

令和2年度 卒業研究論文

題目 GAN 出力画像をフィルタリング
するための複数の CNN を用いた
画像評価に関する研究

指導教員 服部 峻

提出者 室蘭工業大学 情報電子工学系学科

氏名 相場 築

学籍番号 17024001

提出年月日 令和3年2月12日

目次

第 1 章	序論	1
第 2 章	提案手法	2
2.1	処理 1: ぼやけ具合の評価	2
2.2	処理 2: オブジェクト評価	4
2.3	処理 3: ユーザリクエストの反映度評価	4
2.4	処理 4: 総合評価の定式化	5
第 3 章	評価実験	6
3.1	評価実験: 線形結合パラメータの依存性	6
3.1.1	実験用サンプルの概要	7
3.1.2	実験結果	8
3.1.3	考察	9
3.2	評価実験: 2つの閾値への依存性	9
3.2.1	実験結果	10
3.2.2	考察	11
3.3	評価実験: 処理 3のみを用いた線形結合パラメータの依存性	12
3.3.1	実験結果	12
3.3.2	考察	13
3.4	評価実験: 2つの重みパラメータの依存性	13
3.4.1	実験結果	15
3.4.2	考察	17
第 4 章	結論	20
4.1	まとめ	20
4.2	社会的貢献及び技術的貢献について	21
4.3	今後の課題	22
	謝辞	23
	参考文献	24

目次

2.1	提案手法の概観	3
2.2	評価値例: $\text{lap}(img) = 180.69$	3
2.3	評価値例: $\text{lap}(img) = 1959.44$	3
2.4	評価値例: $\text{lap}(img) = 327.04$	3
3.1	精度の低い GAN から得られるような画像	7
3.2	精度が中程度の GAN から得られるような画像	7
3.3	精度が高い GAN から得られるような画像	7
3.4	ユーザが求めるような実際に撮影された画像	7
3.5	キーワード 5 種全体の再現率	10
3.6	キーワード「black body」の再現率	10
3.7	キーワード「blue wings」の再現率	10
3.8	キーワード「curved beak」の再現率	10
3.9	キーワード「red wings」の再現率	10
3.10	キーワード「white belly」の再現率	10
3.11	キーワード 5 種の平均適合率 (α はキーワード毎の最良の値)	11
3.12	キーワード「black body」の適合率 ($\alpha = 0.3$)	11
3.13	キーワード「blue wings」の適合率 ($\alpha = 0.0$)	11
3.14	キーワード「curved beak」の適合率 ($\alpha = 0.4$)	11
3.15	キーワード「red wings」の適合率 ($\alpha = 1.0$)	11
3.16	キーワード「white belly」の適合率 ($\alpha = 0.0$)	11
3.17	処理 3 のみのキーワード 5 種全体の再現率	14
3.18	処理 3 のみのキーワード「black body」の再現率	14
3.19	処理 3 のみのキーワード「blue wings」の再現率	14
3.20	処理 3 のみのキーワード「curved beak」の再現率	14
3.21	処理 3 のみのキーワード「red wings」の再現率	14
3.22	処理 3 のみのキーワード「white belly」の再現率	14
3.23	数式 (3.2) のグラフ	15
3.24	キーワード 5 種全体の 10 番目までの適合率 (α はキーワード毎の固定値)	18

3.25	キーワード 5 種全体の 20 番目までの適合率 (α はキーワード毎の固定値) . . .	18
3.26	キーワード「black body」の 10 番目までの適合率 ($\alpha = 0.3$)	18
3.27	キーワード「black body」の 20 番目までの適合率 ($\alpha = 0.3$)	18
3.28	キーワード「blue wings」の 10 番目までの適合率 ($\alpha = 0.2$)	18
3.29	キーワード「blue wings」の 20 番目までの適合率 ($\alpha = 0.2$)	18
3.30	キーワード「curved beak」の 10 番目までの適合率 ($\alpha = 0.8$)	19
3.31	キーワード「curved beak」の 20 番目までの適合率 ($\alpha = 0.8$)	19
3.32	キーワード「red wings」の 10 番目までの適合率 ($\alpha = 0.8$)	19
3.33	キーワード「red wings」の 20 番目までの適合率 ($\alpha = 0.8$)	19
3.34	キーワード「white belly」の 10 番目までの適合率 ($\alpha = 0.5$)	19
3.35	キーワード「white belly」の 20 番目までの適合率 ($\alpha = 0.5$)	19

表目次

3.1	キーワード 5 種及び全体の適合率と全再現する Top k	8
3.2	数式 (3.3) を用いた時のキーワード 5 種の精度	16

第 1 章

序論

昨今、様々な場所で AI の利用、研究がなされている。有名なものでは、ソフトバンクが開発した人間の感情を認識できるロボットである「Pepper」や、掃除ロボットの「ルンバ」、自動車の自動運転などが挙げられる。また、画像を認識、分類、生成する研究や、小説の冒頭部分を生成するシステムなど多種多様である。特に、本研究で取り上げる「GAN (Generative Adversarial Network)」[1] は画像を生成することに特化しており、絵が描けない人にとっては、欲しい絵の条件を入力すると機械が絵を出力してくれるため、フリー画像を探す手間や、絵を描いてもらう料金がかからずとても有用である。

GAN とはデータを新たに生成できる生成モデルの一種で、主な構造は 2 つに分かれている。1 つ目は「Generator」であり、与えられた画像集合から類似したデータを作る。2 つ目は「Discriminator」と呼ばれ、Generator で作られたデータの真偽を入力データと比較して判別する。これを繰り返すことで、Generator は本物に近いデータを生成できるようになり、さらに、複数の入力データの部分を組み替えて、まったく新しいデータを作ることも可能である。つまり、入力データから特徴を学習し、特徴を含んだデータを新しく生成できる点が GAN の特徴である。既存のデータをマイニングするのとは違い、生成するデータの特徴を指定することでユーザの細かな要求にも対応ができ、有用性が高い。

しかし、GAN の生成画像のクオリティはまばらである。生成されたデータは輪郭のはっきりしない画像から、現実に存在しないような画像でも実際にあると思ってしまう様なものまで生成できるため、画像が他コンテンツに利用できるクオリティであるか否か、定量的に評価するのが難しく、現状の GAN の画像評価は、学習データの知識を有する者が、実際に目で見て判断することが多い。また、ユーザの求める画像に細かな指定があったり、要求する画像の量が多い場合は、要求を完璧に満たす画像を得るために大量の出力を必要とする上、要求通りではない生成画像も大量に含まれることから、生成の度にユーザの目で確認しなければならず、多大な手間が発生する。

そこで、本研究では、GAN によって生成された出力画像を『ぼやけ具合』『生成画像からオブジェクト物体が認識できる度合い』『ユーザリクエストの反映度』の 3 種類の観点から自動的に評価し、ユーザが求めるクオリティの高い画像を効率良くフィルタリングする手法について提案する。

第 2 章

提案手法

GAN はテキストを入力することで画像を生成できる構造 (text-to-image) [2] であると仮定し、学習データは画像とその画像の説明が記述されたものの組の集合とする。このような GAN から得られる大量の生成画像に対する画質評価や、生成対象評価などを自動的に行い、人の評価を介さず選別を行うことで手間を削減し、ユーザのリクエスト (例えば、「黒い体の鳥」) に合った画像を効率的に収集するための手法について提案する。

まず、提案手法の概要について説明する。概観は図 2.1 の通りである。画像とその画像の説明が記述された学習データを用いて構築した GAN から、生成画像を任意の数だけ収集し、生成画像に対して複数の処理を通して評価付けを行う。提案手法には大きく分けて 4 つの処理が存在する。

1. 画像のぼやけている度合いを計算する。
2. 対象物が認識できる度合いを計算する。
3. ユーザのリクエストの反映度を計算する。
4. 処理 1~3 の評価値を用いて総合評価を付ける。

処理 1~3 を行った後、それぞれの処理によって得られる評価値を用いて総合評価を付け、総合評価の順に画像をランキング表示することでユーザのリクエストに合った画像を優先的に表示し効率化を図る。また、「req-CNN」「web-CNN」について、詳細は後述するが、この 2 つの CNN (Convolutional Neural Network) は入力されたリクエストのキーワード (例えば、「blue wings」) 毎に用意するものであり、「blue wings」が入力された場合であれば、2 つの CNN は「青い羽根の鳥」に該当する画像の評価を高く付けるタスクを担う。図 2.1 では、1 つだけ書いてあるが、実際にはリクエストのキーワード毎に構築されるため無数に用意する必要がある。

2.1 処理 1: ぼやけ具合の評価

処理 1 では、生成画像を『ぼやけ具合』の観点で評価付けする。GAN から生成される画像はしばしば色が複雑に混ざり合ったようなぼやけている画像が出力される。このような画像は

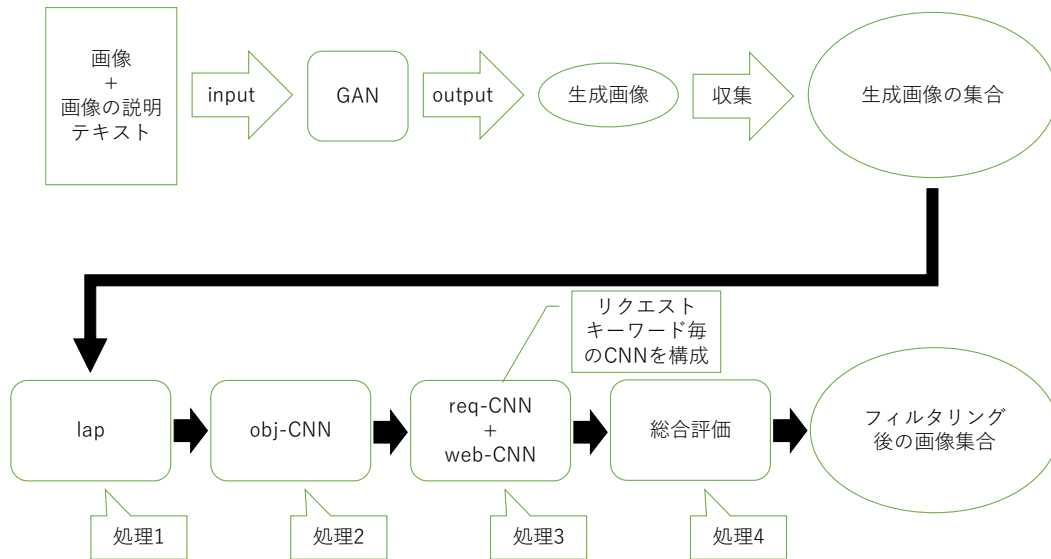


図 2.1 提案手法の概観



図 2.2 評価値例:
 $\text{lap}(img) = 180.69$



図 2.3 評価値例:
 $\text{lap}(img) = 1959.44$



図 2.4 評価値例:
 $\text{lap}(img) = 327.04$

特殊な用途を除き、他コンテンツで利用できるようなクオリティではないため、取り除くことができるように指標を用意する。OpenCV ライブラリ [3] のラプラシアンフィルタを用いることで、画像のエッジ（境界線部分）を検出し、評価対象の生成画像の分散を求め、分散の大きさに比例させて評価値を上げることによって『ぼやけ具合』を表現する。以降、画像 img を入力した時の本節の評価値を $\text{lap}(img)$ と表記する。図 2.2～2.4 のキャプションの数値は、それぞれの画像に処理 1 を行った場合の評価値の例である。図 2.2 は精度が未熟な GAN から出力された画像であり、「鳥のような何か」が写っている程度まで判断できるが、はっきりと断言することが難しいような画像である。図 2.3 は実際に撮影された画像で、誰が見ても「鳥が写っている」と断言できるような画像であり、「鳥と周辺の情景」も輪郭がはっきりと捉えられる。図 2.4 は精度が中程度の GAN から出力された画像であるが、図 2.2 よりも $\text{lap}(img)$ が高く、実際の画像である図 2.3 よりも大幅に低い。つまり、評価値 $\text{lap}(img)$ を用いることで、GAN から生成されるぼやけた画像を除外するために活用できると考えられる。

2.2 処理 2: オブジェクト評価

処理 2 では、GAN に学習させているオブジェクト（例えば、「鳥」）が認識できる度合いについての観点から生成画像毎に評価付けを行う。GAN に限らず現存する AI の殆どが、実用的な段階までの精度にするために学習データや機能を限定的にする場合が多い。本研究で想定している GAN も生成できる画像の大まかなオブジェクトは決まっているものとし、そのオブジェクトが捉えられていない画像の評価を落とすことで、取り除くための指標を用意する。オブジェクトが含まれているかを判断するために GAN と同じ学習データまたは、オブジェクトと同じカテゴリを持つデータセットを学習させた CNN を用意し、生成画像にオブジェクトが捉えられるかを数値的に出すことで評価値を出す。以降、他の CNN と区別するために本節の CNN を「obj-CNN」と呼称し、画像 img に対する obj-CNN から得られる評価値を $obj-CNN(img)$ と表記する。obj-CNN に画像を入力すると、画像のオブジェクトが認識できる確率を $0.0 \leq obj-CNN(img) \leq 1.0$ で返す。本研究の評価実験では、GAN は鳥の種類を複数学習させたものを想定しており、この場合であれば obj-CNN に分析させる要素は『鳥らしいもの』が評価軸に該当する。例として挙げると、3.1.1 節の図 3.2、図 3.4 を入力した時の評価値は、図 3.2 が $obj-CNN(img) = 0.45$ となり、図 3.4 が $obj-CNN(img) = 0.90$ と出力される。オブジェクトが捉えられていないにも関わらず、ユーザの要求が反映されているとは考えにくいいため、このような画像を取り除く目的でここで算出される評価値を 2.4 節での総合評価の数式 (2.3) で利用する。

2.3 処理 3: ユーザリクエストの反映度評価

処理 3 では、ユーザの要求（例えば、「青い羽根の鳥」）が反映されているかの観点から生成画像毎に評価付けを行う。この処理で使用する CNN も構築方法は 2.2 節と同様に GAN に学習させるデータを用いて構築を行う。ユーザの要求はキーワード形式を想定しており、GAN の学習に用いている画像の説明データからキーワードを検索し、ユーザ要求のキーワードが説明文中に含まれている画像だけを集めて CNN（以降、req-CNN と表記）を構築する。これにより、評価対象の生成画像にキーワード要素が含まれているかを req-CNN で分析できるようになる。画像 img に対する req-CNN から得られる画像の評価値 $req-CNN(img)$ を 2.4 節での総合評価の数式 (2.3) で用いることでユーザの要求に近い画像を優先的に表示する。また、本研究の評価実験ではキーワード要素を分析する req-CNN については、インターネットから得られる画像を集めて同じキーワード要素を分析できる CNN（以降、web-CNN と表記）を用意し組み合わせることで、自動的に構築した req-CNN の精度に信憑性が無い場合のためにロバスト性の向上を狙っている。

2.4 処理 4: 総合評価の定式化

これまでに挙げた4種類の評価値を基に、本節では総合評価の計算手法を提案する。まず、2.1節の評価値 $\text{lap}(img)$ と 2.2節の評価値 $\text{obj-CNN}(img)$ の値を閾値で区切って数式 (2.1) と数式 (2.2) のように2値化することで、評価の低い生成画像の表示を制限する。最適な閾値に関しては、3章で検証する。次に、2.3節の req-CNN と web-CNN については、重みパラメータ $0.0 \leq \alpha \leq 1.0$ で線形結合する。最も良い α の値を求めるため、変化させた場合の精度の変化を3.1節で観察する。総合評価 $\text{eval}(img)$ の算出方法を数式 (2.3) に記述する。 $\text{lap}^b(img)$ によって、3.1.1節の図 3.1 のような画像を排除し、 $\text{obj-CNN}^b(img)$ によって、3.1.1節の図 3.2 のようなある程度輪郭を持ちつつも形が歪な画像を排除する。その後、 req-CNN と web-CNN によって、ユーザのリクエストに合っている画像の評価値を高く設定し、総合評価の高い順に並び替えて優先表示を行うというコンセプトで作られている。

$$\text{lap}^b(img) = \begin{cases} 1 & (\text{lap}(img) \geq 300) \\ 0 & (\text{lap}(img) < 300) \end{cases} \quad (2.1)$$

$$\text{obj-CNN}^b(img) = \begin{cases} 1 & (\text{obj-CNN}(img) \geq 0.5) \\ 0 & (\text{obj-CNN}(img) < 0.5) \end{cases} \quad (2.2)$$

$$\text{eval}(img) = \text{lap}^b(img) \cdot \text{obj-CNN}^b(img) \cdot \{(1 - \alpha) \text{req-CNN}(img) + \alpha \text{web-CNN}(img)\} \quad (2.3)$$

$$(0.0 \leq \alpha \leq 1.0)$$

第3章

評価実験

本章では提案したフィルタリング手法に用いられている各種パラメータについての実験を行い、性能の向上に繋がる適切なパラメータ設定を模索していく。

3.1 評価実験: 線形結合パラメータの依存性

本節では、実際に実験を行う際に用いたデータセットや設定について述べた上で、2.3節の req-CNN と web-CNN とを線形結合するパラメータ α の依存性の実験を行い、総合評価値を用いてランキングし、上位 10 件を取り出した場合の適合率と上位 k 件の再現率の結果について述べる。

まず、GAN は『CUB-200-2011』[4] と呼ばれるカリフォルニア大学提供の、鳥の種類が 200 種登録されている 11788 枚の画像データセットと、このデータセットの画像それぞれに説明が書かれているデータ [5] を学習させたものとし、このような GAN から出力される画像をフィルタリングの対象として想定する。次に、処理 2 の obj-CNN は『鳥らしさ』を分析するために、鳥のカテゴリが含まれているデータセットである『cifar10』[6] というデータセットを用いて、鳥と鳥以外とに分けて学習させた。処理 3 の req-CNN は、GAN の学習の際に用いる『CUB-200-2011』とその説明データを使って、ユーザーリクエストとして入力されたキーワードで説明データを基に検索された画像と残りの『CUB-200-2011』からランダムに同数分取り出した画像とを学習させ、『キーワードの反映度』を分析できるようにする。さらに、キーワードに該当するものをウェブから検索して、手動で集めた画像で web-CNN を構築した。

3.1.1 実験用サンプルの概要

実際の GAN から得られる生成画像をリアルタイムに実験するとランダム性に影響されるため、本研究では GAN から得られるような画像を集めた実験用サンプルを作成して、これを用いることとする。画像毎の特徴は以下のように決める。

1. 精度の低い GAN から得られるような画像
2. 精度が中程度の GAN から得られるような画像
3. 精度が高い GAN から得られるような画像
4. ユーザが求めるような実際に撮影された画像

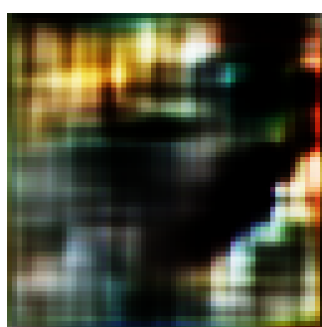


図 3.1 精度の低い GAN から得られるような画像



図 3.2 精度が中程度の GAN から得られるような画像

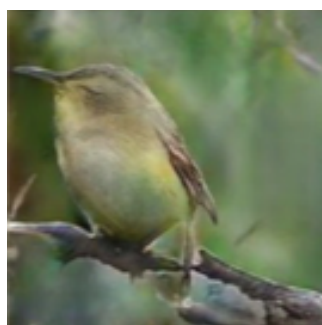


図 3.3 精度が高い GAN から得られるような画像



図 3.4 ユーザが求めるような実際に撮影された画像

図 3.1 のような画像を 16 枚、図 3.2 のような画像を 10 枚、図 3.3 のような画像を 10 枚、図 3.4 のような画像を 41 枚用意し、合計 77 枚の実験用サンプルを作成する。

また、ユーザークエストのキーワードによって結果が変化するため様々な状況での評価を行うために、入力されるキーワードを複数想定して実験を行う。キーワードは「black body」「blue wings」「curved beak」「red wings」「white belly」の 5 種類と仮定し、実験用サンプルの中にキーワード毎の正解として筆者が合っていると判断した画像を各 10 枚ずつ用意した。これらの画像は合計 77 枚の中に含まれている。

表 3.1 キーワード 5 種及び全体の適合率と全再現する Top k

	適合率			全再現する Top k		
	平均値	最大値	最小値	平均値	最大値	最小値
全体	0.50	0.70	0.30	61	77	21
black body	0.70	0.70	0.70	56	56	56
blue wings	0.60	0.60	0.60	23	25	21
curved beak	0.31	0.40	0.30	73	73	73
red wings	0.49	0.50	0.40	76	76	76
white belly	0.42	0.50	0.40	77	77	77

3.1.2 実験結果

各々のキーワードにおける 10 番目までの「適合率」と、何番目の出力ですべての正解が出揃うかの「再現率」を結果として記述する。初めに、 $0.0 \leq \alpha \leq 1.0$ における 10 番目までの適合率の平均と最大値、最小値、さらに再現率が 1.0 に達した際の Top k の値の平均と最大値、最小値についてキーワード毎に記述し、加えてキーワード 5 種の全体についても同様に表 3.1 にまとめる。また、全体の上位 k 件の再現率の変化を図 3.5 に示す。

「black body」を入力した際の 10 番目までの適合率は、線形結合パラメータ α を変化させても 0.7 で一定であり、適合率に変化はなかった。図 3.6 は「black body」を入力した時の再現率を示している。図 3.6 から上位 22 番付近までは、再現率が増加傾向にあるものの全体で見ると線形結合パラメータ α の値に関わらず、すべての正解の画像を拾い上げた Top k の値は表 3.1 を見てもわかる通り 56 番目という結果になった。正解の画像のうち 2 枚が処理 1 もしくは処理 2 の閾値によって、総合評価が 0.0 になっていることがわかり、優先表示から大きく外されてしまったことから、すべての正解を出すまでに後半まで出力する必要があったため、このような結果になったと考えられる。

「blue wings」についても「black body」と同様に、線形結合パラメータ α を変化させても 10 番目の適合率は 0.6 で一定という結果になった。次に図 3.7 の「blue wings」の再現率を示す。図 3.7 から、「black body」の再現率とは違い正解データの総合評価が 0.0 となる評価は付けられておらず、早い段階ですべての正解を拾い上げていることがわかる。しかし、上位 25 番前後までは出力しなければならず、処理 3 の「req-CNN」と「web-CNN」の評価付けにおいて、正解の画像よりも高い評価を付けている画像が存在したことが原因でこのような結果になったと考えられる。

「curved beak」の適合率については、線形結合パラメータ $\alpha = 0.0$ の場合のみ 0.4 という最大値になり、その他の α の値ではすべて 0.3 という結果になった。次に「curved beak」についての再現率を図 3.8 に示す。正解の画像中の 4 枚が処理 1 と処理 2 の閾値により総合評価が 0.0 になり、優先表示から大きく外されていることや、図 3.8 の結果から、前半の推移を見ても残りの正解の画像 6 枚が出力されるまで 34 番目まで出力を要していることと、表 3.1 の平均適合率が低い値であるという点から処理 3 の「req-CNN」と「web-CNN」の分類能力が

低いことがわかる。これは、「curved beak」という鳥のパーツとしても小さな特徴について CNN が規則性を見つけられず、分類が難しくなったためであると考えられる。

「red wings」の適合率は、線形結合パラメータ $\alpha = 0.0$ の場合のみ 0.4 という最小値になり、その他の α の値では 0.5 という結果になった。次に「red wings」についての再現率を図 3.9 に示す。「red wings」については、正解の画像中から 4 枚の画像が総合評価値に 0.0 を付けられており、優先表示から外れている。図 3.9 の結果から線形結合パラメータ $\alpha = 1.0$ の場合で、他の値の時よりも正解を優先していることが確認できる。これは、処理 3 の「req-CNN」で学習用に「red wings」のキーワード検索をして抽出された画像に「red wings」の要素が含まれていない（または、少ない）画像が多かったため精度が出ず、「web-CNN」に重みを置いて用いた結果の方が正解を優先的に表示できたためであると考えられる。

「white belly」の適合率は、線形結合パラメータ $\alpha = 0.0$ と $\alpha = 0.1$ の時に 0.5 の最大値になり、その他の値では 0.4 という結果になった。次に「white belly」についての再現率を図 3.10 に示す。「white belly」については、正解の画像中から 4 枚の画像が総合評価値に 0.0 を付けられており、優先表示から外れている。表 3.1 の結果からは平均適合率が低いことがわかるが、図 3.10 の再現率推移から、28 番目付近まで出力することで線形結合パラメータ α の値に関わらず、比較的早い段階で残りの 6 枚は優先的に表示されていることが確認できる。つまり、処理 1 及び処理 2 の閾値に依って 0.0 と評価付けされ、除外されていることが原因で適合率が下がっていると考えられる。

3.1.3 考察

どのキーワードの適合率の結果を見ても、線形結合パラメータ α を変化させることによる 10 番目までの適合率の大きな変化は見込めない結果となった。これは、処理 3 の「req-CNN」及び、「web-CNN」によるリクエスト反映度の評価によって高い評価値を得られているものの、閾値で 2 値化した lap や obj-CNN の評価によって、正解の画像が各再現率推移の結果から間引かれており、10 番目までの適合率を表示すると、lap, obj-CNN で除外された正解の画像が優先から外れ、10 番以内で連続で正解を当てていたとしても、10 番目まで表示すると総合的に適合率が同じになってしまったためであると考えられる。これは、各キーワードの再現率推移が中間の件数の辺りで大きく間隔を空けてから再現率が 1.0 に収束している点からも見受けられる。さらに全体の再現率の平均である図 3.5 からも、ほとんどのキーワードが正解の画像を除外されていることで後半まで出力する必要がある状態であることが見て取れる。

3.2 評価実験: 2 つの閾値への依存性

3.1 節の「線形結合パラメータ α への依存性」に関する評価実験の結果は、処理 1 の lap, 処理 2 の obj-CNN により正解の画像の一部が除かれていたことが原因で、処理 3 の req-CNN と web-CNN とを線形結合するパラメータ α を変化させても適合率に差が生じず、複数の CNN を組み合わせることによるロバスト性の向上にはつながらなかった。そこで本節では、

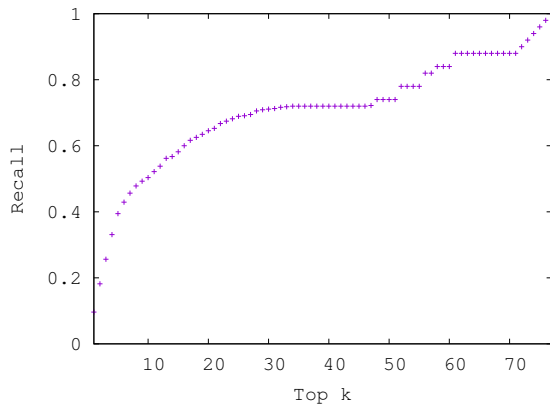


図 3.5 キーワード 5 種全体の再現率

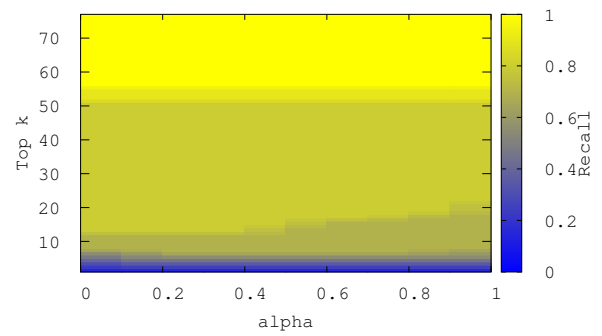


図 3.6 キーワード「black body」の再現率

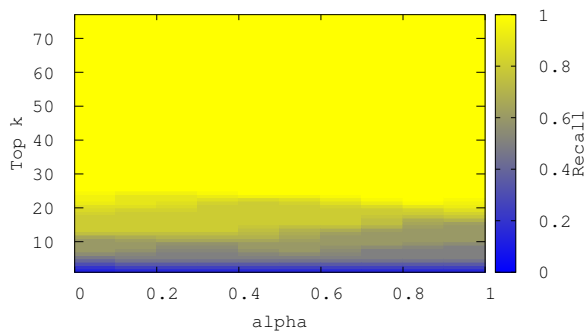


図 3.7 キーワード「blue wings」の再現率

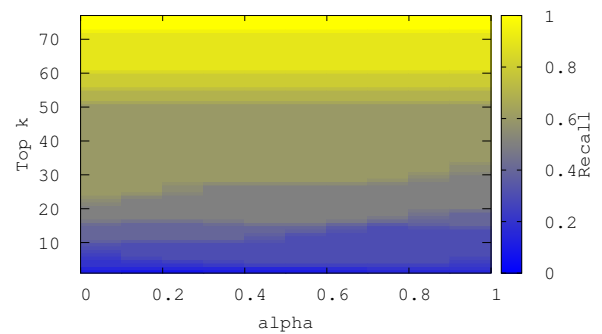


図 3.8 キーワード「curved beak」の再現率

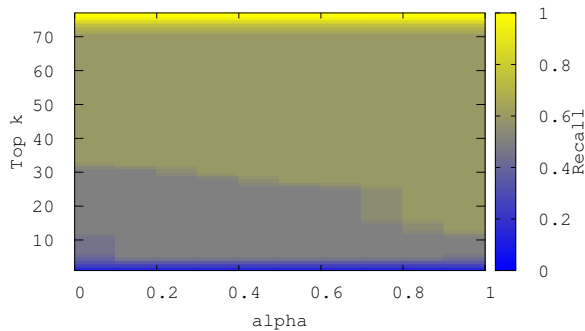


図 3.9 キーワード「red wings」の再現率

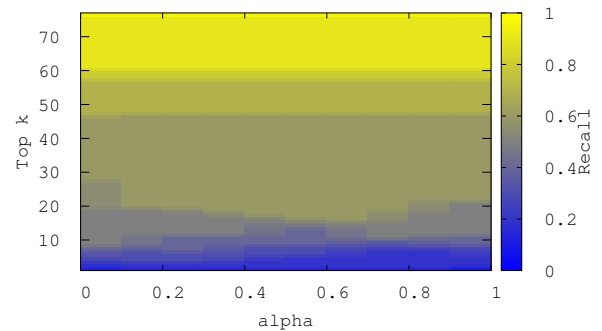


図 3.10 キーワード「white belly」の再現率

各キーワードについて、2.4 節の総合評価の数式 (2.1) 及び数式 (2.2) の閾値を変化させて最も高い適合率となる組を求める。

3.2.1 実験結果

キーワード毎、平均的に適合率の高い値を出している時の線形結合パラメータ α の値に固定して、数式 (2.1) の閾値と数式 (2.2) の閾値を変化させた場合の上位 10 件までの適合率の変化を 2 次元カラーマップとして表示する。各図のラベル「lap th」「obj-CNN th」「Precision」はそれぞれ、「数式 (2.1) の閾値」「数式 (2.2) の閾値」「上位 10 件までの適合率」を示している。

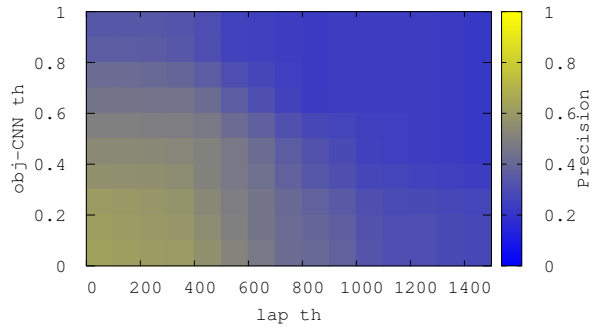


図 3.11 キーワード 5 種の平均適合率 (α はキーワード毎の最良の値)

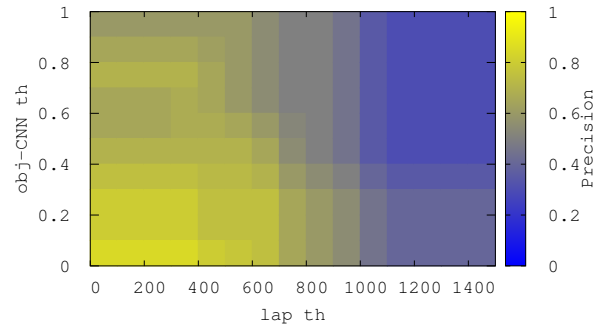


図 3.12 キーワード「black body」の適合率 ($\alpha = 0.3$)

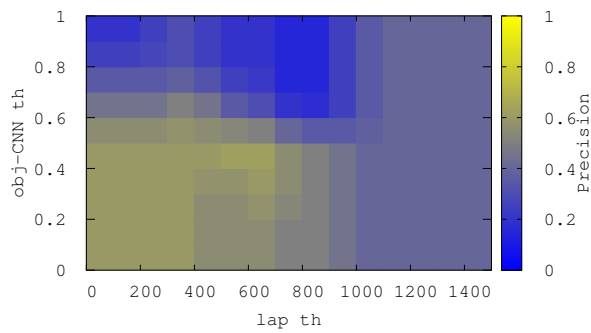


図 3.13 キーワード「blue wings」の適合率 ($\alpha = 0.0$)

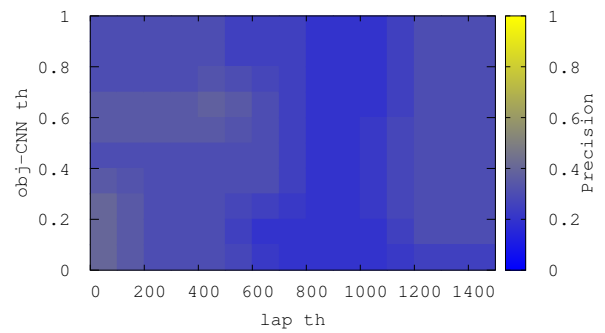


図 3.14 キーワード「curved beak」の適合率 ($\alpha = 0.4$)

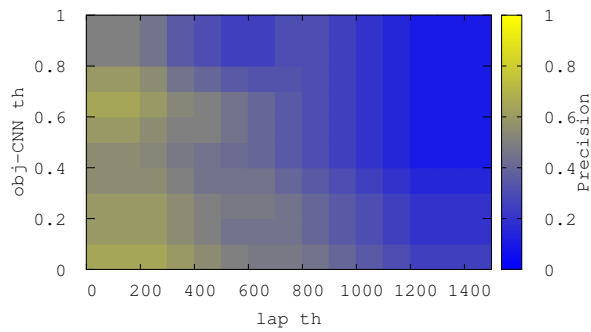


図 3.15 キーワード「red wings」の適合率 ($\alpha = 1.0$)

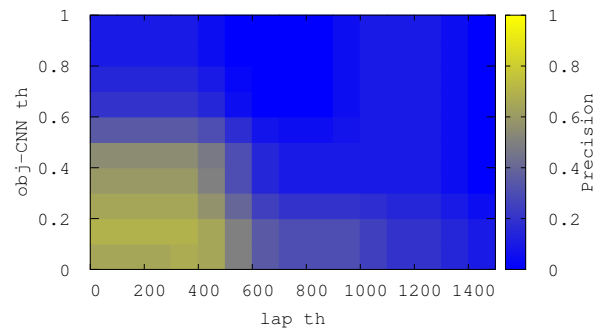


図 3.16 キーワード「white belly」の適合率 ($\alpha = 0.0$)

3.2.2 考察

いずれのキーワードの適合率の結果を見ても「lap th」の値が低い場合に適合率が高くなる傾向が見受けられる。また、「obj-CNN th」についても、その値が0.6付近よりも小さい場合に適合率が高くなる傾向が見受けられ、それ以上になると低く推移していることがわかる。

さらに、キーワード「blue wings」の図 3.13 と「white belly」の図 3.16 から、適合率が比

較的高い分布を示しているのは「lap th」「obj-CNN th」の値がお互いに小さい部分に集中しており、処理1と処理2によって除外する画像の制限を緩める方が優先表示できていることがわかる。これは、3.1節の線形結合パラメータ α の依存性についての実験結果で、正解の画像の総合評価が処理1と処理2によって0.0と付けられ、優先表示できなかった状況に陥っていた結果になったことや図3.11のキーワード5種の平均適合率からも裏付けられる。つまり、処理3の「req-CNN」及び「web-CNN」に画像を与えることで図3.1や図3.2のような画像の評価を低く付ける、もしくは正解の画像に対して較的高い評価を付けることができていることがわかり、処理1や処理2で除外を行うよりも、直接処理3を行う方が優先表示ができていることがわかった。

3.3 評価実験: 処理3のみを用いた線形結合パラメータの依存性

3.2.1節の結果から、「curved beak」を除いたすべてのキーワードについて、図の左下部分に適合率が高くなる傾向が見られる。このことから、「lap th」と「obj-CNN th」の値が低い場合に適合率が高まり、処理1と処理2によって正解の画像が優先表示から除外されていることが示される。そこで、本節では処理1と処理2の評価値を考慮しない場合の総合評価（数式(3.1)）を利用した優先表示を行い、再現率における線形結合パラメータの依存性について再度実験を行うことで、結果が向上するか実験する。

$$\text{eval}(img) = (1 - \alpha) \text{req-CNN}(img) + \alpha \text{web-CNN}(img) \quad (3.1)$$
$$(0.0 \leq \alpha \leq 1.0)$$

3.3.1 実験結果

「black body」の再現率については処理1及び処理2を取り除くことで、僅かに α に優位性が見られる。図3.6と比較すると図3.18の方が α がすべての状態で再現率が1.0に到達するのが早く、さらに、 α が0.2~0.4の間で再現率が上がり始める傾向が見られる。このことから、「black body」については処理1と処理2によって一部の正解の画像の優先表示を妨げていることがわかり、3.1節で見られなかった α による優位性も観測できる。

「blue wings」の再現率については処理1及び処理2を取り除くと、図3.7と比較して図3.19は α がすべての状態で再現率が1.0に到達するのが遅くなるという結果になった。これは、「blue wings」の正解画像のセットに処理1と処理2によって評価値を0.0と付けられたものが存在せず正解以外の画像が正しく取り除かれたため、処理1と処理2を行った方が優先表示の精度が上がったと考えられる。また、図3.7と図3.19が似たような色の分布を観測できる点からも、処理1と処理2によって優先表示の性能が向上していたことが観測できる。

「curved beak」の再現率については図3.20を見ると、処理1及び処理2を取り除くことで α が0.8~1.0の時、僅かに早く再現率が上がり始めることが確認できる。また、0.0~0.1の値の場合で図3.8のすべての状態よりも早く再現率が1.0に到達することが確認できる。このこ

とから、一部 α に優位性が見られ、僅かではあるものの処理 1 と処理 2 を取り除いた場合の方が、優先表示の精度が上がる場合が存在することが観測できる。

「red wings」の再現率については図 3.21 を見ると、処理 1 及び処理 2 を取り除くことで全体で見ても再現率が 1.0 に到達するのが早くなる結果になった。図 3.9 ではすべての正解を表示するまでほぼサンプルのすべてを出力する必要があったが図 3.21 から処理 1 と処理 2 を取り除くことで α が 1.0 に近いほど優先表示の精度が高く、また、 α が 0.6~0.8 で再現率が早く上がり始めることから α についても優位性が確認できる結果になった。

「white belly」の再現率については図 3.22 を見ると、処理 1 及び処理 2 を取り除くことで α が 0.6 の時に再現率が 1.0 に到達するのが早くなることが確認できる。これは図 3.10 でも確認できることから、処理 1 と処理 2 が優先表示の精度を顕著に阻害していることを確認できる結果になった。

3.3.2 考察

すべてのキーワードの処理 1 及び処理 2 を除いた結果を見ると、「blue wings」以外のキーワードの結果から、処理 1 と処理 2 を取り除き、直接処理 3 の結果を用いた総合評価（数式 (3.1)）による優先表示の方がユーザの求めている画像をより早く提供していることが確認できる。これは、3.2.2 節の考察でも述べた通り、処理 1 及び処理 2 の評価によって一部の正解の画像の総合評価値を 0.0 と付けられたため優先表示を妨げていることが原因であると考えられる。さらに、3.1 節では見られなかった線形結合パラメータ α の優位性を各キーワード毎に観測できた。また、唯一正解の画像を取り除いていなかった「blue wings」の結果から、正解の画像セットの中に少しでもぼやけていて、CNN が鳥と断定できないような正解が含まれているキーワードについては処理 3 のみを行う方がユーザの要求に合った画像が出力され、正解として用意されている画像のすべてがぼやけてなく CNN が鳥と判定できるものである場合は処理 1 と処理 2 によって精度が向上していることが確認できる。つまり、平均的には処理 1 及び処理 2 を取り除く方がユーザの要求に合った画像を多く提供できるが、処理 1 と処理 2 が完全に妨げになっていると断言できず、総合評価の算出法を改善することで優先表示の精度が向上する可能性があると考えられる。

3.4 評価実験: 2 つの重みパラメータの依存性

3.3 節の実験結果から総合評価の算出法を改善することで優先表示の精度が向上する可能性があると考えられる。数式 (2.3) では、処理 1 及び処理 2 の閾値のどちらか一方でも評価値が下回るとその評価対象の画像の総合評価は 0.0 になる。これにより、処理 3 の評価値が高い場合でも処理 1 及び処理 2 の評価値が閾値より低いと、優先表示から外れてしまうことで性能の低下を招いていると考えられる。そこで、処理 1 の評価値 $\text{lap}(img)$ の値を 0.0~1.0 の値に調整し、総合評価の算出法を数式 (3.3) のように変化させる。値の調節には sigmoid 関数を基に改変したものを利用し、通常は sigmoid 関数が『 $\frac{1}{1+e^{-x}}$ 』であるのに対し、数式 (3.2) はネ

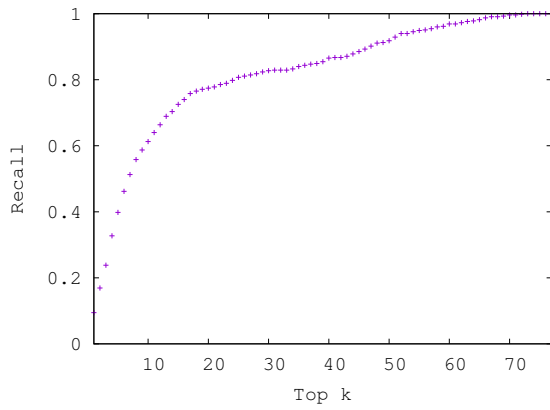


図 3.17 処理 3 のみのキーワード
5 種全体の再現率

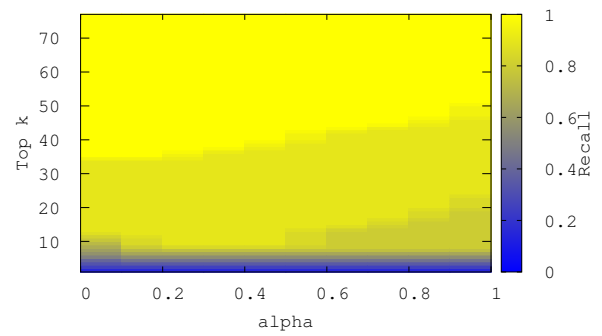


図 3.18 処理 3 のみのキーワード
「black body」の再現率

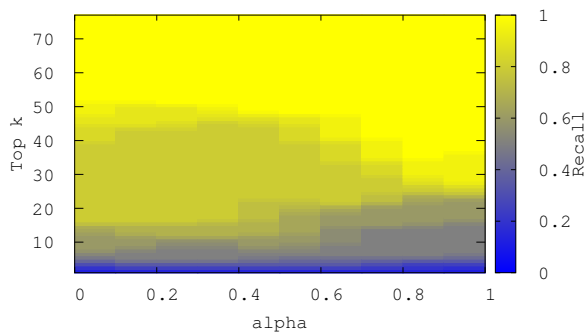


図 3.19 処理 3 のみのキーワード
「blue wings」の再現率

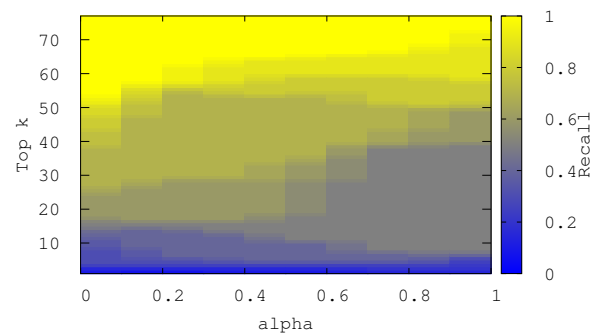


図 3.20 処理 3 のみのキーワード
「curved beak」の再現率

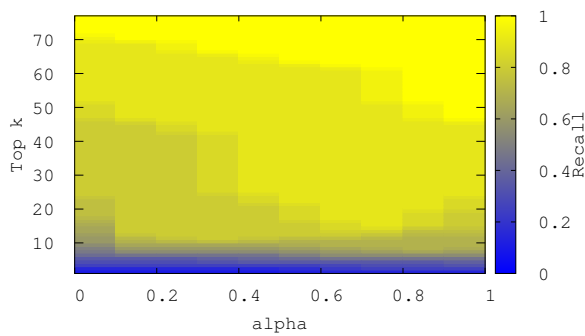


図 3.21 処理 3 のみのキーワード
「red wings」の再現率

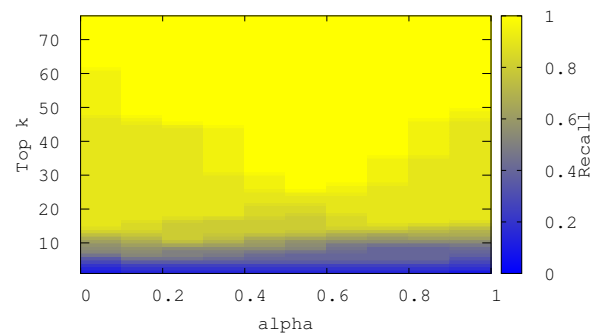


図 3.22 処理 3 のみのキーワード
「white belly」の再現率

イピア数『 e^{-x} 』の部分『 1.01^{-x} 』と変えることで、通常の sigmoid 関数よりも勾配が緩やかになるようにしている。さらに、処理 1 の評価値で基準にすべき値として 300 と設定し、処理 1 の評価値が 300 の時、数式 (3.2) が 0.5 をとるように調節する。そして、調節後の処理 1 の評価値（以降、 $\text{lap-SIG}(img)$ と呼称）とその他の評価値を数式 (3.3) のように結合し、各処理にどの程度の重みを置くかを定める重みパラメータの β と γ を設定する。これにより、重みパラメータの調節によって各処理の重要度を決定しすべての処理の結果を考慮した総合評

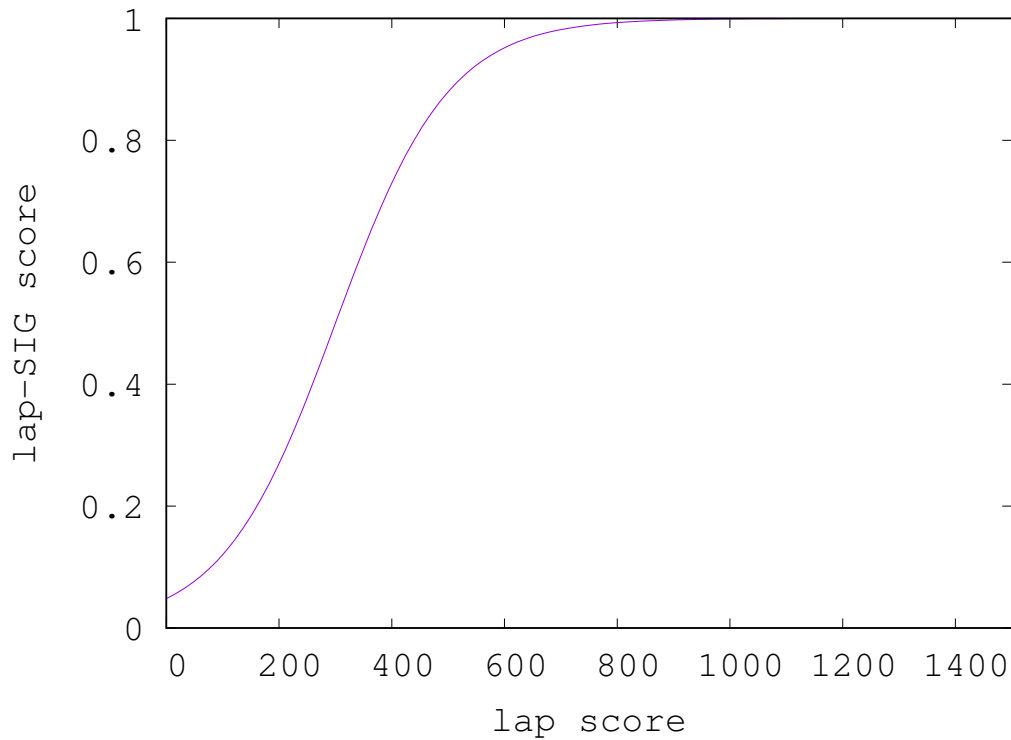


図 3.23 数式 (3.2) のグラフ

価が付けられ、数式 (2.3) のような、一部の処理によって総合評価が極端に下げられる現象を回避することができると考えられる。sigmoid 関数を改変した数式 (3.2) の具体的な軌跡を図 3.23 に示す。本節では、2つの重みパラメータを変化させた場合の適合率の変化を結果として載せる。

$$\text{lap-SIG}(img) = \frac{1}{1 + 1.01^{-(\text{lap}(img)-300)}} \quad (3.2)$$

$$\begin{aligned} \text{eval}(img) = & \text{lap-SIG}(img)^\beta \cdot \text{obj-CNN}(img)^\gamma \\ & \cdot \{(1 - \alpha) \text{req-CNN}(img) + \alpha \text{web-CNN}(img)\}^{1-\beta-\gamma} \quad (3.3) \\ & (0.0 \leq \alpha, \beta, \gamma \leq 1.0, \beta + \gamma \leq 1.0) \end{aligned}$$

3.4.1 実験結果

各キーワード毎に線形結合パラメータである α の値を 3.3 節の結果から 10 番目までの再現率の上昇速度で優れているものに固定し、10 番目までの適合率と 20 番目までの適合率を載せ考察を行う。表 3.2 は数式 (3.3) を用いて優先表示を行った際の最大適合率と、その時の他のパラメータ及び再現率、F 値である。最大適合率を出している時のパラメータが大量に存在するため表 3.2 の各値は一例である。

表 3.2 数式 (3.3) を用いた時のキーワード 5 種の精度

	k	β	γ	α	最大適合率	再現率	F 値
black body	10	0.00	0.00	0.20	0.90	0.90	0.90
	20	0.40	0.10	0.00	0.50	1.00	0.67
blue wings	10	0.50	0.30	0.30	0.80	0.80	0.80
	20	0.20	0.70	0.80	0.50	1.00	0.67
curved beak	10	0.00	0.00	0.70	0.50	0.50	0.50
	20	0.30	0.00	0.40	0.35	0.70	0.47
red wings	10	0.00	0.00	0.10	0.70	0.70	0.70
	20	0.00	0.00	0.60	0.45	0.90	0.60
white belly	10	0.00	0.00	0.20	0.70	0.70	0.70
	20	0.00	0.00	0.00	0.45	0.90	0.60

キーワード「black body」の適合率について、図 3.26 を見ると重みパラメータの β を変化させても 10 番目までの適合率にほとんど変化はなく、 γ の値が 0.0 の時と $\beta + \gamma = 0.9$ の付近の時に適合率が上昇している傾向が見られる。また、 $\beta + \gamma = 1.0$ の時、つまり、req-CNN や web-CNN を無視する時、どの場合においても急激に適合率が低下することがわかる。図 3.27 を見ると、 β が 0.6 の時に微小であるが適合率が高い傾向が見られる。このことから、10 番目までの優先表示において、処理 1 については β の重みを変化させても優先表示に影響しないことがわかり、この影響を受け始めるのは 20 番目以降の表示枚数が増えた時に効果を発揮し始めると考えられる。

キーワード「blue wings」について、図 3.28 を見ると重みパラメータ γ が 0.5~0.7 付近の時に、10 番目の適合率が上昇していることが確認できる。また β については変化させてもほとんど結果が変わっていないことが確認できる。この傾向は図 3.29 でも似たような傾向が表れるが、特に、 $\beta + \gamma = 0.9$ の付近で適合率が高い状態があることを確認できる。つまり、処理 2 については 0.6 程度の重みを設定すると適合率が上昇し、20 番目では処理 3 の重みが 0.1 の時に精度が上昇するという結果になった。

キーワード「curved beak」の適合率について、図 3.30 を見ると重みパラメータの β と、 γ の値の合計が 0.9 の時 ($\beta + \gamma = 0.9$)、つまり、処理 3 の重みが 0.1 の時に適合率が上昇していることがわかる。しかし、 β と γ が共に 0.0 の時でも適合率が上昇している傾向が見られるため、10 番目までを表示する時、処理 3 のみの評価値を用いるか、処理 3 の評価を 0.1 乗した場合の方が、他のパラメータ設定よりも優れていると考えられる。しかし、図 3.31 では、そのような傾向は見られず β が 0.4~0.8 で γ が 0.2 よりも小さい時、適合率が下がる傾向が見られた。これらのことから、処理 3 に限定して総合評価を付けるか、他の処理の評価値と併用する場合は、処理 3 の評価値をより大きく取り扱う方が精度が高くなるという結果になった。

キーワード「red wings」について、図 3.32 を見ると β と γ が共に上昇するにつれ、適合率が下がっていく傾向が見られる。この傾向は、図 3.33 でも同様に見られ、このことから、キーワード「red wings」については、処理 1 と処理 2 を省略し直接処理 3 の評価値を総合評価として優先表示を行う方が精度が高いという結果になった。

キーワード「white belly」について、図 3.34 を見ると全体の傾向としては、キーワード「red wings」と同様に β と γ の値が大きくなるにつれて適合率が下がっていく傾向が見られる。この傾向は図 3.35 も同じ傾向を示している。ただし、図 3.34 では、 β が 0.2~0.5 付近で適合率が上昇し、 β と γ が共に 0.0 の時よりも高いことから、処理 3 のみを総合評価とするよりも、処理 1 の評価値と併用した総合評価を行った方が結果が向上することがわかる。また、図 3.35 については γ が 0.2 の時、僅かに適合率が上昇していることが確認できる。つまり 20 番目までの結果においては処理 2 の評価を微小に用いる方が結果が向上していることがわかる。

3.4.2 考察

キーワード「black body」「blue wings」「curved beak」については、各図の形を観察すると三角形の外側に高い適合率を示す傾向が多く、このことから、重みパラメータの β と γ の合計が 0.9 ($\beta + \gamma = 0.9$) 付近の時、つまり処理 3 の評価値を 0.1 乗し、値を大きく利用すると適合率が高くなることがわかった。この傾向は、キーワード「white belly」でも同様であり、さらに「white belly」については、10 番目まで表示する場合は β が 0.2~0.5 までの値が適切であり、20 番目まで表示する場合は γ が 0.2 の時に、他のパラメータ設定の時よりも適合率が高い傾向を示した。以上より、ほとんどのキーワードにおいては処理 1 と処理 2 の評価値を用いて総合評価を行い、同時に処理 3 の評価値を通常の評価値よりも大きくすることが、バランスよく評価できる方法であると考えられる。ただし、キーワード「red wings」はその限りではなく、処理 1 と処理 2 を含めると適合率が下がる傾向のみが見られ、このことから、CNN が極端に特徴を掴みやすいキーワードであれば処理 1 と処理 2 を含めるとノイズになってしまう性能を妨げる原因になると考えられる。

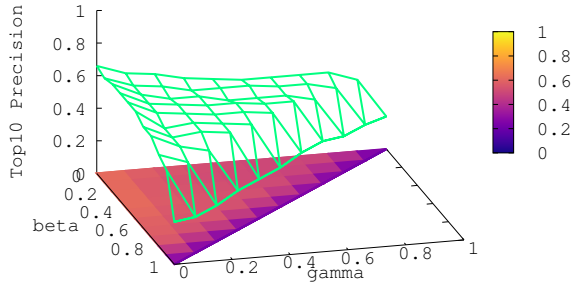


図 3.24 キーワード 5 種全体の 10 番目までの適合率 (α はキーワード毎の固定値)

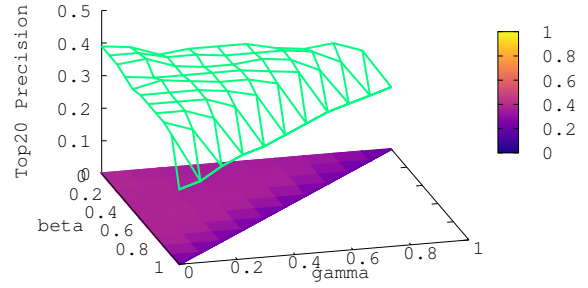


図 3.25 キーワード 5 種全体の 20 番目までの適合率 (α はキーワード毎の固定値)

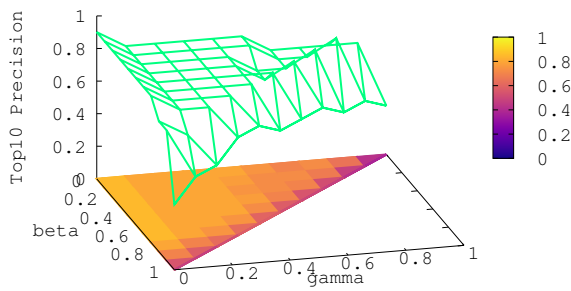


図 3.26 キーワード「black body」の 10 番目までの適合率 ($\alpha = 0.3$)

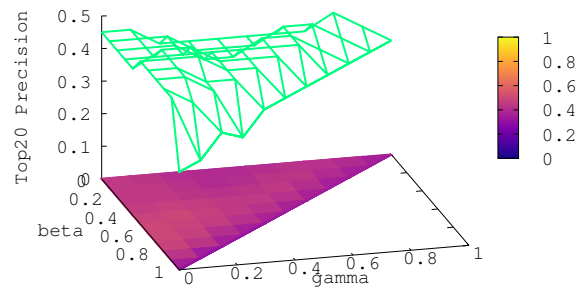


図 3.27 キーワード「black body」の 20 番目までの適合率 ($\alpha = 0.3$)

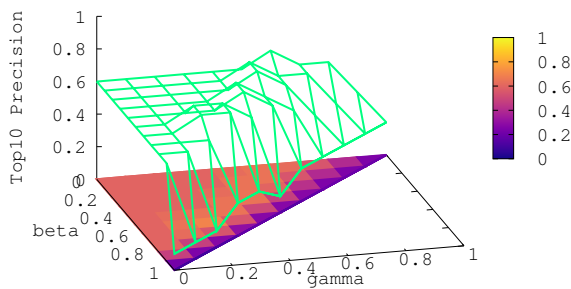


図 3.28 キーワード「blue wings」の 10 番目までの適合率 ($\alpha = 0.2$)

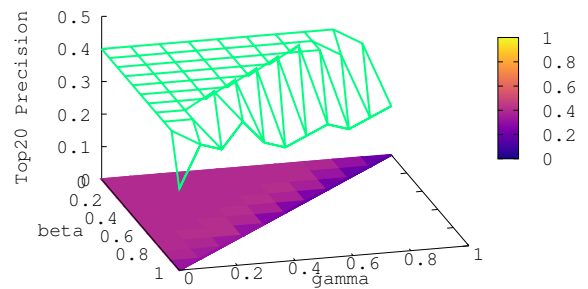


図 3.29 キーワード「blue wings」の 20 番目までの適合率 ($\alpha = 0.2$)

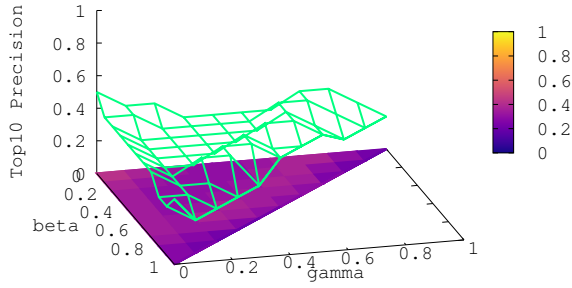


図 3.30 キーワード「curved beak」の 10 番目までの適合率 ($\alpha = 0.8$)

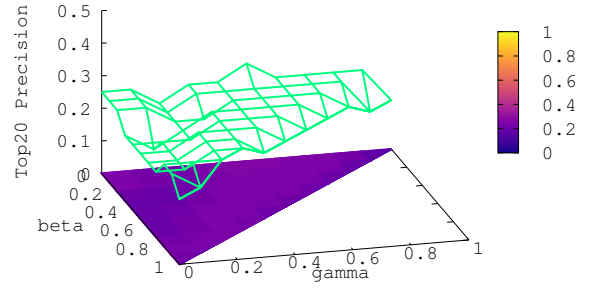


図 3.31 キーワード「curved beak」の 20 番目までの適合率 ($\alpha = 0.8$)

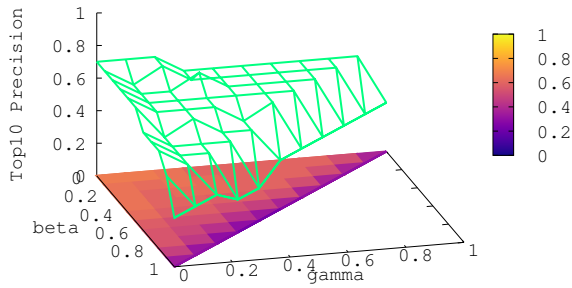


図 3.32 キーワード「red wings」の 10 番目までの適合率 ($\alpha = 0.8$)

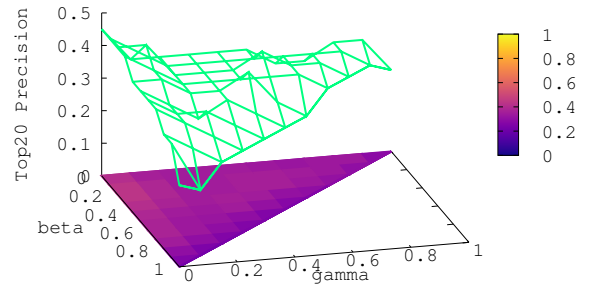


図 3.33 キーワード「red wings」の 20 番目までの適合率 ($\alpha = 0.8$)

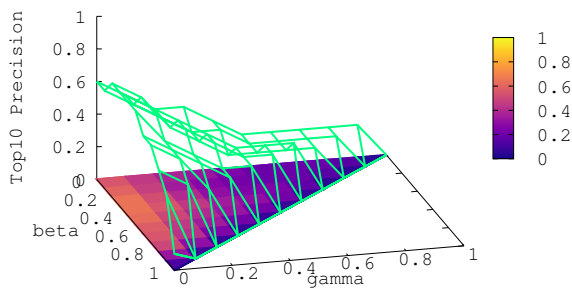


図 3.34 キーワード「white belly」の 10 番目までの適合率 ($\alpha = 0.5$)

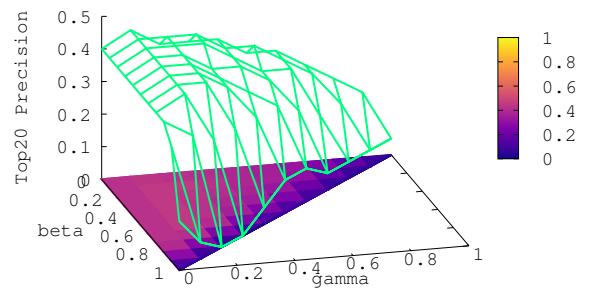


図 3.35 キーワード「white belly」の 20 番目までの適合率 ($\alpha = 0.5$)

第 4 章

結論

本章では本研究の提案手法及び評価実験の結果をまとめ、社会的及び技術的貢献と今後の研究課題についても記述する。

4.1 まとめ

本研究では、GAN から想定される生成画像を処理 1 のぼやけ具合の評価、処理 2 の生成対象物の評価、処理 3 のユーザリクエストの反映度の評価 (2 種類) によって 4 種類の評価値を算出し、処理 4 で 4 種類の評価値を用いた総合評価を行うことで、ユーザのリクエストに合った画像を優先的に表示できるようなフィルタリング手法を提案した。次に、処理 4 で「req-CNN」「web-CNN」の評価値を線形結合するパラメータ α の依存性について数式 (2.3) による実験を行い、10 番目までの適合率を出力させたところ、 α の大きな依存性は認められず、処理 1 と処理 2 の閾値設定に依って適合率が下がっている可能性があるという結果が得られた。そこで、処理 1 と処理 2 の最適な閾値を求めるために、閾値への依存性を実験した結果、図 3.12 と図 3.15 のみ「obj-CNN th」の値が 0.6~0.8 付近で適合率が上昇していることを確認できるが、その他は処理 1 及び処理 2 の閾値に依って正解の画像を除外してしまう場合が多く、数式 (2.3) の評価方法によって、10 番目までの適合率が α の値を変えても変化が微小になるといった結果になり、優先順位を付けても正解を拾い上げることができなくなってしまうといった結果になった。この結果から処理 1 及び処理 2 を除き直接処理 3 を総合評価とした時 (数式 (3.1)) の再現率を出力させた場合、数式 (2.3) を利用した時と比較すると、平均的にユーザの要求を満たす画像を優先的に表示でき、さらにキーワード毎に線形結合パラメータ α の優位的な値を確認できる結果になった。しかし、キーワード「blue wings」のような正解とする画像に少しでもぼやけが含まれるものや、obj-CNN が鳥と判断できないようなものがある場合は、処理 1 及び処理 2 を含めている方が精度が高いことが確認でき、総合評価値の算出方法を改良することで性能が向上する可能性があることがわかった。そこで、処理 1 及び処理 2 の閾値に依って処理 3 の評価に関わらず優先表示から正解の画像を除外してしまう問題を解決するために、すべての評価値を考慮することができ、各処理の重み (信頼性) を調節して総合評価値を付けるような数式 (3.3) を再提案し、重みパラメータ (β, γ) の依存性について

実験を行った。その結果、重みパラメータの β と γ の合計が 0.9 ($\beta + \gamma = 0.9$) 付近の時、つまり処理 3 の評価値を 0.1 乗し、値を大きく利用すると一部のキーワードでは適合率が高くなることがわかった。特に、キーワード「white belly」では、10 番目まで表示する場合は β が 0.2~0.5 までの値が適切であり、20 番目まで表示する場合は γ が 0.2 の時に、他のパラメータ設定の時よりも適合率が高い傾向を示すことから、処理 1 と処理 2 の評価値を用いて総合評価を行い、同時に処理 3 の評価値を通常の評価値よりも大きくすることで、性能が向上すると考えられるという結果になった。また、CNN が極端に特徴を掴みやすいキーワードであればその限りではなく、処理 1 と処理 2 を含めるとノイズになってしまい性能を妨げる原因になると考えられる結果も得られた。

筆者の個人的見解としては、様々なパラメータを調節した際の 10 番目までの適合率の最大値が 0.8 以上、つまり、正解の画像を 10 枚用意した場合には、10 枚中 8 枚がユーザの要求を満たした画像であるか、20 番目までの適合率の最大値が 0.5、つまり、20 番目までにすべての正解の画像が表示されていれば実用的に使用しても問題がないと考えている。実験のために仮定したキーワード 5 種の内この基準を満たしているキーワードは「black body」「blue wings」の 2 種類のみであり、実用化を図るならば最低でも 4 種類のキーワードが基準を満たしている状況が望ましいと考えている。特に本研究で複数利用されている CNN は実際に構築しなければ、結果がどのようなになるか予想することは極めて難しいため、率直に言えば実験の段階で仮定されるキーワードについては、すべて基準を満たしている状況でなければ、実用化した時に多くの欠陥が生じると考えているため、まだまだ研究を進める必要がある。

4.2 社会的貢献及び技術的貢献について

本研究の社会的貢献として、GAN の生成画像に対してその画像を評価、分析する研究は存在せず GAN の生成画像を生成順に表示するよりも提案した現状のフィルタリングを行って優先表示の方が効率的にユーザが求めている画像の提供を期待できる。そして、将来的な社会的貢献としてはしばしば GAN の学習が不安定になる問題の間接的な解決策として、フィルタリング性能が高ければ精度の低い GAN から大量の画像を生成させて評価を行うことで、ユーザが求める画像を提供できる可能性が向上し、求めている画像が出たとしても、ユーザが感覚的に良いと思うことができるような画像を表示できる割合が向上する。

本研究の技術的貢献としては、多くの研究者が GAN 本体の改善について研究するのに対し、本研究では GAN の生成画像をユーザに提供するまでのインターフェースを構築する際に貢献できるような仕組みを意識して提案している点である。具体的に、ラプラシアンフィルタを通した後の画像のエッジについて分散を取得し評価する方法と、GAN の学習したオブジェクトについての評価と、ユーザの要求を反映している度合いについて評価を付けて、これらの評価を複合的に計算する総合評価式を用いることで、GAN の生成画像を独自の観点から評価できるようになっている。そして、将来的な技術的貢献としては、より厳密なぼやけ具合の評価方法並びに、複数の CNN の安定した構築法、加えて分析方法や各パラメータの最適化が実現できれば、GAN の生成画像を対象とした評価だけではなく、似たような観点から画像全般

の数値的評価方法として利用でき、抽象的であった画像の質を評価する一手法としての利用が期待できる。

4.3 今後の課題

今後の課題として、処理1のぼやけ具合を評価する際の評価値 $\text{lap}(img)$ を $0.0 \sim 1.0$ に変換する数式 (3.2) の算出時に基準として 300 という値を用いているが、この値はすべての処理1の画像の評価値を見て、筆者が他のコンテンツで利用できるようなクオリティとしてふさわしいと判断した値であるためこの基準値を調節した場合、結果にどのような影響を及ぼすのかは未知数であり、加えて、sigmoid関数以外の関数を用いて調節することでも結果が変化する可能性があるため実験の必要があると考えられる。また、根本的に処理1の評価の算出法を変えることでより厳密な評価が可能になることも考えられ、処理1についてはまだまだ改善の余地がある。

また、処理2については、本研究では鳥のカテゴリを含むデータセットを用いたが、「cifar10」の画像サイズが 32×32 ピクセルと小さいことで、誤認評価を発生させる原因になっていることも否定できない。さらに、鳥であったにもかかわらず、他のオブジェクト（人の手など）が写っている画像や、鳥の一部（頭や脚のみ）のみが写っている画像などを鳥ではないと誤認し、評価値が下げられている可能性も考えられる。これについては、画像サイズがより大きいデータを利用して学習させることや、様々な写り方をした鳥の画像を用意する必要があると考えられる。

さらに、実験用のデータサンプルについては、キーワード毎の正解の画像を10枚に揃え、総合評価の値でランキングして10番目まで（もしくは20番目まで）の適合率を求めてきたが、実際にGANの生成画像を直接フィルタリングする際は、フィルタリング処理にかかる時間の短縮や、何枚まで画像を表示すればユーザが一度に得る情報量が多過ぎず、ユーザの要求に合った画像を妥当な数表示できるかといったインターフェース的な観点の課題も残されている。

謝辞

本研究に際して、様々なご指導を頂きました服部峻助教並びに荒澤孔明先輩に厚く御礼申し上げます。また、日常の議論を通じて多くの知識や示唆を頂いた服部研究室の皆様にも深く感謝の意を表します。

参考文献

- [1] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications,” *Journal of L^AT_EX Class Files*, vol.14, no.8, pp.1–28 (2015).
<https://arxiv.org/pdf/2001.06937.pdf>
- [2] Jorge Agnese, Jonathan Herrera, Haicheng Tao, Xingquan Zhu, “A Survey and Taxonomy of Adversarial Neural Networks for Text-to-Image Synthesis” *Computer Vision and Pattern Recognition (CVPR)*, arXiv.org, pp.1–26 (2019).
- [3] OpenCV ライブラリ, <https://opencv.org/> (2020).
- [4] Caltech-UCSD Birds-200-2011,
<http://www.vision.caltech.edu/visipedia/CUB-200-2011.html> (2020).
- [5] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, Xiaodong He, “AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks,” *Computer Vision and Pattern Recognition (CVPR)*, arXiv.org, pp.1–9 (2018).
- [6] Alex Krizhevsky, “The CIFAR-10 dataset,”
<https://www.cs.toronto.edu/~kriz/cifar.html> (2020).