

# 疎なネットワーク構造を持つ Deep Learning を用いた 映像分析システム

中尾 堯理<sup>†</sup> 井ノ口 裕也<sup>†</sup> 趙 雄心<sup>‡</sup> 山足 光義<sup>‡</sup>

後藤 央明<sup>†</sup> 佐藤 啓紀<sup>†</sup> 松本 渉<sup>‡</sup>

<sup>†</sup>三菱電機インフォメーションシステムズ株式会社 <sup>‡</sup>三菱電機株式会社

## 1. はじめに

近年、IoT や Deep Learning を代表とする人工知能技術の利用が盛んに進められている。我々は、ビデオカメラを活用した IoT ソリューションとして、低演算量 Deep Learning 技術を用いた映像分析システムを開発した。本システムは、安全・安心や様々な業務を支援すべく、ビデオカメラから取得した映像をリアルタイムに分析し、人の所持物や異常行動の判定、置き去り物検知、領域進入検知などを可能としている。

Deep Learning 技術は、通常 GPU を使用してシステムを構築することが多いが、GPU は CPU と比べて電力消費が激しく、現場への設置には向かない。そこで、本システムでは疎なネットワーク構造を持つ Deep Learning 技術を用いることで演算量を削減し、GPU 無しでのリアルタイム処理を実現した。また、複数の分析結果を組み合わせ、統合することで様々なニーズに対応可能とするとともに、元の画像上にまとめて表示できる機能を備えている。

本稿は、システムの構成と特長について述べる。

## 2. 映像分析システム

映像分析システムは、従来録画したまま何か起こるまでは撮り溜めるだけであったビデオカメラ映像をリアルタイムに分析することで、防犯やおもてなしサービス業務等の支援を行うためのシステムである。

### 2.1 システムの機能構成

本システムの機能構成を図 1 に示す。図に示す通り、ビデオカメラ映像を入力とし、分析結果を描画した映像を出力としている。映像 GW(ゲートウェイ)機能はビデオカメラから映像を取得し各分析機能にフレーム画像を配布する機能を持つ。分析機能は置き・座込・滞留検知、領域内状

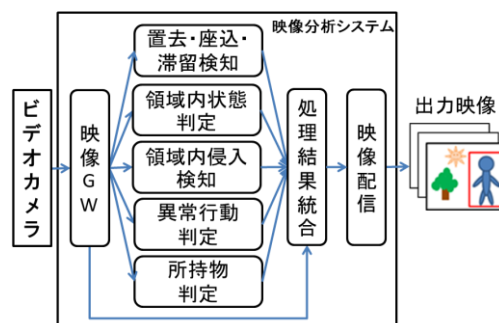


図 1 映像分析システム機能構成図

態判定、領域内進入検知、異常行動判定、所持物判定の 5 種ある。処理結果統合機能では、各分析結果、およびそれらを組み合わせた結果を元の画像上の対象に対し矩形で囲み、ラベルをつけてマージする。マージされた画像は、映像配信機能を用いて配信され、複数のシステム利用者が分析結果を確認できる。

### 2.2 特長

本映像分析システムは、以下の特長を有する。

#### (1) 現場設置を前提としたシステム

映像分析は処理負荷が高くリアルタイムで処理を行うには、GPU が使用されることが多い。一方、ネットワーク帯域から考えると、映像分析サーバはビデオカメラの近く、すなわち現場近辺に置く必要がある。現場近辺に設置する際には、GPU の消費電力や発熱量が問題となる。結果、現場に近い場所で分析処理を行うことが困難であった。本システムは、後述の疎なネットワーク構造を持つ Deep Learning 技術を適用することにより、GPU 無しの処理を実現することで消費電力、発熱量を抑え、現場のすぐ近くで分析結果の出力を行うことを実現した。

#### (2) 機能毎の映像処理結果の統合

システム利用者は分析のスペシャリストではないため、分析結果をわかりやすく示してほしいというニーズが多い。本システムでは後述の処理結果統合機能を用いて、複数の分析結果を組み合わせ、まとめたものを画像に描画し出力す

### A Video Analysis System Using Deep Learning with a Sparse Network Structure

Takamasa Nakao<sup>†</sup>, Yuuya Inokuchi<sup>†</sup>, Zhao Xiongxin<sup>‡</sup>, Mitsuyoshi Yamatari<sup>†</sup>, Hiroaki Gotou<sup>†</sup>, Hiroki Satou<sup>†</sup>, Wataru Matsumoto<sup>‡</sup>

<sup>†</sup>Mitsubishi Electric Information Systems Corporation

<sup>‡</sup>Mitsubishi Electric Corporation

ることが可能である。それにより、1つの画面で複数の分析結果を見ることができるようになる。

### 3. 低演算量 Deep Learning 技術の適用

Deep Learning 技術は、従来手法より高度な推論ができる一方、複雑なモデルを用いて計算するため、学習及び推論処理に必要な演算量・メモリ使用量が膨大になる課題がある。

#### 3.1 疎なネットワーク構造を持つ Deep Learning 技術

本技術は、Deep Learning におけるネットワーク構造と計算方法を効率化する新しいアルゴリズムである。具体的には、圧縮センシング理論を活用し、疎である事前分布の情報を伝播させる為に最適な疎なネットワーク構造とそれに対する計算アルゴリズムを開発した[1]。本手法により、従来型 Deep Learning と比べて認識精度を保ったまま、Deep Learning を低演算量化することが可能となる。演算量をどの程度まで削減できるかは推論対象に依存するが、画像認識の場合、学習・推論処理の演算量及びメモリ使用量を従来比で 1/10 まで削減することができる。低演算量化することによって、処理能力の低い組み込み機器や小型コンピュータ等に搭載することが容易になる。

#### 3.2 性能評価

今回、所持物判定に上記技術を適用することで演算量の削減を行った。表 1 は従来型 Deep Learning と低演算量 Deep Learning を使用した際の処理性能比較表である。ただし、同じ学習と評価データを利用し、認識精度が同じになるように、両手法のネットワーク構造を設計した。

表 1 処理性能比較表

	処理時間 /1 対象
従来型 Deep Learning (CPU)	119.6ms
従来型 Deep Learning (GPU)	12.7ms
低演算量 Deep Learning (CPU)	14.5ms

※GPU: Nvidia GeForce GTX 980

CPU: Intel Core i7 3.5GHz × 1コア

上記技術を用いることによって、従来型 Deep Learning (CPU) に比べ、1/8 以下の処理時間に短縮し、従来型 Deep Learning (GPU) の場合とほぼ同等の処理性能を実現した。

### 4. 処理結果統合機能

5つの分析機能は、処理速度がまちまちであり、それぞれ非同期で動作する。非同期で処理された結果を統合するため、映像 GW で画像だけでは

なく ID を付与し、各機能に画像を配布する。ID を付与することにより、各機能での処理結果の統合が可能となる。

この方式を用いることで、対象画像に対し複数の分析結果をまとめて描画した出力を行うことができる。また、複数の分析結果を組み合わせた検知条件を作成することも可能である。例えば、所持物判定でベビーカーの判定、置去・座込・滞留検知で置去り物の検知を行い、組み合わせることで、置き去りまたは人が押していないベビーカーの検知結果を示すことができる。

(図 2)

様々な検知ニーズに対応するために、本システムでは処理結果統合機能部に検知条件を集中させ、個別の条件を組み込むことができるよう設計している。



図 2 検知結果画像

### 5. まとめ

本稿では、今回開発した映像分析システムについて述べた。ビデオカメラ映像などのデータ量の多いものは集中処理センター等にデータを集めて分析を行うのは困難であり、現場に近い場所で処理を行うには演算量の課題がある。低演算量 Deep Learning を活用することで従来と同等の精度を保ちつつ、演算量を削減しリアルタイムの分析が可能であることを示した。

今後は GPU の小型化が急速に進むと予想される。GPU を使用し、より多くのビデオカメラを収容すると共に、複数ビデオカメラから取得した映像情報を統合して処理を行い、認識精度の向上及びより高度な分析機能を実現できるように開発を進めていく。

### 参考文献

[1]Matsumoto,W., et.al, "A Deep Neural Network Architecture Using Dimensionality Reduction with Sparse Matrices", 23rd International Conference, ICONIP 2016 pp 397-404 (2016)