

ドメイン適応を用いたVision Transformerによる 鍛造部品の不良品検出

Detection of Defective Forged Parts Using Domain Adaptation with Vision Transformer

高木 裕也* 藤田 和弘* 中川 真言† 世継 武志†
Yuya TAKAGI Kazuhiro FUJITA Makoto NAKAGAWA Takeshi YOTSUGI

1 はじめに

現在、鍛造部品の外観検査の多くは人間による目視で行われている。そのため、検査員の違いによる検査精度のバラツキや、人手不足といった課題がある。そこで、画像検査による外観検査の自動化によって、これらの課題を解決することを目的としてVision Transformer(ViT)を用いた鍛造部品の不良品検出について研究を行った。その結果、撮影時期による環境変化や部品の表面テクスチャの変化によって識別率が低下することがわかった[1]。本研究では、ドメイン適応[2]を用いて、異なる時期に撮影された鍛造部品の画像データ間の特徴量のギャップを小さくすることで、学習データとは異なる撮影時期のデータでも識別精度低下を抑える方法を検討し、従来手法との比較実験を行った。

2 データセット

本研究では、高橋金属(株)提供の鍛造部品の底面部と側面部が写っている画像をデータセットとする。側面部について傷や欠け、打痕が見られる不良品画像と、良品画像でラベル付けされている。用いるデータセットの撮影時期とラベル、枚数を、表1に示す。

表 1: 本研究のデータセット

撮影時期	ラベル	枚数
2023/01	不良品	1000
2023/02	良品	1000
2023/03	不良品	744
2023/04	良品	1000
2023/05	良品	1000
2023/05	不良品	595

撮影時期が2023年1月から2023年4月までのデータは、学習用データ2,048枚、検証用データ256枚、評価用データ1,440枚に分割した。また、2023年5月に撮影された良品と不良品のデータは、学習していない撮影時期のデータの識別精度について評価するために用いた。画像の例を図1に

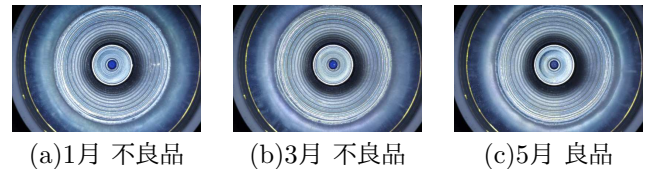


図 1: サンプルデータ

示す。

3 手法

データセットの不良品画像は側面部(ネジ部)についての欠陥だけなので、原画像から側面部のみが写った画像に変換する。原画像を鍛造部品の中心を原点として極座標変換を行い、横軸を角度、縦軸を動径とした278×256画素にリサイズし、上部150画素を除去することで、128×256画素の画像にする。変換後の不良品画像の例を図2に示す。

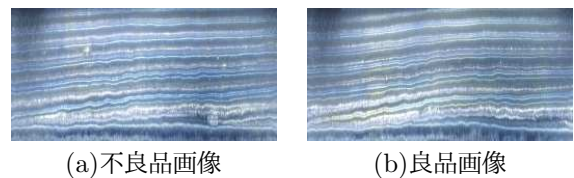


図 2: 変換後の画像

学習用データ2,048枚を用いて、ドメイン適応を用いた学習を行うViT (ViT_DA)と、ドメイン適応を用いずに学習を行うViT、畳み込みニューラルネットワーク (CNN) の3種類の学習モデルを作成する。CNNモデルにはSEResNetを採用した。ViT_DAのモデル構造を図3に示す。ViT_DAの構造は、入力画像の特徴量を出力するViT、良品および不良品のラベル識別用の全結合層、撮影時期のドメイン識別用の全結合層で構成されている。また、本研究の学習用データは、撮影時期でラベルが統一されているため、撮影時期を識別することがラベルを識別することと同義になってしまう。そこで、ドメイン識別用の全結合層を良品画像用と不良品画像用の2種類用意することで、ラベルごとにドメインを識別する構造にした。

*龍谷大学, Ryukoku University

†高橋金属株式会社, Takahashi Metal Industries Co., Ltd.

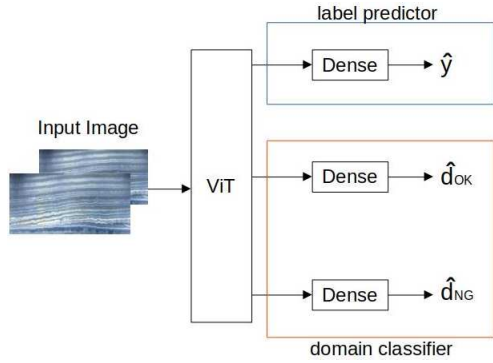


図 3: ドメイン適応を用いて学習するViT

良品画像用ドメイン識別層の損失関数 $L_{d,ok}$ と、不良品画像用ドメイン識別層の損失関数 $L_{d,ng}$ は以下のように定義する。

$$L_{d,ok} = \sum_i d_i^{ok} \log(\hat{d}_i^{ok}) \quad (1)$$

$$L_{d,ng} = \sum_i d_i^{ng} \log(\hat{d}_i^{ng}) \quad (2)$$

ここで、 d_i^{ok} は良品画像の正解ドメイン、 d_i^{ng} は不良品画像の正解ドメイン、 \hat{d}_i^{ok} は良品画像の予測ドメイン、 \hat{d}_i^{ng} は不良品画像の予測ドメインである。

ViTとラベル識別層の損失関数 L_y は以下のように定義する。

$$L_y = \lambda \sum_i y_i \log(\hat{y}_i) - (1 - \lambda)(L_{d,ok} + L_{d,ng}) \quad (3)$$

ここで、 y_i は正解ラベル、 \hat{y}_i は予測ラベル、 λ はドメイン損失の影響を調整するハイパーパラメータである。

4 識別実験結果

作成した学習済みモデルを用いて、良品画像および不良品画像に対して識別を行った。評価用データおよび、2023年5月撮影のデータに対する識別のROC曲線を図4に示す。図4より、学習データと同時期に撮影された評価用データに対しては、AUCの値が0.98から1.0となっており、どのモデルにおいても識別精度が高いことが分かる。一方で、2023年5月に撮影した時期の異なるデータに関しては、全てのモデルにおいて、評価用データに対する性能よりも低くなっている。しかし、ViT_DAのAUCが0.91と他のモデルと比べて高く、本手法が時期の異なるデータに対しても比較的高い性能を維持していることが分かる。

ViT_DAとViTに対して、t-SNEで2次元に次元削減した特徴量の分布を図5に示す。図5より、ドメイン適応を

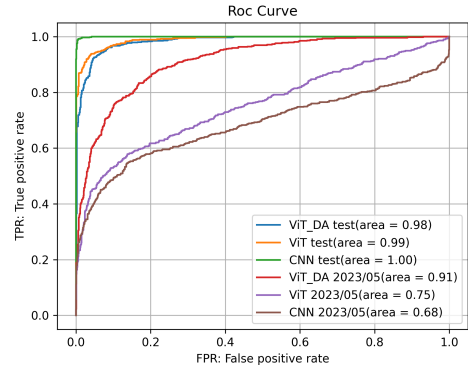


図 4: ROC曲線

用いていないViTは、学習用データと、学習用データとは異なる撮影時期のデータの特徴量が分離されている。一方ViT_DAは、学習用データと2023年5月撮影のデータの分布が重なりあっていることが分かる。

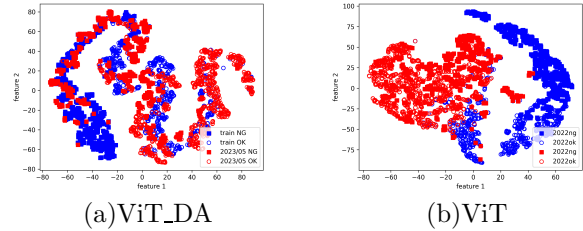


図 5: 特徴量の分布

5 まとめ

本研究では、鍛造部品の側面部について、ドメイン適応を用いたViTによって、撮影時期による環境変化や部品の表面テクスチャの変化による識別率低下を抑えることを試みた。その結果、学習データと学習データとは異なる撮影時期のデータ間の特徴量のギャップを小さくすることができ、従来手法と比べてAUC値の低下を抑制することができた。本稿では撮影時期をドメインとして用いたが、製造時期をドメインとして用いて学習することも今後検討したい。

参考文献

- [1] 高木 裕也, 藤田 和弘, 中川 真言, 世継 武志: “ Vision Transformer を用いた鍛造部品の不良品検出 ”, FIT2023 (2023)
- [2] Li, D., Yang, Y., Song, Y.-Z., and Hospedales, T. M.: “ Deeper, broader and artier domain generalization ”, IEEE International Conference on Computer Vision, pp. 5542-5550 (2017)