Lab: Chihuahua or Muffin

Devin Prikryl-Martin

Houston Community College

ITAI 1378 Computer Vision

Professor Patricia McManus

October 2, 2024

Lab: Chihuahua or Muffin

I approached this lab with confidence and was quickly reminded how much I still have to learn in respect to the technical and programming side of Computer Vision. I think I understand the process at a high level but I'm barely beginning my journey with programming so I got a bit stuck during the lab when running into a fill in the blank errors because I lack fluency.

I understand that we begin the lab by importing our libraries and frameworks. We then build our neural network, beginning with defining the architecture of the classifier. When originally confronted by the first two question marks for input height and width I wasn't sure what exactly this was in reference to initially so I plugged in a 4 for each just to see if it worked without error. When I scrolled a bit farther down I saw the image of the chihuahua with the dimensions 3X224X224 and realized the image height and width was probably referring to this 224X224 pixel image so I went back and adjusted those numbers accordingly. My understanding of the rest of this code cell is extremely limited, but from what I can understand it appears to be structuring the neural network, giving it information about the how the neural network will run, which includes information about the input, output and hidden layers in the network and how information or data will pass through it.

The next cell appears to tell the computer where to run the network, in the GPU or CPU. When confronted with the question mark in this cell I tried several guesses before I resorted to seeking out further information. My first attempt was "load" and my second attempt was "save", more pure stabs in the dark than informed guesses. My third attempt was "MySkynet" because I could see in the previous cell that "class MySkynet(nn.Module):" has a line below it that seems to refer to it as a "very basic neural network" and I used the note "# load our simple neural network" as the context clue to connect those ideas. When that didn't work I took to Claude.ai to see if I could use this resource to help me understand what I was missing. I gave Claude an image of the code cells up to this point and asked what I should input instead of the question mark in order to resolve the error. Claude informed me the correct answer was "MySkynet()" which lead me to try to understand what an empty set of parentheses means in this context, as that is what was missing from my last attempt. (Claude, 2024) This tells me I was on the right track but I lack the technical knowledge that would allow me to apply my suppositions correctly. I learned that an empty set of parentheses in this context probably means we are executing a function call with no arguments, which means you are telling the computer to run a set of instructions defined by a function and you are running it in its default state, because an argument would introduce new information that would change the function. (Yordanov, 2019)

Next in the lab is a section titled "Data and Data Loading" which begins by reminding us that we will need to use some of our data to train the model, but reserve some of the data to validate the accuracy of the model, or in other words demonstrate that the model is able to accurately classify new images it encounters or if it is overfitted to the training data and will only produce accurate results when viewing that specific dataset. The

subsequent two cells appear to show the information about the image data set and then display examples of the labeled training images the model will use to teach itself to identify image features. I understand by reading through the next steps that we must convert the images into tensor objects, which are 3 dimensional arrays of pixels that the neural network is able to understand. The next step is defining "transforms" to convert the images to tensors. Unfortunately, the link explaining more about transforms leads to a non-existent webpage, so I returned to Claude to deepen my understanding. (Claude, 2024) From this resource I learn that transforms are "a series of operations or modifications applied to the input data to prepare them for the neural network" which I understand to be comparable to the preprocessing step in our previous lab during which we normalized the images and converted them to greyscale. I can infer from this that the transforms in the next cell involve resizing the images, converting them to tensors and normalizing them. The cell that follows involves creating the dataset with the applied transforms. I used Claude again to help me figure out what to input here after first trying just "transforms" and running into an error. (Claude, 2024) The next step involves forming dataloaders from datasets and defining the batch size, I chose 32 as a trial number.

We then train the model, which begins with defining the train loop. This involves defining the number of times it trains on the data, which I understand to be represented by "epochs". Each epoch includes a training and a validation set. During this training the data moves forwards through the model, calculates loss, moves backwards through the model to correct mistakes and then updates the model's weights. Then during validation phase it assesses how the model is doing without updating weights. During both phases I keeps track of loss and accuracy. We then define the loss (or "error") function and the optimizer. From what I am able to understand these are both pre-existing algorithms that we call upon to help our model learn from its mistakes.

Finally we train our model and assess the results. The output shows that we ran 3 epochs with decreasing training error and increasing accuracy in each iteration. When we examine the visualization of how our model interpreted the validation set I am impressed to find the Epoch 3/3 accuracy finding of .9333 to be spot on. Only two of the images were incorrectly labeled and we're middle of the road in terms of confidence of the answer. The model seems to be better at understanding the features of the chihuahua and is more confused when it looks at a muffin. I attempted to reach 100% accuracy by adjusting some of the parameters listed in the lab. It didn't take much, simply adding an epoch lead to a 1.000 accuracy by the last epoch. This tells me how vitally important it is for a model to train an appropriate number of times.

There were a number of times during this lab where I felt a bit defeated by my lack of understanding. I am the type of student who has rarely struggled inside of academia because my pursuits have largely been in my comfort zone. Deciding to pursue AI has been an exercise in humility and adaptability for me, and was a decision I made with the intent to grow my capacity for the discomfort that comes with trying to learn things I don't naturally and easily understand. This lab, at times, felt like moving from the shallow end of a pool to

the open ocean. But every time I got a little overwhelmed, I made sure to use a tool, like Claude, to help provide clarity and point me in the right direction. I am so grateful to be learning computer science at a time when these sorts of tools are available.

Although this exercise is a bit on the silly side, because humans don't have much need for help distinguishing a dog from a pastry, I can immediately see how this technology could have extremely beneficial real-world use cases. I could see how this might be useful examining lots of images of potentially cancerous tissue or scanning for various medical anomalies that a human eye might miss. I could also see how this technology could potentially be used in manufacturing to examine the quality of products, especially in conditions that might be unsafe for a human to do the examining, like in extreme temperatures or pressures.

References

Yordanov, V. (2019, February 20). *Python basics: Functions*. Medium. https://towardsdatascience.com/python-basics-functions-ed7c35e194a9

Anthropic. (2024, October 2). Claude. "Review these screenshots and explain what should replace the question mark in the code and why"

Anthropic. (2024, October 2). Claude. "Explain in detail what 'transforms' are in the context of Python and Computer Vision"