

IA : Intel dévoile ses puces NNP Spring Hill et Crest

Mardi, lors de la conférence Hot Chips 2019 à Palo Alto, Intel a [présenté plus avant](#) deux processeurs de sa gamme Nervana Neural Network Processor : NNP-I (nom de code Spring Hill) et NNP-T (Spring Crest). Ils sont le fruit du rachat de [Nervana Systems](#) en 2016.

Développées par le centre R&D d'Intel à Haïfa (Israël), ces puces utilisent l'intelligence artificielle (IA) et le procédé de fabrication de puces en 10 nanomètres du fondeur américain. Le tout pour fournir de « meilleures performances avec une consommation réduite d'énergie », selon les promoteurs de l'offre.

Le processeur Intel [Nervana Neural Network Processor for Inference](#) (NNP-I), pour commencer, a été développé pour le calcul intensif et en temps quasi réel. Il combine coeur de CPU et moteur d'inférence basés sur l'IA.

Il n'est pas encore question de disponibilité générale. Mais, relève [Reuters](#), Facebook utilise déjà des processeurs NNP-I pour certains de ses serveurs.

NPP et Xeon

De son côté, le groupe chinois [Baidu est partenaire](#) de développement de la puce Intel [Nervana Neural Network Processor for Training](#) (NNP-T).

NNP-T s'appuie sur des modèles d'apprentissage profond (deep learning). La puce est conçue pour traiter rapidement des données massives, rares ou complexes et fournir le juste équilibre entre calcul, communication et mémoire, selon Intel.

Son architecture permet d'optimiser la réutilisation de données à la volée.

Dans des environnements datacenters et cloud, ces puces peuvent compléter des [processeurs Xeon Scalable](#) exploités pour des tâches spécifiques d'intelligence artificielle.

« Dans un monde doté d'IA, nous devons adapter les solutions matérielles en une combinaison de processeurs dédiés à des cas d'usage spécifiques », a déclaré Naveen Rao, VP d'Intel et directeur général du groupe produits d'intelligence artificielle. « Il nous faut donc examiner les besoins distinctifs des applications et réduire le temps de latence pour fournir les meilleurs résultats au plus près des données. »

Les puces NPP-T et NNP-I ont été conçues dans ce but.