LIVRE BLANC

IA dans l'analyse vidéo

Étude sur les analyses basées sur le machine learning (apprentissage automatique) et le deep learning (apprentissage profond)

Mars 2021

AXIS COMMUNICATIONS

Table des matières

1 Avant-propos

Les analyses vidéos basées sur l'IA représentent le sujet le plus discuté dans le secteur de la vidéosurveillance. Certaines applications peuvent considérablement accélérer l'analyse de données et automatiser les tâches répétitives. Mais actuellement, les solutions d'IA ne peuvent pas remplacer l'expérience d'un opérateur humain et les compétences de prise de décisions. La force réside plutôt dans une combinaison : tirer profit des solutions d'IA pour améliorer et développer l'efficacité humaine.

Le concept de l'IA intègre les algorithmes de machine learning et de deep learning. Ces deux types d'algorithmes créent automatiquement un modèle mathématique à l'aide d'un nombre considérable de données d'échantillon, (*les données d'apprentissage)*, pour acquérir la capacité de calculer les résultats sans y être particulièrement programmés. Un algorithme d'IA se développe grâce à une procédure répétitive, dans laquelle un cycle de collecte de données d'apprentissage, d'étiquetage des données d'apprentissage, d'utilisation des données étiquetées pour former l'algorithme et de test de l'algorithme appris se répète jusqu'à ce que le niveau de qualité désiré soit atteint. Après cela, l'algorithme est prêt à être utilisé dans les applications d'analyse qui peuvent être achetées et déployées sur un site de surveillance. À ce moment-là, l'ensemble de l'apprentissage est terminé et l'application n'apprendra plus rien de nouveau.

Une tâche habituelle destinée à l'analyse vidéo basée sur l'IA est la détection visuelle de personnes et de véhicules dans un flux de données vidéo et la distinction de chaque élément. Un algorithme de *machine learning* a appris la combinaison de caractéristiques visuelles qui définit ces éléments. Un algorithme de *deep learning* est plus sophistiqué et peut, s'il y a été formé, détecter des objets bien plus complexes. Mais de bien plus importants efforts de développement et d'apprentissage et des ressources de calcul beaucoup plus nombreuses sont également requis lorsque l'application finalisée est utilisée. Pour des besoins de surveillance bien définis, il faut donc examiner si une application de machine learning optimisée et dédiée peut être suffisante.

Le développement des algorithmes et l'augmentation de la capacité de traitement des caméras ont rendu possible l'exécution d'analyses vidéos basées sur l'IA directement sur la caméra (en périphérie de réseau) plutôt que l'exécution de calculs sur un serveur (basés sur un serveur). Cela permet une meilleure fonctionnalité en temps réel car les applications ont un accès immédiat au matériel vidéo non comprimé. Grâce aux accélérateurs matériels dédiés, tels que les MLPU (unité de traitement de machine learning) et les DLPU (unité de traitement de deep learning), dans les caméras, les analyses en périphérie de réseau peuvent être exécutées avec une plus faible consommation d'énergie qu'avec un CPU ou un GPU (unité de traitement graphique).

Avant l'installation d'une application d'analyse vidéo basée sur l'IA, les recommandations du fabricant basées sur les préconditions et limitations connues doivent être soigneusement étudiées et respectées. Chaque installation de surveillance est unique et les performances de l'application doivent être évaluées sur chaque site. Si la qualité s'avère inférieure à celle attendue, des recherches doivent être réalisées à un niveau holistique et ne pas se concentrer uniquement sur l'application d'analyse elle-même. Les performances des analyses vidéos dépendent de nombreux facteurs liés à l'aspect matériel de la caméra, sa configuration, la qualité vidéo, la dynamique de la scène et l'éclairage. Dans de nombreux cas, connaître l'impact de ces facteurs et les optimiser en conséquence permet d'augmenter les performances des analyses vidéos de l'installation.

Puisque l'IA est de plus en plus appliquée au domaine de la surveillance, les avantages de l'efficacité opérationnelle et de nouveaux cas d'utiliser doivent s'équilibrer avec une discussion attentive sur les lieux et les moments où appliquer cette technologie.

2 Introduction

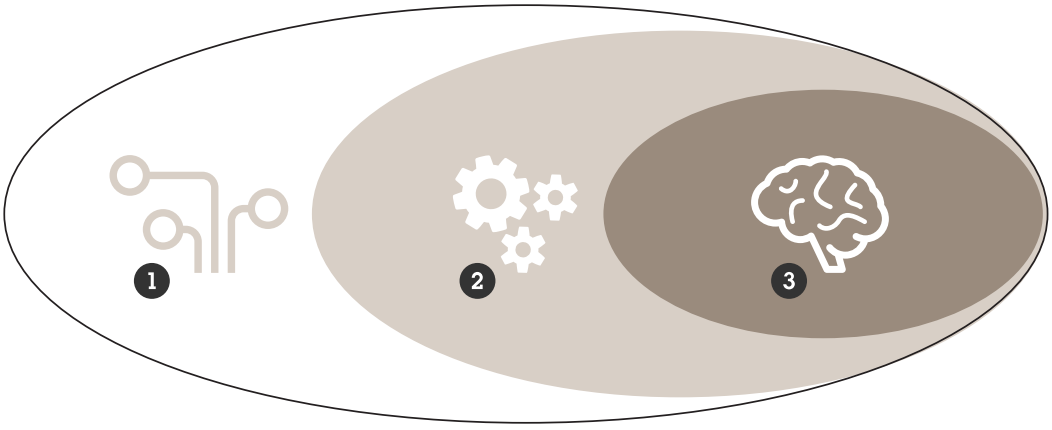
L'IA, l'intelligence artificielle, a été développée et a fait l'objet de débats depuis que les premiers ordinateurs sont été inventés. Tandis que les incarnations les plus révolutionnaires n'existent pas encore, les technologies basées sur l'IA sont largement utilisées de nos jours pour exécuter des tâches clairement définies, telles que la reconnaissance vocale, les moteurs de recherche et les assistants virtuels. L'IA est également de plus en plus employée dans le domaine de la santé dans lequel elle fournit des ressources précieuses, par exemple, les radiodiagnostics et les analyses de lecture d'empreintes rétiniennes.

Les analyses vidéos basées sur l'IA représentent le sujet le plus discuté dans le secteur de la vidéosurveillance et les attentes sont élevées. Il existe des applications sur le marché qui utilisent les algorithmes d'IA pour accélérer efficacement l'analyse de données et automatiser les tâches répétitives. Mais dans un contexte plus large de surveillance, l'IA aujourd'hui et dans un futur proche devra être vue comme un simple élément, parmi tant d'autres, dans la procédure de création de solutions précises.

Ce livre blanc fournit un contexte technologique sur les algorithmes de machine learning et de deep learning et sur la façon dont ils peuvent être développés et appliqués à l'analyse vidéo. Il contient un bref compte-rendu de l'accélération matérielle de l'IA et des arguments pour et contre relatifs à l'exécution d'analyses basées sur l'IA en périphérie de réseau, plutôt que sur un serveur. Ce livre examine également la façon dont les préconditions peuvent être optimisées pour que les analyses vidéos basées sur l'IA soient performantes, en tenant compte d'un important éventail de facteurs.

3 IA, machine learning et deep learning

L'intelligence artificielle (IA) est un vaste concept associé aux machines qui peuvent résoudre des tâches complexes tout en démontrant des traits évidents d'intelligence. Le deep learning (ou apprentissage profond) et le machine learning (ou apprentissage automatique) sont des sous-catégories d'IA.



1 Intelligence artificielle
2 Machine learning
3 Deep learning

3.1 Machine learning

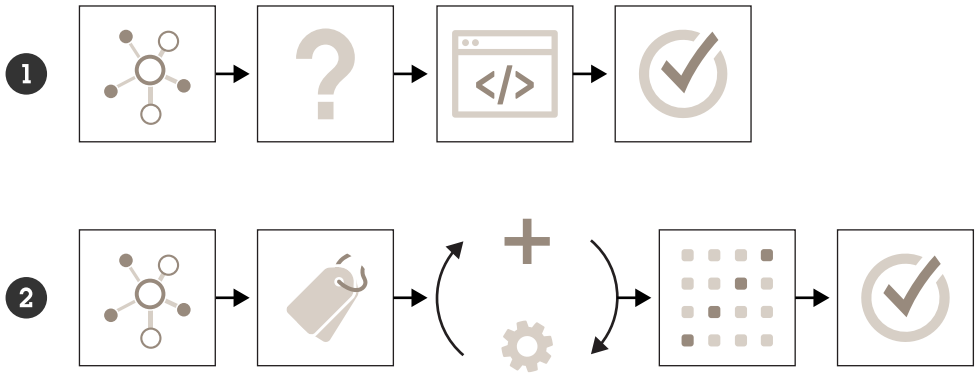
Le machine learning est un sous-ensemble de l'IA qui utilise les algorithmes d'apprentissage statistique pour créer des systèmes capables d'apprendre et de se développer automatiquement pendant la formation sans y être explicitement programmés.

Dans cette partie, nous faisons la distinction entre la programmation traditionnelle et le machine learning dans le contexte de la *vision par ordinateur*, cette discipline qui permet aux ordinateurs de comprendre ce qu'il se passe dans une scène grâce à l'analyse d'images et de vidéos.

La vision par ordinateur à programmation traditionnelle est basée sur des méthodes qui calculent les *caractéristiques* d'une image, par exemple, les programmes informatiques qui recherchent les bords prononcés et les repères de sommet. Ces caractéristiques doivent être définies manuellement par un développeur d'algorithme qui connaît les éléments importants dans les données d'image. Le développeur combine alors ces caractéristiques de l'algorithme pour conclure ce qui se trouve dans la scène.

Les algorithmes de machine learning créent automatiquement un modèle mathématique à l'aide d'un nombre considérable de données d'échantillon, *les données d'apprentissage*, pour acquérir la capacité de prendre des décisions en calculant les résultats sans y être particulièrement programmés. Les caractéristiques sont toujours créées à la main mais la façon dont les caractéristiques sont combinées est apprise par l'algorithme lui-même par l'exposition à un grand nombre de données d'apprentissage étiquetées ou *annotées*. Dans ce livre, nous appelons cette technique d'utilisation de caractéristiques créées à la main dans des combinaisons apprises, *le machine learning classique*.

En d'autres termes, pour une application de machine learning, nous devons apprendre à l'ordinateur à acquérir le programme que nous souhaitons. Les données sont collectées, puis annotées par l'homme, parfois assisté par des pré-annotations effectuées par des ordinateurs serveurs. Le résultat est alimenté dans le système et ce processus continue jusqu'à ce que l'application ait appris suffisamment à détecter ce que nous souhaitons qu'elle détecte, par exemple, un type spécifique de véhicule. Le modèle appris devient le programme. À noter que lorsque le programme est terminé, le système n'apprend plus rien de nouveau.



1 *Programmation traditionnelle :*
Les données sont collectées. Les critères du programme sont définis. Le programme est codé (par l'homme). Terminé.
2 *Machine learning :*
Les données sont collectées. Les données sont étiquetées. Le modèle subit une procédure d'apprentissage itérative. Le modèle appris finalisé devient le programme. Terminé.

L'avantage de l'IA sur la programmation traditionnelle, dans le cadre de la création d'un programme de vision par ordinateur, réside dans la possibilité de traiter un très grand nombre de données. Un ordinateur

peut passer en revue des milliers d'images sans perdre sa concentration, alors que les programmateurs humains se fatigueront et perdront leur concentration au bout d'un moment. Ainsi, l'IA peut rendre l'application considérablement plus précise. Cependant, plus l'application est compliquée, plus il sera difficile pour la machine de produire le résultat attendu.

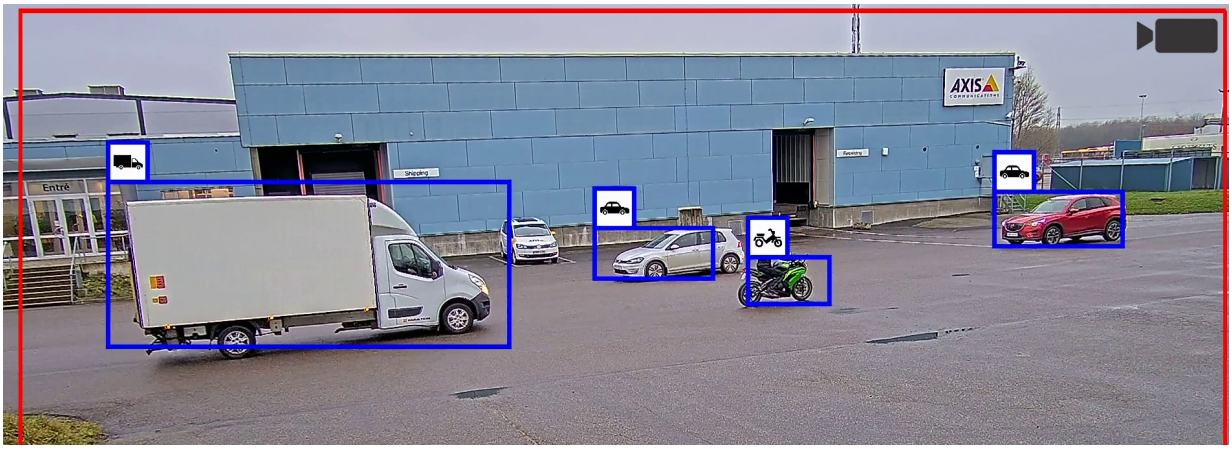
3.2 Deep learning

Le deep learning est une version améliorée du machine learning dans laquelle l'extraction de caractéristiques et la façon dont on combine ces caractéristiques, dans des structures profondes de règles pour produire une sortie, sont apprises d'une façon guidée par les données. L'algorithme peut automatiquement définir les caractéristiques à rechercher dans les données d'apprentissage. Il peut également apprendre des structures très profondes de combinaisons chaînées de caractéristiques.

Le cœur des algorithmes utilisés en deep learning est inspiré par le fonctionnement des neurones et l'utilisation de ceux-ci par le cerveau pour former un niveau supérieur de connaissances en combinant les sorties neuronales dans une hiérarchie profonde, ou un *réseau* de règles chaînées. Le cerveau est un système dans lequel les combinaisons elles-mêmes sont également formées par les neurones, supprimant la distinction entre l'extraction des caractéristiques et la combinaison des caractéristiques, les rendant identiques d'une certaines façon. Ces structures sont simulées par les chercheurs dans ce qu'on appelle *les réseaux de neurones artificiels,* ce qui représente le type d'algorithme le plus utilisé en deep learning. Voir l'annexe du présent document pour un aperçu rapide des réseaux de neurones.

À l'aide des algorithmes de deep learning, il est possible de créer des détecteurs visuels complexes et de les former automatiquement à détecter des objets très complexes, quelles que soient l'échelle, la rotation et les autres variations.

La raison derrière cette souplesse réside dans le fait que les systèmes de deep learning peuvent apprendre à partir d'un plus grand nombre de données, d'une plus grande variété, que les systèmes de machine learning classiques. Dans la plupart des cas, il seront beaucoup plus performants que les algorithmes de vision par ordinateur créés à la main. Cela rend le deep learning particulièrement adapté aux problèmes complexes dans lesquels la combinaison de caractéristiques ne peut être facilement réalisée par des experts humains, telles que la classification d'images, le traitement du langage et la détection d'objets.



La détection d'objets basée sur le deep learning peut classer des objets complexes. Dans cet exemple, l'application d'analyse peut non seulement détecter les véhicules, mais également classer le type de véhicule.

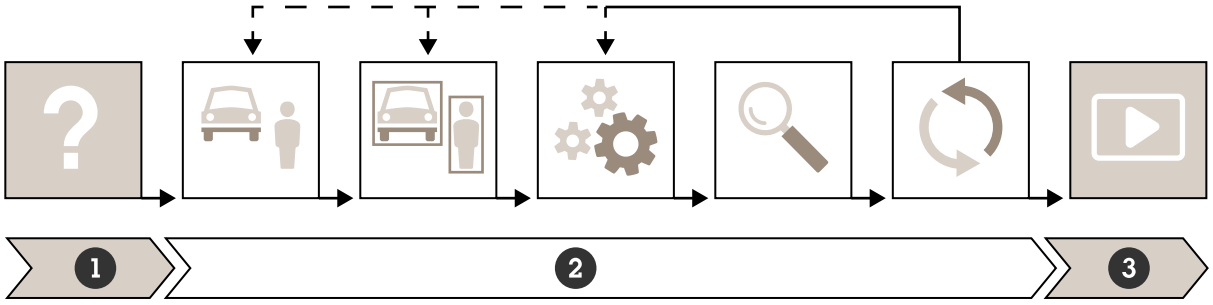
3.3 Machine learning classique contre deep learning

Alors qu'il existe des types similaires d'algorithmes, un algorithme de deep learning utilise généralement une plus large gamme de combinaisons de caractéristiques apprises qu'un algorithme de machine learning classique. Cela signifie que les analyses basées sur l'apprentissage profond peuvent être plus flexibles et peuvent apprendre à exécuter des tâches plus complexes si elles y sont formées.

Cependant, pour des analyses de surveillance en particulier, un algorithme de machine learning classique optimisé et dédié peut être suffisant. Dans un cadre bien spécifié, il peut fournir des résultats similaires à un algorithme d'apprentissage profond tout en nécessitant moins d'opérations mathématiques et peut donc être plus rentable et plus économe en énergie. De plus, il exige beaucoup moins de données d'apprentissage, ce qui réduit considérablement les interventions en matière de développement.

4 Les étapes du machine learning

Le développement d'un algorithme de machine learning suit une série d'étapes et de répétitions, représentées de façon générale ci-dessous, avant qu'une application d'analyse finalisée puisse être déployée. Au cœur d'une application d'analyse, on trouve un ou plusieurs algorithmes, par exemple un détecteur d'objets. Dans le cas d'applications basées sur le deep learning, le cœur de l'algorithme est le modèle d'apprentissage profond.



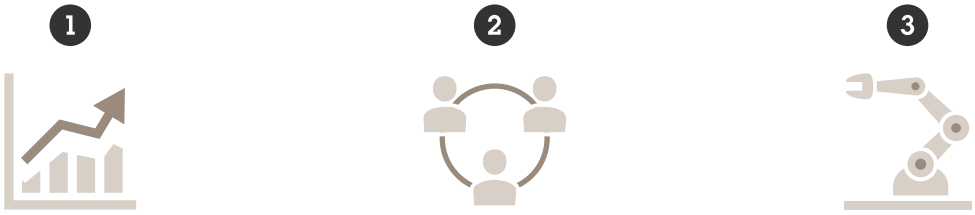
1 *Préparation: Définition de l'objectif de l'application.*
2 *Formation : Collecte de données d'apprentissage. Annotation des données. Formation du modèle. Test du modèle. Si la qualité n'est pas celle attendue, les étapes précédentes sont exécutées à nouveau en un cycle répétitif d'amélioration.*
3 *Déploiement : Installation et utilisation de l'application terminée.*

4.1 Collecte et annotation de données

Pour développer une application d'analyse basée sur l'IA, vous devez collecter un grand nombre de données. En vidéosurveillance, cela se compose habituellement d'images et de clips vidéos d'hommes et de véhicules ou d'autre éléments d'intérêt. Afin de rendre les données reconnaissables pour une machine ou un ordinateur, une procédure d'annotation de données, au cours de laquelle les objets pertinents sont catégorisés et étiquetés, est nécessaire. L'annotation des données est principalement une tâche manuelle et à forte intensité de main-d'œuvre. Les données préparées doivent couvrir une variété suffisamment importante d'échantillons pertinents pour le contexte dans lequel l'application d'analyses sera utilisée.

4.2 formation

La formation, ou l'apprentissage, correspond au moment où le modèle est alimenté par les données annotées et où le cadre d'apprentissage est utilisé pour modifier et améliorer le modèle de façon répétée jusqu'à ce que la qualité souhaitée soit atteinte. En d'autres termes, le modèle est optimisé afin de résoudre la tâche définie. La formation peut être réalisée d'après l'une des trois méthodes principales.



1 *Apprentissage supervisé : le modèle apprend à réaliser des prédictions précises.*
2 *Apprentissage non supervisé : Le modèle apprend à identifier les groupes.*
3 *Apprentissage par renforcement : Le modèle apprend à partir des erreurs*

4.2.1 Apprentissage supervisé

L'apprentissage supervisé est la méthode la plus utilisée actuellement en machine learning. On peut le décrire par l'apprentissage par l'exemple. Les données d'apprentissage sont clairement annotées, ce qui signifie que les données d'entrée sont déjà jumelées avec le résultat de sortie souhaité.

L'apprentissage supervisé nécessite généralement un très grand nombre de données annotées et les performances de l'algorithme formé est directement dépendant de la qualité des données d'apprentissage. L'aspect qualitatif le plus important réside dans le fait d'utiliser un ensemble de données qui représente toutes les potentielles données d'entrée à partir d'une situation de déploiement réelle. Pour les détecteurs d'objets, le développeur doit s'assurer de former l'algorithme avec un large éventail d'images, avec différents exemples d'objets, d'orientation, d'échelles, de situation d'éclairage, d'arrière-plans et de distractions. Uniquement si les données d'apprentissage sont représentatives du cas d'utilisation prévu, l'application d'analyse finale pourra réaliser des prédictions précises aussi lors du traitement de nouvelles données, non vues pendant la phase d'apprentissage.

4.2.2 Apprentissage non supervisé

L'apprentissage non supervisé utilise des algorithmes pour analyser et regrouper des ensembles de données non étiquetés. Ce n'est pas une méthode d'apprentissage habituelle dans le secteur de la surveillance, car le modèle exige un important étalonnage et de nombreux tests tandis que la qualité reste imprévisible.

Les ensembles de données doivent être pertinents pour l'application d'analyse mais ne doivent pas être clairement étiquetés ou marqués. Le travail d'annotation manuelle est éliminé, mais le nombre d'images ou de vidéos requis pour l'apprentissage doit être grandement augmenté, de plusieurs ordres de grandeur. Pendant la phase d'apprentissage, le modèle à former est identifie, épaulé par le cadre d'apprentissage, les caractéristiques communes dans les ensembles de données. Cela lui permet de regrouper les données, pendant la phase de déploiement, conformément aux exemples tout en lui permettant également de détecter les anomalies qui ne correspondent à aucun des groupes appris.

4.2.3 Apprentissage par renforcement

L'apprentissage par renforcement est par exemple utilisé en robotique, dans l'automatisation industrielle et dans la planification stratégique commerciale, mais en raison de la nécessité d'une importante rétroaction,

la méthode a actuellement un usage limité en surveillance. L'apprentissage par renforcement consiste à agir convenablement afin de maximiser l'éventuelle *récompense* dans une situation spécifique, une récompense qui devient plus importante lorsque le modèle fait les bons choix. L'algorithme n'utilise pas de paires d'étiquettes/de données pour l'apprentissage, mais il est plutôt optimisé en testant ses décisions par l'intermédiaire de l'interaction avec l'environnement tout en mesurant la récompense. Le but de l'algorithme est d'apprendre une politique d'actions permettant de maximiser la récompense.

4.3 Tests

Une fois que le modèle est appris, il doit être minutieusement testé. Cette étape contient généralement un partie automatisée complétée par des tests approfondis dans des situations de déploiement dans la vie réelle.

Dans la partie automatisée, l'application est référencée avec de nouveaux ensembles de données, non vus par le modèle au cours de sa formation. Si ces repères ne sont pas là où ils devraient être, le procédure redémarre : de nouvelles données d'apprentissage sont collectées, des annotations sont réalisées ou affinées et le modèle est recyclé.

Lorsque le niveau de qualité souhaité est atteint, un test de terrain démarre. Au cours de ce test, l'application est exposée à des scénarios du monde réel. Leur quantité et leur variation dépendent du champ de l'application. Plus le champ est étroit, moins de variations devront être testées. Plus le champ est large, plus le nombre de tests nécessaires sera élevé.

Les résultats sont à nouveau comparés et évalués. Cette étape peut donc provoquer à nouveau le redémarrage de la procédure. Un autre résultat potentiel pourrait être de définir des préconditions, expliquant un scénario connu dans lequel il n'est pas recommandé d'utiliser l'application ou de ne l'utiliser que partiellement.

4.4 Déploiement

La phase de déploiement est aussi appelée phase de déduction ou de prédiction. *La déduction* ou *la prédiction* correspond à la procédure d'exécution d'un modèle appris de machine learning. L'algorithme utilise ce qu'il a appris au cours de la phase d'apprentissage pour produire sa sortie souhaitée. Dans le contexte des analyses de surveillance, la phase de déduction correspond à l'exécution de l'application sur un système de surveillance qui contrôle des scènes de la vie réelle.

Pour atteindre des performances en temps réel lors de l'exécution d'un algorithme basé sur le machine learning sur des données d'entrée vidéo ou audio, une accélération matérielle spécifique est généralement nécessaire.

5 Analyse en périphérie de réseau

Analyses vidéos hautes performances basées sur un serveur car elles nécessitent plus d'énergie et de refroidissement que ce que peut offrir une caméra. Mais le développement d'algorithme et l'augmentation de la capacité de traitement des dispositifs en périphérie de réseau ces dernières années ont rendu possible l'exécution d'analyses vidéos en périphérie de réseau basée sur l'AI.

Il existe des avantages évidents aux applications d'analyses en périphérie de réseau : elles ont accès à du matériel vidéo non comprimé avec une très faible latence, ce qui permet des applications en temps réel tout en évitant les frais additionnels et la complexité du déplacement des données dans le cloud pour les

calculs. Les analyses en périphérie de réseau s'accompagnent également de moins de frais matériels et de déploiement car les ressources serveur nécessaires dans le système de surveillance sont moindres.

Certaines applications peuvent bénéficier de l'utilisation d'une combinaison de traitement sur serveur et en périphérie de réseau avec le prétraitement sur la caméra et un traitement plus approfondi sur le serveur. Un tel système hybride peut faciliter une mise à l'échelle rentable des applications d'analyse en travaillant sur plusieurs flux de caméra.

6 Accélération matérielle

Tandis que vous pouvez souvent exécuter une application d'analyse spécifique sur plusieurs types de plateformes, l'utilisation de l'accélération matérielle dédiée atteint un niveau de performances bien plus élevé lorsque l'énergie est limitée. Les accélérateurs matériels permettent une mise en œuvre à faible consommation d'énergie des applications d'analyse. Ils peuvent être complétés par des ressources informatiques sur serveur ou cloud le cas échéant.

- **GPU (unité de traitement graphique)**. Les GPU ont été principalement développées pour les applications de traitement graphique mais elles sont également utilisées pour l'accélération de l'IA sur les plateformes de serveur ou de cloud. Alors qu'elles sont également parfois utilisées dans les systèmes intégrés (en périphérie de réseau), les GPU ne sont pas optimales, du point de vue de la faible consommation d'énergie, pour les tâches de déduction de machine learning.
- **MLPU (unité de traitement de machine learning)**. Une MLPU peut accélérer la déduction d'algorithmes spécifiques de machine learning classique pour la résolution de tâches de vision par ordinateur avec une très faible consommation d'énergie. Elle est conçue pour la détection d'objets en temps réel d'un nombre limité de types d'objets simultanés, par exemple, des hommes et des véhicules.
- **DLPU (unité de traitement de deep learning)**. Les caméras équipées d'une DLPU intégrée peuvent accélérer la déduction générale d'algorithme de deep learning avec une faible consommation d'énergie, ce qui permet une classification d'objets plus granulaire.

7 L'IA n'en est qu'au tout début de son développement

Cela peut être tentant de comparer le potentiel d'une solution d'IA et ce qu'un être humain peut réaliser. Tandis que les opérateurs de vidéosurveillance ne peuvent être entièrement vigilants que sur une courte durée, un ordinateur peut continuer à traiter d'importantes quantités de données extrêmement rapidement sans même fatiguer. Mais ce serait une incompréhension fondamentale de supposer que les solutions d'IA pourraient remplacer l'opérateur humain. La véritable force réside dans une combinaison réaliste : tirer profit des solutions d'IA pour améliorer et développer l'efficacité d'un opérateur humain.

Les solutions de machine learning ou de deep learning sont souvent décrites comme disposant de la capacité d'apprendre ou de se développer automatiquement grâce à l'expérience. Mais les systèmes d'IA disponibles aujourd'hui n'apprennent *pas* automatiquement de nouvelles compétences après leur déploiement et ne se souviendront *pas* d'événements spécifiques qui se sont produits. Pour améliorer les performances du système, ils doivent être recyclés avec des données plus précises et de meilleure qualité au cours de sessions d'apprentissage supervisées. L'apprentissage non supervisé nécessite généralement un grand nombre de données afin de générer des groupes et il n'est donc pas utilisé dans les applications de vidéosurveillance. Il est plutôt utilisé aujourd'hui principalement pour l'analyse d'importants ensembles de données afin de détecter des anomalies, dans des transactions financières par exemple. La plupart des

approches mises en avant en tant qu'« autoapprentissage » en termes de vidéosurveillance sont basées sur une analyse des données statistiques et pas sur le recyclage des modèles d'apprentissage profond.

L'expérience humaine bat toujours de nombreuses applications d'analyse basées sur l'IA à des fins de surveillance. En particulier celles qui sont supposées exécuter des tâches très générales et lorsque la compréhension du contexte est essentielle. Une application basée sur le machine learning pourrait réussir à détecter une « personne qui court » si elle y a été formée en particulier mais contrairement à un humain qui peut contextualiser les données, l'application ne peut comprendre la raison pour laquelle la personne court, pour prendre le bus ou pour fuir un officier de police lui courant après. Malgré les promesses des entreprises appliquant l'IA dans leurs applications d'analyse pour la surveillance, l'application ne peut pas encore comprendre ce qu'elle voit sur la vidéo avec la même perspicacité qu'un être humain.

Pour la même raison, les applications d'analyse basées sur l'IA peuvent également déclencher de fausses alarmes ou passer à côté d'une alarme. Habituellement, cela pourrait se produire dans un environnement complexe avec beaucoup de mouvement. Mais il pourrait également s'agir, par exemple, d'une personne qui porte un gros objet, ce qui empêcherait probablement l'application de classifier correctement les caractéristiques humaines.

Les analyses basées sur l'IA aujourd'hui devraient être utilisées pour des fonctions d'assistance, par exemple, pour déterminer approximativement la pertinence d'un incident avant d'alerter un opérateur humain afin de décider de la réponse à donner. Ainsi, l'IA est utilisée pour atteindre l'évolutivité et l'opérateur humain est nécessaire pour évaluer les incidents potentiels.

8 Étude sur les performances optimales des analyses

Afin de guider les attentes qualitatives d'une application d'analyses basée sur l'IA, il est recommandé d'étudier et de comprendre avec soin les préconditions et limites connues, généralement listées dans la documentation de l'application.

Chaque installation de surveillance est unique et les performances de l'application doivent être évaluées sur chaque site. Si la qualité n'atteint pas le niveau escompté ou attendu, il est fortement recommandé de ne pas uniquement concentrer les recherches sur l'application elle-même. Toutes les recherches doivent être réalisées à un niveau holistique car les performances d'une application d'analyses dépendent de tant de facteurs, la plupart pouvant être optimisés si on connaît leur impact. Parmi ces facteur, on trouve par exemple le matériel de la caméra, la qualité vidéo, la dynamique de la scène, le niveau d'éclairage, ainsi que la configuration, la position et le sens de la caméra.

8.1 Capacité à exploiter l'image

On dit souvent que la qualité d'image dépend de la haute résolution et de l'importante sensibilité à la lumière de la caméra. Bien que l'importance de ces facteurs ne puissent être remise en question, il en existe bien sûr d'autres qui sont tout aussi importants pour *la facilité d'utilisation* effective d'une image ou d'une vidéo. Par exemple, un flux vidéo d'excellente qualité de la caméra de surveillance la plus évoluée peut s'avérer inexploitable si la scène n'est pas suffisamment éclairée de nuit, si la caméra n'est pas cadrée ou si la connexion au système est coupée.

Le placement de la caméra devrait être soigneusement étudié avant sa mise en place. Pour que les analyses vidéos s'exécutent comme prévu, la caméra doit être positionnée de telle sorte que la vue sur la scène surveillée soit dégagée et sans obstacles.

La facilité d'utilisation d'une image peut également dépendre du cas d'utilisation. Une vidéo qui semble correcte pour un œil humain pourrait ne pas offrir la qualité optimale pour les performances d'une application d'analyse vidéo. En fait, les nombreuses méthodes de traitement de l'image communément utilisées pour améliorer l'apparence d'une vidéo pour un visionnage par l'homme ne sont pas recommandées pour l'analyse vidéo. Ces méthodes peuvent comprendre par exemple l'application de méthodes de réduction de bruit, des méthodes de plage dynamique étendue et des algorithmes d'exposition automatique.

Les caméras vidéos sont aujourd'hui souvent équipées d'un éclairage IR intégré qui leur permet de fonctionner dans le noir complet. Ce qui représente un avantage certain car cela permet d'installer des caméras dans des sites où l'éclairage est difficile en limitant la nécessité d'installer un éclairage supplémentaire. Cependant, si un site est soumis à de fortes pluies ou à des chutes de neige, il est fortement recommandé de ne pas compter sur la lumière provenant de la caméra ou d'un emplacement à proximité immédiate de la caméra. Une lumière trop importante pourrait se refléter directement à l'arrière de la caméra, contre les gouttes de pluie et les flocons de neige, rendant l'analyse impossible. Par contre, avec la lumière ambiante, les analyses risquent de donner de meilleurs résultats même en cas de très mauvais temps.

8.2 Distance de détection

Il est difficile de déterminer une distance de détection maximum pour une application d'analyse basée sur l'IA, une valeur exacte de fiche technique en mètres ou en pieds ne peut être complètement vraie. La qualité d'image, les caractéristiques de la scène, les conditions météorologiques et les propriétés de l'objet, telles que la couleur et la luminosité, ont un impact important sur la distance de détection. Il est évident, par exemple, qu'un objet lumineux contre un arrière-plan sombre lors d'une journée ensoleillée peut être détecté à des distances plus importantes qu'un objet sombre lors d'une journée pluvieuse.

La distance de détection dépend également de la vitesse des objets à détecter. Pour obtenir des résultats précis, une application d'analyse vidéo doit « voir » l'objet pendant suffisamment longtemps. Cette durée dépend des performances de traitement (fréquence d'images) de la plateforme : plus les performances de traitement sont faibles, plus l'objet doit être visible longtemps pour être détecté. Si le temps d'exposition de la caméra ne correspond pas bien à la vitesse de l'objet, le flou de mouvement qui apparaît sur l'image peut également diminuer la précision de détection.

Les objets rapides peuvent être facilement manqués s'ils passent trop près de la caméra. Une personne qui court loin de la caméra, par exemple, pourrait être correctement détectée, tandis qu'une personne qui court à proximité de la caméra à une vitesse identique pourrait entrer et sortir du champ de vision si rapidement qu'aucune alarme ne serait déclenchée.

Dans les analyses basées sur la détection de mouvement, les objets qui se déplacent vers la caméra, ou qui s'en éloignent, représentent un autre défi. La détection sera particulièrement difficile pour les objets à déplacement lent, qui ne provoqueront que de très faibles changements dans l'image par rapport aux mouvements dans la scène.

Généralement, une caméra haute résolution n'offre pas une plus importante distance de détection. Les capacités de traitement requises pour l'exécution d'un algorithme de machine learning sont proportionnelles à la taille des données d'entrée. Cela signifie que la puissance de traitement requise pour analyser l'entière résolution d'une caméra 4K est au moins quatre fois supérieure à une caméra 1080p. Il est très courant d'exécuter des applications basées sur l'IA sur une plus faible résolution que celle offerte par la caméra ou le flux en raison des limites des capacités de traitement de la caméra.

8.3 Configuration des enregistrements et des alarmes

En raison des différents niveaux de filtres appliqués, les analyses d'objet ne génèrent que très peu de fausses alarmes. Mais les performances des analyses d'objets sont conformes aux prévisions uniquement lorsque les préconditions listées sont toutes remplies. Dans d'autres cas, ces analyses pourraient passer à côté d'événement importants.

S'il n'est pas absolument certain que toutes les conditions seront en permanence remplies, il est par conséquent recommandé d'adopter une approche prudente et de configurer le système afin qu'une classification spécifique des objets ne soit pas le seul déclencheur d'alarme. Cela provoquera plus de fausses alarmes mais réduira également le risque de manquer des événements importants. Lorsque les alarmes et les déclencheurs vont directement dans un centre de contrôle des alarmes, chaque fausse alarme devient très chère. Il est évident que la classification des objets doit être fiable afin de filtrer les alarmes indésirables. Mais la solution d'enregistrement peut et doit toujours être configurée pour ne pas uniquement reposer sur la classification d'objets. Dans le cas d'une véritable alarme manquée, cette configuration vous permet d'évaluer, à partir de l'enregistrement, la raison de cet oubli et ainsi d'améliorer l'installation et la configuration dans leur ensemble.

Si la classification d'objets est réalisée sur le serveur au cours de la recherche d'incident, il est recommandé de configurer le système pour un enregistrement en continu et de ne pas filtrer du tout l'enregistrement initial. L'enregistrement en continu consomme beaucoup d'espace de stockage mais cela est compensé dans une certaine mesure par les algorithmes modernes de compression tels que Zipstream.

8.4 Maintenance

Une installation de surveillance doit être régulièrement entretenue. Des inspections physiques, et pas seulement le visionnage de la vidéo par l'interface VMS, sont recommandées afin de découvrir et supprimer tout élément qui pourrait perturber ou bloquer le champ de vision. C'est également important dans les installations standard d'enregistrement seul, mais cela est d'autant plus essentiel lorsqu'on utilise des analyses.

Dans le contexte de la détection de mouvement vidéo de base, un obstacle ordinaire tel qu'une toile d'araignée qui se balance dans le vent pourrait augmenter le nombre d'alarmes, ce qui provoquerait une augmentation inutile de la consommation d'espace de stockage. Avec l'analyse d'objets, la toile pourrait

tout simplement créer une zone à exclure dans l'espace de détection. Ses fils pourraient obstruer les objets et fortement réduire les chances de détection et de classification.

Les toiles d'araignées pourraient perturber le champ de vision d'une caméra de surveillance.

De la saleté sur le verre avant ou sur la bulle de la caméra ne devrait pas poser de problèmes en pleine journée. Mais en conditions de faible luminosité, la lumière qui atteint une bulle sale sur le côté, par exemple la lumière des phares d'une voiture, pourrait provoquer des reflets imprévus qui pourraient réduire la précision de détection.

La maintenance de la scène elle-même est tout aussi importante que l'entretien de la caméra. Au cours du cycle de service d'une caméra, de nombreux événements peuvent se produire dans la scène sous surveillance. Une simple comparaison de l'image avant/après révèlera les éventuels problèmes. À quoi ressemblait la scène lorsque la caméra a été déployée et à quoi ressemble-t-elle aujourd'hui ? Est-ce que la zone de détection doit être ajustée ? Le champ de vision de la caméra devrait-il être ajusté ou la caméra devrait-elle être déplacé à un emplacement différent ?

9 Confidentialité et intégrité personnelle

Lorsqu'on travaille dans la sécurité et la surveillance, il faut trouver le juste équilibre entre les droits individuels à la confidentialité et à l'intégrité personnelle et l'ambition d'augmenter la sécurité en prévenant les crimes et en permettant les enquêtes scientifiques. Dans le cas d'une utilisation et d'une installation spécifiques, cela exige un examen minutieux et éthique ainsi que la compréhension et l'application de la législation locale. Cela place également des exigences sur la solution afin de garantir la cybersécurité et d'éviter l'accès non intentionnel au matériel vidéo par exemple. Dans le même temps, les analyses en périphérie de réseau et la génération de métadonnées à des fins statistiques peuvent augmenter la protection de la confidentialité uniquement si des données anonymisées sont transmises pour un traitement ultérieur.

Avec l'augmentation des applications d'analyses automatisées dans les systèmes de surveillance, certains aspects nouveaux doivent être pris en compte. Puisque les applications d'analyse s'accompagnent d'un risque de fausses détections, il est important que la procédure de décision implique un opérateur ou un utilisateur expérimenté. On appelle souvent ce phénomène « garder un homme dans la boucle ». De plus, il est important de reconnaître que la décision humaine peut être affectée par la façon dont l'alarme est

générée et présentée. Sans une formation correcte et une connaissance du fonctionnement de la solution d'analyse, on pourrait tirer de mauvaises conclusions.

Une inquiétude supplémentaire pourrait être provoquée par la façon dont les algorithmes de deep learning sont développés et pour certains cas d'utilisation, cela nécessite une approche prudente lorsqu'on applique la technologie. La qualité de ces algorithmes est fondamentalement liée aux ensembles de données, c'est-à-dire les vidéos et les images, utilisées pour former l'algorithme. Des tests ont démontré que si ces supports ne sont pas sélectionnés soigneusement, certains systèmes d'IA pourraient présenter des biais éthiques et de genre dans les détections. Cela a provoqué une discussion ouverte et a donné lieu à des limites et des activités législatives afin de garantir que ces aspects sont abordés au cours du développement des systèmes.

Puisque l'IA est de plus en plus appliquée dans le domaine de la surveillance, il est important de compenser les avantages de l'efficacité opérationnelle et de nouveaux cas d'utilisation éventuels par une discussion attentive sur les lieux et les moments où appliquer la technologie.

10 Annexe

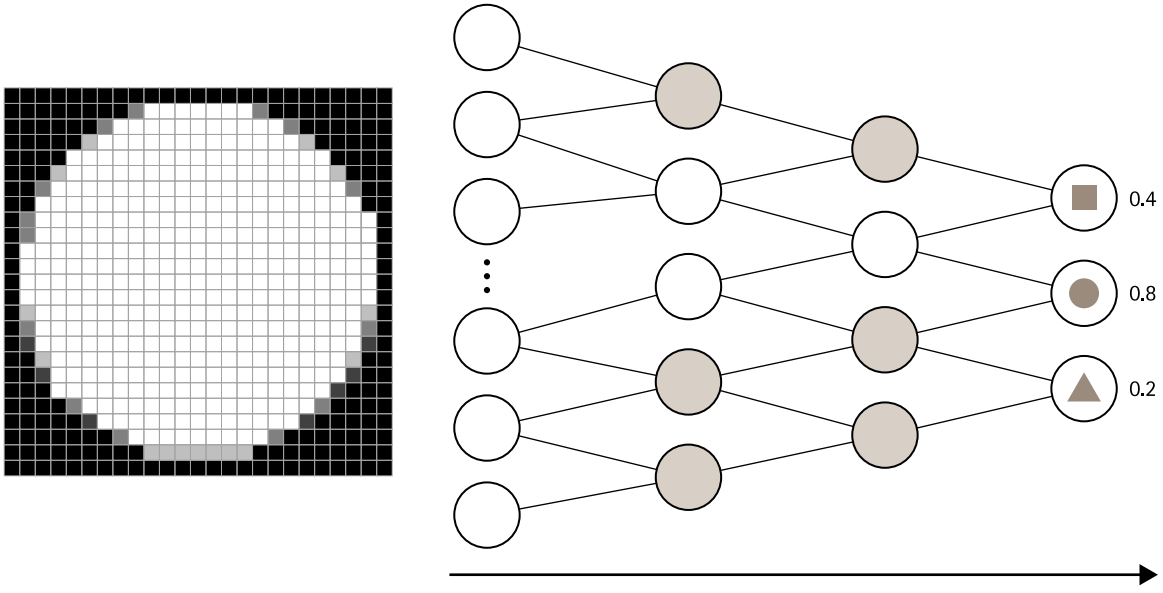
Cette annexe fournit des informations générales sur les réseaux de neurones artificiels qui forment la base du deep learning.

10.1 Réseaux de neurones

Les réseaux de neurones sont une famille d'algorithmes utilisés pour reconnaître les relations dans les ensembles de données grâce à une procédure assez similaire au fonctionnement du cerveau humain. Un réseau de neurones se compose d'une hiérarchie de plusieurs couches de ce qu'on nomme nœuds ou neurones qui sont interconnectés. Les informations passent le long de ces connexions, de la couche d'entrée à la couche de sortie en passant par le réseau.

L'hypothèse de fonctionnement des réseaux de neurones est qu'un échantillon de données d'entrée peut être réduit à un ensemble limité de caractéristiques, de qui crée une bonne représentation des données d'entrée. Ces caractéristiques peuvent alors être combinées et permettront de classer les données d'entrée, par exemple en décrivant le contenu d'une image.

L'illustration ci-dessous montre un exemple dans lequel un réseau de neurones est utilisé pour identifier à quelle classe appartient l'image d'entrée. Chaque pixel de l'image est représenté par un seul nœud d'entrée. Tous les nœuds d'entrée sont couplés aux nœuds de la première couche. Ceux-ci produisent des valeurs de sortie qui sont transmises en tant que valeurs d'entrée à la seconde couche et ainsi de suite. Dans chaque couche, les fonctions de pondération, les valeurs de biais et les fonctions d'activation sont également impliquées dans la procédure.



Exemple d'une image d'entrée (à gauche) et d'un réseau de neurones (à droite). Lorsque la couche de sortie est atteinte, le réseau a conclu les probabilités pour chaque catégorie possible (carré, cercle ou triangle). La catégorie avec la valeur de probabilité la plus élevée représente la forme la plus probable de l'image d'entrée.

Cette procédure est appelée *propagation avant*. En cas de non-correspondance du résultat de la propagation avant, les paramètres du réseau sont légèrement modifiés par l'intermédiaire de la *rétropropagation*. Durant cette procédure d'apprentissage répétitive, les performances du réseau s'améliorent petit à petit.

Après le déploiement, un réseau de neurones n'a, en général, aucun souvenir des pas avant précédents. Cela signifie qu'il ne s'améliore pas avec le temps et qu'il ne peut que détecter les types d'objets, ou résoudre les types de tâches pour lesquels il a été formé.

10.2 Réseaux de neurones à convolution (CNN)

Les réseaux de neurones à convolution (CNN) sont une sous-catégorie de réseaux neuronaux artificiels qui se sont avérés être particulièrement adaptés aux tâches de vision par ordinateur et ils sont au cœur du développement rapide de l'apprentissage profond. Dans le cas de la vision par ordinateur, le réseau est formé pour rechercher automatiquement les caractéristiques distinctives des images, ressemblant à des bords, des coins, des différences de couleurs, permettant d'identifier effectivement les formes des objets dans une image.

La principale opération pour accomplir cette tâche est l'opération mathématique appelée *convolution.* C'est une opération très efficace puisque la sortie de chaque nœud en particulier dépend uniquement d'un environnement limité dans les données d'entrée, qui a été produit par la couche précédente, plutôt que par l'utilisation de l'ensemble du volume de données d'entrée. En d'autres termes, dans un CNN, chaque nœud n'est pas connecté à tous les nœuds de la couche précédente mais uniquement à un petit sous-ensemble. Les convolutions sont complétées par d'autres opérations qui réduisent la taille des données tout en conservant les informations les plus utiles. Comme dans un réseau de neurones artificiels standard, les données deviennent de plus en plus abstraites au fur et à mesure qu'on s'enfonce dans le réseau.

Pendant la phase d'apprentissage, le CNN apprend la meilleure façon d'appliquer les couches. C'est ainsi que les convolutions devraient combiner les caractéristiques à partir de la couche précédente pour que la sortie du réseau corresponde autant que possible avec les annotations des données d'apprentissage. Au cours de la déduction, le réseau de neurones à convolution formé applique alors successivement les couches de convolutions qui résultent de l'apprentissage.

À propos d'Axis Communications

En concevant des solutions qui améliorent la sécurité et les performances de l'entreprise, Axis crée monde plus clairvoyant et plus sûr. En tant qu'entreprise de technologie de réseau et leader de l'industrie, Axis propose des solutions de vidéosurveillance, de contrôle d'accès, d'interphonie et de systèmes audio. Les performances de ces solutions sont améliorées grâce à des applications d'analyse intelligentes et une formation de haute qualité.

Axis emploie près de 4 000 personnes dans plus de 50 pays et collabore avec des partenaires technologiques et d'intégration de systèmes dans le monde entier pour fournir des solutions clients adaptées. Axis a été fondée en 1984 et le siège social se trouve à Lund, en Suède