

深層学習のための 人物画像データ自動収集システム

キヤノン株式会社

木村宇任

kimura.takato@mail.canon

開発における問題点

人物認識モデルの開発において データ収集の効率化が課題

- 課題①：不要な画像を保存するための
ストレージコストが高い
- 課題②：収集した画像を学習に使用するために
膨大なラベル付けコストがかかる

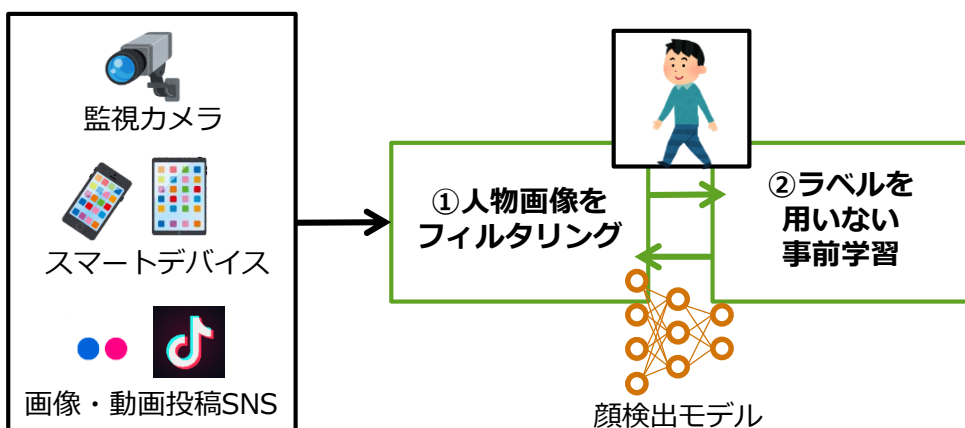
手法・ツールの適用による解決

課題を解決するための人物画像 データ自動収集システムを提案

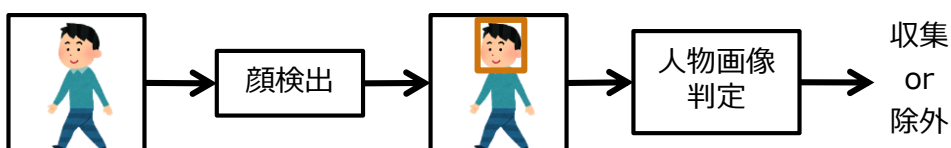
- 施策①：顔検出モデルによって人物が映って
いる画像をフィルタリング
- 施策②：自己教師あり学習によってラベルを
用いずに事前学習モデルを生成する

データ収集システム

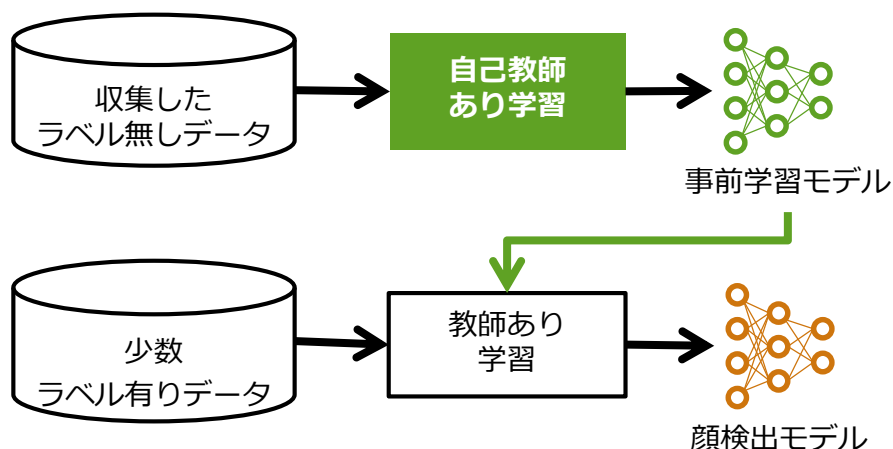
顔検出と自己教師あり学習を用いた
データ収集システムを提案



①人物画像をフィルタリング



②ラベルを用いない事前学習

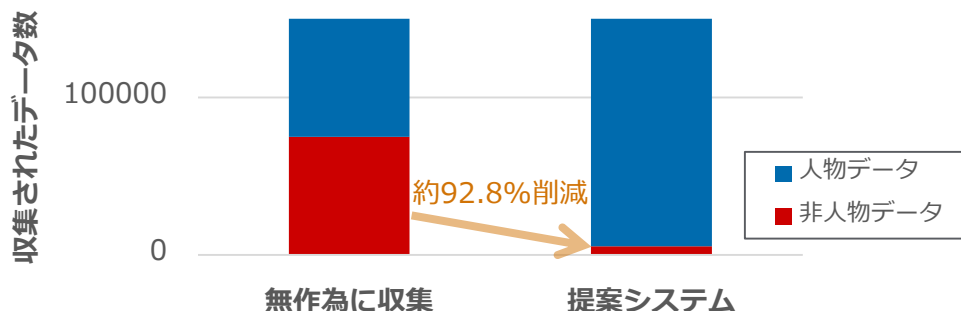


コスト削減効果

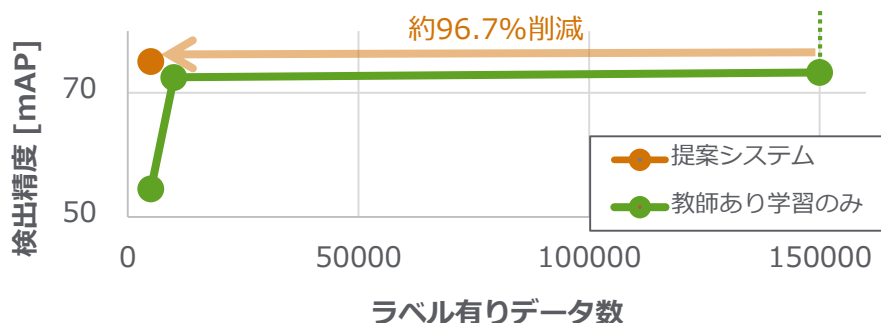
プロトタイプを実装し
課題の解決に効果があることを検証

約30万枚の公開データセットを用いて評価

不要なデータのストレージコストを約**92.8%**削減



ラベル付けコストを約**96.7%**削減



まとめ、今後の展望

人物画像データ自動収集システムを提案し、
課題の解決に効果があることを確認した

実際の収集先からのデータ収集でも
効果があることを確認する