

Person Re-identification における超解像

Super-Resolution for Person Re-identification

伊藤 天詞^{*1} 吉田 龍人^{*1} 藤井 純一郎^{*1}
Tenshi Ito Ryuto Yoshida Junichiro Fujii

^{*1}八千代エンジニアリング株式会社
Yachiyo Engineering Co.,Ltd.

Person Re-identification (ReID) において、人物とカメラが遠いときに画像が低解像度化し、同一人物の類似度が低下する。超解像モデルを取り入れて学習する手法は、人物の姿勢等の変化を含むデータセットでは有効性が示されている一方、超解像による単純な解像度差の抑制が精度向上に寄与しているか明らかでない。そこで本研究では、超解像の基礎的な有効性を検証することを目的として、人物画像のコンテンツ（姿勢、服装等）が固定された撮影距離のみが異なるデータセットを用いて、超解像適用による類似度変化を調査する。分析の結果、超解像による解像度差の抑制が ReID に有効であることを確認した。

1. はじめに

人々のインフラ利用状況をデータ化し分析するためには、異なる時間帯や異なるカメラに写る同一人物を識別する Person Re-identification (ReID) は重要な技術である。実運用では、コスト削減のため低いメモリ容量で高速に動作する ReID モデルの使用が望ましく、近年では様々な軽量モデルが公開されている [He 20, Zhou 19a]。実運用における ReID では、人物とカメラ間の距離（撮影距離）は一定ではなく、人物が遠方に位置するほど取得される人物画像は低解像度化する。これにより、解像度差があるほど同一人物画像間の ReID 特徴量の類似度が低下してしまい [吉田 23]、ReID 精度の低下に繋がる。これに対処するために、超解像 (Super-Resolution: SR) 技術を取り入れて低解像度画像を高解像度化することにより ReID 精度改善を試みる研究が行われている [Zhang 21, Jiao 18, Li 19, Wang 18b]。これら手法は、監視カメラ映像等から作成された公開データセットにおいてその有効性を示している。しかし、こうしたデータセットでは撮影距離の変化に伴い、人物の姿勢や画角、背景の変化が必然的に生じることから、SR による単純な解像度差の抑制が本質的に精度向上に寄与しているか明らかでない。

そこで本研究では、人物画像のコンテンツ（姿勢、服装、背景等）が固定された撮影距離のみが異なるデータセットを用いて、SR 適用による類似度変化の基礎的分析を行う。具体的には、SR 適用有無における ReID 特徴量のコサイン類似度行列を作成し、低解像度画像を SR することにより高解像度画像との類似度が向上するか検証する。また、SR 適用有無における特徴空間の可視化も行い、SR した画像の分布特性を分析する。これにより、SR による解像度差の抑制が ReID に有効であることを示す。

2. 関連研究

本章では、ReID と超解像について述べる。

2.1 Person Re-identification

ReID は、異なる時間帯や異なるカメラに写る同一人物を識別する技術である。深層学習を用いた ReID は、Li ら [Li 14]

連絡先: 伊藤天詞, 八千代エンジニアリング株式会社, 東京都台東区浅草橋 5-20-8 CS タワー 3F 技術開発研究所,
tn-ito@yachiyo-eng.co.jp

によって提案されて以来、高精度なモデル [He 21] だけでなく実用のための効率的で軽量なモデルも提案されている [Luo 19, Zhou 19b, Zhou 21]。Bag of Tricks (BoT) [Luo 19] は、複雑な構造を用いずに単純な ResNet50 [He 16] バックボーンに対して学習率やデータ拡張の工夫、正規化層の追加等の学習テクニックを組み合わせることで、推論時の計算効率を維持した手法である。Omni-Scale Network (OSNet) [Zhou 19b] は、様々なスケールの受容野を扱う構造を持ち、一般的な畳み込みを空間方向畳み込みとチャンネル方向畳み込みに分解して扱うことでパラメータ数を削減した手法である。Zhou ら [Zhou 21] は、OSNet に Instance Normalization (IN) を組み込むことで異なるカメラ間のスタイル差異に対処した OSNet-AIN を提案した。

2.2 超解像

SR は、入力画像の高周波成分を疑似的に復元し、高解像度化する技術である。SR モデルは、高解像度画像 x_{HR} にダウンサンプリング等の前処理を加えた低解像度画像 $x_{HR\downarrow}$ を用いて、低解像度画像 $x_{HR\downarrow}$ を SR した画像 x_{SR} と元の高解像度画像 x_{HR} との誤差が最小となるように学習する。深層学習を用いた SR は、Dong らによる畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を導入した SRCNN [Dong 14] から発展し、モデル構造別に派生している。Transformer ベース手法 [Liang 21] や Diffusion ベース手法 [Saharia 22] は高い性能を発揮するが、パラメータ数が多く処理が低速である。一方で、CNN ベース手法 [Lim 17, Chai 25] や Generative Adversarial Network (GAN) ベース手法 [Wang 18a, Wang 21] は、処理が高速であり低いメモリ容量で動作するため実用的である。GAN ベースの代表的な手法に、実世界の劣化を考慮して SR できる Real-ESRGAN [Wang 21] がある。従来は単純なダウンサンプリングのみで作成した低解像度画像を用いて学習していたが、Real-ESRGAN はブラー、リサイズ、ノイズ、JPEG 圧縮を複数回繰り返し適用することでより現実に近い劣化を模擬した低解像度画像を用いて学習する。

超解像の ReID タスクへの応用

ReID における低解像度人物画像による低精度問題を解決するために、SR を取り入れて精度改善を試みる研究がされている [Zhang 21, Jiao 18, Li 19, Wang 18b]。これら手法は、視

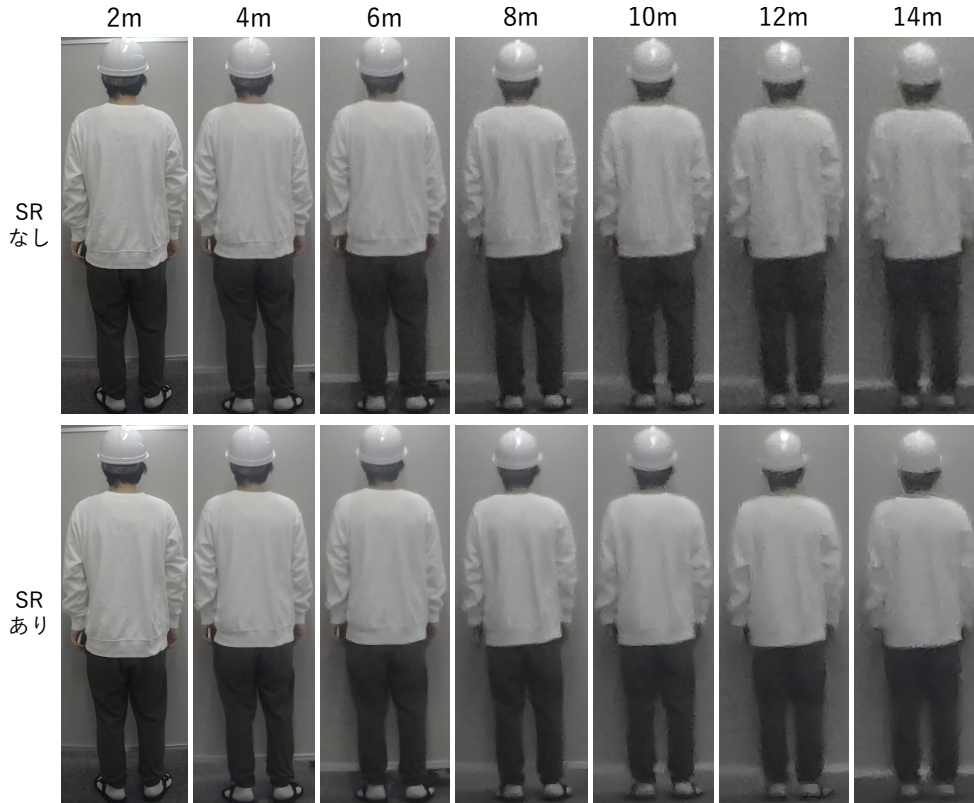


図 1: 各撮影距離における人物画像. 上段が SR なし, 下段が SR ありの人物画像である.

覚的な品質が向上するように SR した画像は ReID モデルにとって重要ではないことを指摘し, ReID 精度が向上するように SR モデル及び ReID モデルを再度学習する. 推論時においても両モデルを使用することから, 実用における導入コストが高い.

3. 実験

本章では, 人物画像のコンテンツが同じで撮影距離のみが異なるデータセット [吉田 23] を用いて, SR により解像度差異を抑制した際の類似度変化及び特徴表現を分析する. 類似度変化の分析では, SR 適用有無におけるコサイン類似度行列を作成し, 各撮影距離間での類似度の違いを可視化する. 実用性を考慮し, ReID モデルには OSNet-AIN, SR モデルには Real-ESRGAN を用いる.

3.1 データセット

吉田らのデータセット [吉田 23] は, 一人の人物を特定の条件下で撮影し作成された. このデータセットは, 撮影距離, カメラ高度, 人物の向き, 服装, 背景の違いによるデータから構成される. 本研究では, 撮影距離の違いによるデータを用いる. これは, $d \in \{2, 4, 6, 8, 10, 12, 14\}m$ のデータから構成され, 撮影距離が遠いほど低解像度な画像となる. 各撮影距離における人物画像を図 1 に示す. また, 被写体の微小な姿勢変化による影響を抑制するために各撮影距離で 5 枚の同一姿勢の同一人物画像が含まれる (解像度は多少異なる). 各撮影距離 $\{2, 4, 6, 8, 10, 12, 14\}m$ における総ピクセル数の平均は, $\{961831, 256508, 122962, 68980, 45013, 30739, 22223\}px$ である.

3.2 実験条件

Real-ESRGAN は公式リポジトリで公開されている学習済み重み (realesr-general-x4v3.pth), OSNet-AIN は torchreid [Zhou 19a] で公開されている学習済み重み (osnet-ain-ms-dc.pth) を用いる. SR モデルにより, 入力画像の縦横をそれぞれ 4 倍に高解像度化する. ReID モデルに入力される画像は, 256×128 にリサイズされる. ReID 特徴量の次元数は 512 である.

3.3 評価方法

コサイン類似度は, 学習済み ReID モデル $\mathcal{F}(\cdot)$ に人物画像 x_q, x_g を入力した際の ReID 特徴量を $\mathbf{f}_q = \mathcal{F}(x_q)$, $\mathbf{f}_g = \mathcal{F}(x_g)$ として, 式 1 で求める.

$$\text{Sim}(x_q, x_g) = \frac{\mathbf{f}_q \cdot \mathbf{f}_g}{\|\mathbf{f}_q\|_2 \|\mathbf{f}_g\|_2} \quad (1)$$

ここで, $\|\cdot\|_2$ は L_2 ノルム, $\cdot$ は内積である. 各撮影距離で 5 枚のデータがあるため, 各撮影距離間のコサイン類似度は $5 \times 5 = 25$ 通りの類似度の平均として算出する. 特徴表現は, ReID 特徴量を UMAP [McInnes 18] を用いて 2 次元に次元圧縮して可視化する.

3.4 実験結果

SR 適用後の人物画像例を図 1 下段に示す. 撮影距離 14m の人物画像に着目すると, SR により背景のノイズが抑制され, 高画質化していることが視覚的に確認できる. SR 適用有無におけるコサイン類似度行列を図 2 に示す. 図 2(e) における右上部分に着目すると, 撮影距離 10m~14m の画像を SR することで撮影距離 2m~6m のような高解像度画像との類似度が向上することが確認できる. 一方で, 下半分に着目すると SR

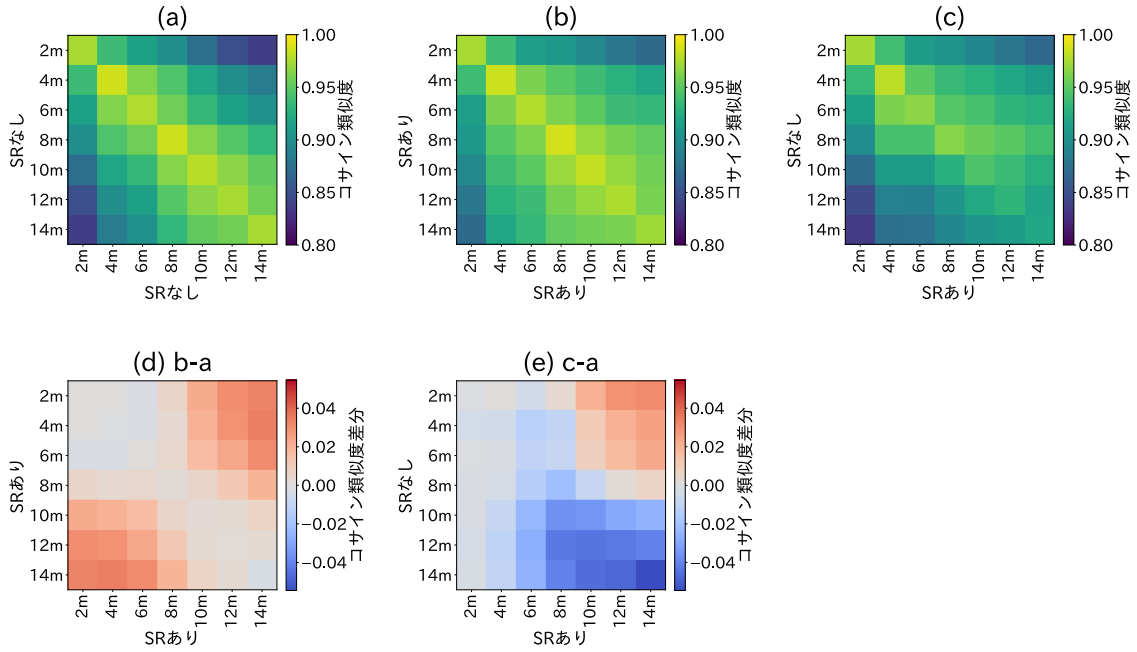


図 2: SR 適用有無におけるコサイン類似度行列. 上段は, 左から (a) 「SR なし」, (b) 「SR あり」, (c) 「縦 SR なし横 SR あり」の行列である. 下段は, 左から (d) 「b と a の差分」, (e) 「c と a の差分」である. 低解像度画像 (遠い撮影距離) を SR した際の高解像度画像 (近い撮影距離) との類似度向上は, e の上三角部分で確認できる.

により元の撮影距離 10 m～14 m の画像との類似度が低下する傾向であることがわかる. また, 撮影距離 2m のような高解像度画像は, SR を適用しても他の撮影距離の画像との類似度がほとんど変化しないことを確認した. 図 2(d) では, SR 画像同士の類似度は基本的に向上する一方で, 撮影距離 14m の SR 画像同士のよう, 多少類似度が低下する場合もあることを確認した.

SR 適用有無での各撮影距離における特徴空間の可視化結果を図 3 に示す. 撮影距離 2m～6m のような高解像度な画像は, SR しても元画像と近い位置に分布している. 一方で, 撮影距離 8m～14m のような比較的低解像度な画像は, SR することで撮影距離が近い高解像度画像の特徴量に近づくことを確認した.

以上より, SR により解像度差異を抑制することで類似度が向上する傾向にあることから, SR は単純な解像度差に対して類似度が低下してしまう ReID において有効であることを確認した.

4. 考察

図 2(e) 下半分において, SR により元の撮影距離 10 m～14 m の画像との類似度が低下する傾向であったが, これは人物画像の解像度差が増大したためであると考えられる. これとは反対に, 対角成分に着目すると, 撮影距離の近い画像は元の解像度が高いことから SR による総ピクセル数の増大幅が大きいにもかかわらず, 類似度低下は少ない. このことから, ある程度高解像度な画像を SR しても ReID 特徴量はそれほど変化しないことが示唆される. また, 図 2(d) では SR 画像同士の類似度が多少低下する場合があることが確認された. これは, SR モデルが高周波成分を復元する際に入力ごとの微細なノイズ等の違いにより異なる細部を生成してしまい, それが ReID モデルに悪影響を与えたことが原因であると考えられる.

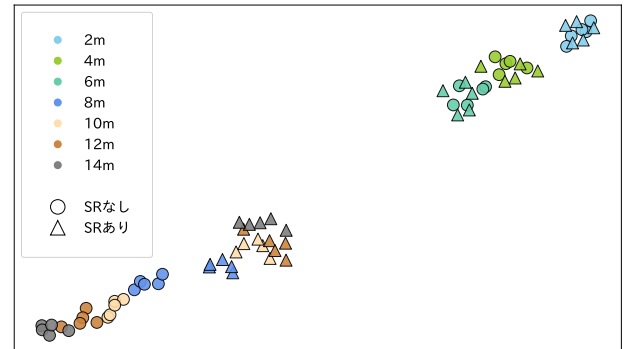


図 3: SR 適用有無での各撮影距離における特徴空間の可視化. 各特徴量は UMAP により 2 次元に次元削減されている.

使用したデータセットは一人の人物のみで構成されていることから, 本結果の一般性については限定的である. より多様なデータを用いた分析は今後の課題とする.

5. おわりに

本研究では, 人物画像のコンテンツが同じで撮影距離のみが異なるデータセットを用いて, SR により解像度差異を抑制した際の ReID における有効性の検証を行った. 実験では, 撮影距離が遠い低解像度画像を SR することにより, 撮影距離が近い高解像度画像との類似度が向上することを確認した. 今後は, 本知見の一般性を確認するために異なるアーキテクチャを持つ SR モデルや ReID モデルを用いた比較検証を行う予定である.

参考文献

- [Chai 25] Chai, X., Zhang, Y., Zhang, Y., Cheng, Z., Qin, Y., Yang, Y., and Song, L.: Distillation-Supervised Convolutional Low-Rank Adaptation for Efficient Image Super-Resolution, in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1431–1440 (2025)
- [Dong 14] Dong, C., Loy, C. C., He, K., and Tang, X.: Learning a deep convolutional network for image super-resolution, in *European conference on computer vision*, pp. 184–199 Springer (2014)
- [He 16] He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 (2016)
- [He 20] He, L., Liao, X., Liu, W., Liu, X., Cheng, P., and Mei, T.: FastReID: A Pytorch Toolbox for General Instance Re-identification, *arXiv preprint arXiv:2006.02631* (2020)
- [He 21] He, S., Luo, H., Wang, P., Wang, F., Li, H., and Jiang, W.: Transreid: Transformer-based object re-identification, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 15013–15022 (2021)
- [Jiao 18] Jiao, J., Zheng, W.-S., Wu, A., Zhu, X., and Gong, S.: Deep low-resolution person re-identification, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32 (2018)
- [Li 14] Li, W., Zhao, R., Xiao, T., and Wang, X.: Deep-reid: Deep filter pairing neural network for person re-identification, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 152–159 (2014)
- [Li 19] Li, Y.-J., Chen, Y.-C., Lin, Y.-Y., Du, X., and Wang, Y.-C. F.: Recover and identify: A generative dual model for cross-resolution person re-identification, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8090–8099 (2019)
- [Liang 21] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R.: Swinir: Image restoration using swin transformer, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844 (2021)
- [Lim 17] Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K.: Enhanced deep residual networks for single image super-resolution, in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144 (2017)
- [Luo 19] Luo, H., Gu, Y., Liao, X., Lai, S., and Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 0–0 (2019)
- [McInnes 18] McInnes, L., Healy, J., and Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction, *arXiv preprint arXiv:1802.03426* (2018)
- [Saharia 22] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M.: Image super-resolution via iterative refinement, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 45, No. 4, pp. 4713–4726 (2022)
- [Wang 18a] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks, in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0 (2018)
- [Wang 18b] Wang, Z., Ye, M., Yang, F., Bai, X., and 0001, S. S.: Cascaded SR-GAN for scale-adaptive low resolution person re-identification., in *IJCAI*, Vol. 1, p. 4 (2018)
- [Wang 21] Wang, X., Xie, L., Dong, C., and Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1905–1914 (2021)
- [Zhang 21] Zhang, G., Ge, Y., Dong, Z., Wang, H., Zheng, Y., and Chen, S.: Deep high-resolution representation learning for cross-resolution person re-identification, *IEEE Transactions on Image processing*, Vol. 30, pp. 8913–8925 (2021)
- [Zhou 19a] Zhou, K. and Xiang, T.: Torchreid: A Library for Deep Learning Person Re-Identification in Pytorch, *arXiv preprint arXiv:1910.10093* (2019)
- [Zhou 19b] Zhou, K., Yang, Y., Cavallaro, A., and Xiang, T.: Omni-scale feature learning for person re-identification, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3702–3712 (2019)
- [Zhou 21] Zhou, K., Yang, Y., Cavallaro, A., and Xiang, T.: Learning generalisable omni-scale representations for person re-identification, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 44, No. 9, pp. 5056–5069 (2021)
- [吉田 23] 吉田龍人, 大久保順一, 藤井純一郎, 高森秀司, 天方匡純: Person Re-identification の実用に向けた DNN モデルの特性評価, *AI・データサイエンス論文集*, Vol. 4, No. 3, pp. 385–392 (2023)