

IA na análise de vídeo

Considerações para análises baseadas em aprendizado de máquina e aprendizado profundo

Março 2021

Sumário

1	Resumo	3
2	Introdução	4
3	IA, aprendizado de máquina e aprendizado profundo	4
	3.1 Aprendizado de máquina	5
	3.2 Aprendizagem profunda	6
	3.3 Aprendizado de máquina clássico versus aprendizado profundo	6
4	Os estágios do aprendizado de máquina	7
	4.1 Coleta e registro de dados	7
	4.2 Treinamentos	7
	4.3 Teste	9
	4.4 Implantação	9
5	Análise baseada na borda	9
6	Aceleração de hardware	10
7	A IA ainda está em seu desenvolvimento inicial	10
8	Considerações para o desempenho analítico ideal	11
	8.1 Aproveitamento de imagem	11
	8.2 Distância de detecção	12
	8.3 Configuração de alarmes e gravação	12
	8.4 Manutenção	13
9	Privacidade e integridade pessoal	13
10	Apêndice	15
	10.1 Redes neurais	15
	10.2 Redes neurais convolucionais (CNN, Convolutional neural networks)	16

1 Resumo

A análise de vídeo baseada em IA é um dos tópicos mais discutidos no setor de videomonitoramento. Alguns dos aplicativos podem acelerar substancialmente a análise de dados e automatizar tarefas repetitivas. Entretanto, as soluções de IA de hoje não podem substituir a experiência do operador humano e as habilidades de tomada de decisão. A vantagem consiste em uma combinação: aproveitar as soluções de IA para melhorar e aumentar a eficiência humana.

O conceito de IA incorpora algoritmos de aprendizado de máquina e algoritmos de aprendizado profundo. Ambos os tipos constroem um modelo matemático automaticamente, usando quantidades substanciais de dados de amostra (*dados de treinamento*), para obterem a capacidade de calcular resultados sem serem especificamente programados para isso. Um algoritmo de IA é desenvolvido por meio de um processo iterativo, no qual um ciclo de coleta de dados de treinamento, identificação dos dados de treinamento, uso dos dados identificados para treinar o algoritmo e teste do algoritmo treinado é repetido até que o nível de qualidade desejado seja alcançado. Depois disso, o algoritmo está pronto para uso em um aplicativo de análise, que pode ser comprado e implantado em uma zona de vigilância. Nesse ponto, todo o treinamento está concluído e o aplicativo não aprenderá nada de novo.

Uma tarefa típica para análises de vídeo baseadas em IA é detectar visualmente humanos e veículos em uma transmissão de vídeo e distingui-los. Um algoritmo de *aprendizado de máquina* aprendeu a combinação de atributos visuais que definem esses objetos. Um algoritmo de *aprendizado profundo* é mais aprimorado e pode – se treinado para isso – detectar objetos muito mais complexos. Mas também requer esforços significativamente maiores para o desenvolvimento e o treinamento, e muito mais recursos computacionais quando o aplicativo finalizado é colocado em uso. Para necessidades de vigilância bem definidas, deve-se considerar se um aplicativo dedicado e otimizado de aprendizado de máquina pode ser suficiente.

O desenvolvimento de algoritmos e o aumento da capacidade de processamento das câmeras possibilitaram análises de vídeo baseadas em IA avançadas diretamente na câmera (baseadas em borda), em vez de realizar os cálculos em um servidor (baseadas em servidor). Isso permite uma melhor funcionalidade em tempo real, pois os aplicativos têm acesso imediato ao material de vídeo não compactado. Com aceleradores de hardware dedicados nas câmeras, como MLPU (unidade de processamento de aprendizado de máquina) e DLPU (unidade de processamento de aprendizado profundo), a análise baseada em borda pode ser implementada com mais eficiência energética do que com uma CPU ou GPU (unidade de processamento gráfico).

Antes de um aplicativo de análise de vídeo baseada em IA ser instalado, as recomendações do fabricante de acordo com as pré-condições e limitações conhecidas devem ser cuidadosamente estudadas e seguidas. Cada instalação de vigilância é única e o desempenho do aplicativo deve ser avaliado em cada local. Se a qualidade for menor do que o esperado, as investigações devem ser feitas em um nível global, e não se concentrar apenas no aplicativo de análise. O desempenho da análise de vídeo depende de muitos fatores relacionados ao hardware da câmera, configuração da câmera, qualidade do vídeo, dinâmica da cena e iluminação. Em muitos casos, conhecer o impacto desses fatores e otimizá-los adequadamente possibilita o aumento de desempenho da análise de vídeo na instalação.

Como a IA tem sido implantada cada vez mais na vigilância, as vantagens da eficiência operacional e os novos casos de uso devem ser avaliados com uma discussão cuidadosa sobre quando e onde implantar essa tecnologia.

2 Introdução

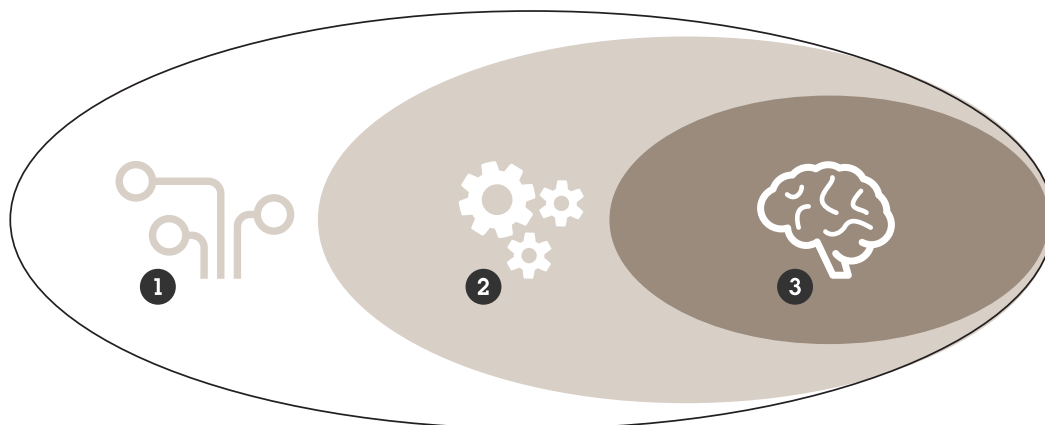
A IA, inteligência artificial, foi desenvolvida e discutida desde que os primeiros computadores foram inventados. Embora os tipos mais revolucionários ainda não tenham sido lançados, atualmente, as tecnologias baseadas em IA são amplamente utilizadas para realizar tarefas claramente definidas em aplicativos, como reconhecimento de voz, mecanismos de pesquisa e assistentes virtuais. A IA também é cada vez mais usada na área de saúde, onde fornece recursos valiosos em, por exemplo, diagnósticos de raios-x e análises de exames de retina.

A análise de vídeo baseada em IA é um dos tópicos mais discutidos no setor de videomonitoramento; e as expectativas são altas. Existem aplicativos no mercado que usam algoritmos de IA para acelerar a análise de dados e automatizar tarefas repetitivas de forma eficiente. Contudo, em um contexto mais amplo de vigilância, a IA, hoje e no futuro próximo, deve ser vista como apenas um elemento, entre vários outros, no processo de construção de soluções precisas.

Este livro branco apresenta um cenário tecnológico sobre os algoritmos de aprendizado de máquina e aprendizado profundo e como eles podem ser desenvolvidos e aplicados para a análise de vídeo. Isso inclui um breve resumo sobre o hardware de aceleração de IA e os prós e contras de executar análises baseadas em IA na borda em comparação com um servidor. O artigo também analisa como as pré-condições para o desempenho da análise de vídeo baseada em IA podem ser otimizadas, considerando uma variedade de fatores.

3 IA, aprendizado de máquina e aprendizado profundo

A inteligência artificial (IA) é um conceito amplo associado a máquinas que podem resolver tarefas complexas enquanto demonstram atributos aparentemente inteligentes. O aprendizado profundo e o aprendizado de máquina são subconjuntos da IA.



- 1 *Inteligência artificial*
- 2 *Aprendizado de máquina*
- 3 *Aprendizagem profunda*

3.1 Aprendizado de máquina

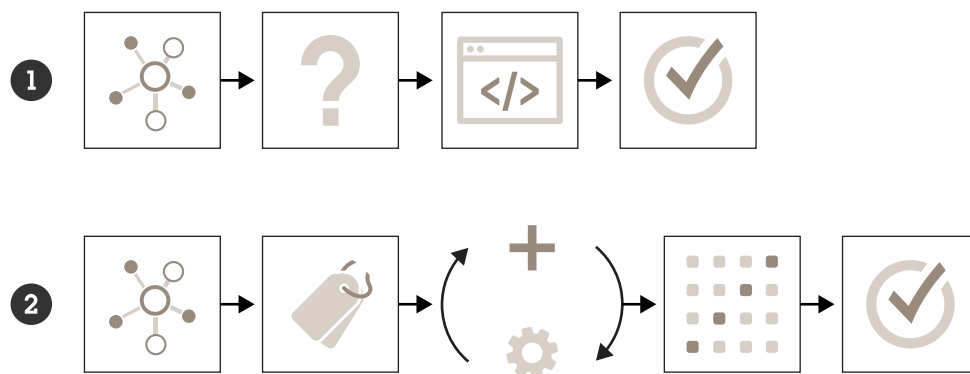
O aprendizado de máquina é um subconjunto da IA que usa algoritmos de aprendizado estatísticos para construir sistemas com a capacidade de aprender e melhorar automaticamente durante o treinamento sem serem explicitamente programados.

Nesta seção, apresentamos uma distinção entre a programação tradicional e o aprendizado de máquina no contexto da *visão computacional* – a disciplina de fazer com que os computadores entendam o que está acontecendo em uma cena por meio da análise de imagens ou vídeos.

A visão computacional tradicionalmente programada é baseada em métodos que calculam os *atributos* de uma imagem, por exemplo, programas de computador que procuram bordas pronunciadas e cantos. Esses atributos precisam ser definidos manualmente por um desenvolvedor de algoritmo que saiba o que é importante nos dados da imagem. O desenvolvedor então combina esses atributos para que o algoritmo determine o que é encontrado na cena.

Os algoritmos de aprendizado de máquina constroem automaticamente um modelo matemático usando quantidades substanciais de dados de amostra – *dados de treinamento* – para obterem a capacidade de tomar decisões calculando resultados sem serem especificamente programados para isso. Os atributos ainda são feitos manualmente, mas como combiná-los é aprendido pelo próprio algoritmo por meio da exposição a grandes quantidades de dados de treinamento identificados ou *registrados*. Neste artigo, essa técnica de usar atributos manuais em combinações aprendidas é chamado de *aprendizado de máquina clássico*.

Em outras palavras, para um aplicativo de aprendizado de máquina, é preciso treinar o computador para obter o programa desejado. Os dados são coletados e registrados por pessoas, às vezes, com a ajuda de um registro prévio de computadores servidores. O resultado é alimentado no sistema e esse processo continua até que o aplicativo tenha aprendido o suficiente para detectar o desejado, por exemplo, um tipo específico de veículo. O modelo treinado se torna o programa. Observe que quando o programa é concluído, o sistema não aprende nada de novo.



- 1 *Programação tradicional:*
Os dados são coletados. Os critérios do programa são definidos. O programa é codificado (por uma pessoa). Pronto.
- 2 *Aprendizado de máquina:*
Os dados são coletados. Os dados são identificados. O modelo passa por um processo de treinamento iterativo. O modelo com treinamento finalizado se torna o programa. Pronto.

A vantagem da IA em relação à programação tradicional – ao construir um programa de visão computacional – é a capacidade de processar dados extensos. Um computador pode analisar milhares de imagens sem perder o foco, enquanto um programador humano ficará cansado e sem atenção depois de um

tempo. Assim, a IA pode tornar o aplicativo significativamente mais preciso. No entanto, quanto mais complexo for o aplicativo, mais difícil será para a máquina produzir o resultado desejado.

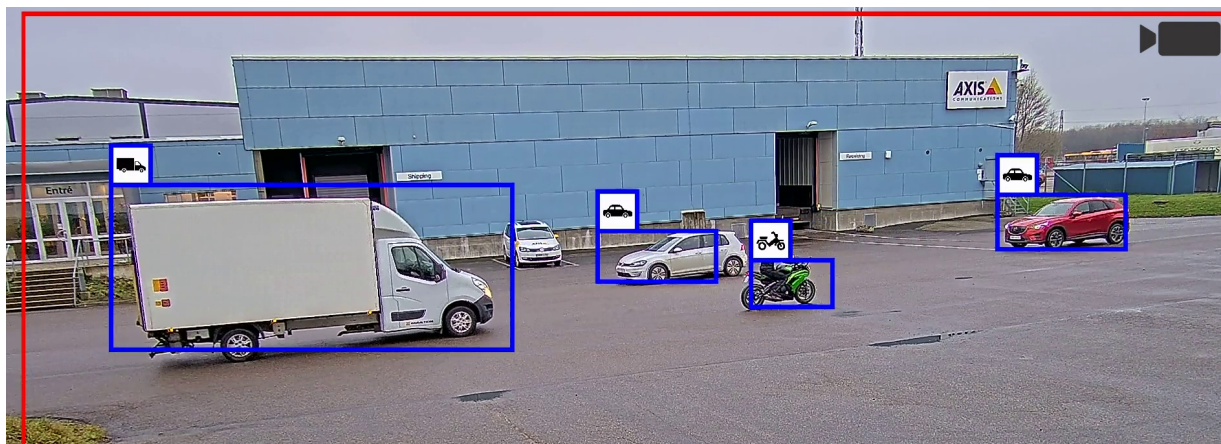
3.2 Aprendizagem profunda

O aprendizado profundo é uma versão aprimorada do aprendizado de máquina, em que tanto a extração de atributos quanto a forma de combiná-los – em estruturas de regras complexas para produzir um resultado – são aprendidos de maneira orientada por dados. O algoritmo pode definir automaticamente quais atributos deve procurar nos dados de treinamento. Ele também pode aprender estruturas muito complexas de combinações de atributos encadeados.

A base dos algoritmos usados no aprendizado profundo é inspirada em como os neurônios funcionam e como o cérebro os usa para formar um conhecimento de nível superior, combinando as saídas dos neurônios em uma hierarquia complexa, ou uma *rede*, de regras encadeadas. O cérebro é um sistema no qual as próprias combinações também são formadas por neurônios, apagando a distinção entre a extração do atributo e a combinação dos atributos, tornando-os iguais em algum sentido. Essas estruturas foram simuladas por pesquisadores em algo chamado *redes neurais artificiais*, que é o tipo de algoritmo mais amplamente usado no aprendizado profundo. Consulte o apêndice deste documento para obter uma breve visão geral sobre as redes neurais.

Com algoritmos de aprendizado profundo, é possível criar detectores visuais intrincados e treiná-los automaticamente para detectar objetos muito complexos, flexíveis a escala, rotação e outras variações.

A razão por trás dessa flexibilidade é que os sistemas de aprendizado profundo podem aprender com uma quantidade muito maior e mais variada de dados do que os sistemas de aprendizado de máquina clássico. Na maioria dos casos, eles superam significativamente os algoritmos de visão computacional manuais. Isso torna o aprendizado profundo especialmente adequado para problemas complexos, em que a combinação de atributos não pode ser facilmente formada por especialistas humanos, como a classificação de imagens, o processamento de linguagem e a detecção de objetos.



A detecção de objetos com base no aprendizado profundo pode classificar objetos complexos. Neste exemplo, o aplicativo de análise não apenas detecta os veículos, mas também pode classificar o tipo de veículo.

3.3 Aprendizado de máquina clássico versus aprendizado profundo

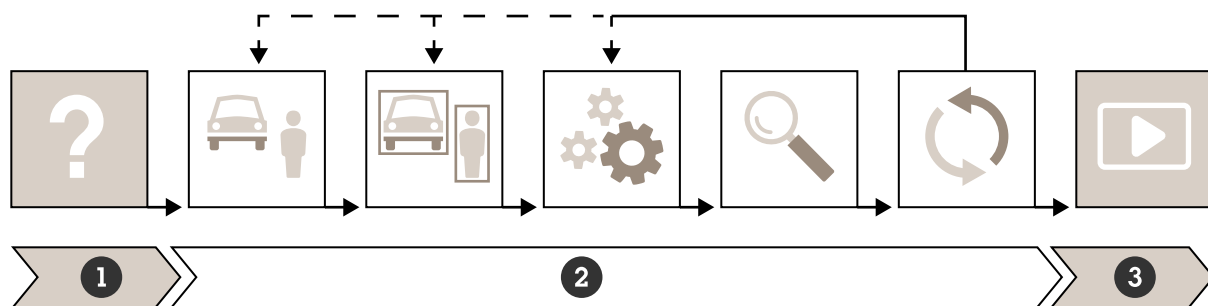
Embora sejam tipos semelhantes de algoritmos, um algoritmo de aprendizado profundo normalmente usa um conjunto muito maior de combinações de atributos aprendidas do que um algoritmo de aprendizado de

máquina clássica. Isso significa que a análise baseada em aprendizado profundo pode ser mais flexível e, se treinada corretamente, aprender a executar tarefas muito mais complexas.

No entanto, para análises de vigilância específicas, um algoritmo de aprendizado de máquina clássico otimizado e dedicado pode ser suficiente. Em um escopo bem especificado, ele pode fornecer resultados semelhantes a de um algoritmo de aprendizado profundo enquanto exige menos operações matemáticas, e, portanto, ser mais econômico e consumir menos energia. Além disso, ele requer muito menos dados de treinamento e isso reduz muito o esforço de desenvolvimento.

4 Os estágios do aprendizado de máquina

O desenvolvimento de um algoritmo de aprendizado de máquina segue uma série de etapas e iterações (apresentadas abaixo de modo genérico), antes que um aplicativo de análise finalizado possa ser implementado. A essência de um aplicativo de análise inclui um ou mais algoritmos, por exemplo, um detector de objetos. No caso de aplicativos baseados em aprendizado profundo, a base do algoritmo é o modelo de aprendizado profundo.



- 1 *Preparação: Definir a finalidade do aplicativo.*
- 2 *Treinamentos: Coletar dados de treinamento. Registrar os dados. Treinar o modelo. Testar o modelo. Se a qualidade não for a esperada, as etapas anteriores serão executadas novamente em um ciclo de melhoria iterativo.*
- 3 *Implantação: Instalar e usar o aplicativo finalizado.*

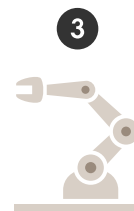
4.1 Coleta e registro de dados

Para desenvolver um aplicativo de análise baseada em IA, é preciso coletar grandes quantidades de dados. No videomonitoramento, isso normalmente inclui imagens e trechos de vídeo de humanos e veículos ou outros objetos de interesse. Para tornar os dados reconhecíveis por uma máquina ou computador, é necessário um processo de registro de dados, em que os objetos relevantes são categorizados e identificados. O registro de dados é, sobretudo, uma tarefa manual e trabalhosa. Os dados preparados precisam cobrir uma variedade grande o suficiente de amostras que sejam relevantes para o contexto em que o aplicativo de análise será usado.

4.2 Treinamentos

Treinamento, ou aprendizagem, é quando o modelo é alimentado com dados registrados e uma estrutura de treinamento é usada para modificá-lo e melhorá-lo iterativamente até que a qualidade desejada seja

alcançada. Ou seja, o modelo é otimizado para resolver a tarefa definida. O treinamento pode ser feito de acordo com um dos três métodos principais.



- 1 *Aprendizagem supervisionada: o modelo aprende a fazer previsões precisas*
- 2 *Aprendizagem não supervisionada: o modelo aprende a identificar agrupamentos*
- 3 *Aprendizagem por reforço: O modelo aprende com os erros*

4.2.1 Aprendizagem supervisionada

A aprendizagem supervisionada é o método mais usado no aprendizado de máquina atualmente. Ela pode ser descrita como aprendizagem por meio de exemplos. Os dados de treinamento são claramente registrados, o que significa que os dados de entrada já estão emparelhados com o resultado de saída desejado.

A aprendizagem supervisionada geralmente requer uma grande quantidade de dados registrados, e o desempenho do algoritmo treinado depende diretamente da qualidade dos dados de treinamento. O aspecto qualitativo mais importante é usar um conjunto de dados que represente todos os potenciais dados de entrada de uma situação real de implantação. Para detectores de objetos, o desenvolvedor deve certificar-se de treinar o algoritmo com uma ampla variedade de imagens, com diferentes instâncias de objetos, orientações, escalas, situações de iluminação, planos de fundo e distrações. O aplicativo de análise final só será capaz de fazer previsões precisas durante o processamento de novos dados, não vistos na fase de treinamento – se os dados de treinamento forem representativos para o caso de uso planejado.

4.2.2 Aprendizagem não supervisionada

A aprendizagem não supervisionada usa algoritmos para analisar e agrupar conjuntos de dados não identificados. Esse não é um método de treinamento comum no setor de vigilância, pois o modelo requer muita calibração e testes, mas a qualidade ainda pode ser imprevisível.

Os conjuntos de dados devem ser relevantes para o aplicativo de análise, mas não precisam ser explicitamente identificados ou marcados. O trabalho de registro manual é eliminado, mas o número de imagens ou vídeos necessários para o treinamento deve ter um aumento significativo, em várias ordens de magnitude. Durante a fase de treinamento, o modelo a ser treinado identifica, com o suporte da estrutura de treinamento, atributos comuns nos conjuntos de dados. Isso permite que ele, durante a fase de implantação, agrupe dados de acordo com os padrões, ao mesmo tempo que permite detectar anomalias que não se encaixam em nenhum dos grupos aprendidos.

4.2.3 Aprendizagem por reforço

A aprendizagem por reforço é usada, por exemplo, em robótica, automação industrial e planejamento de estratégia de negócios, mas devido à necessidade de grandes quantidades de feedback, atualmente, o método tem uso limitado na vigilância. Essa aprendizagem reforçada consiste em tomar as medidas adequadas para maximizar o potencial de *recompensa* em uma situação específica; uma recompensa que aumenta quando o modelo faz as escolhas certas. O algoritmo não usa pares de dados/rótulos para o treinamento, mas é ao invés disso otimizado ao testar suas decisões, por meio da interação com o

ambiente, enquanto avalia a recompensa. O objetivo do algoritmo é aprender uma política para realizar ações que ajudem a maximizar a recompensa.

4.3 Teste

Depois que o modelo é treinado, ele precisa ser amplamente testado. Esta etapa normalmente contém uma parte automatizada, complementada com testes extensivos em situações de implantação reais.

Na parte automatizada, o aplicativo é comparado com novos conjuntos de dados, não analisados pelo modelo durante seu treinamento. Se essas referências não estiverem adequadas, o processo será reiniciado: serão coletados novos dados de treinamento, além de realizar ou refinar os registros e treinar novamente o modelo.

Depois de obter o nível de qualidade desejado, um teste de campo será iniciado. Nesse teste, o aplicativo é exposto a cenários do mundo real. O valor e a variação dependerão do escopo do aplicativo. Quanto mais restrito o escopo, menos variações precisarão ser testadas. Quanto mais amplo for o escopo, mais testes serão necessários.

Os resultados são novamente comparados e avaliados. Esta etapa pode fazer novamente com que o processo seja reiniciado. Outra possibilidade seria definir pré-condições, explicando um cenário conhecido no qual o aplicativo não deve ser usado, ou ser usado apenas parcialmente.

4.4 Implantação

A fase de implantação também é chamada de fase de inferência ou previsão. *Inferência* ou *previsão* é o processo de execução de um modelo de aprendizado de máquina treinado. O algoritmo usa o que aprendeu durante a fase de treinamento para produzir a saída desejada. No contexto da análise de vigilância, a fase de inferência é a execução do aplicativo em um sistema de vigilância que monitora cenas da vida real.

Para obter um desempenho em tempo real ao executar um algoritmo baseado em aprendizado de máquina em dados de entrada de áudio ou vídeo, geralmente, é necessária uma aceleração de hardware específica.

5 Análise baseada na borda

A análise de vídeo de alto desempenho costumava ser baseada em servidor, pois exigia mais energia e resfriamento do que uma câmera poderia oferecer. Entretanto, o desenvolvimento de algoritmos e o aumento da capacidade de processamento de dispositivos de borda nos últimos anos possibilitaram análises de vídeo baseadas em IA avançadas na borda.

Há nítidas vantagens nos aplicativos de análise baseada em borda: eles têm acesso a um material de vídeo não compactado com latência muito baixa, permitindo aplicativos em tempo real e evitando o custo adicional e a complexidade de mover os dados para a nuvem para realizar cálculos. A análise baseada em borda também tem custos mais baixos de hardware e implantação, já que são necessários menos recursos de servidor no sistema de vigilância.

Alguns aplicativos podem se beneficiar do uso de uma combinação de processamento baseado em borda e em servidor, com pré-processamento na câmera e um processamento posterior no servidor. Esse sistema híbrido pode facilitar o dimensionamento econômico de aplicativos de análise, operando em várias transmissões de câmera.

6 Aceleração de hardware

Embora você possa executar um determinado aplicativo de análise em vários tipos de plataformas, quando a energia é limitada, o uso da aceleração de hardware dedicada obtém um desempenho muito maior. Os aceleradores de hardware permitem a implementação de aplicativos de análise com uso eficiente de energia. Eles podem ser complementados por recursos de servidor e computação em nuvem, quando adequado.

- **GPU (unidade de processamento gráfico).** As GPUs foram desenvolvidas principalmente para aplicativos de processamento gráfico, mas também são usadas para acelerar a IA em servidores e plataformas na nuvem. Apesar de, às vezes, serem usadas em sistemas incorporados (borda), as GPUs não são ideais, do ponto de vista da eficiência energética, para tarefas de inferência de aprendizado de máquina.
- **MLPU (unidade de processamento de aprendizado de máquina).** Uma MLPU pode acelerar a inferência de algoritmos específicos do aprendizado de máquina clássico para resolver tarefas de visão computacional com eficiência energética muito alta. Ela é projetada para detecção de objetos em tempo real de um número limitado de tipos de objetos simultâneos, por exemplo, humanos e veículos.
- **DLPU (unidade de processamento de aprendizado profundo).** Câmeras com uma DLPU integrada podem acelerar a inferência geral de algoritmos de aprendizado profundo com alta eficiência energética, permitindo uma classificação de objetos mais granular.

7 A IA ainda está em seu desenvolvimento inicial

É tentador fazer uma comparação entre o potencial de uma solução de IA e o que um ser humano pode alcançar. Enquanto os operadores humanos de videomonitoramento só conseguem estar totalmente alertas por um curto período de tempo, um computador pode continuar processando grandes quantidades de dados com extrema rapidez, sem nunca se cansar. No entanto, seria um equívoco elementar presumir que as soluções de IA poderiam substituir o operador humano. A verdadeira vantagem consiste em uma combinação realista: aproveitar as soluções de IA para melhorar e aumentar a eficiência de um operador humano.

As soluções de aprendizado de máquina ou de aprendizado profundo costumam ser descritas como capazes de aprender ou melhorar automaticamente através da experiência. Mas os sistemas de IA disponíveis hoje *não* aprendem automaticamente novas habilidades após a implantação e *não* se lembrarão de eventos específicos que ocorreram. Para aperfeiçoar o desempenho do sistema, ele precisa ser treinado novamente com dados melhores e mais precisos durante sessões de aprendizagem supervisionada. A aprendizagem não supervisionada normalmente requer muitos dados para gerar agrupamentos e, portanto, não é usado em aplicativos de videomonitoramento. Em vez disso, hoje, ela é usada principalmente para analisar grandes conjuntos de dados, a fim de encontrar anomalias, por exemplo, em transações financeiras. A maioria das abordagens desenvolvidas como "autoaprendizagem" no videomonitoramento é baseada em uma análise de dados estatísticos, e não em realmente retreinar os modelos de aprendizado profundo.

A experiência humana ainda supera muitos aplicativos de análise baseada em IA para fins de vigilância. Principalmente aqueles que devem realizar tarefas muito gerais e onde a compreensão contextual é crítica. Um aplicativo baseado em aprendizado de máquina pode detectar com sucesso uma "pessoa correndo" se for especificamente treinado para isso, mas, ao contrário de um ser humano que pode colocar os dados em contexto, o aplicativo não entende por que a pessoa está correndo – para pegar o ônibus ou fugir do policial correndo nas proximidades? Apesar das promessas das empresas que implantam a IA em seus aplicativos de análise para vigilância, o aplicativo ainda não consegue entender o que vê no vídeo com a mesma percepção que um ser humano.

Pelo mesmo motivo, os aplicativos de análise baseada em IA também podem disparar alarmes falsos ou não acioná-los. Isso normalmente pode acontecer em um ambiente complexo com muito movimento. Mas

também pode ocorrer quando, por exemplo, uma pessoa estiver carregando um objeto grande — obstruindo efetivamente as características humanas para o aplicativo, tornando menos provável uma classificação correta.

A análise baseada em IA hoje deve ser usada de forma auxiliar, por exemplo, para determinar a relevância de um incidente antes de alertar um operador humano para que ele decida sobre a resposta. Dessa forma, a IA é usada para alcançar escalabilidade e o operador humano está presente para avaliar possíveis incidentes.

8 Considerações para o desempenho analítico ideal

Para utilizar a qualidade esperada de um aplicativo de análise baseado em IA, é recomendável estudar com atenção e compreender as pré-condições e limitações conhecidas, normalmente listadas na documentação do aplicativo.

Cada instalação de vigilância é única e o desempenho do aplicativo deve ser avaliado conforme o local. Se a qualidade não estiver no nível esperado ou antecipado, é altamente recomendável não focar a investigação apenas no aplicativo. Todas as investigações devem ser feitas em um nível global, pois o desempenho de um aplicativo de análise depende de muitos fatores, muitos dos quais podem ser otimizados se estivermos cientes do seu impacto. Esses fatores incluem, por exemplo, hardware da câmera, qualidade do vídeo, dinâmica da cena, nível de iluminação, bem como configuração, posição e direção da câmera.

8.1 Aproveitamento de imagem

Frequentemente, a qualidade da imagem depende da alta resolução e da alta sensibilidade da câmera à luz. Embora a importância desses fatores não possa ser questionada, certamente existem outros que são igualmente influentes para a atual ou real *usabilidade* de uma imagem ou um vídeo. Por exemplo, a transmissão de vídeo com a melhor qualidade, da câmera de vigilância mais cara, pode se tornar inútil se a cena não tiver iluminação suficiente à noite, se a câmera tiver sido redirecionada ou se a conexão com o sistema for interrompida.

O posicionamento da câmera deve ser meticulosamente considerado antes da instalação. Para que a análise de vídeo funcione conforme o esperado, a câmera precisa estar posicionada para permitir uma visão clara, sem obstáculos, da cena desejada.

A usabilidade da imagem também pode depender do caso de uso. O vídeo que parece bom ao olho humano pode não ter a qualidade ideal para o desempenho de um aplicativo de análise de vídeo. Na verdade, muitos métodos de processamento de imagem que normalmente são usados para melhorar a aparência do vídeo para visualização humana não são recomendados ao usar análise de vídeo. Isso pode incluir, por exemplo, métodos de redução de ruído aplicados, métodos de amplo alcance dinâmico ou algoritmos de exposição automática.

Hoje, as câmeras de vídeo geralmente têm iluminação infravermelha integrada, o que permite o funcionamento na escuridão total. Isso é positivo, pois permite que as câmeras sejam colocadas em locais de difícil iluminação e reduz a necessidade de instalação de iluminação adicional. No entanto, se houver previsão de chuva forte ou neve em uma instalação, é altamente recomendável não depender da luz proveniente da câmera ou de um local muito próximo à câmera. Muita luz pode ser refletida diretamente de volta para a câmera, em contraste com gotas de chuva e flocos de neve, impossibilitando a análise. Por outro lado, com a luz ambiente, há uma maior chance de que a análise apresente alguns resultados, mesmo em climas desfavoráveis.

8.2 Distância de detecção

É difícil determinar a distância máxima de detecção de um aplicativo de análise baseada em IA — um valor exato de folha de dados em metros ou pés nunca pode ser a verdade absoluta. A qualidade da imagem, as características da cena, as condições climáticas e as propriedades do objeto, como cor e brilho, têm um impacto significativo na distância de detecção. É evidente, por exemplo, que um objeto claro contra um fundo escuro durante um dia ensolarado pode ser detectado visualmente em distâncias muito maiores do que um objeto escuro em um dia chuvoso.

A distância de detecção também depende da velocidade dos objetos a serem detectados. Para obter resultados precisos, um aplicativo de análise de vídeo precisa "ver" o objeto durante um período de tempo bastante extenso. A duração desse período depende do desempenho de processamento (taxa de quadros) da plataforma: quanto menor o desempenho de processamento, mais tempo o objeto precisará estar visível para ser detectado. Se o tempo do obturador da câmera não for bem alinhado à velocidade do objeto, o desfoque por movimento exibido na imagem também poderá diminuir a precisão da detecção.

Objetos rápidos podem se passar despercebidos mais facilmente se estiverem passando perto da câmera. Por exemplo, uma pessoa correndo localizada longe da câmera pode ser bem detectada, enquanto uma pessoa correndo muito perto da câmera na mesma velocidade pode estar dentro e fora do campo de visão tão rapidamente que nenhum alarme será acionado.

Na análise baseada em detecção de movimento, os objetos que se movem diretamente em direção à câmera ou para longe dela representam outro desafio. A detecção será particularmente difícil para objetos que se movem lentamente, causando apenas mudanças muito pequenas na imagem em comparação ao movimento na cena.

Uma câmera de resolução mais alta normalmente não fornece uma maior distância de detecção. Os recursos de processamento necessários para executar um algoritmo de aprendizado de máquina são proporcionais ao tamanho dos dados de entrada. Isso significa que a capacidade de processamento necessária para analisar a resolução total de uma câmera 4K é pelo menos quatro vezes maior do que para uma câmera 1080p. É muito comum executar aplicativos baseados em IA em uma resolução mais baixa do que a câmera ou o fluxo pode oferecer, devido às limitações na capacidade de processamento da câmera.

8.3 Configuração de alarmes e gravação

A análise de objetos gera pouquíssimos alarmes falsos devido aos diversos níveis de filtro que aplica. No entanto, a análise de objetos só funciona como deveria quando todas as pré-condições listadas são atendidas. Em outros casos, eles podem perder eventos importantes.

Se não for absolutamente certo que todas as condições serão atendidas o tempo todo, é recomendável adotar uma abordagem conservadora e configurar o sistema de modo que uma classificação de objeto específica não seja o único acionador de alarme. Isso causará mais alarmes falsos, mas também reduzirá o risco de perder algo importante. Quando os alarmes ou disparadores vão diretamente para um centro de monitoramento de alarmes, cada alarme falso se torna muito caro. Há uma necessidade óbvia de uma classificação de objetos confiável para filtrar alarmes indesejados. Contudo, a solução de gravação ainda pode e deve ser configurada para não depender apenas da classificação do objeto. No caso de um alarme real não acionado, essa configuração permite avaliar, a partir da gravação, o motivo da perda do alarme e, em seguida, melhorar a instalação e a configuração geral.

Se a classificação do objeto for feita no servidor durante uma busca de incidentes, é recomendável configurar o sistema para um registro contínuo e não filtrar o registro inicial de forma alguma. A gravação contínua consome muito armazenamento, mas isso é, até certo ponto, compensado por algoritmos de compactação modernos, como o Zipstream.

8.4 Manutenção

Uma instalação de vigilância deve ser mantida regularmente. As inspeções físicas, e não apenas a visualização do vídeo por meio da interface de VMS, são recomendadas para descobrir e remover quaisquer elementos que possam atrapalhar ou bloquear o campo de visão. Isso é importante também em instalações padrão, apenas de gravação, mas é ainda mais crítico ao usar análises.

No contexto da detecção básica de movimento por vídeo, um obstáculo típico, como uma teia de aranha que balança com o vento, pode aumentar o número de alarmes, resultando em um consumo de armazenamento maior do que o necessário. Com a análise do objeto, a teia basicamente criaria uma zona de exclusão na área de detecção. Seus fios ocultariam objetos e reduziriam muito a chance de detecção e classificação.



As teias de aranha podem atrapalhar o campo de visão de uma câmera de vigilância.

É improvável que uma sujeira no vidro frontal ou o dome da câmera cause problemas durante o dia. Mas em condições de pouca iluminação, a luz que atinge uma bolha suja pela lateral, como por exemplo dos faróis de um carro, pode causar reflexos inesperados que são capazes de diminuir a precisão da detecção.

A manutenção da cena é tão importante quanto a manutenção da câmera. Durante a vida útil de uma câmera, muitas situações podem acontecer na cena que ela está monitorando. Uma comparação simples de imagens de antes e depois revelará os possíveis problemas. Como era a cena quando a câmera foi instalada e como ela está hoje? É necessário ajustar a área de detecção? O campo de visão da câmera deve ser ajustado ou a câmera deve ser movida para um local diferente?

9 Privacidade e integridade pessoal

Trabalhar com segurança e vigilância requer equilibrar os direitos individuais à privacidade e à integridade pessoal com a ambição de aumentar a segurança, prevenindo crimes ou permitindo investigações forenses. No caso de instalação e uso específico, isso requer consideração ética cuidadosa, bem como compreensão e aplicação da legislação local. Também impõe exigências à solução para, por exemplo, garantir a segurança cibernética e impedir o acesso não intencional ao material de vídeo. Ao mesmo tempo, a análise baseada em borda e a geração de metadados para fins estatísticos podem aumentar a proteção da privacidade se somente dados anônimos forem transmitidos para um processamento posterior.

Com a crescente implantação de análises automatizadas em sistemas de vigilância, alguns fatores novos devem ser considerados. Como os aplicativos de análise apresentam o risco de falsas detecções, é

importante que o processo de decisão envolva um operador ou usuário experiente. Isso é frequentemente mencionado como manter um "ser humano informado". Além disso, é importante reconhecer que a decisão humana pode ser afetada pela forma como o alarme é gerado e apresentado. Sem o treinamento adequado e o conhecimento da funcionalidade da solução de análise, podem ser tiradas conclusões erradas.

Uma outra preocupação pode ser a maneira como os algoritmos de aprendizado profundo são desenvolvidos e, para alguns casos de uso, isso requer uma abordagem cautelosa ao implantar a tecnologia. A qualidade desses algoritmos está fundamentalmente ligada aos conjuntos de dados, ou seja, aos vídeos e imagens usados para treinar o algoritmo. Testes demonstraram que, se esse material não for cuidadosamente selecionado, alguns sistemas de IA podem apresentar preconceitos étnicos e de gênero nas detecções. Isso gerou uma discussão aberta e deu origem a restrições legislativas e de atividades para garantir que tais aspectos sejam tratados durante o desenvolvimento dos sistemas.

Como a IA tem sido implantada cada vez mais na vigilância, é importante equilibrar as vantagens da eficiência operacional e os novos casos de uso em potencial com uma discussão cuidadosa sobre quando e onde implantar essa tecnologia.

10 Apêndice

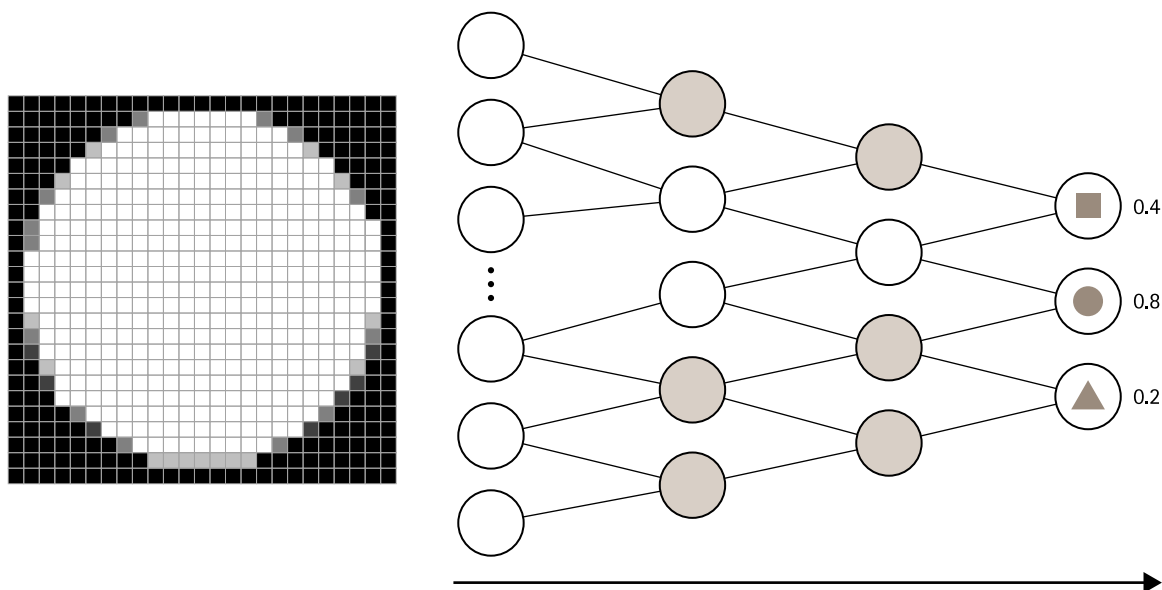
Este apêndice fornece informações básicas sobre as redes neurais artificiais que formam a base do aprendizado profundo.

10.1 Redes neurais

As redes neurais são uma família de algoritmos usada para reconhecer as relações em conjuntos de dados por meio de um processo um tanto semelhante ao funcionamento do cérebro humano. Uma rede neural consiste em uma hierarquia de várias camadas, dos chamados nós ou neurônios, que estão interconectadas, e em informações que estão sendo transmitidas ao longo das conexões, da camada de entrada, através da rede, para a camada de saída.

A premissa para que as redes neurais funcionem é a de que uma amostra de dados de entrada possa ser reduzida a um conjunto limitado de atributos, criando uma boa representação dos dados de entrada. Esses atributos podem então ser combinados e ajudarão a classificar os dados de entrada, por exemplo, descrevendo o conteúdo de uma imagem.

A ilustração abaixo apresenta um exemplo em que uma rede neural é usada para identificar a categoria da imagem de entrada. Cada pixel da imagem é representado por um nó de entrada. Todos os nós de entrada são associados aos nós da primeira camada. Eles produzem valores de saída que são passados como valores de entrada para a segunda camada, e assim por diante. Em cada camada, também estão envolvidos no processo as funções de ponderação, os valores de propensão e as funções de ativação.



Exemplo de uma imagem de entrada (à esquerda) e uma rede neural (à direita). Quando a camada de saída for alcançada, a rede terá concluído as probabilidades para cada categoria possível (quadrado, círculo ou triângulo). A categoria com o valor de probabilidade mais alto será o formato mais provável da imagem de entrada.

Este processo é denominado *forward propagation*. Em caso de incompatibilidade do resultado da propagação direta, os parâmetros da rede serão ligeiramente modificados por meio da *propagação reversa*. Durante esse processo de treinamento iterativo, o desempenho da rede é progressivamente aprimorado.

Após a implantação, uma rede neural, em geral, não terá memória de propagações diretas anteriores. Isso significa que ela não melhora com o tempo e só pode detectar os tipos de objetos ou resolver os tipos de tarefas para as quais foi treinada.

10.2 Redes neurais convolucionais (CNN, Convolutional neural networks)

As *Redes neurais convolucionais* (CNN) são um subtipo de redes neurais artificiais que provaram ser especialmente adequadas para tarefas de visão computacional e estão no centro do rápido progresso do aprendizado profundo. No caso da visão computacional, a rede é treinada para procurar automaticamente características distintas da imagem, como bordas, cantos e diferenças de cores, identificando efetivamente as formas dos objetos em uma imagem.

A principal operação para fazer isso é a operação matemática chamada *convolução*. Ela é uma operação muito eficiente, pois a saída de cada nó individual depende apenas de um ambiente limitado nos dados de entrada, que foram produzidos pela camada anterior, em vez de usar todo o volume de dados de entrada. Em outras palavras, em uma CNN, o nó não está conectado a todos os nós da camada anterior, mas apenas a um pequeno subconjunto. As convoluções são complementadas por outras operações que reduzem o tamanho dos dados enquanto retêm as informações mais úteis. Como em uma rede neural artificial padrão, os dados se tornam cada vez mais abstratos à medida que se aprofundam na rede.

Durante a fase de treinamento, a CNN aprende a melhor forma de aplicar as camadas. Ou seja, como as convoluções devem combinar os atributos da camada anterior para que a saída da rede corresponda o máximo possível com os registros dos dados de treinamento. Durante a inferência, a rede neural convolucional treinada aplica sequencialmente as camadas de convoluções resultantes do treinamento.

Sobre a Axis Communications

A Axis torna possível um mundo mais inteligente e seguro criando soluções para melhorar a segurança e o desempenho dos negócios. Como empresa de tecnologia de rede e líder do setor, a Axis oferece soluções em vigilância por vídeo, controle de acesso, intercomunicação e áudio. Nossas soluções são aprimoradas por aplicativos de análise inteligentes e apoiados por treinamento de alta qualidade.

A Axis tem cerca de 4.000 funcionários dedicados em mais de 50 países e colabora com parceiros de tecnologia e integração de sistemas em todo o mundo para fornecer soluções aos clientes. A Axis foi fundada em 1984 e tem sede em Lund, Suécia