

smart. intégrations mag

www.smartintegrationsmag.com

N°31 - NOVEMBRE/DÉCEMBRE 2018 - 15 €



ÉDITION
SPECIALE

VIDEOPROTECTION & DEEP LEARNING
Les promesses de l'Intelligence Artificielle

18 08:36AM 22s
CAM 237 - CAM 6

SECURITY

UNUSUAL BEHAVIOR

1 ALERT
REAL-TIME VISUALISATION

ALERT



FALL DETECTION

DEEP LEARNING

L'entreprise XXII, spécialisée dans la vision par ordinateur, travaille sur le deep learning pour améliorer, entre autres, la détection des dynamiques humaines : chute, comportement agressif...
Photo : DR

L'apprentissage se poursuit

Appliqués à l'analyse vidéo, les réseaux de neurones sont responsables de gains de performance significatifs. Mais ils demandent encore à être éprouvés.

Dans sa définition la plus triviale, l'intelligence artificielle, ou IA, n'est autre que la capacité d'une machine, soit un ordinateur, à prendre une décision. « Et cela existe depuis 50 ans, s'amuse Jean-Baptiste Ducatez, dirigeant de Foxstream. Maîtrisant l'analyse vidéo et la détection d'intrusion, nous exploitons l'IA depuis plus de 10 ans. » Le sujet n'a rien de nouveau, donc. Mais il anime l'actualité technologique car une technique prometteuse, formalisée dans les années 80, prend aujourd'hui son essor : l'apprentissage approfondi (ou deep learning). C'est une méthode dérivée de l'apprentissage automatique (ou machine learning) et qui repose essentiellement sur des réseaux de neurones, c'est-à-dire des algorithmes qui s'inspirent de l'organisation et du fonctionnement de nos propres cellules cérébrales. Le deep learning est à l'origine des progrès spectaculaires de l'IA

depuis quelques années, que ce soit pour la compréhension du langage naturel, l'une des vertus des assistants de Google, Amazon et consorts, ou la reconnaissance de formes et d'objets dans un flux vidéo. La lecture des plaques d'immatriculation, ayant

adopté l'IA de longue date, est l'une des premières applications à en tirer profit.

William Eldin, XXII. « Les algorithmes d'apprentissage approfondi doivent maigrir pour être déployés à grande échelle. »



Responsable du groupe Computer Vision chez Genetec, qui a acquis le spécialiste Autovu en 2007, Paule Brodeur a pu mesurer cette évolution : « Depuis douze ans, nous utilisons des algorithmes de machine learning pour identifier les plaques et traiter l'image. Mais les performances ont fini par plafonner. Il fallait accomplir un travail gigantesque, ne serait-ce que pour obtenir un gain de 1%. Nous avons ensuite développé quelques blocs d'algorithme avec du deep learning, faute de pouvoir le faire en totalité à cause de la puissance limitée des caméras. D'un coup, les performances ont bondi de 5%. » Une expérience de lecture de plaques, réalisée par les équipes d'AutoVu sur les routes de Montreal en avril dernier, montre que les réseaux de neurones ont un taux de réussite de 96%, contre 87,5% pour des algorithmes classiques. Le deep learning, bien qu'il ne s'applique pas à toutes les fonctions algorithmiques embarquées, est d'ores-et-déjà à l'œuvre dans les dernières versions commerciales des caméras d'Autovu, lesquelles sont conçues en interne.

L'IA et le deep learning sont des aides précieuses pour l'analyse intelligente d'un événement.



Photo : DR



Les caméras Autovu, dédiées à la lecture de plaques minéralogiques, embarqueront l'an prochain des GPUs pour généraliser les fonctions d'apprentissage profond.

Dans d'autres applications, ces réseaux de neurones n'en sont encore qu'à leurs débuts. Foxstream, par exemple, ne les commercialise toujours pas mais poursuit ses expérimentations depuis deux ans en lien avec le CNRS.

■ Partir sur de bonnes bases

En matière de détection d'intrusion, Jean-Baptiste Ducatez entrevoit les bénéfices : « Les algorithmes actuels différencient les silhouettes d'apparence humaine et les objets qui ne le sont pas du tout. Avec le deep learning, le nombre de catégories où classer les objets explose. Non seulement la classification s'affine mais la portée de la détection passe d'une quinzaine de mètres à plus de 200 mètres. Les performances des systèmes progressent de manière significative. » Trois problématiques sont à

résoudre avant que le deep learning ne se démocratise. Il s'agit tout d'abord de se procurer les bases de données, en l'occurrence des images et des vidéos pour le marché de la vidéoprotection, qui serviront à « nourrir » ces algorithmes. Ceux-ci doivent en effet ingurgiter une quantité colossale de données, représentant le maximum de cas de figure, pour apprendre à faire ce pour quoi ils ont été élaborés. Par exemple percevoir un être humain, que celui-ci soit debout, allongé, filmé sous tel ou tel angle, portant un chapeau ou non, etc. « Or, s'il est simple de récupérer un million d'images de chats sur internet, c'est beaucoup plus compliqué pour des scènes de délits ou de crimes, précise Jean-Baptiste Ducatez. L'enjeu a changé. La question n'est plus de savoir si on possède le bon algorithme, mais les bonnes images. »

Concernant Autovu, Genetec est confronté à la même difficulté. « Nous constituons nous-mêmes nos bases, en réalisant des trajets en voiture et en capturant des images, et on fait aussi appel à nos intégrateurs, confie Paule Brodeur. Pour les installations fixes, il vaut mieux des images ressemblantes - même angle de caméra, même luminosité, etc - pour que les performances soient optimales. Ce travail prend du temps. » D'autant plus que ces images doivent être annotées, exercice qui consiste à délimiter et à nommer un objet dans une image, pour que l'algorithme sache l'identifier.

■ Des GPUs forts en calcul

Des bases de données publiques à visée universitaire existent, comme Image-net, mais ne fournissent pas toujours les images appropriées. L'alternative est de faire appel à des partenaires spécialisés, tel Ingedata, chargés de préparer les images qui alimenteront des projets d'IA et de vision par ordinateur. Nul doute que ce travail fastidieux vaudra de l'or, la data étant annoncée comme le carburant de l'économie de demain. Deuxième frein : la puissance de calcul nécessaire pour que ces réseaux de neurones « moulinent » sans ralentissement. Ces trois dernières années, l'usage croissant des GPUs (graphics processing unit) a contribué en partie à répondre à ce besoin : les architectures parallèles de ces circuits graphiques, réservés autrefois à l'image 3D, excellent dans les calculs exigeants sur de gros volumes de données. La puissance requise n'en demeure pas moins énorme pendant la phase d'entraînement de l'algorithme. « Celui-ci doit ensuite maigrir pour être déployé à grande échelle sur des systèmes moins puissants, explique William Eldin, pdg de la start-up XXII, spécialisée dans la vision par ordinateur pour la sécurité et le commerce (identification automatique de produits). Dans cette phase dite d'inférence, qui succède à la phase d'entraînement, l'algorithme qui a obtenu un bon taux de réussite subit donc un régime pour être hébergé sur un serveur local, puis est mis en situation réelle. »

■ Le cloud est hors de prix

Le traitement des images ne pourrait-il pas être exécuté dans le cloud,

qui offre des ressources quasi-illimitées, disponibles à la demande ? En réalité, ce n'est pas une solution toujours souhaitable, les vidéos étant parfois sensibles, ni viable. William Eldin prend l'exemple d'une grande infrastructure de vidéosurveillance dans le transport : « Il faudrait analyser chaque seconde 30 images de 8 millions de pixels, à multiplier par 5000 caméras. Le cloud coûterait des dizaines de millions d'euros chaque mois. Des serveurs locaux, équipés de cartes nVidia (qui domine le marché du GPU, ndlr), sont plus économiques, coûtant entre 50000 et 200000 euros. » C'est du reste l'option choisie par XXII pour implanter ses algorithmes chez sa dizaine de clients. Dans les transports, la détection de bagages abandonnés est le best-seller de la société. L'intégration de ces réseaux de neurones dans les caméras représente bien sûr le Graal. Les CPUs (Central processing unit) embarqués, choisis par les fabricants pour baisser les prix, sont préférables aux GPUs pour une industrialisation massive. Mais ils n'ont pas encore la carrure pour assurer les lourds traitements imposés par le deep learning. Genetec a ainsi décidé de remplacer les CPUs par des GPUs dans ses prochaines caméras Autovu en 2019 et attend « une réduction de moitié des erreurs » précise Paule Brodeur. C'est un choix plus coûteux, mais ce sont des caméras spécifiques à haute valeur ajoutée. Quoi qu'il en soit, « si le suivi de visage est envisageable, les algorithmes très lourds comme ceux employés par Amazon dans ses magasins Amazon Go sont impossibles à porter aujourd'hui dans une caméra, constate William Eldin. Mais dans deux ans, qui sait ? »

Le troisième et dernier aspect qui suscite des interrogations concerne les réseaux de neurones eux-mêmes. « On ne sait pas trop comment ça fonctionne à l'intérieur, admet Jean-Baptiste Ducatez. C'est un peu une boîte noire, dont les résultats sont parfois aberrants. » Un être humain peut ainsi être confondu... avec un cheval. Ces interprétations rocambolesques sont heureusement très rares, mais suffisantes pour inciter à la prudence. « Des erreurs aussi flagrantes ne sont pas tolérables dans le domaine de la sécurité, poursuit Jean-Baptiste Ducatez. Ce n'est que le commencement de cette technologie : le travail de recherche et les phases de vali-

Un NAS pour démocratiser le deep learning

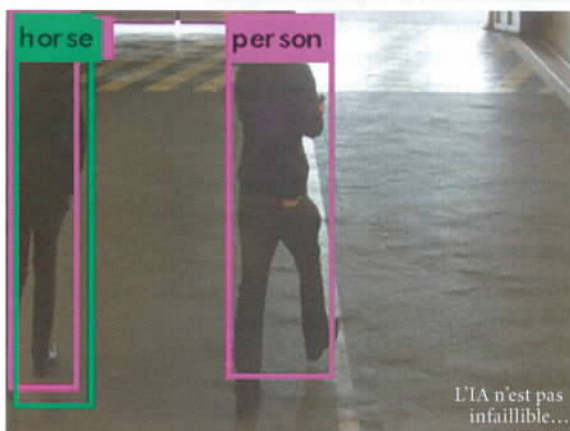
Avec le DS1419dva, un NAS équipé d'un circuit graphique, Synology a l'intention de mettre la vidéosurveillance « augmentée » par deep learning à portée du plus grand nombre



Ce NAS bénéficie d'un circuit GPU de nVidia qui accélère les fonctions de détection basées sur de l'apprentissage profondi.

Deep video analytics ou dva : ces trois initiales sont la signature du DS1419dva de Synology, un NAS à quatre baies dont la sortie est prévue en début d'année 2019. Il ne s'agit pas d'un modèle réservé exclusivement aux applications de vidéosurveillance, comme les NVR de la marque, mais les bénéfices de son moteur d'apprentissage profondi sont illustrés avec des images de caméras installées le long d'une voie publique ou d'un parking. Dans le premier cas, l'algorithme est capable de faire la distinction entre un piéton et un cycliste

ou un motocycliste. Dans le second, il garde la trace d'un individu sur un parking, même quand celui-ci passe derrière le feuillage d'un arbre. Afin d'améliorer la détection, le logiciel Surveillance Station s'appuie sur un GPU Geforce GTX. On ignore encore la puissance embarquée, le nombre de canaux pouvant être pris en charge et toutes les applications envisageables, en rapport avec les types d'algorithmes mis à disposition. On suppose cependant que le prix devrait rester abordable et adapté à des projets de petite ampleur.



L'IA n'est pas infallible...

ation doivent continuer pour fiabiliser les algorithmes. » Lesquels permettront de répondre plus rapidement à des besoins précis. « Il y a quelques années, un maire m'a demandé s'il était possible de détecter

des gros chiens traversant les rues piétonnes, se souvient Jean-Baptiste Ducatez. Mon département R&D a acquiescé, mais le développement d'un algorithme aussi spécifique aurait mobilisé un ingénieur pendant un an. Le maire n'était pas prêt à payer pour cela. Aujourd'hui, sous réserve d'utiliser les bonnes bases d'images, on pourrait répondre à cette demande. Viendront même des algorithmes qui pourront détecter des gros chiens et des bagages abandonnés. » Mais ce n'est pas de sitôt que ces réseaux de neurones seront aussi systémiques que le cerveau humain. « L'homme, prédictif dès qu'il accumule de l'expérience et capable d'analyser beaucoup de choses, n'est pas automatisable » conclut William Eldin.

Frédéric Monflier