

The vote is over, but the fight for net neutrality isn't. Show your support for a free and open internet.

Learn more

tensorflow / tensorflow

[Feature Request]: fold batch norm into convolution weights of graph transform tool #12740

New issue

Closed

ydzhang12345 opened this issue on 1 Sep 2017 · 4 comments



ydzhang12345 commented on 1 Sep 2017 • edited

System information

- Have I written custom code (as opposed to using a stock example script provided in TensorFlow): No
- OS Platform and Distribution (e.g., Linux Ubuntu 16.04): Linux Ubuntu 16.04
- TensorFlow installed from (source or binary): source
- TensorFlow version (use command below): 1.3.0
- Python version: 2.7
- Bazel version (if compiling from source): 0.5.1
- CUDA/cuDNN version: 8.0, 6.0
- GPU model and memory: GTX1080
- Exact command to reproduce:

```
bazel build tensorflow/tools/graph_transforms:transform_graph
bazel-bin/tensorflow/tools/graph_transforms/transform_graph
--in_graph=tensorflow_inception_graph.pb
--out_graph=optimized_inception_graph.pb
--inputs='Mul'
--outputs='softmax'
--transforms='
strip_unused_nodes(type=float, shape="1,299,299,3")
remove_nodes(op=Identity, op=CheckNumerics)
fold_constants(ignore_errors=true)
fold_batch_norms
fold_old_batch_norms'
```

After diving into the generated pb graph, I found that the current fold batch norm ops seems to only fold batch norm into two ops: mul and add, which is not exactly as the readme demonstrated: fold the batch norm into the weights of convolution. If this is the case, it would be better to further merge the mul and add into conv weights.

Assignees

petewarden

Labels

None yet

Projects

None yet

Milestone

No milestone

Notifications

5 participants

michaelisard assigned petewarden on 2 Sep 2017



carlthome commented on 4 Sep 2017

Contributor

On a similar note, what about folding [this](#) as well? I.e. fold across `array_ops.split` .



passerbyd commented on 12 Sep 2017

fold\_batch\_norms/fold\_old\_batch\_norms merges the scale part of a (fused) batch normalization into conv\_op only if the bn op is trailing a conv op. Otherwise nothing is did. And the tool run as expected in this case. You can check its implementation.



**tensorflowbutler** commented 15 days ago


Member

It has been 14 days with no activity and this issue has an assignee.Please update the label and/or status accordingly.



**ydzhang12345** commented 15 days ago

resolved

 **ydzhang12345** closed this 15 days ago