Reyhane Shahrokhian 99521361

HomeWork6 of DeepLearning Course

Dr. DavoodAbadi

Question 1:

***1-1**:*

Convolutional layers can detect local patterns like edges, textures, etc using filters to convolve over the input image. Also they can capture more complex patterns like the eye or face of the cat. They work great at learning hierarchical features in images which lead to a high performance network in this task.

The attention mechanism helps the model to assign weights to different regions of the image, allowing it to pay more attention to important features and ignore irrelevant or distracting information. In this task it will be useful to detect features related to a cat.

***1-2**:*

CNNs work by detecting the presence of features in an image, and using the knowledge of these features to predict whether an object exists. For CNNs, both of these pictures are almost the same. CNN does not encode the relative position of different features. Large filters are required to encode the combination of these features.

Self-Attention modules help to model long-range dependencies without compromising on computational and statistical efficiency. The self-attention module is complementary to

convolutions and helps with modeling long range, multi-level dependencies across image regions.

I use this site.

Question 2:

***2-1**:*

TP: when a model predicts correctly that a sample belongs to class positive.

TN: when a model predicts correctly that a sample belongs to class negative.

FP: when a model predicts wrongly that a sample belongs to class positive but it actually belongs to class negative.

FN: when a model predicts wrongly that a sample belongs to class negative but it actually belongs to class positive.

***2-2**:*

I will use F1-score as the evaluation metric. The reason for that is that the probability of being a criminal among all people is really low but also finding that person is really important for us according to data security.

This issue is exactly the same as the example in our slides which was about detecting a rare disease. If the model predicts everyone is innocent, then the precision will be 1 and Recall will be 0. And if the model predicts no one is innocent, then the precision will be near 0 and Recall will be 1. So they aren't suitable metrics for this problem, however F1-score considers a combination of them which leads to finding the real criminals:).

Question 3:

***3-1**:*

Rotation estimation can be a valuable preprocessing step in classification tasks, particularly when dealing with data that exhibits variability in orientation. It helps improve the model's ability to generalize across different orientations, leading to more robust and accurate classification performance.

***3-2**:*

A one-hot vector is a binary vector representation where all elements are zero except for one, which is marked with a value of 1. One-hot encoding is commonly used to represent categorical variables, especially in machine learning and natural language processing tasks. However, there are some problems with using one-hot encoding:

- High Dimensionality and Sparse:

  As the number of words in our vocabulary increases, the dimensionality of the one-hot encoded vectors also increases. This can lead to a high-dimensional and sparse representation, making it computationally expensive and memory-intensive, especially when dealing with large datasets.

- Lack of meaningful distance:

  One-hot encoding treats all categories as orthogonal and equally distant from each other. It does not capture any semantic similarity or relationships between words of our vocabulary.

***3-3**:*

Word2Vec is often considered a self-supervised algorithm because it learns meaningful representations of words in a continuous vector space without explicit supervision from labeled data. The model leverages the inherent structure and patterns in the data to generate its own supervision signal for training.

Question 4:

***4-1**:*

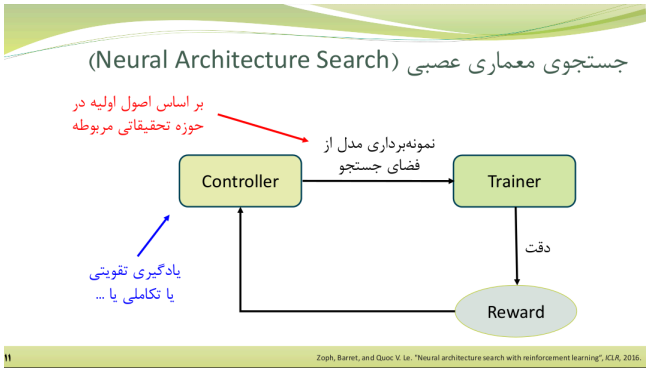
Reinforcement Learning (RL) in Neural Architecture Search (NAS) for setting hyperparameters is an approach that leverages the principles of reinforcement learning to automate the process of finding optimal neural network architectures and their associated hyperparameters.

The first step is to define an objective function or reward signal that quantifies the performance of a neural network on a given task.

Then a search space that encompasses various neural network architectures and hyperparameter configurations should be defined. This space can include options such as the number of layers, types of layers, learning rates, dropout rates, and other relevant hyperparameters.

The agent in RL is responsible for selecting architectures and hyperparameters from the search space. The agent's goal is to maximize the reward signal obtained from the objective function.

The schematic of that is from our slides and the RL agent is called controller:



***4-2**:*

RL-based NAS can automatically search for neural network architectures that are adaptable to different input image sizes. This is important for object detection tasks where images may vary in resolution or aspect ratio. The agent can explore architectures that dynamically adjust to different input sizes, allowing for better generalization and handling of images with various dimensions.

RL-based NAS can explore different architectures with varying numbers of layers to find the optimal depth for the given object detection task. Too few layers may result in insufficient feature representation, while too many layers may lead to overfitting.

Question 5:

The GAN training is inherently unstable because of simultaneous dynamic training of two competing models. Their performances are improved in parallel while epochs paas. So they both became more powerful and that's why the loss of epoch 1 and epoch 100 is approximately the same but the quality of pictures are not the same.