＜修士論文＞

熟練工の知見に基づいた深層学習による鍛造製品の欠陥検出

滋賀大学大学院
データサイエンス研究科
データサイエンス専攻

修了年度　：2023年度
学籍番号　：6022104
氏　　名　：上羽 悠介
指導教員　：飯山 将晃
提出年月日：2024年1月10日

目 次

1　序論

鍛造とは，金属加工技術の一種であり，工業製品に欠かせない技術である．鍛造は自動車をはじめ，農耕機械や建造物などさまざまな分野で使用されており，中でも自動車のような強度及び大量生産が必要な分野で鍛造品が使用されている．自動車部品における鍛造品は高強度という製品の特性上，エンジンやトランスミッション，足回りといった重要部品に用いられることが多い [1]．そのため鍛造品をはじめとする粗形材を供給するメーカに対する品質保証の要求は厳しい．特に製品機能への影響という面から製品外観の品質は重要であり，全数保証が求められている．

そのため，品質保証として外観検査が行われ，一般的には人による目視検査法と呼ばれる検査方法が広く採用されている．この理由として人による目視検査法は欠陥品か正常品かを瞬時に見分ける判断力に優れており，さらにあいまいさを含む柔軟な判断が可能であることから，ある一定の精度で比較的容易に導入できるためである．

また，製造業の現場では熟練工というような，いわゆる匠の技とよばれる熟練した技術を持っている作業者が活躍しており，目視検査においても同様に熟練工が活躍している．目視検査における熟練工は，他の検査員と同じ作業手順で検査を行っていても，経験が浅い検査員では見逃すような，わずかな製品の違和感に気づき，不適合品を検出することができる．このような作業は誰もができる行為ではなく，作業ノウハウや勘・コツといった「暗黙知」として留まっていることが多い．

目視検査法の欠点として，検査員ごとに経験やスキルのばらつきがあること，検査員の疲労などによる見逃しの発生リスクが懸念される．さらに目視検査に必要とする工数は多大なものとなっておりコスト削減の観点でも問題がある．また，目視検査法の性能は熟練工の「暗黙知」に依存しており，働き手不足や技術継承の問題がある．このような問題に対応するため，信頼性・効率性向上を目的とした外観検査の自動化が求められている [2][3]．しかしながら目視検査法から自動検査法への切り替えが試みられているものの，製品及び検査項目の多様性，精度が不十分といった阻害要因によって進んでいないのが現状である [3]．

自動外観検査の手法としてルールベースと呼ばれる手法が開発されてきた．ルールベースの手法として例えば，カメラやレーザー，三次元計測によってデータを取得したのち，二値化処理等の画像処理を施したうえで閾値処理による識別手法がある [1]．しかし，ルールベースの手法では熟練工が持っている「暗黙知」領域の全てをルール化できず，熟練工レ

ベルの精度を達成するのが困難である．そこで，精度向上の解決策として近年では深層学習をはじめとした機械学習手法による工業製品の自動検査化が研究されている [4][5].

機械学習や深層学習は高精度を達成することが可能だが，引き換えにモデルの説明性や解釈性が低下する．本研究では説明性と解釈性について以下のように定義する．説明性は，モデルの出力に対する判定根拠を可視化や用いた特徴量の寄与度によって説明することが可能であることを指す．解釈性については，説明性を与える手法等によって示された判定根拠に対して，人が納得して理解できることを指す用語として使用する．識別モデルを現場導入する場合，現場作業者や品質担当者，さらには会社や顧客といった相手に対して説明責任が問われる．そのため説明性や解釈性は,「モデルを納得して使用できるか」という点で現場導入時の重要な要素となるが，深層学習は説明性や解釈性が低いという問題を抱えている．この問題に対して，特に画像識別モデルにおいては，説明性を付与する手法 [6][7] が提案されている．しかし説明性が付与されたとしても，その説明結果に解釈性がない場合，現場導入がためらわれてしまう要因となる．

さらに問題点として，深層学習モデルは学習データのバイアスから本質的ではない特徴を学習してしまい，人が予期しないような誤識別をしてしまうことがある.「猿」と「ギター」が同時に映っている画像では「人」と誤認識してしまう例 [8] や，人をゴリラと誤認識してしまう例 [9] が報告されている．仮に深層学習モデルの認識精度が高いとしてもこのような誤認識をしてしまうとモデルに対する信頼を損なうこととなり，現場導入に対する妨げとなってしまう．このように人が容易に識別可能な対象を認識モデルが誤識別しない性質のことを本研究では信頼性と定義する．

本研究ではこれらの課題を踏まえて，機械学習及び深層学習手法を用いた，高精度でありながらも解釈性と信頼性の向上を目指した手法を提案し，鍛造品の欠陥検出モデルに適用する．モデルの解釈性及び信頼性向上に対するアプローチとして，熟練工が目視検査を行う際の感覚に基づいた特徴を定義し，これとモデルの注目箇所との一致度を損失関数に組み込む方法を提案する．この損失関数は，熟練工が識別の手がかりとなりうると考えている箇所とモデルの注目箇所が一致している場合に損失が小さくなるように設計する．これにより指示した箇所と注目箇所が一致するように学習され，モデルの注目箇所の誘導によって本質的でない特徴を学習してしまうことによる誤分類を防ぐことができ，結果として解釈性と信頼性の向上が期待できる．さらに識別精度の向上を目指すために，機械学習手法である LightGBM モデル，畳み込み層を用いた比較的シンプルな深層学習モデルと事

前学習済み深層学習モデルの VGG16 をファインチューニングしたモデルを用いてアンサンブルを行う．

本論文の構成は以下のとおりである．第 2 章では工業製品の自動外観検査及び自動化手法の先行例について紹介する．第 3 章では物体識別タスクについての関連研究について紹介する．第 4 章では提案手法の各モデルの説明及びアンサンブル手法の結果について説明する．第 5 章では Grad-CAM を用いた損失関数の設計及び，鍛造製品データに対する実験結果について説明する．第 6 章ではオープンデータセットへ適用した実験結果について紹介し，考察をおこなう．第 7 章では結論と今後の展望について述べる．

2　工業製品の外観検査

2.1　外観品質と外観検査

2.1.1　外観品質

社会の成熟に伴い，品質に対する顧客の要求レベルは高まっており，製品一つひとつの信頼性が求められている．中でも外観品質は製品に本来備わっている性質・性能のうち，見かけにかかわるものであり，製品機能への影響という面から重要視されている．また外観品質は目で見える「実感品質」であり，メーカーとしての信頼を確保する上でも重要な品質項目であるといえる．

2.1.2　外観検査

外観検査は，製品を一つひとつ検査して外観的な不適合があれば排除し，適合品だけを顧客に提供することを目的とした方法である．製造過程上で製品の不良が発生することをゼロに出来ない状況では品質保証のために外観検査は欠かせないものである．

外観検査の方法は一般的に人による「目視検査法」が行われるケースが多い．目視検査法は一般的なレベルの外観検査を比較的容易に実施できることから導入しやすい．またあいまいさを含む判断が可能であり，多様な良否サンプルの用意といった準備工数が少ないことも導入がしやすい理由である．一方で長時間作業による検査精度や検査処理能力の低下によって不適合品の見逃しなど，誤識別が起こる問題がある．この問題解決のために検査員を増員することはひとつの対応方法であるが，人件費の増加やそもそもの労働人口の低下という問題が存在する．また，もう一つの外観検査の方法として，機械によって外観

検査をおこなう「自動検査法」がある．自動検査法の利点として，適切に検査設定を行うことが可能であれば検出の精度が高く，人間のような疲れや個々の能力変動がなく，処理能力の増強も可能である．ただし検査対象によっては適切に検査設定を行うのが困難かつ設定以外の項目については検出が出来ないという問題がある．

2.1.3　判断の誤りと過検出

外観検査における判断の誤りには 2 通りある．一つ目は，適合と判断すべき製品を不適合と判断する場合の「過検出」，二つ目に，不適合と判断すべき製品を適合と判断する場合の「見逃し」がある．過検出は，微小な欠陥レベルの製品を検出し，適合品であるにもかかわらず不適合とする行動である．過検出が頻発すると，製品の歩留まりが悪化し，過剰品質となり適正な品質保証とは言えない．見逃しは，明らかに不良品であると識別できる製品を適合とする行動である．見逃しと過検出はトレードオフの関係にあり，どちらの指標を重視することになるかは目的によって異なる．外観検査においては，見逃しは後工程（顧客の場合が多い）への不適合品の流出を意味するため，品質保証の観点から重視されることが多い．また本研究では製品の外観的な欠点 (キズ，変形，欠損) のことを欠陥と呼び，品質基準を満たしていないものを不適合品或いは不良品と呼ぶ．

2.2　自動外観検査

2.2.1　自動化の阻害要因

目視検査法に対する様々な問題点から，自動検査法への切り替えが求められているが，いくつかの阻害要因によって進まない状況である．[3] によると，費用面及び検査対象・項目の多様性によって生じている課題の他に，精度及び処理能力に関する課題があると報告されている．また，その他に鍛造品の外観検査においては「欠陥レベルが連続的である」「製品表面状況がショットブラスト痕による細かい凹凸により，外観上の見え方に個体差が発生している」といった課題もあり，自動検査法への切替のさらなる障害となっている．

ここで欠陥レベルが連続的とは，欠陥の程度に何か区切りがあり段階的に変化しているわけではなく，連続的であることを指す．図 1 に示すように重大から軽微な欠陥まで，欠陥レベルにグラデーションがあり，適合・不適合に明確に二分することが難しい．この特性によって，分類精度向上や画像に対する教師ラベルの付与などの難易度が高くなっている．

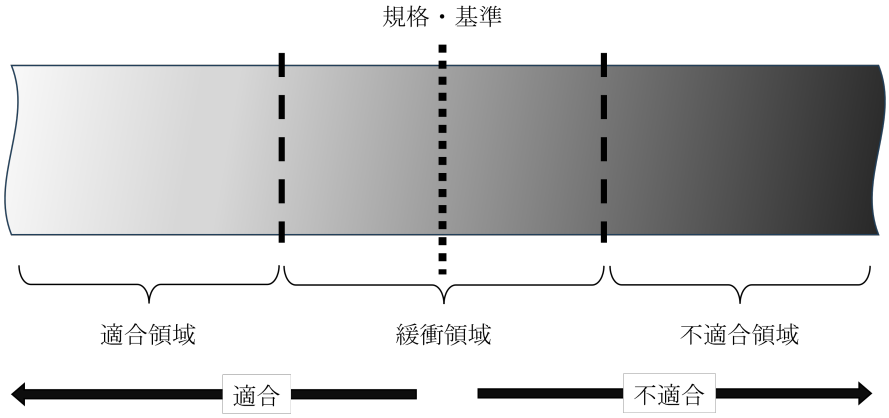


図 1: (参考文献 [2] 図 3.3 を参考に作成) 欠陥レベルの連続的変化と規格・基準と適合・不適合の関係

もう一つの問題点であるショットブラスト痕とは，製品表面の 0.1mm 程度の細かい凹凸のことである．鍛造はいくつかの工法があるが，材料の加工温度で分類した場合，熱間鍛造，冷間鍛造，温間鍛造などに大別される．本研究では熱間鍛造で加工された製品を対象としている．熱間鍛造では材料を高温で加熱するため，製品表面に酸化スケールと呼ばれる被膜が発生する．この被膜を除去するためにショットブラストという細かい鉄球を高速でぶつけて研掃する工程が設けられている．このショットブラスト工程によりショットブラスト痕が発生しているが，ショットブラストの鉄球はランダムに製品にぶつかるため表面凹凸位置もランダムに発生する．またショットブラストの鉄球サイズも摩耗により変動し設備内で様々なサイズの鉄球が混在しているため，凹凸サイズもランダムとなる．このランダム性のために外観上の見え方に個体差が発生し，自動検査法の適用を困難にしている．

2.2.2 ルールベース手法

自動検査法において，識別対象の領域をどのように定義するかはその性能を左右する重要な点である．そのような定義，言い換えればルール，による外観検査の自動化手法は，ルールベースと呼ばれ，設定したルールが適合・不適合を正確に表現することができれば高精度且つ，結果の解釈が容易である欠陥検出システムの構築が可能となる．これまでにも自動車用部材の外観検査自動化は多くの開発事例が報告されている．しかし適用する検査手法の見極めや特徴量の抽出方法，判定ルールの設定にノウハウが必要であるなどの理由から困難であることが多い．

鍛造品の外観不良には欠肉，打痕，バリといった製品の形状変化を伴うものと，汚れ，錆といった外観上のみが変化するものに分けられる．本研究では欠肉と呼ばれる外観不良を対象とするため，製品の形状変化を捉える必要がある．部品表面の形状の変化を捉えるには主にレーザ測定と CCD カメラによる画像計測が用いられることが多い．しかしレーザ

測定は高精度であるものの一般に装置が高額になるため，安価な方法である光学カメラによる画像取得と画像処理を組み合わせた手法が提案されている．

[5] ではコネクティングロッドと呼ばれる自動車用鍛造部品の表面上の欠陥に対して CCD カメラで撮影した画像を用いた手法が紹介されている．提案ではまず対象画像をいくつかのブロック単位に分割し，各ブロックにおける輝度値のヒストグラムを基にブロック間の非類似度を計算，二値化処理を施し検査対象部位を動的に検出する．非類似度を用いているのは，検査対象としている部位は他部位と比較して異なる反射特性を示しているためである．そうして検出した対象部位に対して，ヒストグラム平坦化とガウシアンフィルタによる画像の平滑化を行い，欠陥部位の強調を行う．その後，検査対象領域の各画素に対して，近傍画素を含めた平均値を求め，あらかじめ設定した閾値を下回ったとき，欠陥が原因で明度が低いと考え欠陥候補とする．そして検出された欠陥候補がある一定以上の範囲となった場合に，欠陥として判定する．この手法ではある程度の大きさの欠陥に対しては適切に検出できるが，欠陥が小さい場合に見逃してしまうことが欠点である．この欠点への解決策として欠陥候補となる画素の閾値設定の調整が考えられるが，過検出と見逃しはトレードオフの関係となるため，更なる精度向上のためには，他の計測方法やルールの追加が必要である．

[4] では同じく鍛造部品に対して画像処理による欠陥検出を提案している．提案手法では前処理としてヒストグラム平坦化とガウシアンフィルタを施し，欠陥の強調とショットブラスト痕の軽減を行う．エッジ検出手法である Canny 法と直線検出手法であるハフ変換を併用して欠陥検出を行っている．しかしながらこの手法においても安定性向上の課題が指摘されている．

2.2.3　機械学習・深層学習手法

一方で機械学習・深層学習による外観検査自動化手法も近年多く提案されている．

機械学習とは，訓練データからバターンとルールを学習し，未知のデータに対して予測や分類を行うことが可能な技術である．深層学習（Deep Learning）とは，機械学習の枠組みの中でも，ニューラルネットワークと呼ばれる人の脳の仕組みから着想を得た構造のアルゴリズムを用いた手法のことを指す．画像の分類タスクに利用できる機械学習手法の例として，決定木や Support Vector Machine が挙げられる [10]．しかしこれらの手法は，ルールベース手法と同じく，画像前処理や特徴抽出，特徴削減，および分類器の選択について

ノウハウやチューニングが必要となる．対して，深層学習は認識に必要となる特徴量を学習によって取得することができ，高精度なモデルを獲得することが可能である．特にデータ形式が画像の場合は，畳み込み層を使用した畳み込みニューラルネットワーク (CNN) や自然言語処理モデルの Transformer を画像タスクに用いた Vision Transformer(ViT) 等が手法例として挙げられる [11]．CNN は畳み込み層で入力画像の特徴を学習し，特徴マップと呼ばれる画像特徴量を表現するフィルターを獲得することができる．この畳み込み層をいくつか重ねることにより，様々な特徴量を獲得することが可能となり，ルール化が困難な画像認識タスクに対しても高精度に推論することができる．ただし十分な精度を得るには大量のラベル付きデータが必要である．

大量のラベル付きデータが必要という課題に対して有効な手法として転移学習 (Transfer learning) が挙げられる．転移学習とは，目標ドメイン (terget domain) と呼ばれる新規タスクに対して，元ドメイン (source domain) と呼ばれる関連した別タスクによって学習された結果を再利用する手法である [12]．転移学習の手法はいくつかあり，事前学習モデルを特徴抽出器として使用して SVM やロジスティック回帰などの分類器を目的タスクのデータで学習する方法や，事前学習されたパラメータを使用して，目的タスクのデータで再度学習を行う fineturning と呼ばれる手法がある．転移学習では従来大量の学習データが必要とされていた深層学習において，少量の学習データでも高精度を達成できる重要な技術となっており，特に fineturning は事前学習した画像と一見類似していないようなターゲット画像に対しても，汎用的な特徴量を獲得することで高い分類精度を得ることができる [13].

ラベル付きの学習データが正例，負例ともに手に入る場合は欠陥検出を分類タスクとして解くことが可能だが，まったく欠陥データが手に入らない場合もある．そのような場合は教師なし学習による異常検知手法が用いられる．

AnoGAN[14] は，Generative Adversarial Networks(GAN)[15] をもとに発展した異常検知手法である．GAN は Generator; 生成器と Discriminator; 識別器の 2 つのニューラルネットワークを競い合わせるように学習する画像生成モデルである．AnoGAN は，生成画像と入力画像の残差から計算される残差スコアと，正常画像から学習した特徴量空間における入力画像の特徴量の適合具合を表す判別スコアを用いる．これら 2 つのスコアを用いて異常度スコアとして異常検知を行う．

VAE[16] は AutoEncoder(AE)[17] から発展された手法である．AE は，入力を潜在変数に変換する符号化器と，潜在変数から入力の再構成を行う複合化器の二つから構成され，入

力と出力が同じになるように学習される．VAE では符号化器で生成される潜在変数が正規分布となるように学習される．これにより潜在空間の近い領域に位置するデータ点同士が類似した特性となることが期待される．VAE で学習されたエンコーダーによって得られる特徴量は低次元空間上でデータを表現することができ，KNN や One-Class Support Vector Machine[18] といった異常検知手法に適用することが可能である [19].

3　画像認識と認識根拠の可視化技術

3.1　物体認識

物体認識とは画像中の物体やシーンのカテゴリを求める問題である．物体認識は，一般物体認識と特定物体認識の大きく分けて 2 つのタスクに分類される．一般物体認識は「猫」や「車」といった一般的なクラスの識別を行うものである [20].

一般物体認識は，深層学習の発展及び大規模画像データのオープン化により大幅に精度向上がなされている．特に ImageNet[21] の識別精度を競うコンペティションの ILSVRC では CNN を用いた AlexNet の登場以来深層学習手法のモデルが上位を占めており，2015 年に人間の認識エラー率を下回った [22].

一方の特定物体認識は特定のカテゴリの物体（例えば顔や人物）の認識を行う問題である．顔画像から個人の同定などがこれにあたる.

本研究では扱う画像は鍛造品という一般的ではない対象であるが，未知のデータに対して「良品」「不良品」というカテゴリのどちらに属するかという認識を目的としており，クラス数 2 の一般物体認識として位置づけられる.

3.2　Context Model

物体認識では検出対象の物体に対して，画像全体から局所特徴量を抽出するため背景情報が認識を阻害するノイズとなる [23]．本研究でも背景情報をノイズとして取り扱うアプローチをとっているが，一方で背景情報を利用することで高精度にシーン認識を行う手法も提案されている．そこでは前景物体と背景物体の共起性を利用しており，このような共起性を利用した物体認識のことを Context Model[24] という.

[24] では対象オブジェクト周辺の色の分布やテクスチャといった低レベルの画像情報が物体検出の精度向上に寄与すると示している.

[25] では物体の周辺領域の情報であるローカルコンテキストや，画像全体の情報であるグローバルコンテキストを利用することで，全体画像に対して小さい物体や一部分しか映っていないような大きな物体に対しての識別精度が向上したことを示している．

これら従来研究に対し，本研究が対象とする外観検査では画像に写りこんでいる検査対象以外の物体や背景はほぼ不変であり，また背景シーンの情報が前景には影響しないため，提案手法では背景部分の影響を低減するように処理する．

3.3 Grad-CAM

深層学習の発展により画像認識のさまざまなタスクにおいて精度は大幅に向上した．しかし深層学習モデルや機械学習モデルは高精度であることの引き換えにモデルの予測根拠を人が直観的に理解できないといった問題がある [26]．このような問題はモデルの説明性や解釈可能性が低いといった用語で表現される．[27] では説明性と解釈性の用語について区別して説明されており，本研究でも二つを以下のように区別して使用する．説明性は，モデルの出力に対する判定根拠を可視化や用いた特徴量の寄与度によって説明することが可能であることを指す．解釈性については，説明性を与える手法等によって示された判定根拠に対して，人が納得して理解できることを指す用語として使用する．つまり説明性がある場合でも解釈性を持ち合わせているわけではないことがある．

画像認識の分野においては CNN モデルの説明性を付与する手法として Class Activation Mapping(CAM)[6] が提案されている．CAM は畳み込み層の最終層に global average pooling(GAP) 層を有した CNN モデルに対して利用可能な手法である．クラスアクティベーションマップの生成方法について図 2 に示す．GAP 層は畳み込み層で得られた特徴マップ毎に平均を計算し，出力を行う層である．次に GAP 層から出力された値を後続の分類用 NN に重みを掛けて入力する．それから分類用 NN は適切な重みとなるように学習を行う．その結果，得られた GAP の値に掛かる重みの値が，各分類結果に対する各特徴マップの重要度と考えられる．そして得られた重みを特徴マップに掛けることで，出力結果の判断根拠を可視化することができる．

図 2: (参考文献 [6]figure2 から引用) 上：入力層→畳込層→ GAP 層→出力層からなる予測モデルを考える．最終畳込層直後に GAP 層を実行して，各特徴マップ毎の平均を取る．これにより各特徴マップの値に対する重みと出力結果との対応関係を取ることが可能となる．下：得られた重みを各特徴マップに掛けることで，出力に対してどの特徴マップが重視されたかを可視化することができる．

CAM は GAP 層を持たないモデルに対しては適用することはできない．そのため CAM の拡張手法として Grad-CAM[7] が提案されている．CAM は GAP 層の出力値に対する重みを各特徴マップの重要度として解釈するのに対して，Grad-CAM は CNN の出力に対する特徴量マップの勾配を利用して重要度を計算する．図 3 に Grad-CAM の概要を示す．Grad-CAM における重要度は以下のように計算される．まずネットワークの順方向に入力を行う．そして出力に対する特徴量マップの勾配 $\frac{\partial y^c}{\partial A^k}$ を計算する．勾配 $\frac{\partial y^c}{\partial A^k}$ は出力を各特徴マップで微分したものなので，勾配の大小は出力クラスに対する各特徴マップの影響度を表すものと解釈出来る．このネットワークに対する逆向きの勾配を GAP によって特徴量マップ毎に平均化したものが α_k^c となり，出力クラスに対する重要度を表している．そして α_k^c を係数として各特徴量マップに掛けたものを ReLU 関数に通すことで Grad-CAM $L^c_{Grad-CAM}$ を得る．ここで ReLU 関数を通す理由は，出力 C に対して正の影響を与える特徴量マップのみを可視化するためである

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{GAP}} \frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{2}$$

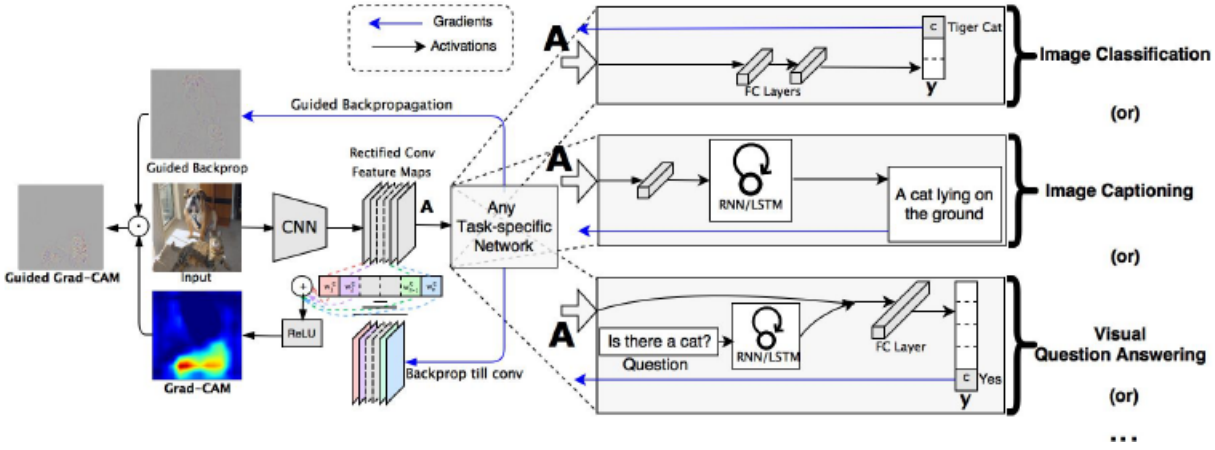


図 3: (参考文献 [7]figure2 から引用) 画像分類の場合，入力画像をネットワークに準伝播した際に得られる出力クラスに対して各特徴量マップ毎の勾配を計算する．計算された勾配を GAP により平均化した後，ReLU 関数を通して正の勾配情報のみを取り出しヒートマップを作成する．

3.4 Grad-CAM の結果を用いた認識精度の向上

3.4.1 共起バイアス

深層学習モデルを用いた学習では，データセットに由来する様々なバイアスから対象を正確に識別することが出来ないことがある．その例として共起バイアス (co-occurrence bias) が上げられる [28][29]．共起バイアスとは，識別対象と同時に出現する確率が高い物体や背景が存在する場合に，識別対象の特徴ではなく同時に出現した物体や背景の特徴を使用して識別することである．例えば [30] では船と水辺が共起しており，船を認識するのに水面や海岸沿いの輪郭の特徴を使って識別していると報告されている．他にも [31] では，口紅が塗られているかを識別するのに口だけではなく、目や眉などのメイクに注目して識別していると報告されている．前述した context-model では，積極的に共起性を利用して識別精度を上げるアプローチであったが，共起関係に信頼性が無いような場合（必ずしも同時に画像に出現することが約束されていない場合）に共起バイアスを学習してしまうとモデルの判定結果の信頼性及び汎化性能が失われてしまう．このような問題に対処するために，注目箇所の誘導についていくつかの研究がされている．

3.4.2 アテンション誘導

[28] や [29] では Grad-CAM の可視化結果とユーザーが指定した領域との重なり具合を損失関数に組み込むことで，共起バイアスが低減されることを報告している．[28] の実験では CelebA データセットを用いた顔属性の分類を実施しているが,「口紅を塗っている」と

「メイクが濃い」などのような共起関係が存在している問題に焦点を当てている．この問題に対して，(1)Grad-CAM による可視化，(2) ユーザーによる注目領域の指定，(3) クロスエントロピーと Grad-CAM の結果を用いた損失の導入，の 3 つのアプローチを提案している．(1) についてはモデルがどこに注目しているかを可視化するために Grad-CAM を利用する．(2) では画像に注釈をつけるためのユーザーインタラクションを提案している．顔のランドマーク検出を行うライブラリを用いて入力画像のセグメンテーションが行われるが，ユーザーは目や口といった顔の領域が分割されたイラスト上で操作するだけで関心領域の指定が完了する UI を作成した．(3) では Grad-CAM で出力された注目領域と (2) で指定した関心領域の重なり具合を組み込んだ損失関数を導入し，ファインチューニングを行う．以下に提案されている損失関数を記載する．

$$Loss = w_a \cdot loss_a + w_g \cdot loss_g \tag{3}$$

$$IoU_{loss} = -ln\left(\frac{G \cup S}{G \cap S}\right) \tag{4}$$

式 3 の $loss_a$ は，バイナリクロスエントロピー損失である．$loss_g$ は式 4 によって計算され，Grad-CAM によって可視化された注目領域 G とユーザー指定領域 S の一致度である．本研究においても式 3，4 を参考にした損失関数を導入した．詳しくは 5 章にて議論する．

4　自動外観検査の実現に向けた検査モデルの構築

欠陥品の検出として適切な手法を比較検討するために，ルールベース手法，機械学習手法，深層学習手法のモデル構築及び，評価結果について記載する．まずは使用したデータセット及び評価指標について説明を行う．

4.1　実験設定

4.1.1　データセットの説明

本研究では株式会社ゴーシューで製造している製品群の中から代表して 1 つの製品を選び、識別対象とする不具合事象を「欠肉」に限定して研究をおこなう．対象製品を限定した理由は，代表として選んだ製品は過去に自動検査法への切り替えを試みた製品であり，画像の撮影環境が整備されているためである．また不具合事象を欠肉に限定した理由は，欠

肉は鍛造製品の中で代表的な不具合事象であり，その他製品に対しても共通して発生し得る不具合事象であるため，水平展開が行いやすいからである．

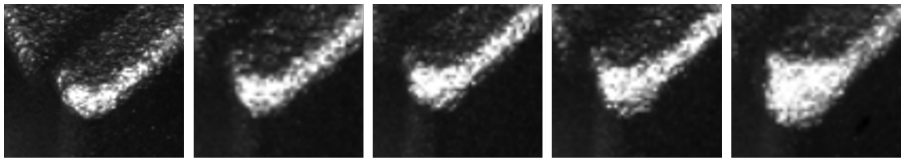
次に不具合事象の欠肉について説明を行う．鍛造は成型用の金型を用いて金属材料を成型する加工方法であるが，材料が金型に充満せずに欠損となることを欠肉という [32]．欠肉の形状に対しては図面規格にて許容されているが，規格を逸脱した場合に不良品となる．

画像の撮影には専用の撮像システムを用いる．撮像システムは照明環境が一定の状態で撮影ができるように筐体の中にカメラと照明を組み込み，外光が入らない構造となっている．上記撮像システムで取得できる原画像のサイズは 600×800[pixels] である．製品全体画像に対して欠陥が発生する領域が極めて小さいため，全体画像の変化に対して欠陥による変化も小さいものとなることから，対象領域のみが写った範囲の抽出を行い，これを学習・評価用データとして使用する．

画像枚数の内訳について表 1 に示す．また train・validation は同一生産ロット，test データは別生産ロットからデータを取得している．図 4 に検査領域のみ切り取った良品・不良品の画像サンプルを示す．

表 1: 画像枚数内訳

種類	良品画像	不良品画像
train	771 枚	174 枚
validation	193 枚	44 枚
test	433 枚	68 枚



(a) (b) (c) (d) (e)

図 4: (a)(b)(c) 良品画像サンプル，(d)(e) 不良画像サンプル

4.1.2 評価指標

モデルの評価指標については適合率 (Precision)，再現率 (Recall)，F 値 (F-score) を用いた．それぞれの計算式を式 5，6，7 に示す．また本研究においては不良品を陽性 (positive)，良品を陰性 (negative) として扱った．

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F - score = \frac{FP}{TP + FP} \tag{7}$$

Precision はモデルが不良と判定した画像のうち，実際に不良だった画像の割合である．Precision は 0 から 1 の範囲の値をとり，1 に近づくほどよい．過剰に不良と判定してしまうモデルの場合，Precision は低くなる．次に Recall は，不良であるデータ全体のうち，モデルによる予測も不良だったものの割合である．こちらも 0 から 1 の値をとり，1 に近づくほどよい．そのため不良品の見逃しが少ないほど数値は高くなる指標である．これらの Precision と Recall はトレードオフの関係にあり，どちらか一方が高くても良いモデルとは言えない．3 つ目の F-score は Precision と Recall の値を調和平均した値である．一般的にはモデル間比較の場合，F-score の高いモデルが Precision と Recall のバランスが取れた良いモデルとされることが多いが，タスクの性質によってどちらかを優先すべき場合がある．本研究では不良品の見逃しは品質不具合となるため，Recall を重視する．

4.1.3 現場導入における精度目標について

上記の評価指標はモデル単一の性能評価や提案手法による効果の評価に使用する．次に自動外観検査の導入時に到達すべき精度目標について説明する．評価には「見逃し率」と「過検出率」を使用した．それぞれは以下計算式で算出される.「見逃し率」は Recall と近い概念だが，すべての不良品のうち，誤って良品と判定した割合である．一方で「過検出率」はすべての良品のうち，誤って不良品と判定した割合である．これらの指標に対する企業目標は「見逃し率」は 0 %,「過検出」は 10 %以下となる．つまり品質不具合となる不良品の見逃しは 0 個を目指し，歩留まり率を表す過検出は 10 %以下であれば，再検査の費

用を含めても自動化の恩恵が得られることを意味している．これらの指標は複数のモデルをアンサンブルした際の最終精度評価時に使用する．

$$\text{見逃し率} = \frac{FN}{TP + FN} \tag{8}$$

$$\text{過検出率} = \frac{FP}{TN + FP} \tag{9}$$

4.2 予測モデル

4.2.1 ベースライン手法

本研究ではまず，ベースラインモデルとしてハンドクラフト画像特徴量によるルールベースモデルの構築を行った．その理由は，ルールベースモデルは使用する特徴量及び判定ルールが人によって作成されるため解釈性に優れており，目標精度に対して満足する場合は運用モデルとして好ましいためである．

本研究では図 5 に示す構造を用いて予測分類を行った．構造内の各ステップについて説明を行う．まず入力画像に対して画像位置補正を行う．位置補正には位相限定相関法を用いた．位相限定相関法は画像をフーリエ変換したのち，位相スペクトルを利用して，画像のマッチングを行う手法である．図 6(b) に位相限定相関法を施した例を示す．(a) 原画像は欠肉領域がやや左側に寄っているが，(b) 位置補正後は画像中心位置に補正されている．位置補正後に欠肉領域の全体が捉えられる矩形の座標を指定して検査領域の抽出を行う．

次に適応的二値化による欠肉面積の抽出を行う．ベースラインモデルでは欠肉面積を特徴量として使用する．適応的二値化を採用した理由は 2 つある．一つ目は，画像によって欠肉領域の画素値が異なる場合があり，単一の閾値を用いる二値化手法ではうまく処理できない場合がある．この問題に対して，適応的二値化は変換画素と周辺画素の平均値を閾値として二値化処理を行う手法であるため，画像毎の画素値の変化に対応できる．二つ目にショットブラスト痕の凹形状の影響を低減するためである．表面の凹形状により，画素明度が周囲と比べて暗くなっている領域が存在する場合があり，その場合画像内で共通の閾値で二値化を行う手法では，欠肉領域が一部欠損した状態で二値化されてしまう問題がある．このような問題に対しても適応的二値化は画素毎に閾値が設定されるため影響が少ない．図 6(c) の例では，(a)，(b) でみられた部分的に暗い箇所に対してもうまく二値化処理がされている．

続いて二値化された画像に対して，モルフォロジー変換のクロージング処理を施す．クロージングは二値画像に対して膨張と縮小を複数回行い．微小なノイズを除去する処理である．適応的二値化である程度ショットブラスト痕の影響を低減して二値化することができるが，図 6(c) のように欠肉領域内に黒い穴があいてしまうことがある．このような場合にクロージングによってこのような黒い穴を埋めることができる．図 6(d) に例を示す．尚，適応的二値化及びクロージングについては各種パラメータ調整が必要だが，本実験では試行錯誤のうえ決定している．

最後に，クロージング後の二値化された欠肉領域の画素値合計を算出し，ある一定の値以上となった場合に欠陥として判定する．

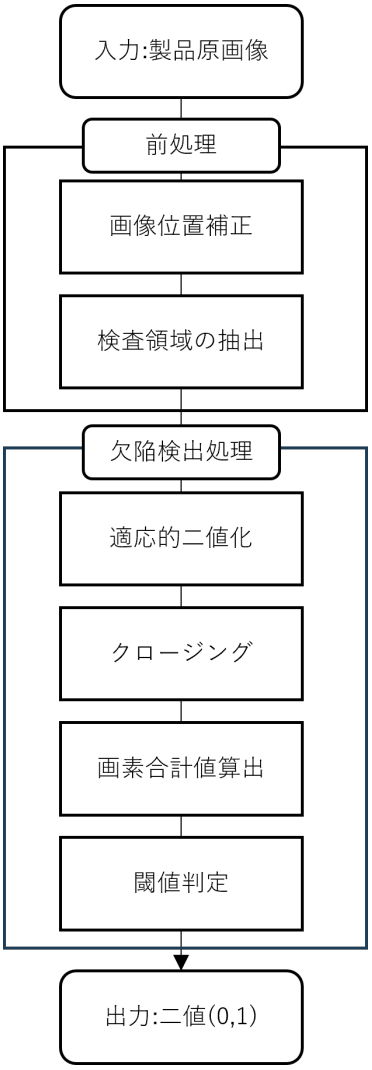


図 5: ベースライン構造

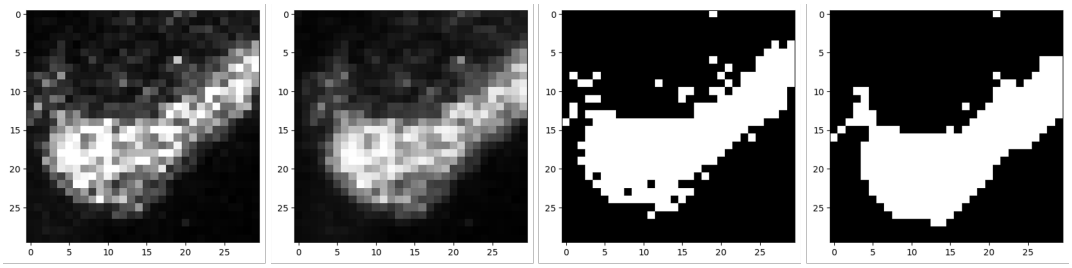


図 6: 各画像処理結果例. (a) 原画像※検査領域のみ表示 (b) 位相限定相関法 (c) 適応的二値化 (d) クロージング

4.2.2 LightGBM モデルの構築

次に高精度化のため機械学習手法の LightGBM[33] の構築を行った．LightGBM は，決定木を使用したブースティングアルゴリズムをベースとしたモデルである．LightGBM を選択した理由は，高精度且つ学習速度が速いこと，feature-importance によって判定に寄与した特徴量を取得できるためである．使用する特徴量については 4.3 節で詳細に述べる．

4.2.3 CNN モデルの構築

続いて深層学習手法の畳み込みニューラルネットワーク (CNN) を用いた分類器を構築した．ルールベース手法及びハンドクラフト画像特徴量を使用した機械学習手法では，人間が作成した特徴量を使用するため，解釈性は比較的高いが，高精度を達成する良い特徴量の作成には試行錯誤が必要である．深層学習手法は特徴量を学習によって獲得するため解釈性は低くなりやすいが，高精度を達成することができる．

本研究で使用したネットワーク構造について図 7 に，各種ハイパーパラメータについて表 2 に示す．またスクラッチで作成したモデルは今後 MyModel と表記する．MyModel の backborn は，畳み込みとバッチ正規化 (BN）を 2 層、MAX プーリング層が 1 層の Layer が 2 層積み重なっている．Head は全結合層が 2 層積み重なって構成されている．出力クラスは「良品:0」「不良品:1」の二値となる．誤差関数は Cross Entropy とした．エポック数は 50 として，validation データの正解率が一番高いエポック時のパラメータを採用した．

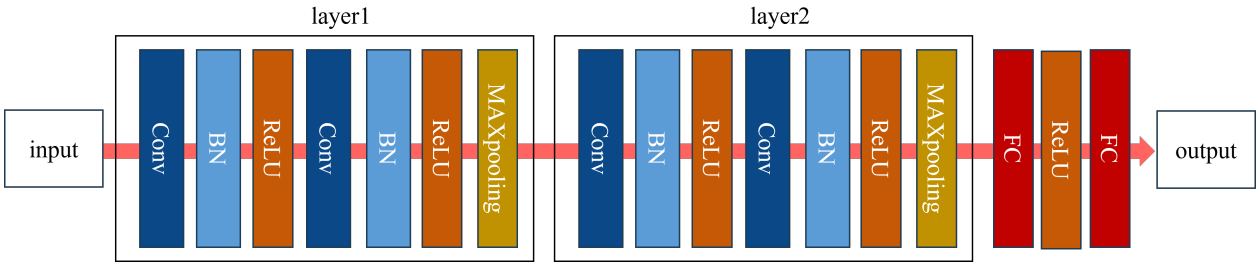


図 7: ネットワーク構造

表 2: 各種パラメータ

層	Filter size	Stride	入力形状	出力形状
畳み込み層 (1)	5 × 5	1 × 1	3,30,30	6,30,30
BN 層 (1)	-	-	6,30,30	6,30,30
ReLU 層 (1)	-	-	6,30,30	6,30,30
畳み込み層 (2)	5 × 5	1 × 1	6,30,30	6,30,30
BN 層 (2)	-	-	6,30,30	6,30,30
ReLU 層 (2)	-	-	6,30,30	6,30,30
プーリング層 (1)	2 × 2	2 × 2	6,30,30	6,15,15
畳み込み層 (3)	3 × 3	1 × 1	6,15,15	16,15,15
BN 層 (3)	-	-	16,15,15	16,15,15
ReLU 層 (3)	-	-	16,15,15	16,15,15
畳み込み層 (4)	3 × 3	1 × 1	16,15,15	16,15,15
BN 層 (4)	-	-	16,15,15	16,15,15
ReLU 層 (4)	-	-	16,15,15	16,15,15
プーリング層 (2)	2 × 2	2 × 2	16,15,15	16,7,7
全結合層 (1)	-	-	784	120
ReLU 層 (5)	-	-	120	120
全結合層 (2)	-	-	120	2
損失関数	Cross Entropy			
最適化アルゴリズム	Adam			
バッチサイズ	32			
epoch 数	50			

4.2.4 VGG16 モデルの構築

次に事前学習済みモデルを用いたファインチューニング手法による識別モデルの構築を行った．複数のモデル候補から予備実験を行い，精度面から ImageNet で事前学習された VGG16 に決定した．予備実験の結果は補足資料に掲載する．VGG16[34] は，畳み込み層が 13 層と，全結合層が 3 層の計 16 層からなるニューラルネットワークである．このモデルは畳み込み層に 3 × 3 の小さな畳み込みフィルタのみを用いていることが特徴である．ファインチューニングによるパラメータ更新は全層を対象としている．

4.3 特徴量設計

4.3.1 ハンドクラフト画像特徴量

LightGBM に用いたハンドクラフト画像特徴量を表 3 に示す．ベースモデルで使用した欠陥面積 (二値) の他に，いくつか作成したものの中から feature-importance で上位のもの

表 3: ハンドクラフト特徴量

特徴量	次元数
欠肉面積 (二値)	1
欠肉面積 (画素値)	1
主成分得点	5
投影面積	12

を選択した．ベースモデルから追加した特徴量についての説明を行う．

4.3.2 欠肉面積 (二値)

ベースラインモデルで用いた欠肉面積について説明する．まず特徴量の作成にあたり、ラベリング作業を担当した製品検査作業者に対して，画像から良否判定を行う際にどの部分を見て判断しているのかヒアリングを実施した．その結果から欠肉部の製品形状の変化具合から良否判定を行っているとの回答が得られたため，その判定方法の形式化を行った．本実験では「欠肉」という欠陥を検査対象としているが，欠肉形状の特性から光の反射が集中しやすい形状となっている．その特性から光の反射がある領域（鏡面反射領域）は欠肉領域と解釈することができ，欠陥判定に有用であると考えられる．また適応的二値化を行うことで，ショットブラスト痕によるノイズや，生産ロットの違いによる製品表面肌の部分的な明るさの差異を吸収できることも期待できる．

4.3.3 欠肉面積 (画素値)

二値化を施した欠陥面積の他に画素値の値をそのまま合計した特徴量を使用した．先述した二値化欠陥面積は，ショットブラスト痕によるノイズ除去や，製品表面肌のばらつきを吸収する目的で二値化を施している．しかし二値化によってハイライトの強弱に関する情報が失われてしまうため，原画像の画素合計値を特徴量として使用した．二値化面積と同じく，欠陥品は欠陥形状によって通常品よりも多く光が反射しているため，画素合計値としては大きくなる．

4.3.4 主成分分析

主成分分析 (PCA) とは多次元データの情報を低次元空間で表現する際に用いられる手法である．相関している変数同士をまとめることができるため，機械学習においては次元圧

縮に使われることがある．画像に対しても主成分分析を応用することができ，顔画像認識の例 [35] では主成分分析で抽出した特徴量 (主成分得点) を利用することによって認識精度の向上が報告されている．

本研究ではまず学習データ群で PCA を実施した．次に test データに対して，学習データで得た PCA モデルを適用して各主成分軸における各データ点の値の大きさを表す主成分得点 (スコア) を取得し，それを特徴量として使用した．また本研究では第 5 主成分までを特徴量として使用した．

4.3.5　投影面積

欠陥の形状を考慮した特徴量として投影面積を使用した．作業者にヒアリングした結果，欠陥の面積以外にも欠肉形状が横に広いことや，縦に広いことといった情報も識別時に考慮していると回答が得られた．これは図面規格が面積ではなく，形状で既定されているためである．そこで縦・横の長さ情報を取得する目的で投影面積特徴量を導入した．以下図 8 に投影面積の算出イメージを記載する．入力画像 (30×30) を縦・横方向に 5 ピクセル刻みに画素値を合計する．その結果，対象品の光の反射特性によって，合計画素値が大きい部分は欠肉範囲が広いという情報を得ることができる．

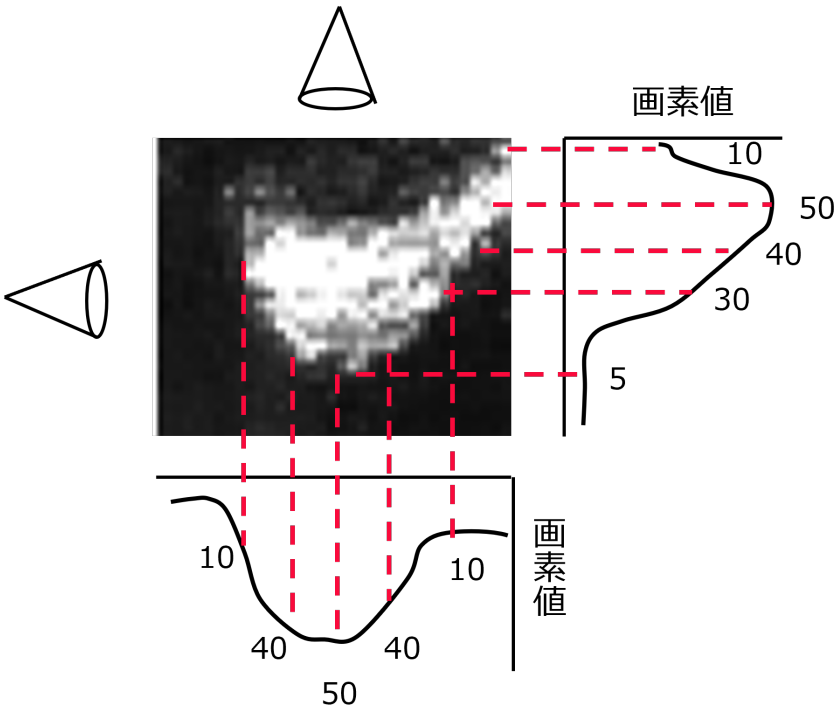


図 8: 投影面積の算出方法

4.4 実験結果

表 4 に構築したモデルの精度及び 3 つのモデルのアンサンブル結果を示す．アンサンブルに用いたモデルは LightGBM，MyModel，VGG16 である．またアンサンブル方法は 3 つのモデル出力の多数決とした．表 4 の結果より，ベースモデルであるルールベースと比べて機械学習・深層学習モデルは F-score が向上しており，認識精度の向上が確認された．またモデル単体の精度では，重視している指標の Recall がベースモデルよりも下がっているモデルもあるが，アンサンブル結果では各種評価指標ともベースモデルを上回る結果となった．

これらの結果より，鍛造製品の自動外観検査において，機械学習手法及び深層学習手法を用いることにより，従来の画像特徴量を使用したルールベース手法から識別精度の改善が行えることが示唆された．

表 4: モデル精度とアンサンブル結果

	Precision	Recall	F-score
ルールベース（baseline）	0.35	1.0	0.52
LightGBM	0.67	0.96	0.79
MyModel	0.58	1.0	0.73
VGG16	0.82	0.94	0.88
アンサンブル	0.71	1.0	0.83

5　熟練工の判断根拠を模した深層学習モデルの構築

5.1　実用化に必要な深層学習モデルの解釈性と信頼性

3 章でも述べたが，機械学習手法及び深層学習モデルは高い予測精度と引き換えに，説明性・解釈性が低いという問題がある．特に深層学習モデルは内部構造が複雑であるためその傾向が顕著である．この問題に対して，Grad-CAM をはじめとするモデルの出力結果に対して説明性を付与する手法が考案されている．しかし説明性が付与されたとしても解釈性が低い場合がある．例を図 9 に示す．図 9 は，Grad-CAM を用いてモデルが識別の際に注目している特徴を可視化したものであるが，共起バイアスによって船を認識する際に水面や海岸沿いの特徴を利用して識別している．このように説明に対して解釈性が得られない場合というのは，判定根拠が人の感覚と違う場合や，本質的な特徴を捉えていないといった場合に生じやすい．さらに本質的な特徴を捉えていないモデルは，人が予期しないような誤判定をしてしまうことがある．このようなモデルは信頼性が低いと判断され実用化の障壁となる．

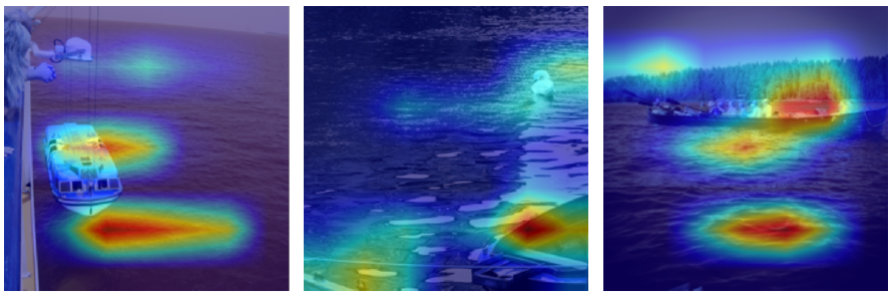


図 9: (参考文献 [29]Figure2 より抜粋)Grad-CAM によるモデルの注目箇所の可視化結果．船を認識するのに水辺の特徴を利用していることが分かる．

5.2　欠肉不良予測に対する Grad-CAM による可視化結果

4 章で実験を行った深層学習モデルの MyModel の識別結果に対して Grad-CAM による可視化を行った．その結果を図 10 に示す．これは誤識別した画像に対する Grad-CAM による可視化結果であるが，判定根拠を示すカラーマップの赤い領域は背景部分を示しており，解釈性を得ることができない．また欠肉領域のサイズ感は熟練工が見た際に大きいと感じるものであり，モデルの信頼性も損ねる結果となっている．

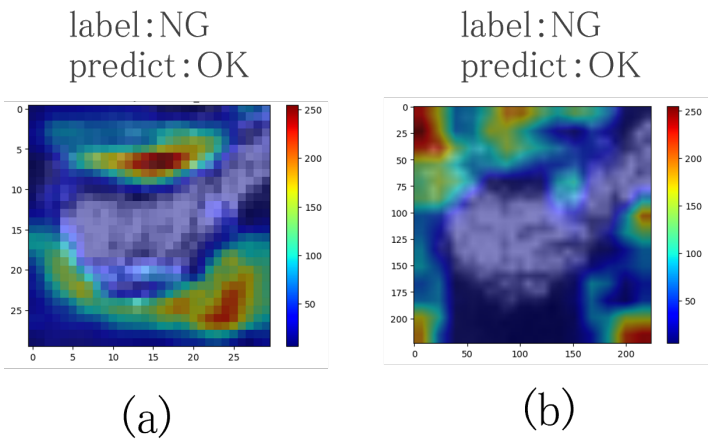


図 10: (a)MyModel 及び (b)VGG16 による鍛造品検査の予測に対する Grad-CAM 可視化結果

5.3 提案手法

共起バイアスの解決策として，提案手法では熟練工が目視検査を行う際の感覚に基づいた特徴領域と Grad-CAM によるモデルの注目領域の可視化結果との一致度を罰則項として加えた損失関数を導入する．これにより人が認知している本質的な特徴を学習するように誘導されることが期待できる．図 11 に示すように，熟練工が目視検査を行う際の感覚に基づいた特徴領域を 0，1 の二値画像として表現する．そしてこの領域と Grad-CAM によるモデルの注目領域の可視化結果との一致度 (Binary Cross Entropy) を罰則項として加える．

提案する損失関数（Grad-CAM 損失）を式 10，11，12 で示す．ここで Grad-CAM による出力画像を G，熟練工の感覚に基づいた欠肉領域の二値化画像を S，G と S のピクセル数を N とする．L_{CE} は通常の Cross Entropy であり，L_G は，S の ij 成分を正解として，G の ij 成分との Binary Cross Entropy を計算したものである．尚 Grad-CAM の出力画像における各ピクセル値は，0 から 1 の範囲で正規化をおこなっている．w_{CE} と w_G はふたつの損失のバランスを調整する役割をもつ．Grad-CAM 損失は熟練工が指定した領域以外の箇所に注目して識別を行った場合に損失が大きくなるため，指定領域への誘導が期待される．

$$L = w_{CE} \cdot L_{CE} + w_G \cdot L_G \tag{10}$$

$$L_{CE} = CrossEntropy \tag{11}$$

$$L_G = H(S,G) = -\frac{1}{N}\sum_i \sum_j [S_{ij} log(G_{ij}) + (1 - S_{ij}) log(1 - G_{ij})] \tag{12}$$

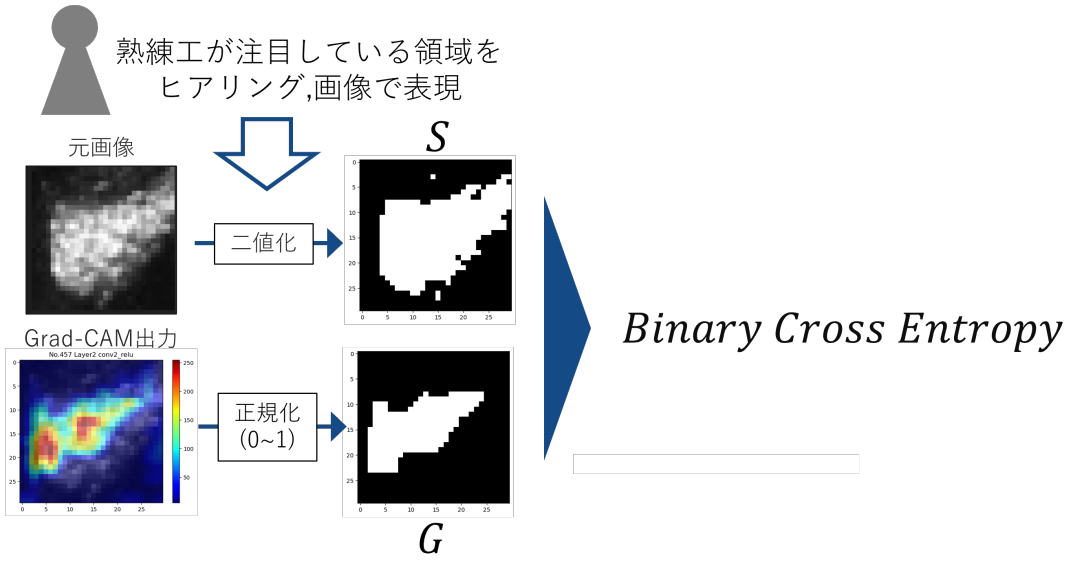


図 11: Grad-CAM 損失関数

5.4 実験設定

train データ，validation データ及び test データは 4 章で使用したデータと同じ枚数，同じ割合で使用した．また評価指標についても同じく Precision，Recall，F-score を使用した．モデル注目箇所の誘導結果の評価については，損失有無時の Grad-CAM 可視化結果を比較することで評価を行った．

使用モデルは，4 章で構築した MyModel を対象とし，損失関数に Grad-CAM 損失を適用した．

Grad-CAM 損失を用いた学習にはあらかじめ指定した特徴領域のマスク画像が必要となる．特徴領域の設定については，4.3 節で使用したハンドメイド特徴量の欠肉面積を熟練検査員が認知している本質的な特徴として利用した．

5.5 実験結果

Grad-CAM 損失を用いて学習したモデルの精度評価結果を表 5 に示す．Grad-CAM 損失を追加していないモデルを Normal，提案手法の Grad-CAM 損失を追加したモデルを GradCAM-Loss と表記している．L_{CE} の係数 w_{CE} については 1.0，L_G の係数 w_G については 0.03 とした．この係数の決定については w_{CE} は 1.0 で固定し，w_G については複数の値をグリッドサーチで調査を行い，validation の精度が最も良かったものを選択した．また，記載している結果については，学習パラメータを初期化して学習を行う実験を 5 回繰り返したものの中から最も Validation の精度（Recall,F-score,Precision の順で優先）の高かったモデルの結果を記載している．

精度の比較による結果は，Grad-CAM 損失の追加により F-score が若干ではあるがモデルの精度が低下したことを確認した．

続いて Grad-CAM による可視化結果について図 12 に示す．図 12 に示したサンプルは，Normal が良品を誤って不良品と誤識別した FP:(偽陽性) から抽出したものの中から，提案手法は正しく識別且つ，人による判定が容易な欠陥サイズのものを選択した．この選択方法の理由については人が判断しやすい欠陥レベルの製品を誤識別した場合にモデルの信頼性が損なわれやすいためである．尚，FN:False Negative(偽陰性) については，Normal モデルの識別結果に test 及び val データに FN が存在しなかったため，記載をしていない．

図 12 において，左のグレースケール画像が入力画像である．中央 2 つのカラーマップ画

像は Grad-CAM の可視化結果であり，赤枠側が予測結果のクラスである．1 つの入力に対して良品，不良品の可視化結果がある理由は，Grad-CAM は予測クラスの可視化だけでなく，任意のクラスに対しても特徴領域の可視化が可能なためである．

結果について確認する．sample1，2 ともに，Normal では欠陥領域ではなく，背景に注目して OK と誤識別してしまっている．対して提案手法は良品を識別する際の注目箇所が欠肉領域の輪郭周辺に分散していることが分かる．NG を識別する際の注目箇所は Normal と大きくは変わらないが，結果として正しく不良品と判定できるようになった．これは Normal では背景の本質的でない特徴に過度に反応して誤識別してしまう共起バイアスの影響を受けていたと考えられる．提案手法ではアテンション誘導によって共起バイアスの影響が低減されたことにより，正しく識別が行えるようになったと考えられる．

表 5: Grad-CAM 損失による精度比較 (MyModel)

	Precision	Recall	F-score
Normal	0.58	1.0	0.73
GradCAM-Loss (wg0.03)	0.56	0.99	0.71

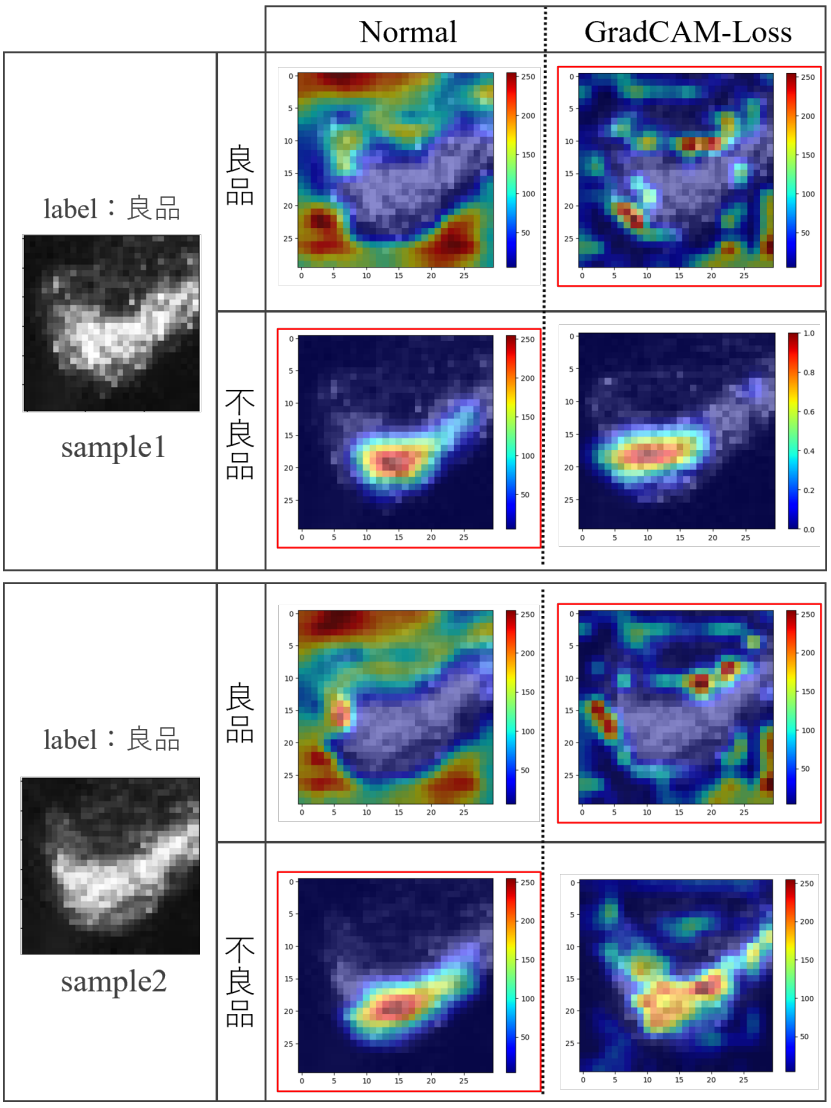


図 12: MyModel:Normal と GradCAM-Loss(提案手法) の Grad-CAM 可視化結果比較: 左に入力画像，中央のカラーマップは Grad-CAM の可視化結果を示す．カラーマップの左右はそれぞれ良品，不良品を識別する際に使用された領域を表し，赤枠で囲まれた側が予測ラベルである．

6　提案損失関数の一般性評価

提案手法の一般性評価としてオープンデータセットを用いた評価を行った．以下で用いた実験設定及び結果について述べる．

6.1　実験設定

データセットとして MS COCO(Microsoft Common Objects in Context)[36] を用いた．MS COCO はオブジェクト検出やセグメンテーションに主に利用されるデータセットであり，約 33 万枚の画像が収録されている．91 のカテゴリがあり，そのうち 82 カテゴリにアノテーション付きインスタンスが付与されている．MS COCO は 2014 年，2015 年，2017 年と段階的にリリースされており，今回は 2017 年リリースのものを使用した．2017 年リリースのデータセットは事前に train データと validation データに分けて用意されているが，事前に分けられた train データを学習用の train と validation に，事前に分けられた validation データを評価用の test に使用した．また MS COCO には複数のカテゴリが用意されているが，問題を簡単にするために使用カテゴリを限定して，二値分類タスクとして実験を行った．

次に使用したカテゴリについて説明する．使用したカテゴリは Boat(カテゴリ ID：9) と Airplane(カテゴリ ID:5) の 2 つを使用した．使用カテゴリの選択理由は，提案手法の効果を検証するために識別対象と共起性のある画像を使用したいという意図があったためである．具体的には Boat カテゴリの画像は水辺との風景と同時に出現する確率が高いと考え，同じく Airplane カテゴリの画像は，背景情報に空が写ることが多いと考えられたためである．

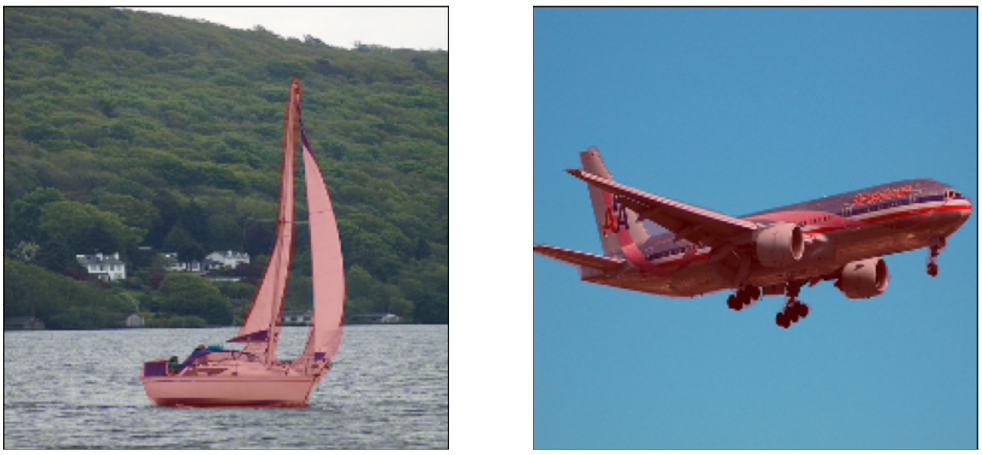
続いてデータの前処理について説明する．MS COCO データは 1 枚の画像につき，複数のカテゴリが付与されていることがあるため，Boat と Airplane カテゴリが重複している画像は除外した．次に複数の同一クラスオブジェクトがアノテーションされている画像は，面積が最大となるオブジェクトに付与されたアノテーション情報のみを利用した．最後に画像サイズを揃えるため，全ての画像を 224×224 のサイズにリサイズした．またマスク画像についても同様のリサイズ処理を行った．以上の前処理を行った上で，最終的に学習及び評価に使用した画像枚数を表 6 に示す．

予測モデルには 4.2.3 節で構築したスクラッチ CNN モデルを使用した．

Grad-CAM 損失の学習に必要となるマスク画像には，5 章で使用した熟練工の知見を表現した 2 値画像の代わりとして，MS COCO データセットにあらかじめ付与されている物体領域のアノテーションデータを使用した．アノテーション領域を可視化した結果を図 13 に示す．このアノテーション領域を物体を識別する際の本質的な特徴として利用した．

表 6: MS COCO 使用データ内訳

種類	Boat	Airplane
train	2033 枚	2076 枚
validation	874 枚	892 枚
test	118 枚	97 枚



(a) (b)

図 13: MS COCO データセットに付与されたアノテーション領域の可視化．赤いハイライトがアノテーション領域を示す.(a)boat，(b)airplane クラスの例

6.2 実験結果

二値分類の結果について表 7 に示す．通常の CE 損失で学習したモデルを Normal，Grad-CAM 損失で学習したモデルを GradCAM-Loss と表記している．L_{CE} の係数 w_{CE} については 1.0， L_G の係数 w_G については 10.0 とした． この係数の決定については w_{CE} は 1.0 で固定し，w_G についてはグリッドサーチを行い，validation の精度が一番良かったものを選択した．評価指標の比較では，F-score が若干であるが向上していることが確認できた．

表 7: 提案損失による精度比較 (Airplane，Boat の二値分類結果)

	Precision	Recall	F-score
Normal	0.66	0.72	0.69
GradCAM-Loss (wg10.0)	0.68	0.80	0.74

次に Grad-CAM による可視化結果を図 14，図 15 に示す．図 14 に示したサンプルは，Normal が誤識別且つ，提案手法では正しく識別できたものの中から，主観となるが対象物が識別しやすい画像を選択した．選択理由としては，人による識別が比較的容易な画像に対してモデルが誤認識した場合に信頼性が損なわれやすいためである．図 15 に示したサンプルは，画像全体のなかで対象物が小さく映っているような，人でも識別の難易度が高いと思われる画像を選択した．このような画像に対してモデルが大きい出力値 (尤度) で誤識別した場合，同じように信頼性が損なわれてしまうと考えられるためである．そのため図 15 に対しては，考察のためにモデルの出力値 (尤度) を追加している．

図 14 の結果については，対象物（label:airplane）の全体が写っている画像に対しての推論結果であり，いずれの結果も Normal は誤識別，提案手法では正しく識別できている例である．Normal では，対象物以外の背景情報にも強く注目している．対して，提案手法では，背景に対する強い注目は軽減され，対象物に反応していることが確認された．つまり注目箇所の誘導により，共起バイアスの影響を低減し，正しく識別することができたと考えられる．

一方で図 15 は，対象物 (label:Boat) が画像中にごくわずかしか映り込んでいない画像に対する可視化結果である．モデルの出力値を確認すると，sample1 の Normal については高い尤度で Boat と識別している．しかしモデルの注目箇所を確認すると Boat の特徴を捉えて識別しておらず，水面の特徴に強く反応して識別している．このような本質的な特徴を捉えていないにもかかわらず，高い出力値で識別するような挙動は，説明責任が必要とされるタスクにおいてモデルの信頼性が低下する要因となってしまう．一方で提案手法についても本質的な特徴について捉えられていないものの，尤度は極端に高いものではなくなっている．この理由は，提案手法は対象物以外の特徴は学習しにくいため，対象物を識別するための特徴が画像に無い場合は尤度も低くなりやすいと考えられる．しかし学習データにある対象物の特徴のみを学習するように誘導するため，入力画像のバリエーションが多いデータ等に対する推論においては識別性能が低下すると考えられる．

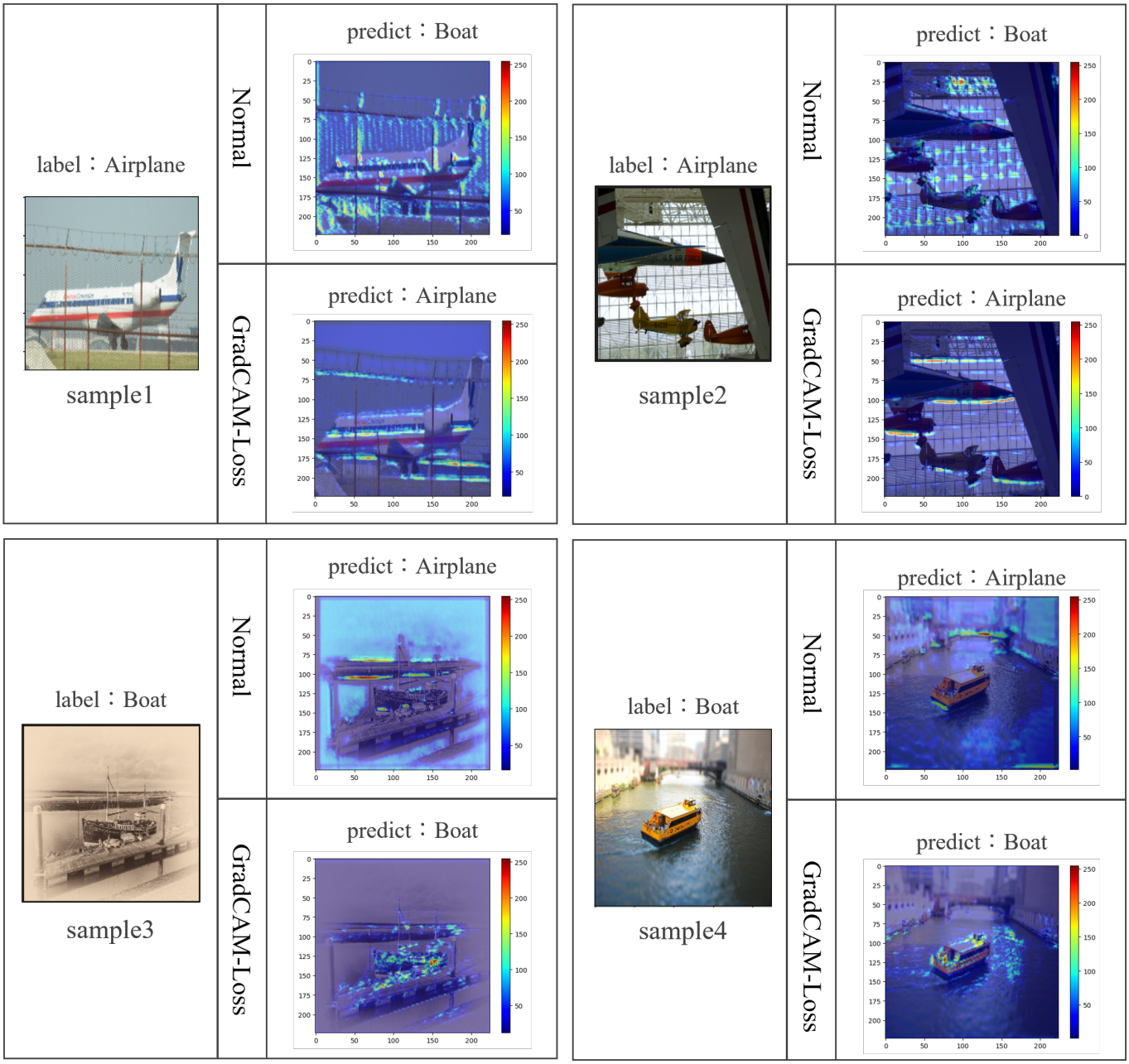


図 14: Airplane，Boat の二値分類結果に対する可視化 (提案手法のみ正解の画像例)

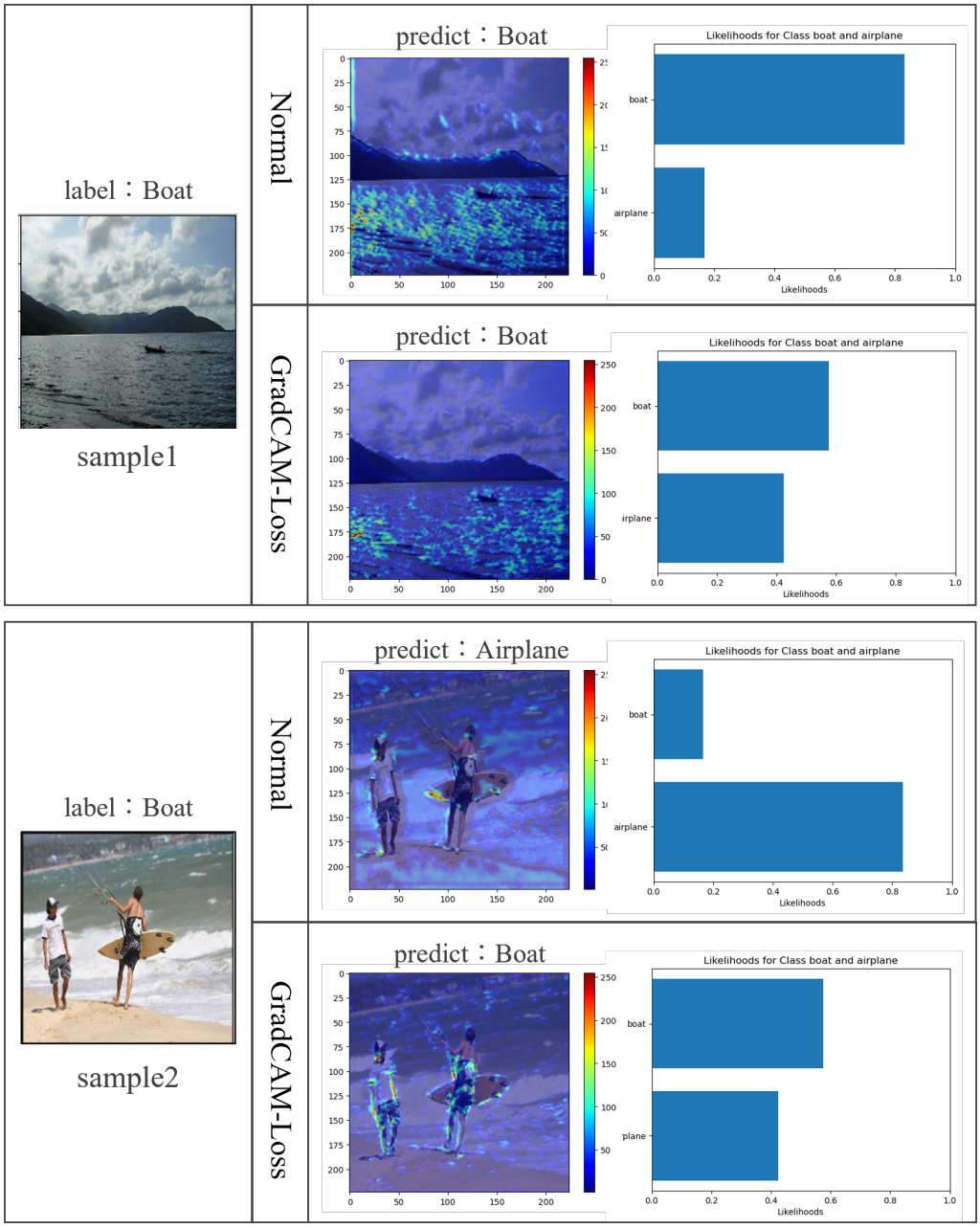


図 15: Airplane，Boat の二値分類結果に対する可視化 (対象物体の識別が困難な場合)

7　結論

7.1　結論

本論文では，鍛造品の外観検査の自動化を目的として，株式会社ゴーシュー提供のデータを用いた画像識別モデルの構築及びモデルの解釈性・信頼性向上のための損失関数の導入を提案した．

第 4 章では，機械学習手法の LightGBM 及びスクラッチ畳み込みニューラルネットワークモデル (MYModel)，事前学習済み VGG16 を用いてアンサンブル学習を行い，従来手法のルールベースモデルとの精度比較について述べた．手元のデータに対しては，アンサンブル学習を行うことで不良品の検出漏れを表す評価指標の Recall を 1.0 で達成しつつ，良品を不良品と誤識別する過検出を表す Precision が 0.71 で識別できることが分かった．また F1 値は 0.83 を達成しており，実運用における目標精度に達成することができた．

第 5 章では，CNN モデルの識別結果に対する解釈性・信頼性を向上させることを目標に，熟練工と呼ばれるベテラン作業員が目視検査を行う際に手がかりとしている製品特徴と，モデルの注目箇所との一致度を組み込んだ損失関数（Grad-CAM 損失）を提案した．熟練工の知見を二値画像で表現を行い，Grad-CAM を用いたモデルの注目箇所の可視化画像との一致度を算出した．一致度が高い場合は損失が低く，反対に一致度が低い場合には損失が高くなるようにすることで，注目箇所の誘導に効果があることが示された．

第 6 章では，提案損失をオープンデータの MS COCO データセットに適用した場合の効果について検証した．Boat と Airplane の二値分類タスクにおいて，提案損失により注目箇所が変化し, モデルの信頼性向上の効果が確認できた．

7.2　今後の課題

今後の課題としては，外観検査の自動化に実際に運用する場合に発生する様々な不具合モード及び多品種に対応できることである．本研究では代表製品の代表不具合である「欠肉」のみを対象としていたため，その他製品の欠肉画像やその他想定される不具合モードの画像についても学習を行う必要がある．また人による目視検査の場合は未知の不具合に対しても「違和感」によって検出できるが，教師あり学習の場合未知の画像に対して検出することができない点も課題である．この点については正常画像のみで学習する異常検知アプローチが必要と考えられる．

また，提案損失の一般性評価については，本研究では効果確認時のサンプリングが筆者の主観によって決定されている．今後は恣意性を排除したサンプリングによる定量的な評価が必要である．さらに，識別対象の本質的な特徴としてデータセットのアノテーション情報を利用したが，鍛造製品の識別のように，識別の難易度や本質的な特徴の設定は各種ドメインに依存される．そのため，一般画像認識の場合でも，どこまでの領域を本質的な特徴とするかの設定については議論の余地を残す．

謝辞

本研究にあたり，指導教員の飯山将晃教授には多大なご指導とご支援を頂きましたこと，心より感謝申し上げます．また岩山幸治准教授には貴重なご意見とご指摘をいただきましたことを深く感謝申し上げます．そして本研究を進めるためにデータを提供いただいた九州精鍛株式会社のご担当者の皆様と，社会人派遣元の株式会社ゴーシューの皆様には研究に関する様々なご助言をいただきましたこと，御礼申し上げます．最後に，データサイエンス研究科の同窓生の皆様，飯山研究室の皆様には様々な議論を交わしていただき，常に刺激をもらいながら研究に励むことができました．ここに感謝申し上げます．

付録

4.2.4 節にてアンサンブルモデルに VGG16 を選択して実験を行ったが，複数の深層学習モデルから選択した旨を述べた．表 8 に予備実験として行ったモデル毎の精度比較結果を示す．尚，精度比較時は validation データに対する F1 値を選定基準として用いた．

アンサンブルに用いるモデルとして，一般的には多様なモデルを用いるほうが良いとされているため，シンプルな畳み込みモデルとして MyModel(fc2 層)，事前学習済みモデルとして VGG16 を選択した．

表 8: 深層学習モデルの精度比較

	Precision	Recall	F-score
MyModel (fc1)	0.932	0.932	0.932
MyModel (fc2)	0.933	0.955	0.944
MyModel (fc3)	0.952	0.909	0.930
ResNet18	0.629	0.886	0.736
ResNet50	0.897	0.591	0.712
VGG16	0.949	0.841	0.892
VGG16(bn)	0.900	0.818	0.856
VGG19	0.854	0.932	0.891
VGG19(bn)	0.889	0.727	0.800
ViT(patch16)	0.756	0.773	0.764

参考文献

[1] 渡邊裕之. 自動車用部材向け外観検査技術. 電気製鋼, Vol. 85, No. 2, pp. 139–148, 2014.

[2] 北廣和雄. 外観品質保証－製品外観の完成度・信頼性を高める考え方と進め方－. 2014.

[3] 公益財団法人ちゅうごく産業創造センター：ものづくり企業の生産現場における検査. ものづくり企業の生産現場における検査の自動化促進可能性調査. 2016.

[4] 山崎達也, 福井暉斗ほか. 画像処理による鍛造部品欠陥検出の検討. 研究報告コンシューマ・デバイス & システム (CDS), Vol. 2018, No. 12, pp. 1–6, 2018.

[5] 西山龍貴, 山崎達也ほか. 画像処理による鍛造部品の外観不良検査手法における誤判定の抑制に関する検討. 研究報告コンシューマ・デバイス & システム (CDS), Vol. 2020, No. 17, pp. 1–4, 2020.

[6] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning Deep Features for Discriminative Localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[7] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual Explanations from Deep Networks via Gradient-based Localization. In *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[8] Jianyu Wang, Zhishuai Zhang, Cihang Xie, Yuyin Zhou, Vittal Premachandran, Jun Zhu, Lingxi Xie, and Alan L. Yuille. Visual Concepts and Compositional Voting. *CoRR*, Vol. abs/1711.04451, , 2017.

[9] WIRED. When It Comes to Gorillas, Google Photos Remains Blind. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/. 2023 年 9 月 18 日閲覧.

[10] Batta Mahesh. Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR).[Internet]*, Vol. 9, No. 1, pp. 381–386, 2020.

[11] Daniel Weimer, Bernd Scholz-Reiter, and Moshe Shpitalni. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP annals*, Vol. 65, No. 1, pp. 417–420, 2016.

[12] 神嶌敏弘. 転移学習. 人工知能, Vol. 25, No. 4, pp. 572–580, 2010.

[13] Yung-Kyun Noh Seunghyeon Kim, Wooyoung Kim and Frank C. Park. Transfer Learning for Automated Optical Inspection. 5 2017.

[14] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In *International conference on information processing in medical imaging*, pp. 146–157. Springer, 2017.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. *Advances in neural information processing systems*, Vol. 27, , 2014.

[16] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[17] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *science*, Vol. 313, No. 5786, pp. 504–507, 2006.

[18] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural computation*, Vol. 13, No. 7, pp. 1443–1471, 2001.

[19] Rong Yao, Chongdang Liu, Linxuan Zhang, and Peng Peng. Unsupervised Anomaly Detection Using Variational Auto-Encoder based Feature Extraction. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–7. IEEE, 2019.

[20] 柳井啓司. 一般物体認識における機械学習の利用. 電子情報通信学会技術研究報告; 信学技報, Vol. 110, No. 76, pp. 103–112, 2010.

[21] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

[22] Princeton University Stanford Vision Lab, Stanford University. ImageNet Large Scale Visual Recognition Challenge (ILSVRC). `https://www.image-net.org/index.php`. 2023年11月10日閲覧.

[23] Maoqing Tian, Shuai Yi, Hongsheng Li, Shihua Li, Xuesen Zhang, Jianping Shi, Junjie Yan, and Xiaogang Wang. Eliminating Background-Bias for Robust Person Re-Identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5794–5803, 2018.

[24] STANLEY BILESCHI LIOR WOLF. A critical view of context. In *nternational Journal of Computer Vision*, pp. 251–256, 2006.

[25] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The Role of Context for Object Detection and Semantic Segmentation in the Wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 891–898, 2014.

[26] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information fusion*, Vol. 58, pp. 82–115, 2020.

[27] Miruna A. Clinciu and Helen F. Hastie. A Survey of Explainable AI Terminology. pp. 8–13, 2019.

[28] Xi Yang, Bojian Wu, Issei Sato, and Takeo Igarashi. Directing DNNs Attention for Facial Attribution Classification using Gradient-weighted Class Activation Mapping. In *CVPR Workshops*, pp. 103–106, 2019.

[29] Yi He, Xi Yang, Chia-Ming Chang, Haoran Xie, and Takeo Igarashi. Efficient Human-in-the-loop System for Guiding DNNs Attention. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 294–306, 2023.

[30] Antonio Torralba and Alexei A Efros. Unbiased Look at Dataset Bias. In *CVPR 2011*, pp. 1521–1528. IEEE, 2011.

[31] Quanshi Zhang, Wenguan Wang, and Song-Chun Zhu. Examining CNN Representations With Respect To Dataset Bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, 2018.

[32] 湯川伸樹. 鍛造. 軽金属, Vol. 58, No. 1, pp. 38–45, 2008.

[33] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in neural information processing systems*, Vol. 30, , 2017.

[34] Karen Simonyan and Andrew Zisserman. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. *arXiv preprint arXiv:1409.1556*, 2014.

[35] 坂野鋭. パターン認識における主成分分析-顔画像認識を例として. 2001.

[36] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755. Springer, 2014.