

## エッジ端末における 8K 超高精細映像物体検出手法の提案と評価 Object Detection in 8K Ultra-High Definition Video on Edge Devices

飯沼 宥光<sup>†</sup>  
Yuko Iinuma

八田 彩希<sup>†</sup>  
Saki Hatta

鶴澤 寛之<sup>†</sup>  
Hiroyuki Uzawa

吉田 周平<sup>†</sup>  
Shuhei Yoshida

山崎 晃嗣<sup>†</sup>  
Koji Yamazaki

### 1. はじめに

エッジ AI はリアルタイムに推論処理を実現できると同時に、プライバシーやセキュリティ保護の観点から注目を集めている。また、8K 超高精細カメラの小型化が進んだことで、8K カメラを搭載したエッジ端末が発表されている[1]。そのため、8K 映像からエッジ端末上で推論処理を行うエッジ AI 技術の需要が高まっていくと予想される。特に、映像中の物体の位置とクラスを推論する物体検出は、被災者の捜索や防犯をはじめ幅広い分野において応用が期待されている。

このような背景から、高精細映像から物体検出を行う手法として、画像を均等に分割して、各分割画像から得られた物体検出結果を合成する均等分割手法[2]や、物体が分布すると推定される領域を切り出して物体検出を適用する適応的切出手法[3-5]が提案されている。均等分割手法を 8K 映像に適用した場合、分割画像数の増加に伴って物体の切断による誤差が累積し、最終的な物体検出精度が低下するという課題がある。また、計算量が増大するためリアルタイムにエッジ端末上で実行することが難しい。一方、適応的切出手法では、物体検出を適用する切出画像数を大幅に削減できる可能性があるものの、場面によっては均等分割手法と同程度の画像数となる場合も考えられる。

そこで、本研究では密度推定を用いた適応的切出手法のもとに、物体検出精度の低下を抑制しながら、従来手法からさらに切出画像数を削減可能な物体検出手法を提案する。

### 2. 提案手法

#### 2.1 提案手法の概要

提案手法の概要を図 1 に示す。まず縮小した入力画像全体から密度推定モデルを用いて物体の分布密度を推定する。この密度分布から物体検出を適用する候補となる領域を矩形として抽出する。さらに、計算量を削減し、エッジ端末上で一定の時間内に処理を完了できるようにするため、2.3 節で述べる矩形選択によって所定の数まで物体検出を適用する矩形を絞り込む。最後に、選択された矩形を入力画像から切り出して物体検出を適用する。

#### 2.2 密度推定と矩形抽出

本手法では、各物体のサイズに合わせて密度分布を与える方法で作成した正解密度マップを用いて、密度推定モデルの学習を行う[6]。この学習方法で得られる密度推定モデルを用いることで、入力画像に映る大小両物体の密度分布が得られるようになる。この密度分布には、小さな物体が密集した領域からは分散が小さく密度の高い分布が、大きな物体に対しては分散が大きく密度の低い分布がそれぞれ推定されるという特徴がある。

<sup>†</sup> 日本電信電話株式会社デバイスイノベーションセンター  
Device Innovation Center, NTT Corporation

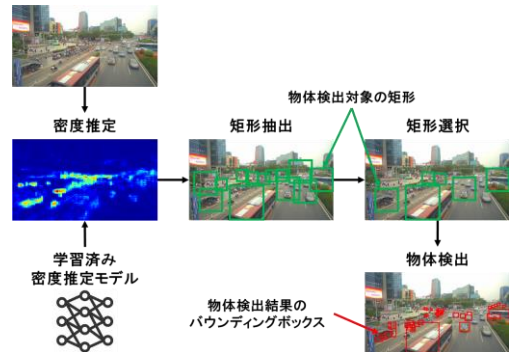


図 1 提案手法の概要

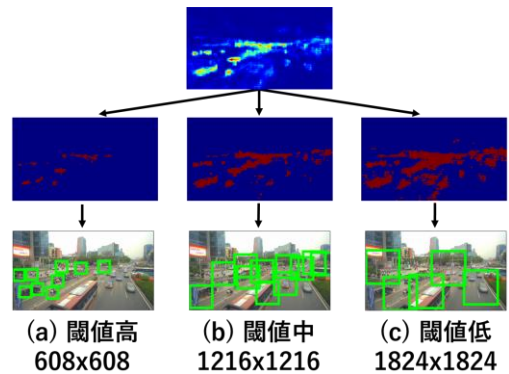


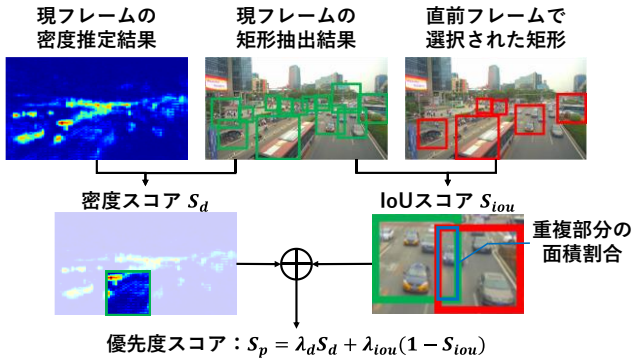
図 2 多段階閾値を用いた矩形抽出方法

そこで、密度推定結果から物体検出を適用する候補となる領域を矩形として抽出する矩形抽出では、推定された密度値に対して多段階の閾値を設定し、高い閾値からは小さな矩形を、低い閾値からは大きな矩形を抽出する。これによって、入力画像に映る物体のサイズに適した矩形を抽出することが可能となる。3 段階の閾値を用いて矩形抽出を行う場合の概要を図 2 に示す。まず、推定された密度マップについて、与えられたそれぞれの閾値以上の値を持つ領域を 1、それ以外を 0 とする 2 値画像を生成する。次に、ピクセル値が 1 の連続領域を抽出する。最後に、各連続領域に対して、それを内包する矩形を割り当てる。これによって、それぞれの閾値に対応する矩形を抽出することが可能となる。

#### 2.3 矩形選択

矩形抽出で得られた物体検出を適用する候補領域の中から、矩形選択によって所定の数の矩形を選択し、その矩形で画像を切り出して物体検出を適用する。選択されなかった領域については、過去フレームで得られた物体検出結果を映像の動きに合わせて座標補正して用いることによって、物体検出精度を保ちながら物体検出を適用する矩形数を削減することができる。

矩形選択では、矩形の重複度を表す IoU(Intersection over Union)スコアと、ある矩形内の物体の密度を表す密度スコ



アを組み合わせて各矩形の優先度スコアを計算する。IoU スコアは、過去フレームで選択された矩形と重複度の小さい矩形を選択するのに用いる。これによって画像全体の各領域に満遍なく物体検出を適用できる。一方、密度スコアはより多くの物体が分布すると推定される領域にある矩形を選択可能であり、映像開始や場面転換の直後において特に有用である。本手法では、密度スコアと IoU スコアの加重和を取り、映像中の場面に合わせた矩形選択を実現する。

具体的な矩形の選択方法を図 3 に示す。まず、現フレームから推定された密度マップ上において、抽出された矩形内の各ピクセルの密度の和を計算し密度スコア  $S_d$  とする。また、現フレームで抽出されたある矩形について、直前フレームで選択された矩形との IoU を計算し、その最大値を IoU スコア  $S_{iou}$  とする。次に、密度スコアと IoU スコアの係数をそれぞれ  $\lambda_d$ 、 $\lambda_{iou}$  として、優先度スコア  $S_p = \lambda_d S_d + \lambda_{iou} (1 - S_{iou})$  を計算する。この時、直前フレームで選択された矩形と重複の小さな矩形を選択するため、 $1 - S_{iou}$  としている。また、1 フレーム目では  $\lambda_d = 1$ 、 $\lambda_{iou} = 0$  とし、2 フレーム目以降では  $\lambda_{iou}$  の比率が大きくなるように、フレームごとに調整する。場面転換が検出された場合には、 $\lambda_d = 1$ 、 $\lambda_{iou} = 0$  にリセットする。最後に、優先度スコアの高いものから順に所定の数の矩形を選択する。

### 3. 評価

#### 3.1 評価方法

25K 映像で構成された PANDA データセット[7]を 8K にリサイズして密度推定モデルの学習、および物体検出精度の評価を行う。密度推定モデルの学習には PANDA の画像セット PANDA-image を利用し、物体検出精度の測定には PANDA-image と重複する映像シーケンスを排除した映像セット PANDA-video の 1306 枚の画像を利用した。また、物体検出モデルには GitHub で公開されている YOLO v3 [8] を用いている。なお、エッジ端末[9]におけるリアルタイム実行を可能とするため、本評価では矩形選択の絞り込み数を 6 とした。

#### 3.2 評価結果

提案手法で密度推定と矩形抽出のみを行った場合、均等分割手法を用いた場合、YOLO v3 で縮小した 8K 画像から物体検出を行った場合の物体検出精度 mAP (%) と、物体検出を適用した分割・切出画像数を表 1 に示す。mAP は高いほど物体検出精度が高いことを表している。また、均等分割手法の画像数は 608 ピクセル四方で 8K 画像を分割した

表 1 物体検出精度と分割・切出画像数の比較

	mAP (%) ※1	分割・切出画像数 ※2
提案手法 (矩形抽出のみ)	23.3	18.9
均等分割 (全体 + 全分割)	22.0	105
YOLO v3 (全体のみ)	0.0	1

※1 信頼度上位500件についてIoU≧50%を正解として計算

※2 提案手法では全画像の平均切出画像数を計算

表 2 矩形選択結果の評価

	mAP (%)	分割・切出画像数
矩形選択なし	23.3	18.9
矩形選択あり	18.6	6
物体検出数 (%) ※		77.2

※ 「物体検出数 (%)」は「矩形選択なし」の場合の検出数に対する「矩形選択あり」の場合の検出数の割合

際の画像数であり、提案手法で矩形選択まで行った場合に示している切出画像数は、1306 枚の評価用画像で得られた切出画像数の平均値である。提案手法で密度推定と矩形抽出までを行った場合には、均等分割手法と比較して平均で 1/5 以下の分割・切出画像数で物体検出に必要な計算量を大幅に削減しながら、1.3%高い mAP を達成している。これは、YOLO v3 では全体画像からほとんど物体が検出できておらず、均等分割手法において物体検出結果の合成処理が機能せず精度が低下したためであると考えられる。

さらに、矩形選択を実施した場合としなかった場合の物体検出精度と物体検出数の割合を表 2 に示す。矩形選択によって切出画像数を 1/3 以下の 6 まで削減しながら、77.2% の物体検出数を維持し、mAP の低下を 4.7% に抑えられている。これによって、精度劣化を抑えながら物体検出を適用する画像数を大幅に削減できることを確認した。

### 4. おわりに

本研究では、適応的切出手法をもとに、矩形抽出と矩形選択によって、精度を保ちながら物体検出を適用する切出画像数を削減する、エッジ端末でリアルタイム実行可能な物体検出手法を提案した。また、提案手法について評価を行い、均等分割手法に比べて 5 分の 1 以下の分割・切出画像数で mAP が 1.3% 向上すること、さらに矩形選択によって 6 矩形まで切出画像数を削減しながら、mAP 低下を 4.7% に抑えられることを確認した。

#### 参考文献

- [1] DJI Inspire 3. Available online: <https://www.dji.com/jp/inspire-3>. (参照 2023-05-25)
- [2] H. Uzawa *et al.*, "High-definition object detection technology based on AI inference scheme and its implementation", IEICE Electronics Express, 2021, Volume 18, Issue 22, Pages 20210323.
- [3] C. Li *et al.*, "Density Map Guided Object Detection in Aerial Images," 2020 IEEE/CVF CVPRW, 2020, pp. 737-746.
- [4] F. Yang *et al.*, "Clustered Object Detection in Aerial Images," 2019 IEEE/CVF ICCV, 2019, pp. 8310-8319.
- [5] J. Zhang *et al.*, "How to Fully Exploit the Abilities of Aerial Image Detectors," 2019 IEEE/CVF ICCVW, 2019, pp. 1-8.
- [6] 飯沼有光 他, "エッジ端末における 8K 超高精細映像 AI 推論処理手法の提案", 電子情報通信学会ソサエティ大会, 2022.
- [7] X. Wang *et al.*, "PANDA: A Gigapixel-Level Human-Centric Video Dataset," 2020 IEEE/CVF CVPR, 2020, pp. 3265-3275.
- [8] A. Bochkovskiy. 2018. Yolo v4, v3 and v2 for Windows and Linux. <https://github.com/AlexeyAB/darknet>. (参照 2023-04-28)
- [9] NVIDIA Jetson Xavier NX. Available online: <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-xavier-nx/>. (参照 2023-05-16)