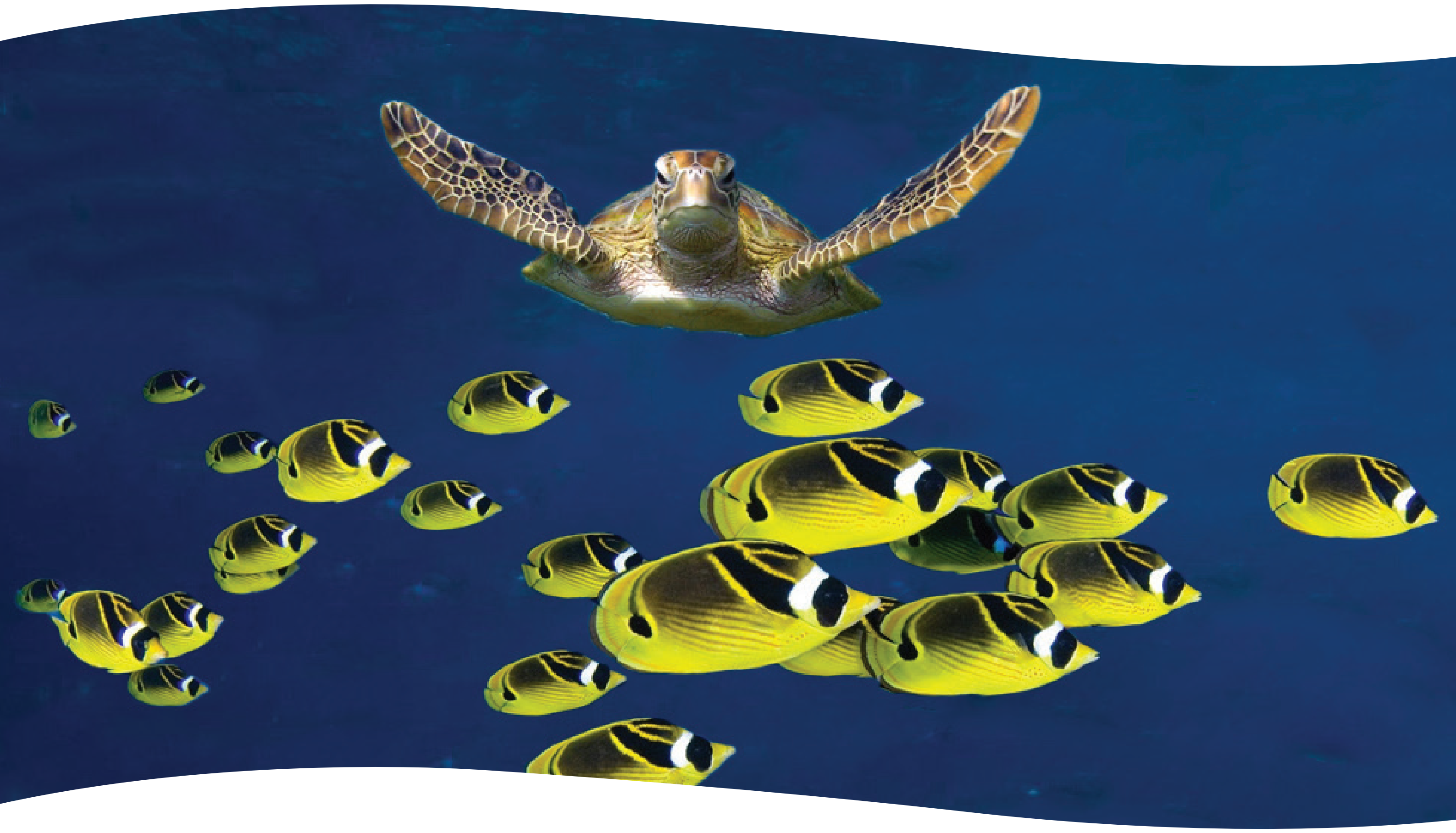
Tropical Water Quality Hub

National Environmental Science Programme

Final Report

Monitoring aesthetic value of the Great Barrier Reef by using innovative technologies and artificial intelligence

Susanne Becken, Rod Connolly, Bela Stantic, Noel Scott, Ranju Mandal and Dung Le



Griffith UNIVERSITY

Monitoring aesthetic value of the Great Barrier Reef by using innovative technologies and artificial intelligence

Susanne Becken[1], Rod Connolly[1], Bela Stantic[1], Noel Scott[1], Ranju Mandal[1] and Dung Le[1]
[1] Griffith Institute for Tourism, Griffith University

Supported by the Australian Government's
National Environmental Science Program
Project 3.2.3 Monitoring aesthetic value of the Great Barrier Reef by using artificial intelligence to score photos and videos

National Library of Australia Cataloguing-in-Publication entry:
978-1-925514-17-9

This report should be cited as:
Becken, S., Connolly, R., Bela Stantic, B., Scott, N., Mandal, R. and Le, D. (2017) *Monitoring aesthetic value of the Great Barrier Reef by using innovative technologies and artificial intelligence.* Report to the National Environmental Science Program. Reef and Rainforest Research Centre Limited, Cairns (48pp.).

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.

Cover photographs: Susanne Becken

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

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AP .... Average Precision
CNN .... Convolutional Neural Network
GBR .... Great Barrier Reef
GBRMPA ...... Great Barrier Reef Marine Park Authority
HoG .... Histogram of Gradients
mAP .... Mean Average Precision
NESP .... National Environmental Science Program
RIMReP .... Reef 2050 Integrated Monitoring and Reporting Program
SPP .... Spatial Pyramid Pooling
TWQ .... Tropical Water Quality

ACKNOWLEDGEMENTS

This project was supported through funding from the Australian Government's National Environmental Science Program (NESP) Tropical Water Quality (TWQ) Hub.

This is the final report of a NESP Tropical Water Quality Hub project (3.2.3) on the "Monitoring aesthetic value of the Great Barrier Reef by using artificial intelligence to score photos and videos". The research was funded to respond to the urgent need to develop a monitoring system for the aesthetic value of the Reef. This research used advanced technology to elicit what environmental and experiential attributes contribute to aesthetic value. A Big Data platform using artificial intelligence was then created to assess large volumes of visitor-supplied imagery.

EXECUTIVE SUMMARY

The aesthetic values of the Great Barrier Reef (GBR), Australia, are under multiple pressures. However, to date there is no agreed and systematic method to measure and approach the aesthetic beauty of a natural site. This project focused on underwater aesthetic value, because it is changing most rapidly and is at acute risk from deterioration of water quality, ocean acidification, coral bleaching, and biodiversity loss.

Capitalizing on the fast-developing area of citizen science and user online generated content, this research drew on publicly available images (in particular through the photo sharing site Flickr) to build a large dataset of GBR images for the assessment of aesthetic value. More specifically, the science in this project was developed in two parallel, but inter-linked research streams. The first stream focused on understanding what attributes determine perceived beauty of the GBR. Eye tracking technology (N= 21 images and 66 participants) and an online survey (N= 705) were used to identify key attributes and measure their relative importance. Second, using different types of neural network analyses (involving deep learning), the second stream developed a computer-based system for the automated identification of marine species (N= 50 species) and a model for automated assessment of image attractiveness (using N=2,500 GBR images).

The results provide empirical evidence for the usefulness of eye tracking statistics, as an objective measure for attention to an image, and in turn perceived beauty. The conjoint analysis conducted on online survey data revealed that all four beauty attributes (fish, coral, turtle and contrast) have significant impacts on people's perceived beauty ranking. On the basis of their relative important weights, fish is more important than the other elements in determining perceived beauty. More specifically, the diversity of sea life shown in a GBR picture increases its perceived beauty, in particular if the coral and fish have vivid colours. The presence of non-vivid fishes has a negative impact while the presence of non-vivid coral has slightly positive impact on the overall picture beauty. Additional analysis also revealed the importance of water quality on image rating.

Not all respondents perceive beauty in the same way, and differences were found, for example, in relation to age. The older the participants are, the more they rely on the fish attribute in evaluating GBR beauty. More research on different cultural groups and their perceptions of aesthetic value would be beneficial.

This research delivered a proof-of-concept of an automated species identification system, and an aesthetic assessment model. For both systems, the detection and assessment scores were high, despite some limitations, such as the availability of only comparatively small datasets for training. Ways for improving the systems and models are suggested, and with some additional expert input and resources, the systems presented in this report could serve as a robust basis for the future implementation of an automated monitoring system of key aspects of the GBR.

The report concludes with six recommendations on the next steps and possible extensions of this work. One recommendation specifically suggests to engage with Reef managers to further advance the use of Big Data and artificial intelligence for cost-effective and real time monitoring.

1.0 INTRODUCTION

1.1 Research Background

The Great Barrier Reef (GBR), as a UNESCO World Heritage site, is inscribed for multiple criteria, including its outstanding heritage value including its significant aesthetic characteristics. The aesthetic value, or beauty of the Reef, is of great importance to Australians and visitors, now and will remain so in the future. Both Tourism Australia and Tourism Events Queensland rely heavily on the GBR to promote Australia and Queensland to visitors domestically and internationally. Research by Tourism Australia (2015) shows that 42% of international visitors rank the GBR as Australia's most appealing tourist attraction. It is therefore no surprise that tourism generates a considerable economic benefit to the region. A 2017 Deloitte Access Economics study revealed that tourism generates an estimated AU$6.4 billion per year and sustains over 64,000 jobs.

Aesthetic value is referred to in several key documents, including in the original World Heritage Area listing of the GBR, the Burra Charter (2013), the Great Barrier Reef Region Strategic Assessment, and the Reef 2050 Long-Term Sustainability Plan. The GBR's "superlative natural beauty" (World Heritage Criterion vii) extends to above and underwater landscapes and ecosystems. Aesthetic value, however, is not well understood, and the 2012 UNESCO mission recommended that further work is required to understand the aesthetic dimensions of the GBR. Such work should consider that aesthetic value is a sensory perception of environmental attributes, influenced by cultural and personal factors.

Aesthetic values, like ecological values, are under multiple pressures. The Great Barrier Reef Outlook Report 2014 (Great Barrier Reef Marine Park Authority [GBRMPA], 2014) considers the underwater beauty of the GBR under threat due to reduction in coral cover and reduced water clarity. Theoretically, attributes associated with aesthetic value largely reflect the environmental values, such as biological diversity (see Johnston et al., 2013). However, to fully understand changes, and potentially declines, in value we require a better understanding of what exactly constitutes aesthetic value for the GBR and how it can be measured.

This project focused on underwater aesthetic value, because it is changing most rapidly and is at acute risk from deterioration of water quality, ocean acidification, coral bleaching, and biodiversity loss (GBRMPA, 2017).

1.2 Project Approach

This research capitalises on the increasing role that citizen science can play in the monitoring of natural assets (Becken et al., 2017; Seresinhe, Moat & Preis, 2017). Users of the GBR share substantial amounts of imagery (photos and videos) via different channels, such as Instagram, twitter, flickr, weibo and youtube. Photographic imagery/videos of users' experiences in the GBR contain information on the environmental and experiential attributes of aesthetic value (Johnston et al., 2013; Johnston & Smith, 2014).The images give important clues about what "matters" to Reef users. It is acknowledged that using material provided by humans is an anthropocentric approach. Thus, such an approach accepts that aesthetic value is a human concept arising from interactions between nature and people.

Aesthetic value, beauty or 'scenicness' are challenging topics to study, and research to date has largely focused on the scenic value of terrestrial ecosystems or urban parks. Quantifying scenicness is challenging, although more recent advances in data processing technology have accelerated progress. 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.

Very little research has focused on the aesthetic value and attributes of coral reefs. Interesting insights can be gained from Haas et al.'s (2015) study that used a computational evaluation of the natural beauty of coral reefs. A standardized computational approach was developed to assess coral reefs, using 109 visual features that were deemed important in defining beauty. These features were derived from the aesthetic appearance of art, and they included colour intensity, diversity of an image, relative size of objects, colour ranges, and types and distribution of specific objects, and texture. The authors found that the mathematical approach (also involving machine learning) was successful and provided a cost effective monitoring tool for marine managers.

Our project was informed by Haas et al.'s (2015) study in terms of its basic approach of using automated and computer-based methods; however, we took a more people-centred approach in identifying key attributes that determine perceptions of beauty. More specifically, to measure which elements of a simulated marine situation attract people's attention and trigger an emotional response, this project employed a set of laboratory-based methods (Scott et al., 2016) for testing visual stimuli. The development of material has drawn on the input by experts, as well as on existing research on marine systems (see Haas et al., 2014). This project sought to measure objects (e.g., fish), attributes (e.g., colour) or relations (e.g., fish with coral) that are important to define beauty, naturalness, discovery and other dimensions in an 'objectified' way.

The research team then used this new objective knowledge to inform the development of computer-based algorithms and deep machine learning methods for automated analysis of imagery. The underlying hypothesis is that it is possible to score large amounts of photographs and videos for their aesthetic values and implement an automated process to achieve this.
In summary, this project addressed the urgent need to understand and monitor the aesthetic value of the Great Barrier Reef. Focusing on the fast-changing underwater systems of the Reef, this research used advanced technology, such as eye tracking, to elicit what environmental and experiential attributes contribute to aesthetic value. A Big Data approach using artificial intelligence has further been created as part of this project to assess a large number of visitor-supplied imagery.

2.0 OVERALL STRUCTURE OF THE PROJECT

The science in this project was developed in two parallel, but inter-linked research streams. In the first part of the project, we focused on obtaining and assessing suitable imagery. The overall goal was to understand better what constitutes 'beauty' in underwater landscapes or vistas (observed through images), and to use automated processes to assess incoming imagery.

The first step involved securing suitable images. Several data sources are available and this project drew on images provided by tourism organisations and the photo-sharing website Flickr. Once images were secured, eye tracking technology helped to test and quantify the correlation between perceived beauty of underwater features and 'attention'. Building on the eye tracking insights, a survey collected data on beauty scores using a large number of images.

This survey generated insight and data for the second research stream that developed a neural network based method for the automated identification of species (i.e. defined through specific 'regions of interest') and other image attributes (see Figure 1 for a conceptual overview). Large volumes of training images were required to train and validate the algorithms and improve the assessment of further underwater images. The specific research objectives and data requirements are discussed in more detail in the following section.

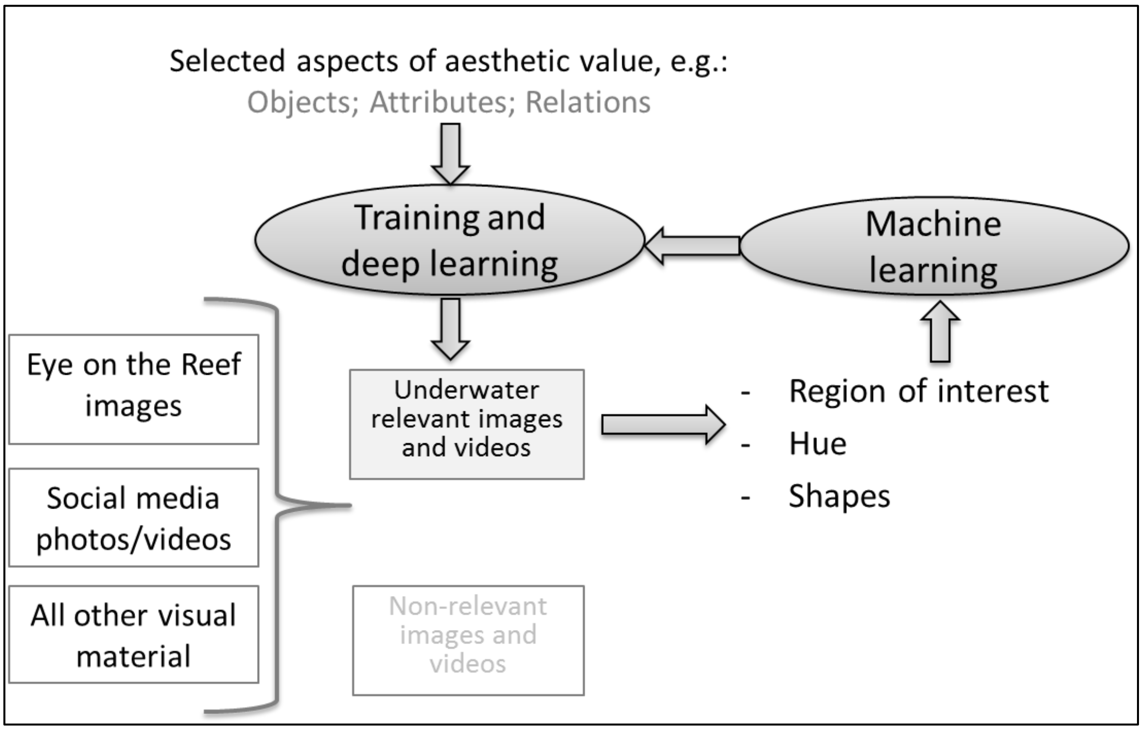


Figure 1: Conceptual overview of how images (i.e. photos and videos from a range of sources) can be selected and scored for beauty, which in turn informs machine and deep learning for future automated scoring.

2.1 Research Streams

The two research streams each followed a specific research objective. The objectives were:

- Objective 1: Determining aesthetic value using eye tracking: This first research stream aimed to assess the environmental and experiential attributes of the underwater Great Barrier Reef based on their aesthetic value.
- Objective 2: Developing algorithmic approach for the recognition of key objects in underwater images of the GBR and also automatically assessing their aesthetic values: This second research stream sought to develop and calibrate computer algorithms that automate the recognition of key objects, attributes and relations relevant to aesthetic value based on analysis of images and videos. The expected outcome was that the algorithms, through an automated, machine-based deep learning process, would be able to assess new imagery and score aesthetic value.

2.2 Data requirements

Both objectives relied on accessing suitable imagery, either in the form of photographs or videos. Photos were required both for the eye tracking experiment and the online survey, and videos (in addition to photographs) were used in the development and training of the proposed object detection/recognition algorithm. To select suitable images, the following characteristics had to be met:

- Were taken from the GBR only, displaying underwater scenery;
- Displayed no humans;
- Were taken from 1-2 metres from objects;
- Were high resolution.

Images came from several sources. First, we accessed the data libraries of Tourism Events Queensland (free to the public: https://teq.queensland.com) and Tourism Tropical North Queensland (free to members of the Tourism Tropical North Queensland tourism organisation http://www.ttnq.org.au).

In addition, and to ensure a much larger and diverse supply of photographs, we have developed an API that allowed us to download Flickr images by keyword (e.g., Great Barrier Reef available at https://www.flickr.com). As per our requirement, we downloaded all images with their metadata (including coordinates where available), and the data were stored in a MongoDB database for future access.

More specifically, the Flickr API consists of a set of callable methods, and some API endpoints. Our Flickr API has two components, given a set of keywords (i.e. Great Barrier Reef), we first downloaded the raw information (i.e. geo location, upload date, tags, machine tags, URL, etc.) of all the images and stored them in Json format in a MongoDB NoSQL database. Finally, the URL information, which is one of the attributes of raw data in Json format, was used to download the actual image media and stored in secured folders on file system organised by year.

Many Flickr users have chosen to offer their work under a Creative Commons license, and we can browse or search through content under each type of license. We have download around 27,000 Flickr images of approximately 63 GB. However, for this present study 2,500 images were selected from the dataset and labelled for use in the online survey (see below, Findings Stream 1).

For the conjoint analysis (see below), several images had to be photo-shopped to add some specific species and, in some cases, existing species had to be removed. In addition, we modified the images by changing brightness and contrast as per requirements for some the attributes to be tested in the experiment. We decided that 256x256 was a suitable standard size for the test images. This helped participants to view easily all nine pictures on a screen and facilitated their ranking process.

In addition, a series of underwater videos was taken at Lady Elliot Island to generate training material for automated fish identification of GBR marine species. Videos were taken by a research assistant from Griffith University, using a hand held camera (GoPro Hero 4). A database was created, whereby files were named to include species name and status of camera action (i.e. active or stationary). The videos were then annotated manually to identify fish species. This information was used to train a neural networks based system for the automated identification of fish species (see Findings Stream 2).

3.0 FINDINGS STREAM 1: UNDERSTANDING AESTHETIC VALUE

3.1 Introduction

This section reports on the first research stream, which aimed to develop a method to monitor the perceived underwater scenic beauty of the GBR. Given the importance of maintaining the natural beauty of the GBR, it is somewhat surprising that there is no agreed approach to evaluating perceived beauty. Some prior work has made recommendations but no objective assessment method is as yet available (Johnston et al., 2013). Development of an objective approach to measure perceived beauty would enable monitoring of the condition of the GBR on this attribute.

Johnston et al. (2013) suggested that there are four main data sources that can be used to examine aesthetic value. These include a) direct expressions of aesthetic values as evident in images and videos taken and posted online by visitors, Reef users and professional photographers, b) survey-based self-report on perceptions of beauty (often focused on tourists), c) mediated evidence of aesthetic values as expressed in promotional materials of the GBR, and d) bespoke consultation data with experts or stakeholders. This research will mainly focus on the first type of data, namely the GBR images provided by visitors, although some promotional material is also taken into consideration.

The findings of this research stream will be presented in two parts. First, the eye tracking study will be discussed, and this is followed by the findings from an online survey to further assess the perceived beauty of a large range of underwater GBR images.

3.2 Eye Tracking

The assessment of beauty is a subjective matter and the basic idea was that eye tracking might provide a more objective way of measuring how different people look at underwater photographs. Thus, to see if eye tracking provided a suitable method we examined the relationship between data from two eye tracking measures and respondents' beauty scores. In other words, the study relates average respondent rankings of the perceived beauty of underwater photographs with objective measures of 'attention'. Results indicate significant correlations between average perceived beauty rankings and average eye tracking fixation count and fixation duration measures.

3.2.1 Aesthetic experience and beauty

There are a number of ways of conceptualizing beauty, which may be grouped into objective and subjective approaches (Lothian, 1999). The first considers that beauty is an objective and intrinsic characteristic of an object. An alternative subjective approach to the conceptualization of beauty considers that an object can have no objective property of beauty and instead that it is the person's perception and interpretation of the object that determines its perceived beauty.

This study adopted the subjective approach to beauty, treating it as a human reaction that varies across different demographics and cultures. Such human reactions are measurable and provide an objective measure of "beauty", independent of respondent bias or the researcher's own beliefs. This second approach is consistent with the empirical aesthetics of Berlyne (1971) whereby neuroscientific evidence is considered to strengthen, complement, and constrain explanation of beauty at a psychological level. This concept of perceived beauty is similar to attractiveness.

Eye tracking techniques can measure visual attention processes in terms of the number or duration of eye-fixations on particular images or the areas of interest (AOI) within an image. Eye tracking has been used to measure preferential attention to emotional pictures and videos. These studies indicate that an emotional picture, either pleasant or unpleasant, is more likely to be fixated than a neutral picture (Simola, Le Fevre, Torniainen & Baccino, 2015). Emotional pictures capture attention during the early stages of picture processing by our brain (Nummenmaa, Hyönä, & Calvo, 2006). Similar effects have been observed between facial attractiveness and fixation duration and in differences in attention as measured by eye fixations to pleasant and unpleasant scenes (Calvo & Lang, 2004).

The literature of underwater aesthetics is limited. One study has looked at people's preferences for, affective responses to and the restorative potential of, different types of public aquaria exhibits (Cracknell, White, Pahl, & Depledge, 2017). Dinsdale (2009) showed that human visual evaluations provided consistent judgment of coral reef status regardless of their previous knowledge or exposure to these particular ecosystems. There is evidence that the evaluations of images of the pristine or damaged coral reefs can be in terms of pleasant or ugly. Coral reef photographs are associated with the "good" end of the evaluation dimension (Dinsdale, 2009). In summary, the beauty of a photo is considered here as a personal judgement based on pleasant emotional reactions to the photo. The perception of beauty causes a reorientation of attention towards the object that is perceived as beautiful. Therefore, the study proposes that there will be significant correlations between attention to and perceived beauty/ugliness of the images of underwater coral reefs and scenes.

3.2.2 Eye tracking method

This study uses two methods to evaluate 21 images of underwater reef scenes obtained from a variety of sources. Eye tracking provides a measure of attention while self-completion questions were used to rank images and obtain an average beauty rating to when viewing. A total of 66 volunteers participated. The participants were recruited using convenience sampling and received a small incentive to participate.

<u>Image rating</u>
Two self-report items were used in this study, one item evaluating the beauty of each picture (1-Not beautiful at all, 10-Very beautiful) and one item evaluating how the picture met subjects' expectations of the GBR where 1= Not at all and 10= Very much.

<u>Attention and beauty</u>
Eye tracking is a useful technique for objective measurement of attention (Scott, Le, Zhang, & Moyle, 2017 Online), by determining when an individual's eye pauses to examine or interpret a component of an advertisement or image (Rayner, Rotello, Stewart, Keir, & Duffy, 2001). In

the present study, the fixation count and total fixation were used to measure attention, considered here as to the degree of attraction for an observer (i.e. a measure of beauty).

Procedure

This first study took place in April and May 2017. Ethical approval was obtained through Griffith University, and respondents provided informed consent. Eye tracking data were collected using Tobii T60 Eye Tracker technology, which requires the respondent to sit in a chair in an upright position and view images on a computer monitor. Participants were free to look at each picture on screen as long as they wanted, and during this time their eye fixations were recorded. After viewing each picture, respondents then rated the beauty of each picture. After completing the eye tracking experiment, participant were interviewed to identify (by pointing) which part of each picture attracted their attention the most, and then to rate the perceived beauty of this AOI on a 10 point scale (1-Not beautiful at all, 10-Very beautiful).

Analysis of eye tracking data

The Tobii eye-tracker provides a record of the direction of the respondent's some 60 times per second and 'maps' this onto a location on the image being viewed. Subsequently, these mapped points analysed to determine fixation count and duration data. In the study, the criteria for a fixation was 250ms. The location the respondent pointed to in the post experiment interview was used to create an AOI, and fixation count and duration data were also estimated. The fixation, image beauty evaluations and post-experiment data for each photo, was then exported to IBM SPSS version 24 where t-tests and correlation analyses were conducted.

3.2.3 Eye tracking results

The data items collected were at picture level (the whole picture's beauty ranking, time to first fixation, total fixation duration, fixation count and total viewing time) as well as at AOI level (AOI's beauty, AOI's time to first fixation, AOI's total fixation duration, AOI's fixation count and AOI's total visit). All scores were calculated based on the average value of 66 participants for 21 photos, except the AOI beauty, which is identified and calculated based on 40 interviewed participants as this procedure was introduced during data collection.

What determines 'picture beauty'?

First, an independent t-test was conducted to determine whether there is a significant difference in eye tracking measures among beautiful (chosen using average respondent beauty score >=5) and ugly pictures (average respondent beauty score <5). All eye tracking measures (fixation count, fixation duration and total time visit) were significantly different between two groups of beautiful and ugly pictures ($p<0.05$), except the so-called "Picture time to first fixation" (Table 1).

Next, a correlation analysis was conducted (Table 2). The average beauty score of each picture is significantly correlated with ratings of whether the pictures met their expectation of the GBR, AOI beauty, picture fixation duration, picture fixation count and picture time visit as well as AOI's time to first fixation. Results show a strong correlation between picture beauty ratings and whether the picture met the respondent's expectations of the beauty of the GBR. This indicates a strong expectation that a GBR experience involves beautiful scenery, and that they have high aesthetic expectations. There is no correlation between picture's beauty and picture time to first fixation. Other correlations were not significant at the 0.05 level. Based on these results a series of regression models were constructed using selected variables with high correlations.

Table 1: Independent T-test of picture groups (beautiful versus ugly)

	Beauty	N	Mean	Std. Deviation	Std. Error Mean	Independent samples test	Levene's test Sig. (p)	T-test Sig. (2-tailed)
Expectation	>= 5	8	7.34	1.822	.644	Equal variances assumed	.000	.000
	< 5	13	2.91	.655	.182	Equal variances not assumed		.000
AOI beauty	>= 5	8	6.60	2.253	.797	Equal variances assumed	.028	.000
	< 5	13	2.70	1.399	.388	Equal variances not assumed		.001
P1st Time	>= 5	8	.03	.023	.008	Equal variances assumed	.160	.800
	< 5	13	.04	.040	.011	Equal variances not assumed		.773
PF Duration	>= 5	8	7.56	1.773	.627	Equal variances assumed	.211	.001
	< 5	13	5.16	1.131	.314	Equal variances not assumed		.006
PFcount	>= 5	8	28.14	5.106	1.805	Equal variances assumed	.139	.001
	< 5	13	20.07	3.754	1.041	Equal variances not assumed		.002
PTimevisit	>= 5	8	7.9242	1.90587	.67383	Equal variances assumed	0.173	0.001
	< 5	13	5.4317	1.18105	.32756	Equal variances not assumed		0.007

Table 2: Correlations between averages of variables (all pictures)

	Picture Beauty	Picture Expec	AOI beauty	Picture 1stTime	PF Duration	PF Count	PTime Visit	AOI 1stTime	AOI FDuration	AOI FCount	AOI TimeV
PBeauty	1	0.997**	0.911**	0.132	0.613*	0.692**	0.603**	0.731**	-0.166	-0.303	-0.178
		0.000	0.000	0.570	0.003	0.001	0.004	0.000	0.471	0.183	0.440
PExpec		1	0.903**	0.147	0.600**	0.679**	0.589**	0.729**	-0.179	-0.312	-0.190
			0.000	0.526	0.004	0.001	0.005	000	0.436	0.168	0.409
AOIbeauty			1	0.124	0.590*	0.648**	0.584**	0.589**	-0.090	-0.224	-0.100
				0.593	0.005	0.001	0.005	0.005	0.697	0.329	0.665
P1stTime				1	-0.073	-0.041	-0.072	0.242	-0.254	-0.240	-0.250
					0.753	0.859	0.756	0.291	0.267	0.295	0.274
PFDuration					1	0.968**	0.999**	0.222	0.635**	0.503*	0.624**
						0.000	0.000	0.335	0.002	0.020	0.002
PFCount						1	0.966**	0.328	0.493*	0.396	0.481*
							0.000	0.146	0.023	0.075	0.027
PTimevisit							1	0.215	0.650*	0.520*	0.639**
								0.349	0.001	0.016	0.002
AOI1stTime								1	-0.424	-0.490*	-0.433
									0.056	0.024	0.050
AOIFDuration									1	0.959**	1.000**
										0.000	0.000
AOIFcount										1	0.962**
											0.000
AOITimevisit											1

*** Correlation is significant at the 0.01 level (2-tailed).*

** Correlation is significant at the 0.05 level (2-tailed).*

A series of regression models helped to determine the relationship between eye tracking measures and individual picture beauty score (rather than with average picture scores). Because picture fixation duration, picture fixation count and picture total time visit are strongly correlated (Table 2 above), they were not included in the same regression model. Among these three individual eye tracking measures, picture fixation count is the best predictor of individual picture beauty (Adjusted $R^2 = 0.451$, $p<0.05$) (Model 2, Table 3).

Three other multivariate regression models including AOI time to first fixation and eye tracking measures were run (model 4 to 6). These models (4-6) offer better prediction power (Adjusted $R^2 >0.70$, $p<0.05$) (Table 3). The best model includes picture fixation count and AOI time to first fixation as predictors of picture beauty (model 5). Therefore, beauty ranking was related to quickly fixating on a part of an image considered beautiful and to fixating more times on the picture. It is worth noting that even though AOI beauty is correlated with picture beauty, this variable is not used for regression models because it is strongly correlated with other independent variables (Table 2 above).

Table 3: Picture beauty explained by eye tracking measures

Model		**Model summary**		**Unstandard-ized coefficients**	**Coefficients Sig.**
		Adjusted R square	**ANOVA sig.**		
1	(Constant)	0.343	0.003	-0.352	.827
	PFDuration			0.850	.003
2	(Constant)	0.451	0.001	-2.131	0.228
	PFCount			0.300	0.001
3	(Constant)	0.330	0.004	-0.257	0.874
	PTimeVisit			0.794	0.004
4	(Constant)	0.720	0.000	-1.106	0.305
	PFDuration			0.658	0.001
	AOI1stTime			8.019	0.000
5	(Constant)	0.737	0.000	-2.004	0.108
	PFCount			0.220	0.000
	AOI1stTime			7.234	0.001
6	(Constant)	0.713	0.000	-1.053	0.334
	PTimeVisit			0.615	0.001
	AOI1stTime			8.080	0.000
Dependent variable: Picture beauty					

<u>AOI and picture image attributes heat maps</u>

The eye tracking study data for individual beautiful and ugly images were analysed to produce fixation heat maps (Figure 2). These heat maps indicate that respondents' eyes fixated on particular parts of an image particularly fish, coral, turtles as well as rubbish. These results share similarities with work done by the CSIRO (pers. Communication, Dr. Nadine Marshall, October 2017), who tested coral cover, coral pattern, topography, fish abundance, and visibility and found topography, fish abundance, and visibility had a significant effect on aesthetic ranking of photos. The results were then used to select coral, fish and turtles as image attributes for testing in the conjoint experiment (Section 3.3).

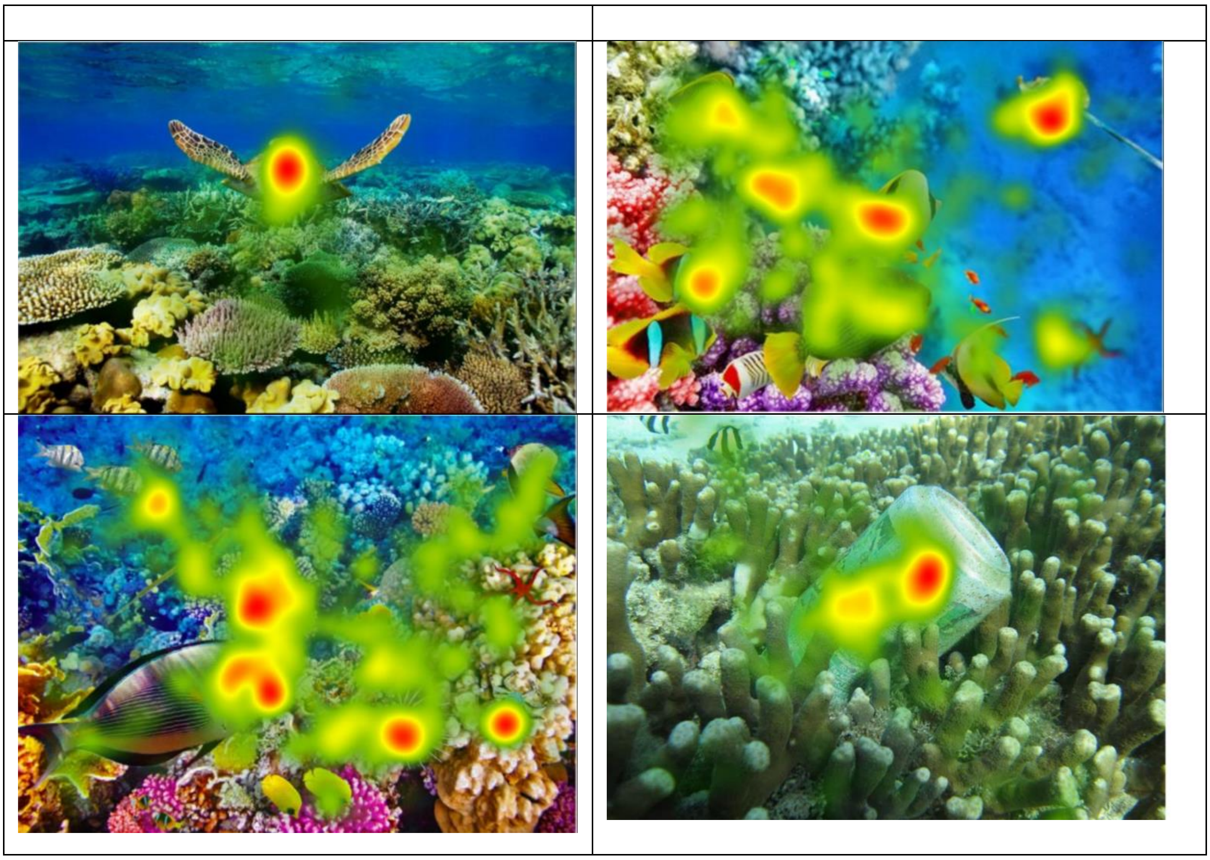


Figure 2: Examples of heat maps using eye tracking technology.

3.2.4 Conclusion from the eye tracking experiment

This study has tested an innovative eye tracking method to measure the relative perceived beauty of the GBR underwater images. The results of this study indicate that there is a significant correlation between eye tracking measures (fixation count, fixation duration and total time visit) and picture beauty scores. These results support previous research that indicates that pleasantness and beauty capture attention (Arriaza, Cañas-Ortega, Cañas-Madueño & Ruiz-Aviles, 2004). This study provides an evidence for the use of eye tracking measures in studying natural beauty. This method therefore suggests a means for objective measurement of the relative beauty of images and potentially for monitoring the aesthetics of environments, such as underwater coral reefs.

This study is an initial study only and is limited in terms of the number of images used (21 photos). Further research is needed to develop this method and compare its results with expert based judgement of aesthetic value (Schirpke, Timmermann, Tappeiner & Tasser, 2016), especially of coral reefs. To this end, the method needs retesting using larger numbers of photos – perhaps 60 pictures (30 beautiful vs 30 ugly pictures). It also would be useful to examine respondents from important visitor target markets with different cultural backgrounds (Asia vs Western) to determine if there are differences in their aesthetic perceptions. It would be useful to also examine the differences of Indigenous and Non-Indigenous Australians' aesthetic perceptions.

In this study only one item was used to measure respondents' self-reported beauty ranking and it may also be useful to use other scales, such as visual appeal.

3.3 Online survey and conjoint analysis

An online survey (N=772, with N=705 usable surveys) was conducted to examine the comparative effect of different types of image attribute on image overall beauty ranking. The image beauty rankings of 2,500 images were also collected and used in developing algorithms and machine learning for image evaluation. This is discussed in the Findings – Stream 2, Section 5.

There are a large number of potential image attributes that could affect overall image ranking. Not all of these could be tested due to the costs of survey data collection. As a result of the preparatory research and sample cost considerations, the project team selected key factors for this study, namely the diversity of marine life, colours and the picture's colour contrast. A conjoint experimental design was used to collect data.

3.3.1 Conjoint survey pre-survey experiment

Following the eye tracking experiments, a pre-survey experiment was conducted in order to confirm whether the presence of an aesthetic object (i.e. fish or turtle) improves people's perceived beauty ranking of the GBR pictures.

Two pictures containing dead-coral only, selected from the pictures used in the previous eye tracking experiment, were photoshopped to add a turtle to the background (Picture 1 and 2 below). Three people rated Picture 1 and three others rated Picture 2 (Figure 3). Picture 1 was rated at an average of 3 out of 10-point beauty scale (1-Not beautiful at all, 10-Very beautiful) which is similar with the beauty score of the original picture (i.e. dead coral picture without turtle) in the previous study. However, the average beauty score of the photoshopped Picture 2 improved to 7 out of 10 in (average score of the original picture in the previous study was 3.6 out of 10). The after-experiment interviews revealed that participants thought picture 1 was not well photoshopped and looked fake while picture 2 was not considered photoshopped.

This pre-test had two important implications: (1) Photoshopped pictures must be well designed and look natural to participants, and (2) the presence of an aesthetic object can improve the perceived beauty of GBR pictures. Based on this initial result, nine pictures were photoshopped for further testing the importance of each aesthetic object (i.e. coral, fish, turtle) on GBR perceived beauty.

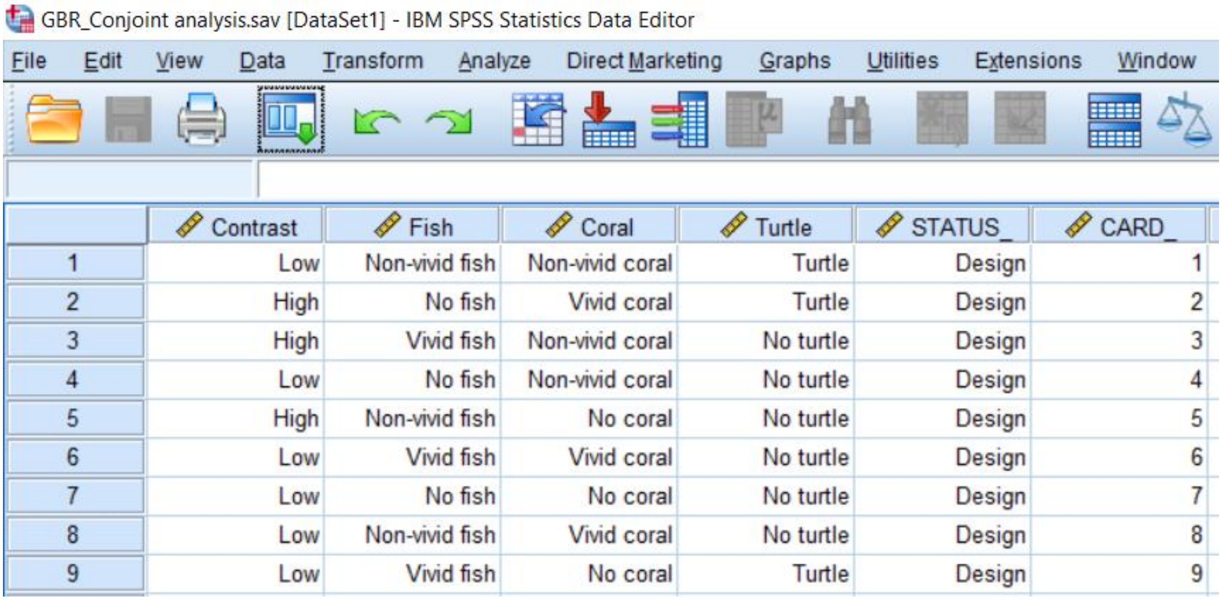

Picture 1 Picture 2

Figure 3: Two photoshopped pictures used in the pre-survey experiment.

Thus, based on a literature review and the eye tracking experiment results, it was possible to select key factors that appear influential in terms of the perceived image beauty. These factors translated into four beauty attributes at different levels as follows.

- Contrast: High vs low
- Coral: No coral, non-vivid coral, vivid coral
- Fish: No fish, non-vivid fish, vivid fish
- Turtle: No turtle, turtle

These four attributes and their levels give rise to 24 possible combinations (2x3x3x2). It would clearly be tedious for respondents to rank so many scenarios, so the Orthoplan subroutine in SPSS was used to produce an orthogonal main-effects design, which ensures the absence of multi-collinearity between attributes (Silayoi & Speece, 2007). Figure 4 shows the nine combinations of attribute level recommended by SPSS Orthoplan. Based on this suggestion, nine photoshopped pictures were created in Photoshop. In order to avoid bias, the attributes (i.e. fish, turtle, and coral) were kept the same across different pictures (Figure 4). These pictures were checked and commented by 10 different people until they were well designed and looked natural.



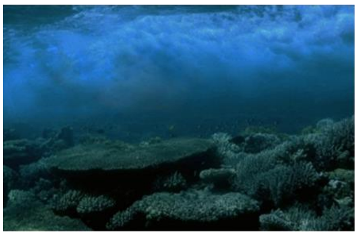
	Contrast	Fish	Coral	Turtle	STATUS_	CARD_
1	Low	Non-vivid fish	Non-vivid coral	Turtle	Design	1
2	High	No fish	Vivid coral	Turtle	Design	2
3	High	Vivid fish	Non-vivid coral	No turtle	Design	3
4	Low	No fish	Non-vivid coral	No turtle	Design	4
5	High	Non-vivid fish	No coral	No turtle	Design	5
6	Low	Vivid fish	Vivid coral	No turtle	Design	6
7	Low	No fish	No coral	No turtle	Design	7
8	Low	Non-vivid fish	Vivid coral	No turtle	Design	8
9	Low	Vivid fish	No coral	Turtle	Design	9

Picture 1

Picture 2

Picture 3

Picture 4



Picture 5

Picture 6

Picture 7

Picture 8

Picture 9

Figure 4: Orthogonal main-effects design and nine images used for conjoint analysis.

3.3.2 Online survey design and method

Some 10,000 underwater pictures of the Great Barrier Reef originated from Flickr, and websites of the Tourism and Events Queensland and Tropical Tourism North Queensland (see further above) were initially considered in this study. Flickr pictures were downloaded using the key words “Great Barrier Reef”. These were manually reviewed and 2,500 pictures were then selected from the downloaded database. These pictures were relabelled from 1 to 2,500 and uploaded to the survey platform Qualtrics.

Survey instrument

The questionnaire used in the Qualtrics survey included four different sections:

- Section 1: Demographic details
  Participants answered demographic questions (gender, age).
- Section 2: Picture rating
  Each participant rated 40 pictures randomly selected from the 2500 pictures. Each picture was rated at least 10 times to enable an average beauty score to be calculated. Based on feedback from a pilot test, participants were asked to take a break of 15 seconds after rating the 20 first pictures. These data were then used to train and calibrate the proposed automatic aesthetic prediction system (see Findings Stream 2).
- Section 3: Conjoint picture ranking (see below for more details)
  Conjoint analysis is a research method used for exploring how different independent variables influence the dependent variable. In this case, the study examines the influences of aesthetic factors on the GBR pictures beauty. As explained above, four aesthetic attributes (i.e. fish, coral, turtle, and contrast) were chosen based on the earlier eye tracking experiment results, literature review and a pre-test experiment. For the conjoint analysis, participants were asked to rank the nine GBR pictures in terms of their beauty (1 - The most beautiful picture to 9 – The least beautiful picture). The pictures were presented to participants in randomised order to avoid any bias related to picture positions.
- Section 4: Open-ended question and past experiences.
  Before completing the questionnaire, participants answered an open-ended question on factors that they considered influenced their evaluation of the GBR images (i.e. "From your point of view, what are the factors that make a picture of the Great Barrier Reef beautiful?"). Respondents were also asked to disclose their travel experiences (at three levels) with respect to different activities, such as diving and snorkelling, on the Great Barrier Reef.

Data collection

The survey was launched online on 4th October 2017 using the online survey company Qualtrics, which provides a population representative of the Australian population. A total of 772 survey completions were recorded after 3 weeks. However, only 705 questionnaires were eligible for data analysis after uncompleted questionnaires were excluded.

3.3.3 Conjoint analysis results

Relative importance of attributes

The conjoint results indicate that fish is the most important attribute in determining consumer perceived beauty. The relative importance weight (i.e. importance value) of this attribute is 32% (Table 4). This means that the presence/absence and colour (vivid versus non-vivid) of a fish makes the most difference of any attribute tested in the overall beauty ranking. Coral is the second most important attribute (25.10%), followed by colour contrast (22.23%) and the presence of a turtle (20.63%).

Utility scores provide more details on how different levels of four beauty attributes influence respondents' beauty evaluation. Vivid fish, vivid coral, the presence of turtle and high colour contrast contribute positively to a GBR picture beauty. The positive utility of 1.822 for vivid fish, 1.114 for colour contrast, 1.012 for the presence of turtle and 1.037 for vivid coral clearly point out the positive impact of these attributes on tourists' perceived beauty. The best combination of a beautiful GBR picture should include vivid fish, vivid coral, and turtle along with a high level of colour contrast.

Non-vivid fish has a negative utility (-0.63) which means that the presence of fishes in a GBR picture may not increase perceived beauty if fish is not of vivid colours. Interestingly, non-vivid coral has a low positive utility (0.043). In contrast with the common concern that coral in unhealthy state (i.e. losing its colours) negatively influence tourist beauty evaluation of the GBR, this study provides an evidence that the presence of non-vivid coral (i.e. dead or bleaching coral) will not negatively impact tourist perceived beauty.

Table 4: Results of the conjoint analysis (n=705)

Attribute	Level	Utility score	Importance value (%)
Fish	No fish	-1.191	32.03
	Non-vivid fish	-0.630	
	Vivid fish	1.822	
Coral	No coral	-1.080	25.10
	Non-vivid coral	0.043	
	Vivid coral	1.037	
Contrast	High	1.114	22.23
	Low	-1.114	
Turtle	No turtle	-1.012	20.64
	Turtle present	1.012	

<u>Impact of other variables on beauty perceptions</u>

ANOVA tests were conducted in order to examine the influences of consumer sociodemographic (age, gender) and travel characteristics (travel and diving/snorkelling experiences) on the relative importance weights (i.e. importance value) of these beauty attributes. The results show that age has an impact on the importance values of fish and coral attributes ($p<0.01$), and travel experience is related to the importance value of coral attribute ($p<0.05$). Older participants appear to rely more on the fish attribute than coral attribute in evaluating the beauty of a GBR picture. Interestingly, regular visitors to the GBR and those respondents who have never been to the GBR give a higher importance weight to coral attribute than tourists who visited the GBR one or two times before (Table 5).

Table 5: ANOVA results on the relative importance weights of beauty attributes (i.e. importance values) among different groups of tourists (%)

	Contrast		**Fish**		**Coral**		**Turtle**	
	Mean	*Sig.*	*Mean*	*Sig.*	*Mean*	*Sig.*	*Mean*	*Sig.*
Age		0.474		0.000*		0.000*		0.584
18-24 years old	23.90		27.32		29.24		19.51	
25-34 years old	22.25		29.36		27.73		20.66	
35-44 years old	22.34		30.88		26.20		20.58	
45-54 years old	22.37		33.77		23.91		19.94	
55-64 years old	21.62		34.76		21.75		21.85	
65+ years old	21.28		35.51		22.13		21.06	
Gender		0.963		0.866		0.966		0.961
Male	22.13		32.17		24.99		20.70	
Female	22.32		31.90		25.20		20.56	
Other	22.73		27.28		27.27		22.72	
Travel experiences		0.348		0.844		0.050*		0.740
Never	22.34		31.82		25.85	*	19.98	
One or two times	22.43		32.33		23.48		21.75	
More than 3 times	20.41		32.21		27.01		20.63	
Diving or		0.255		0.152		0.140		0.254
snorkelling	22.45		32.69		24.77		20.07	
experiences	22.44		31.61		24.63		21.30	
Never	20.68		30.32		27.75		21.23	
One or two times								
More than 3 times								

* Significant at 0.01 level

** Significant at 0.05 level

In order to understand how beauty attributes are related to perceived beauty rankings, further an ANOVA analysis was conducted using the utility scores. Results of the ANOVA analysis showed that only age is significantly related to utility scores while gender, travel experience and diving/snorkelling experience were not significantly related. The utility scores of four attribute levels including no fish, vivid fish, no coral and non-vivid coral (Table 6) differed by age groups. Under-45-year-old participants had lower utility scores than people over 45 years old for two fish attribute levels: no fish (negative impact) and vivid fish (positive impact). This means that the absence of fish results in a lower decrease of perceived beauty and the presence of vivid fish leads to a lower increase of perceived beauty for under-45-year-old participants in comparison with older groups. In contrast, the utility score for young and medium age participants is higher than old participants for the no-coral attribute level. Hence, the absence of coral in a GBR picture decreases the perceived beauty of young and medium age participants more than for older groups.

The results for the attribute level of non-vivid coral are complex. The presence of non-vivid coral in GBR pictures has a negative impact on perceived beauty for young and medium age subjects but a positive impact on perceived beauty of old people. This result may be explained as follow. Participants over 45 years old seem to pay more attention to the health of the GBR. Hence, the presence of non-vivid coral makes them rate a GBR picture less beautiful. In contrast, participants at young and medium age pay less attention to the coral state and the presence of non-vivid coral does not negatively influence their perceived beauty.

Table 6: ANOVA results on utility scores of beauty attributes among different groups of visitors/Reef users (%)

		N	Utility mean	ANOVA sig.
No fish	18-24 years old	88	-.8258	0.000
	25-34 years old	140	-1.0286	
	35-44 years old	125	-1.1680	
	45-54 years old	120	-1.3222	
	55-64 years old	109	-1.3242	
	65+ years old	123	-1.4173	
	Total	705	-1.1915	
Vivid fish	18-24 years old	88	1.4015	0.000
	25-34 years old	140	1.6310	
	35-44 years old	125	1.7947	
	45-54 years old	120	1.9750	
	55-64 years old	109	2.0428	
	65+ years old	123	2.0217	
	Total	705	1.8217	
No coral	18-24 years old	88	-1.2727	0.000
	25-34 years old	140	-1.3024	
	35-44 years old	125	-1.1547	
	45-54 years old	120	-.9639	
	55-64 years old	109	-.8563	
	65+ years old	123	-.9241	
	Total	705	-1.0799	
Non-vivid coral	18-24 years old	88	.1667	0.003
	25-34 years old	140	.2476	
	35-44 years old	125	.0453	
	45-54 years old	120	-.0694	
	55-64 years old	109	-.0765	
	65+ years old	123	-.0678	
	Total	705	.0426	

Open ended question on what determines beauty

At the end of the online survey, respondents were asked (multiple responses accepted) which factors make a picture of the GBR beautiful. Answers were recorded and analysed in a thematic analysis. The concept map generated using Leximancer (software to analyse relationships been occurrences of words) indicates five major themes: colours (359 occurrences), coral and sea features (325), diversity of sea life (175), water quality (167), natural beauty (109) (Figure 5). Three of these factors (i.e. coral and sea animals, colours and diversity of sea life) were included in the conjoint analysis. The last factor, water quality, has not been investigated directly, but may be related to vividness. This should be taken into consideration in future research.

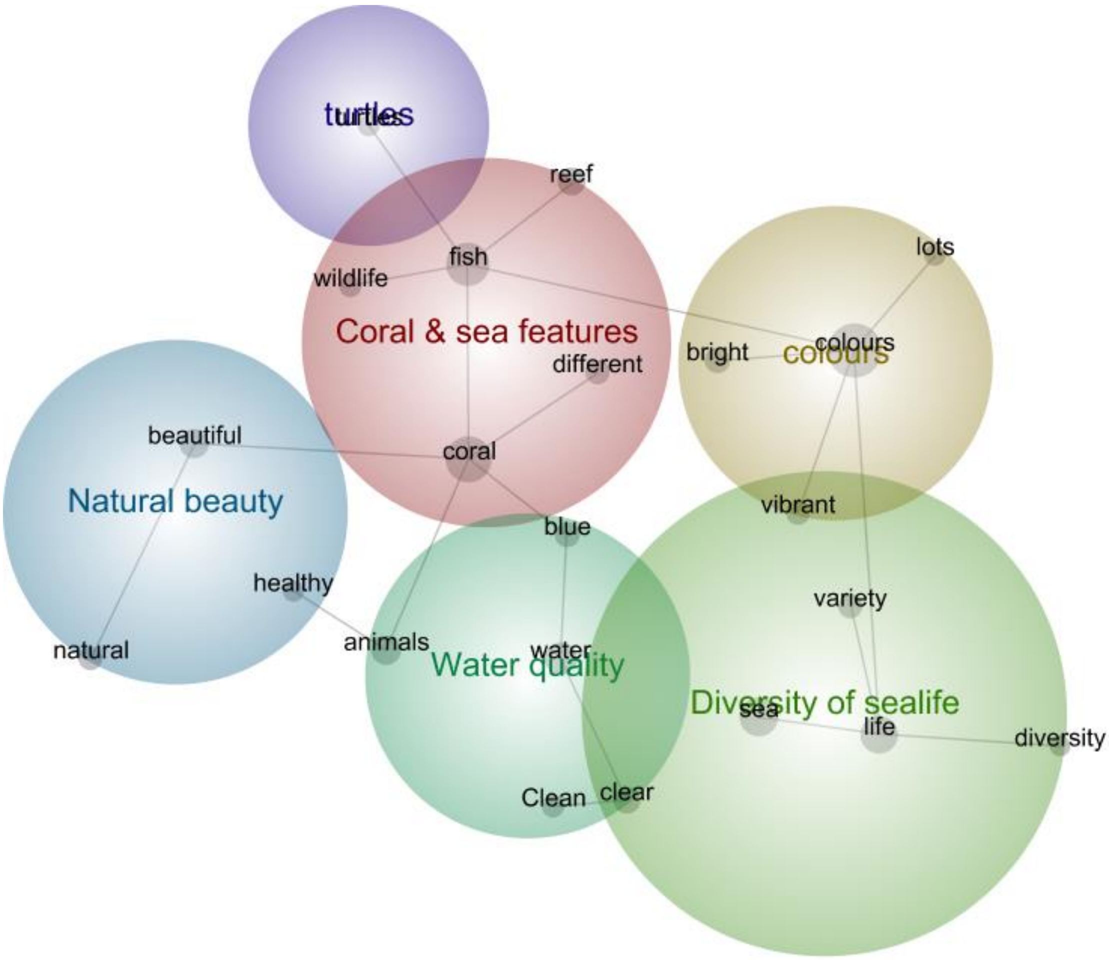


Figure 5: Thematic analysis of participants' beauty factors.

3.3.1 Clusters of beauty rating

Conjoint analysis was conducted for the whole samples (n=705) and provides the mean of importance values and the utility scores of beauty attributes. However, these importance values and utility scores vary among different age groups. This raises the question whether there are different consumer segments representing different types of visitor/user groups who perceive GBR beauty differently. Hence, cluster analysis was performed using the relative importance weights (i.e. importance values) of four beauty attributes (fish, coral, turtle and contrast). The four-cluster solution was found to have a meaningful interpretation (Figure 6). The importance value of the coral attribute was the key factor that distinguish participants of different clusters (i.e. groups), followed by fish (0.8), turtle (0.55) and contrast (0.25).

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	2	1	3	4
Label				
Description				
Size	43.8% (309)	20.1% (142)	19.1% (135)	16.9% (119)
Inputs	Coral 23.44	Coral 46.11	Coral 12.94	Coral 18.13
	Fish 32.06	Fish 20.99	Fish 47.79	Fish 27.27
	Turtle 16.86	Turtle 16.90	Turtle 20.47	Turtle 35.09
	Contrast 27.64	Contrast 16.00	Contrast 18.81	Contrast 19.51

Figure 6: Results of cluster analysis.

Based on our survey, four clusters (segments) of consumers were identified based on their differences in perceiving GBR beauty (i.e. the relative importance weights of beauty attributes). Cluster 2 was the biggest group containing 309 participants (43.8% of the sample). The rest of the samples were divided into three groups of similar size (from 16.9% to 20.1% of the samples) which give the highest important values to a specific beauty attribute: coral (cluster 1), fish (cluster 3) and turtle (cluster 4). This reflects the differences among users regarding how they perceive GBR beauty. Cluster 1 showed a strong preference for coral and was named "coral focus". Similarly cluster 3 was named "fish focus" and cluster 4 "turtle focus". The biggest cluster was named "Diversity focus" as the relative importance weights were similar values for the four beauty attributes (Figure 7).

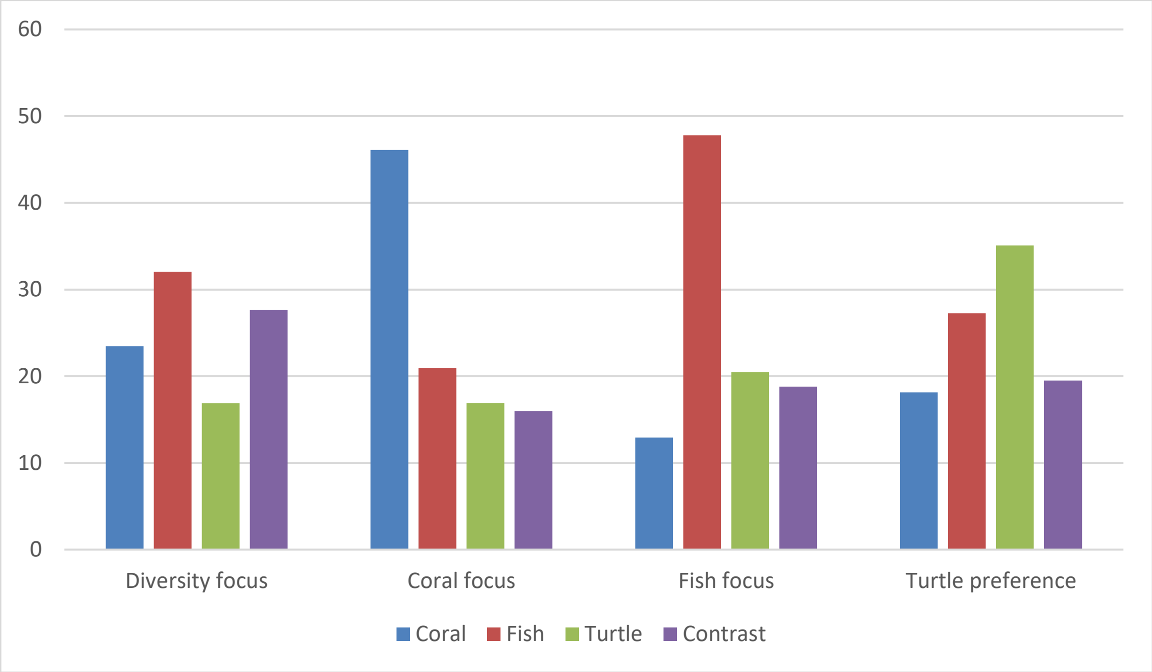


Figure 7: Relative Importance weights (i.e. importance values) of beauty attributes by segments.

Further analysis on these clusters is possible, for example, to understand better how the segments differ in their utility scores. Preliminary analysis reveals that the sign of the utility scores for contrast, fish and turtle attribute levels is the same across four segments, the utility score of non-vivid coral is reverse (positive versus negative). For the coral focus and turtle focus segments (36% of the sample), the negative utility means that the GBR pictures with unhealthy coral (i.e. non-vivid coral which has lost its colours) were rated as less beautiful. In contrast, the presence of unhealthy coral did not have a negative impact on perceived beauty for the diversity focus and coral focus groups (64% of the sample). More detail can be found in Scott et al. (2018 submitted).

3.3.2 Discussion and conclusion from the online survey

The conjoint analysis conducted on online survey data provides important insights into how visitors evaluate the beauty of GBR pictures. The results indicate that all four beauty attributes have significant impacts on visitors' perceived beauty ranking. On the basis of their relative important weights, fish is more important than other factors (i.e. coral, turtle and contrast) in determining perceived beauty (Table 4). More specifically, the diversity of sea life shown in a GBR picture increases its perceived beauty, in particular if the coral and fish have vivid colours. For example, a high ranked beautiful picture should include turtle, vivid fish and vivid coral. The presence of non-vivid fish has a negative impact while the presence of non-vivid coral has slightly positive impact on the overall picture beauty.

The thematic analysis of answers to open ended questions indicated that colours, coral and sea features (325), diversity of sea life (175), water quality (167) were important for ranking beauty. Water quality is of particular concern to Reef managers, and future research should systematically test the impact of different levels of water quality on the perceptions of attractiveness.

There is heterogeneity among the respondents in their beauty evaluation. Respondents' perceived beauty rank varies according to their age. The older the participants are, the more they rely on the fish attribute in evaluating GBR beauty (Table 5).

These findings are relevant to understanding: (1) how GBR beauty is subjectively perceived by visitors or Reef users, and (2) how the impacts of beauty attributes on perceived beauty vary among different segments. First, fish was the most important beauty attribute influencing perceive beauty of the four tested perceived beauty. Second, the impacts of intensity/impact level (e.g., fish, turtle, and contrast) and/or valence (e.g. coral) on perceived beauty vary among different tourist segments. Overall, GBR pictures used in tourism promotion which show a diversity of vivid elements will be rated highly on aesthetic value. Most tourists (49%) belong to diversity-focus segment and give high important values to all four beauty attributes. The coral-focus, fish-focus and turtle-focus segments each have less than 20% of the sample (see Figures 6 and 7).

These results indicate that the presence of unhealthy coral in GBR images (i.e. bleaching coral or dead coral that has lost its vivid colours) may not necessarily have a negative impact on perceived beauty.

4.0 FINDINGS STREAM 2: DEVELOPING MACHINE LEARNING ALGORITHMS

4.1 Introduction

The assessment of natural beauty is a difficult task, but as could be seen from the previous sections, it is possible to provide some kind of quantification of beauty or attractiveness. Asking people to rate images or scenic views provides important insights into subjective beauty, but it is a time consuming task and not practical as a long-term monitoring instrument. Instead, the use of computer-based systems may provide a viable alternative.

The development of algorithms to assess imagery is relatively new. In comparable research on terrestrial landscapes, Seresinhe, Preis and Moat (2017) extracted a large number of features (using the Places Convolutional Neural Network) from about 200,000 images of the Great Britain to examine the key attributes of beautiful outdoor spaces. For example, they found that built structures, such as towers or castles, increased the scenic value of an image. Trees also enhanced the perceived beauty, but less diverse features (e.g. green grass) were less attractive. Building on this type of research, as well as Haas et al.'s (2017) art-based assessment of attractive attributes of coral reef, this research uses neural networks to develop systems for measuring the beauty of the GBR.

Thus, this second research stream aimed to develop machine learning algorithms for two purposes applied to underwater imagery:
1. Automated identification of fish/species, and;
2. Automated assessment of aesthetic scores from an image.

For both purposes, the process involved the creation of training and testing datasets, the development and calibration of computer algorithms, and the testing of precision using independent validation datasets.

In the first part of this research stream, we developed a system to detect the so-called *region of interest* in an image, and then identify/classify those regions (i.e. through bounding boxes that capture the relevant object) into predefined classes (e.g., marine species). To develop a proof-of-concept system, it was decided that the region of interest was defined as different types of fish species that were prevalent in the ecosystems of the Great Barrier Reef or South-East Queensland, Australia. For both types of marine systems, video footage was available as training material (see further below for more details). Thus, the final output of this system was an automated assessment of a number of species of interest, and their frequency of occurrence in the imagery. This system is an important first step for automating the manual assessment or coding of underwater species done by marine scientists.

The second task is different from the first work on automated species identification. Here, we aimed to predict the aesthetic value of underwater images. Two different approaches for the aesthetic assessment are possible, and we conducted experiments using both.
1. Classification: An image can be categorized into either low or high aesthetic value, using a classification task and a machine learning approach.

2. Regression: The image is assessed by an actual score, whereby the system provides a score range (e.g. 1 to 10) for all the test images.

Two different deep learning approaches (architectures) were used to design and implement the research. Deep leaning is a type of machine learning technique that involves a large (multi-layered) Neural Network. In our work, convolutional neural networks were used to solve the above problems (More on Convolutional Networks can be found at http://www.deeplearningbook.org, chapter 9). More detail on the particular methods employed is presented in the following sections.

4.2 Automated detection of fish species in underwater imagery

4.2.1 Background

The use of underwater videos to assess diversity and abundance of fish is becoming more popular amongst marine biologists. It is a common practice for marine scientists to manually assess fish abundance using multiple underwater video cameras (Salman et al., 2016), but such a manual processing of video footage is labour intensive, and as a result very costly.
Instead, an automatic processing of videos can be employed to achieve the objectives in a cost and time-efficient way. Recent advancement in the science of deep neural networks has led to the development of automatic object detection and recognition. For example, our computer bases system can now detect and identify an object in five frames per second. This automatic detection can be applied to recognise fishes and other marine species, whilst dealing with challenges, such as distinguishing the fish from the background (e.g. ocean floor, plants, rocks, complex background, deformation, low resolution and light propagation). The particular detection task in our research is made more complex by the high levels of occlusion (due to schooling by fish), colour and texture of fishes.

4.2.2 Method

An end-to-end deep learning based architecture has been introduced to solve the identification task, and it was found that it outperformed existing state of the art methods. More specifically, our system involved the following. A Region Proposal Network (RPN) was introduced by an object detector termed as Faster R-CNN (Ren et al. 2015), and this was combined with three classification networks and fine-tuned for automatic detection and recognition of fish species from videos taken by Remote Underwater Video Stations (Salman et al., 2016). RPN is basically a small Convolutional Neural Networks (CNN), which takes a convoluted feature space and returns the detected bounding boxes of the *region of interest*. It can be combined with any state-of-the-art classification network to produce detection and classification simultaneously to enable an end-to-end architecture (i.e. fully automatic, see Figure 8 below). The Caffe (Jia at al., 2014) deep learning library was used to develop and conduct all the experiments presented in this report. Caffe library provides a framework with a set of open source library functions (mostly written in C++) and a python interface to design and develop a wide range of customised deep learning models. There are other and more recently introduced open source libraries, such as Tensorflow, but Caffe was one of the earlier libraries and was found to be more stable. We therefore adopted Caffe in our research.

In our experiments, the publicly available Caffe deep learning models for object detectors were used for initial weights and to enable a transfer learning technique. More specifically, to take advantage of all network architectures used in our experiments, a transfer learning technique from ImageNet (Russakovsky et al., 2015) was used during the fine-tuning of our models. We found that this technique achieved a better performance and a faster convergence. Transfer learning technology is very useful in our work because of the data limitations. A deep learning model needs a large number (millions) of images to be trained successfully. Several trained models were made available by other researchers who have undertaken similar research with adequate resources and infrastructure. Using these trained parameters helps at the early stage of the training to have faster convergence.

We combined three different CNN-based classification models with different sizes (small, medium and large) with the RPN network in our experiments to obtain Faster R-CNN models of different sizes (i.e. the number of layers). The proposed RPN model was combined with a classification model to achieve the detection and classification simultaneously. Three classification models were used in our experiments, namely ZF (Zeiler and Fergus, 2013), CNN-M (Chatfield et al., 2014), and VGG-16 (Simonyan & Zisserman, 2014), and their relative performances were compared.

Some details on the system and process are shown in Figure 8.

- The network takes an image as input, whereby the first few layers (represented by a box in Figure 8) are 'feed forward layers' that are used to obtain the Convolutional feature maps.
- The computed features map from the last convolutional layer were used by the RPN network to determine whether the object is present or not, if present the system computes all the bounding boxes and returns N numbers of bounding box co-ordinates.
- On convolution (256 channels), a 3x3 filter (i.e. a 3x3 matrix) filter slides across the map. At each sliding region, a 256 dimensional feature vector is produced and fully connected to a cls layer and a regression layer.
- The cls layer is a 2-class classification layer to determine the presence (or absence) of an object and the regression layer is a bounding box regressor to adjust the position of the bounding box. Anchors are employed to provide the network translation-invariant (each anchor represents a combination of different aspect ratios and scale).
- After receiving proposals, each proposed region on the convolutional feature map is passed into a region of interest pooling layer and the purpose is to get a fixed feature vector output for each region for the fully connected layer.
- Finally, the feature vector from the region of interest is linked to a 'fully connected layer' for the final classification.

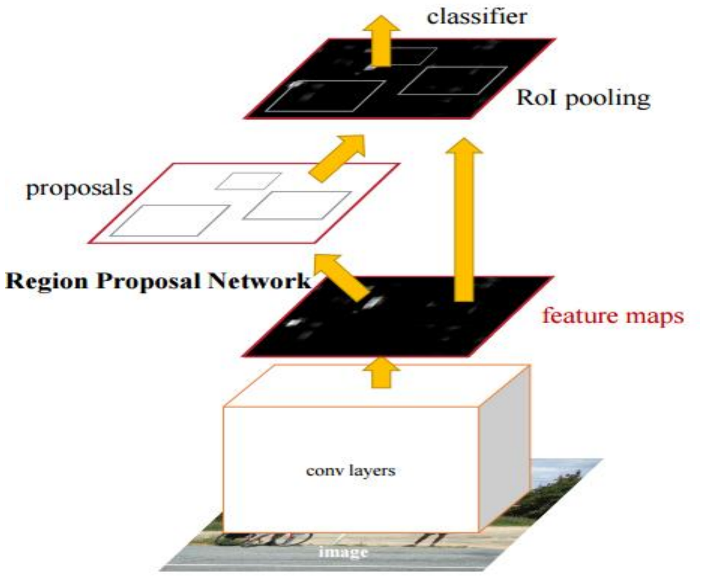


Figure 8: A single and unified deep-learning framework for marine species detection and classification. Convolution layers at the beginning extract features from an image, followed by a Region Proposal Network, which finds the object boundary. Finally, extracted boundaries are classified using classification layers of the network.

4.2.2 Results and discussion – automated species recognition

The above system was trained to detect and identify up to 50 different species available in our datasets. Underwater raw videos footages used in our experiments were provided by the University of Sunshine Coast (Borland et al. 2017; Gilby et al., 2017), and additional videos were recorded specifically for this project.

The dataset used in our experiments contained 4,909 images with 12,365 annotated samples of 49 species of fish and crustaceans plus a turtle. The Vatic interactive video annotation tool (Vondrick et al., 2013) was employed to annotate the data. It was standardized in a PASCAL VOC format (Everingham et al., 2015). The dataset was divided into training, validation, and test sets using a random sampling technique. This division of data is a standard practice for machine learning-based approaches. The training, validation and test set comprised 70%, 10%, and 20% of a dataset, respectively. It is important to note that the data to train the system must be in image form and image frames had to be extracted from the videos.

In the training phase, "gold standard" data are normally used to train the model, by pairing the input with the expected output. In the validation phase, additional data are used. The purpose is to estimate how well the model has been trained and to estimate model properties (e.g. mean error for numeric predictors, classification errors, recall, and precision for IR-models etc.). Finally, test data were used for calculating the test performance of the system. Figure 9 provides examples of input and output images.

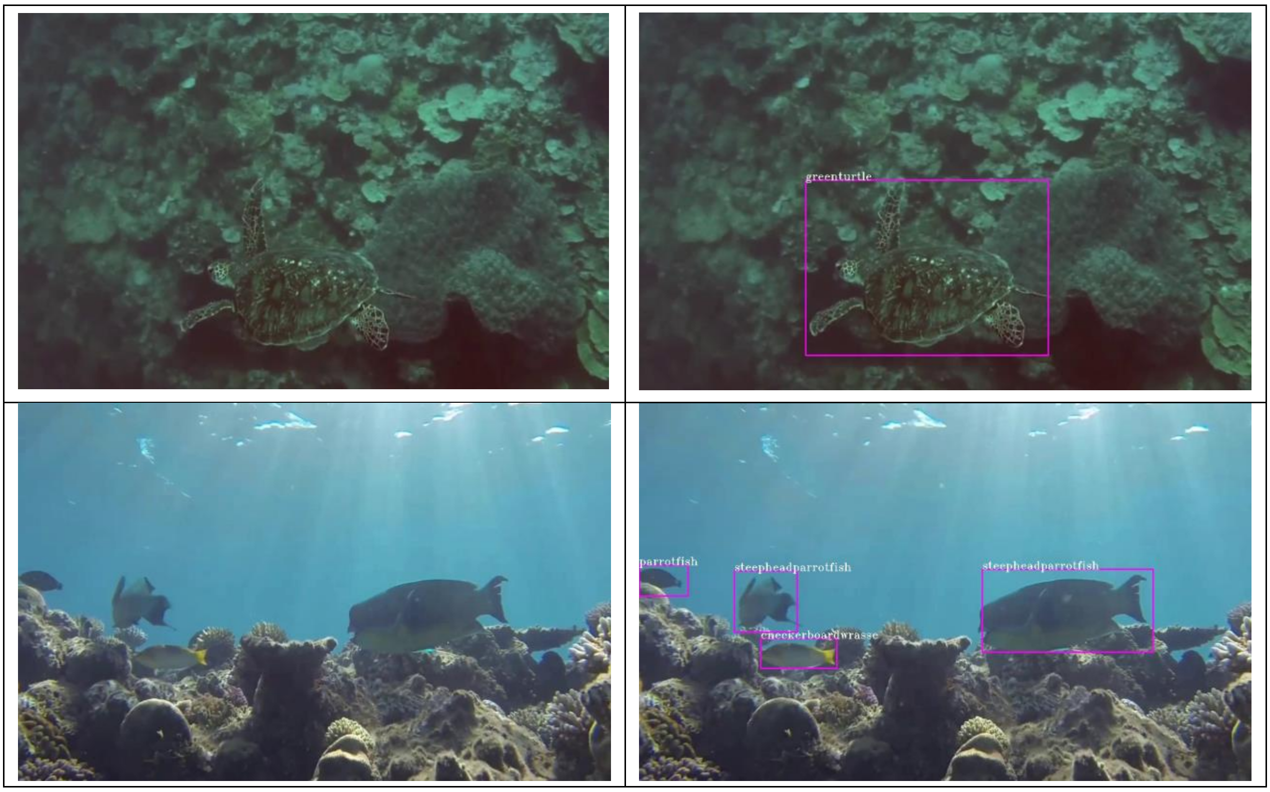


Figure 9: Example of object detection on the GBR dataset. Images on the left are input images and images on the right are output images, after detection. The system can detect multiple species in a single frame.

The detection results of several fish species from two sets of experiments are detailed in Tables 7 and 8. More specifically, the mean Average Precision (mAP) summarizes rankings from multiple queries by averaging the precision. Table 7 shows the eight different fish species from the GBR dataset and precision results.

Table 7: Results from the experiment on the detection dataset of eight GBR species

Species	**Avg Precision**	**Species**	**Avg Precision**
Dark surgeon	1*	**Green Sea turtle**	1
Damselfish	0.8749	**Checkerboard wrasse**	0.9091
Lemon damsel	1	**Steephead parrotfish**	1
Common cleanerfish	1	**Coraltrout**	1

Note*: 1 refers to 100% precision accuracy rate, which can be explained by the small sample (N=8), the nature of the species (with highly discriminatory features) and the fact that for these species the training and test data sets were prepared from the same videos (although different sample frames were selected randomly for training and testing).

Table 8 presents the results obtained from the experiment on whole dataset, namely 50 species South East Queensland surf fish species and eight for the GBR dataset (the ID of a turtle is common to both datasets). To provide more detail, the mAP of the ten most frequent species is reported as well. Table 8 also presents the results obtained from using three network architectures considered in our experiments. Compared with the two other models, the VGG-16 network (Simonyan & Zisserman, 2014) outperformed in terms of precision. For the VGG-16 network a mAP of 82.4% was achieved, and the average time taken for processing an image for detection during testing process was 0.2 seconds (i.e. 5 frames per second). For the

ZF (Zeiler & Fergus, 2013) and CNN-M (Chatfield et al., 2014) models the processing time was 0.1 seconds (i.e. 10 frames per second) for network models, which implies that these systems are capable of processing videos in real-time scenario. The VGG-16 CNN had also performed well in related research on the aesthetics of landscapes (Seresinhe et al., 2017).

Table 8: Results from the experiment on the fish species datasets, comparing the three different models. AP = Average Precision, mAP = mean Average Precision.

Species	AP on VGG16	AP on CNN-M	Ap on ZF
Blue spotted flathead	1	0.818	0.831
Sand whiting	1	0.945	0.909
Smooth golden toadfish	1	1	1
Southern herring	1	0.947	0.867
Smooth nose Wedge fish	0.996	0.989	0.892
Painted grinner	0.986	0.951	0.972
Reticulate whip ray	0.909	0.909	0.909
Starry pufferfish	0.909	0.899	0.807
Swallowtail dart	0.909	0.838	0.994
Common stingray	0.906	0.995	0.97
mAP (whole sample)	0.824	0.769	0.75

The key to precise detection is the extent of training. The training process helps the network model to learn relevant parameters so that it performs better when assessing test data to produce a desired output. During the training phase, a batch of training images are fed into the neural network model for learning and this process continues as long as the learning error rate is in decreasing mode.

In neural networks, the processing of a batch of images is referred to as iteration. To observe this process the accuracy needs to be plotted on a graph, which then helps to understand the behaviour of the learning process. The plot also provides information on the point that gives best results. Figure 10 below shows such a graph and how the mAP improves with an increasing number of iterations during the training process. The Figure also provides insights into the relative performance of the three different network architectures over time (see also Table 8 above for class wise performance). The highest precision measured through mAP was reached after 90,000 iterations for the VGG16 model.

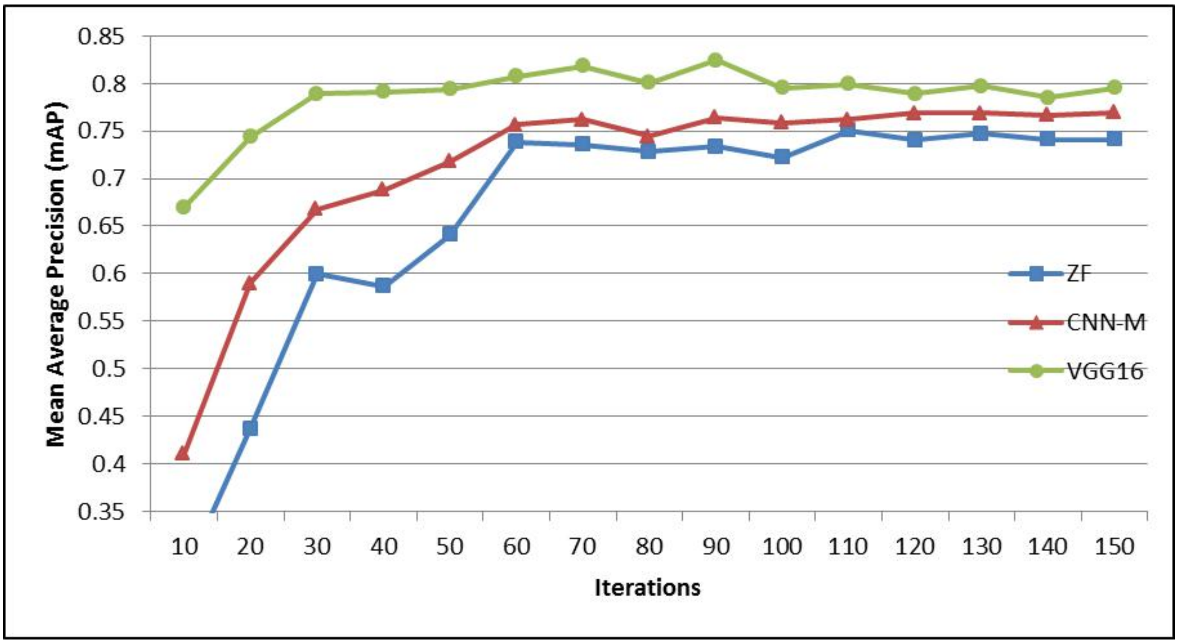


Figure 10: Mean Average Precision on test data using three different models. The x-axis represents iteration in thousands.

4.3 Neural-network based beauty rating of the Great Barrier Reef images

4.3.1 Overview

This part of the project addressed the urgent need to monitor the aesthetic value of the GBR, and it built directly on the eye tracking experiments reported in the findings of research stream 1 earlier in this report. An artificial intelligence-based technique (Large Neural Network model) was used to assess a large number of photos and assign a score of aesthetic value to each image.

Broadly, the task involved to distinguish computationally the aesthetic attributes of an image (see also Haas et al., 2015 for a related assumption). The literature proposes several methods to solve such challenging classification and scoring problems. The earlier approaches can be categorised into two groups, based on visual feature types (hand-crafted features, and deep features based on Convolutional Neural Network). More specifically, the term "hand-crafted" feature refers to properties derived employing various algorithms using the information present in an image. As an example, edges and corners are two simple features that can be extracted from images. A basic edge detector algorithm works by finding areas where the image intensity "suddenly" changes. For example, the shell of a turtle can be identified as an edge. Likewise, the so-called Histogram of Gradients (HoG) (Dalal & Triggs, 2005) is another type of handcrafted feature that can be applied in many different ways. In contrast, Convolutional Neural Network (CNN) based features are learning from the training samples, and they do this by using dimensionality reduction and convolutional filters. We used CNN features, and the procedure initially required to provide NxN (the value of N can vary from such as 1, 3, 7 etc.) matrix filters, which are filled with random real values. The filter values change during the training phase by observing the training image samples that are used as an input.

4.3.2 Method

Deep Convolutional Neural Network based methods have shown promising results for image aesthetics assessment in recent works (Deng et al., 2016; Seresinhe et al., 2017). An Inception module, for example, is a small neural network that has recently delivered very high performance in object classification (Szegedy et al., 2014). However, the Inception module has not been used for the research problem of assessing aesthetics in an image. We propose a novel Deep CNN structure for image aesthetics assessment consisting of three Inception modules and connect intermediate local layers to the global layer using a concatenation process.

More specifically, a spatial pyramid pooling was employed for the final feature computation. A more detailed description of the architecture and different modules is provided below. The network was first trained and then fine-tuned using a large-scale aesthetics assessment AVA dataset (Murray et al., 2012). The AVA dataset has 250,000 images, which is very useful for training such a large deep neural network model.

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.

<u>Inception module</u>
The main idea behind the Inception architecture (see also Szegedy et al., 2014) is to consider how an optimal local sparse[1] structure of a convolutional vision network can be approximated and covered by readily available dense components[2]. The Inception architecture is restricted to pre-determined filter sizes to avoid patch-alignment issues.

The Inception architecture used in this research follows a multi-step process (Figure 11). The Inception module takes an image as input, and applies three different filters of sizes 1x1, 3x3 and 5x5 for convolution in four different channels. Finally, the filters are concatenated into a single channel. For example, if an inception module takes an image of size 28x28 pixels, it returns a 3-dimensional matrix of 256x28x28. All the boxes shown in Figure 11 describe building blocks of Neural Networks, and by combining these boxes this module has been created producing certain features.

[1] Local sparse refers to the pixels representation in a small region of an image that has no pattern.
[2] Dense components means that the sparse values can be stored into a smaller dimension matrix. For example, for an image of 224x224 pixels, the goal might be to reduce the dimension, but without losing any vital information.

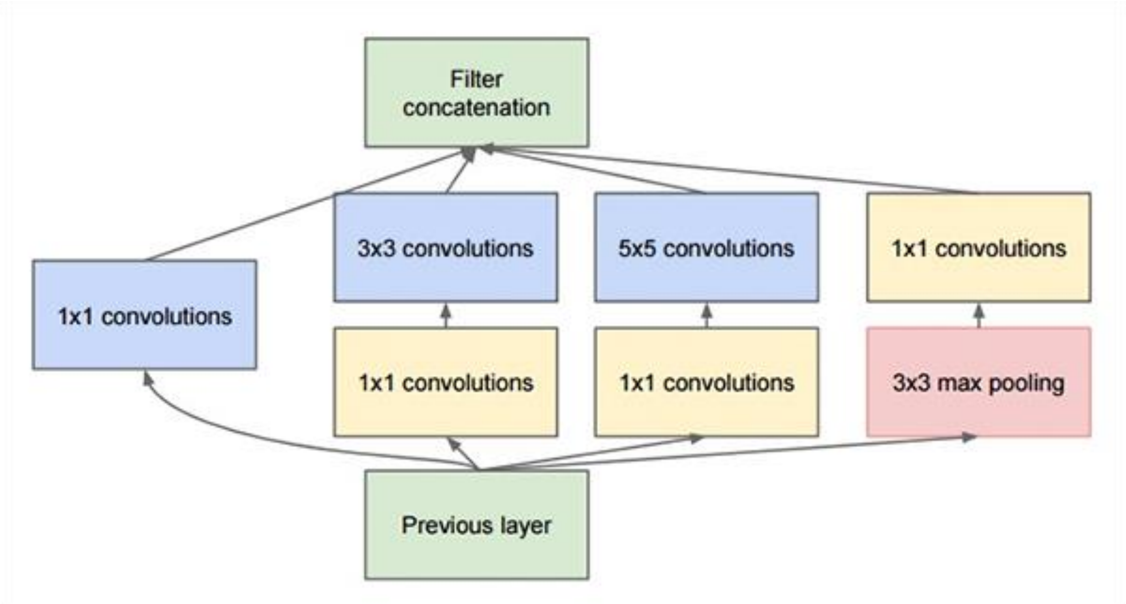


Figure 11: Architecture of Inception module (Source: after Szegedy et al. 2014).

Spatial Pyramid Spooling (SPP) layer

To remove the size constraint (image size) in our network, we place a Spatial Pyramid Pooling (SPP) layer (instead of the fully connected layer) on top of the last convolutional layer. Traditional Neural Networks only accept an image of fixed sizes (e.g., 227x227 or 256x256), because the node size of the fully connected layer in the Neural Network needs to be predefined. The SPP layer pools the features and generates a fixed length output, which finally feeds into the classifiers. It partitions the image into coarser levels and aggregates the local features in them. SPP is able to generate a fixed length feature regardless of the input size. This property helps the architecture to preserve the image composition (i.e. different to CNN architectures, resizing, scaling or cropping are not required). Moreover, uses of random size images during training reduce the chances of overfitting the network. Figure 12 shows how the spatial pooling layer works on multi-task convolutional layers, and Figure 13 illustrates the pooling process with an example.

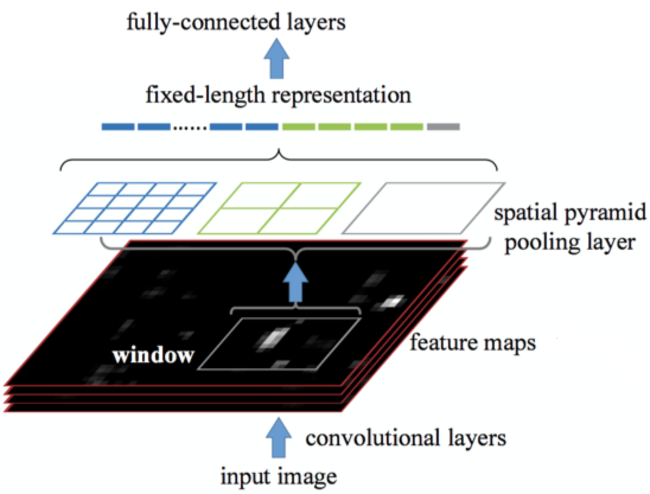


Figure 12: Spatial pooling layer to create the final feature vector in our network.

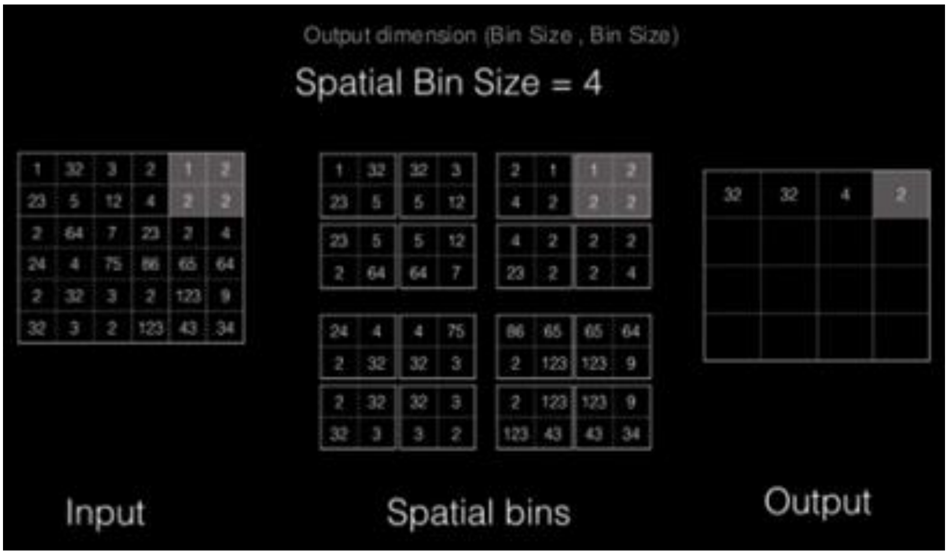


Figure 13: Example of Spatial Pyramid Pooling where the resultant matrix will always be 4x4 regardless of the input matrix size. First, the image is divided into 4 spatial bins and each bin is again divided into 4 spatial bins. Finally, from each 16 (4x4) bins the maximum value is chosen for the final matrix.

<u>The aesthetics score prediction deep learning architecture</u>

Our model has used three inception modules (Inception_a, Inception_b, and Inception_c[3], marked by yellow colour in Figure 14 below). The first and the second inception module are considered to extract local image features, and the final inception layer is then used to extract global image features after one step of average pooling and another step of spatial pyramid pooling.

All the convolution layers and including layers in the sub-module (i.e. Inception module), use rectified linear activation. Three different output vectors from all the three inception modules are finally concatenated after the Spatial Pyramid Pooling. The channels of Inception_a and Inception_b produce 256 sizes feature vectors each as local features, and Inception_c channel produces a vector of sizes 512 as global features. Outputs obtained from these channels are finally passed through a concatenation layer to form 1024 dimension features.

This concatenated layer is followed by a fully connected layer with the same dimension to produce the final feature vector. The final feature vector is fed into a softmax with a loss layer to obtain a binary output, which indicates low or high aesthetic quality of an image in our classification task. To produce a beauty score in another experiment we replaced the softmax layer with a Euclidean loss layer.

[3] For our model we have used the recently published Inception module from GoogleLeNet model as s submodule.

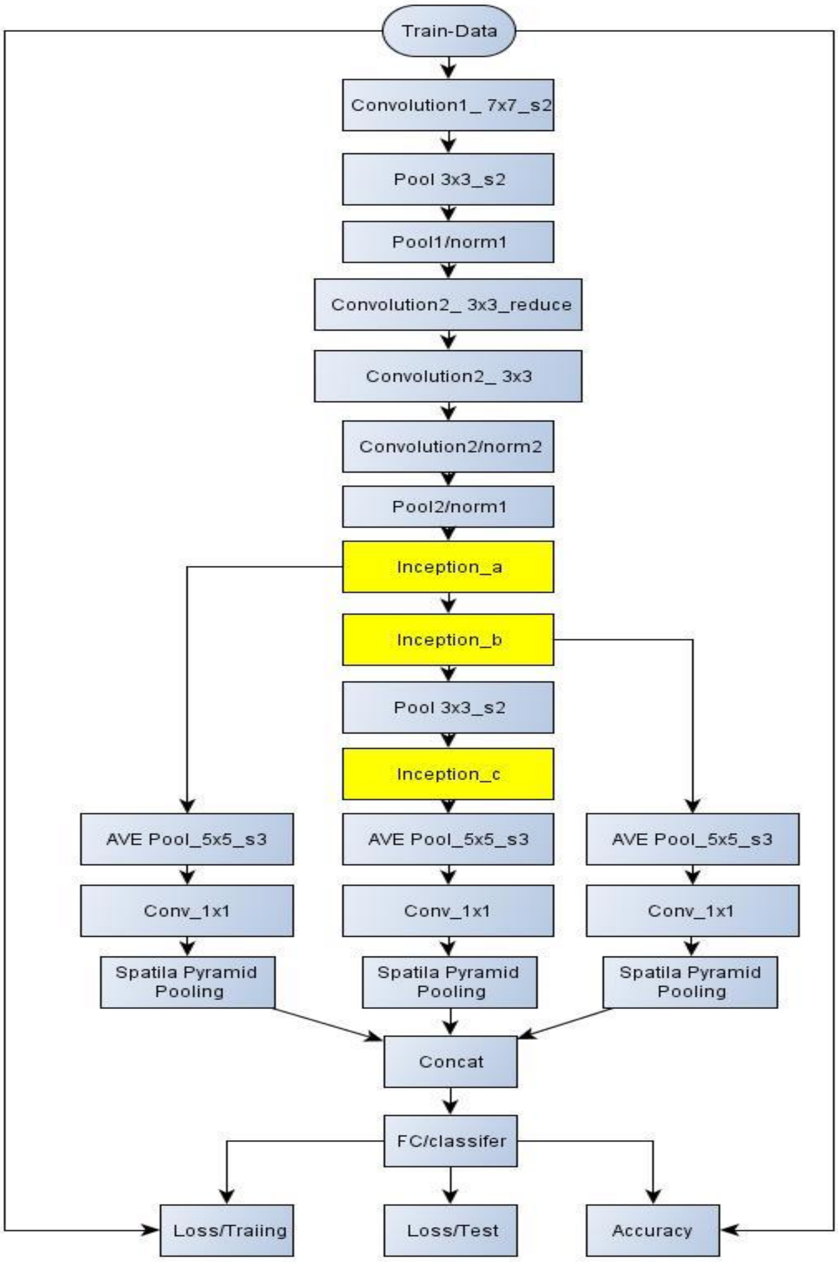


Figure 14: Inception (Szegedy et al. 2014) model-based image aesthetic classification (or score) network architecture.

<u>Datasets</u>

For experimental purposes, both publicly available datasets and dataset developed in-house were used. For the dataset specific to the Great Barrier Reef (i.e. the GBR dataset), we used 2,500 underwater GBR images, which were downloaded from the Flickr social media platform (see earlier in this report). These images were sorted based on the content, and then rated by participants in an online survey for their aesthetic beauty. At least 10 survey participants provided an aesthetic score for each image and the mean score was calculated.

Most of the images (i.e. 80%) served as training material to enable the proposed Neural Network model to learn key feature parameters. The remaining 20% of data were used during the test and validation phases. The validation dataset helps to understand the system performance in terms of accuracy during the training phase, whereas the test set is normally used once the training phase is completed and ready for deployment.

To better understand the distribution of 'beautiful' and 'ugly' pictures in the dataset, Figure 15 presents the number of images with scores above or equal/below 5. More than 2,080 images were scored as highly aesthetic (score >5) and only 420 (score <=5) images were scored as having low aesthetics. Figure 15 also shows how many images of high and low scores were used in each experimental stage.

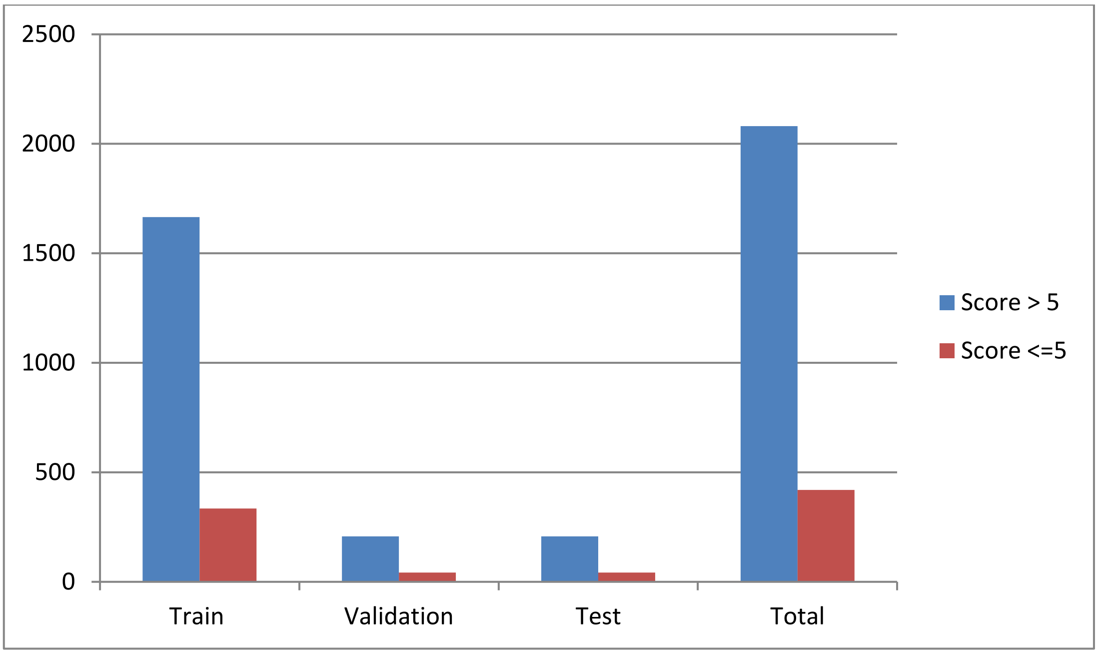


Figure 15: Score distribution of the GBR dataset, containing a total of 2,500 images.

It is important to note that the GBR dataset of 2,500 images 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 (AVA, see Murry et al., 2012). This helped to train the system and allowed us to use the in-house GBR dataset for fine-tuning the algorithm. The detailed dataset description of the AVA is given below.

- AVA1: 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 value classes. By doing this, we obtained 74,056 images in the low aesthetic value class, and 181,447 images in the high aesthetic value class. Approximately 10% (25,549 of 229,954) of images were used for testing the system performance.
- AVA2: In a different experimental setup, and to increase the gap between images with high aesthetic and low aesthetic value, all images were sorted based on their mean scores. Then, the top 10% of images were considered as highly aesthetic and the bottom 10% of the total images were classed as low aesthetic. Thus, 51,100 images (approximately, 20% of the full dataset) then formed the AVA dataset that was used for training.

4.3.3 Results – neural network based beauty scoring

This section presents the results obtained from two different approaches, namely classification and regression. Figure 16 shows that the highest accuracy was obtained at 25 epochs[4] of 84.8% on the GBR dataset. Here epoch means one forward pass and one backward pass of all the training examples available during the training phase. More specifically, during the training process, we need to feed all the training images into the system to update the Neural Network training weights (i.e. the parameters), and this process repeats multiple times to obtain a maximum accuracy. In our experiment, we recorded accuracies in testing (green curve in Figure 16), training loss (orange curve) and validation loss (red curve) after each epoch.

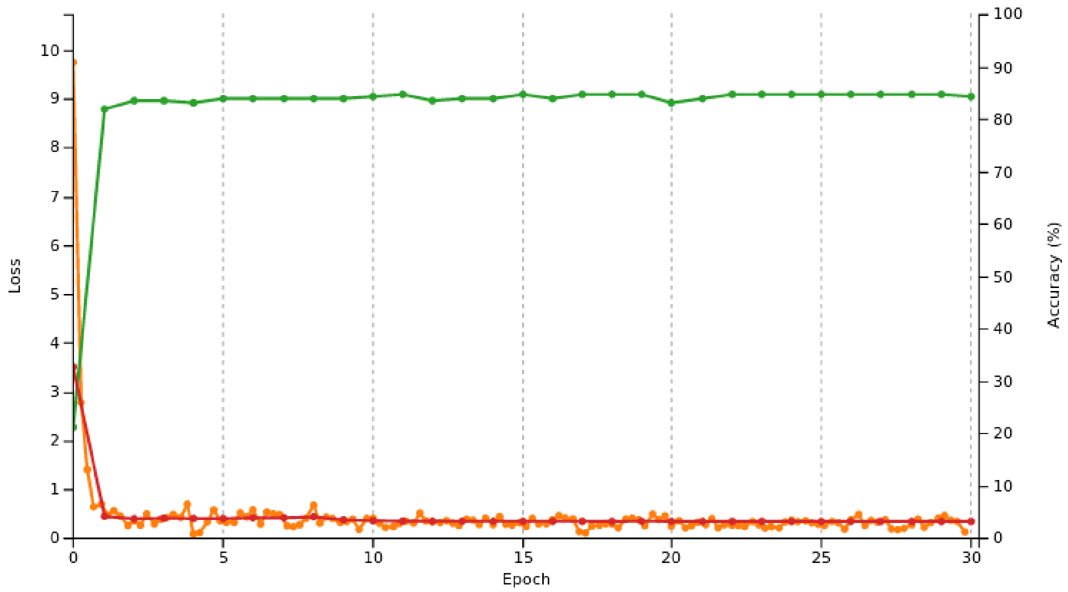
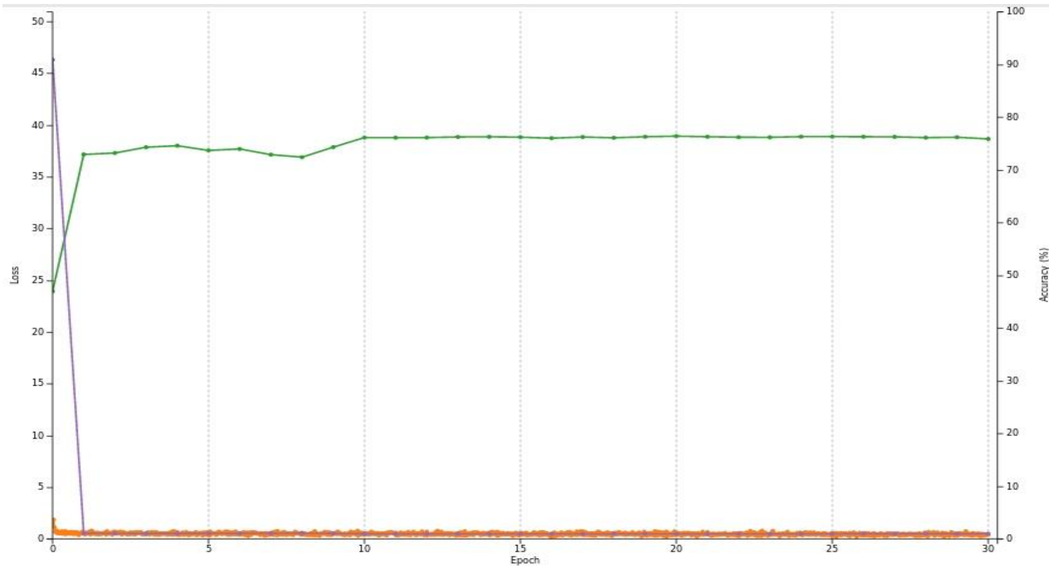


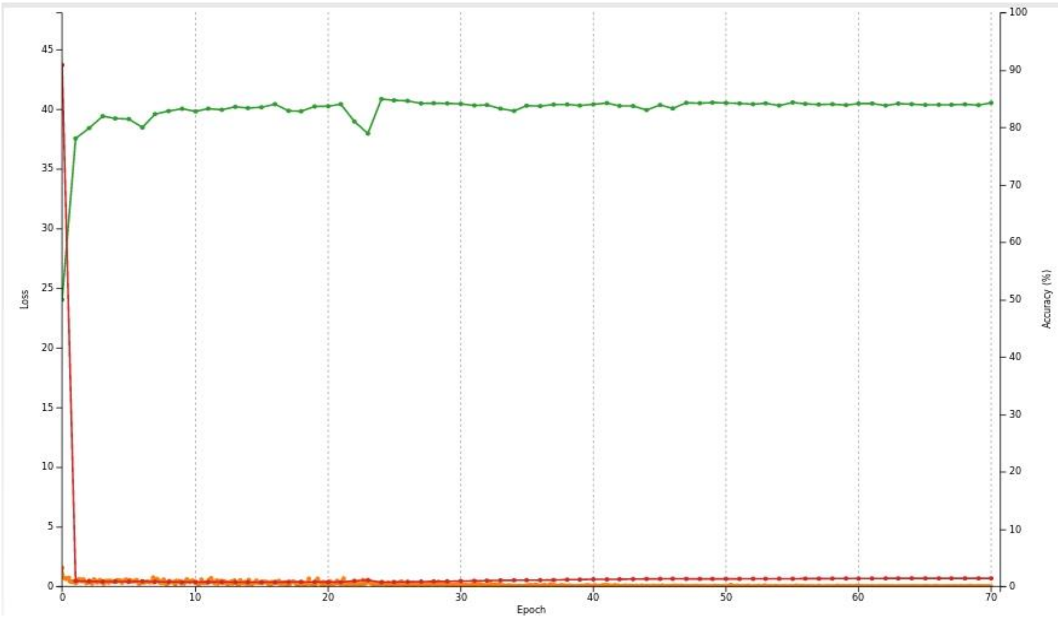
Figure 16: Accuracy achieved over epochs obtained from large scale GBR image in the analysis of aesthetic value.

The accuracy was also tested for the publicly available (AVA) training sets (Murry et al.2012) that were used to enhance our GBR in-house data set. Figures 17 (a) and 17 (b) show the same measures of the system performance, as already presented in Figure 16 for the GBR data. In the AVA1, accuracy of 76.1% was achieved on the full dataset (with little loss), and 84.5% accuracy was achieved for the AVA2 dataset when the training epoch was 30.

[4] Epoch is a term related to Neural Network training. For example, if there is a total of 100 training samples, then the process involves feeding these training data repeatedly into the system. In our experiment we have provided data 30 times, which means 30 epochs.



(a)



(b)

Figure 17: Results obtained for large scale image aesthetic analysis on publicly available AVA dataset. The green curves represent accuracy.

In our regression-based image aesthetic analysis, we computed the aesthetic scores for all the test images. The following tables show some successful (Table 9) and less successful (Table 10) sample score prediction results on the GBR images, along with their surveyed score and the predicted scores. The images that are ranked differently by people versus the machine are of particular interest and indicate a need for more training. Image Q2_887_1.jpg showing a turtle, for example, is perceived as beautiful when rated through the human eye, but less attractive when scored by the machine. The greater weight of the 'charismatic species' of a turtle may have compensated for less vivid colours and limited contrast.

Table 9: Comparison of the aesthetic score results for selected images obtained from the survey and the machine

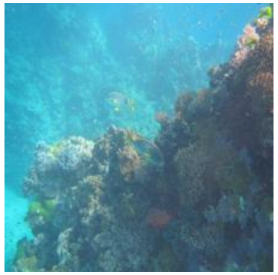
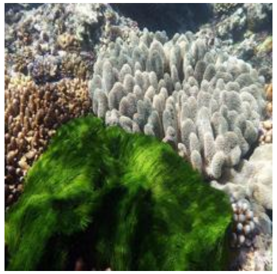
Image Name	 Q1043_1.jpg	 Q718_1.jpg	 Q1125_1.jpg	 Q1224_1.jpg
Score obtained from survey	5.83	5.00	6.83	4.50
Score obtained by machine	5.83	4.99	6.83	4.51

Table 10: Sample images that illustrate a gap between the aesthetic score assigned by participants in a survey and by the machine

Image Name	 Q527_1.jpg	 Q2_883_1.jpg	 Q2_887_1.jpg	 Q2_971_1.jpg
Score obtained from survey	8.4	8.08	7.73	8.72
Score obtained by machine	5.36	5.04	4.63	5.56

In our GBR in-house dataset, 500 images were used to test the machine based scoring system. The proposed system assign one score to each image in a range between 1 and10. We found that 86.6% of images were correctly predicted when we set the threshold to 1.5 (Euclidian distance). The overall performance based on different threshold is presented in Table 11.

Table 11: Overall performance of the image aesthetic assessment on GBR test dataset

Threshold (Euclidean distance)	**1.5**	**1.4**	**1.3**	**1.2**	**1.1**	**<=1**
Number of Samples	433	415	372	372	342	316
Percentage Accuracy	86.6	83	74.4	74.4	68.4	63.2

4.4 Discussion of the computer-based identification and beauty estimation findings

This research delivered a proof-of-concept of an automated species identification system, and an aesthetic assessment model.

The first part of this research focused on fish species detection and recognition. This was deemed important for several reasons. First, it is known that key species influence the value of the Reef experience and the subjective rating of attractiveness. It is not known at this point what these species are, but building on the findings in research Stream 1 it is clear that vivid-coloured fishes are important. Other species, such as turtles, ‘nemo’ (anemone fish), dolphins, and rays (see also for a social media analysis in Becken et al., 2017) are likely to also play an important role in the overall aesthetic value of the Reef. Second, the ability to recognise keystone or indicator species – either for reasons of monitoring aesthetic value or ecological states/changes – is of great importance to Reef managers. Developing automated systems that can identify species in underwater images (photo or video), in real-time, would result in substantial cost savings for environmental monitoring.

The results above indicate that the performance of both the systems is impressive in terms of accuracy. However, there is scope for improvement. For the fish identification system, it is important to extend the system to further species of interest. Here, input from marine managers is needed to ensure that relevant indicator species are captured. The system can also be extended to other ‘features’ or ‘regions of interest’. Extension only depends on the supply of further training material that shows the object of interest (from different angles and under different conditions, such as light) to train the machine and test the system.

Further development is also possible in terms of the output that the system provides. At present, the system can count the number of fishes from an underwater video and identify what they are (as long as they were part of the training set). However, the actual size (height and width) cannot be determined at this point. A more sophisticated stereo camera would be necessary to capture the information that would help to deliver such output variables. Significant modifications to the algorithm would be necessary to extract this information. For certain use cases, such as the monitoring of fisheries, this might be a worthwhile task.

Overall, and before fully implementing the system presented in this report, it will be necessary to create a larger dataset to improve the training and learning process. A larger dataset would result in a more robust system with better detection and recognition accuracy. Essentially, such an expansion is a matter of scaling up, rather than the development of new algorithms.

The assessment of aesthetic value in images will also benefit from a larger image library for training and testing. In our research we used a total of 2,500 images. Whilst the pool of relevant data from Flickr or other sources is much larger, the training involves manual sorting of images and scoring by different people. The requirement in our research was that a minimum number of ten respondents in an online survey had to score an image. The main limitation here was cost of the survey, and with additional resources it is possible to improve the model with a greater number of rated photographs.

Furthermore, it emerged that our dataset was unbalanced as the majority of pictures were rated with high beauty scores. This is possibly not surprising, given that images posted on Flickr are naturally biased towards more attractive photographs that reflect positive experiences. This positive bias has several implications. First, in terms of training of the model, we would have to specifically collect non-pleasant images (this can be achieved). Second, however, it may raise questions about the use of online-supplied imagery by Reef visitors. Do people tend to post beautiful images, and will this make it difficult to determine declines in aesthetic value? Or will this positive bias be consistent across time, so that people always post 'relatively more beautiful' pictures, but the algorithm will recognise that there is a trend of absolute decline against which these postings occur? These are important questions that should be discussed with Reef managers and explored with more specific social science research. A longitudinal assessment of comparing historic images with present ones might also give insights into the (changing or constant) effect of bias.

Finally, the development of the Neural Network based method using survey supplied beauty scores has not taken into consideration the subjective nature of scores and the fact that people from different demographic groups (see also findings in research stream 1) might perceive images differently. Future research could use multivariate regression analysis to explore possible effects and improve the accuracy of the model.

5.0 OVERALL CONCLUSION AND RECOMMENDATIONS

Aesthetic value is an important criterion of the Great Barrier Reef and its World Heritage listing. To date, there is no systematic and consistent methodology on how to assess the 'natural beauty or aesthetic value' (as articulated as part of Criterion vii in the statements of Outstanding Universal Value). Indeed, Johnston et al. (2013) noted that World Heritage arguments are usually led by the other criteria, with the aesthetic attributes relying on rhetorical descriptions of visual aspects of the site, for example, the turquoise waters and scattered islands in the case of the GBR. Johnston et al., therefore, reinforced the importance of finding an approach to assess and monitor the aesthetic value of the GBR at a variety of scales, including underwater, at water level and panoramic.

This research focused on underwater beauty and responds directly to the need for including some form of measurement into the evolving integrated Reef monitoring system (see Addison et al., 2015), as being developed through the Reef 2050 Integrated Monitoring and Reporting Program (RIMReP[5]). One possible source of data for monitoring aesthetic value is the large volume of images supplied by visitors to the Reef. People who spend time at the GBR for a range of reasons (tourism, recreation, work) often post images online to share their experiences.

Capitalising on freely available imagery has become a recent trend in environmentally focused research. In particular, the use of Flickr images has become prominent. As one of the most popular photo sharing sites, Flickr offers considerable potential for a range of analyses, especially since a substantial share of Flickr images also contains location and time stamps. Keeler et al. (2015), for example, used Flickr data to estimate visitation to natural areas in North America, and The Nature Conservancy (2017) drew on Flickr images as one data source to estimate the economic impact of coral reefs tourism. Seresinhe et al. (2017) used Flickr photographs to examine the scenic value of landscapes in Great Britain, using similar methods as developed for this present research on the GBR.

Thus, drawing on online content – or citizen supplied data – could provide a long-term and sustainable avenue for suppling raw data for aesthetic assessments of natural sites. These images can be complemented by more structured data, such as those supplied to GBRMPA's Eye on the Reef program, or photos/videos specifically taken for monitoring purposes by marine scientists.

This research brought together two research stream, that each are essential for the development of a large scale monitoring system. First, we explored what constitutes aesthetic value and what the key attributes might be. As suggested by Johnson and Smit (2014) an experiential lens was taken and using eye tracking technology and survey data, key attributes could be identified. This was then followed by proof-of-concept research to develop computer-

---

[5] The Reef 2050 Integrated Monitoring and Reporting Program is an integral part of the Reef 2050 Plan. It covers all aspects of the GBR's environment: natural and physical attributes, heritage values and its social, economic and cultural aspects. For more information, see: http://www.gbrmpa.gov.au/managing-the-reef/reef-integrated-monitoring-and-reporting-program

based algorithms that help to identify the key species of interest and also to score (at large scale) the aesthetic value of any given GBR underwater image.
Several important findings emerged. First, the results provide a strong evidence that eye tracking measures can be useful in studying natural beauty. In line with the underlying theory (Arriaza et al., 2004), we provide empirical evidence that attractiveness and beauty attract attention. Given the significant correlation between eye tracking measures and picture beauty scores, we suggest that this method provides an objective measurement of the relative beauty of GBR images. The eye tracking also allowed us to test the impact of adding specific attributes that might enhance aesthetic value, for example, a turtle over bleached coral. More testing of this kind could be undertaken to test a range of scenarios (e.g., underwater structures), including a decline of the environmental quality of the Reef or presence/absence of charismatic species.

Second, a survey of Australian residents highlighted the relative importance of four key attributes, namely contrast, fish, coral and turtle. The findings revealed that the diversity of sea life apparent in an underwater photograph of the GBR increases its perceived beauty, in particular if the coral and fish have vivid colours. Maximum attractiveness could be reached by showing a diverse picture that includes a turtle, vivid fish and vivid coral. Interestingly, the presence of non-vivid fish resulted in a negative impact while the presence of non-vivid coral had a slightly positive impact on the overall picture beauty. This means that the presence of bleached or dead coral in GBR images may not necessarily have a negative impact on perceived beauty.

The perception of beauty is a subjective matter, and it is therefore not surprising that the beauty scores differ for different types of consumers. The results of this research reveal, for example, older participants rank the fish attribute higher when evaluating GBR beauty. Also, participants over 45 years old rated a GBR picture with non-vivid coral as less beautiful, possibly because they are more aware of the environmental implications of a bleached coral. Respondents with limited experiences of visiting the GBR (once or twice) relied more on the coral attribute when evaluating GBR beauty than tourists with greater GBR experience or those without experience. Finally, and based on the analysis of answers to open ended question, respondents noted that beauty ranking would depend on colours, coral and sea features (325), diversity of sea life (175), and water quality (167). Water quality, in particular, should be included in future assessment of aesthetic.

The second stream of the research, using different types of Neural Networks to develop machine-based solutions for species identification and aesthetic assessment, delivered promising results. Detection and assessment scores were high, despite some obvious limitations, for example, as a result of a limited number of training data. Ways for improving the systems and models have been outlined above, and we are confident that the proof-of-concept presented in this report can – with additional resources – serve as the basis for the future implementation of a complete automated monitoring system.

Based on the findings from this research we make the following recommendations:

1. Discuss findings with key stakeholders and Reef managers to obtain feedback on the usefulness and viability of the approaches for monitoring aesthetic value, as presented in this report. Obtain feedback on priorities for monitoring and gaps in our current systems.

2. Undertake further testing, using a combination of eye tracking and self-report, to understand better the aesthetic preferences of key user groups, including Traditional Owners.
3. Identify key species for monitoring of aesthetic value and assess whether these also align with other classifications of ecological indicator species. Generate more data material (e.g. videos) to train and test the proposed system for species recognition.
4. Include other key factors in the beauty assessment, including water quality, but also man-made structures that could erode aesthetic value.
5. Extend the methods developed for this research to assess above water aesthetic value of the GBR World Heritage site.
6. Link the measurement of aesthetic value with visitor satisfaction (e.g., through monitoring sentiment expressed in social media) to understand the relationship between potentially declining 'beauty' and tourism value.

REFERENCES

Addison, P., Walshe, T., Sweatman, H., Jonker, M., Anthony, K., MacNeil, A., Thompson, A. & Logan, M. (2015). *Towards an integrated monitoring program: Identifying indicators and existing monitoring programs to effectively evaluate the Long Term Sustainability Plan.* Townsville, Queensland: Australian Institute of Marine Science.

Arriaza, M., Cañas-Ortega, J. F., Cañas-Madueño, J. A., & Ruiz-Aviles, P. (2004). Assessing the visual quality of rural landscapes. *Landscape and Urban Planning, 69*(1), 115-125. doi:http://dx.doi.org/10.1016/j.landurbplan.2003.10.029

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.

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Borland, H.P., Schlacher, T.A., Gilby, B.L., Connolly, R.M., Yabsley, N.A. & Olds, A.D. (2017). Habitat type and beach exposure shape fish assemblages in the surf zones of ocean beaches. *Marine Ecology Progress Series*, 570, 203-211.

Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion, 28*(3), 221-243.

Chatfield, K., Simonyan, K., Vedaldi, A. & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. In *British Machine Vision Conference*. Available (02/11/17) arXiv:1405.3531.

Cracknell, D., White, M. P., Pahl, S., & Depledge, M. H. (2017). A preliminary investigation into the restorative potential of public aquaria exhibits: a UK student-based study. *Landscape Research, 42*(1), 18-32. doi:10.1080/01426397.2016.1243236

Dalal, N. & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection, Proc. IEEE Conf. Computer Vision and Pattern Recognition.

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.

Dinsdale, E. (2009). Linking ecological and perceptual assessments for environmental management: a coral reef case study. *Ecology and Society, 14*(2), 1–17.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1), 98–136.

GBRMPA (2014). *Outlook Report 2014*. GBRMPA: Townsville. Available (24/09/17) http://elibrary.gbrmpa.gov.au/jspui/handle/11017/2855

GBRMPA (2017). *Current conditions on the Reef*. Update 5: 12 April 2017. Available (22/05/17) http://www.gbrmpa.gov.au/managing-the-reef/threats-to-the-reef/climate-change/marine-park-management/current-conditions-on-the-great-barrier-reef

Gilby, B.L., Olds, A.D., Connolly, R.M., Yabsley, N.A., Maxwell, P.S., Tibbetts, I.R., Schoeman, D.S. & Schlacher, T.A. (2017). Umbrellas can work under water: using threatened species as indicator and management surrogates can improve coastal conservation. *Estuarine, Coastal and Shelf Science*, 199,132-140.

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

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. Available (09/10/17) arXiv preprint arXiv:1408.5093.

Johnston, C. & Smith, A. (2014). Beautiful one day: Assessing the world heritage aesthetic values of the Great Barrier Reef. *Historic Environment*, 26(2), 54-71.

Johnston, C., Smith, A. & Dyke, J. (2013). *Defining the Aesthetic Values of the Great Barrier Reef World Heritage Area*. Prepared for the Department for Sustainability, Environment, Water, Population and Communities. Available (12/08/17) https://www.environment.gov.au/system/files/resources/7524507e-37a6-4cff-9195-cb5fe19d8f52/files/gbr-aesthetic-values.pdf

Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T. & Downing, J.A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *The Ecological Society of America*, 13(2), 76-81.

Lothian, A. (1999). Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning, 44*(4), 177-198. doi:http://dx.doi.org/10.1016/S0169-2046(99)00019-5

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

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion, 6*(2), 257-268.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*(3), 219-226.

Ren, S., He, K., Girshick, R. & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of Advances in Neural Information Processing Systems* (NIPS).

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* (IJCV), 115(3), 211–252.

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J. & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, 14(9), 570–585.

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

Scott, N., Le, D., Zhang, R., & Moyle, B. (2017 Online). A review of eye tracking research in tourism. *Current Issues in Tourism.* https://doi.org/10.1080/13683500.2017.1367367

Scott, N., Le, D, Becken, S., and Connelly, R. (2018 Submitted) Measuring perceived beauty of the Great Barrier Reef using eye tracking. *Journal of Sustainable Tourism*.

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

Silayoi, P., & Speece, M. (2007). The importance of packaging attributes: a conjoint analysis approach. *European Journal of Marketing, 41*(11/12), 1495-1517. doi:10.1108/03090560710821279

Simola, J., Le Fevre, K., Torniainen, J., & Baccino, T. (2015). Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials. *Neuroimage, 106*, 21-33. doi:http://dx.doi.org/10.1016/j.neuroimage.2014.11.030

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, 1409/1556, 2014.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going deeper with convolutions. *CoRR*, abs/1409.4842.

The Nature Conservancy (2017). Mapping Ocean Wealth. Available (12/05/17) http://oceanwealth.org/

Tourism Australia (2015). Research underpins new push to promote Australia’s aquatic and coastal experiences. Available (08/04/16) http://www.tourism.australia.com/documents/TASI9055_Coastal_Factsheet_DestAusCon-V6.pdf

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

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

Reef & Rainforest
RESEARCH CENTRE

www.nesptropical.edu.au