ABSTRACT

Wildlife conservation is more important than ever to protect biodiversity and keep the balance of the ecosystem. In recent years, machine learning and deep learning have been spreading in the computer vision field and gained huge success. The advancements in this field have also made contributions to research on wildlife biology. However, considering the fact that wildlife usually lives in complex nature surroundings, research on wildlife biology still confronts a number of challenges. For instance, the identification of wildlife could be disturbed by its habitat. To reduce the effect of living surroundings and ensure that the further analysis can focus on the animals of interest, an effective strategy could be foreground/background segmentation.

Background subtraction is one of the traditional segmentation techniques and includes temporal median filter as well as statistical background modeling. These approaches might perform well in relatively simple scenarios, but they also have many constraints. Deep learning segmentation models have been proven to be effective in previous research, especially the Mask R-CNN of two-stage segmentation model and the YOLACT of one-stage segmentation model, as well as their variants. On the other hand, it requires sufficient ground-truth data of wildlife segmentation to train these models, which poses challenges in view of the limited datasets.

In this thesis, a framework is developed as a tool for automatically segmenting wildlife in video sequences that is not limited to only a few certain species. It contains mainly three components, i.e., a YOLOV5-based detection model "MegaDetector", a foundation segmentation model Segment Anything Model (SAM), and a Video Object Segmentation (VOS) model "Cutie". In addition, a matching procedure and post-processing were implemented to overcome the issue of multiple overlapping animals in video sequences. As both SAM and MegaDetector were trained with extensive datasets, they demonstrate outstanding performance by general segmentation tasks and wildlife detection tasks, and thus the framework directly employed their pre-trained models without fine-tuning and domain adaption.

The framework was tested quantitatively with five high-resolution leopard video clips from the Pan African Programme and achieved a score (Mask IoU between predicted masks and ground-truth masks) of over 85%. Moreover, the framework was tested qualitatively with two YouTube low-resolution videos, which contain multiple overlapping animals. The results are reliable in the majority of cases.

Keywords: Wildlife Conservation, Automatic Segmentation, Foundation Segmentation Model, Video Object Segmentation

ZUSAMMENFASSUNG

Der Schutz von Wildtieren ist heutzutage wichtiger denn je, um die biologische Vielfalt zu schützen und das Ökosystem im Gleichgewicht zu halten. In den letzten Jahren haben Machine Learning und Deep Learning im Bereich der Computer Vision Verbreitung gefunden und dort Erfolge erzielt. Die Fortschritte in diesem Feld haben gleichzeitig auch einen Beitrag zur Wildtierforschung geleistet. Gleichwohl bestehen durch die Tatsache, dass Wildtiere üblicherweise in komplexen Naturumgebungen leben, nach wie vor einige Herausforderungen, mit denen sich die Wildtierbiologie konfrontiert sieht. Beispielsweise könnte etwa die Identifikation von Wildtieren durch ihren Lebensraum gestört werden. Um den Effekt der lebenden Umgebung zu reduzieren und sicherzustellen, dass der Fokus der weiteren Analyse auf den sich im Mittelpunkt des Interesses befindlichen Tieren befindet, könnte Vordergrund/Hintergrund-Segmentierung eine wirksame Strategie sein.

Die Hintergrundsubtraktion gehört zu den traditionellen Segmentierungstechniken und beinhaltet den temporalen Medianfilter sowie die statistische Hintergrundmodellierung. Diese Ansätze könnten eine gute Performanz in relativ einfachen Szenarien aufweisen, wobei sie jedoch auch vielen Einschränkungen unterliegen. Deep Learning Segmentierungsmodelle haben sich in der bisherigen Forschung als effektiv erwiesen, was insbesondere für das Mask R-CNN Modell der zweistufigen Segmentierung und das YOLACT Modell der einstufigen Segmentierung sowie deren Variationen gilt. Andererseits sind ausreichende Ground-Truth-Daten über Wildtiersegmentierung erforderlich, um diese Modelle zu trainieren, was angesichts der begrenzten Datensätze eine Herausforderung darstellt.

Im Rahmen dieser Thesis wird ein Framework als Werkzeug zur automatischen Segmentierung von Wildtieren in Videosequenzen entwickelt, welches sich nicht nur auf einige wenige bestimmte Tierarten beschränkt. Es enthält vornehmlich drei Komponenten, nämlich ein auf YOLOV5 basierendes Detektionsmodell "MegaDetector", ein grundlegendes Segmentierungsmodell "Segment Anything Model" (SAM) sowie ein Video Object Segmentation (VOS) Modell "Cutie". Außerdem wurde ein Matching-Verfahren und eine Nachbearbeitung implementiert, um das Problem von sich mehrfach überlappenden Tieren in Videosequenzen zu lösen. Da sowohl SAM als auch MegaDetector mit umfangreichen Datensätzen trainiert wurden, weisen sie eine herausragende Performanz bei allgemeinen Segmentierungsaufgaben und Wildtierdetektierungsaufgaben auf, weshalb ihre vortrainierten Modelle ohne Fine-tuning und Domänenanpassung direkt im Framework eingesetzt werden.

Das Framework wurde quantitativ mit fünf hochauflösenden Leoparden-Videoclips des Pan African Programme getestet und erreichte einen Score (Masken-IoU zwischen vorhergesagten Masken und Ground-Truth-Masken)

von über 85%. Zusätzlich ist das Framework qualitativ mit zwei niedrigauflösenden YouTube-Videos getestet worden, die mehrere überlappende Wildtiere enthalten. In den meisten Fällen sind die Ergebnisse reliable.

Schlagwörter: Wildtierschutz, Automatische Segmentierung, Grundlegendes Segmentierungsmodell, Video Objekt Segmentierung