Classification of Deep-sea footage

Shashank Anand
Delft University of Technology
Delft, Netherlands
Email: S.Anand-1@student.tudelft.nl

Mihhail Sokolov
Delft University of Technology
Delft, Netherlands
Email: M.Sokolov@student.tudelft.nl

***Abstract*—The purpose of this study is to compare the rule based and neural network based solutions for classifying deep sea footage. Automation of classifying deep sea footage is absolutely essential as manual classification is error prone. In this paper, we describe in detail the methods used by us in evaluating the rule based system and the neural network. As a result of this analysis, we discovered that the rule based system did not perform well when it came to oceans of non-dark blue colour (such as light green). The neural network performed relatively well irrespective of parameters such as ocean colour. Finally, we conclude with describing why the neural network is superior to the rule based system for classifying deep sea footage.**

I. Introduction

Autonomous Underwater Vehicles (AUVs) are often sent on missions that last several hours. A large part of the ocean floor is uncharted and unexplored [1]. The primary focus of a significant amount of AUV missions is the procurement of photographs and video footage from the deep sea for research purposes. Video footage obtained over the course of an AUV mission is several hours long. The footage is later analyzed to determine ”interesting” segments which are noteworthy to researchers of the deep sea, biologists, and other scientists. This footage could also be used in studying ocean contamination/pollution [2]. Typically, analysis of the footage obtained is done by a team of experts who go through the footage manually. This is a very expensive and time consuming process. This manual filtering is also error prone as humans tend to have biases. Therefore, it is of great importance to automate this process to significantly reduce the cost, and to also improve the accuracy of the analysis.

Neural Network based solutions have already been in use in ocean related research for a long time. It is widely used in estimating thickness of the ice in oceans, which can be very useful when it comes to estimating the effects of global warming.[3] [4][5]. These systems are also of high importance in maritime security by means of ship detection from satellite images.[6]. Furthermore, they are also of interest to marine biologists in identifying flora, fauna, and their habitats[7]. The methods used in such research have made use of Regression Trees, Machine Learning methods, and statistical analysis. Convolutional Neural Network (CNN), and Deep Convolutional Neural Network based solutions have only been primarily used in the field of terrain mapping and land segmentation.[8].
Another noteworthy algorithm for large-scale classification is the streaming ensemble algorithm, which is although only suitable for real time classfication.[9]. In this paper, we describe our application of the CNN based strategy for the classification of deep-sea footage. Most existing research in the field is limited to classification of variations of neural networks and machine learning methods.[10][11][12]. In addition to a neural network based solution, an original rule based solution based on pattern recognition is described in this paper. This was done to provide an efficient comparison between the effectiveness of classic rule based pattern recognition and modern CNN based solutions.

A. *Characteristics of Deep-Sea Footage*

Deep sea footage is characterized by mostly static darkness (empty waters), with very sparse interesting objects (such as flora and fauna). The static part of the footage depends highly on the waters in which the footage is obtained from. A large portion of the footage is uninteresting because it is mostly empty. This is a major cause for diminished accuracy when the analysis is manually performed because it is easy for a human to miss certain interesting frames that are interspersed among many uninteresting frames.

B. *Defining What is Interesting*

In our research, we have considered interesting images to be those that contain flora or fauna or any other object of colours that contrast with that of the dark oceanic background as seen in Figure 1 and Figure 2. We have decided to use this criteria for interesting images based on manual inspection of the data set. A detailed manual inspection of the footage resulted in the observation that images could be classified ”interesting” for a variety of reasons (such as containing flora, fauna, inanimate objects that are out of place (such as plastic and other trash), etc) , and ”interesting” images could be important to a variety of research. This led us to conclude that any image that contained objects with colour contrasting that of the background would be interesting.

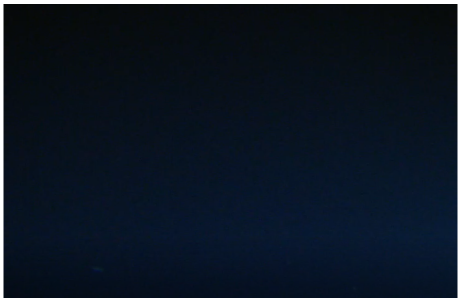

Fig. 1. An average uninteresting image

Fig. 2. An interesting image that could be easily missed by a human

II. Related Work

In this section we will shortly describe the related theory to the methods that we have applied.

A. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a kind of Artificial Neural Networks that are mostly used for image classification and object recognition. They use simple principles to make classification robust to image scaling and translation. CNNs usually use alternating convolution and max-pooling layers followed by a small number of fully connected layers. [13]

B. Image Processing

Image processing was carried out using python's automated image processing libraries. Parameters required for image processing were initially estimated using k-Nearest-Neighbours. This was later replaced in favour of a manual estimation of parameters (described in a later section).

C. Graphical Analysis

A protocol for classifying images as interesting or not for the rule based system was derived primarily via a graphical method. The x and y coordinates of pixels can be plotted on the xy plane, with the z axis corresponding to particular features of pixels. The resulting graph can reveal a lot of patterns in images, especially patterns related to localization of similar features.

D. Evaluation Metrics

In order to evaluate our algorithms we used 3 metrics: accuracy (1), precision (2) and recall (3). These evaluation methods are simple to understand and are widely used to estimate performance of machine learning algorithms.

1) Accuracy: Accuracy is the ratio of the number of correct predictions to the total number of predictions. It can be expressed as:

$$\frac{tp + tn}{tp + fp + tn + fn} \tag{1}$$

where tp stands for True Positives, tn for True Negatives, fp for False Positives and fn for False Negatives.

2) Precision: Precision is the ration of the number of correct positive predictions to the number of total positive predictions. It can be expressed as:

$$\frac{tp}{tp + fp} \tag{2}$$

where tp stands for True Positives and fp for False Positives.

3) Recall: Recall is the ratio of the number of correct positive predictions to the total number of correct predictions. It can be expressed as:

$$\frac{tp}{tp + tn} \tag{3}$$

where tp stands for True Positives and tn for True Negatives. [14]

III. Data Set

In this section we briefly describe the data sets we have used to design our algorithms.

This research used data provided by the Ocean Exploration Trust's Nautilus Exploration Program, Cruise NA095, and NA101. We have used footage from two different expeditions (NA095, and NA101) which took place in the Cascadia Margin, and the Papahānaumokuākea Marine National Monument respectively.

The data set contains a good mix of uninteresting deep sea footage, and interesting footage, which is useful as it helps us identify patterns and have a well-trained model.

This variety helps us to reduce bias in our model.

IV. Methods

In this section we describe in detail the two algorithms we have designed for classification of deep sea footage.

A. Rule-based System

This section will describe the methods involved in classifying images using a rule-based system as opposed to a neural network.

1) Graphical Analysis: The image was first graphically represented to observe any patterns. For this purpose, the z axis was chosen to be

$$C = 65536 * R + 256 * G + B \quad (4)$$

This equation effectively maps each RGB triplet to a unique value. This can now be used as a feature to graph along with the xy coordinates of the pixel occupying the xy plane. Pixel colours were chosen as the z-feature because high contrast regions can be detected very effectively and easily based on the peaks that can be observed from the graph of the image. But, this graph is still not sufficient as the non-peak regions are not stable. This graph can be further improved by plotting the deviation of C about the median C (mean C is not chosen as mean is influenced by outliers to a much higher degree than the median) as the z-coordinate. The resulting graphs are much smoother, and the peaks are more profound.This method has proven to be effective, as seen in the peaks observed in Figures 4, 6, and 8.

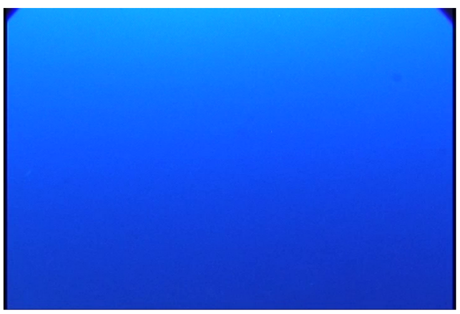

Fig. 3. Image 1

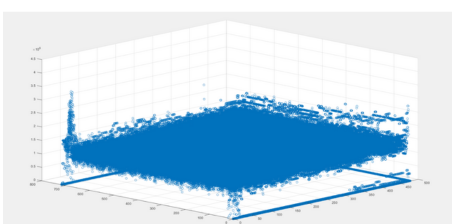

Fig. 4. Graph of Image 1

Fig. 5. Image 2

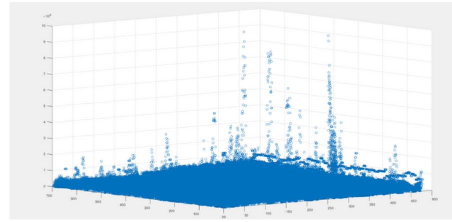

Fig. 6. Graph of Image 2

Fig. 7. Image 3

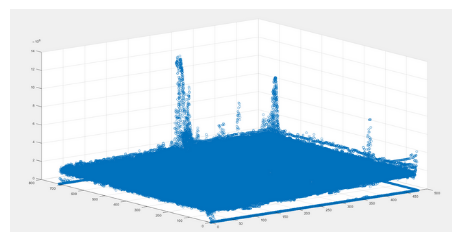

Fig. 8. Graph of Image 3

2) New Classification Protocol: From the graphical analysis described above, our protocol is now adapted for the rule based system and described in greater detail and is more specialised than the rather abstract one described initially. This is because it is necessary to have a solid classification protocol in place for evaluation by a rule based system. Any image that contains a peak is now classified as an interesting image. Images that contain a peak in their graphs signify the existence of high contrast regions. These are interesting images. Any image that does not contain peaks signifies the lack of any high contrast regions. These images are not interesting

3) Classification Algorithm: Now that a pattern has been graphically identified, an algorithm to decide whether an image is interesting or not. First, we must generate the deviations of C about median C for each pixel. Secondly, the existence of peaks should be tested. On discovering a peak, the algorithm terminates and return true. If the algorithm has completed evaluation and has not discovered a single peak, it terminates and returns false. Peaks can be identified by iterating through all possible MxN sub-graphs for all M and N less than x and y. These sub-graphs are called clusters. If the mean of all deviations of C about median C for a given cluster is greater than a fixed constant (called the minimum peak height), the cluster is labelled a peak. As stated earlier, on discovery of a peak, the algorithm terminates and returns true. As it can be observed, the running time of this algorithm is poor. The running time can be significantly improved using certain heuristics. Instead of searching through all M and N, M and N can be set to an initial value, and then grown additively or multiplicatively. Also, M and N can be set to have an upper limit as larger clusters also make themselves visible as smaller clusters.

B. Machine Learning System

This section will describe a method used to classify images using Machine Learning System, more precisely Convolutional Neural Network (CNN).

1) Convolutional Neural Network: CNNs are the most popular way of classifying images because they are good at extracting features from the images using convolutional layers and are robust to translations because of max-pooling layers. Initial method was creating a Convolutional Neural Network and training it on the dataset. Depicted below is the initial architecture of the network used for training (Figure 9).

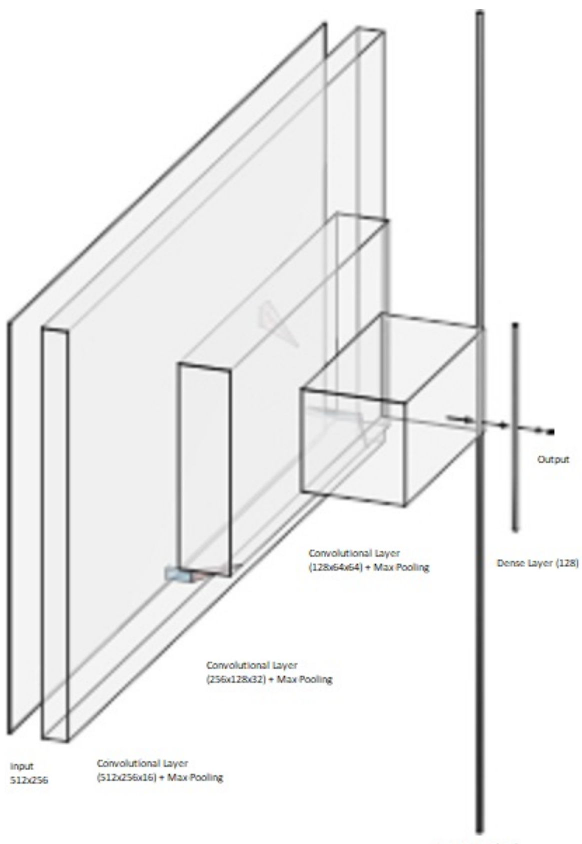

Fig. 9. CNN architecture

This kind of architecture is quite common for basic CNNs.

2) Training: The aforementioned Convolutional Neural Network has 67,133,473 trainable parameters. Therefore even after a long training period the accuracy of the network was very low - approximately 50%. The training graph with accuracy, loss, precision and recall over epochs is depicted below (Figure 10).

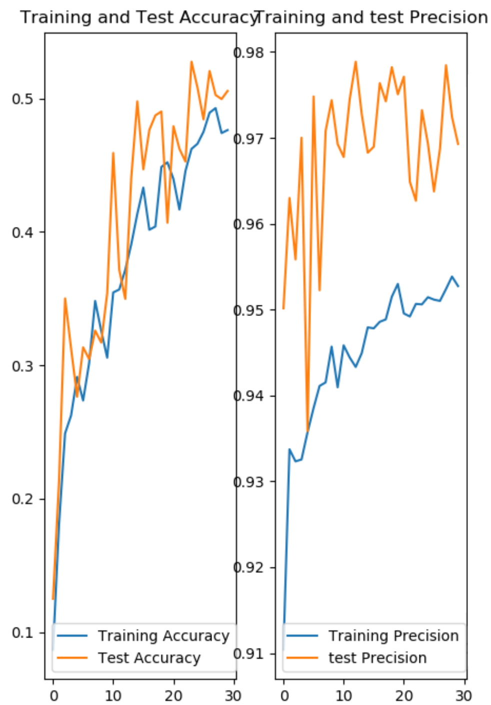

Fig. 10. CNN training graph

From the training graph can be seen that network has not yet converged. However, since it takes such a long to converge, this is not the best method.

3) Transfer Learning: In order to overcome the problem of having too many parameters for a relatively small dataset, we decided to use transfer learning - use pre-trained networks and re-train only last layer on our own data [15]. We have tried and tested several popular networks, all of which were pre-trained on the ImageNet dataset. Below are the networks, their accuracy and their training graphs all trained for 8 epochs:

- MobileNetV2 [16]: 45% accuracy, Figure 11
- DenseNet[17]: 92% accuracy, Figure 12
- InceptionV3[18]: 63% accuracy, Figure 13
- Xception[19]: 93% accuracy, Figure 14
- InceptionResNetV2[20]: 88% accuracy, Figure 15
- VGG19[21]: 93% accuracy, Figure 16

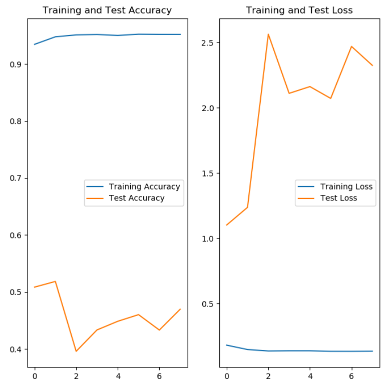

Fig. 11. MobileNetV2 training graph

Fig. 12. DenseNet training graph

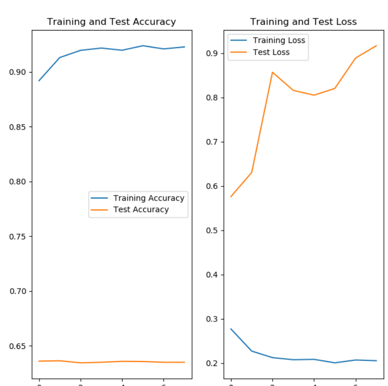

Fig. 13. InceptionV3 training graph

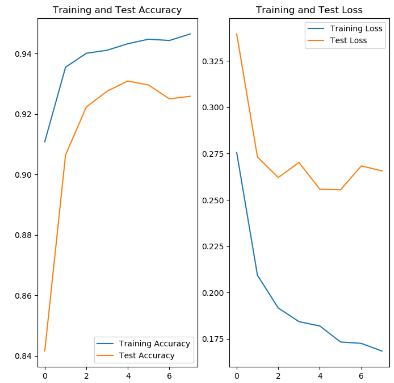

Fig. 14. Xception training graph

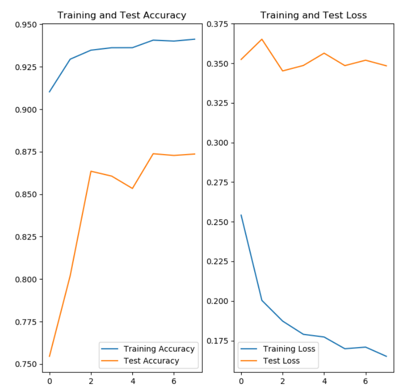

Fig. 15. InceptionResNetV2 training graph

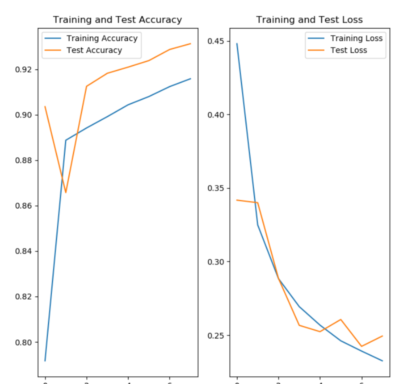

Fig. 16. VGG19 training graph

Out of these networks it appears that VGG19 trains and generalizes the best. Therefore, we decided to use VGG19 as a base for our Machine Learning System.

V. Results

The rule-based system falls short when it comes to oceans of different colours. As can be seen from the graphs, oceans that are not dark-blue do not show localised peaks in areas of high contrast. Therefore the rule based method is only viable for standard dark blue oceans.

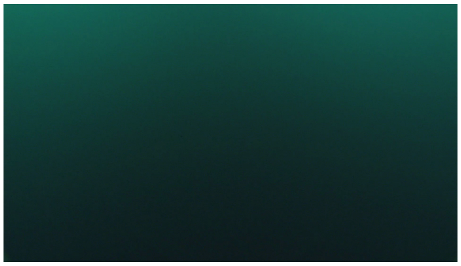

Fig. 17. Light Green ocean

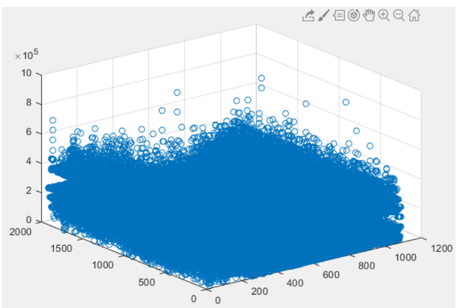

Fig. 18. Does not provide any identifiable patterns

The neural network on the other hand outperforms the rule-based system in all aspects. It has a higher accuracy, and also supports oceans that are not only dark-blue in colour. VGG19 network with re-trained last layer performs the best out of the networks tried. A more detailed training graph of VGG19 with accuracy, loss, precision and recall is given below (Figure 19).

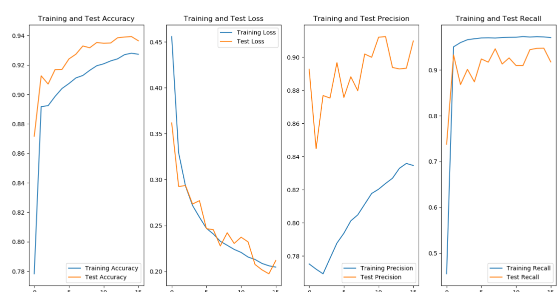

Fig. 19. Detailed VGG19 training graph

The accuracy is approximately 93-94% with quite a high precision and recall.

VI. Conclusion

It has been conclusively established that the neural network is superior to the rule based method in classifying deep sea footage. This is because deep sea footage contains a very high degree of unpredictability which makes it extremely difficult to analyze using a rule based system. For this purpose, a neural network works best as it learns from the supplied data set. We have also established that the neural network should be supplied with footage from not just one expedition, but should contain footage from many varied expeditions in oceans of various colours to maximize accuracy.

References

[1] J. Ontrup, N. Ehnert, M. Bergmann, and T. W. Nattkemper, "Biigle - web 2.0 enabled labelling and exploring of images from the arctic deep-sea observatory hausgarten," in *Oceans 2009 - Europe : Bremen, Germany, 11 - 14 May 2009 ; [International Oceans '09 Conference and Exhibition] / IEEE. Institute of Electrical and Electronics Engineers. - Piscataway, NJ : IEEE.*, vol. 1, 2009. [Online]. Available: http://dx.doi.org/10.1109/OCEANSE.2009.5278332

[2] H. Miyake, H. Shibata, and Y. Furushima, "Deep-sea litter study using deep-sea observation tools," *Interdisciplinary Studies on Environmental Chemistry - Marine Environmental Modeling and Analysis*, 01 2011.

[3] S. Lee, J. Im, J. Kim, M. Kim, M. Shin, H.-C. Kim, and L. Quackenbush, "Arctic sea ice thickness estimation from cryosat-2 satellite data using machine learning-based lead detection," *Remote Sensing*, vol. 8, no. 9, p. 698, 2016.

[4] M. Dawson, A. Fung, and M. Manry, "Sea ice classification using fast learning neural networks," *[Proceedings] IGARSS 92 International Geoscience and Remote Sensing Symposium*.

[5] I. Zaier, C. Shu, T. Ouarda, O. Seidou, and F. Chebana, "Estimation of ice thickness on lakes using artificial neural network ensembles," *Journal of Hydrology*, vol. 383, no. 3-4, p. 330–340, 2010.

[6] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, p. 1174–1185, 2015.

[7] S. Džeroski and D. Drumm, "Using regression trees to identify the habitat preference of the sea cucumber (holothuria leucospilota) on rarotonga, cook islands," *Ecological Modelling*, vol. 170, no. 2-3, p. 219–226, 2003.

[8] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, p. 3954–3962, 2018.

[9] W. N. Street and Y. Kim, "A streaming ensemble algorithm (sea) for large-scale classification," *Proceedings of*

the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 01, 2001.

[10] S. Smoliński and K. Radtke, "Spatial prediction of demersal fish diversity in the baltic sea: comparison of machine learning and regression-based techniques," *ICES Journal of Marine Science*, vol. 74, no. 1, p. 102–111, May 2016.

[11] M. A. Ghorbani, R. Khatibi, A. Aytek, O. Makarynskyy, and J. Shiri, "Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks," *Computers Geosciences*, vol. 36, no. 5, p. 620–627, 2010.

[12] B. A. Malmgren, M. Kucera, J. Nyberg, and C. Waelbroeck, "Comparison of statistical and artificial neural network techniques for estimating past sea surface temperatures from planktonic foraminifer census data," *Paleoceanography*, vol. 16, no. 5, p. 520–530, 2001.

[13] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," *arXiv e-prints*, p. arXiv:1412.6806, Dec. 2014.

[14] M. Hossin and S. M.N, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining Knowledge Management Process*, vol. 5, pp. 01–11, 03 2015.

[15] P. S. Dhillon, D. Foster, and L. Ungar, "Transfer Learning Using Feature Selection," *arXiv e-prints*, p. arXiv:0905.4022, May 2009.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *arXiv e-prints*, p. arXiv:1801.04381, Jan. 2018.

[17] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *arXiv e-prints*, p. arXiv:1608.06993, Aug. 2016.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv e-prints*, p. arXiv:1512.00567, Dec. 2015.

[19] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *arXiv e-prints*, p. arXiv:1610.02357, Oct. 2016.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv e-prints*, p. arXiv:1602.07261, Feb. 2016.

[21] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv e-prints*, p. arXiv:1409.1556, Sep. 2014.