

## 論文

## 機械学習データセットのための時系列動画アノテーションツールの開発

林 勇 樹\*  
追手門学院大学Development of a Time-Series Video Annotation Tool  
for Machine Learning DatasetsYuki HAYASHI  
Otemon Gakuin University

## Abstract

The utilization of video in machine learning has garnered significant attention recently. Recognizing what appears in videos and understanding their actions are crucial tasks, leading to various research efforts in video processing. However, studies involving prolonged videos, such as sports, often need help with unique appearances, movements, and camera angles, making applying general pre-trained models from large datasets difficult. This research focuses on competitive swimming as a specific case, developing a new dataset and a video annotation tool without relying on existing models or datasets. The tool aims for compatibility across devices, ease of annotation even in long-duration videos, and the ability to leverage the unique aspects of competitive swimming. The system developed through this study<sup>3)</sup> enables easy task distribution among annotators beyond the developers, facilitating the construction of necessary structures for the dataset with minimal operations.

キーワード：データセット，時系列，動画アノテーション

Key Word：Dataset, Time Series, Movie Annotation

## 緒言

近年，機械学習における学習データとしての動画活用の需要が高まっている。特に動画内容の認識や理解は様々な応用分野での研究の中心となっている。動画からの情報抽出タスクは，日常のシーン認識からスポーツ分析の現場まで幅広く利用されはじめている<sup>1,2)</sup>。しかし，サッカーや野球に代表されるメジャーなスポーツ以外の種目のような，特殊な環境下で撮影された動画を機械学

習システムのテーマにする場合，一般的な学習済みモデル<sup>3)</sup>やデータセットでは十分なパフォーマンスを発揮できないことが多い。例えば，競泳においては水面上に横向きになった状態で推進するが，多くの学習済みモデルは立位または座位のデータを人間として学習<sup>4)</sup>しており，横向きに推進する水面の人間や水中の人間をうまく認識することができない。このような場合，対象となる競技の見え方，姿勢に特化したデータセットで学習を行うアプローチが必要となる。本研究では，競泳競技に対

\*責任著者：林 勇樹，追手門学院大学，yu-hayashi@otemon.ac.jp

Corresponding Author：Hayashi Yuki, Otemon Gakuin University, yu-hayashi@otemon.ac.jp

する機械学習プロセスに焦点を当て、特殊な環境下で撮影された動画に対する効率的なアノテーションを可能にするツールの開発を行う。既存の解析手法やデータセットに依存せずに構築された独自のデータセットを基に、機械学習モデルの精度向上を目指す。このアプローチにより、スポーツ動画の分析、特に特殊な条件下での動画分析に応用可能な手法を提案することが本研究のねらいである。

## 方法

### 物体認識 (Object Detection) のためのアノテーションツールの開発

アノテーションツールはHTMLとJavaScriptを用いたブラウザベースのシステムとして開発した。Webブラウザを用いて、ページ上でアノテーションを行いたい動画を開くと、十字キーが表示され、一般的なアノテーションツールと同様にアノテートしたい部分を選択することでその座標が画面上の表の動画時間軸に対応する行へ挿入される。一度目のアノテーションでマウスダウンしたところからマウスアップしたところまでのマウスの軌跡により、起点になる座標  $(x,y)$  とアノテーションする区形の横幅  $(w)$ 、縦幅  $(h)$  が決定され、 $(x,y,w,h)$  の四つのデータが時間  $t$  の位置に挿入される。さらに、マウスアップのタイミングで現在の入力欄は隣の選手または次の時間軸に自動で移る仕組みとした。この自動遷移が開発した動画アノテーションツールの最大の強みであり、複数の操作を介在させることなく、連続的に動画の中に映っているスポーツ選手の動作や目的の物体を学習させるためのデータセットを構築していくことができる。特に本システムは競泳競技にフォーカスを当てており、1回のレースに8人から10人が一度に泳ぐことを踏まえ、画面上の表は横に10個の列（選手毎）と縦に動画の時間（フレーム毎）の行数を用意した。

本システムはブラウザベースであるため、一般的なPCに加えて、タブレット端末でも使用可能である。特にアノテーション作業には細かな作業と精度が要求される。その点において、タブレット端末用のペン型デバイスを併用することで（図2）、効率的に高精度な学習データの生成を実現できるように支援することができる。

### クラス分けのためのアノテーション

アノテーションにおいて、物体の座標を認識するタス

クのためのものと、認識した物体が何を行っているのか認識するための分類タスクに使用するかによってデータセットの内容に違いがある。物体が何を表しているかを認識するためには、あらかじめ用意された行動タグを物体に紐つける必要がある。競泳競技の場合、認識したい物体（選手）は競技プールの中で8人から10人の人間であり、その人間が手を動かしたり足を動かしたりする様子を細かくフェーズ事に切り替えて、手のかきにかかる時間や進んだ距離を分析するデータを作成する必要がある。そのようなタスクにおいて、それぞれの姿勢に特徴的な画像上の見え方、特徴を踏まえた学習を行う必要がある。データセット上でのタグ付けの精度と、タグ付けされたデータの数・多様性<sup>5)</sup>は機械学習プロセスの精度向上において重要な意味を持つ。

前述の物体認識のためのデータセットに加えて、認識したもの、すなわちすでに物体認識された座標位置において、そのデータが表している情報に対応するタグを効率よく付加していく方法を検討する必要がある。本システムでは、この課題に対応するため、データセットにタグ付けをする際に、クラスごとに設定されたキーボードのキーの割り当てを行うこととした。競泳競技をターゲットにした本研究の場合、[Sキー]はストローク、[Uキー]は水中局面、[Bキー]は水中局面から陸上局面への遷移タイミング、というようなクラス分けを行う。このクラス分けを行う際には、キーボードの押下イベントと同時に、次の時間軸（フレーム）への移動も可能としており、前述のマウスオフタイミングでの対象入力欄の遷移と同様、使用者の余計な作業をできるだけ割いた操作性を重要視した。（図3）特に本研究で扱った競泳のような連続的な動作を分析・解析する上では、周期的な動作の時間の変動、動きの変動を捉えることが大変重要であり、このタグ付けの質と量が機械学習の結果に違いを出すと考えられる。

またアノテーションを行う際、視覚的な支援システムとして素材のマスク化支援システム（図4）を導入した。

競泳競技は同じ選手が移動する場所、すなわちレーンは定まっているため、アノテーションしている選手のレーン以外の場所にマスクをすることで、アノテーション間違いを防ぐことができた。アノテーションツールへのマスク機能の導入前後ではアノテーション間違いが減り、同じレーンを間違えて再度アノテーションしてしまったというような、作業が徒労に終わることがなくな

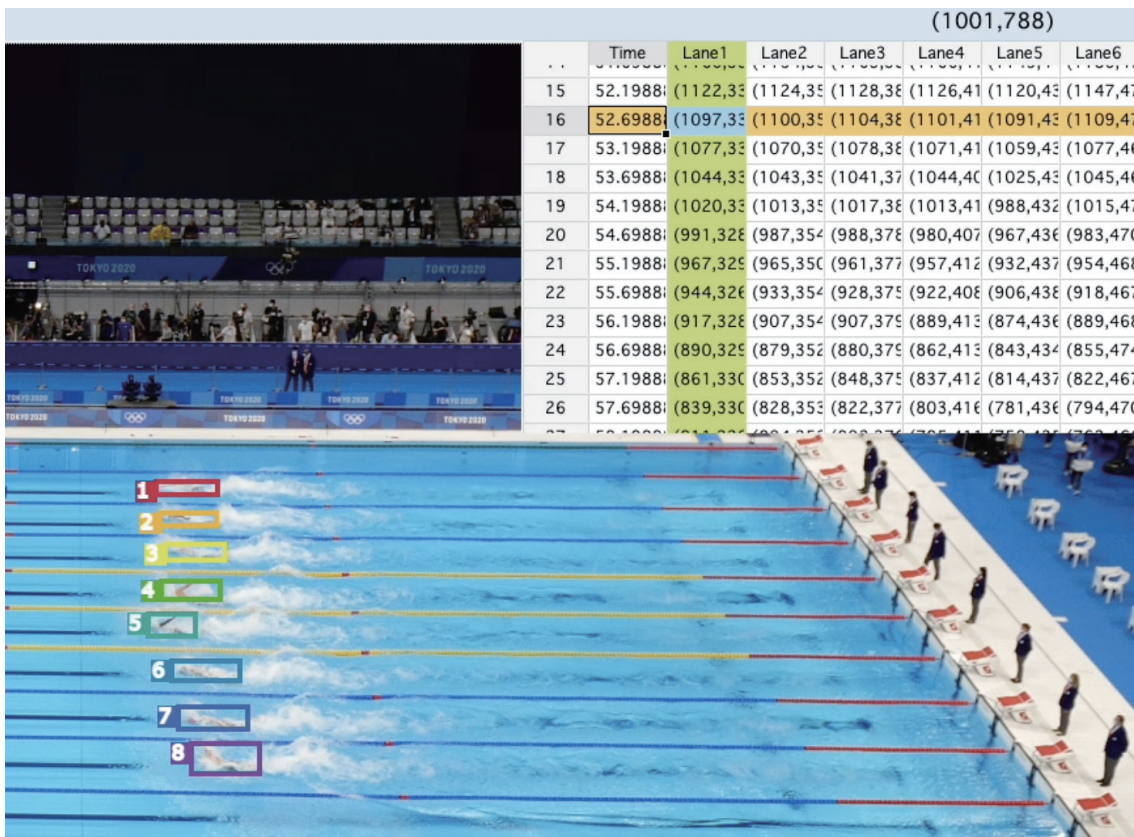


図1 開発したアノテーションソフトウェア

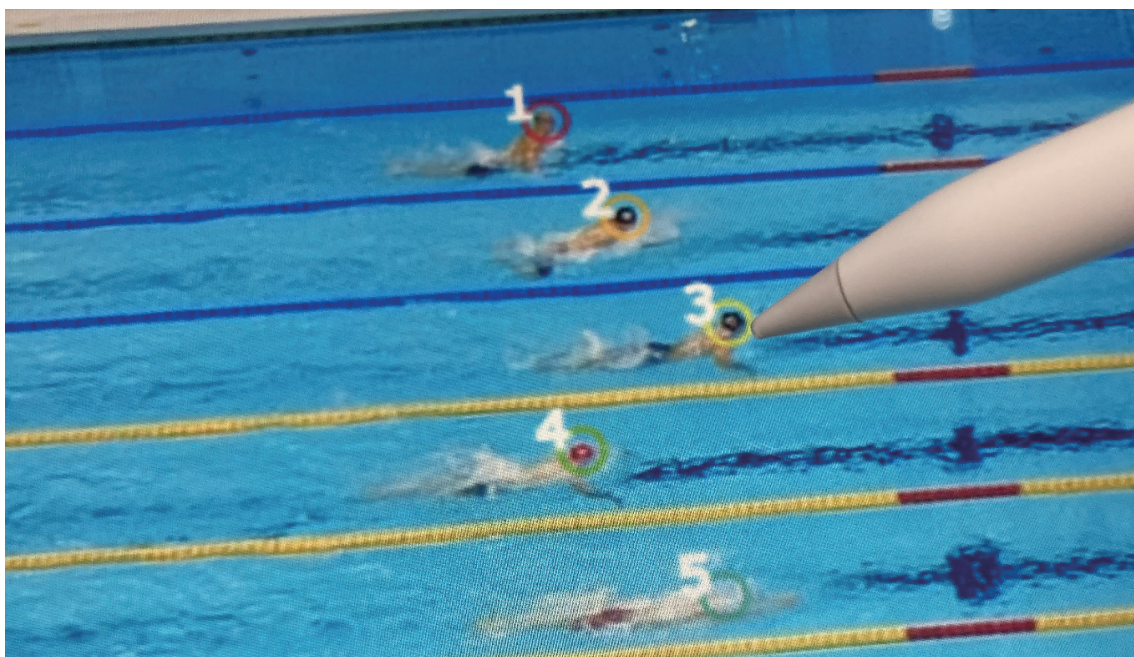


図2 タブレット端末とペン型入力デバイスによるアノテーション

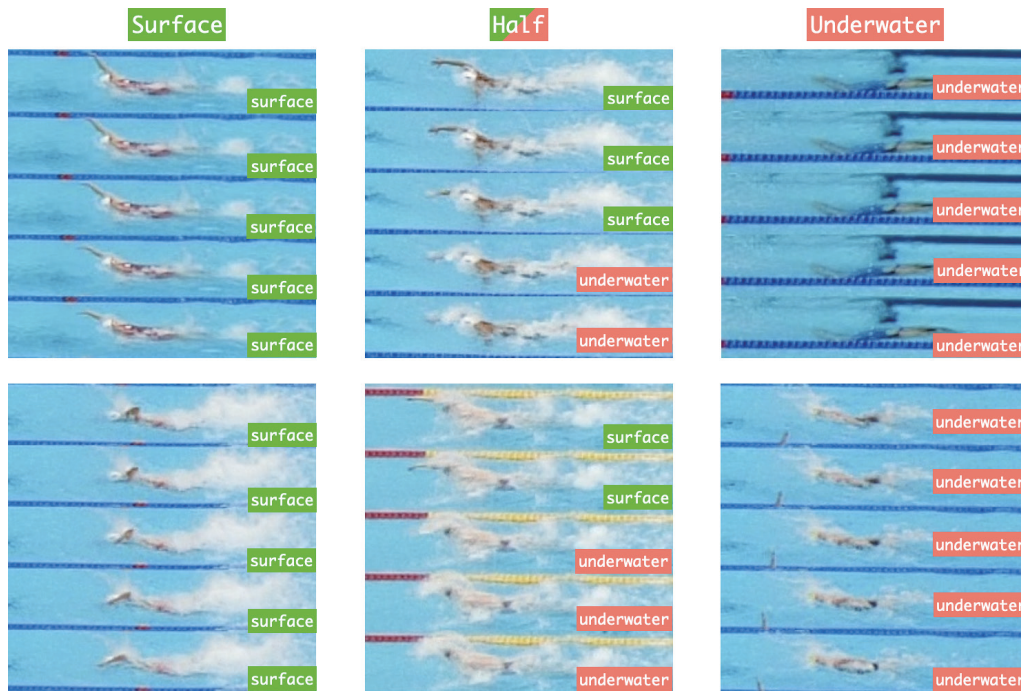


図3 開発したシステムによりタグ付けされた時系列データの一例

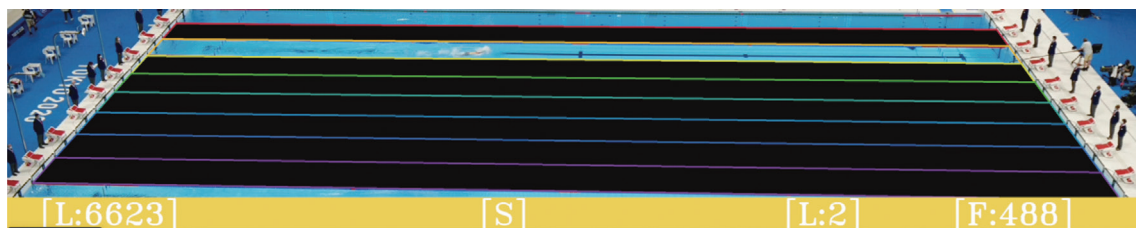


図4 レーン外を分析しないようにするためのマスク

る。本システムではアノテーションされたデータをその場で動画として可視化、再生する機能を有する。

## 結果

本研究により開発されたアノテーションツールを用いて、与えられた動画に対するアノテーションを行った結果、一般的なアノテーションツールを用いたアノテーション時間の短縮とアノテーションしてから機械学習のプロセスを走らせるまでの時間の短縮を行うことができた。競泳競技のような人間の行う複数の泳ぎ方に対して存在する、それぞれに特徴的な動作と似通った動作との種目ごとの差異を学習させるための大量のアノテーションデータを用意することができた。このような機械学習プロセスにおいて、アノテーション時間の短縮が行われ、アノテーター2名で2週間でのアノテーション作業

により、異なる泳者30名を含む8種目、約27万個のタグ付けされたデータセットを作成することができた。

## 考察

アノテーションを進める際に問題になった点として、時系列にアノテーションしていくべきか、一つの時間に対して全ての泳者をアノテーションするかという、操作性についての検討事項である。よし悪しはアノテーターの好みによる部分もあるが、正確なアノテーション、特に複数の泳者に対してその泳者の見た目の特化しすぎないアノテーションを行う上で、同じ時間に複数の泳者を順番にアノテーションしていく手法によって皮膚の色や骨格に依存しない汎用性のあるタグ付けを行うことが可能になった。

当初は五つのクラス分けを一つの泳者に対して実施す

ることを検討していたが、機械学習プロセス精度の検証にあたり、クラス数を削減する必要が出てきた。このような場合にもすでにタグ付けが済んでいるデータをシステム上で機械的に変換することで、容易にクラス数の減少に対応させることができたと考えられる。本システムを用いて行ったアノテーションによるデータ数は、機械学習におけるオリジナルのクラス分けタスクを実行上では十分なデータ量で、アノテーション後の推論プロセスで実施した。結果<sup>6)</sup>は良好であり、本研究開発によってアノテーションツールそのものを自作することが研究の推進に大きく影響したことを確認できた。

## 結論

本研究における動画アノテーションツールの開発によって、機械学習のプロセスの中で重要な役割を果たすデータセットの生成を支援するためのシステムを開発することができた。特に一般的なデータセットを使うことができない特殊な環境を目的とした機械学習プロセスにおいては、データセットの量・質が機械学習の推論結果に大きな差を与えるため、自作、アノテーションツールそのものを自作する意義は大きかったと言える。開発したアノテーションツールは2種類であり、物体認識のためのアノテーションツールと、認識した物体のクラス分けのためのタスクに対応するアノテーションツールの2種類を開発した。物体認識のためのアノテーションツールでは、マウスのマウスオンとマウスアウトのみでアノテーションを進めていくことができ、作業の効率化が図られた。またデバイスを問わず利用できること、さらにタッチディスプレイを利用できる環境においてはペンタブレット、ペン型入力デバイスを用いることで、より詳細にかつスピード感を持ってアノテーションができるようになった。またクラス分けにおいては選手の存在しないレーンにマスクをすること、クラスに対応させたキーボード入力によって現在の姿勢をタグ付けすること、キーボードの入力を動画のコマ送りと紐付けることで、タグ付けのタスクと動画送りのタスクを同時にこなせるシステムを開発した。本研究で開発されたシステムを利用して行われた機械学習プロセスでは、認識や解析の精度は期待された基準を満たし、通常のアノテーションツールや公開されているデータセットよりも、より特殊な環境に適したデータセットを自作できることが確認された。

## 謝辞

本研究は、公益財団法人ヤマハ発動機スポーツ振興財団による YMFS スポーツチャレンジ助成 2022 年度（第 16 期）「機械学習を用いた競泳レース分析の高解像度化による新しいストローク分析パラメータの抽出」の助成を受けたものです。ここに感謝の意を表します。

## 利益相反

本研究において申告すべき利益相反は存在しない。

## 参考文献

1. Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, et al. (2021) ViViT: A Video Vision Transformer, *Computer Vision and Pattern Recognition*, abs/2103.15691
2. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, et al. (2020) End-to-End Object Detection with Transformers, *Computer Vision and Pattern Recognition*, abs/2005.12872
3. TsungYi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, et al. (2014) Microsoft COCO: Common Objects in Context. *Computer Vision and Pattern Recognition*, abs/1405.0312.
4. 岩瀬 慶, 榎堀 優, 吉田 直人, 間瀬 健二 (2019) 人物領域推定と姿勢情報を用いた寝姿体圧画像からの関節位置推定の検討. *情報処理学会研究報告 (Web) IPSJ Technical Report*, 15: 1-7.
5. 金城 敬太. 機械学習における多様性と公平性に関する一考察. (2021) *人工知能学会研究会資料 知識ベースシステム研究会 123 回*, 8-04.
6. 林 勇樹, 仲井 志文, 今井 元太, 松井 健, 機械学習を用いた競泳レース分析システムの開発, (2023) *日本水泳・水中運動学会 2023 年次大会 論文集*, 76-81