

深層学習を用いた鋳造製品画像判別技術の開発

上 菌 剛*, 谷山清吾**

Development of Image Discrimination Technology for Casting Products using Deep Learning

Tsuyoshi UEZONO and Seigo TANIYAMA

工業製品（鋳造加工部品）の外観検査を自動化することを目的として、Deep Learning による良否判別の研究を実施した。ミラーを用いて1カメラで対象製品を撮像するシステムを構築し、Deep Learning に供する多量の画像を取得した。Deep Learning の手法としては深層距離学習を適用し、サンプルの側面、内面、上面（切削面と鋳造面）の各部位においてAIモデルを作成し、未知の画像に対する誤判別率2.6%（19/732枚）を得た。

Keyword : Deep Learning, 機械学習, 距離学習, 外観検査, 鋳造

1. 緒 言

近年、AI（Artificial Intelligence：人工知能）という言葉が当たり前のように使われている。辞書¹⁾には「学習・推論・判断といった人間の知能の持つ機能を備えたコンピューターシステム」とあり、人間の振る舞いを模擬するようなシステムを指す大きな概念のことである。この人工知能における研究課題の一つに機械学習（machine Learning）がある。機械学習とは、「データから反復的に学習し、そこに潜むパターン（特徴）を見つけ出すこと」であり、アルゴリズムとしてサポートベクターマシーン（SVM）、ニューラルネットワーク（NN）、深層学習（Deep Learning）、決定木（DT）、ナイーブベイズ（Naive Bayes）など様々な提案がなされている²⁾。その中で、現在の AI ブームの火付け役となった技術が Deep Learning である。G.Hinton 氏が2012年に論文発表³⁾し、同年に開催された AI 系画像認識コンテストで、初めて Deep Learning を利用したシステムが使われ、圧倒的な認識精度を發揮した。

Deep Learning では、コンピューターが自ら学習データから特徴量を抽出し、予測モデルを構築することが可能であるため、人間が介在して特徴量設計を行う必要がなくなり、以前よりはるかに容易に機械学習を行うことができるようになった。更にコンピューターの性能向上や、2010年代からビッグデータを扱うようになり、膨大な計算リソースを獲得したこと、またインターネットの普及による広範囲で迅速な情報の共有、交換など、技術開発に必要な要素が十分に揃ったことから、AI 技術は飛躍的に向上し本格的な活用が始まることとなった。AI の身近な活用事例として、パンの画像識別を行うベーカリースキャン、キュウリの画像選別システム、医療分野の画像診断、音声認識による AI スピーカー、テキスト生成 AI などがあり、既に

多くの分野で利用が進んでいる^{4)~8)}。

工業製品の生産においても、その活用範囲は幅広いと考える。例えば、製品の的外観検査などでは、多くの場面で画像による自動化が行われている。しかし、中には自動化が難しく、目視による検査に頼らざるを得ない製品などもある。その一例として鋳造製品などの不均一な表面に発生する不良などが挙げられる。このような製品の的外観検査には、Deep Learning によるアプローチが必要なモチーフであると考えた。そこで、本研究では外観検査の自動化が難しく、目視検査を実施している鋳造製品について、Deep Learning 技術を適用して良品、不良品の判別を行う技術の実用性について研究を実施した。

2. 実験方法

2.1 学習と評価に用いた工業製品

実験に供する工業製品（以下、サンプル）は、県内企業が目視検査で月産7万個程出荷する鋳造加工部品（ナット：42mm角）を対象にした（図1）。形状は8角形で、外側の表面に見える素地は砂型を転写した鋳造肌になっており、内側は切削加工により鏡面部分とねじ山が形成されている。発生する不良には、ピンホール、へこみ、こぶ等が



a) 外表面 b) 内側加工面
図1 鋳造加工部品（ナット）

* 地域資源部 ** 生産技術部



図2 代表的な不良（側面）

ある。ピンホールは内部にできた鑄巣が表面に出現したものである。こぶやへこみは、砂型の不具合により発生した不良である。不良の一例を図2に示す。

2.2 画像の取得

一般的に Deep Learning に使用する画像は膨大な枚数が必要である。また、本研究ではサンプル個々の良否判別を行うことから、サンプルの全表面を撮像する必要がある。そこで、効率的に撮像するために、鏡を利用した撮像システムを構築することとした。全体構成のイメージを図3に示す。

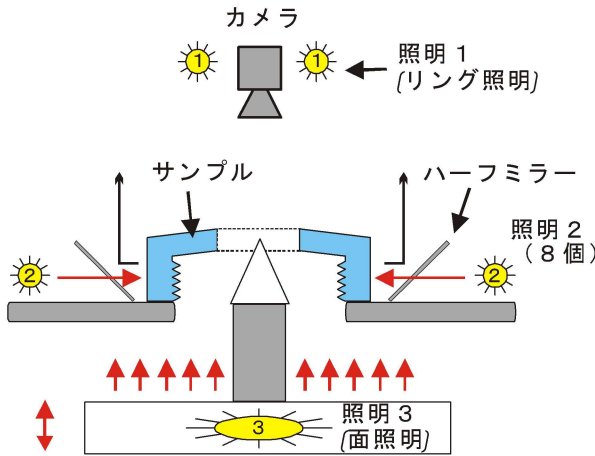


図3 撮像システム（イメージ）

サンプルは、開口部の広い方を下向きに設置し、その上方にカメラを1台設置した。サンプル上面の撮像は、照明1を点灯し、照明の輝度を調整して、直視で切削面と梨地面をそれぞれ撮像した。サンプル内面の撮像は、照明3を点灯し、サンプル中央にコーンミラー（φ10mm）を設置し、その反射像を撮像した。サンプルの高さ18mmに対して、コーンミラーの高さは5mmであるため、上下に可動する機構を付加して4回に分けて撮像した。サンプル側面の撮像は、照明2を点灯し、ハーフミラーの反射像を撮像した。サンプル側面は内面の切削加工のためにチャッキングされて素地が潰れて鏡面となっている部分がある。通常の照明方法では光の乱反射が無いため黒く写り、不良との差別化が困難になる。そこで、ハーフミラーの裏側に照明を設置して落射照明となる構成にし、潰れた鏡面は全反射で明るく写るように工夫した。サンプルは8角形のため、それぞれの面に対して照明とハーフミラーを準備し、1面ずつ8回撮像した。底面の撮像については、生産ラインへの適用

の仕方により撮像環境が大きく異なることが想定されるため、今回の研究では検討しなかった。

2.3 AIアルゴリズム

工業製品の外観検査において、一般的には不良品の発生はそれほど多くないため、不良品の情報が無くても学習できることが望ましい。この解決策としては、教師無し学習でアプローチすることになる。教師無し学習に対して教師有り学習というものがある。これは、多量の良品データと多量の不良品データを用意し、両者の特徴をそれぞれ学習させた後、評価において、入力データがどちらの特徴に近いかで判別を行う手法である。一方、教師無し学習は、良品のみを学習データとして利用し、良品の特徴のみを学習させる。評価において、入力されたデータが良品の特徴を持っているか否かで判断する。

このような異常を検知する分野で、現在主流となっている技術は、Deep Learning に距離学習を付加した深層距離学習 (Deep Metric Learning)⁹⁾ である。距離学習とは、データ間の計量（距離や類似度など）を学習する手法である。直感的には、特徴が似ているデータは近く、そうでないデータは遠くなるように計量を学習する。本研究では深層距離学習を用いて研究を行った。

データを入力して、ある出力を得るAIアルゴリズムを具現化したものを一般的にモデルという。本研究におけるモデルの全体像を図4に示す。入力画像をDNN (Deep Neural Network) に入力し、次に距離学習を適用する処理を付加し、最終的に2つの出力（良品か不良品）を得る。DNNには、MobileNetV2 と呼ばれる開発済みモデルを用い、これに ImageNet（データセットと呼ばれる多種多様な画像データ群）で学習したパラメータを適用しベースとした。

一般的に DNN は多数の層構造となっており、入力層から中間層を経て出力層に向けて情報が流れる。層間ではパラメータ（重みやバイアス）により情報の変換がなされる。MobileNetV2 は、入力層が1層、中間層が155層、出力層が1層の157層で構成される。入力層は224(縦)×224(横)×3(色)=150,528個で、出力層の個数は任意に設定が可能である。Depthwise Separable Covolutionと呼ばれる計算量を減らすための工夫がされており、その名のとおりモ

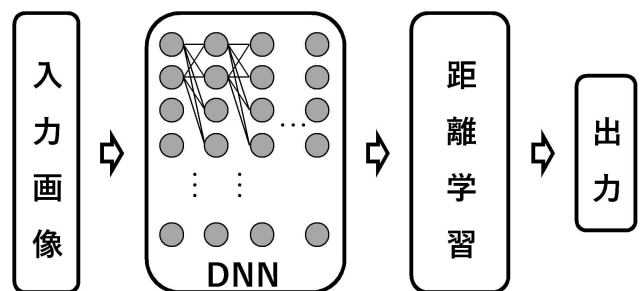


図4 モデルの全体像

バイル端末でも使用できるほど計算量やメモリ使用量が小さいという特徴がある。

次に、距離学習の機能は、L2 Softmax Loss という損失関数¹⁰⁾を使用した。損失関数とは、正解との誤差を表す指標で、L2 Softmax Loss の損失関数は、式(1)、式(2)で示される。Mはバッチサイズで、一度に学習させる画像の枚数であり、例えば16の場合16枚学習させる毎にパラメータを1回更新することになる。Xiは、モデルに入力する画像のデータで、関数f(Xi)はDNN部分からの出力を指す。Cは、出力のカテゴリ数で2(良品と不良品)となる。WとBはモデルのパラメータで、それぞれ重みとバイアスである。式1は、一般的なSoftmax Lossの式で、画像を入力した際に出力される結果と正解との誤差を表しており、この出力が小さくなるように、モデルのパラメータが更新され学習が進む。式2は、本損失関数の特徴で、DNNの出力の大きさを任意の定数αにスケールすることを表している。式1の損失関数は、パラメータが適切に更新され、精度が上がることに伴い小さくなるが、精度が上がっていてもDNNの出力が小さくなることでも小さくなる。つまり損失は小さくなっているが、認識精度が上がらないという状況が生まれる。これを抑制するために式2の制約を加えている。L2 Softmax LossはDNNの最終出力層の直前の情報を利用するため、最終出力層の手前に処理を追加し学習モデルを完成する。評価の際は、L2 Softmax Loss直後の出力を基にLocal Outlier Factor (LOF: データの集まりの中から外れた値を見つけ出す外れ値検知アルゴリズム)を実施し、その値を異常スコアとして判別に利用した。

$$-\frac{1}{M} \sum_{i=1}^M \log \frac{e^{W_{y_i}^T f(x_i) + b_{y_i}}}{\sum_{j=1}^C e^{W_j^T f(x_i) + b_j}} \quad \dots (1)$$

$$\|f(x_i)\|_2 = \alpha, \forall i = 1, 2, \dots, M \quad \dots (2)$$

- M: バッチサイズ
- Xi: 入力画像のデータ
- f(Xi): DNN部分の出力
- C: 出力のカテゴリ数 (2)
- W, B: パラメータ (重みとバイアス)
- α: 任意の定数

本アルゴリズムの検討は、python言語と深層学習フレームワーク「Keras」の開発環境を構築し実施した。学習に利用したPC環境は表1のとおりである。

表1 PC環境

パーツ	性能
CPU	Core i7K-8700K 4.6GHz
GPU	NVIDIA GeForce GTX1080Ti 11GB GDDR5X
メモリ	32GB DDR4

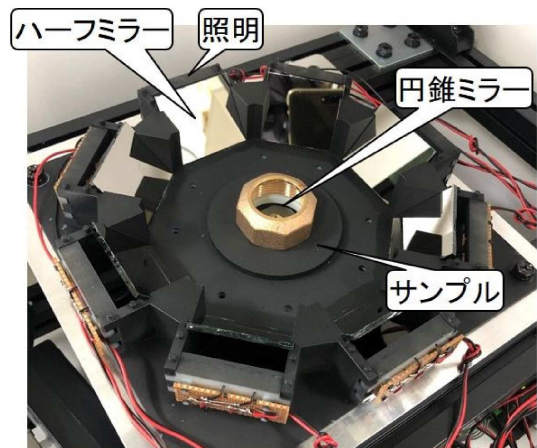
3. 結果及び考察

3.1 撮像システムの構築と画像取得

撮像システムの筐体は、アルミフレームにより外形43×49×80cmの大きさで作成した。サンプルを置くステージは3DプリンタによりLED、ハーフミラーが設置できるように作成した。光の反射を抑えるために、つや消し黒の塗装を施した。ステージ中央の下部には、Z軸ステージを設けてLEDとコーンミラーが取り付けられており、ステージ中央の穴(φ3.5cm)からコーンミラーがせり上がってくる機構とした。ステージから40cm上方にUSBカメラとLEDを設置し、ステージ上の全てのハーフミラーが画角に収まるようにレンズを選択した。カメラとサンプル間は、アームロボット等の搬送機構を動かせる空間を想定して40cmを確保した。ステージの下部には任意のLEDを点灯させるために、リレーを制御するI/Oコントローラ、電源など制御機器を設置した。構築した撮像システムを図5に、使用した代表的な機材を表2に示す。



a) 全体像



b) ステージ上

図5 構築した撮像システム

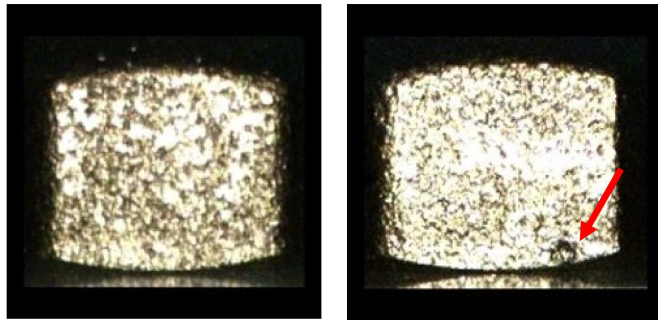
表2 使用機材

パーツ	性能
カメラ	DFKAFUJ003-M12+TBN7.2C 10MP
LED	COB LED 白 3,000K
Z軸ステージ	TSDM40-15X+GSC-01
IOコントローラ	DIO-1616HN-USB
コーンミラー	CNMR-10A Φ10mm
ハーフミラー	透過率30%

サンプルの設置は、ガイド治具を用いて手作業で行った。撮像に供するサンプルは、良品を約800個と、あらかじめ企業で選別した各種不良品を約200個、合計約1,000個準備した。カメラからは全ての面がミラーなどにより見えている状態であるが、最適な画像を得るために14回に分けて撮像した。撮像に必要な時間は1個あたり4.5秒であった。

3.2 側面画像の学習と結果

学習は、3,000枚の良品画像（以下 OK 画像）と、100枚の不良品画像（以下 NG 画像）合計3,100枚（256×256ドットの8bitカラー）を用いた。代表的な画像を図6に示す。図中の矢印は不良部位である（以下同様）。



a) OK画像 b) NG画像
図6 側面の画像

バッチサイズは16, 32, 64, 128の4通り、エポック数は250回までとした。エポックは学習の回数で、1エポックで3,100枚の画像を学習させることになる。画像サイズが大きいため、バッチサイズ256以上は、GPUメモリが不足し実行できなかった。

評価は200枚の OK 画像と200枚の NG 画像を用いた。それぞれ半数は学習に供したもの（以下学習画像）を用い、残り半数は学習に使用していない未知の画像（以下未知画像）として評価した（表3）。

表3 使用画像数 (単位: 枚)

		OK 画像	NG 画像
学 習		3,000	100
評 価	学習画像	100	100
	未知画像	100	100

バッチサイズとエポックの組合せは全て行い、OK 画像と NG 画像の異常スコアのグラフにより最適な条件を選定した。最適な評価結果のグラフを図7に示す。

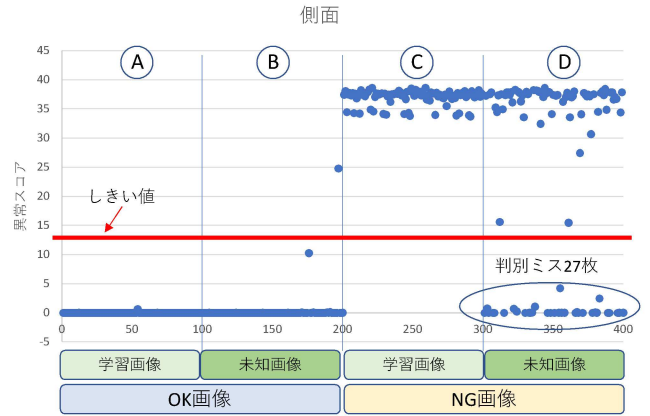


図7 評価結果（側面）

横軸は評価画像400枚分で、左から学習画像（OK）100枚、未知画像（OK）100枚、学習画像（NG）100枚、未知画像（NG）100枚である。説明の便宜上、各領域にA～Dのラベルを割り当てる。

縦軸は異常スコアで、OK 画像と判断できる特徴量を基準（0付近）として、そこからの異常スコアの大きさを示している。期待すべき結果としては、NG 画像のプロットが縦軸の大きな位置（OK 画像の特徴からの違いが大きい）に出現することである。

領域Aは、学習済みの OK 画像であり、全てが0付近にプロットされた。領域Bは、未知の OK 画像であり、AIモデルに初めて入力したことから0付近が多いものの、領域Aと比較して大きな値を持つものもある。領域Cは、学習済みの NG 画像であり、多くは0から離れた位置にプロットされており、OK 画像とはどこか違うと判断していることが判る。領域Dは、未知の NG 画像であり領域Cと比較して同様に NG と判断された画像が多いものの、それほど値の大きくならないものもある。

ここで領域B, Dは、未知画像を判別できるか大きな意味を持つ領域である。両者の間に大きな隔たりがあれば、しきい値を設定して判別が可能になる。本結果では領域Bでは、ほとんどが異常スコア0付近にあり OK 画像と認識していて問題無いが、領域Dでは27枚の画像が異常スコア0付近にあり、正しく NG と判断されていないことが判り、結果としては不十分である。

27枚の画像を確認すると、比較的サイズの小さな不良の画像が多かったことから、AIの判断基準が大きなサイズの不良側にシフトしているものと考えられた。不良画像の中には、比較的大きな不良も含まれており、その画像が学習に使われるとその影響を受け、大きな不良を主とする判断基準が生成されると考えた。そこで、学習画像の見直しを行い、比較的大きな不良画像は学習データから排除する選別を行い、再度学習させた。結果のグラフを図8に示す。

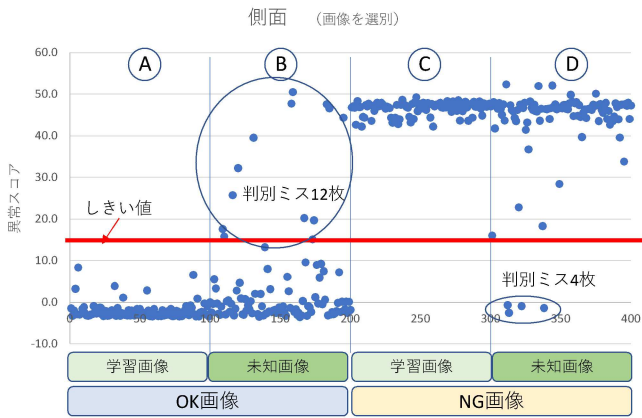


図8 評価結果（側面）（選別後）

領域Dの異常スコア0付近は4枚と減少し、改善が見られたが、領域Bの異常スコアが大きくなるものが多くなった。学習に供する画像により、判別結果が大きく変わることが確認できた。本結果において、適当な異常スコア値でしきい値を設けた結果、NGをOKと判断したものが4枚（偽陰性率4%）と、OKをNGとして判断したものが12枚（偽陽性率12%）、誤判別率としては合計で8.0%となった。偽陰性は市場に出るためクレーム対象になり、偽陽性は不良率が高くなることに繋がる。どちらも低く抑えることは重要である。

誤判別した領域Dの画像を図9に示す。領域Bにおいて、誤判別に繋がる共通した要因を特定することは残念ながらできなかった。領域Dについては、不良部位が周辺部に存在するという共通の要因が考えられた。学習画像にそのような不良画像を追加することで対応ができるのではと考えられた。

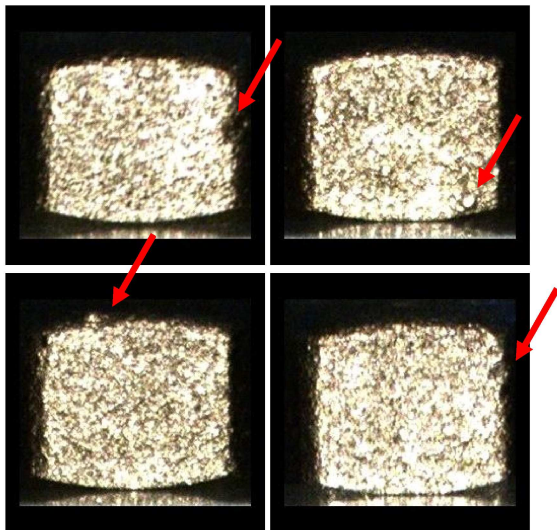


図9 偽陰性（領域D）のNG画像

3.3 内面画像の学習と結果

内面はねじ山加工のため切削面となっている。そのため、鋳造面に発生する不良は存在せず、鋳造時に生成される内部の巣が切削時に出現する不良が主となる。その発生確率

は低いため、不良画像はほとんど得ることができない。そこで、φ0.5とφ1mmのドリルで貫通穴を開け疑似不良のサンプルを作成した。疑似不良と実不良の画像を図10に示す（平面画像）。ほぼ同様の画像が得られており、ドリル穴による不良でも代用可能と判断した。

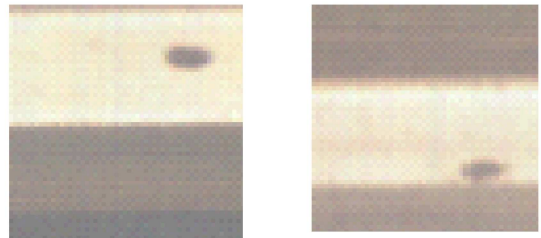


図10 不良画像の比較

内面の画像は、コーンミラーの高さを変えて4回撮像している。そのため、一番上2枚の画像にはねじ山以外の部分も撮像されているがその区別無く学習に供した。4種類の画像を図11に示す。一例として、図11b)に疑似不良が写り込んだ画像を示す。



図11 内面の画像

学習は、3,000枚の良品画像（以下OK画像）と、100枚の不良品画像（以下NG画像）合計3,100枚を用いた。評価は200枚のOK画像と200枚のNG画像を用いた。実験方法は3.2と同様である。最適な評価結果のグラフを図12に示す。

結果の傾向は、側面とほぼ同様で、OKとNGは総じて判別されていることが判る。領域Dに着目すると、異常スコア0付近が1枚あり、適当なしきい値を設けても判別ができない状況であった。偽陰性率は1%で、誤判別率は0.5%となった。

判別ミスの画像を図13に示す。矢印部分が不良箇所であるが、図11b)ほど明確な不良として撮像できなかったことが原因と思われる。

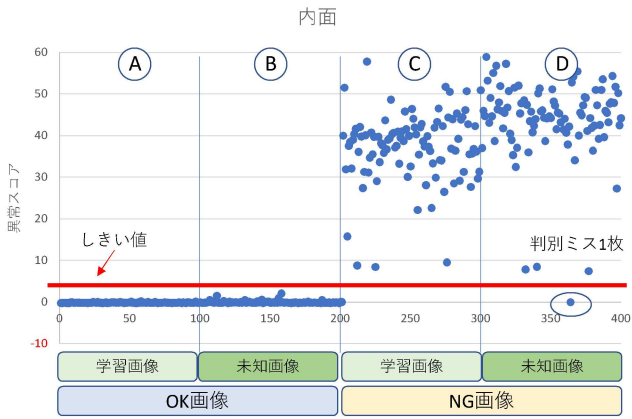


図12 評価結果 (内面)



図13 判別ミスの画像

3. 4 上面画像 (切削面) の学習と結果

上面の切削面には、鋳巣に起因する不良の発生はほとんど無いが、切削そのものの不具合 (切削面が形成されない、もしくは狭い) が散見されるためこれにターゲットを絞った。切削の不具合の画像を図14に示す。



a) 正常 b) 不良
図14 切削面が形成されない例

本不良は発生数も少なく、また疑似的な画像の作成も困難であることから、教師無し学習 (NG 画像を用いない) で検討した。NG 画像を使用しないこと以外は、これまでの実験と同様の手法を用いた。

学習は、800枚の OK 画像のみを用いた。評価は200枚の OK 画像 (半数は未知画像) と 7枚の NG 画像を用いた。用いた NG 画像は切削面が無いもの、狭いものに加えてやや傷、汚れのあるものを準備した。供した NG 画像 7枚と、OK 画像を図15に示す。

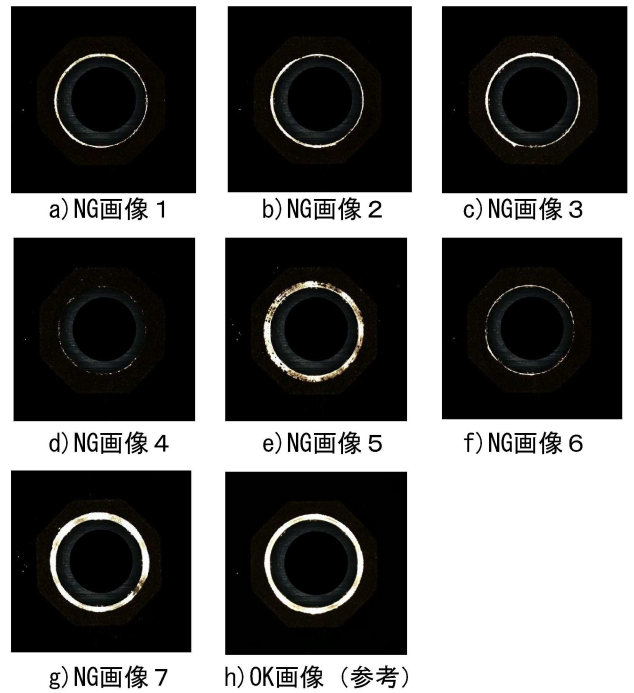


図15 上面 (切削面) の画像

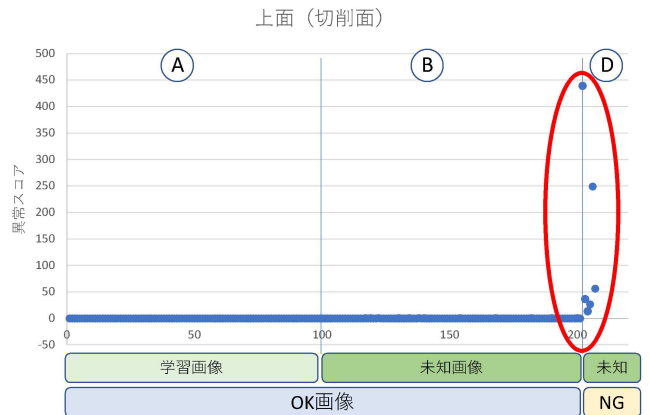


図16 評価結果 (上面 : 切削面)

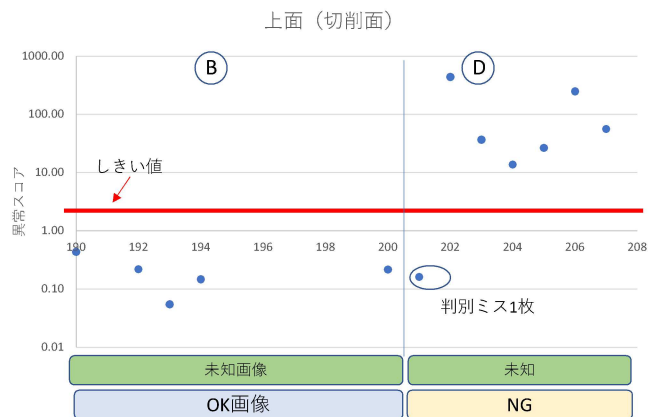


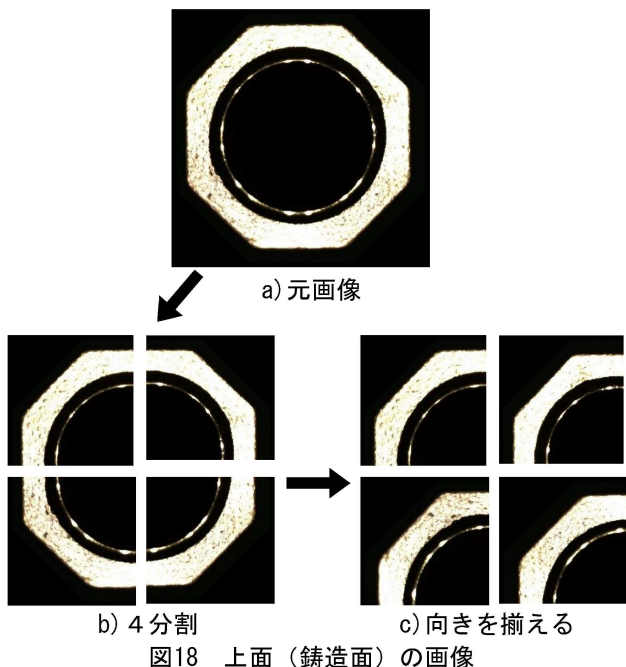
図17 評価結果 (上面 : 切削面) (拡大)

最適な評価結果のグラフを図16に、図16の一部 (丸囲み) を拡大したものを図17 (縦軸対数) に示す。

結果は、未学習の OK 画像については問題無く判別が可能であった。用意した NG 画像 7 枚のうち、1 枚 (図15g)) は判別が困難であった。偽陰性率は14%であった。図15の e) と g) は、正常に作成した切削面に、汚れがあるものと、一部に傷があるものを不良としている。明るく写る部分が広い場合 g) は正常と判断したのではないかと考えられた。

3. 5 上面画像 (鋳造面) の学習と結果

上面の鋳造面について検討した。この面は、砂型の接合部に接する部位があり、また溶解した金属が流れ込む部位もあるため、砂型から取り出した後、手作業で研削処理を行い、その後ブラスト工程で仕上げを行っている。そのため周辺の鋳造面とは明らかに仕上がりが異なる状態のものが多くなり理想的な OK 画像を多く確保することが難しい。できるだけ健全な OK 画像を得るため、取得画像を 4 分割して選別した (図18)。



学習は、185枚の OK 画像と、40枚の NG 画像、合計225枚を用いた。評価は370枚の OK 画像 (半数は未知画像) と、80枚の NG 画像 (半数は未知画像) を用いた。実験方法は 3. 2 と同様である。最適な評価結果のグラフを図19に示す。

結果の傾向は、側面とほぼ同様に、OK と NG は総じて判別されていることが判る。適当なしきい値を設定したとき、領域 B では 4 枚が、領域 D では 1 枚が判別困難であった。偽陽性率は2.2%、偽陰性率は2.5%であった。

誤判別した領域 D の画像を図20に示す。領域 B において、誤判別に繋がる共通した要因を特定することは残念ながらできなかった。領域 D については、側面の考察を勘案すると、不良部位が周辺部に存在するという要因が考えられた。

以上 3. 2 ~ 5 の結果、傾向としては OK 画像、NG 画像の判別はできていると思われる。ただし細かく見ていく

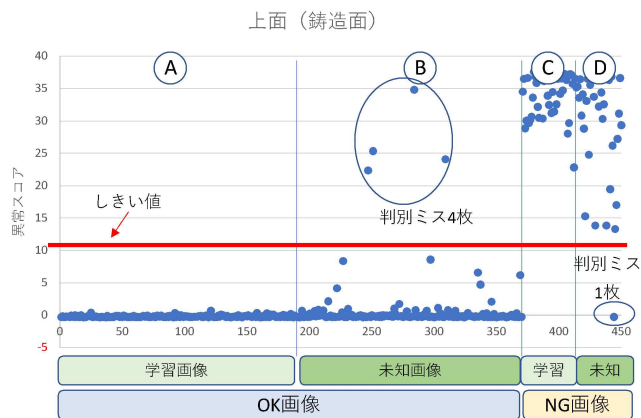


図19 評価結果 (上面 : 鋳造面)

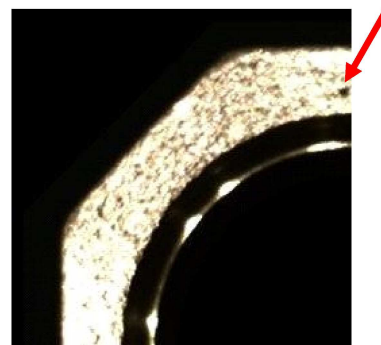


図20 偽陰性 (領域 D) の NG 画像

と誤判別もあり100%とはならなかった。各実験の未知画像に関する結果 (領域 B, D) の誤判別の一覧を表4に示す。単純に母数の総和と、誤判別の数の総和の比ではあるが、全体で2.6%の誤判別率となった。いずれの場合も誤判別があり、更なる改善が必要である。

誤判別について考察を試みるが、AIが自ら作り上げた判別基準であり、人間が理解するルールベースではないため、改善する明確な根拠を示すことは難しかった。ただ、判別が100%でなくても十分実用性はあると考える。本案件のように、外観検査という「機能」ではなく「見た目」が問題になる場合、不良の外部流出というデメリットと、AIによる検査時間の短縮、検査人員の削減などコスト削減の効果などのメリットを比較して、メリットがデメリットを上回れば十分利用が可能である。また、学習の画像選択など

表4 領域 B, D の誤判別一覧 (枚)

		側面	内面	上面 1 (切削面)	上面 2 (鋳造面)	合計
領域 B	誤判別数	8	0	0	4	12
	供試数	100	100	100	185	485
領域 D	偽陽性率	8.0%	0%	0%	2.2%	2.5%
	誤判別数	4	1	1	1	7
領域 D	供試数	100	100	7	40	247
	偽陰性率	4.0%	1.0%	14.3%	2.5%	2.8%
合計	誤判別数	12	1	1	5	19
	供試数	200	200	107	225	732
計	誤判別率	6.0%	0.5%	0.9%	2.2%	2.6%

アプローチするポイントはまだまだあり、更なる認識率の向上は見込めると考える。

4. 結 言

外観検査の自動化が実現していない鑄造加工部品の良否判別を対象に、深層距離学習を適用し、以下の結果を得た。

- (1) 円錐ミラーとハーフミラーを用いることで、1カメラで底面を除くサンプル全景を撮像するシステムを構築し、Deep Learning に供する多量の画像を取得することが可能になった。
- (2) 深層距離学習という手法を用いることにより OK 画像 NG 画像の特徴を学習したモデルを作成した。各実験において評価を行い、未知画像として供した画像数の総和のうち、誤判別の画像の総和は2.6%であった。

参 考 文 献

- 1) 三省堂:”大辞林 第四版”(2019) p.1397
- 2) 秋庭伸也ら:”機械学習図鑑” 翔泳社(2019)
- 3) A. Krizhevsky, I. Sutskever and G. EHinton:”Image Net Classification with Deep Convolutional Neural Networks”, In Proc., NIPS2012 (2012)
- 4) BakeryScan製品ホームページ <https://bakeryscan.com/> (閲覧日 2023-8-8)
- 5) キュウリ生産者が開発、人工知能(AI)を取り入れた自動選別装置 https://agri.mynavi.jp/2019_10_01_90831/ (閲覧日 2023-8-8)
- 6) 「医用画像解析ソフトウェアEIRL aneurysm」のサービス提供を開始 <https://lpxel.net/news/press-release/2020/10026/> (閲覧日 2023-8-8)
- 7) AIスピーカーの仕組みとは?【徹底解説】 <https://nissenad-digitalhub.com/articles/ai-for-speaker/> (閲覧日 2023-8-8)
- 8) 2023年最新版!実際に使えるテキスト生成AIおすすめまとめ! https://blogcake.net/text-ai/#index_id1 (閲覧日 2023-8-8)
- 9) 深層距離学習(Deep Metric Learning)の基礎から紹介 <https://tech-blog.optim.co.jp/entry/2021/10/01/100000> (閲覧日 2023-8-8)
- 10) R Ranjan, CD Castillo, R Chellappa:”L2-constrained softmax loss for discriminative face verification” arXiv preprint arXiv:1703.09507 (2017)