Intel® Optimized Cloud Modules for AWS*: Stable Diffusion Distributed Training

Discover the power of AI model optimization with the Intel® Optimized Cloud Modules for AWS, designed explicitly for the Stable Diffusion model. This hands-on module guides you through fine-tuning the cutting-edge Stable Diffusion v1.5 model for textual-inversion tasks, harnessing the robust capabilities of 4th Generation Intel® Xeon® Scalable Processors on AWS.

Intel ® Advanced Matrix Extensions

Launch an EC2 instance and open it up in a command prompt. You can do so from the AWS console with the instructions that are found here.

If you are using a 4th Geneneration Xeon CPU, you can verify that you have the AMX instruction set by running:

```
lscpu | grep amx
```

and you should see the following flags:

```
amx_bf16 amx_tile amx_int8
```

Intel Extension for PyTorch

The Intel Extension for PyTorch elevates PyTorch performance on Intel hardware with the integration of the newest features and optimizations that have not yet been incorporated into open source PyTorch.

```
import intel_extension_for_pytorch as ipex

unet = ipex.optimize(unet, dtype=weight_dtype)

vae = ipex.optimize(vae, dtype=weight_dtype)
```

GitHub Repo | Docs | Examples

Configure Accelerate

When configuring the multi-CPU setup using accelerate config, you will be prompted with several questions. To select the appropriate answers based on your environment. Make sure you select IPEX and bf16 in the configuration wizard.

```
Do you want to use Intel PyTorch Extension (IPEX) to speed up training on CPU? [yes/NO]:yes
```

```
Do you wish to use FP16 or BF16 (mixed precision)?
bf16
```

GitHub Repo | Docs | GitHub Repo

Next Steps:

All Cloud Modules | GitHub Repo | DevHub Discord

*Names and brands may be claimed as the property of others.