

ディープラーニング軽量化技術の AISION®への適用

坂根 正造 渡辺 孝光

映像IoTシステム「AISION® *1) (アイシオン)*2)」の中核装置である「映像IoT-GW」は画像センシング機能を搭載したエッジデバイスである。この映像IoT-GWにディープラーニング軽量化技術を適用することで、高精度な画像センシングをエッジデバイスで実現可能とした。本稿では、エッジデバイスの特長を活かした映像IoT-GWへのディープラーニング軽量化技術の適用例を紹介する。以下ではエッジデバイスの一般的な性質を説明する場合には単に「エッジデバイス」と表記し、AISIONの特長を説明する場合には「映像IoT-GW」と表記する。

ディープラーニングによる画像センシング

近年IPネットワークカメラは世界的に普及が進み、さまざまな用途で活用されている。その用途の中でも、画像センシングは特に注目を集めている。画像センシングの主な利用目的を表1に示す。

表1 画像センシングの主な利用目的

分類	認識対象	用途
①セキュリティ強化	顔	特定人物を検出
	行動	特定の行動を検知、万引き検知など
②マーケティング	人物	特定エリア内やエリアを通過する人数をカウント
	人物属性	人物の特徴を抽出し、年齢、性別などの属性情報としてデータ化
③作業効率化	物体	特定の物体を検知、忘れ物検知など
	車番	ナンバープレートの文字を抽出、データ化

従来の画像センシングは、監視カメラを用いたセキュリティ強化用途で主に活用されてきたが、近年はマーケティングや作業効率化用途にも活用されるようになり、その需要は今後も拡大を見込まれている。

画像センシングの普及要因の一つに、画像解析技術の進歩がある。特にディープラーニングの登場によりその精度が飛躍的に向上し、より実用性の高い画像センシングの実現が可能となった。

従来の解析技術とディープラーニングの比較を表2に示す。

表2 従来解析技術とディープラーニングの比較

比較項目	従来の解析技術	ディープラーニング
解析方法の学習	人(専門家)が学習データに対する知識をコンピュータに教える	学習データよりコンピュータ自身が学習
学習データ量	少量	大量
学習データ量による精度	学習データ追加による精度向上の見込みが小さい	学習データ追加による精度向上の見込みが大きい
演算量	小	大(特に学習)

ディープラーニングはコンピュータ自身で解析方法を学習して解析精度を上げるために大量の学習データが必要であり、学習に必要な演算量も大きいため計算コストも多大となる。

映像IoT-GWの画像センシング概要

映像IoT-GWではさまざまな画像センシング機能をモジュールとして提供し、「画像センシングモジュール」を選択搭載することで用途に合わせた画像センシングが行える。図1は道路の設計や保守を目的とした車種別の車両通行量計測をする画像センシングモジュールを搭載した映像IoT-GWの動作例である。

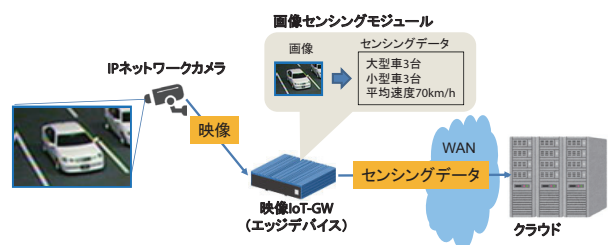


図1 映像IoT-GWでの画像センシング（車両通行量計測）

画像センシングをエッジデバイスで実施した場合とクラウドで実施した場合の特徴の比較を表3に示す。表3の①～③はエッジデバイスの長所であり、④～⑤は短所である。以下でこれらを具体的に説明する。

*1) AISIONは、沖電気工業株式会社の登録商標です。 *2) AISION (アイシオン) : AI(人工知能、eye)とvision(洞察力、先見の明、未来像)を掛け合わせた造語

表3 エッジデバイスとクラウドでの画像センシングの比較

比較項目	エッジデバイス	クラウド
①処理負荷	分散型	集中型
②データ通信量	小	大
③個人情報保護	容易	困難
④画像間連携	困難	容易
⑤高負荷解析	困難	容易

(1) エッジデバイスでの画像センシングの長所

図1では映像IoT-GWに対してIPネットワークカメラから道路を走行している車両の映像が入力される。映像IoT-GWは入力された映像に対して画像センシングを行い車両を抽出し、抽出した車両の車種（大型車/小型車）判別と車両の速度を測定しデータ化する。映像IoT-GWでは画像センシングをすることにより、クラウド（センター装置）の負荷を低減可能である（表3 ①の長所）。また、映像IoT-GWで車種別通行台数や速度だけをデータ化することにより、少量のセンシングデータを集計装置（クラウド）を送信することが可能となる（表3 ②の長所）。車両の映像には運転手の顔やナンバープレートなどの個人情報が含まれている可能性もあるが、映像IoT-GWでは画像センシングを行った後に映像を破棄することで個人情報の流出を防ぐことが可能である（表3 ③の長所）。

(2) エッジデバイスでの画像センシングの短所

画像間連携（表3の④）の例として、人、車両（ナンバープレート）の検出・追跡が挙げられる。人や車両の検出はエッジデバイスでも可能であるが、検出した人や車両がどこから来てどこに行こうとしているのかを追跡する場合は、複数地点の情報が必要となるためクラウドで行った方がよい。高負荷解析（表3の⑤）では、顔認証などの大量の登録データから特定の人物を特定するような場合では、処理能力の高いクラウドで行わせた方が効率的である。

軽量化ディープラーニングの概要

ディープラーニングは「学習」と「推論」のプロセスがある（図2）。「学習」では大量の画像と正解データの組合せからニューラルネットワークのパラメーター（畳み込みフィルターや全結合層の結合重みなど）を最適化する。学習によって得られたパラメーターとニューラルネットワークの構造の組合せを「学習済みモデル」と呼ぶ。

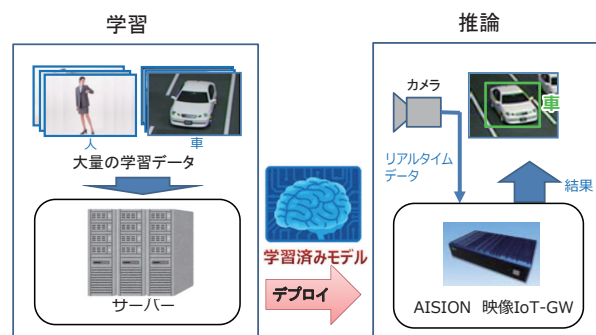


図2 学習とエッジデバイスでの推論

「推論」では学習済みのモデルを用いて入力データを処理する。推論の計算コストやメモリー量はモデルの層の深さやパラメーター数に依存する。よって推論処理をエッジデバイスでする場合、限られた計算リソースで動作させるためにモデルの軽量化が必要になる。

ディープラーニングの推論処理をエッジデバイスで動作させる際に問題になるのがプロセッサの処理能力とメモリー量である。そこで映像IoT-GWではモデルの軽量化に取り組み、ディープラーニングを利用した画像センシングモジュールの搭載を可能とした。以下に開発中の画像センシングモジュールで実施しているディープラーニングの軽量化方法を解説する。

(1) モデルのパラメーター削減

学習済みモデルから推論の精度に影響しない冗長なパラメーターを削除することにより、計算コストやメモリー量の削減を行える。図3に示すように画像識別のディープラーニングは畳み込み層やプーリング層、全結合層で構成される。畳み込み層では画像に複数のフィルターを掛けて特徴抽出をするが冗長なフィルターを削減することで主に計算コストを抑えられる。また全結合層では畳み込み層で抽出した特徴量に基づく分類をするが、全結合層の重みを低ランク近似¹⁾などでパラメーター削減することで主にメモリー量を抑えられる。ここで低ランク近似とは特異値分解などにより行列の次元を圧縮する軽量化手法である。

開発中の画像センシングモジュールでは精度が落ちない範囲で畳み込み層や全結合層のパラメーターを削減することで計算コストとメモリー量を削減する。

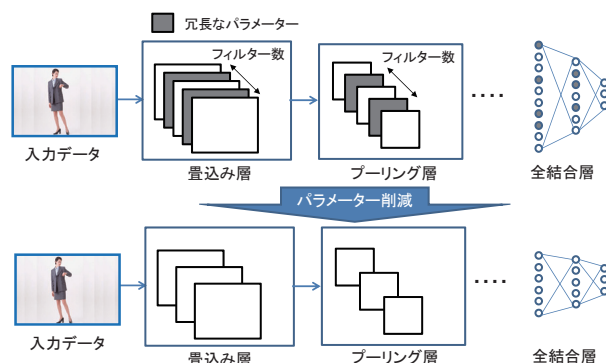


図3 モデルのパラメーター削減イメージ

(2) 低ビット量子化

一般的に32ビットの演算精度でおこなっているディープラーニングの計算を16ビット以下の低ビット演算に置き換えることで、モデルの構造とは関係なくメモリー量の削減や演算の高速化が可能となる。しかし、演算のビット数と推論の性能との間にはトレードオフがあり、低ビット量子化したモデルでは元のモデルに比べて性能劣化が発生する。特に顔画像から年齢を推定するような回帰問題では性能劣化が顕著である。よって、ディープラーニングで推論したい対象により適切なビット量子化を見極める必要がある。

(3) 知識の蒸留

ディープラーニングでは層が深くパラメーターの多いモデルの方が推論精度は上がりやすい。一方で、エッジデバイスでの動作は層が浅くパラメーターが少ない軽量なモデルが適している。そこで図4のように、推論精度は高いが計算コストが大きいモデルを「教師」、エッジデバイスで動作可能な軽量なモデルを「生徒」とし、教師の中間層や出力層の確率を生徒に学習させる「知識の蒸留」²⁾という手法がある。

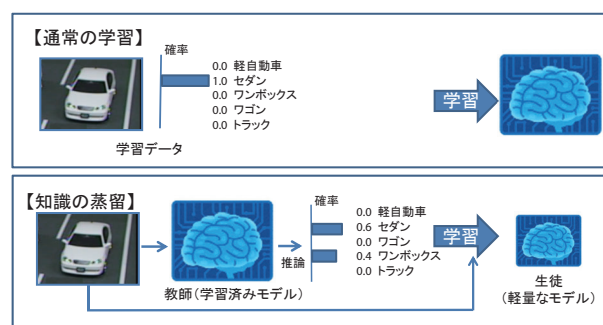


図4 知識の蒸留イメージ

例えば人間が見ても「セダン」か「ワンボックス」なのか判断しづらい車種を通常の学習のように直接「セダン」とは教えず、教師の出力「ワンボックスにも似ているセダン」を教えることで生徒に対し矛盾のない学習が行える。開発中の画像センシングモジュールで知識の蒸留を行なったところ、通常の学習に比べ精度が同等で計算コストを1/3にすることができた。

(4) 周波数重畳

上記の三つは映像IoT-GWの画像センシングモジュールで取り入れているディープラーニング軽量化の一般的な手法であるが、最後にOKI独自の周波数重畳を紹介する。OKIでは入力画像に対する空間周波数の解析結果と輝度情報を重畳した「周波数重畳ディープラーニング」を開発した。これにより悪環境に対して耐性があり、かつ軽量なモデルを実現している(図5)。ここで空間周波数とは画像の濃淡変化の頻度を表す。カメラの焦点から外れた物体はぼやけて映るため低周波成分が多く、焦点が合った物体は鮮鋭に映るため高周波成分が多くなる。悪天候時にカメラのレンズを横切る雪や雨はぼやけて映ることが多いことから、低周波成分を除去することにより、このような条件下での識別精度が向上する。また、画像縮小により失われる高周波成分を重畳することで、入力データのサイズを小さくしても高い識別精度を維持可能である。さまざまな天候下での実画像による車両識別評価では、識別精度を98.0%から99.8%に上げつつ、処理時間を16倍高速化し、メモリー量を1/10に削減可能であることを確認できた。

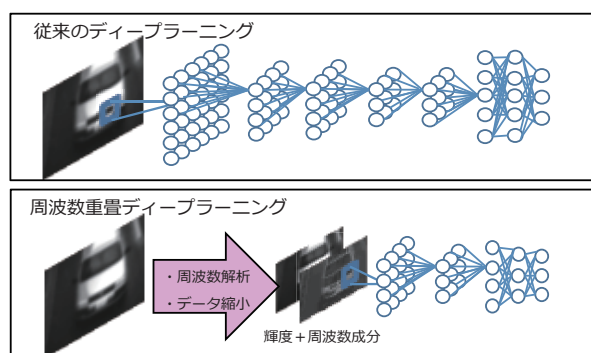


図5 周波数重畳イメージ

AIION映像IoT-GWでのディープラーニングの課題と今後

ここまでディープラーニングの軽量化を述べたが、エッジデバイスはセンサーに近い位置で処理するためリアルタイム性が要求される推論には適しているが、膨大な演

算が必要な学習は向いていないという特徴がある(表4)。

しかし、ディープラーニングを現場に適した精度に改善していくためには、PoC^{*3)}や運用しながら現場のデータを学習していくことが望まれる。

また、頻繁にデータが追加されるような顔画像による個人識別や車両の車種名識別などの用途では、エッジデバイスに対し随時学習済みのモデルの更新が必要になり管理が煩雑になる。

これらの課題に対しOKIでは映像IoT-GWとクラウドの連携を検討している。

表4 エッジデバイスとクラウドの特徴

	エッジ	クラウド
センサーまでの距離	近	遠
演算能力(単体)	低	高
システム管理	分散	一元化
通信データ量	小	大

(1) クラウド連携による現場環境への適合

映像IoT-GWで現場データを取得しクラウドで学習することで現場環境に適合した推論を可能にする。

図6に映像IoT-GWとクラウドを連携した現場環境へのディープラーニングの適合例を示す。映像IoT-GWから現場のデータをクラウドに送信し、クラウドでは複数の映像IoT-GWの現場のデータからモデルの再学習をする。学習したモデルは映像IoT-GWに再配置する。このサイクルを回すことで継続的にモデルを進化させ現場環境に適合した学習済みモデルに改善していく。

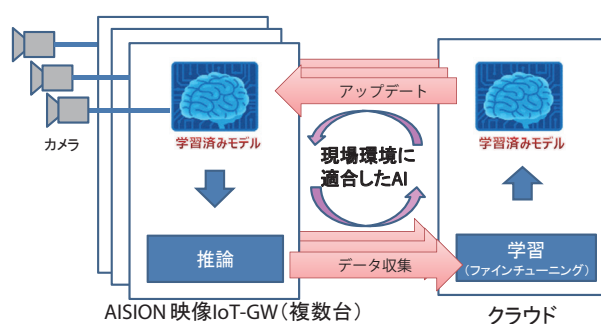


図6 現場環境への適合

(2) クラウド連携による推論処理の分散

推論処理をクラウドと映像IoT-GWで分担することで、推論精度やメンテナンス性を向上させることが可能である。図7に映像IoT-GWとクラウドを連携した分散処理の例を示す。映像IoT-GWでは一次処理の推論を行い(例えば

顔の検出)、クラウドでは層の深いモデルで二次処理の推論(例えば個人識別)を行う。クラウドですべて処理した場合、映像データをクラウドに送信する必要があり通信ネットワークへの負荷の増大を招くが、映像IoT-GWとクラウドで分散処理を行なうことで通信量を減らしエッジデバイス単体での推論に対して高精度化が可能になる。また、学習の更新頻度が高い推論をクラウドで行なうことで映像IoT-GWのメンテナンスも軽減される。

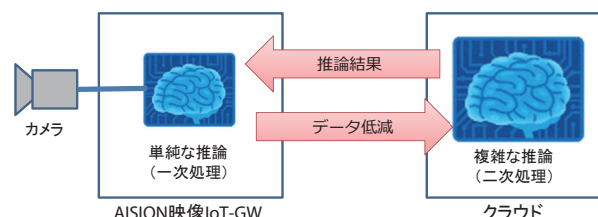


図7 クラウドとの分散処理

まとめ

本稿ではディープラーニング軽量化技術とその映像IoT-GWへの適用を紹介した。映像IoT-GWで提供している画像センシングの全てをディープラーニングに置き換えることは考えていない。ディープラーニングのメリット/デメリットを考慮に入れ、アプリケーションの要件を元に必要な部分にディープラーニングを適用することが重要である。今後は画像センシングの用途拡大のために、ディープラーニング軽量化技術などの高度な解析技術を活用した高精度画像センシングモジュールを提供する予定である。◆◆

参考文献

- 1) J. Xue, J. Li, and Y. Gong: Restructuring of deep neural network acoustic models with singular value decomposition, INTERSPEECH, pp. 2365-2369 (2013)
- 2) Geoffrey Hinton, Oriol Vinyals, and Jeff Dean: Distilling the knowledge in a neural network, arXiv preprint

●筆者紹介

坂根正造: Shozo Sakane. 情報通信事業本部 ネットワークシステム事業部 システム第五部

渡辺孝光: Takamitsu Watanabe. 情報通信事業本部 基盤技術センター 先端技術開発部

*3) PoC(Proof of Concept):新しい概念や理論の実現可能性を検証するための簡単な試行