

Microprocesseurs d'IA de vision de milieu de gamme pour applications périphériques

Par Rolf Horn, ingénieur d'applications chez Digi-Key Electronics

Les applications Edge AI utilisent des algorithmes de vision par ordinateur pour détecter des personnes, des objets ou des anomalies en temps réel. Le traitement des images et des vidéos en périphérie requiert typiquement un microprocesseur (MPU) d'IA dédié à la vision, capable d'interfacer avec des caméras, d'exécuter des modèles d'IA et incluant souvent un accélérateur IA dédié.

L'intégration de capacités d'IA de vision dans un seul dispositif réduit le coût et l'encombrement liés à l'utilisation de composants séparés, ce qui fait des microprocesseurs d'IA de vision de pointe une solution idéale pour les applications embarquées compactes.

Le [microprocesseur RZ/V2N](#) (Figure 1) de [Renesas Electronics Corporation](#) est un microprocesseur d'IA de vision qui présente une basse consommation d'énergie, des performances d'inférence IA élevées, quatre cœurs de processeur Arm® Cortex®-A55 (1,8 GHz), un processeur Arm Cortex-M33 (200 MHz) et deux entrées de caméra via une connexion MIPI.



Figure 1 : Le microprocesseur RZ/V2N de Renesas offre aux concepteurs de nouvelles options pour intégrer l'IA de vision dans les applications périphériques. (Source de l'image : Renesas Electronics Corporation)

Ce microprocesseur de Renesas constitue une solution économique pour les applications périphériques qui requièrent des capacités d'IA modérées à élevées à un prix abordable. Il fait partie de la [série RZ/V](#) de l'entreprise, dont l'objectif est d'offrir une évolutivité étendue, des bureaux intelligents aux drones (Figure 2).



Figure 2 : Positionné comme une offre milieu de gamme de la série RZ/V, le RZ/V2N permet des applications telles que les robots mobiles à usage domestique et les systèmes de surveillance des conducteurs. (Source de l'image : Renesas Electronics Corporation)

Exigences d'un microprocesseur d'IA de vision

Les applications Edge AI fonctionnent souvent dans des dispositifs embarqués, qui peuvent être alimentés par batterie ou fonctionner avec de faibles réserves d'énergie. Par conséquent, les microprocesseurs de vision doivent offrir des capacités d'inférence élevées tout en consommant moins d'énergie que les dispositifs de calcul à haute performance traditionnels.

Le microprocesseur d'IA de vision idéal équilibre les performances, le rendement énergétique, l'intégration, la facilité de développement et la sécurité. Voici un aperçu de certaines des caractéristiques clés à prendre en compte lors de la sélection d'un microprocesseur :

- **Performances d'inférence** : le RZ/V2N fournit jusqu'à 15 TOPS grâce à son accélérateur DRP-AI3 intégré, ce qui le rend adapté aux applications de milieu de gamme telles que les caméras intelligentes, l'inspection industrielle et la robotique en périphérie. Alors que certains systèmes hautes performances, tels que les robots collaboratifs et les drones autonomes, peuvent nécessiter de 80 à 100 TOPS, de nombreuses applications Edge AI fonctionnent correctement avec de 1 à 15 TOPS, selon leur complexité. Le TOPS par watt (TOPS/W) définit le rendement du produit, qui est une mesure du nombre d'opérations pouvant être effectuées par seconde par watt.

Bien que le nombre de TOPS fournisse une indication de base des performances, la vitesse d'inférence réelle peut être considérablement améliorée avec l'ajout d'un accélérateur d'IA dédié, qui délègue les charges de travail d'IA de vision reposant sur des calculs intensifs de matrices et de tenseurs. Cela permet aux systèmes de fonctionner plus rapidement et plus efficacement, avec moins de cycles d'horloge et une consommation d'énergie plus faible.

- **Fonctionnement basse consommation** : de nombreux dispositifs périphériques fonctionnent sur batterie ou dans des limites thermiques strictes. Les microprocesseurs

de vision conçus pour l'Edge AI incluent souvent une mise à l'échelle de fréquence et tension dynamique (DVFS), qui ajuste la consommation d'énergie en fonction des exigences des charges de travail. Associée à des techniques telles que l'élagage des réseaux neuronaux — qui compressent la taille du modèle et réduisent les calculs inutiles — la technologie DVFS permet d'obtenir un rapport TOPS/W plus élevé, améliorant ainsi à la fois le temps de fonctionnement et la durée de vie de la batterie. L'accélérateur DRP-AI3 permet d'éviter le recours à des processeurs graphiques (GPU) énergivores, contribuant ainsi à des rapports TOPS/W plus élevés en périphérie.

- *Traitement d'image sur puce* : les microprocesseurs de vision avec processeurs de signal d'image (ISP) intégrés en option peuvent effectuer des tâches de nettoyage d'image de routine, telles que la correction du niveau de noir, la correction des couleurs, le recadrage et la correction des ombres. Dans les applications de sécurité ou de surveillance, l'ISP peut également pré-filtrer les trames. Par exemple, dans un flux vidéo continu, le système peut rejeter les images statiques et n'envoyer au processeur d'IA que les images de mouvement ou d'activité (par exemple, la détection d'intrus), ce qui réduit les inférences inutiles et permet d'économiser de l'énergie.
- *Mémoire sur puce* : la mémoire est également un facteur important pour les performances et le rendement. Le stockage local des données permet d'éviter la latence et les coûts énergétiques liés à l'accès à la mémoire externe, qui peuvent être importants lors de l'inférence IA en temps réel. Avec 1,5 Mo de SRAM sur puce et la prise en charge de la mémoire LPDDR4X, le RZ/V2N offre un équilibre entre vitesse de traitement interne et options de mémoire extensible.
- *Accélération du déploiement de l'IA* : les kits outils d'IA et les cartes d'évaluation qui incluent des applications et des interfaces préprogrammées peuvent aider les développeurs à prototyper et déployer rapidement des applications d'IA de vision. De plus, le microprocesseur doit être capable de prendre en charge les formats de modèles d'IA standard. Le RZ/V2N est compatible avec les formats de modèles standard, comme ONNX et TensorFlow Lite.
- *Sécurité* : dans les environnements périphériques, chaque capteur ou point d'extrémité peut représenter un vecteur d'attaque potentiel. Il est donc important que les microprocesseurs de vision soient capables de prendre en charge des fonctionnalités de sécurité intégrées telles que le démarrage sécurisé et le chiffrement des chemins de données. Le RZ/V2N inclut des fonctionnalités de démarrage sécurisé et de chiffrement au niveau matériel, et il utilise Arm TrustZone pour isoler les opérations sécurisées, contribuant ainsi à protéger à la fois l'intégrité des modèles et les données d'entrée sensibles.

Fonctionnalités compatibles IA du microprocesseur RZ/V2N

L'accélérateur d'IA propriétaire de Renesas, le DRP-AI3 (Dynamically Reconfigurable Processor), est répertorié à 10 TOPS/W mais peut être amélioré jusqu'à 15 TOPS/W avec un élagage avancé, qui compresse la taille des modèles que le système doit traiter. Cela peut éliminer le recours à un processeur graphique ou à un FPGA distinct.

Le RZ/V2N ne mesure que 15 mm², ce qui en fait une bonne option pour les dispositifs compacts. La combinaison d'un processeur quadricœur, d'un accélérateur d'IA dédié et de la prise en charge de deux entrées de caméra dans un seul dispositif ouvre de nouvelles opportunités aux concepteurs pour intégrer l'IA de vision dans des applications telles que les caméras intelligentes, les dispositifs de sécurité, les robots et même les appareils électroménagers.

Le microprocesseur fonctionne avec une faible consommation d'énergie, ce qui réduit la quantité de chaleur générée, éliminant ainsi le besoin de systèmes de refroidissement et de ventilateurs supplémentaires, et réduisant par conséquent la taille et le coût des systèmes embarqués. Grâce à sa capacité à accueillir deux caméras, il permet aux applications de capturer des images sous deux angles différents et d'améliorer la reconnaissance spatiale. Un système peut effectuer de multiples opérations, comme compter simultanément les voitures dans un parking et reconnaître les plaques d'immatriculation.

Architecture du microprocesseur RZ/V2N

Le microprocesseur RZ/V2N offre une gamme complète de caractéristiques et de fonctions pour le développement d'applications destinées au marché de l'IA de milieu de gamme exigeant une IA hautes performances à un prix abordable (Figure 3).

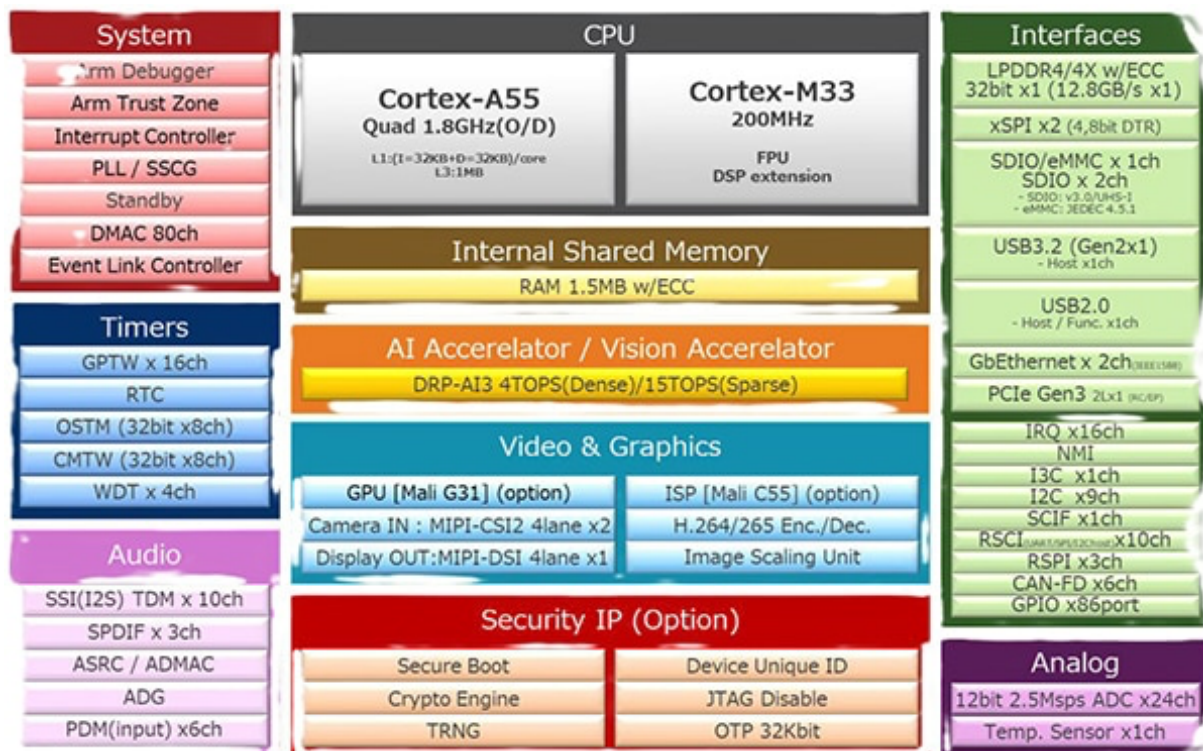


Figure 3 : Schéma de l'architecture RZ/V2N. (Source : Renesas Electronics Corp.)

Les fonctionnalités clés incluent les suivantes :

- **Processeur (CPU) :** l'architecture hybride comprend un Cortex-A55 à quatre cœurs 1,8 GHz, un processeur hautes performances et un Cortex-M33 200 MHz, un cœur basse consommation conçu pour le contrôle en temps réel et les tâches liées à la sécurité.
- **Mémoire interne partagée :** 1,5 Mo de RAM pour la mémoire sur puce avec code de correction d'erreur (ECC), contribuant à l'intégrité des données. Les algorithmes ECC détectent et corrigent les erreurs dans les données, à la fois pendant le stockage et la transmission. La mémoire sur puce de 1,5 Mo permet aux algorithmes d'IA de s'exécuter rapidement, et le RZ/V2N dispose également d'une interface pour la mémoire DDR externe qui peut être ajoutée si davantage de mémoire est nécessaire.
- **Accélérateur d'IA :** le moteur d'IA dédié DRP-AI3 de Renesas permet le traitement rapide des inférences IA, offrant la faible consommation d'énergie et la flexibilité requises par les points d'extrémité.

- *Vidéo et graphisme* : processeur graphique et ISP en option améliorant le traitement des images et le rendu des graphiques.
- *Temporisateurs* : les temporisateurs prennent en charge les opérations en temps réel, essentielles pour la commande moteur et d'autres applications d'automatisation.
- *Bloc audio* : idéal pour les applications audio à plusieurs canaux, telles que les haut-parleurs intelligents et les systèmes d'infodivertissement.
- *Interfaces* : les interfaces mémoire haute vitesse et les périphériques à large bande passante font partie des nombreuses interfaces pouvant être connectées au module de microprocesseur de vision.
- *Bloc analogique* : un convertisseur analogique-numérique (CAN) 12 bits élimine le recours à des CAN séparés dans les systèmes de contrôle ou les applications de surveillance.

Renesas propose également le kit de cartes d'évaluation [RTK0EF0186C03000BJ](#) pour le RZ/V2N, permettant aux concepteurs de prototyper et d'évaluer des applications d'IA de vision (Figure 4). Les concepteurs peuvent également accéder à des applications d'IA couvrant plus de 50 cas d'utilisation dans les applications d'IA et le SDK d'IA de l'entreprise sur GitHub.



Figure 4 : Le kit de cartes d'évaluation pour le RZ/V2N inclut une carte de processeur, une carte d'extension et deux cartes secondaires, ainsi qu'un kit de développement logiciel pour l'IA. (Source de l'image : Renesas Electronics Corp.)

Conclusion

Le RZ/V2N de Renesas est idéal pour les applications Edge AI de milieu de gamme devant fournir des informations basées sur les données avec une latence réduite à des vitesses élevées. Grâce à son format compact et à sa capacité à répondre aux exigences d'inférence tout en fonctionnant avec une faible consommation d'énergie, il convient parfaitement à une large gamme de dispositifs embarqués.



Rolf Horn, ingénieur applications chez DigiKey, fait partie du groupe de support technique européen depuis 2014, avec pour principale responsabilité de répondre à toutes les questions liées au développement et à l'ingénierie des clients finaux en EMEA, ainsi que de rédiger et de corriger des articles et des blogs en allemand sur les plateformes TechForum et de DK. Avant de rejoindre DigiKey, il a travaillé chez plusieurs fabricants dans le domaine des semi-conducteurs, en se concentrant sur les systèmes FPGA embarqués, microcontrôleurs et processeurs pour les applications industrielles et automobiles. Rolf est titulaire d'un diplôme en génie électrique et électronique de l'Université des sciences appliquées de Munich, en Bavière, et a débuté sa carrière professionnelle chez un distributeur local de produits électroniques en tant qu'architecte de solutions système afin de partager ses connaissances et son expertise en constante progression en tant que conseiller de confiance.

Loisirs : passer du temps avec la famille et les amis, voyager dans notre transporteur VW-California et faire de la moto sur une BMW GS 100 de 1988.