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

Zhelin Zheng† 永田 寅臣† 叶谷 相馬† 渡辺 桂吾‡
†山口東京理科大学 ‡岡山大学
E-mail: nagata@rs.socu.ac.jp

1 背景と目的

深層学習においては最適化アルゴリズムが不均衡な量のデータを処理する場合，多数派のカテゴリにほとんどの時間が費やされ，少数派のカテゴリの学習時間が少なくなってしまうという傾向がある．このため，この問題を放置しておくと学習後のモデルは少数カテゴリを多数カテゴリとして誤分類してしまう可能性が高まってしまう．カテゴリの不均衡は，環境，生活，ビジネスなど多くの領域で遭遇するため非常に重要な問題であり，場合によっては，バランスの取れたカテゴリ分布を前提とする標準的な学習手法の性能に対する明らかなボトルネックとなっている [1].

工業製品によっては欠陥や不良品の発生頻度が非常に低いためにそれらの画像収集が容易ではなく，結果的に，畳み込みニューラルネットワーク (CNN) を用いた欠陥検出システムを構築しようとした場合，汎化性の高いシステムの実現が困難となる場合が少なくない．対象となる欠陥が含まれた限られた量の画像を拡張する (増やす) ことを目的に敵対生成ネットワーク (Generative Adversarial Network: GAN) [2] が適用された報告があるものの，画像内の特定領域に注目し，その領域内に限定して類似した特徴を持つ領域画像を生成させるという機能は有していないようである．一方，Stable Diffusion [3, 4] を応用することで，オリジナル画像内のターゲットとなる材料部分の位置，姿勢，色彩を維持したまま，新たなプロンプトに従って異なる特徴を持つ欠陥の領域画像を生成させ，材料部分に描き足すことが可能となる.

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，Stable Diffusion のインペインティング機能を応用して欠陥が含まれた画像の拡張を行うことで，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 実験内容

研究室では工業製品の不良品検出のために CNN, SVM, CAE, FCN, YOLO, FCDD, PatchCore, FastFlow などの深層学習モデルを効率的に設計，訓練，評価できるアプリケーションを MATLAB の AppDesigner 上で開発している [5, 6]．CNN モデルについては，これまでに工業製品，工業材料，培養細胞，金属の火花試験などの写真の分類問題への適用実験を行ってきた中で，VGG19 の転移学習ベースのモデルが常に高い分類精度を発揮できていたために，今回も同様の転移学習による設計方法で CNN モデルを構築することとした.

2.1 Stable Diffusion

Rombach らによって提案された Stable Diffusion は，Diffusion Model (拡散モデル) をベースとした Text-to-Image の画像生成モデルであり，VAE (Variational Auto Encoder) でピクセル画像を潜在空間表現に変換することでモデルの軽量化が図られるとともに，拡散モデルのバックボーンである U-Net を用いた画像生成の条件づけに Text Encoder である Transformer が使用されている [3]．図 1 は Rombach らが提案した潜在拡散モデルを示している．このモデルは大きく 3 つの領域に分けられ，それぞれ Pixel Space, Latent Space, Conditioning と呼ばれる．図左側の Pixel Space の部分を構成する要素は VAE と呼ばれる変分オートエンコーダであり，この部分には画像から潜在変数への変換と潜在変数から画像への変換の役割がある．潜在空間は，VAE エンコーダによって得られた潜在変数を処理する部分であり，この部分のメインは U-Net である．ここで使用されている U-Net は，一般的な U-Net と比較して，“cross-attention”という形で外部条件を与えることができるという利点がある．図右側の Conditioning 領域では，テキストデータをベクトルデータに変換できるテキストエンコーダを用いている．ユーザがテキストで指示を出すと，そのテキストは Conditioning 部のテキストエンコーダによってベクトルに変換され，ベクトル形式で生成画像に指示を与えることができる．なお，Conditioning 部への入力はテキストだけでなく，Images や Semantic Map, Representations なども含まれる．

Jonathan らによって提案された Denoising Diffusion Probabilistic Models (DDPM) [7] は，訓練部分とサン

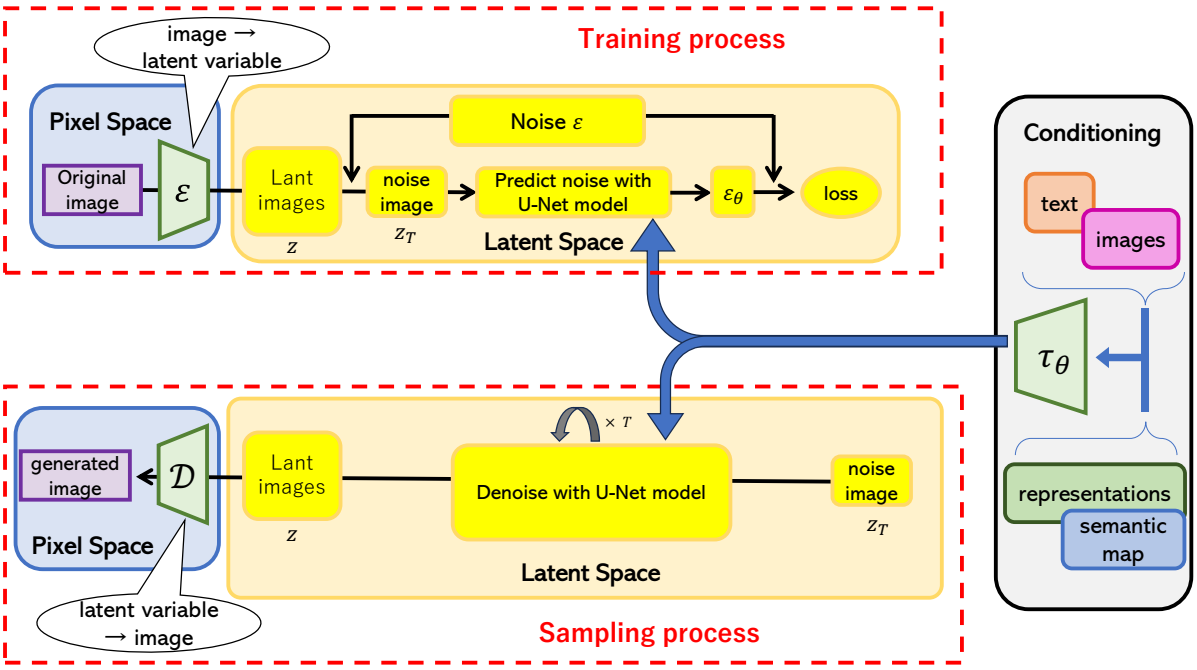

図 1　Structure of Latent Diffusion Model [3].

プリング部分に分けられており，この DDPM を以下の 3 つのサブセクションで検証した報告がある [8, 9].

2.1.1　Stable Diffusion の拡散過程

拡散過程では，原画像 x_0 にガウスノイズを徐々に加え，$x_1, ..., x_T$ 状態の画像を生成する．ここで，x_T は純粋なガウスノイズの状態である．この処理の方程式は次式で与えられる．

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon_t, \quad t = 1, 2, 3, ..., T \tag{1}$$

ここで，ϵ_t は標準正規分布に従うノイズである．$\beta_t \in (0, 1)$ は離散時間 t で加えるノイズの強さを表し，予め決められたパラメータである．信号を $\sqrt{1-\beta_t}$ 倍に減衰したあとにノイズを加える方法を採用すると，次式のように書き換えることができる．

$$q(x_t|x_{t-1}) := N(x_t; \sqrt{1-\beta_T}x_{t-1}, \beta_T I) \tag{2}$$

拡散過程に対して漸化式 (1) を繰り返し用いると，任意の時刻 t における x_t は，純粋ガウス雑音の線形結合を x_0 に加えることによって得られることがわかる．つまり，x_t は x_0 に線形結合 $\epsilon_1, ..., \epsilon_T$ を加えたものである．正規分布の和も正規分布に従うので，純粋ガウス雑音の線形結合 $\epsilon_1, ..., \epsilon_T$ は正規分布に従い，次式のようにまとめることができる．

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon \tag{3}$$

$$q(x_t|x_0) := N(x_t; \sqrt{\overline{\alpha}_t}x_0, \sqrt{1-\overline{\alpha}_t}I) \tag{4}$$

ここで，$\alpha_t = 1 - \beta_T$, $\overline{\alpha}_t = \prod_{t=1}^{T} \alpha_t$, $\epsilon = \prod_{t=1}^{T} \epsilon_t$ である．

2.1.2　Stable Diffusion の逆拡散過程

生成モデルの主役である逆拡散過程の遷移確率 $q(x_{t-1}, x_t)$ そのものを得ることは難しいが，ノイズ付与前の画像 x_0 で条件づけた場合の遷移確率 $q(x_{t-1}|x_t, x_0)$ は，拡散過程で求めた式を利用すれば具体的な表式を得ることが可能となる．順プロセスの結果を使うために，逆プロセスの遷移確率はベイズの定理を使って次のように表される．

$$q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1}, x_0)q(x_{t-1}|x_0)}{q(x_t|x_0)} \tag{5}$$

ここで，

$$q(x_{t-1}|x_t, x_0) \propto \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \tag{6}$$

であり，式 (5) の右辺に現れる 3 つの確率分布は全て拡散過程で現れるもので，これらの具体的な表式は式 (2) と (4) から得られる．さらに，規格化定数の部分を無視すれば次式のようになる．

$$q(x_{t-1}|x_t, x_0) \propto \exp\left[-\frac{1}{2}\frac{(x_t - \sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_T}\right] \tag{7}$$

$$q(x_{t-1}|x_0) \propto \exp\left[-\frac{1}{2}\frac{(x_{t-1} - \sqrt{\overline{\alpha}_{t-1}}x_0)^2}{1-\overline{\alpha}_{t-1}}\right] \tag{8}$$

$$q(x_t|x_0) \propto \exp\left[-\frac{1}{2}\frac{(x_t-\sqrt{\overline{\alpha}_t}x_0)^2}{1-\overline{\alpha}_t}\right] \tag{9}$$

であり，式 (7), (8), (9) を式 (5) に代入し，式 (6) と連立すると，$q(x_{t-1}|x_t, x_0)$ の平均 $\tilde{\mu}_t(x_t, x_0)$ と分散 σ^2 は次式から計算できる．

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}x_t + \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}x_0 \tag{10}$$

$$\sigma^2 = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t \tag{11}$$

つまり，$q(x_{t-1}|x_t, x_0) = N(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \sigma^2 I)$ と書き換えることができる．式 (3) を使えば x_t と x_0 の関係を知ることができ，さらにこの関係を式 (10) に関連付ければ，次のような平均値が得られる．

$$\tilde{\mu}_t(x_t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon\right) \tag{12}$$

また，x_t はモデルの入力とする場合，予測した平均値は次のように定義することができる．

$$\mu(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) \tag{13}$$

ここで，$\epsilon_\theta(x_t, t)$ は x_t から ϵ を予測するための近似値であり，予測値と実際加えた値の誤差は次のセクションで求められる．また，$\epsilon_\theta(x_t, t)$ を使うと x_{t-1} は次のように計算できる．

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z \tag{14}$$

ただし，z は標準正規分布に従う．

2.1.3 Stable Diffusion の損失関数

拡散モデル [10] は，次式で与えられる形式の潜在変数モデルである．

$$p_\theta(x_0) := \int p_\theta(x_{0:T})dx_{1:T} \tag{15}$$

ここで，$x_1, ..., x_T$ は，データ $x_0 \sim q(x_0)$ と同じ次元を持つ潜在変数である．結合分布 $p_\theta(x_{0:T})$ は逆プロセスと呼ばれ，$p(x_T) = N(x_T; 0, I)$ で始まり，学習されたガウス遷移を持つマルコフ連鎖として定義される．

$$p_\theta(x_{0:T}) := p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{16}$$

$$p_\theta(x_{t-1}|x_t) := N(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \tag{17}$$

学習は，次のように負の対数尤度に関する通常の変分境界を最適化することによって実行される．

$$\begin{aligned}
&E\left[-\log p_\theta(x_0)\right] \\
&= -\int dx_0 q(x_0)\log p_\theta(x_0) \\
&\leq -\int dx_0 q(x_0)\left(\int q(x_{1:T}|x_0)\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}dx_{1:T}\right) \\
&= E_{q(x_{0:T})}\left[-\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right] \\
&= E_q\Bigg[D_{KL}\underbrace{(q(x_T|x_0)||p_\theta(x_T))}_{L_T} \\
&+ \prod_{t=2}^{T} D_{KL}\underbrace{(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}_{L_{t-1}} \\
&- \underbrace{\log p_\theta(x_0|x_1)}_{L_0}\Bigg] \\
&= L_T + L_{T-1} + ... + L_0
\end{aligned} \tag{18}$$

ここで，$q(x_T|x_0)$ は学習可能なパラメータを持たず，$p_\theta(x_T)$ は単なるガウスノイズ確率であるため，L_T は学習中は定数となり，無視できる．また，L_0 は最後のノイズ除去ステップの再構成損失であり，これも学習過程では無視できる．式 (12) と (13) を L_{T-1} に代入すると，平均二乗誤差 (MSE) を用いた目標平均と近似値の損失は以下のように計算される．

$$\begin{aligned}
L_{T-1} =& E_{x_0,\epsilon}\left[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(x_t) - \mu(x_t, t)||^2\right] \\
=& E_{x_0,\epsilon}\Big[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\overline{\alpha}_t)}||\epsilon \\
&- \epsilon_\theta(x_t, t)(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, t)||^2\Big]
\end{aligned} \tag{19}$$

2.2 Stable Diffusion を用いた領域画像の拡張

実験では，ブラウザを通して Stable Diffusion の機能を利用できる Stable Diffusion Web UI (SDWU) というツールを用いた．Stable Diffusion には 2 つのタイプがあり，Text-to-Image と Image-to-Image の機能を利用できる．本研究では，限られた数のオリジナルの製品画像から Stable Diffusion で拡張した画像の有用性を検証するために，画像から画像を生成させる機能 Image-to-Image, すなわちサンプリングスクリプトを使用した．オリジナル画像の解像度は 2590×1942 であるが，Web UI の内部処理上の制約により，8 の倍数の画像しか生成できないため，図 2 のようにオリジナ画像からターゲットとなる素材部分が含まれるように 1288×1288 (1288 = 8×161) の領域を切り出し，Cropping 画像と定義した．

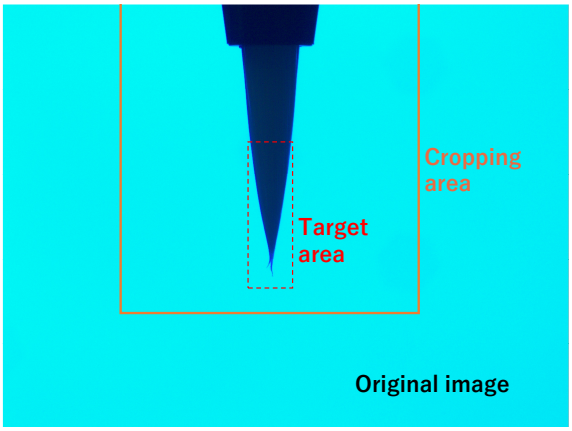

図 2　Cropping area and target area specified in an original image.

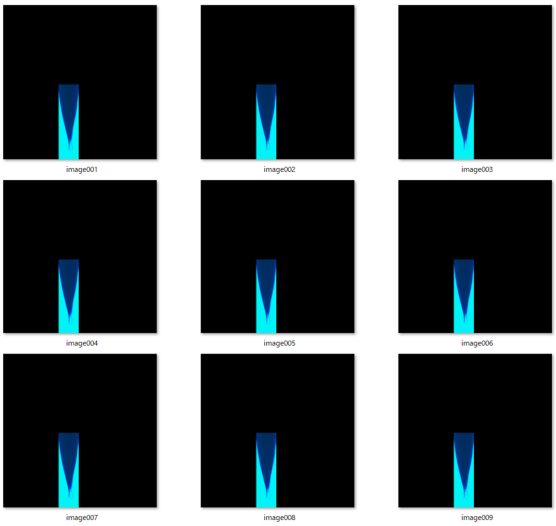

図 3　Target material areas with defects to be augmented by Stable Diffusion Web UI.

図 3 のように描画生成領域を指定することで局所サンプリングができるため，例えば，特定の欠陥の特徴をもとに再構成したい画像の生成が可能になる．Web UI で利用できるサンプリング法は 31 種類ほどあるが，今回は生成の安定さ及び画像の再現性を求めるために “DPM++ 2M Karras” を用いて，Steps: 40, Denoising: 0.2 で良品画像を生成し，Steps: 40, Denoising: 0.4 で不良品画像を生成した．図 4 はオリジナルの不良品画像であり，図 5 には新たに生成した不良品画像の例を示す．なお，図 4 の画像には，今回使用した工業材料特有の細かな欠陥が含まれているが，生成された図 5 の画像内においても本物の欠陥のような特徴を観察することができた．以上の実験結果から，実際の生産ラインにおける発生頻度の低い工業製品や工業材料の画像拡張にも有効であるものと期待される．

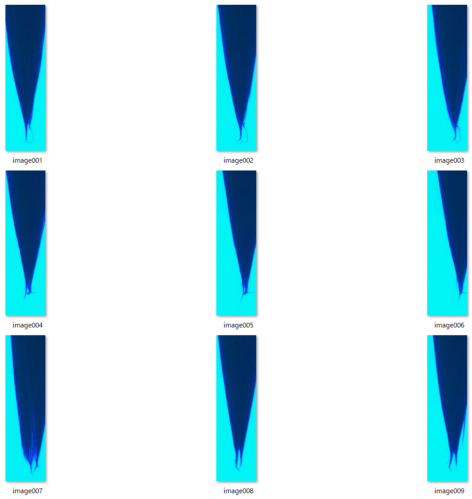

図 4　Original image samples of an industrial material with defects.

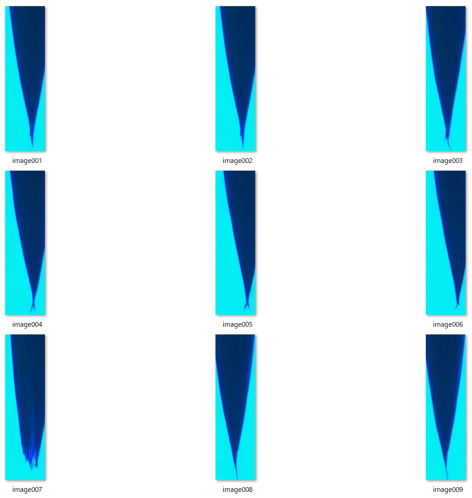

図 5　Images of an industrial material with defects augmented using Stable Diffusion Web UI.

もう一つの例として，図 6 にはオリジナル画像と拡張された画像との差分を示しているが，SDWU はターゲットエリアを中心に画像を再構築できていることが確認できる．なお，この図からは確認しにくいものの，ターゲットエリア以外の部分にもわずかな色彩の変化が起きていたことがわかる．

2.3　比較実験

今回は Stable Diffusion で拡張した画像が CNN の汎化性能にどのような影響を及ぼすのかを評価するために，二つのデータセットを用いて VGG19 の転移学習ベースの CNN モデルを構築した．データセット A はオリジナル画像 (良品: 149, 不良品: 196) のみで構成し，データセット B はデータセット A に対して “DPM++ 2M

図 6　Example of the subtraction between an original image and its augmented image.

表 1　Comparison of classification accuracies

	CNN_A	CNN_B
Trial 1	96.52 %	97.10 %
Trial 2	96.81 %	97.68 %
Trial 3	96.52 %	96.23 %
Trial 4	95.65 %	98.55 %
Trial 5	95.65 %	97.97 %
Mean	96.23 %	97.51 %

Karras"を指定して拡張した画像を加えた良品: 354, 不良品: 475 で構成した．また，データセット A, とデータセット B で訓練した CNN モデルはそれぞれ，CNN_A, と CNN_B とした．さらに，左右反転させたオリジナル画像をテスト用データセット C (良品: 149, 不良品: 196) として，CNN_A, と CNN_C の分類精度を評価した．

表 1 には，CNN_A, と CNN_B をそれぞれ 5 回ずつ試行的に訓練した後，テスト用データセット C に対する分類実験の結果を示している．また，表 2 と表 3 にはそれぞれ，CNN_A と CNN_B のベストなモデルによる混同行列の結果を示す．この実験結果から，SDWU で拡張した画像を訓練に用いた場合，平均分類精度とベストの分類精度はそれぞれ，1.28%, 1.74% 程度改善することができた．

3　おわりに

本研究では，CNN モデルをもとに工業材料の欠陥検出システムを構築していくにあたり，SDWU 上の image-to-image 機能で再構築した拡張画像が分類性能に及ぼす影響を評価するための検証実験を行った．VGG19 の転移学習ベースの二つの CNN モデルでの比較実験の結果，不良品の発生頻度が低いような工業製品の画像拡張に有効であることが確認された．また，本研究では，先行研究 [11, 12] で確認できなかった Recall rate (再現率) の向上も確認された．

表 2　Best classification result by VGG19-based CNN_A model at Trial 2.

True \ Predicted	Anomaly (NG)	Normal (OK)
Anomaly (NG)	143	6
Normal (OK)	5	191

表 3　Best classification result by VGG19-based CNN_B model at Trial 4.

True \ Predicted	Anomaly (NG)	Normal (OK)
Anomaly (NG)	146	3
Normal (OK)	2	192

参考文献

[1] Japkowicz, N., "Learning from imbalanced data sets: a comparison of various strategies," *International Journal of AAAI Workshop Learn*, Imbalanced Data Sets, Vol.68, pp. 10–15, 2000.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y., "Generative adversarial nets," *International Journal of Advances in Neural Information Processing Systems*, Vol. 27, pp. 2672–2680, 2014.

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., "High-resolution image synthesis with latent diffusion models," *Procs. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

[4] Luzi, L., Siahkoohi, A., Mayer, P. M., Casco-Rodriguez, J., Baraniuk, R., "Boomerang: Local sampling on image manifolds using diffusion models," https://doi.org/10.48550/arXiv.2210.12100, 2022.

[5] 永田寅臣，渡辺桂吾，"不良品検出のための畳み込みニューラルネットワークとサポートベクタマシン設計支援ツール"，システム/制御/情報，Vol. 64, No. 8, pp. 304–309, 2020.

[6] 永田寅臣，渡辺桂吾，“畳み込みニューラルネットワーク (CNN)・畳み込みオートエンコーダ (CAE)・サポートベクタマシン (SVM) のための設計支援ツールの開発”，画像ラボ, Vol. 32, No. 12, pp. 20–26, 2021.

[7] Ho, J., Jain, A., Abbeel, P., “Denoising diffusion probabilistic models,” *International Journal of Advances in Neural Information Processing Systems*, Vol. 33, pp. 6840–6851, 2020.

[8] https://qiita.com/iitachi-tdse/items/6cdd706efd0005c4a14a
Accessed 26 February 2024

[9] Avrahami, O., Lischinski, D., Fried, O., “Blended diffusion for text-driven editing of natural images,” *Procs. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18208–18218, 2022.

[10] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., “Deep unsupervised learning using nonequilibrium thermodynamics,” *Procs. of International Conference on Machine Learning*, pp. 2256–2265, 2015.

[11] 鄭 哲霖，永田寅臣，“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), 2023.

[12] Zhelin, Z., Fusaomi, N., Souma, K., Keigo, W, Maki K. Habib, “Design of CNN models for defect detection of an industrial material using image augmentation based on Stable Diffusion,” *Procs. of the 29th International Symposium on Artificial Life and Robotics*, pp. 1180–1184, 2024.