【学位論文審査の要旨】

機械翻訳（MT）は，ある言語から別の言語へテキストを翻訳するタスクである．ニューラル機械翻訳（NMT）は，Google 翻訳などの多くのオンライン翻訳サービスにも導入されているアプローチで，高い翻訳性能を持つ．NMT の核となるのは計算機による言語意味理解で，大量のテキストを用いた学習によってある原言語の文の意味表現を理解し，それを元に目的言語の文を生成することができる．

これに対し，人間は言語，視覚，聴覚といった複数のモダリティを同時に組み合わせて活用することが可能である．NMT も言語だけでなく，他にも補助的なモダリティを取り入れることで，様々な側面から人間レベルの理解に近づけることが可能であると考えられる．

複数のモダリティを組み合わせる手法として，マルチモーダルニューラル機械翻訳（MNMT）は，NMT を拡張し，画像情報を利用して，画像と対になった原言語文を目的言語文に翻訳するものである．画像という補助的なモダリティを用いることで，意味が曖昧な単語や文法的な性別（例えばドイツ語やフランス語の男性名詞や女性名詞）など，テキストの文脈だけでは正しい翻訳ができない状況が数多く存在するため，テキストのみを用いた翻訳よりも精度の高い翻訳が期待できる．

先行研究として，画像全体から抽出したグローバルな視覚的特徴を用いて，テキスト処理のためのニューラルネットワークの隠れ状態を初期化する研究や，画像をグリッドに分けて抽出した局所的な視覚的特徴を用いた研究がある．しかし，最新の研究では，MNMT が期待に反して視覚的特徴を無視することが指摘されている．それは，これら従来から提案されている視覚的特徴の統合の仕方が満足のいくものではないからである．

そこで，本研究では，これらの問 題を解決する 2 つの手法を提案する．一つは領域注視型 MNMT，もう一つは単語領域アライメント誘導型 MNMT（WRA-guided MNMT）と名付けられたものである．本研究の主な貢献は以下の通りである．

1. 領域注視型 MNMT では，物体検出により抽出された意味的な画像領域を利用し，2 つの注視機構を用いて視覚とテキストの特徴を統合することを提案する．
2. WRA-guided MNMT では，MNMT におけるテキストと画像のモダリティ間を橋渡しするため，単語領域アライメント（WRA）という新しい仕組みを提案する．

これらの提案手法をリカレントニューラルネットワーク（RNN）と自己注視ネットワーク（Transformer）という 2 つの NMT のアーキテクチャに実装し，検証を行った．英語・ドイツ語，英語・フランス語に対する実験の結果，提案手法は最先端の MNMT 手法のほとんどを凌駕することが示された．さらに，提案手法はより優れた視覚情報の利用を実現できることを分析により示した．

以上のように，本論文で提案する手法は，画像と言語の情報を組み合わせた自然言語処理の新しい基盤となることが期待されるため，情報科学において重要な意義があると考え

られる．よって，本論文は博士（情報科学）の学位を授与するに十分な価値があるものと認められる．

（最終試験又は試験の結果）

本学の学位規則に従い，最終試験を行った．公開の席上（対面とオンラインのハイブリッド）で論文発表を行い，学内外の教員による質疑応答を行った．また，論文審査委員により本論文及び関連分野に関する試問を行った．これらの結果を総合的に判断した結果，専門科目についても十分な学力があるものと認め，合格と判定した．