

学位論文の要旨

機械学習に基づく画像認識・動画認識・動画解釈に関する研究
Research on Image Recognition, Video Recognition, and Video Interpretation
Based on Machine Learning

氏名 中島 彩奈 印

近年、人工知能(Artificial Intelligence: AI)を用いたシステムは様々な業界で使われている。AIは1956年に誕生し、二度のブームを経て現在第三次ブームとなっている。第一次ブームは1950年代後半から始まり、「推論」や「探索」と呼ばれる技術によりパズルゲームなど明確なルールが存在する問題に対して高い性能を発揮した。その後、1980年代には知識をコンピュータに入力し推論を行うことにより、より高度な推論の可能性があると考えられ第二次AIブームとなった。専門的な知識を蓄え、専門家のように振舞えることからエキスパートシステムと呼ばれ、医療や金融などの多くの分野で採用された。しかし、入力されていない常識的な知識を持たず、一つ一つ一般常識を入力しきることの難しさにより、それ以上には発展しなかった。これら二度のブームが過ぎ、2000年頃の機械学習の登場により現在は第三次AIブームとなっている。機械学習とはAIが自ら学習する仕組みのことである。機械学習では特徴量の入力が必要であり、どのような特徴量を用いて判断を行うかが機械学習の精度に影響する。この特徴量の入力は人の入力によるものであったが、近年のディープラーニング(Deep Learning)の発展により、コンピュータが自動で特徴量を抽出できるようになった。これにより、第一次・第二次AIブームでは実現できなかった問題も実現可能となった。

本論文では、これまでAIの実用化が困難であった産業界における認識問題をテーマとして取り上げ、Deep Learningによる解決を目指す。その方法として、視覚情報を基にした画像認識技術を適用する。これらの特徴の抽出には事前に大量のデータをコンピュータに学習させることが不可欠である。学習により大量のデータから特徴を抽出することで、その特徴をもとにした分類が可能となるが、その詳細は不透明であり、人にも分からない特徴量を導き出して学習していることが多い。また、大量のデータがあれば良いというわけではなく、データの質も認識精度に大きく関わるということが明らかになり、データの質も重要視されるようになった。そのため、機械学習を用いた認識技術では、目的を明確にし、それを実現するためにどのようなデータが必要かを考えてデータを用意しなければならない。

本稿は以下の3つのテーマについて実践した内容を説明し、機械学習に基づく認識技術を適用するために、目的に応じた学習データの作成を行い、問題解決へのアプローチ方法

とその精度について述べる。

(1) 外観検査の自動化

概要： 工業製品の外観検査の自動化を目標とし、検査性能の上昇と人の介入を最小限に抑える検査方法について検討した。筆記具部品の検査においては、不良品発生率の低さから十分な欠陥画像数を確保できないという問題を解決するため、欠陥と背景を合成して欠陥画像数を増大させることで、検出率・的中率が上昇する結果を得た。化粧品ボトルの検査では欠陥の大きさによって良品か不良品か決まるため、物体認識後にセグメンテーションアルゴリズムを併用し、欠陥の大きさによる判定の可能性を見出した。

(2) 害獣認識

概要： **Deep Learning** による物体認識では 100%の精度を保証することは難しく、より信頼性のあるシステム開発のため、物体認識と動体検出のアルゴリズムの組み合わせによる個体数管理のための害獣認識システムを提案した。物体認識と動体検出にはそれぞれ検出が困難なデータがあるが、二つの認識方法を相補的に組み合わせることでそれぞれのデメリットを補い合うことができ、人の負担軽減のための害獣認識手法の有用性を示した。

(3) エッジ AI プロセッサを用いた行動の分類

概要： 監視カメラ等から得られる動画に対し、行動認識技術を用いた解析により、人の行動の分類を行った。リアルタイムでの実行処理を行うシステム実現を目的とし、エッジ AI プロセッサ上で骨格情報の抽出を行い、機械学習モデルの実装に対する処理性能、消費電力などを比較した。また、エッジ AI プロセッサから得た骨格情報に対し、**SOM** を用いた分類を行い、基本行動の認識が可能であることを示した。

上記 3 つのテーマに示す通り、機械学習を用いた認識技術の適用とシステムの信頼性確保のため、各テーマの目的を明確にし、その問題解決の手法を推進した。また、各々のステップにて学会発表を、研究会発表や論文投稿を行い、学術的な成果を挙げることもできた。

学 位 論 文 の 要 旨

機械学習に基づく画像認識・動画認識・動画解釈に関する研究

Research on Image Recognition, Video Recognition, and Video Interpretation
Based on Machine Learning

氏 名 中島 彩奈 印

In recent years, Artificial Intelligence (AI) systems have been used in a variety of industries. AI was born in 1956, and after two booms, it is now in the third boom. The first boom started in the late 1950s, and the techniques called "inference" and "search" demonstrated high performance for problems with clear rules, such as puzzle games. In the 1980s, knowledge was input into computers for inference, and the possibility of advanced inference led to the second AI boom. They were called expert systems because of the specialized knowledge they provided, and were adopted in many fields such as medicine and finance. However, the system did not have any commonsense knowledge that had not been input, and it did not develop further due to the difficulty of inputting common sense one by one. These two booms have passed, and with the development of machine learning in the 2000s, we are now in the third AI boom. Machine learning is a system in which AI learns by itself. Machine learning requires the input of features, and the accuracy of machine learning is affected by what features are used to make decisions. The input of these features used to be done by human input, but with the recent development of deep learning, it has become possible for computers to automatically extract features. As a result, problems that could not be realized in the first and second AI booms can now be realized.

In this paper, we focus on recognition problems in industry, where the application of AI has been difficult, and aim to solve them using Deep Learning. As a method, we apply image recognition technology based on visual data. To extract these features, it is essential to let the computer learn a large amount of data in advance. By extracting features from a large amount of data through learning, it is possible to classify the data based on those features. However, the details of the learning process are unclear, and we often learn by deriving features that are unknown even to humans. In addition, it is not enough to have a large amount of data. The quality of the data is also very important because it is related to the recognition accuracy. Therefore, in recognition technology using machine learning, we must prepare data by clarifying the objective and

considering what kind of data is needed to realize it.

This paper has been practiced on three topics. To apply recognition technology based on machine learning, we create training data according to the purpose, and describe our approach to problem solving and its accuracy.

(1) Automation of Appearance Inspection

abstract: Aiming at the automation of the appearance inspection of industrial products, we studied an inspection method that increases the inspection performance and minimizes human intervention. In the inspection of pen parts, there is a problem that enough defect images cannot be secured due to the low defective rate. As a method to solve this problem, we increased the number of defect images by composing defects and backgrounds and obtained results that increased the detection ratio and target ratio. In the inspection of cosmetic bottles, the size of the defect determines whether the bottle is OK or NG. Therefore, a segmentation algorithm is used in combination with object recognition to determine the size of defects.

(2) Wild Animal Recognition

abstract: To develop a more reliable system, we proposed a wildlife recognition system for the management of populations based on a combination of object recognition and motion detection algorithms. By combining the two detection methods, the disadvantages of each can be compensated for, demonstrating the usefulness of the vermin recognition method for reducing the human task.

(3) Classification of Human Actions Using an Edge AI Processor

abstract: We classified human actions by analyzing videos obtained from surveillance cameras using action recognition algorithms. The skeleton information was extracted on the edge AI processor, and the processing performance and power consumption were compared. In addition, we used SOM to classify the obtained skeletal information, and we showed that it is possible to recognize basic actions.