Semantic Segmentation of Golf Courses for Course Rating Assistance

Mortensen, Jesper Kjærgaard; Soares Matthiesen, Vinicius; González de Frutos, Jacobo; Bujdosó, Kata; Christensen, Jesper Thøger; Møgelmose, Andreas

Published in:
Proceedings - 2023 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2023

DOI (link to publication from Publisher):
10.1109/ICMEW59549.2023.00050

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

Semantic Segmentation of Golf Courses for Course Rating Assistance

1st Jesper Kjærgaard Mortensen
Department of Electronic Systems
Aalborg Univeristy
Aalborg, Denmark
jkjar18@student.aau.dk

2nd Vinicius Soares Matthiesen
Department of Electronic Systems
Aalborg University
Aalborg, Denmark
vmatth19@student.aau.dk

3rd Jacobo González de Frutos
Department of Electronic Systems
Aalborg University
Aalborg, Denmark
jgonza22@student.aau.dk

4th Kata Bujdosó
Department of Electronic Systems
Aalborg University
Aalborg, Denmark
kbujdo22@student.aau.dk

5th Jesper Thøger Christensen
Department of Architecture, Design and Media Technology
Aalborg University
Aalborg, Denmark
jtc@create.aau.dk

6th Andreas Møgelmose
Department of Architecture, Design and Media Technology
Aalborg University
Aalborg, Denmark
anmo@create.aau.dk

***Abstract*—This paper introduces a system to assist golf course raters in determining the difficulty rating of a golf hole. Currently, determining the rating of a given golf hole relies on time-consuming manual measurements on the ground, which we attempt to partially automate. A U-net neural network is trained to classify greens, fairways, tees, bunkers, and water in golf courses, and a course rating assistance system is implemented to measure distances between relevant course parts. Since no public datasets containing golf courses existed prior to this work, we present a new public data set of golf courses created from orthophotos. 1,123 RGB orthophotos for training/validation and 108 RGB orthophotos for testing were gathered from 107 Danish golf courses (58% of Danish courses) during the spring season and manually annotated. The dataset is publicly available on Kaggle[1]. The U-net model accomplished a mean intersection over union (IoU) of 69.6%, mean sensitivity of 78.0%, and mean positive predictive value (PPV) of 84.1%. Based on this automatic analysis of the course images, the course rating assistance system computes 5 crucial distances for course raters to determine a course rating and achieved a mean error of 2.7% and 17.7% for green length and width, as well as a mean error of 3.3% and 4.2% for male/female hole lengths.**

***Index Terms*—Semantic segmentation, U-net, aerial imagery, golf, course rating system**

I. Introduction

In golf, there is a handicap system that allows players with different abilities to compete equally against one another in a game. This is achieved by giving each player a playing handicap that is determined by their skill level (handicap index) and the course rating. The course rating is based on a rating of the features on each golf hole to establish its difficulty. Using this system, players will be allocated more or fewer shots for each hole based on their playing handicap, when playing an entire golf course.

Up until the present time, the rating of golf courses has relied on time-consuming measuring of various elements on each hole by a team of 4 golf course raters. On each hole on an 18-hole course, a course rating team uses the average shot lengths of four theoretical golf player types (scratch male, scratch female, bogey male, bogey female) along with measurements such as fairway width, green size, length to carry[2] crossing obstacles, distance from the center of the target landing zone to lateral obstacles, bunker depth, length of grass, trees, the slope of the green, elevation, topography, etc. [1] The recorded measurements of the golf course are then converted into the actual course rating computation using a complex and confidential set of rules. The manual process is tedious, time-consuming, and limits the precision and consistency of the measurements. Imaging satellite observation systems have vastly improved throughout the 20th and 21st centuries, and have enhanced the acquisition capability of accurate remote sensing data for performing land cover analysis and monitoring green spaces [2]. Extracting features from remote sensing data of urban green spaces and golf courses is similar, and while not all of the above metrics can reliably be extracted from overhead imagery, some are worth investigating.

To optimize the process of rating golf courses, this paper proposes a full implementation of the U-net convolutional neural network (CNN) adjusted for multiclass semantic segmentation, using ImageNet transfer learning. Furthermore, a custom data set of 107 Danish golf courses is used to extract features for measuring the green size, distances to landing zones, distance from landing zones to obstacles, distance from tee area to obstacles, length of the hole, and width of the fairway. These metrics are some of the most time-consuming to collect manually. Topography, elevation change, and depth of bunkers are omitted from the current implementation, as

[1] https://www.kaggle.com/datasets/jacotaco/danish-golf-courses-orthophotos

[2] The point where the ball lands after a stroke before it starts rolling.

they cannot readily be extracted from RGB images.

The contributions of this paper are:

1) Custom data set containing 1,231 orthophotos of various scales (1:1000-1:1500) of 107 Danish golf courses with corresponding segmentation masks.
2) Evaluation of U-net CNN with transfer learning, for classifying layout elements of golf courses such as greens, fairways, tee areas, bunkers, and water hazards to extract relevant features for rating golf course difficulty.
3) Implementation of a course rating assistance system capable of calculating metrics relevant for grading course difficulty.

II. Related Work

Methods for performing land cover analysis from remote sensing data can roughly be divided into four parts: thresholding with one or multiple bands [3], pixel-based classification using handcrafted features [4], [5], object-oriented classification [6], and traditional deep learning. With the advent of CNNs, an increasing adaptation of deep learning methods for semantic segmentation classification of land cover has been studied in recent years to achieve greater accuracy compared to traditional methods with promising results [7]–[9]. Men et al. [7] proposes a solution to extract green space from urban areas obtained by satellite images. This solution is based on the U-net CNN, modified by introducing a convolution block channel attention (CBCA) module to replace the skip connections in the original U-net architecture for better performance. A mean IoU of 94.77% is achieved with the implementation of the CBCA, and a mean IoU of 94.61% using the original U-net CNN, indicating only a slight performance improvement. Wurm et al. [8] propose a semantic segmentation solution for extracting slum areas in satellite images. This method uses the FCN-VGG19 architecture to classify data into four classes: urban areas, vegetation, water, and slums. To explore the capabilities of transfer learning, a pre-trained CNN based on Quickbird satellite imagery was been applied to the Sentinel-2 satellite with greater mapping areas but lower geometric resolution and active SAR imagery from the TerraSAR-x satellite. Applying transfer learning to the Sentinel-2 imagery, the PPV and sensitivity increased from 30 to 55% and from 79 to 85%, respectively, highlighting the advantages of using transfer learning. Ayhan et al. [9] propose a solution for classifying three vegetation covers of similar appearance: trees, shrubberies, and grass, using only RGB images. To perform the classification, the semantic segmentation deep learning method, DeepLabV3+, is utilized on both a high and low-resolution data set and achieves an average classification accuracy of 78% for the low-resolution data set, showcasing the ability to accurately detect similar vegetation using only RGB images.

In this paper, we use a U-net architecture as presented in [7], but without the CBCA module, as the complexity of background features and shadow occlusions are of less significance for golf courses and the CBCA module only contributes to a minuscule difference in increased IoU. We take a pre-training approach similar to [8], as it improves results significantly (though we use ImageNet). Additionally, only RGB images are used as in [9], as it performs well for detecting similar vegetation. Finally, we train on a completely new data set, as no data sets of golf courses existed prior to our work.

III. Materials and Methods

A. Data Collection

Since no public data sets with orthophotos of golf courses exist, all data have been extracted from a national database named "Dataforsyningen[3]" (DF), which provides nationwide coverage of orthophotos for Denmark. A total of 1,123 RGB orthophotos with a resolution of 1600x900 pixels have been gathered for training and validation from 107 Danish golf courses during the spring season since this season is the most recently available in DF. To aim for a diverse data set containing as many courses in Denmark as possible for greater variance, 10 to 20 images have been captured per golf course. The data set is randomly divided into a 70/30 split for training and validation respectively. The images are captured using the program "QGIS[4]" with a scale of 1:1000 as it provides a great spatial resolution, while still capturing a broad portion of the physical layout and features of the golf course. Therefore, each image may contain 1 to 4 golf holes each with some remaining parts of other converging holes, see Fig. 1a.

The data has been annotated by experts to form the ground truth for the semantic segmentation, with a focus on these classes: greens, fairways, tees, bunkers, water, and background, see Fig. 1b. Rough and semi-rough are not annotated, as these cannot be reliably differentiated at a scale of 1:1000. Trees were also ignored, as it is too time-consuming considering their minor contribution to the course rating. Additionally, in cases where some classes are not clearly visible, they are omitted from that particular hole, e.g. fairways that are very difficult to differentiate from the rough. Fig. 2 shows how the different classes are distributed among the data set and illustrates an imbalance between the five classes since some classes are represented more than others.

An additional 108 images have been collected for testing at scales of 1:1000, 1:1250, and 1:1500 to ensure the captured testing images contain the layout of a single full golf hole, as this is essential for extracting measurements for the course rating. Each scale is uniformly represented in the test data. For the testing images, the surroundings of the golf hole have been manually masked by changing the pixel values to black, as the course rating is computed per hole.

The data set for this project has been made publicly available and can be found on Kaggle[5].

[3]https://dataforsyningen.dk/

[4]https://www.qgis.org/en/site/

[5]https://www.kaggle.com/datasets/jacotaco/danish-golf-courses-orthophotos

(a) (b)

Fig. 1: (a) Orthophoto of golf holes. (b) Segmentation mask of the orthophoto. White: Background, Dark-Green: Greens, Light-Green: Fairways, Red: Tees, Yellow: Bunkers, Blue: Waters

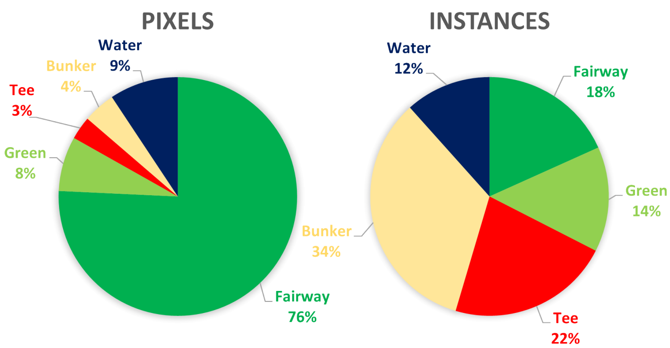


Fig. 2: Distribution of the five classes among the training and validation data set. (left) shows the percentage of pixels per class. (right) shows the percentage of instances per class.

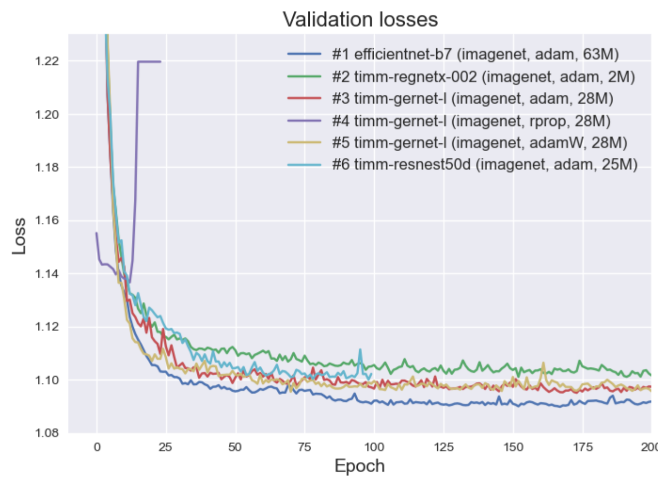


Fig. 3: Validation losses of the pre-trained models when trained on the golf data set.

B. U-net Architecture

U-net was originally built for bio-medical segmentation where large data sets are out of reach, similar to our application. The U-net architecture is divided into two parts: a contracting path (encoder) and an expansive path (decoder), which forms the U-shape of the architecture, hence the name U-net [10].

Our U-net architecture is implemented as presented in the original U-net paper [10], with adjustments to the input and output channels. The input channel is resized to 832x512x3 as 3 channels are required for RGB images, and the output is changed to 832x512x6 to contain the five classes on a golf course and a background class.

C. Training

The model can be trained from scratch with random weights on the collected data set. However, based on the finding from Wurm et al. [8], transfer learning has been implemented by importing weights from previously trained models to the encoder part of U-net, which gives the model an advantage of having weights that are already good at detecting features. The data set used to train a pre-trained model for transfer learning is ImageNet[6], which at this time contains 14 million images.

We compare six different encoder models, all trained with a batch size of 16 on a Linux machine using the NVIDIA A40 GPU with 48 GB GPU memory using the PyTorch framework. Additionally, the models have been trained with a learning rate of 0.0001 on images with 832x512 pixels along with different weights from pre-trained models for transfer learning and different optimizers. The parameters for the pre-trained models consist of the encoder family, encoder, the encoder weights, and the size of the model. To select the best model for semantic segmentation of golf courses, the U-net model has been trained using the various parameters until the loss converges. Fig. 3 presents the validation losses for all 6 pre-trained models when trained on the golf data set using cross-entropy loss as the loss function.

[6]https://www.image-net.org/about.php

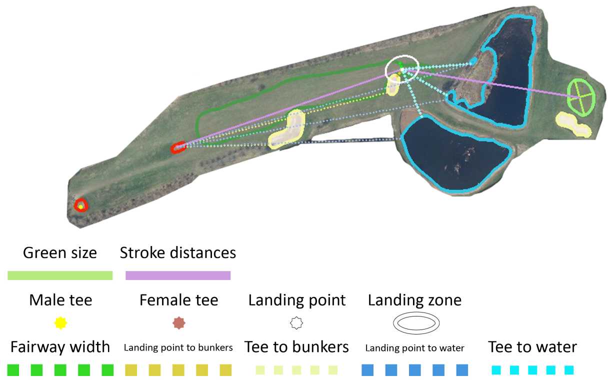


Fig. 4: Illustration of the GUI for a bogey female with all of the relevant features.

D. Course Rating System

The implemented course rating system analyzes a hole from two different player types: the scratch and the bogey player for males and females to get a better evaluation for the hole rating. The scratch player plays with a course handicap of 0 and hits tee shots of 228 [192][7] meters and can reach 431 [365] meters in two shots [1]. The bogey player plays with a course handicap of around 20 and hits tee shots of 183 [137] meters and can reach 338 [256] meters in two shots [1]. The implemented course rating system takes the segmented image from the U-net as input and calculates relevant distances between areas: the fairway's width, the total length of the hole, the distance from the landing point and tee area to obstacles, and the green's length and width. Fig. 4 shows the output, visualizing the above-mentioned distances. The segmentation model struggles to reliably predict tee areas - they are too small and vary too much. Since tee areas are a vital part of a golf course, the user assists the system by marking tee areas manually. The following sections present how each measurement is calculated.

1) Width of the fairway: As presented in the USGA course rating system [1], the fairway width of a golf hole is measured perpendicular to the line of play at the average landing point of the ball measured between the carry stroke length and total stroke length for each of the four player types. The width of the fairway is calculated using the segmentation image from the model, where a circle with the radius corresponding to the player's stroke length is drawn at the player's tee. Intersections between the circle and the fairway are calculated, where the distance between the two most outer intersections is determined as the width of the fairway.

2) Length of the golf hole: To determine the horizontal distance from the tee area to the center of the green along the intended line of play, landing points for the bogey or scratch, male or female are found at the center of the fairway, using a table of static stroke lengths [1]. Based on the landing points, the distance to the next landing point is calculated using the stroke distance. This process is repeated until no landing point with the fairway beyond the played direction can be found. Once this scenario occurs, the distance from the last landing point to the front, center, and back of the green is found and all stroke lengths for each player type are added together to measure the total length of the golf hole.

3) Distance from the landing zone and tee areas to obstacles: On a golf hole, obstacles have an impact on how the hole is played. Thus, it is relevant to know the distance to obstacles within the first stroke range and within a 47 meters radius of the landing point [1]. For this system, only water hazards and bunkers are counted as obstacles. To calculate the landing point to obstacle distance, the center point of the landing zone is designated, making this the reference point. The minimum distance from the reference point to each obstacle is calculated within the specified 47 meters radius [1]. Similarly, the calculation of the distance from the tee areas will be the same except for the difference in having the tee areas as the reference points and the maximum distance limited to the stroke lengths of each respective player type.

4) Length and width of the green: The length of the green is determined as the longest distance between two points on the green, and the width is the length of the line perpendicular to that. The resolution of the orthophotos does not allow the localization of the actual hole, and thus the hole will be designated as the center of the green where the length and width intersect.

IV. Results and Discussion

A. Segmentation

The models have been tested on 108 testing images. To evaluate the model, three metrics are utilized: IoU, sensitivity, and PPV [11]. The results from the models are shown in Table I and a confusion matrix for the classes is shown in Fig. 6.

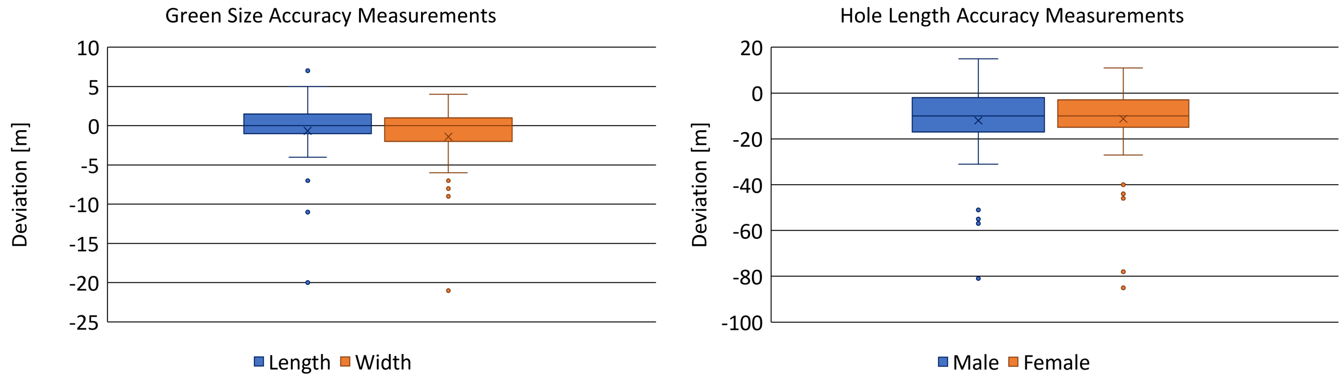
The best model on average for predicting the five classes is model 3 with a mean IoU of 69.6%, mean sensitivity of 78.0%, and mean PPV of 84.1%.

Bunkers have an IoU ranging between 63.3% and 80.4%, making it the most accurate class out of the five. Their color simply stands out compared to other classes. Fig. 6 underscores that bunkers are rarely misclassified.

The green and water classes are generally accurately found too. A common aspect of these three classes is that they are easy to annotate, which results in good model accuracy. However, poorly maintained golf courses can result in multiple greens being predicted in a single golf hole, which is unwanted for the course rating system.

The fairway IoU ranges from 41.9% to 76.6%. The biggest reason for this poor accuracy lies in the course quality, where low-quality courses contribute to it being harder to differentiate between the fairway and the adjacent rough. Not only does this hamper the system performance directly, it also makes annotating the ground truth troublesome, which leads to scenarios of imprecise annotations or complete exclusion of the fairway in the training data. In the testing data, a "best guess" approach has been used for annotating the fairways to

[7] [] represents measurement for female players



(a) (b)

Fig. 5: (a) Results of the green length and width measurements. (b) Results of the length of the hole measurements.

TABLE I: Results of the models containing the mean IoU, Sensitivity, and PPV for all classes. The best value is highlighted.

Results	**Fairway**			**Green**			**Tee**			**Bunker**			**Water**			**Mean**		
Model	IoU (%)	Sens (%)	PPV (%)	IoU (%)	Sens (%)	PPV (%)	IoU (%)	Sens (%)	PPV (%)	IoU (%)	Sens (%)	PPV (%)	IoU (%)	Sens (%)	PPV (%)	IoU (%)	Sens (%)	PPV (%)
1	74.2	79.9	89.7	78.4	87.2	**89.7**	**52.2**	**57.9**	**83.9**	79.7	85.6	92.5	59.4	65.3	**81.7**	68.8	75.2	**86.9**
2	53.2	55.2	**90.8**	69.8	85.2	77.6	23.4	25.6	57.7	78.1	86.3	89.5	54.4	62.3	69.1	55.8	62.9	76.9
3	**76.6**	**84.5**	89.5	80.2	**89.0**	89.1	48.0	53.5	82.2	80.0	**90.5**	87.9	**63.4**	**72.8**	72.0	**69.6**	**78.0**	84.1
4	46.3	39.0	89.0	49.1	55.5	81.2	0.0	0.0	0.0	63.3	80.8	76.8	33.4	34.6	64.9	38.4	44.0	62.4
5	67.4	72.5	88.8	**80.6**	88.0	88.8	48.0	54.1	78.3	**80.4**	87.9	90.9	53.7	59.3	69.7	66.0	72.4	83.3
6	41.9	45.0	70.9	55.9	61.6	81.2	0.0	0.0	0.0	78.1	82.1	**94.3**	60.6	64.7	80.3	47.3	50.7	65.4

ensure that the system is not punished for correct classifications that are mistakenly not marked as ground truth.

Tees are the most difficult class for the model to find, as they occur in various shapes and sizes along with different colors depending on the golf course. Similarly to fairways, the prediction of tees is affected by poor course quality and similarity to adjacent grass. In Table I, the results of tee predictions range between an IoU of 0% and 52.2%. Thus, tee areas must be manually marked by the user in the rating system, since they are essential for extracting features of golf courses.

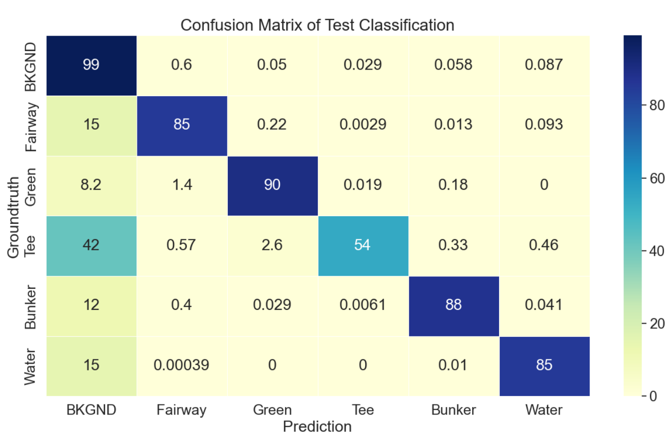


Fig. 6: Confusion matrix of the six classes on the testing images measured in percentage of total class pixels using model 3.

B. Course rating

The measurements from the course rating system are compared with real-life measurements of 108 golf holes to validate its accuracy. However, as only the length of holes and size of the greens are typically provided by the respective golf clubs in Denmark, only these measurements will be evaluated. Fig. 5a shows the deviation between the ground truth measurements of the greens and the predicted measurements. The deviation of hole length is shown in Fig 5b.

The greens in the testing set have a mean length and width of 30.0 and 23.1 meters, and the predicted mean length and width of the greens are 29.2 and 19.0 meters. This corresponds to a mean absolute error of 0.8 m (2.7%) and 4.1 m (17.7%). The model performs well but tends to underestimate the size of greens, especially green widths. Part of the discrepancy in width, however, has roots in the fact that not all golf courses compute their width as the length of a line perpendicular to the length, as is specified in the USGA regulations.

Likewise, hole lengths are often underestimated (fig. 5b), as the mean absolute error across the male and female lengths are 11.2 m (3.3%) and 12.6 m (4.2%). Part of the reason is that the current system does not use any height map information to calculate the distances and instead assumes that the golf holes lie on a flat plane. Additionally, holes are assumed to lie in the center of the green, as the hole is not visible from an orthophoto.

However, some measurements for both genders have substantial deviations from the original length and are displayed as statistical outliers in Fig. 5b. These outliers can be explained by how the landing points are calculated. As presented in

Section III-D, the landing points are calculated with fixed distances until no landing points remain on the fairway. This works very well with straight holes, but holes with the form of a dogleg[8] pose a challenge. As trees have been excluded from the segmentation mask, they are not treated as obstacles. Therefore, the line of play can take otherwise impossible "shortcuts" over the trees on dogleg holes resulting in a distance loss.

Similar distance deviations can also be caused by inaccuracies in the semantic segmentation, where poor predictions of the fairway may influence the intended line of play, as no landing points can be found, forcing the course rating system to measure distances through unplayable terrain or obstacles e.g. trees.

Overall, most of these differences are the result of inaccuracies in the semantic segmentation, which means that when the model's accuracy increases, the distances will surely be more precise.

C. Future Work

Additional data sources containing elevation information should be added to include elevation changes, resulting in better accuracy. Furthermore, as the model was trained on 1,123 images, extensive augmentation or more images could be added to the data set to improve the variation of the images, including images from other seasons, not just from different courses. The entire data set consists of Danish golf courses and the model is trained on Danish terrain. While there is variance between golf courses in Denmark, the model will be most accurate when tested on environments that are similar to Denmark, and will decrease in accuracy if tested on foreign golf courses as the climate and other geographical factors can be vastly different compared to Denmark.

Since the models could not predict tees to a satisfying result, other approaches can be taken into consideration, such as building another model exclusively for tee prediction.

V. Conclusion

As of today, golf course rating still happens manually but thanks to the fast evolution of AI, tasks like this can be more easily assisted than with hard computing systems. In this paper, the main approach is to apply semantic segmentation on orthophotos of Danish golf courses. To achieve that, U-net, a CNN architecture that is commonly used for semantic segmentation, has been implemented. The network has been trained with weights from multiple encoders to get the best model possible. The best model obtained a mean IoU of 69.6%, mean sensitivity of 78.0%, and mean PPV of 84.1% out of six trained models. After obtaining the prediction images from the trained neural network, different types of distances on the result images have been calculated to provide help in course rating. In some cases, the tees of the golf courses are not predicted due to vision difficulties but the golf rater can assign them manually in our system. After manually assigning tees, the systems compute relevant distances with an error of less than 5% (except in the case of green width at 17.7%, which is due to issues with the ground truth data). In summary, the created system provides a human-in-the-loop solution for aiding course rating in golf, allowing significant time savings for course raters.

[8]A dogleg is a hole with an angled fairway, where the green is not completely visible from the tee area.

References

[1] United States Golf Association, "The usga course rating system," 2012.

[2] Wenxue Fu, Jianwen Ma, Pei Chen, and Fang Chen, *Remote Sensing Satellites for Digital Earth*, Springer Singapore, 2020.

[3] Soojeong Myeong, David J. Nowak, and Michael J. Duggin, "A temporal analysis of urban forest carbon storage using remote sensing," *Remote Sensing of Environment*, vol. 101, no. 2, pp. 277–282, 2006.

[4] Quanlong Feng, Jiantao Liu, and Jianhua Gong, "Uav remote sensing for urban vegetation mapping using random forest and texture analysis," *Remote Sensing*, vol. 7, no. 1, pp. 1074–1094, 2015.

[5] Razieh Shojanoori and Helmi ZM Shafri, "Review on the use of remote sensing for urban forest monitoring," *Arboric. Urban For*, vol. 42, no. 6, pp. 400–417, 2016.

[6] Juan P. Ardila, Wietske Bijker, Valentyn A. Tolpekin, and Alfred Stein, "Context-sensitive extraction of tree crown objects in urban areas using vhr satellite images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 15, pp. 57–69, 2012.

[7] Guoqiang Men, Guojin He, and Guizhou Wang, "Concatenated residual attention unet for semantic segmentation of urban green space," *Forests*, vol. 12, no. 11, pp. 1441, 2021.

[8] Michael Wurm, Thomas Stark, Xiao Xiang Zhu, Matthias Weigand, and Hannes Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS journal of photogrammetry and remote sensing*, vol. 150, pp. 59–69, 2019.

[9] Bulent Ayhan and Chiman Kwan, "Tree, shrub, and grass classification using only rgb images," *Remote Sensing*, vol. 12, no. 8, pp. 1333, 2020.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015, pp. 234–241, Springer International Publishing.

[11] Rafael Padilla, Sergio L. Netto, and Eduardo A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.