

Technology Reports (特集)

画像認識

ディープラーニング

AI

サービス向上・社会課題解決に貢献するAI

画像認識技術を用いたゴルフスイング解析サービス「GOLFAI」

サービスイノベーション部 川嶋 克明 酒井 俊樹
 ムハマドヒルミ ビン シャピエン いとう ゆうき
 移動機開発部 MUHD HILMI BIN SHAPIEN 伊藤 友基
 ビジネスクリエーション部 わたべ みずき
 渡部 瑞季

近年、ゴルフスイングを自動で解析可能なゴルフレッスンアプリを使用するユーザが増えている。一方で既存のアプリでは高額な専用のセンサが必要となることが多いため、ゴルフ初中級者にとっては利用のハードルが高い。そこでドコモは、より安価かつ簡易なゴルフスイング解析を提供するため、スマートフォンで撮影された動画像のみを用いて解析可能なサービス「GOLFAI」を開発した。これにより、スマートフォン1台でゴルフスイングを解析できるようになり、ユーザは自身のスイングに対するパーソナライズされたアドバイスを受けることが可能になった。

また、本アプリの開発にあたっては、UI/UX開発プロセスの内製化に取り組むことで、より迅速な提供を可能とした。

1. まえがき

GOLFAI [1] は、ユーザからアップロードされたゴルフスイング動画を画像認識技術^{*1}により自動で解析し、解析結果に応じたアドバイスを無料提供するドコモのサービスである。2020年3月よりApp Storeで提供が開始されており、ゴルフの上達に悩んでいる初中級者をメインターゲットとしている。

ゴルフスクールでは、ゴルフインストラクターか

らゴルフ上達のための自身に合った指導を直接受けることができるが、レッスン料やゴルフスクールに通う時間が負担となることがある。そこで近年では、それらの負担を減らすために、スイングを自動で解析できるSmart Golf Lesson [2] などのゴルフレッスンアプリが増えている。一方で既存のアプリでは高額な専用のセンサをゴルフクラブに取り付ける必要があることが多く、ゴルフ初中級者にとっては利用のハードルが高かった。そこでGOLFAIではディープ

©2021 NTT DOCOMO, INC.

本誌掲載記事の無断転載を禁じます。

本誌に掲載されている社名、製品およびソフトウェア、サービスなどの名称は、各社の商標または登録商標。

^{*1} 画像認識技術：画像処理技術や機械学習技術を用いて、画像を機械に理解させ、意味を取り出す技術。

ラーニング^{*2}を用いた画像認識を採用したことで、専用のセンサの代わりにスマートフォンで撮影された動画のみから体の関節位置やスイング軌道のセンシングを行っている。これにより、従来のアプリより安価かつ簡易にゴルフスイングの解析が可能となる。

ゴルフの上達を支援するためには単にスイングの良し悪しの判定のみではなく、どこが悪いのか、どのように練習すれば改善されるかなど、より具体的な情報をユーザに提示する必要がある。ディープラーニングは、一般的に推定結果に対する根拠の明示が難しいため、単体ではユーザが必要とする情報の取得が難しいことが多い。GOLFAIの場合、センシングにはディープラーニングを活用するものの、スイングの評価には専門家のゴルフレッスンの知見を組み合わせた古典的なルールベースの手法を用いることで、この課題を解決した。

本稿では、GOLFAIの画像認識機能の具体例と、本サービスをユーザに迅速に届けるために実施した

UI/UX^{*3}内製化の取組みについて解説する。

2. GOLFAIのシステム概要

GOLFAIのシステム構成を図1に示す。GOLFAIでは、ユーザはまずスマートフォンでゴルフスイング動画を撮影し、それらを専用クラウドにアップロードする。ユーザからアップロードされた動画に対してクラウド上のサーバが画像認識処理を実施し、画像認識結果をJSON (JavaScript Object Notation)^{*4}形式でアプリ側にレスポンスする。アプリ側では返却された画像認識結果を、動画像に重畳表示、もしくはアドバイスのテキスト表示などの形でゴルフ指導を提供する。

3. 画像認識概要

画像上からゴルフ指導に必要な特徴点を抽出するため、GOLFAIでは物体検出と姿勢推定の2つの

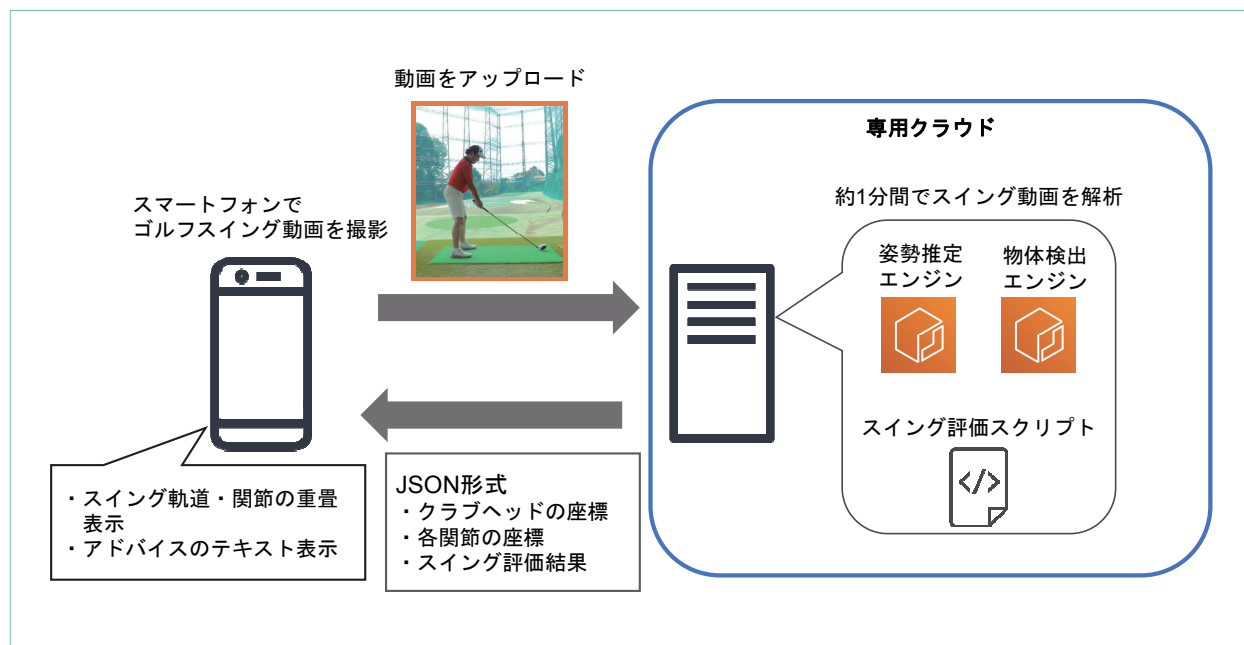


図1 システム構成

^{*2} ディープラーニング：多層のニューラルネットワーク（^{*14}参照）を用いた機械学習の一種。

^{*3} UI/UX：ユーザインタフェース・ユーザエクスペリエンスの略称。

^{*4} JSON：JavaScriptのオブジェクト記述に基づくデータ記述言語。

ディープラーニングを使用している。ディープラーニングの適用にあたり、ゴルフクラブの種類、撮影場所などさまざまな条件下で撮影されたゴルフスイング動画を独自に約1,000本収集している。1つの動画当たり約20秒程度の長さであり、フルHD解像度・120fps (frames per second)*5である。これらのデータを用いてファインチューニング*6を実施することで精度向上を達成した。

また、抽出された特徴点からルールベースの手法により、ゴルフスイングにおける重要なポジションのフレーム*7検出およびスイングタイプの分類を行い、ユーザが解釈しやすい形にした上でアドバイスの提供を可能にしている。

3.1 特徴点検出

(1)物体検出

物体検出は、画像中の事前に定義された物体の位置と種類を推定する技術である。GOLFAIではゴルフクラブヘッドを検出しており、全フレームに対し

て物体検出処理を実施することでスイングの軌道を算出している(図2(a))。

近年、物体検出技術ではディープラーニングを用いた学習ベースの手法が高い精度を達成しており、大きく分けて以下の2つの手法がある。

①Two stage detection framework

物体の候補領域を提案する前処理の後に、物体の種類と領域の座標を推定する2段階の処理を行う手法である。2ステージのフレームワークでは、まずselective search*8やCNN (Convolutional Neural Network)*9で取得される特徴マップ*10を用いて候補領域を生成する。次に各候補領域から特徴量*11を算出し、サポートベクターマシーン*12などの分類器*13を用いて物体の種類や詳細な座標を推定する。代表的なニューラルネットワーク*14にSPPNet [3], Fast RCNN [4], Faster RCNN [5] などがある。

②One stage detection framework

物体の候補領域、種類、領域の座標を1つの



(a)物体検出



(b)姿勢推定

図2 特徴点抽出例

*5 fps：単位時間当りの静止画像の数。

*6 ファインチューニング：一度学習されたモデルのパラメータを初期値に設定し、別のデータセットを用いて再度学習を行い、パラメータを微調整する手法。

*7 フレーム：映像の元になる静止画像の1コマ。

*8 selective search：画像中の類似するピクセルをまとめていく

ことで物体の候補領域を算出する手法。

*9 CNN：多次元配列に対して縦方向と横方向に走査しながら、特定のサイズのベクトルを掛け合わせる処理が導入されたニューラルネットワーク(*14参照)であり、画像認識によく用いられる。

*10 特徴マップ：本稿では入力画像をCNNで処理した結果得られる多次元配列を指す。

ニューラルネットワークで推定する手法である。1ステージのフレームワークでは候補領域の座標と種類を一度の処理で直接推定する。候補領域の生成なども含めたすべての計算を、単一のニューラルネットワークで行うシンプルな処理構造が特徴である。このため、一般的に2ステージのフレームワークより学習や推論が高速になると言われている。代表的なニューラルネットワークにYOLO [6], SSD [7], CornerNet [8] などがある。

GOLFAIではマシンリソースや処理時間を削減するため、図3に示すように、1ステージのフレームワークを採用している。このモデルでは3つの検出レイヤを持っており、それぞれ大中小と異なるサイズの物体を検出する役割をもつため、画像上の物体のサイズに左右されない高精度な物体検出が可能

となっている。

(2)姿勢推定

姿勢推定は、画像に映る人間の関節位置を2次元座標や3次元座標として推定する技術である。推定対象は事前に定義・学習する必要があるが、GOLFAIではゴルフスイングの評価に必要となる肘や手首、膝などの関節の2次元座標を算出している（図2(b)）。姿勢推定技術では、近年ディープラーニングを用いた学習ベースの推定手法の技術開発が進められており、高い精度を達成している。ディープラーニングを用いた姿勢推定技術は、主に以下の2つのアプローチに分類される。

①Top-downアプローチ

物体検出技術を用いて画像中から人物を検出し、検出された人物それぞれから関節を検出する手法である。人物検出、関節の検出を独立して行うシンプルな手法であり、人物検出の精度

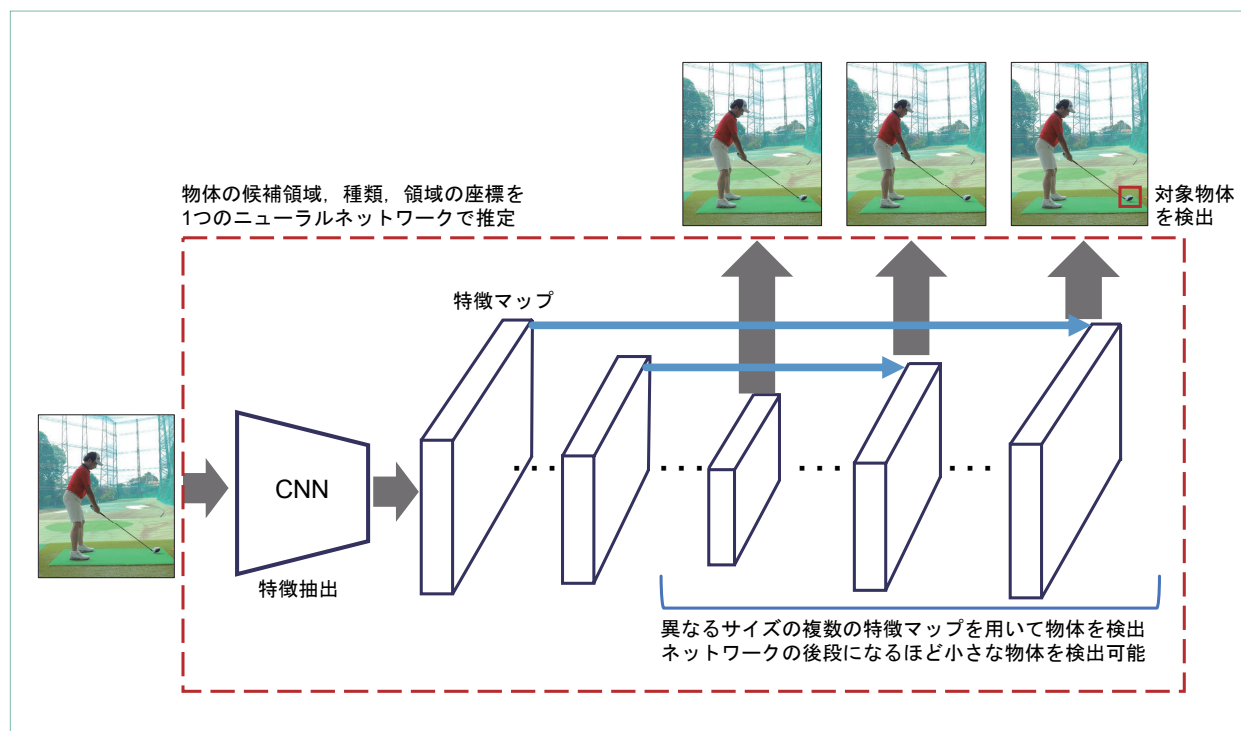


図3 物体検出の処理概要

- *11 特徴量：データから抽出される、そのデータの特徴付けの量（数値）のこと。
- *12 サポートベクターマシン：パターン識別に用いられる機械学習手法の一種。カーネル・トリックと呼ばれる方法で線形分離が不可能な問題に対しても適用可能な手法である。
- *13 分類器：入力をその特徴量に基づいてあらかじめ定められた分

類先のいずれかに分類する装置。

- *14 ニューラルネットワーク：人間の脳の構造を模した数理モデル。

を向上させることで姿勢推定の精度向上にもつながる。一方で、画像中の人物数に比例して計算量が増えるデメリットがある。代表的なニューラルネットワークにDeep Pose [9], Cascaded Pyramid Network [10], High-Resolution Network [11] などがある。

②Bottom-upアプローチ

画像中に存在する人間の関節をすべて検出した後にそれらを人物ごとにつなぎ合わせる手法である。単一のニューラルネットワークを用いて一度に画像中の全人物の関節点を検出するため、人物数によって推論速度が変わりにくい。一方で、関節を検出できたとしても、関節間のつながり合わせが失敗してしまい精度が低くなることがある。代表的なニューラルネットワークにDeep Cut [12], Open Pose [13], PersonLab [14] などがある。

GOLFAIでは処理時間の増加を防ぐためにBottom-upの手法を採用し、関節の位置推定と関節間の繋がり推定の2つのニューラルネットワークで構成されるモデルを使用している（図4）。関節の位置だけでなく、向きの情報も学習することで検出された関節を高精度に人物ごとにつなぎ合わせる事が可能となっている。

3.2 スイング分析

(1)スイングポジションの推定

ゴルフスイングにおいて重要と考えられているアドレス（A）、トップ（T）、インパクト（I）、フィニッシュ（F）の4つのポジションのフレームを検出している。これにより、ユーザは図5のようにA、T、I、Fボタンをタッチするだけで、所望のポジションのフレームに再生位置を移動し、スイングを確認することが可能となる。ゴルフスイングには個

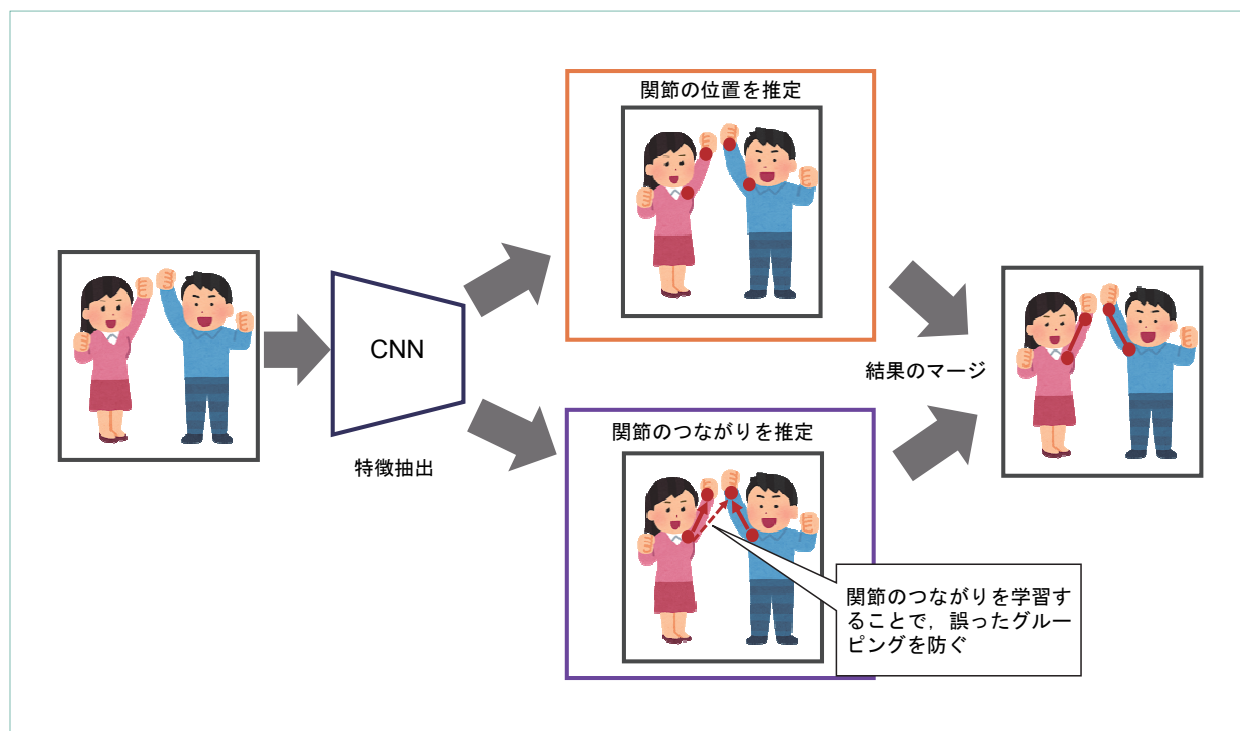


図4 姿勢推定の処理概要

人差はあるものの、その動きには規則性がある事に着目し、ルールベースでポジションを推定する。

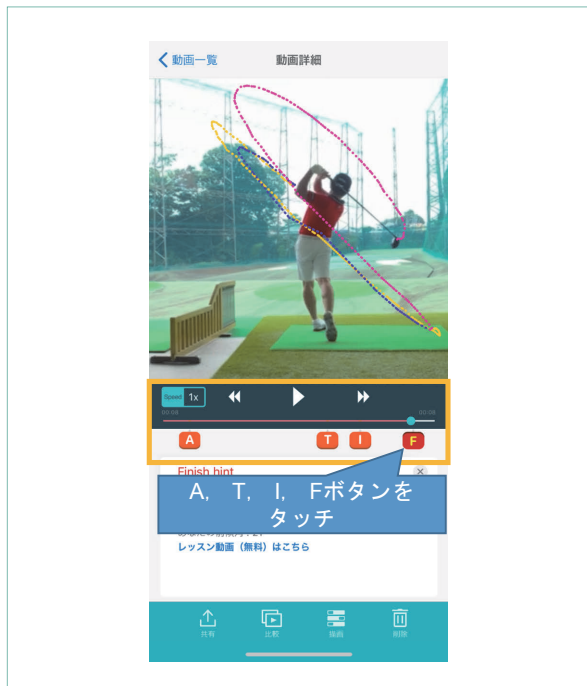


図5 ポジション推定結果例

(2)スイング評価

物体検出、姿勢推定技術によって算出された特徴点を用いてスイングを評価している。専門家とスイングの評価ポイントおよびその評価方法について議論を重ね、独自に画像認識によるスイング評価方法を開発した。専門家の知見を数式に落とし込みルールベースの手法で評価している。以下にGOLFAIにおける評価ポイントの例を示す。

(a)体の前傾角度の維持 (図6(a))

体の前傾角度がスイング中に一定であると良いスイングと言われている。この観点で良いスイングかを評価するため、ポジションごとに腰中央と首の座標を用いて前傾角度を算出し、ポジション間の前傾角度の変化量が一定数値以上だったときにユーザへ改善のアドバイスを提供している。ここで、一定数値は専門家の知見に基づき決定している。

(b)オーバースイング (図6(b))

オーバースイングと呼ばれる過度な振りかぶりは、スイング軌道が不安定になるため一般的



図6 スイング評価例

に良くないと言われている。この観点で良くないスイングかを評価するため、トップポジションのときにゴルフクラブヘッドが体よりボール側かつ、首の位置よりも低い位置にあるときオーバースイングと判定している。

(c)チキンウィング（図6(c)）

フォロースルーの際に左肘（左打ちなら右肘）が引けてしまっているスイングはチキンウィングと呼ばれている。チキンウィングでない正常なスイングの肘の動きをユーザの側方から見たとき、フォロースルーのときには肘が体に隠れて見えないという専門家の知見に基づき、インパクト以降にゴルフクラブヘッドが体を横切るよりも前に左肘（左打ちなら右肘）が画像上から検出された場合にチキンウィングと判定している。

上記のようにスイングの評価に関してはディープラーニングではなく、ルールベースの手法を組み合わせることで、ユーザが解釈しやすい形でアドバイスを実現している。

4. UI/UX検討プロセス内製化によるGOLFAIの開発期間短縮／使いやすさ向上の取り組み

4.1 サービス開発の課題

近年サービスアプリ市場においては競争が激しく、ユーザ獲得のためには、いかに早く、優れたUI/UXをもつアプリを提供するかが課題となっている。そのため、新規技術を活かしたGOLFAIを、UI/UXを最大に考慮しつつ、効率的、かつスピーディに開発し、市場に投入する必要がある。

一方、現状のドコモサービス開発においては、検討開始からサービス提供までに平均で5～6カ月の期間を要するため、今回、UI/UX・デザイン検討お

よびアウトプットを行うプロセス（以下、UI/UXプロセス）の改善に着目した。UI/UXプロセスの改善により、下記2点の課題を解決する。

・UI/UXの解釈上の齟齬

UI/UXを向上させるには、ユーザ提供価値についてステークホルダ間（アプリ開発側）で認識を一致させた上でユーザの反応を意識しながら試行錯誤を繰り返す必要があるが、認識を一致させることが難しい。

・情報伝達の非効率性

ステークホルダ間でのコミュニケーションに十分な配慮が必要であるが、外部のデザイン会社と連携する場合はコミュニケーションを密に取り難く、円滑に進めにくい。

ステークホルダの増加により、意思疎通が煩雑になることやコミュニケーションのタイムロスが要因となり、開発スピードの低下につながる。そのため、「UI/UXプロセス内製化」の取組みをGOLFAI開発において試験的に実施し改善を図ることとした。

4.2 UI/UXプロセス内製化

GOLFAI開発では、UI/UXプロセスにおいて外部のデザイン会社への委託は行わず、開発部門にて内製的に検討および対応を行った。

図7に示すように、全ステークホルダが参加したワークショップで定めたコンセプトをベースとして、機能案・デザイン案を開発部門にて作成・提案した。開発部門がUI/UXプロセスを担うことでステークホルダを減らしたことに加え、企画部門・開発ベンダ双方の状況を把握しているため、「実現したい要望」と「開発負荷」のバランスを取った形で効果的・効率的な提案を行うことができた。また、改善サイクルを回すにあたっても、橋渡し役として企画部門との仕様案のすり合せ・開発ベンダとの実現性検討をスピーディに行った。



図7 UI/UXプロセス内製化による開発工程

具体的には、デザインデータの作成・受渡しに広く活用されるツールであるSketch^{*15} [15] にてデザイン案を作成し、複数回の改善を繰り返しながら最終版を完成させた。操作可能なアプリが無い段階では、プロトタイピング^{*16}を容易に作成できるツールであるPrott^{*17} [16] にて画面遷移を擬似的に再現することで、実操作に近い形でのデザイン確認を開発初期段階から可能とし、改善サイクルの高速化につなげた。

開発部門がUI/UXプロセスを担った結果として、GOLFAIのリリースを4カ月で実現させた。その中でUI/UX検討プロセスにておおむね1カ月ほどの期間短縮寄与があったと考えられる。新規サービスをいち早く世の中にリリースすることを目的とし、かつ、ステークホルダ減少の効果が高い中小規模のプロジェクトに対して、UI/UXプロセス内製化の効果が高いと考えられる。

5. あとがき

本稿では、GOLFAIで使用されている、ゴルフク

ラブヘッドを検出する物体検出技術と、体の関節位置を推定する姿勢推定技術について解説し、またそれらの技術を用いて抽出された特徴点を活用したスイングの評価方法、その技術をいち早くユーザに届けるためのUI/UXプロセス内製化の取組みを解説した。ゴルフスイングの上達を支援するには、スイングの良し悪しの判定のみでなく、どこが悪いのか、改善方法など具体的なアドバイスが必要となる。本サービスでは、最新のディープラーニングを用いた高精度な特徴点検出と専門家の知見を組み合わせることで、ユーザに有益なアドバイスを提供している。今後は逆光などの悪い撮影条件下でも高精度に認識可能な画像認識技術の開発や、サッカーなど他のスポーツへの本技術の活用、UI/UXプロセス内製化の広い適用に関して検討を行う予定である。

文 献

- [1] GOLFAI 公式サイト.
<https://golfai.jp/>
- [2] SONY スマートゴルフレッスン.
<https://smartsports.sony.net/golf/product/1G/JP/ja/>
- [3] K. He, X. Zhang, S. Ren and J. Sun: "Spatial Pyramid

*15 Sketch: オランダのBohemian Coding社がサービス提供しているベクターグラフィックエディターのこと。一般的にアプリやWebのUIデザインツールとして利用されている。

*16 プロトタイピング: 検証・フィードバックのため、検討早期に実働モデルを製作する手法。

*17 Prott: 日本のGoodPatch社がサービス提供しているアプリや

Webデザインに特化したプロトタイピングツールのこと。

- Pooling in Deep Convolutional Networks for Visual Recognition,” In ECCV, 2014.
- [4] R. Girshick : “Fast R-CNN,” In ICCV, 2015.
- [5] S. Ren, K. He, R. Girshick and J. Sun : “Faster R-CNN : Towards Real Time Object Detection with Region Proposal Networks,” In NIPS, 2015.
- [6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi : “You only look once : Unified, real time object detection,” In CVPR, 2016.
- [7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu and A. Berg : “SSD : Single Shot Multibox Detector,” In ECCV, 2016.
- [8] H. Law and J. Deng : “CornerNet : Detecting Objects as Paired Keypoints,” In ECCV, 2018.
- [9] A. Toshev and C. Szegedy : “Deeppose : Human Pose Estimation via Deep Neural Networks,” In CVPR, 2014.
- [10] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu and J. Sun : “Cascaded Pyramid Network for Multi-Person Pose Estimation,” In CVPR, 2018.
- [11] K. Sun, B. Xiao, D. Liu and J. Wang : “Deep High-Resolution Representation Learning for Human Pose Estimation,” In CVPR, 2019.
- [12] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler and B. Schiele : “Deepcut : Joint Subset Partition and Labeling for Multi Person Pose Estimation,” In CVPR, 2016.
- [13] Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh : “Realtime Multiperson 2d Pose Estimation using Part Affinity Fields,” In CVPR, 2017.
- [14] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson and K. Murphy : “PersonLab : Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model,” In ECCV, 2018.
- [15] Sketchホームページ.
<https://www.sketch.com/>
- [16] Prottホームページ.
<https://prottapp.com/ja/>