ディープラーニング向け インテル® ソフトウェア製品の紹介 (II)

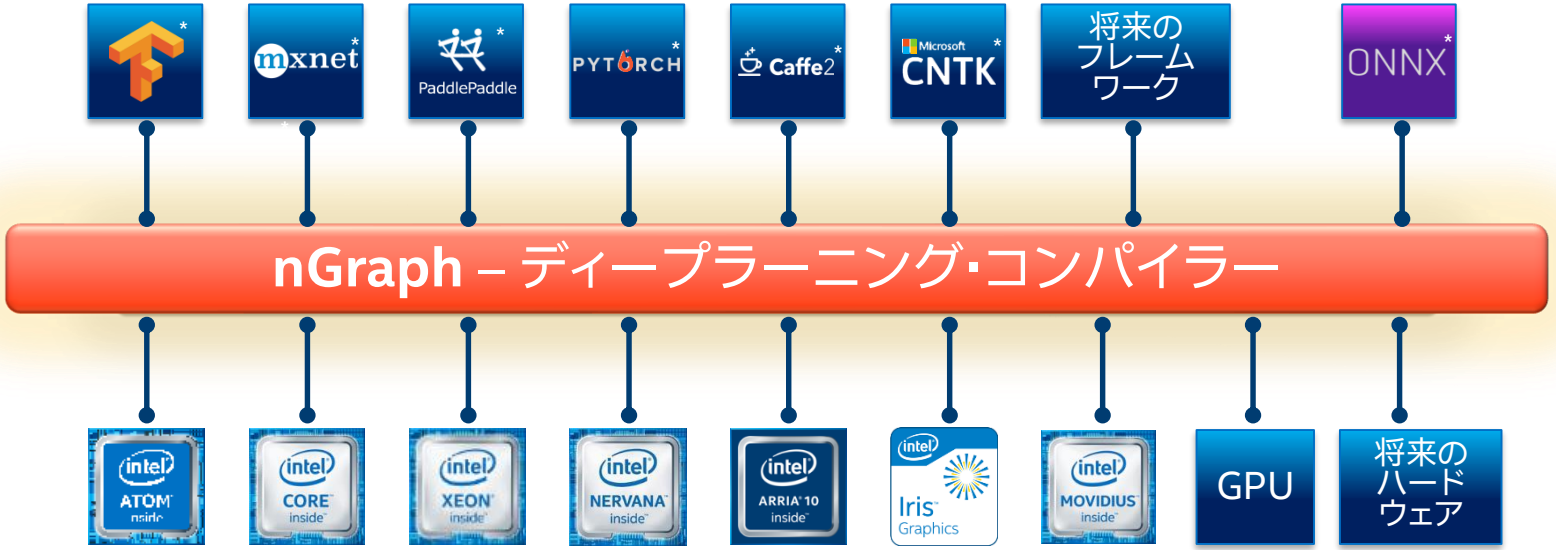
ディープラーニングのツール

nGraph

インテル® nGraph コンパイラー

アルファ版
提供中

さまざまなフレームワークやハードウェア上でモデルの柔軟な実行を可能にするオープンソースのコンパイラー



https://www.intel.ai/ngraph/ (英語)

記載されているすべての製品、コンピューター・システム、日付、および数値は、現在の予想に基づくものであり、予告なく変更されることがあります。

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® パフォーマンス・ライブラリー

ディープ・ニューラル・ネットワーク向けインテル® マス・カーネル・ライブラリー
(インテル® MKL-DNN)
インテル® Machine Learning Scaling Library (インテル® MLSL)

ディープ・ニューラル・ネットワーク向け インテル® マス・カーネル・ライブラリー

(インテル® MKL-DNN)

インテル® MKL-DNN

インテルのオープンソースである
ディープ・ニューラル・ネットワーク向けマス・カーネル・ライブラリー

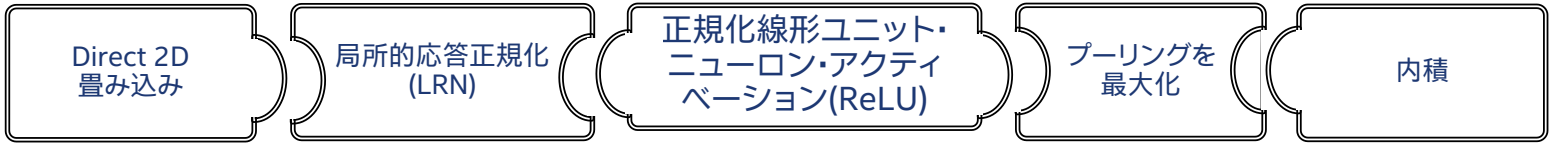
インテルのハードウェア上で最適化するディープラーニング・フレームワークの開発者用

ディストリビューションの詳細

- オープンソース
- Apache* 2.0 ライセンス
- すべてのインテル® ハードウェアで共通の DNN API
- 業界のフレーム統合を推進するために、ディープラーニングのコミュニティーで 迅速なリリースを繰り返す
- インテル® MKL を基に高度なベクトル化とスレッド化による最高性能

github.com/01org/mkl-dnn (英語)

例:



記載されているすべての製品、コンピューター・システム、日付、および数値は、現在の予想に基づくものであり、予告なく変更されることがあります。

最適化に関する注意事項

© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

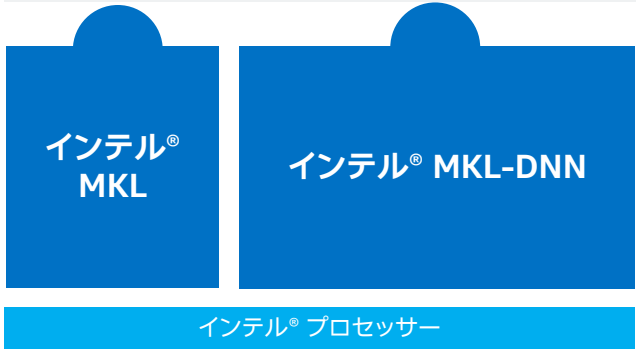
インテル® プロセッサー向け ディープラーニング・ソフトウェア・スタック

ディープラーニングと AI エコシステム には端末とデータセンターのアプリケーションが含まれる

- オープンソース・フレームワーク (TensorFlow*、MXNet*、CNTK、PaddlePaddle*)
- インテルのディープラーニング製品 (neon™ フレームワーク、BigDL、OpenVINO™ ツールキット)
- 内製のユーザー・アプリケーション

インテル® MKL とインテル® MKL-DNN はディープラーニング・アプリケーションをインテル® プロセッサーに最適化

- 仕様の変更にフレームワークの開発者と協同で対応 (TensorFlow*、MXNet*、PaddlePaddle*、CNTK)
- インテルに最適化されたライブラリーを使用 (Caffe、Torch*、Theano*)
- 独自のソリューションを可能



インテル® MKL-DNN はディープラーニング用オープンソースの数値演算ライブラリー (https://github.com/intel/mkl-dnn (英語))

- 広範囲の DNN 関数を実装した高速なオープンソース
- いち早く新しく実験的な機能をテスト
- オープンなコミュニティーへの貢献

インテル® MKL は数値演算アプリケーションに幅広く対応する独自の高性能ライブラリー
ディストリビューション: インテル® レジストレーション・センター、パッケージ・リポジトリー (apt、yum、conda、pip)

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MKL/MKL-DNN を使用した TensorFlow*

インテル® Distribution for Python* の使用

- 例えば MKL_VERBOSE=1 をサポートする多くの NumPy* 計算用にインテル® MKL を使用
- Conda (英語) や YUM (英語)、APT (英語) のパッケージ・マネージャーをサポート

事前にビルドされている TensorFlow* wheel (英語) を使用するか、`bazel build --config=mkl` オプションで TensorFlow* をビルド

- **インテル® VTune™ Amplifier を使用する場合には、ソースプログラムのビルドが必要**
- アフィニティーと内部と外部実行スレッドの数の設定を含む CPU 最適化 (英語) のアドバイスに従う
- インテル® MKL-DNN 関連のさらなる最適化が次のバージョンで予定: 可能であれば最新の TensorFlow* のマスターを使用

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® Distribution of Caffe (インテル® Caffe)

インテルが BVLC Caffe のフォーク (英語) をメンテナンス

CNN向けの最速の CPU フレームワーク

インテル® Xeon® スケーラブル・プロセッサー (開発コード名 Skylake) の数値精度の低い推論 (英語) をサポート

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MKL-DNN 概要

機能:

- 訓練 (float32) と 推論 (float32、int8)
- CNN (1D、2D と 3D)、RNN (plain、LSTM、GRU)
- インテル® プロセッサーに最適化

ポータビリティー:

- コンパイラー: インテル® C++ コンパイラー/Clang/GCC/M
- OS: Linux*、Windows*、macOS*
- スレッド化: OpenMP*、インテル® TBB

プリミティブ	クラス
• (逆)畳み込み • 内積 • Vanilla RNN、LSTM、GRU	計算集約的操作
• プーリング 平均/最大 • バッチ 正規化 • 局所反応正規化 • 活性化関数 (ReLU、Tanh、Softmax、...) • 和	メモリー依存演算
• リオーダー • 連結	データ操作

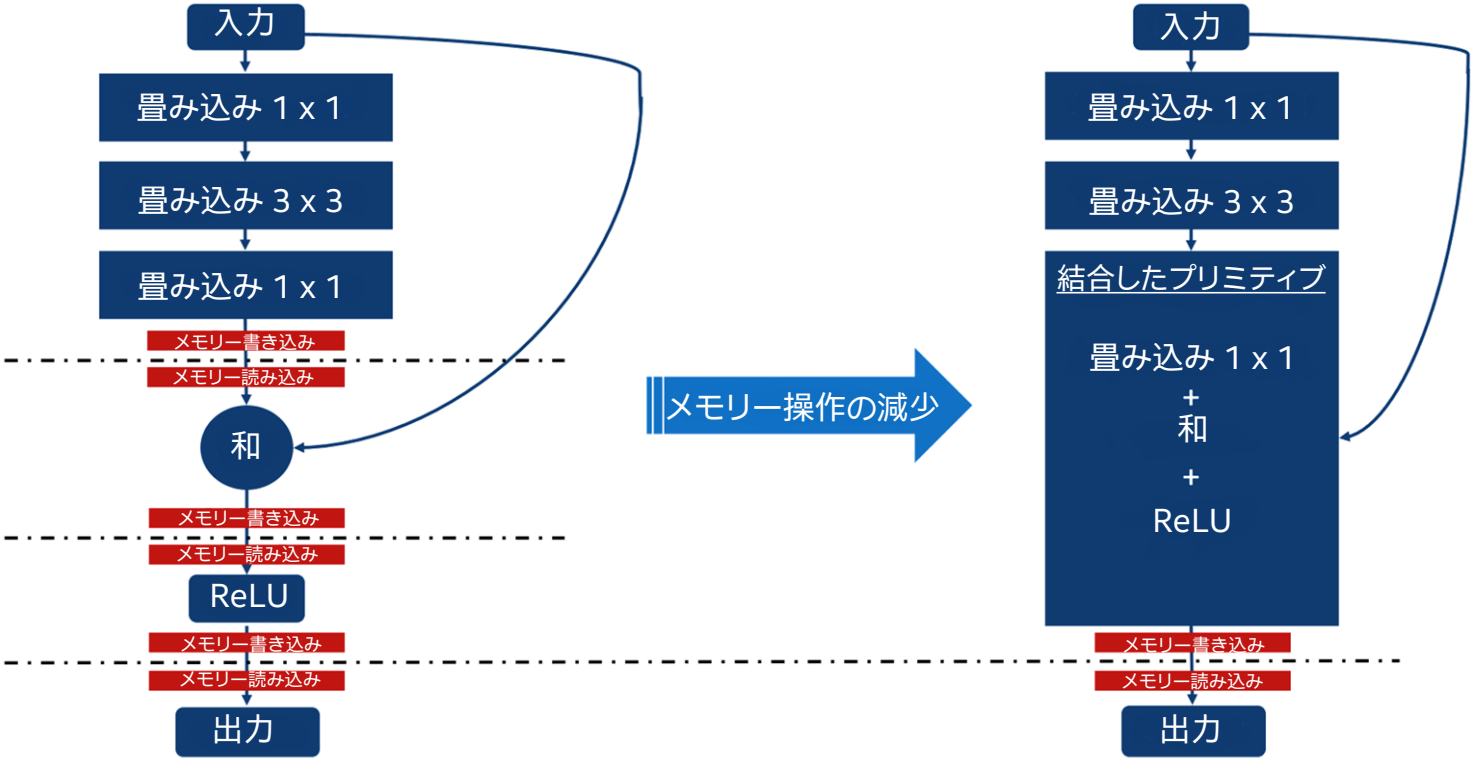
インテル® MKL-DNN を使用しているフレームワーク:

インテル® Caffe、TensorFlow*、MxNet*、PaddlePaddle*、CNTK、OpenVINO™、DeepBench*

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MKL-DNNライブラリーで最適化

レイヤー結合の例



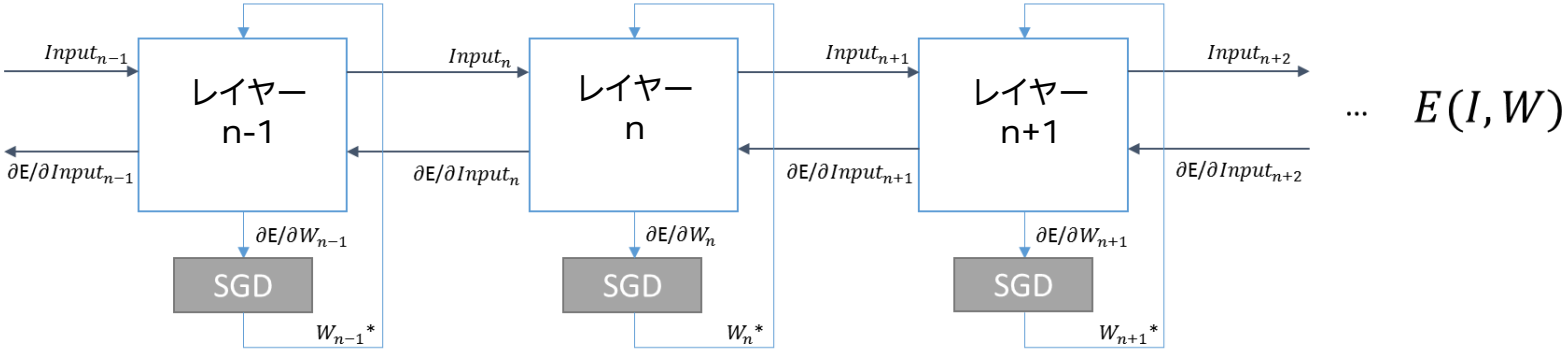
最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® Machine Learning Scaling Library

(インテル® MLSL)

ディープラーニングの訓練

- フォワード伝播: 入力バッチと現在の重みを基に損失関数を計算
- バックワード伝播: (チェーンルールを使って) すべてのレイヤーの重みに関して誤差勾配を計算
- 重みのアップデート: 重みのアップデートに勾配を使用。vanilla SGD、Momentum、Adam などの異なるアルゴリズムが存在。



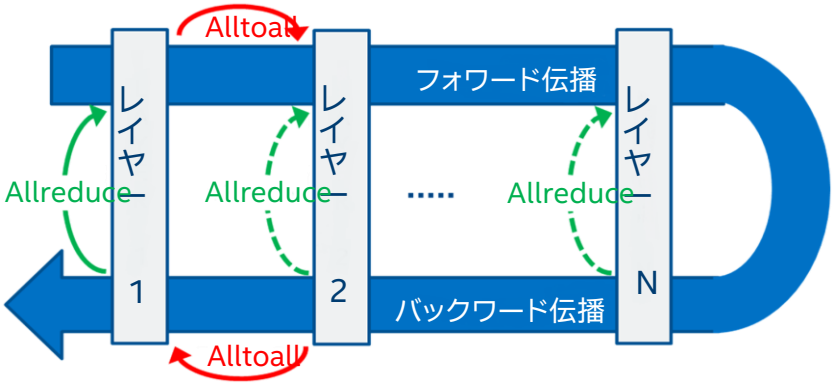
SGD: $W_n^* = W_n - \alpha * \partial E/\partial W_n$ またはバリエーション

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

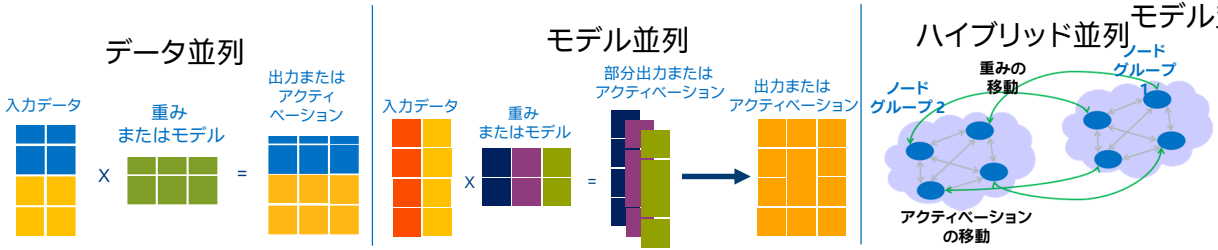
インテル® Machine Learning Scaling Library
(インテル® MLSL)

ディープラーニングのスケールアウト: 必要条件

- ✓ 最適な計算分割の方法を選択
- ✓ バッチサイズの大小に対してスケーラビリティーを有効
- ✓ 通信量の減少
- ✓ 最適な通信アルゴリズムの選択
- ✓ レイテンシー重視の通信を優先
- ✓ CNN、RNN、LSTM などのワークロードをサポート
- ✓ ディープラーニング・フレームワークとの統合



処理の分割方法によって異なる通信
データ並列 = Allreduce (または) Reduce_Scatter + Allgather
モデル並列 = AlltoAll



多くのディープラーニングフレームワーク

多様なネットワークファブリック

イーサネット
Omni-Path　InfiniBand*

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL 2018 Update 3

(https://github.com/intel/MLSL (英語))

ハードウェア

- インテル® Core™ プロセッサー・ファミリー以降
- インテル® Xeon® プロセッサー E5 v3 ファミリー (推奨)
- インテル® Xeon® プロセッサー E7 v2 ファミリー (推奨)
- 第2世代インテル® Xeon Phi™ プロセッサー (開発コード名 Knights Landing)
- メモリー: コアあたり 1GB
  コアあたり 2GB (推奨)
- 必要な空きディスク容量: 1GB

ソフトウェア

OS:

- Red Hat* Enterprise Linux* 6/7
- SuSE* Linux* Enterprise Server 12
- Ubuntu* 16

コンパイラー:

- GNU*: C/C++ 4.4.0 以降
- インテル® C++ コンパイラー for Linux* 16.0 以降

仮想化環境:

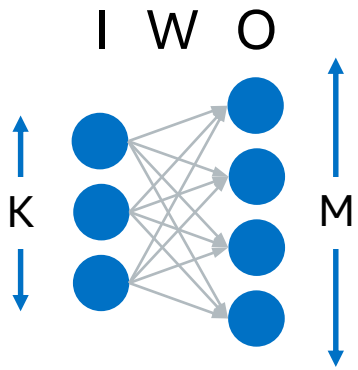
- Docker*
- KVM

Python* インタープリター:

- Python* 2.x

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション

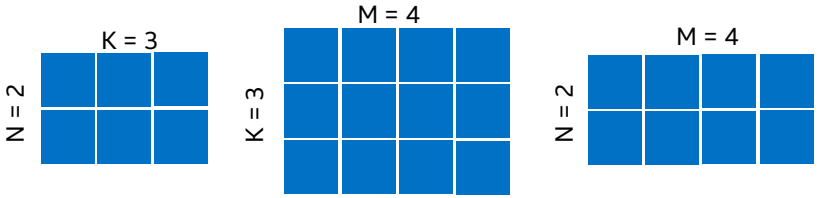


完全に接続されたレイヤー

$I \in R^{NxK}$
入力

$W \in R^{KxM}$
重み
またはモデル

$O \in R^{NxM}$
出力または
アクティベーション



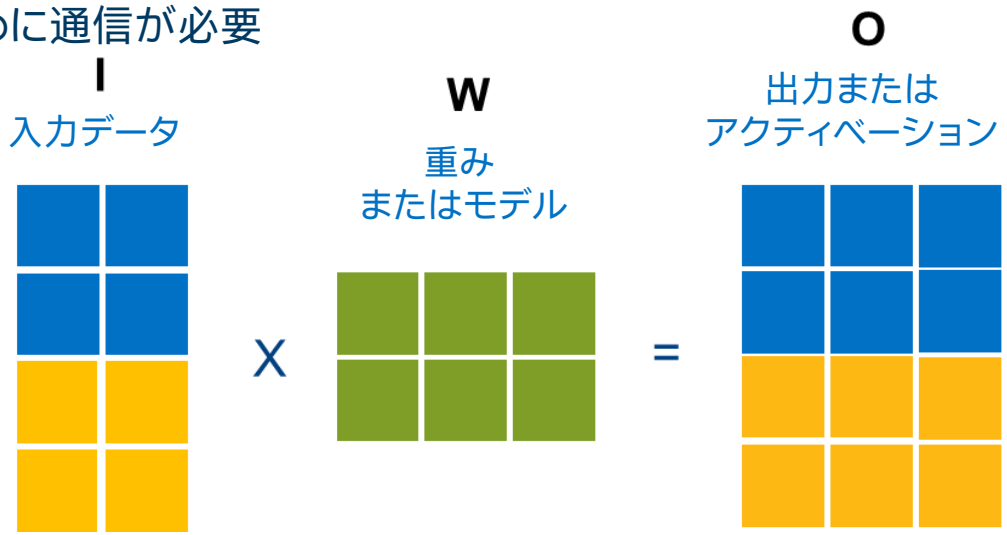
いくつかの並列オプション

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション

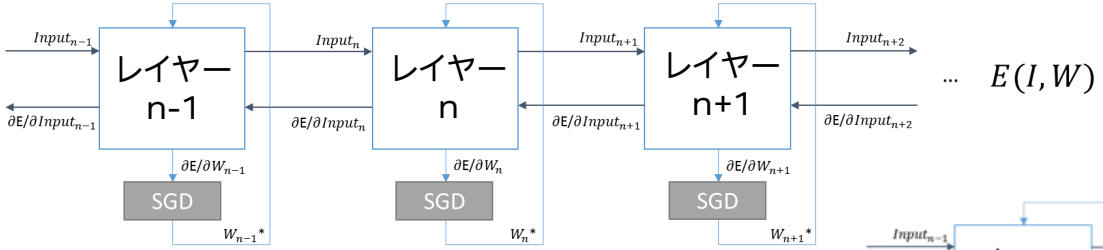
データ並列:

- すべてのノードにモデルをコピー
- 各ノードに分割されたバッチの入力データを送る
- すべてのノードで勾配の平均を得るために通信が必要
- 通信パターンはどちらも可能
  - AllReduce パターン
  - ReduceScatter + AllGather パターン



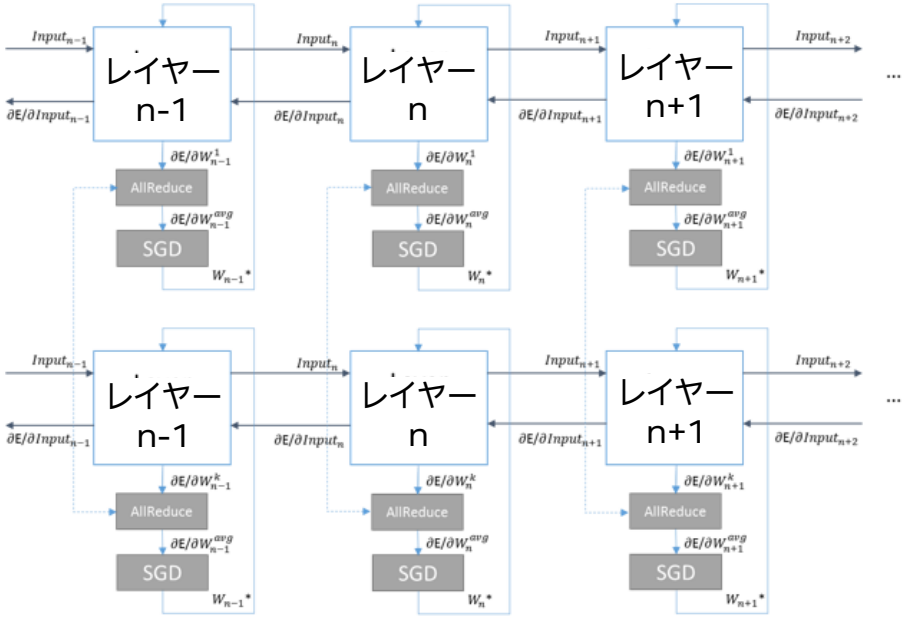
最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション



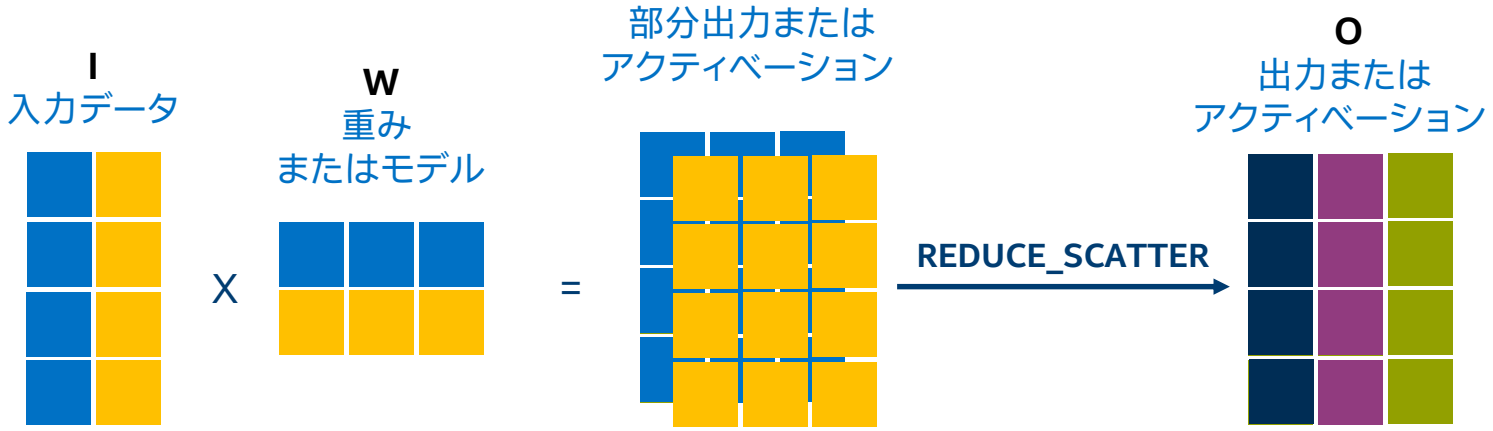
$E(I, W)$

データ並列



最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション

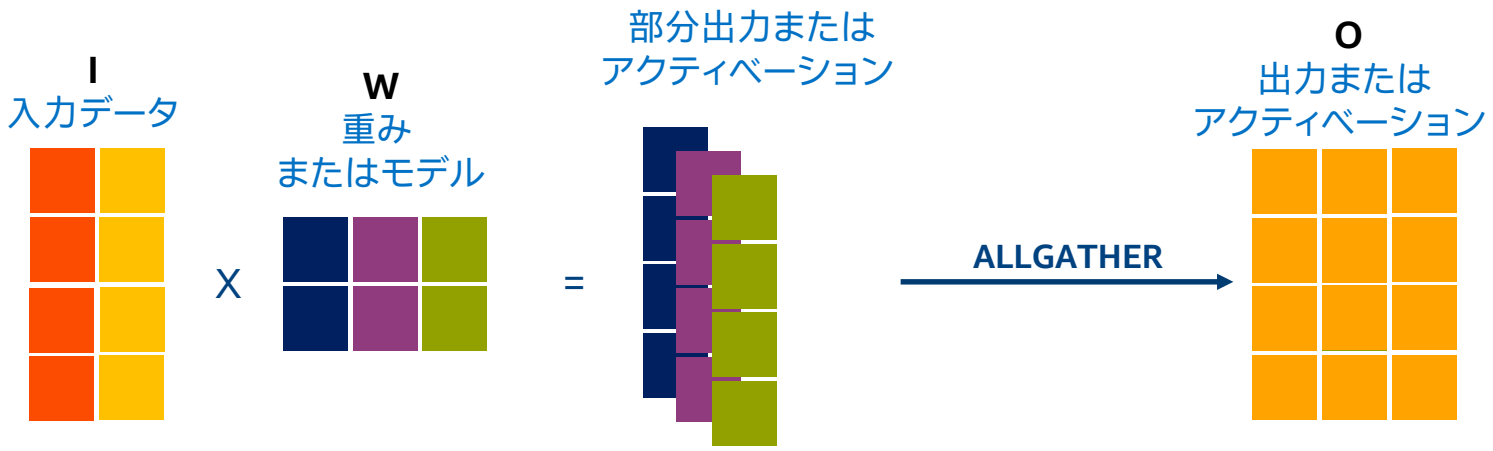


モデルの並列 (#1):

- すべてのノードでモデルを分割
- 入力データのスライスを各ノードに送る
- 次のレイヤーの処理を続けるため部分的なアクティベーションの通信が必要

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション

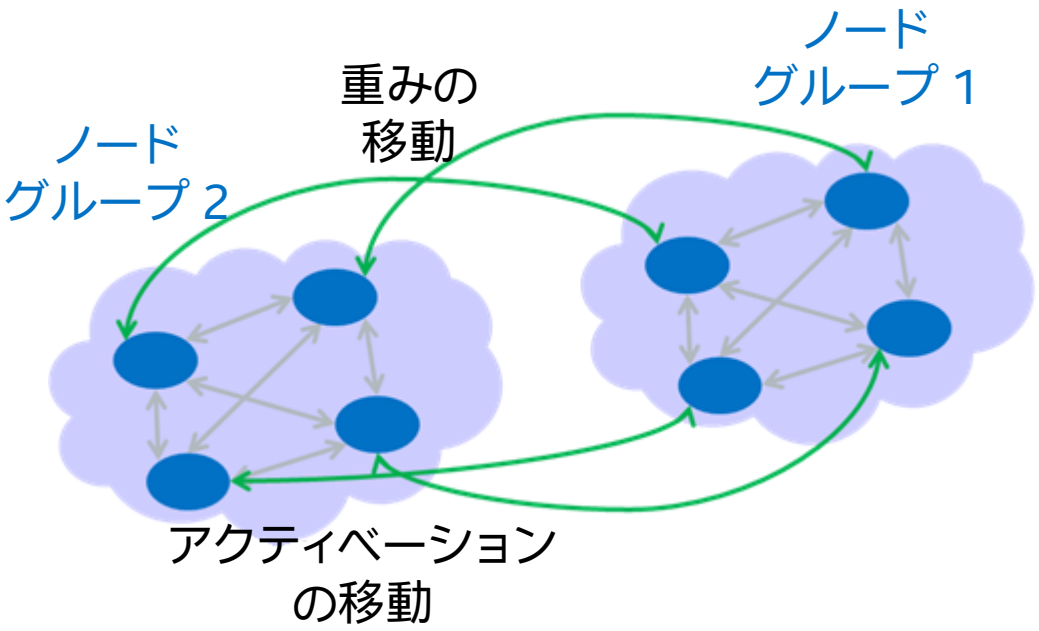


モデルの並列 (#2):

- すべてのノードでモデルを分割
- 入力データの同じバッチを各ノードに送る
- 結果を集めて、さらに処理を続けるために部分的なアクティベーションの通信が必要

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 並列オプション



ハイブリッド並列:

- ノードをグループに分割
- グループ内でモデル並列化
- グループ間はデータ並列化
- 勾配とアクティベーションの両方を通信

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : スケールの並列化

一般的な経験則

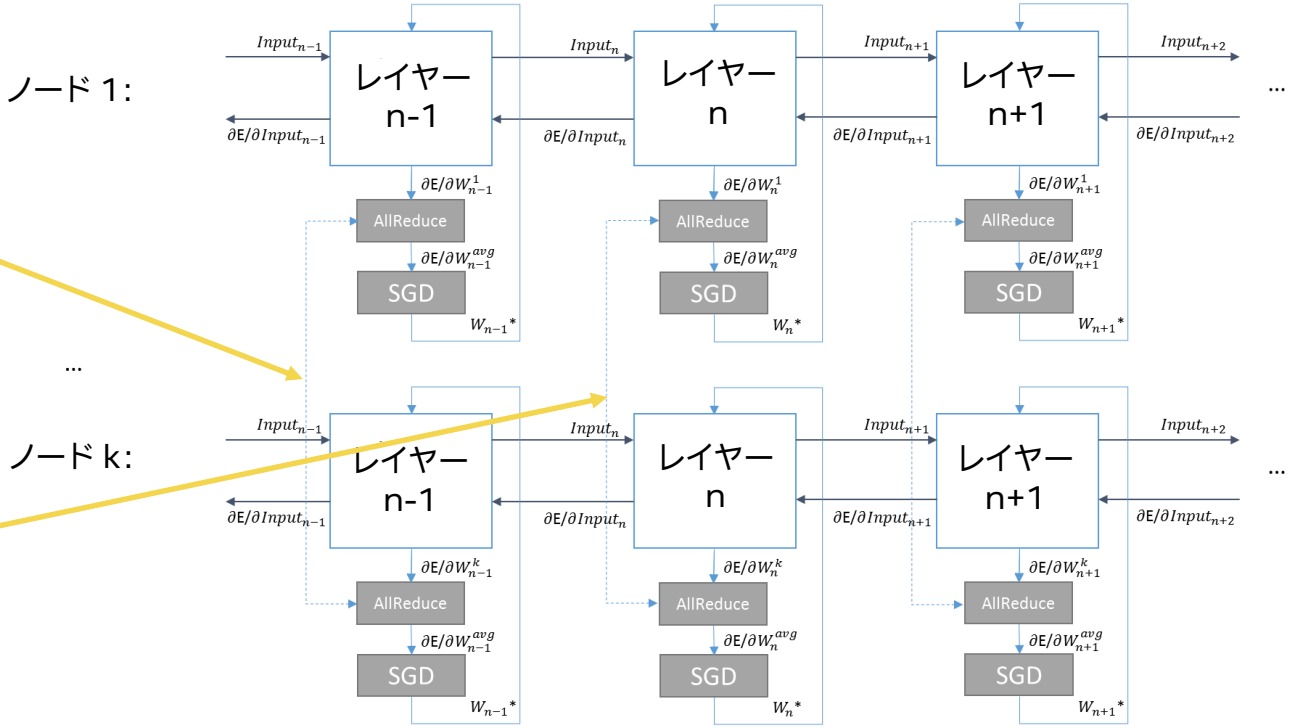
- アクティベーション > 重み の場合にはデータ並列を使用
- 重み > アクティベーション の場合にはモデル並列を使用

データ並列とモデル並列の注意点

- スケール時のデータ配列では アクティベーション << 重み
- スケール時のモデル並列では 重み << アクティベーション
- 通信時間はスケールに影響

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : メッセージの優先度



最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® MLSL : 機能

現バージョンの機能:

- ✓ 非ブロッキング･ディープラーニング･レイヤーと集団インターフェイス
- ✓ Python*/C++/C 対応
- ✓ 非同期通信数列
- ✓ 最適化されたアルゴリズム
- ✓ データ並列、モデル並列、ハイブリッド並列をサポート
- ✓ 量子化を最初にサポート– インテル® Caffe/インテル® MLSL に同梱
- ✓ (環境変数を通して)逆優先順位付けをビルトイン – インテル® Caffe/インテル® MLSL に同梱

将来的な機能 (開発中または調査中):

- ✓ 優先度を明確にする API
- ✓ スパースデータの allreduce
- ✓ 勾配の量子化と圧縮
- ✓ クラウドネイティブ機能

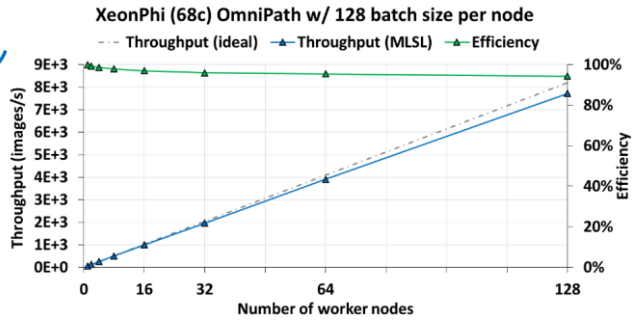
https://github.com/intel/MLSL (英語)

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

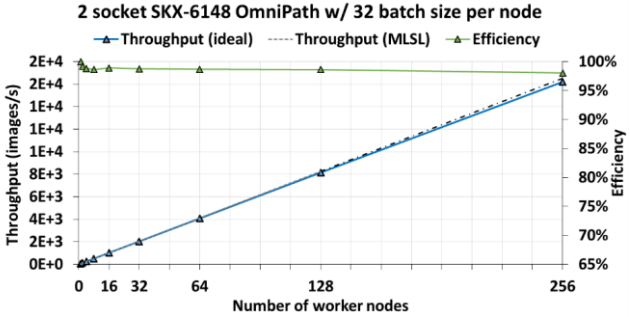
HPC 環境でのスケールアウト

- **インテル® Caffe**: インテル® MLSL ベースのマルチノード・ソリューション **Horovod、nGraph:** 開発中
- インテル® MLSL は Baidu* の DeepBench* に最適化
- SURFSara: ResNet50 time-to-train record (~40 分、768 SKX†) を達成するためにインテル® Caffe/インテル® MLSL を使用[1]
- UC-Berkeley、TACC、UC-Davis: ResNet50 をインテル® Caffe/インテル® MLSL (2048 KNL†) を使用して 14 分で達成[2]

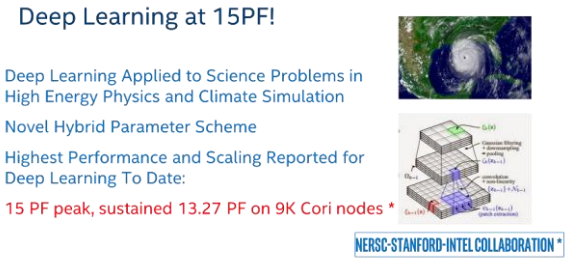
† 開発コード名
[1] https://arxiv.org/pdf/1711.04291.pdf (英語)
[2] https://arxiv.org/pdf/1709.05011.pdf (英語)



IA での TensorFlow* のスケーリング



インテル® Caffe/インテル® MLSL



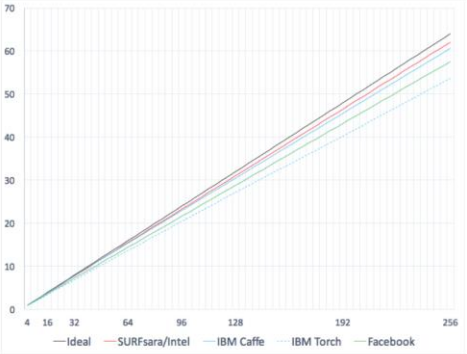
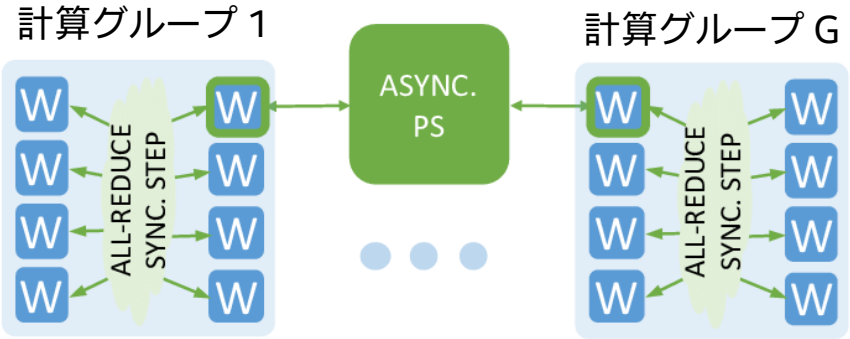


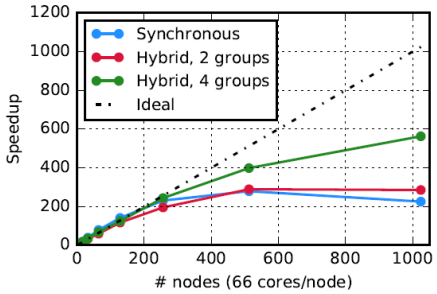
Fig 1. Scaling efficiency on Stampede2 (speedup vs number of workers). This plot starts from scaling on 4 workers, which has a scaling factor of 1.

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

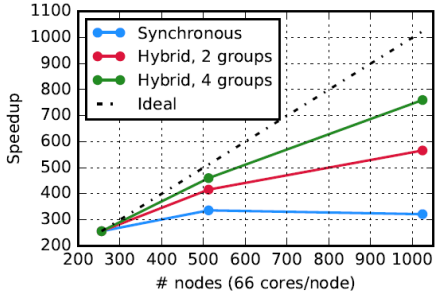
15 ペタフロップス・システム[1]における ディープラーニング

- NERSC、スタンフォード大学、インテルの共同開発
- 分散 SGD への新たなアプローチ : グループ内では同期、グループをまたがるときには非同期動作
- 記録的なスケーリング : 同じモデルを協力して訓練するノード数 (9600 KNL[†])
- ピーク性能 : ~15ペタフロップス
- 通信手段 : ほかのグループとの通信にはインテル® MLSL、グループ内の通信には MPI を使用
- メカニズムはインテル® Caffe/インテル® MLSL で採用





(a) HEP



(b) Climate

† 開発コード名
[1] https://arxiv.org/abs/1708.05256 (英語)

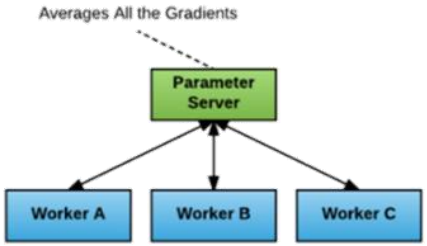
最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

TensorFlow* のスケーリング

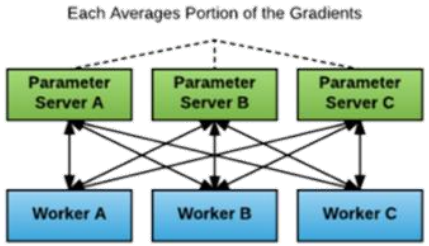
分散 TensorFlow* の比較

パラメーター・サーバーを使った分散 TensorFlow*

パラメーター・サーバーの使用 →



または

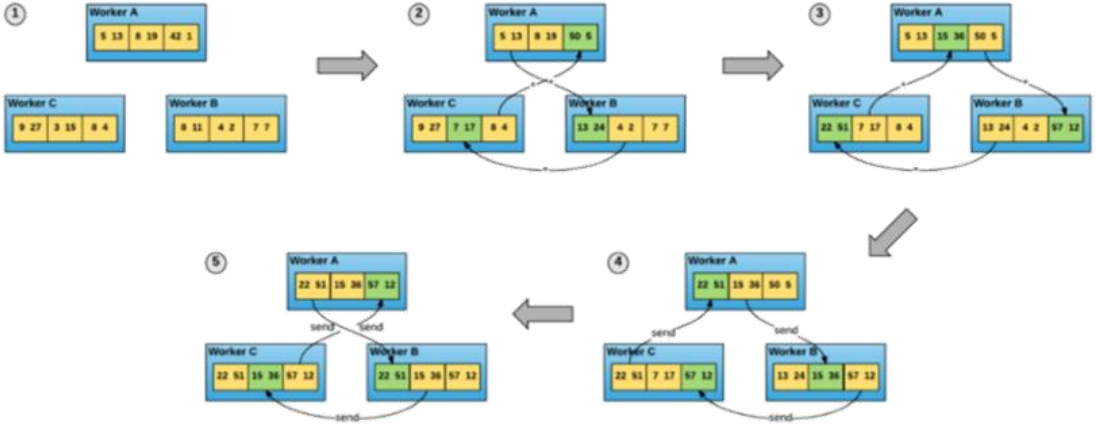


分散訓練処理のパラメーター・サーバー・モデルはパラメーター・サーバーから、それぞれが異なるパフォーマンス・プロファイルを持つワーカーへ異なる重みを設定することが可能

HOROVOD

TensorFlow* 用分散訓練フレームワークに関する Uber* のオープンソース

パラメーター・サーバーなし →

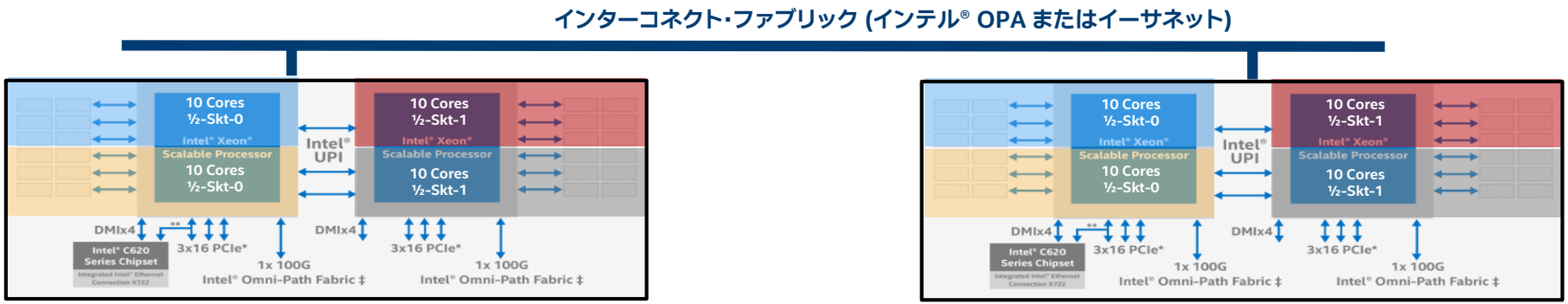


リングの all-reduce アルゴリズムでは、勾配を平均化し、パラメーター・サーバーなしに勾配をすべてのノードに分散させることができる

出典: https://eng.uber.com/horovod/ (英語)

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

分散訓練: Horovod MPI ライブラリーを使ったマルチノード・マルチソケット



マルチノード & マルチソケット上での分散訓練の実行

- パラメーター・サーバーは必要なし
- 各ワーカーノードの各ソケットは 2 つ以上のフレームワーク・ストリームを実行
- Horovod MPI ライブラリー経由でノード間通信

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

Horovod for multinode:

パラメーター・サーバーより (PS):

```
NP=4
PER_PROC=10
HOSTLIST=192.168.10.110
MODEL=inception3
BS=64
BATCHES=100
INTRA=10
INTER=2
```

```
/usr/lib64/openmpi/bin/mpirun --allow-run-as-root -np $NP -cpus-per-proc $PER_PROC –
map-by socket -H $HOSTLIST --report-bindings --oversubscribe -x LD_LIBRARY_PATH python
./tf_cnn_benchmarks.py --model $MODEL --batch_size $BS --data_format NCHW –
num_batches $BATCHES --distortions=True --mkl=True --local_parameter_device cpu –
num_warmup_batches 10 --optimizer rmsprop --display_every 10 --kmp_blocktime 1 –
variable_update horovod --horovod_device cpu --num_intra_threads $INTRA –
num_inter_threads $INTER  --data_dir /home/tf_imagenet --data_name imagenet
```

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

推論

OpenVINO™ ツールキット

(Open Visual Inference & Neural Network Optimization)

OpenVINO™ ツールキットの構成

インテル® ディープラーニング・デプロイメント・ツールキット

モデル・オプティマイザー
変換、最適化

IR

推論エンジン
最適化された推論

20+ 事前訓練済みモデル

コンピューター・ビジョン・アルゴリズム

サンプル

IR = 中間表現フォーマット

従来のコンピューター・ビジョン・ツールとライブラリー

最適化されたライブラリー

OpenCV

OpenVX*

Photography Vision

コードサンプル

インテル® CPU & インテル® グラフィックス向け

メディア / ビデオ / グラフィックスの性能向上

インテル® メディア SDK
オープンソース版

OpenCL*
ドライバー & ランタイム

インテル® グラフィックス搭載 CPU 向け

インテル® FPGA に最適化

FPGA ランタイム環境
(インテル® FPGA SDK for OpenCL*)

Bitstreams

FPGA – Linux* のみ

サポート OS: (64ビット) CentOS* 7.4、Ubuntu* 16.04.3 LTS、Microsoft* Windows® 10、Yocto Project* バージョン Poky Jethro v2.0.3

インテル® アーキテクチャー・ベースのプラットフォームをサポート

intel CELERON inside | intel ATOM inside | intel CORE inside | intel XEON inside | intel ARRIA 10 inside | intel Movidius | intel IRIS Pro GRAPHICS

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。
OpenVX および OpenVX ロゴは Khronos の商標です。
OpenCL および OpenCL ロゴは、Apple Inc. の商標であり、Khronos の使用許諾を受けて使用しています。
Microsoft、Windows、Windows ロゴは、アメリカ合衆国および / またはその他の国における Microsoft Corporation の商標または登録商標です。

インテル® ディープラーニング・デプロイメント・ツールキット

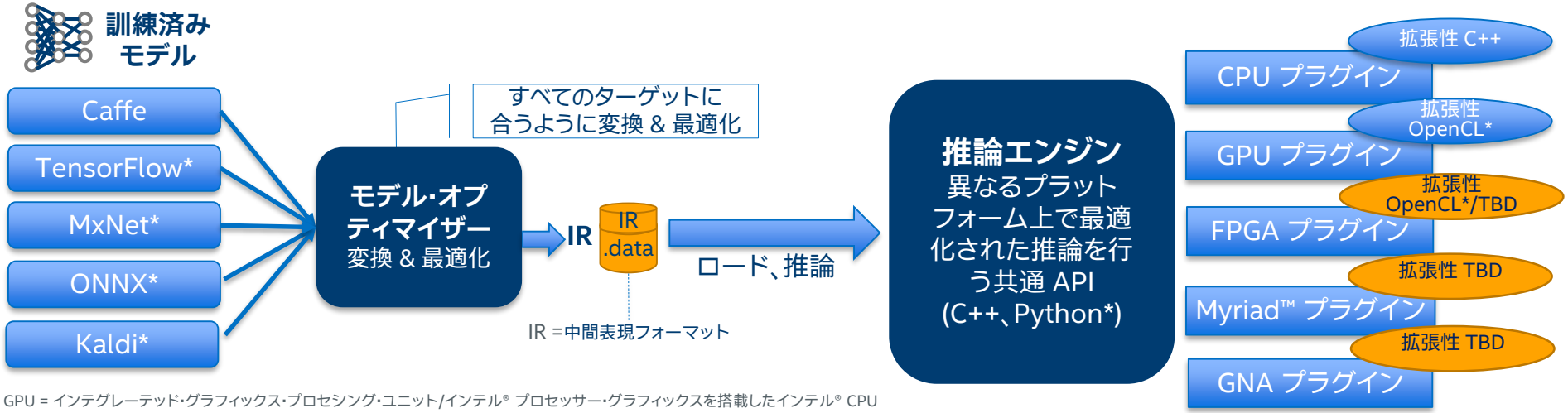
インテル® アーキテクチャーの性能を有効活用

モデル・オプティマイザー

- **What it is**: 準備段階 -> トレーニングされたモデルをインポート
- **Why important**: 従来のトポロジー変換で性能/スペースを最適化。ハードウェアがサポートするデータ型への変換では最大の性能向上が見込まれる

推論エンジン

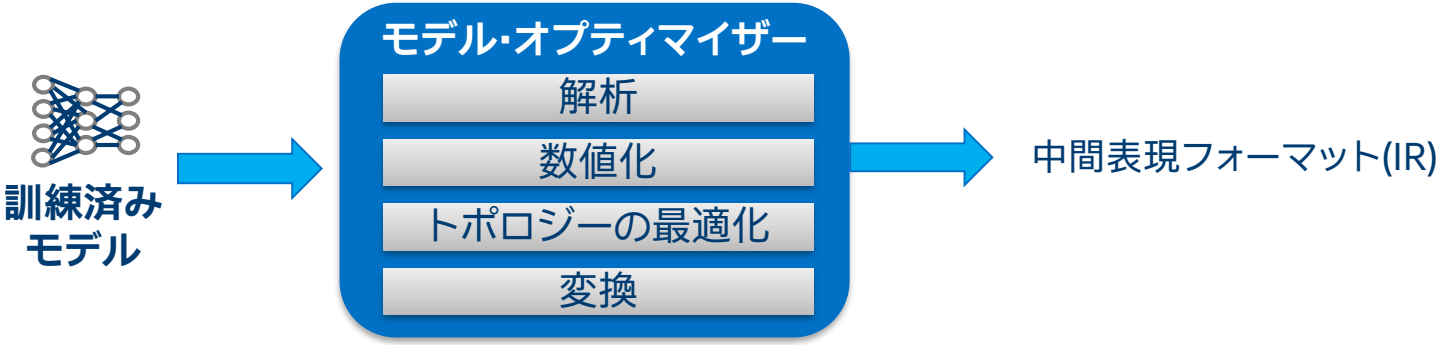
- **What it is**: 高レベルの推論 API
- **Why important**: インターフェイスは各ハードウェア・タイプごとに直接ロードされるプラグインとして実装。プログラマーが複数のコード経路を実装、維持することなく最高の性能を提供することが可能



GPU = インテグレーテッド・グラフィックス・プロセシング・ユニット/インテル® プロセッサー・グラフィックスを搭載したインテル® CPU

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。
OpenCL および OpenCL ロゴは、Apple Inc. の商標であり、Khronos の使用許諾を受けて使用しています。

モデル・オプティマイザーによる性能向上

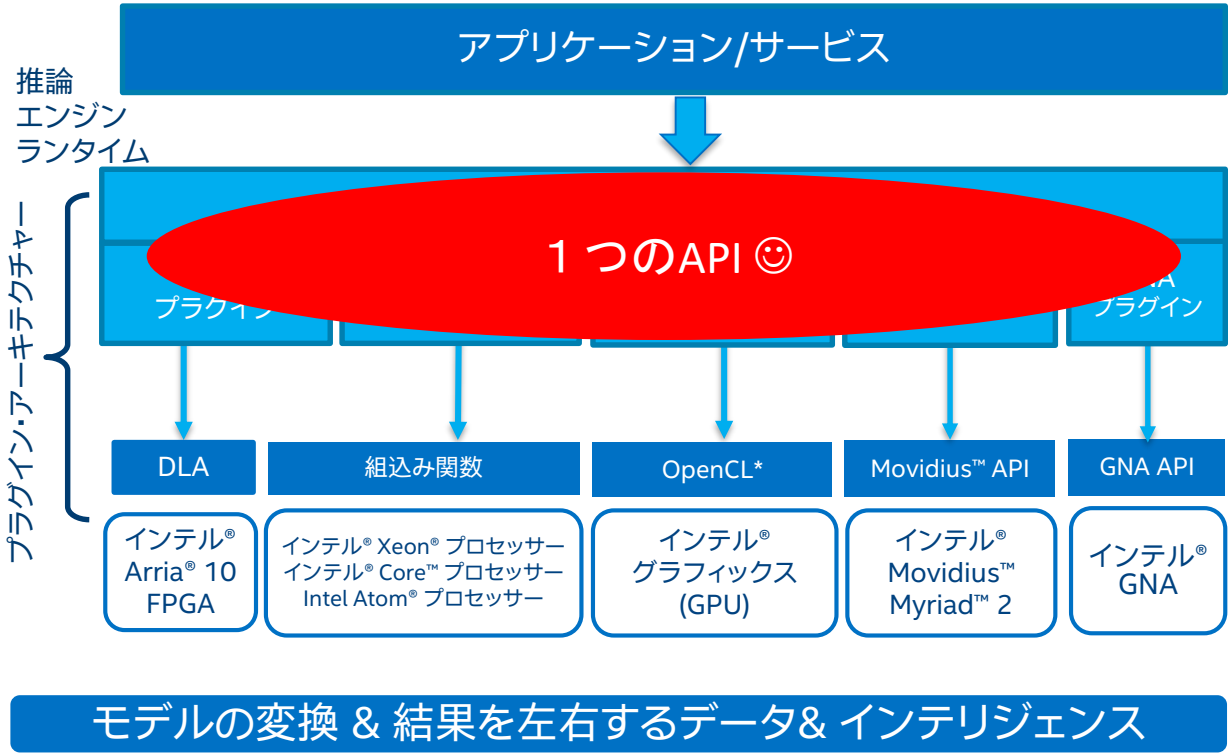


- 使い勝手の良いPython* ベースのワークフローはフレームワークの再構築を要求しない
- サポートされているさまざまなフレームワークからモデルを読み込む - Caffe、TensorFlow*、MXNet*、ONNX*、Kaldi*
- Caffe、MXNet*、TensorFlow* の 100 以上のモデルを検証済み、ONNX* model zoo のすべてのモデルをサポート
- Kaldi* のサポートにより、モデル・オプティマイザーは画像認識以外のネットワークでの推論を拡張
- 中間表現フォーマット (IR) は標準レイヤーを使用しており、ユーザーが提供するカスタムレイヤーは Caffe 不要
- サポートされていないレイヤーの場合には、元のフレームワークにフォールバックが可能だが、オリジナルのフレームワークが必要

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

推論エンジンで最適なモデル性能を実現

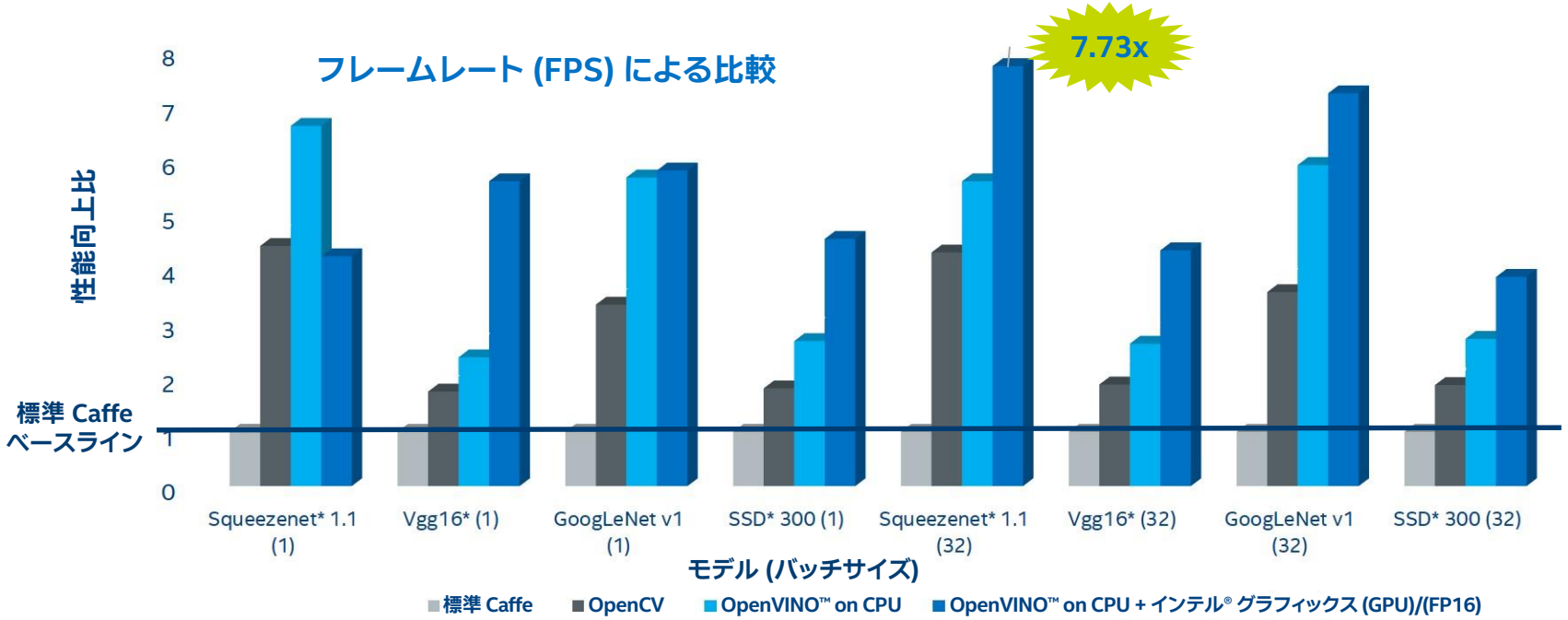
- すべてのインテル® アーキテクチャーをサポートする簡単で統一された API
- 幅広いインテル® アーキテクチャー (CPU/GEN/FPGA) をサポートする最適化された推論
- ヘテロジニアスなハードウェア・タイプにまたがったレイヤーの実行をサポート
- 非同期実行により性能を向上
- 次世代のインテル® アーキテクチャー・ベースのプロセッサー向けの大規模開発



GPU = インテグレーテッド・グラフィックス・プロセシング・ユニット/インテル® プロセッサー・グラフィックスを搭載したインテル® CPU
GNA = ガウス混合モデルおよびニューラル・ネットワーク・アクセラレーター

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。
OpenCL および OpenCL ロゴは、Apple Inc. の商標であり、Khronos の使用許諾を受けて使用しています。

各モデルにおける OpenVINO™ ツールキットとインテル® アーキテクチャーによるディープラーニング・ワークロードの性能向上



インテルのハードウェアではアクセラレーターなしでも高速

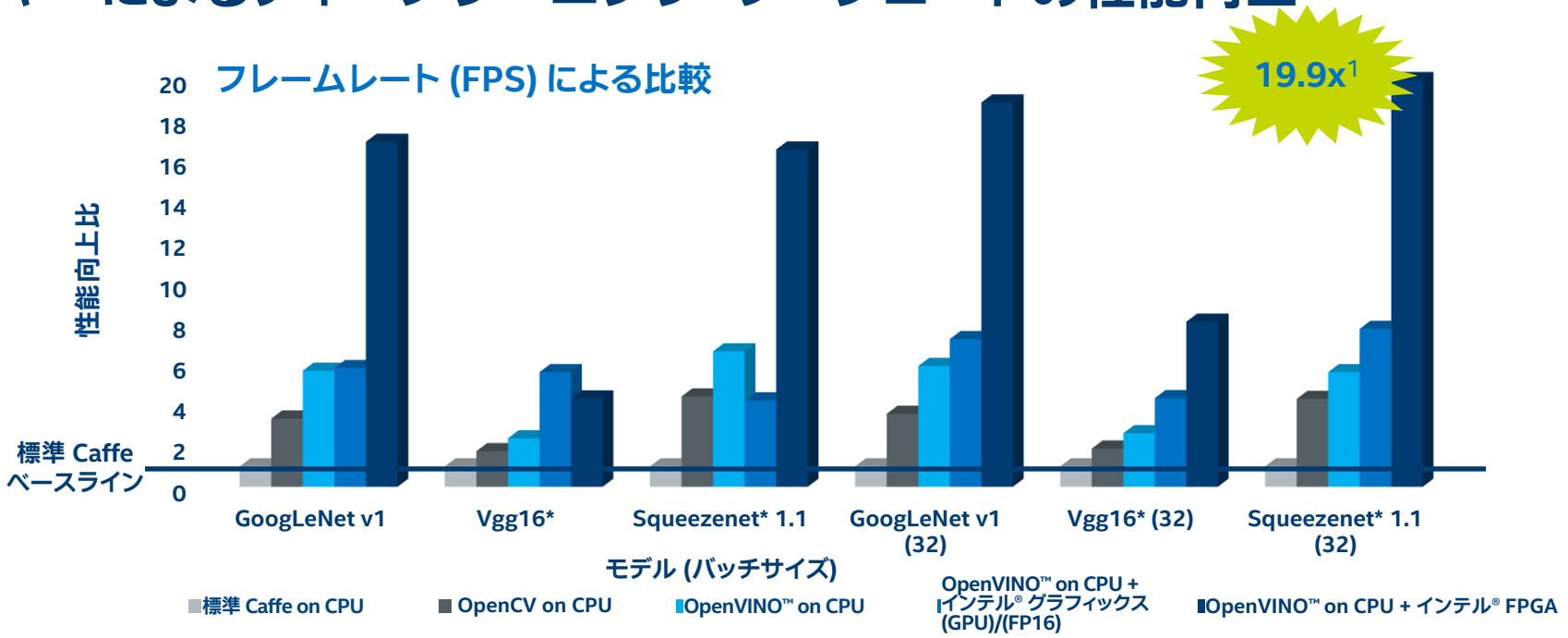
[1]ワークロードに応じて、FP16 の質/解像度は多少影響を受けます。FP32 から FP16 へのパフォーマンス/質の変更が精度に影響することがあります。異なる設定を試し、状況に最も適した設定を選択してください。上記のベンチマーク結果は、追加のテストによって変更が必要になる可能性があります。異なる設定を試し、状況に最も適した設定を選択してください。上記のベンチマーク結果は、追加のテストによって変更が必要になる可能性があります。パフォーマンス結果は 2018 年 4 月 10 日時点のテスト結果に基づいたものであり、公開されている利用可能なすべてのセキュリティー・アップデートが適用されていない可能性があります。詳細については、構成の開示を参照してください。絶対的なセキュリティーを提供できる製品はありません。性能やベンチマーク結果について、さらに詳しい情報をお知りになりたい場合は、www.intel.com/benchmarks (英語) を参照してください。 **システム構成:** 2018 年 4 月 10 日時点のインテルによるテスト。インテル® Core™ i7-6700K プロセッサー @ 2.90GHz 固定、GPU GT2 @ 1.00GHz 固定。社内テスト v312.30 – Ubuntu* 16.04、OpenVINO™ ツールキット 2018 RC4。テストは、使用モデル (公開されているもの)、バッチサイズ、その他の要因などのさまざまなパラメーターに基づきます。異なるモデルを異なるインテル® ハードウェア・ソリューションで高速化できますが、同じインテル® ソフトウェア・ツールが使用されます。

最適化に関する注意事項

© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル® コンパイラーでは、インテル® マイクロプロセッサーに限定されない最適化に関して、他社製マイクロプロセッサー用に同等の最適化を行えないことがあります。これには、インテル® ストリーミング SIMD 拡張命令 2、インテル® ストリーミング SIMD 拡張命令 3、インテル® ストリーミング SIMD 拡張命令 3 補足命令などの最適化が該当します。インテルは、他社製マイクロプロセッサーに関して、いかなる最適化の利用、機能、または効果も保証いたしません。本製品のマイクロプロセッサー依存の最適化は、インテル® マイクロプロセッサーでの使用を前提としています。インテル® マイクロアーキテクチャーに限定されない最適化のなかにも、インテル® マイクロプロセッサー用のものがあります。この注意事項で言及した命令セットの詳細については、該当する製品のユーザー・リファレンス・ガイドを参照してください。注意事項の改訂 #20110804

各モデルにおける OpenVINO™ ツールキットとインテル® アーキテクチャーによるディープラーニング・ワークロードの性能向上



インテル® FPGA でさらに大きな性能向上

[1]ワークロードに応じて、FP16 の質/解像度は多少影響を受けます。FP32 から FP16 へのパフォーマンス/質の変更が精度に影響することがあります。異なる設定を試し、状況に最も適した設定を選択してください。パフォーマンス結果は 2018 年 6 月 13 日時点のテスト結果に基づいたものであり、公開されている利用可能なすべてのセキュリティー・アップデートが適用されていない可能性があります。詳細については、構成の開示を参照してください。絶対的なセキュリティーを提供できる製品はありません。性能やベンチマーク結果について、さらに詳しい情報をお知りになりたい場合は、www.intel.com/benchmarks (英語) を参照してください。**システム構成:** 2018 年 6 月 13 日時点のインテルによるテスト。インテル® Core™ i7-6700K プロセッサー @ 2.90GHz 固定、GPU GT2 @ 1.00GHz 固定。社内テスト v3.15.21 – Ubuntu* 16.04、OpenVINO™ ツールキット 2018 RC4、インテル® Arria® 10 FPGA 1150GX。テストは、使用モデル (公開されているもの)、バッチサイズ、その他の要因などのさまざまなパラメーターに基づきます。異なるモデルを異なるインテル® ハードウェア・ソリューションで高速化できますが、同じインテル® ソフトウェア・ツールが使用されます。

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。
インテル® コンパイラーでは、インテル® マイクロプロセッサーに限定されない最適化に関して、他社製マイクロプロセッサー用に同等の最適化を行えないことがあります。これには、インテル® ストリーミング SIMD 拡張命令 2、インテル® ストリーミング SIMD 拡張命令 3、インテル® ストリーミング SIMD 拡張命令 3 補足命令などの最適化が該当します。インテルは、他社製マイクロプロセッサーに関して、いかなる最適化の利用、機能、または効果も保証いたしません。本製品のマイクロプロセッサー依存の最適化は、インテル® マイクロプロセッサーでの使用を前提としています。インテル® マイクロアーキテクチャーに限定されない最適化のなかにも、インテル® マイクロプロセッサー用のものがあります。この注意事項で言及した命令セットの詳細については、該当する製品のユーザー・リファレンス・ガイドを参照してください。注意事項の改訂 #20110804

OpenVINO™ ツールキットのデモ

最適化に関する注意事項
© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

サマリー/Q&A

さあ、始めよう

最適化に役立つ情報

iSUS (IA Software User Society)

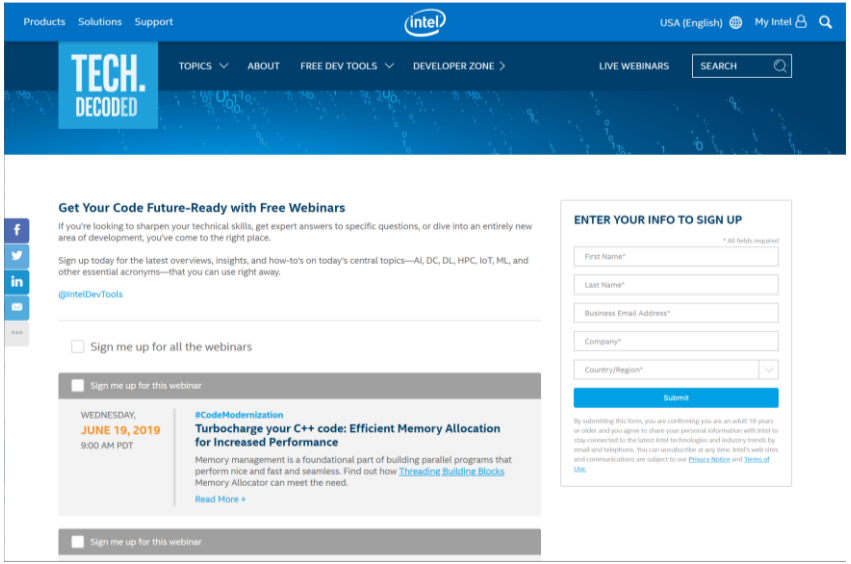
- インテル® ソフトウェア開発製品の日本語版や日本語ヘルプの提供
- インテル コーポレーションのインテル® ソフトウェア・ネットワークの記事の日本語化

インテル® Parallel Studio XE

- 概要、機能、サポート、コードサンプル (英語)
- トレーニング資料、Tech.Decoded ウェビナー (英語)、HOW TO ビデオと記事 (英語)
- お客様の声 (英語) & ケーススタディー
- その他のインテル® ソフトウェア開発製品

インテル® Code Modernization Program

- 概要 (英語)
- ライブ・トレーニング (英語)



TECH ウェビナーの視聴　https://intel.ly/2PdkNhN (英語)

© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

法務上の注意書きと最適化に関する注意事項

パフォーマンス結果は 2018 年 9 月時点のテスト結果に基づいたものであり、公開されている利用可能なすべてのセキュリティー・アップデートが適用されていない可能性があります。詳細については、構成の開示を参照してください。絶対的なセキュリティーを提供できる製品はありません。

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサー用に最適化されていることがあります。SYSmark* や MobileMark* などの性能テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行ったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。詳細については、www.intel.com/benchmarks (英語) を参照してください。

本資料の情報は、現状のまま提供され、本資料は、明示されているか否かにかかわらず、また禁反言によるとよらずにかかわらず、いかなる知的財産権のライセンスも許諾するものではありません。製品に付属の売買契約書『Intel's Terms and Conditions of Sale』に規定されている場合を除き、インテルはいかなる責任を負うものではなく、またインテル製品の販売や使用に関する明示または黙示の保証 (特定目的への適合性、商品性に関する保証、第三者の特許権、著作権、その他、知的財産権の侵害への保証を含む) をするものではありません。

© 2019 Intel Corporation. 無断での引用、転載を禁じます。Intel、インテル、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Arria、Intel Atom、Intel Core、Iris、Stratix、Xeon、Movidius、Myriad、Intel Nervana、OpenVINO は、アメリカ合衆国および/ またはその他の国における Intel Corporation またはその子会社の商標です。

最適化に関する注意事項

インテル® コンパイラーでは、インテル® マイクロプロセッサーに限定されない最適化に関して、他社製マイクロプロセッサー用に同等の最適化を行えないことがあります。これには、インテル® ストリーミング SIMD 拡張命令 2、インテル® ストリーミング SIMD 拡張命令 3、インテル® ストリーミング SIMD 拡張命令 3 補足命令などの最適化が該当します。インテルは、他社製マイクロプロセッサーに関して、いかなる最適化の利用、機能、または効果も保証いたしません。本製品のマイクロプロセッサー依存の最適化は、インテル® マイクロプロセッサーでの使用を前提としています。インテル® マイクロアーキテクチャーに限定されない最適化のなかにも、インテル® マイクロプロセッサー用のものがあります。この注意事項で言及した命令セットの詳細については、該当する製品のユーザー・リファレンス・ガイドを参照してください。
注意事項の改訂 #20110804

© 2019 Intel Corporation. 無断での引用、転載を禁じます。
* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。