Image Classification

Anahita Esfandiaryfard A.ESFANDIARYFARD@STUDIUM.UNICT
Gianluigi Mazzaglia[1] UNI310589@STUDIUM.UNICT.IT

1. Disclaimer

For this project we initially user Medical MNIST dataset but since the accuracy of the model was so high and the difference of accuracy in different layers of the convolution is not really visible we decided to test our model on another dataset called Flowers Recognition as well to compare the result with a different input data and see the performance of the algorithm on a dataset with fewer data and images with lower quality. Since we worked in both dataset and analyzed the result in both models we decided to include both results here.

2. Model Description

First we describe here **Medical MNIST**, a deep model that consists of a 3D encoder, to learn volumetric information, followed by a fully connected layer to perform the classification(see Fig. 5). Initially, input data (3D crop around a lesion) flows trough a 3D encoder, to exploit inter slice volumetric information. Then the extracted features are provided to a capsules-based encoder, which predicts image classes outcome as the name of the body part which the picture is taken.

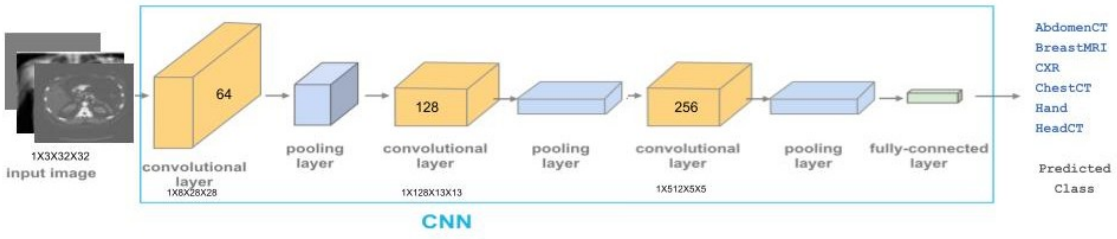

Figure 1: A simple representation of the proposed model.

It presents an initial 3D encoder of three convolution layers, to process the input volume. The three 3D convolution layers have respectively 64, 128 and 512, 5x3x3 kernels applied with a stride of 1 in each dimension and followed by a ReLU activation function. After each convolution layer we added a pooling layer and then we used two fully connected layer. For the second attempt we trained another model based on another dataset to make sure if the result of our first model is correct we used Flowers Recognition dataset, we created a model as same as before whit this difference that we added 2 more convolution layers. The five 3D convolution layers have respectively 8,16,32,64,128 channels,a kernel size of 3 is applied

with a stride of 1 in each dimension and followed by a ReLU activation function and by a Batch Normalization. After each convolution layer we added a pooling layer and then we use 2 fully connected layers and a Classifier that assigns a probability to each of the 5 classes in order to predict the most likely class.

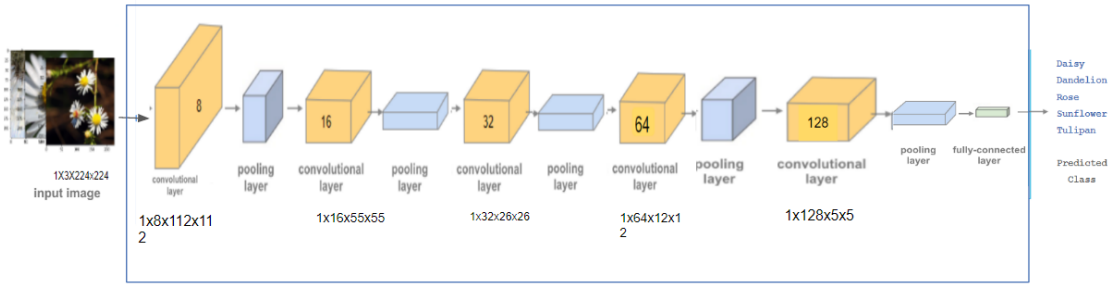

Figure 2: A simple representation of the proposed model.

3. Dataset

The first dataset is provided by the Kaggel, which is a simple MNIST-style medical images in 64x64 dimension, There were originally taken from other datasets and processed into such style. There are 58954 medical images belonging to 6 classes, all including 10000 images and one of the classes (Breast MRI) include 8954 datas. We separated this dataset to three sub dataset using index selection randomly from the original data set in a way guaranty that we have sample from all labels. We used a centerCrop that given tensor image at the center using an input size of 64.

For the second model we used another dataset provided by Kaggle, names Flower Recognition. This dataset contains 4242 images of flowers in 5 classes, daisy with 764 images, dandelion 1052, rose 784, sunflower 733 and tulip 984. The data collection is based on the data flicr, google images, yandex images.That which can be used to recognize plants from the photo. The pictures are divided into five classes: daisy, tulip, rose, sunflower, dandelion.Photos are not high resolution, about 320x240 pixels. Photos are not reduced to a single size, they have different proportions. We used a centerCrop that given tensor image at the center using an input size of 224. We used a Random Horizontal and Vertical flip in order to flip the image with a probability p as a default (p=0.5). We also resized all the image given an equal size of 224.

For both dataset We used 80 percent of the data as train set and then 10 percent for test and 10 percent for validation set

4. Training procedure

For both models we first started to train a model with one conv layer and get the training result and step by step we added another conv layer and we compared the result to make sure if its a good fit and we are avoiding over fitting or under fitting.

The medical models are trained for 10 epochs on a Colab using GPU and CPU. (Most of the time CPU because of the limitation of colab and not letting us to use GPU).Batch size is set to 64 and we used SGD as optimizer. We stoped the training ater 3 layers because the accuracy of the model had no change and was almost equal in more layers.

On the other hand the flower model decided based on the results to be trained with 20 epochs, with batch size of 16 because the dimension of the dataset was smaller in compare to the other one. The learning rate used for both models is equal to 0.01. We decided to stop the training after 5 layers because the loss function started to increase after 5 layers.

5. Experimental Results

Table 2 shows test results for the Medical MNIST model:

Model	Accuracy
– + Layer 1	99.9%
– + Layer 2	99.8%
– + Layer 3	99.8%
Your final model	99.8%

Table 1: Medical MNIST

And above is the confusion matrix for this model:

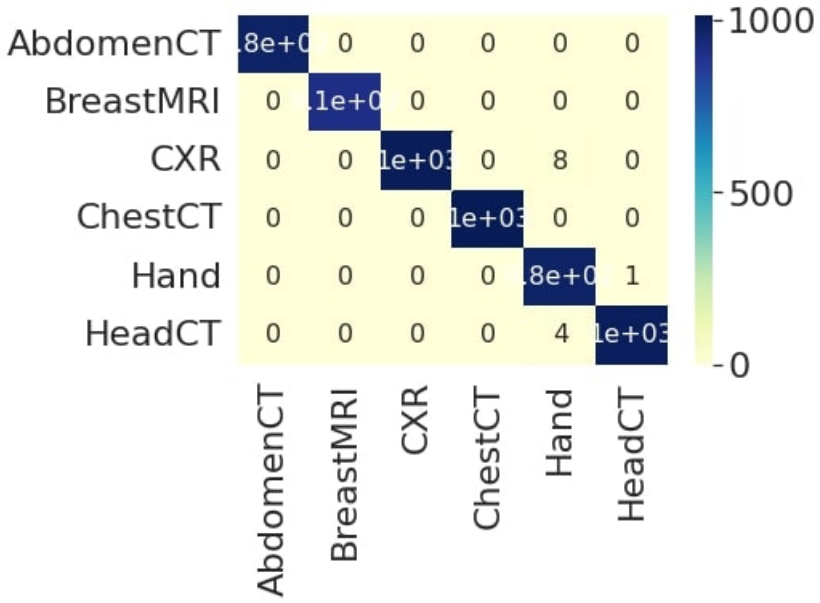

Figure 3: Medical MNIST Confusion Matrix

This table 2 shows test results for the Flowers Recognition model and in the following in the confusion matrix:

Model	Accuracy
– + Layer 1	57.22%
– + Layer 2	63.99%
– + Layer 3	66.29%
– + Layer 4	69.87%
– + Layer 5	73.14%
Your final model	73.14%

Table 2: Flower Recognition

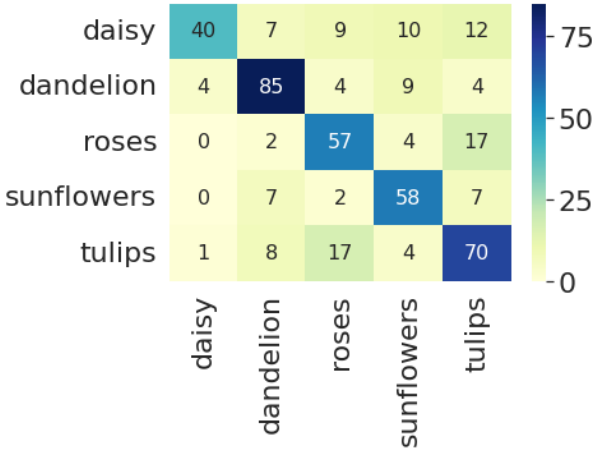

Figure 4: Flower Recognition Confusion Matrix

As we can see from the confusion matrix, the class "daisy" is that one with the lower percentage of images correctly classified and we can see from the image below as an example of daisy which is classified as sunflowers:

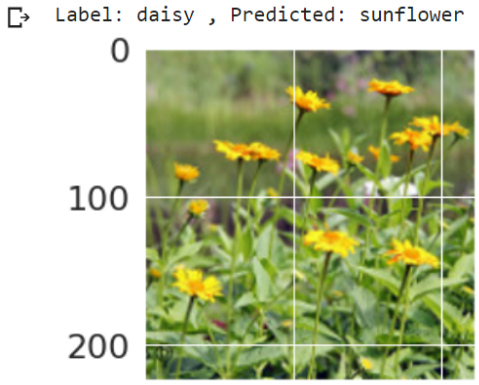

Figure 5: Example of misclassified image