

原著論文

感性語を考慮した敵対的生成ネットワークを用いたロゴマーク生成

佐川 友里香, 萩原 将文

慶應義塾大学大学院

Logo Generation with Generative Adversarial Networks Conditioned on Sentiment Terms

Yurika SAGAWA and Masafumi HAGIWARA

Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Yokohama, Kanagawa 223-8522, Japan

Abstract : Logos are very important to represent the impression of companies and brands. Since many logos are created by designers, there is a problem of high cost both in terms of money and time. Since sentiment terms are important as keywords when designing logos, we propose a logo generation system using sentiment terms. Originally, we should create a dataset of logo images and sentiment terms directly, but it is difficult to collect these logo images and the corresponding sentiment terms. Therefore, we employ the domain adaptation method to the dataset of furniture images and sentiment terms to create dataset of the logos and sentiment terms. Finally, we can generate logos considering 17 sentiment terms using Generative Adversarial Networks. We evaluated the proposed system using MS-SSIM score and confirmed that it can generate logos with diversity. Moreover, we performed subjective experiments to confirm that it can generate logos reflecting sentiment terms.

Keywords : Logo generation, Generative adversarial networks, Sentiment terms

1. はじめに

企業やブランドを象徴するロゴマークは、それらのコンセプトやイメージを伝える手段として重要なものである。現在主流であるロゴマークの作成方法はデザイナーによるもので、これは金銭面・時間面ともにコストが大きいという問題点がある。

近年のニューラルネットワークによる生成モデルの発展は、この問題の解決策となる可能性がある。画像生成と呼ばれる、学習データに存在しない未知の画像を生成するタスクは、深層学習の発展に伴い精度が年々向上している。特に敵対的生成ネットワーク (Generative Adversarial Networks, GANs) [1] を用いた研究の発展は大きい。GANは画像を生成する生成器 (Generator) と画像を識別する識別器 (Discriminator) という2つのニューラルネットワークから構成されている。これら生成器と識別器を敵対的に学習させ、最終的には生成器が学習画像と類似している画像を生成できるようになることをめざしている。

GANを用いた研究としては、解像度が高い著名人の顔画像を生成する研究 [2]、解釈文を入力として画像を生成する研究 [3]、明確な教師画像ペアを必要とせず画像変換を行う研究 [4] などが代表的である。初めてGANを用いてロゴマークを生成した研究は、Sageらの研究 [5] である。彼らはロゴマーク画像のクラスタリングを行い、その結果得られたクラスを教師ラベルとして扱うことで、これらラベルを入力とするロゴマーク生成を実現させた。ただし、入力となるラベルはクラスタリングにより生成されたものであり、

ロゴマークのデザインを考える際に使いやすいものとは言えない。そこで、色情報を元にロゴマークを生成する研究 [6] が存在する。この研究では各画像のピクセル値に対してクラスタリングを行う事でロゴマークの代表色を抽出し、その代表色を入力ラベルとして用いる。そのため、指定した色を用いたロゴマークを生成することが可能となっている。

本論文では、ロゴマークのデザインを行う際にイメージを表すキーワードが重要であると考え、指定した感性語を元にロゴマークを生成するシステムを提案する。ここでニューラルネットワークを用いる場合、一般的に大量の学習用データが必要である。しかしながら、目的に適する大量のデータ、特に感性工学関係のデータは取得が非常に困難である。ロゴマーク自動生成をめざす場合には、ロゴマーク画像とそれに対応した感性語の大量のデータセットが必要となる。そこで本研究では、取得可能なドメインでのデータを異なるドメインに適応させる転移学習の考え方を採用する。具体的には、家具画像・感性語のデータセットを作成し、転移学習によりロゴマーク生成を行うというアプローチを用いる。そして、“かわいい”や“モダン”などのロゴマークのイメージを表すと考えられる感性語17単語を元に、感性語が与えるイメージを反映させたロゴマークの生成を目指す。

本研究の貢献は以下の3点である。

- (1) 家具通販サイトを用いた家具画像・感性語データセットの作成
- (2) 転移学習手法ADDAによるロゴマーク画像・感性語データセットの作成
- (3) 指定した感性語ラベルを考慮したロゴマーク生成

以下では第2章で関連研究、第3章でロゴマーク画像・感性語クラスの決定、第4章でロゴマーク生成、第5章で生成例と評価実験、第6章で結論について述べる。

2. 関連研究

2.1 生成モデル

生成モデルとは、データを生成する過程を数理的にモデル化したもので、確率分布により表される。これら生成モデルのうち、乱数 z からデータ x を生成する生成器 $p(x|z)$ をモデル化する手法は生成器モデルと呼ばれ、Variational AutoEncoder (VAE) [7, 8] と GAN [1] が代表的である。これら2つのネットワークは、潜在変数から画像を生成することができる。VAEは潜在変数が確率分布に従うように学習させることで画像を生成するが、損失関数にL2ロスを使用しているためぼやけた画像が生成されやすいという問題点がある [5]。一方GANは生成器と識別器を同時に競わせて学習する事で、本物らしい画像を生成する事ができる。識別器 D は画像が本物であるか偽物であるか、すなわち学習データ画像 x であるか生成画像 $G(z)$ であるかの真偽判定を行う。それに対して生成器 G は、乱数 z を元に画像を生成し、識別器をだますように学習を行う。つまりGANは以下の損失関数で表されるようなミニマックスゲームを行うように学習を行う [1]。

$$\min_G \max_D L_{adv} = \log D(x) + \log(1 - D(G(z))) \quad (1)$$

しかしGANには、学習が不安定である点、および同じような画像しか生成できない状態 (モード崩壊) に陥る可能性がある点という問題点がある。そこで学習の不安定さ、モード崩壊を改善するため、様々な研究が行われている。GANの学習方法の改善案の提案 [9] や、畳み込み層を用いたGANの提案 [10]、確率分布の距離を測る際に Wasserstein 距離を使用した研究 [11-13] などが存在する。

また、乱数を入力として画像を生成する通常のGANを拡張した研究として、条件付きモデルが提案されている。条件付きモデルの代表的なネットワークがconditional GAN (cGAN) [14] とよばれるもので、クラスラベルを生成器と識別器の入力に与える事で、明示的に画像を生成することができる。また、cGANを改良した研究としてAuxiliary Classifier GAN (ACGAN) [15] も有名である。これは生成器にのみクラスラベルを与え、識別器で画像の真偽判定に加えてクラス分類を行うモデルである。

2.2 ロゴマーク生成

初めてGANを用いてロゴマーク生成を行った研究はSageらの研究 [5] である。この研究の貢献は大きく以下の3つである。

- Large Logo Dataset (LLD) の作成 [16]。
- クラスタリングによる擬似ラベルの作成。
- 擬似ラベルを使用したロゴマーク生成。

しかし、これらのラベルはクラスタリングにより作成されたものであり、人間が直感的に理解しやすいものとは言えない。

そこで理解しやすいラベルとして色情報を用いた研究 [6] が存在する。この研究では12色を用いてロゴマーク画像に対して色クラスを定義し、色ラベルを使用したロゴマーク生成を行っている。

ロゴマークのデザインを行う際には、イメージを表す感性語が重要であると考えられる。そこで本研究ではロゴマーク画像に対して感性語クラスを定義し、感性語ラベルを考慮したロゴマーク生成を目指す。

3. ロゴマーク画像・感性語クラスの決定

GAN [1] を用いた画像生成では、学習に大量の画像が必要となる。そこでインターネット上からスクレイピングを行うことで画像・感性語のデータセット作成を行った。前述のように、本来であればロゴマーク画像と感性語のデータセットを直接作成するべきであるが、関係性が得られる題材の取得が極めて困難であり、まず家具画像・感性語データセットの作成を行った。次に転移学習を行うことでロゴマーク・感性語データセットを作成した。家具における感性的イメージの重要性や、感性語を多く取得できる通販サイトの存在、家具という題材の一般性などを考慮し、家具画像を選択した。

以下の手順によりロゴマーク画像に対して感性語クラスを定義した。

(1) デザイナーズ家具通販サイトより家具画像・感性語組を取得。

(2) Adversarial Discriminative Domain Adaptation (ADDA) [17] を用いてロゴマーク画像に感性語ラベルを付与。

ADDAとは、GANの敵対的に学習させるという学習手法を用いた転移学習手法である。図1にADDAの学習の流れを示す。図1の各フェーズにおいて、パラメータを学習するネットワークを実線、パラメータを固定するネットワークを破線で示している。ADDAで行う処理は大きく3つのフェーズに分けられる。ここで、転移学習元である家具データセットをソースドメイン、転移学習先であるロゴマークデータセットをターゲットドメインと呼ぶ。

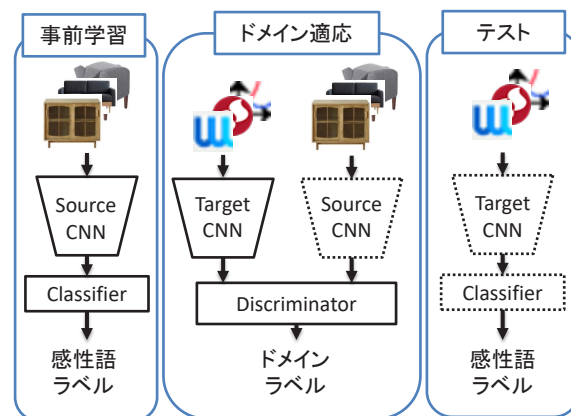


図1 ADDAによる転移学習の流れ

感性語を考慮した敵対的生成ネットワークを用いたロゴマーク生成

- (1) **事前学習**: 事前学習フェーズでは、ソースドメインにおける画像 \mathbf{X}_s , 正解ラベル \mathbf{Y}_s を畳み込みニューラルネットワーク (Convolutional Neural Networks, CNNs) [18] を用いて学習させる。ソース CNN の最終層における特徴空間を $M_s(\mathbf{x}_s)$, CNN の最終出力を $C(M_s(\mathbf{x}_s))$, クラス数を K とすると、損失関数は以下の式で表せる。

$$\min_{M_s, C} L_{cls} = -\sum_{k=1}^K y_{s,k} \log C(M_s(\mathbf{x}_s)) \quad (2)$$

- (2) **ドメイン適応**: ドメイン適応フェーズでは、ソースドメインで学習済みの CNN モデルを用いて、ソースドメインとターゲットドメインの CNN 最終層における特徴空間が類似するように学習させる。この際の学習手法として GAN の学習手法を使用し、識別器によりソース・ターゲットのどちらの特徴空間であるかを分類する。一方ターゲットドメインの CNN モデルは識別器をだますように学習するため、結果的にターゲットドメインの CNN モデルでは、ソースドメインの特徴空間に類似した特徴空間を得ることができる。ターゲットドメイン画像を \mathbf{X}_t , ターゲット CNN の最終層における特徴空間を $M_t(\mathbf{x}_t)$ とすると、損失関数は以下の式で表される。

$$\min_{M_t} \max_D L_{adv} = \log D(M_s(\mathbf{x}_s)) + \log(1 - D(M_t(\mathbf{x}_t))) \quad (3)$$

- (3) **テスト**: テストフェーズではドメイン適応フェーズで学習させたターゲットドメインの CNN モデルを使用し、ターゲット画像にラベルを付与する。

3.1 家具画像・感性語組の取得

情報解析に用いる目的として、デザイナーズ家具通販サイト FLYME [19] から家具画像と説明文を取得し、家具画像・感性語データセットを作成した。家具画像・感性語データセット作成の流れを図2に示す。図2に示した画像・説明文は FLYME のウェブサイトより引用している。具体的な手順としては、まず初めに1つの家具アイテムにつき、3枚の画像と1文の説明文を取得した。そして画像の選定と感性語の選定を行い、1枚の画像につき最大5個の感性語ラベルを付与した。これは、正解ラベルが一意に定まる物体認識タスクなどと異なり、感性語の正解ラベルは一意に定まらないためである。

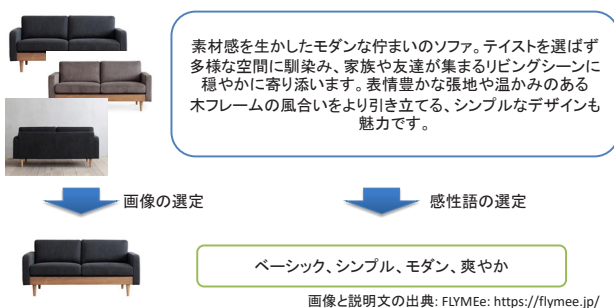


図2 家具画像・感性語データセット作成の流れ

画像の選定では、画像の背景が与える影響を低減するため、背景色が白に近い画像を選ぶ作業を行った。具体的には、背景が白という事は画像の4隅が白色であるとし、4隅のRGB値と白のRGB値の距離の総和が一番小さいものを3画像のうち1枚抽出した。

感性語の抽出は、以下の流れで行った。これは、CNNを用いて感性語予測を行う既存研究 [20, 21] を参考にしたものである。

まず初めに、形態素解析ツールである MeCab [22] を使用し、名詞 (形容動詞語幹) と形容詞 (自立) という感性語となりうる品詞を取得した。その後、不適切なものを以下の基準により削除した。

- 単語となり得ないもの (例: “ザイン” など)
- 物体の名称を表す単語 (例: “商品” など)
- 素材・色を表す単語 (例: “ガラス”, “赤” など)
- 意味が大域的すぎる単語 (例: “良い”, “悪い” など)
- 家具の感性語として明らかに不適切なもの (例: “名高い”, “健やか” など)

その後、表記ゆれと類語の整理を行った。“柔らか”と“柔らかい”などは1単語として扱い、類語の整理には日本語 WordNet [23] を使用した。

3.2 ADDAによる転移学習

作成した家具画像・感性語データセットを元に、ADDAによる転移学習を行った。感性語という曖昧なラベルを扱うため、1枚の画像につき正解ラベルを最大5個付与したマルチラベルデータセットを作成した。そのため、従来のCNNをマルチラベルに対応するよう改良し、感性語予測を行った。マルチラベルに対応させたCNNの構造を図3に示す。CNNでマルチラベルを扱う方法については様々な手法 [24] が提案されているが、Binary Relevance法 [25] という簡潔な手法を用いた。これは識別層を分類クラス数に分割し、それぞれの識別層においてクラスに含まれるか含まれないかの2値分類を行う手法である。

また、ADDAによる転移学習はGANの学習手法を用いているため、学習が不安定であるという問題点がある。学習安定化のため、損失関数に以下2点の変更を加えた。

- Wasserstein GANモデル (WGAN) [10] の使用。
- Feature Matching [9] の導入。

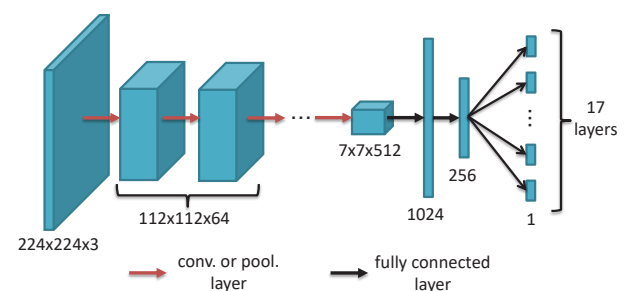


図3 マルチラベルに対応させたCNNモデル

WGANモデルは損失関数にWasserstein距離を用いたモデルである。通常のGANでは確率密度分布の距離を測る指標としてJSダイバージェンスが用いられていたが、勾配消失問題の発生が指摘されていた。そこでWasserstein距離を用いることでその改善を目指したのがWGANモデルである。WGANモデルでは以下の損失関数でWasserstein距離とするため、識別器の重みをクリップする事も行っている。

$$\min_G \max_D L_{adv} = D(\mathbf{x}) - D(G(\mathbf{z})) \quad (4)$$

Feature Matchingはソース・ターゲットドメイン画像を入力した場合の識別器の中間層同士のL2ロスを生成器の損失項に加える手法である。識別器の中間層における出力を関数 f とすると、以下の式で表される。

$$\min_G L_{fm} = \|f(\mathbf{x}) - f(G(\mathbf{z}))\| \quad (5)$$

4. 提案するロゴマーク生成ネットワーク

提案するロゴマーク生成ネットワークを図4に示す。提案ネットワークではロゴマーク生成の既存研究[5]であるSageらのネットワークを用いた。彼らは、layer conditional GANと呼ばれる構造を使用しており、これはラベル情報をすべての層の入力に入力する構造である。ラベル情報として17種類の感性語クラスを用い、各層の入力に結合した。ただし、各層に入力しているラベル情報を考慮しすぎないための工夫としてショートカット接続を持つResNet機構[26]を用いている。提案ネットワークではさらにSageらのモデルを、より強く感性語ラベルを考慮するためにACGANモデル[15]に変更した。また、今回学習に用いたロゴマーク・感性語データセットはマルチラベルデータセットであるため、マルチラベルに対応したACGAN構造にした。具体的には3.2節で述べたBinary Relevance法[25]を用いて、マルチラベルに対応させた。

ACGANは識別器が入力画像の真偽判定に加え、クラス分類も行うモデルである。通常のGANの損失 L_{adv} に加えて、クラス分類における損失 L_{cls} も考慮する。 \mathbf{c} を入力ラベルとすると、損失関数は以下の式で表される。

$$\min_G \max_D L_{adv} = \log D(\mathbf{x}) + \log(1 - D(G(\mathbf{c}, \mathbf{z}))) \quad (6)$$

$$\min_{G,D} L_{cls} = -\sum_{k=1}^K (y_k \log D(\mathbf{x}) + y_k \log D(G(\mathbf{c}, \mathbf{z}))) \quad (7)$$

また、学習を安定化するためWGAN-gradient penalty (WGAN-gp) [13] モデルを用いた。WGAN-gpモデルは損失関数にWasserstein距離を用いたモデルで、学習が進んだ場合に勾配のノルムが1になるという性質を利用したものである。3.2節のADDAによる転移学習では、識別器の重みをクリップしていたが、この手法ではうまく学習できない場合があるという指摘があり、その改善策としてgradient penaltyが導入された[13]。損失関数は以下の式で表される。ただし、 λ は係数、 $\hat{\mathbf{x}}$ は生成画像と学習画像の任意の中間画像である。

$$\min_D \max_G L_{adv} = D(G(\mathbf{z})) - D(\mathbf{x}) + \lambda [\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}}) - 1\|^2] \quad (8)$$

5. 生成例と評価実験

5.1 ロゴマークに適した感性語抽出に関する実験

初めに、3.1節で作成した家具画像・感性語データセットのうち、ロゴマークの感性語として適している感性語の選択を行うため、予備実験を行った。提案システムでは家具画像・感性語データセットにおける感性語ラベルが最終的にロゴマーク生成の入力ラベルとなるため、ロゴマークのイメージを表す感性語として適している単語を選択する必要がある。

そこで、3.1節で作成した家具画像・感性語データセットのうち画像枚数の多い上位50単語に対して、ロゴマークの感性語として適しているかどうかを「当てはまる」「どちらとも言えない」「当てはまらない」の3段階で評価を行った。被験者は20代の男女17名である。実験結果に対し、「当てはまる」をはじめとする3段階の選ばれる確率がともに1/3であるという帰無仮説に基づき、カイ二乗適合度検定を行った。有意水準5%において有意差が得られたものを表1に示す。表1におけるスコアとは、「当てはまる」と回答した被験者の人数である。有意差が得られた感性語のうち、「当てはまる」という回答数の上位24単語を選択し、今後の実験で用いた。

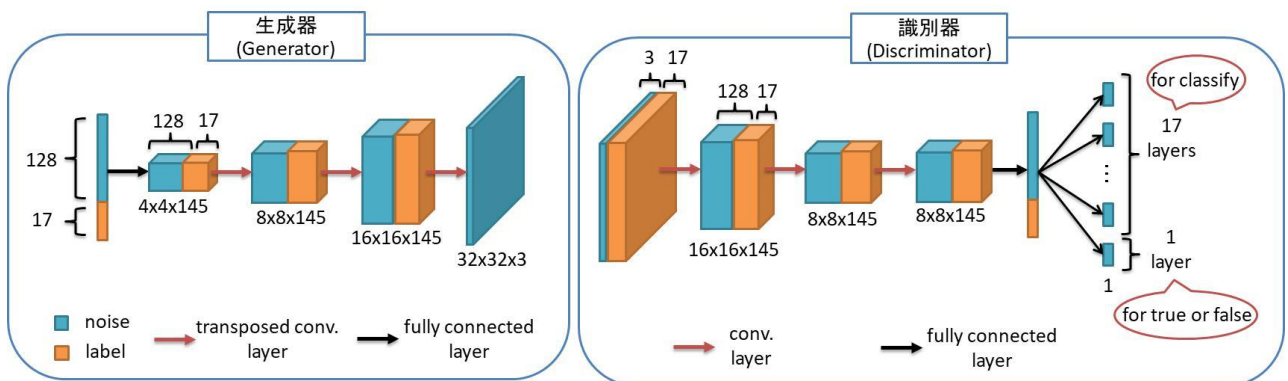


図4 提案するロゴマーク生成ネットワークの構成

感性語を考慮した敵対的生成ネットワークを用いたロゴマーク生成

表1 予備実験において有意差が得られた感性語

感性語	スコア	感性語	スコア	感性語	スコア	感性語	スコア
モダン	16	シック	13	美しい	12	多彩	10
かわいい	16	ゴージャス	13	高級	12	自然	9
シンプル	15	エレガント	13	クラシック	12	楽しい	8
シャープ	15	ユニーク	13	綺麗	11	新鮮	5
明るい	15	爽やか	13	上品	11	心地よい	3
ポップ	15	クール	13	カジュアル	11	フリー	1
ダーク	14	スマート	13	力強い	11	丈夫	1
ダイナミック	14	奇抜	13	不思議	11		

5.2 ADDAによる転移学習

5.2.1 事前学習

ADDAの事前学習フェーズでは、家具画像・感性語データセットを用いてCNNの学習を行う。予備実験の結果より感性語クラス数は24クラスとなり、全画像数は14,050枚となった。CNNモデルにはImageNet Large Scale Visual Recognition Challenge [27] という画像認識精度を競う大会で高精度を記録しているVGG19モデル [28] を使用し、Binary Relevance法 [25] を用いてマルチラベルに対応するよう構造を変更した。評価指標は $top\ K$ 精度を用いた。これはCNNの予測上位 K 個のうち、正解ラベルが含まれている割合を表す指標である。学習枚数は13,500枚、ミニバッチサイズは32、学習回数は50回とした。最適化手法にはAdam [29] を用いた。実験の結果、 $top1$ 精度は66.3%、 $top5$ 精度は90.6%であった。

5.2.2 ドメイン適応

事前学習フェーズにより学習した家具ドメインCNNモデルを用いて、ロゴマークドメインCNNの学習を行う。ロゴマーク画像はSageらの研究で用いられているLLD-iconデータセット [15] を使用した。転移学習を行う際、ソースドメインとターゲットドメインの学習画像枚数は同じである必要があるため、LLD-iconデータセットより13,500枚をランダムに選択し学習した。LLD-iconデータセットの画像はブラウザのファビコンを集めたデータセットで、各画像の大きさは 32×32 である。CNNに入力するため、 224×224 にリサイズして使用した。また、ミニバッチサイズは16、学習回数は50とした。最適化手法にはRMS-Prop [30] を用いた。最適化手法を変更したのは、Wasserstein GANモデルに変更したためである。

5.2.3 テスト

ドメイン適応フェーズにより学習したロゴドメインCNNモデルを用いて、ロゴマーク画像に対して感性語のラベリングを行う。ロゴマークの全画像485,377枚に対してテストを行い、各画像に対して出力上位5単語を取得した。CNNの最終出力は各クラスにおける予測確率を表すため、上位5単語のうち出力確率 p がランダム確率より高いもの、すなわち、 $p>1/24$ を満たすものを選択した。ロゴマーク画像とその出力例を図5に示す。



	感性語	予測確率	感性語	予測確率	感性語	予測確率
top1	モダン	0.999	かわいい	0.996	ユニーク	0.950
top2	シンプル	0.000	美しい	0.003	美しい	0.045
top3	シック	0.000	綺麗	0.000	明るい	0.001
top4	ダーク	0.000	ゴージャス	0.000	奇抜	0.000
top5	美しい	0.000	爽やか	0.000	モダン	0.000

図5 転移学習後のロゴマーク画像とその出力例

5.3 ロゴマーク生成

提案システムでは、5.2.3節において感性語ラベルを付与したロゴマークデータセットのうち画像枚数が1,000枚以上である17クラスを選択し、この17クラスを用いてロゴマーク生成を行った。

作成したロゴマーク画像・感性語データセットの各感性語とその画像枚数を表2に示す。また、その一部を図6に示す。学習画像枚数は485,377枚、バッチサイズは64、学習総数は200,000イテレーションとした。各感性語を元として生成した画像例を図7に示す。ただし、ロゴマークを生成する際のネットワーク構造として、Sageらのネットワークと、SageらのネットワークをACGANに変更したネットワークを用いた場合について実験を行った。図7の各列の上半分がACGANネットワークを使用した場合、下半分がSageらのネットワークを使用した場合における生成例を表している。

表2 ロゴマーク・感性語データセットにおける感性語と画像枚数

感性語	枚数	感性語	枚数
シンプル	230,426	ポップ	11,035
美しい	133,485	ユニーク	9,901
モダン	57,513	ダーク	6,509
シック	45,299	上品	6,252
かわいい	37,689	高級	4,881
ゴージャス	23,184	シャープ	2,576
エレガント	22,379	カジュアル	2,210
綺麗	14,436	爽やか	1,709
明るい	12,458		



図6 作成したロゴマーク・感性語データセットの一部



図7 入力となる感性語ラベルと生成画像 (上: ACGANネットワーク, 下: Sageらのネットワーク)

5.4 定量評価実験

5.4.1 Diversityスコア

GANを用いて画像を生成した場合、モード崩壊と呼ばれる、生成画像が類似してしまう状態に陥ってしまう可能性がある。提案手法モデルがモード崩壊に陥っていないかを確認するため、MS-SSIM指標 [31] を用いたDiversityスコアを評価指標として定量評価を行った。この指標は、Odenaらの研究 [15] で提案された。MS-SSIM指標は画像組に対して類似度を計算する評価指標であるため、MS-SSIMの平均値が低ければ画像が類似していない、すなわち、多様性があると言える。

比較のためにSageらの研究 [5] を再現した。この生成例を図8に示す。入力に与えるクラス数は128で、その他の実験パラメータは提案手法実験時と同じ値を用いた。ただし、生成例の上半分がACGANネットワーク、下半分がSageらのネットワークである。

Sageらの研究による生成画像、提案手法による生成画像におけるDiversityスコアを表3に示す。生成画像からランダムに5万枚を選択し、Diversityスコアを計算した。

Diversityスコアは値が低いほど多様性があると言えるため、提案手法は既存研究に劣る結果となった。この原因とし

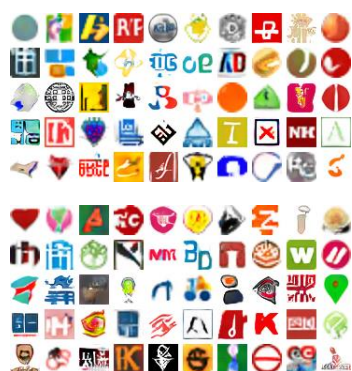


図8 Sageらの研究における生成画像 (上: ACGANネットワーク, 下: Sageらのネットワーク)

表3 MS-SSIM指標を用いたDiversityスコア

モデル	Diversityスコア
Sageらの研究 [5]	0.056 ± 0.102
Sageらの研究 (ACGAN) [5]	0.057 ± 0.103
提案手法 (Sageらのネットワーク)	0.162 ± 0.183
提案手法 (ACGAN)	0.152 ± 0.174

て、画像生成ネットワークに入力するラベルのクラス決定方法があげられる。既存研究では画像特徴を用いたクラスタリングにより128クラスに分類しており、各クラス内には画像特徴的に似ているロゴマークが集まっている。それに対し提案手法では、感性という曖昧な特徴を元に17クラスに分類しているため、各クラス内の画像は既存研究ほど似ている画像が集まっているわけではない。GANによる画像生成は、似ている画像が同一クラスであるほど精度の高い画像が生成できるため、クラス分類手法の違いにより生成画像の精度・多様性に影響があったと考えられる。

ただし、提案手法のDiversityスコアは0.15前後の値であり、図7の生成例をみても様々なロゴを生成できている事がわかる。そのため提案手法モデルはモード崩壊に陥っていない事が示唆される。

5.4.2 CNNを用いた感性語識別率

生成した画像について、CNNを用いて感性語の推定を行った。5.3節で作成したロゴマーク生成用のデータセットを用いて感性語予測のCNNモデルを構築し、識別率を算出した。感性語予測CNNモデルの学習には、ロゴマーク画像480,000枚を使用し、ミニバッチサイズは32、学習回数は5回とした。評価指標にはtopK精度を用いた。

比較のため、感性語予測CNNモデルの学習に用いていないテスト用画像5,000枚でも実験を行った。実験条件を等しくするため、テスト用画像のラベルをシングルラベル化し、各ラベルにつき同じ画像枚数で実験を行った。具体的には、テスト用画像に複数ラベルが付与している場合、画像を複製し1画像につき1ラベルが付与されている状態にした。

テスト用画像と生成画像に対するtop1精度、top5精度を表4に示す。テスト用画像の識別率と比較すると、top1精度、top5精度ともに識別率は下回る結果となったが、17クラスという比較的多いクラス数において、top1精度で約30%、top5精度で約70%の確率で感性語イメージを反映したロゴマーク画像を生成できている事がわかる。

表4 CNNを用いた感性語識別率

画像	top1 [%]	top5 [%]
テスト用画像	59.6	94.1
提案手法 (Sageらのネットワーク)	29.7	71.4
提案手法 (ACGAN)	29.5	72.0

表5 主観評価実験の結果

評価項目		評価値平均
何らかの形・模様を表しているか (**)		0.85
感性語が表すイメージに適しているか	シンプル (**)	3.62
	美しい (**)	3.45
	モダン (**)	3.95
	シック (*)	3.19
	かわいい (**)	3.74
	ゴージャス	2.99
	エレガント (*)	3.35
	綺麗 (**)	3.51
全感性語 (**)		3.48

(*) 有意水準5%で有意差, (**) 有意水準1%で有意差

5.5 主観評価実験

生成したロゴマーク画像に対して、入力した感性語に即したロゴマーク画像を生成できているかどうか検証するため、主観評価実験を行った。本実験では表2に示した画像枚数上位8クラス(シンプル, 美しい, モダン, シック, かわいい, ゴージャス, エレガント, 綺麗)を実験対象とし、各クラス84枚のACGANモデルにおける生成画像について実験した。被験者に生成画像と入力感性語を提示し、以下の2項目について評価を行った。ただし評価項目2については、評価項目1で当てはまるものに対してのみ評価を行った。

- (1) 生成画像が何らかの形・模様を表しているか
 - 0(当てはまらない)と1(当てはまる)の2段階
- (2) 感性語が表すイメージに適しているか
 - 1(当てはまらない)から5(当てはまる)の5段階

本来であれば感性語を考慮したロゴマーク生成の既存研究と比較を行うべきであるが、適する研究やベースラインの設定が困難であり、提案手法のみで実験を行った。被験者は20代の男女14名である。その結果を表5に示す。評価値の期待値(5段階であれば3)と実験結果が同じであるという帰無仮説に基づき、ウィルコクソンの符号順位検定を行った。検定の結果より、全ての感性語における評価値平均において有意差が確認できた。また、クラス別に検定を行うと、有意水準1%においては「シック」「ゴージャス」「エレガント」のクラス以外で有意差が確認できた。被験者からは、「シック・エレガント・ゴージャスは単語の意味定義が曖昧で評価が難しい」という意見が得られた。加えて、図7の学習データセット画像を見ると、3単語の学習画像は画像特徴の類似性が小さく、「シックらしい」などの感性特徴を捉える事が難しかったと考えられる。

しかし、有意差が得られたクラスについての評価値平均はおよそ3.5であり、感性語を考慮したロゴマークを生成できている事が示唆される。

6. 結論

本論文では、感性語を考慮した敵対的生成ネットワークを用いたロゴマーク生成を提案した。本論文の貢献は大きく以下の3つである。

- (1) 家具通販サイトを用いた家具画像・感性語データセットの作成

- (2) 転移学習手法ADDAを用いたロゴマーク画像・感性語データセットの作成

- (3) 感性語ラベルを考慮したロゴマーク生成

感性工学という曖昧で主観が大きな要素をしめる分野では、データ収集も困難な課題の一つである。インターネット上から感性語データセットを作成する手法、別ドメインの感性情報を転移学習により使用する方法の提案も貴重な貢献であると考えられる。

生成画像について評価実験を行い、感性語が与えるイメージを反映したロゴマークが生成可能である事が示された。また、Diversityスコアを用いた実験により、既存研究には劣るものの多様性のある画像を生成できている事が確認された。

今後の展望として、画像・感性語データセットの学習方法の改良が考えられる。本研究ではCNNを用いて画像と感性語の対応関係を学習した。CNNは物体認識タスクにおいて高精度を記録しているが、これは層が深くなるにつれて認識に必要な物体情報のみを残存するためと言われている。そのため物体の形や構造で一意に識別する事ができない感性語推定タスクでは、CNNは適していないと感じた。そこで、CNNを用いるのではなく、感性的特徴を表す指標を用いて画像と感性語の対応関係を学習させることで、感性語推定タスクの精度向上が期待できると考える。

参考文献

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., and Bengio, Y.: Generative adversarial nets, In Advances in Neural Information Processing Systems, pp.2672-2680, 2015.
- [2] Karras, T., Aila, T., Laine, S., and Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation, International Conference on Learning Representations, 2018.
- [3] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H.: Generative adversarial text to image synthesis, 33rd International Conference on Machine Learning, 48, pp.1060-1069, 2016.
- [4] Zhu, J. Y., Park, T., Isola, P., and Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks, IEEE International Conference on Computer Vision, pp.2242-2251, 2017.
- [5] Sage, A., Agustsson, E., Timofte, R., and Van Gool, L.: Logo synthesis and manipulation with clustered generative adversarial networks, IEEE Conference on Computer Vision and Pattern Recognition, pp.5789-5888, 2018.
- [6] Mino, A., and Spanakis, G.: LoGAN: Generating logos with a generative adversarial neural network conditioned on color, arXiv preprint arXiv:1810.10395, 2018.
- [7] Kingma, D. P., and Welling, M.: Auto-encoding variational

- bayes, International Conference on Learning Representations, 2013.
- [8] Rezende, D. J., Mohamed, S., and Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models, 31st International Conference on Machine Learning, 32(2), pp.1278-1286, 2014.
 - [9] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X.: Improved techniques for training GANs, In Advances in Neural Information Processing Systems, pp.2234-2242, 2016.
 - [10] Radford, A., Metz, L., and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434, 2015.
 - [11] Arjovsky, M., Chintala, S., and Bottou, L.: Wasserstein generative adversarial networks, 34th International Conference on Machine Learning, 70, pp.214-223, 2017.
 - [12] Berthelot, D., Schumm, T., and Metz, L.: Improved boundary equilibrium generative adversarial networks, IEEE Access, 6, pp.11342-11348, 2018.
 - [13] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C.: Improved training of wasserstein gans, In Advances in Neural Information Processing Systems, pp.5767-5777, 2017.
 - [14] Mirza, M., and Osindero, S.: Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.
 - [15] Odena, A., Olah, C., and Shlens, J.: Conditional image synthesis with auxiliary classifier GANs, 34th International Conference on Machine Learning, 70, pp.2642-2651, 2017.
 - [16] LLD-large logo dataset-version 0.1, <https://data.vision.ee.ethz.ch/cvl/lll/> (2019.01.23 閲覧).
 - [17] Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T.: Adversarial discriminative domain adaptation, IEEE Conference on Computer Vision and Pattern Recognition, pp.2962-2971, 2017.
 - [18] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D.: Backpropagation applied to handwritten zip code recognition, Neural Computation, 1(4), pp.541-551, 1989.
 - [19] FLYMEe : <https://flymee.jp/> (2018.12.09 閲覧).
 - [20] Chen, T., Borth, D., Darrell, T., and Chang, S.: DeepSentiBank: Visual sentiment concept classification with deep convolutional neural networks, arXiv preprint arXiv:1410.8586, 2014.
 - [21] Shin, A., Ushiku, Y., and Harada, T.: Image captioning with sentiment terms via weakly-supervised sentiment dataset, British Machine Vision Conference, pp.53.1-53.12, 2016.
 - [22] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis, 2004 Conference on Empirical Methods in Natural Language Processing, pp.230-237, 2004.
 - [23] Bond, F., Baldwin, T., Fothergill, R., and Uchimoto, K.: Japanese SemCor: A sense-tagged corpus of Japanese, 6th International Conference of the Global WordNet Association, pp.56-63, 2012.
 - [24] Zhang, M., and Zhou, Z.: A review on multi-label learning algorithms, IEEE Transactions on Knowledge and Data Engineering, 26(8), pp.1819-1837, 2014.
 - [25] Boutell, M. R., Luo, J., Shen, X., and Brown, C. M.: Learning multi-label scene classification, Pattern Recognition, 37(9), pp.1757-1771, 2004.
 - [26] He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016.
 - [27] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L.: ImageNet large scale visual recognition challenge, International Journal of Computer Vision, 115(3), pp.211-252, 2015.
 - [28] Simonyan, K., and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, The International Conference on Learning Representations, 2015.
 - [29] Kingma, D., and Ba, J.: Adam: A method for stochastic optimization, International Conference on Learning Representations, 2015.
 - [30] Tieleman, T., and Hinton, G.: Lecture 6.5 – rmsprop coursera: Neural Networks for Machine Learning, University of Toronto, Technical Report, 2012.
 - [31] Wang, Z., Simoncelli, E.P., and Bovik, A.C.: Multiscale structural similarity for image quality assessment, 37th Asilomar Conference on Signals, Systems & Computers, 2, pp.1398-1402, 2003.



佐川 友里香 (非会員)

2017年 慶應義塾大学理工学部卒業。2019年 同大学大学院開放環境科学専攻修士課程修了。在学中、生成器モデルによる画像生成の研究及び感性工学の研究に従事。



萩原 将文 (正会員)

1982年 慶大・工・電気卒。1987年 同大学院博士課程修了。工博。同年同大助手。現在、同大教授。1991-92年度スタンフォード大学訪問研究員。視覚・言語・感性情報処理とその融合の研究に従事。1990年 IEEE Consumer Electronics Society 論文賞、1996年 日本ファジィ学会著述賞、2004年、2014年 日本感性工学会論文賞、2013年 日本神経回路学会最優秀研究賞受賞。2015-17年 日本知能情報ファジィ学会会長。