

AI時代に注目のGPU百科

ご購入はこちら

松岡 洋

● GPUはディープ・ランニングや仮想通貨で大注目!

現在、深層学習や仮想通貨などでGPU市場がにぎわっており、秋葉原のパーツ・ショップにはさまざまなGPUが並んでいます。NVIDIAのGeForceシリーズやAMDのRadeonシリーズが有名です。

深層学習においては、GPUで学習を行い、その学習済みモデルをスマートフォンに組み込まれたGPUで取り込んで、判定に利用するといった用途も見られます。つまり、「機械学習に使えるGPU」と言っても、サーバからスマートフォンまで、さまざまな規模の品があります。表1に主なGPU製品群を示します。

● NVIDIAだけではない

深層学習の学習工程では、グーグルのフレームワークTensorFlowなどが、NVIDIAのCUDAを使用するため、GeForce GTX TITAN XやGTX 1080が多用されています。

AMDもHIPというCUDAをAMDのGPUカーネルに変換する方法を提供しており、RadeonやRadeon ProでもTensorFlowを動かすことができます。

さらに学習済みモデルを使ってスマートフォンなど

表1 主なGPU製品群

メーカー	製品群	用途
NVIDIA	Tesla	科学計算や機械学習など、サーバ向け
	Quadro	3Dモデリングなど、ワークステーション向け
	GeForce	ゲームや機械学習など、デスクトップ向け
	Jetson	組み込み機器向け
AMD	FirePro	科学計算や機械学習など、サーバ向け
	Radeon Pro	映像制作など、ワークステーション向け
	Radeon	ゲームや機械学習など、デスクトップ向け
	Radeon (SoC)	CPU内蔵GPU
インテル	Iris,HD Graphics (SoC)	CPU内蔵GPU
ARM	Mali (SoC)	CPU内蔵GPU、エッジ・コンピューティング向け

で判定する場合には、OpenCLで記述されたライブラリを使用することもあり、深層学習 = CUDA = NVIDIAではありません。

NVIDIAのGPU

表2にNVIDIAの主なGPUボードの仕様を示します。

通常の科学計算で用いる単精度浮動小数点(FP32)での理論性能はハイエンドのGTX TITAN

Xで11TFLOPSです。この理論性能はコア数、クロック周波数などハードウェアの仕様上到達できる上限という意味で、通常はこの50%前後の計算性能となります。

表2 NVIDIAの主なGPUボードの仕様

型名	アーキテクチャ	コア数	クロック周波数 [MHz]	メモリ容量 [バイト]	メモリ帯域幅 [Gバイト/s]	理論性能値 (FP32) [FLOPS]
GeForce GTX TITAN X	Pascal	3584	1417	12G	480	11T
GeForce GTX 1080	Pascal	2560	1607	8G	224	8T
GeForce GTX 1060	Pascal	1280	1506	6G	192	4T
Jetson TX2	Pascal	256	1300	8G (CPUと共用)	60	750G