Enhancing Aerial Object Detection: Automated Conversion of Axis-Aligned to Polygonal and Rotated Annotations

Daniela L. Freire[1 [0000−0002−5363−3608]],
Andre C. P. L. F. de Carvalho[1 [0000−0002−4765−6459]],
Mateus Antonio Chinelatto[2 [0000−0002−6933−213X]],
Augusto José Peterlevitz[2 [https://orcid.org/0000−0003−0575−9633]], and
Ricardo Dutra da Silva[2 [https://orcid.org/0000−0002−8002−8411]]

[1] University of Sao Paulo, Sao Paulo, Brazil {danielalfrere,andre}@icmc.usp.br
[2] Instituto de Pesquisas Eldorado, Brazil
{mateus.chinelatto, augusto.peterlevitz, ricardo.silva}@eldorado.org.br

Abstract. In the field of computer vision, object detection within aerial imagery presents unique challenges due to the varied angles and complex backgrounds of objects such as buildings and vehicles. Current annotation tools often lack support for accurately delineating these objects, relying primarily on manual bounding box methods which are time-consuming and prone to inconsistency. The study aims to improve the precision and efficiency of annotations in aerial imagery by automating the conversion of axis-aligned annotations into more adaptable polygonal and rotated forms. This approach seeks to refine the quality of training datasets for object detection models without altering the models themselves. Utilising the Segment Anything Model along with advanced object detection models YOLOv8 and YOLOv9, our automated technique transforms traditional annotations to fit oriented bounding boxes and complex segmentation requirements. This method was evaluated through a case study on electrical transmission towers in aerial images. Preliminary results indicate that our automated method not only enhances annotation accuracy but also significantly reduces the manual labour involved, thus lowering the overall costs and time associated with data preparation for object detection training. The success of this approach highlights the potential for broader application and development in automated annotation technologies.

Keywords: Object Detection · Image Segmentation · Machine Learning · YOLO Model · Segment Anything Model · Oriented Bounding Box · Polygonal Annotation · Annotation Conversion.

1 Introduction

Artificial intelligence (AI), focusing on data-centric approaches, has emerged as a critical area of innovation, particularly in computer vision. Our proposed

method, a novel and significant contribution, automates the conversion of axis-aligned annotations into polygonal and rotated annotations. This approach, which prioritises systematic and scalable enhancements to data quality over modifications to the model itself, is particularly crucial in object detection tasks. The precision of annotations, which determine the locations and boundaries of objects in images, is a fundamental aspect of this approach. Correct and consistent annotations are essential, as they directly influence the effectiveness of training and the subsequent performance of object detection models [8].

Manual bounding box methods, the mainstay of annotating images for object detection, pose significant challenges. These challenges are particularly pronounced in aerial imagery scenarios, where objects such as vehicles or buildings often appear at various angles, making the annotation task more complex. Current annotation tools often need more support for oriented bounding boxes (OBBs) and polygonal annotations, which are crucial for accurately delineating inclined objects. Moreover, manual drawing of the bounding boxes is time consuming and often needs to be more consistent and accurate, leading to significant variations in the quality of the annotation [1].

The laborious nature of manual annotations not only incurs high time costs but also profoundly impacts the quality of the resulting models. Inaccurate annotations can lead to poorly trained models, which may fail to detect objects correctly, posing a considerable risk to the reliability and effectiveness of AI applications in crucial fields such as surveillance and remote sensing.

This article introduces an automated technique to transform axis-aligned annotations into polygonal and rotated annotations compatible with advanced object detection and segmentation models such as YOLOv8 and YOLOv9, using the Segment Anything Model (SAM). YOLOv8 and YOLOv9 have been developed to address various object detection and segmentation challenges, including oriented bounding boxes. The presented automated conversion method aims to significantly reduce the time and cost associated with manual annotation processes by leveraging these AI models. SAM, combined with YOLO models, is expected to enhance annotation accuracy and efficiency, offering a reliable solution to the difficulties associated with manual annotation techniques.

Furthermore, we applied this method in a real-case study involving electrical transmission towers in aerial images. Aerial images of transmission towers introduce a unique set of complexities for computer vision researchers, especially in object detection and instance segmentation tasks. These structures exhibit diverse configurations and are predominantly composed of linear elements with varying degrees of sparsity. Additionally, such images are marked by significant variations in background scenery, lighting conditions, and the relative sizes of objects across different captures. This case study not only tests the effectiveness of our proposed method in a challenging real-world scenario but also demonstrates the potential for broader application in similar settings.

The structure of this paper is outlined as follows. Section 2 provides an overview of the background, along with a concise description of the methods and techniques used. The experimental study is detailed and analysed in Section 3.