# Integration Approach for Manual Generated Single Tree Crown Annotations

**QIPENG MEI[1], JANIK STEIER[1] & DOROTA IWASZCZUK[1]**

*Abstract: For an accurate mapping of forest stands, precise object detection at the individual tree level is necessary. Currently, supervised deep learning models dominate this task. To train a reliable model, it is crucial to have a robust model and an accurate tree crown annotation dataset. The current method for generating these datasets still relies on manual annotation. However, the tree crowns exhibit intricate contours. In some instances, trees intersect with each other, and their spatial arrangement is irregular. This leads to inaccurate and incomplete quantity annotations, including the inclusion of multiple tree crowns in a single annotation. Therefore, this study explores a novel approach that integrates the annotations of multiple annotators for the same region of interest and can reduce annotation inaccuracies due to personal preference and bias.*

## 1 Introduction

Single tree crown mapping is an important component of forestry research. It provides data for various environmental assessments, modeling, and management by supplying information such as crown size changes, individual tree placements, and health status. Traditional field measurement of individual tree crowns is a time-consuming process and hampered by access concerns in private areas. In contrast, the remote sensing measurement collects a broader variety of tree cover information, is less expensive, and is not limited by accessibility constraints (ZHAO et al. 2023).

The progress in computer vision has led to the creation of various techniques for the automatic identification and delineation of tree crowns in aerial images, supplanting traditional manual methods reliant on visual interpretation. These approaches predominantly rely on image analysis, with deep learning emerging as a prominent strategy in the realms of object detection and instance segmentation. This dominance is further bolstered by the swift advancements in graphics processing units and the continual refinement of algorithms (HOESER et al. 2020). By harnessing human-provided knowledge, exemplified by manually annotated tree crown labels, supervised deep learning models iteratively optimize their vast number of parameters, which endows them with the capacity to differentiate individual tree crowns, akin to the cognitive processes of humans.

Hence, along with a meticulously crafted deep learning model, an accurately annotated dataset is also of paramount significance. However, the creation of annotation datasets for tree crowns is fraught with several unavoidable challenges. First, unlike other regular surface features (buildings, roads, etc.), the outer contour of the tree crown is extremely complex. Moreover, the overlapping and intermingling of multiple trees pose additional complexities. In terms of spatial distribution, particularly in forested areas, trees exhibit an uneven distribution. Lastly, the quality of images, accompanied by certain surface intricacies like shadows, and the presence of entities resembling trees in appearance (e.g., green belts, lawns) contribute to the

[1] Technical University of Darmstadt, Remote sensing & image analysis group, Franziska-Braun-Str. 7, 64287 Darmstadt, E-Mail: [qipeng.mei, janik.steier, dorota.iwaszczuk]@tu-darmstadt.de

annotator's decision-making challenges. Negotiating these hurdles proves to be a challenging task, even for seasoned experts.
In response to these challenges, this study introduces an innovative approach, proposing the integration of annotations from multiple annotators for the same region of interest (ROI). The objective is to alleviate the detrimental impact stemming from the individual preferences and errors of annotators during the process of single tree crown annotation.
In the ensuing chapters, Chapter 2 provides a comprehensive introduction to the methodology, while Chapter 3 delineates the experimental details. Moving forward, Chapter 4 and Chapter present the results and related evaluations. Lastly, the study culminates in Chapter 6 with a summary and future research prospect.

# 2 Methodology

Deep learning models utilized for instance segmentation typically employ vector data as a carrier in their annotation datasets. Geometric shapes, such as points, lines, and polygons, serve as representations for labels. In contrast to raster data, vector data tends to exhibit smaller file sizes, storing solely the endpoint coordinates of geometric shapes rather than the values of individual pixels. Additionally, the resolution remains constant regardless of scaling ratios.
One idea for integrating vector annotation data involves transforming the vector data into the raster domain for specific numerical operation and subsequently reverting it back. WALTER (2018) proposed a method for amalgamating multiple polygons based on the “Wisdom of the Crowd”. This approach assumes that "if many individuals measure the same object, the average geometry should closely approximate the real geometry" (WALTER 2018). COLLMAR et al. (2023) expanded on this concept to integrate tree crown labels obtained from crowdsourcing through a two-step process. However, these methodologies are centered around polygon integration for a single object and are not applicable when dealing with the adjacency of multiple objects within a single image.
Our work further extends the principle of the “Wisdom of the Crowd”, reconstructing the matrix transformed from vector data, which we named the acquisition matrix. Employing a combination of Markov random field (MRF) and graph cuts algorithm, we achieve integrated annotation, enabling the amalgamation of multiple independent polygons within a single region of interest.

## 2.1 Acquisition matrix

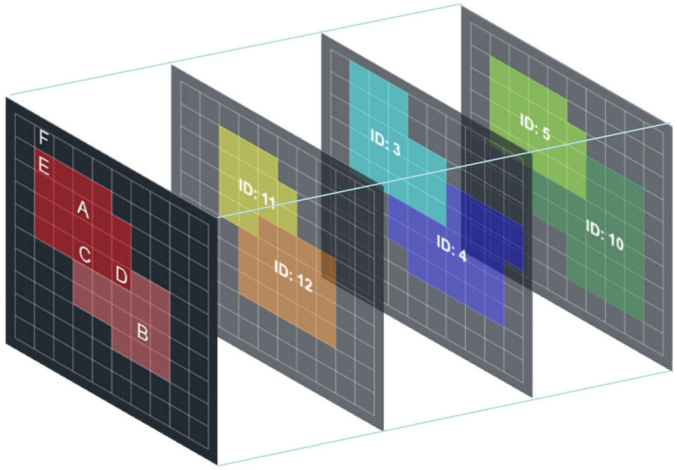
The shape of the acquisition matrix is (N+1, H, W), where N represents the number of the individual annotations related to a ROI, and H and W are the height and width of the corresponding ROI. The additional layer, denoted as 1, serves for further numerical operations. The specific process is as follows:

1. Each individual vector annotation is converted to a raster annotation map with the shape of (1, H, W), where the value of each pixel in the map is the identifier (ID) of the tree crown label, and background pixels are assigned a value of 0.
2. The N layers of the acquisition matrix are formed by stacking these raster images.
3. The expansion layer records the frequency *f* of the current pixel being labeled.

Moreover, through processing the acquisition matrix, potential clusters corresponding to the ROI are identified. These clusters are characterized by their ID sequence, composed of the tree

crown labels assigned by each annotator. Refinement of potential clusters is based on the following assumptions:

- Pixels in the region around the tree center, where most annotators agree on the tree crown, share the same ID-sequence, referred to as the central ID-sequence (CIDS).
- Pixels at the edges of the tree crown may exhibit different ID-sequences due to varying perspectives of one or more annotators, referred to as the edge ID-sequence (EIDS).




| Pixel | Frequency | 1st ID | 2nd ID | 3rd ID |
|---|---|---|---|---|
| A | 3 | 11 | 3 | 5 |
| B | 3 | 12 | 4 | 10 |
| C | 3 | 11 | 3 | 5 |
| D | 3 | 12 | 3 | 5 |
| E | 2 | 0 | 3 | 5 |
| F | 1 | 0 | 3 | 0 |

Fig. 1: Instance of the acquisition matrix

Fig.1 illustrates an instance of the acquisition matrix, where we have specifically identified six key pixels for further explanation. A and B denote the CIDS serving as focal points for most annotators, surrounded by pixels that share the same ID sequences. Meanwhile, C, D, and E represent the EIDS, with the former two having lower recognition levels than A and B but being unequivocally related. Notably, E is situated on a label boundary, deviating from the decision of one annotator. Lastly, F is highly likely to be categorized as background, aligning with the assessment of a single annotator. Subsequent processing entails their assignment to the pertinent clusters.

We obtain anchors for potential clusters by retaining the CIDS. The process is as follow:

1. Retain potential clusters with $f$ exceeding the threshold T, indicating majority annotator agreement.
2. Apply non-maximum suppression considering the abundance of ID-sequences to eliminate redundancies that may point to the same tree crown. This approach is designed to preserve CIDS while eliminating EIDS.

The finally refined potential clusters obtained represent the recognized trees within the corresponding ROI. Subsequently, MRFs are employed to regenerate labels for each individual tree crown.

## 2.2 Markov Random Field

MRF is a probabilistic graphical model used to model the relationship between random variables. It is a graph structure composed of a set of nodes and edges connecting these nodes, where nodes represent random variables, and edges represent the dependency between variables (GEMAN & GRAFFIGNE 1986). In this work, we established a MRF to characterize the annotation map, where each pixel ($x_i$) in the map is a node, connected by edges with its eight neighbors' nodes ($x_j$). Their set is denoted by $\mathcal{N}$ resp. $x_j \in \mathcal{N}(x_i)$. Based on the MRF we implemented the following energy function (QIU et al. 2022):

$$E = \sum_{x_i} \left[ \sum_{x_j \in \mathcal{N}(x_i)} E_p(x_i, x_j) + E_u(x_i) \right] \rightarrow min \quad (1)$$

Here, two terms determine the total energy $E$. $E_p(x_i, x_j)$ is the pairwise potential of $x_i$ and its neighbor $x_j$. To encourage adjacent pixels to be classified into the same cluster (C), pairwise potential is defined as:

$$E_p(x_i, x_j) = \begin{cases} 0, & if\ C(x_i) = C(x_j) \\ 1, & if\ C(x_i) \neq C(x_j). \end{cases} \quad (2)$$

$E_u(x_i)$ is the unary potential of each pixel, representing the energy of classifying the pixel into clusters. It is obtained as follows:

1. Compute the similarity matrix $S(x_i)$ of the ID-sequences between the pixel and all potential clusters, where similarity is quantified as the fraction of the number of identical IDs over the total number of items in the ID-sequence.
2. Convert the similarity matrix into a unary potential matrix $M_u(x_i)$ through the following numerical computation:

$$E_u(x_i) \in M_u(x_i) = min(-log_2 S(x_i), 2048) \quad (3)$$

Subsequently, $E_u(x_i)$ are normalized and scaled to the range of 0 to 2048 based on the maximum unary potential. In this way, each pixel presents high unary potential on clusters with low relevance.

For the established MRF, our objective is to derive the optimal clusters distribution that minimizes the total energy associated with it. Here, we utilize an efficient approximation algorithm based on graph cuts, namely α-expansion moves, to achieve minimum energy estimation (BOYKOV et al. 2001). The final clusters distribution is converted to vector data, which is the integrated annotation.

## 3 Experimentation

To assess the viability of the integration method, we selected eight locations within Frankfurt am Main as ROIs. These areas encompass cemeteries, streets, squares, and backyards within the city area. The trees' arrangement varies from completely irregular (such as cemeteries) to organized (such as streets, squares), reflecting both natural growth and intentional planting (see Fig.2). Each aerial image has dimensions of 512×512 pixels, with a Ground Sampling Distance (GSD) of 20 centimeters per pixel, thereby providing ground surface information across an area of 10485 square meters.

One of the ROIs (H in Fig.2) was annotated twice by the three experts. As for remaining ROIs (A to G in Fig.2), four experts independently annotated the tree crowns. Subsequently, we applied a threshold of 70 to identify the centres of potential clusters (3 for A to G, 4 for H).

Moreover, the integrated and individual manual annotations were compared against the tree cadastre collected in the field. The cadastre, maintained by the parks department of the City of Frankfurt am Main, has been accessible online to the public since 2014, with the latest version updated in August 2023.

It is worth noting that since tree registry records do not have access to private areas, this portion of the tree will be excluded at evaluation. For example, the left edge of B and the left-center part of D in Fig.2.

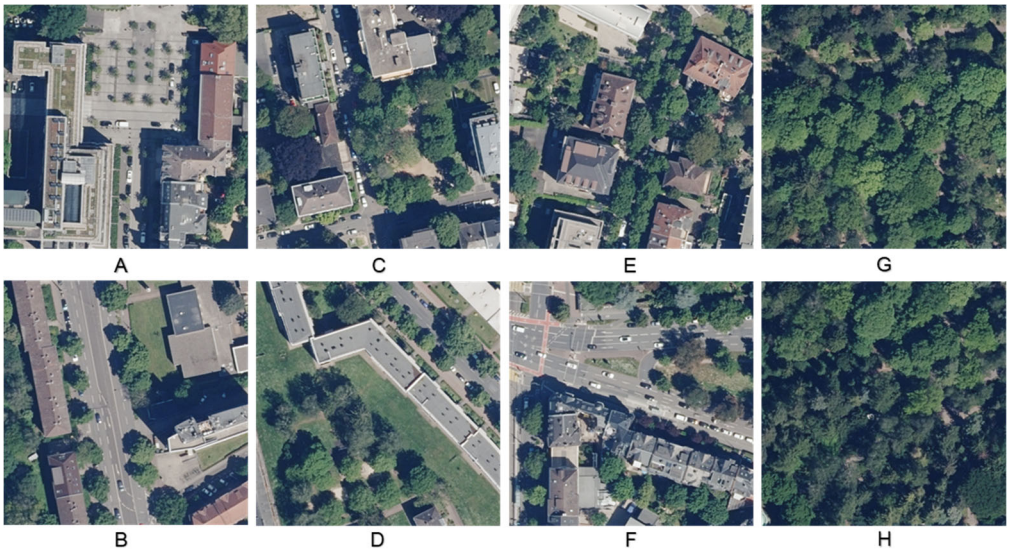



Fig. 2: Aerial images of the experimental region of interest

## 4 Results

In Fig. 3, we present the visualization of annotations for some distinct representative areas, which exemplify specific spatial arrangements. The white lines in the figure represent the integrated annotation, while the differently colored lines correspond to individual manual annotations.

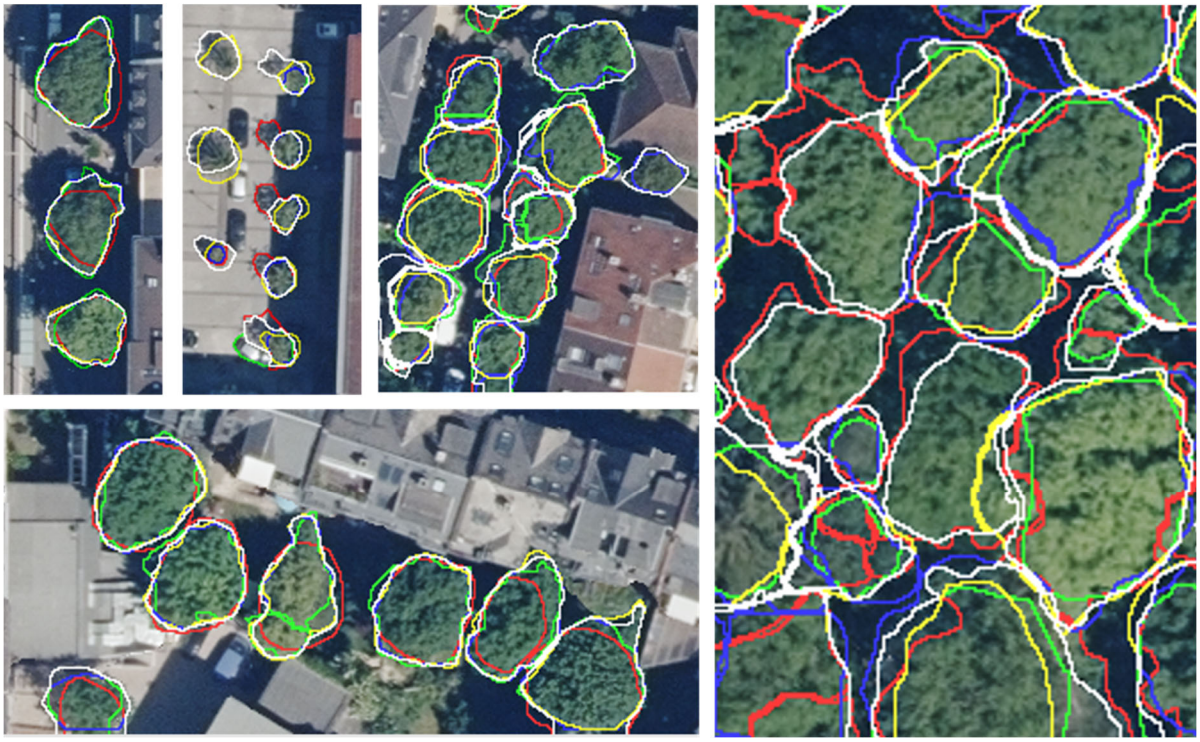


Fig. 3: Annotation visualization in representative areas

From a visual standpoint, the integrated annotation effectively encompasses the tree crowns, capturing the consensus among annotators. Furthermore, each label aligns more closely with the actual edge of the tree crown, providing a more realistic representation from a subjective perspective. In particular, when trees are scattered and distributed individually, tree crown labels of integrated annotation are more representative of the actual situation.

# 5 Evaluation

## 5.1 Intersection over Union

Intersection over Union (IoU) is a common metric used to evaluate the accuracy of object detection, which is defined as the ratio of the area of intersection between the predicted bounding box and the ground truth bounding box to the area of union between them. Here, we use overall IoU to assess the consistency of the annotations with the field measured tree coverage from the tree register. The tree coverage is expressed as a circular shape, calculated from crown diameter.

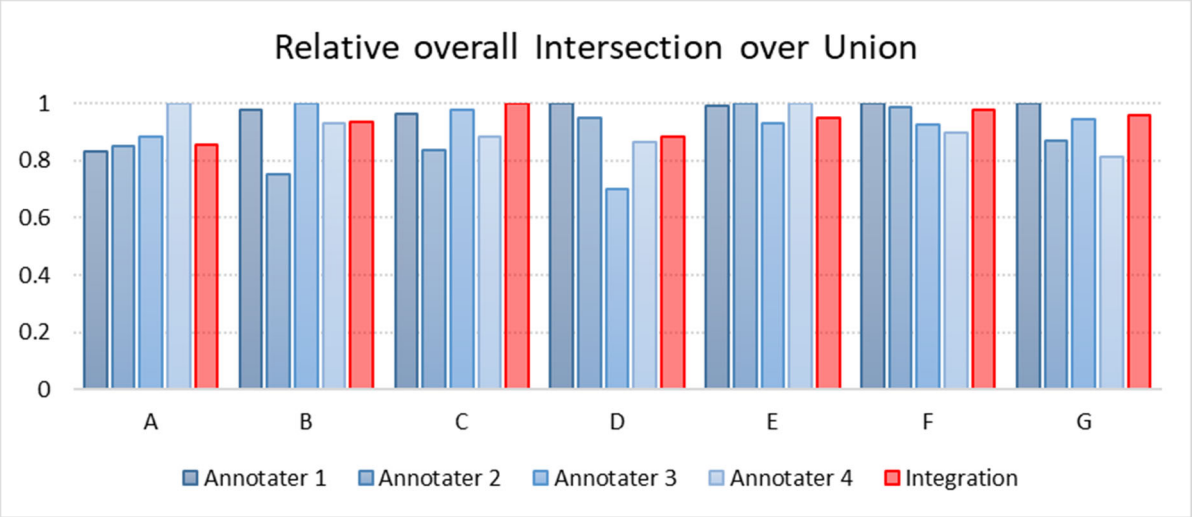



Fig. 4: Relative overall Intersection over Union

Fig. 4 depicts the relative overall IoU of A to G. We represent the relative IoU of the annotations by establishing the optimal value of each ROI as 1, thereby standardizing the remaining annotations. It is apparent that among individual manual annotations for each ROI, there are considerable variations. In contrast, integrated annotation proves effective in alleviating these disparities and aligning IoU closer to, or even reaching, the optimal level.
For H, the overall IoU of 68% in the integrated annotation is higher than the 59% to 65% in the manual single annotations. Fig. 5 shows the IoU distribution for each label in the annotation. It is evident that the labels in integrated annotation tend to demonstrate higher IoU when compared to individual manual annotations.

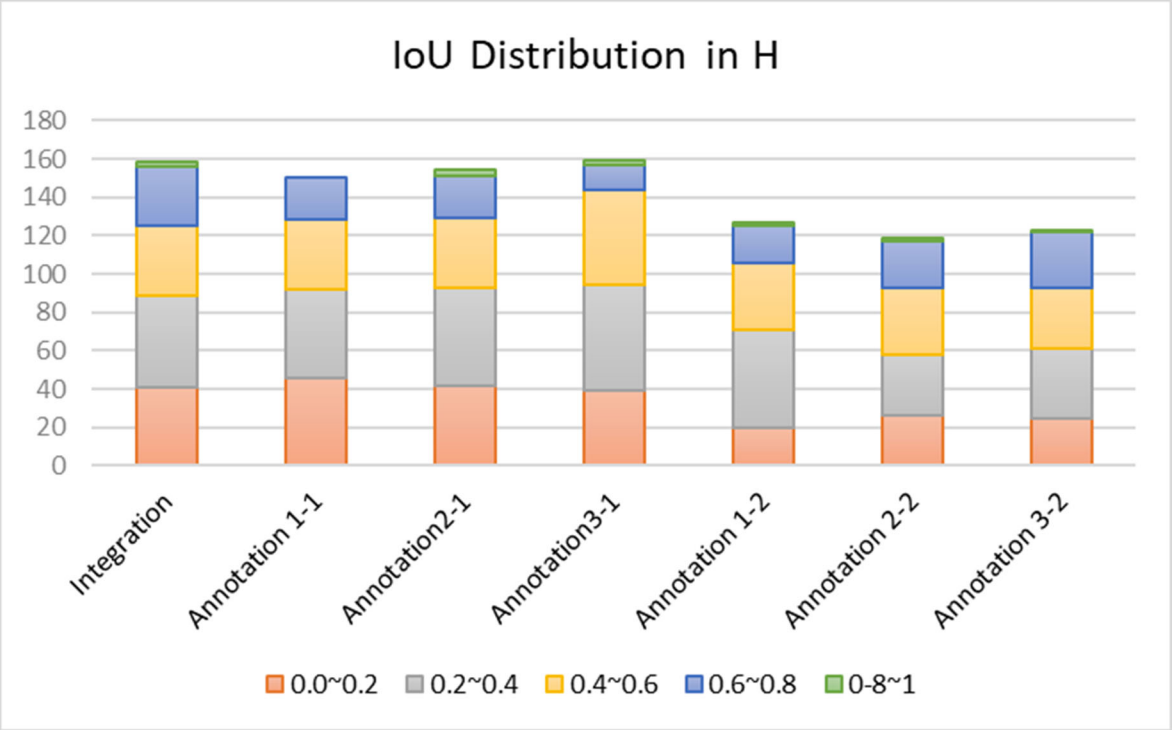



Fig. 5: IoU Distribution in H

## 5.2 Tree center coverage

The ability of the annotation to cover all trees is an indicator worth evaluating. We assessed tree center coverage by verifying the inclusion of tree centers from the tree registry in the annotation (see Fig. 6). In Comparison with the fluctuations in IoU, tree center coverage typically exhibits minimal divergence among annotators. Integrated annotations are influenced by different annotators to converge to a neutral position. For H, the integrated annotation was able to cover 82% of the trees' centers compared to manual single annotations, which ranged from 69% to 79%.

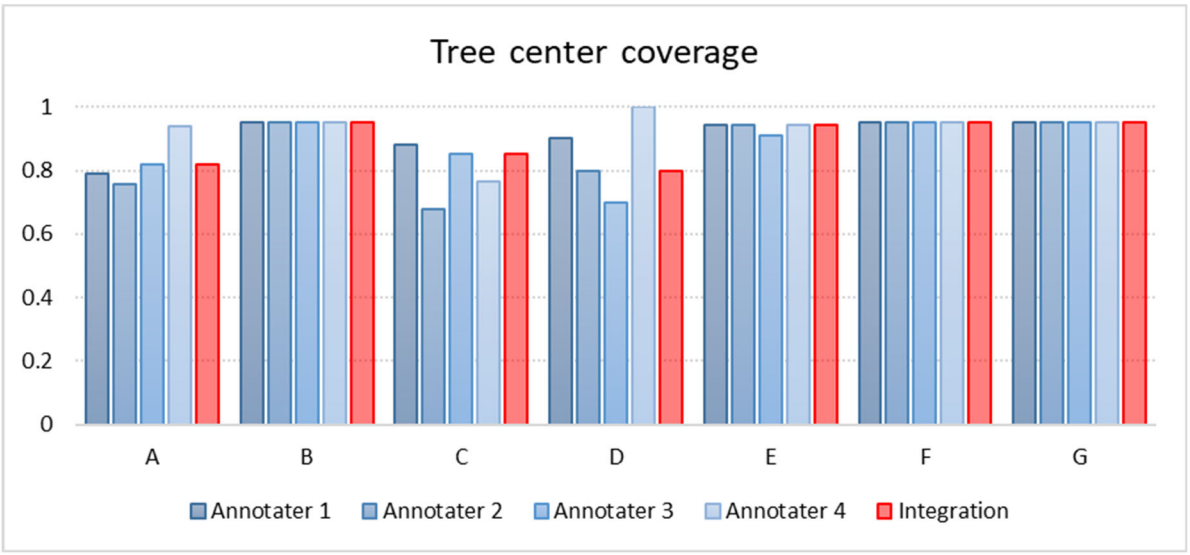



Fig. 6: Tree center coverage

### 5.3 Multiple trees as a single label

Influenced by image quality and tree characteristics, labeling multiple trees as a single entity is a common annotation error. Therefore, this type of error is quantified and presented in Fig. 7. In the figure, Single indicates that the label is associated with only one tree center, while Multiple denotes that the label covers more than one tree center.

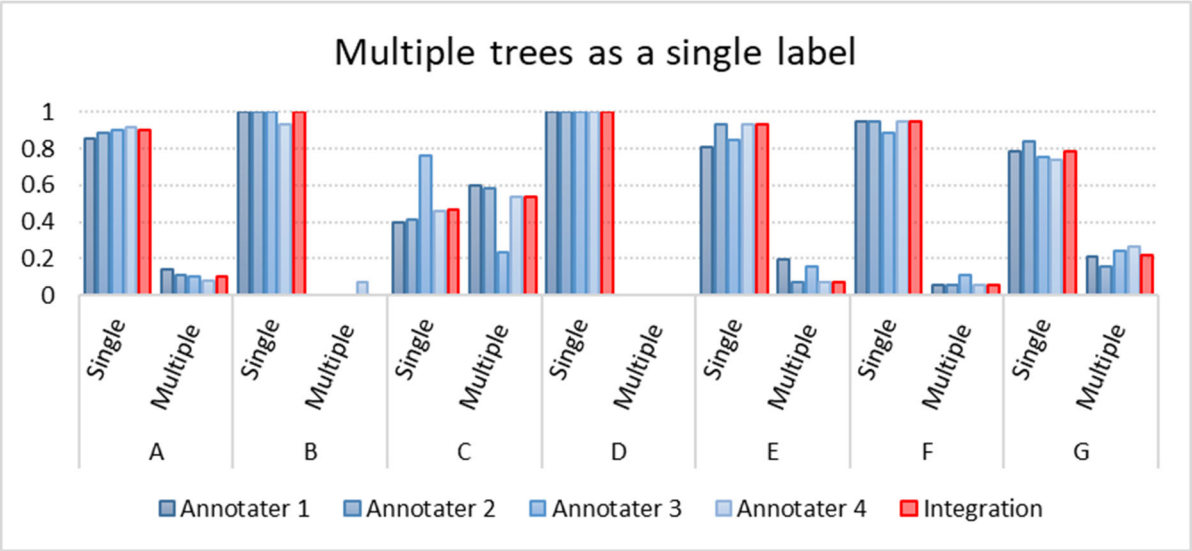



Fig. 7: Multiple trees as a single label

It is worth noting that this error is present in almost all ROIs. However, due to the mutual constraints imposed by multiple annotators, the integration of labels can effectively keep this situation well-controlled within a low range.

### 5.4 Annotation Preferences

Finally, we conducted a quantitative assessment of each annotator's annotation preferences in three dimensions (see Fig. 8): IoU, TCC (tree center coverage) and TSL (tree as single label).

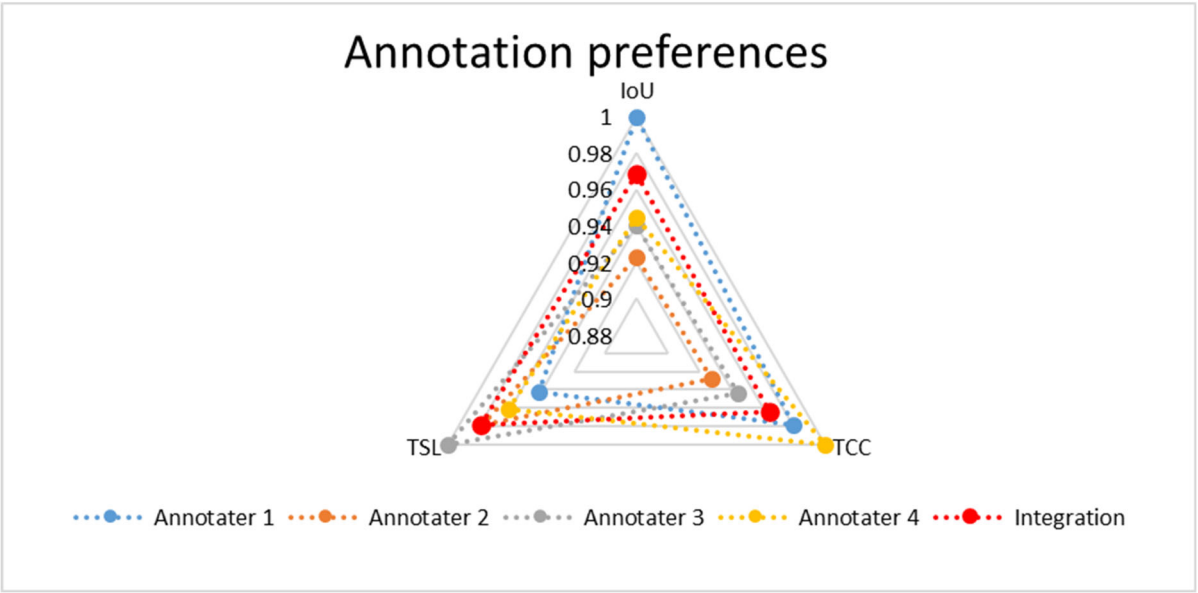



Fig. 8: Annotation preferences

Similar to Section 5.1, we standardize the representation of metrics by defining the optimal value for each dimension as 1, thereby normalizing the assessments of the other annotators. Among these annotators, Annotator 1 delineates the broadest tree coverage in the aerial image, encompassing more trees, albeit with less distinction between individual trees. Conversely, Annotator 3 places greater emphasis on distinguishing individual trees, but overlooks the labelling and contouring of a few trees. Annotator 3 shares a similar annotation preference with Annotator 2. Lastly, Annotator 4 excels at identifying trees in aerial images but falls short in distinguishing and describing contours. In comparison, our integrated annotations, by leveraging the strengths and compensating for the weaknesses of individual annotators, produce an annotation map that is balanced on all three dimensions.

# 6 Conclusion

This study introduces a novel integration approach of annotations that begins by constructing an acquisition matrix to gather annotations from each annotator for the same region of interest. Subsequently, adhering to the wisdom of the crowd principle, a Markov random field is constructed to represent each pixel. Finally, the labels are reassigned to all pixels through graph cuts algorithm.

Experimentation on tree crowns in various regions of Frankfurt am Main demonstrates the method's potential to integrate annotations from different annotators while mitigating the bias and preference from individual annotators. While the integrated annotations are not necessarily optimal for a single metric, they prove to be the most balanced, approaching or reaching optimal values for multiple metrics. Therefore, this approach is adept at generating relatively high-quality annotations in situations where the ground truth is uncertain or unavailable.

In future research, we would attempt to generate a large number of labels by crowdsourcing to verify its capability on a high number of annotations. In addition, possible improvements are annotations on different patterns, i.e. grouping annotators on RGB images, pseudo-red images, and radar images to provide diverse data references for integration. For the assessment of annotation quality, it is meaningful to perform an all-encompassing assessment by consistency checks, e.g. the consistency of pixel colors within a single label, the consistency of the NDVI, and whether to incorporate edge shadowing errors, to assess the performance of the integration method more comprehensively.

# 7 Acknowledgement

This study was conducted as part of the cooperative project ForSens between TU Darmstadt and Karuna Technology UG. We would like to express our sincere gratitude to Karuna Technology for providing the data for this research. The funding was provided by the State of Hesse as a part of "LOEWE funding line 3" HA-Project-No.: 1381/22-86. A special thanks goes to all the annotators.

Furthermore, I am thankful to the China Scholarship Council (CSC) for supporting my personal research, Grant/Award Number: 202308080109.

## 8 References

BOYKOV, Y., VEKSLER, O. & ZABIH, R., 2001: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence, **23**(11), 1222-1239, https://doi.org/10.1109/34.969114.

COLLMAR, D., WALTER, V., KÖLLE, M. & SÖRGEL, U., 2023: From multiple polygons to single geometry: optimization of polygon integration for crowdsourced data, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., **X-1**/W1-2023, 159-166, https://doi.org/10.5194/isprs-annals-X-1-W1-2023-159-2023.

GEMAN, S. & GRAFFIGNE, C., 1986: Markov random field image models and their applications to computer vision. Proceedings of the international congress of mathematicians, **1**.

HOESER, T. & KUENZER, C., 2020: Object detection and image segmentation with deep learning on Earth observation data: A review-part I: Evolution and recent trends. Remote Sensing, **12**(10), 1667, https://doi.org/10.3390/rs12101667.

QIU, K., BULATOV, D. & LUCKS, L. ,2022: Improving Car Detection from Aerial Footage with Elevation Information and Markov Random Fields. Signal Processing and Multimedia Applications. SciTePress., 112-119, https://doi.org/10.5220/0011335900003289.

WALTER, V., 2018: Integration of multiple collected polygons with a raster-based approach. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., **XLII**-4, 679-685, https://doi.org/10.5194/isprs-archives-XLII-4-679-2018.

ZHAO, H., MORGENROTH, J., PEARSE, G. & SCHINDLER, J., 2023: A Systematic Review of Individual Tree Crown Detection and Delineation with Convolutional Neural Networks (CNN). Curr Forestry Rep, **9**, 149-170, https://doi.org/10.1007/s40725-023-00184-3.