

畳み込みニューラルネットワークを用いた アスペクト比歪み画像の修正

L^AT_EX Style file for manuscripts of JSAI 20XX

山根 佐介^{*1} 櫻井 隆平^{*2} 李 周浩^{*3}
Yamane Sasuke Sakurai Ryuhei Lee Joo-Ho

^{*1}立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University

^{*2*}立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University

We propose a method to correct the aspect ratio distortion of the image using the convolution neural network (CNN). The distortion of the aspect ratio means that a change of aspect ratio of an image by adding a stretching process. In the proposed method, we construct an aspect ratio estimation model using distorted images as input. Estimation of the aspect ratio based on image input can be regarded as a regression problem. This regression problem is modeled using CNN. This makes it possible to correct distorted images. In the performance evaluation experiment of the model trained using VOC2012, aspect ratio distortion correction of 100 test images was performed. Our method can accurately correct the distortion within 7.0% of stretch from original images on average. In terms of precision, 85% of distorted images are successfully corrected. As a result of the analytical experiment, it is assumed that the model uses the edge component of the image and performs the aspect ratio estimation.

1. はじめに

本研究でアスペクト比とは画像の拡張度合いのことをいう。画像にはその画像本来の正しいアスペクト比が存在する。正しいアスペクト比とは、画像内の被写体のアスペクト比が正しいものである。アスペクト比と画像の関係の例を図.1に示す。本論文ではアスペクト比の正常な画像を正常画像、アスペクト比の歪んだ画像を歪み画像とする。本研究でのアスペクト比の歪みとは、画像の縦横比を保たない引き伸ばし処理を加え、画像本来のアスペクト比と異なることをいう。アスペクト比の歪んだ画像では画像内の被写体が実世界では存在しない形状となるため、見る人に違和感や不快感を与える。そのため、本来のアスペクト比を維持する必要がある。しかし現在、誰もが画像編集ツールを利用する。そうした状況から誤ってアスペクト比の歪んだ画像を作成してしまう問題がある。

近年、画像のアスペクト比の歪みに着目した研究がなされている。[Li 16] らの研究は畳み込みニューラルネットワーク (CNN) を用い画像のアスペクト比の歪みを考慮した物体検出手法を提案している。また [Niu 10] らは人が歪みに気づかないような歪み画像の作成を目的とした研究を行っている。しかし、本研究の目的は画像のアスペクト比の歪みを正常なアスペクト比に修正することである。歪み画像を正常画像に戻すことを目的とした既存研究は存在していない。

本研究の貢献は、CNN を用いて入力画像のアスペクト比の歪みを推定できることを示すことである。



図 1: アスペクト比 1 の正常画像 (左図) とアスペクト比 2 の歪み画像 (右図) の例

2. CNN を用いたアスペクト比歪み修正手法

2.1 概要

本研究の目的はアスペクト比の歪んだ画像を正しいアスペクト比に修正することである。提案手法では、まず歪み画像のアスペクト比を推定する。その後、推定したアスペクト比を基に歪み画像を正常画像に修正する。そのため、歪み画像を入力、アスペクト比を出力としたアスペクト比推定モデルを構築する。画像入力に基づくアスペクト比の推定は回帰問題として捉えることができるため、この回帰問題を CNN を用いてモデル化する。これにより歪み画像のアスペクト比が推定できる。歪み画像のアスペクト比が分かれば正常画像への修正が可能である。図.2 に提案手法の処理の流れを示す。

CNN の訓練データにはアスペクト比情報がラベル付けされた画像を用いる。本手法では自動で正常画像から訓練データを作成しラベル付けを行う。

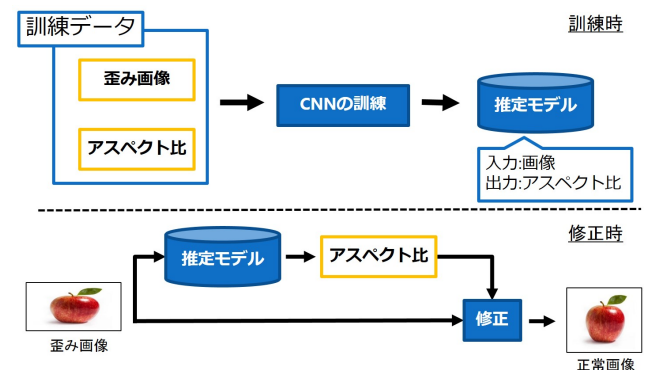


図 2: 提案手法の概要図。図上部:CNN を訓練し推定モデルを作成する処理。図下部:作成した推定モデルを使い歪み画像の修正処理。

2.2 本手法で扱うアスペクト比

本研究で扱うアスペクト比と対数アスペクト比について記述する。アスペクト比とは画像の拡張度合いのことをいう。画

連絡先: 山根佐介, 立命館大学大学院 情報理工学研究科 情報理工学専攻 AIS Lab., 〒525-8577 滋賀県草津市野路東 1-1-1 クリエーションコア 7 F, (077)561-5931, is0240he@ed.ritsumei.ac.jp

像の拡張度合いとは画像内の被写体が正しい形状をしている状態を基準としたその被写体の歪みの度合いのことをいう。そのため画像本来の正しいアスペクト比とは画像内の被写体が正しい形状をしている状態の拡張度合いである。本研究では正常画像のアスペクト比を $r = 1$ と定義する。また、正常よりも縦に伸びていれば $0 < r < 1$ 、横に伸びていれば $1 < r$ とする。アスペクト比と歪みの度合いの関係は式.1 の通りである。

$$\mathbf{x}'_{i,j} = \begin{cases} \mathbf{x}_{ri,j} & (0 < r < 1) \\ \mathbf{x}_{i,j} & (r = 1) \\ \mathbf{x}_{i,j/r} & (1 < r) \end{cases} \quad (1)$$

ここで、 r はアスペクト比である。 $\mathbf{x}'$ が歪み画像、 $\mathbf{x}$ が正常画像、 $\mathbf{x}_{i,j}$ は画像の輝度値、 (i, j) は画像内の座標である。本研究でのアスペクト比は画像の歪みの度合いであるため、アスペクト比は画像の拡張処理のみ変化し切り抜き処理では不変である。

次に、対数アスペクト比について記述する。対数アスペクト比は以下の式で作成される。

$$l = \log_e(r) \quad (2)$$

ここで、 l は対数アスペクト比、 r はアスペクト比である。対数アスペクト比を用いることで、歪みの向きは違うが歪みの度合いは同じである歪み画像らと正常画像との距離が同じとなる。

2.3 アスペクト比推定モデル

提案手法は、アスペクト比推定モデルを CNN を用いてモデル化する。アスペクト比推定モデルは入力された歪み画像の対数アスペクト比を出力する。画像入力に基づく実数値の出力は回帰問題であるため、モデルの損失関数は平均二乗誤差 (MSE) である。提案するアスペクト比推定モデルは 9 層の CNN である。CNN の構成は 1 から 7 層目が畳み込み層、8 層目が大域的最大値プーリング層、9 層目が全結合層である。活性化関数は ReLU 関数を採用する。活性化関数の入力に対して Batch Normalization を適用する。畳み込み層では、 1×1 サイズのバディングを行う。

2.4 訓練データ作成方法

訓練データにはアスペクト比情報がラベル付けされた画像を用いる。本手法では自動で正常画像から訓練データを作成しラベル付けを行う。正常画像には大規模画像データセットを用いる。

2.4.1 訓練画像の作成

訓練画像の作成方法を記述する。作成手順は以下のとおりである。

1. 訓練画像作成のための入力として正常画像を準備
2. ある区間内から一様な確率で対数アスペクト比 l を生成
3. 式.1 より $r = e^l$ である画像を作成

本研究では、 $[\log_e(3^{-1}), \log_e(3)]$ の範囲から対数アスペクト比を生成する。その理由として提案手法は人が作成する歪み画像の修正を目的としており、人が作成する歪み画像はアスペクト比が $[3^{-1}, 3]$ の範囲内であることが多いと考えたからである。

次に、作成した画像を固定サイズに拡張する。CNN の訓練データは固定サイズでなくてはならないからである。具体的な拡張方法について記述する。まず画像サイズが可変である画像の中心を正方形に切り出す。切り出すサイズは画像の短辺を一边とする正方形である。そして、画像から切り出された可変サイズの正方形画像を固定サイズに拡張する。拡張された固定サイズの画像を訓練データとする。提案手法では画像から切り出された可変サイズの正方形画像を 256×256 サイズに拡張する。画像を固定サイズに拡張する工程を図.3 に示す。

2.4.2 正解ラベルの作成

訓練データの正解ラベルについて記述する。正解ラベルは歪み画像の対数アスペクト比とする。正解ラベルは歪み画像が生成されるたびに作成される。

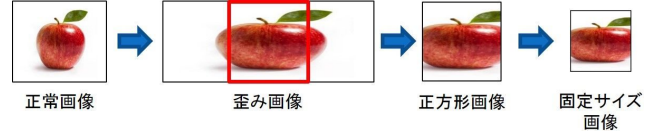


図 3: 画像を固定サイズに拡張する工程

表 1: 使用したコンピュータの仕様

OS	Windows 10 (64 bit)
CPU	Intel Core i7-6700 Processor (8M Cache, 4.00 GHz)
メモリ	16 GB
GPU	NVIDIA GeForce GTX TITAN X (12 GB)

3. 実験

提案手法のアスペクト比推定精度の検証実験とアスペクト比推定モデルの分析実験について記述する。本実験では前章で記述したアスペクト比推定モデルを、PASCAL VOC2012[Everingham] の画像データを用いて訓練した。VOC2012 は一般物体認識タスクに用いられる大規模画像データセットである。データセットの画像データはすべて正常画像である。評価実験では、アスペクト比推定モデルの推定結果を定量的に評価する。分析実験では、アスペクト比の推定に用いられる画像特徴を分析する。今回使用したコンピュータの構成を表.1 に示す。なお、本研究で使用された実験用スクリプトは https://github.com/YamaneSasuke/correct_aspect_ratio から入手できる。

3.1 実験準備

実験を行うにあたり、まずデータセットの画像データ 17100 枚のうち、17000 枚を訓練用、100 枚をテスト用とした。また 17000 枚の訓練用画像のうち、16500 枚をモデルの訓練に、500 枚を検証に使用した。それらの画像は重複しない。検証とはモデルの訓練時に毎エポック、モデルの損失を訓練データ以外のデータで確認することである。この検証結果の最も良かったエポック時のモデルを最終的な訓練済みモデルとして採用する。エポックごとに訓練用画像 17000 枚から歪み画像を作成する。そのため、エポックごとにアスペクト比の異なる歪み画像を使いモデルの訓練を行う。ハイパーパラメータは、学習率 0.01、バッチサイズは 100、エポック数は 1000 である。最適化手法には ADAM を使用した。訓練データの拡張は CNN の性能向上に重要である。データの拡張とは、訓練データ画像に対し適切な画像処理を行い訓練に使用する画像データを増加させる処理である。提案手法では、訓練データに対し、ランダムクロップ処理とランダムフリップ処理を行いデータの拡張を行った。モデルの訓練時間は、108264 秒 (約 30 時間) である。

3.2 アスペクト比推定精度の評価実験

3.2.1 実験方法

提案手法のアスペクト比推定精度を検証する。定量評価として、提案手法の推定結果と正解ラベルの誤差を評価した。ここでいう誤差はモデルによる対数アスペクト比の推定値 $y = CNN(\mathbf{x})$ と真の対数アスペクト比 l の差である。

$$E = l - y \quad (3)$$

まず、テスト用画像から歪み画像を作成しモデルへ入力する。歪み画像は画像ごとに一様な確率でランダム生成した対数アスペクト比から作成される。その入力から得られた推定値と正解ラベルの誤差を評価した。

評価するにあたり人間のアスペクト比歪み修正能力を調査した。調査方法は人間に対しアスペクト比の異なる歪み画像 20 枚を正常画像に修正させ、その誤差の絶対値を計測した。被験者は 15 名である。その結果、人間の誤差の絶対値の平均は $\log_e(1.1303)$ であった。そのため、 $|\log_e(1.1303)|$ を推定成功とみなす閾値として設定した。誤差の絶対値が $\log_e(1.1303)$ 以下の場合に推定成功とみなす。

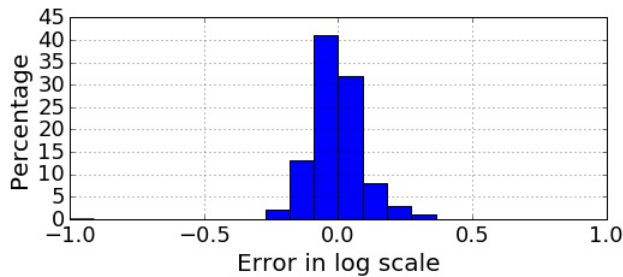


図 4: 100 枚の歪み画像の推定誤差ごとの分布。横軸は誤差、縦軸は割合である。

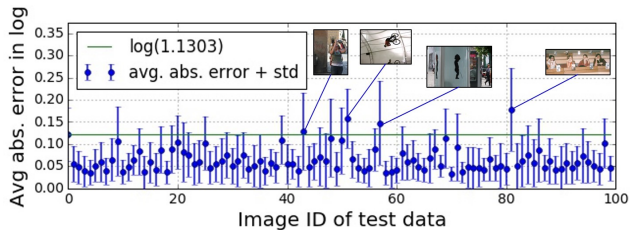


図 5: テスト用画像 1 枚から作成された 21 枚の歪み画像の誤差の絶対値の平均をテスト用画像ごとに表した結果。横軸はテスト用画像、縦軸は誤差の絶対値の平均である。テスト用画像は 100 枚のため横軸は $[0, 100]$ である。緑色の直線のグラフは推定成功とみなす閾値を表している。

3.2.2 実験結果

結果を図 4 に示す。図の横軸は誤差、縦軸は割合を表している。誤差の絶対値の平均値は $\log_e(1.0694)$ である。また 100 枚中 85 枚のテスト用画像で誤差の絶対値が閾値 $\log_e(1.1303)$ 以下であった。

次に、100 枚のテスト用画像 1 枚につき 21 種類のアスペクト比の歪み画像を作成した。21 種類のアスペクト比は $[\log_e(3^{-1}), \log_e(3)]$ までも 21 分割した値である。その計 2100 枚の画像を入力とし、推定値と正解ラベルとの誤差を求めた。

同じテスト用画像から作成された歪み画像 21 枚の誤差の絶対値の平均値をテスト用画像ごとに表したものが図 5 である。横軸がテスト用画像番号、縦軸が 21 枚の歪み画像の誤差の絶対値の平均値である。100 枚中 96 枚のテスト用画像の誤差の絶対値の平均が推定成功とみなす閾値 $\log_e(1.1303)$ 以下であった。同じアスペクト比の歪み画像 100 枚の誤差の絶対値の平均をアスペクト比ごとに表したものが図 6 である。横軸が 21 種類のアスペクト比、縦軸がアスペクト比ごとの 100 枚の歪み画像の誤差の絶対値の平均である。

3.2.3 考察

定量評価においては、図 4 より提案したモデルは誤差の絶対値の平均は $\log_e(1.0694)$ であった。またテスト用画像のうち 85% が推定成功とみなす誤差以下であった。この結果より、提案したモデルがアスペクト比歪み修正に有用であることが分かる。図 5 よりほかのテスト用画像と比べ誤差が大きい画像が存在することが分かる。これらの画像を確認したところ特定方向のエッジが多いことが判明した。そのため、アスペクト比推定において画像のエッジ成分が重要であると推測できる。図 6 よりアスペクト比の値に依らず、すべての種類のアスペクト比の画像において誤差が推定成功とみなす閾値以下である。この結果より、提案手法がアスペクト比の歪みが大きい画像の修正、小さい画像の修正ともに有用であることが分かる。

3.3 アスペクト比推定に用いられる画像特徴の分析実験

本実験は、提案手法の推定値とその入力画像のエッジ成分との関係を分析することが目的である。評価実験の結果から、

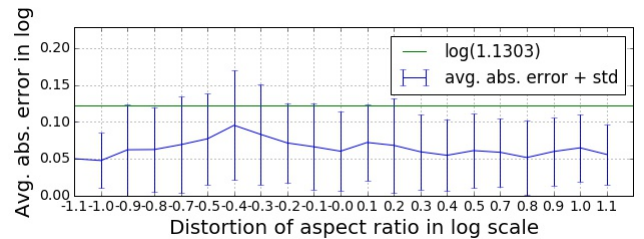


図 6: 同じアスペクト比の歪み画像の誤差の絶対値の平均をアスペクト比ごとに表した結果。横軸はアスペクト比、縦軸は誤差の絶対値の平均である。アスペクト比は $[\log_e(3^{-1}), \log_e(3)]$ までも 21 段階あるため横軸は $[\log_e(3^{-1}), \log_e(3)]$ である。緑色の直線のグラフは推定成功とみなす閾値を表している。

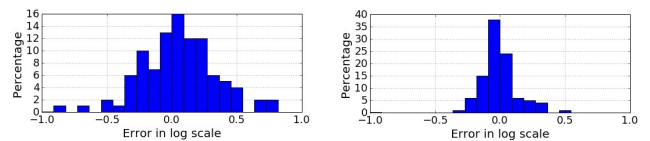


図 7: ボカシ画像の推定誤差 (左図) とエッジ画像の推定誤差 (右図)

画像のエッジ成分がアスペクト比推定に重要ではないかと推測した。画像のエッジ成分とは画像の高周波成分のことである。本実験では、その推測を検証する。実験に用いるモデルは評価実験に用いたものと同じである。

3.3.1 実験方法

実験にはテスト用画像 100 枚を使用した。テスト用画像に対し、画像処理を施しエッジ成分を除去したボカシ画像とエッジ成分のみを抽出したエッジ画像をそれぞれ 100 枚作成した。それらの画像から歪み画像を作成し入力画像として推定結果の誤差を調査した。歪み画像は画像ごとに様な確率でランダム生成した対数アスペクト比から作成される。また、エッジの方向と量の推定結果に対する影響を検証するため、正常画像に対し徐々に特定方向のエッジを足しその推定結果の推移を調査した。エッジを足すとは画像の水平方向、もしくは垂直方向に対して輝度値 0 の幅 2 ピクセルの直線を描画する処理のことである。

3.3.2 実験結果

ボカシ画像をもとに作成した歪み画像を入力とした際の誤差を図 7 左図に示す。100 枚の入力画像の推定誤差の絶対値の平均は $\log_e(1.2471)$ である。また 100 枚中 37 枚のテスト用画像で誤差の絶対値が閾値 $\log_e(1.1303)$ 以下であった。エッジ画像をもとに作成した歪み画像を入力とした際の誤差を図 7 右図に示す。100 枚の入力画像の推定誤差の絶対値の平均は $\log_e(1.1019)$ である。また 100 枚中 71 枚のテスト用画像で誤差の絶対値が閾値 $\log_e(1.1303)$ 以下であった。

エッジの方向と量による推定値の推移を図 8 図上に示す。図上の画像を入力とした時の推定値が画像下にある数値である。この結果から特定方向のエッジを足すことで推定値に変化があることがわかる。また、エッジの量も推定値と関係があると推測できる。

3.3.3 考察

図 7 左図と図 7 右図の結果から提案手法は画像のエッジ成分を用いてアスペクト比を推定していることが分かる。ボカシ画像の推定誤差平均が $\log_e(1.2471)$ であるのに対し、エッジ画像の推定誤差平均は $\log_e(1.1019)$ である。図 8 図上の正常画像を入力した際の推定値は $\log_e(1.0222^{-1})$ である。水平方向のエッジを一つ足した際の推定値は $\log_e(1.0584)$ であり、エッジを足す前と比べ推定値が大きくなった。この意味は、推定モデルが水平方向のエッジが増加した画像の方が水平方向に引き

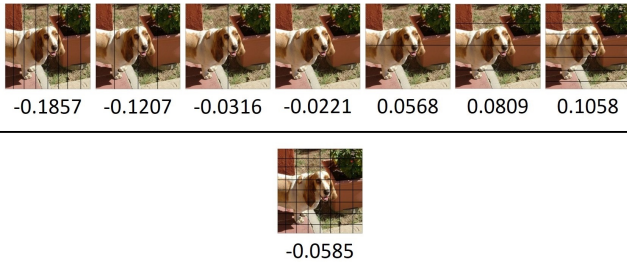


図 8: 図上:エッジの向きと量による推定値の推移。図下:水平方向と垂直方向のエッジを同等に足した場合の推定値

伸ばされた歪み画像であると判断したということである。逆に垂直方向のエッジを一つ足した際の推定値は $\log_e(1.0321^{-1})$ であり、エッジを足す前と比べ推定値が小さくなった。この意味は、推定モデルが垂直方向のエッジが増加した画像の方が垂直方向に引き伸ばされた歪み画像であると判断したということである。この結果からエッジの方向がアスペクト比の歪みの方向と関係していると推測される。加えて、エッジの量が推定値の大小に関係することがわかる。また、図.8 図下の結果より水平方向と垂直方向のエッジを同等に足した場合は正常画像に近い推定値が出力された。これらの結果から、水平方向と垂直方向のエッジの量の割合がアスペクト比の推定に関係していると考えられる。

3.4 歪み画像修正結果

提案手法を用いた歪み画像の修正結果を示す。誤差の絶対値が $\log_e(1.1303)$ 以下である推定成功結果を図.9 に、誤差の絶対値が $\log_e(1.1303)$ 以上である推定失敗結果を図.10 に示す。図内左列が入力した歪み画像、図内中央列が修正結果、図内右列が正解となる正常画像である。

4. まとめ

本研究では、CNN を用いた歪み画像のアスペクト比修正手法を提案した。提案手法では、画像を入力、アスペクト比を出力としたアスペクト比推定モデルを CNN を用いて構築した。訓練データには歪み画像を使用した。VOC2012 をデータセットとして使用したモデルの性能評価実験では、100 枚のテスト用画像のアスペクト比歪み修正を行った。提案手法の推定値の誤差平均は $\log_e(1.0694)$ であった。評価実験においての 85% の歪み画像を修正することに成功した。分析実験の結果、モデルは画像のエッジ成分を使用しアスペクト比推定を行っていると考えられる。

4.1 今後の課題

本研究で、提案したモデルがアスペクト比歪み修正において有用であると証明され、画像の高周波成分であるエッジ情報を用いてアスペクト比の推定を行っているという推測がなされた。この推測をもとに推定モデルの分析をより進めることで、画像のアスペクト比の歪みに関係のある特徴の解明につながると考えられる。特徴の解明は推定モデルの性能を向上させるうえで重要な課題である。

参考文献

[Everingham] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A.: The PASCAL Visual Object Classes Challenge 2012 (VOC2012)

[Li 16] Li, B., Wu, T., Shao, S., Zhang, L., and Chu, R.: Object Detection via End-to-End Integration of Aspect Ratio and Context Aware Part-based Models and Fully Convolutional Networks, *CoRR*, Vol. abs/1612.00534, (2016)

[Niu 10] Niu, Y., Liu, F., Li, X., Bao, H., and Gleicher, M.: Detection of Image Stretching, in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, pp. 93–100, New York, NY, USA (2010), ACM



図 9: 修正結果。誤差が閾値以下の推定成功例。左列が入力した歪み画像、中央列が修正結果、右列が正常画像。



図 10: 修正結果。誤差が閾値以上の推定失敗例。左列が入力した歪み画像、中央列が修正結果、右列が正常画像。