Navigating an Ocean of Video Data: Deep Learning for Humpback Whale Classification in YouTube Videos

UNIVERSITY OF ILLINOIS CHICAGO
Honors College

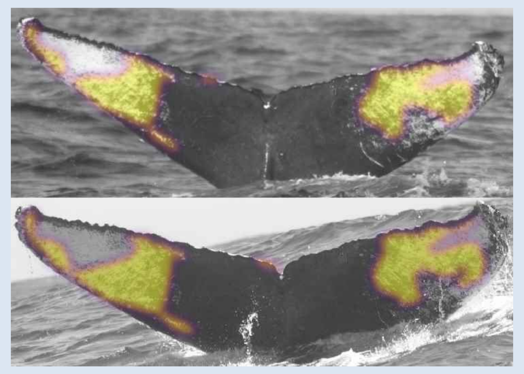
Michelle Ramirez, *University of Illinois at Chicago*

Abstract

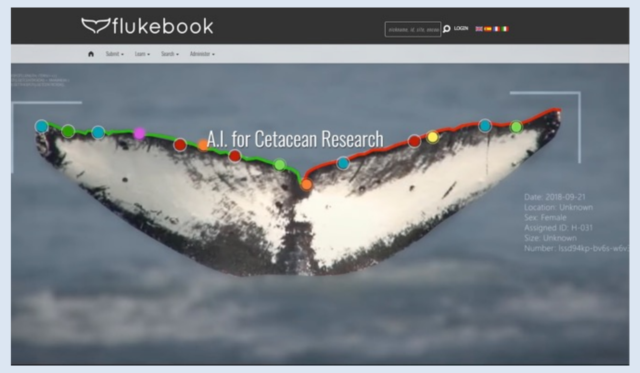
Image analysis technologies empowered by artificial intelligence (AI) have proved images and videos to be an opportune source of data to learn about Humpback Whale (*Megaptera novaeangliae*) population sizes and dynamics via photographed encounters, particularly as threats to Humpback Whales increase with the rise of sea temperatures and human encroachment in oceanic habitats. With the advent of social media, platforms such as YouTube present an abundance of video data across spatiotemporal contexts documenting Humpback Whale encounters from users worldwide. In our work, we focus on automating the classification of YouTube videos as relevant or irrelevant based on whether they document a true Humpback Whale encounter or not via deep learning. Our dataset consists of 203 relevant and 204 irrelevant videos retrieved via a text query made to the YouTube Data API v3 and manually annotated for relevancy. In our approach, we use a CNN-RNN architecture pretrained on the ImageNet dataset for classification of YouTube videos as relevant or irrelevant. Our CNN-RNN model was able to achieve an average 85% accuracy, 84.7% (irrelevant) and 86.6% (relevant) F1 scores, 91.3% (irrelevant) and 82.0% (relevant) precision, and 79.5% (irrelevant) and 92.1% (relevant) recall using five-fold cross validation for evaluation on our dataset. Our model performed well on videos where Humpback Whale encounters consisted of visible whale flukes, breaching, lunging, underwater encounters, and aerial views captured via drone footage. Misclassified videos arose due to issues of visibility, realistic ocean-themed video games, and other marine species in frame. Through our work, we have shown that deep learning can be applied to automate classification of YouTube videos as containing Humpback Whale encounters or not. We hope our work motivates conservation teams to look to social media platforms as a time-efficient source of wildlife monitoring data for Humpback Whales.

Background

- **Rising sea temperatures and human encroachment** on oceanic habitats pose threats to humpback whale populations [1].
- It becomes important to monitor how humpback whale populations respond to changes in food availability, breeding grounds, and seasonal patterns.
- Monitoring humpback whales via traditional tagging and tracking methods becomes difficult for conservation teams due to their **globe-spanning migration patterns** and travel to **remote parts of the ocean [2].**
- Social media platforms, such as **YouTube**, have a **global distribution of users** eager to upload images and videos on all kinds of topics, including humpback whale encounters [3].
- We can use **images** and **AI** to assess humpback whale **population sizes and dynamic**s by **identifying individuals photographed** based on unique markings and features [2].



Source: WildMe



Source: Flukebook

Data Collection

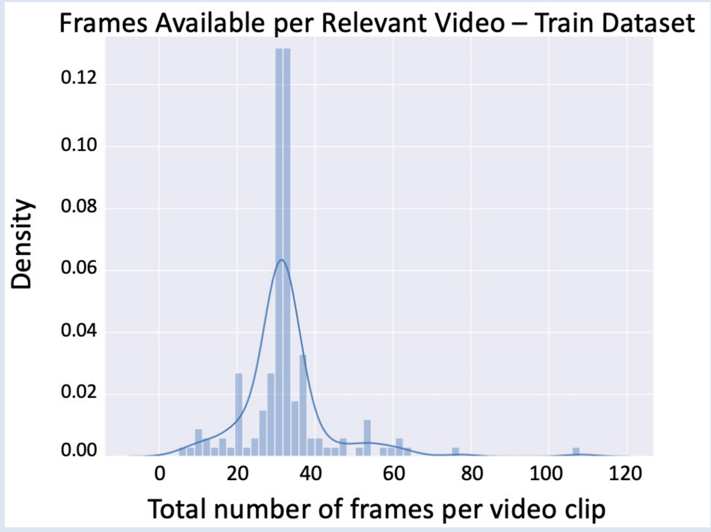
- Collected a total of **407 YouTube videos** (203 relevant, 204 irrelevant) via YouTube Data API (v3) using a text query of "Humpback Whale".
- Relevant humpback whale videos consisted of visible whale flukes, breaching, lunging, underwater encounters, and aerial views captured via drone footage.
- Videos were manually labeled as **relevant** (contains true humpback whale encounter) and **irrelevant** (does not contain true humpback whale encounter).

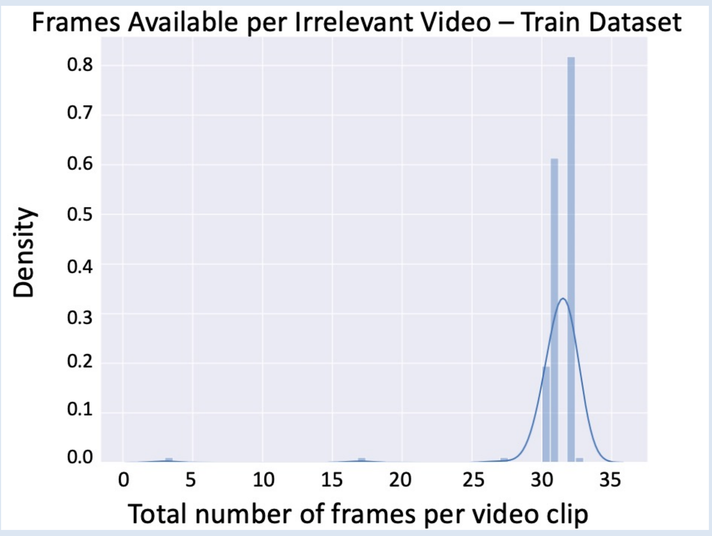
Data Preparation

- Manually identified and annotated **10-20 second "occurrence intervals"** in each **relevant video** where humpback whales were visible.
- For **irrelevant videos**, a random 10-20 second interval was chosen to represent the video.
- **31 image frames** were extracted from each video's intervals. Videos that had fewer than 31 frames were **padded** by replicating the middle frame while **keeping video frames in order**.

Data Preparation (cont.)

We choose 31 frames to extract from each video as the **mean number of frames** available per each (relevant and irrelevant) video reduced to its 10-20 second occurrence interval is **31 frames**.



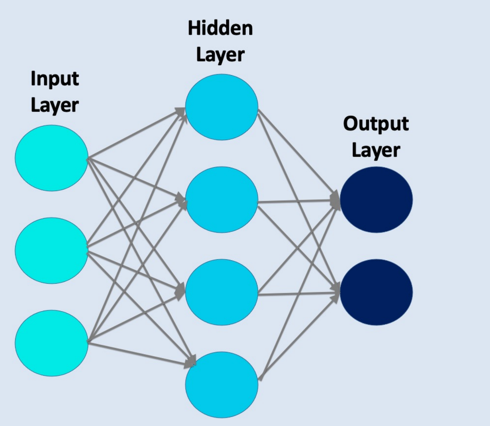


Deep Learning Classification Model

What is Deep Learning?

Deep Learning is a field within machine learning that uses **multiple layers** of **neural networks** for classification tasks.

Neural networks are connected groups of units (neurons) organized in a layer structure, with **different layers applying different transformations** to the data received.



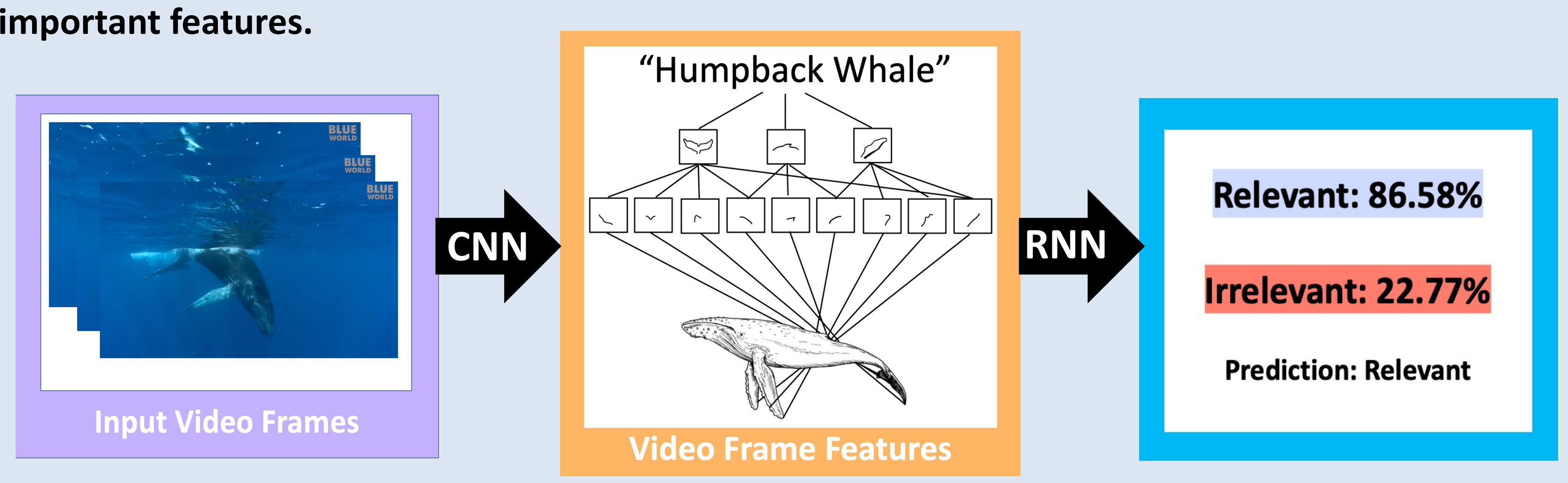
A neural network consists of multiple layers

What is a CNN?

Convolutional neural networks (CNN) are a type of deep neural networks consisting of an input layer, hidden layers, and an output layer. CNNs are especially useful to process **image data**, as the hidden layers perform convolutions that transform images by **emphasizing their important features.**

What is an RNN?

Recurrent neural networks (RNN) are a type of deep neural networks that process sequential data. RNNs cycle the information they process through a loop. When an RNN makes a decision, it considers the current input data and what it has learned from the data it received previously.



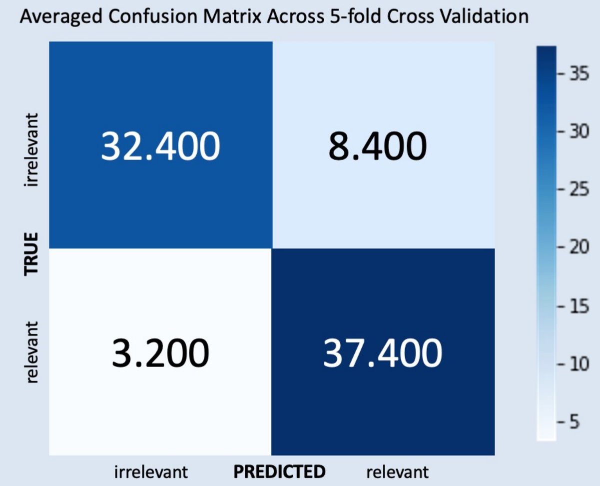
For our classification task, we use both CNNs and RNNs since **videos** are **sequential image data**. After inputting video frames, we use **InceptionV3 pretrained on the ImageNet** dataset as the convolutional base (**CNN**) to extract frame features. An **RNN** is then trained to evaluate frame features in sequential order and **classify the video** as **relevant** (humpback whales present) or **irrelevant** (no humpback whales).

Classification Results

To assess the performance of our model, we used **5-fold cross validation** to partition our videos into train and test sets.

- **Train set**: 162 relevant videos, 163 irrelevant for each fold.
- **Test set:** 41 relevant videos, 41 irrelevant videos per fold.

Our classifier was equally good at correctly classifying relevant and irrelevant videos, seen by the approximately equal number of correctly predicted relevant and irrelevant videos in our confusion matrix (right).



Classification Results (cont.)

Classification Model Performance Metrics

Fold #	Accuracy %	Precision %	Recall %	F1 Scores %
1	89.0	92.1 / 86.4	85.4 / 92.7	88.7 / 89.4
2	87.8	97.0 / 81.6	78.0 / 97.6	86.5 / 88.9
3	86.4	85.4 / 87.5	87.5 / 85.4	86.4 / 86.4
4	84.0	88.9 / 80.0	78.0 / 90.0	83.1 / 84.7
5	81.5	93.3 / 74.5	68.3 / 95.0	78.9 / 83.5
Average	**85.7**	**91.3 / 82.0**	**79.5 / 92.1**	**84.7 / 86.6**

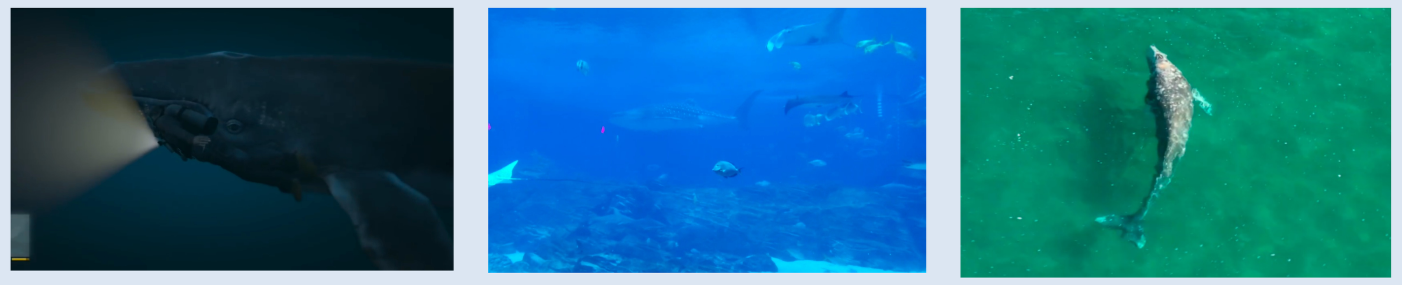
Precision, recall, F1: relevant % / irrelevant %

Videos Correctly Classified as Relevant

Relevant Videos Falsely Predicted "Irrelevant"

Videos Correctly Classified As Irrelevant

Irrelevant Videos Falsely Predicted "Relevant"



Our **model performed well** on videos where Humpback Whale encounters consisted of visible **whale flukes, breaching, lunging, underwater encounters, and aerial views** captured **via drone footage. Misclassified videos** arose due to issues of **visibility, realistic ocean-themed video games, and other closely-resembling marine species in frame.**

Conclusion

Our work has shown that deep learning can be applied to automate classification of YouTube videos as containing humpback whale encounters or not. In **future work**, we recommend **expanding the dataset** to include more videos **of closely-resembling marine species**, including but not limited to blue whales, pilot whales, sperm whales, and grey whales. Including additional videos of closely-resembling species will allow our classifier a greater chance to learn to distinguish relevant and irrelevant videos at the **species level**. We hope our work motivates conservation teams to look to social media platforms as a time-efficient source of wildlife monitoring data for Humpback Whales.

References

[1] Kershaw, J.L., Ramp, C.A., Sears, R., Plourde, S., Brosset, P., Miller, P.J.O. and Hall, A.J. (2021), Declining reproductive success in the Gulf of St. Lawrence's humpback whales (*Megaptera novaeangliae*) reflects ecosystem shifts on their feeding grounds. Glob. Change Biol., 27: 1027-1041. https://doi.org/10.1111/gcb.15466
[2] *Image analysis pipeline: Wild me docs*. Wild Me Docs Blog RSS. (n.d.). Retrieved March 22, 2022, from https://docs.wildme.org/docs/researchers/ia_pipeline
[3] Tuuli Toivonen, Vuokko Heikinheimo, Christoph Fink, Anna Hausmann, Tuomo Hiippala, Olle Järv, Henrikki Tenkanen, Enrico Di Minin, Social media data for conservation science: A methodological overview, Biological Conservation, Volume 233, 2019, Pages 298-315, ISSN 0006-3207, https://doi.org/10.1016/j.biocon.2019.01.023.