

クラウド技術を活用したリモート監視に関する研究開発

—CNNを用いた工業部品の欠陥画像分類—

渡辺 博己 浅井 博次 生駒 晃大 棚橋 英樹

A Study on Remote Monitoring by Using Cloud-Based Technologies - Defect Classification of Industrial Parts Using Convolutional Neural Networks -

Hiroki WATANABE Hirotsugu ASAI Akihiro IKOMA Hideki TANAHASHI

あらまし 本研究では、既存の外観検査装置の欠陥分類機能を高度化するため、検査装置で取得した欠陥画像を、畳み込みニューラルネットワーク（CNN）により欠陥種別に分類するシステムを試作した。システムに実装する学習モデルについては、既存モデルをベースに改良し、学習方法を検討することにより、新たな学習モデルを構築した。また、検査装置で取得した欠陥画像について、学習済みモデルの分類精度を検証する実験を行い、欠陥画像の分類が可能であることを確認した。

キーワード 深層学習、畳み込みニューラルネットワーク（CNN）、工業部品、欠陥画像、画像分類

1. はじめに

検査工程においては、目視検査の負担を軽減するために画像検査装置（以下、検査装置）の導入が進められている。検査作業において、人は検査対象物を目で見て、脳で検査するため、欠陥の認識能力が高く、汎化能力も高い。しかし、人は長時間同じ作業を繰り返すことが困難であるとともに、個人により検査能力にバラツキがある。これに対し従来の検査装置は、カメラで画像を取得し、コンピュータで画像処理により設定されたパラメータに基づいて欠陥を検出・分析する。そのため、ただ単調に指示された作業をひたすら繰り返すことが得意であり、同一性能の別の装置があれば、同一の検査能力を実現することが容易である。しかし、事前に人から与えられた設定以外の作業は困難であるため、人に比べて認識能力や汎化能力が劣り、未だに目視作業に頼らざるを得ない検査対象物が数多く存在する。

一方、深層学習に代表されるAI技術は、これまでの検査業務を劇的に革新することが期待されている。特に、畳み込みニューラルネットワーク^[1]（以下、CNN）に代表される機械学習手法は年々進展し、今や人を超える認識性能を持つとまで言われている^[2,3]。しかし、工業部品の不良率は一般的に低く、欠陥画像を収集することは容易ではない。そのため、深層学習を用いた検査技術においては、欠陥画像データを学習し、欠陥種別を予測するCNNではなく、良品画像データを学習し、良品データとの乖離を検出するAutoEncoder^[4]を利用するケースが多い。

こうした中、自動車部品の検査を請け負う岐阜県内の中小企業から、ある検査対象物について、従来使用している検査装置では検出・分析機能に問題があり、人による目視検査作業の負担が大きいため、AI技術により装置を改善したいとの要望があった。そこで、本研究では、企業には比較的多くの欠陥品が集まり、検査装置により欠陥画像の収集が容易であることから、AI技術では一般的に行われていない欠陥種別の特定について取り組むこととし、検査装置の認識性能向上を目的として、CNNを用いた欠陥画像分類技術を開発することにした。本稿では、開発する画像分類技術について試作システムの構成を述べるとともに、CNNによる欠陥画像分類について実験を行ったので、その結果について報告する。

2. 欠陥画像分類システムの試作

2.1 既存画像検査装置の概要

検査装置における検査対象は円筒部品であり、検査では、部品を回転させた状態で、回転機構と同期の取れたラインスキャンカメラにより部品全周の画像を取得する。この時、1つの対象物を4回転させ、各回転において異なる方向から照明光を照射し、欠陥画像を取得する。

取得した画像については、欠陥の有無を分析し、欠陥があれば、欠陥部分が画像中心となるよう256×256の解像度で、特定のフォルダにビットマップ形式の画像ファイルとして保存する。欠陥の種別によっては、全ての照明光により検出される欠陥もあれば、そうでない欠陥もある。また、

欠陥の分布・大きさによっては、複数の画像で同一の欠陥が記録されることもある。そのため、欠陥の状態により取得される画像数が異なり、1つの検査対象物について複数の欠陥画像ファイルがフォルダに蓄積される。

図1に検査装置で取得した欠陥12種類、正常3種類の画像例を示す。なお、検査装置においては、欠陥画像のみが保存され、異物や汚れの付着した画像以外の正常画像については保存されない仕様となっている。そのため、正常画像を取得する際には、検査装置の欠陥判定しきい値を調整することにより取得した。

2. 2 試作システムの構成

欠陥画像分類技術の導入における企業側の条件としては、検査装置の機能はそのまま残し、試作システムを後付けすることであった。そのため、欠陥画像ファイルが保存されるフォルダを共有化し、検査装置と試作システムをネットワーク接続する構成とすることで、CNNによる欠陥画像分類が可能なシステムを試作した。

図2に構築した試作システムの概要を示す。システムは、対象フォルダを絶えず監視し、検査装置により画像ファイルが保存され、フォルダのファイル数が増加した時に、直ちに増加分のファイル名を取得する。次に、取得したファイル名の画像ファイルからデータを取得し、CNNにより構築した学習モデルに基づいて、欠陥画像の種別を推定する。推定結果については、ディスプレイ上に画像データ、ファイル名とともに欠陥種別を表示する。なお、対象フォルダについては、クラウドストレージ上のフォルダであっても監視可能であり、クラウドサービスとして画像分類機能を提供することも可能である。現状では、画像毎に欠陥種別を出力しているが、最終的には、各対象物から取得した複数枚の画像を用いて予測結果を統合し、対象物毎に欠陥種別に分類する機能を、今後、システムに実装する予定である。

3. 欠陥画像分類モデルの構築

3. 1 分類モデルの検討

CNNの代表的なモデルにはAlexNet^[5]やVGG^[6]、GoogLeNet^[7,8]、ResNet^[9]等があるが、本研究では、単純な構造であるため、改良が容易なVGGの16層モデルをベースとして、分類モデルを検討した。

中小企業がCNNを利用するにあたっては、PCの導入コストを低減するために、学習や分類に係る処理において、認識性能を確保しつつ、計算コストを抑えたモデル構成であることが望ましい。そのため、分類モデルの構成では、まず、VGGの全結合層へのベクトル生成時に使われるFlattenをGlobal Average Pooling^[10]（以下、GAP）に変更したモデルを構成し、モデルの軽量化を図った。また、本研究で分類する欠陥種別は15であり、大規模なデータを取り扱うことはないことから、全結合層を出力層の1層とするモデルを構成した。さらに、畳み込み層が3回連続するブロックに

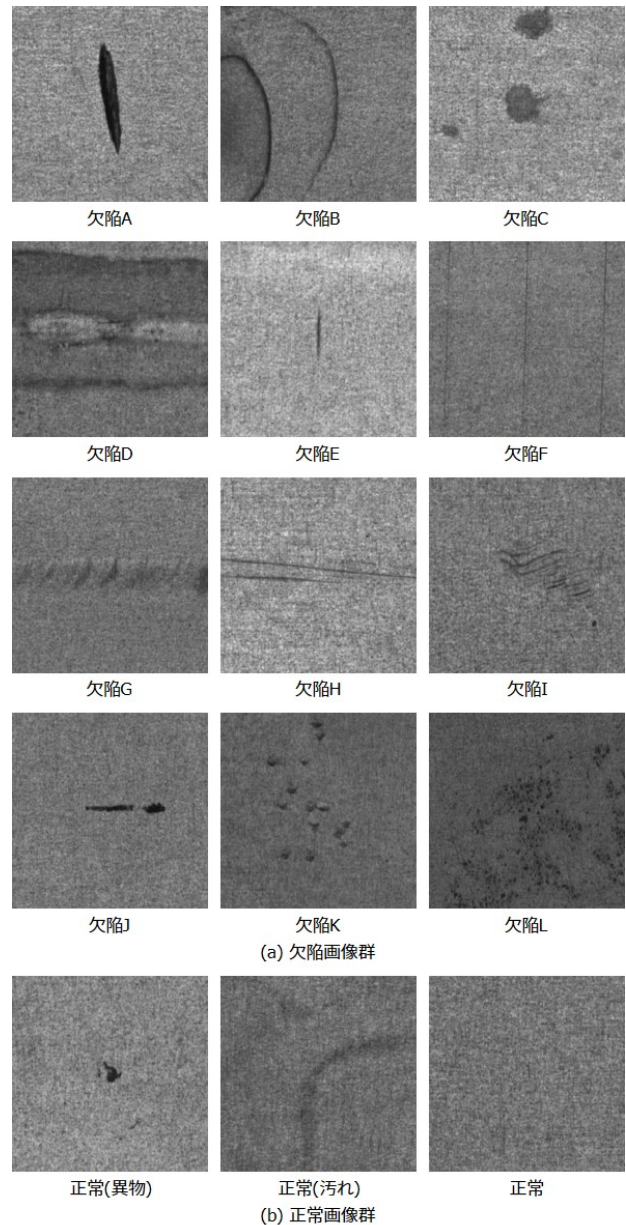


図1 欠陥種別毎の画像例

については、1層目のフィルタサイズを1×1にし、2層目と3層目の出力を重ねる形で統合するモデル（以下、基本モデル）を構成した。表1に、VGGをベースとした各モデルの構成を示す。なお、GAPについては、全結合層でのパラメータ数や過学習を解消するとされており、全結合層をGAPに置き換えた場合においても認識性能が高まることが報告されている^[10]。

次に、基本モデルのGAPの直前のPooling層までを並列に配置し、得られた2つのベクトルを統合した後、1×1の畳み込み層を経てGAPに入力するモデル（以下、双入力モデル）を構成した。2.1節で述べたとおり、欠陥部分は概ね画像中心となるよう取得され、その大きさは種別により異なる。そのため、画像全体と中心付近との2つの画像データを入力データとすることで、大域的な特徴と局所的な特徴が得

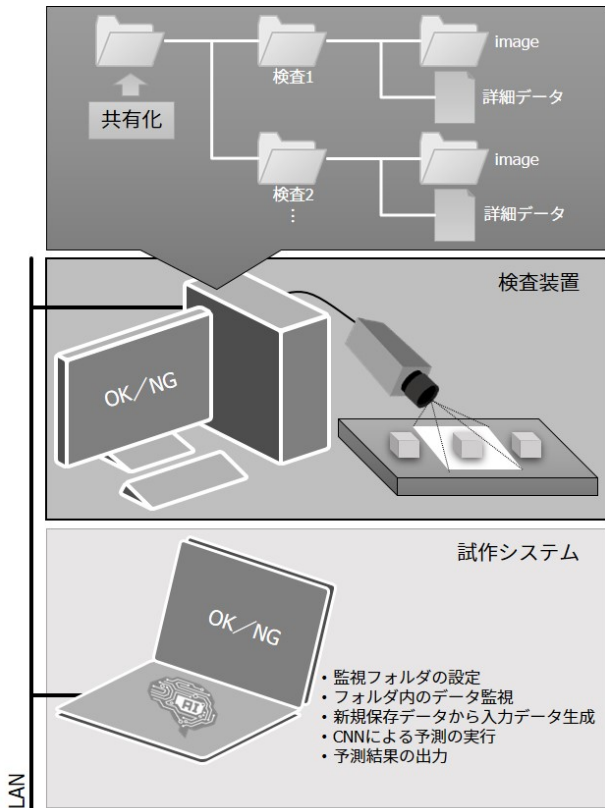


図2 試作システムの概要図

表1 VGGをベースとした各モデルの構成

VGG16	Multi-Conv +GAP+FC	Multi-Conv +GAP	基本モデル
Input128-1	Input128-1	Input128-1	Input128-1
Conv3-64	Conv3-64	Conv3-64	Conv3-64
Conv3-64	Conv3-64	Conv3-64	Conv3-64
Max Pooling	Max Pooling	Max Pooling	Max Pooling
Conv3-128	Conv3-128	Conv3-128	Conv3-128
Conv3-128	Conv3-128	Conv3-128	Conv3-128
Max Pooling	Max Pooling	Max Pooling	Max Pooling
Conv3-256	Conv3-256	Conv3-256	Conv1-256
Conv3-256	Conv3-256	Conv3-256	Conv3-256
Conv3-256	Conv3-256	Conv3-256	Conv3-256
Max Pooling	Max Pooling	Max Pooling	Concatenate
Conv3-512	Conv3-512	Conv3-512	Max Pooling
Conv3-512	Conv3-512	Conv3-512	Conv1-512
Conv3-512	Conv3-512	Conv3-512	Conv3-512
Max Pooling	Max Pooling	Max Pooling	Conv3-512
Conv3-512	Conv3-512	Conv3-512	Concatenate
Conv3-512	Conv3-512	Conv3-512	Max Pooling
Conv3-512	Conv3-512	Conv3-512	Conv1-512
Max Pooling	Max Pooling	Max Pooling	Conv3-512
Flatten	Global Average Pooling	Global Average Pooling	Conv3-512
FC-4096	FC-4096	FC-15	Concatenate
FC-4096	FC-4096	Soft-max	Max Pooling
FC-15	FC-15		Conv1-512
Soft-max	Soft-max		Conv3-512

られることを想定している。表2に双入力モデルの構成を示す。なお、入力データには、元の画像サイズを128×128にリサイズしたグレースケール画像を使用するとともに、中心付近の画像については、元の画像から128×128のサイズで中心部分を切り取った画像を使用する。また、2つの入

表2 双入力モデルの構成

双入力モデル	
(画像全体)	(中心付近)
Input128-1	Input128-1
Conv3-32	Conv3-32
Conv3-32	Conv3-32
Max Pooling	Max Pooling
Conv3-64	Conv3-64
Conv3-64	Conv3-64
Max Pooling	Max Pooling
Conv1-128	Conv1-128
Conv3-128	Conv3-128
Conv3-128	Conv3-128
Concatenate	Concatenate
Max Pooling	Max Pooling
Conv1-256	Conv1-256
Conv3-256	Conv3-256
Conv3-256	Conv3-256
Conv3-256	Conv3-256
Concatenate	Concatenate
Max Pooling	Max Pooling
Conv1-256	Conv1-256
Conv3-256	Conv3-256
Conv3-256	Conv3-256
Concatenate	Concatenate
Max Pooling	Max Pooling
Concatenate	Concatenate
Conv1-256	Conv1-256
Global Average Pooling	Global Average Pooling
FC-15	FC-15
Soft-max	Soft-max

力データを使用することによりパラメータ数が増加することから、各畳み込み層のチャンネル数を半数とすることで、処理コストを軽減することにした。

3. 2 実験と結果

実験で使用する欠陥種別毎の画像データ数を表3に示す。学習フェーズでは、学習データとして各種別240枚、学習時の評価データとして60枚を、これらの画像データからランダムにサンプリングし、学習を繰り返すエポック数を500に設定してモデルの学習を行った。なお、評価データの正解率が200エポックの間に更新されない場合は、学習フェーズを終了するよう設定し、バッチサイズは60とした。予測フェーズでは、モデルの性能を評価するために、表3に示す全ての画像データを用いて、欠陥12クラス、正常3クラスの計15クラスに分類した。

まず、基本モデルの認識性能を評価する実験では、表1に示す各モデルについて、学習・予測フェーズを繰り返す実験を10回試行し、各クラスの適合率、及び再現率の平均値を算出した。実験に使用したモデルのパラメータ数、学習が収束した平均エポック数、学習に要した1エポックあたりの平均処理時間、及び分類結果から算出した平均適合率、再現率を表4に示す。なお、各試行における各モデルへの入力データは全て同一であり、実験で使用したPCの主な仕様は、CPUがCore i7 6700K、メモリが64GB、GPUがGeForce GTX 1080 Ti×2である。結果より、パラメータ数が少ないにも関わらず、基本モデルが最も高い認識性能を持つことが示された。また、パラメータ数を削減することにより、学習に要する処理時間も短縮することができた。

次に、双入力モデルの認識性能を評価するために、基本モデルの各畳み込み層のチャンネル数を半減したモデルを単入力モデルとし、画像全体を入力データとするモデルを

表3 欠陥種別毎のデータ数

種別	データ数	種別	データ数	種別	データ数
欠陥A	677	欠陥B	449	欠陥C	470
欠陥D	789	欠陥E	561	欠陥F	2,256
欠陥G	1,240	欠陥H	319	欠陥I	695
欠陥J	468	欠陥K	654	欠陥L	404
正常(異物)	752	正常(汚れ)	361	正常	503

表4 基本モデルの評価実験の結果

モデル	VGG16	Multi-Conv +GAP+FC	Multi-Conv +GAP	基本モデル
パラメータ数	65.1M	33.7M	14.7M	11.7M
平均エポック数	308.2	426.7	349.3	109.1
平均学習時間	10.8秒/epoch	9.2秒/epoch	8.4秒/epoch	7.9秒/epoch
平均適合率	0.8214	0.8627	0.8901	0.8928
平均再現率	0.8463	0.8832	0.9076	0.9086

単入力（大域）モデル，中心付近を入力データとするモデルを単入力（局所）モデルとして，各モデルの認識性能を評価する実験を行った．実験結果を表5に示す．なお，実験に関する諸条件については，先に述べた実験と同一である．

結果より，単入力（大域）モデルにおける平均適合率，再現率が，基本モデルを上回っていることから，チャンネル数の半減は，認識性能の向上に効果があったと判断できる．しかし，双入力モデルの平均適合率，再現率は単入力（大域）モデルを下回る結果となった．この原因としては，各単入力モデルの特徴が相殺され，調和的に学習が行われていることが予想される．表6に各単入力モデルにおける欠陥種別毎の平均再現率を示すが，欠陥種別により値の大小が交差していることから，大域的，及び局所的な画像データから欠陥種別を推定する方法は有効であると考えられる．そこで，学習方法を見直すことで，再度，双入力モデルの認識性能を評価することにした．

4. モデルの学習方法の検討

4.1 学習方法の検討

双入力モデルについては，学習フェーズにおいて，効果的な学習が行われていないことが考えられる．そのため，2つの異なる方法で学習することにより，モデルの認識性能を比較する実験を行った．

1つ目の学習方法としては，先に学習した2つの単入力モデルの学習済みパラメータを利用する方法で，転移学習と呼ばれる方法を用いる．手順としては，まず，単入力（大域）モデル，単入力（局所）モデルを順に学習する．次に，各単入力モデルのGAPまでのパラメータを双入力モデルに転移する．最後に，各単入力モデルから転移したパラメータは更新せず，2つのネットワークを結合した層以降のみを学習する．

2つ目の方法は，GoogLeNetのように複数の出力を評価す

表5 双入力モデルの評価実験の結果

モデル	基本モデル	単入力(大域)モデル	単入力(局所)モデル	双入力モデル
パラメータ数	11.7M	2.9M	2.9M	6.1M
平均エポック数	110.3	99.2	72.0	90.9
平均学習時間	7.8秒/epoch	3.7秒/epoch	3.7秒/epoch	7.2秒/epoch
平均適合率	0.8921	0.8985	0.8766	0.8915
平均再現率	0.9092	0.9138	0.8967	0.9099

表6 単入力モデルにおける平均再現率

種別	単入力(大域)モデル	単入力(局所)モデル
欠陥A	0.8245	0.8338
欠陥B	0.9555	0.9316
欠陥C	0.9640	0.9540
欠陥D	0.9403	0.8777
欠陥E	0.9668	0.9681
欠陥F	0.9923	0.9881
欠陥G	0.8984	0.8619
欠陥H	0.9461	0.9524
欠陥I	0.7830	0.7291
欠陥J	0.9534	0.9380
欠陥K	0.8628	0.8722
欠陥L	0.8443	0.8383
正常(異物)	0.8774	0.8369
正常(汚れ)	0.9304	0.9067
正常	0.9671	0.9625

る方法で，双入力モデルについて，画像全体，及び中心付近を処理する各ネットワークに単入力モデルと同様に，GAP，出力層を加え，3つの出力誤差を同時に最小化する．つまり，1つの出力については，双入力モデルによる出力誤差を最小化し，残りの2つの出力については，各単入力モデルの出力誤差をそれぞれ最小化するよう学習する．なお，3つの出力からそれぞれ分類結果が得られるが，各単入力モデルの分類結果については評価しない．

4.2 実験と結果

実験においては，3.2節の実験と同様に，前節の学習方法により双入力モデルの学習を行った．また，本実験においても，学習・予測フェーズを10回繰り返す，各クラスにおける適合率，及び再現率の平均値を算出した．実験結果を表7に示す．なお，表中，従来の学習方法を通常学習，パラメータを転移する方法を転移学習，複数の出力誤差を同時に最小化する方法をマルチ学習として記述する．

結果より，通常学習よりも転移学習，マルチ学習による方法が平均適合率，再現率が高くなり，マルチ学習による方法が最も高くなった．マルチ学習については，既にGoogLeNetにより実証されている方法であるが，マルチ学習は転移学習と同様に，各単入力モデルの特徴を踏まえた状態で学習していると考えられる．そのため，複数の入力データを用いる場合は，個々のネットワークの出力誤差を評価可能なモデルを構成することが有効であると考えられる．なお，転移学習については，パラメータ数が少ない

表7 学習方法の評価実験の結果

学習方法	通常学習	転移学習	マルチ学習
パラメータ数	6.1M	0.3M	6.1M
平均エポック数	73.2	106.0	67.9
平均学習時間	7.2秒/epoch	3.4秒/epoch	7.5秒/epoch
平均適合率	0.8923	0.9050	0.9071
平均再現率	0.9105	0.9203	0.9222

ため平均学習時間は短いですが、事前に2つの単入力モデルを学習する必要があることから、全体の学習時間は他の学習方法と比較すると倍以上となる。

5. 欠陥画像分類モデルの検証

最も認識性能が高くなったマルチ学習による双入力モデルについて、画像サイズ、バッチサイズを変更した場合の認識性能を検証した。画像サイズ、バッチサイズは、GPUが実装するメモリ容量の影響を大きく受けるため、これらの値を変更した場合の認識性能への影響を把握することは、深層学習に関する技術をシステムに実装する上で重要である。

画像サイズについては、大域画像のリサイズ後の画像サイズと局所画像の切り出しサイズを同一とし、64~192ピクセルまで32ピクセルずつ変更して、3.2節の実験と同一条件のもと、学習・予測フェーズを10回繰り返し、平均認識率を求める実験を行った。結果を図3に示す。なお、2つの入力データを用いる効果を検証するために、単入力(大域)モデルについても平均認識率を求め、これらの推移を比較した。その結果、160×160までは徐々に認識精度が高くなっていくが、それ以降は横ばいとなる傾向にあることが分かった。また、画像サイズが小さくなるほど2つの入力データを用いる効果が高くなることも確認できた。

次に、大域・局所画像のサイズを128×128、及び160×160として、バッチサイズを30~300まで30ずつ変更し、3.2節の実験と同一条件のもと、学習・予測フェーズを各画像サイズについて10回繰り返し、平均認識率の推移を比較した。結果を図4に示す。バッチサイズの違いによる認識性能への影響は、画像サイズを変更することよりも小さいが、学習データ数に対して剰余が発生するバッチサイズを設定した時は、認識性能が低下する傾向にあることが明らかとなった。

これらの実験より得られた結果から、表8に、画像サイズを160×160、バッチサイズを300とした時の、欠陥種別に対する平均適合率、再現率を示す。また、欠陥種別を欠陥と正常の2クラスとした場合の平均適合率、再現率も示す。種別間で認識精度にバラ付きはあるものの、全体の平均適合率、再現率は、91.62%、93.05%であった。正常・欠陥クラスの場合は、正常クラスの適合率が若干低いですが、全体の平均適合率、再現率は、96.13%、97.84%と高い認識性能が得

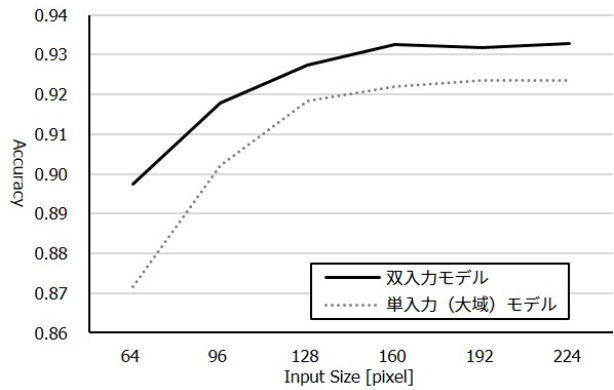


図3 画像サイズによる認識性能の評価

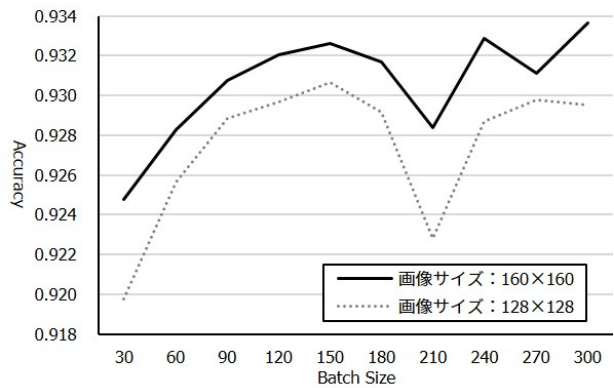


図4 バッチサイズによる認識性能の評価

表8 欠陥種別毎の分類結果

種別	平均適合率	平均再現率	種別	平均適合率	平均再現率
欠陥A	0.8634	0.8579	欠陥	0.9946	0.9861
欠陥B	0.9379	0.9728			
欠陥C	0.9331	0.9768			
欠陥D	0.9703	0.9455			
欠陥E	0.9270	0.9615			
欠陥F	0.9980	0.9941			
欠陥G	0.9885	0.9185			
欠陥H	0.7819	0.9574			
欠陥I	0.9112	0.8145			
欠陥J	0.9268	0.9553			
欠陥K	0.8959	0.8884			
欠陥L	0.8968	0.8682	正常	0.9280	0.9708
正常(異物)	0.9074	0.9105			
正常(汚れ)	0.9208	0.9608			
正常	0.8838	0.9749	全体	0.9613	0.9784
全体	0.9162	0.9305			

られた。得られた結果については、試作システムで学習済みモデルを利用する際には、1つの対象物に対して複数の画像から分類結果を判定するよう実装することから、多数決等による誤分類キャンセル処理等が行えるため、十分な結果であると考えている。なお、分類モデルについては、試作システムでの検証を進めながら、ネットワーク構成も

含めて検討することとしているため、今後も認識性能が高くなるよう改良を図る予定である。

6. まとめ

本研究では、従来の検査装置の認識性能の向上を図るために、CNNを用いた画像分類システムを試作した。また、システムに実装する学習済みモデルを構築するために、VGGをベースとした双入力モデルを構成するとともに、学習方法の違いによるモデルの認識性能について検証実験を行った。

試作したシステムでは、検査装置とネットワーク経由で接続することにより、検査装置の欠陥抽出画像が保存されるフォルダをリアルタイムに監視し、CNNによる画像分類処理を行うことで、検査対象物の欠陥種別を推定することが可能となった。また、構築した学習モデルについては、正常・欠陥の2クラスに分類する検証実験において、平均適合率として正常・欠陥の順に92.80%、99.46%、平均再現率として97.08%、98.61%の認識性能が得られた。

今後は、更に認識性能を向上させるとともに、汎化能力を高めるために、欠陥画像の収集を継続し、画像分類技術の検討を進める。また、既存の検査装置と同時に、試作システムを稼働させた検査業務を実施し、システムの検証実験を行なうとともに、システムの運用上の課題を抽出し、システムの改良を図る予定である。

文 献

- [1] Y. LeCun et al. : “Gradient-based learning applied to document recognition”, In Proc. of the IEEE, vol.86, no.11, pp.2278-2324 (1998)
- [2] S. F. Dodge and L. J. Karam : “A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions”, In CoRR, abs/1705.02498 (2017)
- [3] R. Geirhos et al. : “Comparing deep neural networks against humans: object recognition when the signal gets weaker”, In CoRR, abs/1706.06969 (2017)
- [4] G. E. Hinton and R. Salakhutdinov : “Reducing the dimensionality of data with neural networks”, Science, vol.313, no.5786, pp.504-507 (2006)
- [5] A. Krizhevsky, I. Sutskever and G. E. Hinton : “ImageNet Classification with Deep Convolutional Neural Networks”, In Proc. of NIPS, pp.1097-1105 (2012)
- [6] K. Simonyan and A. Zisserman : “Very Deep Convolutional Networks for Large-Scale Image Recognition”, In CoRR, abs/1409.1556 (2014)
- [7] C. Szegedy et al. : “Going deeper with convolutions”, In CoRR, abs/1409.4842 (2014)
- [8] C. Szegedy et al. : “Rethinking the Inception Architecture for Computer Vision”, In CoRR, abs/1512.00567 (2015)

- [9] K. He et al. : “Deep Residual Learning for Image Recognition”, In CoRR, abs/1512.03385 (2015)
- [10] M. Lin, Q. Chen and S. Yan : “Network in network”, In CoRR, abs/1312.4400 (2014)