Right Whale Recognition

Right whales comprise three species of substantial baleen whales within the genus *Eubalaena*: the North Atlantic right whale (*E. glacialis*), the North Pacific right whale (*E. japonica*), and the Southern right whale (*E. australis)*. [1]

Right Whale Recognition was a computer vision competition hosted by NOAA Fisheries on the Kaggle.com data science platform [2]. This kaggle challenge is to recognize the "Right whales" in order to track and monitor their population as marine biologists routinely engage in manual identification of Right whales during population monitoring, however this process is slow and time consuming [2].

The **aim is to determine which whale is depicted in each image**, categorized by their unique whaleID in the kaggle dataset. Instead of identifying whales by species, the goal is to distinguish between individual whales.

DESIGN OF EXPERIMENT

What is the specific regression/classification/other task you intend to perform?
This is a multi-class classification task to correctly classify Right whales into their respective given classes. In this context, each class refers to an individual whale.

What is the uncertain quantity of interest (or decision)?
Transformers are to NLP what CNNs are to computer vision. I am still exploring vision transformer model vs traditional CNN model approach for this task to identify which performs better. The Vision Transformer (ViT) achieves remarkable performance, consistently matching or even outperforming existing state-of-the-art methods across various image recognition benchmarks [3], however, that is constrained to having trained over larger datasets. The dataset for this task is not very large, so traditional CNN based approaches might still be the best.

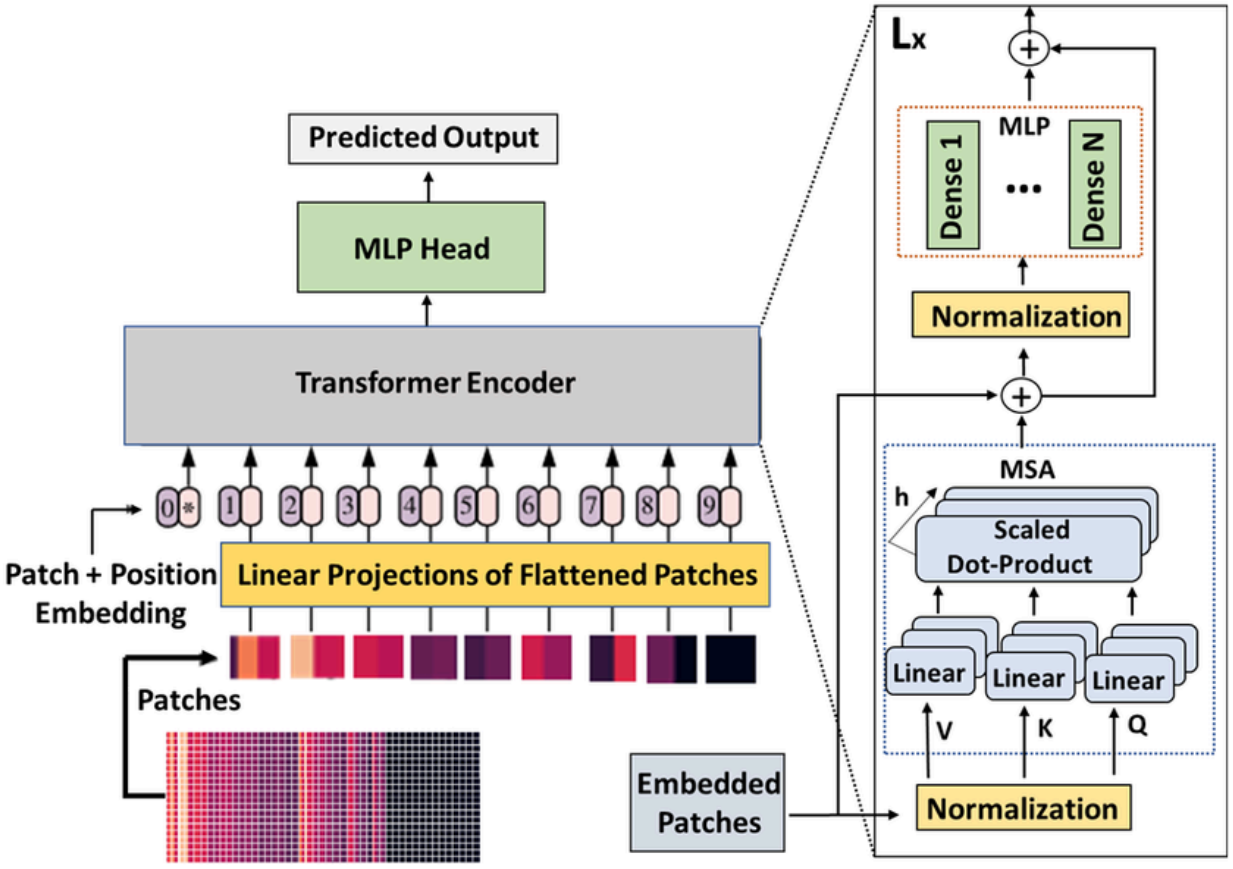
For decision problems, what is the utility function?
Using categorical cross entropy as loss function (i.e. evaluations performed using multi-class logarithmic loss). Loss will be minimized using one of the variations of gradient descent – most probably Adam, but also considering experimenting with other variations(batch, mini-batch) and types(RMSProp, SGD, Adagrad)

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij})$$

What proposed model architectures will you be using?

Apart from traditional CNN based approaches, I'm exploring the Vision Transformer model for classification of right whales. The raw data will be converted to black and white images and then slicked up into patches and fed to the vision transformer for classification. Below is the block diagram of the rough model architecture.

Vision Transformer (ViT) Architecture [4]



How will you evaluate the space of possible model architectures?
The top ranked team solution [8] used an ensemble of classification models trained on passport-like images of the whales' heads [6]. The second position ranked team [7] calculated a bounding box to localize the whale heads and then used classification models that were based on VGG and ResNet. Various neural network models can be used for this classification. CNN model with some variations is the most used approach for training this dataset. Vision transformer model approach for training is one approach I was unable to find with my literature survey. (Not surprising, since Transformers[9] really came into the picture in 2017 and this competition started in 2015 and closed in 2017). Hence, After evaluating existing model architectures for this problem, I'm exploring Vision Transformer models and CNN variations.

How will you measure your model's success?
The main attribute to determine a model's success will be accuracy i.e. how many Right whale images can the model correctly classify into its correct class(whaleID). Also considering exploring the F1 Score metric since dealing with image data.

If prior art exists, how do you expect your model to perform compared to those results, and why?
Typically the Vision Transformer model based approach should perform better than most previously used techniques, however considering the size of the data is relatively small, it may not perform better.

DATA

Identify the specific data sources you will need.
The dataset needed for this task is provided in the kaggle Right Whale Recognition competition [2].

Report the format of the data, how it is tagged, etc.
The training data comprises images and a csv file. The images are of right whales photographed during aerial surveys [2]. The training data in the csv file(`train.csv`) contains the image file names and a corresponding whaleID class.

Here's a short snippet of the training data –

	Image	whaleID
0	w_7812.jpg	whale_48813
1	w_4598.jpg	whale_09913
2	w_3828.jpg	whale_45062
3	w_8734.jpg	whale_74162
4	w_3251.jpg	whale_99558

w_7812.jpg

The total number of classes i.e. unique whaleIDs is 447.
The total number of training samples/images is 4544.
The total number of testing samples/images is 6925.

Connect these details of your data to your experiment. Argue that your data supports the question you are trying to answer/problem you are trying to solve.
This dataset [5] is specially collected to solve the problem of tracking and monitoring Right whales and presented as a task at the kaggle competition [2]. Hence, this dataset is most appropriate for this task(as it was collected with this sole purpose).

REFERENCES

[1] https://en.wikipedia.org/wiki/Right_whale
[2] Christin B. Khan, Shashank, Wendy Kan. (2015). Right Whale Recognition. Kaggle. https://kaggle.com/competitions/noaa-right-whale-recognition
[3] https://medium.com/@faheemrustamy/vision-transformers-vs-convolutional-neural-networks-5fe8f9e18efc
[4] Gufran, Danish & Tiku, Saideep & Pasricha, Sudeep. (2023). VITAL: Vision Transformer Neural Networks for Accurate Smartphone Heterogeneity Resilient Indoor Localization. 10.48550/arXiv.2302.09443.
[5] C. Khan, P. Duley, A. Henry, J. Gatzke2, and T. Cole1, “North atlantic right whale sighting survey (narwss) and right whale sighting advisory system (rwsas) 2013 results summary,” US Dept Commer, Northeast Fisheries Science Center Reference Document, pp. 14–11, 2014.
[6] A. Kabani and M. R. El-Sakka, "Improving Right Whale recognition by fine-tuning alignment and using wide localization network," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 2017, pp. 1-6, doi: 10.1109/CCECE.2017.7946736. Keywords: {Head;Whales;Training;Machine learning;Pattern recognition;Feature extraction;Merging;Whale Localization;Whale Detection;Whale Recognition;Deep Learning;Convolutional Neural Network;Localization;Detection;Recognition;Image Classification},
[7] F. Lau, “Recognizing and localizing endangered right whales with extremely deep neural networks,” http://felixlaumon.github.io/2015/01/08/kaggle-right-whale.html, [Online; accessed 04-August-2016].
[8] R. Bogucki, “Which whale is it, anyway? face recognition for right whales using deep learning,” http://deepsense.io/deep-learning-rightwhale-recognition-kaggle/, [Online; accessed 04-August-2016].
[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems. 6000–6010.