

遠方小物体の検知可能距離を従来比 3 倍以上に拡大する映像監視技術を実証 ～4K リアルタイム解析で、道路・空港などのインフラ監視の広域化・省設備化に貢献～

発表のポイント:

- ◆ NTT 独自の超高精細映像 AI 推論 LSI により、4K 映像を高精細なままりアルタイムに解析し、遠方に小さく映る対象の検知可能距離を従来比 3 倍以上に拡大する映像監視技術を実証しました。
- ◆ カメラ 1 台でより遠方まで監視できる特長を活かし、道路・空港などのインフラ監視の広域化・省設備化への貢献が期待されます。
- ◆ 検知対象の「高精細な切り出し画像」と「動きの情報」を VLM へ伝えることで、1 分間の 4K 事故映像を用いた評価で従来判断できなかった事故を判断できることを確認するとともに、VLM への入力回数を約 98.1%削減しました。

NTT 株式会社(本社:東京都千代田区、代表取締役社長:島田 明、以下「NTT」)は、道路監視や空港監視など、幅広いインフラ監視への活用に向け、自社開発の「超高精細映像 AI 推論 LSI^{※1,2}」により、4K 映像を高精細なままりアルタイムに解析し、遠方に小さく映る対象の検知可能距離を従来比 3 倍以上に拡大するとともに、検知した対象の状況判断まで支援する映像監視技術を実証しました。

本技術では、4K 映像から遠方の小さな対象も高精細に捉え、「高精細な切り出し画像」と位置・速度・移動方向などの「動きの情報」を視覚言語モデル(VLM^{※3})へ入力します。さらに、必要な事象を検知した場合にのみ VLM による解析を実行することで、状況判断に必要な情報を保ちながら、VLM への入力回数や通信量を削減します。1 分間の 4K 事故映像を用いた評価では、4K 映像全体を VLM へ入力する構成では判断できなかった事故を本システムで判断でき、VLM への入力回数を約 98.1%削減しました。

なお、NTT イノベティブデバイス株式会社より、2026 年度内に、本 LSI を搭載した製品(Engineering Sample)の提供開始を予定しています。

本研究成果は、2026 年 11 月 18 日～25 日に開催される NTT R&D FORUM 2026^{※4}に展示予定です。

1. 背景

道路や空港などのインフラ監視では、限られた設備や人員で、広い範囲から事故や異物などの重要な事象を把握することが求められています。こうした監視の効率化に向け、近年では、物体の認識だけでなく、対象の状態や周囲との関係から「何が起きているか」を推論できる視覚言語モデル(VLM)の活用が期待されています。

一方で従来構成では、VLM への入力時またはその前段の物体検出時に画像縮小を伴うため、その過程で遠方の小さな対象などの重要な情報が失われます。このため、4K などの高精細カメラを用いても 1 台あたりの監視範囲を広げられず、後段での状況判断に必要な情報も得られないという課題があります。

2. 研究の成果

NTT は、4K 映像を縮小せず、高精細なままりアルタイムに解析できる独自開発の超高精細映像 AI 推論 LSI を活用し、遠方の小さな対象まで捉えるとともに、検知した対象の状況判断まで支援する映像監視技術を実証しました(図 1)。

本技術では、本 LSI を VLM の前段処理エンジンとして活用し、4K 映像から VLM が見るべき対象を高精細に捉え、その状況を判断するために必要な情報とともに VLM へ伝えることで、VLM への入力データ量を抑えながらも画質は保ち、VLM による状況判断を支援します。

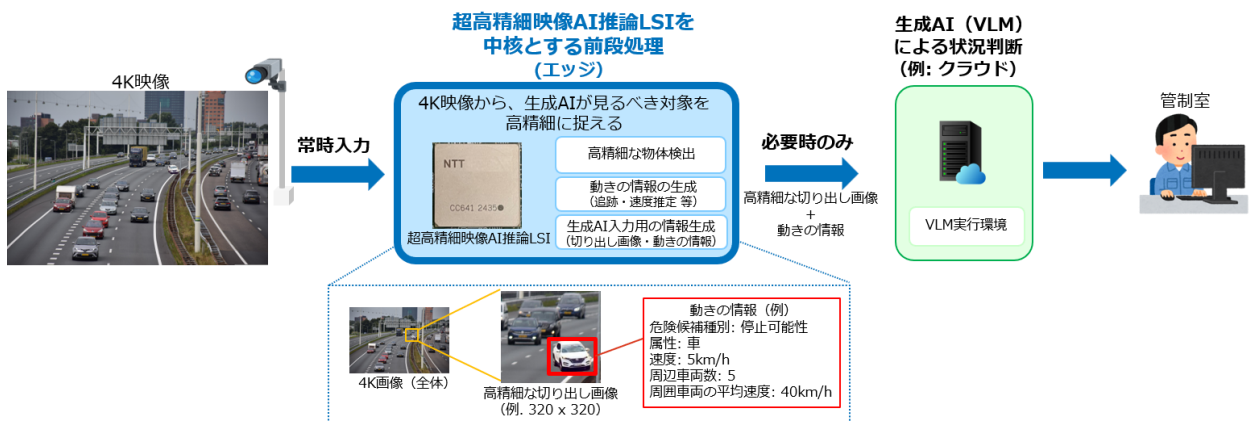


図 1. 超高精細映像 AI 推論 LSI を前段処理に活用した VLM 連携道路監視 PoC システム

3. 技術のポイント

(1) 遠方まで監視範囲を広げる 4K 高精細・リアルタイム解析

NTT が独自開発した超高精細映像 AI 推論 LSI は、4K 映像を縮小せず、1 秒間に 30 フレームでリアルタイムに物体を検出し、遠方に小さく映る対象も高精細に捉えます。4K 映像全体を一般的な物体検出モデルの入力解像度に縮小する方式と比べて、遠方の小さな対象の特徴を維持できるため、検知可能距離を 3 倍以上に拡大できることを確認しています。さらに、検出結果をもとに対象を追跡し、位置・速度・移動方向などの「動きの情報」を推定します。

(2) 「高精細な切り出し画像」と「動きの情報」で VLM の状況判断を支援

検知した対象の種別や動きの情報から、VLM による詳細な状況確認が必要と判定した場合、対象周辺を 4K 解像度を保ったまま切り出した高精細画像と、対象や周辺物体の位置・速度・移動方向などの「動きの情報」をセットで VLM へ入力します。これにより、VLM は切り出し画像から対象の形状や周囲との関係などの細部を確認するとともに、画像だけでは分からない時間的な変化を「動きの情報」から把握し、状況判断を行います。

(3) 必要な場面だけ VLM で解析するイベントドリブン処理

平常時は、消費電力 20W 以下の超高精細映像 AI 推論 LSI が 4K 映像を常時解析し、VLM による詳細な状況判断が必要な事象を検知した場合にのみ、VLM による解析を実行します。このイベントドリブン処理により、映像を継続的に VLM へ入力する構成と比べて、VLM への入力回数を大幅に削減できます。さらに、(2) のとおり、VLM へ入力する情報も高精細な切り出し画像と動きの情報を絞ることで、VLM へ送る情報量を抑え、処理負荷や通信量を低減するとともに、トークン消費量やシステム全体のエネルギー消費の低減も期待できます。

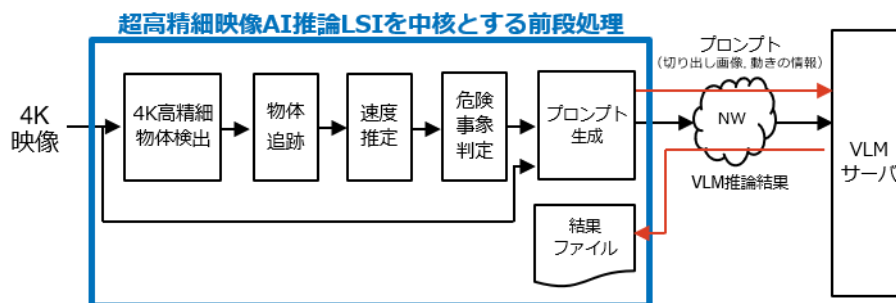


図 2. PoC システムにおける処理フロー詳細

4. 実験結果

事故発生の瞬間を含む 1 分間(1,800 フレーム)の 4K 映像※⁵を用いて、上記技術に基づいて構成した PoC システムの動作を確認しました。評価映像は、画面奥方向に走行する黄色の車両が上下線間の構造物に衝突し、車道を横切るように停止する事故場面を含むものです。

4K 映像全体を VLM へ入力した場合には、入力時の画像縮小によって黄色の車両が小さく潰れ、事故と判断できませんでした(図 3)。これに対し、本システムでは、独自 LSI を中核とする前段処理によって停止した黄色の車両を検知し、その周辺を元の 4K 映像から画素情報を保ったまま 320×320 画素で切り出した画像と、周辺車両を含む動きの情報を VLM へ入力することで、事故と判断できることを確認しました。

また、事故発生前後を含む映像変化の大きい 1 分間においても、VLM への入力回数は 34 回に抑えられました。これは、全 1,800 フレームを VLM へ入力する場合と比べて、入力回数を約 98.1% 削減したことに相当します。さらに、VLM へ入力する画像を対象周辺の局所領域に絞ることで、入力データ量も大幅に削減できることを確認しました。

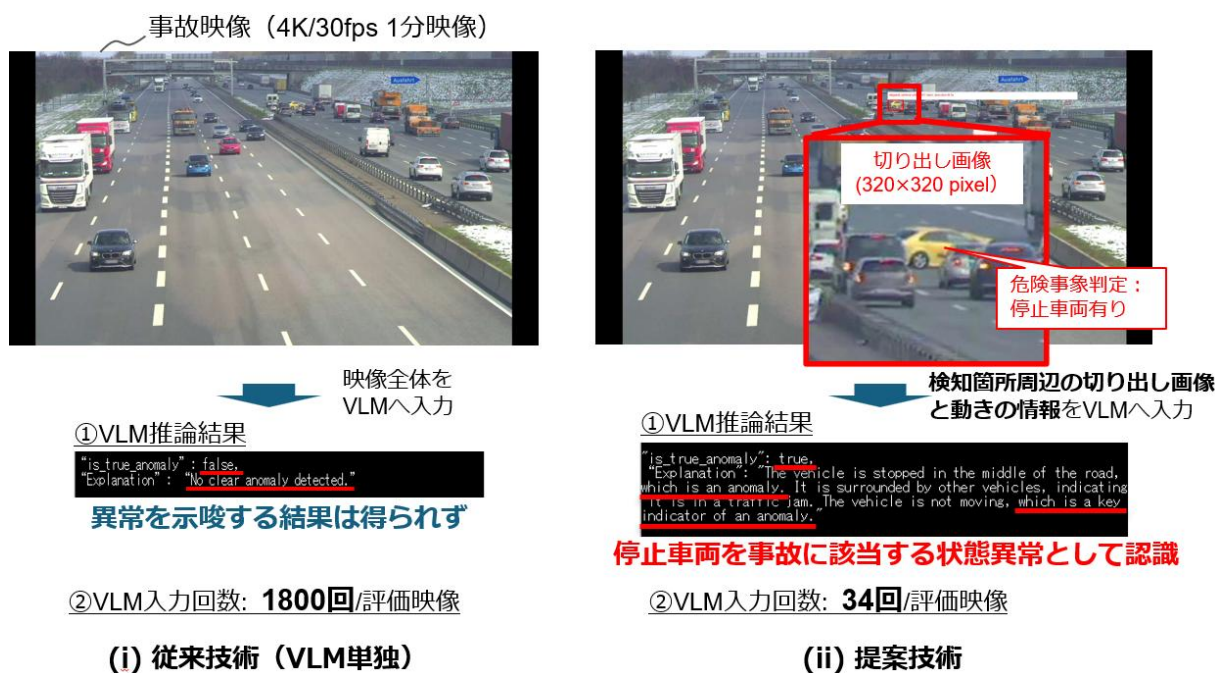


図 3. 評価結果

5. 今後の展開

今後は、道路監視や空港監視など、さまざまな監視分野への適用を進め、1 台のカメラでより遠方まで広範囲に監視できる特長を活かして、監視設備の広域化・省設備化に貢献していきます。さらに、VLM をはじめとする生成 AI との連携を高度化し、高精細映像から後段の解析や判断に必要な情報を効率的に抽出・伝達する「生成 AI の目」となる前段処理技術として、幅広い分野への展開をめざします。

なお、NTT イノベーティブデバイス株式会社より、2026 年度内に、本 LSI を搭載した製品 (Engineering Sample) の提供開始を予定しています。

6. 関連する過去の報道発表

・2025 年 4 月 11 日「超高精細映像からのリアルタイム AI 推論をエッジ/端末上でも実行可能にする AI 推論 LSIを開発 ～AI 推論の解像度制約を 4K に拡張し、リアルタイムかつ低電力での動作を実現～」(<https://group.ntt.jp/newsrelease/2025/04/11/250411a.html>)

【用語解説】

※1. LSI: Large Scale Integration

※2. 超高精細映像 AI 推論 LSI URL: <https://group.ntt.jp/newsrelease/2025/04/11/250411a.html>

※3. VLM: Vision Language Model

※4. NTT R&D FORUM 2026 <https://www.rd.ntt/forum/2026/>



※5 C. Creß et al., “A9-dataset: Multi-sensor infrastructure-based dataset for mobility research,” 2022 IEEE Intelligent Vehicles Symposium (IV), pp. 965–970, 2022. 本データセットの映像を 4K 解像度にアップスケールし、評価映像として利用

■本件に関する報道機関からのお問い合わせ先

NTT 株式会社

IOWN 総合イノベーションセンタ

広報担当

[問い合わせフォームへ](#)