Technology Reports (Special Articles)
Image Recognition Deep Learning AI
Special Articles on AI for Improving Services and Solving Social Problems

GOLFAI: Golf Swing Analysis Service Using Image Recognition Technology

Service Innovation Department Katsuaki Kawashima Toshiki Sakai
Communication Device Development Department Muhd Hilmi Bin Shapien Yuki Ito
Business Creation Department Mizuki Watabe

Golf lesson apps that can automatically analyze one's golf swing are becoming increasingly popular. Many of these apps, however, require high-priced dedicated sensors that can deter their use for beginners and intermediate players. To provide low-cost and simplified golf swing analysis, NTT DOCOMO developed GOLFAI, a service that can analyze golf swings using only the video taken with a smartphone. With GOLFAI, a golf swing can be analyzed with a single smartphone and the user can receive personalized advice on one's swing.

In addition, we speeded up the development and provision of this app by adopting an in-house UI/UX development process.

1. Introduction

GOLFAI [1] is an NTT DOCOMO service that automatically analyzes golf-swing video uploaded by the user through image recognition technology*1 and uses the results of that analysis to provide free advice to the user. It has been available from the App Store since March 2020 as a service targeting mainly beginners and intermediate players who would like to improve their golf swing.

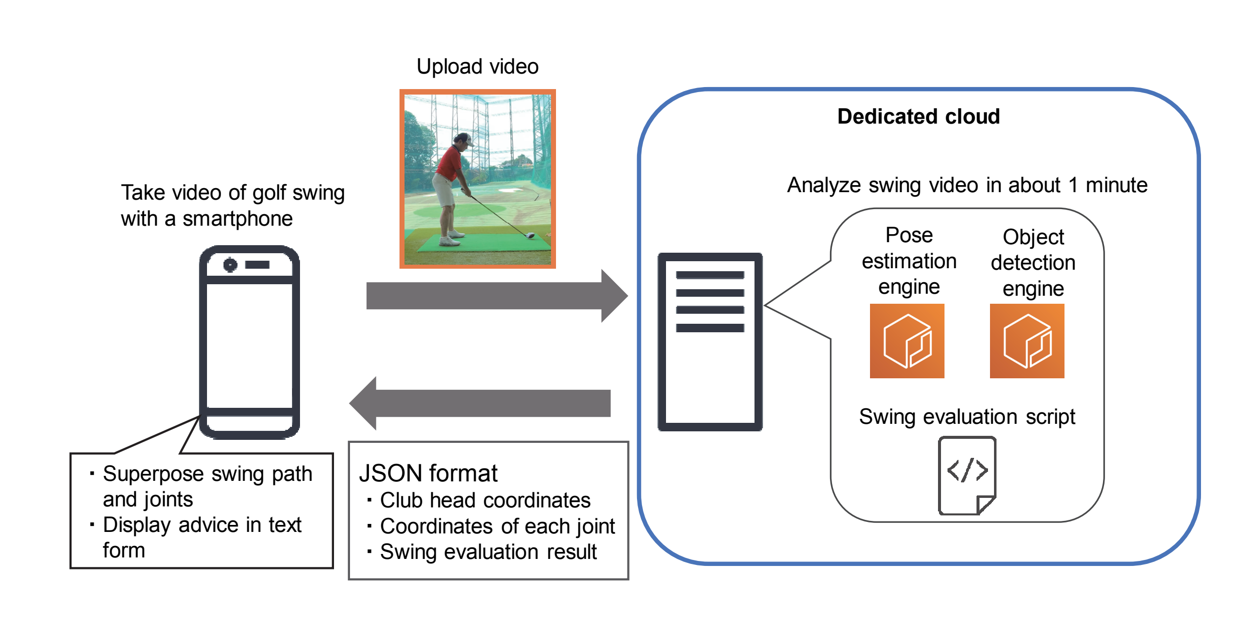
A golfer can attend golf school and receive personalized instruction directly from a golf instructor to improve one's level of play, but lesson fees, travel time to and from the school, etc. can place a burden on the student. To ease this burden, recent years have seen an increasing number of golf lesson apps such as Smart Golf Lesson [2] that can automatically analyze the user's swing. Many of these apps, however, require that high-priced dedicated sensors be attached to the golf club, which has presented a high hurdle to their use for beginners and intermediate players. Against this background, GOLFAI adopts image recognition using deep learning*2 to sense the body's joint positions and swing path from only video images captured by a smartphone as an alternative to dedicated sensors. In this way, GOLFAI enables low-cost and simplified golf-swing analysis compared with existing apps.

To help a person become better at golf, it is not simply a matter of determining whether that person's swing is good or bad—the user of a golf lesson app needs to be presented with more detailed information such as what part of the swing is bad and what kind of practice should improve the swing. In deep learning, it is generally difficult to describe the grounds for reaching a certain inference, and as a result, it is frequently difficult to obtain the information needed by the user simply on the basis of deep learning. GOLFAI, although using deep learning for sensing purposes, solves this problem by using a classic rule-based technique that incorporates the knowledge in expert-taught golf lessons for evaluating a swing.

In this article, we present specific examples of GOLFAI image recognition functions and describe our in-house UI/UX*3 development process that we adopted to speed up the delivery of this service to users.

2. GOLFAI System Overview

The GOLFAI system configuration is shown in **Figure 1**. In GOLFAI, the user begins by taking



- Superpose swing path and joints
- Display advice in text form

JSON format
- Club head coordinates
- Coordinates of each joint
- Swing evaluation result

Figure 1　System configuration

*1 Image recognition technology: Technology for mechanically understanding images and extracting meaning using image processing technology, machine learning technology, etc.

*2 Deep learning: A type of machine learning that uses a multi-layered neural network (see *14).

*3 UI/UX: Abbreviations of "user interface" and "user experience."

video of his or her golf swing using a smartphone and uploading the video to a dedicated cloud. The GOLFAI system then performs image recognition processing against the uploaded video on a server on that cloud and returns the results of image recognition to the app in JavaScript Object Notation (JSON)*4 format. Finally, the app superposes those results of image recognition on the video or provides golf instruction by displaying advice in text form.

3. Image Recognition Overview

GOLFAI uses object detection and pose estimation as two types of deep learning for extracting from images the feature points needed for golf instruction. In applying deep learning, we independently collected about 1,000 clips of golf-swing video taken under a variety of conditions such as type of golf club and shooting location. Each clip is approximately 20 seconds long and is shot with Full HD resolution at 120 frames per second (fps)*5. We succeeded in improving the accuracy of feature-point extraction by performing fine-tuning*6 with this data.

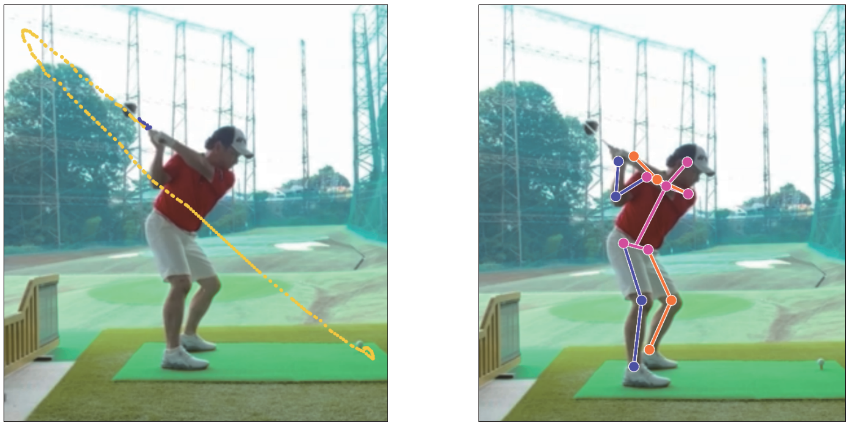
GOLFAI uses extracted feature points and a rule-based technique to detect a frame*7 that corresponds to an important position in a golf swing and to categorize swing type. It then converts this information into a form that is easy for the user to understand and offers advice.

3.1 Feature Point Detection

1) Object Detection

Object detection refers to technology that estimates the position and class of a predefined object within an image. In GOLFAI, the system detects the golf club head and calculates the path of the swing by performing object-detection processing against all frames (**Figure 2** (a)).

In object detection technology, learning-based techniques using deep learning have reached high levels of accuracy in recent years. These techniques


(a) Object detection

(b) Pose estimation

Figure 2 Example of extracting feature points

*4 JSON: A data description language based on JavaScript object notation.

*5 fps: Number of still images per unit time.

*6 Fine-tuning: A technique that sets parameters of a model that has already been learned once to initial values and then uses a different dataset to relearn that model and finely adjust those parameters.

*7 Frame: One of the many single still images that make up video.

can be broadly divided into the following two types.

(1) Two-stage detection framework

After preprocessing that proposes object candidate areas, this technique performs two-stage processing to estimate the class of object and area coordinates. This two-stage framework begins by generating candidate areas using a feature map*8 obtained through a selective search*9 or Convolutional Neural Network (CNN)*10. It then calculates features*11 from each candidate area and estimates object type and detailed coordinates using a support vector machine*12 or other type of classifier*13. In this framework, typical neural networks*14 include SPPNet [3], Fast RCNN [4], and Faster RCNN [5].

(2) One-stage detection framework

This technique estimates object candidate areas and object class and area coordinates using a single neural network. In short, it is a one-stage framework that directly estimates candidate-area coordinates and class in a single process. It has a simple processing structure that can perform all calculations including those for generating candidate areas using a single neural network. For this reason, a one-stage detection framework is said to be faster than a two-stage detection framework in terms of learning and inference. Here, typical neural networks include YOLO [6], SSD [7], and CornerNet [8].

For GOLFAI, we have adopted a one-stage framework as shown in **Figure 3** to reduce machine resources and processing time. This model has three detection layers each having the role of detecting an object of different size, that is, a large, medium, or small object. In this way, it becomes possible to perform high-accuracy object detection regardless of the size of the object in the image.

2) Pose Estimation

Pose estimation is technology for estimating the positions of human joints in an image in the form of two-dimensional or three-dimensional coordinates. The targets of estimation must be defined and learned beforehand, and in GOLFAI, the system calculates the two-dimensional coordinates of the joints needed for evaluating a golf swing such as elbows, wrists, and knees (Fig. 2 (b)). Pose estimation technology has been reaching high levels of accuracy owing to recent advances in learning-based estimation techniques using deep learning. In this regard, pose estimation technology using deep learning can be mainly divided into the following two approaches.

(1) Top-down approach

This is a technique that detects persons within an image using object detection technology and detects the joints of each of those persons. It's a simple technique that performs person detection and joint detection separately and that can improve the accuracy of pose estimation by improving the accuracy of person detection. Its weak point is that its computational cost increases in proportion to the number of persons in the image. Here, typical neural networks include DeepPose [9], Cascaded Pyramid Network [10], and High-Resolution Network [11].

(2) Bottom-up approach

This is a technique that first detects all human joints existing in the image and then

[8] Feature map: In this article, a multidimensional array obtained from the results of processing an input image by a CNN.

[9] Selective search: A technique that calculates object candidate areas by grouping similar pixels in an image.

[10] CNN: A type of neural network (see *14) that introduces processing for multiplying vectors of specific sizes while scanning a multidimensional array in the vertical and horizontal directions. CNN is widely used in image recognition.

[11] Feature: An amount (a numeric value) extracted from data to characterize that data.

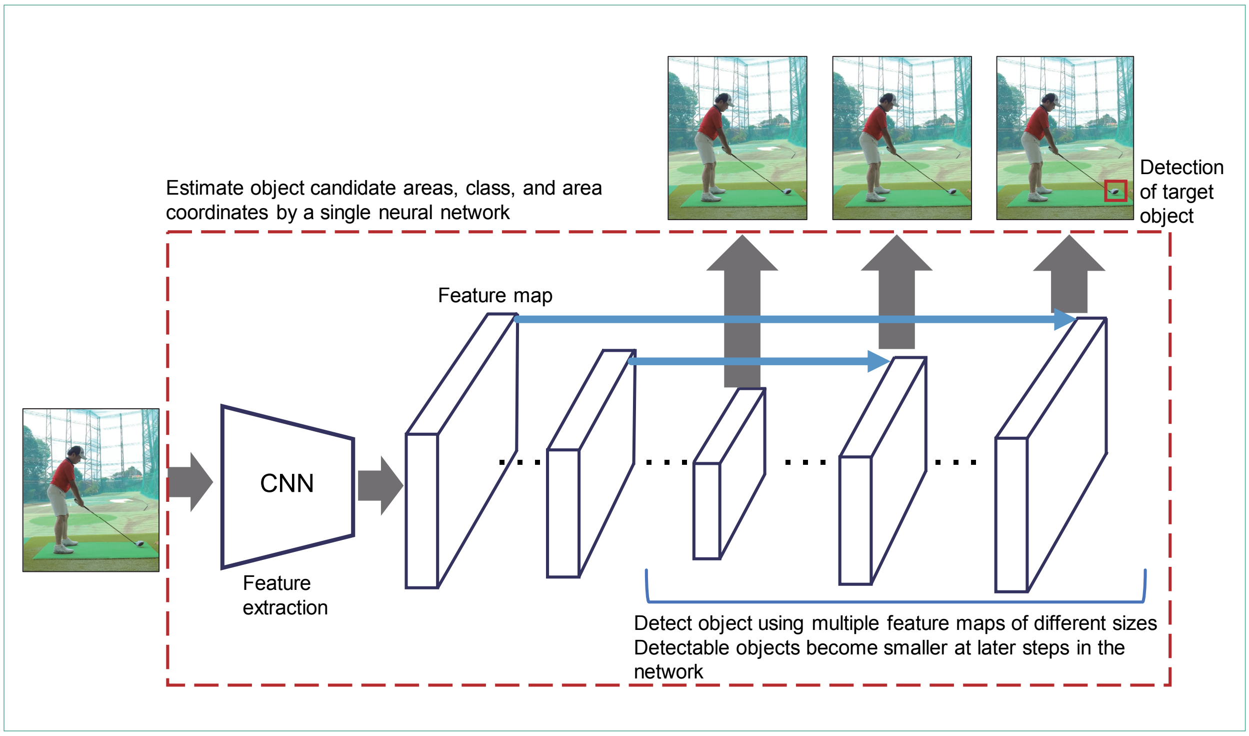

Estimate object candidate areas, class, and area coordinates by a single neural network
Detection of target object
Feature map
CNN
Feature extraction
Detect object using multiple feature maps of different sizes
Detectable objects become smaller at later steps in the network

Figure 3 Overview of object detection processing

connects those joints on a person-by-person basis. Since it detects the joint points of all persons in the image at one time using a single neural network, inference speed is hardly changed by the number of persons present. On the other hand, accuracy may suffer even if joints have been correctly detected if failures occur in connecting the joints. Typical neural networks here include Deep Cut [12], Open Pose [13], and PersonLab [14].

For GOLFAI, we decided to use the bottom-up approach to prevent an increase in processing time and incorporated a model consisting of two neural networks to estimate joint positions and joint connections (**Figure 4**). In this model, learning not only joint position but joint orientation as well enables high-accuracy connection of detected joints for each person.

3.2 Swing Analysis

1) Estimation of Swing Position

GOLFAI detects the frames corresponding to the four positions of address (A), top (T), impact (I), and finish (F) considered to be key components of the golf swing. As shown in **Figure 5**, this enables the user to check one's swing by simply touching the A, T, I, and F buttons to move the playback position to the frame of the desired position. Although individual differences can be found in golf swings, GOLFAI focuses on standard elements of swing movement and estimates position in a rule-based manner.

*12 Support vector machine: A machine-learning method used in pattern recognition. It can be applied even to problems that are not linear separable by the "kernel trick" method.

*13 Classifier: An algorithm that classifies input into one of a number of predetermined classifications based on features.

*14 Neural network: A mathematical model that mimics the structure of the human brain.

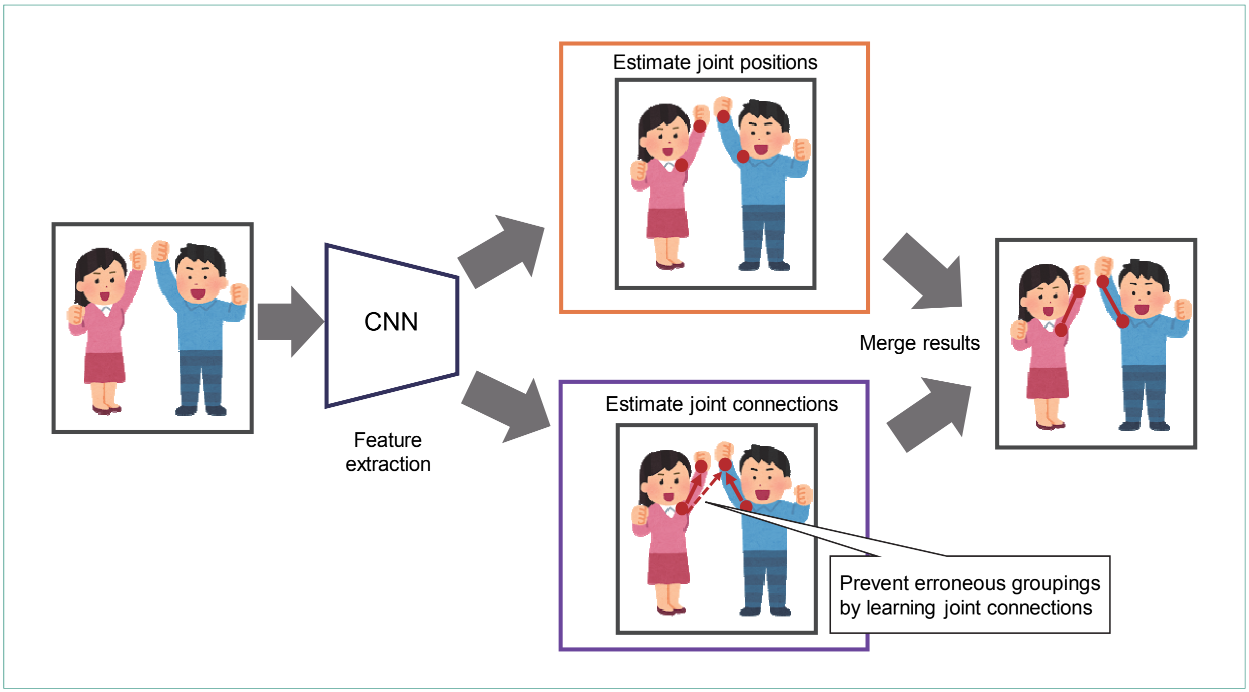

Estimate joint positions
CNN
Feature extraction
Merge results
Estimate joint connections
Prevent erroneous groupings by learning joint connections

Figure 4　Overview of pose estimation processing

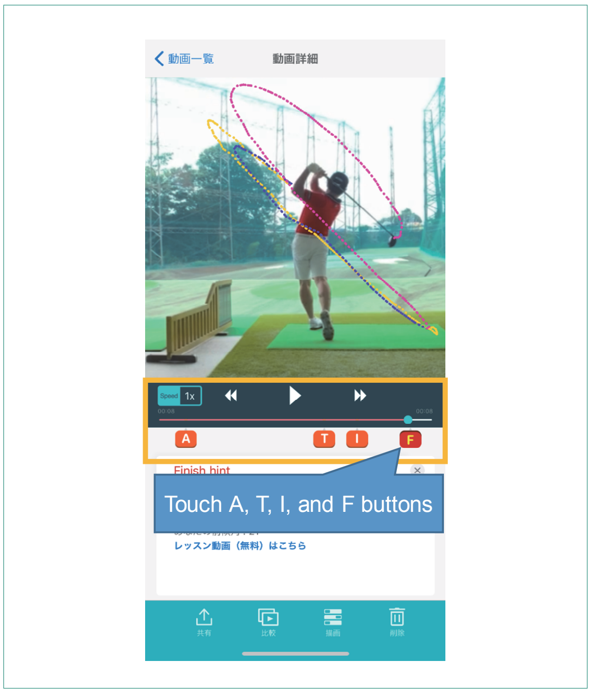


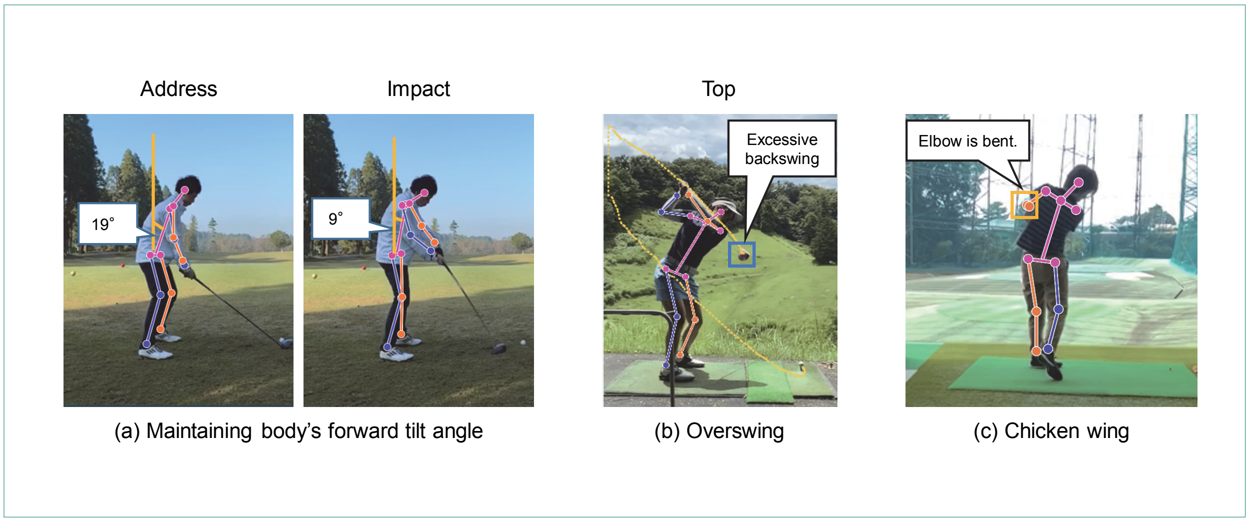
Figure 5　Example of position estimation results

2) Swing Evaluation

This process evaluates a swing using the feature points calculated using object-detection and pose-estimation technologies. We developed a swing evaluation method by holding discussions with experts on swing evaluation points and their evaluation methods and by applying proprietary image recognition techniques. Through these discussions, we broke down expert knowledge into mathematical formulas and used a rule-based technique to evaluate swings. The following presents examples of evaluation points in GOLFAI.

(a) Maintaining body's forward tilt angle (**Figure 6** (a))

It is said that keeping the body's forward tilt angle fixed during a swing makes for a good swing. To evaluate whether a swing is good from this viewpoint, GOLFAI first



(a) Maintaining body's forward tilt angle

(b) Overswing

(c) Chicken wing

Figure 6 Example of swing evaluation

calculates the forward tilt angle for each position using the coordinates of the hip and the neck. Then, if the amount of change in the forward tilt angle between positions should exceed a fixed value, it offers the user advice on how to improve the swing. This fixed value is determined on the basis of expert knowledge.

(b) Overswing (Fig. 6 (b))

It is also said that an excessive backswing called an "overswing" is generally not good since it can induce instability in the swing path. To evaluate whether a swing is bad from this viewpoint, GOLFAI judges that an overswing occurs when the golf club head at top position is located on the ball side relative to the body at a point lower than the neck.

(c) Chicken wing (Fig. 6 (c))

A swing in which the left elbow (right elbow for a left-handed player) is bent in the follow-through is called a "chicken wing." According to experts, the elbow will be hidden by the user's body and out of view during the follow-through when observing elbow movement in a normal swing (no chicken wing) from a position lateral to the user. Based on this knowledge, GOLFAI judges that a chicken-wing swing occurs after impact when the left elbow (right elbow for a left-handed player) is detected in the image before the golf club head traverses the body.

In the above way, GOLFAI can offer the user advice in an easy-to-understand format by incorporating a rule-based technique in swing evaluation instead of using deep learning.

4. Shortening GOLFAI Development Period and Improving Ease-of-use by In-house UI/UX Studies

4.1 Service Development Issues

In a service app market that has become extremely competitive in recent years, how to provide an app with a superb UI/UX as quickly as possible to acquire users has become a major issue. We therefore felt the need to efficiently and quickly develop and market GOLFAI, a service that makes use of new technologies, while giving full consideration to the UI/UX.

However, service development at NTT DOCOMO currently takes five to six months on average from initial studies to service provision, so this time, we focused on studying UI/UX design and improving the output process (hereinafter referred to as "UI/UX process"). Improving the UI/UX process solves the following two issues.

- Conflicts in interpreting UI/UX

  To improve the UI/UX, a process of trial and error must be repeated to reach an agreement among stakeholders (on the app-development side) as to the value provided to the user while taking user response into account. It is difficult, however, to reach a common understanding in this way.

- Inefficiencies in communicating information

  Communication among stakeholders must be sufficiently considered, but when working with an outside design company, maintaining close communication and achieving a smooth development process is difficult.

As the number of stakeholders increases, exchanging information becomes more complicated and more time is taken up by communication activities, all of which can slow down the development process. Taking this into consideration, we tested an "in-house UI/UX process" in GOLFAI development with the aim of improving this process.

4.2 In-house UI/UX Process

In GOLFAI development, we studied and applied an in-house UI/UX process in the development department without consigning any work to an outside design company.

As shown in **Figure 7**, the development department created and proposed functions and a design based on concepts established at a workshop attended by all stakeholders. By assigning the UI/UX process to the development department, the number of stakeholders could be reduced and conditions at both the planning department and development vendor could be readily understood. This made it possible to enforce more effective and efficient proposals in a form that achieved a balance between "service requirements" and "development load." In this way, the development department took on the role of a mediator that could speed up proposals for specifications with the planning department and feasibility studies with the development vendor even when repeating the improvement cycle.

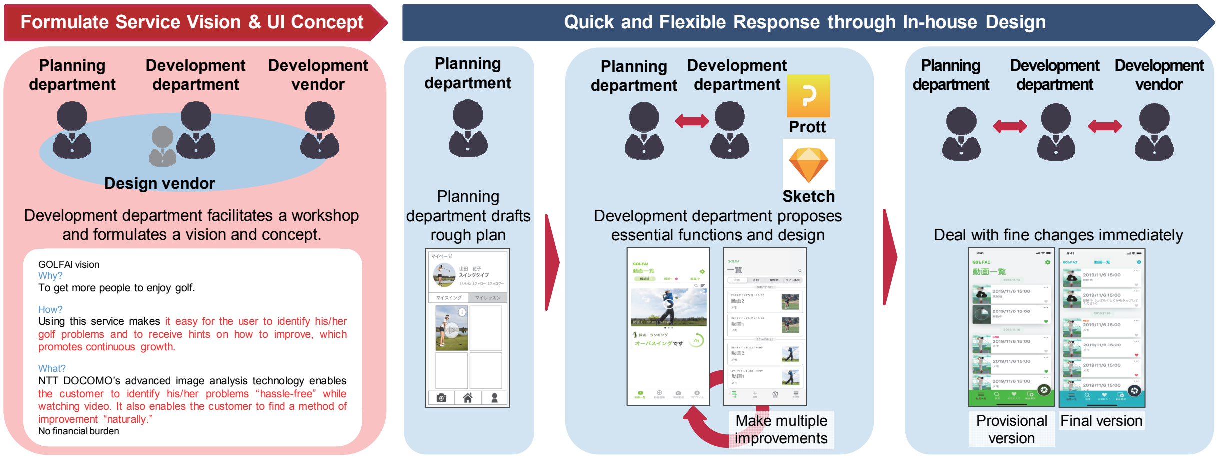
Specifically, the development department created a design proposal using Sketch*15 [15], a tool that is widely used for creating and delivering design data, and completed a final version of GOLFAI after repeating the improvement cycle multiple times. Here, at the stage with no operable app, screen transitions were simulated using Prott*16 [16], a tool that simplifies prototyping*17. This made it possible to check the design in a form close to actual operation from the initial development stage, which had the effect of speeding up the improvement cycle.

As a result of adopting an in-house UI/UX process in the development department, we were able to release GOLFAI in just four months. We

*15 Sketch: A vector graphics editor provided by Bohemian Coding in the Netherlands. It is generally used as a UI design tool for apps and the web.

*16 Prott: A prototyping tool specifically for app and web design provided by Goodpatch Inc. in Japan.

*17 Prototyping: An early sample, model, or release of a product to test and evaluate a concept.



Development department facilitates a workshop and formulates a vision and concept.

GOLFAI vision

Why?

To get more people to enjoy golf.

How?

Using this service makes it easy for the user to identify his/her golf problems and to receive hints on how to improve, which promotes continuous growth.

What?

NTT DOCOMO's advanced image analysis technology enables the customer to identify his/her problems "hassle-free" while watching video. It also enables the customer to find a method of improvement "naturally."

No financial burden

Development department proposes essential functions and design

Figure 7　Development flow by an in-house UI/UX process

consider that the UI/UX study process contributed to this shortening of the development period by about one month. In short, an in-house UI/UX process can be highly effective if the objective is to release a new service as quickly as possible and if the project is small or medium in size in which the effect of reducing the number of stakeholders is high.

5. Conclusion

In this article, we described two key technologies used by GOLFAI: object detection technology for detecting the golf club head and pose estimation technology for estimating the positions of joints on the human body. We also described a swing evaluation method using the feature points extracted by those technologies and explained our adoption of an in-house UI/UX process for speeding up the delivery of this service to users. To help improve a golf swing, it is not sufficient to merely judge whether a swing is good or bad. It is also necessary to give specific advice such as what part of the swing is bad and what method can be used to improve the swing. This service provides the user with helpful advice by combining high-accuracy feature point detection using the latest deep learning techniques and expert knowledge. Going forward, we plan to develop image recognition technology that can achieve high-accuracy recognition even under poor shooting conditions as in a backlit scene. We also plan to study the application of this technology to other sports such as soccer and means of widening the application of this in-house UI/UX process.

REFERENCES

[1] GOLFAI Official Site (In Japanese).

https://golfai.jp/

[2] SONY Smart Golf Lesson (In Japanese).
https://smartsports.sony.net/golf/product/1G/JP/ja/

[3] K. He, X. Zhang, S. Ren and J. Sun: "Spatial Pyramid 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 Multi-person 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 website.
https://www.sketch.com/

[16] Prott website.
https://prottapp.com