

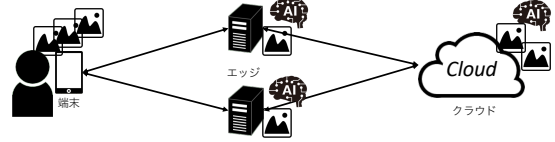
# 分散映像分析システムの消費電力最小化方式の実装と評価

大阪大学 基礎工学部 情報科学科 村田研究室  
川口峻平

2023年2月14日

## 研究背景

- VRやデジタルツインなどの技術が普及
  - サービスのリアルタイム性・映像の分析精度が重要
  - 従来は計算リソースが豊富なクラウド上で分析するシステムを想定
    - ネットワーク上で大量のデータ
    - 通信や計算に要する消費電力の増加
- 分散型の映像分析システムの実現が必要
  - エッジコンピューティング技術に着目
    - クラウドまで流れるトラフィック量の削減による消費電力の削減
  - エッジ・クラウドコンピューティング技術による映像分析
    - リアルタイム性と認識精度の要求を満たしつつ省電力化
    - クラウドのコンピューティングリソースを効率的に利用



## 分散映像分析システムとその消費電力最小化<sup>[10]</sup>

- 分散映像分析システム
  - 端末・エッジ・クラウドに映像を分散させて映像分析
  - 3種類のAIモデルを用いた映像の分析
    - YOLOv3-tiny・YOLOv3・YOLOv3-spp

### 消費電力最小化方式

$$F = -P + \sum_{s \in S} (T_s^{max} - T_s) \alpha I + \sum_{s \in S} (A_s - A_s^{min}) \beta I$$

$$P = \sum_{d \in D^s} \frac{\text{機器の消費電力}}{P^d(t)} + \sum_{n \in N^s} \frac{\text{ネットワークの消費電力}}{P^n(t)}$$

- 機器が分析する映像の割合とAIモデルの組み合わせを最適化して実現
  - 遺伝的アルゴリズムを用いて各機器のフレーム処理割合を決定
  - 遅延と認識精度の制約を満たしつつ消費電力を最小化

[10] 西 英之, 村田 正幸, 長谷川 剛: "分散映像分析システムの消費電力最適化方式の検討", 信学技報, 122, 275, pp. 28-33 (2022).

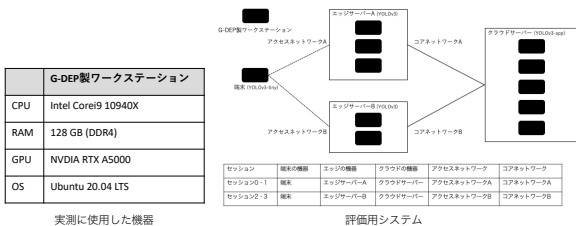
## 研究目的とアプローチ

- 研究目的
  - 分散映像分析システムを対象とした消費電力最小化方式の実装
    - 実機を用いたシステムを対象に最適化を行い効果を検証
  - 消費電力予測モデルとの予測値と実測値の誤差の検証および低減
    - 消費電力をより高精度に予測することで電力最小化の信頼性を向上
    - 従来の予測モデルの改善
- アプローチ
  - 実機を用いたシステムを構築し評価
    - 機器に映像分析タスクを割り当てて消費電力と遅延と認識精度を実測
  - AIモデル・フレーム処理割合・映像セッション数の組み合わせをさまざまに変更して消費電力の実測値を取得
    - 従来の消費電力予測モデル<sup>[10]</sup>の予測値と実測値の誤差の分布を調査
    - GPU消費電力を実測し計算機の消費電力との関連性を調査

[10] 西 英之, 村田 正幸, 長谷川 剛: "分散映像分析システムの消費電力最適化方式の検討", 信学技報, 122, 275, pp. 28-33 (2022).

## 評価用分散映像分析システムの実装と評価

- 評価用のシステム構成
  - 端末1台・エッジ2台・クラウド1台
  - 対象の機器に映像分析タスクを割り当てて実測
  - 機器に割り当てるフレーム処理割合を処理性能に応じて割り当てる
- 実装したシステムで映像分析した時の消費電力を実測
  - 各機器の映像分析時の消費電力の実測値をもとにシステム全体の消費電力を計算
  - ネットワークでの消費電力はシミュレーションから出力されたものを使用



実測に使用した機器

評価用システム

## 従来の予測モデルと実測値

- 消費電力の実測値と予測モデルの予測値には誤差が発生
  - システム全体で2.7%の誤差が発生
    - 端末で10.6%
    - エッジサーバー-Aで1.3%
    - エッジサーバー-Bで0.8%
  - 端末とエッジサーバーの誤差が大きく異なる
    - AIモデルや映像セッション数の違いに依存
    - フレーム処理割合に依存

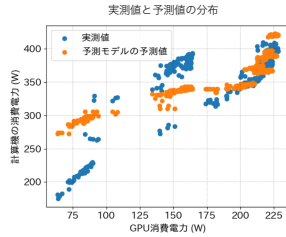
	実測値	予測値	誤差 (絶対値)
端末	376.37 W	336.41 W	39.97 W
エッジサーバー-A	1116.81 W	1102.18 W	14.63 W
エッジサーバー-B	741.88 W	735.83 W	6.04 W
システム全体	2242.24 W	2181.59 W	60.64 W

- 誤差が原因で必ずしも消費電力が最小化されない可能性
  - 誤差の検証
  - 高精度に予測を行う消費電力予測モデルが必要

## 従来の予測モデルと実測値の誤差の検証

7

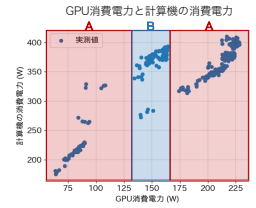
- 誤差の検証方法
  - AIモデル・処理割合・映像セッション数の組み合わせを変えて実測
    - 90パターン (3×3×10)
  - GPU消費電力を同時に実測
    - 映像分析時に消費する電力の多くがGPUで消費されていると仮定
- 検証結果
  - 平均絶対パーセント誤差は9.3%
  - GPU消費電力によって誤差の傾向が変動
    - AIモデルや処理割合の差によるもの
  - GPU消費電力が小さい時に誤差が大きい
    - YOLOv3-tinyを用いた映像分析時
    - GPU利用率に比べてCPU利用率が大きい



## 新たな消費電力予測モデル

8

- GPU消費電力と計算機の消費電力の関係
  - AIモデルや映像セッション数によってクラス分けできる



- 予測モデルの改善
  - GPU消費電力によって (A, B) 2つの場合に分けて予測
  - 実測値から係数と固定の消費電力を決定

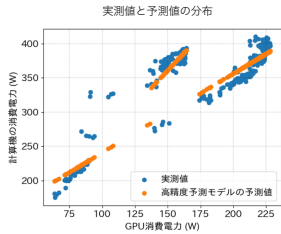
$$P^d(t) = \begin{cases} P_{GPU}^d(t)\gamma^P + P_{fixed_1}^d & (137W \leq P_{GPU}^d(t) \leq 164W) \\ P_{GPU}^d(t)\delta^P + P_{fixed_2}^d & (else) \end{cases}$$

$$\begin{aligned} (\gamma^P, \delta^P) &= (2.04, 1.16) \\ (P_{fixed_1}^d, P_{fixed_2}^d) &= (55.81, 125.60) \end{aligned}$$

## 新たな予測モデルと実測値の誤差

9

- 誤差の検証方法
  - AIモデル・処理割合・映像セッション数の組み合わせを変えて実測
    - 90パターン (3×3×10)
  - GPU消費電力を同時に実測
- 検証結果
  - 平均絶対パーセント誤差は3.5%
  - GPU消費電力に関わらず高精度に予測
    - AIモデルや処理割合の違いによる誤差を低減
    - GPU消費電力が150W付近の場合も高精度に予測
    - 従来モデルに比べてGPU消費電力が小さい時の誤差を低減
  - 誤差は発生
    - CPU消費電力の考慮が必要



## まとめと今後の課題

10

- まとめ
  - 分散映像分析システムの消費電力最小化の効果を実機で検証
  - 従来の消費電力予測モデルの誤差の検証
    - 映像分析時の消費電力にはCPU消費電力も関係があることを確認
  - より高精度に予測を行うための予測モデルの改善
    - 実測値との平均絶対パーセント誤差を3.5%まで低減
- 今後の課題
  - 実アプリケーションを想定した上での省電力効果の検証
    - 端末・エッジ・クラウドの機器を選定して行う
  - CPU消費電力とGPU消費電力を消費電力予測モデルに反映
    - CPU消費電力の実測を同時に行う手法を検討