Batch Size VS Steps per Epoch

The batch size and steps per epoch are both important parameters that affect the training process of a neural network. Let's discuss each of them individually:

Batch size

The batch size refers to the number of samples that are processed by the neural network in one forward and backward pass. During training, the input data is divided into batches, and the gradients are computed and updated based on the average of the loss for each batch. The batch size can have a significant impact on training dynamics and convergence.

- Larger batch sizes can provide computational efficiency by taking advantage of parallel processing and optimized hardware. They can also lead to more stable gradients due to the increased amount of data in each update step. However, larger batch sizes require more memory and might lead to slower convergence and potentially worse generalization.
- Smaller batch sizes can provide better generalization as they introduce more randomness and noise into the training process. They also require less memory, which can be beneficial for models with limited resources. However, smaller batch sizes may result in less stable gradients due to the increased variance in each update step and can lead to slower training convergence.

Steps per epoch

The steps per epoch refers to the number of iterations required to complete one epoch. An epoch is defined as a complete pass through the entire training dataset. In practice, due to memory limitations, it's common to process the data in smaller batches and update the model's parameters multiple times within an epoch. The number of steps per epoch is determined by the dataset size and the batch size.

- If the dataset is larger, and the batch size is relatively small, the number of steps per epoch will be higher. Each step involves processing one batch of data and updating the model's parameters accordingly. The purpose of having multiple steps per epoch is to ensure that the model has seen and learned from a diverse set of data points before completing an epoch.

- If the dataset is smaller or the batch size is larger, the number of steps per epoch will be lower. In extreme cases, where the batch size is equal to the dataset size, there will be only one step per epoch since the entire dataset is processed in a single pass.

It's important to tune both the batch size and steps per epoch based on the available computational resources, memory constraints, and the characteristics of the dataset. Finding the optimal values for these parameters often involves experimentation and can vary depending on the specific deep learning task and architecture.