

プログラム名：「脳情報の可視化と制御による活力溢れる生活の実現」

PM名：山川 義徳

プロジェクト名：「脳ビッグデータ」

委 託 研 究 開 発

実 施 状 況 報 告 書 (成 果)

平成28年度

研究開発課題名：

「脳サーチエンジン」

—脳情報，画像，言語を相互変換可能なシステムの構築—

研究開発機関名：

東京大学 大学院情報理工学系研究科

研究開発責任者

原田 達也

I 当該年度における計画と成果

1. 当該年度の担当研究開発課題の目標と計画

当該年度では、昨年度開発した半教師付きマルチモーダル学習手法を画像、動画、文章等の広範なデータに適用することで、その有効性を検証する。また、脳情報の画像化の要素技術として、情報が削減された画像特徴から元画像を推定、再構成する高精度な手法を開発する。特に手法のコアとなる、画像の局所的特徴間の接続の良さをモデリングするアルゴリズムを考案する。

2. 当該年度の担当研究開発課題の進捗状況と成果

2-1 進捗状況

当初の当該年度の目標は、昨年度開発した半教師付きマルチモーダル学習手法（SBCCA）を広範なデータに適用することであったが、SBCCA では実データ応用にまだ十分な性能が得られないことが分かり、少量の教師データでもより高い分類性能をもつ識別機の学習方法の開発をおこなった。

画像特徴からの元画像の推定では、画像や動画と自然言語間の相互変換といったより挑戦的な課題に着手することにした。具体的には、感情用語を付与可能な画像キャプション生成、ラベル情報からの動画生成、自然言語記述からの画像合成、さらに、視覚的質問応答に関する研究を実施した。

2-2 成果

少量の教師データでも高い分類性能をもつ識別器を実現する研究

近年、大量のラベル付きデータを用いた機械学習によって、様々なモダリティのカテゴリを精度よく識別する手法が実現されている。しかしながら、ラベルが付与された人への提示情報と脳情報とのペア（多種類のモダリティデータ）を収集するには大変なコストがかかる。そこで、注目するモダリティデータとは異なるモダリティのデータから、注目モダリティにも利用可能な共通の識別器を学習できれば、様々なモダリティに対して大量にラベルを付与する必要がなくなる。例えば、Web から集めたラベル付き画像で識別器を学習させ、その識別機で脳情報を分類できれば、ラベル付与された脳情報を収集する必要がなくなる。本研究では、モダリティに依存しない識別器学習のために、ドメイン適合の考え方にもとづき、異なるモダリティから共通の情報を抽出する手法を提案した。本提案手法は Zero-shot 学習に用いられる標準的なデータセットにおいて、高い精度を得られることが分かった。

感情用語を付与可能な画像キャプション生成の研究

従来の画像キャプションタスクは、物体、動作、およびそれらの関係を含む画像の事実に主焦点が当てられていた。しかしながら、画像ではない要素、すなわち形容詞で表現された画像の感情的な情報を説明することはほとんど無視されてきた。そこで本研究では、感情用語を用いた画像キャプション生成の新しい手法を実現した。具体的には、感情に関するデータセットを構築し、このデータセットから感情を識別するネットワークを構築した。この感情を識別するネットワークを通常の画像キャ

プシオン生成のシステムに組み込むことによって、感情用語を含む画像のキャプションを生成している。本提案システムから得られた結果の一部を図1に示す。



図1：画像から感情用語を含むキャプションを生成した例。

ラベル情報からの動画像生成に関する研究

近年、深層学習を利用した静止画像生成に関する研究が行われているが、動画像生成に関する研究はこれまでされていなかった。そこで本研究では、動画像識別器を利用したラベル情報からの動画像生成手法を提案した。具体的には、三次元畳み込みニューラルネットワークの識別誤差と、オートエンコーダの復元誤差の両方を最小化するようにネットワークを学習することで、生成結果がより自然な動画像となるシステムを開発した。図2にラベル情報から生成した動画像例を示す。

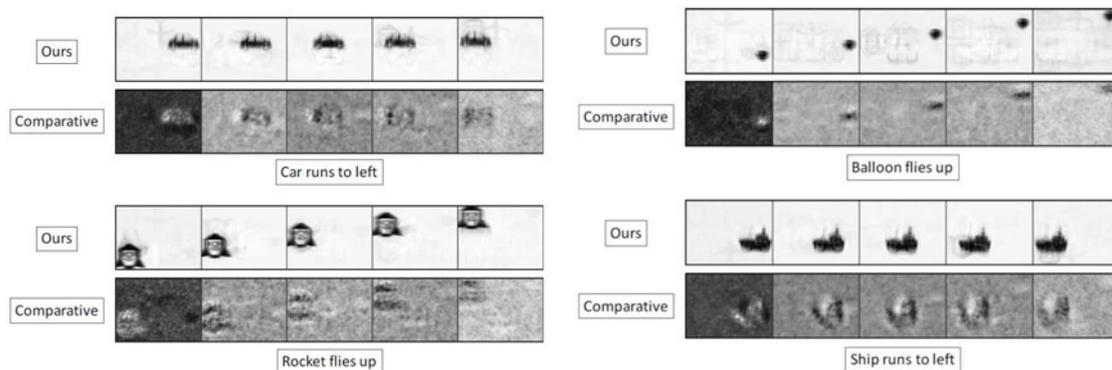


図2：ラベル情報から生成した動画像の例。

自然言語記述からの画像合成の研究

本研究では、与えられた文章からシーン画像を合成するシステムを開発した。具体的には、文章中の単語の関係性を、画像中の物体の関係性に対応付ける階層的構文分析を開発した。提案システムをクリップアートデータと実際の画像データに適用し、クラウドソーシングで評価することで提案手法の有効性を確認した。図3に文章からシーン画像を合成した例を示す。

An irritated gentleman throws tennis ball toward girls near a tall tree.

Three powerboats are racing near the cliff while balloon hovering over them.

Little girl laughing and touching a boy's face as they lean toward each other.

A man and female partner smile for the camera.



図 3 : 文章からのシーン画像の合成例.

視覚的質問応答に関する研究

視覚的質問応答 (VQA) は、画像に関する質問に対して、自然言語で回答する知的システムを構築するタスクである。この VQA タスクにおいて、本研究では画像と質問を共通空間に射影し、射影された各特徴を様々な方法によって統合することで、正確な回答を生成する手法を開発した。提案手法は、VQA challenge 2016 の抽象的なシーンカテゴリ部門において第一位の成績を収めた。

2-3 新たな課題など

少量の教師データでも高い分類性能をもつ識別器を実現する研究において、実験的に高い分類性能が得られているものの、理論的な解析をまだおこなっておらず、また、脳情報に対しても実際に有効に働くかの検証を実施していないために、これらを今後進めていく必要がある。

画像や動画と自然言語との相互変換に関しては、まだ十分な精度が得られているとはいえないために、生成モデルを改善する必要がある。

3. アウトリーチ活動報告

特になし。