

Aesthetic Assessment of the Great Barrier Reef using Deep Learnings

Bela Stantic and Ranju Mandal



Aesthetic Assessment of the Great Barrier Reef using Deep Learning

Bela Stantic¹ and Ranju Mandal¹

¹ School of Information and Communication Technology, Griffith University



Supported by the Australian Government's
National Environmental Science Program
Project 5.5 Measuring aesthetic and experience values using Big Data approaches

© Griffith University, 2020



Creative Commons Attribution

Aesthetic Assessment of the Great Barrier Reef using Deep Learning is licensed by Griffith University for use under a Creative Commons Attribution 4.0 Australia licence. For licence conditions see: <https://creativecommons.org/licenses/by/4.0/>

National Library of Australia Cataloguing-in-Publication entry:
978-1-925514-59-9

This report should be cited as:

Stantic, B. and Mandal, R. (2020) *Aesthetic Assessment of the Great Barrier Reef using Deep Learning*. Report to the National Environmental Science Program. Reef and Rainforest Research Centre Limited, Cairns (30pp.).

Published by the Reef and Rainforest Research Centre on behalf of the Australian Government's National Environmental Science Program (NESP) Tropical Water Quality (TWQ) Hub.

The Tropical Water Quality Hub is part of the Australian Government's National Environmental Science Program and is administered by the Reef and Rainforest Research Centre Limited (RRRC). The NESP TWQ Hub addresses water quality and coastal management in the World Heritage listed Great Barrier Reef, its catchments and other tropical waters, through the generation and transfer of world-class research and shared knowledge.

This publication is copyright. The Copyright Act 1968 permits fair dealing for study, research, information or educational purposes subject to inclusion of a sufficient acknowledgement of the source.

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of the Australian Government.

While reasonable effort has been made to ensure that the contents of this publication are factually correct, the Commonwealth does not accept responsibility for the accuracy or completeness of the contents, and shall not be liable for any loss or damage that may be occasioned directly or indirectly through the use of, or reliance on, the contents of this publication.

Cover photographs: (Front) Masking of photos for training of deep neural networks; (Back) Web portal. Images: Bela Stantic.

This report is available for download from the NESP Tropical Water Quality Hub website:
<http://www.nesptropical.edu.au>

CONTENTS

Contents.....	i
List of Tables.....	ii
List of Figures.....	ii
Acronyms	iv
Abbreviations	iv
Acknowledgements	v
Executive Summary	1
1.0 Introduction	3
1.1 Background.....	3
1.2 Approach	3
1.3 Related Work: Deep Learning	4
1.4 Objective.....	4
2.0 Datasets used in this research	6
2.1 Images from Flickr	6
2.2 AVA Dataset	9
3.0 Proposed Methodology.....	10
3.1 System Requirements.....	10
3.2 Deep learning Models (ResNet and Residual Attention Network).....	11
4.0 Results and Discussion	16
5.0 Proof of concept of a web-based application	22
5.1 Architecture for deployment	22
5.2 Dashboard Functionality	23
6.0 Conclusion	27
References.....	28

LIST OF TABLES

Table 1.	List of python pip packages installed in our server virtual environment as a requirement to run the experiments.	10
Table 2.	Residual Attention vs. Inception ResNet. Residual Attention Network has more parameters to learn 1.53 times more parameters to learn during model training.	14

LIST OF FIGURES

Figure 1.	Four samples of under-water images containing coral and fish from our dataset.	6
Figure 2.	Four samples of above-water images from our original underwater image dataset that were not included in the experimental dataset.	7
Figure 3.	Sample of masked images produced by using Tobii eye-tracking information. Three samples of under-water images containing coral and fish are shown whereby the first row shows the original images and the following row shows the masked images.	8
Figure 4.	Train, Validation and Test scored images of underwater environment were used during experiments.	9
Figure 5.	Residual learning: a building block.	12
Figure 6.	Architecture of the Residual Attention Network Architecture (proposed by He et al., 2016). The output layer is modified to produce an image aesthetic score instead of a class label.	13
Figure 7.	Visual attention can be applied to any kind of inputs, regardless of their shape. In the case of matrix-valued inputs, such as images, the model learns to attend to specific parts of the image. Top and bottom rows in the figure show the original and masked images, respectively.	15
Figure 8.	Training and validation loss graph from Inception ResNet: (a) Training loss on original dataset (i.e. non masked), (b) Validation loss on original dataset, (c) Training loss on masked dataset and (d) Validation loss on masked dataset.	16
Figure 9.	Training and validation loss graph from Residual Attention Network: (a) Training loss on the original dataset (i.e., non-masked), (b) Validation loss on original dataset, (c) Training loss on masked dataset, and (d) Validation loss on the masked dataset.	17
Figure 10.	Predicted aesthetic score of a few sample original images from the GBR dataset using our proposed Inception-ResNet-based aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.	18
Figure 11.	Some predicted aesthetic scores with a significant difference with ground truth from the GBR dataset using our proposed aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.	19
Figure 12.	Predicted aesthetic score of a few sample original images from the GBR masked dataset using our proposed aesthetic assessment model. Predicted	

	Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.	20
Figure 13.	Some predicted aesthetic scores with a significant difference with ground truth from the GBR dataset using our proposed aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image	21
Figure 14.	Proof of concept of a web-based application	22
Figure 15.	Architecture for web deployment.....	23
Figure 16.	Target identification and sentiment, calculated from pasted text.	23
Figure 17.	Examples of semantic network analysis outputs related to GBR Twitter communication as an example of text.	24
Figure 18.	Heatmap that shows the frequency of posts by location.	24
Figure 19.	Examples of additional outputs that can be visualised when mining social media posts of interest to the GBR.....	25
Figure 20.	Prototype option for calculating the beauty score of an image.	26

ACRONYMS

CNN	Convolutional Neural Network
DL	Deep Learning
EMD	Earth mover's distance
GBR	Great Barrier Reef
GBRMPA	Great Barrier Reef Marine Park Authority
GPL	Graphics Processing Unit
NESP	National Environmental Science Program
RIMRep	Reef 2050 Integrated Monitoring and Reporting Program
ROI	Region of Interest
TWQ	Tropical Water Quality

ABBREVIATIONS

abbrev.	abbreviated
A&P	analysis and prediction
assoc.	association
temp.	temporary
abbrev.	abbreviated
A&P	analysis and prediction
assoc.	association
temp.	temporary
abbrev.	abbreviated
A&P	analysis and prediction
assoc.	association
temp.	temporary
abbrev.	abbreviated
A&P	analysis and prediction
assoc.	association
temp.	temporary

ACKNOWLEDGEMENTS

This Technical Background Report details the computer science behind a proposed automated image rating platform. It serves as guidance for future research in the area of using artificial intelligence for environmental monitoring.

The report forms part of a research project on aesthetic value ([Project 5.5](#)) funded by the Australian Government's National Environmental Science Program (NESP) Tropical Water Quality Hub.

EXECUTIVE SUMMARY

A significant investment has been made by the Australian Government for improving reef health, and monitoring changes of the Great Barrier Reef (GBR). The Great Barrier Reef Marine Park Authority (GBRMPA) is responsible for implementing the Reef 2050 Integrated Monitoring and Reporting Program (RIMRep). The Reef is an irreplaceable ecosystem facing a growing combination of threats. Ecological, cultural and aesthetic values of the Great Barrier Reef are under stress and effective ways for monitoring change are required. Monitoring environmental change, however, is costly and the notion of citizen science or collective sensing has attracted increasing attention.

A key objective of the [NESP TWQ Hub Project 5.5](#) was to better understand perceptions of natural beauty as an integral part of heritage value. This project focused on the underwater aesthetic value of the GBR, because of rapid changes in the marine environment. Despite some earlier investment into this research area, there is still a need to improve the tools we use to measure the aesthetic beauty of marine landscapes.

This research drew on images publicly available on the Internet (in particular through the photo sharing site Flickr) to build a large dataset of GBR images for the assessment of aesthetic value. Building on earlier work in [NESP TWQ Hub Project 3.2.3](#), we conducted a survey focused on collecting beauty scores of a large number of GBR images. We have incorporated eye-tracking information into our machine learning model as studying eye movements provides unique insights into what catches our attention and what information we process. Eye-tracking is a sensor technology that makes it possible for a computer to know where a person is looking. An eye tracker can detect the presence, attention, and focus of the user. We captured the attention and focus of the participants during the survey for image ranking.

Thus, the eye-tracking technology and an online survey were used to identify the Region of Interest (ROI) and scores provided by the survey participants. Second, using state-of-the-art deep learning-based architectures, we developed our aesthetic assessment architectures and trained and tested the model. We analyzed the beauty score automatically using a trained model and computed the overall validation accuracy. Here, we investigated how the masked images with attention map performed on image aesthetic assessment. The experimental outcome shows that eye tracking technology-based attention information is useful to improve the performance of our aesthetic assessment model.

In this work, we applied a new loss function EMD (earth mover's distance) to train the aesthetic quality assessment model of images. We show that models with the same Convolutional Neural Network architecture, trained on masked images produce better performances. The proposed model is designed to predict a higher correlation with human aesthetic ratings (obtained from the survey). The correlation-based method predicted the distribution of ratings as a histogram. It performed better compared to classifying images to binary class label (low/high score) or regressing to the mean score. We employed the squared EMD loss that was reported as a performance booster in previously proposed approaches.

As part of our future work, we will exploit the trained models' architecture on combining both images (i.e., masked and non-masked) as input and feed them in parallel to enhance the

prediction score. Our current experimental setup accepts a single image as an input. The computational time complexity can also be improved by optimizing network parameters and weights.

1.0 INTRODUCTION

1.1 Background

The Great Barrier Reef (GBR) is a UNESCO World Heritage site inscribed for its outstanding heritage value. This includes its significant aesthetic characteristics (Criterion vii). Aesthetic value – and the need to monitor it – is explicitly recognised in the Great Barrier Reef 2050 Long-Term Sustainability Plan (e.g. discussed in Becken et al., 2019). The aesthetic beauty of the Reef is important not only to Australians but also to visitors who travel from far to admire the unique landscape and biodiversity. The economic value of the GBR is also substantial, in particular due to its tourism activity (Deloitte Australia, 2017).

There has been some recent interest in researching the aesthetic value of the GBR. Previously, researchers sought to define aesthetic value (Context, 2013), understand its relationship to ecological conditions (Marshall et al., 2019), and identify key environmental attributes to determine aesthetic perceptions (Le et al., 2019). The measurement and monitoring of aesthetic value has also received attention (Marshall et al., 2017; Curnock et al., 2020), so has potential differences across different cultural groups (Le et al., 2020). This report provides additional insights into the use of artificial intelligence for automated rating of the beauty of underwater Reef images. It connects to earlier work but adds a unique dimension of using new technologies for innovative monitoring.

1.2 Approach

This research capitalises on technological advances in information technology, as well as increasing volumes of online data shared by people who visit natural environments (Becken et al., 2017; Becken et al., 2019; Seresinhe, Moat & Preis, 2017). Visitors to the GBR share substantial amounts of imagery (photos and videos) via channels such as Instagram, Twitter, Flickr, Weibo and Youtube. Photographic imagery/videos of users' experiences at the GBR contain information on the environmental and experiential attributes of aesthetic value (Stantic et al., 2019). The images give important clues about what "matters" to Reef users, because people usually take photos (and share them) of situations or sceneries that made an impression of them (in whatever subjective way). The approach of using user-supplied imagery to somehow measure and monitor the aesthetic value of the Reef is a human-centric one that assumes that aesthetic value arises from interactions between nature and people. In other words, areas of the Reef that are not visited by humans will not be measured.

Aesthetic value or landscape beauty are challenging topics to research, and research to date has largely focused on the scenic value of terrestrial ecosystems or urban parks. Quantifying beauty is particularly difficult, especially since it assumes some kind of objective way of capturing it. More recent advances in data processing technology have accelerated progress (e.g. Haas et al., 2015), as has research into the cross-cultural differences of aesthetic value (Le et al., 2020). Using crowdsourced data from the photo-site Flickr, OpenStreetMap and 'Scenic-Or-Not', Seresinhe et al. (2017) found that computer-based models can generate accurate estimates of perceived scenic value of a wide range of natural environments. Interesting insights into how new computer technology can assist in the task of measuring

beauty, can be gained from Haas et al.'s (2015) who presented a computational evaluation of the natural beauty of coral reefs.

We propose our image aesthetic assessment model with a recently proposed loss function EMD (earth mover's distance) (Esfandarani et al., 2017) along with a state-of-the-art classifier to train the aesthetic assessment model. In comparison to our previous approach, we include attention-based masked images and original images along with scores to train our proposed network model and obtained better achieved improved accuracy.

1.3 Related Work: Deep Learning

Image quality assessment and prediction of photo aesthetic values have been a challenging problem in image processing and computer vision. Partly this is because aesthetic assessment has a subjective component (i.e. influenced by individual's feelings, tastes, or opinions). Existing photo aesthetics assessment methods are available (Dhar et al., 2011; Jiang et al., 2010; Lu et al., 2014; Nishiyama et al., 2011; Niu & Liu, 2012; Su et al., 2011; Tang et al., 2013), whereby most are using extraction of visual features and then employing various machine learning algorithms to predict aesthetic values of photos. Aesthetic assessment techniques aim to quantify semantic level characteristics associated with emotions and beauty in images, whereas technical quality assessment deals with measuring low-level degradations such as noise, blur, compression, artefacts, and other elements.

Earlier proposed techniques designed handcrafted aesthetics features according to aesthetics perception of people and photography rules (Bhattacharya et al., 2010; Datta et al., 2006; Dhar et al., 2011) and obtained encouraging results, despite handcrafted feature design for aesthetic assessment being a very challenging task. More robust feature extraction techniques were proposed by other researchers who leveraged more generic image features (e.g. Fisher Vector (Marchesotti et al., 2011; Perronnin et al., 2007; Perronnin et al. 2010) and the bag of visual words (Su et al., 2011). Generic feature-based representation of images is not ideal for image aesthetic assessment as those features are designed to represent natural images in general for object identification or localization, and not specifically the perceived aesthetics of them.

In contrast, and as an alternative approach, Convolutional Neural Network (CNN)-based features are learned from the training samples, and they do this by using dimensionality reduction and convolutional filters. Recent approaches to image aesthetic assessment mostly apply more complex and robust deep Convolutional Neural Networks (CNN) architectures (Kang et al., 2014; Bosse et al., 2016; Bianco et al., 2016). Availability of large-scaled labelled and scored images from online repositories has enabled CNN-based methods to perform better than previously proposed non-CNN approaches (Murray et al. 2012; Ponomarenko et al., 2013). Moreover, having access to pre-trained models (e.g. ImageNet (Russakovsky et al., 2015)) for network training initialization and fine-tuning of the network on subject data of image aesthetic assessment have been proven more effective technique for typical deep CNN approach.

1.4 Objective

The objective of this project is to develop a deep learning-based algorithm for automated aesthetic assessment of underwater images from the Great Barrier Reef region. Besides, the

experiments are further extended using a dataset with masked attention images that are created using participants' eye-tracking information, and a comparative study shows that the system trained and test with the masked images achieved better accuracy. In this technical report, we describe our deep learning architectures, experimental results and key findings.

2.0 DATASETS USED IN THIS RESEARCH

2.1 Images from Flickr

Flickr is an image hosting service and one of the main sources of images for our project. As per our requirement, we downloaded all images and their metadata (including coordinates where available) based on keyword filter such as “Great Barrier Reef”. The Flickr API is available for non-commercial (but commercial use is possible by prior arrangement) use by outside developers. To ensure a much larger and diverse supply of photographs, we have developed a python-based application using Flickr API that allowed us to download Flickr images by keyword (e.g. “Great Barrier Reef” available at <https://www.flickr.com>).

The Flickr dataset documentation includes API methods that we have used along with all the arguments and the values set for searching and downloading the Flickr images. For example, we extracted: user_id, text, minimum upload date, maximum upload date, privacy filter, machine tags, and latitude and longitude (Murray et al., 2012). However, Flickr API methods provide a wide range of images and not all of them are suitable for this research. For example, a search for GBR images yields photographs of underwater and above water landscapes, tourist experiences, and other activities. The focus of this research was on under-water images, which had to be filtered from the downloaded Flickr photos.

From the collected images we identified a total of 5407 relevant images with coral and fish contents out of a total of approximately 55,000 downloaded images. Sample images, as shown in Figure 1, that only contained coral and fish scenes were considered in our experiments, whereas above water images (see Figure 2) as well as images of artificial underwater environments were not included in our dataset.



Figure 1. Four samples of under-water images containing coral and fish from our dataset.

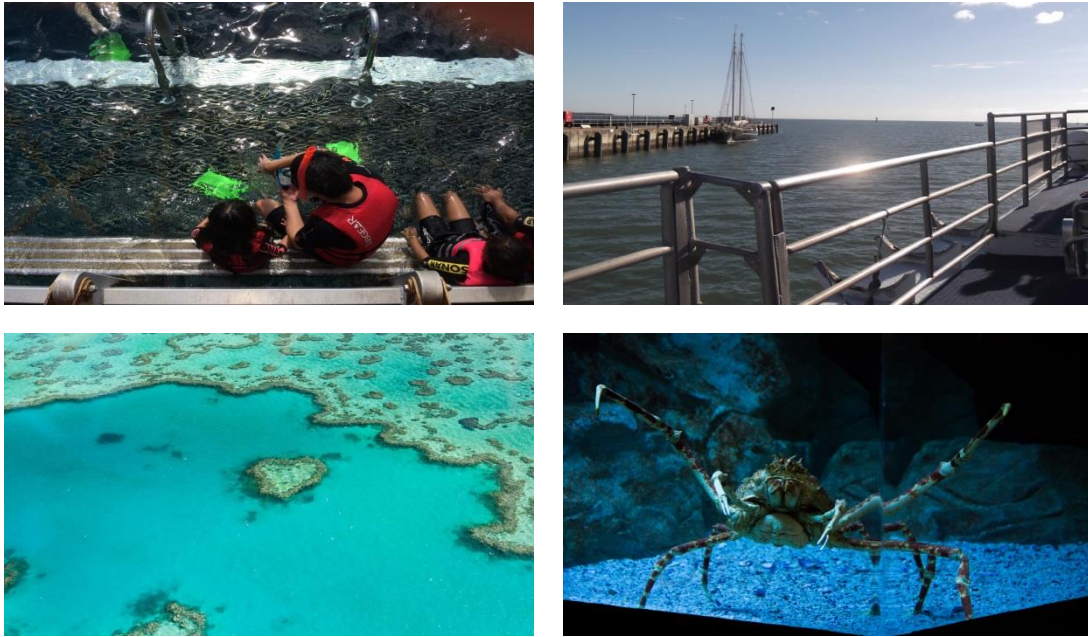


Figure 2. Four samples of above-water images from our original underwater image dataset that were not included in the experimental dataset.

Eye tracking data: The eye tracking information we obtained from the Tobii eye-tracker (see report by Le et al., 2020) provides a record of the direction of respondents' eye movements on an image (on the computer screen) with a rate of 60 times per second. The software 'maps' these locations on the image being viewed. The eye-tracking equipment used in this study is Tobii Pro TX300, available at the laboratory of Griffith Institute for Tourism. Details of the device can be found at the following link: "<https://www.tobii.com/product-listing/tobii-pro-tx300/>".

The images used for eye-tracking experiment were all resized to fit the eye-tracking screen with resolution of 1200x800 pixels. We considered a total of 5407 images for the eye tracking experiment. In this current study we used 3500 images whilst an additional 2500 images had already been used during an earlier NESP TWQ Hub funded research (Project 3.2.3). The eye-tracking technology allowed us to identify 'regions of interest', where research participants showed particular interest (eye focus through fixation time). This in turn allowed us to mask images according to important and secondary content. Three examples of masked images are presented in Figure 3.



Figure 3. Sample of masked images produced by using Tobii eye-tracking information. Three samples of under-water images containing coral and fish are shown whereby the first row shows the original images and the following row shows the masked images.

For experimental purposes, a dataset of Flickr Great Barrier Reef images was developed in-house. Our dataset contains two set of images containing a total of 10,814 ($5407 * 2$) scored images after final pre-processing stage. These images were sorted based on the content and then rated by participants in an online survey for their aesthetic beauty. The scale used was a ten-point Likert scale ranging from ugly (1) to beautiful (10) (Curnock et al., 2020). At least 10 survey participants provided an aesthetic score for each image and the mean score was calculated. Each image has two versions (i.e. original version and a masked version produced using Tobii Pro TX300 technology to capture respondent's attention). In total, 5407 underwater images were scored, and these images were also used to apply the mask based on eye-tracking information (see above). Most of the images (i.e. 90%) served as training material to enable the proposed Neural Network architecture to learn key feature parameters. The remaining 10% of images from the dataset were used as a validation dataset. The validation dataset helps to understand the system performance in terms of accuracy during the training phase. These remaining 10% of the full data were used during the test phase. We divided our data into training 4378 (80%) images, 487 (10% of data from training set) for validation, and 542 (~10%) test images for the experiments. Figure 3 shows the data division ratio we followed during experimentation. Please note same numbers of masked images are also available for experimentation and comparative study.

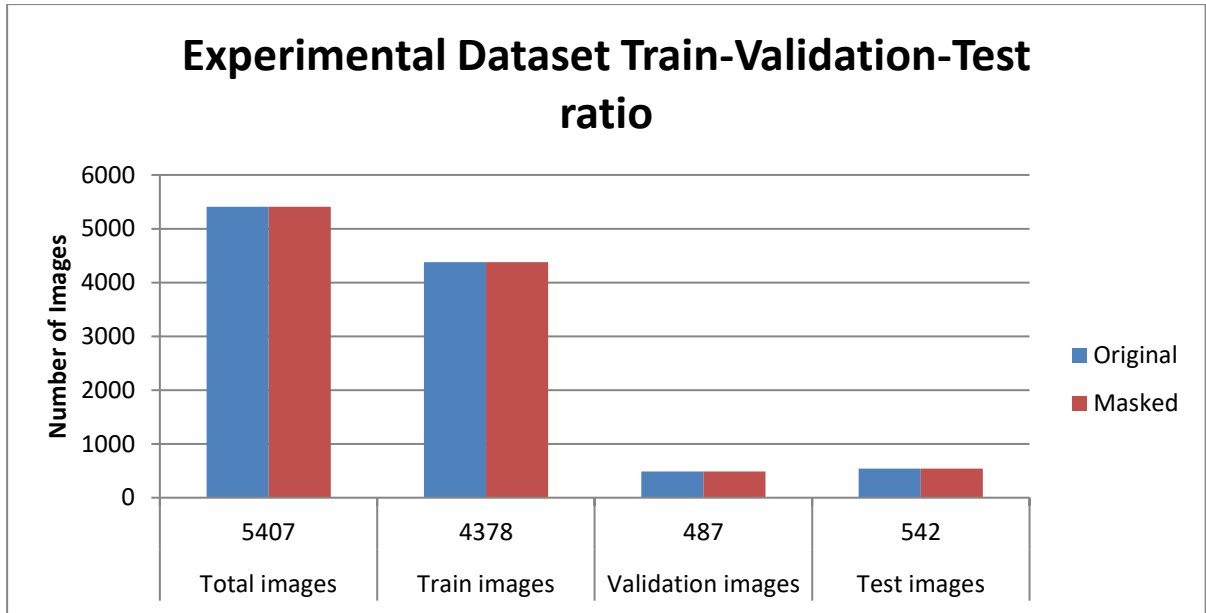


Figure 4. Train, Validation and Test scored images of underwater environment were used during experiments.

2.2 AVA Dataset

The AVA aesthetic dataset is a publicly available benchmark dataset. This dataset contains approximately 255,000 images with their corresponding aesthetic score values. The dataset has been adopted in a wide range of research projects on aesthetic assessment. It is important to note that the GBR dataset we have created has 5407 original images (see above), which is comparatively small for training a multi-layered deep Convolutional Neural Network. It was therefore necessary to complement the GBR data with a large-size, publicly available dataset and the AVA) seemed to be the most suitable one. This helped to train the system and allowed us to use the in-house GBR dataset from Flickr for fine-tuning the algorithm. The detailed dataset description of the AVA can be found in Murry et al. (2012). We adopted the score of 0.5 (mean aesthetic score ranges between 0 and 1) as the threshold value to divide the dataset into high aesthetic value and low aesthetic classes. By using these assumptions, we obtained 74,056 images in the low aesthetic class, and 181,447 images in the high aesthetic class. In total, 229,954 images we used for training our network architecture and 25,549 (approximately 10% images) were used for testing system performance.

3.0 PROPOSED METHODOLOGY

3.1 System Requirements

Data processing and all the experiments of image aesthetic assessment were conducted on our Big data server at Griffith University. Graphics Processing Unit (GPU) is considered as the heart of Deep Learning (DL) because DL (a variant of machine learning) requires extensive mathematical computational (n-dimensional matrix multiplication) resources. GPU is a single-chip processor used for extensive graphical and mathematical computations which are necessary for large deep learning architectures used in our experiments.

The training of our model using publicly available datasets (AVA dataset, see above) used the GPU-accelerated Python library for deep learning such as TensorFlow and Keras. Besides, designing and experimenting with large deep learning models such as AutoML and large size datasets requires multi-GPU systems like Big data servers. High GPU memory usage happens when we partition training on multiple GPUs to enable data parallelism (i.e. having each of our GPUs process a different subset of your data independently).

A total of five powerful GPUs were installed in our server which provides GPU-enabled computational resources in DL-based training and testing stages. The servers' configurations and GPU configurations are as follows: Intel(R) Xeon(R) CPU E5-2609 v3 @ 1.90GHz, AMD Ryzen Threadripper 1900X 8-Core Processor, 8234 MB Memory, a total of 5 GPUs (4 GEFORCE GTX 1080 Ti, 12GB memory each along with NVIDIA CUDA Cores 3584 and one GeForce GTX 1080 GPU with 2560 NVIDIA CUDA Cores).

We have created a python virtual environment known as pip environment in our server for experimentation related to image aesthetic assessment. Pipenv is a dependency manager for Python projects which is similar to Node.js' npm or Ruby's bundler tools. While pip can install Python a package, Pipenv is recommended as it provides a higher-level tool that simplifies dependency management for common use cases. A list of packages installed in the virtual environment is presented in the following Table 1.

Table 1. List of python pip packages installed in our server virtual environment as a requirement to run the experiments.

Package Name	Version	Package Name	Version
absi-py	0.7.1	pandas	0.25.0
astor	0.8.0	path.py	11.5.0
bleach	1.5.0	Pillow	6.0.0
cycler	0.10.0	pip	19.1.1
enum34	1.1.6	pip-autoremove	0.9.1
gast	0.2.2	protobuf	3.9.1
grpcio	1.21.1	Pympler	0.7
h5py	2.9.0	pyparsing	2.4.0
html5lib	0.9999999	python-dateutil	2.8.0

importlib-metadata	0.18	pytz	2019.1
Keras	2.2.4	PyYAML	5.1.1
Keras-Applications	1.0.6	scipy	1.2.0
Keras-Preprocessing	1.0.5	setuptools	41.0.1
kiwisolver	1.1.0	six	1.12.0
llvmlite	0.29.0	tensorboard	1.13.1
Markdown	3.1.1	tensorflow-estimator	1.13.0
matplotlib	3.0.3	tensorflow-gpu	1.13.1
mock	3.0.5	tensorflow-tensorboard	0.4.0
numba	0.44.1	termcolor	1.1.0
numpy	1.17.0	Werkzeug	0.15.5
opencv-python	4.1.0.25	wheel	0.33.4
zipp	0.5.1		

3.2 Deep learning Models (ResNet and Residual Attention Network)

We propose two novel Deep CNN architecture for image aesthetics assessment adapted from the recently published state-of-the-art image classification model (He et al., 2016; Wang et al., 2017). Both the models used in our experiments have been well tested as image classifier for a large number of classes. In experiments, our aim for predictions with higher correlation with human ratings, instead of classifying images to low/high score or mean score regression, the distribution of ratings are predicted as a histogram (Esfandarani and Milanfar, 2017). The squared EMD (Earth Mover's Distance) loss-based assessment was proposed by Esfandarani and Milanfar (2017). It shows a performance boost in the accurate prediction of the mean score. All network models for aesthetic assessment are based on image classifier architectures. Two different architectures (with and without attention mechanism) state-of-the-art networks were explored for the proposed applications. Networks used in our experiments were first trained and then fine-tuned using the large-scale aesthetics assessment AVA dataset. The complete architecture of this project consists of different sub-modules, and each of these sub-modules consists of building blocks, such as pooling, filters, activation functions, and so forth. The following sections provide more information on the sub-modules.

Residual Network (ResNet): Theoretically, neural networks should get better results as added more layers. A deeper network can learn anything that a shallower version of itself can, plus possibly some more parameters. The intuition behind adding more layers to a deep neural network was that more layers progressively learn more complex features (Figure 5). The first, second, and third layers learn features such as edges, shapes, objects, respectively, and so on. He et al. (2016), empirically presented that there is a maximum threshold for depth with the traditional CNN model. As more layers are added, the network yields better results until some point; then as more layers are added, the accuracy starts to drop. The reason behind failures of the very deep CNN is mostly related to optimization function, network weights initialization, or the well-known vanishing/exploding gradient problem. Vanishing gradients are especially blamed, however, He et al. (2016) argue that the use of Batch Normalization ensures that the gradients have healthy norms. In contrast, deeper networks are harder to

optimize due to added difficulty in the process of training; it becomes harder for the optimization to find the right parameters.

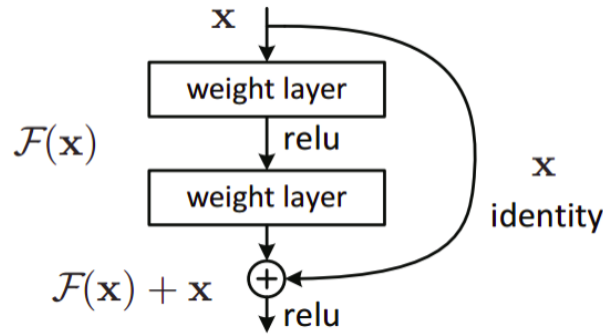


Figure 5. Residual learning: a building block.

The problem of training very deep networks has been attenuated with the introduction of a new neural network layer, the so-called Residual Block (Figure 5). Residual Networks attempt to solve the issue by adding the so-called skip connections. A skip connection is depicted in Figure 5. If, for a given dataset, there are no more things a network can learn by adding more layers to it, then it can just learn the identity mapping for those additional layers. In this way, it preserves the information in the previous layers and will not achieve lesser outcomes than the shallower layers. So, the most important contribution to ResNet architecture is the 'Skip Connection' identity mapping.

This identity mapping does not have any parameters and is essentially there to add the output from the previous layer to the layer ahead. However, sometimes x and $F(x)$ will not have the same dimension. Recall that a convolution operation typically shrinks the spatial resolution of an image, e.g. a 3×3 convolution on a 32×32 image results in a 30×30 image. The identity mapping is multiplied by a linear projection W to expand the channels of shortcuts to match the residual. This allows for the input x and $F(x)$ to be combined as input to the next layer. A block with a skip connection as in the image above is called a residual block, and a Residual Neural Network (ResNet) is just a concatenation of such blocks.

$$y = F(x; W_i) + W_s x$$

The above equation shows when $F(x)$ and x have a different dimensionality such as 32×32 and 30×30 . This W_s term can be implemented with 1×1 convolutions; this introduces additional parameters to the model. The Skip Connections between layers add the outputs from previous layers to the outputs of stacked layers. This results in the ability to train much deeper networks than what was previously possible. He et al. (2016) proposed 1001-layer ResNet network and achieved state-of-the-art performance on benchmark datasets such as CIFAR-10 (4.62 % error), and a 200-layer ResNet on ImageNet also achieved state-of-the-art performance. We specifically used an Inception-ResNet-v2 (Szegedy et al., 2016) as a base model and modified it into aesthetic assessment model to derive aesthetic score. The network is 164 layers deep and designed to classify images up to 1000 object categories. The pertained model has learned rich feature representations for a wide range of images from ImageNet database.

Attention in Neural Networks: Attention mechanisms in neural networks, otherwise known as neural attention or just attention, have recently attracted a lot of interest. Attention mechanisms expand capabilities of neural networks: they allow approximating more complicated functions, or in more intuitive terms, the approach enables focusing on specific parts of the input. They have led to performance improvements in natural language benchmarks, as well as to entirely new capabilities such as image captioning, addressing in memory networks and neural programmers. In general, a neural attention mechanism equips a neural network with the ability to focus on a subset of its inputs (or features): it selects specific inputs.

Let $x \in \mathbb{R}^d$ be an input vector, $z \in \mathbb{R}^k$ a feature vector, $a \in [0,1]^k$ an attention vector, $g \in \mathbb{R}^k$ an attention glimpse and $f_\phi(x)$ an attention network with parameters ϕ . Typically, attention is implemented as

$$a = f_\phi(x) \text{ and } g = a \odot z,$$

where $\odot$ is element-wise multiplication, while z is an output of another neural network $f_\theta(x)$ with parameters θ .

Residual Attention Network: A Residual Attention Network is a convolutional neural network that incorporates both attention mechanism and residual units which can incorporate with state-of-art feed forward network architecture in an end-to-end training fashion (see Figure 6). A Residual Attention Network is constructed by stacking multiple Attention Modules that generate attention-aware features. Residual unit is a basic component that utilizes skip-connections to jump over few layers with nonlinearities and batch normalizations which is the prominent feature is the attention module.

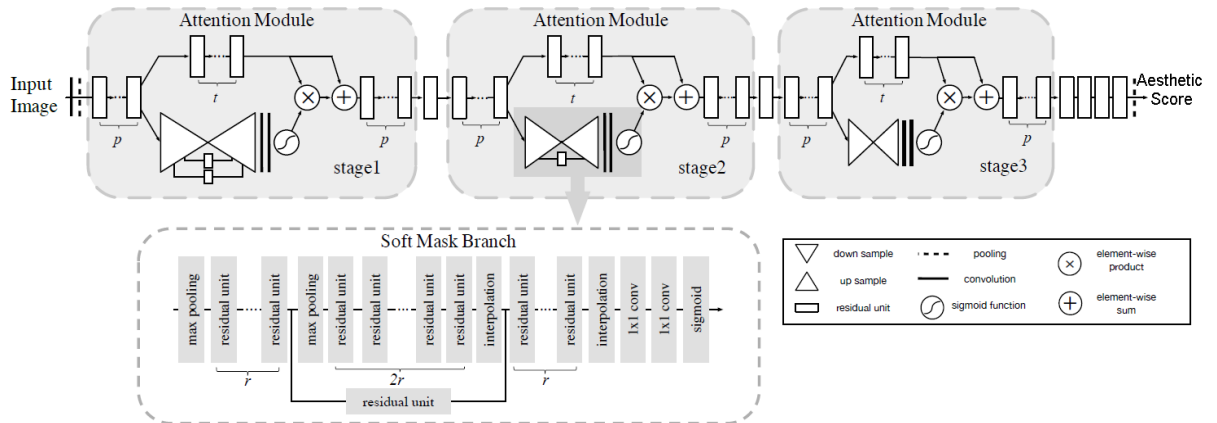


Figure 6. Architecture of the Residual Attention Network Architecture (proposed by He et al., 2016). The output layer is modified to produce an image aesthetic score instead of a class label.

Each Attention Module shown in Figure 6 is divided into two branches: mask branch and trunk branch. The trunk branch performs feature processing with Residual Units and can be adapted to any state-of-the-art network structures. Mask Branch uses bottom-up top-down structure softly weight output features with the goal of improving trunk branch features. The bottom-up step collects global information of the whole image by down sampling (i.e. max pooling) the image.

The top-down step combines global information with original feature maps by up sampling (i.e. interpolation) to keep the output size the same as the input feature map. Inside each Attention Module, bottom-up top-down feedforward structure is used to unfold the feedforward and feedback attention process into a single feedforward process. The attention-aware features from different modules change adaptively as layers are going deeper. Importantly, an attention residual learning to train very deep Residual Attention Networks which can be easily scaled up to hundreds of layers. The experiment also demonstrates that Residual Attention Network is robust against noisy labels.

In the Residual Attention Network approach, pre-activation Residual Unit (He et al., 2016), ResNeXt (Xie et al., 2017) and Inception elements (Szegedy et al., 2016) are used as basic units to construct Attention Module. Given trunk branch output $T(x)$ with input x , the mask branch uses bottom-up top-down structure (Long et al., 2014; Noh et al., 2015; Badrinarayanan et al., 2015) to learn same size mask $M(x)$ that softly weight output features $T(x)$. The bottom-up top-down structure mimics the fast feedforward and feedback attention process. The output mask is used as control gates for neurons of trunk branch similar to a Highway Network (Srivastava et al., 2015). Table 2 shows a comparative study on number of parameters are present in both the architectures we have considered in our experiments.

Table 2. Residual Attention vs. Inception ResNet. Residual Attention Network has more parameters to learn 1.53 times more parameters to learn during model training.

Model	Total parameters	Trainable parameters	Non-trainable parameters
Residual Attention Network (Wang et al., 2017)	83,401,866	83,289,738	112,128
Inception ResNet (Szegedy et al., 2016)	54,352,106	54,291,562	60,544

The goal is to predict the distribution of aesthetic scores for a given image. Ground truth score distribution of human ratings of a given image can be expressed as an empirical probability mass function $M = [m_{s_1} ; : : : m_{s_N}]$ with $s_1 < s_i < s_N$, where s_i denotes the i^{th} score bucket, and N denotes the total number of score buckets. In the GBR dataset $N = 10$,

$s_1 = 1$ and $s_N = 10$, Given the distribution of ratings as M , the mean score is defined as:

$$\mu = \sum_{i=1}^N s_i \times m_{s_i}$$

and the standard deviation of the score is computed as

$$\sigma = \sum_{i=1}^N (s_i - \mu)^2 \times m_{s_i}$$

We qualitatively compare images by the mean and standard deviation of scores instead of the only mean. We computed the mean and standard deviation for each image present in our dataset, and compare them with the mean and standard deviation obtained from the predicted distribution. The earth mover's distance (EMD) can effectively measure the distance between

two probability distributions over a region. This is also known as Wasserstein metric. We applied the EMD function to compute the distance between the ground truth aesthetic score distribution and predicted probability score.

We have also used Transfer learning, which is a powerful technique that lets deep learning architecture achieve state-of-the-art results with smaller datasets or less computational power. We took advantage of pre-trained models that have been trained on similar, larger AVA dataset (Murry et al., 2012). Our model learned via transfer learning and by optimizing parameter weights on previous learning, it can generally reach higher accuracy with much fewer data and computation time than models without transfer learning. In our experiment, we used a training versus test ratio of 80% and 20%, respectively from the data to fine-tune the model to achieve the desired accuracy.



Figure 7. Visual attention can be applied to any kind of inputs, regardless of their shape. In the case of matrix-valued inputs, such as images, the model learns to attend to specific parts of the image. Top and bottom rows in the figure show the original and masked images, respectively.

4.0 RESULTS AND DISCUSSION

This section presents the results obtained from two different approaches, namely Inception-ResNet and Residual Attention Network, over the 30 training epochs. Here epoch means one forward pass and one backward pass of all the training samples available during the training phase. More specifically, during the training process, we had to feed all the training images into the system to update the Neural Network training weights (i.e., the parameters), and this process repeated 30 times to obtain maximum accuracy. In our experiment, we recorded training loss (orange curve on Figure 8a. and 8c), validation loss (orange curve in Figure 8b and 8c.) after each epoch. Figures 8a and 8c show that 0.133 and 0.128 reported training losses were achieved from our experiment on original and masked datasets, respectively. We obtained the best validation loss of 0.100 and 0.09 from the original and masked datasets, respectively, on the GBR dataset.

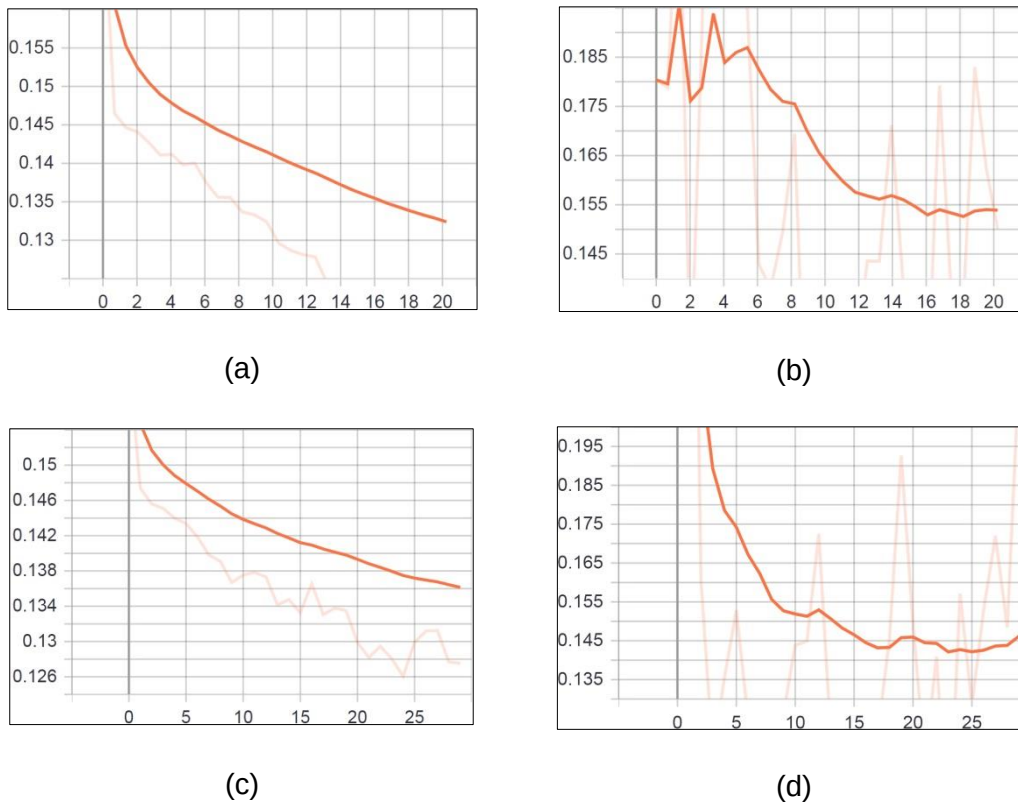


Figure 8. Training and validation loss graph from Inception ResNet: (a) Training loss on original dataset (i.e. non masked), (b) Validation loss on original dataset, (c) Training loss on masked dataset and (d) Validation loss on masked dataset.

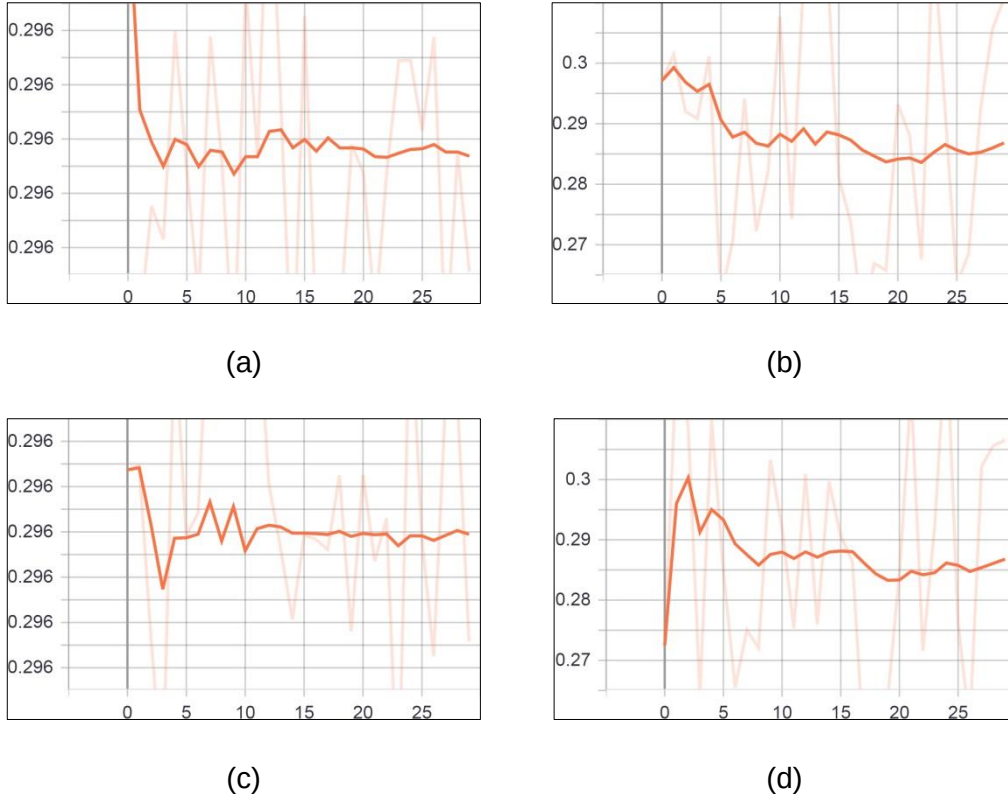


Figure 9. Training and validation loss graph from Residual Attention Network: (a) Training loss on the original dataset (i.e., non-masked), (b) Validation loss on original dataset, (c) Training loss on masked dataset, and (d) Validation loss on the masked dataset.

Training and validation losses were also recorded during the experiments using the Attention ResNet model. Training loss (orange curve on Figures 9a. and 9c), validation loss (orange curve on Figures 9b and 9c.) after each epoch, and we run this experiment up to 30 epochs. Figures 9a and 9c show that 0.2962 and 0.2961 reported training loss were achieved from our experiments on the original and masked dataset respectively. We obtained the best validation loss of 0.253 and 0.252 from the original and masked datasets respectively on the GBR dataset. From the experiments, it was observed that the Inception-ResNet model trained with masked images outperformed other models with 0.09187 validation loss. The Residual Attention Model's best validation loss was 0.252 compared to 0.09187 validation loss obtained by Inception-ResNet. Also, we observed that the model (validation loss was 0.09187) trained with attention masked images outperformed the model (validation loss was 0.10) trained with non-mask images.



(a) 5.67 ± 2.47 (5.67 ± 2.75)



(b) 7.62 ± 2.16 (7.70 ± 2.28)



(b) 6.53 ± 2.3 (6.54 ± 2.016)



(c) 6.98 ± 2.24 (7.0 ± 2.25)



(d) 8.017 ± 2.03 (8.058 ± 1.92)



(e) 5.30 ± 2.59 (5.28 ± 2.34)

Figure 10. Predicted aesthetic score of a few sample original images from the GBR dataset using our proposed Inception-ResNet-based aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.



(a) 6.41 ± 2.37 (4.0 ± 2.73)

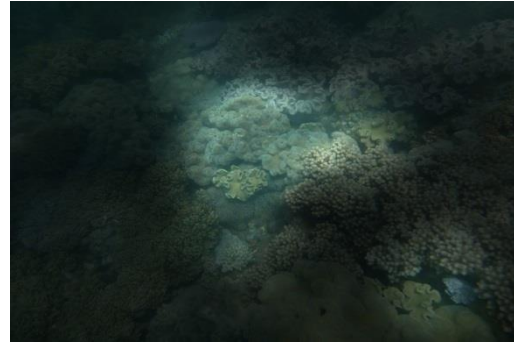


(b) 6.02 ± 2.40 (9.0 ± 1.22)

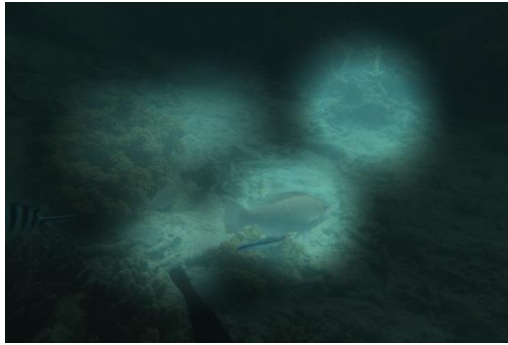
Figure 11. Some predicted aesthetic scores with a significant difference with ground truth from the GBR dataset using our proposed aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.



(a) 6.86 ± 2.29 (6.86 ± 2.82)



(b) 5.37 ± 2.69 (5.38 ± 2.55)



(c) 4.81 ± 2.63 (4.11 ± 2.55)



(d) 7.77 ± 1.89 (8.44 ± 1.25)



(e) 5.02 ± 2.67 (4.75 ± 1.98)



(f) 7.23 ± 2.20 (7.70 ± 1.85)

Figure 12. Predicted aesthetic score of a few sample original images from the GBR masked dataset using our proposed aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image.



(a) 2.85 ± 0.71 (7.23 ± 1.92)



(b) 3.11 ± 2.30 (8.5 ± 0.92)

Figure 13. Some predicted aesthetic scores with a significant difference with ground truth from the GBR dataset using our proposed aesthetic assessment model. Predicted Mean $\pm$ Standard Deviation (and Ground truth Mean $\pm$ Standard Deviation) scores are shown below each image

5.0 PROOF OF CONCEPT OF A WEB-BASED APPLICATION

To make the system of automated beauty scoring accessible, the team developed a proof of concept for how it could be deployed on web. The website is accessible at:

Proof of concept: <https://bigdata.ict.griffith.edu.au:4000/>

The website brings together research on the experiential value using Twitter data (see Stantic et al., 2019) and automated image rating, as explained in this present technical report. The main page of the prototype is shown in Figure 14. The key features will be explained briefly below.

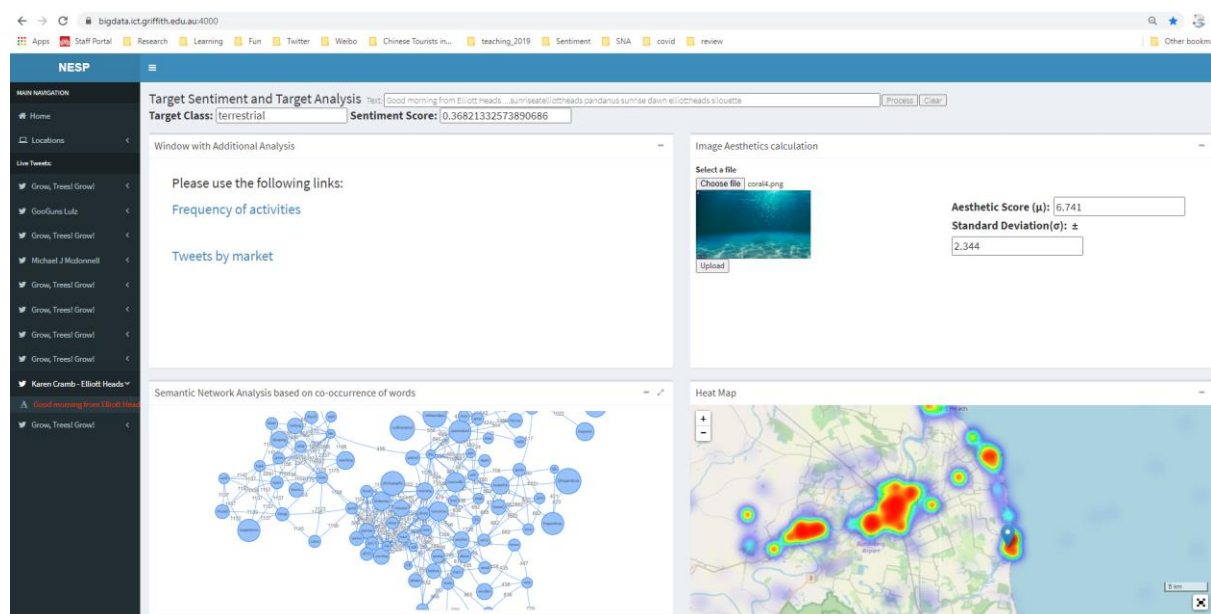


Figure 14. Proof of concept of a web-based application

5.1 Architecture for deployment

We have created a Flask application which is available on our bigdata server. Flask is a micro-framework and using flask we can create a python-based application. It does not include many of the tools that more full-featured frameworks might provide, but it exists as a module that we can import into our projects to enable initialization of a web application.

Next, we created a file that serves as the entry point for our application. This tells our uWSGI (gunicorn in Figure 15) server how to interact with the application. Once the uWSGI application server is up and running, waiting for requests on the socket file in our project directory. Finally, we needed to configure Nginx to pass the web requests to that socket using the uWSGI protocol.

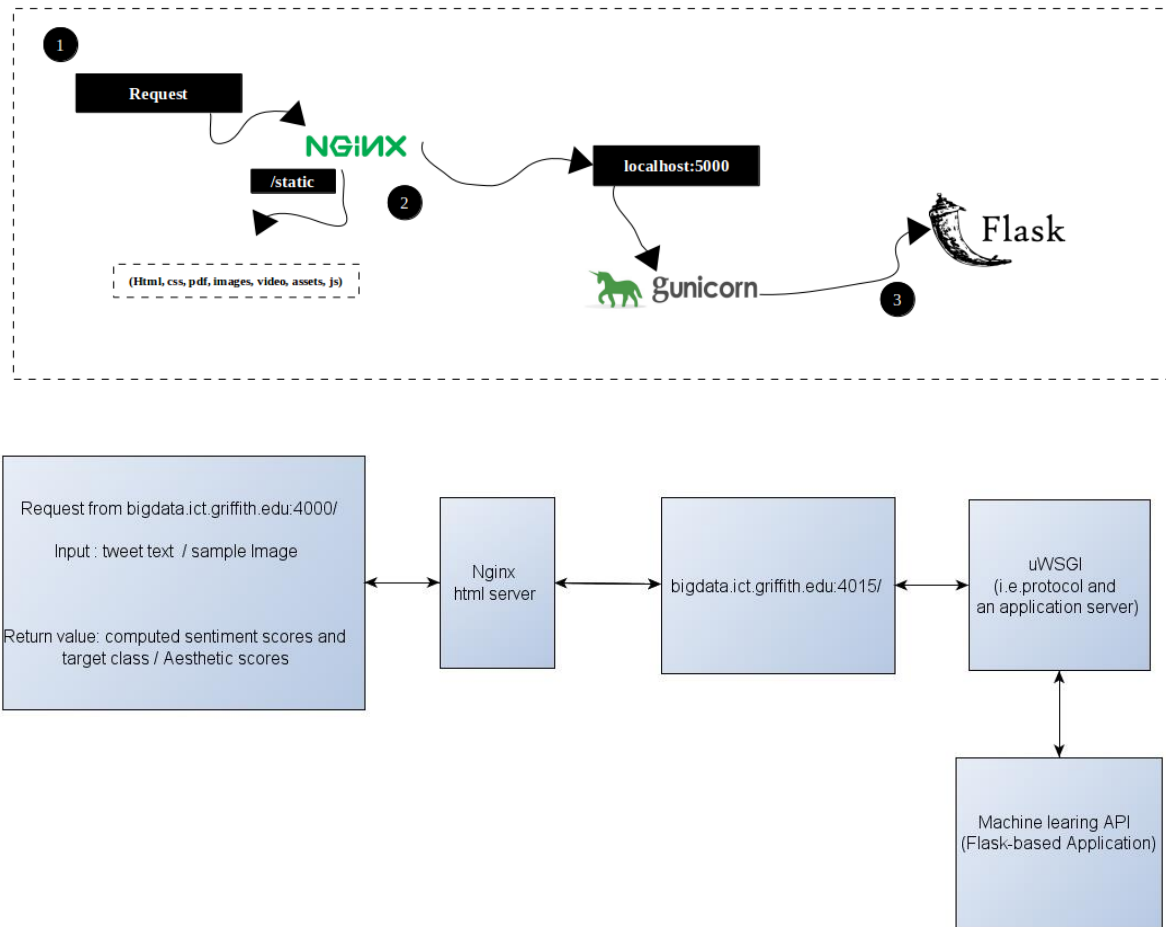


Figure 15. Architecture for web deployment.

5.2 Dashboard Functionality

As explained in Stantic et al. (2019), Twitter text – or any other text for that matter – can be used to identify GBR-related targets and sentiment. There are 11 targets and sentiment is measured on a scale between -1 (negative) and +1 (positive) (Figure 16).

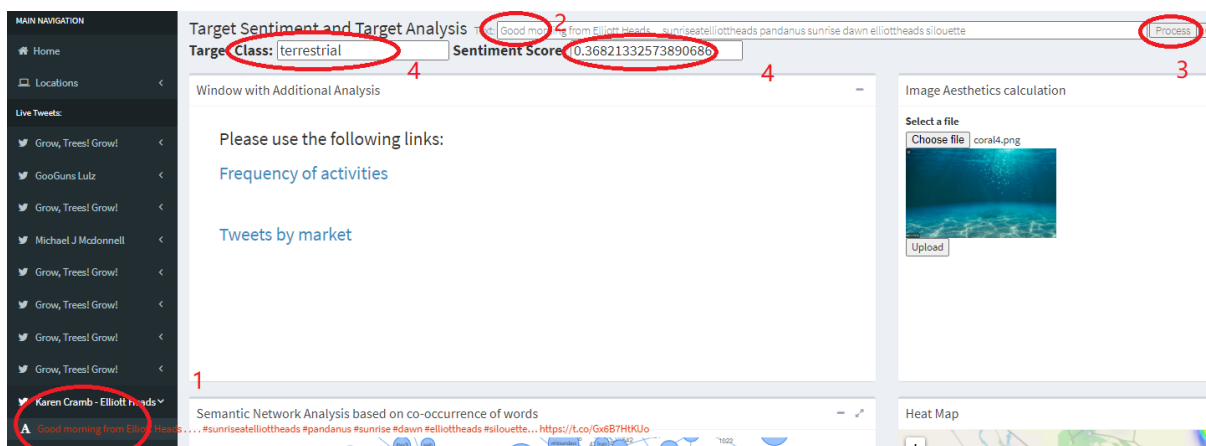


Figure 16. Target identification and sentiment, calculated from pasted text.

In addition, the interface provides the outputs of a Semantic Network Analysis which is based on co-occurrence of words in a particular text. It is possible to zoom in by clicking on any particular word (see also two examples in Figure 17).

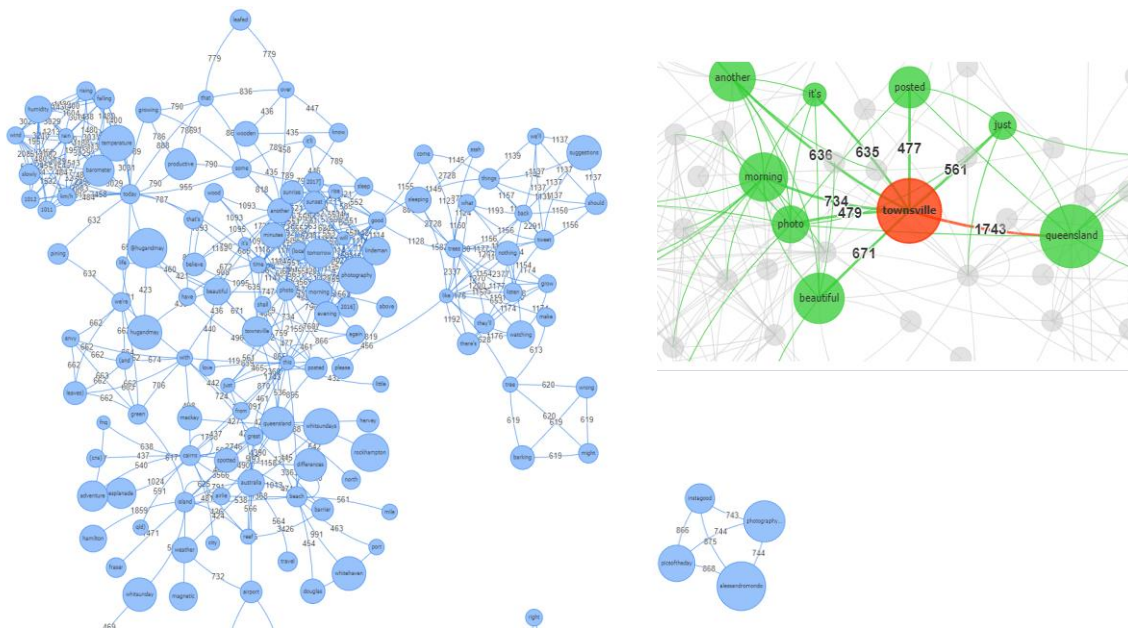


Figure 17. Examples of semantic network analysis outputs related to GBR Twitter communication as an example of text.

Further, the output provides a heat map based on location of posting. This works for text that is associated with metadata on geolocations. About 17% of Twitter posts in the GBR catchment were found to have coordinates attached to them (Becken et al., 2017).

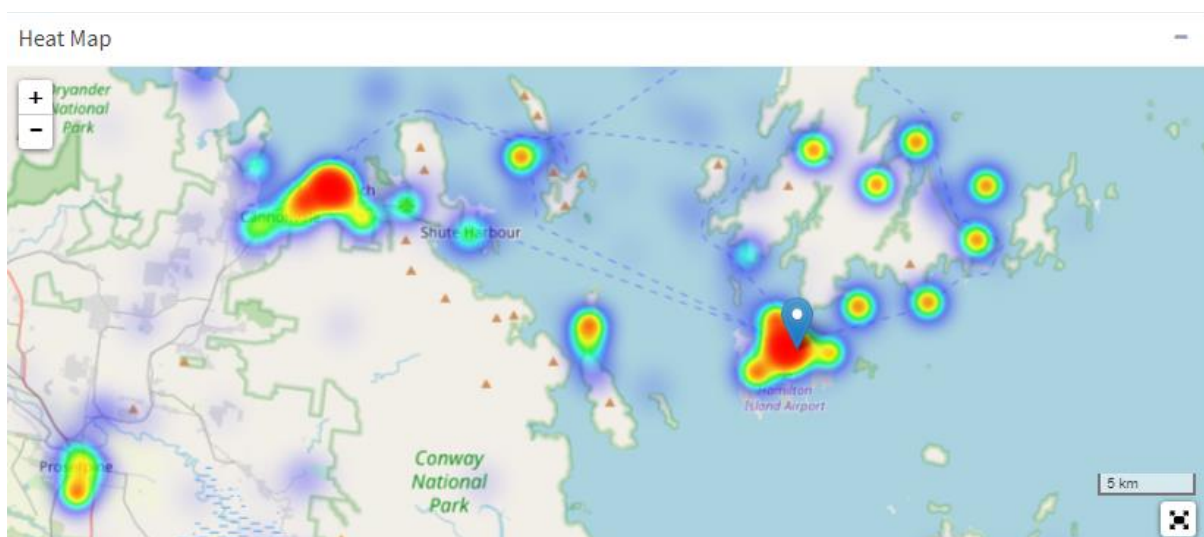


Figure 18. Heatmap that shows the frequency of posts by location.

Additional analysis can be undertaken based on various ways of filtering and presenting the data, for example:

- Post by market (origin of the social media user)
- Activity (e.g. by keywords, relevant to water based activities)
- A Combination of both.

Screenshots of example outputs are shown below in Figure 19.

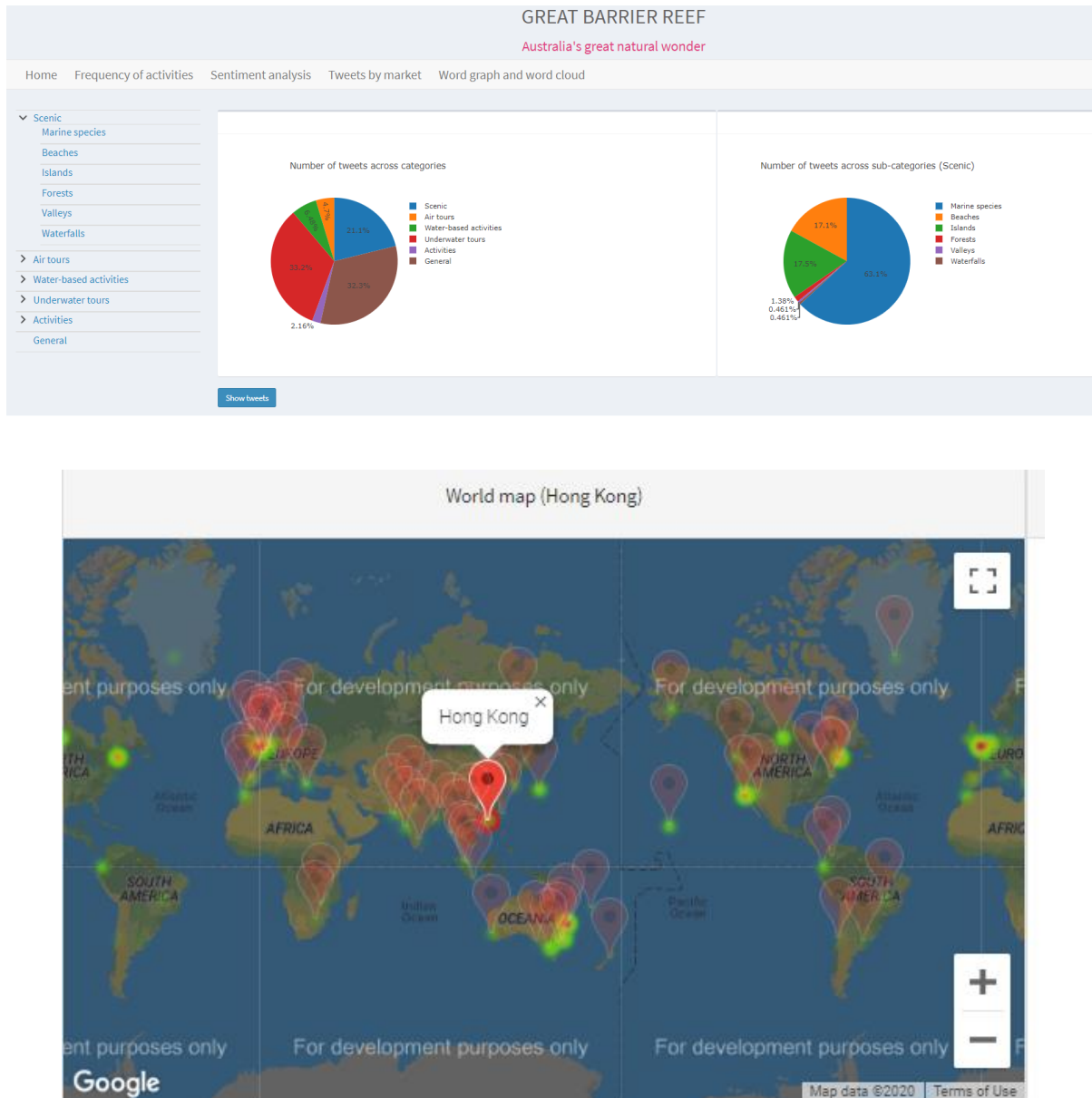


Figure 19. Examples of additional outputs that can be visualised when mining social media posts of interest to the GBR.

Finally, the website provides the opportunity to calculate the aesthetic score of an image that can be uploaded by the user. Figure 20 shows the example of an underwater image with a relatively high score of 6.862 (out of 10) and the standard deviation 2.37 (calculated as elaborated in section *Residual Attention Network* on page 14).

Image Aesthetics calculation

Select a file

Choose file coral.png

Upload

Aesthetic Score (μ) 6.862

Standard Deviation(σ): $\pm$ 2.347

Figure 20. Prototype option for calculating the beauty score of an image.

6.0 CONCLUSION

The objective of the work was to study the effectiveness of real human attention obtained using an eye-tracking device on deep Convolutional Neural Network architecture for automatic image aesthetic assessment and predicting an aesthetic score for images. The significance of the aesthetic evaluation system was studied in detail from the literature. A state-of-the-art deep Residual Attention Network architecture and Inception-ResNet were adapted to our requirement for the modelling of our GBR aesthetic assessment task. A wide range of experiments was conducted using a comprehensive analysis of performance is presented. In our experiment, Inception-ResNet-v2 learned rich aesthetic feature representations from the GBR dataset especially, with masked attention images and improved 0.9% accuracy compared to the model trained with non-mask images. Our model was able to effectively predict the distribution of aesthetic ratings, rather than just the mean scores. The prediction of distribution is more accurate quality prediction with higher correlation to the ground truth aesthetic score distribution. As part of our future work, we will exploit and use mask and non-mask images in parallel fashion to achieve better prediction accuracy.

REFERENCES

- Becken, S., Connolly, R.M., Chen, J. & Stantic, B. (2019). A hybrid is born: integrating collective sensing, citizen science and professional monitoring of the environment. *Ecological Informatics*, 52, 35-45.
- Becken, S., Stantic, B., Chen, J., Alaei, A. & Connolly, R. (2017). Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. *Journal of Environmental Management*, 203, 87-97.
- Bhattacharya, S., Sukthankar, R. & Shah, M. (2010). A framework for photoquality assessment and enhancement based on visual aesthetics. *ACM International Conference on Multimedia*, 271–280.
- Bianco, S., Celona, L., Napoletano, P., & Schettini, R. (2016). On the use of deep learning for blind image quality assessment. *arXiv preprint arXiv:1602.05531*.
- Bosse, S., Maniry, D., Wiegand, T., & Samek, W. (2016). A deep neural network for image quality assessment. *Computer Vision and Pattern Recognition*, 3773–3777.
- Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion*, 28(3), 221-243.
- Context. (2013). Defining the Aesthetic Values of the Great Barrier Reef. Retrieved from Canberra, Australia:
- Curnock, M., Pert, P., Smith, A., Molinaro, G., & Cook, N. (2020). Designing an Aesthetics Long-Term Monitoring Program (ALTMP) for the Great Barrier Reef: Technical report to the National Environmental Science Programme (Tropical Water Quality Hub). Report to the National Environmental Science Program. Reef and Rainforest Research Centre Limited, Cairns (87pp.).
- Datta, R., Joshi, D., Li, J. & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. *European Conference on Computer Vision*, 288–301.
- Deloitte Access Economics (2017). At what price? The economic, social and icon value of the Great Barrier Reef. Brisbane, Australia.
- Deng, Y., Loy, C.C. & Tang, X. (2016). Image aesthetic assessment: An experimental survey. Available (20/12/18) *arXiv preprint arXiv:1610.00838*.
- Dhar, S., Ordonez, V., & Berg, T. (2011). High level describable attributes for predicting aesthetics and interestingness. *Computer Vision and Pattern Recognition*, 1657–1664.
- Wang, F., Jiang, M., Qian, C. Yang, S., Li, C., Zhang, H., Wang, X. and Tang, X. (2017), "Residual attention network for image classification," In *Proc. Computer Vision and Pattern Recognition (CVPR)*, , pp.6450–6458.
- Noh, H., Hong, S. and Han, B. (2015), "Learning deconvolution network for semantic segmentation," *CoRR*, vol. abs/1505.04366, [Online]. Available: <http://arxiv.org/abs/1505.04366>
- Esfandarani H. T. and Milanfar, P., (2017), "NIMA: neural image assessment," *CoRR*, vol. abs/1709.05424. [Online]. Available: <http://arxiv.org/abs/1709.05424>
- Haas, A. F., Guibert, M., Foerschner, A., Co, T., Calhoun, S., George, E., Hatay, M., Dinsdale, E., Sandin, S.A., Smith, J.E., Vermeij, M.J.A., Felts, B., Dustan, P., Salamon, P. & Rohwer, F. (2015). Can we measure beauty? Computational evaluation of coral reef aesthetics. *PeerJ* 3: e1390 <https://doi.org/10.7717/peerj.1390>
- Long, J., Shelhamer, E. and Darrell, T., (2014), "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038. [Online]. Available: <http://arxiv.org/abs/1411.4038>

- Jiang, W., Loui, A., & Cerosaletti, C. (2010). Automatic aesthetic value assessment in photographic images. *International Conference on Multimedia and Expo (ICME)*, 920–925.
- He, K., Zhang, X., Ren, S. and Sun, J., (2016), “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- He, K. , Zhang, X., Ren, S. and Sun, J. , (2016) , “Identity mappings in deep residual networks,” *CoRR*, vol. abs/1603.05027. [Online]. Available: <http://arxiv.org/abs/1603.05027>
- Kang, L., Ye,P., Li, Y., & Doermann, D. (2014). Convolutional neural networks for no-reference image quality assessment. *Computer Vision and Pattern Recognition*, 1733–1740.
- Le, (Jenny) Dung, Becken, S. & Whitford, M. (2020). A cross-cultural investigation of the great barrier reef aesthetic value. *NESP TWQ*. <https://nesptropical.edu.au/wp-content/uploads/2020/09/NESP-TWQ-Project-5.5-Technical-Report-2.pdf>
- Lu, X., Lin,Z., Jin, H., Yang,J.,& Wang, J. Z. (2014). Rapid: Rating pictorial aesthetics using deep learning. *ACM International Conference on Multimedia*, 457–466.
- Marchesotti, L., Perronnin, F., Larlus, D., & Csurka, G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. *International Conference on Computer Vision*,1784–1791.
- Marshall, N., Marshall, P., & Smith, A. (2017). Managing for Aesthetic Values in the Great Barrier Reef: Identifying indicators and linking Reef Aesthetics with Reef Health. Cairns, Australia. Retrieved from <https://nesptropical.edu.au/wp-content/uploads/2018/09/NESP-TWQ-Project-3.2.4-Final-Report.pdf>.
- Marshall, N., Marshall, P., Curnock, M., Pert, P., Smith, A., & Visperas, B. (2019). Identifying indicators of aesthetics in the Great Barrier Reef for the purposes of management. *PLoS One*, 14(2), e0210196.
- Murray, N., Marchesotti, M. & Perronnin, F. (2012). AVA: A Large-Scale Database for Aesthetic Visual Analysis. Available (09/10/17) <http://refbase.cvc.uab.es/files/MMP2012a.pdf>
- Nishiyama, M., Okabe, T., Sato, I., & Sato, Y. (2011). Aesthetic quality classification of photographs based on color harmony. *Computer Vision and Pattern Recognition*, 33–40.
- Niu, Y. & Liu,F. (2012). What makes a professional video? a computational aesthetics approach. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22(7), 1037–1049.
- Perronnin F. & Dance, C. (2007) Fisher kernels on visual vocabularies for image categorization. *Computer Vision and Pattern Recognition*,1–8.
- Perronnin, J. S. F., & Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. *European Conference on Computer Vision*, 143–156.
- Ponomarenko, N., Ieremeiev, O., Lukin, V., Jin, L., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., and Kuo, C. C. (2013). A new color image database tid2013: Innovations and results. *International Conference on Advanced Concepts for Intelligent Vision Systems - Volume 8192*, 402–413.
- Srivastava, R. K. , Greff, K. and Schmidhuber, J., (2015) , “Training very deep networks,” *CoRR*, vol. abs/1507.06228. [Online]. Available: <http://arxiv.org/abs/1507.06228>
- Xie, S., Girshick, R., Dollr, P., Tu, Z. and He, K., (2017), “Aggregated residual transformations for deep neural networks,” in *In Proc. Computer Vision and Pattern Recognition (CVPR)*, , pp. 5987–5995.

- Schirpke, U., Timmermann, F., Tappeiner, U., & Tasser, E. (2016). Cultural ecosystem services of mountain regions: Modelling the aesthetic value. *Ecological Indicators*, 69, 78-90. doi: <https://doi.org/10.1016/j.ecolind.2016.04.001>
- Seresinhe, C.H., Moat, H.S. & Preis, T. (2017). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, 0(0) 1–16. DOI: 10.1177/0265813516687302
- Seresinhe, C.I., Preis, T. & Moat, H.S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society Open Science*, 4: 170170. <http://dx.doi.org/10.1098/rsos.170170>
- Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, 1409/1556, 2014.
- Stantic, B., Mandal, R. & Chen, J. (2019). Target Sentiment and Target Analysis. Technical Report. <https://nesptropical.edu.au/wp-content/uploads/2020/02/NESP-TWQ-Project-5.5-Technical-Report-1.pdf>
- Su, H.-H., Chen, T.-W., Kao, C.-C., Hsu, W. H., & Chien.S.-Y. (2011). Scenic photo quality assessment with bag of aesthetics-preserving features. *ACM International Conference on Multimedia*, 1213– 1216.
- Szegedy, C. Ioffe, S. and Vanhoucke, V. (2016), “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” *CoRR*, vol. abs/1602.07261. [Online]. Available: <http://arxiv.org/abs/1602.07261>
- Tang, X., Luo, W., & Wang, X. (2013). Content-based photo quality assessment. *IEEE Transactions on Multimedia*, vol. 15(8), 1930–1943.
- Badrinarayanan, V., Handa, A. and Cipolla, R. (2015), “Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixelwise labelling,” *CoRR*, vol. abs/1505.07293. [Online]. Available: <http://arxiv.org/abs/1505.07293>
- Vondrick, C., Patterson, D. & Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation,” *International Journal of Computer Vision*, 1–21, 10.1007/s11263-012-0564-1. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0564-1>
- Xue, W., Zhang, L., & Mou, X. (2013) Learning without human scores for blind image quality assessment. *Computer Vision and Pattern Recognition*, 995–1002.
- Zeiler, M. D. & Fergus, R. (2013). Visualizing and understanding convolutional networks. *CoRR*, Vol. abs/1311.2901. Available (12/12/17) <http://arxiv.org/abs/1311.2901>

← → ↻ bigdata.ict.griffith.edu.au:4000 🔍 ⭐ ⚙️ ⏸ Paused ⋮

NESP

MAIN NAVIGATION

- Home
- Locations
 - Cairns
 - Townsville
 - Arlie Beach
 - Mackay
- Live Tweets:
 - ♥ AUBREY BLACK
 - ♥ JML
 - ♥ Danilo
 - ♥ Same Weather
 - ♥ GooGuns Lutz
 - ♥ Hani Al-Hakim
 - ♥ GooGuns Lutz
 - ♥ MITHS
 - ♥ Joseph R
 - ♥ AUSSIECAMPFIREKITCH

Target Sentiment and Target Analysis

Text: Follow the Sun * bundaberg australia sunset sunrise sun pretty beautiful red orange sky

Target Class: Sentiment Score:

Process Clear

Window with Additional Analysis

Please use the following links:

[Frequency of activities](#)

[Tweets by market](#)

Image Aesthetics calculation

Select a file

Choose File coral.png

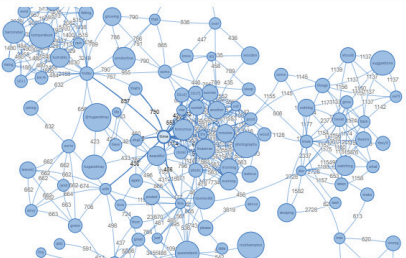


Upload

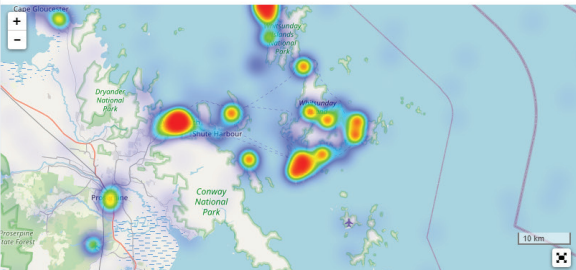
Aesthetic Score (μ):

Standard Deviation(σ): $\pm$

Semantic Network Analysis based on co-occurrence of words



Heat Map



10 km