

# GANにより学習されたデザイン知識を利用した自動デザイン機構

石山 航平<sup>†</sup> 牛尼 剛聡<sup>‡</sup>

<sup>†</sup>九州大学大学院芸術工学府 〒815-0032 福岡県福岡市南区塩原 4-9-1

<sup>‡</sup>九州大学大学院芸術工学研究会 〒815-0032 福岡県福岡市南区塩原 4-9-1

E-mail: <sup>†</sup> ishiyama.kouhei.842@s.kyushu-u.ac.jp, <sup>‡</sup> ushiama@design.kyushu-u.ac.jp

**あらまし** 本論文では、機械学習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。近年、機械学習を利用した画像生成手法の一つとして敵対的生成ネットワーク (GAN) が注目されている。一般的に、GANを利用した画像生成では、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。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、入力に応じて適切なデザインを生成するということが困難である。この問題を解決するために、本研究では、大量のグラフィックデザインの画像を用いて複数のデザインの観点からGANを学習させることで得られる学習済みの識別器を利用して、与えられた入力素材を利用して自動的にデザインされた画像を評価することが可能になり、グラフィックデザインを自動生成する手法を提案する。さらに、被験者による主観評価による評価実験によって提案手法の有効性を示す。

**キーワード** 敵対的生成ネットワーク(GAN), 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。特に敵対的生成ネットワーク(GAN)[1]の出現により、機械学習技術を利用し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の研究は様々な分野で行われているが、グラフィックデザインは自動生成が強く求められる分野の一つである。現在、広告やポスター等のグラフィックデザインのほとんどは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、デザイナーの負担を軽減することが期待できる。また、ユーザの嗜好や目的に応じて、適切なコンテンツを低コストで利用可能となることが期待される。

機械学習を利用したコンテンツの自動生成の代表的な手法であるGANでは、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る。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コンテンツが教師データであるか、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。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生成を行うことができるようになる。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数多く提案されているが、与えられる画像や文字を利用して、画素レベルの画像生成を行うのは

困難なため、そのほとんどの研究が各素材のレイアウト情報のみに注目し、レイアウトの生成を行っている[2,3,5]。そのため、本研究では、レイアウトのみではなく、配色や各素材の組み合わせなどにも注目した総合的なグラフィックでデザインの生成を目指す。

また、レイアウト生成の研究ではレイアウト情報についてアノテーションされたデータを用いてモデル構築を行っているが、本研究では汎用性を高めるために、教師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、結果として与えられる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教師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拡張可能となる利点がある。

本論文で提案する手法の概要を図1に示す。本論文ではグラフィックデザインの対象としてバナー広告を例題として利用する。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や色等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

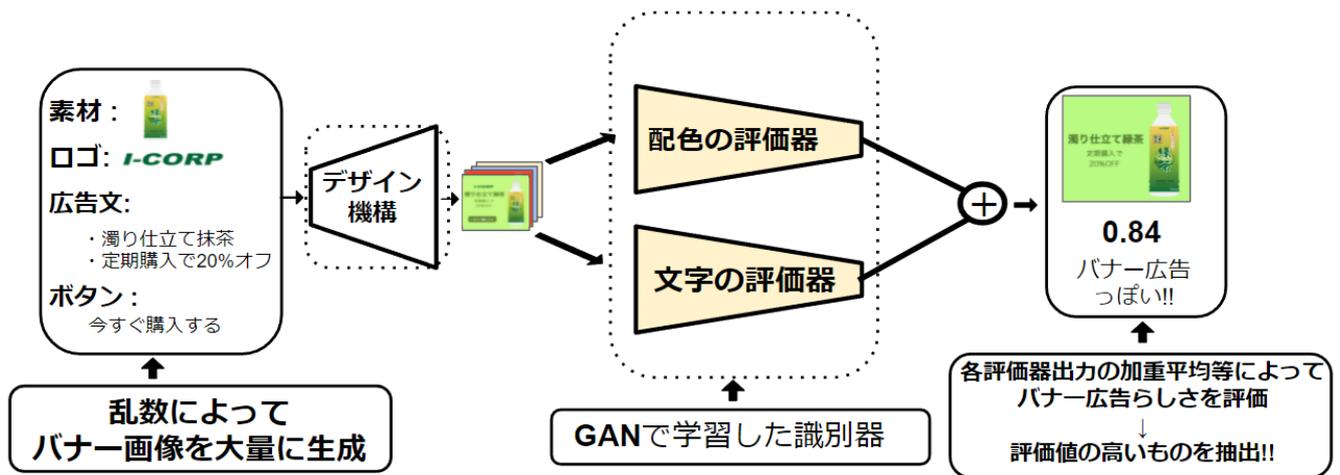


図2 バナー生成システムの全体像

## 2. 関連研究

これまで、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郡からピクセルレベルの画像を生成するワイヤーフレームレンダリング機構を取り入れ、その画像を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、教師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予め取得できることを前提としている。しかし、グラフィックを自動生成するためにそのグラフィックを構成するためのレイアウト情報が取得可能であるとは限らない。例えば、

インターネット上に存在する広告画像に関しては、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イメージを教師データとして広告を生成することができれば、より汎用性が高い広告生成手法となる。本論文では、各要素のレイアウト情報が不明な完成画像を教師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、これらの手法は入力画像のみの特徴を用いており、「バナー広告の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できない。

## 3. 提案手法

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。本論文では、生成の対象とする画像をバナー広告とする。本手法では、ユーザがバナー広告の構成要素として以下を明示的に入力として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 *Design* は、入力された構成要素を利用し

てデザインを決定する関数である．具体的には，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以下の表 1 に表すベクトル  $p$  に基づいて，画像  $I$  を生成する．

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

ユーザが入力したバナー広告画像の構成要素はそれぞれ，バナー画像上での位置，サイズ，背景色，前景色等の情報をデザインパラメータとして扱う．このとき，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になる．本手法では，デザインパラメータの決定は，デザイン機構によって行われる．デザイン機構では，乱数によってそれぞれのパラメータを決定してバナー画像候補を生成する．ただし，各素材の重複は許さないものとする．

評価器  $Evaluate$  は画像  $I$  のデザインの適切さを評価しスコア  $s$  を返す関数である．

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

デザイン機構によって生成されたバナー広告画像候補に対して，評価機構が，バナー広告画像としての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．本研究では，インターネット上に存在する大量のバナー広告の適切さを，「配色」，「文字のレイアウト」の 2 種類の観点から捉え，それぞれの観点からの適切さを評価する評価器を利用する．評価器は，GAN を利用して学習済みの識別器を用いる．

大量のバナー広告を学習データとして，GAN によって学習が行われた識別器の出力は，入力データが本物のバナー広告であると判断した確率である．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いると考える．そこで，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 3) ．

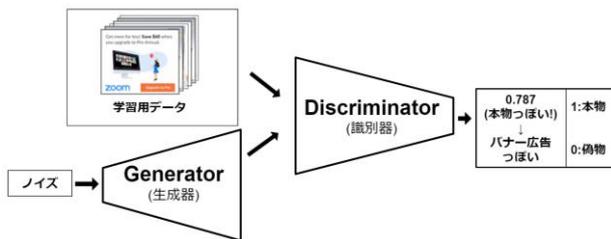


図 3 GAN の学習と識別器

表 1 デザイン機構によって乱数によって決定されるパラメータ群

乱数によって決定されるパラメータ	
背景	色 (R, G, B)
広告文	文字色 (R, G, B) フォントサイズ
素材画像	座標 (x, y)
	拡大倍率
	座標 (x, y)
ロゴ	拡大倍率
	座標 (x, y)
ボタン	色 (R, G, B)
	座標 (x, y)

#### 4. 評価器の学習

本手法では，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．

グラフィックデザインの良し悪しを決める構成要素として，写真や画像などのビジュアル，文字，色，レイアウト，線・飾りなど様々な観点が存在する [11]．本研究においては色と文字に注目し，それらのレイアウトについてデザイン的に適切かどうかの評価を自動的に行うことを目指す．

それぞれの観点からのデザイン的な妥当性を学習するため，学習データに前処理を加えることによって，「配色」，「文字のレイアウト」という 2 つのデザインの観点レイアウト情報について評価を行う評価器を構築する．最終的にそれぞれの評価器の出力に対してすべての出力の和をとり，最終的な評価値を計算する．

画像を対象とする GAN は様々な手法が提案されているが，本研究では，GAN のモデルとして PG-GAN (Progressive Growing of GAN) [12] を利用する．PG-GAN 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 の手法である．

また，すべての評価器の学習において識別器損失関数  $L$  として式 (1) に示す 2 値クロスエントロピー誤差を用いる．ここで， $y$  は識別機での出力を表し， $t$  は教師画像の場合は 1 であり，生成機で生成された画像の場合は 0 となる．

$$L = -t \log y - (1 - t) \log(1 - y) \quad (1)$$

##### 4.1 配色評価器

配色評価器は，入力として与えられる画像に対して，全体の配色の適切さに関して評価を行う．配色評価器は深層ニューラルネットワークモデルとして実現し，GAN の識別機として学習を行う．ここでは，PG-GAN で既存のバナー画像を学習する段階において， $8 \times 8$  画素の低解像度での生成時の識別器を用いる．低解像度

での生成時の識別器は全体の配色に関して広告としての適切さを評価できるようになる。ここで利用する GAN の概念図を図 4 に示す。

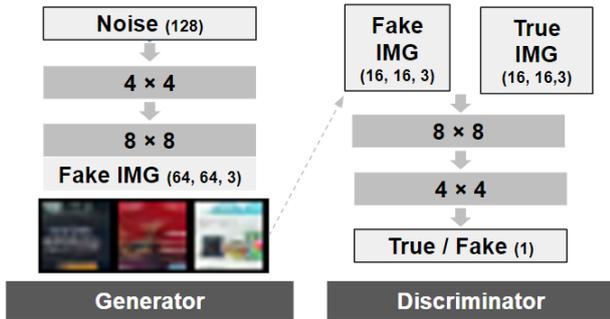


図 4 配色評価器のための PG-GAN

#### 4.2 文字レイアウト評価器

文字レイアウト評価器は、入力として与えられる画像の広告文やボタン画像の内容文といった文字の大きさと配置に関してバナー広告としての妥当性という観点から評価を行う。評価器作成のため、入力として与えられる画像から文字領域を表すマスク画像に変換する。マスク画像への変換の概要を以下の図 5 に示す。



図 5 マスク画像への変換

本研究では、文字領域の矩形検出のために Google Cloud Vision API の fullTextAnnotation1 の Paragraph のモデルを利用する。文字領域として認識された部分を 1, それ以外の領域を 0 とすることで 2 階調のマスク画像を生成する。作成された実際のバナー広告画像の文字領域を表すマスク画像を学習データとして、生成器がバナー広告画像の文字領域であるマスク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文字レイアウトに関して広告としての妥当性という点で評価できるようになる。ここで利用する GAN の機構を図 6 に示す。

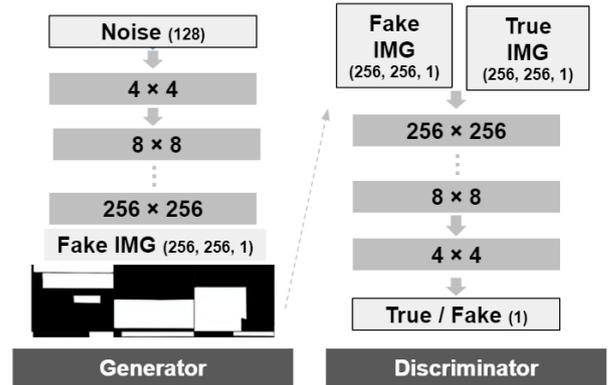


図 6 文字レイアウト評価器のための PG-GAN

### 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 Moat2 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

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 PyTorch3 を用いた。

#### 5.2 評価器の実装

4 章で述べた DCGAN を学習データのバッチサイズ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画像の例を図 7, 図 8 に示す。

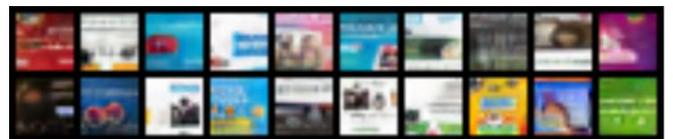


図 7 配色の評価器の学習のための PG-GAN の 4000 エポックにおける生成器の出力画像例

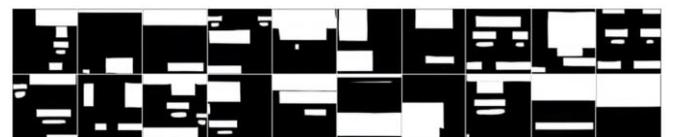


図 8 文字レイアウトの評価器の学習のための

1 “高密度ドキュメントのテキスト検出のチュートリアル”, <https://cloud.google.com/vision/docs/fulltext-annotations> (参照 2021-12-21)

2 “Measure consumer attention across channels, devices, and platforms”, <https://www.moat.com/> (参照 2021-12-21)

3 “PyTorch”, <https://pytorch.org/> (参照 2021-12-21)

PG-GAN の 60000 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 2 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と色、ボタンの位置と大きさと色、と背景を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 10,000 枚生成した。上記で得られた識別器をバナー広告画像の評価器として、それを 8×8 画素にリサイズしたものを配色の評価器の入力、前述の文字領域のみのマスク画像に変換したものを文字レイアウトの評価器の入力として評価を行った。それぞれの評価器において評価値が高かった画像上位 5 件、及び評価値が低かった画像の下位 5 件を図 9～12 に示す。

表 2 デザイン機構に入力した素材群

素材画像	
ロゴ画像	<b>I-CORP</b>
広告文	濁り仕立て緑茶
ボタン内容文	ボタン内容文



図 9 配色評価器における評価上位 5 件



図 10 配色評価器における評価下位 5 件



図 11 文字レイアウト評価器における評価上位 5 件



図 12 文字レイアウト評価器における評価下位 5 件

5.4 被験者実験

5.1 で生成したバナー画像に対する印象を評価するために、被験者を利用した主観評価実験を行った。対象としたのは、配色評価値と文字評価器とそれぞれを足した総合評価器による評価値の上位と下位それぞれ 500 件から無作為に 4 件ずつを抽出した合計 24 件の画像の比較を行った。

対象とするバナー画像を対象に、被験者に対してアンケートによる主観評価を行った。具体的には、ランダムに並べ替えられたバナー画像に対して、以下の質問にそれぞれ 5 段階評価で回答をしてもらった。ここでは、数字が大きいほど良いものとした。被験者の人数は 41 名である。

- (1) 全体の配色は適切だと思いますか？
- (2) テキスト要素のレイアウト（素材画像は無視）は適切だと思いますか？
- (3) 全体的にバナー広告のデザインとして適切だと思いますか？

5.5 被験者実験

被験者実験によるそれぞれの評価器ごとの結果を図 13 に示す。図 13 には 3 種類の結果が含まれており、左から以下の内容を表している。

1. 配色評価器での出力値の上位、または下位のそれぞれ 4 件の画像に対して「全体の配色は適切だと思いますか？」という質問に対する返答の集計結果の平均をとったもの
2. 文字レイアウト評価器での出力値の上位、または下位のそれぞれ 4 件の画像に対して「テキスト要素のレイアウトは適切だと思いますか？」という質問に対する返答の集計結果の平均をとったもの
3. 総合評価器での出力値の上位、または下位のそれぞれ 4 件の画像に対して「全体的にバナー広告のデザインとして適切だと思えますか？」という質問に対する返答の集計結果の平均をとったもの

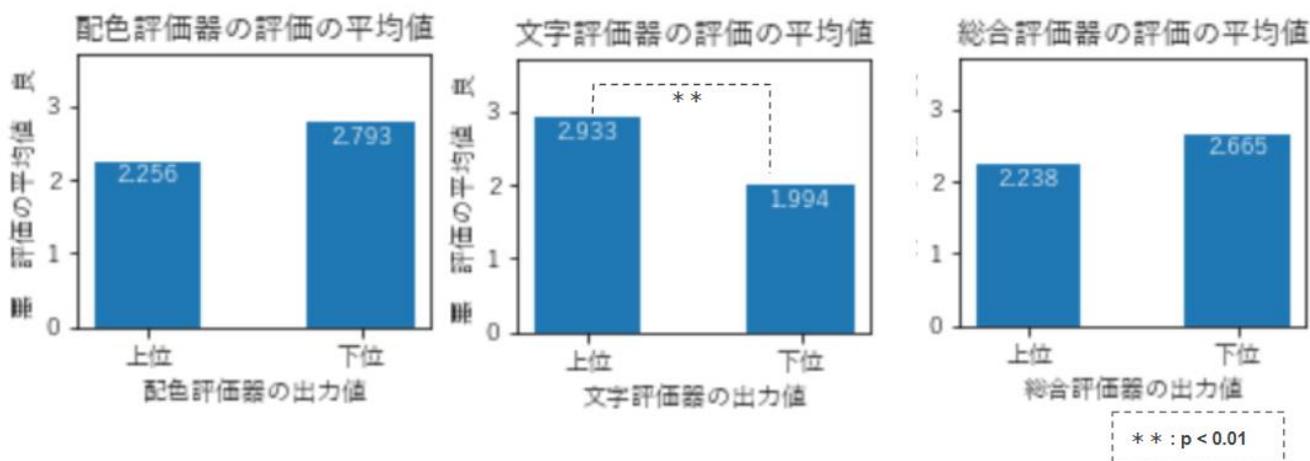


図 13 被験者実験の結果

文字レイアウト評価器に関しては、評価器の出力と被験者による主観評価には相関がみられた。これらについて、t 検定によって評価器の出力が高い画像群と低い画像群に対する主観評価の平均値の差について検証した。両者において p 値は  $p < 0.001$  となり、優位水準 0.01 を下回り、有意差が示された。

しかし、配色評価器と総合評価器においては、評価器の出力と被験者による主観評価では逆の相関がみられ、期待した結果は得られなかった。今回の実験では 7200 枚という比較的少ない画像から評価器を構築したため、うまく評価できなかった可能性がある。今後はもう少しデータ数を増やして精度をあげていきたい。

## 6 まとめ

本研究では、バナー広告を対象として入力された素材に対してグラフィックデザインを生成するシステムを提案した。具体的には、バナー広告画像を学習データとして PG-GAN によって学習された識別器を評価器として、乱数によって生成される入力素材を利用したグラフィックデザインの評価を行い、グラフィックデザインとして妥当なものを抽出する。

PG-GAN によって「配色」、「文字レイアウト」という 2 点を評価する個別の評価器を GAN で学習した識別器を構築し、評価実験を行い、文字レイアウトについては有効性を示すことができたが、配色評価器また総合評価でも有効性を示すことができなかった。今後はデータ数を増やして、精度向上を目指していきたい。

また、デザイン機構については現在、パラメータを乱数によって決定しているが、深層強化学習などの手法によってグラフィックデザイン生成用いるデザインパラメータの潜在分布の最適化をしたいと考えている。

## 謝辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

## 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] KARRAS, Tero, et al. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [13] SELVARAJU, Ramprasaath R., et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*. 2017. p. 618-626.
- [14] Kingma, Diederik P and Ba, Jimmy Lei. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014
- [15] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks . arXiv preprint arXiv:1511.06434, 2015. 2, 3
- [16] 石山航平, 牛尼剛聡, “敵対的生成ネットワークを利用したバナー広告の自動生成手法”, DEIM Forum 2020,2020.3