画像生成における感情価と覚醒度を用いた印象操作方法の検討

大野ひとみ† 大田美空† 中田景子† 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，制作者の意図をテキストで表現することにより画像を自動生成できる画像生成技術が急激に発展している．しかし，制作者の意図を必要十分な形でテキストとして表現することは一般ユーザには容易ではない．そこで本研究は，画像生成技術の中でも Stable Diffusion [1] を対象に，メインの対象物を表す簡易なテキストと共に，意図する印象を表現するものとして，感情を 2 次元で表す値である感情価（Valence）と覚醒度（Arousal）を条件として与えることにより，特定の対象物に対して所望の印象を付与した画像を簡単に生成する手法を提案する．

2 提案手法

提案手法は，メインの対象物を o とするとき，“*a photo of [o]*” という簡易なテキストと感情価と覚醒度 (v, a) を入力とし，Stable Diffusion により画像 I_{out} を生成する．Stable Diffusion では，CLIP guidance [2] と呼ばれる，意味が類似した画像とテキストに対し，類似度の高い特徴量ベクトルを抽出するよう学習されたモデルである CLIP を活用し，I_{out} に対して，その画像特徴量ベクトルを $\boldsymbol{f}^g$ に近づけるという制約を持たせる方法が提案されている．そこで，共起しやすい名詞と形容詞の対 [4]，形容詞 [3]，画像 [5] のそれぞれに対し感情価と覚醒度がラベル付けされた既存データベースである DB_{na}, DB_{va}^{adj}, DB_{va}^{I} を用いて $\boldsymbol{f}^g$ を設定する以下の 3 種類の手法を提案する．

手法 1：o と共起しやすい形容詞を DB_{na} より選択し，DB_{va}^{adj} を参照し (v, a) に最も近い形容詞のテキスト特徴量ベクトルを $\boldsymbol{f}^g$ とする．

手法 2：DB_{va}^{I} から (v, a) に近い感情価，覚醒度を持つ画像を N 枚選択し，これらの平均画像特徴量ベクトルを $\boldsymbol{f}^g$ とする．

手法 3：CLIP guidance なしで生成した画像の特徴量ベクトルを $\boldsymbol{f}^{org}$ とする．DB_{va}^{I} から $\boldsymbol{f}^{org}$ に基づき類似した画像を N 枚選択し，これらの平均画像特徴量ベクトルと，2) で得た画像特徴量ベクトルの差を感情変換ベクトル $\boldsymbol{v}$ とし，$\boldsymbol{f}^{org} + \boldsymbol{v}$ を $\boldsymbol{f}^g$ とする．

3 実験

図 1 に，対象物 o を“*car*”，(v, a) を $(1.1, 5.1)$, $(6.4, 5.0)$, $(4.0, 1.7)$ とし，各手法で生成した画像，及び各生成画像に対し DB_{va}^{I} から $\boldsymbol{f}^{org}$ に基づき類似した画像を N 枚選択し，これらの平均の感情価，覚醒度を示す．ただし手法 2，3 と共に $N = 10$ とした．生成画像の質は手法 1 が最もよいが，全体として，与えた (v, a) に応じて生成画像が変化することが分かった．また，物体に応じて $\boldsymbol{f}^g$ を決定する手法 3 の方が手法 2 に比べて，より質の高い画像が生成された．

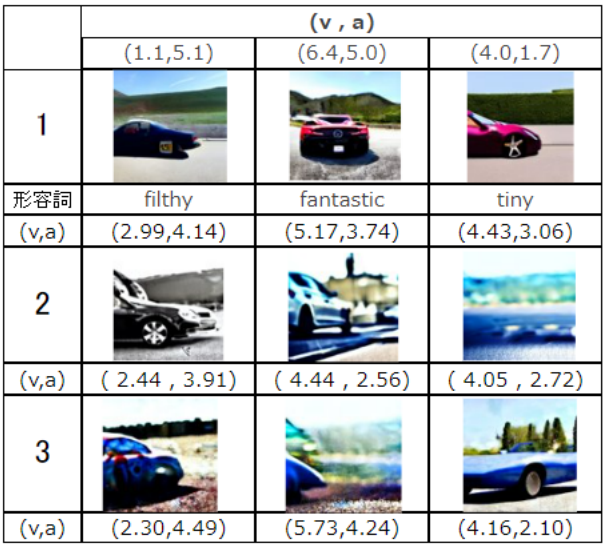

図 1　生成画像の例

4 むすび

本稿は，画像に意図する印象を条件として与えることにより，特定の対象物に対して所望の印象を付与した画像をある程度生成可能であるが，画像の質が低い画像もあり，改善の余地があることを示した．

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

参考文献

[1] R. Rombach, et al., “High-Resolution Image Synthesis with Latent Diffusion Models,” CVPR, 2022.
[2] A. Nichol, et al., ”GLIDE:Guided Language to Image Diffusion for Generation and Editing,” arXiv:2112.10741. -
[3] S. M. Mohammad, ”Obtaining Reliable Human Ratings of Valence, Arousal, and Dominancefor 20,000 English Words,” ACL, pp.174-184, 2018.
[4] D. Borth, et al., ”Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs,” ACMMM, pp.223–232, 2013.
[5] B. Kurdi , et al., ”Introducing the Open Affective Standardized Image Set (OASIS),” Behavior Research Methods, 49(2), pp.457–470,2016.
[6] A. Radford, et al., ”Learning Transferable Visual Models from Natural Language Supervision,” ICML, 2021.