

画像教師データ収集のためのビジュアルアーツの設計と開発

松井 直

郭 清蓮

概要: 教師データの収集は、ディープラーニング結果を決定する重要なプロセスであるが、大変な時間と労力が要する作業である。例えば、ビデオ映像から男女を分類するための教師データを作成する場合は、まず、ビデオ映像か人物が存在するフレームを見つけてさなければならない。また、男女を判断するために有効な情報として、人物の頭部や服装に関連する部分を切り取らなければならない。さらに、教師データとして、画像のサイズとコントラストを調整する必要もある。最終的に、画像に適切なファイル名を付け、分類が分かるフォルダに保存することも欠かせなければならない。一つのソフトウェアを利用して、このような一連の処理をマウスクリックやメニュー選択や矢印キー操作などの作業で、ビジュアル的な環境下において、簡単に行うことができれば、各分野の研究者にとっては、とても有意義なことである。我々の研究はこのようなソフトウェアの構築と実現を目標としている。

キーワード: 画像収集, 教師データ収集, ビジュアルアーツ, 可視化

Visual arts design and development for image teacher data collection

NAOKI MATSUI

QINGLIAN GUO

Abstract: The collection of image teacher data is an important process for determining the accuracy and results of neural networks, but this collection work requires a lot of time and effort. For example, if you want to create a teacher data to classify men and women from video footage, you must first find the frame where the video footage or person is present. Moreover, it is necessary to cut the part related to the head and clothes of the person as effective information to judge the man and woman. In addition, it is necessary to adjust the size and the contrast of the image as teacher data. Finally, it is necessary to put the appropriate file name in the image, and to store it in the folder which understands the classification. By using one software, it is very meaningful for the researcher in each field if such a series of processing can be done easily in the visual environment by the work such as the mouse click, the menu selection, and the arrow key operation. Our research is aimed at building and realizing such software.

Keywords: Image Collection, Teacher Data collection, Visual Arts, Visualization

1. はじめに

ディープラーニングは、多くの技術の進歩や大規模データセット、計算機資源がもたらした飛躍的なブレイクスルーにより、音声認識、コンピュータビジョン、そして機械翻訳のような分野で大きな成果を上げてきた。ディープラーニングと強化学習の組み合わせは、コントロールだけではなく戦略とプランニングの領域でもうまくいくことが示されているといわれている[1]。

ディープラーニングは自動的にあらゆる物事を学び、どんな質問にも的確に答えるようになることが期待されているが、確実な成果を得るには、課題を明確に定めてあると同時に、大量な教師データを準備しなければならない。教師データとは、人間が用意した問題に対して答えがセットになっているデータのことである。教師データを繰り返し学習することによって、ディープラーニング自体が出力結果の正否を判断しながら最適化をおこなう。

教師データを作成するために、人間がデータを1つ1つ分類し、意味のあるラベルを付ける作業が必要である。例えば、図1の画像に対して、C1_A1_V4のような文字ラベルが付けられた(MAS and MuHavi data)。人物の動き認識を目的としているため、A1が人物のインデックス番号、C1などはアクションの番号、V4はカメラの番号である[2]。



図1 ラベル付けの例[2]

よって、良質な教師データの作成は、高精度なAI作成のカギを握ると言われている。良質な教師データを作成するために以下のようなツールが開発されている。

「Cloud Video Intelligence API」は、動画中の対象を自動認識してラベル付けることができるアプリである。男女分類の学習データを集めたいと仮定する。その場合、

固定カメラで動いている対象物を背景差分などの手法で抽出できる。その対象物をカメラの設置角などの情報をもとに、顔認識し、大きさを推定する。その大きさを元に、人間である可能性が高い画像を集めることができる。自動的に楽に教師データを作成できそうであるが、ディープラーニングの精度を95%以上にあげるために、特殊な角度から撮影をおこない、手動で人間の部分を取り出し、教師データとして作成する価値はある。ネガティブデータを集めるためにも、手動でデータ作成が重要である。例えば、男性と女性の違いは、服装、持ち物、髪型にも見える。このようなネガティブサンプルを十分に集めることは、高精度を実現する上で必要になる[3]。

ラベリング AndroidApp は、辛いラベル付け作業をとっても有効なスマホアプリである[4]。Adobe Sensei は、ディープラーニングを使った自動的に画像検索やタグ付け、顔認識を活用した加工、画像領域で建物などのラベル付け、オーディエンスのセグメント化、アトリビューション、自然言語処理（デジタル文書のテキスト認識、トピックのモデリング、センチメント分析）など、クリエイティブだけではなく、マーケティングまで広範囲で支援するアプリである[5]。この二つのアプリが必要されている理由は、誰でもデータセットを自分で用意できる環境がほしいからである。手動で独自の学習データの収集ができれば、ディープラーニングにおいても独自の成功を成し遂げるだろう。

従来型の仕事と同様、効率的に効果的に行う環境がほしい。そのために教師データの作成をサポートするシステムが重要となる。本研究のシステムは、パソコンを利用して、手軽に教師データを作成することを支援する。Processing言語を利用して開発を行う、以下のような機能を備えている。

- ・映像から画像を収集する：映像から目的となるものを写っているフレームを自動的に見つけ、そのフレームを画像として収集・保存する

- ・タイピングを避ける：ビジュアル的なインターフェースを提供することによって、ファイル名を入力する作業を最小限にする。新しいカテゴリを追加する際、一回だけ名前変数を初期化すればよい。メニュー、ボタン、チェックボックス、スライドバーなどを活かし、マウス操作でファイルの選択、切り換え、ラベル付などを実現する。

- ・切り取り作業を可視化する：画像から有効な部分だけ切り取り、サイズを揃え、新しい名前を付け、教師データとして保存する。大量な画像に対して、この一連の作業を行うことが不可欠であるが、可視化とインターフェースを利用して、作業のストレスを減らすことができる。カテゴリ・

visualization をベースに、マウスクリックだけの操作に単純化したいと考える。

- ・カテゴリ visualization：画像データのカテゴリが思ったより多く、高い精度のディープラーニング結果を期待するならば、教師データを再分化する、階層的な構造を作る必要がある。人分析を例にすれば、正面から自動的に正確に人物を認識できるが、側面、頭上、後ろ姿からも、人物を認識できるようにするために、各カテゴリの教師データを漏れなく用意しなければならない。データに何らかの偏見があれば、それに基づいて訓練されたシステムもそうなる。Google は、そんな厄介で深刻な問題を引き起こしかねない状態を、“Equality of Opportunity” [機会均等] と名付けた方法を用いて系統的に回避しようとしている。

カテゴリを可視化することは、教師データを作成ツールには必要なサポート機能、作業の負荷を軽減させる有効な手段だと考えられる。カテゴリが多くなると、木構造でカテゴリを管理することが難しくなるが、カテゴリの可視化は一層困難になる。と言うのは、限られたスクリーンスペースに置いて、元画像、抽出画像、インターフェース、カテゴリなどを同時に表示したいからである。動的に変化する、構造と色を活かして複雑なカテゴリを分かりやすく表示することは、本研究の特徴である。

- ・冗長な作業を避ける：ラベル付けにはその分野の専門家が必要な場合がある。複数人でラベル付けしている場合、人によって異なる教師データが作られる可能性がある。このような冗長な作業も避けたい。できる限り普遍的なラベルを付けられるために、同じ元画像から作られた教師データなら、比較・統一させる機能を提供し、一枚のデータに自動的に纏めることができる。

- ・水増しデータ作成機能：データの水増し（Data Augmentation）は、元の学習データに線形変換を加えてデータの数を増やすことをいう。線形変換とは、左右反転、位置をずらす、ノイズを増す、コントラストを変えるなどのズーム・変形したりする処理である。データの水増しは、少ないデータでも学習精度を上げるためのテクニックであるので、これをサポートする機能も提供したい。

2. システム構成

構築したシステムは、次の通りである。

(1) システムのトップ画面

ツールを実行するとトップ画面として、図2が表示される。

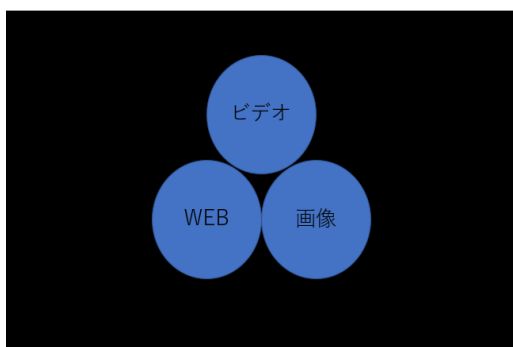


図2 ツールトップ画面

ツールのトップ画面は図2のように、3つのボタンで構成されており3つのモードがある。そのモードは、「ビデオモード」、「WEBモード」、「画像モード」である。「ビデオモード」では、主に動画のフレーム抽出をおこない、「WEBモード」では、WEBから画像を取得し、「画像モード」では、画像を加工し名前を付けて保存することができる。

(2) ビデオ画面

「ビデオボタン」をクリックすると、図2の画面に移行する。

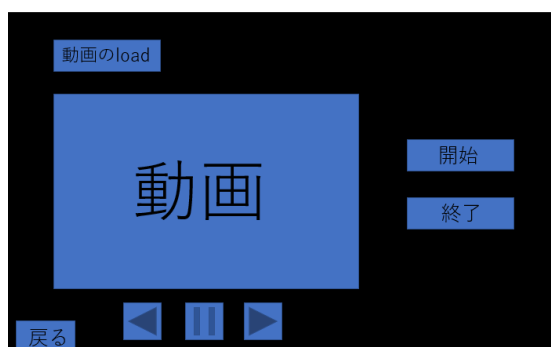


図3 ビデオ画面

ビデオモードでは、主に動画からフレームの抽出をおこなう。初めに、画面左上にある動画 load ボタンをクリックすると図3の画面に移行しフレーム抽出をおこないたい動画を選択することができる。次に動画を選択した後に画面下側にある動画動停止ボタンをクリックすることで、動画をスタートさせることができる。最後に、画面右側にある開始ボタンをクリックすることで任意のタイミングでフレーム抽出をおこない終了ボタンを押すことでフレーム抽出を終了することができる。

抽出したフレームは、TIF ファイルとして frames フォルダに番号順にファイル名を付けて保存される。左下の戻るボタンをクリックするとトップ画面に戻ることができる。

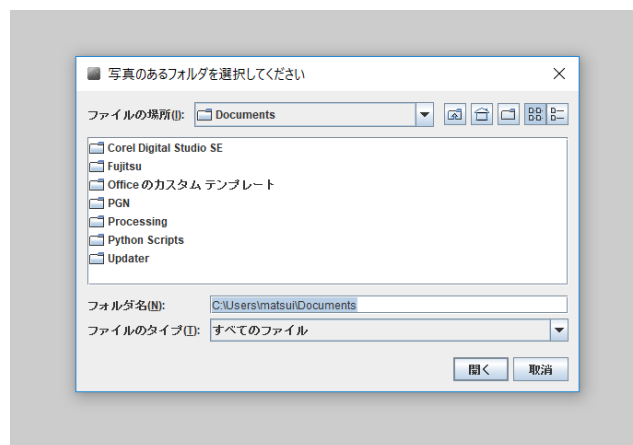


図4 動画選択画面

(3) WEB 画面

「WEB ボタン」をクリックすると WEB モードに移行し、WEB から画像を取り込むことができる。

(4) 画像画面

「画像ボタン」をクリックすると、図4の画面に移行する。



図5 画像画面

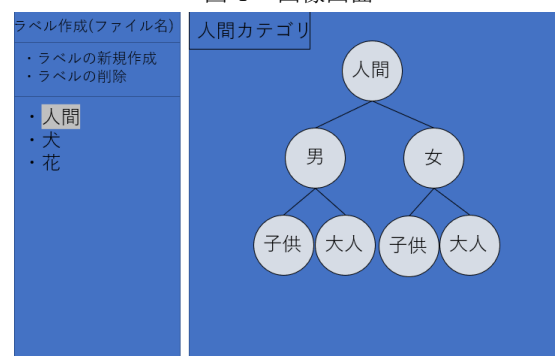


図6 ラベル作成画面

画像モードでは、主に画像から必要な部分だけを学習データとして効率よく取り出し適切なラベル付け(名前)をできる。初めに、画面左上にある画像 load ボタンを押し画像があるフォルダを選択する動画モードの図3と同じような画面に移行する。次に、画像に適切なラベルを付けるために画面中央上にあるラベル作成タブを押すと図5のような

画面に移行する。この画面では、ラベルの作成と削除をおこなうことができる。

例えば、人間を認識する AI を作ろうとする場合、簡単に構成を考えると図 5 のようになる。この場合ラベル名を付けるとすると人間で男、子供の時、人間は英語で Human, 男は Man, 子供は Child となるので頭文字を取り HMC となる。さらに、学習データは大量に必要となるので HMC の後には順番に数字を振り分けるようになっている (HMC-001~HMC-100 などとなる)。

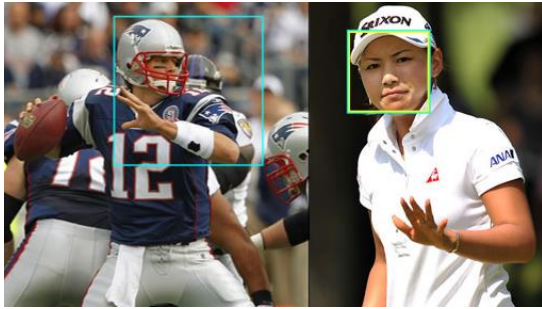


図 6 (左)画像の領域指定, (右)顔を自動的に検出

ラベル設定をおこなった後は、画面右下に作成したラベルがカテゴリとして木構造など可視化した状態で表示され、そのラベルがどのような構成でできているかを瞬時に見て分かるようになる。

最後に、図 6 のように画像から学習データとして使用したいところを水色の四角部分を任意に合わせて画像を取り出すか、人間の顔として認識した場合自動的にあわせて画像を取り出す機能がある。その後画面中央下にある保存ボタンを押すかラベルを作成したときにショートカットキーを設定することで保存したい場合にそのキーを押すことによって設定されたラベル名で保存することができ、保存された画像は PNG ファイルとして指定のサイズにリサイズして保存先に設定したフォルダに保存される。保存された画像は画面右上にファイル名とともに表示される。また、画面下にある左右キーを押すと同一フォルダにある画像を連続して処理することができる。画面右下にある戻るボタンでツールトップ画面に戻ることができる。

3. カテゴリとラベルの可視化

本ツールでは、図 7 のようにカテゴリを作成する時内部的に 2 つ以上の階層を持つ木構造となる。そのカテゴリを可視化することによって、関係性を客観的に見ることができ、簡単にラベルを変更できるようになっている。そこで作成したカテゴリを使用して、図 4 のような限られたスペースに効率よく表示し、任意にラベルを選べる表示方法を検討し、ツール内で使用者に合った表示方法を選択できる

ように 3 つの方法を設計開発した。

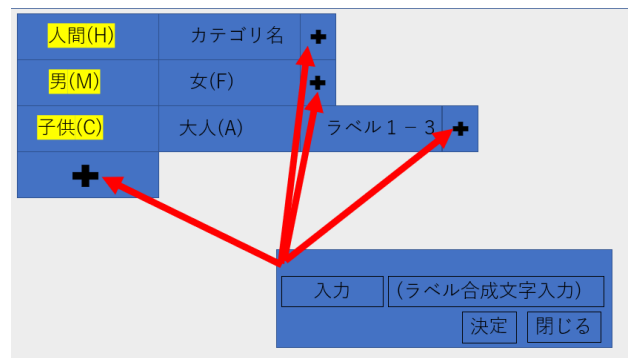


図 7 カテゴリとラベル作成

3.1 ラベル作成画面

図 7 よりカテゴリとラベルを作成する場合、初めにカテゴリ欄にあるプラスタグをクリックし図 7 の右下の入力欄に左側に名称右側にラベル名を作るための文字を入力する。図 7 の場合はカテゴリ名に「人間(H)」その下位の階層に「男(M)」と「女(F)」さらに、「男(M)」の下位の階層に「子供(C)」と「大人(A)」があり、「女(F)」タブをクリックするところにも下位の階層に「子供(C)」と「大人(A)」が存在している。

3.2.1 ラベルの可視化表現 part1

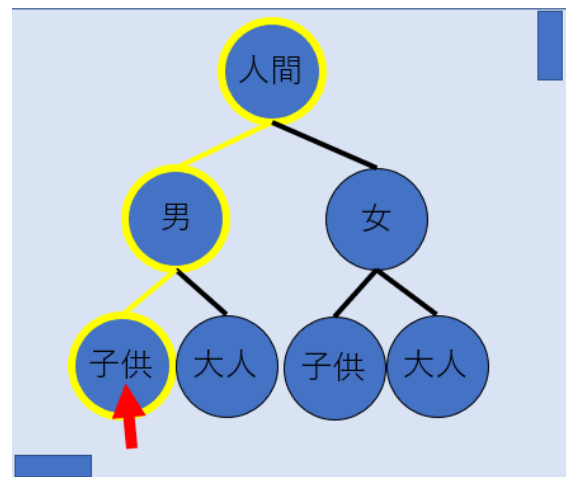


図 8 木構造型表示

図 8 のように木構造型表示は、全体の構成や縦と横の関係性が分かりやすい表示方法だが、構成が多くなると木全体が大きくなり全体が見えにくくなり表示領域を圧迫してしまう欠点がある。

3.2.2 ラベルの可視化表現 part2

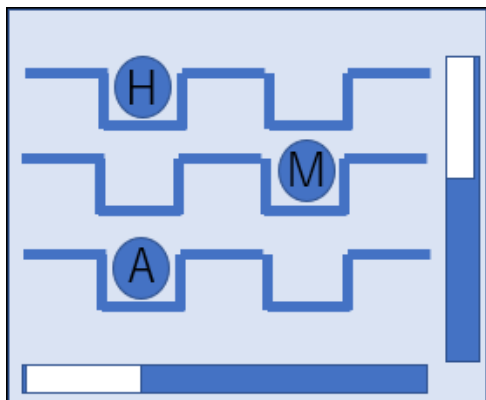


図9 段差型表示

図9のような段差型カテゴリ表示は、シンプル性を重視した表示方法である。木構造型と比べると全体的に表示領域が小さくなるが、表示方法がシンプルになるのでどこに何があるかを把握しなければいけない欠点がある。

3.2.3 ラベルの可視化表現 part3

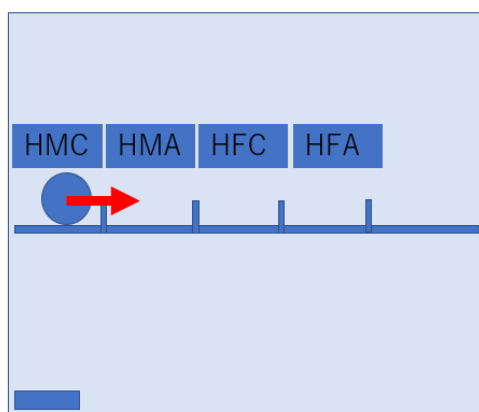


図10 スクロール型表示

図10のようなスクロール型表示は、段差型をさらにシンプルにした表示方法である。図10のようにラベル名を表示し、使用したいラベル名の領域に円をスクロールすることでそのラベル名を使用できる。この表示方法では、全体の関係性が分からなくなるのと横に領域が大きくなる欠点がある。

4. まとめ

本研究では、ニューラルネットワークなどで使用する画像教師データの収集を効率よくおこなうための自作ビジュアルーツについての設計と開発について述べてきた。従来の複数個のツールを用いて学習データの作成や毎回おこなわなければいけなかった名前を付けフォルダを選び保存

する手間などを削減することができた。

今後の展望として、ツールのビジュアル面の向上や画像の切り取り方法の追加、コンストラクタの設定などが上げられる。

参考文献

- [1] “セミナー情報 | パナソニックの人工知能研究開発 - Panasonic”.
<http://tech-ai.panasonic.com/jp/seminar/bengio2016>, (参照 2018-4-13).
- [2] “NEW MuHAVi “uncut” (November 2014) - velastin”.
<http://velastin.dynu.com/MuHAVi-MAS>, (参照 2018-4-13).
- [3] “Cloud Video Intelligence API で動画の自動認識を試す | そるでぶろぐ”.
<https://devlog.arksystems.co.jp/2017/03/13/2635>, (参照 2018-4-13).
- [4] “深層学習の訓練用画像のラベルをスマホで作成 [紹介] - Qiita”.
<https://qiita.com/MOKSckp/items/7a12efbf5b23632862eb>, (参照 2018-4-13).
- [5] “デザイナーの新しい常識！「Adobe Sensei」が告げる次世代デザイン環境の到来 - Adobe Creative Station”.
<https://blogs.adobe.com/creativestation/general-adobe-sensei-next-generation-design-environment>, (参照 2018-4-13).

正誤表

下記の箇所に誤りがございました。お詫びして訂正いたします。

訂正箇所	誤	正
3 ページ ～5 ページ 図番号	図 2 ビデオ画面～図 1 0 スクロール 型表示	図 3 ビデオ画面～図 1 1 スクロール 型表示
3 ページ 8 行目～5 ペ ージ 図番号	各ページの図番号	各ページの図番号を 1 つずつ繰り上 げ.