

# GANにより学習された視覚的重要度に基づく自動デザイン機構

石山航平<sup>1</sup> 牛尼剛聡<sup>1</sup>

あらまし 本論文では、敵対的生成ネットワーク(GAN)を利用してグラフィックデザインを自動生成する手法を提案する。GANを利用した画像生成において、学習済みの生成器に対してノイズを入力として画像を生成するのが一般的である。しかし、グラフィックデザインのように入力として用いる素材画像やテキスト等がユーザによって指定される場合、それらを利用してGANにより適切なグラフィックデザインを生成することは困難である。グラフィックデザインは配置、配色など様々なデザイン要素から構成されるが、本研究ではレイアウトに注目する。本論文では、訓練データとなるグラフィックデザインを、それらの視覚的重要度を表したマップ画像に変換し、マップ画像にしてGANによって生成器と識別器を学習させる。そして、学習済みの識別器を利用して、指定された素材を利用して自動的に生成したレイアウトの評価を行う。提案手法の有効性を評価するために、被験者を利用した主観評価による実験の結果、提案手法の有効性が示された。

**キーワード**：敵対的生成ネットワーク(GAN)、自動デザイン、バナー広告、生成モデル、画像生成、深層学習

## 1. はじめに

近年、AI・機械学習技術の進歩により、コンピュータを用いたコンテンツの自動生成に関する研究が活発化している。その中でも、敵対的生成ネットワーク(GAN)[1]の出現により、機械学習技術に基づいたコンテンツ生成は飛躍的な進歩を遂げている。

コンテンツの自動生成が期待される分野には様々なものが存在する。その代表的な分野の一つに、グラフィックデザインがある。現在、広告やポスターなどといったグラフィックデザインのほとんどは専門的な知識と経験を持ったプロのデザイナーによって制作されている。機械学習技術によってデザインの自動生成を行うことができれば、ユーザの特性や目的に応じて、最適なコンテンツを提供可能になる等様々な効果が期待できる。

コンテンツの自動生成を行う代表的な手法の一つであるGANは、生成器と識別器という2種類のニューラルネットワークから構成され、生成器はノイズからコンテンツを生成するニューラルネットワークであり、識別器は入力されたデータが本物のデータであるか、生成器で生成されたデータであるかを見分けるニューラルネットワークである。これらの生成器と識別器が敵対的に学習を行うことによって、生成器が高品質なコンテンツを生成できるようになる。

これまでも、GANを利用してグラフィックデザインを生成する研究がいくつか提案されている。それらは訓練データとして利用する画像のレイアウト情報に関して人手によるアノテーションデータを用いてモデル構築を行っている。

本論文で提案する手法では、訓練データとして与える画像データは、そこに含まれる各素材要素について配置や色などのデザインパラメータに関する情報が得られることを前提とせず、アノテーションが行われていない生の画像デ

ータのみを用いる。これにより、既存の大量の広告画像データを、そのまま訓練データとして利用できるため、様々な分野の様々な目的のグラフィック生成に適用可能という利点がある。

本論文で提案する手法の概要を図1に示す。本論文では対象とするグラフィックデザインの種類としてバナー広告を例題として利用する。具体的には、商品画像やイメージ画像といった素材画像と広告文を入力として、それらの素材の特徴に応じて、広告文の配置や大きさを自動的にデザインし、バナー広告として適切な広告画像を生成する。



図1 提案手法の概要

## 2. 関連研究

これまで行われてきた、グラフィックデザインを対象として、レイアウトの自動生成や、背景画像に対するテキストの自動配置に関する研究が活発に行われてきた。本章では、それらの研究について述べ、本研究の位置づけを明確にする。

### 2.1 レイアウトの自動生成に関する研究

Sreekanth[2]らは、遺伝的アルゴリズムを用いて、バナー広告のレイアウトを生成する手法を提案している。この手法では、バナー広告を構成する要素の重複や、要素間の距離に関する評価指標を定義し、それらを遺伝的アルゴリズムによって最適化することによって、バナー広告のレイアウトを自動的に決定する。

Jianan[3]らは、GANを利用して、様々なグラフィックデ

ザインのレイアウトの生成を行う手法を提案している。この手法では、各要素を配置するためのパラメータを生成する生成器と配置に関するパラメータ群からレイアウト情報を表すワイヤフレーム画像に変換し、畳み込みニューラルネットワーク[4]によって本物か偽物かを判別する識別器の機構を利用している。

Xinru[5]らは、入力に応じた最適な雑誌のレイアウトの自動生成手法を提案している。この手法では、入力素材の特徴を VGG16[6]や Word2vec[7]によってエンコードし、それらを生成器と識別器に組み込み、生成器のノイズを入力に応じた条件付き分布からサンプリングする。

上記の手法は全て、訓練データとするグラフィックに関して、それを構成する画像やテキストの要素の位置やサイズ等のレイアウト情報が利用できることを前提としている。しかし、グラフィックコンテンツを自動生成するために、そのグラフィックコンテンツを構成するレイアウト情報が利用可能であるとは限らない。特にインターネット上に存在する広告画像を訓練データに利用しようとしても、それらに関しては、構成要素のレイアウト情報をデータとして利用できない場合が多い。レイアウト情報がわからない結果としてのグラフィックコンテンツを訓練データとして利用することができれば、より汎用性が高いグラフィックコンテンツ生成手法となる。本論文では、各要素のレイアウト情報が与えられていない生画像を訓練データとした手法を提案する。

## 2.2 テキストの自動配置に関する研究

Yang[8]らは、テキスト要素が背景画像の被写体部分に重複しないという条件の下で、テキストを適切な位置に配置するために画像中の注意をひきやすい視覚的特性を持った領域を表す顕著性マップ[9]を用いて、既存のテキストのレイアウトのテンプレートの集合から最適なテキスト配置を行う手法を提案している。また、大峠[10]らは、背景画像にニューラルネットワークの物体検出手法におけるネットワークの出力である物体マップを利用して、テキストを配置する位置を決定する手法を提案している。しかし、これらの手法は「グラフィックデザインとしての妥当性」という観点から生成した画像を評価することができない。

## 3. 提案手法

### 3.1 全体像

本論文では、生成の対象とするグラフィックコンテンツの種類をバナー広告とする。本論文で提案するバナー画像生成手法の全体像を図2に示す。提案手法では、ユーザがバナー広告の構成要素として以下を入力として明示的に与えることを前提とする。

- 素材画像
- ロゴ画像
- 広告文

### ● ボタン画像内容文

デザイン機構は、入力された構成要素を利用してグラフィックコンテンツの候補を生成する。形式的には、構成要素の位置、大きさ、色等に関するパラメータからなるベクトル  $\mathbf{p}$  に基づいて、画像  $I$  を生成する。

$$I = \text{Design}(\mathbf{p})$$

$$\mathbf{p} = \begin{pmatrix} \text{素材画像の座標} \\ \text{素材画像の拡大率} \\ \text{ロゴ画像の座標} \\ \text{ロゴ画像の拡大率} \\ \text{広告文の座標} \\ \text{広告文の拡大率} \\ \text{ボタン画像の座標} \end{pmatrix}$$

評価器 *Evaluate* は画像  $I$  のデザインのレイアウトの適切さを評価しスコア  $\mathbf{s}$  を返す関数である。

$$\mathbf{s} = \text{Evaluate}(I)$$

システムは、評価器が出力したスコアが高いものを、バナー広告画像としてユーザに提示する。

### 3.2 デザイン機構

ユーザが入力したバナー広告画像の構成要素に関してはそれぞれのバナー広告画像上での位置、幅、高さがデザインパラメータとなる。すなわち、バナー広告をデザインするという処理は、それぞれの構成要素に対して適切なデザインパラメータを決定することであると考えられる。提案手法では、デザインパラメータの決定は、デザイン機構が行う。デザイン機構は、乱数によってそれぞれのパラメータを決定してバナー広告画像候補を生成する。表1に、本研究で対象とするバナー広告画像の構成要素と、そのパラメータを示す。

### 3.3 評価機構

デザイン機構が生成したバナー広告画像候補に対して、評価機構が、デザインの適切さを評価する。デザインの適切さは様々な観点から考えることができる。本論文では、インターネット上に存在する大量のバナー広告画像のデザインの適切さを文字やボタンや画像類などの素材のレイアウト、文字の大きさや配置に関するレイアウトの観点から捉え評価する。

提案手法において評価器は、GAN を利用して学習させた識別器を用いる。本研究ではバナー広告画像はデザイナーがデザインの専門的な知識を利用して作成されておりデザインとして適切なものであると考える。このとき、デザイン機構が自動的に生成したバナー広告画像が、既存のバナー広告画像と区別することができなければ、その候補はバナー広告画像として適切な性質を有する可能性が高いと考える。

GAN を利用して識別器と生成器の学習を繰り返すことによって、学習済みの識別器は入力画像に対して、潜在的

な「バナー広告のデザインとしての妥当性」を評価可能となる (図 2) .

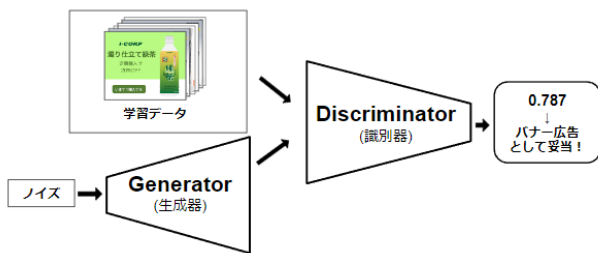


図 2 GAN の学習と識別器

#### 4. レイアウト評価機構の構築

提案手法では, GAN による学習済みの識別器を, 生成された画像がバナー広告画像としての妥当性を評価する評価器として利用する.

評価機構は, 入力として与えられる文字やビジュアルを含む各素材の大きさと配置に関してバナー広告画像としての妥当性という観点から評価を行う. 評価機構を構築するため, 入力として与えられるグラフィックデザインの画像から視覚的に重要度を示した重要度マップに変換する(図 3). マップ画像の変換に事前学習済みの Visual Importance Model[12]を用いる. これは人間がグラフィックデザインを見たときの注目領域を深層学習を用いて構築したもので, 同様のものとして前述の顕著性マップがあるが, これは主に自然画像を対象としていて, 事前知識を含まないボトムアップ効果によるモデルであるため Visual Importance Model の方がより適切にバナー広告画像のレイアウトの特徴を表していると考えられる.

訓練データとするバナー広告画像の視覚的重要度を表すマップ画像を訓練データとして, 画像を生成するように GAN を学習させる. そこで得られる識別器は素材のレイアウトに関してバナー広告画像としての妥当性という点が評価できるようになると考えられる.

GAN に関してはいくつかのモデルが提案されている. 本研究では, GAN のモデルとして PGGAN(Progressive Growing of GAN)[13]を利用する. PGGAN とは学習の過程で生成ネットワークと識別ネットワークに層を追加しながら, 対応する画像の解像度を上げていく GAN の手法であり, 式(1)に示す損失関数を用いて識別器の学習を行う.

$$L = \underbrace{\mathbb{E}_{\hat{x} \sim \mathbb{P}_g} [D(\hat{x})] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)]}_{\text{critic loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x} \sim \mathbb{P}_g} \left[ \left( \|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right]}_{\text{Gradient penalty}} \cdot (1)$$



図 3 評価器の学習のための PGGAN の 300000 エポックにおける生成器の出力画像例

ここで,  $D(x)$ は画像  $x$  に対する識別器の予測結果であり,  $\mathbb{P}_g$ は生成器によって生成された画像集合,  $\mathbb{P}_r$ は訓練データとなる画像集合である.

本手法では学習済みの識別器に対して, (1)式 critic loss を評価値算出に用いる. 実際にデザイン機構によって生成される画像  $I$  を識別器に入力したときの出力の値と, 学習に用いたバナー画像を変換した重要度マップの識別器の出力の平均のユークリッド距離をそれぞれの画像の評価値  $s$  として用いる(式 2).

$$s = \left| [D(I)] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] \right| \quad (2)$$

そのため, 本手法において  $s$  の値は小さいほど望ましく, バナー広告のレイアウトとして妥当だといえる. 評価機構のための PGGAN の機構を以下に示す.

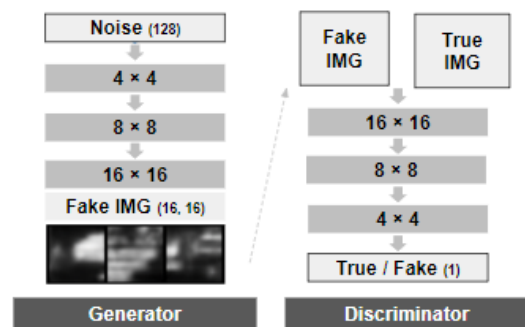


図 4 評価機構構築のための PG-GAN

#### 5. 評価実験

##### 5.1 バナー生成手法

提案手法の有効性を評価するために評価実験を行った. 学習用のデータセットは Web 上に存在する解像度  $300 \times 250$  画素のバナー画像 7200 枚である. バナー画像の収集には Moat1を利用した. また実験に使用したコンピュータの性能は以下の通りである.

OS : Ubuntu18.04LTS 64bit

1 “Measure consumer attention across channels, devices, and platforms”, <https://www.moat.com/> (参照 2022-08-08)

CPU : Intel(R) Xeon(R) Silver 4110  
GPU : NVIDIA Geforce RTX 2080 Ti  
Memory : 94GB

また、実装には python の深層学習フレームワークである PyTorch2を用いた。

## 5.2 評価器の実装

4章で述べた PGGAN をバッチサイズ 4 でそれぞれ配色評価器と文字レイアウト評価器の学習を行った。学習の際の最適化関数は Adam[14]を用いた。生成器で出力された Visual Importance を表すマップ画像の例を図 5 に示す。

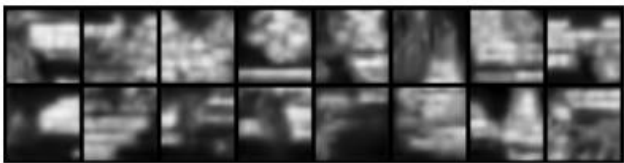


図 5 評価器の学習のための PG-GAN の 300000 エポックにおける生成器の出力画像例

## 5.3 バナー生成結果

提案手法におけるデザイン機構に表 1 に示す素材を入力として、素材画像の位置と大きさ、ロゴ画像の位置と大きさ、広告文のフォントサイズと位置、ボタンの位置と大きさをすべて乱数によってそれぞれが重複を許さない形で決定してバナー画像を 40 枚生成した。

表 1 デザイン機構に入力した素材群

素材画像	
ロゴ画像	
広告文	濁り仕立て緑茶
ボタン内容文	ボタン内容文

## 5.4 被験者実験 1 の手法

本研究では提案手法の有効性を示すため主観評価による被験者実験を行った。5.3 で作成したランダムなレイアウトのバナー画像 40 枚を対象に「バナー広告のレイアウトとして適切だと思いますか?」という質問に対して 5 段階評価を行ってもらった。被験者の人数は 42 名である。

## 5.5 被験者実験 1 の結果

5.4 の被験者実験の結果と提案手法のそれぞれの評価値の相関について 40 枚の画像の内、主観評価でのレイアウトの評価が高かった 10 枚の画像について、5.2 で構築を

した提案手法の評価機構の出力から求められる評価値の順位との相関を Normalized Discounted Cumulative Gain (nDCG)[15] を用いて表した。比較対象としては前述の顕著性マップによって変換を行ったバナー画像を PGGAN を用いて評価器の構築を行ったもの(Saliency + PGGAN)と、Visual Importance Model によって変換を行い DCGAN によって評価器の構築を行ったもの(Visim + DCGAN)を用いた。結果は以下の表 2 に示すとおり提案手法(Visim + PGGAN)の nDCG の値が一番高くなり有効性を示すことができた。

表 2 評価実験結果

Model	nDCG
Visim + PGGAN (提案手法)	0.67
Saliency + PGGAN	0.49
Visim + DCGAN	0.46

## 5.6 実験 2

5.3 の方法でバナー画像を 10000 枚生成し、5.2 で構築をした提案手法の評価機構に入力した際の出力として評価値の高かったもの 5 枚と低かったもの 5 枚を以下の図 6, 7 に示す。提案手法において予測値の値が高かったものは各要素が適度に大きく均衡のとれたレイアウトの画像を抽出できており、予測値の値が低かったものは各要素が小さく不規則に配置されており非常に見難いレイアウトとなった。



図 6 提案手法の評価機構における予測値上位 5 件



図 6 提案手法の評価機構における予測値下位 5 件

## 6. まとめ

本手法は、Visual Importance を表す画像を GAN で学習した識別器を用いてバナー広告のレイアウトの自動評価を実現し、評価実験によりその有効性を示すことができた。

今後は自動生成に向けて、配色などにも考慮した評価機構を個別に構築し、より総合的に評価を行えるように評価機構の構築を検討している。

**謝辞** 本研究は JSPS 科研費 19H04219 の助成を受けたもの

2 “PyTorch”, <https://pytorch.org/> (参照 2022-08-08)

です。

## 参考文献

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Neural Information Processing Systems*, pp.
- [2] Sreekanth Vempati, Korah T Malayil, Sruthi V, Sandeep R, "Enabling Hyper-Personalisation: Automated Ad Creative Generation and Ranking for Fashion e-Commerce", *Workshop on Recommender Systems in Fashion*, arXiv preprint arXiv:1511.06434, 2019
- [3] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, Tingfa Xu, "LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators", *ICLR 2019*
- [4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proc. of the IEEE*, 86(11), pp.2278- 2324, 1998
- [5] Xinru Zheng, Xiaotian Qiao, Ying Cao, Rynson W.H. Lau "Content-aware Generative Modeling of Graphic Design Layouts", *SIGGRAPH2019*, 2019
- [6] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", *ICLR*, 2015
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S corrado, and Jeff Dean, "Distributed Representations of Words and Phrases and their Compositionality. in proc", *Neural Information Processing System 2013*, 3111-3119
- [8] XUYONG YANG, TAO MEI, YING-QING XU, YONG RUI, SHIPENG LI, "Automatic Generation of Visual-Textual Presentation Layout", *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2016
- [9] Itti, Laurent, Christof Koch, and Ernst Niebur. "A model of saliency-based visual attention for rapid scene analysis." *IEEE Transactions on pattern analysis and machine intelligence* 20. 11, pp. 1254-1259, 1998
- [10] 大峠 和基, 大谷 まゆ, "バナー制作のための背景を考慮した自動テキスト配置", *人工知能学会 第34回全国大会 2020*
- [11] 伊藤博臣. *エンジニアのための理論でわかるデザイン入門*. インプレス, 2017, 22p.
- [12] Bylinskii, Zoya, et al. "Learning visual importance for graphic designs and data visualizations." *Proceedings of the 30th Annual ACM symposium on user interface software and technology*. 2017.
- [13] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
- [14] Kingma, Diederik P and Ba, Jimmy Lei. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412. 6980, 2014
- [15] Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4), 422-446.