平成28年度茨城大学工学部情報工学科
卒業研究論文

画像キャプション生成における複数形表現の統一

平成29年2月10日

茨城大学　工学部情報工学科
13T4076F　西　友佑

指導教員：新納浩幸教授

平成 28 年度茨城大学工学部情報工学科卒業研究

画像キャプション生成における複数形表現の統一

氏名：13T4076F 西　友佑
指導教員：新納 浩幸 教授

本論文では，画像キャプション生成における生成文の品質向上を目的とし，その実現のために，訓練データ中の複数形表現を統一することを試みる．具体的には， two dogs や three cars と言った基数を用いた複数形表現を，基数を用いないより単純な表現への統一を行う．これにより，生成文における基数を使用していた部分の誤りが訂正され，品質の向上が期待できる．

入力された画像のキャプション（説明文）を生成する「画像キャプション生成」の研究は従来より活発に行われてきた．この画像キャプション生成に関する問題の一つとして，生成したキャプションの定量的な評価が難しい点が挙げられる．これは，正しいキャプションが一つに定まらないために起こる問題である．例えば，表現の言い換えや説明の詳しさによって，正しい説明文は無数に存在しうる．そして，その文章が正しいかどうかは主観的にしか判断できない．本論文では，文章の詳しさという点に着目し，生成されるキャプションをより単純にすることによって誤りを減らし，生成文の品質を向上させる．具体的な手法としては，基数などを用いた複数形表現をより単純な複数形表現へ統一する．これによって，生成されるキャプションの誤りを減らすことを目指している．

先に述べたような誤りの解消は，生成された後のキャプションを直接修正することでも解決可能に見える．しかし，深層学習を用いた画像キャプション生成では，深層学習の内部を詳しく検証することが難しいためにどの部分で誤りが発生しているかを特定することが困難である．また，時系列的な影響を受ける文章生成において，生成された後の一部分を変更しただけでは，文中の変更した箇所以降の部分において更なる誤りが発生する可能性がある．そのため，本研究では学習に使用する訓練データにおける複数形表現を予め書き換えておき，書き換えたデータを用いて学習を行う．

実験では，訓練データにMSCOCOのデータセットを用い，テストデータにはネットなどから収集した100枚の画像データを使用した．そして，書き換えられた訓練データで学習したシステムによるキャプションと，書き換え以前の訓練データで学習したシステムによるキャプションとを比較し，主観的に評価を行った．結果，キャプションの大きく変化した画像数は100枚中39枚となり，そのうち改悪された画像数が14枚であるのに対し，25枚の画像においてキャプションの改善が認められた．

目次

論文要旨 i

第1章 序論 1
1.1 概要 1
1.2 構成 2

第2章 画像キャプション生成 3
2.1 画像キャプション生成とは 3
2.2 Python 3
2.3 Chainer 4
2.4 ニューラルネットワーク 4
2.5 CNN 6
2.6 RNN 7

第3章 複数形表現の統一 10
3.1 複数形表現の問題 10
3.2 TreeTagger 10

第4章 実験 12
4.1 実験設定 12
4.2 MSCOCOデータセット 12
4.3 結果 13

第5章 考察 17

第6章 結論 18

謝辞 19

参考文献 20

付録 21

第1章

序論

1.1 概要

本論文では画像キャプション生成における生成文の品質向上を目的とし，その実現のために訓練データ中の複数形表現を統一する．今回使用した画像キャプション生成のモデルは，畳み込みニューラルネットワークと再帰的ニューラルネットワークを組み合わせた深層学習モデルである．この深層学習モデルを学習させる際に使用する訓練データに改良を加えることで最終的な生成結果の品質を向上させることが目的である．改良を加える点は，two dogs や three cars といった基数を用いた複数形表現についてである．これらの表現を基数を用いず，より単純な表現へと統一することで，訓練データを改良し，改良された訓練データを用いて学習を行うことにより，生成文の品質の向上が期待できる．

画像キャプション生成についての研究は従来より活発に行われてきた[1]．これは，入力された画像のキャプション（説明文）を生成する研究である．画像キャプション生成は多くの問題を抱えており，そのうちの一つに生成したキャプションを定量的に評価することが難しいという点が挙げられる．すでに幾つかの研究で定量的に評価することの出来る数式などが提案されており，BLUE評価値[6] を用いる方法などが有名であるが，これらの方法には様々な問題があり，決定的な評価方法はまだ存在しない．また，画像に対する正しい説明文が一意に決定しないという点も定量的な評価を難しくする要因である．例えば，犬が1匹写っている画像に対して，「犬がいる」と説明した文章も「茶色の犬が芝生の上に寝転んでいる」と説明した文章も正しい説明文となり得るのである．また，複数の牛が写っている画像などでは，「いくらかの牛がいる」や「牛の群れがいる」,「牛が13匹いる」など，複数形の表現を変えただけでも様々な説明文が考えられる．本論文では，この点に着目し，生成されるキャプションをより単純なものにすることによって誤りを減らし，生成文の品質を向上させることを考えた．具体的な方法としては，基数などを用いた複数形表現を統一，単純化することで生成文の誤りを減らすことを目指した．

先に述べたような生成されたキャプションの誤りの解消は，生成された後のキャプションを直接修正することによっても実現可能に思える．しかし，深層学習を用いた画像キャプション生成において，深層学習の内部を詳しく検証することは難しく，どの部分で誤りが発生したのかを特定するのは困難である．また，時系列的な影響（次の語が前の語に左右される）を受ける文章生成において，生成語の文章の一部分を変更しただけでは文中の他の箇所において誤りが発生する可能性がある．そのため，本研究では学習に使用する訓練データを予め書き換えておき，そのデータを用いて学習を行う．

実験では，訓練データにMSCOCOのデータセットを使用し，テストデータにはネットなどから収集した100枚の画像データを使用した．そして，書き換えられた訓練データで学習したシステムによるキャプションと，書き換え前のデータで学習したシステムによるキャプションとを比較し，主観的に評価付けをした．結果，キャプションが大きく変化した画像数は100枚中39枚であり，そのうち改悪された画像数が14枚，改

善された画像が25枚となった．

本論文は言語処理学会第23回年次大会(NLP2017)にて発表予定である[11]．

1.2　構成

本論文では，画像キャプション生成における生成文の品質向上のため，訓練データの改良を行う．2章で，画像キャプション生成に用いられている技術について説明を行う．3章では，複数形表現の統一について，具体的な手法を説明する．

第2章

画像キャプション生成

2.1 画像キャプション生成とは

画像キャプション生成とは，入力された画像からその画像についてのキャプションを生成する研究である．画像キャプション生成を含む画像認識という分野では，多くの研究が行われてきた．最初は，画像に何が写っているのかを自動で認識するための研究があり，知識ベースやモデルベースと言った手法が提案され，その他様々な手法で画像に対する認識の研究が行われてきた．そうして，徐々に物体に写っている物体の認識精度は高まっていったが，テーブル，ボール，車といった単語がわかるだけで，それらの関係性（テーブルの下にボールがあるのか，テーブルの上にボールがあるのか）までは認識することができなかった．そこで，画像をより深く理解するために，画像キャプション生成という研究が行われるようになっていった．論文[1]では初めて画像のみの入力でキャプション生成を行っている．しかし，実際に文章自体を１から生成しているのではなく，データベース内の最適な1文を利用するという形である．画像キャプション生成のアプローチとしては，大きく２パターンあり，それは既存の文章を利用する手法と新規で文章を生成する手法である．既存の文章を利用する手法は，ニュース記事についている画像について，画像を説明している適当な文章を利用するといった手法であったり，入力である新規の画像の特徴量と似た特徴量を持つ画像についているキャプションを利用する手法などがあげられる．新規で文章を生成する手法としては，予め主語述語などというテンプレートを作っておき，単語を当てはめることで生成する手法や，完全に1から全ての文章を生成する手法などがある．ただし，どちらの手法を用いる場合においても，基となるデータセットにない組み合わせについては表現することが出来ないという大きな問題はついて回る．

そして，近年キャプションを新規で生成する手法として，深層学習を用いたモデルの提案[10]が発表された．これは，画像の物体認識などの分野で成果を挙げていたCNNと呼ばれるニューラルネットワークと，翻訳などに使用されていたRNNと呼ばれるニューラルネットワークを組み合わせることで，CNNから得られる画像の特徴量を入力としてRNNを使って文章を生成するというモデルである．このモデルでは，ニューラルネットワークの学習を行っておけば，画像の入力のみで，学習した単語を組み合わせた新規のキャプションを生成することが出来るモデルである．今回の実験では，先に述べた論文[10]にて提案されている，畳み込みニューラルネットワーク（CNN）と再帰的ニューラルネットワーク（RNN）を組み合わせた深層学習モデルを用いて画像のキャプション生成を行っている．

2.2 Python

Pythonとは，プログラミング言語の1種である．Pythonは言語処理の分野で多く用いられているプログラミング言語で，機械学習やその他様々なパッケージを利用可能な，拡張性にとんだ言語である．また，Cなどの言語に比べて記述が簡単であるという利点もある．今回は，言語処理の分野で多く使われているこ

とに加えて，後述するChainerと呼ばれる深層学習用のライブラリであるChainerがPythonに対応していたため，この言語を使用しての実装を行った．

実験に使用したPythonのバージョンは2.7.0であり，Anaconda2.4.0を使ってインストールを行った．Anacondaとは，Pythonの配布形態の1つで，numpyやscipyといったライブラリを一括でインストールすることが出来るものである．Python2.x，3.x の両バージョンに対応しており，LinuxやMac，Winなど様々なOSにも対応している．様々なOSにまたがって開発をする場合や，開発の環境を一定にしたい場合などに便利である．

2.3　Chainer

Chainerとは，ニューラルネットワークを記述するためのライブラリである．Preferred Networksが開発したこのライブラリは，他のニューラルネットワーク用ライブラリよりも記述の仕方や動作についての習得が容易であり，実際に動かすまでの時間が短いのが特徴である．また，導入に際しても比較的簡単であるため，このライブラリを使用することとした．以下にChainerの公式ページのURLを記す．

`http://chainer.org/`

実験では，バージョン1.17.0を使用した．論文執筆時での最新バージョンは1.20.0.1であった．

2.4　ニューラルネットワーク

ニューラルネットワークとは線形識別モデルの 1 種で，動物の神経組織（ニューロン）を模して作られたモデルである．ニューラルネットワークの動作を説明するために以下の図2.1を用いる．

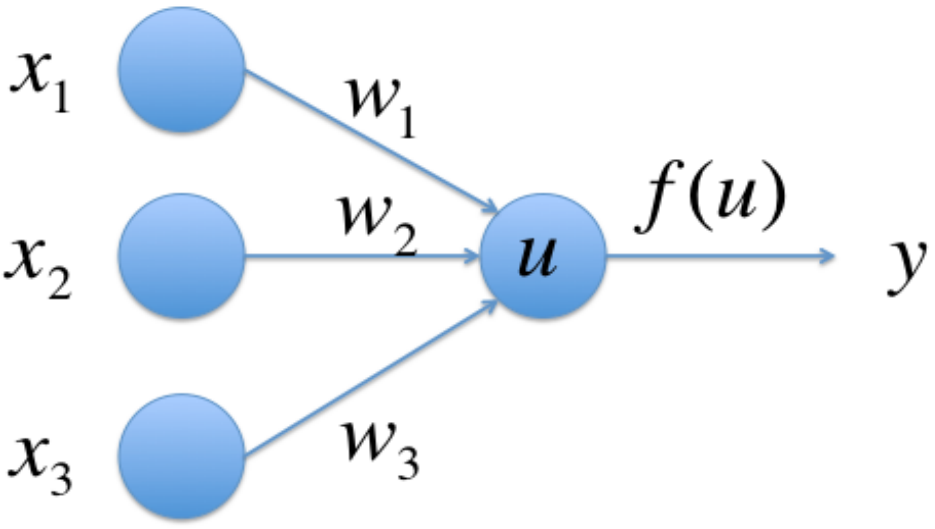

図2.1　単純パーセプトロン

この図は，ニューラルネットワークの中でもシンプルな単純パーセプトロンと呼ばれるモデルである．入力をx_1，x_2，x_3，出力を y とする単純パーセプトロンである．u は以下の式2.1で表される．

$$u = x_1 w_1 + x_2 w_2 + x_3 w_3 \tag{2.1}$$

この時，$f(u)$は活性化関数と呼ばれる関数である．これは，入力の和uから何かしらの出力を出す関数で，後で説明する誤差逆伝播法を用いるために微分可能な関数を設定する．

ここで，x_1からx_3の特徴量からそのデータがスパムメールであるかを判定することを考える．この時，スパムメールであるかどうかの条件は以下の通りとする．

$$\begin{cases} u > 0 : (\text{スパムメールである}) \\ u < 0 : (\text{スパムメールでない}) \end{cases}$$

ここでuの数式を見ると，w_1からw_3がそれぞれの特徴量の影響力を調整していると考えられる．例えば，スパムメールである場合にx_1が深く関係しているのであれば，それに合わせてx_1の影響がuに出やすいようにw_1を大きくする．逆に関係がないようであれば影響を小さくするためにw_1で制限をかけるというように操作を行う．このw_1からw_3を重みとよび，これらをまとめたWを重みベクトルと呼ぶ．この重みベクトルを調整することで学習を行う．

学習には損失関数と呼ばれるものを用いる．これは，出力された結果がどれだけ教師データとかけ離れているのかを表す関数で，基本的に誤差が大きければ大きいほど大きな値を返すものである．そのため，損失関数が返す値が最小になれば学習が終了したものと考えられる．損失関数には0-1損失，絶対誤差損失，二乗誤差損失，対数損失など様々な種類があり，問題にあった損失関数を選択することでより良い結果が得られる．また，学習には学習率と呼ばれるものが用いられる．これは1度の修正でどの程度重みの修正を行うかという割合である．大きすぎる場合は最適解を見逃してしまう場合があるので，ある程度小さい値を用いる．ただし，あまりにも小さすぎると1度の修正で変化する量が小さくなり，計算回数が増えてしまうため，注意が必要である．

活性化関数とはパーセプトロンの出力を決定する関数である．上で説明したような例ではあまり効果が無いが，多層パーセプトロンの場合には微分可能な連続した関数が用いられる．活性化関数として使用されているものには，ステップ関数，シグモイド関数，ReLU，双曲線正接関数などが挙げられる．これも問題によって適切な活性化関数を用いることでより良い結果が得られる．例では二値分類を行っているので，uから教師データのラベルに合わせた数値を出力するステップ関数を用いるのが良いと考えられる（教師データにおいてスパムメールであれば1，スパムメールでなければ-1のラベルがついている場合は，$u > 0$で1を出力，$u < 0$で-1を出力する）．この活性化関数を通して出てきたものが単純パーセプトロンの出力yとなる．

単純パーセプトロンを組み合わせて層を増やしたものが多層パーセプトロンと呼ばれるニューラルネットワークである．これは，単純パーセプトロンからの出力を別の単純パーセプトロンへと入力することで多層としているニューラルネットワークである．単純パーセプトロンでは線形識別可能な問題しか解けなかったが，多層パーセプトロンを用いることで，線形識別不可能な問題を解決することが出来るようになった．これにより，ニューラルネットワークの活躍の幅が増えることとなった．

誤差逆伝播法について説明を行う．これは，多層パーセプトロンのように複数の層を持つニューラルネットワークに用いられる，重み更新（学習）のための手法である．ここでは，誤差逆伝播法の概要を説明するために，ニューラルネットワークを活性化関数$f(x)$とし，層1つでの計算を$g(x)$とする．この場合，1層のニューラルネットワークは入力xが層$g(x)$を通り，活性化関数$f(x)$で出力が決定されるので$f(g(x))$という合成関数の形で表せる．逆に，出力から入力を導出するには微分をしていけば良いことになる．具体的には，活性化関数の出力をfについて偏微分して$\frac{\partial f}{\partial g}$，さらに$g$について偏微分して$\frac{\partial f}{\partial g}\frac{\partial g}{\partial x}$ と表せる．これが誤差逆伝播法の基礎となる．さらに，これを3層のNNとすると，$f(g(g(x)))$となる．この時，$g(x) = Wx$，$\frac{\partial g}{\partial x} = W$とすると，誤差逆伝播法を適用した場合以下の式2.2のとおりとなる．

$$\frac{f(g(g(x)))}{\partial x} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial g}\frac{\partial g}{\partial g}\frac{\partial g}{\partial x} = \frac{\partial f}{\partial g}WWW = \frac{\partial f}{\partial g}W^3 \tag{2.2}$$

この式より層が増えるに連れて重みベクトルWの乗数が増えていくことがわかる．このことから，層数が増えていった場合に，重みベクトルが1以下である場合には重みベクトルが極端に小さくなってしまう問題が発生し（勾配消失問題），重みベクトルが1以上であった場合には極端に巨大になってしまうという問題が発生した（勾配爆発問題）．そのため，深い層（多層）のニューラルネットワークは実現が難しかった．この問題について解決方法が模索されていたが，ReLUと呼ばれる活性化関数を用いることで勾配消失問題については解決することができている．

2.5　CNN

CNNとは畳み込みニューラルネットワーク（Convolutional Neural Network）の略称である．このニューラルネットワークは主に画像の解析に用いられる．

CNNは名前の通り，入力された情報を畳み込み，つまり圧縮して学習を行う．CNNという技術は生物工学における，視覚が情報をどのようにして処理しているのかという観点から開発されている．脳は，物体を見る際にまず光を網膜に投影する．それに反応した細胞が視神経に情報を伝えるのだが，その際に光を受け取る細胞と視神経とは疎に結合している．つまり，複数の細胞の情報を視神経でまとめているのである．これが畳み込みの発想のもととなっている．具体的には，50×50のサイズの白黒の画像があった場合に，5×5の範囲ごとにフィルタを掛けて，フィルタ内において，白である領域が過半数なら白を出力，逆に過半数が黒ならば黒を出力すると言った作業を行う．この作業を，フィルタを少しずつずらしながら行うことによって，元の画像よりも小さな白黒画像のデータが出力される．これが畳み込みである．この時，白黒の画像が光を受けた細胞，フィルタを通して出てきた出力が視神経が受け取る情報と対応付けることが出来る．この畳込みを行うことによって，画像認識において大きな問題であった，画像の歪みや位置のズレなどによる誤差を吸収することが出来る．こうして出来たものを画像の特徴量と言い，この特徴量を利用することで，様々な応用が可能になる．

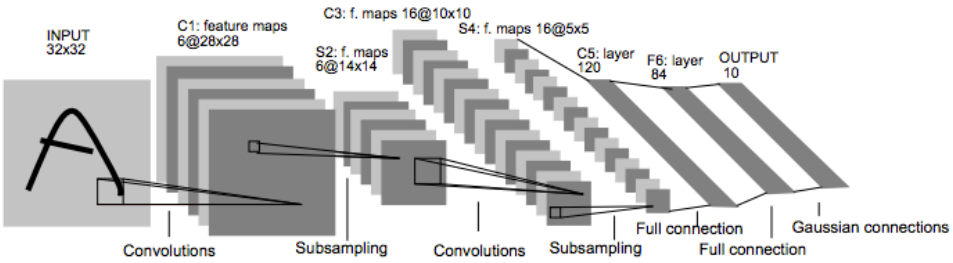

図2.2　手書き文字認識におけるCNN（図は論文[3]よりの引用）

この図では，手書き文字認識におけるCNNのモデルが示されている．入力として，32×32のサイズの手書き文字画像を入力し，画像を6枚28×28の特徴量マップに畳み込んでいる．この時，特徴量のサイズが元の画像よりも小さくなるのは，畳み込みによって複数ドットの情報が1つのドットとして表現されるためである．この図のモデルでは，畳み込みが2回とサブサンプリング（プーリング）を2回それぞれ交互に行っている．これにより，徐々に特徴量を絞り込んでいくのである．これらの工程の後に，それぞれの特徴量を結合していき，最終的に10次元の情報として出力している．出力されるのは，その文字がどの文字である可能性が高いのかという確率になる．この図では，文字がAである確率が一番高ければ正解，そうでなければ間違いというように結果を評価することが出来る．つまり，画像の解像度を徐々に落としていくことで，文字の特徴をわかりやすくしていく作業であると言える．ただし，解像度を落とす頻度が多すぎたり，

1度に落とす解像度が大きすぎる場合には，うまく特徴を取ることが出来ない可能性もあるため，モデルを構築する際には慎重に行う必要がある．

CNNに限らず，ニューラルネットワークとは，学習にコストがかかる技術である．ニューラルネットワークの学習とは，訓練データ（特徴量など）の入力で訓練データの回答と同じ出力が出るように各層での重み付けを調整することを言い，そのための方法として前節で触れた誤差逆伝播法が多く用いられている．ニューラルネットワークを学習させ，重みを調整することで入力と同じものが出力される，つまり回答が未知のデータであっても判別ができるニューラルネットワークになるのである．学習には，純粋な計算能力はもとより，多数のニューラルネットワークの層を記憶・管理する必要があるため，一定以上のメモリが必要となる．また，画像処理のニューラルネットワークでは，一般的に次元数や層の数が多くなりがちであり，他のニューラルネットワークの学習に比べてメモリの使用量が多くなりやすい．最近ではサイズの大きな画像や画素数が大きい画像に対しての学習も多く行われており，訓練データの肥大化に合わせて，ニューラルネットワーク自体も巨大な物を使用する傾向があり，学習のためのコストは徐々に増大している．そのため，画像の特徴量のみを必要とする，画像を認識して何か別のことに使用するといった研究（画像キャプション生成など）では，すでに学習済みのニューラルネットワークのモデルを使用することがしばしばある．

今回の画像キャプション生成の深層学習モデルでは，画像をCNNで畳み込んで特徴量として捉えることで，RNNを使った文章生成のための入力としている．実験ではCNNから出力される特徴量の次元は4096次元であった．また，使用したCNNは学習のためのコストを抑えるために，論文[9]で発表されているものを用いた．このモデルは以下のURLにて公開されている．

`https://gist.github.com/ksimonyan/3785162f95cd2d5fee77#file-readme-md`

2.6　RNN

RNNとは再帰型ニューラルネットワーク(Recurrent Neural Network)の略称である．このニューラルネットワークは主に時系列的なデータの学習に用いられる．時系列的なデータとは，前のデータに次のデータの値が左右される文章や音声，動画といったデータのことである．

RNNとは再帰的（並列）にニューラルネットワークを構築することにより，時系列的なデータに対しての学習を可能にしたモデルである．シンプルな形のRNNでは，再帰を重ね続けることによって，学習がうまく行えなくなる問題が発生していたが，LSTM（Long Short Time Memory）とよばれる手法を組み込んだLSTMRNNという型を用いることによって，その問題を克服した．これによって，長い時系列データに対する学習が可能になったことで注目を集めている．今回の文章生成に用いたのもLSTMを利用したRNNである．また，多くのニューラルネットワークでは，入力や出力が固定次元であり，系列の量がそれぞれ違う場合（文章ごとに長さが違うなど）において適用することは難しかったが，RNNではそのようなデータに対しても適用することが可能である．

図2.3は実験で実際に使用したCNNとRNNによる画像キャプション生成モデルである．一番左の画像が入力されてLSTMと書かれたマスに入力が繋がっている．この部分がCNNによる画像の特徴量抽出を表している．LSTMの詳しい説明は後述するが，RNNの種類の1つである．RNNではニューラルネットワークを再帰的にして時系列的な処理を行う．この図においてRNNの部分はCNNからの出力を受け取った部分から右のまとまりを指す．これは，再帰的に定義されたRNNをわかりやすくするために並列のものとして展開した図である．

この図における学習の流れを辿りながらRNNによる文章生成を説明する．この時，学習用の訓練データ

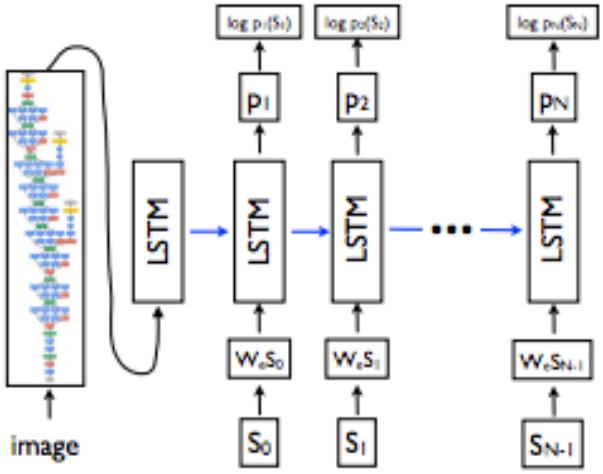

図2.3　実験で使用したCNNとRNNによる画像キャプション生成モデル（図は論文[10]より引用）

として，画像とそれに付随する正解のN-1単語からなるキャプション($S_1 \sim S_{N-1}$)が与えられている．また，入力の際にこの文章には始端記号S_0と終端記号S_Nが付与される．まず，文章生成に先立って，前情報である画像の特徴量をLSTMへ入力する．そして，並列化されたニューラルネット（RNN）に文章の始端記号となるS_0を入力する．これに重みW_eをかけたものをベクトルとしてニューラルネットの層へと入力する．（このときLSTMについてはひとまず入力に対して出力を返す装置と考えておく）LSTMを通って出てきたp_1は入力S_0と事前に入力された画像の特徴量に対する出力であり，意味としては次に生成される単語の確率の集合となる．最後に，出力された確率のコストを最小化($\log p_1 S_1$)することで，最適な単語を出力する．これによって文章の先頭の単語が出力される．

次に，先頭の単語を出力する際にLSTMで得られた情報を隣のS_1が入力となるLSTM（第2単語を生成するLSTM）へと入力する．S_1はS_0の時同様重みW_eをかけてベクトル化された後に，前のRNN（S_0が入力されたニューラルネットワーク）の情報を持つLSTMへと入力される．そして，S_0のLSTMから渡された情報と$W_e S_1$からp_2が生成され，最終出力$\log p_2 S_2$となる．これを繰り返していき，文章の最後の単語S_{N-1}までをニューラルネットワークに入力し，終端記号S_Nが出力される．こうして文章の生成を行った後，出力と入力データの間違いからRNN用の誤差逆伝播法でニューラルネットワークを学習させていく．

具体例を出して説明する．”This is a pen.” という文章と，ペンが1本写った画像がある場合に，まずは画像をCNNに通してペンが写っているという特徴量を抽出し，始端記号S_0が入力されるニューラルネットワークへ入力する．この時出力として確率の集合p_1が出力される．この確率とS_1である単語 This とのコストを計算し一番小さなものが出力となる(This がp_1に含まれていればそれが最小)．次に，S_0を入力したニューラルネットワークのLSTMの情報と，単語 This (S_1)を入力し，p_2を生成し単語 is (S_2)とのコストを計算する．この作業を繰り返していき，入力が最終単語 pen で出力が終端記号S_Nとなったところで生成を終了する．その後，出てきた単語列と訓練データとの差から重みの調整（学習）をおこなう．実際に生成を行う場合は，訓練データのときのように元になるデータがないため画像の特徴量と始端記号を入力した後，前のニューラルネットワークで生成された出力（単語）を次のニューラルネットワークの入力として，終端記号が出力されるまで続けていく．

RNNの学習には通常のニューラルネットワークのような誤差逆伝播法ではなく，BPTT（Back-Propagation Through Time）とよばれる時系列的な構造も考慮した逆伝播の学習方法が採られてい

る．この方法は，RNNを展開した際に，1つのニューラルネットワークの出力が次のニューラルネットワークの入力になっていることから，層の深いニューラルネットワークとみなして誤差逆伝播法を適用する方法である．RNNでは通常のニューラルネットワークとは違い，中間層のデータも次のニューラルネットワークの入力として使用しているため，ニューラルネットワークが出力した系列の微分と，中間層のデータの系列の微分とを用いて誤差逆伝播法を適用する．しかし，この方法は通常の誤差逆伝播法と同様に勾配消失問題が発生する場合がある，そのために提案されたのがLSTMである．このBPTTにも年々改良が加えられている．また，BPTT以外の学習方法もいくつかあるため，データの特性や用途に応じて使い分けることが必要である．

LSTMとは，RNNの中間層に用いられるユニットの種類である．このユニットは記憶の保持を目的としたユニットである．RNNでは時系列的なデータを扱うことが出来るが，シンプルなタイプのRNNでは，時系列が長くなるに連れて勾配消失問題と呼ばれる問題が発生し，誤差逆伝播法が出来ず，上手く学習を行う事ができなかった（数ステップ以上の長さの学習は難しかった）．LSTMはその問題を解決するために，メモリセルと呼ばれる機構を持ち，これにより長期的な記憶の保持が出来るようになっている．しかし，保持し続けるだけでは，誤ったデータも保持し続けてしまう事があるため，それをリセットする忘却機構などが備え付けられている．これらの機構により，LSTMを用いたタイプのRNNは時系列の長い情報に対しても学習を行うことが出来るようになった．説明したのはLSTMの中でも比較的シンプルなタイプであるが，現在でもLSTMには様々な改良が加えられており，様々な種類のLSTMが提案されている．

今回の画像キャプション生成の深層学習モデルでは，CNNによって出された特徴量と，訓練データの文章を文章ごとに単語に分解したものを，RNNの入力とし，中間層（隠れ層）で512次元に変換し，出力層で入力された文章の単語数と同じ次元で出力している．

第3章

複数形表現の統一

3.1 複数形表現の問題

複数形における問題点とは，複数の対象が写っている画像（犬や人間などが複数写り込んでいるもの）に対して生成されたキャプションにおいて，主語が単数になっていたり，複数形になっていても数が間違っていたりする[*1]ことである．これら間違いの多くは基数を含んだ複数形表現を生成した場合に現れる．

この問題を解決するために，学習に用いられるデータセットにおける複数形の表現をより単純な形に書き換え，複数形の表現方法を統一する．

実験では，MSCOCOのAnnotationつき画像データセットを用いた．ただし，MSCOCOで直接配布されているデータセットではなく論文[2]で使用され，以下のURLで公開されているデータセットを用いた．

`http://cs.stanford.edu/people/karpathy/deepimagesent`

訓練データにおける複数形表現を統一した場合の，生成文の品質の変化を調べるために，変更を加えないデータをデフォルトデータセットとして，それを含め5パターンの訓練データを用意した．その内訳は以下のとおりである．

1. デフォルトのデータセット
2. 1について，頭文字の表記ゆれを取り除くために全ての文章を小文字にしたもの
3. 2について，複数形の文章について，基数の部分を some に置換したもの
4. 3について，some となっている部分を a group of にしたもの
5. 4について，a couple of となっている表現を a group of に統一したもの

これら，2から5のパターンのデータセットを用いて学習を行った深層学習モデルによって生成されたキャプションと，デフォルトデータセットを用いて学習を行ったモデルによって生成されたキャプションと主観的に比べることで品質が向上したかどうかを調べる．

3.2 TreeTagger

パターン3を作成するにあたって，英文の品詞を形態素解析によって調べるためにTreeTagger[7, 8]を用いた．これは，シュトゥットガルト大学のHelmut Schmidが開発したツールで，英語やフランス語，ドイツ語など多様な言語に対しての形態素解析を行うことが出来るツールである．また，多種のプログラミング言語にも対応しており，PythonやJava，Rubyなどで使用できる．実験では置換を行うプログラムを

[*1] 画像には3つ写っているのに2つと書かれるなど

Pythonで書いていたため，Pythonを使用して実装を行った．具体的には，訓練データにおける全文章をTreeTaggerによって形態素解析し，1つの文章中に基数（Cardinal number：品詞コードはCD）が存在し，名詞の複数形（noun plural ：品詞コードはNNS）が存在する文章[*2]に対して，品詞コードがCDである単語についての置換を行った．この際，単語が one の場合は，単純化のために置換可能であると想定した単語 some に置換することが出来ないため省いている．また，TreeTaggerを用いたのはパターン3をを作成する時のみであり，パターン4,5については，Pythonの正規表現を用いて， some や a couple of に合致した部分を置換している．

[*2] Two dogs are running in the yard といった文章

第4章

実験

4.1 実験設定

今回の実験では，全ての訓練データのパターンに対して，RNNの学習の回数を100回とし，学習を行ったモデルから生成されたキャプションを結果として主観的に評価を行った．学習したモデルからキャプションを生成する際に使用したテスト用のデータには，学習時に使用されていない画像を100枚用意した．このうち，50枚を複数の対象が写っている画像，もう50枚を対象が1つであったり風景のみの画像とした．これは，複数形表現の品質が向上したかどうかだけでなく，それ以外の表現についても品質の向上，もしくは低下が見られるかどうかの検証を行うためである．

実験に用いた訓練データの調整は以下のような手順で行った．

1. 訓練データの全キャプションを小文字に変換（パターン2の作成）
2. TreeTaggerを用いての品詞の確認・基数の置換（パターン3の作成）
3. 単語someの置換（パターン4の作成）
4. 慣用句 a couple of の置換（パターン5の作成）

また，実行環境は以下の表4.1のとおりである．この環境下において，GPUを用いてRNNの学習100回にかかった時間は14時間程度であった．

表4.1 実行環境

OS	Ubuntu 16.04 LTS
GPU	GeForce GTX 750 Ti
Python	2.7.0 Anaconda 2.4.0
Chainer	1.17.0

4.2 MSCOCOデータセット

MSCOCO（Microsoft Common Objects in Context）データセットとは画像に対してキャプションが付与されているデータセットであり，Microsoftによって提供されている．このデータセットに対する詳しい説明はMSCOCOの論文[4]にかかれているので今回は省略し，実験で使用したMSCOCOデータセットに変更を加えて作成されたデータセットについて説明する．

実験で使用したデータセットの内容は，画像枚数が123,287枚，それらに付与された総キャプション数が616,767文となっている．TreeTaggerを用いて全文章についての形態素解析を行い，カウントを行った結

果，デフォルトデータセットのうち複数形が含まれている文章は284,549文あり，さらにそのうちの65,643文が基数を含むキャプションであった．この基数を含むキャプションの基数部分を some に置き換えたものがパターン3のデータセットとなる．ただし，複数形が含まれており基数も存在するが，その基数が one のみである場合は some に置き換えることができないため，省いている．そのため，実際に置換を行ったキャプション数は63,794文であった．

次に，パターン4を作成するためにキャプション中に some を含むキャプションに対して置換を行った．some を a group of に置換したキャプション数は86,173文であった．この数は，デフォルトデータセットの時点で some が使用されていた文章も含むため，パターン3で置換した文章数よりも多くなっている．また，置換した後にカウントを行っているため，デフォルトデータセットの時点で some が含まれており，パターン3の作成時に some が増えた文章については1つとして数えられている．

そして，デフォルトデータセットのうち a couple of が使用されていたキャプションが9,433文，a group of が使用されていたキャプションが19,359文あった．パターン5を作成するためにこの9,433文に含まれている a couple of を全て a group of に置換した．

以上から，パターン5における a group of を含むキャプション数は，113,072文となった．置換したキャプションの総和よりも最終的な a group of を含むキャプション数が少ないのは，パターン2からパターン5まで順に変換していく間に一つのキャプションに変換すべき語がいくつかあるキャプション[*1]があったため，重複してカウントされたものと考えられる．

以下に各パターンを生成するために変更した文章数を単語ごとにまとめたものを表4.2として示す．

表4.2　単語ごとの変更文章数

変更した単語	変更した文章数
基数	65643
some	86173
a couple of	9433

4.3　結果

実験の結果として，キャプション生成に用いた画像と，それぞれの3章で説明した各パターンのモデルにおける生成文を5文ずつ以下に示す．

図4.1において，それぞれのパターンでの生成結果を以下に示す．

パターン1を用いた場合

a bird that is standing in the water
a flock of birds standing on top of a beach
a couple of birds standing on top of a beach
a bird that is standing on a beach
a flock of birds sitting on top of a beach

[*1] some が使用されているかつ a couple of が使用されているキャプションなど．

図4.1　生成文がより良くなった場合

パターン2を用いた場合

a bird that is standing in the water
a bird flying over a body of water
a bird standing on top of a sandy beach
a couple of birds that are standing in the water
a bird standing on top of a body of water

パターン3を用いた場合

a bird flying over a body of water
a bird that is standing in the water
a couple of birds that are standing in the water
a bird that is sitting on a rock
a bird that is standing on a beach

パターン4を用いた場合

a group of ducks swimming in a body of water
a group of birds sitting on top of a sandy beach
a group of birds sitting on a rock in the water
a group of birds are standing in the water
a black and white bird sitting on a rock

パターン5を用いた場合

a group of ducks swimming in a body of water
a group of ducks are swimming in the water
a group of birds that are standing in the water
a group of ducks that are standing in the water
a group of ducks are swimming in a pond

図4.2において，それぞれのパターンでの生成結果を以下に示す．

図4.2　生成文が悪くなった例

パターン1を用いた場合

a brown and white dog laying on a bed
a dog laying on top of a bed
a black and white dog laying on a bed
a brown and white dog laying on top of a bed
a close up of a dog laying on a bed

パターン2を用いた場合

a close up of a dog wearing a hat
a close up of a dog laying on a couch
a close up of a dog wearing a santa hat
a black and white cat sitting on a chair
a close up of a dog sitting on a chair

パターン3を用いた場合

a dog that is sitting in the grass
a dog that is laying down in the grass
a dog that is sitting on a couch
a dog that is laying down on a bench
a close up of a dog on a leash

パターン4を用いた場合

a dog that is sitting in the grass
a dog that is sitting on a couch
a dog that is sitting on a chair
a dog that is laying down on a couch
a close up of a dog laying on a couch

パターン5を用いた場合

a close up of a cat laying on a bed
a close up of a cat laying on a couch
a close up of a dog laying on a couch
a close up of a dog laying on a bed
a cat that is laying down on a couch

100枚の画像に対して，デフォルトデータセットでの生成文と各訓練データのパターンでの生成文を比較した結果を表4.3に示す．

表4.3　デフォルトデータセットとの比較

	良くなった生成文数	悪くなった生成文数
パターン2	13	11
パターン3	10	12
パターン4	15	13
パターン5	25	14

この結果より，置換したキャプション数が多くなっていくに連れて変化したキャプション数が増加していることがわかる．また，パターン4までは良くなったキャプション数と悪くなったキャプション数にほとんど差がないが，パターン5では良くなったキャプションがより多くなっている．パターン3では良くなったキャプション数よりも悪くなったキャプション数のほうが多くなっているが，生成されたキャプションを見ると，置換して増加したはずの単語some がほとんど使用されていなかった（500文中1,2文程度であった）．

パターン4とパターン5では，置換した a group of やa couple of の表現が多く出現していた．その影響でキャプションが良くなった場合が多く見られたが，副作用として主語となる対象の認識が誤っているものも見られた[*2]．

[*2] 数匹の猫が写っている画像を羊と表現するなど．

第5章

考察

深層学習モデルを用いた画像キャプション生成において，間違ったキャプションが生成された場合，その原因がどこにあるのかを特定することは容易ではない．これは本実験でも確認できている．つまり，本実験では訓練データの書き換えを行うことで書き換えた部分はもちろん，書き換えられていない部分においても誤りの改善が見られた．これは，深層学習における文章生成の部分において，前の語を踏まえた上で次の語の確率を求めるという仕組みに深く関係があると考えられる．例えば，深層学習での文章生成では，訓練データにおいて two という単語の次に来る単語が dogs と cars しかなかった場合には， two のあとには，dogs か cars が続く傾向があるのだとモデルは学習してしまう．そのため，2匹の羊が写っている画像を入力として与えた場合に，two という単語が生成されたとしても，その次に sheep という単語が出現しづらくなる．今回の実験では two に当たる部分，つまり複数形の表現を統一したことによって，次に来る語の選択肢が大幅に増えた．それにより，書き換えを行った部分以外でも誤りが発生しにくくなったと考えられる．

また，本研究はキャプションの言語が英語であることを前提としている．キャプションの言語が日本語になった場合，単数や複数の表現はもともと曖昧であるため，本研究で取ったアプローチは使えない．本研究は文章の粒度を荒くするという点に着目し，具体的な改善策として，英語での文章生成では複数形表現の統一という手法をとった．そのため，日本語での画像キャプション生成[5]の場合には，文章の粒度を荒くするための手法を改めて考える必要がある．

第6章

結論

本論文では，画像キャプション生成の複数形表現に注目することで，生成されるキャプションの品質の向上を行った．具体的には，訓練データのキャプションの複数形表現を統一した．MSCOCOデータセットを訓練データに用いた実験では，生成されるキャプションの品質の向上が確認できた．

画像キャプション生成において，間違ったキャプションを生成した場合，その原因がどこにあるのかを特定するのは容易ではない．本実験の結果はそれを示唆している．今後はこの点について考察して行きたい．また，本研究での訓練データのキャプションの内容の粒度を荒くするというアイデアを日本語のキャプション生成でも試したい．

謝辞

卒業研究を進めるにあたり，熱心にご指導いただいた情報工学科の新納教授に感謝いたします．また，多くのご指摘をいただきました自然言語処理研究室の皆様に感謝します．

参考文献

[1] Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, and David Forsyth. Every picture tells a story: Generating sentences from images. In *European Conference on Computer Vision*, pp. 15–29. Springer, 2010.

[2] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, 2015.

[3] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755. Springer, 2014.

[5] Takashi Miyazaki and Nobuyuki Shimizu. Cross-lingual image caption generation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1780–1790, 2016.

[6] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318. Association for Computational Linguistics, 2002.

[7] Helmut Schmid. Improvements in part-of-speech tagging with an application to german. In *In proceedings of the acl sigdat-workshop*. Citeseer, 1995.

[8] Helmut Schmid. Probabilistic part-ofispeech tagging using decision trees. In *New methods in language processing*, p. 154. Routledge, 2013.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3156–3164, 2015.

[11] 西友佑, 新納浩幸, 古宮嘉那子, 佐々木稔. 画像キャプション生成における複数形表現の統一. 言語処理学会第23回年次大会(NLP2017), p. to appear, 2017.

付録

付録として，パターン５で学習した場合のキャプションとデフォルトデータセットで学習した場合のキャプションを比較した際に，変化が大きかったものを入力した画像とともに以下に示す．

生成結果が改善された例

図1　改善された例1

図1についての生成結果（デフォルトデータセット）

a horse that is standing in the grass
a cow that is standing in the grass
a herd of cattle grazing on a lush green field
a couple of cows are standing in a field
a horse that is standing in a field

図1についての生成結果（パターン５）

a horse that is standing in the grass
a group of cows that are standing in the grass
a group of cows are standing in a field
a group of cows standing in a grassy field
a group of cows are grazing in a field

図2　改善された例2

図2についての生成結果（デフォルトデータセット）

a bird perched on top of a tree branch
a bird that is sitting on a branch
a bird is perched on a tree branch
a bird sitting on top of a tree branch
a bird that is perched on a branch

図2についての生成結果（パターン５）

a bird perched on top of a tree branch
a group of birds sitting on a tree branch
a group of birds sitting on top of a tree branch
a bird sitting on top of a tree branch
a group of birds perched on a tree branch

図3　改善された例3

図3についての生成結果（デフォルトデータセット）

a man riding on the back of a brown horse
a woman riding on the back of a brown horse
a horse that is standing in the grass
a horse that is standing in the dirt
a man riding on the back of a horse

図3についての生成結果（パターン５）

a horse that is standing in the grass
a brown horse standing on top of a lush green field
a horse that is standing in the dirt
a close up of a horse in a field
a brown and white horse standing in a field

図4　改善された例4

図4についての生成結果（デフォルトデータセット）

a cow that is standing in the grass
a cow that is standing in the dirt
a couple of cows that are standing in the grass
a couple of horses standing next to each other
a couple of cows are standing in a field

図4についての生成結果（パターン５）

a group of cows that are standing in the dirt
a group of cows standing next to each other
a group of horses that are standing in the dirt
a group of horses standing next to each other
a group of cows that are standing in the grass

図5　改善された例5

図5についての生成結果（デフォルトデータセット）

a bird that is standing in the water
a flock of birds standing on top of a beach
a couple of birds standing on top of a beach
a bird that is standing on a beach
a flock of birds sitting on top of a beach

図5についての生成結果（パターン５）

a group of ducks swimming in a body of water
a group of ducks are swimming in the water
a group of birds that are standing in the water
a group of ducks that are standing in the water
a group of ducks are swimming in a pond

図6　改善された例6

図6についての生成結果（デフォルトデータセット）

a bird that is standing in the grass
a bird that is standing in the sand
a bird that is standing on a rock
a bird that is standing in the dirt
a bird that is sitting in the grass

図6についての生成結果（パターン５）

a group of birds are flying in the sky
a group of birds flying over a body of water
a group of birds that are flying in the air
a group of birds that are flying in the sky
a group of birds are flying in the air

図7　改善された例7

図7についての生成結果（デフォルトデータセット）

a dog holding a frisbee in its mouth
a dog with a frisbee in its mouth
a dog that is laying down in the grass
a dog that is sitting in the grass
a dog that is standing in the grass

図7についての生成結果（パターン５）

a dog that is laying down on a bench
a brown and white dog laying on the ground
a brown dog laying on top of a lush green field
a group of dogs that are laying on a bench
a brown and white dog laying on a bed

図8　改善された例8

図8についての生成結果（デフォルトデータセット）

a man riding on the back of a brown horse
a horse that is standing in the grass
a horse that is standing in the dirt
a man riding on the back of a horse
a man standing next to a brown horse

図8についての生成結果（パターン５）

a horse that is standing in the grass
a horse that is standing in the dirt
a brown horse standing on top of a lush green field
a brown and white horse standing next to a fence
a brown and white horse standing in a field

図9　改善された例9

図9についての生成結果（デフォルトデータセット）

a man riding on the back of a brown horse
a horse that is standing in the grass
a horse that is standing in the dirt
a man riding on the back of a horse
a couple of horses that are standing in the grass

図9についての生成結果（パターン５）

a group of horses that are standing in the grass
a horse that is standing in the grass
a group of cows that are standing in the grass
a group of horses standing next to each other
a horse that is standing in the dirt

図10　改善された例10

図10についての生成結果（デフォルトデータセット）

a teddy bear sitting on top of a bed
a teddy bear is sitting on a chair
a cat laying on top of a wooden table
a teddy bear sitting on top of a table
a brown teddy bear sitting on a chair

図10についての生成結果（パターン５）

a close up of a cat laying on a bed
a cat that is laying down on a bed
a cat laying on top of a laptop computer
a close up of a cat on a bed
a group of stuffed animals sitting on top of a bed

図11　改善された例11

図11についての生成結果（デフォルトデータセット）

a bunch of broccoli growing on a tree
a bunch of broccoli growing in a tree
a close up of a bunch of broccoli
a plant that is growing in a tree
a close up of a green broccoli plant

図11についての生成結果（パターン５）

a vase filled with flowers on a table
a vase filled with flowers sitting on top of a table
a vase filled with flowers sitting on a table
a vase filled with purple flowers on a table
a vase filled with flowers on top of a table

図12　改善された例12

図12についての生成結果（デフォルトデータセット）

a horse that is standing in the grass
a man riding on the back of a brown horse
a horse standing on top of a lush green field
a horse that is standing in the dirt
a horse standing on top of a grass covered field

図12についての生成結果（パターン５）

a horse that is standing in the grass
a horse that is standing in the dirt
a cow that is standing in the grass
a cow that is standing in the dirt
a group of cows that are standing in the grass

図13　改善された例13

図13についての生成結果（デフォルトデータセット）

a close up of a person holding a cell phone
a man holding a cell phone in his hand
a close up of a person wearing a suit and tie
a close up of a man wearing a suit and tie
a man that is standing in front of a building

図13についての生成結果（パターン５）

a group of people sitting around a table
a group of people sitting at a table
a man sitting on a couch with a laptop
a group of people standing around a table
a group of people sitting around a table with laptops

図14　改善された例14

図14についての生成結果（デフォルトデータセット）

a couple of men standing next to each other
a couple of people that are standing in the grass
a man holding a cell phone in his hand
a man and a woman sitting on a bench
a man holding a cell phone in his hands

図14についての生成結果（パターン５）

a group of men standing next to each other
a group of people standing next to each other
a group of women standing next to each other
a man and a woman standing next to each other
a man and a woman sitting on a bench

図15　改善された例1

図15についての生成結果（デフォルトデータセット）

a dog that is standing in the grass
a couple of cows that are standing in the dirt
a couple of cows that are standing in the grass
a dog that is standing in the dirt
a dog that is sitting on a bench

図15についての生成結果（パターン５）

a group of cows that are standing in the grass
a group of cows that are standing in the dirt
a group of cows that are standing in the snow
a group of dogs that are standing in the grass
a group of dogs standing next to each other

図16　改善された例16

図16についての生成結果（デフォルトデータセット）

a close up of a dog on a couch
a couple of dogs sitting on top of a couch
a close up of a dog laying on a bed
a close up of a dog on a bed
a dog laying on top of a couch

図16についての生成結果（パターン５）

a group of dogs standing next to each other
a group of cows that are standing in the dirt
a group of cows that are standing in the grass
a group of cows that are standing in the snow
a group of dogs that are standing in the grass

図17　改善された例17

図17についての生成結果（デフォルトデータセット）

a man riding a horse in a field
a couple of people that are standing in the grass
a man is riding a horse in a field
a man standing next to a brown horse
a man standing next to a horse in a field

図17についての生成結果（パターン５）

a group of people that are standing in the grass
a group of people that are sitting on a bench
a group of people standing next to each other
a group of people standing on top of a lush green field
a group of people standing on top of a sandy beach

図18　改善された例18

図18についての生成結果（デフォルトデータセット）

a herd of cattle grazing on a lush green field
a cow that is standing in the grass
a couple of cows that are standing in the grass
a couple of cows are standing in a field
a herd of cows grazing on a lush green field

図18についての生成結果（パターン５）

a herd of cattle grazing on a lush green field
a group of cows that are standing in the grass
a herd of sheep grazing on a lush green field
a group of cows are standing in a field
a herd of cattle grazing on a lush green hillside

図19　改善された例19

図19についての生成結果（デフォルトデータセット）

a man riding on the back of a brown horse
a woman riding on the back of a brown horse
a man riding on the back of a horse
a woman riding on the back of a horse
a horse that is standing in the dirt

図19についての生成結果（パターン５）

a horse that is standing in the dirt
a horse that is standing in the grass
a group of horses that are standing in the dirt
a brown horse standing on top of a lush green field
a group of horses standing next to each other

図20　改善された例20

図20についての生成結果（デフォルトデータセット）

a man riding on the back of a brown horse
a horse that is standing in the dirt
a man riding on the back of a horse
a black and white photo of a horse
a couple of horses standing next to each other

図20についての生成結果（パターン５）

a horse that is standing in the dirt
a horse that is standing in the grass
a brown horse standing on top of a lush green field
a brown horse standing on top of a sandy beach
a group of horses standing next to each other

図21　改善された例21

図21についての生成結果（デフォルトデータセット）

a cow that is standing in the grass
a couple of cows that are standing in the grass
a couple of cows are standing in a field
a herd of cattle grazing on a lush green field
a black and white cow standing in a field

図21についての生成結果（パターン５）

a group of cows that are standing in the grass
a group of cows are standing in a field
a group of cows standing in a grassy field
a group of cows standing in a field
a group of cows standing next to each other

図22　改善された例22

図22についての生成結果（デフォルトデータセット）

a couple of cows that are standing in the grass
a cow that is standing in the grass
a herd of cattle grazing on a lush green field
a couple of cows are standing in a field
a herd of sheep grazing on a lush green field

図22についての生成結果（パターン５）

a group of cows that are standing in the grass
a group of cows are standing in a field
a group of cows standing in a grassy field
a group of sheep standing on a lush green field
a group of cows are standing in the grass

図23　改善された例23

図23についての生成結果（デフォルトデータセット）

a yellow fire hydrant sitting in the grass
a red fire hydrant sitting in the grass
a yellow fire hydrant in a grassy field
a red fire hydrant in a grassy field
a yellow fire hydrant in a grassy area

図23についての生成結果（パターン５）

a close up of a vase of flowers
a close up of a vase filled with flowers
a close up of a vase with flowers in it
a vase filled with flowers sitting on a table
a close up of a vase on a table

図24　改善された例24

図24についての生成結果（デフォルトデータセット）

a dog that is standing in the grass
a dog that is laying down in the grass
a dog that is laying on the ground
a brown teddy bear sitting on top of a couch
a dog that is laying on the floor

図24についての生成結果（パターン５）

a group of cows that are standing in the grass
a group of cows standing next to each other
a group of dogs standing next to each other
a group of cows that are standing in the dirt
a group of dogs that are standing in the grass

図25　改善された例25

図25についての生成結果（デフォルトデータセット）

a red fire hydrant sitting in the grass
a yellow fire hydrant sitting in the grass
a red fire hydrant in a grassy field
a red fire hydrant in a grassy area
a yellow fire hydrant in a grassy field

図25についての生成結果（パターン５）

a close up of a vase of flowers
a close up of a vase on a table
a vase filled with flowers on a table
a vase filled with purple flowers on a table
a close up of a vase filled with flowers

生成結果が改悪された例

図26　改悪された例1

図26についての生成結果（デフォルトデータセット）

a dog that is standing in the grass
a dog with a frisbee in its mouth
a dog that is laying down in the grass
a brown and white dog laying in the grass
a brown and white dog playing with a frisbee

図26についての生成結果（パターン５）

a group of sheep that are standing in the grass
a close up of a sheep in a field
a group of sheep standing next to each other
a group of sheep standing on top of a grass covered field
a group of sheep standing on top of a lush green field

図27　改悪された例2

図27についての生成結果（デフォルトデータセット）

a couple of sheep standing next to each other
a close up of a sheep in a field
a teddy bear sitting on a wooden bench
a couple of sheep standing on top of a lush green field
a couple of sheep standing on top of a grass covered field

図27についての生成結果（パターン５）

a group of teddy bears sitting next to each other
a close up of a group of stuffed animals
a close up of a teddy bear in a field
a close up of a stuffed animal in a field
a brown teddy bear sitting on top of a table

図28　改悪された例3

図28についての生成結果（デフォルトデータセット）

a sheep that is standing in the grass
a herd of sheep grazing on a lush green field
a sheep that is standing in a field
a close up of a sheep in a field
a herd of sheep standing on top of a lush green field

図28についての生成結果（パターン５）

a group of sheep that are standing in the grass
a group of sheep standing next to each other
a group of sheep standing on top of a lush green field
a group of sheep standing on a lush green field
a group of sheep standing next to each other in a field

図29　改悪された例4

図29についての生成結果（デフォルトデータセット）

a brown and white dog laying on a bed
a dog laying on top of a bed
a black and white dog laying on a bed
a brown and white dog laying on top of a bed
a close up of a dog laying on a bed

図29についての生成結果（パターン５）

a close up of a cat laying on a bed
a close up of a cat laying on a couch
a close up of a dog laying on a couch
a close up of a dog laying on a bed
a cat that is laying down on a couch

図30　改悪された例5

図30についての生成結果（デフォルトデータセット）

a close up of a cat on a table
a close up of a cat in a sink
a close up of a cat in a bowl
a cat sitting on top of a toilet seat
a cat that is sitting on top of a toilet

図30についての生成結果（パターン５）

a close up of a bird on a table
a close up of a bird in a bowl
a close up of a bird on a plate
a close up of a bird on a branch
a close up of a cat on a table

図31　改悪された例6

図31についての生成結果（デフォルトデータセット）

a dog that is standing in the grass
a dog with a frisbee in its mouth
a couple of sheep standing on top of a lush green field
a dog standing in the grass with a frisbee
a close up of a dog in a field

図31についての生成結果（パターン５）

a group of sheep that are standing in the grass
a group of sheep standing on a lush green field
a group of sheep are standing in a field
a group of sheep standing on top of a lush green field
a herd of sheep grazing on a lush green field

図32　改悪された例7

図32についての生成結果（デフォルトデータセット）

a large building with a clock on it
a clock on the side of a building
a large clock on the side of a building
a clock that is on the side of a building
a building with a clock on top of it

図32についての生成結果（パターン５）

a group of people standing next to each other
a group of people standing in front of a building
a group of people sitting on a bench
a group of people sitting around a table
a man standing in front of a building

図33　改悪された例8

図33についての生成結果（デフォルトデータセット）

a close up of a cat on a table
a close up of a cat laying on a table
a close up of a dog in a room
a close up of a dog on a table
a close up of a dog laying on a bed

図33についての生成結果（パターン５）

a group of sheep standing next to each other
a group of polar bears standing next to each other
a group of sheep that are standing in the grass
a group of sheep that are standing in the dirt
a group of sheep that are standing in the snow

図34　改悪された例9

図34についての生成結果（デフォルトデータセット）

a bird that is standing in the water
a bird flying over a body of water
a bird that is standing on a beach
a flock of birds flying over a body of water
a couple of birds standing on top of a beach

図34についての生成結果（パターン５）

a person on a surfboard in the water
a group of birds flying over a body of water
a group of people sitting on a boat in the water
a bird sitting on top of a tree branch
a group of ducks swimming in the water

図35　改悪された例10

図35についての生成結果（デフォルトデータセット）

a brown and white dog laying on top of a bed
a couple of dogs laying on top of a bed
a brown and white dog laying on a bed
a close up of a dog laying on a bed
a dog laying on top of a bed

図35についての生成結果（パターン５）

a group of brown bears standing next to each other
a group of cows that are standing in the grass
a group of cows standing next to each other
a group of brown and white cows standing next to each other
a group of cows that are standing in the dirt

図36　改悪された例11

図36についての生成結果（デフォルトデータセット）

a horse that is standing in the grass
a herd of cattle grazing on a lush green field
a herd of cattle grazing on a lush green hillside
a herd of cattle standing on top of a lush green field
a couple of cows are standing in a field

図36についての生成結果（パターン５）

a group of elephants that are standing in the grass
a group of elephants standing next to each other
a group of elephants standing in a field
a group of elephants that are standing in the dirt
a group of elephants are standing in a field

図37　改悪された例12

図37についての生成結果（デフォルトデータセット）

a couple of giraffes that are standing in the grass
a couple of giraffe standing next to each other
a couple of giraffes are standing in a field
a herd of cattle grazing on a lush green field
a couple of giraffes standing in a field

図37についての生成結果（パターン５）

a group of elephants that are standing in the grass
a group of elephants standing in a field
a group of elephants are standing in a field
a group of elephants standing next to each other
a group of elephants standing in a grassy field

図38　改悪された例13

図38についての生成結果（デフォルトデータセット）

a cat laying on top of a wooden table
a cat sitting on top of a wooden table
a close up of a cat on a table
a cat laying on top of a wooden bench
a cat sitting on top of a wooden bench

図38についての生成結果（パターン５）

a group of sheep standing next to each other
a group of sheep that are standing in the grass
a group of sheep standing next to each other on a field
a group of sheep that are standing in the dirt
a close up of a group of sheep in a pen

図39　改悪された例14

図39についての生成結果（デフォルトデータセット）

a bird that is standing in the water
a bird that is sitting in the water
a bird that is standing on a rock
a black and white bird standing in the water
a couple of birds that are standing in the water

図39についての生成結果（パターン５）

a bird that is standing in the water
a group of ducks swimming in a body of water
a group of birds that are standing in the water
a group of ducks are swimming in the water
a close up of a bird on a beach