

## D-08 画像生成により拡張したデータを用いたリアルタイム人物再同定

伊福佑耶\*, 新井康平\*\*, 小田まり子\*

(\*久留米工業大学大学院工学研究科電子情報システム工学専攻、\*\*佐賀大学)

## 1. 概要

本研究では、リアルタイム人物再同定(RE-ID)と精度向上を目的として、画像生成AIによるデータ拡張を導入した。生成画像によるデータ拡張を比較実験として実施し、リアルタイム性能と特徴抽出性能を評価した。実験の結果、GPU環境およびCPU環境において、生成画像を用いた拡張による精度改善とリアルタイム性の確保が確認された。今後の展望として、より多くのOpen-setデータを使用したランキング精度の検証を行い、ポーズ変化や姿勢変化の影響を考慮した実験を実施する。

## 2. 目的・背景

近年、監視カメラやスマート施設における人物再同定は、セキュリティや行動解析の分野で重要性を増している。特に、複数カメラにまたがる人物の追跡や人物の識別精度向上は、多様な応用において不可欠である。しかし、Re-IDモデルの性能向上には大量かつ多様な学習データが必要であり、視点、オクルージョン、ポーズの変化、不確実な映像シーケンスなど、多くの不確実な課題がある<sup>[1]</sup>。このため、データ不足による汎化性能の低下が大きな課題となっている。さらに、近年提案されているTransformerをベースとしたTransReID<sup>[2]</sup>などの最先端手法は高精度である一方、計算コストが大きく、リアルタイム処理には不向きである。本研究では、画像生成技術を用いたデータ拡張により、データ不足を原因とする汎化性能の低下を改善しつつ、軽量かつ高速なRe-IDシステムの開発を目指す。これにより、実環境でのリアルタイム人物再同定の実現を目的とする。

## 3. 提案手法

本研究では、検出から検索・可視化まで一貫した再識別パイプラインを構築する。学習段階ではパラメータ数が少なくMarket-1501データセットでRank-1を94.8%達成しているOSNet<sup>[3]</sup>を用いた特徴表現をTriplet Lossとクロスエントロピーで最適化する。人物の登録段階で、学習したOSNetの特徴抽出器を用いて、参照画像から画像拡張を行い、ベクトル表現を取得、FAISS(Facebook AI Similarity Search)に代表ベクトル、pk1ファイルに個人のIDを保存する。推論時はYOLOv11で対象領域を検出し、OSNet特徴に基づくFAISSからL2距離の最近傍探索で高速に照合する。図1にシステムの構成図を示す。

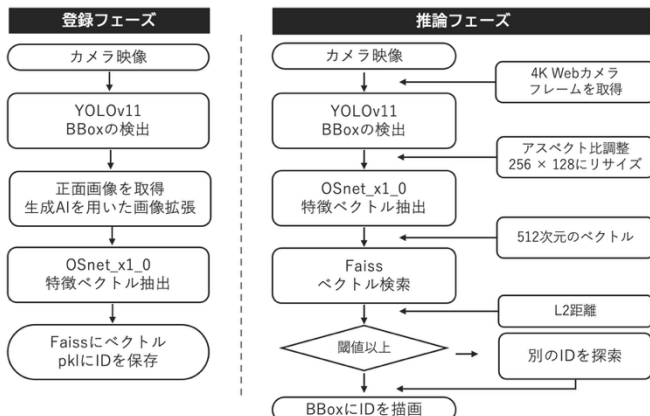


図1. システム構成図

## 4. 実験

本研究の実験では、GPUとしてNVIDIA RTX 3070 Ti、CPUとしてIntel Core i9-12900Kを使用した。学習データとしては、Market-1501と独自収集データ(1名分)を使用した。また、画像生成AIによるデータ拡張を行った場合と行わない場合での比較実験を行った。画像の生成にはGoogle DeepMindのVeo3を使用した。テキストプロンプトとして「添付画像の人物を一周回転させた動画」といった記述を与えて生成した。

## 5. 結果

## 5-1. データ拡張による本人照合の安定性評価

本研究ではデータ拡張の有効性を確かめるために本人照合性能を評価した。テストデータには別カメラで撮影した独自収集データ(1名分)とMarket-1501を使用した。結果を表1に示す。

表1 登録画像1枚の時と、画像拡張時の性能実験結果

登録方法	Threshold	TPR	FAR
1枚のみ	-0.7063	0.0000	0.0100
1枚のみ	-0.7873	0.1000	0.3984
1枚のみ	-0.8344	0.6000	0.7561
拡張画像	-0.7250	0.1500	0.0100
拡張画像	-0.7309	0.2500	0.0081
拡張画像	-0.7556	0.8000	0.0813

FAR≈1%(False Acceptance Rate)におけるTPR(True Positive Rate)が0.15となり1枚のみの場合(TPR=0)に比べてTPRが高く、本人照合性能が向上することが確認された。

## 5-2. 推論フェーズのリアルタイム性評価

GPU・TensorRT・CPUで性能実験を行った結果を表2に示す。

表2 実行環境ごとの性能実験結果

実行環境	YOLO GFLOPs	OSNet GFLOPs	平均FPS	平均レイテンシ (ms)	p95レイテンシ (ms)
GPU	3.31	1.01	7.15	135.81	187.90
TensorRT	3.31	1.01	7.20	135.31	186.00
CPU	3.31	1.01	4.47	219.68	280.73

GPUとTensorRTでは同等の性能を示し、平均FPSは約7.2、平均レイテンシは約135msであった。一方、CPUでは処理性能が大きく低下し、平均FPSは4.47、平均レイテンシは約220msと遅延が顕著に増加した。

## 6. まとめ

本研究では、リアルタイム人物再同定の実現を目指し、画像生成AIを用いた新規ビューでのデータ拡張の有効性を確認した。GPU環境では平均7.15FPS、平均レイテンシ135msを達成し、実時間応答として許容範囲内であることを確認した。ただし、フレームレートは監視カメラのリアルタイム処理には不足する可能性があり、今後はさらなる高速化が課題である。また、より多くのOpen-setデータを用いたランキング精度の検証を行うことで、実環境における実用性向上を目指す。

## 参考文献

- [1] Khawar Islam, "Deep Learning for Video-based Person Re-Identification: A Survey", arXiv :2303.11332, Oct 2024.
- [2] Shuting He, Hao Luo, et al., "TransReID: Transformer-based Object Re-Identification", arXiv, :2102.04378 Mar 2021.
- [3] Kaiyang Zhou, et al., "Omni-Scale Feature Learning for Person Re-Identification", arXiv :1905.00953, Dec 2019.