

Right Whale Recognition

WORK STATEMENT: Everything I submit here is my own work except the work for which I have provided a citation such as an “import” statement for the respective library or a link to the source documentation that I referred to.

What question did you answer with your investigation and/or what problem you solved?

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*). [14]

Right Whale Recognition was a computer vision competition hosted by NOAA Fisheries on the Kaggle.com data science platform [15]. 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 [15].

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.

What prior related work exists

A convolutional neural network achieves state-of-the-art performance on many image datasets such as the MNIST digit classification [1] and the ImageNet large scale classification challenge [2], [3]. This kaggle competition challenge [4] can be categorized as an image classification task and is similar to the face recognition problem; except that this challenge expects to develop a face recognition model for whales. A lot of the solutions for this competition have a similar strategy to the solution suggested in [5], which was inspired by the work of [6]. The team positioned second [7] employed a multi-stage approach, initially regressing a bounding box to pinpoint the whale's head, followed by an alignment process. Their classification models drew upon the Visual Geometry Group network (VGG) [8] and ResNet [9]. Conversely, the team achieving the highest score in this dataset [10] utilized an ensemble of classification models trained on passport-like images of whale heads. Other work includes [11], [12], [13].

How your approach differs from prior related work

Most solutions for this challenge solve it using traditional CNNs and/or incorporating some sort of ensemble learning techniques. I perform the training and inference using the Vision Transformer(ViT) Model. I am resizing the data to 256x256px, performing some preprocessing to get the appropriate input format for the vision transformer(the hugging face ViT implementation takes as input images in the PIL format.). And then I am providing this transformed data to the ViT for training. (Note: Earlier I was planning to convert images to black and white before feeding to the ViT, however upon further inspection, I concluded that to be unnecessary.)

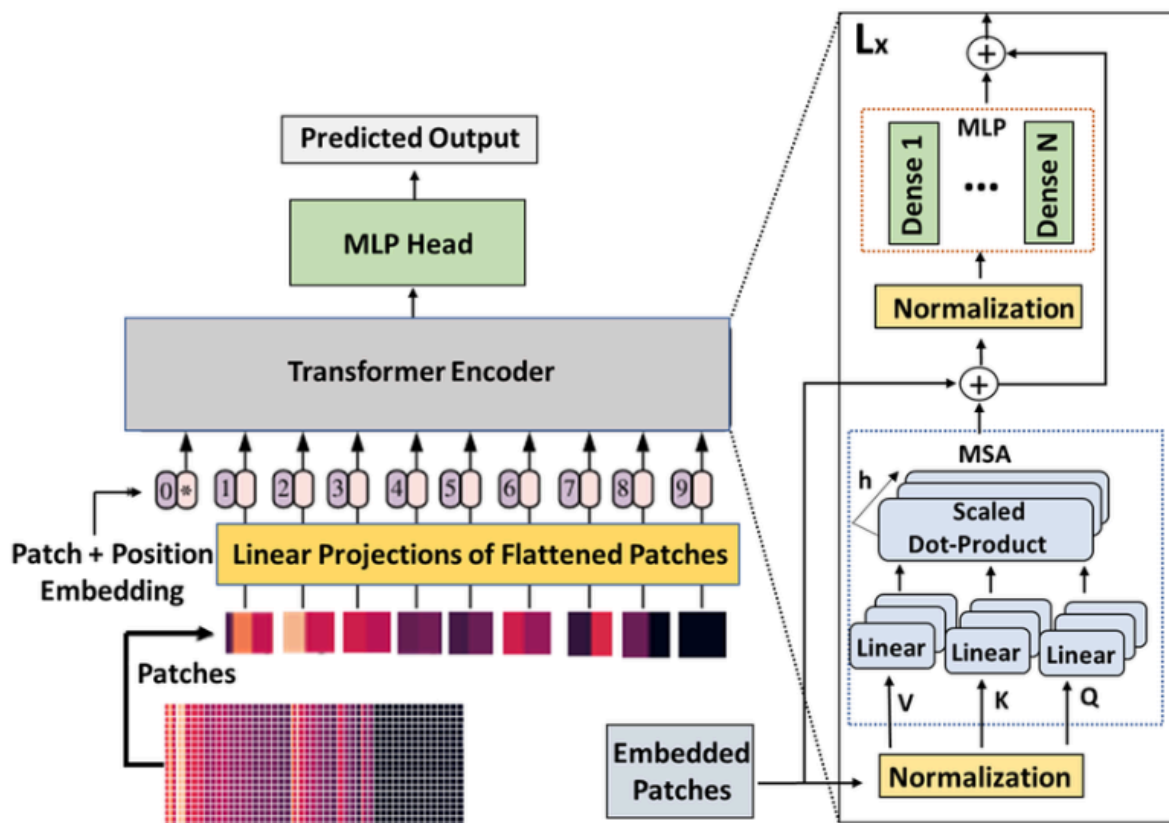


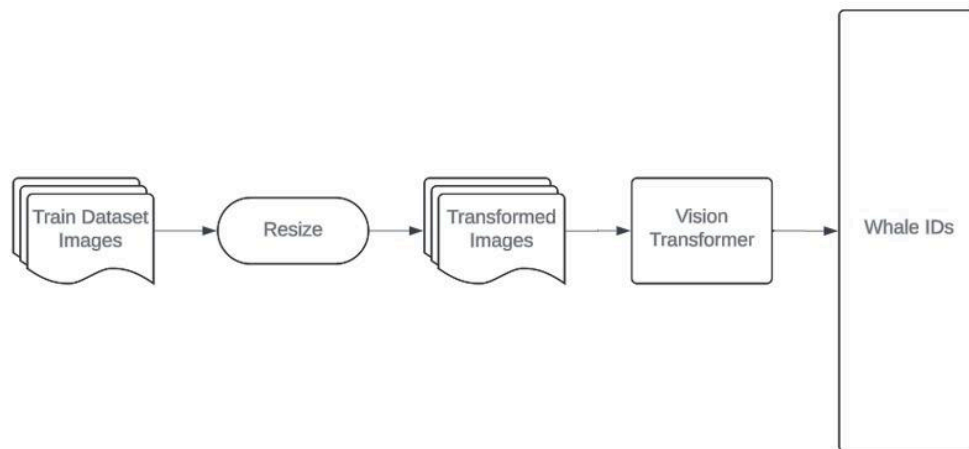
Fig. 1 Vision Transformer Architecture [16]

Your design of experiment

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. I have split the training data into 90% for training and 10% for validation. The training data comprises images and a csv file. The training data in the csv file(train.csv) contains the image file names and a corresponding

whaleID class. I have converted each of the class labels to an integer between 0-446 (since total 447 classes).

Vision Transformer Approach



I have performed data preparation(resizing and preparing according to the input format requirements for the vision transformer) in the following file:

https://github.com/insp7/rightWhaleRecognition/blob/master/prepare_data_for_vision_transformer.ipynb.

After that I provided this data to a pretrained ViT model and then trained it according for this use case for 21 epochs. I have trained multiple times to try and get the best validation accuracy and make sure my model doesn't overfit. I came to the conclusion that training for 21 epochs for this use case and configurations seems to work the best. The training took less than an hour on an RTX 3070 GPU. The training source file:

https://github.com/insp7/rightWhaleRecognition/blob/master/train_vit.ipynb

For testing, I performed inference on the test sample images using the image classification pipeline (from the official hugging face documentation). The inference took about 1 and half min on an RTX 3070 and about an hour on my laptop(without a discrete GPU).

What data you used and how you sourced it

The data was sourced from the Kaggle competition challenge “Right Whale Recognition” [4]

What neural network model(s) you trained



Initially I trained using a traditional CNN to just check my score. I got a public score of about 34 which put me around 268/365 rank on the leaderboard. After that I trained a vision transformer model, and got a score of around 6, which put me around 180~200/365 on the leaderboard.




NOTE: I have not received an official rank for this competition as my submission is considered late due to the competition being finished in 2017.

How I measured SUCCESS/RESULTS

Since this was a kaggle competition, I had no actual test set labels to verify exactly which predictions the model got right/wrong. Hence I had to upload a submission file for which I got a competition score. Here are the exact competition scores, (NOTE: Less is better)

Traditional CNN Approach




	naive-cnn-2.csv Complete (after deadline) · 7d ago			34.41408	34.42473
	y_pred.csv Complete (after deadline) · 7d ago · dimensions: (6925, 448) -- 6924 rows + 1 header row			34.37252	34.23943

267	▼ 3	moby		34.40023	6	8y
	Sample Submission Benchmark			34.41408		
268	▲ 2	Inicalo		34.41408	1	9y

Solution ranked around ~ 268/364 teams.

(Note: Since the competition is finished, I do not have a rank for my submission. Hence, attaching a screenshot for reference.

Better score

Submission and Description		Private Score ⓘ	Public Score ⓘ	Selected
<div> <div>All</div> <div>Successful</div> <div>Selected</div> <div>Errors</div> </div>				Recent ▾
	test.csv Complete (after deadline) · 1h ago	6.10255	6.10255	<input type="checkbox"/>
	naive-cnn-2.csv Complete (after deadline) · 7d ago	34.41408	34.42473	<input type="checkbox"/>
	y_pred.csv Complete (after deadline) · 7d ago · dimensions: (6925, 448) -- 6924 rows + 1 header row	34.37252	34.23943	<input type="checkbox"/>

New Solution Ranked ~184/364.

Future work (optional, if any exists)

There is a scope for training this ViT model on further preprocessed data. I had explored one of the popular approach that a lot of the top 10 solutions performed. That was to first extract the whale bonnet and blowhead coordinates and then perform an affine transformation so that all whale heads look in one direction. And then train on this transformed data so that the network code learns even more properly as now variations will be more prominent.

REFERENCES

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.
- [3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A.

Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[4] Kaggle, “Right whale recognition,” <https://www.kaggle.com/c/noaairight-whale-recognition>, [Online; accessed 19-January-2016].

[5] A. Thomas, “whale-2015,” <https://github.com/anlthms/whale-2015>, 2015, [Online; accessed 19-January-2016].

[6] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[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] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *arXiv preprint arXiv:1512.03385*, 2015.

[10] 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].

[11] 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}

[12] Bogucki R, Cygan M, Khan CB, Klimek M, Milczek JK, Mucha M. Applying deep learning to right whale photo identification. *Conserv Biol*. 2019 Jun;33(3):676-684. doi: 10.1111/cobi.13226. Epub 2018 Nov 28. PMID: 30259577; PMCID: PMC7380036.

[13] Right whale recognition using convolutional neural networks
<https://arxiv.org/abs/1604.05605>

[14] https://en.wikipedia.org/wiki/Right_whale

[15] Christin B. Khan, Shashank, Wendy Kan. (2015). Right Whale Recognition. Kaggle.
<https://kaggle.com/competitions/noaa-right-whale-recognition>

[16] Gufran, Danish & Tiku, Saideep & Pasricha, Sudeep. (2023). VITAL: Vision Transformer Neural Networks for Accurate Smartphone Heterogeneity Resilient Indoor Localization. 10.48550/arXiv.2302.09443.

APPENDIX

Source Code: <https://github.com/insp7/rightWhaleRecognition>