator can be easily produced using ImageDataGenerator API in keras. ImageDataGenerator produces batches of image data with real-time data augmenta- tion.

4.0.2 Experimental Setup

In our experiment Google Inc. was used in execution of the program where the online cloud server known as colab platform with the following specification XLA_CPU device: CPU, XLA_GPU device: Tesla P100-PCIE-16GB its used to speed up the computing use of GPU model. Where the XLA stands for accelerated linear algebra. Dropout layer is used to handle the overfitting of image during the execution of the model. In the weights of the network where there was initialized random pick and trained by adam gradient descent optimizer method associated with softmax loss function. In every experiment there was 70 epoch. i. 2D-CNN: The mostly used architecture in the computer vision field based on image classification, face detection, semantic segmentation and object detection is known as 2D classification neural network. It is popular on the performance of separating the spa- tial information from the image. The PCA method for spectral dimension is been used by 2D-CNN model for revealing the attachment and connection between variable and connec- tion between sample such as generating new hypotheses, finding and evaluating patterns together with detecting outliers.

VGG 16

Vgg 16 it means there is 16 layers in a network excluded the pooling layer on its opera- tions as 16 layers included exactly.It holds the number of filter that is within the convolution layer associated with the activation function.The implementation of VGG-16 is greater that involves 16 weight layers that consist of 13 convolution layer with filter size of 3 x 3 and full-connected layers. Initially the convolution layer posses of 96 filter of size 11 x 11 alongside a slide of 4 pixel and padding with 2 pixel. The convolution layers are been set to 1 pixel from other stride s and padding. From the next in order is convolution layer that has 256filter of size 5 x 5. On the third, fourth and fifth convolution layers posses 384, 384 and 256 filters with size 3 x 3 accordingly. Within the convolution layer their stride and padding which are fixed to 1 pixel.

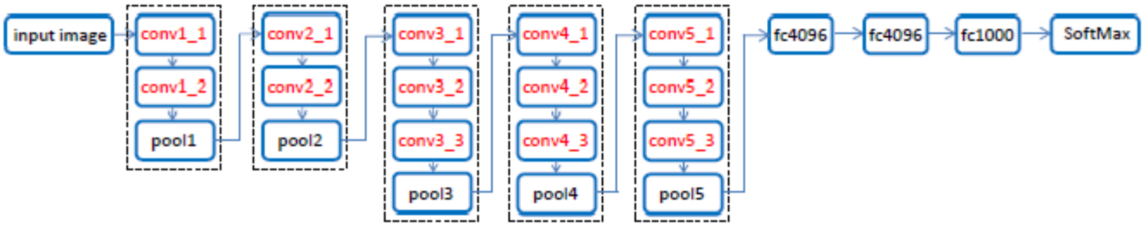

Figure 4.2 The diagram above is the Vgg 16 architecture.

Most of the convolution layer are in split into 5 batches and each batches is associated with pooling layer. The pooling layer (mix pooling) is carried out on a 2 x 2 window of stride 2. In the first batch the number of filters of convolution layer batch it starts from 64 then after it increases by factor of 2 after every pooling layer till it reaches 512. The Vgg 16 make use of small convolution filter that comprise of many convolution layers. The Vgg consist of 138M parameters that solves the difficulties on complex computational requirements.

DATASET NETWORK PARAMETERS:

Layer (type)	Output Shape	Param #
conv 2d (Conv 2D)	(None, 220, 220, 32),	896
max_pooling2d (MaxPooling2D)	(None, 110, 110, 32),	0
conv2d_1(Conv2D)	(None, 110, 110, 64),	18496
max_pooling2d_1 (MaxPooling2)	(None, 55, 55, 64),	0
conv2d_2 (Conv2D)	(None, 55, 55, 128),	73856
max_pooling2d_2 (MaxPooling2)	(None, 27, 27, 128),	0
Conv2d_3 (Conv2D)	(None, 27, 27, 128),	147584
Average_pooling2d (AveragePo)	(None, 13, 13, 128),	0
Conv2d_4 (Conv2D)	(None, 13, 13, 128),	147584
max_pooling2d_3 (MaxPooling2)	(None, 6, 6, 128),	0
flatten (Flatten)	(None, 4608),	0
dropout (Dropout)	(None, 4608),	0
dense (Dense)	(None, 512),	2359808
Dense_1 (Dense)	(None, 87),	
Total params: 2,792,855 Trainable params: 2,792,855 Non-trainable param: 0		

EVALUATION CRITERIA

The experimental results are described and analyzed at the dataset level. Initially we present per class accuracy and then after we present the confusion matrix for every dataset.

The reported accuracy is the evaluated metric for classification model. The confusion matrix outputs relay on a matric that describe the complete achievement of the model. It describes the correctly and incorrectly classified sample at per class level. The calculation of accuracy of matrix is by taking sum of the value lying across the"main diagonal"divide by the total numbers of samples.

$$\text{Accuracy} = \frac{Correctly\ classified\ samples\ Total\ Number\ of\ samples}{Total\ Number\ of\ samples}$$

Where the correctly classified sample are cases except the predicted outcomes are the same as actual truth label. The Average Accuracy is given by:

$$\text{Average Accuracy} = \frac{1}{C}\sum_{i=1}^{C}(x)$$

Where c is the number of classes and x is the percentage of correctly classified pixel in a single class.

The Results from deep convolutional neural network.

Python's VGG16 function is been used to fit the VGG16 model associated with weights from ImageNet. VGG16 requires two channel images as input then after the grayscale image repeats two time in order to obtain two channel input. The augmentation happens to insure the gemstone has label that possess 3000 images from the two channel image that are been expanded in the training set, then after the VGG 16 deep convolutional neural network model from the scratch it achieves a 60% accuracy which demonstrates the 80% training sample of about 3000 training sample and 20% test set. The test accuracy and training from different association of sample and the size of augmented are displayed below on the figure below; Where by the sample type as there is increase of image augmented, the training accuracy raise, never less there is an improved initially the test accuracy; never less when utilizing the same number of augmented images the sample size is raise which appears to deduct the training accuracy and raise the test accuracy.

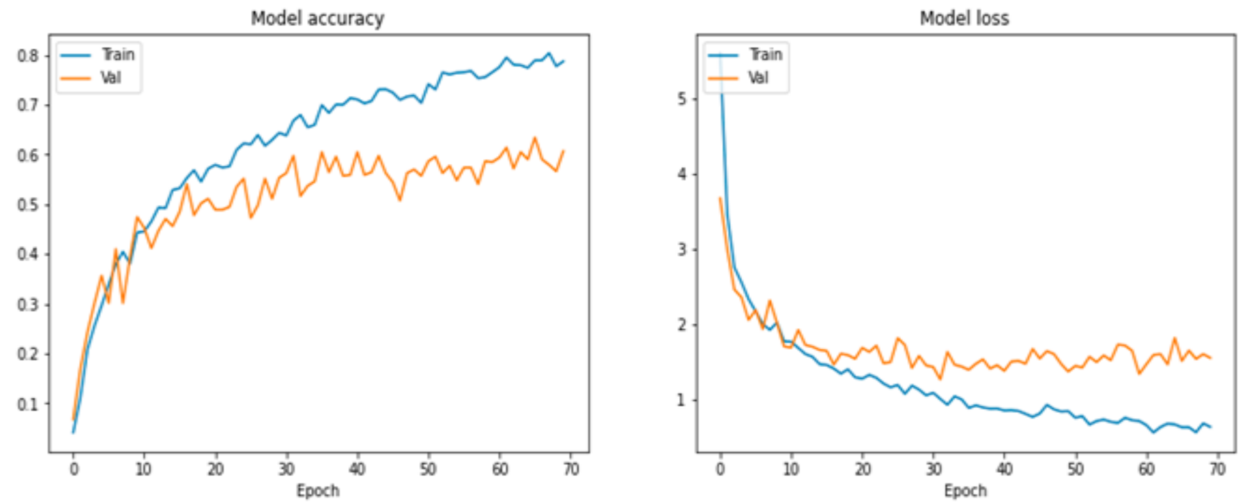

figure 4.3: (a)Model Accuracy (b) Model loss with 30% train data on dataset.

We have plot accuracy of model against size of epoch (train and test) and then after we plot the loss of model against size of epoch (train and test) The epoch illustrates and displays the number of times the algorithm shows the whole dataset. In every time the algorithm has counter the dataset makes the epoch to be complete and when it happens the epoch is large to feed in to the memory immediately split it in various smaller batches. Where the batch size is always the factor of 2. The estimation is given by; iteration per epoch = number of passes, each pass using batch size number of examples. For example if we have -2200 (80%) training sample and batch size is 32 later on it will take 70 iteration to complete 1 epoch.

Canny Edge implementation:

The image displayed below explains the work of canny algorithm that illustrate from the original image, canny edge, image with bounding box, cropped image. The canny edge algorithm is the most popular algorithm that used in detecting the edges of objects present in an image. The Canny edge gives out the outcomes of its principle operation using low error rate, well localization of edge point and one response to a single edge. On canny edge there is converting of image to gray scale which is known as grayscale conversation where there is the high determination of intensity of gradient that occur when the color is changed. The non-maximum suppression is been illustrated on the image magnitude that generate the outcomes in thickness edge that experience the thin of the edge that triggers to perform maximum suppression of the thin out of the edges.

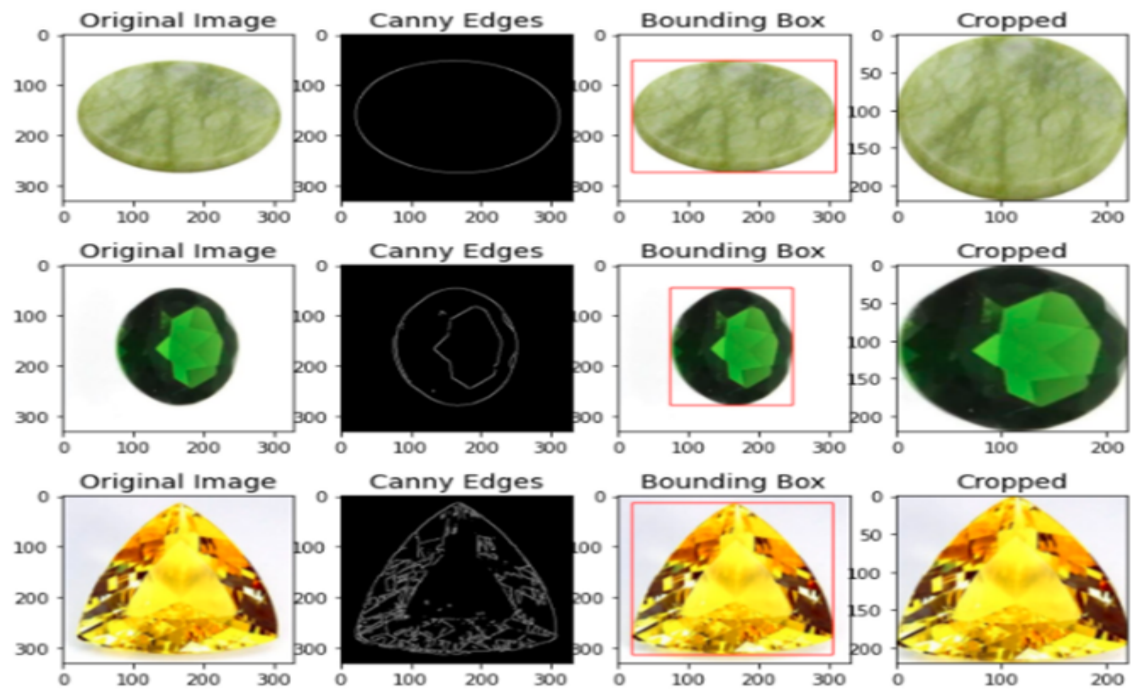

Figure 4.4: Canny edge algorithm, cropping and fitting is been well illustrated on our results.

Misclassification

On our model it happen the mistake of labeling the image samples. These manner is also known as classification errors. These only happen when the classifiers will mostly make wrong prediction, when exactly is posses high confidence of accuracy acceptation. Our model that has make a closer detection before the final pick of the inaccurate class been predicted. The fig:4.5 demonstrate the misclassification outcomes.

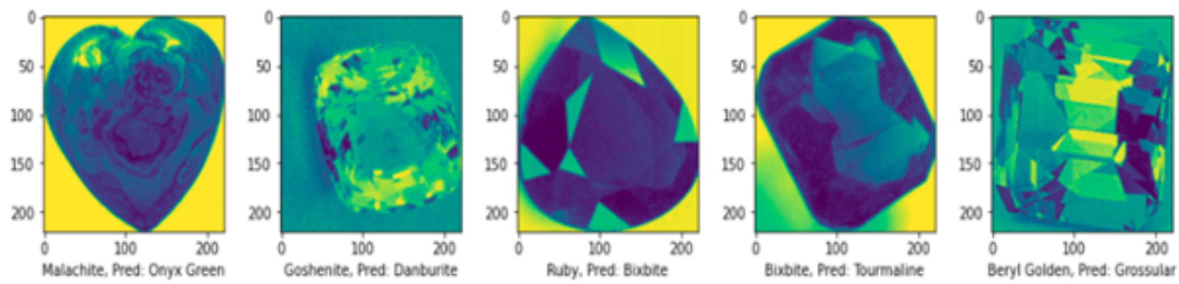

Figure 4.5 : (a) Misclassified Illustration:

In spite of been invaded by noise, the image picture can not easily be identified within the prediction. Misclassification it happens by base softmax function it holds the probability of the high set of the target class and few other samples which assist the tension of the classification between them. The misclassification is intentionally deviating pixel to be in a separate ways than on what is been intended to be displayed. The misclassification happens when there is existence of capacity of the error in the evaluation process for the given for variable that are arranged. The misclassification happens when the sample images are wrongly grouped into specific group for the reason of observation error.

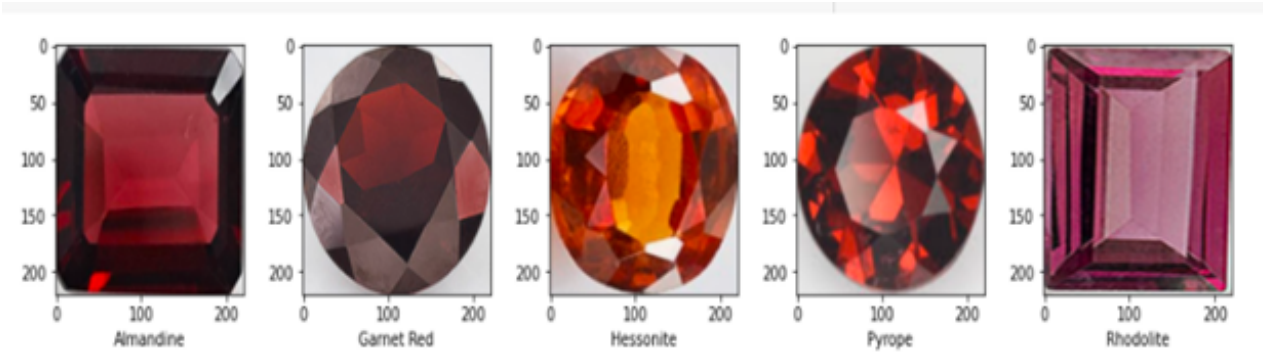

Figure 4.6: (b) Connected Classified illustration:

The illustration on fig:4.6 is the well predicted image sample that are intended to be displayed that can be easily be predicted and has no noise within it. The labelling is done correctly according to the name of the gemstone sample that relates simultaneous with the its origin of the sample color of gemstone. The corrected display above demonstrate on how the clean image has come from removing the intrusion and restoring from images. What happened on these corrected images is that within the convolution neural network there is a way of splitting the noise from the image by feed-forward within the convolution neural network. These situation is been handled as a plain discriminative learning problem. The main factors for using convolution neural network are because of the deep architecture that is active to increase the capability and volatile for exploiting features. In spite of that many development have been handled on regularization and learning skills when it comes on leaning skills that will train convolution neural network that is comprise of variant ReLu, batch normalization and residual learning which assist the CNN to increase the training process and modify the denoising performance.

4.2.3 Actual Prediction:

The actual intent prediction is been displayed from the test folder that contain with a label, class which model predicted and actual class. The mechanism that has illustrated the intended prediction is been discussed that the intended gemstone images are been classi- fied with the accurate expectation. The expectation starts from its naming thus all of the gemstone image are been arranged according to its original class then its name comes after that. The defect of an improved canny algorithm is when it won’t be able to detect edge when the sample image has almost the same color as background; that makes the array edge to have zero value then it make use of the original image. From the name given on the gemstone it indicate that the correctly fitted image sample by naming is in black color and the naming with dark red color are the images that are been misclassified. The actual prediction is been demonstrated on fig:4.7

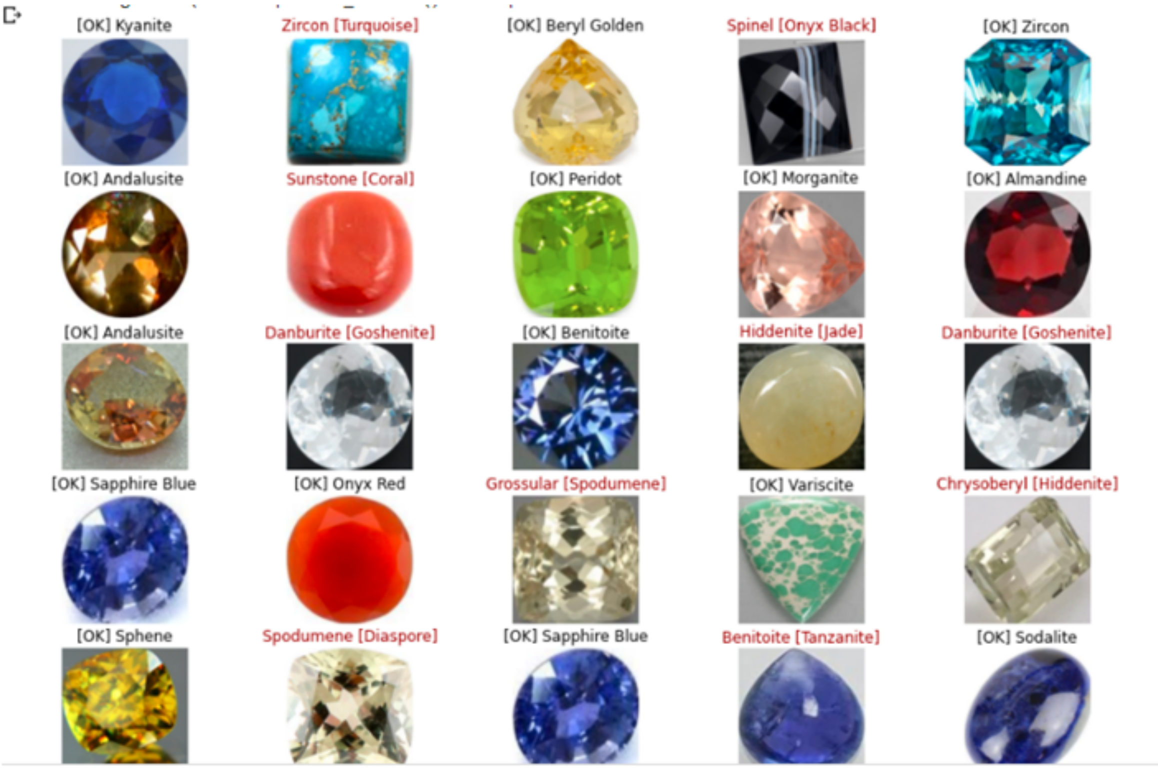

Figure 4.7: Final prediction of our gemstone model

5: Conclusion

5.1 Conclusion and Recommendation

On this chapter we provide the conclusion and recommendation on this research. We propose the convolutional neural network, max pooling, and average pooling that involve Adam optimizer and canny edge embedded with cross entropy. The proposed Convolutional neural network introduce max pooling, average pooling that is embedded by spatial size for reduction of sample images. The convolutional neural network is been comprise of conv2D, max pooling layer, average pooling and dense that involve convolutional and non-linear layer. The dropout and batch normalization are engaged on the dropout on the hidden and the layers which are visible in the neural network that is been implemented in hidden.

We recommend more contribution on our research work that is based on the gemstone multiclass classification using convolutional neural network. The ideology relays on our model performance on Geological dataset that can be easily been phished; if it won't be handled on the level that relates on the performance of our model.

Bibliography:

1. V. B. da Silva, M. Ritt, J. B. da Paz Carvalho, M. J. Brusso, J. T. da Silva, and A. L. Zanatta, "A real-valued genetic algorithm for gemstone cutting," in *2012 XXXVIII Conferencia Latinoamericana En Informatica (CLEI)*, pp. 1–8, IEEE, 2012.
2. R. S. Castro Rios, "Researching of the deep neural network for amber gemstone clas- sification," Master's thesis, Universitat Politècnica de Catalunya, 2018.
3. K.Sfikas,T.Theoharis,andI.Pratikakis,"Exploitingthepanoramarepresentationfor convolutional neural network classification and retrieval.," *3DOR*, vol. 6, p. 7, 2017.

4. E. Chen, X. Wu, C. Wang, and Y. Du, "Application of improved convolutional neural network in image classification," in *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pp. 109–113, IEEE, 2019.
5. W.Rong,Z.Li,W.Zhang,andL.Sun,"Animprovedcannyedgedetectionalgorithm," in *2014 IEEE international conference on mechatronics and automation*, pp. 577–582, IEEE, 2014.
6. R. Biswas and J. Sil, "An improved canny edge detection algorithm based on type-2 fuzzy sets," *Procedia Technology*, vol. 4, pp. 820–824, 2012.
7. D.Lunga,S.Prasad,M.M.Crawford,andO.Ersoy,"Manifold-learning-basedfeature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 55–66, 2013.
8. T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral remotely sensed imagery—a nonlinear time series analysis approach," in *2007 IEEE In- ternational Geoscience and Remote Sensing Symposium*, pp. 1556–1560, IEEE, 2007.
9. J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
10. C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE transactions on geoscience and remote sensing*, vol. 44, no. 10, pp. 2786–2803, 2006.
11. B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1139–1155, 2009.
12. P. Hu, X. Liu, Y. Cai, and Z. Cai, "Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 3, pp. 452–456, 2018.
13. Q. Gao, S. Lim, and X. Jia, "Hyperspectral image classification using joint sparse model and discontinuity preserving relaxation," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 78–82, 2017.
14. E.D.Cubuk,B.Zoph,D.Mane,V.Vasudevan,andQ.V.Le,"Autoaugment:Learning augmentation policies from data," *arXiv preprint arXiv:1805.09501*, 2018.
15. L. Pan, S. Pouyanfar, H. Chen, J. Qin, and S.-C. Chen, "Deepfood: Automatic multi-class classification of food ingredients using deep learning," in *2017 IEEE 3rd international conference on collaboration and internet computing (CIC)*, pp. 181–189, IEEE, 2017.
16. B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 340–344, 2018.
17. L. Zhang, B. Xiong, and G. Kuang, "Convolution neural network with sample stretch for sar target recognition," in *2018 China International SAR Symposium (CISS)*, pp. 1– 5, IEEE, 2018.
18. J. Wu, "Introduction to convolutional neural networks," *National Key Lab for Novel Software Technology. Nanjing University. China*, vol. 5, p. 23, 2017.
19. Z. Eaton-Rosen, F. Bragman, S. Ourselin, and M. J. Cardoso, "Improving data augmentation for medical image segmentation," 2018.

20. X. Zhang, C. Xv, M. Shen, X. He, and W. Du, “Survey of convolutional neural network,” in *2018 International Conference on Network, Communication, Computer En- gineering (NCCE 2018)*, Atlantis Press, 2018.
21. B. Wang and S. Fan, “An improved canny edge detection algorithm,” in *2009 Second International Workshop on Computer Science and Engineering*, vol. 1, pp. 497–500, 2009.
22. H. Zhao, G. Qin, and X. Wang, “Improvement of canny algorithm based on pavement edge detection,” in *2010 3rd International Congress on Image and Signal Processing*, vol. 2, pp. 964–967, IEEE, 2010.
23. H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu, “Robust classification with convolutional prototype learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3474–3482, 2018.
24. X. Wang, S. Zhang, Z. Lei, S. Liu, X. Guo, and S. Z. Li, “Ensemble soft-margin softmax loss for image classification,” *arXiv preprint arXiv:1805.03922*, 2018.
25. X. Wang, S. Zhang, Z. Lei, S. Liu, X. Guo, and S. Z. Li, “Ensemble soft-margin softmax loss for image classification,” *arXiv preprint arXiv:1805.03922*, 2018.
26. A. U. Ruby, D. I. Prasannavenkatesan Theerthagiri, and Y. Vamsidhar, “Binary cross entropy with deep learning technique for image classification,” *International Journal*, vol. 9, no. 4, 2020.
27. E. A. Babushkina, L. V. Belokopytova, A. M. Grachev, D. M. Meko, and E. A. Vaganov, “Variation of the hydrological regime of bele-shira closed basin in southern siberia and its reflection in the radial growth of larix sibirica,” *Regional Environmental Change*, vol. 17, no. 6, pp. 1725–1737, 2017.
28. G.Keren,S.Sabato,andB.Schuller,“Fastsingle-classclassificationandtheprinciple of logit separation,” in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 227–236, IEEE, 2018.
29. G.Keren,S.Sabato,andB.Schuller,“Fastsingle-classclassificationandtheprinciple of logit separation,” in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 227–236, IEEE, 2018.
30. A. Naser, M. Tantawi, H. A. Shedeed, and M. F. Tolba, “Eeg based epilepsy detection using approximation entropy and different classification strategies,” in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 92–97, IEEE, 2017.
31. A. Tuerk and S. Young, “Polynomial softmax functions for pattern classification,” 2001.
32. D. Nyasaka, J. Wang, and H. Tinega, “Learning hyperspectral feature extraction and classification with resnext network,” *arXiv preprint arXiv:2002.02585*, 2020.
33. L.Mou,P.Ghamisi,andX.X.Zhu,“Deeprecurrentneuralnetworksforhyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.
34. X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, “Stock market prediction on high-frequency data using generative adversarial nets,” *Mathematical Problems in Engineering*, vol. 2018, 2018.
35. X. Sun, F. Zhou, J. Dong, F. Gao, Q. Mu, and X. Wang, “Encoding spectral and spatial context information for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2250–2254, 2017.
36. S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, “Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks,” *IEEE*

Transactions on Geoscience and Remote Sensing, vol. 55, no. 8, pp. 4520–4533, 2017.
37. Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.
38. Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.
39. C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9201–9222, 2019.
40. M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, and F. Pla, "Capsule networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2145–2160, 2018.
41. R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image clas- sification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2693–2705, 2017.
42. H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyper- spectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1259–1270, 2017.
43. R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5384–5394, 2019.
44. X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sens- ing data classification based on convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 937–949, 2017.
45. S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, "Two-stream deep architecture for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2349–2361, 2017.
46. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
47. L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5585– 5599, 2017.
48. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.
49. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
50. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.