

AI開発のコスト・スピードの課題を解決する
アノテーション済データセットの活用方法



株式会社ブライセン

2022.02

目次：

1. 背景
2. アノテーションを取り巻く課題
3. ライセンスの取り組み・自社ソリューションと解決策
4. データセット販売
 - バウンディングボックス (BB)
 - セマンティックセグメンテーション (SS)
5. ライセンスデータセットを用いた転移学習検証結果
 - 検証結果
 - 検証方法
6. 会社概要・お問い合わせ先

1. 背景

人工知能技術（AI）は、1950年代にはじめて出現して以来、60年以上にわたって世界で急速に発展・応用されており、各国で研究開発が活発に進められています。昨今では、AIを活用した新たなサービスが様々展開され、その社会的なニーズの高まりに伴い、AI開発において欠かすことのできない教師データ作成、アノテーションの需要も大きく増加しています。

2. アノテーションを取り巻く課題

AI開発には、膨大な学習データが必要不可欠で、データの質や量によって結果の精密度も変化します。分析モデルは、「大量のデータを学習し、法則性などを学習することで、正解がまだわからない新たなデータを新たにインプットして、学習時に決められたルールやパターンを元に認識・予測する」ので、アノテーションの「精度（正答率）」がより正確な分析モデルを作るためには必要不可欠です。¹正確な分析モデルを作るためにアノテーションは重要な作業になります。アノテーションはデータの数に比例して、時間・費用・人員が必要で、AI開発を行う企業にとって、大きな課題になっています。

ブラウザ上には既にアノテーション済の教師データも公開されています。アノテーション済のものを使用することで、効率良く正答率を上げることができます。しかしながら、これまでブラウザ上に公開されているデータのほとんどが、商用利用を禁止しているため、膨大なデータの収集・処理を初めから行わなければいけません。また、既存のデータは国内データに対応しておらず、国内での商用利用には品質をクリアしていないことがあります。

¹ ANINOW. (2020, September 11). 機械学習におけるデータの重要性とは？基礎から実用化まで徹底解説！ | AI 専門ニュースメディア.AINOW. Retrieved January 7, 2022, from <https://ainow.ai/2020/09/16/228069>

3. ブライセンの取り組み・自社ソリューションと解決策

近年、ドライブレコーダーにAIが搭載され始め、新たなサービス開発や自動運転に伴う走行データの社会的ニーズが高まっています。

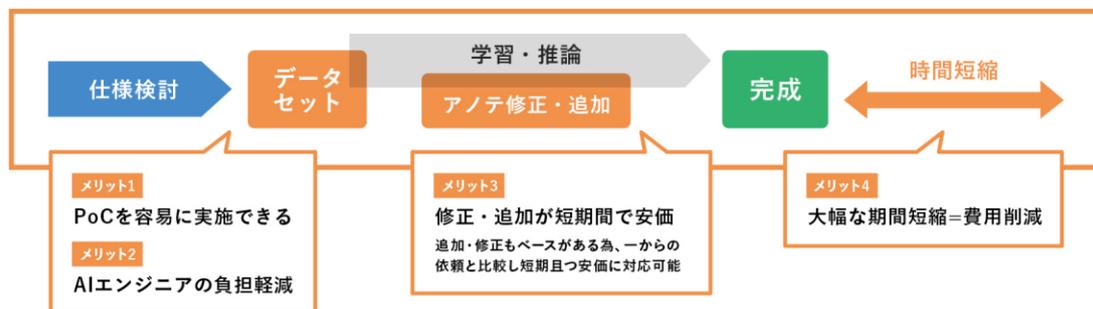
そこで、(株)ブライセンではAIに欠かせない教師データをデータセット化、商用としての使用が可能なデータセット販売を開始しました。高精度 3D 地図データなど位置情報技術を駆使するロケーションインテリジェンスカンパニーのジオテクノロジー株式会社（旧：インクリメントP(株)）のデータを活用し、AI用のアノテーション済みのドライブレコーダー走行画像データセットを販売します。現在ブラウザ上に存在する教師データは商用利用が禁止されているケースもある日本国内では、弊社が業界初の試みです。

アノテーション済みのデータセットを活用することで、機械学習におけるデータの準備費用・それに掛かる時間を大幅に削減することができます。（図1）既存のデータセットをベースに、必要なアノテーションを修正・追加することでブライセンのデータセットは、商用利用を目的とする企業・開発者に国内の自動車走行データの提供を実現可能にしました。

■ 通常時



■ データセット利用時



【図1：AIモデル完成までの通常とデータセット販売を使用した場合の比較】

ブライセンでは、6年前より自動走行データのアノテーション作業を行ってきた経験をいかし、全国のドライブレコーダー走行画像を保有している、インクリメントP社との協業により、迅速かつ安価に大量のアノテーション済走行データセットの販売を実現しました。アノテーション済みのデータセットを活用することでPDCAを迅速かつ安価に回すことを可能とし、AIモデル開発における課題のひとつ、「POC 地獄」の解決方法を提案します。

販売開始から多様な業種（大手携帯キャリア・保険会社・建築会社など）の国内企業から相談を受け、教師データの販売の販売価値と多方面での需要の高さを確信しています。また、今後は走行データのみならず、衛星や産業用途向けのデータセットなど、分野の拡充を予定しています。

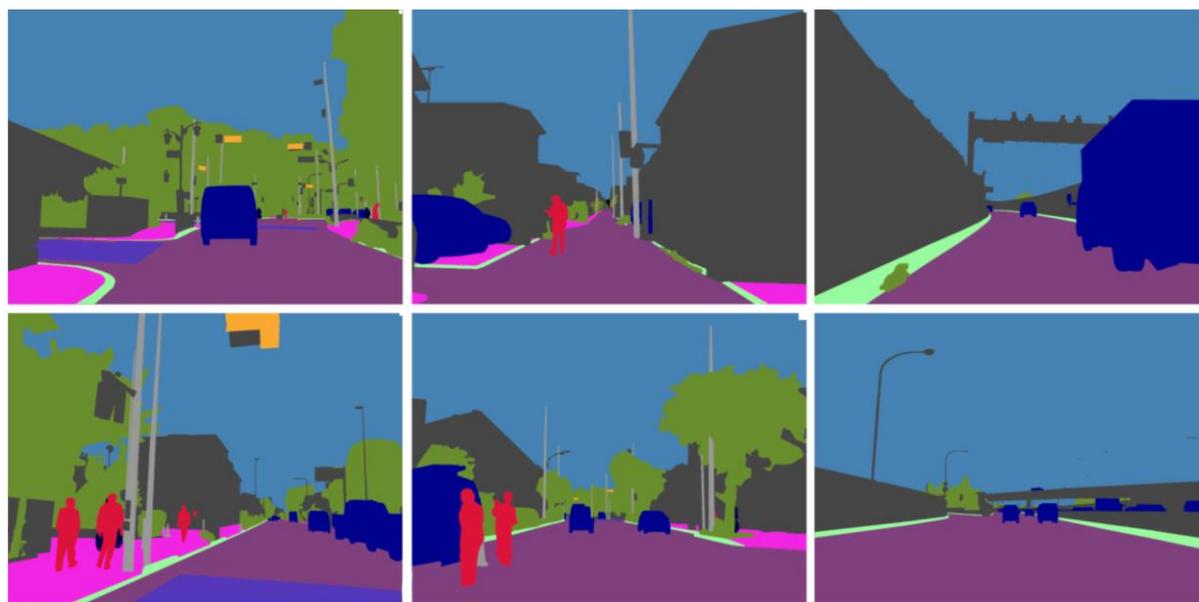
4. データセット販売

■ バウンディングボックス (BB)



項目	内容	
アノテーション種類	バウンディングボックス (BB)	
対象物 (8クラス)	信号機	一時停止標識
	一時停止路面表示	一時停止線：白線
	自動車	人
	バイカー	横断歩道
枚数	10,000枚	
費用	500,000円	
画角	1600×1200	
提供物	・ XML (PascalVOC形式)	
	・ 元画像	
	・ 短形合成画像 (PNG) ※参考用	
提供形式	ブルーレイディスク	

■セマンティックセグメンテーション (SS)



項目	内容			
アノテーション種類	セマンティックセグメンテーション (SS)			
対象物 (20クラス)	信号機	一時停止標識	一時停止路面標示	一時停止線：白線
	自動車	人	バイカー	横断歩道
	車道	歩道	人工物	自然物
	空	ポール	縁石	自動車：実線
	自動車：破線	自動車以外：実線	自動車以外：破線	その他：背景
枚数	3,000枚			
費用	1,000,000円			
画角	1600×1200			
提供物	・アノテーションデータ (JSON)			
	・塗りつぶし画像 (PNG)			
	・元画像			
	・クラス色 (RGB) 定義ファイル			
提供形式	ブルーレイディスク			

※データの利用規約

- ・ご利用可能な範囲は日本国内のみになります。海外へのデータのご提供及び使用はご遠慮ください。
- ・同業他社様へのデータの提供は出来かねます。ご了承ください。

5. ブライセデータセットを用いた転移学習検証結果

オープンAIモデルYOLOv3を用いて、ブライセン保有のデータセットを実際のAIモデルに学習させ、以下のような効果を検証しました。

■検証結果

- (1) ブライセンによるアノテーション済みデータセットが、オープンソースのAI (YOLOv3) の転移学習 (※1) 用教師データとして有効に機能し、新たなオブジェクトを認識することができた
- (2) ブライセンによるアノテーション済みデータセットは、オリジナルのYOLOv3より信号機や停止標識などのオブジェクトをより高く認識することができた

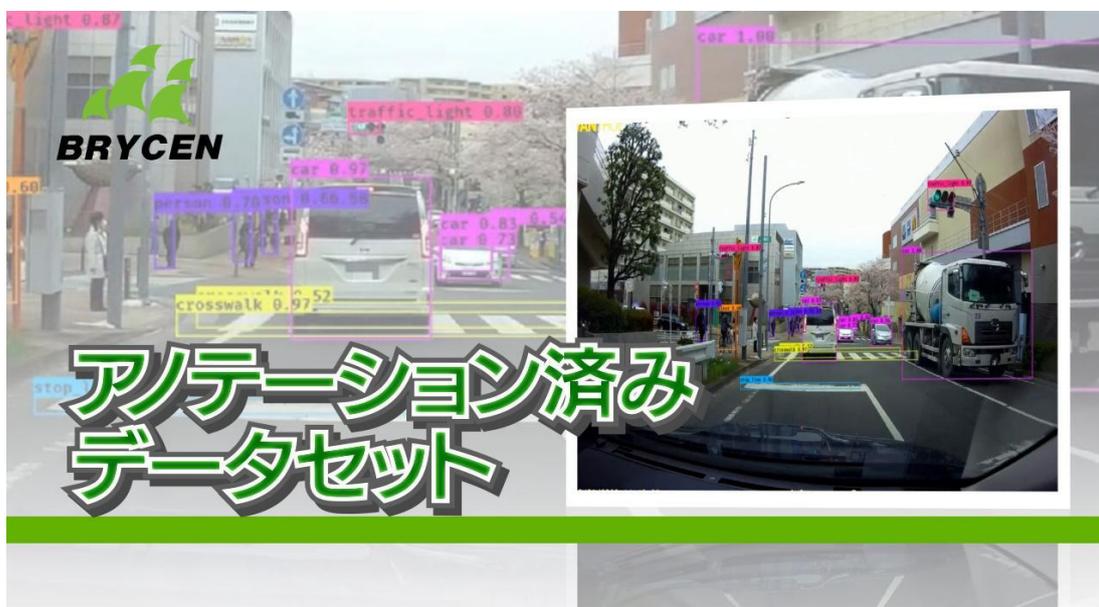
※1 転移学習とは：既存のAIの学習ノウハウを活用することで新しいAIの作成を効率的に行う手法。

■検証方法

- AIモデル：オープンソースのAI (YOLOv3)
- 比較対象：【YOLOv3の学習済みモデル】と【ブライセンアノテーション済みデータセットで転移学習したモデル】の交通系対象物の認識結果を比較

検証結果の詳細は、以下の動画をご覧ください。

[ブライセン走行データセットの有効性検証動画 \(YOLOv3転移学習\)](#)



6. 会社概要・お問い合わせ先

株式会社ブライセンについて

ブライセングループは、コロンブスのように、既成概念にとらわれることなく、最先端の技術とボーダレスなチームで世界を前進させるグローバル開発集団です。カメラソリューションを得意とし、組み込み系・業務・WEB アプリ開発まで幅広く手掛けています。7年前から行っているアノテーション事業では、アノテーションマネージメントサービスを強みとし、GDPR対応やデータセット販売など、新たな取り組みを展開しています。

本社

〒104-6591 東京都中央区明石町 8-1 聖路加タワー30F

代表者:代表取締役社長 藤木 優

設立:1986 年 4 月

URL: <https://www.brycen.co.jp/>



本件に関するお問い合わせ先

エンベデッド開発部 エンベデッド営業部

お問い合わせ: <https://annotation.brycen.co.jp/contact/>

TEL:03-6264-7222