ホワイトペーパー

AIによるビデオ分析

機械学習（マシンラーニング）と深層学習（ディープラーニング）に基づく分析に関する考慮事項

3月 2021

目次

1 まとめ

AIベースのビデオ分析は、映像監視業界で最大級の話題となっています。一部のアプリケーションを活用することで、データ分析を著しく高速化し、反復的なタスクを自動化することができます。しかし、今日のAIソリューションで、人間のオペレーターの経験と意思決定スキルを置き換えることはできません。AIと人間の能力を組み合わせることで強みが生まれます。つまり、AIソリューションを活用することで、人間の効率を改善および向上できるということです。

AIの概念には、機械学習アルゴリズムと深層学習アルゴリズムが含まれます。双方とも、大量のサンプルデータ（*トレーニングデータ*）が使用されて数学モデルが自動的に構築され、特別にプログラムしなくても結果が計算されます。AIアルゴリズムは、反復プロセスを通じて開発されます。つまり、トレーニングデータの収集、トレーニングデータのラベル付け、ラベル付けされたデータを使用したアルゴリズムのトレーニング、トレーニングされたアルゴリズムのテストというサイクルが、目的の品質レベルに達するまで繰り返されるのです。目的レベルに達したアルゴリズムは、監視サイトで展開できる市販の分析アプリケーションで使用できるようになります。この時点で、すべてのトレーニングが完了しているため、アプリケーションがこれ以上新たな事柄を学習するということはありません。

AIベースのビデオ分析の一般的な役割は、ビデオストリーム内の人間と車両を視覚的に検知し、どちらが人間でどちらが車両かを区別することです。*機械学習*アルゴリズムには、こうした対象を定義する視覚的特徴の組み合わせが学習により組み込まれています。*深層学習*アルゴリズムはより精巧で、トレーニング次第では、はるかに複雑な物体を検知することができます。しかし、これには開発とトレーニングにかなりの労力を要し、完成したアプリケーションを使用する際により高い演算リソースが必要となります。そのため、監視ニーズを明確に指定できる場合は、最適化された専用機械学習アプリケーションで十分かどうかを検討する必要があります。

アルゴリズムの開発とカメラの処理能力の向上により、サーバーで計算を実行する（サーバーベース）のではなく、高度なAIベースのビデオ分析をカメラで直接実行（エッジベース）できるようになりました。これにより、アプリケーションから非圧縮のビデオ素材に直ちにアクセスできるようになるため、より優れたリアルタイム機能が実現します。MLPU（機械学習処理ユニット）やDLPU（深層学習処理ユニット）などの専用ハードウェアアクセラレーターにより、CPUやGPU（グラフィックス処理ユニット）を利用するよりもエッジベース分析をより電力効率よく実装することが可能となります。

AIベースのビデオ分析アプリケーションをインストールする前に、既知の前提条件や制限に基づくメーカーの推奨事項を注意深く検討し、これに従う必要があります。すべての監視設備は独特であるため、アプリケーションのパフォーマンスを各サイトで評価する必要があります。期待通りの品質が得られない場合は、分析アプリケーション自体だけに焦点を当てるのではなく、全体的なレベルで調査を行う必要があります。ビデオ分析のパフォーマンスは、カメラハードウェア、カメラ構成、ビデオ品質、シーンダイナミクス、照明に関連する多くの要因に依存します。多くの場合、こうした要因による影響を把握し、それに応じて最適化を図ることで、設置場所におけるビデオ分析パフォーマンスを向上させることができます。

監視にますますAIが導入されている現状を踏まえ、技術を適用する場所と時期を注意深く検討し、バランスを取って運用効率と新たなユースケースの利点を活用する必要があります。

2 はじめに

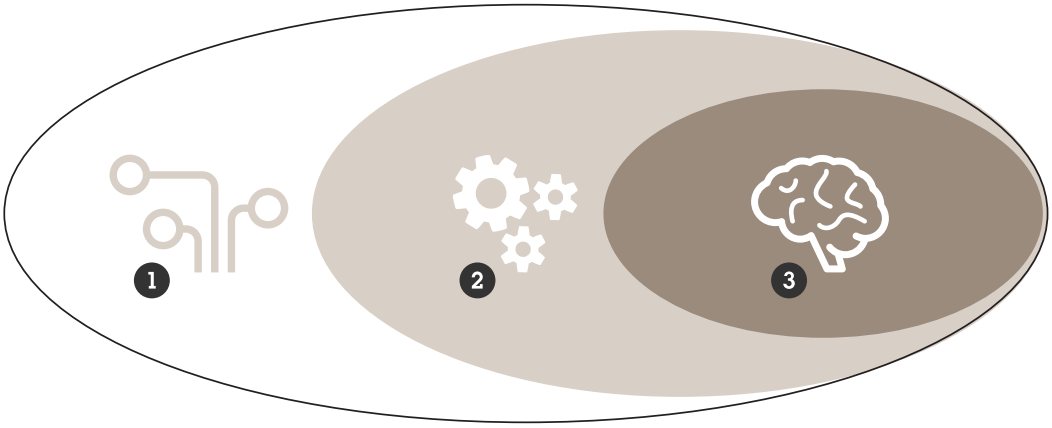
人類初のコンピューターが発明されて以来、人間はAI（人工知能）の開発に取り組み、これについて議論してきました。人類最大の革新技術が達成されたとは言えないまでも、今日、音声認識、検索エンジン、仮想アシスタントなどのアプリケーションで、明確に定義されたタスクを実行するためにAIベースの技術が広く利用されています。X線診断や網膜スキャン分析など、貴重なリソースを提供するヘルスケア分野でもAIがますます導入されるようになりました。

AIベースのビデオ分析は、映像監視業界で最大級の話題となっており、これに対する期待も高まっています。AIアルゴリズムを使用してデータ分析を正常に高速化し、反復タスクを自動化するアプリケーションが市場に出回っています。しかし、監視環境をより広範に捉えた場合、現在だけでなく近い将来も含め、AIは正確なソリューションを構築するプロセスにおける単なる一要素に過ぎません。

本ホワイトペーパーでは、機械学習と深層学習のアルゴリズムに関する技術的背景、およびこれを開発してビデオ分析に適用する方法について解説します。これには、AIアクセラレーターハードウェアの簡単な説明、およびサーバーベースではなくエッジベースでAIベースの分析を実行することの長所と短所に関する説明が含まれます。また、さまざまな要因を考慮に入れて、AIベースのビデオ分析パフォーマンスの前提条件を最適化する方法についてもご説明します。

3 AI、機械学習、深層学習

人工知能（AI）とは、一見インテリジェントな特性を示しながら複雑なタスクを解決できるマシンに関連する幅広い概念で、深層学習と機械学習はAIのサブセットとなります。



1 AI
2 機械学習
3 深層学習

3.1 機械学習

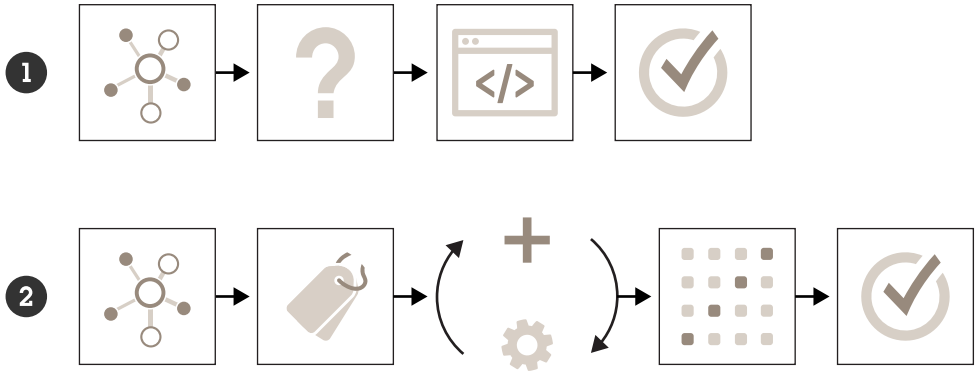
機械学習はAIのサブセットです。これは、統計的学習アルゴリズムを使用して、明示的にプログラムすることなく、トレーニングにより自動的に学習して改善する機能を備えたシステムを構築する手法です。

このセクションでは、画像や動画を分析することで、シーンで発生している現象をコンピューターに理解させる分野「*コンピュータービジョン*」という観点において、従来型のプログラミングと機械学習を区別して考察します。

従来型のプログラミングによるコンピュータービジョンは、画像の*特徴*を計算する方法に基づいています。たとえば、明確な縁や端点を探させるといったコンピュータープログラミングです。こうした機能の場合は、画像データで重要な点を理解しているアルゴリズム開発者が手動でこれを定義する必要があります。次に、開発者はこうした機能を組み合わせて、シーンで検知された要素を結論付けるアルゴリズムを構築します。

機械学習アルゴリズムの場合は、大量のサンプルデータ（*トレーニングデータ*）を使用して数学モデルを自動的に構築し、特別にプログラムしなくても、結果を計算して決定を下すことができます。それでも機能は手動で構築する必要がありますが、大量のラベル付きのトレーニングデータまたは*注釈付き*のトレーニングデータを供給することで、アルゴリズム自体にこれらの機能を組み合わせる方法を学習させます。本ホワイトペーパーでは、手動で加えられた機能を、学習した組み合わせで使用するこの手法を、*古典的機械学習*と呼びます。

つまり、機械学習アプリケーションの場合、必要なプログラムが実行されるように、コンピューターをトレーニングする必要があるということです。収集されたデータに人間が注釈を付けます。サーバーコンピューターにより、補助的に事前の注釈付けが行われる場合もあります。そして、結果がシステムに送られます。アプリケーションで十分な学習が行われ、特定タイプの車両など、必要な対象物が検知されるようになるまでこのプロセスが継続されます。トレーニングを受けたモデルがプログラムとなります。プログラムが終了すると、システムでこれ以上新たな学習が行われることはありません。



1 *従来型のプログラミング：*
データの収集 -プログラム基準の定義 -（人間による）プログラムのコード化 -終了
2 *機械学習：*
データの収集 -データのラベル付け -モデルの反復トレーニングプロセス -完成したトレーニング済みモデルがプログラムとして誕生 -終了

コンピュータービジョンプログラムの構築において、従来型のプログラミングよりも広範なデータを処理できるという点がAIのメリットとなります。数千に上る画像を処理しなければならない場合、人間のプログラマーはしばらくすると疲れが出て注意散漫になり得ますが、コンピューターなら集中を失うことなく一貫して作業を継続することができます。そのため、AIにより、著しく正確なアプリケーションを構築することができます。しかし、アプリケーションが複雑になるほど、マシンだけで目的の結果を達成させるのが難しくなります。

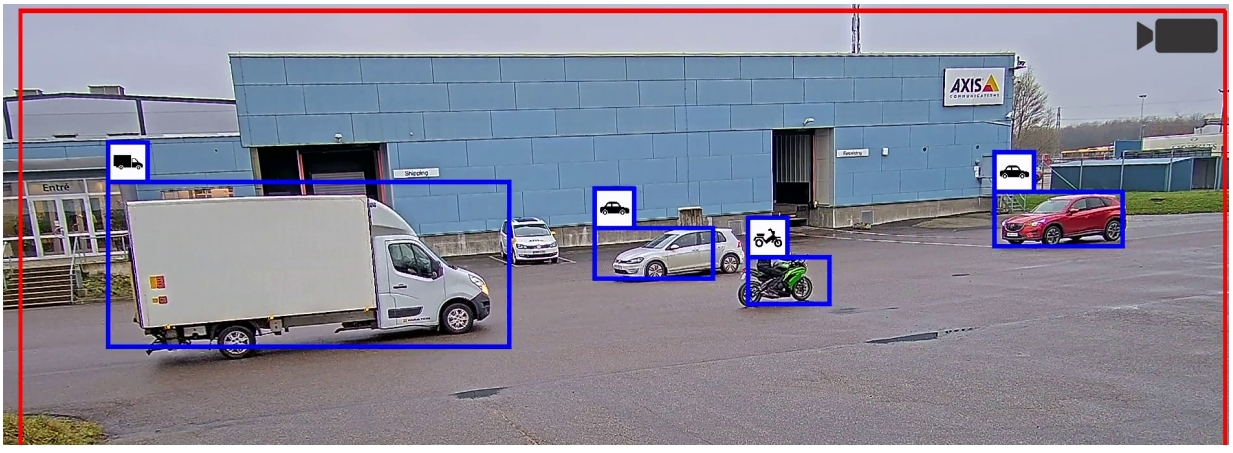
3.2 深層学習

深層学習と機械学習は両方とも、機能を抽出し、これらの機能を組み合わせ、そして深い規則構造に従って結果を生成する方法をデータ駆動型の方法で学習させる手法ですが、深層学習は機械学習よりも精巧です。アルゴリズムにより、トレーニングデータで検索する機能が自動的に定義されるだけでなく、機能の連鎖した組み合わせを非常に深い構造で学習することが可能です。

深層学習で使用されるアルゴリズムのコアは、ニューロンが機能する仕組み、つまり脳内における鎖律の深階層（*ネットワーク*）でニューロンの出力が組み合わせられてより高度なレベルの知識が形成される方法からヒントを得たものです。脳はニューロンの組み合わせにより構成されるシステムで、機能抽出と機能の組み合わせの区別がなく、ある意味これらが同化されています。研究者等がこの構造を模し、深層学習で最も広く使用されているタイプのアルゴリズム「*人工ニューラルネットワーク*」と呼ばれるコンピューティングシステムを開発しました。ニューラルネットワークの概要については、本ホワイトペーパーの付録を参照してください。

深層学習アルゴリズムを使用することで、複雑な視覚検知器を構築し、これを自動的にトレーニングすることで、縮拡張や回転といった変化に柔軟に対応でき、非常に複雑な物体を検知できるシステムを実現することができます。

こうした柔軟性を実現できるのは、深層学習システムでは、古典的機械学習システムよりもはるかに大量かつ多様なデータから学習することが可能であるためです。ほとんどの場合、これは手動で構築されたコンピュータービジョンアルゴリズムよりも著しく優れています。そのため、深層学習は、画像分類、言語処理、物体検知など、人間の専門家が機能の組み合わせを簡単に形成できない複雑な問題に特に適しています。



深層学習に基づく物体検知により、複雑な物体を分類することができます。たとえば、分析アプリケーションで車両を検知できるだけでなく、車両のタイプも分類することが可能となります。

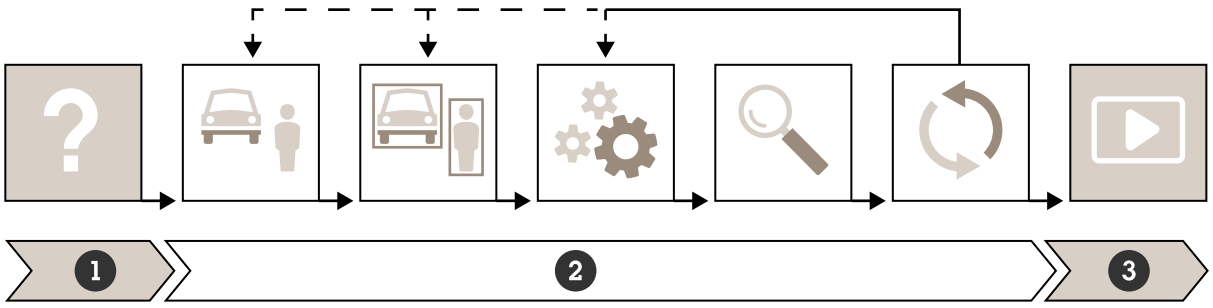
3.3 古典的機械学習と深層学習

両方のアルゴリズムのタイプは類似していますが、深層学習アルゴリズムでは、通常、古典的機械学習アルゴリズムよりもはるかに多くの学習済み機能の組み合わせが使用されます。つまり、深層学習ベースの分析のほうが柔軟性が高く、トレーニング次第では、はるかに複雑なタスクを実行する方法を学習できるということです。

しかし、特定の監視分析には、最適化された専用の古典的機械学習アルゴリズムで十分な場合があります。範囲を明確に指定できる場合は、古典的機械学習アルゴリズムで深層学習アルゴリズムと同様の結果を得ることができます。機械学習のほうが数学的な操作が少なくて済むため、コスト効率が高く、使用電力も削減することが可能となります。また、必要なトレーニングデータもはるかに少なくて済むことから、開発の労力も大幅に軽減されます。

4 機械学習の開発手順

機械学習アルゴリズムの開発には、最終的な分析アプリケーションを展開するまでに、一連の手順と反復が関与します。大まかに視覚化された以下の図をご覧ください。分析アプリケーションの中核には、1つ以上のアルゴリズムが存在します。たとえば、物体検知器などです。深層学習ベースのアプリケーションの場合、アルゴリズムの中核は深層学習モデルとなります。



1 *準備：アプリケーションの目的を定義します。*
2 *トレーニング：トレーニングデータを収集し、データに注釈を付け、モデルをトレーニングし、そしてモデルをテストします。期待通りの品質が達成されなかった場合は、反復的な改善サイクルで、同じ手順を繰り返します。*
3 *展開：完成したアプリケーションをインストールして、使用します。*

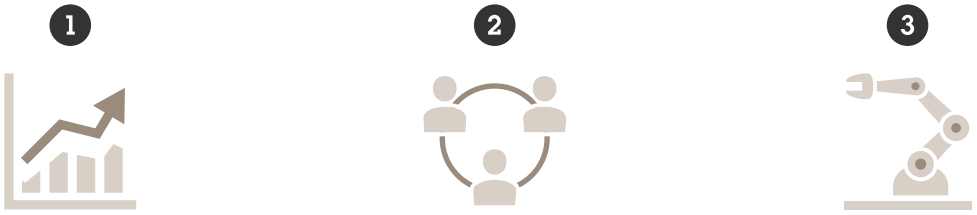
4.1 データの収集とデータの注釈付け

AIベースの分析アプリケーションを開発するには、大量のデータを収集する必要があります。映像監視の場合、通常、これには人間や車両、または関連性の高い他の物体の画像やビデオクリップが含まれます。マシンやコンピューターでデータが認識されるようにするには、関連する物体を分類してラベルを付けるデータ注釈プロセスが必要となります。データ注釈は主に手動で行われるため、手間のかかる作業となります。分析アプリケーションが使用される状況に関連付けられるサンプルを十分にカバーできるだけのデータの種類を準備する必要があります。

4.2 トレーニング

モデルに注釈付きデータを供給し、トレーニングフレームワークを使用して、目的の品質に達するまでモデルを繰り返し変更および改善するプロセスは、トレーニングまたは学習

と呼ばれます。つまり、定義されたタスクを解決できるようにモデルを最適化する工程です。トレーニングは、以下3つの主要な方法のいずれかに従って実施することができます。



1 教師あり学習：モデルに学習させ、正確な予測を行うことができるようにします。
2 教師なし学習：モデルに学習させ、クラスターを識別できるようにします。
3 強化学習：モデルに間違いから教訓を学ばせます。

4.2.1 教師あり学習

今日の機械学習で最も一般的に使用されている方法が教師あり学習です。これは「実例から学習する方法」と言えます。トレーニングデータには明確な注釈が付けられています。つまり、入力データはすでに目的の出力結果と対になっているということです。

通常、教師あり学習には、非常に大量の注釈付きデータが必要となり、トレーニング済みアルゴリズムのパフォーマンスはトレーニングデータの品質に直接的に依存します。最も重要な品質の側面として、実際の展開状況からのすべての潜在的な入力データを表すデータセットを使用することが挙げられます。物体検知器の場合、さまざまな物体の例、向き、スケール、照明の状況、背景、障害要素など、開発者はさまざまな画像を使用してアルゴリズムをトレーニングする必要があります。トレーニングデータが予定のユースケースに適切に一致している場合にのみ、最終的な分析アプリケーションで、トレーニング段階では表れていなかった新規データを処理する際に正確な予測を行うことができるようになります。

4.2.2 教師なし学習

教師なし学習では、アルゴリズムを使用して、ラベルの付いていないデータセットを分析およびグループ化します。この学習方法では、モデルに多くの較正とテストが必要になるにも関わらず、品質が依然として予測できない可能性があるため、監視業界では一般的には使用されていないトレーニング方法です。

データセットは分析アプリケーションに関連している必要がありますが、明確にラベル付けまたはマーク付けする必要はありません。手動で注釈を付ける作業は不要となりますが、トレーニングに必要な画像やビデオの数が数桁増えます。トレーニング段階では、トレーニングフレームワークでサポートし、トレーニング対象のモデルにデータセット内の一般的な機能を識別させます。これにより、展開段階において、パターンに従ってデータをグループ化すると同時に、学習したグループのいずれにも該当しない異常を検知できるようになります。

4.2.3 強化学習

強化学習はロボット工学、産業オートメーション、ビジネス戦略計画などで使用されますが、大量のフィードバックが必要となるため、今日、これは監視分野ではあまり使用されていません。強化学習とは、正しい選択を行うと得られる報酬が増えるという特定の環境内で、適切な行動を取ることで潜在的な*報酬*が最大化されるということをモデルが学ぶ学習の一種です。アルゴリズムのトレーニングではラベル付きデータは使用しませんが、その代わりに、報酬を測定しながら環境との相互作用を通じてその決定をテス

トすることで最適化を図ります。アルゴリズムにおける目標は、報酬を最大化する行動のポリシーを学習することにあります。

4.3 テスト

モデルのトレーニングが完了したら、これを徹底的にテストする必要があります。通常、この段階には、実際の展開状況での広範なテストで補完される自動の部分が含まれます。

自動の部分では、トレーニングではモデルに供給していない新規データセットを用いて、アプリケーションをベンチマークします。こうしたベンチマークテストで期待通りの成果が達成されなかった場合は、新規トレーニングデータの収集、注釈の作成または改良、モデルの再トレーニングというプロセスを最初から繰り返すことになります。

必要な品質レベルに達した場合は、現場テストを開始します。このテストでは、アプリケーションを実際のシナリオで試します。量と偏差は、アプリケーションの範囲によって異なります。範囲が狭いほど、テストする必要のある偏差が少なくなります。範囲が広いと、より多くのテストを行う必要があります。

結果を再び比較し、評価します。この段階で、プロセスを最初からやり直さなければならないという事態も発生し得ます。アプリケーションの使用が推奨されていない、または部分的にしか推奨されていない既知のシナリオを説明して前提条件を定義するという結果となる場合もあります。

4.4 展開

展開段階は、推論段階または予測段階とも呼ばれます。*推論*または*予測*は、トレーニング済みの機械学習モデルを実行するプロセスです。アルゴリズムでは、トレーニング段階で学習した内容が使用され、目的の出力が生成されます。監視分析の場合、推論段階では、実際のシーンを監視する監視システムでアプリケーションを実行します。

音声やビデオの入力データで機械学習ベースのアルゴリズムを実行する際にリアルタイムのパフォーマンスを実現するには、通常、特定のハードウェアアクセラレーションが必要となります。

5 エッジベースのビデオ分析

高性能ビデオ分析では、カメラが提供できるよりも多くの電力と冷却機能が必要となることから、以前はサーバーベースでした。しかし、近年のアルゴリズム開発とエッジデバイスの処理能力の向上により、高度なAIベースのビデオ分析をエッジで実行できるようになりました。

エッジベースの分析アプリケーションには明らかなメリットがあります。非常に低い遅延で非圧縮のビデオ素材にアクセスできるため、演算のためにデータをクラウドに移動するコストと複雑性を回避しながら、リアルタイムのアプリケーションを実現することができます。エッジベースの分析では、監視システムに必要なサーバーリソースが少なくて済むため、ハードウェアと展開のコストも削減することが可能となります。

エッジベースの処理とサーバーベースの処理を組み合わせた一部のアプリケーションには、カメラで前処理を実行し、サーバーでさらに処理できるというメリットがあります。こうしたハイブリッドシステムを活用して、複数のカメラストリームで作業することで、分析アプリケーションでコスト効率の高いスケーリングを促進することができます。

6 ハードウェアアクセラレーション

多くの場合、特定の分析アプリケーションをいくつかのタイプのプラットフォームで実行することができますが、専用のハードウェアアクセラレーションを使用することで、電力が制限されている場合にはるかに高いパフォーマンスを実現することが可能となります。ハードウェアアクセラレーターにより、分析アプリケーションの実装の電力効率を向上することができます。必要に応じて、サーバーとクラウドの演算リソースで補完することが可能です。

- **GPU（グラフィックス処理ユニット）**。GPUは主にグラフィックス処理アプリケーション用に開発されたものですが、サーバーとクラウドプラットフォームでAIを高速化するためにも使用されます。内蔵システム（エッジ）でも使用されることがありますが、電力効率の観点から、GPUは機械学習の推論タスクには最適とは言えません。
- **MLPU（機械学習処理ユニット）**。MLPUにより、特定の古典的機械学習アルゴリズムの推論を加速し、非常に高い電力効率でコンピュータービジョンタスクを解決することができます。これは、人や車両など、限られた数の物体の種類を同時にリアルタイムで検知するように設計されています。
- **DLPU（深層学習処理ユニット）**。DLPUが組み込まれたカメラでは、一般的な深層学習アルゴリズムの推論が高い電力効率で加速されるため、より詳細な物体分類が可能となります。

7 AIは未だ初期開発段階

今日、AIソリューションの可能性と人の能力が比較されがちになっています。人間の映像監視担当者が監視画面を注視できる時間はかなり短いのに比べて、コンピューターなら継続的に大量データを非常に迅速に処理することができます。しかし、AIソリューションが人間のオペレーターに取って代わるという考え方は根本的に間違っています。双方を組み合わせることで得られる現実的なメリットに真の強みがあるのです。つまり、AIソリューションを活用することで、人間のオペレーターの効率を改善および向上できるということです。

多くの場合、機械学習や深層学習ソリューションは、経験を通じて自動的に学習または改善する機能を備えたソリューションと考えられています。しかし、現在利用されているAIシステムは、展開後に自動的に新しいスキルを学習することも、発生した特定のイベントを記憶することも*できません*。システムのパフォーマンスを向上させるには、「教師あり学習」により、適切で正確なデータを使用してシステムを再トレーニングする必要があります。通常、クラスターを生成するために大量のデータが必要となる「教師なし学習」は、映像監視アプリケーションでは使用されません。今日、これは主に大規模なデータセットを分析して異常を見つけるために使用されています。たとえば、金融取引などです。映像監視システムの「自己学習」として推進されているほとんどのアプローチは、深層学習モデルを実際に再トレーニングする手法ではなく、統計データ分析に基づいています。

監視目的においては、多くのAIベースの分析アプリケーションよりも人間の経験のほうが勝っています。特に、非常に一般的なタスクを実行し、前後状況の理解が重要となるケースでは、人間の能力のほうが優れているのです。機械学習ベースのアプリケーションが特別にそのようにトレーニングされていれば、「走っている人物」を正常に検知できますが、前後状況を踏まえてデータを把握できる人間とは異なり、アプリケーションではその人物が走っている理由を理解することができません。単にバスの時間に遅れそうになって走っているのか、警官に追われて走っているのかが分からないわけです。監視向け分析アプリケーションにAIを導入した製品について過度な宣伝がなされることもありますが、アプリケーションでは、人間と同程度の洞察をもって、リモートで撮影されたビデオの被写体を理解することはまだできないのです。

これと同じ理由により、AIベースの分析アプリケーションでは誤警報がトリガーされたり、アラームが見過ごされたりする可能性があります。これは通常、撮影シーン内の動きが多い複雑な環境で発生する可能性があります。また、大きな物体を運んでいる人物についても同様のことが言えます。この場合、アプリケーションでは人間の特性が効果的に捉えられず、正しい分類がなされる可能性が低下します。

今日のAIベースの分析は、補助的な方法で使用するべきです。たとえば、人間のオペレーターが対応を判断する前に、事態の重要性や関連性を大まかに判断するために使用することができます。このように、現在AIはより広範な目的で使用されるようになり、人間のオペレーターが潜在的な事態を評価するためのツールとしての役割を果たしています。

8 最適な分析パフォーマンスに関する考慮事項

AIベースの分析アプリケーションの品質を良好に把握するには、アプリケーションの資料などに通常記載されている既知の前提条件と制限を注意深く調べて理解することが勧められます。

すべての監視設備は独特であるため、アプリケーションのパフォーマンスを各サイトで評価する必要があります。期待通りの品質水準に達しない場合は、アプリケーションだけに焦点を当てて調査しないでください。分析アプリケーションのパフォーマンスは非常に多くの要因により左右されるため、全体的なレベルですべてを調べる必要があります。影響の要因を認識できれば、そのほとんどを最適化することが可能となります。こうした要因には、カメラハードウェア、ビデオ品質、シーンダイナミクス、照明レベル、カメラの構成・位置・方向などが含まれます。

8.1 画像の有用性

画質は、カメラの高解像度と高光感度に依存するとよく言われます。こうした要素は間違いなく重要ですが、画像やビデオの実際の*有用性*に同様の影響をもたらす要素は他にも存在します。たとえば、夜間に十分な照明がない場合やカメラの向きが変えられた場合、またはシステムの接続に損傷がある場合などは、最高級の監視カメラによる最高品質のビデオストリームでも役に立たなくなる可能性があります。

展開する前に、カメラの配置を慎重に検討する必要があります。ビデオ分析を期待通りに実行するには、障害物がなく、目的のシーンを鮮明に表示できるようにカメラを配置する必要があります。

画像の有用性もユースケースによって異なります。人間の目には良好に見えるビデオ画像でも、ビデオ分析アプリケーションのパフォーマンスという観点からは最適な品質ではない可能性もあります。実際、ビデオの外観を向上して人間の目に良好に映るようにするために一般的に使用されている多くの画像処理方法は、ビデオ分析には向いていません。これには、ノイズ除去法、ワイドダイナミックレンジ法、自動露出アルゴリズムなどが含まれます。

多くの場合、今日のビデオカメラには、完全な暗闇でも撮影が可能となるIR照明が統合されています。これにより、カメラを照明条件の悪い場所にも設置でき、施設の照明装置を増やす必要がなくなるため、これはメリットのある機能です。しかし、降雨や降雪の多いサイトでは、カメラ自体の照明機能やカメラに非常に近い場所に設置された照明器具に頼らないことが強く勧められます。雨滴や雪片に反射する光が大量にカメラに取り込まれると、分析が不可能となる場合があります。悪天候の場合でも、周囲光を使用するほうが、分析でより良好な結果が得られる可能性が高くなります。

8.2 検知距離

AIベースの分析アプリケーションの場合、その最大検知距離を把握することは容易ではありません。メートルやフィート単位でデータシートに示されている数値が完全に正しいとは言えないのです。画質、シーンの特性、気象条件、色や輝度などの物体の特質により、検知距離に大きな影響がもたらされます。たとえば、雨天の中に存在する暗い色の物体よりも、晴天環境で暗い背景に明るい色の物体がある場合のほうが、はるかに遠い距離から視覚的に検知できることは明らかです。

検知距離は、検知する物体の動きの速度によっても変わります。正確な結果を得るには、ビデオ分析アプリケーションで物体が十分な時間「認識」される必要があります。その時間の長さは、プラットフォームの処理パフォーマンス（フレームレート）によって異なります。処理パフォーマンスが低いほど、物体が見えている状態が長く持続される必要があります。カメラのシャッター時間が物体の速度とうまく一致していないと、動きによる画像のブレが発生します。そうなると、検知精度が低下する可能性があります。

速度の速い物体がカメラの近くを通り過ぎた場合、これは見落とされる確率が高くなります。たとえば、カメラから遠く離れた場所で走っている人物は十分に検知されると考えられますが、同じ速度でもカメラに非常に近い場所で走っている人物は、カメラの視野に出入りする速度が速すぎるため、アラームがトリガーされない可能性があります。

動体検知に基づく分析では、一直線にカメラに向かって移動してくる物体やカメラから離れていく物体により別の課題が発生します。動きの遅い物体は、シーン全体の動きに対して画像にごくわずかな変化しか生じないため、特に検知が困難となります。

通常、高解像度カメラの検知距離はそれほど長くありません。機械学習アルゴリズムを実行する上で必要な処理能力は、入力データのサイズに比例します。つまり、4Kカメラのフル解像度を分析するには、1080pカメラの場合よりも少なくとも4倍高い処理能力が必要になるということです。カメラの処理能力に制限があるため、AIベースのアプリケーションを、カメラやストリームが提供できる解像度よりも低い解像度で実行することは非常に一般的です。

8.3 アラームと録画の設定

さまざまに異なるフィルターレベルを適用できるため、物体分析では誤警報がほとんど発生しません。しかし、物体分析は、一覧されている前提条件がすべて満たされている場合にのみ正常に実行されます。他のケースでは、重要なイベントを見逃す可能性があります。

そのため、すべての条件が常に満たされることが確実でない場合は、保守的なアプローチを取り、特定の物体分類のみがアラームトリガーとならないようにシステムを設定することが勧められます。そうすると、誤警報が発生する確率が高まりますが、重要な事態を見逃すリスクは軽減されます。アラームやトリガーがアラーム監視センターに直接送信される場合は、誤警報が起こるたびに非常に高額な出費が発生します。不要なアラームを排除するためには、明確に信頼性の高い物体分類が必要となります。しかし、物体分類だけに依存することにならないように、録画ソリューションを設定する必要があります。このように設定することで、発信されるべきアラームが鳴らなかった場合は、録画をチェックして、アラームが発信されなかった理由を評価し、全体的な設置と構成を改善することができます。

事態の検索中にサーバー上で物体分類が行われる場合は、最初の録画をフィルタリングせずに、継続的に録画されるようにシステムを構成することが勧められます。継続的に録画することで大量のストレージが消費されますが、Zipstreamのような最新の圧縮アルゴリズムを活用することで、ある程度問題を軽減することができます。

8.4 メンテナンス

監視設備は定期的に整備する必要があります。VMSインターフェース経由でビデオを表示するだけでなく、物理的な検査を実施して、カメラ視野を妨害する可能性のある要素を発見して除去することが勧められます。標準型の録画のみの設備の場合も定期的な整備が重要となりますが、分析を使用する場合はさらにこの重要性が増します。

基本的なビデオ動体検知の観点からは、風に揺れるクモの巣など、一般的な障害物によって誤警報が増加することで、必要以上にストレージの消費量が増える可能性があります。オブジェクト分析の場合は、クモの巣により、基本的に除外範囲が生成されます。クモ糸により物体が覆い隠されるため、検知と分類の能力が大幅に低下します。

クモの巣により、監視カメラの視野が妨害される可能性があります。

日中は前面ガラスの汚れやレンズの気泡により問題が発生する可能性はほとんどありません。しかし、低光量条件下では、車のヘッドライトなど、汚れた気泡に光が当たることで不測の反射が発生し、検知精度が低下する可能性があります。

カメラのメンテナンスと同様に、シーン関連のメンテナンスも重要となります。カメラの耐用期間中に、監視対象のシーンで多くの変化が発生する可能性があります。前後の画像を単純に比較することで、潜在的な問題を把握することができます。カメラを配置したときのシーンの状況と現在の状況の変化、検知ゾーンを調整する必要性の有無、カメラの視野を調整する必要性の有無、カメラを別の場所に移動する必要性の有無などを検討してください。

9 プライバシーと個人の完全性

セキュリティと監視においては、プライバシーと個人の完全性に関する個人的権利と、犯罪の防止とフォレンジック調査の実施により安全性を向上するという抱負の度合いのバランスを取る必要があります。特定の設置状況やユースケースでは、慎重な倫理的考慮を図ること、および現地の法律を理解してこれに準拠することが必要となります。また、サイバーセキュリティを確保し、ビデオ素材への意図しないアクセスを防止するなど、ソリューションに対する要件も発生します。同時に、エッジベースの分析を用いて、統計目的でメタデータを生成する場合に、匿名化されたデータのみを後処理に送信することで、プライバシー保護を強化できると考えられます。

自動分析が導入された監視システムが増加していることで、いくつかの新しい側面を考慮に入れる必要性が発生しています。分析アプリケーションには誤検知のリスクがあるため、意思決定プロセスには経験豊富なオペレーターかユーザーが関与することが重要となります。多くの場合、これは「人間参加型」と呼ばれます。 また、アラームがどのように生成・表示されたかに対する考慮が人間の決定に含まれていることを確認することが重要です。分析ソリューションの機能に関して適切なトレーニングを受け、その機能を正確に理解していないと、間違った結論が導き出される可能性があります。

深層学習アルゴリズムがどのように開発されたかを考えると、追加の懸念が生じる可能性があります。一部のユースケースでは、技術を適用する際に慎重なアプローチが必要となります。こうしたアルゴリズムの品質は、基本的に、アルゴリズムのトレーニングに使用されたデータセット、つまりビデオと画像に左右されます。素材が慎重に選択されていないと、一部のAIシステムでは検知において民族的バイアスと性別バイアスの両方が介在する可能性があることがテストで実証されています。そのため、これについてはオープンな議論が発生しています。また、システムの開発中にこうした側面に確実に対処することを目的として、立法上の制限や活動の両方に対する問題も提起されています。

監視でAIが利用される確率が高まる中、技術を適用する場所と時期について有意義な議論を適度に実施しながら、運用効率と新たな潜在的ユースケースのメリットを利用していく必要があります。

10 付録

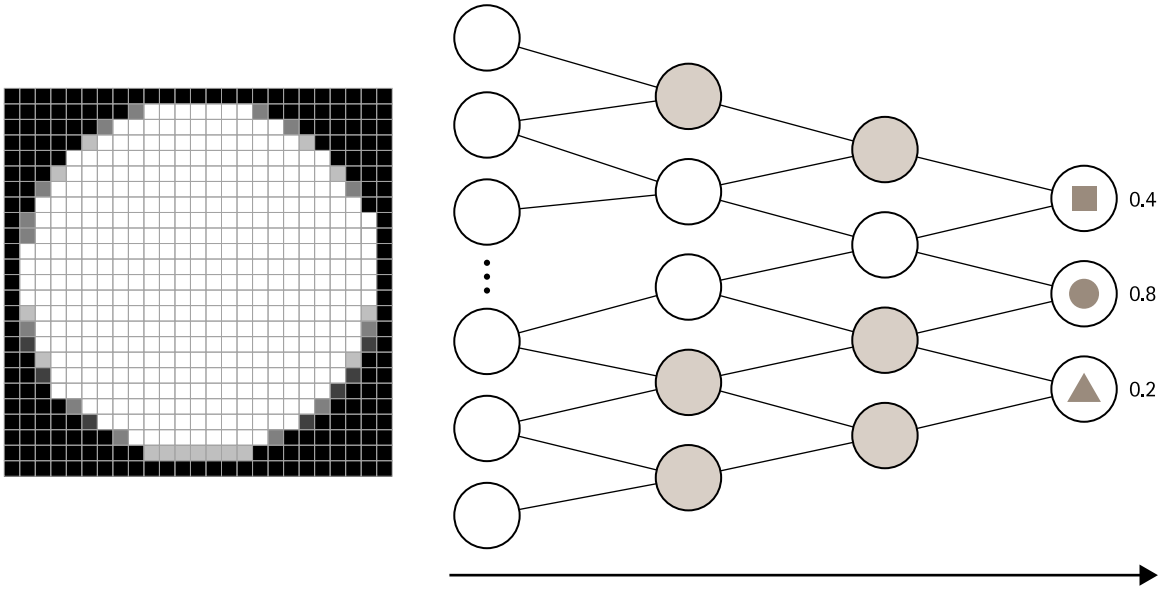
付録には、深層学習の基盤を形成する人工ニューラルネットワークに関する背景情報が含まれています。

10.1ニューラルネットワーク

ニューラルネットワークとは、脳機能に見られるいくつかの特性に類似したプロセスを通じて、データセット内の関係を認識するために使用される一種のアルゴリズムです。ニューラルネットワークは、相互接続されたいわゆるノードまたはニューロンの複数の階層で構成されています。情報は接続に沿って、入力層からネットワークを経由して出力層に伝達されます。

ニューラルネットワークが正常に機能するには、入力データサンプルを有限の機能セットに減少して、入力データの適切な表現を生成できることが前提条件となります。こうした機能が組み合わせられて、入力データが分類されます。たとえば、画像のコンテンツを説明するといった事柄です。

下図には、ニューラルネットワークで、入力画像が属するクラスが識別される方法の例が示されています。画像の各ピクセルは、1つの入力ノードで表されます。すべての入力ノードは、第1層のノードに結合されます。これにより出力値が生成され、これが入力値として第2層に伝達されます。各層では、重み関数、バイアス値、活性化関数がプロセスに関与します。



入力画像（左）とニューラルネットワーク（右）の例。出力層に到達すると、ネットワークでは、考えられる各カテゴリ（正方形、円、三角形）の確率が結論付けられます。確率値が最も高いカテゴリが、入力画像の最も可能性の高い形状となります。

このプロセスは*順伝播（フォワードプロパゲーション）*と呼ばれるものです。順伝播の結果が一致しない場合は、*誤差逆伝播法（バックプロパゲーション）*によりネットワークパラメータがわずかに変更されます。この反復トレーニングプロセス中に、ネットワークのパフォーマンスが徐々に向上します。

一般的に、展開後はニューラルネットワークには前のフォワードパスからのメモリーは存在しません。つまり、これが経時的に改善されることはなく、このニューラルネットワークでは、トレーニング済みの物体タイプの検知またはタスクタイプの解決のみが可能となるということです。

10.2畳み込みニューラルネットワーク（CNN）

畳み込みニューラルネットワーク（CNN）は、人工ニューラルネットワークの一種です。これは、コンピュータービジョンタスクに特に適していることが実証されており、深層学習の急速な進歩の中核を成しています。コンピュータービジョンの場合、ネットワークは縁、角、色の違いといった画像の特徴を自動的に検索するようにトレーニングされており、実際に画像全体の物体の形状を識別することが可能です。

これは主に*畳み込み*と呼ばれる数学演算により達成されます。個々のノードの出力は、入力データ容量全体ではなく、前の層により生成された入力データの限られた範囲にのみ依存するため、これは非常に効率的な操作と言えます。言い換えると、CNNでは、各ノードは前の層のすべてのノードに接続されているのではなく、小さなサブセットにのみ接続されています。畳み込みは、最も有用な情報を保持しながらデータサイズを縮小する他の操作によって補完されます。標準型の人工ニューラルネットワークと同様に、ネットワークに深く入り込むほどデータが抽象化されます。

CNNでは、トレーニング段階で、最善の層の適用方法が学習されます。つまり、畳み込み処理において、前の層の機能を組み合わせて、ネットワークの出力をトレーニングデータの注釈と可能な限り一致させる方法をシステムが学習するわけです。トレーニング済みの畳み込みニューラルネットワークでは、推論段階でトレーニングの結果である畳み込みの層が順次適用されます。

T10163986/ja/M2.2/2103

Axis Communicationsについて

Axisは、セキュリティの向上とビジネスの新しい推進方法に関する洞察を提供するネットワークソリューションを生み出すことで、よりスマートでより安全な世界の実現を目指しています。ネットワークビデオ業界をけん引するリーダーとして、Axisは映像監視、インテリジェントアプリケーション、アクセスコントロール、インターコム、音声システムなどに関連する製品とサービスを提供しています。Axisは50ヶ国以上に3,800人を超える熱意にあふれた従業員を擁し、世界中のパートナーと連携することで、カスタマーソリューションをお届けしています。Axisは1984年に創業し、スウェーデン・ルンドに本社を構えています。

より詳しい情報は*axis.com*をご覧ください。