Dutkat: A Multimedia System for Catching Illegal Catchers in a Privacy-Preserving Manner

Tor-Arne S. Nordmo
tor-arne.s.nordmo@uit.no
UiT The Arctic University of Norway
Tromsø, Norway

Aril B. Ovesen
UiT The Arctic University of Norway
Tromsø, Norway

Håvard D. Johansen
UiT The Arctic University of Norway
Tromsø, Norway

Michael A. Riegler*
SimulaMet
Oslo, Norway

Pål Halvorsen†
SimulaMet
Oslo, Norway

Dag Johansen
UiT The Arctic University of Norway
Tromsø, Norway

ABSTRACT

Fish crime is considered a global and serious problem for a healthy and sustainable development of one of mankind's important sources of food. Technological surveillance and control solutions are emerging as remedies to combat criminal activities, but such solutions might also come with impractical and negative side-effects and challenges. In this paper, we present the concept and design of a surveillance system in lieu of current surveillance trends striking a delicate balance between privacy of legal actors while simultaneously capturing evidence-based footage, sensory data, and forensic proofs of illicit activities. Our proposed novel approach is to assist human operators in the 24/7 surveillance loop of remote professional fishing activities with a privacy-preserving Artificial Intelligence (AI) surveillance system operating in the same proximity as the activities being surveyed. The system will primarily be using video surveillance data, but also other sensor data captured on the fishing vessel. Additionally, the system correlates with other sources such as reports from other fish catches in the approximate area and time, etc. Only upon true positive flagging of specific potentially illicit activities by the locally executing AI algorithms, can forensic evidence be accessed from this physical edge, the fishing vessel. Besides a more privacy-preserving solution, our edge-based AI system also benefits from much less data that has to be transferred over unreliable, low-bandwidth satellite-based networks.

CCS CONCEPTS

• **Computer systems organization** → **Distributed architectures**;
• **Computing methodologies** → **Activity recognition and understanding**; **Computer vision problems**; • **Security and privacy** → **Tamper-proof and tamper-resistant designs**.

*Also affiliated with UiT The Arctic University of Norway, Norway

†Also affiliated with Oslo Metropolitan University, Norway

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

ICDAR '21, August 21–24, 2021, Taipei, Taiwan

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8529-9/21/08...\$15.00

<https://doi.org/10.1145/3463944.3469102>

KEYWORDS

distributed system, AI, privacy, multimodal data analysis

ACM Reference Format:

Tor-Arne S. Nordmo, Aril B. Ovesen, Håvard D. Johansen, Michael A. Riegler, Pål Halvorsen, and Dag Johansen. 2021. Dutkat: A Multimedia System for Catching Illegal Catchers in a Privacy-Preserving Manner. In *Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval (ICDAR '21)*, August 21–24, 2021, Taipei, Taiwan. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3463944.3469102>

1 INTRODUCTION

Global fisheries contribute to food security for millions of people worldwide. It is estimated that the sea produces 17% of the current production of edible meat globally, and that this percentage will increase dramatically as the world's population grows [3]. Capture fisheries play a pivotal role for guaranteeing food security from the sea, but illegal fishing and over-exploitation are widespread problems that negatively impact the sustainability of wild fish stocks. Examples of such illegal activities include fishing vessels that operate without valid licenses, fishing on stocks that are depleted and close to extinction, fishing with illegal gears, and negligence in submitting correct catch data. These problems occur globally, but are perhaps particularly prominent around the poorer coastal countries that cannot launch effective counter-measures to combat such organized illegal activities.

In some geographical areas, like the northern Atlantic, evidence-based regulatory efforts have been shown to be efficient for maintaining sustainable harvesting of fish populations. A requirement, though, is that such regulations are combined with thorough fishing vessel monitoring, surveillance, and control. For instance by requiring vessels to have Automatic Identification Systems (AIS) that track position installed, have them maintain and submit logs to the authorities when fish gear is launched, when catch is landed on deck, and what species and volumes were caught, and where the catch will be delivered for further processing. Fishing vessels are also subject to an inspection at any time and without notification, both while at sea or when docking, by either the coast guard or other public inspectors.

Though such operational control regimes make cheating difficult, they are far from flawless. Failures to report a catch is one example of how criminals can cheat, under-reporting fish quantities landed

another. The monetary gain for such illegal activities can be relatively profitable; UN estimates that illegal fishing activities amount to figures in the order of billions USD a year, and involves corruption and other financial crimes, such as large-scale tax evasion and money-laundering [14].

Some argue that digital surveillance technologies can mitigate fishery crime problems. Automatic 24/7 video-surveillance systems and sensors on board the fishing vessels at sea are particularly touted as technological solutions. The Danish Fisheries Authority is already equipping tenfold fishing vessels from the country's fleet with such equipment, a mutually agreed decision with volunteering fishermen. Norway is also planning mandatory deployment of similar surveillance cameras as one of several remedies to manage, control, and combat illegal fishing activities [13].

At a first glance, such video surveillance might sound as a plausible approach to combat crimes, but comes with a significant disadvantage with regard to privacy. Suddenly, somebody can be watching every fisherman while working on deck, which can be considered personally invasive. The proportionality law principle is also at stake striking a balance between capturing potential illicit behaviour versus intruding on personal spheres of tenfold thousand of fishermen at work. In addition, the analysis of video data in combination with other sensory data and external data is a challenge by itself. Combining different modalities for analysis comes with challenges regarding how to combine them, how to determine the importance of the different input sources, etc. Scale and performance are obvious technical problems considering that fishing vessels in the order of 1,000s will have to transfer voluminous multimedia data over satellite links or radio-based technologies to mainland operational control centers. And, imagine the tedious and labor-intensive task if this data is to be inspected in real-time by human operators.

In this paper we propose *Dutkat*: a technical solution that balances the conflicting properties of surveillance, performance, and privacy. The system provides multimodal analysis of collected data sources, where at the same time the privacy of individuals is preserved, yet secure and reliable ground truth documentation of illegal activities are captured and stored for forensics purposes. The main idea is to replace humans in the parsing of multi-sensor data and thus guarantee confidentiality and privacy, while at the same time being able to analyze and detect efficiently that suspicious activities might occur.

2 SYSTEM ARCHITECTURE

We have devised an architecture for monitoring and storing surveillance data, coupled with algorithmic control of specific physical activities in remote spatial locations. Figure 1 illustrates this distributed architecture where (1) edge nodes located in the physical proximity of interest contain physical monitoring sensors, storage, and computation devices. These edge nodes are connected to a (2) more trusted, stationary hub structure federating edge data for persistent storage. This back-end server hub structure is located in a resource rich and secure proximity and typically connects to multiple edge nodes. Different (3) analysis components can be connected to this central hub, but it is also possible to connect such analysis components directly into data streams or on edge nodes.

Figure 1: System architecture. The colors indicate where the computation takes place, with blue on the vessel and orange on a mainland control site.

Finally, different (4) front-end user-specific modules for operational surveillance purposes, forensics, control operations and the like are provided.

Our main approach is to replace a human being in the analysis loop of a remote video stream, with an algorithm performing the analysis in the same proximity where video and related sensor data are captured. We enhance and complement video data with other sensor data, particularly from edge located Internet of Things (IoT) devices. Captured surveillance data is not sent remotely for human inspection, which could potentially violate privacy regulations, but is continuously analyzed by local AI algorithms. Consider this algorithm as resembling a human inspector locally on each of the vessels involved. On average, the current control regime in the coastal area of Norway has resources to inspect and control in the order of 5% of 1,000 fold fishing vessels, while our approach theoretically scales this to close to 100%. In practice, our system is intended to be deployed first on the bigger trawlers, which are large vessels already fully equipped with digital technology.

We are building a first prototype implementation of this architecture. This is of *Dutkat*, a distributed, privacy-preserving surveillance system where a mainland control center is connected to a number of mobile edge devices. *Dutkat* is intended to be deployed on off-shore trawlers with license to harvest fish in the Arctic Sea. We organize

the different software modules to belong in four different planes stacked upon each other. In the bottom is a (1) data monitoring plane, next a (2) data collection and storage plane, an (3) analysis plane, and a (4) user control plane. Notice that particularly plane 2 and 3 have software modules located both on mainland and on fishing vessels.

Specifically, we target detection of suspicious activities related to quota exploitation and discard of catch, but also to estimate catch volume and species distribution to compare against sales notes. This will also provide automatic documentation for fishermen. The design allows us to analyze multiple data sources, and reduces unnecessary transmission of data over low-bandwidth satellite networks.

Video and sensor equipment will be placed on deck to detect anomalous movement and actions, and cameras will also be placed in the fish processing pipeline for catch analysis. AIS positional data will also be recorded and sent to the mainland. The analysis will be performed on the fishing vessel, and the data and results will be securely stored on the vessel. An aggregated/summarized result will be sent to the mainland for potential flagging of the vessel.

The subsystem on land will analyze AIS data and sales notes, and the results of this will be combined with the results sent from the fishing vessels. The inspection authorities will then have a graphical user interface where they can see which vessels are flagged for inspection, before the vessel even lands.

3 DISCUSSION AND RELATED WORK

The fishing domain and use of distributed AI solutions is a domain we are very familiar with. Over three decades ago, we built a series of StormCast [7–10], probably the first sensor network prototype ever with distributed expert-systems built for trawlers and coast guard ships connected through satellite communication links. StormCast was predicting suddenly erupting hazardous weather situations at sea over large areas. A mainland public version of Stormcast has been continuously operational since 1993¹.

Dutkat will consist of multiple subsystems which are dependent on different models to perform multimodal analysis. There have been multiple works on the sub-problems of privacy-preserving analytics, human activity recognition, fish analysis, and tabular analysis.

Fitwi et al. [5] describe a system for masking private information in video frames from surveillance cameras by doing detection and filtering on the edge. In their paper, they discuss the advantages and disadvantages of a cloud-based surveillance system vs. an edge-based approach. They argue that some pre-processing (i.e., motion detection for filtering of scenes) can happen on the edge, but that deep learning-based object detection methods are too computationally intensive to perform on the edge. They introduce a chaos-based encryption scheme that is used on the data before it is sent to a server for processing. There, they perform window and people detection and face recognition. A notable difference though is that our edge is relatively resourceful where entire computers can be deployed for edge analysis. Currently, our select edge platform we are exploring consists of a cluster of Nvidia Jetson Xavier NX computers.

D'souza et al. [4] describe a similar system that uses object detection for surveillance camera video streams, and whitelists classes of objects that should *not* be censored. Our system consists of many similar components that allow us to preserve privacy. However, a major difference between Dutkat and these two works is our relative greedy use of bandwidth. The papers above describe ways of reducing bandwidth requirements and encrypting data before transfer, whereas we perform the data-intensive video analysis on the edge, and only send minimal information to the mainland site.

Rodriguez-Moreno et al. [15] describe the challenges and state-of-the-art methods for performing activity recognition from video. One needs to recognize the relevant areas that contain humans in the frames, and a variety of techniques can be applied. Hand-crafted feature extraction has been the traditional approach, while in more recent times, deep learning-based solutions have been preferred. In the more specific area of recognizing suspicious activity, the task becomes more difficult. In the Tripathy et al. [16] survey, they show usage of object detection, tracking, and activity recognition in different use-cases, however, in all these cases it is relatively simple to find examples of suspicious activity on video, as the settings are more conventional. In our case, this is not so, and we need to potentially apply transfer learning on simulated suspicious activity.

Marais [12] describes the difficulties of analyzing tabular data with ML techniques; numerical data is usually ordered and continuous, but can be discrete, while categorical data is usually unordered and discrete. This can make feature interactions difficult to learn. This is highly relevant to the analysis of the sales notes, as they are a composite of numerical and categorical features, therefore representing the data in a fitting manner is crucial.

Automatic fish analysis using ML can be difficult depending on what you aim to accomplish. All image-based ML methods are dependent on training data that reflects a large variety of conditions, e.g. lighting and water clarity. Segmentation can be particularly difficult with overlapping fish. Alsmadi et al. [1] discuss how fish classification has been done historically with traditional ML techniques, and different datasets and the data requirements of deep learning methods. We are definitely confident that deep learning methods can achieve state-of-the-art results, given a good dataset. Weight estimation of fish from images can be difficult to achieve, due to not having a 3D view and due to the density of the fish being dependent on species, feed, age, etc. Konovalov et al. [11] apply a LinkNet-34 segmentation network to segment fish in images. They then compare mathematical models for estimating weight of fish based on length, height, and/or segmented area, vs. using a convolutional network to directly estimate weight via regression. Garcia et al. [6] do measurement estimation of fish, but based on a stereoscopic view. They show that using stereo imaging makes segmentation easier in the case of overlapping fish. They utilize Mask RCNN to segment the fish, and then estimate the length of the fish via morphological operations. This could then be used in a mathematical model for weight estimation.

Baltruvisaitis et al. [2] discuss five major challenges of multimodal data analysis in ML. Of these, representation, i.e., exploiting complementarity and redundancy between different data sources, and fusion, i.e., joining information from different modalities to form a prediction are highly relevant to our use-case. The ML pipeline utilizes complementarity in e.g. the AIS and sales notes

¹<https://weather.cs.uit.no/>

data to predict some anomalous activities, and requires prediction from different models to form a final judgement on whether a fishing vessel should be inspected or not.

4 RESEARCH DIRECTIONS

The described system is addressing several aspects of a ML system pipeline. Due to the specific requirements of privacy, but also the challenges by the potential sheer volume of data and the use case itself, different possible research directions have been identified.

R1: Transparent AI system The system built for the use case needs to be transparent in all aspects. This means that the data and its distribution, how it was collected and annotated, how the analysis is performed, and how the results are presented need to be open and comprehensible. To achieve this explainability, the AI system needs to be created that is beyond most of the current black-box systems. Such a system will need to support ways of understanding data distribution, tackle domain shifting (e.g., seasonal differences in the data), and provide explainable and interpretable results.

R2: Multimodal data analysis The analysis requirements for such a system come with different challenges. One of the most important is to be able to analyze data from different sources combined. This in itself is a challenging task and not researched very well, especially in modern, deep learning based AI systems. For such a system, new methods within the field of deep learning for efficient combination of data that go beyond simple late or early fusion need to be researched and developed. This cross-data analysis also includes ways to explain the importance of the different data sources to the outcome of the analysis.

R3: Privacy-preserving AI Analysis involving data gathered from human beings potentially imposes new risks or negative consequences for individuals. Privacy is certainly at risk when, for instance, video data is collected for manual inspections or algorithmic analysis. As such, compliance issues must be a first-order design principle in next-generation novel surveillance systems like ours. This important concern for individuals right for privacy is partially covered by emerging regulations like, for instance, the EU General Data Protection Regulation (GDPR), but this is just one step in the right human-centric direction. AI is processing privacy-sensitive data, where proper directives and regulations need to come into existence. This includes that legal aspects must be integral to our computer science research.

R4: System architectures Figure 1 shows the complexity of building a complete multimodal, ML-based data analysis pipeline distributed over a large variety of computing nodes ranging from large machines ashore to smaller mobile devices on the boats at sea with limited connectivity and power. The system must integrate all the above components in an efficient way. Some data needs to be (pre)processed and filtered on the edge of the system, and more complex anomaly analyzes will be performed ashore. Developing such a system opens several research questions like what data to collect and at which frequency; placement of the processing; amount of edge processing; off-shore communication; etc.

R5: Continuous learning The described use-case comes with the specific challenge of time and location-based changes of key features. Based on the location of the boat and the season of the year the amount of fish, type of fish, and size will change. Models and datasets need to reflect this change. Ways of including this into the analysis need to be researched. Possible starting points are transfer learning, data distribution shift, and continuous learning algorithms.

R6: Data Currently most surveillance datasets are taken in cities. The chance that these datasets will provide a basis for analysis on fishing vessels is, put modestly, rather low. Specific datasets will be needed that is collected under the specific conditions on the boats. In addition, seasonal and day and night shifts should be taken into account. Regarding the fish analysis, current fish datasets are often specifically collected under perfect conditions and in areas with different types of fish. To be able to build a practically useful system one will need a fish dataset that is close to the real world conditions on the boat (fish dataset in the wild). We also conjecture the need for small data analytic datasets, one where small datasets can be useful for specific situations (relative to the typical big datasets found in, for instance, supervised learning scenarios).

The discussed research directions only present a small number of challenges and possible directions that arise, but includes what we identified as the most important ones to build a solid base for future improvements.

5 CONCLUSION

Fraud in the fishing industry is a major international problem, and it is difficult to determine when it takes place without infringing on the privacy of fishermen. Thus, we have proposed Dutkat: a distributed privacy-preserving surveillance and documentation system for detecting when crime potentially occurs. Realizing such a system requires multimodal analysis on a variety of data that is both captured on the fishing vessel, and that is federated on the mainland from other sources.

We conjecture that Dutkat is based on a more general architecture that is applicable in different use-case scenarios. The properties for such use-cases include some sort of continuous sensory input at a (mobile) edge node for real-time or post-analysis. This edge node can be a plane, an army vehicle, or a modern car, but also even a human being like a policeman equipped with multimedia life-logging devices. For example, in self-driving cars there are multiple safety features that require constant surveillance of both the physical environment, but also the driver and passengers. In order to ensure that the privacy of the user and others in the proximity of the car is maintained, the general architecture described will be further explored in such use-case scenarios.

ACKNOWLEDGMENTS

This work is partially funded by the Research Council of Norway project numbers 275516 and 263248, and Lab Nord-Norge ("Samfunnsloftet"). We particularly acknowledge contributions from Kim H. Andreassen, Arne E. Karlsen, Nandor Knust, Jon F. Mikalsen, Magnar Pedersen, Jon P. Rui, and Kjetil Robertsen.

REFERENCES

- [1] Mutasem K. Alsmadi and Ibrahim Almarashdeh. 2020. A survey on fish classification techniques. *Journal of King Saud University - Computer and Information Sciences* (2020). <https://doi.org/10.1016/j.jksuci.2020.07.005>
- [2] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multi-modal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 2 (2018), 423–443.
- [3] Christopher Costello, Ling Cao, Stefan Gelcich, Miguel Á Cisneros-Mata, Christopher M Free, Halley E Froehlich, Christopher D Golden, Gakushi Ishimura, Jason Maier, Ilan Macadam-Somer, et al. 2020. The future of food from the sea. *Nature* 588, 7836 (2020), 95–100.
- [4] Sandeep Dsouza, Victor Bahl, Lixiang Ao, and Landon P Cox. 2020. Amadeus: Scalable, Privacy-Preserving Live Video Analytics. *arXiv preprint arXiv:2011.05163* (2020).
- [5] Alem Fitwi, Yu Chen, Sencun Zhu, Erik Blasch, and Genshe Chen. 2021. Privacy-Preserving Surveillance as an Edge Service Based on Lightweight Video Protection Schemes Using Face De-Identification and Window Masking. *Electronics* 10, 3 (2021), 236.
- [6] Rafael Garcia, Ricard Prados, Josep Quintana, Alexander Tempelaar, Nuno Gracias, Shale Rosen, Håvard Vågstøl, and Kristoffer Løvall. 2020. Automatic segmentation of fish using deep learning with application to fish size measurement. *ICES Journal of Marine Science* 77, 4 (2020), 1354–1366.
- [7] Gunnar Hartvigsen and Dag Johansen. 1990. Co-operation in a distributed artificial intelligence environment—the stormcast application. *Engineering Applications of Artificial Intelligence* 3, 3 (1990), 229–237.
- [8] Dag Johansen. 1993. A distributed approach to the design of applications. In *Proceedings of ICCI'93: 5th International Conference on Computing and Information*. IEEE, 195–201.
- [9] Dag Johansen and Gunnar Hartvigsen. 1994. Convenient abstractions in Storm-Cast applications. In *Proceedings of the 6th workshop on ACM SIGOPS European workshop: Matching operating systems to application needs*. 11–16.
- [10] Dag Johansen, Kjetil Jacobsen, Nils P Sudmann, J Lauvset Kare, and Werner Vogels. 1997. Using Software Design Patterns to Build Distributed Environmental Monitoring Applications. (1997).
- [11] Dmitry A Konovalov, Alzayat Saleh, Dina B Efremova, Jose A Domingos, and Dean R Jerry. 2019. Automatic weight estimation of harvested fish from images. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 1–7.
- [12] Jan André Marais. 2019. *Deep learning for tabular data: an exploratory study*. Ph.D. Dissertation. Stellenbosch: Stellenbosch University.
- [13] Ministry of Trade, Industry and Fisheries. 2019. Framtidens Fiskerikontroll. *NOU 19:21* (2019).
- [14] United Nations Office on Drugs and Crime. 2016. Fisheries Crime. (2016).
- [15] Itsaso Rodríguez-Moreno, José María Martínez-Otseta, Basilio Sierra, Igor Rodríguez, and Ekaitz Jauregi. 2019. Video activity recognition: State-of-the-art. *Sensors* 19, 14 (2019), 3160.
- [16] Rajesh Kumar Tripathi, Anand Singh Jalal, and Subhash Chand Agrawal. 2018. Suspicious human activity recognition: a review. *Artificial Intelligence Review* 50, 2 (2018), 283–339.