深層学習モデルから抽出した特徴ベクトルの画像検索精度と計算時間に関する評価について発表します．

1821005，吉岡拓郎です．よろしくお願いします．

研究背景です．

SNSの普及により写真や画像の投稿が盛んになっており，

大量の画像や写真が蓄積されています．

ユーザが目的の画像に辿り着くために画像検索機能の重要性が増しています．

本研究で重要となる特徴ベクトルについて先に説明させていただきます．

CNN (Convolutional Neural Network)の登場により，画像検索機能は向上した

深層学習モデルの中間層から抽出した特徴ベクトルを利用した画像検索方式を提案します．

ここで中間層から抽出される特徴ベクトルには，意味情報が保存されていると仮定します．意味情報とは，画像を認識する際に，その判断材料となる情報

例で示した画像は、5×5で25画素なので、特徴ベクトルの要素が25個。

この特徴ベクトルは、25次元であるといえる

関連研究については，1つ目は，AlexNetの構造についてです．

2つ目は，特徴ベクトルの抽出について関連されたことが書かれています．

3つ目は，特徴ベクトルの類似度を測る際の手段についてです．

4つ目は，次元の呪いについての関連研究です

次に，研究課題についてです．

特徴ベクトルは，画像認識において，特徴ベクトルが高次元になるほど検索精度が良くなるが計算時間が増加してしまいます．一方，低次元になるほど計算時間は早くなるが検索精度が落ちてしまうことがわかっています．

望ましい検索精度と計算時間を考慮した場合の最適な次元数が明らかになっていません．

次に，研究目的についてです．

本研究では，最適な次元数の特徴ベクトルの分析手法を提案します．

実験結果から，分析手法により，最適な次元数を得ることが可能であるか示します．

深層学習モデルの中間層の次元数を変化させて，異なる次元数の特徴ベクトルを抽出します．

下記の図のようにモデルの全結合層部分の値を変化させてそこから特徴ベクトルを抽出します．

検索精度は，画像検索手法の一つでもあるユークリッド距離を用いる．ベクトル間のユークリッド分離が小さい程類似性が高いとする

提案する分析手法についてです．

STEP-1で基準となるモデルを作成します．その後，STEP-2で中間層の次元数を変化させたモデルを複数作成します．

STEP-3で各モデルを学習させます．STEP-4でSTEP-3で作成した各深層学習モデルの中間層から特徴ベクトルを抽出します．

STEP-5で異なる次元数の各特徴ベクトルを画像検索評価プログラムに読込，評価を行います．

実験環境についてです．

モデルの学習，特徴ベクトルを抽出する際の画像集合に関して，CIFAR-10データセットを使用しました．

作成した深層学習モデルはこの7つになります．それぞれ識別層手前の全結合層の値を100,500,1000,2000,3000,4096,8192と変更しています．

モデルの構成はこのようになっています．13，15層目の次元数の値を変化させます．

特徴ベクトルの抽出は，15層目から行います．

実験1では，各次元数の特徴ベクトルの検索精度と計算時間についての評価を行います．

実験2では，各ラベルの正答率，検索結果上位に表示された画像の共通点を評価します．

実験1では，画像検索精度と計算時間の両方の観点から最も良い結果だった次元数を明確にすることを目的とする．

実験方法1についてです．

画像検索精度を調査するため，基準となる画像と同じラベルを数え，最も正答率の良い特徴ベクトルを求めます．画像検索をする際の計算時間を計測します．

実験1の結果です．

青い棒が正答率，左横に％が書いてあります．オレンジの線が計算時間，右横に時間が書いてあります．次元数8192の特徴ベクトルが最も良い検索精度となりました．

計算時間は，次元数が増えるのと比例して時間がかかるようになりました．

そのため次元数8192は，最も検索精度が良いが計算時間は，最もかかることが確認できました．

実験1の結果から各特徴ベクトルの正答率は決していいものではなかった．その理由が各ラベルの正答率に差があるからではないかと考えた．

そのことから実験2では，ラベルの検索精度の違いを調査することを目的とします．

また，検索結果上位に表示された画像について視覚的な共通点を評価する．特徴ベクトルの持つ意味情報について調査さすることも目的とします

実験2については，それぞれの特徴ベクトルから各ラベルの正答率を出し，ラベルによる正答率の違いをグラフに表します．

検索上位に表示された画像の共通点について評価を行います．

実験2の結果の各ラベルの正答率になります．

この図から車，トラックが正答率が良く，鳥，猫，鹿などの正答率が悪くなっていることがわかりました．正答率が悪いラベルの影響で全体の検索精度が落ちてしまっていることがわかりました．

ここでは，最も正答率の良かったラベル車の画像と最も悪かったラベル猫の画像の検索上位に表示された画像をランキング形式で3件を例として表示しています．

利用したデータセットの画像の解像度が低いことから、もともと画像がそこまで多くの意味情報を持っていない可能性が考えられました．

ただ、ランキング形式で画像を表示したところ，車などは直線などがはっきりしているために、解像度が少なくても意味情報を多く保持できていて、猫などは曲線や色の変化が多く、解像度の低さによってそれが表現されておらず、意味情報があまりないのではないかと考えました。

次元数8192が最も正答率が良かったのですが，計算時間がかかりすげていることから，次元数1000が，次元数，計算時間の両方の観点から最も良かった．本研究の分析手法から最適な次元数を導き出すことが出来ました．

次に，全体を通して検索精度が出ていなかったことに関し，その理由として，ラベルによって検索精度が異なり，検索精度が低いラベルが複数確認できた．

対象物がはっきり写っている画像は検索結果に類似度の高い画像が表示された

今後の展望として，本研究で行った提案手法を応用させるために，ユークリッド距離以外での類似性の評価，異なる深層学習モデル構造を使用した特徴ベクトル抽出を行うことで，異なる画像検索精度や得られる意味情報の調査をすることができるのではないかと考えています．

画像特徴ベクトルを画像検索に適用することで対象物の形や色といった面から画像を検索できると予想する．