

Name:

Hanliang Jiang

Netid:

hj33

CS 441 - HW 5: Deep Learning and Applications

Complete the sections below. You do not need to fill out the checklist. **Be sure to select all relevant pages in Gradescope.**

Total Points Available

[] / 150

1. Applications of AI
 1. Describe the applications [] / 15
 2. Positive impact [] / 7
 3. Negative impact [] / 8
2. Fine-tune Model for Pets Classification
 1. Qs about ResNet-34 structure [] / 10
 2. Epochs vs Loss Plots [] / 10
 3. Best Performance / Question [] / 10
3. CLIP: Contrastive Language-Image Pretraining
 1. Test CLIP zero-shot performance [] / 20
 2. Test CLIP linear probe performance [] / 10
 3. KNN on CLIP features [] / 10
4. Stretch Goals
 1. Compare word tokenizers [] / 20
 2. Implement/train custom network [] / 30

1. Answer “Applications of AI” Questions

1. **How is AI used in that application area?** What is the problem that AI is trying to solve, and what are the key AI/ML technologies involved? What are the technical challenges? (100+ words)

AI is used in areas like manufacturing where robots take place of human workers to produce products. In this case, AI tools use control theory and reinforcement learning to make robots finish the tasks like human beings do. In this field, the biggest challenge might be visualization and location, which is of crucial importance if the robot engage in works that demand accuracy. AI is also used in human computer interaction areas. Siri is a good example, which involves large language models to provide good experience making people's lives easier. Face recognition is another example which uses computer vision methods to identify any human beings or users. Deep learning methods are used in both of these methods, and one technical challenge is that the devices like cell phones are power demanding, so engineers need to come up with more power efficient methods for these detection and interaction.

2. What is the actual or potential positive impact? Who is impacted? (50+ words)

Consumers and users might be a potentially impacted object. Users benefit from better service and personalized experience provided by AI tools. They can save their time working on redundant and boring work with the help of automation tools. They can also get information faster with the help of large language models. In this sense, everyone who uses modern technology benefits.

3. What is the actual or potential negative impact? Who is impacted? (50+ words)

Educators might suffer from the impact of AI tools, they can easily let their students think independently to get the right answers to questions, but AI tools like GPT have been so powerful that students cannot get a better answer working themselves, they need to cooperate with AI tools to get a better grade, and worse yet, it is hard for educators to detect and determine if a submitted piece of homework is done by the students themselves. Besides, many people like sales person might lose their jobs because their work can be done better by AI tools.

4. **What are your sources?** (include full citations and links if available) [required; -15 pts if not provided] Format is not as important as being clear about what source is used.

<https://medium.com/@careervira.community/most-important-advantages-and-disadvantages-of-machine-learning-in-2024-9d56ee451622#:~:text=One%20of%20the%20main%20benefits,manual%20labor%2C%20and%20improve%20efficiency.>

This source is used to inspire my opinions.

2. CNN: Image Classification

1. Answer these questions about the network structure of your model based on ResNet-34

1.1. How many parameters are there in total? (xx.x million)

21303653

1.2. Which of these layers **do not** have trainable parameters? (choose more than 1)

- a) Convolutional
- b) BatchNorm
- c) ReLU
- d) Max pooling
- e) Fully connected

c,d

1.3. True or false: In layers 1-4, whenever the feature map is downsampled by a factor of 2, the number of features is doubled.

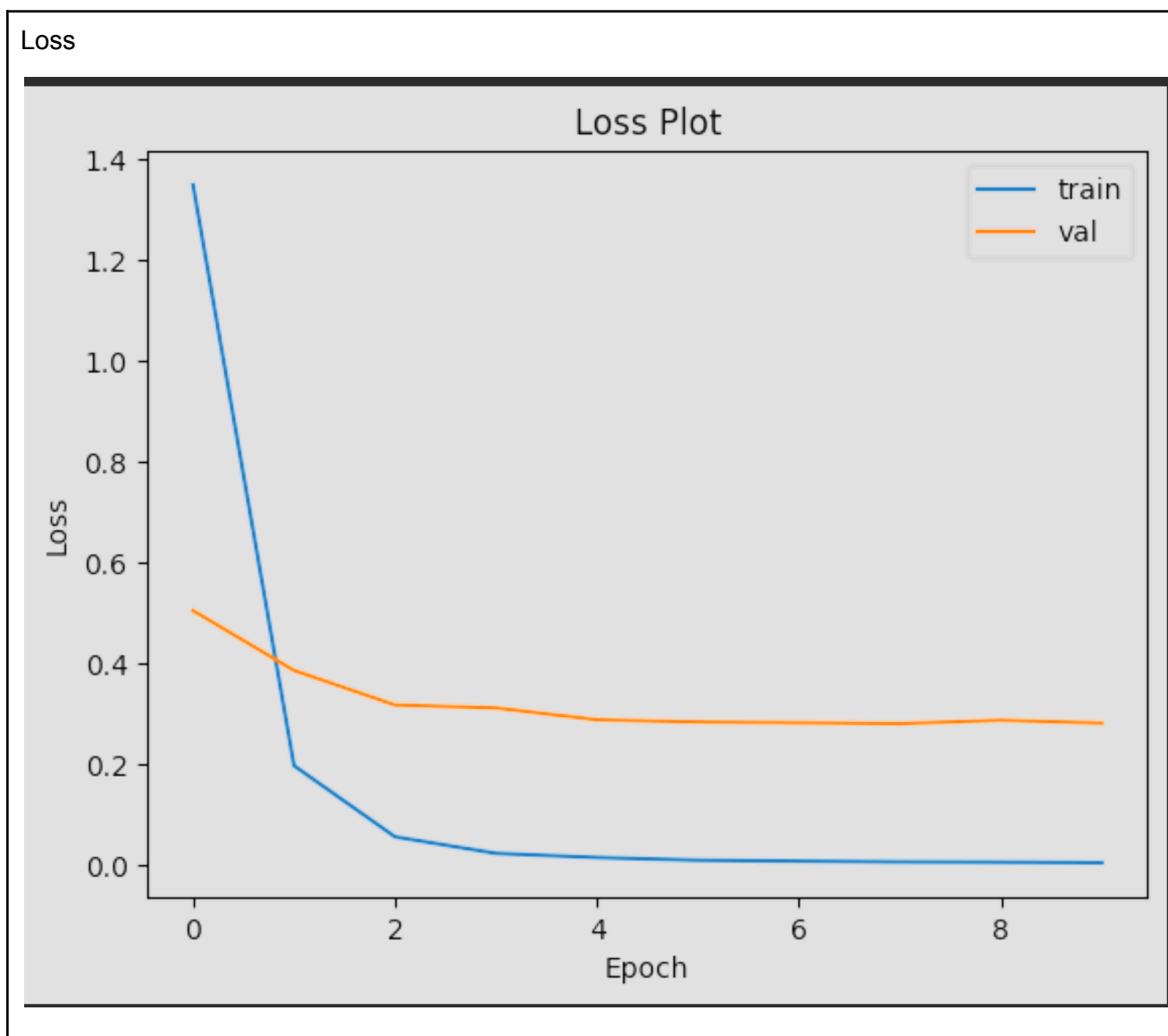
False

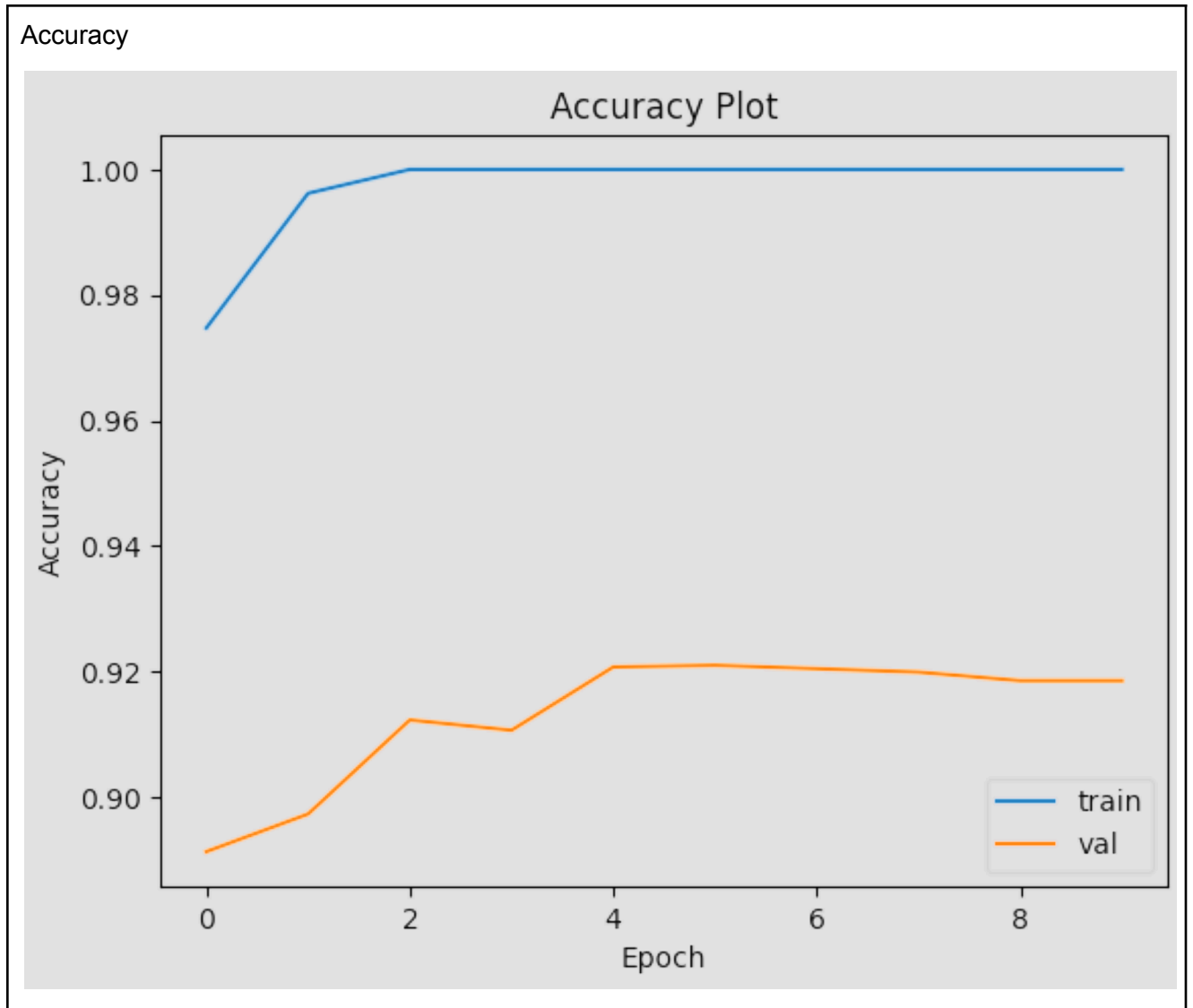
1.4. Which of these are applied immediately before the final fully connected layer? (choose one)

- a) Convolutional
- b) BatchNorm
- c) ReLU
- d) Max pooling
- e) Average pooling
- f) Fully connected

e

2. Plot accuracy and loss for at least 10 epochs





3. Best accuracy / question

Your best val (test_set) accuracy:

0.92095938947942
22

True or False: Once the training accuracy reaches 100%, it's not possible to improve the model with further training.

False

3. CLIP: Contrastive Language-Image Pretraining

1. CLIP zero-shot performance

Your test accuracy (xx.x%)

68.6%

What is the key idea that provides zero-shot ability to CLIP? (choose one)

- a. The visual model is trained to predict the most likely word based on a large dataset of labeled images.
- b. A text model is trained to map words into a vector and a visual model to map patches into an equal sized vector, such that the vectors of image and its textual description are much more similar than those of non-corresponding images and descriptions.
- c. CLIP learns to generate the most likely textual description, given the image.

b

2. CLIP linear probe performance

Your test accuracy (xx.x%)

93.4%

3. KNN on CLIP feature

Your test accuracy: (xx.x%)

85.2%

Best K:

7

4. Stretch Goals

a. Compare word tokenizers

Report encodings for “I am learning about word tokenizers. They are not very complicated, and they are a good way to convert natural text into tokens.” and one additional sentence of your choice. 20 points for reporting trained encodings of at least two models. 10 points for one model. You must train the models on WikiText-2 (should be included in notebook code).

I have problem getting the raw files, I went the office hour but still didn't get the solution

b. Custom network implementation and evaluation

i. Display the structure of the network you implemented

```
...
=====
- layer0                :                9600
- layer1                :               148224
- layer2                :               378112
- layer3                :              1509888
- layer4                :             6034432
- gap                   :                  0
- fc                    :               18981
=====
--Total                  :             8099237 params    --
Network(
  (layer0): Sequential(
    (0): Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3))
    (1): MaxPool2d(kernel_size=3, stride=2, padding=1, dilation=1, ceil_mode=False)
    (2): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (3): ReLU()
  )
  (layer1): Sequential(
    (0): ResBlock(
      (conv1): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (shortcut): Sequential()
      (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (bn1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
      (bn2): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    )
    ...
  )
  (gap): AdaptiveAvgPool2d(output_size=1)
  (fc): Linear(in_features=512, out_features=37, bias=True)
)
```

ii. Plot accuracy and loss

iii. Best test accuracy:

Acknowledgments / Attribution

Asked professor for help regarding fine tuning.