KI-gestütztes Videomonitoring aquatischer Ökosysteme - Chancen, Grenzen und Praxisanforderungen

Die manuelle Verarbeitung von Fischvideos von Unterwasserkameras erfordert einen erheblichen Personalaufwand, was den Einsatz von videobasiertem Monitoring bisher einschränkt. Fortschritte in der Computer Vision ermöglichen es, mit Methoden der künstlichen Intelligenz (KI) die Umweltbedingungen automatisch zu klassifizieren, das Vorhandensein von Fischen zu erkennen, die Arten zu klassifizieren sowie ihre Gesamtkörperlänge und Schwimmrichtung zu schätzen. Trotz des Versprechens der KI, diese arbeitsintensiven Aufgaben zu automatisieren, gibt es auch erhebliche technische und praktische Herausforderungen. In diesem Artikel stellen wir wichtige Erkenntnisse, Leitlinien und Grenzen für den Einsatz von KI für das automatisierte Monitoring von Fischen mithilfe von Videomonitoring vor.

Jeffrey A. Tuhtan, Bernd Mockenhaupt, Frederic de Schaetzen, Rodrigo W. Issa, Adrien Bertaud, Damien Sidler, Henning Müller und Christian Haas

1 Einleitung

KI-Methoden zum Erkennen von Fischen, zur Klassifizierung von Fischarten und zur Abschätzung ihrer Größe und Ermittlung der Schwimmrichtung wurden in den letzten zwei Jahrzehnten entwickelt [1], [2]. Sie haben sich bislang jedoch nicht als Standardverfahren in der fischökologischen Videoanalyse etabliert. Die stärkste Konzentration von Forschung und praktischer Anwendung KI-basierter Überwachungssysteme findet sich bislang im marinen Bereich [3], insbesondere bei Forschungsinstituten und Technologieanbietern aus dem Bereich der Meeresbiologie und Aquakultur [4]. Dies führte in der Vergangenheit dazu, dass Forschungsgelder und industrielle Entwicklungen primär auf maritime Videomonitoring-Systeme ausgerichtet waren. Aufgrund der Fortschritte bei der Entwicklung kostengünstigerer Digitalkameras hat der Einsatz von Unterwasserkameras in den letzten zwei Jahrzehnten zugenommen, wobei der Schwerpunkt im Videomonitoring an Fischaufstiegsanlagen (FAA) liegt [5], [6]. Darüber hinaus haben die in den europäischen Ländern umgesetzten regulatorischen Rahmenbedingungen, wie die Wasserrahmenrichtlinie (WRRL) der Europäischen Union, den Bedarf an Fischbestandsuntersuchungen und -Überwachungen mit robusten Videomonitoring-Lösungen erhöht.

Aufgrund der Konvergenz erschwinglicherer Unterwasserkameras und der gestiegenen regulatorischen Nachfrage, sie für die flussbasierte Überwachung von Fischen einzusetzen, hat der Einsatz dieser Systeme stark zugenommen [7], [8]. Ein struktureller Nachteil kamerabasierter Systeme liegt in der erforderlichen manuellen Videoauswertung durch Fachpersonal, um Informationen, wie die Fischart, Fischgröße, Schwimmrichtung oder besondere Verhaltensmuster, zu dokumentieren. Die manuelle Auswertung von Unterwasservideos kann äußerst zeitintensiv sein und stellt ein zentrales Hemmnis für den flächendeckenden Einsatz videobasierter Monitoringverfahren dar [9].

Daher wird in einigen Ländern an einer automatisierten Auswertung von Unterwasservideos aus FAA und anderen zu überwachenden Wanderkorridoren von Fischen mithilfe von KI gearbeitet [10], [11], [12], [13].

Die aktuellen Fortschritte in der KI für Videomonitoring in Fließgewässern basieren weitgehend auf dem YOLO-Framework (You Only Look Once) [13], das die Objekterkennung und Klassifizierung von Fischen auf jedem Frame (Einzelbild des Videos) durchführt. Zusätzlich zu diesem Ansatz wurden auf Fische spezialisierte KI-Modelle entwickelt und getestet, mit dem Ziel, die Klassifizierungsgenauigkeit zu verbessern. KI-basierte Methoden zur Ermittlung der Schwimmrichtung erfordern darüber hinaus eine Objektverfolgung, die es notwendig macht, dass jedes Objekt mit einer eindeutigen Kennung versehen wird und dass das KI-Modell Fische über die Dauer des Videos kontinuierlich identifizieren sowie verfolgen kann. Da die KI-Methoden mehrere Algorithmen zur Erkennung, Klassifizierung und Verfolgung enthal-

Kompakt

- KI-gestützte Systeme ermöglichen die automatische Erkennung von Fischen, deren Artklassifizierung sowie die Abschätzung ihrer Körperlänge und Schwimmrichtung.
- Lokale Unterwasser-Umweltbedingungen, insbesondere Trübungen und Biofilmaufwuchs, haben einen starken negativen Einfluss auf die KI-Ergebnisse für die Klassifizierung der Fischart und Größenschätzung
- Die Qualität der Annotationen durch menschliche Bewerter hat erheblichen Einfluss auf die Modellgüte - sie wird in der Praxis jedoch häufig unterschätzt.

ten, erfordern sie erhebliche Fachkenntnisse in Computer Vision und Programmierung, um sie anzupassen, und dies stellt eine der Hauptbarrieren für die praktische Anwendung durch nicht-technisches Fachpersonal dar. Neben dieser Expertise stellt der Einsatz von KI-Methoden zur Automatisierung des kamerabasierten Fischmonitorings auch standortspezifische Anforderungen.

Damit KI-Monitoringlösungen an jedem Standort effektiv funktionieren, müssen die Nutzer zunächst eine beträchtliche Menge an Unterwasservideos von einer Vielzahl lokaler Umweltbedingungen sammeln, darunter Trübung, schwaches Licht, Überbelichtung, Schmutz und Blasen sowie Biofilmaufwuchs durch z. B. Algen [14], und auch mit unterschiedlichen Kameras aufgenommen werden, da auch diese die Klassifikation stark beeinflussen. Diese Umgebungsbedingungen sind bei fast jeder Installation relevant und wirken sich stark auf die visuellen Eigenschaften der Videos selbst aus, was zu einer unterschiedlichen KI-Leistung führen kann. Es ist erwähnenswert, dass die Videoqualität in fest installierten Infrastrukturen, wie Kameratunneln, die in einer FAA installiert sind, in der Regel höher ist als bei eigenständigen Kameras, die in frei fließenden Flussabschnitten angebracht sind. Neben den Herausforderungen, die sich durch die sich verändernden lokalen Umweltbedingungen ergeben, ist die Tatsache, dass Artenvielfalt und die altersspezifische Größe von Fischen an den Anlagen nicht homogen verteilt sind. Das bedeutet, dass viele Arten und Größenklassen in den Videos, die bei jeder Anlage aufgenommen werden, über- oder unterrepräsentiert sein können. Dies entspricht dem in der Fachliteratur etablierten Begriff des Klassenungleichgewichts (class imbalance) [15]. Um diese Herausforderungen zu bewältigen, ist es notwendig, Videos an anderen, idealerweise ähnlichen Standorten zu sammeln, um eine ausreichende Datenbasis für das Training robuster KI-Modelle zu schaffen.

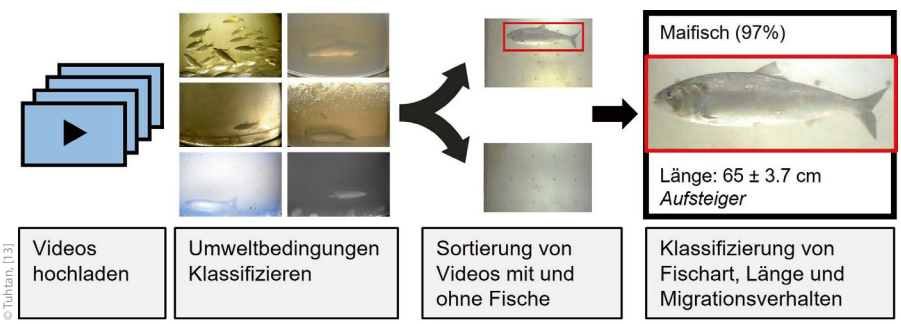
Sobald eine ausreichende Menge an Videos gesammelt wurde, wird das KI-Modell trainiert und getestet, und die Modellparameter können weiter optimiert werden, um die Erkennungs- und Klassifizierungsleistung zu verbessern. Der letzte Schritt, die Validierung, erfolgt anhand einer Reihe von Videos, die während des Trainings und Testens nicht verwendet wurden, und bietet dem Benutzer eine angemessene Schätzung der realen Leistung. Falls Kameras oder die Umgebung (Licht, Biofilmaufwuchs, mehr Sedimente, Wasserstand) sich ändern, müssen die Modelle eventuell nachtrainiert werden, um weiter eine gute Qualität zu gewährleisten. In diesem Artikel skizzieren wir die wichtigsten Schritte, die erforderlich sind, um sicherzustellen, dass ein robustes und zuverlässiges KI-Modell für die Fischdetektion, Klassifizierung der Fischart, Schätzung der Fischgröße und Ermittlung der Schwimmrichtung (und Rheotaxis beim Abstieg) implementiert werden kann. Des Weiteren werden die grundlegenden Modellierungsschritte skizziert und praxisrelevante Empfehlungen formuliert.

2 Methodik

Generell und weitgehend unabhängig vom konkret verwendeten Modell folgt der Einsatz von KI-Methoden einer Reihe von Modellierungsschritten (**Bild 1**). Der erste Schritt besteht darin, die lokalen Umgebungsbedingungen zu identifizieren, insbesondere solche wie Trübung oder Videos bei schlechten Lichtverhältnissen, die typischerweise mit einer geringeren Modellleistung verbunden sind. Als nächstes müssen innerhalb jedes Videos Fische erkannt werden, und Teile der Videos ohne Fische können in der ersten Stufe vom KI-Modell identifiziert und aus der weiteren Verarbeitung entfernt werden. Dies ist wichtig, da dadurch ein Großteil der von den Modellen benötigten Rechenleistung nicht beeinträchtigt und somit die Gesamtverarbeitungszeit beschleunigt wird. Die restlichen Videos mit Fischen werden dann auf Frame-by-Frame-Basis (oder mittels key frames) für die Arten- und Größenklassifizierung aufbereitet. An Stellen, an denen sich die Fische kontinuierlich durch die Videos bewegen, werden schließlich Tracking-Algorithmen ausgeführt, um die Schwimmrichtung der Fische zu bewerten. Diese Informationen können verwendet werden, um ihre Migrationsbewegungen weiter anzuzeigen.

Unabhängig von der Wahl des KI-Modells ist es zunächst notwendig, die Daten zu sammeln und zu organisieren, die zum Trainieren, Testen und Validieren des Modells benötigt werden. Um dies auf eine Weise zu tun, die einen möglichst reproduzierbaren Vergleich zwischen den Modellen ermöglicht, und um die Daten effektiv für das Benchmarking der Modellergebnisse zu teilen, empfehlen wir, die folgenden grundlegenden Leitlinien zu befolgen:

- Metadaten der Kamera, einschließlich des Standorts, des Namens und einer eindeutigen Kennung, Informationen über den Kameratyp, die Videoauflösung und die verwendete Bildrate (z. B. Koblenz, Deutschland, SiteID_A23, Sony IMX485, 3840x2160 px, 15 FPS).
- Videodateien, die sequenziell benannt werden, indem sie ihre Zeitstempel für den Start und das Ende der Aufzeichnung verwenden: Jahr-Monat-day_hour-Minute-Sekunde (z. B. eine Videodatei mit dem Namen „2025-09-21_10-34-33_10-41-26“).
- Extrahierte Frames aus ausgewählten Videos, die nach Videotyp, Art und Schwimmrichtung in Ordnern abgelegt werden (z. B. ein Ordner mit dem Namen „fisch_forelle_flussaufwärts“).



© Tuhtan [13]

Bild 1: Überblick über den KI-basierten Arbeitsablauf zur Überwachung von Videofischen

- Beschreibung der Umgebungsbedingungen innerhalb jedes ausgewählten Rahmens (Frames), gespeichert als Metadatendatei, die jeden extrahierten Rahmennamen mit einem Satz von Umgebungsbedingungen verknüpft (z. B. wird frame_01 mit „hoher Trübung“ und „geringem Biofilmaufwuchs“ in Verbindung gebracht).
- Begrenzungsboxen, die den Fischkörper in jedem Frame zusammen mit der Artenklassen-ID erfassen, die in der Regel im Dateiformat YOLO*.txt gespeichert werden: [class_ID x_center y_center width height].
- Frames, die einer Datenbilderweiterung unterzogen werden (z. B. Bilder werden gedreht, Rauschen wird hinzugefügt, Unschärfe).
- Trennen Sie die Trainings- und Validierungsdaten in separate Ordner, um eine Vermischung von Datasets zu vermeiden (z. B. ein Ordner mit dem Namen „train“ und ein anderer mit dem Namen „valid“).

Die Modellleistung wird häufig anhand des F1-Scores bewertet, der bewertet, wie gut die KI die verschiedenen Fälle korrekt identifiziert hat. In diesem Zusammenhang könnte ein Fall das Vorhandensein von Fischen, die richtige Art, Größenklasse und/oder Schwimmrichtung sein. Der F1-Score wird als harmonisches Mittel aus Precision und Recall berechnet (Gl. (1)). Precision (Gl. (2)) befasst sich mit der Frage: „Wie viele von all den Fällen, von denen das KI-Modell sagte, dass sie korrekt identifiziert wurden, waren tatsächlich richtig?“. Recall (Gl. (3)) bewertet die Frage „Wie viele von allen Fällen, die im Datensatz vorhanden sind, hat die KI tatsächlich gefunden?“

Um die Precision und den Recall zu berechnen, werden vier Werte benötigt, die aus den Ergebnissen entnommen werden müssen: Richtig Positiv (True Positive), Falsch Positiv (False Positive), Richtig Negativ (True Negative) und Falsch Negativ (False Negative). In Bezug auf die Fischdetektion ist ein „Richtig Positive“ die korrekte Erkennung eines Fisches durch die KI in einem Frame, ein „Falsch Positive“ ist eine KI-Erkennung eines Fisches, obwohl kein Fisch vorhanden war; ein „Richtig Negativ“ ist, wenn kein Fisch vorhanden ist und die KI korrekterweise keinen Fisch erkennt, und ein „Falsch Negativ“ ist, wenn die KI keinen Fisch im Frame findet, aber ein Fisch vorhanden war. Daraus ergibt sich:

F1-Score = 2 × (Precision × Recall) / (Precision + Recall) (1)

Precision = Richtig Positiv / (Richtig Positiv + Falsch Positiv) (2)

Recall = Richtig Positiv / (Richtig Positiv + Falsch Negativ) (3)

Das harmonische Mittel gewichtet sowohl Precision als auch Recall gleichermaßen, wenn also einer dieser Werte sehr niedrig ist, ist auch der F1-Score relativ niedrig. Dies verhindert, dass das KI-Modell einen hohen F1-Score erreicht, indem es nur bei Precision oder Recall hervorragend ist, während es den jeweils anderen ignoriert. Ein Erkennungsmodell, das beispielsweise alle Fische in den Videoframes korrekt erkennt (nur True Positives ohne False Negatives), aber auch die Hälfte der Videoframes ohne Fische als „Fisch“ markiert, (False Positives), hätte einen perfekten Recall, aber eine schlechte Precision, was zu einem insgesamt mittelmäßigen F1-Score führt. Zusätzlich zum F1-Score ist es empfehlenswert, eine Konfusionsmatrix zu erstellen, die die Anzahl der richtig positiven, richtig negativen, falsch positiven und falsch negativen Ergebnisse für jeden Modellierungsschritt darstellt, um die Modellleistung im Detail genauer bewerten zu können, vor allem wenn auch mehrere Sorten Fische unterschieden werden sollen. Generell ist zu beachten, dass eine größere Anzahl von falsch positiven Ergebnissen - im Gegensatz zu falsch negativen - für Fischüberwachungssysteme wünschenswert ist, um sicherzustellen, dass so wenig Fische wie möglich unentdeckt bleiben.

Diese Metriken sind interessant, um die Leistung über einzelne Bilder hinweg zu bewerten. Man sollte jedoch bedenken, dass bei Fischwanderungen die Endanforderung darin besteht, Videos zu verarbeiten, in denen das Vorbeiziehen eines Fisches über mehrere Bilder hinweg erfolgt; man kann die Metriken also auch per Fisch berechnen und nicht nur per frame. Insbesondere ist beim Vorbeiziehen eines Fisches an der Kamera dieser beim Ein- und Ausschwimmen nicht vollständig sichtbar, was bei diesen Bildern zu schlechten Ergebnissen führen kann. Die zuvor beschriebenen Metriken können auch auf der globaleren Ebene des Durchgangs eines Fisches angewendet werden (**Bild 2**).

tatsächliche Klasse \ ermittelte Klasse (*n* = 100)	positiv	negativ
positiv	46 richtig positiv	4 falsch negativ
negativ	11 falsch positiv	39 richtig negativ

© Tuhtan

Bild 2: Beispiel für eine Konfusionsmatrix für ein Fischdetektionsmodell, das die richtig positiven, richtig negativen, falsch positiven und falsch negativen Ergebnisse zeigt: In diesem Beispiel gibt es insgesamt 100 Videos, in der Hälfte der Videos kommen Fische vor, in der anderen Hälfte nicht; das Modell identifiziert 46/50 Videos mit Fischen und 39/50 Videos ohne Fische richtig, die restlichen Videos sind entweder falsch negativ (4) oder falsch positiv (11). Dies zeigt, dass das Modell verbessert werden muss, um die Anzahl der falsch-positiven Ergebnisse zu verringern

2.1 Umgebungsbedingungen

Trotz der bekannten Probleme optischer Systeme, die mit Umgebungsbedingungen, wie Trübung, Biofilmaufwuchs (Biofilmbildung von Mikroorganismen wie Algen und Bakterien) und Ablagerungen, verbunden sind, gibt es nur wenige Beispiele für die explizite Einbeziehung dieser Bedingungen in die KI-Verarbeitung. Um dies zu beheben, empfehlen wir, dass jeder Frame, der zum Trainieren, Testen und Validieren des KI-Modells verwendet wird, mindestens eine binäre Kennzeichnung enthält (oder mit mehr als zwei Kategorien für den Grad der Trübung), die angibt, ob der Frame Trübung aufweist oder nicht. Diese Kennzeichnung kann als numerischer Wert von „0“ für keine Trübung und „1“ für Frames mit erheblicher Trübung dargestellt werden. So kann das Modell unter verschiedenen Bedingungen bewertet werden und die Metriken mit und ohne Trübung zu berechnen. Es ist denkbar,

Tabelle 1: Empfohlene Einstufungskriterien für die fünf Arten von Umweltbedingungen, die sich negativ auf KI-basierte Fischüberwachungsmodelle auswirken (Quelle: Tuhtan)

Schweregrad-Klasse	Niedrig	Mittel	Hoch
Biofilmaufwuchs	<25 % des Rahmens sind durch Biofilmaufwuchs (an beliebiger Stelle) beeinträchtigt	25 % - 50 % des Rahmens sind durch Biofilmaufwuchs beeinträchtigt	>50 % des Rahmens sind durch Biofilmaufwuchs beeinträchtigt
Blasen/Schmutz	Ein Teil des Rahmens, aber <25 % sind durch Blasen und/oder Schmutz beeinträchtigt	25 % - 50 % des Rahmens sind durch Blasen und/oder Schmutz beeinträchtigt	>50 % des Rahmens sind durch Blasen und/oder Schmutz beeinträchtigt
Schwache Beleuchtung	10 % bis 25 % des Rahmens beinhalten zu dunkle oder schwarze Pixel	25 % bis 50 % des Rahmens beinhalten zu dunkle oder schwarze Pixel	>50 % des Rahmens beinhalten zu dunkle oder schwarze Pixel
Überbelichtung	10 % bis 25 % des Rahmens beinhalten zu helle oder weiße Pixel (an einer beliebigen Stelle im Bild)	25 % bis 50 % des Rahmens beinhalten zu helle oder weiße Pixel	>50 % des Rahmens beinhalten zu helle oder weiße Pixel
Trübung	Das Bild ist etwas eingetrübt, aber die Hintergrundmerkmale der Umgebung sind deutlich zu erkennen	Das Bild ist eingetrübt und die Hintergrundmerkmale sind kaum erkennbar	Das Bild ist stark eingetrübt, und die Hintergrundmerkmale sind nicht erkennbar
Klar	Wird nur vergeben, wenn keine Trübung erkannt wird, um eine Sortierung von Videos mit allgemein guter Bildqualität zu ermöglichen.		

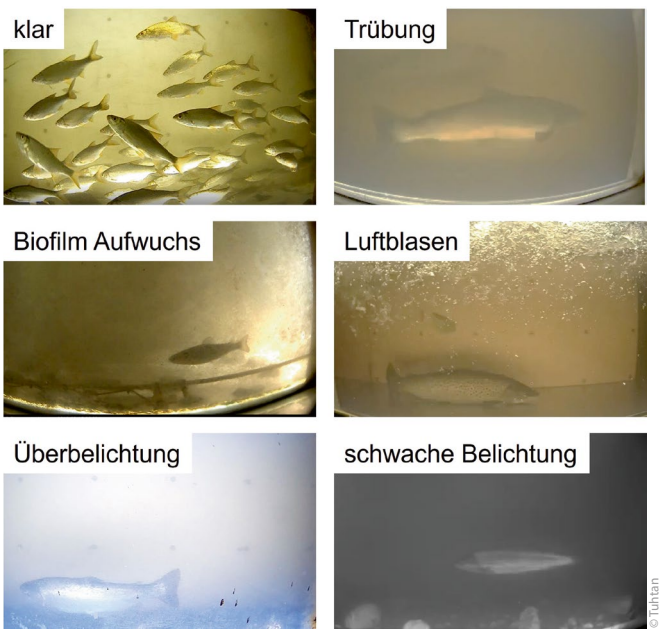
zwei spezialisierte Modelle zu haben: eines mit Trübung und eines ohne. In der Inferenz bedeutet dies jedoch, dass die Trübung automatisch erkannt wird, um das entsprechende Modell auszuwählen, und somit ein weiterer algorithmischer Schritt hinzugefügt wird

Um die Gesamtleistung des KI-Modells über ein breiteres Spektrum von Umgebungsbedingungen hinweg weiter zu verbessern, empfehlen wir jedoch, dass die Systeme eine detailliertere Analyse auf der Grundlage von **Tabelle 1** einbeziehen. Diese fünf empfohlenen Umgebungsbedingungenklassen können schnell in die Metadaten für jedes Bild eingepflegt werden und umfassen bestimmte Mengen an Biofilmaufwuchs, Blasen/Ablagerungen, geringe Beleuchtung, Überbelichtung und Trübung, wie in **Bild 3** gezeigt. Ein zusätzlicher Metadatenwert „klar" kennzeichnet Sequenzen mit besonders hoher Bildqualität. Diese Klasse wird vom KI-Modell automatisch zugewiesen, wenn keine Trübung festgestellt wird. Sie ist nützlich für das Aussortieren von Videos in der Nachbearbeitung mit der höchsten Gesamtqualität und kann bevorzugt zur Annotation von Arten und Größen verwendet werden, bei denen die Gesamtgenauigkeit der menschlichen Bewerter am höchsten sein kann. Die Modelle sollten jedoch immer unter möglichst realistischen Bedingungen trainiert werden, um ein breites Spektrum an Situationen abzudecken, die auftreten können.

2.2 KI-Anwendungen im Fischmonitoring

2.2.1 Fischdetektion

Die Fischdetektion bezieht sich ausschließlich auf das Auffinden des Objektes „Fisch" im Bild. Die Klassifizierung beschreibt im Anschluss die Zuordnung zu einer Art oder Klasse. Die gebräuchlichsten Modellierungsframeworks für die Fischdetektion basieren auf den You-Only-Look-Once-Algorithmen (YOLO-Algorithmen) [16], [17]. Das YOLO-Framework partitioniert Eingabebilder (hier einzelne Videobilder) in ein räumliches Gitter, das auf das Bild angewendet wird, wobei jede Gitterzelle gleichzeitig Begrenzungsbox-Koordinaten, Objektkonfidenz und Klassenwahrscheinlichkeiten durch einen einzigen Convolutional-Neural-Network-Durchgang vorhersagt. Eine Begrenzungsbox ist eine rechteckige Form, die den Körper des Fisches umreißt. Sie wird sowohl vom menschlichen Bewerter verwendet, um einen Bereich eines Videobildes zu definieren, der einen Fisch enthält, als auch von den meisten KI-Modellen, um anzugeben, wo in jedem Bild ein Fisch erkannt und/oder klassifiziert wurde. Die allgemeine Recheneffizienz und der hohe Optimierungsgrad von YOLO sowie die große Menge an frei verfügbaren Codebeispielen und Benutzerdokumentationen haben es zum am häufigsten eingesetzten Verfahren zur Fischdetektion gemacht.



© Tuhtan

Bild 3: Beispiele für die am häufigsten auftretenden Umweltbedingungen: klar, Trübung, Biofilm Aufwuchs, Luftblasen, Überbelichtung und schwache Belichtung, diese spezifischen Bedingungen wurden ausgewählt, weil bekannt ist, dass sie sich negativ auf die Leistung von KI-basierten Fischüberwachungsmodellen auswirken

Es ist erwähnenswert, dass es mehrere Alternativen zu YOLO gibt, um Fische in Bildern zu erkennen, jede mit unterschiedlichen Stärken und Schwächen. Gängige alternative Methoden zu YOLO arbeiten in zwei Schritten: Zuerst werden Bereiche identifiziert, die Objekte enthalten könnten, dann wird bestimmt, um welche Objektklasse es sich handelt (z. B. ein Fisch oder Totholz, Luftblasen, Blätter usw.). Beispiele hierfür sind R-CNN und seine verbesserten Versionen, wie Faster R-CNN [18], das zwar genauer als YOLO sein kann, aber tendenziell langsamer verarbeitet wird.

Neuere Modelle nutzen Transformer-Architekturen [19], wie sie auch bei Sprachmodellen verwendet werden, zur Mustererkennung in Bildsequenzen, wodurch einige der manuellen Feinabstimmungen, die YOLO und andere Modelle erfordern, entfallen. Abhängig von der Erfahrung mit Computer-Vision-Software können auch hochgradig angepasste Algorithmen für Websites mit begrenzter Rechenleistung entworfen und implementiert werden. Da jedoch die Gesamtgenauigkeit und Benutzerfreundlichkeit für Fischüberwachungs-KI-Systeme am wichtigsten sind, werden YOLO und herkömmliche zweistufige Methoden derzeit für die Mehrheit der Benutzer empfohlen. Um Fische zu erkennen, ist es auch möglich, anstatt eine Region vorzuschlagen, die Frames binär zu klassifizieren, um zu sagen, ob sich darin Fische befinden. Ein Vorteil dieses Ansatzes ist die Vorauswahl von Videosequenzen für aufwändigere Verarbeitungen in einem Kontext, in dem nur eine Minderheit von Frames einen Fisch enthält. Modelle, die verwendet werden können, sind zum Beispiel MobileNet, GhostNet oder RestNet.

2.2.2 Klassifizierung der Fischart

Im Gegensatz zur Fischdetektion gibt es für die Klassifizierung der Fischart eine weitaus größere Auswahl an geeigneten und offen verfügbaren Modellen. Hybride Ansätze, die mehrere Techniken kombinieren, erzielen in der Regel bessere Ergebnisse als monolithische Modellarchitekturen bei der Klassifizierung der Fischart (man spricht häufig von Ensemble Learning oder Model Fusion). Zu den bisher erfolgreichsten Architekturen gehören Convolutional Neural Networks (CNN) wie ResNet [20]. Auch verwendet werden CNN in Kombination mit Support-Vector-Machine-Klassifikatoren (SVM-Klassifikatoren) oder der Einsatz von Transfer-Learning, wobei Konzepte aus großen Datensätzen erlernt werden, um die Leistung bei einer neuen Aufgabe zu verbessern. Studien, die Mask R-CNN verwendeten, erzielten starke F1-Werte von etwa 90 %, während ein ResNet-50-Klassifikator mit Transferlernen bei 12 Spezies F1-Werte von etwa 80 % erreichte [21], [22]. Eine andere Studie, bei der mehrere Bilder aus einem Video verwendet wurden, um Meerforellen (*Salmo trutta*) von anderen Klassen zu separieren, erreichte eine Genauigkeit von 97 % [11]. Transfer-Learning, inkrementelles Lernen und hierarchische Klassifizierungsstrategien haben sich als besonders effektiv erwiesen, wenn mit begrenzten und stark unausgewogenen Datensätzen gearbeitet wird, wie sie in Flussfischarten-Datensätzen üblich sind. Die große Variation in der Leistung spiegelt die signifikanten Auswirkungen der Umweltbedingungen und spezifischer Datensatzmerkmale auf den Gesamterfolg der Klassifizierung der Fischart wider.

Es ist zu beachten, dass die Klassifizierungsleistung zwischen den verschiedenen Fischarten erheblich variiert, wobei häufig vorkommende und visuell unterscheidbare Arten (z. B. Flussbarsche (*Perca fluviatilis*)) höhere Genauigkeitsraten erreichen als seltene Arten. Studien berichten immer wieder, dass subtile Unterschiede zwischen den Arten, Überlappung der Fischkörper (z. B. wenn viele Fische gleichzeitig passieren), und Klassenungleichgewichte die größten systematischen Herausforderungen für die Genauigkeit der Klassifizierung einer Fischart darstellen. Seltene Arten sind aufgrund unzureichender Trainingsbeispiele besonders problematisch, während andere Arten aufgrund ähnlicher morphologischer Merkmale in der Regel eine geringere Klassifizierungsleistung aufweisen (z. B. Lachssmolts (*Salmo salar*) und junge Bachforellen (*Salmo trutta*)) oder kleinwüchsige Fische, die sich leicht in komplexe, natürliche Unterwasserumgebungen einfügen. Aufgrund dieser Herausforderungen erfordern KI-Modelle möglicherweise eine artspezifische Optimierung und eine sorgfältige manuelle Verifizierung für eine bessere Speziesidentifizierung.

2.2.3 Abschätzung der Fischgröße

Stereokamerasysteme sind der gebräuchlichste Ansatz zur Abschätzung der Fischgröße (**Bild 4**). Dies liegt daran, dass die Verwendung von zwei Kameras eine direkte Abschätzung der Entfernung eines erkannten Objekts zum Kamerasensor ermöglicht. Anhand dieser Informationen ist es dann möglich, die Größe der Begrenzungsboxen des detektierten Fisches zu ermitteln. Die Genauigkeitsraten von Stereokamerasystemen können je nach verwendeter Methode und Umgebungsbedingungen immer noch erheblich variieren und reichen von weniger als 1 % Fehler in gut kalibrierten kontrollierten Umgebungen mit klaren Wasserbedingungen bis hin zu 7 % bis 11 % Fehler unter schwierigeren Feldbedingungen [23], [24]. Monokulare oder Einzelkamerasysteme können die Größe des Fisches nicht direkt berechnen und stützen sich stark auf mehrstufige KI-Modelle, die versuchen, die Entfernung des Objekts zur Kamera abzuschätzen, oder die auf großen Mengen von menschlich annotierten Fischgrößen für eine feste

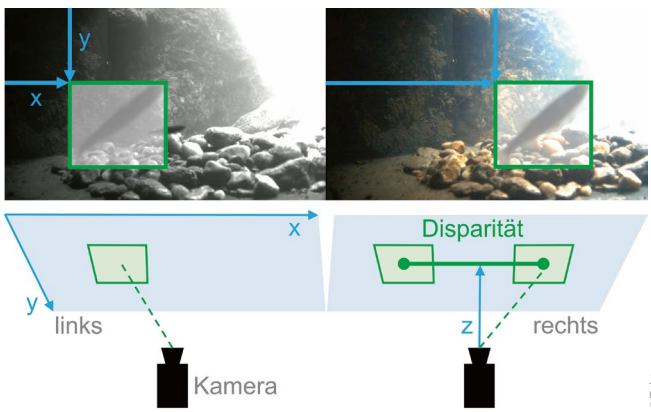


Bild 4: Die Stereokameras (links und rechts) sind in der Lage, die Größe des Fisches aufgrund der unterschiedlichen Position des Begrenzungsbox-Rahmens in jedem der Bilder zu schätzen, dieser Unterschied wird als Disparität bezeichnet und ergibt sich aus der Berechnung des Versatzes der Pixelpositionen (x, y) des Begrenzungsbox-Rahmens, der denselben Fisch in beiden Bildern umgibt; da der Abstand zwischen den einzelnen Kameras bekannt ist, ist die Disparität direkt proportional zum Abstand, z zwischen den Kameras und dem Fisch, mit Hilfe der Triangulation und der Kenntnis des Abstands z ist es dann möglich, die Länge des Fischkörpers innerhalb der Begrenzungsbox direkt zu schätzen

Installation basieren, bei der Größenklassen in 5-cm-Schritten mit Fehlern von +/-5 cm bis 20 cm oder mehr möglich sind.

Es gibt auch klare Zielkonflikte zwischen verschiedenen Ansätzen in Bezug auf die Anforderungen an die Bearbeitung und die Komplexität der Implementierung. Deep-Learning-basierte Stereosysteme bieten eine hochautomatisierte und robuste Schätzung der Fischgröße, erfordern jedoch hohe Rechenressourcen, eine komplexe Kalibrierung und umfangreiche Trainingsdaten. Daher können „Stereo+KI-Methoden" vielversprechende Ergebnisse bei Einzelbildern liefern, sind aber derzeit wahrscheinlich nicht nützlich für große Mengen an kontinuierlichem Videomaterial, wie sie beim Fisch-Monitoring anfallen. Im Allgemeinen erfordern herkömmliche Stereokamera-basierte Ansätze nur eine geringe oder gar keine Verwendung von KI-Modellen und können eine hohe Genauigkeit erreichen, wenn die Kameras mit Zielen bekannter Größe in festen Abständen von der Kamera kalibriert werden. In der Praxis sind diese Systeme jedoch aufgrund der Anforderung, mit zwei Kameras aufzunehmen, selten. Derzeit bieten monokulare Kamerasysteme die flexibelste und kostengünstigste Hardware; es ist jedoch zu erwarten, dass sie im Vergleich zu Stereokamerasystemen im Allgemeinen wesentlich weniger genaue Schätzungen der Fischgröße liefern.

2.2.4 Ermittlung der Schwimmrichtung

Derzeitige Methoden zur Abschätzung der Schwimmrichtung von Fischen in Flussumgebungen mit Unterwasserkameras können Stereo- oder Einzelkamerasysteme verwenden. Stereobildgebende Verfahren verwenden zwei oder mehr Kameras, um dreidimensionale Fischbewegungen mit hoher räumlicher Genauigkeit zu rekonstruieren, mit Fehlern von nur 2 cm und Geschwindigkeitsschätzungen innerhalb von 2 Körperlängen pro Sekunde [25]. Einzelkamerasysteme können auch mit Methoden, wie dem Bayes'schen Multiple-Blob-Tracker, getrackt werden und bieten eine kostengünstigere sowie einfachere Alternative zu Stereosystemen. Beide Methoden können Partikelfilter und Vorder-/Hintergrundmodellierung verwenden, um die Bewegung und Richtung der Fische zu verfolgen und eine Genauigkeit von bis zu 81 % zu erreichen [26]. Die Nachverfolgung funktioniert am besten in geometrisch begrenzten Umgebungen wie FAA und erfordert möglicherweise mehr Daten in natürlichen Umgebungen, in denen Fische hinter Felsen oder Holzresten hervorkommen können.

Ähnlich wie bei allen zuvor diskutierten Modellen beeinflussen Umgebungsbedingungen, wie Trübung, Lichtverhältnisse und Strömungsvariabilität, die Wirksamkeit von Tracking-Methoden erheblich. Stereo- und KI-basierte Systeme bieten zwar eine hohe Genauigkeit und Robustheit, sind aber mit einer höheren Betriebskomplexität und höheren Kosten verbunden. Im Gegensatz dazu sind Ansätze zum Fischtracking mit einer einzelnen Kamera einfacher, aber weitaus weniger präzise.

3 Diskussion

3.1 Empfehlungen für KI-basiertes Fischmonitoring

Um die automatisierte Identifizierung von Fischarten in Flussumgebungen effektiv zu implementieren, wird empfohlen, etablierte Deep-Learning-Modelle zu priorisieren, die in mehreren Computer-Vision-Anwendungen verwendet werden, insbesondere solche, die auf YOLO und Convolutional Neural Networks basieren. YOLOv5 und YOLOv3 haben eine hohe Erkennungs- und Klassifizierungsleistung gezeigt, wobei die F1-Werte über 90 % liegen. CNN-basierte Modelle, insbesondere in Kombination mit Support-Vektor-Maschinen oder verbessert durch Transferlernen, haben ebenfalls eine hohe Klassifizierungsgenauigkeit gezeigt, die bei der Klassifizierung der Fischart manchmal über 95 % liegt. Vorsicht ist geboten, da viele der YOLO-Implementierungen Lizenzbeschränkungen für die (kommerzielle) Nutzung unter der AGPL haben. Ähnliche Modelle sind RTMdet [27] und D-Fine [28] mit einer Apache-Lizenz.

Vielfältige und gut annotierte Datensätze mit mehr als zweitausend framebasierten Annotationen pro Art werden dringend empfohlen. Die Verwendung mehrerer menschlicher Bewerter ist ebenfalls unerlässlich, da alle Individuen eigene Ansichten haben, die dann von den Modellen „gelernt" und zu einem Artefakt werden, das sonst nach Abschluss der Annotationen schwer zu erkennen und zu korrigieren ist. Um dieses Problem zu lösen, wird empfohlen, dass mindestens zwei Personen eine Teilmenge des gesamten Annotationsmaterials erhalten, die redundant annotiert ist. Anschließend kann ein Vergleich zwischen zwei oder mehr menschlichen Annotatoren durchgeführt werden, um festzustellen, welche Unterschiede (z. B. Größenschätzung) vorhanden sind. Ein Korrekturalgorithmus, wie z. B. Regression, kann auf die Modellergebnisse angewendet werden, um die Abweichungen der Bewerter zu korrigieren.

Unter Berücksichtigung der Schätzung der Fischgröße bieten kalibrierte Stereokameras die einfachsten und zuverlässigsten Systeme. Einzelkamerasysteme können verwendet werden, um Größenklassenbewertungen an einer begrenzten Anzahl von Standorten durchzuführen, an denen die Videoqualität hoch ist und eine große Menge an von Menschen annotierten Größendaten zur Verfügung steht. Mit diesen Daten können benutzerdefinierte KI-Modelle mit modelleigenen Anpassungen auf standort- und artspezifischen Charakteristika erstellt werden. Ähnlich wie bei der Größenschätzung ermöglichen stereobasierte Systeme eine genauere Schwimmpfad-Rekonstruktion des Fisches, was die Leistung des KI-Modells bei der korrekten Ermittlung der Schwimmrichtung erhöht. Um einem einzelnen aufsteigenden Fisch eine durchgehende Spur adäquat zuordnen zu können, sind Videobildraten von 15 Bildern pro Sekunde oder höher erforderlich. Für den Abstieg in einem Bypass müssen 50 Bilder und mehr erreicht werden.

Offen zugängliche Benchmark-Datensätze, wie Fish4Knowledge, sind ein nützlicher Anfang für den Vergleich von KI-Modellen, aber neuere und größere Benchmark-Datensätze europäischer Fischarten werden von der Gemeinschaft dringend benötigt, um die wachsende Anzahl von KI-Modellen für das Fischmonitoring zu testen, zu trainieren und zu validieren, um einen zuverlässigen Quervergleich von Ansätzen unter Verwendung eines standardisierten Datensatzes zu gewährleisten, der unterschiedliche Umweltbedingungen einbezieht und berücksichtigt.

3.2 Die Notwendigkeit Umweltbedingungen zu berücksichtigen

Umgebungsbedingungen und Materialunterschiede (Typ der Kamera, Auflösung, Größe des Videotunnels) spielen eine kritische, aber stark unterschätzte Rolle für die Leistung von KI-Modellen.

Studien zeigen immer wieder, dass kontrollierte Umgebungen mit konsistenter Beleuchtung und hochauflösender Videoerfassung deutlich bessere Ergebnisse liefern als uneingeschränkte, sich dynamisch verändernde Flusslebensräume. In diesem Zusammenhang hat die Verwendung einer Infrarot-Hintergrundbeleuchtung zwei Vorteile. Erstens wird das Licht bei starker Trübung nicht von den Partikeln reflektiert und zweitens schwimmen die Fische unbeeinflusst, was bei weißem Licht nicht der Fall ist. Daher kann der Einsatz hochwertiger Unterwasserkameras und die Gewährleistung stabiler Lichtverhältnisse sowie die regelmäßige Entfernung von Biofilmaufwuchs die Genauigkeit des KI-Modells für die Erkennung, die Klassifizierung von Arten, die Größenschätzung und die Beurteilung der Schwimmrichtung erheblich und konsistent verbessern. Wenn die Umgebungsbedingungen nicht explizit als Teil des KI-Systems berücksichtigt werden, ist mit signifikanten Leistungseinbußen in allen Verarbeitungsstufen zu rechnen.

4 Zusammenfassung und Ausblick

Dieser Beitrag bietet einen systematischen Überblick über KI-basiertes Videomonitoring von Fischen in Fließgewässern und unterstreicht ihr Potenzial, die Artenerkennung, Klassifizierung, Größenschätzung und Ermittlung der Schwimmrichtung zu automatisieren. Die Implementierung von Deep-Learning-Modellen, insbesondere YOLO und Convolutional Neural Networks (CNN), hat eine hohe Leistungsfähigkeit gezeigt, insbesondere wenn sie von vielfältigen, gut annotierten Datensätzen unterstützt wird. Jüngste Fortschritte bei eingebetteter Hardware (Edge-Devices) mit geringem Stromverbrauch könnten neue Möglichkeiten bieten, kleinere, begrenztere KI-Modelle in Echtzeit mit den Kamera-Videostreams direkt auszuführen, sie erfordern jedoch weitere Tests und Entwicklungen. Umgebungsbedingungen, wie Trübung, Beleuchtung und Biofilmaufwuchs, beeinflussen die Genauigkeit des KI-Modells erheblich, so dass sie explizit in Trainings- und Validierungsworkflows einbezogen werden müssen. Stereokamerasysteme werden für eine präzise Größenschätzung und Ermittlung der Schwimmrichtung empfohlen, während Einzelkamera-Setups eine bessere Zugänglichkeit auf Kosten einer erheblich geringeren Genauigkeit bieten. Datenerfassungsprotokolle, einschließlich der Verwendung von Multi-Rater-Annotationen, sollten verwendet werden, um die Auswirkungen der Verzerrung durch menschliche Bewerter zu mildern.

Ein weiterer Aspekt, der bei Langzeitkamerasystemen zu berücksichtigen ist, ist die Datenspeicherung. Ein Beispiel: Eine Unterwasser-Fischzählkamera, die mit 50 Farbbildern pro Sekunde in HD-Auflösung (1 920 × 1 080) und H.264-Komprimierung filmt, erzeugt ein Video mit einer geschätzten Bitrate von etwa 8 Mbit/s. Dies ergibt eine Dateigröße von etwa 187 MB für eine einzelne 3-minütige Aufnahme. Wenn das System kontinuierlich in 3-Minuten-Segmenten aufzeichnet, würde es pro Tag etwa 480 Videos erzeugen, was täglich etwa 90 GB Speicherplatz erfordern würde. Im Laufe der Zeit ergibt dies etwa 630 GB pro Woche, 2,7 TB pro Monat (bei 30 Tagen) und etwa 32 TB pro Jahr. Diese Schätzungen bieten eine nützliche Grundlage für die Planung der Datenspeicherinfrastruktur für langfristige Überwachungsanwendungen.

Um die Gesamtleistung der algorithmischen Automatisierung zu messen und gleichzeitig dem Biologen die volle Kontrolle über seine Ergebnisse zu garantieren (no blackbox), ist es wichtig, ihm eine grafische Benutzeroberfläche zur Verfügung zu stellen, mit der er in den Prozess eingreifen (bzw. neu trainieren) kann. Es empfiehlt sich, das h_{vid}/h_{man}-Verhältnis zu verfolgen, d. h. die Anzahl der verarbeiteten Videostunden pro Stunde verbleibender menschlicher Arbeit (Datenverarbeitung + biologische Validierung der Ergebnisse). Mit zuverlässigen und gut trainierten Algorithmen sind über 500 Videos pro Stunde möglich, so dass man ein Jahr in wenigen Tagen verarbeiten kann. Es ist zu beachten, dass in einem Abschnitt der oberen Forellenregion mit nur einer Art und einigen hundert Fischpassagen pro Jahr die Automatisierung deutlich einfacher zu realisieren ist als in einem sehr artenreichen Gewässer mit hunderttausenden Fischbewegungen im Jahr.

KI-gestütztes Videomonitoring bietet großes Potenzial zur Automatisierung oder Teilautomatisierung fischökologischer Untersuchungen. Neben der Weiterentwicklung robuster Modelle müssen insbesondere Umweltbedingungen, Datenqualität und Annotatoreneinflüsse systematisch berücksichtigt werden, um verlässliche Ergebnisse zu erzielen. Die Entwicklung standardisierter europäischer Datensätze zur Qualitätssicherung und dem Vergleich unterschiedlicher Modelle bleibt eine Grundvoraussetzung, ebenso wie die enge Verzahnung technischer Entwicklung mit der praktischen Anwendung.

Dank

Wir möchten uns bei allen direkt und indirekt Beteiligten bedanken, die bei der Erstellung dieser gemeinsamen Arbeit über den aktuellen Stand und die Zukunft der KI für automatische Fischzähler mitgewirkt haben. Die Beiträge von Bernd Mockenhaupt, Christian Haas und Jeffrey Tuhtan wurden teilweise von der Bundesanstalt für Gewässerkunde finanziert. Der Beitrag von Jeffrey Tuhtan wurde außerdem teilweise durch das Stipendium PRG2198 des Estnischen Forschungsrats finanziert.

Autoren

Assoc. Prof. Dr.-Ing. Jeffrey A. Tuhtan
Technische Universität Tallinn, Abt. für Computersysteme
Gruppe für Umweltsensorik und -aufklärung
Akadeemia tee 15A-111
12618 Tallinn, Estland
jeffrey.tuhtan@ttu.ee

Dipl.-Biol. Bernd Mockenhaupt
Bundesanstalt für Gewässerkunde
Am Mainzer Tor 1
56068 Koblenz
mockenhaupt@bafg.de

Dr. sc. ETHZ Frederic de Schaetzen
Kraftwerke Oberhasli AG
Grimselstrasse 19
3682 Innertkirchen, Schweiz
frederic.deschaetzen@kwo.ch

Rodrigo W. Issa, M. Sc.
Fraunhofer-Institut für Graphische Datenverarbeitung IGD
Joachim-Jungius-Straße 11
18059 Rostock
rodrigo.wilhelm.issa@igd-r.fraunhofer.de

Adrien Bertaud, M. Sc.
HES-SO Valais/Wallis - Swiss digital center
Technopole 10
3960 Sion, Schweiz
adrien.bertaud@hevs.ch

Dipl. nat. Damien Sidler
FishLab GmbH
Rue de l'Industrie 23
1950 Sion, Schweiz
damien.sidler@fishlab.ch

Prof. Dr. Henning Müller
Informatics Insitute, HES-SO
Rue du Technopole 3
3960 Sierre, Schweiz
henning.mueller@hevs.ch

Dipl.-Ing. Christian Haas
I am Hydro GmbH
Leopoldstraße 1
78112 St. Georgen
christian.haas@iamhydro.com

Literatur

[1] Salman, A.; Siddiqui, S. A.; Shafait, F. et al.: Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. In: ICES J. Mar. Sci. 77 (2020), Heft 4, 1 295-1 307 (doi.org/10.1093/icesjms/fsz025).
[2] Marrable, D.; Barker, K.; Tippaya, S. et al.: Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning. In: Front. Mar. Sci. 9 (2022), Art. 944 582 (doi.org/10.3389/fmars.2022.944582).
[3] Nadarajan, G.; Yang, C.-L.; Chen-Burger, Y.-H.; Tolosana-Calasanz, R.; Rana, O. F.: Analysing quality of resilience in Fish4Knowledge video analysis workflows. In: Proc. IEEE/ACM Int. Conf. Utility Cloud Comput. (2013), S. 405-410 (doi.org/10.1109/UCC.2013.81).
[4] Ditria, E. M.; Lopez-Marcano, S.; Jinks, E. L.; Brown, C. J.; Connolly, R. M.: Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. In: Rev. Aquacult. 14 (2022), Heft 2, S. 904-916 (doi.org/10.1111/raq.12726).
[5] Mader, H.; Käfer, S.: FishCam & FishNet - fischökologisches Monitoring 4.0. In: Österr. Wasser-Abfallwirtsch. 72 (2020), S. 129-141 (doi.org/10.1007/s00506-020-00654-z).
[6] Haas, C.; Meyer, M.; Schmalz, M.; Tuhtan, J. A.: Kamerabasiertes Monitoring von Fischen in Deutschland und der Schweiz. In: WasserWirtschaft 114 (2024), Heft 10, S. 19-29 (doi.org/10.1007/s35147-024-2401-1).
[7] Haas, C.; Thumser, P.; Mockenhaupt, B.; Völker, F.: Das System Vaki-Riverwatcher als Möglichkeit für ein Langzeitmonitoring von Fisch-Migration in Fischaufstiegsanlagen. In: WasserWirtschaft 108 (2018), S. 41-48 (doi.org/10.1007/s35147-018-0186-9).
[8] Quaranta, E.; Bejarano, M. D.; Comoglio, C. et al.: Digitalization and real-time control to mitigate environmental impacts along rivers: Focus on artificial barriers, hydropower systems and European priorities. In: Sci. Total Environ. 875 (2023), Art. 162 489 (doi.org/10.1016/j.scitotenv.2023.162489).
[9] Vondrick, C.; Patterson, D.; Ramanan, D.: Efficiently scaling up crowdsourced video annotation. In: Int. J. Comput. Vis. 101 (2013), S. 184-204 (doi.org/10.1007/s11263-012-0564-1).
[10] Mockenhaupt, B.; Tuhtan, J.: BfG-Projekt Smart Fish Counter: Automatisierte Bewertung von Art, Größe und Migrationsverhalten durch Künstliche Intelligenz. In: WasserWirtschaft 113 (2023), S. 26-30 (doi.org/10.1007/s35147-023-1782-x).
[11] Tödtmann, H.; Vahl, M.; Freiherr von Lukas, U.; Ullrich, T.: Time-unfolding object existence detection in low-quality underwater videos using convolutional neural networks. In: VISAPP - Proc. 15th Int. Joint Conf. Comput. Vis., Imaging Comput. Graph. Theory Appl. (2020), Vol. 5, S. 370-377 (doi.org/10.5220/0008962803700377).
[12] Jansen, A.; van Bodegraven, S.; Esparon, A. et al.: Monitoring tropical freshwater fish with underwater videography and deep learning. In: Mar. Freshw. Res. 75 (2024), MF23166 (doi.org/10.1071/MF23166).
[13] Al Muksit, A.; Hasan, F.; Bhuiyan Emon, M. F. H. et al.: YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment. In: Ecol. Inform. 72 (2022), Art. 101 847 (doi.org/10.1016/j.ecoinf.2022.101847).
[14] Soom, J.; Pattanaik, V.; Leier, M.; Tuhtan, J. A.: Environmentally adaptive fish or no-fish classification for river video fish counters using high-performance desktop and embedded hardware. In: Ecol. Inform. 72 (2022), Art. 101 817 (doi.org/10.1016/j.ecoinf.2022.101817).
[15] Johnson, J. M.; Khoshgoftaar, T. M.: Survey on deep learning with class imbalance. In: J. Big Data 6 (2019), S. 27 (doi.org/10.1186/s40537-019-0192-5).
[16] Vijayalakshmi, M.; Sasithradevi, A.: AquaYOLO: Advanced YOLO-based fish detection for optimized aquaculture pond monitoring. In: Sci. Rep. 15 (2025), Srt. 6 151 (doi.org/10.1038/s41598-025-89611-y).
[17] Jalal, A.; Salman, A.; Mian, A.; Shortis, M.; Shafait, F.: Fish detection and species classification in underwater environments using deep learning with temporal information. In: Ecol. Inform. 57 (2020), Art. 101 088 (doi.org/10.1016/j.ecoinf.2020.101088).
[18] Yu, C.; Fan, X.; Hu, Z. et al.: Segmentation and measurement scheme for fish morphological features based on Mask R-CNN. In: Inf. Process. Agric. 7 (2020), Heft 4, S. 523-534 (doi.org/10.1016/j.inpa.2020.01.002).
[19] Gong, B.; Dai, K.; Shao, J.; Jing, L.; Chen, Y.: Fish-TViT: A novel fish species classification method in multi water areas based on transfer learning and vision transformer. In: Heliyon 9 (2023), Heft 6, Art. e16 761 (doi.org/10.1016/j.heliyon.2023.e16761).
[20] He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep Residual Learning for Image Recognition. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, 2016, S. 770-778 (doi.org/10.1109/CVPR.2016.90).
[21] Mathur, M.; Vasudev, D.; Sahoo, S. et al.: Crosspooled FishNet: transfer learning based fish species classification model. In: Multimed. Tools Appl. 79 (2020), S. 31 625-31 643 (doi.org/10.1007/s11042-020-09371-x).
[22] Ergün, E.: SwinFishNet: A Swin Transformer-based approach for automatic fish species classification using transfer learning. In: PLoS One 20 (2025), Heft 5, Art. e0 322 711 (doi.org/10.1371/journal.pone.0322711).
[23] Silva, C.; Aires, R.; Rodrigues, F.: A compact underwater stereo vision system for measuring fish. In: Aquacult. Fish. 9 (2024), Heft 6, S. 1 000-1 006 (doi.org/10.1016/j.aaf.2023.03.006).
[24] Ubina, N. A.; Cheng, S.-C.; Chang, C.-C. et al.: Intelligent underwater stereo camera design for fish metric estimation using reliable object matching. In: IEEE Access 10 (2022), S. 74 605-74 619 (doi.org/10.1109/ACCESS.2022.3185753).
[25] Li, Y.; Tan, H.; Deng, Y.; Zhou, D.; Zhu, M.: A two-stage 3D multi-fish tracking model using patch-based underwater stereo matching. In: Biosystems Engineering 250 (2025), S. 144-157 (doi.org/10.1016/j.biosystemseng.2024.12.011).
[26] Morais, E. F.; Campos, M. F. M.; Padua, F. L. C.; Carceroni, R. L.: Particle filter-based predictive tracking for robust fish counting. In: Proc. SIBGRAPI (2005), Natal, Brazil, S. 367-374 (doi.org/10.1109/SIBGRAPI.2005.36).
[27] Lyu, C.; Zhang, W.; Huang, H. et al.: RTMDet: An Empirical Study of Designing Real-Time Object Detectors. Preprint, arXiv:2212.07784 (2022) (arxiv.org/abs/2212.07784).
[28] Peng, Y.; Li, H.; Wu, P. et al.: D-FINE: Redefine Regression Task in DETRs as Fine-grained Distribution Refinement. Preprint, arXiv:2410.13842 (2024) (arxiv.org/abs/2410.13842).

DOI dieses Beitrags: http://doi.org/10.1007/s35147-025-2583-1

Jeffrey A. Tuhtan, Bernd Mockenhaupt, Frederic de Schaetzen, Rodrigo W. Issa, Adrien Bertaud, Damien Sidler, Henning Müller and Christian Haas

AI video monitoring of aquatic organisms - opportunities, limits and best practice recommendations

Manual processing of fish videos from underwater cameras can be very time-consuming, which currently limits the use of video-based monitoring. Recent advances in computer vision allow artificial intelligence (AI) methods to automatically classify environmental conditions, detect the presence of fish, classify species and estimate their total body length and swimming direction. However, despite the promises of AI to automate these labor-intensive tasks, there remain important technical and practical challenges. In this article, we present key findings, guidelines and limitations for the use of AI for automated monitoring of river fish using underwater camera videos.