ISS-P-039

テキストによる広告画像生成のための真正画像と生成画像の比較調査

島村礼子† 大田美空† 新田直子†
† 武庫川女子大学 生活環境学部 情報メディア学科

1. はじめに

近年，テキストを用いた画像生成技術が著しく向上しているが，所望の画像を生成するために適切なテキストを予め決定する必要がある．広告画像は，一般に特定の広告対象に対して所望の印象を持たせるよう制作されると考えられるが，意図する印象を広告対象に与えるため，具体的にどのような画像を生成すればよいかを決定することは一般ユーザには困難である．本稿では，画像のキャプション生成技術に着目し，専門家が制作した既存の広告画像（以下，真正画像と呼ぶ）を説明するテキストを生成することにより，専門家がどのような内容の画像を制作しているか調査する．また，簡易なテキストから自動生成した画像（以下，生成画像と呼ぶ）に対して同様に画像内容を説明するテキストを生成し，真正画像に対するテキストとの比較により，その差異を調査する．

2. 調査方法

広告画像の特性は広告対象により異なると考えられるため，特定の対象に対して専門家が制作した広告画像を真正画像とする．真正画像に対し，既存の画像のキャプション生成手法を用いて，内容を表すテキストを生成する．一方，広告画像を生成するために考えられる簡易なテキストとして，x を広告対象とした ’an advertisement of x’ を入力し，既存の画像生成手法を用いて，x の広告画像を生成する．生成画像に対し，同様に内容を表すテキストを生成する．真正画像と生成画像を表すテキストを構成する単語からストップワードを除去し，ステミングを行った後，各単語の出現パターンを比較する．

3. 調査結果

既存の広告画像データセット [1] のうち広告対象が alcohol の画像からランダムに選択した 100 枚を真正画像とし，x を alcohol として，Stable Diffusion [3] により 100 枚画像を生成した．真正画像と生成画像に対して BLIP [2] で生成したテキストにおける単語の頻度分布を図 1 に示す．生成画像に対しては’alcohol’ が最も頻出した一方，真正画像に対しては，’beer’, ’wine’, ’vodka’, ’whiskey’ などのより具体的な酒類を表す単語が頻出した．また，真正

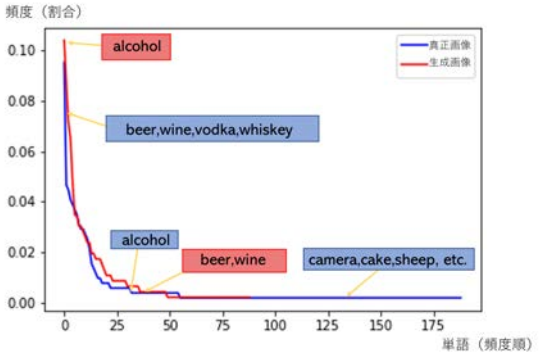

図 1　真正画像と生成画像に対する単語の頻度分布

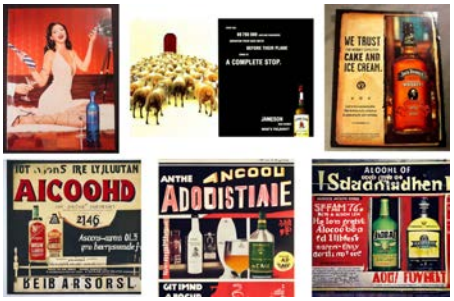

図 2　真正画像（上）と生成画像（下）の例

画像に対しては，alcohol と関係のない単語も含めた約 2.1 倍の単語が出現し，71%の単語は単一の画像に対してのみ出現した．一方，生成画像に対しては 55%の単語が複数の画像に対して出現し，図 2 にも示すように，生成画像は真正画像よりも多様性が低いことが分かる．

4. むすび

本稿では, 広告画像を対象に，簡易なテキストからでもある程度の品質の画像が生成可能であるが，専門家が制作する真正画像に比べ，均質なものとなることを示した．今後は，より多様な画像を生成するためのテキストの決定方法を検討したい．

謝辞
本研究の一部は，JST CREST JPMJCR20D3，科学研究費補助金基盤（C）22K12074 の助成による．

参考文献

[1] Z. Hussain, et al., “Automatic Understanding of Image and Video Advertisements,” CVPR, 2017.
[2] J. Li, et al., “BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation,” ICML, 2022.
[3] R. Rombach, et al., “High-Resolution Image Synthesis with Latent Diffusion Models,” CVPR, 2022.