Hugging Face

Evan Wong, Tanner Scott, Aaron Stern

Hugging Face Transformers

Hugging Face is a collaborative open-source platform for building, training, and deploying machine learning (ML) and natural language processing (NLP) models.

Computer Vision

A computing field that seeks to allow computers to identify and interpret objects in visual media, typically images and video. As a specific form of artificial intelligence, it attempts to replicate human ability to automate tedious and time-consuming tasks.

Tasks Explored

Image Classification
Image Segmentation

Object Detection
Video Classification

Resolving Issues

- **#27341** Generate: should softmax be upcasted to .float()?
  - Often operations within models are unstable at lower precisions, so many times these operations, like softmax, are forced to run at 32-bit floating point precision.
  - One contributor asks if this casting to 32-bit FP was necessary as this cast may increase running time.
  - Created tests of running time and accuracy. Running time did take longer but accuracy was not significantly changed.
  - The casting seems not to be necessary, but the difference is negligible, so the other contributors decide not to make any changes.

```
Default Times (softmax w/ default dtype param):
Median: 19.004344940185547
Mean: 22.11885714530945
Standard Deviation: 11.286693376217634
Minimum: 12.001037596756625
Maximum: 46.010732650756836

Forced Times (softmax w/ float32 dtype param):
Median: 21.004796028137207
Mean: 23.721716165542603
Standard Deviation: 11.330417161274704
Minimum: 11.969327926635742
Maximum: 43.009757999560547
```

```
Mismatches: 24706
Avg diff: 0.000000e+00
Max diff: 9.536743e-07
Time 16bit: 5.357883e-06 seconds
Time 32bit: 1.586733e-05 seconds
```

- **#29790** Missing preprocessor_config.json file after training segformer model
  - Trainer class was originally designed for text-only models, so image preprocessing was necessary
  - Parameter name was updated from tokenizer to image_processor
  - Default image_processor is now included in the trainer

Documentation Improvements

#29817 Video Classification Task Guide Using Undeclared Variables
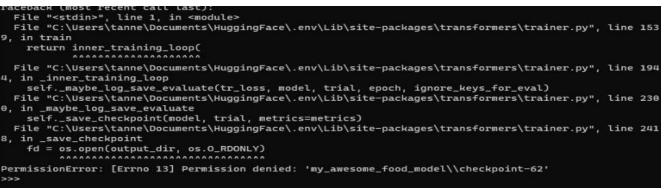- Missing Dependencies: Pillow
- Undeclared Variables: all_video_file_paths, dataset_root_path
- Deprecated Modules: torchvision.transforms.functional_tensor module is deprecated in 0.15 and will be **removed in 0.17**

```
>>> class_labels = sorted({str(path).split("/")[2] for path in all_video_file_paths})
>>> label2id = {label: i for i, label in enumerate(class_labels)}
>>> id2label = {i: label for label, i in label2id.items()}

>>> print(f"Unique classes: {list(label2id.keys())}.")

# Unique classes: ['ApplyEyeMakeup', 'ApplyLipstick', 'Archery', 'BabyCrawling', 'BalanceBeam
```

#29382 Permission Denied Error When Saving Checkpoint During Training
- Permission error occurred during training for image classification.
- Bug exists in code that saves checkpoints during training.
- Known issue on the forum, but not resolved in the repository.
- Another contributor argued that atomic checkpointing is unmaintainable and had the feature removed, solving the issue.

```
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "C:\Users\tanne\Documents\HuggingFace\.env\Lib\site-packages\transformers\trainer.py", line 153
9, in train
    return inner_training_loop(
           ^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\tanne\Documents\HuggingFace\.env\Lib\site-packages\transformers\trainer.py", line 190
4, in _inner_training_loop
    self._maybe_log_save_evaluate(tr_loss, model, trial, epoch, ignore_keys_for_eval)
  File "C:\Users\tanne\Documents\HuggingFace\.env\Lib\site-packages\transformers\trainer.py", line 238
0, in _maybe_log_save_evaluate
    self._save_checkpoint(model, trial, metrics=metrics)
  File "C:\Users\tanne\Documents\HuggingFace\.env\Lib\site-packages\transformers\trainer.py", line 241
8, in _save_checkpoint
    fd = os.open(output_dir, os.O_RDONLY)
         ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
PermissionError: [Errno 13] Permission denied: 'my_awesome_food_model\\checkpoint-62'
```

Goals

- Help resolve reported problems to improve the platform for users.
- Improve documentation to make it easier for new users to learn how to use the platform, in turn improving the accessibility of the platform for everyone.

Issues Worked On

- **#13244** Tapas tokenization Different from TensorFlow Code
- **#21110** Add support for BLIP and GIT in image-to-text and VQA pipelines.
- **#27341** Generate: should softmax be upcasted to .float()?
- **#28427** RagRetriever download too much data and won't stop
- **#29382** Permission Denied Error When Saving Checkpoint During Training
- **#29790** Missing preprocessor_config.json file after training segformer model
- **#29817** Video Classification Task Guide Using Undeclared Variables
- **#30058** Image Segmentation Guide missing dependencies/imports and overwritten variables
- **#30059** Image Classification Guide Missing Dependency and Addition for Local Models
- **#30060** Zero-Shot Image Classification Guide Missing Dependency and Addition for Local Inputs

Challenges

- The steep learning curve of understanding the platform and its repository.
- Issues that we would be capable of resolving are closed quickly.
- Out-of-date documentation is hard to follow.
- Finding missing dependencies when following documentation means we must completely restart the process.