Image Memorability
Helping Alzheimer's patients

António Carlos Lima
DEI – Departamento de Engenharia Informatica
da Universidade de Coimbra
FCTUC – University of Coimbra,
3030-290 Coimbra, Portugal
aclima@student.dei.uc.pt

João Miguel Simões
DEI – Departamento de Engenharia Informatica
da Universidade de Coimbra
FCTUC – University of Coimbra,
3030-290 Coimbra, Portugal
jbsimoes@student.dei.uc.pt

***Abstract*—The aim of this project was to provide Alzheimer's patients help by facilitating the remembering the main events at the end of any given day. This said, it is unthinkable to just show a huge amount of pictures to these patients and expect them to even retain any of them, this calls for a special pre-selection mechanism, a filter. In this sense, throughout this paper we will discuss the tool that we came up with to filter images based on how memorable they appear to be. As such, we based our knowledge on previous studies and research material and present to you a different and potentially more advantageous approach.**

Keywords: Image memorability, Alzheimer's disease, Artificial Neural Networks, Features, Objects

I. Introduction

As has been known throughout the ages, we are continuously bombarded by visual information, some of which we can deeply care for and other that we simply discard into the void that is forgetfulness. Some attempts have been made to try and understand *what makes an image memorable* and if it can be reproduced artificially. For some individuals, e.g. Alzheimer's patients, it's rather difficult and in some cases nearly impossible to select which images are more or less important. A tool that could achieve a filter similar to that which most of us use, would greatly improve their lives.

Previous research on this topic identified several characteristics that influence memorability. However, they all focused on particular aspects such as the contribution of colour, objects statistics and simple features.

Through other works [*Oliva et al.* [1]] , we can gather a considerable amount of mapped inputs and outputs for different approaches to how image memorability might, to some extent, work, i.e. we now have a rudimentary and incomplete gist of how memorable some visual information is or isn't.

This in turn leads us to believe that a neural network might be an efficient classifier for this problem albeit, this approach can only surface now that we have the above mentioned input-output pairings.

The remainder of this paper is our attempt at reconciling various findings and results of other researchers in order to better understand image memorability, in turn helping Alzheimer's patients in an ability that they have lost: *remembering*.

II. State of the art

The problem with most approaches to image memorability so far is that they require too vast amounts of time to process and generate results. Our aim was to make the most of what has already been accomplished, namely already classified images from *Oliva's et al.* [1] datasets and try to improve upon conclusions for faster results. Since an Alzheimer's patient can't go through nor remember all 5000 pictures taken with a Microsoft SenseCam like device at the end of the day, a fast and generalised solution should be applied in order to filter out the least memorable images and re-focus the patients attention on those that are actually memorable.

Thus, we came up with the idea of training a couple of Neural Networks to achieve such objective. Since Neural Networks are good at generalising and can assess thousands of inputs very fast they were our optimal candidate for the problem. They also possess the remarkable ability to stored after having been properly trained and retain their knowledge.

III. Agent's architecture

As we mentioned before, our work consisted in making the most out of neural networks to determine the memorability of a certain image. Therefore, we decided to implement an agent composed by two neural networks. The first one is responsible for detecting the *features/objects* that might comprise an image. The second one uses the detected features in order to predict the *memorability* of the image. Figure 1 is a visual representation of the agent's architecture. Our agent's purpose is to help Alzheimer's patients remembering the most important events that occurred throughout the day.

The input of the *Feature Detection network* consists of grayscale images with 256x256 pixels. The output is a vector representing which objects/features are detected in the image. The *Memorability Classification network* receives the detected features and attributes a memorability value to the image.

IV. Memorability

In this section we will discuss image memorability, more specifically which characteristics make an image more or less

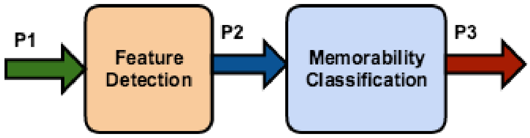

Fig. 1. Agent's architecture

memorable. In addition, we will explain a new approach to predict the memorability of a certain image.

A. Image Memorability

A lot of factors were investigated in order to understand the reason why some images are more memorable than others. For this purpose, the work of *Isola et al.* [1] studied the influence of colour, simple features, object statistics, object semantics, and scene semantics in image memorability. This way image characteristics like colour, simple features and object statistics are lowly correlated with human memory. However, object understanding is necessary to human picture memory [2]. In addition, they verified that objects without semantics are not likely to be memorable and that there is a substantially high correlation between object/scene semantics and human memory [2], [3] . For example, an image with a person is more memorable than an image of a landscape.

Based on these findings, it is possible to match an object and its memorability. Consequently, an image's predicted memorability can be determined by calculating the (positive or negative) contribution of each object.

B. A new approach to image memorability

In this work, we propose a different approach to predict memorability of a certain image, based on previous research [*Isola et al.* [1]] and the power of neural networks. For this purpose, we made use of the *Neural Network Toolbox for Matlab* to implement the classifiers.

The first step to achieve our classifiers consisted of gathering images and corresponding features from *Isola's et al.* [1] dataset. Next, we analysed which type of Neural Network would be more appropriate for feature detection and, after some experiments and research, we decided to use a *Fitting Network*. Fitting Networks are Feed-forward neural networks used to fit an input-output relationship.

Following up, the *Memorability classification network* was trained using image feature dataset [*Isola et al.* [1]] as input and the corresponding image memorability score as target. Thus, we achieve an association between the detected features and an image's memorability. It should be noted that the images' memorability scores were obtain from classification performed with the SVM [Isola et al. (data do paper)].

The dataset used on this work is composed by 2400 images obtained from Isola et al. (data do paper). Although, only 2222 are labelled with the corresponding features. Our Agent's architecture can now be filled with more concrete information:

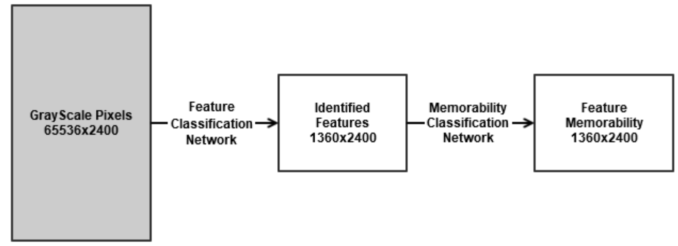

Fig. 2. Explanation of a complete Image Memorability Test

V. Experiments

In the current section we will present the experiments performed with the agent described before. All obtained results will be presented and discussed in this section.

A. Experimental Setup

The first step of our experimented consisted on construct the dataset to be used in each neural network, based on experimental data and results from *Isola et al.* [1].

We started by gathering a set of images and the corresponding features diving it into training, validation and test sets. Also, we considered an universe of 1360 distinct objects that can be identified in an image.

The next step, consisted on training a neural network to be able to identify which objects are present on a given image. With 2400 images of which 2222 are fully pre-classified (features and memorability) we could now proceed by dividing these into training and validation sets. We split the pre-classified set of images into two 70-30% sets for training and validation, respectively. We also possessed a set of 8000 images of which 2400 were used for testing our classifiers.

The Feature Classifier's output is a matrix with 1360 rows by N columns, N being the number of images. Each cell indicates how many times that feature was detected in the i^{th} image.

The Feature Classifier's input is a matrix with 256x256 rows by N columns, N being the number of images. Originally, the images had a RGB colourmap but that proved too heavy for our computers to process, so we decided to convert each one to their grayscale counterpart reducing the size of training data set to about one third. We believe that converting each image from RGB to grayscale doesn't significantly affect the process of object detection and permits a serious decreasing of the training time, but we can't test this theory with our current computers.

As mentioned previously, we studied which type of Neural Network would be more suitable for this classification problem and concluded that a *Fitting Network* was a good choice. Another issue that became apparent was the number of hidden layers and neurons. For most of non-linearly separated problems, one hidden layer is enough and more than one tend to only slow down and over-fit the results so we decided to use only one. Regarding the number of neurons in each hidden layer we decided to use the minimum number of neurons to represent our features:

$$\lceil log_2(features) \rceil$$

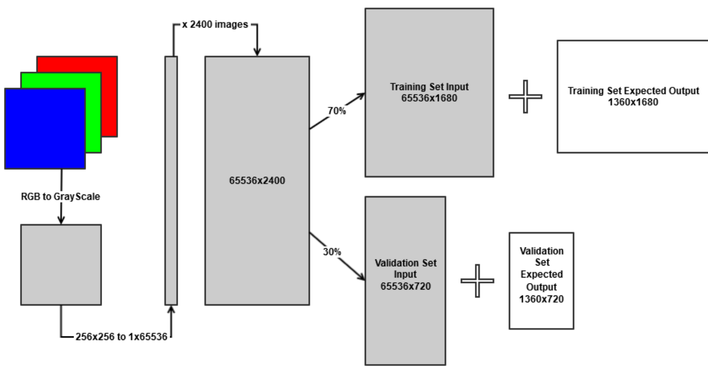

Fig. 3. Explanation of Image Decomposition and Feature Detection

B. Results and Discussion

In this section we will show the results we obtained after running our tests. We'll start with the results portraying to the feature/object detection and compare them to *Isola et al.* [1] original results for the same datasets. Afterwards, we'll analyse the memorability classifier's results.

1) Feature detection: After training the network to detect features we validated said training and compared the results with their corresponding outputs by analysing their performances and similarities. We soon realised that the act of simply counting the number of times a feature was present was too complex of a task for such a small dataset, so we decided to include, for comparison reasons, a duplicate and identical setup for each of our experiments but in which we merely trained the *Feature Classification Network* to detect the presence of features instead of counting them. As we can observe from Fig. 4 and Fig. 5, the results became vastly better in all aspects. We also tried to better understand the results we were achieving and processed the data by means of rounding its values after it had been classified.

2) Memorability Classification: The obtained memorability classification results were acceptable. Fig. 6 shows two of the most memorable images classified by the neural network. We can easily see high semantics value objects, for example *"human figure"*. This result meets the work of Isola et all[*]. Fig. 7 shows the least memorable images classified by the system. Is possible to observe low semantic objects as for example a landscape. These results are what we expected and coincide with what had been previously said about object semantics. However, we noticed that there were some outliers in our results, namely images with some high value semantic objects were being classified as having low memorability. One other interesting aspect is that the numeric memorability results

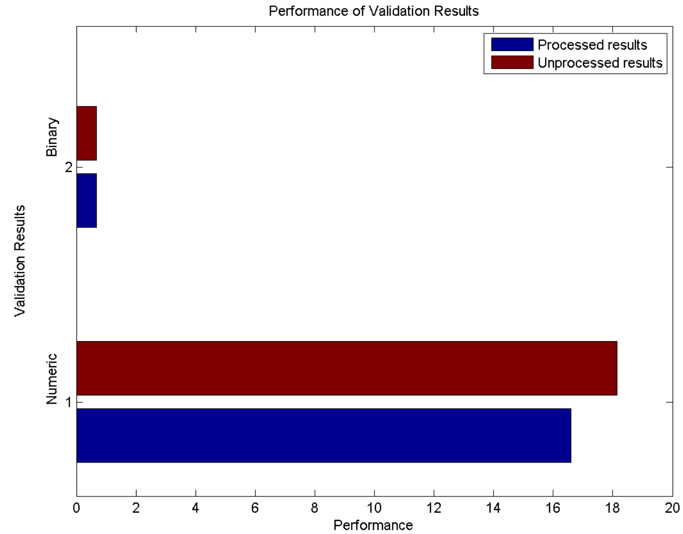

Fig. 4. Feature Detection Network Performances

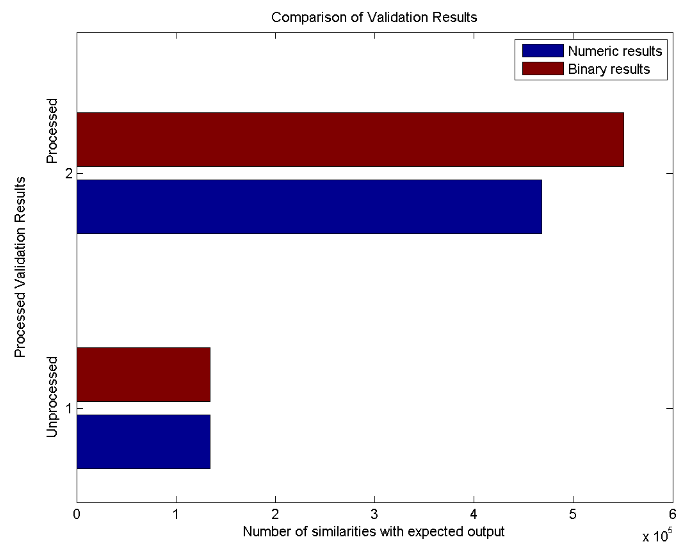

Fig. 5. Feature Detection Network Similarities

seems to be more forgiving than their binary counterpart, that is, the numeric results only have positive values whilst the binary results quite a few (very) negative values.

3) A pattern emerges: During the results' analysis we were surprised by an interesting occurrence. A pattern could be viewed in the distribution of the detected features for the most memorable images (Fig. 10), but the same couldn't be said for the least memorable images (Fig. 11. Comparing the two figures we verified that the most memorable images have a common, more clear, profile. If we look at Fig. 10 the image is linear along the vertical axis for some ranges of features. Contrariwise, images of Fig. 11 aren't nearly as linear along the vertical axis. We believe that this is an interested outcome that should be studied in a future work, especially how a broader dataset might influence said pattern's linearity.

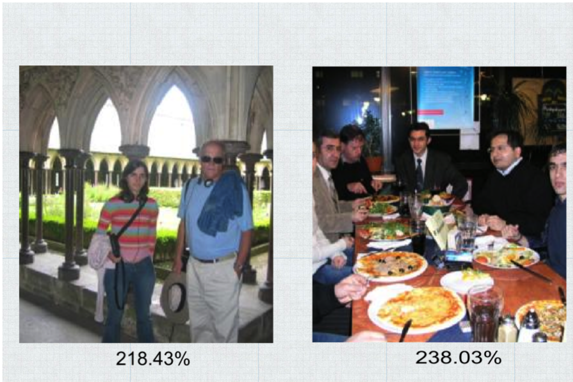

Fig. 6. Most memorable images classified by neural network

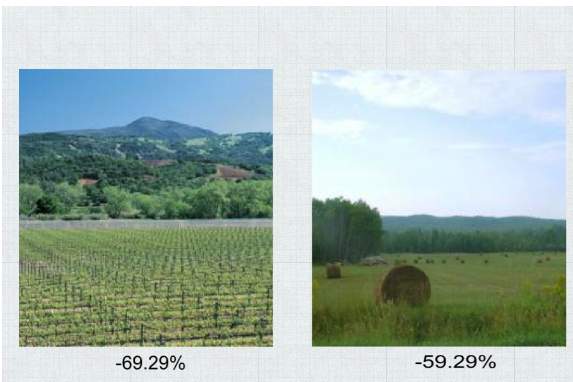

Fig. 7. Least memorable images classified by neural network

VI. Conclusions

As seen before, the image classification's performance decreases when an approach based on neural networks doesn't have a vast enough dataset. In fact, this is an expected result since neural networks have a lower classification performance compared to SVMs. Nevertheless, neural networks are more efficient when it comes to time, in fact our complete process (training, validation and testing), albeit resource demanding, was faster than the other projects that we based our inputs on. Another advantage of neural networks is their generalising capability. Unfortunately, *Isola et al.* [1] only classified the memorability of 1111 images restricting the number of samples we could use on our second half of the training process. We believe that with a larger amount of classified samples our performance would increase significantly and our results would have substantially better quality.

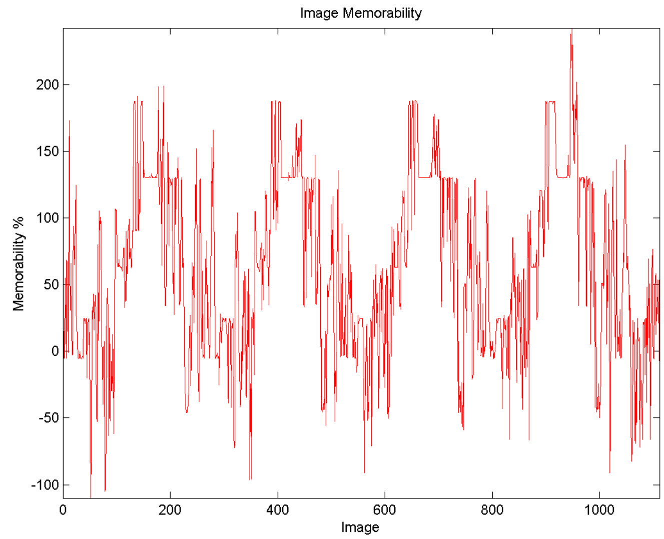

Fig. 8. Binary Image Memorability Results

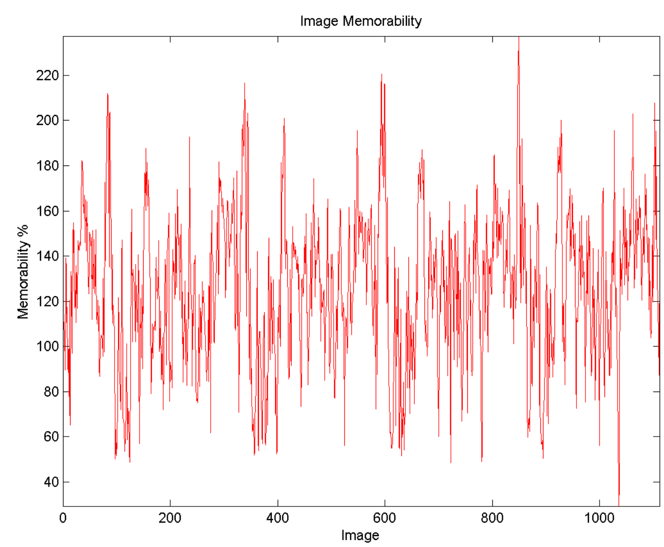

Fig. 9. Numeric Image Memorability Results

References

[1] P. Isola, J. Xiao, A. Torralba, A. Oliva, What makes an image memorable?, IEEE Conf. Comput. Vision Pattern Recognit., p. 145-152, 2011. doi:10.1109/CVPR.2011.5995721

[2] W. Koutstaal and D. Schacter. Gist-based false recognition of pictures in older and younger adults. Journal of Memory and Language, 1997. 148, 149

[3] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Conceptual distinctiveness supports detailed visual long-term memory for realworld objects. JEP:G, 2010. 145, 148, 149

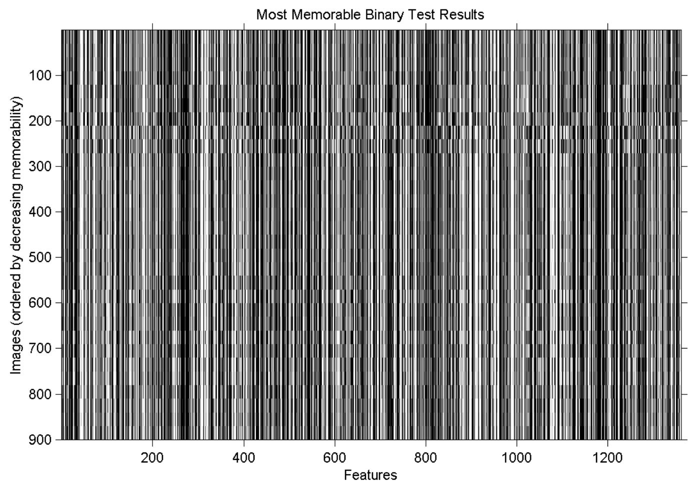

Fig. 10. Features distribution in the 30 most memorable images

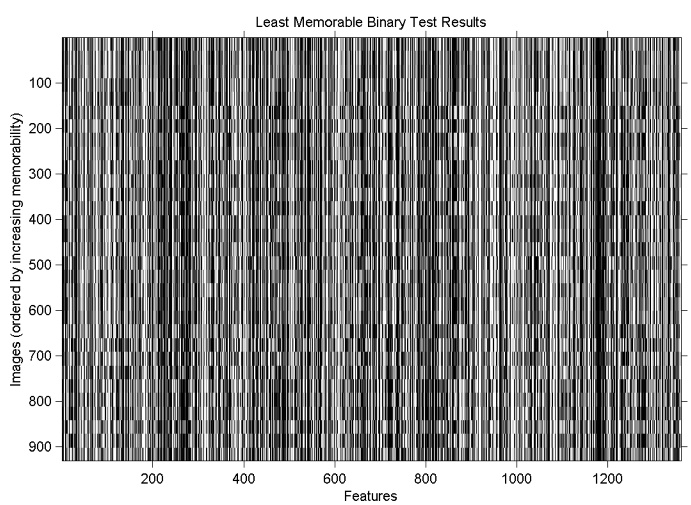

Fig. 11. Features distribution in the 30 least memorable images