

AI時代に 画像の知識が必要な理由

奥村 明弘

人工知能と上手に付き合うには

● AIを学習させるだけのエンジニアでいいのか？

ディープ・ラーニングなどのAIが話題になっています。開発現場では「猫も杓子も」とエンジニアが駆り出され、強力なコンピュータやクラウドなどを使って、マシン・パワーに任せて学習を繰り返しています。しかし、その学習に使っているアルゴリズムは他社製のブラックボックスで、駆り出されたエンジニアはただただ学習素材を作って学習を繰り返しているだけです(図1)。これは技術的な仕事とは言えず、単にコンピュータの奴隷となっているだけだと思います。皆さんや皆さんの職場はそうになっていませんか。

人工知能と上手に付き合っていくには2つのコツがあります。

- ① 100点を狙わない
- ② AIとの知恵比べに勝つ(図2)

● ① 100点を狙わないアプローチ

100点を狙わずビジネスとして成功している例とし

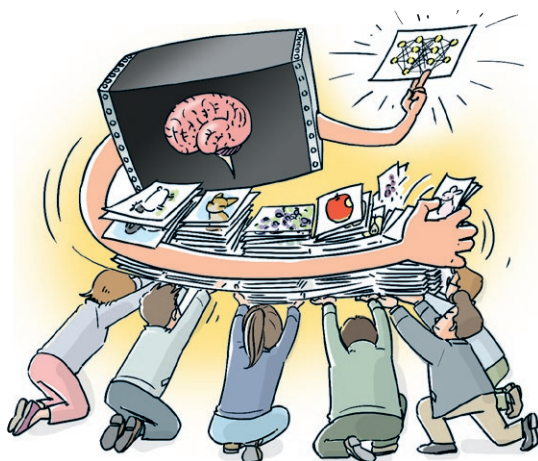


図1 AIを提供されたフレームワークで学習させるだけのエンジニアでいいか？

ては、CMでも有名なSansanの名刺管理サービスや、inahoの自動野菜収穫ロボット(写真1)などが挙げられます。

Sansanの名刺管理サービスはクラウドによるものですが、実は印刷された文字ですらAIは100%の認識ができていないのです。Sansanのすごいところは、ローテクなのですが、サーバの裏にオペレータがいて、AIがうまく読み取れない文字は人手で修正するのです。これで読み取り率100%を実現しています。

同じようにinahoの収穫ロボットは、最初から100%を狙わず、80%でも収穫してくれるのならロボットに任せて、残りは人手で補うという設計思想です。「80%も自動でやってくれるならば、その分だけ助かるでしょ」ということです。

● ② AIとの知恵比べに勝つアプローチ…画像やアルゴリズムの本質の理解が必須

AIとの知恵比べに勝つ方法は、画像データが出来るまでの本質的なことやディープ・ラーニングが何を行っているかを理解して、AIが何をしているのかを推論し解決策を考えることです。実はAIのアルゴリ

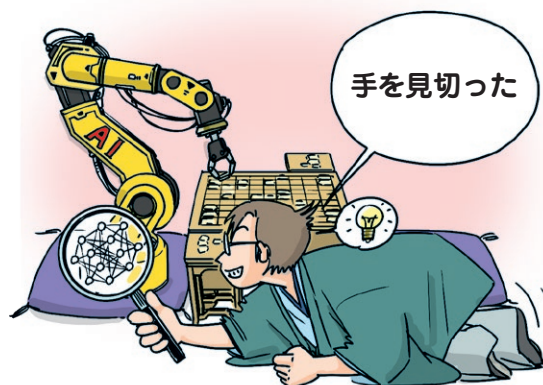


図2 本質を理解してAIとの知恵比べに勝つというアプローチ



写真1 人工知能は80点取ればOKと割り切った野菜収穫ロボット inahoの自動野菜収穫ロボット。残りは人手で収穫する。「かながわビジネスオーデション2020」で神奈川県知事賞を受賞した(写真提供: inaho)

ズムを開発している人ですら、AIが何をしているか正確には分かっていません。

これをスポーツのトレーナやコーチに例えてみます。優秀なスポーツのトレーナやコーチは人間の身体をよく勉強しています。例えば肩や肘を痛めた場合などは、無理をせずインナ・マッスルのトレーニングなどを行い、けがをしにくい身体を作ったり選手のパフォーマンスを上げたりしています。無能なトレーナやコーチは、精神論で片付けようとして、選手をつぶしてしまうことがあるでしょう。身体の筋肉や関節の機能、そして食事や効果的なトレーニング方法を理解することは一流アスリートを育てる上で重要なことです。

ディープ・ラーニングによる画像認識においても同じことが言えて、画像データがどのようにして作られているのかを理解することは重要なのです。

大丈夫？ 画像データのほとんどは“ニセモノ”

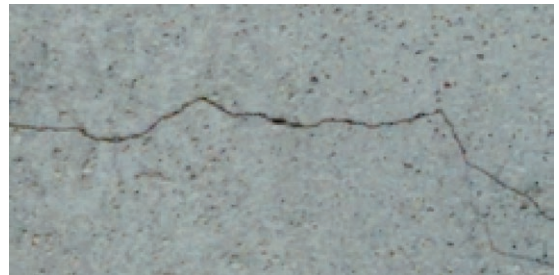
● ①ウェブから手に入る画像は圧縮済み

インターネット、YouTube、地上デジタル放送、DVDなどの普及によって世の中の画像データは100%と言えるほど圧縮されていて、非圧縮の画像を手に入れることが難しくなっています。

8Kの解像度が欲しいと言いながら、高価な8Kカメラで撮影してHEVC(H.265)などで圧縮をかけて劣化させた画像を送っているのですから、ちょっと矛盾しています。また、ネット上にある静止画のほとんどはJPEGなどのコーデックで圧縮されていて、情報量は元画像の5%程度しかなく、張りぼての造形物のようなものです。例として、写真2に拡大したJPEG画像とRAW現像画像(非圧縮)とを比較します。



(a) JPEG (元の情報の5%程度しか持たない)



(b) RAW現像した画像(イメージ)

写真2 画像データのほとんどはニセモノ①…ウェブから入手できる画像は圧縮済み



(a) 発色をよくした画像



(b) ホントの画像(現像処理まで終わらせたRAW画像)

写真3 画像データのほとんどはニセモノ②…カメラで撮影した画像はエフェクトだらけ

第1部

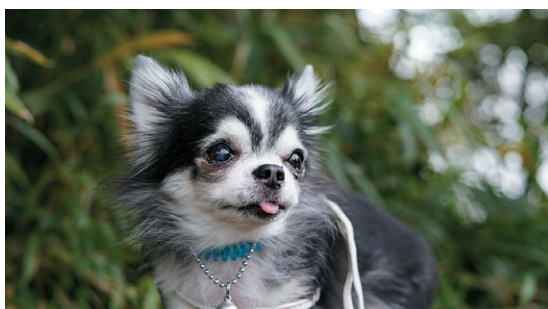
画像の基礎

第2部

画像処理プログラミング

特設

YouTube時代の
撮影&処理術



(a) 例1…犬以外はボケている



(b) 例2…遠くがボケている

写真4 画像データのほとんどはニセモノ③…レンズによるゆがみ

● ②カメラはエフェクトだらけ

我々の持っているカメラも同じです。民生用カメラの宣伝文句として、「自然なボケ感、鮮やかな色、よりきれいに」などのキャッチ・コピーがあったと思いますが、これらのコピーはまさに、「カメラにおいてリアリティを追求しておらず、人が好むような画質を目指している」ことの証明です。

例えば、カラー・マネジメントと言われる色彩の強調や、シャープネスやガンマ・カーブによって輪郭強調されたインパクトのある画像に加工されています。

特に色については、カラー・テレビが登場した時代から、記憶色や期待色という話があって、実際の色よりも強調した色を付けないとお客様に満足してもらえないといった事情があります(写真3)。

● ③レンズによるゆがみ

カメラだけでなくレンズにも同じことが言えます。フォーカスを合わせた被写体以外はレンズの被写界深度の効果を利用してボケた画像になるように設計していて、このボケ方のきれいさを競ったりしているのです(写真4)。

たまにLEDライトや木漏れ日がかきれいに同心円状にボケている写真を目にしますが、人間の目はこのように見えません。これらはあくまでもレンズによるエフェクトなのです。



写真5 広角レンズで得られる画像はゆがんでいるため距離や速度を正確に測れない

近年の高解像度化に伴い、局所的な解像度よりも広い範囲を高解像度で撮影したいというニーズが高まっています。写真5はドライブ・レコーダの広角レンズで得られた画像ですが、横断歩道などがゆがんで曲がって見えます。このようにゆがんだ映像では、せっかく撮影した画像でも距離やスピードが正確に測れず、証拠能力として弱いものとなるため、歪曲収差を換算して距離を求めたりしています。

▶自動運転でも問題になるレンズのゆがみ

スバルのアイサイトのように、ステレオ・カメラで距離を計測して自動ブレーキ制御を行うシステムが実用化されています。このステレオ・カメラによる3次元計測の分野においても、レンズの歪曲収差や製造ばらつきが問題となっています。これらの対処として画像処理による高精度な歪曲収差補正やキャリブレーションが行われています。また、このような歪曲収差によるゆがみは自動運転におけるセンタ・ラインの検出でも問題になります。

つまり、今のミラー・レスなど民生用カメラは、認識や計測向けには設計されていないのです。とは言うものの民生用カメラのコスト・パフォーマンスは非常に高く、これほど低価格、高解像度、高性能、小型なカメラはありません。この民生用カメラをどう使いこなしていくかが鍵かと思います。

画像のことを知らないAIが出来ない理由

● ディープ・ラーニングの画像処理のイメージ

最近ですが、ゼネコンやインフラ関係の方から、「ドローン検査に使いたいのがうまく撮影できない」、「もっと高性能で小型なカメラが欲しい」、「もっと楽に撮影したい」、「検査にディープ・ラーニングを導入しているのだが、昼と夜では結果が変わってしまう。何とかしてほしい」などといった問い合わせを受けま

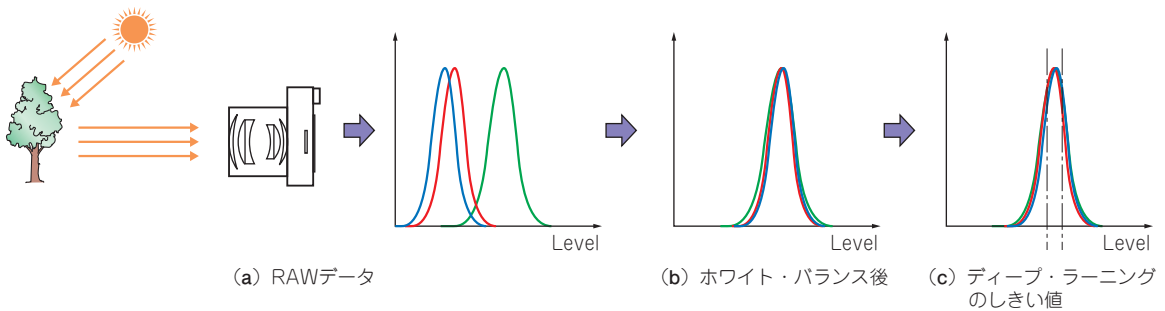


図3 カメラの撮影結果を学習させるということは、カメラが決めたホワイト・バランス後の値を学習してしきい値を決めていることになる

す。そもそも高解像度に撮影すること自体が難しいのです。例えるならこれは、おもちゃのカメラに放送局のカメラよりも厳しい性能を要求しているようなものです。

ディープ・ラーニングとは、さまざまな特徴量のしきい値を自動的に学習で求める技術なのです。また、カメラのホワイト・バランスは、人間が照明光に合わせて色を調整しているので、人の視覚特性に合わせるために付いている機能なのです。ディープ・ラーニングでカメラの撮影結果を学習させるということは、カメラが決めたホワイト・バランス後の値を学習して「しきい値」を決めているということになります。カメラがホワイト・バランスを間違ったり、違うカメラを使ったり設定を変えたりしたら、ディープ・ラーニングで学習したしきい値を超えることもあり得るのです(図3)。ホワイト・バランスからしきい値までを脳で認識している人間とは異なることになります。

● 画像の成り立ちは絶対に知っておいた方が良い

このようにカメラがどうやって画像データを作り、ディープ・ラーニングが何を学習しているのかを知ったうえで、ディープ・ラーニングを使えば、そのアプローチに大きな差が出て来ます。AI(機械学習)やディープ・ラーニングによる超解像なども可能なのですが、画像の劣化要因分析など画像の本質的な知識が必要になってきます。

特集では画像処理の本質を演習形式で

● できるだけ始めやすい実験環境を選ぶ

Windows 10から利用可能になったLinux環境 WSL (Windows Subsystem for Linux) をインストールして、開発環境を構築してみました(図4)。WindowsとLinuxをシームレスに動かすことができ、WindowsとLinuxどちらのソフトウェアも使え、GUIとCUIのどちらのインターフェースも使える良いところ取りと言ってよいでしょう。本格的なディープ・ラーニング

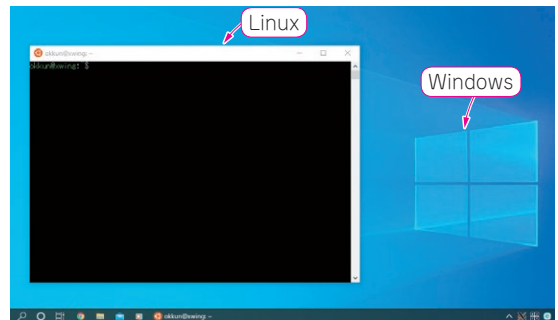


図4 特集では基本的な画像処理のプログラミングをできるだけ用意しやすい環境で実体験できるようにする
WindowsとLinuxがシームレスに使えるWSL

を行おうとするなら仮想でないLinux環境が必要になるのですが、そのLinux環境に慣れるためのトレーニングとしても、面倒なインストールやアンインストールのハードルはかなり低いのでWSLを使って見るのはよいと思います。

● 演習形式でプログラミング

特集では、カメラやディスプレイの画像信号処理を解説します。その後、それらを演習形式で、C/C++でプログラミングします。具体的には、RAW画像から現像した未加工の非圧縮画像を、非圧縮のまま動画像として出力させてみます。

画像処理は、テレビ、パソコン、写真など、さまざまな分野で成長し、分業化されてきたため、歴史は長いのですが画像処理の基礎的・本質的なことを熟知している人は意外と少ないと思います。本特集はこれから画像処理を学ぼうとしている新入社員だけでなく、今、画像処理開発に携わっている方々にも役立つのではないかと思います。

おくむら・あきひろ