

Object Detection TPU Inference Exporter



This package contains SavedModel Exporter for TPU Inference of object detection models.

Usage

This Exporter is intended for users who have trained models with CPUs / GPUs, but would like to use them for inference on TPU without changing their code or re-training their models.

Users are assumed to have:

- `PIPELINE_CONFIG` : A `pipeline_pb2.TrainEvalPipelineConfig` config file;
- `CHECKPOINT` : A model checkpoint trained on any device;

and need to correctly set:

- `EXPORT_DIR` : Path to export SavedModel;
- `INPUT_PLACEHOLDER` : Name of input placeholder in model's `signature_def_map`;
- `INPUT_TYPE` : Type of input node, which can be one of 'image_tensor', 'encoded_image_string_tensor', or 'tf_example';
- `USE_BFLOAT16` : Whether to use bfloat16 instead of float32 on TPU.

The model can be exported with:

```
python object_detection/tpu_exporters/export_saved_model_tpu.py \  
  --pipeline_config_file=<PIPELINE_CONFIG> \  
  --ckpt_path=<CHECKPOINT> \  
  --export_dir=<EXPORT_DIR> \  
  --input_placeholder_name=<INPUT_PLACEHOLDER> \  
  --input_type=<INPUT_TYPE> \  
  --use_bfloat16=<USE_BFLOAT16>
```