

ネットワーク通信を用いた高精細映像の並列分散解析手法

Parallel distributed image analysis method through network communication

植村 宥喜[†]

水谷 后宏^{‡,§}

Yuki Uemura

Kimihiro Mizutani

1 はじめに

近年、映像技術の発展により、撮影映像の高解像度化が進んでいる。具体的には、現在、主流となっている、4K、8K 映像よりも遥かに高精細な 16K、32K 映像の利活用が近い将来に実現される兆しがある。高精細映像を画像解析する際、CNN[1]、SSD[2]、OpenCV[3]、YOLOv3[4] といった深層学習を用いて解析を行う。しかし、これらの手法は、解析対象となる映像の解像度を下げて (例:YOLOv3 の場合、 416×416) から解析することとなるため、高精細映像を直接解析する際、解析対象となるオブジェクトのサイズが縮小されるため、精度の高い画像解析を実現することは難しい。そのため、これらのアルゴリズムを用いて高精細映像を解析するためには、1つの画像を複数の画像に区切って解析する必要があるが、1台のコンピュータにて解析するためには、分割数によって多大な解析コストがかかることは明白である。そこで複数のコンピュータを用いて分割した複数の映像を並列的に処理するシステムを提案し、高精細映像を一台のコンピュータの解析コストを抑えつつ高精度に解析する手法を提案し、その基礎評価を行う。

2 画像解析アルゴリズムの概要

本システムで利用する画像解析アルゴリズムについて、YOLOv3[4] を採用することにする。その理由として、様々なコンピュータのアーキテクチャに対応したライブラリが提供されており、システムを効率的に構築できるためである。YOLOv3[4] では、一枚の画像を W 個 \times H 個のセルから構成されるグリッド分割を行い、各セルに対して予め登録してあるカテゴリ (クラス) に応じてクラス識別を実施し、物体検出を実行している。また、複数のクラスに属する物体があった場合に NMS(Non-Maximum Suppression) と呼ばれるパラメータである IoU 値 (画像の重なり度合いを表す値) と閾値を用いて物体検出の重なりをなくす手法が用いられている。よって本研究では、YOLO の発展技術である YOLOv3 を用いることとする。

3 提案手法の特徴

高精細映像を分割し、個々の分割された映像を解析する際、分割された領域に跨る解析対象 (例:人) が、分断される可能性があり、これによって、解析精度が低下する場合が想定される。本稿では画像解析をオブジェクト認識として扱うことにし、本問題を解決することを目指す。また、分割された映像を複数のコンピュータに送信する際、画像解析の対象とならない、つまり、認識したいオブジェクトが存在しない分割映像は、通信コストを高める可能性がある。本稿では、本通信コストの問題についても解決することを目指す。図 1 に、画像の分割と解析についての概要を記載する。提案手法では、解析対象の部分のみを分割することで画像解析の対象とならない画像を省くことができる。これにより画像全体を分割した場合よりも通信コストを下げるができることと考える。



図 1 画像の分割と解析

4 提案手法の詳細

提案手法では、分割時にオブジェクトが切れる問題に対して、スライディングウィンドウ分割 (SW 分割) を実施する。スライディングウィンドウ分割では、 x , y の座標空間上にて各座標空間 W サイズ毎に矩形に区切り、任意の解析対象 i の矩形領域を $(x_i, y_i, x_i + W, y_i + W)$ と定義する時、矩形を x もしくは y 座標上でスライドさせて、分割する。この時、分割した矩形同士が重ならなかつ、画像全体が同じ矩形サイズで分割可能とした場合、 $x_i + W = x_{i+1}$ 及び $y_i + W = y_{i+1}$ が成立する。スライディングウィンドウ分割では、矩形で分割する際に、矩形間の重なりを容認する。すなわち、 $x_i + W \neq x_{i+1}$ および $y_i + W \neq y_{i+1}$ が成立する。このようにすることで、認識すべきオブジェクトが分断されることを防ぐこととする。

[†] 近畿大学大学院総合理工学研究科, Graduate School of Science and Engineering Research, Kindai University

[‡] 近畿大学情報学部, Faculty of Informatics (KDIX), Kindai University

[§] 近畿大学情報学研究所, Cyber Informatics Research Institute, Kindai University

次に、画像解析の対象とならない、つまり、認識したいオブジェクトが存在しない分割映像を解析対象から省くため、DeepLabv3[5]を用いたオブジェクト群の判定を実施し、オブジェクト群のみを分割対象とすることとした。DeepLabv3とは、セマンティックセグメンテーションを実現するニューラルネットワークであり、物体のクラス(例：人)が存在する領域を特徴化する(例：塗りつぶし)する特性がある。これを用いて、まず解析対象となる映像から、認識したいオブジェクト群の特徴量マップを取得する。特徴量マップから、塗りつぶされたオブジェクト群に対して、二値化処理を実施し、塗りつぶされた領域を強調させる。ここで、塗りつぶされた部分領域を包含する矩形を抽出し、その矩形に対して、スライディングウィンドウ分割を行い、各分割映像を複数のコンピュータに送信し、解析することにする。塗りつぶされた部分領域を矩形で近似し、矩形間の距離(例：中心距離)に近いものを階層的にクラスタリングし、より多くの矩形をカバーする矩形を生成し、それをスライディングウィンドウ分割の対象とする。しかし、矩形間の距離が大きく離れている場合、分割対象となる矩形領域が大きくなる可能性がある。そのため、矩形 i, j に対して、中心座標を c_i, c_j と定義し、クラスタリングするか否かの境界値を T とした時、 $|c_i - c_j| < T$ となる場合のみ、両矩形を包含する矩形を生成することとした。最終的にスライディングウィンドウ分割された映像は、個々のコンピュータに送信され、解析を分散的に実施し、その結果を集約することで、解析を完了とする。

5 実験と評価

図2に、提案手法を用いた画像分割とその分割処理についての手順を示す。セグメンテーションでは、DeepLabv3[5]を用い、映像の二値化やスライディングウィンドウ分割にはOpenCV[3]を利用した。また、被写体の数に応じて、以下に示すように6パターンの手法による比較を行った。

1. 元映像を直接解析する
2. 元映像を等分割して解析する
3. 元映像をSW分割して解析する
4. 部分分割し、直接解析する
5. 部分分割し、等分割して解析する
6. 部分分割し、SW分割して解析する

なお、映像の分割は、 1000×1000 単位の矩形で分割することとし、スライディングウィンドウ分割におけるスライド幅は500pxとした。また、実験対象となる複数の8K写真を用いて、解析対象となるオブジェクトを正確に検出

できるかをYOLOv3[4]を用いて、解析することにした。

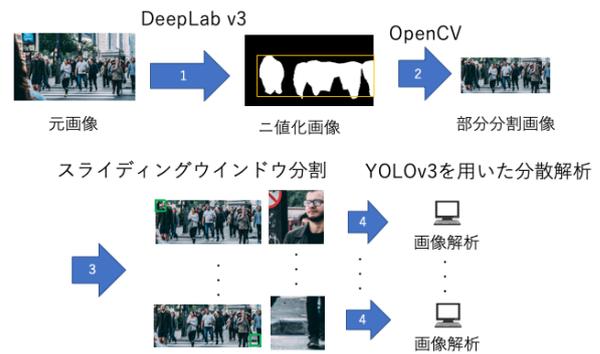


図2 画像分割の流れ

5.1 直接解析と分割を伴う解析の比較

まず、図3のように一枚につき1人が映る人物写真[6]を縦横に均等に並べて、一枚の画像として図4のように400人~6400人を作成した。これらの画像をそれぞれ 10000×10000 の大きさの画像に貼り付けたものを解析対象となる画像とし、実験を行なった。この実験では、高精細映像を解析する際、直接解析する場合と分割して解析する場合でどちらが高精度に解析できるかを比較する。



図3 並べられた人物写真

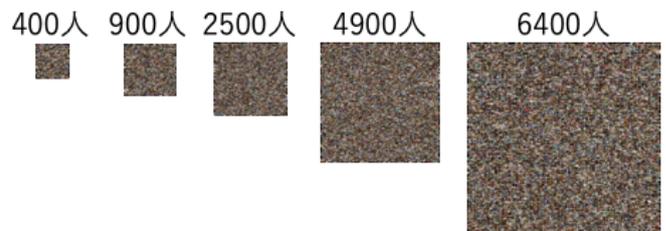


図4 作成した人物写真

この検証の結果を図5に示す。これらの結果から、元映像もしくは部分分割映像を直接解析する手法は、オブジェクト数が増えるほど検出できなくなることが明らかになった。また、分割を行う手法では、検出数では大きな差は見られなかったが、直接解析する手法よりも高精度に解析できていた。この検証によって、8K以上の映像に対して、そのまま解析するよりも分割して解析を行う方が、より高精度に解析することができると示された。したがって、8K以上の高精細映像に対しての解析では、分割を行なって解析する手法の優位性が示された。

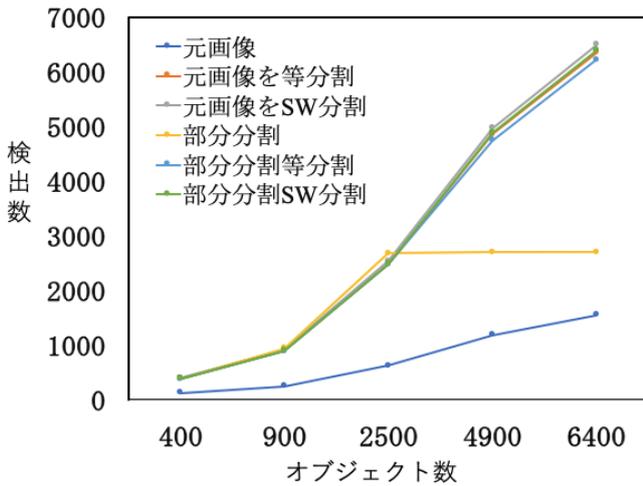


図5 認識精度の比較

5.2 検出数と通信コストの比較

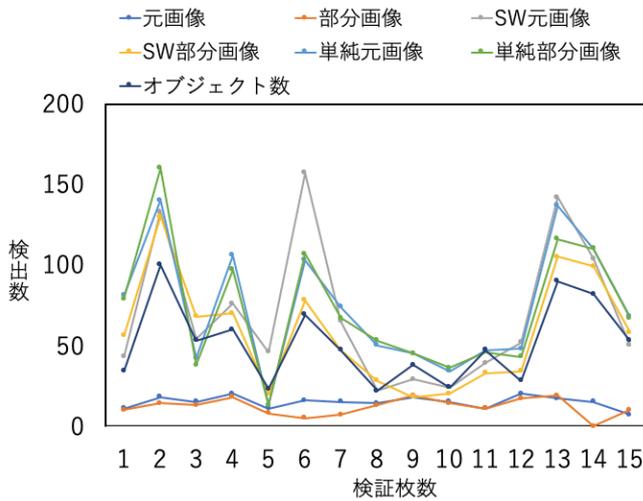


図6 認識精度の比較

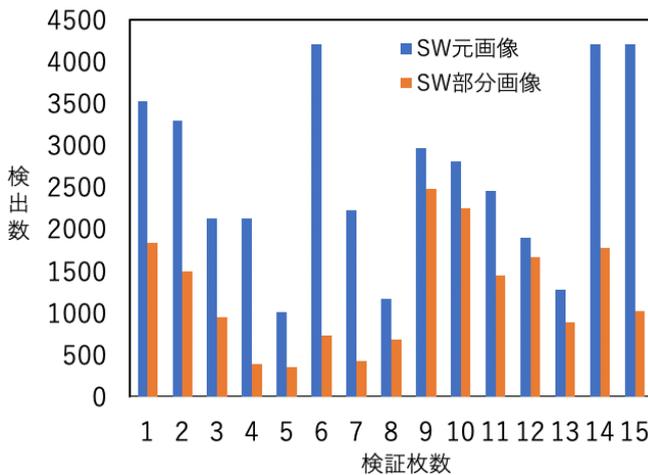


図7 通信コストの比較

次に、解析対象となる画像は、被写体として人・群衆が写っている 3089×2048 以上の高精細の画像である。今回比較する内容は2つである。1つ目は、どの手法がより正解のオブジェクト数に近いのである。2つ目は、元画像とオブジェクト群に対しスライディングウィンドウ分割

を行なった際、通信コストがどうなるかである。本実験結果を図6、図7に示す。実験結果より、オブジェクト群を抽出し、スライディングウィンドウ分割を行い解析する手法が、誤認識を2~30個に抑え、より正解のオブジェクト数に近い結果を出力した。また、画像全体をスライディング分割する手法に対して、提案手法では被写体が集中的に写っている部分のみに着目し、その部分を分割することで、分割枚数を75枚~982枚削減できることがわかった。この結果により、画像を分散的に解析する際の通信回数や画像の通信トラフィックの削減に貢献できるとわかった。ただし、画像に含まれる解析対象の配置分布によって、実験結果が大きく変わる場合がある。例えば、局所的に多くの解析対象が集合している場合、分割後の画像サイズは元の画像サイズより大幅に小さくなる。一方で、画像全体に解析対象が分散して配置されている場合は、画像全体を分割することになり、分割数を削減することができない。今後は、解析対象の配置に応じて分割方法を決定する方法を模索する必要があると考える。

6 おわりに

本論文では、画像を分散的に解析する際の、データの送信に掛かるトラフィック量を削減すべく、解析対象となる箇所を選定し、その部分のみを解析する手法を提案し、6パターンの手法による基礎評価を実施した。具体的には、解析対象が固まっている部分をセマンティックセグメンテーションを用いて特徴化し、その部分に対して、スライディングウィンドウ分割を用いて分割した。そして、分割した画像をそれぞれ解析することで、分散解析におけるトラフィックの削減に貢献した。今後は精度と解析速度のトレードオフ等を検証する予定である。

謝辞

本研究は JSPS 科研費 20K19791 の助成を受けたものです。

参考文献

- [1] Ross Girshick "Fast R-CNN" in Proc. ICCV, pp. 1440 - 1448, 2015.
- [2] Wei Liu, et al. "SSD: Single Shot MultiBox Detector" in Proc. ECCV, pp. 21 - 37, 2016.
- [3] OpenCV, URL : <https://opencv.org>
- [4] YOLOv3, URL : <https://pjreddie.com/darknet/yolo/>
- [5] Christian Szegedy, et al. "Rethinking the Inception Architecture for Computer Vision" arXiv:1512.00567, 2015.
- [6] Labeled Faces in the Wild, URL: <http://vis-www.cs.umass.edu/lfw/>