

A Survey on Image Data Augmentation for Deep Learning

著者

Connor Shorten, Taghi M. Khoshgoftaar

Abstract

ディープ畳み込みニューラルネットワークは、多くのコンピュータビジョンタスクにおいて非常に優れた性能を発揮しています。しかし、これらのネットワークは、オーバーフィットを避けるためにビッグデータに大きく依存しています。**オーバーフィットとは、ネットワークが学習データを完全にモデル化するために、非常に高い分散を持つ関数を学習してしまう現象のことです。**しかし、残念ながら、医療画像解析など、多くのアプリケーション領域ではビッグデータを利用できないことが多いのが現状です。本調査では、限られたデータの問題をデータ空間で解決するData Augmentationに焦点を当てています。**データオーグメンテーションは、トレーニングデータセットのサイズと品質を向上させ、それらを用いてより優れたディープラーニングモデルを構築できるようにする一連の技術を含んでいます。**

本調査では、幾何学的変換、色空間の拡張、カーネルフィルタ、画像の混合、ランダム消去、特徴空間の拡張、敵対的トレーニング、生成的敵対ネットワーク、ニューラルスタイル転送、メタ学習などの画像拡張アルゴリズムを取り上げています。本調査では、GANに基づく増強手法の応用を中心に取
り上げている。本論文では、オーグメンテーション技術に加えて、テスト時間の増大、解像度の影
響、最終データセットのサイズ、カリキュラム学習など、データオーグメンテーションの他の特徴に
ついて簡単に議論する。本調査では、データオーグメンテーションのための既存の手法、有望な開
発、データオーグメンテーションを実施するためのメタレベルの決定を提示する。読者は、データ
オーグメンテーションがどのようにモデルのパフォーマンスを向上させ、限られたデータセットを拡張
してビッグデータの能力を活用することができるかを理解することができるでしょう。

Introduction

ディープラーニングモデルは、識別タスクにおいて信じられないほどの進歩を遂げてきました。これは、ディープネットワークアーキテクチャの進歩、強力な計算能力、ビッグデータへのアクセスによって促進されてきました。ディープニューラルネットワークは、畳み込みニューラルネットワーク（CNN）の開発により、画像分類、物体検出、画像分離などのコンピュータビジョンタスクへの応用に成功しています。これらのニューラルネットワークは、画像の空間特性を保持するパラメータ化された疎結合のカーネルを利用しています。畳み込み層は、画像の空間分解能を順次ダウンサンプルしながら、特徴マップの深さを拡大していきます。この一連の畳み込み変換により、手作業で作成されたものよりもはるかに低次元で、より有用な画像の再表現を作成することができる。CNNの成功は、ディープラーニングをコンピュータビジョンのタスクに適用することへの関心と楽観的な考えに拍車をかけています。

深層畳み込みネットワークをコンピュータビジョンのタスクに適用することで、現在のベンチマークを向上させようとする研究は、多くの分野で行われています。**これらのモデルの一般化能力を向上させることは、最も困難な課題の一つである。一般化可能性とは、以前に見たデータ (訓練データ) と見たことのないデータ (テストデータ) で評価したときのモデルの性能差のことを指します。一般化可能性が低いモデルは、訓練データにオーバーフィットしています。**オーバーフィットを発見する1つの方法は、トレーニング中の各エポックでのトレーニングと検証の精度をプロットすることです。下のグラフは、学習エポックの精度を可視化したときのオーバーフィットの様子を示しています[図1]。

Signs of Overfitting (Training Error and Testing Error)



Desired convergence of Training and Testing Error

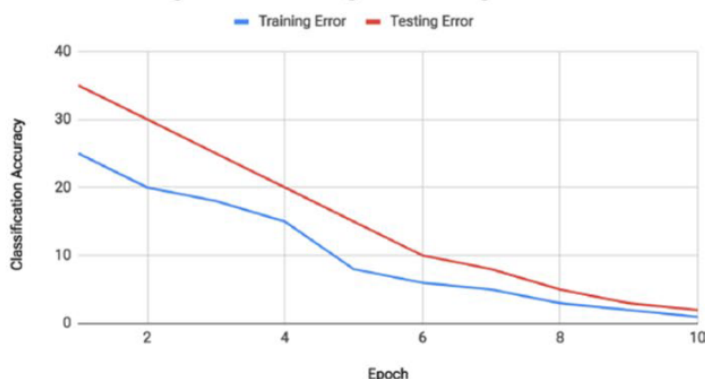


図1: 左側のプロットは、トレーニング率が減少し続けると検証エラーが増加し始める変曲点を示しています。トレーニングの増加により、モデルはトレーニング・データにオーバーフィットし、トレーニング・セットと比較してテスト・セットでのパフォーマンスが悪くなっています。対照的に、右側のプロットは、トレーニングとテストの誤差の間の望ましい関係を持つモデルを示しています。

有用なディープラーニングモデルを構築するためには、検証誤差が訓練誤差とともに減少し続けなければなりません。データオーグメンテーションは、これを達成するための非常に強力な方法です。拡

張されたデータは、可能なデータポイントのより包括的なセットを表すので、トレーニングセットと検証セット、および将来のテストセット間の距離を最小限に抑えることができます。

本調査の焦点である**データオーグメンテーションは、オーバーフィッティングを減らすために開発された唯一の手法ではありません。** 以下の数段落では、ディープラーニングモデルのオーバーフィッティングを回避するために利用可能な他の解決策を紹介します。このリストは、読者にデータオーグメンテーションの文脈をより広く理解してもらうことを目的としています。

一般化性能を向上させるための他の多くの戦略は、モデルのアーキテクチャ自体に焦点を当てています。これにより、AlexNet [1]からVGG-16 [2]、ResNet [3]、Inception-V3 [4]、DenseNet [5]に至るまで、より複雑なアーキテクチャが次々と開発されてきました。ドロップアウト正則化、バッチ正規化、転移学習、事前学習などの機能的なソリューションは、より小さなデータセットへの適用のためにディープラーニングを拡張するために開発されてきました。これらのオーバーフィッティングの解決策の簡単な説明を以下に示す。ディープラーニングにおける正則化手法の完全な調査は、Kukackaら[6]によってまとめられています。これらのオーバーフィッティングの解決法についての知識は、読者に他の既存のツールについての情報を与え、データ拡張とディープラーニングの高レベルのコンテキストをフレーミングすることになります。

- ドロップアウト[7]は、訓練中に実行されて選択されたニューロンの活性化値をゼロにする正則化手法である。この制約により、ネットワーク内の少数のニューロンの予測能力に頼るのではなく、よりロバストな特徴を学習するようにネットワークを強制する。Tompsonら[8]は、この考えを、個々のニューロンではなく特徴マップ全体をドロップアウトする「空間ドロップアウト」を持つ畳み込みネットワークに拡張した。
- バッチ正規化 [9] は、レイヤ内の活性化のセットを正規化する別の正規化技術です。正規化は、各活性化からバッチ平均を減算し、バッチ標準偏差で除算することで機能します。この正規化技術は、標準化とともに、ピクセル値の前処理における標準的な技術です。
- 転移学習 [10, 11] は、オーバーフィットを防ぐためのもう一つの興味深いパラダイムです。転移学習は、ImageNet [12]のような大規模なデータセット上でネットワークを学習し、それらの重みを新しい分類タスクの初期重みとして使用することで動作します。一般的には、完全に接続された層を含むネットワーク全体をコピーするのではなく、畳み込み層の重みだけをコピーします。多くの画像データセットは、ビッグデータでよりよく学習される低レベルの空間特性を共有しているため、この方法は非常に効果的です。転移されたデータ領域間の関係を理解することは、現在進行中の研究課題である[13]。Yosinskiら[14]は、転移可能性が、主に高次の層のニューロンの特殊化と共適応ニューロンの分割の難しさによって負の影響を受けることを発見している。
- 事前学習[15]は、概念的には転移学習と非常によく似ている。プレトレーニングでは、ネットワークアーキテクチャを定義し、ImageNet [12]のような大規模なデータセットで学習する。これが転移学習と異なる点は、VGG-16 [2]やResNet [3]のようなネットワークアーキテクチャを、重みと同様に転移しなければならない点である。事前学習により、大規模データセットを用いた重み

の初期化が可能になる一方で、ネットワークアーキテクチャの設計に柔軟性を持たせることが可能になる。

- ワンショット学習とゼロショット学習 [16, 17] アルゴリズムは、非常に限られたデータでモデルを構築するための別のパラダイムである。ワンショット学習は、顔認識アプリケーションで一般的に使用されている[18]。ワンショット学習へのアプローチとしては、ネットワークが1つまたは数個のインスタンスでしか訓練されていなくても画像分類が可能なような距離関数を学習するシヤムネットワーク[19]を使用する方法がある。もう一つの非常に人気のあるワンショット学習のアプローチは、メモリ拡張ネットワークス[20]の使用である。ゼロショット学習はより極端なパラダイムで、ネットワークがWord2Vec [21]やGloVe [22]のような入力と出力のベクトル埋め込みを使用して、記述的属性に基づいて画像进行分类するものである。

上記の手法とは対照的に、**データオーグメンテーションは、問題の根源である学習データセットからオーバーフィットにアプローチします。これは、オーグメンテーションによって元のデータセットからより多くの情報を抽出できるという前提の下で行われる。**これらの拡張は、データの反りやオーバーサンプリングによって、学習データセットのサイズを人為的に増大させます。データ拡張は、既存の画像をラベルが保存されるように変換します。これには、幾何学的変換や色変換、ランダム消去、敵対的トレーニング、ニューラルスタイル転送などの拡張が含まれます。オーバーサンプリング拡張では、合成インスタンスを作成し、それをトレーニングセットに追加します。これには、画像の混合、特徴空間の拡張、ジェネラティブな敵対的ネットワーク（GAN）が含まれます。オーバーサンプリングとデータ・ワープは、互いに排他的な二分法ではありません。例えば、GANサンプルは、データセットをさらに膨らませるために、ランダムクロッピングと重ねることができます。最終的なデータセットのサイズ、テスト時間の増大、カリキュラム学習、解像度の影響に関する決定は、この調査では「画像データ増大の設計上の考慮事項」のセクションで取り上げています。個々の補強技術の説明は、「画像データ補強の技術」のセクションで列挙します。[図 2]にデータオーグメンテーションの簡単な分類を示します。



図2：図中の色付きの線は、対応するメタ学習スキームがどのデータ増強手法を使用しているかを示しており、例えば、ニューラル・スタイル・トランスファーを用いたメタ学習はニューラル増強[36]でカバーされています。

画像増強技術について議論する前に、問題の背景を整理し、そもそも画像認識を困難なものにしている理由を考えることが有用です。猫と犬のような古典的な識別例では、画像認識ソフトウェアは、視点、照明、オクルージョン、背景、スケールなどの問題を克服しなければなりません。**データオーグメンテーションの課題は、これらの並進不変性をデータセットに焼き込むことで、結果として得られるモデルがこれらの課題にもかかわらず良好なパフォーマンスを発揮するようにすることです。**

データセットが大きいほど、より良いディープラーニングモデルが得られるというのは、一般的に受け入れられている考え方です[23, 24]。しかし、膨大なデータセットを組み立てるのは、データの収集とラベル付けの手作業のため、非常に困難な作業になる可能性があります。限られたデータセットは、医療画像解析において特に一般的な課題です。ビッグデータを考慮すると、深層畳み込みネット

ワークは、Esteveら[25]によって実証されたように、皮膚病変の分類などの医用画像解析タスクに対して非常に強力であることが示されています。このことから、肝臓病変の分類、脳スキャンの解析、皮膚病変の分類などの医用画像解析タスク[26]でのCNNの使用に触発され、継続的な研究が行われています。研究されている画像の多くはコンピュータ断層撮影（CT）や磁気共鳴画像（MRI）スキャンから得られたものであるが、どちらも収集には高価で手間がかかる。疾患の希少性、患者のプライバシー、ラベリングには医療専門家が必要であること、医療画像処理に必要な費用と手作業が必要であることなどから、大規模な医用画像データセットを構築することは非常に困難である。これらの障害により、医用画像分類の応用の観点から、画像データオーグメンテーション、特にGANベースのオーバーサンプリングに関する多くの研究が行われてきました。

データオーグメンテーションの有効性に関する多くの研究では、ベンチマークのために人気のある学術的な画像データセットを利用しています。これらのデータセットには、MNISTの手書き数字認識、CIFAR-10/100、ImageNet、tiny-imagenet-200、SVHN（ストリートビューハウスナンバー）、Caltech-101/256、MIT places、MIT-Adobe 5K dataset、Pascal VOC、Stan-ford Carsなどがあります。最もよく議論されているデータセットは、CIFAR-10、CIFAR-100、ImageNetである。オープンソースのデータセットの拡大により、研究者はデータオーグメンテーション技術の性能結果を比較するための様々なケースを手に入れることができます。ImageNetのようなこれらのデータセットのほとんどはビッグデータに分類されます。多くの実験では、限られたデータ問題をシミュレートするために、データセットのサブセットに自分自身を制約しています。

限られたデータセットに焦点を当てることに加えて、クラスの不均衡の問題と、Data Augmentationがどのようにして有用なオーバーサンプリングの解決策になるかについても検討します。 **クラス不均衡とは、多数派と少数派のサンプルの比率が偏っているデータセットのことです。** Leevyら[27]は、データタイプ間の高クラス不均衡に対する既存の解決策の多くを記述しています。我々の調査では、画像データにおけるクラスバランスのオーバーサンプリングが、どのようにしてデータオーグメンテーションを用いて行われるかを示します。

ディープラーニングやニューラルネットワークモデルの多くの側面は、人間の知性との比較を描いています。例えば、転移学習の人間の知性の逸話は、音楽の学習で例示されています。2人の人がギターの弾き方を学ぼうとしていて、1人がすでにピアノの弾き方を知っている場合、ピアノを弾く人の方が早くギターを弾けるようになる可能性が高いと思われます。音楽学習と同様に、ImageNet画像を分類できるモデルは、ランダムな重みを持つモデルよりも、CIFAR-10画像上でより良いパフォーマンスを発揮する可能性が高い。

データオーグメンテーションは想像力や夢想到に似ている。人間は経験に基づいて異なるシナリオを想像します。想像力は、私たちの世界をよりよく理解するのに役立ちます。GANやNeural Style Transferのようなデータオーグメンテーションの手法は、画像をよりよく理解できるように画像の改変を「想像」することができます。本論文の残りの部分は以下のように構成されている。読者にデータ拡張とディープラーニングの歴史的背景を与えるために、簡単な「背景」が提供されています。

「画像データ拡張技術」では、実験結果とともに各画像拡張技術について詳細に説明しています。"Design considerations for image Data Augmentation"では、テスト時間の増大や画像解像度の影響など、増大処理の追加的な特徴について議論しています。本論文は、事前に提出された資料の「ディスカッション」、「今後の課題」、「結論」で締めくくられています。

2. 背景

データ・ワーピングという形での画像拡張はLeNet-5[28]に見られます。これはCNNを手書き文字の分類に応用した最初の例のひとつである。データオーグメンテーションはオーバーサンプリングのアプリケーションでも研究されている。**オーバーサンプリングとは、不均衡なクラス分布を再サンプリングして、モデルがインスタンスを大多数のクラスタイプとしてラベル付けすることに過度に偏らないようにする技術である。** ランダムオーバーサンプリング (ROS) は、望ましいクラス比率が得られるまで、少数派クラスからランダムに画像を複製する素朴な手法です。インテリジェントなオーバーサンプリング技術は、Chawlaら[29]によって開発されたSMOTE (Synthetic Minority Over-sampling Technique) に遡ります。SMOTEとBorderline-SMOTE[30]の拡張版は、k-Nearest Neighborsを介して既存のインスタンスから新しいポイントを補間することで、新しいインスタンスを作成します。この技術の主な目的は、クラスの不均衡による問題を緩和することであり、SMOTEは主に表形式およびベクトルデータに使用されました。

Krizhevskyら[1]によって開発されたAlexNet CNNアーキテクチャは、ImageNetデータセットに畳み込みネットワークを適用することで、画像分類に革命をもたらしました。**彼らの実験では、データセットのサイズを2048倍に拡大するためにデータ増強が使われています。**これは、元の画像から224×224のパッチをランダムに切り取り、水平方向にフィッティングし、PCAカラーオーグメンテーションを用いてRGBチャンネルの強度を変更することで行われました。このデータ拡張により、ディープニューラルネットワークを学習する際の過剰合を減らすことができた。その結果、モデルのエラー率が1%以上減少したという。

それ以降、2014年にGANが登場し[31]、2015年にはNeural Style Transfer[32]、2017年にはNeural Architecture Search (NAS) [33]が発表されました。DCGAN、CycleGAN、Progressively-Growing GAN[34]といったGANの拡張については、それぞれ2015年、2017年、2017年に様々な作品が発表されました。Neural Style Transferは、2016年にJohnsonらがPerceptual Lossesを開発したことで、スピードアップしました[35]。NASからのメタ学習の概念をData Augmentationに適用することは、Neural Augmentation [36]、Smart Augmentation [37]、Auto-Augment [38]といった作品がそれぞれ2017年、2017年、2018年に発表され、ますます盛んになっています。

Deep Learningを医用画像に適用することは、2012年にCNNが普及して以来、人気のあるアプリケーションです。2017年にEstevaら[25]が皮膚科医レベルの皮膚がん検出を実証したことで、Deep Learningと医用画像はますます注目を集めました。

医用画像処理におけるGANの使用については、Yiらによる調査でよく知られています[39]。この調査では、CTノイズ除去[40]、加速された磁気共鳴画像[41]、PETノイズ除去[42]などの再構成におけるGANの使用、および網膜血管系セグメンテーションにおける超解像GANのアプリケーション[43]をカバーしています。さらに、Yiら[39]は、脳のMRI合成[44, 45]、肺がん診断[46]、高解像度の皮膚病変の合成[47]、胸部X線異常分類[48]などの医療画像アプリケーションにおけるGAN画像合成の利用を取り上げています。GANベースの画像合成Data Augmentationは、2018年にFrid-Adarら[49]が肝臓病変のclassificationに使用しました。Tisは、古典的なオーグメンテーションを用いた場合の感度78.6%、特異度88.4%から、GANベースのData Augmentationを用いた場合には感度85.7%、特異度92.4%と分類性能を向上させた。

今回取り上げた拡張機能の多くは、画像認識モデルの改良に焦点を当てています。画像認識とは、入力画像から「犬」や「猫」などの出力ラベルを予測するモデルです。

しかし、画像認識の結果を、YOLO[50]、R-CNN[51]、Fast R-CNN[52]、Fast R-CNN[53]などのアルゴリズムによる物体検出や、U-Net[55]などのアルゴリズムによるセマンティック・セグメンテーション[54]など、他のComputer Visionタスクに拡張することは可能です。

3. Image Data Augmentation techniques

データオーグメンテーションの有効性を最初に示したのは、水平方向のフィッティング、色空間のオーグメンテーション、ランダムクロッピングなどの単純な変換でした。これらの変換は、画像認識タスクの課題となっている前述の不変性の多くを符号化します。ここでは、幾何学的変換、色空間変換、カーネルコンバータ、画像の混合、ランダム消去、特徴空間の拡張、敵対的学習、GANベースの拡張、ニューラル・スタイル・トランスファー、メタ学習などの拡張方法を紹介します。このセクションでは、各拡張アルゴリズムがどのように機能するかを説明し、実験結果を報告し、拡張技術の短所について議論する。

3.1. 基本的な画像操作に基づくデータ拡張 幾何学的変換

このセクションでは、幾何学的な変換やその他多くの画像処理機能に基づいたさまざまな拡張機能について説明します。以下に説明する拡張機能のクラスは、その実装の容易さによって特徴付けられる。これらの変換を理解することは、データ拡張技術のさらなる研究のための有用な基盤となるでしょう。

また、異なる幾何学的なオーグメンテーションについて、その適用の「安全性」の観点から説明する。データ拡張法の安全性とは、変換後にラベルを保持する可能性を意味します。例えば、猫と犬のようなImageNetの課題では、rotationやfipsは一般的に安全ですが、6と9のような数字認識の課題では

安全ではありません。ラベルを保存しない変換は、予測に自信がないことを示す応答を出力するモデルの能力を強化する可能性があります。しかし、これを実現するためには、オーグメンテーション後のラベルをリファインドする必要がある[56]。もし、非ラベル保存変換後の画像のラベルが[0.5 0.5]のようなものであれば、モデルはよりロバストなコンフデンシー予測を学習することができます。しかし、安全でないData Augmentationごとにrefinedラベルを構築するのは、計算量が多いプロセスです。

オーグメンテーション後のデータにリファインドされたラベルを構築することは困難であるため、オーグメンテーションの「安全性」を考慮することが重要である。この安全性はドメインに依存しており、一般化可能な拡張ポリシーを開発する上での課題となっている（一般化可能な拡張を見つけるためのさらなる調査についてはAutoAugment [38]を参照）。Tereは、ある歪みの大きさをラベルを変更する変換にならない画像処理関数はありません。これは、データに応じた拡張機能の設計と、一般化可能な拡張ポリシーの開発という課題を示すものである。このことは、以下に挙げる幾何学的な拡張機能に関して重要な考慮点となります。