

# 工業製品の画像拡張への生成系 AI Stable Diffusion の応用と VAE モデリングツールの開発

F124612 Zheng Zhelin

永田研究室

## 1. 目的

近年、少子高齢化に伴う労働力不足が深刻化する製造業においては、検査工程の自動化と高精度化が望まれている。しかし、製造現場では多品種少量生産体制が進展し、不良品の発生頻度の低さから訓練データを十分に確保できないことが少なくない。深層学習において、CNNなどの学習モデルに不均衡なデータが与えられる場合、訓練アルゴリズムは多数派のカテゴリに多くの時間を費やすため、結果的に少数派のカテゴリの学習に割かれる時間が少なくなるという傾向がある。このため、訓練後のモデルも少数派のカテゴリを多数派のカテゴリとして誤分類してしまう可能性が高まる。本研究では、不良品の発生頻度の低い工業材料を対象とし、局所領域に存在する特徴の生成が可能となった生成系AIモデルStable Diffusionを応用して欠陥が含まれた画像の拡張を行った。また、代表的生成系AIの一つであり、Stable Diffusionにも内蔵されているVariational Autoencoder (VAE) を構築・解析・改良できるVAEモデリングツールを開発した。さらに、Stable DiffusionとVAEで生成された画像を用いることで欠陥検出用CNNモデルの汎化性能の向上を試みたので報告する。

## 2. 実験方法

対象製品の元画像のサイズは $2590 \times 1942$ ピクセルであったが、Stable Diffusionでは内部処理上の制約により、8の倍数の画像サイズのみ生成が可能となっている。このため、対象となる材料部分(Target area)を変更範囲と指定し、この領域が含まれる $1288 \times 1288$ ピクセル(Cropping area)をオリジナル画像から切り取り、この対象領域での画像拡張を行った。次に、画像自体だけではなく、潜在空間を持つ生成モデルの理論的基盤であるVAEに着目した。Kingmaらによって提案されたVAEモデルは、潜在空間からデータを再構成・生成できる利点から、異常検知、特徴抽出、データ圧縮など幅広い分野で利用されている。一方で、VAEの内部学習過程、特にKLD損失とMSE損失のバランス調整は複雑であり、体系的な理解は容易ではない。本研究では、MATLAB上でVAEモデルを開発し、KLD損失とMSE損失それぞれに基づく学習過程を独立に解析できるようにした。

VAEはエンコーダおよびデコーダのネットワークから構成されるため、これら2つのネットワークを訓練する必要がある。従来の学習法では、KLD損失とMSE損失からなる全損失 $L$ でエンコーダ部とデコーダ部を総合的に学習させていた。しかしながら、この場合、デコーダ部が潜在変数 $z$ を無視してしまうようなPosterior Collapseの現象を避けながら、これら2つの損失を収束させていくことが困難な場合に直面することがあった。この課題に対して、本研究では、図1に示す3つの勾配 $E1$ ,  $E2$ , 及び $D$ を独立に用いて、エンコーダ部、エンコーダ部、及びデコーダ部の学習可能な重みパラメータをそれぞれ更新する方法を提案した。勾配 $E1$ はKLD損失のみを、また、勾配 $E2$ と $D$ はMSE損失のみを用いて自動微分により算出している。

## 3. 結果と考察

実験で用いたオリジナル画像、Stable Diffusionで再構築された画像、それら2つの画像の差分画像の例を図2に示す。外枠のオレンジ色で囲まれた部分はオリジナル画像からカットされた領域であり、赤色で囲まれた部分は対象となる材料部分を含む目標処理領域である。また、オリジナル画像に近い画素値のところは白っぽく、異なるところは黒っぽくなるようにグレー階調の差分図を出力している。この結果から、Stable Diffusionでは範囲指定されたエリアのみに画像を再構築できることが確認された。

また、図3には設計したVAEモデルを用いて拡張した画像の例を示す。(a)は訓練に用いている25枚のオリジナル画像の中の1枚である。

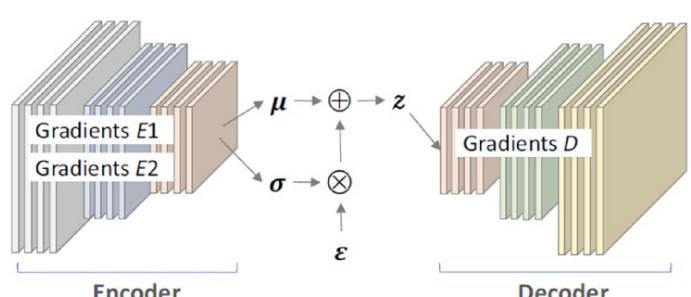


Fig. 1 Three gradients for updating learnable weight parameters in encoder and decoder networks of VAE.

(b)は従来の学習法で訓練したVAEモデルから生成された画像例であり、(c)は本研究で提案した学習法で訓練したVAEモデルから生成された画像例である。開発したVAEモデルを使って可視化した訓練プロセスから考察すると、従来法ではMSE損失とKLD損失を同時に使いながらそれらを同時に収束させていくため、「KLD損失を目指値まで収束できてもMSE損失がまだ大きく、生成される画像のリアル感が低い」、「MSE損失を目指値まで収束できてもKLD損失が小さくなりすぎてPosterior Collapse現象が発生してしまう」といった課題があった。

一方、提案手法ではKLD損失とMSE損失を独立に制御することで、2つの損失を同時に目標値まで収束することが可能となった。その結果、図3(c)のように、提案手法で訓練したVAEモデルは訓練データと同じような欠陥の特徴を持ちながら、訓練データには含まれない新たな画像を生成できていることが確認された。特に、オリジナル画像とは異なった部位にリアルな欠陥が生成できていることが観察された。また、Stable DiffusionとVAEそれぞれを用いて生成させた画像をCNNモデルの訓練に用いることで、両方のCNNモデルの分類精度を改善できることも確認された。さらに、提案手法で訓練したVAEモデルについては、入力画像と出力画像の差分から算出した平均異常スコアを評価指標として、教示なし学習による異常検知にも適用できることが示された。

## 学会発表

- 1) 鄭, 永田, Stable Diffusion を用いた画像拡張による工業材料の欠陥検出用CNNモデルの性能改善, 第31回インテリジェントシステムシンポジウム FAN2023 講演論文集, Th-A3-3(1-3), 九州大学椎木講堂, 2023.
- 2) 鄭, 永田, 叶谷, 渡辺, 工業材料の欠陥検出用CNNモデルの性能改善のためのStable Diffusionの応用, 第30回画像センシングシンポジウム(SSII2024), IS1-12, 5 pages, 2024.
- 3) Z. Zhelin, N. Fusaomi, K. Hirohisa, W. Keigo, and K. H. Maki, Variational Autoencoder Modeler for Systematically Analyzing Training Process, *Procs. of 2024 63rd Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 748–751, 2024.
- 4) Z. Zhelin, N. Fusaomi, O. Akimasa, K. Hirohisa, W. Keigo, K. H. Maki and S. A. G. Ahmad, Application of Variational Autoencoder Modeler for Systematically Analyzing Training Process –Its Application to Image Generation and Anomaly Detection–, *Procs. of The 30th International Symposium on Artificial Life and Robotics (AROB 30th 2025)*, pp. 1123–1128, 2025.
- 5) Z. Zhelin, N. Fusaomi, T. Hisami, N. Hitoshi, K. Hirohisa, O. Akimasa, I. Takeshi, W. Keigo, K. H. Maki, S. A. G. Ahmad, Modeling Support System of Variational Autoencoder for Image Augmentation and Anomaly Detection, *Procs. of 18th International Congress on Advanced Applied Informatics*, pp. 1–7, July 13–19, 2025.
- 6) Z. Zhelin, N. Fusaomi, O. Akimasa, K. Hirohisa, W. Keigo, K. H. Maki and S. A. G. Ahmad, Systematic Evaluation of Variational Autoencoder Training Process with Independent Control of Encoder and Decoder, *Procs. of The 31st International Symposium on Artificial Life and Robotics (AROB 31st 2026)*, 6 pages, 2026.

## Abstract

In manufacturing environments where it is difficult to obtain real defect samples due to low occurrence frequency, the resulting severe data imbalance often causes deep learning models to overlook minority classes and misclassify them as majority ones. To address this issue, in this study, we focused on industrial materials in which defective samples occur only infrequently, and we employed Stable Diffusion—a generative AI model capable of synthesizing localized defects—to augment images containing such realistic defects. We also constructed, analyzed, and improved a Variational Autoencoder (VAE) using our developed VAE modeler, which is one of the powerful deep learning models for generative AI and is incorporated in Stable Diffusion. Furthermore, using images generated by both Stable Diffusion and a VAE model, we evaluated and confirmed the enhancement of the generalization performance of a convolutional neural network (CNN) model designed for defect detection.

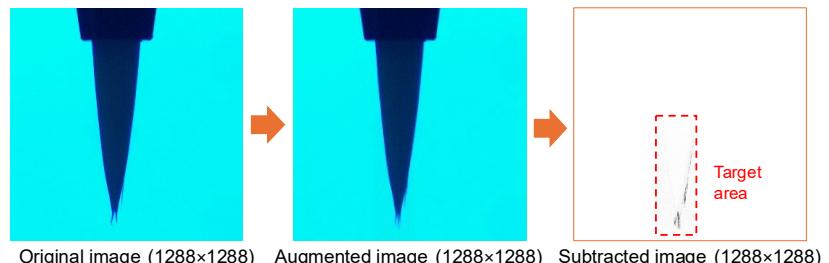


Fig. 2 Example of subtraction between original image and augmented image by Stable Diffusion.

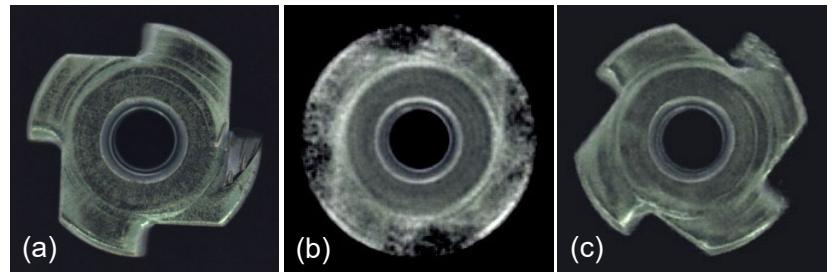


Fig. 3 Examples of (a) original image (a), (b) generated image by conventional VAE, and (c) generated image by proposed VAE.